الإحصاء التطبيقى

دكتورة
سهير فهمى حجازى
أستاذ الإحصاء

دكتور
محمود الدرينى
مدرس الإحصاء

كلية التجارة - جامعة طنطا

الإحصاء التطبيقي

دكتورة
سمير فهمي حجازي
أستاذ الإحصاء
كلية التجارة - جامعة طنطا

دكتور
محمود الدريني
مدرس الإحصاء
كلية التجارة - جامعة طنطا

الشركة المصرية لإعادة التأمين
المكتبة

٢٠٠٣-٢٠٠٤

بسم الله الرحمن الرحيم

تقديم

يتضمن هذا الكتاب الطرق والأساليب الإحصائية التي تستخدم بكثرة في الأبحاث العلمية والتطبيقية في مجالات (على سبيل المثال لا الحصر) الاقتصاد والتجارة ، والزراعة ، والتعليم ، وإدارة الإنتاج ، وإدارة الأفراد ، إلخ

تتدرج موضوعات هذا الكتاب لتشمل أساسيات الإحصاء التطبيقي ، والإحصاء الاستدلالي . يتضمن عرض الفكرة التي يستند عليها مفهوم توزيعات المعاينة ، ثم يتناول موضوع التقدير واختبار معلمات المجتمع : الوسط الحسابي ، النسبة ، الفرق بين وسطي مجتمعين مستقلين وغير مستقلين . ثم موضوع تحليل التباين ، الذي اكتفى بعرضه في اتجاه واحد . في هذا الكتاب أيضاً ، تم عرض بعض الاختبارات اللامعلمية (كا²) لاختبار الاستقلال بين متغيرين ، واختبارات جودة المطابقة . وقد تم تخصيص البابين الأخيرين لعرض أسلوب تحليل الارتباط والانحدار الخطي البسيط والمتعدد ، واختبارات المعنوية الخاصة بهما

يتميز هذا الكتاب بتنوع تطبيقاته في مجالات متعددة ، ويعطي عرضاً متدرجاً لكيفية تناول مشكلة إحصائية معينة ، وكيفية اتخاذ قرار بشأنها

تميز أيضاً أسلوب عرض الكتاب بربطه ببرمجة ميني تاب MiniTab Package ، حيث يتم عرض الأسلوب الإحصائي ومعادلاته الحسابية ، ثم كيف يمكن اختصار جهد العمليات الحسابية باستخدام برمجيات في الإحصاء . وقد تم اختيار

الحزمة الإحصائية Minitab لسهولة استخدامها وسهولة تفسير مخرجاتها ، ولأن مخرجاتها تتوافق مع أسلوب عرض هذا الكتاب بشكل يجعل فهم جميع مخرجات برامجها أمراً سهلاً . وسوف يتم شرح هذه المخرجات وربطها بالمعادلات التي تم تغطيتها في أسلوب التحليل اليدوي . وهذا يجعل القارئ يكتسب الإحساس والتصور للحسابات الآلية ويستطيع أن يحكم عليها بالجودة وتتميز بالدقة والسرعة ، ويكون باستطاعته تحديد الأسلوب الإحصائي المناسب لمشكلته ، وفهم نتائج تحليل البيانات وتفسير مدلولاتها .

ويتطلب فهم معظم موضوعات هذا الكتاب ، معرفة القارئ بمبادئ علم الإحصاء الوصفي ، ومبادئ حسابات نظرية الاحتمالات .

وقد قام المؤلف الأول بعرض الجزء الأول والذي يتضمن الأبواب من الأول حتى الباب الخامس ، وقد قام المؤلف الثاني بعرض الجزء الثاني والذي يتضمن البابين السادس والسابع .

والله نسأل أن نكون قد وفقنا في عرض موضوعات هذا الكتاب .

وعلى الله قصد السبيل ...

سبتمبر ٢٠٠٣

المؤلفان

الجزء الأول

تأليف

أ.د/ سمير فهمي حجازي

أستاذ الإحصاء

الباب الأول

التقدير

ESTIMATION

الباب الأول

التقديــــــر

ESTIMATION

يعرض هذا الباب أسس الإحصاء الاستنباطي والإحصاء الذي يلجأ إليه الباحث ويستخدمها في التعرف على مؤشرات المجتمع باستخدام نتائج العينة ، فأسلوب الاستنباط هام جداً ، حيث أنه يصعب في معظم المواقف دراسة جميع مفردات المجتمع للتعرف على مؤشراته ومعالمه Parameters لأسباب تتعلق بالوقت والجهد والتكلفة.

وينقسم الإحصاء الاستنباطي فرعين هما : التقدير واختبارات الفروض. ويهدف التقدير إلى حساب تقريبي إلى قيمة المعلمة في المجتمع باستخدام الإحصاء الذي تم حسابه من العينة ، والذي يسمى إحصاء العينة Sample Statistics . أما اختبارات الفروض فتهدف إلى التوصل إلى نتائج وقرارات خاصة بمعلمة المجتمع التي يعتقد العلم بها ، ويود الباحث التأكد من القيمة التي يفترض معرفتها ، باستخدام بيانات العينة.

يعرض هذا الباب الطرق الإحصائية المستخدمة في تقدير متوسط المجتمع Population Mean باستخدام متوسط العينة ، وتقدير النسبة في المجتمع باستخدام النسبة المحسوبة من العينة.

وفي الباب الثاني سنعرض أساليب اختبارات الفروض الخاصة بمتوسط مجتمع ، النسبة في المجتمع ، التباين في المجتمع . والباب الذي يليه يعرض طرق اختبارات الفروض للفرق بين متوسطي مجتمعين والنسبة في مجتمعين.

وطرق التقدير التي يتم عرضها في هذا الباب لها تطبيقات عملية عديدة فمثلا سوف نستطيع أن نقدر نسبة أعضاء النادي الذين يؤيدون ترشيح الفرد (أ) ليكون كرئيسا للنادي الرياضي ، وتعتمد هذه النسبة على سحب عينة من أعضاء النادي واستطلاع رأيهم في المرشحين لرئاسة النادي . وتعتمد الطرق التي سيتم عرضها في هذا الباب والتباين الملائم على نظرية النهاية المركزية . ويعرض هذا الباب أيضا كيف يمكن للباحث أن يحسب حجم العينة اللازم ليكون تقديره ذات خصائص معينة . ويعرض أيضا كيفية تقدير الفرق بين متوسطي مجتمعين والنسبة من مجتمعين .

وبدراسة موضوعات هذا الباب ، سوف يستطيع القارئ بعد دراسة هذا الباب أن يصل إلى حلول لبعض المشاكل في المجالات التجارية والإدارية .

١/١ مقدمة

أسلوب الحصر الشامل أسلوب مكلف ومجهد ، وأسلوب العينات أسلوب عملي ومنخفض التكلفة . وبدرجة معينة من الدقة ، يمكن إستخدام أساليب الإحصاء الإستنباطي الذي يعتمد على إحصاء العينة في تقدير معلمة المجتمع . سوف يهتم هذا الباب بتقدير .

١ - متوسط المجتمع (μ) باستخدام إحصاء العينة ($\bar{س}$) .

٢ - النسبة في المجتمع (ن) باستخدام النسبة في العينة (ل) .

٣ - الفرق بين متوسطي مجتمعين ($\mu_1 - \mu_2$) باستخدام الفرق بين متوسطي عينتين ($\bar{س}_1 - \bar{س}_2$) .

٤ - الفرق بين النسبة في مجتمعين مستقلين $(P_1 - P_2)$ باستخدام الفرق بين النسبة في عينتين $(p_1 - p_2)$.

ويبدأ موضوع التقدير بعرض التعريفات الهامة

١ - المقدر Estimator

هو إحصاءة أو معادلة تستخدم لتقدير المعلمة .

وأمثلة المقدرات ، الوسط الحسابي للعينة ويرمز له بالرمز $\bar{x}$ ، ومن الأمثلة أيضاً النسبة في العينة التي سوف يرمز لها بالرمز p .

٢ - التقدير Estimate

هو القيمة العددية للمقدر ، والوسط الحسابي المحسوب من عينة ما $\bar{x}$ هو أحد التقديرات الممكنة .

٣ - التقدير بنقطة Piont Estimate

هو قيمة تقدير واحدة ، مثلاً تقدير متوسط الدخل في المجتمع بـ ٤٠٠ جنيه شهرياً .

٤ - التقدير بمدى Interval Estimate

هو تقدير ذا مدى ، أي تقدير له حد أدنى وحد أعلى ، مثلاً تقدير متوسط دخل الأسرة في المجتمع ما بين ٣٥٠ ، ٩٠٠ جنيه شهرياً .

١/٢/١ معايير إختيار المقدر

حيث يمكن للمقدر جيد ؟ وما هي خصائصه ؟ يعتمد الاختيار على ثلاثة معايير أساسية هي : عدم التحيز ، الكفاءة النسبية ، الاتساق . وسنعرف فيما يلي المقصود بهذه المعايير

١ - عدم التحيز Unbiasedness

يكون المقدر [illegible] غير متحيز [illegible] معلمة المجتمع ، جميع [illegible] العينات الممكنة [illegible] القيمة [illegible] متوسط [illegible] المجتمع [illegible] معلمة المجتمع [illegible] توزيعات المعاينة) أي أن :

القيمة المتوقعة ($\bar{س}$) = μ أو ت ($\bar{س}$) = μ

وأيضاً متوسط توزيع معاينة النسبة يساوي النسبة في المجتمع أي أن

القيمة المتوقعة ($\hat{ل}$) = ل أو ت ($\hat{ل}$) = ل

أي أن الوسط الحسابي المحسوب من العينة ($\bar{س}$) والنسبة المحسوبة من العينة $\hat{ل}$ مقدرات غير متحيزة لمعلمتي المجتمع μ ، ل .

٢ - الكفاءة النسبية Relative Efficiency

المقدر غير المتحيز يكون ذا كفاءة نسبية حين يكون خطأه المعياري Standard Error أقل من أي مقدر اخر غير متحيز لمعلمة المجتمع ويبغي متخذوا القرارات التوصل إلي مقدر دقيق ، ينتج عنه فترة تقدير « ضيقة Narrow » .

وبفرض ثبات المتغيرات الأخرى التي سوف يتم عرضها ، كلما «صغر» الخطأ المعياري كلما «ضاقت» فترة الثقة للتقدير . فحين يكون المجتمع طبيعياً في توزيعه ، وجد أن الوسط الحسابي للعينة ، وكذلك الوسيط المحسوب من العينة مقدران غير متحيزان لمعلمة المجتمع μ .

ولكن وجد أن الخطأ المعياري للوسيط أكبر من الخطأ المعياري للوسط الحسابي .

لذا ، فعند مقارنة الوسط الحسابي بالوسيط ، نجد أن الوسط الحسابي يتمتع بالكفاءة النسبية ، حيث أن خطأه المعياري أقل . لذا فهو المقدر الجيد .

٣ - الإتساق Consistency

يكون المقدر مقدراً متسقاً إذا كانت قيمته مقاربة جداً لقيمة المعلمة المراد تقديرها . وبهذا يكون :

ح [قيمة المقدر = قيمة المعلمة] = قريبة جداً من الواحد ، أي أن :

$$\text{نهـا}_{ن \leftarrow \infty} \text{ ح [قيمة المقدر = قيمة المعلمة] = ١}$$

ووجد أن الخطأ المعياري للمقدر $\bar{س}$ "ع$_{\bar{س}}$" يتناقص مع زيادة حجم العينة حيث أن :

$$ع_{\bar{س}} = \frac{ع}{\sqrt{ن}}$$

وأيضاً الخطأ المعياري للنسبة $\hat{ل}$ يتناقص مع زيادة حجم العينة حيث أن :

$$ع_{\hat{ل}} = \sqrt{\frac{\hat{ل}\,(\,١ - \hat{ل}\,)}{ن}}$$

∴ المقدران : $\bar{س}$ ، $\hat{ل}$ مقداران غير متحيزان ومتسقان ، ويتمتعان بكفاءة نسبية عالية لمعلمتي المجتمع μ ، ل .

٢/١ التقدير لنقطة Point Estimate

يتم سحب عينة عشوائية ، ويتم حساب الوسط الحسابي البسيط للظاهرة محل الدراسة باستخدام بيانات العينة . أو يتم حساب النسبة للظاهرة محل الدراسة . ويكون هذا أو ذاك هو التقدير لنقطة .

ويستخدم الوسط الحسابي للعينة كتقدير "لنقطة" لمتوسط المجتمع حيث أنه مقدراً غير متحيزاً ، متسقاً ويتمتع بكفاءة نسبية .

١- التقدير لنقطة لمتوسط المجتمع = $\bar{س}$

ويعاب على هذا التقدير أن إحتمال أن يكون هذا التقدير $\bar{س}$ مساويا لمعلمة المجتمع صغير - وللتوضيح نعرض المثال التالى :

توزيع المعاينة التالى ناتج من مجتمع متوسطه = ٦ وقيم مفرداته (صفر ، ٢، ٤ ، ٦ ، ٨ ، ١٠ ، ١٢) - وحجم العينة = ٥ .

وينتج توزيع المعاينة من ٢١ عينة مختلفة ، حيث : ^{٧}ق$_{٥}$ = ٢١ عينة

والأوسطة الحسابية لكل عينة معطى فى جدول (٢-١) التالى :

ويوجد ٢١ عينة (ن = ٥) يمكن سحبها من المجتمع (ن = ٧) . وكل عينة من هذه العينات احتمال سحبها لأن تكون العينة المختارة = $\frac{١}{٢١}$.

ويعمل توزيع تكرارى للأوسطة الحسابية الناتجة نصل إلى جدول (١-١).

جدول (١-١) : الأوساط الحسابية لجميع العينات الممكنة

من مجتمع حجمه ن = ٧ وحجم العينة = ٥

نتائج العينة	المتوسط ($\bar{س}$)
٠ ، ٢ ، ٤ ، ٦ ، ٨	٤,٠
٠ ، ٢ ، ٤ ، ٦ ، ١٠	٤,٤
٠ ، ٢ ، ٤ ، ٦ ، ١٢	٤,٨
٠ ، ٢ ، ٤ ، ٨ ، ١٠	٤,٨
٠ ، ٢ ، ٤ ، ٨ ، ١٢	٥,٢
٠ ، ٢ ، ٦ ، ٨ ، ١٠	٥,٢
٠ ، ٢ ، ٤ ، ١٠ ، ١٢	٥,٦
٠ ، ٢ ، ٦ ، ٨ ، ١٢	٥,٦
٠ ، ٤ ، ٦ ، ٨ ، ١٠	٥,٦
٠ ، ٢ ، ٦ ، ١٠ ، ١٢	٦,٠
٠ ، ٤ ، ٦ ، ٨ ، ١٢	٦,٠
٢ ، ٤ ، ٦ ، ٨ ، ١٠	٦,٠
٠ ، ٢ ، ٨ ، ١٠ ، ١٢	٦,٤
٠ ، ٤ ، ٦ ، ١٠ ، ١٢	٦,٤
٢ ، ٤ ، ٦ ، ٨ ، ١٢	٦,٤
٠ ، ٤ ، ٨ ، ١٠ ، ١٢	٦,٨
٢ ، ٤ ، ٦ ، ١٠ ، ١٢	٦,٨
٠ ، ٦ ، ٨ ، ١٠ ، ١٢	٧,٢
٢ ، ٤ ، ٨ ، ١٠ ، ١٢	٧,٢
٢ ، ٦ ، ٨ ، ١٠ ، ١٢	٧,٦
٤ ، ٦ ، ٨ ، ١٠ ، ١٢	٨,٠
	١٢٦,٠

$$\mu_{\bar{س}} = \frac{١٢٦}{٢١} = ٦ = \mu$$

حيث $\mu_{\bar{س}}$ هو الوسط الحسابي للأوساط الحسابية .

جدول (١-٢) : التوزيع التكرارى للأوساط الحسابية في الجدول السابق

الوسط الحسابي	التكرار
٤,٠	١
٤,٤	١
٤,٨	٢
٥,٢	٢
٥,٦	٣
٦,٠	٣
٦,٤	٣
٦,٨	٢
٧,٢	٢
٧,٦	١
٨,٠	١
	٢١

ونلاحظ ان : إحتمال (الوسط الحسابي = ٦) = $\frac{٣}{٢١}$ ، أى أن :

∴ ح [$\bar{س}$ = ٦] = $\frac{٣}{٢١}$ = ٠,١٤٣

يعني أنه مع تكرار عمليـة المعاينة لاحجام عينـة حجمها (٥) مفردات من مجتمع حجمه ٧ مفردات ، يكون إحتمال إحتـواء التقدير على معلمة المجتمع الحقيقية مساويًا ٠,١٤٣ ، وتعتبر هذه النسبة نسبة بسيطة جداً .

٣/١ التقدير لفترة Interval Estimates

يمكـــن أن يكـــون إحتمال أن يحوي التقـــديـــر معلمة المجتمـــع الحقيقية كبيراً . فالتقدير لفترة يحتوي على «مدى» للتقدير . فهو يحتوي عـلى قيم صغيـرة (حد أدنى) وقيـم كبيرة (حـد أعلى) يحتمل أن يقـع بينهـما المعلـمة الحقيقية للمجتمع .

مـثلاً ، إذا تــم مراجعــــة بعض بوالص التأمين على الســيارات ، ووجد أن متوسـط تكلفة كـل شكوى ٢٧٣٨ جنيهاً ، وإذا أردت تقدير قيمة تكلفة كل شكوى بإضافة ٥٠ جنيهاً إلى ٢٧٣٨ وطــرح ٥٠ جنيهاً من ٢٧٣٨ جنيهاً .

(وسوف يتم عرض كيفية إختيار ٥٠ جنيهاً التي يتم إضافتها وطرحها وهو ما يسمى خطأ التقدير . وسوف نرمز له بالرمز خ) . إذن للتقدير لفترة تتبع الآتي :

> فترة تقدير لمتوسط المجتمع : س̄ ± خ
>
> حيث : خ = خطأ التقدير

ولشرح مبادئ للتقديـر لفترة ، تستخدم توزيع المعاينـــة المعطـــى فـــى جدول (١-١) . وبفرض أنك إخترت ٤٫٠ كمقدار لخطأ التقدير ، في هذه الحالـــة ، نحسب الفتـرة : (س̄ ± خ) لجميـع العينات كما هو معطى في جدول (١-٣) .

جدول (١-٣): توزيع المعاينة للأوساط الحسابية للعينة
من مجتمع متوسطه ٦ وفترة الثقة المحسوبة من $\bar{س}$ ± ٠,٤

الوسط الحسابي	ك	الفترة من $\bar{س}$ ± ٠,٤	الفترة التي تحتوي μ = ٦,٠
٤,٠	١	٤,٠ ± ٠,٤ = ٣,٦ إلى ٤,٤	لا
٤,٤	١	٤,٤ ± ٠,٤ = ٤,٠ إلى ٤,٨	لا
٤,٨	٢	٤,٨ ± ٠,٤ = ٤,٤ إلى ٥,٢	لا
٥,٢	٢	٥,٢ ± ٠,٤ = ٤,٨ إلى ٥,٦	لا
٥,٦	٣	٥,٦ ± ٠,٤ = ٥,٢ إلى ٦,٠	نعم
٦,٠	٣	٦,٠ ± ٠,٤ = ٥,٦ إلى ٦,٤	نعم
٦,٤	٣	٦,٤ ± ٠,٤ = ٦,٠ إلى ٦,٨	نعم
٦,٨	٢	٦,٨ ± ٠,٤ = ٦,٤ إلى ٧,٢	لا
٧,٢	٢	٧,٢ ± ٠,٤ = ٦,٨ إلى ٧,٦	لا
٧,٦	١	٧,٦ ± ٠,٤ = ٧,٢ إلى ٨,٠	لا
٨,٠	١	٨,٠ ± ٠,٤ = ٧,٦ إلى ٨,٤	لا

ونجد أن :

احتمال أن فترة الثقة تحتوى على متوسط المجتمع μ

$$= \frac{٣}{٢١} + \frac{٣}{٢١} + \frac{٣}{٢١} = \frac{٩}{٢١}$$

وهذا فيه تحسين عن التقدير لنقطة ، حيث كان هذا الإحتمال مساويا $\frac{٣}{٢١}$
ولتحسين الإحتمال ، أكثر ، إذا إخترنا خطأ التقدير = ٠,٨ ، تحسب الفترة
بحساب [سَ ± ٠,٨] والنتائج معطاه فى جدول (١-٤) .

جدول (١-٤) : توزيع المعاينة للأوساط الحسابية للعينة
من مجتمع متوسطه ٦ – وفترة الثقة المحسوبة سَ ± ٠,٨

متوسط العينة	ك	الفترة سَ ± ٠,٨	الفترة التي تحتوى μ = ٦ ؟
٤,٠	١	[٤,٨ – ٣,٢]	لا
٤,٤	١	[٥,٢ – ٣,٦]	لا
٤,٨	٢	[٥,٦ – ٤,٠]	لا
٥,٢	٢	[٦,٠ – ٤,٤]	نعم
٥,٦	٣	[٦,٤ – ٤,٨]	نعم
٦,٠	٣	[٦,٨ – ٥,٢]	نعم
٦,٤	٣	[٧,٢ – ٥,٦]	نعم
٦,٨	٣	[٧,٦ – ٦,٠]	نعم
٧,٢	٢	[٨,٠ – ٦,٤]	لا
٧,٦	١	[٨,٤ ٦,٨]	لا
٨,٠	١	[٨,٨ – ٧,٢]	لا

ونجد أن :

إحتمال أن تحتوى الفترة على معلمة المجتمع $\mu = ٦$

$$\frac{٢}{٣٦} + \frac{٣}{٣٦} + \frac{٣}{٣٦} + \frac{٣}{٣٦} + \frac{٢}{٣٦} = \frac{١٣}{٢١} = ٠,٦١٩ \quad ٦١,٩ \%$$

∴ يتميز التقدير لفترة عن التقدير لنقطة في أن إحتمال أن يحتوى الفترة على القيمة الحقيقية لمعلمة المجتمع يكون كبيراً ، ويعتمد هذا الإحتمال على المقادر خ

ونعرض التقدير لفترة ، للوسط الحسابي في الحالات الآتية :

١ - إذا كان الإنحراف المعيارى في المجتمع σ معلوماً .

٢ - إذا كان الإنحراف المعيارى في المجتمع σ غير معلوماً

ثم نعرض كيفية إنشاء فترة ثقة للنسبة في المجتمع .

١/٣/١ فترات الثقة للوسط الحسابي

(أ) الإنحراف المعيارى في المجتمع المعروف

يستخدم المصطلح «الثقة» ليصف التقدير لفترة . ففى المثال السابق ، حصلنا على ثقة ٦١,٩ ٪ بأن الفترة س̄ ± ٠,٨ ، والمحسوبة من عينة عشوائية واحدة حجمها ٥ مفردات ، تحتوى على معلمة المجتمع الحقيقية ($\mu = ٦$) . وتعبر النسبة ٦١,٩ ٪ عن مستوى الثقة في هذه الفترة المكونة . بمعنى آخر ٦١,٩ ٪ هى درجة «مصداقية» فترة الثقة المكونة التى تحتو على معلمة المجتمع μ .

فى الجزء الذى تم عرضه ، تم إختيار ٤ ، ٥ ، ٨ ، ٠ كأخطاء للتقدير فى هذا الجزء سوف نعرض إستخدام نظرية الحد المركزى لحساب أخطاء التقدير : خ ، وذلك بعد أن نحدد مستوى الثقة المطلوب .

∴ فترة الثقة عبارة عن «مدى» يحتمل أن يقع داخله المعلمة المراد تقديرها . وهذا الإحتمال يسمى «مستوى الثقة» ويرمز له بالرمز $(١ - \alpha)$ ٪ .

ولإنشاء فترة ثقة للوسط الحسابى فــى المجتمع ، تستخدم نظريــة الحد المركزى التى تستخدم لحساب إحتمال أن يقــع متوسط العينة على بعد معين من متوسط المجتمع .

فإذا كان الخــطأ المعيــارى للوسط الحسـابى $\sigma_{\bar{\text{س}}}$ ، يقرب تــوزيع المعاينـة بالتوزيع الطبيعى ، ونجد أن :

$$\text{ح}\left[-\text{ى} \le \frac{\bar{\text{س}} - \mu}{\sigma_{\bar{\text{س}}}} \le \text{ى}\right]$$

حيث : $\bar{\text{س}}$ = متوســط العينة وهذا الإحتمال موضح فى شكل (١-٦) .
ويمكن صياغة الإحتمال اعلاه كالآتى .

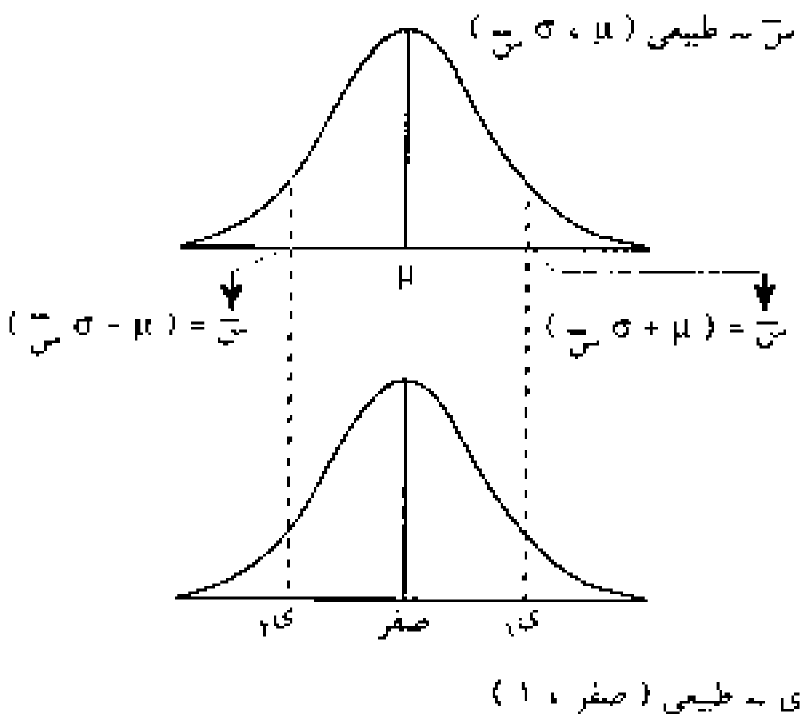
$$\text{ح}\left\{-\text{ى}\,\sigma_{\bar{\text{س}}} \le \bar{\text{س}} - \mu \le \text{ى}\,\sigma_{\bar{\text{س}}}\right\}$$

بطرح س من المتباينة :

$$ح\{\bar{س} + ى\,\sigma_{\bar{س}} \ge \mu - \ge (-\bar{س} - ى\,\sigma_{\bar{س}})\}$$

بالضرب × - ١ ، نجد أن :

$$\boxed{ح\{\bar{س} + ى\,\sigma_{\bar{س}} \ge \mu \ge \bar{س} - ى\,\sigma_{\bar{س}}\}}$$



لذا فعند أى قيمة للمتغير المعيارى ى ، يمكن أن نحدد إحتمال أن القيمة الحقيقية لمعلمة المجتمع تحتوى فى فترة الثقة (الوسط الحسابى ± ج) .

حيث ج = ى × الخطأ المعيارى للمقدر .

ومن دراستنا السالفة وجد أن :

المساحة بين (ى = − ١,٩٦ ، ى = + ١,٩٦) = ٩٥ ٪

أى أن ٩٥ ٪ من المشاهدات للمتغير العشوائى الذى يتبع التوزيع الطبيعى تقع بين القيمة المعيارية ى = ١,٩٦ ، القيمة المعيارية ى = − ١,٩٦

فإذا عوضنا عن قيمة ى = ١,٩٦ ، نجد أن :

ح [$\bar{س} + ١,٩٦ \, \sigma_{\bar{س}} \geq \mu \geq \bar{س} - ١,٩٦ \, \sigma_{\bar{س}}$] = ٠,٩٥ .

ويمكن القول إذن أن :

الفترة [$\bar{س} \pm ١,٩٦ \, \sigma_{\bar{س}}$] سوف تحوى معلمة المجتمع «μ» بإحتمال قدره ٩٥ ٪ .

وعمليًا فإنه يتم سحب عينة واحدة ، ويقدر منها فترة الثقة . ولكى نفرض أنك قمت بسحب عينات عديدة من المجتمع محل الدراسة . وقمت بتقدير فترات الثقة لكل عينة تم سحبها . إذا كررت هذه التجربة مرات عديدة، نجد أن ٩٥ ٪ تقريبًا من الفترات التى تم تكوينها تحتوى على معلمة المجتمع الحقيقية . وهذا هو المقصود بأننا على ثقة ٩٥ ٪ بأن هذه الفترة سوف تحوى متوسط المجتمع .

∴ مستوى الثقة Confidence Level هو احتمال أن طريقة التقدير سوف ينتج عنها فترات تشمل قيمة معلمة المجتمع إذا تم التقدير باستخدام عينات مختلفة من ذات المجتمع ولكن بنفس الحجم .

∴ إذا حددت مستوى الثقة ٩٠ ٪ ، معنى هذا أن فترات الثقة لـ ٩٠ ٪ من العينات الممكنة سوف تشمل قيمة μ . بمعنى آخر إذا كان حجم المجتمع = ٢٠ وحجم العينة = ٥ .

∴ عدد العينات الممكنة = $^{٢٠}ق_{٥}$ = $\frac{١٦ \times ١٧ \times ١٨ \times ١٩ \times ٢٠}{١ \times ٢ \times ٣ \times ٤ \times ٥}$ = ١٥٥٠٤ عينة

ومستوى ثقة ٩٠ ٪ ، يشير إلى أنه ، إذا تم حساب فترة الثقة لكل من عدد ١٥٥٠٤ عينة ، سوف نجد أن معلمة المجتمع الحقيقية محتوية في ١٥٥٠٤ × $\frac{٩٠}{١٠٠}$ = ١٣٩٥٤ عينة تقريباً .

ومستوى ثقة ٩٥ ٪ ، يشير إلى أنه ، إذا تم حساب فترة الثقة لكل من ١٥٥٠٤ عينة ، فسوف نجد أن معلمة المجتمع الحقيقية سوف تكون محتوية ١٥٥٠٤ × ٠,٩٥ = ١٤٧٢٩ عينة تقريباً .

ويجب أن متخذي القرارات يحددون مقدماً مقدار «عدم الثقة» في التقدير وعدم الثقة ينتج عنها إهمال التقديرات أقل من القيمة الحقيقية ، والتقديرات أكبر من القيمة الحقيقية ، ونجد أن :

١ – مستوى الثقة ، (١ – ∝) ٪ يمثل الجزء الأوسط في توزيع المعاينة .

٢ – المساحة في كل «ذيل» من توزيع المعاينة هي مساحة متماثلة في الذيلين ومساحة كل ذيل = ∝ / ٢ ٪ وتمثلا معاً نسبة عدم الثقة (∝) ٪ .

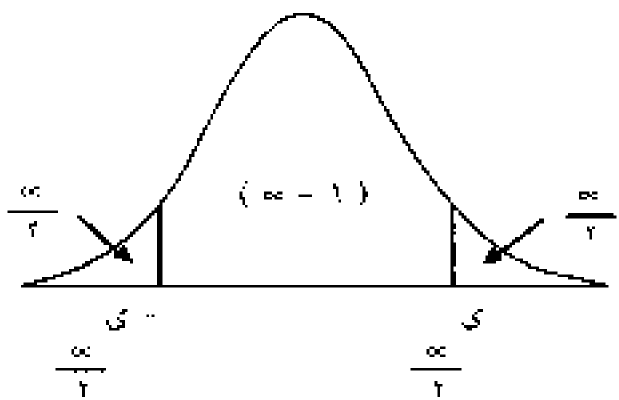
ومستويات الثقة الشائعة الإستخدام هى ٩٠ ٪ ، ٩٥ ٪ ، ٩٩ ٪ .

∴ (١ - ∝) ٪ = ح (فترة الثقة تشمل معلمة المجتمع) .

، ∝ ٪ = ح (فترة الثقة لا تشمل معلمة المجتمع)

∴ المساحة فى منتصف توزيع المعاينة = (١ - ∝)

مساحة كل ذيل = $\frac{\infty}{٢}$ ٪



والمعادلة الآتية تلخص كيفية تكوين فترة ثقة للوسط الحسابي في المجتمع في حالة ما إذا كان الإنحراف المعياري في المجتمع معلوماً ، وحجم العينة كبير لكي يسمح بتطبيق نظرية الحد المركزي .

فترة ثقة للوسط الحسابي في المجتمع في حالة σ معروفة :

$$\bar{س} \pm ى_{\alpha/2} \times \sigma_{\bar{س}}$$

حيث : (١ - α) : مستوى الثقة

$ى_{\alpha/2}$: متغير ى المعياري الذي يجعل مساحة الذيل الأيمن = $\alpha/2$

$\sigma_{\bar{س}}$ = الخطأ المعياري للوسط الحسابي $\dfrac{\sigma}{\sqrt{ن}}$

ونجد أن قيمة $ى_{\alpha/2}$ المصاحبة لفترة ثقة (١ - α) ٪ كالتالي :

مستوى الثقة (١ - α)	α	$ى_{\alpha/2}$
٠,٩٠	٠,١٠	١,٦٤
٠,٩٥	٠,٠٥	١,٩٦
٠,٩٩	٠,٠١	٢,٥٨

إنشاء فترة ثقة للوسط الحسابي في المجتمع

إذا كان الإنحراف المعيارى في المجتمع معروفاً ، وحجم العينة كبيراً ، تكون فترات الثقة كالتالي :

١ - إسحب عينة عشوائية بسيطة واحسب الوسط الحسابي .

٢ - حدد مستوى الثقة ، وحدد قيمة $ى_{\alpha/2}$ من جدول التوزيع الطبيعي المعيارى .

٣ - إحسب الخطأ المعيارى للوسط الحسابي ($\sigma_{\bar{س}}$) كالآتي .

(أ) إذا كان المجتمع غير محدود :

$$\sigma_{\bar{س}} = \frac{\sigma}{\sqrt{ن}}$$

(ب) إذا كان المجتمع محدود .

$$\sigma_{\bar{س}} = \frac{\sigma}{\sqrt{ن}} \sqrt{\frac{ن - ن}{ن - ١}}$$

٤ - إحسب فترة الثقة : $\bar{س} \pm ى_{\alpha/2} \quad \sigma_{\bar{س}}$

وفى بعض الأحيان يكون المطلوب هو تقدير فترة ثقة ، ليس للوسط الحسابي فى المجتمع ولكن «مجموع قيم المجتمع» .

مثلاً إذا أردت أن تقدر جملة المستحق على الحسابات بإنشاء فترة ثقة . في هذه الحالة نقوم بتقدير الوسط الحسابي للمستحق على الحسابات أولاً ، وذلك بإنشاء فترة ثقة عند مستوى الثقة المطلوب .

وتقوم بعد ذلك بضرب الحد الأعلى والحد الأدنى لفترة الثقة في حجم المجتمع .

∴ لتقدير فترة ثقة للمجموع : يكون حدى فترة الثقة :

الحد الأعلى = ن $\left[\bar{س} + ى_{\alpha/2} \quad \sigma_{\bar{س}} \right]$

الحد الأدنى = ن $\left[\bar{س} - ى_{\alpha/2} \quad \sigma_{\bar{س}} \right]$

تطبيق (١-١)

أراد صاحب شركة سياحة أن يقدر متوسط الوقت الذى يقضيه المندوب لعمل الترتيبات اللازمة لأحد الرحلات .

١ - طلبت من مدير الفرع أن يسحب عينة عشوائية من ٤٠ رحلة ثم إعدادها وملاحظة الوقت المستغرق لإنهاء الترتيبات .

وأوضح مدير الفرع أن المتوسط هو ٢٣,٤ دقيقة .

٢ - وغبت فى تقدير متوسط الزمن بإنشاء فترة ثقة ٩٥ ٪ . وأشارت الدراسات السابقة أن الزمن المستغرق يتبع التوزيع الطبيعى . وكان الاعتقاد أن حجم عينة مكونة من ٤٠ رحلة يعتبر حجماً كبيراً ويمكن من تطبيق نظرية الحد المركزى .

∵ مستوى الثقة ٩٥ ٪ ∴ $ى_{0.05/2}$ = ١,٩٦

٣ - أشارت الدراسات السابقة أن الإنحراف المعياري للزمن المستغرق هو ٩,٨ دقيقة .

$$\therefore \quad \sigma_{س} = \frac{\sigma}{\sqrt{ن}} = \frac{٩,٨}{\sqrt{٤٠}} = ١,٥$$

٤ - احسب فترة ثقة كالتالي : $س \pm ك_{\alpha/2} \times \sigma_{س}$

$= ٢٣,٤ \pm ١,٩٦ \times ١,٥ = [\ ٢٠,٥$ إلى $٢٦,٣\ ]$

∴ يقدر الزمن المستغرق ما بين ٢٠,٥ إلى ٢٦,٣ دقيقة . بدرجة ثقة ٩٥ ٪ . وهذا معناه أنك واثق ٩٥ ٪ من أن متوسط الزمن المستغرق لترتيب الرحلات في المجتمع سوف يتراوح ما بين ٢٠,٥ إلى ٢٦,٣ دقيقة

تطبيق(٦-٢)

أراد أحد مخططي إدارة المرور تقدير متوسط عدد السيارات التي تمر كل خمس دقائق والتي تعبر طريقاً معيناً فقام بالآتي :

١ - تم إختيار ٤٠ فترة زمنية ، طول كل منها خمس دقائق ووجد أن متوسط عدد السيارات التي تعبر طريق مصر / الاسكندرية الصحراوى هو ١٢٨,٣ سيارة ، في كل خمس دقائق .

٢ - أراد المخطط أن يكون على مستوى ثقة ٩٠ ٪ .

∴ $(١ - \alpha) = ٩٠$ ٪ ، $\alpha = ١٠$ ٪ ، $ك_{\alpha/2} = ١,٦٤$

٣ - دلت الأبحاث السابقة علي أن الإنحراف المعيارى لعدد السيارات العابرة كل خمس دقائق على طريق مصر/الاسكندرية الصحراوى هو ٤٣,٦ سيارة .

$$\therefore \quad \sigma_{\bar{س}} = \frac{\sigma}{\sqrt{ن}} = \frac{٣٠,٦}{\sqrt{١٠}} = ٩,٦ \text{ سيارة}$$

٤ - فترة الثقة ٩٠ ٪ = $\bar{س} \pm ى_{\alpha/٢} \; \sigma_{\bar{س}}$

$$= ١٢٨,٣ \pm ١,٦٤ \times ٦,٩$$

$$= ١٢٨,٣ \pm ١١,٣ = [\, ١١٧ \text{ إلى } ١٣٩,٦ \,]$$

∴ انــت علـى ثقة ٩٠ ٪ بأن متوســط عدد السيارات التي تعبر طريق مصر / الاسـكندرية الصحراوي كل خمـس دقائق ما بين ١١٧ إلى ١٤٠ سيارة .

١/٣/١ العوامل التي تؤثر على دقة فترة الثقة

كلمـا زاد المدى ، كلمـا زادت درجة الدقة . والمـدى هو الفرق بـين الحد الأعلى والحد الأدنى لفترة الثقة .

∴ المدى = الحد الأعلى لفترة الثقة – الحد الأدنى لفترة الثقة

$$\text{المدى} = (\bar{س} + ى_{\alpha/٢} \; \sigma_{\bar{س}} \;-\; \bar{س} \pm ى_{\alpha/٢} \; \sigma_{\bar{س}})$$

$$٢ \cdot ى_{\alpha/٢} \; \sigma_{\bar{س}}$$

ونجد أن المدى يتأثر بقيمة ى$_{\alpha/٢}$ ، وقيمة $\sigma_{\bar{س}}$.

العامل الذي يؤثر على $ي_{\alpha/2}$ هو مستوى الثقة - وكلما "إرتفع" مستوى الثقة ، كلما أصبحت قيمة $ي_{\alpha/2}$ "أكبر" [قارن الجدول]

∴ إذا رغبنا في فترة ثقة مداها صغير يجب ألا نختار مستوى ثقة مرتفع .

والعامل الذي يؤثر على المدى أيضاً قيمة $\sigma_{\bar{س}} = \frac{\sigma}{\sqrt{ن}}$ التي تتأثر بحجم حجم العينة (ن) . كلما كان حجم العينة كبيراً كلما أدى ذلك إلى "صغر" قيمة $\sigma_{\bar{س}}$ فمثلاً :

$$\sigma = ٥ \qquad ن = ١٠٠ \qquad \sigma_{\bar{س}} = \frac{٥}{\sqrt{١٠٠}} = ٠,٥$$

$$\sigma = ٥ \qquad ن = ٤٩ \qquad \sigma_{\bar{س}} = \frac{٥}{\sqrt{٤٩}} = ٠,٧١$$

$$\sigma = ٥ \qquad ن = ٢٥ \qquad \sigma_{\bar{س}} = \frac{٥}{\sqrt{٢٥}} = ١,٠$$

وبالتالي نجد أنه مع ثبات العوامل الأخرى [(١ - α) ، σ] فإن درجة ثقة التقدير يعتمد على حجم العينة .

وكلما صغر حجم العينة ، كلما "إنخفضت" درجة الدقة وكلما أصبحت فترة الثقة أكثر إتساعاً .

ولكن زيادة حجم العينة يؤدي إلى زيادة التكلفة (تكلفة جمع البيانات) فيجب المقايضة بين زيادة الدقة وإرتفاع التكلفة .

وأيضاً ، كلما صغر حجم الإنحراف المعياري (مع ثبات العوامل الأخرى (١ - α) ، ن) كلما أدى ذلك إلى صغر قيمة σ س وبالتالي صغر المدى .

∴ درجة الدقة تعتمد على (مع إفتراض ثبات العوامل الأخرى)

١ - حجم العينة .   ٢ - مستوى الثقة .

٣ - الإنحراف المعياري .

وعند إتخاذ القرارات يكون المطلوب فترة ثقة دقيقة ، ودرجة ثقة عالية ، وتكلفة معاينة صغيرة . في جزء قادم سوف نعرض كيفية تحديد الحجم الأمثل للعينة .

