الإحصاء الحيوى
وتصميم التجارب

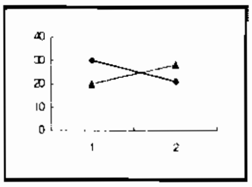
أ. د. عبد الحليم عشماوى
أ. د. صـــلاح جـــلال
أ. د. محمد حسين صادق

المكتبة الأكاديمية
شركة مساهمة مصرية
ISO 9002

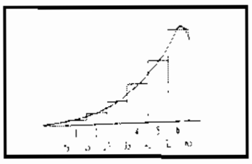
الإحصاء الحيوى

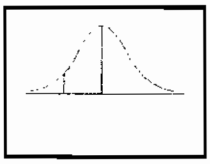
وتصميم التجارب

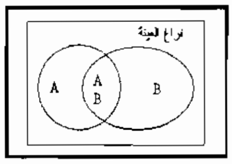
ANOVA



GLM



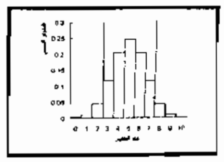




Σ

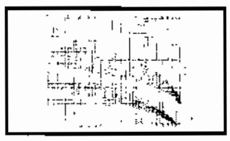
σ^2

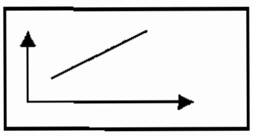
χ^2



μ

α



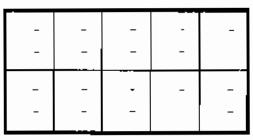


β

أ.د. عبد الحليم عشماوى

أ.د. صلاح جلال

أ.د. محمد حسين صادق



المكتبة الأكاديمية

شركة مساهمة مصرية

ولابد من ذكر أن نواة هذا المؤلف كانت فى كتاب آخر صدر فى ١٩٨٩شارك فيه اثنان من مؤلفى هذا الكتاب بجانب أ.د. عصام الطويل رحمه الله.

ويتقدم المؤلفون بخالص الشكر والتقدير إلى الأستاذ **أحمد أمين** رئيس مجلس إدارة المكتبة الأكاديمية على مجهوداته المخلصة في مجال النشر والمهندس **حمدي قنديل** مدير النشر فلهما خالص التقدير والثناء.

ويأمل المؤلفون أن يكونوا قد وفقوا لتقديم مادة للقارئ والباحث العربى تعينه على تفهم علم الإحصاء وتطبيقة تطبيقا سليما.

والله الموفق.

القاهرة فى يوليو ٢٠٠٨

المؤلفون

مع تحيات د. سلام حسين عويد الهلالي

https://scholar.google.com/citations?user=t1aAacgAAAAJ&hl=en

salamalhelali@yahoo.com

https://www.facebook.com/salam.alhelali

https://www.facebook.com/groups//Biothesis

https://www.researchgate.net/profile//Salam_Ewaid

07807137614

المحتويات

الصفحة

الباب ٥: المعاينة

الباب ٦: اختبارات الفروض

الباب ٧: اختبارات البيانات العددية: مربع كاى (χ^2)

الباب ٨: المقارنة بين متوسطين

الباب ٩: تحليل التباين أحادى الاتجاه

الباب ١٠: العلاقة بين متغيرين: الانحدار والارتباط البسيطان

الباب ١١: تحليل التباين متعدد التقسيمات

الباب ١٢: الافتراضات الخاصة بتحليل التباين

الباب ١٣: تحليل التباين ثنائى الاتجاه فى حالة عدم تساوى تكرار الفئات

الباب ١٤: الانحدار الخطى المتعدد

الباب ١٥: تصميم التجارب

ملحق أ

١ الإحصاء: تعريف ومبادئ

١- مقدمة

٢- طبيعة البيانات

٣- العشيرة (المجتمع) والعينة

٤- برامج الحزم الجاهزة للتحليل الإحصائى

١-١ مقدمة

يمكن تعريف علم الإحصاء بأنه "العلم الذى ينمى ويستخدم أكفأ الطرق فى تجميع وتبويب وترجمة البيانات الكمية بطريقة تمكن من معرفة حجم الخطأ فى الاستنتاجات والتقديرات واتخاذ القرارات فى حالات عدم التأكد وذلك بواسطة الفهم العلمى المبنى على أسس الاحتمالات"، وعلى هذا فإن الإحصاء هو أحد فروع الرياضيات التطبيقية.

وتعتبر الإحصاء فى نفس الوقت تكنولوجية الطريق العلمى، فهى بذلك تمنح الباحث السبيل والتقنيات التى يستعملها فى الوصول إلى استنتاجات معينة وتصميمات محددة من التجارب التى يجريها. وقد أوضح R. A. Fisher أن الإحصاء تشتمل على ثلاثة مجالات للدراسة وهى :

(١) دراسة العشائر (المجتمعات)

(٢) دراسة الاختلافات أو التباينات

(٣) دراسة طرق تلخيص البيانات

(١) **دراسة العشائر (المجتمعات):** فكرة العشائر لا تنطبق فقط على الأفراد أو الأشياء، فأى قياس أو مشاهدة تتكرر لا نهائياً فإن مجموع هذه النتائج المتحصل عليها تكون عشيرة (مجتمع) population من هذه القياسات. هذه العشائر هى مجال الدراسة الخاصة بنظرية الإحصاء، وأن حساب المتوسطات والخطأ القياسى هى محاولات لدراسة هذه العشائر.

(٢) **دراسة الاختلافات أو التباينات:** قيام الإحصاء بدراسة العشائر يؤدى مباشرة إلى دراسة التباين. فالعشيرة التى تتكون من أفراد متطابقة فى كل النواحى يمكن وصفها عن طريق وصف أحد هذه الأفراد فقط وعددها. والعشائر التى تقوم الإحصاء بدراستها دائما ما تظهر تبايناً فى إحدى النواحى أو أكثر، والإحصاء يدرس هذا التباين بدلا من أن يعتبره مصدرا لعدم الدقة يراد التخلص منه.

(٣) **دراسة طرق تلخيص البيانات:** وهذه عبارة عن دراسة طرق تلخيص البيانات المتوفرة عن العشائر فى عدد قليل من القيم ذات معنى تمكن من توصيف هذه العشائر، فمن المستحيل أن يحتفظ عقل أى باحث بجميع البيانات عن العشيرة أو حتى العينة التى يدرسها. وفى كثير من الأحيان فإنه يمكن توصيف العشيرة التى تنتسب إليها العينة.

١-٢ طبيعة البيانات

عند دراسة أى موضوع فإنه من الضرورى الحصول على معلومات خاصة به وذلك من خلال إجراء مشاهدات عليه وبالتالى الحصول على قدر من البيانات data (والمفرد datum) عن هذا الموضوع. وعادة ما تكون مجموعات البيانات المتاحة للتحليل الإحصائى مختلفة الأحجام sizes والأشكال shapes وبالتالى فإن طريقة تحليل هذه البيانات تختلف من حالة إلى أخرى حسب طبيعة هذه البيانات. ويمكن أن تقسم البيانات بعدة طرق منها أن تقسم حسب نوعيتها type إلى بيانات وصفية qualitative أو بيانات كمية quantitative، أو تقسم البيانات حسب طريقة قياسها scale of measurement إلى بيانات اسمية nominal أو بيانات رتبية ordinal أو بيانات فترية interval أو بيانات نسبية ratio. وهذه التقسيمات تعتمد فى النهاية على طبيعة المتغيرات محل الدراسة.

١-٢-١ تقسيم البيانات حسب نوعيتها Type of data

أ - البيانات الوصفية Qualitative data

وهى تلك البيانات التى تقع فى أقسام categories كما لو أنه قد قسمت المشاهدات إلى نوعين (مثلا حيوانات بيضاء أو سوداء) وفى هذه الحالة فإنه سوف يعطى عداً لأعداد الحيوانات من هذين النوعين (وقد يكون أكثر من نوعين).

ب - البيانات الكمية Quantitative data

وهى تلك البيانات الناتجة من قياسات أو تقديرات عددية. فالقياسات تعطى المتغيرات variables . وهنا يظهر نوعان من المتغيرات: الأول قد يمثل عدد الأفراد فى كل تركيب وراثى من مجموعة من التراكيب الوراثية وهذه لا تأخذ إلا قيما (أو وحدات) كاملة وهذه تعرف بالبيانات أو التوزيعات المتقطعة discrete أو غير المستمرة discontinuous، والنوع الآخر قد تكون البيانات فيه تمثل الاختلافات الملاحظة فى الأوزان لمجموعة من الحملان عند فطامها أو لإنتاج اللبن من الأبقار أو الدخل أو الدرجة التى يحصل عليها الطالب فى امتحان ما...الخ. الاختلافات هنا تكون مستمرة continuous ويعرف التوزيع للمتغير فى هذه الحالة بأنه مستمر.

١-٢-٢ تقسيم البيانات حسب طريقة قياسها Scale of measure

أ - البيانات الاسمية Nominal (categorical) data

وهى تلك البيانات الوصفية qualitative والتى تكون على هيئة أقسام ليس لها ترتيب order (rank) أو بناء structure معين أو معنى عددى، وبالتالى لا يمكن إجراء أى عمليات حسابية عليها سوى العد counting. ومن أمثلة تلك البيانات: الجنس gender (ذكور و إناث، والتى يمكن أن يرمز لها مجازا بالحروف M، F أو 0، 1 عند استخدامها فى التحليل الإحصائى)، اللون (مثلا أحمر، أخضر ...). وفى هذه الأمثلة لا يهم أن تكون الذكور أولاً يليها الإناث أو اللون الأحمر يليه الأخضر ... وهكذا.

ويمكن تحليل هذه النوعية من البيانات إحصائيا باستخدام المنوال mode أو مربع كاى χ^2 أو الإحصاءات اللامعلمية non-parametric statistics.

ب - البيانات الرتبية Ordinal data

وهى البيانات التى تم ترتيبها بحيث إن الرتبة الأعلى تمثل قيمة أعلى من قيمة الرتبة الأقل مع الأخذ فى الاعتبار أن الفرق بين قيم الرتب ليس بالضرورة أن يكون متساويا بمعنى أنها تعطى مفهوما أكبر من أو يساوى أو أصغر من فقط. ويمكن أن يطلق عليها أيضا أنها بيانات وصفية. ومثال ذلك أن تقسم أفراد العشيرة (المجتمع) طبقا لمستوى المعيشة إلى طبقات وفيها يمكن القول أن الأفراد الذين ينتمون إلى طبقة فوق المتوسط دخولهم أعلى من الذين ينتمون إلى الطبقة المتوسطة الدخل ولكن لا يمكن تقدير ذلك بطريقة كمية (فلا يمكن أن يقال إن الطبقة الأولى أعلى من الثانية بمقدار 20% مثلا).

ويمكن تحليل هذه النوعية من البيانات إحصائيا باستخدام الوسيط (الوسط) median أو المنوال mode أو ارتباط الرتب rank correlation أو الإحصاءات اللامعلمية non-parametric statistics.

ج - البيانات الفترية (فى فترات) Interval data

وهى تلك البيانات الكمية quantitative والتى يمكن ترتيبها وإجراء العمليات الحسابية عليها مثل الجمع والطرح والضرب والقسمة. ومن أمثلة هذه النوعية من البيانات التواريخ dates ودرجة الحرارة temperature. الفرق بين

وحدات هذه النوعية من البيانات له معنى ولكن النسبة ليس لها معنى، فمثلا فى بيانات درجات الحرارة temperature فإن

$10^{\circ} = 20^{\circ} - 10^{\circ} = 30^{\circ} - 20^{\circ}$ ، ولكن $20^{\circ}/10^{\circ}$ لا تعنى ضعف الحرارة (السخونة).

ويمكن فى مثل هذه النوعية من البيانات حساب المتوسط mean، الانحراف القياسى (المعيارى) standard deviation، الارتباط correlation البسيط والمتعدد، الانحدار regression البسيط والمتعدد، تحليل التباين analysis of variance، التحليل العاملى factor analysis. بمعنى أنه يمكن استخدام الإحصاءات المعلمية parametric statistics.

د - البيانات النسبية Ratio data

وهى تلك البيانات التى تنتمى للبيانات الكمية والتى يمكن وضعها على هيئة كسر اعتيادى، بسط ومقام بشرط أن المقام لا يساوى الصفر. وليس الترتيب ومقدار المسافة هما العنصرين المهمين فقط ولكن أيضا عندما تكون النسبة بين أى قياسين هى الهدف الأساسى للقياس. ومن أمثلة ذلك محيط الصدر والوزن فإنه من المنطقى القول إن 50 كج ضعف 25 كج، وهذه المقولة صحيحة بغض النظر عن وحدات الوزن إذا كانت كيلوجرام أو جرام، وهذا بسب وجود الصفر الطبيعى natural zero لهذه النوعية من البيانات. وتحتوى مجموعة البيانات النسبية على أعداد صحيحة موجبة أو سالبة أو مساوية للصفر أو كسر اعتيادى أو كسر عشرى أو نسبة مئوية. ويوجد عدد لا نهائى من الأعداد النسبية المحصورة بين أى عددين نسبيين. وتجرى جميع العمليات الحسابية على هذه النوعية من البيانات.
ويمكن عرض وتلخيص وتحليل هذه النوعية من البيانات إحصائيا بنفس الطرق السابق ذكرها فى البيانات الفترية.

١-٣ العشيرة (المجتمع) Population والعينة Sample

فى بعض الأحيان، ولكن نادراً ما يمكن قياس كل القيم التى تنتمى إلى موضوع معين، وفى هذه الحالة فإنه يتم التعامل مع ما يعرف بالعشيرة (المجتمع) population وهى تشمل جميع القيم التى تنتمى لوضع ما محدد بتوصيفات معينة. وعلى هذا الأساس فإن العشيرة تحتوى على كل القيم الخاصة بالمتغير الذى يمثل صفة ما أو قياسا معينا. وإذا أمكن قياس جميع هذه القيم فإن العشيرة من الممكن وصفها دون خطأ. وقد تكون العشيرة محددة finite مثال ذلك جميع الأفراد الناشئة من حدوث طفرة

جديدة لم تظهر مسبقا (مثلاً)، وبالتالى من الممكن عد وتوصيف جميع أفراد هذه العشيرة. أو قد تكون عشيرة غير محددة infinite وهو الأكثر شيوعا و يمثلها على سبيل المثال عشيرة الجاموس البرى فى أفريقيا أو نباتات القمح المزروعة فى مصر.

أما العينة sample فهى جزء من العشيرة (وقد تشمل العشيرة كلها). والغرض من استخدام العينة هو الحصول على استدلالات عن العشيرة. وعلى ذلك فإنه من المرغوب فيه أن تحدد العشيرة بقدر من الدقة حتى يمكن الحصول منها على العينة الممثلة والتى تعطى تقديرات سليمة. وفى هذا المجال فإن الأرقام التى تصف العشيرة تسمى بمعالم العشيرة population parameters ومثال هذه المعالم المتوسط الحسابى (μ) وهو المتوسط الذى يدخل فى تحديده كل القيم التى تنتمى إلى هذه العشيرة. ومن ناحية أخرى فإن الأرقام المقابلة والتى تصف أو تحسب من العينة تعرف بالإحصاءات أو التقديرات estimates or sample statistics مثل متوسط العينة $\overline{X}$ وهى تعتبر تقديرات لمعالم العشيرة.

وتبحث الإحصاء فى العلاقة الرياضية بين الإحصاءات والمعالم فى صورة احتمالية بمعنى أنها تحدد مدى الثقة فى أن يكون الإحصاء المحسوب من العينة تقديرا لمعلم العشيرة.

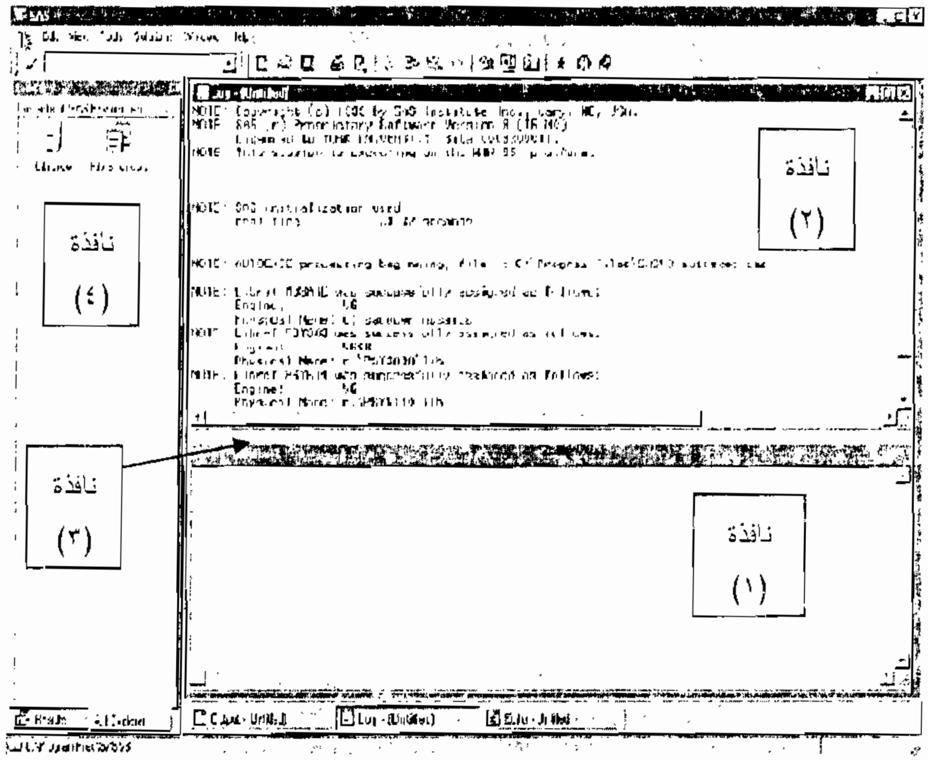
١-٤ برامج الحزم الجاهزة للتحليل الإحصائى

Ready packages for statistical analysis

يوجد عدد كبير من حزم البرامج الجاهزة والتى يمكن استخدامها فى التحليل الإحصائى ومنها برنامج SAS وبرنامج SPSS وهما الأكثر شيوعا مقارنة بالحزم الأخرى. وهذه الحزم بعضها متاح مجانا أو بتكلفة بسيطة وبعضها مكلف. وفى هذا الجزء سوف يتم عرض شرح مختصر ومبسط عن كيفية استخدام كل من البرنامجين SAS و SPSS فى التحليل الإحصائى.

١-٤-١ برنامج SAS (Statistical Analysis System)

عند بداية تشغيل برنامج SAS يظهر أربع نوافذ رئيسية (وهناك نوافذ أخرى يمكن للمستخدم فتحها). الشكل ١-١ يبين النوافذ الرئيسية لبرنامج SAS



شكل ١-١ النوافذ الرئيسية لبرنامج SAS

نافذة (١) نافذة تحرير البرنامج Program editor window وتستخدم هذه النافذة لكتابة خطوات البرنامج وقد تحتوى أيضا على البيانات data المراد تحليلها.

نافذة (٢) نافذة المسجل Log window ويظهر بهذه النافذة رسائل الملاحظات والأخطاء التى قد تكون فى خطوات البرنامج المكتوب بواسطة المستخدم. ويجب أن يتأكد المستخدم من عدم ظهور أى أخطاء بهذه النافذة.

نافذة (٣) نافذة المخرجات Output window ويظهر بها نتائج التحليل بعد تشغيل البرنامج الذى كتبه المستخدم.

نافذة (٤) نافذة المستكشف Explorer window ويظهر بها مكتبة برنامج SAS ويمكن بها عمل تحرير للمخرجات قبل طباعتها.

ويمكن للمستخدم فتح أى من هذه النوافذ (أفقيا أو رأسيا) أو غلقها أو إعادة فتحها حسب الرغبة.

يتميز برنامج SAS بوجود بعض المفردات vocabulary والجمل المركبة بطريقة معينة syntax خاصة به، يقوم المستخدم لهذا البرنامج بإعداد البيانات المطلوب تحليلها والسؤال المطلوب الإجابة عليه من هذه البيانات بطريقة معينة على هيئة جمل statements وخطوات باستخدام لغة SAS التى تعتمد على كلمات مفتاحية keywords ويتم وضع كل ذلك بترتيب معين. وتقسم الجمل والخطوات التى توضع فى برنامج SAS إلى مجموعتين:

١ – مجموعة خطوات البيانات Data steps

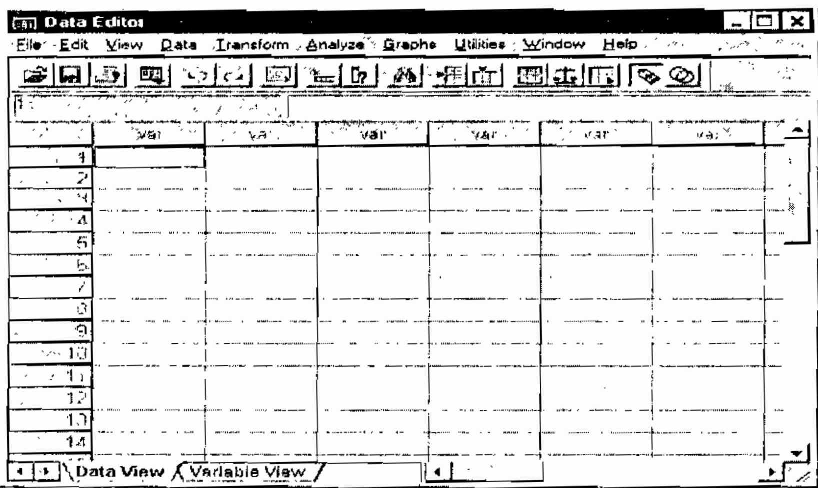
٢ – مجموعة خطوات التحليل Proc steps (Procedure steps)

يوجد داخل كل مجموعة من هاتين المجموعتين جمل لابد من وضعها لكى يعمل البرنامج بطريقة صحيحة وجمل أخرى اختيارية لا تؤثر فى عمل البرنامج.

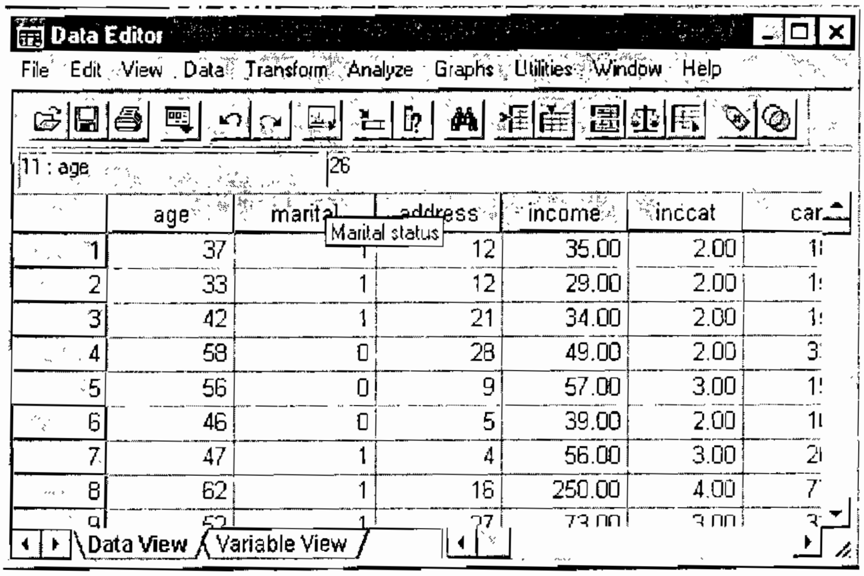
سوف يتم فى الأبواب التالية شرح كيفية استخدام هذا البرنامج تبعا للأمثلة المعروضة فى كل باب.

١-٤-٢ برنامج SPSS (**S**tatistical **P**ackage for **S**ocial **S**ciences)

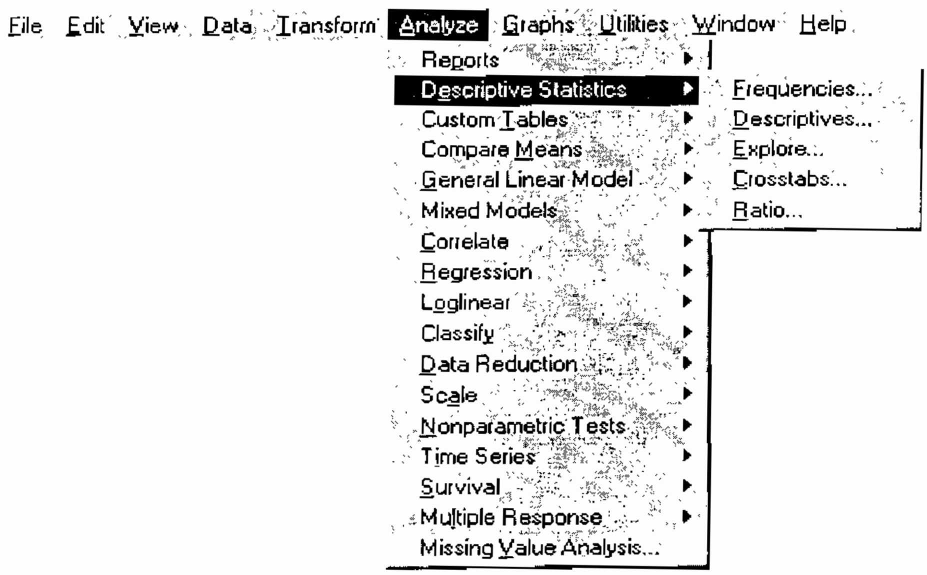
يختلف برنامج SPSS عن برنامج SAS بوجود العديد من النوافذ والقوائم المنسدلة فى SPSS والتى تسهل من طريقة الاستخدام حيث إن المستخدم ليس فى حاجة لكتابة خطوات البرنامج كما فى برنامج SAS. هذا بالإضافة لوجود نافذة لإدخال البيانات والمتغيرات كتلك النافذة الموجودة ببرامج الجداول الإلكترونية (مثلاً برنامج EXCEL). عند بداية تشغيل البرنامج تظهر النافذة التالية:



يمكن البدء فى إدخال البيانات بنفس طريقة الجداول الإلكترونية. ويمكن التحكم فى أسماء المتغيرات ومواصفاتها عن طريق فتح نافذة Variable View أسفل الشاشة. بعد الانتهاء من إدخال البيانات تظهر الشاشة التالية:

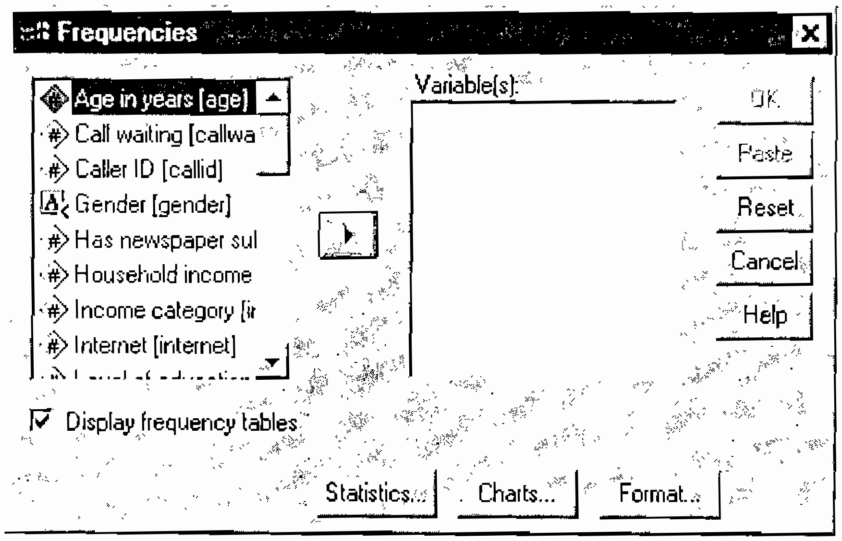


بعد التأكد من صحة إدخال البيانات يمكن البدء فى التحليل الإحصائى عن طريق القوائم المنسدلة من اختيار Analyze فتظهر الشاشة التالية:

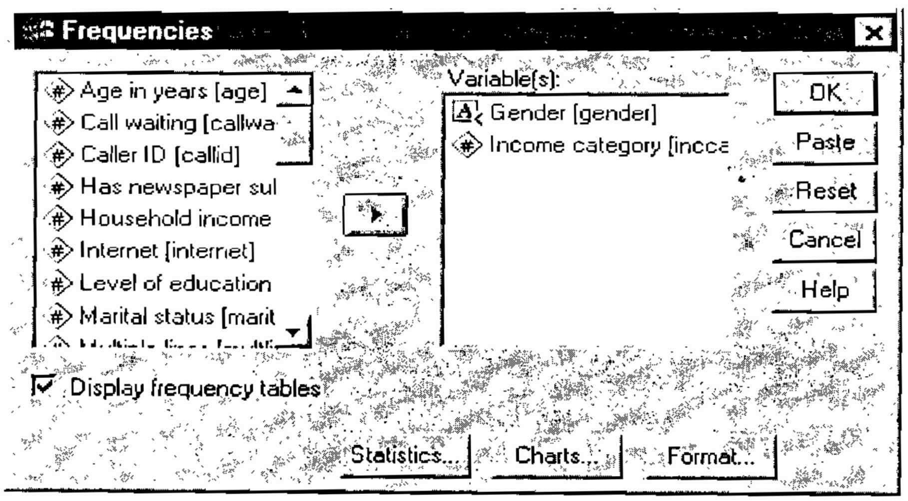


هذه الشاشة تحتوى على جميع التحليلات الإحصائية التى يمكن عملها بواسطة هذا البرنامج

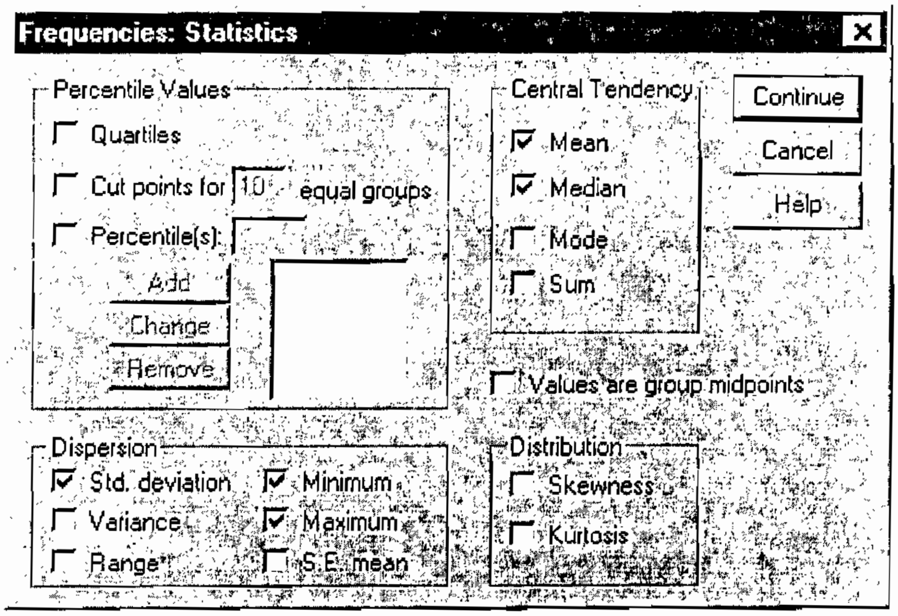
على سبيل المثال عند اختيار Descriptive Statistics ثم Frequencies تظهر شاشة أخرى كالتالى:



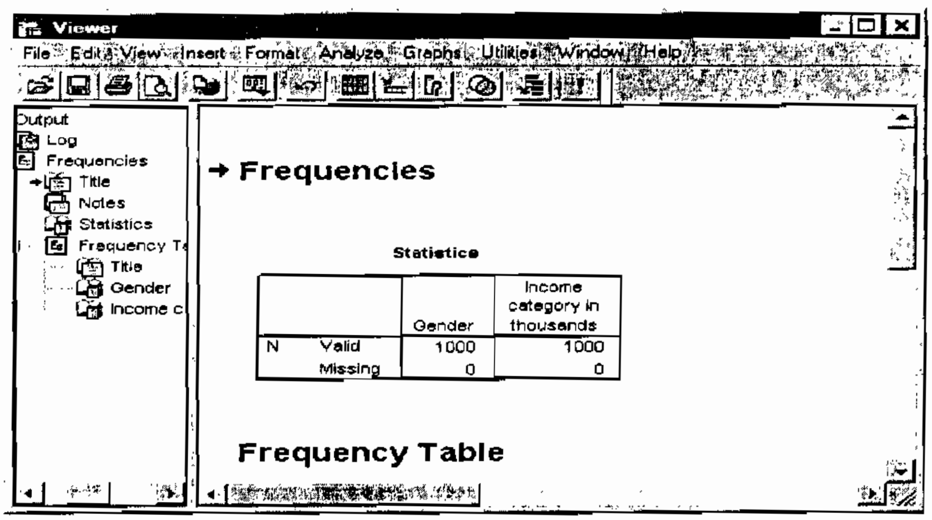
من قائمة المتغيرات يمكن اختيار المتغيرات المرغوب فى عمل تكرارات لها كالتالى:

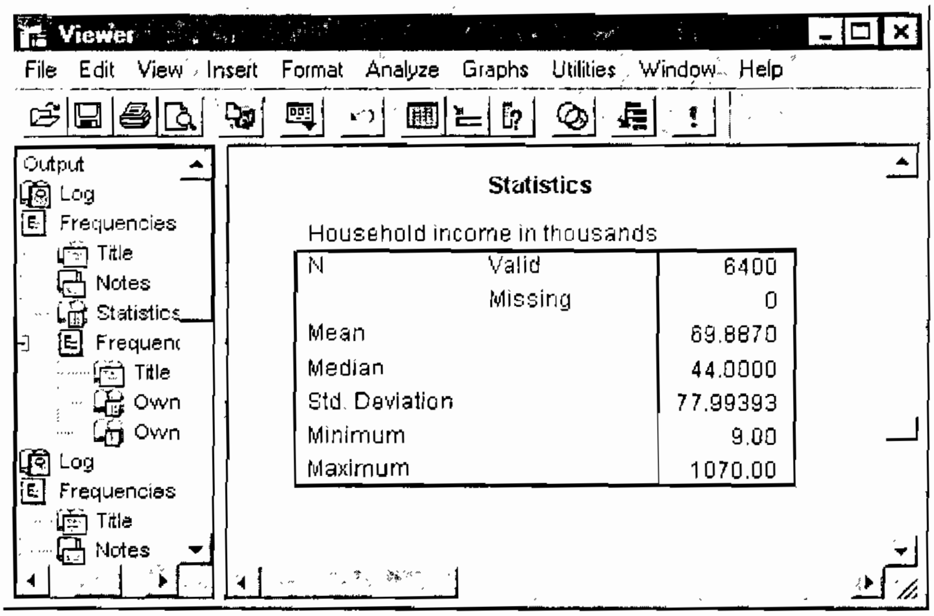


يمكن تحديد الخيارات المطلوبة للتكرارات عن طريق شاشة Statistics فتظهر الشاشة التالية:

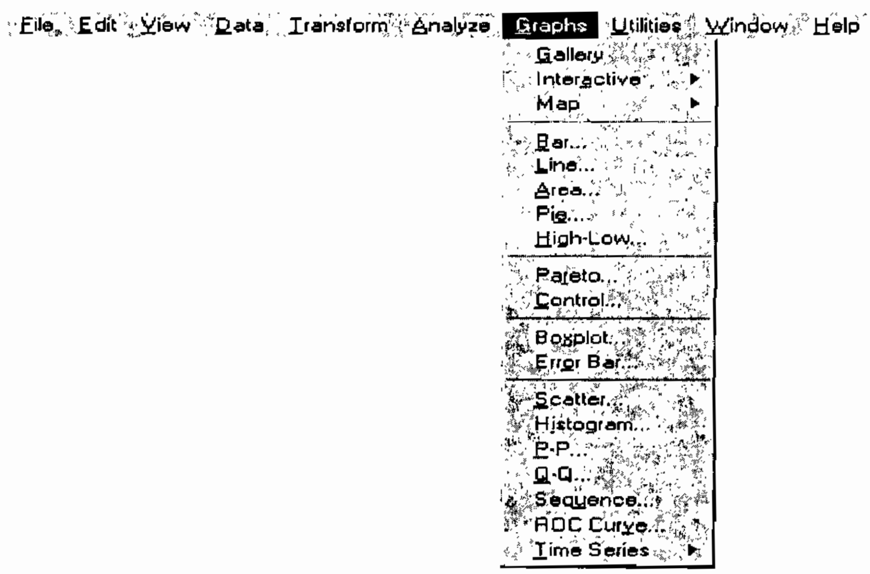


عند الانتهاء من تحديد الخيارات المطلوبة تظهر النتائج كالتالى:

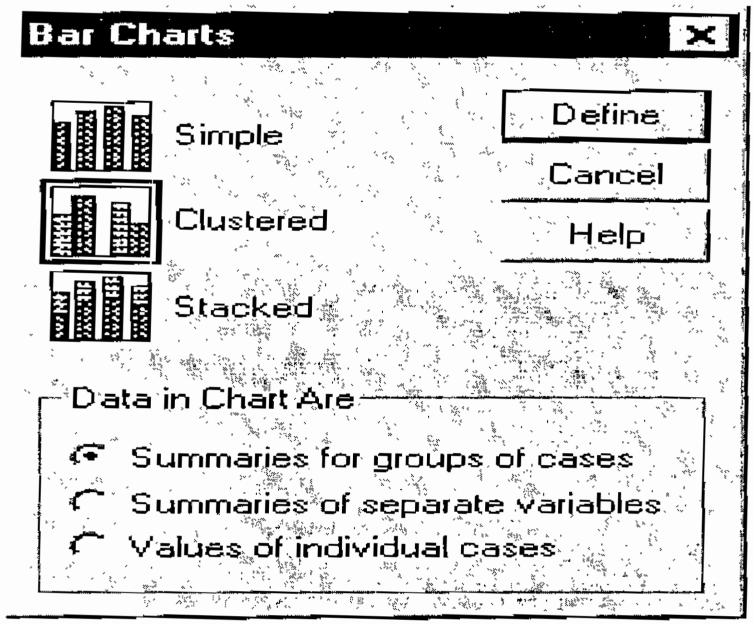




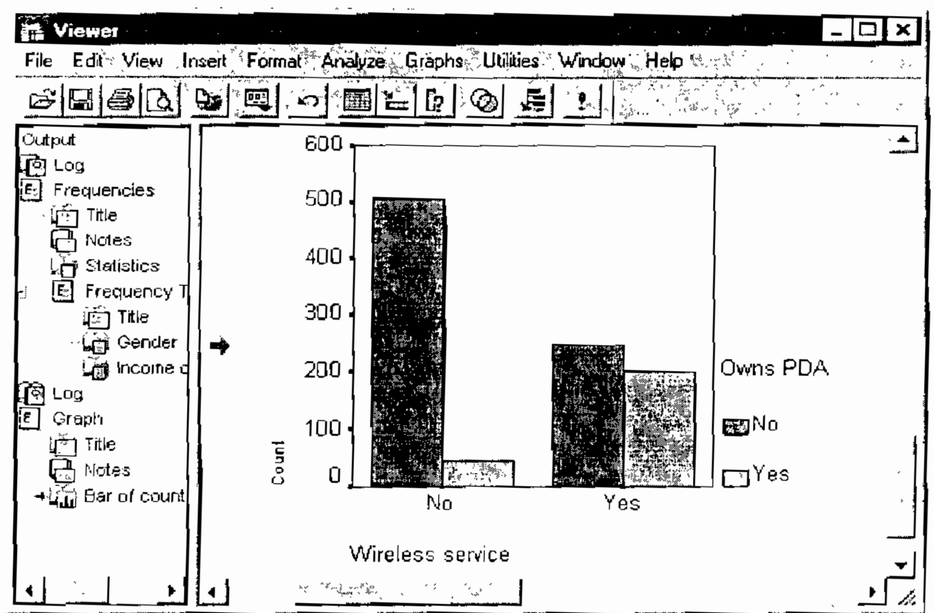
وعند الرغبة فى عمل أشكال توضيحية كوسيلة لعرض البيانات يتم اختيار نافذة الأشكال البيانية Graphs كالتالى:



عند اختيار شكل الأعمدة تظهر الشاشة التالية:



وبعد الانتهاء من الخيارات يظهر الشكل التالي:



مع تحيات د. سلام حسين الهلالي salamalhelali@yahoo.com

توضح الأمثلة السابقة أنه يمكن عمل الكثير من التحليلات الإحصائية باستخدام برنامج SPSS بطريقة سهلة للمستخدم.

ومن المفيد ملاحظة أنه يمكن بسهولة استخدام أحد برامج الجداول الإلكترونية مثل EXCEL فى تجهيز البيانات المطلوب تحليلها وتخزينها فى ملفات مستقلة ثم فتح هذه الملفات بواسطة SAS أو SPSS بدون الحاجة إلى إعادة إدخال هذه البيانات مرة أخرى.

٢ عرض البيانات Presentation of Data

١- مقدمة

٢- العرض الجدولى

٣- العرض البيانى

٢-١ مقدمة

غالبا ما يتوافر لدى الباحث كم من البيانات data تم الحصول عليها بعدة طرق مثل العد enumeration أو الحصر survey (والتى تعتمد أساسا على إعداد قوائم استبيانات questionnaire للإجابة عليها من الأفراد موضع الدراسة)، وقد تكون البيانات عبارة عن عمل تعداد census (مثل تعداد السكان أو تعداد عشيرة حيوانية معينة) وقد تجمع البيانات من سجلات records أو بإجراء التجارب experiments وتسجيل بياناتها ونتائجها وغير ذلك من الطرق. وهذه المجموعة من البيانات قد يصعب الاستفادة منها إلا إذا تم تبويبها التبويب المناسب. ومن ذلك التبويب الجغرافى كأن تقسم البيانات تبعا لمواقعها الجغرافية، أو التبويب الزمنى وفيه تقسم البيانات حسب وحدة زمنية معينة كتصنيف الطلبة طبقا لسنوات تخرجهم أو تصنيف الحيوانات تبعا لموسم الولادة صيفاً أو شتاء. وقد يكون التبويب وصفيا كأن تصنف البيانات طبقا لصفة محددة كاللون أو مستوى التعليم (أمى، أساسى، ثانوى، جامعى وفوق الجامعى). أما التبويب الكمى ففيه تقسم البيانات إلى فئات كتقسيم الأفراد حسب الوزن أو الطول (أقل من 50 كج، ومن 50 – 60 كج و أكثر من 60 كج، مثلاً) ويمكن أن تعرض البيانات بعد تبويبها إما فى صورة جداول tables أو فى صورة بيانية charts أو قد يستخدم العرض التصويرى pictorial presentation. وقد كان لتطور الحاسب الشخصى (PC) وحزم البرامج الجاهزة ready software packages مثل Excel و PowerPoint دوراً كبيراً فى هذا الصدد. كما قد تعرض البيانات فى صورة قيم تسمى بالإحصاءات statistics كأن يعبر عن مجموعة من الأرقام بالمتوسط الحسابى مثلا أو بالانحراف المعيارى وسوف يتم تناول ذلك فيما بعد. وقد تعرض البيانات بطريقة أو أكثر من الطرق سالفة الذكر.

٢-٢ العرض الجدولى Tabular presentation

هو أحد الطرق لعرض البيانات ومن أكثرها انتشاراً. ويتم وضع البيانات وعرضها جدولياً بعد جمعها وتلخيصها وتصنيفها حتى يمكن تتبعها واستيعابها. وعندما تحتوى العينة على عدد كبير من المشاهدات سواء كانت البيانات متقطعة أو مستمرة فإنها توضع فى جدول يبين التكرارات لكل قيمة عددية numerical value أو لكل فئة class. والجداول التى تعرض فيها البيانات إما بسيطة وهى التى تحتوى على عمودين فقط أو مركبة وتشتمل على أكثر من عمودين. والجداول تختزل فيها كمية البيانات المتاحة إلى صورة أكثر فهماً كما أنها ضرورية ولازمة لتمثيل البيانات بيانياً ومنها أيضا يمكن حساب الإحصاءات المختلفة كالمتوسط مثلا بجهد أقل عما لو استخدمت البيانات فى صورتها الأصلية ولاسيما فى عهد ما قبل الحاسبات الإلكترونية. يجب أن تتسم الجداول بالسهولة والوضوح ويكون لكل منها عنوان

واضح وقصير وموجز يوضح للقارئ نوعية البيانات وأحيانا كيفية تقسيمها، يجب أن تتسم عناوين الأعمدة بالوضوح. ويمكن وضع ملاحظات تذييلية footnotes لتوضيح أى رموز أو اختلاف فى الجدول، هذا بالإضافة إلى ملاحظة خاصة بمصدر البيانات source.

٢-٢-١ الجدول التكرارى frequency table للبيانات الوصفية (المتقطعة)

مثال ٢-١

يمثل جدول ٢-١ توزيع الجاموس فى بعض بلدان العالم كمثال على البيانات المتقطعة.

جدول ٢-١ توزيع الجاموس فى بعض بلدان العالم (العدد بالألف رأس)

البلد	العدد	البلد	العدد
بنجلاديش	850	إيران	550
البرازيل	1,095	إيطاليا	205
بلغاريا	9,737	باكستان	62,300
الصين	22,745	الفلبين	3,327
مصر	3,920	تايلاند	1,770
الهند	98,000		

المصدر: منظمة الأغذية والزراعة (FAOSTAT, 2007)

يلاحظ من هذا الجدول أن الهند تحتل المرتبة الأولى من حيث تعداد الجاموس بها تليها باكستان. وبالتالى فطريقة عرض البيانات تمكن القارئ من استيعاب بعض المعلومات بطريقة سريعة.

مثال ٢-٢

يمثل جدول ٢-٢ تعداد السكان فى بعض البلدان العربية حسب ما جاء بمنظمة الصحة العالمية لعام ٢٠٠٥.

جدول ٢-٢ تعداد السكان فى بعض البلدان العربية (بالألف نسمه) لعام ٢٠٠٥

الدولة	عدد السكان	الدولة	عدد السكان
مصر	74,033	ليبيا	5,853
السودان	36,233	الأردن	5,703
المغرب	31,478	الإمارات	4,496
العراق	28,807	لبنان	3,577
السعودية	24,573	الكويت	2,687
اليمن	20,975	عمان	2,567
سوريا	19,043	قطر	813
تونس	10,102	البحرين	727

المصدر: منظمة الصحة العالمية (WHO, 2007)

يلاحظ من الجدول أن مصر بها أكبر تعداد سكانى بين الدول العربية والبحرين هى أقل دولة تعداداً.

٢-٢-٢ العرض الجدولى للبيانات المستمرة

وفيها يقسم المدى range (وهو يمثل الفرق بين أكبر قيمة وأقل قيمة) إلى عدد من الفئات classes وتوزع البيانات على هذه الفئات للحصول على عدد المرات الذى يوجد بكل فئة ويعبر عنه بالتكرار frequency. ويتوقف عدد الفئات ومدى كل منها على طبيعة البيانات ومدى الاختلافات range of variation وعدد المشاهدات number of observations والدقة precision المطلوبة لحساب الإحصاءات من الجدول والدرجة الكافية sufficient degree لتلخيص البيانات بحيث تظهر الاتجاه العام general trend. ويلاحظ أنه بزيادة عدد الفئات بغرض الحصول على درجة عالية من الدقة، قد ينتج عنه عدم وضوح الاتجاه العام بحيث تبرز الانحرافات البسيطة أو الشاذة وعلى ذلك فيجب التوفيق ما بين هذين الاعتبارين.

يراعى أن يكون تبويب البيانات فى مجاميع متقاربة، والقاعدة التى قد تستخدم لتحديد مدى الفئات هى أن تكون فترة الفئة class interval مساوية لربع قيمة الانحراف المعيارى للمتغير σ وهذه تمثل درجة دقة عالية لحساب الإحصاءات المطلوبة من العينة، وإن كان التمثيل البيانى فى مثل هذه الحالة لا يعبر بوضوح عن الاتجاه العام. ويعتبر 1/3 إلى 1/2 الانحراف المعيارى قدرا مناسبا لفترة الفئة يكفل

عرضاً بيانياً ملائماً لمعظم البيانات ويكون النقص فى الدقة اللازمة لحساب الإحصاءات نتيجة لذلك قليلا لدرجة يمكن التغاضى عنها. وعادة يتراوح عدد الفئات ما بين 10 و 20 فئة. ويلزم تقدير الانحراف المعيارى حيث يكون غير معروف وقت إعداد الجدول. وقد أعد Tippett (1926) جدولا يبين العلاقة بين المدى فى العينة والانحراف المعيارى للعشيرة، ويبين جدول ٢ ملحق أ هذه العلاقة. ومن المفضل أن تكون حدود الفئات واضحة ومحددة حتى يسهل توزيع البيانات. كما يفضل أن يكون الحد الأدنى لفترة الفئة الأولى أقل قليلا من أصغر قيمة فى العينة. ويلاحظ عند وقوع بعض المشاهدات على حدود الفئات أن تقسم بالتساوى على الفئتين المشتركتين فى ذلك الحد boundary، أما إذا كان عددها فردياً فالمشاهدات التى ترتيبها فردى تعطى لأحد الأقسام، والتى ترتيبها زوجى تعطى للقسم الآخر. ويمكن تحديد الفئات بحيث تبدأ برقم عشرى واحد أقل من ذلك الذى يتواجد فى المشاهدات. فمثلاً إذا كانت البيانات أرقاماً صحيحة فإن مدى الفئات يبدأ برقم عشرى واحد، فإذا كان المدى الكلى 20 مثلا وحددت الفئات المستخدمة بعشر فئات، فإن مدى الفئة $20 \div 10 = 2$ وحدة. وبالتالى فإن الفئات سوف تبدأ من 9.5 – 11.5 بدلاً من 10 – 12 وهذا يضمن مثلاً أن القيمة التى مقدارها 12 بالضبط توضع فى فئة محددة وهى الفئة التى حداها 11.5 – 13.5.

وتوجد أكثر من طريقة لتوزيع البيانات على الفئات وذلك بعد تحديد حجم فترة الفئة ووضع حدود الفئات المختلفة منها:

الطريقة الأولى:

تتلخص فى وضع شرطة رأسية لكل ملاحظة فى القسم المناسب على أن تكون الشرطة الخامسة فى وضع مائل على ومتقاطعة مع الشرط الأربع وبالتالى يسهل العد وعلى ذلك يتم تكوين حزم من 5 أفراد. وبتجميع العلامات لكل قسم يمكن الحصول على التكرار المقابل له، مثال ذلك /// تعبر عن 3 أما ~~////~~ فتعبر عن 5 ... وهكذا.

الطريقة الثانية:

وتعتبر من أفضل الطرق لتوزيع البيانات وفيها تكتب كل مشاهدة على بطاقة بحجم مناسب مع التأكد من ذلك ثم يكتب مدى الفئات على بطاقات أخرى، وتقسم بطاقات المشاهدات على الأقسام المختلفة ويمكن التأكد من أن لكل مجموعة بطاقات تتبع القسم الذى تقع فيه فعلا.

الطريقة الثالثة:

تستخدم فيها الحاسبات الإلكترونية، ويتم التصنيف آليا إلى المجاميع المختلفة، وتفضل هذه الطريقة إذا ما توفرت تلك الحاسبات وإذا ما كان حجم البيانات ضخماً.

وتحسب مراكز الفئات وذلك بقسمة حاصل جمع حدى الفئة الدنيا على 2 حيث يعطى مركز الفئة الدنيا ثم يجمع مدى الفئات بعد ذلك ليعطى مراكز الفئات المتتالية، وتعد بعد ذلك جداول تشتمل على حدود الفئات ومراكزها والتكرار المقابل لكل منها.

مثال ٢-٣

كون جدول تكرارى للبيانات التالية التى تمثل الأوزان بالجرام لعدد 54 فأرا من الذكور عند عمر 60 يوما.

14.1	12.5	13.2	7.1	5.7	13.7	15.7	12.8	8.1
11.5	13.5	11.8	14.7	17.8	14.0	14.1	12.8	8.3
13.4	13.0	15.6	7.8	6.6	13.4	11.7	5.7	9.0
14.9	13.1	4.7	14.9	15.6	10.7	11.4	11.8	9.2
7.0	14.8	14.6	15.1	15.1	15.0	17.2	11.8	9.4
13.4	15.4	6.6	14.8	18.4	14.5	12.1	16.2	6.6

المدى: $18.4 - 4.7 = 13.7$

حجم العينة 54 ومن جدول ٢ ملحق أ فإن

$$\frac{\sigma}{\text{range}} = \frac{\sigma}{13.7} = 0.22$$
$$\therefore\ \sigma = 3$$

ويكون $\frac{1}{4}$ تقدير σ هو 0.8 و $\frac{1}{3}$ التقدير = 1 و $\frac{1}{2}$ التقدير = 1.5

ويمكن أخذ مدى الفئة 1.5 لتكوين حدود الأقسام الحقيقية بحيث يكون عدد الأرقام العشرية أكثر بواحد من تلك المسجل بها البيانات. ويلاحظ أن حاصل جمع تكرارات الفئات يمثل العدد الكلى للمشاهدات وقد تم حساب مراكز الفئات على النحو التالى:

مركز الفئة الأولى: $\frac{5.95 + 4.45}{2} = \frac{10.4}{2} = 5.2\ \text{gm}$

وبالتالى فإن مركز الفئة الثانية: $5.2 + 1.5 = 6.7\ \text{gm}$

ومركز الفئة الثالثة: $6.7 + 1.5 = 8.2$ gm

... وهكذا. وعليه يمكن تشكيل الجدول التالى:

التكرار	العلامات	مركز الفئة	حدود الفئات Class boundaries	مراحل الفئات (حدود الأقسام المسجلة)
3	///	5.2	4.45 – 5.95	4.5 – 6.0
5	~~////~~	6.7	5.95 – 7.45	6.0 – 7.5
3	///	8.2	7.45 – 8.95	7.5 – 9.0
3	///	9.7	8.95 – 10.45	9.0 – 10.5
7	// ~~////~~	11.2	10.45 – 11.95	10.5 – 12.0
10	~~////~~ ~~////~~	12.7	11.95 – 13.45	12.0 – 13.5
12	// ~~////~~ ~~////~~	14.2	13.45 – 14.95	13.5 – 15.0
8	/// ~~////~~	15.7	14.95 – 16.45	15.0 – 16.5
2	//	17.2	16.45 – 17.95	16.5 – 18.0
1	/	18.7	17.95 – 19.45	18.0 – 19.5
54	**المجموع**			

ويلاحظ عند إجراء الحسابات، كما سيتضح من الباب الثالث، أن كل المشاهدات التى تقع فى فئة معينة تعطى كلها قيمة مركز الفئة، فإذا كان المدى المستخدم فى الفئات صغيرا فإنه بالتالى يمكن استخدام الجداول التكرارية فى تمثيل المتغيرات ذات الطابع الاستمرارى مثل الأوزان أو الإنتاج وخلافه. وقد يصاحب ذلك بعض الفقد فى التفاصيل من حيث معاملة التوزيع المستمر كما لو كان توزيعاً متقطعاً. ولكن هناك فائدة فى تلخيص النتائج وأن الفقد يكون صغيراً ما دامت الفئات حدودها ضيقة. ومن المستحسن فى حالة الجداول التكرارية ألا تكون هناك فئات مفتوحة النهاية.

وقد تقسم التكرارات الخاصة بكل فئة على مجموع التكرارات لتعطى التكرار النسبى relative frequency distribution لكل فئة، ومن هذه التكرارات النسبية ينشأ التوزيع التكرارى النسبى. ويمكن التعبير عن التكرارات النسبية فى صورة نسب مئوية percentage distribution بالضرب فى 100. وهذه طرق مختلفة للتعبير عن التكرارات للفئات المختلفة.

٢-٢-٣ الجداول التكرارية المتجمعة Cumulative frequency tables

قد يكون الغرض من تبويب وعرض البيانات هو معرفة الأفراد الذين يقل دخلهم مثلاً عن قيمة معينة أو عدد الأفراد الذين يزيد دخلهم عن حد معين أو عدد الطلبة الحاصلين علي أقل من 45 درجة. وللإجابة عن مثل تلك الأسئلة يلزم الحصول على التوزيع التكرارى المتجمع سواء كان صاعداً (التوزيع التكرارى المتجمع الصاعد) أو هابطاً (التوزيع التكرارى المتجمع الهابط).

مثال ٢-٤

كون جدول التوزيع التكرارى المتجمع الصاعد والهابط للبيانات التى فى مثال ٢-٣

التوزيع التكرارى المتجمع الصاعد		التوزيع التكرارى المتجمع الهابط	
الفئة	التكرار	الفئة	التكرار
أقل من 6.0	3	أكثر من 4.5	54
أقل من 7.5	8	أكثر من 6.0	51
أقل من 9.0	11	أكثر من 7.5	46
أقل من 10.5	14	أكثر من 9.0	43
أقل من 12.0	21	أكثر من 10.5	40
أقل من 13.5	31	أكثر من 12.0	33
أقل من 15.0	43	أكثر من 13.5	23
أقل من 16.5	51	أكثر من 15.0	11
أقل من 18.0	53	أكثر من 16.5	3
أقل من 19.5	54	أكثر من 18.0	1

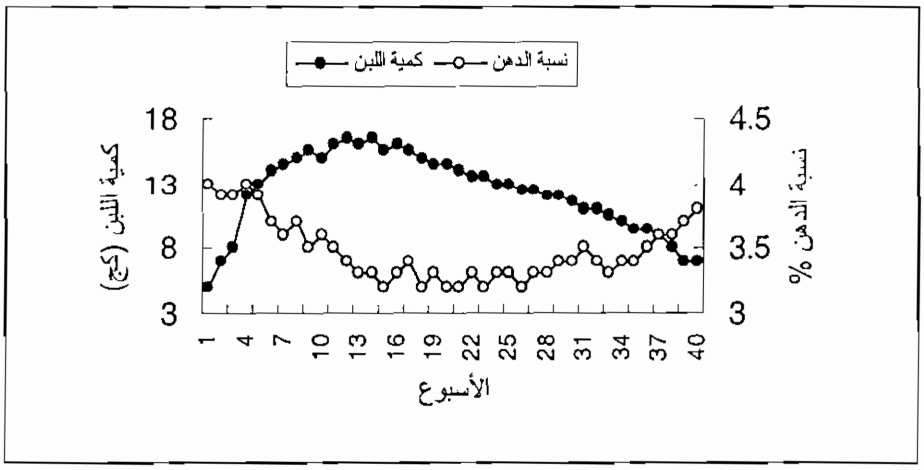
٢-٣ العرض البيانى Graphic presentation

يوجد عدة طرق للعرض البيانى، وتفضيل طريقة على أخرى يعتمد على نوعية البيانات والفكرة من عرضها. ولكن بصفة عامة يتميز العرض البيانى بأنه يجذب النظر إليه ويعطي فكرة سريعة بطريقة سهلة وواضحة وخاصة إذا ما كان عدد المشاهدات كبيراً. يعتبر العرض البيانى أكثر سهولة من الجداول فى عرض البيانات لإمكان متابعتها بمجرد النظر ولتعلقها بالذهن. أما ما يؤخذ عليه فهو أنه يمثل اتجاها عاما وليس بالدقة الكافية التى يعتمد عليها فى استخراج بعض الإحصاءات.

٢-٣-١ الخط البيانى Line chart

يستخدم لعرض البيانات الكمية quantitative data عن طريق رسم العلاقة بين متغيرين حيث يمثل المحور السينى أحد المتغيرين والذى قد يعبر عنه بالمتغير المستقل independent variable، بينما يمثل المحور الصادى المتغير الآخر والذى يطلق عليه المتغير التابع dependent variable، وعند الرغبة فى دراسة أكثر من متغيرين بالنسبة لمتغير آخر مشترك فإن المتغير المشترك يمثل على المحور السينى بينما المتغيرات الأخرى تمثل على المحور الصادى إذا كانت بنفس وحدات القياس حيث يلزم تدرج واحد. إذا اختلفت وحدات القياس فإنه يمكن أخذ محورين رأسيين يمثل كل منهما أحد المتغيرين ويرسم لكل متغير خط بيانى يبين شكل العلاقة بينه وبين المتغير المشترك. ويجب دائما بيان وحدات القياس على المحاور.

ويمثل شكل ٢-١ العلاقة بين المتوسط اليومى لكل من كمية اللبن ونسبة الدهن خلال 40 أسبوعاً من موسم الحليب فى أحد مزارع الألبان.



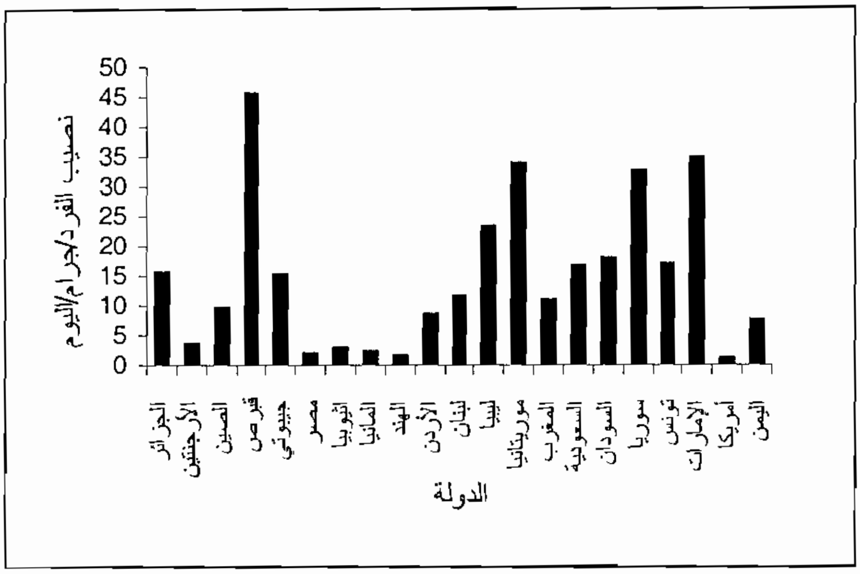
شكل ٢-١ المتوسط اليومى لكل من كمية اللبن ونسبة الدهن خلال 40 أسبوعاً من موسم الحليب فى أحد مزارع الألبان.

٢-٣-٢ الأعمدة البيانية Bar charts

تستخدم الأعمدة البيانية للتعبير البيانى عن العلاقة بين متغيرين وتستخدم فى البيانات الوصفية qualitative data. وهى عبارة عن أعمدة أو مستطيلات قواعدها متساوية و تمثل الصفة الوصفية ويمثل ارتفاعاتها التكرار المقابل لكل منها.

وقد تستخدم التكرارات الفعلية actual frequencies أو التكرارات النسبية relative frequencies. والتكرارات تمثل فى الواقع احتمالات وتوضح كمساحات areas باستخدام الأعمدة أو يعبر عنها بارتفاعات heights باستخدام الخطوط الرأسية. ويلاحظ أن تكون الأعمدة أو الخطوط علي مسافات متساوية.

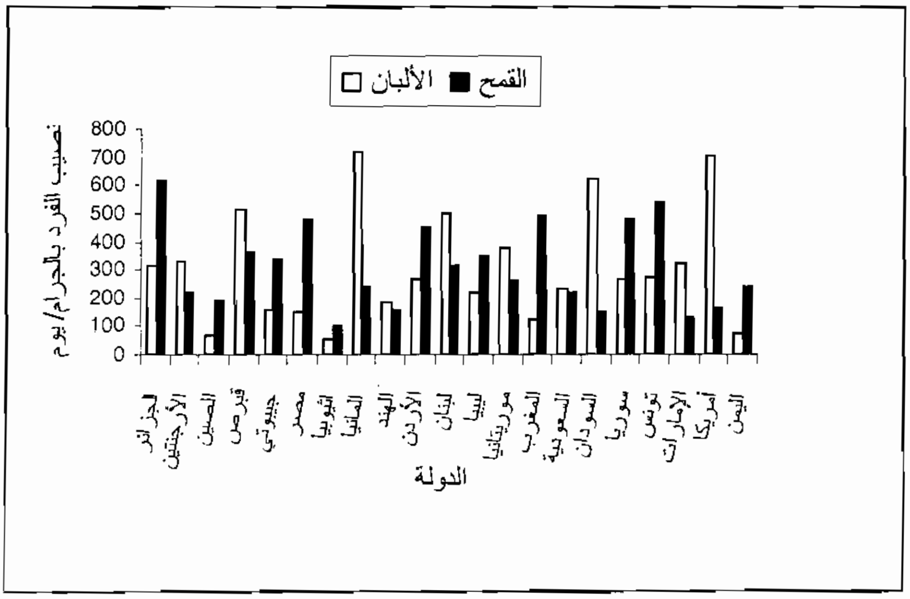
الشكلين ٢-٢، ٢-٣ يوضحان النصيب اليومى للفرد من لحوم الضأن والمعز والألبان والقمح فى بعض دول العالم من بينها مصر.



شكل ٢-٢ النصيب اليومى للفرد من لحوم الضأن والمعز فى بعض دول العالم من بينها مصر (المصدر: FAOSTAT, 2007).

٢-٣-٣ الرسوم الدائرية Pie chart

إذا كانت البيانات المطلوب عرضها تمثل مجموعاً أو نسبة مقسمة إلى أجزاء كأوجه التصرف فى الدخل أو مصادر اللبن أو اللحوم المنتجة فى مصر أو توزيع الطلبة على السنوات الدراسية المختلفة فى إحدى الكليات أو عدد العمليات المختلفة التى تم إجراؤها بإحدى المستشفيات مقسمة تبعا لنوعية العملية الجراحية، فإنه يمكن تمثيل المتغير بدائرة (فطيرة pie) تقسم إلى قطاعات بما يتناسب مع أجزاء المتغير باعتبار أن كل 1% من المجموع أو النسبة يمثلها 3.6° حيث إن الزاوية المركزية للدائرة 360° وتميز القطاعات بالتظليل أو التلوين.



شكل ٢-٣ النصيب اليومى للفرد من الألبان والقمح فى بعض دول العالم من بينها مصر (المصدر: FAOSTAT, 2007).

مثال ٢-٥

ينتج الجاموس فى مصر 1.3 مليون طن لبنا سنويا فى حين أن بقية حيوانات اللبن تنتج 0.7 مليون طن . عبر عن ذلك بيانيا.

كمية اللبن المنتجة سنويا فى مصر $2 = 0.7 + 1.3$ مليون طن.

كمية اللبن المنتجة من الجاموس بالنسبة للإنتاج الكلى $(1.3/2)(100) = 65\%$.

أما بقية الحيوانات الأخرى تنتج 35% من جملة الإنتاج.

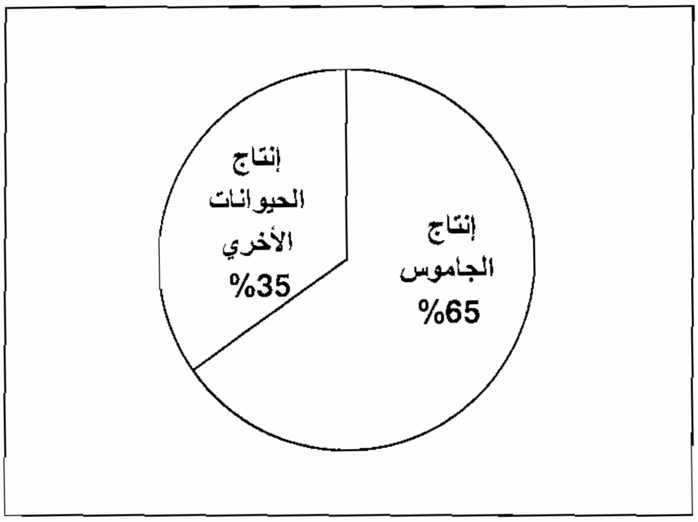
فإذا ما استخدمت الدائرة للتعبير عن ذلك بيانيا:

قطاع الجاموس يخصه $(65)(3.6) = 234^{\circ}$

بقية حيوانات اللبن يخصها $(35)(3.6) = 126^{\circ}$

المجموع $= 360^{\circ}$

ويبين شكل ٢-٤ التمثيل البيانى باستخدام الدائرة



شكل ٢-٤ إنتاج اللبن من الجاموس فى مصر بالنسبة لإنتاج اللبن الكلى

٢-٣-٤ التمثيل البيانى للعرض الجدولى

٢-٣-٤-١ المدرج التكرارى Histogram

يعتبر المدرج التكرارى من الطرق الشائعة الاستخدام لعرض البيانات المستمرة، وفيه تمثل فترات الفئات علي المحور الأفقى والتكرارات تمثل على المحور الرأسى بحيث تمثل مستطيلات يتناسب طولها مع التكرار عندما تكون فترات الفئات متساوية. أما عندما تكون فترات الفئات غير متساوية فتعدل التكرارات لطول فئة معينة (أو أى مدى ثابت) لكل الفئات ويرسم المدرج التكرارى بالأطوال الفعلية للفئات بالتكرارات المعدلة. ويمكن بمجرد النظر للمدرج التكرارى معرفة طبيعية التوزيع بسهولة.

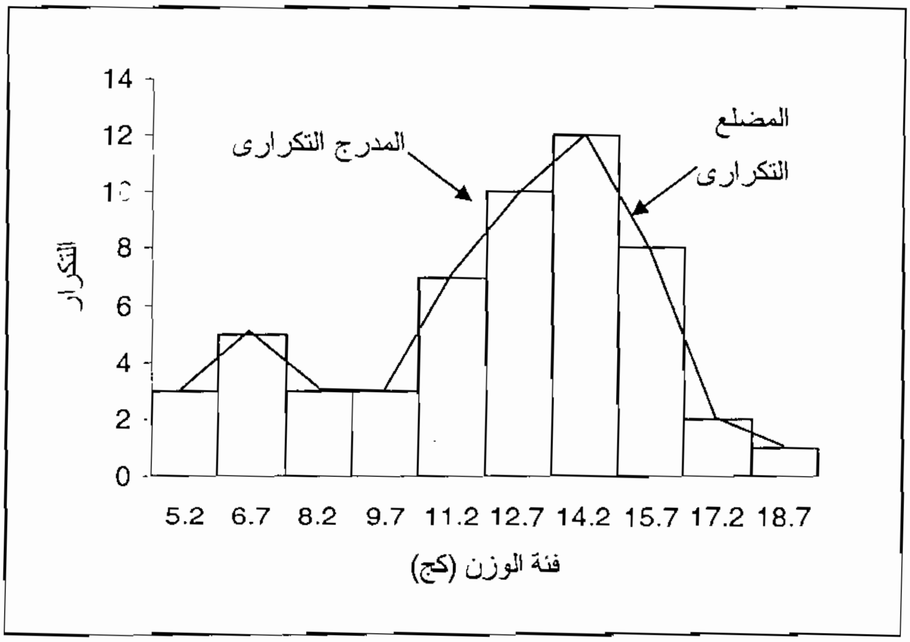
٢-٣-٤-٢ المضلع التكرارى Frequency polygon

ويمكن الحصول عليه من توصيل النقط التى تمثل إحداثياتها الأفقية مراكز الفئات وتمثل إحداثياتها الرأسية التكرارات المقابلة لكل فئة وذلك بخطوط مستقيمة. كما يمكن الحصول عليه من المدرج التكرارى بأن تضاف فئتان كل منهما يساوى الصفر إحداهما قبل الفئة الأولى والأخرى بعد الفئة الأخيرة وتنصف أعلى المستطيلات والتى تمثل مراكز الفئات ثم يتم التوصيل بين تلك النقط بخطوط مستقيمة أيضا. ويستخدم المضلع التكرارى للمقارنة بين توزيعين تكراريين أو أكثر ويلاحظ أن مساحة المدرج التكرارى هى نفسها مساحة المضلع التكرارى.

مثال ٢-٦

مثل الجدول التكرارى فى مثال ٢-٣ بيانيا باستخدام المدرج التكرارى والمضلع التكرارى.

يبين شكل ٢-٥ المضلع التكرارى والمدرج التكرارى.



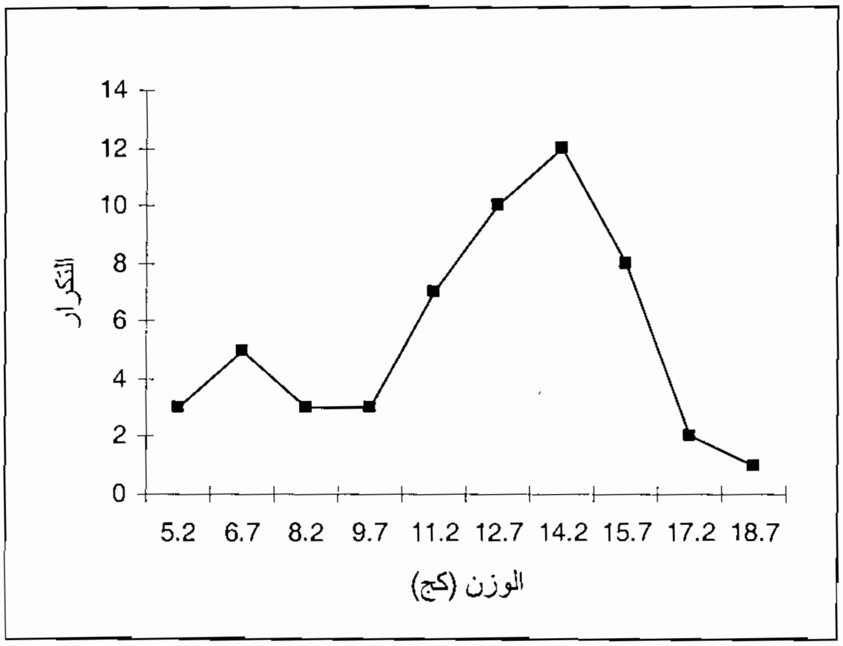
شكل ٢-٥ المدرج والمضلع التكرارى لأوزان الفئران الذكور عند عمر 60 يوماً.

٢-٣-٤-٣ المنحنى التكرارى Frequency curve

يتم التوصيل ما بين النقط التى استخدمت فى رسم المضلع التكرارى لتأخذ شكل لمنحنى، ولا يشترط أن يمر المنحنى بجميع النقط. ويؤخذ عليه أن المساحة الواقعة تحت المنحنى لا تساوى مساحة كل من المضلع والمدرج التكرارى. كما يمكن تمثيل كل من التوزيع التكرارى المتجمع الصاعد أو الهابط بمنحنى.

مثال ٢-٧

ارسم المنحنى التكرارى الذى يمثل البيانات التى فى مثال ٢-٣.



شكل ٢-٦ المنحنى التكرارى لتوزيع أوزان الفئران الذكور عند عمر 60 يوماً.

وتلعب حزم البرامج الجاهزة مثل برنامج EXCEL دورا كبيرا لتسهيل الحصول علي مختلف الأشكال والرسومات البيانية التى تم شرحها فى هذا الباب بالإضافة إلى أشكال أخرى لم يتم ذكرها.

تمارين الباب الثانى

٢-١ البيانات التالية تمثل أوزان الميلاد للعجول الذكور بالكيلوجرام من الجاموس الناتجة فى إحدى المحطات.

22	24	22	23	23	23	27	23	27	30
20	20	22	22	22	22	23	24	23	23
20	30	23	23	18	20	19	20	20	21
20	25	25	25	20	25	20	27	25	25

والمطلوب تمثيل هذه البيانات بيانيا باستخدام المدرج والمضلع والمنحنى التكرارى. ومن جدول التوزيع التكرارى مثل المنحنى المتجمع الصاعد والمنحنى المتجمع الهابط.

٢-٢ الجدول التالى يمثل عدد ونوعية العمليات الجراحية والتى تم إجراؤها فى أحد المستشفيات خلال أحد الأعوام.

نوع العملية	عدد الحالات
صدر	20
عظام	45
عيون	58
جراحة عامة	98
باطنة	115
أعصاب	23

والمطلوب تمثيل هذه البيانات بيانيا باستخدام كل من الرسوم الدائرية والأعمدة البيانية.

٢-٣ تم سؤال 50 طالباً وطالبة عن المسافة بالكيلومتر من المنزل حتى مقر الكلية وكانت الإجابات كالتالى:

6	5	3	24	15	15	6	2	1	3
5	10	9	21	8	10	9	14	16	16
10	21	20	15	9	4	12	27	10	10
3	9	17	6	11	10	12	5	7	11
5	8	22	20	13	1	8	13	4	18

والمطلوب:

أ - عمل جدول التوزيع التكرارى باستخدام 3-1 كفئة أولى.

ب - أوجد حدى الفئة الأولى.

ج - أوجد مدى الفئة.

د - استخرج كل من جدول التوزيع التكرارى الصاعد والهابط.

هـ ارسم كل من المدرج التكرارى والمضلع التكرارى.

٣ مقاييس النزعة المركزية والتشتت

Measures of central tendency and dispersion

١ – مفهوم التجميع

٢ – مقاييس النزعة المركزية

٣ – مقاييس التشتت

٤ – تباين المتوسطات

٥ – معامل الاختلاف

٦ – العلاقة بين المدى والانحراف القياسى

٧ – استخدام برنامج SAS للحصول على الإحصاءات التوصيفية

٨ – التشفير

٩ – الجداول التكرارية

يتطلب عرض البيانات فى كثير من الأحيان ضرورة الحصول على أرقام توضح وتصف العشيرة أو العينة محل الدراسة بطريقة واضحة وليست مبهمة وتجعل منها شيئا منفرداً يختلف عن العينات أو العشائر الأخرى. وكما ذكر سابقاً فإنه لا يمكن عمليا وصف كل الأفراد فى عينة ما أو عشيرة ما، ولكن يتم الحصول من هذه البيانات على الأرقام التى توصف بها العينة (الإحصاءات statistics) والعشيرة (المعالم parameters). وغالبا ما يحتاج حساب هذه الأرقام إلى عمليات رياضية تتعامل مع البيانات الموجودة فى العينة موضع الاعتبار والتى قد تشملها جميعاً أو قد تشمل، فى بعض الأحيان، بعضاً منها فقط. وحتى يمكن التعامل مع البيانات بواسطة العمليات الحسابية فإنه يلزم الإلمام بمفهوم التجميع والذى غالباً ما يستعمل فى الحسابات الإحصائية.

٣-١ مفهوم التجميع Summation notation

فى علم الإحصاء كثيراً ما يتم جمع قيم البيانات أو مربعاتها ... الخ. فعند الرغبة فى قياس متوسط أوزان مجموعة من العجول عند بدء تسمينها مثلا فإنه بدلا من الإشارة إلى عملية التجميع عن طريق ذكرها حرفياً، فإن هناك طريقة متبعة لتوضيح ذلك باستعمال إشارات معينة لتدل على عملية التجميع وهى حرف $\sum$ اليونانى وينطق سيجما (جدول ١٩ ملحق أ). وهذه الإشارة تعنى التجميع، فإذا أريد مثلاً جمع أوزان 10 عجول فبدلاً من وضع ذلك فى صورة $Y_1 + Y_2 + \cdots + Y_{10}$ حيث Y_1 تمثل قيمة الوزن للمشاهدة الأولى و Y_2 تمثل قيمة الوزن للمشاهدة الثانية ... وهكذا، فإنه يمكن الإشارة إلى ذلك باستعمال إشارة التجميع $\sum_{i=1}^{10} Y_i$ حيث i تعتبر فى هذه الحالة تمييزاً يشير إلى رقم أو ترتيب المشاهدات ويوضع تحت وفوق علامة التجميع المدى الذى يتم عليه ذلك بقيمتيه الدنيا والعليا، على الترتيب. ففى الحالة السابقة يتم التجميع من المشاهدة رقم $i = 1$ إلى المشاهدة الأخيرة $i = 10$.

مثال ٣-١

إذا كانت أوزان مجموعة عجول الفريزيان بالكيلوجرام عند بدء تسمينها هى كالتالى:

160,205, 200, 158, 195, 195, 240, 215, 165, 180, 210 كج

فإن قيمة $\sum_{i=1}^{11} Y_i = 2123$ ، وهذا يمثل $2123 = 160 + \cdots + 180 + 210$

وإذا افترض أن المطلوب هو تجميع القيم الثلاث الأولى فقط أى $i = 1,2,3$ فإنه يشار إلى ذلك $\sum_{i=1}^{3} Y_i$ وتصبح القيمة = 555 كج، أى $555 = 165 + 180 + 210$.

وعند افتراض أن عدد الأفراد أو المشاهدات فى عينة يساوى n وأريد التجميع لكل أفراد العينة فإنه يشار إلى ذلك $\sum_{i=1}^{n} Y_i$ حيث تعنى i فى هذه الحالة $1,2,\cdots,n$

ويمكن استعمال علامة التجميع فى تجميع دالات للمتغير Y حيث يمكن الحصول على قيمة $\sum_{i=1}^{n} Y_i^2$ وهذه تعنى الحصول على مربع كل قيمة من قيم المتغير Y_i ثم يتم تجميع هذه القيم المربعة على المشاهدات من $i = 1$ حتى $i = n$. ففى المثال السابق $\sum_{i=1}^{i=11} Y_i^2$ تعنى $416189 = (160)^2 + \cdots + (180)^2 + (210)^2$. ويمكن استعمال أى من الحروف كتمييز ولكن هناك شبه عرف بين الإحصائيين على استعمال الحروف اللاتينية البادئة من i ثم j ثم k وهكذا.

وفى كثير من الأحيان إذا كان التجميع على كل القيم الموجودة فى العينة فإنه يمكن التغاضى عن وضع المدى الذى يتم التجميع عليه ويكتفى بالإشارة إلى $\sum Y_i$ أو حتى يمكن إهمال التمييز كأن تستعمل $\sum Y$ فقط. وهو ما قد يشاهد فى خلال هذا الكتاب عند توضيح المعادلات أو إثباتها.

مثال ٣-٢

إذا كانت قيمة $Y_1 = 7$ ، $Y_2 = 3$ ، $Y_3 = 5$ فأوجد ما يلى:

(أ) $\sum Y_i$ (ب) $\sum_{i=1}^{3} 2Y_i^2$ (ج) $\sum_{i=2}^{3} (Y_i - i)$

الحل:

(أ)
$$\sum Y_i = Y_1 + Y_2 + Y_3 = 7 + 3 + 5 = 15$$

(ب)
$$\sum_{i=1}^{3} 2Y_i^2 = 2(7)^2 + 2(3)^2 + 2(5)^2 = 98 + 18 + 50 = 166$$

(ج)
$$\sum_{i=2}^{3} (Y_i - i) = (3-2) + (5-3) = 3$$

ويمكن استعمال مفهوم التجميع أيضا للتعامل مع متغيرين بدلا من متغير واحد.

مثال ٣-٣

إذا كانت قيم $Y_1 = 2$ ، $Y_2 = 3$ ، $Y_3 = 5$

وكانت قيم $X_1 = 4$ ، $X_2 = 2$ ، $X_3 = 5$ فأوجد قيمة:

(أ) $\sum_{i=1}^{3} X_i Y_i$ ، (ب) $\left(\sum_{i=2}^{3} X_i\right)\left(\sum_{i=1}^{2} Y_i^2\right)$

الحل:

(أ)
$$\sum_{i=1}^{3} X_i Y_i = (4)(2) + (2)(3) + (5)(5) = 8 + 6 + 25 = 39$$

(ب)
$$\left(\sum_{i=2}^{3} X_i\right)\left(\sum_{i=1}^{2} Y_i^2\right) = (X_2 + X_3)(Y_1^2 + Y_2^2) = (2+5)(2^2 + 3^2)$$
$$= (7)(4+9) = (7)(13) = 91$$

وهناك بعض القواعد الأساسية للتعامل مع مفهوم التجميع وهى:

قاعدة ٣-١

تجميع حاصل جمع متغيرين أو أكثر يساوى حاصل جمع تجميعاتهما وعلى ذلك:

$$\sum_{i=1}^{n} (X_i + Y_i + Z_i) = \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i + \sum_{i=1}^{n} Z_i$$

قاعدة ٣-٢

إذا كان ث (C) ثابتا فإن مجموع حاصل ضرب المتغير فى الثابت يساوى قيمة الثابت فى مجموع قيم المتغير أى:

$$\sum_{i=1}^{n} CY_i = C\sum_{i=1}^{n} Y_i$$

قاعدة ٣-٣

إذا كان ث (C) ثابتاً فإن تجميع الثابت من 1 إلى n يساوى حاصل ضرب n فى قيمة الثابت أى:

$$\sum_{i=1}^{n} C = nC$$

مثال ٣-٤

إذا كانت قيمة $Y_1 = 3$ ، $Y_2 = -1$ ، $X_1 = 2$ ، $X_2 = 4$

فأوجد قيمة $\sum_{i=1}^{2}(3X_i - Y_i + 4)$

الحل:

$$\begin{aligned}\sum_{i=1}^{2}(3X_i - Y_i + 4) &= \sum_{i=1}^{2}3X_i - \sum_{i=1}^{2}Y_i + \sum_{i=1}^{2}4\\ &= 3\sum_{i=1}^{2}X_i - \sum_{i=1}^{2}Y_i + (2)(4)\\ &= (3)(2+4) - (3-1) + 8 = 24\end{aligned}$$

٣-٢ مقاييس النزعة المركزية "التمركز" Measures of central tendency

الجداول أو الرسوم البيانية يمكن أن توضح شكلاً عاماً لتوزيع متغير ما وتركيز قيم بيانات هذا المتغير. إلا أنه يوجد مقاييس عددية تدل على توزيع المتغير بطريقة كمية دون الحاجة إلى الرجوع إلى الرسوم البيانية.

تعريف ٣-١

أى مقياس يوضح مركز center أو موقع location مجموعة من البيانات مرتبة حسب قيمها يسمى بمقياس للمركزية. وأهم هذه المقاييس وأكثرها استخداماً وشيوعاً هو المتوسط الحسابى average أو arithmetic mean.

تعريف ٣-٢

إذا كانت مجموعة البيانات $Y_1, Y_2, \cdots, Y_N$ تمثل عشيرة حجمها N مشاهدة فإن متوسط العشيرة (μ) population mean يحسب من المعادلة:

$$\mu = \frac{\sum_{i=1}^{N} Y_i}{N} \qquad (٣-١)$$

ولكن فى أغلب الأحيان فإنه من غير الممكن أو العملى قياس كل البيانات المتعلقة بعشيرة ما حجمها N وبالتالى يتم الحصول على عينة محددة العدد، ومن هذه العينة يتم حساب متوسط العينة.

تعريف ٣-٣

إذا كانت البيانات $Y_1, Y_2, \cdots, Y_n$ تمثل عينة محدودة حجمها n فإن متوسط العينة $\overline{Y}$ sample mean يمكن حسابه من المعادلة:

$$\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} \qquad (٣-٢)$$

فإذا كانت العينة ممثلة تمثيلا جيداً للعشيرة المأخوذة منها، فإن متوسط العينة يعتبر ممثلا (تقديرا لنقطة point estimate) لمتوسط العشيرة.

مثال ٣-٥

فى البيانات المذكورة فى مثال ٣-١ احسب متوسط العينة والتى حجمها $n = 11$

الحل:

$$\overline{Y} = \frac{\sum_{i=1}^{11} \overline{Y}}{n} = (210 + 180 + \cdots + 160)/11 = 193 \ \ \text{kg}$$

ومن الملاحظ هنا أن جميع القيم الموجودة فى العينة قد استخدمت فى حساب المتوسط ومعنى هذا أن المتوسط تقدير شامل ويسمى هذا التقدير بأنه تقدير كاف sufficient من الناحية الرياضية.

وخاصيتان هامتان للمتوسط هما:

- مجموع انحرافات قيم المشاهدات عن متوسطها يساوى صفرا أى $\sum(Y-\overline{Y})=0$
- مجموع مربعات فروق المشاهدات عن متوسطها هو أدنى ما يمكن، أى أن القيمة $\sum(Y-\overline{Y})^2$ تكون أدنى ما يمكن عما لو استخدمت أى قيمة أخرى غير $\overline{Y}$. وتعرف هذه الخاصية بـــ "المربعات الصغرى least squares ".

وهناك مقاييس أخرى للنزعة المركزية يكون استعمالها أقل من المتوسط خاصة عند معالجة البيانات البيولوجية أو الاقتصادية أو الاجتماعية، ومع ذلك ففى حالات كثيرة فإن تلك المقاييس قد تصف العشيرة أو العينة بطريقة أكثر تفهما أو أكثر واقعية.

تعريف ٣-٤

الوسيط (الوسط) median لمجموعة من البيانات المرتبة حسب قيمها هو القيمة الوسطية (فى حالة وجود عدد فردى odd للمشاهدات) أو المتوسط الحسابى للقيمتين الوسيطتين (فى حالة وجود عدد زوجى even للمشاهدات).

وعلى ذلك فإن الوسط هو القيمة التى تقسم توزيع البيانات المرتبة تنازليا أو تصاعدياً إلى قسمين، فهو يقسم التوزيع بحيث إن نصف القيم يقع قبله والنصف الآخر يقع بعده.

تعريف ٣-٥

المنوال mode لمجموعة من البيانات هو تلك القيمة التى تظهر أكثر عددا من المرات أو لها أعلى تكرار.
وقد يوجد مجموعة من البيانات ليس لها منوال بمعنى عدم وجود أى قيمة تتكرر أكثر من مرة واحدة أو قد يكون هناك أكثر من قيمة لها نفس التكرار وفى هذه الحالة يعتبر التوزيع ذا منوالين bimodal أو أكثر multimodal.

مثال ٣-٦

احسب كل من الوسيط والمنوال للبيانات المذكورة فى المثال ٣-١
الحل:

إذا رتبت البيانات فى العينة ترتيبا تنازليا أو تصاعديا فإن العينة تبدو كالتالى:

240, 215, 210, 205, 200, 195, 195, 180, 165, 160, 158

فإن قيمة الوسيط 195 kg وهى التى تقسم البيانات إلى قسمين متساويين أن هناك 5 قيم أكبر منها وخمس قيم أخرى أقل منها.

أما قيمة المنوال فهى أيضاً 195 kg حيث إنها هى القيمة الوحيدة التى تكررت فى العينة مرتين.

ومن الملاحظ فى المثال السابق أن قيمتى الوسيط والمنوال متساويتان وتقتربان من قيمة المتوسط الحسابى. وتعتبر هذه الظاهرة من مميزات البيانات التى تتبع فى توزيعها التوزيع الطبيعى normal distribution.

وتلخيصاً لما سبق فإنه يمكن المقارنة بين المقاييس الثلاثة وهى المتوسط والوسيط والمنوال، فالمتوسط هو أكثر المقاييس استخداماً فى الإحصاء خاصة فى معالجة البيانات البيولوجية. ويرجع ذلك إلى أن توزيع متوسطات العينات عادة ما يكون معروفاً من الناحية الرياضية وبالتالى فإن الطرق المستخدمة فى الاستدلال الإحصائى تكون المبنية على أساس المتوسط الحسابى وبالتالى يمكن معرفة مقدار الخطأ فى التقديرات وتوزيعه. وكما سبق القول فإن المتوسط يعتبر تقديراً لمتوسط العشيرة الأكثر تمثيلا لها وذلك لأنه يستعمل كل البيانات فى العينة بالإضافة إلى أن المتوسط يتمتع بأقل قدر من الخطأ (التباين) minimum variance. أما ما يعاب على المتوسط فهو أنه يتأثر تأثراً كبيراً بالقيم المتطرفة (أى العالية جداً أو المنخفضة جداً outliers)، وذلك لدخول مثل هذه القيم فى حسابه. أما الوسيط فهو سهل فى حسابه ولا يتأثر بالقيم غير الطبيعية وبالتالى يعطى فكرة أوضح عن غالبية القيم فى العينة عن المتوسط، ولكن يعاب عليه طبعاً أنه لا يستعمل كل بيانات العينة فهو غير كاف، وأيضاً فإن تباين قيمة العينات المختلفة يكون أكبر من حالة المتوسط. وعلى ذلك فإن المتوسط يكون أكثر ثباتا من الوسيط وبالتالى فالمتوسط يمثل متوسط العشيرة بقدر أكبر من الدقة أما المنوال فهو أقل المقاييس كفاءة وهو يكون غير ذى نفع فى حالة العينات صغيرة الحجم. وفى بعض الحالات فإن المنوال قد لا يوجد أو يكون متعددا. إلا أنه فى حالة البيانات العديدة فيكون له نفع خاص فى البيانات العدية مثل طول دورة الشبق وعدد الخلفه فى البطن فى بعض أنواع الحيوانات مثله مثل الوسيط فى ذلك.

وهناك مقاييس أخرى للمركزية يمكن تعريفها هنا فى هذا المجال وهى أقل استخداماً من سابقتها، مثال ذلك متوسطات الأرباع quartiles، الأعشار deciles والمئويات percentiles وهى تلك النقاط التى تقسم التوزيع إلى أرباع أو أعشار أو مئويات على التوالى. وعل ذلك فيمكن القول بأن الوسط يمثل الربع الثانى والعشر الخامس والمئوية الخمسين.

تعريف ٣-٦

المتوسط الهندسى لعدد (n) من القيم هو الجذر النونى لحاصل ضرب هذه القيم أى:

$$G = \sqrt[n]{Y_1 . Y_2 \cdots Y_n} = \left(Y_1 . Y_2 .... Y_n\right)^{1/n} \qquad (٣-٣)$$

(ملاحظة: بدلاً من بيان الضرب تفصيلياً يمكن التعبير عن هذه العملية باستخدام πY_i حيث π هى من الحروف اليونانية وينطق باى).

وبالنسبة للبيانات العدية count data فإنه يمكن حساب ما يعرف بالمتوسط التوافقى harmonic mean وهو يستخدم فى حالة إيجاد متوسط عدد المشاهدات فى الفئات المختلفة أو التقسيمات المختلفة subclasses.

تعريف ٣-٧

المتوسط التوافقى لعدد (n) من القيم هو مقلوب متوسط مقلوب القيم فى كل فئة (يرمز له بالرمز H) أى:

$$\frac{1}{H} = \frac{1}{n} \sum_i \left(\frac{1}{Y_i}\right) \qquad (٣-٤)$$

وتستعمل هذه المتوسطات أيضاً فى حساب متوسطات النسب ratios والمعدلات rates.

مثال ٣-٧

احسب المتوسط الهندسى للقيم التالية والتى تمثل معدل الزيادة العددية فى بعض القطعان.

$$1.61,\ 1.4,\ 1.5,\ 1.27,\ 1.22,\ 1.25,\ 1.3,\ 1.2$$

الحل:

$$G = \sqrt[8]{(1.2)(1.3)(1.25)(1.22)(1.27)(1.5)(1.4)(1.61)}$$
$$= 1.337$$

مقارنة بالمتوسط الحسابى وهو 1.344

مثال ٣-٨

احسب متوسط عدد العجول فى عمر سنة الموجودة فى ثمانية قطعان من البيانات التالية:

$$81, 70, 75, 64, 61, 63, 65, 60$$

الحل:

$$H = 1 \div \frac{1}{H} = 1 \div \frac{1}{8}\left(\frac{1}{60} + \frac{1}{65} + \frac{1}{63} + \frac{1}{61} + \frac{1}{64} + \frac{1}{75} + \frac{1}{70} + \frac{1}{81}\right)$$

$$= 1 \div \frac{1}{8}(0.1199) = 66.72 \quad \text{عجل}$$

ويعتبر المتوسط التوافقى فى حالة الأعداد أكثر تمثيلا مقارنة بالمتوسط الحسابى لنفس البيانات والذى قيمته 67.38 عجل.

صندوق ٣-١

مقاييس النزعة المركزية:

المتوسط : $\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$

الوسيط (الوسط): القيمة الوسطية فى العينة

المنوال: القيمة الأكثر تكرارا فى العينة

المتوسط الهندسى: $G = \sqrt[n]{Y_1 . Y_2 ... Y_n} = (Y_1 . Y_2 \cdots Y_n)^{1/n}$

المتوسط التوافقى: $H = 1 \div \frac{1}{n} \sum_i \left(\frac{1}{Y_i} \right)$

٣-٣ مقاييس التشتت (التباين) Measures of dispersion or variation

إن مقاييس المركزية لا توضح مدى توزيع البيانات حول القيم المركزية للتوزيع. فقد تتماثل مجموعة من القيم فى قيمة متوسطها أو منوالها ومع ذلك تختلف اختلافاً بيناً فى توزيع القيم حول المتوسط.

مثال ٣-٩

إذا أخذت ثلاث مجموعات من الأرقام التالية:

(أ) 30, 31, 32, 33, 34

(ب) 34, 32, 28, 29, 37

(ج) 48, 45, 32, 20, 15

لاحظ أن كل من المجموعات الثلاث لها متوسط يساوى 32 فى حين أن القيم فى المجموعة (أ) كلها تقترب من القيمة المركزية (المتوسط) للمجموعة حيث أكبر شذوذ أو انحراف عن هذه القيمة هو وحدتان بالموجب أو بالسالب فى حين أنه فى المجموعة (ب) هناك انحرافات أكبر حيث تبتعد إحدى القيم عن القيمة المتوسطة بخمس وحدات. أما بالنسبة للمجموعة الأخيرة (ج) فإنها أكثرها تشتتاً حيث بلغ انحراف إحدى القيم عن المتوسط 16 وحدة. وهكذا فإنه لا يمكن تحديد أى شكل للتوزيع لقيم المتغير محل الدراسة بمعرفة المتوسط فقط.

وكما أن هناك عدداً من المقاييس استخدمت لقياس مركزية (أو موقع) التوزيع فإنه يوجد أيضاً أكثر من مقياس لتشتت قيم المتغير أو تباينها حول مركزها.

تعريف ٣-٨

المدى range عبارة عن الفرق بين أعلى قيمة وأقل قيمة يأخذها المتغير. وهو يدل على مدى ابتعاد القيم عن المتوسط، وطبيعى أنه كلما شذت القيم عن مركزها كلما اتسع المدى. وفى المثال (٣-٩) كان المدى للمجموعة (أ) $34-30=4$ وحدات فى حين أنه فى المجموعة (ج) كان المدى فيها $48-15=33$ وحدة.

والمدى يعتبر مقياساً سريعاً يمكن منه أخذ فكرة مبدئية عن التشتت فى العينة المدروسة ولكن طبعاً يعاب عليه أنه غير دقيق وأنه لا يستخدم فى حسابه إلا على أعلى وأقل القيم فى التوزيع. وبذا فهو غير ذى فائدة تقريباً فى حالة العينات أو العشائر الكبيرة حيث يتأثر بالقيم الشاذة تأثرا كبيرا.

تعريف ٣-٩

متوسط الانحرافات average deviation لمجموعة من البيانات عددها (N) هو مجموع الانحرافات المطلقة عن المتوسط مقسوما على العدد (N) أى:

$$\frac{\sum_{i=1}^{N}|Y_i-\mu|}{N} \qquad (٣-٥)$$

مثال: ٣-١٠

احسب متوسط الانحرافات لكل من العينتين أ ، ج فى المثال (٣-٩)

الحل:

المتوسط فى كل من العينتين يساوى 32 وحدة وعلى ذلك يكون:

متوسط الانحرافات فى (أ)

$$\frac{|-2|+|-1|+\text{صفر}+|1|+|2|}{5}=\frac{6}{5}=1.2$$

متوسط الانحرافات فى (ج)

$$\frac{|16|+|13|+\text{صفر}+|-12|+|-17|}{5}=\frac{58}{5}=11.6$$

وعلى ذلك فإنه كلما كانت القيم فى البيانات أكثر انحرافاً عن القيمة الوسطية فإن متوسط الانحرافات يزداد. يعاب على هذا القياس أن معالجته الرياضية تصبح غير سهلة ولكنه لا يضخم من قيمة الانحرافات الشاذة التى تحدث فى القيم.

تعريف ٣-١٠

تباين قيم متغير عشوائى هو متوسط مربع انحرافات قيم المتغير عن متوسط العشيرة ويرمز له بالرمز σ^2

النتائج:

فى نافذة المخرجات output window نحصل على النتائج كالتالى:

OUTPUT FROM MEANS PROCEDURE
Analysis Variable: BWEIGHT

N	Mean	Minimum	Maximum	Range	Variance
16	25.13	21.00	35.00	14.00	11.45

Std Dev	Std Error	CV
3.38	0.85	13.47

لحساب الوسيط والمنوال يلزم استخدام proc univariate بدلا من proc means فتظهر النتائج أن الوسيط = 24.5 و المنوال = 26

٣-٨ التشفير Coding

قد يكون من الملائم فى كثير من الحالات عند إجراء الحسابات أن تجرى عملية تشفير للبيانات بغرض تسهيل العمليات الحسابية وذلك بمعاملتها بطريقة رياضية (جبرية) كأن يتم مثلا طرح قيمة معينة من جميع البيانات الواردة فى العينة أو إضافة قيمة لها ... الخ. وقد يكون ملائما فى بعض الأحيان أيضاً إجراء عمليات أخرى مثل الضرب أو القسمة على ثابت أو حتى استخراج الجذور أو الحصول على لوغاريتمات الأعداد. كل ذلك لا يحدث تغيراً فى طريقة حساب الإحصاءات مادامت البيانات جميعها عوملت بنفس المعاملة وأخذ هذا التشفير فى الحسبان عند نهاية عملية الحسابات وذلك بغرض إرجاع القيم المحسوبة إلى حالتها (أو قيمتها) الأصلية وهو ما يعرف بعملية عكس التشفير decoding. وتختلف عملية عكس التشفير حسب نوع المعاملة الجبرية أو الحسابية التى أجريت على البيانات.

فى حالة الجمع أو الطرح لقيمه ثابتة adding or subtracting فإن هذه العملية تؤثر على قيمة المتوسط فقط دون التباين. ولإرجاع قيمة المتوسط المحسوبة إلى قياسه الأساسى يتم عكس ما تم إحداثه من عمليات حسابية وذلك للحصول على القيم الحقيقية actual values، فإذا طرح ثابت من جميع القيم فإن عملية عكس التشفير تتم عقب حساب المتوسط بإضافة نفس القيمة المطروحة إلى التقدير الناتج، والعكس صحيح فى حالة الجمع. أما بالنسبة لقيمة التباين وبالتالى الانحراف القياسى ... الخ، فإن قيمتها لا تتأثر بتلك العملية نظراً لأن تلك القيم تحسب على أساس انحراف من قيمة المتوسط.

وهنا تجدر الإشارة إلى بعض القواعد الجبرية التى تحكم العلاقات بين المتغير والثابت وحساب القيم منهما.

قاعدة ٣-٥

$$\mu_{(Y-C)} = \mu_y - C$$

الإثبات:

بما أن تقدير قيمة μ يتم عن طريق تقدير المتوسط $\overline{Y}$ فيكون كالتالى:

$$\begin{aligned}\mu_{(Y-C)} &= \sum (Y_i - C)/n \\ &= \left(\sum Y_i - \sum C\right)/n \\ &= \left(\sum Y_i - nC\right)/n \\ &= \frac{\sum Y_i}{n} - \frac{nC}{n} = \overline{Y} - C\end{aligned}$$

وبالتالى لعكس التشفير يتم إضافة قيمة الثابت C إلى التقدير الناتج من حساب متوسط القيم المشفرة coded values.

قاعدة ٣-٦

$$V(Y_i - C) = V(Y_i)$$

حيث إن تباين الثابت يساوى صفراً لعدم تغير قيمه، أى أن $V(C) = 0$ ، أى أنه عند عملية التشفير لا يؤثر ذلك على التباين وبالتالى الانحراف القياسى وذلك لأن الحساب يتم من المتوسط فعند ذلك يمكن وضع قيمة $(Y_i - C) = Y_i$ ويصبح حساب التباين للقيم المشفرة للمتغير الجديد Y_i' وبالتالى فإن المتوسط سوف يحسب من مراكز القيم الجديدة، وعلى ذلك فلا يؤثر ذلك على التباين والانحراف القياسى.

مثال ٣-١٧

من مجموعة البيانات التالية اطرح ثابت قيمته 50 ثم احسب كل من المتوسط والانحراف القياسى.

$$51, 59, 57, 55, 53$$

الحل:

بعد الطرح تصبح القيم $1, 9, 7, 5, 3$

المتوسط $\overline{Y} = \frac{25}{5} = 5$

بإضافة الثابت فإن المتوسط الفعلى يكون $50 + 5 = 55$

التباين

$$\frac{3^2 + 5^2 + ... + 1^2 - \frac{(25)^2}{5}}{5-1} = \frac{165 - 125}{4} = 10$$

الانحراف القياسى $\sqrt{10} = 3.162$ و هذه القيمة لا تتأثر بالتشفير.

فى حالة القيم التى تشفر عن طريق قسمتها (أو ضربها) فى رقم ثابت فإنه تبعاً لذلك يتم حساب المتوسط على القيم المشفرة وبعد ذلك يتم الحصول على قيمة المتوسط الحقيقى بعكس التشفير وذلك بالضرب فى الثابت (أو القسمة عليه).

أما بالنسبة للتباين فإنه حسب قاعدة ٣-٤ فإن تباين الثابت مضروبا فى المتغير (أو مقسوماً عليه) يساوى تباين المتغير مقسوماً على مربع الثابت (أو مضروباً فيه). ويلزم فى هذه الحالة عكس عملية التشفير عند الرغبة فى الحصول على القيم الحقيقية. وعلى ذلك ففى التباين يضرب فى (أو يقسم على) مربع الثابت، أما فى حالة الانحراف القياسى فإنه فقط لعكس عملية التشفير يضرب فى (أو يقسم على) الثابت.

مثال ٣-١٨

القيم التالية تمثل تركيز عنصر نادر فى تحليلات العينات وهى:

0.0049, 0.0049, 0.0048, 0.0045, 0.0044, 0.0047

المطلوب حساب التباين والانحراف القياسى ومعامل الاختلاف باستخدام التشفير ثم حساب القيم الحقيقية.

الحل:

(١) يتم ضرب القيم فى ثابت قيمته 10000 للتخلص من العلامة العشرية وتصبح الأرقام صحيحة.

(٢) يتم طرح ثابت آخر قيمته 40 من جميع القيم وتصبح القيم 9, 9, 8, 5, 4, 7

(٣) يتم حساب المتوسط والانحراف القياسى باستخدام القيم المشفرة.

(٤) للحصول على القيم الحقيقية يتم عكس التشفير.

المتوسط المشفر $7 = \frac{42}{6}$، المتوسط الحقيقى $0.0047 = (7+40)X\frac{1}{10000}$

التباين المشفر $\frac{316-6(7)^2}{6-1} = \frac{22}{5} = 4.4$

التباين الحقيقى $4.4X\frac{1}{(10000)^2} = 4.4X10^{-8}$

الانحراف القياسى الحقيقى $\sqrt{4.4X10^8} = 2.0976X10^{-4}$

معامل الاختلاف المشفر $\frac{2.0976}{7}X100 = 29.97\%$

معامل الاختلاف الحقيقى $\frac{2.0976X10^{-4}}{0.0047}X100 = 4.46\%$

٣-٩ الجداول التكرارية Frequency table

يكون من الأوفق فى كثير من الأحيان أن تلخص البيانات المتحصل عليها بحيث توضع فى جداول تحتوى على تكرار كل قيمة تحدث فى العينة أو توضع كل مجموعة قيم مع بعضها فى فئة معينة ثم يذكر عدد المرات التى حدثت بها هذه الفئة. ويكون ذلك مفيداً فى حالة زيادة عدد المشاهدات فى العينة. فإذا كانت القيم المشاهدة تقع فى عدد قليل من الأقسام (الفئات) وكل هذه الأقسام ممثلة كما فى عدد الخلفة فى البطن الواحدة فى المعز مثلاً حيث تتراوح حصرياً ما بين 1 أو 5 فإن كل فئة تمثل أحد القيم ويوضع أمامها عدد الحالات التى شوهدت من هذه القيم، وهذا يمثل جدولاً تكرارياً. أما إذا كانت قيم المتغير لها صفة الاستمرارية إلى حد ما مثل عدد الخلفة فى البطن فى الأرانب (التى تتراوح ما بين 1 و 12) أو نسب الإصابة بمرض معين فى مجاميع الحيوانات، فإن القيم توضع فى فئات تشمل كل منها أكثر من قيمة. ووضع البيانات بهذه الطريقة مع بيان لعدد المشاهدات التى تدخل ضمن كل فئة أو داخل حدود كل تقسيم هو ما يعرف بالجداول التكرارية. وغالباً ما يمكن تمثيل هذه الجداول التكرارية أيضاً، وكما سبق الإشارة إليه، فى صورة أشكال بيانية (هستوجرامات). وفى الجداول التكرارية توضع البيانات فى عدد معقول من الأقسام أو الفئات cells or classes وذلك يضمن عدم فقدان الأساس فى تلخيص البيانات، وهذه عادة ما يتراوح عددها بين

10–20 فئة. وتعرف الفئة على أنها تصنيف توضع فيه كل القيم التى تقترب من بعضها. وبناء على المدى الذى تأخذه البيانات عامة، فإنه يحدد مراحل الفئات أو حدود الفئات أو الخلايا cell intervals or class boundaries التى تحدد الحد الأدنى والأعلى لكل فئة، وهذه تصمم بحيث تكون هذه الحدود أعلى من قيم المشاهدات حتى لا يكون هناك أى تداخل بين الأقسام.

وتحدد نقاط تمثل مراكز الفئات أو الخلايا cell midpoint وهى عبارة عن متوسط مدى الفئات. وعند إجراء الحسابات فإن كل المشاهدات التى تقع فى فئة معينة تعطى كلها قيمة مركز الفئة، فإذا كان المدى المستخدم فى الفئات صغيراً فإنه بالتالى يمكن استخدام الجداول التكرارية فى تمثيل المتغيرات ذات الطابع الاستمرارى مثل الأوزان أو الإنتاج وخلافه. وقد يصاحب ذلك بعض الفقد فى التفاصيل من حيث معاملة التوزيع المستمر كما لو كان توزيعاً متقطعاً، ولكن هناك فائدة فى تلخيص النتائج وأن الفقد يكون صغيرا ما دامت الفئات حدودها ضيقة. ومن المستحسن فى حالة الجداول التكرارية ألا تكون هناك فئات مفتوحة النهاية كأن توضع القيم الشاذة فى فئة نهائية، فمثلاً توضع عدد الخلفة فى البطن فى الأرانب 20 فأكثر، حيث إن عدد الأمهات التى تعطى أكثر من 20 خلفة فى البطن يكون قليلاً وقد تشمل 22 أو 23 أو أكثر، فى حين أن تكون الفئات الأخرى الوسطية لها مدى حوالى 2 مثلاً من 8–10، 10–12... وهكذا.

وعموماً لتجميع البيانات فى جداول تكرارية فإنه يجب:

أولاً: تحديد عدد الفئات التى سوف تبوب فيها البيانات، ويفضل أن يتراوح عددها بين 10، 20 فئة فى حالة ما إذا كان المتغير مستمرا أو فى عدد ملائم من الفئات فى حالة المتغيرات المتقطعة؛ وذلك حتى لا تفقد قدراً كبيراً من المعلومات، وليس عدداً كبيراً جداً حتى لا تفقد ميزة التجميع، ويلاحظ أنه كلما قل عدد البيانات قل عدد الفئات، ويقل بذلك احتمال دخول القيم الشاذة فى العينة.

ثانياً: يتم تقسيم المدى المحسوب من البيانات، وهو الفرق بين أعلى قيمة وأقل قيمة فى العينة، على عدد الفئات وهو ما تم إيضاحه سابقاً أنه يحسن أن يتراوح بين 10، 20 فئة. وهذا يعطى مدى الفئة وهو ما يجب أن يتسع لكل القيم.

ثالثاً: يتم تحديد المدى الخاص بالفئة الدنيا من بين كل الفئات. ومن الملائم والمعمول به فى بعض الحالات أن تبدأ الفئة الدنيا برقم يمثل مضاعفات لمدى الفئة ولو أن هذا لا يعتبر قاعدة لبدء الفئات. ومن المهم هنا أنه لتحديد حدود الفئات أن تبدأ برقم عشرى واحد أقل من ذلك الذى يتواجد فى المشاهدات، فمثلاً إذا كانت الفئة أرقاماً صحيحة فإن المدى للفئات يبدأ برقم عشرى واحد. فإذا كان المدى الكلى

مثلا 20 وحدة وحدد عدد الفئات المستخدمة بعشرة فئات فإن مدى الفئة يصبح $20/10 = 2$ وحدة. وبالتالى فإن الفئات سوف تبدأ من $9.5 - 11.5$ بدلاً من $10 - 12$ وبذلك فإن القيم 10 بالضبط سوف توضع فى فئة معينة حيث إن الفئة السابقة قد تكون $8 - 10$ وهذا لا يوضح موضع الرقم 10 .

رابعاً: بعد تحديد الفئة الدنيا يضاف إلى حدود الفئات اتساع (أو مدى) الفئات فتعطى حدود الفئة التى تليها ... وهكذا.

خامساً: تحسب مراكز الفئات وذلك بقسمة حاصل جمع مدى الفئة الدنيا على 2 حيث يعطى مركز الفئة الدنيا ثم يجمع مدى الفئات ليعطى مراكز الفئات المتتالية.

سادساً: يوضع أمام مركز كل فئة فى الجداول عدد الأفراد والمشاهدات التى تنتمى إلى هذه الفئة وهو ما يعرف بتكرار الفئة class frequency ويرمز له بالرمز f ويعطى حاصل جمع تكرارات الفئات مجموع المشاهدات.

وفى هذا المجال من الممكن أن تقسم التكرارات الخاصة بكل فئة على مجموع التكرارات معطية بذلك التكرار النسبى relative frequency distribution لكل فئة ومن هذه التكرارات النسبية ينشأ التوزيع التكرارى النسبى. وقد يمكن التعبير عن التكرارات النسبية فى صورة مئوية percentage distribution بالضرب فى 100 وهذه طرق مختلفة للتعبير عن التكرارات للفئات المختلفة.

وهناك طرق أخرى أيضا للتعبير عن التوزيع التكرارى للفئات المختلفة كما سبق ذكره فى الباب الثانى فهناك التوزيع التكرارى المتجمع cumulative frequency distribution. وفى هذا المجال يكون من المرغوب فيه معرفة عدد الأفراد التى تقل أو تزيد عن قيمة معينة. وعلى ذلك يتم تجميع تكرار كل فئة مع تكرارات جميع الفئات التى تسبقها أو تليها.

مثال ٣-١٩

فى دراسة عن العمر عند أول ولادة فى أبقار الفريزيان وجد أن العمر يتراوح بين 19 إلى 44 شهراً. ولقد رئى أن تقسم الأعمار إلى فترات تحتوى كل منها على 3 أشهر. على هذا الأساس وحتى لا يحدث هناك لبس فى وضع القيم بدأ بعمر 18 شهراً إلى 45 شهراً لمراحل الفئات. وبما أن الأعمار للأبقار كانت مقيدة إلى أقرب شهر كامل فإنه بذلك تصبح حدود الفئات أقل من المراحل برقم عشرى واحد وبذلك بدء فى حدود الفئات من 17.5 وحتى 44.5 شهراً.

يتم بعد ذلك تحديد مراكز الفئات وذلك بقسمة حدى الفئة على 2 وفى خانة أخرى من الجدول يوضع فيها تكرار كل فئة من الفئات أى عدد المشاهدات التى تتبع تلك الفئة. وقد يضاف إلى ذلك خانات أخرى توضح توزيعاً تكرارياً تكاملياً أو توزيعاً تكرارياً نسبياً ... الخ.

ويوضح جدول ٣-١ تكوين جدول تكرارى لعدد 962 عجلة موضحاً فيه أعمار الأبقار عند أول ولادة لها.

جدول ٣-١ حساب المتوسط والتباين من جداول التوزيع التكرارى

(١) مراحل الفئات	(٢) حدود الفئات	(٣) مراكز الفئات Y	(٤) التكرار f	(٥) الانحراف d	(٦) الانحراف X التكرار fd	(٧) مربع الانحراف X التكرار fd^2
21-18	20.5-17.5	19	6	-3	-18	54
24-21	23.5-20.5	22	20	-2	-40	80
27-24	26.5-23.5	25	187	-1	-187	187
30-27	29.5-26.5	28	340	0	0	0
33-30	32.5-29.5	31	181	1	181	181
36-33	35.5-32.5	34	129	2	258	516
39-36	38.5-35.5	37	65	3	195	585
42-39	41.5-38.5	40	27	4	108	432
45-42	44.5-41.5	43	7	5	35	175
المجموع			**962**		**532**	**2210**

حيث $d_i = \frac{Y_i - 28}{3}$ وتعتبر 28 وسطا فرضيا، أما قيمة 3 فتعبر عن مدى الفئة.

فمثلاً $-3 = \frac{-9}{3} = \frac{19 - 28}{3} = d_1$ مركز انحراف الفئة الأولى.

مثال ٣-٢٠

حساب المتوسط والانحراف القياسي من الجداول التكرارية.

تتلخص الطريقة فيما يلي:

١- تقدر الفئة التى تحتوى على المتوسط وفى كثير من الأحيان، خاصة فى حالة التوزيع الذى يقترب من الطبيعى، فإنه غالباً ما تستخدم تلك الفئة التى تمثل أعلى تكرار على أنها الفئة التى قد تحتوى على المتوسط (ولا يعتبر ذلك أمراً ضرورياً حيث إن طريقة الحساب سوف تصحح لذلك).

٢- يكون عمود (٥) فى جدول ٣-١ للانحرافات deviations ويرمز لها بالرمز d. وفى هذه الخانة يوضع قيمة الصفر فى تلك الفئة التى اعتبرت أنها تحتوى على المتوسط. ويعنى هذا أنه قد طرح من كل قيمة من الفئات المذكورة فى الجدول قيمة تساوى قيمة مركز الفئة التى افترض أنها تحوى المتوسط.

٣- توضع قيماً موجبة أو سالبة تمثل الانحرافات التى تسبق أو تلى الفئة التى تحتوى المتوسط (عمود ٥)، فإذا كانت الفئة أكبر توضع قيمة موجبة وفى حالة الأقل توضع قيمة سالبة. وهذا يعنى أنه قد قسمت كل مركز فئة على مراحل الفئات المستخدمة أى أن وحدات القياس تصبح ممثلة بوحدات لمراحل الفئات، فمثلاً تصبح القيم السالبة $-1,-2,\cdots$ فى حين تصبح القيم الموجبة $1,2,\cdots$. وتعبر d عن انحراف القيمة عن المتوسط الفرضى مقسوماً على مدى الفئة.

٤- يتم الحصول على القيمة الجبرية لحاصل ضرب تكرار كل فئة فى قيمة انحرافها أى fd وتوضع فى عمود مستقل. بعد ذلك يتم الجمع الجبرى لكل القيم أى $\sum fd$، ويجدر الإشارة هنا إلى أنه لو كانت قيمة هذا المجموع صفر فإن ذلك يعنى أن المتوسط الحقيقى كان يساوى قيمة مركز الفئة التى اعتبر أنها تحتوى المتوسط.

٥- يتم حساب المتوسط الحقيقى للعينة وذلك من المعادلة التالية:

$$\overline{Y} = M + \frac{i\sum fd}{\sum f} \qquad (١٦-٣)$$

حيث

M : هى مركز الفئة التى افترض أنها تحتوى المتوسط.

i : هى مرحلة الفئة.

$\sum f$: هى مجموع عدد المشاهدات فى العينة.

٦- يتم تربيع قيم الانحرافات فى كل فئة d^2 ثم ضربها فى تكرار كل فئة للحصول على قيمة fd^2 وتوضع هذه فى عمود آخر أو يتم ذلك بضرب قيمة $(d)(fd)$ فى كل فئة (عمود ٧). ثم بالجمع (وهنا كل القيم تكون موجبة نظراً لأنه قد ربعت قيم d) للحصول على قيمة $\sum fd^2$ أى مجموع حاصل ضرب مربعات الانحرافات فى التكرارات ويتم حساب الانحراف القياسى للعينة من المعادلة التالية:

$$S^2 = \frac{i^2[\sum fd^2 - (\sum fd)^2 / \sum f]}{\sum f - 1} \qquad (١٧-٣)$$

ومنها فإن الانحراف القياسى S عبارة عن الجذر التربيعى للتباين.

لاحظ أنه قد تختلف قيمة المتوسط المحسوب بهذه الطريقة اختلافاً بسيطاً عن قيمة المتوسط. إذا تم حسابه من المعادلة (٣-٢) وذلك لأنه فى المعادلة (٣-١٦) أعطيت كل القيم التى تنتمى إلى فئة معينة نفس القيمة والتى تمثل مركز الفئة الواقعة فيها وهذا قد يختلف عن استخدام كل قيمة بمفردها. ولكن فى حالة البيانات الكثيرة العدد فإن الاختلافات فى الحساب تصبح صغيرة ويمكن تجاهلها ولا توازى الجهد الزائد المبذول فى عمليات الحساب، خاصة إذا كانت أعداد الفئات كبيرة وبالتالى مراحلها ليست متسعة جداً. ونفس الشيء طبعاً سوف يحدث بالنسبة لقيمة الانحراف القياسى.

مثال ٣-٢١

استكمالا للمعلومات الموجودة فى المثال ٣-١٩ فإنه يتم حساب المتوسط $\overline{Y}$ من ضرب القيمة الموجودة فى عمود (٤) وتمثل التكرارات فى قيم العمود (٥) وتمثل الانحرافات وهذه مذكورة فى العمود (٦) والتى يتم تجميعها فى نهاية هذا العمود ويرمز لها بالرمز $\sum fd$. وبتطبيق المعادلة (٣-١٦) يمكن الحصول على قيمة المتوسط كالتالى:

بما أن قيمة $M = 28$ ، $i = 3$

إذا $$\overline{Y} = 28 + \frac{(3)(532)}{962} = 29.66$$

أى أن متوسط العمر عند الولادة الأولى كان 29.66 شهراً

ومن استخدام العمود (٧) يمكن تطبيق المعادلة (٣-١٧) لحساب التباين كالتالى:

$$S^2 = 3^2\left[\frac{2210-(532)^2/962}{961}\right] = 17.942$$

ويكون الانحراف القياسى: $S = \sqrt{17.942} = 4.236$

أى أن الانحراف القياسى لهذه الصفة وهى العمر عند أول ولادة يساوى 4.236 شهراً وهنا تجدر الإشارة إلى أن هذه الطريقة قد وفرت وقتاً كثيراً كان سوف يستخدم فى معالجة البيانات فردياً. ويمكن حساب معامل الاختلاف كالتالى:

$$C.V. = \frac{4.236}{29.66} X100 = 14.3\%$$

تمارين الباب الثالث

٣-١ لديك بيانات تمرين ٢-١ والمطلوب:

أ- حساب مقاييس النزعة المركزية لمجموعة البيانات وهى كل من المتوسط الحسابى، الوسيط ، المنوال، المتوسط الهندسى.

ب- احسب المدى من خلال جدول ٢ ملحق أ، ما هو تقديرك لقيمة الانحراف القياسى.

ج- احسب التباين والانحراف القياسى.

د- احسب قيمة معامل الاختلاف وكل من تباين المتوسط والخطأ القياسى.

٣-٢ إذا كان القطاع المأخوذ فى عظمة الفخذ دائرياً وأن مساحته تقدر بواسطة المعادلة م = ط نق^٢ حيث ط = 3.14 فأوجد المتوسط والتباين ومعامل الاختلاف لمساحة المقطع إذا كانت القيم التالية تمثل نصف القطر بالسنتيمتر:

2.1, 2.9, 2.8, 2.4, 3.6, 2.8, 2.9, 3.5, 3.1

٣-٣ إذا كانت قيمة $\overline{Y} = 40$ ومعامل الاختلاف 20.7% وحجم العينة $n = 16$ فرداً. احسب مجموع المربعات الكلى غير المصحح والخطأ القياسى.

٣-٤ البيانات التالية توضح عدد بويضات لطفيل ما والتى لوحظت فى مجموعة من العينات والمطلوب حساب كل من المتوسط والانحراف والخطأ القياسى لها.

حدود الفئة	التكرار	حدود الفئة	التكرار
5 - 1	3	40 - 36	25
10 - 6	5	45 - 41	22
15 - 11	7	50 - 46	19
20 - 16	18	55 - 51	6
25 - 21	32	60 - 56	6
30 - 26	45	65 - 61	3
35 - 31	38	70 - 66	1

٣-٥ إذا كان متوسط الفرق فى ضغط الدم قبل بدء العلاج بدواء معين وبعده يساوى 2 مم فى تجربة ما قيس فيها ضغط الدم على 12 فرداً من مرضى ضغط الدم وشوهدت 11 قيمة فقط من هذه الفروق فوجدت كالتالى:

3, 2, 8, 2, 1, -3, -7, 10, 1, 7, 8

احسب القيمة الغائبة والانحراف القياسى ومعامل التباين وتباين المتوسط والخطأ القياسى.

٣-٦ اكتب القيم التالية بصورة كاملة (مفكوك الكمية):

(أ) $\sum_{i=5}^{10} b_i^2$ (ب) $\sum_{i=2}^{4}(C_i - 1)$ (ج) $\sum_{m=1}^{5} 5(X_m - 3)$

٣-٧ إذا كانت قيمة $X_1 = 4$ ، $X_2 = 3$ ، $X_3 = 6$ ، $X_4 = -1$

فأوجد قيمة ما يلى:

(أ) $\sum_{i=1}^{4} X_i^2 - 6$ (ب) $\sum_{i=1}^{4}(X_i - 3)$

(ج) $\sum_{i}^{4}(X_i + 1)^2$ (د) $\sum_{i=1}^{4} X_i^2 - 4\left(\frac{\sum_{i}^{4} X_i}{4}\right)^2$

٤ الاحتمال والتوزيعات الاحتمالية

Probability and probability distributions

١- تعريف الاحتمال

٢- قواعد الاحتمال

٣- الاحتمال الشرطى

٤- الاحتمالات اللاحقة

٥- التوزيعات الاحتمالية

٦- مقدمة عن التباديل والتوافيق

٧- توزيع "ذو الحدين"

٨- توزيع بواسون

٩- التوزيع الطبيعى

١٠- التوزيع الطبيعى القياسى

١١- توزيع المتوسطات

١٢- تقريب توزيع "ذو الحدين" بالتوزيع الطبيعى

من منا لا يستخدم كلمة احتمال probability فى حياته اليومية ولكن قد تكون فى عبارات مختلفة فيقال إن هذين الزوجين قد ينجبان ذكراً أو أن هذا الطالب يكاد يكون من المؤكد نجاحه فى مادة ما أو أن هذا الطفل الصغير قد يصير طويلاً وغير ذلك كثير. وهذه العبارات إما مؤكدة الحدوث أو مؤكدة فى عدم حدوثها أو تقع بين هاتين الفئتين. وتقوم نظرية الاحتمالات probability theory بترجمة هذه العبارات الوصفية إلى قياسات كمية حيث إنها الأكثر دقة، فيقال مثلاً إن احتمال الحصول على ذكر يساوى النصف أى 50% واحتمال أن هذا الطالب المؤكد نجاحه هو الواحد الصحيح أو 100% ... وهكذا. ويطلق على إنجاب ذكر أو نجاح الطالب فى مادة ما أو أن يكون الطفل طويلاً "حدث" event وفى مثل هذه الحالة يسمى حدث بسيط simple event، وقد يكون الحدث مركباً compound event كما فى حالة إلقاء حجرى نرد مرة واحدة فإن احتمال الحصول على الوجه العلوى 6 للحجرين هو احتمال الحصول على حدث مركب. وللحصول على نسبة حدوث هذا الحدث أو أى حدث فقد تجرى تجربة experiment والتى قد تشتمل على عدة محاولات trials وعندما تجرى محاولة يتحصل على عدد مختلف من الحالات الممكن ظهورها وتسمى كل الحالات المختلفة الممكن حدوثها (جميع النواتج الممكنة) all possible outcomes. فمثلاً عند إلقاء حجر النرد فإن أى رقم من الأرقام التالية يظهر على وجهه العلوى: 1 أو 2 أو 3 أو 4 أو 5 أو 6 وتمثل هذه الأحداث كل النواتج الممكن ظهورها. ويعبر عن كل الأحداث الممكنة بفراغ العينة sample space وقد يمثل بمحور واحد one dimensional space كما هو الحال عند إلقاء حجر نرد مرة واحدة حيث سيكون هناك محور واحد عليه الأرقام من 1 إلى 6، وقد يمثل فراغ العينة بمحورين two dimensional space كما فى حالة إلقاء حجرى نرد مرة واحدة فالمحور السينى مثلاً يمثل حجر النرد الأول والمحور الصادى يمثل الحجر الثانى وقد يمثل فراغ العينة بأكثر من محورين.

وقد بدأت دراسة الاحتمالات من عدة مئات من السنين وكان هناك اهتمام بها وخاصة في الألعاب الميسرية التى تتوقف على الصدفة chance. وتلعب الاحتمالات دوراً مهماً فى الإحصاء وفى اختبارات الفروض الإحصائية tests of hypotheses وفى التقديرات estimations وغيرها الكثير.

٤-١ تعريف الاحتمال

هناك نوعان من التعريفات أحدهما هو التعريف الكلاسيكى والآخر هو التعريف التجريبى.

٤-١-١ التعريف الكلاسيكى Classical definition

وفيه يعرف احتمال حدوث حدث ما وليكن هذا الحدث E مثلاً بأنه النسبة بين عدد مرات ظهور هذا الحدث المرغوب فيه desired cvent (n) إلى عدد كل النواتج الممكنة (m)، ويعبر عن ذلك كما يلى:

$$\Pr(E) = \frac{n}{m} \qquad (٤-١)$$

ومن هذا التعريف فإن أقل قيمة للاحتمال هى الصفر عندما يكون عدد مرات ظهور الحدث المرغوب فيه مساوياً للصفر وأعلى قيمة للاحتمال هى الواحد الصحيح عندما يكون عدد مرات ظهور الحدث المرغوب فيه (n) مساوياً لعدد كل النواتج الممكنة (m).

مثال ٤-١

ما هو احتمال الحصول على صورة (H) عند رمى قطعة نقود مرة واحدة ؟

وللحصول على قيمة هذا الاحتمال فإن عدد مرات ظهور الحدث المرغوب فيه هو 1 والعدد الكلى للحالات الممكن حدوثها هو 2 وعلى ذلك فإن احتمال الحصول على صورة يساوى ½ ، مع ملاحظة أن قطعة النقود هذه يجب أن تكون غير متحيزة fair، بمعنى عند رمى هذه القطعة فإنها تستقر إما على الصورة أو على الكتابة بنفس درجة التوقع.

ويجب ملاحظة أنه عند استخدام هذا التعريف الكلاسيكى للاحتمال فإنه يشترط أن تكون الأحداث متنافية mutually exclusive، أى حدوث أى منها يمنع حدوث الأخرى وأن يكون لكل حدث نفس الفرصة فى الظهور equally likely.

مثال ٤-٢

ما هو احتمال الحصول على ولد Jack عند سحب ورقة من أوراق اللعب (الكوتشينة) ؟

عدد الأوراق التى عليها Jack = 4

العدد الكلى للأوراق = 52 و هذا يمثل كل الأحداث الممكنة

إذا احتمال الحصول على Jack $4 \div 52 = 0.0769$

مثال ٤-٣

فى المثال ٤-١ ما هو احتمال الحصول على كتابة (T) ؟

عدد مرات الحصول على كتابة عند رمى قطعة واحدة من النقود = 1 والعدد الكلى الممكن حدوثه هو 2 أى أن:

$$m = 2 \ ، \ \acute{n} = 1$$

إذا احتمال الحصول على كتابة: $1 \div 2 = 0.5$

يتضح مما سبق عدة حقائق حول الاحتمالات:

١- إذا أجريت محاولة لها k من الحالات الممكنة والمتساوية فى إمكانية حدوثها فإن احتمال الحصول على أى منها هو:

$$\Pr(E_i) = \frac{1}{k}, i = 1,2,...,k \qquad (٤-٢)$$

كما ظهر ذلك فى الأمثلة ٤-١، ٤-٣.

٢- احتمال حدوث حدث ما E_i أكبر من أو يساوى الصفر وأقل من أو يساوى الواحد الصحيح أى أن:

$$0 \le \Pr(E_i) \le 1 \qquad (٤-٣)$$

٣- مجموع الاحتمالات لكل الأحداث الممكنة يساوى الواحد الصحيح:

$$\sum_i \Pr(E_i) = 1 \qquad (٤-٤)$$

فحاصل جمع احتمال الحصول على صورة واحتمال الحصول على كتابة عند رمى قطعة من النقود يساوى الواحد الصحيح وعند إلقاء حجر نرد مرة واحدة فإن احتمال الحصول على أى رقم يساوى $\frac{1}{6}$ ومجموع الاحتمالات لكل الأحداث الممكنة هو الواحد الصحيح أيضاً ... وهكذا.

٤- احتمال عدم الحصول على حدث ما وليكن $E_i = k$ يساوى الواحد الصحيح مطروحاً منه احتمال الحصول على هذا الحدث أى أن:

$$\Pr(E_i \ne k) = 1 - \Pr(E_i = k) \qquad (٤-٥)$$

مثال ٤-٤

عند إلقاء حجر نرد مرة واحدة ما هو احتمال عدم الحصول على الرقم 6 ؟

احتمال الحصول على الرقم 6 $\frac{1}{6} = 0.166$

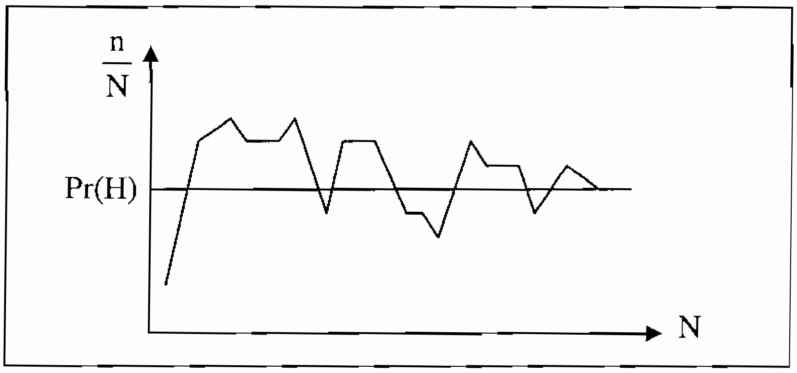
واحتمال عدم الحصول على الرقم 6 $1 - \frac{1}{6} = \frac{5}{6} = 0.834$

٤-١-٢ التعريف التجريبي Experimental or empirical definition

إذا أجريت تجربة من عدة محاولات في كل محاولة يرمى مثلاً عدد من قطع النقود N، ظهرت الصورة (H) في n منها فإنه عندما تؤول N إلى ∞ (ما لانهاية) فإن النسبة $\frac{n}{N}$ تعبر عن قيمة الاحتمال أي أن:

$$Pr(H) = \lim_{N \to \infty} \frac{n}{N} \qquad (٤-٦)$$

ويمكن تمثيل ذلك بيانيا بالشكل ٤-١ إذا كان المحور السيني يمثل N والمحور الصادي يمثل النسبة $\frac{n}{N}$.



شكل ٤-١ التمثيل البياني للتعريف التجريبي للاحتمال

ومن هذا التعريف التجريبي للاحتمال يتضح أنه كلما زادت N كلما أمكن الحصول على قيمة دقيقة للاحتمال بعيداً عن تأثير الصدفة chance التي قد تلعب دوراً كبيراً إذا كانت N صغيرة.

مثال ٤-٥

إذا كان عدد العجول الذكور المولودة 4900 علماً بأن العدد الكلي 10000 عجل من الذكور والإناث. فما احتمال الحصول على عجل ذكر ؟

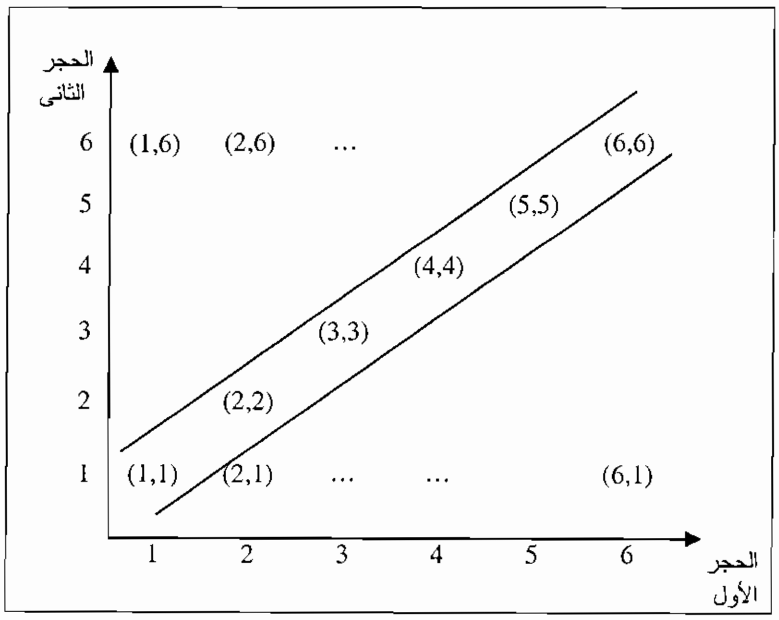
احتمال الحصول على عجل ذكر $\frac{4900}{10000} = \frac{49}{100} = 0.49$

وأبسط طريقة للحصول على قيمة الاحتمال لحدوث حدث ما وليكن E_i هو كتابة كل الحالات المختلفة الممكن ظهورها ثم تحصر عدد الحالات التى يظهر فيها الحدث المرغوب E_i وباستخدام تعريف الاحتمال الكلاسيكى السابق الإشارة إليه يمكن حساب قيمة الاحتمال المراد تقديره.

مثال ٤-٦

عند إلقاء حجرى نرد ما احتمال الحصول على الوجهين متشابهين؟

افترض أن الإحداث السينى يمثل الحجر الأول والإحداث الصادى يمثل الحجر الثانى حيث تكون كل النقط الممكن ظهورها 36 وعدد المرات التى يظهر فيها الوجهان متشابهين 6 مرات وعليه فإن احتمال الحصول على الوجهين متشابهين هو: $\frac{1}{6} = \frac{6}{36}$ كما هو واضح من شكل ٤-٢



شكل ٤-٢ النقط الممكن ظهورها عند إلقاء حجرى نرد مرة واحدة

ولكن حساب الاحتمال بتلك الطريقة قد يحتاج إلى جهد ووقت لحصر كل الحالات المختلفة الممكن حدوثها ولذا تستخدم قواعد الاحتمالات التى وضعت لتوفر الوقت والجهد.

٤-٢ قواعد الاحتمال Rules of probability

٤-٢-١ قاعدة الجمع Addition rule

إذا ألقى حجرا نرد غير متحيزين مرة واحدة. فما احتمال الحصول على وجهين متشابهين (هو نفس المثال ٤-٦)؟ ويمكن التعبير عن ذلك كالآتي:

ما قيمة احتمال الحصول على: (1, 1) أو (2, 2) أو (3, 3) أو (4, 4) أو (5, 5) أو (6, 6) ؟

لاحظ أنه إذا أُلقى حجرى نرد فإن الحصول على (1,1) على كلا الوجهين للحجرين يمنع ظهور الحالات الأخرى الممكنة. والحصول على (2,2) يعنى أيضا عدم ظهور الحالات الأخرى، وهكذا فإن ظهور أى حدث يمنع ظهور الأحداث الأخرى. ويعبر عن تلك الخاصية بأنها أحداث متنافية mutually exclusive ويكون احتمال الحصول على هذه الأحداث المتنافية والتى لها نفس الفرصة فى الظهور هو مجموع احتمالات كل منها ويعبر عن ذلك كما يلى:

$$\begin{aligned}\Pr[(1,1) \text{ or } (2,2) \cdots \text{ or } (6,6)] &= \Pr(1,1) + \Pr(2,2) + \cdots + \Pr(6,6) \\ &= \frac{1}{36} + \frac{1}{36} + \cdots + \frac{1}{36} = \frac{1}{6}\end{aligned}$$

وهى نفس النتيجة السابق الحصول عليها فى مثال ٤-٦. ويمكن تعميم هذه القاعدة لأى من الأحداث المتنافية أى:

$$\Pr(E_1 \text{ or } E_2 \cdots \text{ or } E_k) = \Pr(E_1) + \Pr(E_2) + \cdots + \Pr(E_k) \qquad (٤-٧)$$

أما إذا كانت الأحداث غير متنافية (وهذا شرط أساسى يلزم توفره) فإن النتيجة سوف تختلف وسوف تكون الإجابة خطأ، والمثال التالى يبين ذلك:

مثال ٤-٧

ما احتمال الحصول على وجهين متشابهين أو أن يكون مجموع الوجهين هو 2 عند إلقاء حجرى نرد مرة واحدة ؟

عند النظر للرسم البيانى (شكل٤-٢) فى المثال ٤-٦ وحصر عدد النقط التى تمثل الحصول على الوجهين متشابهين (6 نقاط) أو أن يكون مجموع الوجهين هو 2 (نقطة واحدة) وعليه فإن الاحتمال عبارة عن: $\frac{6}{36}+\frac{1}{36}=\frac{7}{36}$ وهى قيمة غير مضبوطة للاحتمال. وللحصول على القيمة المضبوطة تستخدم العلاقة التالية:

$$\Pr(A \text{ or } B) = \Pr(A) - \Pr(B) - \Pr(AB) \qquad (٤-٨)$$

وتعبر Pr(AB) عن الاحتمال المشترك joint probability وهو احتمال الحصول على الحدث A والحدث B معاً أى نسبة عدد المرات التى يظهر فيها الحدث A والحدث B معاً إلى العدد الكلى الممكن حدوثه. وقد يعبر عن المعادلة (٤-٨) كما يلى:

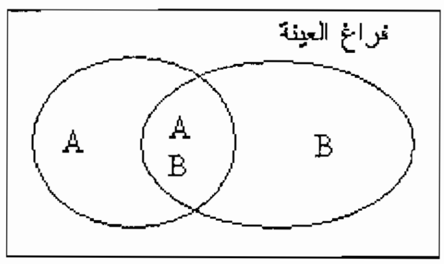
$$\Pr(A \cup B) = \Pr(A) - \Pr(B) - \Pr(A \cap B) \qquad (٤-٩)$$

حيث تمثل $\cup$ اتحاد union بينما تمثل $\cap$ تقاطع intersection.

وتكون الإجابة الصحيحة للمثال ٤-٧ كالتالى:

$$\frac{6}{36}+\frac{1}{36}-\frac{1}{36}=\frac{1}{6}$$

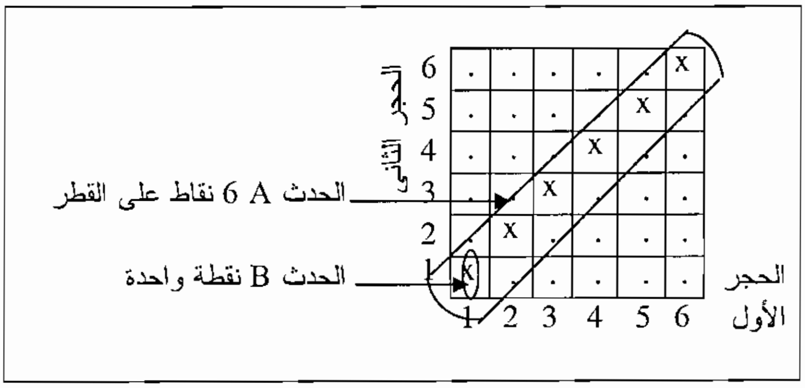
ويمكن التعبير عن المعادلة (٤-٩) بالشكل ٤-٣ المعروف باسم Venn diagram



شكل ٤-٣ تخطيط (شكل) فن Venn diagram

حيث إن احتمال الحصول على الحدث A أو الحدث B هو احتمال الحصول على الحدث A مضافاً إليه احتمال الحصول على الحدث B مطروحاً منه احتمال الحصول على كليهما معاً.

ولتوضيح هذا الشكل فإنه فى المثال ٤-٧ يمكن التعبير عن كل النقط الممكنة وعددها 36 بفراغ العينة sample space ويعبر عن عدد النقط التى يظهر فيها الوجهان متشابهين بالحدث A وهى 6 نقاط وعدد النقط التى فيها مجموع الوجهين يساوى 2 بالحدث B وعددها نقطة واحدة ويمكن بيان ذلك بالشكل ٤-٤ التالى:



شكل ٤-٤ عدد النقاط التى يمثلها الحدث A وعدد النقاط التى يمثلها الحدث B

وفى هذه الحالة فإن احتمال الحصول على الحدث B هو نفسه احتمال الحصول على الحدثين A، B معاً.

٤-٢-٢ قاعدة الضرب Multiplication rule

عند إجراء عدد من المحاولات المستقلة independent trials فإن احتمال الحصول على حدثين مستقلين معاً هو حاصل ضرب احتمال الحصول على الحدث الأول فى احتمال الحصول على الحدث الثانى أى:

$$\Pr(E_1 \text{ and } E_2) = \Pr(E_1) \cdot \Pr(E_2) \qquad (٤-١٠)$$

و للتعميم فإن:

$$\Pr(E_1 \text{ and } E_2 \text{ and} \ldots E_k) = \Pr(E_1) \cdot \Pr(E_2) \cdot \ldots \cdot \Pr(E_k) \qquad (٤-١١)$$

حيث إن الأحداث من E_1 إلى E_k أحداث مستقلة independent events وتعرف بأنها الأحداث التى إذا حدث أى منها لا يؤثر على ظهور الأحداث الأخرى.

ويلاحظ هنا استخدام كلمة and وهى تعنى مع فى حين أنه فى حالة الأحداث المتنافية استعملت كلمة or وتعنى أو.

مثال ٤-٨

عند إلقاء حجرى نرد معاً، ما احتمال الحصول على رقم 6 على الوجهين؟

حيث إن ظهور رقم 6 على وجه حجر النرد الأول مستقل ولا يمنع ظهور الرقم 6 أيضاً على وجه حجر النرد الثانى أى أنهما حدثان مستقلان فإن:

احتمال الحصول على $(6,6)$ عند رمى حجرى النرد معاً: $\frac{1}{6}\times\frac{1}{6}=\frac{1}{36}$

وهى نفس قيمة الاحتمال لو استخدمت الطريقة المبسطة السابق الإشارة إليها كما فى المثال ٤-٦ حيث إن عدد النقط التى فيها الوجهين $(6,6)$ هو واحد والعدد الكلى 36 وبالتالى فهى تعطى نفس النتيجة.

مثال ٤-٩

فى دراسة الوراثة فى النباتات، عند تزاوج الفرد الخليط وتركيبه الوراثى Aa مع نفسه (التلقيح الذاتى)، ما احتمال الحصول على فرد تركيبه AA؟

الحل:

الفرد Aa يعطى نوعين من الجاميطات هما A، a وبالتالى فإن

احتمال الحصول على جاميطة $A = 0.5$

احتمال الحصول على جاميطة $a = 0.5$

وحيث إن الفرد AA ينتج عندما يحصل على A من الأب وأيضا يحصل على A من الأم، وحيث إن حصوله على أى منهما مستقل ولا يؤثر فى احتمال حصوله على الجاميطة الأخرى فإن احتمال الحصول على فرد تركيبه AA $0.25 = (0.5)(0.5)$

مثال ٤-١٠

فى مثال ٤-٩ ما هو احتمال الحصول على فرد تركيبه الوراثى Aa؟

الحل:

استخدام قاعدتى الضرب والجمع معاً يعطى قيمة هذا الاحتمال. فالفرد الذى تركيبه Aa قد ينتج من جاميطة A من الأب وجاميطة a من الأم أو قد ينتج من اتحاد جاميطة A من الأم مع جاميطة a من الأب. لاحظ أن الاحتمالين متنافيان وعليه

فإن احتمال الحصول على هذا الفرد هو: $\frac{1}{2}\times\frac{1}{2}+\frac{1}{2}\times\frac{1}{2}=\frac{1}{2}$

مثال ٤-١١

فى المثال ٤-٩، إذا كان الجين A سائداً على a ويتسبب فى طول الأفراد فما احتمال الحصول على فرد طويل إذا كان كل من AA ، Aa أفراداً طويلة بينما aa أفراداً قصيرة ؟

الحل:

الفرد الطويل إما أن يكون تركيبه الوراثى AA أو تركيبه Aa وكما سبق احتمال الحصول على التركيب الوراثى Aa هو ½ واحتمال الحصول على AA هو ¼، وعلى ذلك فإن احتمال الحصول على فرد طويل $\frac{3}{4}=\frac{1}{4}+\frac{1}{2}$

ويمكن الحصول على نفس النتيجة بالطريقة المبسطة التالية:

		جاميطات الأب	
		a	A
جاميطات الأم	A	Aa طويل	AA طويل
	a	aa قصير	Aa طويل

ومنها يتضح وجود 3 طويل وواحد قصير وعليه فإن احتمال الحصول على فرد طويل = ¾.

مثال ٤-١٢

عند تزاوج فردين تركيبهما الوراثى BB، bb فما احتمال:

١ - الحصول على فرد تركيبه BB؟

٢ - الحصول على فرد تركيبه Bb؟

احتمال الحصول على فرد تركيبه BB يساوى احتمال أن يحصل هذا الفرد على B من كل من الأم والأب ، وحيث إنه يمكن أن يحصل على B من الأب BB فإنه يلزم أن يحصل على B (الجاميطة الأخرى) من الأب الآخر وهذا مستحيل لأن الفرد الآخر bb ولا يعطى إلا نوعاً واحداً من الجاميطات وهو b وعليه فإن:

احتمال الحصول على فرد تركيبه BB هو $(1)(0) = 0$، ويطلق على هذا الحدث حدثاً مستحيلا impossible event.

أما احتمال الحصول على فرد تركيبه Bb:

$$\Pr(Bb) = \Pr(B \text{ from one parent}) \cdot \Pr(b \text{ from the other parent}) = (1)(1) = 1$$

ويطلق على هذا الحدث حدثاً مؤكداً sure event.

صندوق ٤-١

قانون جمع الاحتمالات:

إذا كان هناك حدثان متنافيان فإن احتمال حدوث أحدهما أو الآخر هو حاصل جمع احتمالهما منفردا

$$\Pr(E_1 \text{ or } E_2) = \Pr(E_1) + \Pr(E_2)$$

قانون ضرب الاحتمالات:

إذا كان هناك حدثان مستقلان فإن احتمال حدوث الأول و الثانى هو حاصل ضرب احتمالهما منفردا

$$\Pr(E_1 \text{ and } E_2) = \Pr(E_1) \cdot \Pr(E_2)$$

٤-٣ الاحتمال الشرطى Conditional probability

إذا كانت هناك محاولتان غير مستقلتين non-independent أى متتابعتين فإن احتمال حدوث الحدث E_1 فى أول محاولة والحدث E_2 فى ثانى محاولة معاً هو:

$$\Pr(E_1E_2) = \Pr(E_1) \cdot \Pr(E_2 / E_1) \qquad (٤-١٢)$$

الحد الثانى من الطرف الأيمن يطلق عليه الاحتمال الشرطى:

$$\Pr(E_2 / E_1) = \Pr(E_2 \text{ given that } E_1 \text{ has happened})$$

أى احتمال حدوث الحدث E_2 علما بأن الحدث E_1 قد حدث.

ومن العلاقة (٤-١٢) يمكن استنتاج أن:

$$Pr(E_2 / E_1) = Pr(E_2E_1) / Pr(E_1) \qquad (٤-١٣)$$

وأيضاً:

$$Pr(E_1 / E_2) = Pr(E_2E_1) / Pr(E_2) \qquad (٤-١٤)$$

مع توفر شرط أن $Pr(E_1)$ وأيضاً $Pr(E_2)$ لا تساويان الصفر.

فإذا كان $Pr(E_1 / E_2) = Pr(E_1)$ فهذا معناه أن الحدثين E_1، E_2 مستقلان.

مثال ٤-١٣

صندوق به 5 كرات حمراء، 10 كرات بيضاء. إذا سحب كرتان، ما احتمال الحصول على الأولى بيضاء (W) والكرة الثانية حمراء (R) ؟ علماً بأن الكرة التى تسحب لا ترد مرة أخرى إلى الصندوق without replacement ؟.

احتمال الحصول على الكرة الأولى بيضاء $Pr(W) = \frac{10}{15}$

احتمال الحصول على الكرة الثانية حمراء لو كانت الكرة الأولى بيضاء $Pr(R/W) = \frac{5}{14}$.

إذاً احتمال الحصول على الكرة الأول بيضاء والثانية حمراء هو:

$$Pr(WR) = Pr(W) Pr(R / W) = \frac{10}{15} x \frac{5}{14} = \frac{5}{21}$$

٤-٤ الاحتمالات اللاحقة (صيغة بيز)

Aposteriori probability (Bayes' formula)

كان الحديث فيما سبق عن الاحتمالات المسبقة apriori probability والتى كانت تعنى على سبيل المثال أنه إذا سحبت كرة من صندوق به كرات حمراء وأخرى بيضاء فالمطلوب هو معرفة احتمال أن تكون الكرة حمراء (مثلا). أما فى حالة الاحتمالات اللاحقة للحدث aposteriori probability فهى تعنى أنه إذا سحبت كرة وكانت فعلا حمراء وكان هناك عدد من الصناديق بكل منها كرات حمراء وبيضاء فما هو الصندوق الأكثر احتمالاً فى أن تكون الكرة قد سحبت منه ؟ ولإيجاد قيمة ذلك

الاحتمال تستخدم صيغة بيز Bayes' formula، ولمزيد من الدراسة يمكن الرجوع إلى واحد أو أكثر من المراجع المدونة فى نهاية هذا الكتاب.

مثال ٤-١٤

إذا كان يوجد صندوقان الأول به 5 كرات حمراء، 4 كرات بيضاء، والثانى به 12 كرة حمراء وكرة واحدة بيضاء، علماً بأن احتمال السحب من الصندوق الأول يساوى احتمال السحب من الصندوق الثانى وسحبت كرة واحدة وكانت حمراء فما احتمال أن تكون الكرة الحمراء مسحوبة من الصندوق الأول ؟

فى هذه الحالة تستخدم Bayes' formula حيث:

$$\Pr(B_i/E) = \frac{\Pr(B_i)\ \ \Pr(E/B_i)}{\sum \Pr(B_i)\ \ \Pr(E/B_i)} \qquad (٤-١٥)$$

حيث:

$\Pr(B_i/E)$ تعنى احتمال السحب من الصندوق B_i علماً بأن الكرة التى سحبت كانت حمراء E فى هذا المثال.

$\Pr(B_i)$ تعنى احتمال السحب من الصندوق B_i

$\Pr(E/B_i)$ تعنى احتمال سحب كرة حمراء E علماً بأن الصندوق B_i هو الذى سحبت منه الكرة.

وعلى ذلك:

احتمال سحب كرة حمراء من الصندوق الأول $\Pr(E/B_1) = \frac{5}{9}$

احتمال سحب كرة حمراء من الصندوق الثانى $\Pr(E/B_2) = \frac{12}{13}$

احتمال أن يكون السحب من الصندوق الأول = احتمال أن يكون السحب من الصندوق الثانى

$$\Pr(B_1) = \Pr(B_2) = \frac{1}{2}$$

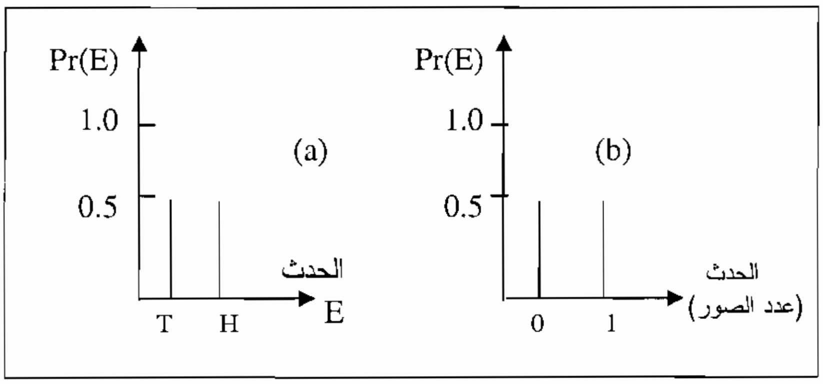
ومن العلاقة (٤-١٥) فإن احتمال أن تكون الكرة الحمراء قد سحبت من الصندوق الأول هو:

$$Pr(B_1/E) = \frac{\frac{1}{2}\times\frac{5}{9}}{\frac{1}{2}\times\frac{5}{9}+\frac{1}{2}\times\frac{12}{13}} = \frac{65}{173}$$

واحتمال أن تكون الكرة الحمراء قد سحبت من الصندوق الثانى $= \frac{108}{173}$

٤-٥ التوزيعات الاحتمالية Probability distributions

عند رمى قطعة نقود غير متحيزة fair يوجد احتمالان: احتمال الحصول على صورة Pr(H) أواحتمال الحصول على كتابة Pr(T) وكل منهما يساوى النصف. والاحتمالات المختلفة للأحداث الممكن حدوثها موضحة بشكل ٤-٥.



شكل ٤-٥ احتمالات الأحداث التى يمكن حدوثها عند رمى قطعة نقود

ويطلق على احتمالات كل الأحداث الممكنة بالتوزيع الاحتمالى. ويمكن أن يعبر عن التوزيع الاحتمالى عن طريق تكوين متغير جديد dummy variable يمثل عدد حالات النجاح مثلاً ولتكن عدد الصور التى تظهر فى كل ناتج من كل النواتج الممكنة، ففى مثل هذه الحالة فإن عدد الصور قد تكون مساوية للصفر باحتمال ½ أو مساوية للواحد الصحيح باحتمال قدره ½ أيضاً. ويمكن تمثيل ذلك بما يلى:

النواتج	عدد مرات ظهور صورة	الاحتمال
H	1	½
T	0	½

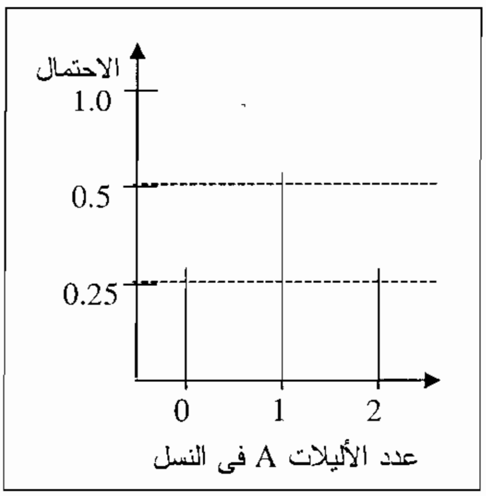
بمعنى أن

$$\Pr(X = x_i) = \frac{1}{2} \text{ , } x_i = 0,1$$

أى احتمال أن المتغير X يأخذ قيمة معينة ولتكن x_i يساوى النصف. وأيضاً فى حالة تزاوج Aa مع نفسه فإن احتمالات كل النواتج الممكنة للحدث تكون ما يسمى بالتوزيع الاحتمالى والذى يكون كالتالى:

$$\Pr(X = x_i) = \frac{1}{4} \qquad \text{for } x_i = 0,\ 2$$
$$= \frac{1}{2} \qquad \text{for } x_i = 1$$

ويمكن تمثيل ذلك بيانيا بالشكل ٤-٦ كالتالى:



شكل ٤-٦ التوزيع الاحتمالى للنسل الناتج من تزاوج Aa مع نفسه

وفى مثل تلك الحالات يعبر عن $\Pr(X = x_i)$ بدالة كثافة الاحتمال أو الدالة الاحتمالية probability or density function.

مع تحيات د. سلام حسين الهلالي salamalhelali@yahoo.com

أما إذا كان المطلوب معرفة احتمال الحصول على فرد يحتوى على أليل a على الأكثر فإن ذلك يمكن أن يعبر عنه بالآتى:

$$F(X) = Pr(X \le x_i) \quad , x_i \le 1$$

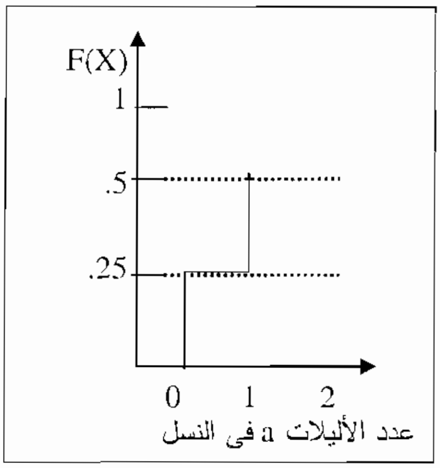
و حيث إن x_i إما أن تساوى الصفر أو الواحد فى هذه الحالة فإن:

$$F(X) = Pr(x_i = 0) + Pr(x_i = 1) = \sum Pr(x_i)$$

وتسمى هذه الدالة بالدالة التجميعية cumulative distribution function وهى تمثل احتمال الحصول على aa أو Aa حيث إن بكل منها الأليل a وفى هذه الحالة فإن:

$$F(X) = Pr(X = 0) + Pr(X = 1)$$
$$= \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$$

ويمكن تمثيل ذلك بيانيا بالشكل ٤-٧



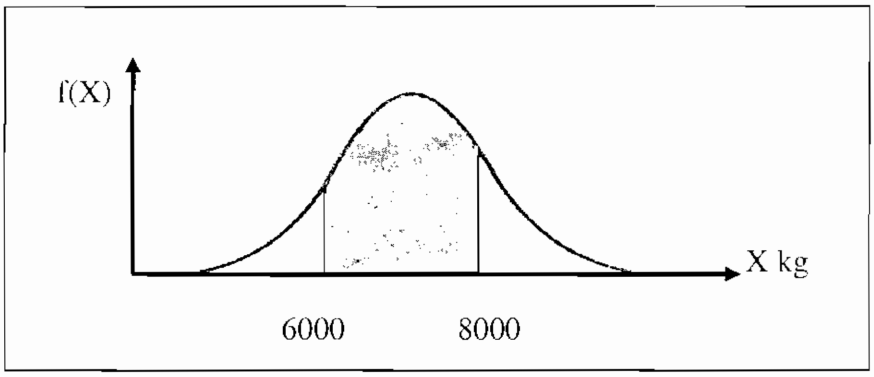
شكل ٤-٧ الدالة التجميعية للحصول على فرد يحتوى على a على الأكثر

والصورة العامة للدالة التجميعية للمتغيرات المتقطعة هى:

$$F(X) = Pr(X \le x_i)$$
$$= \sum_{x_i} Pr(X = x_i)$$

وأقصى قيمة يمكن أن تأخذها الدالة التجميعية هى الواحد الصحيح وأقل قيمة هى الصفر.

والتوزيع الاحتمالى قد يكون لمتغير متقطع discrete variable كما هو الحال فى الأمثلة السابقة وكما فى حالة قطعة النقود وأوراق اللعب والتلقيح الذاتى للفرد Aa. وقد يكون التوزيع الاحتمالى لمتغير متصل (مستمر) continuous variable ومن أمثلة المتغيرات المستمرة القياسات الكمية. ومن أمثلة تلك القياسات: وزن الأفراد، الطول، مستوى هرمون معين فى الدم وكمية اللبن التى تعطيها البقرة. فاحتمال أن تعطى بقرة ما بين 6000 إلى 8000 كج من اللبن فى الموسم هو المساحة المظللة بالنسبة للمساحة الكلية تحت المنحنى (شكل ٤-٨) ويمكن الحصول عليها بحساب التكامل.



شكل ٤-٨ التوزيع الاحتمالى لمتغير متصل (كمية اللبن بالكيلوجرام)

فإذا كانت دالة هذا التوزيع هى f(x) فإن الاحتمال يكون:

$$\int_{6000}^{8000} f(x)\,dx$$

أما الصورة العامة للدالة التجميعية فى التوزيعات المستمرة فهى:

$$\int_{-\infty}^{X} f(x)\,dx$$

لاحظ إن احتمال الحصول على نقطة محددة بالضبط فى التوزيعات المستمرة يساوى صفراً (احتمال الحصول على بقرة تعطى 3000 كج لبن بالضبط يساوى صفر حيث إن $\int_{3000}^{3000} f(x)\,dx = 0$). بينما هذا الاحتمال فى التوزيعات المتقطعة قد يساوى الصفر وقد يكون أكبر من الصفر ويعتبر هذا فرقاً أساسيا بين التوزيعات المتقطعة والتوزيعات المستمرة.

ويقع تحت هذين القسمين من التوزيعات عدد كبير من التوزيعات الهامة. وسوف تقتصر الدراسة الحالية على توزيع "ذو الحدين" و "بواسون" كأمثلة عن التوزيعات المتقطعة والتوزيع الطبيعى من التوزيعات المستمرة.

٤–٦ مقدمة عن التباديل Permutations والتوافيق Combinations

٤–٦–١ التباديل

التباديل تعرف بأنها عدد الطرق التى يمكن أن يقع بها حدث ما مع مراعاة الترتيب، فإذا كان هناك مجموعة مكونة من 3 أفراد هم C , B , A ويراد اختيار اثنين منهم أحدهما رئيس والآخر نائب للرئيس فإن هذين الفردين قد يكونان بست طرق:

$$AB , BA , AC , CA , BC , CB$$

مع ملاحظة أن AB ليست هى BA ففى الأولى A أولاً بينما B ثانياً بينما فى BA فإن B أولاً، A ثانياً ويعبر عن ذلك بما يلى:

تباديل n من الأشياء مأخوذة r فى كل مرة هو:

$$P(n,r) = \frac{n!}{(n-r)!} \qquad (٤-١٦)$$

وهذه المعادلة يمكن فكها كالتالى:

$$P(n,r) = \frac{n(n-1)\cdots(n-r+1)(n-r)(n-r-1)\cdots 3x2x1}{(n-r)(n-r-1)\cdots 3x2x1}$$
$$= n(n-1)\cdots(n-r+1)$$

وعلى ذلك فإن عدد الطرق التى يمكن أن ترتب بها 3 أفراد مأخوذة 2 فى كل مرة هو:

$$P(3,2)=\frac{3!}{(3-2)!}$$

لاحظ أن:

$n!$ يطلق عليها مفكوك n وهو عبارة عن

$$\begin{aligned} n! &= n(n-1)! \\ &= n(n-1)(n-2)! \\ &= n(n-1)(n-2)...3x2x1 \end{aligned}$$

فمثلا $5!=5x4!=5x4x3x2x1=120$

ومفكوك الصفر يساوى الواحد أى $0!=1$

وفيما يلى بعض القواعد الجبرية الخاصة بالتباديل:

قاعدة ١:

إذا كانت A تحدث فى m من المرات، B تحدث فى n من المرات وكان كل من A، B مستقلين فإن عدد الطرق التى يمكن أن يحدث بها A، B معاً هو حاصل ضرب m فى n.

قاعدة ٢:

إذا كانت A ممكن أن تحدث فى m من المرات، B يمكن أن تحدث فى n من المرات وكان A , B حدثين متنافيين فإن A أو B ممكن حدوثهما فى $m+n$ من المرات.

قاعدة ٣:

فى حالة المجموعات التى يكون داخلها تشابه تستخدم المعادلة:

$$P(n,n_1,n_2,\cdots,n_k)=\frac{n!}{n_1!\,n_2!\cdots n_k!} \qquad (١٧-٤)$$

مثال ٤-١٥

عند إلقاء حجرى نرد، ما عدد طرق التى يمكن أن يحدث بها الحدثان معاً؟

عدد طرق $36=(6)(6)$ (قاعدة ١)

مثال ٤-١٦

عند رمى حجر نرد، ما عدد الطرق التى يمكن الحصول منها على 1 أو 6؟

عدد الطرق $1+1=2$ (قاعدة ٢).

مثال ٤-١٧

ما عدد التباديل فى الأحرف التى يمكن أخذها من كلمة STATISTICS (10 أحرف) ؟

فى كلمة STATISTICS يوجد الأحرف التالية:

$$3S = n_1 , 3T = n_2 , 2I = n_3 , 1A = n_4 , 1C = n_5$$

وبالتالى
$$n = n_1 + n_2 + n_3 + n_4 + n_5 = 10$$

وبتطبيق القاعدة ٣ فإن
$$P(10,3,3,2,1,1) = \frac{10!}{3!3!2!1!1!} = 50400$$

٤-٦-٢ التوافيق

تعرف التوافيق على أنها عدد الطرق التى يمكن بها تكوين مجموعة حجمها r من n من الأفراد دون مراعاة الترتيب ويعبر عنها:

$$nCr \text{ or } C_r^n \text{ or } C(n,r) \text{ or } \binom{n}{r} = \frac{n!}{(n-r)!r!} \qquad (\text{٤-١٨})$$

مثال ٤-١٨

إذا كان هناك مجموعة من 3 أفراد ويراد تكوين لجنة من فردين بغض النظر عن ترتيبهما، فما عدد الطرق الممكنة ؟.

عدد الطرق الممكنة للحصول على لجنة مكونة من فردين من 3 أفراد هو 3 حيث إن:

$$C_2^3 = \frac{3\,!}{(3-2)\,!\,2\,!} = 3$$

فإذا كانت الأفراد هى A B C فإن اللجنة قد تكون AB أو AC أو BC ويلاحظ من المعادلة (٤-١٨) أن:

$$C_r^n = P(n,r)/r!$$

ومن أهم القواعد عند دراسة التوافيق لاستخدامها في الاحتمالات أن

$$C_r^n = C_{n-r}^n \qquad (١٩-٤)$$

وعلى ذلك فإن: C_3^5 هى نفسها C_2^5 حيث إن: $C_3^5 = 10 = C_{5-3}^5 = C_2^5$

وهناك علاقة بين التباديل والتوافيق و نظرية الاحتمالات، فقد سبق تعريف احتمال حدث ما بأنه عدد مرات ظهور هذا الحدث على عدد المرات الممكنة، والتباديل والتوافيق يسهلان الحصول على ذلك. والآن افترض وجود كيس به 4 كرات حمراء (R)، 3 كرات بيضاء (W)، وسحبت 3 كرات بدون إحلال، فما احتمال الحصول على الكرات الثلاث حمراء ؟ ولإيجاد مثل هذا الاحتمال باستخدام القواعد السابق ذكرها عن الاحتمالات فإن احتمال أن تكون الكرة الأولى حمراء والثانية حمراء والثالثة حمراء هو باستخدام العلاقة (٤-١٢) بعد تعميمها هو:

$$Pr(3R) = \frac{4}{7} x \frac{3}{6} x \frac{2}{5} = \frac{4}{35}$$

ويمكن الحصول على نفس النتيجة لو استخدمت التوافيق، فإن عدد الطرق التى يمكن الحصول بها على أى 3 كرات من بين 7 هو $C_3^7 = \frac{7!}{4!\,3!}$، وعدد الطرق التى يمكن الحصول بها على 3 كرات حمراء من بين 4 كرات حمراء هو $C_3^4 = \frac{4!}{3!\,1!}$، وباستخدام العلاقة (٤-١) فإن احتمال الحصول على 3 كرات حمراء عند سحب ٣ كرات هو: $Pr(3R) = \frac{C_3^4}{C_3^7} = \frac{4}{35}$ أى أن التوافيق تستخدم فى حساب عدد الطرق التى يمكن أن يحدث بهم حدث ما وعدد الطرق الكلية والتى منها يمكن حساب قيمة الاحتمال.