تماريـــن

[إذا لم يحدد حجم المجتمع في التمرينات التالية . إفترض أن حجم المجتمع غير محدود]

١ - أراد مدير شركة مبيعات بالتقسيط أن يقدر متوسط الرصيد المستحق على كل عميل .

قام بسحب عينة مكونة من ٤٥ حساباً . ووجد أن متوسط الرصيد المستحق = ٢٣١,٣٤ جنيه . وحيث أن متوسط الرصيد المستحق متغير من شهر إلى آخر فقد دلت الأبحاث أن الإنحراف المعياري للرصيد المستحق ثابت عند ٨٤,٣٧ جنيهاً . قدر :

١ - متوسط الرصيد المستحق في المجتمع بإنشاء فترة ثقة ٩٥ ٪ .

٢ - أوجد جملة الرصيد المستحق إذا كان عدد العملاء الذين يتعاملون بنظام التقسيط ٤٨٦٤ عميلاً .

٢ - أراد مدير إدارة المستخدمين تقدير متوسط الزمن الذي يصل فيه الموظفون متأخرين إلى عملهم .

قام بسحب ٣٥ كارت من كروت الحضور والانصراف ، ووجد أن متوسط زمن التأخير ١٢,٤ دقيقة . وحيث أن متوسط التأخير يختلف من يوم إلى آخر ، وجد أن الانحراف المعياري للتأخير ثابت عند ٤,٦ دقيقة

قدّر زمن التأخير الحقيقي بمستوى ثقة ٩٠ ٪

٣ - أراد قسم التسويق بإحدى شركات العطور معرفة متوسط دخول المشترين لأحد أنواع العطور .

تم إختيار ٧٥ مشتريًا ووجد أن متوسط دخولهم ٢٣٤٨٣ جنيهًا سنويًا

فإذا عرف أن الانحراف المعياري للدخول ٥٣١٩ جنيهًا .

قدّر متوسط الدخول لهذه الفئة من المشترين ، بإنشاء فترة ثقة ٩٠ ٪ .

٤ - أراد مسئول في أحد إدارات شركة التأمين معرفة متوسط عدد الأيام المطلوبة للإنتهاء من مطالبات المشتركين .

تم تتبع ٤٠ شكوى ، إختيرت بطريقة عشوائية . ووجد أن متوسط عدد الأيام المطلوبة ٥,٣ يوم عمل .

(أ) إذا كان الإنحراف المعياري لزمن الإنتهاء من مطالبات شكاوى العملاء هو ٢,٢ يومًا . قدر متوسط عدد الأيام وذلك بإنشاء فترة ثقة ٩٥ ٪

(ب) ما هي الإفتراضات اللازمة حتى تكون فترة الثقة أعلاه صحيحة .

٥ - أراد أحد المديرين في محلات عمر أفندي تقدير قيمة خسائر المسروقات اليومية

سحب عينة من سجلات المسروقات ووجد أن متوسط قيمة المسروقات ١٤٣,٤٩ جنيه . إذا عرف أن الإنحراف المعياري لهذه الخسائر ٢٣,٨ جنيه .

قدر المتوسط الحقيقي للمصروفات اليومية بإنشاء فترة ثقة ٩٠ ٪ ، إذا كانت أحجام العينات : (أ) ٤٠ (ب) ٥٠ (جـ) ٦٠ .

٦ - قارن بين النتائج التي حصلت عليها في (أ) ، (ب) ، (جـ) في السؤال (٥) أعلاه . ما هو أثر زيادة حجم العينة ؟

٧ - قام مدير تسويق في أحد شركات إنتاج الدفايات بتقدير متوسط عمر الدفايات من طراز 300 A .

سحب عينة عشوائية من ٣٥ دفاية ، واتضح أن متوسط عدد ساعات تشغيل الدفايات ١٦٣٤ ساعة . إذا علم أن الإنحراف المعياري لساعات تشغيل الدفايات ٢٨ ساعة .

قدر المتوسط الحقيقي في المجتمع بإنشاء فترة ثقة

(أ) ٩٠ ٪ (ب) ٩٥ ٪ (جـ) ٩٩ ٪

وقارن الإجابات التي حصلت عليها في (أ) ، (ب) ، (جـ)

٨ - أرادت شركة كروت إئتمان أن تقدر متوسط عدد مرات إستخدام كروت الإئتمان . سحبت عينة عشوائية من ٥٠ حسابا لحاملي الكروت ، ووجد أن متوسط عدد مرات الإستخدام ٣٫٥ مرة في الشهر

إذا علم أن الإنحراف المعياري لعدد مرات الإستخدام ٠٫٧ . إنشيء فترة ثقة ٩٥ ٪ لتقدير متوسط عدد مرات الإستخدام إذا كان عدد حاملي كروت الإئتمان : (أ) ١٨٣٩٧ . (ب) ١٨٤٠ . (جـ) ١٨٤ .

وقارن بين إجابتك لـ (أ) ، (ب) ، (جـ) أعلاه .

وماذا يحدث لفترة الثقة إذا كان حجم المجتمع غير محدود ؟

٩ - قدر جملة عدد مرات الإستخدام في السؤال (٨) أعلاه للحالات (أ) ، (ب) ، (جـ) . وذلك بإنشاء فترة ثقة ٩٥ ٪ .

٢/٣/١ فترة الثقة لمتوسط المجتمع

(ب) الإنحراف المعياري في المجتمع غير معروف

في العرض السابق تم تقدير متوسط المجتمع بإنشاء فترة ثقة إذا كان الإنحراف المعياري في المجتمع معروفاً . فى معظم الدراسات والتطبيقات ، لا يكون الإنحراف المعياري في المجتمع معروفاً . ويلزم تقديره لإنشاء فترة ثقة لمتوسط المجتمع ، والذى لا يكون أيضاً معروف .

في حالة عدم معرفة الإنحراف المعياري للمجتمع ، يستخدم الإنحراف المعياري المحسوب من بيانات العينة كتقدير للإنحراف المعياري في المجتمع . يقدر بعد ذلك الخطأ المعياري لتوزيع معاينة الوسط الحسابي $خ_{\bar{س}}$

وللتمييز بين الخطأ المعياري للوسط الحسابي المحسوب من العينة عن المحسوب من المجتمع ، نستخدم الرمز $خ_{\bar{س}}$ للدلالة على الخطأ المعياري المحسوب من العينة . والذى يحسب باستخدام احد المعادلتين الآتيتين :

(١) حالة المجتمع غير المحدود (ن > ٣٠) أو ن/ن < ٠.٠٥ .

$$خ_{\bar{س}} = \frac{ع}{\sqrt{ن}}$$

(٢) حالة المجتمع المحدود (ن ≤ ٣٠) أو ن/ن ≥ ٠.٠٥ .

$$خ_{\bar{س}} = \frac{ع}{\sqrt{ن}} \sqrt{\frac{ن - ن}{ن - ١}}$$

والخطأ المعيارى المحسوب من العينة هو المقدر الجيد للخطأ المعيارى فى المجتمع ، حيث أنه يتسم بعدم التحيز ، والاتساق ، والكفاءة النسبية العالية .

توزيع ت وفترة الثقة

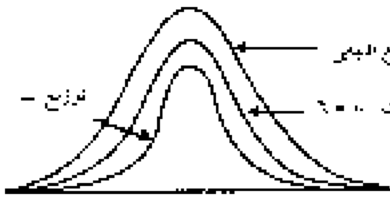
قدم W. S. Gosset (١٨٧٦ - ١٩٣٧) توزيع ت الاحتمالى فى عام ١٩٠٨ . وساعد هذا التوزيع فى حل مشكلة التوزيع الاحتمالى ، فى حالة عدم معرفة الانحراف المعيارى . فوجد Gosset أنه إذا قمت بسحب عينة عشوائية من مجتمع يتبع التوزيع الطبيعى فإن :

$$\frac{\bar{س} - \mu}{ع_{\bar{س}}}$$

تتبع توزيع ت الاحتمالى . ويشار إلى توزيع ت بتوزيع Student's t

وتوزيع ت يشبه إلى حد كبير ، التوزيع الطبيعى المعيارى ، ولكن فى توزيع ت ، يلعب حجم العينة دوراً فى تحديد القيم الاحتمالية . فتوزيع ت ، ليس توزيع معيارى محدد مثل التوزيع الطبيعى . فهو متماثل ، مثله مثل التوزيع الطبيعى ولكن درجة انتشاره أكبر على الإحداثى السينى . فالتوزيع الطبيعى المعيارى متوسطه صفراً ، وانحرافه المعيارى ١ ، وتوزيع ت متوسطه صفر أيضاً ، ولكن انحرافه المعيارى أكبر من ١ ، فهو توزيع منبسط Flatter عن التوزيع الطبيعى . ولكن إذا زاد حجم العينة عن ٣٠ يقترب توزيع ت مع التوزيع الطبيعى ، من حيث الشكل ودرجة الانبساط كما يظهر فى الشكل التالى .

جدول توزيع ت وحساب قيم ت



يتم الحصول على قيم ت من جدول توزيع ت (جدول ٢) وتمثل الأعمدة مساحة ذيل واحد من التوزيع .

ولإنشاء فترة ثقة ، تكون مساحة كل ذيل تحت توزيع المعاينة = ع/٢ ، مثلاً : عند إنشاء فترة ثقة ٩٥ ٪ ، تكون المساحة الوسطى في توزيع المعاينة = ٩٥ ٪ والمساحة المتبقية على طرفي توزيع المعاينة = (١ - ٠,٩٥) = ٠,٠٥ ، في $\frac{ع}{٢}$ = ٠,٠٥ / ٢ = ٠,٠٢٥ .

وتمثل الصفوف في الجدول ما يسمى بدرجات الحرية Degrees of Freedom . ودرجات الحرية هي عدد المشاهدات الحرة في العينة ، الممكن تحديد قيمها إذا علمنا بعض خصائص العينة : مثلاً ، الإنحراف المعياري في العينة ، يمكن حسابه إذا علمنا س̄ ، ويكون إذن عدد المشاهدات الحرة هي (ن-١) ، ويشار إلى درجات الحرية بالرمز "df" .

وعند تقدير متوسط المجتمع باستخدام توزيع ت تكون درجات الحرية = (ن-١) . وللحصول على قيمة ت التي تجعل الطرف الأيمن = ع/٢ ، نحدد درجات الحرية ثم نقرأ القيمة المعطاة عند تقاطع الصف (df) مع العمود (ع/٢) .

مثلاً : لدرجات حرية :

(ن − ١) = ١٩ ، $\frac{\alpha}{٢}$ = ٠,٠٢٥ ، ت١٩ ، ٠,٠٢٥ = ٢,٠٩٣

(ن − ١) = ٧ ، $\frac{\alpha}{٢}$ = ٠,٠٠٥ ، ت٧ ، ٠,٠٠٥ = ٣,٤٩٩

(ن − ١) = ٧ ، $\frac{\alpha}{٢}$ = ٠,٠٥ ، ت١٤ ، ٠,٠٥ = ١,٧٦١

ولكن ما هي قيمة ت عند درجات حرية ٤٣ ؟ في هذه الحالة ، إما حساب قيمة ت باستخدام أحد برامج الكمبيوتر ، أو إستخدام درجات حرية ٤٠ أو إستخدام قيم تقريبية باستخدام القيم من التوزيع الطبيعي المعياري . ويفضل إستخدام القيم التقريبية من جدول التوزيع الطبيعي المعياري في حالة درجات حرية ت غير موجودة في جدول ت ، ن ≥ ٣٠ حيث أن قيم ى المعيارية تعطي تقريباً جيداً لقيمة ت ، إذا كانت درجات الحرية ≥ ٣٠ .

فترة الثقة لمتوسط المجتمع باستخدام توزيع ت

تعود الأهمية العملية لتوصل Gossett لتوزيع ت ، في أنه إذا لم يكن الإنحراف المعياري في المجتمع معلوماً ، ومجتمع الظاهرة محل الدراسة يتبع التوزيع الطبيعي ، يمكن تقدير متوسط الظاهرة في المجتمع بإنشاء فترة ثقة وإستخدام المعادلة الآتية :

فترة ثقة لمتوسط المجتمع ، σ غير معروفة ، مجتمع الظاهرة يتبع التوزيع الطبيعي

$$\bar{س} \pm ت_{\frac{\alpha}{2}} \times ع_{\bar{س}}$$

بدرجات حرية = ن - ١

فإنشاء فترة الثقة باستخدام التوزيع الطبيعي المعياري أو توزيع ت يعتمد على :

١ - الإنحراف المعياري في المجتمع معلوم أم غير معلوم .

٢ - توزيع الظاهرة في المجتمع وعما إذا كان طبيعيًا أو يقترب من التوزيع الطبيعي .

٣ - حجم العينة

والإختيار بين ت ، ى معطى في جدول (١ - ٥) . وعمليًا يصعب معرفة الإنحراف المعياري في المجتمع ، حيث أنه يعتمد على متوسط المجتمع الذي لا نعرفه ونود أن نقدره . ولهذا فإنشاء فترات الثقة لمتوسط المجتمع يعتمد أساسًا على توزيع ت .

جدول (١ - ٥) : الإختيار بين ت ، ى في تقدير متوسط المجتمع الذي يتبع التوزيع الطبيعي

حجم العينة	σ معروفة	σ غير معروفة
كبير	ى	ى(*)
صغير	ى	ت
ى(*) يستخدم ى كتقريب لقيمة ت		

تطبيق (١-٣)

قامت شركة تصنيع مراتب سوست بسحب ٢٠ سوستة من سوست المراتب وقياس عدد الكيلوات اللازمة لضغط السوستة بوصة واحدة .

١ - أظهر فحص مراقبة الإنتاج لبيانات العينة أن متوسط الكيلوجرامات ٢٣,٨ والإنحراف المعيارى ٢,٤ كيلوجرام .

٢ - أردت أن تقدر متوسط الكيلوجرامات اللازمة لضغط السوستة بوصة واحدة في المجتمع بإنشاء فترة ثقة ٩٥ ٪ . ووجدت أن توزيع قوة ضغط السوستة فى المجتمع يتبع التوزيع الطبيعي ، وبدرجات حرية (٢٠-١) = ١٩ . وعند ح/٢ = ٠,٠٢٥ فإن قيمة ت٠,٠٢٥ = ٢,٠٩٣ .

٣ - يتم حساب الخطأ المعيارى للوسط الحسابي فى العينة كالآتى

$$ع_{\bar{س}} = \frac{ع}{\sqrt{ن}} = \frac{٢,٤}{\sqrt{٢٠}} = ٠,٥$$

٤ - بالتعويض بهذه القيم فى معادلة إنشاء فترة الثقة .

$$\bar{س} \pm ت_{\frac{ح}{٢}} \times ع_{\bar{س}}$$

$$= ٢٣,٨ \pm ٢,٠٩٣ \times ٠,٥$$

$$= ٢٣,٨ \pm ١,٠٥ \quad = \quad [\ ٢٢,٨ \text{ إلى } ٢٤,٨\ ]$$

∴ يقدر متوسط قوة الضغط بأنه يتراوح ما بين [٢٢,٨ إلى ٢٤,٨] كيلو جرام بدرجة ثقة ٩٥ ٪ . أى أنك تثق بنسبة ٩٥ ٪ بأن متوسط

قوة الضغط الذي ينتج عند سقوط السيارة بمقدار بوصة واحدة يتراوح بين [٢٢.٨ إلى ٢٤.٨] كيلو جرام

تطبيق (٤-٦)

أراد أحد المديرين لشركة تجارية معرفة متوسط المبيعات في الشركة، فقام بسحب عينة من ١٢ حساباً من حسابات العملاء، ووجد أن متوسط المبيعات الأسبوعية ٢١٣٨٤ جنيه بانحراف معياري ٣٥٧ جنيه، ودلت الخبرات السابقة أن المبيعات تتبع التوزيع الطبيعي، والمطلوب إيجاد فترة ثقة ٩٥٪ .

ن = ١٢ ، د.ح = ١١ ، ت (٠.٠٢٥ ، ١١)

ت جدولية = ٢.٢٠١

$$ع_{س} = \frac{ع}{\sqrt{ن}} = \frac{٣٥٧}{\sqrt{١٢}} = ١٠٣$$

فترة الثقة: $\bar{س} \pm ت \frac{ع}{\sqrt{ن}}$

٢١٣٨٤ ± ٢.٢٠١ × ١٠٣

[٢١١٥٧ إلى ٢١٦١١] جنيه

تماريـــــن

١٠- أوجد قيم ت .

(أ) ن = ٢٥	مستوى الثقة ٩٥ ٪
(ب) درجات الحرية = ٨	مستوى الثقة ٩٠ ٪
(جـ) ن = ٢٠	مستوى الثقة ٩٩ ٪
(د) درجات الحرية = ١٢	مستوى الثقة ٨٠ ٪

١١- أراد مدير أحد مصانع الإلكترونيات تقدير متوسط الزمن اللازم لتجميع أحد المنتجات . فقام بسحب عينة عشوائية وكان الزمن اللازم للتجميع بين مفردات هذه العينة هو ٢٫١ ، ٢٫٥ ، ٢٫٠ ، ٢٫٢ ، ٢٫١ ، ٢٫١ ، ٢٫٢ ، ٢٫٣ ، ١٫٩ ، ٢٫٠ ، ٢٫٤ . بفرض أن زمن التجميع يتبع التوزيع الطبيعي . قدر متوسط زمن التجميع بإنشاء فترة ثقة ٩٥ ٪

١٢- إذا أردت أن تقدر زمن الانتظار للدفع الذي يقضيه المشترى من أحد محلات السوبر ماركت حتى يقوم بالسداد بقيمة مشترواته . سحبت عينة عشوائية حجمها ٩ عملاء وكان زمن الانتظار للدفع هو ٨ ، ٩ ، ١٠ ، ٧ ، ٨ ، ٩ ، ٨ ، ١١ ، ١٠ . قدر متوسط زمن الانتظار للدفع بإنشاء فترة ثقة : ٩٥ ٪ ، ٩٠ ٪ .

١٣- إذا رغب تاجر عقارات تقدير متوسط أسعار شقق التمليك . فوجد أسعار عشرة شقق اختيرت بطريقة عشوائية كالآتي :

٨٣٫٤٠٠ ، ٨٦٫٨٠٠ ، ٨٢٫٩٠٠ ، ٨٥٫١٠٠ ، ٨٦٫٥٠٠

٨٨٫٣٠٠ ، ٨٤٫٥٠٠ ، ٨٩٫٦٠٠ ، ٨٢٫٠٠٠ ، ٨٦٫٥٠٠

إنشيء فترة ثقة ٩٥ ٪ لتقدير متوسط سعر شقق التمليك

١٤ - إذا وجد أن متوسط إنفاق ١٢ أسرة ، اختيرت بطريقة عشوائية ، على الوجبات السريعة ٣٠,١٨ جنيه أسبوعياً .

إذا كان الإنفاق يتبع التوزيع الطبيعي ، استخدم مستوى ثقة ٩٠ ٪ ، قدر المتوسط الحقيقي للإنفاق إذا كان :

(أ) ع = ٦,٣ جنيه ، (ب) ع = ٦,٣ جنيه

- قارن بين النتيجة في (أ) ، (ب)

٤/٨ فترة ثقة للنسبة في المجتمع

Confidence Interval for A Population Proportion

يستخدم مؤشر النسبة بصفة عامة في الإحصاءات المتعلقة بقياس نسب ظاهرة معينة ، منتشرة بين فئة معينة . فمثلاً ، لقياس نسبة المدخنين بين طلبة الجامعة مثلاً ، أو نسبة المشترين لسلعة معينة ، أو نسبة من يشاهدون برنامج تليفزيوني معين ، أو نسبة المشترين لصحيفة يوم الجمعة ، لصحيفة الأهرام بالنسبة لعدد مبيعات الصحف بصفة عامة .

وفي الأبحاث العلمية المتعلقة بالنسبة نحصل على نسبة الظاهرة في المجتمع في مدى معين . فمثلاً ، تتراوح نسبة قراء الأهرام عند الجامعة ما بين ٨٥ ٪ إلى ٩٠ ٪ من جملة القراء . هذا المدى ، يعطي أخيراً ، عن تطبيق لفترات الثقة للنسبة في المجتمع .

وتقدير النسبة في المجتمع ، مثله مثل تقدير الوسط الحسابي في المجتمع ، يعتمد على نظرية الحد المركزي . فتوزيع معاينة النسبة في العينة يمكن أن يقترب بالتوزيع الطبيعي إذا توافرت الشروط الآتية :

(أ) ن > ٣٠ (ب) ن $\hat{ل}$ ≥ ٥ (جـ) ن (١ - $\hat{ل}$) ≥ ٥

فإذا توافرت الثلاثة شروط أعلاه ، يمكن إستخدام الخطأ المعيارى للنسبة والتوزيع الطبيعى التقريبى لتوزيع معاينة النسبة حيث :

$$ح \left[ى = \frac{\hat{ل} - ل}{\sigma_{\hat{ل}}} \sim ى \right] \approx ط (٠ ، ١)$$

حيث : $\hat{ل}$ = النسبة المحسوبة من العينة = $\frac{م}{ن}$

م = عدد مفردات العينة الذى يتوفر فيه الظاهرة محل الدراسة ، عدد المدخنين مثلاً .

ن = حجم العينة .

ل = النسبة فى المجتمع . وهى نسبة غير معروفة ونريد تقديرها

وحيث أننا بصدد إنشاء فترة ثقة للنسبة فى المجتمع التى تقترب بالتوزيع الطبيعى ، فعند مستوى ثقة ٩٥ ٪ نجد أنه ، عند $\frac{\alpha}{٢}$ = ٠,٠٢٥ ، ى~٠,٠٢٥~ = ١,٩٦ .

ومن المعادلة أعلاه نجد أن :

$$ح \left[\hat{ل} - ١,٩٦ \; ع_{\hat{ل}} < ل < \hat{ل} + ١,٩٦ \; ع_{\hat{ل}} \right] = ٩٥ ٪$$

حيث : ع~$\hat{ل}$~ هو الخطأ المعيارى للنسبة ويحسب بأحد المعادلتين الآتيتين طبقاً لمعرفة حجم المجتمع من عدمه ، وطبقاً لنسبة حجم العينة (ن) إلى حجم المجتمع (ن) .

(١) المجتمع غير المحدود   أو ن/ن ≤ ٥ ٪

$$\hat{\text{ع}}_{\hat{\text{ل}}} = \sqrt{\frac{\hat{\text{ل}}\,(\,1 - \hat{\text{ل}}\,)}{\text{ن}}}$$

(٢) المجتمع المحدود   أو ن/ن > ٥ ٪

$$\hat{\text{ع}}_{\hat{\text{ل}}} = \sqrt{\frac{\hat{\text{ل}}\,(\,1 - \hat{\text{ل}}\,)}{\text{ن}}}\;\sqrt{\frac{\text{ن} - \text{ن}}{\text{ن} - 1}}$$

خطوات تكوين فترة ثقة للنسبة في المجتمع :

١ - سحب عينة عشوائية من المجتمع محل الدراسة ، إحسب النسبة في العينة

$$\hat{\text{ل}} = \frac{\text{ى}}{\text{ن}}$$

٢ - تأكد من الشروط الآتية : ن ≥ ٣٠ ، ن $\hat{\text{ل}}$ ≥ ٥ ، ن (١ - $\hat{\text{ل}}$) ≥ ٥

٣ - حدد مستوى الثقة وحدد قيمة $\text{ى}_{\frac{\alpha}{2}}$ من جدول التوزيع الطبيعي المعياري ، بمستوى ثقة (١ - α) ٪

٤ - إحسب الخطأ المعياري للنسبة من العينة ، طبقا لما هو موضح أعلاه

٥ - مقدار فترة الثقة (١ - α) ٪ كالآتي : ل^ ± ع $_{\alpha/2}$ ع$_{ل^}$

فترة ثقة (١ − α) ٪ لتقدير النسبة في المجتمع .
ل^ ± ع $_{\alpha/2}$ ع$_{ل^}$

تطبيق (١-٥)

إذا طلب منك كتابة تقرير عن مقدرة العملاء على شراء بعض السلع المعمرة ، وأردت أن تقدر نسبة العملاء الذين يمكنهم شراء سلع معمرة بالتقسيط قيمتها ٥٠٠ جنيه أو أقل .

١ - تم سحب عينة عشوائية من العملاء حجمها ١٠٠ عميل ، ووجدت أن نسبة من يستطيع تقسيط سلع قيمتها ٥٠٠ جنيه أو أقل ٢٢ ٪ .

٢ - وحيث أن :

(أ) ن ≥ ٣٠

(ب) ن ل^ = ١٠٠ × ٠,٢٢ = ٢٢ < ٥

(جـ) ن (١ - ل^) = ١٠٠ × ٠,٧٨ = ٧٨ < ٥

∴ يمكننا استخدام التوزيع الطبيعي كتقريب لتوزيع المعاينة للنسبة في العينة

٣ - بمستوى ثقة ٩٥ ٪ نجد أن : ع$_{٠,٠٢٥}$ = ١,٩٦

٤ - حيث أن مجتمع العملاء الذين يشترون السلع المعمرة ، مجتمع غير محدود ، وأن حجم العينة = ١٠٠ يمكن النظر إليه كأقل من ٥ ٪ من المجتمع ، ويمكننا إذن ، تجاهل معامل التصحيح $\sqrt{\dfrac{\text{ن} - \text{ن}}{\text{ن} - ١}}$ وحساب الخطأ المعياري للنسبة باستخدام المعادلة :

$$\text{ع}_{\text{ل}} = \sqrt{\frac{\hat{\text{ل}}\,(١ - \hat{\text{ل}})}{\text{ن}}} = \sqrt{\frac{٠,٢٢ \times ٠,٧٨}{١٠٠}} = ٠,٠٤١$$

٥ - فترة الثقة للنسبة : $\hat{\text{ل}} \pm ١,٩٦ \times \text{ع}_{\text{ل}}$

$$= ٠,٢٢ \pm ١,٩٦ \times ٠,٠٤١$$

$$٠,٢٢ \pm ٠,٠٨ = [\,٠,١٤ \text{ إلى } ٠,٣٠\,]$$

٦ - نسبة العملاء الذين يمكنهم تقسيط ٥٠٠ جنيه فو أقل تتراوح ما بين ١٤ ٪ إلى ٣٠ ٪ من العملاء

تطبيق (١-٦)

أرادت شركة معينة معرفة آراء العمال تجاه سياسة الحوافز المتبعة وسمحت ميزانية الأبحاث في الشركة بعمل مقابلة شخصية مع ٨٠ عاملا من جملة عدد العمال ، والبالغ عددهم ٣٨٤ عامل .

١ - وجدت أن ٦٧ عاملاً من الذين تمت مقابلتهم يشجعون سياسة الحوافز الجديدة .

∴ $\hat{\text{ل}} = \frac{٦٧}{٨٠} = ٠,٨٤$

٢ – $\text{ن} = ٨٠ > ٣٠$ ، $\text{ن}\,\hat{\text{ل}} = ٨٠ \times ٠,٨٤ = ٦٧ \geq ٥$ ،

$\text{ن}\,(١ - \hat{\text{ل}}) = ٨٠ \times ٠,١٦ = ١٢,٨ \geq ٥$

وحيث أن الثلاثة شروط أعلاه محققة ، نستطيع إستخدام التوزيع الطبيعي كتقريب لتوزيع معاينة النسبة في العينة .

٣ – عند مستوى ثقة ٩٩ ٪ ، $\alpha = ١ - ٠,٩٩ = ٠,٠١$ ، نجد أن : $\text{ذ}_{٠,٠٠٥} = ٢,٥٨$

٤ – نحسب الخطأ المعياري للنسبة في العينة : حيث أن المجتمع محدود ،

وأيضاً : $\frac{\text{ن}}{\text{N}} = \frac{٨٠}{٣٨٤} = ٢١$ ٪ تقريباً

∴ $\text{ع}_{\hat{\text{ل}}} = \sqrt{\frac{\hat{\text{ل}}\,(١ - \hat{\text{ل}})}{\text{ن}}}\ \sqrt{\frac{\text{N} - \text{ن}}{\text{N} - ١}}$

$= \sqrt{\frac{٠,٨٤ \times (١ - ٠,٨٤)}{٨٠}}\ \sqrt{\frac{٣٨٤ - ٨٠}{٣٨٤ - ١}} = ٠,٠٤$

٥ – فترة الثقة ٩٩ ٪ = $\hat{\text{ل}} \pm ٢,٥٨\ \text{ع}_{\hat{\text{ل}}}$

$= ٠,٨٤ \pm ٢,٥٨ \times ٠,٠٤ = [\,٠,٧٤ \text{ إلى } ٠,٩٤\,]$

∴ نسبة من يؤيدون سياسة الحوافز الجديدة تتراوح ما بين ٧٤ ٪ إلى ٩٤ ٪ .

٥/١ العوامل التي تؤثر علي دقة فترة الثقة

تقاس دقة فترات الثقة بإتساع المدى بين حدى فترة الثقة . والمدى هو الفرق بين الحدين الأعلي والأدنى لفترة الثقة . وكلما «صغر» المدى : كلما إزدادت دقة فترة الثقة .

$$\text{المدى} = (\hat{\text{ل}} + \text{ي}_{\alpha/٢}\,\hat{\text{ع}}_{\hat{\text{ل}}}) - (\hat{\text{ل}} - \text{ي}_{\alpha/٢}\,\hat{\text{ع}}_{\hat{\text{ل}}})$$

$$= ٢\,\text{ي}_{\alpha/٢}\,\hat{\text{ع}}_{\hat{\text{ل}}}$$

∴ «يضيق» المدى إذا صغرت قيم $\text{ي}_{\alpha/٢}$ أو $\hat{\text{ع}}_{\hat{\text{ل}}}$ أو كلاهما .

١ - العامل الذى يؤثر علي قيمة $\text{ي}_{\alpha/٢}$ هو مستوى الثقة . وكلما زاد مستوى الثقة ، كلما زادت قيمة $\text{ي}_{\alpha/٢}$ ، كما تم عرضه من قبل .

٢ - العامل الذى يؤثر علي قيمة $\hat{\text{ع}}_{\hat{\text{ل}}}$ هو حجم العينة ن حيث :

$$\hat{\text{ع}}_{\hat{\text{ل}}} = \sqrt{\frac{\hat{\text{ل}}\,(١ - \hat{\text{ل}})}{\text{ن}}}$$

وكلما زاد حجم العينة ، كلما انخفضت قيمة $\hat{\text{ع}}_{\hat{\text{ل}}}$ (مع ثبات العوامل الأخرى) .

فترة الثقة للنسبة ، حالة العينات الصغيرة

إذا لم تتحقق الشروط الثلاث المعطاة أعلاه وهي :

(١) $٣٠ \le \text{ن}$

(٢) $٥ < \hat{\text{ل}}\,\text{ن}$

(٣) $٥ \le (١ - \hat{\text{ل}})\,\text{ن}$

لا نستطيع إستخدام التوزيع الطبيعي كتقريب لتوزيع معاينة النسبة في العينة

إذا كان حجم العينة صغير ، نستخدم توزيع ذات الحدين Binomial distribution لإنشاء فترة ثقة للنسبة (ولكن يصعب شرح الطريقة هنا)

ولا يمكن إستخدام توزيع ت الذي يستخدم فقط لتقدير متوسط المجتمع وليس النسبة في المجتمع .

تماريـــن

١٥- تم سحب عينة عشوائية حجمها ١٠٠ عميل من مجموع العملاء ١٨٣٢٠ بغرض معرفة آرائهم نحو خدمة توصيل الطلبات للمنازل ووافق ٧٣ عميلاً منهم على هذه الخدمة .

(أ) إستخدم مستوى ثقة ٩٥ ٪ لتقدير نسبة من يؤيدون خدمة التوصيل للمنازل في المجتمع .

(ب) إستخدم مستوى ثقة ٩٥ ٪ لتقدير نسبة من لا يؤيدون خدمة توصيل الطلبات للمنازل .

(جـ) قارن بين حدي الثقة لما حصلت عليه في (أ) ، (ب) أعلاه .

١٦- في إدارة المراقبة على الإنتاج تم فحص ٢٠٠ قميص ووجد أن ٨ منهم معيبين . إستخدم فترة ثقة ٩٩ ٪ لتقدير نسبة المعيب في الإنتاج .

١٧- لاحظ بعض المتجين أن طلب عملائهم الدائمين لمنتجاتهم في إنخفاض شديد . تم فحص طلبات يوم معين وعددهم ٣٨٤ طلباً ووجد أن ٢٧٢ طلباً تم طلبها من العملاء الدائمين .

إذا اعتبرت أن الطلبات في هذا اليوم تمثل عينة عشوائية من جميع الطلبات قدر نسبة الطلبيات من العملاء الدائمين ، استخدم مستوى ثقة ٩٠ ٪ .

١٨- شعر عملاء أحد البنوك أن مصاريف إصدار كروت الائتمان ٥ جنيهاً سنوياً تعد مبلغاً كبيراً . سحبت عينة عشوائية حجمها ٤٠ عميلاً ووافق ١٨٣ منهم على أن مصاريف الإصدار مرتفعة .

إستخدم مستوى ثقة ٩ ٪ وقدر نسبة العملاء الذين يعتبرون مصاريف الإصدار مرتفعة .

١٩- عند مراجعة حسابات كروت إئتمان عينة من ١٠٠ عميل وجد أن ٨٣ منهم ملتزمون بدفع الحدود الدنيا المطلوبة وفي مواعيدها . إستخدم مستوى ثقة ٩٥ ٪ وقدر هذه النسبة في مجتمع حسابات كروت الائتمان وافرض أن

(أ) عدد حسابات كروت الإئتمان ٢٤٣ حساباً .

(ب) عدد الحسابات ٨،٢٤٣ حساباً .

(جـ) قارن بين فترتي الثقة في (أ) ، (ب) أعلاه .

٢٠- في إستفتاء ٢٠٠٠ طالب في أحد المدن ، لمعرفة آرائهم في الدروس الخصوصية ٤٣ ٪ منهم ، أبدوا أهمية الدروس الخصوصية للارتفاع بمستوى التعليم

إستخدم فترة ثقة ٩٩ ٪ لتقدير نسبة الطلبة الذين يؤيدون الدروس الخصوصية وفترة ثقة لتقدير نسبة الطلبة الذين لا يؤيدون الدروس الخصوصية .

٦/١ حجم العينة الأمثل Optimal Sample Size

تتحدد حجم العينة الذي يقوم الباحث بسحب بعدة عوامل أهمها التكلفة . فكل مفردة يتم سحبها يصاحبها تكلفة معينة ، يتحملها الباحث أو الجهة التي تقوم بالدراسة . والعينات الكبيرة ترفع من دقة التقدير ولكنها ترفع أيضاً من التكلفة . والعينات الصغيرة لا تصاحبها دقة عالية ولكنها لا تتكلف كثيراً . فيجب على الباحث الموازنة بين التكلفة والدقة . ويجب سحب الحجم الأمثل الذي يضمن مستوى معين من الدقة . والحجم الأمثل هو أصغر حجم عينة ينتج عنه فترة ثقة محدد لها مقدما حجم خطأ بمستوى ثقة معين .

ويجدر أن نتذكر أن خطأ التقدير هو هذا المقدار الذي يضاف مرة ويطرح مرة أخرى من المقدر ، وسبق وأن رمزنا له بالرمز خ .

فإذا كان الغرض هو تقدير الوسط الحسابي في المجتمع فإن :

خطأ التقدير = $ى_{\alpha/2} \times \frac{\sigma}{\sqrt{ن}}$	σ معروفة
$= ت_{\alpha/2} \times \frac{ع}{\sqrt{ن}}$	σ غير معروفة

وإذا كان الغرض هو تقدير النسبة في المجتمع فإن :

خطأ التقدير = $ى_{\alpha/2} \times \sqrt{\frac{\hat{ل}(١-\hat{ل})}{ن}}$

إن كان :

ن ≥ ٣٠ ، ن $\hat{ل}$ ≥ ٥ ، ن $(١-\hat{ل})$ ≥ ٥

وبصفة عامة فإن :

$$\text{خ} = \text{خطأ التقدير} = \frac{\text{المدى}}{٢}$$

$$= \frac{(\text{الحد الأعلى لفترة الثقة} - \text{الحد الأدنى لفترة الثقة})}{٢}$$

١/٦/١ حجم العينة الأمثل لتقدير متوسط المجتمع

يحدد بالمعادلة الآتية :

$$\text{ن}^{*} = \left[\frac{\text{ذ ص/٢} \times \text{ع}}{\text{خ}} \right]^{٢}$$

حيث : ن* : هو حجم العينة الأمثل

خ : خطأ التقدير ، ع : الإنحراف المعيارى للعينة

وهنا نجد تعارض فى المعادلة أعلاه . حيث أن حجم العينة الأمثل ن* يعتمد على ع : الإنحراف المعيارى فى العينة التى لم يتم تحديدها بعد . وأحد الحلول هو الإعتماد على الإنحراف المعيارى من عينة سابقة . وفى أبحاث التسويق ، يقدر الإنحراف المعيارى من نتائج أبحاث سابقة . وفى أحوال أخرى ، يقوم الباحث بعمل إختبار إستطلاعى Piolt سابقاً للبحث للتعرف على قيمة ع .

ويوجد بعض الأساليب لتقدير قيمة ع مثلاً ، إذا كان التوزيع معتدلاً أو شبه معتدل ، يكون مدى البيانات تقريباً من [μ - ٣ σ إلى μ + ٣ σ] ، أي أن الفرق بين القيمة العظمى للبيانات والقيمة الصغرى يساوي تقريباً ٦ σ .

$$\therefore \quad ع = \frac{\text{القيمة العظمى} - \text{القيمة الدنيا}}{٦} = \frac{\text{المدى}}{٦}$$

تطبيق (١-٧)

في تطبيق (١-٣) إذا أراد المسئول عن الأبحاث أن يقدر متوسط الكيلو جرامات اللازمة لضغط السوستة ١ بوصة بإنشاء فترة ثقة ٩٥ ٪ . وأراد أن يكون تقدير متوسط الكيلو جرامات بخطأ ٢/١ كجم أو أقل .

ما حجم العينة الأمثل ؟

توصلنا من قبل لفترة ثقة – باستخدام عينة من ٢٠ مرتبة وكانت فترة الثقة [٢٣,٨ ± ١,٠٥] أي من [٢٢,٨ كجم إلى ٢٤,٨ كجم] وكان خطأ التقدير ١,٠٥ كجم ولتخفيض خطأ التقدير إلى ٢/١ كجم وتحديد مستوى ثقة ٩٥ ٪ ، يجب أن نزيد حجم العينة .

وحجم العينة الأمثل ، لمستوى ثقة ٩٥ ٪ ، وخطأ = ٢/١ كجم ، أي أن :

خ = ٠,٥ ، ع = ٢,٤ :

$$ن^* = \left[\frac{١,٩٦ \times ٢,٤}{٠,٥}\right]^٢ = ٨٨,٥ \text{ مفردة} = ٨٩ \text{ مفردة تقريباً}$$

∴ يجب سحب عينة حجمها ٨٩ مفردة لكي نصل إلى درجة الدقة المطلوبة .

٣/٦/١ حجم العينة الأمثل لتقدير النسبة في المجتمع

يتبـع نفس منطق الوصول إلى الحجم الأمثل للعينة لتقدير النسـبة في المجتمع ، ويـجب تحديد حجم الخطأ الـذي يوده الباحث . وتـستخـدم المعادلة الآتية :

$$\text{ن}^{*} = \frac{\hat{\text{ل}}\,(1-\hat{\text{ل}})\,\text{ى}^{2}_{\alpha/2}}{\text{خ}^{2}}$$

$$= \hat{\text{ل}}\,(1-\hat{\text{ل}}) \left[\frac{\text{ى}_{\alpha/2}}{\text{خ}}\right]^{2}$$

في حالة :

$\text{ن} \geq 30$ ، $\text{ن}\,\hat{\text{ل}} \geq 5$ ، $\text{ن}\,(1-\hat{\text{ل}}) \geq 5$.

وأكبر حجم عينة يمكن سحبه يكون عند $\hat{\text{ل}} = 1/2$ ، وتكون المعادلة أعلاه كالتالي :

$$\text{ن}^{*} = \left[\frac{\text{ى}_{\alpha/2}}{2\,\text{خ}}\right]^{2}$$

والمعادلة الأخيرة تعطي أكبر حجم عينة يمكن سحبه . وتعاني معادلة الحجم الأمثل من مشكلة تقدير $\hat{\text{ل}}$ قبل سحب العينة .

تطبيق (١-٨)

أراد مدير إدارة أفراد أن يقيم آراء الموظفين عن سياسة الإعلان عن الوظائف الشاغرة في لوحة إعلانات موجودة بكافيتيريا المؤسسة . وأرادت الإدارة مستوى ثقة ٩٠ ٪ أن تجعل خطأ التقدير ٥ ٪ . ما حجم العينة اللازم سحبه لاستطلاع رأى الموظفين في السياسة الجديدة ؟ ، علماً بأنه ظهر في استطلاع سابق أن نسبة مؤيدى السياسة الجديدة ٨٤ ٪ .

مستوى ثقة = $(١ - \alpha) = ٩٠ \%$ $\therefore$ $\alpha = ١٠ \%$ $\text{ى}_{٠٫٠٥} = ١٫٦٥$

$\text{خ} = ٠٫٠٥$ ، $\hat{\text{ل}} = ٠٫٨٤$

$$\text{ن}^* = \hat{\text{ل}}\,(١ - \hat{\text{ل}}) \left[\frac{\text{ى}_{\alpha/٢}}{\text{خ}} \right]^٢$$

$$= ٠٫٨٤ \times ٠٫١٦ \left[\frac{١٫٦٥}{٠٫٠٥} \right]^٢$$

$= ٠٫٨٤ \times ٠٫١٦ \times ١٠٨٩ = ١٤٦$ مفردة تقريباً

أما إذا كانت النسبة $\hat{\text{ل}}$ غير معروفة من استطلاعات سابقة – فيمكن إستخدام معادلة أكبر حجم عينة – ويكون أكبر حجم عينة :

$$\text{ن}^* = \left[\frac{\text{ى}_{\alpha/٢}}{٢\,\text{خ}} \right]^٢ = \left[\frac{١٫٦٥}{٢ \times ٠٫٠٥} \right]^٢$$

$= (١٦٫٥)^٢ = ٢٧٢$ مفردة تقريباً

تمارين

٢١- أراد باحث تقدير متوسط الإنفاق السنوى لحاملى كروت الإئتمان . وذلك بإنشاء فترة ثقة ٩٥ ٪ وحدد الخطأ بـ ٢٠ جنيه أو أقل .

ما حجم العينة الواجب سحبها ، إذا أظهرت الدراسات السابقة أن الإنحراف المعيارى للإنفاق السنوى هو ٨٩ جنيه ؟

٢٢- فى تمرين (١) أعلاه ما حجم العينة الواجب سحبها إذا كان الخطأ ١٥ جنيه أو أقل ؟

٢٣- أراد باحث أن يقدر الطلبيات اليومية بإنشاء فترة ثقة ٩٥ ٪ . وأراد ألا يختلف التقدير عن القيمة الحقيقية أكثر من ١٥ طلبًا . وأشارت الدراسات إلى أن توزيع الطلبات اليومية يتبع التوزيع الطبيعى بإنحراف معيارى ٣٤,٢ طلبًا . ما حجم العينة الواجب سحبها ؟

٢٤- أراد باحث فى قسم مراقبة الجودة أن يقدر نسبة المعيب فى إنتاجه . بإنشاء فترة ثقة ٩٠ ٪ وإذا أراد ألا يختلف تقديره عن النسبة الحقيقية بأكثر من ٠,٠١٥ . ما حجم العينة الواجب سحبها ؟

(أ) بفرض أن نسبة المعيب ٤ ٪

(ب) إذا لم نعرف نسبة المعيب .

٢٥- أراد بنك أن يقدر نسبة عملاء التقسيط الذين يقومون بسداد الأقساط فى مواعيدها . وسمح بحد أقصى للخطأ ١,٨ ٪ واستخدم مستوى ثقة ٩٥ ٪ . وكانت هذه النسبة ٨٧,٥ ٪ فى دراسة سابقة - ما حجم العينة الواجب سحبها ؟

٧/١ فترات الثقة للفرق بين وسطي مجتمعين مستقلين

Confidence Interval for the Difference Between Two Population Means : Independent Samples

غالبًا ما يود متخذوا القرارات مقارنة مجتمعين . مثلاً ، ما الفرق بين متوسط الإنفاق للرجال والسيدات ؟ ما الفرق بين متوسط تكلفة بناء المتر المربع من الإسكان الفاخر في القاهرة والإسكندرية ؟ في الحالة الأولى ، إذا وجد فرق بين متوسط إنفاق الرجال ومتوسط إنفاق السيدات ، فإن هذا يعكس وجود علاقة بين الإنفاق والنوع . وفي الحالة الثانية ، إذا وجد فرق بين متوسط تكلفة البناء في القاهرة والإسكندرية ، فإن هذا يعكس وجود علاقة بين التكلفة ومحافظة البناء .

إن إنشاء فترات الثقة للفرق بين وسطين ، تعتبر احد الأساليب التي يمكن إستخدامها لإكتشاف وجود علاقة بين متغيرين . ويوجد أساليب عديدة لإختبار وجود إرتباط بين متغيرين ، وفترات الثقة هي أحد هذه الأساليب .

ولعند إنشاء فترة ثقة للفرق بين متوسطي مجتمعين مستقلين ، نفرض أن العينة الأولى ، حجمها $ن_1$ سحبت بطريقة عشوائية من المجتمع الأول ، ومتوسط المجتمع الأول μ_1 ، والعينة الثانية حجمها $ن_2$ سحبت بطريقة عشوائية من المجتمع الثاني ، ومتوسط المجتمع الثاني μ_2 .

١ - إحسب $\bar{س}_1$ ، $\bar{س}_2$ ، حيث $\bar{س}_1$ هو الوسط الحسابي البسيط من العينة الأولى ، $\bar{س}_2$ هو الوسط الحسابي البسيط من العينة الثاني ، ($\bar{س}_1 - \bar{س}_2$) هو التقدير لنقطة للفرق بين متوسط المجتمع الأول ومتوسط المجتمع الثاني . وهو تقدير غير متحيز وكفء ومتسق للفرق بين متوسطي المجتمعين ($\mu_1 - \mu_2$) .

٢- يحسب [illegible] حيث [illegible] هو الانحراف المعياري المحسوب من العينة الأولى [illegible] هو الانحراف المعياري المحسوب من العينة الثانية

٣- تحدد مستوى الثقة [illegible]

٤- يحسب الخطأ المعياري للفرق بين متوسطي العينتين [illegible]

٥- يحسب فترة الثقة [illegible] كالآتي

فترة الثقة [illegible]

[illegible]

في المعادلة (٣) أعلاه [illegible] يجب أن [illegible] ومحصور بين [illegible] بين (١ ، ٦) [illegible] يعني معظم [illegible] غير معروفتين لذا يستخدم العمود الأخير في جدول (١ ، ٦)

الافتراض الأساسي لصحة فترة الثقة بالطريقة أعلاه [illegible] أن المجتمع يتبع التوزيع الطبيعي. أما إذا كان المجتمع لا يتبع التوزيع الطبيعي [illegible] ولكن حجم العينتين كبيراً [illegible] نظرية الحد المركزي [illegible] يمكن أن يقرب توزيع معاينة الفرق بين متوسطي العينتين بالتوزيع الطبيعي. أما إذا كان حجم العينتين صغيراً فإنه يجب [illegible] الإحصاء [illegible] الجدول (١ ، ٦) يلخص ذلك.

جدول (١ - ٩) : الاختبارات المناسبة عند تقدير الفرق بين متوسطي مجتمعين مستقلين

المجتمع	حجم العينة	σ_1 ، σ_2 معلومتان	σ_1 ، σ_2 غير معلومتين
يتبع التوزيع الطبيعي	كبير صغير	Z Z	Z t
لا يتبع التوزيع الطبيعي	كبير صغير	Z ** لا معلمي	Z [illegible]

وفي الخطوة (٤) : الخطأ المعياري للفرق بين متوسطين مستقلين هو الانحراف المعياري لتوزيع معاينة الفرق بين متوسطي عينتين مستقلتين.

وعند استخدام اختبار ت نفترض أن التباين في المجتمعين متساوي (أي أن التباين في المجتمع الأول يساوي التباين في المجتمع الثاني). وتحت صحة هذا الافتراض يجب تقدير التباين لكل من المجتمعين. هذا التقدير عبارة عن متوسط حسابي مرجح لتباين العينتين، وأوزان الترجيح هي درجات الحرية، وينتج عن ذلك متوسطا يعكس تباين كل من المجتمعين، ويرمز لهذا التباين المجمع Pooled Variance بالرمز $ع^2_م$ حيث

$$ع^2_م = \frac{(ن_1 - ١)\, ع_1^2 + (ن_2 - ١)\, ع_2^2}{ن_1 + ن_2 - ٢}$$

ويستخدم هذا التباين المجمع لحساب تقدير للخطأ المعياري للفرق بين متوسطي العينتين :

$$ع_{\bar{س}_1 - \bar{س}_2} = \sqrt{\frac{ع_م^2}{ن_1} + \frac{ع_م^2}{ن_2}}$$

$$= \sqrt{ع_م^2 \left[\frac{1}{ن_1} + \frac{1}{ن_2}\right]} = ع_م \sqrt{\frac{1}{ن_1} + \frac{1}{ن_2}}$$

وإذا كانت العينتان كبيرتين ($ن_1 \geq 30$ ، $ن_2 \geq 30$) ، فإن توزيع ت يمكن أن يقرب بالتوزيع الطبيعي

تطبيق (١-٩)

إذا كان من المعروف أن متوسط إنفاق السيدات على الملابس أكبر من متوسط إنفاق الرجال فإذا رغبت تقدير هذا الفرق بإنشاء فترة ثقة .

١ – قمت بسحب عينة عشوائية حجمها ١٥ سيدة ، وعينة عشوائية حجمها ١٢ رجلاً . وسجلت الآتي :

السيدات	الرجـــــال
$\bar{س}_1 = ٢٧,٤٦$	$\bar{س}_2 = ٢٤,٢٤$
$ع_1 = ٢,١٠٠$	$ع_2 = ٣,٠١$
$ن_1 = ١٥$	$ن_2 = ١٢$

٢ – إذا اخترت مستوى ثقة ٩٠ ٪ . α = ١٠ ٪

. $ت_{٠,٩٥}$ بدرجات حرية ((١٥ + ١٢ – ٢) = ٢٥) = ١,٧٨

٣ – دلت الدراسات السابقة على أن توزيعًا الإنفاق للسيدات والرجال يقترب من التوزيع الطبيعي وأن التباين في المجتمعين متجانس .

تقدر الخطأ المعياري كالتالي :

$$ع_م^{٢} = \frac{(ن_١ - ١)\,ع_١^{٢} + (ن_٢ - ١)\,ع_٢^{٢}}{ن_١ + ن_٢ - ٢}$$

$$= \frac{(١٥ - ١) \times (٢,١٠)^{٢} + (١٢ - ١) \times (٣,١)^{٢}}{١٥ + ١٢ - ٢} = ٦,٩٨$$

$$\therefore \quad ع_{\bar{س}_١ - \bar{س}_٢} = \sqrt{٦,٩٨ \left[\frac{١}{١٥} + \frac{١}{١٢} \right]} = ١,٠٢$$

٤ – تقدر فترة الثقة كالتالي .

$$(\bar{س}_١ - \bar{س}_٢) \pm ت_{٠,٩٥} \times ع_{\bar{س}_١ - \bar{س}_٢}$$

$$= (٣٧,٤٦ - ٣٤,٣٤) \pm ١,٧٠٨ \times ١,٠٢$$

= ١,٤٨ إلى ٤,٩٦

وهذا معناه أن إنفاق الرجال أقل من إنفاق السيدات بمقدار ١,٤٨ إلى ٤,٩٦ جنيه .

٨/٦ فترات الثقة للفرق بين نسبتين في مجتمعين

Confidence Interval for the Difference of Two Proportions

أراد أحد خبراء التسويق أن يقدر الفرق بين نسبة إستهلاك السيدات الريفيات ونسبة إستهلاك السيدات الحضريات لأحد مساحيق التنظيف . وذلك قبل قيامه بحملته الإعلانية . وذلك بغرض معرفة هل يوجد علاقة بين إستهلاك هذا المسحوق وكون السيدة من الريف أو من الحضر .

يمكن تقدير الفرق بين نسبة إستهلاك الريفيات ونسبة إستهلاك الحضريات ومعرفة المدى الذي يتراوح داخله هذا الفرق بإنشاء فترة ثقة للفرق بين النسبتين . وتعتمد فترة الثقة للفرق بين نسبتين على توزيع المعاينة للفرق بين نسبتين محسوبتين من عينتين مستقلتين . ويتبع الخطوات التالية

١ - حساب الفرق بين نسبتين من عينتين مستقلتين هو ($\hat{\text{ل}}_{١} - \hat{\text{ل}}_{٢}$) حيث $\hat{\text{ل}}_{١}$ النسبة المحسوبة من العينة الأولى ، $\hat{\text{ل}}_{٢}$ هي النسبة المحسوبة من العينة الثانية :

$$\hat{\text{ل}}_{١} = \frac{\text{م}_{١}}{\text{ن}_{١}} \quad , \quad \hat{\text{ل}}_{٢} = \frac{\text{م}_{٢}}{\text{ن}_{٢}}$$

حيث م١ ، م٢ هما عدد المفردات التي ينطبق عليها الظاهرة محل الدراسة في العينة الأولى والعينة الثانية على التوالي .

والفرق بين النسبتين ($\hat{\text{ل}}_{١} - \hat{\text{ل}}_{٢}$) هو التقدير غير المتحيز ، والمتسق للفرق بين النسبتين في المجتمعين (ل١ - ل٢)

٢ - إذا كان :

$$\text{ن}_1 \ge 30 \quad ، \quad \text{ن}_2 \ge 30$$

$$\text{ن}_1 \hat{\text{ل}}_1 = \text{م}_1 \ge 5 \quad ، \quad \text{ن}_2 \hat{\text{ل}}_2 = \text{م}_2 \ge 5$$

$$\text{ن}_1 (1 - \hat{\text{ل}}_1) \ge 5 \quad ، \quad \text{ن}_2 (1 - \hat{\text{ل}}_2) \ge 5$$

يمكن تقدير فترة الثقة للفرق بين (ل١ - ل٢) بإتباع الآتي :

٣ - حدد مستوى الثقة (١ - ∝) ٪ ، وأوجد قيمة $\text{ى}_{\alpha/2}$.

٤ - يحسب الخطأ المعياري للفرق بين نسبتين من عينتين مستقلتين ، ويرمز له بالرمز $\text{ع}_{\hat{\text{ل}}_1 - \hat{\text{ل}}_2}$

حيث :

$$\text{ع}_{\hat{\text{ل}}_1 - \hat{\text{ل}}_2} = \sqrt{\frac{\hat{\text{ل}}_1 (1 - \hat{\text{ل}}_1)}{\text{ن}_1} + \frac{\hat{\text{ل}}_2 (1 - \hat{\text{ل}}_2)}{\text{ن}_2}}$$

وتكون فترة الثقة (١ - ∝) ٪ :

$$(\hat{\text{ل}}_1 - \hat{\text{ل}}_2) \pm \text{ى}_{\alpha/2} \times \text{ع}_{\hat{\text{ل}}_1 - \hat{\text{ل}}_2}$$

تطبيق (١-١٠)

إهتمت وزارة العمل منذ سنوات بمعرفة نسبة البطالة بين الذكور . ولكن في السنوات الأخيرة لوحظ إزدياد عمالة الإناث ، وظهور أيضاً البطالة بين الإناث . فاهتمت الوزارة بتقدير الفرق بين نسبة البطالة بين الذكور والإناث .

١ - ومن مجتمع قوة العمل ، اختير ٤٥٠ من الذكور ، ٤٥٠ من الإناث . ومجتمع قوة العمل يشمل : العاملون فعلاً ، المنتظرون لدورهم في التعيين ، والباحثين عن العمل ، والفئتين الأخيرتين يشكلا "البطالة" . وجد ٢١ من الذكور ، ٥٣ من الإناث يقعا في الفئتين الثانية أو الثالثة (بطالة) .

$$\therefore \quad \hat{ل}_١ = \frac{٢١}{٤٥٠} = ٠,٠٤٧ \qquad \hat{ل}_٢ = \frac{٥٣}{٤٥٠} = ٠,١١٨$$

٢ - للتأكد من أن توزيع المعاينة للفرق بين النسبتين في العينتين يمكن أن يقرب باستخدام التوزيع الطبيعي ، نجد أن :

(أ) $ن_١ = ٤٥٠ \geq ٣٠$ ، $ن_٢ = ٤٥٠ \geq ٣٠$

(ب) $م_١ = ٢١ \geq ٥$ ، $م_٢ = ٥٣ \geq ٥$

(جـ) $ن_١ (١ - \hat{ل}_١) = ٤٢٩ \geq ٥$ ، $ن_٢ (١ - \hat{ل}_٢) = ٣٩٧ \geq ٥$

∴ تتوافر الثلاثة شروط للعينتين ، ويمكن إستخدام التوزيع الطبيعي كتقريب لتوزيع معاينة الفرق بين النسبتين .

٣ - إذا اخترت فترة ثقة ٨٠ ٪ ، ∴ α = ٢٠ ٪ ، ى ٠,١٠ = ١,٢٨

٤ - تحسب الخطأ المعيارى للفرق بين النسبتين :

$$\text{ع}_{\hat{\text{ل}}_1 - \hat{\text{ل}}_2} = \sqrt{\frac{\hat{\text{ل}}_1(1-\hat{\text{ل}}_1)}{\text{ن}_1} + \frac{\hat{\text{ل}}_2(1-\hat{\text{ل}}_2)}{\text{ن}_2}}$$

$$= \sqrt{\frac{٠,٠٤٧(١-٠,٠٤٧)}{٤٥٠} + \frac{٠,١١٨(١-٠,١١٨)}{٤٥٠}} = ٠,٠١٨$$

∴ فترة الثقة للفرق بين (ل١ - ل٢) =

$$(\hat{\text{ل}}_1 - \hat{\text{ل}}_2) \pm ١,٢٨ \times \text{ع}_{\hat{\text{ل}}_1 - \hat{\text{ل}}_2}$$

$$= (٠,٠٤٧ - ٠,١١٨) \pm ١,٢٨ \times ٠,٠١٨$$

= [- ٠,٠٩٤ إلى - ٠,٠٤٨]

أى أننا نتوقع أن نسبة البطالة للإناث أقل من نسبة البطالة للذكور بنسبة تتراوح ما بين ٤,٨ ٪ إلى ٩,٤ ٪ . وذلك بمستوى ثقة ٨٠ ٪ . وهذا معناه وجود علاقة بين «النوع» و «البطالة» .

تماريـــــن

٢٦- أراد مدير مبيعات أن يقدر الفرق بين متوسط قيمة المبيعات اليومية في فرعين مختلفين . سحب المدير عينة من الفرع رقم (١) وتمثل هذه العينة مبيعات عشرين يومًا . ووجد أن متوسط المبيعات ٢٣٨٢ جنيهًا بإنحراف معيارى ٨٩٣ جنيه . وسحب عينة من الفرع رقم (٢) ، تمثل مبيعات خمسة عشر يومًا ووجد أن متوسط مبيعات هذا الفرع في ١٥ يومًا ٢٨٩٦ جنيهًا بإنحراف معيارى ٥٩١ جنيه .

بفرض أن شرط تجانس التباين محقق . إحسب فترة ثقة ٩٠ ٪ لتقدير الفرق بين متوسط مبيعات الفرع الأول والفرع الثانى .

٢٧- يوجد شركتان شحن متخصصتان في إنهاء إجراءات الشحن . وأراد أحد العملاء أن يختار أحدهما ، وهي الشركة التي تستغرق وقتًا أقصر لإنهاء الإجراءات . قام العميل بفحص عينة من سجلات الشركة الأولى ووجد أن عدد الأيام المستغرقة لإنهاء إجراءات شحن ٩ أوامر هى : { ٣ ، ٢ ، ٢ ، ٥ ، ٤ ، ٢ ، ٣ ، ٣ ، ٤ } يومًا . وسحبت عينة من سجلات الشركة الثانية ووجد أن عدد الأيام المستغرقة لإنهاء إجراءات ٩ أوامر هى : { ٢ ، ٢ ، ٦ ، ٤ ، ٣ ، ٥ ، ٥ ، ٤ ، ٢ } يومًا .

استخدم مستوى ثقة ٩٥ ٪ لتقدير الفرق بين متوسط عدد الأيام المستغرقة لإنهاء إجراءات الشحن في الشركتين .

بماذا تنصح العميل ؟ إختيار الشركة الأولى أم الثانية ؟

٢٨- تسلمت شركة عمالة ، عدد كبير من طلبات التشغيل بعد إعلانها في بعض الوظائف الخالية . سحبت الشركة عينة عشوائية من طلبات الذكور المتقدمين للوظائف ، حجمها ٤٠ طلبًا ووجد أن متوسط عدد سنوات التعليم في هذه العينة ١٤,٣ سنة بإنحراف معيارى ١,٤ سنة . وسحبت

عينة من طلبات الإناث المقدمات، حجمها ٣٥ طلبًا، ووجد أن متوسط عدد سنوات التعليم فى هذه العينة ١٦,٤ سنة بإنحراف معياري ٢,٣ سنة .
بفرض أن مجتمعى الذكور والإناث يتبعا التوزيع الطبيعى وان شرط تجانس التبايـن محقق . إحسب فترة ثقة ٩٠ ٪ لتقدير الفرق بين متـوسط عدد سنوات التعليم في المجتمعين .

٢٩ إذا أردت أن تقدر الفرق بين متوسط الزمن الـذى يستغرقه موردان للوفاء بطلبياتهم ، سحبت عينة عشوائية من عشر طلبيات للمورد الأول ، ووجد أن متوسط الزمـن الذي يستغرقه المورد الأول للوفاء بـطلبياته ١١,٣ يومًا بإنحراف معيارى ١,٢ يوم . وسحبت عينة عشوائية من ١٢ طلبية للمورد الثاني ، ووجد أن متوسط الزمن الذي يستغرقه المورد الثانى للوفاء بطلبياته ٩,٤ يومًا بإنحراف معيارى ٢,٤ يوم .
بإستخدام مستوى ثقة ٩٥ ٪ ، قدر الفرق بـين متوسط الزمن المستغرق فى الوفاء بالطلبيات . هل يدلك هـذا الفرق على إختيار أحد الموردين ؟ وأى مورد تختار ؟

٣٠- أراد مدير إحدى الـقنوات التليـفزيونية الإخبارية معرفة الفرق بـين نسبة مشاهـدى هذه القنـاة من الذكور ومـن الإناث . إختار عينة عـشوائية من الذكور حجمها ٢٠٠ مشاهد ، ووجد أن من بينهم ٤٣ مشاهد يتابع هذه القناة الإخبارية . وإختار عينة عشوائية من الإناث حجمها ٢٠٠ مشاهد ، ووجد أن من بينهم ٣٥ سـيدة تتابع هذه القناة الإخبارية . بمسـتوى ثقـة ٩٥ ٪ ، قدر الفرق بين النسـبتين .

الباب الثاني

إختبارات الفروض : مجتمع واحد

الباب الثاني

إختبارات الفروض : مجتمع واحد

يوجد نوعان من الإحصاء الإستنباطي : التقدير وإختبارات الفروض . وكما عرضنا في الباب الأول ، يستخدم أسلوب التقدير عند تقدير معلمة لايوجد علم بقيمتها، وتستخدم بيانات العينة لتقدير هذه القيمة . أما أسلوب إختبارات الفروض فيستخدم حين يكون عند الباحث «تخمين» أو «إحساس» بقيمة المعلمة ، ويريد الباحث إختبار هذه القيمة المفترضة . باستخدام بيانات العينة .

مثلا ، إذا قامت شركة بشراء تأمين صحي على مستخدميها ، وأرادت معرفة متوسط عدد الأيام التي يتم فيها «دخول» الموظف المستشفى للعلاج في هذه الحالة يستخدم أسلوب التقدير ، ويستخدم الاسلوب الذى تم عرضه في الباب السابق لمعرفة متوسط المدة داخل المستشفى .

أما إذا قامت الشركة بمراجعة السجلات ووجدت أن متوسط عدد الأيام داخل المستشفى للعلاج ثمانية أيام أو أكثر . هنا يجب إستخدام أسلوب إختبارات الفروض ، لإختبار هل متوسط الإقامة داخل المستشفى «حقًا» ثمانية أيام أم أكثر أم أقل .

ويعتمد أسلوبا التقدير واختبارات الفروض على نظرية الحد المركزى . في هذا الباب سوف يتم عرض إختبارات الفروض الخاصة بمتوسط مجتمع واحد : كيفية وضع الإختبار ، وكيف يتم الإختبار . وفي الباب التالي نعرض إختبارات الفروض الخاصة بتساوى نسبة أو نسبتين أو أكثر من عدة مجتمعات مستقلة .

١/٢ إختبارات الفروض

إذا أردت معرفة هل تحتوى أقراص الفيتامينات من نوع معين على ٥ ميللجرام من الزنك أم لا ؟ . حيث أنه لا يمكن تم إختبار جميع الأقراص المنتجة لمعرفة عما إذا كان كل قرص يحتوى على ٥ ميلليجرام زنك أم لا (إختبارا مدمرًا) ، هنا نجد أن أسلوب المعاينة هو الأسلوب العملى والفعال .

تسمح إختبارات الفروض بالوصول إلى نتائج عن المتوسط «الحقيقي» في المجتمع ، (ففي المثال أعلاه ، متوسط الزنك فى أقراص الفيتامينات) ، باستخدام نتائج عينه . ولكن الإعتماد على العينة ، يعرض الباحث إلى خطأ الوصول إلى نتائج غير سليمة عن متوسط المجتمع .

فى مثالنا ، بفرض أنك قمت بسحب عينة من ٣٥ قرصاً ، ووجدت أن متوسط الزنك ٣,٧ ميلليجرام . يوجد سببان يشرحا لماذا كان متوسط الزنك ٣,٧ ميلليجرام، في حين أن المفروض أن يكون ٥ ميلليجرام ؟

١- قد تكون هذه النتيجة لأن القيمة المفترضة (٥ ميلليجرام) إفتراض غير صحيح . وهنا نرفض أن متوسط الزنك ٥ ميلليجرام .

٢- التعليل الثاني هو أن الفرق بين المتوسط الذى حصلت عليه من العينة ، والمتوسط المفترض ، يرجع فقط إلى أخطاء المعاينة . بمعنى أنه يمكن أن يكون متوسط «هذه» العينة ٣,٧ ميلليجرام ، فى حين أن المتوسط في المجتمع فعلا ٥ ميلليجرام . إذا قبلنا هذا التعليل ، فإن هذا معناه أننا لن نرفض أن متوسط الزنك في هذه الفيتامينات ٥ ميلليجرام .

واختبارات الفروض تسمح للباحث أن يقرر عما إذا كان هذا الإختلاف بين قيمة المتوسط في العينة وقيمة المتوسط في المجتمع يرجع إلى أخطاء المعاينة ، أو أن هذا الإختلاف بسبب أن المتوسط المفترض «فعلاً» فرض خاطيء .

٢/٢ الهيكل الأساسي لاختبارات الفروض

سنعرض في هذا الجزء الخطوط العريضة لإختبارات الفروض . وفي الأجزاء التالية سيتم عرض كيفية إجراء إختبارات الفروض المتعلقة بالوسط الحسابي أو النسبة أو التباين في مجتمع واحد . يلخص منهج إختبارات الفروض بصفة عامة في الخطوات الآتية :

١- صياغة الفروض .

٢- الحصول على نتائج عينة .

٣- إختيار الإختبار الإحصائي المناسب ويسمى «إحصائية الإختبار» لاختبار الفرض في (١) أعلاه . وتحديد مستوى المعنوية والمنطقة الحرجة .

٤- حساب الخطأ المعياري .

٥- حساب قيمة إحصائية الإختبار .

٦- إتخاذ قرار قبول أو رفض الفرض العدمي .

١- الخطوة الأولى : صياغة الفروض

الفرض هو «إعتقاد» بالنسبة لمجتمع أو أكثر . في إختبارات الفروض ، عليك أن تحدد فرضين هما : الفرض العدمي Null Hypothesis والفرض البديل Alternative Hypothesis .

والفرض العدمي هو «الإعتقاد» بالنسبة للمجتمع . وهو الفرض الذي يرد الباحث أن يختبره ، ويعتقد أنه «إعتقاداً» صحيحاً إلى أن يثبت عكس ذلك . وهو الفرض «القائم» إلا إذا أثبتت نتائج العينة أنه خاطئ .

مثلاً ، الإعتقاد بأن متوسط الزنك في الفيتامينات هو ٥ ميلليجرام . وهذا الإعتقاد يكون صحيحاً إلى أن تثبت نتائج العينة إما تأكيد صحته أو خطأه .

والفرض البديل ، يحدد قيمة مختلفة لمتوسط المجتمع عن القيمة المعتقد أنها صحيحة (قيمة الفرض العدمي) .

مثلاً ، في مثال الزنك أعلاه ، الفرض البديل هو أن متوسط الزنك في المجتمع (مجتمع الفيتامينات المنتجة من هذا النوع) يختلف عن ٥ ميلليجرام .

وسوف نستخدم الرمز «ف٠» ليوضح فرض العدم ، «ف١» ليوضح الفرض البديل . ففي المثال السابق ، تصاغ الفروض كالآتي :

$ف_٠ : \mu = ٥$ ميلليجرام

$ف_١ : \mu \neq ٥$ ميلليجرام

ويجب أن تصاغ هذه الفروض في مرحلة سابقة لتحليل البيانات ، حتى لا تكون الفروض متحيزة .

ويمكن صياغة الفروض باستخدام المتباينات « ≤ » أو « ≥ » كالتالي .

$ف_٠ : \mu \leq ٥$ ميلليجرام

$ف_١ : \mu > ٥$ ميلليجرام

وطريقة التحليل ، تفترض أن الفرض العدمي فرضاً صحيحاً إلى أن يثبت خطأه . فسواء كان الفرض العدمي : $\mu = ٥$ أو $\mu \leq ٥$ أو $\mu \geq ٥$ فإن الإختبار يفترض أن $\mu = ٥$. والفرض الذي تختاره يعتمد على الإختبار والظروف المحيطة به .

الإختبارات أعلاه ، توضح خاصيتين للفروض العدمية والفروض البديلة :

١- الفرض العدمي يصاحبه علامه = دائما .

٢- الفرض العدمي والفرض البديل مكملان ، بمعنى أنه إذا توصلت إلى أن الفرض العدمي فرض خاطئ ، فإن هذا معناه أن الفرض البديل فرض صحيح . أي إذا رفضنا الفرض العدمي : $\mu \leq ٥$ فإن هذا معناه أن الفرض البديل : $\mu > ٥$ فرض صحيح .

كيف تصاغ الفروض ؟ يعتمد على الظروف المحيطة بموضوع البحث . ولكن إختيار الفروض يمكن أن تقسم إلى خطوتين هما :

١- حدد إذا كنت ستكتب الفرض البديل مصاحباً بعلامة "$\neq$" أم "$>$" أم "$<$" . والفرض البديل (في معظم الإختبارات) هو فرض الباحث وهو الفرض الذي يود الباحث أن يثبت صحته .

٢- إذا كان مناسب ، قرر عما إذا كان "$>$" أم "$<$" متباينة مناسبة للفرض البديل ، وهذا الإختبار يسمى إختبار في إتجاه واحد . إختر "$\neq$" للفرض البديل ، إذا كنت ستتخذ قرارا معينا إذا كان متوسط المجتمع يختلف عن القيمة المفترضة وعن الإعتقاد السائد بشأن متوسط المجتمع . والإختبار الذي يصاحب الفرض البديل فيه علامة "$\neq$" يسمى إختبار في إتجاهين

٢- الخطوة الثانية : الحصول على نتائج العينة

إسحب عينة عشوائية ، واحسب بعض إحصاءات العينة . قرر بعد ذلك عما إذا كنت مستمر في إختبارك أم أنك تؤيد القيمة تحت الفرض العدمي .

إذا أيدت بيانات العينة صحة الفرض العدمي أي أن قيمة إحصاء العينة تقع في إتجاه الفرض العدمي ، فإنه لا داعي للإستمرار في الإختبار . وتكون النتيجة هي أنك لا تستطيع أن ترفض القيمة تحت الفرض العدمي ،وأن الفرض العدمي يكون فرضاً صحيحاً . والقول بأنك «تقبل» الفرض العدمي يعني أنك لا تستطيع رفضه .

٣- الخطوة الثالثة : إختيار إحصائية الإختبار ومستوى المعنوية وتحديد المنطقة الحرجة

الإختبار الإحصائي أو إحصائية الإختبار هو مؤشر يُحسب من بيانات العينة. وإختيار الإختبار الإحصائي المناسب لإختبار متوسط المجتمع هو نفس إختبار ت أو إختبار ى المستخدمين في تقدير متوسط المجتمع بإنشاء فترات ثقة والذي تم عرضه في الباب الأول .

∴ إحصائية الإختبار لاختبار متوسط المجتمع هو :

$$ى = \frac{\bar{س} - \mu_.}{\sigma_{\bar{س}}}$$

$$أو \quad ت = \frac{\bar{س} - \mu_.}{ع_{\bar{س}}}$$

حيث : $\mu_.$: هو قيمة متوسط المجتمع تحت صحة الفرض العدمي .

$\sigma_{\bar{س}}$ = الخطأ المعياري لمتوسط العينة (σ معروفة)

$$= \frac{\sigma}{\sqrt{ن}}$$

$خ_{\bar{س}}$ = الخطأ المعياري لمتوسط العينة (σ غير معروفة)

$$= \frac{خ}{\sqrt{ن}}$$

وقبل أن نشرح المقصود بمستوى المعنوية Significance Level . نوضح بأن الاختبار الإحصائي ، بصفة عامة ، له أربعة نتائج مختلفة . نتيجتان قرارهما صائب . ونتيجتان قرارهما خطأ .

القراران الصحيحان ينتجا من :

١- قبول الفرض العدمي وهو فرض صحيح .

٢- رفض الفرض العدمي وهو فرض خاطئ .

والقراران الخطآن ينتجا من :

١- رفض الفرض العدمي وهو فرض صحيح .

٢- قبول الفرض العدمي وهو فرض خاطئ .

والخطأ الأول يسمى خطأ من النوع الأول Type I error ، والخطأ الثاني يسمى خطأ من النوع الثاني Type II error . وجدول (٢-١) يوضح نتائج الإختبارات الإحصائية .

واحتمال رفض الفرض العدمي وهو فرض صحيح يسمى مستوى المعنوية ،

وهو احتمال الخطأ من النوع الأول ، ويرمز له بالرمز α ٪ . أما إحتمال القرار الصحيح بقبول الفرض العدمي وهوصحيح فيساوى (١ - α) ٪ ويرمز للخطأ من النوع الثاني بالرمز β ٪ ، وهو إحتمال قبول الفرض العدمي وهو فرض خاطئ . أما احتمال القرار الصحيح برفض الفرض العدمي وهو فرض خاطئ فيساوي (١ - β) ٪ ، ويعبر عن قوة الإختبار Power of the test . واحتمالات نتائج الإختبارات الإحصائية معطاه في جدول (٢-٢) .

جدول (٢-١) : نتائج الإختبارات الإحصائية

القرار	طبيعة الفرض العدمي (ف.)	
	صحيح	خاطئ
قبـــول	قرار صحيح	خطأ من النوع الثاني
رفـــض	خطأ من النوع الأول	قرار صحيح

جدول (٢-٢) : نتائج الإختبارات الإحصائية وإحتمالاتها

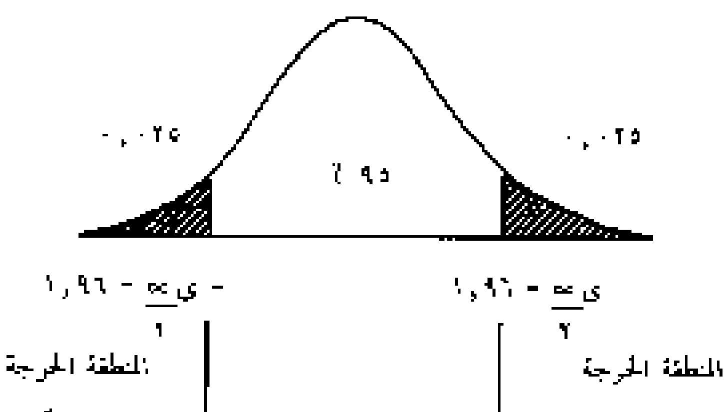
القرار	طبيعة الفرض العدمي (ف.)	
	صحيح	خاطئ
قبـــول	قرار صحيح (١ - α) ٪	خطأ من النوع الثاني β ٪
رفـــض	خطأ من النوع الأول α٪	قرار صحيح (١-β) ٪ = قوة الإختبار

ويهدف صانعوا القرار إلى تخفيض إحتمالات الخطأ . ولكن الخطأ من النوع الأول (α) يسير بطريقة عكسية مع الخطأ من النوع الثاني (β) . فكلما

زادت قيمة ∝ ، انخفضت قيمة β والعكس صحيح . ولكن اتفق على أن تأخذ ∝ القيم ١٪ أو ٥٪ أو ١٠٪ .

والقيمة الحرجة للإختبار الذى يتبع التوزيع الاحتمالى ى أو ت يتم الحصول عليها من جدول التوزيع الطبيعي المعيارى أو توزيع ت الإحتمالى .

وللحصول على القيمة الحرجة ، يجب تحديد ∝ إذا كان الإختبار الإحصائي في اتجاهين ، بمعنى أن الفرض البديل يصاحبه علامة "≠" . تحدد المنطقة الحرجة بقيمة ى أو ت التى تجعل طرفا المنحنى = ∝/٢ . فإذا كانت ∝ = ٥٪ ، ∝/٢ = ٠٫٠٢٥ ، المنطقة الحرجة تكون قيمة ى أو ت التي تجعل المساحة أو الإحتمال أكبر منها = ٠٫٠٢٥ ، والإحتمال أقل منها = ٠٫٠٢٥ ، فإذا كان الإختبار إختبار ى ، نحدد المنطقة الحرجة كما هو معطى فى شكل (٢-١) التالى :



شكل (٢-١) : المنطقة الحرجة لإختبار ى في إتجاهين ، ∝ = ٥٪

وبالتالي نجد أنه عند :

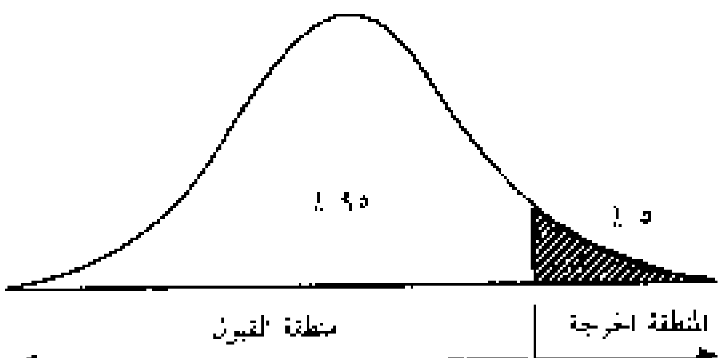
$$\alpha = ١\% \qquad (١ - \alpha)\% = ٩٩\% \qquad ى_{٠,٠٠٥} = ٢,٥٨ \qquad \pm ٢,٥٨$$

$$\alpha = ١٠\% \qquad (١ - \alpha)\% = ٩٠\% \qquad ى_{٠,٠٥} = ١,٦٥ \qquad \pm ١,٦٥$$

وفي حالة إختبار ت ، تتحدد المنطقة الحرجة بالقيمة $\pm$ ت$_{\alpha/٢}$ بدرجات الحرية المناسبة للإختبار .

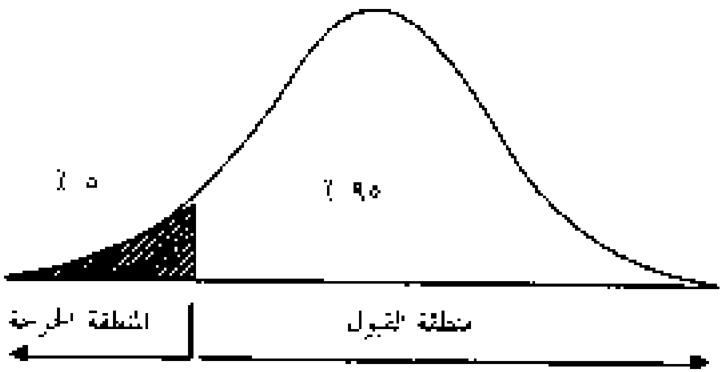
أما إذا كان الإختبار في إتجاه واحد ، بمعنى أن الفرض البديل يصاحبه علامه ">" أو "<" ، تحدد المنطقة الحرجة بقيمة ى أو ت التي تجعل الطرف الأيمن (في حالة >) أو الطرف الأيسر (في حالة <) مساوية α ٪ .

فإذا كانت α = ٥ ٪ والفرض البديل يصاحبه علامة ">" ، تحدد المنطقة الحرجة بقيمة ى أو ت التي تجعل الذيل الأيمن = ٥ ٪ كما هو موضح في شكل (٢-٢) :



شكل (٢-٢) : المنطقة الحرجة لإختبار ى ، إتجاه واحد ناحية اليمين ، α = ٥ ٪

وإذا كانت α = ٥ ٪ ، والفرض البديل يصاحبه علامة "<" ، تحدد المنطقة الحرجة بقيمة ى أو ت التي تجعل الذيل الأيسر = ٥ ٪ كما هو موضح في شكل (٢-٣) :



شكل (٢-٣) : المنطقة الحرجة لإختبار ى ، إتجاه واحد ناحية اليسار ، α = ٥ ٪

الخطوة الرابعة : حساب الخطأ المعياري

الخطأ المعياري في حالة σ معروفة :

$$\sigma_{\bar{س}} = \frac{\sigma}{\sqrt{ن}}$$

والخطأ المعيارى في حالة σ غير معروفة .

$$ع_{\bar{س}} = \frac{ع}{\sqrt{ن}}$$

الخطوة الخامسة : حساب إحصائية الإختبار Test Statistic

بصفة عامة إحصائية الإختبار يأخذ الشكل :

$$\text{إحصائية الإختبار} = \frac{\text{التقدير لنقطة} - \text{القيمة تحت الفرض العدمي}}{\text{الخطأ المعياري للتقدير لنقطة}}$$

ولإختبار قيمة مفترضة للوسط الحسابي في المجتمع μ. ، تحسب إحصائية الإختبار كالتالي :

$$ي = \frac{\bar{س} - \mu_{.}}{\sigma_{\bar{س}}} \quad \text{أو} \quad ت = \frac{\bar{س} - \mu_{.}}{خ_{\bar{س}}}$$

حيث μ. : الوسط الحسابي في المجتمع تحت صحة الفرض العدمي

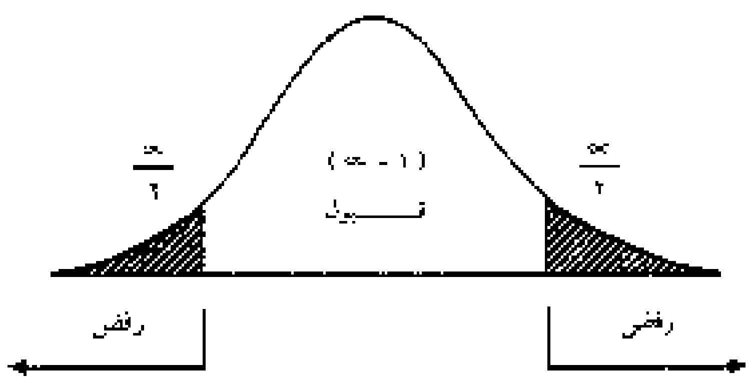
والبسط في إحصائية الإختبار اعلاه (ي أو ت) يوضح الفرق بين المتوسط الناتج من العينة والقيمة المفترضة للوسط الحسابي في المجتمع . وإحصائية الإختبار تعطي نسبة هذا الفرق إلى الخطأ المعياري للوسط الحسابي .