٤-٧ توزيع "ذو الحدين" Binomial Distribution

من التوزيعات المتقطعة discrete distribution والتى تختص بالمتغيرات المتقطعة discrete variables. وكما هو واضح من اسم التوزيع فإن المتغير قد يأخذ

أحد حالتين إما حالة نجاح فى حالة وقوعه أو حالة فشل فى حالة عدم وقوعه. فإذا وجد متغير، X مثلاً، وكان له الدالة الاحتمالية:

$$Pr(X = r; n, p) = C_r^n p^r q^{n-r} \ , \ r = 0, 1, 2, \cdots, n \qquad (٤-٢٠)$$

فإن هذا المتغير يتبع توزيع "ذو الحدين" بالمعلمين n، p.

٤-٧-١ مفكوك "ذو الحدين" Binomial expansion

$(a+b)^n$ يتبع القاعدة التالية:

$$(a+b)^n = a^n + \frac{n!}{(n-1)!1!} a^{n-1} b + \frac{n!}{(n-2)!2!} a^{n-2} b^2 + \ldots + \frac{n!}{1!(n-1)!} a b^{n-1} + b^n \qquad (٤-٢١)$$

إذا كانت $n = 1$ فإن المفكوك يساوى $a + b$

و إذا كانت $n = 2$ فإنه يصبح $a^2 + 2ab + b^2$

و إذا كانت $n = 3$ فإنه يصبح $a^3 + 3a^2b + 3ab^2 + b^3$

و إذا كانت $n = 4$ فإنه يصبح $a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$

وهكذا. ويمكن الحصول على معاملات coefficients حدود "ذو الحدين" وهى كما فى حالة $n = 1$ مثلاً 1,1، وفى حالة $n = 2$ هى 1,2,1، وفى حالة $n = 3$ تكون 1,3,3,1 ويمكن استخدام قاعدة باسكال Pascal rule للحصول على تلك المعاملات كما يلى:

n											
1					1		1				
2				1		2		1			
3			1		3		3		1		
4		1		4		6		4		1	
5	1		5		10		10		5		1
6											

... وهكذا ...

وعلى ذلك فإذا كان هناك عدد n من المحاولات كل محاولة إما حالة نجاح (باحتمال p) أو حالة فشل (باحتمال q) فإن توزيع "ذو الحدين" يمكن الحصول عليه من مفكوك "ذو الحدين" أى:

$$(p+q)^n = p^n + np^{n-1}q + \ldots + npq^{n-1} + q^n$$
$$= \sum_{r=0}^{n} C_r^n p^{n-r} q^r = \sum_{r=0}^{n} C_r^n p^r q^{n-r} \qquad (٤-٢٢)$$

ويطلق على القيمة $C_r^n p^{n-r} q^r$ الحد العام للتوزيع ويعرف بأنه الدالة الاحتمالية للتوزيع ومنه يمكن حساب احتمال الحصول على $n-r$ حالة نجاح فى عينة حجمها n أو عند إجراء n من المحاولات المتكررة مع ملاحظة:

١ – مجموع احتمالى النجاح والفشل يساوى الواحد الصحيح أى $p+q=1$.

٢ – إذا كانت n هى عدد المحاولات المتكررة أو حجم العينة فإن عدد الحدود يساوى $n+1$.

٣ – مجموع أسى q، p فى أى حد من الحدود يساوى n .

٤ – الحدود متنافية بمعنى أن حدوث أى منها يمنع الباقى من الحدوث.

٥ – مجموع الاحتمالات لكل الحدود يساوى الواحد الصحيح.

٦ – يلاحظ أن p وهى احتمال الحصول على حالة نجاح ثابتة لا تتغير من محاولة إلى أخرى وبالتالى q، وهى احتمال الفشل، لا تتغير.

مثال ٤–١٩

عند رمى قطعة من النقود مرة واحدة فإن الحالات التى يمكن أن تظهر هى صورة H أو كتابة T واحتمال كل منها عبارة عن:

$$\Pr(H) = \Pr(T) = \tfrac{1}{2}$$

وهذا هو مفكوك "ذو الحدين" عندما تكون $n=1$ حيث $(p+q)^1 = p+q$

الاحتمال	عدد مرات ظهور كتابة	عدد مرات ظهور صورة	عدد الطرق	النواتج
½	0	1	1	H
½	1	0	1	T

وعند رمى قطعتين من النقود مرة واحدة أو رمى قطعة واحدة من النقود مرتين فإن:

الاحتمال	عدد مرات ظهور كتابة	عدد مرات ظهور صورة	عدد الطرق	النواتج
Pr(2H) = ¼	0	2	1	HH
Pr(HT) = ½	1	1	2	HT TH
Pr(2T) = ¼	2	0	1	TT

وهذا هو مفكوك "ذو الحدين" عندما تكون $n = 2$ هو

$$(p+q)^2 = p^2 + 2pq + q^2$$

وعند رمى ٣ قطع من النقود مرة واحدة أو رمى قطعة من النقود ثلاث مرات فإن:

الاحتمال	عدد مرات ظهور كتابة	عدد مرات ظهور صورة	عدد الطرق	النواتج
$\Pr(3H) = 1/8$	0	3	1	HHH
$\Pr(2H, 1T) = 3/8$	1	2	3	HHT HTH THH
$\Pr(1H, 2T) = 3/8$	2	1	3	HTT THT TTH
$\Pr(3T) = 1/8$	3	0	1	TTT

مع تحيات د. سلام حسين الهلالي salamalhelali@yahoo.com

مثال ٤-٢٠

إذا سحبت عينة من 5 نباتات وكان احتمال الحصول على نبات طويل = $\frac{3}{4}$
و احتمال الحصول على نبات قصير = $\frac{1}{4}$، احسب الاحتمالات الآتية:

١- كل النباتات طويلة.

٢- ٣ نباتات طويلة.

٣- الحصول على كلتا الصفتين فى العينة.

حيث إن الحدثين احتمالهما $p=\frac{3}{4}$، $q=\frac{1}{4}$ لصفتى الطول والقصر، على الترتيب ويتبعان توزيع "ذو الحدين" بحجم عينة $n = 5$، فإن:

١ - احتمال كل النباتات طويلة:

$$C_5^5 p^5 q^{5-5} = p^5 = \left(\frac{3}{4}\right)^5 = 0.2373$$

٢ - احتمال ٣ نباتات طويلة:

$$C_3^5 p^3 q^5 = \frac{5!}{3!2!}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^2 = (10)\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^2 = 0.2637$$

٣ - احتمال الحصول على كلتا الصفتين فى العينة: وهذه تشمل الحدود التى تظهر فيها كلا الصفتين (صفة الطول وصفة القصر) وهى كل الحدود فى مفكوك "ذو الحدين" فيما عدا الحدين الأول والأخير أى:

$$= C_1^5 pq^4 + C_2^5 p^2 q^3 + C_3^5 p^3 q^2 + C_4^5 p^4 q$$

$$= (5)\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^4 + (10)\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^3 + (10)\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^2 + (5)\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right) = 0.9998$$

ويمكن الحصول على نفس النتيجة إذا علم أن مجموع الاحتمالات = 1 وبالتالى فإن احتمال الحصول على كلا الصفتين هو:

$$1 - p^5 - q^5 = 1 - \left(\frac{3}{4}\right)^5 - \left(\frac{1}{4}\right)^5 = 0.9998$$

٤-٧-٢ متوسط وتباين "ذو الحدين"

تظهر معادلة (٤-٢٢) مفكوك "ذو الحدين" ومنها فإن الحد الأول هو احتمال الحصول على n حالة نجاح وصفر حالة فشل، والحد الثانى هو احتمال الحصول على $(n-1)$ حالة نجاح وحالة فشل واحدة، والحد الأخير هو احتمال الحصول على صفر حالة نجاح ،n حالة فشل. وبالتالى فإن متوسط هذا التوزيع يمكن حسابه كما يلى:

X	f	fX
n	p^n	np^n
$n-1$	$np^{(n-1)}q$	$n(n-1)p^{(n-1)}q$
$n-2$	$\frac{n(n-1)}{2!}p^{(n-2)}q^2$	$\frac{n(n-1)(n-2)}{2!}p^{(n-2)}q^2$
$\vdots$	$\vdots$	$\vdots$
1	$npq^{(n-1)}$	$npq^{(n-1)}$
0	q^n	0
إجمالى	**1**	$np^n + n(n-1)p^{(n-1)}q + \ldots + npq^{(n-1)} + 0$

وحيث إن: $\overline{X} = \frac{\sum fX}{\sum f} = \sum fX$ (لاحظ أن $\sum f = 1$) وبالتالى:

$$\overline{X} = np^n + n(n-1)p^{(n-1)}q + \cdots + npq^{(n-1)}$$
$$= np\left[p^{(n-1)} + (n-1)p^{(n-2)}q + \cdots + q^{(n-1)}\right]$$

بوضع $m = (n-1)$ فإن:

$$\overline{X} = np\left[p^m + mp^{m-1}q + \cdots + q^m\right]$$
$$= np(p+q)^m$$

وبما أن $(p+q) = 1$ فيصبح المتوسط:

$$\overline{X} = np \qquad (٤-٢٣)$$

وهو يمثل متوسط عدد الأفراد الحاملة للصفة والتى تمثل p احتمال حدوثها (الطول مثلاً). وبنفس المفهوم فإن متوسط عدد الأفراد الحاملة للصفة الأخرى هو nq.

ويمكن إثبات أن التباين:

$$\sigma^2 = npq \qquad (٤-٢٤)$$

ويتوقف متوسط "ذو الحدين" على حجم العينة واحتمال الحصول على حالة النجاح (مثلاً) فيزداد بزيادتها أو بزيادة أحدهما. أما التباين فيتوقف على حجم العينة وقيمة pq التى تكون أكبر ما يمكن عندما تكون $p = q = \frac{1}{2}$.

وعند الرغبة فى التعبير عن هذه القيم فى صورة نسب وليست وحدات فيكون المتوسط p ,q والتباين $\frac{pq}{n}$.

مثال ٤-٢١

احسب المتوسط والتباين للمثال ٤-٢٠

متوسط عدد الأفراد الحاملة لصفة الطول: $np = 5x\frac{3}{4} = 3.75$

متوسط عدد الأفراد الحاملة لصفة القصر: $nq = 5x\frac{1}{4} = 1.25$

التباين : $npq = 5x\frac{3}{4}x\frac{1}{4} = 0.9375$

مع تحيات د. سلام حسين الهلالي salamalhelali@yahoo.com

٤-٨ توزيع بواسون Poisson distribution

يعتبر هذا التوزيع من التوزيعات التى تختص بالمتغيرات المتقطعة نادرة الحدوث وعندما تحدث يمكن أن تحدث n من المرات، مثال ذلك عدد الميكروبات فى لتر من ماء نقى أو فى طبق من أطباق بترى أو عدد الأشخاص المرضى بمرض نادر الحدوث أو اضمحلال أثر مادة مشعة بمرور الوقت. كما يستخدم توزيع بواسون فى نظرية الصفوف، حيث أنه من المعروف أن "الوصول لطلب الخدمة" يتبع هذا التوزيع، مثال ذلك عدد الأشخاص الواردين إلى متجر معين لشراء سلعة مثلاً فى وقت معين.

وتوزيع بواسون هو نهاية limit توزيع "ذو الحدين" عندما p تؤول إلى الصفر $(p \to 0)$ وعندما n تؤول إلى ما لا نهاية $(n \to \infty)$ وعلى ذلك فإن $\mu = np$ يكون مقداراً ثابتاً تقريباً ويكون: $\sigma^2 = np(1-p) = np$ حيث إن p صغيرة جداً أى أن المتوسط لهذا التوزيع يكون مساوياً لتباينه. ويتبين أن احتمال الحصول على r حالة نجاح يؤول إلى:

$$\Pr(r) = \frac{e^{-\mu}\mu^r}{r!} \quad r = 0, 1, \cdots \qquad (٤-٢٥)$$

وهى صورة مبسطة لحساب احتمال الحصول على r حالة نجاح حيث: $e = 2.7183$ و μ هى متوسط عدد مرات النجاح r فى العينة محسوباً من العشيرة.

وجدول ٤-١ يبين احتمال الحصول على r من النجاحات محسوباً بتوزيع بواسون عندما تكون $\mu = 1$ مقارنةً بتوزيع "ذو الحدين" عندما تكون $n = 100$ ، $p = 0.01$ ، وعندما تكون $n = 25$ ، $p = 0.04$ ومتوسط توزيع "ذو الحدين" فى كلتا الحالتين يساوى الواحد الصحيح أيضاً. ويلاحظ من هذا الجدول أن توزيع بواسون يكون أكثر مطابقة لتوزيع "ذو الحدين" عندما يكون حجم العينة كبيراً.

مثال ٤-٢٢

إذا كان متوسط عدد العينات من مرض نادر الحدوث هو 25 لكل 10000 فرد فما احتمال الحصول على صفر حالة وفاة فى عينة من 100 فرد؟ وما احتمال الحصول على 3 حالات وفاة ؟

متوسط العينة
$$\mu = np = \frac{25}{10000} x 100 = 0.25$$

احتمال الحصول على صفر حالة وفاة:

$$Pr(r=0)=\frac{e^{-\mu}\mu^{0}}{r!}=c^{-\mu}=c^{-0.25}=0.779$$

احتمال الحصول على 3 حالة وفاة:

$$Pr(r=3)=\frac{e^{-\mu}\mu^{3}}{r!}=\frac{2.718^{-.25}}{3x2x1}x0.25^{3}=0.002$$

جدول ٤-١ مقارنة بين توزيع بواسون وتوزيع "ذو الحدين"

احتمال الحصول على r من النجاحات محسوبا بـ			
r	توزيع بواسون	توزيع "ذو الحدين"	
		p = 0.01, n = 100	**p = 0.04, n = 25**
0	0.3679	0.3660	0.3604
1	0.3679	0.3697	0.3754
2	0.1839	0.1849	0.1877
3	0.0613	0.0610	0.0600
4	0.0153	0.0149	0.0137
5	0.0031	0.0029	0.0024
6	0.0005	0.0005	0.0003
≥7	0.0001	0.0001	0.0000
Total	**1.0000**	**1.0000**	**0.9999**

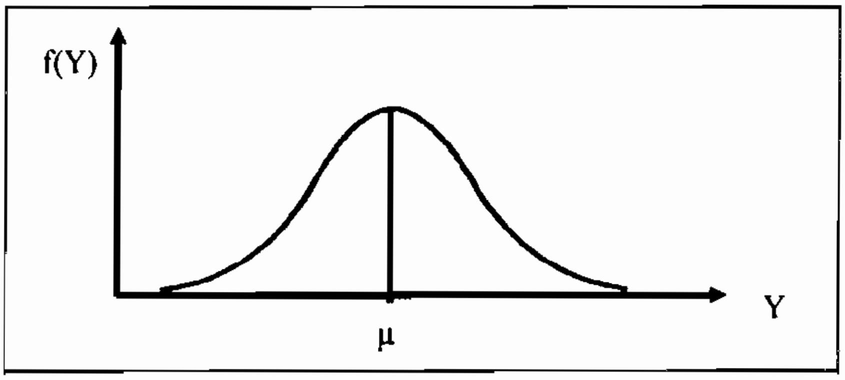
المصدر: Snedecor and Cochran (1987)

٤-٩ التوزيع الطبيعي Normal distribution

من أهم التوزيعات الخاصة بالصفات ذات التوزيع المستمر continuous وهو أكثرها شيوعاً ويتبعه العديد من الصفات البيولوجية (الكمية).

والتوزيع الطبيعي يتحدد بمعلمين هما المتوسط μ والانحراف المعياري σ. وتتحدد قيم المتغير نظرياً في المدى من $-\infty$ إلى $+\infty$.

ودالة التوزيع الطبيعي تأخذ الشكل الناقوسي وقد يطلق عليه المنحنى الطبيعي للأخطاء normal curve of error وقد يطلق عليه أيضا Gussian or Laplacian curve نسبة إلى مكتشفيه، ولو أن De Moivre اكتشفه أيضاً فى نفس الوقت. الشكل ٤-٩ يوضح المنحنى الطبيعى بمتوسط μ وانحراف معيارى σ.



شكل ٤-٩ التوزيع الطبيعى بمتوسط μ وانحراف معيارى σ

ومعادلة المنحنى الطبيعى الذى متوسطه μ وانحرافه المعيارى σ هى $f(Y)$ والتى تمثل التكرار النسبى relative frequency أو الارتفاع height على المحور الصادى المقابل لقيم Y هى:

$$f(Y,\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{Y-\mu}{\sigma}\right)^2} \qquad (٤-٢٦)$$

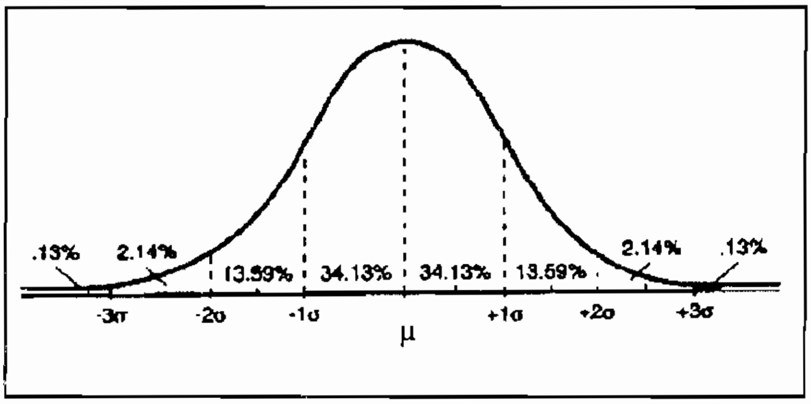
حيث: $\pi = 3.1416$ (وهى القيمة "ط" فى العربية)، $e = 2.7183$ ، $-\infty \le Y \le +\infty$

ومن أهم خصائص التوزيع الطبيعى أنه متماثل حول نقطة المنتصف μ وأن النقطتين $\mu \pm \sigma$ تحصران فيما بينهما حوالى 68% من المساحة تحت المنحنى أى $\frac{2}{3}$ مساحة هذا التوزيع تقريباً وأيضاً النقطتين $\mu \pm 2\sigma$ تحصران فيما بينهما حوالى 95% من المساحة، بينما $\mu \pm 3\sigma$ فإنهما يحصران 99.7% من المساحة. الشكل ٤-١٠ يظهر قيم المساحات المختلفة.

ويرجع الاهتمام بالتوزيع الطبيعى إلى عدة أسباب:

١- يسهل تحويل أى توزيع طبيعى مهما كان متوسطه وانحرافه المعيارى إلى توزيع طبيعى قياسى standardized normal distribution والذى وضعت

له جداول (جدول ٣ ملحق أ) يمكن منها حساب القيم المختلفة للاحتمالات فى أقل وقت وبأقل جهد ودون الحاجة إلى استخدام حساب التكامل.



شكل ٤-١٠ العلاقة بين المساحة تحت المنحنى الطبيعى وكل من المتوسط والانحراف المعيارى.

٢- كثير من المتغيرات الكمية كالأطوال والأوزان (مثلا) تتخذ فى توزيعها عادة شكل التوزيع الطبيعى.

٣- وحتى لو كانت هذه البيانات لا تتوزع طبيعياً فإنه يمكن أحياناً بواسطة التحويرات المختلفة transformations لمقاييس هذه البيانات أن تتوزع طبيعياً كأن يأخذ الجذر التربيعى مثلاً $\sqrt{X}$ أو لوغاريتم القيم $\log X$.

٤- بعض التوزيعات الأخرى تقترب فى توزيعها من التوزيع الطبيعى تحت ظروف معينة كأن يزداد حجم العينة n مثلاً كما فى توزيع "ذو الحدين" وخاصة عندما تقترب p من النصف.

٥- لأى توزيع متغير X تباينه محدد finite فإن توزيع متوسطات العينات المسحوبة من عشيرة هذا المتغير تقترب فى توزيعها من التوزيع الطبيعى وتزداد قرباً كلما زاد حجم العينة n، وتعرف هذه الخاصية بنظرية الحد المركزى central limit theorem ويكون تباين التوزيع الطبيعى للمتوسطات هو σ^2/n ومتوسطه μ أيضاً. ولهذا القانون أهمية خاصة فى الاختبارات الإحصائية للمقارنة بين متوسطات العينات.

٤-١٠ التوزيع الطبيعي القياسي Standard normal distribution

يحدد شكل أى توزيع طبيعي معلمان هما المتوسط والانحراف المعيارى للقيم، وعلى ذلك فقد يوجد عدد لانهائي من التوزيعات.

والتوزيع الطبيعي القياسي هو التوزيع الطبيعي الذى متوسطه يساوى الصفر وتباينه يساوى الواحد الصحيح واصطلح على ذلك بكتابة $N \sim (0,1)$. وأنشئت جداول يمكن منها حساب المساحات المحصورة بين المتوسط وقيمة معينة منه (جدول ٣ ملحق أ) ومن ذلك يسهل الحصول على قيم الاحتمالات المختلفة. ويمكن تحويل أى توزيع طبيعي مهما كان متوسطه وانحرافه المعياري إلى التوزيع الطبيعي القياسي باستخدام العلاقة التالية:

$$Z = \frac{Y-\mu}{\sigma} \qquad (٤-٢٧)$$

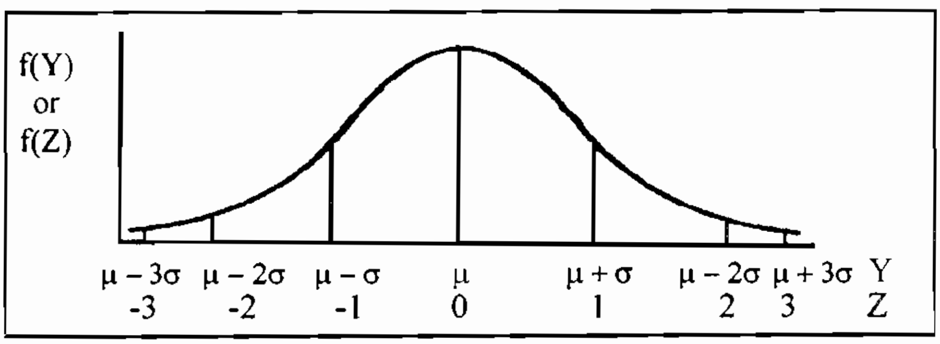
ومنها:

$$Y = \mu + Z\sigma \qquad (٤-٢٨)$$

فإذا كانت Y تتوزع بمتوسط μ وانحراف معياري σ فإن Z تتوزع بمتوسط يساوى الصفر وانحراف معيارى يساوى الواحد الصحيح وتكون دالة هذا التوزيع:

$$f(Z, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} \qquad (٤-٢٩)$$

أى أن القيم Z هى عبارة عن انحرافات المتغير الأساسى Y عن متوسطها مقدرة بوحدات الانحراف القياسى للمتغير الأساسى. والشكل ٤-١١ يبين العلاقة بين المنحنى الطبيعى والمنحنى الطبيعى القياسى



شكل ٤-١١ العلاقة بين المنحنى الطبيعى والمنحنى الطبيعى القياسى

مثال ٤-٢٣

إذا كانت صفة إنتاج اللبن فى أبقار قطيع ما تتوزع طبيعيا بمتوسط 3000 kg لبن فى الموسم وانحراف معيارى 250 kg احسب ما يلى:

١ - احتمال الحصول على بقرة إنتاجها من اللبن 3050 kg بالضبط.

٢- احتمال الحصول على بقرة إنتاجها يتراوح بين 2700, 3500 kg لبن فى الموسم.

٣- احتمال الحصول على بقرة إنتاجها على الأقل 2500 kg لبن فى الموسم.

٤- احتمال الحصول على بقرة إنتاجها على الأكثر 2500 kg لبن فى الموسم.

٥- إذا استبعد أقل 10% من الأبقار، ما هى حدود الإنتاج التى تحصر أقل 10% من الأبقار.

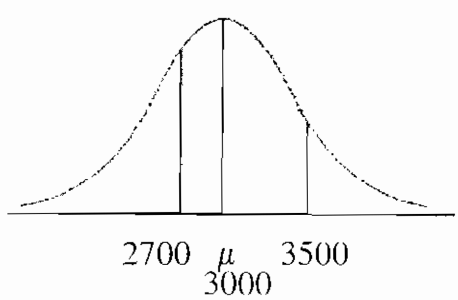
٦- ما حدود الإنتاج التى تحصر أعلى 15% من الأبقار.

٧- إذا كان فى قطيع ما عدد 1000 بقرة فما عدد الأبقار التى يتراوح إنتاجها بين 2700, 3500 kg لبن فى الموسم.

لحساب مثل تلك الاحتمالات فى حالة التوزيع الطبيعى يجدر بادئ ذى بدء رسم شكل تخطيطى للتوزيع مع بيان المطلوب ليسهل الحساب.

١- احتمال الحصول على بقرة إنتاجها من اللبن 3050 kg بالضبط يساوى الصفر وهى تمثل المساحة المنحصرة فوق القيمة 3050 kg والممثلة بنقطة على المنحنى ليس لها أبعاد حيث إن صفة إنتاج اللبن فى هذه الحالة تتوزع توزيعا مستمرا.

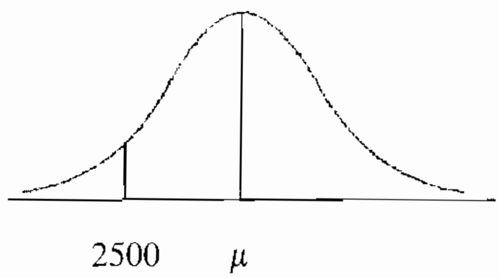
٢- احتمال الحصول على بقرة إنتاجها يتراوح بين 2700, 3500 kg لبن فى الموسم:



$$\Pr(2700 \le X \le 3500) = \Pr\left(\frac{2700-3000}{250} \le \frac{X-\mu}{\sigma} \le \frac{3500-3000}{250}\right)$$
$$= \Pr(-1.2 \le Z \le 0) + \Pr(0 \le Z \le 2)$$
$$= 0.3849 + 0.4772 = 0.8621$$

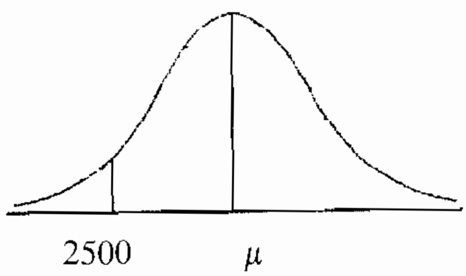
بما أن التوزيع متماثل فإن قيمة الاحتمال ما بين (الصفر، 1.2-) هى نفسها قيمة الاحتمال ما بين (الصفر، 1.2) جدول ٣ ملحق أ.

٣- احتمال الحصول على بقرة إنتاجها على الأقل 2500 kg لبن فى الموسم:



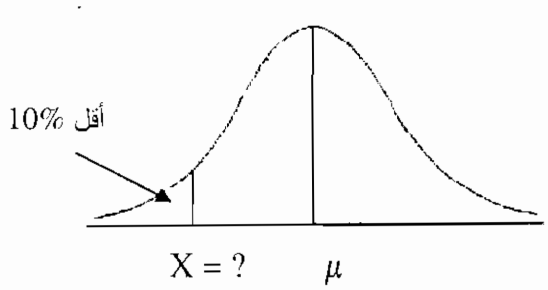
$$\Pr(X \ge 2500) = \Pr\left(\frac{X-\mu}{\sigma} \ge \frac{2500-3000}{250}\right)$$
$$= \Pr(Z \ge -2) = 0.5 + \Pr(0 \le Z \le 2)$$
$$= 0.5 + 0.4772 = 0.9772$$

٣- احتمال الحصول على بقرة إنتاجها على الأكثر 2500 kg لبن فى الموسم:



$$\Pr(X \le 2500) = \Pr\left(\frac{X-\mu}{\sigma} \le \frac{2500-3000}{250}\right)$$
$$= \Pr(Z \le -2)$$
$$= 1 - \Pr(Z \ge -2) = 1 - 0.9772 = 0.0228$$

٥- حدود الإنتاج التى تحصر أقل %10 من الأبقار:



إذا كانت أقل %10 من الأبقار تمثل %10 من المساحة تحت المنحنى (على يسار المتوسط) فإن باقي المساحة تمثل %90 وبالتالى يتم البحث فى جدول ٣ ملحق أ عن قيمة Z التى تقابل مساحة قدرها $0.9 - 0.5 = 0.4$ والتى تساوى $Z = 1.281$، هذه القيمة تم الحصول عليها بالاستكمال interpolation.

$$Z = \frac{X-\mu}{\sigma}$$
$$-1.281 = \frac{X-3000}{250}$$
$$X = 2680 \text{ kg}$$

إذاً حدود الإنتاج التى تحصر أقل %10 من الأبقار هى كل قيمة أقل من 2680 kg

٦- حدود الإنتاج التى تحصر أعلى %15 من الأبقار:

أعلى %15 من الأبقار تمثل %15 من المساحة تحت المنحنى (على يمين المتوسط) وبالتالى فإن باقى المساحة تمثل %85، يتم البحث فى جدول ٣ لملحق أ عن قيمة Z التى تقابل مساحة قدرها $0.85 - 0.5 = 0.35$ وهى $Z = 1.0365$

مع تحيات د. سلام حسين عويد الهلالي

https://scholar.google.com/citations?user=t1aAacgAAAAJ&hl=en

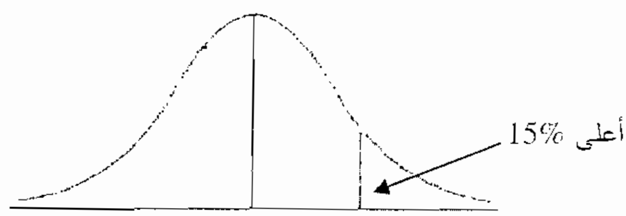
salamalhelali@yahoo.com

https://www.facebook.com/salam.alhelali

https://www.facebook.com/groups/Biothesis/

https://www.researchgate.net/profile/Salam_Ewaid/

07807137614



$$\mu \qquad X = ?$$

$$Z = \frac{X - \mu}{\sigma}$$

$$1.0365 = \frac{X - 3000}{250}$$

$$X = 3259 \text{ kg}$$

إذاً حدود الإنتاج التى تحصر أعلى 15% من الأبقار هى كل قيمة أعلى من 3259 kg

٧ – عدد الأبقار التى يتراوح إنتاجها بين2700, 3500 kg لبن فى الموسم إذا كان العدد الكلى للأبقار 1000 بقرة:

سبق حساب احتمال الحصول على بقرة يتراوح إنتاجها بين 2700 ، 3500 kg وكان $Pr(2700 \le X \le 3500) = 0.8621$

إذا عدد الأبقار $1000 \times 0.8621 = 862$

مع تحيات د. سلام حسين الهلالي salamalhelali@yahoo.com

صندوق ٤-١

توزيع "ذو الحدين":

$$Pr(X = r; n, p) = C_r^n p^r q^{n-r} \quad , \quad r = 0, 1, 2, ..., n$$

المتوسط = np ، التباين = npq

توزيع بواسون:

$$Pr(r) = \frac{e^{-\mu}\mu^r}{r!}$$

المتوسط = التباين

التوزيع الطبيعى:

$$f(Y, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y-\mu}{\sigma}\right)^2}$$

حيث μ هي المتوسط، σ هى الانحراف المعيارى

التوزيع الطبيعى القياسى:

إذا كانت Y تتوزع طبيعيا بمتوسط μ وتباين σ^2 فإن Z تتوزع طبيعيا بمتوسط صفر وتباين يساوى الواحد الصحيح حيث إن

$$Z = \frac{Y-\mu}{\sigma}$$

مع تحيات د. سلام حسين الهلالي salamalhelali@yahoo.com

٤-١١ توزيع المتوسطات Distribution of means

إذا سحبت عينة sample من عشيرة بإحدى طرق المعاينة sampling methods وقدر لها المتوسط الحسابى وليكن $\overline{X}$ وكرر سحب العينات وفى كل مرة يحسب المتوسط منها فإنه يتكون عشيرة من المتوسطات متوسطها μ وتباينها σ^2/n. وحتى يمكن حساب قيم الاحتمالات المختلفة فإنه يمكن تحويل هذا التوزيع مهما كان متوسطه وتباينه إلى التوزيع الطبيعى القياسى بالعلاقة التالية:

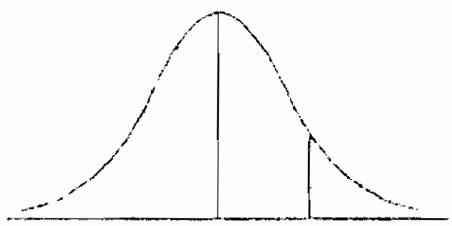
$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} \qquad (٤-٣٠)$$

حيث: $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$ وتسمى $\sigma_{\overline{X}}$ الخطأ القياسى standard error.

مثال ٤-٢٤

فى مثال ٤-٢٣ إذا سحبت عينة عددها 25 بقرة. أحسب احتمال أن يزيد متوسطها عن 3090 kg لبن فى الموسم ؟

$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{250}{\sqrt{25}} = 50\ \text{kg}$ وباستخدام معادلة (٤-٣٠) فإن:



$\mu = 3000 \quad \overline{X} = 3090$

$$\Pr(\overline{X} \geq 3090) = \Pr\left(\frac{\overline{X} - \mu}{\sigma_{\overline{X}}} \geq \frac{3090 - 3000}{50}\right)$$

$= \Pr(Z \geq 1.8)$

$= 0.5 - \Pr(0 \leq Z \leq 1.8) = 0.0359$

٤-١٢ تقريب توزيع "ذو الحدين" بالتوزيع الطبيعى

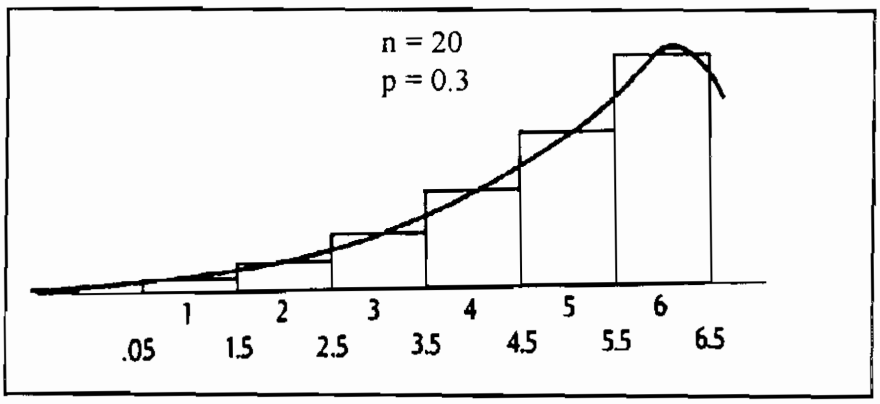
عند زيادة حجم العينة n لصفة تتبع توزيع "ذو الحدين" فإن هذا التوزيع يقترب من التوزيع الطبيعى بمتوسط $\mu = np$ وانحراف معيارى $\sigma = \sqrt{npq}$ ويزداد هذا التوزيع قرباً من التوزيع الطبيعى كلما اقتربت p من النصف (إذا كانت p أقل من النصف وكان $\mu \geq 15$ يستخدم التوزيع الطبيعى التقريبى بدلا من استخدام التوزيع الأصلى وهو "ذو الحدين" والمتوسط يقل عن ذلك إذا كانت p تساوى النصف).

ولهذا التقريب أهمية كبرى فى جعل حساب الاحتمالات أكثر سهولة عندما تكون n كبيرة ولكن يجب إجراء تصحيحاً لعدم الاستمرارية correction for discontinuity حيث إن توزيع "ذو الحدين" من التوزيعات المتقطعة كما سبق ذكره بينما التوزيع الطبيعى من التوزيعات المستمرة ويكون التصحيح بطرح ½ من القيمة المطلقة للفرق بين r (عدد مرات النجاح) والمتوسط كما فى العلاقة التالية:

$$Z = \frac{|r-\mu| - \frac{1}{2}}{\sigma} \qquad (٤-٣١)$$

حيث أن: $\mu = np$ ، $\sigma = \sqrt{npq}$

والشكل التالى يوضح العلاقة بين التوزيع الطبيعى وتوزيع "ذو الحدين".



شكل ٤-١٢ العلاقة بين التوزيع الطبيعى وتوزيع "ذو الحدين"

مثال ٤-٢٥

فى قطيع مكون من 400 بقرة متوقع ولادتها، احسب احتمال الحصول على:

١- 200 ذكر.

٢- 230 ذكر على الأقل.

٣- 170 أنثى على الأكثر.

فى هذا المثال يوجد نوعان من جنس المولود إما أن يكون ذكراً وإما أن يكون أنثى، واحتمال كل منهما يساوى النصف، وبهذا فإن صفة الجنس تتبع توزيع "ذو الحدين" وللحصول على إجابة مضبوطة فإن:

$$\begin{aligned}\Pr(200\,\text{males}) &= C_{200}^{400}\left(\frac{1}{2}\right)^{200}\left(\frac{1}{2}\right)^{200}\\ &= \frac{400!}{200!200!}\left(\frac{1}{2}\right)^{400}\\ &= \frac{400\text{x}399\text{x}\cdots\text{x}201\text{x}200!}{200!200\text{x}199\text{x}\cdots\text{x}3\text{x}2\text{x}1}\left(\frac{1}{2}\right)^{400}\\ &= \frac{400\text{x}399\text{x}\cdots\text{x}201}{200\text{x}199\text{x}\cdots\text{x}3\text{x}2\text{x}1}\left(\frac{1}{2}\right)^{400}\end{aligned}$$

١- واضح أنه من الصعب حساب قيمة هذا الاحتمال. وحيث إن متوسط "ذو الحدين" عبارة عن $np = 200$ واحتمال الحصول على حالة نجاح (الحصول على ذكر مثلاً فى هذه الحالة) يساوى النصف فإنه يمكن استخدام التوزيع الطبيعى التقريبى للحصول على قيم الاحتمالات المختلفة فى مثل هذه الحالة. ويبين الشكل ٤-١٣ توزيع "ذو الحدين" لهذا المثال.

وحيث إن:

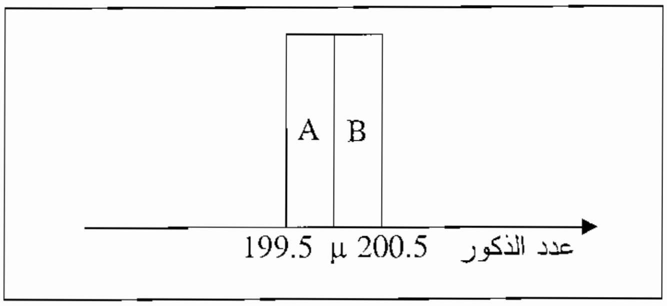
$\mu = np = 400\text{x}\frac{1}{2} = 200$ ، $\sigma = \sqrt{npq} = \sqrt{400\text{x}\frac{1}{2}\text{x}\frac{1}{2}} = 10$ ، فإن

$$Z_1 = \frac{200.5 - 200}{10} = 0.05$$

وقيمة هذا الاحتمال هو 0.0199، وقيمة الاحتمال المقابلة لـ Z_2 هى: 0.0199 أيضاً.

وحيث إن $A = B$ حيث إن التوزيع الطبيعى متماثل (شكل ٤-١٣) فإن مساحة $A + B$ تساوى ضعف مساحة A.

أى أن احتمال الحصول على 200 ذكر عبارة عن $(2)(0.0199) = 0.0398$



شكل ٤-١٣ توزيع "ذو الحدين" لعدد الذكور

٢- احتمال الحصول على 230 ذكر على الأقل:

$$Z = \frac{|r - \mu| - \frac{1}{2}}{\sigma} = \frac{|230 - 200| - \frac{1}{2}}{10} = \frac{29.5}{10} = 2.95$$

إذاً احتمال الحصول على 230 ذكر على الأقل هو $0.5 - 0.4984 = 0.0016$

٣- احتمال الحصول على 170 أنثى على الأكثر. أى احتمال الحصول على 170 أو 169 أو 168 أو ... أو صفر أنثى. وهذا الاحتمال هو نفسه احتمال الحصول على 230 أو 231 أو ... 400 ذكراً، لأن هذه الاحتمالات تقع تحت توزيع "ذو الحدين" أصلاً، وحيث إن حالة النجاح فى هذا المثال هى الحصول على ذكر وتم حساب المتوسط بناء على ذلك (مع ملاحظة أن متوسط الذكور هو نفسه متوسط الإناث حيث إن $p = q$ فى هذه الحالة) وقد حسبت قيمة هذا الاحتمال قبل ذلك مباشرة أى أن:

احتمال الحصول على 170 أنثى على الأكثر هو نفسه الحصول على 230 ذكر على الأقل $= 0.0016$.

تمارين الباب الرابع

٤-١ إذا ألقت قطعة من النقود عشرة مرات فما احتمال الحصول على (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10) صورة ؟ (احتمال الصورة = احتمال الكتابة = 0.5).

٤-٢ إذا كان هناك 4 أحداث (A, B, C, D) هى كل الأحداث الممكنة الناتجة من تجربة وكلها متنافية ويوجد 4 إجابات لقيم الاحتمالات مبينة بالجدول التالى. اختبر أى الإجابات يمثل الإجابة الصحيحة:

	احتمال			
الحالة	**A**	**B**	**C**	**D**
أ	0.34	0.29	0.17	0.30
ب	0.17	0.59	0.21	0.03
ج	0.24	0.05	0.18	0.33
د	$\frac{13}{80}$	$\frac{1}{2}$	$\frac{11}{40}$	$\frac{3}{8}$

٤-٣ ما عدد الطرق التى يمكن بها لأسرة أن تنجب 3 ذكور و 2 أنثى ؟

٤-٤ عند إلقاء حجرى نرد ما احتمال أن يكون مجموع الوجهين 5 على الأقل ؟

٤-٥ أسرة مكونة من 5 أطفال ما احتمال أن يكون الأطفال من كلا الجنسين ؟

٤-٦ إذا كان لديك صندوقان أحدهما: (أ) به 4 كرات حمراء، 6 كرات بيضاء، والآخر (ب) به 5 كرات حمراء، 6 كرات بيضاء. سحبت كرة من كل صندوق، ما احتمال أن تكون الكرتان من لون واحد ؟

٤-٧ إذا كان لديك الأرقام التالية: (3, 2, 1, 0) فما الأعداد الثلاثية الممكن تكوينها دون تكرار الرقم أكثر من مرة فى العدد الواحد وبحيث يكون فى خانة العشرات دائماً الرقم 3 ؟

٤-٨ احسب: C_2^5 ، C_5^6 ، C_3^5

٤-٩ إذا نتج طلوقة من تزاوج فردين طبيعيين ولكن حاملين لأليل أوتوسومى مميت، فما احتمال أن يكون هذا الطلوقة حاملا لهذا الأليل، مع العلم أنه عندما لقح هذا الطلوقة 5 بقرات طبيعية حاملة لهذا الأليل نتج عن هذا التلقيح 5 أبناء طبيعية ؟

٤-١٠ إذا كان لديك 3 صناديق:

صندوق (أ) به 2 كرة بيضاء، 3 كرة حمراء، كرة واحدة سوداء،

صندوق(ب) به 3 كرات بيضاء، 4 كرة حمراء، 2 كرة سوداء،

صندوق(ج) به 4 كرات بيضاء، 2 كرة حمراء، 3 كرات سوداء.

سحبت كرتان من أحد الصناديق وكانت إحداهما حمراء والأخرى سوداء، فما احتمال أن تكون الكرتان المسحوبتان من الصندوق الأول ؟

٤-١١ إذا كان اللون الأسود سائداً سيادة تامة على اللون الأحمر فى سلالة أبقار الأبردين أنجس، طلوقة أسود اللون لقح بقرة سوداء فأنتجا عجلاً أحمر، احسب الاحتمالات الآتية:

أ- أن يكون العجل القادم من نفس الفردين أحمر اللون.

ب- إذا أنتج هذا التزاوج أربع عجول، فما احتمال أن يكون من بينهما 3 عجول حمراء؟

٤-١٢ إذا كان اللون الأجوتى يسود على اللون الأبيض فى الأرانب ويتوقف على زوج واحد من العوامل الوراثية، وفى تزاوج بين فردين خليطين تم الحصول على خمسة أفرد، فما احتمال الحصول على:

١- فردين أجوتى.

٢- على الأقل 3 أفراد أجوتى.

٣- أن يكون الأفراد الخمسة أجوتى.

٤- أن يظهر اللونان فى النسل.

٤-١٣ فى توزيع "ذو الحدين"، إذا كانت $n = 10$، $p = q = 0.5$، احسب الاحتمالات الآتية:

١- الحصول على 4 حالات نجاح على الأقل

٢- الحصول 4 حالات نجاح على الأكثر

٣- الحصول على 3 حالات نجاح

٤-١٤ باستخدام البيانات فى (٤-١٣) احسب الاحتمالات مستخدماً التوزيع الطبيعى التقريبى، مرة عندما تكون $p = 0.5$ وأخرى عندما $p = 0.75$ ثم قارن النتائج التى حصلت عليها بنتائج التمرين السابق. وما تستنتج ؟

٤-١٥ فى تزاوج بين فردين خليطين (Aa)، تم الحصول على 144 فرداً. فما احتمال أن يكون:

(أ) عدد الأفراد الخليطة بين 71 ، 73 فرداً.

(ب) عدد الأفراد الأصلية aa يزيد عن 42 فرداً.

(ج) متوسط عدد الأفراد aa بين 36-36.5 فرداً.

٤-١٦ إذا كانت صفة الطول تتوزع طبيعياً بمتوسط 50 cm وتباين $64\ cm^2$، أوجد ما يلى:

١- احتمال الحصول على فرد طوله 50 cm

٢- القيم التى تحصر فيما بينها أطول 20% من الأفراد

٣- احتمال الحصول على:

أ- فرد طوله بين 45, 60 cm

ب- فرد طوله أقل من 40 cm

٥ المعاينة Sampling

١- مقدمة

٢- الخطوات الرئيسية التى يجب إتباعها فى المعاينة

٣- طرق المعاينة

٤- المعاينة العشوائية البسيطة

٥- مصادر الخطأ فى المعاينة

٥-١ مقدمة

يهتم علم الإحصاء بدراسة العشائر بما تتسم به من معالم بغرض الدراسة أو التخطيط كتقدير حجم القوى العاملة والبطالة أو الإنتاج الزراعى أو الصناعى أو استهلاك الكهرباء أو الغاز أو التركيب العمرى للسكان أو نسبة عدد الذكور إلى الإناث فى كل مرحلة عمرية وغيرها الكثير. ويمكن الحصول على تلك المعلومات عن طريق:

١- الحصر الشامل كإجراء حصر شامل أو تعداد لمجتمع سكان مصر (مثلاً) فى توقيت معين أو عند الرغبة فى الحصول على معلومات حول العديد من التقسيمات الفرعية فى العشيرة حيث يكون حجم العينة كبيراً وضخماً بحيث يمثل الحصر الشامل أفضل حل أو أن يكون الباحث ليس على علم تام بطبيعة مفردات العشيرة التى يرغب فى دراستها.

٢- سحب عينة ما بطريقة ما (تسمى تقنية المعاينة sampling technique) كأن تؤخذ عينة من دم مريض لفحصها حيث إن سحب كل دم المريض يؤدى إلى الوفاة، أى أن استخدام العينة فى مثل تلك الحالة ضرورة.

٣ - الحصر الشامل والعينة معاً حيث يتيح استخدام البيانات والنتائج المتحصل عليها مقارنة نتائج الحصر التام بالعينة وقياس مدى دقة الأخيرة.

والعينة لها فوائد كثيرة منها:

١ - أنها أقل تكلفة حيث يستخدم جزء صغير من العشيرة المطلوب دراستها ولكنه ممثل لها.

٢ - يستغرق الحصول على البيانات المطلوبة عن طريق العينة وقتاً أقل مقارنة بالحصر الشامل وخاصة عندما يكون هناك ضرورة ملحة فى سرعة الحصول على البيانات لكى يتسنى السرعة فى اتخاذ القرار.

٣ - يتطلب الحصول على بيانات تفصيلية، وقد تكون معقدة بعض الشىء، إلى استخدام عدادين مدربين جيداً قد لا تكفى الأعداد المتاحة منهم لتغطية العشيرة ككل.

٤ - ونظراً لأنه يمكن استخدام أفراد عددهم أقل فى حالة العينات عنه فى حالة التعداد التام فيمكن تدريبهم التدريب الجيد وبذلك فإن كفاءتهم تكون أعلى وبالتالى تكون البيانات المتحصل عليها أكثر دقة.

والسؤال الذى يتبادر إلى الذهن: ما هو حجم العينة الأمثل optimum sample size الذى يمكن اختياره؟ فمعروف أن كلما زاد حجم العينة أدى ذلك إلى زيادة دقة التقدير accuracy ولكن ذلك يتطلب أيضاً إنفاقاً أكثر ووقتاً أطول، ومن هنا كان لابد من إجراء موازنة بين دقة التقدير والاقتصاد economy عند تحديد حجم العينة.

٥-٢ الخطوات الرئيسية التى يجب إتباعها فى المعاينة

١- تحديد السؤال (أو الأسئلة) المراد الإجابة عليها فيما يتعلق بالعشيرة المراد دراستها.

٢- تحديد العشيرة التى ستسحب منها العينة تحديداً دقيقاً ومعرفة كل مفرداتها بدقة. وقد لا تكون العشيرة الخاضعة للمعاينة study population هى نفس العشيرة المطلوب دراستها target population ولكن تستخدم الأولى لتعميم نتائجها على الثانية، كأخذ عينة من فئران التجارب لإجراء دراسة عليها ثم تعميم نتائج تلك الدراسة على الإنسان مثلاً.

٣- تحديد البيانات المراد جمعها أو قياسها بدقة وقد يكون ذلك بالمقابلة الشخصية أو بالفحص الطبى مثلاً أو بإجراء تجربة أو بتوجيه أسئلة باستخدام التليفون أو البريد أو بعض منها أو كلها. ويجب قبل اختيار العينة تقسيم العشيرة إلى أجزاء تسمى وحدات المعاينة sampling units منفصلة عن بعضها تماماً بحيث تشمل العشيرة بأكملها. وقد تكون الوحدة مصباحاً كهربائياً أو حقلاً أو مساحة معينة من الأرض محددة الأبعاد وهذه القائمة من جميع وحدات المعاينة يطلق عليها "إطار frame".

٤- تحديد حجم العينة الأمثل طبقاً لتكاليف وحدة المعاينة ودرجة الدقة المطلوبة.

٥- إجراء اختبار مبدئى pre-test حيث يمكن عن طريق تحليل مبدئى pre-analysis أن تعدل طريقة توجيه بعض الأسئلة أو حذف بعضها أو إضافة أسئلة أخرى أو وضع شروط أو قيود معينة عند إجراء التجربة أو جمع معلومة معينة.

٦- فى حالة المعاينة من أجل المسح فإنه يجب تنظيم العمل الميدانى وهو كل ما يختص بإدارة الأعمال من تدريب الأشخاص الذين سيقومون بجمع البيانات وكذلك الإشراف عليهم ووضع الخطط اللازمة فى حالة الفشل فى الحصول على معلومة معينة.

٧- استخدام الطرق الإحصائية المناسبة فى عرض وتلخيص وتحليل البيانات المتحصل عليها ومناقشتها وتفسيرها والاستدلال منها على معالم العشيرة المطلوب دراستها.

٨- يمكن استخدام ما سبق ذكره فى ٧- فى تقدير حجم العينة الأمثل فى الدراسات القادمة كما يمكن التعرف على المشاكل التى واجهت العدادين للتغلب عليها وإيجاد الحلول اللازمة لها وأيضاً الاستفادة منها فى تيسير العمل وبدقة وفى أقل فترة زمنية.

والإخلال بأى من الخطوات الرئيسية السابقة يؤدى إلى الحصول على بيانات غير دقيقة وبالتالى استنتاجات واستدلالات ليست سليمة بالضرورة.

صندوق ٥-١

المعاينة هى وسيلة عملية لدراسة العشائر لعدم إمكان قياس جميع أفراد العشيرة و/أو الاقتصاد فى التكلفة. فالعينة تؤخذ بطريقة محددة ولابد أن يتوافر فيها صفات معينة (سيتم شرح هذه الطرق والشروط فى الفصول التالية) حتى يمكن سحب التقديرات والقرارات المستمدة منها على العشيرة ككل فهي "منه وإليه"

٥-٣ طرق المعاينة Sampling methods

٥-٣-١ المعاينة الاحتمالية Probability sampling

يجرى السحب بتحديد احتمالات الاختيار لكل من الوحدات ثم سحب هذه الوحدات واحدة تلو الأخرى أو فى مجموعات حتى تستكمل العينة. قد تكون السحبات المتتالية فى العينة الاحتمالية مع إعادة الوحدات المختارة فى السحبات السابقة sampling with replacement أو بدون إرجاع sampling without replacement. وحيث إن المعاينة الاحتمالية تخضع لقوانين الاحتمالات وعليه فإنه يمكن حساب أخطاء المعاينة sampling errors وبالتالى مقدار الدقة المتوقعة فى التقديرات.

٥-٣-٢ المعاينة غير الاحتمالية

اختيار عينة بطريقة غير عشوائية. ومن أمثلتها: العينة التى تتكون مفرداتها من متطوعين لاختبار دواء معين فقد لا يوافق الأفراد الذين سيقع عليهم الاختيار، فى حالة

المعاينة الاحتمالية، لإجراء مثل ذلك الاختبار. وفى المعاينة غير الاحتمالية لا يستطيع الباحث قياس دقة النتائج المتحصل عليها حيث إنها لا تخضع لقوانين الاحتمالات كما أنها تؤدى فى الغالب إلى التحيز.

٥-٤ المعاينة العشوائية البسيطة Simple random sampling

هى إحدى صور المعاينة الاحتمالية. فإذا كان حجم العشيرة N ويراد سحب عينة عشوائية بسيطة حجمها n بدون إحلال فإنه يتم سحب وحدات تلك العينة واحدة تلو الأخرى، ويكون احتمال اختيار وحدة من الوحدات من السحب الأول مساوية n/N واحتمال اختيار الوحدة الثانية $(n-1)/(N-1)$ والثالثة $(n-2)/(N-2)$ ... وهكذا، والوحدة الأخيرة فى العينة $1/(N-n+1)$، مع مراعاة أن السحب يتم بدون إحلال without replacement أى أن وحدة المعاينة التى يتم سحبها لا تعاد مرة أخرى إلى العشيرة المسحوبة منها، وبالتالى فإن احتمال سحب وحدات تلك العينة والتى حجمها n من n سحبه بدون إحلال هو

$$(n/N)\cdot[(n-1)/(N-1)]\cdots[1/(N-n+1)] = 1/C_n^N \qquad (٥-١)$$

حيث: $C_n^N = \dfrac{N!}{(N-n)!n!}$ هى عدد التوافيق الممكنة لسحب n من وحدات المعاينة من عشيرة حجمها N.

وبالتالى فإن احتمال سحب أى عينة حجمها n من عشيرة حجمها N هو مقلوب عدد العينات المتساوية فى الحجم والتى لها أيضاً نفس الفرصة فى الاختيار. ويتم سحب مفردات العينة بأن ترقم العشيرة التى سيسحب منها العينة $1,2,\cdots,N$ ثم يتم سحب المفردات التى تكون تلك العينة إما بواسطة جداول الأعداد العشوائية أو عن طريق بعض البرامج الإحصائية باستخدام الحاسبات الآلية أو حتى تلك المتوفرة فى بعض الحاسبات.

ولتوضيح ذلك افترض وجود عشيرة مكونة من 6 وحدات معاينة يرمز لها a, b, c, d, e & f والمطلوب سحب عينة حجمها = 2 من هذه العشيرة فتكون كل العينات الممكن سحبها والتى لها نفس الحجم ونفس الاحتمال:

ab, ac, ad, ae, af, bc, bd, be, bf, cd, ce, cf, de, df, ef

أى عدد العينات الممكن سحبها هو $C_2^6 = \dfrac{6!}{2!4!} = 15$

وبالتالى فإن احتمال سحب أى عينة هو $1/15$، وكل وحدة معاينة تتكرر عدداً متساوياً من المرات فى العينات وهو 5 مرات فى هذه الحالة ويعبر عنه رياضياً بما يلى:

$$C_{n-1}^{N-1} = \frac{(N-1)!}{[(N-n)!(n-1)!]} \qquad (٢-٥)$$

وفى المثال السابق فإن عدد المرات التى يتكرر فيها أى حرف $_{(6-1)}C_{(2-1)} = 5$.

وقد يتم سحب وحدات المعاينة مع الإحلال وفى هذه الحالة فإن احتمال سحب أى مفردة يكون متساوياً ومقداره n/N، وفى هذه الحالة قد تتكرر نفس المفردة وهو أمر قد يكون غير منطقياً مع ملاحظة أن التقديرات وتبايناتها والتى تقدر من العينة تكون أبسط فى حالة المعاينة مع الإحلال عنها فى حالة المعاينة بدون إحلال.

٥-٤-١ اختيار عينة عشوائية بسيطة

يوجد العديد من جداول الأعداد العشوائية والمتوفرة فى كثير من كتب الإحصاء وهى عبارة عن الأرقام $0,1,2,\cdots,9$ موضوعة بطريقة عشوائية. ويمكن اختيار عينة عشوائية بسيطة باستخدام هذه الجداول بإحدى طريقتين وذلك بعد ترقيم وحدات المعاينة فى العشيرة هكذا: $1,2,\cdots,N$.

الطريقة الأولى:

إذا كان حجم العشيرة N وليكن 530 والمطلوب اختيار عينة عشوائية حجمها n ولتكن 10 وحدات. يتم اختيار إحدى الصفحات عشوائياً من جدول الأعداد العشوائية (جدول ١ ملحق أ) وتختار نقطة البداية فى الصفحة المختارة عشوائياً بأن يوضع الإصبع على أى نقطة للبدء منها. وحيث إن حجم العشيرة مكون من 3 أرقام فيتم اختيار 3 أعمدة ولتكن نقطة البداية السطر 00 والأعمدة $60,61,62$ فتكون لدينا الأعداد من 000 إلى 999 وبالتالى تحذف كل الأعداد التى هى أكبر من 530 وأيضاً العدد 000. يتم استعراض أرقام الأعمدة الثلاثة من أعلى إلى أسفل ويتم اختيار الأعداد العشرة التى تمثل العينة وإن لم تستكمل الأعداد العشرة ممكن استخدام الأعمدة التالية للأعمدة السابقة وهى الأعمدة $65,66,67$ وهكذا من أعلى إلى أسفل أيضاً. وإذا تكرر أى عدد فإنه يحذف.

وفي هذه الطريقة عند اختيار عينات مختلفة يجب اختيار نقط بداية جديدة عشوائياً. وبإتباع هذه الطريقة فإن مفردات العينة والتى حجمها 10 هى:

520, 514, 368, 359, 356, 512, 012, 104, 344 & 091

وتكون الأعداد الثلاثية المرفوضة 14 من جملة الأعداد الثلاثية 24 أى أن نسبة الرفض 58% تقريباً.

الطريقة الثانية:

تفضل عندما يكون حجم العشيرة أقل من 500 وليكن 125 مثلاً. تختار عشوائياً نقطة البداية وليكن الصف 00 وثلاثة أعمدة ولتكن 50,51,52. فى هذه الطريقة تحذف الأعداد 000 ويطرح 200 من جميع الأعداد ما بين 201 و 400، ويطرح 400 من جميع الأعداد والتى تقع بين 401 و 600، ويطرح 600 من جميع الأعداد بين 601 و 800، ويطرح 800 من جميع الأعداد من 801 و 999، ويتم رفض أى عدد أو باق أكبر من 125 وأى عدد يتكرر أكثر من مرة.

وبالتالى تكون مفردات العينة المسحوبة هى:

103, 068, 112, 024, 116, 116, 005, 025, 003, 002 & 059

ويلاحظ رفض العدد 116 لتكراره وأن نسبة الرفض فى هذه العينة 13/23 أى 57%.

وعند استخدام الطريقة الثانية فى حالة $N = 320$ (مثلاً) يلاحظ طرح 400 من كل عدد بين 401 و 800 ورفض جميع الأعداد التى هى أكبر من 800 وأيضاً العدد 800 وكل البواقى التى هى أكبر من 320. وطرح 800 بين 801 و 999 حتى لا يعطى احتمال قبول أعلى للبواقى بين 001 و 199 مما يعطيه للبواقى بين 200 و 320.

وإذا استخدمت الطريقة الأولى فى اختيار العينة والتى حجمها 10 من عشيرة حجمها 125 ولو فرضت نفس نقطة البداية السابقة مع حذف كل الأعداد التى هى أكبر من 125 وأيضاً 000 فإن مفردات العينة تكون:

103, 112, 116, 124, 056, 107, 026, 099, 065, 012

وتطلب السحب 90 رقما ثلاثياً لسحب 10 أرقام ثلاثية أى نسبة الرفض بلغت 80/90 أى حوالى 88% وهى أكبر بكثير مقارنة بالطريقة الثانية لسحب العينة من نفس العشيرة ويلاحظ هنا استخدام الأعمدة 50,51,52 والأعمدة 55,56,57 وأيضاً الأعمدة 60,61,62 وكانت نقطة البداية السطر 00.

٥-٤-٢ تقدير معالم العشيرة

يرمز للقيم التى تأخذها وحدات العشيرة بالرموز $Y_1, Y_2, \cdots, Y_N$ حيث حجم العشيرة N والوحدات التى تأخذها العينة بالرموز $Y_1, Y_2, \cdots, Y_n$ حيث n حجم العينة ويلاحظ أن Y_1 فى العينة ليست بالضرورة هى Y_1 فى العشيرة وقد تكون مجرد صدفة.

وفيما يلى أهم المعالم التى يتم تقديرها وسوف يرمز لمعالم العشيرة بالحروف الكبيرة capital letters بينما التقديرات من العينة بالحروف الكبيرة أيضاً مضافاً عليها علامة (^).

	من العينة	من العشيرة
المتوسط	$\hat{\bar{Y}} = \sum_{i=1}^{n} Y_i / n = Y./n$	$\bar{Y} = \sum_{i=1}^{N} Y_i / N = Y./N$
المجموع الكلى للعشيرة	$\hat{Y}. = N\hat{\bar{Y}}$ $= N(\sum Y_i / n)$	$Y. = Y_1 + Y_2 + \ldots + Y_N$ $= \sum Y_i$
التباين (متغير مستمر)	$\hat{S}^2 = \sum (Y_i - \hat{\bar{Y}})/(n-1)$ تقدير غير متحيز لـ S^2	$\sigma^2 = \sum (Y_i - \bar{Y})^2 / N$ $S^2 = \sum (Y_i - \bar{Y})^2/(N-1)$

تباين المتوسط لعينة عشوائية بسيطة

التباين المقدر من العينة

$$S^2_{\hat{\bar{Y}}} = \frac{\hat{S}^2}{n}\left(\frac{N-n}{N}\right) = \frac{\hat{S}^2}{n}(1-f) \qquad (٥-٣)$$

التباين الحقيقى (الفعلى)

$$\sigma^2_{\bar{Y}} = \frac{S^2}{n}\left(\frac{N-n}{N}\right) = \frac{S^2}{n}(1-f) \qquad (٥-٤)$$

الإثبات:

حيث أن

$$n\,(\hat{\bar{Y}} - \bar{Y}) = (Y_1 - \bar{Y}) + (Y_2 - \bar{Y}) + \cdots + (Y_n - \bar{Y}) \qquad (١)$$

$$E(Y_1 + Y_2 + \cdots + Y_n) = \frac{n}{N}(Y_1 + Y_2 + \cdots + Y_N) \qquad (٢)$$

إذاً

$$E[(Y_1 - \bar{Y})^2 + \cdots + (Y_n - \bar{Y})^2] = \frac{n}{N}[(Y_1 - \bar{Y})^2 + \cdots + (Y_N - \bar{Y})^2] \qquad (٣)$$

كما أن

$$\begin{aligned} &E\,[(Y_1 - \bar{Y})\,(Y_2 - \bar{Y}) + (Y_1 - \bar{Y})\,(Y_3 - \bar{Y}) + \cdots + (Y_{n-1} - \bar{Y})\,(Y_n - \bar{Y})] \\ &= \frac{n\,(n-1)}{N\,(N-1)}\,[(Y_1 - \bar{Y})\,(Y_2 - \bar{Y}) + (Y_1 - \bar{Y})\,(Y_3 - \bar{Y}) \\ &\quad + \cdots + (Y_{N-1} - \bar{Y})(Y_N - \bar{Y})] \qquad (٤) \end{aligned}$$

ويلاحظ أن المجموع يحتوى على عدد $\frac{n\,(n-1)}{2}$ من الحدود على اليسار وعلى $\frac{N\,(N-1)}{2}$ من الحدود على اليمين. وبتربيع المعادلة (١) واستخدام (٣) و (٤) فإن

$$(n^2)\,E\,(\hat{\bar{Y}} - \bar{Y})^2 = \frac{n}{N}\left\{ \begin{aligned} &(Y_1 - \bar{Y})^2 + \cdots + (Y_N - \bar{Y})^2 \\ &+ \left[\begin{aligned} &\frac{2\,(n-1)}{N-1}\,[(Y_1 - \bar{Y})\,(Y_2 - \bar{Y}) + \cdots \\ &+ (Y_{N-1} - \bar{Y})\,(Y_N - \bar{Y})] \end{aligned} \right] \end{aligned} \right\}$$

وبإتمام المربعات فإن:

$$(n^2)\,E\,(\hat{\bar{Y}} - \bar{Y})^2 = \frac{n}{N}\left\{ \begin{aligned} &\left(1 - \frac{n-1}{N-1}\right)[(Y_1 - \bar{Y})^2 + \cdots + (Y_N - \bar{Y})^2] \\ &+ \left(\frac{n-1}{N-1}\right)[(Y_1 - \bar{Y}) + \cdots + (Y_N - \bar{Y})]^2 \end{aligned} \right\}$$

ويلاحظ أن الحد الثانى والذى بداخل القوس المربع يساوى صفر، وبقسمة الطرفين على n^2 فإن

$$V(\hat{\bar{Y}}) = E(\hat{\bar{Y}} - \bar{Y})^2 = \left(\frac{N-n}{nN(N-1)}\right)\left(\sum_{i=1}^{N}(Y_i - \bar{Y})^2\right)$$

$$= \frac{S^2}{n}\left(\frac{N-n}{N}\right) = \frac{S^2}{n}(1-f)$$

وبالتالى فإن الخطأ المعيارى لـــ $\hat{\bar{Y}}$

$$\sigma_{\hat{\bar{Y}}} = \frac{S}{\sqrt{n}}\sqrt{(1-f)} \qquad (٥-٥)$$

وتباين $\hat{Y}.$

$$V(\hat{Y}.) = V(N\hat{\bar{Y}}) = \frac{N^2S^2}{n}(1-f) \qquad (٥-٦)$$

وبالتالى فإن الخطأ المعيارى لـــ $\hat{Y}.$

$$\sigma_{\hat{Y}.} = \frac{NS}{\sqrt{n}}\sqrt{(1-f)} \qquad (٥-٧)$$

ملحوظة: فى حالة تقدير التباين من العينة يستخدم $\hat{S}^2$ بدلاً من S^2

ويلاحظ أن الكمية $(N-n)/N$ تعرف بمعامل التصحيح للعشيرة المحدودة finite population correction factor والتى يمكن إعادة كتابتها على الصورة $1-(\frac{n}{N}) = 1-f$ حيث يطلق على $f = \frac{n}{N}$ كسر المعاينة sampling fraction ويمكن إهمال كسر المعاينة إذا كان أقل من 5% . ويشير بعض الإحصائيين إلى أنه يمكن إهمـــال كســـــر المعاينة حتى وإن وصل إلى 10% .

٥-٤-٣ خصائص التقديرات

كما ذكر من قبل فإن الهدف من المعاينة هو تقدير معالم العشيرة من العينة المسحوبة، وحتى توفى العينة بهذا الهدف يجب أن تتوافر فى هذه التقديرات خصائص معينة هى:

١- عدم التحيز unbiasdness
٢- الاتساق consistency
٣- أقل تباين minimum variance
٤- الكفاية sufficiency

وتعتمد دقة أى تقدير من العينة على الطريقة التى يحسب بها هذا التقدير من العينة وأيضاً على خطة المعاينة وهو ما يطلق عليه دقة المعاينة العشوائية البسيطة. وفيما يلى أهم تلك الخصائص للمتوسط الحسابى.

١ – عدم التحيز Unbiasedness

وهذا يعنى أن التقدير المحسوب من العينة $\hat{\bar{Y}}$ يكون غير متحيز لمعلم العشيرة $\bar{Y}$. ويمكن توضيح ذلك كالتالى:

حيث إن

$$E(\hat{\bar{Y}}) = \frac{\sum \hat{\bar{Y}}}{{}_N C_n} = \frac{\sum (Y_1 + Y_2 + \cdots + Y_n)}{(n)\dfrac{N!}{n!(N-n)!}}$$

وحيث إن كل Y_i تظهر فى عدد من العينات مقداره ${}_{N-1}C_{n-1}$ من معادلة (٥-٢) وبالتالى فإن:

$$\begin{aligned}\sum (Y_1 + Y_2 + \cdots + Y_n) &= {}_{N-1}C_{n-1}\,(Y_1 + Y_2 + \cdots + Y_N)\\ &= \frac{(N-1)\,!}{(n-1)\,!\,(N-n)!}\,(Y_1 + Y_2 + \cdots + Y_N)\end{aligned}$$

$$\begin{aligned}E\,(\hat{\bar{Y}}) &= \left(\frac{(N-1)!}{(n-1)\,!\,(N-n)\,!}\right)\left(\frac{n\,!\,(N-n)\,!}{n\,N\,!}\right)(Y_1 + Y_2 + \cdots + Y_N)\\ &= \frac{Y_1 + Y_2 + \cdots + Y_N}{N} = \bar{Y}\end{aligned}$$

أى أن $\hat{\bar{Y}}$ هو تقدير غير متحيز لمتوسط العشيرة. ومنه فإن $\hat{Y}_. = N\,\hat{\bar{Y}}$ هو تقدير غير متحيز لمجموع العشيرة $Y_.$

٢ – الاتساق Consistency

وهذا يعنى أن $\hat{\bar{Y}}$ و $N\hat{\bar{Y}}$ هما تقديرين متسقين لمتوسط ومجموع العشيرة على الترتيب. فالمتوسط $\hat{\bar{Y}}$ يقال إنه تقدير متسق consistent لمتوسط العشيرة $\bar{Y}$ إذا كان احتمال القيمة المطلقة للفرق بين القيمة الحقيقية والتقدير أكبر من قيمة معينة ولتكن ε

مهما تكن يؤول إلى الصفر كلما آلت n إلى مالا نهاية $n \to \infty$ ويمكن التعبير عن ذلك رياضياً بالصورة $\Pr(\hat{\overline{Y}} - \overline{Y}) > \varepsilon) \to 0$ as $n \to \infty$. ويمكن إثبات ذلك باستخدام لا متساوية تشيبتشف Tchebychef 's inequality.

٣ - الكفاءة Minimum variance

تدل على أن المتوسط له أقل تباين وهو أيضاً يتحقق.

٤ - الكفاية Sufficiency

يقال أن المتوسط كاف sufficient statistic إذا كان يأخذ فى الاعتبار عند حسابه كل قيم وحدات المعاينة وهذا أيضاً يتحقق.

مثال ٥-١

فى عشرة من 5 وحدات كانت قيم Y_i هى 13,9,6,3,5 . احسب متوسط العينة فى كل من العينات العشوائية البسيطة الممكنة والتى حجم كل منها = 2 . ثم تحقق من أن $\hat{\overline{Y}}$ هو تقدير غير متحيز لمتوسط العشيرة. احسب التباين لكل من العينات وتحقق من أن $E(\hat{S}^2) = S^2$.

عدد العينات التى حجم كل منها 2 هو $_5C_2 = \dfrac{5\,!}{3\,!\,2\,!} = 10$

احتمال سحب أى عينة عشوائية $\dfrac{1}{_5C_2} = 1 \div \dfrac{5!}{3!2!} = \dfrac{1}{10}$

ويبين جدول ٥-١ العينات العشوائية البسيطة الممكنة وتباينها. وباستخدام بيانات هذا الجدول فإن:

المتوسط العام لمتوسطات العينات: $\dfrac{4 + 5.5 + \cdots + 11}{10} = 7.2$

وهو يساوى متوسط العشيرة 7.2

وبالتالى فإن $\hat{\overline{Y}}$ هو تقدير غير متحيز لمتوسط العشيرة.

تباين العشيرة:

$$S^2 = \sum_{c=1}^{5}(Y_i - \overline{Y})^2 / (N-1)$$

$$= \frac{(5-7.2)^2 + (3-7.2)^2 + \cdots + (13-7.2)^2}{4} = 15.2$$

$$E(\hat{S}^2) = \frac{2 + 0.5 + \cdots + 8}{10} = 15.2$$

وبالتالى فإن $E(\hat{S}^2) = S^2$

جدول ٥-١ العينات العشوائية البسيطة الممكنة وتباينها لمثال ٥-١

العينة	**المتوسط** $\hat{\overline{Y}}$	**التباين** $\hat{S}^2 = \sum(Y_i - \hat{\overline{Y}})/(n-1)$
3, 5	4.0	2.0
6 , 5	5.5	0.5
9 , 5	7.0	8.0
13, 5	9.0	32.0
6, 3	4.5	4.5
9, 3	6.0	18.5
13, 3	8.0	50.0
9, 6	7.5	4.5
13, 6	9.5	24.5
13, 9	11.0	8.0

مثال ٥-٢

البيانات التالية تمثل الدخل الشهرى بالجنيه للعامل لعينة عشوائية بسيطة حجمها 30 مسحوبة من مصنع به 250 عامل

328	326	337	330	330	333	339	329	331	349
336	338	324	333	340	334	341	329	329	340
344	335	345	332	326	335	336	342	335	337

والمطلوب تقدير لمتوسط العشيرة وقيمته الكلية ثم حساب تباين المتوسط والخطأ المعيارى له وللقيمة الكلية.

إذا رمز للدخل الشهرى للعامل رقم i بـ Y_i

فإن $$\hat{\bar{Y}} = \sum_{i=1}^{30} Y_i / n = (328 + 339 + \cdots + 337)/30 = 10043/30 = 334.77 \text{ LE}$$

والتباين $$\hat{S}^2 = \frac{1}{n-1}\left[\sum Y_i^2 - \frac{(\sum Y_i)^2}{n}\right] = 36.87$$

وتباين المتوسط $$\hat{S}^2_{\bar{Y}} = \frac{\hat{S}^2}{n}\left(\frac{N-n}{N}\right) = \frac{36.87}{30}\left(\frac{250-30}{250}\right) = 1.08$$

الخطأ المعيارى للمتوسط $$\hat{S}_{\bar{Y}} = \sqrt{1.08} = 1.04 \text{ LE}$$

مجموع الدخل الشهرى للعمال فى المصنع $$N\hat{\bar{Y}} = 250 \text{ x } 334.77 = 83693 \text{ LE}$$

تباين القيمة الكلية:

$$\hat{S}^2_{\hat{Y}.} = \frac{N^2 S^2}{n}\left(\frac{N-n}{N}\right) = \frac{(250)^2(36.87)^2(250-30)}{(30)(250)} = 67594.998$$

الخطأ المعيارى للقيمة الكلية $$\hat{S}_{\hat{Y}.} = \sqrt{67594.998} = 259.99 \text{ LE}$$

٥–٤–٤ حدود الثقة Confidence limits

لحساب حدى الثقة لمتغير ما يجب معرفة توزيع هذا المتغير. والكثير من المتغيرات التى نتعامل معها تتبع التوزيع الطبيعى أو على الأقل تميل أو تقترب من هذا التوزيع حيث إن التقريب الطبيعى ملائم لمعظم التطبيقات العملية من مصادر عدة. وقد وجد من الدراسات الخاصة بنظرية الاحتمالات أنه كلما زاد حجم العينة اقترب توزيع متوسطات العينات المسحوبة من هذه العشيرة إلى أن يكون طبيعياً وذلك يتحقق فى العشائر اللانهائية والتى لها انحراف معيارى منته. أما فى حالة العشائر المحددة فقد يحتاج الأمر إلى حجم أكبر للعينة يخضع لشروط لازمة وكافية والتى ينتهى تحتها توزيع متوسط عينة إلى التوزيع الطبيعى (Hajck, 1960). ويفشل التقريب الطبيعى عندما تحتوى العشيرة على وحدات متطرفة حيث تؤثر أيضاً على زيادة تباين العينة وانخفاض الدقة، وفى هذه الحالة قد يكون أخذ تعداد كامل أكثر حكمة أو عزل تلك الوحدات الشاذة إذا كان عددها ليس بالكثير، وقد يخفض إزاحة هذه

الوحدات المتطرفة من الالتواء ويقترب بالتوزيع من التوزيع الطبيعى. كما يمكن اللجوء إلى المعاينة الطبقية stratified sampling.

وعلى العموم فإن حدود الثقة لمتوسط العشيرة ومجموعها تكون كما يلى:

للمتوسط:

$$\hat{\bar{Y}} \pm \frac{t\,\hat{S}}{\sqrt{n}}\sqrt{1-f} \qquad (٥-٨)$$

وللمجموع:

$$N\hat{\bar{Y}} \pm \frac{t\,N\,\hat{S}}{\sqrt{n}}\sqrt{1-f} \qquad (٥-٩)$$

حيث t هى قيمة t الجدولية بمستوى معنوية معين وليكن α ودرجات حرية $n-1$ والتى حسب منها التباين.

مثال ٥-٣

فى المثال ٥-٢ احسب مدى الثقة لمتوسط دخل العامل فى المصنع عند مستوى معنوية $\alpha = 0.05$.

بتطبيق المعادلة (٥-٨) وحيث إن قيمة $t_{(.05,29)} = 2.045$ (جدول ٤ ملحق أ) فإن حدا الثقة هما:

$$334.77 \pm 2.045\,(1.04) = 334.77 \pm 2.13$$

وبالتالى فإن الحد الأعلى للثقة هو 336.9 جنيهاً و الحد الأدنى للثقة هو 332.6 جنيهاً

٥-٤-٥ تقدير نسبة (نسبة مجموعتين أو متوسطين)

قد تكون الكمية المراد تقديرها من العينة العشوائية البسيطة هى نسبة بين متغيرين يتغير كل منهما من وحدة معاينة إلى أخرى. ففى الإحصاءات المنزلية قد يكون من المفيد على سبيل المثال معرفة متوسط عدد أجهزة التكييف لكل شقة أو متوسط عدد الأطفال الذكور فى كل أسرة. ولكى تحصى هذه المفردات يمكن أن تسجل فى الشقة رقم i (حيث $i = 1, 2, \cdots, n$): عدد أجهزة التكييف (Y) وأيضاً عدد الحجرات (X) (مع اعتبار أن الصالة حجرة من الحجرات) وبالتالى يكون:

متوسط العشيرة:

$$R = \frac{\sum_{i=1}^{N} Y_i}{\sum_{i=1}^{N} X_i} \qquad (٥-١٠)$$

والتقدير من العينة:

$$\hat{R} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i} = \frac{\hat{\bar{Y}}}{\hat{\bar{X}}} \qquad (٥-١١)$$

وتظهر النسب أيضاً فى العديد من التطبيقات الأخرى فمثلا نسبة المساحة المزروعة بالذرة الشامية إلى مجموع المساحة فى المزرعة أو نسبة عدد الذكور إلى عدد الإناث فى الأقسام المختلفة فى إحدى الكليات الجامعية. ويلاحظ أن كلاً من البسط والمقام فى هذه الحالات متغير من عينة إلى أخرى وهذا يؤدى إلى أن يكون التوزيع الاحتمالى أكثر تعقيداً. وفي العينات ذات الحجم الصغير نسبياً يلاحظ أن $\hat{R}$ غير متناظر ويعد تقديراً منحازاً لــ R (عند تطبيق خصائص التقدير الجيد). لكن فى العينات كبيرة الحجم فإنه يمكن اعتبار أن R تتوزع طبيعياً.

تباين $\hat{R}$:

$$V(\hat{R}) \cong \left(\frac{1-f}{n\bar{X}^2}\right)\left(\frac{\sum_{i=1}^{N}(Y_i - RX_i)^2}{N-1}\right) \qquad (٥-١٢)$$

حيث $R = \bar{Y}/\bar{X}$ هى نسبة متوسطى العشيرة و $f = n/N$

وفى حالة استخدام العينة فإن:

$$\frac{\sum_{i=1}^{n}(Y_i - \hat{R}X_i)^2}{n-1}$$ يعتبر تقديراً للمقدار $$\frac{\sum_{i=1}^{N}(Y_i - RX_i)^2}{N-1}$$

والخطأ المعيارى لـــ $\hat{R}$

$$S(\hat{R})=\left(\frac{\sqrt{1-f}}{(\sqrt{n})(\bar{X})}\right)\sqrt{\frac{\sum_{i=1}^{n}(Y_i-\hat{R}X_i)^2}{n-1}} \qquad (١٣-٥)$$

وإذا كان $\bar{X}$ غير معروف يعوض عنه بتقدير من العينة $\hat{\bar{X}}$

كما يمكن حساب $S(\hat{R})$ كما يلى:

$$S(\hat{R})=\left(\frac{\sqrt{1-f}}{(\sqrt{n})(\bar{X})}\right)\sqrt{\frac{\sum Y_i^2-2\hat{R}\sum Y_iX_i+\hat{R}^2\sum X_i^2}{n-1}} \qquad (١٤-٥)$$

مثال ٥-٤

الجدول التالى يبين عدد كل من الأبناء الذكور والإناث لعدد 20 أسرة من العاملين بإحدى الكليات الجامعية اختيروا بطريقة عشوائية من بين 500 أسرة.

رقم الأسرة	عدد أفراد الأسرة m	عدد الذكور X_1	عدد الإناث X_2	رقم الأسرة	عدد أفراد الأسرة m	عدد الذكور X_1	عدد الإناث X_2
١	2	1	1	١١	3	0	3
٢	2	2	0	١٢	3	1	2
٣	4	2	2	١٣	3	0	3
٤	4	1	3	١٤	4	2	2
٥	0	0	0	١٥	4	2	2
٦	3	3	0	١٦	3	2	1
٧	4	3	1	١٧	2	0	2
٨	3	2	1	١٨	5	3	2
٩	5	2	3	١٩	4	2	2
١٠	3	0	3	٢٠	3	1	2

والمطلوب تقدير:

١- متوسط عدد الذكور ومتوسط عدد الإناث للأسرة الواحدة والخطأ المعيارى لهما.

٢- متوسط عدد الأبناء والخطأ المعيارى له بغض النظر عن الجنس للأسرة الواحدة.

٣- النسبة المئوية والخطأ المعيارى للعدد الكلى من الذكور إلى العدد الكلى من الإناث.

من البيانات:

$$n = 20 \qquad \sum X_1 = 29 \qquad \sum X_2 = 35 \qquad \sum X_1 X_2 = 43$$

$$N = 500 \qquad \sum X_1^2 = 63 \qquad \sum X_2^2 = 81$$

١- متوسط عدد الذكور للأسرة الواحدة:

$$\overline{X}_1 = \frac{\sum X_1}{n} = \frac{29}{20} = 1.45$$

الخطأ المعيارى لمتوسط عدد الذكور للأسرة الواحدة:

$$\hat{S}_{\overline{X}_1} = \left(\frac{\sqrt{1-f}}{\sqrt{n}}\right)\sqrt{\frac{\sum X_1^2 - \frac{(\sum X_1)^2}{n}}{n-1}} = \left(\frac{\sqrt{1-\frac{20}{500}}}{\sqrt{20}}\right)\sqrt{\frac{63-\frac{(29)^2}{20}}{19}} = 0.23$$

وإذا أهمل معامل التصحيح للعشيرة المحدودة حيث إنه يمثل $(20/500)(100) = 4\%$ فقط فإن:

$$\hat{S}_{\overline{X}_1} = \frac{1}{\sqrt{n}}\sqrt{\frac{\sum X_1^2 - \frac{(\sum X_1)^2}{n}}{n-1}} = 0.235$$

الفرق بين التقديرين 0.005، أى أنه فى حالة إهمال معامل التصحيح للعشيرة المحدودة فإن الخطأ المعيارى يزيد بمقدار 2.17%.

متوسط عدد الإناث للأسرة الواحدة:

$$\overline{X}_2 = \frac{\sum X_2}{n} = \frac{35}{20} = 1.75$$

الخطأ المعيارى لمتوسط عدد الإناث للأسرة الواحدة:

$$\hat{S}_{\overline{X}_2}=\left(\frac{\sqrt{1-f}}{\sqrt{n}}\right)\sqrt{\frac{\sum X_2^2-\frac{(\sum X_2)^2}{n}}{n-1}}=\left(\frac{\sqrt{1-\frac{20}{500}}}{\sqrt{20}}\right)\sqrt{\frac{81-\frac{(35)^2}{20}}{19}}=0.22$$

٢- متوسط عدد الأبناء من الذكور والإناث معاً للأسرة الواحدة:

$$\overline{Y}=\frac{\sum m_i}{n}=\frac{2+2+\cdots+3}{20}=\frac{64}{20}=3.2$$

تباين متوسط عدد الأبناء

$$\hat{S}^2_{\hat{\overline{Y}},}=[\sum m^2-\frac{(\sum m)^2}{n}]/(n-1)=[(2)^2+(2)^2+\cdots+(3)^2-\frac{(64)^2}{20}]/19$$
$$=1.326$$

وعليه فإن الانحراف المعيارى لمتوسط عدد الأبناء

$$\hat{S}_{\hat{\overline{Y}}}=\sqrt{\frac{\hat{S}^2_{\hat{\overline{Y}}}}{n}(1-f)}=\sqrt{\left(\frac{1.326}{20}\right)\left(1-\frac{20}{50}\right)}=0.25$$

٣- نسبة العدد الكلى من الذكور إلى العدد الكلى من الإناث:

$$\hat{R}=(\sum X_1/\sum X_2)=29/35=0.829$$

الانحراف المعيارى:

$$S(\hat{R})=\frac{\sqrt{1-f}}{(\sqrt{n})\,(\overline{X}_2)}\sqrt{\frac{\sum X_1^2-2\hat{R}\sum X_1X_2+\hat{R}^2\sum X_2^2}{n-1}}$$
$$=\frac{\sqrt{0.96}}{(\sqrt{20})(1.75)}\sqrt{\frac{63-2(0.829)(43)+(0.829)^2(81)}{19}}=0.2$$

٥-٤-٦ معاينة النسب والنسب المئوية

قد يكون الغرض من المعاينة هو التعامل مع متغيرات لا تتبع التوزيعات المستمرة وإنما تنتمى إلى التوزيعات المتقطعة discrete distribution ومثال ذلك تلك المتغيرات التى تأخذ القيم فيها صفراً أو واحداً كتقسيم العشيرة إلى مدخنين وغير مدخنين، وقد تقسم العشيرة إلى عدة فئات A, B, C مثلا وقد يكون الاهتمام بالفئة A فيأخذ المتغير فى هذه الحالة القيمة واحد وفيما عدا ذلك من فئات يأخذ المتغير فيها القيمة صفراً ... وهكذا. وقد يكون المتغير أصلا مستمراً ولكن تم تقسيمه إلى فئات لا تداخل بينها وبالتالى يصبح متقطعاً كتقسيم أعمار العشيرة إلى فئات: أقل من 20 سنة ومن 20 إلى أقل من 30 ... وهكذا. والمطلوب فى مثل تلك الحالات تقدير العدد الكلى أو النسبة المئوية لوحدات فى العشيرة تتسم بخاصية مميزة أو تقع فى فئة معرفة. وفيما يلى مثال يوضح ذلك المفهوم.

افترض أن كل وحدة فى العشيرة تقع فى أحد الفئتين "ف" أو مكملتها " ف' " وبالتالى يكون:

عدد الوحدات من الفئة "ف" فى		نسبة الوحدات من الفئة "ف" فى	
العينة	العشيرة	العينة	العشيرة
a	A	$\hat{P} = a/n$	$P = A/N$

أى أن $\hat{P}$ تقدير لـــ P وتقدير العينة لـــ A هو $N\hat{P}$ أى $N(a/n)$، ومثل هذا المتغير يتبع توزيع "ذو الحدين" فى العشائر اللانهائية وقد يتبع التوزيع فوق الهندسى hypergeometric distribution فى العشائر المحدودة. وعلى العموم فإن توزيع "ذو الحدين" قد يؤخذ به تقريبياً.

تباين التقديرات:

إذا كان المتغير Y محل الدراسة يأخذ قيماً إما صفر أو واحد صحيح وبتطبيق ما سبق دراسته عند دراسة المعاينة للمتغيرات المستمرة فإن:

من العشيرة:

$$\overline{Y} = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{A}{N} = P \quad ، \quad Y. = \sum_{i=1}^{N} Y_i = A$$

ومن العينة:
$$\hat{\bar{Y}} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{a}{n} = \hat{P} \quad ، \quad \hat{Y}_. = \sum_{i=1}^{n} Y_i = a$$

التباين من العشيرة:
$$S^2 = \frac{\sum (Y_i - \bar{Y})^2}{N-1} = \frac{\sum_{i=1}^{N} Y_i^2 - N\bar{Y}^2}{N-1} \qquad (٥-١٥)$$
$$= \frac{1}{N-1}(NP - NP^2) = \frac{N}{N-1}P(1-P) = \frac{N}{N-1}PQ$$

التباين من العينة:
$$\hat{S}^2 = \frac{n}{n-1}\hat{P}\hat{Q} \qquad (٥-١٦)$$

تباين $\hat{P}$:
$$S_{\hat{P}}^2 = E(\hat{P} - P)^2 = \frac{S^2}{n}\left(\frac{N-n}{N}\right) = \frac{PQ}{n}\left(\frac{N-n}{N}\right) \qquad (٥-١٧)$$

ومن العينة:
$$\hat{S}_{\hat{P}}^2 = \hat{P}\hat{Q}\frac{N-n}{N(n-1)} \qquad (٥-١٨)$$

وإذا كانت N كبيرة جداً بالنسبة لـ n فإن $\hat{S}_{\hat{P}}^2 = \hat{P}\hat{Q}/(n-1)$ وتباين $\hat{A}$ حيث $\hat{A} = N\hat{P}$ وهو المجموع المقدر لوحدات الفئة "ف" وبالتالى:
$$S_{\hat{A}}^2 = \frac{N^2}{n}\left(\frac{N-n}{N-1}\right)PQ \qquad (٥-١٩)$$

ومن العينة:
$$\hat{S}_{\hat{A}}^2 = \frac{N(N-n)}{n-1}\hat{P}\hat{Q} \qquad (٥-٢٠)$$

ويعتبر $\hat{S}_{\hat{A}}^2$ تقديراً غير متحيز لـ $\hat{S}_{\hat{A}}$

مثال ٥-٥

ينتج مصنع 100 ألف جهاز تليفزيون فى الشهر. وفى عينة من 100 جهاز وجد أن 5 أجهزة منها بها عيوب فى الصناعة. فما هو تقدير العدد الكلى لأجهزة التليفزيون التى بها عيوب فى الصناعة فى الشهر ؟. وما هو الخطأ المعيارى لهذا التقدير؟.

$N = 100000 \qquad n = 100$

نسبة الأجهزة التى بها عيوب فى الصناعة: $\hat{P} = a/n = 5/100 = 0.05$

تقدير العدد الكلى للأجهزة التى بها عيوب فى الصناعة:

$$\hat{A} = N\hat{P} = (100000)(0.05) = 5000$$

الخطأ المعيارى للتقدير: $\hat{S}_{\hat{A}} = \sqrt{(100000)(99900)(.05)(.95)/(99)} = 2189.33$

وحيث إن $n/N = 100/100000 = 0.001$ وهى تمثل 0.1% وبالتالى يمكن وضع N بدلاً من $N-n$

وعليه فإن: $\hat{S}_{\hat{A}} = (N)\sqrt{\hat{P}\hat{Q}/(n-1)} = 2190.2$

وهذه القيمة لا تختلف كثيراً عن القيمة السابق حسابها. وإذا كان حجم العينة كبيراً فإنه يمكن أيضاً وضع n بدلاً من $n-1$

وبالتالى تكون المعادلة $\hat{S}_{\hat{A}} = (N)\sqrt{\hat{P}\hat{Q}/n}$

حدود الثقة للـ P

حيث إن المتغير يتبع فعلاً التوزيع فوق الهندسى وبالتالى فإن أحد الأشكال هو التقريب الطبيعى حيث تتوافر طرق عديدة لحساب حدود الثقة.

$$\hat{P} \pm \left[Z\sqrt{1-f}\sqrt{\hat{P}\hat{Q}/(n-1)} + \frac{1}{2n} \right] \qquad (٥-٢١)$$

حيث $f = n/N$ و Z قيمة المتغير الطبيعى الموافق لاحتمال الثقة. أما $1/2n$ وهو الحد الأخير فى الطرف الأيمن فهو تصحيح من أجل الاستمرارية continuity

مثال ٥-٦

في المثال ٥-٥ احسب حدى الثقة لـــ P عند مستوى معنوية 5%

بتطبيق المعادلة (٥-٢١) فإن حدى الثقة هما

$$0.05 \pm \left[1.96\sqrt{.999}\sqrt{(.05)(.95)/99} + \frac{1}{(2)(100)}\right] = 0.5 \pm .0479$$

وبالتالي فإن الحد الأعلى هو .0979 والحد الأدنى هو .0021.

ويلاحظ هنا تأثير $\hat{P}$ عند حساب الخطأ المعياري، ولكي يقل الفرق بين الحدين الأدنى والأعلى في مثل هذه الحالة فيجب زيادة حجم العينة المسحوبة من العشيرة.

٥-٤-٧ تقدير النسب في المعاينة العنقودية

Estimation of proportions in cluster sampling

قد تمثل كل وحدة معاينة عنقوداً من العناصر عند تقدير نسبة العناصر الواقعة في الفئة "ف". مثال ذلك إذا كانت وحدة المعاينة هي الأسرة وكانت العناصر هنا هو عدد كل من الأبناء الذكور والإناث في كل أسرة أو أن تكون وحدة المعاينة صناديق البرتقال المصدرة إلى الخارج والعناصر هي أعداد البرتقال التالفة والسليمة في كل صندوق ... الخ. وقد تكون أعداد العناصر متساوية وقد لا تكون في كل وحدة من وحدات المعاينة.

أولاً: عندما تكون أعداد العناصر متساوية في كل وحدات المعاينة ولتكن m:

إذا كانت نسبة العناصر التي تنتمي إلى الفئة "ف" هي $P_i = a_i / m$ فإن نسبة وحدات العينة الواقعة في "ف" هي:

$$\hat{P} = \frac{\sum a_i}{n\,m} = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\hat{P}_i \qquad (٥-٢٢)$$

وبالتالي فإن التباين الحقيقي

$$\sigma_{\hat{P}}^2 = \left(\frac{1-f}{n}\right)\left(\frac{\sum_{i=1}^{N}(P_i - P)^2}{N-1}\right) \qquad (٥-٢٣)$$

وتقديره من العينة هو :

$$\hat{\sigma}_P^2 = \left(\frac{1-f}{n}\right)\left(\frac{\sum_{i=1}^{n}(P_i - \hat{P})^2}{n-1}\right) \qquad (٥-٢٤)$$

مثال ٥-٧

من بين 1000 صندوق معبأة بالبرتقال ومعدة للتصدير سحبت 5 صناديق بطريقة عشوائية بسيطة وقدر عدد البرتقالات التالفة فى كل صندوق علماً بأن كل صندوق يحتوى على مائة برتقالة وكان عدد الثمار التالفة هى 3,1,0,2,1 فى الصناديق المسحوبة. قدر العدد الكلى للثمار التالفة مع حساب الانحراف المعيارى.

$$\hat{P} = \frac{\sum a_i}{n\,m} = \frac{1+2+0+1+3}{(5)(100)} = 0.014$$

العدد الكلى للثمار التالفة فى كل الصناديق المعدة للتصدير:

$$(1000)(100)(0.014) = 14000$$

ويمكن تجاهل معامل التصحيح للعشيرة المحدودة حيث إن $f = 5/1000$ وقيمتها ضئيلة جداً، وبالتالى فإن:

$$\hat{\sigma}_P^2 = \frac{\sum(P-\hat{P})^2}{n(n-1)} = [\sum P_i^2 - (\sum P_i)^2/n]/[n(n-1)]$$

$$\hat{P}_1 = 1/100\ ,\ \hat{P}_2 = 2/100\ ,\ \hat{P}_3 = 0\ ,\ \hat{P}_4 = 1/100\ ,\ \hat{P}_5 = 3/100$$

$$\hat{\sigma}_P^2 = \frac{.0015 - .00098}{(5)(4)} = 0.000026$$

تباين العدد الكلى من الثمار التالفة:

$$(Nm)^2\hat{\sigma}_P^2 = [(1000)(100)]^2(0.000026) = 260000$$

الانحراف المعيارى للعدد الكلى من الثمار التالفة:

$$\sqrt{(Nm)^2\hat{\sigma}_P^2} = \sqrt{260000} = 509.9$$

ثانياً: عندما تكون أعداد العناصر غير متساوية فى وحدات المعاينة ويرمز لكل بالرمز m_i فإن:

$$\hat{P}_i = a_i / m_i \qquad (٢٥-٥)$$

ونسبة الوحدات فى العينة الواقعة فى الفئة "ف" هى:

$$\hat{P} = \sum a_i / \sum m_i \qquad (٢٦-٥)$$

وبالتالى فإن تباين $\hat{P}$:

$$\sigma_P^2 \cong \left(\frac{1-f}{n\overline{M}^2}\right)\left(\frac{\sum (a_i - Pm_i)^2}{N-1}\right) \qquad (٢٧-٥)$$

حيث P هى نسبة عناصر الفئة "ف" فى العشيرة و $\overline{M} = \sum^{N} m_i / N$ هو متوسط عدد العناصر فى العنقود الواحد والعبارة البديلة هى

$$\sigma_P^2 \cong \left(\frac{1-f}{n}\right)\sum^{N}\left(\frac{m_i}{\overline{M}}\right)^2\left(\frac{(P_i - P)^2}{N-1}\right) \qquad (٢٨-٥)$$

ويقدر هذا التباين من العينة كما فى المعادلة التالية

$$\hat{\sigma}_P^2 = \left(\frac{1-f}{(n)(\overline{m})^2}\right)\left(\frac{\sum a_i^2 - 2\hat{P}\sum a_i m_i + \hat{P}^2 \sum m_i^2}{n-1}\right) \qquad (٢٩-٥)$$

حيث $\overline{m} = (\sum m)/n$ هو متوسط عدد العناصر للعنقود الواحد فى العينة.

مثال ٥-٨

من بيانات المثال ٥-٤ قدر تباين نسبة الذكور وقارنه مع التقدير "ذو الحدين" للتباين.

فى حالة توزيع "ذو الحدين": $\hat{P} = 29/64 = 0.453$ ، $n = 64$

$$\hat{\sigma}_{\hat{P}}^2 = \hat{P}\hat{Q}/n = 0.00387$$

ومن أجل علاقة النسبة لاحظ وجود 20 عنقوداً أى أن: $n = 20$

العدد الكلى لوحدة المعاينة (الأسرة) i: = m_i

عدد الذكور فى الأسرة i: = a_i

وبالتالى $\overline{m} = (\sum m)/n = 64/20 = 3.2$

$$\sum m_i^2 = 320 \text{ ، } \sum a_i m_i = 106 \text{ ، } \sum a_i = 63$$

وبتطبيق المعادلة (٥-٢٩) مع تجاهل معامل التصحيح

$$\hat{\sigma}_{\hat{P}}^2 = \frac{63-(2)(0.453)(106)+(0.543)^2(230)}{(20)(3.2)^2(19)} = 0.00364$$

مما سبق يتضح أن التباين 0.00387 فى حالة توزيع "ذو الحدين" وهو أكبر بقليل عنه فى حالة النسبة وقد تكون أقل. وقد تكون هذه الفروق راجعة لنسبة الجنس والتى تقترب من 0.5 والتى قد لا تكون فى بعض الأسر.

صندوق ٥-٢

تنقسم طرق اختيار العينة إلى:

طرق احتمالية: وهى التى يختار فيها المجرب العينة بطريقة عشوائية تماماً مستخدماً بذلك جداول الأرقام العشوائية أو الحاسبات.

طرق غير احتمالية: وفيها يكون أفراد العينة متطوعين مثلاً لاختبار عقار معين.

والطرق الاحتمالية هى الأفضل وينصح بإتباعها كلما أمكن ذلك.

وإذا اختيرت العينة الاحتمالية بطريقة جيدة وبأعداد مناسبة فإن التقديرات المتحصل عليها تتميز بأنها غير متحيزة ومتسقة ولها أقل تباين وكافية. وهذه الصفات الأربع تم شرحها فى هذا الجزء من المؤلف.

٥-٤-٨ تقدير حجم العينة Estimation of sample size

تبين مما سبق أن اختيار عينة كبيرة جداً يتضمن هدراً للمصادر بينما العينة الصغيرة جداً تقلل من فائدة النتائج. وفى الحقيقة لا يمكن اتخاذ قرار سليم دائماً فى اختيار الحجم المناسب للعينة حيث إن المعلومات الكافية غير متوفرة دائماً.

أولاً: تحديد حجم العينة عند معاينة النسب

كما سبق فإن الوحدات تقسم إلى فئتين "ف" ، "ف' " فإذا كان الخطأ المسموح به (d) فى النسبة المقدرة لوحدات الفئة "ف" وهى $\hat{P}$ وإنه يمكن تجاوز الخطأ الفعلى للمقدار d ولكن باحتمال صغير α فإن المطلوب تحقيق $\Pr\left|(\hat{P}-P)\right| \geq d = \alpha$ وذلك فى معاينة عشوائية بسيطة مع افتراض أن $\hat{P}$ تتوزع طبيعياً وحيث إن:

$$\sigma_{\hat{P}} = \sqrt{\frac{PQ}{n}}\sqrt{\frac{N-n}{N-1}}$$

وعليه فإن العلاقة التى تربط n بالدرجة المطلوبة للدقة هى:

$$d = (Z)\left(\sqrt{\frac{PQ}{n}}\sqrt{\frac{N-n}{N-1}}\right)$$

حيث Z هى قيمة المتغير الطبيعى المعيارى التى تقطع مساحة قدرها $\alpha/2$ من كل من الذيلين (جدول ٣ ملحق أ) ومنها فإن:

$$n = \frac{Z^2\hat{P}\hat{Q}}{d^2}\Bigg/\left[1+\frac{1}{N}\left(\frac{Z^2PQ}{d^2}-1\right)\right]$$

وإذا كانت N كبيرة واستخدمت $\hat{P}$ بدلاً من P فإن التقريب يعطى

$$n_\circ = \frac{Z^2\hat{P}\hat{Q}}{d^2} = \frac{\hat{P}\hat{Q}}{V} \qquad (٥-٣٠)$$

حيث التباين المرغوب لنسبة العينة

$$V = \frac{\hat{P}\hat{Q}}{n_\circ}$$

وعند التطبيق تحسب أولاً $n_\circ$ وإذا كانت $n_\circ / N$ صغيرة يمكن إهمالها ويمكن اعتبار أن $n_\circ$ تقريباً لا بأس به لتقدير حجم العينة n أما إذا كان غير ذلك فإنه يمكن حساب n من العلاقة

$$n = \frac{n_\circ}{1 + (n_\circ - 1)/N} \qquad (٣١-٥)$$

مثال ٥-٩

إذا كانت $\hat{P} = 0.4$ ، $d = 0.05$ ومستوى المعنوية $\alpha = 0.05$ وكان حجم العشيرة $N = 3000$. احسب حجم العينة فى حالة المعاينة العشوائية البسيطة.

بما إن $Z = 2$ (تقريباً، جدول ٣ ملحق أ) فإن

$$n_\circ = \frac{(2)^2 (0.4)(0.6)}{(0.05)^2} = 384$$

وحيث إن $n_\circ / N = 0.128$ أى 12.8% وبالتالى لا يمكن إهمالها فإن:

$$n = \frac{n_\circ}{1 + (n_\circ - 1)/N} = \frac{384}{1 + (384 - 1)/3000} = 340.5 \cong 341$$

ثانياً: تقدير حجم العينة العشوائية البسيطة فى حالة البيانات لمتغير مستمر

إذا كان متوسط عينة عشوائية بسيطة هو $\hat{\bar{Y}}$ والمطلوب هو تحقيق المعادلة الاحتمالية التالية

$$P\left(\left|\frac{\hat{\bar{Y}} - \bar{Y}}{\bar{Y}}\right| \geq r\right) = P\left(\left|\frac{N\hat{\bar{Y}} - N\bar{Y}}{N\bar{Y}}\right| \geq r\right) = P(|\hat{\bar{Y}} - \bar{Y}|) \geq r\bar{Y} = \alpha$$

حيث α قيمة احتمالية صغيرة وبافتراض أن $\hat{\bar{Y}}$ تتوزع طبيعياً و r الخطأ النسبى فى تقديرات مجموع أو متوسط العشيرة. وحيث أن:

$$\sigma_{\hat{\bar{Y}}} = \sqrt{\frac{N-n}{N}}\left(\frac{S}{\sqrt{n}}\right)$$

وبالتالى فإن

$$r\overline{Y} = Z\sigma_{\hat{\overline{Y}}} = Z\left(\sqrt{\frac{N-n}{N}}\right)\left(\frac{S}{\sqrt{n}}\right)$$

حيث Z هى قيمة المتغير الطبيعى المعيارى التى تقطع مساحة قدرها $\propto/2$ من كل من الذيلين. ويمكن إيضاح أن:

$$n = \left(\frac{ZS}{r\overline{Y}}\right)^2 \Bigg/ \left[1+\frac{1}{N}\left(\frac{ZS}{r\overline{Y}}\right)^2\right]$$

ويلاحظ أن المعادلة تحتوى على $S/\overline{Y}$ وهو معامل اختلاف العشيرة وتخمينه أسهل من تخمين S نفسها.

وكتقدير مبدئى فإن

$$n_\circ = \left(\frac{ZS}{r\overline{Y}}\right)^2 \qquad (٥-٣٢)$$

وإذا كانت $n_\circ / N$ كبير فإن

$$n = \frac{n_\circ}{1+(n_\circ / N)} \qquad (٥-٣٣)$$

وقد تقدر $n_\circ$ كما يلى:

$$n_\circ = \frac{Z^2 S^2}{d^2} \qquad (٥-٣٤)$$

حيث d هو الخطأ المطلق فى $\hat{\overline{Y}}$ بدلاً من الخطأ النسبى r أى أن $d = r\overline{Y}$.

مثال ٥-١٠

فى عشيرة حجمها 500 إذا كان متوسط متغير ما $\overline{Y} = 22$ والتباين $S^2 = 90$. فما هو حجم العينة العشوائية البسيطة لتقدير $\overline{Y}$ فى حدود خطأ مقدراه 10% وذلك باستثناء فرصة واحدة من عشرين.

باستثناء فرصة واحدة من عشرين تعنى أن مستوى المعنوية $\alpha = 0.05$ وبالتالى $Z \doteq 2$ تقريباً، $d = r\overline{Y} = (0.1)(22) = 2.2$ وبالتالى فإن:

$$n_\circ = \frac{(2)^2 (90)}{(2.2)^2} = 74.3$$

وحيث إن $n_\circ / N = 74.3 / 500$ أى 14.9% وبذلك لا يمكن إهمالها فإن

$$n = \frac{74.3}{1 + 74.3/500} = 64.7 \cong 65$$

وعلى القارئ أن يستنتج بنفسه أنه كلما زاد الخطأ المسموح به كلما قل حجم العينة والعكس صحيح بينما عكس ذلك فى حالة التباين حيث يزداد حجم العينة كلما زاد حجم التباين والعكس صحيح.

ملحوظة هامة:

إذا كان هناك بيانات لأكثر من متغير مستمر فى العشيرة وتم حساب n فى كل متغير فإن n الأكبر هى التى تؤخذ فى الاعتبار عند تحديد حجم العينة.

ومما سبق يتضح ضرورة إيجاد طريقة لتقدير تباين العشيرة حيث إنه أحد العوامل الأساسية فى تقدير حجم العينة وفيما يلى بعض هذه الطرق:

١- تؤخذ العينة على مرحلتين حيث تؤخذ عينة عشوائية بسيطة حجمها n_1 للحصول على $\hat{S}_1^2$ تقديراً للـــ S^2 أو $\hat{P}_1$ تقديراً لـــ P ومن ثم يمكن الحصول على n. وهذه الطريقة أجدر بالثقة ولكن يؤخذ عليها أنها تؤدى إلى طول فترة المعاينة.

٢- باستخدام نتائج مسح إحصائى استكشافى صغير غالباً ما يكون مقتصراً على جزء من العشيرة يسهل تناوله أو أنه يهدف للكشف عن حجم بعض المشاكل.

٣- استخدام نتائج مسوح إحصائية سابقة وهنا ترجع أهمية نشر أية معلومات حول الانحراف المعيارى تم الحصول عليها من مسوح سابقة أو أن تكون فى متناول يد الباحثين حتى يستطيعوا استخدامها عند الحاجة إليها.

٤- بعملية تخمين حول طبيعة العشيرة وذلك باستخدام بعض التوزيعات الرياضية البسيطة لتقدير التباين.

وهناك الكثير من الطرق التى تستخدم فى تقدير حجم العينة لا يتسع المجال لذكرها ويمكن الرجوع إلى أحد المراجع التى تتناول تقنية المعاينة الإحصائية.

ويوجد العديد من طرق المعاينة فبالإضافة إلى المعاينة العشوائية البسيطة التى تم تناولها فيما سبق توجد المعاينة العشوائية الطبقية stratified random sampling

والمعاينة النمطية systematic sampling والمعاينة العنقودية cluster sampling وغيرهم ولكن لن يتسع المجال لتناولهم فى هذا المؤلف.

٥-٥ مصادر الخطأ فى المعاينة

١- عدم التحديد الدقيق للعشيرة المراد معاينتها.

٢- الاختيار غير السليم لطريقة المعاينة.

٣- أخطاء نتيجة عدم الاستجابة non-response وهى تعنى الفشل فى مقياس بعض الوحدات فى العينة المختارة وذلك قد يرجع إلى الإخفاق فى تحديد موقع بعض وحدات العينة أو زيارتها أو أن يكون المنزل مغلقاً وليس به الأشخاص المعنيون أو أن يكون الأشخاص المعنيون بالبحث غير قادرين على الإجابة، فقد تكون المعلومات المطلوب الإجابة عليها غير متوفرة أو من غير الممكن الإدلاء بها، وقد ترجع عدم الاستجابة أيضاً إلى أن هناك أشخاصاً يرفضون إجراء مثل تلك المقابلات أو الإدلاء ببيانات عنهم وعن ممتلكاتهم ...الخ. وفى مثل تلك الأحوال تكون الزيارات المتكررة أحد الحلول الممكنة للتغلب على تلك الأخطاء، كما أن التوعية بضرورة وأهمية إجراء مثل تلك المسوح الإحصائية من الأهمية بمكان، كما تشكل الصياغة الجيدة للسؤال وطريقة تقديمه إمكانية الإجابة عليه بدقة وبدون تردد.

٤- أخطاء القياس نتيجة عدم دقة أداة القياس أو أن الأشخاص المعنيين لا يعطون إجابات دقيقة أو قد يكون بعضهم متحيزاً لإجابات معينة دون غيرها أو قد يكون هناك ارتباط بين أخطاء القياس بين وحدات العينة فى اتجاه ما قد يكون موجباً أو سالباً. وهنا يلزم إيجاد طريقة ما لتحديد القيمة الصحيحة وقد يكون إعادة القياس بطريقة مستقلة أكثر دقة وقد يكون تدريب القائمين بالحصر تدريباً دقيقاً وأن يكونوا من ذوى الخبرة حتى يمكنهم الحصول على البيانات المطلوبة بالدقة الكافية. وقد يكون إعادة قياس عينات جزئية أحد الحلول لتفادى أخطاء القياس. وقد تكون سرية البيانات المتحصل عليها ضرورة وخاصة عندما تكون الأسئلة عن واقع شخص كالسرقة مثلاً أو استخدام بعض العقاقير الطبية.

٥- أخطاء نتيجة عدم المعالجة الإحصائية السليمة للبيانات المتحصل عليها.

تمارين الباب الخامس

٥-١ فى عشيرة من 5 وحدات كانت قيمها 2 ,3 ,5 ,6 ,1 وسحبت عينة عشوائية بسيطة من وحدتين بدون إحلال، فما هو عدد العينات العشوائية البسيطة الممكن سحبها وما هو احتمال سحب كل منها ؟ احسب المتوسط والتباين وتحقق من أن المتوسط والتباين المحسوب من العينة هما تقديران غير متحيزين لمتوسط وتباين العشيرة.

٥-٢ إذا كان متوسط عدد كرمات العنب فى الأراضي المستصلحة حديثاً 900 فى الفدان. وكان متوسط إنتاج الكرمة الواحدة 13 كج بتباين مقداره 30 كج مربع. احسب كمية الإنتاج فى مساحة قدرها 100 فدان (فدان = $4200\ m^2$) والتى هى حجم العشيرة وانحرافه المعيارى إذا علمت أن المعاينة كانت عشوائية بسيطة وكان حجم العينة خمسة أفدنة. قدر حدود الثقة لكمية الإنتاج الكلية أيضاً بمستوى ثقة 95% .

٥-٣ سحبت عينة عشوائية بسيطة من 5 من العاملين في أحد المصانع وكانت دخولهم كما يلى: 350 ,400 ,200 ,330 ,500 جنيهاً مصرياً فى الشهر. فما هو متوسط دخل العامل فى المصنع إذا كان به 1000 عامل ثم احسب تباينه. ثم قدر مجموع الدخل المتحصل عليه لجميع العاملين فى المصنع فى الشهر الواحد وكذلك انحرافه المعيارى.

٥-٤ فى دراسة لاستهلاك الأصناف المختلفة للمنظفات الصناعية للملابس فى مدينة القاهرة سجلت المعلومات من السيارات المنتظرة فى إشارات المرور. حلل هذه الطريقة من المعاينة.

٥-٥ إذا كانت $\hat{P} = 0.25$ ، $d = 0.02$ ومستوى المعنوية 5% وحجم العشيرة هو 2000. احسب حجم العينة العشوائية البسيطة.

٥-٦ أعد حساب حجم العينة فى مثال ٥-١٠ فى حدود خطأ مقداره 5%، 15%، 20% وذلك بمستوى معنوية 5% . ماذا تستنتج؟

٥-٧ الجدول التالى يمثل الدخل الشهرى بالجنيه لـــ 10 أسر مصرية أخذت بطريقة المعاينة العشوائية البسيطة من بين 200 أسرة وأيضاً تم رصد عدد أفراد الأسرة المدخنين وغير المدخنين:

الأسـرة	عدد الأفراد غير المدخنين	عدد الأفراد المدخنين	المنفق على الدروس الخصوصية	الدخــل
١	3	0	260	1300
٢	2	1	100	500
٣	5	1	400	2100
٤	2	0	250	1200
٥	3	0	180	900
٦	3	2	180	850
٧	4	1	270	1200
٨	3	2	200	900
٩	5	1	800	3000
١٠	2	3	500	2400

احسب النسبة المئوية للدخل المنفق على الدروس الخصوصية و خطأه المعيارى.

٥-٨ فى التمرين السابق المطلوب مقارنة علاقة النسبة بعلاقة التوزيع الثنائى وفقاً لحالة التدخين من عدمه. واحسب التباين فى كل.

٦ اختبارات الفروض Tests of hypotheses

١- التوزيع العينى للمتوسطات

٢- حدود الثقة

٣- اختبارات الفروض الإحصائية

٤- تحديد العدد الأمثل للمشاهدات فى التجربة (حجم العينة)

هناك نوعان من التقديرات:

الأول: تقديرات محددة بنقطة point estimation وهى التى يقدر فيها معلم parameter العشيرة بقيمة واحدة مثال ذلك تقدير متوسط العشيرة μ من العينة بواسطة $\overline{Y}$ أو تباين العشيرة σ^2 بواسطة S^2 ... وهكذا.

الثانى: تقديرات الفترة interval estimation وهى تقدير للمدى الذى يمكن أن يتراوح فيه تقدير المعلم نفسه باحتمال معين.

وسوف يتم فى هذا الباب تغطية موضعين رئيسيين:

١- بعد الحصول على التقديرات المعينة إلى أى مدى تكون الثقة فى هذه التقديرات ؟

٢- إلى أى مدى يمكن إرجاع النتائج المتحصل عليها إلى الصدفة المحضة، وإلى أى مدى يمكن اعتبار النتائج راجعة لفارق أو فوارق حقيقية بافتراض فرضاً معيناً ؟

وكلا المفهومين مهمان فى مجال الاستنباطات الإحصائية. وسوف يتم تغطية هذين الموضعين فيما يتعلق بالمتوسطات والتباين.

٦-١ التوزيع العينى للمتوسطات Sampling distribution of means

إذا كان هناك متغير ما، وليكن Y، ومهما كان توزيعه، أى أنه ليس بالضرورة أن يكون طبيعياً، ولكن له متوسط μ وتباين σ^2 وسحبت منه عينات عشوائية كبيرة الحجم كل منها n وحسب المتوسط لكل منها فإن هذه المتوسطات تتوزع طبيعياً (مهما كان توزيع المتغير الأصلى) بمتوسط قدره μ وتباين σ^2/n حيث n هى عدد الأفراد أو المشاهدات التى حسب على أساسها المتوسطات. والجذر التربيعى لتباين المتوسط يطلق عليه الانحراف المعيارى للمتوسط standard deviation of the mean أو الخطأ القياسى standard error.

مثال ٦-١

كانت أوزان سبع عجلات جاموسى عند عمر ثلاثة أشهر بالكيلوجرام هى:

112, 105, 91, 95, 96, 108, 90

فإن تقدير المتوسط: $\overline{Y} = 99.6$ kg، وتقدير التباين: $S_Y^2 = 75.62$ kg^2، وتقدير تباين المتوسط: $S_{\overline{Y}}^2 = 75.62/7 = 10.8\,\text{kg}^2$.

أى أنه إذا أخذ العديد من العينات حجم كل منها 7 حيوانات وحسب متوسط المتوسطات والتباين بينها فإنه من المتوقع أن يكونا 10.8,99.6 على التوالى، ويبدو من العلاقة (٣-١٣)، أنه كلما زاد حجم العينة المقدر منها المتوسط كلما قل تباين المتوسط وهذا منطقى، حيث إنه فى غالب الأحيان لا يعرف بالضبط قيمة σ_Y^2 ولكن يكون هناك تقدير له هو S_Y^2 وبالتالى يكون تقدير تباين المتوسط $S_{\bar{Y}}^2 = \sigma_Y^2 / n$.

وهنا تجدر الإشارة إلى نقطة مهمة عادة ما تكون غير واضحة لكثير من الباحثين وهى أنه لا علاقة بين قيمة σ_Y^2 المقدرة من عينة وبين حجم العينة ذاتها، أى أنه ليس بالضرورة أن تعطى العينة الكبيرة تقديراً أصغر σ_Y^2 طالما أن كل العينات مأخوذة عشوائياً من نفس العشيرة ولكن الذى له علاقة وثيقة بحجم العينة هو مدى الوثوق بالتباين المقدر هذا حيث يزيد الوثوق به كلما زاد حجم العينة، وكذلك تباين المتوسط حيث يقل كلما زاد حجم العينة. وسوف يتم التركيز فى معالجة المتوسط فى هذا الباب على التوزيع الطبيعى.

٦-٢ حدود الثقة Confidence limits

تركزت المناقشات فى معظم الأبواب السابقة على طريقة حساب تقديرات estimation المعالم parameters فى العشيرة وأن تكون هذه التقديرات غير متحيزة وهذه يطلق عليها تقديرات محددة بنقطة point estimates. وفى هذا الباب سوف يتم مناقشة مدى الوثوق فى هذا التقدير المحسوب أو بمعنى آخر مدى احتمال أن يكون التقدير ممثلاً تمثيلاً حقيقياً للعشيرة، فنادراً ما يعرف معالم العشيرة ولكن تقدر بهذه التقديرات.

٦-٢-١ حدا الثقة للمتوسط فى حالة معرفة التباين الحقيقى

افترض جدلاً أن هناك عشيرة حيث μ، σ_Y (المتوسط والانحراف المعيارى الحقيقيان) معلومان، وبفرض أن القيم تتوزع توزيعا طبيعيا. وأخذ من هذه العشيرة العديد من العينات كل منها حجمه n مشاهدة، فسيكون التباين بين هذه المتوسطات σ^2 / n، ويمكن القول بأن المدى من $\mu - (1.96)\frac{\sigma}{\sqrt{n}}$ إلى $\mu + (1.96)\frac{\sigma}{\sqrt{n}}$ يتوقع أن يحتوى على 95% من متوسطات العينات المأخوذة، أى أنه إذا أخذ 100 عينة مثلاً يتوقع أن يقع متوسط 95 منها فى هذا المدى. كما يتوقع أن المدى من $\mu - (2.58)\frac{\sigma}{\sqrt{n}}$ إلى $\mu + (2.58)\frac{\sigma}{\sqrt{n}}$ يحتوى على 99% من العينات. (القيمتان 1.96 و 2.58 مأخوذتان من جدول ٣ ملحق أ). ويمكن التعبير عن ذلك رياضياً كما يلى:

$$\Pr\left(-1.96 \le \frac{\overline{Y}-\mu}{\sigma/\sqrt{n}} \le +1.96\right) = 0.95$$

وبضرب الطرفين فى $\sigma/\sqrt{n}$

$$\Pr\left(-1.96\sigma/\sqrt{n} \le (\overline{Y}-\mu) \le +1.96\sigma/\sqrt{n}\right) = 0.95$$

ومنها:

$$\Pr\left(-1.96\sigma/\sqrt{n} \le (\mu - \overline{Y}) \le +1.96\sigma/\sqrt{n}\right) = 0.95$$

وبإضافة $\overline{Y}$ لكل الأطراف ينتج:

$$\Pr\left(\overline{Y}-1.96\sigma/\sqrt{n} \le \mu \le \overline{Y}+1.96\sigma/\sqrt{n}\right) = 0.95 \qquad (٦-١)$$

وبالمثل فإن:

$$\Pr\left(\overline{Y}-2.58\sigma/\sqrt{n} \le \mu \le \overline{Y}+2.58\sigma/\sqrt{n}\right) = 0.99 \qquad (٦-٢)$$

وتسمى (٦-١) حدى الثقة للمتوسط لـ 5% بينما (٦-٢) حدى الثقة لـ 1% ويمكن تعميم التعبيرين السابقين كما يلى:

$$\Pr\left(\overline{Y}-(Z_{\alpha/2})(\sigma)/\sqrt{n} \le \mu \le \overline{Y}+(Z_{\alpha/2})(\sigma)/\sqrt{n}\right) = 1-\alpha \qquad (٦-٣)$$

وكلما ضاق الفرق بين حدى الثقة كلما زادت الثقة فى المتوسط المحسوب. فمثلاً إذا كان حدا الثقة 5% لمتوسط ما هما 32&34 kg معنى هذا أن احتمال أن يحتوى المدى بين 32&34 kg متوسط العشيرة μ هو 95%، بينما إذا كان لمتوسط آخر حدا الثقة 5% هما 38&28 kg فإن احتمال أن يحوى هذان الحدان المتباعدان نسبياً المتوسط الحقيقى هو 95%، وبالتالى فإن الثقة تكون أكبر فى المتوسط الأول. وإذا فرض أن حدى الثقة لمتوسط ثالث هما 33.5&32.5 kg يكون هو أدقها جميعاً إلى أن يأخذ حدا الثقة القيمة النظرية $33+0$ إلى $33-0$ أى ينطبق التقدير $\overline{Y}$ على قيمة μ للعشيرة نفسها وهذا مطلق الثقة فى التقدير. وهناك خطأ شائع يجب العمل على تجنبه وهو أنه كثيراً ما يقال إن احتمال وقوع المتوسط الحقيقى بين حدى الثقة هو مثلاً 95%، والصحيح أن يقال إن احتمال أن يحتوى حدا الثقة المتوسط الحقيقى هو 95% وذلك لأن المتوسط الحقيقى ثابت لا يتغير ومعنى القول إن هناك احتمال 95% فى أن يقع المتوسط بين الحدين أن هناك احتمال 5% أن يقع خارجها وذلك غير ممكن لثبوت قيمته كما سبق القول.

مثال ٦-٢

إذا فرض أن σ_Y^2 لوزن الفطام فى سلالة أغنام $25\ kg^2$ وقدر المتوسط من عينة حجمها 12 حملا بمقدار $18\ kg$.

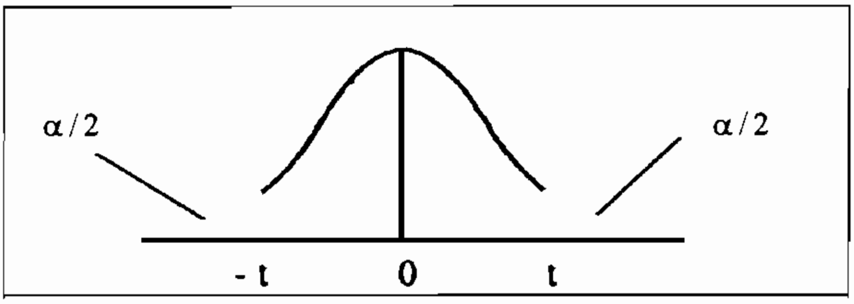
يكون حدا ثقة 5% لهذا المتوسط هما: $18 \pm \left((1.96)(\frac{5}{\sqrt{12}})\right) = 18 \pm 2.83$، أى أن الحد الأدنى هو $15.17\ kg$ والحد الأعلى هو $20.83\ kg$

بينما يكون حدا ثقة 1% هما $18 \pm \left((2.58)(\frac{5}{\sqrt{12}})\right) = 18 + 3.72$، أى أن الحد الأدنى هو $14.28\ kg$ والحد الأعلى هو $21.72\ kg$

٦-٢-٢ حدا الثقة للمتوسط فى حالة عدم معرفة تباين العشيرة

فى الفصل السابق افترض معرفة قيمة تباين العشيرة σ^2 وهذا فرض يسمح باستخدام جداول التوزيع الطبيعى القياسى (جدول ٣ ملحق أ). ولكن نادراً ما يتحقق مثل هذا الشرط وإن كان من الممكن استخدام جداول التوزيع الطبيعى على أى الأحوال إذا كان حجم العينة كبيراً (أكبر من 100 مثلا). فإذا كانت $\overline{Y}$ تتوزع طبيعياً فإن القيمة $\overline{Y} - \mu$ تتوزع طبيعياً، وبقسمتها على ثابت $\sigma_{\overline{Y}}^2$ أى $(\overline{Y} - \mu)/\sigma_{\overline{Y}}^2$ فإن هذه القيمة تتوزع أيضاً طبقاً للتوزيع الطبيعى. فإذا لم تكن القيمة الحقيقة للتباين σ_Y^2 معلومة يمكن استخدام القيمة المقدرة لها بدلاً منها أى S_Y^2، وفى هذه الحالة لا تصبح القيمة $(\overline{Y} - \mu)/S_{\overline{Y}}$ موزعة حسب التوزيع الطبيعى ولكنها تتوزع حسب توزيع "t" (جدول ٤ ملحق أ) والمعروف باسم "student distribution" الذى وضعه "Gosset" والذى كان يوقع بحوثه باسم "Student" ويشبه توزيع "t" التوزيع الطبيعى فى أنهما متناظران حول المتوسط ولكن يختلفان عن بعضهما فى أن المساحة (وبالتالى الاحتمال) المحددة بين قيم "t" تختلف باختلاف حجم العينة معبراً عنها بدرجات الحرية degrees of freedom (df). وفى الحالات البسيطة تكون درجات الحرية مساوية لـ $(n-1)$. ويتطابق التوزيعان عندما تكون n كبيرة جداً. ومن الوجهة العملية يمكن استخدام التوزيع الطبيعى إذا ما زادت n عن 100 كما سبق القول. وجدول ٤ ملحق أ يبين الاحتمال معبراً عنه بنسبة المساحة تحت المنحنى فى كلا الذيلين الخارجة عن نقطة t معينة. وهكذا تكون القيمة الحرجة critical value لـ t عند احتمال 5% (ويطلق على هذا الاحتمال الرمز α) ودرجات حرية 7 مثلاً 2.365 (جدول ٤ ملحق أ) وتكتب هكذا $t_{(.05,7)} = 2.365$، وحيث إن هذا الجدول يمثل الذيلين معاً فإنه يعنى مجموع المساحتين الخارجتين عن القيمتين ± 2.365 هو

5%، وحيث إن المنحنى متناظر الشطرين فتكون المساحة ما بعد 2.365+ هى 2.5% أى $\alpha/2$ والمساحة ما قبل 2.365− هى 2.5%. وعلى القارئ أن يتحقق من إن $t_{(.05,10)} = 2.228$، $t_{(.01,15)} = 2.947$، $t_{(.05,\infty)} = 1.96$. ويلاحظ أن القيمة الأخيرة 1.96 هى نفس القيمة للمنحنى الطبيعى عند احتمال 0.025 على كل من طرفى التوزيع الطبيعى. والشكل ٦-١ يوضح توزيع t



شكل ٦-١ توزيع t عند درجات حرية df ومستوي معنوية α

ولحساب حدى الثقة فى حالة عدم معرفة σ^2 ولكن المتاح هو تقدير له وهو S^2، فإنه يمكن إعادة كتابة التعبير (٦-٣) ولكن مع استبدال t مكان Z، S مكان σ أى:

$$\Pr\left(\overline{Y} - t_{\alpha,(n-1)}S_{\overline{Y}} \le \mu \le \overline{Y} + t_{\alpha,(n-1)}S_{\overline{Y}}\right) = 1-\alpha \qquad (٦-٤)$$

ويلاحظ القارئ هنا أن القيمة t عند α وليس عند قيمة $\alpha/2$ كما هو فى حالة التوزيع الطبيعى (٦-٣) وذلك لأن جدول التوزيع الطبيعى يحسب المساحات فى جانب واحد فقط من المتوسط بينما t يحسبها على كلا الجانبين كما سبق القول.

مثال ٦-٣

البيانات التالية تمثل وزن الميلاد بالكيلوجرام لعدد 12 عجل جاموسى والمطلوب هو حساب حدى الثقة للمتوسط عند 5% ، 1%

25, 28, 31, 30, 28, 22, 26, 30, 22, 28, 31, 32

المتوسط $\overline{Y} = 27.75\text{kg}$ ، التباين $S^2_{\overline{Y}} = 11.48\ \text{kg}^2$ ، $n = 12$ ، $df = 11$
وبالكشف فى جدول t عند درجات حرية 11 فإن $t_{.05,\,11} = 2.201$، $t_{.01,\,11} = 3.106$، وبالتالى يكون حدا الثقة عند 5% هما:

$$27.75 \pm 2.2\sqrt{\frac{11.48}{12}} = 27.75 \pm 2.15$$ ، أى أن الحدين هما $29.9, 25.6$ kg

ويكون الحدان عند 1% هما:

$$27.75 \pm 3.106\sqrt{\frac{11.48}{12}} = 27.75 \pm 3.04$$ ، أى أن الحدين هما $30.79,\ 24.71$ kg

٦-٢-٣ حدا الثقة للتباين:

تتوزع النسبة $(n-1)S^2/\sigma^2$ طبقا لتوزيع مربع كاى χ^2. ويمكن الرجوع إلى العديد من المراجع لتفهم ذلك رياضيا (مثلاً Socal and Ralf 1973). وبطريقة مشابهة لتلك التى اتبعت عند استنباط حدى الثقة للمتوسط يمكن كتابة التعبير التالى:

$$\Pr\left[\chi^2_{(1-\alpha/2),(n-1)} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{(\alpha/2),(n-1)}\right] = 1-\alpha$$

ويلاحظ أن قيم χ^2 ليست متماثلة على الجانبين ذلك لأن توزيع مربع كاى نفسه ليس متناظر الشطرين ومن التعبير السابق يمكن كتابة:

$$\Pr\left[\frac{(n-1)S^2}{\chi^2_{(\alpha/2),(n-1)}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{(1-\alpha/2),(n-1)}}\right] = 1-\alpha$$

وحيث إن $(n-1)S^2$ = مجموع مربعات الانحرافات المصححة $\sum y^2$ فإنه يمكن كتابة التعبير السابق كما يلى:

$$\Pr\left[\frac{\sum y^2}{\chi^2_{(\alpha/2),(n-1)}} \le \sigma^2 \le \frac{\sum y^2}{\chi^2_{(1-\alpha/2),(n-1)}}\right] = 1-\alpha \qquad (٦-٥)$$

مثال ٦-٤

حساب حدى الثقة للتباين فى مثال ٦-٣ عند كل من المستويين $1\%, 5\%$

فى هذا المثال $S^2_Y = 11.48$ بدرجات حرية 11 أى أن $\sum y^2 = 126.28$، وباستخراج قيم χ^2 المطلوبة من جدول ٦ ملحق أ، فيكون:

الحد الأدنى للثقة للتباين عند 5% هو:

$$126.28 \div \chi^2_{(.025),(11)} = 126.28 \div 21.92 = 5.76\ \text{kg}^2$$

والحد الأعلى للثقة للتباين عند 5% هو:

$$126.28 \div \chi^2_{(.975),(11)} = 126.28 \div 3.82 = 33.06\ \text{kg}^2$$

والحد الأدنى للثقة عند مستوى 1% هو:

$$126.28 \div \chi^2_{(.005),(11)} = 126.28 \div 26.76 = 4.72\,\text{kg}^2$$

والحد الأعلى للثقة عند مستوى 1% هو:

$$126.28 \div \chi^2_{(.995),(11)} = 126.28 \div 2.60 = 48.57\ \text{kg}^2$$

صندوق ٦-١

- التباين للعشيرة σ^2_Y: مجموع مربعات الانحرافات عن المتوسط مقسوماً على حجم العشيرة
- التباين من العينة S^2_Y: مجموع مربعات الانحرافات عن متوسط العشيرة مقسوماً على حجم العينة ناقص واحد
- σ_Y: الانحراف القياسي للعشيرة
- S_Y: الانحراف القياسي للعينة
- $\sigma_{\overline{Y}}$: الخطأ القياسي إذا كان تباين العشيرة معروف
- $S_{\overline{Y}}$: الخطأ القياسي إذا كان تباين العشيرة غير معروف ويقدر من التباين المحسوب من العينة

٦-٣ اختبارات الفروض الإحصائية Tests of hypotheses

من أهم وظائف علم الإحصاء الحيوى أنه بواسطته يمكن اختبار فرض ما اختباراً موضوعياً مع وضع العبارات الاحتمالية له. فالأسلوب الموضوعى المثالى لاتخاذ قرار ما هو أن تضع فرضا معينا يمكن اختباره ومعه بدائله وعندما يختبر الفرض إما أن يرفض أو لا يرفض وفى حالة رفضه قد يقبل الفرض البديل.

افترض أن عشيرة ما احتمال الذكور بها = احتمال الإناث = 0.5 وجيء بعينة من 10 حيوانات فكان منها 8 من جنس، و 2 من الجنس الآخر. والسؤال: هل هذه العينة تتبع فعلاً هذه العشيرة ؟، أو هل هذه العينة مأخوذة عشوائياً من العشيرة وإن أى اختلاف فى النسبة بينها وبين العشيرة هى اختلافات راجعة للصدفة المحضة ؟.

إذا أخذ مفكوك "ذو الحدين" لهذه الحالة $(p+q)^{10}$ حيث p احتمال الحصول على ذكر q = احتمال الحصول على أنثى = 0.5، يكون كما يلى:

$$p^{10}+10p^9q+45p^8q^2+...+45p^2q^8+10pq^9+q^{10}$$

ويبين الجدول التالى احتمالات حدوث الأعداد المختلفة من الذكور والإناث

عدد ذكور	عدد إناث	الاحتمال تحت فرض $p = q = 0.5$
0	10	0.000977
1	9	0.009766
2	8	0.043945
3	7	0.117188
4	6	0.205078
5	5	0.246094
6	4	0.205078
7	3	0.117188
8	2	0.043945
9	1	0.009766
10	0	0.000977
المجموع		**1.000**

يتضح من هذا الجدول أن هناك فعلاً احتمال أن تكون هذه العينة ممثلة للعشيرة حيث إنه بالصدفة المحضة يمكن الحصول على العينة بمثل هذا الانحراف أو أكبر فى طرفى التوزيع باحتمال قدره $0.109 = 0.000977 + 0.009766 + (0.043945)(2)$ أى حوالى 0.11 وهو احتمال ضئيل نسبياً. ويعنى هذا أنه إذا اتخذ قراراً بأن هذه العينة شاذة بالنسبة للعشيرة، ولذا فهى ليست تابعة لها، نكون قد وقعنا فى خطأ قدره 11% وهو احتمال أن تكون العينة تابعة للعشيرة ولكن الصدفة المحضة هى التى عملت على الحصول على نسبة $(8:2)$. أيضاً إذا كان القرار أن هذه العينة تتبع العشيرة وأن الانحرافات بينها وبين العشيرة راجعة للصدفة قد يكون هذا قراراً يحتمل الخطأ والصواب طبقاً لحقيقة الأمر.

ويمكن النظر للموضوع بطريقة أخرى. عشيرة ما يراد اختبار ما إذا كانت $p = q = 0.5$ فأخذت عينة عشوائية سابقة الوصف فما هو القرار ؟

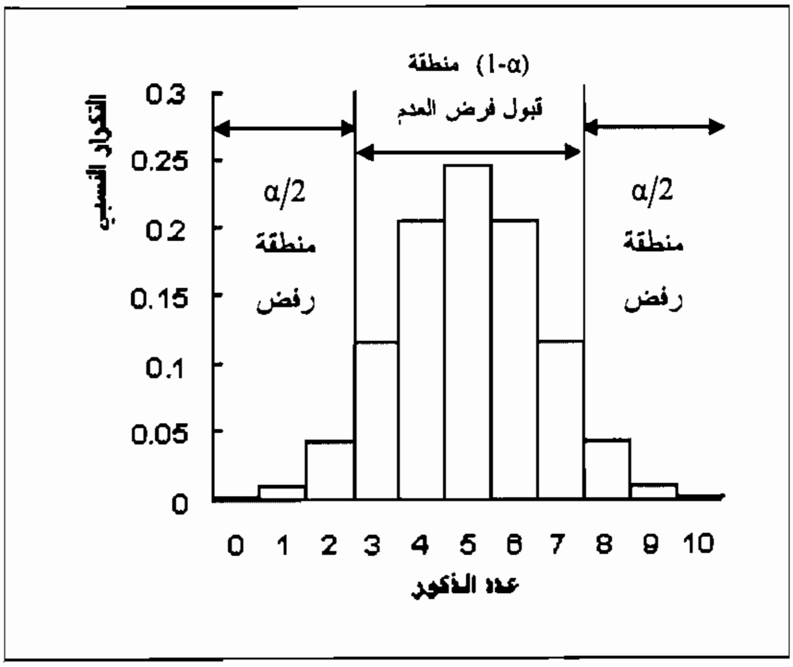
لهذا يوضع فرض يطلق عليه فرض العدم لأسباب ستتجلى فيما بعد (ويرمز له بالرمز H_o) ويمكن كتابته: $H_0 : p = q = 0.5$

وعادة ما يحدد الباحث النسبة التى يسمح بها لقراره أن يكون فى خطأ من هذا النوع، فإذا وضع الباحث لنفسه خطأ مسموحاً قدره 0.11، مثلاً فإنه سيقرر أن الفرض خاطئ فقط فى الحالات $(0-10)$، $(1-9)$، $(2-8)$، $(8-2)$، $(9-1)$، $(10-0)$. (شكل ٦-٢).

أما إذا أراد الباحث أن يكون أكثر تحوطاً بالنسبة لهذا الخطأ ولا يود أن يقرر أن النسبة $p : q = 1:1$ ، إلا فى الحالات $(10-0)$، $(0-10)$ فإنه فى هذه الحالة يكون قد وضع خطأ لنفسه قدره $0.001954 = (0.000977)(2)$، ولكن بتقليله لهذا النوع من الخطأ يكون قد وقع فى خطأ آخر وهو أنه من الممكن أن يقرر عدم رفض فرض العدم بينما هو خاطئ وتسمى القيم التى على أساسها يرفض أو يقبل الباحث فرض العدم بالقيم الحرجة وهى تحدد تحت المنحنى أو الهيستوجرام مساحات تسمى بمناطق الرفض أو المناطق الحرجة.

ويبين ٦-٢ المنطقة الحرجة أو منطقة الرفض إذا ما اتخذ الباحث قيمة هذا الخطأ بقدر 0.11 أى كأنه يقول فقط إذا كانت العينة بدرجة انحراف قدره $(2-8)$ أو أكثر (أى $1-9$ أو $0-10$) فإن فرض العدم سوف يرفض. أما إذا كان الانحراف أقل من هذا أى ($3-7$، $4-6$ أو $5-5$) فإن فرض العدم لا يرفض، ويسمى هذا النوع من الخطأ "خطأ من النوع الأول type I error" ويعرف بأنه احتمال رفض فرض العدم وهو صحيح. كما جرت العادة علي تسميته مستوى المعنوية significance level ويرمز له أيضاً بالرمز α. وغالباً ما يستخدم فى المجالات البيولوجية مستويان

للمعنوية هما 0.05 ، 0.01، ويطلق على الفروق التى تزيد قيمتها عن المستوى الأول (0.05) بأنها معنوية بينما تلك التى تزيد عن قيم المستوى الثانى (0.01) يطلق عليها معنوية جداً، وإن كانت تستخدم مستويات أخرى طبقاً لمقتضيات الدراسة. ونظراً للتوسع فى استخدام الحزم الجاهزة من البرامج الإحصائية مع أجهزة الحاسب فإنه يكتفى بكتابة كلمة معنوى يتلوها مباشرة مستوى المعنوية بين قوسين.

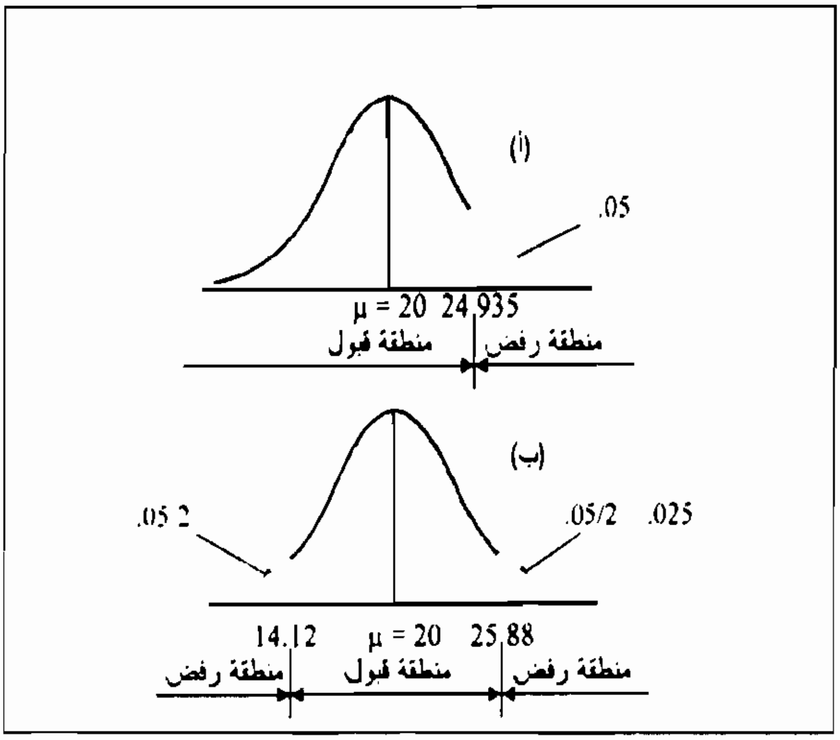


شكل ٦-٢ مناطق الرفض (الحرجة) والقبول فى حالة $\alpha = 0.11$

مثال ٦-٥

افترض وجود متغير يتبع توزيعاً متصلاً (وليس متقطعاً كما فى المثال السابق) فى عينة من 4 حملان قدر منها المتوسط بمقدار 25 kg وكان تباين العشيرة الحقيقى $\sigma^2_Y = 36\,kg^2$. وكان فرض العدم أن $H_0 : 20$ kg.

تباين المتوسط $\sigma^2_{\bar{Y}} = 36/4 = 9\ kg^2$، وبالتالى $\sigma_{\bar{Y}} = 3$ kg. وإذا فرض أن توزيع الأوزان يتبع توزيعاً طبيعياً فإنه يمكن تمثيل الحالة بالشكل ٦-٣ (أ) للاختبار من طرف واحد و ٦-٣ (ب) للاختبار من طرفين كالتالى:



شكل ٦-٣ المناطق الحرجة على التوزيع الطبيعى ($\alpha = 0.05$)

(أ) اختبار من طرف واحد

(ب)اختبار من طرفين

احتمال الحصول على متوسط 25 kg أو أعلى من عشيرة متوسطها 20 kg يمكن استخراجها من جدول التوزيع الطبيعى (جدول ٣ ملحق أ) حيث $Z = (25 - 20)/3 = 1.67$ والمساحة التى تحدها القيمة وما بعدها هى 0.0475، أى أنه إذا كان فعلاً المتوسط يتبع هذه العشيرة فإن احتمال ذلك هو 0.0475، وبعبارة أخرى فإن هناك هذا القدر من الاحتمال ليكون المتوسط لعينة من نفس العشيرة وعليه إذا قرر الباحث رفض فرض العدم القائل بأن متوسط العشيرة 20 kg أو $H_0 : \mu_0 = 20$ أو $H_0 : \mu_0 = \mu$ ومنها $H_0 : \mu_0 - \mu = 0$ (من هنا جاء لفظ العدم null) فيكون قد ارتكب خطأ من النوع الأول $\alpha = 0.0475$.

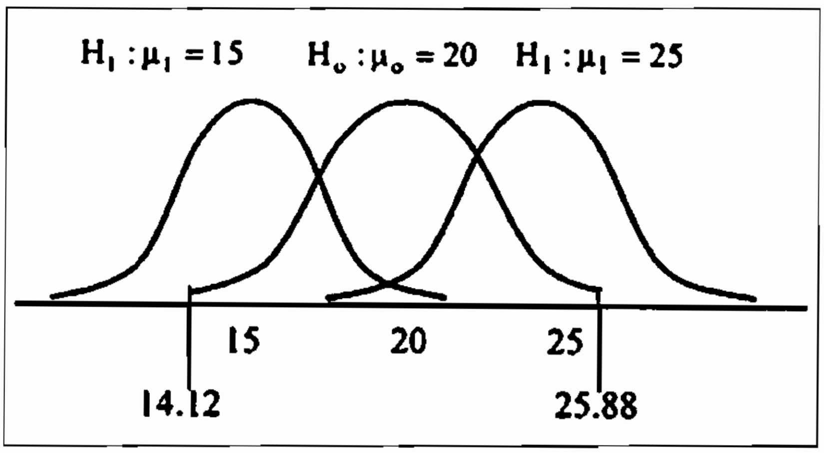
وإذا كان الباحث قد حدد لنفسه خطأ من النوع الأول بمقدار 5% على جانب واحد من المتوسط أى أنها 5% من المساحة الكلية، فيكون حد المنطقة الحرجة هو $(1.645)\sigma_{\bar{Y}} = (1.645)(3) = 4.935$ kg أكبر من المتوسط أى:

$20 + 4.935 = 24.935$ kg

أما إذا حدد مساحة 5% على جانبى المتوسط فيكون الحد الأدنى $20 - (3)(1.96) = 14.12$ kg والحد الأقصى $20 + (3)(1.96) = 25.88$ kg، أى أنه سيقرر رفض فرض العدم إذا قل المتوسط عن 14.12 kg أو زاد عن 25.88 kg وما بين هاتين القيمتين فإنه لن يستطيع رفض فرض العدم أن $\mu = \mu_0$.

والذى يحدد ما إذا كان اختبار المعنوية من جهة واحدة one–tail أو من جهتين two–tail هو مدى توافر المعلومات المسبقة لدى الباحث، فإذا كان لدى الباحث من المعلومات المسبقة ما يجعله يعتقد أنه إن لم يكن الفرق $\mu - \mu_0 = 0$، فإنه يتوقع أن يكون أكبر من صفر فهذا يؤدى إلى اختبار من جهة واحدة ولكن فى غياب مثل هذا الاعتقاد (أى قد يكون μ أكبر من μ_0 أو μ_0 أكبر من μ) فإن هذا يؤدى إلى اختبار من الجهتين.

بديهى أنه إذا رفض فرض العدم H_0 أن يكون هناك فرض بديل alternative hypothesis يمكن قبوله. مثلاً فى مثال ٦–٥ فرض العدم H_0 كان $\mu_0 = 20$ وإذا رفض الباحث هذا الفرض فإنه سيكون على استعداد لقبول فرض بديل H_1 أن المتوسط الحقيقى يبعد 5 kg عن هذا المتوسط أى إما 15 kg أو 25 kg شكل ٦–٤، وإذا حدد الباحث أيضاً المنطقة الحرجة α بالقيمة 5% كما هو فى الشكل ٦–٣ وتحدها القيمتان 14.12، 25.88 وكما هو موضح أيضاً بشكل ٦–٤.



شكل ٦–٤ فرض العدم والفروض البديلة تحت $\alpha = .05$ (.025 على كل جانب)

تمثل المساحة المظللة على المنحنى قيمة $\alpha = 5\%$ من مجموع المساحة الكلية (2.5% على كل من جانبى المنحنى) وبالتالى فإن الباحث إذا لم يفرض فرض العدم لقيم بين 20، 25.88، قد يقع فى خطأ وهو أن هناك احتمال أن هذه القيم تتبع المنحنى $H_1 : \mu = 25$ ، ويمكن صياغة السؤال بطريقة أخرى إذا فرض أن $\mu = 25$ ما احتمال الحصول على قيمة 25.88 أو أقل (وهى القيمة التى عندها لا يرفض الفرض $H_0 : \mu = 20$ ؟ هذا الاحتمال عبارة عن:

$$= 0.5 + \Pr\left(Z = \frac{25.88 - 25}{\sigma_{\bar{y}}}\right) = 0.5 + \Pr\left(Z = \frac{25.88 - 25}{3}\right)$$

$$= 0.5 + \Pr(Z = 0.29) = 0.5 + 0.114 = 0.614$$

وذلك باستخدام جداول المساحات تحت التوزيع الطبيعى. معني ذلك أنه إذا قبل الفرض $H_0 : \mu = 20$ ، فإنه سوف يكون هناك خطأ قدره 0.614 أن تنتمى العينة للعشيرة التى متوسطها 25 kg. ويسمى هذا النوع من الخطأ (خطأ من النوع الثانى type II error) ويرمز له بالرمز β وهو أيضا يحدده الباحث بالقيمة التى يريدها وذلك عند إجراء التجربة واختيار العينة. وعليه يمكن تمثيل احتمال خطأ النوع الأول α من عدمه واحتمال خطأ النوع الثانى β من عدمه بالشكل التالى:

		فرض العدم H_0	
		صحيح	**خاطئ**
القرار	**يرفض**	خطأ النوع الأول Type I Error (α)	صحيح$(1 - \beta)$
	لا يرفض	صحيح $(1 - \alpha)$	خطأ النوع الثانى Type II Error (β)

وتمثل القيمة $(1 - \beta)$ احتمال أن يرفض الباحث فرض العدم H_0 عندما يكون خاطئاً. وهذا التعبير يطلق عليه قوة الاختبار power of the test، وكلما كان الاختبار قوياً كلما تمكن الباحث من رفض فرض العدم عندما يكون غير صحيح وفى حالة العكس أن الاختبار الضعيف يكلف الباحث الكثير للبحث عن الفروق أو الاختلافات التى قد تكون فعلا موجودة ولكن لضعف الاختبار فإنه لا يتمكن من رفض فرض

العدم والإعلان عن معنوية هذه الفروق ويكون فى هذا إهدار للإمكانيات البحثية. وفى المثال السابق حسبت قيمة β على منحنى$H_1: \mu = 25$ وحيث أن كلا من المنحنيين $\mu = 15$، $\mu = 25$ متماثلين فإن β تحت المنحنى $H_1: \mu = 15$ هى نفسها 0.61 .

فرض فى المعالجة السابقة اختبار من جهتين two–tail ولكن إذا كان لدى الباحث ما يجعله يعتقد أنه إن لم يكن $\mu = 20$ فإنه سوف يكون 25 (أى أنه يعتقد أن المعاملة سوف تؤدى إلى زيادة المتوسط) ففى هذه الحالة يصبح الاختبار ممثلا لجانب واحد وتكون α كلها خاصة بالاختبار على جانب واحد من المتوسط ويكون حدها $(1.645)(\sigma_{\bar{Y}})$ من المتوسط الفرضى أى $20 + (1.645)(3) = 24.935$ kg ويكون احتمال الحصول على مثل هذه القيمة أو أقل على منحنى $H_1: \mu = 25$ هو:

$$= 0.5 - \Pr\left(Z = \frac{24.935 - 25}{3}\right) = 0.5 - \Pr(Z = 0.022) = 0.5 - 0.0088 = 0.4912$$

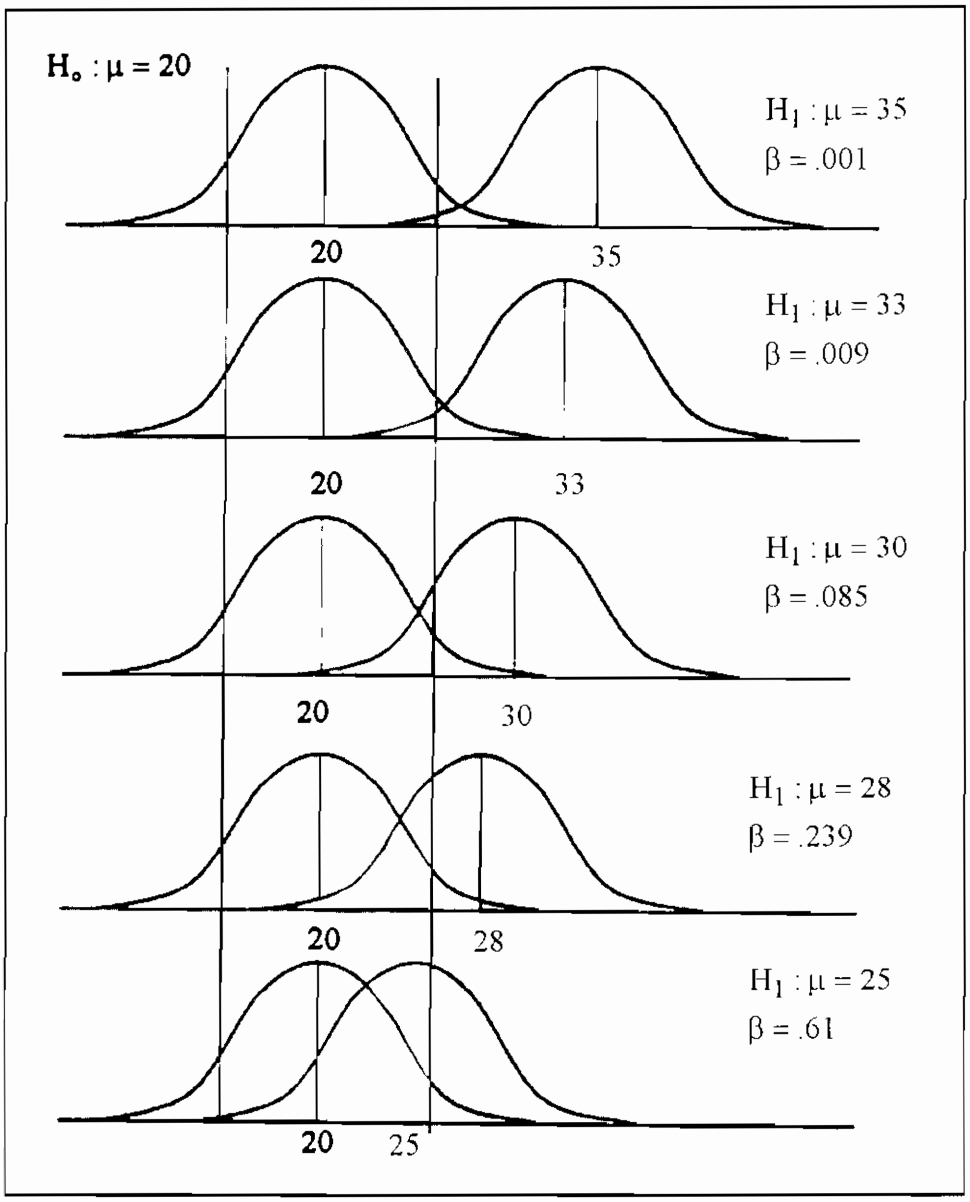
وهذه هى قيمة β، وتكون قوة الاختبار عبارة عن: $1 - 0.491 = 0.509$

وهناك علاقة وثيقة بين قيمة H_1 وقوة الاختبار فكلما كانت القيمة البديلة التى يمكن أن تقبل فى حالة رفض فرض العدم بعيدة عن $Ho: \mu_0 = \mu$ (أى المتوسط الفرضى) كلما قلت β وزادت قوة الاختبار أى كلما تمكن الباحث من أن يرفض هذا الفرض ويعلن أن المتوسط μ_1 يختلف عن μ_0 الفرضى وهذا منطقى ويبين شكل ٦–٥ هذه العلاقة بوضوح. فعند $\alpha = 5\%$ ‏(2.5% على كل جانب) وفى حالة ما إذا كان $H_0: \mu = 20$ فإن $H_1: \mu = 25$ أعطت $\beta = 0.61$، بينما إذا كانت $H_1: \mu = 28$ فإن β تصبح 0.239 وإذا بلغت $H_1: \mu = 35$ فإن β تصبح 0.001 أى أن قوة الاختبار فى هذه الحالة تصبح $(1 - 0.001 = 0.999)$، والشكل ٦–٥ يبين أنه كلماً زاد بعد H_0 عن H_1 ازاد المنحنيان انفصالا عن بعضها، وأمكن التمييز بوضوح بين H_0 ، H_1 وبالتالى زادت قوة الاختبار والعكس صحيح لدرجة أنه إذا تطابقت H_0 ، H_1 تصبح قوة الاختبار مساوية لقيمة α.

وبدراسة شكلى ٦–٤، ٦–٥ تتضح العلاقة العكسية بين كل من α، β فإذا زيدت قيمة α أى تحرك الخط الرأسى 25.88 يساراً ويؤدى هذا إلى انكماش المساحة المظللة الممثلة لقيمة β والعكس صحيح.

من المعلوم أن تفرطح المنحنى الطبيعى يرجع إلى التباين فإذا زاد التباين زاد التفرطح والعكس صحيح. ومن شكل ٦–٥ يبدو أنه كلما زاد التفرطح كلما ازداد تداخل منحنيا H_0 ، H_1 وبالتالى كلما زادت قيمة β. ومن هذا يمكن استنباط القاعدة المهمة فى التجريب الإحصائى أنه كلما قل التباين كلما أمكن اختبار الفروض بقوة

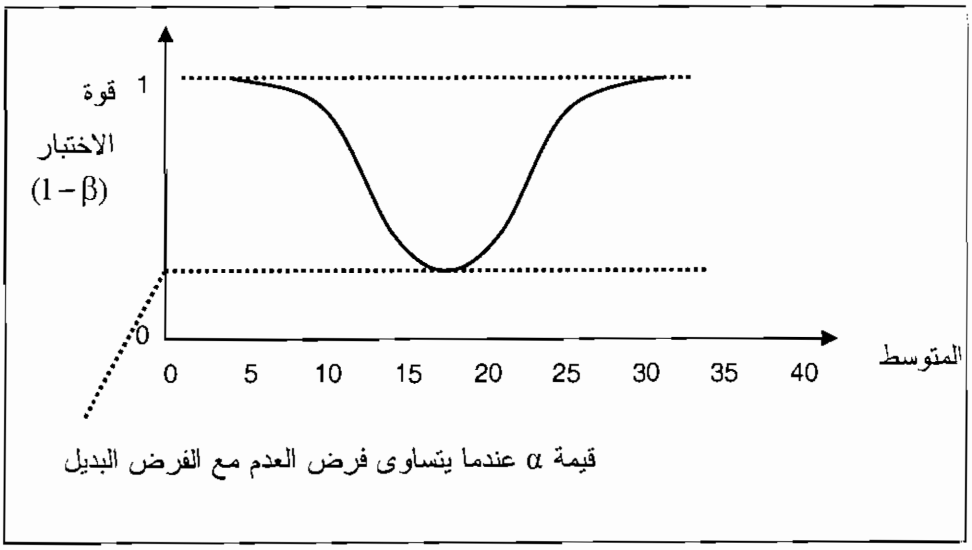
أكبر وزادت مقدرة المجرب على فصل H_0 من H_1. ولتقليل قيمة $\sigma_{\bar{Y}}^2$ هناك عدة طرق منها زيادة الوحدات التجريبية حيث إن $\sigma_{\bar{Y}}^2 = \sigma_Y^2 / n$، وأيضاً عن طريق اتباع التصميمات الإحصائية المناسبة.



شكل ٦-٥ علاقة قيمة الفرض البديل بخطأ النوع الثانى β، الخطوط الرئيسية تمثل 5% (2.5% على كل جانب)

ويلاحظ أنه عند حساب خطأ النوع الأول α (مثلاً شكل ٦-٤) حسب احتمال أن تكون القيمة 25.88 أو أقل وهى المساحة المحددة بين القيمة الحرجة 25.88 إلى نهاية المنحنى يساراً، والتى افترض أنها تقع كلها فى منطقة القبول. وهذا ليس صحيحا تماما لأنه من المعلوم أن حدود التوزيع الطبيعى النظرية هى $\pm\infty$ وبالتالى فإن الطرف الأيسر سيستمر نظريا ليعبر منطقة القبول إلى المنطقة الحرجة عن يسار $H_0 : \mu = 20$. ولكن هذه المساحة متناهية فى الصغر ويمكن تجاهلها.

قد يتساءل البعض لماذا حددنا بالذات أن $H_1 : \mu_1$ يبعد 5 kg عن المتوسط فى كلا الاتجاهين عند حساب خطأ النوع الثانى؟ وهذا سؤال فى محله إذ إنه نادراً ما يعلم المجرب بالضبط أن يحدد متوسطات الفروض البديلة ولكن هنا قد افترضت لترسيخ مفهوم خطأ النوع الثانى وقوة الاختبار. والقول الأعم أن المجرب يود اختبار بحيث تكون قوته أكبر ما يكون على مختلف فروض بديله ممكنة، فمثلاً إذا رسمت منحنيات شكل ٦-٥ على هيئة رسم بيانى (شكل ٦-٦) مع وضع القوة $(1-\beta)$ على الإحداثى الصادى ينتج ما يسمى بمنحنى قوة الاختبار power curve ومنه يمكن القول إن قوة الاختبار تكاد تقترب من الواحد الصحيح إذا ما زاد الفرق بين μ فى H_0، H_1 عن 7، وعند فرق أقل من هذا تنخفض القوة إلى أن تصبح مساوية لقيمة α عندما تنطبق قيمة μ عند كل من H_0، H_1. ويقل اتساع الوادى الناشئ لها من المنحنى فى هذا الشكل كلما قلت قيمة σ_Y^2.



شكل ٦-٦ منحنى قوة الاختبار

افترض فى المعالجات السابقة أن التباين الحقيقى σ_Y^2 للعشيرة معلوماً، وهذا افتراض نادراً ما يتحقق ولكن عادة σ_Y^2 غير معلوم ويستخدم بدلا منه تقديراً له هو S_Y^2 المقدرة من العينة وكل ما قيل عن الحالة حيث σ_Y^2 معلوماً يقال أيضاً فى حالة استخدام S_Y^2 بخلاف واحد ألا وهو استخدام جدول t بدلا من جدول التوزيع الطبيعى Z ويكون المدخل إلى القيم فى الجدول مقابلاً لعدد درجات الحرية فى العينة التى قدر على أساسها S_Y^2.

صندوق ٦-٢

الخطأ من النوع الأول α: احتمال أن يرفض فرض العدم وهو صحيح

الخطأ من النوع الثاني β: احتمال ألا يرفض فرض العدم وهو خطأ

قوة الاختبار $(1-\beta)$: 1 – خطأ النوع الثانى

٦-٤ تحديد العدد الأمثل للمشاهدات فى التجربة (حجم العينة)

العدد الأمثل للتجربة هو أقل عدد ممكن من الوحدات التجريبية (حيوانات مثلاً) الذى يحقق للمجرب اختبار فروق معينة تحت احتمال لكل من α، β محددتين. والسؤال الذى كثيراً ما يوجهه المجرب إلى الإحصائى هو كم حيوانا أو قطعة زراعية أو وحدة تجريبية يجب أن تتناوله التجربة حتى تنتج هذه التجربة المعلومات التى يريدها المجرب ؟ ولإجابة موضعية عن السؤال يجب توافر كل المعلومات التالية:

١- تباين الخطأ σ_e^2 فى التجربة أو فى مثل هذا النوع من التجارب. وإن لم يكن يعلمها المجرب فعليه البحث عنها فى المراجع للحصول على قيمتها المتوقعة فى مثل هذه التجارب. إذا تعذر عليه ذلك أيضا، يمكن توقع قيمتها من معرفة معامل الاختلاف C.V. فى مثل هذه التجارب والمتوسط الذى يتوقعه.

٢- قيمة احتمال خطأ النوع الأول α الذى يكون المجرب على استعداد لتقبله.

٣- قوة الاختبار أى $(1-\beta)$ التى يريد المجرب أن تحققها التجربة.

٤- الفرق بين المعاملات أو الأقسام الذى يود أن يعلنه الباحث معنويا.

مثلا هل يود الباحث أن يعلن فرقاً معنوياً بقدر 5 kg فى وزن فطام الحملان؟ أم أنه يريد أن يزيد من حساسية اختباره ليتمكن من أن يعلن الفرق 0.5 kg معنويا؟ هذا هو قرار المجرب نفسه يقرره طبقاً لخبرته وما يريده من وراء التجربة المقامة. ويعبر عن الفرق هذا فى صورة مجموع مربعات انحرافات المتوسطات المتوقعة عن متوسطها. فمثلا إذا شاء المجرب أن يعلن الفرق بين معاملة متوسطها 30 kg وأخرى متوسطها 35 kg معنوياً فيكون متوسط المعاملتين 32.5 ويكون مجموع مربعات انحرافات المعاملات عن متوسطها $12.5 = (35-32.5)^2 + (30-32.5)^2$. وإذا زاد عدد المعاملات عن اثنتين أصبح من الصعب تخيل مجموع المربعات هذا وعلاقته بالفرق بين متوسطى أى معاملتين. فإذا رمز لمتوسط المعاملة بالرمز μ_j والمتوسط العام بالرمز μ فإن مجموع مربعات انحرافات المتوسطات عن متوسطها هو $\sum(\mu_j-\mu)^2$ ويمكن كتابته $(k-1)(d^2)/2$ حيث k هى عدد المعاملات، d هو الفرق المراد اختباره ففى المثال السابق $k=2$، $d=5$ وتكون:

$$(k-1)((d^2)/2 = (2-1)(5)^2/2 = 12.5$$

وإذا كان هناك 4 معاملات مثلا وقدر الفرق المراد اختباره أيضا 5 فيكون هذا التعبير:

$$(4-1)(5)^2/2 = 37.5$$

وقد وصف Tang (1938) طريقة تحديد حجم العينة. وتفترض هذه الطريقة أن الوحدات التجريبية تتوزع طبيعيا بتباين واحد هو σ_e^2، وفى هذه الحالة يمكن تعريف القيمة φ (فاى):

$$\varphi = \sqrt{\frac{\sum(\mu_j-\mu)^2/k}{\sigma_e^2/n}} \qquad (٦-٦)$$

حيث k هى عدد المعاملات، $\sum(\mu_j-\mu)^2$ هو مجموع مربعات انحرافات المتوسطات عن متوسطها، σ_e^2 هى تباين الخطأ، n هى العدد الوحدات التجريبية فى كل معاملة. ويجب معرفة تصميم التجربة بالتقريب حتى يمكن منه حساب درجات الحرية للخطأ تحت أى n. وكل محددات هذه القيمة هى معطيات يعلمها الباحث إلا n. فهو يفترض قيمة لها ثم يحسب φ ، ومن الرسوم البيانية (١ إلى ٩ ملحق ب) يحدد إذا كانت التجربة تحت α معينة ستحقق له قوة الاختبار المطلوبة، فإذا كانت قوة الاختبار أقل مما حدده المجرب عليه أن يجرب n أكبر وإذا كانت أكبر مما حدد فيجرب n أصغر وهكذا بالتذبذب بين قيم n يحدد العدد الذى يجب أن يكون فى كل معاملة. ومدخل هذه المنحنيات هو v_1 أى درجات الحرية بين المعاملات وهى $(k-1)$.