فإذا كان هذا الفرق = صفر ، فهذا معناه أن الوسط الحسابي في العينة يساوي القيمة المفترضة في المجتمع . ونتوقف عند هذه النقطة لأن بيانات العينة تؤيد الفرض العدمي . وكلما زاد هذا الفرق كلما اوضح ذلك بُعد متوسط بيانات العينة عن متوسط المجتمع . ويجب أن يقيم هذا الفرق كنسبة من الخطأ المعياري .

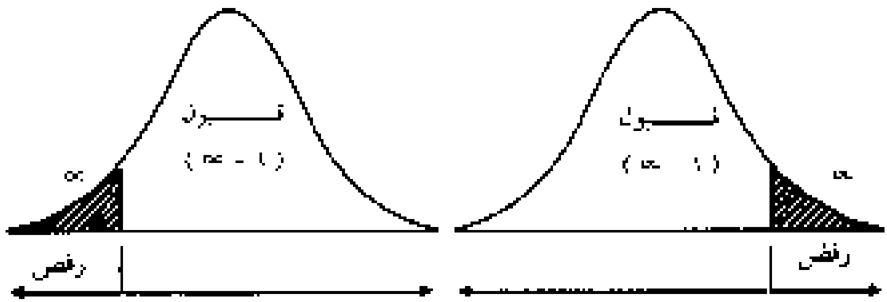
الخطوة السادسة : إتخاذ قرار بشأن الفرض العدمي

تقارن القيمة المحسوبة من الخطوة الخامسة (إحصائية الإختبار) بالمنطقة الحرجة (الخطوة الثالثة) : وترفض الفرض العدمي ،إذا وقعت قيمة إحصائية الإختبار داخل المنطقة الحرجة . وتقبل الفرض العدمي ، إذا كانت خارج المنطقة الحرجة .

ونلاحظ أن منطقة القبول مساحتها (١ - ∝) ٪ ، وهي مكملة لمنطقة الرفض التي مساحتها (∝) ، كما يتضح من شكل (٢-٤) ، (٢-٥) .



شكل (٢-٤) : مناطق القبول والرفض للاختبار في إتجاهين



شكل (٢-٥). مناطق القبول والرفض للاختبار في إتجاه واحد

٣/٢ اختبارات الوسط الحسابي

سوف نقوم في هذا الجزء بتطبيق طرق الإختبار الأساسية التي تم عرضها في الجزء السابق .

وطريقة صياغة الفروض على درجة كبيرة من الأهمية ، حيث أن الفروض توضح ماذا نود أن نختبر . وفي التطبيقات التالية ،سوف نركز على صياغة الفروض ، وكيفية تحديد منطقة قبول الفرض العدمي ، ومنطقة أو مناطق الرفض . والخطوات التي سوف نقوم باتباعها هي :

١- صياغة الفرض العدمي والفرض البديل .

٢- سحب عينة عشوائية ، وحساب الوسط الحسابي البسيط والانحراف المعياري باستخدام بيانات العينة .

٣- إختيار إحصائية الإختبار ، وهو إما إختبار ز أو إختبار ت وتحديد مستوى المعنوية والمنطقة الحرجة .

٤- حساب الخطأ المعيارى للوسط الحسابى .

٥- حساب قيمة الإختبار الإحصائي المحدد في الخطوة (٣) أعلاه .

٦- الوصول إلى قرار بشأن متوسط المجتمع ، كالآتي:

إتجاه الفرض البديل	رفض الفرض العدمي إذا كانت :
$\neq$	| قيمة ت المحسوبة | > : $ت_{\alpha/2}$ |
>	قيمة ت المحسوبة < − $ت_{\alpha}$
<	قيمة ت المحسوبة > $ت_{\alpha}$

حيث $ت_{\alpha}$ هى القيمة الجدولية من توزيع ت الإحتمالي التى تجعل ذيل منحنى $ت_{\alpha}$ الأيمن مساوى α .

وتشمل الخطوة (٣) أعلاه ، إختبار الإختبار الإحصائي وتحديد عما إذاكان الإختبار في إتجاه واحد أو فى إتجاهين .

والإختبار الإحصائى لإختبار متوسط المجتمع إما إختبار ى أوإختبار ت .

حيث :

$$\text{ى} = \frac{\bar{\text{س}} - \mu_{.}}{\sigma_{\bar{\text{س}}}} \quad \text{أو} \quad \text{ت} = \frac{\bar{\text{س}} - \mu_{.}}{\text{خ}_{\bar{\text{س}}}}$$

وإختبار أي منهما تتوقف على نفس الإعتبارات التي تم عرضها في باب التقدير ونوجزها في الآتي :

١- إذا كانت البيانات في المجتمع تتبع التوزيع الطبيعي ، والانحراف المعياري في المجتمع (σ) معروفاً . فإن الإختبار الإحصائي هو إختبار ى سواء كان حجم العينة كبيرا أم صغيرا .

٢- إذا كانت البيانات في المجتمع تتبع التوزيع الطبيعي ، والإنحراف المعياري في المجتمع (σ) غير معروفًا ، وحجم العينة كبير (ن ≥ ٣٠) فإنه طبقا لنظرية الحد المركزي فإن $\bar{س}$ تتبع التوزيع الطبيعي بمتوسط وإنحراف معياري $\frac{ع}{\sqrt{ن}}$ ، ويكون الإختبار الإحصائي هو إختبار ى .

٣- إذا كانت البيانات في المجتمع تتبع التوزيع الطبيعي ، والإنحراف المعياري في المجتمع (σ) غير معروفًا وحجم العينة صغير (ن < ٣٠) فإن توزيع معاينة الوسط الحسابي للعينة يتبع توزيع ت الإحتمالي بدرجات حرية (ن - ١) . ويكون الإختبار الإحصائي هو إختبار ت .

٤- إذا كانت البيانات في المجتمع لا تتبع التوزيع الطبيعي ، و σ غير معروفة ، وحجم العينة كبير فإن إختبار ى يقرب التوزيع الإحتمالي للوسط الحسابي في العينة (نظرية الحد المركزي) . ويكون الإختبار الإحصائي هو إختبار ى .

٥- إذا كانت البيانات في المجتمع لا تتبع التوزيع الطبيعي ، و σ غير معروفة وحجم العينة صغير فإنه يجب إتباع بعض الإختبارات اللامعلمية ، مثل إختبار Wilcoxon Signed-ranks . (يرجع إلى كتب الاحصاء غير المعلمي) .

٦ - إذا كانت البيانات في المجتمع تعكس إلتواءاً ، يجب إستخدام الوسيط وليس الوسط الحسابي البسيط كمقياس للنزعة المركزية . ويجب إتباع الإختبارات اللامعلمية مثل إختبار الإشارات Sign test وهو أحد الإختبارات اللامعلمية (يرجع إلي كتب الإحصاء غير المعلمي) .

تطبيق (٢-١)

إهتمت إحدى شركات تصنيع المربات بوزن العبوات المنتجة . فكان إنتاجها لعبوات ٢٥٠ جم . وكان إحدى مهام مدير الإنتاج متابعة الإنتاج ، وإيقاف الإنتاج إذا كان وزن العبوات يختلف كثيراً عن ٢٥٠ جم . ويجب في هذه الحالة إتخاذ قرار معين بشأن الإنتاج .

١- في هذه الحالة ، نستخدم علامة ≠ في الفرض البديل وتكون صياغة الفروض كالآتي :

ف. : μ = ٢٥٠ ؛ ف١ : μ ≠ ٢٥٠

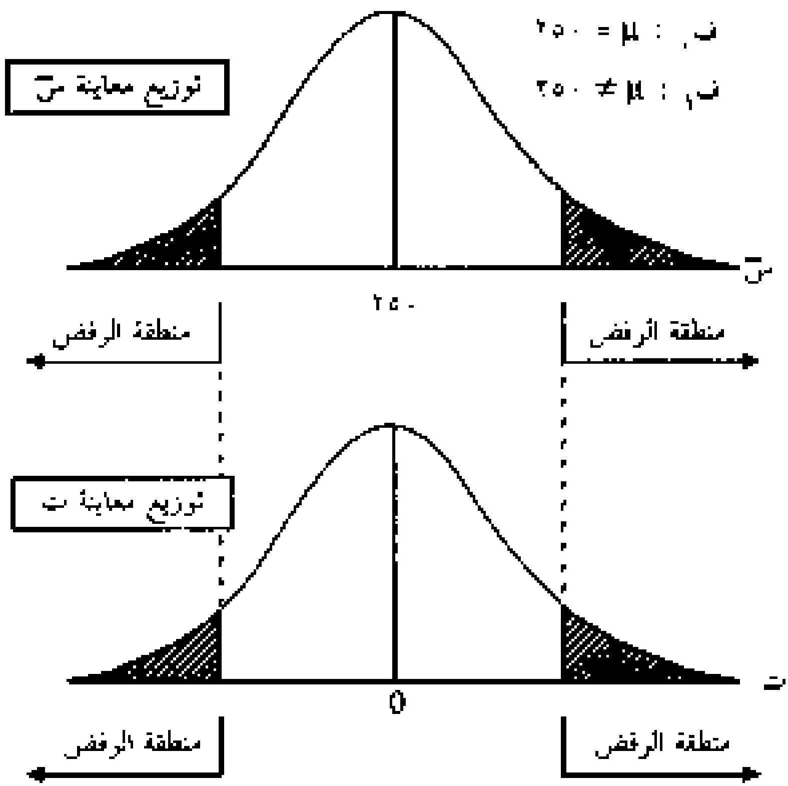
٢- تم سحب عينة من العبوات عددها ٢٠ عبوة ، ووجد أن متوسط وزن العبوة ٢٥٥,٩٧ جم بانحراف معياري ١٧,٩٧جم .

∴ متوسط العينة لا يؤيد الفرض العدمي ، حيث أن متوسط العينة يختلف عن ٢٥٠ جم . لذا يجب الإستمرار في خطوات إختبارات الفروض .

٣- دلت الأبحاث السابقة أن توزيع الأوزان في المجتمع يمكن أن يقرب بالتوزيع الطبيعي ، حيث أنه متماثل وذو منوال واحد .

حيث أن ن = ٢٠ حجم صغير ، ولا يمكن أن نطبق عليه نظرية الحد المركزي . إذن ، إختبار ت هو الإختبار المناسب .

إذا قررت أن مستوى المعنوية ∝ = ١٠ ٪ ، وحيث أن الفرض البديل يصاحبه علامة ≠ ويوجد إذن منطقتان رفض ، مساحة كل منها = ∝/ ٢ = ٥٪ كما هو موضح في شكل (٢-٦) :



شكل (٢-٦) مناطق القبول والرفض لتوزيع معاينة س̄ وتوزيع معاينة ت.

وتحدد المنطقة الحرجة بقيمة ت٠٥/ ٢ ، ت٩٥/ ٢ بدرجات حرية (ن-١) = (٢٠ - ١) = ١٩

∴ بدرجات حرية ١٩ نجد أن القيمة المطلقة $|ت_{٠,٠٥}|$ = ١,٧٢٩

(شكل (٣-٧)) .

٤- الخطأ المعياري لمتوسط العينة = $ع_{\bar{س}} = \frac{ع}{\sqrt{ن}} = \frac{١٧,٩٧}{\sqrt{٢٠}} = ٤,٠٢$

٥- الإختبار الإحصائي :

$$ت = \frac{\bar{س} - \mu}{ع_{\bar{س}}} = \frac{٢٥٥,٩٧ - ٢٥٠}{٤,٠٢} = ١,٤٩$$

٦- حيث أن قيمة ت المحسوبة أقل من قيمة $|ت_{\alpha/٢}|$ ، أي أن :

$(|١,٧٢٩| > ١,٤٩)$

∴ لا نرفض الفرض العدمي بأن متوسط وزن عبوات المربى = ٢٥٠ جم.

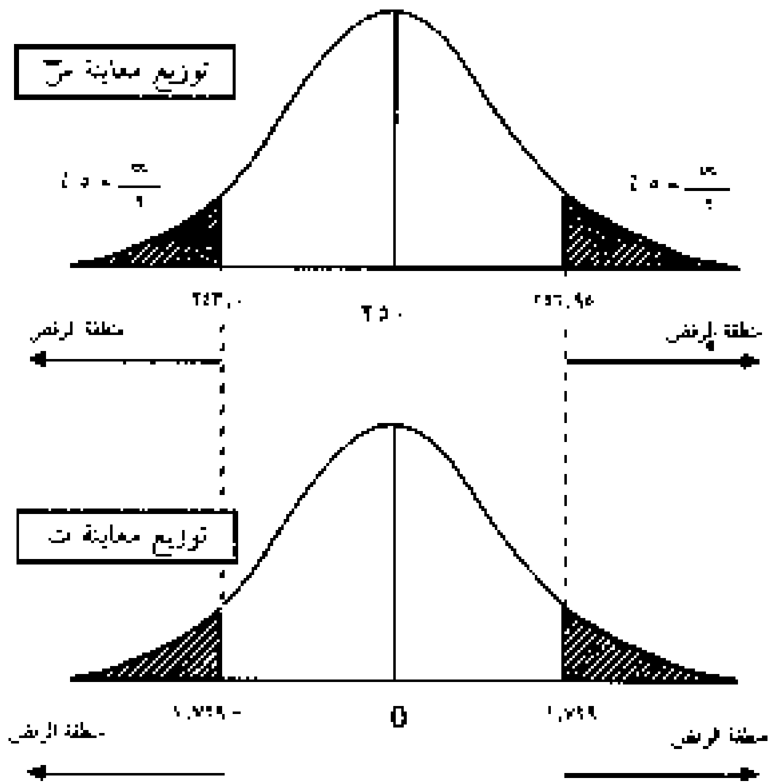
ويمكن التعبير عن مناطق الرفض في التطبيق أعلاه بدلالة $\bar{س}$ كالتالي :

منطقة الرفض هي المنطقة أكبر من μ + $ت_{\alpha/٢} \times ع_{\bar{س}}$

والمنطقة أقل من μ − $ت_{\alpha/٢} \times ع_{\bar{س}}$

∴ منطقة الرفض في حالة الإختبار في إتجاهين هي المنطقة أكبر من :

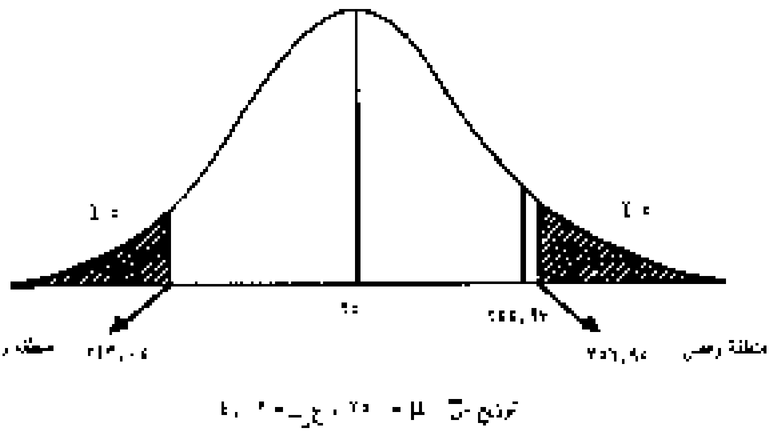
(٢٥٠ + (١,٧٢٩ × ٤,٠٢) = ٢٥٦,٩٥ جم)



شكل (٢-٧) : مناطق القبول والرفض للاختبار في تطبيق (٣-١)

والمنطقة أقل من :

(٢٥٠ - (١,٧٢٩ × ٤,٠٢) = ٢٤٣,٠٥ جم



شكل (٢-٨) : توزيع س̄ ، ومناطق القبول والرفض ، α = ٥ ٪

تطبيق (٢-٣)

إدعت إحدى المؤسسات التي تقدم دورات تدريبية في الحاسبات أن خريجها يحصلوا على أجر سنوي يزيد عن ٢٦ ألف جنيه سنوياً .

١- صياغة الفروض :

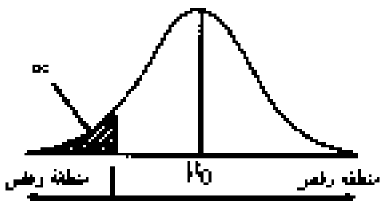
ف. : μ ≥ ٢٦٠٠٠ جنيه ، ف١ : μ < ٢٦٠٠٠ جنيه

٢- إختارت المؤسسة عينة حجمها ٤٥ خريجاً من جملة الخريجين البالغ عددهم ١٣٣ خريجاً . وأوضحت نتائج العينة أن متوسط الأجر السنوي ٢٥,٧٧٧ جنيه بانحراف معياري ١٠٩١ جنيه

وهنا لا نقبل الفرض العدمي حيث أن متوسط العينة يختلف عن المتوسط للمجتمع ، وقيمة س̄ تقع في نطاق الفرض البديل .

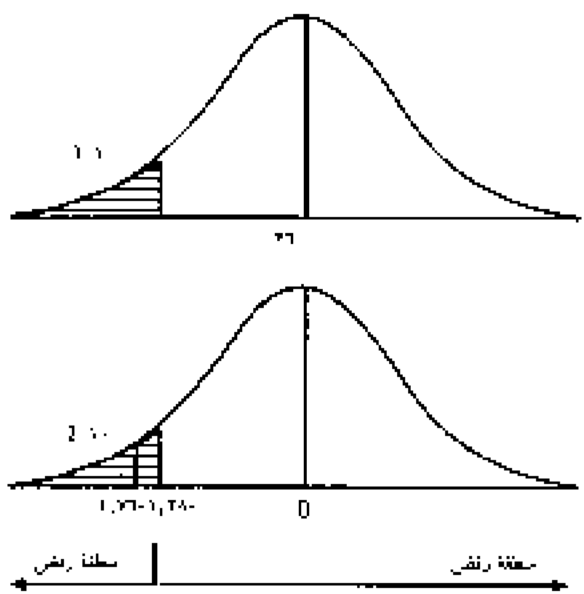
٣- حيث أن الفرض البديل يصاحبه علامه "<" ، إذن يوجد منطقة رفض واحدة ، وتقع على يسار التوزيع . والاختبار اختبار في اتجاه واحد ، كما هو موضح في الشكل (٢-٩) .

حيث أن حجم العينة كبير ،
وبفرض أن البيانات فى المجتمع تتبع التوزيع الطبيعى ، إذن الإختبار الإحصائي هو اختبار ى .



شكل (٢-٩) : مناطق القبول والرفض ، إختبار فى إتجاه واحد (ناحية اليسار)

وتحدد المنطقة الحرجة باستخدام قيمة ى التى تجعل الطرف الايسر مساوياً ∝ . وقيمة ى التى تحدد المنطقة الحرجة هى - ى٠،٠٥ = - ١،٢٨ . كما هو موضح فى شكل (٢-١٠) .



شكل (٢-١٠) : مناطق القبول والرفض للإختبار فى تطبيق (٢-٢)

٤- حيث ان نسبة العينة إلى المجتمع أكبر من ٥٪ (٤٥ ÷ ١١٣ = ٣٩,٨٪) ، إذن نحسب الخطأ المعياري للوسط الحسابي باستخدام المعادلة :

$$ع_{\bar{س}} = \frac{ع}{\sqrt{ن}} \sqrt{\frac{ن - ن}{ن - ١}} = \frac{١٠٩١}{\sqrt{٤٥}} \sqrt{\frac{١١٣ - ٤٥}{١١٣ - ١}} = ١٢٧$$

٥- الإختبار الإحصائي هو إختبار ى (حيث أن حجم العينة كبير) :

$$ى = \frac{\bar{س} - \mu}{\sigma_{\bar{س}}} = \frac{٢٥٧٧٧ - ٢٦٠٠٠}{١٢٧} = - ١,٧٦$$

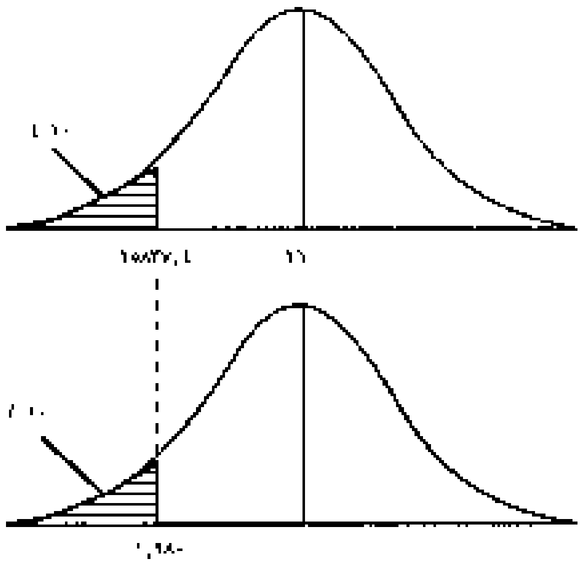
٦- حيث ان القيمة المحسوبة (- ١,٧٦) تقع داخل منطقة الرفض ، إذن ، نرفض الفرض العدمي ، ونصل إلى النتيجة بأن متوسط الأجر أقل من ٢٦٠٠٠جنيه .

ويمكن حساب المنطقة الحرجة بدلالة س كالآتى :

المنطقة الحرجة هي المنطقة أقل من أو يساوى (μ - ١,٢٨ × $ع_{\bar{س}}$)

اى $\bar{س} \leq$ (٢٦٠٠٠ - ١,٢٨ × ١٢٧) $\leq$ ٢٥٨٣٧,٤٤ جنيه .

وشكل (٢-١١) يوضح ذلك :



شكل (٢-١٦) : مناطق القبول والرفض لتوزيع س̄ ، توزيع ى ، ٩٥٪ ، ٥٪

تطبيق (٢-٣)

أرادت إحدى إدارات المرور أن تقرر عما إذا كان تقاطع شارعين معينين يحتاجا إلى إشارة مرور أم لا . وسوف يتم وضع إشارة مرور إذا كان ضغط المرور في هذا التقاطع ١٥٠ سيارة أو أكثر في الساعة .

١ - صياغة الفروض :

وحيث أن إضافة إشارة مرور مكلف ماديًا فإن الإدارة تود أن تثبت أن ضغط المرور لا يستلزم إضافة إشارة مرور أى أن حركة المرور أقل من ١٥٠ سيارة في الساعة . يكون الفرض العدمي والفرض البديل كالآتي :

ف. : $\mu \geq$ ١٥٠ (وهو ما ترغب الإدارة نفيه)

ف١ : $\mu <$ ١٥٠ (وهو ما ترغب الإدارة في تأكيده)

١٢ - تم متابعة حركة المرور في عشر ساعات ووجد أن عدد السيارات المارة في كل ساعة كالآتي :

١٤٠ ١٤٣ ١٦٢ ١٢٥ ٩٣ ١٦٠ ١٣٢ ١٠٨ ١٧٦ ١٠٩

$$\therefore \bar{\text{س}} = \frac{\text{مجـ س}}{\text{ن}} = \frac{\text{١٤٠ + ١٤٣ + ... + ١٠٩}}{\text{١٠}}$$

$$= \frac{\text{١٣٤٨}}{\text{١٠}} = \text{١٣٤,٨ سياره}$$

$$\therefore \text{ع}^2 = \frac{\text{مجـ س}^2 - \frac{(\text{مجـ س})^2}{\text{ن}}}{\text{ن} - \text{١}}$$

$$\text{مجـ س}^2 = \text{١٤٠}^2 + \text{١٤٣}^2 + \text{.....} + \text{١٧٦}^2 + \text{١٠٩}^2$$

$$= \text{١٨٨١١٢}$$

$$\therefore \text{ع}^2 = \frac{\text{١٨٨١١٢} - \frac{(\text{١٣٤٨})^2}{\text{١٠}}}{\text{١٠} - \text{١}} = \frac{\text{١٨٨١١٢} - \text{١٨١٧١٠,٤}}{\text{٩}}$$

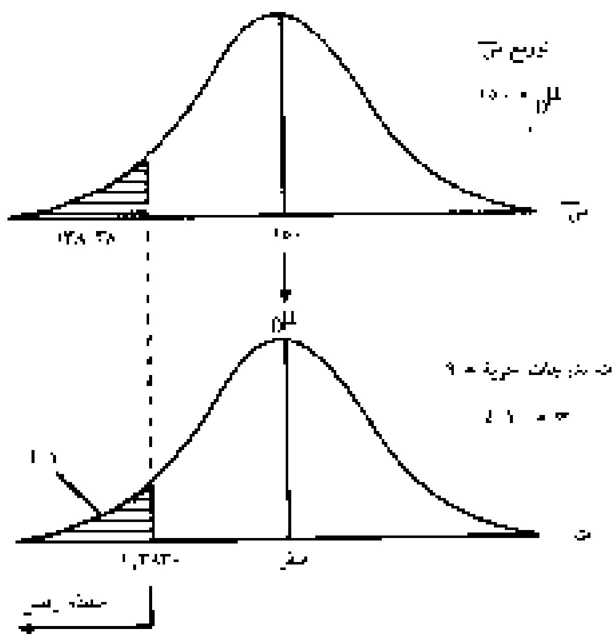
$$= \frac{\text{٦٤٠١,٦}}{\text{٩}} = \text{٧١١,٢٨٩} \qquad \therefore \text{ع} = \text{٢٦,٦٧ سيارة}$$

٣- دلت الخبرة السابقة أن توزيع مرور السيارات فى الساعة يتبع التوزيع الطبيعى ، ويكون الإختبار الإحصائى المناسب هو إختبارى ت بدرجات حرية = (ن - ١) = (١٠ - ١) = ٩ .

بمستوى معنوية ١٠ ٪ ، نجد أن المنطقة الحرجة هى المنطقة < - ت٫١٠ < أى ≥ -١٫٣٨ كما يتضح فى الشكل (٢-١٢) .

٤- الخطأ المعيارى لمتوسط العينة : $ع_{\bar{س}} = \frac{ع}{\sqrt{ن}} = \frac{٢٦,٦٧}{\sqrt{١٠}} = ٨,٤$

٥- إحصائية الإختبار : $ت = \frac{\bar{س} - \mu}{\sigma_{\bar{س}}} = \frac{١٥٠ - ١٣٤,٨}{٨,٤} = - ١,٨$



شكل (٢-١٢) : مناطق القبول والرفض للإختبار فى تطبيق (٢-٣)

حيث أن القيمة المحسوبة تقع داخـــل منطـــقة الرفض .

٥- يرفض الفرض العدمي ونصل إلى النـــتيجة بأن ضغط المرور في الساعة لا يـــستدعى إضافة إشارة مرور

الوصول إلى قرار بدلالة س̄

نرفض الفرض العدمي إذا كان متوسط عدد الـــسيارات المارة س̄ :

$$\leq \mu - ت \times ع_{\bar{س}}$$

$$\leq ١٥٠ - ١,٣٨٣ \times ٨,٤$$

$$\leq ١٣٨,٣٨$$

شكل (٢-١٢) يوضح أن قيمة ت الجدولية = ت٢٩، ١٠,٠ = ١,٣٨٣ يقابلـــها قيمة س̄ = ١٣٨,٣٨ . والقيمتان تـــفصلا بين مناطق القـــبول ومـــناطق الرفض .

وقيمة س̄ المحـــسوبة من العينة ١٣٤,٨ يقابلها قيمة ت المعيارية = ١,٨ .

تطبيق (٢-٤)

أراد مدير الأفراد بـــإحدى الهيئات الحكومـــية معرفة متوسط عـــدد الأسابيع التي ينتظرها طالب الوظيفة حتى إنهاء إجراءات التعيين . ومن المعروف أن هذه المدة تزيد على ٢٠ أسبوعًا .

لإجراء الإختبار الخاص بهذا الهدف ، قام هذا المدير بالخطوات الآتية :

١- صياغة الفروض :

الإعتقاد السائد بأن هذه المدة تزيد على ٢٠ أسبوعًا :

∴ ف. : $\mu \leq$ ٢٠ (وهو ما يود الباحث نفيه)

ف١ : $\mu >$ ٢٠ (وهو ما يود الباحث تأكيده)

٢- قام المسئول بسحب عينة عشوائية من ٥٠ موظف ، تسلموا العمل في سنوات سابقة ، ووجد أن زمن إنتظارهم لإنهاء إجراءات التوظيف كان متوسطه ٢١,٧ أسبوعاً بانحراف معيارى ٧,٤ أسبوع .

هنا وجد المسئول أن سَ (٢١,٧) أكبر من القيمة ٢٠ أسبوع وهذا معناه أن سَ (٢١,٧) تقع في داخل منطقة القبول ، حيث أن منطقة الرفض على يسار التوزيع ، فإننا نتوقف عند هذه الخطوة ونصل إلى القرار بعدم رفض الفرض العدمى .

المعنوية الإحصائية

١- إذا تم رفض الفرض العدمى ، فإن الإختبار يكون إختبارا معنويا إحصائياً Statistically Significant .

٢- الإختبار الإحصائى "حساس" بالنسبة لحجم العينة . فكلما زاد حجم العينة كلما "صغرت" قيمة الخطأ المعيارى وكلما أدى ذلك إلى "إرتفاع" قيمة الإختبار الإحصائى ، وزيادة قيمة الإختبار الإحصائى تؤدى إلى إحتمال أكبر لرفض الفرض العدمى .

٣- **الفرق المعنوى إحصائيا** : هو الفرق بين قيمة س (التقدير لنقطة) وقيمة المعلمة تحت صحة الفرض العدمى (μ_0) ، منسوبا إلى قيمة الخطأ

المعياري . فعند رفض الفرض العدمي يقال أن هناك فرقاً معنوياً إحصائياً بين متوسط العينة والقيمة المفترضة للمجتمع .

٤- **الفرق المعنوى عمليا** : هو الفرق بين قيمة معلمة المجتمع المفترضة ، تحت صحة الفرض العدمي ، وقيمة أخرى محددة بمتخذى القرارات أو الفرق العملي الذي يدفع إلى إجراء الإختبار الإحصائي . مثلا ، في تطبيق (٢-٢) قد يرى متخذوا القرار أن الفرق بين (٢٥,٧٧٧ - ٢٦٠٠٠) يستحق إجراء إختبارات المعنوية الإحصائية . وربما لا يروا أن هذا الفرق يستدعي إجراء الإختبار . وربما يصلوا إلى أن فرق ٢٢٣ جنيه سنويا لا يستحق كل هذا ، ويقبلوا الفرض العدمي بأن متوسط الأجر = ٢٦٠٠٠ سنويا. هنا يسمى الفرق «غير معنويا عمليا» ، بالرغم من أن الإختبار وجد أنه معنوى إحصائيا .

وفي تطبيق (٢-١) تم رفض الفرض العدمي : ف. : μ = ٢٥٠ جم أى أن الإختبار معنوى إحصائيا . ولكن قد يرى متخذوا القرارات أن الفرق بين متوسط العينة ومتوسط المجتمع المفترض (٢٥٠,٩٧ - ٢٥٠) = ٠,٩٧ جم فرق صغير جدا ، ولا يستحق توقف الإنتاج إلى أن يتم ضبط المعدات . وهنا نجد أن ٠,٩٧ فرق غير معنوى عملي .

تمارين

١- الإختبار فى اتجاه واحد ، والمنطقة الحرجة على يسار توزيع المعاينة . أوجد المنطقة باستخدام توزيع ت فى الحالات الآتية :

(أ) درجات الحرية = ٢٣ ، α = ١٠٪

(ب) درجات الحرية = ١٧ ، α = ٥٪

٢- الإختبار في إتجاهين إوجـــد المنطقة الحرجة باستخدام توزيع ت في الحالات الآتية :

(أ) درجات الحرية = ٧٤ ، ∝ = ٪١

(ب) درجات الحرية = ٢٤ ، ∝ = ٪٥

٣- إوجـــد المنطقة الحـــرجة باستخدام توزيع ت ، لاختبار الفـــرض العدمي : لم ≥ ٣٠٠ في الحالات الآتية :

(أ) ن = ٢١ ، ∝ = ٪١٥

(ب) ن = ٢١ ، ∝ = ٪٥

٤ - إوجـــد المنطقة الحـــرجة باستخدام توزيع ت ، لاختبار الفـــرض العدمي : لم ≤ ٢٦ في الحالات الآتية :

(أ) ن = ٤٣ ، ∝ = ٪٥

(ب) ن = ١٠ ، ∝ = ٪١٠

٥- يهتم المهندسون بمعرفة سرعة دوران المراوح الكهربائية لتوليد الكهرباء . فإذا كان متوسط سرعة الدوران ١٤ ميل في الساعة .

حاول بعض المهندسين التنفيذين تغيير موقع المراوح الكهربائيـة ووضعها فوق جبل ، واختبر سرعة دوران ٤٥ مروحه ووجد ان متوسط السرعه ١٤,٩ ميل في الساعة ، بانحراف معياري ٣,٨ ميل في الساعة .

(أ) إختبر الفرض العدمي أن الموقع الجديد غير مناسب لزيادة سرعة الدوران .

(ب) ما هي الإفتراضات ، لصحة الإختبار اعلاه .

٦- إدعت بعض شركات البناء والتعمير بأنهم يمكنهم الإنتهاء من عمليات تشطيب أي منزل في مدة لا تزيد عن ٩٠ يوما . أردت أن تتأكد من صحة هذا الإدعاء لعدد من المنازل التي تم تشطيبها بمعرفة هذه الشركة وحصلت على عدد الأيام التالية التي لزمت لتشطيب عدد ٢٤ منزلا :

٩٢	٧٣	٧٥	٧٩	٨٠	٨٢	٨٤	٨٦	٨٧	٨٨	٨٩	٩٠
١١٨	٩٣	٩٥	٩٨	٩٩	١٠١	١٠٣	١٠٤	١٠٦	١٠٩	١١٠	١١٣

إفترض أن توزيع عدد أيام تشطيب المنازل يقترب من التوزيع الطبيعي المتماثل وذو منوال واحد . إستخدام ∝ = ٥ ٪ قم بالإختبار الإحصائي المناسب لقبول الإدعاء أعلاه من عدمه ، في حاله .

(أ) الباحث يتفق مع إدعاء هذه الشركة .

(ب) الباحث يشك في إدعاء هذه الشركة .

٧- في شركات الأعمال الكبرى ، يعين فرد إضافي للرد على التليفونات إذا كانت فترة انتظار العميل للرد عليه أكثر من ٨ ثواني . وفي متابعة لزمن إنتظار الرد على العميل في ٢٥ مكالمة ، إتضح أن متوسط فترة الإنتظار ٩,٢ ثانية بانحراف معياري ٤,٣ ثانية .

استخدام ∝ = ٥ ٪ ، إختبر الفرض العدمي بعدم الحاجة إلي تعيين موظف إضافي .

٨- بالإشارة إلى تمرين (٧) أعلاه ، أعد الإختبار ، بفرض أن ن = ٦٠ ، ن = ١٢٠ ، وقارن بين النتائج في (٧) وفي (٨) . ما هي النتيجة التي توصلت إليها بشأن زيادة حجم العينة ؟

٩- في التمرين (٥) ، إذا كان حجم العينة ٧٥ مروحة . أعد الإختبار . وما هي نتيجة زيادة حجم العينة ؟

١٠- بفرض أن الفرض العدمي هو : $\mu \geq 50$ ، ومتوسط العينة ٤٥ :

(أ) ما هي إشارة القيمة الحرجة ؟

(ب) ماهي إشارة قيمة الإختبار الإحصائي ؟

(جـ) ما هي نتيجة الإختبار . قبول أو عدم قبول الفرض العدمي ؟

إعط شرح لإجاباتك لـ (أ) ، (ب) ، (جـ) .

٤/٢ إختبار النسبة (العينات الكبيرة)

يحتاج متخذوا القرارات في غالبية الأحيان معرفة النسبة الخاصة بظاهرة معينة في المجتمع . مثلا ، في مجال التسويق : ماهي نسبة العملاء في المجتمع التي تعرف منشأ المنتج ؟ وفي مجال التصنيع : ما هي النسبة في المجتمع التي تفضل السيارات الصغيرة ؟ وفي مجال التعليم - ما هي النسبة في المجتمع التي تؤيد التعليم الخاص ؟ ، يمكن تقدير هذه النسب باستخدام أسلوب التقدير باستخدام فترات الثقة التي تم عرضها في الباب السابق .

في هذا الباب ، سوف نختبر صحة نسبة (معروفة) مقدماً . مثلا نختبر عما إذا كانت النسبة في المجتمع التي تعرف منشأ منتج معين تساوي ٣٠٪ أو أكثر ؟ وعما إذا كانت نسبة من يفضلون السيارات الصغيرة الحجم أقل من ٤٠٪ ، وعما إذا كانت نسبة من يؤيد التعليم الخاص مساوية ٢٥٪ أو أقل .

وخطوات إجراء إختبار النسبة من مجتمع واحد تتبع الخطوات الآتية .

١ - صياغة الفروض : تأخذ الفروض أحد الأشكال الآتية :

(أ) $ف_{٠} : ل = ل_{٠}$ ، $ف_{١} : ل \neq ل_{٠}$

(ب) $ف_{٠} : ل \leq ل_{٠}$ ، $ف_{١} : ل > ل_{٠}$

(جـ) $ف_{٠} : ل \geq ل_{٠}$ ، $ف_{١} : ل < ل_{٠}$

حيث $ل_{٠}$ هي النسبة المفترضة في المجتمع .

٢ - سحب عينة عشوائية ، وحساب النسبة في العينة : $\hat{ل}$. وعلى الباحث ، هنا ، أن يقرر هل يستمر في الإختبار أم يتوقف عند هذه الخطوة . فإذا كانت $\hat{ل}$ في إتجاه النسبة تحت صحة الفرض العدمي ، تقبل الفرض العدمي ونتوقف عند هذه الخطوة ، وإلا نستمر في الإختبار .

٣ - تحقق من أن إختبار ى هو الإختبار المناسب ، وذلك إذا كان :

(أ) $ن \geq ٣٠$ ، (ب) $ن ل_{٠} \geq ٥$ ، (جـ) $ن (١ - ل_{٠}) \geq ٥$

إذا كان إختبار ى هو الإختبار المناسب ، تحدد المنطقة الحرجة بالقيمة $\pm ى_{ح/٢}$ في حالة الإختبار في الإتجاهين ، وبالقيمة $(+ ى_{ح})$ في حالة الإختبار في إتجاه واحد ومنطقة الرفض على يمين التوزيع . وبالقيمة $(- ى_{ح})$ في حالة الإختبار في اتجاه واحد ومنطقة الرفض على يسار التوزيع .

٤- إحسب الخطأ المعياري للنسبة :

(١) المجتمع غير محدود ، ه/ن < ٠,٠٥

$$\text{ع}_{\text{ل}} = \sqrt{\frac{\text{ل}_{٠} \,(\,١ - \text{ل}_{٠}\,)}{\text{ه}}}$$

(٢) المجتمع محدود ، ه/ن > ٠,٠٥

$$\text{ع}_{\text{ل}} = \sqrt{\frac{\text{ل}_{٠} \,(\,١ - \text{ل}_{٠}\,)}{\text{ه}}} \sqrt{\frac{\text{ن} - \text{ه}}{\text{ن} - ١}}$$

٥- إحسب الإختبار الإحصائي :

الإختبار الإحصائي :

$$\text{ى} = \frac{\hat{\text{ل}} - \text{ل}_{٠}}{\text{ع}_{\text{ل}}}$$

حيث ى تتبع التوزيع الطبيعي المعياري .

٦- .تخذ قرار بشأن الفرض العدمي ، كالآتي :

علامة الفرض البديل	رفض الفرض العدمي إذا كانت .
$\neq$	$\lvert$ ى المحسوبة $\rvert > \lvert$ ى$_{\alpha/2} \rvert$
$<$	ى المحسوبة $< -$ ى$_{\alpha}$
$>$	ى المحسوبة $>$ ى$_{\alpha}$

ويلاحظ ان إختبار النسبة يشبه إلي حد كبير إختبار الوسط الحسابي . ففي الإختبارين ، يحسب الإختبار الإحصائي بنسبه الفرق بين إحصاء العينة والقيمة تحت صحة الفرض العدمي ، منسوب إلي الخطأ المعياري لإحصاء العينة .

إختبار النسبة | إختبار المتوسط

$$\text{ى} = \frac{\hat{\text{ل}} - \text{ل}_{٠}}{\sigma_{\hat{\text{ل}}}} \qquad\qquad \text{ت} = \frac{\bar{\text{س}} - \mu}{\sigma_{\bar{\text{س}}}}$$

والفرق في بسط معادلة النسبة ، يوضح الفرق بين النسبة في العينة والنسبة المفترضة في المجتمع ، هذا الفرق يجب أن يكون فرقا منسوباً إلى الخطأ المعياري للنسبة من العينة . وهنا يجب أن نوضح أن إختبار ت لا يستخدم لإختبار النسبة في المجتمع . فإذا كان حجم العينة صغير ، أو لم يتحقق الثلاث شروط أعلاه ، بمعنى أن $\text{ن} > ٣٠$ ، $\text{ن ل}_{٠} > ٥$ ، $\text{ن} (١ - \text{ل}_{٠}) > ٥$ ، فيجب إختبار النسبة باستخدام توزيع ذات الحدين أو تعديل إختبار الإشارة Sing test ليناسب إختبار النسبة في حالة العينات الصغيرة (إختبار الإشارة هو أحد الإختبارات اللامعلمية) .

تطبيق (٢-٥)

عرف أن أحد شركات شحن البضائع تتسبب في مفقودات وتلف بنسبة ٢,٣ ٪ . أراد أحد الموردين إختيار شركة شحن أخرى لمدة ستة شهور ومعرفة هل نسبة المفقودات والتلف في الشركة الجديدة أقل من ٢,٣ ٪ أم لا .

١- صياغة الفروض تكون كالآتي :

ف٠ : ل ≥ ٠,٠٢٣

ف١ : ل < ٠,٠٢٣

٢- بعد نهاية الشهور ، وجد أن عدد الطرود التي تم شحنها ٥٧٣ طردا وكان عدد الطرود المفقودة والتالفة ١٠ .

تكون شواهد العينة إذن أن النسبة في العينة = ١٠ / ٥٧٣ = ٠,٠١٧ ، ومنه القيمة ٠,٠١٧ تختلف وأقل من القيمة تحت صحة الفرض العدمي وتقع في إتجاه منطقة الرفض . لذا نستمر في الاختبار .

٣- حجم العينة = ٥٧٣ > ٥

ن ل٠ = ٥٧٣ × ٠,٠٢٣ = ١٣ ≥ ٥

ن (١ - ل٠) = ٥٧٣ × (١ - ٠,٠٢٣) = ٥٦٠ ≥ ٥

∴ العينة كبيرة ، ويقترب توزيع معاينة النسبة من العينة بالتوزيع الطبيعي .

إذا أخترنا α = ٥ ٪ نجد أن (ي٠,٠٥) = (- ١,٦٥)

∴ القيمة الحرجة هي منطقة الرفض وهي المنطقة أقل من أو تساوي (-١,٦٥) .

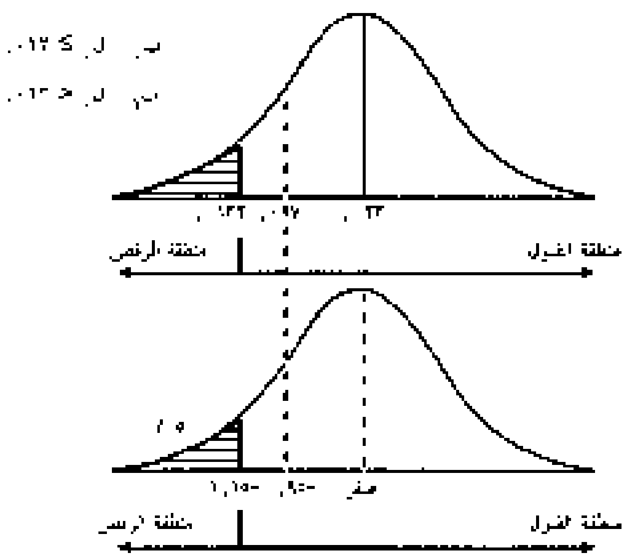
٤- الخطأ المعياري للنسبة في العينة (مجتمع غير محدود) :

$$ع_{\hat{ل}} = \sqrt{\frac{ل_0\,(1 - ل_0)}{ن}} = \sqrt{\frac{0{,}023\,(1-0{,}023)}{573}} = 0{,}0063$$

٥- الإختبار الإحصائي :

$$ي = \frac{\hat{ل} - ل_0}{ع_{\hat{ل}}} = \frac{0{,}017 - 0{,}023}{0{,}0063} = -\,0{,}95$$

٦- حيث أن قيمة ي المحسوبة (- ٠,٩٥) أكبر من قيمة ي الحرجة (- ١,٦٥) وتقع في منطقة القبول (شكل (٢-١٣)) ، إذن ، لا نرفض الفرض العدمي ، ونستطيع القول أن النسبة في الشركة الثانية أكبر من أو تساوي ٠,٠٢٣ ، وليست أقل من ذلك .



شكل (٢-١٣) : مناطق القبول والرفض في التطبيق (٢-٥)

ولمعرفة القيمة الحرجة بدلالة لَ نطبق المعادلة :

ل. = ى − ع × عِ لَ

= ٠,٠٢٣ − ١,٦٥ × ٠,٠٠٦٣ = ٠,٠١٣٢

∴ نرفض الفرض العدمي إذا كانت لَ (النسبة المحسوبة من العينة أقل من ٠,٠١٣٢ . وحيث أن لَ (٠,٠١٧) أكبر من قيمة لَ الحرجة (٠,٠١٣٢) ،

∴ لا نرفض الفرض العدمي (شكل (٣–١٧) .

<u>تطبيق (٢–٦)</u>

فى تطبيق (٢–٥) ، إذا تم مراجعه عدد أكبر من الطرود التى تم شحنها بواسطة الشركة الجديدة . فتم مراجعة ٤٤١٧ طردا ووجد أن عدد الطرود المفقوده والتالفة ٧٥ طردا . هل هناك شواهد تؤيد الفرض العدمى ؟ والفرض البديل ؟

نجد أن الفروض تصاغ بنفس الطريقة السابقة :

ف. : ل ≤ ٠,٠٢٣

ف١ : ل > ٠,٠٢٣

لَ = ٧٥ / ٤٤١٧ = ٠,٠١٧

وهى نفس النسبة السابقة ولكن تختلف قيمة الخطأ المعيارى حيث ان حجم العينة قد إختلف .

$$\text{ع}_{\text{ل}} = \sqrt{\frac{\text{ل}_{.}(1 - \text{ل}_{.})}{\text{ن}}} = \sqrt{\frac{0,023\,(1 - 0,023)}{4417}} = 0,0023$$

وقيمة الإختبار الإحصائى :

$$\text{ي} = \frac{\hat{\text{ل}} - \text{ل}_{.}}{\text{ع}_{\text{ل}}} = \frac{0,017 - 0,023}{0,0023} = -2,61$$

وحيث أن ي المحسوبة (- ٢,٦١) تقع داخل منطقة الرفض ، يكون القرار : رفض الفرض العدمى وقبول الفرض البديل بأن نسبة الظروف المفقودة والتالفة فى الشركة الجديدة اقل من ٢,٣٪ .

ولحساب قيمة ل^ الحرجة :

$$= \text{ل}_{.} - \text{ي}_{\text{م}} \times \text{ع}_{\text{ل}}$$

$$= 0,023 - 1,65 \times 0,0023 = 0,019$$

وحيث أن ل^ < قيمة ل^ الحرجة (٠,٠١٧ < ٠,٠١٩)

∴ نرفض الفرض العدمى ، حيث أن ٠,٠١٧ تقع داخل منطقة رفض الفرض العدمى .

من نتائج تطبيق (٢-٥) ، (٢-٦) يتضح أن إختبار النسبة حساس بالنسبة لحجم العينة . وكلما زاد حجم العينة ، كلما إنخفض قيمة الخطأ المعيارى ، وارتفعت قيمة الإختبار الإحصائى .

تطبيق (٢-٧)

حذرت إحدى شركات الأدوية بأن العقار الجديد له أعراض جانبية ويسبب ارتفاع في ضغط الدم يجب عدم تعاطيه . وطلب مدير هذه الشركة بمتابعة بعض المرضى المتعاطين لهذا العقار وملاحظة ضغط الدم . وأن العقار سوف يصرح به إذا تسبب في ارتفاع ضغط الدم لنسبة ٤ ٪ من المتعاطين على الأكثر .

١- صياغة الفروض

حيث أن شركة الأدوية لا ترغب في أن يكون لهذا العقار أعراض جانبية ، ولا ترغب أن تكون نسبة الأعراض الجانبية مرتفعة ، يكون الفروض كالآتي :

$ف_0$: ل $\geq$ ٠,٠٤ (ما تود الشركة نفيه)

$ف_1$: ل $<$ ٠,٠٤ (ما تود الشركة تأكيده)

٢- بمراجعة سجلات المرضى الذين تعاطوا العقار لمدة ٦ شهور ، وجد أن ١٣ مريضاً من جملة ٧٢٨ مريضاً تعاطو العقار تسبب لهم إرتفاع في ضغط الدم .

$$\therefore \hat{ل} = \frac{١٣}{٧٢٨} = ٠,٠١٨$$

وحيث أن $\hat{ل}$ (٠,٠١٨) أقل من $ل_0$ (٠,٠٤) $\therefore$ نستمر في الإختبار خاصة وأن قيمة $\hat{ل}$ تقع في اتجاه منطقة الرفض .

٣- إذا اخترنا α = ٠,٠٠٢ . وهذا المستوى من المعنوية يجعل إحتمال إستمرار إستخدام العقار ونسبه الأعراض الجانبية كبير ، إحتمالا بسيطاً . أي احتمال قبول الفرض العدمي ($ل_0$ $\geq$ ٠,٠٤) وهو فرض صحيح (قرار صائب) ، إحتمال كبير .

وحيث أن الإختبار في اتجاه واحد ، والمنطقة الحرجة على يسار التوزيع .
نجد أن القيمة الحرجة = $(-\text{ى}_{٠,٠٠٢}) = (-٢,٨٨)$.

ونستخدم التوزيع الطبيعي للأسباب الآتية :

$$\text{ن} = ٧٢٨ \geq ٣٠$$

$$\text{ن ل} = ٧٢٨ \times ٠,٠٤ = ٢٩ \geq ٥$$

$$\text{ن}(١-\text{ل}) = ٧٢٨(١-٠,٠٤) = ٦٩٩ \geq ٥$$

٤ - الخطأ المعياري للنسبة (المجتمع غير محدود) .

$$\text{ع}_{\hat{\text{ل}}} = \sqrt{\frac{\text{ل}(١-\text{ل})}{\text{ن}}} = \sqrt{\frac{٠,٠٤(١-٠,٠٤)}{٧٢٨}} = ٠,٠٠٧$$

٥ - الإختبار الإحصائي :

$$\text{ى} = \frac{\hat{\text{ل}} - \text{ل}}{\text{ع}_{\hat{\text{ل}}}} = \frac{٠,٠١٨ - ٠,٠٤}{٠,٠٠٧} = -٣,١٤$$

وحيث أن قيمة ى المحسوبة (− ٣,١٤) أقل من قيمة ى الحرجة (−٢,٨٨) ∴ ترفض الفرض العدمي بمستوى معنوية ٠,٠٠٢ ، ونستطيع القول أن نسبة الاعراض الجانبية من هذا العقار أكبر من ٤ ٪ ، وبالتالي لا يجب أن يعطى تصريحاً لهذا العقار .

ولحساب قيمة $\hat{\text{ل}}$ الحرجة :

$$\hat{\text{ل}} = \text{ل} - \text{ى}_{\text{ن}} \times \text{ع}_{\hat{\text{ل}}}$$

$$= ٠,٠٤ - ٢,٨٨ \times ٠,٠٠٧ = ٠,٠١٩٨$$

ونرفض الفرض العدمي إذا كانت لـ^ أقل من لـ^ الحرجة .

وحيث أن لـ^ (١٨ ، ٠) < لـ^ الحرجة (١٩٨ ، ٠) .

∴ نرفض الفرض العدمي ، ونستطيع القول أن للعقار أعراض جانبية يتأثر بها نسبة أكبر من ٤ ٪ .

تمارين

١١- قامت شركة تسويق عن طريق البريد بإرسال كوبونات خصم لعملائها . وكان إعتقاد هذه الشركة أنه إذا إستجاب ٢٠ ٪ أو أكثر من العملاء لكوبونات الخصم الخصم يكون وسيط الإعلان عن طريق البريد ناجحًا ، وإلا فسوف تضطر الشركة للبحث عن وسيط آخر لتسويق منتجاتها .

تم إرسال الكوبونات إلى مائة عميل ، واستجاب لهذه الكوبونات ٢٣ عميلا .

إستخدم ∝ = ١ ٪ . واتخذ قرارا بشأن وسيلة التسويق .

١٢- حدث قلق في قسم حسابات العملاء لأحد الشركات ، لأن نسبة الفواتير المتأخرة الدفع كانت كبيرة . فاتفق على أنه إذا زادت هذه النسبة عن ١٠ ٪ سوف تتخذ الشركات إجراءات جديدة للتحصيل . تم سحب عينة من ٧٥ حسابا ووجد أن ٩ منهم عليهم فواتير متأخرة

قم بصياغة الفروض واستخدم ∝ = ١ ٪ لتقدم النصح للشركة : هل تتخذ إجراءات جديدة للتحصيل أم لا ؟

١٣- قررت شركة بيع ملابس أنها سوف تتسع في مساحة عرض ملابسها إذا كان أكثر من ٣٥ ٪ من مبيعاتها تزيد قيمتها عن ١٥٠ جنيها . تم سحب عينة عشوائية من مبيعات ٧٥ عميلا ، ووجد أن ٢٨ عميلا تزيد مبيعاتهم عن ١٥٠ جنيه .

هل تنصح الشركة بالتوسع في مساحة العرض ؟ استخدم ∝ = ١٠ ٪

١٤- إذا كان الفرض العـدمي أن النسبة في المجتمع ٨٥, أو أكثر وأكدت نتائج العينة أن النسبة في العينة ٨٧, ٠ .

اشرح لماذا تقبل الفرض العدمي بمقارنة النسبتين أعلاه .

١٥- إذا عرف أن نسبة أخطاء المراجعة في أحد البنوك ٧,٣ . تم سحب ١٥٠ ملفا من ملفات القروض ووجد أن ٨ منهم بهم أخطاء . استخدم مستوى معنوية ٢٪ وقرر هل أخطاء المراجعة في هذا البنك ٧,٣٪ على الأكثر أم لا ؟

٥/٢ اختبار تباين المجتمع

كما أمكننا إختبار قيمة مفترضة لمتوسط المجتمع وللنسبة في مجتمع ، يمكننا أيضا إختبار تباين المجتمع . مثلا إذا أردت التحقق من أن المورد قد قام بتوريد ألواح ألمنيوم لا يختلف سمكها عن ٠,٦ ملليمتر ، أو أن بعض أطباق الصدف المصنعه يدويًا لا يختلف أقطارها بأكثر من ٠,١٥ سم .

يمكننا إختبار درجة الإختلاف هذه باستخدام التباين . والتباين يعكس إختلاف القيم عن الوسط الحسابي . والإختبار الذي يستخدم لاختبار التباين هو إختبار كا^{٢} Chi - squar test .

ولقد توصل كارل بيرسون Karl Pearson إلى أن العينات التي تم سحبها من مجتمع ذو توزيع طبيعي ، بانحراف معياري σ ، فإن النسبة :

$$\frac{(ن - ١)\, ع^{٢}}{\sigma^{٢}}$$

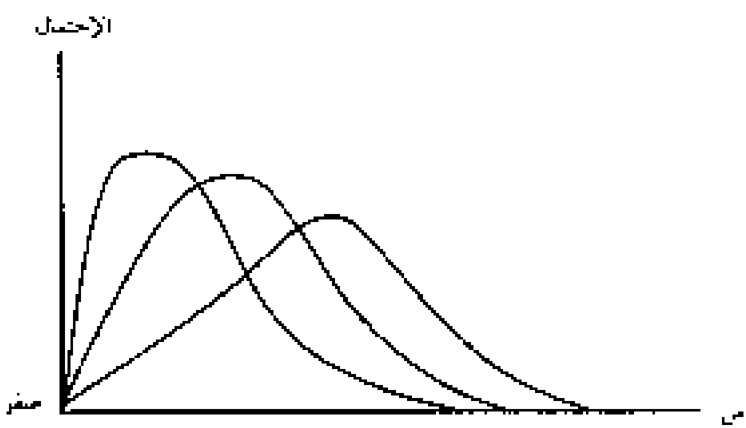
يمكن أن توصف بتوزيع كا^2 بدرجات حرية = (ن - ١) . وهذه النسبة هي أساس الإختبار الذى يستخدم لاختبار أن التباين فى المجتمع يساوى قيمة معينة مفترضة .

وتوزيع كا^2 ، مثله مثل توزيع ت عبارة عن عائلة من التوزيعات ، يعتمد على درجات الحرية . ولا يوجد توزيع فعلى لتوزيع كا^2 أو توزيع ت ، ولكن يوجد عدد من التوزيعات . وكلما زادت درجات الحرية كلما « إقترب » توزيع كا^2 من التوزيع الطبيعى .

ولكن توزيع كا^2 يختلف عن توزيع ت من زاويتين :

١- جميع قيم كا^2 موجبه

٢- توزيع كا^2 توزيع موجب الإلتواء فى حين أن توزيع ت توزيع متماثل . ولكن درجة الإلتواء فى توزيع كا^2 تتلاشى كلما زادت درجات الحرية كما يتضح فى شكل (٢-١٤) .



شكل (٢-١٤) : توزيع كا^2 بدرجات حرية مختلفة

وسوف نعرض كيفية إستخدام جدول توزيع كا^{2} بعد عرضنا لخطوات إختبار التباين من مجتمع واحد .

حين تتبع الظاهرة في المجتمع التوزيع الطبيعي ، تكون خطوات إختبار أن تباين المجتمع يساوي قيمة معينة كالآتي :

١ - صياغة الفروض :

$$ف_{0} : \sigma^{2} = \sigma_{0}^{2}$$

$$ف_{1} : \sigma^{2} \neq \sigma_{0}^{2}$$

أو $ف_{0} : \sigma^{2} \leq \sigma_{0}^{2}$ ضد $ف_{1} : \sigma^{2} > \sigma_{0}^{2}$

أو $ف_{0} : \sigma^{2} \geq \sigma_{0}^{2}$ ضد $ف_{1} : \sigma^{2} < \sigma_{0}^{2}$

حيث σ_{0}^{2} هي القيمة المفترضة لتباين المجتمع . والصيغة الأولى تعكس إختبارا في اتجاهين ، والصيغتان الأخيرتان تعكسا إختبارين كل في اتجاه واحد . وفي إختبارات كا٢ بصفة عامة ، الفرض العدمي ، هو الفرض الذي يود الباحث قبوله (بعكس الإختبارات التي تم عرضها حتى الآن) .

٢ - يسحب عينة عشوائية من المجتمع حجمها ن ، ويحسب التباين من بيانات هذه العينة ، حيث :

$$ع^{2} = \frac{مجـ (س - \bar{س})^{2}}{ن - 1} = \frac{ن\ مجـ س^{2} - (مجـ س)^{2}}{ن (ن - 1)}$$

وهنا تقرر عما إذا كنت تستمر في خطوات الإختبار التالية أم تتوقف . إذا كانت ع^{٢} تقع في نطاق الفرض العدمي ، يقبل الفرض العدمي عند هذه النقطة ونتوقف ولا نستمر في خطوات الإختبار ، وإذا كانت ع^{٢} لا تقع في نطاق الفرض العدمي نستمر بالخطوة (٣) التالية .

٣- حدد قيمة α ، وحدد قيمة كا^{٢} الحرجة باستخدام جدول كا^{٢} الإحتمالي بدرجات حرية (ن − ١) . طبقًا لإتجاه الفرض العدمي والفرض البديل .

٤- إحسب قيمة إختبار كا^{٢}

$$كا^{٢} = \frac{(ن - ١)\ ع^{٢}}{\sigma^{٢}} = \frac{مجـ\ (س - \bar{س})^{٢}}{\sigma^{٢}}$$

٥- اتخذ قرار بشأن الفرض العدمي كالآتي :

اتجاه الفرض البديل	رفض ف. إذا كانت :
≠	القيمة المحسوبة > كا$^{٢}_{\alpha/٢}$ أو < كا$^{٢}_{(١-\alpha/٢)}$
>	القيمة المحسوبة < كا$^{٢}_{(١-\alpha)}$
<	القيمة المحسوبة > كا$^{٢}_{\alpha}$

قيم كا^2 الحرجة

جدول توزيع كا^2 الإحتمالي معطى في الملحق (جدول (٣)) . تمثل الصفوف في هذا الجدول درجات الحرية (d.f) ، وتشير الأعمدة إلى مساحة الذيل الأيمن في التوزيع . وقيمة كا^2 المطلوبة هي القيمة المعطاه عند تقاطع الصف مع العمود . ولتوضيح كيف يستخدم الجدول ، بفرض أن المطلوب هو قيمة كا^2 التي تجعل الذيل الأيمن ٥ ٪ ، درجات حرية (d.f) = ٥ ، نجد أن قيمة كا^2 الحرجة ١١,٠٧ . وقيمة كا^2 التي تجعل الذيل الأيمن = ٥ ٪ ، df = ٢١ مساوية ٣٢,٦٧١ .

ولإختبار الفرض في اتجاه واحد توضع α في الذيل الأيمن أو الذيل الأيسر للمنحنى . ولكن المنحنى غير متماثل ، كيف نحصل على قيمة كا^2 التي تجعل الذيل الأيسر = α ؟

حيث أن الجدول يعطي مساحة الذيل الأيمن . نجد أن قيمة كا^2 التي تجعل مساحة الذيل الأيسر من المنحنى = α هي نفس قيمة كا^2 التي تجعل مساحة الذيل الأيمن من المنحنى = $(1 - \alpha)$.

فإذا كان اتجاه الإختبار إلى اليسار يعني أن المنطقة الحرجة تقع على يسار التوزيع ، نجد أن القيمة الحرجة التي تحدد مساحة الذيل الأيسر α = كا$^2_{(1-\alpha)}$.

وإذا كان الإختبار في اتجاهين تكون القيمة الحرجة التي تجعل مساحة الذيل الأيمن $\alpha/2$ = كا$^2_{\alpha/2}$ ، والقيمة الحرجة التي تجعل مساحة الذيل الأيسر $\alpha/2$ = كا$^2_{(1-\alpha/2)}$.

تطبيق (٢-٨)

تعتمد شركة تصنيع سيارات ، إشتهرت بدقة إنتاجها ، على دقة قطر الرومان بلى لسياراتها . وأنه يجب أن يكون الانحراف المعيارى لقطر الرومان بلى المنتجة أصغر ما يمكن . وحددت الآتي : إذا رفضت الفرض العدمي وكانت ع أقل من ,٠٢٥ ملليمتر إستخدم الرومان بلي المنتجة ، وإذا كانت ع > ,٠٢٥ ملليمتر ، لا تستخدم الرومان بلى المنتجة ، ولكن إسحب مفردات إضافية من الإنتاج وأعد الإختبار .

∴ الفرض العدمي : $ف_٠ : \sigma^2 \geq (,٠٢٥)^2$

الفرض البديل : $ف_١ : \sigma^2 < (,٠٢٥)^2$

تم سحب عينة من ٢٠ رومان بلي ، وقيست الأقطار ، وتم حساب الانحراف المعيارى ، ووجد أنه مساويا ,٠١٨ ملليمتر . والإختبار الإحصائي الواجب إستخدامه :

$$كا^2 = \frac{ع^2(ن - ١)}{\sigma^2}$$

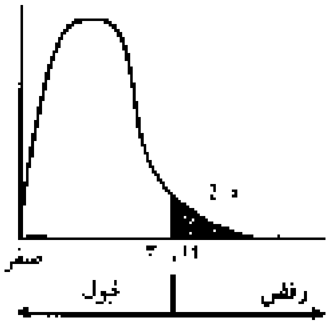
حيث التوزيع الاحتمالى لإحصائية الإختبار يتبع توزيع كا² الإحتمالي بدرجات حرية (٢٠-١) = ١٩ .

باستخدام جدول توزيع كا² ، بدرجات حرية (٢٠-١) = ١٩ ، α = ٥٪ نجد أن المنطقة الحرجة على يمين التوزيع ، والقيمة الحرجة = ٣٠,١٤ .

نحسب قيمة الإختبار الإحصائي

$$كا^{٢} = \frac{(ن - ١)\,ع^{٢}}{\sigma^{٢}}$$

$$= \frac{(٢٠ - ١)\,(٠,١٨)^{٢}}{(٠,٢٥)^{٢}} = ٩,٨٥$$



حيث أن قيمة $كا^{٢}$ المحسوبة (٩,٨٥) < القيمة الجدولية (٣٠,١٤)

∴ نقبل الفرض العدمي بأن الانحراف المعياري لأقطار الرومان يلى أقل من ٠,٢٥ .

تطبيق (٢-٩)

حاول مدير أحد البنوك معرفة درجة الإختلاف في مدد زمن الإنتظار في طابور خدمة العملاء في البنك . ورأى أنه إذا تحرك طابور خدمة العملاء بسرعة منتظمة ، فإن العملاء لا يتذمرون ، ولكن إذا لم يتحرك طابور خدمة العملاء بسرعة منتظمة فإن هذا ينتج عنه حالة من التذمر وعدم الرضا عن خدمات البنك بصفة عامة . وكان هدف مدير البنك أن يكون الانحراف المعياري لزمن الإنتظار خمس ثوان أو أقل .

تكون الفروض إذن :

$ف_{٠} : \sigma^{٢} \geq (٥)^{٢}$ ثوان

$ف_{١} : \sigma^{٢} < (٥)^{٢}$ ثوان

لاحظ موظف الشباك أن زمن الخدمة لعدد من العملاء كالآتي (بالثانية) :

٢٩ ، ٣٢ ، ٢٦ ، ٣٤ ، ٤٢ ، ٣٤ ، ٢٣

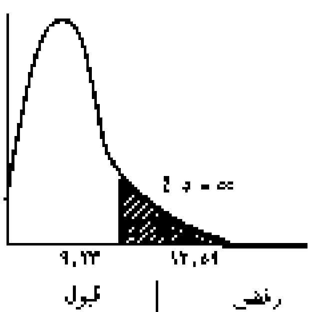
من هذه البيانات وجد أن :

$$\text{ن} = ٧ \quad ، \quad \text{مجـ س}^{٢} = ٧١٤٦ \quad ، \quad \text{مجـ س} = ٢٢٠$$

$$\therefore \text{ع} = \sqrt{\frac{\text{مجـ س}^{٢} - \frac{(\text{مجـ س})^{٢}}{\text{ن}}}{\text{ن} - ١}} = \sqrt{\frac{\text{ن مجـ س}^{٢} - (\text{مجـ س})^{٢}}{\text{ن} - ١}}$$

$$= \sqrt{\frac{٧ \times ٧١٤٦ - (٢٢٠)^{٢}}{(٧-١)}} = ٦,٢$$

بفرض أن زمن الإنتظار يمكن أن يقرب بالتوزيع الطبيعي إحصائية الاختبار لإختبار الفرض أعلاه هو :



$$\text{كا}^{٢} = \frac{(\text{ن} - ١)\,\text{ع}^{٢}}{\sigma^{٢}}$$

$$= \frac{(٧ - ١)\,(٦,٢)^{٢}}{(٥)^{٢}} = ٩,٢٣$$

حيث كا^{٢} يتبع توزيع كا^{٢} الإحتمالي بدرجات حرية (ن-١) .

وحيث أن المنطقة الحرجة ناحية اليمين ∴ القيمة الحرجة هي قيمة كا^{٢} بدرجات حرية = (ن - ١) = (٧ - ١) = ٦ وبمستوى معنوية ∝ = ٥٪ ، نجد أن القيمة الحرجة = ١٢,٥٩ .

وحيث أن قيمة كا^{٢} المحسوبة (٩,٢٣) تقع داخل منطقة القبول ، لا ترفض الفرض العدمي ، وهذا يدل على كفاءه موظفي الشباك حيث أن الانحراف المعياري لزمن الإنتظار أقل من ٥ ثوان .

تمارين

١٦- إوجد قيم كا^{٢} الحرجة للحالات الآتية :

(أ) ن = ٢٣ مساحة الذيل الأيمن = ٥٪

(ب) ن = ٩ مساحة الذيل الأيسر = ١٪

(جـ) ن = ١٧ مساحة الذيل الأيمن = ٥٪ ، والذيل الأيسر = ٥٪

(د) ن = ٤ مساحة الذيل الأيسر ٥٪

١٧- الآتي مجموعة من الفروض البديلة ، إوجد قيم كا٢ الحرجة للحالات الآتية :

(أ) ف١ : σ > ٨ ن = ١٣ α = ٥٪

(ب) ف١ : $\sigma \neq$ ١,٢ ن = ١٠ α = ١٠٪

(جـ) ف١ : σ < ٣,٦ ن = ٢١ α = ٥٪

(د) ف١ : $\sigma \neq$ ١١,٨ ن = ٨ α = ١٪

(هـ) ف١ : $\sigma \neq$,٤ ن = ١٦ α = ٥٪

(و) ف١ : σ > ٧,٤ ن = ١٩ α = ١٪

(ز) ف١ : $\sigma \neq$ ٠,٨٤ ن = ١٠ α = ١٠٪

١٨ - أنت مسئول عن مراقبة الشريط اللاصق (السيلوتيب) . وطبقاً للمعايير الدولية المتفق عليها ، يجب ألا يزيد درجة إختلاف المادة اللاصقة على الشريط عن ٠,٠٦ من المليمتر . قمت بسحب عينة عشوائية من الشرائط اللاصقة (السيلوتيب) ، حجمها ١٩ شريطاً واتضح من بيانات هذه العينة : أن الانحراف المعياري لثقل المادة اللاصقة = ٠,٠٨٤ ملليمتر .

إستخدم ∝ = ١٠٪ ، واختبر الفرض أن المادة اللاصقة لا تقابل المواصفات الدولية .

١٩- أراد مسئول في هيئة المواصلات السلكية واللاسلكية أن يكون متوسط الزمن المستغرق من موظف الهيئة الذى يعمل فى البحث فى الدليل أن يكون ٣٨ ثانية أو أقل ، وبانحراف معيارى ٦ ثوان أو أقل

قمت بملاحظة عينة عشوائية من بعض موظفي الدليل الجدد ، وسجلت الأزمنة التالية :

٣٨	٤٣	٣٠	٢٨	٣٤	٤٠	٣٨	٢٨	٢٦	٣٦
٤٠	٣٦	٣٣	٣٠	٣١	٢٩	٣٥	٣٨	٣٩	٣٩

(أ) إستخدم ∝ = ٥٪ إختبر عما إذا كان الإنحراف المعيارى للزمن المستغرق يقابل المواصفات كما وضعتها الهيئة .

(ب) ما هي إفتراضات الإختبار في (أ) أعلاه ؟

٢٠- أرادت شركة أدوية الإتساع في معدات إنتاجها ، خاصة الألات التى تشكل الكبسولات . إدعى مُصنّع آلة التشكيل أن الانحراف المعيارى لشكل الكبسولات المنتجة ٠,١٢٥ مليجرام أو أقل .

وقبل أن تقرر الشركة الشراء من هذا المصنع أو غيره . قامت بملاحظة ٢٠ كبسولة منتجة . ووجدت أن الانحراف المعياري لشكل الكبسولات ١٦٥, ملليمتراً .

إستخدم $\propto$ = ٥ ٪ واختبر الفرض العدمي :

(أ) ف ِ : $\sigma^2 \leq$ ١٢٥,

(ب) ف ِ : $\sigma^2 =$ ١٢٥,

(جـ) ف ِ : $\sigma^2 \geq$ ١٢٥,

٢١- في التمرين ٥ أعلاه ، أعد الإختبارات المطلوبة باعتبار أن عدد الكبسولات المنتجة = ١٠ بانحراف معياري ١٦٥,٠ ملليجرام .

✍ ✍ ✍

الباب الثالث

إختبارات الفروض : مجتمعين

الباب الثالث

إختبارات الفروض ، مجتمعين

في الباب السابق ، تم عرض كيفية إختبار الفروض الخاصة بوسط حسابي واحد والتباين من مجتمع واحد . فى هذا الباب نعرض إختبارات الفروض المتعلقة بالفرق بين متوسطى عينتين مستقلتين أو غير مستقلتين ، وأيضاً الفرق بين نسبتين من مجتمعين . وسوف نعرض في الملحق آخر هذا الباب ، كيفية إستخدام برمجية Minitab في التقدير واختبارات الفروض التي تم عرضها فى الأبواب الأول حتى الثالث .

يستخدم صانعوا القرار بكثرة إختبارات الفروض باستخدام عينتين ، خاصة عند الحاجة إلى مقارنة إختيارين أمامهم . فمثلاً هل متوسط المبيعات للسوق المحلي أعلى من متوسط المبيعات للتصدير ؟ هل نسبة المعيب فى صناعة بعض إطارات السيارات أعلى من إطارات مصنعة بواسطة شركة أخرى ؟ أن أساليب الإستدلال التى سوف يتم عرضها ، تسهل في عملية إتخاذ القرار ، وتجعل عملية إتخاذ القرار عملية علمية مستندة على أساس علمي ، وليس بالمشاهدة والحكم الشخصي فقط . .

ويجب هنا أن نفرق بين العينات المستقلة والعينات غير المستقلة . فالعينات تكون مستقلة إذا كان سحب كل مفردة مستقل تماماً عن سحب المفردة الأخرى. وسحب مفردات العينة الأولى مستقل تماماً عن سحب مفردات العينة الثانية . أما العينات غير المستقلة فهي عينات يعتمد سحب مفردات العينة الثانية على سحب مفردات العينة الأولى . وقد تكون نفس المفردات فى العينتين ، مثلاً درجات إمتحان أعمال السنة ، ودرجات آخر الفصل الدراسى لنفس الطلبة ، تمثل عينتين غير مستقلتين ، فنستخدم نفس المفردات ولكن قراءتين

لكل مفردة . وكمثال آخر مقارنة أطوال الآباء بأطوال الأبناء من سن ٢٠ سنة فأكثر . هنا نقوم بعملية مقارنة ، طول كل أب مع طول ابنه ، والعينتان في هذا المثال عينتان غير مستقلتان . أما إذا سحبنا عينة عشوائية من إطارات خط إنتاج معين ، وعينة عشوائية من إطارات خط إنتاج آخر ، فالعينتان مستقلتان .

١/٣ خطوات إختبارات الفروض

خطوات إختبار وجود فرق بين متوسطي مجتمعين أو النسبة في مجتمعين ست خطوات هي :

١ – صياغة الفروض .

٢ – الحصول على بيانات العينتين .

٣ – إختيار الإختبار الإحصائي ، وتحديد مستوى المعنوية ، والمنطقة الحرجة

٤ – حساب الخطأ المعياري .

٥ – حساب قيمة الإختبار الإحصائي .

٦ – إتخاذ قرار بشأن صحة الفرض العدمي .

ويجب ملاحظة أنه في الخطوة (الأولى) : وهي خطوة صياغة الفروض ، أن الفرض العدمي دائماً تصاحبه علامة = أو ≤ أو ≥ . فنقطة التساوي دائماً مصاحبة للفرض العدمي . والفرض البديل هو المكمل للفرض العدمي ، فتصاحبه ≠ أو < أو > ، ولا يصاحبه أبداً علامة «التساوي» ، ويمكن النظر إلى الفرض البديل على خطوتين تبدأ بسؤال .

هل من المناسب إستخدام ≠ بدلاً من < أو > ؟

١ – إذا كان غير مناسباً ، دائماً إستخدام < أو > .

٢ – يتم إختيار ≠ حين تكون درجة الإختلاف عن التساوي تؤدي إلى إتخاذ قرارات مختلفة .

٢/٣ اختبار الفرق بين متوسطي مجتمعين : العينات المستقلة

عادة ما يرغب متخذي القرارات مقارنة متوسطي مجتمعين - مثلاً - مقارنة هل متوسط الوقت الذي يقضيه المشاهدون «الذكور» لمشاهدة القناة الإخبارية هو نفس متوسط الوقت الذي يقضيه المشاهدات «الإناث» لنفس القناة الإخبارية ؟ هل متوسط إنفاق الرجال على الملابس أقل من إنفاق السيدات ؟ هل متوسط الأعمار في الجامعات هو نفس متوسط الأعمار في المعاهد ؟ هل متوسط سن الزواج في الريف أقل من متوسط سن الزواج في الحضر ؟ هل متوسط قيم المبيعات في شهر رمضان أعلى من متوسط قيم المبيعات في شهر جمادى الأولى؟

وهكذا ، في أي موقف يرغب متخذ القرار المقارنة بين متوسطي مجتمعين ، يقوم بعملية إختبار للفرق بين متوسطي المجتمعين .

ويعتمد الإختبار الإحصائي الذي يستخدم لإختبار الفرق بين متوسط مجتمعين مستقلين على :

١ - حجم العينتين .

٢ - هل المجتمعان يتبعان التوزيع الطبيعي أم لا ؟

٣ - هل الإنحراف المعياري للمجتمعين معروف أم لا ؟

٤ - هل الإنحراف المعياري للمجتمعين متساويان أم لا ؟

واختيار الإختبار الإحصائي معطى في جدول (٣ - ١) . وفي حالة إذا كان المجتمعان لا يتبعا التوزيع الطبيعي ، والإنحراف المعياري غير معلوم ، وحجم

العينة ليس كبيراً ، ولا يمكن من إستخدام نظرية الحد المركزى ، يجب إستخدام بعض الأساليب والطرق اللامعلمية .

جدول (٣-١) : الإختيار بين إختبار ى وإختبار ت لإختبار الفرق بين متوسطى مجتمعين مستقلين

المجتمع	حجم العينة	σ_1 ، σ_2 معلومتان	σ_1 ، σ_2 غير معلومتان
موزع توزيعاً طبيعياً	كبير	ى	ى
	صغير	ى	ت
غير موزع توزيعاً طبيعياً	كبير	ى	ى
	صغير	طرق لا معلمية	طرق لا معلمية

ويتبع الإختبار الست خطوات السابق عرضهم . ولكن تعتمد الخطوات الست على أن :

١ - الظاهرة محل الدراسة فى المجتمعين تتبع التوزيع الطبيعى ، أو يمكن أن تقرب بالتوزيع الطبيعى .

٢ - تباين الظاهرة فى المجتمعين متساوى أى أن : $\sigma_1^2 = \sigma_2^2$ ، وهذا الافتراض يسمى إفتراض " تجانس التباين Homogeniety of Variance " .

وفى حالة عدم تحقق هذا الشرط ، (شرط تجانس التباين) يمكن إستخدام إختبار ت ، ولكن يجب تعديل درجات حرية الإختبار ، ولن نتعرض لهذه الحالة فى هذا الكتاب . وسوف نحدد الإختبار فى حالة ما إذا كان شرط تجانس التباين محقق .

وكما تم عند عرض إنشاء فترات الثقة للفرق بين متوسطين ، فإن تحقق شرط تجانس التباين ، يُمكن الباحث من حساب التباين المجمع ($ع_م^2$) ، وهذا التباين المجمع ، هو المقدر غير المتحيز ، والمتسق ، والكفء لتباين المجتمع الأول وتباين المجتمع الثاني . وهذا التباين المجمع هو التباين الذي يستخدم لحساب الخطأ المعياري للفرق بين المتوسطين . فيعتمد إختبار ت ، لإختبار الفرق بين وسطين على نفس الإفتراض ، وهو إفتراض تجانس التباين .

وتكون خطوات إختبار ت لإختبار الفرق بين متوسطي مجتمعين مستقلين كالآتي :

١ - صياغة الفروض :

ف. : $\mu_1 = \mu_2$ أو $(\mu_1 - \mu_2) =$ صفر

$\mu_1 \leq \mu_2$ أو $(\mu_1 - \mu_2) \leq$ صفر

$\mu_1 \geq \mu_2$ أو $(\mu_1 - \mu_2) \geq$ صفر

ف١ : هو المكمل للفروض العدمية أعلاه .

٢ - يتم سحب عينتين مستقلتين احجامهما $ن_1$ ، $ن_2$ من مجتمعين مستقلين .

ويتم حساب : $\bar{س}_1$ ، $\bar{س}_2$ ، $ع_1$ ، $ع_2$ حيث :

$\bar{س}_1$ ، $\bar{س}_2$ هما الوسطان الحسابيان من العينة الأولى والعينة الثانية على التوالي ، $ع_1$ ، $ع_2$ هما الإنحراف المعياري المحسوبان من بيانات العينة الأولى وبيانات العينة الثانية على التوالي .

٣ - تحدد الإختبار الإحصائي المناسب ، ويحدد مستوى المعنوية (∝) والمنطقة الحرجة . فإذا كان الإختبار في إتجاهين يكون هناك قيمتين حرجتين عما ت بدرجات حرية ($ن_1 + ن_2 - 2$) التي تجعل مساحة الذيل الأيمن مساوياً ∝/ ٢ ، والتي تجعل مساحة الذيل الأيسر مساوياً ∝/ ٢ .

وإذا كان الإختبار في إتجاه واحد ، ناحية اليمين ، تكون القيمة الحرجة هي قيمة ت بدرجات حرية ($ن_1+ن_2-2$) التي تجعل مساحة الذيل الأيمن = ∝ .

وإذا كان الإختبار في إتجاه واحد ، ناحية اليسار ، تكون القيمة الحرجة هي قيمة ت بدرجات حرية ($ن_1+ن_2-2$) التي تجعل مساحة الذيل الأيسر = ∝

٤ - حساب الخطأ المعياري للفرق بين متوسطي عينتين ويوجد هنا حالتان .

(أ) الخطأ المعياري للفرق بين متوسطي عينتين : σ_1^2 ، σ_2^2 غير معلومتان .

$$ع_{\bar{س}_1 - \bar{س}_2} = \sqrt{ع_م^2 \left(\frac{1}{ن_1} + \frac{1}{ن_2} \right)}$$

حيث :

$$ع_م^2 = \frac{(ن_1 - 1) ع_1^2 + (ن_2 - 1) ع_2^2}{ن_1 + ن_2 - 2}$$

(ب) الخطأ المعياري للفرق بين متوسطي عينتين : σ_1^2 ، σ_2^2 معلومتان .

$$\sigma_{\bar{س}_1 - \bar{س}_2} = \sqrt{\left(\frac{\sigma_1^2}{ن_1} + \frac{\sigma_2^2}{ن_2} \right)}$$

٥ – إحسب قيمة الإختبار الإحصائي ، ويعتمد الإختبار الإحصائي على σ_1 ، σ_2 كالتالي :

(أ) إذا كان σ_1 ، σ_2 غير معلومتان ويفترض تساويهما يكون الإختبار الإحصائي :

$$ت = \frac{(\bar{س}_1 - \bar{س}_2) - (\mu_1 - \mu_2)}{ع_{\bar{س}_1 - \bar{س}_2}}$$

حيث : $(\mu_1 - \mu_2)$ ، هو قيمة الفرق تحت الفرض العدمي ، في معظم الأحيان يكون مساويًا للصفر ، وفي بعض الأحيان يختلف عن الصفر .

(ب) إذا كان σ_1 ، σ_2 معلومتان ويفترض تساويهما ، يكون الإختبار الإحصائي :

$$ى = \frac{(\bar{س}_1 - \bar{س}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{س}_1 - \bar{س}_2}}$$

وتحت شرط تجانس التباين (التساوي) في المجتمعين : $\sigma_1^2 = \sigma_2^2$ يكون :

$$\boxed{\sigma_{\bar{س}_1 - \bar{س}_2} = \sqrt{\left(\frac{\sigma_1^2}{ن_1} + \frac{\sigma_2^2}{ن_2}\right)}}$$

٦ – إتخاذ قرار بشأن صحة الفرض العدمي ، كالآتي :

إتجاه ف$_1$	يرفض الفرض العدمي إذا كانت :
$\neq$	القيمة المحسوبة $\lvert ت \rvert >$ القيمة الجدولية $\lvert ت_{\alpha/2} \rvert$
$>$	القيمة المحسوبة ت $>$ القيمة الجدولية $ت_{\alpha}$
$<$	القيمة المحسوبة ت $<$ القيمة الجدولية $- ت_{\alpha}$

حيث ت بدرجات حرية ($ن_١ + ن_٢ - ٢$) .