مثال ٦-٦

مجرب لديه ٤ معاملات ويود أن يعلن فروقاً مثل تلك التى بين المتوسطات 26, 21, 16, 17 معنوية عند $\alpha = 0.05$، $(1-\beta) = 0.9$، وأن تقدير تباين الخطأ $\sigma_e^2 = 49$. ما عدد الحيوانات الأمثل فى كل معاملة، إذا علم أن التصميم هو كامل العشوائية؟

المحاولة الأولى:

$n = 5$ ومنها سيكون العدد الكلى للحيوانات $20 = (4)(5)$ أى 19 درجة حرية منها 3 بين المعاملات (إذاً $v_1 = 3$) والباقى 16 داخل المعاملات أو الخطأ أى أن $(v_2 = 16)$، ذلك طبقا للتصميم التجريبى الموضوع. متوسط المعاملات كلها 20 بالتعويض فى المعادلة (٦-٦)

$$\varphi = \sqrt{\frac{\left[(26-20)^2 + (21-20)^2 + (16-20)^2 + (17-20)^2\right]/4}{49/5}} = \frac{3.94}{3.13} = 1.26$$

وبدخول رسم بيانى (٣ ملحق ب) عند $v_1 = 3$، $v_2 = 16$، $\alpha = 0.05$ يتضح أن هذا العدد يحقق قوة اختبار أقل من 0.5 وعليه لابد من زيادة n، عدد الحيوانات فى كل معاملة.

المحاولة الثانية:

$n = 15$ منها $v_2 = 56$، $\varphi = 2.18$ وطبيعى فإن v_1 ثابتة. وهذا يحقق قوة اختبار أعلى من 0.9 فيتم اختيار n أقل.

المحاولة الثالثة:

$n = 15$ ومنها $v_2 = 48$، $\varphi = 2.03$، وهذه أيضاً تحقق قوة اختبار أعلى من 0.9 فتختار n أقل.

المحاولة الرابعة:

$n = 12$ ومنها $v_2 = 44$ ، $\varphi = 1.9$ وهذه تعطى قوة اختبار مساوية 0.9 تقريباً فيكون العدد الأمثل هو 12 فى كل معاملة.

والقيمة $\sum(\mu_j-\mu)^2 = 26$ يمكن الحصول عليها لو أن المجرب قام بها بصورة أسهل تخيلا وهى أنه إذا أراد أن يختبر فرقاً قدره 6.43 بين متوسطات المعاملات وهذا ينتج مجموع مربعات انحرافات قدره $(3)(6.43)^2 \div 2 = 62.02$

ويجب أن يستوضح القارئ لنفسه أن حجم العينة يزداد إذا انخفضت α، إذا انخفضت β، إذا قل الفرق المراد اختباره و إذا زادت σ_e^2 والعكس صحيح. وعند استخدام الرسوم البيانية يمكن تقريب v_2 إلى أقرب قيمة لها فى المنحنى.

وإن كانت الطريقة السابقة والتى تستخدم المنحنيات مفيدة جدا فى حساب قوة الاختبار عند كل توليفة من σ^2 و α وفروق بين المتوسطات، إلا أنه توجد جداول تسهل هذه العملية إذا كان الهدف فقط هو حساب العدد الأمثل من الوحدات التجريبية. ففى جدول ١٨ ملحق أ لكى يمكن تحديد العدد الأمثل يجب معرفة:

١- عدد المعاملات.

٢- أقصى فرق بين متوسطى معاملتين.

٣- σ_e، α، β

ففى المثال السابق عدد المعاملات $r = 4$ وأقصى فرق هو $26-10=16$ و $\sigma_e = 7$ و $\alpha = 0.05$ و $\beta = 0.9$، ومنه يمكن تحديد القيمة

$$\frac{\Delta}{\sigma} = \frac{10}{7} = 1.43$$

وإذا فرض أن هذه القيمة هى 1.5 (لتواجدها فى الجدول) فإن العدد الأمثل يكون 14 وهو رقم قريب إلى حد ما من الرقم 12 المحسوب بالطريقة المطولة. أيضا فى هذه الطريقة يمكن التعبير عن الفروق فى صورة انحرافات قياسية وليست مطلقة وفى هذه الحالة لا يتطلب معرفة تقدير للانحراف المعيارى (Neter et al 1996).

تمارين الباب السادس

٦-١ البيانات التالية تمثل محصول لبن الموسم الأول لمجموعة الأبقار بالكيلوجرام والذى يفترض أنه يتوزع طبيعياً. فإذا عملت أن $\sigma_Y^2 = 384400\ kg^2$ فى العشيرة ، احسب حدى ثقة 95%، 99% للمتوسط

3564, 3684, 3624, 3213, 3534,

4515, 3647, 2943, 3700, 2568

٦-٢ أعد التمرين (٦-١) مع استخدام S_Y^2 بدلا من σ_Y^2.

٦-٣ البيانات التالية تمثل المتوسط اليومى لإنتاج اللبن بالكيلوجرام لمجموعة من الأبقار، احسب تقدير التباين σ_Y^2 واحسب حدى الثقة له عند 95% و99% مع افتراض التوزيع الطبيعى لهذه الصفة

10.4, 15.6, 19.8, 17.6, 14.3, 12.6,

14.2, 12.2, 14.5, 17.2, 14.9, 13.2

٦-٤ قدر متوسط درجات الطلبة فى مادة الإحصاء فكان 73.2 درجة وذلك لعينة عشوائية عددها 144 طالبا. فإذا كان الانحراف القياسى لعشيرة الطلبة بالكلية فى هذه المادة 8 درجات، احسب حدى الثقة عند 95% و 99% لهذا المتوسط.

٦-٥ قدر طول الجسم فى عينة عشوائية عدد 200 سمكة من أسماك البلطى فكان 17.5 سم، فإذا كان الانحراف القياسى لطول الأسماك $\sigma_Y = 2.75\ cm$ فى عشيرة البلطى، احسب حدى الثقة عند 95% و 99% لهذا المتوسط.

٦-٦ فى عشيرة من الجاموس كان تباين وزن الميلاد بها $\sigma_Y^2 = 25\ kg^2$. تم أخذ عينة عشوائية وكانت أوزانها كالتالى:

34.1, 33.1, 30.1, 32, 28.1, 32.4, 32.5, 31.2, 34.4,

32.4, 25.6, 35.5, 31.5, 31.4, 38.4

حدد الباحث قيمة $\alpha = 5\%$ (2.5% على كل جانب)، وأن فرض العدم $H_0 : \mu = 35$

(أ) احسب قيمة β للفرض البديل $H_1 : \mu_1 = 27$ ، والفرض البديل $H_1 : \mu_1 = 38$.

(ب) إذا فرض الباحث $\alpha = 0.01$ (0.005 على كل جانب) أعد حساب β

(ج) أعد حسابات β إذا كانت σ_Y^2 غير معلومة واستخدم الباحث مكانها S_Y^2 المقدرة من العينة.

(د) قدر قوة الاختبار فى كل الأحوال السابقة.

٦-٧ مجرب يود أن يعلن فرقاً معنوياً قدره 2 kg بين متوسطات 4 معاملات عند $\alpha = 0.05$ ، $\beta = 0.2$ فإذا علم أن $\sigma_e^2 = 36$ وأن التصميم التجريبى هو تام العشوائية completely random، ما هو العدد الأمثل من الحيوانات فى كل معاملة؟

٦-٨ مجرب يود أن يعلن فرقاً معنويا بين معاملتين متوسطهما 40, 30 kg عند $\alpha = 0.05$ فإذا علم أن تباين هذه الصفة $\sigma^2 = 64\ kg^2$ وأن بكل معاملة 10 حيوانات، ما هو احتمال أن يتمكن المجرب من إعلان ما يريد إعلانه؟

٧ اختبارات البيانات العددية: مربع كاى (χ^2)

Enumeration data tests: Chi-Square (χ^2)

١- مقدمة

٢- توزيع مربع كاى χ^2

٣- حساب χ^2 من البيانات العددية

٤- العلاقة بين χ^2 و t

٥- استخدامات χ^2 على البيانات المستمرة

٦- استخدامات χ^2 على البيانات غير المستمرة

٧-١ مقدمة

يستخدم توزيع "ذو الحدين" لمعالجة البيانات غير المستمرة أو المتقطعة discontinuous or discrete data إحصائياً إذا كان عدد الفئات classes محدوداً باثنين، وقد يكون توزيع "بواسون" أكثر موافقة فى أحيان أخرى. أما إذا زاد عدد الفئات عن اثنين (حالة متعددة الحدود multinomial) فإنه فى هذه الحالة يستخدم المفكوك $(p+q+r+\cdots)^n$ بدلاً من $(p+q)^n$ كما فى حالة توزيع "ذو الحدين". ويكون تقدير المتوسط، والتباين ... الخ، من تقديرات المعالم الإحصائية، والتى تلزم لتحديد ما إذا كانت هناك فروق معنوية للانحرافات عن التكرارات المتوقعة expected frequencies أكثر تعقيداً، وعلى ذلك فإنه يفضل استخدام اختبارات المعنوية المعتمدة على توزيع مربع كاى فى مثل هذه البيانات غير المستمرة.

٧-٢ توزيع مربع كاى Chi-Square distribution

إذا كان X متغيراً عشوائياً يتوزع طبيعياً بمتوسط μ وتباين σ^2 بحيث إن σ^2 أكبر من الصفر فإن $Z=(X-\mu)/\sigma$ تتبع توزيعاً طبيعياً متوسطه الصفر وتباينه يساوى الواحد الصحيح أى أن $Z\sim N(0,1)$ و $Z^2=(X-\mu)^2/\sigma^2$ تتوزع طبقاً لتوزيع مربع كاى بدرجة حرية واحدة أى أن:

$$Z^2 \sim \chi^2 \ , \ df = 1$$

ومعنى ذلك أنه إذا أخذت مشاهدة كانحراف من متوسط العشيرة مقسومة على الانحراف المعيارى للعشيرة لينتج ما يسمي بالانحراف الطبيعى القياسى standard normal deviate فإن مربع تلك الكمية $(X-\mu)^2/\sigma^2$ يتوزع حسب توزيع مربع كاى بدرجة حرية واحدة.

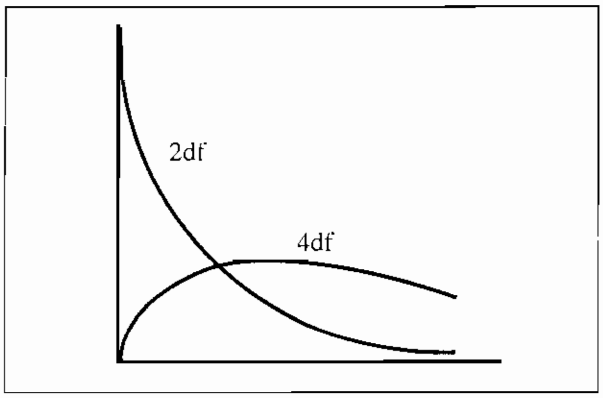
وإذا كان هناك متغير Y يتبع توزيع مربع كاى بدرجات حرية m، ومتغير آخر Z يتبع توزيع مربع كاى أيضا بدرجات حرية n فإن مجموع المتغيرين $Y+Z=W$ يتبع توزيع مربع كاى بدرجات حرية مقدارها $m+n$ على أن يكون المتغيران مستقلين.

وعلى ذلك إذا كانت $X_1, X_2, \cdots, X_n$ تمثل عينة عشوائية حجمها n مأخوذة من عشيرة تتوزع طبيعياً بمتوسط μ وتباين σ^2 فإن المتغير $Y=\sum\frac{(X-\mu)^2}{\sigma^2}$ له توزيع مربع كاى بدرجات حرية مقدارها n.

وحيث إن $\sigma^2 = \frac{\Sigma (X-\mu)^2}{df}$ كما سبق ذكره فى الباب الثالث فإن $df = \frac{\Sigma (X-\mu)^2}{\sigma^2}$ والتى تساوى أيضاً $\chi^2_{(n)}$ المتوقعة أى أن $df = \chi^2_{(n)}$

ويتضح أن هذا التوزيع له علاقة وثيقة جداً وتامة بدرجات الحرية، ومعلم parameter هذا التوزيع هو درجات الحرية حيث متوسط توزيع مربع كاى = n والتباين لهذا التوزيع ضعف متوسطة أى 2n، وقيمة χ^2 تتراوح بين الصفر وما لا نهاية (∞) فهى دائماً موجبة حيث تحتوى على مجموع مربعات.

ويلاحظ أنه لكل عدد من درجات الحرية يوجد توزيع لمربع كاى بعضها مبين فى شكل ٧-١.



شكل ٧-١ توزيع مربع كاى لدرجات حرية مختلفة 2 و 4

ويلاحظ من شكل ٧-١ أن توزيع مربع كاى يقترب من التماثل كلما زادت درجات الحرية وجدول ٦ ملحق أ يبين قيم χ^2 لعدة توليفات من الاحتمالات ودرجات الحرية.

وحيث إن تباين العشيرة يكون عادة غير معروف، ويستخدم تقدير له بالنقطة هو $S^2 = \frac{\Sigma (X-\overline{X})^2}{n-1}$ فإن $\Sigma(X-\overline{X})^2 = (n-1)\,S^2$ حيث $\Sigma(X-\overline{X})^2$ هى مجموع مربع الانحرافات عن المتوسط، والتى تتكون من $(n-1)$ من الانحرافات المستقلة متوسطاتها تساوى الصفر، وبقسمة هذه الانحرافات على σ حتى تكون تبايناتها مساوية للوحدة فإن $\frac{(n-1)S^2}{\sigma^2}$ تتبع توزيع χ^2 وتستخدم فى حساب فترة الثقة لتباين

العشيرة σ^2. وتوزيع χ^2 بدرجة حرية واحدة له علاقة مباشرة بالتوزيع الطبيعى فقيمة χ^2 بدرجة حرية واحدة ومستوى معنوية 5% هى 3.24 (جدول ٦ ملحق أ). أى أن $\Pr(\chi^2 \geq 3.84) = 0.05$.

ومن تعريف χ^2 أنها عبارة عن مربع انحراف متغير عشوائى يتوزع طبيعياً بمتوسط صفر وتباين يساوى الواحد الصحيح أى أن $Z = \sqrt{3.84} = 1.96$، واحتمال أن Z أكبر من أو تساوى 1.96 (جدول ٣ ملحق أ) عبارة عن

$0.05 - 0.4750 = 0.025 = (0.5)(0.05)$

(حيث إن توزيع χ^2 فى الجهة الموجبة فقط بينما توزيع Z من الجهتين). وعلى ذلك فإن جدول التوزيع الطبيعى القياسى مركز فى سطر واحد فى جدول χ^2 لدرجة حرية واحدة، والتوزيع الطبيعى القياسى يمتد من $-\infty$ إلى $+\infty$ أما توزيع χ^2 فهو مربع Z، ولذا فإنه يمتد من صفر إلى ∞.

٧-٣ حساب χ^2 من البيانات العددية

تستخدم المعادلة التالية لإيجاد قيمة χ^2 فى حالة البيانات العدية count data

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad (٧-١)$$

حيث

O_i تمثل العدد المشاهد observed number للفئة i،

E_i تمثل العدد المتوقع expected number للفئة i،

k تمثل عدد الفئات classes

والتى اشتقت فى حالة توزيع "ذو الحدين" كما يلى:

إذا ما أخذ عدد الولادات فى أحد المستشفيات، فإنه من السهل معرفة عدد المواليد الذكور X من عدد المواليد الكلى n، والمتوقع أن يكون np بتباين قدره npq (كما ذكر فى الباب الرابع)، وإذا ما رمز للعدد المشاهد بالرمز O والعدد المتوقع بالرمز E فإن

$$\chi^2 = \frac{(X-\mu)^2}{\sigma^2} = \frac{(X-np)^2}{npq} = \frac{(O-E)^2}{np(1-p)} = \frac{(O-E)^2}{np} + \frac{(O-E)^2}{n(1-p)}$$

حيث إن $p + q = 1$،

و $\frac{1}{p}+\frac{1}{1-p}=\frac{(1-p)+p}{p\,(1-p)}=\frac{1}{p\,(1-p)}$

وبالتالى فإن:

$$\chi^2=\frac{(O_A-E_A)^2}{E_A}+\frac{(O_B-E_B)^2}{E_B}$$

وحيث إن E_A و E_B هما العدد المتوقع فى الفئتين فإن $\chi^2=\Sigma\frac{(0-E)^2}{E}$ وهى الصورة المستخدمة عادة فى الحسابات.

ودرجات الحرية مقدارها $(k-1)$، وتفقد درجة حرية من عدد الفئات k حيث

$$E_i=np_i$$
$$\sum_i E_i=\sum_i np_i=n\sum_i p_i=n$$

حيث إن $\sum P_i=1$

فإذا كان $k=2$ فيكون هناك حرية اختيار فئة واحدة أما الفئة الأخرى فهى غير مستقلة إذا علم أن n تمثل مجموع الفئتين.

مثال ٧–١

فى عينة من 1000 كتكوت وجد 470 كتكوتاً ذكراً، احسب قيمة χ^2 إذا كانت النسبة الجنسية 1:1.

حيث إن عدد الذكور المشاهد 470

فإن عدد الإناث المشاهد $1000-470=530$ وبالتالى فإن:

	العدد المشاهد O_i	العدد المتوقع E_i	$(O_i - E_i)$	$(O_i - E_i)^2$
ذكور	470	$(1000)(.05) = 500$	$470 - 500 = -30$	900
إناث	530	500	$530 - 500 = 50$	900
المجموع	1000	1000	0	

وبالتالى فإن

$$\chi^2 = \sum_i \frac{(O_i - E_i)}{E_i} = \frac{900}{500} + \frac{900}{500} = 3.6$$

بدرجة حرية واحدة.

٧-٤ العلاقة بين χ^2 و t

كما سيأتى فيما بعد (الباب الثامن) فإن اختبار t يمثل النسبة بين الانحراف عن المتوسط إلى الانحراف المعيارى للفرق، ومن الفصل السابق فإنه إذا كان هناك فئتان A و B والعدد المشاهد فى كل x_1 و x_2 والعدد المتوقع np و nq على التوالى فإن

$$\chi^2 = \frac{(x_1 - np)^2}{np} + \frac{(x_2 - nq)^2}{nq} = \frac{(x_1 - np)^2}{npq}$$

حيث إن $p + q = 1$ وإن $(x_2 - nq)^2 = (x_1 - np)^2$

وبأخذ الجذر التربيعى للطرفين فإن

$$\sqrt{\chi^2} = \sqrt{\frac{(x_1 - np)^2}{npq}} = \frac{x_1 - np}{\sqrt{npq}}$$

وهى عبارة عن الفرق مقسوما على الانحراف المعيارى للفرق أى أن:

$$\sqrt{\chi^2} = t$$

وبالتالى فإن

$$t^2 = \chi^2 \qquad (٧-٢)$$

أى أن χ^2 تساوى مربع t.

مثال ٧-٢

إذا كان عدد النباتات الطويلة 150، والقصيرة 30 فهل الانحرافات عن النسبة المتوقعة 3:1 معنوية ؟

طبقاً لتوزيع "ذو الحدين" فإن:

العدد المتوقع للنباتات الطويلة: $np = \frac{3}{4}(150+30) = 135$

الانحراف المعيارى: $\sqrt{npq} = \sqrt{180 \text{ x } \frac{3}{4} \text{ x } \frac{1}{4}} = 5.81$

$$t = \frac{150-135}{5.81} = \frac{15}{5.81} = 2.582^{*} > 1.96$$

حيث 1.96 هى قيمة t الجدولية عند مستوى معنوية 5% ودرجات حرية ∞ (جدول ٤ ملحق أ). استخدمت t مجازاً بدلاً من التوزيع الطبيعى حيث إن Z = t عند درجات حرية ∞.

أما إذا استخدمت χ^2 فإن:

الفئة	O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$
النباتات الطويلة	150	135	15	225
النباتات القصيرة	30	45	-15	225
إجمالى	180	180	zero	450

وقيمة $\chi^2 = \frac{225}{135} + \frac{225}{45} = 6.667^{*} > 3.84$ حيث 3.84 هى قيمة χ^2 عند مستوى معنوية 5% ودرجة حرية واحدة (جدول ٦ ملحق أ). لاحظ أن $t^2 = \chi^2$ أى أن $(2.582)^2 = 6.667$.

٧-٥ استخدامات χ^2 على البيانات المستمرة

من أهم استخدامات χ^2 على البيانات المستمرة:

١- يستخدم χ^2 فى اختبار تجانس معاملات الارتباط لعدد من العينات كما سيأتى ذكره فى الباب العاشر.

٢- يستخدم أيضاً فى حساب فترة أو حدود ثقة confidence interval لتباين العشيرة كما ذكر فى الباب السادس وهى طريقة دقيقة تماماً.

٣- فى اختبار تجانس التباين كما سيأتى ذكره فى الباب الثانى عشر.

٤- فى مدى ملائمة توزيع ما للتوزيعات المستمرة Goodness of fit for continuous distribution لمعرفة ما إذا كانت مجموعة من البيانات تتوزع حسب توزيع معين، وليكن التوزيع الطبيعى، وللمقارنة بين التوزيعين فإن التكرارات المتوقعة يلزم حسابها، ولحساب التكرارات المتوقعة فإنه يلزم إيجاد احتمال الحصول على كل فترة والتى يمكن الحصول عليها من جدول التوزيع الطبيعى القياسى (جدول ٣ ملحق أ). أما التباين والمتوسط فيقدران من البيانات المتاحة، ويعتبران تقديرين غير متحيزين (μ, σ^2) ويتضح ذلك من المثال التالى.

مثال ٧-٣

هل البيانات التى فى مثال ٣-١٩ تتفق والتوزيع الطبيعى ؟

حدود الفئات	التكرارات (العدد المشاهد) O_i	الاحتمال	العدد المتوقع E_i	$(O_i - E_i)$	$(O_i - E_i)^2$
أقل من 20.5	6	0.0154	14.81	-8.81	77.62
20.5-23.5	20	0.0581	55.89	-35.89	1288.09
23.5-26.5	187	0.1531	147.28	39.72	1577.68
26.5-29.5	340	0.2574	247.62	92.38	8534.06
29.5-32.5	181	0.2646	254.55	-73.55	5409.60
32.5-35.5	129	0.1676	161.23	-32.23	1038.77
35.5-38.5	65	0.0655	63.01	1.99	3.96
أكثر من 38.5	34	0.0183	17.60	16.40	268.96
المجموع	962	1.0000	961.99	0.01	

ويجب ألا يختلف مجموع الاحتمالات عن الواحد الصحيح، وأيضاً ألا يختلف مجموع الأعداد المتوقعة عن مجموع الأعداد المشاهدة، وألا يختلف مجموع

الانحرافات $\sum(O_i - E_i)$ عن الصفر إلا فى حدود خطأ التقريب. ولقد استخدم المتوسط $\bar{X} = 29.66$ كتقدير لمتوسط العشيرة. واستخدم $S^2 = (4.236)^2$ كتقدير لتباين العشيرة أى أن $S = 4.236$ مع ملاحظة أنه لحساب الاحتمال للخلية الأولى اعتبر أن حدودها من $-\infty$ إلى أقل من 20.5، أما احتمال القيمة الأخيرة، فتم حسابه على أساس حدود الفئة أكبر من 38.5 وبالتالى:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} = \frac{77.62}{14.81} + \frac{1288.09}{55.89} + \cdots + \frac{268.96}{17.6} = 116.5^{**}$$

درجات الحرية= عدد الفئات − 1 − عدد الثوابت المقدرة أى $5 = 2 - 1 - 8$

χ^2 الجدولية بمستوى معنوية 5% ودرجات حرية 5 = 11.07

χ^2 الجدولية بمستوى معنوية 1% ودرجات حرية 5 = 15.09

وحيث إن χ^2 المحسوبة أكبر من χ^2 الجدولية عند مستوى معنوية 1% فإن البيانات لا تتفق والتوزيع الطبيعى بخطأ من النوع الأول مقداره 1%.

ويجب ملاحظة أنه إذا كان العدد المتوقع للفئات الأولى قليلاً فتدمج فى فئة واحدة، وكذلك الفئات الثلاث أو الأربع الأخيرة فتدمج فى فئة واحدة أخيرة بحيث لا يقل عدد التكرارات المتوقعة فى الفئة عن 5−10 كحد أدنى، وبالتالى ففى المثال السابق تم دمج الفئتين الأخيرتين فى فئة واحدة أخيرة حيث إن العدد المتوقع للفئة الأخيرة قبل الدمج كان 2.5 (ويرجع السبب فى ألا يقل العدد عن 5، ويفضل ألا يقل عن 10، إلى أن توزيع χ^2 يعتمد أساساً على التوزيع الطبيعى ولكن χ^2 مشتق هنا من توزيع "ذو الحدين" والذى يقترب من التوزيع الطبيعى كلما زادت n).

٧-٦ استخدامات χ^2 على البيانات غير المستمرة

قد يستخدم أيضاً χ^2 فى حالة البيانات العددية enumeration data وهى تلك البيانات التى تحتوى على أعداد تقع فى فئات معرفة جيداً well-defined classes كعدد الذكور، وعدد الإناث فى مدينة ما مثلاً.

وعندما يستخدم توزيع χ^2 فى مثل تلك البيانات فإنه يكون عادة مرتبط باختبار جودة الملائمة goodness of fit، وتستخدم أيضاً المعادلة (٧-١)، وفيها العدد المشاهد يشير إلى العدد الملاحظ فى الخلية أما العدد المتوقع فيشير إلى العدد المتوقع طبقاً لفرض العدم، وهو القيمة النظرية، ومجموع الانحرافات $\sum(O_i - E_i)$ تساوى الصفر فى حدود خطأ التقريب أما درجات الحرية فتختلف حسب الحالة.

مثال ٧-٤

فى المثال ٧-١ اختبر الفرض القائل بأن النسبة الجنسية 1:1

فرض العدم:

الذكور : الإناث = 1:1

الفرض البديل:

الذكور : الإناث ≠ 1:1

وحيث إن $3.6 = \chi^2$ ودرجات الحرية $1 = 1 - 2$

وقيمة χ^2 الجدولية بدرجة حرية واحدة و 5% مستوى معنوية = 3.84 وهى أكبر من قيمة χ^2 المحسوبة، وبالتالى لا يمكن رفض فرض العدم، وبالتالى فإن النسبة الجنسية هى 1:1 والانحراف عنها إنما يرجع إلى الأخطاء العشوائية أو الصدفة.

ويمكن الحصول على χ^2 بطريقة مباشرة دون الحاجة إلى حساب العدد المتوقع المقابل لكل عدد مشاهد فى حالة قسمين اثنين حيث:

$$\chi^2 = \frac{(r_2 n_1 - r_1 n_2)^2}{r_1 r_2 \, (n_1 + n_2)}$$

وذلك إذا كان هناك فئتين A و B والعدد المشاهد n_1، n_2 على التوالى، ونسبة B : A هى $r_2 : r_1$

وحيث إن χ^2 توزيع مستمر فإنه يتم التصحيح للاستمرارية correction for continuity حتى يكون أكثر ملائمة، وهذا يتيح الحصول على احتمالات أكثر دقة من جداول χ^2 وعلى ذلك فإن Yates اقترح تصحيحاً للاستمرارية يطبق فى حالة ما إذا كانت هناك درجة حرية واحدة، وهذا التصحيح يجعل التوزيع الفعلى للبيانات المتقطعة قريباً جداً من توزيع χ^2 والذى أساسه انحرافات معتدلة طبيعية normal deviates، ويكون التقريب للقيم المطلقة للانحرافات حيث يخفض بمقدار نصف أى أن:

$$\text{Adjusted } \chi^2 = \Sigma \frac{\left(|O_i - E_i| - \frac{1}{2}\right)^2}{E_i} \qquad (٣-٧)$$

وهذا التعديل يؤدى إلى خفض قيمة χ^2 المحسوبة، وعلى ذلك ففى اختبارات الفروض الإحصائية يكون لهذا التعديل فائدة عندما تكون χ^2 الغير معدلة كبيرة عن χ^2 الجدولية عند مستوى المعنوية المرغوب.

إذا كان عدد الخلايا أكثر من اثنين يتبع أيضاً نفس المفهوم فى إيجاد قيمة χ^2 وتكون درجات الحرية هى عدد الفئات أى k مطروحاً منها واحد.

مثال ٧-٥

من تزاوج AaBb مع aabb كانت الأعداد المشاهدة للنسل الناتج كما يلى:

التركيب الوراثى	aabb	aaBb	Aabb	AaBb
العدد الملاحظ	65	30	25	70

فإذا كانت النسبة المتوقعة للتراكيب الوراثية المختلفة هى 1:1:1:1 فهل تختلف الأعداد المشاهدة عن الأعداد المتوقعة ؟ اختبر ذلك إحصائياً.

التركيب الوراثى	**العدد الملاحظ**	**العدد المتوقع E_i**	$(O_i - E_i)$	$(O_i - E_i)^2$
AaBb	70	47.5	22.5	506.25
Aabb	25	47.5	-22.5	506.25
aaBb	30	47.5	-17.5	306.25
aabb	65	47.5	17.5	306.25
المجموع	**190**	**190**	**zero**	

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} = \frac{506.25}{47.5} + \frac{506.25}{47.5} + \frac{306.25}{47.5} + \frac{306.25}{47.5} = 34.2^{**}$$

وحيث إن فرض العدم هو أن نسب التراكيب الوراثية الأربعة 1:1:1:1 ودرجات الحرية $4-1=3$ و χ^2 من الجدول عند 1% ودرجات حرية 3 تساوى 11.3 وبالتالى يرفض فرض العدم، أى أن التراكيب الوراثية لا تتفق والنسبة 1:1:1:1 .

٧-٦-١ اختبار التجانس لعينات من خليتين Test of homogeneity

كثير من البيانات العددية يختص بوجود أو عدم وجود صفة أو ملاحظة معينة وبالتالى فإن الجداول التى بها الأعداد الملاحظة فى خليتين هى الأكثر عرضاً

للبيانات، وقد يكون هناك عدة عينات بنفس القدر من المعلومات، وقد يكون من المرغوب فيه معرفة ما إذا كانت هذه العينات متجانسة حتى إذا كان الأمر كذلك فإنه يمكن تجميعهم، والحصول على تقدير موحد لنسبة العشيرة كتقدير النسبة الجنسية فى مفرخة باستخدام عدة عينات، وقد يكون من المرغوب فيه اختبار نسبة عامة common ratio ولتكن 1:3 مثلاً.

مثال: ٧-٦

إذا كانت البيانات التالية تمثل عدد بذور البسلة الملساء والمجعدة الناتجة فى الجيل الأول من تزاوج فردين خليطين.

	العائلة						
	١	٢	٣	٤	٥	٦	**المجموع**
عدد البذور الملساء	240	38	264	125	95	44	**804**
عدد البذور المجعدة	70	20	72	35	35	20	**252**

اختبر الفرض القائل بأن نسبة البذور الملساء إلى المجعدة هى 1:3 .

الجدول التالى يبين العدد المشاهد والمتوقع لكل عائلة والفروق بينهما لكل من البذور الملساء والمجعدة

العائلة	**البذور الملساء**			**البذور المجعده**		
	العدد المشاهد	**العدد المتوقع**	**الفرق**	**العدد المشاهد**	**العدد المتوقع**	**الفرق**
١	240	232.5	7.5	70	77.5	-7.5
٢	38	43.5	-5.5	20	14.5	5.5
٣	264	252.0	12.0	72	84.0	-12.0
٤	125	120.0	5.0	35	40.0	-5.0
٥	95	97.5	-2.5	35	32.5	2.5
٦	44	48.0	-4.0	20	16.0	4.0
	806	**793.5**	**12.5**	**252**	**264.5**	**-12.5**

وتحسب χ^2 الكلية $(\text{Total}\,\chi^2)$ كالتالى:

$$\text{Total}\,\chi^2 = \frac{(7.5)^2}{232.5} + \frac{(-7.5)^2}{77.5} + \cdots + \frac{(-4)^2}{48} + \frac{(4)^2}{16} = 8.458$$

بدرجات حرية $6 = (6)(2-1)$

وإذا اعتبر أن كل البذور الملساء معاً، وكل البذور المجعدة كوحدة فإن χ^2 المجمعة $(\text{pooled}\,\chi^2)$ يمكن الحصول عليها كالتالى:

$$\text{pooled}\,\chi^2 = \frac{(12.5)^2}{793.5} + \frac{(-12.5)^2}{264.5} = 0.7876$$

بدرجات حرية $1 = 1 - 2$ وهى غير معنوية.

وبطرح χ^2 المجمعة من χ^2 الكلية يمكن الحصول على χ^2 لعدم التجانس heterogeneity أى $8.458 - 0.7876 = 7.67$ بدرجات حرية $5 = 1 - 6$ وهى غير معنوية حيث إن χ^2 الجدولية عند درجات حرية 5 ومستوى معنوية 5% = 11.07. ويتضح أن العينات متجانسة وتنطبق النسبة 1:3 عليها.

ملحوظة: إذا كانت χ^2 لعدم التجانس معنوية فإنه من المنطقى الرجوع إلى χ^2 لكل عينة منفردة $\text{individual}\,\chi^2$.

٧-٦-٢ اختبار التداخل باستخدام مربع كاى χ^2 Interaction

قد تقسم البيانات العددية تبعاً لعدة متغيرات فمثلا قد يكون الفرد ذكراً أو أنثى، وقد يكون مدخناً أو غير مدخن، وقد يكون مثقفاً أو غير مثقف وهكذا، أى يوجد متغيران فى كل تقسيم فى كل حالة. وتوضع البيانات العددية المقسمة تبعاً لأكثر من متغير فى جدول يسمى contingency table فإذا كان هناك متغيران، ويراد اختبار استقلالية المتغيرين عن بعضهما، أى مثلاً معرفة ما إذا كانت نسبة المتعلمين إلى غير المتعلمين فى الذكور هى نفس نسبة المتعلمات إلى غير المتعلمات فى الإناث (مع وجود بعض الأخطاء العشوائية). وإذا كان لا يوجد استقلال، يطلق على الاختبار، اختبار χ^2 للتداخل وتكون البيانات على هيئة جدول عدد صفوفه r وعدد أعمدته c أى $r \times c$.

فإذا كانت $r = 3$ و $c = 4$ فإن الجدول سيكون 3×4 أى أنه يوجد متغيران، بالمتغير الأول 3 أقسام والثانى 4 أقسام، ومن المرغوب فيه اختبار استقلالية المتغيرين. فإذا كان المتغير A له 3 أقسام A_1, A_2, A_3 ومتغير آخر B أقسامه B_1, B_2, B_3, B_4 فإن درجات الحرية الكلية $(3)(4) - 1 = 11$. وهذا يتضح من الجدول التالى:

المتغير الأول A	المتغير الثانى B: B_1	B_2	B_3	B_4	
A_1	x	x	x	.	x
A_2	x	x	x	.	x
A_3	.	.	.	.	.
	x	x	x	.	

وقد فقدت درجة حرية نتيجة أن المجموع الكلى لابد وأن يساوى العدد الكلى للأفراد، وأيضاً أخذ المجموع الهامشى لأعداد كل من B_1, B_2, B_3, B_4 فإنه يمكن تحديد 3 منها اختيارياً، ولهذا أيضاً 3 درجات حرية، وإذا أخذ المجموع الهامشى للأعداد فى كل A_1, A_2, A_3 فإنه يمكن اختيار 2 منها أى درجات الحرية هنا تساوى 2، وبالتالى فإنه بالنسبة للجدول ككل فإنه يمكن اختيار أى فئتين داخل الصفوف، وأى 3 فئات داخل الأعمدة لتنتج درجات الحرية التى مجموعها 6 أى 3x2 وعلى ذلك فإن درجات الحرية للتداخل = (عدد الصفوف – 1) (عدد الأعمدة – 1) أى 3x2 بمعنى أن:

$$df = (r-1)(c-1) \qquad (٧-٤)$$

وهذا ناتج من:

درجات الحرية للتداخل = (درجات الحرية الكلية) – (درجات الحرية للصفوف) – (درجات الحرية للأعمدة)

بمعنى أن:

$$\begin{aligned} df &= (rc-1)-(r-1)-(c-1) \\ &= rc-1-r+1-c+1 \\ &= rc-r-c-1 \\ &= (r-1)(c-1) \end{aligned}$$

ويجب ملاحظة أن اختبار الاستقلالية اختبار متماثل، فإذا كان هناك متعلمون وغير متعلمين فى العامل الأول وذكور وإناث فى العامل الثانى وهو الجنس فإن نسبة الذكور إلى الإناث فى المتعلمين إلى نسبة الذكور إلى الإناث فى غير المتعلمين تعطى نفس النتيجة كما لو كان فرض العدم هو اختبار ما إذا كانت نسبة المتعلمين إلى غير المتعلمين فى الذكور تساوى نسبة المتعلمين إلى غير المتعلمين فى الإناث وتستخدم الكمية التالية أيضاً:

$$\sum_{i}\frac{(0_i - E_i)^2}{E_i} = \sum_{i.j}\frac{(n_{ij} - E_{ij})^2}{E_{ij}} \qquad (٥-٧)$$

والتى تتوزع تقريباً كتوزيع χ^2 بدرجات حرية = (عدد الصفوف – 1) (عدد الأعمدة – 1). وتحسب القيم المتوقعة E_{ij} لكل قيمة مشاهدة n_{ij} بافتراض أن فرض العدم صحيح وهو استقلالية المتغيرين كما يلى:

$$E_{ij} = \frac{n_{i.}\,n_{.j}}{n_{..}} = \frac{\text{(مجموع الصف) (مجموع العامود)}}{\text{المجموع الكلي}}$$

حيث $n_{i.}$ مجموع الصف الذى تقع فيه الخلية و $n_{.j}$ مجموع العمود الذى تقع فيه أيضاً الخلية و $n_{..}$ يمثل العدد الكلى.

أما إذا كانت هناك نسب متوقعة مسبقة فيحسب على أساسها العدد المتوقع لكل خلية.

مثال ٧-٧

فى المثال ٧-٦ هل نسبة البذور الملساء إلى المجعدة هو نفسه بفرض عدم وجود نسبة متوقعة عن طريق حساب الأعداد المتوقعة فى كل خلية أولاً.

العائلة	ملساء		مجعده		المجموع
	مشاهد	متوقع	مشاهد	متوقع	
١	240	236.16	70	73.84	310
٢	38	44.19	20	13.81	58
٣	264	255.97	72	80.03	336
٤	125	121.89	35	38.11	160
٥	95	99.03	35	30.96	130
٦	44	48.76	20	15.24	64
المجموع	**806**		**252**		**1058**

حيث

العدد المشاهد فى العائلة الأولى وبذوره ملساء = 240

والعدد المتوقع المقابل له $236.16 = \frac{806 \text{ x } 310}{1058}$ ... وهكذا

فرض العدم:

البذور الملساء : المجعدة فى العائلة الأولى= البذور الملساء : المجعدة فى العائلة الثانية ... وهكذا. وقيمة χ^2 بدرجات حرية $5 = (6-1)(2-1)$ تكون

$$\chi^2 = \frac{(240-236.16)^2}{236.16} + \frac{(70-73.84)^2}{73.84} + \cdots + \frac{(20-15.24)^2}{15.24} = 7.795$$

وهى غير معنوية وبالتالى لا يرفض فرض العدم، وبالتالى ليس هناك سبب للاعتقاد بأن النسبة من عائلة إلى أخرى تختلف باختلاف العائلة، أى أن كلاً من العائلة وحالة البذور مستقلان.

وإذا كانت χ^2 معنوية فإن الطريقة الشائعة لمعرفة منبع المعنوية هو حذف الصف (الصفوف) أو العمود (الأعمدة) التى يعتقد أنها السبب فى المعنوية (أى التى تكون فيها الأعداد شاذة نتيجة لسبب ما مثل الوفيات أو جين مميت ... الخ) ثم يجرى الاختبار مرة أخرى ... وهكذا.

توجد حالات خاصة من الجدول الذى عدد صفوفه r وعدد أعمدته c منها: عندما يكون عدد الصفوف يساوى عدد الأعمدة (c) أى جدول c x c فإذا كانت الأفراد تقسم تبعاً لخاصية A ولها (A_1, A_2) و B ولها (B_1, B_2) وكانت الأعداد المشاهدة كما يلى:

	الخاصية A		
الخاصية B	**A_1**	**A_2**	**المجموع**
B_1	n_{11}	n_{21}	**$n_{.1}$**
B_2	n_{12}	n_{22}	**$n_{.2}$**
المجموع	**$n_{1.}$**	**$n_{2.}$**	**$n_{..}$**

فإن درجات الحرية $1 = (2-1)(2-1)$ و χ^2 فى هذه الحالة الخاصة عبارة عن

$$\chi^2 = \frac{[(n_{11})(n_{22}) - (n_{12})(n_{21})]^2 (n_{..})}{(n_{1.})(n_{2.})(n_{.1})(n_{.2})} \qquad (٧-٦)$$

وهى معادلة شائعة الاستعمال وتختصر كثيراً من الوقت.

وفى حالة جدول 2×2 فإنه يوجد درجة حرية واحدة؛ ولذا فإن χ^2 المعدلة يمكن الحصول عليها من المعادلة التالية

$$\text{Adjusted}\chi^2 = \frac{[|(n_{11})(n_{22}) - (n_{12})(n_{21})| - n_{..}/2]^2 (n_{..})}{(n_{1.})(n_{2.})(n_{.1})(n_{.2})} \qquad (٧-٧)$$

ملاحظة هامة:

χ^2 فى جميع الحالات يجب أن تحسب على أساس الأعداد الفعلية actual numbers ولا تحسب مطلقاً على أساس نسب مئوية ... الخ.

هذا وقد استحدثت طرق لتحليل χ^2 بشكل يماثل تحليل التباين، أى أنه توجد عدة مصادر للاختلافات وتقسم χ^2 الكلية طبقاً لهذه المصادر آنياً. والطرق المتاحة الآن لتحليل linear log models تشترط تساوى الأعداد واتزان تركيب البيانات وليس فى هذا المؤلف مجال للخوض فيها.

صندوق ٧-١

تستخدم χ^2 فى تحليل البيانات واختبار النظريات الفرضية الخاصة بالبيانات العددية (أى العدية أى التى تُعد).

لإجراء اختبارات χ^2 لابد أن يكون هناك أعداد فعلية ولا يجوز إجراء هذا الاختبار على نسب أو نسب مئوية.

بجانب وجود الأعداد المشاهدة يجب أن يكون هناك أعداد متوقعة تحسب طبقاً لفرض العدم كنسبة 3:1 أو 1:1 أو توزيع الأعداد طبقاً للتوزيع الطبيعى مثلا، إن لم يوجد مثل هذه الأعداد المتوقعة فهناك نوع واحد من χ^2 يمكن حسابه وهو الذى يختبر الفرض أن نسب الأعداد فى المستويات المختلفة لمتغير ما هى نسب واحدة.

يستخدم توزيع χ^2 في اختبار الفروض الخاصة بالتباين وحساب حدود ثقته.

تمارين الباب السابع

٧-١ فئتان A و B والعدد المشاهد n_1 و n_2 على الترتيب. أثبت أن:

(أ) $\chi^2 = \frac{(n_2 - r n_1)^2}{r(n_1 + n_2)}$ إذا كانت النسبة المتوقعة للفئات A : B هى r : 1

(ب) $\chi^2 = \frac{(s n_1 - r n_2)^2}{rs(n_1 + n_2)}$ إذا كانت النسبة المتوقعة هى r : s

(ج) $\chi^2 = \frac{(n_1 - n_2)^2}{(n_1 + n_2)}$ إذا كانت النسبة المتوقعة هى 1 : 1

٧-٢ إذا توافرت البيانات التالية فى جدول 2 x 2 والتى تمثل الأعداد المشاهدة

الخاصية الثانية	الخاصية الأولى		
	A_1	A_2	المجموع
B_1	n_{11}	n_{21}	$n_{.1}$
B_2	n_{12}	n_{22}	$n_{.2}$
المجموع	$n_{1.}$	$n_{2.}$	$n_{..}$

أثبت أن

$$\chi^2 = \frac{[(n_{11})(n_{22}) - (n_{12})(n_{21})]^2 (n_{..})}{(n_{1.})(n_{2.})(n_{.1})(n_{.2})}$$

٧-٣ فى قطيع كانت أعداد الذكور والإناث كالتالى:

العائلة	ذكور	إناث
١	50	50
٢	48	54
٣	21	20
٤	18	18
٥	40	40
٦	40	42
٧	8	8
٨	24	22

احسب:

١- χ^2 فى حالة غياب نسب فرضية

٢- χ^2 إذا كانت النسبة الجنسية 1 : 1 واختبر تجانس البيانات

٣- ناقش النتائج المتحصل عليها فى ١- ، ٢-

٧-٤ إذا كان عدد الذكور إلى الإناث فى كليتين جامعيتين A و B كالمبين فى الجدول التالى

	الكلية الجامعية	
	A	B
عدد الذكور	300	122
عدد الإناث	40	500

فهل نسبة الذكور إلى الإناث متساوية فى الكليتين ؟

٧-٥ A ، B طريقتان لحفظ السائل المنوى، لقحت 40 بقرة بواسطة السائل المنوى المحفوظ بالطريقة A فأخصبت 26 بقرة بينما لقحت 60 بقرة بالسائل المنوى المحفوظ بالطريقة الثانية فأخصبت 54 بقرة. هل هناك تأثير لطريقة الحفظ على الخصوبة؟ اختبر ذلك إحصائياً مع بيان الفروض المختبرة والبديلة.

٨ المقارنة بين متوسطين

Comparison between two means

١- طريقة العينتين المستقلتين

٢- طريقة العينة المزدوجة

٣- طبيعة الفرق بين الطريقتين

فى كثير من التجارب يكون الفرق بين تأثير المعاملات وليس أثر المعاملات نفسها هو الأكثر أهمية بالنسبة للمجرب. فمثلاً إذا كان هناك دواء معين يقلل من الإصابة بمرض ما أو يؤدى إلى إحداث تغيير فسيولوجى معين كخفض لمستوى مادة معينة فى الدم أو لزيادة معدل الاستفادة من الغذاء ... الخ؛ فإن هناك طريقتين لإجراء مثل هذه المقارنات.

٨-١ طريقة العينتين المستقلتين Independent samples method

وتستخدم هذه الطريقة عندما يراد مقارنة الفرق بين متوسطى عشيرتين (معاملتين) كل منهما قد تم الحصول عليها بطريقة عشوائية من داخل العشيرة كما يحدث عند مقارنة نظامين لتغذية العجول حديثة الولادة أو لتسمين الحملان بعد فطامها. وهذا النوع من التجريب هو الأكثر استخداما فى التجارب حيث لا يوجد أساس لوجود تشابه بين فردين من أفراد العشيرة أكثر من بقية الأفراد (كما سيتضح فى الفصل التالى). وفى هذه الحالة فإن التجربة تتم عن طريق ما يعرف بمقارنة المجموعتين group comparison حيث توزع المعاملتان عشوائيا على أفراد العينة. ثم يقارن ويختبر الفرق فى النهاية بين متوسطى المجموعتين.

٨-٢ طريقة العينة المزدوجة Paired sample method

وفى هذا النوع من العينات تختار أزواج من الأفراد المتشابهة فى كثير من الجوانب مثال ذلك التوائم من نفس البطن والجنس فى الأرانب أو الأبقار أو نصفى ورقة أو ورقتان فى نبات واحد أو أصيصان فى نفس الموقع ... الخ.

وفى هذا المجال فإن أحد أفراد كل زوج يتلقى عشوائياً أحد المعاملتين ويتلقى الفرد الآخر المعاملة الأخرى.

وفى بعض الحالات التى تستخدم فيها هذه العينات بكثرة يستخدم نفس الفرد فى مرحلتين مختلفتين مثل قياس ضغط الدم للفرد قبل وبعد إعطائه دواء معيناً self-pairing والفرق بين القيم على أفراد الزوج الواحد يعبر عن الفرق بين المعاملات، وتؤخذ هذه الفروق على أنها المشاهدات فى التجربة، وتكون مجموعة الفروق بين أفراد الأزواج المختلفة مجموعة البيانات التى يتم تحليلها لاستخلاص المقارنة المطلوبة بين المعاملتين. وبذلك يكون عدد المشاهدات فى التجربة مساويا لعدد الفروق أى n وليس بعدد الأفراد كما هو الحال فى النوع الأول.

٨-٣ طبيعة الفرق بين الطريقتين

إن الفرق بين الطريقتين يتوقف على طبيعة التباين (V) للفرق بين متغيرين. فالتباين للفرق بين المتغيرين X_1 ، X_2 يتبع المعادلة العامة التالية:

$$V(X_1 - X_2) = V(X_1) + V(X_2) - 2\,Cov(X_1, X_2) \qquad (٨-١)$$

ففى حالة كون القيم مستقلة عن بعضها البعض أى أن X_1 ، X_2 متغيران عشوائيان. أى فى حالة العينات المستقلة فإن تباين الفرق يساوى مجموع التباين لكل من X_1 ، X_2 حيث إنه يمكن إثبات هذه العلاقة كالتالى:

$$V(X_1 - X_2) = E[(X_1 - X_2) - (\mu_1 - \mu_2)]^2 \qquad (٨-٢)$$

الرمز E فى المعادلة (٨-٢) هو اختصار لكلمة القيمة المتوقعة expected value وهى من الناحية الرياضية تعنى القيمة المتوقعة للكمية التى بين الأقواس، وبما أن القيمة المتوقعة لأى متغير هى متوسط العشيرة المأخوذ منها هذا المتغير، فإن E تعنى متوسط الكمية داخل القوس.

المعادلة (٨-٢) يمكن إعادة كتابتها كالتالى:

$$\begin{aligned} V(X_1 - X_2) &= E[(X_1 - X_2) - (\mu_1 - \mu_2)]^2 \\ &= E[(X_1 - \mu_1) - (X_2 - \mu_2]^2 \\ &= E(X_1 - \mu_1)^2 + E(X_2 - \mu_2)^2 - 2E[(X_1 - \mu_1)(X_2 - \mu_2)] \end{aligned}$$

ومن التوقعات الرياضية فإن القيمة المتوقعة للحدين الأول والثانى من الجانب الأيمن للمعادلة السابقة عبارة عن:

$$E(X_2 - \mu_2)^2 = \sigma_2^2 \ ، \ E(X_1 - \mu_1)^2 = \sigma_1^2$$

أما الحد الثالث فهو عبارة عن:

$$E(X_1 - \mu_1)(X_2 - \mu_2) = Cov(X_1, X_2)$$

حيث Cov هى تعبير عن التغاير covariance بين المتغيرين X_1 ، X_2. وفى حالة كون X_1 ، X_2 مأخوذين عشوائياً من العشيرة، فإن الحد الثالث أى التغاير يصبح مساوياً للصفر ويصبح تباين الفرق بين X_1 ، X_2 يساوى مجموع التباين لكليهما. أى:

$$V(X_1 - X_2) = V(X_1) + V(X_2) \qquad (٣-٨)$$

فإذا كانا مأخوذين من نفس العشيرة (دون تأثير للمعاملة على التباين) فإنه فى هذه الحالة:

$$V(X_1 - X_2) = 2\sigma_X^2 \qquad (٤-٨)$$

وبتطبيق هذا المفهوم على الفرق بين متوسطين $(\overline{X}_1 - \overline{X}_2)$ وحيث إنه قد سبق توضيح أن تباين المتوسط للعينات التى عدد أفرادها n عبارة عن σ^2 / n فإنه ينتج أن تباين الفرق بين المتوسطين:

$$V(\overline{X}_1 - \overline{X}_2) = \frac{2\sigma^2}{n} \qquad (٥-٨)$$

وبناء على ما تقدم وفى حالة مقارنة الأزواج تكون قيمة $Cov(X_1, X_2)$ فى (٨-١) دائما موجبة وذلك لأنه كما ذكر سابقاً إذا كانت عملية الأزواج ناجحة فإن أفراد نفس الزوج تكون متشابهة، فلو كانت القيمة $(X_1 - \mu_1)$ موجبة فإن $(X_2 - \mu_2)$ أيضا تكون موجبة ويكون الناتج موجباً. أما إذا كانت قيمة $(X_1 - \mu_1)$ سالبة فإن قيمة $(X_2 - \mu_2)$ تتجه لتصبح سالبة أيضا، وبالتالى يصبح الناتج موجباً. ولذا فإن تباين الفرق بين متوسطين يكون أقل من مجموع التباين فى (٨-٣) حيث تصبح قيمة التغاير covariance موجبة ويكون نتيجة ما تقدم أن يقل حجم تباين متوسط الفروق عن تباين الفرق بين متوسطين مما يجعل التجربة أكثر حساسية فى كشف الفروق بين تأثير المعاملات.

مثال ٨-١

البيانات التالية تمثل الزيادة فى الوزن بالجرام بعد نهاية التجربة فى 10 أزواج من الأرانب مأخوذة من نفس البطن لمعاملتين غذائيتين مختلفتين فى نسبة البروتين فيهما:

رقم الزوج	عليقة ١	عليقة ٢	الفرق D
١	403	234	169
٢	117	130	-13
٣	234	182	52
٤	221	143	78
٥	91	78	13
٦	130	65	65
٧	260	221	39
٨	195	195	0
٩	104	91	13
١٠	156	143	13
المجموع	**1911**	**1482**	**429**
المتوسط	**191.1**	**148.2**	**42.9**

الحل:

$$H_1 : \mu_d \neq 0 \quad ، \quad H_\circ : \mu_d = 0$$

عدد الفروق: 10، مجموع مربع الفروق: $\sum D^2 = 43771$

مجموع مربع الفروق المصحح: $\sum d^2 = 43771 - \dfrac{(429)^2}{10} = 25366.9$

تباين الفروق: $S_d^2 = \dfrac{\sum d^2}{n-1} = \dfrac{25366.9}{9} = 2818.54$

الخطأ القياسى لمتوسط الفروق: $S_{\bar{d}} = \sqrt{S_d^2/n} = \sqrt{2818.54/10} = 16.79\ g$

وبالتالى فإنه لاختبار معنوية هذه الفروق فيما إذا كانت تختلف عن قيمة الصفر أى أن فرض العدم هو: $H_\circ : \mu_d = 0$ وعليه فإن الاختبار للمعنوية يأخذ الصورة العامة وهى:

$$t = \frac{\bar{d} - \mu_d}{S_{\bar{d}}} \qquad (٨-٦)$$

وبالتالى تكون قيمة t المحسوبة: $t = \dfrac{42.9 - 0}{16.79} = 2.56$

مع تحيات د. سلام حسين الهلالي salamalhelali@yahoo.com

قيمة t الجدولية (جدول ٤ ملحق أ) عند درجات حرية 9 ومستوى معنوية 0.05 هى $t_{(9,.05)} = 2.26$ ، وبالتالى يرفض فرض العدم، أى أن هناك فرقاً معنوياً بين العليقتين فى تأثيريهما على الزيادة فى الوزن.

ماذا يمكن أن يتحقق من زيادة فى حساسية التجربة عند اتباع طريقة الأزواج مقارنة بطريقة العينات المستقلة؟ من الواضح أنه يمكن الحصول على معلومات توضح مقارنة بين الطريقتين. فإذا فرض أنه فى جدول البيانات تم تحليل كل من الأفراد المنتمية إلى كل من المعاملتين على أساس أن كلا من العينتين تمثل أفراداً مستقلة لا ترتبط مع بعضها بأى تشابه، وبالتالى فإن تحليل التجربة سوف يتم على أساس حساب التباين المجمع pooled variance من العينتين بصفة مستقلة ثم اختبار الفرق بين متوسطى المعاملتين حسب فرض العدم بتساوى متوسطى المعاملتين وليس عدم وجود اختلاف لمتوسط الفروق عن الصفر كما فى الحالة الأولى ويكون التحليل بالتالى كما يلى:

العينة الأولى (العليقة ١)

المجموع: $\sum X_1 = 1911$ ، مجموع المربعات غير المصحح $\sum X_1^2 = 445653$

مجموع المربعات المصحح: $\sum x_1^2 = 445653 - \frac{(1911)^2}{10} = 80460.9$

التباين $S_1^2 = 80460.9/9 = 8940.1$

العينة الثانية (العليقة ٢)

المجموع: $\sum X_2 = 1482$ ، مجموع المربعات غير المصحح $\sum X_2^2 = 251134$

مجموع المربعات المصحح: $\sum x_2^2 = 251134 - \frac{(1482)^2}{10} = 31501.6$

التباين $S_1^2 = 31501.6/9 = 3500.18$

ولكن يمكن الحصول على قيمة واحدة للتباين تمثل التباين المرجح بدرجات الحرية ولكن يلزم أن يجرى أولاً اختبار التجانس test of homogeneity باستخدام اختبار F بأن يقسم التباين الأكبر على التباين الأصغر وتقارن القيمة الناتجة بقيمة F الجدولية (جدول ٥ ملحق أ) بدرجات حرية كل من البسط والمقام، فإذا كانت القيمة المحسوبة أقل من الجدولية فهذا يعنى أن التباينين متجانسان و عليه فإن اختبار التجانس:

$$F = 8940.1/3500.18 = 2.55$$

قيمة F الجدولية عند درجة حرية 9 لكل من البسط والمقام ومستوى معنوية 5% تساوى 3.18 وهى أكبر من F المحسوبة وبالتالى فإن التباينين متجانسان ويحسب التباين المجمع pooled variance كالتالى:

$$S_P^2 = \frac{S_1^2(n_1-1)+S_2^2(n_2-1)}{(n_1-1)+(n_2-1)}$$

$$= \frac{\sum x_1^2 + \sum x_2^2}{n_1+n_2-2} = \frac{80460.9+31501.6}{18} = 6220.14$$

وتباين الفرق بين متوسطين:

$$S^2_{(\overline{X}_1-\overline{X}_2)} = S_P^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right) = 6220.14\left(\frac{1}{10}+\frac{1}{10}\right) = \frac{2(6220.14)}{10}$$

والخطأ القياسى للفرق بين المتوسطين: $S_{(\bar{x}_1-\bar{x}_2)} = \sqrt{\frac{2(6220.14)}{10}} = 35.27$

وبما أن فرض العدم بتساوى المعاملتين يكون فى هذه الحالة: $H_\circ : \mu_1 - \mu_2 = 0$ والفرض البديل $H_1 : \mu_1 - \mu_2 \neq 0$ فإنه لاختباره بواسطة اختبار t تكون صورته كالتالى:

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_{(\bar{x}_1-\bar{x}_2)}} \qquad (٨-٧)$$

وبالتالى تكون t المحسوبة: $t = \frac{191.1-148.2}{35.27} = 1.216$

وبالكشف عن معنوية هذا الفرق عند مستوى 0.05 ودرجات حرية $18=(n_1+n_2-2)$، فإن قيمة t الجدولية 2.101، وبالتالى يصبح الفرق غير معنوى، أى أنه لا يمكن رفض فرض العدم بتساوى متوسطى المعاملتين فى تأثيرهما على الزيادة فى الوزن.

مثال ٨-٢

استخدام برنامج SAS لإجراء اختبار t فى أزواج للبيانات المذكورة فى مثال ٨-١

```
DATA PAIRED;
INPUT RATION1 RATION2 @@;
DIFF=RATION1-RATION2;
CARDS;
403 234 117 130 234 182 221 143 91 78
130 65 260 221 195 195 104 91 156 143
PROC MEANS N MEAN STD STDERR T PRT;
VAR RATION1 RATION2 DIFF;
TITLE 't-test in paired comparison';
Run;
```

نتيجة التحليل:

t-test in paired comparison

The MEANS Procedure

Variable	N	Mean	Std Dev	Std Error	t Value	Pr > |t|
RATION1	10	191.1000000	94.5521020	29.9000000	6.39	0.0001
RATION2	10	148.2000000	59.1623003	18.7087621	7.92	<.0001
DIFF	10	42.9000000	53.0899656	16.7885212	2.56	0.0309

مثال ٨-٣

استخدام برنامج SAS لإجراء اختبار t فى مجاميع للبيانات المذكورة فى مثال ٨-١

```
DATA GROUPS;
INPUT TRT GAIN @@;
CARDS;
1 403 1 117 1 234 1 221 1 91 1 130 1 260 1 195 1 104 1 156
2 234 2 130 2 182 2 143 2 78 2 65 2 221 2 195 2 91 2 143
PROC TTEST COCHRAN;
CLASS TRT;
VAR GAIN;
TITLE 't-test in GROUPS';
Run;
```

لاحظ أن اختيار COCHRAN يستخدم عند عدم تجانس التباين وذلك لتقدير قيمة مستوى المعنوية لاختبار t المقرب (Steel and Torrie, 1980).

نتيجة التحليل:

t-test in GROUPS

The TTEST Procedure

Variable	Method	Variances	DF	t Value	Pr > |t|
GAIN	Pooled	Equal	18	1.22	0.2396
GAIN	Satterthwaite	Unequal	15.1	1.22	0.2425
GAIN	Cochran	Unequal	9	1.22	0.2548

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
GAIN	Folded F	9	9	2.55	0.1786

مما سبق يتضح أن إجراء عملية الإزواج قد أدى إلى تخفيض فى حجم الخطأ القياسى لاختبار الفرق بين المعاملتين ولقد نتج هذا عن استخدام أفراد متشابهة مع بعضها مبدئياً مما قلل من الاختلافات داخل الأزواج أى أن معامل الارتباط فيما بينهما مرتفعاً.

ويمكن القول إن تجارب الأزواج فى حالة نقص الاختلافات داخل الأزواج عن الاختلافات بين الأفراد العشوائية تؤدى إلى زيادة فى كفاءة التجربة للإحساس بالفروق بين المعاملات، أو بمعنى آخر أنه باستخدام العينات المستقلة كان يستوجب للحصول على نفس القدر من الكفاءة (مقدرة على أساس الخطأ القياسى للفرق بين المتوسطين) استخدام عدد أكبر من المشاهدات فى التجربة. ولكن فى نفس الوقت فإن درجات الحرية المستخدمة فى نظام العينات المستقلة يكون ضعف درجات الحرية المتاحة للخطأ فى حالة الأزواج، وأيضا قد يؤدى استخدام أزواج المشاهدات إلى الحصول على تباين أكبر للفرق بين المتوسطين إذا ما كان الارتباط موجباً بين أزواج المشاهدات التى تستخدم فى التجريب كما وأن إهمال الأزواج وإجراء التجارب كما لو كانت مستقلة مع وجود الارتباط يؤدى إلى فقد معلومات قيمة وزيادة فى الأعباء كان يمكن توفيرها.

صندوق ٨-١

يستخدم اختبار t بطريقة العينتين المستقلتين إذا لم يكن التشابه بين الفرد في معاملة وفرد في المعاملة الأخرى أكثر من التشابه بين الأفراد في العينة عموماً.

بينما يستخدم اختبار t بطريقة الأزواج إذا كان التشابه بين فرد في معامله وفرد في المعاملة الأخرى أكبر من التشابه بين الأفراد عموماً.

وبالإشارة إلى اختبار التشابه السابق ذكره إذا كانت F المحسوبة معنوية عند مستوى المعنوية المتعارف عليه فى مجالات الزراعة المختلفة وهو 5% فإنه بالتالى لا يمكن استخراج قيمة متجانسة للتباين ولكن تكون خطوات التحليل كما يلى:

بعد حساب المتوسط والتباين لكل من المجموعتين واختبار التجانس والذى أدى إلى رفض الفرض القائل بأن التباين متجانس فإن:

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \qquad (٨-٨)$$

تقارن قيمة t المحسوبة هذه بقيمة t المعدلة والتى تستخرج من جدول t (جدول ٤ ملحق أ) بعد حساب قيمة جديدة لدرجات الحرية تسمى effective df وهى مأخوذة عن Satterthwaite (1946) حيث:

$$\text{effective df} = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{(S_1^2/n_1)^2/(n_1-1) + (S_2^2/n_2)^2/(n_2-1)} \qquad (٩-٨)$$

ثم يستكمل التحليل الإحصائى كما سبق.

تمارين الباب الثامن

٨-١ التالى هى درجات 8 طلاب التى حصلوا عليها فى مادتين مختلفتين، هل درجات الطلاب متساوية للمادتين ؟

رقم الطالب	المادة الأولى	المادة الثانية
١	42	84
٢	41	28
٣	13	34
٤	99	69
٥	80	38
٦	88	28
٧	30	5
٨	100	80

٨-٢ لمقارنة الدرجات التى تحصل عليها الطلاب فى مادتين مختلفتين، أخذ 10 طلاب بطريقة عشوائية وأعطوا اختبار المادة الأولى، كما أخذ 10 طلاب آخرين وأعطوا اختبار المادة الثانية. وكانت درجات الطلاب كما يلى

المادة الأولى		المادة الثانية	
رقم الطالب	الدرجة	رقم الطالب	الدرجة
١	75	١	34
٢	35	٢	47
٣	73	٣	99
٤	49	٤	92
٥	50	٥	20
٦	100	٦	78
٧	41	٧	84
٨	17	٨	27
٩	94	٩	80
١٠	56	١٠	87

هل الدرجات متساوية فى المادتين ؟

٨-٣ البيانات التالية تمثل الزيادة الكلية في الوزن بالكيلوجرام فى ذكور الحملان من الولادة التوأمية والمغذاة على عليقتين أ ، ب

رقم زوج التوائم	عليقة أ	عليقة ب
١	14.0	14.4
٢	15.2	13.7
٣	15.7	14.7
٤	14.2	12.8
٥	14.8	14.2
٦	14.9	14.3
٧	14.1	13.5
٨	13.0	13.7
٩	14.6	13.1
١٠	16.8	11.9

والمطلوب اختبار فرض العدم بتساوى كمية الزيادة الكلية فى الوزن فى العليقتين.

٨-٤ اختبر فرض العدم بأنه لا يوجد فروق بين متوسطى العليقتين فى التمرين السابق بغض النظر عن العلاقة بين فردى نفس الزوج ثم قارن النتائج المتحصل عليها مع نتائج التمرين السابق.

٩ تحليل التباين أحادى الاتجاه

One way analysis of variance

١- مقدمة

٢- المتغير والعامل والمعاملة والمستوى

٣- افتراضات نموذج تحليل التباين

٤- ماذا يراد عادة من تحليل التباين ؟

٥- مفهوم وكيفية إجراء تحليل التباين

٦- النموذج الرياضى

٧- اختبار جميع الفروق بين أزواج المتوسطات

٨- المقارنات المستقلة (المتعامدة)

٩- تحليل التباين عند عدم تساوى المكررات

١٠- النموذج التجميعى الخطى

١١- تقدير مكونات التباين فى حالة اختلاف حجم العينات

١٢- معامل الارتباط الداخلى (الجوانى)

١٣- العينات داخل العينات

٩-١ مقدمة

تنحصر أهم الأسس فى دراسة الموضوعات ذات الطابع العلمى فى دراسة الاختلافات بين الظواهر المختلفة. فبدون الاختلافات أو التباينات لا توجد طريقة لدارسة العلاقة بين المتغيرات، أى أن دراسة التباين variance فى البيانات يعطى دلالة على مدى تشتتها وفى نفس الوقت أيضاً يوضح مدى اتساق هذه البيانات أو تجانسها homogenous من عدم اتساقها hetcrogeneous.

٩-٢ المتغير والعامل والمعاملة والمستوى
Variable, factor, treatment, level

فى الواقع المتغير والعامل والمعاملة والمستوى هى ألفاظ ذات معانى متداخلة، فالمتغير هو المشاهدة التى قد تأخذ قيماً مختلفة. وقد يكون هذا المتغير مستقلاًindependent variable، أى لا يفترض أنه يتأثر بمتغير آخر فى نطاق التجربة. أو قد يكون المتغير تابعاً dependent variable، أى يتأثر بمتغير آخر فى نطاق التجربة، وقد يطلق عليه أحياناً متغير الاستجابة response variable.

أما العامل فهو كثيراً ما يستخدم بمفهومه العام مثل قول العوامل المؤثرة على الإنتاج، العوامل المؤثرة على خصوبة الحيوان ... الخ. ولكن فى المجال الإحصائى فالعامل يقصد به مؤثر (أى متغير مستقل) له عدة مستويات levels، مثال ذلك إذا وضع جنس الحيوان فى نموذج model حيث المتغير التابع هو النمو والمتغير المستقل هو الجنس، والجنس فى هذه الحالة يعتبر عاملاً يأخذ عدة مستويات: ذكر، أنثى، مخصى ...الخ. وقد يكون العامل متغيراً يرغب المجرب فى دراسة أثره على متغير تابع مثل دراسة تأثير مستويات مختلفة من البروتين فى العليقة على نمو الحيوان. وقد يرغب المجرب فى دراسة أثر أكثر من عامل فمثلاً أثر البروتين (P) وله ثلاثة مستويات وأثر إضافة فيتامين معين (V) وله مستويين. فى هذه الحالة هناك عاملان هما البروتين والفيتامين وتجربة كل من مستويات البروتين مع كل من مستويات الفيتامين يطلق عليه تجارب عاملية factorial experiments. والتجربة فى الواقع عبارة عن 6 معاملات هى $P_1V_1, P_1V_2, P_2V_1, P_2V_2, P_3V_1, P_3V_2$ أو معاملة واحدة لها 6 مستويات. إذن المعاملة هى عامل له مستويات ولكن ليس كل عامل معاملة حيث ممكن أن يكون العامل هو الموقع أو العمر أو الحظيرة، وإن كانت هذه كلها يمكن جعلها عاملاً إذا قصدت مباشرة بالدراسة ولها مستوياتها المقصودة بالبحث. إذا يمكن القول أن كل عامل أو معاملة متغير ولكن ليس كل متغير عاملاً أو معاملة. وسوف تستخدم هذه التعبيرات تبادليا فى هذا المؤلف.

هذا وقد سبقت الإشارة فى الباب الثامن إلى استخدام التباين أو مشتقاته فى اختبار الفرض الخاص بتساوى متوسطى عشيرتين. حيث تم استخدام اختبار فرض العدم بواسطة اختبار(t). ولكن فى كثير من الحالات يفترض وجود أكثر من عشيرتين ويراد اختبار الفرض الخاص بالفروق فيما بينها. فمثلا قد يكون الاهتمام مركزاً فى اختبار اختلاف مستويات مكون ما من العليقة على نمو الحيوانات. فإذا افترض وجود أكثر من عشيرتين، فإنه يمكن إجراء سلسلة من اختبارات (t) فى كل منها يتم اختبار الفرق بين متوسط عشيرة ما والمتوسطات الأخرى. فعلى سبيل المثال لو كانت هناك ثلاثة متوسطات (ثلاثة عشائر مفترضة) فسوف يتم عمل ثلاث اختبارات، ولو كان هناك 4 متوسطات يتم عمل ستة اختبارات للفروق، ... وهكذا تتزايد هذه الاختبارات بزيادة عدد المتوسطات المختبرة. وفى نفس الوقت فإن كل اختبار (t) سوف يكون مرتبطاً بمستوى قدره α لاحتمال حدوث خطأ من النوع الأول type I error، وعلى هذا الأساس فإنه إذا أجرى العديد من هذه الاختبارات فإنه من المتوقع الوصول إلى قرار خاطئ فى بعض من هذه الاختبارات متوقفاً على درجات الاحتمال.

وتحليل البيانات بطريقة تحليل التباين يعتبر امتداداً لاختبار(t) والخاص باختبار فرض العدم عن المتوسطات. وهذه الطريقة تسمح باختبار واحد وباحتمال خطأ واحد حجمه α أن يجرى اختبارات تجيب على الأسئلة الخاصة بالمتوسطات واختلافها، كأن يكون السؤال عما إذا كانت البيانات توضح أن العشائر المفترضة تختلف فيما بينها، وأيضاً إذا كانت هذه الاختلافات تعتبر معنوية ولا ترجع إلى الصدفة.

ويتلخص الغرض من تحليل التباين فى تقسيم التباين الكلى بين البيانات إلى مكونات ذات مغزى والتى تقيس مصادر الاختلافات.

ففى حالة اختبار الفروق بين المعاملات الموجودة فى العليقة وتأثيرها على اختلاف النمو فى الأفراد، فإن تحليل التباين سوف يقوم بتقسيم التباين الكلى إلى جزء يرجع إلى الخطأ التجريبى experimental error والجزء الثانى يرجع إلى الخطأ التجريبى بالإضافة إلى أى تباين يرجع إلى اختلاف مستويات العامل فى العليقة. فإذا كان فرض العدم صحيحاً، أى أنه توجد اختلافات بين معدلات النمو على المستويات المختلفة للعليقة، فإن كلا من الجزئين المقسم إليهما التباين يعطى تقديراً مستقلا للخطأ التجريبى. وبالتالى فإن الاختبار يقوم على أساس مقارنة لكل من مصدرى التباين أو الاختلاف، وأن مستوى معنوية هذا الاختلاف سوف يوضح عن طريق نسبة التقديرين للتباين لبعضهما، وهذه تتبع توزيعاً يعرف بتوزيع ف F distribution.

الأساس الرياضى لتحليل التباين قدمه أحد أئمة علم الإحصاء وهو العالم الإنجليزى R. A. Fisher، ولقد قام العالم الأمريكى G. W. Snedecor بدراسة

الناحية الرياضية للنسبة بين تقديرات التباين وتوزيعها وسمى بتوزيع F نسبة إلى اسم Fisher.

٩-٣ افتراضات نموذج تحليل التباين Assumption in analysis of variance

إذا أريد دراسة أثر الزيادة فى كمية البروتين فى العليقة والتى ستكون مصحوبة بنمو أسرع، فإن فرض العدم هنا هو أن متوسط النمو فى الحيوانات التى لم تتلق المادة فى المعاملة تتساوى مع تلك التى حصلت على المادة، وبدرجات متزايدة، أى أن جميع المتوسطات للمجموعات التى تلقت المعاملات المختلفة متساوية.

يعتبر تحليل التباين أساساً اختباراً للمعنوية، حيث يكون المطلوب، بناء على العينة، هو تقدير تواجد من عدم تواجد علاقة بين المعاملة والاستجابة فى العشائر المختلفة. وهناك بعض الافتراضات فى عملية أخذ العينة تعتبر أساسية حتى يصبح تحليل التباين صحيحاً من ناحية الاستقراءات التى سيتم الوصول إليها من العينة. وهذه الافتراضات هى:

أولا: أن الأفراد فى العينة يجب أن تكون مستقلة independent وعشوائية random فى اختيارها. وهذا يعنى أن هناك استقلالية فى كون أى فرد يدخل ضمن العينة وهذا أيضاً يوضح أنه ليست هناك علاقة بين اختيار فرد ما فى العينة واختيار فرد آخر. ومن خصائص العينات العشوائية ليس فقط أن الأفراد لها نفس الفرصة فى الظهور فى العينة ولكن أيضاً إعطاء أى توليفة من الأفراد نفس الفرصة للاختيار.

ثانياً: تجانس التوزيع أو الانتشار وهو ما يعرف homoscedasticity بمعنى آخر أن العشائر المختلفة المأخوذ منها العينة لها نفس التباين common variance. وهذا يعنى أنه، حتى ولو وجدت اختلافات بين المتوسطات فى العشيرة المؤثر عليه بالمعاملات فإن التشتت أو التباين داخلها لم يتغير وأن تباينات القيم غير مرتبطة مع بعضها.

ثالثاً: حيث إن تحليل التباين يفترض إجراء اختبارات معنوية وحتى يمكن إجراء مثل هذه الاختبارات فإنه يفترض أن الأفراد والمشاهدات المأخوذة من عشيرة واحدة تختلف عن بعضها اختلافاً طبيعياً (الأخطاء) وأن هذه الاختلافات مأخوذة من عشيرة تتبع التوزيع الطبيعى normal distribution بمتوسط صفر وتباين قدره σ^2.

وعموماً فإن هذه الافتراضات إذا لم تتحقق violated بوضوح فإن تحليل التباين وبالتالى اختبار المعنوية سوف يؤدى إلى الوصول إلى نتائج أو قرارات متحيزة

ويصبح هناك عدم ثقة فى الاستقراءات المأخوذة لتطبيقها على العشيرة حيث يمثل الهدف الأساسى من استخدام العينة وتحليل نتائجها. وسوف تتم مناقشة أهم هذه الافتراضات وطرق معالجتها فى حالة عدم تحققها فى الباب الثانى عشر.

٩-٤ ماذا يراد عادة من تحليل التباين؟

عند أخذ الحالة السابقة والخاصة بدراسة تأثير إضافة مادة ما إلى العليقة على نمو الأفراد، فإذا كان هناك ثلاثة مستويات لهذه المادة وتم توزيع أفراد العينة عشوائيا على المستويات الثلاثة فإن ذلك سوف يؤدى إلى وجود ثلاثة متوسطات يمثل كل منها عشيرة تتلقى نوعاً من المعاملة، والمطلوب معرفة ما إذا كانت هذه العشائر تختلف عن بعضها علماً بإنه من المفترض أن تأثير المعاملة سوف لا يؤثر على التباينات بين مجموعة الأفراد التى تتلقى نفس المعاملة. وبالتالى يوجد ثلاثة متوسطات يرمز لها بالرمز $\overline{Y}_i$ حيث i تمثل مستوى المعاملة $i = 1, 2, 3$.

وهناك ثلاث حالات يمكن تصويرها بيانياً فى الشكل ٩-١ (أ، ب، ج) والذى يبين العلاقة الممكن تصورها بين مجموعة من العينات أو المواقف والعشيرة المأخوذ منها هذه العينات.

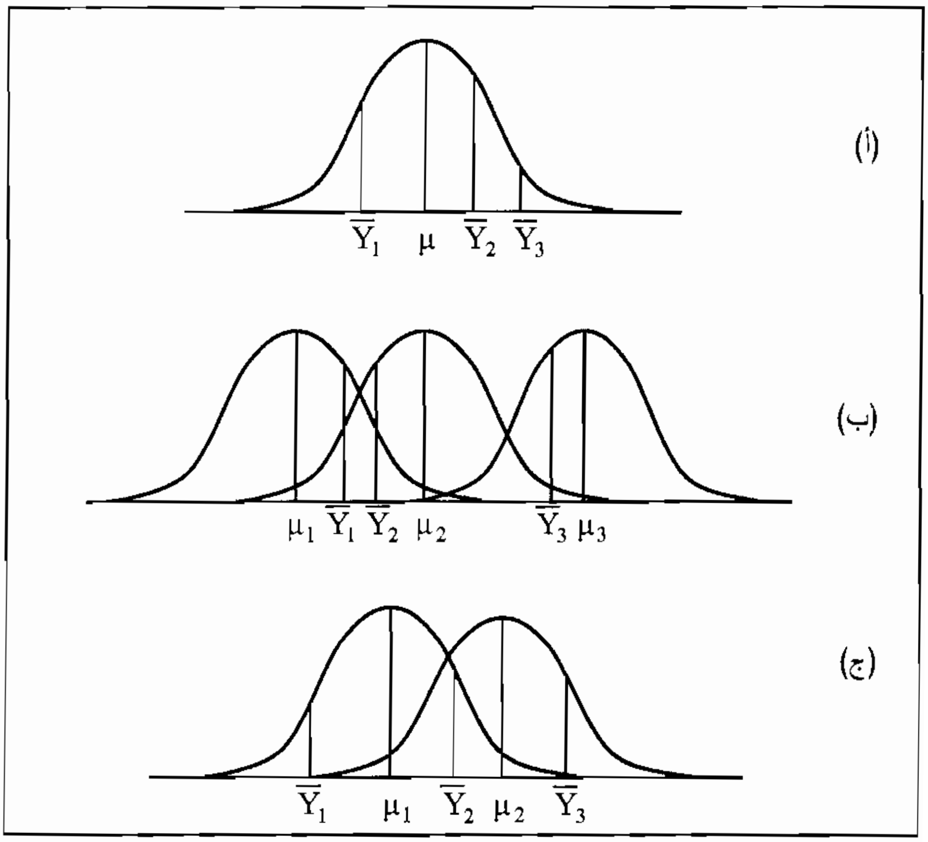
الحالة (أ): يوجد توزيع طبيعى واحد وهذا يمثل عشيرة واحدة. أما بالنسبة للقيم المختلفة الموضحة للمتوسط $\overline{Y}_i$ فإنها تمثل متوسطات العينات أو المعاملات الثلاثة والتى تعتبر مختلفة فقط للأسباب الطبيعية لاختلاف نقطة التقدير point estimate والتى تحسب من كل عينة لتمثيل متوسط العشيرة. أى أنه ليس هناك سبب للقول بأن هذه النقاط تعتبر مختلفة عن بعضها ولكنها جميعاً مأخوذة من نفس العشيرة والتى متوسطها μ.

الحالة (ب): فهى مختلفة حيث إن هناك ثلاث عشائر كل منها ممثل بتوزيع طبيعى وقيم $\overline{Y}_3$ ، $\overline{Y}_2$ ، $\overline{Y}_1$ تعتبر نقاطاً تقديرية لمتوسط الثلاثة عشائر المختلفة μ_3 ، μ_2 ، μ_1. وبما أن الثلاثة متوسطات للعينات (المعاملات) مختلفة عن بعضها فإن هذا يعنى وجود اختلافات بين متوسطات العشائر الثلاثة.

الحالة (ج): هناك متوسطان لعينتين تمثلان متوسطاً لعشيرة واحدة μ_1 فى حين أن متوسط المعاملة الثالثة $\overline{Y}_3$ فإنه يعتبر تقديراً لمتوسط العشيرة μ_2 وهو يختلف عن الأول كنتيجة لتأثير المعاملة.

ومن نفس المنطلق فإنه يمكن توسيع مدى فرض العدم ليشمل أى عدد من المتوسطات المفترضة أو بالتالى أى عدد من مستويات العامل. وفى كل هذه الحالات فإن الفرض المقابل هو أنه ليست كل المتوسطات متساوية. وعلى هذا الأساس يتم

رفض فرض العدم فى حالة اختلاف ولو واحد فقط من متوسطات العشيرة عن بقية القيم. وهذا هو أساس تحليل التباين وبالتالى اختبار المعنوية.



شكل ٩-١ العلاقة الممكن تواجدها بين مجموعة من العينات والعشيرة المأخوذ منها العينات

ويمكن تحليل التباين من استخدام كل البيانات الموجودة فى التجربة (الدراسة) دون النظر إلى المسبب، لتقسيم التباين الكلى إلى تأثير العامل المختبر (المعاملات) والخطأ التجريبى. ثم بعد ذلك مقارنة مصادر التباين بواسطة اختبار F.

٩-٥ مفهوم وكيفية إجراء تحليل التباين

سوف يستخدم المثال العددى التالى لتوضيح مفهوم وكيفية إجراء تحليل التباين والبيانات تمثل 4 معاملات مختلفة لحفظ مادة معينة (قد تكون المعاملات الأربعة

عبارة عن 4 مواد حفظ مختلفة أو 4 مستويات مختلفة من نفس مادة الحفظ) وتمثل الأرقام عدد الساعات التى تبقى فيها المادة صالحة قبل أن تفسد:

	مادة الحفظ			
المكررات	أ	ب	ج	د
١	7	5	3	10
٢	8	3	4	7
٣	7	3	3	9
المجموع	**22**	**11**	**10**	**26**
المتوسط	**7.33**	**3.67**	**3.33**	**8.67**

فإذا افترض أنه لا توجد اختلافات بين الأربعة معاملات فإن 12 مشاهدة تتوزع حول متوسط واحد μ وتباين مشترك فيما بينها σ^2.

وكما سبق فإنه يمكن من خلال تحليل التباين الحصول على أكثر من تقدير للتباين فى العشيرة. وبالتالى يوجد ثلاثة تقديرات يمكن الحصول عليها لتقدير التباين σ^2.

أولاً: إذا افترض أن فرض العدم صحيح، فإن هذا يعنى أن كل البيانات الموجودة فى التجربة مأخوذة من نفس العشيرة. وبالتالى يمكن الحصول على تقدير للتباين الكلى من مجموع المربعات الكلى المصحح وهى:

$$\sum y^2 = \sum Y_i^2 - \frac{(\sum Y)^2}{n} = (7^2 + 8^2 + \cdots + 9^2) - \frac{(69)^2}{12}$$
$$= 469 - 396.75 = 72.25$$

ومجموع المربعات المصحح هذا له 11 درجة حرية. وبالتالى يصبح متوسط المربعات $72.25/11 = 6.568$ وهو أول تقدير للتباين σ^2.

ثانياً: التقدير الثانى للتباين فى العشيرة σ^2 يمكن الحصول عليه عن طريق حساب مجموع المربعات المصحح لكل من الأربعة معاملات بنفس الطريقة التى سبق شرحها فى الباب الثامن وحساب ما سبق وأطلق علية التباين المجمع pooled variance، وبالتالى:

$$S_p^2 = \sum y_1^2 + \sum y_2^2 + \sum y_3^2 + \sum y_4^2 \qquad (١-٩)$$
$$= 0.67 + 2.67 + 0.67 + 4.67 = 8.68$$

عدد درجات حرية هذا المجموع هو $t(n-1)$ حيث t هى عدد المعاملات 4، n عدد الأفراد (المكررات داخل المعاملة)، أى أن درجات الحرية 8 درجات حرية. وقسمة مجموع المربعات هذا على درجات الحرية يعطى تقدير $\sigma^2 = 1.085$ من داخل المعاملات. ويعرف هذا بالتباين داخل المعاملات within treatment mean square.

وطبعاً من الممكن ملاحظة أنه فى كل حالة تم حساب مجموع انحرافات أفراد العينة عن متوسطها. أى أنه فى هذه الحالة لا يتأثر هذا التقدير باختلاف متوسطات المعاملات أو عدم اختلافها. وبما أن هذه القيم مقاسة على الأفراد، أى كل وحدة تلقت المعاملة ككل فإن تقدير التباين من داخل المعاملات يعتبر قياساً للخطأ التجريبى للتجربة.

ثالثاً: أما التقدير الثالث والأخير فهو الذى يقيس الاختلافات بين المعاملات أو تأثيراتها ككل. إن متوسطات المعاملات يعتبر كل منها تقديرا لمتوسط العشيرة. وهذه المتوسطات لها تباين حول متوسط العشيرة يساوى σ^2/n حيث n هى عدد المشاهدات داخل كل معاملة (عينة). وبالتالى فإنه إذا تم حساب مجموع المربعات للانحرافات المأخوذة من متوسطات العينات وبضربه فى n يعطى مجموع مربعات الانحرافات بين هذه المتوسطات والمتوسط العام، وفى المثال فإن متوسطات المعاملات هى 7.33, 6.67, 3.33,8.67 وبالتالى فإن مجموع مربع الانحرافات يمكن حسابه كالتالى:

$$(3)\left[(7.33)^2+(3.67)^2+(3.33)^2+(8.67)^2-\frac{(23)^2}{4}\right]$$
$$=(3)(21.2056)=63.6168$$

ومجموع المربعات المحسوب من المتوسطات له 3 درجات حرية وبالتالى فإن تقدير التباين من المتوسطات يكون $63.6168/3 = 21.21$.

وعادة ما يتم حساب مجموع المربعات بين المعاملات من معادلة (٩-٤) والتى سيأتى ذكرها فيما بعد، وعلى ذلك تصبح العمليات الحسابية لحساب مجموع المربعات المصحح بين المعاملات كالتالى:

$$\frac{(22)^2+(11)^2+(10)^2+(26)^2}{3}-\frac{(69)^2}{12}$$
$$=\frac{1381}{3}-\frac{4761}{12}=63.583$$

وهى تساوى نفس القيمة المذكورة سابقاً باستثناء بعض أخطاء التقريب. وعموما فإن المعادلة (٩-٤) هى الأكثر دقة وذلك لعدم إجراء التقريب لكل متوسط عند حساب المربعات منها. كما أنه عادة، يتم حساب مجموع المربعات داخل المعاملات بطرح المربعات بين المعاملات من مجموع المربعات الكلى فى هذا النوع من تحليل التباين وذلك نتيجة لخاصية مهمة فى تحليل التباين وهى خاصية التجميع لمجاميع المربعات وأيضا لدرجات الحرية المقابلة concept of additivity of sums of squares and degrees of freedom (هذه الخاصية تظل ثابتة وصحيحة فى تحليل البيانات للتقسيم الأحادى وأيضا فى حالة التقسيمات الأخرى المتعددة، إذا كانت الأعداد فى الفئات إما متساوية أو متناسبة proportional، ولكن هذه الخاصية تفقد صلاحيتها إذا اختلفت التكرارات فى التقسيمات المتعددة مما يستوجب إجراء تعديلات على تحليل البيانات الخاصة بتحليل التباين).

وعلى هذا الأساس فإن مجموع المربعات داخل المعاملات هو مجموع المربعات الكلى مطروحاً منه مجموع المربعات بين المعاملات.

نظرية ٩-١

$$\sum_i \sum_j (Y_{ij} - \overline{Y}_{..})^2 = n\sum_i (\overline{Y}_{i.} - \overline{Y}_{..})^2 + \sum_i \sum_j (Y_{ij} - \overline{Y}_{i.})^2$$

إثبات النظرية .

بإضافة $\overline{Y}_{i.}$ مرة وطرحه مرة أخرى وإعادة الترتيب ينتج أن:

$$\begin{aligned}
\sum_i \sum_j (Y_{ij} - \overline{Y}_{..})^2 &= \sum_i \sum_j \left[(\overline{Y}_{i.} - \overline{Y}_{..}) + (Y_{ij} - \overline{Y}_{i.})\right]^2 \\
&= \sum_i \sum_j \left[(\overline{Y}_{i.} - \overline{Y}_{..})^2 + 2(\overline{Y}_{i.} - \overline{Y}_{..})(Y_{ij} - \overline{Y}_{i.}) + (Y_{ij} - \overline{Y}_{i.})^2\right] \\
&= \sum_i \sum_j (\overline{Y}_{i.} - \overline{Y}_{..})^2 + 2\sum_i \sum_j (\overline{Y}_{i.} - \overline{Y}_{..})(Y_{ij} - \overline{Y}_{i.}) + \sum_i \sum_j (Y_{ij} - \overline{Y}_{i.})^2
\end{aligned}$$

وبما أن مجموع الانحرافات على كل المدى يساوى صفر، فإن الحد الأوسط يصبح صفر حيث إن:

$$\sum_i (\overline{Y}_{i.} - \overline{Y}_{..}) = \sum_j (\overline{Y}_{ij} - \overline{Y}_{i.}) = 0$$

لاحظ أن الحد الأول لا يحتوى على أفراد العينة ولكنه عبارة عن مربع الفرق بين متوسط كل معاملة والمتوسط العام، وبالتالى يمكن كتابته على النحو التالى:

$$\sum_i \sum_j (\overline{Y}_{i.} - \overline{Y}_{..})^2 = n \sum_i (\overline{Y}_{i.} - \overline{Y}_{..})^2$$

وبالتالى فإن:

$$\sum_i \sum_j (Y_{ij} - \overline{Y}_{i.})^2 = \sum_i \sum_j (Y_{ij} - \overline{Y}_{..})^2 - n \sum_i (\overline{Y}_{i.} - \overline{Y}_{..})^2 \qquad (٩-٢)$$

حيث i هى رتبة المعاملات وعددها t فى حين أن j هو رتبة الأفراد أو المشاهدات داخل كل معاملة وعددها n. ففى النظرية (٩-١) نجد أن الحد الأيسر هو مجموع المربعات الكلية total sum of squares أما الجانب الأيمن فحده الأول هو مجموع المربعات بين المعاملات (الفئات) between classes sum of squares، والحد الثانى هو مجموع المربعات داخل المعاملات within classes sum of squares، وهذا الأخير يطلق عليه فى كثير من الحالات الخطأ error ويعنى الخطأ التجريبى experimental error.

٩-٦ النموذج الرياضى Mathematical model

لقد سبق الحديث عن الافتراضات التى يجب توافرها ليصبح تحليل التباين صحيحاً. هذه الافتراضات تختص بالمشاهدات التى يجرى تحليلها. ويمكن تعريف (وصف) كل مشاهدة بنموذج رياضى يحوى تلك التأثيرات التى يفترض أنها تؤدى إلى حدوث الاختلافات بين المشاهدات بعضها مع البعض. والنموذج الرياضى للتقسيم أحادى الجهة يعبر عنه كالتالى:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

حيث:

μ هى المتوسط العام للعشيرة المأخوذ منها هذه المشاهدات، كأن يكون عشيرة أوزان الميلاد للعجول من سلالة معينة أو نسبة البيض الذى يفقس من سلالة معينة من الدجاج أو إجمالى الدخل أو الإنفاق الأسرى شهرياً أو سنوياً فى محافظة معينة ... الخ. أى أن μ تمثل الجزء المشترك بين كل الأفراد التى ينطبق عليها تعريف هذه العشيرة.

α_i تعبر عن تأثير الفئة (المستوى) class أو المعاملة والمفترض أنها تؤثر على كل الأفراد التى تتبع أو الواقعة تحت تأثير هذا العامل و i تمثل عدد تلك الفئات ويأخذ القيم المختلفة من 1 إلى t العدد الكلى لتلك الفئات. وعلى هذا الأساس فإنه يفترض أن متوسط أى معاملة أو فئة $\mu_i = \mu + \alpha_i$ وأن القيمة المتوسطة لجميع هذه الفئات أى متوسط قيم μ_i يساوى μ، المتوسط العام. ويتبع ذلك أن مجموع قيم α_i يساوى الصفر أى أن $\sum \alpha_i = 0$، أى أن كل قيمة من قيم α_i هى عبارة عن انحراف متوسط الأفراد التابعة لهذه الفئة عن المتوسط العام.

$\in_{ij}$ هذا الجزء من النموذج تم تسميته من قبل بالخطأ والذى افترض أنه مأخوذ من توزيع طبيعى للأخطاء (الاختلافات) بمتوسط قدره صفر وتباين σ^2 ويرمز لذلك بأن $\in_{ij} \sim NID(0, \sigma^2)$ أى أنها تتوزع مستقلة عن بعضها وطبيعيا حول متوسط للاختلافات قدره صفر ونفس التباين σ^2.

Y_{ij} تعبر عن المشاهدة التى تحت فئة (مستوى) المعاملة i وترتيبها j بين المكررات والتى عددها n.

ومن الطبيعى أن الفئات المختلفة لها متوسطات مختلفة μ_i ولكن الاختلافات أو التباينات داخل كل فئة يفترض أن لها قيمة واحدة هى σ^2.

وهذه الطريقة فى وصف المشاهدة Y_{ij} تقترح أيضاً أنه إذا لم يكن هناك تأثير للمعاملات، أى أن μ_i كلها متساوية، فإن النموذج الرياضى يختزل إلى الصورة الأكثر سهولة، أى إذا كان فرض العدم صحيحاً وهى:

$$Y_i = \mu + \in_i \quad , i = 1, 2, \ldots, tn$$

وبالتالى فإن تقسيم الأفراد أو المشاهدات إلى أقسام أو فئات كان تقسيماً عشوائياً لا يدل فى الواقع على وجود اختلافات بين هذه الفئات ويصبح المصدر الوحيد للتباينات هى تلك الاختلافات الطبيعة بين الأفراد داخل العشيرة ككل.

كان الاهتمام فى هذا النموذج ينحصر فى تأثير المعاملات، فهل هذه المعاملات تؤدى إلى وجود اختلاف فى المتوسطات أم لا؟ أى هل ينشأ عن ذلك عشائر مختلفة كل منها له متوسط مختلف μ_i؟ ولاستبيان ذلك فإن الاهتمام يكون بالأساس إجراء اختبار للمعنوية لتوضيح ما إذا كانت كل المتوسطات متساوية أم لا، وهو ما سبق إيضاحه أنه يتم بواسطة اختبار F.

لاستخلاص النتائج drawing inference من النموذج الرياضى أو بمعنى آخر لاستقراء النتائج يجب أن يكون واضحاً لدى المجرب كيفية اختيار المعاملات أو الأقسام المعبر عن أثرها بـ α فى النموذج لأن كيفية الاختيار هذه ستحدد الاستنتاج المنشود، وللوصول إلى ذلك يلزم التفرقة بين الأثر الثابت fixed effect والأثر العشوائى random effect.

صندوق ٩-١

- تحليل التباين هو وسيلة لتقسيم التباين الكلى المشاهد وإرجاع كل قسم إلى عامل معين. ولإجراء التحليل هذا ليس ضرورياً فرض أى توزيع معين للمشاهدات.
- يحتوى جدول تحليل التباين على نفس مصادر التباين المعبر عنها فى النموذج الرياضى لا أكثر ولا أقل.
- بينما إجراء التحليل نفسه لا يحتاج إلى فرض توزيع معين إلا أنه لإجراء اختبارات المعنوية يجب أن تكون المشاهدات موزعة توزيعاً طبيعياً حتى يمكن استخدام اختبار F.

الأثر الثابت fixed effect

إذا كان لدى المجرب عدد من المعاملات ويرغب فى دراسة أثار والفروق بين هذه المعاملات بعينها دون غيرها فإنه يطلق على أثار المعاملة آثار ثابتة ولا يجوز أن تنسحب نتائج التحليل على غير تلك المعاملات بعينها. مثال ذلك إذا أراد المجرب دراسة أثر إضافة مكون معين للعليقة أو دراسة الفرق فى الدخل بين محافظة أ ومحافظة ب ... وهكذا، فإنه فى هذه الحالة ليس للمعاملة مكون تباين variance component ولكنه مجرد قيمة مربعة لأن مستويات المعاملة (أو الأقسام) لم تؤخذ بطريقة عشوائية ويرمز له بالرمز K^2 وليس σ^2 كما سيأتى فيما بعد. ويكون الاهتمام الأساسى فى هذه الحالة هو دراسة المتوسطات وليس التباين بينها.

الأثر العشوائى random effect

وفيه يختار المجرب مستويات المعاملة بطريقة عشوائية من عدة مستويات محتملة. فمثلاً إذا أراد المجرب دراسة ما إذا كان هناك فروق بين السلالات فى نموها، فإنه يختار عدداً من السلالات عشوائياً من عشيرة من السلالات ويجرى التحليل على هذه السلالات. وفى هذه الحالة يمكن للمجرب أن يعمم نتائجه على عشيرة السلالات التى أخذت منها العينة ويكون لبين السلالات مكون تباين σ^2 لأنه تباين حقيقى وليس مجرد قيمة مربعة كما هو الحال فى الأثر الثابت. وفى حالة الأثر العشوائى فإنه عادة ما يكون الهدف الأساسى من التحليل هو دراسة التباين فى عشيرة السلالات وليس متوسطات السلالات المختارة عشوائياً.

وتظل طريقة حساب مجموع المربعات ومتوسطها ودرجات الحرية واحدة فى كل من نماذج الأثر الثابت ونماذج الأثر العشوائى. الاختلافات فقط تكون فى اختبارات فرض العدم وتقدير مكونات التباين ولاسيما عندما يكون تحليل التباين متعدد الاتجاهات وليس فى اتجاه واحد. وسيتضح ذلك فى معالجات قادمة.

وإذا كانت كل العوامل فى النموذج ذات أثر ثابت يطلق عليه نموذج ثابت fixed mode (نموذج من النوع الأول model I). أما إذا كانت أثار العوامل فى النموذج عشوائية فيطلق عليه نموذج عشوائى random model (نموذج من النوع الثانى model II). وفى حالة ما إذا احتوى النموذج على بعض العوامل ذى أثر ثابت وأخرى ذى أثر عشوائى سمى بالنموذج المختلط mixed model (نموذج من النوع الثالث model III).

لاحظ أنه فى جميع أنواع النماذج السابقة يعتبر المتوسط العام ذا أثر ثابت والخطأ ذا أثر عشوائى.

مثال ٩-١

البيانات التالية تمثل الأوزان النهائية لذكور الحملان المسمنة لفترة 120 يوما بعد الفطام على خمسة علائق مختلفة. والمراد معرفة ما إذا كانت الاختلافات بين العلائق تؤدى إلى اختلاف الوزن باعتبار أن الحيوانات قد بدأت من أوزان متقاربة. (تم طرح 40 كيلوجرام من الأوزان لسهولة الحساب).

	العليقة					
	هـ	د	ج	ب	أ	
	7	2	3	9	5	
	6	3	5	7	4	
	9	4	2	8	8	
	4	1	3	6	6	
	7	4	7	9	3	
132	**33**	**14**	**20**	**39**	**26**	$Y_{i.}$ **المجموع**
5.28	**6.6**	**2.8**	**4.0**	**7.8**	**5.2**	$\overline{Y}_{i.}$ **المتوسط**

الحل التفصيلى:

(١) فرض العدم:

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$، أى أن المتوسطات للعلائق متساوية.

الفرض البديل:

$H_1: \mu_i \neq \mu_{i'}$، أى أنه على الأقل يوجد متوسطان غير متساويين. ويحدد مستوى المعنوية، أى حجم منطقة الرفض $\alpha = 0.05$ مثلا.

(٢) النموذج الرياضى:

$$Y_{ij} = \mu + \tau_i + \in_{ij}$$

حيث:

Y_{ij} = وزن الحمل j تحت المعاملة i ، حيث $i = 1, \cdots, 5$ أى 5 معاملات ، $j = 1, \cdots, 5$ ، أى 5 مكررات فى كل معاملة،

μ = المتوسط العام،

τ_i = تأثير المعاملة i كانحراف عن المتوسط العام،

$\in_{ij}$ = الخطأ التجريبى $\in_{ij} \sim NID(0, \sigma^2)$، أى يتوزع طبيعياً بمتوسط 0، وتباين σ^2.

(٣) الحسابات:

(أ) مجموع المربعات الكلى المصحح: Total sum of squares (TSS)

$$TSS = \sum_i \sum_j Y_{ij}^2 - \frac{\left(\sum_i \sum_j Y_{ij}\right)^2}{tn} = \sum_i \sum_j Y_{ij}^2 - \frac{(Y_{..})^2}{tn} \qquad (٩\text{-}٣)$$

$$= 5^2 + 4^2 + ... + 7^2 - \frac{(132)^2}{25} = 834 - 696.96 = 137.04$$

وله 24 درجة حرية، والقيمة $(Y..)^2 / tn$ تعرف بمعامل التصحيح correction factor (CF)

(ب) مجموع المربعات بين المعاملات (العلائق)

Between treatments sums of squares (BSS)

$$BSS = \sum_i \frac{Y_{i.}^2}{n} - CF \qquad (٩\text{-}٤)$$

$$= \frac{26^2 + 39^2 + ... + 33^2}{5} - 696.96 = 79.44$$

وله 4 درجات حرية.

(ج) مجموع المربعات داخل المعاملات (الخطأ)

Within treatment sums of squares (WSS)

$$WSS = TSS - BSS \qquad (٩\text{-}٥)$$

بمعنى طرح مجموع المربعات بين المعاملات من مجموع المربعات الكلى المصحح أى $137.04 - 79.44 = 57.6$ ، وله 20 درجة حرية.

وعادة ما يتم تلخيص وعرض البيانات الخاصة بتحليل التباين فى جدول يعرف جدول تحليل التباين ويرمز له بالرمز ANOVA table، ويوضح جدول ٩-١ هذا التحليل.

جدول ٩-١ تحليل التباين لبيانات أوزان الحملان المسمنة على 5 علائق مختلفة

ANOVA table

مصدر التباين Source of variation (SOV)	درجات الحرية Degrees of freedom (df)	مجموع المربعات Sum of squares (SS)	متوسط المربعات Mean square (MS)	F المحسوبة Computed F
بين المعاملات Treatments	(t-1) = 4	79.44	19.86	6.9
داخل المعاملات Error	t(n-1) = 20	57.6	2.88	
الكلى المصحح C-Total	**(tn – 1) = 24**	**137.04**		

(٤) الاختبار:

بما أن القيمة المحسوبة لاختبار F

$$= \frac{\text{متوسط المربعات بين المعاملات}}{\text{متوسط المربعات داخل المعاملات}} = 6.9$$

وبمقارنتها بقيمة F من جدول F (جدول ٥ ملحق أ) التى يتم الكشف فيها عن القيمة التى تمثل هذه النسبة بين متوسطى المربعات عند مستوى المعنوية المحدد قبل التجربة α أى منطقة الرفض. إذا كانت قيمة F المحسوبة أكبر من تلك الموجودة فى الجدول، فإن فرض العدم يرفض.

ويلاحظ أن جدول F يختلف عن جدول t لأن جدول F به مدخلان، أحدهما أفقى يمثل عدد درجات الحرية بين الفئات أو المعاملات والآخر رأسى يمثل درجات حرية الخطأ.

وفى المثال السابق إذا حددت منطقة الرفض $\alpha = 0.05$ وعند درجات حرية بين المعاملات = 4 وداخل المعاملات = 20 فإن قيمة F الجدولية $F_{(4,20,0.05)} = 2.87$ ، أى أن هناك سببا لرفض فرض العدم وقبول الفرض البديل وهو أن متوسطات المعاملات مختلفة عن بعضها.

مثال ٩-٢

يمكن الحصول على نفس الحل للمثال ٩-١ باستخدام برنامج SAS كالتالي:

```
DATA LAMB;
 INPUT TRT $ WEIGHT @@;
CARDS;
A 5 A 4 A 8 A 6 A 3 B 9 B 7 B 8 B 6 B 9 C 3 C 5 C 2 C 3 C 7
D 2 D 3 D 4 D 1 D 4 E 7 E 6 E 9 E 4 E 7
PROC ANOVA;
CLASS TRT;
MODEL WEIGHT = TRT;
RUN;
```

لاحظ:

استخدمت العلامات @@ حتى يمكن قراءة البيانات بالتتابع بدلاً من وضع كل معاملة فى سطر منفصل.

استخدمت علامة $ للدلالة على أن رمز المعاملة a, b , c , d, e تقرأ كحروف هجائية وليس كأرقام.

تم وضع البيانات بحيث إن كل رقم يسبقه رمز المعاملة الخاصة به.

استخدمت proc anova للحصول على تحليل التباين، مع ملاحظة أن أثر المعاملة لابد وأن يوضع فى الـــ class.

النتائج:

The SAS System

Analysis of Variance Procedure

Dependent Variable: WEIGHT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	79.4400000	19.8600000	6.90	0.0012
Error	20	57.6000000	2.8800000		
Corrected Total	24	137.0400000			

لاحظ أنه قد تم حذف بعض نتائج البرنامج والتى سوف تذكر فيما بعد.

٩-٧ اختبار جميع الفروق بين أزواج المتوسطات (فصل المتوسطات)

Test of all differences between pairs of means -pairwise comparisons

بعد إجراء اختبار F ووجود أساس لرفض فرض العدم بعدم وجود فروق معنوية بين متوسطات المعاملات وذلك فى حالة المعاملات الثابتة، أى أن البيانات تتبع عشائر مختلفة فى متوسطاتها ولكن لها نفس التباين، وأيضا فى حالة عدم وجود مقارنات comparisons معينة مختارة مقدما لإجرائها فإنه قد يكون من المناسب إجراء اختبارات بين كل زوج من المتوسطات، وبالتالى تحديد أى المتوسطات تختلف عن بعضها وأيها لا يختلف، وهذا ما يطلق عليه بفصل المتوسطات. وقد تكون هذه المقارنات مقررة مقدما قبل بدء التجربة وتسمى المقارنات فى هذه الحالة apriori comparisons أو آجلا بعد الحصول على النتائج الفعلية وتسمى aposteriori comparisons. وهذا يعنى أنه لو أن F معنوية فإنها لا تعنى بالتالى أن جميع المتوسطات مختلفة عن بعضها البعض. فإن بعض الفروق بين المتوسطات تعتبر فروقا حقيقية فى حين أن بعض الفروق تمثل اختلافات ولا تمثل فروقا حقيقية بين العشائر واختبار F فى هذه الحالة يمثل متوسطا لمعنوية الفروق ولكن لا يوضح معنوية فروق معينة. كما قد تتصادف بعض الحالات فى أن تكون F غير معنوية ولكن يكون هناك فرق بين بعض أزواج المتوسطات.

٩-٧-١ اختبار أقل فرق معنوى The least significant difference (LSD)

إذا لم يوجد اتجاه معين لإجراء مقارنات محددة فإنه يمكن استخدام اختبار يعرف باختبار أقل فرق معنوى (LSD)، وهو بالأساس يعتبر اختبار t على مستوى معنوية محدد. ويعتبر هذا الاختبار من أسهل طرق فصل المتوسطات وأكثرها شيوعا.

وحيث إن الانحراف القياسى للفرق بين متوسطين فى العشيرة يقدر من المعادلة (٨-٥) أى $\sqrt{2\sigma^2/n}$ وحيث إن التباين فى العشيرة بين الأفراد المعاملة بنفس المعاملة وهو σ^2 يتم تقديره من تحليل التباين بواسطة S^2 أى متوسطات المربعات داخل المعاملات وعليه فان التقدير estimate للانحراف المعيارى للفرق بين متوسطى معاملتين هو $\sqrt{2\sigma^2/n}$ وأن قيمة S^2 هذه لها درجات حرية الخطأ.

وفى المثال ٩-١ وجد أن $S^2 = 2.88$ ولها 20 درجة حرية فى الخطأ وحيث إن كل متوسط يقدر من $n = 5$ مشاهدات، فإن تقدير الخطأ القياسى للفرق بين متوسطين هو $\sqrt{2\sigma^2/n} = \sqrt{2(2.88)/5} = 1.073$، وأن قيمة t الجدولية على مستوى معنوية 5% ودرجات حرية 20 يساوى 2.086، فإنه حتى يكون الفرق بين أى زوج من

المتوسطات معنويا يجب أن تزيد قيمته عن $2.24 = (2.086)(1.073)$ وهى تمثل الكمية التى يطلق عليها أقل فرق معنوى وهى:

$$.\ LSD = t\left(\sqrt{\frac{2S^2}{n}}\right) \qquad (٩-٦)$$

وهذه القيمة تمثل أقل كمية يجب أن يصل إليها الفرق بين متوسطين حتى يقال إن هذا الفرق حقيقى (طبعا بمستوى معنوية محدد). وإذا أجرى الاختبار على مستوى 1% فقط فتصبح القيمة $3.05 = (2.845)(1.073)$، ويكون الفرق بين متوسط المعاملة (ب) وهى أعلاها ومتوسط المعاملة (د) وهى أقلها $7.8 - 2.8 = 5$ kg يعتبر فرقاً معنويا على مستوى 1%، وأيضا الفرق بين متوسطى المعاملتين هـ ، د $6.6 - 2.8 = 3.8$ kg. فى حين أن الفرق بين متوسطى المعاملتين أ، ج $5.2 - 4 = 1.2$ kg لا يمثل فرقا حقيقياً.

وعند الرغبة فى إجراء مقارنة بين كل زوجين من المتوسطات فإنه يمكن ترتيب المتوسطات ترتيبا تنازليا فى جدول، ثم يتم بعد ذلك طرح أقل المتوسطات قيمة من أعلاها ومن ثم المتوسط الذى يليه فى الترتيب التصاعدى ... وهكذا. ويتم مقارنة كل من الفروق الناتجة بالقيم المحسوبة لأقل فرق معنوى LSD وإعلان الفروق الحقيقية التى تزيد قيمتها عن قيمة الاختبار. وتسهيلا على ذلك فإنه إذا كان أحد الفروق غير معنوى فإن الفرق التالى لن يكون أيضا معنويا وعليه يقل عدد المقارنات المحسوبة.

ويمكن توضيح ذلك فى جدول ٩-٢ لمقارنة متوسطات المعاملات الخمس فى مثال ٩-١.

جدول ٩-٢ اختبار معنوية الفروق بين أزواج المتوسطات بواسطة اختبار LSD

المتوسط	الفروق (كج)			
	- د	- ج	- أ	- هـ
7.8 = ب	5.0**	3.8**	2.6	1.2
6.6 = هـ	3.8	2.6	1.4	
5.2 = أ	2.4	1.2		
4.0 = ج	1.2			

ويتم مقارنة الفروق الموضحة فى الجدول بالقيم المحسوبة LSD على مستوى المعنوية المطلوب. وتبين الاختلافات المعنوية بوضع علامة (**) على الفروق المعنوية على مستوى 1% أما إذا كان المطلوب إجراء الاختبار على مستوى 5% فيمكن وضع علامة (*) على الفروق المعنوية.

وهناك اعتراضات كثيرة على استخدام اختبار LSD دون تمييز لإجراء الاختبار بين جميع أزواج المتوسطات. وذلك لأنه قد يجرى عدد أكبر من الاختبارات أكثر مما تتحمله المعلومات المتوفرة فى التجربة وبالتالى يتم الوصول إلى قرارات يعتقد أنها صحيحة فى حين قد لا يتوافر لها نفس القدر من الاحتمال للحكم الخاطئ.

فإذا كانت هناك تجربة بها 3 معاملات فإن هناك درجتى حرية للفروق بين المتوسطات فى هذه الحالة فى حين أن هناك 3 فروق ممكنة فيما بينها. وإذا كان هناك 5 معاملات فإن عدد المقارنات بين أزواج المتوسطات الممكنة يبلغ 10 فروق فى حين أنه توجد 4 درجات حرية. وفى حالة 10 معاملات فإنه يجرى 45 مقارنة فى حين أنه توجد 9 درجات حرية فقط.

ففى حالة مقارنة 3 معاملات أ، ب، ج، فإن هناك 3 مقارنات ممكنة، ولكن إذا أخذت مقارنة (أ، ب) بالإضافة إلى المقارنة (ب، ج) أى (أ – ب) + (ب – ج) والتى تعطى (أ – ب + ب + ج) أى (أ – ج) والتى يطلق عليها المقارنة الثالثة وعلى ذلك فالأخيرة غير مستقلة عن المقارنتين السابقتين وذلك لاشتراك المعاملة (ب) فى كل من المقارنتين.

ولقد أوضح Cochran and Cox (1957) أنه فى حالة إجراء المقارنة بين أعلى متوسط وأقل متوسط فى تجربة بها أكثر من معاملتين فإن الفرق بينهما إذا تم اختباره بواسطة LSD فإنه قد يعلن عن فروق معنوية فى حين أنها ليست حقيقية، أى رفض فرض العدم فى حالات أكثر من الحقيقية. ويرجع ذلك إلى أنه عند إجراء كل المقارنات فإن حجم منطقة الرفض α يصبح كبيرا وليس عند المستوى المحدد من قبل.

فمثلا إذا كان المجرب يريد إجراء اختبار بين أعلى وأقل المتوسطات على مستوى معنوية 5% فإنه فى الحقيقة يكون حجم α فى حالة 3 معاملات 13% . أما إذا زاد عدد المعاملات وأصبح 5 معاملات، فإنه يصل إلى 29% . وفى حالة 10 معاملات يصبح احتمال وجود فرق معنوى غير حقيقى 60%، أى أنه قد يكون من الأوفق فى هذه الحالة إجراء الاختبار برمى قطعة نقود وإعلان الفرق معنويا عند الحصول على الصورة وغير معنوى عند الحصول على الكتابة حيث إنه فى هذه الحالة يصبح احتمال الخطأ 50% وهو أقل من 60% المذكورة. هذا النوع من

الاختبارات التى تتم بعد الحصول على النتائج يعرف بالمقارنات التى توجهها البيانات المتحصل عليها ولذلك فهى غير سليمة وغالبا منحازة. وحيث إن اختبار LSD يعرض التجربة والمقارنات لأكبر قدر من الخطأ من النوع الأول بالنسبة لبقية الطرق المعروضة هنا وبالتالى فهو أكثر الطرق قوة فى الاختبار (Kemp, 1975).

أما إذا كانت المقارنات التى يرغب المجرب فى تقييمها قد خططت من قبل إجراء التجربة بحيث أن تكون مستقلة عن بعضها، أى أن بعضها لا يعبر عنه فى صورة دالة خطية من المقارنات الأخرى، فإن هذه المقارنات تعرف بالمقارنات المستقلة independent or orthogonal وسوف يتم توضيحها فى تفصيل لاحق.

وهذا النوع من المقارنات يمكن اختباره بواسطة LSD دون أن يؤثر زيادة عدد المعاملات فى التجربة فى زيادة احتمالات الخطأ من النوع الأول α، فمثلا إذا كانت هناك تجربة بها 6 معاملات فإن هناك 6 متوسطات هى $\overline{Y}_6, \overline{Y}_5, \overline{Y}_4, \overline{Y}_3, \overline{Y}_2, \overline{Y}_1$. فإذا قورن $(\overline{Y}_5 - \overline{Y}_6), (\overline{Y}_3 - \overline{Y}_4), (\overline{Y}_1 - \overline{Y}_2)$ مثلا، يلاحظ أن كل معاملة لم تظهر فى أكثر من مقارنة واحدة وبالتالى فإن هذه المقارنات تعتبر مستقلة عن بعضها البعض.

وقد سبق إيضاح أنه فى حالة عدم معنوية اختبار F فإنه لا ينصح بإجراء أية مقارنات بين المتوسطات إلا إذا كانت هذه المقارنات مخططة من قبل إجراء التجربة، وأن التجربة قد صممت لاستبيان هذه الفروق، عندئذ قد يصبح استعمال LSD فى المقارنات اختباراً صحيحاً. ويوضح Fisher (1935) أنه فى حالة عدم معنوية اختبار F فإنه لا يوجد أى دليل يؤيد رفض فرض العدم وأنه يجب الحذر فى إعلان فروق معنوية لبعض المقارنات. ولقد أدى ذلك إلى تعضيد الكثير من الإحصائيين عدم إجراء اختبار LSD فى حالة عدم معنوية اختبار F لتوفير قدر من الحماية ضد الخطأ من النوع الأول.

مثال ٩-٣

استخدمت 4 علائق rations فى تغذية الحملان بعد فطامها وحسب معدل الزيادة اليومى فى الوزن بالجرام خلال فترة التجربة حيث استخدم 5 حملان فى كل معاملة. ويريد القائم بالتجربة اختبار الفرق بين كل من المعاملتين 1، 4 وأيضاً المعاملتين 2 ، 3 وكانت معدلات الزيادة فى الوزن بالجرام كالتالى:

المعاملة	المكررات					المجموع	المتوسط (جم)
١	110	98	84	42	104	**438**	**87.6**
٢	122	224	60	178	126	**710**	**142.0**
٣	84	194	162	190	184	**814**	**162.8**
٤	338	274	338	170	308	**1428**	**285.6**
						3390	**169.5**

الحل:

الشخص القائم بالتجربة لديه فكرة مبدئية عن تأثير المعاملات المستخدمة فى التجربة وذلك لأنه قام بوضع فرض لاختبار المعاملات حيث إنه سوف يقارن المعاملة ٢ مع ٣ وكذلك ١ مع ٤، وذلك لأنه يعتقد أن هناك سبباً لوجود اختلافات بين كل زوج من هذه المعاملات. ولكنه فى نفس الوقت يرى تجريب المعاملات الأربع بفرض زيادة مدى التجربة وجعلها أوسع وهو ما سوف يناقش فيما بعد.

فرض العدم: $H_o: \mu_i = \mu_i' = \mu$

الفرض البديل: $H_1: \mu_2 \neq \mu_3$ و $\mu_1 \neq \mu_4$

النموذج الرياضى: $Y_{ij} = \mu + \tau_i + \epsilon_{ij}$

الافتراضات: $\epsilon_{ij} \sim NID(0, \sigma^2)$

الاشتراطات: $\sum \tau_i = 0$

معامل التصحيح: $C.F. = \frac{(\sum Y_{ij})^2}{tn} = \frac{(3390)^2}{20} = 574605$

مجموع المربعات الكلى المصحح:

$$(110)^2 + (98)^2 + ... + (308)^2 - C.F. = 151175$$

مجموع المربعات للعلائق:

$$\frac{(438)^2 + ... + (1428)^2}{5} - C.F. = 104939.8$$

مجموع المربعات الداخلى (الخطأ) $151175 - 104939.8 = 46235.2$

والجدول ٩-٣ يمثل جدول تحليل التباين

جدول ٩-٣ تحليل التباين للبيانات الخاصة بمثال ٩-٣

SOV	df	SS	MS	F - test
بين العلائق Ration	3	104939.8	34979.9**	12.1
داخل العلائق Error	16	46235.2	2889.7	
الكلى C-Total	**19**	**151175**		

وبما أن F الجدولية $F_{(3,16,0.01)} = 5.29$ فإن النتائج توضح وجود اختلافات بين المعاملات ويرفض فرض العدم وبالتالى يمكن إجراء اختبار LSD لتوضيح أى من المقارنات المخططة معنوية.

وبما أن المعاملات متساوية فى تكراراتها فإن:

الخطأ القياسى للفرق بين متوسطى معاملتين $g\ 34 = \sqrt{\dfrac{2(2889.7)}{5}}$

وحيث إن اختبار F قد أجرى على مستوى معنوية 0.01 فإنه بالتالى سوف يتم اختبار LSD على نفس المستوى . وحيث قيمة t الجدولية = 2.921

إذا أقل فرق معنوى $LSD = (2.921)(34) = 99.3\ g$

المقارنة الأولى (عليقة ٢،٣): $162.8 - 142 = 20.8\ g$ والفرق غير معنوى.

المقارنة الثانية (عليقة ٤، ١): $285.6 - 87.6 = 198\ g$ والفرق معنوى (0.01).

وقد يكون من الملائم أيضاً هنا توضيح أنه إذا افترض أنه يراد إجراء اختبار لمعنوية جميع الفروق (وهو ما لا يعتبر اختباراً جيداً فى هذه الحالة) يتضح أن المعاملة الرابعة كانت الوحيدة التى يختلف متوسطها عن بقية المعاملات. أما المعاملات الأخرى فهى غير مختلفة عن بعضها. ويمكن توضيح ذلك دون كتابة جداول لكل الفروق. وذلك بترتيب المتوسطات تنازلياً أو تصاعدياً مثلا ثم ربط المتوسطات التى لا تختلف معنوياً عن بعضها بخط أسفلها أما التى لا ترتبط ببعضها فهذا يدل على اختلافها معنوياً كما يلى:

(١)	(٢)	(٣)	(٤)	
87.6	143	162.8	285.6	المتوسطات

كما أن هناك صورة أخرى لتوضيح ذلك بأن يوضع نفس الحرف الهجائى على كل المعاملات التى لا تختلف معنوياً ووضع حرف آخر على تلك التى تختلف عن بعضها كالآتى:

(١)	(٢)	(٣)	(٤)	
B 87.6	B 143	B 162.8	A 285.6	المتوسطات

وطريقة العرض هذه تصلح فى حالة اختلاف عدد الوحدات التجريبية فى المعاملات.

وعموماً فإن اختبار LSD يعتبر اختبارا معقولا وسهلا وقوياً بالمفهوم الإحصائى وذلك لإجراء المقارنات بقيمة واحدة فقط وكما إنه فى حالة المقارنات المخططة والمستقلة فإن الاختبار يعتبر صحيحاً، حيث الاختبار يعتبر اختبار مقارنة بين متوسطين comparisonwise test.

مثال ٩-٤

استخدام برنامج SAS لإجراء اختبار LSD لبيانات مثال ٩-١

```
DATA LAMB;
 INPUT TRT $ WEIGHT @@;
CARDS;
A 5 A 4 A 8 A 6 A 3 B 9 B 7 B 8 B 6 B 9 C 3 C 5 C 2 C 3 C 7
D 2 D 3 D 4 D 1 D 4 E 7 E 6 E 9 E 4 E 7
PROC ANOVA;
CLASS TRT;
MODEL WEIGHT = TRT;
MEANS TRT/ LSD;
RUN;
```

لاحظ:

تم استخدام برنامج SAS بنفس الخطوات المذكورة فى مثال ٩-٢ مع إضافة اختيار means trt/LSD والذى يعنى طلب حساب LSD لمتوسطات المعاملات.

ولابد أن يأتى هذا الاختيار بعد الـــ model وان تشتمل الـــ class على الأثر المطلوب حساب LSD له.

إذا لم يذكر مستوى المعنوية فى البرنامج فإن LSD الناتجة سوف تكون عند 0.05 ويمكن إضافة 0.01 عند الرغبة فى حساب LSD عند مستوى معنوية 1% كالتالى: means trt/ LSD alpha = 0.01;

النتائج:

The SAS System

Analysis of Variance Procedure

Dependent Variable: WEIGHT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	79.4400000	19.8600000	6.90	0.0012
Error	20	57.6000000	2.8800000		
Corrected Total	24	137.0400000			

Analysis of Variance Procedure
T tests (LSD) for variable: WEIGHT

NOTE: This test controls the type I comparison wise error rate not the experiment wise error rate.

Alpha= 0.05 df= 20 MSE= 2.88
Critical Value of T= 2.09
Least Significant Difference= 2.2389

Means with the same letter are not significantly different.

T Grouping		Mean	N	TRT
A		7.800	5	b
B	A	6.600	5	e
B	C	5.200	5	a
D	C	4.000	5	c
D		2.800	5	d

بمعنى أن المتوسطات التى يصحبها نفس الحرف الهجائى لا تختلف معنوياً عن بعضها.

وهذه نفس النتائج التى سبق الحصول عليها.

٩-٧-٢ اختبار توكى Tukey's test

نظراً لما قيل عن الـــ LSD ومن تعريضه التجربة والمقارنات إلى خطأ أكبر من النوع الأول، حاول Tukey أن يجد حلا لعلاج هذا باختبار أطلق عليه HSD (honestly significant difference) ويستخدم فيه جداول معدلة عن جدول (t) (جدول ١٥ ملحق أ). ويتصف اختبار توكى هذا بأنه يعطى قدرا أكبر من الحماية من خطأ النوع الأول، وعليه فهو يعرض التجربة والمقارنات إلى خطأ أكبر من النوع الثانى (أى لا يرفض فرض العدم بينما هو غير صحيح) وبالتالى فهو أقل قوة فى إعلان الفروق المعنوية إذا وجدت. وطريقة حسابه تماثل إلى حد كبير طريقة حساب LSD مع استخدام قيمة Q بدلا من t الأصلية كما تستخدم فى اختبار توكى S^2/n بدلا من $2S^2/n$ حيث دمجت $\sqrt{2}$ فى الجدول لتوفر خطوة على الحاسب وعليه تكون:

$$HSD = Q_\alpha \sqrt{\frac{S^2}{n}} \qquad (٩-٧)$$

من جدول ٩-٣ الخاص بتحليل التباين لمثال ٩-٣ $S^2 = 2889.7$ له 16 درجة حرية، وكان عدد المعاملات 4 وبكل معاملة 5 مشاهدات، ومن جدول ١٥ ملحق أ تكون $Q_\alpha = 4.05$، أى:

$$HSD = (4.05)\left(\sqrt{\frac{2889.7}{5}}\right) = 97.36$$

وواضح أن HSD عند مستوى 5% أكبر من LSD عند نفس المستوى والذى كان

$$.\,LSD = (2.12)\left(\sqrt{\frac{(2)(2889.7)}{5}}\right) = 72.08$$

وباختبار الفروق بين كل معاملتين طبقاً لاختبار توكى، تقارن كل من هذه الفروق بالقيمة HSD=97.36 . فإذا تساوى زاد الفرق المحسوب عن 97.36 كان هذا الفرق معنوياً وإذا قل عن 97.36 لا يكون معنوياً، وتكون الفروق بين متوسطات المعاملات كالتالى:

٤ – ١ = 198.0 جرام وهى تزيد عن 97.36، الفرق معنوى عند 5%
٤ – ٢ = 143.6 جرام وهى تزيد عن 97.36، الفرق معنوى عند 5%
٤ – ٣ = 122.8 جرام وهى تزيد عن 97.36، الفرق معنوى عند 5%
٣ – ١ = 75.2 جرام وهى تقل عن 97.36، الفرق غير معنوى عند 5%
٣ – ٢ = 20.8 جرام وهى تقل عن 97.36، الفرق غير معنوى عند 5%
٢ – ١ = 54.4 جرام وهى تقل عن 97.36، الفرق غير معنوى عند 5%

ويلاحظ أن الفرق (٣) – (١) = 75.2 وهو معنوى عند 5% باستخدام LSD بينما هذا الفرق غير معنوى باستخدام اختبار توكى.

٩-٧-٣ اختبار المدى المتعدد الجديد لـ (دنكن):

Duncan's new multiple range test

وجد أن LSD يحقق أكبر قدر من قوة الاختبار بينما يعرض التجربة والمقارنات إلى أكبر قدر من خطأ النوع الأول، بينما HSD أو اختبار توكى على نقيض ذلك يعطى حماية أكبر ضد خطأ النوع الأول ولكنه أقل قوة فى إعلان الفروق المعنوية. لذا فإن دنكن (Duncan, 1955) حاول إيجاد حل وسط بين الاثنين بانيا نظريته على ما يلى:

إذا أخذ الفرق بين أعلى متوسط وأقل متوسط فإن هذا الفرق يحتمل أن يحتوى على أخطاء عشوائية أكبر من أى فرق آخر بين متوسطين. لذلك لابد من توفر قدر من الحماية ضد خطأ النوع الأول متناسباً مع درجة الفرق هذه، أى ترتيب المتوسطات بالنسبة لبعضها. ففى مثال ٩-٣ أعلى متوسط كان للمعاملة (٤) وأقل متوسط كان للمعاملة (١)، وعليه فالفرق بينهما 198 g يكون أكبر عرضة للخطأ العشوائى أكبر من الفرق بين معاملتين متجاورتين فى ترتيبهما مثل (٤) – (٣) = 122.8 g، (٣)-(٢) = 20.8 g ... وهكذا. وعليه فكل فرق يختبر على أساس قيمة تختلف حسب بعد المتوسطات عن بعضها.

ويكون هذا الاختبار ملائما عند الرغبة فى إجراء مقارنات بين جميع أزواج المتوسطات. ويمكن إجراء هذا الاختبار سواء كانت F معنوية أم غير معنوية.

وتتلخص طريقة اختبار دنكن فى مقارنة المدى بين متوسطين بعد الترتيب التصاعدى بقيمة تعرف بالمدى significant studentized range ويرمز له بالرمز

SSR وحيث إن قيمة $LSD = t\sqrt{\frac{2S^2}{n}}$ أى أنها تساوى $t\sqrt{2S^2_{\bar{Y}}}$، وحيث إن قيمة t تستعمل لاختبار الفرق بين متوسطين وليس المدى فيما بينهما فقد صممت قيم أخرى تشابه قيمة t وتتزايد بتزايد المدى الذى تحتويه المقارنة المختبرة، ويلاحظ أنةً فى حالة ما إذا كان المدى يحتوى على متوسطين متجاورين فى الترتيب فإن القيمة الجديدة تتطابق مع قيمة t الجدولية.

ولقد تم ضرب القيم الجديدة فى $\sqrt{2}$ لأنه ثابت لينتج ما يعرف بقيم SSR وهى الموجودة فى جدول ١٦ ملحق أ. وهذا الجدول له مدخلان، الأفقى منها يدل على عدد المتوسطات الداخلة فى المدى الذى تم فيه المقارنة وتبدأ بالرقم 2، 3، 4 ... وهكذا. أما الرأسى فإنه يعطى عدد درجات الحرية المحسوب منها الخطأ القياسى $S_{\bar{Y}}$ من جدول تحليل التباين. وهناك قيمتان لــ SSR أحداهما خاصة بمستوى (0.05)، والأخرى خاصة بمستوى (0.01). وهنا تجدر الإشارة إلى أن هذه المستويات لا تمثل احتمال حدوث خطأ من النوع الأول α كما هو الحال فى اختبار LSD حيث إن هذا ليس اختبارا للمعنوية ولكن يسمى هذا المستوى بمستوى الحماية من حدوث الخطأ الخاص بالمقارنة التى تشمل المدى special protection level.

وبتطبيق اختبار دنكن على البيانات الخاصة بمثال ٩-٣ ومن جدول تحليل التباين الخاص به فإن قيمة الخطأ القياسى $\sqrt{2889.7/5} = 24.04$.

ولإجراء اختبار دنكن لكل أزواج المتوسطات للعلائق المختلفة يتم ترتيبها تنازلياً وأخذ الفروق بينها بنفس طريقة الجدول ٩-٢، والتى يوضحها الجدول ٩-٤.

جدول ٩-٤ اختبار الفروق بين المتوسطات بواسطة اختبار دنكن فى مثال ٩-٣

المتوسط	الفروق		
	المتوسط – (١)	المتوسط – (٢)	المتوسط – (٣)
(٤) **285.6**	198.0*	143.6*	122.8*
(٣) **162.8**	75.2*	20.8	-
(٢) **142.0**	54.4	-	
(١) **87.6**	-		

وبالتالى فعند الرغبة فى إجراء الاختبارات على مستوى 0.05 فإن قيم SSR حسب عدد المتوسطات P فى المدى وعند درجات حرية الخطأ 16 تكون:

P	2	3	4
SSR	3.00	3.15	3.23

وعلى هذا الأساس فإنه لمقارنة متوسط المعاملة (٤) بمتوسط المعاملة (١) يحتوى المدى على 4 متوسطات وتصبح قيمة اختبار دنكن:

$$(SSR)(S_{\overline{Y}}) = (3.23)(24.04) = 77.65$$

وبما أن القيمة أقل من الفرق المشاهد = 198 إذا يعلن المدى معنويا. أما فى حالة مقارنة متوسط المعاملة (٤) بمتوسط المعاملة (٢) فإن المدى يحتوى 3 متوسطات وتصبح القيمة $75.73 = (24.04)(3.15)$ ، وفى حالة مقارنة مدى يحوى متوسطين فقط تصبح القيمة $72.12 = (24.04)(3.0)$ ، ... وهكذا.

ويلاحظ هنا أن المقارنة تتم بأكثر من قيمة فى حين أنه فى حالة LSD كانت هناك قيمة واحدة لإجراء المقارنة بها.

ويلاحظ أنه فى حالة وجود متوسطين فقط فإن نتائج الاختبارات الثلاثة لابد وأن تكون متطابقة؛ وذلك لأن هناك درجة حرية واحدة بين المتوسطين، وأيضاً هناك فرق واحد دون غيره. بينما إذا وجد ثلاث معاملات بينها درجتان حرية وبينها ثلاث مقارنات هى: (أ – ب)، (أ – ج)، (ب – ج). وهنا قد يبدأ الاختلاف فى نتائج الطرق الثلاث، وكلما زاد عدد المعاملات كلما كانت هناك فرصة أكبر لأن تختلف الطرق الثلاث عن بعضها. والجدول التالى يبين عدد المعاملات وعدد المقارنات المستقلة (أى درجات الحرية) وعدد المقارنات الممكنة بين هذه المتوسطات متوسطات:

عدد المعاملات	عدد المقارنات المستقلة = (عدد المعاملات – 1) = درجات حرية المعاملات	عدد المقارنات الممكنة = $\frac{(\text{عدد المتوسطات})(\text{عدد المتوسطات} - 1)}{2}$
2	1	1
3	2	3
4	3	6
5	4	10
6	5	15
7	6	21
10	9	45

وواضح أنه كلما زاد عدد المعاملات كلما زاد الفرق والنسبة بين عدد المقارنات الممكنة ودرجات الحرية، وهذه مدعاة أكبر لأن تختلف الطرق الثلاث عن بعضها.

ويوجد طرق أخرى لفصل المتوسطات ولكل مواصفات ومزايا وعيوب. إلا إن هذه الطرق الثلاث السابق ذكرها هى أكثرها شيوعاً فى التجريب الزراعى وعلى المجرب أن يقرر أيها يستخدم طبقاً لما يريده من الحماية أو السماح تجاه نوعى الخطأ. وعموماً يمكن القول إن LSD أكثرها قوة وأكثرها خطأ من النوع الأول وعلى النقيض منها اختبار توكى HSD وبينهما هو اختبار دنكن (Kemp, 1975; Gill, 1973; Balaam, 1963). وغنى عن البيان أنه يمكن حساب كل من توكى و دنكن على مستوى (0.01).

مثال ٩-٥

استخدام برنامج SAS لإجراء اختبارات توكى و دنكن لبيانات مثال ٩-٣

```
DATA LAMB;
INPUT TRT GAIN @@;
CARDS;
1 110 1 98 1 84 1 42 1 104 2 122 2 224 2 60 2 178 2 126
3 84 3 194 3 162 3 190 3 184 4 338 4 274 4 338 4 170 4 308
PROC ANOVA;
CLASS TRT;
MODEL GAIN = TRT;
MEANS TRT/ TUKEY DUNCAN;
RUN;
```

لاحظ:

لم تستخدم علامة $ فى input وذلك لتمييز مستويات المعاملة بالأرقام بدلا من الحروف.

استخدم كل من اختيار TUKEY و DUNCAN مع اختيار means.

النتائج المتحصل عليها عند مستوى معنوية 0.05 ويمكن تغيير ذلك إلى مستوى 0.01 كما سبق ذكره.

النتائج:

```
Analysis of Variance Procedure
Class Level Information
Class   Levels   Values

TRT       4      1 2 3 4
Number of observations in data set = 20
```

Dependent Variable: GAIN

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	104939.800	34979.933	12.11	0.0002
Error	16	46235.200	2889.700		
Corrected Total	19	151175.000			

Duncan's Multiple Range Test for variable: GAIN
NOTE: This test controls the type I comparison wise error rate, not the experiment wise error rate

Alpha= 0.05 df= 16 MSE= 2889.7

Number of Means	2	3	4
Critical Range	72.07	75.58	77.77

Means with the same letter are not significantly different.

Duncan Grouping	Mean	N	TRT
A	285.60	5	4
B	162.80	5	3
B	142.00	5	2
B	87.60	5	1

Tukey's Studentized Range (HSD) Test for variable: GAIN

NOTE: This test controls the type I experiment wise error rate, but generally has a higher type II error rate than REGWQ.

Alpha= 0.05 df= 16 MSE= 2889.7
Critical Value of Studentized Range= 4.046
Minimum Significant Difference= 97.27
Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	TRT
A	285.60	5	4
B	162.80	5	3
B	142.00	5	2
B	87.60	5	1

صندوق ٩-٢

- يقصد بفصل المتوسطات هو اختبار أى من المتوسطات تختلف عن بعضها البعض.
- ومن العديد من الطرق الإحصائية المستخدمة فى فصل المتوسطات والتى تم تناولها فى هذا المؤلف:
  ١- أقل فرق معنوى LSD
  ٢- اختبار توكى HSD
  ٣- اختبار دنكن DMRT
- يعتبر LSD أكثر الطرق الثلاث إظهاراً للفروق المعنوية؛ لذا فهو أقواها ولكنه يرتكب خطأ من النوع الأول (أى رفض فرض العدم بينما يكون صحيحاً) هو الأكبر قيمة بين الثلاثة. لذا وكحماية من هذا النوع من الخطأ يفضل عدم استخدام الـ LSD إلا إذا كان اختبار F فى تحليل التباين معنوياً.
- أما اختبار توكى فهو على النقيض من الـ LSD لذا فهو يرتكب خطأ أكبر من النوع الثانى أى إلى عدم رفض فرض العدم مع أنه خطأ.
- أما دنكن فهو بين الاثنين ويمكن استخدامه حتى لو لم يعط اختبار F فى تحليل التباين قيماً معنوية.

٩-٨ المقارنات المستقلة (المتعامدة) Orthogonal comparisons

كما ذكر سابقاً أنه باستعمال اختبارات المعنوية سواء فى ذلك LSD أو اختبار دنكن فإن هناك عدداً من المقارنات بين المتوسطات أكثر من عدد درجات الحرية الموجودة للاختلافات بين مستويات المعاملات وبالتالى لا يمكن أن تكون هذه المقارنات مستقلة عن بعضها.

سميت المقارنات بالمتعامدة orthogonal نتيجة للمفهوم الرياضى الذى يعنى أنه فى المثلث قائم الزاوية فإن مربع الوتر = مجموع مربعى الضلعين الآخرين

(فيثاغورث). أما فى حالة المثلثات الحادة أو المنفرجة، يكون هناك زيادة أو نقص لمربع الوتر على مربعى الضلعين الآخرين، حيث الزيادة أو النقص تحددها العلاقة بين المسقط وأحد الضلعين، وهذا يعنى عدم استقلال التأثيرات. وهذا إلى حد كبير مفهوم التغاير covariance بين تأثير القوى المختلفة.

فإذا كان هناك تجربة بها عدد t من المعاملات فإنه يوجد $(t-1)$ درجات حرية للفروق بين المعاملات وأيضاً نفس العدد من المقارنات المستقلة فيما بينها أما بالنسبة للعدد الكلى الممكن للمقارنات بين أزواج المعاملات فإنه يصبح $t(t-1)/2$.

والمقارنات المستقلة تمكن من تقسيم مجموع المربعات إلى أجزاء مستقلة عن بعضها كل منها يمثل مقارنة ما لها درجة حرية واحدة. فمثلا إذا كان هناك مجموعة من المعاملات لإضافة مكون أو أكثر بكميات أو نسب مختلفة فإنه يمكن مبدئياً قياس مدى اختلاف المستويات عن بعضها ككل أو مقارنة مستوى بآخر. كما وأنه فى بعض الحالات يكون هناك أكثر من عامل واحد ولكل عامل أكثر من مستوى واحد، وفى هذه الحالة فإن المقارنات المستقلة تمكن من بيان تأثير توليفات combinations مختلفة من العوامل على الناتج النهائى. وفى حالات أخرى فإن المقارنات المستقلة تمكن من معرفة شكل العلاقة هل هى علاقة خطية linear أم أنها علاقة غير خطية non linear or curve linear.

ويتوقف تقديم مجموع المربعات بين المعاملات إلى أجزائه على تكوين توليفة (دالة) خطية linear combination (function) من متوسطات المستويات المختلفة المراد إدخالها فى المقارنة (يفضل استخدام المجاميع بدلا من المتوسطات عند تساوى أعداد المشاهدات فى كل مستوى).

ففى المثال ٩-٣ كان هناك 4 علائق مختلفة يراد معرفة تأثيرها على معدل الزيادة اليومية فى الوزن. ولو افترض أن المعاملات الأربع تمثل عاملين مختلفين هما فيتامين معين وله مستويان (إضافة الفيتامين وعدم إضافته) أما العامل الآخر هو مضاد حيوى وله أيضاً مستويان (إضافة المضاد الحيوى وعدم إضافته) وعلى ذلك فسوف يكون هناك أربع معاملات للمستويات هى عدم إضافة كل من الفيتامين والمضاد الحيوى (معاملة ١)، إضافة الفيتامين فقط (معاملة ٢)، إضافة المضاد الحيوى فقط (معاملة ٣) و إضافة كل من الفيتامين والمضاد الحيوى (معاملة ٤).

عند الرغبة فى تقييم توليفة خطية لدراسة المقارنة بين إضافة أى من المادتين بالمقابل عدم الإضافة على الإطلاق فهذا يعنى مقارنة بين متوسط المعاملة الأولى مع متوسط المعاملات الثلاث الأخرى وتصبح التوليفة الخطية بين المتوسطات الأربعة كالتالى:

$$L=\overline{Y}_1-\left(\frac{1}{3}\overline{Y}_2+\frac{1}{3}\overline{Y}_3+\frac{1}{3}\overline{Y}_4\right)$$

ويلاحظ أن كل من المتوسطات قد دخل فى هذه التوليفة بمعامل coefficient معين، فالمعامل لمتوسط المعاملة (١) هو 1، أما المعامل لمتوسطات المعاملات الثلاث الأخرى هو $-\frac{1}{3}$. ويمكن إعادة كتابة التوليفة السابقة بضرب المعاملات فى 3 للتخلص من الكسور وتصبح:

$$L=3\overline{Y}_1-\left(\overline{Y}_2+\overline{Y}_3+\overline{Y}_4\right)$$

وبذلك تكون المعاملات $(3,-1,-1,-1)$. والصورة العامة التى يمكن بها وضع التوليفة الخطية هى:

$$L=\lambda_1\overline{Y}_1+\lambda_2\overline{Y}_2+\cdots+\lambda_t\overline{Y}_t \qquad (٩-٨)$$

حيث قيم λ_i هى المعامل الخاص بكل متوسط لمستوى $\overline{Y}_i$. وحتى تصبح التوليفة ممثلة لمقارنة تمكن من حساب مجموع مربعات خاص بها فإن مجموع المعاملات لابد من أن يساوى صفراً أى $\sum\lambda_i=0$. ففى المعادلة (٩-١٢) مجموع المعاملات يساوى صفر $\{[(3)+(-1)+(-1)+(-1)]=0\}$. ويمكن التعبير عن التوليفة فى صورة المجاميع للمستويات بدلا من المتوسطات فتصبح:

$$L=\lambda_1T_1+\lambda_2T_2+\ldots+\lambda_tT_t \qquad (٩-٩)$$

وينطبق على هذه المقارنة نفس ما ينطبق على المتوسطات. ويمكن حساب مجموع المربعات الخاص بأى مقارنة من خلال الدالة الخطية الخاصة بها. فإذا استخدمت المتوسطات يصبح مجموع المربعات للمقارنة:

$$\text{Sum of squares due to a comparison}=\frac{nL^2}{\sum\lambda_i^2} \qquad (٩-١٠)$$

أما إذا استخدمت المجاميع فإن مجموع المربعات الخاص بالمقارنة يصبح:

$$\text{Sum of squares due to a comparison}=\frac{L^2}{n\sum\lambda_i^2} \qquad (٩-١١)$$

وهذا ينتج طبعاً عن العلاقة بين المجموع والمتوسط.

وبالطبع من الممكن تكوين أكثر من مقارنة (دالة خطية) بين المتوسطات أو المجاميع كل منها مستقل عن الأخرى ولها مجموع مربعات مستقل. وعدد هذه المقارنات يساوى عدد درجات الحرية بين المستويات، أى (t-1) وحتى تصبح مقارنتان مستقلتين أو متعامدتين orthogonal فإنه لابد من توافر شرطين:

أ- أن مجموع المعاملات لكل من المقارنتين يساوى صفراً أى: $\sum_{i} \lambda_i = 0$.

ب- أن مجموع حاصل ضرب المعاملات لكل من المقارنتين يساوى صفراً أى $\sum \lambda_i \lambda_j = 0$ حيث λ_i وهو المعامل للمتوسط $\overline{Y}_i$ فى المقارنة الأولى فى حين أن λ_j هو المعامل لنفس المتوسط فى المقارنة الثانية. وبناء على هذين الشرطين، يمكن تقسيم أو تجزئة مجموع المربعات بين المستويات إلى أجزاء مستقلة أو متعامدة خاصة بمقارنة ما.

ومن ذلك يتضح أنه إذا أجريت جميع المقارنات المستقلة الممكنة بين المستويات والتى عددها $(t-1)$ وجمعت هذه الأجزاء فإنها تعطى مجموع المربعات بين المعاملات دون تداخل فيما بينها.

عند تطبيق ما سبق ذكره على المثال ٩-٣ فإنه يمكن حساب عدد من المقارنات بين المجاميع أو المتوسطات الخاصة بالمعاملات عددها يساوى عدد درجات الحرية وهو 3 = 1-4 درجات حرية أى 3 مقارنات مستقلة ممكنة. المقارنة الأولى التى سبق ذكرها وهى اختلاف مجموعة المقارنة control عن المعاملات التى أضيف فيها الفيتامين والمضاد الحيوى. وباستخدام القيم الموجودة فى هذا المثال وتطبيق المعادلة (٩-٨) فإن: $L_1 = 3(87.6) - (142 + 162.8 + 285.6) = -327.6$.

وبتطبيق المعادلة (٩-١٠) يصبح مجموع المربعات الخاص بهذه المقارنة:

$$\frac{(5)(-327.6)^2}{(3)^2 + (-1)^2 + (-1)^2 + (-1)^2} = 44717.4$$

وحيث إن مجموع المربعات للمقارنة مرتبط بدرجة حرية واحدة خاصة بها وعلى ذلك فإن متوسط المربعات لنفس المقارنة يصبح نفس القيمة (القسمة على واحد) وبالتالى يمكن اختبار معنوية هذه المقارنة بواسطة اختبار F بقسمة متوسط المربعات الخاص بها على متوسط مربعات الخطأ من جدول ٩-٣ أى $F = 44717.4 / 2889.7 = 15.5$. وهذا الاختبار يقارن بقيمة F الجدولية والتى قيمتها $F_{(1,16,0.01)} = 8.53$، وهذا يعنى أن الفرق بين نمو الحملان المعاملة والحملان فى مجموعة المقارنة control كان معنوياً.

المقارنة الثانية والتى يمكن اختبارها قد تكون الفرق بين إضافة أى من الفيتامين أو المضاد الحيوى منفردين وبين إضافتهما معا. وفى هذه الحالة يتم استبعاد مجموعة المقارنة control وتصبح المقارنة كالتالى: $L_2 = 2Y_4 - (Y_2 + Y_3)$، وقيمتها: $1332 = [(2)(1428) - (710 + 814)]$، ويكون مجموع المربعات الخاص بها:

$$\frac{(1332)^2}{(5)[(2)^2 + (-1)^2 + (-1)^2]} = 59140.8$$

يلاحظ هنا استخدام المجاميع بدلاً من المتوسطات وبالتالى تطبيق المعادلة (٩-١١) للحصول على مجموع المربعات. ويتم اختبار هذه المقارنة بنفس الطريقة أى أن: $F = 59140.8 / 2889.7 = 20.5$ وهى معنوية أيضاً عند 1%

وتطبيقاً للشروط الخاصة بكون هاتين المقارنتين مستقلتين orthogonal، يتم تكوين جدولاً يوضع فيه المجاميع والمعاملات كما فى جدول٩-٥.

جدول ٩-٥ المعاملات للمقارنات المختلفة المستقلة

مجموع المربعات للمقارنة	$n\Sigma\lambda_i^2$	$\Sigma\lambda_i^2$	$\Sigma\lambda_i$	المعاملة ٤	٣	٢	١	
	-	-	-	1428	814	710	438	المجموع
44717.4	60	12	0	-1	-1	-1	3	المقارنة الأولى Comparison 1
59140.8	30	6	0	2	-1	-1	0	المقارنة الثانية Comparison 2
1081.6	10	2	0	0	1	-1	0	المقارنة الثالثة Comparison 3
104939.8								

وكما ذكر سابقاً فإن عدد المقارنات المستقلة يساوى عدد درجات الحرية بين المستويات، وبالتالى فقد تم عمل مقارنتين. أما المقارنة الأخيرة والثالثة كما تظهر فى الجدول هى مقارنة بين إضافة كل من الفيتامين بمفرده أو إضافة المضاد الحيوى بمفرده. وقد وجد أن مجموع المربعات المحسوب لهذه المقارنة 1081.6 وعند إجراء اختبار F لهذه المقارنة $1081.6 / 2889.7 = 0.37$، أى أنها كانت غير معنوية.

ومن الواضح من جدول ٩-٦ أن مجموع المربعات للمقارنات الثلاث تجمع مجموع مربع الانحرافات المحسوب بين المعاملات فى الجدول ٩-٣ ويرجع ذلك إلى أن كلاً منها مستقلة عن الأخرى. كما يتضح أن مجموع حاصل ضرب العوامل لكل مقارنتين يساوى صفراً فمثلا بالنسبة للمقارنتين (1, 2):

$$(3)(0)+(-1)(-1)+(-1)(-1)+(-1)(2)=0$$

وبالنسبة للمقارنتين (1, 3):

$$(3)(0)+(-1)(-1)+(-1)(1)+(-1)(0)=0$$

وهكذا ...

ويمكن إعادة كتابة جدول ٩-٣ الخاص بمثال ٩-٣ بصورة أكثر تفصيلا فى جدول ٩-٦ كما يلى:

جدول ٩-٦ جدول تحليل التباين مع الإشارة إلى مجموع المربعات للمقارنات المستقلة

SOV	df	SS	MS
TRT	3	104939.8	34979.9^{**}
Comparison 1	1	44717.4	44717.4^{**}
Comparison 2	1	59140.8	59140.8^{**}
Comparison 3	1	1081.6	1081.6
Error	16	1081.6	2889.7
C - Total	**19**	**151175.0**	

ويطلق على اختبار F فى هذه الحالة لاختبار معنوية المقارنات، اختبار F المستقل orthogonal F test.

كما يمكن أيضا استخدام اختبار t المستقل orthogonal t test. ولإجراء اختبار t المستقل تستخدم المعادلة التالية:

$$t=\frac{\left|\Sigma\lambda_i\overline{X}_i\right|}{\sqrt{(\Sigma\lambda_i^2)\frac{S^2}{n}}} \qquad (٩-١٢)$$

حيث S^2 من جدول ٩-٣ تساوى 2889.72 ، $n = 5$ وبالتالى تكون t للمقارنة الأولى:

$$t = \frac{|(3)(87.6) + (-1)(142) + (-1)(162.8) + (-1)(285.6)|}{\sqrt{\left[(3)^2 + (-1)^2 + (-1)^2 + (-1)^2\right]\frac{2889.7}{5}}} = \frac{|-327.6|}{83.28} = 3.93$$

و t للمقارنة الثانية:

$$t = \frac{|(2)(285.6) - (142 + 162.8)|}{\sqrt{\left[(2)^2 + (-1)^2 + (-1)^2\right]\frac{2889.7}{5}}} = \frac{266.4}{58.89} = 4.52$$

وهكذا بالنسبة للمقارنة الثالثة.

وتقارن قيمة t المحسوبة بقيمة t الجدولية التى قيمتها $t_{(16,\ 0.01)} = 2.92$ وعليه فإن الفرق للمقارنة الأولى معنوى، أى أن الفرق بين متوسط المعاملة الأولى مع متوسط المعاملات الثلاث الأخرى مؤكد وليس راجع للصدفة بدرجة ثقة 99%، وأيضاً الفرق بين إضافة أى من الفيتامين أو المضاد الحيوى وبين إضافتهما معاً أيضاً معنوى ... وهكذا.

ويلاحظ أن قيمة F للمقارنة الأولى تساوى مربع قيمة t لنفس المقارنة وكذلك العلاقة بين القيمة F وقيمة t لكل من المقارنة الثانية والمقارنة الثالثة وذلك لوجود درجة حرية واحدة فى البسط فى اختبار F حيث $F = t^2$. ويلاحظ أيضا أن النتائج متطابقة باستخدام اختبار t أو اختبار F فى هذه الحالة فيما عدا خطأ التقريب.

ويلاحظ أنه فى اختبار t المستقل يختار المجرب مجموعة مقارنات مستقلة واحدة فقط والتى تعكس ما يريده ويفضل تصميم مثل هذه المقارنات عند بدء التجربة وليس بعد الحصول على النتائج وحساب المتوسطات حتى لا يحصل المجرب على نتائج متحيزة.

٩-٨-١ المقارنات المستقلة باستخدام برنامج SAS

يمكن استخدام برنامج SAS لعمل المقارنات المستقلة باستخدام اختيار contrast مع PROC GLM أى طريقة النموذج الخطى العام والتى سوف يتم شرحها فيما بعد. وباستخدام بيانات مثال ٩-٣ يمكن كتابة البرنامج كالتالى:

```
DATA LAMB;
INPUT TRT GAIN @@;
*---- TRT 1: CCONTROL, 2: VET, 3:AB, 4:VET AND AB ----;
*---- CONTROL: NO VET. ------------;
CARDS;
1 110 1 98 1 84 1 42 1 104 2 122 2 224 2 60 2 178 2 126
3 84 3 194 3 162 3 190 3 184 4 338 4 274 4 338 4 170 4 308
PROC GLM;
CLASS TRT;
MODEL GAIN = TRT;
CONTRAST 'CONTROL VS OTHER' TRT 3 -1 -1 -1;
*----- MEANS TRT1 VERSUS OTHER TREATMENTS-----;
CONTRAST 'VETORAB VS BOTH' TRT 0 -1 -1 2;
*----- MEANS TRT1AND TRT3 VERSUS TRT4 --------;
CONTRAST 'VET VS AB' TRT 0 -1 1 0;
*----- MEANS TRT2 VERSUS TRT3 -----------;
RUN;
```

لاحظ:

تم استخدام proc glm بدلاً من proc anova حيث إنه لا يمكن استخدام اختيارcontrast مع proc anova.

يوضع اختيار contrast يتبعه اسم للمقارنة المرغوب عملها، ويمكن وضع أى اسم ذى معنى للمجرب. يلى اسم المقارنة المعاملات الخاصة بتلك المقارنة.

النتائج:

```
              General Linear Models Procedure
                 Class Level Information
                 Class    Levels    Values

                  TRT       4       1 2 3 4
          Number of observations in data set = 20
              General Linear Models Procedure

Dependent Variable: GAIN
                                Sum of        Mean
Source                  DF     Squares       Square     F Value   Pr > F
TRT                      3   104939.800    34979.933     12.11    0.0002
CONTROL VS OTHER         1    44717.400    44717.400     15.47    0.0012
VETORAB VS BOTH          1    59140.800    59140.800     20.47    0.0003
VET VS AB                1     1081.600     1081.600      0.37    0.5493
Error                   16    46235.200     2889.700
Corrected Total  19    151175.000
```

٩-٩ تحليل التباين عند عدم تساوى التكرارات

كثيرا لا تتساوى أعدد المكررات أو المشاهدات فى العينات أو المعاملات المختلفة فى حالة البيانات التى يتم تجميعها من مصادرها. أما فى حالة التجارب التى يخطط لها ويقوم بها المجرب فى كثير من المجالات فإن العينات تكون غالبا متساوية من حيث المكررات.

ففى حالة تساوى المعاملات فى أعداد المشاهدات بها فإن طريقة الحساب التى استخدمت للحصول على مجاميع المربعات فى فصل ٩-٦ كانت بقسمة مجموع مربعات مجاميع المعاملات على عدد المكررات داخل أى معاملة، وهو ما كان متساوياً، مطروحاً منه معامل التصحيح، وأيضا تم حساب درجات الحرية للخطأ من اعتبار أن قيمة n متساوية وبالتالى كان العدد يساوى $t(n-1)$ كما فى جدول ٩-١.

أما فى حالة عدم تساوى العينات فى مكرراتها فيصبح العدد الكلى للمشاهدات هو المجموع الكلى لها بتجميع عددها فى كل معاملة أى يصبح $\sum_i n_i$ حيث تمثل n_i عدد المشاهدات فى المعاملة i. وبما أن عدد المعاملات لا يتغير وهو t فإن عدد درجات الحرية للمجموع وداخل المعاملات يصبح$\left(\sum n_i - 1\right)$ ، $\left(\sum n_i - t\right)$ على الترتيب. فإذا افترض أن $N = \sum n_i$ عندئذ يصبح عدد درجات الحرية الكلى فى التجربة $N-1$ وعدد درجات الحرية داخل المعاملات (الخطأ) $N-t$.

وكما ذكر مسبقا فإن اختلاف عدد المعاملات فى تحليل التباين للبيانات أحادية التقسيم لا يؤثر على خاصية التجميع ولكن يلاحظ أن طريقة الحسابات تختلف قليلا كما يلى:

معامل التصحيح $CF = \dfrac{(Y..)^2}{N}$ ، حيث $N = \sum n_i$

$$TSS = \sum_i \sum_j Y_{ij}^2 - CF \qquad (٩-١٣)$$

وهو مجموع المربعات الكلى. وبالتالى مجموع المربعات بين المعاملات يكون:

$$BSS = \sum_i \frac{Y_{i.}^2}{n_i} - CF \qquad (٩-١٤)$$

وبالطبع وحيث لا تتأثر خاصية التجميع فى هذا النوع من التحليل يمكن الحصول على مجموع المربعات داخل المعاملات (الخطأ) WSS بالطرح كما سبق فى المعادلة (٩-٩) أو ، بطريقة مباشرة:

$$WSS = \sum_i \sum_j Y_{ij}^2 - \sum_i \frac{Y_{i.}^2}{n_i} \qquad (١٥-٩)$$

والمعادلة الأخيرة مهمة حيث توضح مفهوم الخطأ. وكثيراً ما يكون هناك احتياج إلى استخدامها عندما يراد حساب الخطأ مباشرة. ويلاحظ أن خاصية التجميع تنعدم فى الحالات الأكثر تعقيداً كما سيأتى فيما بعد. كما أن اختبار F فى هذه الحالة لا يتأثر حيث إن متوسط المربعات بين المعاملات له $(t-1)$ درجات حرية ومتوسط المربعات للخطأ له $(N-t)$ درجات حرية. أما التعديلات التى يجب الإشارة إليها فى هذا المجال فهى خاصة بإجراء اختبارات المعنوية للفروق بين كل متوسطين إما بواسطة اختبار LSD أو بواسطة اختبار دنكن وفيما يلى توضيح التعديلات المطلوبة فى كل حالة.

أولا: فى حالة اختبار LSD حيث يلزم تقدير الانحراف المعيارى للفرق بين متوسطى أى معاملتين فى التجربة وبفرض المعاملتين k, i فإن الخطأ القياسى لمقارنة متوسطى المعاملتين هو:

$$S_{(\overline{Y}_i - \overline{Y}_k)} = \sqrt{S^2\left(\frac{1}{n_i} + \frac{1}{n_k}\right)} \qquad (١٦-٩)$$

حيث S^2 هو متوسط المربعات للخطأ كالسابق و n_i ، n_k هما عدد المشاهدات أو التكرارات لكل من المعاملتين i و k على التوالى. وبالتالى تصبح قيمة $LSD = (t)S_{(\overline{Y}_i - \overline{Y}_k)}$ وتتغير هذه القيمة بتغير المتوسطات للمعاملات المقارنة أما درجات الحرية الخاصة بقيمة t فهى لا تزال درجات حرية الخطأ فى جدول تحليل التباين $(N-t)$.

ثانيا: فى حالة إجراء اختبار دنكن سبق إيضاح أن قيمة الاختبار DMRT هى $(SSR)(S_{\overline{Y}})$ وحيث إنه لا توجد قيمة ثابتة للخطأ القياسى $S_{\overline{Y}}$ وذلك لتغير التكرارات وعلى ذلك فإنه لمقارنة متوسطى معاملتين مختلفتين فى تكراراتهما تستخدم قيمة S^2 فى المعادلة. وبذلك يصبح الاختبار كالتالى:

$$DMRT = SSR\sqrt{\frac{S^2}{2}\left(\frac{1}{n_i} + \frac{1}{n_k}\right)} \qquad (١٧-٩)$$

وسبب القسمة على $\sqrt{2}$ يرجع إلى وجود الكمية داخل القيم الجدولية كما سبق إيضاحه.

مثال ٩-٦

أجريت تجربة من 3 معاملات أ، ب، ج على الحملان وكانت أوزان الحملان بالكيلوجرام فى المعاملات الثلاث كالتالى:

المعاملة	المكررات					المتوسط
أ	32.5	32.5	33.0			**32.67**
ب	35.0	34.0	34.5	37.0	33.5	**34.80**
ج	30.5	31.5	31.0	31.5		**31.12**

جدول تحليل التباين

SOV	df	SS	MS	F
TRT	2	30.85	15.29^{**}	1.96
Error	9	8.15	0.906	
C - Total	**11**	38.73		

وعند إجراء اختبار LSD بين المتوسطات تكون المقارنات كالتالى:

قيمة اختبار LSD	المقارنة	
	الخطأ القياسى	القيمة
$(.727)(2.262) = 1.644$ الفرق غير معنوى	$\sqrt{0.906\left(\frac{1}{3}+\frac{1}{4}\right)} = 0.727$	أ − ج = 1.55
$(.695)(2.262) = 1.57$ الفرق غير معنوى	$\sqrt{0.906\left(\frac{1}{5}+\frac{1}{3}\right)} = 0.695$	ب − أ = 2.13
$(.639)(2.262) = 1.45$ الفرق معنوى	$\sqrt{0.906\left(\frac{1}{4}+\frac{1}{3}\right)} = 0.639$	ب − ج = 3.68

وعند إجراء اختبار الفروق باستخدام اختبار Duncan، ترتب الفروق بين متوسطات المعاملات تنازليا فتكون كالتالى 31.12, 32.67, 34.8 للمعاملات ب، أ، ج، على التوالى. يحوى المدى المختبر للفرق بين المعاملتين ب، ج 3 متوسطات

وبالتالى قيمة SSR عند 9 درجات حرية للخطأ تساوى 3.34 وتكون قيمة اختبار دنكن على مستوى معنوية 5% كالتالى:

$$SSR = 3.34\sqrt{\frac{0.906}{2}\left(\frac{1}{5}+\frac{1}{4}\right)} = 1.508$$

وبمقارنته بالفرق بين المتوسطين ب، ج وقدره 3.68 kg يتضح أن الفرق معنوى كما سبق إيضاحه.

ويجدر الإشارة هنا أنه فى حالة عدم تساوى عدد الوحدات التجريبية فى المعاملات ممكن أن يحدث فرق أكبر بين متوسطين فى أحدهما أو كليهما أعداد قليلة تكون غير معنوية بالمقارنة بفرق أصغر ولكن بين متوسطين فى أحدهما أو كليهما أعداد أكبر، وفى هذه الحالة لا تصلح طريقة ربط المتوسطات غير معنوية الفرق بخط واحد ولكن تستعمل الحروف للدلالة على معنوية الفروق.

٩-١٠ النموذج التجميعى الخطى The linear additive model

بعد تقديم كيفية إجراء تحليل التباين وماهيته وبعض الافتراضات الخاصة بالأخطاء errors وتوزيعها فإنه يوجد بعض الافتراضات الخاصة بتأثير المعاملات أيا كان نوعها. وكما سبق ذكره، فإن النموذج الرياضى لأى مشاهدة يتكون من تجميع ثلاثة أجزاء وهى المتوسط العام μ وانحراف متوسط المعاملة عن المتوسط العام α_i وأخيرا الخطأ الطبيعى للمشاهدة حول متوسط المجموعة المنتمية إليها. وكما سبق يوجد نوعان من الافتراضات الخاصة بأثر المعاملة وهما إما أن يكون أثر المعاملة ثابتاً fixed أو أن يكون أثرها عشوائياً random.

أولا: التأثير الثابت Fixed effect (Model I)

فى هذه الحالة يفترض أن مجموع التأثيرات $\sum\alpha_i = 0$، وأنه لو أعيدت التجربة فإنه يتوقع الحصول على نفس المجموعة من التأثيرات فى التجربة الجديدة. ويكون النموذج الرياضى بالتالى وكما سبق ذكره فى (٩-٦) هو:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \qquad (٩-١٨)$$

ثانيا: التأثير العشوائى Random effect (Model II)

إذا كان تأثير المعاملات a_i عشوائياً فإن ذلك يعنى أن المستويات التى فى التجربة تعتبر عينة مأخوذة من عشيرة كبيرة لا نهائية من تلك المستويات، وأن متوسط هذه العشيرة هو صفر، حيث إنها أيضا تعتبر انحرافات عن المتوسط العام μ.

وفى هذه الحالة فإنه عند إعادة التجربة يتم الحصول على عينة جديدة من عشيرة المستويات هذه.

وبما أن مجموعة المستويات فى هذا النموذج مأخوذة عشوائياً من عشيرة المستويات والذى متوسطه يساوى صفرا فإن المستويات أيضاً لها تباين σ_a^2 وهو ما يراد أيضا تقدير قيمته من تحليل نتائج التجربة.

بمعنى أنه فى هذه الحالة لا يكون الاهتمام بإيجاد اختبار لمدى معنوية التأثيرات لمجموعة محددة من المستويات كما فى النموذج الأول، ولكن الاهتمام يكون أكثر بتقدير مدى ما تشارك به هذه المستويات فى التباين الكلى حيث إنها عنصر من عناصر التباين بين المشاهدات المختلفة. وعلى هذا الأساس يصبح متوسط المربعات mean squares عبارة عن جزئين، الأول راجع للتباينات الناتجة من عشيرة المستويات المستخدمة فى التجربة σ_a^2 والثانى هو التباين الراجع لاختلاف المشاهدات داخل المستوى الواحد من المعاملات وهى σ_e^2 أو الخطأ التجريبى.

وبذا يصبح التباين بين مشاهدتين مأخوذتين عشوائيا من نفس المستوى (المعاملة) يرجع إلى التباين الداخلى σ_e^2 (الخطأ) فى حين أن التباين بين مشاهدتين مأخوذتين عشوائيا من مستويين مختلفين يرجع إلى كلا التباينين أى $\sigma_e^2 + \sigma_a^2$. وبذلك فإنه فى حالة النموذج العشوائى Model II فإنه يفترض أن الخطأ $\in_{ij}$ يتوزع طبيعيا بمتوسط صفر وتباين σ_e^2 ويرمز لذلك $\in \sim NID(0,\sigma_e^2)$ وأيضا فإن مستويات المعاملات a_i تتوزع طبيعيا بمتوسط صفر وتباين σ_a^2 ويرمز لذلك $a \sim NID(0\ ,\ \sigma_a^2)$. ويكون النموذج فى هذه الحالة:

$$Y_{ij} = \mu + a_i + \in_{ij} \qquad (١٩-٩)$$

وهناك أوجه للشبه والاختلاف بين كل من النموذجين يمكن تلخيصها فيما يلى:

أ- فى النموذج I يفترض أن α_i تأثيرها ثابت وبالتالى فإن $\sum\alpha_i = 0$، وكذلك $\in \sim NID(0,\sigma_e^2)$، ويراد تقدير قيمتها واختبار معنوية الفروق بين هذه المجموعة المحددة من المستويات.

ب- فى النموذج II يفترض أن a_i تأثيرها عشوائى وبالتالى $a \sim NID(0\ ,\ \sigma_a^2)$ وكذلك $\in \sim NID(0,\sigma_e^2)$، وبالتالى فإن هناك اهتماماً واضحاً تجاه تقدير حجم σ_a^2.

ج- اختبار فرض عدم $H_0: \mu_i = \mu$ يتماثل مع اختبار الفرض بأن $H_0: \sigma_a^2 = 0$ حيث إن في كلتا الحالتين α_i أو a_i لابد أن تساوى صفراً، وأن اختبار F يعتبر صحيحاً في كلتا الحالتين لفرض العدم.

د- متوسط المربعات S^2 داخل المعاملات يعتبر تقديراً لقيمة σ_e^2 في العشيرة، في حين أنه يمكن إثبات أنه في حالة اختلاف المتوسطات μ_i فإن متوسط المربعات بين المعاملات في النموذج I يعتبر تقديراً غير منحاز للكمية $\sigma_e^2 + n\sum\alpha_i^2/(t-1)$، حيث كما سبق $\alpha_i = \mu_i - \mu$، أى أنه بالإضافة إلى التباين العادى بين الأفراد المعاملة بنفس المعاملة فهناك تأثير اختلاف متوسطات المعاملات والذى يتم تربيعه ثم يحسب متوسطه بالقسمة على درجات الحرية $(t-1)$ وبالطبع فإن الكمية α_i هى كمية ثابتة فى هذا النموذج. أما بالنسبة للنموذج II فإن متوسط المربعات للخطأ هو نفسه في حالة النموذج I، أما متوسط المربعات بين المعاملات فهو تقدير للكمية $\sigma_e^2 + n\,\sigma_a^2$ حيث σ_a^2 هى مكون التباين بين المعاملات. وحيث n تمثل عدد الأفراد أو المشاهدات داخل كل مستوى فإنه يمكن حساب تقدير σ_a^2 من S_a^2 من تحليل التباين كالتالى:

مكون التباين بين المعاملات:

$$S_a^2 = \frac{\text{MS between} - \text{MS within}}{n} \qquad (٩-٢٠)$$

وبالتعويض بالقيم المتوقعة لكل من متوسطى المربعات تكون معادلة التوقع:

$$S_a^2 = \frac{(\sigma_e^2 + \sigma_a^2) - \sigma_e^2}{n} \cong \sigma_a^2 \qquad (٩-٢١)$$

وبالتالى فإن S_a^2 المحسوبة من التحليل تعتبر تقديراً لقيمة σ_a^2 لمكون التباين بين المجموعات.

وفى المثال المذكور لتحليل التباين فى جدول ٩-١ وبافتراض أن الاختلافات بين العلائق المختلفة تأثيرها عشوائى، أى النموذج الرياضى II، فإن التقدير المتحصل عليه لتباين الخطأ $\sigma_e^2 = 2.88$، متوسط المربعات بين المعاملات 19.86 وهو تقدير للقيمة $\sigma_e^2 + n\,\sigma_a^2$ وحيث $n = 5$، عدد الحملان داخل كل عليقة، وعلى ذلك فإن تقدير مكون التباين الراجع لأثر العليقة $\sigma_a^2 = \frac{19.86 - 2.88}{5} = 3.396$. أما إذا كان تأثير المعاملات ثابتاً (نموذج I) فإن توقع متوسط المربعات يكون $\sigma_e^2 + nK_a^2$.

٩-١١ تقدير مكونات التباين فى حالة اختلاف حجم العينات

فى كثير من الحالات والتى يكون التعامل فيها مع بيانات فى مجالات خاصة مثل تربية الحيوان أو الدواجن أو الأسماك أو فى دراسة النواحى الوراثية والاجتماعية فى حالة المجتمعات الإنسانية، أو الدراسات الوراثية فى النبات، يكون حجم العائلات (أى عدد أفراد العائلة) فى معظمها مختلفا من عائلة إلى أخرى وهذه العائلات التى تدخل فى العينة تعتبر جزءاً عشوائياً من عشيرة كبيرة من تلك العائلات. وبالتالى فإن الاختلافات فيما بينها سواء كانت وراثية أو اجتماعية تعتبر نموذجاً لما تعني به التأثيرات العشوائية، أى النموذج الرياضى II. والذى يمكن التعبير عنه رياضياً كما يلى:

$$Y_{ij} = \mu + f_i + \in_{ij} \quad , i = 1,2,...,t \quad j = 1,2,...,n$$

وبالتالى تكون افتراضات تحليل التباين كما سبق ($f \sim NID(0,\sigma_f^2)$، $\in \sim NID(0,\sigma_e^2)$) حيث تمثل f_i فى هذه الحالة الاختلافات العشوائية الموجودة بين العائلات سواء كانت وراثية كنتيجة لتأثير العوامل الوراثية الموروثة من الأب (الطلوقة) كما فى دراسات تربية الحيوان، أو أنها تأثيرات اجتماعية للعائلة ككل نتيجة لعوامل تعليمية أو ثقافية أو اقتصادية مختلفة كما فى الدراسات الإنسانية، وقد تكون نتيجة لاختلاف التراكيب الوراثية للخطوط المختلفة lines كما فى دراسات تربية النباتات.

وحيث إن حجم العائلات أو الخطوط يختلف من حالة إلى أخرى بحيث إن عدد المشاهدات المدروسة فى العائلة f_i هو n_i وبالتالى فإن حجم التجربة كلها $N = \sum n_i$ ولا يؤثر ذلك على تحليل التباين كما سبق إيضاحه فى حالة لنموذج I. ولكن الاختلاف هنا هو أن متوسط المربعات بين المجموعات (العائلات) يعتبر تقديراً غير منحاز للكمية $\sigma_e^2 + n_\circ \sigma_f^2$ حيث $n_\circ$ فى هذه الحالة هى المتوسط التوافقى لعدد أفراد المجموعة الواحدة وهو يساوى:

$$n_\circ = \frac{1}{t-1}\left(N - \frac{\sum n_i^2}{N}\right) \qquad (٩-٢٢)$$

وبناء على ذلك فإنه يمكن تقدير قيمة مكون التباين بين المجموعات كما سبق باستخدام المعادلة (٩-٢٠) مع استبدال قيمة n فى المقام بقيمة $n_\circ$ المحسوبة من المعادلة (٩-٢٢) كالتالى:

$$S_a^2 = \frac{\text{MS between} - \text{MS within}}{n_\circ} \qquad (٩-٢٣)$$

ويعتبر التقدير المتحصل عليه لمكون التباين بين المجموعات σ_a^2 تقديراً غير منحاز إلا أنه فى حالة عدم تساوى العينات لا يعتبر بنفس الكفاءة كما هو الحال فى حالة التساوى. ولذا فإنه ينصح، إذا كان بالإمكان، استخدام عينات ذات أحجام متساوية وذلك لسهولة معرفة التوزيع الرياضى لها فى هذه الحالة.

مثال ٩-٧

الجدول التالى يوضح إنتاج البيض الناتج لمجاميع من الدجاج معبراً عنه كنسبة مئوية حيث كل مجموعة تمثل بنات ديك واحد أى أنصاف شقيقات (قد لا تعتبر النسب المئوية مثالاً جيداً لتحليل التباين نظراً لأنه عند عدم تساوى العينات فإن اعتبار أن σ_e^2 ثابتة أو موحدة لكل المعاملات قد لا يكون صحيحاً تماماً إلا أنه نظراً لأن أغلب النسب تقع فى المدى من 25-75% فإن الفروق تكون ضئيلة، كما سيتضح ذلك عند دراسة التحويرات transformation فيما بعد).

رقم الديك	النسبة المئوية لإنتاج البيض (للبنات)							عدد البنات
١	59.7	60.0	70.5	66.4				4
٢	76.2	69.3	66.6	73.5				4
٣	62.5	71.3	52.3	65.0	73.5			5
٤	75.6	58.0	71.0	49.0	69.0			5
٥	64.3	57.3	44.3	62.5	59.5	50.0		6
٦	62.7	50.0	69.6	34.0	69.0	52.3		6
٧	45.0	58.3	69.1	72.9	77.8	75.2		6
٨	71.8	58.4	24.7	55.7	50.0	68.3	57.5	7
٩	66.8	46.3	50.5	68.0				4
إجمالى								47

وبالطبع فإن الاختلاف بين ما يورثه الآباء للأبناء يعتبر كميات عشوائية مأخوذة من عشيرة كبيرة من هذه الاختلافات الوراثية وبالتالى فإنه باستخدام النموذج الثانى II:

$$Y_{ij} = \mu + s_i + \epsilon_{ij} \quad , i = 1,2,...,9 \;\; j = 1,2,...,n_i$$

حيث تعتبر s_i انحراف وراثة الأب i عن المتوسط العام μ، $\in_{ij}$ هو الاختلافات الطبيعية بين الأفراد والتى ترجع إلى أسباب وراثية وأخرى بيئية. وكلا الكميتين عشوائيتان أى $s \sim NID(0, \sigma_s^2)$، $\in \sim NID(0, \sigma_e^2)$ ويجري التحليل الإحصائى لهذه البيانات للحصول على المعلومات التالية التى يوضحها جدول تحليل التباين التالى:

ANOVA table

SOV	df	SS	MS	EMS
Between parents	8	1320.65	165.04	$\sigma_e^2 + n_\circ \sigma_s^2$
Within parents	38	4685.39	123.30	σ_e^2
C - Total	**46**	**6006.04**		

EMS: Expected mean squares

وحيث إنه يكون هناك اهتمام فى هذه الحالة بتقدير إلى أى مدى توجد تباينات راجعة إلى تأثير اختلاف الآباء فإن ذلك يستوجب الحصول على تقدير لمكون التباين σ_s^2 الخاص بتأثير الأب وبالتالى يجب حساب n_o طبقا للمعادلة (٩-٢٧):

$$n_o = \frac{1}{8}\left[47 - \frac{(4^2 + 4^2 + ... + 7^2 + 4^2)}{47}\right] = 5.197$$

وهذه لا تختلف كثيراً عن قيمة $\overline{n} = 5.22$ ولكنها أدق. وبالتعويض فى (٩-٢٣) يمكن الحصول على تقدير σ_s^2 كالتالى:

$$\sigma_s^2 = \frac{165.04 - 123.3}{5.197} = 8.024$$

علماً بأن التقدير المحسوب لمكون التباين داخل المجموعات هو نفسه متوسط المربعات داخل المجموعات، أى: $\sigma_e^2 = 123.3$.

مثال ٩-٨

استخدام برنامج SAS لحل بيانات مثال ٩-٧ والحصول على مكونات التباين.

```
DATA EX10111;
INPUT SIRE EGG @@;
CARDS;
1 59.7 1 60 1 70.5 1 66.4 2 76.2 2 69.3 2 66.6 2 73.5
3 62.5 3 71.3 3 52.3 3 65 3 73.5 4 75.6 4 58 4 71 4 49 4 69 5 64.3
5 57.3 5 44.3 5 62.5 5 59.5 5 50 6 62.7 6 50 6 69.6 6 34 6 69 6 52.3
7 45 7 58.3 7 69.1 7 72.9 7 77.8 7 75.2 8 71.8 8 58.4 8 24.7 8 55.7
8 50 8 68.3 8 57.5 9 66.8 9 46.3 9 50.5 9 68
PROC VARCOMP METHOD = TYPE1;
CLASS SIRE;
MODEL EGG = SIRE;
RUN;
```

لاحظ:

استخدمت proc varcomp للحصول على مكونات التباين للمتغيرات العشوائية في النموذج، في هذه الحالة مكون الأب والخطأ التجريبي.

استخدمت method = type 1، وهي التي تحدد طريقة حساب مجموع المربعات حيث يوجد طرق أخرى مثل MIVQUE0 و ML (maximum-likelihood). شرح هذه الطرق والاختلافات بينها يقع خارج نطاق هذا المؤلف.

النتائج:

```
             Variance Components Estimation Procedure
                     Class Level Information

                   Class    Levels    Values

                   SIRE        9      1 2 3 4 5 6 7 8 9

             Number of observations in data set = 47

Dependent Variable: EGG

Source              DF      Type I SS        Type I MS

SIRE                 8    1320.65378723    165.08172340

Error               38    4685.73600000    123.30884211

Corrected Total     46    6006.38978723
```

Source	Expected Mean Square
SIRE	Var(Error) + 5.1968 Var(SIRE)
Error	Var(Error)

Variance Component	Estimate
Var(SIRE)	8.03817982
Var(Error)	123.30884211

٩-١٢ معامل الارتباط الداخلى (الجوانى) Intra-class correlation

فى تحليل التباين أحادى التقسيم حيث يوجد عدد من الأفراد فى كل فئة من الفئات وعندما يكون مكون التباين بين المجموعات أكبر من الصفر $(\sigma_a^2 > 0)$ فإن هذا يدل على أن أفراد المجموعة أو الفئة الواحدة تتشابه مع بعضها. وفى النموذج II لتحليل التباين فإن عدم ثبوت فرض العدم الخاص بأن $\sigma_a^2 = 0$ يوضح أن هناك سبباً للتشابه بين أفراد المجموعة الواحدة أكثر من أفراد لا تنتمى لنفس المجموعة.

وقد افترض أن كل قيم المشاهدات Y_{ij} لها نفس المتوسط μ ونفس التباين σ_e^2، ولكن أى فردين من نفس الفئة i يكون بينهما معامل ارتباط معين وهذا المعامل يطلق عليه اسم معامل الارتباط الداخلى (الجوانى) Intra – class correlation وهو يقيس مدى التشابه بين الأفراد المنتمية إلى فئة معينة لسبب ما ولكن دون أن تكون هناك علاقة سببية (صفة معينة تؤدى إلى أن يميز أحد الأفراد بوضع يختلف عن الأفراد الآخرى).

فمثلاً لا يوجد سبب عند قياس التشابه بين الأخوة الأشقاء حيث لا يؤثر أحدهما على الآخر كما هو الحال فى حالات أخرى مثل التشابه بين الأب وابنه حيث يمرر الأب نصف مورثاته إلى النسل مباشرة.

ويمكن تقدير معامل الارتباط الداخلى بالقيمة r_I من جدول تحليل التباين حيث إنه يمكن جبرياً إثبات أن متوسط المربعات بين الفئات BMS هو تقدير للكمية:

$$\sigma^2\left[1+(n-1)\rho_I\right] \qquad (٩-٢٤)$$

وأن متوسط المربعات داخل الفئات WMS فى هذه الحالة هو تقدير للكمية:

$$\sigma^2(1-\rho_I) \qquad (٩-٢٥)$$

حيث n عدد الأفراد داخل الفئة وأن σ^2 هو التباين للأفراد و ρ هو معامل الارتباط الجوانى بين الأفراد وبالتالى فإن $(BMS - WMS)$ تعطى تقديراً للكمية $n\sigma^2\rho_I$ ومن هذا تتضح العلاقة:

$$r_I = \frac{BMS - WMS}{BMS - (n-1)(WMS)} \qquad (٢٦-٩)$$

وبالتعويض عن مكونات كل من متوسطات المربعات تعطى العلاقة:

$$r_I = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_a^2 + \hat{\sigma}_e^2} \qquad (٢٧-٩)$$

وتوقعاً فإن الحد الأدنى لمعامل الارتباط الداخلى هو صفر ولكن فى الواقع قد تنشأ القيم السالبة لمعامل الارتباط الداخلى من وجود قيمة أكبر لمتوسط المربعات داخل الفئات أكبر من متوسط المربعات بين الفئات وقد يحدث ذلك فى بعض الحالات، مثلاً فى الأرانب سينشأ تنافس بين أفراد العش الواحد أو البطن الواحدة وسوف تدفع باستمرار الأفراد الكبيرة الأفراد الصغيرة عن حصولها على الغذاء وبذلك ينتج أن التباين فى الأوزان داخل المجموعات أكبر من التباين بين المجموعات مما قد يعطى قيماً سالبة لقيمة r_I.

٩-١٣ العينات داخل العينات Sample within samples

يمكن تصور وضع تقسم فيه كل عينة إلى تحت عينات sub-samples ويمكن أيضاً لهذه الأخيرة أن تقسم إلى تحت تحت عينات sub-subsamples ... وهكذا. وبذلك تنشأ تقسيمات داخل تقسيمات مما يعطى توزيعاً متفرعاً أو عنقودياً hierarchal or nested. أو قد يكون العامل الرئيسى أو التقسيم الأكبر هو عامل ذو تأثير ثابت fixed كما سبق. فى حين أن العوامل التى فى التقسيمات التالية له تكون مختارة عشوائياً random، وبذلك ينشأ ما يعرف بالنموذج الخليط أى أن بعض التأثيرات ثابتة والأخرى عشوائية. أو قد تكون أعداد المكررات (المشاهدات) داخل العينات أو تحت العينات غير متساوية مما يؤدى إلى الحصول على بيانات غير متزنة.

٩-١٣-١ التقسيم العنقودى: حالة تمام العشوائية للعوامل

Nested or hierarchal classifications

كما سبق إذا قسمت كل عينة إلى تحت عينات sub-samples وإلى تحت تحت عينات sub-subsamples أو تقسيمات داخل تقسيمات فإن ذلك سوف يعطى توزيعاً

متفرعاً أو عنقودياً hierarchal or nested. فلو افترض مثلاً أنه توجد مجموعة من الحظائر يوضع فى كل منها ديك واحد ومعه مجموعة من الدجاج (الأمهات) وتعطى كل أم عدداً من البيض الذى يفقس ليعطى الكتاكيت (النسل) فإن أى كتكوت يمكن أن يوضع فى عينة تمثل نسل إحدى الأمهات وينشأ أيضاً عن ذلك مجاميع من النسل كل منها تنتمى لديك معين. أو قد تؤخذ مجموعة من العينات من أكثر من حقل أو موقع ثم يجرى على كل عينة من العينات المأخوذة من الحقل أكثر من تقدير واحد للمكونات المطلوب الكشف عنها مثل نسبة الملوحة أو الكالسيوم... الخ. وعلى هذا الأساس فقد يكون عدد العينات من كل حقل متساوياً كما أن عدد التقديرات لكل عينة متساوياً. فى حين فى المثال الخاص بالدجاج قد يختلف عدد الأمهات لكل ديك ويختلف أيضاً عدد النسل الناتج من كل أم.

فى مثل هذه التجارب العنقودية فإن بعض العوامل تكون محتواة داخل عوامل أخرى كتحت فئات sub classes وفى هذه الحالة لا يمكن بالتالى إجراء مقارنات لتحت الفئات على كل مستويات الفئات classes.

مثال ٩-٩

لوحظ فى حديقة ما فى منطقة صناعية ظهور بقع مرضية على أوراق الأشجار وهنا قد تختلف الأشجار عن بعضها كما أن الأوراق قد تختلف أيضاً عن بعضها، ولذا فإنه يجرى عد لعدد البقع على كل من نصفى أوراق مأخوذة عشوائياً من أشجار مختارة أيضاً عشوائياً من الحديقة لتقدير حجم الإصابة مثلاً. فإذا افترض أخذ 3 شجرات ثم أخذ 4 أوراق من كل شجرة وتم عد لعدد البقع على كل من نصفى الورقة، فيكون شكل التوزيع كالتالى:

	الشجرة الأولى				الشجرة الثانية				الشجرة الثالثة			
الأوراق	١	٢	٣	٤	١	٢	٣	٤	١	٢	٣	٤
أنصاف الأوراق	١ ٢	١ ٢	١ ٢	١ ٢	١ ٢	١ ٢	١ ٢	١ ٢	١ ٢	١ ٢	١ ٢	١ ٢

وبذا يصبح حجم التجربة الكلى فى هذه الحالة عبارة عن عدد أنصاف الأوراق التى تم عليها العد (المقياس أو المشاهدة) وهى 24 = (2)(4)(3) نصف ورقة. ويوضح الجدول التالى القيم المشاهدة لعدد البقع فى هذا المثال:

		الشجرة الأولى				الشجرة الثانية				الشجرة الثالثة			
الورقة		١	٢	٣	٤	١	٢	٣	٤	١	٢	٣	٤
نصف الورقة	١	11	5	6	4	15	9	11	12	17	10	10	8
	٢	9	5	4	2	17	8	10	11	16	11	9	10
المجموع للورقة		**20**	**10**	**10**	**6**	**32**	**17**	**21**	**23**	**33**	**21**	**19**	**18**
المجموع للشجرة		**46**				**93**				**91**			
المجموع الكلى		**230**											

والنموذج الرياضى للتقسيم العنقودى (عاملين أو تقسيمين):

$$Y_{ijk} = \mu + \tau_i + \ell_{ij} + \epsilon_{ijk} \qquad (٢٨-٩)$$

حيث μ تمثل المتوسط العام و τ_i تأثير اختلاف الأشجار عن بعضها وقد يرجع ذلك إلى اختلافات فى المكان فى الحقل أو اختلافات مصدرها وراثى مثلاً، ℓ_{ij} هى اختلاف الأوراق عن بعضها والذى يكون نتيجة لمواقع غير محددة أو لأسباب عشوائية أخرى و ε_{ijk} وهو الخطأ الطبيعى والذى لا يؤول إلى أى من العوامل فى النموذج وهو الذى يستخدم فى اختبار فرض العدم. وفى المثال السابق i تمثل الأشجار حيث $i=1, 2, 3$ و j تمثل الأوراق حيث $j = 1, 2, 3, 4$ وأنصاف الأوراق يمثلها k حيث $k = 1, 2$. أما افتراضات النموذج فهى كتلك الخاصة بالنموذج العشوائى كنتيجة لوضع التجربة وبذلك يكون $\tau \sim N(0,\sigma_\tau^2)$، $\ell \sim N(0,\sigma_\ell^2)$، $\in \sim N(0,\sigma_e^2)$ أى أن النموذج تام العشوائية فى تأثير العوامل الداخلة فيه.

ولإجراء التحليل تستخدم المجاميع الموجودة أسفل جدول البيانات فى حساب مجاميع المربعات حسب النموذج الرياضى المفترض ويلاحظ هنا أنه ليست هناك مجاميع للصفوف حيث إن أنصاف الأوراق لا تمثل عاملاً من عوامل الاختلافات ولكن هى قياسات عشوائية على أقل وحدات عينية، أى أن الوحدة التجريبية experimental unit هى فى هذه الحالة الورقة الكاملة، أما نصف الورقة فتعتبر وحدة تعيين sampling unit يجرى عد البقع على كل من نصفى الورقة لاختبار مدى كفاءة عملية العد ودقة القياس. ويجرى التحليل كالتالى:

معامل التصحيح:

$$CF = \frac{(Y_{...})^2}{(i)(j)(k)} = \frac{(230)^2}{(3)(4)(2)} = 2204.17$$

مجموع المربعات الكلى:

$$\sum\sum\sum Y_{ijk}^2 - CF = 16^2 + 14^2 + \cdots + 10^2 - 2204.17 = 375.83$$

مجموع المربعات بين الأشجار:

$$\sum_i \frac{Y_{i..}}{(j)(k)} - CF = \frac{(46)^2 + (93)^2 + (91)^2}{(4)(2)} - 2204.17 = 176.58$$

مجموع المربعات بين الأوراق:

$$\frac{\sum\sum Y_{ij.}^2}{(k)} - CF = \frac{(20)^2 + (10)^2 + \ldots + (18)^2}{2} - 2204.17 = 362.83$$

ومجموع المربعات الأخير هذا أى 362.83 يشمل كلا من الاختلافات بين الأشجار والتى كان مجموع مربعاتها 176.58 بالإضافة إلى الاختلافات بين الأوراق داخل الشجرة الواحدة، ويمكن حساب مجموع المربعات بين الأوراق داخل الأشجار (وتكتب بين الأوراق/الأشجار) كالتالى:

$$\frac{\sum\sum Y_{ij.}^2}{(k)} - \frac{\sum Y_{i..}^2}{(j)(k)} = 362.83 - 176.58 = 186.25$$

ويمكن حساب مجموع المربعات للعينات، أى بين العينات داخل الأوراق داخل الأشجار (وتكتب بين العينات/الأوراق/الأشجار) كالتالى:

$$\sum\sum\sum Y_{ijk}^2 - \frac{\sum\sum Y_{ij.}^2}{(j)(k)} = 375.83 - 362.83 = 13$$

ويمكن تلخيص النتائج المتحصل عليها فى جدول تحليل التباين التالى مع العلم بأن التباين فى الفئة الأكبر تحتوى إضافة إلى مصدر التباين الخاص بها التباينات الناشئة من الفئات الأقل منها وبالتالى فإنه يمكن أيضاً كتابة متوسط المربعات المقدر أمام كل متوسط مربعات محسوب.

ANOVA table

SOV	df	SS	MS	EMS
بين الأشجار	2	176.58	88.290	$\sigma_e^2 + 2\sigma_\ell^2 + 8\sigma_\tau^2$
بين الأوراق/الأشجار	9	186.25	20.694	$\sigma_e^2 + 2\sigma_\ell^2$
بين العينات/الأوراق/الأشجار	12	13.00	1.083	σ_e^2
الكلى المصحح	**23**	**375.83**		

ومن الجدول يمكن حساب مكونات التباين لكل من التأثيرات الداخلة فى النموذج وتبعاً لمتوسط المربعات المقدر.

فتقدير مكون التباين للأوراق عبارة عن:

$$\sigma_\ell^2 = \frac{\sigma_e^2 + 2\sigma_\ell^2 - \sigma_e^2}{2} = \frac{20.694 - 1.083}{2} = 9.81$$

فى حين أن تقدير مكون التباين بين الأشجار عبارة عن:

$$\sigma_\tau^2 = \frac{\sigma_e^2 + 2\sigma_\ell^2 + 8\sigma_\tau^2 - \sigma_e^2 + 2\sigma_\ell^2}{8} = \frac{88.29 - 20.694}{8} = 8.45$$

ولاختبار الفرضيات الخاصة بمكونات التباين فإنه يستخدم اختبارF بحيث يكون متوسط المربعات المستخدم فى المقام يحتوى على كل ما يتوقع وجوده فى متوسط المربعات للبسط ما عدا الجزء المرغوب فى اختباره، أى أنه لاختبار معنوية σ_ℓ^2 فإن $F = \frac{20.694}{1.083} = 19.1$ أى $\frac{\sigma_e^2 + 2\sigma_\ell^2}{\sigma_e^2}$، وعلى ذلك فتحت فرض العدم الخاص بأن $\sigma_\ell^2 = 0$ تصبح القيمة النظرية لاختبار $F = 1$ ، وأيضاً بالنسبة لاختبار الفرض الخاص بأن $\sigma_\tau^2 = 0$ فإن $F = \frac{88.29}{20.694} = 4.266$. وبالتالى فإنه بالكشف عن قيمة F الجدولية $F_{(0.01,9,2)} = 4.39$ فى الحالة الأولى، يتضح أن قيمتها معنوية جداً أى أن الاختلافات بين الأوراق داخل الأشجار كانت معنوية ويرفض الفرض الخاص بأن $\sigma_\ell^2 = 0$.

وأيضاً تكون الاختلافات بين الأوراق معنوية عند مستوى معنوية 5% وبالتالى فإن الخطأ القياسى لمتوسط المعاملة (الأشجار) $S_{\overline{Y}} = \sqrt{\frac{20.694}{8}} = 1.608$ والخطأ القياسى للفرق بين متوسطى معاملتين هو $\sqrt{\frac{2(20.694)}{8}} = 2.274$.

٩-١٣-٢ حالة النموذج الخليط Mixed model

يكون العامل الرئيسى أو التقسيم الأكبر فى كثير من الحالات هو عاملا ذا تأثير ثابت fixed كما سبق. فى حين أن العوامل التى فى التقسيمات التالية له تكون مختارة عشوائياً random، وبذلك ينشأ ما يعرف بالنموذج الخليط أى أن بعض التأثيرات ثابتة والأخرى عشوائية. فلو افترض فى الدواجن مثلا وجود عدد من الحظائر، وبكل حظيرة مجموعة من الأمهات المختارة عشوائياً وأن القياسات تتم على النسل الناتج، فإن تأثير الحظائر فى هذه الحالة يكون ثابتاً حيث إن هذه هى كل المستويات المستخدمة فى التجربة فى حين أن الأمهات تأثيرها عشوائى نظراً لاختيار التزاوجات بهذه الطريقة وأيضاً النسل الناتج منها تكون الاختلافات فيما بينها عشوائيةً. ويكون شكل النموذج الرياضى فى هذه الحالة كالتالى:

$$Y_{ijk} = \mu + \alpha_i + d_{ij} + \in_{ijk} \qquad (٩-٢٩)$$

حيث تمثل α_i التأثير الثابت للحظائر وينطبق عليها الاشتراط $\sum \alpha_i = 0$. فى حين أن σ_d^2 تمثل التأثير العشوائى للأمهات داخل الحظائر وهى تتبع التوزيع الطبيعى بمتوسط مقداره صفر وتباين σ_d^2. بينما تمثل $\in_{ijk}$ التأثير العشوائى للاختلافات بين النسل الناتج وهو الخطأ الحقيقى والذى يتوزع طبيعيا بمتوسط صفر وتباين σ_e^2.

خطوات حساب مجاميع المربعات وخلافه هى نفسها الخاصة بالنموذج السابق إلا أنه فى العمود الخاص بمتوسط المربعات المتوقع للعامل الثابت α_i تستبدل القيمة σ_a^2 بالقيمة $\frac{\sum \alpha_i^2}{(a-1)}$ حيث $(a-1)$ تمثل عدد درجات الحرية للتقسيم الثابت.

وفى النموذج الخليط فإن متوسط أى فئة من الفئات الخاصة بالتأثير الثابت فى التقسيم الأعلى عبارة عن:

$$\overline{Y}_{i..} = \mu + \overline{\alpha}_i + \overline{d}_{i.} + \overline{\in}_{i..} \qquad (٩-٣٠)$$

حيث تمثل $\overline{d}_{i.}$ و $\overline{\in}_{i..}$ متوسط التأثيرات العشوائية الداخلة فى الفئة i. وعلى هذا الأساس فإن التباين الخاص بأى متوسط $\overline{Y}_{i..}$ يكون كالتالى:

$$V(\overline{Y}_{i..}) = \frac{\sigma_d^2}{d} + \frac{\sigma^2}{nd} = \frac{1}{nd}(\sigma^2 + n\sigma_d^2) \qquad (٩-٣١)$$

وبذا فإن متوسط المربعات للأمهات داخل (الحظائر) هو التقدير غير المتحيز للكمية $\sigma^2 + n\sigma_d^2$. وبالتالى فإن الخطأ القياسى عبارة عن الجذر التربيعى للتباين المعطى فى (٩-٣١).

٩-١٣-٣ حالة عدم تساوى حجم العينة Unequal numbers

فى تجارب وبيانات الإنتاج الحيوانى وغيره قلما أن تتوافر أعداد الحيوانات أو المشاهدات المتساوية سواء فى العينات أو العينات داخل العينات. فهناك اختلافات فى عدد التزاوجات لكل طلوقة واختلافات فى عدد الأبناء الناتجة من كل أم (فى الأرانب أو الدواجن على سبيل المثال) وبذلك يمكن القول بأن بيانات الإنتاج الحيوانى غالباً ما تكون غير متزنة.

ولقد استنبطت الوسائل الحسابية لتحليل التباين فى حالة عدم تساوى حجم العينات وتحت العينات كما وأن الطرق التى يمكن بها حساب مكونات التباين أيضاً فى تلك الحالات قد استخرجت أيضاً، ولو أنها إلى حد ما أكثر صعوبة عنها فى حالات تساوى حجم العينات وتحت العينات.

أما بالنسبة لطريقة الحساب فهى لا تختلف كثيراً عن الطريقة التى اتبعت فى تحليل التباين فى فصل ٩-١١.

مثال ٩-١٠

فى تجربة على نمو الكتاكيت بالجرام (مطروحاً منه 100 جرام) والناتجة من تزاوج ديكين مختارين عشوائياً مع مجموعة من الإناث فى فترة الأسبوعين الأولين تم الحصول على النتائج التالية:

	الديك أ			الديك ب				
الأمهات	١	٢	٣	٤	٥	٦	٧	٨
النمو فى النسل	7	10	8	6	14	8	12	16
		8	4	8	9	6	18	12
		5		8		4	20	14
						8		11
								12
								15
المجموع للأم	**7**	**23**	**12**	**22**	**23**	**26**	**50**	**80**
المجموع للديك		**42**				**201**		
العدد		**6**				**18**		

النموذج الرياضى:

$$Y_{ijk} = \mu + s_i + d_{ij} + \in_{ijk} \qquad (٩-٣٢)$$

وقد افترض أن جميع العوامل فى النموذج عشوائية فيما عدا المتوسط وكذلك $s \sim N(0,\sigma_s^2)$ و $d \sim N(0,\sigma_d^2)$ و $\in \sim N(0,\sigma_e^2)$.

العدد الكلى للنسل: $N = 24$

المجوع الكلى: $\sum_i \sum_j \sum_k Y_{ijk} = 243\,g$

مجموع المربعات الكلى: $\sum_i \sum_j \sum_k Y_{ijk}^2 = 2893$

مجموع المربعات المصحح: $\sum_i \sum_j \sum_k Y_{ijk}^2 - CF = 2893 - 2460.375 = 432.625$

مجموع المربعات بين الآباء: $\sum_i \frac{Y_{i..}^2}{n_i} - CF = 2538.5 - 2460.375 = 78.125$

مجموع المربعات بين الأمهات داخل الآباء:

$$\sum_i \sum_j \frac{Y_{ij}^2}{n_{ij}} - \sum_i \frac{Y_{i..}^2}{n_i} = 2792.167 - 2538.5 = 253.667$$

أو $$3371.792 - 78.125 = 253.667$$

مجموع المربعات بين النسل داخل الأمهات داخل الآباء:

$$\sum_i \sum_j \sum_k Y_{ijk}^2 - \sum_i \sum_j \frac{Y_{ij.}^2}{n_{ij}} = 2893 - 2792.167 = 100.833$$

أو $$2893 - 2460.375 - (78.125 + 253.667) = 100.833$$

وبناء عليه يمكن وضع النتائج المتحصل عليها من التحليل فى جدول تحليل التباين التالى:

SOV	df	SS	MS	EMS
Sires	1	78.13	78.12	$\sigma_e^2 + k_2\sigma_d^2 + k_3\sigma_s^2$
Dams/sires	6	253.67	42.28	$\sigma_e^2 + k_1\sigma_d^2$
Offspring /Dams/ Sires	16	100.73	6.30	σ_e^2
C-Total	**23**	**432.62**		

ويلاحظ من جدول تحليل التباين وحساب مجاميع المربعات السابقة أن مجموع المربعات بين الأمهات داخل الطلائق أى الأمهات التى تتزاوج مع أحد الطلائق يساوى مجموع المربعات بين الأمهات بدون النظر إلى مجاميع الطلائق مطروحاً منه مجموع المربعات بين الطلائق. وأيضاً درجات الحرية فهناك 8 أمهات عدد درجات الحرية فيما بينها 7 وبالتالى درجة الحرية للأمهات داخل الطلائق $7 - 1 = 6$ حيث إن هناك درجة حرية واحدة بين الطلائق. ولكن بناء على أن عدد الأفراد (المشاهدات) فى كل مستوى غير متساوى، أى أن عدد النسل لكل أم غير متساوى، وأيضاً عدد الأمهات التى تزاوجت مع كل أب وبالتالى فإن حجم عائلة كل طلوقة متغير. ويجب حساب قيم k_1, k_2, k_3 التى تظهر فى العمود الأخير من جدول تحليل التباين حتى يتسنى تقدير مكونات التباين. ولقد تم عرض الحالة التى يختلف فيها عدد الأفراد فى التقسيم الأحادى فى الفصل ٩-١١ حيث تم حساب الكمية $n_\circ$ على أساس المتوسط

التوافقى لعدد الأفراد فى الفئة. ولقد أعطى (1969) Gaylor and Hartwell القيم المتوقعة لمتوسطات المربعات فى حالة التحليل العنقودى وكيفية تقدير مكونات التباين. وعموماً فإن القيم المطلوبة لمتوسط الأعداد هى:

$$k_1 = \frac{1}{df(dams)}\left(N... - \sum_i \frac{\sum_j n_{ij}^2}{n_{i.}} \right) \qquad (٣٣-٩)$$

$$k_2 = \frac{1}{df(sires)}\left(\sum_i \frac{\sum_j n_{ij}^2}{n_{i.}} - \frac{\sum_i \sum_j n_{ij}^2}{N..} \right) \qquad (٣٤-٩)$$

$$k_3 = \frac{1}{df(sires)}\left(N... - \frac{\sum_i n_{i.}^2}{N...} \right) \qquad (٣٥-٩)$$

وبتطبيق هذه المعادلات على المثال ٩-١٠ فإن حسابها كالتالى:

$$k_1 = \left(\frac{1}{6}\right)\left[24 - \left(\frac{1^2+3^2+2^2}{6} + \frac{3^2+2^2+4^2+3^2+6^2}{18}\right)\right] = 2.9259$$

$$k_2 = \left(\frac{1}{1}\right)\left[\begin{array}{l}\left(\dfrac{1^2+3^2+2^2}{6} + \dfrac{3^2+2^2+4^2+3^2+6^2}{18}\right) \\ -\left(\dfrac{1^2+3^2+...+6^2}{24}\right)\end{array}\right] = 2.7778$$

$$k_3 = \left(\frac{1}{1}\right)\left[24 - \frac{6^2+18^2}{24}\right] = 9$$

وبالتالى فيمكن تقدير مكونات التباين حسب جدول تحليل التباين والمعادلات التالية:

$$\sigma_d^2 = \frac{MSD - MSW}{k_1} \qquad (٣٦-٩)$$

وبالتعويض

$$\sigma_d^2 = \frac{42.278 - 6.6302}{2.9859} = 12.2957$$

أما تقدير قيمة σ_S^2 فيمكن حسابه عن طريق التعويض فى المعادلة التالية:

$$\sigma_s^2 = [MS_s - MS_w - k_1/k_2(MS_d - MS_w)]/k_3 \qquad (٣٧-٩)$$

وبالتعويض

$$\sigma_s^2 = \frac{1}{9}\left[78.125 - 6.302\left(\frac{2.778}{2.9259}\right)(42.278 - 6.302)\right] = \frac{37.6682}{9} = 4.1851$$

هذه التقديرات لمكونات التباين تعتبر غير متحيزة. أما فى حالة النموذج الذى يكون فيه العامل الأساسى (الطلائق فى هذه الحالة) ذات تأثير ثابت fixed فإن اختبار F لفرض العدم الخاص بالطلائق يكون $78.125/42.278 = 18.5$ لا يمكن أن يستخدم نظراً لاختلاف القيم المحسوبة لكل من k_1 وk_2 ولكن يشكل بسطا ومقاما لهذا الاختبار.

هذا ويمكن أن يمتد مثل هذا النوع من التحليلات إلى مستويات أو عوامل تدخل ضمن التحليل العنقودى كأن تكون مثلاً الطلائق ممثلة داخل مزارع مختلفة ... وهكذا.

مثال ٩-١١

يمكن استخدام برنامج SAS لحل مثال ٩-١٠ كالتالى:

```
DATA NESTED;
INPUT SIRE $ DAM PROG @@;
CARDS;
A 1 7 A 2 10 A 2 8 A 2 5 A 3 8 A 3 4
B 4 6 B 4 8 B 4 8 B 5 14 B 5 9 B 6 8
B 6 6 B 6 4 B 6 8 B 7 12 B 7 18 B 7 20
B 8 16 B 8 12 B 8 14 B 8 11 B 8 12 B 8 15
```

```
PROC VARCOMP METHOD = TYPE1;
CLASS SIRE;
MODEL PROG  = SIRE DAM(SIRE);
RUN;
```

لاحظ الطريقة التى كتب بها الـ model حيث توضح أن الأمهات موزعة داخل الطلائق (dam(sire وهذا هو سبب أن أرقام الأمهات للطلوقة الأول تختلف عن أرقام الأمهات للطلوقة الثانى.

نتائج التحليل:

Variance Components Estimation Procedure
Class Level Information

Class	Levels	Values
SIRE	2	A B
DAM	8	1 2 3 4 5 6 7 8

Number of observations in data set = 24

Dependent Variable: PROG

Source	DF	Type I SS	Type I MS
SIRE	1	78.12500000	78.12500000
DAM(SIRE)	6	253.66666667	42.27777778
Error	16	100.83333333	6.30208333
Corrected Total	23	432.62500000	

Source	Expected Mean Square
SIRE	Var(Error) + 2.7778 Var(DAM(SIRE)) + 9 Var(SIRE)
DAM(SIRE)	Var(Error) + 2.9259 Var(DAM(SIRE))
Error	Var(Error)

Variance Component	Estimate
Var(SIRE)	4.18541960
Var(DAM(SIRE))	12.29549051
Var(Error)	6.30208333

صندوق ٩-٣

افتراض عشوائية أو ثبات المؤثرات لا يؤثر على طريقة تقسيم درجات الحرية أو حساب مجموع المربعات أو المتوسطات ولكن يؤثر فقط علي محتويات متوسط المربعات المتوقع (EMS) وبالتالى على اختبار الفروض فى تحليل التباين.

تمارين الباب التاسع

٩-١ ثلاثة أصناف من البطاطس تمت زراعتها فى حقل متماثل مقسم إلى 12 قطعة وكانت النتائج المتحصل عليها لوزن المحصول الناتج كالتالى:

أ- 18, 20, 14, 11

ب- 18, 28, 17, 22

ج- 9, 18, 6, 7

المطلوب: اختبار الفرض الخاص بتساوى المحصول فى الأصناف الثلاثة. وما هى الافتراضات والاشتراطات المطلوبة لإجراء التحليل؟ اذكر النموذج الرياضى المستخدم.

٩-٢ العينات التالية تمثل المحصول الناتج من استخدام خمسة أنواع من الأسمدة. والمطلوب: اختبار فرض تساوى المتوسطات فى المحصول واختبار أيها تختلف عن الأخرى بواسطة استخدام اختبار LSD.

أ	ب	ج	د	هـ
21	35	45	32	45
35	12	60	53	29
32	27	33	29	31
28	41	36	42	22
14	19	31	40	36
47	23	40	23	29
25	31	43	35	42
48	20	48	42	30

٩-٣ فى تجربة ما فى الدواجن استخدم فيها 9 ديوك حيث تزاوج كل من الديوك مع 4 أمهات وقيس الإنتاج لعدد 5 أبناء من كل تزاوج مأخوذة عشوائياً. وكان متوسط الإنتاج للبيض فى التجربة 62.75 بيضة/فرد. وكان مجموع المربعات الكلى غير المصحح $\sum Y^2 = 773905.75$. استكمل بناء على ذلك جدول تحليل التباين التالى وقدر مكون التباين لكل من الديوك والأمهات والخطأ.

SOV	df	SS	MS
Sires		4732.0	
Dams/sires			397.5
Progeny/Dams/ Sires			

٩-٤ أعطيت النموذج الرياضى $Y_{ij} = \mu + \tau_i + \in_{ij}$ حيث $i = 1,\cdots,4$ و $j = 1,2$ وكان متوسط المعاملات هو 8 ,5 ,3 ,2 للمعاملات من 1 إلى 4 على التوالى، وكان مجموع المربعات الكلى غير المصحح = 212

المطلوب:

١- إجراء تحليل التباين واختبار فرض العدم بتساوى المعاملات.

٢- إجراء المقارنات وتقسيم التباين الخاص بالمقارنة.

(أ) المعاملة 1 ضد بقية المعاملات

(ب)المعاملة 3 ضد المعاملة4

(ت)المعاملة 2 ضد المعاملتين 3، 4.

٩-٥ من جدول تحليل التباين التالى المطلوب تقدير قيمة الارتباط الداخلى إذا علمت أن قيمة $n_\circ = 5.197$

SOV	df	SS
Sires	8	1273.3
Within sires		
C-Total	**46**	**5872.1**

٩-٦ فيما يلى تقدير محتوى الفوسفور فى أربعة أصناف من أشجار الفاكهة

الصنف				
١	0.40	0.35	0.47	0.50
٢	0.70	0.65	0.79	0.84
٣	0.80	0.60	0.66	0.73
٤	0.75	0.57	0.70	0.80

أ- اختبر فرض العدم بعدم وجود فروق معنوية بين الأصناف الأربعة عند مستوى معنوية 5% .

ب- اختبر فرض العدم بأن متوسط الصنف الأول لا يختلف معنوياً عن متوسط الأصناف الأخرى.

٩-٧ جربت معاملتان غذائيتان لمعرفة تأثيرهما على وزن العرف فى الديوك بالجرام. وكانت البيانات كما يلى (علماً بأن الديوك كانت متماثلة إلى حد كبير فى أوزانها وأعمارها وحالتها الصحية):

معاملة (أ)	129	115	72	125	127	68	130	125
معاملة (ب)	70	71	79	55	75	74	90	

والمطلوب اختبار فرض العدم بتساوى متوسطى المعاملتين باستخدام

أ- اختبار t

ب- اختبار F ، تحقق من أن $F = t^2$

٩-٨ جرب نوعان من المبيدات على كمية المحصول الناتج من 30 فدان فى كل وكان متوسط إنتاج الفدان والانحراف المعيارى كما يلى (علماً بأنه قد تركت 30 فدان أخرى كمجموعة ضابطة control):

	العدد	المتوسط	الانحراف المعيارى للعينة
مبيد (١)	30	91	7.5
مبيد (٢)	30	88	8.2
مجموعة ضابطة	30	75	8.8

والمطلوب:

١- كتابة النموذج الرياضى مع النص على الافتراضات

٢- اختبر فروض العدم التالية عند مستوى 5% :

$H_\circ : \mu_1 = \mu_2 = \mu_3$

$H_\circ : (\mu_1 + \mu_2)/2 = \mu_3$

$H_\circ : \mu_1 = \mu_2$

١٠ العلاقة بين متغيرين: الانحدار والارتباط البسيطان

Simple regression and correlation

١ - مقدمة

٢ - معادلة خط الانحدار

٣ - استخدام المصفوفات فى الانحدار

٤ - مصادر الاختلاف فى الانحدار الخطي

٥ - القيم المعدلة لأثر انحدار Y على X

٦ - الانحرافات المعيارية وحدود الثقة للتقديرات المختلفة

٧ - اختبار معنوية معامل الانحدار

٨ - المقارنة بين معاملى الانحدار

٩ - التوزيع ذو المتغيرين

١٠ - العلاقة بين معامل التحديد وخطأ التقدير

١١ - تقييم ملائمة نموذج التحليل

١٢ - الارتباط البسيط

١٣ - العلاقة بين معامل الانحدار ومعامل الارتباط

١٤ - اختبار معنوية معامل الارتباط وتقدير حدود الثقة له

١٥ - اختبار تساوى معاملى ارتباط

١٦ - اختبار تجانس عدد من معاملات الارتباط

١٧ - السبب والأثر فى تحليل الارتباط والانحدار

١٨ - تباين الدالة الخطية

١٩ - ارتباط الرتب

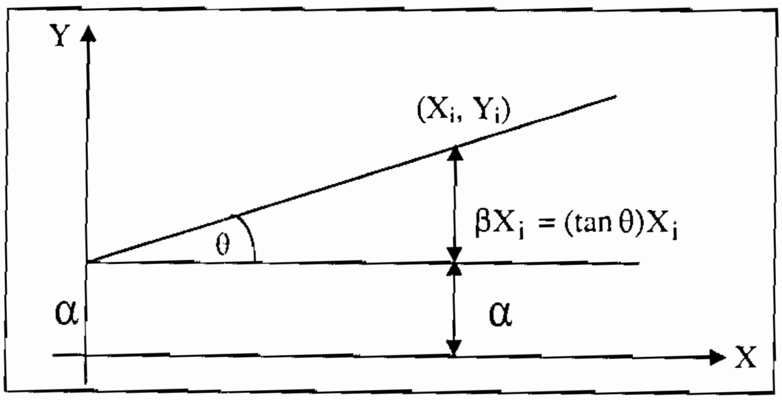
١٠-١ مقدمة

يطلق على الانحدار regression الارتداد أو الاعتماد ويستخدم لدراسة العلاقة بين متغيرين أحدهما Y والذى يعتمد فى قيمته على متغير آخر X. ويطلق على Y المتغير التابع أو المعتمد dependent variable بينما يطلق على X المتغير المستقل independent variable وقد يسمى أيضاً بالمتغير التفسيرى explanatory variable أى الذى يفسر التغيرات فى Y. ويوجد أمثلة كثيرة لذلك منها دراسة العلاقة بين الأسمدة X وكمية المحصول Y، دراسة العلاقة بين العمر عند أول ولادة X ومحصول اللبن Y، دراسة العلاقة بين الطول X والوزن Y، دراسة العلاقة بين محيط الصدر X والوزن Y. وعادة ما يكون المتغير المستقل هو المتغير الأسهل فى القياس عن المتغير التابع. ويعبر عن العلاقة بين المتغيرين رياضياً $Y = f(X)$، أى أن Y دالة X ويعبر عنها إحصائياً بالانحدار.

وقد تعبر المعادلة $Y = 3 + 5X$ على سبيل المثال عن العلاقة بين المتغيرين، وهى تمثل معادلة خط مستقيم صورته العامة:

$$Y = \alpha + \beta X \qquad (١٠-١)$$

حيث α هى الجزء المقطوع من محور الصادات intercept، أى قيمة Y عندما تكون قيمة X مساوية للصفر، β تمثل ميل slope الخط المستقيم، أى ظل tan الزاوية θ التى يصنعها المستقيم فى الاتجاه الموجب لمحور السينات. كما يطلق على β معامل الانحدار أو الاعتماد egression coefficient. وشكل ١٠-١ يمثل تلك العلاقة بيانياً.

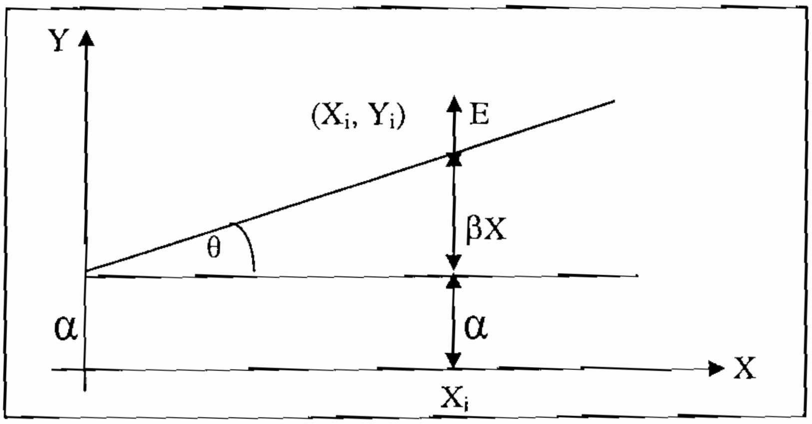


شكل ١٠-١ تمثيل العلاقة $Y = \alpha + \beta X$ بيانياً

ومن الشكل يلاحظ أن لكل قيمة من المتغير المستقل X قيمة تقابلها للمتغير التابع Y وتقع كل قيم Y على خط مستقيم واحد.

ولدراسة العلاقة بين متغيرين وليكن أحدهما وزن الحيوان بالكيلوجرام والآخر العمر بالأسبوع، فإن أول ما يجب عمله هو تمثيل تلك العلاقة بيانياً على أن يمثل المحور السينى المتغير المستقل، وهو العمر فى مثل هذه العلاقة، ويمثل المحور الصادى الوزن فى شكل يعرف بشكل الانتشار scatter diagram ومنه يتضح إذا كانت العلاقة خطية أو غير ذلك. وإذا كانت العلاقة خطية، أى أن النقط فى شكل الانتشار تتجمع حول خط مستقيم، حيث غالباً أن العلاقة لا تتحقق تماماً بمعنى أن بعض النقط تكون فوق الخط والبعض تحته، وقد تكون هناك عدة نقط على الخط. والمتغير التابع Y يمكن التعبير عنه بالعلاقة (١٠-٢) والشكل ١٠-٢ التاليين:

$$Y = \alpha + \beta X + \in \qquad (١٠-٢)$$



شكل ١٠-٢ تمثيل العلاقة $Y = \alpha + \beta X + \in$ بيانياً

وتمثل $\in$ الخطأ الخاص المصاحب للمتغير Y ويطلق عليه الخطأ العشوائى أو التجريبى وقد يرجع إلى:

١ - خطأ فى قياس المتغير Y errors of measurement

٢ - قد يكون هناك متغيرات أخرى غير X تؤثر على Y ولكن أهملت باعتبار أن X هو المتغير الأساسى محل الدراسة، ويعبر عنها بالمتغيرات المحذوفة omitted variables والتى تدخل ضمن مكونات الخطأ.

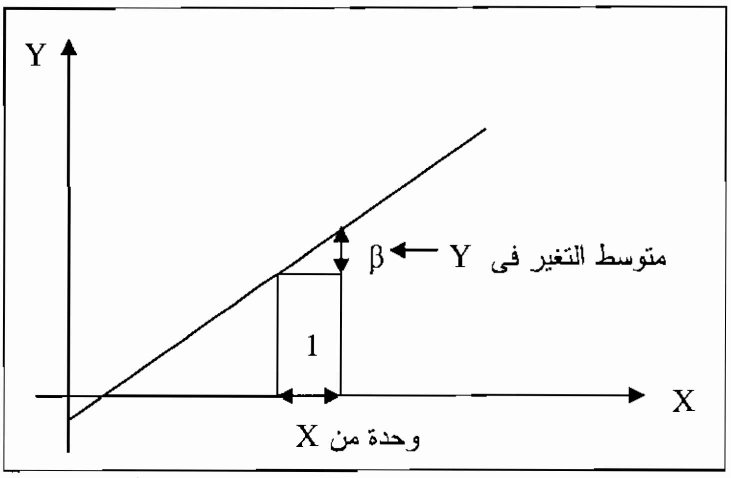
والعلاقة (١٠-٢) هنا تمثل علاقة حقيقية true relationship والهدف هو تقدير معالم هذه العلاقة وهما α و β. و هذه العلاقة مثل للبيانات البيولوجية biological data لوجود عنصر الخطأ والذى لا يمكن التحكم فيه.

١٠-٢ معادلة خط الانحدار

يعرف معامل انحدار المتغير التابع Y على المتغير المستقل X بأنه متوسط التغير فى المتغير التابع Y عندما يتغير المتغير المستقل X بمقدار الوحدة وذلك فى مدى معين للمتغير X، ويرمز له بالرمز β_{YX}.

فمثلاً عند دراسة العلاقة بين الوزن بالكيلوجرام والعمر باليوم فى عجول التسمين كان معامل اعتماد (انحدار) الوزن على العمر هو 0.8 كج/يوم، أى $\hat{\beta}_{YX} = 0.8$ كج/يوم. حيث $\hat{\beta}_{YX}$ هو تقدير غير متحيز لمعامل الانحدار فى العشيرة. وهذا يعنى أنه بزيادة العمر بمقدار يوم فإن الوزن يزيد بمقدار 0.8 كج.

أما إذا كان معامل الانحدار $\hat{\beta}_{YX} = -3$ وحدة من Y / وحدة من X، فإن هذا يعنى أنه بزيادة X بمقدار الوحدة فإن Y تنقص بمقدار 3 وحدات. ومعامل الانحدار لابد وأن يكون مميزاً وقيمته تتراوح بين $-\infty$ إلى $+\infty$ ويمثل شكل ١٠-٣ تعريف معامل الانحدار هندسياً.



شكل ١٠-٣ تمثيل معامل الانحدار هندسياً

وترجع أهمية دراسة الانحدار إلى:

١ - معرفة ما إذا كانت Y تعتمد على X وللحصول على مقياس لتلك العلاقة.

٢ - التنبؤ بقيم Y عند معرفة قيم X، وتفسير Y بواسطة X (وذلك فى مدى معين).

٣ - تحديد شكل منحنى الانحدار regression curve

٤ – معرفة الأخطاء الحقيقية الموجودة فى التجربة بعد التعديل (التصحيح) لأثر المتغير المستقل.

٥ – اختبار الفروض التى قد يضعها الباحثون حول العلاقة بين المسبب cause وتأثيره effect.

قد لا يفضل لسبب أو لآخر دراسة كل أفراد العشيرة وعلى ذلك يكتفى بدراسة العلاقة بين المتغيرين فى عينة أو جزء (عينة) من هذه العشيرة وذلك بغرض الوصول إلى تقدير لكل من α، β والتى يعبر عنها بنموذج التقدير estimated model التالى:

$$Y = a + bX + e \qquad (١٠-٣)$$

حيث b, a هما تقديران غير متحيزين لكل من α، β السابق الإشارة إليهما فى (١٠-٢)، e تقدير لعنصر الخطأ غير المعروف. ولكى تتم عملية التقدير يلزم مجموعة من المشاهدات للمتغير X وما يقابلها من المتغير Y.

وفى حالة الانحدار الخطى البسيط simple linear regression يمكن رسم الخط بنقطتين وفى هذه الحالة لا يمكن تقدير خطأ، أما إذا كان هناك مجموعة من النقط فيلزم توفيق خط يسمى خط الانحدار regression line يمثل هذه العلاقة ومعادلته هى:

$$\hat{Y} = a + bX \qquad (١٠-٤)$$

وتقدير كل من b ,a يكون بإحدى الوسيلتين التاليتين أو غيرهما:

١- حساب الفروق (الأخطاء) بغض النظر عن الإشارة مع محاولة جعل مجموعها أقل ما يمكن. أو

٢- حساب مجموع مربع الأخطاء $\sum e^2$ من النموذج (١٠-٤)، حيث $e_i = Y_i - a - bX_i$ وجعله أقل ما يمكن. وهذه الطريقة هى الأكثر شيوعاً واستخداماً وتعرف بطريقة المربعات الصغرى least squares method.

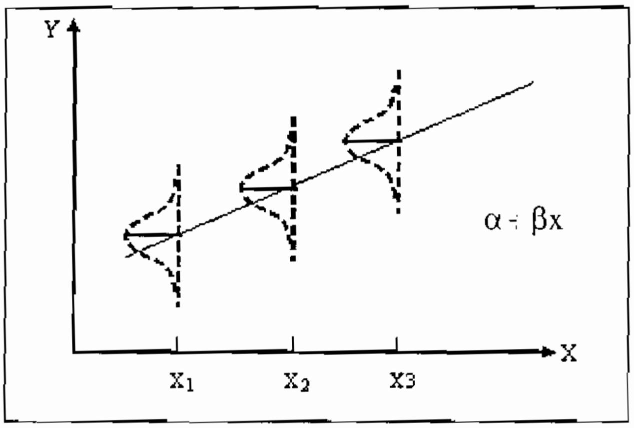
ولكى يتم ذلك توضع افتراضات لازمة لعملية التقدير وهى:

١- المتغير X ثابت fixed فى المعاينات المتكررة repeated sampling أى ليس له توزيع احتمالى.

٢- لا يوجد ارتباط بين كل من المتغير المستقل X ومكون الخطأ e.

٣- لكل قيمة للمتغير X_i يوجد توزيع للمتغير Y متوسطه يقع على خط الانحدار $\hat{\mu}_{y.x_i} = a + \hat{b}X_i$، وتختلف المتوسطات للتوزيعات ولكن لها نفس التباين $\sigma^2_{Y.X}$ وشكل ١٠-٤ يبين ذلك.

٤- الأخطاء e_i مستقلة وتتوزع طبيعياً بمتوسط يساوى الصفر وتباين $\sigma^2_{Y.X}$ وهذا الفرض مهم عند اختبار معنوية الانحدار.



شكل ١٠-٤ التوزيع الطبيعى لقيم Y حول خط الانحدار $\alpha + \beta x$ لبعض قيم X

ولتطبيق طريقة المربعات الصغرى، أى لجعل مجموع مربعات الأخطاء $\sum e^2$ أقل ما يمكن، يستخدم كل من التفاضل الجزئى وتفاضل دالة الدالة على النحو التالى:

بما أن: $Y_i = a + bX_i + e_i$

إذاً $e_i = Y_i - a - bX_i$

وبتربيع الطرفين: $e_i^2 = (Y - a - bX_i)^2$

وبالجمع لكل قيم e_i من 1 إلى n حيث n عدد أزواج المشاهدات:

$$\sum e_i^2 = \sum (Y - a - bX_i)^2$$

بالتفاضل الجزئى لكل من a ، b ومساواة الناتج بالصفر:

$$\frac{\partial \sum e^2}{\partial a} = 2\sum (Y_i - a - bX_i)(-1) = 0$$

$$\therefore\ \sum Y_i = na + b\sum X_i \qquad (١٠-٥)$$

$$\frac{\partial \sum e^2}{\partial b} = 2\sum (Y_i - a - bX_i)(-X_i) = 0$$

$$\therefore\ \sum X_i Y_i = a\sum X_i + b\sum X_i^2 \qquad (١٠-٦)$$

وتسمى المعادلتان (١٠-٥)، (١٠-٦) بالمعادلتين الاعتياديتين normal equations وبحل هاتين المعادلتين آنياً يمكن الحصول على:

$$a = \overline{Y} - b\overline{X} \qquad (١٠-٧)$$

$$b_{yx} = \frac{\sum XY - (\sum X)(\sum Y)/n}{\sum X^2 - (\sum X)^2/n} \qquad (١٠-٨)$$

ويمكن الوصول إلى صورة أخرى لمعامل الانحدار b إذا تم التعبير عن قيم المتغيرين كانحراف عن متوسطهما، أى استخدام y بدلاً من Y حيث $y = Y - \overline{Y}$ ، x بدلاً من X حيث $x = X - \overline{X}$. وبالتالى يمكن استنتاج أن:

$$b_{yx} = \frac{\sum xy}{\sum x^2} \qquad (١٠-٩)$$

وإذا قسم كل من بسط ومقام العلاقة (١٠-٩) على درجات الحرية n-1 فإن:

$$b_{yx} = \frac{\sum xy/(n-1)}{\sum x^2/(n-1)} = \frac{Cov(X,Y)}{V(X)} \qquad (١٠-١٠)$$

ويعرف $Cov(X,Y)$ بالتغاير covariance وهو عبارة عن الجزء من التباين المشترك joint variance بين قيم المتغيرين X، Y وقيمته قد تكون موجبة أو سالبة أو صفر.

ومعامل الانحدار b أو b_{yx} (معامل انحدار Y على X) يأخذ أى قيمة ويستمد إشارته من إشارة التغاير.

وتكون معادلة الخط المستقيم (معادلة الانحدار) عبارة عن:

$$\hat{Y} = a + bX \qquad (١٠-١١)$$

وبالتعويض عن a بقيمتها $a = \overline{Y} - b\overline{X}$

$$\hat{Y} = \overline{Y} - b\overline{X} + bX = \overline{Y} + b(X - \overline{X}) = \overline{Y} + bx \qquad (١٠-١٢)$$

أى أن:

$$\hat{y} = \hat{Y} - \overline{Y} = bx \qquad (١٠-١٣)$$

ومن خصائص خط الانحدار:

١ – مجموع الانحرافات عن خط الانحدار يساوى صفر.

٢ – مجموع مربع الانحرافات عنه أقل ما يمكن.

٣ – نقطة تقاطع المتوسطين $(\overline{X}, \overline{Y})$ تقع على هذا الخط.

مثال ١٠-١

فى دراسة للعلاقة بين الطول X بالسنتيمتر والوزن Y بالكيلوجرام فى عشيرة ما، كانت البيانات كالتالى:

الزوج	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠
الطول	150	175	170	160	150	155	152	180	154	165
الوزن	50	77	65	63	61	62	55	75	55	68

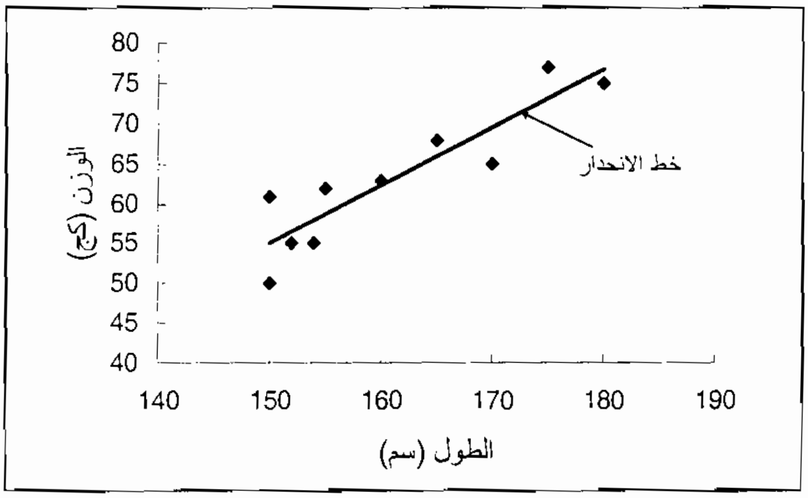
المطلوب:

١ – تمثيل تلك العلاقة بيانياً.

٢ – حساب معامل انحدار الوزن على الطول

٣ – كتابة معادلة خط الانحدار ورسم خط الانحدار رسماً دقيقاً.

التمثيل البيانى يمثله شكل ١٠-٥، ويلاحظ من دراسة شكل الانتشار أن العلاقة بين الطول والوزن علاقة أقرب ما تكون إلى الخطية (فى المدى المدروس). وعلى ذلك يمكن تقدير معادلة الخط المستقيم (معادلة الانحدار) $\hat{Y} = a + bX$ الذى يمثل هذه العلاقة بحيث يكون مجموع الانحرافات عن هذا الخط مساوية للصفر ومجموع مربع الانحرافات عن نفس الخط أقل ما يمكن، أى أن التقدير يتم بطريقة المربعات الصغرى.



شكل ١٠-٥ انحدار الوزن على الطول

وحيث إن الوزن قد يعتمد على الطول فإن الوزن Y هو المتغير التابع والطول X هو المتغير المستقل. ومن المعطيات:

$n = 10$

$\overline{X} = 161.1 \qquad \sum X = 1611 \qquad \sum X^2 = 260595$

$\overline{Y} = 63.1 \qquad \sum Y = 631 \qquad \sum XY = 102415$

وبتطبيق المعادلة (١٠-٨) فإن:

$$b_{yx} = \frac{102415 - \dfrac{(1611)(631)}{10}}{260595 - \dfrac{(1611)^2}{10}} = \frac{760.9}{1062.9} = 0.716 \ \text{kg/cm}$$

وحيث إن: $a = \overline{Y} - b\overline{X}$ فإن $a = 63.1 - (161.1)(0.716) = -52.228$ kg

ولرسم خط الانحدار (شكل ١٠-٥) تؤخذ نقطة تقاطع متوسطى المتغيرين والجزء المقطوع من المحور الصادى a ويتم التوصيل بين النقطتين للحصول على خط الانحدار الذى معادلته ستكون كما يلى:

$$\hat{Y} = -52.228 + 0.716X$$

وتسمى معادلة الخط المستقيم هذه معادلة الانحدار ويطلق عليها أيضاً معادلة التنبؤ prediction equation والتى منها يمكن حساب أى قيمة متوقعة لـ Y إذا علمت قيمة X، فمثلاً القيمة المتوقعة للوزن إذا كان الطول 185 سم هى:

$$\hat{Y} = -52.228 + (0.716)(185) = 80.232 \text{ kg}$$

والقيمة المتوقعة للوزن إذا كان الطول 172 سم هى:

$$\hat{Y} = -52.228 + (0.716)(172) = 70.924 \text{ kg}$$

ويمكن أن تقدر القيمة المتوقعة من (١٠-١٢) كما يلى:

$$\begin{aligned}\hat{Y} &= 63.1 + 0.716(x) = 63.1 + (0.716)(172 - 161.1) \\ &= 63.1 + (0.716)(10.9) = 70.904\end{aligned}$$

والفرق الناتج بين 70.924 ، 70.904 راجع لخطأ التقريب rounding error.

مثال ١٠-٢

يمكن حل مثال ١٠-١ باستخدام PROC REG فى برنامج SAS، وتستخدم هذه الطريقة عند الرغبة فى الحصول على تقدير كل من a (الجزء المقطوع من المحور الصادى intercept) و b (معامل الانحدار slope) بالإضافة إلى تحليل التباين والذى سوف يتم تناوله فى الأجزاء التالية.

```
DATA WEIGHT;
INPUT HEIGHT WEIGHT @@;
CARDS;
150 50 175 77 170 65 160 63 150 61
155 62 152 55 180 75 154 55 165 68
PROC REG;
MODEL WEIGHT = HEIGHT;
RUN;
```

لاحظ أنه يمكن إضافة بعض الخيارات إلى الـ model منها على سبيل المثال:

Model weight = height / XPX I;

وهذا يؤدى إلى الحصول على مصفوفة $X'X$ ومعكوسها، أى $(X'X)^{-1}$

نتيجة التحليل:

The REG Procedure
Model: MODEL1
Dependent Variable: weight

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	544.70676	544.70676	34.53	0.0004
Error	8	126.19324	15.77416		
Corrected Total	9	670.90000			

الجزء المقطوع من المحور الصادى (a)

Root MSE	3.97167	R-Square	0.8119
Dependent Mean	63.10000	Adj R-Sq	0.7884
Coeff Var	6.29425		

معامل الانحدار b

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-52.22693	19.66572	-2.66	0.0290
height	1	0.71587	0.12182	5.88	0.0004

لاحظ وجود بعض النتائج الأخرى والتى سوف يتم تناولها فى الأجزاء التالية.

١٠-٣ استخدام المصفوفات فى الانحدار

استخدام المصفوفات له كثير من الميزات من أهمها أن التعامل الرياضى فيها يكون أكثر اختصاراً وأوضح رؤية، أى أنه بمجرد كتابة المشكلة وحلها عن طريق المصفوفات فإنه يمكن تطبيق الحل على أى مشكلة بغض النظر عن عدد الحدود فى معادلة الانحدار. والخطوات التالية تبين كيفية حل مثال ١٠-١ باستخدام المصفوفات.

سبق عرض المعادلة (١٠-٣) وهى $Y = a + bX + e$، هذه المعادلة يمكن إعادة كتابتها باستخدام المصفوفات كالتالى:

$$Y = X\beta + \varepsilon \qquad (١٠-١٤)$$

وباستخدام بيانات المثال ١٠-١ يمكن تكوين المصفوفات التالية:

$$Y = \begin{bmatrix} 50 \\ 77 \\ \vdots \\ 55 \\ 68 \end{bmatrix}, X = \begin{bmatrix} 1 & 150 \\ 1 & 175 \\ \vdots & \vdots \\ 1 & 154 \\ 1 & 165 \end{bmatrix}, \beta = \begin{bmatrix} a \\ b \end{bmatrix}, \text{and } \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_9 \\ \varepsilon_{10} \end{bmatrix}$$

حيث:

Y : متجهة vector حجمه 10 x 1 يحتوى على قيم المتغير التابع

X : مصفوفة حجمها 10 x 2 تحتوى على العوامل المستقلة

β : متجهة حجمه 2 x 1 يحتوى على المعالم المراد تقديرها وهى a، b

ε : متجهة حجمه 10 x 1 يمثل الخطأ

وبتطبيق المعادلة (١٠-١٤) $Y = X\beta + \varepsilon$ يمكن الحصول على:

$$Y = \begin{bmatrix} 50 \\ 77 \\ \vdots \\ 55 \\ 68 \end{bmatrix} = \begin{bmatrix} a + 150\ b \\ a + 175\ b \\ \vdots \\ a + 154\ b \\ a + 165\ b \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_9 \\ \varepsilon_{10} \end{bmatrix}$$

$$(10 \times 1) \qquad (10 \times 1) \qquad (10 \times 1)$$

المعادلتان الاعتياديتان normal equations (١٠-٥)، (١٠-٦) يمكن التعبير عنهما بالمصفوفات كالتالى:

$$X'X\hat{\beta} = X'Y \qquad (١٠-١٥)$$

وبالتعويض:

$$X'X = \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \\ 150 & 175 & \cdots & 154 & 165 \end{bmatrix} \begin{bmatrix} 1 & 150 \\ 1 & 175 \\ \vdots & \vdots \\ 1 & 154 \\ 1 & 165 \end{bmatrix} = \begin{bmatrix} 10 & 1611 \\ 1611 & 260595 \end{bmatrix}$$

$$(2 \times 10) \qquad (10 \times 2) \qquad (2 \times 2)$$

والصورة العامة لهذا الجزء:

$$X'X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$$

$$(2 \times n) \qquad (n \times 2) \qquad (2 \times 2)$$

$$X'Y = \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \\ 150 & 175 & \cdots & 154 & 165 \end{bmatrix} \begin{bmatrix} 50 \\ 77 \\ \vdots \\ 55 \\ 68 \end{bmatrix} = \begin{bmatrix} 631 \\ 102415 \end{bmatrix}$$

$$(2 \times 10) \qquad (10 \times 1) \qquad (2 \times 1)$$

$$X'Y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

$$(2 \times n) \qquad (n \times 1) \qquad (2 \times 1)$$

وبالتالى تكون الصورة العامة للمعادلة (١٠-١٥) كالتالى:

$$\begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} \qquad (١٠-١٦)$$

وبالتعويض

$$\begin{bmatrix} 10 & 1611 \\ 1611 & 2600595 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 631 \\ 102415 \end{bmatrix}$$

حل المعادلة (١٠-١٦) سوف يعطى تقديراً لكل من a، b بطريقة المربعات الصغرى. ولعمل ذلك لابد من الحصول على معكوس inverse المصفوفة التى فى الجانب الأيسر لهذه المعادلة أى إيجاد $(X'X)^{-1}$ وبالتالى يكون:

$$\hat{\beta} = (X'X)^{-1}X'Y \qquad (١٧-١٠)$$

بمعنى أن:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} \qquad (١٨-١٠)$$

ويمكن الحصول على هذا المعكوس كالتالى:

$$(X'X)^{-1} = \begin{bmatrix} \dfrac{\sum X_i^2}{n\sum (X_i - \overline{X})^2} & \dfrac{-\overline{X}}{\sum (X_i - \overline{X})^2} \\ \dfrac{-\overline{X}}{\sum (X_i - \overline{X})^2} & \dfrac{1}{\sum (X_i - \overline{X})^2} \end{bmatrix} \qquad (١٩-١٠)$$

وبأخذ عامل مشترك للجانب الأيمن تصبح المعادلة (١٠-١٩) كالتالى:

$$(X'X)^{-1} = \frac{1}{n\sum (X_i - \overline{X})^2} \begin{bmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{bmatrix} \qquad (٢٠-١٠)$$

وبالتعويض

$$(X'X)^{-1} = \begin{bmatrix} 24.517358 & -0.151566 \\ -0.151566 & 0.0009408 \end{bmatrix}$$

وبتطبيق المعادلة (١٠-١٨) يكون تقدير كل من a، b كالتالى:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 24.517358 & -0.151566 \\ -0.151566 & 0.0009408 \end{bmatrix} \begin{bmatrix} 631 \\ 102415 \end{bmatrix} = \begin{bmatrix} -52.22693 \\ 0.7158717 \end{bmatrix}$$

وهى نفس النتائج المتحصل عليها سابقاً مع بعض الاختلافات نتيجة للتقريب. ومن ذلك تتضح إمكانية وسهولة تطبيق طريق المصفوفات على أى أمثلة أخرى باستخدام نفس المعادلتين (١٠-١٨) و (١٠-٢٠).

مثال ١٠-٣

حل مثال ١٠-١ باستخدام طريقة **IML** فى برنامج SAS، وتستخدم هذه الطريقة عند الرغبة فى استخدام المصفوفات فى الحل

```
PROC IML;
X={1 150, 1 175, 1 170, 1 160, 1 150,
     1 155, 1 152, 1 180, 1 154, 1 165};
  XP = X`;
XPX = XP*X;
    Y= {50, 77, 65, 63, 61, 62, 55, 75, 55, 68};
YP = Y`;
XPY = XP*Y;
XPXINV=INV(XPX);
SOL = XPXINV*XPY;
PRINT XP YP XPX XPY XPXINV SOL;
QUIT;
```

لاحظ:

طريقة كتابة المصفوفات حيث تكتب جميع عناصر المصفوفة داخل قوسين من النوع{ }.

تنتهى عناصر كل صف بفصله عادية.

استخدام علامة ` للحصول على المقلوب transpose (X`) واستخدام INV للحصول على المعكوس inverse.

لابد أن ينتهى البرنامج بكلمة quit.

نتائج التحليل:

XP

	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9	COL10
ROW1	1	1	1	1	1	1	1	1	1	1
ROW2	150	175	170	160	150	155	152	180	154	165

YP

	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9	COL10
ROW1	50	77	65	63	61	62	55	75	55	68

XPX

	COL1	COL2
ROW1	10	1611
ROW2	1611	260595

XPY

	COL1
ROW1	631
ROW2	102415

XPXINV

	COL1	COL2
ROW1	24.517358	-0.151566
ROW2	-0.151566	0.0009408

SOL

	COL1
ROW1	-52.517358
ROW2	0.7158717

صندوق ١٠-١

- عندما يكون هناك متغيران أحدهما وليكن Y يعتمد على متغير آخر وليكن X فيعرف الأول بالتابع والثاني بالمستقل.
- ولكل من المتغيرين تباين ولكن بينهما أيضاً تباين مشترك يطلق عليه التغاير covariance.
- يمكن تقنين العلاقة بين هذين المتغيرين باستخدام معادلة الخط المستقيم باعتبار X على المحور السينى و Y على المحور الصادى وحيث إن العلاقة بين المتغيرين عادة ما تكون غير تامة أى أن Y تحتوى على جزء غير راجع إلى X وهو ما يطلق عليه الخطأ وبالتالى فإن معادلة الخط المستقيم يضاف إليها مكون يعبر عن هذا الخطأ وتكون معادلة الانحدار هى:

$$Y = a + bX + e$$

حيث a ، b ، ثابتان، e الخطأ فى المشاهدة. كما هو الحال فى معادلة الخط المستقيم فإن a تمثل الجزء المقطوع من المحور الصادى بينما b هى ظل زاوية تقاطع الخط المستقيم مع المحور السينى.

- يقدر معامل انحدار Y على X من

$$b_{YX} = \frac{\sum XY - (\sum X \sum Y)/n}{\sum X^2 - (\sum X)^2/n}$$

١٠-٤ مصادر الاختلاف فى الانحدار الخطى

بالرجوع إلى البيانات التى فى مثال ١٠-١ وجد أنه عندما كانت قيمة X للفرد الأول 150 سم فإن قيمة Y له 50 كج، بينما قيمة X للفرد الثانى 175 سم وقيمة Y له 77 كج. ومعنى ذلك أن الوزن أثقل للطول الأكبر، كما يلاحظ أنه بينما كانت قيمة X للفرد الخامس هى أيضاً 150 سم كان الوزن المقابل لها 61 كج. أى أن هناك أيضاً فرقاً فى الوزن حتى لو تساوت الأطوال. ويطلق على هذا الفرق غير المعروف مسبباته خطأ error أو من خط الانحدار from regression. وبالتالى فإن مصادر الاختلاف فى الانحدار الخطى جزء منه راجع إلى اعتماد الوزن على الطول والجزء الآخر خاص بكل فرد. ويمكن بيان ذلك رياضياً كما يلى:

فى الانحدار البسيط الذى يمثل العلاقة بين المتغيرين Y,X، يمكن التعبير عن أى مشاهدة بالنموذج المبين فى (١٠-٢) أى:

$$Y = \alpha + \beta X + \in$$

حيث تمثل $\alpha + \beta X$ متوسط المتغير Y والتى تقابل قيمة محددة للمتغير X ويرمز له μ_{yx} وتمثل $\in$ مكون الخطأ error component، أى أن قيمة Y تمثل مجموع متغيرين أحدهما المتوسط والآخر الخطأ العشوائى. وفى النموذج التقديرى estimated model فإن:

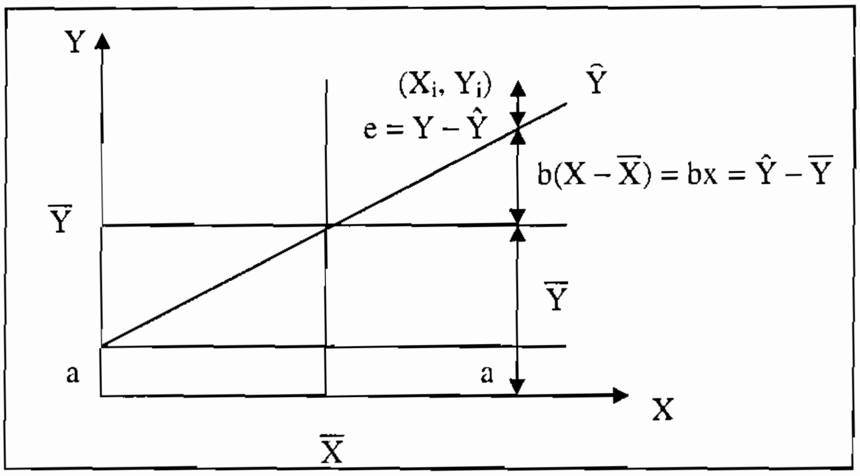
$$Y = a + bX + e$$
$$Y = \overline{Y} + b(X - \overline{X}) + e$$
$$Y = \overline{Y} + (\hat{Y} - \overline{Y}) + (Y - \hat{Y})$$

أى أن:

$$(Y - \overline{Y}) = (\hat{Y} - \overline{Y}) + (Y - \hat{Y}) \qquad (١٠-٢١)$$

ويمكن توضيح ذلك بالشكل ١٠-٦.

ويتضح من الشكل ١٠-٦ إن انحراف القيمة عن متوسطها عبارة عن مجموع حدين. الأول يمثل انحراف القيمة المتوقعة عن المتوسط، والتى تمثل الجزء الراجع إلى اعتماد Y على X ويسمى due to regression أى $(\hat{Y} - \overline{Y}) = b(X - \overline{X}) = bx$. أما الحد الثانى فهو يمثل انحراف القيمة Y عن القيمة المتوقعة $\hat{Y}$ والواقعة على خط الانحدار، وهذا الحد يمثل الخطأ العشوائى ويسمى from regression، أى $e = d_{Y.X} = Y - \hat{Y} = Y - a - bX = y - bx$.



شكل ١٠-٦ مصادر الاختلاف فى Y

ويلاحظ أن مجموع انحرافات قيم $\hat{Y}$ عن المتوسط يساوى صفر، أى $\sum(\hat{Y} - \overline{Y}) = 0$، ومجموع انحرافات Y عن خط الاعتماد أيضاً يساوى صفر أى $\sum(Y - \hat{Y}) = 0$.

وبالتالى من المعادلة (١٠-٢١) يمكن إثبات أن:

$$\sum(Y - \overline{Y})^2 = \sum(\hat{Y} - \overline{Y})^2 + \sum(Y - \hat{Y})^2 \qquad (١٠-٢١)$$

أى أن:

TSS = RSS + ESS

حيث:

TSS عبارة عن مجموع المربعات الكلى total sum of squares،

RSS عبارة عن مجموع المربعات الراجع لانحدار Y على X،
sum of squares due to regression

ESS عبارة عن مجموع المربعات عن خط الانحدار،
sum of squares from regression

هذه القيم يمكن الحصول عليها كالتالى:

$$RSS = \sum(\hat{Y} - \overline{Y})^2 = b^2 \sum x^2 = b \sum xy = \frac{(\sum xy)^2}{\sum x^2}$$

$$ESS = \sum y^2 - RSS = \sum y^2 - \frac{(\sum xy)^2}{\sum x^2}$$

والجدول ١٠-١ يبين مصادر الاختلاف في Y لبيانات مثال ١٠-١

جدول ١٠-١ مصادر الاختلاف فى قيم Y وتقسيم التباين إلى مصادره المختلفة

X	Y	$\hat{Y}$	$Y - \overline{Y}$	$\hat{Y} - \overline{Y}$	$e = Y - \hat{Y}$	$(\hat{Y} - \overline{Y})^2$	$(Y - \hat{Y})^2$
150	50	55.2	-13.1	-7.9	-5.2	62.41	27.04
175	77	73.1	13.9	10.0	3.9	100.00	15.21
170	65	69.4	1.9	6.3	-4.4	39.69	19.36
160	63	62.3	-0.1	-0.8	0.7	0.64	0.49
150	61	55.3	-2.1	-7.9	5.8	62.41	33.64
155	62	58.7	-1.1	-4.4	3.3	19.36	10.89
152	55	56.6	-8.1	-6.5	-1.6	42.45	2.56
180	75	76.6	11.9	13.5	-1.6	182.25	2.56
154	55	58.0	-8.1	-5.1	-3.0	26.01	9.00
165	68	65.9	4.9	2.8	2.1	7.84	4.41
1611	**631**	**631**	**0.0**	**0.0**	**0.0**	**542.86**	**125.16**

مثال ١٠-٤

قسم مجموع المربعات الكلى فى Y إلى مكوناته مستخدماً بيانات المثال ١٠-١.

مجموع المربعات الكلى:

$$TSS = \sum y^2 = \sum Y^2 - \frac{(\sum Y)^2}{n} = 40487 - \frac{(631)^2}{10} = 670.9$$

مجموع المربعات الراجع إلى الانحدار:

$$RSS = \sum(\hat{Y} - \overline{Y})^2 = b^2 \sum x^2 = b \sum xy$$

$$\sum x^2 = \sum X^2 - \frac{(\sum X)^2}{n} = 260595 - \frac{(1611)^2}{10} = 1062.9$$

$$RSS = (0.716)^2 (1062.9) = 544.9$$

مجموع مربعات الخطأ (من خط الانحدار):

$$ESS = \sum(Y - \hat{Y})^2 = TSS - RSS = 670.9 - 544.7 = 126.2$$

وقيم كل من RSS و ESS هى نفس القيم المتحصل عليها فى جدول ١٠-١ والاختلافات بينها راجع إلى التقريب.

يمكن الحصول على نفس النتائج باستخدام المصفوفات كالتالى:

$$TSS = Y'Y - n\overline{Y}^2 \qquad (٢٢-١٠)$$

$$RSS = b'X'Y - n\overline{Y}^2 \qquad (٢٣-١٠)$$

$$ESS = TSS - RSS = Y'Y - b'X'Y \qquad (٢٤-١٠)$$

وباستخدام بيانات مثال ١٠-١

مجموع المربعات الكلى:

$$TSS = Y'Y - n\overline{Y}^2 = \begin{bmatrix} 50 & 77 & \cdots & 55 & 68 \end{bmatrix} \begin{bmatrix} 50 \\ 77 \\ \vdots \\ 55 \\ 68 \end{bmatrix} - \left[(10)(63.1)^2\right] = 670.9$$

و هذا المجموع له 9 درجات حرية أى (n-1).

مجموع المربعات الراجع إلى الانحدار:

$$RSS = \begin{bmatrix} -52.22693 & 0.71587 \end{bmatrix} \begin{bmatrix} 631 \\ 102415 \end{bmatrix} - \left[(10)(63.1)^2 \right] = 544.534$$

وهذا المجموع له درجة حرية واحدة.

مجموع مربعات الخطأ (من خط الانحدار):

$$ESS = TSS - RSS = 670.9 - 544.534 = 126.357$$

و هذا المجموع له 8 درجات حرية أى (n-2).

ويعتبر متوسط مجموع مربعات الخطأ، $(ESS)/(n-2)$، تقدير غير متحيز للتباين σ^2 ويرمز له بالرمز $S^2_{y.x}$.

لاحظ أن نفس النتائج تم الحصول عليها عند استخدام اختيار PROC REG فى برنامج SAS فى فصل ١٠-٤ مع ملاحظة وجود خطأ التقريب.

١٠-٥ القيم المعدلة لأثر انحدار Y على X (Adjusted Y)

كما ذكر سابقاً $\quad Y = \overline{Y} + bx + e$

وعلى ذلك فقيمة Y المعدلة، ويرمز لها Y_A، أى بعد إزالة الجزء الراجع للانحدار bx من قيمة Y هى:

$$\text{Adjusted } Y = Y_A = \overline{Y} + e = Y - bx \qquad (١٠-٢٥)$$

وتفيد القيم المعدلة فى المقارنة بين الأفراد أو بين المتوسطات المعدلة، أى بعد إزالة أثر المتغير المستقل. فمثلاً إذا كان معامل انحدار الوزن على العمر هو 0.8 كج لكل يوم وكان هناك حيوانان الأول عمره 140 يوم ووزنه 200 كج والثانى عمره 168 يوم ووزنه 220 كج، فأيهما أثقل وزناً؟ وللإجابة على ذلك يجب أولا إزالة أثر الاختلاف فى العمر للحيوانين. فإذا كان متوسط العمر هو 105 يوماً ومتوسط الوزن 190 كج فإن:

$$Y_{A_1} = Y_1 - bx_1 = 200 - 0.8(140 - 105) = 172 \text{ kg}$$

$$Y_{A_2} = Y_2 - bx_2 = 220 - 0.8(168 - 105) = 169.6 \text{ kg}$$

ومعنى ذلك أنه لو أن الحيوانين كانا عند نفس العمر (105 يوم مثلاً) لكان التوقع أن يزن الحيوانان 172 و 169.6 كج، على التوالى. وعلى ذلك فالحيوان الأول أكثر وزناً من الحيوان الثانى بعد إزالة العمر وبالتالى يكون هو الأثقل وزناً.

وتجدر الإشارة هنا أنه بحساب معامل الانحدار فإنه يمكن استخراج كل العلاقة بين المتغيرين Y,X وما يترك فى Y بعد هذه العلاقة وهو e فإن العلاقة بينه وبين كل من المتغيرات الأخرى تساوى صفراً أى أن: $\mathrm{Cov}(e,X) = \mathrm{Cov}(e,\hat{Y}) = 0$.

ويمكن التأكد من ذلك بحساب كل من $\sum e\hat{Y}$ و $\sum eX$ من جدول ١٠--١ والتى ستكون مساوية للصفر أيضاً.

١٠-٦ الانحرافات المعيارية وحدود الثقة للتقديرات المختلفة

Standard deviation and confidence limits of estimates

١٠-٦-١ الانحرافات المعيارية

سبق اعتبار أن Y تقدير غير متحيز لـــ $\mu_{y.x}$ وكذلك $\hat{Y},b,a$ تقديرات غير متحيزة لكل من $\mu_{Y.X},\beta,\alpha$ على التوالى.

تباين $\overline{Y}$ هو: $\sigma^2_{\overline{Y}} = \dfrac{\sigma^2}{n}$ وذلك من المعادلة (٣-١٣) بالباب الثالث.

تباين b:

$$\sigma_b^2 = V\left(\frac{\sum xy}{\sum x^2}\right)$$
$$= \frac{1}{(\sum x^2)^2}\left[V(x_1y_1 + x_2y_2 + \ldots + x_ny_n)\right]$$
$$= \frac{1}{(\sum x^2)^2}\left[x_1^2V(y_1) + x_2^2V(y_2) + \cdots + x_n^2V(y_n)\right]$$

وحيث إن: $V(y_1) = V(y_2) = \cdots = V(y_n) = \sigma^2$ ، فإن:

$$\sigma_b^2 = \frac{1}{(\sum x^2)^2}\left[x_1^2\sigma^2 + x_2^2\sigma^2 + \cdots + x_n^2\sigma^2\right]$$

$$= \frac{1}{(\sum x^2)^2}\sigma^2\sum x^2 = \frac{\sigma^2}{\sum x^2} \qquad (١٠-٢٦)$$

تباين الجزء المقطوع من محور الصادات α :

يمكن بنفس المفهوم إثبات أن:

$$\sigma_a^2 = \sigma^2\left(\frac{1}{n} + \frac{\overline{X}^2}{\sum x^2}\right) \qquad (١٠-٢٧)$$

وتباين متوسط العشيرة الذى يقابل قيمة محددة لـــ X أى $\mu_{y.x}$ هو:

$$\sigma_{\mu_{y.x}}^2 = \sigma^2\left(\frac{1}{n} + \frac{x_i^2}{\Sigma x_i^2}\right) \qquad (١٠-٢٨)$$

وتباين القيمة المتوقعة التى تقابل قيمة محددة لـــ X هى:

$$\sigma_{\hat{Y}}^2 = \sigma^2\left(1 + \frac{1}{n} + \frac{x_i^2}{\Sigma x_i^2}\right) \qquad (١٠-٢٩)$$

حيث القيمة المتوقعة $\hat{Y}$ هى تقدير لنقطة جديدة لـــ Y التى تقابل قيمة محددة لـــ X. وبأخذ الجذر التربيعى للمعادلات السابقة يمكن الحصول على الانحراف المعيارى للتقديرات المختلفة.

فمثلاً الانحراف المعيارى لمعامل الانحدار SE_b أو S_b هو

$$S_b = \frac{S_{y.x}}{\sqrt{\sum x^2}} \qquad (١٠-٣٠)$$

حيث $S_{y.x}^2$ هى تقدير لـــ σ^2 علماً بأن $S_{y.x}^2 = ESS/(n-2)$ كما ذكر من قبل. و ESS هى مجموع مربعات الخطأ و n عدد أزواج المتغيرات (المشاهدات).

وتعبر $S_{y.x}$ عن الخطأ القياسى للتقدير أو الانحراف المعيارى لـــ Y باعتبار أن X ثابتة.

ومما سبق يمكن كتابة مصفوفة التباين والتغاير variance-covariance matrix للتقديرات المختلفة كالتالى:

$$V(\hat{\beta}) = \begin{bmatrix} V(a) & Cov(a,b) \\ Cov(a,b) & V(b) \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\sigma^2 \sum X_i^2}{n\sum(X_i - \overline{X})^2} & -\dfrac{\overline{X}\sigma^2}{\sum(X_i - \overline{X})^2} \\ -\dfrac{\overline{X}\sigma^2}{\sum(X_i - \overline{X})^2} & \dfrac{\sigma^2}{\sum(X_i - \overline{X})^2} \end{bmatrix} \qquad (١٠-٣١)$$

وبأخذ عامل مشترك σ^2 فإن المعادلة (١٠-٣١) تصبح:

$$V(\hat{\beta}) = (X'X)^{-1}\sigma^2 \qquad (١٠-٣٢)$$

وللحصول على تقدير لمتوسط العشيرة (باستخدام جبر المصفوفات) عند قيمة معينة ولتكن X_o ضع $X'_o = [1 \quad X_o]$، وبالتالى يمكن تقدير متوسط العشيرة عند قيمة X_o كما يلى:

$$\hat{\mu}_{y.x} = [1 \quad X_o]\begin{bmatrix} a \\ b \end{bmatrix} = X'_o b = b'X_o$$

وحيث إن تباين $\hat{\mu}_{y.x}$ عبارة عن

$$V(\hat{\mu}_{y.x}) = V(a) + 2X_o cov(a,b) + X_o^2 V(b)$$

فإن

$$V(\hat{\mu}_{y.x}) = [1 \quad X_o]\begin{bmatrix} V(a) & cov(a,b) \\ cov(a,b) & V(b) \end{bmatrix}\begin{bmatrix} 1 \\ X_o \end{bmatrix}$$

$$= X'_o (X'X)^{-1}\sigma^2 X_o$$

$$= X'_o (X'X)^{-1} X_o \sigma^2$$

وهذه مطابقة تماماً للمعادلة (١٠-٢٩).

عندما تكون $X_0 = 156$ (فى مثال ١٠-١) فإن

$$\hat{\mu}_{y.x} = [1 \quad 156]\begin{bmatrix} -52.22693 \\ 0.71587 \end{bmatrix} = 59.45$$

$$V(\hat{\mu}_{y.x}) = X_0'(X'X)^{-1}X_0\sigma^2_{y.x}$$

وحيث إن $\sigma^2_{y.x} = 15.77416$ تعتبر تقديراً غير متحيز لـ $\sigma^2_{y.x}$ فإن

$$V(\hat{\mu}_{y.x}) = [1 \quad 156]\begin{bmatrix} 24.517358 & -0.151566 \\ -0.151566 & 0.0009408 \end{bmatrix}\begin{bmatrix} 1 \\ 156 \end{bmatrix}(15.77416)$$
$$= 1.956841$$

وبالتالى

$$S_{\hat{\mu}_{Y.X}} = \sqrt{1.956841} = 1.398871$$

تقدير تباين القيمة المتوقعة $\hat{Y}$ عند قيمة معينة للمتغير X:

عندما يكون الغرض هو التنبؤ prediction بقيمة المتغير التابع Y عند قيمة معينة للمتغير X ولتكن X_0 (وليس التقدير estimation كما هو الحال عند تقدير متوسط العشيرة عند نفس النقطة X) فإن القيمة المتوقعة تكون هى نفسها مساوية لمتوسط العشيرة عند تلك النقطة ولكن بتباين أكبر هو:

$$V(\hat{Y}_{X=X_o}) = [1 + X_0'(X'X)^{-1}X_0]\sigma^2_{y.x}$$

وتستخدم $S^2_{Y.X}$ فى حالة عدم معرفة $\sigma^2_{Y.X}$

١٠-٦-٢ حدود الثقة لكل من a، b، $\hat{\mu}_{Y.X}$، $\hat{Y}$

بنفس المفهوم الذى استخدم لإيجاد حدود الثقة لكل من المتوسط والتباين فى الباب السادس فصل ٦-٢ يمكن إيجاد حدود الثقة كما يلى:

حدا الثقة لــ a هما:

$$(\overline{Y} - b\overline{X}) \pm t_\alpha S_{y.x}\sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum x^2}} \qquad (١٠-٣٣)$$

حدا الثقة لــ b هما:

$$b \pm t_\alpha \frac{S_{y.x}}{\sqrt{\sum x^2}} \qquad (١٠-٣٤)$$

حدا الثقة لــ $\hat{\mu}_{Y.X}$: أى حدا الثقة لمتوسط العشيرة الذى يقابل قيمة معينة لــ X (حيث $\mu_{Y.X} = \overline{Y} + bx$) هما:

$$\hat{Y} \pm t_\alpha S_{y.x}\sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum x^2}} \qquad (١٠-٣٥)$$

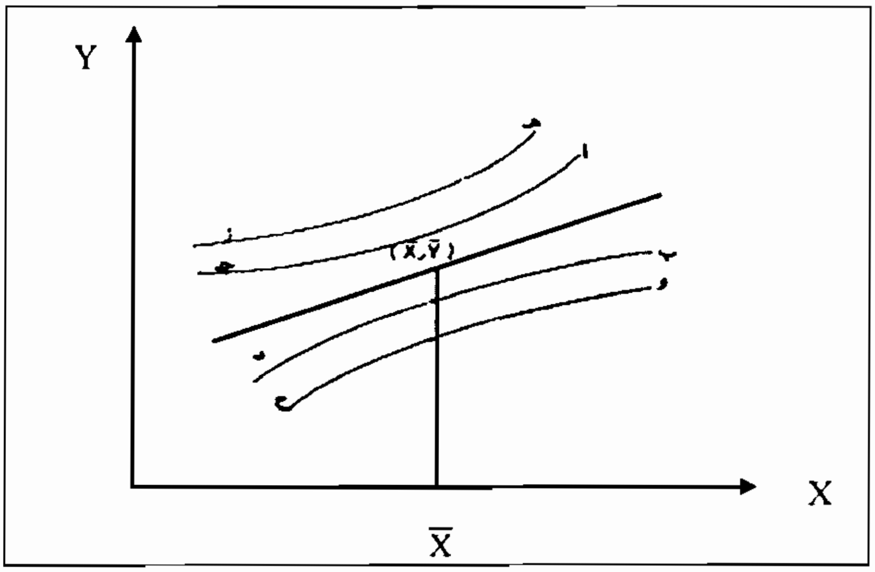
حيث $\hat{Y} = \overline{Y} + bx$.

وهذه تمثل حدى الثقة لخط الانحدار. وهذه القيم تشكل خطين وتسمى المساحة المحصورة بينهما بحزام الثقة confidence belt عند مستوى احتمالى α (مستوى المعنوية).

أما حدا الثقة للقيمة المتوقعة $\hat{Y}$ فهما:

$$(\overline{Y} - bx) \pm t_\alpha S_{y.x}\sqrt{1 + \frac{1}{n} + \frac{\overline{X}^2}{\sum x^2}} \qquad (١٠-٣٦)$$

وشكل ١٠-٧ يبين حزام الثقة حول $\hat{\mu}_{Y.X}$ وحزام الثقة حول $\hat{Y}$.



شكل ١٠-٧ حزام الثقة حول $\hat{\mu}_{Y.X}$ وهو أ ب ج د وحزام الثقة حول $\hat{Y}$ وهو هـ و ز ح.

وفى المعادلات السابقة t_α هى قيمة t الجدولية بدرجات حرية $n-2$ ومستوى معنوية α.

مثال ١٠-٥

احسب حدى الثقة لمعامل اعتماد Y على X باستخدام بيانات المثال ١٠-١ علماً بأن مستوى المعنوية 5%.

حدا الثقة هما:

$$b \pm t_\alpha \frac{S_{y.x}}{\sqrt{\sum x^2}}$$

حيث

$$S_{Y.X} = \sqrt{\frac{\sum y^2 - [(\sum xy)^2/(\sum x^2)]}{(n-2)}} \ ، \ \sum x^2 = \sum X^2 - \frac{(\sum X)^2}{n} \ ،$$

$$\sum y^2 = \sum Y^2 - \frac{(\sum Y)^2}{n}$$

وبالتعويض فإن:

$$\sum x^2 = 260595 - \frac{(1611)^2}{10} = 1062.9$$

$$\sum y^2 = 40487 - \frac{(631)^2}{10} = 670.9$$

$$\sum xy = 102415 - \frac{(1611)(631)}{10} = 760.9$$

$$S_{Y.X} = \sqrt{\frac{670.9 - [(760.9)^2/(1062.9)]}{(10-2)}} = \sqrt{\frac{670.9 - 544.7}{8}} = 3.97$$

$$S_b = \frac{S_{y.x}}{\sqrt{\sum x^2}} = \frac{3.97}{\sqrt{1062.9}} = 0.122$$

وبالتالى فإن حدى الثقة لمعامل الاعتماد (الانحدار) هما:

$$0.72 \pm (2.306)(0.122) = 0.72 \pm 0.28$$

أى أن الحد الأدنى للثقة هو 0.44 والحد الأعلى للثقة هو 1.00

مثال ١٠-٦

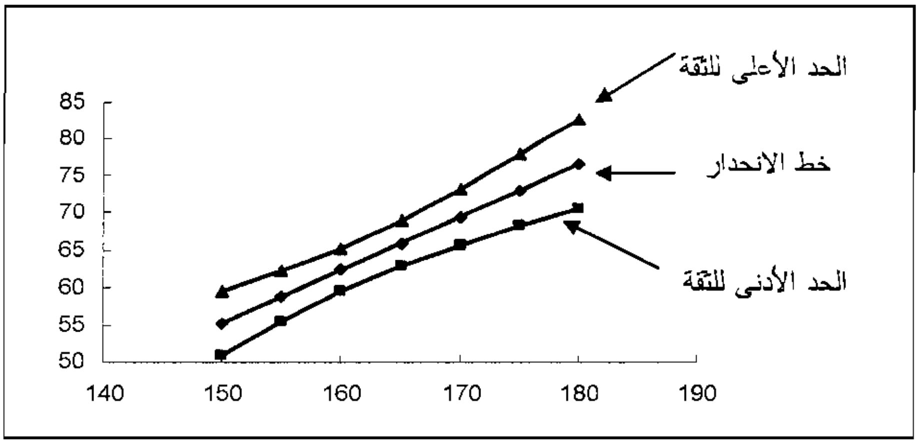
احسب حدود الثقة لخط الانحدار بمستوى معنوية 5% باستخدام بيانات المثال ١٠-١، وماذا تعنى ثم بين ذلك بيانياً.

لحساب حدود الثقة لخط الانحدار يلزم أولاً حساب حدود الثقة لعدد من $\mu_{y.x}$ ثم تمثل هذه النقط بيانياً ويتم التوصيل فيما بينها فينتج خطين أحدهم يمثل الحد الأعلى لخط الانحدار والآخر يمثل الحد الأدنى لخط الانحدار.

حيث إن قيمة t الجدولية بمستوى معنوية 5% ودرجات حرية 8 هى 2.306 وقيمة $S_{y.x} = 3.97$ وباستخدام المعادلة (١٠-٣٥) فإنه يمكن الحصول على النقاط التالية:

حدود الثقة		$\hat{Y} \pm t_{\alpha} S_{y.x} \sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum x^2}}$	$\hat{Y}$	X
الحد الأعلى	الحد الأدنى			
59.40	50.90	55.15 ± 4.25	55.15	150
62.09	55.37	58.73 ± 3.36	58.73	155
65.22	59.40	62.31 ± 2.91	62.31	160
68.99	62.79	65.89 ± 3.10	65.89	165
73.29	65.65	69.47 ± 3.82	69.47	170
77.91	68.19	73.05 ± 4.86	73.05	175
82.68	70.58	76.63 ± 6.05	76.63	180

ويمكن تمثيل حدود الثقة لخط الانحدار لهذه البيانات بيانياً بالشكل التالى:



وحزام الثقة هذا يعنى أن احتمال أن يحوى الحزام خط الاعتماد الحقيقى $(\alpha + \beta X)$ هو 0.95.

١٠-٧ اختبار معنوية معامل الانحدار

يمكن اختبار الفرض القائل بأن معامل انحدار العشيرة β يساوى الصفر، أى لا توجد علاقة بين المتغيرينX, Y باستخدام اختبار t كما يلى:

فرض العدم: $H_o : \beta = 0$. وتحت صحة فرض العدم فإن $\beta = 0$ وبالتالى:

$$t = \frac{b - \beta}{S_b} = \frac{b\sqrt{\sum x^2}}{S_{y.x}}$$

وتقارن هذه القيمة بقيمة t الجدولية بدرجات حرية $n-2$ ومستوى معنوية وليكن α.

أما إذا كان فرض العدم هو $Ho : \beta = c$ حيث c هى قيمة معينة لمعامل انحدار العشيرة، فيكون الفرض البديل أن $H_1 : \beta \neq c$ وتكون

$$t = \frac{b - c}{S_b}$$

مثال ١٠-٧

اختبر معنوية معامل انحدار الوزن على الطول لبيانات **مثال ١٠-١** باستخدام اختبار t.

فرض العدم: $H_o : \beta = 0$ ومن مثال ١٠-٥ $S_b = S_{y.x}/\sqrt{\sum x^2} = 0.12177$
إذاً

$$t = \frac{b - \beta}{S_b} = \frac{0.71587}{0.12177} = 5.88$$

ومن جدول t تظهر القيم $t_{(8,0.05)} = 2.306$ ، $t_{(8,0.01)} = 3.350$

وبالتالى يرفض فرض العدم وبالتالى فإن معامل الانحدار يختلف عن الصفر معنوياً، أى أن العلاقة بين Y ، X معنوية بدرجة ثقة 99%.

يمكن اختبار معنوية معامل الانحدار باستخدام تحليل التباين عن طريق تكوين جدول تحليل التباين ANOVA table. فكما سبق وذكر أن مجموع المربعات الكلى فى Y يقسم إلى مكونين أحدهما راجع إلى الانحدار due to regression والآخر عن

خط الانحدار deviations from regression. وبالتالى فإن جدول ١٠-٢ لتحليل لتباين يمكن تكوينه كالتالى:

جدول ١٠- ٢ تحليل التباين باستخدام تحليل الانحدار

SOV مصدر التباين	df	SS مجموع المربعات	MS متوسط المربعات	F
Due to regression راجع للانحدار	1	$RSS = b'X'Y$		$RSS/S^2_{y.x}$
From regression من الانحدار	n-2	$ESS = \sum y^2_{y.x} = Y'Y - b'X'Y$	$S^2_{y.x}$	
C. Total الكلى المصحح	**n-1**	$\sum y^2 = Y'Y - n\overline{Y}^2$		

وتقارن قيمة F المحسوبة بقيمة F الجدولية بدرجات حرية 1، $n-2$ ومستوى معنوية وليكن α.

وحيث إن t هى $(b\sqrt{\sum x^2})/(S_{y.x})$ فإن قيمة t^2 هى $(b^2\sum x^2)/(S^2_{y.x})$ وهى نفس قيمة F أى أن $F = t^2$ فى حالة الانحدار البسيط.

مثال ١٠-٨

كون جدول تحليل التباين واختبر معنوية معامل الانحدار باستخدام بيانات المثال ١٠-٤.

لاحظ أن جدول تحليل التباين المطلوب قد تم الحصول عليه باستخدام برنامج SAS عند حل مثال ١٠-٢، والجدول كان كالتالى:

SOV	df	SS	MS	F
Due to regression	1	544.7	5.447	34.52**
From regression	8	126.2	15.775	
C. Total	**9**	**670.9**		

قيم F الجدولية $F_{(1,8,0.05)} = 5.32$ ، $F_{(1,8,0.01)} = 11.26$

قيمة F المحسوبة أكبر من قيمة F الجدولية عند مستوى معنوية 5%، 1% .

إذاً العلاقة معنوية بدرجة ثقة 99% بين المتغيرين X و Y.

وحيث إن قيمة t المستخرجة فى المثال ١٠-٧ هى 5.88 ومربعها هو 34.57 ، وهى نفس قيمة F (مع مراعاة الاختلاف البسيط بين مربع قيمة t وقيمة F نتيجة للتقريب) وكلتا الطريقتين متطابقتان تماماً.

صندوق ١٠-٢

- معامل الانحدار هو متوسط التغير فى المتغير التابع Y لكل وحدة تغير فى المتغير المستقل X. وهو قيمة مميزة (أى كج، سم، أردب، وحدة سماد ... الخ) وتتراوح قيمته بين $+\infty$ و $-\infty$.
- بينما تتيح دراسة الانحدار بحث العلاقة بين متغيرين يفترض أن أحدهما مستقل وثابت (أى بدون توزيع احتمالى) والآخر تابع وله توزيع احتمالى (طبيعى مثلا) فإنه أيضا يمكن من:
  - تقسيم التباين الكلى $\sum y^2$ فى Y إلى جزء راجع إلى التغير فى X وهو $b\sum xy$ وجزء آخر راجع إلى عوامل غير X وهو $\sum y^2 - b\sum xy$.
  - استنباط معادلة الانحدار $\hat{Y} = \alpha + bX$ وبذلك يمكن التنبؤ بأى قيمة للمتغير التابع عند أى قيمة للمتغير المستقل فى حدود مدى الدراسة.
  - اختبار المعنوية للعلاقة b_{YX} وكذلك استنباط حدود الثقة عند مستوى احتمالى معين.

١٠-٨ المقارنة بين معاملى الانحدار

يستخدم اختبار t لمقارنة معاملى الانحدار فى عينتين لمعرفة ما إذا كان هذان المعاملان هما تقدير لنفس معامل انحدار العشيرة β أم لا بنفس المفهوم المستخدم فى مقارنة المجاميع group comparison، كما هو موضح فى الباب الثامن ويكون ذلك كما يلى:

$$H_o : \beta_1 = \beta_2 \quad \text{بمعنى} \quad H_o : \beta_1 - \beta_2 = 0$$

وتحت صحة فرض العدم فإن الكمية $\beta_1 - \beta_2 = 0$ تتبع توزيع t بدرجات حرية $[(n_1 - 2) + (n_2 - 2)]$ وبالتالى فإن

$$t = \frac{(b_1 - b_2) - (\beta_1 - \beta_2)}{\sqrt{S^2_{b_1 - b_2}}} = \frac{b_1 - b_2}{S_{b_1 - b_2}} \qquad (١٠-٣٧)$$

حيث

$S^2_{b_1 - b_2} = S^2_p / \sum x_1^2 + S^2_p / \sum x_2^2$ ،

$\sum x_1^2$ مجموع مربع الانحرافات عن المتوسط بالنسبة للمتغير X_1 فى العينة الأولى،

$\sum x_2^2$ مجموع مربع الانحرافات عن المتوسط بالنسبة للمتغير X_2 فى العينة الثانية،

S^2_p تمثل التباين المشترك pooled variance والذى يعبر عنه كما يلى

$$S^2_p = \frac{(\sum y_1^2 - b_1^2 \sum x_1^2) + (\sum y_2^2 - b_2^2 \sum x_2^2)}{(n_1 - 2) + (n_2 - 2)}$$

أى مجموع المربعات عن خط الإعتماد فى العينة الأولى + مجموع المربعات عن خط الاعتماد فى العينة الثانية مقسوماً على مجموع درجات الحرية للخطأ فى العينتين.

كما يستخدم أيضاً تحليل التباين لإجراء هذا الاختبار ويمكن الرجوع فى ذلك إلى Steel and Torrie (1980).

مثال ١٠-٩

قياسان Y,X أخذا على أفراد عينتين وكانت البيانات التالية:

العينة الأولى:

$n_1 = 10$ ، $b_1 = 1.26$ وحدة من Y لكل وحدة من X

مجموع المربعات عن خط الاعتماد = 2.3 ، $\sum x_1^2 = 17.5$

العينة الثانية:

$n_2 = 10$ ، $b_2 = 1.35$ وحدة من Y لكل وحدة من X

مجموع المربعات عن خط الاعتماد = 4 ، $\sum x_2^2 = 19$

اختبر الفرض القائل بأن معاملى الانحدار هما نفس التقدير لمعامل انحدار العشيرة.

الحل:

$$H_o : \beta_1 - \beta_2 = 0 \text{ ، } S^2_{y.x} = \frac{2.3+4}{10+10-4} = 0.39$$

$$S^2_{b_1-b_2} = 0.39\left(\frac{1}{17.75}+\frac{1}{19}\right) = 0.04 \text{ وبالتالى } S_{b_1-b_2} = \sqrt{0.04} = 0.2$$

$$\therefore t = \frac{|1.35-1.26|}{0.2} = 0.045$$

قيمة t الجدولية $t_{(16,0.05)} = 2.12$ وهى أكبر من t المحسوبة (0.45) وبالتالى لا يرفض فرض العدم وعلى هذا فإن b_2, b_1 هما تقديران لنفس معامل انحدار العشيرة β وتحسب β العامة للمثال لأنه ليس هناك سبب لوجود معاملى انحدار ولكن معامل انحدار واحد.

١٠-٩ التوزيع ذو المتغيرين

فى كثير من الحالات يكون لكل من المتغيرين Y,X توزيع احتمالى بمعنى أن أزواج المتغيرين كل منهما مسحوب عشوائياً. فمثلاً كمية اللبن التي تعطيها البقرة فى الموسم والعمر عند أول ولادة لكل بقرة إذا سحبت 20 بقرة عشوائياً وتم تسجيل العمر

وكمية اللبن لكل بقرة. أو عند دراسة العلاقة بين عدد لطع دودة ورق القطن وكمية المحصول فى عدد من القطع اختيرت عشوائيا. فى مثل هذه الحالات تكون العينات مسحوبة من عشيرة ذات متغيرين، وفى هذه الحالة لا توضع قيود أو شروط على أى من المتغيرين، كأن يكون أحدهما ثابتاً fixed مثلاً. ويمكن حساب معاملين للانحدار أحدهما معامل انحدار المتغير الأول مثلاً على المتغير الثانى، والآخر معامل انحدار المتغير الثانى على المتغير الأول. ولو أنه فى كثير من الأحيان يكون الاهتمام مقصوراً على حساب معامل انحدار واحد له معنى معين. وفى العينات المسحوبة من عشيرة ذات متغيرين فإنه يمكن تلخيص بيانات العينة المسحوبة وذلك بحساب متوسط وتباين كل من المتغيرين على حدة وكذلك حساب التغاير بينهما ويمكن أن يعبر عن التغاير بمعامل التحديد coefficient of determination أو الجذر التربيعى له والذى يعرف بمعامل الارتباط البسيط كما سيتضح فيما بعد.

مثال ١٠-١٠

فى عينة من 8 أزواج كانت البيانات التالية:

رقم المشاهدة	١	٢	٣	٤	٥	٦	٧	٨
Y_1	6	4	0	10	14	2	12	8
Y_2	9	3	2	10	15	8	11	12

احسب معامل انحدار Y_1 على Y_2 وكذلك معامل انحدار Y_2 على Y_1 مع تقسيم التباين فى كل من المتغيرين إلى مكوناته.

الحل:

$$\sum Y_1 = 56 \qquad \overline{Y}_1 = 7 \qquad \sum Y_1^2 = 560 \qquad \sum y_1^2 = 168$$

$$\sum Y_2 = 70 \qquad \overline{Y}_2 = 8.75 \qquad \sum Y_2^2 = 748 \qquad \sum y_2^2 = 135.5$$

$$\sum Y_1 Y_2 = 620 \qquad \sum y_1 y_2 = 130$$

معامل انحدار Y_1 على Y_2 0.959 وحدة من Y_1 لكل وحدة من Y_2 ناتج من

$$b_{Y_1Y_2} = \frac{\sum y_1y_2}{\sum y_2^2} = \frac{130}{135.5} = 0.959$$

معامل انحدار Y_2 على Y_1 0.774 وحدة من Y_2 لكل وحدة من Y_1 ناتج من

$$b_{Y_2Y_1} = \frac{\sum y_1y_2}{\sum y_1^2} = \frac{130}{168} = 0.774$$

تباين المتغير الأول: $S_{Y_1}^2 = \frac{168}{7} = 24$ والانحراف المعيارى 4.9

تباين المتغير الثانى: $S_{Y_2}^2 = \frac{135.5}{7} = 19.36$ والانحراف المعيارى 4.4

التغاير بين المتغيرين: $\text{cov}(Y_1, Y_2) = \frac{130}{7} = 18.57$

إذا اعتبر أن Y_2 هو المتغير التابع فإن مجموع المربعات الراجع لانحدار Y_2 على Y_1 (RSS):

$$RSS = \frac{(\sum y_1y_2)^2}{\sum y_1^2} = \frac{(130)^2}{168} = 100.6$$

مجموع المربعات الكلى للمتغير Y_2: $TSS = 135.5$

مجموع المربعات عن خط انحدار Y_2 على Y_1:

$$ESS = 135.5 - 100.6 = 34.9$$

وعلى ذلك فإن:

$$\frac{RSS}{TSS} = \frac{(\sum y_1y_2)^2/\sum y_1^2}{\sum y_2^2} = \frac{(130)^2/168}{135.5} = 0.74$$

أما إذا اعتبر أن Y_1 هو المتغير التابع فإن مجموع المربعات الراجع لانحدار Y_2 على Y_1:

$$RSS = \frac{(\sum y_1 y_2)^2}{\sum y_2^2} = \frac{(130)^2}{135.5} = 124.7$$

مجموع المربعات عن خط انحدار Y_1 على Y_2:

$$ESS = 168 - 124.7 = 43.3$$

وعلى ذلك فإن:

$$\frac{RSS}{TSS} = \frac{(\sum y_1 y_2)^2 / \sum y_1^2}{\sum y_1^2} = \frac{(130)^2/135.5}{168} = 0.74$$

أما حاصل ضرب معاملى الانحدار

$$(b_{Y_1 Y_2})(b_{Y_2 Y_1}) = (0.959)(0.774) = 0.74$$

وبالتالى يوجد معادلتان للانحدار يمثلان خطى الانحدار يمر كل منهما بنقطة تقاطع متوسطى المتغيرين ولا يشترط أن ينطبق الخطين كل منهما على الآخر، إذ لا يتحقق ذلك إلا إذا كان مجموع مربع الانحرافات عن خط الانحدار مساوياً للصفر أى أن جميع النقط واقعة على خط الانحدار بدون أى أخطاء كما يلاحظ أن معامل انحدار Y_2 على Y_1 ليس هو مقلوب معامل انحدار Y_1 على Y_2.

معادلة خط انحدار Y_2 على Y_1 هى:

$$\hat{Y}_2 = a_1 + b_{Y_2 Y_1} Y_1$$

حيث:

$a_1 = \overline{Y}_2 - b_{Y_2 Y_1} \overline{Y}_1$ وبالتعويض $a_1 = 8.75 - (0.774)(7) = 3.33$

وعلى ذلك فإن معادلة خط الانحدار $\hat{Y}_2 = 3.33 + 0.774 Y_1$

أما معادلة خط انحدار Y_1 على Y_2 فهى:

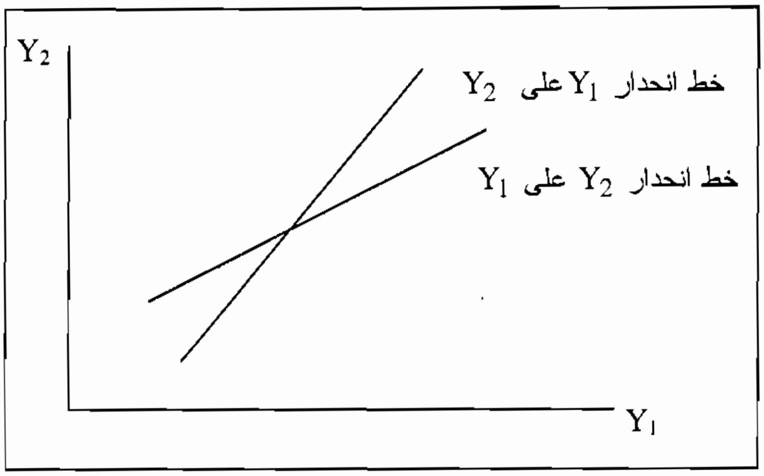
$$\hat{Y}_1 = a_2 + b_{Y_1 Y_2} Y_2$$

حيث:

$a_2 = \overline{Y}_1 - b_{Y_1 Y_2} \overline{Y}_2$ وبالتعويض $a_2 = 7 - (0.959)(8.75) = -1.39$

وعلى ذلك فإن معادلة خط الانحدار $\hat{Y}_2 = -1.39 + 0.9594 Y_2$

ويبين الشكل التالي خطى الانحدار.



ويلاحظ أيضاً أن النسبة بين مجموع المربعات الراجعة للانحدار إلى مجموع المربعات الكلى عندما كان Y_2 هو المتغير التابع هى نفسها عندما كان Y_1 هو المتغير التابع، أى أن هذه النسبة والتى تساوى 0.74 هى نفسها بغض النظر عن كون أى المتغيرين هو التابع وهذه النسبة تساوى حاصل ضرب معاملى الانحدار أى أن:

$$(b_{Y_1Y_2})(b_{Y_2Y_1}) = \frac{(\sum y_1y_2)^2/\sum y_1^2}{\sum y_2^2} = \frac{(\sum y_1y_2)^2/\sum y_2^2}{\sum y_1^2}$$

ويرمز لهذه الكمية بالرمز r^2 ويطلق عليها معامل التحديد coefficient of determination أى أن:

$$r^2 = (b_{Y_1Y_2})(b_{Y_2Y_1}) = \frac{(\sum y_1y_2)^2}{(\sum y_1^2)(\sum y_2^2)} \qquad (١٠-٣٨)$$

والجذر التربيعى لمعامل التحديد (r) يطلق عليه معامل الارتباط (التلازم) البسيط. وكثيراً ما يعرف معامل الارتباط بأنه معامل اعتماد قياسى standard regression، أى أنه إذا قيست المشاهدات فى كل من المتغيرين بوحدات قياسية (أى مقسومة على انحرافها المعيارى) كان معامل اعتماد Y على X فى هذه الحالة يساوى معامل اعتماد X على Y يساوى الارتباط بينهما، فإذا رمز لمعامل الاعتماد القياسى هذا بالرمز $\mathring{b}_{YX}$ فإن:

$$b_{YX}^{\circ} = \frac{\sum \frac{x}{\sigma_x} \sum \frac{y}{\sigma_y}}{\sum \frac{x^2}{\sigma_x^2}} = \frac{\sum xy}{\sum x^2} \cdot \frac{\sigma_x^2}{\sigma_x \sigma_y}$$

$$b_{YX}^{\circ} = b_{yx} \frac{\sigma_x}{\sigma_y} = b_{xy} \frac{\sigma_y}{\sigma_x} = r \qquad (١٠-٣٩)$$

وسوف يتم فيما بعد شرح معامل الارتباط بالتفصيل.

١٠-١٠ العلاقة بين معامل التحديد r^2 وخطأ التقدير $S_{y.x}$

سبق بيان أن مجموع مربع الانحرافات عن خط الانحدار وكان رمزه ESS أو $d_{y.x}^2$ هو

$$ESS = \sum d_{y.x}^2 = \sum y^2 - \frac{(\sum xy)^2}{\sum x^2} = \sum y^2 \left(1 - \frac{(\sum xy)^2}{\sum x^2 \sum y^2}\right)$$

وحيث إن:

$$(b_{Y_1 Y_2})(b_{Y_2 Y_1}) = \frac{(\sum y_1 y_2)^2}{(\sum y_1^2)(\sum y_2^2)} = r^2$$

$$ESS = \sum y^2 (1 - r^2) \qquad (١٠-٤٠)$$

وحيث إن ESS هى مجموع مربعات فإن قيمتها تساوى صفر على الأقل ولكى يتحقق ذلك يلزم أن تكون قيمة r^2 ما بين الصفر والواحد الصحيح أى أن:

$$\left(0 \le r^2 \le 1\right) \qquad (١٠-٤١)$$

وهذه إحدى خصائص معامل التحديد.

وعندما تكون n أى حجم العينة كبيراً فإن $n-1$ تساوى تقريباً $n-2$ وعلى ذلك فإن العلاقة (١٠-٤٠) تصبح:

$$\frac{ESS}{n-2} \cong \frac{\sum y^2}{n-1}(1-r^2)$$

أى أن:

$$S_{y.x}^2 \cong S_y^2(1-r^2) \qquad (١٠-٤٢)$$

وبأخذ الجذر التربيعى للطرفين:

$$S_{y.x} \cong S_y\sqrt{(1-r^2)} \qquad (١٠-٤٣)$$

حيث $S_{y.x}$ تمثل خطأ التقدير أما S_y فتمثل الانحراف المعيارى للمتغير Y.

وأيضاً:

$$\frac{S_{y.x}^2}{S_y^2} \cong 1-r^2 \qquad (١٠-٤٤)$$

$$r^2 \cong 1-\frac{S_{y.x}^2}{S_y^2} \cong \frac{S_y^2 - S_{y.x}^2}{S_y^2} \qquad (١٠-٤٥)$$

وعندما تكون $r^2 = 1$، وهو الحد الأقصى لقيمة معامل التحديد، فإن: $S_{y.x}^2$ لابد وأن تساوى الصفر. وعلى ذلك فإن قيمة r^2 تساوى صفر فى حالة عدم وجود علاقة بين المتغيرين أى عندما يكون التباين هو نفسه التباين عن خط الانحدار وكما أن r^2 تساوى الواحد الصحيح عندما تكون جميع النقط واقعة على خط الانحدار.

وإذا وصف معامل التحديد r^2 كما سبق بأنه النسبة من التباين فى Y والتى ترجع إلى اعتماد Y على X فإن المقدار $(1-r^2)$ يمثل النسبة من التباين فى أحد المتغيرين الخالية من تأثير المتغير الثانى. وعلى سبيل المثال إذا كانت $r^2 = 0.74$ (كما فى مثال ١٠-١٠) فمعنى ذلك أن 74% من الاختلاف أو التباين فى أحد المتغيرين ترجع إلى الاختلافات فى قيم المتغير الآخر وأن $1-0.74 = 0.26$ من التباين راجعة إلى الخطأ العشوائى أو التباين الغير منسوب إلى تباين قيم المتغير الثانى.

ويمكن اختبار مدى معنوية معامل التحديد باختبار فرض العدم $H_0 : \rho = 0$ ضد الفرض البديل $H_o : \rho \neq 0$ حيث ρ عبارة عن معامل ارتباط العشيرة. ولاختبار ذلك الفرض تستخدم t بدرجات حرية $n-2$ كالتالي:

$$t = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} \qquad (٤٦-١٠)$$

وعلى الرغم من أن معامل التحديد هو المقياس الأكثر استخداماً، إلا أن هذا لا يعطي دليلا كافياً على أن r^2 فعلا تفسر قيمة الاختلافات في المتغير التابع والراجعة إلى المتغير المستقل. ولذلك لابد من تقييم مدى ملائمة evaluating the fit النموذج المستخدم في التحليل.

١٠-١١ تقييم ملائمة نموذج التحليل Evaluating the fit

حتى الآن تم شرح النتائج الأساسية والتي تستخدم للوصول إلى استنتاجات فيما يتعلق بمفهوم النموذج الخطي البسيط. وهذه النتائج تكون صالحة وذات معنى حتى الآن فيما يتعلق فقط بالخطأ التجريبي (أو المتبقي residual) في النموذج المستخدم. وكثيراً ما يجب تحليل مكونات هذا المتبقي من خلال عمل بعض الرسومات البيانية والتحليلية. وكما سبق فإن كبر قيمة معامل التحديد ومعنوية اختبار t له لا يؤكدان أن النموذج المستخدم يعتبر هو الأكثر ملائمة لتحليل البيانات محل الدراسة. والمثال التالي يوضح ذلك. ولابد من الأخذ في الاعتبار أن التقييم الدقيق للمتبقي له أهمية في التأكيد على المحافظة على الافتراضات الخاصة باستخدام نظرية المربعات الصغرى.

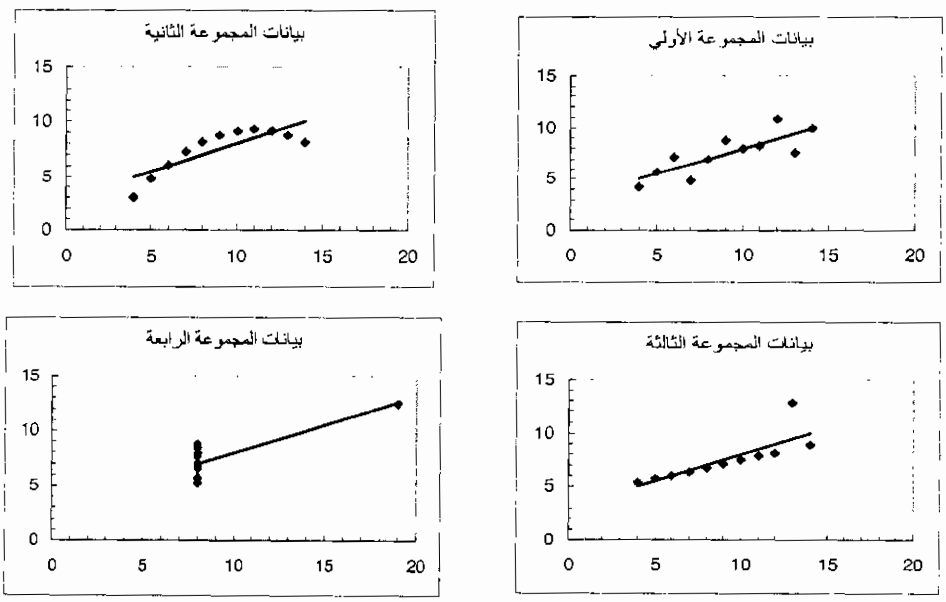
مثال ١٠-١١

يوضح جدول ١٠-٣ أربعة مجموعات من البيانات لمتغيرين X، Y وهذه المجموعات لها نفس الإحصاءات. والشكل ١٠-٨ يوضح الشكل الانتشاري لهذه البيانات مع خط الانحدار لكل مجموعة (المصدر: Chatterjee and Price, 1991).

يظهر تحليل بيانات هذا المثال باستخدام معادلة الانحدار الخطي البسيط أن معادلة الانحدار للمجموعات الأربع من البيانات متطابقة وهي $\hat{Y} = 3 + 0.5X$ ومعامل التحديد قيمته $r^2 = 0.67$ للأربعة مجموعات من البيانات. ولكن الشكل الانتشاري١٠-٨ وخط الانحدار لمجموعات البيانات كل على حده توضح أن التحليلات التي اعتمدت على المتوسط ومعامل الانحدار ومعامل التحديد لم تنجح جميعها في الكشف عن الاختلافات بين أنماط المجموعات الأربع من البيانات وبالتالي يكون التحليل غير صحيح.

جدول ١٠-٣ أربعة مجموعات من البيانات لمتغيرين X، Y لها نفس المتوسط

	مجموعة ١		مجموعة ٢		مجموعة ٣		مجموعة ٤	
	X_1	Y_1	X_2	Y_2	X_3	Y_3	X_4	Y_4
	10	8.04	10	9.14	10	7.46	8	6.58
	8	6.95	8	8.14	8	6.77	8	5.76
	13	7.58	13	8.74	13	12.74	8	7.71
	9	8.81	9	8.77	9	7.11	8	8.84
	11	8.33	11	9.26	11	7.81	8	8.47
	14	9.96	14	8.10	14	8.84	8	7.04
	6	7.24	6	6.13	6	6.08	8	5.25
	4	4.26	4	3.10	4	5.39	19	12.50
	12	10.84	12	9.13	12	8.15	8	5.56
	7	4.82	7	7.26	7	6.42	8	7.91
	5	5.68	5	4.74	5	5.73	8	6.89
المتوسط	**9**	**7.5**	**9**	**7.5**	**9**	**7.5**	**9**	**7.5**



شكل ١٠-٨ شكل انتشارى مع خط الانحدار للأربعة مجموعات من بيانات مثال ١٠-١١.

لاحظ أن المخالفات الصغيرة small violations والخاصة بالتحليل بطريقة المربعات الصغرى لا تؤثر بدرجة كبيرة فى استنتاجات التحليل بينما مخالفات نموذج التحليل المستخدم تؤدى إلى تغيير جذرى فى استنتاجات التحليل.

يعتبر تحليل المتبقى analysis of residual طريقة سهلة للكشف عن مدى ملائمة نموذج تحليل الانحدار. ويوجد طريقتان لتحليل المتبقى: الطريقة الأولى تعتمد على الرسم البيانى حيث يمثل المحور الصادى القيم القياسية للمتبقى والمحور السينى عبارة عن القيم المتوقعة $\hat{Y}$ أو قيم المتغير المستقل X_i، أما الطريقة الثانية فهى تعتمد على تقسيم المتبقى إلى جزئين، الجزء الأول يسمى الخطأ النقى pure error والجزء الثانى يطلق عليه عدم الكفاية lack of fit.

١٠-١١-١ تحليل المتبقى بطريقة الرسم البيانى

من المعادلتين (١٠-٣) و (١٠-٤) وجدول تحليل التباين ١٠-٢ يمكن حساب قيمة الانحراف (المتبقى) عن خط الاعتماد لكل قيمة من المتغير المعتمد Y_i كالتالى:
$e_i = Y_i - \hat{Y}_i$ وهذه القيم يمكن تحويلها إلى قيم قياسية باستخدام $e_{is} = e_i / S_{Y.X}$.

بصفة عامة عندما يتم اختيار النموذج الصحيح فإن القيم القياسية للمتبقى تتراوح بين 2 إلى 2- وتتوزع عشوائياً حول الصفر. والرسم البيانى لابد أن لا يظهر نمط واضح للاختلافات فى هذه القيم.

مثال ١٠-١٢

استخدم بيانات المثال ١٠-١١ لرسم القيم المتبقية القياسية e_{is} على المحور الصادى وكل من القيم المتوقعة $\hat{Y}_i$ وقيم المتغير المستقل X_i للمجموعات الأربع من البيانات. $S_{y.x} = 1.24$ ، $a = 3$ ، $b_{y.x} = 0.5$ للمجموعات الأربع من البيانات.

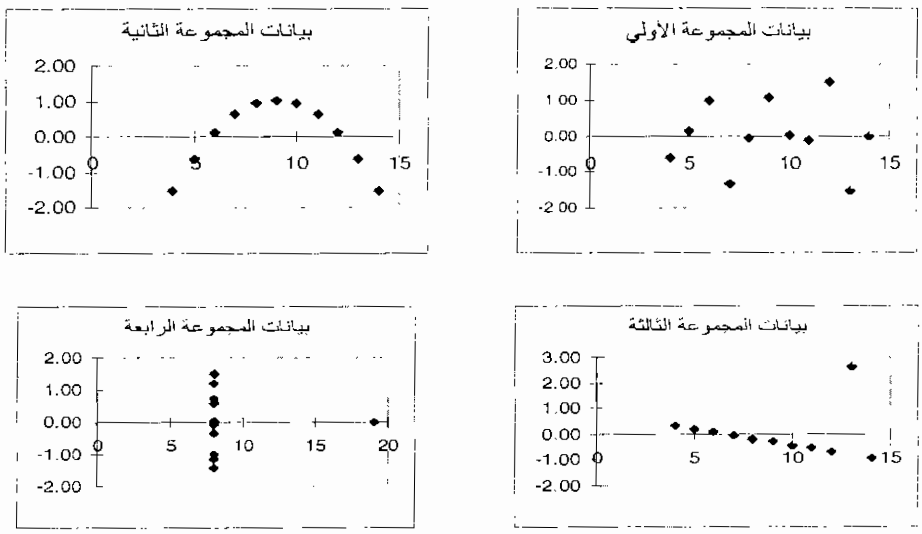
يوضح جدول ١٠-٤ قيم المتغير المستقل و القيم المتوقعة والقيم المتبقية القياسية لمجموعات البيانات الأربع المذكورة فى مثال ١٠-١١، والشكلين ١٠-٩ و ١٠-١٠ يوضحان القيم المتبقية القياسية على المحور الصادى وقيم المتغير المستقل والقيم المتوقعة على المحور السينى للمجموعات الأربع، على الترتيب.

يتضح من الشكلين ١٠-٩ و ١٠-١٠ أن بيانات المجموعة الأولى فقط هى التى يمكن تحليلها بالنموذج $Y = a + bX + e$ وكان توزيع القيم المتبقية القياسية لهذه المجموعة يبدو عشوائياً حول الصفر وتراوحت قيمه بين ± 2 ولم يظهر توزيع هذه القيم أى نمط معين بتغير قيم العامل المستقل أو القيم المتوقعة. بينما الأشكال البيانية الخاصة بمجموعات البيانات الثانية والثالثة والرابعة أظهرت جميعها نمطاً معيناً غير

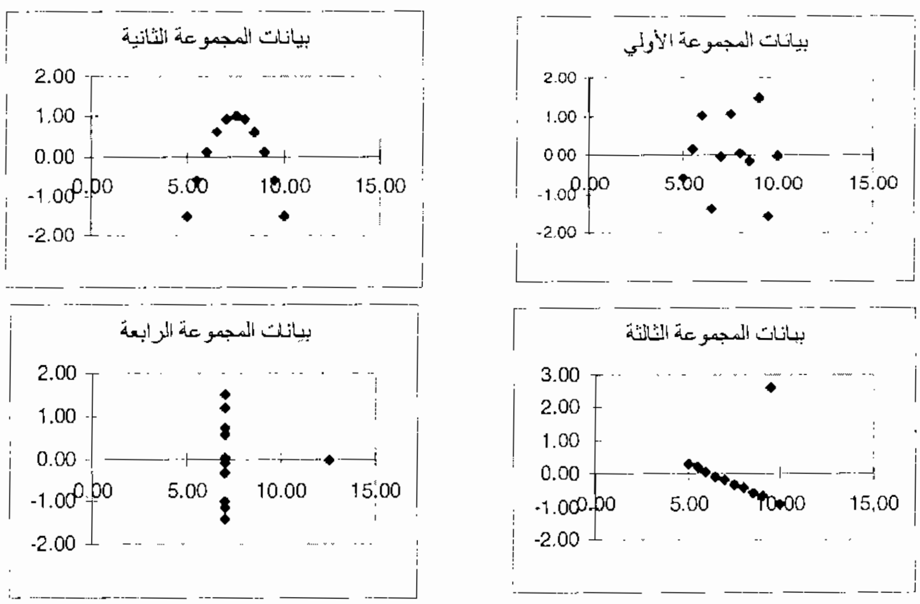
عشوائى لتوزيع القيم المتبقية القياسية بتغير سواء العامل التابع أو القيم المتوقعة مما يدل على عدم صلاحية النموذج $Y = a + bX + e$ فى تحليل هذه المجموعات من البيانات ولابد من البحث عن نموذج آخر غير النموذج الخطى البسيط. ومعنى هذا أن هناك ربما عوامل أخرى تؤثر على المتغير التابع و/أو العلاقة بين X و Y ليست بالخطية ولكنها قد تكون أكث تعقيداً.

جدول ١٠-٤ قيم المتغير المستقل و القيم المتوقعة والقيم المتبقية القياسية للمجموعات الأربع من البيانات المذكورة فى مثال ١٠-١١

مجموعة ٤			مجموعة ٣			مجموعة ٢			مجموعة ١		
e_{4s}	$\hat{Y}_4$	x_4	e_{3s}	$\hat{Y}_3$	x_3	e_{2s}	$\hat{Y}_2$	x_2	e_{1s}	$\hat{Y}_1$	x_1
-0.34	7.0	8	-0.44	8.0	10	0.92	8.0	10	0.03	8.0	10
-1.00	7.0	8	-0.19	7.0	8	0.92	7.0	8	-0.04	7.0	8
0.57	7.0	8	2.62	9.5	13	-0.61	9.5	13	-1.55	9.5	13
1.49	7.0	8	-0.32	7.5	9	1.03	7.5	9	1.06	7.5	9
1.19	7.0	8	-0.56	8.5	11	0.61	8.5	11	-0.14	8.5	11
0.03	7.0	8	-0.94	10.0	14	-1.54	10.0	14	-0.03	10.0	14
-1.41	7.0	8	0.06	6.0	6	0.11	6.0	6	1.00	6.0	6
0.00	12.5	19	0.32	5.0	4	-1.54	5.0	4	-0.60	5.0	4
-1.16	7.0	8	-0.69	9.0	12	0.11	9.0	12	1.49	9.0	12
0.74	7.0	8	-0.06	6.5	7	0.61	6.5	7	-1.36	6.5	7
-0.09	7.0	8	0.19	5.5	5	-.061	5.5	5	0.15	5.5	5



شكل ١٠-٩ القيم المتبقية القياسية على المحور الصادى وقيم المتغير المستقل على المحور السينى للأربعة مجموعات من البيانات.



شكل ١٠-١٠ القيم المتبقية القياسية على المحور الصادى والقيم المتوقعة على المحور السينى للمجموعات الأربع من البيانات.

١٠-١١-٢ طريقة الخطأ النقى وعدم الكفاية Lack of fit and pure error

كما سبق فإن $e_i = Y_i - \hat{Y}_i$ تمثل المتبقى عند X_i وهذه تمثل القيمة التى تختلف بها القيمة الحقيقية Y_i عن القيمة المتوقعة $\hat{Y}_i$. وأيضا سبق إيضاح أن $\sum e_i = 0$. تشتمل قيمة المتبقى هذه على جميع المعلومات المتاحة والتى فشل النموذج المستخدم فى التحليل فى توظيفها لتفسير التباين فى المتغير المعتمد Y. هذا المتبقى يمكن تقسيمه إلى جزئين الأول يعرف بالخطأ النقى pure error، كما سبق، ويرمز له بالرمز S_e^2 والجزء الثانى يعرف بعدم الكفاية lack of fit ويرمز له بالرمز MS_L وهذا الجزء يمثل عدم ملائمة النموذج المستخدم فى تحليل البيانات محل الدراسة. يجب ملاحظة أنه يلزم وجود مشاهدات متكررة لكل من X، Y حتى يمكن تحليل المتبقى. ولحساب ذلك افترض وجود بيانات مكررة يمكن تعريفها كالتالى:

$Y_{11}, Y_{12}, ..., Y_{1n_1}$ عبارة عن عدد n_1 من المشاهدات المكررة عند X_1

$Y_{21}, Y_{22}, ..., Y_{2n_2}$ عبارة عن عدد n_2 من المشاهدات المكررة عند X_2

Y_{ju} عبارة عن المشاهدة u عند X_j حيث $u = 1, 2, \cdots, n_j$

$Y_{m1}, Y_{m2}, ..., Y_{mn_m}$ عبارة عن عدد n_m من المشاهدات المكررة عند X_m

بمعنى أن $n = \sum_{j=1}^{m} \sum_{u=1}^{n_j} 1 = \sum_{j=1}^{m} n_j$ مشاهدة.

مساهمة n_1 من المشاهدات عند X_1 فى مجموع مربعات الخطأ النقى عبارة عن مجموع المربعات الداخلى لقيم Y_{1u} حول متوسطهم $\overline{Y}_1$ بمعنى أن:

$$\sum_{u=1}^{n_1}(Y_{1u} - \overline{Y}_1)^2 = \sum_{u=1}^{n_1} Y_{1u}^2 - n_1\overline{Y}_1^2 = \sum_{u=1}^{n_1} Y_{1u}^2 - \left(\sum_{u=1}^{n_1} Y_{1u}\right)^2 / n_1 \quad (١٠-٤٧)$$

تجميع pooling مجموع المربعات الداخلية الراجعة إلى كل المكررات يؤدى إلى الحصول على مجموع المربعات الراجع إلى الخطأ النقى

$$\sum_{j=1}^{m} \sum_{u=1}^{n_j} (Y_{ju} - \overline{Y}_j)^2 \qquad (١٠-٤٨)$$

ودرجات الحرية عبارة عن:

$$n_e = \sum_{j=1}^{m} (n_j - 1) = \sum_{j=1}^{m} n_j - m \qquad (٤٩-١٠)$$

وبما أن مجموع مربعات الخطأ النقى هو جزء من مجموع المربعات للمتبقى فإنه يمكن التعبير عن المتبقى للمشاهدة u عند X_i كالتالى:

$$Y_{ju} - \hat{Y}_j = (Y_{ju} - \overline{Y}_j) - (\hat{Y}_j - \overline{Y}_j)$$

لاحظ أن جميع القيم المكررة عن أى X_i سوف يكون لهما نفس القيمة المتوقعة $\hat{Y}_j$. بتربيع جانبى المعادلة (١٠-٤٩) والجمع على كل من u و j يمكن الحصول على:

$$\sum_{j=1}^{m}\sum_{u=1}^{n_j} (Y_{ju} - \hat{Y}_j)^2 = \sum_{j=1}^{m}\sum_{u=1}^{n_j} (Y_{ju} - \overline{Y}_j)^2 + \sum_{j=1}^{m}\sum_{u=1}^{n_j} (\hat{Y}_{ju} - \overline{Y}_j)^2 \qquad (٥٠-١٠)$$

لاحظ أن الجانب الأيسر من هذه المعادلة عبارة عن مجموع مربعات المتبقى، أما الجانب الأيمن فيحتوى على مجموع مربعات الخطأ الحقيقى وهو

$$\sum_{j=1}^{m}\sum_{u=1}^{n_j} (Y_{ju} - \overline{Y}_j)^2$$

أما الجزء الثانى فهو مجموع مربعات عدم الكفاية .

مثال ١٠-١٣

الجدول التالى يوضح عدد 24 مشاهدة بعضها مكرر. ومعادلة الانحدار الخطى البسيط لها $\hat{Y} = 1.436 + 0.338X$. يوضح جدول ١٠-٥ تحليل التباين لهذه البيانات، (المصدر: Draper and Smith, 1981).

مشاهدة	X	Y	مشاهدة	X	Y	مشاهدة	X	Y	مشاهدة	X	Y
١	1.3	2.3	٧	3.3	1.8	١٣	4.7	5.4	١٩	5.3	2.1
٢	1.3	1.8	٨	3.7	3.7	١٤	4.7	3.2	٢٠	5.7	3.4
٣	2.0	2.8	٩	3.7	1.7	١٥	4.7	1.9	٢١	6.0	3.2
٤	2.0	1.5	١٠	4.0	2.8	١٦	5.0	1.8	٢٢	6.0	3.0
٥	2.7	2.2	١١	4.0	2.8	١٧	5.3	3.5	٢٣	6.3	3.0
٦	3.3	3.8	١٢	4.0	2.2	١٨	5.3	2.8	٢٤	6.7	5.9

جدول ١٠-٥ تحليل بيانات مثال ١٠-١٣

SOV	df	SS	MS	F
Due to regression	1	6.326	6.926	6.569
From regression	22	21.192	$0.963 = S^2$	

الخطوة التالية هى حساب مجموع مربعات الخطأ النقى وبالطرح من مجموع مربعات الخطأ الراجع للانحدار يمكن الحصول على مجموع مربعات عدم الكفاية.

١ - مجموع مربعات الخطأ النقى لمكررات القيم Y عند $X = 1.3$ عبارة عن $(0.5)(2.3 - 1.8)^2 = 0.125$ وهذه القيمة لها درجة حرية واحدة.

٢ - مجموع مربعات الخطأ النقى لمكررات القيم Y عند $X = 4.7$ عبارة عن

$$(5.4)^2 + (3.2)^2 + (1.9)^2 - (3)[(5.4 + 3.2 + 1.9)/3]^2$$
$$= 43.01 - (10.5)^2/3 = 6.26$$

وهذه القيمة لها 2 درجة حرية.

ويمكن تلخيص حسابات مجموع مربعات الخطأ النقى فى الجدول التالى:

df	$\sum_{u=1}^{n_j}(Y_{ju} - \overline{Y}_j)^2$	مستوى X
1	0.125	1.3
1	0.845	2.0
1	2.000	3.3
1	2.000	3.7
2	0.240	4.0
2	6.260	4.7
2	0.980	5.3
1	0.020	6.0
11	**12.470**	**المجموع**

وبالتالى يمكن إعادة كتابة جدول تحليل التباين ١٠-٥ فى جدول آخر ١٠-٦ والذى يظهر مجموع مربعات عدم الكفاية ومجموع مربعات الخطأ النقى بدرجات حرية = درجات الحرية من خط الاعتماد – درجات حرية عدم الكفاية، أى $11 = 11 - 22$

جدول ١٠-٦ تحليل التباين لبيانات مثال ١٠-١٣ مع إظهار قيمة مجموع مربعات كل من عدم الكفاية والخطأ النقى

SOV	df	SS	MS	F
Due to regression	1	6.326	6.926	6.569
From regression	22	21.192	$0.963 = S^2$	
Lack of fit	11	8.722	$0.793 = MS_L$	0.699
Pure error	11	12.470	$1.134 = S_e^2$	
C. Total	**23**			

لاحظ أن F المحسوبة لعدم الكفاية حسبت من $F = 0.793/1.134 = 0.699$ وهى غير معنوية بمستوى $\alpha = 0.05$ وبالتالى فإن النموذج المستخدم يعتبر كافياً لتفسير التباين فى قيم Y.

مثال ١٠-١٤

يمكن استخدام برنامج SAS لإجراء اختبار عدم الكفاية lack of fit لبيانات مثال ١٠-١٣.

```
DATA LACK;
INPUT X Y @@;
CARDS;
1.3 2.3 1.3 1.8 2 2.8 2 1.5 2.7 2.2 3.3 3.8
3.3 1.8 3.7 3.7 3.7 1.7 4 2.8 4 2.8 4 2.2
4.7 5.4 4.7 3.2 4.7 1.9 5 1.8 5.3 3.5 5.3 2.8
5.3 2.1 5.7 3.4 6 3.2 6 3 6.3 3 6.7 5.9
PROC SORT;
BY Y;
PROC RSREG;
MODEL Y = X / COVAR = 1 LACKFIT;
RUN;
```

لاحظ:

لابد من أن يتم ترتيب البيانات على المتغير التابع ولذلك استخدم اختيار PROC SORT باستخدام Y كمتغير للترتيب.

استخدم اختيار PROC RSREG (quadratic response surface) بدلا من PROC REG.

استخدم اختيار COVR = 1 مع النموذج للإشارة إلى أن هناك متغيراً مستقلاً واحداً فقط.

استخدم اختيار lackfit مع النموذج للحصول على مجموع المربعات الراجع لكل من الخطأ النقى وعدم الكفاية.

النتائج:

The RSREG Procedure

Response Surface for Variable Y

Response Mean	2.858333
Root MSE	0.981503
R-Square	0.2298
Coefficient of Variation	34.3383

Regression	DF	Type I Sum of Squares	R-Square	F Value	Pr > F
Covariates	1	6.324667	0.2298	6.57	0.0178
Linear	0	0	0.0000	.	.
Quadratic	0	0	0.0000	.	.
Crossproduct	0	0	0.0000	.	.
Total Model	1	6.324667	0.2298	6.57	0.0178

Residual	DF	Sum of Squares	Mean Square	F Value	Pr > F
Lack of Fit	11	8.723666	0.793061	0.70	0.7183
Pure Error	11	12.470000	1.133636		
Total Error	22	21.193666	0.963348		

صندوق ١٠-٣

- تسمى نسبة التباين الراجع للاعتماد إلى التباين الكلى ($b\sum xy/\sum y^2$) بمعامل التحديد r^2.
- لا يجب أخذ ارتفاع r^2 ومعنويتها دليلا على أن النموذج المفترض قد استخلص كل المعلومات فى Y، ففى كثير من هذه الحالات تكون مازالت معلومات فى المتبقى لم يستخلصها النموذج.
- فى الحالات التى يكون فيها أكثر من قيمة للمتغير المعتمد Y لنفس قيمة المتغير المستقل X، فإنه يمكن تقسيم مجموع مربعات المتبقى إلى جزئين:

  جزء راجع لكفاية النموذج الإحصائى المستخدم والآخر خطأ نقى، وبهذا يمكن أخذ فكرة عن مدى ملائمة النموذج الإحصائى وعن القدر الذى استخلص به المعلومات فى Y.

١٠-١٢ الارتباط البسيط Simple correlation

(معامل الارتباط لبيرسون Pearson correlation coefficient)

يقيس معامل الارتباط شدة العلاقة بين متغيرين أو هو مقياس لدرجة تغير متغيرين معاً وهو تقدير غير متحيز لمعامل ارتباط المتغيرين فى العشيرة ويحسب من المعادلة التالية:

$$r_{Y_1Y_2} = \frac{\sum(Y_1-\overline{Y}_1)(Y_2-\overline{Y}_2)}{\sqrt{\sum(Y_1-\overline{Y}_1)^2}\sqrt{(Y_2-\overline{Y})^2}} = \frac{\sum y_1y_2}{\sqrt{\sum y_1^2}\sqrt{\sum y_2^2}} \qquad (١٠-٥١)$$

وبقسمة البسط والمقام على درجات الحرية $n-1$ فإن:

$$r_{Y_1Y_2} = \frac{(\sum y_1y_2)/(n-1)}{\sqrt{(\sum y_1^2)/(n-1)}\sqrt{(\sum y_2^2)/(n-1)}}$$

$$= \frac{\text{cov}(Y_1, Y_2)}{(S_{Y_1})(S_{Y_2})} \qquad (٥٢-١٠)$$

حيث S_{Y_1} هى الانحراف المعيارى للمتغير Y_1، S_{Y_2} هى الانحراف المعيارى للمتغير Y_2.

وفى مثال ١٠-١٠ فإن معامل الارتباط $r_{Y_1Y_2} = \frac{130}{\sqrt{168}\sqrt{135.5}} = 0.86$

ومعامل الارتباط هو الجذر التربيعى لحاصل ضرب معاملى الانحدار، أى أنه عبارة عن المتوسط الهندسى لهما أى:

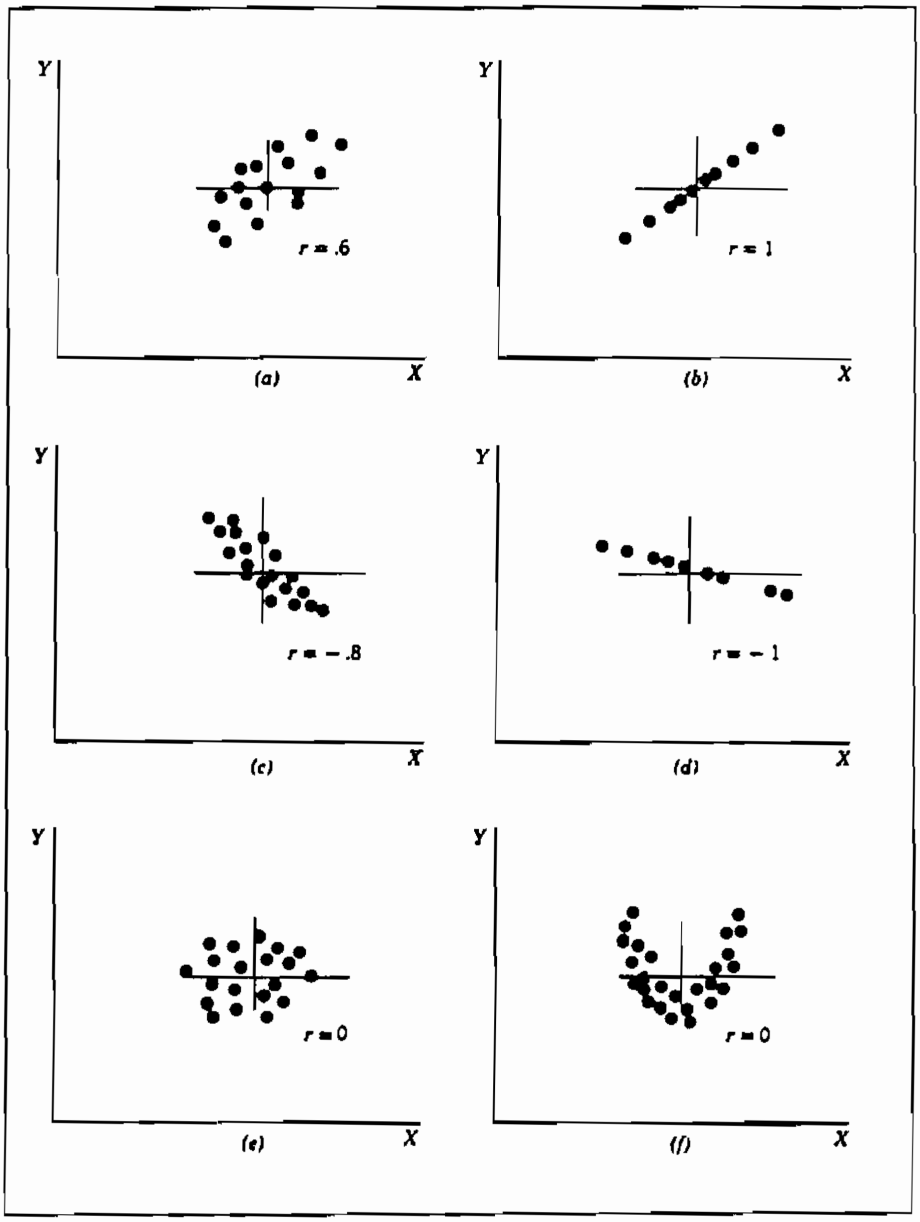
$$r = \sqrt{(b_{Y_1Y_2})(b_{Y_2Y_1})} \qquad (٥٣-١٠)$$

وفى مثال ١٠-١٠ فإن معامل الارتباط $r = \sqrt{(0.959)(0.774)} = 0.86$

وإذا كان معامل التحديد، وهو مربع معامل الارتباط، تنحصر قيمته بين الصفر والواحد الصحيح فإن معامل الارتباط تنحصر قيمته بين 1+ ، 1- أى أن:

$$(-1 \le r \le +1) \qquad (٥٤-١٠)$$

ويكون معامل الارتباط موجباً عندما تكون القيم الكبرى للمتغير الأول تقابلها القيم الكبرى للمتغير الثانى، أى أن تغير المتغيرين يكون فى نفس الاتجاه. ويكون معامل الارتباط سالباً عندما تكون القيم الكبرى لأحد المتغيرين تقابلها القيم الصغرى للمتغير الآخر. ويمكن من الأشكال الانتشارية (١٠-١١ a، b، c، d، e، f) الاستدلال على الارتباط. الشكل (b) يمثل علاقة خطية كاملة موجبة $r = 1$، والشكل (d) يمثل علاقة خطية كاملة سالبة $r = -1$. الشكل (a) يمثل علاقة غير كاملة موجبة $r = 0.6$ والشكل (c) يمثل علاقة غير كاملة سالبة $r = -0.8$. عند مقارنة الشكلين (e) و (f) يلاحظ أن الارتباط قيمته صفر فى الحالتين ولكن الشكل (f) يظهر بوضوح وجود علاقة قوية بين X و Y رغم أن $r = 0$ ومعنى ذلك أن معامل الارتباط فى هذه الحالة لا يعنى عدم وجود علاقة بين المتغيرين ولكنه يعنى عدم وجود علاقة خطية.



شكل ١٠-١١ حالات مختلفة تمثل العلاقة بين متغيرين X و Y

وعلى عكس معامل الاعتماد الذى هو قيمة مميزة فإن معامل الارتباط كمية مجردة أو مطلقة، أى مستقلة عن وحدات القياس. كما يمكن حسابه بوحدات مختزلة

لكل من المتغيرين ولا يلزم لها التعديل بعد ذلك. والملاحظ من دراسات البيانات البيولوجية أن معامل الارتباط نادراً ما يكون أعلى من 0.9.

مثال ١٠-١٥

احسب معامل انحدار Y_2 على Y_1 وكذا معامل انحدار Y_1 على Y_2 ومعامل الارتباط بينهما للبيانات التالية، ثم مثل العلاقة بين المتغيرين بيانياً.

$:Y_1$	1	4	8	0	2
$:Y_2$	2	5	9	1	3

من البيانات:

$$n = 5 \qquad \sum Y_1 = 15 \qquad \sum Y_1^2 = 85$$

$$\sum Y_1 Y_2 = 100 \qquad \sum Y_2 = 20 \qquad \sum Y_2^2 = 120$$

$$b_{Y_1Y_2} = \frac{\sum Y_1Y_2 - \frac{\sum Y_1 \sum Y_2}{n}}{\sum Y_1^2 - \frac{(\sum Y_1)^2}{n}} = \frac{100 - \frac{(15)(20)}{5}}{120 - \frac{(20)^2}{5}} = 1$$

$$b_{Y_2Y_1} = \frac{\sum Y_1Y_2 - \frac{\sum Y_1 \sum Y_2}{n}}{\sum Y_2^2 - \frac{(\sum Y_2)^2}{n}} = \frac{100 - \frac{(15)(20)}{5}}{85 - \frac{(15)^2}{5}} = 1$$

معامل انحدار Y_1 على Y_2 عبارة عن واحد صحيح، أى وحدة من Y_1 لكل وحدة من Y_2، وكذلك معامل انحدار Y_2 على Y_1.

معادلتى خطى الانحدار هما:

$$\hat{Y}_1 = -1 + Y_2$$

$$\hat{Y}_2 = 1 + Y_1$$

ومعامل الارتباط هو:

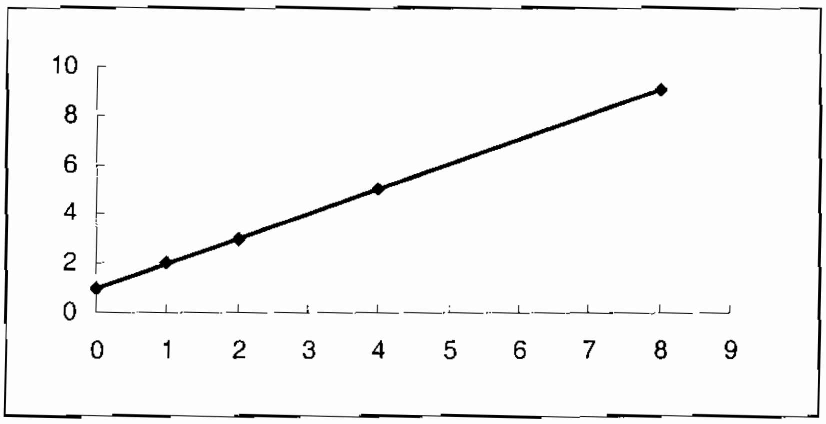
$$r_{Y_2Y_1} = \frac{\sum Y_1Y_2 - \frac{\sum Y_1 \sum Y_2}{n}}{\sqrt{\sum Y_1^2 - \frac{(\sum Y_1)^2}{n}}\sqrt{\sum Y_2^2 - \frac{(\sum Y_2)^2}{n}}}$$

$$= \frac{100 - \frac{(15)(20)}{5}}{\sqrt{85 - \frac{(15)^2}{5}}\sqrt{120 - \frac{(20)^2}{5}}} = 1$$

وكما سبق يمكن استخدام معاملى الانحدار فى حساب معامل الارتباط كالتالى:

$$r = \sqrt{(b_{Y_1Y_2})(b_{Y_2Y_1})} = \sqrt{(1)(1) = 1}$$

والشكل ١٠-١٢ يمثل العلاقة بين المتغيرين بيانياً.



شكل ١٠-١٢ التمثيل البيانى لخطى الانحدار فى مثال ١٠-١٥

لاحظ أن جميع النقط تقع على خط واحد، خط انحدار Y_2 على Y_1 هو نفسه خط الانحدار Y_1 على Y_2 وهذا لا يحدث إلا عندما يكون الارتباط تاماً بين المتغيرين وهى الحالة التى فيها تكون القيم المتوقعة predicted values هى نفسها القيم الفعلية actual values حيث $S^2_{Y_1Y_2} = 0$.

مثال ١٠-١٦

استخدام برنامج SAS فى حساب معامل انحدار Y_2 على Y_1 وكذا معامل انحدار Y_1 على Y_2 ومعامل الارتباط بينهما لبيانات مثال ١٠-١٥.

```
DATA RC;
INPUT Y1 Y2 @@;
CARDS;
1 2 4 5 8 9 0 1 2 3
PROC REG;
MODEL Y1 = Y2;
MODEL Y2 = Y1;
PROC CORR;
RUN;
```

لاحظ:

كل معامل انحدار يراد حسابه يوضع فى نموذج model منفصل .
استخدام PROC CORR لحساب معامل الارتباط.

النتائج

The REG Procedure
Model: MODEL1

Dependent Variable: Y1
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	40.00000	40.00000	Infty	<.0001
Error	3	0	0		
Corrected Total	4	40.00000			

Root MSE	0	R-Square	1.0000		
Dependent Mean	3.00000	Adj R-Sq	1.0000	Coeff Var	0

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-1.00000	0	-Infty	<.0001
Y2	1	1.00000	0	Infty	<.0001

Dependent Variable: Y2

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	40.00000	40.00000	Infty	<.0001
Error	3	0	0		
Corrected Total	4	40.00000			

Root MSE	0	R-Square	1.0000
Dependent Mean	4.00000	Adj R-Sq	1.0000
Coeff Var	0		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.00000	0	Infty	<.0001
Y1	1	1.00000	0	Infty	<.0001

The CORR Procedure

2 Variables: Y1 Y2

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Y1	5	3.00	3.16228	15.00	0	8.00
Y2	5	4.00	3.16228	20.00	1.00	9.00

Pearson Correlation Coefficients, N = 5
Prob > |r| under H0: Rho=0

	Y1	Y2
Y1	1.00000	1.00000
Y2	1.00000	1.00000

١٠-١٣ العلاقة بين معامل الانحدار ومعامل الارتباط

١- حيث إن معامل انحدار Y_2 على Y_1 هو

$$b_{Y_2Y_1} = \frac{\operatorname{cov}(Y_1, Y_2)}{S_{Y_1}^2}$$

فإنه بضرب كل من البسط والمقام فى S_{Y_2}

$$b_{Y_2Y_1} = \left(\frac{\operatorname{cov}(Y_1, Y_2)}{S_{Y_1}^2}\right) \cdot \left(\frac{S_{Y_2}}{S_{Y_2}}\right) = \left(\frac{\operatorname{cov}(Y_1, Y_2)}{(S_{Y_1})(S_{Y_2})}\right)\left(\frac{S_{Y_2}}{S_{Y_1}}\right)$$

$$\therefore b_{Y_2Y_1} = (r)\left(\frac{S_{Y_2}}{S_{Y_1}}\right) \qquad (١٠-٥٥)$$

وحيث إن $S_{y_1} = \sqrt{(\sum y_1^2)/(n-1)}$ ، $S_{y_2} = \sqrt{(\sum y_2^2)/(n-1)}$

$$\therefore b_{Y_2Y_1} = r\sqrt{(\sum y_2^2)/(\sum y_1^2)} \qquad (١٠-٥٦)$$

أى أن:

$$r = b_{Y_2Y_1}\sqrt{(\sum y_1^2)/(\sum y_2^2)} \qquad (١٠-٥٧)$$

وبالمثل فإن:

$$r = b_{Y_1Y_2}\sqrt{(\sum y_2^2)/(\sum y_1^2)} \qquad (١٠-٥٨)$$

٢- حيث إن معادلة خط انحدار Y_2 على Y_1 هى $\hat{Y}_2 = \overline{Y}_2 + b(Y_1 - \overline{Y}_1)$
وبالتعبير عن Y_1 ، Y_2 بوحداتهم القياسية standard units أى:

$$Y_2^{\circ} = \frac{Y_2 - \overline{Y}_2}{S_{Y_2}} \quad , \quad Y_1^{\circ} = \frac{Y_1 - \overline{Y}_1}{S_{Y_1}}$$

وبالتعويض فى معادلة خط الانحدار عن Y_1 ، Y_2 بوحداتهم القياسية فإن

$$\hat{Y}_2^{\circ} S_{y_2} = b_{Y_2 Y_1} Y_1^{\circ} S_{y_1}$$

$$\hat{Y}_2^{\circ} = b_{Y_2 Y_1} \frac{S_{Y_1}}{S_{Y_2}} Y_1^{\circ}$$

$$\hat{Y}_2^{\circ} = rY_1^{\circ} \qquad (١٠-٥٩)$$

حيث $\hat{Y}_2^{\circ}$ تمثل القيمة المتوقعة معبراً عنها بوحدات قياسية. وباستخدام الوحدات القياسية يصبح معامل الارتباط r هو نفسه معامل الانحدار والاختلافات بينهما تتلاشى.

٣- سبق بيان أن معامل الارتباط هو المتوسط الهندسى لمعاملى الانحدار، أى:

$$r = \pm\sqrt{(b_{Y_1 Y_2})(b_{Y_2 Y_1})}$$

وخطى الانحدار يتطابقان عندما تكون $r = \pm 1$، ويكونان قريبين جداً من بعضهما عندما يكون معامل الارتباط قريب من ± 1.

١٠-١٤ اختبار معنوية معامل الارتباط وتقدير حدود الثقة له

إذا سحب عدد كبير من العينات العشوائية حجم كل منها n من عشيرة تتبع التوزيع الطبيعى لمتغيرين معامل الارتباط بينهما $\rho = 0$ وحسب معامل الارتباط لكل عينة، فإن توزيع معاملات ارتباط العينات يقترب من التوزيع الطبيعى ويزداد اقتراباً منه بزيادة حجم العينة. وبالتالى فإن التوزيع العينى لمعاملات الارتباط يكون متماثلاً أيضاً حتى عندما يكون توزيع أحد المتغيرين غير طبيعى كأن يكون قيماً ثابتة fixed مثلاً يختارها الباحث على أن يكون توزيع المتغير الآخر طبيعياً. أما إذا كان معامل ارتباط العشيرة $\rho \neq 0$ فإن التوزيع العينى لمعاملات الارتباط لا يتبع التوزيع الطبيعى وإنما يكون التوزيع ملتويا skewed مهما كان حجم العينات. ويمثل شكل ١٠-١٣ التوزيع العينى لمعاملات الارتباط المسحوبة من عشيرة معامل الارتباط يساوى صفراً وكذلك التوزيع العينى لمعاملات ارتباط مسحوبة من عشيرة معامل الارتباط بها لا يساوى الصفر.