ولاحظ أوجه التشابه بين إختبار ت للفرق بين متوسطين وإختبار ت لاختبار متوسط واحد .

متوسطين	متوسط واحد
$ت = \dfrac{(\bar{س}_١ - \bar{س}_٢) - (\mu_١ - \mu_٢)_٠}{خ_{\bar{س}_١ - \bar{س}_٢}}$	$ت = \dfrac{\bar{س} - \mu_٠}{خ_{\bar{س}}}$
درجات الحرية = $(ن_١ - ١) + (ن_٢ - ١)$	درجات حرية = $(ن - ١)$

وفى الإختبارين يقيم الفرق بين نتائج العينة والقيم المفترضة كنسبة من الخطأ المعيارى المناسب .

تطبيق(٣-١)

اراد مدير إنتاج فى إحدى مصانع السيارات معرفة أى نوع من الإطارات يستخدم : هل يستخدم النوع العادى ؟ أم يستخدم النوع المميز ؟ وقرر إذا كانت فرق الكيلو مترات التى تتحملها الإطارات يزيد على ١٢,٥٠٠ الف كم فإنه يفضل النوع المميز ، أما إذا كان هذا الفرق أقل من ١٢,٥٠٠ الف كم فسوف يستخدم النوع العادى . وحيث أن مدير الإنتاج يود أن يكون النوع المميز هو المفضل ، وضع فروضه كالأتى :

$ف_٠ : (\mu_١ - \mu_٢)_٠ \geq ١٢,٥٠٠$ ألف كم

$ف_١ : (\mu_١ - \mu_٢)_٠ < ١٢,٥٠٠$ ألف كم

حيث μ_1 متوسط سريان الإطارات المميزة في مجتمع الإطارات المميزة .

μ_2 متوسط سريان الإطارات العادية في مجتمع الإطارات العادية .

٢ - قام باحث في هذا المصنع ، بسحب عينة عشوائية من الإطارات المميزة حجمها ١٦ إطاراً ، وتم إستخدامها حتى إهلاكها . وتم سحب عينة عشوائية من الإطارات العادية حجمها ٢٠ إطاراً وتم إستخدامها حتى إهلاكها . وكانت النتائج كالتالي :

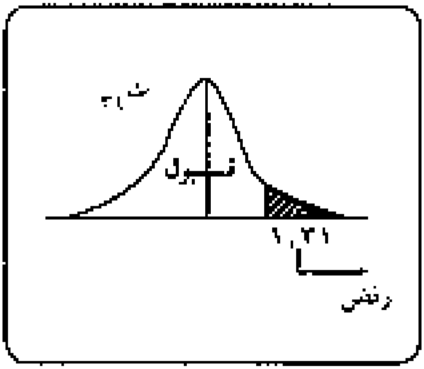
	الإطارات العادية	الإطارات المميزة
الوسط الحسابي	٣١,٤٨٤	٤٤,٠٧٤
التبايـــــــن	٤٣٨٤	٦٠٧١
ن	٢٠	١٦

وهنا لاحظ الباحث بوجود فرق بين الوسط الحسابي للإطارات المميزة والوسط الحسابي للإطارات العادية ، وهذا الفرق يساوي (٤٤,٠٧٤ - ٣١,٤٨٤) = ١٢,٥٩٠ ألف كم بين النوعين . حيث أن هذا الفرق أكبر من القيمة تحت الفرض العدمي ، ويقع في إتجاه الفرق البديل . نستمر في الإختبار لمعرفة هل هذا الفرق (١٢٥٩٠ - ١٢٥٠٠) = ٩٠ كم معنوياً أم غير معنوياً .

٣ - بفرض ان توزيعا الإطارات في المجتمعين يتبعا التوزيع الطبيعي ، وحيث أن حجما العينتان صغيرا (ن١ < ٣٠ ، ن٢ < ٣٠) . يكون الإختبار الإحصائي هو إختبار ت .

وبمستوى معنوية ١٠ ٪ ، ودرجات حرية = (ن١ + ن٢ - ٢) = (٢٠ + ١٦ - ٢) = ٣٤ نجد أن القيمة الحرجة هي ت٣٤ = ١,٣١

٤ - تحسب قيمة الإختبار الإحصائي :



$$\text{ت} = \frac{(\bar{\text{س}}_1 - \bar{\text{س}}_2) - (\mu_1 - \mu_2)}{\text{ع}_{\bar{\text{س}}_1 - \bar{\text{س}}_2}}$$

$$\text{ع}_{\bar{\text{س}}_1 - \bar{\text{س}}_2} = \sqrt{\text{ع}^2_{\text{م}} \left(\frac{1}{\text{ن}_1} + \frac{1}{\text{ن}_2} \right)}$$

$$\text{ع}^2_{\text{م}} = \frac{\text{ع}^2_1 (\text{ن}_1 - 1) + \text{ع}^2_2 (\text{ن}_2 - 1)}{\text{ن}_1 + \text{ن}_2 - 2}$$

$$= \frac{٣٢٤٨ \times (٢٠ - ١) + ٦٠٧١ \times (١٦ - ١)}{١٦ + ٢٠ - ٢} = ٥١٠٨$$

$$\therefore \text{ع}_{\bar{\text{س}}_1 - \bar{\text{س}}_2} = \sqrt{٥١٠٨ \left(\frac{١}{١٦} + \frac{١}{٢٠} \right)} = ٢٤,٠٠ \text{ كم}$$

$$\therefore \text{ت} = \frac{(٤١,٧٤ - ٣١,٤٨٤) - ١٢٠}{٢٤} = ٣,٧٥$$

وحيث أن قيمة ت المحسوبة تقع داخل منطقة الرفض .

∴ نفرض الفرض العدمي ، وننصح مدير الإنتاج بالتحول إلى النوع المميز حيث أن فرق الكيلومترات تزيد على ١٢,٥٠٠ ألف كم .

الفرق الحرج لإختبار في إتجاهين

يحدد الفرق الحرج الذي يحدد منطقة قبول الفرض العدمي والذي إذا كان الفرق $(\bar{س}_1 - \bar{س}_2)$ خارج مساحة ، يرفض الفرض العدمي إذا كان $(\bar{س}_1 - \bar{س}_2)$ خارج الفرق الحرج ، حيث الفرق الحرج في حالة الإختبار في إتجاهين :

الفرق الحرج = $(\mu_1 - \mu_2) \pm ت_{\alpha/2} \times ع_{\bar{س}_1 - \bar{س}_2}$

ت بدرجات حرية = $(ن_1 + ن_2 - ٢)$ مستوى معنوية = α %

الفرق الحرج لإختبار في إتجاه واحد (ناحية اليمين) يعطى بالمعادلة :

الفرق الحرج = $(\mu_1 - \mu_2) + ت_{\alpha} \times ع_{\bar{س}_1 - \bar{س}_2}$

ت بدرجات حرية = $(ن_1 + ن_2 - ٢)$ ، ومستوى معنوية = α %

إذا كان الفرق بين $(\bar{س}_1 - \bar{س}_2)$ أكبر من الفرق الحرج ، يرفض الفرض العدمي :

والفرق الحرج لاختبار في إتجاه واحد (ناحية اليسار) يعطى بالمعادلة :

$$\text{الفرق الحرج} = (\mu_1 - \mu_2)_0 - \text{ت}_{\nu} \times \text{ع}_{\bar{\text{س}}_1 - \bar{\text{س}}_2}$$

ت بدرجات حرية = ($ن_1 + ن_2 - ٢$) ، ومستوى معنوية = α %

إذا كان الفرق بين ($\bar{\text{س}}_1 - \bar{\text{س}}_2$) اقل من الفرق الحرج يرفض الفرض العدمي بالمعادلة :

في تطبيق (٣-١) ، نجد أن :

$$\text{الفرق الحرج} = (\mu_1 - \mu_2)_0 + \text{ت}_{\nu} \times \text{ع}_{\bar{\text{س}}_1 - \bar{\text{س}}_2}$$

∴ الفرق الحرج = ١٢,٥٠٠ + ١,٣١ × ٢٤ = ١٢,٥٣١ ألف كم

اى أنه إذا زاد الفرق بين ($\bar{\text{س}}_1 - \bar{\text{س}}_2$) عن ١٢,٥٣١ ألف كم ، نرفض الفرض العدمي .

وحيث أن $\bar{\text{س}}_1 - \bar{\text{س}}_2$ = ٤٤,٠٧٤ − ٣١,٤٨٤ = ١٢,٥٩٠

∴ $\bar{\text{س}}_1 - \bar{\text{س}}_2$ > ١٢,٥٣١

∴ القرار : رفض الفرض العدمي .

تطبيق(٣-٢)

سحبت عينة عشوائية حجمها ١٤ علبة من علب المشروبات الغازية الدايت (ذات السعرات الحرارية المنخفضة) من إنتاج الشركة (أ) ، ووجد أن متوسط عدد السعرات ٢٣ سعراً حرارياً ، بإنحراف معياري ٣ سعر حراري .

وسحبت عينة عشوائية حجمها ١٦ علبة من علب المشروبات الغازية الدايت من إنتاج الشركة (ب) ، ووجد أن متوسط السعرات الحرارية ٢٥ سعراً بإنحراف معيارى ٣ سعر حرارى .

باستخدام α = ١ ٪ ، هل يختلف متوسط السعرات الحرارية فى الشركتين؟ بفرض أن توزيع السعرات الحرارية يتبع التوزيع الطبيعى فى الشركتين ، وبإفتراض أن تباين السعرات الحرارية فى الشركتين متساوى .

١ - صياغة الفروض :

$$ف_. \quad (\mu_1 - \mu_2)_. \quad = \quad صفراً$$

$$ف_1 \quad (\mu_1 - \mu_2)_. \quad \neq \quad صفراً$$

حيث μ_1 هو متوسط السعرات الحرارية فى إنتاج المشروبات الغازية الدايت فى الشركة (أ) ، μ_2 هو متوسط السعرات الحرارية فى إنتاج المشروبات الغازية الدايت فى الشركة (ب) .

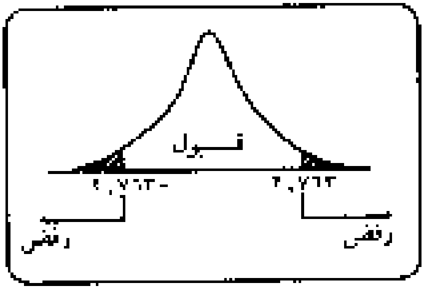
٢ - وصل الباحث إلى البيانات الآتية :

$ن_1 = ١٤$ $\quad \bar{س}_1 = ٢٣$ $\quad ع_1 = ٣$

$ن_2 = ١٦$ $\quad \bar{س}_2 = ٢٥$ $\quad ع_2 = ٤$

٣ - حيث أن حجم العينتين صغيرتين ($ن_1$ < ٣٠ ، $ن_2$ < ٣٠) ، ويفترض تجانس تباين الشركتين .

∴ الإختبار الإحصائى هو إختبار ت بدرجات حرية ($ن_1 + ن_2 - ٢$) = (١٤ + ١٦ - ٢) = ٢٨ ، وعند α = ١ ٪ ، والإختبار فى إتجاهين ، نجد أن ت الحرجة = ± ٢,٧٦٣ .



٤ - نحسب قيمة الإختبار الإحصائي :

$$\text{ت} = \frac{(\bar{\text{س}}_1 - \bar{\text{س}}_2) - (\mu_1 - \mu_2)}{\text{ع}_{\bar{\text{س}}_1 - \bar{\text{س}}_2}}$$

$$\text{ع}_{\bar{\text{س}}_1 - \bar{\text{س}}_2} = \sqrt{\text{ع}_{\text{ف}}^2 \left(\frac{1}{\text{ن}_1} + \frac{1}{\text{ن}_2} \right)}$$

$$\text{ع}_{\text{ف}}^2 = \frac{(\text{ن}_1 - 1)\,\text{ع}_1^2 + (\text{ن}_2 - 1)\,\text{ع}_2^2}{\text{ن}_1 + \text{ن}_2 - 2}$$

$$= \frac{(14 - 1)(13)^2 + (16 - 1)(4)^2}{14 + 16 - 2} = 12{,}7949$$

$$\therefore \quad \text{ع}_{\bar{\text{س}}_1 - \bar{\text{س}}_2} = \sqrt{12{,}7949 \left(\frac{1}{14} + \frac{1}{16} \right)} = 1{,}3067$$

$$\therefore \quad \text{ت} = \frac{(23 - 25) - \text{صفر}}{1{,}3067} = -1{,}531$$

حيث أن قيمة ت المحسوبة (- ١,٥٣١) تقع داخل منطقة القبول (- ٢,٧٦٣ إلى ٢,٧٦٣٥) .

∴ لا نرفض الفرض العدمي ، ونصل إلى قرار بأنه لا يوجد إختلاف في متوسط السعرات الحرارية في إنتاج الشركتين .

الفرق الحرج لإختبار في إتجاهين

لمعرفة الفرق الحرج ، نجد أنه يوجد منطقتا رفض ، باستخدام توزيع ت ، وجد أن منطقتا الرفض هي المنطقة ≤ ٢,٧٦٣ ، والمنطقة ≥ - ٢,٧٦٣ .

∴ يوجد فرق حرج باستخدام توزيع ($\bar{س}_1 - \bar{س}_2$) يقابل القيمة ٢,٧٦٣ ، وفرق حرج يقابل القيمة - ٢,٧٦٣ .

$$\text{الفرق الحرج} = (\mu_1 - \mu_2) \pm ت \times ع_{\bar{س}_1 - \bar{س}_2}$$

∴ الفـرق الحـــرج الأيمـن = (صفر) + ٢,٧٦٣ × ١,٣٠٦٧

= ٣,٦١ سعراً حرارياً

الفـرق الحـــرج الأيســـر = (صفر) - ٢,٧٦٣ × ١,٣٠٦٧

= - ٣,٦١ سعراً حرارياً

إذا كان الفرق المطلق | $\bar{س}_1 - \bar{س}_2$ | أكبر من | ٣,٦١ | نرفض الفرض العدمي .

وحيث أن الفرق المطلق هو | ($\bar{س}_1 - \bar{س}_2$) | = | (٢٣ - ٢٥) | = - ٢ سعر حراري وهو أقل من الفرق الحرج المطلق (٣,٦١) .

∴ لا نرفض الفرض العدمي ، ونصل إلى قرار بأنه لا يوجد إختلاف في المتوسط في إنتاج الشركتين من المشروبات الغازية الدايت .

تماريـــــن

١ – أعطيت بيانات عينتين مستقلتين كالآتي :

$ن_١ = ٢٢$ $\qquad \bar{س}_١ = ١٢,٥$ $\qquad ع_١ = ٣,٧٥$

$ن_٢ = ١٨$ $\qquad \bar{س}_٢ = ١٤,٦$ $\qquad ع_٢ = ٣,١٠$

(أ) ما هو مقدار ($\mu_٢ - \mu_١$) ؟

(ب) إنشيء فترة ثقة ٩٥ ٪ للفرق ($\mu_٢ - \mu_١$) .

(جـ) إختبر :

$ف_٠$: ($\mu_١ - \mu_٢$) = صفر ضد $ف_١$: ($\mu_١ - \mu_٢$) $\neq$ صفر .

٢ – إنشيء فترة ثقة ٩٥ ٪ للفرق بين ($\mu_١ - \mu_٢$) ، واختبر صحة الفرض العدمي : $ف_٠$: ($\mu_١ - \mu_٢$)$_٠ \leq ٣$ مستخدمًا البيانات الآتية :

$ن_١ = ٢٠$ $\qquad \bar{س}_١ = ٣٣,٧٥$ $\qquad ع_١ = ٥,٢٥$

$ن_٢ = ٢٣$ $\qquad \bar{س}_٢ = ٢٨,٥٠$ $\qquad ع_٢ = ٤,٥٥$

$\alpha = ٥ ٪$

٣ – تم الحصول على البيانات الآتية من عينتين مستقلتين ، سحبا بطريقة عشوائية من مجتمعين يتبعا التوزيع الطبيعي ، بتباين متساوي :

العينة الأولى : ٢٧ ٣١ ٢٥ ٣٣ ٢١ ٣٥ ٢٨

٣٠ ٢٦ ٢٥ ٣١ ٣٣ ٣٠

العينة الثانية : ٢٤ ٢٨ ٢٢ ٢٥ ٢٤ ٢٢

٢٩ ٢٦ ٢٥ ٢٨ ١٩ ٢٩

(أ) ما هو مقدار (μ_1 − μ_2) ؟

(ب) إنشيء فترة ثقة ٩٩ ٪ لتقدير الفرق بين (μ_1 − μ_2) .

(جـ) إختبر الفرض العدمي :

ف. : $\mu_1 \leq \mu_2$ ، ؛ α = ١ ٪

(٤) إعتقد مدير مبيعات أحد فروع المحلات التجارية الكبرى أن إنفاق السيدات أكثر من إنفاق الرجال على الملابس .

قام بسحب عينة عشوائية من العملاء الرجال ، حجمها ٢٥ رجلاً ، ووجد أن متوسط إنفاقهم الربع سنوى على الملابس ٧٣ جنيهًا بإنحراف معيارى ١٧,٥ جنيهًا . وسحب عينة عشوائية من العملاء السيدات ، حجمها ٢٠ سيدة ، ووجد أن متوسط الإنفاق الربع سنوي على الملابس ٨٧ جنيهًا ، بإنحراف معياري ١٤,٤ جنيهًا .

المطلوب :

(أ) قدر الفرق فى الإنفاق الربع سنوى ، بإنشاء فترة ثقة ٩٥ ٪ .

(ب) ما هو الفرق الحرج بين إنفاق السيدات وإنفاق الرجال إستخدام α = ٥ ٪ .

(جـ) هل تتفق مع مدير المبيعات بأن فرق الإنفاق بين السيدات والرجال يزيد على خمسة جنيهات .

(٥) إدعت شركة أدوية (أ) أن منتجها يزيل الآلام بسرعة أكبر من منتج شركة آخر . قام باحث باختبار هذا الإدعاء . فقام بإعطاء العقار (أ) من إنتاج الشركة الأولى إلى ٢٥ مريضاً وقام بقياس سرعة زوال الألم بالدقيقة . ووجد أن متوسط الزمن لزوال الألم ٤٤ دقيقة بإنحراف معياري ١٣ دقيقة .

وقام باختيار ٢٢ مريضاً آخر ، وأعطى لهم العقار (ب) من إنتاج شركة أخرى ، ووجد أن متوسط الزمن لزوال الألم ٤٩ دقيقة بإنحراف معياري ١١ دقيقة .

(أ) إنشيء فترة ثقة للفرق بين ($\mu_1 - \mu_2$) . α = ٥ ٪

(ب) هل تتفق مع إدعاء الشركة (أ) ؟ α = ٥ ٪ .

(جـ) ما هو الفرق الحرج اللازم لرفض الفرض العدمي ؟ α = ٥ ٪ .

٣/٣ إختبار الفرق بين متوسطي مجتمعين : عينتين غير مستقلتين

بفرض أن قسم الأبحاث في إحدى شركات البترول ، قد توصل إلى سائل يضاف إلى تانك وقود السيارة ومن شأنه أن يزيد من كفاءة الموتور .

أراد قسم الأبحاث إختبار الفرض اعلاه . فقام باختيار ٣٥ سيارة وكان البنزين فيها ٩٠ عادى ، وجربت السيارات لعدد ٢٠٠٠ كم . وقيس عدد الكيلو مترات المعطاة لكل صفيحة بنزين ، ثم تم إضافة السائل إلى تانك الوقود قبل تموينه بالبنزين ٩٠ . وقيس عدد الكيلومترات المعطاة لكل صفيحة بنزين .

المثال أعلاه هو مثال لما يسمى «قبل - بعد» التجربة ، أو ما يسمى « الأزواج المتوافقة Matched Pair » . في هذا المثال مفردات التجربة «بعد» غير مستقل عن مفردات التجربة «قبل» ، لأن نفس العبارات قد قيست «قبل» إجراء التجربة ، «بعد» إجراء التجربة .

وتنشأ «عينة الأزواج المتوافقة» إذن في حالتين :

١ - تختار نفس المفردات لـ «قبل Pre» و لـ «بعد Post» التجربة .

٢ - أن تكون أزواج القيم من مجتمعين ، ولكن أزواج القيم مترابطة - مثلاً إختيار عينة عشوائية من الأزواج ، والتعرف على عمر الزوج ثم إختيار زوجاتهم ، ومعرفة أعمارهن ، ومقارنة متوسط عمر الزوج بمتوسط عمر الزوجة .

وبالتالي ، يمكن النظر إلى «إختبار الأزواج المتوافقة» ، كإختبار «العلاقة» بين متغيرين ، مثلاً ، العلاقة بين سن الزوج وسن الزوجة ، أو هل الفرق بين عمر الزوج وعمر الزوجة يشكل فرقاً معنوياً أم لا .

ويمكن النظر إلى إختبار «قبل - بعد» كإختبار لدراسة «الأثر» أي أثر إجراء تجربة معينة ، وهل هذا الأثر أو هذه الإضافة معنوية أو غير معنوية .

خطوات الإختبار

سوف نشير إلى فرق متوسطي المجتمعين بالرمز ف ، حيث المدون الأسفل «ف» يشير إلى الفرق . ونشير إلى الفرق بين متوسطي العينتين بالرمز ف̄ ، حيث :

$$\bar{ف} = \frac{مجـ ف}{ن} \quad , \quad ف = (س_1 - س_2)$$

وكامتداد لنظرية الحد المركزي ، فإن توزيع معاينة الوسط الحسابي للفرق ($\bar{ف}$) يتبع التوزيع الطبيعي بمتوسط $\mu_ف$ وإنحراف معياري $ع_{\bar{ف}}$ ، حيث :

$$ع_{\bar{ف}} = \frac{ع_ف}{\sqrt{ن}} \qquad \text{مجتمع غير محدود}$$

$$ع_ف = \frac{مجـ (ف - \bar{ف})^2}{ن - 1}$$

وحيث أن الإنحراف المعياري في المجتمع $\sigma_ف$ لا يكون معروفاً ، نجد أن الإختبار الإحصائي لإختبار الفرض العدمي هو اختبار ت بدرجات حرية (ن - ١) .

وخطوات إختبار الفرق بين متوسطي مجتمعين غير مستقلين :

١ - صياغة الفروض :

$ف_0 : (\mu_1 - \mu_2) = \mu_ف =$ صفر

$ف_1 : \quad \mu_ف \neq$ صفر

ويمكن أن يصاحب الفرض البديل علامة " > " أو علامة " < " إختبار في إتجاه واحد .

٢ - سحب عينتين غير مستقلتين - إحسب الفرق بين أزواج القيم ($س_١$ - $س_٢$) ، وإحسب متوسط الفرق : $\bar{ف}$ ، $ع_ف$ ، وقرر عما إذا كنت ستستمر في الإختبار أم لا .

ولا تستمر في الإختبار إذا كان $\bar{ف}$ تقع في نطاق الفرض العدمي ، أما إذا كان ف خارج نطاق الفرض العدمي ، فتستمر في الإختبار .

٣ - حدد الإختبار الإحصائي ت ، مستوى المعنوية ، والمنطقة الحرجة للإختبار الإحصائي حيث درجات الحرية = (ن - ١) .

٤ - إحسب الخطأ المعياري لمتوسط الفرق :

$$ع_{\bar{ف}} = \frac{ع_ف}{\sqrt{ن}}$$

٥ - إحسب الإختبار الإحصائي :

$$ت = \frac{\bar{ف} - م_ف}{ع_{\bar{ف}}}$$

حيث $م_ف$ هو متوسط الفرق تحت الفرض العدمي .

٦ - الوصول إلى قرار بشأن رفض الفرض العدمي أو عدم رفضه كالآتي :

إتجاه $ف_١$	رفض الفرض العدمي إذا كنت :
$\neq$	القيمة المحسوبة | ت | > | $ت_{ح/٢}$ | الجدولية
>	القيمة المحسوبة ت > + $ت_{ح}$ الجدولية
<	القيمة المحسوبة ت < - $ت_{ح}$ الجدولية

ولاحظ أوجه التشابه بين إختبار الفرق بين متوسطي عينتين غير مستقلتين، واختبار الوسط الحسابي من مجتمع واحد .

متوسط الفرق	وسط حسابي واحد
$\text{ت} = \dfrac{\bar{\text{ف}} - \mu_{\text{ف}}}{\text{ع}_{\bar{\text{ف}}}}$	$\text{ت} = \dfrac{\bar{\text{س}} - \mu}{\text{ع}_{\bar{\text{س}}}}$
درجات الحرية = (ن - ١)	درجات الحرية = (ن - ١)

والإختباران يتبعا توزيع ت الإحتمالي بدرجات حرية (ن-١)، والمنطقة الحرجة واحدة بالنسبة للاختبارين عند نفس مستوى المعنوية α ، وفي نفس الإتجاه .

الفرق الحرج

للوصول إلى قيمة الفرق بين ($\text{س}_1 - \text{س}_2$) حيث يرفض الفرض العدمي إذا كان الفرق المحسوب أكبر منه أو أصغر منه ، يحسب الفرق الحرج كالآتي .

إختبار في إتجاهين

$$\text{الفرق الحرج} = \mu_{\text{ف}} \pm \text{ت}_{\alpha/2} \times \text{ع}_{\bar{\text{ف}}}$$

إختبار في إتجاه واحد (يمين)

$$\text{الفرق الحرج} = \mu_{\text{ف}} + \text{ت}_{\alpha} \times \text{ع}_{\bar{\text{ف}}}$$

إختبار في إتجاه واحد (يسار)

$$\text{الفرق الحرج} = \mu_{\text{ف}} - \text{ت}_{\alpha} \times \text{ع}_{\bar{\text{ف}}}$$

حيث $\mu_{\text{ف}}$ هو الفرق تحت صحة الفرض العدمي

تطبيق (٣-٣)

أراد مدير إدارة الأفراد تطبيق سياسة العمل في ساعات غير ثابتة (مرنة) . وقبل التطبيق لهذه السياسة ، أراد أن يتعرف هل هذه السياسة الجديدة أدت إلى غياب أقل أم لا .

١ – وافق مدير الشركة على هذه السياسة ، ورأى تطبيقها إذا أدت إلى تقليل عدد أيام الغياب بيومين على الأقل . لذا قام الباحث بصياغة الفروض كالآتي :

ف. : م ف ≤ ٢ ؛ ف١ : م ف > ٢

حيث ف : تمثل الفرق بين عدد أيام الغياب في السياسة الجديدة وعدد أيام الغياب في السياسة الحالية .

٢ – قام الباحث باختيار ٦ من المحاسبين الذين يؤيدوا سياسة الساعات المرنة . وتم مراجعة أيام غيابهم في ثلاثة أشهر قبل تطبيق السياسة المرنة عليهم . وتم مراجعة أيام الغياب في ثلاثة أشهر بعد تطبيق السياسة المرنة للعمل . ووصل إلى البيانات الآتية :

رقم المحاسب	عدد أيام الغياب		ف
	السياسة الحالية	السياسة الجديدة	
١	١٤	١٠	٤ –
٢	٩	٦	٣ –
٣	١١	٧	٤ –
٤	٥	٠	٥ –
٥	٨	٢	٦ –
٦	٩	٣	٦ –
			٢٨ –

$$\therefore \quad \bar{ف} = \frac{-٢٨}{٦} = -٤,٧$$

حيث أن ف ـ = ـ ٤,٧ يوماً ، أى أن متوسط الفرق يقع في إتجاه الفرض البديل ، إذن نستمر في خطوات الإختبار .

٣ - بفرض أن توزيع الفروق يتبع أو يمكن أن يقرب بالتوزيع الطبيعي ، وحجم العينة صغير ، والإنحراف المعياري للفروق في المجتمع غير معروف .

∴ الإختبار الإحصائي هو : إختبار ت . بدرجات حرية = (ن - ١) = (٦ - ١) = ٥ ، ∝ = ٥ ٪ القيمة الحرجية = - ت ٠,٠٥ = - ٢,٠١٥ .

٤ - الخطأ المعيارى لمتوسط الفروق :

$$ع_{\bar{ف}} = \frac{ع_ف}{\sqrt{ن}} \quad \cdot \quad \sqrt{\frac{ع_ف^{٢}}{ن}}$$

$$ع_ف^{٢} = \frac{مجـ (ف - \bar{ف})^{٢}}{ن - ١} = \frac{ن \, مجـ ف^{٢} - (مجـ ف)^{٢}}{ن (ن - ١)}$$

$$= \frac{٦ \times ١٣٨ - (٢٨)^{٢}}{٦ \times (٦ - ١)} = ١,٤٦٦$$

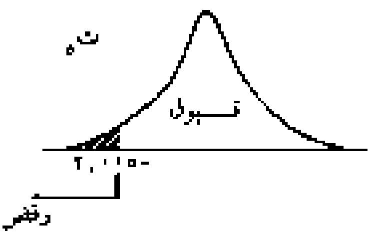
حيث :

مج ف^{٢} = ٤^{٢} + ٣^{٢} + ٤^{٢} + ٥^{٢} + ٦^{٢} + ٦^{٢} = ١٣٨

مج ف = ٢٨

∴ $ع_{ف} = \sqrt{١,٤٦٦} = ١,٢١$

∴ $ع_{\bar{ف}} = \frac{١,٢١}{\sqrt{٦}} = ٠,٤٩$



٥ - نحسب قيمة الإختبار الإحصائي :

$$ت = \frac{\bar{ف} - الفرق}{ع_{\bar{ف}}} = \frac{٢ - ٤,٧}{٠,٤٩} = -١٣,٦٧$$

حيث أن القيمة المحسوبة ت = (-١٣,٦٧) > القيمة الجدولية (-٢,٠١٥)

∴ نرفض الفرض العدمي ، ونستطيع القول أن الفرق في عدد أيام الغياب أقل من يومين ، وأن السياسة الجديدة المرنة تؤدي إلى عدد أيام غياب أقل بيومين على الأقل .

الفرق الحرج

لمعرفة عدد أيام الغياب الحرجة التي بمقارنتها بالفرق الحقيقي نستطيع قبول أو رفض الفرض العدمي :

الفرق الحرج = لـف – تـ × عـف = ٢ – ٢,٦٥ × ٠,٤٩ = ٠,٧٠ يوماً

∴ نقارن الفرق الحرج بالفرق الفعلي (–٤,٧) حيث أن الفرق الفعلي أقل من الفرق الحرج

∴ نرفض الفرض العدمي . ونصل إلى نفس النتيجة أعلاه .

تماريـــــن

٦ – أدعى أحد محاسبي الضرائب (أ) ، أن طريقة حساباته تقلل من الضرائب المستحقة بـ ٣٠٠ جنيه ، تم سحب عينة عشوائية للتأكد من صحة هذا الإدعاء، وطلب من محاسب آخر (ب) تقدير الضرائب المستحقة على كل مفردة من مفردات العينة ، ومقارنتها بتقدير المحاسب (أ). وكانت النتائج كالآتي :

العميل	المحاسب (ب)	المحاسب (أ)
١	١٤٧٨٣	١٤٣١٧
٢	٤٢١٨	٤١٧٣
٣	٩٠٢٦	٨٨٣٧
٤	٦٣٨٢	٦١١٧
٥	١٣٨٣١	١٢ ٢٠
٦	٧١٣٤	٦٢١٦
٧	٩٢ ٤	٨٧٢١
٨	٤٦٤٢	٥٠١٧

(أ) هل تؤيد إدعاء المحاسب (أ) ؟ إستخدم ∝ = ٥ ٪

(ب) ما هو الفرق الحرج في التمرين أعلاه ؟ وما معناه ؟

(جـ) إعط توزيع الفرق ، وتوزيع الإختبار الإحصائي ، وحدد المنطقة الحرجة على التوزيعين .

٧ - أردت أن تقدر الفرق فى الحوافز السنوية بين السيدات والرجال العاملين بأحد شركات المحاسبية الخاصة الكبرى ، قمت بسحب ٧ من المحاسبين السيدات ، وتم إختيار ٧ مفردات من المحاسبين الرجال يتوافقون مع مفردات السيدات من حيث عدد سنوات الخبرة ، ساعات العمل ، دقة العمل ، أيام الغياب وخلافه . وحصلت على البيانات الآتية

رقم المفردة	الحوافز السنوية	
	الرجال	السيدات
١	١٤٥٠	١٢٠٠
٢	٧٥٠	٨٠٠
٣	١١٥٠٠	١٠٠٠
٤	٨٠٠	٨٠٠
٥	٦٧٥	٧٢٥
٦	٧٧٥	٨٢٥
٧	١٢٢٥	١٢٠٠

(١) إستخدام ∝ = ٥ ٪ واختبر الفرض بأن الحوافز متساوية للرجال والنساء.

(٢) ما هي الافتراضات التي يجب إفتراضها حتى يكون الإختبار الإحصائي صحيحًا .

(٣) إعط توزيع الفروق وتوزيع الإختبار الإحصائي وحدد المنطقة الحرجة على التوزيعين .

٨ - في تجربة طبية ، قبل إعطاء العقار وبعد إعطاء العقار ، تم الحصول على النتائج الآتية :

رقم المريض	١	٢	٣	٤	٥
قبل	٨٣	٦٠	٥٥	٩٩	٧٧
بعد	٩٢	٧١	٥٦	١٠٤	٨٩

(أ) المطلوب إختبار الفرض العدمي بعدم وجود أثر للعقار ، استخدم α = ١٠ ٪ .

(ب) إعط توزيع الفروق وتوزيع الإختبار الإحصائي وحدد مناطق القبول والرفض على التوزيعين .

(جـ) إختبر الفروض العدمية الآتية :

(١) $\mu_{\text{قبل}}$ = ٥ ، (٢) $\mu_{\text{بعد}}$ = ٥٠

وقارن بين ما حصلت عليه بإجابتك في (أ) أعلاه .

٩ - اعطيت تجربة لعينتين مرتبطتين وكانت النتائج كالتالي

رقم المفرده	١	٢	٣	٤	٥	٦	٧
س	٥٥	٦٨	٤٠	٥٥	٧٥	٥٢	٤٩
ص	٤٤	٥٤	٢٥	٥٦	٦٢	٣٨	٣٨

(أ) إختبر : $\mu_{\text{س}} \le$ ٦٠ α = ٥ ٪

(ب) إختبر : $\mu_{\text{ص}} \le$ ٦٠ α = ٥ ٪

(جـ) إختبر : ($\mu_{\text{س}} - \mu_{\text{ص}}$) = صفر α = ٥ ٪

(د) إعط توزيع الفروق وتوزيع الإختبار الإحصائي في (أ) ، (ب) ، (جـ) أعلاه ووضح على التوزيعين مناطق القبول ومناطق الرفض ، مستخدمًا α = ٥ ٪ .

٤/٣ اختبار الفرق بين نسبتين

يحاول متخذوا القرارات في كثير من المواقف مقارنة النسبة في مجتمعين. مثلاً هل تختلف نسبة مشاهدة المسلسلات التليفزيونية بين الرجال والنساء ؟ هل تختلف نسبة الأمية بين الريف والحضر ؟ هل تختلف نسبة مرض الفشل الكلوي بين الريف والحضر ؟ هل تختلف نسبة المعيب في إنتاج الشركة (أ) عن إنتاج الشركة (ب)؟

كل هذه أسئلة ، تتطلب قياس الفرق بين النسبتين ، ومعرفة هل هذه الفروق جوهرية أم لا . في حالة وجود إختلاف بين النسبتين ، أو أحد النسب أكبر من النسبة الأخرى ، فإن هذا معناه وجود علاقة بين المجتمعين والظاهرة المقاسة . مثلاً إذا وجد أن نسبة المدخنين بين الرجال أقل من ذات النسبة بين السيدات ، فإن هذا معناه وجود علاقة بين التدخين والنوع . وإذا وجد إختلاف بين نسبة مرض الفشل الكلوي بين الريف والحضر ، فإن هذا معناه وجود علاقة بين الفشل الكلوي والمنطقة التي نشأ فيها الفرد . وهكذا.

واختبار الفرق بين نسبتين ، يعتمد على التوزيع الاحتمالي للفرق بين نسبتين من عينتين مستقلتين . فإذا رمزنا للنسبة في العينة الأولى بالرمز $\hat{ن}_1$ والنسبة في العينة الثانية بالرمز $\hat{ن}_2$ حيث :

$$\hat{ن}_1 = \frac{س_1}{ن_1} \quad ، \quad \hat{ن}_2 = \frac{س_2}{ن_2}$$

، $م_1$ = عدد المشاهدات التي يتطبق عليها الظاهرة في العينة الأولى

، $م_2$ = عدد المشاهدات التي يتطبق عليها الظاهرة في العينة الثانية

، $ن_1$ = حجم العينة الأولى ، $ن_2$ = حجم العينة الثانية .

وطبقًا لنظرية الحد المركزي فإن :

$$ى = \frac{(\hat{ل}_1 - \hat{ل}_2) - (ل_1 - ل_2)_.}{ع_{\hat{ل}_1 - \hat{ل}_2}}$$

حيث $(ل_1 - ل_2)_.$ هو القيمة تحت صحة الفرض العدمي . وتكون نظرية الحد المركزي صحيحة في إختبار الفرق بين نسبتين : إذا تحققت الشروط الآتية :

(١) $ن_1 \geq 30$ ، $ن_2 \geq 30$

(٢) $ن_1 \hat{ل}_1 \geq 5$ ، $ن_2 \hat{ل}_2 \geq 5$

(٣) $ن_1 (1-\hat{ل}_1) \geq 5$ ، $ن_2 (1-\hat{ل}_2) \geq 5$

والخطأ المعياري $ع_{\hat{ل}_1 - \hat{ل}_2}$ يقدر بأحد طريقتين ، حسب طريقة صياغة الفرض العدمي :

١ إذا كان الفرض العدمي $ل_1 \leq ل_2$ أي أن $ل_1 - ل_2 \leq$ صفر أو $ل_1 \geq ل_2$ أي أن $ل_1 - ل_2 \geq$ صفر في هذه الحالة يفترض إحتمال عدم تساوي النسبتين في المجتمع ، ويقدر الخطأ المعياري باستخدام المعادلة :

$$\text{ع}_{\hat{\text{ل}}_1 - \hat{\text{ل}}_2} = \sqrt{\frac{\hat{\text{ل}}_1 (1 - \hat{\text{ل}}_1)}{\text{ن}_1} + \frac{\hat{\text{ل}}_2 (1 - \hat{\text{ل}}_2)}{\text{ن}_2}}$$

٢ – إذا كان الفرض العدمي $\text{ل}_1 - \text{ل}_2$ = صفر ، يقدر الخطأ المعياري على مرحلتين . أولاً نقوم بحساب النسبة العامة باستخدام العينتين ، ثم نحسب الخطأ المعياري للفرق بين النسبتين باستخدام النسبة العامة ، $\hat{\text{ل}}$ كالآتي :

$$(1)\quad \hat{\text{ل}} = \frac{\text{ن}_1 \hat{\text{ل}}_1 + \text{ن}_2 \hat{\text{ل}}_2}{\text{ن}_1 + \text{ن}_2} = \frac{\text{س}_1 + \text{س}_2}{\text{ن}_1 + \text{ن}_2}$$

$$(2)\quad \text{ع}_{\hat{\text{ل}}_1 - \hat{\text{ل}}_2} = \sqrt{\frac{\hat{\text{ل}} (1 - \hat{\text{ل}})}{\text{ن}_1} + \frac{\hat{\text{ل}} (1 - \hat{\text{ل}})}{\text{ن}_2}}$$

$$= \sqrt{\hat{\text{ل}} (1 - \hat{\text{ل}}) \left(\frac{1}{\text{ن}_1} + \frac{1}{\text{ن}_2} \right)}$$

وتكون خطوات إختبار الفرق بين نسبتين :

١ – صياغة الفروض :

ف_0 : $\text{ل}_1 = \text{ل}_2$ أو $(\text{ل}_1 - \text{ل}_2)$ = صفر

ف_1 : $\text{ل}_1 \neq \text{ل}_2$ أو $(\text{ل}_1 - \text{ل}_2) \neq$ صفر

حيث الإختبار في إتجاهين . ويمكن أن يكون الإختبار في إتجاه واحد ناحية اليمين أو ناحية اليسار .

٢ - سحب عينتين مستقلتين ، وإحسب $\hat{ل}_1$ ، $\hat{ل}_2$.

٣ - تحقق من أن : (١) $ن_1$ ، $ن_2$ $\geq$ ٣٠

(٢) $ن_1 (1 - \hat{ل}_1) \geq 5$ ، $ن_2 (1 - \hat{ل}_2) \geq 5$

(٣) $ن_1 \hat{ل}_1 \geq 5$ ، $ن_2 \hat{ل}_2 \geq 5$

إذا تحققت الشروط أعلاه ، يكون الإختبار الإحصائي هو إختبار ى .

حدد مستوى المعنوية وحدد القيمة الحرجة $ى_{\alpha/2}$ أو $ى_{\alpha}$ أو $- ى_{\alpha}$.