صندوق ١٠-٤

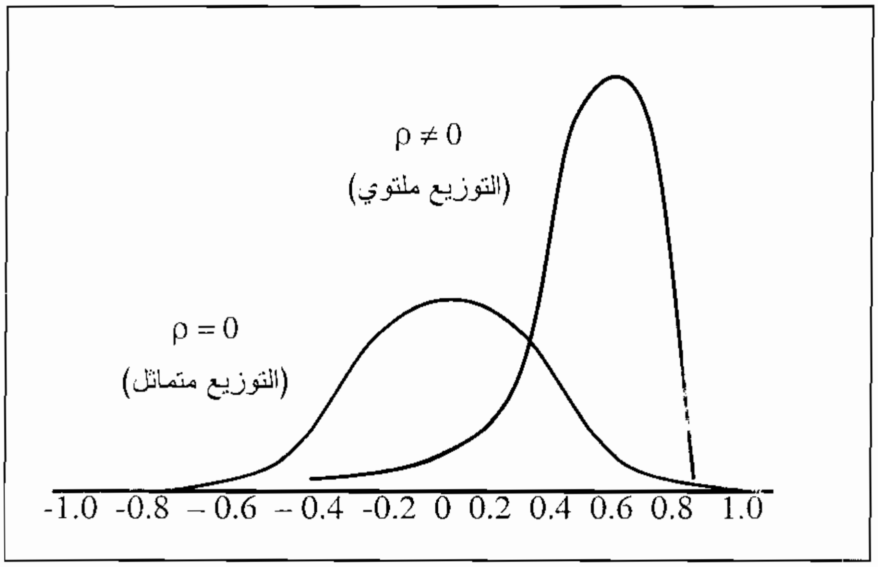
- معامل الارتباط (r) بين متغيرين هو مقياس لشدة العلاقة بينهما. تتراوح قيمته بين 1+، 1- وكلما بعدت قيمته عن الصفر ازدادت شدة العلاقة بين المتغيرين.
- معامل الارتباط كمية مطلقة لا تميز.
- العلاقة الرياضية بين معاملى الاعتماد والارتباط وثيقة جدا، ويمكن تعريف الأخير بأنه الجذر التربيعى لحاصل ضرب معامل اعتماد X على Y ومعامل اعتماد Y على X . ويعرف معامل الارتباط كذلك بأنه معامل الانحدار القياسى أو المعيارى standard (أى بين المتغيرين بعد تعبيرهما $(Y-\overline{Y})/\sigma_Y$) لأحد المتغيرين على الآخر

$$\begin{aligned} r &= b_{Y_1Y_2} / \sqrt{\sum y_2^2 / \sum y_1^2} \\ &= b_{Y_2Y_1} / \sqrt{\sum y_1^2 / \sum y_2^2} \\ &= \frac{\sum y_1 y_2}{\sqrt{(\sum y_1^2)(\sum y_2^2)}} \end{aligned}$$

- يسمى مربع معامل الارتباط (r^2) بمعامل التحديد أى النسبة من تباين المتغير المرتبط بالمتغير الآخر أو الممكن تفسيرها من قبل المتغير الآخر.
- تختبر معنوية r فى حالة فرض العدم $(\rho = 0)$ عن طريق اختبار t أو مباشرة عن طريق جداول خاصة.
- فى حالة ما إذا كان فرض العدم $(\rho \neq 0)$ فإن هذه الحالة تحتاج إلى تحوير خاص.

١٠-١٤-١ اختبار معنوية معامل الارتباط عندما يكون معامل الارتباط فى العشيرة يساوى صفراً $(\rho = 0)$

يستخدم فى مثل هذه الحالة، والتى يكون فيها توزيع r متماثلا حول المنتصف (كما فى شكل ١٠-١٣) طريقتان لاختبار معنوية العلاقة بين متغيرين. الأولى اختبار معنوية الانحدار باستخدام اختبار t ودرجات حرية $n-2$ ومستوى معنوية وليكن α. والثانية اختبار معنوية معامل الارتباط باستخدام الجداول الخاصة بذلك. وكل من الاختبارين يعطى نفس النتيجة ويمكن استخدام أى منهما.



شكل ١٠-١٣ التوزيع العينى لمعاملات الارتباط

أولا: استخدام اختبار t:

$$H_o : \rho = 0$$

$$t = \frac{b}{S_b}$$

بدرجات حرية $n-2$ ومن العلاقة (١٠-٥٦)

$$b = r\frac{S_{Y_2}}{S_{Y_1}}$$

$$t = r\frac{S_{Y_2}}{S_{Y_1}} / S_b = r\frac{S_{Y_2}}{S_{Y_1}} / \sqrt{\left[\frac{\sum y_2^2 - r^2 \sum y_2^2}{(n-2)\sum y_1^2}\right]}$$

$$= \frac{rS_{Y_2}}{S_{Y_1}}\sqrt{\frac{(n-2)\sum y_1^2}{\sum y_2^2 (1-r^2)}}$$

وحيث إن:

$$\frac{S_{y_2}}{S_{y_1}} = \sqrt{\frac{\sum y_2^2}{\sum y_1^2}}$$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \qquad (٦٠-١٠)$$

حيث يعبر عن $\frac{(1-r^2)}{(n-2)}$ بتباين معامل الارتباط.

مثال ١٠-١٧

فى مثال ١٠-١٠ اختبر معنوية معامل الارتباط إذا كان معامل ارتباط العشيرة يساوى صفراً.

حيث $n = 8$ ، $r = 0.86$ ، درجات الحرية $n-2 = 8-2 = 6$ وحيث إن:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{(0.86)(\sqrt{6})}{\sqrt{1-(0.86)^2}} = 8.09$$

وبما إن $t_{(6,0.05)} = 2.447$ ، $t_{(6,0.01)} = 3.707$ وهذه القيم أعلى من قيمة t المحسوبة (8.09) وبالتالى يرفض فرض العدم وهذا يعنى أن العلاقة بين المتغيرين معنوية بدرجة ثقة 99%.

ثانيا: وضعت جداول بمستوى معنوية 5% وأيضاً بمستوى معنوية 1% لاختبار معنوية معامل الارتباط مباشرة وذلك بدرجات حرية $n-2$ (جدول ٧ ملحق أ)، وتتوقف نتيجة الاختبار على حجم العينة وعلى قيمة r فمثلاً:

العينة الأولى: $r_1 = 0.3$ $n_1 = 10$ درجات الحرية = 8

العينة الثانية: $r_2 = 0.3$ $n_1 = 100$ درجات الحرية = 98

العينة الثالثة: $r_3 = 0.7$ $n_3 = 10$ درجات الحرية = 8

على ذلك عند درجات حرية 8 كان معامل الارتباط والذى قيمته 0.3 غير معنوى بينما معامل الارتباط وقيمته 0.7 فهو معنوى بمستوى معنوية 5% عند نفس درجات الحرية. بينما معامل الارتباط (0.3) والذى كان غير معنوى عند درجات حرية 8 يكون معنوياً بمستوى معنوية 1% عند درجات حرية 98، ويجب ملاحظة أنه بزيادة درجات الحرية فإن الحد الأدنى لمعامل الارتباط المعنوى يقل.

والعلاقة المعنوية بين المتغيرين لا تعنى بالضرورة أن أحد المتغيرين هو سبب التغير فى المتغير الآخر (أى علاقة سببية) فقد يكون السبب راجع إلى مصادر أخرى. وإذا كانت $r = 0.3^{**}$ وبالرغم من أن معامل الارتباط معنوى جداً – أى رفض فرض العدم القائل بأنه ليس هناك علاقة بين المتغيرين وقبول الفرض البديل بوجود علاقة بينهما – إلا أن هذا يعنى أن $(0.3)^2 = 0.09$ من تباين Y_2 يرجع للعلاقة بين المتغيرين Y_1 ، Y_2 وأن $(1 - 0.09) = 0.91$ من التباين خالى من تأثير تلك العلاقة ويرجع إلى مصادر أخرى ذكرت فيما سبق عند الحديث عن عنصر الخطأ فى فصل (١٠-١). وتبين الاختبارات الإحصائية أن هناك علاقة خطية بميل لا يساوى الصفر ومعنوى. ولكن هذا لا يعنى أن معظم الاختلافات فى Y_2 هى نتيجة انحدارها على Y_1 بل قد يكون هناك مصادر أخرى للتباين.

وجدير بالملاحظة أنه إذا كان معامل انحدار b_{YX} معنوياً عند مستوى معنوية معين كان أيضاً كل من b_{XY} ، r معنوياً عند نفس المستوى.

١٠-١٤-٢ اختبار معنوية معامل الارتباط عندما يكون معامل ارتباط العشيرة لا يساوى صفراً $(\rho \neq 0)$

لإجراء مثل هذا الاختبار فإنه يفترض أن يكون توزيع كل من المتغيرين طبيعياً، أى يشترط أن تكون أزواج المتغيرات فى العينة تمثل عينة عشوائية مأخوذة من عشيرة ذات متغيرين وهى السابق الإشارة إليها. وإذا كان ρ لا يساوى صفرا فإن التوزيع العينى لمعاملات الارتباط يكون ملتويا ولا يجوز استخدام t والتى تستخدم فقط عندما يكون التوزيع متماثلاً أى عندما تكون $\rho = 0$. ولقد وجد Fisher أنه بتحويل r إلى قيمة أخرى هى Z تصبح الأخيرة ذات توزيع طبيعى. وتحسب Z كما يلى:

$$Z_r = \frac{1}{2}[\log_e(1+r) - \log_e(1-r)] \qquad (٦١-١٠)$$

حيث $\log_e$ هو اللوغاريتم الطبيعى أى ln ، وتتوزع Z طبيعياً بمتوسط Z_ρ حيث $Z\rho = \frac{1}{2}\log_e\frac{1+\rho}{1-\rho}$ وانحراف معيارى $\sigma_Z = \frac{1}{\sqrt{n-3}}$.

ووضعت جداول لتحويل r إلى ما يقابلها من قيم Z (جدول ٨ ملحق أ) وأخرى للتعبير عن قيم Z بما يقابلها من قيم Z (جدول ٩ ملحق أ) ويمكن تغير أى منهما للأخرى بدرجة كافية من الدقة.

وعلى ذلك فالكمية $\frac{Z_r - Z_\rho}{\sqrt{n-3}}$ تتوزع طبيعيا أيضاً. وتستخدم جداول التوزيع الطبيعى أو جدول t عند درجات حرية ∞ ويفضل ألا يقل حجم العينة عن 50 حتى يكون معامل الارتباط أكثر وثوقا reliable.

مثال ١٠–١٨

أخذت عينة عشوائية من عشيرة ذات متغيرين وكان معامل الارتباط 0.5، اختبر ما إذا كانت هذه العينة مسحوبة من عشيرة معامل الارتباط بها 0.6 علماً بأن حجم العينة كان 50 فرداً بدرجة ثقة 95% .

$H\ :\rho = 0.6$

إذا كانت $r = 0.5$ ، فإن قيمة Z المقابلة لها هى $Z = 0.549$

وعند قيمة $\rho = 0.6$ قيمة Z المقابلة لها هى $Z\rho = 0.693$

وحيث إن $\sigma_Z = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{50-3}} = 0.146$

وبالتالى

$$\frac{|Z_r - Z_\rho|}{\sigma_Z} = \frac{|0.549 - 0.693|}{0.146} = 0.99$$

وهذه القيمة (0.99) تقارن بقيمة $t_{(\infty,0.05)} = 1.96$. وعلى ذلك لا يرفض فرض العدم بأن معامل ارتباط العينة 0.5 لا يختلف عن معامل ارتباط العشيرة 0.6 أى أن هذه العينة مسحوبة من عشيرة معامل الارتباط بها 0.6.

١٠-١٤-٣ تقدير حدود الثقة لمعامل ارتباط العشيرة باستخدام بيانات العينة

حيث إن التوزيع العينى لمعاملات الارتباط يصبح ملتوياً إذا كان معامل الارتباط فى العشيرة لا يساوى الصفر وحيث إن Z تتوزع طبيعياً بغض النظر عن حجم العينة بمتوسط Z_ρ وتباين $\frac{1}{n-3}$ فإن حدود الثقة لـــ Z_ρ هى:

$$\Pr(Z_r - Z_{\alpha/2}\sigma_Z \le Z_\rho \le Z_r + Z_{\alpha/2}\sigma_Z) = 1-\alpha \qquad (١٠-٦٢)$$

يمكن أن يعبر عن قيمة $Z_{\alpha/2}$ بأنها قيمة t عند مستوى معنوية α ودرجات حرية ∞.

وباستخدام (جدول ٩ ملحق أ) يمكن إيجاد قيم r المقابلة والتى تمثل حدود الثقة لمعامل ارتباط العشيرة.

مثال ١٠-١٩

ما هى حدود الثقة لمعامل ارتباط العشيرة إذا كان معامل ارتباط العينة 0.425 وحجم العينة 35 وذلك بدرجة ثقة 99% ؟

بما أن $r = 0.425$ وباستخدام (جدول٩ ملحق أ) فإن $Z_r = 0.454$

$$\sigma_Z = \frac{1}{\sqrt{35-3}} = \frac{1}{32} = 0.177 \text{ ،} \quad t_{(\infty,0.01)} = 2.576$$

$\therefore$ حدى الثقة لـــ Z_ρ هما:

$$0.454 \pm (2.576)(0.177) = 0.454 \pm 0.455$$

أى هما 0.001- يمثل الحد الأدنى و 0.909+ يمثل الحد الأعلى. وبتحويل Z إلى قيم r المقابلة يكون الحد الأدنى 0.001- والحد الأعلى 0.721 بدرجة ثقة 99% .

لاحظ أن حدى الثقة ليسا على نفس المسافة على كلا الجانبين لمعامل ارتباط العينة حيث التوزيع غير متماثل كما أنه إذا زاد حجم العينة فإن فترة الثقة تنكمش والعكس صحيح.

١٠-١٥ اختبار تساوى معاملى ارتباط لكل من عينتين مأخوذتين من نفس العشيرة معامل الارتباط بها لا يساوى صفراً

أى اختبار فرض العدم بأن الفرق بين معاملى الارتباط يساوى الصفر. ولكى يتم ذلك تحول كل من قيمتى معامل الارتباط إلى قيم Z المقابلة ثم تختبر معنوية الفرق بين قيمتى Z وما ينطبق على Z فهو ينطبق على r.

أى تحول r_1 إلى Z_1 حيث تباين Z_1 هو $\sigma^2_{Z_1} = \frac{1}{n_1 - 3}$

و تحول r_2 إلى Z_2 وتباين Z_2 هو $\sigma^2_{Z_2} = \frac{1}{n_2 - 3}$

وعلى ذلك فإن $\sigma^2_{Z_1 - Z_2} = \sigma^2_{Z_1} + \sigma^2_{Z_2}$ والانحراف المعيارى للفرق بين Z_1 و Z_2 هو

$$S_{Z_1 - Z_2} = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}} \qquad (١٠-٦٣)$$

وحيث إن $\frac{Z_1 - Z_2}{S_{Z_1 - Z_2}}$ تتبع التوزيع الطبيعى فيمكن مقارنتها بقيمة t عند مستوى معنوية α ودرجات حرية ∞.

مثال ١٠-٢٠

فى تجربة كانت البيانات كالتالى:

العينة الأولى: $n_1 = 15$ ، $r_1 = 0.32$

العينة الثانية: $n_2 = 35$ ، $r_2 = 0.65$

اختبر ما إذا كان معاملا الارتباط للعينتين مسحوبين من نفس العشيرة التى معامل الارتباط بها لا يساوى صفراً.

بما أن $r_1 = 0.32$ فإن $Z_1 = 0.332$ ،

$r_2 = 0.65$ فإن $Z_1 = 0.775$

والتباين لكل من Z_1 و Z_2 عبارة عن $\sigma^2_{Z_1} = \frac{1}{12}$ ، $\sigma^2_{Z_2} = \frac{1}{32}$

والانحراف المعيارى للفرق بينهما من العلاقة (١٠-٦٣)
أى أن:

$$S_{Z_1 - Z_2} = \sqrt{\frac{1}{12} + \frac{1}{32}} = 0.338$$

وبالتالى فإن

$$t = \frac{|0.332 - 0.775|}{0.338} = 1.31$$

وتقارن قيمة t المحسوبة بالقيمة $t_{(\infty, 0.05)} = 1.96$ وبالتالى فإن Z_1 و Z_2 من نفس العشيرة وبالتالى فإن معاملى الارتباط هما لعينتين من نفس العشيرة.

١٠-١٦ اختبار تجانس عدد من معاملات الارتباط وتجميعهم فى قيمة واحدة كتقدير أكثر صلاحية لمعامل ارتباط العشيرة

إذا كانت العينات مأخوذة من نفس العشيرة تحول قيم معاملات الارتباط للعينات المختلفة إلى قيم Z المناظرة لها ثم يختبر الفرض بأن معاملات ارتباط العينات هى تقدير لنفس معامل ارتباط العشيرة باستخدام المعادلة التالية:

$$\sum W_i (Z_i - \overline{Z}_W)^2 \qquad (١٠-٦٤)$$

والتى تساوى

$$\sum W_i Z_i^2 - (\sum W_i Z_i)^2 / \sum W_i \qquad (١٠-٦٥)$$

والتى تتوزع حسب توزيع مربع كاى χ^2 بدرجات حرية = (عدد معاملات ارتباط العينات - 1) حيث

$$W_i = 1/\sigma^2_{Z_i} = (n_i - 3) \qquad (١٠-٦٦)$$

أى أن W_i هى مقلوب تباين Z_i وتستخدم كنوع من عوامل الوزن weight حيث يعطى وزن أكبر للعينات كبيرة الحجم ووزن أقل للعينات الأقل حجما. وإذا كانت قيمة مربع كاى غير معنوية فهذا يعنى أن قيم Z متجانسة وبالتالى فإن قيم r للعينات تكون مأخوذة من نفس العشيرة ويحسب المتوسط $\overline{Z}_W$ حيث إنه يساوى مجموع قيم Z الموزونة بمقلوب تبايناتها على مجموع الأوزان أى أن:

$$\overline{Z}_W = \frac{\sum W_i Z_i}{\sum W_i} \qquad (١٠-٦٧)$$

ثم تستخرج قيمة r المقابلة لقيمة $\overline{Z}_W$ وتكون كتقدير لمعامل ارتباط العشيرة وتباين هذا المتوسط هو:

$$\sigma^2_{\overline{Z}_W} = \frac{1}{\sum(n_i - 3)} \qquad (١٠-٦٨)$$

وبذا يكون انحرافه المعيارى

$$\sigma_{\overline{Z}_W} = \sqrt{\frac{1}{\sum(n_i - 3)}} \qquad (١٠-٦٩)$$

ولقد وجد Fisher أن قيم Z تكون متحيزة biased فكل منها أكبر بهذه الكمية:

$$\frac{\rho}{2(n_i - 1)} \qquad (١٠-٧٠)$$

ويظهر أثر هذا التحيز واضحاً عند الرغبة فى الحصول على تقدير لمعامل الارتباط فى العشيرة من عدد كبير من معاملات الارتباط للعينات المختلفة، حيث تتجمع هذه الكمية الصغيرة لكل ويظهر تأثيرها؛ ولذا يجب التصحيح لهذا التحيز بالعلاقة التالية:

$$\text{corected}\, Z_i = \hat{Z}_i = Z_i - \left(\frac{\rho}{2(n_i - 1)}\right) \qquad (١٠-٧١)$$

وإذا لم يكن معامل الارتباط فى العشيرة معروفاً، وهذا ما يحدث غالباً، تستخدم r المقابلة $\overline{Z}_W$ بدلاً من ρ ثم يحسب متوسط قيم Z المصححة والموزونة كما يلى:

$$\overline{Z}^{\circ}_W = \frac{\sum W_i Z^{\circ}_i}{\sum W_i} \qquad (١٠-٧٢)$$

وتكون قيمة r المقابلة لقيمة $\overline{Z}^{\circ}_W$ هى معامل الارتباط فى العشيرة.

وإذا كان حجم العينات لا يختلف كثيراً فقد يهمل الوزن ويحسب المتوسط حيث يقسم مجموع قيم Z المقابلة لقيم r على عددها ثم تستخرج قيمة r المقابلة.

مثال ١٠-٢١

البيانات التالية مأخوذة من عشيرة ما:

العينة	معامل الارتباط	حجم العينة
١	0.62	30
٢	0.59	5
٣	0.65	12
٤	0.60	17

هل هذه العينات مأخوذة من عشيرة معامل الارتباط فيها وليكن ρ، وما هو التقدير الأكثر صلاحية لهذا المعامل؟

العينة	معامل الارتباط r_i	حجم العينة	Z_i	الوزن $W_i = n_i - 3$	$W_i Z_i$	$W_i Z_i^2$
1	0.62	30	0.725	27	19.575	14.192
2	0.59	5	0.678	2	1.356	0.919
3	0.65	12	0.775	9	6.975	5.406
4	0.60	17	0.693	14	9.702	6.723
المجموع		**64**		**52**	**37.608**	**27.240**

ولاختبار فرض العدم بأن معاملات الارتباط في العينات المختلفة لها نفس معامل الارتباط في العشيرة فإن:

حيث إن
$$\chi^2 = \sum(n_i - 3)Z_i^2 - \frac{[\Sigma(n_i - 3)Z_i]^2}{\Sigma(n_i - 3)}$$

وبالتعويض $\chi^2 = 27.240 - \frac{(37.608)^2}{52} = 0.04$ بدرجات حرية $(4 - 1 = 3)$.

والقيمة الجدولية $\chi^2_{(3, 0.05)} = 7.81$، وعلى ذلك فإن χ^2 المحسوبة غير معنوية أي أن العينات التي قدر فيها معاملات الارتباط مأخوذة من نفس العشيرة. ويكون التقدير الأكثر صلاحية لمعامل الارتباط في العشيرة هو القيمة المقابلة $\bar{Z}_W$.

وحيث إن $\bar{Z}_W = \frac{\sum W_i Z_i}{\sum W_i} = \frac{37.608}{52} = 0.72$ فإن قيمة r التى تقابلها هى 0.62.

ولتصحيح التحيز لتقدير معامل الارتباط فى العشيرة يستخدم التقدير 0.62 بدلاً من ρ فى الكمية $\frac{\rho}{(2)(n_i - 1)}$ حيث إن معامل الارتباط فى العشيرة غير معروف كالتالى:

العينة	حجم العينة	الوزن $W_i = n_i - 3$	r_i	Z_i	$\frac{\rho}{2(n_i - 1)}$	Z المصححة Z°	$W_i Z_i^\circ$ المصححة والموزونة
١	30	27	0.62	0.725	0.011	0.714	19.278
٢	5	2	0.59	0.678	0.078	0.600	1.200
٣	12	9	0.65	0.775	0.028	0.747	6.723
٤	17	14	0.60	0.693	0.019	0.674	9.436
المجموع	**64**	**52**					**36.637**

وبالتالى فإن $\bar{Z}_W^\circ = 36.637/52 = 0.704$ وقيمة r المقابلة هى 0.6 والتى تعتبر تقديراً غير متحيزاً لمعامل الارتباط فى العشيرة.

١٠-١٧ السبب والأثر فى تحليل الارتباط والانحدار

Cause and effect in regression and correlation analysis

عند تفسير نتائج تحليل الانحدار والارتباط لابد أن يكون من الواضح تماماً أنه ليس بالضرورة أن نتائج التحليل تدل على السبب والأثر cause and effect، فعلى سبيل المثال فى إحدى الدراسات بلغ معامل الارتباط بين دخل الأسرة واستهلاك اللحوم فى فترة زمنية معينة 0.6)، هذا ليس بالضرورة دليلا على أن زيادة دخل الأسرة يؤدى إلى زيادة استهلاك اللحوم أو زيادة استهلاك اللحوم يؤدى إلى زيادة دخل الأسرة ولكن هذا المعامل يدل على أن المتغيرين يتغيران سوياً فى نفس الاتجاه وذلك قد يكون بسبب وجود متغير ثالث، وليكن ارتفاع مستوى المعيشة بصفة عامة، والذى يؤثر فى كل من الدخل واستهلاك اللحوم. وفى حالة ما إذا أمكن تثبيت هذا العامل

الثالث وتتعامل معه بطريقة إحصائية معينة فإنه فى هذه الحالة فقط فإن معامل الارتباط قد يكون ذا مدلول معين. وهذا هو الهدف من دراسة الارتباط المتعدد أو الجزئى والذى سوف يتم تناوله فيما بعد. وبالمثل عند دراسة الانحدار البسيط فإن الحصول على معامل انحدار معنوى لا يدل أيضاً على السبب والأثر إلا إذا أخذت عوامل أخرى فى الاعتبار.

وعلى الرغم من أن نتائج دراسة الانحدار والارتباط البسيط ليست دليلا على السبب والأثر فإن تلك النتائج ربما تعطى بعض الاقتراحات التى قد تساعد فى دراسة السبب والأثر فمثلاً وجد أن التدخين له علاقة ارتباط عالية بالإصابة بسرطان الرئة مما أوحى إلى ضرورة دراسة السبب والأثر بين المتغيرين لاستيضاح العلاقة البيوكيميائية بينهما.

١٠-١٨ تباين الدالة الخطية Variance of a linear function

إذا كانت $L_1 = Y_1 + Y_2$ فإن تباين L_1 عبارة مجموع التباينين لكل من Y_1 و Y_2.

أى أن:

$$\sigma^2_{L_1} = \sigma^2_{y_1} + \sigma^2_{y_2} \qquad (\text{٧٣-١٠})$$

وذلك إذا كان Y_1 و Y_2 غير مرتبطين uncorrelated. وأيضاً فإن تباين الفرق بين هذين المتغيرين يساوى مجموع التباينين. أى أنه إذا كانت $L_2 = Y_1 - Y_2$ فإن:

$$\sigma^2_{L_2} = \sigma^2_{y_1} + \sigma^2_{y_2} \qquad (\text{٧٤-١٠})$$

أى أن تباين الفرق يساوى تباين المجموع فى حالة المتغيرات غير المرتبطة. ويمكن إثبات ذلك كما يلى:

حيث إن التباين لأى متغير وليكن L هو:

$$V(L) = \frac{\sum (L - \mu_L)^2}{N}$$

وحيث إن $L = Y_1 + Y_2$ فإن $\mu_L = \mu_{Y_1} + \mu_{Y_2}$ وبالتالى:

$$V(L) = \frac{1}{N}\sum[Y_1 + Y_2 - (\mu_{Y_1} + \mu_{Y_2}]^2$$
$$= \frac{1}{N}\sum[(Y_1 - \mu_{Y_1}) + (Y_2 - \mu_{Y_2})]^2$$
$$= \frac{1}{N}[\sum(Y_1 - \mu_{Y_1})^2 + \sum(Y_2 - \mu_{Y_2})^2 + 2\sum(Y_1 - \mu_{Y_1})(Y_2 - \mu_{Y_2})$$
$$= \frac{\sum(Y_1 - \mu_{Y_1})^2}{N} + \frac{\sum(Y_2 - \mu_{Y2})^2}{N} + 2\frac{\sum(Y_1 - \mu_{Y_1})(Y_2 - \mu_{Y_2})}{N}$$

وبالتالى فإن

$$\sigma_L^2 = \sigma_{y_1}^2 + \sigma_{y_2}^2 + 2\operatorname{cov}(y_1, y_2) \qquad (١٠-٧٥)$$

وإذا كان المتغيران غير مرتبطين فإن الحد الأخير (التغاير covariance) يساوى صفراً.

وبالمثل يمكن إثبات أن تباين الفرق يساوى مجموع التباينين إذا كان المتغيران غير مرتبطين.

أما إذا كان المتغيران مرتبطين فإن تباين المجموع كما فى (١٠-٧٥) وحيث إن

$$\rho = \frac{\operatorname{cov}(y_1, y_2)}{\sigma_{y_1}\sigma_{y_2}}$$

أى أن $\operatorname{cov}(y_1, y_2) = \rho\sigma_{y_1}\sigma_{y_2}$، وعلى ذلك تصبح المعادلة (١٠-٧٥) على الصورة التالية:

$$\sigma_{L_1}^2 = \sigma_{y_1}^2 + \sigma_{y_2}^2 + 2\rho\sigma_{y_1}\sigma_{y_2} \qquad (١٠-٧٦)$$

وعلى ذلك فإن الارتباط الموجب يزيد من تباين المجموع عن مجموع التباين بالمقدار $2\rho\sigma_{y_1}\sigma_{y_2}$ والارتباط السالب يخفضه بنفس الكمية، أى

$$\sigma_{L_1}^2 = \sigma_{y_1}^2 + \sigma_{y_2}^2 - 2\rho\sigma_{y_1}\sigma_{y_2} \qquad (١٠-٧٧)$$

لاحظ هنا أن الارتباط الموجب ينقص تباين الفرق عن مجموع التباينين بينما الارتباط السالب يزيده عن مجموع التباينين. وفى تجارب الأزواج paired experiments في تحليل t فإن الهدف الأساسى هو إيجاد الارتباط الموجب بين فردى الزوج وبالتالى ينقص تباين الفرق عن مجموع التباينين بالكمية $2\rho\sigma_{y_1}\sigma_{y_2}$

وبنفس المفهوم يمكن إثبات أن:

$$\sigma^2_{(a_1Y_1+a_2Y_2)} = a_1^2\sigma^2_{y_1} + a_2^2\sigma^2_{y_2} + 2\rho a_1 a_2 \sigma_{y_1}\sigma_{y_2} \qquad (\text{٧٨-١٠})$$

حيث إن a_1 ، a_2 ثابتان.

وللتعميم : إذا كانت L دالة خطية بالصورة $L = a_1Y_1 + a_2Y_2 + \ldots + a_nY_n$ فإن تباين L هو:

$$\begin{aligned}\sigma^2_L &= a_1^2\sigma^2_{y_1} + a_2^2\sigma^2_{y_2} + \ldots + 2a_1a_2\rho_{y_1y_2}\sigma_{y_1}\sigma_{y_2} \\ &\quad + \ldots + 2a_{(n-1)}a_n\rho_{y_{(n-1)}\cdot y_n}\sigma_{y_{(n-1)}}\sigma_{y_n} \qquad (\text{٧٩-١٠}) \\ &= \sum a_i^2\sigma^2_{y_i} + \sum_i\sum_{j\neq i} a_ia_j\rho_{y_iy_j}\sigma_{y_i}\sigma_{y_j}\end{aligned}$$

وإذا كانت المتغيرات غير مرتبطة فإن الحد الثانى من المعادلة فى الطرف الأيمن يساوى صفراً وتصبح المعادلة:

$$\sigma^2_L = \sum a_i^2\sigma^2_{y_i} \qquad (\text{٨٠-١٠})$$

وإذا كان هناك دالتان خطيتان فى نفس المتغيرات: ولتكن الدالة الأولى $L_3 = a_1Y_1 + a_2Y_2$ والدالة الثانية $L_4 = b_1Y_1 + b_2Y_2$ ، فإن التغاير بينهما:

$$\begin{aligned}\text{cov}(L_3, L_4) &= a_1b_1\sigma^2_{y_1} + a_2b_2\sigma^2_{y_2} \\ &\quad + (a_1b_2 + a_2b_1)\text{cov}(Y_1, Y_2) \qquad (\text{٨١-١٠})\end{aligned}$$

وللتعميم إذا كانت:

$L_2 = b_1Y_1 + \ldots + b_nY_n$ ، $L_1 = a_1Y_1 + \ldots + a_nY_n$

فإن:

$$\text{cov}(L_1, L_2) = \sum a_i b_i \sigma_i^2 + \sum_i \sum_{j>i} (a_i b_j + a_j b_i)\text{cov}(Y_i, Y_j) \qquad (١٠-٨٢)$$

١٠-١٩ ارتباط الرتب Rank correlation

في حالة المتغيرات التي لا تتبع توزيع معين أو التي تتبع توزيعاً غير معروف المعالم (الثوابت) فإن حساب r كتقدير لمعامل الارتباط في العشيرة يكون غير صالح وخاصة عندما يكون التوزيع ذو المتغيرين بعيداً عن التوزيع الطبيعي ولذلك فإن إجراء تحوير للمتغيرين أملاً في أن يكون توزيعهما المشترك وثيق الشبه للتوزيع الطبيعي ذي المتغيرين يجعل تقدير ρ ممكنا في بعض الأحيان وغير ممكن في البعض الآخر. ولكن هل المتغيران مرتبطان وهل يتغيران في نفس الاتجاه أم في اتجاهين متضادين؟ وقد تم فيما سبق إيضاح أنه في حالة اختبار فرض العدم بأن معامل الارتباط في العشيرة يساوي صفراً فإنه يمكن استخدام r على أن يتوزع أحد المتغيرين طبيعياً، أما عندما يكون توزيع المتغيرين غير طبيعي فإن أحسن طريقة للإجابة على السؤال السابق هو ترتيب كل من المتغيرين (تنازلياً وتصاعدياً) ثم يختبر التوافق بين الترتيبين. وفي حالة البيانات غير المرتبة فإن أول خطوة هي ترتيب كل من المتغيرين كل منهما على حده ويستخدم معامل ارتباط الرتب r_s والذي وضعه Spearman حيث:

$$r_s = 1 - \frac{6\Sigma d_i^2}{n(n^2 - 1)} \qquad (١٠-٨٣)$$

حيث $d_i = Y_{1_i} - Y_{2_i}$ ، Y_{1_i} ، Y_{2_i} هما رتبة الفرد i في كل من المتغير الأول والثاني على التوالي. فإذا كانت Σd_i^2 تساوي صفراً فإن $r_s = \pm 1$.

ومعامل ارتباط الرتب تنحصر قيمته بين 1- في حالة عدم التوافق المطلق discordance و 1+ في حالة التوافق التام complete concordance. ولاختبار معنوية معامل ارتباط الرتب للعينات التي حجمها عشرة أزواج أو أقل يستخدم الحد الأدنى لقيمة r_s المعنوية على مستوى 5% ومستوى 1% كما وصفه Kendall (1970) في جدول ١٠-٧، حيث إن توزيع r وتوزيع r_s متساويان في حالة اختبار فرض العدم $r = 0$.

ويلاحظ أنه عندما يكون حجم العينة 4 مثلاً فإن الحد الأدنى لمعنوية r_s يكون أكبر من الواحد الصحيح عند مستوى 5% وأيضاً 1% وبالتالي فإن معنوية معامل

ارتباط الرتب فى مثل هذه الحالة لن تتحقق حيث إن الحد الأعلى له يساوى الواحد الصحيح.

ويستخدم معامل ارتباط الرتب فى الحالات التى يتعذر فيها قياس المتغيرات بطريقة كمية وإنما تعطى رتباً أو درجات، ولقد استنبط Kendall مقياساً آخر لدرجة التوافق بين المتغيرين لن يتم تناوله هنا فى الوقت الحالى.

جدول ١٠-٧ معنوية معامل ارتباط الرتب للعينات التى حجمها عشرة أزواج أو أقل

حجم العينة	الحد الأدنى لمعنوية r_s 5%	1%
٤ أو أقل	-	-
٥	1.000	-
٦	0.886	1.000
٧	0.786	0.929
٨	0.738	0.857
٩	0.683	0.817
١٠	0.648	0.781
١١ أو أكثر	**يستخدم جدول ٧ ملحق أ بدرجات حرية** $n-2$	

مثال ١٠-٢٢

حساب معامل ارتباط الرتب واختبار معنويته.

رقم البقرة	الترتيب حسب كمية اللبن	الترتيب حسب المطابقة للنموذج	الفرق d_i	d_i^2
١	3	2	1	1
٢	4	3	1	1
٣	1	1	0	0
٤	2	5	-3	9
٥	5	4	1	1
المجموع			**0**	**12**

بتطبيق المعادلة (١٠-٨٣) فإن

$$r_s = 1 - \frac{(6)(12)}{(5)(25-1)} = 0.4$$

ومن القيمة الجدولية (حجم العينة = 5) يتضح أن معامل ارتباط الرتب غير معنوى أى ليس هناك علاقة بين الترتيبين.

تمارين الباب العاشر

١٠-١ فى عينة ما وجدت القيم التالية للعلاقة بين الوزن Y_2 بالكيلوجرام والارتفاع Y_1 بالسنتيمتر:

$$n = 12 \qquad \sum Y_1 = 228 \qquad \sum Y_2 = 450$$

$$\sum y_1 y_2 = -542.5 \qquad \sum y_1^2 = 961 \qquad \sum y_2^2 = 1225$$

المطلوب حساب:

١ - معامل انحدار الوزن على الارتفاع ومعامل انحدار الارتفاع على الوزن.

٢ - معامل الارتباط.

٣ - تحليل التباين فى الوزن إلى مكوناته.

٤ - ايجاد قيمة $\hat{Y}_2$ عندما تكون Y_1 تساوى 30 ,19 ,22 سنتيمتر.

٥ - تمثيل العلاقة بين المتغيرين بيانياً.

١٠-٢ فى التمرين السابق اختبر الفرض بأن معامل انحدار الوزن على الارتفاع يساوى صفراً بمستوى معنوية 1% .

١٠-٣ فى التمرين ١٠-١ اختبر الفروض الآتية:

١ - $H_\circ : \rho = 0$

٢ - $H_\circ : \rho = -3$

١٠-٤ فى تجربة تسمين كان يتم وزن 20 حمل أسبوعياً فإذا كان رمز الوزن Y والأسابيع X وتوافرت البيانات التالية:

$$\sum Y = 400 \qquad \sum Y^2 = 9600 \qquad \sum XY = 5160$$

$$\sum X = 240 \qquad \sum X^2 = 3024$$

المطلوب حساب:

١ - معامل الانحدار b_{yx} ومعامل الارتباط بين المتغيرين.

٢- معامل التحديد.

٣- اختبر فرض العدم $H_0: \beta = 0$ ، $H_0: \rho = 0$. وضح جدول تحليل التباين.

٤- اكتب معادلة التنبؤ وارسم خط الانحدار رسماً دقيقاً.

١٠-٥ ما هي حدود الثقة لكل من الانحدار والجزء المقطوع من محور الصادات في تمرين ١٠-٤ وما هي حدود الثقة لخط الانحدار؟ بين ذلك بيانياً.

١٠-٦ إذا توافرت البيانات التالية:

العينة الأولى : معامل الارتباط = 0.325		حجم العينة = 25
العينة الثانية : معامل الارتباط = 0.34		حجم العينة = 15
العينة الثالثة : معامل الارتباط = 0.31		حجم العينة = 22

هل هذه العينات مسحوبة من عشيرة واحدة وما هو تقديرك لمعامل الارتباط؟. احسب الاختبارات الإحصائية اللازمة.

١٠-٧ إذا كانت البيانات التالية مسحوبة من عشيرة طبيعية ذات متغيرين

Y_1	0.1	1.0	1.1	1.7	1.8	2.1	2.1	2.3
Y_2	0.7	2.7	1.5	3.0	1.8	2.8	2.8	2.3
Y_1	2.6	2.7	2.9	3.4	4.4	4.4	4.9	6.1
Y_2	2.7	4.4	2.6	3.2	2.7	3.8	4.0	6.6

وكان معامل ارتباط العشيرة يساوى 0.7 ومعامل اعتماد العشيرة Y_2 على Y_1 يساوى 1

المطلوب حساب:

١- معامل الارتباط بين المتغيرين ومعامل اعتماد Y_2 على Y_1 وارسم خط الانحدار رسماً دقيقاً.

٢- معامل التحديد.

٣- اختر أقل أربع قيم للمتغير Y_1 وقيم Y_2 المقابلة لها وأيضاً أكبر أربع قيم للمتغير Y_1 وقيم Y_2 المقابلة لتكوين عينة من 8 أزواج لهذه العينة. احسب المطلوب فى ١- وقارن النتائج المتحصل عليها.

٤- قسم كل من قيم Y_2 إلى مكوناتها ثم احسب $\Sigma e_1 Y_1$، $\Sigma e_1 Y_2$ وقارن بين $\Sigma e_1 Y_1$، $\sum e_1^2$. اشرح معنى هذه النتائج.

١٠-٨ بين كيف أن تباين معامل الارتباط = $\frac{1-r^2}{n-2}$

حيث r معامل الارتباط و n حجم العينة (عدد أزواج المشاهدات).

١٠-٩ قام اثنان من المحكمين بترتيب 10 بقرات حسب حالتها الجسمية كما يلى:

رقم البقرة	المحكم الأول	المحكم الثانى
١	8	10
٢	9	8
٣	7	9
٤	5	3
٥	1	4
٦	4	2
٧	2	1
٨	3	5
٩	6	6
١٠	10	7

احسب معامل ارتباط الرتب واختبر معنويته.

١١ تحليل التباين متعدد التقسيمات

Multi-way analysis of variance

فى كثير من الحالات يكون التصميم أحادى الاتجاه أو التقسيم الأحادى للبيانات قد لا يشكل الوضع الأمثل لاستبيان حقيقة الأمور من اختبار للفروض الخاصة بتأثير معاملة معينة.

وينشأ ذلك إذا تعرف المجرب على أن هناك عاملاً آخر بجانب اختلاف المعاملات قد يدخل فى التأثير على نتيجة (قيم) البيانات المتحصل عليها. فمثلاً قد يكون لجنس الحملان تأثير على الأوزان عند الفطام مما قد يؤثر على اختبار الفرض الخاص بمعاملات غذائية معينة تتم فى فترة الرضاعة أو فى استيضاح الفروق بين السلالات. كما أن الأصناف المختلفة من محصول ما قد تكون سبباً للاختلافات يؤثر على دقة الحكم على استبيان الاختلافات فى كمية المحصول بين معاملات التسميد، أو أن المقياس المستخدم فى قياس عملية معينة يختلف باختلاف القائم على عملية القياس نفسها.

كل هذا يؤدى إلى عدم تجانس الوحدات التجريبية والتى كانت الأساس فى استخدام التقسيم الأحادى للبيانات وبالتالى أيضاً يمكن النظر إلى التقسيم متعدد الفئات على أنه يوسع مدى التجربة حيث إنه يقيس الفروق بين المعاملات المختلفة تحت ظروف لعوامل أخرى، وبالتالى يوسع فى مدى الاستدلال.

١١-١ التقسيم ثنائى الاتجاه Two-way classification

تأسيساً على ما سبق فإن من المشاهدات ما يمكن تصنيفها بناءً على أساسين للتصنيف، ويعرف هذا بالتقسيم ثنائى الاتجاه. وبالتالى فإن البيانات يمكن تبويبها على هيئة مستطيل بحيث تمثل الصفوف rows أحد أسس التصنيف (إحدى الاتجاهين) وتمثل الأعمدة columns التصنيف الآخر (الجهة الثانية). فمثلاً قد تكون الصفوف هى مستويات المعاملات المختلفة فى حين تمثل الأعمدة جنس الحيوان أو عمره أو سلالته أو القائمين بالقياس.

وكمثال يوضح التقسيم ثنائى الاتجاه فإن الجدول ١١-١ يمثل النتائج المتحصل عليها من تجربة استخدمت فيها ثلاثة أصناف من القمح أدخلت فى تجربة لتقدير تأثير أربعة معاملات من التسميد على المحصول الناتج حيث تمثل كل مشاهدة أو خلية من الاثنتى عشرة مشاهدة توليفة معاملة treatment combination وهى فى هذه الحالة يمثلها رقم واحد هو قيمة المحصول الناتج فى هذه الخلية Y_{ij} والتى تمثل التقاء المعاملة التسميدية i بالصنف من القمح j حيث يمكن تمثيل المشاهدات أو الخلايا cells فى جدول ١١-٢.

جدول ١١-١ محصول القمح الناتج فى تجربة تشمل 3 أصناف، 4 معاملات

معاملات التسميد	أصناف القمح أ	ب	ج	المجموع	المتوسط
١	64	72	74	**210**	**70**
٢	55	57	47	**159**	**53**
٣	59	66	58	**183**	**61**
٤	58	57	53	**168**	**56**
المجموع	**236**	**252**	**232**	**720**	**60**

جدول ١١-٢ التقسيم الثنائى الاتجاه (مشاهدة واحدة فى كل خلية)

الصفوف Rows	الأعمدة Column 1	2	...	J	...	C	Total	Mean
1	Y_{11}	Y_{12}	...	Y_{1j}	...	Y_{1c}	$Y_{1\cdot}$	$\overline{Y}_{1\cdot}$
2	Y_{21}	Y_{22}	...	Y_{2j}	...	Y_{2c}	$Y_{2\cdot}$	$\overline{Y}_{2\cdot}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	Y_{i1}	Y_{i2}	...	Y_{ij}	...	Y_{ic}	$Y_{i\cdot}$	$\overline{Y}_{i\cdot}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
r	Y_{r1}	Y_{r2}	...	Y_{rj}	...	Y_{rc}	$Y_{r\cdot}$	$\overline{Y}_{r\cdot}$
Total	$Y_{\cdot 1}$	$Y_{\cdot 2}$	...	$Y_{\cdot j}$	...	$Y_{\cdot c}$	$Y_{\cdot\cdot}$	
Mean	$\overline{Y}_{\cdot 1}$	$\overline{Y}_{\cdot 2}$	...	$\overline{Y}_{\cdot j}$	...	$\overline{Y}_{\cdot c}$		$\overline{Y}_{\cdot\cdot}$

ويلاحظ من جدول ١١-٢ أن كل قيمة أو مشاهدة Y_{ij} يمكن تحديدها عن طريق تحت حرفين two subscripts حيث يمثل الحرف الأول المعاملة (الصف) الذى تنتمى إليه المشاهدة فى حين يمثل الحرف الثانى العمود الذى تنتمى إليه المشاهدة. وفى جدول ١١-١ يلاحظ أن المشاهدة Y_{13} تمثل معاملة السماد الأولى وصنف القمح الثالث وقيمتها 74، فى حين أن Y_{42} تمثل المشاهدة التى تنتمى للمعاملة الرابعة فى الصنف الثانى وقيمتها 57.

وفى هذا الجزء سوف يتم عرض المعادلات التى تمكن من تقدير واختبار ما إذا كانت التباينات فى المحصول ترجع إلى اختلاف المعاملات التسميدية أم إلى اختلاف أصناف القمح المنزرع أم إلى الاختلافات الداخلية (الخطأ) أو إلى خليط من مصادر التباين هذه، ومن جدول ١١-٢ يلاحظ أن هناك عدداً rc من قيم Y_{ij} والتى يفترض أنها تتبع توزيعاً طبيعياً بمتوسط قدره μ وتباين σ^2 وأنها مستقلة عن بعضها. ويمثل $Y_{i.}$ ، $\overline{Y}_{i.}$ كلا من المجموع والمتوسط للصف i، فى حين أن $Y_{..}$ ، $\overline{Y}_{..}$ تمثل نفس القيمتين فى العمود j على التوالى، وأن الكميتين $Y_{.j}$ ، $\overline{Y}_{.j}$ تمثلان المجموع الكلى للقيم ومتوسطها العام على التوالى.

وحيث إن متوسط الصف $\overline{Y}_{i.}$ هو تقدير غير متحيز للمتوسط فى العشيرة $\mu_{i.}$ وأيضاً متوسط العمود $\overline{Y}_{.j}$ هو تقدير للمتوسط فى العشيرة $\mu_{.j}$ فإنه يمكن اختبار فروض العدم الخاصة بكل منها.

ولإجراء اختبار أن الفروق بين المشاهدات ترجع إلى التباين بين الصفوف (التقسيم الأول) فإن:

فرض العدم هو أن كل المتوسطات فى العشيرة متساوية أى:

$$H_\circ : \mu_{1.} = \mu_{2.} = \cdots = \mu_{i.} \qquad (١-١١)$$

والفرض البديل H_1 أن ليست كل المتوسطات متساوية أو على الأقل يوجد متوسط واحد يختلف عن باقى المتوسطات.

وفى نفس الوقت يمكن أيضاً اختبار الفرض الخاص بالاختلافات بين الأعمدة (التقسيم الثانى) أى أن:

فرض العدم هو أن كل المتوسطات فى العشيرة متساوية أى:

$$H_\circ : \mu_{.1} = \mu_{.2} = \cdots = \mu_{.j} \qquad (٢-١١)$$

والفرض البديل H_1 أن ليست كل المتوسطات متساوية أو على الأقل يوجد متوسط واحد يختلف عن باقى المتوسطات.

١١-٢ النموذج الرياضى Mathematical model

النموذج الرياضى الذى يستخدم فى حالة التقسيم ثنائى الجهة تكون صورته كما فى المعادلة التالية:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \qquad (٣-١١)$$

حيث تمثل الكمية μ المتوسط العام للعشيرة فى حين أن الكمية α_i هى تأثير الصف i ، β_j هى التأثير الخاص بالعمود j وأن الكمية ε_{ij} هى الخطأ الطبيعى والذى يفترض أن قيمته تكون مستقلة وتتبع توزيعاً طبيعياً متوسطه الصفر وتباينه σ_e^2 أى أن $\varepsilon \sim NID(0, \sigma_e^2)$.

ويفترض فى هذا النموذج الرياضى أن تأثير كل من الصف والعمود على المشاهدة يكون تجميعياً additive وسوف يناقش لاحقا كيف يجرى تحليل التباين فى حالة عدم توافر هذا الشرط وكيفية اختباره.

ويلاحظ من جدول ١١-٢ أن المعادلة الخاصة بالمتوسط العام هى مجموع للمعادلات الخاصة بالصفوف وأيضا وفى نفس الوقت هى مجموع للمعادلات الخاصة بالأعمدة، وعلى ذلك فإنه للحصول على حلول solutions لتأثيرات كل من الصفوف والأعمدة يلزم وضع بعض الاشتراطات conditions or restrictions لذلك، وأكثر تلك الاشتراطات استخداماً هو اشتراط أن $\sum_i \alpha_i = \sum_j \beta_j = 0$ حيث كل من α_i ، β_j تمثل انحرافا عن المتوسط العام μ.

وعليه عند تطبيق هذه الاشتراطات على النموذج (١١-٣) يمكن الحصول على المعادلات التالية:

$$\mu_i = \frac{\sum_j (\mu + \alpha_i + \beta_j)}{c} = \mu + \alpha_i \qquad (٤-١١)$$

$$\mu_j = \frac{\sum_i (\mu + \alpha_i + \beta_j)}{r} = \mu + \beta_j \qquad (٥-١١)$$

وبالتالى تصبح فروض العدم (١١-١) ، (١١-٢) كالتالى:

$$H_\circ : \alpha_1 = \alpha_2 = \cdots = \alpha_r \qquad (٦-١١)$$

والفرض البديل H_1 : على الأقل هناك قيمة واحدة من α_i لا تساوى صفراً.

$$H_\circ : \beta_1 = \beta_2 = \cdots = \beta_c \qquad (٧-١١)$$

والفرض البديل H_1 : على الأقل هناك قيمة واحدة من β_j لا تساوى صفراً.

وكل من الاختبارين الخاصين بفروض العدم السابقة مبنى على أساس الحصول على مقارنة تقديرات مستقلة للتباين العام للعشيرة σ_e^2 ويتأتى ذلك عن طريق تقسيم مجموع المربعات الكلية فى العينة إلى ثلاثة أجزاء مستقلة حسب المعادلة التالية:

$$\sum_i \sum_j (\overline{Y}_{ij} - \overline{Y}_{..})^2$$

$$= c \sum_{i=1}^{r} (\overline{Y}_{i.} - \overline{Y}_{..})^2 + r \sum_{j=1}^{c} (\overline{Y}_{.j} - \overline{Y}_{..})^2 + \sum_i \sum_j (Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})^2$$

(١١-٨)

ويمكن إثبات ذلك بإعادة ترتيب الكمية داخل القوس أى:

$$\begin{aligned}
\sum_i^r \sum_j^c (Y_{ij} - \overline{Y}_{..})^2 &= \sum_i \sum_j [(\overline{Y}_{i.} - \overline{Y}_{..}) + (\overline{Y}_{.j} - \overline{Y}_{..}) + (Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})]^2 \\
&= \sum_i \sum_j (\overline{Y}_{i.} - \overline{Y}_{..})^2 + \sum_i \sum_j (\overline{Y}_{.j} - \overline{Y}_{..})^2 \\
&\quad + \sum_i \sum_j (Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})^2 \\
&\quad + 2\sum_i \sum_j (\overline{Y}_{i.} - \overline{Y}_{..})(\overline{Y}_{.j} - \overline{Y}_{..}) \\
&\quad + 2\sum_i \sum_j (\overline{Y}_{i.} - \overline{Y}_{..})(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}) \\
&\quad + 2\sum_i \sum_j (\overline{Y}_{.j} - \overline{Y}_{..})(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})
\end{aligned}$$

ولكن بما أنه قد سبق إثبات أن مجموع أى انحرافات عن المتوسط يساوى صفراً فإن كل الكميات التى لا يوجد بها التربيعات (أى حاصل ضرب الانحرافات) تساوى صفراً أى أن:

$$\sum_i^r \sum_j^c (Y_{ij} - \overline{Y}_{..})^2 = c\sum_i (\overline{Y}_{i.} - \overline{Y}_{..})^2 + r\sum_j (\overline{Y}_{.j} - \overline{Y}_{..})^2 + \sum_i \sum_j (Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})^2$$

ويلاحظ أن الثلاثة مجاميع للمربعات التى ينقسم إليها مجموع المربعات الكلى يؤول كل منها إلى مصدر من مصادر الاختلافات، ويمكن بالتالى الرمز إلى المعادلة (١١-٨) فى صورة مكونات مجاميع المربعات بالمعادلة التالية:

$$TSS = RSS + CSS + ESS \qquad (١١-٩)$$

وكل منها يمثل جزءاً من مجاميع المربعات فى (١١-٨) حيث:

مجموع المربعات الكلى: $\text{Total sum of squares (TSS)} = \sum_i \sum_j (Y_{ij} - \overline{Y}_{..})^2$

مجموع المربعات للصفوف: $\text{Row sum of squares (RSS)} = c\sum_i (\overline{Y}_{i.} - \overline{Y}_{..})^2$

مجموع المربعات للأعمدة: $\text{Column sum of squares (CSS)} = r\sum_j (\overline{Y}_{.j} - \overline{Y}_{..})^2$

مجموع المربعات للخطأ:

$$\text{Error sum of squares (ESS)} = \sum_i \sum_j (Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})^2$$

وهنا تجدر الإشارة إلى أنه يوجد ثلاثة تقديرات مستقلة لقيمة σ_e^2 من كل من مجاميع المربعات الثلاثة وهى تلك التى بين الصفوف، بين الأعمدة والخطأ. فأول تقدير هو من بين الصفوف RSS وهذا مبنى على أساس المتوسطات الخاصة بالصفوف وانحرافاتها عن المتوسط العام وله درجات حرية $(r-1)$ وعلى ذلك فإن التقدير الأول لقيمة التباين σ^2 يقدر من $S_R^2 = RSS/(r-1)$ وهذا يعتبر تقديراً غير منحاز لقيمة σ^2 إذا كان فرض العدم صحيحاً أى $\alpha_1 = \alpha_2 = \cdots = \alpha_r = 0$. ولكنه ى حالة ما إذا كان واحداً أو أكثر من قيم α لا تساوى صفراً فإن القيمة المحسوبة S_R^2 تصبح أكبر من قيمة σ^2.

وبنفس المنطق فإن التقدير الثانى يحسب من متوسط المربعات الخاص بالأعمدة ى $S_C^2 = CSS/(c-1)$ حيث إن هناك $(c-1)$ درجات حرية. وهو أيضاً تقدير غير منحاز لقيمة σ^2 فى حالة صحة فرض العدم أى $\beta_1 = \beta_2 = \cdots = \beta_c$. أما إذا كانت إحدى قيم β لا تساوى صفراً فإن التقدير الناتج S_C^2 يعطى قيمة أكبر من قيمة σ^2.

والتقدير الثالث لقيمة σ^2 هو الذى يحسب من مجموع مربعات الخطأ ESS وله $(r-1)(c-1)$ درجات حرية وهو أيضاً مستقل عن التقديرين السابقين وبذلك فإن

القيمة المحسوبة له هى $S_E^2 = ESS/(r-1)(c-1)$. وهذا التقدير غير متحيز ولا يتأثر بأى من فرضى العدم.

ولاختبار أى من الفرضين السابقين الخاصين بتأثيرات الصفوف أو الأعمدة فإن متوسط المربعات المراد اختباره سواء S_R^2أو S_C^2 تختبر بالمقارنة بـــ S_E^2وذلك بقسمة متوسط المربعات المختبر على S_E^2وهذا يعطى قيمة لها توزيع F فى حالة ما إذا كان فرض العدم صحيحاً وقيمة F الجدولية بمستوى المعنوية المفترض.

ففى حالة اختبار تأثير الصفوف يكون $F_R = S_R^2 / S_E^2$ بدرجات حرية $(r-1)$ للبسط و $(r-1)(c-1)$ للمقام. أما فى حالة اختبار تأثير الأعمدة تكون قيمته $F_C = S_C^2 / S_E^2$ بدرجات حرية $(c-1)$ للبسط و $(r-1)(c-1)$ للمقام.

وكما ذكر سابقاً بالباب التاسع فى (٩-٣) فإن مجموع المربعات الكلى TSS يحسب أولاً من المشاهدات. ومن مجاميع الصفوف والأعمدة يتم حساب RSS، CSS ثم بالطرح يحسب مجموع المربعات للخطأ كالتالى:

$$ESS = TSS - RSS - CSS \qquad (١٠-١١)$$

ويعتبر هذا الإجراء صحيحاً، أى حساب مجموع المربعات للخطأ بالطرح طالما أن عدد الأفراد فى كل من الخلايا cells أو تحت الفئات subclasses متساوى أو متناسب proportional. ومن الواضح أن عدد درجات الحرية للخطأ أيضاً يمكن حسابها بالطرح مثل مجموع المربعات حيث:

$$(r-1)(c-1) = (rc-1) - (r-1) - (c-1) \qquad (١١-١١)$$

ويمكن أيضاً توضيح الكيفية والمعادلات التى يتم بواسطتها حساب مجاميع المربعات المصححة الكلية وللصفوف وللأعمدة من المجاميع كالتالى:

$$TSS = \sum_i \sum_j Y_{ij}^2 - \frac{Y_{..}^2}{rc} \qquad (١٢-١١)$$

$$RSS = \sum_i \frac{Y_{i.}^2}{c} - \frac{Y_{..}^2}{rc} \qquad (١٣-١١)$$

$$CSS = \sum_j \frac{Y_{.j}^2}{r} - \frac{Y_{..}^2}{rc} \qquad (١٤-١١)$$

وبالتالى يمكن تلخيص حسابات مجاميع المربعات وتحليل التباين الثنائى التقسيم كما في جدول ١١-٣.

جدول ١١-٣ تحليل التباين ثنائى التقسيم (مشاهدة واحدة فى كل خلية)

SOV	df	SS	MS	F
بين متوسطات الصفوف (المعاملات)	$(r-1)$	RSS	$S_R^2 = \dfrac{RSS}{r-1}$	$F_R = \dfrac{S_R^2}{S_E^2}$
بين متوسطات الأعمدة (القطاعات)	$(c-1)$	CSS	$S_C^2 = \dfrac{CSS}{c-1}$	$F_C = \dfrac{S_C^2}{S_E^2}$
الخطأ	$(r-1)(c-1)$	ESS	$S_E^2 = \dfrac{ESS}{(r-1)(c-1)}$	
المجموع	$(rc-1)$	**TSS**		

مثال ١١-١

باستخدام البيانات فى جدول ١١-١ اختبر فرضى العدم التاليين

١- أنه لا توجد فروق بين متوسطات المحصول لمعاملات التسميد المختلفة المجربة.

٢- أنه لا توجد فروق بين متوسطات المحصول لأصناف القمح المختلفة المزروعة.

الحل:

لمعاملات التسميد:

$$H_\circ : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$$

H_1 : على الأقل واحدة من قيم α لا تساوى صفر

لأصناف القمح:

$$H_\circ : \beta_1 = \beta_2 = \beta_3 = 0$$

H_1 : على الأقل واحدة من قيم β لا تساوى صفر

وبافتراض أن الاختبار سيجرى على مستوى معنوية 5%

١- حسب معادلة (١١-١٢) يحسب مجموع المربعات الكلى TSS كالتالى:

$$TSS = (64)^2 + (55)^2 + \cdots + (53)^2 - \frac{(720)^2}{12} = 662$$

٢- حسب معادلة (١١-١٣) يحسب مجموع المربعات بين الصفوف RSS كالتالى:

$$RSS = \frac{(210)^2 + (159)^2 + (183)^2 + (168)^2}{3} - \frac{(720)^2}{12} = 498$$

٣- حسب المعادلة (١١-١٤) يحسب مجموع المربعات بين الأعمدة CSS كالتالى:

$$CSS = \frac{(236) + (252) + (232)}{4} - \frac{(720)^2}{12} = 56$$

٤- حسب المعادلة (١١-١٠) يحسب مجموع المربعات للخطأ ESS كالتالى:

$$ESS = 662 - (498 + 56) = 108$$

ويلخص بالتالى تحليل التباين للبيانات فى جدول ١١-٤.

جدول ١١-٤ تحليل التباين للبيانات المذكورة فى جدول ١١-١

SOV	df	SS	MS	F
بين معاملات التسميد	3	498	166	$\frac{166}{18} = 9.22^*$
بين أصناف القمح	2	56	28	$\frac{28}{18} = 1.56$
الخطأ	6	108	18	
الكلى	**11**	**662**		

ومن نتائج تحليل التباين يتضح:

١- أنه لا يمكن قبول فرض العدم الخاص بتساوى تأثير معاملات التسميد على متوسط المحصول حيث إن قيمة F المحسوبة تزيد عن القيمة الجدولية والتى قيمتها $F_{(3, 6, .05)} = 4.76$.

٢- أنه لا يوجد ما يؤدى إلى الاعتقاد بأن هناك اختلافاً بين متوسط محصول أصناف القمح المجربة وبالتالى لا يمكن رفض فرض العدم الخاص بذلك لأن قيمة F المحسوبة تقل عن القيمة الجدولية والتى قيمتها $F_{(2, 6, .05)} = 5.14$.

١١-٣ المقارنة بين المتوسطات Group mean comparisons

إن المقارنة بين المتوسطات والتى سبق مناقشتها فى حالة التقسيم الأحادى بالباب لتاسع تنطبق على هذا التقسيم أيضاً، فلقد وجد أنه باختبار فرض العدم الخاص بتأثير لمعاملات التسميدية قد تم رفضها وبالتالى فإنه يتم قبول الفرض البديل أى أنه هناك على الأقل واحدة من المعاملات التسميدية متوسطها يختلف عن المتوسط العام. ولاختبار أى الفروق تعتبر مختلفة يجرى مثلاً اختبار LSD ومن فصل ٩-٧-١ فإن ختبار LSD كان

$$LSD = t\sqrt{\frac{2\,S^2}{n}}$$

حيث S^2 هو متوسط المربعات للخطأ كما أشير سابقاً. ومـــن جدول ١١-٣ يتم لتعويض عن قيمة الاختبار حيث $t_{(6, .05)} = 2.447$ وتصبح قيمة LSD هى:

$$LSD = (2.447)\sqrt{\frac{2(18)}{3}} = 8.48$$

وبعد ترتيب المتوسطات تنازلياً حيث كان أعلاها متوسط المعاملة الأولى وأقلها متوسط المعاملة الثانية ويمكن إيجاد الفروق بين المتوسطات فى الجدول التالى:

المعاملة	المعاملة	(١)	(٣)	(٤)	(٢)
المعاملة	المتوسطات	70	61	56	53
		الفروق			
(٢)	53	17*	8	3	
(٤)	56	14*	5		
(٣)	61	9*			

ومن الجدول السابق عند مقارنة قيمة اختبار LSD بالفروق بين المتوسطات يتضح أن الفروق والتى تصل إلى المعنوية هى تلك التى توجد بين متوسط المعاملة (١) وكل من المتوسطات الأخرى أما بقية الفروق بين المتوسطات فلم تكن كذلك.

ويمكن الحصول على مجموع المربعات الخاص بالمقارنة المستقلة orthogonal comparison الخاصة بالمعاملة الأولى ضد بقية المعاملات وذلك باستخدام ما سبق توضيحه فى الفصل ٩–٨ باستخدام المعادلة (٩–٨) لحساب قيمة المقارنة L وباستخدام المعادلة (٩–٩) لحساب مجموع المربعات الراجع لهذه المقارنة. وبذا تصبح المقارنة:

$$L = (3)(210) - (159 + 183 + 168) = 120$$

ومجموع المربعات الخاص بالمقارنة L $= 400 = \dfrac{(120)^2}{3(12)}$

وباختبار مجموع المربعات هذا ضد الخطأ يتضح أن $F = \dfrac{400}{18} = 22.2^{**}$ بدرجة حرية واحدة للبسط و 6 درجات حرية للخطأ وهى معنوية جداً.

إن استخدام أكثر من صنف من أصناف القمح فى التجربة لمعرفة تأثير التسميد فى هذا النوع من التحليل، والذى يعرف بتصميم القطاعات العشوائية الكاملة complete randomized block design، والذى تمثل فيه أصناف القمح القطاعات blocks بالإضافة إلى المعاملات التسميدية treatments يؤدى إلى زيادة مدى التجربة وقدرتها، وهو ما سوف يتم تناوله لاحقاً فى الباب الخامس عشر. حيث يلاحظ أن المعاملات التسميدية قد اتضحت نتائجها بتطبيقها على أكثر من صنف من أصناف القمح والتى لا تدخل اختلافاتها ضمن الاختلافات بين المعاملات التسميدية وبذلك يتحقق زيادة مدى التجربة على مجال أوسع فى التجريب.

مثال ١١–٢

استخدام برنامج SAS لتحليل البيانات فى جدول ١١–١ واختبار فروض العدم التاليين والمقارنة بين المتوسطات.

```
DATA TWOWAY;
INPUT FERTILIZ VARIETY $  CROP @@;
CARDS;
1 A 64 2 A 55 3 A 59 4 A 58
1 B 72 2 B 57 3 B 66 4 B 57
1 C 74 2 C 47 3 C 58 4 C 53
```

```
PROC GLM;
CLASS FERTILIZ VARIETY;
MODEL CROP = FERTILIZ VARIETY / SS3;
MEANS FERTILIZ /LSD;
CONTRAST 'FIRST FERTILIZ VS OTHERS' FERTILIZ 3 -1 -1 -1;
RUN;
```

النتائج:

The GLM Procedure

Class Level Information

Class	Levels	Values
FERTILIZ	4	1 2 3 4
VARIETY	3	A B C

Number of observations 12

The GLM Procedure

Dependent Variable: CROP

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	554.0000000	110.8000000	6.16	0.0234
Error	6	108.0000000	18.0000000		
Corrected Total	11	662.0000000			

R-Square	Coeff Var	Root MSE	CROP Mean
0.836858	7.071068	4.242641	60.00000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
FERTILIZ	3	498.0000000	166.0000000	9.22	0.0115
VARIETY	2	56.0000000	28.0000000	1.56	0.2856

The GLM Procedure
t Tests (LSD) for CROP

NOTE: This test controls the Type I comparisonwise error rate, not the xperimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	6
Error Mean Square	18
Critical Value of t	2.44691
Least Significant Difference	8.4764

Means with the same letter are not significantly different.

t Grouping	Mean	N	FERTILIZ
A	70.000	3	1
B	61.000	3	3
B	56.000	3	4
B	53.000	3	2

The GLM Procedure

Dependent Variable: CROP

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
FIRST FERTILIZ VS OTHERS	1	400.0000000	400.0000000	22.22	0.0033

١١-٤ التداخل Interaction

لفهم ومعرفة طبيعة التداخل افترض أنه فى تجربة ما لمعرفة تأثير كل من مستوى الطاقة (عالية و منخفضة) ومستوى البروتين (عالى و منخفض) على الزيادة فى وزن الحيوان كانت البيانات التالية:

	مستوى الطاقة	
مستوى البروتين	عالى	منخفض
عالى	8	6
منخفض	5	4

ومنها يتضح أنه عندما تغير مستوى البروتين من منخفض إلى عالٍ أدى هذا إلى زيادة فى الوزن مقدارها وحدتان عندما كان مستوى الطاقة منخفضاً بينما كان هذا التغير مقداره ثلاث وحدات عندما كان مستوى الطاقة عالياً ومعنى ذلك أن مدى تأثير التغير فى البروتين يعتمد على مستوى الطاقة.

أما إذا كانت البيانات المتحصل عليها هى كما يلى:

	مستوى الطاقة	
مستوى البروتين	عالى	منخفض
عالى	7	6
منخفض	5	4

فمعنى ذلك أن تغير البروتين من منخفض إلى عالٍ أدى إلى زيادة مقدارها وحدتين سواء كان مستوى الطاقة منخفضاً أو عالياً.

فى الحالة الأولى يكون هناك تداخل بين العاملين أما فى الحالة الأخيرة فلا يوجد تداخل، أى أنه يمكن أن يعرف التداخل بأنه الفرق بين الفروق أفقياً أو رأسياً، فإذا كان هذا الفرق مساوياً للصفر فليس هناك تداخل، أما إذا كان هذا الفرق لا يساوى الصفر فهذا يعنى وجود تداخل أى $1 = (4-6) - (5-8)$ فى الحالة الأولى، وفى الحالة الثانية $0 = (4-6) - (5-7)$ فهذا يعنى عدم وجود تداخل. كما يمكن أن يؤخذ الفروق أفقياً أيضاً وكلاهما يعطى نفس النتيجة.

ويمكن أن توضيح ذلك بما يلى فى حالة عاملين أ، ب لكل منهما مستويان.

	أ		
ب	مستوى ١	مستوى ٢	المتوسط
مستوى ١	μ_{11}	μ_{12}	$\overline{\mu}_{1\cdot}$
مستوى ٢	μ_{21}	μ_{22}	$\overline{\mu}_{2\cdot}$
المتوسط	$\overline{\mu}_{\cdot 1}$	$\overline{\mu}_{\cdot 2}$	$\overline{\mu}_{\cdot\cdot}$

فإذا كان $\mu_{11} - \mu_{12} = \mu_{21} - \mu_{22}$

أو $\mu_{11} - \mu_{21} = \mu_{12} - \mu_{22}$

أى $\mu_{11} - \mu_{21} - \mu_{12} + \mu_{22} = 0$، فإن هذا يعنى عدم وجود تداخل.
أما إذا لم تتحقق صحة هذه المتساويات فإن هذا يعنى وجود تداخل بين العاملين.

وفى حالة عدم وجود تداخل فإن تأثير كل من العاملين يكون تجميعياً additive وعليه فإن μ_{ij} (فى الحالة العامة) يمكن أن يعبر عنه كما يلى:

$$\begin{aligned}\mu_{ij} &= \overline{\mu}_{\cdot\cdot} + (\overline{\mu}_{i\cdot} - \overline{\mu}_{\cdot\cdot}) + (\overline{\mu}_{\cdot j} - \overline{\mu}_{\cdot\cdot})\\ &= \overline{\mu}_{i\cdot} + \overline{\mu}_{\cdot j} - \overline{\mu}_{\cdot\cdot}\end{aligned}$$

حيث

$\bar{\mu}..$: المتوسط العام

$\bar{\mu}_i.-\bar{\mu}..$: تأثير المستوى i فى العامل الأول

$\bar{\mu}._j-\bar{\mu}..$: تأثير المستوى j فى العامل الثانى

أى إذا كانت $\mu_{ij}-\bar{\mu}_i.-\bar{\mu}._j+\bar{\mu}..=0$ فإن هذا يعنى عدم وجود تداخل.

أما إذا كانت هذه الكمية لا تساوى صفراً فهى تعنى تأثير التداخل (التداخل بين العاملين).

١١-٥ تحليل التباين فى اتجاهين مع وجود تداخل

Two-way classification analysis of variance with interaction

فى حالة وجود أكثر من مشاهدة observation فى كل خلية فإن مصادر الاختلاف تحتوى على مصدر آخر وهو التداخل كما سيتضح فيما بعد. فإذا كان هناك عاملان A ، B والعامل الأول A له r مستوى أما العامل الثانى فله c مستوى فيكون عدد الخلايا (r)(c)، فإذا كان بكل خلية n ملاحظة فيكون عدد الملاحظات الكلى (r)(c)(n) ويعبر عن كل ملاحظة بالنموذج التالى:

$$Y_{ijk}=\mu+A_i+B_j+(AB)_{ij}+e_{ijk}$$

حيث

$i=1,,2,\cdots,r$ ، $j=1,,2,\cdots,c$ ، $k=1,,2,\cdots,n$ ،

و $(AB)_{ij}$ تمثل تأثير التداخل بين العاملين B, A.

وبقية مكونات النموذج فهى كما سبق تعريفها.

ويسمى كل من أثر A وأثر B بالأثر الرئيسى main effcct بينما أثر AB بأثر التداخل interaction effect.

ويمكن تمثيل المشاهدات أو الخلايا cells فى التقسيم ثنائى الاتجاه فى حالة وجود أكثر من مشاهدة فى كل خلية كالتالى (القيم التى بين الأقواس تمثل مجاميع الخلايا):

		B						
		B_1	...	B_j	...	B_c	**Sum**	**Men**
		Y_{111}	...	Y_{1j1}	...	Y_{1c1}		
		Y_{112}	...	Y_{1j2}	...	Y_{1c2}		
	A_1	$\vdots$	...	$\vdots$	...	$\vdots$		
		Y_{11n}	...	Y_{1jn}	...	Y_{1cn}		
		$(Y_{11}.)$	...	$(Y_{1j}.)$	...	$(Y_{1c}.)$	$Y_1..$	$\overline{Y}_1..$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
		Y_{i11}	...	Y_{ij1}	...	Y_{ic1}		
		Y_{i12}	...	Y_{ij2}	...	Y_{ic2}		
A	A_i	$\vdots$	...	$\vdots$	...	$\vdots$		
		Y_{i1n}	...	Y_{ijn}	...	Y_{icn}		
		$(Y_{i1}.)$	...	$(Y_{ij}.)$	...	$(Y_{ic}.)$	$Y_i..$	$\overline{Y}_i..$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
		Y_{r11}	...	Y_{rj1}	...	Y_{rc1}		
		Y_{r12}	...	Y_{rj2}	...	Y_{rc2}		
	A_r	$\vdots$	...	$\vdots$	...	$\vdots$		
		Y_{r1n}	...	Y_{rjn}	...	Y_{rcn}		
		$(Y_{r1}.)$	...	$(Y_{rj}.)$	...	$(Y_{rc}.)$	$Y_r..$	$\overline{Y}_r..$
Sum		$Y._1.$	...	$Y._j.$	...	$Y._c.$	$Y...$	
Mean		$\overline{Y}._1.$	...	$\overline{Y}._j.$	...	$\overline{Y}._c.$		$\overline{Y}...$

ويقدر النموذج السابق كما يلى:

$$Y_{ijk} = \overline{Y}... + (\overline{Y}_i.. - \overline{Y}...) + (\overline{Y}._j. - \overline{Y}...)$$
$$+ (\overline{Y}_{ij}. - \overline{Y}_i.. - \overline{Y}._j. + \overline{Y}...) + (Y_{ijk} - \overline{Y}_{ij}.)$$

حيث

$\overline{Y}...$: هو تقدير للمتوسط μ

و $(\overline{Y}_i.. - \overline{Y}...)$: تأثير المستوى i من العامل A

و $(\overline{Y}._j. - \overline{Y}...)$: تأثير المستوى j من العامل B

و $(\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...})$: تأثير التداخل بين العاملين

أما $(Y_{ijk} - \overline{Y}_{ij.})$: تقدير للخطأ الخاص بالملاحظة Y_{ijk}

ويقدر مجموع المربعات الكلى المصحح (TSS) كما يلى:

$$TSS = \sum_i \sum_j \sum_k (Y_{ijk} - \overline{Y}_{...})^2$$

$$= \sum_i \sum_j \sum_k Y_{ijk} - \frac{(Y_{...})^2}{rcn} \qquad (١٥-١١)$$

والحد الأخير فى (١١-١٥) يعرف بمعامل التصحيح CF أى

$$CF = \frac{(Y_{...})^2}{rcn}$$

ويقسم مجموع المربعات الكلى المصحح (TSS) إلى ما يلى:

مجموع المربعات بين مستويات العامل A (SSA) Sum of squares between A's حيث

$$SSA = cn \sum_i (\overline{Y}_{i..} - \overline{Y}_{...})^2 = \sum_i \frac{Y_{i...}^2}{cn} - CF \qquad (١٦-١١)$$

ومجموع المربعات بين مستويات العامل B (SSB) Sum of squares between B's حيث

$$SSB = rn \sum_j (\overline{Y}_{.j.} - \overline{Y}_{...})^2 = \sum_j \frac{Y^2_{.j.}}{rn} - CF \qquad (١٧-١١)$$

ومجموع المربعات للتداخل (SSI) Interaction sum of squares حيث

$$SSI = \sum_i \sum_j \frac{Y^2_{ij.}}{n} - CF - SSA - SSB \qquad (١٨-١١)$$

أما المكون الأخير وهو مجموع المربعات للخطأ (SSE) Error sum of squares حيث

$$SSE = \sum_i \sum_j \sum_k (Y_{ijk} - \overline{Y}_{ij.})^2 = \sum_i \sum_j \sum_k Y_{ijk}^2 - \sum_i \sum_j \frac{Y_{ij.}^2}{n} \qquad (١٩-١١)$$

ويمكن تقدير (SSE) بالطرح أيضاً كما يلي:

$$SEE = TSS - (SSA + SSB + SSI) \qquad (٢٠-١١)$$

ويمثل جدول ١١-٥ تحليل التباين فى حالة عاملين مع وجود تداخل.

وتختلف قيمة متوسط المربعات المتوقع (EMS) حسب نوع النموذج المفترض فذا كان تأثير كل من A ، B عشوائياً فإن النموذج يكون عشوائياً بينما إذا كان كل من التأثيرين ثابتاً فإن النموذج يكون ثابتاً. أما إذا كان أحدهما ثابتاً والآخر عشوائياً فن النموذج يكون خليطاً mixed. وفى كل هذه الحالات فإن مجموع مربعات الانحرافات ودرجات الحرية متشابهة.

ودرجات الحرية للتداخل فهى دائماً مساوية لحاصل ضرب درجات الحرية للعوامل المتداخلة.

ولإجراء اختبارات المعنوية فإن هذا يتوقف على نوع النموذج كما هو مبين بالأسهم فى جدول ١١-٥ لكل الحالات الممكنة.

ويجب ملاحظة أنه فى حالة معنوية التداخل يعطى له أهمية أكبر من معنوية كل من العاملين الأساسيين. أما إذا كان التداخل غير معنوى تصبح الأهمية لمعنوية كل من العاملين الرئيسيين. وفى هذه الحالة هناك رأيان: الأول هو إضافة مجموع المربعات وأيضاً درجات الحرية للتداخل إلى مجموع المربعات ودرجات حرية الخطأ ثم تجرى اختبارات المعنوية وهذا يؤدى إلى زيادة درجات حرية الخطأ وبالتالى زيادة حساسية التجربة (أى قوة الاختبار). أما الرأى الآخر فيترك جدول تحليل التباين كما هو محتوياً على السطر الخاص بالتداخل. واحتمال الخطأ من النوع الأول فى الحالة الأولى يكون مشروطاً بعدم وجود تداخل والذى اختبر عند α ، β معينين.

وفى الحالات التى يكون فيها التداخل معنوياً يجب الحذر، كما سبق القول، عند تسير آثار العوامل الأساسية main effects ويمكن توضيح ذلك عند النظر للشكل ١١-١ (أ، ب، ج).

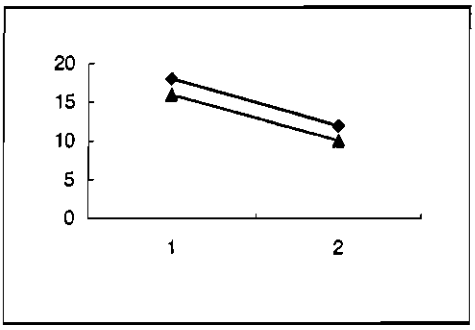
جدول ١١-٥ جدول تحليل التباين فى حالة عاملين مع وجود تداخل

SOV	df	SS	MS	EMS: A,B random	EMS: A,B fixed	EMS: A fixed, B random	EMS: A random, B fixed
Between A's	r-1	SSA	SSA/(r-1)	$\sigma_e^2 + n\sigma_{AB}^2 + cn\sigma_A^2$	$\sigma_e^2 + cn\,K_A^2$	$\sigma_e^2 + n\sigma_{AB}^2 + cnK_A^2$	$\sigma_e^2 + cn\,\sigma_A^2$
Between B's	c-1	SSB	SSB/(c-1)	$\sigma_e^2 + n\sigma_{AB}^2 + m\sigma_B^2$	$\sigma_e^2 + m\,K_B^2$	$\sigma_e^2 + m\sigma_B^2$	$\sigma_e^2 + n\sigma_{AB}^2 + mK_B^2$
A x B	(r-1) (c-1)	SSI	SSI/(r-1)(c-1)	$\sigma_e^2 + n\sigma_{AB}^2$	$\sigma_e^2 + n\,K_{AB}^2$	$\sigma_e^2 + n\sigma_{AB}^2$	$\sigma_e^2 + n\sigma_{AB}^2$
Error	rc(n-1)	SSE	SSE/rc(n-1)	σ_e^2	σ_e^2	σ_e^2	σ_e^2
Total	rcn-1	TSS					

$$K_{AB}^2 = \frac{n\Sigma_i \Sigma_j (AB)_{ij}^2}{(r-1)(c-1)},\ K_B^2 = \frac{m\ \Sigma_j\ B_j^2}{c-1},\quad K_A^2 = \frac{cn\,\Sigma_i\ A_i^2}{r-1}$$

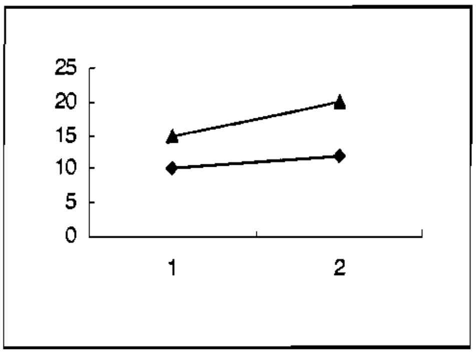
حالة (أ)

	سلالة أ	سلالة ب
بيئة ١	18	16
بيئة ٢	12	10



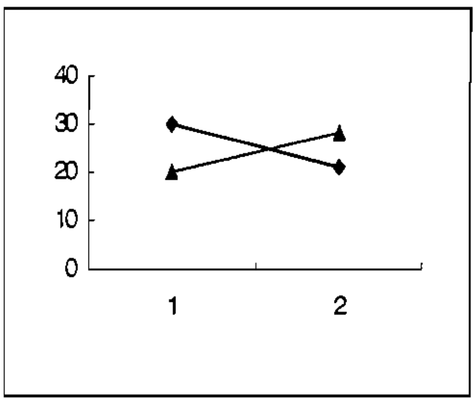
حالة (ب)

	معاملة أ	معاملة ب
موقع ١	10	15
موقع ٢	12	20



حالة (ج)

	درجة حرارة أ	درجة حرارة ب
درجة رطوبة ١	30	20
درجة رطوبة ٢	21	28



شكل ١١-١ حالات مختلفة من التداخل بين عاملين

فى الحالة (أ) لا يوجد تداخل ويمكن التعميم بسهولة بالنسبة لآثار العوامل الرئيسية وأى أثر كل من البيئة والسلالة. أى يمكن القول إن البيئة (١) أعلى من البيئة (٢) والسلالة (أ) أعلى من (ب) دون تحفظ.

وفى الحالة (ب) الفرق بين المعاملتين فى الموقع الأول هو $-5 = 15 - 10$ والفرق فى الموقع الثانى $-8 = 20 - 12$، أى يمكن القول إن المعاملة (أ) أقل من (ب)، أى يمكن القول إن المعاملة (أ) أقل من (ب) والموقع (١) أقل من (٢) ولكن بتحفظ حيث إن هناك فرقاً بين الفرقين وإذا أريد التعميم بالنسبة للآثار الرئيسية فيجب ألا يتعدى المدى المستويات من المعاملات المستخدمة فى التجربة.

بينما فى الحالة (ج) لا يجوز التعميم مطلقاً حيث إن درجة الحرارة (أ) تعطى قراءة أعلى عند درجة رطوبة (أ)، وأدنى عند درجة الرطوبة الأخرى بعكس درجة الحرارة (ب).

مثال ١١-٣

ثلاث معاملات تغذية أ، ب، ج جربت على مجموعتين من الحملان ذكوراً وإناثاً بحيث كان هناك حيوانان من كل جنس فى كل عليقة وكانت نتائج التجربة كالتالى (بعد أخذ وسط فرضى). والمطلوب إجراء تحليل التباين لمصادره المختلفة مع إجراء اختبارات المعنوية اللازمة.

	المعاملات الغذائية			
الجنس	**أ**	**ب**	**ج**	**المجموع**
ذكور	2	2	4	**18**
	3	2	5	
إناث	1	2	2	**11**
	1	2	3	
المجموع	**7**	**8**	**14**	**29**

مجموع المربعات الكلى المصحح TSS :

$$TSS = 2^2 + 3^2 + ... + 2^2 + 3^2 - \frac{(29)^2}{12} = 14.92$$

مجموع المربعات بين الجنسين SSS :

$$SSS = \frac{(18)^2 + (11)^2}{6} - \frac{(29)^2}{12} = 4.09$$

مجموع المربعات بين المعاملات SST :

$$SST = \frac{7^2 + 8^2 + (14)^2}{4} - \frac{(29)^2}{12} = 7.17$$

مجموع المربعات للتداخل SSI :

$$SSI = \frac{5^2+4^2+9^2+2^2+4^2+5^2}{2} - \frac{(29)^2}{12} - 4.09 - 7.17 = 2.16$$

مجموع المربعات للخطأ SSE :

$$SSE = 14.92 - (4.09 + 7.17 + 2.16) = 1.5$$

وإذا كان النموذج ثابتاً فإن جدول تحليل التباين واختبارات المعنوية تكون كما يلى:

ANOVA table

SOV	df	SS	MS	EMS
بين الجنسين (S)	1	4.09	4.09^{**}	$\sigma_e^2 + 6K_s^2$
بين المعاملات (T)	2	7.17	3.59^{**}	$\sigma_e^2 + 4K_t^2$
التداخل (S x T)	2	2.16	1.08	$\sigma_e^2 + 2K_{st}^2$
الخطأ	6	1.5	0.25	σ_e^2
الكلى	**11**	**14.92**		

ولاختبار معنوية الجنس فإن قيمة F المحسوبة $F = \frac{4.09}{0.25} = 16.36$ وهى أكبر من قيمة F الجدولية عند درجات حرية 1، 6 ومستوى معنوية 1% أى أن تأثير الجنس معنوى جداً.

أما قيمة F المحسوبة لإجراء اختبار لمعنوية المعاملات فهى $F = \frac{3.59}{0.25} = 14.36$ وهى أيضاً أكبر من قيمة F الجدولية عند درجات حرية 2، 6 ومستوى معنوية 1% .

أما قيمة F للتداخل فهى $F = \frac{1.08}{0,25} = 4.32$ وهى أقل من القيمة الجدولية عند مستوى معنوية 5% ودرجات حرية 2، 6 أى أنه ليس هناك تداخل.

وعلى ذلك فإن فرض العدم بتساوى متوسطى الجنسين يرفض ويقبل الفرض البديل الذى ينص على أن هناك فروقاً معنوية بدرجة ثقة 99% بين الجنسين.

مع تحيات د. سلام حسين الهلالي salamalhelali@yahoo.com

أما فرض العدم الخاص بتساوى متوسطات المعاملات فيرفض أيضاً ويقبل الفرض البديل الذى ينص على أن بعض هذه المتوسطات يختلف، أى أن الفروق بين المعاملات معنوى وغير راجع للصدفة بينما فرض العدم الثالث والخاص بتساوى متوسطات الخلايا فلا يمكن رفضه.

ولمعرفة أى أزواج المتوسطات يختلف عن بعضه معنوياً فإنه يمكن إجراء اختبار دنكن كما يلى:

متوسط المعاملة الأول (أ) $\frac{7}{4} = 1.75$

متوسط المعاملة الثانية (ب) $\frac{8}{4} = 2.00$

متوسط المعاملة الثالثة (ج) $\frac{14}{4} = 3.50$

ويكون ترتيب المتوسطات (ج) ثم (ب) ثم (أ)

والخطأ القياسى $S_{\overline{Y}} = \sqrt{\frac{0.25}{4}} = 0.25$

والفرق بين متوسط المعاملة (ج) ومتوسط المعاملة (أ) يساوى 1.75 وهو أكبر من قيمة دنكن والتى تساوى $0.895 = (3.58)(0.25)$ حيث المدى هنا 3 ودرجات حرية الخطأ تساوى 6، أما الفرق بين متوسط المعاملة (ج) ومتوسط المعاملة (ب) فهو 1.5 وهو أيضاً أكبر من قيمة دنكن والتى تساوى $0.865 = (3.46)(0.25)$ حيث المدى فى هذه الحالة 2 وعلى ذلك فإن متوسط المعاملة ج يختلف معنوياً عن كل من متوسط المعاملة (أ) ومتوسط المعاملة (ب). أما الفرق بين متوسط المعاملة (ب) ومتوسط المعاملة (أ) فهو 0.25، أقل من قيمة دنكن 0.865 حيث المدى 2 ونفس درجة الحرية (أى 6) وعلى ذلك فالفرق بينهما غير معنوى.

ويمكن أن يستنتج من هذه التجربة وتحليلها أنه إذا كانت المعاملات الثلاث متوافرة فإن المعاملة (ج) ستعطى أفضل النتائج، أما فى حالة الاضطرار لاستخدام إحدى المعاملتين (أ)، (ب) فإن اختيار أى منهما يتوقف على عوامل أخرى كتكاليف كل معاملة مثلاً حيث إن الفرق بينهما غير معنوى.

مثال ١١-٤

استخدام برنامج SAS مثال ١١-٣

```
DATA LAMB;
INPUT RATION $ SEX $  RESPONSE @@;
CARDS;
A M 2 A M 3 A F 1 A F 1
B M 2 B M 2 B F 2 B F 2
C M 4 C M 5 C F 2 C F 3
PROC GLM;
CLASS RATION SEX;
MODEL RESPONSE = SEX RATION SEX*RATION/SS3;
MEANS RATION/DUNCAN;
RUN;
```

لاحظ

يمكن وضع نموذج التحليل بطريقة مختصرة عند الرغبة فى وضع التداخل فى النموذج كالتالى MODEL RESPONSE = RATION | SEX

نتائج التحليل:

The GLM Procedure

Class Level Information

Class	Levels	Values
RATION	3	A B C
SEX	2	F M

Number of observations 12

The GLM Procedure

Dependent Variable: RESPONSE

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	13.41666667	2.68333333	10.73	0.0059
Error	6	1.50000000	0.25000000		
Corrected Total	11	14.91666667			

R-Square	Coeff Var	Root MSE	RESPONSE Mean
0.899441	20.68966	0.500000	2.416667

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SEX	1	4.08333333	4.08333333	16.33	0.0068
RATION	2	7.16666667	3.58333333	14.33	0.0052
RATION*SEX	2	2.16666667	1.08333333	4.33	0.0685

The GLM Procedure

Duncan's Multiple Range Test for RESPONSE

NOTE: This test controls the Type I comparisonwise error rate, not the xperimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	6
Error Mean Square	0.25

Number of Means	2	3
Critical Range	.8651	.8966

Means with the same letter are not significantly different.

Duncan Grouping	Mean	N	RATION
A	3.5000	4	C
B	2.0000	4	B
B	1.7500	4	A

صندوق ١١-١

- الأثر الرئيسى هو الفرق بين متوسطات مستويات العوامل بينما التداخل هو الفرق بين الفروق بين متوسطات عامل ما عند مستويات عامل آخر.
- فى غياب التداخل يمكن التعميم بالنسبة للعوامل الرئيسية ولكن فى وجوده قد ينتفى تماما معنى الآثار الرئيسية، لذا يجب الحذر عند تفسير العوامل الرئيسية فى وجود التداخل.

تمارين الباب الحادى عشر

١١-١ من البيانات التالية كون واملأ جدول تحليل التباين واختبر الفروض الإحصائية التالية مع كتابة النموذج الرياضى model والنص على الافتراضات علماً بأن مجموع المربعات الكلى الغير مصحح 820 :

أ- لا توجد فروق بين العلائق

ب- لا توجد فروق بين السلالات

ج- لا توجد فروق تداخل بين العلائق والسلالات

	العليقة		
	أ	ب	ج
السلالة I	3	2	0
	2	3	1
	6	2	3
السلالة II	3	-1	2
	6	4	-2
	7	4	4
السلالة III	5	6	7
	7	5	2
	7	7	5
السلالة IV	5	9	1
	6	7	7
	5	0	4

١١-٢ A ، B معاملتان أجريتا على ذكور وإناث الحيوانات بين أى الحالات فيها تداخل بين الجنس والمعاملة فى كل من الحالات الآتية:

أ- متوسط المعاملة A فى الذكور 40 وفى الإناث 40 ومتوسط المعاملة B فى الذكور 20 وفى الإناث 10

ب- متوسط المعاملة A فى الذكور 40 وفى الإناث 30 ومتوسط المعاملة B فى الذكور 20 وفى الإناث 10

ج- متوسط المعاملة A فى الذكور 5 وفى الإناث 10 ومتوسط المعاملة B فى الذكور 10 وفى الإناث 5

د- متوسط المعاملة A فى الذكور 5 وفى الإناث 10 ومتوسط المعاملة B فى الذكور 7 وفى الإناث 5

١١-٣ البيانات التالية تمثل إدرار الحليب اليومى بالكيلوجرام لأبقار من ثلاثة تراكيب وراثية مختلفة تحت نظامين مختلفين للإيواء، المطلوب كتابة النموذج الكامل وتقسيم التباين إلى مصادره المختلفة واختبار فروض العدم الملائمة مع محاولة تفسير النتائج تفسيراً بيولوجياً.

	التركيب الوراثى		
	فريزيان	أبقار خليط	أبقار بلدية
حظائر مسقوفة	23	18	4
	33	18	9
	34	15	4
	40	15	7
	26	12	2
	24	23	12
حظائر مكشوفة	10	4	2
	26	3	7
	13	13	7
	23	14	8
	18	13	4
	8	13	5

١٢ الافتراضات الخاصة بتحليل التباين

Assumptions of analysis of variance

١- عشوائية المعاينة

٢- استقلالية المشاهدات

٣- تجانس التباين

٤- طبيعية التوزيع

٥- التجمعية

٦- تحوير البيانات

كما سبق إيضاحه فى البابين التاسع والحادى عشر وكما سيتضح فيما بعد فى الباب الخامس عشر أنه لإجراء تحليل التباين وتكوين الجدول الخاص به فإن الافتراض الوحيد المطلوب هو أن تكون درجات الحرية بكل مصدر مستقلة خطياً أى linearly independent وهذا ما حققه الشرط الخاص بأن مجموع تأثير المعاملات يساوى صفراً مثلاً، ولهذا الشرط تفقد درجة حرية واحدة، ولكن الأمر يختلف عندما يراد اختبار أى فرض أو فروض خاصة بالعينة المجرى عليها التجربة. والتى عادة ما يستخدم فيها جداول F ، t أو χ^2 ... الخ. فلإجراء اختبار هذه الفروض باستخدام تلك الجداول لابد أن يستوثق المجرب أن هناك عدة شروط تتوافر فى المادة التجريبية. وأهم هذه الشروط هى: عشوائية المعاينة، استقلال المشاهدات، تجانس التباينات، طبيعية التوزيع، والتجمعية. وسوف يتم فى هذا الباب شرح كل من هذه الاشتراطات، كيفية الاختبار لها، كيفية معالجة البيانات إذا لم تتوافر إذا أمكن، وما الذى يمكن أن يحدث لنتائج اختبار الفروض إذا لم تتوافر مثل هذه الافتراضات.

١٢-١ عشوائية المعاينة Random sampling

هذا شرط أساسى لصحة اختبار أى فرض. فالعينة التى سوف يجرى عليها اختبار الفرض لابد أن تكون مأخوذة عشوائياً من عشيرة محددة التعريف والذى سوف يردّ الاستنباط عليها. فإذا أراد المجرب أن يختبر أداء حملان سلالة ما على أربع علائق مثلاً يجب أولاً أن تكون هذه الحملان تمثل عينة عشوائية من حملان هذه السلالة ثم يجب أن توزع هذه الحملان عشوائياً على المعاملات الأربع دون تدخل من المجرب، وأحسن وسيلة لإجراء هذا هى استخدام جداول الأرقام العشوائية (جدول ١ ملحق أ) أو كتابة رقم كل حمل على ورقة ثم خلط هذه الأوراق جيداً ثم سحب كل مجموعة عشوائياً من مجموعة الأوراق أو من خلال الحاسب. ودون هذا الإجراء قد يتسرب إلى التجربة بعض العوامل التى تسبب غياب العشوائية عن غير قصد من المجرب، افترض أن المجرب سيسحب من الحظيرة أول مجموعة من الحملان ليعطيها المعاملة الأولى، وثانى مجموعة ليعطيها المعاملة الثانية ... وهكذا حتى المجموعة المتبقية يعطيها للمعاملة الرابعة. فى مثل هذا الإجراء قد يتسرب كما سبق القول بعض العوامل المحددة للتعشية الكاملة حيث قد تكون المجموعة الأولى من الحملان هى الأهدأ طبعاً وربما الأضعف جسماً وهكذا بالتدريج فى المجاميع حتى تكون المجموعة الأخيرة التى عينت للمعاملة الرابعة هى أكثر الحملان نشاطاً (حيث لم يتمكن من الإمساك بها أولاً) وربما الأكثر حيوية ونشاطاً، وفى هذا تحيز للنتائج، فالفروق بين المعاملات ربما ستحتوى أيضاً على فرق فى الحملان نفسها وهذا طبعاً لم يقصده المجرب. حتى وإن كان ذلك "السيناريو" المذكور مبالغ فيه ولكن لن يكلف المجرب جهداً كبيراً إذا هو أجرى التعشية بالطرق سالفة الذكر.

أيضاً كثيراً ما يتم وضع الحيوانات المولودة، فى محطة تجارب مثلاً، فى قوائم حسب تاريخ ميلادها مثلاً، فإذا أخذ المجرب المجموعة الأولى للمعاملة الأولى والثانية للثانية ... وهكذا، فإنه يدخل عامل العمر بطريقة غير مقصودة، وهذا أيضاً قد يؤدى إلى تحيز نتائج التجربة. وهكذا يمكن سرد عديد من الأمثل المتوقعة وغير المتوقعة التى قد تؤدى إلى تحيز النتائج دون قصد من المجرب. وصمام الأمان فى هذا هو إجراء التعشية إجراءاً سليماً، وليس للتعشية العامة اختبار معين، ولكن يمكن اختبار مثلاً إذا كانت المعاملات موزعة عشوائياً على عامل معين مثلاً كالوزن أو العمر ... الخ. وعدم توافر هذا الشرط يؤدى إلى حالة يكون المجرب غير واثق فيها من إرجاع الظواهر إلى مسبباتها الحقيقية والتى تحيز النتائج.

١٢-٢ استقلالية المشاهدات Independence of observations

المقصود هنا باستقلالية المشاهدات هو استقلال الخطأ فى هذه المشاهدات أى e_{ij} فى النموذج الإحصائى مثلاً $Y_{ij} = \mu + a_i + e_{ij}$ ، فهذه e_{ij} كلها لجميع المشاهدات يجب أن تكون مستقلة عن بعضها ومن توزيعات متطابقة. فإذا فرض أن المجرب لديه أربعة حيوانات فى كل معاملة وكل معاملة وضعها فى حظيرة، وفى كل معاملة يوضع كمية ثابتة من العليقة، والمطلوب اختبار الفرق بين العلائق. ففى هذه الحالة يمكن أن ينشأ وضع هو أنه إذا استهلكت بعض الحيوانات فى المعاملة الواحدة نصيباً أكبر من العليقة، فإن البعض الآخر سيكون نصيبه أقل بالضرورة وينشأ عن هذا ارتباطاً سالباً بين قيم الانحرافات الـــ e_{ij}'s لهذه المجموعة، وفى ذلك حيدة عن الاستقلال وستكون هذه الحيدة عن استقلال الـــ e_{ij} بصورة أكبر إذا ما كانت مساحات أماكن الغذاء (المداود) محدودة الأمر الذى سيؤدى إلى تصارع (تناطح) الحيوانات والأقوى هو الذى سيفوز بنصيب أكبر من العليقة على حساب الأضعف وهو ما يطلق عليه bullying effect. ولهذا يجب أن يفطن المجرب لمثل هذه الأمور مسبقاً، وأن يحاول علاجها مسبقاً طبقاً لإمكانياته كأن يعطى لكل حيوان كمية من العليقة مستقلاً عن بقية الحيوانات إذا سمح المكان بهذا أو يكثر من المساحات المخصصة للتغذية "الطوايل" أو تقدم العليقة مجزأة على عدة مرات إذا سمحت أغراض التجربة بذلك أو قص قرون الحيوانات حتى تقلل من هذا التأثير.

وهناك نوع من ترابط الأخطاء دأب كثير من باحثى الإنتاج الحيوانى على تجاهله وهو الارتباط الموجب بين المشاهدات على مستوى الوقت وذلك عندما تؤخذ المشاهدة (الوحدة التجريبية) على نفس الحيوانات على فترات زمنية. فعدم أخذ هذا الإجراء فى الحسبان عند التخطيط أو التحليل يؤدى إلى ارتباط بين الأخطاء. فالحيوان الذى يقاس بنفسه عدة مرات فى الشتاء ثم فى الربيع ثم فى الصيف ثم فى الخريف، إذا كان تنفسه

سريعاً بالنسبة لبقية الحيوانات فى أحد هذه المواسم، يحتمل أن يكون سريعاً نسبياً أيضاً فى بقية المواسم الآخرى وفى هذا ارتباط بين الأخطاء. ومن هذا أيضاً ظاهرة المعاومة فى الأشجار والتى ينشأ عنها ارتباط سالب ... الخ، وسيعالج هذا الموقف فى الباب الخامس عشر. وقد ينشأ الارتباط أيضاً فى حالة عدم التعشية السليمة. ويمكن أخذ فكرة عن هذا الارتباط من حساب e_{ij} لكل مشاهدة فهذه القيم يجب ألا تأخذ اتجاهاً معيناً فى القيمة أو الإشارة أو التكرار. فمثلاً لا يجب أن تكون هناك عدة قيم سالبة تتبعها أخرى موجبة، أو تكون قيمها متدرجة إلى أعلى أو إلى أسفل إذا رتبت المشاهدات عشوائياً أو تكون منتظمة التكرار ويتوالى السالب منها والموجب فى نظام معين. وهناك اختبارات للكشف عن استقلالية الأخطاء منها اختبار ديربن واطسون Dubrin/Watson statistic. وليس هناك من أسلوب لمعالجة البيانات حتى يتوافر فيها شرط الاستقلال هذا، ولكن الذى يجب عمله هو التعشية السليمة مع اختيار التصميم الإحصائى الملائم وعدم توافر شرط الاستقلالية يؤثر تأثيراً ملحوظاً على اختبار t و F.

١٢-٣ تجانس التباين Homogeneity of variance

والمقصود بالتباين هنا هو تباين الخطأ e فى النموذج الإحصائى، ففى الأبواب السادس والتاسع والعاشر فإن تجانس التباين كان شرط مسبق لإجراء اختبارات الفروض. فإذا كان الهدف هو اختبار الفرق بين معاملتين مثلاً طبقاً للنموذج $Y_{ij} = \mu + a_i + e_{ij}$ فالفرض هنا يعنى أنه إذا قدر التباين داخل المعاملة الأولى بقيمة $\hat{\sigma}^2_{e1}$ وداخل المعاملة الثانية بقيمة $\hat{\sigma}^2_{e2}$ فإن هذين التقديرين هما تقدير لقيمة مشتركة واحدة هى $\hat{\sigma}^2_e$ أى أن الأخطاء فى المعاملة الأولى وتلك التى فى المعاملة الثانية يتبعان (أو مأخوذتان من) نفس العشيرة.

وهناك عدة طرق لاختبار فرض تجانس التباينات أى

$$H_\circ : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma_e^2$$

١٢-٣-١ اختبار بارتلت لتجانس التباين

Bartlett's (1937) test for homogeneity of variance

فى هذا الاختبار تقدر قيمة B كما يلى:

$$B = \frac{2.3026}{C}\{[\sum(n_i - 1)]\log_{10} \overline{S}^2 - \sum(n_i - 1)\log_{10} S_i^2\} \qquad (١٢-١)$$

حيث

$$C = 1 + \frac{1}{3(k-1)}\left[\Sigma\frac{1}{n_i - 1} - \frac{1}{\Sigma(n_i - 1)}\right]$$

S_i^2 : تقدير غير متحيز لتباين (σ_i^2) المعاملة i

$\overline{S}^2$: متوسط التباين

$\log_{10}$: اللوغاريتم للأساس 10، وإذا استخدم اللوغاريتم الطبيعى natural logarithm (الأساس e) فإن عامل الضرب 2.3026 لا يستخدم فى المعادلـــة (١٢-١)

n_i : عدد المشاهدات فى المعاملة i

K : عدد التباينات

والإحصاء B يكون موزعاً حسب χ^2 بدرجات حرية $(k-1)$.

مثال ١٢-١

إذا كانت أوزان الحملان فى ثلاث مجموعات كالتالى:

المجموعة الأولى: 20، 15.2، 19، 13.1، 14، 19.1

المجموعة الثانية: 20.7، 21.7، 16.9، 15.9، 16.2، 21

المجموعة الثالثة: 15.9، 16.7، 13.6، 22.8، 20.6، 17

اختبر الفرض أن التباينات الثلاث لهذه المجموعات متساوية أى:

$$H_o : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$$

يمكن إجراء اختبار بارتلت لتجانس التباين كالتالى:

المجموعة	مجموع المربعات SS	درجات الحرية $(n_i - 1)$	متوسط المربعات S^2	$\log_{10} S^2$	$(n_i - 1)$ x $\log_{10} S_i^2$	$1/(n_i - 1)$
الأولى	44.4333	5	8.8867	0.9487	4.7437	0.2
الثانية	35.6133	5	7.1227	0.8527	4.2632	0.2
الثالثة	55.9333	5	11.1867	1.0487	5.2435	0.2
المجموع	**135.9799**	**15**			**14.2504**	**0.6**

$$\bar{S}^2 = \sum(n_i - 1)\sigma_i^2 = TSS/N = 135.9799/15 = 9.0653$$

$$C = 1 + \frac{1}{(3)(3-1)}\left[0.6 - \frac{1}{15}\right] = 1.0889$$

$$\log_{10}\bar{S}^2 = \log_{10}\left(\frac{135.9799}{15}\right) = \log_{10}(9.0653) = 0.95738$$

$$B = \frac{2.3026}{1.0889}[(15)(0.95738) - 14.2504] = 0.2332$$

والقيمة 0.2332 تتوزع تبعا لتوزيع χ^2 بدرجات حرية 2، وهى غير معنوية حيث إن قيمة χ^2 الجدولية (جدول ٦ ملحق أ) عند درجات حرية 2 ومستوى معنوية 5% تساوى 5.99 وبالتالى لا يمكن رفض فرض العدم، أى أن التباينات متجانسة طبقاً لاختبار بارتلت.

١٢-٣-٢ اختبار F_{max} (Hartley, 1950)

وهذا اختبار بسيط فى إجرائه، ويمكن إجراؤه باستخدام جدول ١٠ ملحق أ حيث إن:

$$F_{max} = \frac{\text{largest variance of k treatments}}{\text{lowest variance of k treatments}} = \frac{\sigma^2 \text{ largest}}{\sigma^2 \text{ lowest}}$$

ويكشف عن قيمة F_{max} عند مستوى معنوية وليكن α ، k عدد المعاملات ودرجات حرية $(n-1)$ حيث n: عدد الملاحظات فى كل معاملة (جدول ١٠ ملحق أ)، فإذا زادت المحسوبة عن الجدولية أو تساوت القيمتان أدى ذلك إلى رفض فرض العدم وتكون التباينات غير متجانسة.

فى مثال ١٢-١

$$F_{max} = \frac{11.1867}{7.1227} = 1.57$$

وهى أقل من القيمة الجدولية 10.8 عند عدد معاملات 3 ودرجات الحرية 5 ومستوى معنوية 5% وعليه فلا يرفض فرض العدم ويستنتج من ذلك تجانس التباينات

وهى نفس النتيجة التى سبق الحصول عليها باختبار Bartlett ، ولكن الأخير أكثر كفاية من الأول. وإذا اختلف عدد المشاهدات فى كل معاملة فيمكن استخدام n للمعاملة الأكثر عدداً، وهذا سيؤدى إلى رفض فرض العدم أكبر مما يجب، أى يجعل استنتاج المجرب أكثر تحفظاً.

١٢-٣-٣ اختبار كوكران Chochran (1941)

وفيه تحسب

$$C = \frac{\sigma^2 \text{ largest}}{\sum_{j=1}^{k} \sigma_j^2}$$

حيث $\sigma^2 \text{ largest}$ هو أعلى تقدير للتباين بين عدد k من المعاملات وأن $\sum_{j=1}^{k} \sigma_j^2$ هو مجموع التباينات كلها وتقارن قيمة C المحسوبة بالقيمة المستخرجة من الجدول عند k، $(n-1)$. ولقيمة C جداول خاصة للكشف عليها (جدول ١١ ملحق أ). وفى مثال ١٢-١ فإن

$$C = \frac{11.1867}{8.8867 + 7.1227 + 11.1867} = 0.4113$$

حيث k = 3، $5 = (n-1)$، وبالكشف فى جدول ١١ ملحق أ يظهر أن القيمة 0.4113 أقل من القيمة الجدولية 0.7071 وعليه لا يرفض فرض العدم ويستنبط من ذلك أن التباينات متجانسة.

وجدير بالذكر أن اختبار F ، t لهما درجة كبيرة من الصمود والصلاحية حتى لو كان هناك تجاوز معقول عن تجانس التباينات، وأن الاختبارات الثلاثة المذكورة أعلاه حساسة لأى انحراف عن فرض تجانس التباينات لذا فإنه ليس من الإجراءات الروتينية أن يختبر المجرب لتجانس التباينات إلا إذا كان هناك انحراف واضح عن التجانس أو أسباب أخرى تدعو المجرب إلى هذا.

أما إذا أوضحت الاختبارات انحرافاً واضحاً عن فرض التجانس فقد يكون هذا مدعاة لأن يحاول المجرب تحوير transformation قياسات التجربة، وهذا الموضوع سيناقش فى نهاية هذا الباب.

١٢-٤ طبيعية التوزيع Normally distributed

ويقصد أن الأخطاء e_{ij}'s موزعة توزيعاً طبيعياً. ويلاحظ أن الحيدة البسيطة عن التوزيع الطبيعى لا يكون لها تأثير كبير على اختبار t ، F ولكن يكون الأثر كبيراً إذا كان الالتواء skewness حاداً، أى أن معظم المشاهدات نحو يمين المنحنى أو نحو يساره. أو لو كان التفرطح kurtosis شديداً. ويمكن التعرف على هذا برسم البيانات كما سبق فى الأبواب السابقة.

كما يمكن التوقع بأن التوزيع لن يكون طبيعياً من طبيعة البيانات نفسها فمثلاً البيانات العددية وليست القياسية غالباً مالا تتوزع حسب التوزيع الطبيعى وكذلك البيانات التى تؤخذ على هيئة نسبة مئوية تبعد قيمها عن 50% كثيراً ... الخ وسيتم مناقشة هذا عند الحديث عن التحويرات transformations.

١٢-٤-١ اختبار الالتواء Skewness

يكون منحنى التوزيع الطبيعي متماثلاً حول منتصفه. ولكن إذا تركزت المشاهدات أكثر إلى يمين المنحنى يطلق عليه ملتوى إلى اليسار وإذا كانت إلى اليسار يطلق عليه ملتوى إلى اليمين (شكل ١٢-١).

شكل ١٢-١ التوزيع الملتوى إلى اليسار والتوزيع الملتوى إلى اليمين

ويقال أن التوزيع التكرارى موجب الالتواء إذا كان المتوسط أكبر من المنوال، ويقال أن التوزيع سالب الالتواء إذا كان المتوسط أقل من المنوال.

ويقدر معامل الالتواء بالمعادلة التالية:

$$\text{skew} = \frac{\text{mean - mode}}{\text{standard deviation}}$$

وحيث إنه أحيانا يصعب تقدير الــ mode وإن

$$\text{mean} - \text{mod e} = 3(\text{mean} - \text{median})$$

فإن

$$\text{skew} = \frac{3(\text{mean} - \text{mode})}{\text{standard deviation}}$$

والمثال التالى يبين كيفية حساب واختبار الالتواء باستخدام العزوم حول المتوسط.

مثال ١٢-٢

جدول ١٢-١ يمثل الدرجة التى حصل عليها الطلبة فى احدى المواد الدراسية عدد 186 طالب، هل توزيع هذه الصفة فى هذه العينة يتوزع حسب التوزيع الطبيعى؟

جدول ١٢-١ درجات الطلبة وتكراراتها فى إحدى المواد الدراسية

u^4	u^3	u^2	u	التكرار f	الحد الأدنى للقسم
625	125	25	5	6	5
2401	343	49	7	4	7
6561	729	81	9	7	9
14641	1331	121	11	9	11
28561	2197	169	13	21	13
50625	3375	225	15	23	15
83521	4913	289	17	23	17
130321	6859	361	19	33	19
194481	9261	441	21	19	21
279841	12167	529	23	15	23
390625	15625	625	25	17	25
531441	19683	729	27	5	27
707281	24389	841	29	2	29
923521	29791	961	31	2	31

اختبار الالتواء:

$$h_1 = \frac{\Sigma fu}{n} = 17.634 \qquad \sum fu = 3280$$

$$h_2 = \frac{\Sigma fu^2}{n} = 341.258 \qquad \sum fu^2 = 63474$$

$$h_3 = \frac{\Sigma fu^3}{n} = 7060.086 \qquad \sum fu^3 = 1313176$$

$$h_4 = \frac{\Sigma fu^4}{n} = 153919.194 \qquad \sum fu^4 = 28628970$$

$$m_2 = \frac{\Sigma(X-\overline{X})^2}{n} = h_2 - h_1^2 = 30.30$$

$$m_3 = \frac{\Sigma(X-\overline{x})^3}{n} = h_3 - 3h_1h_2 + 2h_1^3 = -26.2795$$

$$\sqrt{b_1} = m_3 / m_2\sqrt{m_2} = -26.2795 / 30.30\sqrt{30.30} = -0.1576$$

وفى العينات التى تتبع التوزيع الطبيعى فإن الكمية $\sqrt{b_1}$ تتوزع حسب التوزيع الطبيعى بمتوسط صفر وانحراف معيارى قدره $\sqrt{6/n}$ أى أن الانحراف المعيارى هنا $\sqrt{6/186} = 0.18$.

والقيمة المقدرة $\sqrt{b_1}$ هى (0.158 -) وهى أقل من الانحراف المعيارى وعليه لا يرفض الفرض: أن هذه العينة الالتواء فيها منعدم. وتدل إشارة b على اتجاه الالتواء سالباً كان أم موجباً، فالسالب يدل على تراكم المشاهدات نحو القيم المرتفعة من التوزيع، والموجب يدل على تراكم المشاهدات نحو القيم المنخفضة.

وجدير بالذكر أن ما ذكر من أن الكمية $\sqrt{b_1}$ تتوزع طبيعياً بمتوسط صفر وانحراف معيارى $\sqrt{6/n}$، وهذا الافتراض قريب من الصحة عندما تكون n قدرها 150 أو أكثر. وللعينات التى يتراوح عددها بين 25 إلى 200 فإن جدول ١٢ ملحق أ يعطى مستويات المعنوية 1%، 5% فى جهة واحدة one tail. ففى المثال السابق فإن $b_1 = -0.158$ أقل من تلك التى عند 1%، 5% . ولعينة مكونة من 200 فرد فإن القيمة الحرجة عند 1%، 5% هى 0.280, 0.403 على التوالى، وعليه لا يرفض فرض العدم بأن التوزيع طبيعى وهو مطابق لنفس الاستنتاج السابق.

ويتبين من جدول ١٢-١ أن الأرقام قد تصل إلى كميات كبيرة جداً خاصة تحت عمودى u^3 ، u^4 ويمكن اختزال قيمة u إما بطرح ثابت أو القسمة على ثابت أو الاثنين معاً وذلك لأن الكمية $\sqrt{b_1}$ وكذلك مقياس التفرطح (كما سيأتى فيما بعد) مستقلان عن وحدة القياس وأى اختزال لا يؤثر فى حساباتهما.

١٢-٤-٢ اختبار التفرطح Test of kurtosis

كما قيس الالتواء بواسطة متوسط القيمة $(y-\mu)^3$ فى العشيرة وهى المسماة بالعزم الثالث، فإن التفرطح يقاس بالعزم الرابع $(y-\mu)^4$ مقسوماً على مربع التباين σ^4. وإذا كانت العينة تتبع حقاً التوزيع الطبيعى فإن القمة المتوقعة لهذه النسبة = 3، فإذا كانت أكبر من 3 دل هذا على أن الشكل الناقوسى مدبب القمة أكبر مما يجب وإن قل عن 3 دل على أن الناقوس مستوى القمة أكثر مما يجب.

وبالعودة إلى مثال ١٢-٢ وجدول ١٢-١ يمكن حساب ما يلى:

$$m_4 = h_4 - 4h_1h_3 + 6h_1^2h_2 - 3h_1^4 = 2545.7576$$

$$b_2 = m_4 / m_2^2 = 2.77288$$

$$g_2 = b_2 - 3 = 2.77288 - 3 = -0.2271$$

تتوزع الكمية g_2 تقريباً حسب التوزيع الطبيعى بمتوسط قدره صفر وانحراف معيارى قدره $\sqrt{24/n}$ أى $\sqrt{24/186} = 0.359$. الانحراف الذى قدره (0.2271 -) لا يختلف معنوياً عن الصفر عند مستوى 5% وعليه يقبل الفرض القائل أن $g_2 = 0$، أى أن التفرطح هو ذلك المفروض أن يكون فى حالة التوزيع الطبيعى. وكون قيمة انحراف التفرطح سالبة (أى أقل من 3) فإن ذلك يدل على أن منحنى العينة مستوى القمة قليلاً ولكن ليس أكثر مما يجب تحت فرض 5% فى حالة التوزيع الطبيعى.

ويعتبر إجراء الاختبار على g_2 مجاز حيث إنه يتطلب أن تكون n أكبر من 1000 حتى يقترب توزيع g_2 من التوزيع الطبيعى. وجدول ١٣ ملحق أ يعطى قيماً مقربة أدق لكل من مستويات المعنوية 1%، 5% عندما يكون حجم العينة بين 200، 1000، وحيث إن توزيع g_2 ملتوى فإن التوزيع معطى لكل من جهتى التوزيع مستقلاً. ففى المثال السابق حيث كانت $b_2 = 2.77288$ وحجم العينة 186 فإن b_2 أعلى من القيمة الحرجة الدنيا فى الجدول (القيمة الحرجة عند حجم العينة 200 هى 2.37، 2.51 عند 1%، 5% على التوالى)، وعليه لا يرفض فرض التوزيع الطبيعى وهو نفس الاستنتاج السابق.

أما إذا قل عدد العينة عن 200 فإن Geary (1936) استنبط جداول بها القيم الحرجة للاختبارات حتى عدد العينة 11 وهى توافق إلى حد كبير قيم g_2 المشروحة سابقاً (انظر أيضاً Snedecor and Chochran, 1987).

مثال ١٢-٣

حل مثال ١٢-٢ باستخدام برنامج SAS لإجراء اختبار الالتواء والتفرطح

```
DATA GRADES;
INPUT GRADE REP @@;
CARDS;
5 6 7 4 9 7 11 9 13 21 15 23 17 23 19 33
21 19 23 15 25 17 27 5 29 2 31 2
PROC MEANS N MEAN STDERR SKEWNESS KURTOSIS T PRT;
VAR GRADE;
FREQ REP;
RUN;
* ---  USING UNIVARIATE PROCEDURE  ----;
PROC UNIVARIATE;
VAR GRADE;
FREQ REP;
RUN;
```

لاحظ

يمكن إجراء اختبار الالتواء والتفرطح إما باستخدام اختيار PROC MEANS أو اختيار PROC UNIVARIATE.

نتائج التحليل

The MEANS Procedure

Analysis Variable : GRADE

N	Mean	Std Error	Skewness	Kurtosis	t Value	Pr > |t|
186	17.6344086	0.4046066	-0.1568986	-0.2075433	43.58	<.0001

The UNIVARIATE Procedure

Variable: GRADE

Freq: REP

Moments

N	186	Sum Weights	186
Mean	17.6344086	Sum Observations	3280
Std Deviation	5.51809788	Variance	30.4494042
Skewness	-0.1568986	Kurtosis	-0.2075433
Uncorrected SS	63474	Corrected SS	5633.13978
Coeff Variation	31.2916526	Std Error Mean	0.40460657

١٢-٥ التجمعية Additivity

وهذه الخاصية أو الشرط قد يأخذ أكثر من معنى. ففى حالة تحليل التباين ذى الاتجاهين بدون تكرار المشاهدة كما فى القطاعات العشوائية مثلاً وإذا كانت المعاملات ثابتة (نموذج ١) فإنه يجب فرض غياب التداخل بين المعاملة والقطاعات حتى يتسنى اختبارات الفروض الخاصة بالمعاملة باستخدام الخطأ من تحليل التباين. معنى ذلك أن أثر المعاملة يجمع على أثر القطاع ليحدثا أثرهما على المشاهدة. ولكن إذا ضرب الاثنان مثلاً فتكون خاصية التجمعية قد فقدت. وإذا وجد التداخل فعلاً وتم تجاهله ولم تجرى الاختبارات اللازمة لمعرفة مدى أهميته، فإن اختبار الفروض الخاصة بالمعاملات يصبح غير كفء واختباراً ضعيفاً. ولاختبار وجود مثل هذا التداخل فى حالة عدم وجود تكرار للمشاهدة داخل كل معاملة وكل قطاع سيتم مناقشته عند مناقشة تصميم القطاعات العشوائية. أما إذا تكررت المشاهدة فإنه يمكن اختبار خاصية التجمعية هذه لأنه سوف يكون هناك نوعان من الخطأ أحدهما يطلق عليه خطأ تجريبى experimental error وهو فى حقيقته التداخل بين القطاعات والمعاملات وآخر يطلق عليه خطأ عينى sampling error وهو مقياس للتباين بين مشاهدات فى نفس القطاع ونفس المعاملة. وناتج قسمة متوسط مربعات الخطأ التجريبى على متوسط مربعات الخطأ العينى يعطى F لاختبار فرض عدم وجود تداخل بين القطاعات والمعاملات، وهذا أيضاً سيأتى شرحه تفصيلاً عند تقديم تصميم القطاعات العشوائية.

وفى التجارب الحيوية أو الزراعية يمكن أن ينتج التداخل من كثير من العلاقات بين المعاملات وطريقة تأثيرها على الكائن موضوع التجربة مثل ظاهرتا التعاضد synergism أو التعارض interference ويوجد أيضاً فى الوراثة كما هو معروف بالتفوق بين الجينات epistasis والمثال الموضح فى جدول ١٢-٢ يبين هذا التداخل.

ويتضح من جدول ١٢-٢ أن المشاهدة $a_1b_1 = 2$ عندما يكون سلوك المعاملات تجمعى وهى حاصل جمع $a_1 + b_1$ ... وهكذا لبقية المشاهدات. وفى هذه الحالة فإن الفرق بين أى عمودين ثابت فى أى صف وكذلك الفرق بين أى صفين ثابت فى كل عمود.

بينما إذا كان السلوك ضربياً فإن المشاهدة a_3b_2 مثلاً $15 = (5)(3)$، ويكون الفرق بين أى صفين غير متساو فى الأعمدة، وكذلك الفرق بين الأعمدة فى الصفوف المختلفة. وإذا علم الباحث أن مثل هذا السلوك سائد بين بيانات التجربة فيمكن علاج الأمر وتحوير البيانات، وذلك بأخذ لوغاريتمات القيم، الأمر الذى يوفر خاصية التجمعية كما هو واضح من الجدول السابق.

جدول ١٢-٢ مثال للتداخل بين العوامل وتوافر خاصية التجميعية

العامل B		العامل A			ملاحظات
		$a_1 = 1$	$a_2 = 2$	$a_3 = 3$	
$b_1 = 1$	**تجمعى**	2.0	3.0	4.00	آثار تجميعية
	ضربى	1.0	2.0	3.00	آثار مضروبة
	لوغاريتم الضربى	0.0	0.3	0.48	تحوير لوغاريتمى
$b_2 = 5$	**تجمعى**	6.0	7.0	8.00	آثار تجميعية
	ضربى	5.0	10.0	15.00	آثار مضروبة
	لوغاريتم الضربى	0.7	1.0	1.18	تحوير لوغاريتمى

وفى حالة تحليل التباين ذى الاتجاه الواحد والذى يفترض فيه النموذج

$$Y_{ij} = \mu + t_i + e_{ij}$$

حيث μ هى المتوسط و t_i أثر المعاملة و e_{ij} الخطأ، يمكن أن يحدث ألا تتوافر خاصية التجمعية كأن يكون النموذج الممثل للواقع مثلاً هو:

$$Y_{ij} = \mu \ t_i \ e_{ij}$$

أى مكونات النموذج مضروبة فى بعضها بدل أن تكون مجموعة على بعضها، وينتج عن هذا أيضاً عدم تجانس التباين.

ويمكن تلافى هذا الفرق للاشتراطات اللازمة لتحليل التباين، وذلك بإجراء التحوير المناسب كما هو مبين فى جدول ١٢-٣.

جدول ١٢-٣ مثال لنموذج فى اتجاه واحد مضروب العوامل والتحوير المناسب بفرض أن $\mu = 1$

	المشاهدات الفعلية			تحوير لوغاريتمى		
	$t_1 = 2$	$t_2 = 4$	$t_3 = 6$	t_1	t_2	t_3
$e_{i1} = 1$	2	4	6	0.3010	0.6021	0.7782
$e_{i2} = 2$	4	8	12	0.6021	0.9031	1.0792
$e_{i3} = 3$	6	12	18	0.7782	1.0792	1.2553
$e_{i4} = 4$	8	16	24	0.9031	1.2041	1.3802
S^2	6.67	26.67	60	0.0682	0.0682	0.0682

وواضح من جدول ١٢-٣ أنه إذا حللت النتائج بفرض نموذج تجميعى فإن هذا سوف يؤدى إلى تباينات غير متجانسة داخل المعاملات تراوحت قيمتها من 6.67 إلى 60، بينما إذا أمكن التعرف على أن النموذج هو ضربى وعليه إجراء التحوير المناسب الذى تنتج عنه تباينات متجانسة تماماً، هذا بالطبع لأن المثال توضيحى بحت. وجدير بالذكر أن كثيراً من اشتراطات تحليل التباين مرتبطة ببعضها بمعنى أنه فى المثال السابق وجد أن البيانات الأصلية بياناتها مختلفة، وأن الخطأ مرتبط بقيمة المشاهدة، وأن خاصية التجمعية غير متوفرة، وبعمل التحوير المناسب تم علاج هذه النواحى الثلاث.

١٢-٦ تحوير البيانات Data transformations

المقصود بالتحوير هو إجراء نفس التغيير على كل البيانات الأولية بحيث تغير بعض خصائص البيانات بينما تستبقى الخصائص الأخرى كما هى. ويجرى التحوير عند القيام بتحليل التباين للأغراض الرئيسية التالية:

١- توفير تجانس التباينات؛

٢- توفير طبيعية التوزيع؛

٣- توفير خاصية التجمعية.

وكما سبق ذكره أن هناك عدة اشتراطات يجب توافرها حتى يمكن اختبار الفروض، وفى حالة عدم توافر بعض أو كل هذه الفروض فإنه ينصح أحياناً بإجراء تحوير فى البيانات أملا فى التوصل إلى شكل محدد من البيانات يتوافر فيه الاشتراطات المطلوبة ولكن يظل محتفظاً بخصائص البيانات من فروق بين المعاملات

والاتجاهات الأخرى المطلوب دراستها. ولأن اختبار F أقل تأثراً بعدم توافر شرطى طبيعية التوزيع وتجانس البيانات، فإن إجراء التحوير للغرض الثالث هو أدعى من إجرائه للغرضين الأولين. والغرض الثالث هذا هام جداً توافره خصوصاً فى بعض التصميمات الإحصائية التى قد لا يتوافر فيها تكرار المشاهدة مثل تصميم القطاعات العشوائية randomized block.

وكما سبق القول، فإن التحوير ممكن أن يعالج أكثر من نقص فى آن واحد، كما فى جدول ١٢-٣، وإن كان من الممكن إجراء التحوير المناسب لمعالجة بعض النواقص كما فى حالة ارتباط التباين بمتوسط المعاملة، إلا أنه ليست القاعدة أن يجد المجرب التحوير المناسب فى كل الحالات. ويذكر (Kirk, 1995) أنه لا يوجد تحوير مناسب إذا وجدت أى من الحالات التالية:

١- متوسط المعاملات متساوية تقريباً ولكن التباينات غير متجانسة.

٢- متوسطات المعاملات تختلف بطريقة مستقلة عن التباينات.

٣- تجانس التباينات ولكن شكل توزيع متوسطات المعاملات غير متجانس.

ونظرياً إذا علم السلوك الرياضى للمشاهدات تماماً، فإنه يمكن إيجاد التحوير المناسب ولكن قلما يتوافر مثل هذا الوضع فى الحياة العملية. وهناك عدة طرق لمحاولة إيجاد التحوير المناسب ولكن سيتم سرد التحويرات التى وجدت مناسبة فى حالات معينة. وفى حالة عدم إمكان التوصل إلى التحوير المناسب يمكن إتباع طرق تحليلية غير معتمدة على توزيع معين وهى ما يطلق عليها الإحصاءات اللامعلمية nonparametric statistics أو الإحصاءات حرة التوزيع distribution-free statistics وهى تخصص أبعد من مجال هذا المؤلف.

١٢-٦-١ تحوير الجذر التربيعى Square root transformation

يتناسب فى بعض أنواع البيانات التباين مع متوسط المعاملة كما هو الحال في توزيع بواسون حيث $\mu = \sigma^2$. وينشأ مثل هذا التوزيع عندما يكون المتغير التابع (أو المشاهدة) ناتج عن عد أفراد أو أشياء أو أحداث والتى احتمال حدوثها بسيط مثل عدد بويضات طفيل معين فى أمعاء حيوان معالج ضد الطفيليات أو عدد الحيوانات المنوية الشاذة فى سائل منوى محفوظ أو عدد خلايا معينة فى الدم ... الخ. فى مثل هذه الحالات يحور كل رقم إلى الجذر التربيعى له، وإن كانت هناك أصفار فيكون التحوير باستخدام $Y = \sqrt{Y + 0.5}$ ثم يجرى التحليل واختبارات الفروض على البيانات المحورة. وترد متوسطات المعاملات وانحرافاتها القياسية فقط إلى وحداتها الأصلية بإحداث تحوير عكسى.

وجدول ١٢-٤ يبين هذا التحوير لعدد الحيوانات المنوية الشاذة فى مساحة معينة من الشريحة تحت الميكروسكوب لثلاثة مخففات.

جدول ١٢-٤ تأثير ثلاثة مخففات على عدد من الحيوانات المنوية الشاذة فى مساحة محددة من الشريحة

	البيانات الأصلية y			البيانات المحورة $y = \sqrt{y + 0.5}$		
	المخفف			المخفف		
	الأول	الثانى	الثالث	الأول	الثانى	الثالث
	6	12	23	2.55	3.54	4.85
	0	8	12	0.71	2.92	3.54
	8	4	12	2.92	2.12	3.54
	4	8	20	2.12	2.92	4.53
	4	14	12	2.12	3.81	3.54
المتوسط	**4.4**	**9.2**	**15.8**	**2.084**	**3.06**	**4.00**
S^2	**8.80**	**15.20**	**28.2**	**0.7016**	**0.4265**	**0.4095**

المتوسطات محسوبة من البيانات المحورة ثم مردودة إلى وحداتها الأصلية، أى يربع المتوسط ثم يطرح منه 0.5، هى 3.84، 8.86، 15.5 على التوالى).

وواضح من هذا المثال أن أعلى تباين كان أعلى أكثر من ثلاث مرات أى $28.2/8.8 = 3.2$ قدر أقل تباين قبل التحوير بينما أصبح أعلى تباين أقل من ضعف أقل تباين أى $0.7016/0.4095 = 1.7$.

١٢-٦-٢ التحوير اللوغاريتمى Logarithmic transformation

يستخدم هذا التحوير عندما يكون هناك تناسب بين الانحرافات المعيارية والمتوسطات، ويجرى التحوير كما يلى:

$$y' = \log_{10} Y$$

أو $y' = \log_{10}(Y+1)$ عندما يكون هناك قيم قدرها صفراً أو صغيرة جداً.

ويفيد هذا التحوير إذا ما كانت معاملة معينة مثلاً تزيد عن معاملة أخرى بنسبة ثابتة، وليس بقدر ثابت فمثلاً فى جدول ١٢-٢ القيمة a_2 دائماً ضعف القيمة a_1 (الفرق بينهما 1 تحت b_1، 5 تحت b_2) وأن a_3 ثلاثة أضعاف a_1 و50% أعلى من a_2 (الفرق 2، 1 تحت b_1 على التوالى ، 10، 5 تحت b_2 على التوالى). وذلك ينطبق تحت b1, b2. لذا فإن التحوير اللوغاريتمى حول الزيادة بالنسبة إلى زيادة بالإضافة فأصبح الفرق بين a_2 و a_1 هو 1 فى كل الحالات ... وهكذا. وكذلك فى جدول ١٢-٣ فالمشاهدات الفعلية فى t_3 كانت دائماً ثلاث مرات قدر t_1، 50% أعلى من t_2 ولكن بعد التحوير اللوغاريتمى أصبح هذا الفرق 0.311، وهكذا فى جميع الأحوال.

ويستخدم التحوير اللوغاريتمى أيضاً عندما يكون المتغير التابع هو زمن الاستجابة وتكون البيانات ملتوية إلى يمين التوزيع. مثل فترة الاستجابة reaction time وهى الفترة ما بين إدخال الذكر إلى الإناث والتلقيح.

١٢-٦-٣ التحوير بالمقلوب Inverse transformation

ويستخدم فى مثل حالات التحوير اللوغاريتمى عندما يكون هناك تناسب بين متوسطات المعاملات وانحرافاتها المعيارية، ويكون التحوير $y' = 1/Y$. وعندما تكون إحدى قيم Y مساوية للصفر فإن $y' = 1/(Y+1)$.

١٢-٦-٤ تحوير مقلوب جيب الزاوية Arcsin or angular transformation

يفيد هذا التحوير عندما يكون المتوسط والتباين متناسبين وأن التوزيع يتبع "ذو الحدين". وتتحقق مثل هذه الظروف عندما يكون العدد ثابتاً مثلاً والمتغير هو عدد مرات النجاح من العدد الكلى. كأن يعطى للذكر عدد ثابت من الإناث ثم يسجل عدد الإناث المخصبة للمقارنة بين الذكور المختلفة. كذلك يفيد هذا التحوير عندما تكون البيانات مسجلة على هيئة نسب مئوية وخصوصاً إذا كانت أقل من 30% أو أعلى من 70% مثل النسبة المئوية للرطوبة فى اللحم أو نسبة الدهن فى الذبيحة ... الخ. أما نسبة التصافى فى الذبيحة مثلاً فهى لا تحتاج عادة إلى تحوير، لأن النسبة المئوية لها لا تتخطى الحدين 30, 70% ويجرى هذا التحوير عن طريق $y' = 2 \arcsin \sqrt{y}$ حيث y معبر عنها كنسبة. ويعطى جدول ١٤ ملحق أ القيم المحورة من 0.001 وحتى 0.999 ويقترح أن تعطى $1/2n$ أو $1/4n$ فى حالة y تساوى صفر، $(1-1/2n)$ فى حالة y = 1، حيث n هى عدد المشاهدات عندما تكون n أقل من 50.

مثال ١٢-٤

البيانات التالية تمثل نسبة مئوية للأنسجة الدهنية إلى الوزن الكلى لإحدى القطعيات فى ذبيحة الأغنام تحت ثلاثة أنظمة غذائية محتلفة.

جدول ١٢-٥ نسبة الأنسجة الدهنية فى إحدى القطعيات بذبيحة الأغنام فى ثلاثة معاملات

	البيانات الأصلية			البيانات المحورة		
	معاملة ١	**معاملة ٢**	**معاملة ٣**	**معاملة ١**	**معاملة ٢**	**معاملة ٣**
	0.10	0.14	0.28	0.6435	0.7670	1.1152
	0.12	0.10	0.33	0.7075	0.6435	1.2239
	0.09	0.18	0.35	0.6094	0.8763	1.2661
	0.11	0.16	0.28	0.6761	0.8230	1.1152
	0.11	0.15	0.23	0.6761	0.7954	1.0004
	0.15	0.12	0.23	0.7954	0.7075	1.0004
المتوسط	**0.113**	**0.142**	**0.283**	**0.6847**	**0.7688**	**1.202**
S^2	**0.0004**	**0.0008**	**0.0025**	**0.0041**	**0.0069**	**0.0122**

ويتبين من جدول ١٢-٥ أن $F_{max} = 0.0025/0.0004 = 6.25$ وهى معنوية وتدل على عدم تجانس التباينات.

بينما بعد تحوير البيانات $F_{max} = 0.0122/0.0041 = 3.0$ وهى غير معنوية.

مثال ١٢-٥

حل مثال ١٢-٤ باستخدام برنامج SAS

```
DATA TRANS;
INPUT TRT FAT @@;
TFAT = 2*ARSIN(SQRT(FAT));
CARDS;
1 0.1 1 0.12 1 0.11 1 0.11 1 0.15
2 0.14 2 0.1 2 0.18 2 0.16 2 0.15 2 0.12
3 0.28 3 0.33 3 0.35 3 0.28 3 0.23 3 0.23
PROC MEANS MEAN VAR;
```

```
VAR FAT TFAT;
BY TRT;
RUN;
```

نتائج التحليل

The MEANS Procedure

--- trt=1 ---

Variable	Mean	Variance
FAT	0.1180000	0.000370000
TFAT	0.6997288	0.0033720

--- trt=2 ---

Variable	Mean	Variance
FAT	0.1416667	0.000816667
TFAT	0.7687848	0.0069335

--- trt=3 ---

Variable	Mean	Variance
FAT	0.2833333	0.0024667
TFAT	1.1201828	0.0121622

تمارين الباب الثانى عشر

١٢-١ فى إحدى التجارب لقياس عدد بيض الديدان الاسطوانية فى أمعاء أربع سلالات من الأغنام وجدت النتائج التالية علماً بأن كل سلالة ممثلة بعدد عشر تقديرات للبيض.

السلالة	متوسط عدد البيض	الانحراف المعيارى S
١	2250	210
٢	12100	4051
٣	8650	2025
٤	9160	2516

أ- افحص العلاقة بين المتوسط والانحراف المعياري

ب- اختبر ما إذا كانت التباينات متجانسة

١٢-٢ البيانات التالية لطول فترة الحمل فى الجاموس المصرى، افحص ما إذا كانت هذه العينة تتبع التوزيع الطبيعى باستخدام اختبار الالتواء والتفرطح.

الحد الأدنى للقسم (يوم)	التكرار
296	10
301	18
306	71
311	131
316	156
321	99
326	24
331	6
إجمالى	**515**

١٢-٣ التالى هو عدد البكتيريا فى ١ سم^٣ من اللبن لثلاث بقرات فى ثلاث فترات مختلفة

رقم البقرة	عند الحلب	بعد ٢٤ ساعة من الحلب	بعد ٤٨ ساعة من الحلب
١	12000	14000	57000
٢	13000	20000	65000
٣	21500	31000	106000

١- احسب المتوسط والتباين لكل من الثلاث فترات وافحص العلاقة بين التباين والمتوسط. حور البيانات لوغاريتمياً ثم أعد حساب التباين والمتوسط وقارن بين الحالتين، ثم ناقش النتائج.

٢- حلل التباين على كل من البيانات الأصلية وتلك المحورة وقارن النتائج.

١٣ تحليل التباين ثنائى الاتجاه
فى حالة عدم تساوى تكرار الفئات

Two-way analysis of variance
in case of unequal subclass frequencies

١- مقدمة

٢- الطرق التقريبية لتحليل التباين فى حالة عدم تساوى تكرار الفئات

٣- طريقة المتوسطات غير الموزونة

٤- طريقة الأعداد المتوقعة للفئات

٥- الطرق المضبوطة لتحليل التباين فى حالة عدم تساوى الأعداد فى الفئات

٦- طريقة موائمة الثوابت للتأثيرات الثابتة للعوامل

٧- طريقة المربعات الموزونة للمتوسطات

٨- تنويه عن تحليل البيانات غير متناسبة التكرار

١٣-١ مقدمة

فى كثير من الحالات، يجد الباحث نفسه أمام حالات تختلف فيها أعداد المشاهدات في الفئات subclasses المختلفة المقسمة إليها البيانات ويظهر ذلك فى حالة البيانات المجمعة من السجلات، وبدرجة أقل فى حالة التجارب المخطط لها. فمثلاً فى البيانات المجمعة لدراسة ناتج الحليب من سلالة غالبا ما تكون أعداد الأبقار فى مراحل العمر وعدد مواسم الحليب تختلف من فئة إلى أخرى وأيضاً تختلف الأعداد فى المحطات المختلفة. وفى دراسات الأغنام والمعز تختلف الأعداد فى الجنسين ويختلف عدد التوائم عن الحملان المفردة ... الخ.

١٣-٢ الطرق التقريبية لتحليل التباين فى حالة عدم تساوى تكرار الفئات[١]

فى حالات كثيرة من وجود الأعداد غير المتساوية وغير المتزنة (غير المتناسبة) يرغب المجرب فى تحليل البيانات بطريقة سريعة تعطى دلالات عن تأثير العوامل المأخوذة فى الاعتبار فى النموذج الرياضى ولو أنها ليست تامة الدقة ولكنها تقريبية إلى حد ما. هذه الطرق تتميز بأنها أقل كثيراً فى التعقيد من ناحية طرق الحساب المستخدمة والتى قد تستلزم معرفة ببعض نواحى الرياضيات التى قد لا تتوافر للبعض، وفى نفس الوقت تعطى دلالة كافية عن تأثير العوامل المختلفة على مصادر الاختلافات الناشئة فى البيانات، أو أن هذه الطرق قد يتم استخدامها مبدئياً للحصول على الاتجاهات trends المعينة للبيانات انتظاراً لاستخدام الطرق الأخرى الأكثر تعقيداً فى مراحل تالية من التجربة أو عند توافر جميع البيانات، ويلاحظ أنه فى بعض الحالات قد يستوجب إجراء طرق التحليل الأكثر تعقيداً باستخدام الحاسبات الإلكترونية مما لا يتوافر فى بعض الحالات.

وهناك طريقتان تقريبيتان يمكن استخدامها فى حالة تحليل البيانات غير المتزنة وغير المتساوية وهما:

أ- طريقة المتوسطات غير الموزونة Unweighted means

ب-طريقة الأعداد الفئوية المتوقعة Expected subclass numbers

[١] يعتبر هذا فصل - إلى حد بعيد - تاريخياً، حيث يبين تطور هذا النوع من تحليل البيانات عندما لم يكن هناك حواسب إلكترونية ولا حتى حاسبات متقدمة. ولكنه يبين بوضوح علاقة التكرار فى الفئات باستقلال المشاهدات وتهاوى نظرية تساوى مجموع المربعات عندما تكون التكرارات غير متساوية وغير متناسبة.

وكلا الطريقتين تستخدمان نظرية التجميع الخاصة بتحليل التباين addition theorem وهى الخاصية التى تتمتع بها البيانات المتزنة (المستقلة)، وبالتالى، فإن الحسابات لمجاميع المربعات المختلفة تصبح أقل تعقيداً.

وهناك ملاحظة فى هذه النظم وهى، أنها لا تصلح عندما يكون عدد الأفراد فى أى من الفئات يساوى صفراً وسيتضح ذلك فيما بعد.

١٣-٣ طريقة المتوسطات غير الموزونة Method of unweighted means

وتعطى هذه الطريقة نتائج دقيقة إلى حد كبير لتحليل التباين إذا كانت الاختلافات بين عدد المشاهدات فى الفئات المختلفة ضيقة أى فى حدود 2:1 أو 3 ويتناقص مدى دقة precision النتائج كلما ازدادت التباينات بين أعداد المشاهدات فى الفئات المختلفة subclasses.

وطريقة المتوسطات غير الموزونة هى عبارة عن تحليل تباين لمتوسطات الفئات وبالتالى، فهى عبارة عن تحليل تباين لبيانات متساوية، حيث لكل فئة متوسط واحد، وعلى ذلك، فإن تقسيم مجموع المربعات الكلى total sum of squares إلى أجزاء خاصة بالتأثيرات الرئيسية (الأساسية) main effects وتأثير التداخل بينهما interaction يعتبر مشروعاً حيث أن نظرية التجميع لمجموع المربعات تكون صحيحة. ويلاحظ فى هذه الطريقة أن مجاميع المربعات للتأثيرات الأساسية والمحسوبة على أساس المتوسطات تختلف عن مجموع المربعات الداخلى (الخطأ) within sum of squares (error) والذى يكون محسوباً على أساس البيانات الفردية individual observations ويتم حساب مجموع المربعات للخطأ من داخل الفئات على أساس انحراف قيم المشاهدات داخل كل فئة عن متوسط هذه الفئة $\sum_i\sum_j\sum_k(Y_{ijk}-\overline{Y}_{ij.})^2$ والطريقة التى تستعمل حسابياً هى أن يحسب مجموع المربعات الكلى الغير مصحح:

$$\text{uncorrected total sum of quares} = \sum_i\sum_j\sum_k Y_{ijk}^2 \qquad (١٣-١)$$

ويحسب مجموع المربعات بين الفئات غير المصحح:

$$\text{uncorrected between subclasses sum of quares} = \sum_i\sum_j Y_{ij.}^2 \qquad (١٣-٢)$$

حيث $\sum_k Y_{ijk} = Y_{ij.}$

وبالتالى فإنه بطرح الكمية الثانية من الأولى تعطى مجموع المربعات داخل الفئات، أى

$$\sum_i\sum_j\sum_k Y_{ijk}^2 - \sum_i\sum_j Y_{ij.}^2$$

بمعنى أن مجموع مربعات الخطأ = مجموع المربعات الكلى – مجموع المربعات بين الفئات.

$$\text{Within SS} = \text{Total SS} - \text{Subclasses SS} \qquad (١٣-٣)$$

مما تقدم يتضح أن، متوسط المربعات داخل الفئات يتم حسابه بقسمة مجموع المربعات داخل الفئات على درجات الحرية الخاصة به والذى يساوى $(N - rc)$ حيث r تمثل عدد الصفوف (أحد التأثيرات الرئيسية)، c تمثل عدد الأعمدة (التأثير الرئيسى الآخر فى البيانات ثنائية التقسيم)، وهو بالتالى غير صحيح لاستخدامه فى اختبار متوسطات المربعات الخاصة بالتأثيرات الرئيسية أو التداخل المحسوبة على أساس متوسطات الفئات وليس المشاهدات الفردية.

ولتصحيح ذلك إما أن يصحح ليصبح متوسطات المربعات داخل الفئات محسوباً على أساس المتوسطات أو تصحيح متوسطات المربعات الخاصة بالتأثيرات الرئيسية والتداخل لتصبح محسوبة على أساس المشاهدات الفردية. وبما أن هناك متوسط مربعات واحد داخل الفئات فى مقابل عدة متوسطات مربعات للتأثيرات الأخرى، فعادة ما يتم إجراء التصحيح لمتوسط المربعات داخل الفئات. ويكون ذلك بضرب متوسطات المربعات المحسوب على أساس المشاهدات الفردية بمقلوب المتوسط التوافقى harmonic mean لعدد الأفراد فى الفئات، وبذلك يضرب متوسط المربعات داخل الفئات فى الكمية $1/n_h$ كما فى المعادلة (١٣-٤)

$$\frac{1}{n_h} = \frac{1}{rc}\left[\frac{1}{n_{11}} + \frac{1}{n_{12}} + \cdots + \frac{1}{n_{rc}}\right] \qquad (١٣-٤)$$

حيث n_{ij} تمثل عدد المشاهدات فى الفئة ij

مثال ١٣-١

تمثل بيانات جدول ١٣-١ عدد الحملان ومتوسط وزنها عند الميلاد فى دراسة عن تأثير كل من اختلاف سلالة الحمل (٤ سلالات) ونوع الولادة (مفردة أو توأمية).

جدول ١٣-١ وزن الميلاد بالكيلوجرام للحملان وأعدادها فى السلالات المختلفة

السلالة		نوع الولادة: مفردة	نوع الولادة: توأمية	المجموع
العواسى	المتوسط	4.400	3.980	**8.380**
	العدد	35	59	
النجدى	المتوسط	4.448	3.743	**8.191**
	العدد	25	35	
الدوربر	المتوسط	4.314	3.585	**7.899**
	العدد	14	26	
الحرى	المتوسط	3.300	2.713	**6.013**
	العدد	5	25	
المجموع		**16.462**	**14.021**	**30.483**

ويمكن التعبير عن قيمة المتوسط لأى من الفئات فى الجدول السابق بالنموذج الرياضى التالى:

$$\overline{Y}_{ij.} = \mu + \beta_i + \tau_j + (\beta\tau)_{ij} + \overline{\xi}_{ij.} \qquad (١٣-٥)$$

حيث β_i تمثل تأثير السلالة، τ_j تمثل تأثير نوع الولادة، $(\beta\tau)_{ij}$ هو التداخل بين التأثيرين على متوسط وزن الميلاد للفئة. ويشترط للحصول على الحلول ما يلى:

$$\sum_i \beta_i = \sum_j \tau_j = \sum_i (\beta\tau)_{ij} = \sum_j (\beta\tau)_{ij} = 0 \qquad (١٣-٦)$$

$$\overline{\xi}_{ij.} = \sum_k \xi_{ijk} / n_{ij} \qquad (١٣-٧)$$

كما أن هناك افتراضاً لإجراء تحليل التباين واختبار الفروض الخاصة بالتأثيرات الرئيسية وهى أن $\overline{\xi} \sim NID(0, \sigma_I^2)$. ومن خلال هذا النموذج يمكن حساب كل من مجموع المربعات الخاصة بتأثير السلالة ونوع الولادة والتداخل فيما بينهما. وهناك نموذج رياضى آخر يمكن به التعريف بمحتوى المشاهدة الواحدة وهو:

$$Y_{ijk} = \mu + \beta_i + \tau_j + (\beta\tau)_{ij} + \xi_{ijk} \qquad (١٣-٨)$$

والاشتراطات لهذا النموذج كسابقتها فى النموذج (١٣-٦) ولكن يفترض أن $\xi \sim NID(0,\sigma_e^2)$.

وهذا النموذج (١٣-٨) هو المستخدم لحساب مجموع المربعات داخل الفئات وأن σ_e^2 هى التباين للمشاهدة فى العشيرة، فى حين أن σ_I^2 هى التباين لمتوسط الفئة فى العشيرة، ولذا فإنه يتم ضرب قيمة التوقع الخاص بـ σ_e^2 بالكمية $1/n_h$ للحصول على التوقع الخاص بـ σ_I^2 وبالتالى يمكن إجراء اختبارات المعنوية. وتكون الحسابات المطلوبة لإجراء تحليل التباين بطريقة المتوسطات غير الموزونة كما يلى:

١- من حسابات سابقة مبنية على أساس المشاهدات الفردية وجد أن قيمة متوسط المربعات داخل الفئات (within MS) يساوى 0.285

٢- معامل التصحيح CF (من جدول ١٣-١): $(30.483)^2/8 = 116.152$

٣- مجموع المربعات بين الفئات Subclasses SS:

$$(4.4)^2 + (3.98)^2 + \cdots + (2.713)^2 - CF = 2.556$$

٤- مجموع المربعات للسلالات Breeds SS:

$$[(8.38)^2 + (8.191)^2 + (7.899)^2 + (6.013)^2]/2 - CF = 1.782$$

٥- مجموع المربعات لنوع الولادة Type of birth SS:

$$[(16.462)^2 + (14.021)^2]/4 - CF = 0.7448$$

٦- مجموع المربعات للتداخل Interaction SS:
= مجموع المربعات للفئات- (مجموع المربعات للسلالات + مجموع المربعات لنوع الولادة) = 0.0299

٧- مقلوب المتوسط التوافقى:

$$\frac{1}{n_h} = \frac{1}{8}\left[\frac{1}{35} + \frac{1}{59} + \ldots + \frac{1}{25}\right] = 0.05821$$

وهنا يتضح، ضرورة أن تكون كل الفئات بها فرد واحد على الأقل لكى يمكن حساب المتوسط التوافقى، أى تجنب أن يكون مقام أحد هذه الكسور صفر.

٨- متوسط المربعات للخطأ = (متوسط المربعات داخل الفئات) $(1/n_h)$:

$$(0.285)(0.05821) = 0.01659$$

وبالتالى، يمكن استكمال البيانات فى جدول ١٣-٢ لتحليل التباين للمتوسطات الغير موزونة إذا افترض أن تأثير كل من السلالات ونوع الولادة ذات تأثير ثابت fixed effect.

جدول ١٣-٢ تحليل التباين لبيانات جدول ١٣-١ باستخدام طريقة المتوسطات غير الموزونة

SOV	df	SS	MS	F
السلالات (B)	3	1.7820	0.59	35.54**
نوع الولادة (T)	1	0.7448	0.74	44.58**
التداخل (B x T)	3	0.0299	0.01	0.60^{NS}
الخطأ Error	214		0.0166	
المجموع	221			

ولاختبار الفروض الخاصة بتأثير كل من السلالة ونوع الولادة فى النموذج الرياضى الذى تم افتراضه، وفى حالة التأثيرات الثابتة للعوامل fixed effects فإن متوسط المربعات الخاصة بكل من التأثيرات الرئيسية والتداخل يختبر مقابل متوسط المربعات للخطأ.

وكما سبق إيضاحه فى فصل ١١-٥ بأنه إذا كانت التأثيرات الرئيسية ذات تأثير عشوائى، فإنه لاختبار متوسط المربعات لها يكون باستخدام متوسط المربعات الخاص بالتداخل بين العاملين. أما إذا كان أحد التأثيرات عشوائياً والآخر ثابتاً، فيتم اختبار متوسط المربعات الخاص بالعامل العشوائى التأثير بمقارنته بمتوسط مربعات الخطأ، أما العامل الثابت فيختبر مقابل متوسط المربعات للتفاعل. وفى جميع الأحوال يختبر متوسط المربعات للتداخل بين العوامل مع متوسط مربعات الخطأ، ويجدر الملاحظة، أنه فى معظم التجارب يود الباحث معرفة تباين القيمة الفردية فى مثل هذه التجارب، هذه القيمة لا يمكن التعرف عليها من الجدول إلا إذا ذكرت قيمة n_h.

١٣-٤ طريقة الأعداد المتوقعة للفئات

Method of expected subclass numbers

يلاحظ فى كثير من التجارب التى تجرى أن أعداد المشاهدات فى الفئات المختلفة تكون غير متناسبة disproportionate، فى حين يعتقد المجرب أن عنده من الأسباب

ما يقنعه بأن الأعداد الخاصة بالفئات المختلفة في العشيرة التى أخذت منها العينة (التجربة) تتمتع بتناسب فى الأعداد فيما بينها. فمثلاً قد تنطبق نسبة الذكور إلى الإناث على ما هو معروف من تساوى نسبة الجنسين فى العشيرة، أو أن عدد الحملان التوأمية والمفردة التى يحصل عليها المجرب فى بياناته، لا تنطبق عليها النسبة المعروفة فى إحدى سلالات الأغنام، وبناء على ذلك، فيكون السبب فى اختلاف النسب المشاهدة للفئات المختلفة فى العينة عنها فى العشيرة يرجع إلى المعاينة sampling variation أى اختلافات العينات عن بعضها وعن نسب العشيرة، ففى هذه الحالة يمكن استخدام طريقة التحليل المبنية على أساس الأعداد المتوقعة للفئات كما هى فى العشيرة. ويكون ذلك بأن تعدل البيانات (الأعداد أو النسب للفئات المختلفة) لتصبح متناسبة وبالتالى، فإن ما حدث من عدم التطابق لنظرية التجميع additivity لمجاميع المربعات يعود ثانية، وتصبح مجاميع المربعات تجميعية فى تحليل التباين.

ومبدئياً يلزم أن يختبر المجرب ما إذا كان الافتراض بأن الأعداد فى الفئات المختلفة مأخوذة من العشيرة التى بها أعداد الفئات متناسبة من عدمه ويكون ذلك بإجراء اختبار مربع كاى χ^2 test (الباب السابع). فإذا افترض أن عدد المشاهدات فى الفئة التابعة للصف i، والعمود j هى n_{ij} فإن القيمة المتوقعة $E(n_{ij})$ بناء على النسب الموجودة فى العشيرة يمكن حسابها كالتالى:

$$E(n_{ij}) = (n_i.)(n._j)/n.. \qquad (٩-١٣)$$

حيث $n_i.$ هو مجموع المشاهدات فى الصف i ، $n._j$ هو مجموع المشاهدات فى العمود j ، $n..$ هو العدد الكلى للمشاهدات فى التجربة. والقيمة $E(n_{ij})$ هى ما يتوقعه المجرب للفئة ij إذا كانت ينطبق عليها نفس النسب الموجودة فى العشيرة الأصلية. وبالتالى، فإن قيمة χ^2 التى تختبر ما إذا كانت العينات فعلاً متناسبة تحسب على أساس:

$$\chi^2 = \sum_i \sum_j \frac{[n_{ij} - E(n_{ij})]^2}{E(n_{ij})} \qquad (١٠-١٣)$$

وهى الصورة العامة لاختبار مربع كاى.

وتتم مقارنة القيمة المتحصل عليها لمربع كاى χ^2 بالقيمة الجدولية عند مستوى المعنوية المفترض، وبدرجات حرية تساوى $(r-1)(c-1)$ حيث c, r هما عدد الصفوف والأعمدة على التوالى. فإذا كانت قيمة χ^2 المحسوبة معنوية فإن ذلك يدل على أن العينة مأخوذة من عشيرة أعداد الفئات فيها ليست متناسبة. أى أن السبب فى

كون الأعداد فى العينة غير متناسبة لا يرجع إلى التباين فى العينات ولكن لطبيعة العشيرة نفسها.

وبناء على ذلك فإن تحليل التباين لا يمكن إجراؤه بهذه الطريقة. وعموماً فإن اختبار χ^2 يعتبر اختباراً مبدئياً. أى أنه سوف يكون له بعض التأثير على مستوى المعنوية التى يتم بها اختبار كل من التأثيرات الرئيسية والتداخل لاحقاً، ولو أنه يمكن، إلى حد ما، عدم أخذ هذا التأثير فى الاعتبار. ولإجراء تحليل التباين باستخدام طريقة الأعداد المتوقعة للفئات بفرض أن قيمة اختبار χ^2 المبدئية كانت غير معنوية فإنه يتم حساب لكل فئة من الفئات subclass مجموعاً total جديداً على أساس القيمة المحسوبة للعدد المتوقع لهذه الفئة والتى عرفت على أساس $E(n_{ij})$ بضربها فى قيمة المتوسط الحقيقى المحسوب لهذه الفئة، وبالتالى يكون هناك تكرار أو عدد جديد لكل فئة هو قيمة $E(n_{ij})$ ومجموعا جديدا يساوى $\overline{Y}_{ij}[E(n_{ij})]$، أما بالنسبة لمجموع المربعات لداخلى within sum of squares فإنه يتم حسابه كما سبق فى (١٣-٣) أى من المشاهدات الأصلية داخل الفئات ويتم بعد ذلك إجراء حساب مجاميع المربعات للتأثيرات وللتداخل بالطريقة المتبعة عادة باستخدام المجاميع للفئات المعدلة للنسب.

مثال ١٣-٢

يوضح جدول ١٣-٣ بيانات أوزان الحملان من الذكور والإناث لأربعة سلالات مختلفة عند عمر حوالى 170 يوما.

وكما ذكر فإنه يتم إجراء اختبار مبدئى لمعرفة ما إذا كانت الأعداد المتحصل عليها فى التجربة تتبع حقيقة العشيرة التى بها أعداد متناسبة للفئات ويتم ذلك بحساب الأعداد المتوقعة $E(n_{ij})$ كما فى معادلة (١٣-٩).

فمثلاً: $(74)(32)/143 = 16.56$ ، $(69)(44)/143 = 21.23$ ... وهكذا.

ومن هذه الأعداد يتم حساب قيمة χ^2 كما فى المعادلة (١٣-١٠)

$$\chi^2 = \frac{(19-16.56)^2}{16.56} + \frac{(13-15.44)^2}{15.44} + \cdots + \frac{(22-19.78)^2}{19.78} = 1.69$$

بدرجات حرية $(2-1)(4-1) = 3$

وعند مقارنة قيمة χ^2 المحسوبة بقيمة χ^2 من جدول ٦ ملحق أ عند مستوى معنوية 5% أى $[\chi^2_{(0.05,3)} = 7.81]$ يتضح أنها غير معنوية.

جدول ١٣-٣ أوزان الحملان من أربعة سلالات مقسمة حسب الجنس لحسابات طريقة الأعداد المتوقعة للفئات

السلالة		الجنس				
		إناث		**ذكور**		
		المشاهد	**النسبى**	**المشاهد**	**النسبى**	**المجموع**
الدوربر	n_{ij}	19	16.56	13	15.44	**32**
	$\overline{Y}_{ij}$	30.37		27.88		
	$Y_{ij\cdot}$		502.93		430.47	**933.40**
العواسى	n_{ij}	24	22.77	20	21.23	**44**
	$\overline{Y}_{ij}$	27.00		30.15		
	$Y_{ij\cdot}$		614.79		640.08	**1254.87**
الحرى	n_{ij}	12	13.45	14	12.55	**26**
	$\overline{Y}_{ij}$	16.71		18.14		
	$Y_{ij\cdot}$		224.75		227.66	**452.41**
النجدى	n_{ij}	19	21.22	22	19.78	**41**
	$\overline{Y}_{ij}$	29.63		26.32		
	$Y_{ij\cdot}$		628.75		520.61	**1149.36**
المجموع	$\sum_i n_{ij}$	**74**		**69**		**143**
	$\sum_i Y_{ij}$		**1971.22**		**1818.82**	**3790.04**

وفى جدول ١٣-٣ تحسب $\overline{Y}_{ij} = Y_{ij\cdot}$ مضروبة فى العدد النسبى $E(n_{ij})$ أى مثلاً 502.93 فى الخلية الأولى = (30.37)(16.56) وأن 520.61 فى الخلية الأخيرة = (26.32)(19.78) ... وهكذا.

وبناء على ذلك فإنه يتم استكمال تحليل التباين باستخدام هذه الطريقة، وكما سبق فإنه يتم حساب مجموع مربعات الخطأ من البيانات الأصلية من داخل الفئات subclasses والغير مذكورة فى جدول ١٣-٣.

مجموع مربعات الخطأ: within SS = 2721.1852

ويتم بعد ذلك استخدام البيانات الموجودة فى جدول ١٣-٣ لاستكمال تحليل التباين باستخدام الأعداد المتوقعة والمجاميع الجديدة المحسوبة وهى كالتالى:

معامل التصحيح:

$$CF = \frac{(3790.04)^2}{143} = 100450.4$$

مجموع المربعات بين الفئات Subclasses SS:

$$\text{Subclasses SS} = \frac{(502.923)^2}{16.56} + \cdots + \frac{(520.6)^2}{19.78} - CF = 2940.6$$

مجموع المربعات بين السلالات SS (B) :

$$SS\ (B) = \frac{(933.39)^2}{32} + \cdots + \frac{(1149.36)^2}{41} - CF = 2656.6$$

مجموع المربعات بين الجنسين SS (S) :

$$SS\ (S) = \frac{(1971.22)^2}{72} + \frac{(1818.82)^2}{69} - CF = 2.7$$

مجموع السربعات للتداخل SS (S x B) :

$$SS\ (S \times B) = 2940.7 - (2656.71 + 2.77) = 281.3$$

ويتم عرض النتائج المتحصل عليها فى جدول تحليل التباين (جدول ١٣-٤) التالى:

جدول ١٣-٤ تحليل التباين للبيانات فى مثال ١٣-٢ باستخدام طريقة الأعداد المتوقعة للفئات

SOV	df	SS	MS
بين السلالات	3	2656.6	885.5**
بين الجنسين	1	2.7	2.7
التداخل	3	281.3	93.8**
الخطأ(أو داخل الفئات)	135	2721.2	20.16
المجموع	**142**		

وباختبار الفرض الخاص بمعنوية التداخل بين العاملين بقسمة متوسط المربعات الخاص بالتداخل على الخطأ وجد أنه معنوى جداً وهو اختبار مباشر. أما بالنسبة لاختبار التأثيرات الرئيسية فإنه نتيجة لعدم تساوى الأعداد تختلف قيم المعاملات لكل من مكونات التباين ولقد أوضح Searle (1986) مكونات كل من متوسطات المربعات فى هذه الحالة حسب ما إذا كانت العوامل الرئيسية فى النموذج ثابتة fixed أو عشوائية random أو مختلطة mixed وبالتالى يمكن تكوين المقام الملائم لإجراء اختبار F حسب النموذج الرياضى والعامل الرئيسى الذى يجرى اختباره.

وفى التجربة السابقة إذا كان النموذج الرياضى المستخدم يفترض أن تأثير العوامل الرئيسية ثابت فعليه يمكن إجراء اختبار F باستخدام متوسط المربعات للخطأ ولو أنه يعتبر اختباراً غير دقيق. ونتيجة لكون التداخل بين العوامل تأثيرها معنوياً فإنه لا يصبح هناك معنى كبير للآثار الرئيسية، ولكن قد ينظر إلى حجم تأثير كل من العاملين ومشاركته فى مجموع المربعات. ويتضح هذا فى اختلاف حجم التأثير للعامل الأول مقارنة بالعامل الثانى.

صندوق ١٣-١

إذا لم يؤخذ فى الاعتبار عدم تناسب أعداد المشاهدات (الوحدات التجريبية) فى الفئات فإنه يتسبب عنها مشكلتان هما: عدم استقلالية المشاهدات وخرق لقاعدة مجموع المربعات. ولتفادى هذا ذكرت طريقتان تقريبيتان هما طريقة المتوسطات غير الموزونة وطريقة الأعداد المتوقعة للفئات. وكلتا الطريقتين يحققان هدف تجاوز المشكلتين السابق ذكرهما. ولكنهما طريقتين تقريبيتين والثانية منهما تتطلب فروضاً معينة مثل أن عدم التناسب فى الأعداد مرجعه عشوائى وليس بسبب معاينة غير موفقة أو بسبب التجربة نفسها (مثل نفوق حيوانات نتيجة للمعاملة مثلاً).

١٣-٥ الطرق المضبوطة لتحليل التباين فى حالة عدم تساوى الأعداد فى الفئات

Exact methods for analysis of variance for data with unequal subclass frequencies

إن طرق تحليل التباين المضبوطة للبيانات التى تتصف بعدم تساوى الأعداد فى فئاتها المختلفة تحتاج إلى حسابات معقدة بعض الشئ، ويحتاج بعضها إلى المعرفة بنواحى معينة من الرياضيات مثل جبر المصفوفات، خاصة عندما يزداد عدد التقسيمات classifications أو العوامل factors وبالتالى تتفاوت الأعداد فى الفئات المختلفة تفاوتاً كبيراً، كما وأن هناك طرق تحليل مضبوطة تتبع فى حالة احتواء النموذج الرياضى على تداخلات بين العوامل أو فى حالة عدم وجودها.

وسوف يتم استعراض بعض الطرق المستخدمة فى التحليل فى حالة وجود عاملين أو أساسين للتقسيم أى حالة two-way classifications (والتى يمكن توسيع مداها لتشمل حالات وجود أكثر من عاملين) ففى هذه الحالة ينشأ اتجاهين للتقسيم أى عاملان A، B ولكل منهما عدد من المستويات، فإذا افترض أن عدد المستويات لعامل الأول A تساوى r أى يصبح هناك $A_1, A_2, \cdots, A_r$ وأن عدد المستويات لعامل الثانى B تساوى c أى أن هناك $B_1, B_2, \cdots, B_c$ فيكون العدد الكلى للفئات subclasses هو rc وهو حاصل ضرب مستويات A_i فى مستويات B_j. أما بالنسبة لعدد المشاهدات فى كل فئة فيختلف، وبذلك تكون قيم n_{ij} مختلفة. وبذا ينشأ الشكل التالى للبيانات كما فى جدول ١٣-٥.

حيث تكون التعويضات الموجودة فى الجدول كالتالى:

عدد المشاهدات فى الفئة ij: n_{ij}

مجموع قيم المشاهدات فى الفئة ij: $Y_{ij.} = \sum_k Y_{ijk}$

قيمة المشاهدة k فى الفئة ij: Y_{ijk}

عدد المشاهدات فى المستوى A_i: $n_{i.} = \sum_j n_{ij}$

مجموع قيم المشاهدات فى المستوى A_i: $Y_{i..} = \sum_j \sum_k Y_{ijk}$

عدد المشاهدات فى المستوى B_j: $n_{.j} = \sum_i n_{ij}$

جدول ١٣-٥ التقسيم الثنائى فى حالة اختلاف الأعداد الفئوية

		B						
		B_1	B_2	...	B_j	...	B_c	**Total**
A	A_1	$Y_{11\cdot}$	$Y_{12\cdot}$	...	$Y_{1j\cdot}$	...	$Y_{1c\cdot}$	$Y_{1\cdot\cdot}$
		n_{11}	n_{12}	...	n_{1j}	...	n_{1c}	$n_{1\cdot}$
	A_2	$Y_{21\cdot}$	$Y_{22\cdot}$	...	$Y_{2j\cdot}$	...	$Y_{2c\cdot}$	$Y_{2\cdot\cdot}$
		n_{21}	n_{22}	...	n_{2j}	...	n_{2c}	$n_{2\cdot}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	A_i	$Y_{i1\cdot}$	$Y_{i2\cdot}$	...	$Y_{ij\cdot}$	...	$Y_{ic\cdot}$	$Y_{i\cdot\cdot}$
		n_{i1}	n_{i2}	...	n_{ij}	...	n_{ic}	$n_{i\cdot}$
	⋮	⋮	⋮	...	⋮	⋮	⋮	⋮
	A_r	$Y_{r1\cdot}$	$Y_{r2\cdot}$	...	$Y_{rj\cdot}$	...	$Y_{rc\cdot}$	$Y_{r\cdot\cdot}$
		n_{r1}	n_{r2}	...	n_{rj}	...	n_{rc}	$n_{r\cdot}$
Total		$Y_{\cdot 1\cdot}$	$Y_{\cdot 2\cdot}$	...	$Y_{\cdot j\cdot}$	...	$Y_{\cdot c\cdot}$	$Y_{\cdot\cdot\cdot}$
		$n_{\cdot 1}$	$n_{\cdot 2}$	...	$n_{\cdot j}$	...	$n_{\cdot c}$	$n_{\cdot\cdot}$

مجموع قيم المشاهدات فى المستوى B_j: $\quad Y_{\cdot j\cdot} = \sum_i \sum_k Y_{ijk}$

عدد المشاهدات الكلية: $\quad n_{\cdot\cdot} = \sum_i \sum_j n_{ij} = \sum_i n_{i\cdot} = \sum_j n_{\cdot j}$

مجموع قيم المشاهدات الكلية: $\quad Y_{\cdot\cdot\cdot} = \sum_i \sum_j \sum_k Y_{ijk}$

وبالطبع يمكن الحصول على متوسط كل فئة أو كل مستوى بقسمة المجموع الخاص بالفئة أو المستوى على عدد المشاهدات لهذا المجموع.

وسوف يتم التطرق إلى الطرق المستخدمة فى تحليل التباين لهذا النوع من البيانات.

١٣-٦ طريقة موائمة الثوابت للتأثيرات الثابتة للعوامل

Method of fitting constants for fixed effects

تستخدم هذه الطريقة فى الحالات التى لا يفترض النموذج الرياضى فيها وجود التداخل بين العوامل المؤثرة على تباين قيم المشاهدات بل فقط التأثيرات الرئيسية main effects، ولا يلغى هذا إمكانية حساب التداخل من نتائج تحليل العينة وبالتالى اختبارها. ويعرف هذا النوع من التحليل أيضاً بطريقة المربعات الصغرى least squares analysis. وهذه الطريقة فى التحليل للبيانات غير المتساوية وغير المتناسبة أكثر الطرق المضبوطة استخداماً ويمكن لتسهيل العمليات الحسابية البدء بالتحليل المبدئى للبيانات ويكون كالتالى:

١- يحسب مجموع المربعات للفئات Subclasses SS وهو

$$\sum_i \sum_j Y_{ij.}^2 / n_{ij} - CF \qquad (١٣-١١)$$

حيث معامل التصحيح هو $CF = Y_{...}^2 / n_{..}$. ومجموع المربعات هذا له درجات حرية تساوى $(rc-1)$.

٢- يحسب مجموع المربعات الخاص بتأثير العامل الأول A وبإهمال أى تأثير للعامل الثانى B أى $(A, \text{ignoring } B)$ وله $(r-1)$ درجات حرية

$$\text{SS A, ignoring B} = \sum_i \frac{Y_{i..}^2}{n_{i.}} - CF \qquad (١٣-١٢)$$

٣- بنفس الطريقة يحسب مجموع المربعات الخاص بتأثير العامل الثانى مع إهمال تأثير العامل الأول $(B, \text{ignoring } A)$ وله $(c-1)$ درجات حرية

$$\text{SS B, ignoring A} = \sum_j \frac{Y_{.j.}^2}{n_{.j}} - CF \qquad (١٣-١٣)$$

٤- أخيراً يتم حساب مجموع المربعات داخل الفئات within subclasses SS وهذه الكمية لها درجات حرية تساوى $(n_{..} - rc)$

$$\text{within subclasses SS} = \sum_i \sum_j \sum_k Y_{ijk}^2 - \sum_i \sum_j \frac{Y_{ij.}^2}{n_{ij}} \quad (١٤-١٣)$$

أو يطرح مجموع المربعات للفئات من مجموع المربعات الكلى بعد تصحيح كليهما.

وبالطبع فإن مجموع المربعات الكلى $\text{Total SS} = \sum_i \sum_j \sum_k Y_{ijk}^2 - CF$

وله درجات حرية تساوى $(n..-1)$

وكما سبق فإن النموذج الرياضى المستخدم فى هذا التحليل يكون كالتالى:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \xi_{ijk} \quad (١٥-١٣)$$

حيث $i = 1, 2, \cdots, r$ ، $j = 1, 2, \cdots, c$ ، $k = 1, 2, \cdots, n$

ويفترض حتى يمكن إجراء اختبارات المعنوية أن $\xi \sim NID(0, \sigma^2)$. ولإيجاد حل لتأثير العوامل، أو بمعنى آخر حتى يمكن حل المعادلات الاعتيادية normal equations والتى تنتج من جعل مجموع المربعات للانحرافات أقل ما يمكن يجب وضع الاشتراط التالى:

$$\sum_i \alpha_i = \sum_j \beta_j = 0 \quad (١٦-١٣)$$

أو أى اشتراط آخر يمكن من حل المعادلات الطبيعية (مثلاً $\alpha_1 = \beta_1 = 0$).

وبناء على نظرية أقل مجموع مربعات theory of least squares فإنه للحصول على التقديرات للثوابت التى تعطى أقل مجموع مربعات للخطأ أن يتم إجراء التفاضل للنموذج الرياضى (١٣-٥) أى أن المطلوب هو الحصول على تقديرات للثوابت تجعل الكمية

$$\sum_{ijk} e_{ijk}^2 = \sum [Y_{ijk} - \mu - \alpha_i - \beta_j]^2 \quad (١٧-١٣)$$

أقل ما يمكن.

ويمكن الحصول على المعادلات الاعتيادية normal equations بإجراء التفاضل الجزئى بالنسبة للمعالم ومساواة المعادلة بالصفر والتى تعطى الحلول أو التقديرات

للثوابت بعد تطبيق الاشتراطات السابق ذكرها والتى يمكن تلخيصها فى جدول ١٣-٦

وبالنظر إلى المعادلات الاعتيادية يتضح أن المعامل coefficient لكل من الثوابت فى معادلة رقم 1 يساوى مجموع المعاملات لنفس الثابت فى المعادلات من رقم 2 وحتى رقم $(r+1)$، وبنفس الطريقة فمعامل المعادلة رقم 1 يساوى مجموع المعاملات للمعادلات من رقم $(r+2)$ إلى $(r+c+1)$، ونفس الشيء ينطبق على الطرف الأيمن من المعادلات. وبالتالى لا يمكن إيجاد حلول لهذه المعادلات الآنية نظراً لوجود هذه العلاقة فيما بينها أى أنها تصبح معادلات غير مستقلة.

وحتى يمكن إيجاد حلول لهذه المعادلات الآنية (الاعتيادية) تطبق الاشتراطات السابق ذكرها فى (١٣-١٦) أو بعض الاشتراطات الأخرى والتى بتطبيقها يقل عدد المعادلات فى كل تقسيم classification معادلة واحدة، وكذلك عدد الثوابت المراد تقديرها، وبالتالى يصبح هناك معادلة واحدة للمتوسط μ، $(r-1)$ معادلات للمستويات الخاصة بالعامل α، $(c-1)$ معادلات للمستويات الخاصة بالعامل β أى يكون العدد الكلى للمعادلات الاعتيادية هو $(r+c-1)$.

فمثلاً إذا كان عدد المستويات للعامل α هو $r=3$ مستويات هو $\alpha_1,\alpha_2,\alpha_3$ فى حين أن عدد المستويات للعامل β هو مستويين $c=2$ وهما β_1,β_2 فإن العدد الكلى للمعادلات الاعتيادية الناتجة وهو 6 معادلات (واحدة للمتوسط + 3 للعامل α + 2 للعامل β)، أما بعد تطبيق الاشتراطات فيصبح العدد الكلى للمعادلات هو 4 (معادلة المتوسط + 2 للعامل α + معادلة واحدة للعامل β).

وعموماً يطلق على المعادلات الناتجة من تطبيق هذه الاشتراطات بالمعادلات الاعتيادية المختزلة reduced normal equation. وكنتيجة لحل هذه المعادلات المختزلة آنياً يمكن الحصول على التقديرات (الحلول) للثوابت التى تحقق أقل مربعات خطأ ممكن.

فإذا كان عدد المعادلات المطلوب حلها عدداً قليلاً فى حدود 3 أو 4 معادلات فإنه يمكن حلها بالطرق العادية لحل المعادلات الخطية، أما إذا زاد العدد عن ذلك، فإن جبر المصفوفات يصبح أنسب الطرق للحصول على تلك الحلول (الآن يجرى هذا بالحاسب الآلى فى أقل من لمح البصر).

ويلاحظ أيضاً فى جدول ١٣-٦ أن المعاملات للتقديرات المختلفة فى الجدول أى a_i، b_j تكون متناظرة حول الخط المحورى الرئيسى leading diagonal والذى يمثل القيم التى تتساوى فيها قيمة j, i (ويبدأ من الصف الأول والعمود الأول ثم الصف الثانى والعمود الثانى ... وهكذا حتى الصف الأخير والعمود الأخير). فمثلاً

جدول ١٣-٦ مصفوفة المعادلات الاعتيادية

المعلم	Normal equation المعادلة الاعتيادية										
	الطرف الأيسر (LHS)										
parameter	$\hat{\mu}$	a_1	a_2	$\cdots$	a_r	b_1	b_2	$\cdots$	b_c	الطرف الأيمن (RHS)	رقم المعادلة
μ	$n_{..}$	$n_{1.}$	$n_{2.}$	$\cdots$	$n_{r.}$	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.c}$	$Y_{...}$	1
α_1	$n_{1.}$	$n_{1.}$				n_{11}	n_{12}	$\cdots$	n_{1c}	$Y_{1..}$	2
α_2	$n_{2.}$		$n_{2.}$			n_{21}	n_{22}	$\cdots$	n_{2c}	$Y_{2..}$	3
$\vdots$	$\vdots$			$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
α_r	$n_{r.}$				$n_{r.}$	n_{r1}	n_{r2}	$\cdots$	n_{rc}	$Y_{r..}$	$r+1$
β_1	$n_{.1}$	n_{11}	n_{21}	$\cdots$	n_{r1}	$n_{.1}$				$Y_{.1.}$	$r+2$
β_2	$n_{.2}$	n_{12}	n_{22}	$\cdots$	n_{r2}		$n_{.2}$			$Y_{.2.}$	$r+3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$			$\vdots$		$\vdots$	$\vdots$
β_c	$n_{.c}$	n_{1c}	n_{2c}	$\cdots$	n_{rc}				$n_{.c}$	$Y_{.c.}$	$r+c+1$

المعامل للصف الأول والعمود الثانى $j = 2, i = 1$ هو n_{12} وهى نفس القيمة للصف الثانى والعمود الأول $j = 1, i = 2$ ... وهكذا. ويسمى هذا الترتيب للمعاملات فى الجدول بالمصفوفة المتناظرة symmetrical matrix.

وتعرف المصفوفة السابقة الذكر بمصفوفة التباين والتغاير للمعاملات Variance-covariance coefficient matrix

ويلاحظ فى هذا الجدول أنه داخل كل تقسيم من التقسيمات تكون المعاملات الخارجة عن الخط المحورى الرئيسى off-diagonal تساوى صفراً. فمثلاً معاملات $a_2, a_3, \cdots, a_r$ فى معادلة α_1 تساوى صفراً. وأيضاً معاملات $b_2, b_3, \cdots, b_c$ فى معادلة β_2 تساوى صفراً أيضاً.

أما RHS فهى اختصار لـ Right Hand Side أى الجانب الأيمن وهى ببساطة تمثل مجاميع الأقسام المقابلة. فالمقابل لمعادلة المتوسط هو المجموع العام بينما $Y_1..$ هو مجموع المعاملة الأولى ... وهكذا.

وكما سبق فإنه بحل المعادلات الاعتيادية المختزلة يمكن الحصول على التقديرات للثوابت فى النموذج وهى قيم $\hat{\mu}$، a_i، b_j حيث تمثل هذه قيم الثوابت μ، α_i، β_j على التوالى.

ويمكن إثبات أن الاختزال reduction فى مجموع المربعات الكلى الراجع لموائمة النموذج الرياضى الكامل (١٣-١٥) يحسب كالتالى:

الاختزال فى مجموع المربعات الكلى نتيجة لموائمة جميع التقديرات μ، α_i، β_j Reduction in total sum of squares due to fitting all constants ويرمز له $R(\mu, a_i, b_j)$ وهو يساوى:

$$R(\mu, a_i, b_j) = \hat{\mu}Y... + a_1Y_1.. + a_2Y_2.. + \cdots + a_rY_r.. + b_1Y._1. + b_2Y._2. + \cdots + b_cY._c. \qquad (١٣-١٨)$$

والتى يمكن اختصارها كالتالى:

$$R(\mu, a_i, b_j) = \hat{\mu}Y... + \sum_i a_iY_i.. + \sum_j b_jY._j. \qquad (١٣-١٩)$$

ويعتبر مجموع مربعات الخطأ أو المتبقى المتحصل عليه من مواءمة هذه التقديرات للثوابت فى النموذج الرياضى هو أقل مجموع مربعات ممكن، نظراً لأنه قد استخدمت جميع الثوابت.

وبتطبيق نفس الطريقة العامة يمكن مواءمة نماذج مختزلة لأى مشاهدة تتضمن تأثير عامل واحد من العاملين والنماذج التى تطبق تكون كالتالى:

$$Y_{ijk} = \mu + \alpha_i + \xi_{ijk} \qquad (٢٠-١٣)$$

$$Y_{ijk} = \mu + \beta_j + \xi_{ijk} \qquad (٢١-١٣)$$

وبالتالى يمكن حساب الاختزال reduction فى مجموع المربعات نتيجة لموائمة النموذجين (١٣-٢٠)، (١٣-٢١). ويعبر عنها كالتالى $R(\mu,\alpha_i)$ ، $R(\mu,\beta_j)$ على التوالى.

ويلاحظ أنه بمواءمة النموذج الرياضى

$$Y_{ijk} = \mu + \xi_{ijk} \qquad (٢٢-١٣)$$

يحسب الاختزال فى مجموع المربعات نتيجة لهذا العامل، ويعبر عنه $R(\mu)$ وهذه الكمية تساوى

$$R(\mu) = \frac{(\Sigma Y_{ijk})^2}{n..} = CF \qquad (٢٣-١٣)$$

أى أنها تساوى قيمة معامل التصحيح، وعلى هذا الأساس فإنه يمكن باستخدام المعادلتين (١٣-١٩)، (١٣-٢٣) حساب الاختزال فى مجموع المربعات كنتيجة لمواءمة تأثير كل من العاملين A, B ويمكن التعبير عنها كالتالى:

$$R(a_i, b_j) = \hat{\mu}Y... + \sum_i a_i Y_i.. + \sum_j b_j Y._j. - CF \qquad (٢٤-١٣)$$

والآن يمكن استخلاص الاختزال الإضافى فى مجموع المربعات الراجع لمواءمة α_i والتى يرمز لها بالرمز $R'(a_i)$ وذلك عن طريق طرح معامل التصحيح من الاختزال فى مجموع المربعات الناتج من مواءمة النموذج $\mu + \alpha_i$ فى (١٣-٢٠)، وهو بالضبط يساوى ما تم الحصول عليه فى التحليل المبدئى أى مجموع المربعات الخاص بتأثير العامل A بإهمال أى تأثير للعامل B حسب المعادلة (١٣-١٢) وطبعاً نفس الشئ ينطبق على $R'(b_j)$. ومن جميع ما تقدم يمكن الآن، حساب تحليل التباين النهائى ووضعه فى جدول يمكن منه اختبار فرض العدم الخاص بالتأثيرات الرئيسية للعوامل A, B بناء على النموذج الرياضى الخالى من تأثير التداخل بين العوامل (١٣-١٥). ولقد سبق إيضاح أنه يمكن حساب مجموع المربعات الخاص بالتداخل بين

العاملين A، B من الحسابات الخاصة بهذا النموذج الرياضى بطريقة غير مباشرة ولو أنه غير وارد كجزء من التأثيرات، وبالتالى فإنه يمكن اختبار متوسط المربعات للتداخل لاختبار مدى تعبير النموذج (١٣-١٥) عن البيانات فى أحقية عدم وجود تداخل بين التأثيرات.

ومن المفترض أنه حتى يكون هناك تداخل بين العوامل وبالتالى يمكن حسابه مباشرة أن يتم تبنى النموذج الرياضى الخاص بوجود هذا التأثير، كما سبق إيضاحه، فى (١٣-٨) وبالاشتراطات السابق ذكرها فى (١٣-٦) أى أنه يتم مواءمة النموذج والحصول على الاختزال الكلى فى مجموع المربعات أى $R[\mu, a_i, b_j, (ab)_{ij}]$ وبالتالى $R[a_i, b_j, (ab)_{ij}]$ بطرح معامل التصحيح.

ولكن يلاحظ أن هذا الأخير هو بالضبط ما سبق الحصول عليه فى (١٣-١١) أى مجموع المربعات بين الفئات subclasses SS أى أنه يمكن الحصول عليه بالطرح. وعلى هذا الأساس يمكن عمل جدول تحليل التباين النهائى للبيانات كما سيوضح فى جدول ١٣-٧.

ومن هذا الجدول يمكن اختبار كل من التأثيرات الرئيسيـــة للنموذج (١٣-١٥) كما يمكن أيضاً اختبار معنوية التداخل بين العوامل، وبالطبع فإن اختبار التأثيرات يصبح صحيحاً فقط فى حالة عدم معنوية التداخل فى العشيرة.

ويتم اختبار معنوية التداخل بين العوامل حسب فرض العدم: $H_\circ : \alpha\beta = 0$، بقسمة متوسط مربعات التداخل على متوسط مربعات الخطأ أى:

$$F = (MS)_{AB} / (MS)_E$$

وذلك لكل أنواع التأثيرات سواء كانت ثابتة أو عشوائية أو مختلطة كما فى جدول ١٣-٧، وتكون درجات الحرية للاختبار هى درجات الحرية المقابلة لكل متوسط مربعات.

أما لاختبار كل من التأثيرات الرئيسية A، B فإنها أيضاً يتم اختبارها باستخدام متوسط المربعات للخطأ $(MS)_E$ حيث أنه فى هذه الطريقة هناك افتراض مبدئى أنه لا يوجد تداخل بين تأثير العوامل فى العشيرة، وبذا يصبح متوسط المربعات للخطأ هو الأساس للاختبار سواء لفرض العدم $H_\circ : \alpha = 0$ أو $H_\circ : \beta = 0$ وبدرجات الحرية المقابلة لكل أى $(r-1)$ أو $(c-1)$ للبسط على التوالى ودرجـــات حـــريـــة $(n_{..} - rc)$ للمقام.

وهنا قد تجدر الملاحظة بأنه فى حالة عدم واقعية افتراض عدم وجود تداخل بين العوامل يصبح اختبار فرض العدم الخاص بالتأثيرات الرئيسية غير صحيح تماماً،

جدول ١٣-٧ جدول كامل لبيانات محللة بطريقة مواءمة الثوابت*

SOV	df	SS	MS
A باستبعاد B A eliminating B	r 1	R(a,b) SSB ignoring A R(a) (١٣-١٣) - (٢٤-١٣)	R(a)/(r 1) $(MB)_A$
B باستبعاد A B eliminating A	c 1	R(a,b) SSA ignoring B R(b) (١٢-١٣) - (٢٤-١٣)	R(b)/(c 1) $(MB)_B$
التداخل بين A، B AB interaction	(r 1)(c 1)	R[a,b,(ab)] R(a,b) R(ab) (٢٤-١٣) - (١١-١٣)	R(ab)/[(r 1)(c 1)] $(MB)_{AB}$
داخل الفئات Within subclasses	n.. rc	Total SS subclasses SS SEE (١٤-١٣)	SSE/(n.. rc) $(MS)_E$

* الأرقام تشير إلى أرقام المعادلات التى يمكن استخدامها فى الحسابات

ويلزم إتباع طرق أخرى لتحليل البيانات لا تتضمن أصلاً هذا الافتراض الخاص بالتداخل في العشيرة مثل طريقة المربعات الموزونة للمتوسطات method of weighted squares of means.

مثال ١٣-٣

البيانات التالية في جدول ١٣-٨ تمثل كيفية إجراء تحليل التباين باستخدام طريقة مواءمة الثوابت في التقسيم الثنائي الجهة. وتمثل البيانات أوزان الفطام للحملان من 4 سلالات ومقسمة حسب جنس الحمل.

جدول ١٣-٨ وزن الفطام للحملان من 4 سلالات مقسمة حسب الجنس

الجنس		الإناث b_1	الذكور b_2	المجموع
العواسي a_1	n_{ij}	24	20	**44**
	$\sum Y_{ijk}$	648.0	603.0	**1251**
	$\sum Y_{ijk}^2$	17619.5	18688.0	**36307.5**
النجدي a_2	n_{ij}	19	22	**41**
	$\sum Y_{ijk}$	563.0	579.0	**1142**
	$\sum Y_{ijk}^2$	17155.5	15976.5	**33132**
الدوربر a_3	n_{ij}	19	13	**32**
	$\sum Y_{ijk}$	577.0	362.5	**939.5**
	$\sum Y_{ijk}^2$	17888.0	10483.75	**28371.75**
الحري a_4	n_{ij}	12	14	**26**
	$\sum Y_{ijk}$	200.5	254.0	**454.5**
	$\sum Y_{ijk}^2$	3459.75	4636.5	**8096.25**
المجموع		**74**	**69**	**143**
		1988.5	**1798.5**	**3787**
		56122.75	**49784.75**	**105907.5**

ومن هذا الجدول يمكن استنباط المعادلات الاعتيادية الخاصة بهذا المثال كما هو موضح في جدول ١٣-٦ وهذه المعادلات حسب النموذج الرياضي

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \xi_{ijk}$$

وكما سبق بافتراض عدم وجود تداخل بين العوامل ستكون المعادلات الاعتيادية كالتالى:

المتوسط	$143\hat{\mu} + 44a_1 + 41a_2 + 32a_3 + 26a_4 + 74b_1 + 69b_2 = 3787$
العواسى (a_1)	$44\hat{\mu} + 44a_1 + 24b_1 + 20b_2 = 1251$
النجدى (a_2)	$41\hat{\mu} + 41a_2 + 19b_1 + 22b_2 = 1142$
الدوربر (a_3)	$32\hat{\mu} + 32a_3 + 19b_1 + 13b_2 = 939.5$
الحرى (a_4)	$26\hat{\mu} + 26a_4 + 12b_1 + 14b_2 = 454.5$
الإناث (b_1)	$74\hat{\mu} + 24a_1 + 19a_2 + 19a_3 + 12a_4 + 74b_1 = 1988.5$
الذكور (b_2)	$69\hat{\mu} + 20a_1 + 22a_2 + 13a_3 + 14a_4 + 69b_2 = 1798.5$

وهناك العديد من الاشتراطات التى يمكن فرضها لجعل المعادلات الاعتيادية ذات حل منها أن $a_1 + a_2 + a_3 + a_4 = 0$ ، $b_1 + b_2 = 0$. أو اشتراط آخر هو افتراض أى من a_1 أو a_2 أو a_3 أو a_4 مساويا للصفر وكذلك أى من b_1 أو b_2 مساوياً للصفر. وبالتالى يمكن الحصول على المعـادلات الاعتيادية المختـزلة reduced normal equations وذلك بطرح أحد المعادلات من كل تقسيم. مثلاً طرح معادلة a_4 من كل من a_1 ، a_2 ، a_3 وأيضا طرح معادلة b_2 من معادلة b_1 . وعن طريق الطرح من الصفوف والأعمدة فى كل معادلة يمكن الحصول على المصفوفة التالية:

المتوسط	$\hat{\mu}$	a_1	a_2	a_3	a_4	b_1	b_2	
μ	143	44	41	32	26	74	69	= 3787
α_1	18	44	0	0	-26	12	6	= 796.5
α_2	15	0	41	0	-26	7	8	= 687.5
α_3	6	0	0	32	-26	7	-1	= 485
β_1	5	4	-3	6	-2	74	-69	= 190

يلى ذلك طرح العمود a_4 من كل من العمود a_1 و a_2 و a_3 وأيضاً طرح العمود b_2 من العمود b_1 ، للحصول على المصفوفة التالية:

$$
\begin{array}{c|c|c|c}
 & \begin{matrix}\hat{\mu} & a_1 & a_2 & a_3 & b_1\end{matrix} & \text{RHS} & \\
 & \begin{bmatrix} 143 & 18 & 15 & 6 & 5 \\ 18 & 70 & 26 & 26 & 6 \\ 15 & 26 & 67 & 26 & -1 \\ 6 & 26 & 26 & 58 & 8 \\ 5 & 6 & -1 & 8 & 143 \end{bmatrix} & \begin{bmatrix} 3787.0 \\ 796.5 \\ 687.5 \\ 485.0 \\ 190.0 \end{bmatrix} & \begin{matrix} \text{المتوسط } (\mu) \\ \text{العواسى } (\alpha_1) \\ \text{النجدى } (\alpha_2) \\ \text{الدوربر } (\alpha_3) \\ \text{الإناث } (\beta_1) \end{matrix} \\
 & X'X & X'Y &
\end{array}
$$

ويلاحظ فى المصفوفة الأخيرة والتى يعبر عنها بـــ $X'X$ ما يلى:

١- أن المصفوفة متناظرة حول القطر الرئيسى leading diagonal أى أن قيمة العنصر الموجود فى الصف الثانى والعمود الثالث (26) هى نفس قيمة العنصر الموجود فى الصف الثالث والعمود الثانى (26) ... وهكذا.

٢- القيم الموجودة على الخط القطرى الرئيسى لابد وأن تكون موجبة، وذلك لكونها تمثل مجموع مربعات أما القيم الموجودة خارج هذا القطر off-diagonal فيمكن أن تكون موجبة أو سالبة، وذلك لأنها تمثل مجموع حاصل ضرب cross products.

٣- أن قيمة العنصر فى الصف الأول والعمود الأول وهى معامل المتوسط $\hat{\mu}$ تمثل العدد الكلى للأفراد، وكذلك قيمة الجانب الأيمن (RHS) والذى يعبر عنه بـــ $X'Y$ لهذه المعادلة تمثل المجموع الكلى للقيم.

٤- تبعاً لتطبيق الاشتراطات لإيجاد الحلول للمعادلات فإن العلاقة الموجودة بين معادلة المتوسط $\hat{\mu}$ والمعادلات الخاصة بكل تأثير والتى كانت متمثلة فى أن معامل الثوابت فى معادلة المتوسط يساوى مجموع معاملات الثوابت فى المعادلة الخاصة بكل تأثير، وكذلك RHS قد اختفت وبالتالى أصبحت المعادلات مستقلة عن بعضها أى ذات حل.

وهناك أكثر من طريقة لحل المعادلات الاعتيادية المختزلة آنياً simultaneous solving منها حل المعادلات عن طريق جبر المصفوفات matrix algebra وبالتالى يلزم الحصول على مقلوب المصفوفة matrix inverse. وعلى هذا الأساس فإن مقلوب المصفوفة المختزلة $X'X$ وهو $(X'X)^{-1}$ يساوى:

$$\begin{bmatrix} 0.007309 & -0.001607 & -0.001228 & 0.000546 & -0.000227 \\ -0.001607 & 0.018690 & -0.004521 & -0.006128 & -0.000417 \\ -0.001228 & -0.004521 & 0.019575 & -0.006724 & 0.000746 \\ 0.000546 & -0.006128 & -0.006724 & 0.023098 & -0.001101 \\ -0.000227 & -0.000417 & 0.000746 & -0.001101 & 0.007085 \end{bmatrix}$$

وللحصول على تقديرات للثوابت فإن $\hat{\beta} = (X'X)^{-1}X'Y$ حيث $\hat{\beta}$ هى المتجهة vector: $(\hat{\mu} \quad a_1 \quad a_2 \quad a_3 \quad b_1)$ وبالتالى يمكن الحصول على التقديرات التالية للثوابت فى النموذج الرياضى:

١- المتوسط العام: $\hat{\mu} = 25.777$ kg

٢- تأثير السلالات (عددها 3 فى المصفوفة المختزلة):

$a_1 = 2.643$، $a_2 = 2.086$ ، $a_3 = 3.557$ للعواسى والنجدى والدروبر على التوالى. وبتطبيق الاشتراطات (١٣-١٦) يمكن حساب قيمة a_4 حيث:

$$a_4 = -(a_1 + a_2 + a_3) = -8.286$$

وبالتالى يمكن حساب متوسط كل من السلالات بجمع التأثير الخاص بتلك السلالة على قيمة المتوسط العام للحصول على المتوسطات التالية:

العواسى $25.777 + 2.643 = 28.42$

النجدى $25.777 + 2.086 = 27.862$

الدوربر $25.777 + 3.557 = 29.334$

الحرى $25.777 + (-8.286) = 17.49$

٣- تأثير جنس الحمل (عددها واحد فى المصفوفة المختزلة) $b_1 = 0.132$ kg للإناث، وبتطبيق الاشتراط السابق فإن قيمة $b_2 = -0.132$، وعلى هذا الأساس يكون متوسط الإناث عند الفطام $25.777 + 0.132 = 25.909$ ومتوسط الذكور $25.777 + (-0.132) = 25.645$

ويمكن حالياً البدأ فى حساب مجاميع المربعات حتى يمكن استكمال تحليل التباين:

١- مجموع المربعات الكلية الغير مصحح: $\sum_{ijk} Y_{ijk}^2 = 105907.5$

٢- معامل التصحيح:

$$CF = \frac{(Y...)^2}{n} = \frac{(3787)^2}{143} = 100289.2937$$

٣- مجموع المربعات الكلي المصحح:

$$TSS = \sum_{ijk} Y_{ijk}^2 - CF = 105907.5 - 100289.2937 = 5618.2063$$

٤- مجموع المربعات بين الفئات:

$$\text{Subclasses SS} = \sum_{ij} \frac{Y_{ij}^2}{n_{ij}} - CF = \frac{(648)^2}{24} + \frac{(603)^2}{20} + \cdots + \frac{(254)^2}{14} - CF$$
$$= 2897.0211$$

٥- مجموع المربعات الخاص بالعامل الأول (a) بإهمال تأثير العامل الثانى (b):

$$\text{SSA ignoring B} = \sum_{i} \frac{Y_{i..}^2}{n_{i.}} - CF = \frac{(1251)^2}{44} + \cdots + \frac{(454.5)^2}{20} - CF$$
$$= 2615.9313$$

٦- مجموع المربعات للعامل الثانى (b) بإهمال تأثير العامل الأول (a):

$$\text{SSB ignoring A} = \sum_{j.} \frac{Y_{.j.}^2}{n_{.j}} - CF = \frac{(1798.5)^2}{69} + \frac{(1988.5)^2}{74} - CF = 23.2194$$

٧- مجموع المربعات داخل الفئات (الخطأ) Within subclasses SS:

$$\text{Within subclasses SS} = \sum_{ijk} Y_{ijk}^2 - \sum_{ij} \frac{Y_{ij.}^2}{n_{ij}} = 105907.5 - 103186.3148$$
$$= 2721.1852$$

ويستكمل تحليل التباين باستعمال طريقة مواءمة الثوابت واستخدام تقديرات الثوابت السابق الحصول عليها.

٨- يحسب الاختزال فى مجموع المربعات لمواءمة جميع الثوابت μ، a_i، b_j حسب المعادلة (١٣-١٩):

$$R(\mu, a_i, b_j) = \hat{\mu}\, Y... + \sum_i \hat{a}_i\ Y_{i..} + \sum_j \hat{b}_j\ Y_{.j.} = \hat{\beta}(X'Y)$$

$$= \begin{bmatrix} 25.777 & 2.643 & 2.086 & 3.557 & 0.132 \end{bmatrix} \begin{bmatrix} 3787 \\ 796.5 \\ 687.5 \\ 485 \\ 190 \end{bmatrix}$$

$$= (25.77)(3787) + (2.643)(796.5) + \cdots + (0.132)(190) = 102907.6872$$

ويفضل فى الحسابات السابقة أن يتم استخدام قيم الثوابت بدون تقريب أى 6 أرقام عشرية على الأقل.

٩- يحسب الاختزال فى مجموع المربعات نتيجة لمواءمة العوامل غير المتوسط حسب المعادلة (١٣-٢٤):

$$R(a_i, b_j) = R(\mu, a_i, b_j) - CF = 102907.6872 - 100289.2937 = 2618.3935$$

١٠- مجموع المربعات للتداخل بين العاملين A، B:

$$\begin{aligned} \text{SS interaction (A x B)} &= R[a_i, b_j,] - R(a_i, b_j) \\ &= 2897.0211 - 2618.3935 = 278.6276 \end{aligned}$$

١١- مجموع المربعات للعامل A باستبعاد B أى $R(a_i)$:

$$\begin{aligned} \text{SSA eliminating B } &= R(a_i) = R(a_i, b_j) \text{ - SSB ignoring A} \\ &= 2618.3935 - 23.2194 = 2595.1741 \end{aligned}$$

١٢- مجموع المربعات للعامل B باستبعاد A أى $R(b_j)$:

$$\begin{aligned} \text{SSB eliminating A } &= R(b_j) = R(a_i, b_j) \text{ - SSA ignoring B} \\ &= 2618.3935 - 23.2194 = 2595.1741 \end{aligned}$$

ويمكن تلخيص النتائج فى جدول ١٣-٩ لتحليل التباين كالتالى:

جدول ١٣-٩ تحليل التباين النهائى لبيانات مثال ١٣-٨

SOV	df	SS	MS	F
Factor A, $R(a_i)$	$c-1=3$	2595.174	865.058**	42.9
Factor B, $R(b_j)$	$r-1=1$	2.462	2.462	< 1
Interaction A x B	(r-1)(c-1) = 3	278.628	92.88**	4.6
Error	n.. – rc = 135	2721.185	20.16	
Total	**n.. – 1 = 142**			

ويتضح من نتيجة اختبار F أنه ولو أن النموذج المفترض كان لا يحتوى على التأثير الخاص بالتداخل بين العاملين A، B إلا أن تأثيره كان معنوياً عند مستوى ($P<0.01$) وبالتالى فإنه كان من الأوفق عند افتراض وجود هذا التأثير فى النموذج مسبقاً أن تستخدم طريقة المربعات الموزونة للمتوسطات (الفصل ١٣-٧) method of weighted squares of means. أو أنه كنتيجة لوجود تداخل معنوى بين العوامل فإنه لا يصح اختبار التأثيرات الرئيسية للعوامل main effects بل أنه يجب استخدام طريقة المربعات الموزونة للمتوسطات لاستكمال تحليل التباين. ومن المهم أن يلاحظ أنه إذا كانت البيانات تحتوى على بعض الفئات subclasses الخالية من الأفراد أو الفارغة empty cells فإنه لا يمكن استخدام غير طريقة مواءمة الثوابت السابق ذكرها حيث أن الطرق الأخرى تحتاج إلى حساب الكمية $1/n_{ij}$ وهو ما لا يمكن استخدامه فى تلك الحالات.

وهناك الكثير من الإحصائيين ينصحون بأنه فى حالة استخدام طريقة مواءمة الثوابت تختبر معنوية التداخل بين العوامل ثم بناء على نتيجة هذا الاختبار يتم اختبار التأثيرات الرئيسية. وهو أن يتم اختبار متوسط مربعات التداخل على مستوى معنوية منخفض أى حوالى ($P<0.25$) حتى لا يحدث تأثيراً قوياً على مستوى معنوية الاختبارات اللاحقة للتأثيرات الرئيسية.

١٣-٧ طريقة المربعات الموزونة للمتوسطات

Method of weighted squares of means

كما ذكر سابقاً، فإنه إذا وجد أنه باختبار متوسط المربعات للتداخل بين العوامل كان معنوياً فإنه لا يصبح اختبار التأثيرات الرئيسية غير منحاز. وطريقة المربعات الموزونة للمتوسطات تقدم اختباراً غير منحاز للتأثيرات الرئيسية فى حالة معنوية

التداخل، ويتم الوزن حسب مقلوب تباين متوسطات الخلايا والذى يكون محصلة لعدد الأفراد فى كل خلية.

ويتم تحليل متوسطات الخلايا أو الفئات cell or subclass means حيث يعرف متوسط أى خلية

$$\overline{Y}_{ij.} = \sum_k Y_{ijk} / n_{ij} \qquad (١٣-٢٥)$$

والنموذج الرياضى المفترض فى هذه الحالة لكل قيمة يحتوى على التداخل بين التأثيرات الرئيسية أى $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \xi_{ijk}$ حيث $i = 1, 2, \cdots, r$، $j = 1, 2, \cdots, c$، $k = 1, 2, \cdots, n_{ij}$، وقد سبق الحديث عن الافتراضات الخاصة بهذا النموذج. وإتباع طريقة التحليل بالمربعات الموزونة للمتوسطات مبنيا على أساس ما أشير إليه سابقاً فى أن تباين متوسط القيم يساوى تباين القيمة الواحدة مقسوماً على عدد القيم أى أنه إذا افترض أن:

$$\overline{Y}_{ij.} = \left[\sum_j \left(\frac{Y_{ij\cdot}}{n_{ij}} \right) / c \right] \qquad (١٣-٢٦)$$

فإن التباين لهذا المتوسط

$$V(\overline{Y}_i..) = V\left[\sum_j \left(\frac{Y_{ij\cdot}}{n_{ij}} \right) / c \right] = \frac{1}{c^2} \left[\sum_j V\left(\frac{Y_{ij\cdot}}{n_{ij}} \right) \right] \qquad (١٣-٢٧)$$

وبافتراض أن التباينات للخلايا كلها متجانسة homogeneous وقيمة كل منها σ^2 فإن (١٣-٢٧) تصبح:

$$\frac{1}{c^2} \sum_j \frac{\sigma^2}{n_{ij}} = \frac{\sigma^2}{c^2} \sum_j \frac{1}{n_{ij}} = \frac{\sigma^2}{\omega_i} \qquad (١٣-٢٨)$$

حيث

$$\frac{1}{\omega_i} = \frac{1}{c^2} \sum_j \frac{1}{n_{ij}} \qquad (١٣-٢٩)$$

وتمثل c عدد الأعمدة أى مستويات العامل B، وتعرف القيمة ω_i (وتنطق Omega) على أنها الوزن الخاص بالمتوسط $\overline{Y}_{i..}$ وسوف يتم استخدام هذا الوزن لحساب مجموع المربعات للتأثير الرئيسى للعامل A.

وبنفس الطريقة توجد قيمة:

$$V(\overline{Y}_{.j.}) = \sigma^2 / \nu_j \qquad (١٣-٣٠)$$

حيث

$$\frac{1}{\nu_j} = \frac{1}{r^2} \sum_i \frac{1}{n_{ij}} \qquad (١٣-٣١)$$

وتمثل r عدد الصفوف أو مستويات العامل A، وتكون ν_j (وتنطق Nu) هى الوزن الخاص بالمتوسط $\overline{Y}_{.j.}$ وتستخدم فى حساب مجموع المربعات للتأثير الرئيسى للعامل B.

وبناء على ذلك فإنه للحصول على مجموع المربعات الخاص بالتأثيرات الرئيسية بالنسبة للعامل A فإن مجموع المربعات الموزون الخاص به يمكن حسابه كالتالى:

$$SSA = \sum_i \omega_i \overline{Y}_{i..}^2 - \frac{(\sum_i \omega_i \overline{Y}_{i..})^2}{\sum_i \omega_i} \qquad (١٣-٣٢)$$

أما مجموع المربعات الموزون للعامل B فيكون كالتالى:

$$SSB = \sum_j \nu_j \overline{Y}_{.j.}^2 - \frac{(\sum_j \nu_j \overline{Y}_{.j.})^2}{\sum_j \nu_j} \qquad (١٣-٣٣)$$

ويلاحظ أن عملية الوزن باستخدام مقلوب الأعداد لكل فئة تؤدى إلى أن المشاهدات التى يكون تباينها أقل (نظراً لكثرة العدد فى الفئة) تكون الدقة فى تقديرها أكبر وبالتالى فإنه يلزم إعطاؤها وزناً أكبر عند حساب مجموع المربعات.

ولاستكمال تحليل التباين فإن التحليل يبدأ باستخدام طريقة مواءمة الثوابت المذكورة سابقاً، ويحسب من هذا مجموع المربعات بين الفئات between subclasses sum of squares كما فى (١٣-١١) ثم يحسب الاختزال فى مجموع المربعات كنتيجة لمواءمة تأثيرى كل من العاملين الرئيسين A، B كما فى (١٣-٢٤) وبالتالى يمكن الحصول على مجموع المربعات للتداخل بطرح الثانى من الأول كما سبق. بعد ذلك يتم اختبار متوسط المربعات للتداخل، فإن وجد أنه معنوى لا يستكمل التحليل بطريقة مواءمة الثوابت بل يتم الحصول على مجاميع المربعات للتأثيرات الرئيسية بالطريقة التى شرحت فى هذا الجزء. هذا مع العلم بأنه فى أى من الطرق لابد من حساب مجموع المربعات داخل الفئات within-subclasses sum of squares أولاً كما فى (١٣-١٤) حتى يمكن اختبار معنوية التداخل باستخدام متوسط المربعات داخل الفئات، ويستكمل تحليل التباين بعد ذلك بالطريقة العادية لاختبار متوسط المربعات لكل من التأثيرين الرئيسين.

صندوق ١٣-٢

طريقة موائمة الثوابت وطريقة المربعات الصغرى الموزونة: هما طريقتان مضبوطتان لحساب تحليل التباين، الأولى عندما يكون التداخل بين العوامل الرئيسية غير معنوى والثانية عندما يكون هذا التداخل معنوياً.

والفكرة فيهما أنهما يقدران المعالم بما يحقق أن الخطأ فى النموذج هو أقل ما يمكن لذا أطلق عليهما "أقل مربعات للخطأ" وإن كانت الطريقتان كثيرة المعادلات والحساب ولكن أساسهما الرياضى بسيط ويجرى التحليل بواسطة الحواسب فى ومضة عين.

مثال ١٣-٤

باستخدام البيانات الموجودة فى جدول ١٣-٨ والخاصة بأوزان الحملان عند الفطام يتضح من تحليل البيانات باستخدام طريقة مواءمة الثوابت (١٣-١٥) أن مجموع المربعات للتداخل المحسوب كان 278.6276 وله 3 درجات حرية، وبالتالى فإن متوسط المربعات للتداخل كان 92.88 وهى قيمة معنوية على مستوى (P<0.01) كما هو واضح من جدول ١٣-٩، وعليه فإن هذا يستوجب عدم استكمال تحليل البيانات بهذه الطريقة لاختبار التأثيرات الرئيسية للعوامل. ويجب استخدام طريقة المربعات الموزونة للمتوسطات، ويوضح جدول ١٣-١٠ الحسابات المطلوبة لاستكمال تحليل البيانات.

وبناء على نتائج جدول ١٣-١٠ فإنه يمكن حساب مجموع المربعات للتأثيرات الرئيسية للعاملين على النحو التالى:

١- باستخدام المعادلة (١٣-٣٢).

$$\text{Breed SS} = 101587.887 - \frac{(3737.468)^2}{141.14} = 2617.595$$

٢- بتطبيق المعادلة (١٣-٣٣)

$$\text{Sex SS} = 89800.014 - \frac{(3483.224)^2}{135.11} = 3.083$$

ويمكن تلخيص نتائج تحليل التباين فى الجدول التالى:

SOV	df	SS	MS	F
Breed (B)	3	2617.595	872.532	43.28**
Sex (S)	1	3.083	3.083	< 1
B x S	3	278.628	92.876	4.61**
Within	135	2721.185	20.157	
Total	**142**			

جدول ١٣-١٠ الحسابات المطلوبة لاستكمال تحليل التباين للبيانات التى فى جدول ١٣-٨

السلالة		إناث	ذكور	$\sum_j(1/n_{ij})$	ω_i	$\overline{Y}_{i\cdot\cdot}$	$(\omega_i)(\overline{Y}_{i\cdot\cdot})$	$(\omega_i)(\overline{Y}_{i\cdot\cdot})^2$
	n_{ij}	24	20					
العواسى	$1/n_{ij}$	0.04167	0.05	0.09167	43.63478	28.575	1246.8638	35629.133
	$\overline{Y}_{ij\cdot}$	27.0	30.15					
	n_{ij}	19	22					
النجدى	$1/n_{ij}$	0.05263	0.04545	0.09808	40.78303	27.9749	1140.9011	35629.133
	$\overline{Y}_{ij\cdot}$	29.6316	26.31818					
	n_{ij}	19	13					
الدوربر	$1/n_{ij}$	0.05263	0.07692	0.12955	30.87611	29.1265	899.3130	26193.84
	$\overline{Y}_{ij\cdot}$	30.3684	27.8846					
	n_{ij}	12	14					
الحرى	$1/n_{ij}$	0.08333	0.07143	0.15476	25.84647	17.4256	450.3902	7848.3201
	$\overline{Y}_{ij\cdot}$	16.7083	18.1429					
$\sum_j(1/n_{ij})$		0.23026	0.2438	المجموع	141.14039		3737.468	101587.887
v_j		69.4867	65.6276	135.1143				
$\overline{Y}_{\cdot j\cdot}$		25.9271	25.6239					
$(v_j)(\overline{Y}_{\cdot j\cdot})$		1801.5886	1681.635	3483.224				
$(v_j)(\overline{Y}_{\cdot j\cdot})^2$		46709.967	43090.047	89800.014				

مثال ١٣-٥

استخدام برنامج SAS لحل بيانات الجدول التالي والتي تمثل عدد الأيام اللازمة لإنبات germination ثلاثة أنواع varieties من بذور نبات معين في نوعين من التربة soil باستخدام عدد مختلف من صواني الإنبات pots (Searle, 1987).

نوع التربة	نوع بذور النبات ١	٢	٣
١	6	13	14
	10	15	22
	11		
٢	12	31	18
	15		9
	19		12
	18		

```
DATA GERMIN;
INPUT VARIETY SOIL DAYS @@;
CARDS;
1 1 6 1 1 10 1 1 11 2 1 13 2 1 15 3 1 14 3 1 22
1 2 12 1 2 15 1 2 19 1 2 18 2 2 31 3 2 18 3 2 9 3 2 12
PROC GLM;
CLASS VARIETY SOIL;
MODEL DAYS = VARIETY SOIL VARIETY*SOIL/SS3;
LSMEANS VARIETY SOIL VARIETY*SOIL/STDERR;
RUN;
```

لاحظ:

لابد من استخدام SS3 (النوع الثالث لمجموع المربعات) في حالة عدم تساوي التكرارات داخل الفئات والذي يمثل الحل المضبوط بطريقة الاختزال reduction.

استخدام اختيار LSMEANS للحصول على متوسطات أقل مجموع مربعات والتي تكون أدق من المتوسط الحسابي وهي تعني حساب القيمة المتوقعة لمتوسطات الفئات كما لو كانت البيانات متزنة أخذا في الاعتبار القيمة المتوسطة للعوامل الرئيسية. استخدام اختيار STDERR للحصول على الخطأ المعياري لهذه المتوسطات.

النتائج التحليل

The GLM Procedure

Class Level Information

Class	Levels	Values
VARIETY	3	1 2 3
SOIL	2	1 2

Number of observations 15

Dependent Variable: DAYS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	400.0000000	80.0000000	6.00	0.0103
Error	9	120.0000000	13.3333333		
Corrected Total	14	520.0000000			

R-Square	Coeff Var	Root MSE	DAYS Mean
0.769231	24.34322	3.651484	15.00000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
VARIETY	2	192.1276596	96.0638298	7.20	0.0135
SOIL	1	123.7714286	123.7714286	9.28	0.0139
VARIETY*SOIL	2	222.7659574	111.3829787	8.35	0.0089

Least Squares Means

VARIETY	DAYS LSMEAN	Standard Error	Pr > |t|
1	12.5000000	1.3944334	<.0001
2	22.5000000	2.2360680	<.0001
3	15.5000000	1.6666667	<.0001

SOIL	DAYS LSMEAN	Standard Error	Pr > |t|
1	13.6666667	1.4054567	<.0001
2	20.0000000	1.5315610	<.0001

VARIETY	SOIL	DAYS LSMEAN	Standard Error	Pr > |t|
1	1	9.0000000	2.1081851	0.0021
1	2	16.0000000	1.8257419	<.0001
2	1	14.0000000	2.5819889	0.0004
2	2	31.0000000	3.6514837	<.0001
3	1	18.0000000	2.5819889	<.0001
3	2	13.0000000	2.1081851	0.0002

١٣-٨ تنويه عن تحليل البيانات غير متناسبة التكرار

تبين فيما سبق مدى تعقيد تحليل البيانات غير المتزنة وإن كانت الحاسبات وبرمجياتها قد يسرت كثيرا هذه التحليلات. ولكن يلزم الحذر عند تفسير البيانات غير متناسبة التكرار من حيث ما إذا كان الاختلاف فى التكرار يعكس التكرار الفعلى فى العشيرة أم إنه نتيجة لأثر العوامل المختلفة محل الدراسة كأن تؤدى معاملة معينة إلى ارتفاع نسبة النفوق مثلا. ويستلزم الأمر كثيرا من الحذر إذا كانت هناك خلايا فارغة empty cells ليست بها مشاهدات وهذا يؤثر على تقدير التداخلات بين العوامل وقد يصل الأمر إلى التأثير أيضا على تقدير العوامل الرئيسية main effects مما قد يستوجب تجزئة البيانات إلى أجزاء يحلل كل منها مستقلا أو بطريقة ما بين التحليل العنقودى nested analysis والتحليل المتعامد crossed analysis، والمثال التالى يوضح هذا المفهوم.

مثال ١٣-٦

افترض وجود عاملان A، B وبكل منهما ٤ مستويات. وفى الجدول التالى الخلايا التى بها علامة x هى خلايا بها مشاهدات أما الخلايا التى لا يوجد بها هذه العلامة فهى خلايا فارغة لا يوجد بها أى مشاهدات.

	A_4	A_3	A_2	A_1		
جزء I	x	x			B_1	
	x	x			B_2	
			x	x	B_3	جزء II
			x	x	B_4	

مثل هذه البيانات لا يمكن تحليلها كوحدة واحدة حيث لا يمكن مقارنة A_1 أو A_2 مع A_3 أو A_4 لأن المقارنتين A_1، A_2 يقعان تحت مستويات مختلفة من

B (B_3، B_4) والمقارنتين A_3، A_4 يقعان تحت B_1، B_2. ولكن يمكن مقارنة A_1 مع A_2 وكذلك مقارنة A_3 مع A_4. وتسمى مثل هذه البيانات بالبيانات غير المتصلة disconnected data والتى تتبع ظاهرة تسمى "الاتصالية" connectedness.

ويتم تحليل مثل هذه البيانات بتقسيمها إلى جزئين وليكن جزء I ويشمل أربع خلايا هى A_3B_1، A_3B_2، A_4B_1، A_4B_2 وجزء II ويشمل أربع خلايا هى A_1B_3، A_1B_4، A_2B_3، A_2B_4. ويكون التحليل كما يلى:

الجزء الأول I

SOV	df
A	1
B	1
A x B	1
الكلى	3

الجزء الثانى II

SOV	df
A	1
B	1
A x B	1
الكلى	3

ويمكن ضم التحليلين معا ليصبحا

SOV	df
A داخل الأجزاء	2
B داخل الأجزاء	2
A x B داخل الأجزاء	2

وهذا ما أطلق عليه خليط بين التحليل العنقودى والتحليل المتعامد.

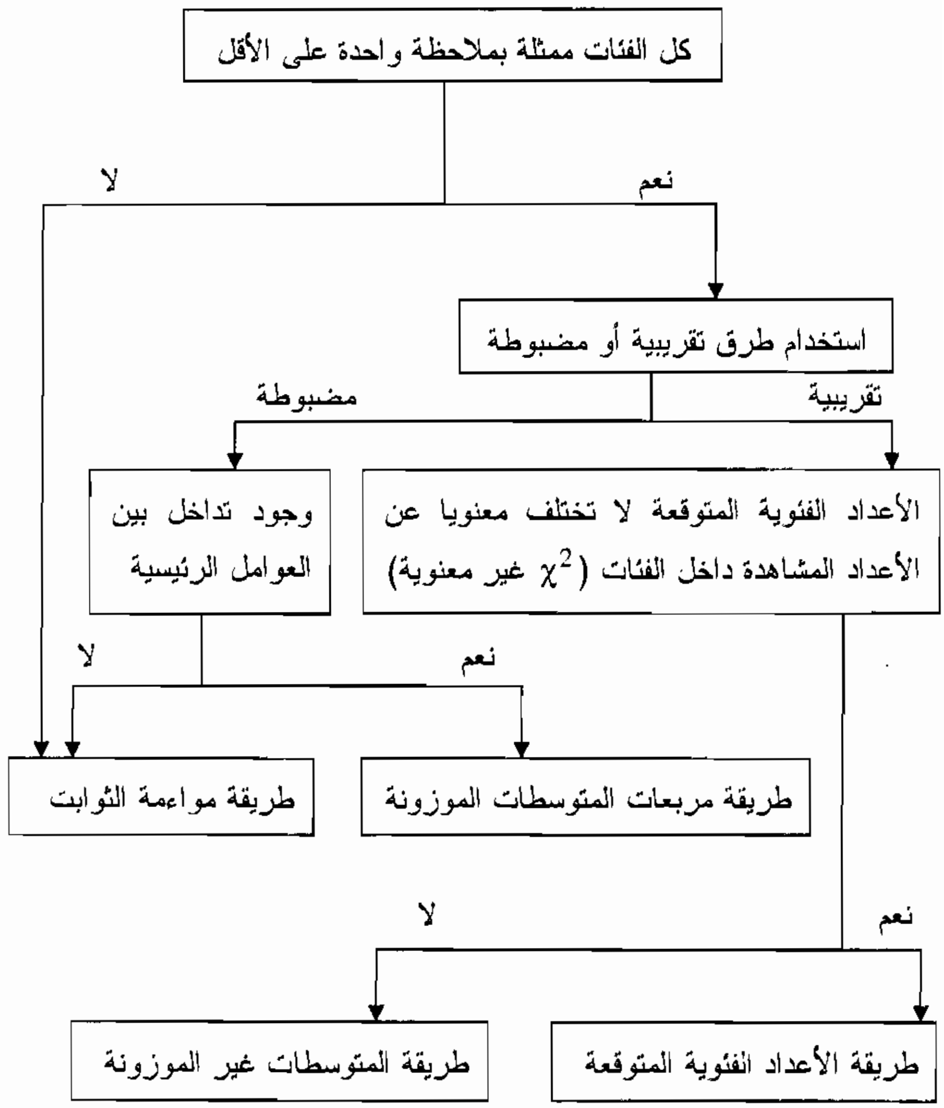
أما إذا فرض أن الخلية A_2B_2 أو A_3B_3 بأى منهما ولو فرد واحد تصبح البيانات متصلة connected data فإنه يجوز مقارنة كل العوامل السابقة ولكن ليست التداخلات كلها ويصبح التحليل:

SOV	df
A	3
B	3
A x B	2
الكلي	8

لاحظ أن درجات الحرية للتداخل ليست حاصل ضرب درجات حرية العوامل الرئيسية بسبب وجود الخلايا الفارغة.

أما إذا زاد الاتصال بوجود خلايا أخرى بها مشاهدات فإنه يمكن تقدير بعض التداخلات بدرجة أكبر ... وهكذا. وكلما زاد عدد الخلايا المحتوية على مشاهدات زادت الاتصالية وزاد التحليل رصانة وأمكن تقدير عدد أكبر من المعالم الإحصائية. ولمزيد من التفاصيل يمكن الرجوع إلى Searle, 1987.

ويبين شكل ١٣-١ هيكل توضيحى للطرق المستخدمة فى تحليل التباين فى حالة عدم تساوى تكرار الفئات.



شكل ١٣-١ هيكل توضيحى للطرق المستخدمة فى تحليل التباين فى حالة عدم تساوى تكرار الفئات

تمارين الباب الثالث عشر

١٣-١ البيانات المدرجة فى الجدول التالى تمثل متوسط وزن الذكور بالكيلوجرام عند عمر 8 أسابيع لأربع سلالات من الدجاج ناتجة من 3 دفعات فى أوقات مختلفة من السنة (لاحظ اختلاف عدد الطيور التى تصل إلى نهاية التجربة).

السلالة		دفعة الفقس		
		١	٢	٣
أ	الوزن	1.317	1.493	1.425
	العدد	18	11	11
ب	الوزن	1.641	1.675	1.471
	العدد	22	9	6
ج	الوزن	1.879	2.128	1.894
	العدد	17	8	9
د	الوزن	1.591	1.811	1.713
	العدد	26	7	18

فإذا كان مجموع المربعات الكلى المصحح 13.1525 . كون واملأ جدول تحليل التباين بطريقة المتوسطات غير الموزونة وقارنها بطريقة الأعداد المتوقعة للفئات.

١٣-٢ جربت ثلاث طرق محددة لتدريس الإحصاء على مجموعة من الطلبة والطالبات وفى نهاية التجربة كانت نتائج الاختبار كما يلى:

	طلبة	طالبات
طريقة ١	5، 9	7، 8، 3
طريقة ٢	4، 5، 7	6، 4، 9
طريقة ٣	2، 1، 4	

كون واملأ جدول تحليل التباين. اجر اختبارات المعنوية المختلفة عند مستوى معنوية 5% مع كتابة النموذج الإحصائى.

١٤ الانحدار الخطى المتعدد

Multiple linear regression

١ – مقدمة

٢ – الانحدار المتعدد فى حالة متغيرين مستقلين

٣ – حدود الثقة لمعاملى الانحدار β_1 ، β_2

٤ – تقسيم الاختلافات فى Y إلى مكوناتها

٥ – تحليل التباين واختبار F

٦ – اختيار أفضل معادلة انحدار

٧ – الارتباط المتعدد والجزئى

١٤-١ مقدمة

عند دراسة الانحدار الخطى البسيط كان هناك متغير مستقل واحد (X) يؤثر على متغير تابع (Y)، ولكن كثيراً ما تكون هناك عدة متغيرات مستقلة تؤثر على المتغير التابع. فعند دراسة العوامل التى تؤثر على كمية اللبن مثلاً أو وزن الحملان عند الفطام أو غيرهما يكون هناك عوامل كثيرة وليس عاملاً واحداً. فإذا ما أخذ فى الاعتبار متغير مستقل واحد فقط وأهملت باقى المتغيرات المؤثرة المستقلة فهذا من شأنه أن يؤدى إلى زيادة مكون الخطأ مما قد يؤدى إلى نتائج بعيدة عن الدقة حيث تفقد اختبارات المعنوية حساسيتها وذلك راجع إلى زيادة خطأ التقدير $S_{y.x}$ السابق الإشارة إليه فى الباب العاشر. وحيث أن اختبار t يعبر عن النسبة بين معامل الانحدار (b) إلى الانحراف المعيارى له $[(S_{y.x})/(\sqrt{\sum x^2})]$، والتى تقل بزيادة الانحراف المعيارى، مما قد يؤدى إلى نتائج غير معنوية. فى حين أنها قد تكون معنوية إذا لم يتم إغفال المتغيرات الأخرى. إلا أن هناك نقطة أخرى يجب وضعها فى الاعتبار وهى درجات الحرية، فزيادة عدد المتغيرات المستقلة يؤدى إلى نقص فى درجات حرية الخطأ وهذا أيضاً يؤدى إلى نقص فى دقة الاختبار إذا ما كانت بعض هذه المعاملات غير معنوية. والتركيز فى هذا الباب سوف يكون على دراسة الانحدار والارتباط المتعددين اللذين يختصان بوصف العلاقات الخطية بين أكثر من متغيرين لتقدير التأثير المتحد combined effect لعدة متغيرات مستقلة على متغير واحد، ودراسة إمكانية التنبؤ بواسطة معادلة الانحدار التى تعطى أدق قيمة للمتغير التابع (Y) بدلالة المتغيرات المستقلة المعروفة، مع إهمال بعض العوامل المستقلة ذات التأثير الضئيل وغير المعنوى والتى بإهمالها لن تتأثر نتيجة التنبؤ. وكذلك دراسة إمكانية ترتيب العوامل المستقلة حسب أهميتها بناءاً على اختبارات المعنوية.

وسوف تتناول الدراسة فى هذا الباب العلاقة بين ثلاث متغيرات فقط (اثنين مستقلين وآخر تابع) وذلك بهدف إعطاء نبذة مختصرة عن طريقة التقدير ومفهوم كل من معاملات الانحدار المتعددة والجزئية وكذلك كل من معاملات الارتباط الكلية والجزئية واختبارات المعنوية، علما بأنه فى حالة أكثر من ثلاث متغيرات فليس هناك من مبادئ جديدة ولكنها امتداد لها.

١٤-٢ الانحدار المتعدد فى حالة متغيرين مستقلين

يمكن التعبير عن المتغير التابع Y الذى يؤثر فيه متغيرين مستقلين X_1، X_2 بالنموذج التالى:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \qquad (١٤-١)$$

حيث β_1 تمثل معامل الانحدار (الاعتماد) الجزئى partial regression coefficient للمتغير Y على المتغير X_1، أى متوسط مقدار التغير فى Y عندما تتغير X_1 بمقدار الوحدة مع ثبات X_2. وتمثل β_2 معامل الانحدار الجزئى للمتغير Y على المتغير X_2، أى متوسط مقدار التغير فى Y عندما تتغير X_2 بمقدار الوحدة مع ثبات X_1. أما بقية مكونات النموذج فهى كما سبق تعريفها فى الباب العاشر.

ويعرف الانحدار المتعدد multiple regression بأنه متوسط التغير فى Y بمقدار β_1، β_2 نتيجة تغير كل من X_1، X_2 بمقدار الوحدة، على الترتيب.

ولتقدير معالم (ثوابت) parameters هذا النموذج (١٤-١) يستخدم النموذج التقديرى التالى:

$$\hat{Y} = a + b_1X_1 + b_2X_2 \qquad (١٤-٢)$$

حيث a، b_1، b_2 تقديرات غير متحيزة لكل من α، β_1، β_2 على الترتيب. ويستخدم فى ذلك طريقة المربعات الصغرى التى سبق الإشارة إليها وبالتالى:

$$Y = a + b_1X_1 + b_2X_2 + e \qquad (١٤-٣)$$

١٤-٢-١ طريقة المربعات الصغرى Method of least squares

من النموذج (١٤-١):

$$\varepsilon = (Y - \alpha - \beta_1X_1 - \beta_2X_2)$$

$$\varepsilon^2 = \sum(Y - \alpha - \beta_1X_1 - \beta_2X_2)^2$$

$$Q = \sum\varepsilon^2 = \sum(Y - \alpha - \beta_1X_1 - \beta_2X_2)^2 \qquad (١٤-٤)$$

حيث $\sum\in^2$ تعبر عن مجموع مربعات الخطأ، أى مجموع مربع الانحرافات عن خط الانحدار. وتقدر α، β_1، β_2 بحيث يتحقق الشرطان التاليان:

١ - مجموع الأخطاء يساوى صفر، أى $\sum\varepsilon = 0$

٢- مجموع مربعات الخطأ $\sum\in^2$ أقل ما يمكن، ويكون ذلك باستخدام التفاضل الجزئى للمقدار $\sum\in^2$ بالنسبة لكل من α، β_1، β_2 كل على حدة، وفى كل مرة توضع نتيجة التفاضل مساوية للصفر كما يلى:

$$\frac{\partial Q}{\partial \alpha}\Big|_{a,b_1,b_2} = 2\sum(Y - a - b_1X_1 - b_2X_2)(-1) = 0$$

$$\frac{\partial Q}{\partial \beta_1}\Big|_{a,b_1,b_2} = 2\sum(Y - a - b_1X_1 - b_2X_2)(-X_1) = 0$$

$$\frac{\partial Q}{\partial \beta_2}\Big|_{a,b_1,b_2} = 2\sum(Y - a - b_1X_1 - b_2X_2)(-X_2) = 0$$

وتتكون مجموعة المعادلات التالية:

$$\begin{aligned}\sum Y &= na + b_1\sum X_1 + b_2\sum X_2\\ \sum X_1Y &= a\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2\\ \sum X_2Y &= a\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2\end{aligned} \qquad (١٤-٥)$$

وتسمى هذه المجموعة من المعادلات بالمعادلات الاعتيادية أو الطبيعية normal equations. وبحل هذه المعادلات آنياً يمكن الحصول على تقدير لكل من α، β_1، β_2 كالتالى:

$$a = \overline{Y} - b_1\overline{X}_1 - b_2\overline{X}_2 \qquad (١٤-٦)$$

وبالتعبير عن X، Y كانحرافات عن متوسطاتهما فإن:

$$b_1 = \frac{\sum x_1y\sum x_2^2 - \sum x_2y\sum x_1x_2}{\sum x_1^2\sum x_2^2 - (\sum x_1x_2)^2} \qquad (١٤-٧)$$

$$b_2 = \frac{\sum x_2y\sum x_1^2 - \sum x_1y\sum x_1x_2}{\sum x_1^2\sum x_2^2 - (\sum x_1x_2)^2} \qquad (١٤-٨)$$

ويمكن إعادة كتابة المعادلة (١٤-٢) باستخدام جبر المصفوفات كالتالى:

$$Y = X\beta + \varepsilon \qquad (١٤-٩)$$

حيث:

$Y' = [Y_1 \quad Y_2 \quad \cdots \quad Y_n]$ ، $\varepsilon' = [\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_n]$ ،

$$X = \begin{bmatrix} 1 & X_{11} & X_{21} \\ 1 & X_{12} & X_{22} \\ \vdots & \vdots & \vdots \\ 1 & X_{1n} & X_{2n} \end{bmatrix} \quad ,\beta' = [\alpha \quad \beta_1 \quad \beta_2]$$

ومن (١٤-٩):

$$Q = \sum \varepsilon^2 = \varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta) = Y'Y - \beta'X'Y - Y'X\beta + \beta'X'X\beta$$
$$= Y'Y - 2\beta'X'Y + \beta'X'X\beta$$

حيث إن $Y'X\beta = \beta'X'Y$ لأن كل منهما عبارة عن مصفوفة $1x1$.

وباستخدام طريقة المربعات الصغرى فإن:

$$\frac{\partial Q}{\partial \beta_2}|_{\hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0$$

ومنها:

$$X'X\hat{\beta} = X'Y$$

وبالتالى:

$$\hat{\beta} = (X'X)^{-1}X'Y \qquad (١٤-١٠)$$

حيث إن $(X'X)^{-1}$ عبارة عن مقلوب $(X'X)$ كما سبق شرحه فى الباب العاشر.

١-٢-٢ الافتراضات الخاصة بدراسة العلاقة بين أكثر من متغيرين

١ - المتغيرات المستقلة مقاسة بدون خطأ وليس لها توزيعات احتمالية وعددها يكون أقل من حجم العينة (عدد المشاهدات).

٢ - الأخطاء فى قيم المتغير التابع مستقلة وتتوزع طبيعياً بمتوسط يساوى الصفر وتباين يساوى σ_e^2. وهذا الافتراض مهم وضرورى فى حالات اختبارات المعنوية.

٣ - لا يوجد أى ارتباط بين أى من الأخطاء وكل من المتغيرات المستقلة. أى أن $cov(X_i, e_j) = 0$ لكل من i ، j.

مثال ١٤-١

من البيانات التالية أحسب معادلة الانحدار المتعدد.

4	7	8	6	2	:Y
5	4	3	2	0	:X_1
4	3	0	1	2	:X_2

من البيانات السابقة يمكن حساب ما يلى:

$\sum X_1 = 14$ $\quad \sum X_2 = 10$ $\quad \sum Y = 27$

$\sum X_1^2 = 54$ $\quad \sum X_2^2 = 30$ $\quad \sum Y^2 = 169$

$\sum x_1^2 = 14.8$ $\quad \sum x_2^2 = 10$ $\quad \sum y^2 = 23.2$

$\sum X_1X_2 = 34$ $\quad \sum X_1Y = 84$ $\quad \sum X_2Y = 47$

وبالتعويض فى المعادلات الاعتيادية فى (١٤-٥):

$$5a + 14b_1 + 10b_2 = 27$$

$$14a + 54b_1 + 34b_2 = 84$$

$$10a + 34b_1 + 30b_2 = 47$$

ومن هذه المعادلات:

$a = 5$

$b_1 = 1.125$ وحدة من Y لكل وحدة من X_1 باعتبار X_2 ثابتة.

$b_2 = -1.375$ وحدة من Y لكل وحدة من X_2 باعتبار X_1 ثابتة.

وبالتالى فإن معادلة التنبؤ:

$$\hat{Y} = 5 + 1.125X_1 - 1.375X_2$$

وإذا اعتبر أن b_1، b_2 معاملا انحدار بسيط فإن معامل انحدار Y على X_1

$$b_{y.x_1}=\frac{\Sigma X_1Y-\frac{\Sigma X_1\Sigma Y}{n}}{\Sigma X_1^2-\frac{(\Sigma X_1)^2}{n}}=\frac{84-\frac{(14)(27)}{5}}{54-\frac{(14)^2}{5}}=0.57$$

أى 0.57 وحدة من Y لكل وحدة من X_1. بينما معامل الانحدار الجزئى بين Y و X_1 هو 1.125 مع تثبيت X_2.

وبنفس الطريقة معامل انحدار Y على X_2:

$$b_{y.x_2}=\frac{\Sigma X_2Y-\frac{\Sigma X_2\Sigma Y}{n}}{\Sigma X_2^2-\frac{(\Sigma X_2)^2}{n}}=\frac{47-\frac{(10)(27)}{5}}{30-\frac{(10)^2}{5}}=-0.7$$

أى 0.7- وحدة من Y لكل وحدة من X_2. بينما معامل الانحدار الجزئى بين Y، X_2 باعتبار X_1 ثابتة هو 1.375-. ومن ذلك يتضح أن معاملات الانحدار البسيطة مختلفة القيمة عن معاملات الانحدار الجزئية.

١٤-٣ حدود الثقة لمعاملى الانحدار β_1 ، β_2

حدى الثقة لــ β_1 هما:

$$b_1 \pm tS_{b_1} \qquad (١١-١٤)$$

حدى الثقة لــ β_2 هما:

$$b_2 \pm tS_{b2} \qquad (١٢-١٤)$$

حيث:

$$S_{b_1}=\sqrt{\frac{S^2\sum x_2^2}{\sum x_1^2\sum x_2^2-(\sum x_1x_2)^2}} \text{ ، } S_{b_2}=\sqrt{\frac{S^2\sum x_1^2}{\sum x_1^2\sum x_2^2-(\sum x_1x_2)^2}} \text{ ،}$$

S_b، S_{b_2} هما الانحراف القياسى standard error لكل من معاملى الانحدار b_1، b_2 على الترتيب، t هى قيمة t الجدولية بدرجات حرية $(n-k-1)$ حيث n هى

عدد المشاهدات، k عدد معاملات الانحدار المقدرة ومستوى معنوية وليكن α، S^2 هى متوسط مربعات الخطأ.

١٤-٤ تقسيم الاختلافات فى Y إلى مكوناتها

حيث إن:

$$Y = a + b_1X_1 + b_2X_2 + e$$

$$\hat{Y} = a + b_1X_1 + b_2X_2$$

فإنه يمكن أيضاً التعبير عن Y بعد التعويض عن a بقيمتها حيث:

$$Y = \overline{Y} + b_1(X_1 - \overline{X}_1) + b_2(X_2 - \overline{X}_2) + e$$

وأيضا فإن:

$$\hat{Y} = \overline{Y} + b_1(X_1 - \overline{X}_1) + b_2(X_2 - \overline{X}_2)$$

وعلى ذلك فإن:

$$Y - \overline{Y} = b_1(X_1 - \overline{X}_1) + b_2(X_2 - \overline{X}_2) + e$$

$$\hat{Y} - \overline{Y} = b_1(X_1 - \overline{X}_1) + b_2(X_2 - \overline{X}_2)$$

حيث إن:

$(Y - \overline{Y})$ عبارة عن انحراف قيمة Y عن متوسطها الحسابى.

$(\hat{Y} - \overline{Y})$ هو انحراف القيمة المتوقعة المقابلة لقيمة معينة لـــ X عن المتوسط الحسابى للمتغير Y.

$(Y - \hat{Y})$ هو انحراف القيمة Y عن القيمة المتوقعة.

أى أنه يمكن تقسيم انحراف القيمة عن متوسطها إلى مكونين أحدهما راجع إلى انحدار (اعتماد) Y على كل من X_1، X_2 والذى يطلق عليه الجزء الراجع للانحدار due to regression، والمكون الأخر راجع إلى الانحراف عن خط الانحدار والذى يطلق عليه عن خط الانحدار from regression، ويعبر عن $(Y - \hat{Y})$ أو (e) أو $(d_{y \cdot x_1 x_2})$.

مثال ١٤-٢

قسم الاختلافات فى Y إلى مكوناتها فى المثال ١٤-١

$(\hat{Y}-\overline{Y})$	$(Y-\hat{Y})$	$(Y-\overline{Y})$	$\hat{Y}$	X_2	X_1	Y	
-3.150	-0.250	-3.4	2.250	2	0	2	
0.475	0.125	0.6	5.875	1	2	6	
2.975	-0.375	2.6	8.375	0	3	8	
-0.025	1.625	1.6	5.375	3	4	7	
-0.275	-1.125	-1.4	5.125	4	5	4	
0	**0**	**0**	**27.000**	**10**	**14**	**27**	**المجموع**

١٤-٥ تحليل التباين واختبار F

تعتبر الكمية $\sum(Y-\hat{Y})^2/(n-k-1)$ (والتى يرمز لها بالرمز S^2 ويعبر عنها بمتوسط مربعات الخطأ أو بمتوسط المربعات الراجع إلى الانحراف عن خط الانحدار) تقديرا غير متحيز لـ σ_e^2 وذلك فى الانحدار المتعدد حيث يقسم مجموع مربعات انحرافات قيم المتغير التابع Y عن متوسطها إلى مكونين:

١ - مجموع المربعات الراجع إلى خط الانحدار SS due to regression

٢ - مجموع المربعات عن خط الانحدار SS from regression

وتبين المعادلة التالية ذلك حيث إن:

$$(Y-\overline{Y})=(\hat{Y}-\overline{Y})+(Y-\hat{Y})$$

ومنها:

$$\begin{aligned}\sum(Y-\overline{Y})^2 &= \sum(\hat{Y}-\overline{Y})^2+\sum(Y-\hat{Y})^2+2\sum(\hat{Y}-\overline{Y})(Y-\hat{Y})\\ &= \sum(\hat{Y}-\overline{Y})^2+\sum(Y-\hat{Y})^2\end{aligned}$$

حيث إن الحد الأخير $[2\sum(\hat{Y}-\overline{Y})(Y-\hat{Y})]$ يساوى الصفر.

أى أن:

$$TSS = RSS + ESS$$

حيث:

$$TSS=\sum y^2=\sum(Y-\overline{Y})^2$$

$$RSS=\sum(\hat{Y}-\overline{Y})^2=\sum\hat{y}^2=b_1\sum x_1y+b_2\sum x_2y$$

$$ESS=\sum(Y-\hat{Y})^2=\sum y^2-RSS$$

مثال ١٤-٣

قسم مجموع المربعات فى Y إلى مكوناته فى مثال ١٤-٢.

مجموع المربعات الكلى TSS:

$$\sum y^2 = 169 - \frac{(27)^2}{5} = 23.2$$

ومجموع المربعات الراجع إلى خط الانحدار RSS:

$$\begin{aligned} RSS &= \sum(\hat{Y} - \overline{Y})^2 \\ &= b_1 \sum x_1 y + b_2 \sum x_2 y \\ &= b_1\left[\sum X_1 Y - \frac{\sum X_1 \sum Y}{n}\right] + b_2\left[\sum X_2 Y - \frac{\sum X_2 \sum Y}{n}\right] \\ &= 0.57[84 - (14)(27)/5] - 0.7[47 - (10)(27)/5] = 19.075 \end{aligned}$$

مجموع المربعات الراجع إلى الانحراف عن خط الانحدار ESS:

$$ESS = 23.2 - 19.075 = 4.125$$

وكما كان الحال فى الاعتماد البسيط فإنه يمكن الحصول على ESS أيضاً بالحساب حيث إن: $ESS = \sum(Y - \hat{Y})^2$، ومن مثال ١٤-٢ فإن:

$$ESS = (-0.25)^2 + (0.125)^2 + \cdots + (-1.125)^2 = 4.125$$

وهى نفس النتيجة التى تم الحصول عليها بالطرح دون الرجوع إلى تقدير القيم المتوقعة ثم حساب انحرافات القيم الفعلية عن قيمها المتوقعة والتربيع والجمع على كل مفردات العينة.

ولاختبار معنوية الانحدار المتعدد باستخدام اختبار F يستخدم الجدول التالى:

SOV	df	SS	MS	F
راجع إلى الانحدار Due to reg.	k	RSS	RSS/k	$\frac{RSS/k}{S^2}$
عن خط الانحدار From reg.	$n-k-1$	ESS	$S^2 = ESS/(n-k-1)$	
Total الكلى	$n-1$	$\sum y^2$		

وقيمة F المحسوبة تقارن بقيمة F الجدولية بدرجات حرية (k)، $(n-k-1)$ ومستوى معنوية α، وفى هذه الحالة يعتبر اختبار مشترك join test يختبر معنوية معاملات الانحدار معاً وبعد ذلك يمكن تحديد أى معاملات الانحدار هذه تختلف معنوياً عن الصفر.

مثال ١٤-٤

فى مثال ١٤-١ كون جدول تحليل التباين واختبر الفرض $H_\circ : \beta_1 = \beta_2 = 0$.

SOV	df	SS	MS	F
راجع إلى الانحدار	2	19.075	9.5375	4.62^{ns}
عن خط الانحدار	2	4.125	2.0625	
الكلى	**4**	**23.2**		

^{ns} (P> 0.05)

وقيمة F المحسوبة (4.62) أقل من قيمة F الجدولية بدرجات حرية 2، 2 ومستوى معنوية 5% وبالتالى لا يكون هناك مبرراً لرفض فرض العدم أى أن العلاقة بين Y وكل من X_1، X_2 معاً غير معنوية.

ولاختبار معنوية أثر إضافة متغير ثالث عند دراسة العلاقة بين متغيرين حتى يمكن إهماله أو إضافته إلى النموذج فإنه يمكن إجراء الاختبار التالى:

حيث أن النموذج فى حالة العلاقة بين متغيرين اثنين فقط:

$$\hat{Y} = a' + b_1'X_1$$

والنموذج فى حالة العلاقة بين ٣ متغيرات:

$$\hat{Y} = a + b_1X_1 + b_2X_2$$

ويحسب مجموع المربعات الراجع للانحدار RSS فى كل من النموذجين والفرق بينهما يكون راجعاً إلى إدخال المتغير الثالث إلى النموذج وله درجة حرية واحدة. وعلى ذلك فمتوسط مربعاته هو نفسه مجموع مربعاته والذى يقسم على متوسط مربعات الخطأ وذلك لاختبار معنوية معامل الانحدار الجزئى Y على X_2 ويمكن توضيح ذلك بالمثال التالى.

مثال ١٤-٥

إذا كان مجموع المربعات الراجعة للخطأ = 3، $n = 50$، $\sum x_1 y = 2$، $\sum x_1^2 = 16$ ، متوسط المربعات الراجعة للانحدار = 0.5. كون جدول تحليل التباين واختبر فرض العدم $H_o: \beta_1 = \beta_2 = 0$، ثم اختبر إذا كانت β_2 معنوية أم لا.

SOV	df	SS	MS	F
راجع إلى الانحدار	2	1	0.5	8.33^{**}
عن خط الانحدار	47	3	0.06	

وحيث إن قيمة F المحسوبة معنوية بدرجة ثقة 99% فهذا يعنى أن العلاقة بين المتغيرات، العلاقة بين Y من ناحية و X_1، X_2 معاً من ناحية أخرى، معنوية إحصائياً وليست راجعة إلى الصدفة أو الأخطاء العشوائية.

ولاختبار فرض العدم $H_o: \beta_2 = 0$ يحسب مجموع المربعات الراجعة للانحدار Y على X_1، X_2 معاً ثم يحسب مجموع المربعات الراجعة للانحدار Y على X_1 وبالطرح يمكن الحصول على مجموع المربعات الراجعة نتيجة إضافة المتغير X_2 حيث $\sum(x_1y)^2 / \sum x_1^2 = (2)^2/16 = 0.25$، وبالتالى فإن مجموع المربعات نتيجة إضافة X_2 إلى النموذج $1 - 0.25 = 0.75$، ويكون جدول تحليل التباين كالتالى:

SOV	df	SS	MS	F
الراجع للانحدار على X_1، X_2	2	1		
الراجع للانحدار على X_1 بغض النظر عن X_2	1	0.25		
الراجع للانحدار نتيجة لإضافة X_2	1	0.75	0.75	$\frac{0.75}{0.06} = 12.25$
عن خط الانحدار	47	3	0.06	

وعلى ذلك يرفض فرض العدم الخاص بعدم وجود علاقة بين المتغيرين Y، X_2 فى وجود X_1 ، أى أنه توجد علاقة معنوية بينهما بدرجة ثقة 99% وبالتالى فإنه لا يمكن إغفال إضافة المتغير X_2 إلى النموذج الإحصائى.

مثال ١٤-٦

حل مثال١٤-١ باستخدام برنامج SAS.

```
DATA MULTREG;
INPUT X1 X2 Y @@;
CARDS;
0 2 2 2 1 6 3 0 8
4 3 7 5 4 4
PROC REG;
MODEL Y = X1 X2;
RUN;
```

نتائج التحليل:

Model: MODEL1

Dependent Variable: Y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	19.07500	9.53750	4.624	0.1778
Error	2	4.12500	2.06250		
C Total	4	23.20000			

Root MSE	1.43614	R-square	0.8222
Dep Mean	5.40000	Adj R-sq	0.6444
C.V.	26.59520		

Parameter Estimates

Parameter Variable	DF	Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	5.000000	1.30725995	3.825	0.0621
X1	1	1.125000	0.42912910	2.622	0.1199
X2	1	-1.375000	0.52205808	-2.634	0.1190

وتوجد عدة طرق تمكن من تقرير إبقاء أو استبعاد متغير مستقل معين من معادلة الانحدار المتعدد. ومن هذه الطرق طريقة الاستبعاد الخلفى backward elimination، طريقة الاختيار الأمامى forward selection وطريقة الخطوة خطوة stepwise procedure. وفى جميع هذه الطرق يلزم استخدام أى من برامج التحليل الإحصائى مثل SAS أو غيره حيث إنه من الصعب تنفيذ هذه الطرق باستخدام الآلات الحاسبة العادية.

١٤-٦ اختيار أفضل معادلة انحدار Selecting the best regression equation

لجعل معادلة الانحدار المتعدد أكثر فائدة فى أغراض التنبؤ بسلوك المتغير التابع عند تغير أى من المتغيرات المستقلة فإنه لابد من أن تتضمن معادلة الانحدار كل المتغيرات المستقلة الممكنة والتى يمكن أن تؤثر على المتغير التابع. فلو فرض وجود متغير تابع وعدد 10 متغيرات مستقلة معنى ذلك أنه يمكن تكوين عدد $[{}_{10}C_1+{}_{10}C_2+{}_{10}C_3+\cdots+{}_{10}C_{10}]$ أى 1023 معادلة انحدار مختلفة، وهذا يجعل من الصعب مقارنة هذه المعادلات ببعضها البعض، هذا بالإضافة إلى تكلفة الحصول على المعلومات الخاصة بكل المتغيرات المستقلة. وبالتالى فإن المتغيرات ذات التأثير المعنوى فقط هى التى يمكن أن تتضمنها معادلة الانحدار وبالتالى اختيار أفضل معادلة انحدار.

١٤-٦-١ طريقة الاستبعاد الخلفى Backward elimination method

يتم فى هذه الطريقة وضع جميع المتغيرات المستقلة فى معادلة الانحدار ثم محاولة اختيار أفضل معادلة انحدار تحتوى على أقل عدد ممكن من المتغيرات المستقلة. وتجرى هذه الطريقة كالتالى:

١- تقدير معادلة انحدار تتضمن كل المتغيرات المستقلة.

٢- حساب قيمة F الجزئية (partial F) لكل متغير اشتملت عليه معادلة الانحدار والتى تعتمد فى حسابها على متوسط مجموع المربعات الجزئى partial mean square.

٣- تقارن أقل قيمة من قيم F الجزئية (بفرض أنها F_L) بقيمة F الجدولية (بفرض أنها F_0) بمستوى معنوية سبق اختياره. وبالتالى:

أ - إذا كانت قيمة F_L أصغر من قيمة F_0 يتم إلغاء المتغير المستقل قرينها. ثم يعاد حساب معادلة الانحدار مرة أخرى بدون هذا المتغير وتكرر نفس الخطوات مرة أخرى.

ب- إذا كانت قيمة F_L أكبر من أو تساوى قيمة F_0 فإن معادلة الانحدار فى هذه الحالة تمثل المعادلة النهائية.

١٤-٦-٢ طريقة الاختيار الأمامي forward selection method

فى هذه الطريقة يتم إضافة المتغيرات المستقلة الواحد تلو الآخر كالتالى:

١- تحسب قيم معاملات الارتباط البسيط بين المتغير التابع وكل من المتغيرات المستقلة. يتم اختيار أول متغير مستقل وهو المتغير الذى له أعلى معامل ارتباط بسيط مع المتغير التابع (افترض أنه X_H) بمستوى معنوية p_H، يقارن مستوى المعنوية هذا مع مستوى معنوية سبق تحديده وليكن SLE (significant level of entry) وبالتالى:

أ – إذا كانت قيمة p_H أكبر من قيمة SLE فإن المتغير X_H قرين هذه القيمة لا يمكن وضعة فى معادلة الانحدار وبالتالى لا يمكن حساب معادلة الانحدار فى هذه الحالة.

ب– إذا كانت قيمة p_H أقل من أو تساوى قيمة SLE فإن المتغير X_H يمكن أن تشتمل علية معادلة الانحدار.

٢ – يتم حساب معاملات الانحدار الجزئية بين المتغير التابع وكل من المتغيرات المستقلة المتبقية والمصححة للمتغير(أو المتغيرات) التى تضمنتها معادلة الانحدار مع حساب احتمالات معنوياتها. يتم اختيار المتغير المستقل الذى له أعلى معامل ارتباط جزئى مع المتغير المعتمد (وليكن X_H) بمستوى معنوية p_H ، وبالتالى:

أ – إذا كانت قيمة p_H أكبر من قيمة SLE فإن المتغير X_H قرين هذه القيمة لا يمكن وضعة في معادلة الانحدار وبالتالي لا يوضع هذا المتغير في معادلة الانحدار وتكون معادلة الانحدار متضمنة المتغير السابق فقط.

ب– إذا كانت قيمة p_H أقل من أو تساوى قيمة SLE فإن المتغير X_H يمكن أن تشتمل علية معادلة الانحدار. ويتم تقييم المتغير التالى ... وهكذا.

١٤-٦-٣ طريقة الخطوة خطوة stepwise method

هذه الطريقة هى نفس طريقة الاختيار الأمامى مع الاختلاف فإن المتغير المستقل لذى يتم اختياره ليس بالضرورة أن يظل باقيا فى معادلة الانحدار حيث أنه يتم إعادة قييم المتغيرات التى سبق أن تضمنتها معادلة الانحدار عند كل مرة يضاف إليها متغير جديد أى باستخدام طريقة الاستبعاد الخلفى.

مثال ١٤-٧

أورد Draper and Smith (1981) بيانات عن استخدام أربعة مواد كيماوية مختلفة (X_1، X_2، X_3، X_4) معبرا عنها كنسبة من الخامات التى تستخدم فى صناعة الأسمنت وإثرها على كمية الطاقة الناتجة (Y) بالكالورى لكل جرام أسمنت

مصنع. والمطلوب تقدير أفضل معادلة انحدار متعدد باستخدام طريقة الاستبعاد الخلفى backward elimination وطريقة الاختيار الأمامى forward sclection وطريقة الخطوة خطوة stepwise procedure وكانت البيانات كالتالى:

X_1	X_2	X_3	X_4	Y
7	26	6	60	78.5
1	29	15	52	74.3
11	56	8	20	104.3
11	31	8	47	87.6
7	52	6	33	95.9
11	55	9	22	109.2
3	71	17	6	102.7
1	31	22	44	72.5
2	54	18	22	93.1
21	47	4	26	115.9
1	40	23	34	83.8
11	66	9	12	113.3
10	68	8	12	109.4

الحل باستخدام البرنامج الإحصائى SAS

```
DATA CEMENT;
INPUT X1-X4 Y @@;
CARDS;
7  26 6 60 78.5 1 29 15 52 74.3 11 56 8 20 104.3
11 31 8 47 87.6 7 52 6  33 95.9 11 55 9 22 109.2
3 71 17 6 102.7 1 31 22 44 72.5 2 54 18 22 93.1
21 47  4 26 115.9 1  40 23 34 83.8 11 66 9 12 113.3
10 68 8 12 109.4
PROC CORR NOPROB NOSIMPLE; VAR Y; WITH X1 X2 X3 X4;
PROC STEPWISE;
MODEL Y = X1 X2 X3 X4/F B STEPWISE;
RUN;
```

لاحظ:

١- استخدام أمر prco corr لحساب معامل الارتباط البسيط بين المتغير المعتمد وكل من المتغيرات المستقلة رغم عدم الحاجة لوضع أمر لحساب ذلك ولكن وضع الأمر بغرض إظهار قيم معاملات الارتباط. لاحظ استخدام noprop عند عدم الرغبة فى مستويات المعنوية المصاحبة لكل معامل ارتباط. استخدام nosimple لعدم طبع معلومات توصيف البيانات مثل المتوسط، المجموع ... الخ لعدم الحاجة لهذه المعلومات.

٢ – استخدام proc stepwise عند الرغبة في اختيار أفضل معادلة انحدار.

٣ – يكتب أمر model مع وضع جميع المتغيرات المستقلة فى النموذج وإضافة الاختيارات المختلفة للثلاث طرق السابق شرحها حيث b تمثل backward و f تمثل forward مع العلم أنه يمكن كتابة الكلمة كاملة بدلا من الاختصار.

٤ – يمكن إضافة اختيارات مختلفة لدرجة المعنوية إلى أمر model وذلك عند الرغبة فى تغيير مستوى المعنوية فى الحالات المختلفة.

نتائج التحليل:

```
                    Correlation Analysis

        4 'WITH' Variables:  X1       X2       X3       X4
        1 'VAR'  Variables:  Y
            Pearson Correlation Coefficients  / N = 13

                                    Y
                     X1          0.73072

                     X2          0.81625

                     X3         -0.53467

                     X4         -0.82131
```

Forward Selection Procedure for Dependent Variable Y

```
Step 1 Variable X4 Entered   R-square = 0.67454196    C(p)= 138.73083349
              DF     Sum of Squares    Mean Square       F       Prob>F
Regression     1     1831.89616002   1831.89616002    22.80     0.0006
Error         11      883.86691690     80.35153790
Total         12     2715.76307692
```


مع تحيات د. سلام حسين الهلالي salamalhelali@yahoo.com

```
          Parameter    Standard        Type II
Variable   Estimate    Error         Sum of Squares     F      Prob>F
INTERCEP  117.56793118  5.26220651   40108.47690796   499.16   0.0001
X4         -0.73816181  0.15459600    1831.89616002    22.80   0.0006

Bounds on condition number:          1,            1
----------------------------------------------------------------------
Step 2   Variable X1 Entered  R-square = 0.97247105 C(p) = 5.49585082

            DF  Sum of Squares   Mean Square     F     Prob>F
Regression   2    2641.00096477  1320.50048238   176.63  0.0001
Error       10      74.76211216     7.47621122
Total       12    2715.76307692

          Parameter      Standard       Type II
Variable   Estimate      Error        Sum of Squares    F       Prob>F
INTERCEP  103.09738164   2.12398361 17614.67006622  2356.10   0.0001
X1          1.43995828   0.13841664   809.10480474   108.22   0.0001
X4         -0.61395363   0.04864455    1190.92463664 159.30   0.0001

Bounds on condition number:     1.064105,     4.256421
----------------------------------------------------------------------
Step 3  Variable X2 Entered R-square = 0.98233545    C(p) = 3.01823347

            DF  Sum of Squares   Mean Square     F    Prob>F
Regression   3    2667.79034752   889.26344917   166.83 0.0001
Error        9      47.97272940     5.33030327
Total       12    2715.76307692

          Parameter    Standard       Type II
Variable  Estimate     Error        Sum of Squares    F       Prob>F
INTERCEP  71.64830697  14.14239348  136.81003409    25.67     0.0007
X1         1.45193796   0.11699759  820.90740153   154.01     0.0001
X2         0.41610976   0.18561049   26.78938276     5.03     0.0517
X4        -0.23654022   0.17328779    9.93175378     1.86     0.2054

Bounds on condition number:     18.94008,     116.3601
----------------------------------------------------------------------
No other variable met the 0.5000 significance level for entry into the
model.
----------------------------------------------------------------------
     Summary of Forward Selection Procedure for Dependent Variable Y
       Variable    Number  Partial    Model
Step   Entered       In     R**2      R**2       C(p)      F        Prob>F
 1       X4          1     0.6745    0.6745   138.7308  22.7985    0.0006
 2       X1          2     0.2979    0.9725     5.4959 108.2239    0.0001
 3       X2          3     0.0099    0.9823   3.0182     5.0259    0.0517
----------------------------------------------------------------------
```

Backward Elimination Procedure for Dependent Variable Y

Step 0 All Variables Entered R-square = 0.98237562 C(p) = 5.00000000

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	4	2667.89943757	666.97485939	111.48	0.0001
Error	8	47.86363935	5.98295492		
Total	12	2715.76307692			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	62.40536930	70.07095921	4.74551686	0.79	0.3991
X1	1.55110265	0.74476987	25.95091138	4.34	0.0708
X2	0.51016758	0.72378800	2.97247824	0.50	0.5009
X3	0.10190940	0.75470905	0.10909005	0.02	0.8959
X4	-0.14406103	0.70905206	0.24697472	0.04	0.8441

Bounds on condition number: 282.5129, 2489.203

Step 1 Variable X3 Removed R-square = 0.98233545 C(p) = 3.01823347

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	3	2667.79034752	889.26344917	166.83	0.0001
Error	9	47.97272940	5.33030327		
Total	12	2715.76307692			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	71.64830697	14.14239348	136.81003409	25.67	0.0007
X1	1.45193796	0.11699759	820.90740153	154.01	0.0001
X2	0.41610976	0.18561049	26.78938276	5.03	0.0517
X4	-0.23654022	0.17328779	9.93175378	1.86	0.2054

Bounds on condition number: 18.94008, 116.3601

Step 2 Variable X4 Removed R-square = 0.97867837 C(p) = 2.67824160

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	2	2657.85859375	1328.92929687	229.50	0.0001
Error	10	57.90448318	5.79044832		
Total	12	2715.76307692			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	52.57734888	2.28617433	3062.60415610	528.91	0.0001
X1	1.46830574	0.12130092	848.43186034	146.52	0.0001
X2	0.66225049	0.04585472	1207.78226562	208.58	0.0001

```
Bounds on condition number:      1.055129,      4.220516
------------------------------------------------------------------------
All variables left in the model are significant at the 0.1000 level.
------------------------------------------------------------------------
  Summary of Backward Elimination Procedure for Dependent Variable Y

       Variable    Number  Partial    Model
Step   Removed       In      R**2      R**2       C(p)       F        Prob>F
  1     X3           3     0.0000    0.9823    3.0182    0.0182    0.8959
  2     X4           2     0.0037    0.9787    2.6782    1.8633    0.2054
------------------------------------------------------------------------

              Stepwise Procedure for Dependent Variable Y

Step 1  Variable X4 Entered  R-square = 0.67454196 C(p) =138.73083349

           DF  Sum of Squares    Mean Square      F       Prob>F
Regression  1   1831.89616002  1831.89616002   22.80    0.0006
Error      11    883.86691690    80.35153790
Total      12   2715.76307692

          Parameter        Standard      Type II
Variable  Estimate         Error       Sum of Squares     F        Prob>F

INTERCEP  117.56793118    5.26220651  40108.47690796  499.16    0.0001
X4         -0.73816181    0.15459600   1831.89616002   22.80    0.0006

Bounds on condition number:            1,             1
------------------------------------------------------------------------

Step 2  Variable X1 Entered  R-square = 0.97247105  C(p) = 5.49585082

            DF  Sum of Squares   Mean Square      F      Prob>F
Regression   2   2641.00096477  1320.50048238   176.63   0.0001
Error       10     74.76211216     7.47621122
Total       12   2715.76307692

          Parameter    Standard       Type II
Variable  Estimate     Error        Sum of Squares     F        Prob>F
INTERCEP  103.09738164  2.12398361  17614.67006622  2356.10    0.0001
X1          1.43995828  0.13841664    809.10480474   108.22    0.0001
X4         -0.61395363  0.04864455   1190.92463664   159.30    0.0001

Bounds on condition number:     1.064105,      4.256421
------------------------------------------------------------------------

Step 3  Variable X2 Entered  R-square = 0.98233545 C(p) = 3.01823347
```

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	3	2667.79034752	889.26344917	166.83	0.0001
Error	9	47.97272940	5.33030327		
Total	12	2715.76307692			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	71.64830697	14.14239348	136.81003409	25.67	0.0007
X1	1.45193796	0.11699759	820.90740153	154.01	0.0001
X2	0.41610976	0.18561049	26.78938276	5.03	0.0517
X4	-0.23654022	0.17328779	9.93175378	1.86	0.2054

Bounds on condition number: 18.94008, 116.3601

Step 4 Variable X4 Removed R-square = 0.97867837 C(p) = 2.67824160

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	2	2657.85859375	1328.92929687	229.50	0.0001
Error	10	57.90448318	5.79044832		
Total	12	2715.76307692			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	52.57734888	2.28617433	3062.60415609	528.91	0.0001
X1	1.46830574	0.12130092	848.43186034	146.52	0.0001
X2	0.66225049	0.04585472	1207.78226562	208.58	0.0001

Bounds on condition number: 1.055129, 4.220516

All variables left in the model are significant at the 0.1500 level.
No other variable met the 0.1500 significance level for entry into the model.

Summary of Stepwise Procedure for Dependent Variable Y

Step	Variable Entered	Variable Removed	Number In	Partial R**2	Model R**2	C(p)	F	Prob>F
1	X4		1	0.6745	0.6745	138.7308	22.7985	0.0006
2	X1		2	0.2979	0.9725	5.4959	108.2239	0.0001
3	X2		3	0.0099	0.9823	3.0182	5.0259	0.0517
4		X4	2	0.0037	0.9787	2.6782	1.8633	0.2054

١٤-٧ الارتباط المتعدد والجزئي Multiple and partial correlation

يفترض، كما هو الحال في حالة الارتباط البسيط، أن المتغيرات عشوائية أي في عينة عشوائية مسحوبة من عشيرة متعددة المتغيرات تتوزع طبيعيا multivariate normal distribution، وعلى ذلك فلن يكون الاهتمام بأي من المتغيرات مستقل وأي منها تابع.

ويعرف معامل الارتباط الجزئي بين متغيرين بأنه هو الارتباط بين هذين المتغيرين في مجموعة من الأفراد أو المشاهدات كل منها له نفس المتغير الثالث (أو المتغيرات الأخرى في حالة أكثر من ثلاث متغيرات)، أي معامل الارتباط بين الأول والثاني على اعتبار أن المتغير الثالث ثابت أو بعد التصحيح لتأثيره. وقيمة هذا المعامل تنحصر أيضاً بين 1- ، 1+.

ومعاملات الارتباط الجزئية في حالة ثلاث متغيرات هي:

$$r_{1,2.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)}\sqrt{(1-r_{23}^2)}}$$

$$r_{1,3.2} = \frac{r_{13} - r_{12}r_{32}}{\sqrt{(1-r_{12}^2)}\sqrt{(1-r_{32}^2}}$$

$$r_{2,3.1} = \frac{r_{23} - r_{21}r_{31}}{\sqrt{(1-r_{12}^2)}\sqrt{(1-r_{13}^2)}} \qquad (١٣-١٤)$$

حيث $r_{12} = r_{21}$ ... وهكذا. ومعنى $r_{12.3}$ أنه معامل الارتباط الجزئي بين المتغير الأول والمتغير الثاني مع بقاء المتغير الثالث ثابتاً.

ومن ذلك فإنه لحساب معاملات الارتباط الجزئية بين ثلاث متغيرات عشوائية يلزم حساب معاملات الارتباط البسيطة بين كل زوج منهم أي r_{12}، r_{13}، r_{23}. ويستخدم جدول ٧ ملحق أ لاختبار معنوية معاملات الارتباط البسيط. أما معامل الارتباط المتعدد R فيقيس شدة العلاقة بين أحد المتغيرات وجميع المتغيرات الأخرى معاً وقيمته تنحصر بين الصفر والواحد الصحيح. ويمكن حساب معامل الارتباط المتعدد بين المتغير 1 من جهة والمتغيرين 2، 3 من جهة أخرى كما يلي:

$$R_{1(2,3)} = \sqrt{1-(1-r_{12}^2)(1-r_{13.2}^2)} \qquad (١٤-١٤)$$

كما يحسب من مجموع المربعات فى تحليل الانحدار حيث:

$$R = \sqrt{\frac{RSS}{TSS}} \qquad (١٥-١٤)$$

وحيث إن:

$$R^2 = RSS/TSS$$
$$= \sum \hat{y}^2 / \sum y^2$$

فإن:

$$\sum \hat{y}^2 = R^2 \sum y^2 = RSS$$

وعليه فإن مجموع المربعات عن خط الانحدار هو:

$$ESS = TSS - RSS = \sum y^2 - R^2 \sum y^2 = (1-R)^2 \sum y^2$$

وتكون قيمة F لاختبار فرض العدم أن المتغيرين 2، 3 معاً لا يؤثران على المتغير 1 هى:

$$F = \frac{R^2 \sum y^2 / k}{(1-R^2)\sum y^2 / (n-k-1)} = \left(\frac{n-k-1}{k}\right)\left(\frac{R^2}{1-R^2}\right) \qquad (١٦-١٤)$$

وبالتالى فإنه يستخدم أيضاً اختبار F لاختبار معنوية معامل الارتباط المتعدد.

مثال ١٤-٨

استخدم بيانات المثال ١٤-١ لحساب معاملات الارتباط الجزئية بين المتغيرين الأول والثالث ومعامل الارتباط المتعدد بين المتغير 1 من جهة مع 2، 3 من جهة أخرى واختبر معنوية كل منهم ولسهولة التعبير أعد تسمية y بـــ X_3.

حساب كل من معاملات الارتباط البسيط كما يلى:

$$r_{13} = \frac{\sum X_1 X_3 - \sum X_1 \sum X_3 / n}{\sqrt{\sum X_1^2 - \frac{(\sum X_1)^2}{n}} \sqrt{\sum X_3^2 - \frac{(\sum X_3)^2}{n}}}$$

$$= \frac{84 - \frac{(27)(14)}{5}}{\sqrt{54 - \frac{(14)^2}{5}}\sqrt{169 - \frac{(27)^2}{5}}} = 0.45$$

وبالمثل

$$r_{23} = -0.46 \text{ ، } r_{12} = 0.49$$

معامل الارتباط الجزئى:

$$r_{13.2} = \frac{0.45 - (0.49)(-0.46)}{\sqrt{1 - (0.49)^2}\sqrt{1 - (-0.46)^2}} = 0.87$$

ودرجات الحرية 2

أما معامل الارتباط المتعدد (معادلة ١٤-١٤):

$$R_{1(23)} = \sqrt{1 - [1 - (0.49)^2][(1 - (0.87)^2]} = 0.91$$

ومنها $R^2 = 0.82$ بمعنى أن حوالى 82% من التباين فى المتغير Y يمكن تفسيره بالنموذج الخطى فى المتغيرين المستقلين X_1 ، X_2.

ويمكن الحصول على معامل الارتباط المتعدد أيضا باستخدام الجذر التربيعى لناتج قسمة مجموع المربعات الراجعة للانحدار على مجموع المربعات الكلى كالتالى:

$$R_{1(23)} = \sqrt{\frac{19.075}{23.2}} = 0.91$$

ولاختبار معنوية معامل الارتباط المتعدد تحسب قيمة F حيث:

$$F = \frac{5 - 2 - 1}{2}\left[\frac{(0.91)^2}{1 - (0.91)^2}\right] = 4.8$$

والفرق بين قيمة F والتى تساوى 4.8 فى حالة الارتباط المتعدد، 4.6 فى حالة الانحدار المتعدد إنما يرجع إلى خطأ التقريب.

وقيمة F المحسوبة أقل من قيمة F الجدولية وعلى ذلك فالعلاقة بين المتغيرات الثلاثة غير معنوية.

ويمكن حساب R^2_{adj}، أى R^2 المعدلة لعدد المتغيرات التفسيرية (المستقلة)، من المعادلة التالية:

$$R^2_{adj} = 1 - \left(\frac{n-1}{n-k-1}\right)\left(1 - R^2\right) = 1 - \left(\frac{5-1}{5-2-1}\right)(1 - 0.82) = 0.64$$

صندوق ١٤-١

- الانحدار المتعدد هو امتداد طبيعى للانحدار البسيط حيث يكون هناك متغير تابع واحد (Y) وعدة متغيرات مستقلة $(X_1, X_2, \cdots, X_p)$ بدلا من واحد فقط. وتصبح معادلة التنبؤ $\hat{Y} = a + b_1X_1 + b_2X_2 + \cdots + b_pX_p$ حيث b_1 هى معامل الانحدار الجزئى للمتغير Y على المتغير المستقل X_1 ... وهكذا لبقية الـــ b's.
- وبالتوازى مع الانحدار البسيط فإن التباين الكلى فى Y يقسم إلى أجزاء كل منها يرجع إلى متغير مستقل بعينه. ويحسب مجموع المربعات الراجع لمتغير مستقل معين بحساب المتبقى فى النموذج بدون هذا التغير منقوصا منه المتبقى فى النموذج بعد إضافة هذا المتغير.
- يلاحظ أن حاصل جمع مجموع انحرافات هذه الأجزاء الراجعة للانحدار لا تساوى مجموع المربعات الراجع إلى الاعتماد على كل المتغيرات المستقلة مجتمعة إلا إذا كانت المتغيرات المستقلة مستقلة عن بعضها تماما (أى معامل الارتباط بين كل اثنين منها = صفر).
- وحيث أنه فى الحالة الواقعية يمكن أن يزداد عدد المتغيرات المستقلة كثير، فهناك طرق للمساعدة فى تقرير أى من المتغيرات المستقلة يمكن حذفه من النموذج الإحصائى دون فقد معلومات فقدا معنويا. من هذه الطرق: الاستبعاد الخلفى، الاختيار الأمامى، الخطوة خطوة.

تمارين الباب الرابع عشر

١٤-١ أورد Hill (1982) التمرين التالى حيث أخذت عينة من 24 دجاجة وكانت X_1 تمثل الوزن الحى بالمائة جرام، X_2 تمثل عمق الصدر بالسنتيمتر وتمثل Y الوزن بعد الطبخ بالمائة جرام وكانت البيانات كالتالى:

رقم الدجاجة	Y	X_1	X_2	رقم الدجاجة	Y	X_1	X_2
١	6.03	16.44	8.3	١٣	7.55	18.40	9.0
٢	6.09	16.76	8.5	١٤	5.83	15.46	7.5
٣	6.67	17.22	8.8	١٥	5.93	15.20	8.2
٤	6.25	16.04	8.5	١٦	6.65	17.15	8.7
٥	5.65	15.74	8.5	١٧	6.24	15.60	8.1
٦	6.91	18.06	9.0	١٨	5.50	16.95	8.0
٧	7.81	20.60	8.7	١٩	7.22	17.22	8.1
٨	5.68	16.20	9.0	٢٠	7.17	18.36	8.7
٩	7.14	17.30	8.6	٢١	5.81	17.50	8.0
١٠	7.10	18.85	8.1	٢٢	7.07	16.66	8.0
١١	6.29	17.63	8..0	٢٣	5.51	14.88	8.6
١٢	6.70	17.78	8.6	٢٤	5.77	16.25	8.2

المطلوب:

١- تقدير معادلة الانحدار واختبار معنوية معاملات الانحدار المختلفة.

٢- تقدير معاملات الارتباط الجزئية والارتباط المتعدد واختبار معنوية كل منها.

١٤-٢ أجرى Ostle, 1963 تجربة على الأرانب لدراسة المتغيرات التى قد يكون لها علاقة بدرجة تصلب الشرايين degree of atherosclerosis (Y) والتى قدرت بقيم صحيحة تتراوح بين 0 إلى 4 وأخذت قياسات للمتغيرات التالية:

X_5	X_4	X_3	X_2	X_1	Y	رقم الأرنب
18	0.90	2.46	424	30	2	١
10	0.91	2.39	313	30	0	٢
30	0.95	2.75	243	35	2	٣
21	0.95	2.19	365	35	2	٤
39	1.00	2.67	396	43	3	٥
19	0.79	2.74	356	43	2	٦
56	1.26	2.55	346	44	3	٧
28	0.95	2.58	156	44	0	٨
42	1.10	2.49	278	44	4	٩
21	0.88	2.52	349	44	1	١٠
56	1.29	2.36	141	44	1	١١
24	0.97	2.36	245	44	1	١٢
27	1.01	2.15	395	44	1	١٣
45	1.11	2.56	297	45	3	١٤
20	0.94	2.62	310	45	2	١٥
35	0.96	3.39	151	45	3	١٦
15	0.88	3.57	370	45	4	١٧
64	1.47	1.98	379	45	4	١٨
31	1.05	2.06	463	45	3	١٩
60	1.32	2.45	316	45	4	٢٠
36	1.08	2.25	280	45	4	٢١
59	1.36	2.20	139	49	0	٢٢
37	1.13	2.05	245	49	4	٢٣
25	0.88	2.15	373	49	1	٢٤
54	1.18	2.15	224	51	3	٢٥
33	1.16	2.10	677	51	4	٢٦
59	1.40	2.10	424	51	4	٢٧
30	1.05	2.10	150	51	0	٢٨

حيث

X_1 : متوسط الجرعة اليومية للكولسترول بالجرام

X_2 : متوسط كولسترول سيرم الدم بالمليجرام لكل 100 مل.

X_3 : وزن الجسم الابتدائى initial body weight بالكيلوجرام.

X_4 : النسبة بين الوزن النهائى والوزن الابتدائى للجسم.

X_5 : متوسط كمية الغذاء المأكول بالجرام لكل كيلوجرام وزن ابتدائى.

المطلوب:

١ – إجراء تحليل الانحدار المناسب.

٢ – ما هى أفضل معادلة انحدار عند مستويات معنوية مختلفة ولتكن 5%، 15% .

١٥ تصميم التجارب

Design of Experiments

مع تحيات د. سلام حسين الهلالي salamalhelali@yahoo.com

١٥-١ مقدمة

يعتبر تصميم التجارب أحد الخطوات الرئيسية التى تتبع عند إجراء أى بحث علمى حيث إنه يُمكن الباحث من الحصول على المعلومات اللازمة للإجابة على السؤال أو الأسئلة محل الدراسة بطريقة منطقية. و يتناول التصميم التجريبى experimental design هذا الأمر بدءاً بوضع المعلومات المرغوب الحصول عليها فى صورة أسئلة حتى يمكن الحصول فعليا على هذه المعلومات واتخاذ قرارات بشأنها بعبارات احتمالية. وأسلوب التصميم التجريبى أساساً هو مقارنة واختبار الاختلافات (التباين) بين الوحدات التجريبية. وعندما تكون هذه الاختلافات صغيرة بالنسبة لمجموعة معينة فإنه نادراً ما يكون هناك حاجة إلى تصميم تجريبى معقد والعكس صحيح.

ويمكن القول إن الاصطلاح "تصميم التجارب" يقصد به أنه الخطة المعينة والمحددة لتوزيع ووضع الوحدات التجريبية experimental units بالنسبة للمعاملات والمؤثرات المختلفة.

ويتناول تصميم التجارب أنشطة رئيسية متطلبة فى البحث العلمى وهى بالترتيب حسب أسبقية استخدامها:

١- تحديد أهداف التجربة وهذه الخطوة تؤدى إلى صياغة أسئلة محددة يلزم الإجابة عليها وهذه الأسئلة سوف تحدد اتخاذ القرارات فيما يتعلق بالخطوات التالية.

٢- تكوين الفروض (النظريات) الإحصائية statistical hypotheses (فرض العدم null والفرض البديل alternative). فرض العدم null hypothesis هو مقولة عن واحد أو أكثر من معالم العشيرة population parameters، ونادراً ما تكون الفروض الإحصائية هى نفسها الفروض البحثية ولكن يمكن وضع الثانية فى صورة الأولى حتى يمكن اختبارها. فقد يكون السؤال البحثى مثلا هو: هل يؤثر مستوى الطاقة فى عليقة الحملان على معدل نموها؟. وعند الرغبة فى اختبار ذلك إحصائياً فإنه يتم وضع السؤال فى صورة نظرية إحصائية فرضية يطلق عليها فرض العدم H_o وهذه قد تكون متوسط المستوى الأول = متوسط المستوى الثانى = ... = متوسط المستوى الأخير. ثم يتم اختبار هذا الفرض، والنتيجة ستكون إما رفض هذا الفرض (أى أنه يوجد فرق بين متوسطات هذه المستويات سواء جميعها أو بعضها) أو عدم القدرة على رفضه (أى لا يوجد فروق بين المتوسطات) وذلك باحتمالات معينة فى وجود فروض بديلة معينة.

٣- تحديد قواعد اتخاذ القرارات مثل احتمال خطأ النوع الأول واحتمال خطأ النوع الثانى والفروق المرغوب إعلانها معنوية.

٤- تحديد جميع مصادر الاختلاف sources of variation وهذه تتضمن:

أ- تحديد المعاملات ومستويات كل معاملة (وقد سبق الحديث فى الباب التاسع عن المتغير والعامل والمعاملة والمستوى).

ب- تحديد الوحدات التجريبية experimental units، والوحدة التجريبية يقصد بها أصغر وحدة يتم وضعها تحت نفس مستوى المعاملة محل الدراسة. وبالتالى قد تكون الوحدة التجريبية عبارة عن حيوان واحد أو مجموعة حيوانات فى حظيرة واحدة أو دجاجة واحدة أو مجموعة دجاج داخل قفص واحد أو بطارية واحدة أو مجموعة أسماك فى حوض واحد أو حوض plot واحد فى أرض زراعية. وتحديد نوعية الوحدة التجريبية يكون مهم جدا لتحديد نوعية المكررات replicates. والمكررات هى عبارة عن مجموعة من الوحدات التجريبية تقع تحت نفس مستوى المعاملة. فإذا كانت الوحدة التجريبية هى حيوان واحد فإن المكررات عبارة عن مجموعة الحيوانات التى تخضع لنفس مستوى المعاملة. أما إذا كانت الوحدة التجريبية هى مجموعة دجاج فى قفص واحد أو مجموعة أسماك فى حوض واحد أو حوض فى أرض زراعية فإنه يلزم تكرار أقفاص الطيور أو أحواض السمك أو أحواض الزراعة التى تخضع لنفس مستوى المعاملة. ولابد هنا من التفرقة بين المكررات replicates والقياسات المتكررة repeated measurements حيث أن الأخيرة عبارة عن تكرار نفس القياس على نفس الوحدة التجريبية مثل أخذ أكثر من عينة دم على نفس الحيوان فى نفس الوقت أو فى أوقات مختلفة.

ج- تحديد عوامل القطاعات blocking factors، العوامل المزعجة noise factors، والمتغايرات covariates إن وجدت.

٥- اختيار طريقة التوزيع العشوائى للوحدات التجريبية على مستويات المعاملة أو ما يسمى بطريقة التعشية randomization.

٦- تحديد القياسات المطلوب أخذها على الوحدات التجريبية وطريقة قياسها.

٧- توصيف نموذج model التحليل.

٨- حساب عدد المشاهدات observations المطلوب أخذها.

٩- جمع البيانات طبقاً للخطة الموضوعة مسبقا.

١٠- تحليل البيانات طبقاً للخطة الموضوعة مسبقا.

١١- اتخاذ القرارات الخاصة بالفروض الإحصائية (صحتها من عدمه) طبقاً للقواعد الموضوعة مع استنباط احتمالات أن تكون هذه العبارات خاطئة.

ومما سبق يتضح وجود مبدأين أساسيين فى تصميم التجارب (Fisher, 1960) هما:

١- التعشية randomization

٢- التكرار replication

وهما الأساس الذى يختلف فيه تصميم تجريبى عن آخر، كما سيتضح فيما بعد، وهذان المبدآن لازمان أيضاً للحصول على تقدير سليم لخطأ التباين (أو الخطأ التجريبى) اللازم لاختبار الفروض واتخاذ قرارات باحتمالات خطأ محسوبة كما سبق التنويه إليه فى الباب السادس.

وعادة ما يتم إجراء التجارب لسبب أو أكثر من الأسباب التالية والتى سوف يختلف فى كل منها طريقة صياغة فرض العدم null hypothesis والفرض البديل alternative hypothesis:

١- تحديد الأسباب الرئيسية لوجود اختلافات فى متغير الاستجابة response variable الذى تم قياسه تحت ظروف تجريبية معينة.

٢- تحديد الظروف التى تؤدى إلى الحصول على أعلى maximum أو أقل minimum قيمة لمتغير الاستجابة.

٣- مقارنة ما تحقق فى متغير الاستجابة عند مستويات مختلفة للعوامل التى يتحكم فيها الباحث.

٤- الحصول على نموذج رياضى mathematical model بغرض التنبؤ prediction بسلوك متغير الاستجابة مستقبلا.

١٥-٢ المتغيرات Variables

المتغير variable عبارة عن الشىء الذى يتم قياسه أو التحكم فيه أو التعامل معه بطريقة معينة أثناء إجراء التجارب. وعادة ما يطلق عليه المتغير العشوائى random variable. ويوجد عدة أنواع من المتغيرات وهى المتغير المستقل independent variable والمتغير التابع dependent variable والمتغير المضايق (المزعج) nuisance variable.

المتغير المستقل هو المتغير الذى يقع تحت سيطرة الباحث والذى سوف يؤثر على متغير آخر. أما المتغير التابع (يسمى أحيانا متغير الاستجابة response variable) فهو المتغير الذى يتم قياسه أثناء إجراء التجربة ويتأثر بالمتغير المستقل ويعكس أثره. فمثلا عند الرغبة فى دراسة أثر العمر على الوزن، فالعمر متغير مستقل والوزن متغير تابع، أو عند دراسة أثر معاملة معينة على العمر عند النضج الجنسى فتكون لمعاملة هى المتغير المستقل والعمر هو المتغير التابع، أو دراسة الاختلاف فى درجات الإجهاد الحرارى على مستوى هرمون معين فى الدم فتكون درجات الإجهاد لحرارى هى المتغير المستقل ومستوى الهرمون فى الدم هو المتغير التابع ... وهكذا. وعلى المجرب أن يتحرى الدقة فى اختيار متغيراته. فالمتغير التابع الذى عادة ما يريد لمجرب دراسته وكيفية تأثره بالعوامل الأخرى يجب أن يختار بحيث يكون حساساً بالدرجة التى يريدها المجرب وسهل القياس بقدر الإمكان. أما اختيار مستويات المتغير لمستقل فقد تعتمد على نتائج تجارب سابقة أو اعتبارات نظرية. وأحيانا يكون من لمفيد عمل تجربة استطلاعية لتحديد مستويات العامل المستقل (بمعنى مستويات لمعاملة) وذلك قبل إجراء التجربة الفعلية.

المتغيرات المضايقة (المزعجة) nuisance variables هى متغيرات عادة تمثل مصادر للتباين غير مرغوب فيها ولا تمثل أى رغبة مباشرة لدى المجرب لقياس واختبار تأثيرها ولكنها موجودة فى التجربة ولها تأثير على المتغير المستقل. لذلك لابد من أخذها فى الاعتبار حتى يتم تجنب أثرها على احتمال طمس أو إخفاء أثر المتغير لمستقل المرغوب قياسه أصلا. فإذا أراد مجرب مثلا أن يدرس أثر مستوى التغذية على التسمين فى الحملان وأن المجرب هذا اشترى حملانه من الأسواق، هذه الحملان غالباً ما تختلف فى أوزانها عند الشراء وهذا الاختلاف فى الوزن قد يؤثر على أدائها فى التسمين لذلك لابد من أخذ هذه الاختلافات فى الوزن عند بداية التسمين حتى لا يخفى أثر الاختلاف فى مستويات التغذية. وفى تجارب الإنتاج الحيوانى كثيراً ما تكون لمتغيرات المضايقة هذه ممثلة فى جنس الحيوان، عمر الحيوان ووزنه، عمر أم لحيوان فى حالات الحيوانات صغيرة السن، مكان شراء وتنشئة هذا الحيوان ... الخ. وفى التجارب الزراعية الأخرى قد يكون المتغير المضايق عبارة عن السنة، درجة لحرارة أو الرطوبة ... الخ. ويجب التعامل مع المتغيرات المضايقة بطريقة تسمح بأن ينجلى الأثر الرئيسى للمتغير أو المتغيرات المستقلة المرغوب دراستها. ويمكن التعامل مع المتغيرات المضايقة بالطرق التالية:

١- تحكم تجريبى experimental control وذلك من خلال:

أ- تثبيت الوحدات التجريبية عضوياً physical control إن أمكن كأن تؤخذ كل الحيوانات من نفس الجنس أو نفس الوزن أو مشتراة من نفس المنطقة ... الخ.

ب- توزيع الوحدات التجريبية عشوائيا على كل مستويات المعاملات المختلفة وبالتالى يتم توزيع المصادر المعلومة وغير المعلومة أو التحيز bias بطريقة عشوائية على كل التجربة وبالتالى لا يكون التأثير موزع على مستوى واحد أو عدد محدود من مستويات المعاملة بمعنى أن لا يكون أثرها متحيزاً عند تقدير المتوسطات وتباين الخطأ.

ج- وضع مستوى يمثل المتغير المضايق ضمن مستويات المعاملة عند تصميم التجربة.

٢- تحكم إحصائى statistical control وذلك عن طريق وضع المتغير المضايق عند التحليل كمغاير covariate أو يطلق عليه متغير ملازم concomitant variable إذا كان مستمراً.

١٥-٣ كفاءة التصميم التجريبى Efficiency of experimental design

من أوائل الأسئلة التى تواجه الباحث فى مجال معين هو أى من التصميمات الإحصائية يؤدى إلى الحصول على خلاصة للبحث محل الدراسة بأكبر قدر من الكفاءة الممكنة. ويمكن تعريف الكفاءة فى هذه الحالة بعدة طرق، فقد تعرف الكفاءة بمدى الوقت اللازم لتجميع بيانات التجربة، أو تكاليف تجميع هذه البيانات، أو نسبة بين المعلومات المتحصل عليها إلى تكلفة جمع هذه البيانات ... الخ. وعادة ما يصعب الحصول على قيمة رقمية لكفاءة تصميم تجريبى معين بصورة مطلقة ولكن يمكن الحصول على الكفاءة النسبية لتصميم بالإشارة إلى تصميم أخر. وهذا هو الأكثر واقعية فكفاءة التصميم التجريبى ١ إلى التصميم التجريبى ٢ يمكن حسابها من المعادلة التالية (Kirk, 1968)

$$\text{Relative efficiency of design 1 to design 2} = \frac{\left(\dfrac{n_2 C_1}{\sigma_1^2}\right)\left(\dfrac{df_1+1}{df_1+3}\right)}{\left(\dfrac{n_1 C_2}{\sigma_2^2}\right)\left(\dfrac{df_2+1}{df_2+3}\right)}$$

حيث:

n : عدد الوحدات التجريبية فى كل من التصميمين،

σ^2 : تباين الخطأ المقدر فى كل من التصميمينن،

df : درجات حرية الخطأ فى كل من التصميمين،

C : تكلفة الوحدة التجريبية فى كل من التصميمين.

ومن هذه المعادلة يتضح بصورة أساسية أنه كلما زاد تباين الخطأ لتصميم ما كلما قت كفاءته النسبية وكذلك كلما زادت تكلفة الوحدة التجريبية فيه قلت كفاءته النسبية.

١٥-٤ تحديد حجم العينة Determination of sample size

بمجرد أن يتم توصيف المتغيرات المستقلة والمعتمدة فإنه لابد من تحديد حجم العينة. والسؤال عن حجم العينة مهم ومتكرر، وقد تم تناول هذا الموضوع بإسهاب فى البابين الخامس والسادس.

١٥-٥ التصميمات التجريبية Experimental designs

التصميمات التجريبية عديدة جداً ومتشعبة بقدر تشعب المواقف التجريبية ذاتها ولا قبل لهذا المؤلف أن يتناولها جميعاً بالتفصيل فقد أصبح لكل فرع من العلوم تصميماته الأكثر استخداماً فيه بجانب أن هذا المؤلف لم يقصد به أصلا أن يكون شاملا فى تصميم التجارب. ففى هذا الكتاب سيتم تناول بعض التصميمات الإحصائية الأكثر شيوعاً والتى تعتبر مقدمة لهؤلاء المتخصصين فى هذا الفرع ولمن أراد مزيداً من المعلومات فعليه الرجوع إلى المراجع الأكثر تخصصاً.

والقاعدة العامة التى يجب أن تتبع هى أن أفضل التصميمات التجريبية لوضع ما هو أسهلها. فهى سهلة التخطيط سهلة التنفيذ وأيضاً نتائجها أسهل فى التفسير وتعقيداتها الرياضية أقل. ويجب ألا يلجأ إلى التصميم الأكثر تعقيداً إلا لحاجة محددة وليس لتعود أو هواية. وعند إلقاء نظرة عامة على التصميمات التجريبية المتاحة للمجرب فإن أبسطها على الإطلاق هو عندما تكون المادة التجريبية متجانسة تماماً. وهنا يستخدم المجرب تصميم تام التعشية completely randomized design، وفيه تقسم الوحدات التجريبية على عدد المعاملات (أو ما يراد دراسته) وتوزع الوحدات التجريبية على هذه المعاملات عشوائياً. وعندما لا تتوافر الوحدات التجريبية المتجانسة تماماً يبدأ التعقيد تدريجياً.

وهناك عدة نقاط يجب أن تؤخذ فى الاعتبار عند المفاضلة بين أنواع التصميمات التجريبية المختلفة لاختيار أفضل تصميم منها:

١- نوعية البيانات اللازمة لاختبار فرض العدم:

أ- عدد مستويات المعاملة اللازم استخدامها

ب- كيفية اختيار مستويات المعاملة هل بناء على معلومات مسبقة apriori basis أم يتم الاختيار عشوائيا لمستويات المعاملة من ضمن عدد من المستويات يمثل عشيرة مستويات المعاملة.

ج- هل يلزم استخدام التجارب العاملية factorial experiments والتى من خلالها يمكن تقدير أثار التداخل؟

د- مدى أهمية المستويات المختلفة للمعاملات المختلفة بالنسبة للباحث، فإنه يتم فى بعض التصميمات التضحية بقوة اختبار أثر معين فى مقابل زيادة قوة اختبار أثر أخر.

٢- مدى كفاية عدد الوحدات التجريبية لاختبار فرض العدم:

أ- مدى عشوائية اختيار الوحدات التجريبية من عشيرة تمثل الوحدات التجريبية والتى يرغب الباحث فى دراستها.

ب-مدى إمكانية تقسيم الوحدات التجريبية إلى قطاعات blocks.

ج- مدى ملائمة الوحدة التجريبية للوضع تحت أكثر من مستوى للمعاملة.

د- مدى المخاطرة التى يمكن أن تتعرض لها الوحدات التجريبية من حيث الإصابة أو الموت وبالتالى محدودية استخدام نوعيات معينة من الوحدات التجريبية.

٣- مدى قوة التصميم التجريبى لاختبار فرض العدم:

أ- حجم أثر المعاملة والذى يرغب المجرب فى قياسه عمليا.

ب- النتائج التى قد تترتب على إهمال كل من خطأ النوع الأول والنوع الثانى.

٤- مدى كفاءة التصميم التجريبى فى اختبار فروض العدم:

أ- مدى التحسين فى كفاءة التصميم عند استخدام قطاعات من الوحدات التجريبية المتجانسة فى كل قطاع مقارنة بالتوزيع العشوائى لعدد أكبر من الوحدات التجريبية على المستويات المختلفة للمعاملة.

ب-مدى إمكانية زيادة كفاءة التصميم التجريبى عن طريق استخدام عدد أكبر من الوحدات التجريبية أو عن طريق التحكم التجريبى experimental control.

ج- مدى إمكانية زيادة كفاءة التصميم التجريبى عن طريق أخذ مقياس أو أكثر على علاقة بالمتغير التابع وبالتالى إمكانية استخدام طريقة تحليل الانحدار regression analysis.

د- مدى إمكانية زيادة كفاءة التصميم التجريبى عن طريق استخدام تصميم أكثر تعقيدا وما يتطلبه ذلك من زيادة فى الوقت اللازم لتصميم وتنفيذ وتحليل التجربة.

١٥-٦ التصميم تام التعشية Completely Randomized Design

هذا التصميم هو أبسط التصميمات الإحصائية على الإطلاق من حيث توزيع الوحدات التجريبية على مستويات المعاملة وكذلك من حيث تحليل نتائج التجربة. ويتطلب هذا التصميم:

١- معاملة واحدة لها مستويين أو أكثر.

٢- أن تكون الوحدات التجريبية متجانسة، كأن تكون قطعة أرض متجانسة فى خصوبتها أو أن تكون مجموعة الحيوانات التى ستجرى عليها التجربة جميعها من نفس السلالة ونفس الجنس ونفس العمر تقريباً ... وهكذا. ويتم توزيع الوحدات التجريبية عشوائيا على مستويات المعاملة بحيث أن كل وحدة تجريبية تقع تحت مستوى واحد فقط من مستويات المعاملة.

١٥-٦-١ التعشية Randomization

إذا كان هناك عدد k من مستويات المعاملة المرغوب دراستها، فتقسم الوحدات التجريبية عشوائياً إلى k من الأقسام ثم يعطى كل قسم أحد المستويات. وعادة ما تكون هذه الأقسام متساوية فى عدد الوحدات التجريبية.

فإذا فرض أن مجرب يود تجريب 4 مستويات لمعاملة معينة (أى $k = 4$) على 32 حيواناً. جميع هذه الحيوانات متجانسة فى كل ما يمكن أن يدركه المجرب، وفيما قد يؤثر على استجابة الحيوانات لمستويات المعاملة. وبالتالى يمكن استخدام التصميم تام التعشية بأن يقسم المجرب الحيوانات بطريقة عشوائية تماماً إلى أربعة أقسام كل منها به 8 حيوانات (أى $n = 8$) ثم يحدد مستوى معين من الأربع مستويات لكل قسم. ويمكن تمثيل ذلك بشكل ١٥-١.

المستويات levels			
٤	٣	٢	١
Y_{41}	Y_{31}	Y_{21}	Y_{11}
Y_{42}	Y_{32}	Y_{22}	Y_{12}
Y_{43}	Y_{33}	Y_{23}	Y_{13}
Y_{44}	Y_{34}	Y_{24}	Y_{14}
Y_{45}	Y_{35}	Y_{25}	Y_{15}
Y_{46}	Y_{36}	Y_{26}	Y_{16}
Y_{47}	Y_{37}	Y_{27}	Y_{17}
Y_{48}	Y_{38}	Y_{28}	Y_{18}

شكل ١٥-١ التصميم تام التعشية

١٥-٦-٢ النموذج الإحصائى Statistical Model

النموذج الإحصائى هو عبارة عن تعبير رياضى عن العوامل التى تؤثر فى المشاهدة Y طبقاً لافتراضات التجربة ولابد أن يعكس النموذج العلاقة بين متغير الاستجابة (المعتمد) والمتغير الرئيسى (المستقل) والمسئول عن إحداث تغير فى معامل الاستجابة. وقد سبق شرح كيفية كتابة النموذج الإحصائى أو الرياضى فى الباب التاسع.

ويمكن كتابة النموذج الرياضى لهذا التصميم كما يلى:

متغير الاستجابة = ثابت + أثر المعاملة + خطأ

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

حيث: $i = 1,2,3,4$ أى $k = 4$ ، $j = 1,2,\cdots,8$ أى $n = 8$

فالقيمة Y_{11} تعنى أنها المشاهدة الأولى فى المستوى الأول للمعاملة (أى الحيوان الأول فى المستوى الأول)، Y_{12} تمثل الحيوان الثانى فى المستوى الأول، بينما Y_{43} تمثل الحيوان الثالث فى المستوى الرابع للمعاملة ... وهكذا. ومن الواضح أن هذا التصميم لا يخرج عن كونه بيانات منظمة فى اتجاه واحد، كتلك التى تم مناقشتها وتحليل مثلها فى الباب التاسع. وكل قيمة من قيم Y هى محصلة لمتوسط عام μ

يضاف إليه أثر المعاملة τ_i بجانب خطأ عشوائى ممثل فى ε_{ij} وهو الخاص بكل مشاهدة منفردة. والمثال التالى يوضح كيفية تحليل هذا النوع من التصميمات.

مثال ١٥-١

فى تجربة لدراسة أثر إضافة فيتامين ب على النمو فى الدجاج، جربت ثلاث مستويات من الفيتامين: 20 ,10 ,0 مجم لكل كج من وزن الجسم، وجرب كل مستوى على 4 طيور. ويوضح جدول ١٥-١ أوزان الطيور عند عمر ٩ أسابيع.

جدول ١٥-١ وزن الطيور (كج) عند ٩ أسابيع من العمر بعد معاملتها بفيتامين ب

مستوى المعاملة مجم/كج من وزن الجسم	الوزن	المجموع	المتوسط
0	1.43, 1.22, 1.24, 1.00	4.89	1.2225
10	1.60, 2.00, 1.83, 1.88	7.31	1.8275
20	2.18, 1.92, 1.89, 1.80	7.79	1.9475
الكلى		**19.99**	**1.66**

وطبقاً للنموذج الإحصائى فإن التباين بين هذه القيم يمكن إرجاعه إلى المتوسط (أى معامل التصحيح) بالإضافة إلى تباين راجع إلى المعاملة t بالإضافة إلى تباين راجع إلى الخطأ أو الصدفة e وهو الشئ الذى لا يمكن للتجربة أن تفسره أكثر من هذا، كأن تكون الطيور مختلفة وراثياً إلى حد ما أو أى عوامل أخرى. وعلى هذا فيكون:

مجموع مربعات الانحرافات الكلية عن المتوسط:

$$= \sum\sum Y_{ij}^2 - \frac{(\sum\sum Y_{ij})^2}{nk}$$

$$= 1.43^2 + 1.22^2 + \cdots + 1.89^2 + 1.80^2 - \frac{(19.99)^2}{12} = 1.465$$

والمكون الأخير فى المعادلة يطلق عليه معامل التصحيح كما سبق ذكره من قبل.

مع تحيات د. سلام حسين الهلالي salamalhelali@yahoo.com

مجموع المربعات بين المعاملات:

$$= \sum_{i.} \frac{Y_{i.}^2}{n} - CF = \frac{4.89^2 + 7.31^2 + 7.79^2}{4} - \frac{(19.99)^2}{12} = 1.208$$

مجموع المربعات الراجعة للخطأ أو المتبقى فى هذه الحالة:

$$= 1.465 - 1.208 = 0.257$$

وهذه القيمة الأخيرة هى نفسها الممكن الحصول عليها بحساب مجموع مربعات الانحرافات داخل كل معاملة (كل له 3 درجات حرية) ثم جمعها أى:

$$= \sum_{i}[(\sum_{j} Y_{ij}^2 - \frac{Y_{i.}^2}{n})]$$

$$= \left(1.43^2 + \cdots + 1.00^2 - \frac{4.89^2}{4}\right) + \left(1.60^2 + \cdots + 1.88^2 - \frac{7.31^2}{4}\right)$$

$$+ \left(2.18^2 + \ldots + 1.80^2 - \frac{7.79^2}{4}\right) = 0.257$$

ويمكن وضع النتائج فى جدول تحليل التباين analysis of variance (ANOVA) كما فى جدول ١٥-٢.

جدول ١٥-٢ تحليل التباين (ANOVA) لتجربة دراسة أثر إضافة فيتامين ب إلى العليقة على تسمين الدجاج.

SOV	df	SS	MS	EMS
بين المعاملات	$k-1$ $3-1=2$	1.208	0.604**	$\sigma_e^2 + 4\sigma_t^2$
داخل المعاملات أو الخطأ أو المتبقى	$k(n-1)$ $3(4-1)=9$	0.257	0.029	σ_e^2
الكلى عن المتوسط	$kn-1$ $(3)(4)-1=11$	1.465		

١٥-٦-٣ الافتراضات اللازمة لإجراء تحليل التباين

لا يمكن إجراء تحليل التباين رياضياً إلا إذا فرضت افتراضات معينة على التقديرات للمعالم المنصوص عليها فى النموذج الإحصائى. وهناك العديد من هذه الافتراضات التى يمكن وضعها، ولكن على المجرب أن يختار أنسبها وأكثرها سهولة فى التفسير عند الحصول على النتائج. ففى المثال السابق ١٥-١ يمكن القول إن المجرب يريد فقط تقدير الفروق بين مستويات المعاملة وبعضها البعض، أو بين كل مستوى ومستوى معاملة معينة أخرى، وهذا هو الأهم من الناحية التجريبية. ومن الناحية النظرية فإن المراد الأول غير قابل للتحقيق لأسباب قد تبدو فيما بعد لبعض القراء، وقد يتطلب شرحها تفصيلاً لأساليب أبعد من مجال هذا المؤلف. بينما يمكن تقدير المطلب الثانى كما سيتضح فيما بعد. وإذا وضعت كل المعالم المراد تقديرها فى معادلات آنية لمحاولة حلها يمكن الحصول على ما يسمى بالمعادلات الاعتيادية normal equations. ففى المثال السابق وطبقا للنموذج فإن المجرب يود تقدير المتوسط μ وأثر كل من المستويات الثلاثة للمعاملة حيث يعرف أثر مستوى المعاملة بأنه متوسط مستوى المعاملة مطروحا منه المتوسط العام. فالمجموع الكلى 19.99 وطبقاً للنموذج محصلة لجمع 12 متوسط μ وأربعة أثار لكل مستوى. بينما يكون مجموع الأربعة طيور فى مستوى المعاملة محصلة لجمع أربعة متوسطات بالإضافة إلى أربعة أثار من المستوى الأول ... وهكذا. ويمكن وضع ذلك فى المعادلات الاعتيادية كما يلى:

$$14\hat{\mu} + 4t_1 + 4t_2 + 4t_3 = 19.99 \qquad (١)$$
$$4\hat{\mu} + 4t_1 = 4.89 \qquad (٢)$$
$$4\hat{\mu} + 4t_2 = 7.31 \qquad (٣)$$
$$4\hat{\mu} + 4t_3 = 7.79 \qquad (٤)$$

وتسمى المعادلة (١) بمعادلة المتوسط والمعادلة (٢) بمعادلة المستوى الأول والمعادلة (٣) بمعادلة المستوى الثانى والمعادلة (٤) بمعادلة المستوى الثالث. وواضح أنه، طبقا للنموذج، فإن لكل معلومة فيه يراد تقديرها يوجد معادلة اعتيادية. ويمكن النظر إلى هذه المعادلات على إنها معادلات آنية أى أن هناك أربعة مجاهيل (μ، t_1، t_2، t_3) ويراد تقديرها من أربعة معادلات. ولكن بالتدقيق فى المعادلات يتضح أنها غير مستقلة عن بعضها non-orthogonal linearly، بمعنى أن أحد المعادلات محصلة للثلاثة الباقية فمثلا (١) = (٢) + (٣) + (٤) أو (٢) = (١) − (٣) − (٤) ... وهكذا. ولذا فهذه المعادلات لا يمكن حلها آنيا بالطرق العادية. ولكن مثلاً إذا وضع المجرب شرط أو افتراض constraint أن مجموع آثار المعاملات مساوى

للصفر أى $\sum \hat{t}_i = 0$ فتصبح هذه المعادلات ممكنة الحل. ومثل هذا الافتراض لن يعقد من إمكانية المجرب فى تفسير نتائجه لأن الفروق بين المعاملات ستظل كما هى لا تتغير بالرغم من هذا الافتراض.

وبتطبيق هذا الفرض على المعادلة (١) يصبح:

$$12\,\hat{\mu} + 4(t_1 + t_2 + t_3) = 12\hat{\mu} = 12.99$$

ومنها:

$$\hat{\mu} = \frac{19.99}{12} = 1.66\,\text{kg}$$

وبالتالى يمكن حساب أثر المستوى الأول للمعاملة باستخدام المعادلة (٢) كالتالى:

$$t_1 = \frac{4.89}{4} - 1.66 = -0.44\,\text{kg}$$

وأثر المستوى الثانى للمعاملة من المعادلة (٣):

$$t_2 = \frac{7.31}{4} - 1.66 = 0.16\,\text{kg}$$

وأثر المستوى الثالث للمعاملة من المعادلة (٤):

$$t_3 = \frac{7.79}{4} - 1.66 = 0.29\,\text{kg}$$

ويلاحظ أن $t_1 + t_2 + t_3 = 0$ (فيما عدا فروق التقريب). وبدون هذا الشرط وشرط أخر أن $\sum e_{ij} = 0$ حيث e_{ij} هى تقدير ε_{ij} لما أمكن الحصول على هذه التقديرات.

ويمكن القول بأن كل فرض كهذا يفقد من أجله درجة حرية واحدة. ففى المثال السابق كان مجموع المربعات حول المتوسط له 11 درجة حرية لأنه يشترط أن مجموع الانحرافات حول المتوسط يساوى صفراً بمعنى أن $\sum (Y_{ij} - \mu) = 0$. وبين المعاملات يفقد درجة حرية نتيجة للشرط أن $\sum t_i = 0$ ... وهكذا.

عموماً فإن الافتراضات الخاصة بهذا التصميم هى $\sum t_i = \sum e_{ij} = 0$. ويمكن للمجرب أن يضع افتراضات أخرى مثل إن أحد المعاملات = صفر ... الخ.

١٥-٦-٤ اختبارات الفروض الإحصائية

قبل إجراء اختبارات الفروض الإحصائية يجب على المجرب:

١- تقدير ما إذا كانت المعاملات فى التجربة عشوائية أم ثابتة، وهذا مهم فى كيفية صياغة الفرض، وأيضاً فى استخدام النتائج المتحصل عليها. فالافتراض بأن المعاملة ثابتة يعنى أن المجرب يحصر اهتمامه الأساسى فى هذه المعاملات بعينها دون غيرها ولا يمكن بذلك تعميم النتائج على معاملات أخرى لم تدخل في نطاق التجربة. فإذا كان المجرب يدرس الفروق بين سلالات مثلا، واختار السلالات أ، ج، د بعينها، فالنتائج المتحصل عليها تنحصر فقط على هذه السلالات الثلاثة دون غيرها، ويمكن للمجرب أن يرتبها طبقا لقيم متوسطاتها. بينما إذا كان لدى المجرب العديد من السلالات ويود دراسة الفروق بين هذه السلالات واختار عشوائياً ثلاثة منها وكانت بينها فروق فهذا ينطبق على العشيرة الأصلية التي أخذت منها السلالات عشوائياً، أى أن هذه السلالات مختلفة. بينما إذا لم يجد فرقاً فيمكن التعميم بأن هذه السلالات لا تختلف عن بعضها البعض.

وفى حالة الافتراض بأن المعاملات ثابتة فإن فرض العدم يكون:

$$H_\circ : \tau_1 = \tau_2 = \tau_3$$

وهو نفسه:

$$H_\circ : \mu_1 = \mu_2 = \mu_3$$

حيث μ_1، μ_2، μ_3 تمثل متوسطات مستويات المعاملة.

والفرض البديل H_1 هو أن بعض أوكل متوسطات مستويات المعاملة غير متساوية.

وفى حالة ما إذا كانت المعاملات عشوائية فإن:

$$H_\circ : \sigma_t^2 = 0$$

$$H_1 : \sigma_t^2 \neq 0$$

أى أنه فى حالة عشوائية المعاملات فإن المجرب يكون أكثر اهتماماً بالتباين بين المعاملات عن الفرق بين المتوسطات، حيث إن هذه المتوسطات تعنى قليلاً لأنها لأقسام مأخوذة عشوائيا. وإن كانت طريقة اختبار الفروض الإحصائية واحدة في الإثنين بالنسبة لهذا التصميم.

وتوقع متوسطات مربعات الانحرافات (EMS) فى جدول ١٥-٢ وهى للمعاملات العشوائية. أما فى حالة المعاملات الثابتة فإنها تكون نفس الشئ ماعدا استبدال القيمة K_t^2 بالقيمة σ_t^2 حيث إن:

$$K_t^2 = \frac{\sum t_i^2}{k-1}$$

٢- قبل الإجراء الفعلى لاختبارات الفروض يجب على المجرب مراجعة الافتراضات اللازمة لذلك مثل أن تتوزع e_{ij} فى النموذج حسب التوزيع الطبيعى، وتجانس التباين واستقلالية الخطأ كما سبق إيضاحه فى الباب الثانى عشر. وبعد توافر هذه الافتراضات فإن الفروض الإحصائية يمكن اختبارها باختبار F.

ومن جدول ١٥-٢ فإن F المحسوبة $F = 0.604 \div 0.029 = 20.83$ بدرجات حرية البسط والمقام أى 2، 9 أما F الجدولية $F_{(2,9,0.01)} = 8.02$ وبالتالى يرفض فرض العدم. فإذا كانت المعاملات ثابتة يستنتج المجرب أن هناك فروقاً بين متوسطات المعاملات. أما إذا كانت المعاملات عشوائية فيكون الاستنتاج أن هناك تبايناً بين عشيرة المعاملات التى أخذت منها هذه العينة من المعاملات الثلاث.

١٥-٦-٥ طرق فصل المتوسطات Mean Separation Procedures

عادة ما يرغب المجرب، كما سبق الإشارة فى الباب التاسع، أن يذهب بتحليل تجربته إلى تفاصيل أكثر من مجرد الانتهاء عند المقولة أن متوسطات المعاملات تختلف عن بعضها وذلك في حالة المعاملات الثابتة. فغالباً ما يهتم المجرب بأن يقارن أى من هذه المعاملات أفضل أو أسوأ من معاملة أخرى بعينها، وهذا ما يطلق عليه فصل المتوسطات. وقد يستخدم اختبار t المستقل أو اختبار F المستقل أو LSD أو أى من الاختبارات السابق الإشارة إليها فى الباب التاسع.

١٥-٦-٥-١ اختبار t المستقل Orthogonal t Test

كما سبق بيانه أن بين كل عدد (t) من المعاملات يوجد فقط عدد $(t-1)$ من المقارنات المستقلة بين هذه المعاملات. ففى المثال ١٥-١ حيث هناك ٣ معاملات فإن عدد المقارنات المستقلة هو اثنين فقط. فالمقارنات كلها هى:

أ- المعاملة (١) - المعاملة (٢) أى: $1.22 - 1.83 = -0.61$ kg

ب- المعاملة (١) – المعاملة (٣) أى: $1.22 - 1.95 = -0.73$ kg

ج- المعاملة (٢) – المعاملة (٣) أى: $1.83 - 1.95 = -0.12$ kg

وأن هناك مقارنتين مستقلتين فقط بمعنى أن أى مقارنة من الثلاث مقارنات السابقة يمكن حسابها من المقارنتين الأخريين:

المقارنة أ : ب – ج، أى: $-0.73 - (-0.12) = -0.61$ kg

المقارنة ب: أ + ج، أى: $-0.61 + (-0.12) = -0.73$ kg

المقارنة ج : ب – أ، أى: $-0.73 - (-0.61) = -0.12$ kg

ويمكن للمجرب أن يصمم مقارناته المستقلة بين المعاملات بحيث تعكس ما يريده، ففى المثال السابق إذا أراد المجرب مقارنتين مستقلتين فيجرى ما يلى:

١- توضع معاملات coefficients ولتكن λ_i لكل معاملة لتعكس ما يريده.

٢- يختبر إذا كانت مجموع المعاملات يساوى صفر، أى $\sum \lambda_i = 0$ لكل مقارنة.

٣- يختبر إذا كان مجموع حاصل ضرب معاملات المقارنتين يساوى صفر أيضا، أى $\sum \lambda_i \lambda_{i'} = 0$ حيث i إحدى المقارنات، i' مقارنة أخرى.

فمثلا:

المعاملة		(١)	(٢)	(٣)
المتوسط		1.22	1.83	1.95
مجموعة مقارنات أولى	λ_i	1	-0.5	-0.5
	$\lambda_{i'}$	0	1	-1
مجموعة مقارنات ثانية	λ_i	1	0	-1
	$\lambda_{i'}$	1	-2	1
مجموعة مقارنات ثالثة	λ_i	1	-1	0
	$\lambda_{i'}$	1	0	-1

ففى مجموعة المقارنات الأولى يعنى المجرب أن يحسب الفرق بين المعاملة (١) ومتوسط المعاملتين (٢)، (٣) فى المقارنة الأولى وأن يحسب الفرق بين المعاملة (٢) والمعاملة (٣) في المقارنة الثانية. وبالتالي فإن:

مجموع معاملات المقارنة الأولى: $\sum\lambda_i = 1 + (-0.5) + (-0.5) = 0$

مجموع معاملات المقارنة الثانية: $\sum\lambda_{i'} = 0 + 1 + (-1) = 0$

مجموع حاصل ضرب معاملات المقارنتين:

$$\sum\lambda_i\lambda_{i'} = (1)(0) + (-0.5)(1) + (-0.5)(-1) = 0$$

إذن هاتان المقارنتان مستقلتان عن بعضهما. أما مجموعة المقارنات الثانية فإن المجرب يعنى مقارنة المعاملة (١) بالمعاملة (٣) فى المقارنة الأولى ويعنى مقارنة متوسط المعاملتين (١)، (٣) بالمعاملة (٢) وبالتالى فإن:

مجموع معاملات المقارنة الأولى: $\sum\lambda_i = 1 + 0 + (-1) = 0$

مجموع معاملات المقارنة الثانية: $\sum\lambda_{i'} = 1 + (-2) + 1 = 0$

مجموع حاصل ضرب معاملات المقارنتين:

$$\sum\lambda_i\lambda_{i'} = (1)(1) + (0)(-2) + (-1)(1) = 0$$

وبالتالى فإن مقارنتى هذه المجموعة، أيضا، مستقلتان. أما مجموعة المقارنات الثالثة فإن المجرب يرغب فى مقارنة المعاملة (١) بالمعاملة (٢) ومقارنة المعاملة (١) بالمعاملة (٣) وبالتالى فإن:

مجموع معاملات المقارنة الأولى: $\sum\lambda_i = 1 + (-1) + 0 = 0$

مجموع معاملات المقارنة الثانية: $\sum\lambda_{i'} = 1 + (0) + (-1) = 0$

مجموع حاصل ضرب معاملات المقارنتين:

$$\sum\lambda_i\lambda_{i'} = (1)(1) + (-1)(0) + (0)(1) \neq 0$$

وبهذا لم يتحقق شرطا استقلال المقارنات. ولذا فإن مقارنتى المجموعة الثالثة غير مستقلتين non-orthogonal comparisons or contrasts.

وفى اختبار t المستقل يختار المجرب مجموعة مقارنات مستقلة واحدة فقط والتي تعكس ما يريده. فمثلا بالنظر إلى مجموعة المقارنات الأولى فإن هناك منطقاً واضحاً فى تصميم هذه المقارنة، ففى المقارنة الأولى يحسب المجرب الفرق بين الطيور التى أعطيت الفيتامين (أى العليقتين ٢، ٣) وتلك التى لم تعط أى فيتامين (أى المعاملة ١، والتى تسمى كنترول فى هذه الحالة). بينما فى المقارنة الثانية فإن المجرب يحسب الفرق بين المعاملتين (٢) و(٣)، بمعنى أى المستويين أدى إلى نمو أكبر، مستوى 10 مجم أو مستوى 20 مجم.

إجراء اختبار t المستقل

باستخدام المعادلة التالية:

$$t = \frac{\sum \lambda_i \overline{Y}_i}{\sqrt{\frac{\sum \lambda_i^2}{n_i} S^2}}$$

حيث S^2 من جدول ١٥-٢ هى متوسط المربعات داخل المعاملات أو الخطأ أو المتبقى (0.029)، $n = 4$ أى عدد الطيور فى كل معاملة.

وبالتالى فإن t للمقارنة الأولى:

$$t = \frac{(1)(1.2225) + (-0.5)(1.8275) + (-0.5)(1.9475)}{\sqrt{\frac{0.029}{4}\left[1^2 + (-0.5)^2 + (-0.5)^2\right]}} = \frac{0.665}{0.104} = 6.394$$

و t للمقارنة الثانية:

$$t = \frac{(0)(1.2225) + (1)(1.8275) + (-1)(1.9475)}{\sqrt{\frac{0.029}{4}\left[0^2 + 1^2 + (-1)^2\right]}} = \frac{0.12}{0.12} = 1.00$$

والمقارنة الأولى هى اختبار لفرض العدم أن متوسط المعاملة (١) – [متوسطى المعاملتين (٢)، (٣)] = صفر، والفرض البديل أن الفرق لا يساوى صفراً، بمعنى أن:

فرض العدم: $H_o : \mu_1 = (\mu_2 + \mu_3)/2$

والفرض البديل: $H_1 : \mu_1 \neq (\mu_2 + \mu_3)/2$

بينما فرض العدم للمقارنة الثانية هو أن متوسط المعاملة (٢) – متوسط المعاملة (٣) = صفر والفرض البديل أن الفرق لا يساوى صفراً، أى:

فرض العدم: $H_\circ : \mu_2 - \mu_3 = 0$

والفرض البديل: $H_1 : \mu_2 - \mu_3 \neq 0$

وتختبر قيمة t أمام درجات الحرية لـ S^2 من جدول ١٥-٢ والتى تساوى 9 تحت مستوى $\alpha = 5\%$ أى أن $t_{(9, 0.05)} = 2.262$. فى الحالة الأولى t معنوية، وفى الحالة الثانية غير معنوية. أى أن هناك زيادة معنوية فى وزن الجسم نتيجة لإضافة الفيتامين، ولكن الفرق بين إضافة 10 ملجم وإضافة 20 مجم غير معنوى.

١٥-٦-٥-٢ تقسيم التباين بين المعاملات وإجراء اختبار F المستقل

من جدول ١٥-٢ مجموع المربعات بين المعاملات هو 1.208 بدرجتى حرية. وإذا كانت المقارنتان مستقلتين، فإنه يمكن تقسيم التباين بين المعاملات إلى أجزاء كل منها يقابل مقارنة، وبحيث أن مجموع هذه الأجزاء يساوى مجموع المربعات بين المعاملات

$$\text{Sum of squares due to a comparison} = \frac{n\left(\sum \lambda_i \overline{Y}_i\right)^2}{\sum \lambda_i^2} \qquad (١٥-١)$$

ومن المثال ١٥-١ فإن:

مجموع المربعات للمقارنة الأولى:

$$= \frac{4[(1)(1.225) + (-0.5)(1.8275) + (-0.5)(1.9475)]^2}{1^2 + (-0.5)^2 + (-0.5)^2} = \frac{4(0.665)^2}{1.5} = 1.179$$

مجموع المربعات للمقارنة الثانية:

$$= \frac{4[(0)(1.225) + (1)(1.8275) + (-1)(1.9475)]^2}{0^2 + (1)^2 + (-1)^2} = \frac{4(0.12)^2}{2} = 0.029$$

وواضح أن مجموع المربعات بين المعاملات هو: $= 1.179 + 0.029 = 1.208$

وعليه يمكن إعادة كتابة جدول تحليل التباين ١٥-٢ أكثر تفصيلاً كما يلى:

SOV	df	SS	MS
بين المعاملات	2	1.208	0.604**
المقارنة الأولى	1	1.179	1.179**
المقارنة الثانية	1	0.029	0.029
الخطأ	9	0.257	0.029
الكلى عن المتوسط	11	1.465	

ومن هذا الجدول يمكن اختبار معنوية المقارنة الأولى بواسطة:

$$F_9^1 = \frac{1.179}{0.029} = 40.6554$$

واختبار معنوية المقارنة الثانية بواسطة:

$$F_9^1 = \frac{0.029}{0.029} = 1.0$$

وهما يقودان لقرارات متطابقة تمام التطابق لاختبارى t المقابلين، ولأن F هنا بدرجة حرية واحدة للبسط فإن للمقارنة الأولى:

$$F = t^2 = (6.394)^2 = 40.655$$

وللمقارنة الثانية

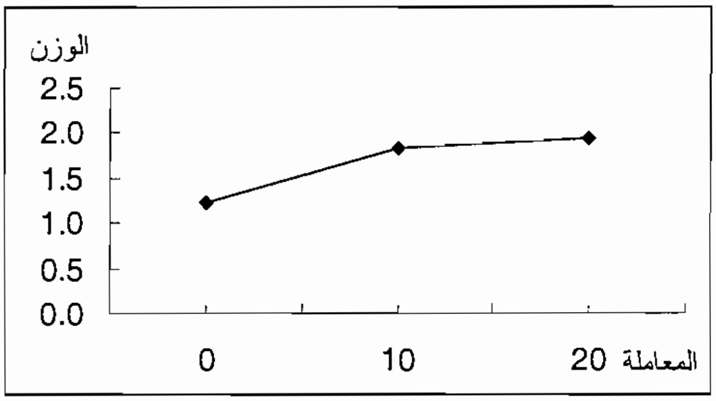
$$F = t^2 = 1$$

ما عدا أخطاء التقريب. كما يمكن إتباع طرق فصل المتوسطات الأخرى مثل LSD أو دنكن أو توكى سابقة الشرح.

١٥-٦-٦ الحدود المتعددة المستقلة (المتعامدة) Orthogonal Polynomials

يمكن التمييز بين نوعين من المعاملات هما معاملات وصفية وأخرى كمية. ففى المعاملات أو الأقسام الوصفية، كل قسم يختلف فى صفته، مثلا كالفرق بين السلالات أو المواقع أو اختلافات راجعة إلى الجنس أو الطلائق ... الخ. بينما المعاملات الكمية فإنها تختلف عن بعضها فى الدرجة كأن تكون المعاملات 20، 30، 40 درجة حرارة مئوية أو 100، 150، 200 كج سماد للفدان أو كما فى المثال ١٥-١ حيث يعطى للطيور 0، 10، 20 مجم من الفيتامين لكل كج من وزن الجسم. وفى كلا النوعين من

المعاملات فإنه يمكن إتباع طرق فصل المتوسطات السابق شرحها. إلا أنه فى حالة المعاملات الكمية هناك وسيلة إضافية لدراسة تأثير المعاملات، فكثيراً ما يود المجرب معرفة العلاقة بين مستوى المعاملة والاستجابة لها ونوع هذه العلاقة، هل هى علاقة خطية أم غير ذلك. ويمكن إجراء هذا بيسر من خلال الحدود المتعددة المستقلة orthogonal polynomial. ففى المثال ١٥-١ قد يتساءل المجرب هل العلاقة بين كمية الفيتامين ونمو الطيور خطية فقط؟ أى أنه بزيادة الفيتامين بوحدة واحدة يزداد النمو بنفس القدر بغض النظر عن مستوى الفيتامين. أم أن الزيادة فى النمو هذه تكون سريعة عند المستويات المنخفضة من الفيتامين ثم تقل سرعة النمو بعد ذلك. فواضح من شكل ١٥-٢ أن بزيادة الفيتامين من 0 إلى 10 مجم فإن معدل الزيادة فى النمو أعلى منه عند زيادة الفيتامين من 10 إلى 20 مجم.



شكل ١٥-٢ العلاقة بين إضافة الفيتامين ومتوسط وزن الجسم في الدجاج

فهناك علاقة خطية ولا شك بمعنى أن هناك زيادة فى الوزن بزيادة الفيتامين، ولكن هذه العلاقة ليست هكذا فقط، لأن معدل التغيير فى الوزن نفسه يتغير. والحدود المتعددة المستقلة orthogonal polynomial تفيد فى:

١- تحديد نوعية هذه العلاقات.

٢- فصل مجموع المربعات بين المعاملات إلى أجزاء يقابل كل منها نوع العلاقة المحددة.

٣- اختبار معنوية هذه العلاقات.

٤- حساب معامل اعتماد المتغير التابع (وهو فى مثال ١٥-١ يعبر عن الوزن) على المتغير المستقل (وهو المعاملة).

ففى حالة وجود مستويين من المعاملة يوجد بينهما درجة حرية واحدة وهى لا تقيس إلا الاتجاه الخطى linear فقط. بينما إذا كان هناك ثلاث معاملات بينهما درجتان حرية فإنه يمكن تقسيمها لتقيس أحدهما الاتجاه الخطى والأخرى لتقيس الاتجاه من الدرجة الثانية (أو التربيعى) quadratic. وفى حالة أربع معاملات بينها ثلاث درجات حرية يمكن تقسيمها إلى اتجاه خطى وآخر تربيعي وأخر من الدرجة الثالثة (أو تكعيبى) cubic ... وهكذا. ولإجراء هذا على مثال ١٥-١ تستخرج معاملات coefficients الحدود المتعددة المستقلة من جدول ١٧ ملحق أ وهى:

مستوى المعاملة	0	10	20
المتوسط	1.2225	1.8275	1.9475
Linear	-1	0	1
Quadratic	1	-2	1

ويلاحظ أنها مستقلة ويتوافر فيها الشرطان $\sum\lambda_i = 0$ ، $\sum\lambda_i\lambda_{i'} = 0$ كما سبق الاشارة فى ١٥-٦-٥-١، وطبقاً للمعادلة (١٥-١) فإن مجموع المربعات الراجعة للأثر الخطى linear :

$$\text{Sum of squares due to linear effect} = \frac{n\left(\sum\lambda_i \overline{Y}_i\right)^2}{\sum\lambda_i^2}$$

$$= \frac{4[(1.225)(-1)+(1.8275)(0)+(1.9475)(1)]^2}{(1)^2+(0)^2+(-1)^2} = 1.051$$

مجموع المربعات الراجعة للأثر التربيعى quadratic :

$$= \frac{4[(1.2225)(1)+(1.8275)(-2)+(1.9475)(1)]^2}{(1)^2+(-2)^2+(1)^2} = 0.157$$

ومجموع الأثرين: $= 1.051 + 0.157 = 1.208$

وهو نفس مجموع المربعات بين المعاملات فى جدول ١٥-٢، ولكنه قسم بطريقة أخرى ويمكن وضع النتائج فى الصورة التالية:

SOV	df	SS	MS
بين المعاملات	2	1.208	0.604**
خطى L	1	1.051	1.051**
الدرجة الثانية Q	1	0.157	0.157*
الخطأ	9	0.257	0.029
الكلى عن المتوسط	11	1.465	

ويمكن اختبار فرض العدم أن الاتجاه الخطى = صفر بواسطة:

$$F = \frac{1.051}{0.029} = 36.24$$

وهذه القيمة معنوية عند 0.01 .

وكذلك اختبار فرض العدم أن الاتجاه من الدرجة الثانية = صفر بواسطة:

$$F = \frac{0.157}{0.029} = 5.41$$

وهذه القيمة معنوية عند 0.05. وبهذا يستنتج أن الاتجاه الخطى موجود أى أن هناك بصفة عامة زيادة فى الوزن مصاحبة للزيادة فى الفيتامين ويستنتج أيضا أن الزيادة فى الوزن هذه تتغير معنوياً بزيادة مستوى الفيتامين كما استدل على ذلك بمعنوية المستوى الثانى من العلاقة.

ويمكن حساب معامل انحدار الوزن على المعاملة كما يلى:

$$b = \frac{\sum \lambda_i \overline{Y}_i}{\sum \lambda_i^2} \qquad (١٥-٢)$$

وعليه يكون معامل الانحدار الخطى:

$$b_L = \frac{0.73}{2} = 0.365 \text{ kg/10 mg}$$

ومعامل الانحدار التربيعى:

$$b_Q = \frac{-0.49}{6} = -0.082 \text{ kg/10 mg}$$

ويمكن التعامل مع b_L، b_Q بنفس كيفية التعامل مع معامل الانحدار البسيط من حيث الخطأ القياسى واختبارات المعنوية ... الخ.

و b_L تعنى أنه فى المتوسط يزيد وزن الطيور بقدر 0.365 كج لكل 10 مجم زيادة فى الفيتامين بينما b_Q تعنى أن هذه الزيادة الخطية تنقص بقدر 0.082 كج / 10 مجم، وذلك واضح من شكل ١٥-٢، واختبار معنوية معاملى الانحدار هذين هو نفسه اختبار المعنوية لمجموع المربعات الراجع إلى الخطى وإلى الدرجة الثانية المبينة فى الجدول السابق.

والمعاملات المبينة في جدول ١٧ ملحق أ تستخدم فقط إذا كان الفرق بين المستويات متساوى أى 0، 10، 20 مثلا. أما إذا كان الفرق بين المستويات غير متساوى كأن يكون 5، 10، 30 مثلا فيمكن إتباع نفس المبادئ، ولكن بطرق حسابية أكثر تعقيداً لن يتم التطرق لها فى هذا المؤلف.

وليس معنى أن متعددة الحدود المستقلة تمكن من تجزئة درجات حرية ومجموع مربعات المعاملة أنه لابد أن يجرى هذا وبالتجزىء الكامل فى كل حالة. فمثلا إذا وجد 6 معاملات بينها 5 درجات حرية فإنه يمكن فصل المكون الخطى، ومكون الدرجة الثانية (التربيعى)، ومكون الدرجة الثالثة (التكعيبى) كل على حدة، بينما يترك مكون الدرجة الرابعة ومكون الدرجة الخامسة مع بعضهما بدرجتى حرية وكثيراً ما يصعب تفسير المكونات الأعلى من الدرجة الثالثة تفسيراً بيولوجياً.

١٥-٦-٧ متوسط المربعات المتوقعة Expected Mean Squares (EMS)

كمتطلب لإجراء اختبار F، ومتطلب لتقدير ثوابت أخرى فى العشيرة سيتضح بعضها فيما بعد، عادة ما يود المجرب أن يرجع متوسط المربعات mean squares أو مجموع المربعات إلى مكوناته أو بمعنى آخر التعبير عنه بدلالة التباينات (أو القيم المربعة) التى تم ذكرها فى النموذج الإحصائى. ففى المثال ١٥-١ يمكن القول إن التباين بين الطيور التابعة لنفس المعاملة تختلف فيما بينها فى e_{ij} وبالتالى فإن التباين بين طيور نفس المعاملة يحتوى فقط على σ_e^2 طبقا للنموذج الإحصائى. أما الاختلاف بين متوسط معاملة ومتوسط معاملة أخرى (وبفرض أن المعاملة هنا عشوائية) فإن هذا الفرق جزء منه مرجعه إلى أثر المعاملة نفسها t، وجزء آخر مرجعه إلى الاختلاف فى قيمة e_{ij} (الخطأ). وعليه فإن التباين بين المعاملات مرجعه إلى σ_e^2، σ_t^2 والمطلوب هنا حساب توقع متوسط المربعات معبراً عنه بدلالة التباينات أو القيم المربعة المنصوص عليها فى النموذج.وفيما يلى قواعد تساعد على هذا الاستنباط بصفة عامة ثم مثال لأبسط الحالات حتى يمكن للقارئ المبتدئ متابعته وتركت الحالات الأكثر تشعباً إلى مجال آخر.

قواعد حساب توقع متوسط المربعات (أو توقع مجموع المربعات):

١- عبر عن مجموع المربعات sum of squares بدلالة الثوابت فى النموذج الأصلى.

٢- اتبع قواعد حساب مربع المجموع ومجموع المربعات فمثلا

$$\begin{aligned}(A+B-C)^2 &= A^2+B^2+C^2+2AB-2AC-2BC\\ &= A^2+B^2+C^2+2(AB-AC-BC)\end{aligned}$$

ومثلاً

$$\sum_{i\neq i'}^{n} a_i a_{i'} + \sum_{i=1}^{n} a_i^2 = \left(\sum_{i=1}^{n} a_i\right)^2$$

$$\left(\sum_{i=1}^{n} a_i + \sum_{j=1}^{m} b_j\right)^2 = \sum_{i=1}^{n} a_i^2 + \sum_{i\neq i'} a_i a_{i'} + \sum_{j=1}^{m} b_j^2 + \sum_{j\neq j'} b_j b_{j'} + 2\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j$$

٣- استنبط توقعات التعبير السابق.

٤- القيمة المتوقعة لتغاير عناصر النموذج مساوية للصفر، لأنه عادة ما يفترض أن المشاهدات مستقلة عن بعضها.