٤ - إحسب قيمة الخطأ المعياري للفرق بين نسبتي عينتين : $ع_{\hat{ل}_1 - \hat{ل}_2}$

٥ - إحسب قيمة الإختبار الإحصائي :

$$ى = \frac{(\hat{ل}_1 - \hat{ل}_2) - (ل_1 - ل_2)}{ع_{\hat{ل}_1 - \hat{ل}_2}}$$

٦ - اتخذ قرار بشأن صحة الفرض العدمي ، حيث :

إتجاه الفرض البديل	يرفض الفرض العدمي إذا كانت :
$\neq$	$\mid$ ى المحسوبة $\mid$ > $ى_{\alpha/2}$ الجدولية
>	ى المحسوبة > + $ى_{\alpha}$ الجدولية
<	ى المحسوبة < - $ى_{\alpha}$ الجدولية

ولاحظ أوجه التشابه بين إختبار النسبة في مجتمع واحد وإختبار الفرق بين نسبتى مجتمعين :

نسبة واحدة	نسبتين
$\text{ى} = \dfrac{\hat{\text{ن}} - \text{ل}_{.}}{\text{خ}_{\hat{\text{ل}}}}$	$\text{ى} = \dfrac{(\hat{\text{ن}}_{1} - \hat{\text{ل}}_{2}) - (\text{ل}_{1} - \text{ل}_{2})_{.}}{\text{خ}_{\hat{\text{ل}}_{1} - \hat{\text{ل}}_{2}}}$

وفى الحالتين يقيم الفرق بين نتائج العينة والقيمة المفترضة كنسبة من الخطأ المعيارى .

تطبيق (٣-٤)

رات وزارة التأمينات معرفة هل نسبة إصابة العمل واحدة بين الرجال والنساء أم لا . قامت الوزارة بسحب عينة عشوائية حجمها ١٠٠ من الرجال وعينة عشوائية أخرى حجمها ٨٠ من السيدات . ووجدت أن إصابة العمل في عينة الرجال ١٦ حالة وفي عينة السيدات ٩ حالات .

فقام مسئول الأبحاث في الوزارة بالآتي :

١ - صياغة الفروض : نسبة إصابة العمل واحدة في المجتمعين :

$\text{ف}_{.} : \text{ل}_{1.} = \text{ل}_{2.} \quad \text{أو} \quad (\text{ل}_{1} - \text{ل}_{2})_{.} = \text{صفر}$

$\text{ف}_{1} : \text{ل}_{1.} \neq \text{ل}_{2.} \quad \text{أو} \quad (\text{ل}_{1} - \text{ل}_{2})_{.} \neq \text{صفر}$

٢ - وجد ان نسبتى إصابة العمل فى العينتين كالآتى :

$$\hat{ل}_1 = \frac{١٦}{١٠٠} = ٠,١٦ \qquad \hat{ل}_2 = \frac{٩}{٨٠} = ٠,١١$$

وحيث أن النسبتين ليستا متساويتين (٠,١٦ ≠ ٠,١١) ، والفرق بين النسبتين المفترضتين يختلف عن الصفر ، حيث : (٠,١٦ - ٠,١١) = ٠,٠٥ . وهذا الفرق يقع فى إتجاه الفرض البديل ، لذا قرر مسئول الأبحاث الإستمرار في خطوات الإختبار .

٣ - حيث أن :

$$ن_1 = ١٠٠ \geq ٣٠ \quad ، \quad ن_2 = ٨٠ \geq ٣٠$$

$$ن_1 \hat{ل}_1 = ١٠٠ \times ٠,١٦ = ١٦ \geq ٥ \quad ، \quad ن_2 \hat{ل}_2 = ٨٠ \times ٠,١١ = ٩ \geq ٥$$

$$ن_1 (١ - \hat{ل}_1) = ١٠٠ \times ٠,٨٤ = ٨٤ \geq ٥ \quad ،$$

$$ن_2 (١ - \hat{ل}_2) = ٨٠ \times ٠,٨٩ = ٧١ \geq ٥$$

∴ الإختبار الإحصائي المناسب هو إختبار ى ، وعند α = ١٠ ٪ نجد أن المنطقة الحرجة هي المنطقة ≤ ١,٦٥ ، ≥ - ١,٦٥ .

٤ - نحسب الخطأ المعيارى $ع_{\hat{ل}_1 - \hat{ل}_2}$

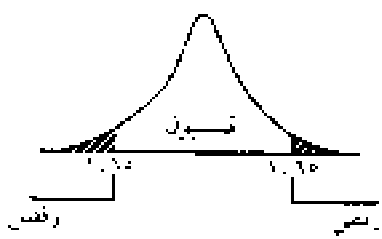
$$ع_{\hat{ل}_1 - \hat{ل}_2} = \sqrt{\hat{ل} \, (١ - \hat{ل}) \left(\frac{١}{ن_1} + \frac{١}{ن_2} \right)}$$

$$[illegible] = \frac{[illegible]}{[illegible]} = \frac{[illegible]}{١٨} = [illegible]$$

$$[illegible] = \sqrt{[illegible]\left(\frac{١}{[illegible]} + \frac{١}{[illegible]}\right)}$$

$$= \sqrt{[illegible]} = [illegible]$$

[illegible] - الاختبار الإحصائي



$$ى = \frac{[illegible]}{[illegible]}$$

$$ى = \frac{([illegible]) - صفر}{٠,٠٢٣} = ٠,٩٦$$

٦ - حيث أن قيمة ى المحسوبة (٠,٩٦) تقع داخل منطقة القبول ، إذن لا نرفض الفرض العدمي ، ونستطيع أن نقول أن النسبة متساوية في المجتمعين.

الفرق الحرج

١ - إذا كان الإختبار في إتجاهين

الفرق الحرج = $(\text{ل}_1 - \text{ل}_2) + \text{ى}_{\alpha/2} \times \hat{\text{ع}}_{\hat{\text{ل}}_1 - \hat{\text{ل}}_2}$

٢ - إذا كان الإختبار في إتجاه واحد ناحية (يمين)

الفرق الحرج = $(\text{ل}_1 - \text{ل}_2) + \text{ى}_{\alpha} \times \hat{\text{ع}}_{\hat{\text{ل}}_1 - \hat{\text{ل}}_2}$

٣ - إذا كان الإختبار في إتجاه واحد ناحية (يسار)

الفرق الحرج = $(\text{ل}_1 - \text{ل}_2) - \text{ى}_{\alpha} \times \hat{\text{ع}}_{\hat{\text{ل}}_1 - \hat{\text{ل}}_2}$

في المثال أعلاه :

الفرق الحرج = $(\text{ل}_1 - \text{ل}_2) \pm \text{ى}_{\alpha/2} \times \hat{\text{ع}}_{\hat{\text{ل}}_1 - \hat{\text{ل}}_2}$

= صفر ± ١,٦٥ × ٠,٠٢٥ = ± ٠,٠٨٥٨

وحيث أن الفرق المشاهد هو (٠,١٦ - ٠,١١) = ٠,٠٥

∴ الفرق المشاهد (٠,٠٥) < الفرق الحرج (٠,٠٨٥٨)

∴ لا نرفض الفرض العدمي . ويتم رفض الفرض العدمي إذا كان الفرق المشاهد ≥ ٠,٠٨٥٨ .

تطبيق (٣-٥)

تعرف إدارات المستخدمين بصفة عامة ، أن نسبة السيدات اللاتي لا يرضين عن ترقياتهم الوظيفية أعلى من نسبة الرجال غير الراضين بحوالي ٢٠ ٪ . أراد مدير الإدارة بحث هذا الموقف ، وهل نسبة السيدات غير الراضيات أعلى فعلاً ٢٠ ٪ عن نسبة الرجال غير الراضين :

فتم صياغة الفروض كالآتي :

$$ف_٠ : ن_١ - ن_٢ \leq ٠,٢٠ \quad ، \quad ف_١ : ن_١ - ن_٢ > ٠,٢٠$$

حيث $ن_١$ هي نسبة السيدات غير الراضيات ، $ن_٢$ نسبة الرجال غير الراضين .

وقام الباحث بسحب عينة من السيدات حجمها ٧٥ سيدة ، وعينة من الرجال حجمها ٧٥ رجلاً . ووجد أن عدد السيدات غير الراضيات في هذه العينة ٥٤ سيدة ، وعدد الرجال غير الراضين ٤٨ رجلاً . فقام بحساب الآتي :

$$\hat{ن}_١ = \frac{٥٤}{٧٥} = ٠,٧٢ \quad ، \quad \hat{ن}_٢ = \frac{٤٨}{٧٥} = ٠,٦٤$$

ووجد الباحث أن الفرق بين النسبتين في العينتين (٠,٧٢ - ٠,٦٤) = ٠,٠٨ تقع في إتجاه الفرض البديل . فاستمر في خطوات الإختبار :

وحيث أن :

$ن_1 \geq ٣٠$ ، $ن_2 \geq ٣٠$

$ن_1 \hat{ل}_1 \geq ٥$ ، $ن_2 \hat{ل}_2 \geq ٥$

$ن_1 (١ - \hat{ل}_1) \geq ٥$ ، $ن_2 (١ - \hat{ل}_2) \geq ٥$

∴ الإختبار الإحصائى هو إختبار ى ، ومستوى المعنوية التى اختارها هى ٥ ٪ والإختبار فى إتجاه واحد ناحية اليسار .

∴ المنطقة الحرجة هى المنطقة ≤ - ١,٦٤ .

الخطأ المعيارى للإختبار الإحصائى :

حيث أن الفرض العدمى : ف. : $ل_1 - ل_2 \leq$ صفر

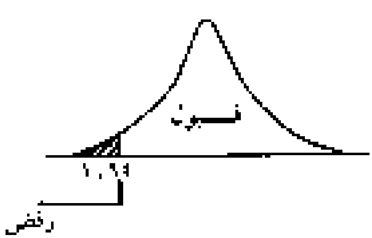
$$∴ \; ع_{\hat{ل}_1 - \hat{ل}_2} = \sqrt{\frac{\hat{ل}_1 (١ - \hat{ل}_1)}{ن_1} + \frac{\hat{ل}_2 (١ - \hat{ل}_2)}{ن_2}}$$

$$= \sqrt{\frac{٠,٧٢ (١ - ٠,٧٢)}{٧٥} + \frac{٠,٦٤ (١ - ٠,٦٤)}{٧٥}}$$

$$= ٠,٠٧٥٨٩ = ٠,٠٧٦$$

$$∴ \; ى = \frac{(٠,٧٢ - ٠,٦٤) - ٠,٢}{٠,٠٧٦}$$

$$= - ١,٥٨$$



وحيث أن قيمة ى المحسوبة (– ١,٥٨) تقع في منطقة القبول ، إذن لا نرفض الفرض العدمي ونستطيع القول أن نسبة السيدات غير الراضيات عن ترقياتهن أعلى من نسبة الرجال غير الراضين بـ ٢٠ ٪ على الأقل .

الفرق الحرج

الإختبار في إتجاه واحد ناحية اليسار

∴ الفرق الحرج = (ل$_1$ – ل$_2$)$_0$ – ى$_\alpha$ × ع$_{\hat{ل}_1 - \hat{ل}_2}$

= ٠,٢ – ١,٦٥ × ٠,٠٧٦

= ٠,٠٧٤٦ = ٠,٠٧٥ تقريباً

وحيث أن الفرق المشاهد ($\hat{ل}_1$ – $\hat{ل}_2$) = ٠,٠٨ > ٠,٠٧٥

∴ نقبل الفرض العدمي ، ونصل إلى النتيجة السابقة .

تمارين

١٠ – في عينة عشوائية حجمها ٢٠٠ طالب ، وجد أن من يؤيد نظام الفصلين الدراسيين ٣٨ طالباً . وفي عينة عشوائية من ٢٢٠ طالبة وجد أن من يؤيد هذا النظام ٧٦ طالبة . هل نسبة من يؤيد نظام الفصلين الدراسيين واحدة بين الطلبة والطالبات ؟ إستخدم α = ٥ ٪ .

١١ – سحبت عينتان عشوائيتان حجم كل منهما ٥٠٠ مفردة . وسئلت مفردات العينتين عن جدوى الخدمة العامة بالنسبة للخريجين فوجد أن :

م$_1$ = ١٤٠ ، م$_2$ = ١٩٢

هـــل تعتقد أن نسبة من يؤيد الخدمة العامة واحدة في المجتمعين ؟ إستخدم ∝ = ٥ ٪ .

١٢- في عينة من ١٠٠٠ مفردة وجد أن م١ = ٢٨٠ وفي عينة من ١٢٠٠ مفردة وجد أن م٢ = ٣٩٦ .

إختبر الفرض العدمي بأن ل١ - ل٢ ≤ ٠٫٠٦ ، وإستخدم ∝ = ٥ ٪ .

وضح توزيع (لˆ١ - لˆ٢) ، وتوزيع الإختبار الإحصائي موضحاً المنطقة الحرجة على التوزيعين .

١٣- أرادت شركة متخصصة في تقديم الوجبات السريعة تقدير الفرق بين نسبة من يروا أن وجباتهم ممتازة في فرعين من فروع الشركة .

سحبت عينة عشوائية من الفرع (١) مكونة من ٢٠٠ فرد وأوضحت نتائج العينة أن ١١٤ فرداً رأوا أن الخدمة والوجبة ممتازتين .

وسحب عينة عشوائية من الفرع (٢) مكونة من ٢٥٠ فرداً وأوضحت نتائج هذه العينة أن ١٥٥ فرداً رأوا أن الخدمة والوجبة ممتازتين .

(أ) قدر الفرق بين النسبتين بإنشاء فترة ثقة ٩٥ ٪ .

(ب) اختبر تساوي النسبتين في الفرعين ∝ = ٥ ٪ .

(جـ) إعط الفرق الحرج وإشرح معناه . وإعط توزيع الفرق وتوزيع الإختبار الإحصائي وحدد المناطق الحرجة على التوزيعين .

١٤- أرادت شركة تصنيع معجون الأسنان مقارنة نسبة مستخدمي نوعين من منتجاتها . في عينة من ٥٠٠ مستخدم ، أوضح ٢٠ ٪ منهم إستخدامهم للنوع الأول . وفي عينة من ٤ مستخدم ، أوضح ١٣ ٪ إستخدامهم للنوع الثاني

(أ) أنشيء فترة ثقة ٩٠ ٪ لتقدير الفرق بين نسبة مستخدمي النوعين .

(ب) إختبر الفرض العدمي :

ف. : ل١ - ل٢ ≥ ٠٫٠٥ ، حيث ل١ هو نسبة مستخدمي معجون الأسنان للنوع الأول ، ل٢ هو نسبة مستخدمي معجون الأسنان للنوع الثاني .

(جـ) إعط شكل توزيع الفرق بين النسبتين وشكل التوزيع الإحتمالي للإختبار الإحصائي .

ووضح على الشكلين مناطق القبول ومناطق الرفض .

٥/٣ اختبار تساوي تباينين

عرضنا كيفية إختبار تساوي متوسطين ، أي ف. : $\mu_1 = \mu_2$ ، في الباب السابق . ويوجد إفتراض هام لإجراء الإختبار ، بالأسلوب الذي تم عرضه ، وهو إفتراض أن التباين في المجتمعين متساويين ، أي إفتراض تجانس التباين ، وتحت صحة هذا الإفتراض أمكننا حساب التباين المجتمع (ع^{2}م) بإستخدام بيانات العينتين ، و (ع^{2}م) هو قيمة مقدر σ^2 للمجتمعين .

∴ ع^{2}م : مقدر جيد لتباين المجتمعين إذا كان : $\sigma_1^2 = \sigma_2^2$.

والمقدر الجيد هو مقدر غير متحيز ، ومتسق وكفء .

أما إذا كان $\sigma_1^2 \neq \sigma_2^2$ أي أن شرط تجانس التباين غير محقق ، فإن الإختبار كما تم عرضه ، لا يكون الإختبار الصحيح ، ويلزم إستخدام إختبار لا معلمي .

التوزيع الاحتمالي لنسبه تباينين لعينتين مستقلتين ، سحبا بطريقة عشوائية من مجتمعين يتبعا التوزيع الطبيعي ، يتبع توزيع ف F-distribution الاحتمالي.

إذن الإختبار الإحصائي ، لاختبار تساوي تباينين من مجتمعين مستقلين هو نسبه تباين العينتين .

ومع أن إختبار F يشترط أن يكون توزيع الظاهرة محل الدراسة في المجتمعين ، يتبع التوزيع الطبيعي ، إلا أنه يمكن أيضا إستخدامه طالما أن التوزيع ذو منوال واحد Single-Peaked وحجم العينتين متساوي تقريبا .

اختبار ف F-test

يتبع الإختبار ست خطوات هي :

١ - صياغة الفروض .

$$ف_{.} : \sigma^{٢}_{١} = \sigma^{٢}_{٢} = \sigma^{٢}_{.}$$

أو . $$ف_{.} : \frac{\sigma^{٢}_{١}}{\sigma^{٢}_{٢}} = ١$$

ضد $$ف_{١} : \frac{\sigma^{٢}_{١}}{\sigma^{٢}_{٢}} \neq ١ \text{ ، أو : } < ١ \text{ ، أو : } > ١$$

٢- سحب عينتين مستقلتين وحساب تباين كل عينة $ع^2_1$ ، $ع^2_2$. وفى هذه الخطوة تـــقارن { $ع^2_1$ ، $ع^2_2$ } ، فإذا كانت النسبة في إتجاه الفرض العدمي (= ≤ ≥) نتوقف عند هذه الخطـوة ، وإذا كانت فى اتجاه الفرض البديل ، نستمر في الاختبار .

٣- يحدد إذا كان الاختبار فى اتجاهين او في إتجاه واحد ، ويحدد مستوى المعنوية ∝ . والمنطقة الحرجة تتحدد من توزيع F الاحتمالي بدرجات حرية فى البسط وفى المقام كالآتى :

درجات حرية البسط = (حجم العينة فى بسط المعادلة - ١)

درجات حرية المقام = (حجم العينة في مقام المعادلة - ١)

وتحدد المنطقـة الحرجة باستخدام جدول تـوزيع ف (جدول (٤) ، جدول (٥)) . ونسـتخدم (∝/ ٢) إذا كان الاختبـار في إتجاهين ، (∝) إذا كان الاختبار فى إتجاه واحد .

٤- نحسب قيمة الاختبار الإحصائى :

$$ف = \frac{ع^2_1}{ع^2_2}$$

او باستخدام الجدول التالى :

الفرض العدمي	الفرض البديل	اختبار ف
$\sigma_{١}^{٢} = \sigma_{٢}^{٢}$	$\sigma_{١}^{٢} \neq \sigma_{٢}^{٢}$	$\frac{\text{تباين العينة (الاكبر)}}{\text{تباين العينة (الأصغر)}}$
$\sigma_{١}^{٢} \geq \sigma_{٢}^{٢}$	$\sigma_{١}^{٢} < \sigma_{٢}^{٢}$	$\frac{\text{ع}_{١}^{٢}}{\text{ع}_{٢}^{٢}}$
$\sigma_{١}^{٢} \leq \sigma_{٢}^{٢}$	$\sigma_{١}^{٢} > \sigma_{٢}^{٢}$	$\frac{\text{ع}_{٢}^{٢}}{\text{ع}_{١}^{٢}}$

وباتباع الجدول أعلاه ، وهو وضع التباين الاكبر في بسط معادلة F ، تقع المنطقة الحرجة على اليمين دائما . وذلك حيث أن قيمة F المحسوبة تكون دائما أكبر من الواحد الصحيح . وبالتالي يكون إهتمامنا بمنطقة الرفض على اليمين فقط ، حتى إذا كان اتجاه الإختبار ناحية اليسار ، كما سيتضح في الخطوة (٥) التالية .

٥ الوصول إلى قرار بشأن الفرض العدمي كالآتي .

علامة الفرض البديل	رفض الفرض العدمي إذا كانت :
$\neq$	قيمة ف المحسوبة > (ف$_{ح/٢}$)
> أو <	قيمة ف المحسوبة > (ف$_{ح}$)

القيم الحرجة من توزيع ف

توزيع ف ، مثله كمثل توزيع كا^2 ، توزيع ملتو ناحية اليمين . ولكن تختفي درجة الإلتواء مع زيادة درجات الحرية .

وجميع قيم ف موجبة فتوزيع ف يمتد من الصفر إلى ما لا نهاية . و ف عائلة من التوزيعات : مثل توزيعات ت ، وتوزيع كا^2 . ويتحدد شكل التوزيع ، ودرجة التواءه بدرجة حرية البسط ودرجة حرية المقام .

وفي توزيع ف يوجد زوجاً من درجات الحرية ، هي درجة حرية التباين في بسط معادلة ف ، ودرجات حرية التباين في مقام معادلة ف .

ويستخدم الرمز فـ ي لتشير إلى قيمة ف الحرجة والإختبار في اتجاه واحد ، والرمز فـ ي/٢ ليشير إلى قيمة ف الحرجة والإختبار في اتجاهين .

جدولي توزيع ف (جدول ٤ ، ٥) . جدول (٤) يعطي القيم الحرجة ، إذا كان مساحة الذيل الأيمن في التوزيع مساوياً ٥ ٪ . وجدول (٥) يعطي القيم الحرجة ، التي تجعل الذيل الأيمن مساوياً ١ ٪ .

ولاستخدام أي من الجدولين (٤) أو (٥) : نحدد درجة حرية البسط (df numerator) ومحدده في الأعمدة ، ونحدد درجة حرية المقام (df denomintor) ومحدده في الصفوف . وتقاطع عمود درجة حرية البسط مع صف درجة حرية المقام نصل إلى القيمة الحرجة .

مثلاً فـ و بدرجات حرية (١٠ ، ٨) = ٣,٣٤ (جدول (٤))

البسط

المقام

ف$_{٠,٠٥}$ بدرجـــات حريـــة (٨ ، ١٠) = ٣,٠٧ (جدول (٤))

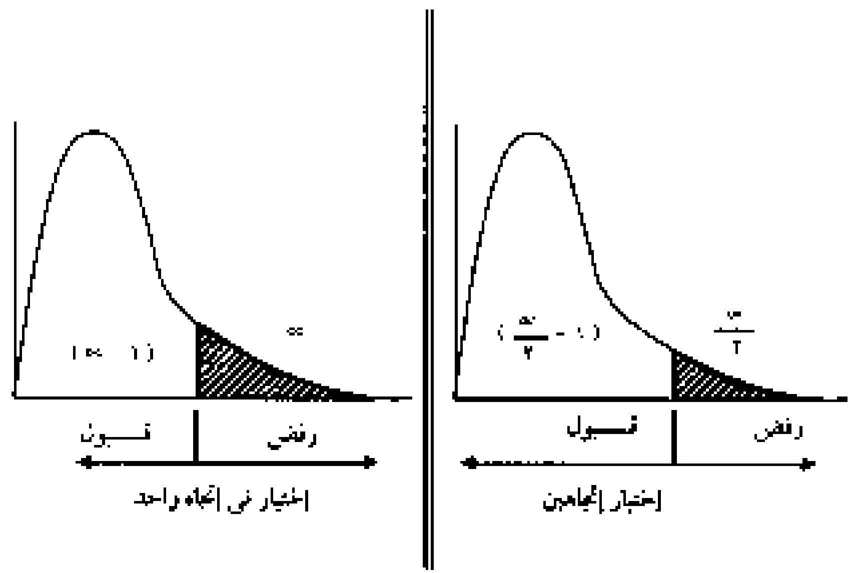
البســط ← ٨

المقـــام ← ١٠

ف$_{٠,٠١}$ بدرجـــات حريـــة (١٢ ، ٦) = ٧,٧٢ (جدول (٥))

ف$_{٠,٠١}$ بدرجـــات حريـــة (٦ ، ١٢) = ٤,٨٢ (جدول (٥))

وتحدد المنطقة الحرجة ، حسب الإستراتيجية المتبعة ، وهي وضع تباين العينة الأكبر في البسط كالتالي :



تطبيق (٣-٦)

تهتم مصانع مواد اللصق ، بقوة تحمل المادة اللاصقة ، وتتم المقارنة بين مادة لاصقة وأخرى بمقارنة متوسط قوة التحمل للمادتين . ويرى أحد المصانع ، أنه أيضا لا يرغب في أن تكون درجة الإختلاف في قوة التحمل كبيرة في المادة المفضلة التي يستخدمها .

ورأى المصنع مقارنة درجة الإختلاف في قوة التحمل في مادة لاصقة مستخدمة حاليا ، ومادة أخرى مطورة ، ويأمل أن تكون درجة الإختلاف في قوة التحمل في المادة المطورة أقل من درجة الإختلاف في قوة التحمل للمادة المستخدمة حاليا .

فتم صياغة الفروض كالتالي :

ف. : $\sigma_1^2 \le \sigma_2^2$ ، ف١ : $\sigma_1^2 > \sigma_2^2$

حيث σ_2^2 هو تباين قوة التحمل في النوع الجديد .

، σ_1^2 هو تباين قوة التحمل في النوع الحالي .

وقام المصنع في استخدام المادة اللاصقة للصق أرجل أربعة كراسي ، مستخدما المادة المطورة ، وفي أرجل أربع كراسي مستخدما المادة الحالية . وكانت قوة التحمل (بالكجم) إلى أن تتكسر الأرجل كالآتي :

المادة الجديدة	المادة الحالية
٢٨,٤	٢٧,٤
٢٩,٤	٣٠,٢
٢٨,٧	٢٩,٨
٢٨,٥	٣١,٠

فقام الباحث بحساب تباين قوة التحمل مستخدما المادة اللاصقة الجديدة وكانت كالآتي :

$$\text{مجـ س}^{\text{٢}} = \text{٣٢٩٥,٥} \qquad\qquad \text{مجـ س} = \text{١١٤,٨}$$

$$\therefore\ \text{ع}^{\text{٢}} = \frac{\text{ن مجـ س}^{\text{٢}} - (\text{مجـ س})^{\text{٢}}}{\text{ن (ن} - \text{١)}}$$

$$= \frac{\text{٤ (٣٢٩٥,٥)} - (\text{١١٤,٨})^{\text{٢}}}{\text{٤ (٤} - \text{١)}} = \text{,٢٥}$$

وتم حساب تباين قوة التحمل ، مستخدما المادة اللاصقة الحالية كالتالي :

$$\text{مجـ س}^{\text{٢}} = \text{٣٥١١,٨} \qquad\qquad \text{مجـ س} = \text{١١٨,٤}$$

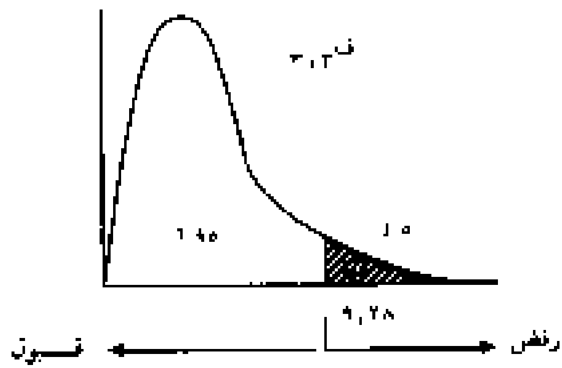
$$\therefore\ \text{ع}^{\text{٢}} = \frac{\text{ن مجـ س}^{\text{٢}} - (\text{مجـ س})^{\text{٢}}}{\text{ن (ن} - \text{١)}}$$

$$= \frac{\text{٤ (٣٥١١,٨)} - (\text{١١٤,٨})^{\text{٢}}}{\text{٤ (٤} - \text{١)}} = \text{٢,٣٩}$$

وهنا نجد أن نسبة تباين العينتين $= \dfrac{\text{,٢٥}}{\text{٢,٣٩}} < \text{١}$ وتقع فى اتجاه الفرض البديل . إذن لا نستطيع قبول الفرض العدمى ، عند هذه النقطة . ونستمر فى الاختبار .

بفرض أن قوة تحمل المادة اللاصقة ، تتبع التوزيع الطبيعى ، نحسب قيمة الاختبار الإحصائى $= \text{ف} = \dfrac{\text{التباين الأكبر}}{\text{التباين الأصغر}} = \dfrac{\text{٢,٣٩}}{\text{,٢٥}} = \text{٩,٦٥}$

والمنطقة الحرجة : $\text{ف}_{\text{,٠٥}}$ بدرجات حرية (٣ ، ٣) = ٩,٢٨



وحيث أن القيمة المحسوبة (٩,٦٥) ≤ ف الحرجة (٩,٢٨) .

∴ نرفض الفرض العدمي ، ونقبل الفرض البديل بأن تباين المادة اللاصقة الجديدة أقل من تباين المادة اللاصقة الحالية .

تطبيق (٣-٧)

في تطبيق (٣-١) تم مقارنة متوسط عدد الكيلو مترات التي تتحملها الإطارات العادية بتلك التي تعطيها الإطارات المميزة . ووجد أن تباين عينة الإطارات العادية = ٤٣٨٤كم٢ وتباين عينة الإطارات المميزة = ٦٠٧١ كم٢ . وتم رفض الفرض العدمي بأن الفرق في الكيلو مترات أقل من ١٢٥٠٠ كم٢ .

والآن نود أن نختبر الفرض الذي إستند عليه هذا الإختبار ، وهو إختبار تجانس تباين المجتمعين . وكان حجم عينة الإطارات العادية ٢٠ إطارا وحجم عينة الإطارات المميزة ١٦ إطارا .

تكون الفروض كالآتي :

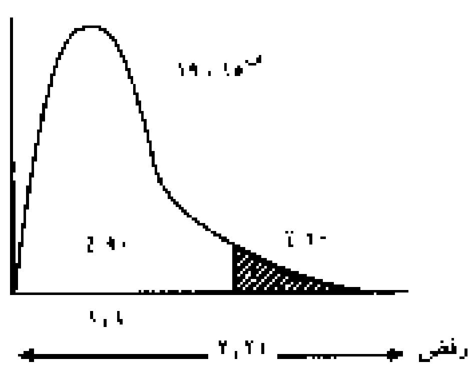
$$ف_{٠} : \sigma_{١}^{٢} = \sigma_{٢}^{٢} \quad \text{أو} \quad \frac{\sigma_{١}^{٢}}{\sigma_{٢}^{٢}} = ١$$

$$ف_{١} : \sigma_{١}^{٢} \neq \sigma_{٢}^{٢} \quad \text{أو} \quad \frac{\sigma_{١}^{٢}}{\sigma_{٢}^{٢}} \neq ١$$

حيث $\sigma_{١}^{٢}$: تباين الإطارات المميزة ، $\sigma_{٢}^{٢}$: تباين الإطارات العادية

∴ الإختبار في إتجاهين ، وباعتبار أن α = ١٠ ٪ . الإختبار الإحصائي لاختبار تساوي تباينين هو اختبار ف ، وتكون القيمة الحرجة إذا تم وضع التباين الأكبر في البسط هي : ف٠,٠٥ (١٥ ، ١٩) = ٢,٢١

$$\text{الإختبار الإحصائي : ف} = \frac{\text{التباين الأكبر}}{\text{التباين الأصغر}} ، \frac{٦٠٧١}{٤٣٤٨} = ١,٤٠$$



∴ نقبل الفرض العدمي بأن التباين متجانس في المجتمعين ، ويكون استخدام ع^{٢}م كمقدار لتباين المجتمعين صحيحاً .

تمارين

١ - أوجد قيم ف الحرجة فى الحالات الآتية :

(أ) α = ٥ ٪ درجة حرية البسط = ٨ درجة حرية المقام = ١٢

(ب) α = ٥ ٪ درجة حرية البسط = ١٢ درجة حرية المقام = ٨

(جـ) α = ١ ٪ درجة حرية البسط = ١٠ درجة حرية المقام = ١٥

(د) α = ١ ٪ درجة حرية البسط = ١٥ درجة حرية المقام = ٢٠

(هـ) α = ١ ٪ درجة حرية البسط = ٥ درجة حرية المقام = ٧

(و) α = ٥ ٪ درجة حرية البسط = ١٤ درجة حرية المقام = ١٤

(ز) α = ١ ٪ درجة حرية البسط = ١٦ درجة حرية المقام = ٢٢

٢ - من المجتمع (أ) سحبت عينة عشوائية حجمها ٢٠ مفردة ، وكان تباين هذه العينة = ٨,٤ ، ومن المجتمع (ب) سحبت عينة عشوائية حجمها ١٥ مفردة ، وكان التباين = ٩,٨٥ .

إعط درجات حرية البسط ودرجات حرية المقام وقيمة ف الحرجة للفروض العدمية التالية :

(أ) $\sigma^2_{\text{أ}} = \sigma^2_{\text{ب}}$ α = ١ ٪

(ب) $\sigma^2_{\text{أ}} \leq \sigma^2_{\text{ب}}$ α = ٥ ٪

(جـ) $\sigma^2_{\text{أ}} \geq \sigma^2_{\text{ب}}$ α = ٥ ٪

٣- من المجتمع (أ) تم سحب عينة عشوائية حجمها ١٤ مفرده ، وأظهرت نتائج العينة أن الوسط الحسابي = ١٠٠ والتباين = ١٢ .

ومن المجتمع (ب) ، تم سحب عينة عشوائية حجمها ١٧ مفرده ، وأظهرت نتائج العينة أن الوسط الحسابي = ١١٨ والتباين ٢٧,٦ .

إعط درجات حرية البسط ودرجات حرية المقام وقيمة ف الحرجة للفروض العدمية التالية :

(أ) $\sigma^2_{\text{أ}} = \sigma^2_{\text{ب}}$ $\qquad \propto = ١٠\%$

(ب) $\sigma^2_{\text{أ}} \leq \sigma^2_{\text{ب}}$ $\qquad \propto = ٥\%$

(جـ) $\sigma^2_{\text{أ}} \geq \sigma^2_{\text{ب}}$ $\qquad \propto = ٥\%$

إستخدام برمجية Minitab في التقدير

وإختبارات الفروض للأوسطة الحسابية والنسب

يبدأ أي تحليل إحصائي بالبيانات ، حيث أن البيانات هي المادة الخام لأي تحليل إحصائي . والبيانات بنوعيها الوصفي والكمي يمكن أن تكون مدخلات برمجية Minitab ، إلا أننا سوف نركز في هذا الملحق على البيانات الكمية .

ويتم إدخال البيانات في صحيفة عمل Worksheet . ويمكن أن يشمل البرنامج الواحد عدداً من صحائف العمل (يعتمد العدد على ذاكرة الكمبيوتر المستخدم) . وتشمل صحيفة العمل عدداً من الخلايا ، تمثل كل خلية رقماً للصف ورقماً للعمود ويرمز للأعمدة Columns بالرمز C ثم رقم العمود بمعنى أن C1 هو العمود الأول ، C2 هو العمود الثاني وهكذا . أما الصفوف فتمثل المشاهدات وترقم الصفوف من ١ إلى عدد المشاهدات

إذن ، لبدء العمل ببرمجية Minitab :

١ - تقوم بإدخال البيانات ، كل متغير في عمود معين ،

٢ - نسمي المتغير باسم يميزه عن غيره ، ويكون ذا معنى للمستخدم ، بدلاً من C2 مثلاً . ونعطي إسم المتغير بالحروف اللاتينية في الصف أسفل رقم العمود وأعلى من أول صف .

والآن بفرض أن البيانات قد تم إدخالها ، وتم تسميتها ، وبفرض إستخدام المثال التالي :

أرادت شركة تأمين تقدير الفرق بين متوسط سرعة قيادة الرجال وسرعة قيادة السيدات على الطرق السريعة . سحبت عينة عشوائية من ١٦ رجل ووجد أن سرعة القيادة على الطرق السريعة هو :

٩٥ ٩٠٠ ٨٩ ١١٠ ٩٠ ٨٩ ٧٤ ١١٢ ١١٣ ٩٧

٩٤ ٨٦ ٩٠ ٩٥ ٩٣ ٩٦

وسحبت عينة عشوائية من ١٤ سيدة ، ووجد أن سرعة القيادة هي :

٨١ ٨٥ ٨٥ ٩٦ ٩٠ ١٠٢ ١١٠

٦٤ ٧٨ ٧٣ ٩٢ ٨٥ ٦٣ ٧٧

١ - تقدير متوسط سرعة الرجال على الطرق السريعة (∝ = ٥ ٪) .

٢ - إختبار متوسط سرعة قيادة الرجال على الطرق السريعة = ٨٠ كم/ساعة .

٣ - تقدير متوسطة سرعة قيادة السيدات (∝ = ٥ ٪) .

٤ - إختبار أن متوسط سرعة السيدات على الطرق السريعة ٧٠ كم/ساعة .

٥ - إختبار أن متوسط سرعة قيادة الرجال متساوية بين الرجال والسيدات .

٦ - قدر الفرق بين سرعة قيادة الرجال وسرعة قيادة السيدات

في برمجية Minitab يتم التقدير بإنشاء فترات ثقة في نفس برنامج إختبار الفروض . ولكن يجب تعريف البرمجية بالبرنامج المطلوب تنفيذه . إذن لتنفيذ المطلوب في (١) ، (٢) أعلاه ، تقوم بالآتي :

١ - من القائمة الرئيسية إختر

Stat → Basic Statistics ⟶ 1 - Sample t

٢ - يظهر صندوق حوار Dialog Box إختار بالماوس . المتغير C1 أو men وذلك بالنقر مرتين بالماوس في صندوق Test mean . أعط متوسط سرعة قيادة الرجال (الفرض العدمي) . μ = ٨٠ ادخل إذن 80 .

Test mean	80	OK

٣ - تظهر النتائج الآتية :

One-Sample T: men

```
Test of mu = 80 vs mu not = 80

Variable          N       Mean      StDev    SE Mean
men              16      94.56      10.05       2.51

Variable          95.0% CI                  T        P
men         (   89.21,    99.92)         5.80    0.000
```

وبفحص مخرجات برمجية Minitab نجد الآتي :

١ – عـــــدد المشاهـــــدات = ١٦ (N = 16)

٢ – متوسط Mean سرعة قيادة الرجال = ٩٤ كم

٣ – الانحراف المعياري StDev للسرعة = ١٠,٠٥

٤ – $ع_{\bar{س}}$ SE Mean = ٢,٥١ = $\frac{١٠,٠٥}{\sqrt{١٦}}$

٥ – فترة ثقة ٩٥ ٪ 95.0 % CI لمتوسط سرعة قيادة الرجال :

$$٨٩,٢١ \leq \mu \leq ٩٩,٩٢$$

٦ – قيمة اختبار T ت = ٥,٨٠

٧ – P = الإحتمال الشاهد ويمثل الاحتمال أو المساحة أكبر من قيمة ت المشاهدة .

∴ ح (ت > ٥,٨٠) = ٠٠٠٠

وإذا كانت P < ٥ ٪ نرفض الفرض العدمي ، وإذا كان أكبر من ٥ ٪ نقبل الفرض العدمي .

∴ **القرار** : رفض الفرض العدمى بأن السرعة تختلف عن ٨٠ كم/ ساعة .

ونستطيع إختبار الفرض العدمي باستخدام فترة الثقة كالتالى .

إذا إحتوت فترة الثقة القيمة تحت صحة الفرض العدمى نقبل الفرض العدمى .

وإذا كانت قيمة الفرض العدمي خارج نطاق فترة الثقة نرفض الفرض العدمى .

في المخرجات أعلاه ، فترة الثقة (٩٥ ٪) هي :

$$٨٩,٢١ \le \mu \le ٩٩,٩٢$$

ونجد أن : ف٠ : μ = ٨٠

حيث أن ٨٠ خارج نطاق فترة الثقة .

∴ نرفض الفرض العدمى .

وبنفس الأسلوب نجيب على المطلوب فى (٤) ، (٥) أعلاه ونجد أن المخرجات هى :

One-Sample T: women

```
Test of mu = 70 vs mu not = 70

Variable         N       Mean      StDev    SE Mean
women           14      85.07      13.61       3.64

Variable              95.0% CI              T        P
women          (   77.22,    92.93)       4.14   0.001
```

ويتبع نفس أسلوب التحليل أعلاه .

∴ فترة ثقة ٩٥ ٪

٧٧,٢٢ ≤ μ_2 ≤ ٩٢,٩٣

قيمة ت = ٤,١٤ ؛ P = ٠,٠٠١ وهي أقل من ٥ ٪

∴ نرفض الفرض العدمي بمستوى معنوية ٥ ٪

المطلوب في (٦):

لاختبار الفرق بين متوسطي سرعة قيادة الرجال وسرعة قيادة السيدات .

ف٠ : $\mu_1 - \mu_2$ = صفر

١ – ادخل البيانات في عمود C1 (men) ، C2 (women)

٢ – اختر

Stat → Basic Statistics
→ 2 - Sample t

٣ – في صندوق الحوار اختر

- Sample in different Columns

First [men]

Second [women]

[√] Assume equal Variances

أي أن التباين متساوي في المجتمعين ، ثم :

[OK]

يظهر المخرجات الآتية :

Two-Sample T-Test and CI: men, women

```
Two-sample T for men vs women

         N      Mean     StDev   SE Mean
men     16      94.6      10.0       2.5
women   14      85.1      13.6       3.6

Difference = mu men - mu women
Estimate for difference:  9.49
95% CI for difference: ([illegible], [illegible])
T-Test of difference = 0 (vs not =): T-Value = [illegible]  P-
Value = [illegible]  DF = [illegible]
```

ويفحص المخرجات أعلاه . نلاحظ المقاييس الوصفية لعينة الرجال وعينة السيدات

١ – متوسط سرعة قيادة الرجال = ٩٤,٦ كم/ساعة .

٢ – متوسط سرعة قيادة السيدات = ٨٥,١ كم/ساعة

٣ – الإنحراف المعياري (رجال) = ١٠,٠ كم/ساعة

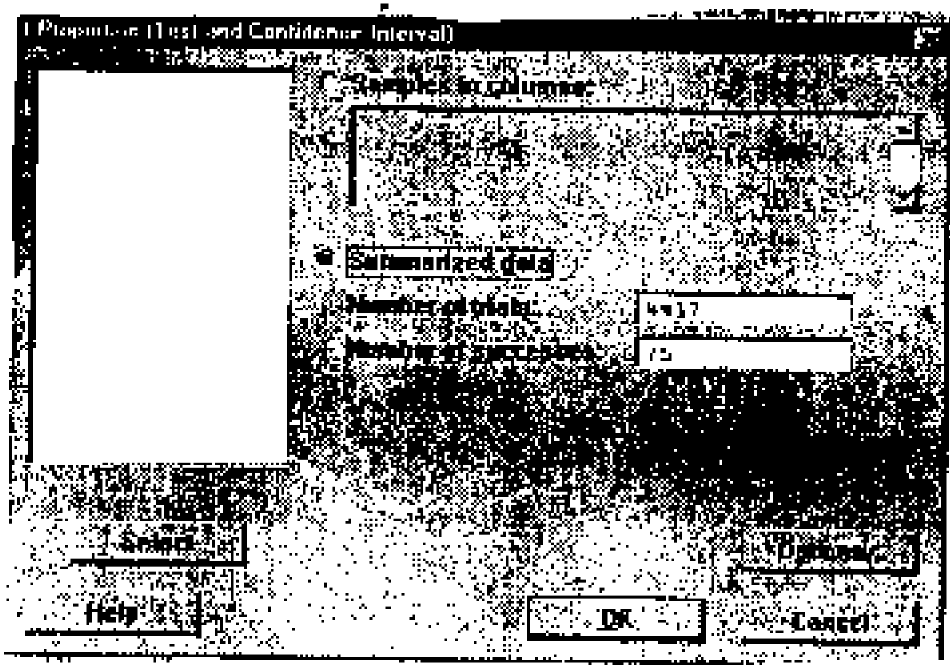
٤ – الإنحراف المعياري (سيدات) = ١٣,٦ كم/ساعة .