٥- عند تقدير توقع متوسط المربعات EMS لمؤثر ثابت fixed effect فإن المتوقع فى الـــ EMS المتداخلة مع مؤثر عشوائى random effect آخر لا تساوى صفراً. بينما يكون توقع التداخل هذا فى المؤثر العشوائى مساوياً للصفر. لو أن العاملين (المؤثرين) ثابتين فإن القيمة المتوقعة للتداخل فى الـــ EMS تكون مساوية للصفر.

ويمكن توضيح ذلك باستخدام أبسط أنواع النماذج وبفرض أن المعاملة عشوائية:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

حيث

$$\varepsilon_{ij} \sim NID(0, \sigma_e^2) \ ، j=1,2,\cdots,n \ ، i=1,2,\cdots,k$$

وأن t عشوائية بتباين قدره σ_t^2 أى $\tau_{ij} \sim NID(0, \sigma_t^2)$

مجموع المربعات بين المعاملات

$$= n\sum_{i=1}^{k}\left(\frac{\sum_{j=1}^{n} Y_{ij}}{n} - \frac{\sum_{ij} Y_{ij}}{nk}\right)^2 = n\sum_{i=1}^{k}(\overline{Y}_{i.} - \overline{Y}_{..})^2$$

حيث تعبر $\overline{Y}_{i.}$ عن متوسط المعاملة i، $\overline{Y}_{..}$ المتوسط العام.

$$= n\sum_{i=1}^{k}[(\mu + t_i + \frac{1}{n}\sum_{j=1}^{n} e_{ij}) - (\mu + \frac{1}{k}\sum_{i=1}^{k} t_i + \frac{1}{nk}\sum_{ij} e_{ij}]^2$$

$$= n\sum_{i=1}^{k}[(t_i - \frac{1}{k}\sum_{i=1}^{k} t_i) + (\frac{1}{n}\sum_{j=1}^{n} e_{ij} - \frac{1}{nk}\sum_{ij} e_{ij})]^2$$

$$= n\sum_{i=1}^{k}[(t_i^2 + \frac{1}{k^2}\sum_{i=1}^{k} t_i^2 - 2\frac{1}{k} t_i^2)$$

$$+ (\frac{1}{n^2}\sum_{j=1}^{n} e_{ij}^2 + \frac{1}{n^2k^2}\sum_{ij} e_{ij}^2 - 2\frac{1}{n^2k}\sum_{j=1}^{n} e_{ij}^2)] + \text{other products}$$

وبأخذ التوقع فإن مجموع المربعات المتوقع بين المعاملات

(expected sum of squares between treatments)

$$= n\sum_{i}[(\sigma_t^2 + \frac{\sigma_t^2}{k} - 2\frac{\sigma_t^2}{k}) + (\frac{\sigma_e^2}{n} + \frac{1}{nk} - \frac{2}{nk}\sigma_e^2)]$$

$$= nk[\sigma_t^2(1 - \frac{1}{k}) + \sigma_e^2(\frac{1}{n} - \frac{1}{nk})]$$

$$= n\sigma_t^2(k-1) + \sigma_e^2(k-1)$$

ويحسب توقع مجموع المربعات للخطأ (أو بين أفراد نفس المعاملة)

sums of squares due to error (or between individuals within treatment)

$$= \sum_{i}^{k}\sum_{j}^{n}(Y_{ij} - \overline{Y}_{i.})^2 = \sum_{i}^{k}\sum_{j}^{n}(\mu + t_i + e_{ij} - \mu - t_i - \frac{1}{n}\sum_{j=1}^{n} e_{ij})^2$$

$$= \sum_{i}\sum_{j}(e_{ij} - \frac{1}{n}\sum_{j=1}^{n} e_{ij})^2$$

وتكون القيمة المتوقعة:

$$E[\sum_{i}\sum_{j}(e_{ij} - \frac{1}{n}\sum_{j=1}^{n} e_{ij})^2] = \sum_{i}\sum_{j}(\sigma_e^2 + \frac{n}{n^2}\sigma_e^2 - \frac{2}{n}\sigma_e^2)$$

$$= \sum_{i}\sum_{j}[\sigma_e^2(1 - \frac{1}{n})]$$

$$= \sigma_e^2 k(n-1)$$

ويمكن وضع النتائج على هيئة جدول ١٥-٣ التالى:

جدول ١٥-٣ متوسط الانحرافات المتوقعة في أبسط صورة (معاملات عشوائية)

SOV	df	مجموع المربعات المتوقعة Expected sum of squares	متوسط المربعات المتوقعة Expected mean squares (EMS)
بين المعاملات	$k-1$	$\sigma_e^2(k-1) + \sigma_t^2 n(k-1)$	$\sigma_e^2 + n\sigma_t^2$
الخطأ Error (بين الوحدات التجريبية داخل المعاملات)	$k(n-1)$	$\sigma_e^2 k(n-1)$	σ_e^2
الكلى	$kn-1$		

ويمكن الحصول على متوسط المربعات المتوقعة EMS بقسمة مجموع المربعات المتوقعة على درجات الحرية كما فى جدول ١٥-٣. وبما أنه قد افترض أن أثر المعاملات عشوائى random فإن هناك تباين حقيقى بين مستويات المعاملة المختلفة.

ولكن إذا افترض أن أثر المعاملات ثابت fixed فلن يكون هناك تباين حقيقى، ولكنه مجرد قيم مربعة، وفى الحالة البسيطة هذه يظل كل شىء على ما هو عليه إلا أنه تستبدل σ_e^2 بالرمز k_t^2 لتدل على أنها مجرد قيم مربعة.

ويمكن وصف ما تم التوصل إليه فى جدول ١٥–٣ بأن التباين بين أى طائرين فى نفس مستوى المعاملة يتوقع أن يكون σ_e^2 بينما التباين بين أى مستويين مختلفين يتوقع أن يكون $\sigma_e^2 + n\sigma_t^2$ والـــ n توضح أنه فى كل معاملة فإن الوحدة التجريبية (الطائر) تتكرر عدد n من المرات. وبتطبيق هذا الجزء على مثال ١٥–١ يمكن الحصول على تقدير لكل من σ_e^2 و σ_t^2 والتى يطلق عليها اصطلاح مكونات التباين variance components كما هو موضح بجدول ١٥–٢.

وبالتالى فإن: $\sigma_e^2 = 0.029$ ، $\sigma_t^2 = \dfrac{0.604 - 0.029}{4} = 0.144$

وجدير بالذكر أنه يمكن القول بأن جملة التباين المتوقع للقيمة Y_{ij} والتى قد يرمز لها بـــ σ_P^2 أو σ_T^2 هو: $\sigma_P^2 = \sigma_t^2 + \sigma_e^2 = 0.144 + 0.029 = 0.173$

ومن فوائد مكونات التباين variance components أنها تفيد فى معرفة الأهمية النسبية لكل مصدر من مصادر التباين المختلفة إلى المصادر الأخرى فالأهمية النسبية للاختلافات بين المعاملات عبارة عن:

$$\frac{\sigma_t^2}{\sigma_p^2} = \frac{0.144}{0.029 + 0.144} = 0.83$$

حيث أن σ_P^2 تمثل التباين الكلى. بينما الأهمية النسبية للاختلافات داخل المعاملات عبارة عن:

$$\frac{\sigma_e^2}{\sigma_p^2} = \frac{0.029}{0.029 + 0.144} = 0.17$$

ونظرية ومفهوم وتطبيق مكونات التباين لها شأن هام جداً فى مجال الوراثة الكمية وتربية الحيوان وتربية النبات.

مع تحيات د. سلام حسين الهلالي salamalhelali@yahoo.com

١٥-٦-٨ استخدام برنامج SAS فى تحليل التباين للتصميم تام التعشية

يمكن استخدام برنامج SAS لتحليل البيانات المتحصل عليها من تجارب التصميمات تامة التعشية وذلك للحصول على تحليل التباين والتفرقة بين المتوسطات وحساب مكونات التباين ولكن لابد من ملاحظة الأتى:

١- يعطى برنامج SAS أربع أنواع من مجموع المربعات ويطلق عليها Type I SS، Type II SS، Type III SS، Type IV SS وذلك فى حالة تحليل التباين.

٢- يمكن إجراء تحليل التباين باستخدام طريقتين:

أ- استخدام طريقة PROC ANOVA وهذه تستخدم فقط فى حالة البيانات المتزنة والأربعة أنواع من مجموع المربعات متساوية ولا يوجد فرق بينها.

ب- استخدام طريقة PROC GLM وهى تصلح للاستخدام فى جميع أنواع البيانات سواء متزنة أو غير متزنة. فى حالة ما إذا كانت البيانات متزنة فإن الأربع أنواع من مجموع المربعات متساوية. أما إذا كانت البيانات غير متزنة مع عدم وجود خلايا مفقودة missing cells فإن Type III SS يتساوى مع Type IV SS. وهذا هو مجموع المربعات المطلوب والذى سوف يختلف عن Type I SS، Type II SS. فى حالة وجود خلايا مفقودة فإن تقدير Type III SS يختلف عن تقدير Type IV SS. وفى هذه الحالة فإن اختبارات Type III SS لها خاصية الاستقلال orthogonal property أما اختبارات Type IV SS فلها خاصية الاتزان balancing property.

٣- فى حالة البيانات غير المتزنة، سواء عدم تساوى الوحدات التجريبية أو وجود خلايا مفقودة فإنه يلزم تقدير متوسطات أقل مجموع مربعات least-squares means (وقد يسمى population marginal means) وليس المتوسطات الحسابية arithmetic means. متوسطات أقل مجموع مربعات عبارة عن القيمة المتوقعة لمتوسط الفئة class أو تحت الفئة subclass كما لو كانت البيانات متزنة مع الأخذ فى الاعتبار القيمة المتوسطة لجميع المتغايرات covariates إذا وجدت. ويستخدم اختيار LSMEAN لتقدير هذه المتوسطات مع PROC GLM.

٤- عند الرغبة فى حساب المقارنات المستقلة orthogonal comparisons يمكن عمل ذلك باستخدام اختيار CONTRAST والذى يأتى بعد النموذج model فى طريقة PROC GLM.

٥- تستخدم PROC VARCOMP عند الرغبة فى حساب مكونات التباين variance components. وفى هذه الحالة يلزم تحديد طريقة حساب هذه المكونات حيث إنه يوجد عدة طرق للحساب منها على سبيل المثال TYPE1 و MIVQUE0 و ML (maximum likelihood). وهذه الطرق خارج نطاق هذا الكتاب.

مثال ١٥-٢

حل مثال ١٥-١ باستخدام

PROC ANOVA و PROC GLM و PROC VARCOMP

```
DATA POULTRY;
INPUT VITAMIN GROWTH @@;
CARDS;
0 1.43 0 1.22 0 1.24 0 1
10 1.6 10 2 10 1.83 10 1.88
20 2.18 20 1.92 20 1.89 20 1.8
*-----  PROC ANOVA ----------;
PROC ANOVA;
CLASS VITAMIN;
MODEL GROWTH = VITAMIN;
MEANS VITAMIN/LSD;
*---- PROC GLM -----------------;
PROC GLM;
CLASS VITAMIN;
MODEL GROWTH = VITAMIN/SS3;
MEANS VITAMIN/LSD;
CONTRAST 'TRT 1 VS AV TRT 1 & 2' VITAMIN 1 - 0.5  -0.5;
CONTRAST 'TRT 2 VS 3' VITAMIN 0 1 -1;
*----- PROC VARCOMP --------;
PROC VARCOMP METHOD = TYPE1;
CLASS VITAMIN;
MODEL GROWTH = VITAMIN;
RUN;
```

لاحظ

عند استخدام اختيار MEANS فإن المتغيرات التى تأتى مع هذا الاختيار لابد أن تكون مذكورة مع أمر CLASS.

يمكن استخدام اختبار (t) أو أقل فرق معنوى LSD أو دنكن DUNCAN أو توكي Tukey للتفرقة بين المتوسطات فى حالة معنوية اختبار F. وجميع هذه الاختبارات تأتى بعد اختيار MEANS.

عند استخدام طريقة PROC VARCOMP لتقدير مكونات التباين، يأتn بعدها طريقة التقدير إما TYPE1 أو ML أو M1VQUE0.

يكتفى بكتابة الأمر RUN مره واحدة فى نهاية البرنامج.

نتائج التحليل:

The ANOVA Procedure

Class Level Information

Class	Levels	Values
VITAMIN	3	0 10 20

Number of observations 12

Dependent Variable: GROWTH

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1.20806667	0.60403333	21.15	0.0004
Error	9	0.25702500	0.02855833		
Corrected Total	11	1.46509167			

R-Square	Coeff Var	Root MSE	GROWTH Mean
0.824567	10.14460	0.168992	1.665833

Dependent Variable: GROWTH

Source	DF	Anova SS	Mean Square	F Value	Pr > F
VITAMIN	2	1.20806667	0.60403333	21.15	0.0004

t Tests (LSD) for GROWTH

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	9
Error Mean Square	0.028558
Critical Value of t	2.26216
Least Significant Difference	0.2703

Means with the same letter are not significantly different.

t Grouping	Mean	N	VITAMIN
A	1.9475	4	20
A	1.8275	4	10
B	1.2225	4	0

The GLM Procedure

Class Level Information

Class	Levels	Values
VITAMIN	3	0 10 20

Number of observations 12

Dependent Variable: GROWTH

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1.20806667	0.60403333	21.15	0.0004
Error	9	0.25702500	0.02855833		
Corrected Total	11	1.46509167			

R-Square	Coeff Var	Root MSE	GROWTH Mean
0.824567	10.14460	0.168992	1.665833

Dependent Variable: GROWTH

Source	DF	Type III SS	Mean Square	F Value	Pr > F
VITAMIN	2	1.20806667	0.60403333	21.15	0.0004

t Tests (LSD) for GROWTH

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	9
Error Mean Square	0.028558
Critical Value of t	2.26216
Least Significant Difference	0.2703

Means with the same letter are not significantly different.

t Grouping	Mean	N	VITAMIN
A	1.9475	4	20
A	1.8275	4	10
B	1.2225	4	0

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
TRT 1 VS AV TRT 1 & 2	1	1.17926667	1.17926667	41.29	0.0001
TRT 2 VS 3	1	0.02880000	0.02880000	1.01	0.3415

Variance Components Estimation Procedure

Class Level Information

Class	Levels	Values
VITAMIN	3	0 10 20

Number of observations 12

Dependent Variable: GROWTH
Type 1 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square
VITAMIN	2	1.208067	0.604033	Var(Error) + 4 Var(VITAMIN)
Error	9	0.257025	0.028558	Var(Error)

Type 1 Estimates

Variance Component	Estimate
Var(VITAMIN)	0.14387
Var(Error)	0.02856

١٥-٦-٩ معامل الارتباط الجوانى Intra-class Correlation

ينشأ بين الأفراد أو الوحدات التجريبية التى تتبع نفس مستوى المعاملة أو القسم ارتباط بالمضاهاة بأفراد لا تنتمى لنفس مستوى المعاملة أو القسم. وإذا كانت هذه المعاملات أو الأقسام عشوائية فإن النسبة $\sigma_t^2/(\sigma_e^2+\sigma_t^2)$ والتى أطلق عليها Fisher المسمى intra-class correlation بمعنى معامل الارتباط الجوانى (الداخلى) يمكن أن تقيس شدة هذا الارتباط.

ومعامل الارتباط الجوانى بين الطيور التى تنتمى لنفس مستوى المعاملة فى مثال ١٥-١ هو

$$r_I = \frac{0.144}{0.029 + 0.144} = 0.83$$

١٥-٦-١٠ تحليل تصميم تام التعشية فى حالة غياب مشاهدة أو أكثر

إذا فقدت وحدة تجريبية أو أكثر فإنه يتم تعديل التحليل بدرجة طفيفة كما يلى:

١- مجموع المربعات غير المصحح كما هو.

٢- معامل التصحيح يحسب بتربيع المجموع الكلى مقسوما على عدد الوحدات التجريبية منقوصا منها عدد الوحدات التى فقدت.

٣- للحصول على مجموع المربعات بين المعاملات يربع مجموع كل معاملة ويقسم على عدد الوحدات التجريبية فى هذه المعاملة ثم تجمع ويطرح منها معامل التصحيح.

٤- يبقى تحليل التباين كم هو فيما عدا إنقاص درجات الحرية للخطأ بعدد الوحدات المفقودة.

٥- تحسب k (فى مكونات التباين) على أنها المتوسط التوافقى لعدد الوحدات داخل كل معاملة.

بينما إذا فقدت جميع الوحدات التجريبية التابعة لمعاملة معينه فتحذف هذه المعاملة تماما من التجربة ولا سبيل لاسترجاع المعلومات عنها.

صندوق ١٥-١

- يختلف تصميم إحصائى عن آخر فى الطريقة التى يتم بها تعشية الوحدات التجريبية.
- تصميم تام العشوائية هو أبسط التصميمات الإحصائية والذى يجب إتباعه فى حالة عدم وجود دواعى لاستخدام تصميمات أخرى أكثر تعقيدا.
- يشترط فى التصميم تام العشوائية تجانس الوحدات التجريبية.
- يمكن حساب المتوسطات من أى تصميم ولكن لا يجوز اختبار معنوية المتوسطات إلا فى حالة توافر شروط معينة أهمها أن الخطأ فى النموذج الإحصائى يتوزع توزيعا طبيعيا، أنه متجانس وأنه مستقل من وحدة تجريبية إلى أخرى.
- فى حالة عدم توافر شرط أو أكثر من الشروط السابقة فإنه يجب اللجوء إلى طرق إحصائية أخرى.

١٥-٧ تصميم القطاعات العشوائية الكاملة

Complete Randomized Block Design

ويطلق عليه أيضاً القطاعات العشوائية Randomized Block، فى التصميم السابق (تام التعشية) كان الشرط لاستخدامه هو تجانس الوحدات التجريبية. وفى كثير من الأحيان لا يتوافر هذا الشرط وبالتالى فإن عدم التجانس هذا سوف يؤدى إلى إخفاء أثر المعاملات التى يرغب المجرب فى دراستها بالإضافة إلى أن هذه الاختلافات تعتبر من المتغيرات المزعجة nuisance variables والتى يمكن تقليل أثرها عن طريق استخدام تصميم القطاعات العشوائية الكاملة. وبصفة عامة فإن هذا التصميم يستخدم فى حالة:

١- وجود معاملة واحدة لها مستويين تجريبيين أو أكثر.

٢- إمكانية توزيع الوحدات التجريبية على هيئة قطاعات blocks بحيث أن الاختلافات بين الوحدات التجريبية داخل كل قطاع within blocks أقل من تلك التى بين القطاعات among blocks.

٣- التوزيع العشوائى لمستويات التجربة على الوحدات التجريبية داخل كل قطاع.

ويمكن الوصل إلى تجانس الوحدات التجريبية داخل كل قطاع بعدة طرق تعتمد على نوعية الدراسة. فمثلا عندما يكون هناك عدم تجانس فى اتجاه واحد فقط كما فى حالة إجراء تجربة لاختبار مجموعة أصناف جديدة من محصول معين وقطعة الأرض المتاحة للتجربة تقع بجوار قناة رى الأمر الذى يسئ إلى صرف التربة المجاورة للقناة مباشرة، بينما يتحسن الوضع تدريجياً كلما تم الابتعاد عن القناة، كما هو موضح فى شكل ١٥-٣. فلو أن المجرب تجاهل هذا الأمر تماماً وصمم تجربته بالتصميم كامل العشوائية فقد يتكرر صنف مرة أو أكثر في جزء سيئ التربة بالصدفة المحضة أو في جزء جيد من التربة، ويعزى الأداء في هذه الحالة إلي أداء الصنف نفسه مع العلم بأن التربة كانت عاملاً مؤثراً في التجربة. لذا يلجأ المجرب إلى تقسيم قطعة الأرض (أو مادته التجريبية) إلى قطاعات blocks، كل منها متجانس بقدر الإمكان. ثم توزع الأصناف عشوائياً في كل قطاع على حدة، وبذلك يضمن المجرب أنه في كل مستوى من المستويات جودة التربة ممثلة في القطاعات المختلفة وأن كل صنف أو معاملة سوف يمثل. لذا فإن تقسيم المادة التجريبية إلى قطاعات يكون عمودياً على اتجاه الاختلاف كما فى الشكل ١٥-٣. وفى هذه الحالة فإن العامل الأساسى الذى يحدد عدد القطاعات هو الحجم الذى يعتبر متجانساً بقدر الإمكان. وقد لا يكون الاتجاه ثابتاً فى كل المادة التجريبية كأن تزيد خصوبة التربة ثم تقل ثم تزيد مثلاً. وفى هذه الحالة

أيضاً يمكن تقسيم المادة التجريبية إلى قطاعات بحيث فى النهاية يكون كل قطاع متجانساً بقدر الإمكان.

تزداد جودة التربة فى هذا الاتجاه

←

قطاع ٤	قطاع ٣	قطاع ٢	قطاع ١		قناة رى

شكل ١٥-٣ تكوين القطاعات عمودياً على اتجاه عدم التجانس

وفى التجارب الحيوانية قد يختلف العمر فى الحيوانات الأمر الذى قد ينعكس على أداء الحيوانات فى تجارب التسمين وبالتالى فإنه يمكن ترتيب الحيوانات حسب أعمارها وتقسيمها إلى قطاعات (أعمار) بحيث تكون كل مجموعة متجانسة فى العمر إلى حد ما وداخل كل مجموعة توزع مستويات المعاملة عشوائياً. أو قد يكون هناك عدة سلالات مثلا ويراد تجربة علائق تسمين فتعتبر كل سلالة بمثابة قطاع توزع داخله مستويات المعاملة عشوائياً.

وهذا التصميم شائع الاستعمال جداً، وإن كان أصعب درجة من التصميم تام العشوائية إلا أنه ما زال سهل التنفيذ والتحليل والاستنباط.

أما سبب وصف هذا التصميم "بالكاملة complete" لأنه فى بعض الأحيان يكون عدد المعاملات كبير بحيث لا يسمح بوضعها كلها فى قطاع واحد لضمان التجانس. وفى هذه الحالة تقسم المعاملات بطريقة معينة بحيث يحتوى القطاع على جزء فقط من المعاملات وفى هذه الحالة يسمى التصميم "بغير الكاملة incomplete".

١٥-٧-١ التعشية Randomization

بعد تحديد عدد القطاعات يقسم كل قطاع إلى أجزاء متساوية بعدد مستويات المعاملة ثم توزع مستويات المعاملة عشوائياً فى كل قطاع على حدة. ويجب التنويه بوجوب استقلالية عملية التعشية من قطاع إلى آخر. فلو فرض أن هناك 5 مستويات من معاملة معينة (أ – ب – ج – د – هـ) وحدد عدد القطاعات بثلاثة فإنه من الممكن أن تكون الخريطة الواقعية للتعشية كما يلى:

القطاعات		
III	II	I
أ	ب	د
ج	أ	ب
هـ	ج	أ
د	هـ	ج
ب	د	هـ

١٥-٧-٢ النموذج الإحصائى

$$Y_{ij} = \mu + b_i + t_j + e_{ij}$$

حيث Y_{ij} هى المشاهدة التابعة للمعاملة j فى القطاع i، b_i أثر القطاع i، t_j أثر المعاملة j ، e_{ij} الخطأ لهذه المشاهدة. والتحليل فى هذه الحالة شبيه للتحليل ذى الاتجاهين والذى تمثل فيه المعاملة كاتجاه والقطاع كاتجاه آخر (الباب الحادى عشر).

١٥-٧-٣ تحليل القطاعات العشوائية

ليس فى التحليل مبادئ جديدة أكثر من تلك التى تم تناولها عند مناقشة تحليل التباين ذى الاتجاهين (الباب الحادى عشر) وعليه فيكون:

مجموع المربعات الكلية عن المتوسط:

$$\sum y^2 = \sum_{ij} Y_{ij}^2 - \frac{(\sum Y_{ij})^2}{nk}$$

مجموع المربعات بين القطاعات:

$$\sum_i (\sum_j Y_{ij})^2 / k - \frac{(\sum Y_{ij})^2}{nk}$$

مجموع المربعات بين المعاملات:

$$\sum_j (\sum_i Y_{ij}^2) / n - \frac{(\sum Y_{ij})^2}{nk}$$

مثال ١٥-٣

مجرب يود دراسة تأثير أربعة إضافات غذائية كمنشطات نمو على التسمين فى الدجاج بالمقارنة بعدم إضافة أى منشطات. أى أن المجرب لدية معاملة لها خمسة

مستويات هى A، B، C (بدون إضافات للمقارنة)، D، E. وكان لدى المجرب ثلاث بيوت تسمين هذه البيوت قد تختلف فيما بينها من حيث الموقع وتيارات الهواء ... الخ.

قسم المجرب كل بيت إلى خمسة أجزاء متساوية ووزع الخمسة معاملات على لخمسة أجزاء عشوائياً فى كل بيت. ووضع فى كل قطاع عدد متساوى من دجاج للحم المتشابه بحيث كان كل قطاع به 1000 طائر. وفى نهاية فترة التسمين تم وزن لدجاج فى كل جزء جملة واحدة، أى أن الوحدة التجريبية هى مجموعة الطيور فى لجزء الواحد. ويبين جدول ١٥-٤ البيانات التى حصل عليها المجرب.

جدول ١٥-٤ نتائج تجربة أثر منشطات النمو على التسمين في الدجاج والمصممة كقطاعات عشوائية، وزن الدجاج بالكيلوجرام.

المعاملة	القطاع ١	٢	٣	المجموع	المتوسط
A	2540	1905	2275	6720	2240
B	2792	2227	2712	7731	2577
C (المقارنة)	1965	1910	1990	5865	1955
D	2498	2388	2533	7419	2473
E	2480	1920	1960	6090	2030
المجموع	**12275**	**10350**	**11200**	**33825**	
المتوسط	**2455**	**2070**	**2240**		**2255**

مجموع المربعات الكلية غير المصحح

$$\sum_{ij} Y_{ij} = (2540)^2 + (2792)^2 + \ldots + (1960)^2 = 77886949$$

معامل التصحيح

$$CF = \frac{(\sum_{ij} Y_{ij})^2}{nk} = \frac{(33825)^2}{15} = 76275375$$

مجموع المربعات الكلية (عن المتوسط) $= 77886949 - 76275375 = 1611574$

مجموع مربعات القطاعات

$$= \frac{(12275)^2 + (10350)^2 + (11200)^2}{5} - CF = 372250$$

مجموع مربعات المعاملات

$$= \frac{(6720)^2 + (7731)^2 + (5865)^2 + (7419)^2 + (6090)^2}{3} - CF = 876174$$

مجموع مربعات الخطأ $= 1611574 - (372250 + 876174) = 363150$

ويمكن تلخيص نتائج التحليل فى جدول التباين ١٥-٥ التالى:

جدول ١٥-٥ تحليل التباين لتجربة دراسة أثر منشطات النمو على التسمين فى الدجاج والمصممة كقطاعات عشوائية

SOV	df	SS	MS
بين القطاعات Between Blocks	$(n-1) = 2$	372250	186125.0
بين المعاملات Between treatments	$(k-1) = 4$	876174	219043.5*
الخطأ Error	$2 \times 4 = 8$	363150	45393.8
المربعات الكلية عن المتوسط		1611574	

F للقطاعات $F = 186152/45393.8 = 4.1$ ،

F للمعاملات $F = 219043.5/45393.8 = 4.8$

ولاختبار الفرض أنه لا يوجد فروق معنوية بين المعاملات فإن $F = 4.8$ معنوية عند مستوى $\alpha = 0.05$ أى أن فرض العدم يرفض بينما الفروق بين القطاعات غير معنوية.

١٥-٧-٤ طرق فصل المتوسطات

يمكن فصل متوسطات المعاملات بواسطة أى من اختبار LSD أو اختبار HSD (توكى) أو اختبار دنكن Duncan.

١٥-٧-٤-١ اختبار LSD

يتم حساب قيمة LSD عند مستوى معنوية 0.05 و 8 درجات حرية (وهي درجات حرية الخطأ) حيث $t = 2.036$ وبالتالى فإن:

$$LSD = 2.306\sqrt{\frac{2(45393.8)}{3}} = 401.2$$

ترتب قيم متوسطات مستويات المعاملة تنازليا حيث الأكبر فالأصغر فتكون B، D، A، E، C

يتم حساب الفروق بين أزواج المتوسطات ثم مقارنة الفرق الناتج بالقيمة المحسوبة لـــ LSD وهى 401.2 مع وضع علامة (*) للدلالة على معنوية الفرق بين المتوسطين من عدمه كالتالى:

$$B - C = 2577 - 1955 = 622 \text{ kg}^*$$

$$B - E = 2577 - 2030 = 547 \text{ kg}^*$$

$$B - A = 2577 - 2240 = 337 \text{ kg}$$

$$B - D = 2577 - 2473 = 104 \text{ kg}$$

$$D - C = 2473 - 1955 = 518 \text{ kg}^*$$

$$D - E = 2473 - 2030 = 443 \text{ kg}^*$$

$$D - A = 2473 - 2240 = 133 \text{ kg}$$

$$A - C = 2240 - 1955 = 285 \text{ kg}$$

$$A - E = 2240 - 2030 = 210 \text{ kg}$$

$$E - C = 2030 - 1955 = 75 \text{ kg}$$

١٥-٧-٤-٢ اختبار توكى (HSD)

الانحراف القياسى

$$= \sqrt{\frac{45393.8}{3}} = \sqrt{15131.25} = 123 \text{ kg}$$

$$HSD = 4.89\sqrt{\frac{45393.8}{3}} = 601.5 \text{ kg}$$

وطبقاً للـــ HDS فإن الفرق بين المعاملة B والمعاملة C وهو 622 كج معنوياً وبقية الفروق غير معنوية عند مستوى 0.05 .

١٥-٧-٤-٣ اختبار دنكن Duncan

المعاملتين	الفرق المعنوى	الفرق المشاهد
B – C	3.52 x 123 = 433 kg	622*
B – E	3.47 x 123 = 433 kg	547*
B – A	3.39 x 123 = 433 kg	337
B – D	3.26 x 123 = 433 kg	104
D – C	3.47 x 123 = 433 kg	518*
D – E	3.39 x 123 = 433 kg	443
D – A	3.26 x 123 = 433 kg	133
A – C	3.39 x 123 = 433 kg	285
A – E	3.26 x 123 = 433 kg	210
E – C	3.26 x 123 = 433 kg	75

وجدير بالذكر أنه كثيراً ما يراد معرفة إذا كانت المعاملة ككل قد أثرت فى أداء هذه الطيور من عدمه، ويمكن اختبار هذا بمضاهاة متوسط الأربعة منشطات ضد المقارنة (C) كما يلى:

فرض العدم: متوسط المعاملات = متوسط المقارنة

متوسط المعاملات A، B، D، E عبارة عن

$$\begin{aligned}\overline{Y}_t &= (\overline{Y}_A + \overline{Y}_B + \overline{Y}_D + \overline{Y}_E)/4 \\ &= (22440 + 2577 + 2473 + 2030)/4 = 2330 \text{ kg}\end{aligned}$$

ولاختبار فرض العدم يلزم حساب تباين هذا المتوسط

$$\begin{aligned}V(\overline{Y}_t) &= V[(\overline{Y}_A + \overline{Y}_B + \overline{Y}_D + \overline{Y}_E)/4] \\ &= (1/16)V(\overline{Y}_A + \overline{Y}_B + \overline{Y}_D + \overline{Y}_E) \\ &= (1/16)[V(\overline{Y}_A) + V(\overline{Y}_B) + V(\overline{Y}_D) + V(\overline{Y}_E)]\end{aligned}$$

وحيث إن التباين بكل متجانس (فرضاً) فإن:

$$S^2_{\overline{Y}_t} = \frac{1}{16}\left(\frac{4S^2}{3}\right) = \frac{S^2}{12} = \frac{\text{Error MS}}{12}$$

وهذه القيمة فى المثال ١٥-٣ عبارة عن:

$$S^2_{\overline{Y}_t} = 45393.8/12 = 3782.8$$

وبالتالى:

$$t = \frac{\overline{Y}_c - \overline{Y}_t}{\sqrt{S^2_{\overline{Y}_t} + S^2_{\overline{Y}_c}}} = \frac{2330 - 1955}{\sqrt{3782.8 + 15131.25}} = \frac{375}{137.53} = 2.73^*$$

وهذه القيمة معنوية عند مستوى 0.01 وعليه يرفض فرض العدم ويستخلص أن ضافة المنشط بصفة عامة يزيد من النمو.

وللحصول على تقديرات للمتوسطات وإمكانية تحليل التباين يلزم الافتراض أن

$$\sum_{i=1}^{3} b_i = \sum_{j=1}^{5} t_j = \sum_{ij} e_{ij} = 0$$

وذلك فى المعادلات الاعتيادية NE's المذكورة فى جدول ١٥-٦.

وبتطبيق هذه الافتراضات يمكن تقدير كل من:

المتوسط العام $\hat{\mu} = 233825/15 = 2255 \text{ kg}$

أثر القطاع الأول $\hat{b}_1 = \frac{12275}{5} - 2255 = 200 \text{ kg}$

أثر القطاع الثانى $\hat{b}_2 = \frac{10350}{5} - 2255 = -185 \text{ kg}$

أثر القطاع الثالث $\hat{b}_3 = \frac{11200}{5} - 2255 = -15 \text{ kg}$

أثر المعاملة A $\hat{t}_1 = \frac{6720}{3} - 2255 = -15 \text{ kg}$

أثر المعاملة B $\hat{t}_2 = \frac{7731}{3} - 2255 = 322 \text{ kg}$

أثر المعاملة C $\hat{t}_3 = \frac{5865}{3} - 2255 = -300 \text{ kg}$

أثر المعاملة D $\hat{t}_4 = \frac{7419}{3} - 2255 = 218 \text{ kg}$

أثر المعاملة E $\hat{t}_5 = \frac{6090}{3} - 2255 = -225 \text{ kg}$

جدول ١٥-١٠ متوسط المربعات المتوقعة واختبارات المعنوية عند تكرار الوحدة التجريبية فى القطاعات العشوائية

SOV	متوسط المربعات المتوقعة وطريقة الاختبار مبينة بالأسهم			
	كلاهما عشوائى	كلاهما ثابت	B ثابتة، t عشوائية	B عشوائية، t ثابتة
السلالات b	$\sigma_e^2 + 4\sigma_{bt}^2 + 12\sigma_b^2$	$\sigma_e^2 + 12K_b^2$	$\sigma_e^2 + 4\sigma_{bt}^2 + 12K_b^2$	$\sigma_e^2 + 12\sigma_b^2$
المعاملات t	$\sigma_e^2 + 4\sigma_{bt}^2 + 16\sigma_t^2$	$\sigma_e^2 + 16K_t^2$	$\sigma_e^2 + 16\sigma_t^2$	$\sigma_e^2 + 4\sigma_{bt}^2 + 16K_t^2$
الخطأ التجريبى	$\sigma_e^2 + 4\sigma_{bt}^2$	$\sigma_e^2 + 4K_{bt}^2$	$\sigma_e^2 + 4\sigma_{bt}^2$	$\sigma_e^2 + 4\sigma_{bt}^2$
الخطأ العينى	σ_e^2	σ_e^2	σ_e^2	σ_e^2

١٥-٧-٥ حساب الكفاءة النسبية لتصميم القطاعات العشوائية الكاملة

يسمح تصميم القطاعات العشوائية الكاملة للمجرب بتقليل الاختلافات بين الوحدات التجريبية عن طريق فصل جزء من التباينات الكلية والراجع إلى القطاعات وبالتالى خفض قيمة الخطأ التجريبى. ولقد سبق ذكر أن كفاءة أى تصميم تتناسب أساسا تناسبا عكسياً مع قيمة الخطأ التجريبى، وإلى درجة أقل تتناسب طردياً مع درجات الحرية للخطأ. وبعد إجراء وتنفيذ تصميم إحصائى معين قد يتساءل المجرب: هل استفادت التجربة فعلا من التصميم المتبع بالنسبة إلى تصميم آخر كان يمكن إتباعه؟. ففى المثال ١٥-٣ ما الذى استفادته التجربة من كونها صممت على هيئة قطاعات عشوائية كاملة بالنسبة مثلا إلى ما لو اتبع التصميم تام التعشية. ويمكن حساب هذه الكفاءة النسبية relative efficiency (RE) كما يلى:

$$RE = \frac{(n-1)M_B - n(k-1)M_E}{(nk-1)M_E} \qquad (١٥-٣)$$

حيث

MB = متوسط مربعات القطاعات،

ME = متوسط مربعات الخطأ،

n = عدد القطاعات،

k = عدد مستويات المعاملة

وبالتعويض عن هذه الرموز بقيمها فى مثال ١٥-٣ فإن كفاءة تصميم القطاعات العشوائية الكاملة إلى التصميم تام العشوائية:

$$RE = \frac{(3-1)(186125) - 3(5-1)(45393.8)}{(3 \times 5-1)(45393.8)} = 1.44$$

ومعنى هذا أن كفاءة التجربة زادت بمقدار 44% عما إذا كانت أجريت باستخدام التصميم التام العشوائية.

١٥-٧-٦ تقدير القيم الغائبة missing values في تصميم القطاعات العشوائية الكاملة

فى بعض الأحيان تفقد واحدة أو أكثر من الوحدات التجريبية لسبب أو آخر خارجاً عن إرادة المجرب ومن غير أن تكون المعاملات سببا فى هذا الفقد. ففى المحاصيل

الحقلية مثلا قد تأكل الحيوانات أحد الوحدات التجريبية أو قد تروى بالخطأ مما يجعلها غير صالحة لغرض التجربة ... الخ. وفى التجارب الحيوانية قد يموت الحيوان لسبب غير وارد فى التجربة أو قد تقتل الحيوانات البرية بعض حيوانات التجربة ... الخ. فلو فرض فى المثال ١٥-٣ أن المعاملة D فى القطاع (٢) قد فقدت لسبب ما ويود المجرب أن يحلل النتائج بعد تقدير هذه القيمة المفقودة، فعلية إتباع الخطوات التالية:

١- تقدر قيمة للوحدة التجريبية الغائبة من المعادلة التالية:

$$\hat{Y}_m = \frac{k(t_m) - n(b_m) - Y_{..}}{(k-1)(n-1)}$$

حيث

$\hat{Y}_m$ = تقدير القيمة الغائبة،

k = عدد المعاملات،

n = عدد القطاعات،

t_m = مجموع المعاملة التي بها الوحدة التجريبية الغائبة،

b_m = مجموع القطاع الذي به الوحدة التجريبية الغائبة،

$Y_{..}$ = المجموع الكلى.

وفى المثال الحالى فإن القيمة المفقودة للمعاملة D فى القطاع (٢) والتى تأخذ القيمة 2388 فى جدول ١٥-٤ سيتسبب عن فقدها أن مجموع المعاملة D يصبح $t_m = 7419 - 2385 = 5031$، وإن القطاع (٢) سيكون مجموعة $b_m = 7962$ وإن المجموع الكلى سيصبح 31436، وعليه تكون القيمة المفقودة

$$\hat{Y}_m = \frac{(5)(5031) - (3)(7962) - 31437}{(5-1)(3-1)} = 2200.5$$

٢- توضع القيمة المقدرة هذه مكان الوحدة التجريبية المفقودة ويجرى تحليل التباين عادياً تماماً وكأن شيئاً لم يحدث.

٣- توضع النتائج فى جدول تحليل التباين العادى مثل جدول ١٥- ٥ تماماً غير إن درجات حرية كل من التباين الكلى والخطأ تنقص درجة حرية واحدة هى عدد الوحدات الغائبة ويصبح الجدول كالتالى:

SOV	df	SS	MS
بين القطاعات	2	463312.5	231656.2*
بين المعاملات	4	803799	200949.8*
الخطأ	7	327400	46771.4
الكلى	13	1594511.5	

٤- تجرى اختبارات المعنوية كالمعتاد إلا أنه عند مقارنة معاملة وحداتها التجريبية ناقصة فإن التباين بين متوسطيهما هو:

$$M_E\left(\frac{2}{n}+\frac{k}{n(n-1)(k-1)}\right)=46771.4\left(\frac{2}{3}+\frac{5}{3(3-1)(5-1)}\right)=40924.95$$

وإذا فقدت وحدتان تجريبيتان أ ، ب مثلا فتفرض قيمة وسطى لإحداهما ولتكن أ مثلا ثم على أساسها تقدر ب ثم تقدر أ من ب ثم جولة ثانية لتقدير ب من أ ... وهكذا حتى يتقارب تقديريين متتاليين لكل من أ، ب. ويتبع نفس الإجراء فى تحليل التباين إلا أنه ينقص درجتان حرية من كل من الخطأ والتباين الكلى هذا ويجب مراعاة ألا يكون الفقد مركز فى معاملة معينة أو فى قطاع معين مما قد يجعل المعلومات المتحصل عليها من الوحدات المتبقية قليلة القيمة. وفى حالة توافر حاسب إليكترونى فإن تقدير القيم المفقودة، يصبح سهلا ولاسيما فى حالة فقد أكثر من وحدتين تجريبيتين حيث تتعقد فيها طريقة التقدير نسبيا.

١٥-٧-٧ اختبار توكى للتجمعية Tukey's Test for additivity

كما سبق شرحه فى الباب الثانى عشر فإن هناك شروطاً معينة يجب توافرها حتى يتمكن المجرب من إجراء اختبارات الفروض وأن أحد هذه الشروط هى أن تكون العوامل فى النموذج الإحصائى تجمعيه. ففى تصميم القطاعات العشوائية والذى وصف نموذجه بأنه:

$$Y_{ij}=\mu - b_i - t_j - e_{ij}$$

بمعنى أن كل من μ، b، t، e تحدث أثرها بأن يجمع كل أثر على الأخر فلا تكون العلاقة مثلا μbt أو μb^t ... وهكذا.

ويمكن باستخدام اختبار Tukey (1949) لإجراء اختبار ما إذا كانت الآثار تجمعيه من عدمه. ومن فوائد هذا الاختبار أنه:

١- يساعد فى معرفة إذا كانت هناك حاجة إلى تحوير transformation البيانات.

٢- قد يساعد فى اقتراح تحوير معين.

٣- يختبر ما إذا كان التحوير أدى الهدف المرجو منه.

ولن يتم التطرق إلى نظرية هذا الاختبار ولكن سيتم تطبيق الطريقة على مثال ١٥-٣ فى جدول ١٥-٧.

جدول ١٥-٧ اختبار توكى لتجمعية العوامل فى النموذج الإحصائى لتجربة أثر منشطات النمو على الزيادة فى وزن الدجاج.

المعاملة	القطاع ١	القطاع ٢	القطاع ٣	المجموع $Y_{.j}$	المتوسط $\bar{Y}_{.j}$	d_j	$w_j = Y_{ij}d_i$
A	2540	1905	2275	6720	2240	-15	**121450**
B	2792	2227	2712	7731	2577	322	**105725**
C	1965	1910	1990	5865	1955	-300	**9800**
D	2498	2388	2533	7419	2473	218	**19825**
E	2480	1920	1960	6090	2030	-225	**115450**
المجموع $Y_{i.}$	**12275**	**10350**	**11200**	**33825**		**0.0**	**372250**
المتوسط $\bar{Y}_{i.}$	**2455**	**2070**	**2240**		**2255**		
d_i	**200**	**-185**	**-15**	**0.0**			

لاحظ فى جدول ١٥-٧ أن $d_i = \bar{Y}_{i.} - \bar{Y}_{..}$ ، $d_j = \bar{Y}_{.j} - \bar{Y}_{..}$ ، (لاحظ أن: $\sum d_i = \sum d_j = 0$) ، $w_j = \sum Y_{ij} d_i$

فمثلا

$d_{(i=1)} = 2455 - 2255 = 200$ ، $d_{(j=1)} = 2240 - 2255 = -15$ ،

$w_1 = (2540)(200) - (1905)(-185) - (2275)((-15) = 121450$ ... الخ.

لاحظ أن $\sum w_j = 372250$ وهذه القيمة تساوى أيضا

$$\sum w_j = (12275)(200) - (10350)(-185) - (11200)(-15)$$

$$N = \sum w_j d_j = (121450)(-15) - (1055725)(322) - (9800)(-300)$$
$$+ (19825)(218) + (115450)(-225) = 7627300$$

$$\sum d_i^2 = (200)^2 + (-185)^2 + (-15)^2 = 74450$$

$$\sum d_j^2 = (-15)^2 + (322)^2 + (-300)^2 + (218)^2 + (-225)^2 = 292058$$

حساب مجموع المربعات الناجم عن عدم التجمعية

$$\frac{N^2}{\sum d_i^2 \sum d_j^2} = \frac{(7627300)^2}{(74450)(292058)} = 2676$$

ويمكن تطوير جدول تحليل التباين ١٥-٦ كما يلى فى جدول ١٥-٨.

جدول ١٥-٨ تحليل التباين لإجراء اختبار التجمعية.

SOV	df	SS	MS
بين القطاعات	2	372250	186125
بين المعاملات	4	876174	219043.5
الخطأ	8	363150	
الغير تجمعية	1	2676	2676
المتبقى	7	360474	51496
الكلى	14	1611574	

ويختبر فرض أنه لا يوجد هناك تجمعية بواسطة اختبار F حيث البسط هو متوسط المربعات للغير تجمعية والمقام هو متوسط المربعات للمتبقى والذى له 7 درجات حرية. وبالتالى $F = 2676/51496 = 0.052$ وهى غير معنوية، ومن هذا لا يمكن رفض هذا الفرض أى أن الآثار فى النموذج الإحصائى تحدث أثارها بصفة تجمعية.

١٥–٨ تصميم القطاعات العشوائية مع تكرار الوحدات التجريبية

وإن كان تصميم القطاعات العشوائية السابق وصفه، والذى فيه تمثل كل وحدة تجريبية مرة واحدة فقط فى كل قطاع، هو الأكثر شيوعاً فى التجارب الحقلية وبعض التجارب الحيوانية إلا أنه فى كثير من التجارب الحيوانية فإن الوحدة التجريبية تتكرر أكثر من مرة فى نفس المعاملة والقطاع. فمثلا قد تتمثل القطاعات فى الأعمار وداخل كل من الأعمار توزع المعاملات عشوائية بحيث أن كل منها يمثل بحيوانين أو أكثر. فى هذه الحالة وإن كان التصميم أساسا كالذى سبق وصفه إلا أنه جد عنصر جديد هنا هو الفرق بين وحدتين فى نفس القطاع ولهما نفس المعاملة. هذا النوع من الخطأ يطلق عليه خطأ عينى (أى من العينة) sampling error مقارنة بالخطأ الآخر الذى ذكر فى مثال ١٥–١ حيث يطلق عليه خطأ تجريبى experimental error. ويصبح النموذج الإحصائى:

$$Y_{ijk} = \mu - b_i - t_j - e_{ij} - E_{ijk}$$

حيث تمثل μ، b_i، t_j، e_{ij} كما سبق، بينما تمثل E_{ijk} الخطأ العينى.

مثال ١٥–٤

أجريت تجربة لدراسة أثر نسبة العليقة الخشنة إلى العليقة المركزة على وزن الحملان بالكيلوجرام عند عمر ١٠ أسابيع، وكان هناك أربع سلالات، والتى اعتبرها المجرب كقطاعات، وثلاث نسب (المعاملات) كما فى جدول ١٥–٩.

فى النموذج $i = 4$، $j = 3$، $k = 4$

مجموع المربعات الكلية $= 31^2 + 35^2 + \cdots + 11^2 - (960)^2 / 48 = 3302$

مجموع المربعات بين السلالات (أى بين القطاعات)

$$= \frac{(300)^2 + (216)^2 + (216)^2 + (228)^2}{12} - \frac{(960)^2}{48} = 408$$

مجموع المربعات بين المعاملات

$$= \frac{(480)^2 + (256)^2 + (224)^2}{16} - \frac{(960)^2}{48} = 2432$$

جدول ١٥-٩ بيانات مثال ١٥-٤ لدراسة أثر نسبة العليقة الخشنة إلى المركزة على نمو حملان أربع سلالات من الأغنام

المعاملة خشنة : مركزة	السلالة أ	ب	ج	د	المجموع
٣٠ : ٧٠	31	33	24	34	
	35	27	27	29	
	35	28	28	29	
	35	28	29	28	
المجموع	**136**	**116**	**108**	**120**	**480**
٤٥ : ٥٥	19	17	9	16	
	22	16	16	12	
	28	12	14	12	
	27	11	13	12	
المجموع	**96**	**56**	**52**	**52**	**256**
٦٠ : ٤٠	23	15	17	20	
	15	7	16	12	
	16	11	13	13	
	14	11	10	11	
المجموع	**68**	**44**	**56**	**56**	**224**
الإجمالى	**300**	**216**	**216**	**228**	**960**

وحسابياً مجموع مربعات الخطأ التجريبى عبارة عن التداخل interaction بين القطاعات (أى السلالات) والمعاملات وكما سبق فى الباب الحادى عشر فإن الخطأ التجريبى:

$$= \frac{(136)^2 + (116)^2 + \cdots + (56)^2}{4} - \frac{(960)^2}{48} - 408 - 2432 = 112$$

أما الخطأ العينى sampling error فإنه يحسب إما بطرح كل شئ من مجموع المربعات الكلية أى

$$= 3302 - 408 - 2432 - 112 = 350$$

وتكون درجات الحرية له

$$= 47 - 3 - 2 - 6 = 36$$

أو أنه يحسب بجمع مجموعات المربعات الكلية بين كل أربعة حيوانات تابعة لنفس السلالة ونفس المعاملة أى:

$$= [31^2 + \cdots + 35^2 - \frac{(136)^2}{4}] + [33^2 + \cdots + 28^2 - \frac{(116)^2}{4}] + \cdots$$

$$+ [20^2 + \cdots + 11^2 - \frac{(56)^2}{4}] = 350$$

ودرجات الحرية له $(12)(4 - 1) = 36$

ويمكن تلخيص نتائج تحليل التباين لمثال ١٥-٤ (حالة تكرار الوحدة التجريبية) كالتالى:

SOV	df	SS	MS
القطاعات	3	408	136
المعاملات	2	2432	1216
الخطأ التجريبى	6	112	18.7
الخطأ العينى	36	350	9.7
الكلى	**47**	**3302**	

ويبقى السؤال أى الخطأين يستخدم فى اختبارات المعنوية؟ وللإجابة يلزم الأخذ فى الاعتبار ما قيل سالفاً فى الباب الحادى عشر من حيث أن العوامل المختبرة هل هى عشوائية random أو ثابتة fixed، فإذا كان العاملان عشوائيين فيختبر معنويتهما بقسمة كل من متوسط المربعات للعامل على متوسط المربعات للخطأ التجريبى. بينما إذا كانا ثابتين فيقسم كل من متوسط مربعات العامل على متوسط المربعات للخطأ العينى. ويلخص جدول ١٥-١٠ متوسط المربعات المتوقعة EMS واختبارات المعنوية تحت الافتراضات المختلفة.

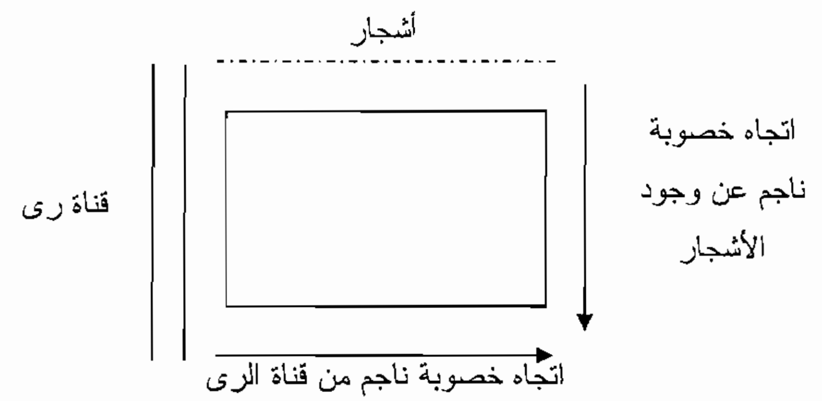
وفى محاولة لجعل اختبار المعنوية أكثر قوة قد يلجأ بعض المجربين إلى اختبار معنوية الخطأ التجريبى أولاً وذلك بقسمة متوسط المربعات التابعة له على تلك التابعة للخطأ العينى فإذا لم تكن F معنوية استدل من هذا على أنهما لا يختلفان عن بعضهما باحتمال خطأ معين وعليه فيقوم المجرب بضم مجموع مربعاتهما سوياً وكذلك درجات حريتهما ويعتبرهما كأنهما خطأ واحد. وفى المثال المذكور سيكون حاصل جمع

مجموعى مربعات الخطأين 462 = 350 - 112 بدرجات حرية 42 = 36 - 6 وهذا يعطى متوسط مربعات قدره 11 .

ولا ينصح بمثل هذا الإجراء إلا إذا كانت درجات الحرية للخطأ التجريبى فعلا مخفضة وإن المجرب فى حاجة حقيقية إلى درجات حرية أكبر حتى تزيد من قوة اختبار المعنوية. وحتى عند إتباع هذا الإجراء فيجب على المجرب أن يكون واعيا إلى أن مستوى المعنوية الذى يجرى عنده الاختبارات قد تغير. ذلك لأنه أصبح مشروطاً بعدم معنوية الفرق بين الخطأين. حيث أنه عندما لا يرفض الفرض بأن الخطأين متساويين فإنه يفعل ذلك عند درجة احتمال معين أى أنه ليس رفضاً مطلقاً وهذا من شأنه أن يزيد من قيمة α عند استخدام الخطأ (بعد ضمهما) لاختبارات المعنوية فى التجربة. وطبقاً للخطأ الذى سيتقرر فإنه هو الذى يستخدم فى فصل المتوسطات ... الخ.

١٥–٩ تصميم المربع اللاتينى Latin Square Design

تم التطرق فى ١٥–٧ إلى تصميم القطاعات العشوائية حيث أن هناك اتجاها واحدا للاختلاف بين الوحدات التجريبية والذى تكون فيه القطاعات عمودية على هذا الاتجاه. وطبيعى أن يفكر المجرب أنه ربما يكون هناك أكثر من اتجاه لهذه الاختلافات. فمثلا إذا فرض أن هناك قطعة أرض ويراد إجراء التجربة عليها وأحد جوانب هذه الأرض محدود بقناة رى مما سينجم عنه اختلاف فى الخصوبة فى اتجاه معين. ويمكن أيضاً تصور أن الجانب العمودى على هذا الجانب ملاصق لصف من الأشجار العالية التى تؤثر على نمو النباتات بحجبها لضوء الشمس والعناصر الغذائية بدرجات متفاوتة تتوقف على بعد النبات عن الأشجار. كما فى شكل ١٥–٤.



شكل ١٥–٤ الاختلاف فى المادة التجريبية الناجم عن مؤثرين فى اتجاهين مختلفين.

جدول ١٥-١٠ متوسط المربعات المتوقعة واختبارات المعنوية عند تكرار الوحدة التجريبية فى القطاعات العشوائية

SOV	متوسط المربعات المتوقعة وطريقة الاختبار مبينة بالأسهم			
	كلاهما عشوائى	كلاهما ثابت	B ثابتة، t عشوائية	B عشوائية، t ثابتة
السلالات b	$\sigma_e^2+4\sigma_{bt}^2+12\sigma_b^2$	$\sigma_e^2+12K_b^2$	$\sigma_e^2+4\sigma_{bt}^2+12K_b^2$	$\sigma_e^2+12\sigma_b^2$
المعاملات t	$\sigma_e^2+4\sigma_{bt}^2+16\sigma_t^2$	$\sigma_e^2+16K_t^2$	$\sigma_e^2+16\sigma_t^2$	$\sigma_e^2+4\sigma_{bt}^2+16K_t^2$
الخطأ التجريبى	$\sigma_e^2+4\sigma_{bt}^2$	$\sigma_e^2+4K_{bt}^2$	$\sigma_e^2+4\sigma_{bt}^2$	$\sigma_e^2+4\sigma_{bt}^2$
الخطأ العينى	σ_e^2	σ_e^2	σ_e^2	σ_e^2

وعليه فلو أن المجرب استخدم التصميم العشوائى التام لكانت هناك فرصة أن تتوزع المعاملات بصورة غير متوازنة بالنسبة لمستويات الخصوبة الناجمة عن القناة أو عن الأشجار وإذا استخدم تصميم القطاعات العشوائية معتبراً الأشجار فقط فإن المعاملات تتوزع بصورة غير متزنة بالنسبة لعدم تجانس الخصوبة الناتج عن قناة الرى ... وهكذا. لذا فإن المجرب حريص على أن تمثل معاملاته بدرجة متوازنة بالنسبة لعدم تجانس الخصوبة الناتج عن وجود الأشجار وأيضاً الناتج عن قناة الرى. وتصميم المربع اللاتينى يوفر للمجرب هذا الشرط . وبصفة عامة فإن هذا التصميم يستخدم فى حالة:

١- وجود معاملة واحدة لها مستويين تجريبيين أو أكثر.

٢- لابد من عدم وجود تداخل بين الصفوف والأعمدة ومستويات المعاملة للمربع.

٣- تساوى عدد الصفوف والأعمدة مع عدد مستويات المعاملة. وهذا الاتزان من الصعب الوصول إليه فى المربعات التى تكون أكبر من ٨ x ٨ حيث تزداد المساحة كثيرا مما يهدد بزيادة عدم التجانس لعوامل أخرى غير تلك المعتبرة فى الدراسة.

٤- التوزيع العشوائى لمستويات المعاملة على الصفوف والأعمدة بشرط أن كل مستوى من مستويات المعاملة يظهر مرة واحدة فى الصف ومرة واحدة فى العمود.

١٥-٩-١ التعشية Randomization

إذا استخدم شكل ١٥-٤ للشرح فإنه يمكن تقسيم قطعة الأرض عمودياً على اتجاه عدم التجانس الناجم عن قناة الرى وبهذا تنشأ قطاعات سوف يصطلح تسميتها باسم الأعمدة columns، وإذا كونت قطاعات عمودية على اتجاه عدم التجانس الناجم عن الأشجار سينشأ قطاعات يطلق عليها صفوفاً rows. وشرط من شروط استخدام تصميم المربع اللاتينى، كما سبق ذكره، هو أن يكون عدد الأعمدة يساوى عدد الصفوف يساوى عدد مستويات المعاملة. يتم توزيع مستويات المعاملة على الأعمدة والصفوف بطريقة عشوائية بشرط أن كل مستوى يمثل مرة فى كل صف ومرة فى كل عمود. وسيرمز لمستويات المعاملة بالحروف اللاتينية. المربع القياسى standard square هو ذلك المربع الذى تترتب فيه المعاملات فى العمود الأول أو فى الصف الأول ترتيباً تنازلياً أو تصاعدياً. فإذا كان هناك معاملة ذات مستويين فإنه يوجد مربع قياسى واحد هو:

A	B
B	A

وفى حالة معاملة لها ثلاثة مستويات فإنه يوجد مربع قياسى واحد هو:

A	B	C
B	C	A
C	A	B

لاحظ أن كل معاملة ممثلة مرة فى كل عمود ومرة فى كل صف، وبتبادل مواقع كل من الأعمدة فيما بينها وكل من الصفوف فيما بينها في المربع ٣ × ٣ يمكن الحصول على ١٢ مربعاً هى المربعات الممكنة كما يلى:

A	B	C
B	C	A
C	A	B

A	C	B
B	A	C
C	B	A

B	C	A
C	A	B
A	B	C

B	A	C
C	B	A
A	C	B

C	B	A
A	C	B
B	A	C

C	A	B
A	B	C
B	C	A

A	B	C
C	A	B
B	C	A

A	C	B
C	B	A
B	A	C

B	C	A
A	B	C
C	A	B

B	A	C
A	C	B
C	B	A

C	B	A
B	A	C
A	C	B

C	A	B
B	C	A
A	B	C

لاحظ أن المربع الأول فقط من هذه المربعات هو القياسى أما باقى المربعات فجميعها غير قياسية. كما أنه فى كل منها مستويات المعاملة ممثلة مرة فى كل عمود ومرة فى كل صف.

أما فى حالة المربع ٤ × ٤ فإنه يوجد عدد أربعة مربعات قياسية ولكل منها عدد من المربعات الممكن تكوينها وهى $4!(4-1)! = 144$، وبالتالى فإن مجموع المربعات الممكن تكوينها هى ٥٧٦ مربعا منها ٤ قياسية هى:

مربع (١)

د	ج	ب	أ	
A	B	C	D	أ
B	D	A	C	ب
C	A	D	B	ج
D	C	B	A	د

مربع (٢)

د	ج	ب	أ	
A	B	C	D	أ
B	A	D	C	ب
C	D	B	A	ج
D	C	A	B	د

مربع (٣)

A	B	C	D	أ
B	A	D	C	ب
C	D	A	B	ج
D	C	B	A	د

مربع (٤)

A	B	C	D	أ
B	C	D	A	ب
C	D	A	B	ج
D	A	B	C	د

فى حالة المربع ٥ × ٥ فإن هناك ٥٦ مربعاً قياسياً وأن جميع المربعات الممكنة عددها ١٦١٢٨٠ مربعاً. ويعطى Fisher and Yates (1949) قوائم المربعات القياسية لكل من ٤ × ٤، ٥ × ٥ ومشتقاتها كما أعطى Kitagarwa and Miome (1953) جميع المربعات ٤ × ٤ الممكنة. ويقترح Fisher and Yates (1949) ما يلى عند إجراء التعشية للمربعات المختلفة:

١- فى المربع اللاتينى ٢ × ٢ رتب عشوائياً أعمدة المربع القياسى.

٢- فى المربع اللاتينى ٣ × ٣ وزع عشوائياً ترتيب الثلاث أعمدة فى المربع القياسى وكذلك الصفين الآخرين. أو اختر عشوائياً أحد المربعات الإثنى عشر.

٣- فى المربع اللاتينى ٤ × ٤ اختر أحد الأربع مربعات القياسية عشوائياً ثم وزع عشوائياً ترتيب الأعمدة وكذلك الثلاث صفوف الأخيرة. أو يختار عشوائياً أحد المربعات من الـ ٥٧٦ مربعاً. فمثلا إذا اختير احد المربعات القياسية عشوائياً وكان المربع (٢) توزع الأعمدة عشوائياً. وليكن هذا

الترتيب د أ ب ج ثم توزع الصفوف الثلاث الأخيرة عشوائياً كأن تصبح ج د ب كما يلى:

تعشية الأعمدة:

ج	ب	أ	د	
B	C	D	A	أ
A	D	C	B	ب
D	B	A	C	ج
C	A	B	D	د

وبتعشية الصفوف الثلاث الأخيرة يمكن الحصول على المربع العشوائى:

ج	ب	أ	د	
B	C	D	A	أ
D	B	A	C	ج
C	A	B	D	د
A	D	C	B	ب

ويلاحظ أيضاً أنه بعد التعشية هذه ما زال المربع اللاتينى يحتفظ بالخاصية الأساسية أن كل معاملة ممثلة مرة واحدة فى كل عمود وفى كل صف.

٤- فى المربع اللاتينى ٥ × ٥ يختار أحد الـــ ٥٦ مربعاً قياسياً ثم توزع الأعمدة عشوائيا وأيضاً الأربع صفوف الأخيرة.

ولإجراء التعشية لمربعات أكبر من ٥ × ٥ يحال القارئ إلى مرجع Federer (1963) السابق أو إلى Fisher and Yates (1949).

١٥-٩-٢ النموذج الإحصائى Statistical Model

بجانب المتوسط والخطأ التجريبى هناك ثلاث مؤثرات هى الصفوف والأعمدة ومستويات المعاملة. وإذا فرض أن عدد المعاملات = عدد الصفوف = عدد المعاملات = n فإنه يمكن توصيف النموذج الإحصائى كما يلى:

$$Y_{ijk} = \mu - r_i - c_j - t_k - e_{ijk}$$

حيث

Y_{ijk} : المشاهدة على الوحدة التجريبية فى العمود j والصف i والمعاملة k،

μ : المتوسط،

e_{ijk} : الخطأ لهذه المشاهدة،

$i = j = k = 1, 2, \cdots, n$ ، أى أن عدد الصفوف = عدد الأعمدة = عدد المعاملات $= n$

والافتراضات اللازمة لتحليل التباين وحساب المتوسطات هى أن

$$\sum r_i = \sum c_j = \sum t_k = \sum e_{ijk} = 0$$

ودرجات الحرية للتباين الكلى من المتوسط = $(n^2 - 1)$

ودرجات الحرية لكل من الأعمدة والصفوف والمعاملات = $(n - 1)$

ودرجات الحرية للخطأ = $(n - 1)(n - 2)$

مثال ١٥-٥

أراد مجرب فى أحد البلاد الحارة أن يختبر أربعة أنواع من الهوايات فى حظائر ماشية اللبن وكان لديه أربعة حظائر، والأبقار عنده مقسمة إلى تلك التى تحلب فى موسمها الأول أو الثانى أو الثالث أو الأكثر من ذلك. صمم المجرب تجربته باعتبار موسم الحليب هو الأعمدة والحظائر هى الصفوف وأنواع الهوايات هى المعاملات (أى كمربع لاتينى ٤×٤). وكان إنتاج اللبن بالكيلوجرام كما هو مبين فى جدول ١٥-١١.

ويمكن إجراء التحليل كما يلى:

مجموع المعاملات: $A = 17610$ ، $B = 15850$ ، $C = 16110$ ، $D = 14240$

متوسط المعاملات: $A = 4402.5$ ، $B = 3962.5$ ، $C = 4027.5$ ، $D = 3560$

مجموع المربعات الكلية عن المتوسط

$$= (1870)^2 + (4280)^2 + \cdots + (4490)^2 - \frac{(63810)^2}{16} = 15006043.7$$

جدول ١٥-١١ إنتاج اللبن بالكيلوجرام لأبقار مختلفة المواسم فى أربعة تصميمات من الحظائر بها أربعة أنواع من الهوايات.

	الموسم الأول	الموسم الثانى	الموسم الثالث	الموسم الرابع	المجموع	المتوسط
الحظيرة الأولى	1870 D	4280 C	3660 B	4840 A	**14650**	**3662.5**
الحظيرة الثانية	2150 C	4500 B	4880 A	4410 D	**15940**	**3985.0**
الحظيرة الثالثة	3200 B	4140 A	5010 D	5320 C	**17670**	**4417.5**
الحظيرة الرابعة	3750 A	2950 D	4360 C	4490 B	**15550**	**3887.5**
المجموع	**10970**	**15870**	**17910**	**19060**	**63810**	
المتوسط	**2742.5**	**3967.5**	**4477.5**	**4765**		**3988.1**

مجموع المربعات بين الأعمدة

$$= \frac{(10970)^2 + \cdots + (19060)^2}{4} - \frac{(63810)^2}{16} = 9580118.7$$

مجموع المربعات بين الصفوف

$$= \frac{(14650)^2 + \cdots + (15550)^2}{4} - \frac{(63810)^2}{16} = 1202118.7$$

مجموع المربعات بين المعاملات

$$= \frac{(17610)^2 + \cdots + (14240)^2}{4} - \frac{(63810)^2}{16} = 1428818.7$$

مجموع مربعات الخطأ

$$= 15006043.7 - 9580118.7 - 1202118.7 - 1428818.7 = 2794987.69$$

ويمكن تلخيص النتائج فى جدول ١٥-١٢ لتحليل التباين

جدول ١٥-١٢ تحليل التباين لتجربة تأثير نوع الهوايات على إنتاج الأبقار والمصممة كمربع لاتينى ٤ x ٤

SOV	df	SS	MS
الأعمدة (موسم الحليب)	3	9580118.7	3193372.9
الصفوف (الحظائر)	3	1202118.7	400706.2
المعاملات (الهوايات)	3	1428818.7	476272.9
الخطأ	6	2794987.6	465831.3
الكلى عن المتوسط	**15**	**15006043.7**	

ويمكن اختبار فرض العدم أن مواسم الحليب متساوية أى عمود ١ = عمود ٢ = عمود ٣ = عمود ٤ بواسطة:

$$F = 3193372.9 / 465831.3 = 6.86$$

وهى معنوية ويرفض الفرض.

واختبار فرض العدم أن الحظائر كلها متساوية أى صف ١ = صف ٢ = صف ٣ = صف ٤ بواسطة:

$$F = 400706.2 / 465831.3 = 0.86$$

وهى غير معنوية ولا يرفض فرض العدم.

كما يمكن اختبار فرض العدم أن المعاملات متساوية بواسطة:

$$F = 476272.9 / 465831.3 = 1.02$$

وهى غير معنوية، وعليه لا يرفض فرض العدم أى أنه لا يوجد ما يدل على أن الهوايات مختلفة عن بعضها.

كما يمكن إجراء فصل متوسطات مواسم الحليب المعنوية بالطرق السابق شرحها حيث إن:

الخطأ القياسى لمتوسط موسم الحليب عبارة عن

$$S_{\overline{Y}_{j.}} = \sqrt{\frac{465831.3}{4}} = 341.3$$

والخطأ القياسى للفرق بين متوسطين هو

$$S_d = \sqrt{\frac{2(465831.3)}{4}} = 482.6$$

ويمكن بسهولة كتابة المعادلات الاعتيادية للتجربة السابقة وحل تلك المعادلات وذلك بوضع افتراضات إضافية مثل $\sum c_i = \sum r_j = \sum t_k = 0$.

وبالتالى يمكن الحصول عل:

$$\hat{\mu} = \frac{63810}{16} = 3988.125 \text{ kg}$$

أثر العمود الأول $$\hat{c}_1 = \frac{10970}{4} - 3988.125 = -1245.625 \text{ kg}$$

أثر العمود الثانى $$\hat{c}_2 = \frac{15870}{4} - 3988.125 = -20.625 \text{ kg}$$

أثر العمود الثالث $$\hat{c}_3 = \frac{17910}{4} - 3988.125 = +489.375 \text{ kg}$$

أثر العامود الرابع $$\hat{c}_4 = \frac{19060}{4} - 3988.125 = +776.875 \text{ kg}$$

أثر الصف الأول $$\hat{r}_1 = \frac{14650}{4} - 3988.125 = -325.625 \text{ kg}$$

أثر الصف الثاني $$\hat{r}_2 = \frac{15940}{4} - 3988.125 = -3.125 \text{ kg}$$

أثر الصف الثالث $$\hat{r}_3 = \frac{17670}{4} - 3988.125 = +429.375 \text{ kg}$$

أثر الصف الرابع $$\hat{r}_4 = \frac{15550}{4} - 3988.125 = -100.625 \text{ kg}$$

أثر المعاملة الأولى $$\hat{t}_1 = \frac{17610}{4} - 3988.125 = +414.375 \text{ kg}$$

أثر المعاملة الثانية $\hat{t}_2 = \frac{15850}{4} - 3988.125 = -25.625$ kg

أثر المعاملة الثالثة $\hat{t}_3 = \frac{16110}{4} - 3988.125 = +39.375$ kg

أثر المعاملة الرابعة $\hat{t}_4 = \frac{14240}{4} - 3988.125 = -428.125$ kg

استخدام اختيار PROC GLM لحل مثال ١٥-٥ والتفرقة بين المتوسطات.

```
DATA LATIN;
INPUT SEASON BARN FAN $ MILK @@;
CARDS;
1 1 D 1870 1 2 C 2150 1 3 B 3200 1 4 A 3750
2 1 C 4280 2 2 B 4500 2 3 A 4140 2 4 D 2950
3 1 B 3660 3 2 A 4880 3 3 D 5010 3 4 C 4360
4 1 A 4840 4 2 D 4410 4 3 C 5320 4 4 B 4490
PROC GLM;
CLASS SEASON BARN  FAN;
MODEL MILK = SEASON BARN FAN/SS3;
MEANS SEASON BARN FAN/DUNCAN;
RUN;
```

نتائج التحليل:

The GLM Procedure

Class	Levels	Values
SEASON	4	1 2 3 4
BARN	4	1 2 3 4
FAN	4	A B C D

Number of observations 16

Dependent Variable: MILK

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	12211056.25	1356784.03	2.91	0.1031
Error	6	2794987.50	465831.25		
C-Total	15	15006043.75			

R-Square	Coeff Var	Root MSE	MILK Mean
0.813743	17.11376	682.5183	3988.125

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SEASON	3	9580118.750	3193372.917	6.86	0.0230
BARN	3	1202118.750	400706.250	0.86	0.5109
FAN	3	1428818.750	476272.917	1.02	0.4464

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	6
Error Mean Square	465831.3

Number of Means	2	3	4
Critical Range	1181	1224	1245

Means with the same letter are not significantly different.

Duncan Grouping	Mean	N	SEASON
A	4765.0	4	4
A	4477.5	4	3
A	3967.5	4	2
B	2742.5	4	1

Duncan's Multiple Range Test for MILK

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	6
Error Mean Square	465831.3

Number of Means	2	3	4
Critical Range	1181	1224	1245

Means with the same letter are not significantly different.

Duncan Grouping	Mean	N	BARN
A	4417.5	4	3
A	3985.0	4	2
A	3887.5	4	4
A	3662.5	4	1

Duncan's Multiple Range Test for MILK

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	6
Error Mean Square	465831.3

Number of Means	2	3	4
Critical Range	1181	1224	1245

Means with the same letter are not significantly different.

Duncan Grouping	Mean	N	FAN
A	4402.5	4	A
A	4027.5	4	C
A	3962.5	4	B
A	3560.0	4	D

١٥-٩-٣ كفاءة المربع اللاتينى

كفاءة المربع اللاتينى بالنسبة للتصميم تام العشوائية

$$= \frac{M_r - M_c - (n-1)M_e}{(n+1)M_e}$$

حيث M_r، M_c، M_e، n هى متوسط المربعات لكل من الصفوف والأعمدة والخطأ وعدد الصفوف، على التوالى. وبالتطبيق على البيانات المذكورة فى مثال ١٥-٥ فإن هذه النسبة تساوى:

$$= \frac{400706.2 - 3193372.9 - (4-1)(465831.3)}{(4+1)(465831.3)} = 2.14$$

أى أن تصميم المربع اللاتينى أكفأ من التام العشوائية بنسبة 214% .

بينما كفاءة تصميم المربع اللاتينى بالنسبة لتصميم القطاعات العشوائية مع اعتبار أن الأعمدة هى القطاعات تقاس بالمعادلة:

$$= \frac{M_c - (n-1)M_e}{nM_e}$$

بينما إذا اعتبر أن الصفوف هى القطاعات فإن المعادلة تصبح:

$$= \frac{M_r - (n-1)M_e}{nM_e}$$

وبالتطبيق على البيانات المذكورة فى مثال ١٥-٥ فإن النسبة الأولى تساوى:

$$= \frac{3193372.9 - (4-1)(465831.3)}{(4)(465831.3)} = 2.46$$

والمعادلة الثانية تساوى:

$$= \frac{400706.2 - (4-1)(465831.3)}{(4)(465831.3)} = 0.97$$

وذلك الفرق الكبير بين النسبتين راجع إلى الاختلافات الواسعة بين الأعمدة أى مواسم الحليب كما هو واضح من جدول تحليل التباين ١٥-١٢.

١٥-٩-٤ تقدير القيم الغائبة

تقدر القيمة الغائبة $\hat{Y}_m$ باستخدام المعادلة

$$\hat{Y}_m = \frac{n(C_m - R_m - T_m) - 2Y...}{(n-1)(n-2)}$$

حيث C_m ، R_m ، T_m هى مجموع كل من العمود والصف والمعاملة التى تقع بها القيمة الغائبة، $Y...$ المجموع الكلى ، n عدد الصفوف.

ففي البيانات المذكورة فى مثال ١٥-٥ إذا فرض أن المعاملة D فى العمود الرابع والصف الثانى غائبة (قيمتها 4410) يصبح

$$Y... = 59400 \text{ ، } T_m = 9830 \text{ ، } R_m = 11530 \text{ ، } C_m = 14650$$

وتكون القيمة المقدرة لها

$$\hat{Y}_m = \frac{4\,(14650 - 11530 - 9830) - (2)\,(59400)}{(4-1)\,(4-2)} = 4206.7 \text{ kg}$$

توضع هذه القيمة مكان القيمة الغائبة ويعاد حساب تحليل التباين مع خفض درجات حرية الخطأ بدرجة واحدة إذا كانت درجات الحرية تسمح بذلك. ويكون تباين الفرق بين متوسطين أحدهما المعاملة ذات القيمة الغائبة:

$$= \sqrt{M_C\left[\frac{2}{n} + \frac{1}{(n-1)\,(n-2)}\right]}$$

وكما سبق القول فى حالة تصميم القطاعات العشوائية فإنه إذا فقدت أكثر من قيمة (قيمتان مثلا) فإنهما يقدران أولا بقيم وسطى ثم تثبت إحداها وتقدر الأخرى ثم تثبت هذه وتقدر الأولى وتعاد الكرة حتى يثبت تقديران متتاليان فتكون القيمة هى التقدير. ويجرى التحليل عاديا وتنقص درجات حرية الخطأ بعدد المشاهدات الغائبة إذا كان يسمح بذلك، فمثلا عندما يكون المربع ٣ × ٣ فإن درجات حرية الخطأ 2 وبالتالى إذا فقد قيمتان فلن تبقى هناك درجات حرية للخطأ.

١٠-٩-٥ تعدد المشاهدات فى تصميم المربع اللاتينى

فى مثال البيانات المذكورة فى ١٥-٥ كانت كل معاملة فى صف أو فى عمود ممثلة ببقرة (أو وحدة تجريبية) واحدة ولكن فى بعض الأحيان يضع المجرب وحدتين تجريبيتين بدلاً من واحدة وذلك لزيادة قوة اختبار الفرق بين المتوسطات. وكما اتضح فى تصميم القطاعات العشوائية فإن تكرار الوحدة التجريبية يجعل للتجربة خطأين أحدهما تجريبي experimental error والآخر عينى sampling error.

مثال ١٥-٦

أورد Kirk (1995) مثالاً به أربعة أصناف من إطارات السيارات وكان الهدف هو اختبار الفرق بينها. والمتغير هنا هو سمك قشرة الإطار بعد استخدامه لمسافة 16000 كيلومتر. وحيث أن استهلاك الإطار ممكن أن يتأثر بنوع العربة وأيضاً بموضع الإطار فيجب ضمان أن كل نوع إطار مجرب مع كل عربة وفى كل موقع، أى يمين أمامى وخلفى ويسار أمامى وخلفى. صممت التجربة على هيئة مربع لاتينى ٤ × ٤ وبكل خلية مشاهدتين كالتالى:

		موقع الإطار فى السيارة			
		أمامى يمين	أمامى يسار	خلفى يمين	خلفى يسار
		b_4	b_3	b_2	b_1
نوع السيارة	a_1	10 ،7 b_4	7 ،5 b_3	4 ،2 b_2	1 ،3 b_1
	a_2	2 ،6 b_1	10 ،8 b_4	8 ،6 b_3	3 ،5 b_2
	a_3	4 ،4 b_2	3 ،2 b_1	9 ،9 b_4	5 ،7 b_3
	a_4	6 ،6 b_3	3 ،3 b_2	3 ،2 b_1	11 ،8 b_4

ويمكن تلخيص البيانات كما يلى:

نوع السيارة	موقع الإطار فى السيارة								المجموع
	b_4		b_3		b_2		b_1		
a_1	17	t_4	12	t_3	6	t_2	4	t_1	**39**
a_2	8	t_1	18	t_4	14	t_3	8	t_2	**48**
a_3	8	t_2	5	t_1	18	t_4	12	t_3	**43**
a_4	12	t_3	6	t_2	5	t_1	19	t_4	**72**
المجموع	**45**		**41**		**43**		**43**		**172**

مجموع المعاملات: $t_1 = 22$ ، $t_2 = 28$ ، $t_3 = 50$ ، $t_4 = 72$.

مجموع المربعات الكلية عن المتوسط

$$= 3^2 + 1^2 + \cdots + 6^2 + 6^2 - \frac{(172)^2}{32} = 235.5$$

مجموع المربعات بين الأعمدة (مواقع الإطارات)

$$= \frac{(43)^2 + \cdots + (45)^2}{8} - \frac{(172)^2}{32} = 1$$

مجموع المربعات بين الصفوف (العربات)

$$= \frac{(39)^2 + \cdots + (42)^2}{8} - \frac{(172)^2}{32} = 5.25$$

مجموع المربعات بين المعاملات (أنواع الإطارات)

$$= \frac{(22)^2 + \cdots + (72)^2}{8} - \frac{(172)^2}{32} = 194.5$$

مجموع مربعات الخطأ التجريبى

$$=\frac{4^2+8^2+\cdots+8^2+12^2}{2}-\frac{(172)^2}{32}-1.0-5.25-194.5=2.75$$

مجموع مربعات الخطأ العينى $= 235.5 - 1 - 5.25 - 194.5 - 2.75 = 32$

وهو فى نفس الوقت = بين الوحدات التجريبية داخل الخلايا

$$=(3^2+1^2-\frac{4^2}{2})+(5^2+3^2-\frac{8^2}{2})+\cdots+(6^2+6^2-\frac{12^2}{2})$$

أى أنه يمكن حساب الخطأ العينى مباشرة وليس بالطرح.

ويمكن تلخيص النتائج فى جدول تحليل التباين التالى:

SOV	df	SS	MS
الأعمدة (مواقع الإطارات)	$(n-1)=3$	1.0	0.333
الصفوف (العربات)	$(n-1)=3$	5.25	1.750
المعاملات (أنواع الإطارات)	$(n-1)=3$	194.5	64.833
الخطأ التجريبى	$(n-1)(n-2)=6$	2.75	0.458
الخطأ العينى	$n^2(2-1)=16$	32.0	2.0
الكلى عن المتوسط	$2n^2-1=31$	**235.5**	

وفى حالة افتراض أن العوامل كلها ثابتة فإنه يمكن اختبار فروض العدم التالية:

١- عدم وجود فروق بين مواقع الإطارات بواسطة $F=\frac{0.333}{0.458}=0.73$ وهى غير معنوية فلا يرفض فرض العدم.

٢- عدم وجود فروق بين العربات بواسطة $F=\frac{1.750}{0.458}=3.82$ وهى أيضاً غير معنوية.

٣- عدم وجود فروق بين أنواع الإطارات يمكن اختباره بواسطة $F=\frac{64.833}{0.458}=141.56$ وهى معنوية جداً، فيرفض فرض العدم هذا.

وإذا كان التفاعل بين كل من الأعمدة والصفوف والأعمدة والمعاملات والصفوف والمعاملات غير موجود فإن الخطأ التجريبي والخطأ العينى هما تقديران لنفس الشئ أى لتباين الخطأ σ_e^2 وأيضاً

$$F = \frac{\text{متوسط مربعات الخطأ التجريبى}}{\text{متوسط مربعات الخطأ العينى}}$$

هى اختبار جزئى للتجمعية addivity ولكن يمكن أيضاً استخدام Tukey's test لاختبار التجمعية بصورة أكمل ويرجع القارئ فى هذا إلى Kirk (1968).

ويمكن ضم كلا من الخطأين فى واحد إذا لم يكونا مختلفين عن بعضهما معنوياً لتعضيد درجات الحرية بغية زيادة قوة الاختبار وهذا الإجراء خاضع لنفس الضوابط التى نوقشت فى فصل ١٥-٨.

أما فصل المتوسطات فيتبع نفس الإجراءات السابقة.

١٥-٩-٦ تصميم المربع اللاتينى اليونانى Graceo-Latin Square Design

تتم التعشية فى المربع اللاتينى بطريقة تضمن أن كل العوامل الثلاث (العمود والصف والمعاملة) تتكرر مع بعضها الآخر بدرجة متساوية. ولكن ممكن تخيل أن هناك أربعة عوامل يراد تعشيتها. ففى البيانات المذكورة فى مثال ١٥-٤ حيث كان يوجد ٤ حظائر، ٤ مواسم حليب، ٤ معاملات. وبافتراض أنه يوجد ٤ سلالات ويراد إجراء التعشية بحيث أن كل معاملة تتكرر مرة فى كل حظيرة ولكل موسم حليب وفى كل سلالة. فإذا رمز لمواسم الحليب بالأعمدة والحظائر بالصفوف والمعاملات بالحروف اللاتينية A، B، C، D كما سبق، ورمز للسلالات الأربع بالحروف اليونانية α، β، γ، δ فإنه يمكن وضعها جميعاً فى تصميم لاتينى يونانى كما يلى:

أعمدة

A α	B β	C γ	D δ	
B δ	A γ	D β	C α	صفوف
C β	D α	A δ	B γ	
D γ	C δ	B α	A β	

وواضح من هذا الترتيب أن α ممثلة مرة فى كل عمود ومرة فى كل صف ومرة مع كل معاملة (أى A، B، C، D) وكذلك كل من β, γ, δ. وكذلك A ممثلة مرة فى كل عمود ومرة فى كل صف ومرة فى كل حرف يونانى (أى سلالة) ... وهكذا. وجدير بالذكر أنه ليس لكل عدد من المعاملات مربع لاتينى يونانى.

١٥-٩-٧ تصميمات أخرى تتبع عائلة تصميم المربع اللاتينى

من أهم ما يتميز به تصميم المربع اللاتينى أنه يضبط الاختلافات بين المادة التجريبية فى اتجاهين فى وقت واحد وعليه يمكن إزالة جزء من التباين والذى بدون هذا التصميم سيتبقى فى الخطأ وبالتالى قد يقلل من قوة الاختبار فى التجربة. ولكن لاحظ أن تصميم المربع اللاتيني محدد بأن يكون عدد المعاملات = عدد الأعمدة = عدد الصفوف، وهذا قيد أيضاً فإذا قل عدد المعاملات عن ٤ ستنخفض درجات الحرية للخطأ بدرجة ملحوظة وهذا يضعف من قوة الاختبار. فمثلا فى حالة وجود ٣ معاملات فإن درجات الحرية للخطأ هى 2 بينما إذا كان عدد المعاملات ٢ فإنه لا يوجد درجات حرية للخطأ. بينما إذا زاد عدد المعاملات كثيراً فإنه لابد من زيادة عدد الوحدات التجريبية مما قد يخلق عدم تجانس فى المادة التجريبية هذه، وهذا أمر غير مرغوب فيه. لذا فإن تصميم المربع اللاتينى بالطريقة التى وصف بها حتى الآن لا ينصح بإتباعه إلا إذا كان عدد المعاملات يتراوح بين ٤، ١٠. ماذا إذن فى حالة وجود عدد قليل من المعاملات؟

مثال ١٥-٧

هناك اختباران أ، ب يراد مقارنة الدرجات التى تحصل عليها التلاميذ فى كل منهما. وهناك فترتان وكل تلميذ سيختبر فى الفترتين.

يمكن تخيل تصميم المربع اللاتينى ٢ × ٢ التالى:

	التلميذ (١)	التلميذ (٢)
الفترة الأولى	أ	ب
الفترة الثانية	ب	أ

وواضح أن فى مثل هذا التصميم الإحصائى البسيط لن يتبقى من درجات الحرية شئ لاختبارات المعنوية. ويمكن للمجرب أن يكرر مثل هذا المربع عدة مرات. وقد

تكون كل المربعات من نفس المجموعة المتجانسة من التلاميذ أو قد يكون كل مربع ممثلا لمتغير آخر (مثل العمر) ويجرى تقسيم المادة التجريبية عليه وهذا الأخير هو خليط بين تصميم القطاعات العشوائية والمربع اللاتينى. أى أنه يتم تقسيم المادة التجريبية عمودياً على اتجاه العمر مثلا ثم داخل كل قطاع بدلاً من التوزيع عشوائياً فإنها توزع لتكون مربعاً لاتينياً. والتصميم الأول يطلق عليه تصميم العبور أو تصميم التغير change over design أو cross over design. بينما الترتيب الثانى يطلق عليه تصميم مجموعة مربعات لاتينية. و الجدول التالى يمثل الدرجات التى تحصل عليها التلاميذ فى الاختبارين.

التلاميذ

	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠	١١	١٢	مجموع
الفترة الأولى	أ 66	ب 46	أ 48	ب 20	أ 56	ب 28	أ 62	ب 16	أ 16	ب 34	أ 36	ب 52	480
الفترة الثانية	ب 56	أ 42	ب 40	أ 22	ب 24	أ 54	ب 40	أ 22	ب 26	أ 28	ب 26	أ 52	432
مجموع	122	88	88	42	80	82	102	38	42	62	62	104	912

مجموع الاختبار أ، ب علي التوالى 504، 408

ولو فرض أن كل التلاميذ متجانسون من حيث ما يمكن أن يؤثر على النتيجة فيكون التحليل كما يلى:

مجموع المربعات الكلية عن المتوسط

$$= (66)^2 + (56)^2 + \cdots + (52)^2 + (104)^2 - \frac{(912)^2}{24} = 5352$$

مجموع المربعات بين التلاميذ

$$= \frac{(122)^2 + (88)^2 + \cdots + (104)^2}{2} - \frac{(912)^2}{24} = 4032$$

مجموع المربعات الراجعة لترتيب أخذ الاختبار

$$= \frac{(480)^2 + (432)^2}{12} - \frac{(912)^2}{24} = 96$$

مجموع المربعات الراجعة إلى الاختبار

$$= \frac{(504)^2 + (408)^2}{12} - \frac{(912)^2}{24} = 384$$

مجموع مربعات الخطأ $= 5352 - 4032 - 96 - 384 = 840$

ويمكن تلخيص النتائج فى جدول تحليل التباين التالى:

SOV	df	SS	MS
بين التلاميذ	11	4032	366.5
بين الترتيبين	1	96	96.0
بين الاختبارين	1	384	384.0
الخطأ	10	840	84.0
الكلى عن المتوسط	23	5352	

وفى حالة فرض أن تأثير الترتيبين والاختبارين ثابتان فإن اختبار فرض العدم لكل منهما هو $F = \frac{384}{84} = 4.57$ ، $F = \frac{96}{84} = 1.14$ على التوالى وكلاهما غير معنوى عند مستوى 5%.

وبهذه الطريقة أمكن توفير عدد كاف من درجات الحرية للخطأ لإجراء اختبار معنوية ذى قوة مناسبة وقد أطلق على هذا التصميم العبور أو Cross-over لأنه إذا نظر إلى مثال ١٥-٧ وتم توصيل أ فى التلميذ الأول مع أ فى التلميذ الثانى وكذلك ب للتميذ الأول مع ب للتلميذ الثانى فإن الخطين يتقاطعان.

مثال ١٥-٨

حل مثال ١٥-٧ باستخدام برنامج SAS

```
DATA COVLATIN;
INPUT STUDENT PERIOD TEST $ GRAD @@;
CARDS;
1 1 A 66 2 1 B 46 3 1 A 48 4 1 B 20
5 1 A 56 6 1 B 28 7 1 A 62 8 1 B 16
9 1 A 16 10 1 B 34 11 1 A 36 12 1 B 52
1 2 B 52 2 2 A 42 3 2 B 40 4 2 A 22
5 2 B 24 6 2 A 54 7 2 B 40 8 2 A 22
9 2 B 26 10 2 A 28 11 2 B 26 12 2 A 52
PROC GLM;
CLASS STUDENT PERIOD TEST;
MODEL GRAD = STUDENT PERIOD TEST/SS3;
RUN;
```

نتائج التحليل:

The GLM Procedure
Class Level Information

Class	Levels	Values
student	12	1 2 3 4 5 6 7 8 9 10 11 12
period	2	1 2
test	2	a b

Number of observations 24

Dependent Variable: grad

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	4512.000000	347.076923	4.13	0.0153
Error	10	840.000000	84.000000		
Corrected Total	23	5352.000000			

R-Square	Coeff Var	Root MSE	grad Mean
0.843049	24.11882	9.165151	38.00000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
student	11	4032.000000	366.545455	4.36	0.0138
period	1	96.000000	96.000000	1.14	0.3102
test	1	384.000000	384.000000	4.57	0.0582

والآن افترض أنه فى المثال السابق أن مجموعة التلاميذ لم تكن متجانسة فى لعمر الشئ الذى قد يؤثر على نتيجة الاختبار. وإذا فرض أن هناك مقدرة من الاستفادة من الاختبار الأول عند حل الاختبار الثانى وأن هذه المقدرة تتأثر بعمر التلميذ وجب إذن أن تصمم التجربة كمجموعة من المربعات اللاتينية. افترض الآن أن المجرب حصل على 12 تلميذاً قسمها إلى 6 مجاميع عمرية كل منها يمثل مربع لاتينى كما يلى:

	(١)	(١٢)	مجموع
	أ	ب	
	66	52	118
	ب	أ	
	56	52	108
مجموع الأعمدة	122	104	
مجموع المربعات	226		

	(٢)	(٧)	مجموع
	أ	ب	
	42	40	82
	ب	أ	
	46	62	108
مجموع الأعمدة	88	102	
مجموع المربعات	190		

	(٣)	(٦)	مجموع
	أ	ب	
	48	28	76
	ب	أ	
	40	54	94
مجموع الأعمدة	48	82	
مجموع المربعات	130		

	(٥)	(١٠)	مجموع
	أ	ب	
	56	34	90
	ب	أ	
	24	28	52
مجموع الأعمدة	80	62	
مجموع المربعات	142		

	(٤)	(١١)	مجموع
	أ	ب	
	22	26	48
	ب	أ	
	20	36	56
مجموع الأعمدة	42	62	
مجموع المربعات	104		

	(٨)	(٩)	مجموع
	أ	ب	
	22	26	48
	ب	أ	
	16	16	32
مجموع الأعمدة	38	42	
مجموع المربعات	80		

وعلى أساس هذه الافتراضات والتصميم الموضوع يكون التحليل كما يأتى:

مجموع المربعات الكلى عن المتوسط = ذلك للتصميم السابق = 5352

مجموع المربعات بين المربعات

$$=\frac{(226)^2+\cdots+(80)^2}{4}-\frac{(912)^2}{24}=3708$$

مجموع المربعات بين الأعمدة (التلاميذ) داخل المربعات

$$= \left[\frac{(122)^2 + (104)^2}{2} - \frac{(226)^2}{4}\right] + \cdots + \left[\frac{(42)^2 + (38)^2}{2} - \frac{(80)^2}{4}\right] = 324$$

مجموع المربعات بين الصفوف (الفترات) داخل المربعات

$$= \left[\frac{(118)^2 + (108)^2}{2} - \frac{(226)^2}{4}\right] + \cdots + \left[\frac{(32)^2 + (48)^2}{2} - \frac{(80)^2}{4}\right] = 716$$

مجموع المربعات بين الاختبارين = ذلك للتصميم السابق = 384

مجموع مربعات الخطأ $= 5352 - 3708 - 324 - 716 - 384 - 220 = 220$

ويمكن تلخيص النتائج فى جدول تحليل التباين التالى:

SOV	df	SS	MS
بين المربعات	5	3708	741.6**
بين التلاميذ داخل المربعات	6(2-1) = 6	324	54
بين الفترات داخل المربعات	6(2-1) = 6	716	119.3
الاختبار	1	384	384*
الخطأ	5	220	44
الكلى عن المتوسط	23	5352	

وإذا فرض أن العوامل فى النموذج الإحصائى ثابتة فإنه يمكن اختبار فروض العدم التالية:

١- عدم وجود فروق بين الأعمار (المربعات) بواسطة: $F = \dfrac{741.6}{44} = 16.8$
وهى معنوية باحتمال 0.01 ويرفض فرض العدم.

٢- عدم وجود فرق بين الاختبارين بواسطة: $F = \dfrac{384}{44} = 8.7$ وهى معنوية باحتمال 0.05 ويرفض فرض العدم.

مثال ١٥-٩

حل مثال ١٥-٨ باستخدام برنامج SAS

```
DATA COVLATIN;
INPUT SQUARE STUDENT PERIOD TEST $ GRAD @@;
CARDS;
1 1 1 A 66 2 2 1 B 46 3 3 1 A 48 5 4 1 B 20 4 5 1 A 56 3 6 1 B 28
2 7 1 A 62 6 8 1 B 16 6 9 1 A 16 4 10 1 B 34 5 11 1 A 36 1 12 1 B 52
1 2 2 B 56 2 2 2 A 42 3 3 2 B 40 5 4 2 A 22 4 5 2 B 24 3 6 2 A 54
2 7 2 B 40 6 8 2 A 22 6 9 2 B 26 4 10 2 A 28 5 11 2 B 26 1 12 2 A 52
PROC GLM;
CALSS SQUARE STUDENT PERIOD TEST;
MODEL GRAD = SQUARE STUDENT(SQUARE)
             PERIOD(SQUARE) TEST/SS3;
RUN;
```

نتائج التحليل:

```
                 The GLM Procedure
              Class Level Information
             Class      Levels   Values

        SQUARE        6     1 2 3 4 5 6
        STUDENT      12     1 2 3 4 5 6 7 8 9 10 11 12
        PERIOD        2     1 2
        TEST          2     A B

         Number of observations    24

Dependent Variable: GRAD
                           Sum of
Source            DF      Squares     Mean Square   F Value   Pr > F
Model             18   5132.000000  285.111111      6.48     0.0240
Error              5    220.000000   44.000000
Corrected Total   23   5352.000000
        R-Square    Coeff Var    Root MSE    GRAD Mean
        0.958894    17.45592     6.633250    38.00000

Source               DF Type III SS  Mean Square  F Value Pr > F
SQUARE                5  3708.000000 741.600000  16.85   0.0038
STUDENT(SQUARE)       6  324.000000   54.000000   1.23   0.4203
PERIOD(SQUARE)        6  716.000000  119.333333   2.71   0.1466

TEST                  1  384.000000  384.000000   8.73   0.0317
```


salamalhelali@yahoo.com مع تحيات د. سلام حسين الهلالي

من الواضح أنه أمكن التغلب على نقص درجات الحرية للخطأ فى حالة العدد القليل من المعاملات فى تصميم المربع اللاتينى وذلك بإتباع إما التصميم العبورى أو تصميم مجاميع المربعات اللاتينية وكل له اشتراطاته. ومن الواضح أيضا أنه قد تم تجاوز الحد فى تحليل نفس البيانات طبقاً لتصميمين إحصائيين مختلفين وذلك بسبب أن التحليل يجب أن يتبع التصميم الواقعى فقط دون التلاعب به بعد الحصول على البيانات، ولكن كان المقصود هنا معالجة البيانات بطريقتين مختلفتين بافتراض أن كلا من التحليلين كان هو المقصود فعلا. وكثيراً ما يستخدم التصميم العبورى أو تصميم مجاميع المربعات اللاتينية فى تجارب الإنتاج الحيوانى خاصة فى ماشية اللبن، حيث يقسم منحنى الحليب إلى فترات متساوية بعدد المعاملات ثم تكون مجاميع من الأبقار كل منها بعدد المعاملات. كلما كانت منحنيات الحليب جميعها متوازنة فى الأبقار المختلفة (أى متساوية فى المثابرة) أمكن استخدام التصميم العبورى ولكن إذا اختلفت أشكال منحنيات الحليب فى الحيوانات المختلفة تقسم الحيوانات إلى مجاميع كل منها متشابهة فى شكل منحنيات الحليب ويكون عدد كل مجموع مساوى لعدد المعاملات أو مضاعفاتها وتكون منها مربعات لاتينية. أما كون منحنى الحليب لنفس الحيوان يقسم إلى عدة أقسام (وحدات تجريبية) بعدد المعاملات فإن هذا يستوجب أن يكون هناك فاصلا زمنياً مناسباً بين كل معاملتين حتى لا تؤثر المعاملة السالفة فى المعاملة الحالية، وهذه الفترة يحددها المجرب طبقاً للمعلومات البيولوجية، ولكن إذا شك المجرب فى أن المعاملة السالفة لن يزول أثرها تماماً فهناك من الأساليب والتصميمات الإحصائية التى تسمح بتقدير الأثر المتبقى residual effect من المعاملة السالفة على المعاملة الحالية وإزالته إحصائياً (Cochran and Cox, 1950).

١٥-١٠ تصميمات القطاعات المنشقة Split Plot Designs

فى كثير من التجارب التى يستخدم فيها تصميم القطاعات العشوائية الكاملة قد يكون من غير الممكن توزيع كل توليفات العوامل محل الدراسة داخل قطاع واحد. وهذا قد يحدث عندما تحتاج بعض هذه العوامل عدد كبير من الوحدات التجريبية لتقييمها بينما يوجد عوامل أخرى تحتاج إلى عدد أقل من الوحدات التجريبية. كما أنه قد تتفاوت فى بعض الأحيان درجة الدقة التى يود أن يوليها المجرب لدراسة ومقارنة العوامل المختلفة. ويقصد باصطلاح القطاعات المنشقة أن التجربة عبارة عن تجربه عاملية factorial experiment مع وجود مزج confounded بين العامل الرئيسى main effect والقطاعات blocks ولكن مع وجود عدد كاف من المكررات replicates لكى يمكن تقدير هذه العوامل. فإذا فرض أن مجرباً يود اختبار ٤ هجن من دجاج اللحم. تقييم الهجن هذا يحتاج إلى أعداد كبيرة من الطيور، وليكن 600 طائر من كل من الهجن. وإذا كان لدى المجرب ٥ حظائر فإنه يمكنه أن يصمم

التجربة مبدئيا كقطاعات عشوائية حيث تمثل كل حظيرة بقطاع ويقسم كل قطاع عشوائياً إلى ٤ أجزاء بكل جزء منها عدد 150 طائرا، كل جزء يخص أحد الهجن. فإذا تم إجراء التوزيع عشوائيا ومستقلا لكل حظيرة (أى القطاع) وكانت الوحدة التجريبية هنا هى 150 طائر يصبح تحليل التباين كما يلى:

SOV	Df
القطاعات (الحظائر)	4
الهجن	3
الخطأ	12
الكلى	**19**

ويمكن استغلال الإمكانيات التجريبية إلى أكفأ حد فقد يود المجرب تجربة بعض علائق التسمين وعددها 3 على هذه الطيور. وبالتالي يقوم بتقسيم طيور كل هجين فى كل حظيرة إلى ثلاثة تحت أقسام sub-plot متساوية كل منها 50 طائر ويتم توزيع كل من هذه العلائق عليها عشوائياً. وتجرى هذه التعشية مستقلة تماماً فى كل الهجن والحظائر، ويمكن التعبير عن هذا التصميم التجريبى بالشكل التالى حيث القطاعات تمثل الحظائر (٥ قطاعات) والأقسام الرئيسية تمثل الهجن ويرمز لها بالحروف أ، ب، ج، د وتحت ٣ أقسام تمثل العلائق ويرمز لها بالأرقام ١، ٢، ٣:

الحظائر (أو القطاعات أو المكررات)

الأولى	الثانية	الثالثة	الرابعة	الخامسة
هجين أ	هجين ج	هجين د	هجين د	هجين ب
٢ ١ ٣	١ ٢ ٣	٣ ٢ ١	٢ ١ ٣	٢ ١ ٣
هجين ج	هجين أ	هجين أ	هجين ب	هجين ج
١ ٣ ٢	١ ٣ ٢	١ ٢ ٣	٣ ٢ ١	٢ ١ ٣
هجين ب	هجين د	هجين ج	هجين ج	هجين أ
٢ ١ ٣	٢ ١ ٣	٣ ٢ ١	٣ ٢ ١	١ ٣ ٢
هجين د	هجين ب	هجين ب	هجين أ	هجين د
١ ٣ ٢	١ ٣ ٢	٢ ٣ ١	٢ ١ ٣	٢ ١ ٣

ويمكن النظر إلى هذه التجربة كتجربتين اجريا على نفس المادة التجريبية. الأولى تجربة اختبارات الهجن والثانية اختبارات العلائق وتداخلها مع الهجن. العاملان

المدروسان وهما الهجن والعلائق يسمى أحدهما الأقسام الرئيسية main plots وهى الهجن والآخر تحت الأقسام sub-plots وهى العلائق.وهذا يتحدد من أى العاملين معشى داخل الآخر. فالعامل المعشى داخل الآخر يسمى بالتحت قسم بينما العامل المعشى داخله العامل الآخر يكون هو العامل الرئيسى. ويلاحظ هنا أنه يمكن مقارنة أى عليقتين (تحت قسم) ببعضهما داخل نفس الهجين بينما لا يمكن مقارنة هجينين داخل نفس العليقة وهذا ينعكس على المقارنات بين مستويات العوامل. فواضح أنه بمقارنة عليقتين تحت نفس الهجين تكون المقارنة أدق من مقارنة هجينين لأنهما ليسا تحت نفس العليقة وهذا ينعكس على اختبارات المعنوية التى ستجرى من جدول تحليل التباين.

وهذا التصميم شائع الاستخدام فى التجارب الحقلية التى يتطلب بعض عواملها مساحة أكبر من الأخرى، فمثلا اختبارات الجرارات تحتاج إلى مساحات كبيرة بينما اختبارات مسافات الزراعة قد تتطلب مساحة أقل بينما تجارب التسميد قد تتطلب مساحة أقل ... وهكذا. كما يستخدم هذا التصميم فى التجارب الحيوانية (وأحياناً دون وعى كاف) عندما يستمر أخذ المشاهدات على نفس الحيوان لفترة زمنية continuous in time فيما يعرف بالقياسات المتكررة repeated measures وهذه تتطلب عناية معينة سيتم تناولها فيما بعد.

١٥-١٠-١ النموذج الإحصائى

بغض النظر مؤقتاً عن تحت الأقسام فإنه يوجد تجربة مصممة كقطاعات عشوائية كما سبق توضيحه فى ١٥-١٠ وجدول تحليل التباين. والآن إذا أدخل فى الصورة العامل الجديد فإنه سيمثل مصدراً جديداً للتباين وكذلك تداخله مع العامل الرئيسى وسيكون لهما خطأ آخر غير ذلك التابع للعامل الرئيسى. ويمكن كتابة النموذج الرياضى للتجربة المبينة كما يلى:

$$Y_{ijk} = \mu - b_i - h_j - e_{1ij} - r_k - (hr)_{jk} - e_{2ijk}$$

حيث

μ تمثل المتوسط العام، b_i تمثل تأثير القطاع (الحظيرة) حيث $i = 1,\cdots,5$ أو n، h_j تمثل تأثير الهجين حيث $j = 1,2,3,4$ أو m، e_{1ij} تمثل الخطأ الأول بينما r_k تمثل تأثير العليقة حيث $k = 1,2,3$ أو p، $(hr)_{jk}$ تمثل التداخل بين العليقة k والهجين j ، e_{2ijk} تمثل الخطأ الثانى.

مثال ١٥-١٠

أورد Federer (1963) فى مثال عدد البذور النابتة من كل 100 بذرة فى تجربة مصممة كقطاعات منشقة لاختبار 8 أصناف نبات V (والتى تحتاج إلى وحدات تجريبية أكثر) و 4 معاملات للبذور للإنبات r (والتى تحتاج إلى مادة تجريبية أقل). وفى هذه التجربة يراد اختبار الأصناف كعامل رئيسى ومعاملات البذور كعامل تحت رئيسى على نسبة الإنبات فى البذور، وكانت النتائج كالتالى:

المكررة الثالثة						المكررة الثانية						المكررة (القطاع) الأولى					
المجموع					الصنف	المجموع					الصنف	المجموع					الصنف
92	r_1 45	r_2 16	r_3 24	r_4 7	V_7	80	r_1 40	r_4 9	r_2 15	r_3 16	V_4	97	r_4 6	r_1 66	r_3 13	r_2 12	V_1
127	r_4 12	r_1 54	r_2 25	r_3 36	V_8	78	r_4 8	r_2 16	r_3 16	r_1 38	r_4 V_7	89	r_3 20	r_2 8	r_1 51	r_4 10	V_3
102	r_2 8	r_4 16	r_3 29	r_1 49	V_6	104	r_2 20	r_3 28	r_1 41	15	V_8	88	r_2 4	r_3 19	r_4 13	r_1 52	V_4
91	r_4 11	r_2 16	r_1 52	r_3 12	V_5	86	r_3 10	r_1 51	r_4 12	r_2 13	V_5	109	r_1 59	r_4 8	r_2 14	r_3 28	V_6
84	r_4 7	r_2 11	r_1 59	r_3 7	V_4	98	r_2 10	r_3 13	r_1 63	r_4 12	V_1	86	r_2 20	r_1 45	r_4 12	r_3 9	V_5
116	r_1 63	r_4 10	r_2 14	r_3 29	V_3	67	r_4 4	r_2 11	r_1 47	r_3 5	V_2	145	r_4 15	r_1 77	r_3 27	r_2 26	V_2
101	r_3 11	r_1 70	r_4 7	r_2 13	V_1	127	r_1 66	r_4 21	r_3 32	r_2 8	V_6	102	r_1 49	r_3 30	r_2 14	r_4 9	V_8
110	r_4 15	r_3 11	r_1 66	r_2 18	V_2	141	r_4 14	r_1 81	r_3 30	r_2 16	V_3	109	r_1 56	r_4 15	r_3 26	r_2 12	V_7
823					**مجموع**	**781**					**مجموع**	**825**					**مجموع**

ويمكن كتابة النموذج الرياضى للتجربة كما يلى:

$$Y_{ijk} = \mu - b_i - v_j - e_{1ij} - t_k - (vt)_{jk} - e_{2ijk}$$

حيث

μ تمثل المتوسط العام،

b_i تمثل تأثير القطاع حيث $i = 1, 2, 3$،

v_j تمثل تأثير الصنف حيث $j = 1, 2, \cdots, 8$ ،

e_{1ij} تمثل الخطأ أ،

t_k تمثل تأثير المعاملة k حيث $k = 1, 2, 3, 4$ ،

$(vt)_{jk}$ تمثل التداخل بين الصنف j والمعاملة k،

e_{2ij} تمثل الخطأ ب.

ومن بيانات التجربة يمكن عمل الملخص التالى:

المعاملات	الأصناف								المجموع	المتوسط
	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8		
r_1	199	190	195	151	148	174	139	144	**1340**	**55.8**
r_2	35	55	38	30	49	24	44	59	**334**	**13.9**
r_3	37	43	79	42	31	89	66	94	**481**	**20.0**
r_4	35	34	34	29	35	51	30	36	**274**	**11.4**
المجموع	**296**	**322**	**346**	**252**	**263**	**238**	**279**	**333**	**2429**	
المتوسط	**24.7**	**26.8**	**28.8**	**21.0**	**21.9**	**28.2**	**23.2**	**27.7**		**25.3**

ويمكن حساب مجموع المربعات للمصادر المختلفة من التباين كما يلى:

الكلى عن المتوسط:

$$= (12)^2 + (13)^2 + \cdots + (11)^2 + (15)^2 - \frac{(2429)^2}{96} = 36736.24$$

بين المكررات:

$$= \frac{(825)^2 + (781)^2 + (823)^2}{32} - \frac{(2429)^2}{96} = 38.58$$

بين الأصناف:

$$= \frac{(296)^2 + (322)^2 + \cdots + (279)^2 + (333)^2}{12} - \frac{(2429)^2}{96} = 763.16$$

الخطأ أ: وهو حسابياً يساوى التداخل بين الأصناف والمكررات (أو القطاعات)

$$=\frac{(97)^2+(89)^2+\cdots+(101)^2+(110)^2}{4}-\frac{(2429)^2}{96}-38.58-763.16=1377.25$$

بين المعاملات:

$$=\frac{(1340)^2+\cdots+(274)^2}{24}-\frac{(2429)^2}{96}=30774.28$$

التداخل بين الأصناف والمعاملات:

$$=\frac{(199)^2+(190)^2+\cdots+(30)^2+(36)^2}{3}-\frac{(2429)^2}{96}$$
$$-763.16-30774.28=2620.13$$

لخطأ ب (المتبقى):

$$=36736.24-38.58-763.16-1377.25-30774.28-2620.13=1162.84$$

ويمكن تكوين جدول تحليل التباين ١٥-١٣.

وبفرض أن كلا من الأصناف والمعاملات ثابتة فتجرى اختبارات فروض العدم كما يلى:

فرض العدم: الفروق بين الأصناف = صفر، $F=\frac{109.02}{98.3}=1.1$

وهى غير معنوية ولا يرفض فرض العدم.

فرض العدم: الفروق بين المعاملات = صفر، $F=\frac{10258.09}{24.23}=423.36$

وهى معنوية جداً ويرفض فرض العدم.

فرض العدم: التداخل = صفر، $F=\frac{124.77}{24.23}=5.15$

وهى معنوية ويرفض فرض العدم.

جدول ١٥-١٣ تحليل التباين لتجربة مصممة كقطاعات منشقة بعدد مكررات $n = 3$، عدد أقسام رئيسية (أصناف) $m = 8$، عدد أقسام تحت رئيسية (معاملات) $p = 4$

SOV	df	SS	MS
الأقسام الرئيسية:	$mn - 1 = 23$		
مكررات	$n - 1 = 2$	38.58	19.29
أصناف (V)	$m - 1 = 7$	763.16	109.02
خطأ أ	$(m - 1)(n - 1) = 14$	1377.25	98.38
الأقسام تحت رئيسية:	$mn(p - 1) = 72$		
معاملات (r)	$p - 1 = 3$	30774.28	10258.09**
R X V	$(m - 1)(p - 1) = 21$	2620.13	124.77*
خطأ ب	$m(n - 1)(p - 1) = 48$	1162.84	24.23
الكلى	$mnp - 1 = 95$	**36736.24**	

لاحظ أن الخطأ القياسى لكل من متوسطات المعاملات يمكن الحصول عليه كالتالى:

$$\sqrt{\frac{MS\,error(b)}{nm}} = \sqrt{\frac{24.23}{3x8}} = 1.005$$

أما الخطأ القياسى لكل من متوسطات الأصناف فيمكن الحصول عايه كالتالى:

$$\sqrt{\frac{MS\,error(a)}{np}} = \sqrt{\frac{98.38}{3x4}} = 2.863$$

وهناك ملاحظتان يجدر الإشارة إليهما بخصوص جدول ١٥-١٤:

أولاً: متوسط مربعات الخطأ (ب) أقل من متوسط مربعات الخطأ (أ) وهذا متوقع لأن المقارنات بين الأقسام تحت رئيسية تتم تحت نفس القسم الرئيسى.

ثانياً: درجات الحرية للخطأ (أ) أقل دائماً عن تلك للخطأ (ب).

وكلا الملاحظتين تؤديان إلى أن تكون اختبارات معنوية الأقسام الرئيسية (أى الأصناف) أقل دقه وأقل قوة من اختبارات كل من الأقسام تحت الرئيسية والتداخل بين المعاملات الرئيسية والمعاملات تحت الرئيسية وهذا ما أشير له عند تقديم هذا التصميم.

ولفصل المتوسطات بفرض أن العوامل ثابتة يكون هناك عدة احتمالات كما يلى:

مقارنة معاملة رئيسية بأخرى (أى الأصناف فى المثال السابق) مثلا $(V_1 - V_2)$ فن الفرق بين المتوسطين هنا له خطأ قياسى قدره:

$$= \sqrt{\frac{2\,MSerror(a)}{np}} = \sqrt{\frac{(2)(98.38)}{3x4}} = 4.049$$

والفرق بين متوسطين لمعاملتين تحت رئيسيتين (أى المعاملة فى المثال السابق) فى نفس القسم أى $(t_1 - t_2)$ مثلا له خطأ قياسى:

$$= \sqrt{\frac{2\,MSerror(b)}{nm}} = \sqrt{\frac{(2)(24.23)}{3x8}} = 1.421$$

والفرق بين متوسطى معاملتين رئيسيتين (أصناف) عند نفس المعاملة تحت ارئيسية أى $(V_2t_1 - V_1t_1)$ مثلا لها خطأ قياسى:

$$= \sqrt{\frac{2[(p-1)MSerror(b) + MSerror(a)]}{np}}$$

$$= \sqrt{\frac{2[(4-1)(24.23) + 98.38]}{3x4}} = 5.34$$

ولمقارنات أخرى يرجع إلى Federer (1963).

وبمقارنة كفاءة تصميم القطاعات المنشقة بالقطاعات العشوائية مثلا ستظهر أن لها نفس درجة الكفاءة بصفة عامة بالنسبة للعاملين معا إلا أن تصميم القطاعات العشوائية يستقطع من كفاءة العامل الرئيسى ليزيد من كفاءة العامل تحت الرئيسى وللاستزادة فى هذا الموضوع يرجع إلى Federer (1963), Cochran and Cox (1957) وكذلك لحساب القيم الغائبة.

مثال ١٥-١١

حل مثال ١٥-١٠ باستخدام برنامج SAS

```
DATA SPLITP;
INPUT REP VARITY TRT PLANT @@;
CARDS;
1 1 1 66 1 1 2 12 1 1 3 13 1 1 4 6 1 2 2 26 1 2 3 27 1 2 1 77
1 2 4 15 1 3 1 51 1 3 2 8 1 3 3 20 1 3 4 10 1 4 1 52 1 4 2 4
1 4 3 19 1 4 4 13 1 5 1 45 1 5 2 20 1 5 3 9 1 5 4 12 1 6 1 59
1 6 2 8 1 6 3 28 1 6 4 14 1 7 1 56 1 7 2 12 1 7 3 26 1 7 4 15
1 8 1 49 1 8 2 14 1 8 3 30 1 8 4 9 2 1 1 63 2 1 2 10 2 1 3 13
2 1 4 12 2 2 1 47 2 2 2 11 2 2 3 5 2 2 4 4 2 3 1 81 2 3 2 16
2 3 3 30 2 3 4 14 2 4 49 2 4 3 16 2 4 1 40 2 4 2 15 2 5 1 51
2 5 2 13 2 5 3 10 2 5 4 12 2 6 1 66 2 6 2 8 2 6 3 32 2 6 4 21
2 7 1 38 2 7 2 16 2 7 3 16 2 7 4 8 2 8 1 41 2 8 2 20 2 8 3 28
2 8 4 15 3 1 1 70 3 1 2 13 3 1 3 11 3 1 4 7 3 2 1 66 3 2 2 18
3 2 3 11 3 2 4 15 3 3 1 63 3 3 2 14 3 3 3 29 3 3 4 10 3 4 1 59
3 4 2 11 3 4 3 7 3 4 4 7 3 5 1 52 3 5 2 16 3 5 3 12 3 5 4 11
3 6 1 49 3 6 2 8 3 6 3 29 3 6 4 16 3 7 1 45 3 7 2 16 3 7 3 24
3 7 4 7 3 8 1 54 3 8 2 25 3 8 3 36 3 8 4 12
PROC GLM;
CLASS REP VARITY TRT;
MODEL PLANT = REP VARITY REP*VARITY TRT
                    VARITY*TRT/SS3;
TEST H = REP E = REP*VARITY / HTYPE = 3 ETYPE = 3;
TEST H = VARITY E = REP*VARITY / HTYPE = 3 ETYPE = 3;
MEANS VARITY / DUNCAN E =REP*VARITY ETYPE = 3;
MEANS TRT/DUNCAN ETYPE=3;
LSMEANS VARITY / E = REP*VARITY STDERR;
LSMEANS TRT/STDERR;
RUN;
```

لاحظ

تم اختبار الأصناف والمكررات باستخدام خطأ (أ) والذى يمثل فى النموذج بـــ REP*VARITY. واستخدم فى هذه الحالة أمر TEST والذى لابد أن يذكر معه H والتى تمثل المتغير المراد اختباره وكذلك E والتى تمثل الخطأ المراد استخدامه فى الاختبار.

تم وضع نوع type مجموع المربعات المراد استخدامه، وقد استخدم فى هذا البرنامج النوع الثالث type III والذى لابد أن يحدد لكل من H و E باستخدام HTYPE=3 و ETPYPE=3.

عند الرغبة فى استخدام اختبار الفصل بين المتوسطات، وهو Duncan فى هذا المثال، لابد من تحديد الخطأ المراد استخدامه والذى إذا لم يحدد فإن البرنامج سيستخدم خطأ النموذج، والذى يمثل الخطأ (ب) فى المثال السابق. لذلك عند الرغبة فى التفرقة بين متوسطات الأصناف فلابد من استخدام الخطأ (أ) والذى يعبر عنه باختيار E=REP*VARITY وكذلك تحديد نوعه والذى عبر عنه باختيار ETYPE = 3.

استخدم اختيار LSMEANS للحصول متوسطات أقل المربعات (سبق شرح هذا المفهوم من قبل) للعوامل التى تذكر بعد هذا الاختيار (الصنف والمعامله فى هذا المثال) وكذلك الحصول على الخطأ القياسى لهذه المتوسطات باستخدام اختيار STDERR مع ذكر نوع الخطأ الذى سوف يستخدم فى حساب الخطأ القياسى عن طريق استخدام اختيار E=REP*VARITY لحساب الخطأ القياسى للأصناف والذى إذا لم يحدد فإن البرنامج سيستخدم خطأ النموذج (وهو الخطأ ب) وهذا من الأخطاء الشائعة فى مثل هذه الحالات.

نتائج التحليل

The GLM Procedure
Class Level Information

Class	Levels	Values
REP	3	1 2 3
VARITY	8	1 2 3 4 5 6 7 8
TRT	4	1 2 3 4

Number of observations 96

Dependent Variable: PLANT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	47	35573.40625	756.88098	31.24	<.0001
Error	48	1162.83333	24.22569		
Corrected Total	95	36736.23958			

R-Square	Coeff Var	Root MSE	PLANT Mean
0.968346	19.45279	4.921960	25.30208

Source	DF	Type III SS	Mean Square	F Value	Pr > F
REP	2	38.58333	19.29167	0.80	0.4568
VARITY	7	763.15625	109.02232	4.50	0.0006
REP*VARITY	14	1377.25000	98.37500	4.06	0.0001
TRT	3	30774.28125	10258.09375	423.44	<.0001
VARITY*TRT	21	2620.13542	124.76835	5.15	<.0001

Tests of Hypotheses Using the Type III MS for REP*VARITY as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
REP	2	38.5833333	19.2916667	0.20	0.8241
VARITY	7	763.1562500	109.0223214	1.11	0.4100

Duncan's Multiple Range Test for PLANT

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	14
Error Mean Square	98.375

Number of Means	2	3	4	5	6	7	8
Critical Range	8.685	9.100	9.357	9.530	9.653	9.742	9.810

Means with the same letter are not significantly different.

Duncan Grouping	Mean	N	VARITY
A	28.833	12	3
A	28.167	12	6
A	27.750	12	8
A	26.833	12	2
A	24.667	12	1
A	23.250	12	7
A	21.917	12	5
A	21.000	12	4

Duncan's Multiple Range Test for PLANT

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	48
Error Mean Square	24.22569

Number of Means	2	3	4
Critical Range	2.857	3.005	3.102

Means with the same letter are not significantly different.

Duncan Grouping	Mean	N	TRT
A	55.833	24	1
B	20.042	24	3
C	13.917	24	2
C	11.417	24	4

Least Squares Means

Standard Errors and Probabilities Calculated Using the Type III MS for REP*VARITY as an Error Term

VARITY	PLANT LSMEAN	Standard Error	Pr > |t|
1	24.6666667	2.8632004	<.0001
2	26.8333333	2.8632004	<.0001
3	28.8333333	2.8632004	<.0001
4	21.0000000	2.8632004	<.0001
5	21.9166667	2.8632004	<.0001
6	28.1666667	2.8632004	<.0001
7	23.2500000	2.8632004	<.0001
8	27.7500000	2.8632004	<.0001

TRT	PLANT LSMEAN	Standard Error	Pr > |t|
1	55.8333333	1.0046910	<.0001
2	13.9166667	1.0046910	<.0001
3	20.0416667	1.0046910	<.0001
4	11.4166667	1.0046910	<.0001

١٥-١٠-٢ أشكال أخرى من تصميم القطاعات المنشقة

قد تتطلب الحاجة إلى أن يؤخذ عامل ثالث ويعشى داخل العامل الثانى (غير الرئيسى) لنفس الدواعى التي سبق ذكرها من حجم المادة التجريبية المتطلب ودرجة الدقة المنشودة ... الخ. وفى هذه الحالة يسمى التصميم القطاعات المنشقة المنشقة Split split-plot design وهكذا إلى أى درجة توجبها التجربة.

كما لوحظ فى تحليل المثال ١٥-١٠ أن العامل الرئيسى وهو الأصناف كان يتبع تصميم القطاعات العشوائية بينما العامل تحت الرئيسى يتبع التصميم تام العشوائية. ولكن هذا ليس هو الحال بالضرورة فى جميع الحالات. فقد يصمم العامل الرئيسى مثلا تصميما تام العشوائية والتحت رئيسى كتام العشوائية أيضاً. كأن يكون هناك مثلا ثلاث سلالات وبكل سلالة عشر حيوانات مختارة عشوائياً، إلى هنا والتجربة تصميمها تام العشوائية. وإذا طبقت معاملات على الحيوانات عشوائياً فإن هذه المعاملات تعتبر تحت أقسام، فكلا العاملين الرئيسى والتحت رئيسى يتبعان التصميم تام العشوائية. وقد يكون العامل الرئيسى تابعا لتصميم المربع اللاتينى بينما العامل التحت رئيسى يتبع العشوائى التام، أو أيضاً المربع اللاتينى. وهكذا يمكن مزج أى تصميمات بالنسبة للعامل الرئيسى والعامل التحت رئيسى طبقاً لمقتضيات المادة التجريبية.

ويبين الشكل التالى تجربة قطاعات منشقة لأربعة مستويات للعامل الرئيسى A مصممة كمربع لاتينى ٤ × ٤ وثلاث مستويات للعامل B كتحت رئيسى يتبع التصميم تام العشوائية.

	عمود ١	عمود ٢	عمود ٣	عمود ٤
صف ١	a_4: b_2, b_3, b_1	a_3: b_3, b_1, b_2	a_2: b_3, b_2, b_1	a_1: b_1, b_3, b_2
صف ٢	a_3: b_1, b_3, b_2	a_2: b_1, b_2, b_3	a_1: b_3, b_2, b_1	a_4: b_3, b_1, b_2
صف ٣	a_2: b_1, b_2, b_3	a_1: b_2, b_1, b_3	a_4: b_1, b_2, b_3	a_3: b_2, b_3, b_1
صف ٤	a_1: b_2, b_1, b_3	a_4: b_1, b_2, b_3	a_3: b_1, b_3, b_2	a_2: b_2, b_1, b_3

ويمكن وضع تحليل التباين لمثل هذا التصميم كما يلى:

SOV	df
العامل الرئيسي:	$(a^2 - 1) = 15$
أعمدة	$(a - 1) = 3$
صفوف	$(a - 1) = 3$
معاملة A	$(a - 1) = 3$
خطأ أ	$(a - 1)(a - 2) = 6$
العامل تحت الرئيسي	$a^2(b - 1) = 32$
معاملة B	$(b - 1) = 2$
تداخل A x B	$(a - 1)(b - 1) = 6$
خطأ ب	$a(a - 1)(b - 1) = 24$
الكلي	$a^2b - 1 = 47$

وهناك تصميم على علاقة بتصميمات القطاعات المنشقة يطلق عليه split block وفيه يكون كلا العاملين على قدم المساواة ولا يعشى أحدهما داخل الآخر ولكن كل يعشى على الآخر وفي هذه الحالة فإن التجربة تركز أكثر على التداخل بين العاملين. فإذا فرض أن هناك عاملين A (4 مستويات)، B (3 مستويات)، وزع A عشوائياً على المادة التجريبية بغض النظر عن B، ووزع B عشوائياً على نفس المادة التجريبية بغض النظر عن A ويصبح شكل التصميم كما يلي:

مكررة ١

	a_1	a_4	a_2	a_3
b_2				
b_1				
b_3				

مكررة ٢

	a_2	a_3	a_1	a_4
b_3				
b_2				
b_1				

ويكون تحليل التباين كما يلي:

SOV	df
مكررات	1
بين A	3
خطأ أ	3
بين B	2
خطأ ب	2
تداخل A x B	6
خطأ ج	6

ويختبر A، B، A × B على خطأ أ، خطأ ب، خطأ ج على التوالى.

١٥-١٠-٣ تنويه عن استخدام تصميم القطاعات المنشقة فى التجارب على الإنسان والحيوان والمحاصيل الحقلية المستديمة

فى التصميم الأساسى للقطاعات المنشقة يكون العامل الرئيسى معشى وأيضاً العامل تحت الرئيسى معشى داخل العامل الرئيسى. وهذا يؤدى إلي الاعتقاد بتجانس مصفوفة التباين والتغاير الخاصة بمستويات المعاملة تحت الرئيسية. فمثلا إذا كانت B هى العامل تحت الرئيسى فيتوقع وجود شرط أن تباينات المجاميع b_1، b_2 متساوية وكذلك تغاير كل b مع الأخرى. هذا الشرط مطلوب وضروري لإجراء اختبار F سليم. تخيل أن باحث يقارن بين سلالتين من الحيوانات من حيث إنتاج الحيوانات المنوية ويأخذ عشوائياً 10 حيوانات من كل سلالة، وهو أيضاً يدرس أثر المواسم المختلفة الأربعة على إنتاج السائل المنوى على هذه الحيوانات فإن هذا ما أطلق عليه تصميم القطاعات المنشقة المستمر فى الزمن Split plot design continuous in time وشاع استخدامه فى هذا النوع من التجارب حيث تعتبر السلالات (أو أى أقسام أخرى) العامل الرئيسى والمواسم العامل تحت الرئيسى. وفى التجربة سابقة الشرح يكون تحليل التباين كالتالى:

SOV	df
بين السلالات	1
بين الحيوانات داخل السلالات (خطأ أ)	18
بين المواسم	3
التداخل بين السلالات والمواسم	3
المتبقى (خطأ ب)	54
الكلى	**79**

وعادة ما يستخدم خطأ (أ) لاختبار السلالات وهذا صحيح تماماً. ولكن عند النظر إلى المواسم فإنها لم تعش بالمعنى السابق شرحه. هذه التعشية هى التى تضمن تجانس مصفوفة التباين والتغاير بين مجاميع المعاملة تحت رئيسية (المواسم). وللفرد أن يظن أن التغاير بين موسمين متتاليين على نفس الحيوان سيكون أعلى من التغاير بين موسمين غير متتاليين. وإذ صح هذا الظن فإن المجرب يكون قد تجاوز الحدود بهذا التصميم وعليه:

أولا: أن يختبر مصفوفة التباين والتغاير هذه،

ثانيا: أنه ليس هناك أثر باقى فى الحيوان فى مجموعة حالية من المجموعة السابقة. وسوف يتم تناول هذا الاختبار فيما بعد.

وقد حذر كل من Rowell and Walter (1976) ،Allen et al (1983) من مثل هذا الاستخدام الجائر وتطرقا إلى أساليب لمعالجة مثل هذا النوع من البيانات، وهذه المعالجة تختلف باختلاف المواقف. ومن هذه الأساليب استخدام الـ multivariate analysis polynomial أو تحليل التغاير covariance analysis.

ومع وجوب التحفظ الشديد حتى فى إتباع تصميم القطاعات المنشقة فى البيانات المستمرة فى الزمن continuous in time وتخصيص خطأ مستقل لكل من العوامل المدروسة فإن كثيراً من البحوث تجرى تحليلاتها بافتراض تمام استقلالية المشاهدات وتجانس مصفوفة تباينها وتغايرها وهذا تجاوز أبشع يجب العمل على تلافيه تماماً. والمثل السيئ لهذا هو أن تحلل التجربة السابقة كما يلى:

df	SOV
1	بين السلالات
3	بين المواسم
3	التداخل بين السلالات والمواسم
72	الخطأ
79	**الكلى**

١٥- ١١ تحليل المشاهدات المتكررة على نفس الوحدة التجريبية

يقصد بالمشاهدات المتكررة بصفة عامة أنها البيانات التى تعبر عن الاستجابة والتى تقاس على الوحدة التجريبية فى فترات متعددة. إذا تكررت المشاهدة مرتين على نفس الفرد فإن تحليل مثل هذه البيانات قد تم شرحه سابقاً "اختبار ("t" للأزواج). ولكن فى هذا الفصل سيتم شرح طرق التحليل فى حالة ثلاثة مشاهدات أو أكثر. وقد نال هذا الموضوع أهمية مؤخراً فقط، مثلاً أنظر Davis (2002).

فكثير من التجارب تصمم على أن تقاس استجابة الفرد على فترات لمعرفة تطور هذه الاستجابة. وتجارب أخرى تصمم على أخذ المشاهدة على نفس الفرد واستخدام هذا الفرد عدة مرات لتسجيل مشاهدات أخرى. مثال الحالة الأولى أن يحقن حيوان بهرمون ويؤخذ عليه عدة مشاهدات لتسجيل تركيز الهرمون على فترات زمنية معينة.

ومثال الحالة الثانية أن تقاس درجة حرارة الحيوان فى الصيف ثم تقاس درجة حرارته ثانياً وثالثاً ورابعاً فى الخريف والشتاء والربيع. فمثلا مجرب يريد أن يدرس أثر نظامين لإيواء الأبقار (نظام مسقوف وآخر مكشوف) على درجة حرارة الحيوان (درجة حرارة المستقيم) خلال اليوم (عند الساعات ٦، ١٢، ١٨، ٢٤)، المجرب لديه ٣٢ حيواناً متماثلاً. ويقوم بتقسيم الحيوانات إلى ٨ (= ٢ نظام × ٤ أوقات) مجاميع عشوائية ويخصص لكل توليفة نظام إيواء ووقت ٤ حيوانات كما فى الشكل التالى:

	الساعة ٦		**الساعة ١٢**		**الساعة ١٨**		**الساعة ٢٤**	
إيواء مسقوف	١	٥	٨	٢٥	٣	٢٢	١٢	١٣
	١٧	١٨	١٥	٢١	٩	٢٦	٣٢	٢
إيواء مكشوف	٦	٢٠	٢٣	٣٠	١٠	٢٤	٤	١٦
	٧	١٩	١٤	٢٩	١١	٢٧	٢٨	٣١

ويصبح توزيع درجات الحرية (وبالتالى التحليل) كما يلى:

df	SOV
1	نظام الإيواء
3	الوقت
3	نظام الإيواء x الوقت
24	الخطأ (حيوان داخل توليفة الإيواء والوقت)
31	**الكلى**

ولكن بفرض أن المجرب قرر ألا يعشى الحيوانات على الوقت وتم قياس درجات الحرارة فى الأربع أوقات على نفس الحيوان فيصبح التصميم كما هو موضح بالشكل التالى:

أرقام الحيوانات تحت نظام الإيواء المسقوف

الوقت	٢	٥	٧	١٢	١٥	١٦	١٧	١٨	٢١	٢٢	٢٣	٢٤	٢٧	٢٩	٣٠	٣١
٦																
١٢																
١٨																
٢٤																

أرقام الحيوانات تحت نظام الإيواء المكشوف

الوقت	١	٣	٤	٦	٨	٩	١٠	١١	١٣	١٤	١٩	٢٠	٢٥	٢٦	٢٨	٣٢
٦																
١٠																
١٨																
٢٤																

ويكون شكل التحليل كالتالى:

SOV	df
نظام الإيواء	1
الحيوانات (داخل الإيواء) = خطأ أ	30
الوقت	3
نظام الإيواء x الوقت	3
الخطأ = خطأ ب	90
الكلى	**127**

الفرق بين التحليلين السابقين هو لب تحليل المشاهدة المتكررة. ففى التحليل الأول كانت الوحدة التجريبية هى الحيوان ولكن فى التحليل الثانى كانت الوحدة التجريبية هى القياس فى الوقت المعين ويوجد منه ١٢٨ قياس.

ويلاحظ التشابه بين التحليل الثانى وتصميم القطاعات المنشقة مع فارق هام هو أنه لم تتم التعشية على الوقت على خلاف القطاعات المنشقة حيث تتم مثل هذه التعشية. بمعنى أن المتغير الأول له مستويين هما نظاما الإيواء وفيه تعشى الحيوانات على الإيواء حيث يمثل مصدر التباين الحيوانات داخل الإيواء الخطأ (أ). أما المتغير الثانى وهو الوقت وله الخطأ الأخير ذو الـــ ٩٠ درجة حرية وهو ما قد يطلق عليه الخطأ (ب) ولم تتم التعشية على الوقت على خلاف القطاعات المنشقة حيث تتم مثل هذه التعشية، أى أن نفس الحيوان سيقاس عليه الأربع أوقات. والسؤال هنا هل غياب التعشية أمر مسموح به أو مرغوب؟ من المهم معرفة أن هذا أمر مجافى تماماً للفرض بين جميع المشاهدات مستقلة، فكون المشاهدة تؤخذ على نفس الحيوان يخلق ارتباطاً محتملاً بين هذه المشاهدات وهذا إخلال بالفرض الأساسى، إذن كيف التصرف فى مثل هذه الحالة؟

١٥-١١-١ مميزات ومضار المشاهدات المتكررة

١- المشاهدات المتكررة هى إخلال صريح لاختبارات فرض العدم (أى استقلالية المشاهدات) ولكن لا يمكن التغاضى عن هذا فى التحليل إلا بعد إتمام بعض الاختبارات التى سيتم شرحها فيما بعد.

٢- واضح جداً أن المشاهدات المتكررة تزيد كثيراً من درجات الحرية المتاحة للمتغير الذى لم يتم عليه التعشية وعليه فإن قوة اختبار فرض العدم تزيد، (فى المثال السابق 24 درجة حرية للخطأ فى التحليل الأول مقارنة بـــ 90 درجة فى التحليل الثانى).

٣- تثبيت قياس المعاملة فى أوقات مختلفة على نفس الحيوان يزيل أثر الحيوان وربما يوضح بدرجة أكبر فرق المعاملات. وهذا نفس المنطق الذى من أجله يستخدم اختبار t فى حالة الأزواج.

إذا توافرت الاشتراطات اللازمة لإتباع التحليل الثانى فإن هذا سيؤدى إلى خفض عدد الوحدات التجريبية إلى حد كبير وبالتالى خفض تكاليف التجربة. ويبين الشكل ١٥-٥ الاستراتيجية التى يمكن إتباعها عند وجود بيانات ذات مشاهدات متكررة.

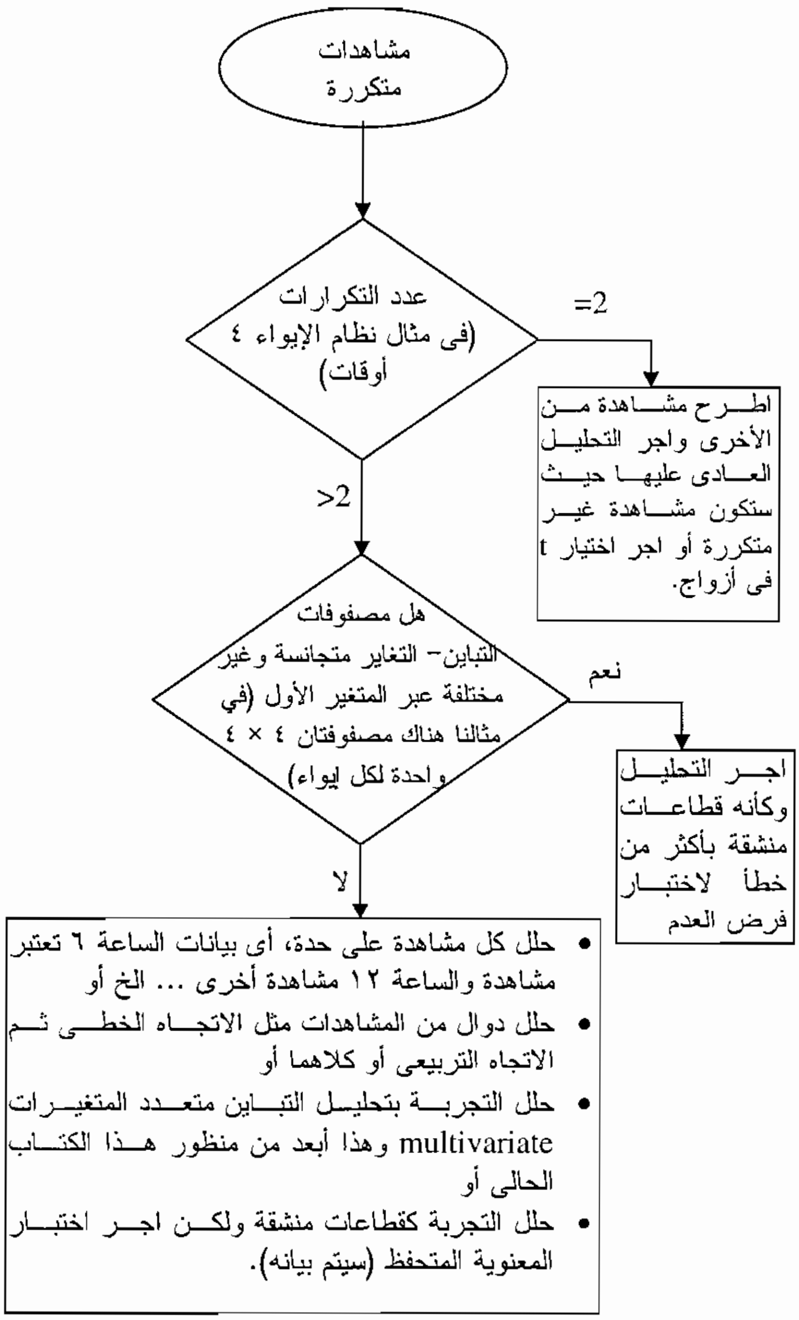
مثال ١٥-١٢

يقيس مجرب النشاط الجنسى لكباش بالمدة الزمنية التى يأخذها الكبش لاعتلاء النعجة عند إدخال نعجة شائعة عليه. والتالى المدد بالدقائق لثلاث تراكيب وراثية فى الأربعة فصول للسنة. وقد اختيرت الكباش عشوائيا من كل سلالة وأخذت المشاهدات على نفس الكبش فى الفصول الأربعة وكانت البيانات كالتالى:

السلالة	الحيوان	فصل السنة			
		خريف	شتاء	ربيع	صيف
مرينو	١	44	39	10	19
	٢	31	23	35	15
	٣	48	39	18	22
	٤	19	36	25	28
	٥	51	38	27	9
خليط	١	23	24	12	26
	٢	17	21	3	26
	٣	16	17	14	25
	٤	10	16	13	2
	٥	16	13	19	2
أوسيمى	١	36	16	9	30
	٢	21	44	19	23
	٣	21	14	23	26
	٤	20	34	25	38
	٥	48	23	17	21

ويكون تحليل التباين لبيانات مثال ١٥-١٢، إذا افترض تصميم القطاعات المنشقة، كما يلى:

SOV	df	SS	MS
السلالة	2	1833.23	916.62
الحيوانات/السلالة (خطأ أ)	12	600.00	50.00
الفصل	3	1016.98	338.99
السلالة x الفصل	6	770.77	128.46
خطأ ب	36	3426.90	95.16



شكل ١٥-٥ استراتيجية تحليل البيانات ذات المشاهدة المتكررة

مثال ١٥-١٣

حل مثال ١٥-١٢ باستخدام برنامج SAS

```
DATA  SHEEP;
INPUT  BREED $ SEASON $ ANIMAL TIME  @@;
*-- M = MERINO ---   C = CROSS ---   O = OSSIMI ----;
* -- A = AUTUMN --- W = WINTER --- S = SPRING ---
   SR =  SUMMER ---;
CARDS;
M A 1 44 M A 2 31 M A 3 48 M A 4 19 M A 5 51
M W 1 39 M W 2 23 M W 3 39 M W 4 36 M W 5 38
M S 1 10 M S 2 35 M S 3 18 M S 4 25 M S 5 27
M SR 1 19 M SR 2 15 M SR 3 22 M SR 4 28 M SR 5 9
C A 1 23 C A 2 17 C A 3 16 C A 4 10 C A 5 16
C W 1 24 C W 2 21 C W 3 17 C W 4 16 C W 5 13
C S 1 12 C S 2 3 C S 3 14 C S 4 13 C S 5 19
C SR 1 26 C SR 2 26 C SR 3 25 C SR 4 2 C SR 5 2
O A 1 36 O A 2 21 O A 3 21 O A 4 20 O A 5 48
O W 1 16 O W 2 44 O W 3 14 O W 4 34 O W 5 23
O S 1 9 O S 2 19 O S 3 23 O S 4 25 O S 5 17
O SR 1 30 O SR 2 23 O SR 3 26 O SR 4 38 O SR 5 21
PROC GLM;
CLASS BREED SEASON ANIMAL;
MODEL TIME = BREED ANIMAL(BREED) SEASON
                 BREED*SEASON/SS3;
TEST H = BREED E = ANIMAL(BREED)/ HTYPE  = 3 ETYPE = 3;
MEANS BREED / DUNCAN  E = ANIMAL(BREED)  ETYPE = 3;
MEANS SEASON / DUNCAN ETYPE=3;
LSMEANS BREED / E = ANIMAL(BREED)  STDERR;
LSMEANS SEASON / STDERR;
RUN;
```

نتائج التحليل

```
              The GLM Procedure
           Class Level Information
        Class      Levels    Values
       BREED         3       C M O
       SEASON        4        A S SR W
       ANIMAL        5        1 2 3 4 5
       Number of observations    60
```

Dependent Variable: TIME

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	23	4220.983333	183.521014	1.93	0.0376
Error	36	3426.000000	95.166667		
Corrected Total	59	7646.983333			

R-Square	Coeff Var	Root MSE	TIME Mean
0.551980	41.83849	9.755340	23.31667

Source	DF	Type III SS	Mean Square	F Value	Pr > F
BREED	2	1833.233333	916.616667	9.63	0.0004
ANIMAL(BREED)	12	600.000000	50.000000	0.53	0.8839
SEASON	3	1016.983333	338.994444	3.56	0.0235
BREED*SEASON	6	770.766667	128.461111	1.35	0.2612

Tests of Hypotheses Using the Type III MS for ANIMAL(BREED) as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
BREED	2	1833.233333	916.616667	18.33	0.0002

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	12
Error Mean Square	50

Number of Means	2	3
Critical Range	4.872	5.100

Means with the same letter are not significantly different.

Duncan Grouping	Mean	N	BREED
A	28.800	20	M
A	25.400	20	O
B	15.750	20	C

Duncan's Multiple Range Test for TIME

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	36
Error Mean Square	95.16667

Number of Means	2	3	4
Critical Range	7.224	7.595	7.836

Means with the same letter are not significantly different.

Dunean Grouping		Mean	N	SEASON
	A	28.067	15	A
	A	26.467	15	W
B	A	20.800	15	SR
B		17.933	15	S

Least Squares Means

Standard Errors and Probabilities Calculated Using the Type III MS for ANIMAL(BREED) as an Error Term

BREED	TIME LSMEAN	Standard Error	Pr > |t|
C	15.7500000	1.5811388	<.0001
M	28.8000000	1.5811388	<.0001
O	25.4000000	1.5811388	<.0001

SEASON	TIME LSMEAN	Standard Error	Pr > |t|
A	28.0666667	2.5188181	<.0001
S	17.9333333	2.5188181	<.0001
SR	20.8000000	2.5188181	<.0001
W	26.4666667	2.5188181	<.0001

١٥-١١-٢ اختبار صلاحية التحليل على أساس القطاعات المنشقة

يلزم اختبار تساوى مصفوفات التباين-التغاير وكذلك اختبار تماثل symmetry تلك المصفوفات حتى يكون تصميم القطاعات المنشقة صالحا valid للاستخدام فى حالة تحليل البيانات ذات المشاهدات المتكررة.

أولا: اختبار تساوى مصفوفات التباين-التغاير

١- تكوين مصفوفة تباين - تغاير var-cov matrix مصححة للمتوسط للمشاهدات المتكررة داخل كل مستوى أساسى (السلالات فى مثال ١٥-١٢).

٢- أى سوف يكون هناك المصفوفات A_1، A_2، A_3 للسلالات الثلاثة كما يلى:

$$A_1 = \begin{bmatrix} 178.3 & 42.00 & -49.50 & -60.45 \\ 42.00 & 46.5 & -50.00 & 9.75 \\ -49.6 & -50.00 & 89.5 & -21.25 \\ -60.45 & 9.75 & -21.25 & 51.3 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 21.3 & 14.15 & -3.85 & 40.90 \\ 14.15 & 18.7 & -16.55 & 44.70 \\ -3.85 & -16.55 & 33.7 & -46.05 \\ 40.90 & 44.70 & -46.05 & 168.2 \end{bmatrix}$$

$$A_3 = \begin{bmatrix} 154.7 & -61.8 & -48.40 & -38.15 \\ -61.8 & 160.2 & 26.6 & 3.85 \\ -48.40 & 26.60 & 38.8 & 11.30 \\ -38.15 & 3.85 & 11.30 & 45.3 \end{bmatrix}$$

فالقيمة 178.3 مثلاً هى مجموع مربعات قيم المرينو المصححة في الخريف مقسوماً على درجات الحرية، أى

$$= [44^2 + 31^2 + \cdots + 51^2 - \frac{193^2}{5}]/4$$

والقيمة 42.00 هى مجموع حاصل ضرب مشاهدات الخريف فى مشاهدات الشتاء للمرينو مقسوماً على درجات الحرية، أى

$$= [(44)(39) + (31)(23) + \cdots + (51)(38) - \frac{(193)\,(175)}{5}]/4$$

٣- تحسب المصفوفة الكلية A_{pooled} وكل عنصر فيها هو متوسط العناصر الثلاثة المقابلة فى A_1، A_2، A_3 وهى:

$$A_{pooleed} = \begin{bmatrix} 118.100 & -1.883 & -33.917 & -19.233 \\ -1.883 & 75.1333 & -13.483 & 19.433 \\ -33.917 & -13.483 & 54.000 & -18.667 \\ -19.233 & 19.433 & -18.667 & 88.2667 \end{bmatrix}$$

حيث

$118.100 = \dfrac{178.3 - 21.3 - 154.7}{3}$، وهكذا لبقية العناصر.

٣ – حساب قيمة محدد determinant للمصفوفات A_1، A_2، A_3، A_{pooled} وللحصول على ذلك للمثال ١٥–١٢ افترض:

عدد مستويات المتغير الأساسى (السلالات) p حيث $p = 3$

عدد مستويات المتغير الثانوى (الفصول) q حيث $q = 4$

عدد الوحدات داخل كل متغير أساسى n حيث $n = 5$

عدد الوحدات فى كل مستويات المتغير الأساسى N حيث $N = np = 15$

وبالرمز لمحدد المصفوفة بالخطين الرأسيين يمكن الحصول على المحددات التالية:

$$|A_1| = 2,368,612.38$$

$$|A_2| = 104,166.57$$

$$|A_3| = 17,040,426.05$$

$$|A_{pooled}| = 25,873,858.91$$

٤– حساب قيمة M_1 و E_1 حيث

$$M_1 = (N - p)\ \ln\left|A_{pooled}\right| - \left[\sum_{i}^{p}(n_i - 1)\ln|A_i|\right]$$

$$E_1 = \frac{2q^2+3q-1}{6(q+1)(p-1)}\left(\sum_i^p \frac{1}{n_i-1} - \frac{1}{N-p}\right)$$

حيث ln هو اللوغاريتم الطبيعي. وبالتعويض فى مثال ١٥-١٢ فإن

$$M_1 = (15-3)(17.069) - 4(14.678 - 11.554 - 16.651) = 33.296$$

$$E_1 = \frac{2(4^2)+3(4)-1}{6(4+1)(3-1)}\left[\left(\frac{1}{4}+\frac{1}{4}+\frac{1}{4}\right) - \frac{1}{15-3}\right]$$
$$= \frac{2(16)+11}{(30)(2)}\left(\frac{3}{4}-\frac{1}{12}\right) = (1.394)(0.896) = 3.463$$

٥ – حساب χ_1^2 ودرجات الحرية لها حيث

$$\chi_1^2 = (1-E_1)M_1$$

ودرجات الحرية

$$df_1 = \frac{q(q+1)(p-1)}{2}$$

وبالتعويض فى مثال ١٥-١٢

$$\chi_1^2 = (1-3.463)(33.296) = 3.463$$

ودرجات الحرية

$$df_1 = \frac{4(4+1)(3-1)}{2} = 20$$

قيمة مربع كاى غير معنوية وعليه لا يرفض فرض العدم ويستخلص تساوى التباين– التغاير داخل السلالات.

ثانيا: اختبار تماثل symmetry المصفوفات

١– تكوين مصفوفة تسمى A_{ave} وذلك بأخذ متوسط كل العناصر القطرية من المصفوفة A_{pooled} لتشكل كل العناصر القطرية فى المصفوفة A_{ave}، ثم يؤخذ متوسط جميع القيم خارج القطر من المصفوفة A_{pooled} (نصف المصفوفة فقط) لتشكل بقية عناصر المصفوفة A_{ave}.

بالتعويض فى مثال ١٥-١٢ تكون القيم القطرية للمصفوفة A_{ave} عبارة عن

$$=\frac{118.1 - 75.1333 - 54 - 88.2667}{4} = 83.875$$

والقيم خارج القطر عبارة عن

$$=\frac{(-1883 - 33.917 - 19.233 - 13.483 - 19.433 - 18.667)}{6} = -11.292$$

وبالتالى تكون A_{ave} عبارة عن

$$A_{ave} = \begin{bmatrix} 83.875 & -11.292 & -11.292 & -11.292 \\ -11.292 & 83.875 & -11.292 & -11.292 \\ -11.292 & -11.292 & 83.875 & -11.292 \\ -11.292 & -11.292 & -11.292 & 83.875 \end{bmatrix}$$

٢- حساب محدد A_{ave}

فى مثال ١٥-١٢ يكون المحدد $|A_{ave}| = 43,094,771.06$

٣- حساب قيمة M_2 و E_2 حيث

$$M_2 = -(N-P)\ \ln \frac{|A_{pooled}|}{|A_{ave}|}$$

$$E_2 = \frac{q(q+1)^2(2q-3)}{6(N-P)(q-1)(q^2+q-4)}$$

بالتعويض فى مثال ١٥-١٢

$$M_2 = -(15-3)\ \ln(\frac{25,873,858.91}{43,094,771.06}) = -12\ \ln\ (0.600) = 6.122$$

$$E_2 = \frac{4(4+1)^2(8-3)}{6(15-3)(4-1)(16+4-4)} = \frac{500}{3456} = 0.145$$

٤- حساب χ_2^2 ودرجات الحرية لها حيث

$$\chi_2^2 = (1 - E_2)M_2$$

ودرجات الحرية

$$df_2 = \frac{q^2 + q - 4}{2}$$

وبالتعويض فى مثال ١٥-١٢

$$\chi_2^2 = (1 - 0.145)(6.122) = 5.234$$

ودرجات الحرية

$$df_2 = \frac{4^2 + 4 - 4}{2} = 8$$

وقيمة مربع كاى عند درجات الحرية هذه غير معنوية وعليه لا يرفض فرض العدم: أن المصفوفات الثلاثة متماثلة. وعليه يمكن تبنى تحليل القطاعات المنشقة واختبار معنويات الفروق بنفس الطريقة السابق شرحها عند مناقشة القطاعات المنشقة.

وفى حالة إذا بينَ التحليل أن مصفوفات التباين- التغاير غير متماثلة أو غير متساوية فيجرى تحليل التباين لقطاعات منشقة ولكن تضرب قيمة درجات الحرية للخطأ فى القيمة $\theta = [1/(1 - q)]$ وهذا من شأنه أن يجعل اختبار المعنوية أكثر تحفظا (أى يرفض فرض العدم بدرجة أقل أى الاختبار المتحفظ) يرجع فى هذا إلى Geisser and Greenhouse (1958) و Kirk (1995) .

١٥-١٢ تحليل التغاير Analysis of Covariance

فى كل التصميمات السابق شرحها كان هناك متغير واحد يقاس على الوحدة التجريبية وهذا المتغير يتم تحليل التباين فيه لمعرفة أثر العوامل المذكورة فى النموذج model. ففى مثال ١٥-١ كان هذا المتغير هو النمو فى الدجاج والعامل المؤثر هو كمية الفيتامين فى العليقة. بينما فى مثال ١٥-٤ كانت الصفة هى محصول اللبن بينما المؤثرات هى نوع الهوايات والحظائر وموسم الحليب ... وهكذا. وفى كثير من الأحيان يكون أحد العوامل المؤثرة على المتغير المدروس ليس بمتغير متقطع discrete ولكنه متغير متصل continuous أى متغير مستقل متصل. فعلى سبيل المثال إذا كانت التجربة هى دراسة أثر ثلاثة علائق على معدل التسمين فى الحملان، فإنه عادة ما يكون من غير الممكن الحصول على حملان متساوية فى وزنها الابتدائى. وبتوزيع الحملان عشوائياً على المعاملات الثلاث فإنه من المتوقع، ولكن ليس هناك ضمان، أن تتساوى المعاملات فى متوسط أوزانها الابتدائية، خصوصا إذا

صندوق ١٥-٢

تكرار المشاهدة على نفس الوحدة التجريبية وتجاهل هذا عند التحليل يعتبر أمراً مخلاً بافتراضات تحليل التباين (أساساً استقلال المشاهدات). إذا كان التكرار على الوقت أو المكان فإن هيكل البيانات يشبه تلك التابع للقطاعات المنشقة ولكن الأخيرة إذا صممت جيداً فإن المشاهدات بها تكون مستقلة. فى مثل هذه الحالات يمكن تحليل تجارب المشاهدات المتكررة على أنها قطاعات منشقة فقط بعد إتمام اختبارين هما تساوى مصفوفات التباين- التغاير وكذلك تماثلها. إذا لم يرفض فرض العدم بتجانسها أو تساويها- فإنه يمكن تحليل البيانات على أنها قطاعات منشقة أما إذا رفض أحد فرضى العدم (أى التساوى والتجانس) يمكن إجراء التحليل كقطاعات منشقة ولكن يستخدم اختبار معنوية متحفظ.

كانت الأعداد كبيرة، ولكن سيظل هناك تباين بين حملان المعاملة الواحدة نتيجة لاختلاف أوزانها الابتدائية. هذا التباين من شأنه أن يزيد من تباين الخطأ وبالتالى الإنقاص من كفاءة التجربة أو قوة الاختبار. هذا بجانب الاحتمال القائم أن يتدخل المجرب لفرض تساوى المتوسطات بتحريك الحملان من معاملة إلى أخرى بغرض تسوية المتوسطات الابتدائية والذى سيؤدى إلى تحيز النتائج وهذا أمر غير مرغوب فيه.

وتحليل التغاير ليس بالتصميم فى حد ذاته ولكنه وسيلة لضبط الخطأ وزيادة كفاءة التجربة ويمكن أن يكون مصاحباً لأى تصميم من التصميمات سالفة الذكر.

مثال ١٥-١٤

أراد مجرب أن يختبر الفروق بين ثلاثة علائق تسمين أ ، ب ، ج فاشترى 24 حملا من الأسواق ووزعها عشوائياً على الثلاث معاملات. وحيث أن الحملان لم تكن متساوية فى الأوزان عند الشراء، وحيث أنه من المعلوم أن الوزن الابتدائي في التسمين ممكن أن يكون له أثر على معدل زيادة الحيوان فى الوزن فإن المجرب أراد

أن يزيل أثر الوزن الابتدائى هذا على الوزن النهائى فى تجربته. والنتائج التالية تمثل التجربة حيث X هى الوزن الابتدائى للحمل بينما Y هى وزنه النهائى (هذا المثال يمثل تحليل التغاير فى تصميم تام العشوائية أو تقسيم أحادى الاتجاه).

الحيوان	عليقة أ		عليقة ب		عليقة ج	
	X	Y	X	Y	X	Y
١	24.5	38.5	25.6	40.5	22.6	29.9
٢	22.5	33.4	16.5	19.1	25.8	38.0
٣	25.5	38.0	18.5	22.3	24.5	29.0
٤	18.4	32.3	16.4	20.0	21.9	24.8
٥	16.9	24.4	20.4	25.1	18.9	19.2
٦	19.1	28.0	27.4	42.2	20.8	21.7
٧	23.2	42.5	14.4	15.1	22.8	28.4
٨	19.2	27.1	21.2	33.6	15.9	17.0
المجموع	**169.3**	**264.2**	**160.4**	**217.9**	**173.2**	**208.0**
المتوسط	**21.16**	**33.03**	**20.05**	**27.23**	**21.65**	**26.0**

إجمالى قيم X = 502.9 بمتوسط 20.95 كج وإجمالى قيم Y = 690.1 بمتوسط 28.75 كج.

وتحلل هذه البيانات مرة للمتغير Y ومرة للمتغير X ومرة ثالثة للمتغيرين X ، Y معاً كما يلى:

المتغير X:

مجموع المربعات الكلية

$$= (24.5)^2 + (22.5)^2 + \cdots + (15.9)^2 - \frac{(502.9)^2}{24} = 296.060$$

مجموع المربعات بين المعاملات

$$= \frac{(169.3)^2 + (160.4)^2 + (173.2)^2}{8} - \frac{(502.9)^2}{24} = 10.761$$

داخل المعاملات $= 296.060 - 10.761 = 285.299$

المتغير Y :

مجموع المربعات الكلية

$$= (38.5)^2 + (33.4)^2 + \cdots + (17)^2 - \frac{(690.1)^2}{24} = 1561.80$$

مجموع المربعات بين المعاملات

$$= \frac{(264.2)^2 + (217.9)^2 + (208)^2}{8} - \frac{(690.1)^2}{24} = 225.006$$

داخل المعاملات $= 1561.18 - 225.006 = 1336.174$

المتغيرين X ، Y معاً

هنا يكمن الجديد فى هذا الجزء. والجديد هنا بسيط جداً حيث إنه بدلا من تربيع كل من X أو Y بمفردها فإن أحدهما يضرب فى الآخر للحصول على التغاير بينهما وتحليله كما يلى:

مجموع حاصل الضرب الكلى

$$= (24.5)\ (38.5) + (22.5)\ (33.4) + \cdots + (15.9)\ (17) - \frac{(690.1)\ (502.9)}{24}$$

$$= 15043.95 - 14460.47 = 583.48$$

مجموع حاصل الضرب بين المعاملات

$$= \frac{(169.3)\ (264.2) - (160.4)\ (217.9) - (173.2)\ (208.0)}{8} - \frac{(690.1)(502.9)}{24}$$

$$= 2.757$$

مجموع حاصل الضرب داخل المعاملات $= 5843.48 - 2.757 = 580.723$

وجدير بالذكر أنه بينما لا يمكن لتباينات كل من X و Y أن تأخذ قيماً سالبة، فإن التغاير بينهما ممكن أن يأخذ قيماً سالبة حسب نوع العلاقة بينهما. ويمكن تلخيص النتائج المتحصل عليها حتى الآن فى جدول ١٥-١٤. ومن هذا الجدول يلاحظ:

جدول ١٥-١٤ تحليل التغاير فى تجربة تسمين الحملان

SOV	df	مجموع المربعات أو حاصل الضرب SS or SCP			التباين الراجع إلى الاعتماد	المتبقى بعد استقطاع التباين الراجع إلى الاعتماد		
		X	Y	XY		df	SS	MS
بين المعاملات	2	10.761	225.006	2.757				
داخل المعاملات	21	285.299	1336.174	580.723	1182.055	20	154.118	7.706
الكلى	**23**	**296.060**	**1561.180**	**583.480**	**1149.932**	**22**	**411.248**	
لاختبار المعاملات						2	257.130	128.565**

١- إذا اختبرت معنوية الفروق بين متوسطات المعاملات فى الوزن الابتدائى (X) فإنها لا تختلف عن الصفر وهذا متوقع نتيجة للتوزيع العشوائى للحملان على المعاملات.

٢- مع وجود فروق مشاهدة فى الوزن النهائى (Y) بين المعاملات (33، 27.2، 26 كج) إلا أن اختبار F بغض النظر عن التغاير بين X ، Y عبارة عن:

$$F = \frac{225.006/2}{1336.174/21} = 1.77$$

وهى قيمة غير معنوية. وهذا يوضح أن التجربة قاصرة على أن تبين معنوية الفروق الموجودة.

٣- هناك ارتباط واضح بين الوزن الابتدائى X والوزن النهائى Y على مستوى البيانات كلها بغض النظر عن المعاملات. ومعامل الارتباط هذا مقداره:

$$r_{xy} = \frac{583.48}{\sqrt{(296.06)(1561.18)}} = 0.86$$

وعليه يمكن القول أن هناك تباينات فى Y مرتبطة مع X وهذه التباينات تتبقى فى لخطأ التجريبى أن لم تزل بأسلوب معين. هذا الأسلوب هو التغاير حيث يستقطع من هذا الخطأ الجزء الذى يرجع إلى كون Y مرتبطة مع X وبالتالى متأثرة بها. وقد سبق فى الباب العاشر بيان كيفية حساب التباين فى Y الراجع إلى اعتماده أو ارتداده على متغير آخر X بواسطة القيمة $(\sum xy)^2 / \sum x^2$.

وعليه فالتباين فى Y النقى من تأثير X عليها هو التباين فى Y بعد طرح التباين الراجع إلى تأثير X. أى أن مجموع مربعات خطأ Y بعد استقطاع تأثير X عبارة عن

$$= 1336.174 - \frac{(580.723)^2}{285.299} = 154.118$$

وهذا له 20 درجة حرية حيث استقطعت درجة حرية واحدة للارتداد.

أما لحساب التباين بين متوسطات المعاملات فى Y بعد استقطاع تأثير X عليها فإن سطرى المعاملة والخطأ يجمعان. وفى هذا التصميم يعطى حاصل جمعهما سطر الكلى. ويستقطع الجزء الراجع إلى اعتماد أو ارتداد Y على X من التباين فى Y أى:

$$=1561.18-\frac{(583.48)^2}{296.06}=411.248$$

وهذا له 22 درجة حرية حيث استقطعت درجة حرية واحدة من الكلى (23) نتيجة للارتداد.

ويحسب مجموع المربعات بين المعاملات بعد إزالة أثر X على Y عن طريق $411.248 - 257.13 = 154.118$ وله درجتان حرية بمتوسط مربعات قدره 128.565.

ولاختبار فرض العدم أن متوسطات المعاملات فى الوزن النهائى متساوية:

$$F=\frac{128.565}{7.706}=16.684$$

بدرجات حرية 2، 20 . وعليه يرفض فرض العدم حيث إن قيمة F الجدولية المناسبة هى 5.85 عند مستوى معنوية 1%. وبهذا فإن استخدام تحليل التغاير زاد من مقدرة التجربة على إعلان معنوية الفروق إذا وجدت وهذا يعنى أن قوة التجربة زادت.

وجدير بالذكر أنه إذا أريد حساب علاقات معينة بين X، Y فإنها تقدر أساساً من السطر الخاص بالخطأ. فمثلا إذا أريد حساب الارتباط بين الوزن الابتدائى X والوزن النهائى Y فإنه

$$r_{xy}=\frac{580.723}{\sqrt{(285.299)(1336.174}}=0.94$$

بينما اعتماد Y على X فإنه

$$b_{y.x}=\frac{580.723}{285.299}=2.04\ \text{kg/kg}$$

والنموذج الإحصائى لمثل هذا التصميم (تصميم تام العشوائية أو تقسيم أحادى الاتجاه) دون اعتبار التغاير هو

$$Y_{ij} = \mu - t_i - e_{ij}$$

أما النموذج فى حالة اعتبار التغاير يكون

$$Y_{ij} = \mu - t_i - bx_{ij} - e_{ij}$$

حيث تمثل b معامل ارتداد Y على X_{ij} وتمثل x_{ij} انحراف الوزن الابتدائى للحمل من المتوسط للوزن الابتدائى. والافتراضات الخاصة بتحليل التغاير هى نفسها تلك الافتراضات الخاصة بكل من تحليل التباين وتحليل الانحدار والتى سبق الإشارة إليهما.

ويمكن فصل المتوسطات عن بعضها بالطرق سابقة الشرح ولكن مع اعتبار أن كل مقارنة سوف يكون لها خطأ قياسى مختلف نتيجة لاختلاف متوسطها فى المتغير X. فمثلا الخطأ القياسى للفرق بين المعاملتين أ ، ب هو الخطأ القياسى بين متوسطى الوزن النهائى بعد التعديل للوزن الابتدائى كالتالى:

الوزن المعدل $= \overline{Y}_i - b_{yx}(\overline{X}_i - \overline{X})$

الوزن المعدل للمعاملة أ $= 33.03 - (2.04)(21.16 - 20.95) = 32.60$ kg

الوزن المعدل للمعاملة ب $= 27.23 - (2.04)(20.05 - 20.95) = 29.07$ kg

الوزن المعدل للمعاملة ج $= 26.00 - (2.04)(21.65 - 20.95) = 24.57$ kg

والفروق بين هذه القيم هى فروق خالية من أثر الوزن الابتدائى.

وتباين الفرق بين الوزنين المعدلين للمعاملتين أ ، ب يساوى

$$= \sqrt{\sigma_e^2\left[\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{\sum x^2}(\overline{X}_{1.} - \overline{X}_{2.})^2\right]}$$

$$= \sqrt{7.706\left[\frac{1}{8} + \frac{1}{8} + \frac{1}{285.299}(21.16 - 20.05)^2\right]} = \sqrt{1.962} = 1.4$$

وهكذا لبقية الفروق وعلى أساس هذه القيم تقاس معنوية الفروق أما بدنكن أو توكى أو خلافة.

وفى هذا المثال اعتبر أن هناك معامل ارتداد b_{yx} واحد يمثل القيم فى المعاملات الثلاثة، ولكن يمكن افتراض أن هناك ثلاث معاملات ارتداد مختلفة واحدة لكل معاملة وتكملة التحليل على هذا النحو.

مثال ١٥-١٥

حل مثال ١٥-١٤ باستخدام برنامج SAS

```
DATA ANLCOV1;
INPUT TRT $ X Y @@;
CARDS;
A 24.5 38.5 A 22.5 33.4 A 25.5 38 A 18.4 32.3 A 16.9 24.4
A 19.1 28 A 23.2 42.5 A 19.2 27.1 B 25.6 40.5 B 16.5 19.1
B 18.5 22.3 B 16.4 20 B 20.4 25.1 B 27.4 42.2 B 14.4 15.1
B 21.2 33.6 C 22.6 29.9 C 25.8 38 C 24.5 29 C 21.9 24.8
C 18.9 19.2 C 20.8 21.7 C 22.8 28.4 C 15.9 17
PROC GLM;
CLASS TRT;
MODEL X = TRT/SS3;
PROC GLM;
CLASS TRT;
MODEL Y = TRT/SS3;
PROC GLM;
CLASS TRT;
MODEL Y = X TRT/SOLUTION;
RUN;
```

لاحظ:

استخدمت طريقة PROC GLM للحصول على ثلاثة تحليلات منفصلة لكل من X بمفردها فى نموذج، Y بمفردها فى نموذج ثم Y مع وضع X كتغاير. وفى الحالة الأخيرة لا توضع X فى الـ CLASS ويوضع مع النموذج اختيار SOLUTION للحصول على تحليل التغاير وتقدير معامل الانحدار. كما يلاحظ طريقة وضع الثلاثة نماذج حيث أنه لا يمكن استخدام PROC GLM إلا مرة واحدة فقط لكل نموذج.

نتائج التحليل:

The GLM Procedure
Class Level Information

Class	Levels	Values
TRT	3	A B C

Number of observations 24

Dependent Variable: X

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	10.7608333	5.3804167	0.40	0.6779
Error	21	285.2987500	13.5856548		
Corrected Total	23	296.0595833			

R-Square	Coeff Var	Root MSE	X Mean
0.036347	17.59016	3.685872	20.95417

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRT	2	10.76083333	5.38041667	0.40	0.6779

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	225.005833	112.502917	1.77	0.1951
Error	21	1336.173750	63.627321		
Corrected Total	23	1561.179583			

R-Square	Coeff Var	Root MSE	Y Mean
0.144126	27.74093	7.976674	28.75417

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRT	2	225.0058333	112.5029167	1.77	0.1951

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1407.060161	469.020054	60.86	<.0001
Error	20	154.119423	7.705971		
Corrected Total	23	1561.179583			

R-Square	Coeff Var	Root MSE	Y Mean
0.901280	9.654125	2.775963	28.75417

Source	DF	Type I SS	Mean Square	F Value	Pr > F
X	1	1149.932120	1149.932120	149.23	<.0001
TRT	2	257.128041	128.564021	16.68	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
X	1	1182.054327	1182.054327	153.39	<.0001
TRT	2	257.128041	128.564021	16.68	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	-18.06833933 B	3.69100615	-4.90	<.0001
X	2.03548911	0.16434775	12.39	<.0001
TRT A	8.01730094 B	1.39029203	5.77	<.0001
TRT B	4.49428258 B	1.41267082	3.18	0.0047
TRT C	0.00000000 B	.	.	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

لاحظ أن النموذج الأول أعطى تحليل التباين للمتغير X والنموذج الثانى أعطى تحليل التباين للمتغير Y وكلا النموذجين لم يأخذا فى الاعتبار العلاقة بين المتغيرين. أما النموذج الثالث فقد أخذ فى الاعتبار العلاقة بين المتغيرين وأعطى نوعين من مجموع المربعات، الأول هو SS Type I (1149.932) والذى يمثل التباين الكلى الراجع إلي اعتماد Y على X أى $[(\sum xy)^2 / \sum x^2]$. أما الثانى فهو SS Type III (1182.06) والذى يمثل التباين الراجع إلى اعتماد Y على X داخل المعاملات. الخطأ الناتج من النموذج الثالث (154.119) يمثل بعد استقطاع التباين الراجع إلى الاعتماد، بمعنى أنه مصحح للتغاير، وله 20 درجة حرية. ومجموع المربعات الناتج من النموذج الثالث لبين المعاملات يمثل مجموع المربعات بين المعاملات بعد استقطاع التباين الراجع إلى الاعتماد.

مثال ١٥-١٦

التالى يمثل الوزن بالكيلوجرام والعمر باليوم لعجول وعجلات من سلالتى أبقار الهيرفورد و السانتاجيرترودس. المطلوب مقارنة أوزان العجول والعجلات وكل من السلالتين عند أعمار ثابتة (هذا المثال يمثل تحليل التغاير فى حالة تصميم القطاعات العشوائية أو التقسيم ثنائى الاتجاه).

	عجول		عجلات		المجموع	
	العمر	الوزن	العمر	الوزن		
	X	Y	X	Y	X	Y
الهيرفورد						
	375	373	359	352		
	369	355	380	330		
	367	355	365	348		
	357	366	379	350		
	353	310	356	289		
المجموع	**1821**	**1759**	**1839**	**1669**	**3660**	**3428**
السنتاجيرترودس						
	355	332	354	340		
	338	362	331	293		
	314	311	318	268		
	310	340	308	300		
	309	316	335	315		
المجموع	**1626**	**1661**	**1646**	**1516**	**3272**	**3177**
الكلى	**3447**	**3420**	**3485**	**3185**	**6932**	**6605**

أولاً: تحليل العمر (X)

مجموع المربعات الكلية

$$= (375)^2 + (369)^2 + \cdots + (335)^2 - \frac{(6932)^2}{20} = 11320.8$$

مجموع المربعات بين الجنسين

$$= \frac{(3447)^2 + (3485)^2}{10} - \frac{(6932)^2}{20} = 72.2$$

مجموع المربعات بين السلالتين

$$= \frac{(3660)^2 + (3272)^2}{10} - \frac{(6932)^2}{20} = 7527.2$$

مجموع مربعات التداخل بين السلالة والجنس

$$= \frac{(1821)^2 + (1839)^2 + (1626)^2 + (1646)^2}{5}$$

$$- \frac{(6932)^2}{20} - 72.2 - 7527.2 = 7599.6 - 7599.4 = 0.2$$

مجموع مربعات الخطأ $= 11320.8 - 72.2 - 7527.2 - 0.2 = 3721.2$

ثانياً: تحليل الوزن (Y)

مجموع المربعات الكلية

$$= (373)^2 + (355)^2 + \cdots + (315)^2 - \frac{(6605)^2}{20} = 15805.75$$

مجموع المربعات بين الجنسين

$$= \frac{(3420)^2 + (3185)^2}{10} - \frac{(6605)^2}{20} = 2761.25$$

مجموع المربعات بين السلالتين

$$= \frac{(3428)^2 + (3177)^2}{10} - \frac{(6605)^2}{20} = 3150.05$$

مجموع مربعات التداخل بين السلالة والجنس

$$= \frac{(1759)^2 + (1669)^2 + (1661)^2 + (1516)^2}{5}$$

$$- \frac{(6605)^2}{20} - 2761.25 - 3150.05 = 6062.55 - 5911.3 = 151.25$$

مجموع مربعات الخطأ

$$= 15805.75 - 2761.25 - 3150.05 - 151.25 = 9743.20$$

ثالثاً: تحليل تغاير X مع Y

مجموع حواصل الضرب الكلية

$$= (375)(373) + (369)(355) + \cdots + (335)(315) - \frac{(6932)(6605)}{20} = 7682.0$$

مجموع حواصل الضرب بين الجنسين

$$= \frac{(3447)(3420) - (3485)(3185)}{10} - \frac{(6932)(6605)}{20} = -446.5$$

مجموع حواصل الضرب بين السلالتين

$$= \frac{(3660)(3428) - (3272)(3177)}{10} - \frac{(6932)(6605)}{20} = 4869.4$$

مجموع حواصل الضرب للتداخل بين الجنس والسلالة

$$= \frac{(1821)(1759) + (1839)(1669) + (1626)(1661) + (1646)(1516)}{5}$$

$$- \frac{(6932)(6605)}{20} - (-446.5) - 4869.4 = -5.5$$

الخطأ $= 7682 - (-446.5) - (4869.4) - (-5.5) = 3264.6$

ويمكن تلخيص النتائج كما في جدول ١٥-١٥.

والتالى يبين كيفية الحصول على مجموع المربعات المتبقى بعد استقطاع الجزء الراجع إلى الاعتماد:

جدول ١٥-١٥ تحليل نتائج واختبارات المعنوية لتجربة الفرق بين جنس وسلالة أبقار اللحم (مثال ١٥-١٦)

SOV	df	مجموع المربعات أو حاصل الضرب SS or SCP			المتبقى فى تباين Y بعد استقطاع الجزء الراجع إلى الاعتماد		
		X	Y	XY	df	SS	MS
بين الجنسين	1	72.2	2761.25	-446.5			
بين السلالتين	1	7527.2	3150.05	4869.4			
الجنس x السلالة	1	0.2	151.25	-5.5			
الخطأ	16	3721.2	9743.20	3264.6	15	6879.17	
الكلى	**19**	**11320.8**	**15805.75**	**7682.0**			
لاختبار الجنسين	17	3793.4	12504.45	2818.1	16	10410.90	
فرق الجنسين معدل للعمر					1	3531.73	3531.73
لاختبار السلالتين	17	11248.4	12893.25	8134.0	16	7011.35	
فرق السلالتين معدل للعمر					1	132.18	132.18
لاختبار التداخل بين الجنس والسلالة	17	3721.4	9894.45	3259.1	16	7040.22	
فروق التداخل معدلة للعمر					1	161.05	161.05

مجموع المربعات المتبقى بعد استقطاع الجزء الراجع إلى الاعتماد للخطأ

$$= 9743.2 - \frac{(3264.6)^2}{3721.2} = 6879.17$$

مجموع مربعات المتبقى بعد استقطاع الجزء الراجع إلى الاعتماد لاختبار بين الجنسين

$$= 12504.45 - \frac{(2818.1)^2}{3793.4} = 10410.90$$

مجموع مربعات المتبقى بعد استقطاع الجزء الراجع إلى الاعتماد لاختبار بين السلالتين

$$= 12893.25 - \frac{(8134)^2}{11248.4} = 7011.30$$

مجموع مربعات المتبقى بعد استقطاع الجزء الراجع إلى الاعتماد لاختبار التداخل بين الجنس والسلالة

$$= 9894.45 - \frac{(3259.1)^2}{3721.4} = 7040.22$$

أما مجموع المربعات للفروق المعدلة للعمر يتحصل عليها كما يلى:

للجنسين

$$= 10410.90 - 6879.17 = 3531.73$$

للسلالتين

$$= 7011.35 - 6879.17 = 132.18$$

للتداخل

$$= 7040.22 - 6879.17 = 161.05$$

وتحسب F لكل مصدر تباين يراد اختباره بقسمة متوسط مربعات هذا المصدر معدلا للعمر على متوسط الخطأ معدل للعمر أيضاً.

مع تحيات د. سلام حسين الهلالي salamalhelali@yahoo.com

لاختبار معنوية الجنسين تكون قيمة F $= 7.70 = \frac{3531.73}{458.61}$، وللسلالتين هى $0.29 = \frac{132.18}{458.61}$ والتداخل بينهما هى $0.35 = \frac{161.05}{458.61}$

وواضح أن تلك التابعة للجنسين هى المعنوية فقط.

وإذا أريد حساب معامل ارتداد الوزن على العمر:

$$b_{yx} = \frac{3264.6}{3721.2} = 0.88 \text{ kg/day}$$

ومعامل الارتباط

$$r = 0.54$$

ويمكن تلخيص النتائج كما فى جدول ١٥-١٧.

وبنفس الطريقة يمكن تطبيق تحليل التغاير على التصميمات الأكثر تعقيداً.

مثال ١٥-١٧

حل مثال ١٥-١٦ باستخدام برنامج SAS

```
DATA ANLCOV2;
INPUT BREED $ SEX $ X Y @@;
CARDS;
H M 375 373 H M 369 355 H M 367 355 H M 357 366
H M 353 310 H F 359 352 H F 380 330 H F 365 348
H F 379 350 H F 356 289 S M 355 332 S M 338 362
S M 314 311 S M 310 340 S M 309 316 S F 354 340
S F 331 293 S F 318 268 S F 308 300 S F 335 315
PROC GLM;
CLASS SEX BREED;
MODEL X = SEX BREED SEX*BREED/SS3;
PROC GLM;
CLASS SEX BREED;
MODEL Y = SEX BREED SEX*BREED/SS3;
PROC GLM;
CLASS SEX BREED;
MODEL Y = X SEX BREED SEX*BREED/SOLUTION;
RUN;
```

نتاج التحليل

The GLM Procedure
Class Level Information

Class	Levels	Values
SEX	2	F M
BREED	2	H S

Number of observations 20

Dependent Variable: X

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	7599.60000	2533.20000	10.89	0.0004
Error	16	3721.20000	232.57500		
Corrected Total	19	11320.80000			

R-Square	Coeff Var	Root MSE	X Mean
0.671295	4.400003	15.25041	346.6000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SEX	1	72.200000	72.200000	0.31	0.5851
BREED	1	7527.200000	7527.200000	32.36	<.0001
SEX*BREED	1	0.200000	0.200000	0.00	0.9770

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	6062.55000	2020.85000	3.32	0.0467
Error	16	9743.20000	608.95000		
Corrected Total	19	15805.75000			

R-Square	Coeff Var	Root MSE	Y Mean
0.383566	7.472191	24.67691	330.2500

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SEX	1	2761.250000	2761.250000	4.53	0.0491
BREED	1	3150.050000	3150.050000	5.17	0.0371
SEX*BREED	1	151.250000	151.250000	0.25	0.6250

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	8926.57589	2231.64397	4.87	0.0102
Error	15	6879.17411	458.61161		
Corrected Total	19	15805.75000			

R-Square	Coeff Var	Root MSE	Y Mean
0.564768	6.484548	21.41522	330.2500

Source	DF	Type I SS	Mean Square	F Value	Pr > F
X	1	5212.805102	5212.805102	11.37	0.0042
SEX	1	3422.288065	3422.288065	7.46	0.0154
BREED	1	130.437179	130.437179	0.28	0.6016
SEX*BREED	1	161.045549	161.045549	0.35	0.5623

Source	DF	Type III SS	Mean Square	F Value	Pr > F
X	1	2864.025895	2864.025895	6.24	0.0246
SEX	1	3531.721930	3531.721930	7.70	0.0142
BREED	1	132.176603	132.176603	0.29	0.5992
SEX*BREED	1	161.045549	161.045549	0.35	0.5623

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	46.90280555 B	114.5656308	0.41	0.6880
X	0.87729765	0.3510597	2.50	0.0246
SEX F	-32.50919058 B	13.6167738	-2.39	0.0306
SEX M	0.00000000 B	.	.	.
BREED H	-14.61460819 B	19.2586901	-0.76	0.4597
BREED S	0.00000000 B	.	.	.
SEX*BREED F H	11.35091906 B	19.1548689	0.59	0.5623
SEX*BREED F S	0.00000000 B	.	.	.
SEX*BREED M H	0.00000000 B	.	.	.
SEX*BREED M S	0.00000000 B	.	.	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

تمارين الباب الخامس عشر

١٥-١ أخذت مجموعة من 4 كباش من كل مربى من أربعة من مربى أغنام الصوف وقيس وزن الجزة بالكيلوجرام ووجد كما يلى:

المربى الأول	4.3	4.1	4.6	3.4
المربى الثاني	5.0	8.5	4.0	4.5
المربى الثالث	4.7	5.0	5.0	6.5
المربى الرابع	4.0	6.0	5.6	4.8

أ- أكتب النموذج الإحصائى الذى يمثل التباينات فى هذه البيانات.

ب- أجر تحليل التباين واختبر ما إذا كان وزن الجزة يختلف من مربى لآخر.

ج- أجر اختبار LSD ، HSD ودنكن على المتوسطات.

١٥-٢ أجريت تجربة على أثر ٣ معاملات فى قطاعين على وزن العجول بالكيلوجرام عند عمر 100 يوماً، وكانت النتائج كما يلى:

	القطاع الأول	القطاع الثانى
المعاملة ١	87.3	81.2
	79.1	92.3
	75.4	79.1
	89.1	84.5
المعاملة ٢	69.5	72.8
	69.1	80.4
	75.9	73.6
	73.2	86.4
المعاملة ٣	74.1	74.5
	57.3	75.5
	65.9	65.4
	71.8	68.6

أ- أكتب النموذج الإحصائى الذي يصف هذه البيانات وحلل التباين طبقا للنموذج الإحصائى. ما نوع التصميم؟

ب- اختبر الفروض الإحصائية المختلفة.

ج- أجر اختبارات LSD ، HSD ودنكن على متوسطات المعاملات.

د- قسم مجموع المربعات بين المعاملات إلى مجموع المربعات بين المعاملة الأولى ومتوسط المعاملتين الثانية والثالثة – واختبر لمعنوية الفروق.

١٥-٣ البيانات التالية تمثل أوزان جزات الصفوف بالكيلوجرام فى أغنام الرامبوبية التى عوملت بأربع معاملات مختلفة فى أربع مزارع (الصفوف) وأربع طرق رعاية مختلفة (الأعمدة) حيث تدل الحروف على المعاملات المختلفة.

	رعاية ١	رعاية ٢	رعاية ٣	رعاية ٤
	أ	د	ج	ب
مزرعة ١	3.4	4.1	4.3	4.1
	3.5	4.6	3.4	3.8
	ب	أ	د	ج
مزرعة ٢	4.4	2.7	4.2	3.8
	3.7	2.7	5.3	4.1
	ج	ب	أ	د
مزرعة ٣	3.2	4.8	3.5	4.3
	3.8	3.8	3.3	4.6
	د	ج	ب	أ
مزرعة ٤	3.9	3.4	4.0	3.3
	3.7	4.3	4.2	3.8

أ- اكتب النموذج الإحصائى وحلل التباين لمصادرة المختلفة واختبر معنوية الفروض الإحصائية المناسبة. ما نوع التصميم؟

ب- إذا كانت المعاملات هى أربعة مستويات مختلفة من الإضافات بالعليقة أ: 5، ب : 20، ج : 10، د : 15 مجم/كج عليقة. قسم مجموع المربعات بين المعاملات إلى خطى وتربيعى وتكعيبى ثم اختبر معنوية كل منها. قدر

معامل الارتداد الخطى للزيادة فى الوزن على الإضافة بالعليقة مع حساب خطأ القياس.

٥٠-٤ يراد مقارنة أوزان طلائق من سلالتى السانتاجيرتيرودس والهيرفورد ولكن وجد أن هناك فروق فى أعمار هذه الطلائق كيف تجرى هذه المقارنة مع التصحيح لفروق العمر من البيانات التالية:

			سانتاجيرترودس			
العمر (يوم)	446	412	484	413	473	474
الوزن (كج)	537	488	501	448	491	453
			هيرفورد			
العمر (يوم)	432	468	488	416	396	448
الوزن (كج)	354	353	393	382	354	390

قدر معامل ارتداد الوزن على العمر من جدول تحليل التغاير واحسب متوسط وزن كل سلالة معدلا للعمر.

٥١-٥ البيانات التالية تمثل درجات 12 طالب تم توزيعهم عشوائياً بالتساوى على كل من مقررين دراسيين A ، B على مدى اختبارات دورية خلال الفصل الدراسى علماً بأن الست طلاب هم أنفسهم الذين تعددت عليهم الاختبارات. المطلوب اختبار هل هناك فرق فى الدرجات التى تحصل عليها الطلاب من مقرر إلى آخر؟ هل درجات الطلاب تتغير طبقاً لعدد الاختبارات التى يتعرض لها الطالب.

		الاختبار الأول	الاختبار الثانى	الاختبار الثالث
المقرر الأول A	الطالب ١	42	99	100
	الطالب ٢	25	100	71
	الطالب ٣	42	55	95
	الطالب ٤	66	82	83
	الطالب ٥	39	68	47
	الطالب ٦	33	67	83

		الاختبار الأول	الاختبار الثانى	الاختبار الثالث
المقرر الثانى B	الطالب ٧	98	55	100
	الطالب ٨	60	56	91
	الطالب ٩	13	52	30
	الطالب ١٠	42	17	57
	الطالب ١١	43	89	32
	الطالب ١٢	56	61	100

١٥-٦ التالى هى أوزان بالكيلوجرام وأعمار باليوم لعجول وعجلات تتبع ثلاث سلالات مختلفة من الأبقار فى معاملتين. والمطلوب اختبار فرق الأوزان بين العجول والعجلات وبين السلالات معدلة للعمر مع حساب معامل ارتداد الوزن على العمر واختبارات فروض العدم الخاصة بالجنس والسلالة والتداخل بينهما.

	معاملة ١		معاملة ٢	
	العمر	الوزن	العمر	الوزن
براهما	451	352	486	375
	471	326	430	343
	438	285	475	368
	459	328	471	377
سانتاجيرتيرودس	478	368	446	537
	460	352	412	488
	485	345	484	501
	480	315	413	448
شارولية	474	493	484	553
	410	403	457	474
	484	407	467	480
	448	470	448	536

ملحق أ

١- الأرقام العشوائية

٢- الانحراف المعيارى (σ) مقسوماً على المدى (R) لأحجام عينة مختلفة

٣- المساحة تحت التوزيع الطبيعى (المساحة من صفر إلى z)

٤- توزيع t (اختبار من طرفين)

٥- توزيع F

٦- توزيع مربع كاى χ^2

٧- القيم المعنوية لمعاملات الارتباط البسيط (r) والمتعدد (R)

٨- قيم z بدلالة r

٩- قيم r بدلالة z

١٠- توزيع الإحصاء F_{max}

١١- القيم الحرجة لاختبار Cochran لتجانس التباين

١٢- اختبار معنوية الالتواء (اختبار من طرف واحد)

١٣- اختبار معنوية التفرطح

١٤- تحوير مقلوب جيب الزاوية Arcsin

١٥- قيم توكى

١٦- قيم دنكن

١٧- معاملات الحدود المتعددة

١٨- تحديد العدد الأمثل لتحليل التباين (نموذج ثابت)

١٩- الحروف اليونانية

جدول ١ – الأرقام العشوائية

	00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
00	54463	22662	65905	70639	79365	67382	29085	69831	47058	08186
01	15389	85205	18850	39226	42249	90669	96325	23248	60933	26927
02	85941	40756	82414	02015	13858	78030	16269	65978	01385	15345
03	61149	69440	11286	88218	58925	03638	52862	62733	33451	77455
04	05219	81619	10651	67079	92511	59888	84502	72095	83463	75577
05	41417	98326	87719	92294	46614	50948	64886	20002	97365	30976
06	28357	94070	20652	35774	16249	75019	21145	05217	47286	76305
07	17783	00015	10806	83091	91530	36466	39981	62481	49177	75779
08	40950	84820	29881	85966	62800	70326	84740	62660	77379	90279
09	82995	64157	66164	41180	10089	41757	78258	96488	88629	37231
10	96754	17676	55659	44105	47361	34833	86679	23930	53249	27083
11	34357	88040	53364	71726	45690	66334	60332	22554	90600	71113
12	06318	37403	49927	57715	50423	67372	63116	48888	21505	80182
13	62111	52820	07243	79931	89292	84767	85693	73947	22278	11551
14	47534	09243	67879	00544	23410	12740	02540	54440	32949	13491
15	98614	75993	84460	62846	59844	14922	48730	73443	48167	34770
16	24856	03648	44898	09351	98795	18644	39765	71058	90368	44104
17	96887	12479	80621	66223	86085	78285	02432	53342	42846	94771
18	90801	21472	42815	77 408	37390	76766	52615	32141	30268	18106
19	55165	77312	83666	36028	28420	70219	81369	41943	47366	41067
20	75884	12952	84318	95108	72305	64620	91318	89872	45375	85436
21	16777	37116	58550	42958	21460	43910	01175	87894	81378	10620
22	46230	43877	80207	88877	89380	32992	91380	03164	98656	59337
23	42902	66892	46134	01432	94710	23474	20423	60137	60609	13119
24	81007	00333	39693	28039	10154	95425	39220	19774	31782	49037
25	68089	01122	51111	72373	06902	74373	96199	97017	41273	21546
26	20411	67081	89950	16944	93054	87687	96693	87236	77054	33848
27	58212	13160	06468	15718	82627	76999	05999	58680	96739	63700
28	70577	42866	24969	61210	76046	67699	42054	12696	93758	03283
29	94522	74358	71659	62038	79643	79169	44741	05437	39038	13163
30	42626	86819	85651	88678	17401	03252	99547	32404	17918	62880
31	16051	33763	57194	16752	54450	19031	58580	47629	54132	60631
32	08244	27647	33851	44705	94211	46716	11738	55784	95374	72655
33	59497	04392	09419	89964	51211	04894	72882	17805	21896	83864
34	97155	13428	40293	09985	58434	01412	69124	82171	59058	82859
35	98409	66162	95763	47420	20792	61527	20441	39435	11859	41567
36	45476	84882	65109	96597	25930	66790	65706	61203	53634	22557
37	89300	69700	50741	30329	11658	23166	05400	66669	48708	03887
38	50051	95137	91631	66315	91428	12275	24816	68091	71710	33258

تكملة ...

... تابع جدول ١ – الأرقام العشوائية

	00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
39	31753	85178	31310	89642	98364	02306	24617	09609	83942	22716
40	79152	53829	77250	20190	56535	18760	69942	77448	33278	48805
41	44560	38750	83635	56540	64900	42912	13953	79149	18710	68618
42	68328	83378	63369	71381	39564	05615	42451	64559	97501	65747
42	68328	83378	63369	71381	39564	05615	42451	64559	97501	65747
43	46939	38689	58625	08342	30459	85863	20781	09284	26333	91777
44	83544	86141	15707	96256	23068	13782	08467	89469	93842	55349
45	91621	00881	04900	54224	46177	55309	17852	27491	89415	23466
46	91896	67126	04151	03795	59077	11848	12630	98375	52068	60142
47	55751	62515	21108	80830	02263	29303	37204	96926	30506	09808
48	85156	87689	95493	88842	00664	55017	55539	17771	69448	87530
49	07521	56898	12236	60277	39102	62315	12239	07105	11844	01117
50	64249	63664	39652	40646	97306	31741	07294	84149	46797	82487
51	26538	44249	04050	48174	65570	44072	40192	51153	11397	58212
52	05845	00512	78630	55328	18116	69296	91705	86224	29503	57071
53	74897	68373	67359	51014	33510	83048	17056	72506	82949	54600
54	20872	54570	35017	88132	25730	22626	86723	91691	13191	77212
55	31432	96156	89177	75541	81355	24480	77243	76690	42507	84362
56	66890	61505	01240	00660	05873	13568	76082	79172	57913	93448
57	41894	57790	79970	33106	86904	48119	52503	24130	72824	21627
58	11303	87118	81471	52936	08555	28420	49416	44448	04269	27029
59	54374	57325	16947	45356	78371	10563	97191	53798	12693	27928
60	64852	34421	61046	90849	13966	39810	42699	21753	76192	10508
61	16309	20384	09491	91588	97720	89846	30376	76970	23063	35894
62	42587	37065	24526	72602	57589	98131	37292	05967	26002	51945
63	40177	98590	97161	41682	84533	67588	62036	49967	01990	72308
64	82309	76128	93965	26743	24141	04838	40254	26065	07938	76236
65	79788	68243	59732	04257	27084	14743	17520	95401	55811	76099
66	40538	79000	89559	25026	42274	23489	34502	75508	06059	86682
67	64016	73598	18609	73150	62463	33102	45205	87440	96767	67042
68	49767	12691	17903	93871	99721	79109	09425	26904	07419	76013
69	76974	55108	29795	08404	82684	00497	51126	79935	57450	55671
70	23854	08480	85983	96025	50117	64610	99425	62291	86943	21541

تكملة ...

مع تحيات د. سلام حسين الهلالي salamalhelali@yahoo.com

... تابع جدول ١ – الأرقام العشوائية

	00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
71	68973	70551	25098	78033	98573	79848	31778	29555	61446	23037
72	36444	93600	65350	14971	25325	00427	52073	64280	18847	24768
73	03003	87800	07391	11594	21196	00781	32550	57158	58887	73041
74	17540	26188	36647	78386	04558	61463	57842	90382	77019	24210
75	38916	55809	47982	41968	69760	79422	80154	91486	19180	15100
76	64288	19843	69122	42502	48508	28820	59933	72998	99942	10515
77	86809	51564	38040	39418	49915	19000	58050	16899	79952	57849
78	99800	99566	14742	05028	30033	94889	53381	23656	75787	59223
79	92345	31890	95712	08279	91794	94068	49337	88674	35355	12267
80	90363	65162	32245	82279	79256	80834	06088	99462	56705	06118
81	64437	32242	48431	04835	39070	59702	31508	60935	22390	52246
82	91714	53662	28373	34333	55791	74758	51144	18827	10704	76803
83	20902	17646	31391	31459	33315	03444	55743	74701	58851	27427
84	12217	86007	70371	52281	14510	76094	96579	54853	78339	20839
85	45177	02863	42307	53571	22532	74921	17735	42201	80540	54721
86	28325	90814	08804	52746	47913	54577	47525	77705	95330	21866
87	29019	28776	56116	54791	64604	08815	46049	71186	34650	14994
88	84979	81353	56219	67062	26146	82567	33122	14124	46240	92973
89	50371	26347	48513	63915	11158	25563	91915	18431	92978	11591
90	53422	06825	69711	67950	64716	18003	49581	45378	99878	61130
91	67453	35651	89316	41620	32048	70225	47597	33137	31443	51445
92	07294	85353	74819	23445	68237	07202	99515	62282	53809	26685
93	79544	00302	45338	16015	66613	88968	14595	63836	77716	79596
94	64144	85442	82060	46471	24162	39500	87351	36637	42833	71875
95	90919	11883	58318	00042	52402	28210	34075	33272	00840	73268
96	06670	57353	86275	92276	77591	46924	60839	55437	03183	13191
97	36634	93976	52062	83678	41256	60948	18685	48992	19462	96062
98	75101	72891	85745	67106	26010	62107	60885	37503	55461	71213
99	05112	71222	72654	51583	05228	62056	57390	42746	39272	96659

تكملة ...

... تابع جدول ١ – الأرقام العشوائية

	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99
00	59391	58030	52098	82718	87024	82848	04190	96574	90464	29065
01	99567	76364	77204	04615	27062	96621	43918	01896	83991	51141
02	10363	97518	51400	25670	98342	61891	27101	37855	06235	33316
03	86859	19558	64432	16706	99612	59798	32803	67708	15297	28612
04	11258	24591	36863	55368	31721	94335	34936	02566	80972	08188
05	95068	88628	35911	14530	33020	80428	39936	31855	34334	64865
06	54463	47237	73800	91017	36239	71824	83671	39892	60518	37092
07	16874	62677	57412	13215	31389	62233	80827	73917	82802	84420
08	92494	63157	76593	91316	03505	72389	96363	52887	01087	66091
09	15669	56689	35682	40844	53256	81872	35213	09840	34471	74441
10	99116	75486	84989	23476	52967	67104	39495	39100	17217	74073
11	15696	10703	65178	90637	63110	17622	53988	71087	84148	11670
12	97720	15369	51269	69620	03388	13699	33423	67453	43269	56720
13	11666	13841	71681	98000	35979	39719	81899	07449	47985	46967
14	71628	73130	78783	75691	41632	09847	61547	18707	85489	69944
15	40501	51089	99943	91843	41995	88931	73631	69361	05375	15417
16	22518	55576	98215	82068	10798	86211	36584	67466	69373	40054
17	75112	30485	62173	02132	14878	92879	22281	16783	86352	00077
18	80327	02671	98191	84342	90813	49268	95441	15496	20168	09271
19	60251	45548	02146	05597	48228	81366	34598	72856	66762	17002
20	57430	82270	10421	00540	43648	75888	66049	21511	47676	33444
21	73528	39559	34434	88596	54086	71693	43132	14414	79949	85193
22	25991	65959	70769	64721	86413	33475	42740	06175	82758	66248
23	78388	16638	09134	59980	63806	48472	39318	35434	24057	74739
24	12477	09965	96657	57994	59439	76330	24596	77515	09577	91871
25	83266	32883	42451	15579	38155	29793	40914	65990	16255	17777
26	76970	80876	10237	39515	79152	74798	39357	09054	73579	92359
27	37074	65198	44785	68624	98336	84481	97610	78735	46703	98265
28	83712	06514	30101	78295	54656	85417	43189	60048	72781	72606
29	20287	56862	69727	94443	64936	08366	27227	05158	50326	59566
30	74261	32592	86538	27041	65172	85532	07571	80609	39285	65340
31	64081	49863	08478	96001	18888	14810	70545	89755	59064	07210
32	05617	75818	47750	67814	29575	10526	66192	44464	27058	40467
33	26793	74951	95466	74307	13330	42664	85515	20632	05497	33625
34	65988	72850	48737	54719	52056	01596	03845	35067	03134	70322
35	27366	42271	44300	73399	21105	03280	73457	43093	05192	48657
36	56760	10909	98147	34736	33863	95256	12731	66598	50771	83665

تكملة ...

... تابع جدول ١ – الأرقام العشوائية

	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99
37	72880	43338	93643	58904	59543	23943	11231	83268	65938	81581
38	77888	38100	03062	58103	47961	83841	25878	23746	55903	44115
39	28440	07819	21580	51459	47971	29882	13990	29226	23608	15873
40	63525	94441	77033	12147	51054	49955	58312	76923	96071	05813
41	47606	93410	16359	89033	89696	47231	64498	31776	05383	39902
42	52669	45030	96279	14709	52372	87832	02735	50803	72744	88208
43	16738	60159	07425	62369	07515	82721	37875	71153	21315	00132
44	59348	11695	45751	15865	74739	05572	32688	20271	65128	14551
45	12900	71775	29845	60774	94924	21810	38636	33717	67598	82521
46	75086	23537	49939	33595	13484	97588	28617	17979	70749	35234
47	99495	51434	29181	09993	38190	42553	68922	52125	91077	40197
48	26075	31671	45386	36583	93459	48599	52022	41330	60651	91321
49	13636	93596	23377	51133	95126	61496	42474	45141	46660	42338
50	32847	31282	03345	89593	69214	70381	78285	20054	91018	16742
51	16916	00041	30236	55023	14253	76582	12092	86533	92426	37655
52	66176	34037	21005	27137	03193	48970	64625	22394	39622	79085
53	46299	13335	12180	16861	38043	59292	62675	63631	37020	78195
54	22847	47839	45385	23289	47526	54098	45683	55849	51575	64689
55	41851	54160	92320	69936	34803	92479	33399	71160	64777	83378
56	28444	59497	91586	95917	68553	28639	06455	34174	11130	91994
57	47520	62378	98855	83174	13088	16561	68559	26679	06238	51254
58	34978	63271	13142	82681	05271	08822	06490	44984	49307	61717
59	37404	80416	69035	92980	49486	74378	75610	74976	70056	15478
60	32400	65482	52099	53676	74648	94148	65095	69597	52771	71551
61	89262	86332	51718	70663	11623	29834	79820	73002	84886	03591
62	86866	09127	98021	03871	27789	58444	44832	36505	40672	30180
63	90814	14833	08759	74645	05046	94056	99094	65091	32663	73040
64	19192	82756	20553	58446	55376	88914	75096	26119	83898	43816
65	77585	52593	56612	95766	10019	29531	73064	20953	53523	58136
66	23757	16364	05096	03192	62386	45389	85332	18877	55710	96459
67	45989	96257	23850	26216	23309	21526	07425	50254	19455	29315
68	92970	94243	07316	41467	64837	52406	25225	51553	31220	14032
69	74346	59596	40088	98176	17896	86900	20249	77753	19099	48885
70	87646	41309	27636	45153	29988	94770	07255	70908	05340	99751
71	50099	71038	45146	06146	55211	99429	43169	66259	97786	59180
72	10127	46900	64984	75348	04115	33624	68774	60013	35515	62556
73	67995	81977	18984	64091	02785	27762	42529	97144	80407	64524
74	26304	80217	84934	82657	69291	35397	98714	35104	08187	48109
75	81994	41070	56642	64091	31229	02595	13513	45148	78722	30144
76	59537	34662	79631	89403	65212	09975	06118	86197	58208	16162

تكملة ...

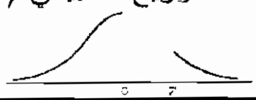
... تابع جدول ١ – الأرقام العشوائية

	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99
77	51228	10937	62396	81460	47331	91403	95007	06047	16846	64809
78	31089	37995	29577	07828	42272	54016	21950	86192	99046	84864
79	38207	97938	93459	75174	79460	55436	57206	87644	21296	43393
80	88666	31142	09474	89712	63153	62333	42212	06140	42594	43671
81	53365	56134	67582	92557	89520	33452	05134	70628	27612	33738
82	89807	74530	38004	90102	11693	90257	05500	79920	62700	43325
83	18682	81038	85662	90915	91631	22223	91588	80774	07716	12548
84	63571	32579	63942	25371	09234	94592	98475	76884	37635	33608
85	68927	56492	67799	95398	77642	54913	91583	08421	81450	76229
86	56401	63186	39389	88798	31356	89235	97036	32341	33292	73757
87	24333	95603	02359	72942	46287	95382	08452	62862	97869	71775
88	17025	84202	95199	62272	06366	16175	97577	99304	41587	03686
89	02804	08253	52133	20224	68034	50865	57868	22343	55111	03607
90	08298	03879	20995	19850	73090	13191	18963	82244	78479	99121
91	59883	01785	82403	96062	03785	03488	12970	64896	38336	30030
92	46982	06682	62864	91837	74021	89094	39952	64158	79614	78235
93	31121	47266	07661	02051	67599	24471	69843	83696	71402	76287
94	97867	56641	63416	17577	30161	87320	37752	73276	48969	41915
95	57364	86746	08415	14621	49430	22311	15836	72492	49372	44103
96	09559	26263	69511	28064	75999	44540	13337	10918	79846	54809
97	53873	55571	00608	42661	91332	63956	74087	59008	47493	99581
98	35531	19162	86406	05299	77511	24311	57257	22826	77555	05941
99	28229	88629	25695	94932	30721	16197	78742	34974	97528	45447

جدول ٢ – الانحراف المعيارى (σ) مقسوماً على المدى (R) لأحجام عينة مختلفة

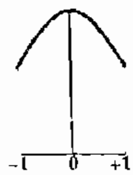
n	σ / R	n	σ / R
2	0.886	17	0.279
3	0.591	18	0.275
4	0.486	19	0.271
5	0.430	20	0.268
6	0.395	30	0.244
7	0.370	50	0.222
8	0.351	70	0.208
9	0.337	100	0.200
10	0.325	150	0.189
11	0.315	200	0.182
12	0.307	300	0.172
13	0.300	400	0.169
14	0.294	500	0.164
15	0.288	700	0.159
16	0.283	1000	0.154

جدول ٣ – المساحات تحت التوزيع الطبيعي (المساحة من صفر إلي z)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.000	0.004	0.008	0.012	0.016	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4260	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000									

جدول ٤ – توزيع t (اختبار من طرفين)



df الخطأ	احتمال الحصول على قيمة أكبر (مع إهمال الإشارة)								
	0.500	**0.400**	**0.200**	**0.100**	**0.050**	**0.025**	**0.010**	**0.005**	**0.001**
1	1.000	1.376	3.078	6.314	12.706	25.452	63.657		
2	0.816	1.061	1.886	2.920	4.303	6.205	9.925	14.089	31.598
3	.765	0.978	1.638	2.353	3.182	4.176	5.841	7.453	12.941
4	.741	.941	1.533	2.132	2.776	3.495	4.604	5.598	8.610
5	.727	.920	1.476	2.015	2.571	3.163	4.032	4.773	6.859
6	.718	.906	1.440	1.943	2.447	2.969	3.707	4.317	5.959
7	.711	.896	1.415	1.895	2.365	2.841	3.499	4.029	5.405
8	.706	.889	1.397	1.860	2.306	2.752	3.355	3.832	5.041
9	.703	.883	1.383	1.833	2.262	2.685	3.250	3.690	4.781
10	.700	.879	1.372	1.812	2.228	2.634	3.169	3.581	4.587
11	.697	.876	1.363	1.796	2.201	2.593	3.106	3.497	4.437
12	.695	.873	1.356	1.782	2.179	2.560	3.055	3.428	4.318
13	.694	.870	1.350	1.771	2.160	2.533	3.012	3.372	4.221
14	.692	.868	1.345	1.761	2.145	2.510	2.977	3.326	4.140
15	.691	.866	1.341	1.753	2.131	2.490	2.947	3.286	4.073
16	.690	.865	1.337	1.746	2.120	2.473	2.921	3.252	4.015
17	.689	.863	1.333	1.740	2.110	2.458	2.898	3.222	3.965
18	.688	.862	1.330	1.734	2.101	2.445	2.878	3.197	3.922
19	.688	.861	1.328	1.729	2.093	2.433	2.861	3.174	3.883
20	.687	.860	1.325	1.725	2.086	2.423	2.845	3.153	3.850
21	.686	.859	1.323	1.721	2.080	2.414	2.831	3.135	3.819
22	.686	.858	1.321	1.717	2.074	2.406	2.819	3.119	3.792
23	.685	.858	1.319	1.714	2.069	2.398	2.807	3.104	3.767
24	.685	.857	1.318	1.711	2.064	2.391	2.797	3.090	3.745
25	.684	.856	1.316	1.708	2.060	2.385	2.787	3.078	3.725
26	.684	.856	1.315	1.706	2.056	2.379	2.779	3.067	3.707
27	.684	.855	1.314	1.703	2.052	2.373	2.771	3.056	3.690
28	.683	.855	1.313	1.701	2.048	2.368	2.763	3.047	3.674
29	.683	.854	1.311	1.699	2.045	2.364	2.756	3.038	3.659

تكملة ...

... تابع جدول ٤ - توزيع t (اختبار من طرفين)

df الخطأ	احتمال الحصول على قيمة أكبر (مع إهمال الإشارة)								
	0.500	**0.400**	**0.200**	**0.100**	**0.050**	**0.025**	**0.010**	**0.005**	**0.001**
30	.683	.854	1.310	1.697	2.042	2.360	2.750	3.030	3.646
35	.682	.852	1.306	1.690	2.030	2.342	2.724	2.996	3.591
40	.681	.851	1.303	1.684	2.021	2.329	2.704	2.971	3.551
45	.680	.850	1.301	1.680	2.014	2.319	2.690	2.952	3.520
50	.680	.849	1.299	1.676	2.008	2.310	2.678	2.937	3.496
55	.679	.849	1.297	1.673	2.004	2.304	2.669	2.925	3.476
60	.679	.848	1.296	1.671	2.000	2.299	2.660	2.915	3.460
70	.678	.847	1.294	1.667	1.994	2.290	2.648	2.899	3.435
80	.678	.847	1.293	1.665	1.989	2.284	2.638	2.887	3.416
90	.678	.846	1.291	1.662	1.986	2.279	2.631	2.878	3.402
100	.677	.846	1.290	1.661	1.982	2.276	2.625	2.871	3.390
120	.677	.845	1.289	1.658	1.980	2.270	2.617	2.860	3.373
∝	.675	.842	1.282	1.645	1.960	2.241	2.576	2.807	3.291

جدول ٥ – توزيع F بمستوى معنوية
5% (القيمة العلوية)، 1% (القيمة السفلية)

df المقام	درجات حرية df البسط											
	1	2	3	4	5	6	7	8	9	10	11	12
1	161 4,052	200 4,999	216 5,403	225 5,625	230 5,764	234 5,859	237 5,928	239 5,981	241 6.022	242 6,056	243 6,082	244 6,106
2	18.51 98.49	19.00 99.00	19.16 99.17	19.25 99.25	19.30 99.30	19.33 99.33	19.36 99.36	19.37 99.37	19.38 99.39	19.39 99.40	19.40 99.41	19.41 99.42
3	10.13 34.12	9.55 30.82	9.28 29.46	9.12 28.71	9.01 28.24	8.94 27.91	8.88 27.67	8.84 27.49	8.81 27.34	8.78 27.23	8.76 27.13	8.74 27.05
4	7.71 21.20	6.94 18.00	6.59 16.69	6.3 15.98	6.26 15.52	6.16 15.21	6.09 14.98	6.04 14.80	6.00 14.66	5.96 14.54	5.93 14.45	5.91 14.37
5	6.61 16.26	5.79 13.27	5.4 12.06	15.1 11.39	5.05 10.97	4.95 10.67	4.88 10.45	4.82 10.29	4.78 10.15	4.74 10.05	4.70 9.96	4.68 9.89
6	5.99 13.74	5.14 10.92	4.76 9.78	4.53 9.15	4.39 8.75	4.28 8.47	4.21 8.26	4.15 8.10	4.10 7.98	4.06 7.87	4.03 7.79	4.00 7.72
7	5.59 12.25	4.74 9.55	4.35 8.45	4.12 7.85	3.97 7.46	3.87 7.19	3.79 7.00	3.73 6.84	3.68 6.71	3.63 6.62	3.60 6.54	3.57 6.47
8	5.32 11.26	4.46 8.65	4.07 7.59	3.84 7.01	3.69 6.63	3.58 6.37	3.50 6.19	3.44 6.03	3.39 5.91	3.34 5.82	3.31 5.74	3.28 5.67
9	5.12 10.56	4.26 8.02	3.86 6.99	3.63 6.42	3.48 6.06	3.37 5.80	3.29 5.62	3.23 5.47	3.18 5.35	3.13 5.26	3.10 5.18	3.07 5.11
10	4.96 10.04	4.10 7.56	3.71 6.55	3.48 5.99	3.33 5.64	3.22 5.39	3.14 5.21	3.07 5.06	3.02 4.95	2.97 4.85	2.94 4.78	2.91 4.71
11	4.84 9.65	3.98 7.20	3.59 6.22	3.36 5.67	3.20 5.32	3.09 5.07	3.01 4.88	2.95 4.74	2.90 4.63	2.86 4.54	2.82 4.46	2.79 4.40

تكملة ...

... تابع جدول ٥ – توزيع F

df المقام	14	16	20	24	30	40	50	75	100	200	500	∞
	درجات حرية df البسط											
1	245 6,142	246 6,169	248 6,208	249 6,234	250 6,261	251 6,286	252 6,302	253 6,323	253 6,334	254 6,352	254 6,361	254 6,366
2	19.42 99.43	19.43 99.44	19.44 99.45	19.45 99.46	19.46 99.47	19.47 99.48	19.47 99.48	19.48 99.49	19.49 99.49	19.49 99.49	19.50 99.50	19.50 99.50
3	8.71 26.92	8.69 26.83	8.66 26.69	8.64 26.60	8.62 26.50	8.60 26.41	8.58 26.35	8.57 26.27	8.56 26.23	8.54 26.18	8.54 26.14	8.53 26.12
4	5.87 14.24	5.84 14.15	5.80 14.02	5.77 13.93	5.74 13.83	5.71 13.74	5.70 13.69	5.68 13.61	5.66 13.57	5.65 13.52	5.64 13.48	5.63 13.46
5	4.64 9.77	4.60 9.68	4.56 9.55	4.53 9.47	4.50 9.38	4.46 9.29	4.44 9.24	4.42 9.17	4.40 9.13	4.38 9.07	4.37 9.04	4.36 9.02
6	3.96 7.60	3.92 7.52	3.87 7.39	3.84 7.31	3.81 7.23	3.77 7.14	3.75 7.09	3.72 7.02	3.71 6.99	3.69 6.94	3.68 6.90	3.67 6.88
7	3.52 6.35	3.49 6.27	3.44 6.15	3.41 6.07	3.38 5.98	3.34 5.90	3.32 5.85	3.29 5.78	3.28 5.75	3.25 5.70	3.24 5.67	3.23 5.65
8	3.23 5.56	3.20 5.48	3.15 5.36	3.12 5.28	3.08 5.20	3.05 5.11	3.03 5.06	3.00 5.00	2.98 4.96	2.96 4.91	2.94 4.88	2.93 4.86
9	3.02 5.00	2.98 4.92	2.93 4.80	2.90 4.73	2.86 4.64	2.82 4.56	2.80 4.51	2.77 4.45	2.76 4.41	3.73 4.36	3.72 4.33	3.71 4.31
10	2.86 4.60	2.82 4.52	2.77 4.41	2.74 4.33	2.70 4.25	2.67 4.17	2.64 4.12	2.61 4.05	2.59 4.01	2.56 3.96	2.55 3.93	2.54 3.91
11	2.74 4.29	2.70 4.21	2.65 4.10	2.61 4.02	2.57 3.94	2.53 3.86	2.50 3.80	2.47 3.74	2.45 3.70	2.42 3.66	2.41 3.62	2.40 3.60

تكملة ...

... تابع جدول ٥ – توزيع F

df المقام	درجات حرية df البسط 1	2	3	4	5	6	7	8	9	10	11	12
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76	2.72	2.69
	9.33	6.93	5.95	5.41	5.06	4.82	4.65	4.50	4.39	4.30	4.22	4.16
13	4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.77	2.72	2.67	2.63	2.60
	9.07	6.70	5.74	5.20	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.56	2.53
	8.86	6.51	5.56	5.03	4.69	4.46	4.28	4.14	4.03	3.94	3.51	3.43
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55	2.51	2.48
	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45	2.42
	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.61	3.55
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45	2.41	2.38
	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.45
18	4.41	3.59	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34
	8.28	6.01	5.09	4.58	4.25	4.01	3.85	3.71	3.60	3.51	3.44	3.37
19	4.38	3.52	3.13	2.90	2.14	2.63	2.55	2.48	2.43	2.38	2.34	2.31
	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30
20	4.35	3.49	3.14	2.87	2.71	2.60	2.52	4.45	2.40	2.35	2.31	2.28
	8.10	5.85	4.94	4.43	4.10	3.87	3.71	3.56	3.45	3.37	3.30	3.23
21	4.32	3.47	3.07	2.8	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25
	8.02	5.78	4.87	4.37	4.04	3.81	3.65	3.51	3.40	3.31	3.24	3.17
22	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.35	2.30	2.26	2.23
	7.94	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12

تكملة ...

تابع جدول ٥ - توزيع F

df المقام	درجات حرية df البسط											
	14	16	20	24	30	40	50	75	100	200	500	∝
12	2.64 4.05	2.60 3.98	2.54 3.86	2.50 3.78	2.46 3.70	2.42 3.61	2.40 3.56	2.36 3.49	2.35 3.46	2.32 3.41	2.31 3.38	2.30 3.36
13	2.55 3.85	2.54 3.78	2.50 3.67	2.42 3.59	2.38 3.51	2.34 3.42	2.32 3.37	2.28 3.30	2.26 3.27	2.24 3.21	2.22 3.18	2.21 3.16
14	2.48 3.70	2.44 3.62	2.39 3.51	2.35 3.43	2.31 3.34	2.27 3.26	2.24 3.21	2.21 3.14	2.19 3.11	3.06 2.10	3.02 2.08	3.00 2.07
15	2.43 3.56	2.39 3.48	2.33 3.36	2.29 3.29	2.25 3.20	2.21 3.12	2.18 3.07	2.15 3.00	2.12 2.97	2.92 2.04	2.89 2.02	2.87 2.01
16	2.37 3.45	2.33 3.37	2.28 3.25	2.24 3.18	2.20 3.10	2.16 3.01	2.13 2.96	2.09 2.98	2.07 2.86	2.04 2.80	2.02 2.77	2.01 2.75
17	2.33 3.35	2.29 3.27	2.23 3.16	2.19 3.08	2.15 3.00	2.11 2.92	2.08 2.86	2.04 2.79	2.02 2.76	1.99 2.70	1.97 2.67	1.96 2.65
18	2.29 3.27	2.25 3.19	2.19 3.07	2.15 3.00	2.11 2.91	2.07 2.83	2.04 2.78	2.00 2.71	1.98 2.68	1.95 2.62	1.93 2.59	1.92 2.57
19	2.26 3.19	2.21 3.12	2.15 3.00	2.11 2.92	2.07 2.84	2.02 2.76	2.00 2.70	1.96 2.63	1.94 2.60	1.91 2.54	1.90 2.51	1.88 2.49
20	2.23 3.13	2.18 3.05	2.12 2.94	2.08 2.86	2.04 2.77	1.99 2.69	1.96 2.63	1.92 2.56	1.90 2.53	1.87 2.47	1.85 2.44	1.84 2.42
21	2.20 3.07	2.15 2.99	2.09 2.88	2.05 2.80	2.00 2.72	1.96 2.63	1.93 2.58	1.89 2.51	1.87 2.47	1.84 2.42	1.82 2.38	.81 2.36
22	2.18 3.02	2.13 2.94	2.07 2.83	2.03 2.75	1.98 2.67	1.93 2.58	1.91 2.53	1.87 2.46	1.84 2.42	1.81 2.37	1.80 2.33	.78 2.31

تكملة ...

... تابع جدول ٥ – توزيع F

df المقام	درجات حرية df البسط 1	2	3	4	5	6	7	8	9	10	11	12
23	4.28 7.88	3.42 5.66	3.03 4.76	2.80 4.26	2.64 3.94	2.53 3.71	2.45 3.54	2.38 3.41	2.32 3.30	2.28 3.21	2.24 3.14	2.20 3.07
24	4.26 7.82	3.40 5.61	3.01 4.72	2.78 4.22	2.62 3.90	2.51 3.67	2.43 3.50	2.36 3.37	2.30 3.25	2.26 3.17	2.22 3.09	2.18 3.03
25	4.24 7.77	3.38 5..57	2.99 4.68	2.76 4.18	2.60 3.86	2.49 3.63	2.41 3.46	2.34 3.32	2.28 3.21	2.24 3.13	2.20 3.05	2.16 2.99
26	4.22 7.72	3.37 5.53	2.98 4.64	2.74 4.14	2.59 3.82	2.47 3.59	2.39 3.42	2.32 3.29	2.27 3.17	2.22 3.49	2.18 3.02	2.15 2.96
27	4.21 7.68	3.35 5.49	2.96 4.60	2.73 4.11	2.57 3.79	2.46 3.56	2.37 3.'39	2.30 3.26	2.25 3.14	2.20 3.06	2.16 2.98	2.13 2.93
28	4.20 7.64	3.34 5.45	2.95 4.57	2.71 4.07	2.56 3.76	2.44 3.53	2.36 3.36	2.29 3.23	2.24 3.11	2.19 3.03	2.15 2.95	2.12 2.90
29	4.18 7.60	3.33 5.42	2.93 4.54	2.70 4.04	2.54 3.73	2.43 3.50	2.35 3.33	2.28 3.20	2.22 3.08	2.18 3.00	2.14 2.92	2.10 2.87
30	4.17 1.56	3.32 5.39	2.92 4.51	2.69 4.02	2.53 3.70	2.42 3.47	2.34 3.30	2.27 3.17	2.21 3.06	2.16 2.98	2.12 2.90	2.09 2.84
32	4.15 7.50	3.30 5.34	2.90 4.46	2.67 3.97	2.51 3.66	2.40 3.42	2.32 3.25	2.25 3.12	2.19 3.01	2.14 2.94	2.10 2.86	2.07 2.80
34	4.13 7.44	3.28 5.29	2.88 4.42	2.65 3.93	2.49 3.61	2.38 3.38	2.30 3.21	2.23 3.08	2.17 2.97	2.12 2.89	2.08 2.82	2.05 2.76
36	4.11 7.39	3.26 5.25	2.86 4.38	2.63 3.89	2.48 3.58	2.36 3.35	2.28 3.18	2.21 3.04	2.15 2.94	2.10 2.86	2.06 2.78	2.03 2.72

تكملة ...

... تابع جدول ٥ - توزيع F

df	درجات حرية df البسط											
المقام	14	16	20	24	30	40	50	75	100	200	500	∞
23	2.14 2.97	2.10 2.89	2.04 2.78	2.00 2.10	1.96 2.62	1.91 2.53	1.88 2.48	1.84 2.41	1.82 2.37	1.79 2.32	1.77 2.28	1.76 2.26
24	2.13 2.93	2.09 2.85	2.02 2.74	1.98 2.66	1.94 2.58	1.89 2.49	1.86 2.44	1.82 2.36	1.80 2.33	1.76 2.27	1.74 2.23	1.73 2.21
25	2.11 2.89	2.06 2.81	2.00 2.70	1.96 2.62	1.92 2.54	1.87 2.45	1.84 2.40	1.80 2.32	1.77 2.29	1.74 2.23	1.72 2.19	1.71 2.17
26	2.10 2.86	2.05 2.77	1.99 2.66	1.95 2.58	1.90 2.56	1.85 2.41	1.82 2.36	1.78 2.28	1.76 2.25	1.72 2.19	1.70 2.15	1.69 2.13
27	2.08 2.83	2.03 2.74	1.97 2.63	1.93 2.55	1.88 2.47	1.84 2.38	1.80 2.33	1.76 2.25	1.74 2.21	1.71 2.16	1.68 2.12	1.67 2. 0
28	2.06 2.80	2.02 2.71	1.96 2.60	1.91 2.52	1.87 2.44	1.81 2.35	1.78 2.30	1.75 2.22	1.72 2.18	1.69 2.13	1.67 2.09	1.65 2.06
29	2.05 2.77	2.00 2.68	1.94 2.57	1.90 2.49	1.85 2.41	1.80 2.32	1.77 2.27	1.73 2.19	1.71 2.15	1.68 2.10	1.65 2.06	1.64 2.03
30	2.04 2.74	1.99 2.66	1.93 2.55	1.89 2.47	1.84 2.38	1.79 2.29	1.76 2.24	1.72 2.16	1.69 2.13	1.66 2.07	1.64 2.03	1.62 2.01
32	2.02 2.70	1.97 2.62	1.91 2.51	1.86 2.42	1.82 2.34	1.76 2.25	1.74 2.20	1.69 2.12	1.67 2.08	1.64 2.02	1.61 1.98	1.59 1.96
34	2.00 2.66	1.95 2.58	1.89 2.47	1.84 2.38	1.80 2.30	1.74 2.21	1.71 2.15	1.67 2.08	1.64 2.04	1.61 1.98	1.59 1.94	1.57 1.91
36	1.98 2.62	1.93 2.54	1.87 2.43	1.82 2.35	1.78 2.26	1.72 2.17	1.69 2.12	1.65 2.04	1.62 2.04	1.59 1.94	1.56 1.90	1.55 1.87

تكملة ...

... تابع جدول ٥ - توزيع F

df المقام	درجات حرية df البسط											
	1	2	3	4	5	6	7	8	9	10	11	12
38	4.10	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02
	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.91	2.82	2.75	2.69
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07	2.04	2.00
	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.88	2.80	2.73	2.66
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.02	1.99
	7.27	5.15	4.29	3.80	3.49	3.26	3.10	2.96	2.86	2.77	2.70	2.64
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98
	7.24	5.12	4.26	3.78	3.46	3.24	3.07	2.94	2.84	2.75	2.68	2.62
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.09	2.04	2.00	1.97
	7.21	5.10	4.24	3.76	3.44	3.22	3.05	2.92	2.82	2.73	2.66	2.60
48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.08	2.03	1.99	1.96
	7.19	5.08	4.22	3.74	3.42	3.20	3.04	2.90	2.80	2.71	2.64	2.58
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02	1.98	1.95
	7.17	5.06	4.20	3.72	3.41	3.18	3.02	2.88	2.78	2.70	2.62	2.56
55	4.02	3.17	2.78	2.54	2.38	2.27	2.18	2.11	2.05	2.00	1.97	1.93
	7.12	5.01	4.16	3.68	3.37	3.15	2.98	2.85	2.75	2.66	2.59	2.53
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92
	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.02	1.98	1.94	1.90
	7.04	4.95	4.10	3.62	3.31	3.09	2.93	2.79	2.70	2.61	2.54	2.47
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.01	1.97	1.93	1.89
	7.01	4.92	4.08	3.60	3.29	3.07	2.91	2.77	2.67	2.59	2.51	2.45

تكملة ...

... تابع جدول ٥ - توزيع F

df المقام	درجات حرية df البسط											
	14	16	20	24	30	40	50	75	100	200	500	∝
38	1.96 2.59	1.92 2.51	1.85 2.40	1.80 2.32	1.76 2.22	1.71 2.14	1.67 2.08	1.63 2.00	1.60 1.97	1.57 1.90	1.54 1.86	1.53 1.84
40	1.95 2.56	1.90 2.49	1.84 2.37	1.79 2.29	1.74 2.20	1.69 2.21	1.66 2.05	1.61 1.97	1.59 1.94	1.55 1.88	1.53 1.84	1.51 1.81
42	1.94 2.54	1.89 2.46	1.82 2.35	1.78 2.26	1.73 2.17	1.68 2.08	1.64 2.02	1.60 1.94	1.57 1.91	1.54 1.85	1.51 1.80	1.49 1.78
44	1.92 2.52	1.88 2.44	1.81 2.32	1.76 2.24	1.72 2.15	1.66 2.06	1.63 2.00	1.58 1.92	1.56 1.88	1.52 1.82	1.50 1.78	1.48 1.75
46	1.91 2.50	1.87 2.42	1.80 2.30	1.75 2.22	1.71 2.13	1.65 2.04	1.62 1.98	1.57 1.90	1.54 1.86	1.51 1.80	1.48 1.76	1.46 1.72
48	1.90 2.48	1.86 2.40	1.79 2.28	1.74 2.20	1.70 2.11	1.64 2.02	1.61 1.96	1.56 1.88	1.53 1.84	1.50 1.78	1.47 1.73	1.45 1.70
50	1.90 2.46	1.85 2.39	1.78 2.26	1.74 2.18	1.69 2.10	1.63 2.00	1.60 1.94	1.55 1.86	1.52 1.82	1.48 1.76	1.46 1.71	1.44 1.68
55	1.88 2.43	1.83 2.35	1.76 2.23	1.12 2.15	1.67 2.06	1.61 1.96	1.58 1.90	1.52 1.82	1.50 1.78	1.46 1.71	1.43 1.66	1.41 1.64
60	1.86 2.40	1.81 2.32	1.75 2.20	1.70 2.12	1.65 2.03	1.59 1.93	1.56 1.87	1.50 1.79	1.48 1.74	1.44 1.68	1.41 1.63	1.39 1.60
65	1.85 2.37	1.80 2.30	1.73 2.18	1.68 2.09	1.63 2.00	1.57 1.90	1.54 1.84	1.49 1.76	1.46 1.71	1.42 1.64	1.39 1.60	1.37 1.56
70	1.84 2.35	1.79 2.28	1.72 2.15	1.67 2.07	1.62 1.98	1.56 1.88	1.53 1.82	1.47 1.74	1.45 1.69	1.40 1.62	1.37 1.56	1.35 1.53

تكملة ...

... تابع جدول ٥ - توزيع F

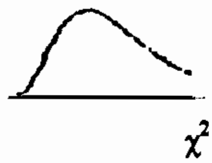
df المقام	درجات حرية df البسط											
	1	2	3	4	5	6	7	8	9	10	11	12
80	3.96 6.96	3.11 4.88	2.72 4.94	2.48 3.56	2.33 3.25	2.21 3.04	2.12 2.87	2.05 2.74	1.99 2.64	1.95 2.55	1.91 2.48	1.88 2.41
100	3.94 6.90	3.09 4.82	2.70 3.98	2.46 3.51	2.30 3.20	2.19 2.99	2.10 2.82	2.03 2.69	1.97 2.59	1.92 2.51	1.88 2.43	1.85 2.36
125	3.92 6.84	3.07 4.78	2.68 3.94	2.44 3.47	2.29 3.17	2.17 2.95	2.08 2.79	2.01 2.65	1.95 2.56	1.90 2.47	1.86 2.40	1.83 2.33
150	3.91 6.91	3.06 4.75	2.67 3.91	2.43 3.44	2.27 3.14	2.16 2.92	2.07 2.76	2.00 2.62	1.94 2.53	1.89 2.44	1.85 2.37	1.82 2.30
200	3.89 6.76	3.04 4.71	2.65 3.88	2.41 3.41	2.26 3.11	2.14 2.90	2.05 2.73	1.98 2.60	1.92 2.50	1.87 2.41	1.83 2.34	1.80 2.28
400	3.86 6.70	3.02 4.66	2.62 3.83	2.39 3.36	2.23 3.06	2.12 2.85	2.03 2.69	1.96 2.55	1.90 2.46	1.85 2.37	1.81 2.29	1.78 2.23
1000	3.85 6.66	3.00 4.62	2.61 3.80	2.38 3.34	2.22 3.04	2.10 2.82	2.02 2.66	1.95 2.53	1.89 2.43	1.84 2.34	1.80 2.26	1.76 2.20
$\propto$	3.84 6.64	3.99 4.60	2.60 3.78	2.37 3.32	2.21 3.02	2.09 2.80	2.01 2.14	1.94 2.51	1.88 2.41	1.83 2.32	1.79 2.24	1.75 2.18

تكملة ...

... تابع جدول ٥ - توزيع F

df المقام	درجات حرية df البسط											
	14	16	20	24	30	40	50	75	100	200	500	∝
80	1.82 2.32	1.77 2.24	1.70 2.11	1.65 2.03	1.60 1.94	1.54 1.84	1.51 1.78	1.45 1.70	1.42 1.65	1.38 1.57	1.35 1.52	1.32 1.49
100	1.79 2.26	1.75 2.19	1.68 2.06	1.63 1.98	1.57 1.89	1.51 1.79	1.48 1.73	1.42 1.64	1.39 1.59	1.34 1.51	1.30 1.46	1.28 1.43
125	1.77 2.23	1.72 2.15	1.65 2.03	1.60 1.94	1.55 1.85	1.49 1.75	1.45 1.68	1.39 1.59	1.36 1.54	1.31 1.46	1.27 1.40	1.25 1.37
150	1.76 2.20	1.71 2.12	1.64 2.00	1.59 1.91	1.54 1.83	1.47 1.72	1.44 1.66	1.37 1.56	1.34 1.51	1.29 1.43	1.25 1.37	1.22 1.33
200	1.74 2.17	1.69 2.09	1.62 1.97	1.57 1.88	1.52 1.79	1.45 1.69	1.42 1.62	1.35 1.53	1.32 1.48	1.26 1.39	1.22 1.33	1.19 1.28
400	1.72 2.12	1.67 2.04	1.60 1.92	1.54 1.84	1.49 1.74	1.42 1.64	1.38 1.57	1.32 1.47	1.28 1.42	1.22 1.32	1.16 1.24	1.13 1.19
1000	1.70 2.09	1.65 2.01	1.58 1.89	1.53 1.81	1.47 1.71	1.41 1.61	1.36 1.54	1.30 1.44	1.26 1.38	1.19 1.28	1.13 1.19	1.08 1.1
∝	1.69 2.07	1.64 1.99	1.57 1.87	1.52 1.79	1.46 1.69	1.40 1.59	1.35 1.52	1.28 1.41	1.24 1.36	1.17 1.25	1.11 1.15	1.00 1.00

جدول ٦ توزيع مربع كاى χ^2



درجات الحرية	احتمال الحصول على قيمة أعلى من 0.995	0.990	0.975	0.950	0.900	0.750	0.500
1	....	....	....	....	0.02	0.10	0.45
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34
15	4.60	5.23	6.27	7.26	8.55	11. 04	14.34
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34
21	8.03	8.90	10.28	11. 59	13.24	16.34	20.34

تكملة ...

... تابع جدول ٦- توزيع مربع كاى χ^2

درجات الحرية	احتمال الحصول على قيمة أعلى من 0.995	0.990	0.975	0.950	0.900	0.750	0.500
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34
23	9.26	10.20	11.69	13.09	14.85	18.14	22.34
24	9.89	10.86	12.40	13.85	15:66	19.04	23.34
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34
27	11.81	12.88	14.57	16.15	18.11	21.75	26.34
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34
29	13.12	1'4.26	16.05	17.71	19.77	23.57	28.34
30	13.79	14.95	16.79	18.49	20.60	24.48	29.34
40	20.71	22.16	24.43	26.51	29.05	33.66	39.44
50	27.99	29.71	32.36	34.76	37.69	42.94	49.33
60	35.53	37.48	40.48	48.49	46.46	52.29	59.33
70	43.28	45.44	48.76	51.74	55.83	61.70	69.33
80	51.17	53.54	57.15	60.39	64.28	71.14	79.33
90	59.20	61. 75	65.65	69.13	73.29	80.62	89.33
100	67.33	70.06	94.22	77.93	82.36	90.13	99.33

تكملة ...

... تابع جدول ٦– توزيع مربع كاى χ^2

درجات الحرية	احتمال الحصول على قيمة أعلى من 0.250	0.100	0.050	0.025	0.010	0.005
1	1.32	2.71	3.84	5.02	6.63	7.88
2	2.77	4.61	5.99	7.38	9.21	10.60
3	4.11	6.25	7.81	9.35	11.34	12.84
4	5.39	7.78	9.49	11.14	13.28	14.86
5	6.63	9.24	11.07	12.83	15.09	16.75
6	7.84	10.64	12.59	14.45	16.81	18.55
7	9.04	12.02	14.07	16.01	18.48	20.28
8	10.22	13.36	15.51	17.53	20.09	21.96
9	11.39	14.68	16.92	19.02	21.67	23.59
10	12.55	15.99	18.31	20.48	23.21	25.19
11	13.70	17.28	19.68	21.92	24.72	26.76
12	14.85	18.55	21.03	23.34	26.22	28.30
13	15.98	19.81	22.36	24.74	27.69	29.82
14	17.12	21.06	23.68	26.12	29.14	31.32
15	18.25	22.31	25.00	27.49	30.58	32.80
16	19.37	23.54	26.30	28.85	32.00	34.27
17	20.49	24.77	27.59	30.19	33.41	35.72
18	21.60	25.99	28.87	31.53	34.81	37.16
19	22.72	27.20	30.14	32.85	36.19	38.58
20	23.83	28.41	31.41	34.17	37.57	40.00
21	24.93	29.62	32.67	35.48	38.93	41.40

تكملة ...

... تابع جدول ٦- توزيع مربع كاى χ^2

درجات الحرية	احتمال الحصول على قيمة أعلى من					
	0.250	**0.100**	**0.050**	**0.025**	**0.010**	**0.005**
22	26.04	30.81	33.92	36.78	40.29	42.80
23	27.14	32.01	35.17	38.08	41.64	44.18
24	28.24	33.20	36.42	39.36	42.98	45.56
25	29.34	34.38	37.65	40.65	44.31	46.93
26	30.43	35.56	38.89	41.92	45.64	48.29
27	31.53	36.74	40.11	43.19	46.96	49.64
28	32.62	37.92	41.34	44.46	48.28	50.99
29	33.71	39.09	42.56	45.72	49.59	52.34
30	34.80	40.26	43.77	46.98	50.89	53.67
40	45.62	51.80	55.76	59.34	63.69	66.77
50	56.33	63.17	67.50	71.42	76.15	79.49
60	66.98	74.40	79.08	83.30	88.38	91.95
70	77.58	85.53	90.53	95.02	100.42	104.22
80	88.13	96.58	101.88	106.63	112.33	116.32
90	98.64	107.56	113.14	118.14	124.12	128.30
100	109.14	118.50	124.34	129.56	135.81	140.17

جدول ٧ – القيم المعنوية لمعاملات الارتباط البسيط (r) والمتعدد (R)

عند مستوى 5% (القيمة العلوية)، 1% (القيمة السفلية)

df الخطأ	عدد المتغيرات المستقلة 1	2	3	4	df الخطأ	عدد المتغيرات المستقلة 1	2	3	4
1	.997 1.000	.999 1.000	.999 1.000	.999 1.000	13	.514 .641	.608 .712	.664 .755	.703 .785
2	.950 .990	.975 .995	.983 .997	.987 .998	14	.497 .623	.590 .694	.646 .737	.686 .768
3	.878 .959	.930 .976	.950 .983	.961 .987	15	.482 .606	.574 .677	.630 .721	.670 .752
4	.811 .917	.881 .949	.912 .962	.930 .970	16	.468 .590	.559 .662	.615 .706	.655 .738
5	.754 .874	.836 .917	.874 .937	.898 .949	17	.456 .575	.545 .647	.601 .691	.641 .724
6	.707 .834	.795 .886	.839 .911	.867 .927	18	.444 .561	.532 .633	.587 .678	.628 .710
7	.666 .798	.758 .855	.807 .885	.838 .904	19	.433 .549	.520 .620	.575 .665	.615 .698
8	.632 .765	.726 .827	.777 .860	.811 .882	20	.423 .537	.509 .608	.563 .652	.604 .685
9	.602 .735	.697 800	.750 .836	.786 .861	21	.413 .526	.498 .596	.522 .641	.592 .674
10	.576 .708	.671 .776	.726 .814	.763 .840	22	.404 .515	.488 .585	.542 .630	.582 .663
11	.553 .684	.648 .753	.703 .793	.741 .821	23	.396 .505	.479 .574	.532 .619	.572 .652
12	.532 .661	.627 .732	.683 .773	.722 .802	24	.388 .496	.470 .565	.523 .609	.562 .642

تكملة ...

... **تابع جدول ٧** - القيم المعنوية لمعاملات الارتباط البسيط (r) والمتعدد (R)

عند مستوى 5% (القيمة العلوية)، 1% (القيمة السفلية)

df	عدد المتغيرات المستقلة			
	1	2	3	4
25	.381	.462	.514	.553
	.487	.555	.600	.633
26	.374	.454	.506	.545
	.478	.546	.590	.624
27	.367	.446	.498	.536
	.470	.538	.582	.615
28	.361	.439	.490	.529
	.463	.530	.573	.606
29	.355	.432	.482	.521
	.456	522	.565	.598
30	.349	.426	.476	.514
	.449	.514	.558	.591
35	.325	.397	.445	.482
	.418	.481	.523	.556
40	.304	.373	.419	.455
	.393	.454	.494	.526
45	.288	.353	.397	.432
	.372	.430	.470	.501
50	.273	.336	.379	.412
	.354	.410	.449	.479
60	.250	.308	.348	.380
	.325	.377	.414	.442
70	.232	.286	.324	.354
	.302	.351	.386	.413
80	.217	.269	.304	.332
	.283	.330	.362	.389
90	.205	.254	.288	.315
	.267	.312	.343	.368
100	.195	.241	.274	.300
	.254	.297	.327	.351
125	.174	.216	.246	.269
	.228	.266	.294	.316
150	.159	.198	.225	.247
	.208	.244	.270	.290
200	.138	.172	.196	.215
	.181	.212	.234	.253
300	.113	.141	.160	.176
	.148	.174	.192	.208
400	.098	.122	.139	.153
	.128	.151	.167	.180
500	.088	.109	.124	.137
	.115	.135	.150	.162
1000	.062	.077	.088	.097
	.081	.096	.106	.115

جدول ٨ – قيم z بدلالة r

r	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
.0	0.000	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080	0.090
.1	.100	.110	.121	.131	.141	.151	.161	.172	.182	.192
.2	.203	.213	.224	.234	.245	.255	.266	.277	.288	.299
.3	.310	.321	.332	.343	.354	.365	.377	.388	.400	.412
.4	.424	.436	.448	.460	.472	.485	.497	.510	.523	.536
.5	.549	.563	.576	.590	.604	.618	.633	.648	.662	.678
.6	.693	.709	.725	.741	.758	.775	.793	.811	.829	.848
.7	.867	.887	.908	.929	.950	.973	.996	1.020	1.045	1.071
.8	1.099	1.127	1.157	1.188	1.221	1.256	1.293	1.333	1.376	1.422

r	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
.90	1.472	1.478	1.483	1.488	1.494	1.499	1.505	1.510	1.516	1.522
.91	1.528	1.533	1.539	1.545	1.551	1.557	1.564	1.570	1.576	1.583
.92	1.589	1.596	1.602	1.609	1.616	1.623	1.630	1.637	1.644	1.651
.93	1.658	1. 666	1.673	1.681	1.689	1.697	1.705	1.713	1.721	1.730
.94	1.738	1.747	1.756	1.764	1.774	1.783	1.792	1.802	1.812	1.822
.95	1.832	1.842	1.853	1.863	1.874	1. 886	1.897	1.909	1.921	1.933
.96	1.946	1.959	1.972	1.986	2.000	2.014	2.029	2.044	2.060	2.076
.97	2.092	2.109	2.127	2.146	2.165	2.185	2.205	2.227	2.249	2.273
.98	2.298	2.323	2.351	2.380	2.410	2:443	2.477	2.515	2.555	2.599
.99	2.646	2.700	2.759	2.826	2.903	2.994	3.106	3.250	3.453	3.800

جدول ٩ – قيم r بدلالة z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.000	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080	0.090
0.1	.100	.110	.119	.129	.139	.149	.159	.168	.178	.187
0.2	.197	.207	.216	.226	.236	.245	.254	.264	.273	.282
0.3	.291	.300	.310	.319	.327	.336	.345	.354	.363	.371
0.4	.380	.389	.397	.405	.414	.422	.430	.438	.446	.454
0.5	.462	.470	.478	.485	.493	.500	.508	.515	.523	.530
0.6	.537	.544	.551	.558	.565	.572	.578	.585	.592	.598
0.7	.604	.611	.617	.623	.629	.635	.641	.647	.653	.658
0.8	.664	.670	.675	.680	.686	.691	.696	.701	.706	.711
0.9	.716	.721	.726	.731	.735	.740	.744	.749	.753	.757
1.0	.762	.766	.770	.774	.778	.782	.786	.790	.793	.797
1.1	.800	.804	.808	.811	.814	.818	.821	.824	.828	.831
1.2	.834	.837	.840	.843	.846	.848	.851	.854	.856	.859
1.3	.862	.864	.867	.869	.872	.874	.876	.879	.881	.883
1.4	.885	.888	.890	.892	.894	.896	.898	.900	.902	.903
1.5	.905	.907	.909	.910	.912	.914	.915	.917	.919	.920
1.6	.922	.923	.925	.926	.928	.929	.930	.932	.933	.934
1.7	.935	.937	.938	.939	.940	.941	.942	.944	.945	.946
1.8	.947	.948	.949	.950	.951	.952	.953	.954	.954	.955
1.9	.956	.957	.958	.959	.960	.960	.961	.962	.963	.963
2.0	.964	.965	.965	.966	.967	.967	.968	.969	.969	.970
2.1	.970	.971	.972	.972	.973	.973	.974	.974	.975	.975
2.2	.976	.976	.977	.977	.978	.978	.978	.979	.979	.980
2.3	.980	.980	.981	.981	.982	.982	.982	.983	.983	.983
2.4	.984	.984	.984	.985	.985	.985	.986	.986	.986	.986
2.5	.987	.987	.987	.987	.988	.988	.988	.988	.989	.989
2.6	.989	.989	.989	.990	.990	.990	.990	.990	.991	.991
2.7	.991	.991	.991	.992	.992	.992	.992	.992	.992	.992
2.8	.993	.993	.993	.993	.993	.993	.993	.994	.994	.994
2.9	.994	.994	.994	.994	.994	.995	.995	.995	.995	.995

جدول ١٠- توزيع الإحصاء F_{max}

$$F_{max} = (\hat{\sigma}^2_{largest} / \hat{\sigma}^2_{smallest})$$

df for S_i^2	α	k = number of variances										
		2	3	4	5	6	7	8	9	10	11	12
4	.05	9.60	15.5	20.6	25.2	29.5	33.6	37.5	41.4	44.6	48.0	51.4
	.01	23.2	37.0	49.0	59.0	69.0	79.0	89.0	97.0	106.0	113.0	120.0
5	.05	7.15	10.8	13.7	16.3	18.7	20.8	22.9	24.7	26.5	28.2	29.9
	.01	14.9	22.0	28.0	33.0	38.0	42.0	46.0.	50.0	54.0	57.0	60.0
6	.05	5.82	8.38	10.4	12.1	13.7	15.0	16.3	17.5	18.6	19.7	20.7
	.01	11.1	15.5	19.1	22.0	25.0	27.0	30.0	32.0	34.0	36.0	37.0
7	.05	4.99	6.94	8.44	9.70	10.8	11.8	12.7	13.5	14.3	15.1	15.8
	.01	8.89	12.1	14.5	16.5	18.4	20.0	22.0	23.0	24.0	26.0	27.0
8	.05	4.43	6.00	7.18	8.12	9.03	9.78	10.5	11.1	11.7	12.2	12.7
	.01	7.50	9.9	11.7	13.2	14.5	15.8	16.9	17.9	18.9	19.8	21.0
9	.05	4.03	5.34	6.31	7.11	7.80.	8.41	8.95	9.45	9.91	10.3	10.7
	.01	6.54	8.5	9.9	11.1	12.1	13.1	13.9	14.7	15.3	16.0	16.6
10	.05	3.72	4.85	5.67	6.34	6.92	7.42	7.87	8.28	8.66	9.01	9.34
	.01	5.85	7.4	8.6	9.6	10.4	11.1	11.8	12.4	12.9	13.4	13.9
12	.05	3.28	4.16	4.79	5.30	5.72	6.09	6.42	6.72	7.00	7.25	7.48
	.01	4.91	6.1	6.9	7.6	8.2	8.7	9.1	9.5	9.9	10.2	10.6
15	.05	2.86	3.54	4.01	4.37	4.68	4.95	5.19	5.40	5.59	5.77	5.93
	.01	4.07	4.9	5.5	6.0	6.4	6.7	7.1	7.3	7.5	7.8	8.0
20	.05	2.46	2.95	3.29	3.54	3.76	3.94	4.10	4.24	4.37	4.49	4.59
	.01	3.32	3.80	4.30	4.60	4.90	5.10	5.30	5.50	5.60	5.80	5.90
30	.05	2.07	2.40	2.61	2.78	2.91	3.02	3.12	3.21	3.29	3.36	3.39
	.01	2.63	3.00	3.30	3.40	3.60	3.70	3.80	3.90	4.00	4.10	4.20
60	.05	1.67	1.85	1.96	2.04	2.11	2.17	2.22	2.26	2.30	2.33	2.36
	.01	1.96	2.20	2.30	2.40	2.40	2.50	2.50	2.60	2.60	2.70	2.70
∝	.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

جدول ١١– القيم الحرجة لاختبار Cochran لتجانس التباين

$$C = (\text{largest } S^2 / \sum S_i^2)$$

df for S_i^2	α	k = number of variances										
		2	3	4	5	6	7	8	9	10	15	20
1	.05	.9985	.9669	.9065	.8412	.7808	.7271	.6798	.6385	.6020	.4709	.3894
	.01	.9999	.9933	.9676	.9279	.8828	.8376	.7945	.7544	.7175	.5747	.4799
2	.05	.9750	.8709	.7679	.6838	.6161	.5612	.5157	.4775	.4450	.3346	.2705
	.01	.9950	.9423	.8643	.7885	.7218	.6644	.6152	.5727	.5358	.4069	.3297
3	.05	.9392	.7977	.6841	.5981	.5321	.4800	.4377	.4027	.3733	.2758	.2205
	.01	.9794	.8831	.7814	.6957	.6258	.5685	.5209	.4810	.4469	.3317	.2654
4	.05	.9057	.7457	.6287	.5441	.4803	.4307	.3910	.3584	.3311	.2419	.1921
	.01	.9586	.8335	.7212	.6329	.5635	.5080	.4627	.4251	.3934	.2882	.2288
5	.05	.8772	.7071	.5895	.5065	.4447	.3974	.3595	.3286	.3029	.2195	.1735
	.01	.9373	.7933	.6761	.5875	.5195	.4659	.4226	.3870	.3572	.2593	.2048
6	.05	.8534	.6771	.5598	.4783	.4184	.3726	.3362	.3067	.2823	.2034	.1602
	.01	.9172	.7606	.6410	.5531	.4866	.4347	.3932	.3592	.3308	.2386	.1877
7	.05	.8332	.6530	.5365	.4564	.3980	.3535	.3185	.2901	.2666	.1911	.1501
	.01	.8988	.7335	.6129	.5259	.4608	.4105	.3704	.3378	.3106	.2228	.1748
8	.05	.8159	.6333	.5175	.4387	.3817	.3384	.3043	.2768	.2541	.1815	.1422
	.01	.8823	.7107	.5897	.5037	.4401	.3911	.3522	.3207	.2945	.2104	.1646
9	.05	.8010	.6167	.5017	.4241	.3682	.3259	.2926	.2659	.2439	.1736	.1357
	.01	.8674	.6912	.5702	.4854	.4229	.3751	.3373	.3067	.2813	.2002	.1567
16	.05	.7341	.5466	.4366	.3645	.3135	.2756	.2462	.2226	.2032	.1429	.1108
	.01	.7949	.6059	.4884	.4094	.3529	.3105	.2779	.2514	.2297	.1612	.1248
36	.05	.6602	.4748	.3720	.3066	.2612	.2278	.2022	.1820	.1655	.1144	.0879
	.01	.7067	.5153	.4057	.3351	.2858	.2494	.2214	.1992	.1811	.1251	.0960
∞	.05	.5813	.4031	.3093	.2513	.2119	.1833	.1616	.1446	.1308	.0889	.0675
	.01	.6062	.4230	.3251	.2644	.2229	.1929	.1700	.1521	.1376	.0934	.0709

جدول ١٢ – اختبار معنوية الالتواء

اختبار من طرف واحد (حيث $\sqrt{b_1} = m_3 / m_2\sqrt{m_2}$)

حجم العينة n	نقطة النسبة 5%	1%	الانحراف المعيارى	حجم العينة n	نقطة النسبة 5%	1%	الانحراف المعيارى
25	0.711	1.061	0.4354	100	0.389	0.567	0.2377
30	.661	0.982	.4052	125	.350	.508	.2139
35	.621	0.921	.3804	150	.321	.464	.1961
40	.587	0.869	.3596	175	.298	.430	.1820
45	.558	0.825	.3418	200	.280	.403	.1706
50	.533	0.787	.3264	250	.251	.360	.1531
60	.492	0.723	.3009	300	.230	.329	.1400
70	.459	0.673	.2806	350	.213	.305	.1298
80	.432	0.631	.2638	400	.200	.285	.1216
90	.409	0.596	.2498	450	.188	.269	.1147
100	0.389	0.567	0.2377	500	0.179	0.255	0.1089

جدول ١٣– اختبار معنوية التفرطح

(حيث $b_2 = m_4 / m_2^2$)

حجم	الأعلى		الأدنى		حجم	الأعلى		الأدنى	
العينة n	1%	5%	5%	1%	العينة n	1%	5%	5%	1%
50	4.88	3.99	2.15	1.95	600	3.54	3.34	2.70	2.60
75	4.59	3.87	2.27	2.08	650	3.52	3.33	2.71	2.61
100	4.39	3.77	2.35	2.18	700	3.50	3.31	2.72	2.62
125	4.24	3.71	2.40	2.24	750	3.48	3.30	2.73	2.64
150	4.13	3.65	2.45	2.29	800	3.46	3.29	2.74	2.65
					850	3.45	3.28	2.74	2.66
200	3.98	3.57	2.51	2.37	900	3.43	3.28	2.75	2.66
250	3.87	3.52	2.55	2.42	950	3.42	3.27	2.76	2.67
300	3.79	3.47	2.59	2.46	1000	3.41	3.26	2.76	2.68
350	3.72	3.44	2.62	2.50					
400	3.67	3.41	2.64	2.52	1200	3.37	3.24	2.78	2.71
450	3.63	3.39	2.66	2.55	1400	3.34	3.22	2.80	2.72
500	3.60	3.37	2.67	2.57	1600	3.32	3.21	2.81	2.74
550	3.57	3.35	2.69	2.58	1800	3.30	3.20	2.82	2.76
600	3.54	3.34	2.70	2.60	2000	3.28	3.18	2.83	2.77

جدول ١٤- تحوير مقلوب جيب الزاوية Arcsin

$$\varphi = 2 \arcsin \sqrt{X}$$

X	φ	X	φ	X	φ	X	φ	X	φ
.001	.0633	.041	.4078	.36	1.2870	.76	2.1177	.971	2.7993
.002	.0895	.042	.4128	.37	1.3078	.77	2.1412	.972	2.8053
.003	.1096	.043	.4178	.38	1.3284	.78	2.1652	.973	2.8115
.004	.1266	.044	.4227	.39	1.3490	.79	2.1895	.974	2.8177
.005	.1415	.045	.4275	.40	1.3694	.80	2.2143	.975	2.8240
.006	.1551	.046	.4323	.41	1.3898	.81	2.2395	.976	2.8305
.007	.1675	.047	.4371	.42	1.4101	.82	2.2653	.977	2.8371
.008	.1791	.048	.4418	.43	1.4303	.83	2.2916	.978	2.8438
009	.1900	.049	.4464	.44	1.4505	.84	2.3186	.979	2.8507
.010	.2003	.050	.4510	.45	1.4706	.85	2.3462	.980	2.8578
.011	.2101	.06	.4949	.46	1.4907	.86	2.3746	.981	2.8650
.012	.2195	.07	.5355	.47	1.5108	.87	2.4039	.982	2.8725
.013	.2285	.08	.5735	.48	1.5308	.88	2.4341	.983	2.8801
.014	.2372	.09	.6094	.49	1.5508	.89	2.4655	.984	2.8879
.015	.2456	.10	.6435	.50	1.5708	.90	2.4981	.985	2.8960
.016	.2517	.11	.6761	.51	1.5908	.91	2.5322	.986	2.9044
.017	.2615	.12	.7075	.52	1.6108	.92	2.5681	.987	2.9131
.018	.2691	.13	.7377	.53	1.6308	.93	2.6062	.988	2.9221
.019	.2766	.14	.7670	.54	1.6509	.94	2.6467	.989	2.9315
.020	.2838	.15	.7954	.55	1.6710	.95	2.6906	.990	2.9413
.021	.2909	.16	.8230	.56	1.6911	.951	2.6952	.991	2.9516
.022	.2978	.17	.8500	.57	1.7113	.952	2.6998	.992	2.9625
.023	.3045	.18	.8763	.58	1.7315	.953	2.7045	.993	2.9741
.024	.3111	.19	.9021	.59	1.7518	.954	2.7093	.994	2.9865
.025	.3176	.20	.9273	.60	1.7722	.955	2.7141	.995	3.0001

تكملة ...

... تابع – جدول ١٤ تحوير مقلوب جيب الزاوية Arcsin

$$\varphi = 2 \arcsin \sqrt{X}$$

X	φ	X	φ	X	φ	X	φ	X	φ
026	.3239	.21	.9521	.61	1.7926	.956	2.7189	.996	3.0150
.027	.3301	.22	.9764	.62	1.8132	.957	2.7238	.997	3.0320
028	.3363	.23	1.0004	.63	1.8338	.958'	2.7288	.998	3.0521
.029	.3423	.24	1.0239	.64	1.8546	.959	2.7338	.999	3.0783
.030	.3482	.25	1.0472	.65	1.8755	.960	2.7389		
.031	.3540	.26	1.0701	.66	1.8965	.961	2.7440		
.032	.3597	.27	1.0928	.67	1.9177	.962	2.7492		
.033	.3654	.28	1.1152	.68	1.9391	.963	2.7545		
.034	.3709	.29	1.1374	.69	1.9606	.964	2.7598		
.035	.3764	.30	1.1593	.70	1.9823	.965	2.7652		
.036	.3818	.31	1.1810	.71	2.0042	.966	2.7707		
.037	.3871	.32	1.2025	.72	2.0264	.967	2.7762		
.038	.3924	.33	1.2239	.73	2.0488	.968	2.7819		
.039	.3976	.34	1.2451	.74	2.0715	.969	2.7876		
.040	.4027	.35	1.2661	.75	2.0944	.970	2.7934		

جدول ١٥- قيم توكى باحتمال 5%

df الخطأ	عدد المعاملات									
	2	3	4	5	6	7	8	9	10	11
1	18.00	27.00	32.80	37.20	40.50	43.10	45.40	47.30	49.10	50.60
2	6.09	8.33	9.80	10.89	11.73	12.43	13.03	13.54	13.99	14.39
3	4.50	5.91	6.83	7.51	8.04	8.47	8.85	9.18	9.46	9.72
4	3.93	5.04	5.76	6.29	6.71	7.06	7.35	7.60	7.83	8.03
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.51
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.21
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82	4.92
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73	4.82
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73
120	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56	4.64
∝	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55

تكملة ...

... **تابع جدول ١٥**- قيم توكى باحتمال 5%

df الخطأ	عدد المعاملات								
	12	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	51.90	53.20	54.30	55.40	56.30	57.20	58.00	58.80	59.60
2	15.08	15.58	15.38	15.65	15.91	16.14	16.36	16.57	16.77
3	9.35	10.16	10.35	10.52	10.69	10.84	10.98	11.12	11.24
4	8.21	8.37	8.52	8.67	8.80	8.92	9.03	9.14	9.24
5	7.32	7.47	7.60	7.72	7.83	7.93	8.03	8.12	8.21
6	6.79	6.92	7.03	7.14	7.24	7.34	7.43	7.51	7.59
7	6.43	6.55	6.66	6.76	6.85	6.94	7.02	7.10	7.17
8	6.18	6.29	6.39	6.48	6.57	6.65	6.73	6.80	6.87
9	5.98	6.09	6.19	6.28	6.36	6.44	6.51	6.58	6.64
10	5.83	5.93	6.03	6.11	6.19	6.27	6.34	6.40	6.47
11	5.71	5.81	5.90	5.98	6.06	6.13	6.20	6.27	6.33
12	5.61	5.71	5.80	5.88	5.95	6.02	6.09	6.15	6.21
13	5.53	5.63	5.71	5.79	5.86	5.93	5.99	6.05	6.11
14	5.46	5.55	5.64	5.71	5.79	5.85	5.91	5.97	6.03
15	5.40	5.49	5.57	5.65	5.72	5.78	5.85	5.90	5.96
16	5.35	5.44	5.52	5.59	5.66	5.73	5.79	5.84	5.90
17	5.31	5.39	5.47	5.54	5.61	5.67	5.73	5.79	5.84
18	5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79
19	5.23	5.31	5.39	5.46	5.53	5.59	5.65	5.70	5.75
20	5.20	5.28	5.36	5.43	5.49	5.55	5.61	5.66	5.71
24	5.10	5.18	5.25	5.32	5.38	5.44	5.49	5.55	5.59
30	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.47
40	4.90	4.98	5.04	5.11	5.16	5.22	5.27	5.31	5.36
60	4.81	4.88	4.94	5.00	5.06	5.11	5.15	5.20	5.24
120	4.71	4.78	4.84	4.90	4.95	5.00	5.04	5.09	5.13
$\propto$	4.62	4.68	4.74	4.80	4.85	4.89	4.93	4.97	5.01

جدول ١٦ - قيم دنكن

df الخطأ	α	عدد المتوسطات في المدى 2	3	4	5	6	7	8
1	.05	18.00	18.00	18.00	180.0	18.00	18.00	18.00
	.01	90.00	90.00	90.00	90.00	90.00	90.00	90.00
2	.05	6.09	6.09	6.09	6.09	6.09	6.09	6.09
	.01	14.0	14.0	14.0	14.0	14.0	14.0	14.0
3	.05	4.50	4.50	4.50	4.50	4.50	4.50	4.50
	.01	8.26	8.50	8.60	8.70	8.80	8.90	8.90
4	.05	3.93	4.01	4.02	4.02	4.02	4.02	4.02
	.01	6.51	6.80	6.90	7.00	7.10	7.10	7.20
5	.05	3.64	3.74	3.79	3.83	3.83	3.83	3.83
	.01	5.70	5.96	6.11	6.18	6.26	6.33	6.40
6	.05	3.46	3.58	3.64	3.68	3.68	3.68	3.68
	.01	5.24	5.51	5.65	5.73	5.81	5.88	5.95
7	.05	3.35	3.47	3.54	3.58	3.60	3.61	3.61
	.01	4.95	5.22	5.37	5.45	5.53	5.61	5.69
8	.05	3.26	3.39	3.47	3.52	3.55	3.56	3.56
	.01	4.74	5.00	5.14	5.23	5.32	5.40	5.47
9	.05	3.20	3.34	3.41	3.47	3.50	3.52	3.52
	.01	4.60	4.86	4.99	5.08	5.17	5.25	5.32
10	.05	3.15	3.30	3.37	3.43	3.46	3.47	3.47
	.01	4.48	4.73	4.88	4.96	5.06	5.13	5.20
11	.05	3.11	3.27	3.35	3.39	3.43	3.44	3.45
	.01	4.39	4.63	4.77	4.86	4.94	5.01	5.06
12	.05	3.08	3.23	3.33	3.36	3.40	3.42	3.44
	.01	4.32	4.55	4.68	4.76	4.84	4.92	4.96
13	.05	3.06	3.21	3.30	3.35	3.38	3.41	3.42
	.01	4.26	4.48	4.62	4.69	4.74	4.84	4.88
14	.05	3.03	3.18	3.27	3.33	3.37	3.39	3.41
	.01	4.21	4.42	4.55	4.63	4.70	4.78	4.83

تكملة ...

... تابع جدول ١٦ – قيم دنكن

df الخطأ	α	عدد المتوسطات في المدى 2	3	4	5	6	7	8
15	.05	3.01	3.16	3.25	3.31	3.36	3.38	3.40
	.01	4.17	4.37	4.50	4.58	4.64	4.72	4.77
16	.05	3.00	3.15	3.23	3.30	3.34	3.37	3.39
	.01	4.13	4.34	4.45	4.54	4.60	4.67	4.72
17	.05	2.98	3.13	3.22	3.28	3.33	3.36	3.38
	.01	4.10	4.30	4.41	4.50	4.56	4.63	4.68
18	.05	2.97	3.12	3.21	3.27	3.32	3.35	3.37
	.01	4.07	4.27	4.38	4.46	4.53	4.59	4.64
19	.05	2.96	3.11	3.19	3.26	3.31	3.35	3.37
	.01	4.05	4.24	4.35	4.43	4.50	4.56	4.61
20	.05	2.95	3.10	3.18	3.25	3.30	3.34	3.36
	.01	4.02	4.22	4.33	4.40	4.47	4.53	4.58
22	.05	2.93	3.08	3.17	3.24	3.29	3.32	3.35
	.01	3.99	4.17	4.28	4.36	4.42	4.48	4.53
24	.05	2.92	3.07	3.15	3.22	3.28	3.31	3.34
	.01	3.96	4.14	4.24	4.33	4.39	4.44	4.49
26	.05	2.91	3.06	3.14	3.21	3.27	3.30	3.34
	.01	3.93	4.11	4.21	4.30	4.36	4.41	4.46
28	.05	2.90	3.04	3.13	3.20	3.26	3.30	3.33
	.01	3.91	4.08	4.18	4.28	4.34	4.39	4.43
30	.05	2.89	3.04	3.12	3.20	3.25	3.29	3.32
	.01	3.89	4.06	4.16	4.22	4.32	4.36	4.41
40	.05	2.86	3.01	3.10	3.17	3.22	3.27	3.30
	.01	3.82	3.99	4.10	4.17	4.24	4.30	4.34
60	.05	2.83	2.98	3.08	3.14	3.20	3.24	3.28
	.01	3.76	3.92	4.03	4.12	4.17	4.23	4.27
100	.05	2.80	2.95	3.05	3.12	3.18	3.22	3.26
	.01	3.71	3.86	3.93	4.06	4.11	4.17	4.21
∝	.05	2.77	2.92	3.02	3.09	3.15	3.19	3.23
	.01	3.64	3.80	3.90	3.98	4.04	4.09	4.14

تكملة ...

... تابع – جدول ١٦ قيم دنكن

df	α	عدد المتوسطات في المدى						
		9	10	12	14	16	18	20
1	.05	18.0	18.0	18.0	18.0	18.0	18.0	18.0
	.01	90.0	90.0	90.0	90.0	90.0	90.0	90.0
2	.05	6.09	6.09	6.09	6.09	6.09	6.09	6.09
	.01	14.0	14.0	14.0	14.0	14.0	14.0	14.0
3	.05	4.50	4.50	4.50	4.50	4.50	4.50	4.50
	.01	9.00	9.00	9.00	9.10	9.20	9.30	9.30
4	.05	4.02	4.02	4.02	4.02	4.02	4.02	4.02
	.01	7.20	7.30	7.30	7.40	7.40	7.50	7.50
5	.05	3.83	383	383	3.83	3.83	3.83	3.83
	.01	6.44	6.50	6.60	6.60	6.70	6.70	6.80
6	.05	3.68	3.68	3.68	3.68	3.68	3.68	3.68
	.01	6.00	6.00	6.10	6.20	6.20	6.30	6.30
7	.05	3.61	3.61	3.61	3.61	3.61	3.61	3.61
	.01	5.73	5.80	5.80	5.90	5.90	6.00	6.00
8	.05	3.56	3.56	3.56	3.56	3.56	3.56	3.56
	.01	5.51	5.50	5.60	5.70	5.70	5.80	5.80
9	.05	3.52	3.52	3.52	3.52	3.52	3.52	3.52
	.01	5.36	5.40	5.50	5.50	5.60	5.70	5.70
10	.05	3.47	3.47	3.47	3.47	3.47	3.47	3.48
	.01	5.24	5.28	5.36	5.42	5.48	5.54	5.55
11	.05	3.46	4.46	3.46	3.46	3.46	3.47	3.48
	.01	5.12	5.15	5.24	5.28	534	5.38	5.39
12	.05	3.44	3.46	3.46	3.46	3.46	3.47	3.48
	.01	5.02	5.07	5.13	5.17	5.22	5.24	5.26
13	.05	3.44	3.45	3.45	3.46	3.46	3.47	3.47
	.01	4.94	4.98	5.04	5.08	5.13	5.14	5.15
14	.05	3.42	3.44	3.45	3.46	3.46	3.47	3.47
	.01	4.87	4.91	4.96	5.00	5.04	5.06	5.07

تكملة ...

... تابع – جدول ١٦ قيم دنكن

df الخطأ	α	عدد المتوسطات في المدى						
		9	10	12	14	16	18	20
15	.05	3.42	3.43	3.44	3.45	3.46	3.47	3.47
	.01	4.81	4.84	4.90	4.94	4.97	4.99	5.00
16	.05	3.41	3.43	3.44	3.45	3.46	3.47	3.47
	.01	4.76	4.79	4.84	4.88	4.91	4.93	4.94
17	.05	3.40	3.42	3.44	3.45	3.46	3.47	3.47
	.01	4.72	4.75	4.80	4.83	4.86	4.88	4.89
18	.05	3.39	3.41	3.43	3.45	3.46	3.47	3.47
	.01	4.68	4.71	4.76	4.79	4.82	4.84	4.85
19	.05	3.39	3.41	3.43	3.44	3.46	3.47	3.47
	.01	4.64	4.67	4.72	4.76	4.79	4.81	4.82
20	.05	3.38	3.40	3.43	3.44	3.46	3.46	3.47
	.01	4.61	6.65	4.69	4.73	4.76	4.78	4.79
22	.05	3.37	3.39	3.42	3.44	3.45	3.46	3.47
	.01	4.57	4.60	4.65	4.68	4.71	4.74	4.75
24	.05	3.37	3.38	3.41	3.44	3.45	3.46	3.47
	.01	4.53	4.57	4.62	4.64	4.67	4.70	4.72
26	.05	3.36	3.38	3.41	3.43	3.45	3.46	3.47
	.01	4.50	4.53	4.58	4.62	4.65	4.67	4.69
28	.05	3.35	3.37	3.40	3.43	3.45	3.46	3.47
	.01	4.47	4.51	4.56	4.60	4.62	4.65	4.67
30	.05	3.35	3.37	3.40	3.43	3.44	3.46	3.47
	.01	4.45	4.48	4.54	4.58	4.61	4.63	4.65
40	.05	3.33	3.35	3.39	3.42	3.44	3.46	3.47
	.01	4.37	4.41	4.46	4.51	4.54	4.57	4.59
60	.05	3.31	3.33	3.37	3.40	3.43	3.45	3.47
	.01	4.31	4.34	4.39	4.44	4.47	4.50	4.53
100	.05	3.29	3.32	3.36	3.40	3.42	3.45	3.47
	.01	4.25	4.29	4.35	4.38	4.42	4.45	4.48
∝	.05	3.26	3.29	3.34	3.38	3.41	3.44	3.47
	.01	4.17	4.20	4.26	4.31	4.34	4.38	4.41

جدول ١٧ – معاملات الحدود المتعددة المتعامدة (r عدد المعاملات)

r	Polynomial	المعاملات										مجموع مربعات المعاملات
3	Linear	-1	0	1								2
	Quadratic	1	-2	1								6
4	Linear	-3	-1	1	3							20
	Quadratic	1	-1	-1	1							4
	Cubic	-1	3	-3	1							20
5	Linear	-2	-1	0	1	2						10
	Quadratic	2	-1	-2	-1	2						14
	Cubic	-1	2	0	-2	1						10
	Quartic	1	-4	6	-4	1						70
6	Linear	-5	-3	-1	1	3	5					70
	Quadratic	5	-1	-4	-4	-1	5					84
	Cubic	-5	7	4	-4	-7	5					180
	Quartic	1	-3	2	2	-3	1					28
7	Linear	-3	-2	-1	0	1	2	3				28
	Quadratic	5	0	-3	-4	-3	0	5				84
	Cubic	-1	1	1	0	-1	-1	1				6
	Quartic	3	-7	1	6	1	-7	3				154
8	Linear	-7	-5	-3	-1	1	3	5	7			168
	Quadratic	7	1	-3	-5	-5	-3	1	7			168
	Cubic	-7	5	7	3	-3	-7	-5	7			264
	Quartic	7	-13	-3	9	9	-3	-13	7			616
	Quintic	-7	23	-17	-15	15	17	-23	7			2184
9	Linear	-4	-3	-2	-1	0	1	2	3	4		60
	Quadratic	28	7	-8	-17	-20	-17	-8	7	28		2772
	Cubic	-14	7	13	9	0	-9	-13	-7	14		990
	Quartic	14	-21	-11	9	18	9	-11	-21	14		2002
	Quintic	-4	11	-4	-9	0	9	4	-11	4		468
10	Linear	-9	-7	-5	-3	-1	1	3	5	7	9	330
	Quadratic	6	2	-1	-3	-4	-4	-3	-1	2	6	132
	Cubic	-42	14	35	31	12	-12	-31	-35	-14	42	8580
	Quartic	18	-22	-17	3	18	18	3	-17	-22	18	2860
	Quintic	-6	14	-1	-11	-6	6	11	1	-14	6	780

جدول ١٨ – تحديد العدد الأمثل لتحليل التباين (نموذج ثابت r عدد المعاملات)

	Power $1-\beta = .70$															
	$\Delta/\sigma = 1.0$				$\Delta/\sigma = 1.25$				$\Delta/\sigma = 1.50$				$\Delta/\sigma = 1.75$			
r	α				α				α				α			
	.2	**.1**	**.05**	**.01**	**.2**	**.1**	**.05**	**.01**	**.2**	**.1**	**.05**	**.01**	**.2**	**.1**	**.05**	**.01**
2	7	11	14	21	5	7	9	15	4	6	7	11	3	4	6	9
3	9	13	17	25	6	9	11	17	5	7	8	12	4	5	7	10
4	11	15	19	28	7	10	13	19	5	7	9	13	4	6	7	10
5	12	17	21	30	8	11	14	20	6	8	10	14	5	6	8	11
6	13	18	22	32	9	12	15	21	6	9	11	15	5	7	8	12
7	14	19	24	34	9	13	16	22	7	9	11	16	5	7	9	12
8	15	20	25	35	10	13	16	23	7	10	12	17	6	7	9	13
9	15	21	26	37	10	14	17	24	7	10	12	17	6	8	9	13
10	16	22	27	38	11	14	18	25	8	10	13	18	6	8	10	14

	Power $1-\beta = .70$											
	$\Delta/\sigma = 2.0$				$\Delta/\sigma = 2.5$				$\Delta/\sigma = 3.0$			
r	α				α				α			
	.2	**.1**	**.05**	**.01**	**.2**	**.1**	**.05**	**.01**	**.2**	**.1**	**.05**	**.01**
2	3	4	5	7	2	3	4	5	2	3	3	5
3	3	4	5	8	3	3	4	6	2	3	3	5
4	4	5	6	8	3	4	4	6	2	3	4	5
5	4	5	6	9	3	4	5	6	3	3	4	5
6	4	5	7	9	3	4	5	7	3	3	4	5
7	4	6	7	10	3	4	5	7	3	3	4	5
8	5	6	7	10	3	4	5	7	3	3	4	5
9	5	6	8	10	3	4	5	7	3	4	4	6
10	5	6	8	11	4	5	6	7	3	4	4	6

تكملة ...

... **تابع جدول ١٨** – تحديد العدد الأمثل لتحليل التباين

(نموذج ثابت r عدد المعاملات)

Power $1-\beta = .80$																
	$\Delta/\sigma = 1.0$				$\Delta/\sigma = 1.25$				$\Delta/\sigma = 1.50$				$\Delta/\sigma = 1.75$			
r	**α**				**α**				**α**				**α**			
	.2	**.1**	**.05**	**.01**	**.2**	**.1**	**.05**	**.01**	**.2**	**.1**	**.05**	**.01**	**.2**	**.1**	**.05**	**.01**
2	10	14	17	26	7	9	12	17	5	7	9	13	4	5	7	10
3	12	17	21	30	8	11	14	20	6	8	10	14	5	6	8	11
4	14	19	23	33	9	13	15	22	7	9	11	16	5	7	9	12
5	16	21	25	35	10	14	17	23	8	10	12	17	6	8	9	13
6	17	22	27	38	11	15	18	25	8	11	13	18	6	8	10	13
7	18	24	29	39	12	16	19	26	9	11	14	18	7	9	10	14
8	19	25	30	41	12	16	20	27	9	12	14	19	7	9	11	15
9	20	26	31	43	13	17	21	28	9	12	15	20	7	9	11	15
10	21	27	33	44	14	18	21	29	10	13	15	21	8	10	12	16

Power $1-\beta = .80$												
	$\Delta/\sigma = 2.0$				$\Delta/\sigma = 2.5$				$\Delta/\sigma = 3.0$			
r	**α**				**α**				**α**			
	.2	**.1**	**.05**	**.01**	**.2**	**.1**	**.05**	**.01**	**.2**	**.1**	**.05**	**.01**
2	3	4	6	8	3	3	4	6	2	3	4	5
3	4	5	6	9	3	4	5	7	3	3	4	5
4	4	6	7	10	3	4	5	7	3	3	4	5
5	5	6	7	10	4	4	5	7	3	4	4	6
6	5	7	8	11	4	5	6	8	3	4	4	6
7	5	7	8	11	4	5	6	8	3	4	5	6
8	6	7	9	12	4	5	6	8	3	4	5	6
9	6	7	9	12	4	5	6	8	3	4	5	6
10	6	8	9	12	4	5	6	8	3	4	5	6

تكملة ...

... **تابع جدول ١٨** – تحديد العدد الأمثل لتحليل التباين

(نموذج ثابت r عدد المعاملات)

Power 1 – β = .90

r	Δ/σ = 1.0				Δ/σ = 1.25				Δ/σ = 1.50				Δ/σ = 1.75			
	α				α				α				α			
	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01
2	14	18	23	32	9	12	15	21	7	9	13	15	5	7	8	12
3	17	22	27	37	11	15	18	24	8	11	13	18	6	8	10	13
4	20	25	30	40	13	16	20	27	9	12	14	19	7	9	11	15
5	21	27	32	43	14	18	21	28	10	13	15	20	8	10	12	15
6	22	29	34	46	15	19	23	30	11	14	16	21	8	10	12	16
7	24	31	36	48	16	20	24	31	11	14	17	22	9	11	13	17
8	26	32	38	50	17	21	25	33	12	15	18	23	9	11	13	17
9	27	33	40	52	17	22	26	34	13	16	18	24	9	12	14	18
10	28	35	41	54	18	23	27	35	13	16	19	25	10	12	14	19

Power 1 – β = .90

r	Δ/σ = 2.0				Δ/σ = 2.5				Δ/σ = 3.0			
	α				α				α			
	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01
2	4	6	7	10	3	4	5	7	3	3	4	6
3	5	7	8	11	4	5	6	8	3	4	5	6
4	6	7	9	12	4	5	6	8	3	4	5	6
5	6	8	9	12	4	5	6	9	4	4	5	7
6	7	8	10	13	5	6	7	9	4	4	5	7
7	7	9	10	13	5	6	7	9	4	5	5	7
8	7	9	11	14	5	6	7	9	4	5	6	7
9	8	9	11	14	5	6	8	10	4	5	6	7
10	8	10	11	15	5	7	8	10	4	5	6	7

تكملة ...

... تابع جدول ١٨ – تحديد العدد الأمثل لتحليل التباين

(نموذج ثابت r عدد المعاملات)

Power $1-\beta=.95$																
	$\Delta/\sigma=1.0$				$\Delta/\sigma=1.25$				$\Delta/\sigma=1.50$				$\Delta/\sigma=1.75$			
r	α				α				α				α			
	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01
2	18	23	27	38	12	15	18	25	9	11	13	18	7	8	10	14
3	22	27	32	43	14	18	21	29	10	13	15	20	8	10	12	16
4	25	30	36	47	16	20	23	31	12	14	17	22	9	11	13	17
5	27	33	39	51	18	22	25	33	13	15	18	23	10	12	14	18
6	29	35	41	53	19	23	27	35	13	16	19	25	10	12	14	19
7	30	37	43	56	20	24	28	36	14	17	20	26	11	13	15	19
8	32	39	45	58	21	25	29	38	15	18	21	27	11	14	16	20
9	33	40	47	60	22	26	30	39	15	19	22	28	12	14	16	21
10	34	42	48	62	22	27	31	40	16	19	22	29	12	15	17	21

Power $1-\beta=.95$												
	$\Delta/\sigma=2.0$				$\Delta/\sigma=2.5$				$\Delta/\sigma=3.0$			
r	α				α				α			
	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01
2	5	7	8	11	4	5	6	8	3	4	5	6
3	6	8	9	12	5	6	7	9	4	4	5	7
4	7	9	10	13	5	6	7	9	4	5	5	7
5	8	9	11	14	5	6	7	10	4	5	6	7
6	8	10	11	15	6	7	8	10	4	5	6	8
7	8	10	12	15	6	7	8	10	4	5	6	8
8	9	11	12	16	6	7	8	11	5	5	6	8
9	9	11	13	16	6	8	9	11	5	6	6	8
10	9	11	13	17	6	8	9	11	5	6	7	8

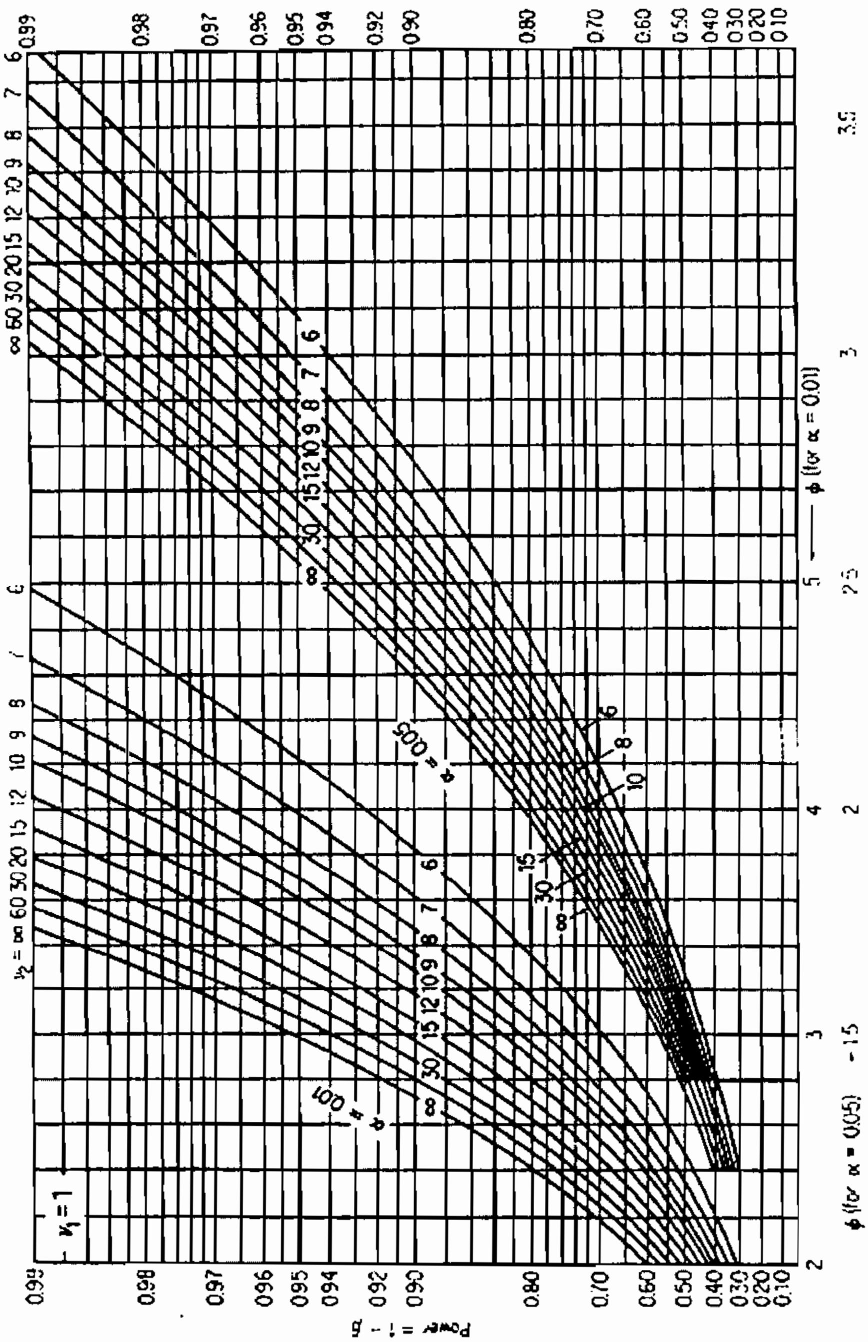
جدول ١٩ – الحروف اليونانية

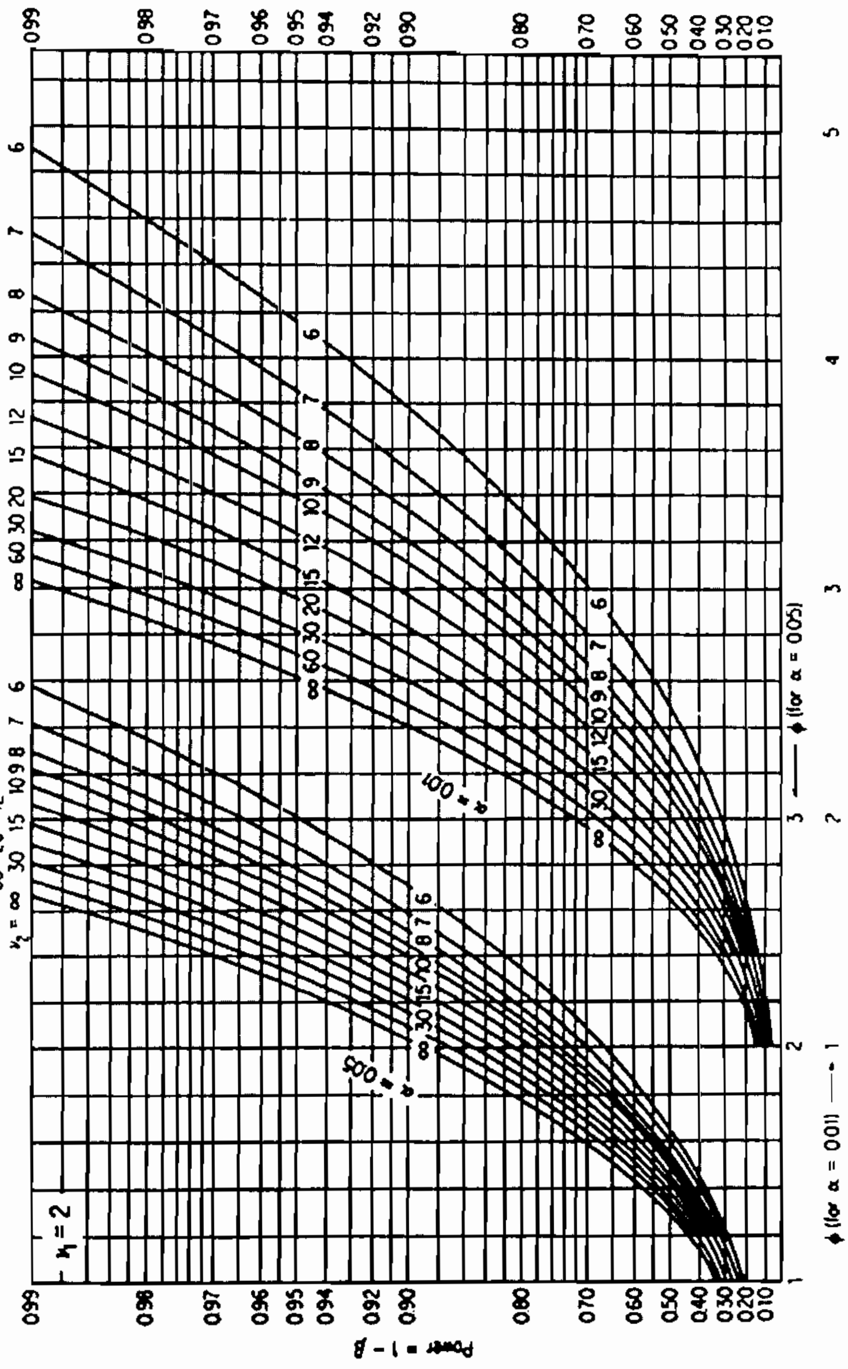
Greek letters

الحرف الكبير upper case	الحرف الصغير lower case	الاسم name	الحرف الكبير upper case	الحرف الصغير lower case	الاسم name
ν	Ν	Nu	Α	α	Alpha
ξ	Ξ	Xi	Β	β	Beta
ο	Ο	Omicorn	γ	Γ	Gamma
π	Π	Pi	δ	Δ	Delta
ρ	Ρ	Rho	ε	Ε	Epsilon
σ	Σ	Sigma	ζ	Ζ	Zeta
τ	Τ	Tau	η	Η	Eta
υ	Υ	Upsilon	θ	Θ	Theta
φ	Φ	Phi	ι	Ι	Iota
χ	Χ	Chi	κ	Κ	Kappa
ψ	Ψ	Psi	λ	Λ	Lambda
ω	Ω	Omega	μ	Μ	Mu

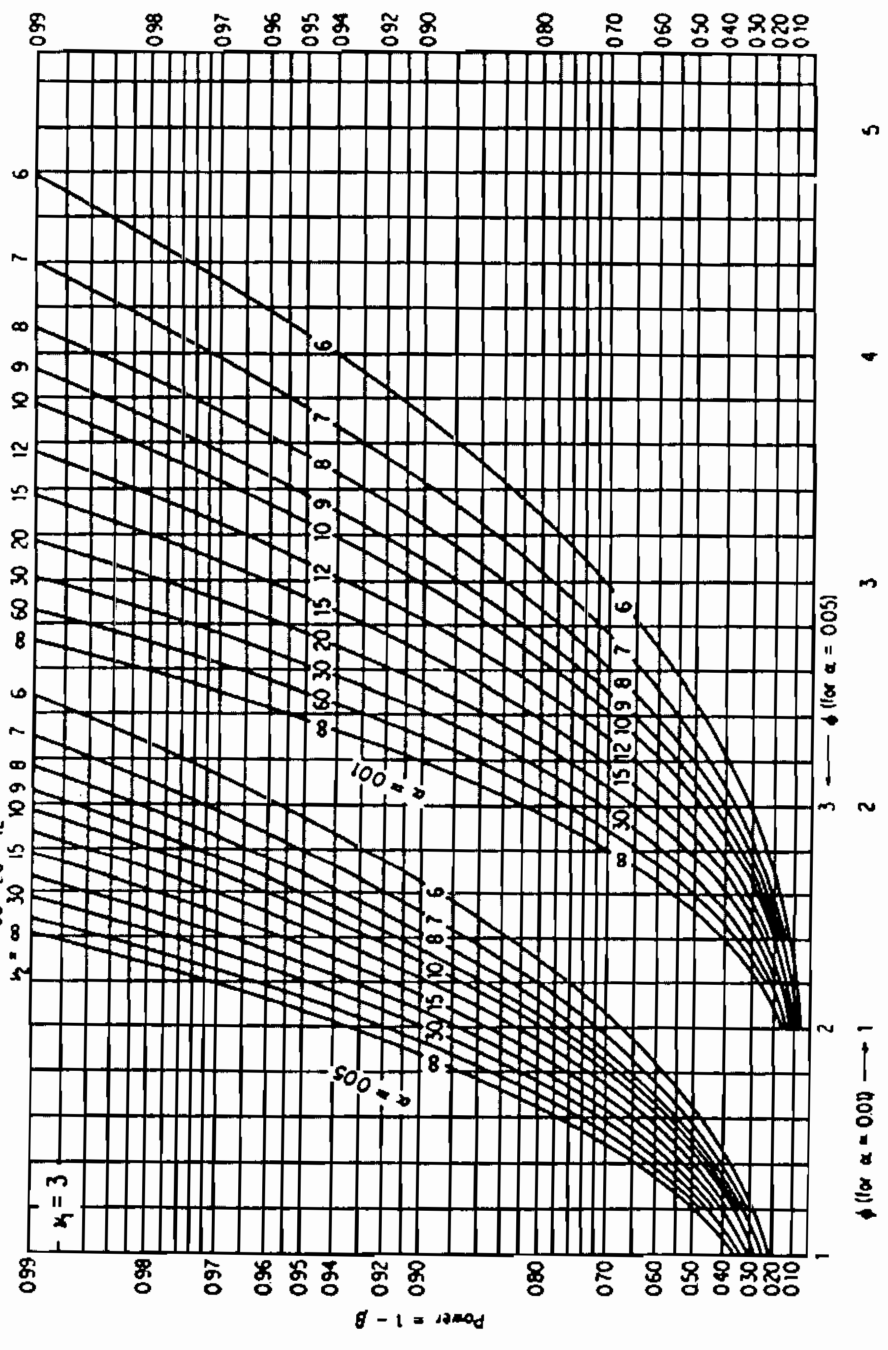
ملحق ب

أشكال تحديد الحجم الأمثل للتجربة

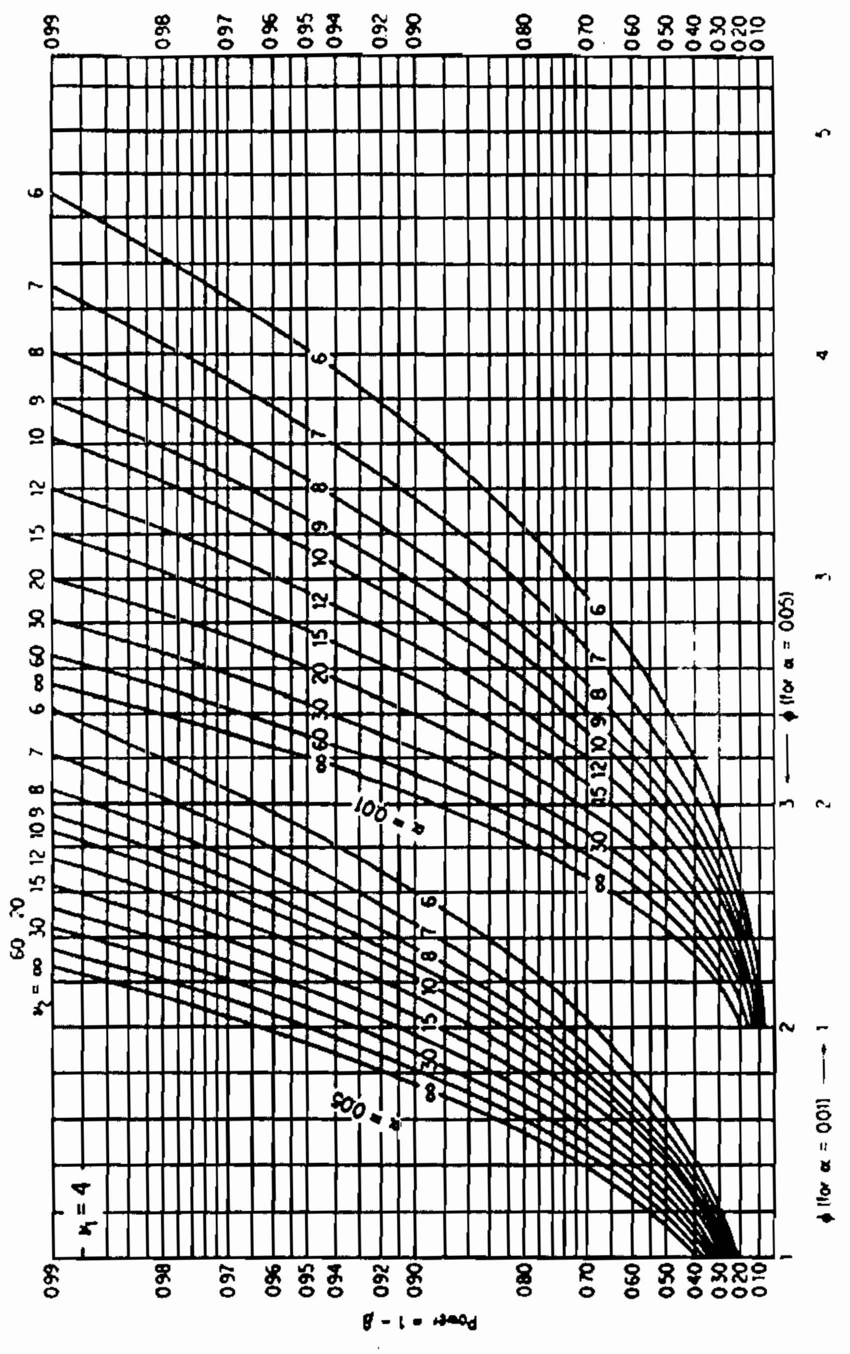

Power = 1 - β
φ (for α = 0.05)
φ (for α = 0.01)
$\nu_1 = 1$
ν_2 = ∞ 60 30 20 15 12 10 9 8 7 6
α = 0.01
α = 0.05



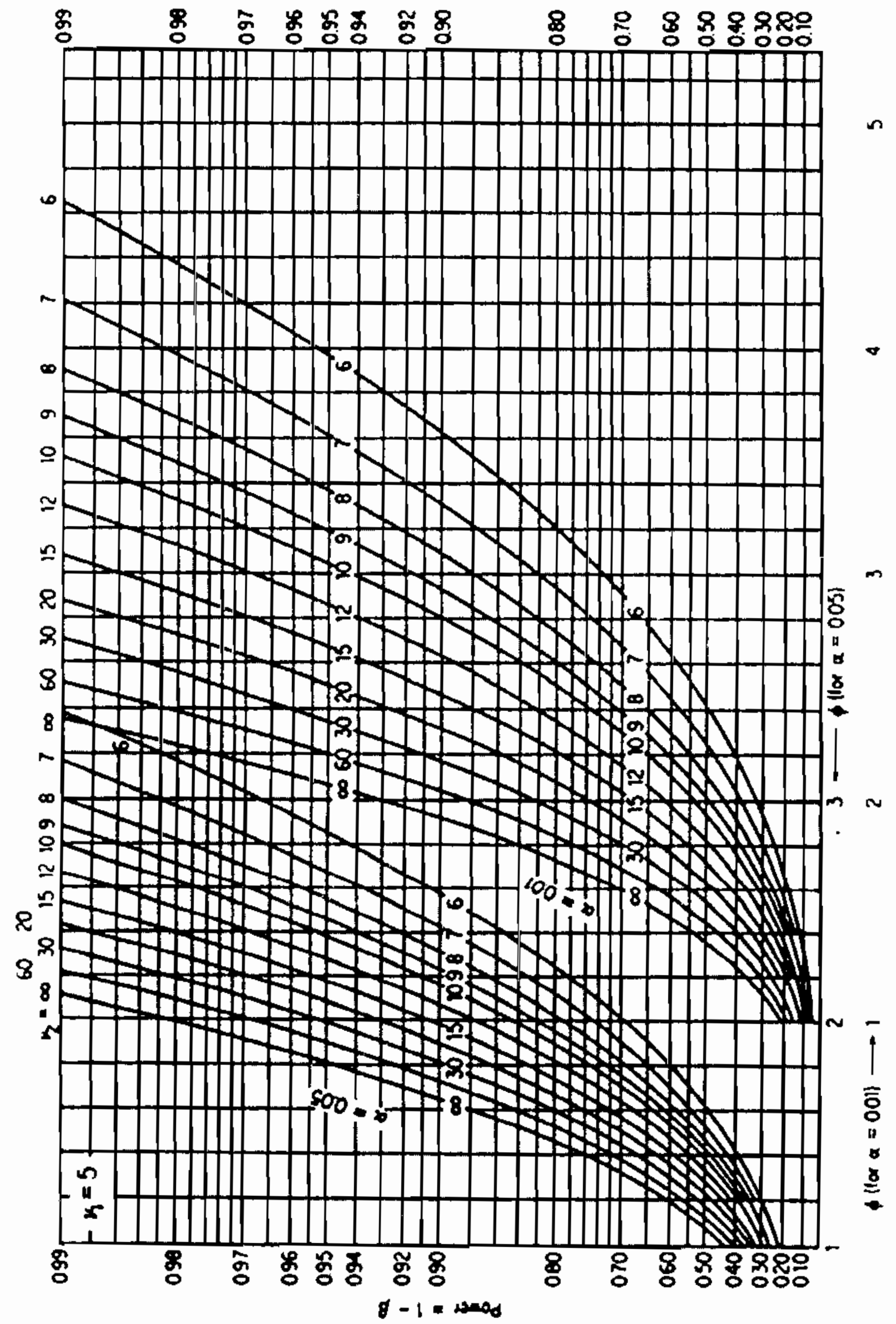
Power = 1 − β
$\nu_1 = 2$
$\nu_2 = \infty$
α = 0.05
α = 0.01
φ (for α = 0.01) ——→
φ (for α = 0.05)



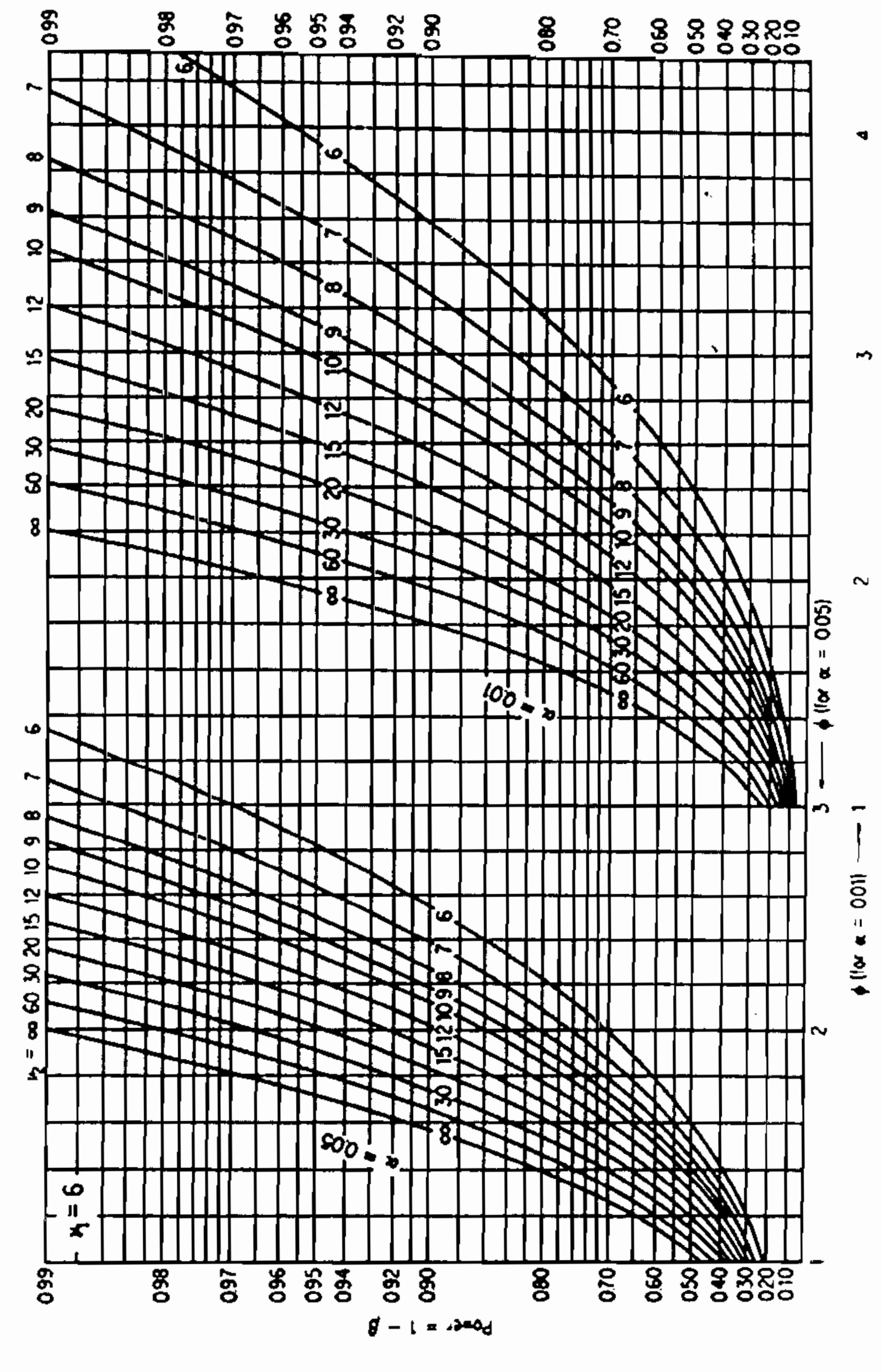
$\nu_1 = 3$
Power = 1 − β
φ (for α = 0.01)
φ (for α = 0.05)
α = 0.05
α = 0.01



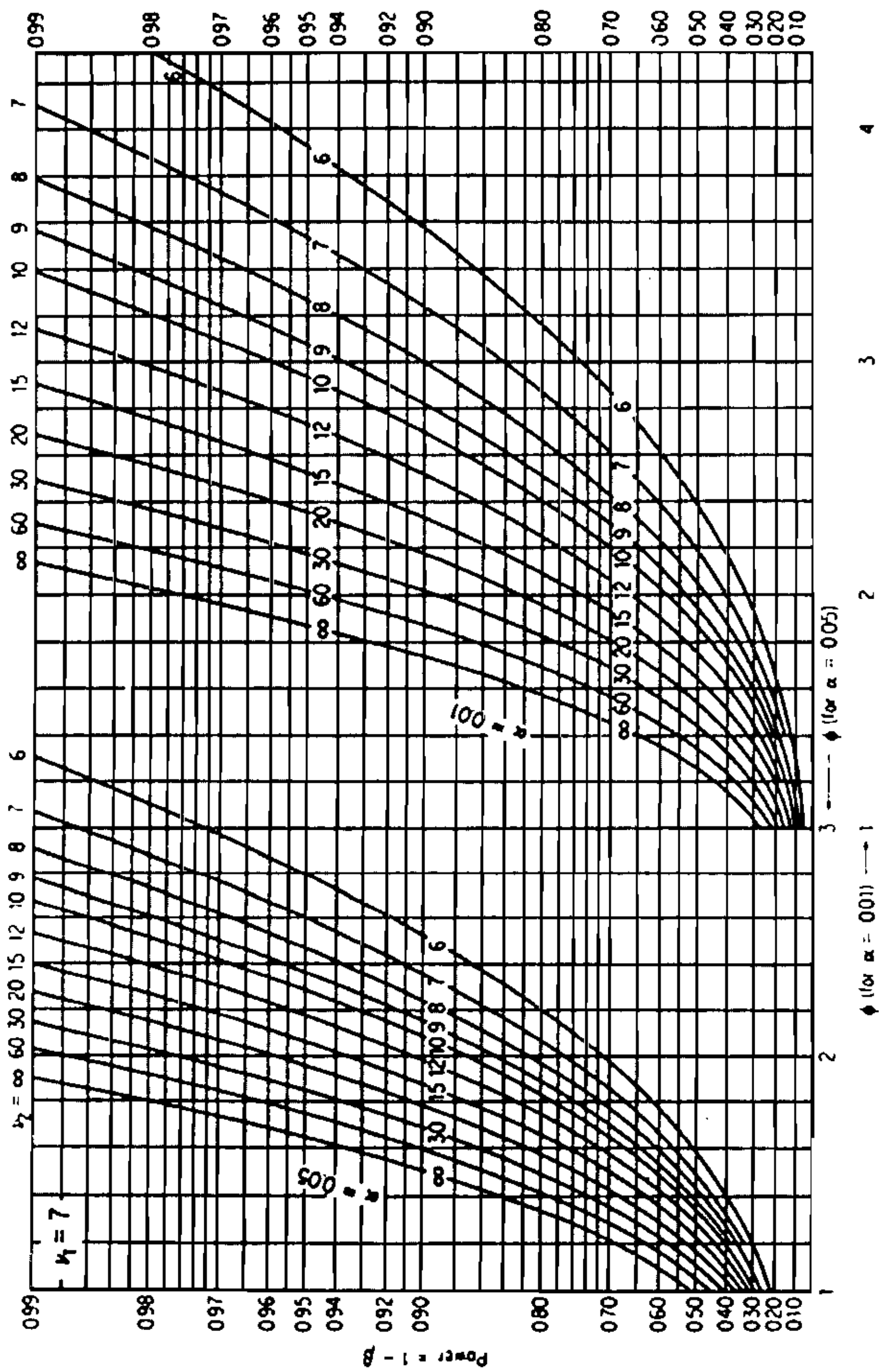
$\nu_1 = 4$
$\nu_2 = \infty$ 60 30 20 15 12 10 9 8 7 6
$\alpha = 0.05$
$\alpha = 0.01$
Power = 1 − β
φ (for α = 0.01)
φ (for α = 0.05)



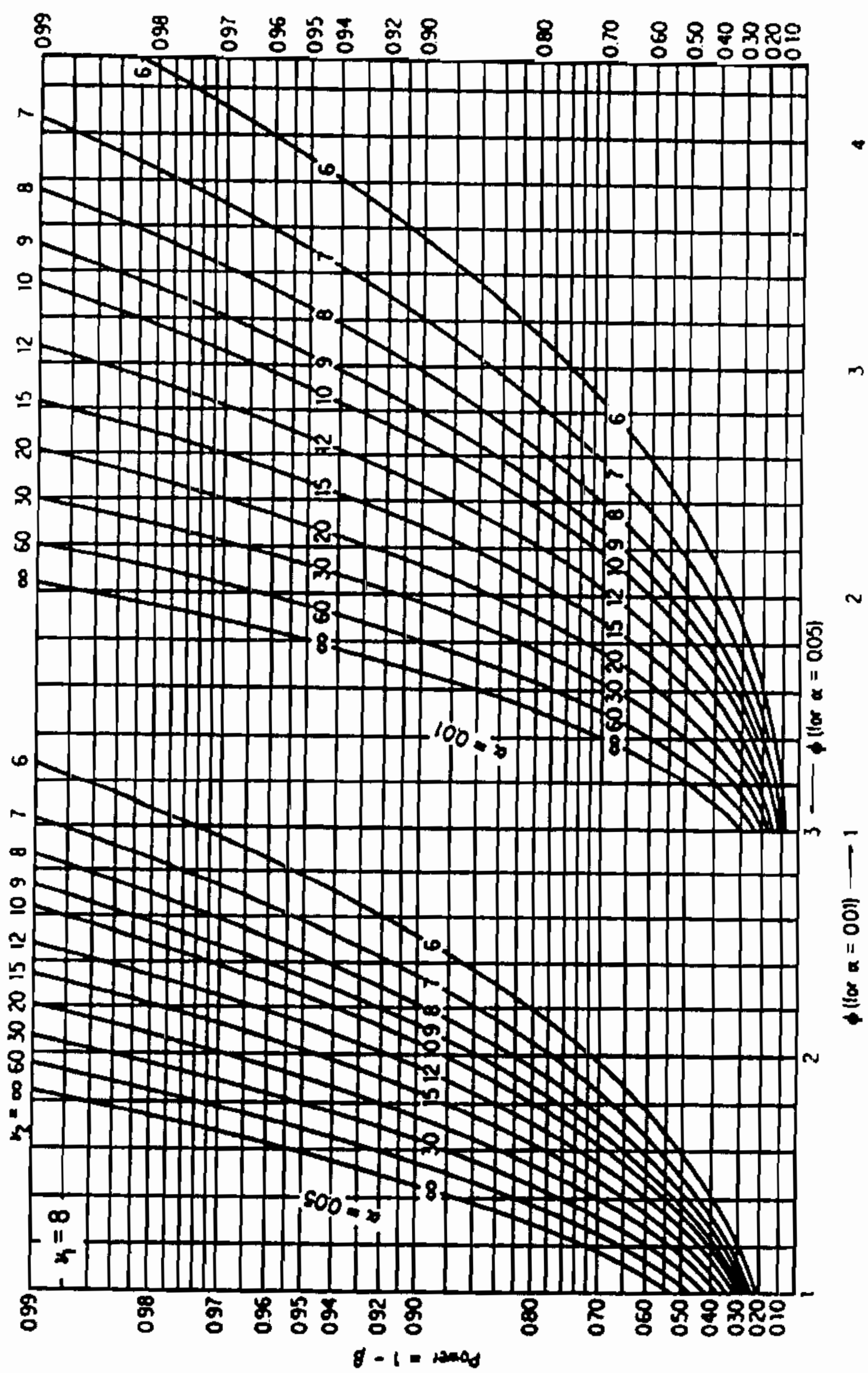
$\nu_1 = 5$
Power = 1 - β
φ (for α = 0.01)
φ (for α = 0.05)
α = 0.05
α = 0.01



Power = 1 − β
ϕ (for α = 0.01)
ϕ (for α = 0.05)
ν1 = 6
ν2 = ∞
α = 0.05
α = 0.01



$\nu_1 = 7$
$\nu_2 = \infty$
$\alpha = 0.05$
$\alpha = 0.01$
Power = 1 − β
φ (for α = 0.01)
φ (for α = 0.05)



$\nu_1 = 8$
Power = 1 - β
φ (for α = 0.01)
φ (for α = 0.05)
α = 0.05
α = 0.01

ملحق ج

مقدمة فى جبر المصفوفات

Introduction to matrix algebra

المصفوفة عبارة عن مجموعة من العناصر elements مرتبة على هيئة صفوف وأعمدة. وحجم المصفوفة order or dimension يعتمد على عدد الصفوف (r) وعدد الأعمدة (c) بها. إذا افترض أن عدد الصفوف = n وعدد الأعمدة = m فإنه يمكن التعبير عن المصفوفة كالتالى:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1,m-1} & a_{1,m} \\ a_{21} & a_{22} & \cdots & a_{2,m-1} & a_{2,m} \\ . & . & & . & . \\ . & . & \cdots & . & . \\ . & . & & . & . \\ a_{n-1,1} & a_{n-1,2} & \cdots & a_{n-1,m-1} & a_{n-1,m} \\ a_{n,1} & a_{n,2} & \cdots & a_{n,m-1} & a_{n,m} \end{pmatrix}$$

وقد تكون عناصر المصفوفة عبارة عن أعداد scalars أو تعبيرات رياضية mathematical expressions أو مصفوفات أخري matrices داخل مصفوفة كبيرة. والأمثلة التالية تعبر عن بعض صور لهذه المصفوفات:

$$A = \begin{pmatrix} 7 & 12 & -2 & 15 \\ -13 & 3 & 4 & 55 \\ 9 & 1 & -4 & 30 \end{pmatrix} ، \quad B = \begin{pmatrix} x & y+1 & x+y+z \\ a-b & (c)\log(b) & e \\ \sqrt{x-b} & (m+n)/n & h \end{pmatrix} ،$$

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$$

وقد تكون المصفوفة عبارة عن صف واحد فقط به عدة أعمدة وتسمى row vector أو عبارة عن عمود واحد فقط به عدة صفوف وتسمى column vector أو عبارة عن عنصر واحد فقط وتسمى scalar.

وعناصر القطر diagonal elements عبارة عن القيم التى تقع على قطر المصفوفة فقط أما باقى العناصر بالمصفوفة فتسمى عناصر خارج القطر off-diagonal elements. مثلا المصفوفة A_{3x3} التالية:

$$A = \begin{bmatrix} 5 & 3 & 2 \\ 7 & 1 & 4 \\ 4 & 6 & -1 \end{bmatrix}$$

قيم القطر بهذه المصفوفة هى 5، 1، 1- أما القيم الأخرى بالمصفوفة تعتبر قيم خارج القطر.

أنماط المصفوفات Types of matrices

المصفوفة المربعة Square matrix هى المصفوفة التى يتساوى بها عدد الصفوف والأعمدة.

المصفوفة المستطيلة Rectangular matrix هى المصفوفة التى لا يتساوى بها عدد الصفوف مع عدد الأعمدة.

المصفوفة المتماثلة أو المتناسقة Symmetric matrix هى مصفوفة مربعة يتساوى فيها العنصر الموجود بالصف i والعمود j مع العنصر الموجود بالصف j والعمود i .

المصفوفة القطرية Diagonal matrix هى مصفوفة مربعة ومتماثلة جميع العناصر بها مساوية للصفر ماعدا العناصر القطرية. ويمكن أن تكتب هذه المصفوفة بالصورة التالية: $D = diag(d_1\, d_2 \ldots d_n)$، حيث d_i تعبر عن قيم العناصر الموجدة على قطر المصفوفة.

مصفوفة الوحدة Identity matrix هى مصفوفة مربعة متماثلة جميع قيم العناصر بها تساوى صفر فيما عدا القيم القطرية و التي جميعها عبارة عن الواحد الصحيح. ويرمز لهذه المصفوفة بالرمز I. وهى حالة خاصة للمصفوفة القطرية.

المصفوفة الصفرية Null matrix هى مصفوفة مكونة من أى عدد من الصفوف والأعمدة ولكن جميع قيم العناصر بها مساوية للصفر. ويرمز لها بالرمز 0.

مصفوفة الواحدات Only 1's matrix هى مصفوفة مكونة من أى عدد من الصفوف والأعمدة ولكن جميع قيم العناصر بها مساوية للواحد الصحيح. ويرمز لها بالرمز J

مصفوفة مثلثية Triangular matrix هى مصفوفة مربعة بها أما جميع العناصر الموجودة أعلى القطر أو الموجودة أسفل القطر مساوية للصفر، وفى

الحالة الأولى تسمى مصفوفة مثلثية دنيا lower triangular matrix وفي الحالة الثانية تسمى مصفوفة مثلثية عليا upper triangular matrix.

مصفوفة ثلاثية الأقطار Tridiagonal matrix وهي مصفوفة مربعة جميع قيمها مساوية للصفر باستثناء القيم الموجدة على القطر والقيم الموجودة مباشرة على يمين ويسار القطر فقط.

مقلوب المصفوفة Transposition

ويقصد به تحويل المصفوفة الأصلية إلى مصفوفة جديدة عن طريق تحويل صفوف المصفوفة الأصلية إلى أعمدة في المصفوفة الجديدة، وبالتالي إذا كان حجم المصفوفة الأصلية $n \times m$ تصبح المصفوفة الجديدة $m \times n$ وإذا كان رمز المصفوفة الأصلية A فإن المصفوفة الجديدة يرمز لها بالرمز A'.

جمع وطرح المصفوفات Addition and subtraction of matrices

يمكن جمع و/أو طرح المصفوفات فقط في حالة ما إذا تساوى حجم المصفوفات المراد جمعها أو طرحها، ونتيجة الجمع أو الطرح عبارة عن مصفوفة لها نفس حجم المصفوفات محل الجمع أو الطرح. بمعنى إذا كانت $A = \{a_{ij}\}$ $B = \{b_{ij}\}$ فإن:

$$A \pm B = \{a_{ij} \pm b_{ij}\}$$

مثال:

$$A = \begin{pmatrix} 4 & 5 & 3 \\ 6 & 0 & 2 \end{pmatrix} \text{، } B = \begin{pmatrix} 1 & 0 & 2 \\ 3 & 4 & 1 \end{pmatrix}$$

$$A + B = \begin{pmatrix} 4+1 & 5+0 & 3+2 \\ 6+3 & 0+4 & 2+1 \end{pmatrix} = \begin{pmatrix} 5 & 5 & 5 \\ 9 & 4 & 3 \end{pmatrix} = B + A$$

الطرح هو الجمع مع ملاحظة أن إحدى المصفوفتين تم ضرب جميع عناصرها في (-1) بمعنى:

$$A + (-1)B = \begin{pmatrix} 3 & 5 & 1 \\ 3 & -4 & 1 \end{pmatrix}$$

ضرب المصفوفات Multiplication of matrices

يمكن ضرب المصفوفات فقط في حالة ما إذا تساوى عدد الأعمدة في المصفوفة الأولى مع عدد الصفوف في المصفوفة الثانية. فمثلا إذا كانت المصفوفة C حجمها $p \times q$ والمصفوفة D حجمها $m \times n$، فإنه يمكن ضرب المصفوفتين فقط في حالة ما

إذا كانت $q = m$ وبالتالى يكون حجم المصفوفة الناتجة من ضرب المصفوفتين $C \times D$ هو $p \times n$. لاحظ بصفة عامة أن $C \times D \neq D \times C$، وغالبا لا يمكن عمل $D \times C$ نتيجة لعدم التوافق. بمعنى:

إذا كانت $C_{p \times q} = \{C_{ij}\}$، $D_{m \times n} = \{d_{ij}\}$ مع توافر شرط أن $q = m$ فإن:

$$CD_{pxn} = \left\{ \sum_{k=1}^{m} c_{ik} d_{kj} \right\}$$

مثال

$$D = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 3 & -1 \end{pmatrix} ، C = \begin{pmatrix} 6 & 4 & -3 \\ 3 & 9 & -7 \\ 8 & 5 & -2 \end{pmatrix}$$

فإن حاصل ضرب C x D:

$$CD = \begin{pmatrix} 6(1)+4(2)-3(3) & 6(1)+4(0)-3(-1) \\ 3(1)+9(2)-7(3) & 3(1)+9(0)-7(-1) \\ 8(1)+5(2)-2(3) & 8(1)+5(0)-2(-1) \end{pmatrix} = \begin{pmatrix} 5 & 9 \\ 0 & 10 \\ 12 & 10 \end{pmatrix}$$

مقلوب المصفوفة Matrix inversion

مقلوب المصفوفة المربعة A، والذى يرمز له بالرمز A^{-1}، عبارة عن مصفوفة عندما يتم ضربها أو يضرب فيها المصفوفة الأصلية ينتج مصفوفة الوحدة I. بمعنى أن:

$$A^{-1}A = I \quad و \quad AA^{-1} = I$$

ويمكن قلب أى مصفوفة بالطريقة المعتادة عندما يكون محددها determinant لا يساوى الصفر. أما إذا كان المحدد يساوى صفر فهناك طرق أخرى خارج نطاق هذا الكتاب للحصول على المقلوب، ومن أشهرها الحصول على المقلوب العام generalized inverse. ويوجد برامج كثيرة للحاسب تمكن من الحصول على المقلوب

ومنها برنامج SAS عن طريق استخدام PROC IML. والمثال التالى يوضح كيفية الحصول على المقلوب.

للحصول على المقلوب يتم حساب المحدد أولاً، وقاعدة حساب المحدد هى:

المصفوفة $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ محددها هو $|A| = a_{11}a_{22} - a_{21}a_{22}$، وإذا كانت المصفوفة أكبر من 2 x 2 فالقاعدة العامة للحصول على المحدد هى:

$$|A| = \sum_{J=1}^{n} (-1)^{i+j} a_{ij} |M_{ij}|$$

حيث n عبارة حجم المصفوفة A، M_{ij} عبارة عن تحت مصفوفة ناتجة من حذف صف i وعمود j من المصفوفة A. لاحظ أنه يمكن استخدام أى صف لأن النتيجة سوف تكون واحدة. فإذا كانت A عبارة عن:

$$A = \begin{pmatrix} 6 & -1 & 2 \\ 3 & 4 & -5 \\ 1 & 0 & -2 \end{pmatrix}$$

فإن محددها يكون

$$\begin{aligned} |B| &= 6\begin{vmatrix} 4 & -5 \\ 0 & -2 \end{vmatrix} - 1(-1)\begin{vmatrix} 3 & -5 \\ 1 & -2 \end{vmatrix} + 2\begin{vmatrix} 3 & 4 \\ 1 & 0 \end{vmatrix} \\ &= 6(-8) + 1(-1) + 2(-4) \\ &= -57 \end{aligned}$$

وبالتالى يمكن الحصول على مقلوب المصفوفة A. ويتم ذلك عن طريق الحصول على عدد من تحت المصفوفات تسمى signed minors والتى سوف تؤدى إلى الحصول على ما يعرف بـــ adjoint matrix ومنها يمكن الحصول على المقلوب كالتالى:

$M_{11} = +1\begin{pmatrix} 4 & -5 \\ 0 & -2 \end{pmatrix}$ ومنها $|M_{11}| = -8$،

$$M_{12}=-1\begin{pmatrix}3 & -5\\ 1 & -2\end{pmatrix}$$ ومنها $|M_{12}|=+1$ ،

$$M_{13}=+1\begin{pmatrix}3 & 4\\ 1 & 0\end{pmatrix}$$ ومنها $|M_{13}|=-4$ ،

$$M_{21}=+1\begin{pmatrix}-1 & 2\\ 0 & -2\end{pmatrix}$$ ومنها $|M_{21}|=-2$ ،

$$M_{22}=+1\begin{pmatrix}6 & 2\\ 1 & -2\end{pmatrix}$$ ومنها $|M_{22}|=-14$ ،

$$M_{23}=-1\begin{pmatrix}6 & -1\\ 1 & 0\end{pmatrix}$$ ومنها $|M_{23}|=-1$ ،

$$M_{31}=+1\begin{pmatrix}-1 & 2\\ 4 & -5\end{pmatrix}$$ ومنها $|M_{31}|=-3$

$$M_{32}=-1\begin{pmatrix}6 & 2\\ 3 & -5\end{pmatrix}$$ ومنها $|M_{32}|=-36$ ،

$$M_{33}=+1\begin{pmatrix}6 & -1\\ 3 & 4\end{pmatrix}$$ ومنها $|M_{33}|=+27$

وبالتالى تكون الـــ adjoint matrix

$$M_A=\begin{pmatrix}-8 & 1 & -4\\ -2 & -14 & -1\\ -3 & 36 & 27\end{pmatrix}$$

والمقلوب عبارة عن:

$$A^{-1}=|A|^{-1}M'_A$$

وبالتعويض يكون مقلوب المصفوفة A هو

$$A^{-1} = \frac{1}{-57}\begin{pmatrix} -8 & -2 & -3 \\ 1 & -14 & 36 \\ -4 & -1 & 27 \end{pmatrix}$$

ولمزيد من المعلومات عن جبر المصفوفات يمكن الرجوع إلى Searle (1982).

المراجع References

١- مراجع باللغة العربية

٢ - مراجع باللغة الإنجليزية

١- مراجع باللغة العربية

- أحمد عبادة سرحان (١٩٦٥): مذكرات فى الإحصاء البيولوجى، دار المعارف.
- أحمد عبادة سرحان (١٩٦٨): مقدمة فى طرق التحليل الإحصائى، دار الكتب الجامعية، الإسكندرية.
- أحمد عبادة سرحان وثابت محمود أحمد (١٩٦٩): تصميم وتحليل التجارب، دار الكتب الجامعية، الإسكندرية.
- السيد سعد قاسم ولطفى هندى (١٩٦٤): مبادئ الإحصاء التجريبى، دار المعارف.
- السيد سعد قاسم، مسعد زكى الحفنى، عبد المجيد أبو المجد وجمال الدين حسن (١٩٧٥): الإحصاء التطبيقى فى العلوم الزراعية، دار المعارف.
- صلاح جلال، عصام الطويل وعبد الحليم عشماوى (١٩٨٨): الإحصاء الحيوى ومقدمة فى تصميم التجارب (الجزء الأول والجزء الثالث) مركز التنمية البشرية والمعلومات، القاهرة.
- صلاح جلال، عصام الطويل وعبد الحليم عشماوى (١٩٨٩): لإحصاء الحيوى ومقدمة فى تصميم التجارب (الجزء الثانى) مركز التنمية البشربة والمعلومات، القاهرة.
- محمد على بشر ومحمد ممدوح الروبى (١٩٧٠): مقدمة فى طرق الإحصاء وتصميم التجارب، دار المعارف، الإسكندرية.
- محمد محمد أبو العلا ومحمد على بشر (١٩٦٧): مبادئ الإحصاء وتصميم التجارب، دار المعارف، الإسكندرية.

٢- المراجع باللغة الإنجليزية

- Allen, D. B., J. H. Burton and J. D. Hott (1983). J. Anim. Sci., 57: 765.

- Balam, L. N. (1963). J. Stat., 5: 62.

- Bartlett, M. S. (1937). J. R. Stat. Soc., 4: 137.

- Chatterjee, S. and B. Price (1991). Regression Analysis by Example. 2nd ed. John Wiley & Sons, New York.

- Cochran, W. G. (1941). Ann. Eugen., 11: 47.

- Cochran, W. G. (1977). Sampling Techniques. 3rd ed. John Wiley & Sons, New York.

- Cochran, W. G. and G. M. Cox (1957). Experimental Designs. 2nd ed., John Wiley & Sons, New York.

- Davis, C. D. (2002). Statistical methods for the Analysis of Repeated Measurements. Springer-Verlag, New York.

- Draper, N. and H. Smith (1981). Applied Regression Analysis. 2nd ed. John Wiley & Sons, New York.

- Duncan, D. B. (1955). Biometrics, 11: 1.

- EXCEL (2002). EXCEL for Windows, Microsoft Corporation, Calif. USA.

- FAO (2007). FAOSTAT, Available online at http://www.faostat.fao.org, Accessed Jan., 2008.

- Federer, W. T. (1963). Experimental Designs. McMillan Company, New York.

- Fisher, R. A. (1921). Metron, 1: 3.

- Fisher R. A. (1966) The Design of Experiments. 8th ed., Hafner Publishing, New York.

- Fisher, R. A. and F Yates (1949). Statistical Tables for Biological, Agriculture and Medical Research, Oliver and Boyd.

- Gaylor, D. W. and T. D. Hartwell (1969). Biometrics, 25: 427.

- Geisser, S. and S. W. Greenhouse (1958). Annals of Mathematical Statistics, 29: 885-891.

- Gill, J. L. (1973). J. Dairy Sci., 56: 973.

- Hajek, J. (1960). Math. Inst. Hungarian. Acd. Sci. 5: 361.

- Hartley, H. (1950). Biometrica, 37: 308.

- Hill, W. G. (1982). Unpublished data. Edinburgh. U.K.

- Hinkelmann, K. and O. Kempthorn (2005). Design and Analysis of Experiments: Advanced Experimental Design. Vol. 2, John Wiley & Sons, New York.

- Hog, R. V. and A. T. Craig (1978). Introduction to mathematical statistics, McMillan Company, New York.

- Johanson, J. (1919). Genetics., 4:307.

- Kemp, K. E. (1975). J. Dairy Sci., 58:1374.

- Kendall, M. G. (1970). Rank Correlation Methods. 3rd ed. Charles Griffin, London.

- Kirk, R. E. (1995). Experimental Design: Procedures for the Behavioral Sciences. 3rd ed. Monterey Calf.: Brooks/Cole Publishing.

- Kitagarwa, T. and M. Mitome (1953). Tables for the design of Factorial Experiments. Baifukam Co., Ltd. Tokyo, Japan.

- Nelder, J. A. (1976). Hypothesis testing in linear models. The Amr. Stat. 30: 103.

- Neter, J., M. H. Kutner, C. J. Nachtsheim and W. Wasserman (1996). Applied Linear Statistical Models. 4th ed. Richard D. Irwin, Inc. Chicago, Ill., USA.

- Ostle, B. (1963). Statistics in Research. 2nd ed. The Iowa State University Press, Ames, Iowa, USA.

- Rowell, J. G. and D. E. Watters (1976). J. Agric. Res. 87: 423.
- SAS (1999). SAS/STAT User's guide. Release 8 edition. Cary, N.C.
- Satterthqaite, F. E. (1946). Biom. Bull. 2: 110.
- Searle, S. R. (1971). Linear Models. John Wiley & Sons, New York.
- Searle, S. R. (1982). Matrix Algebra Useful for Statistics. John Wiley & Sons, New York.
- Searle, S. R. (1987). Linear Models for Unbalanced Data. John Wiley & Sons, New York.
- Snedecor, G. W. and W. G. Cochran. (1987). Statistical Methods. 4th Printing, The Iowa State University Press, Ames, Iowa, USA.
- SPSS (2002). User's Guide, Release 11.50.0, SPSS Inc., Chicago, IL. USA.
- Steel, R. G. an J. H. Torrie (1980). Principles and procedures of Statistics. A Biometrical Approach. 2nd ed., McGraw-Hill Company, Inc., USA.
- Tang, P. C. (1938). Stat. Res. Mem., 2: 126.
- Tippett, L. H. C. (1926). Biometrica , 17: 364.
- Tukey, J. W. (1949). Biometrics, 5: 232.
- WHO (2007). World Health Organization, Data and statistics, Available online at http://www.who , Accessed Jan., 2008.
- Yamane, T. (1967). Elementary Sampling Theory. Prantice-Hall Inc , Englewood. Cliffs, N.J., USA.

فهرس موضوعى أبجدى

عن المؤلفين

د. عبد الحليم عشماوى. عمل معيدا فى جامعة عين شمس بعد تخرجه من كلية الزراعة جامعة القاهرة عام ١٩٦٨. حصل على دبلوم الدراسات العليا فى الإحصاء ١٩٧٤ من جامعة القاهرة وماجستير ١٩٧٥ و دكتوراه ١٩٨١ فى تربية الحيوان من جامعة عين شمس. تدرج فى الألقاب العلمية من معيد إلى أستاذ عام ١٩٩١ بجامعة عين شمس والتى مازال يعمل بها حتى الآن. خلال عمله بجامعة عين شمس سافر فى مهمة علمية ١٩٨٢ كأستاذ زائر لمدة عام فى معهد وراثة الحيوان بأدنبرة، المملكة المتحدة. كما درس الإمكانيات المتطورة للحاسبات ونظم المعلومات فى كل من هولندا والولايات المتحدة الأمريكية من خلال برامج التبادل الثقافى. عمل أستاذ للإحصاء الحيوى بجامعة الملك عبد العزيز بالمملكة العربية السعودية ١٩٩٤-٢٠٠٢. أشرف على العديد من طلبة الماجستير والدكتوراه وله العديد من الأبحاث المنشورة فى مجلات علمية عالمية ومحلية. حائز على جائزة الدولة التشجيعية فى العلوم الزراعية عام ١٩٩٢ ونوط الامتياز من الطبقة الأولى عام ١٩٩٥ من جمهورية مصر العربية.

د. صلاح جلال. تخرج فى جامعة الإسكندرية عام ١٩٥٧، تخصص إنتاج حيوانى وعمل معيدا بنفس الجامعة إلى أن أوفد فى بعثة إلى الولايات المتحدة الأمريكية عام ١٩٥٩. حصل على ماجستير تربية حيوان من جامعة تكساس A&M ثم الدكتوراه فى تربية الحيوان من جامعة ولاية أيوا عام ١٩٦٥ بالولايات المتحدة الأمريكية وكانت الإحصاء تخصصه الثانى. عمل باحثاً بمعهد الصحراء بمصر حتى عام ١٩٨٦ حيث التحق بكلية الزراعة جامعة عين شمس متدرجاً فى الألقاب بها من مدرس إلى أستاذ، وهى نفس الكلية التى مازال يعمل بها حتى الآن. عمل فى مهمة علمية ١٩٧٢-١٩٧٣ كأستاذ زائر بقسم الإحصاء فى جامعة ويست فيرجينيا بالولايات المتحدة الأمريكية. أعير إلى منظمة الأغذية والزراعة FAO للعمل كخبير فى تربية الحيوان ١٩٧٥-١٩٨٢ ثم الخبير الإقليمى لإنتاج وصحة الحيوان بالمكتب الإقليمى للشرق الأدنى ١٩٩٢-١٩٩٦ ثم المقر الرئيسى للمنظمة بروما ١٩٩٦-٢٠٠٠ كخبير للموارد الوراثية الحيوانية. عمل فى مهمات استشارية فى تربية وإنتاج الحيوان فى بلدان عديدة فى أفريقيا، أسيا، أوربا وأمريكا. عمل فى مهمات استشارية عديدة فى الهيئات التنفيذية والبحثية والأكاديمية بمصر. قام بتدريس مواد تربية الحيوان ووراثة العشائر والإنتاج الحيوانى والإحصاء فى الجامعات المصرية وبعض الجامعات الأمريكية. أشرف على عشرات من طلاب الدراسات العليا فى الماجستير والدكتوراه مصريين وعرب وأفارقة وله ما يزيد على ١٠٠ بحثا منشورا فى الدوريات العلمية العالمية والإقليمية والمصرية. عضو فى الجمعيات والمحافل العلمية والأكاديمية والمهنية فى مصر وأوروبا والولايات المتحدة الأمريكية.

د. محمد حسين صادق. تخرج فى جامعة عين شمس عام ١٩٧٢ تخصص إنتاج حيوانى وعمل معيداً فى نفس الجامعة. أوفد فى بعثة إلى جامعة ولاية أيوا بالولايات المتحدة الأمريكية عام ١٩٨٣ للحصول على دكتوراه فى تربية ووراثة الحيوان كتخصص رئيسى وإحصاء حيوى كتخصص فرعى. تدرج فى الألقاب العلمية حتى لقب أستاذ بجامعة عين شمس والتى مازال يعمل بها حتى الآن. عمل كأستاذ مساعد بجامعة الملك فيصل بالمملكة العربية السعودية ١٩٩٥-١٩٩٩. عمل مستشاراً للعديد من الشركات الخاصة والحكومية فى مصر والسعودية. قام بتدريس العديد من مقررات الإحصاء وتربية ووراثة الحيوان والإنتاج الحيوانى وتطبيقات الحاسب الآلى فى مجال الإنتاج الحيوانى والتحليل الإحصائى فى مصر والسعودية. قدم استشارات إحصائية لكثير من الباحثين فى المجالات الحيوية المختلفة فى مصر والسعودية. اشترك فى العديد من المشاريع البحثية الممولة من جهات محلية ودولية مع تولى مسئولية التحليل الإحصائى لهذه المشاريع. أشرف على العديد من طلاب الماجستير والدكتوراه مصريين وعرب وله العديد من الأبحاث المنشورة فى الدوريات العلمية العالمية والمصرية وهو عضو فى العديد من الجمعيات العلمية.