٥ – $ع_{\bar{س}}$ (رجال) = $\frac{ع}{\sqrt{ن}} = \frac{١٠,٠}{\sqrt{١٦}} = ٢,٥$

٦ – $ع_{\bar{س}}$ (سيدات) = $\frac{ع}{\sqrt{ن}} = \frac{١٣,٦}{\sqrt{١٤}} = ٣,٦$

٢ تقدير النسبة [illegible]

[illegible]



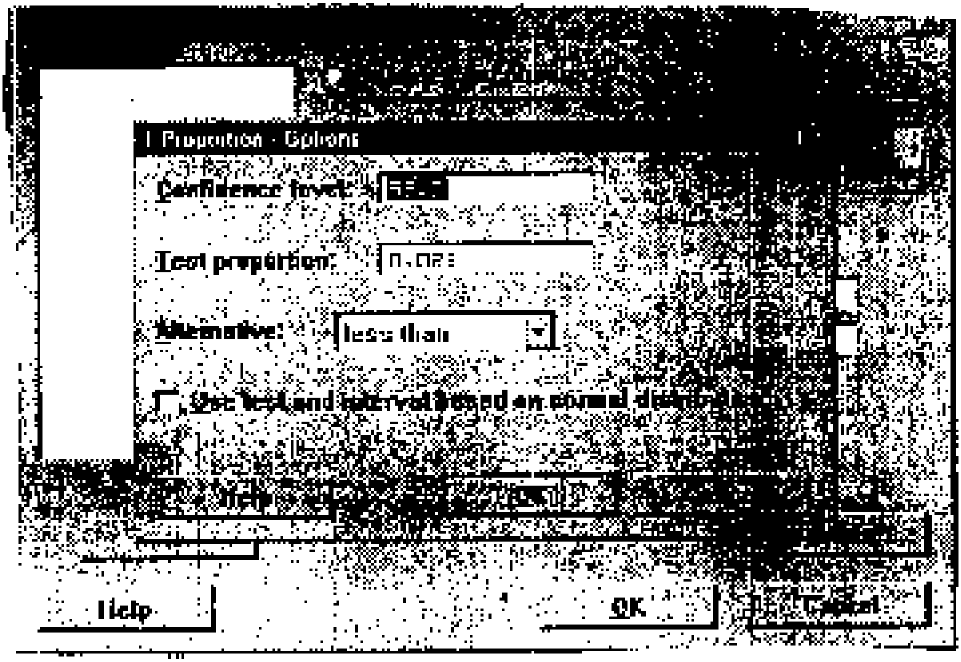
حيث : عـــدد المحاولات Number of Trial = عدد المشاهـدات = (عدد الطرود) .

عدد مرات النجاح Number of Successes = عدد المفردات التي ينطبق عليهم الظاهرة

-- عدد الطرود التالفة

* إذا تم النقر على OK تعطى المخرجات فترة الثقة (٩٥ ٪)

* إذا تم النقر على Options يظهر الصندوق التالى .



حيث : ندخل مستوى الثقة (٩٥ ٪)

وقيمة ل . = ٠.٠٢٣

والفرض البديل > less than

[illegible] OK ثم OK تظهر المخرجات التالية :

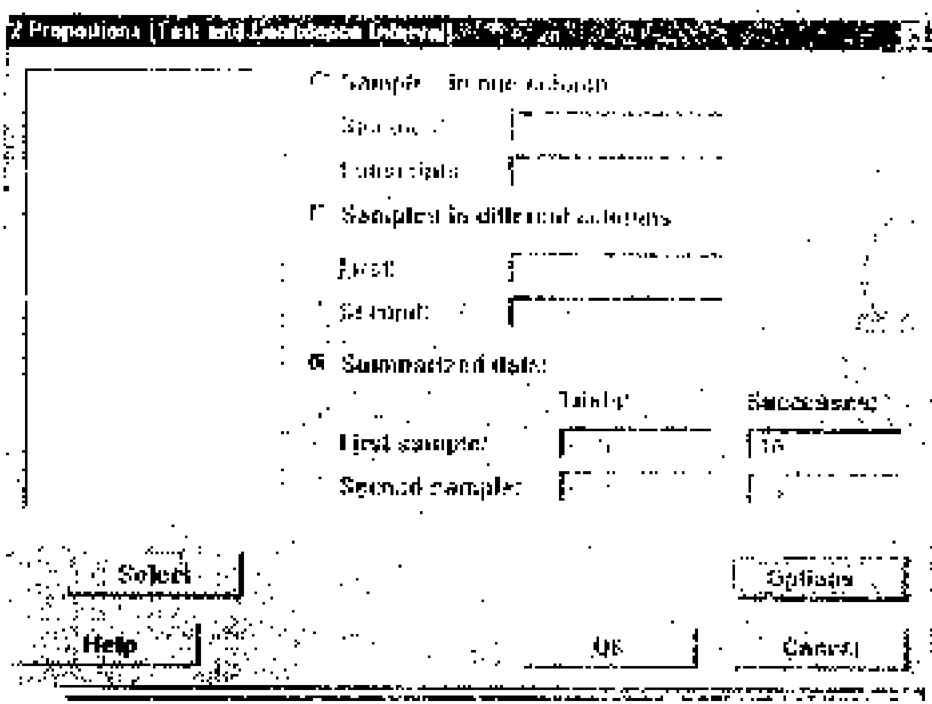
Test and CI for One Proportion

Test of p = 0.023 vs p < 0.023

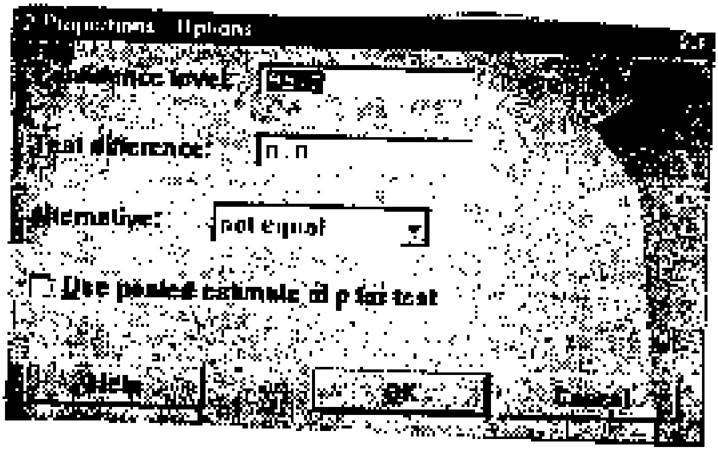
Sample	X	N	Sample p	95.0% Upper Bound	Exact P-Value
	[illegible]	[illegible]	[illegible]	0.020539	0.003

تقدير واختبار الفرق بين نسبتين

Stat → Basic Statistics



* اختر **Options** يظهر صندوق الحوار التالي .



حيث (١) مستوى الثقة = 95.0

(٢) Test difference = الفرق تحت صحة الفرض العدمي

$ل_١ - ل_٢ =$ صفر

(٣) الفرض البديل Alternative هو لا يساوي not equal .

* انقر على **OK** يظهر صندوق الحوار السابق ، وبنقر على **OK** تظهر المخرجات الآتية .

Test and CI for Two Proportions

Sample	X	N	Sample p
1	[illegible]	[illegible]	[illegible]
2	[illegible]	[illegible]	[illegible]

[illegible]
[illegible]
[illegible] P-Value = 0.002

- حيث P [Sample] = ل^ ∴ $\hat{ل}_1$ = ١٩٠, ، $\hat{ل}_2$ = ٣٢٣,
- $\hat{ل}_1$ − $\hat{ل}_2$ = التقدير غير المتحيز للفرق بين النسبتين في المجتمعين = ١٩, − ٣٢٣, = − ١٣٣,
- فترة ثقة ٩٥ ٪ للفرق بين النسبتين .

  − ٠٥, ≤ ل١ − ل٢ ≤ − ٢١٥,
- الاختبار الإحصائي (ى) أو Z = − ٣,١٦
- الاحتمال المشاهد = ٠٠٢,
- لاختبار : ل١ − ل٢ = صفر ضد ل١ − ل٢ ≠ صفر

(أ) إذا إحتوت فترة الثقة على الصفر نقبل الفرض العدمي .

وحيث أن فترة الثقة لا تحوى الصفر ، إذن نرفض الفرض العدمي .

(ب) الاحتمال المشاهد = ٠٠٢, > ٠٥,

∴ نرفض الفرض العدمي

(جـ) تقارن قيمة Z بالقيم ± ١,٩٦ حيث أن |Z| < |١,٩٦|

∴ نرفض الفرض العدمي .

الباب الرابع

تحليل التباين

ANALYSIS OF VARIANCE

الباب الرابع

تحليل التبايـــن

ANALSIS OF VARIANCE

فى هذا الباب سوف نعرض إمتداداً للباب السابق ، وهو كيفية إختبار تساوى عدة متوسطات من مجتمعات مستقلة . وهو إمتداد لاختبار تساوى متوسطى مجتمعين مستقلين ، مثلا إختبار هل يتساوى متوسط الإنفاق فى محافظات الوجه البحرى ؟ هل يتساوى عدد ساعات الإستذكار لطلبه ٦ كليات مختلفة ؟ هل يتساوى متوسطات عدد الكيلو مترات لكل صفيحه بنزين لأربعة انواع مختلفة من السيارات ؟ ويسمى الأسلوب المستخدم للتحليل أعلاه بأسلوب تحليل التباين .

والإختبارات أعلاه تعتمد على الإختبار الإحصائى ف ، أو "F" وهذا الإختبار "F" نسبه إلى الإنجليزى رائد الإحصائى فى العالم Sir R. A. Fisher (١٨٩٠ - ١٩٦٢) . والإختبار الإحصائى F يتبع نفس منهج الإختبارات السابقة وهى مقارنة قيمة F المحسوبة بقيمة F الجدولية .

١/٤إختبار تساوى متوسطين أو أكثر : تحليل التباين (ANOVA)

هل متوسط المبيعات اليومية فى مطاعم الوجبات السريعة فى خمسة فروع متساوية أم لا ؟ هل متوسط زمن الإنتظار للوفاء بالطلبيات متساوى بالنسبة لأربعه موردين الذين نتعامل معهم ؟ هل متوسط المبيعات متساو ، إذا اتبعت أربعة طرق مختلفة للإعلان ؟ كل هذه أسئلة يُردّ عليها بعد إجراء إختبارات تحليل التباين .

ما هو أسلوب تحليل التباين ؟

تحليل التبايـن هو الأسلوب المستخدم لإختبار الفرض الـعدمي بتساوي عدة أوسطة حسابية .

وعند عرض أسلوب تحليل التباين ، نستخدم الرمز : س ر و ليمثل قيمة المفرده رقم ر في العينـة و . فيوجد عدة عينات ، كـل عينة تمـثل « مجمـوعة أو معالجة » معينة . فمثلا عينة تمـثل الفرع (أ) ، عينة تمثل الفرع (٢) ، وعينة تمثل الفـرع (٣) ، . . . ، (٤) ، . . . (٥) ، في مطاعم الوجبات الـسريعة . أو عينـة تمثل الزمن المـستغرق لـلمورد رقـم (١)، المورد رقم (٢)، المـورد رقم (٣). . . وهكذا :

وتمثل بيانات كل عينة في عمود ، كالآتي : مثلا :

العينة الأولى	العينة الثانية	العينة الثالثة
س١١	س٢١	س٣١
س١٢	س٢٢	س٣٢
س١٣	س٢٣	س٣٣
س١٤	س٢٤	س٣٤
مجـ س ر١	مجـ س ر٢	مجـ س ر٣

وهذه العينات يطلق عليها «معالجات Treatments»

وتكون خطوات إستخدام أسلوب تحليل التباين كالآتى :

١ - صياغة الفروض :

الفرض العدمي ف. : $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_م = \mu$

الفرض العدمي ف١ : المتوسطات ليست كلها متساوية .

بمعنى وجود على الأقل متوسطان مختلفان .

٢- يتم سحب عدد (م) عينة مستقلة . ونحسب الأوسطة الحسابية لكل عينة :

$$\bar{س}_1 \quad ، \quad \bar{س}_2 \quad ، \quad \ldots \quad ، \quad \bar{س}_م$$

حيث :

$$\bar{س}_1 = \frac{مجـ\ س_1}{ن_1} \quad ، \quad \bar{س}_2 = \frac{مجـ\ س_2}{ن_2} \quad ، \quad \bar{س}_م = \frac{مجـ\ س_م}{ن_م}$$

٣- يحدد مستوى المعنوية α . ونحدد درجات حرية البسط ودرجات حرية المقام كالآتى :

درجة حرية البسط = (عدد المعالجات - ١) = م - ١

درجة حرية المقام = (عدد المشاهدات الكلية - عدد المعالجات)

= مجـ ن - م

نحدد المنطقة الحرجة باستخدام جدول توزيع ف .

∴ توجد ف ج بدرجات حرية (م - ١ ، (مجـ ن - م))

ويلاحظ أن ∝ هي مساحة الذيل الأيمن فى توزيع ف . ومع أن الاختبار في الجانبين ، إلا أنه في أسلوب تحليل التباين يوضع الخطأ من النوع الأول (∝) جميعه في الطرف الأيمن من التوزيع

٤ - نحسب مجموع المربعات «بين» المعالجات :

$$\text{مجموع المربعات « بين »} = \text{مجـ}\ ن_و \left(\bar{س}_و - \bar{\bar{س}} \right)^{٢}$$

$$\text{حيث } \bar{\bar{س}} \text{ هو الوسط الحسابى العام } \therefore \quad \frac{\text{مجـ مجـ } س_{ر\,و}}{\text{مجـ } ن_و} = \frac{\text{مجـ } ن_و\ \bar{س}_و}{\text{مجـ } ن_و}$$

٥ - نحسب مجموع المربعات « داخل » المعالجات :

$$\text{مجموع المربعات « داخل »} = \text{مجـ}_و \left(ن_و - ١ \right) ع^{٢}_و$$

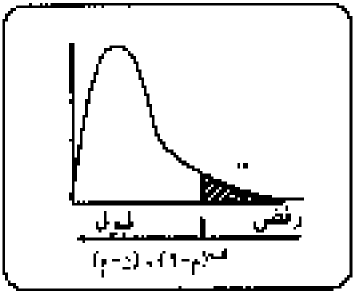
$$\text{حيث ، } ع^{٢}_و \text{ هو تباين العينة رقم و}$$

$$ن_و = \text{حجم العينة رقم و}$$

وهذا المجموع = مجموع المربعات «داخل» العينة الأولى + مجموعات المربعات «داخل» العينة الثانية + ... + مجموع المربعات «داخل» العينة م

$$\text{٦ - نحسب التباين « بين » المعالجات} = \frac{\text{مجموع المربعات « بين » المعالجات}}{م - ١}$$

٧ - نحسب التباين «داخل» المعالجات = $\dfrac{\text{مجموع المربعات «داخل» المعالجات}}{\text{مجـ}\ ن_{و} - م}$



٨ - نحسب الاختبار الإحصائي :

$$ف = \dfrac{\text{التباين «بين» المعالجات}}{\text{التباين «داخل» المعالجات}}$$

٩ - نصل إلى قرار بشأن صحة الفرض العدمي كالآتي :

نرفض الفرض العدمي إذا كانت قيمة ف المحسوبة ≥ ف الجدولية (الخطوة ٣) .

تقسيم الاختلافات الكلية

يمكن النظر إلى الخطوة (٤) ، (٥) كتحليل للاختلافات الكلية في البيانات كالآتي :

$$\text{مجـ مجـ}\ (س_{رو} - \bar{\bar{س}})^{2} = \text{مجموع الاختلافات الكلية}$$

= مجموع مربعات انحرافات القيم عن الوسط الحسابي العام .

$$= \text{مجـ مجـ}\ [\,س_{رو} - \bar{س}_{و} + \bar{س}_{و} - \bar{\bar{س}}\,]^{2}$$

$$= \text{مجـ مجـ}\ (س_{رو} - \bar{س}_{و})^{2} + \text{مجـ مجـ}\ (\bar{س}_{و} - \bar{\bar{س}})^{2}$$

$$+\ ٢\ \text{مجـ مجـ}\ (س_{رو} - \bar{س}_{و})\,(\bar{س}_{و} - \bar{\bar{س}})$$

= مجـ مجـ (سـرو − سـر)^2 + مجـ نو (سـر − سـ)^2

+ ٢ مجـ (سـر − سـو) مجـ (سـرو − سـر)

= مجموع المربعات «داخل» المعالجات + مجـموع المربعات «بين» المعالجات

حيث أن . ٢ مجـ (سـر − سـ) مجـ (سـرو − سـر)

= ٢ (صفر) مجـ (سـرو − سـر) = صفراً

∴ مجموع الاختلافات الكلية = مجموع الاختلافات «بين»

+ مجموع الاختلافات «داخل»

تقسيم درجات الحرية الكلية

وبالمثل ، نجد أن مجموع درجة الحرية للمجموع Total ينقسم كالآتي

مجموع درجات حرية المجموع = درجة حرية «بين» المعالجات + درجة حرية «داخل» المعالجات

ن − ١ = م − ١ + (ن − م)

حيث ن = مجـ نو ، م = عدد المعالجات

ونلاحظ الآتي أيضا:

١ - عند قسمة مجموع المربعات «بين» المعالجات على درجات الحرية المصاحبة ينتج تباين «بين» المعالجات . ويسمى «متوسط المربعات بين المعالجات» .

٢ - عند قسمة مجموع المربعات «داخل» المعالجات ، وتسمى مجموع مربعات الأخطاء (فى معظم الأحوال) على درجات الحرية المصاحبة ، ينتج تباين «داخل» المعالجات ، أو تباين الأخطاء ، ويسمى «متوسط» مربعات الأخطاء .

٣ - حيث أن قسمة تباينين مستقلين يتبع توزيع ف الاحتمالي

$$\therefore \text{ف} = \frac{\text{متوسط المربعات «بين» المعالجات}}{\text{متوسط مربعات «داخل» المعالجات}}$$

بدرجات حرية (البسط = م - ١ ، المقام = ن - م)

ونلخص ما سبق فى الجدول التالي :

جدول تحليل التباين في اتجاه واحد

عدد المعالجات = م ، $ن = ن_1 + ن_2 + \dots + ن_م$

مصدر الاختلاف	مجموع المربعات	درجات الحرية	متوسط المربعات	ف
بين المعالجات	$\text{مجـ } ن_ر (\bar{س}_ر - \bar{\bar{س}})^2$	م - ١	$\frac{\text{مجموع المربعات بين المعالجات}}{\text{م - ١}}$	$\frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط المربعات داخل المعالجات}}$
داخل المعالجات	$\text{مجـ مجـ } (س - \bar{س}_ر)^2$	ن - م	$\frac{\text{مجموع المربعات داخل المعالجات}}{\text{ن - م}}$	
الكلى	$\text{مجـ مجـ } (س - \bar{\bar{س}})^2$	ن - ١	$\frac{\text{مجموع المربعات الكلى}}{\text{ن - ١}}$	

تطبيق (٤-١)

في تجربة لتحديد أثر برامج معالجات النصوص Word Processors . اختيرت ثلاثة برامج مختلفة واختير ٢٤ سكرتيرة قسموا بطريقة عشوائية إلى ثلاث مجموعات ، ثم وزع أحد البرامج الثلاثة ، بطريقة عشوائية إلى كل مجموعة . ثم طلب من كل سكرتيرة كتابة نص مقالة معينة (أعطي نفس النص جميع المجموعات) ، ثم سجلت المدة المستغرقة بالدقيقة لكتابة هذا النص لكل سكرتيرة ، وحسب متوسط الكلمات في الدقيقة الواحدة .

وكانت النتائج كالتالي :

نوع البرنامج

١	٢	٣
٤٤	٤٠	٥٤
٣٩	٣٧	٥٠
٣٣	٣٨	٤٩
٥٦	٥٣	٥٥
٤٣	٣٨	٤٥
٤٧	٤٥	٤٩
٥٦	٤١	٦٦
٥٨	٦٠	٥٦

هل تعتقد أن سرعة الكتابة تتأثر بنوع البرنامج أم لا ، استخدم مستوى ثقة ٥ ٪

الحل :

المطلوب هو اختبار الفرض العدمي البديل بأن متوسط السرعة يختلف باختلاف نوع البرنامج أي أن .

ف٠ : $\mu_1 = \mu_2 = \mu_3 = \mu$

ف١ : جميع المتوسطات ليست متساوية

$$\text{الاختبار الإحصائي ف} = \frac{\text{متوسط المربعات بين المعالجات}}{\text{متوسط المربعات داخل المعالجات}}$$

، ف يتبع التوزيع الاحتمالي للمتغير ف بدرجات حرية = (م - ١) ، (ن - م) ،

ولحساب الاختبار الإحصائي يلزم حساب :

(١) مجموع المربعات بين المعالجات = مجـ ن ر (سَ ر - سَ=)

مجـ س١ = ٤٤ + ٣٩ + ... + ٥٨ = ٣٧٦

ن١ = ٨ ∴ سَ١ = $\frac{٣٧٦}{٨}$ = ٤٨

مجـ س٢ = ٤٠ + ٣٧ + ... + ٦٠ = ٣٥٢

ن٢ = ٨ ∴ سَ٢ = $\frac{٣٥٢}{٨}$ = ٤٤

مجـ س٣ = ٥٤ + ٥٠ + ... + ٥٦ = ٤٢٤

ن٣ = ٨ ∴ سَ٣ = $\frac{٤٢٤}{٨}$ = ٥٣

مجـ مجـ س ر و = ٣٧٦ + ٣٥٢ + ٤٢٤ = ١١٥٢

ن = ن$_{١}$ + ن$_{٢}$ + ن$_{٣}$ = ٢٤

$$\therefore \bar{\bar{\text{س}}} = \frac{١١٥٢}{٢٤} = ٤٨$$

(١) مجموع المربعات بين المعالجات = مجـ ن$_{ر}$ ($\bar{\text{س}}_{ر}$ − $\bar{\bar{\text{س}}}$)^{٢}

= ن$_{١}$ ($\bar{\text{س}}_{١}$ − $\bar{\bar{\text{س}}}$)^{٢} + ن$_{٢}$ ($\bar{\text{س}}_{٢}$ − $\bar{\bar{\text{س}}}$)^{٢} + ن$_{٣}$ ($\bar{\text{س}}_{٣}$ − $\bar{\bar{\text{س}}}$)^{٢}

حيث ن$_{١}$ = ن$_{٢}$ = ن$_{٣}$ = ٨

$\bar{\text{س}}_{١}$ = ٤٧ ، $\bar{\text{س}}_{٢}$ = ٤٤ ، $\bar{\text{س}}_{٣}$ = ٥٣ ، $\bar{\bar{\text{س}}}$ = ٤٨

∴ مجموع المربعات بين المعالجات = ٨ (٤٧ − ٤٨)^{٢} + ٨ (٤٤ − ٤٨)^{٢} + ٨ (٥٣ − ٤٨)^{٢}

= ٨ { ١ + ١٦ + ٢٥ }

= ٨ × ٤٢ = ٣٣٦

(٢) مجموع المربعات داخل المعالجات = مجـ مجـ (س ر و − $\bar{\text{س}}_{ر}$)^{٢}

= مجـ$_{و}$ (س$_{١و}$ − $\bar{\text{س}}_{١}$)^{٢} + مجـ$_{و}$ (س$_{٢و}$ − $\bar{\text{س}}_{٢}$)^{٢} + مجـ$_{و}$ (س$_{٣و}$ − $\bar{\text{س}}_{٣}$)^{٢}

= مجموع المربعات داخل المعالجة الأولى + مجموع المربعات داخل المعالجة الثانية + مجموع المربعات داخل المعالجة الثالثة

حيث مجموع المربعات داخل المعالجة = $\text{مجـ} \left(\text{س}_{ر} - \overline{\text{س}}_{ر} \right)^{٢}$

مجموع المربعات داخل المعالجة الأولى = $\text{مجـ} \left(\text{س}_{١} - \overline{\text{س}}_{١} \right)^{٢}$

$$= \text{مجـ س}^{٢} - \frac{(\text{مجـ س})^{٢}}{\text{ن}}$$

$$\text{مجـ س}_{١}^{٢} = ٤٤^{٢} + ٣٩^{٢} + \ldots + ٥٨^{٢} = ١٨٢٤٠$$

$$\text{مجـ س}_{١} = ٣٧٦$$

$$\therefore \text{مجموع المربعات داخل المعالجة الأولى} = ١٨٢٤٠ - \frac{(٣٧٦)^{٢}}{٨}$$

$$= ٥٦٨$$

وبالمثل نجد أن

$$\text{مجـ س}_{٢}^{٢} = ١٥٩٧٢ \quad , \quad \text{مجـ س}_{٢} = ٣٥٢$$

$$\therefore \text{مجموع المربعات داخل المعالجة الثانية} = ١٥٩٧٢ - \frac{(٣٥٢)^{٢}}{٨} = ٤٨٤$$

$$، \text{مجـ س}_{٣}^{٢} = ٢٢٧٦٠ \quad \text{مجـ س}_{٣} = ٤٢٤$$

$$\therefore \text{مجموع المربعات داخل المعالجة الثالثة} = ٢٢٧٦٠ - \frac{(٤٢٤)^{٢}}{٨} = ٢٨٨$$

$$\therefore \text{مجموع المربعات داخل المعالجات} = ٥٦٨ + ٤٨٤ + ٢٨٨ = ١٣٤٠$$

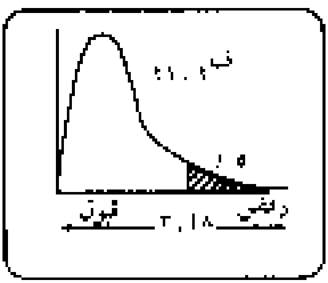
ونلخص الحسابات السابقة في جدول تحليل التباين الآتي :

جدول تحليل التباين لاختبار سرعة الكتابة (تطبيق (٤-١))

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	ف
بين المعالجات	٢	٣٣٦	١٦٨	٢،٦٣
داخل المعالجات	٢١	١٣٤٠	٦٣،٨١	
المجموع	٢٣	١٦٧٦		

∴ المنطقة الحرجة هي المنطقة ≤ ٣،٤٨

وحيث أن : ف المحسوبة أقل من ف الجدولية ∴ لا نرفض الفرض العدمي ، ونستطيع القول أن سرعة الكتابة لا تتأثر بنوع البرنامج المستخدم .



بمستوى معنوية α = ٥٪ ، القيمة الحرجة هي قيمة ف ، بدرجات حرية (٢ ، ٢١) = ٣،٤٨

تطبيق (٤-٢)

بفرض أنك قمت بتعيين ٨ مستخدمين جدد في قسم الحاسب الآلي . وقد تم تدريب هؤلاء المستخدمين ، وأن خطتك إعطاء حوافز مالية ومكافأة إلى المدرب الذي نجح في تدريب هؤلاء المستخدمين الجدد

ويقاس نجاح التدريب ، بإنتاجية المتدربين ، وقيست الإنتاجية بعدد الصفحات التي يستطيع المتدرب إدخالها على برنامج Word 7 .

تكون الفروض إذن :

$ف_0 : \mu_1 = \mu_2 = \mu_3$

$ف_1$: المتوسطات ليست جميعها متساوية .

حيث μ_1 ، μ_2 ، μ_3 هو متوسط إنتاجية المتدربين وهو متوسط عدد الصفحات التي يعطيها المتدربين للمدرب الأول والثاني والثالث على التوالي.

تم الحصول على البيانات الآتية من المتدربين الثمانية ($ن_1$ = ٢ ، $ن_2$ = ٣ ، $ن_3$ = ٣) .

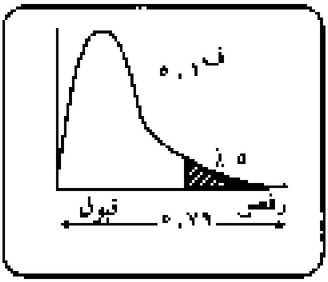
	المدرب الأول	المدرب الثاني	المدرب الثالث
	١٢	١٩	١٩
	١٧	٢١	٢٠
		٢٠	٢٤
مجموع عدد الصفحات	٢٩	٦٠	٦٣
متوسط عدد الصفحات	١٤,٥	٢٠	٢١

∴ $\bar{س}_1$ = ١٤,٥ صفحة ، $\bar{س}_2$ = ٢٠ صفحة

$\bar{س}_3$ = ٢١ صفحة

حيث أن متوسطات العينة غير متساوية ، إذن نستمر فى الإختبار . ونطبق اسلوب تحليل التباين .

$$\text{الإختبار الإحصائى هو إختبار ف} = \frac{\text{متوسط المربعات "بين"}}{\text{متوسط المربعات " داخل "}}$$



حيث ف يتبع التوزيع الإحتمالى ف بدرجات حرية ((م - ١) ، (ن - م)) = (٣ - ١ ، ٨ - ٣) = (٢ ، ٥) ، وبمستوى معنوية ٥٪ نجد أن $ف_{٠,٠٥}$ (٢ ، ٥) = ٥,٧٩

نحسب متوسط المربعات "بين" = $\text{مجـ}\ ن_و\ (\bar{س}_و - \bar{\bar{س}})^٢$

$\bar{س}_١ = ١٤,٥$ $\qquad$ $\bar{س}_٢ = ٢$ $\qquad$ $\bar{س}_٣ = ٢١$

$$\bar{\bar{س}} = \frac{٢٩ + ٦٠ + ٦٣}{٢ + ٣ + ٣} = ١٩$$

$\therefore$ $\text{مجـ}\ ن_و\ (\bar{س}_و - \bar{\bar{س}})^٢ = ٢\ (١٤,٥ - ١٩)^٢ + ٣\ (٢٠ - ١٩)^٢$

$+ ٣\ (٢١ - ١٩)^٢ = ٥٥,٥$

مجموع المربعات "داخل" المعالجات = مجموع مربعات المعالجة الأولى

\+ مجموع مربعات المعالجة الثانية

\+ مجموع مربعات المعالجة الثالثة

مجموع المربعات «داخل» المعالجة الأولى = مجـ (س $_1$ − $\overline{\text{س}}_1$)^2

$$= \text{مجـ س}_1^2 - \frac{(\text{مجـ س}_1)^2}{\text{ن}_1}$$

مجـ س$_1^2$ = ١٢^2 + ١٧^2 = ٤٣٣ ، مجـ س$_1$ = ٢٩

∴ مجموع المربعات «داخل» المعالجة الأولى = ٤٣٣ − $\frac{(٢٩)^2}{٢}$ = ١٢,٥

وبالمثل نجد أن : مجـ س$_2^2$ = ١٢٠٢ مجـ س$_2$ = ٦٠

∴ مجموع المربعات «داخل» المعالجة الثانية = ١٢٠٢ − $\frac{(٦٠)^2}{٣}$ = ٢

، مجـ س$_3^2$ = ١٣٣٧ مجـ س$_3$ = ٦٣

∴ مجموع المربعات «داخل» المعالجة الثالثة = ١٣٣٧ − $\frac{(٦٣)^2}{٣}$ = ١٤

∴ مجموع المربعات داخل المعالجات = ١٢,٥ + ٢ + ١٤ = ٢٨,٥

ونصل إلى جدول تحليل التباين التالي :

جدول تحليل التباين لاختبار الإنتاجية (متوسط عدد الصفحات)

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	ف
بين المعالجات	٥٥,٥	٢	٢٧,٧٥	٤,٨٧
داخل المعالجات	٢٥,٥	٥	٥,٧	
الكلي	٨٤	٧		

∴ ف المحسوبة = $\frac{٢٧,٧٥}{٥,٧}$ = ٤,٨٧

وحيث أن ف المحسوبة (٤,٨٧) أقل من ف$_{ج}$ (٢ ، ٥) = ٥,٧٩

∴ لا نرفض الفرض العدمي ، ونستطيع القول أن الإنتاجية متساوية ويستحق المدربون مكافأة متساوية .

٢/٤ تحليل التباين : مقارنة متوسطي مجتمعين

في الباب السابق تم عرض كيفية إختبار تساوي أو عدم تساوي متوسطين في مجتمعين مستقلين – وكان الفرض العدمي والفرض البديل كالآتي :

ف$_{.}$: $\mu_{1} = \mu_{2}$

ف$_{1}$: $\mu_{1} \neq \mu_{2}$

والإختبار الإحصائي الذي تم عرضه :

$$\text{ت} = \frac{\bar{س}_{1} - \bar{س}_{2}}{ع_{\bar{س}_{1} - \bar{س}_{2}}} \quad .... \quad (1)$$

حيث $ع_{\bar{س}_{1} - \bar{س}_{2}} = \sqrt{ع^{2}_{م} \left(\frac{1}{ن_{1}} + \frac{1}{ن_{2}} \right)}$

، $ع^{2}_{م}$ هو التباين المجمع ، تحت صحة فرض تجانس التباين .

وفي هذا الباب ، ومع عرض أسلوب تحليل التباين ، نجد أن الفرض العدمي والفرض البديل هما

ف. : $\mu_1 = \mu_2$

ف. : $\mu_1 \neq \mu_2$

ويكون الإختبار الإحصائى :

$$ف = \frac{\text{متوسط المربعات بين م المعالجات}}{\text{متوسط المربعات داخل المعالجات}} \quad \ldots \quad (٢)$$

ف يتبع التوزيع الاحتمالى ف بدرجـات حريـة (٢ - ١ ، ن - ٢) = (١ ، ن - ٢) حيث م فى هذه الحالة الخاصة = ٢

ما العلاقة إذن بين الإختبارين (١) ، (٢) ؟ نستعرض هذه العلاقة من خلال المثال التالى .

تطبيق (٤-٣)

فى تطبيق (٤ - ٢) بفرض وجود مدربين فقط هما المدرب رقم (٢) ، المدرب رقم (٣) ، نجد أن إنتاجية المدربين هما :

مدرب (٢)	مدرب (٣)
١٩	١٩
٢١	٢٠
٢٠	٢٤
٦٠	٦٣

$\therefore \bar{س}_1 = \frac{٦٠}{٣} = ٢٠$ $\qquad$ $\bar{س}_2 = \frac{٦٣}{٣} = ٢١$

وحيث أن $\bar{س}_1 \neq \bar{س}_2$ $\therefore$ نستمر فى الإختبار ، إما باستخدام إختبار ت أو باستخدام أسلوب تحليل التباين واختبار ف .

أولاً: إختبار ت

باستخدام المعادلة (١) أعلاه :

$$\text{ت} = \frac{(\bar{\text{س}}_1 - \bar{\text{س}}_2) - (\mu_1 - \mu_2)}{\text{ع}_{\bar{\text{س}}_1 - \bar{\text{س}}_2}}$$

$$\text{ع}_{\bar{\text{س}}_1 - \bar{\text{س}}_2} = \sqrt{\text{ع}_{\text{م}}^2 \left(\frac{1}{\text{ن}_1} + \frac{1}{\text{ن}_2} \right)}$$

$$\text{ع}_{\text{م}}^2 = \frac{(\text{ن}_1 - 1)\,\text{ع}_1^2 + (\text{ن}_2 - 1)\,\text{ع}_2^2}{\text{ن}_1 + \text{ن}_2 - 2}$$

$$= \frac{\text{مجـ}(\text{س}_1 - \bar{\text{س}}_1)^2 + \text{مجـ}(\text{س}_2 - \bar{\text{س}}_2)^2}{\text{ن}_1 + \text{ن}_2 - 2}$$

$\text{مجـ}(\text{س}_1 - \bar{\text{س}}_1)^2$ = مجموع المربعات داخل المعالجة الأولى

$$= \text{مجـ}\,\text{س}_1^2 - \frac{(\text{مجـ}\,\text{س}_1)^2}{\text{ن}_1}$$

$$= ١٢٠٢ - \frac{(٦٠)^2}{٣} = ٢$$

$\text{مجـ}(\text{س}_2 - \bar{\text{س}}_2)^2$ = مجموع المربعات داخل المعالجة الثانية

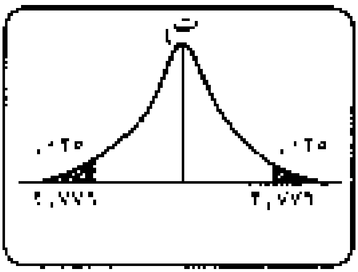
$$= \text{مجـ}\,\text{س}_2^2 - \frac{(\text{مجـ}\,\text{س}_2)^2}{\text{ن}_2}$$

$$= ١٣٣٧ - \frac{(٦٣)^2}{٣} = ١٤$$

$$\therefore \quad ع_م^{٢} = \frac{٢ + ١٤}{٣ + ٣ - ٢} = ٤$$

$$\therefore \quad ع_{\bar{س}_١ - \bar{س}_٢} = \sqrt{٤ \left(\frac{١}{٣} + \frac{١}{٣} \right)} = ١,٦٣$$

$$\therefore \quad ت = \frac{(٢٠ - ٢١) - صفر}{١,٦٣} = ,٦١٣٥$$



المنطقة الحرجة هي المنطقة $\leq$ ت$_{٠,٠٢٥}$ بدرجات حرية = ٤ هي المساحة أكبر من ٢,٧٧٦ والمنطقة $\geq$ − ت$_{٠,٠٢٥}$ بدرجات حرية ٤ هي المساحة أقل من − ٢,٧٧٦

∴ القرار : نقبل الفرض العدمي بتساوي متوسطي المجتمع .

ثانيا : باستخدام أسلوب تحليل التباين

باستخدام المعادلة (٢) أعلاه :

$$ف = \frac{\text{متوسط المربعات "بين" المعالجات}}{\text{متوسط المربعات "داخل" المعالجات}}$$

مجموع المربعات بين المعالجات = $مجـ \; ن_و \, (\bar{س}_و - \bar{\bar{س}})^{٢}$

$$\bar{س}_١ = ٢٠ \qquad \bar{س}_٢ = ٢١ \qquad \bar{\bar{س}} = \frac{٦٠ + ٦٣}{٣ + ٣} = ٢٠,٥$$

∴ مجموع المربعات "بين" المعالجات =

$$٣ \, (٢٠ - ٢٠,٥)^{٢} + ٣ \, (٢١ - ٢٠,٥)^{٢} = ١,٥$$

∴ مجموع المربعات «داخل» المعالجات = مجموع المربعات داخل المعالجة الأولى
+ مجموع المربعات داخل المعالجة الثانية

مجموع المربعات «داخل» المعالجة الأولى = مجـ $(س_١ - \bar{س}_١)^٢$

$$= ١٢٠٢ - \frac{(٦٠)^٢}{٣} = ٢$$

مجموع المربعات «داخل» المعالجة الثانية = مجـ $(س_٢ - \bar{س}_٢)^٢$

$$= ١٣٣٧ - \frac{(٦٣)^٢}{٣} = ١٤$$

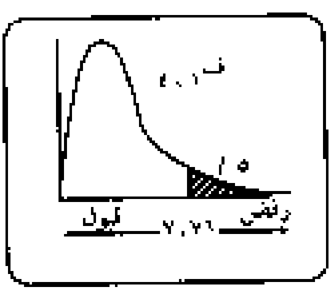
∴ مجموع المربعات «داخل» المعالجات = ٢ + ١٤ = ١٦

ونصل إلى جدول تحليل التباين الآتي :

جدول تحليل التباين للفرق بين متوسطي معالجتين

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	ف
بين المعالجات	١	١,٥	١,٥	٠,٣٧٥
داخل المعالجات	٤	١٦	٤	
الكلي	٥	١٧,٥		

وقيمة ف الحرجة بمستوى معنوية ٥٪ ودرجات حرية (١ ، ٤) = ٧,٧١ وحيث أن قيمة ف المحسوبة (٠,٣٧٥) < قيمة ف الحرجة (٧,٧١)



∴ لا يرفض الفرض العدمي ، ونستطيع القول بأن الإنتاجية واحدة بين المجموعتين .

وبعد عرض المثال السابق نصل إلى الحقائق التالية :

١ - قيمة ت المحسوبة = ٠,٦١٢٣٥ ، قيمة ف المحسوبة = ٠,٣٧٥

∴ ف المحسوبة = (ت المحسوبة)^{٢}

٢- قيمة ت$_{٠,٠٢٥}$ الجدولية (إختبار في اتجاهين) بدرجات حرية (٤) = ٢,٧٧٦

قيمة ف$_{٠,٠٥}$ بدرجات حرية (١ ، ٤) = ٧,٧١

∴ ف$_{٠,٠٥}$ بدرجات حرية (١ ، ٤) = (ت$_{٠,٠٢٥}$)^{٢} بدرجات حرية ٤ .

وينطبق هذا طالما أن درجات حرية ف في البسط = ١ ، ودرجات حرية ف في المقام = درجة حرية ت ، وذلك باستخدام مستوى معنوية ∝ في اختبار ف ، ومستوى معنوية $\frac{∝}{٢}$ في اختبار ت .

٣ - قيمة التباين المجمع في إختبار ت = ٤

متوسط المربعات «داخل» المعالجات في اختبار ف = ٤

∴ ع$^{٢}_{م}$ باستخدام المعادلة (١) = متوسط المربعات «داخل» المعالجات في تحليل التباين ، باستخدام المعادلة (٢) .

إذن ، عند إختبار تساوي متوسطين في مجتمعين مستقلين ، يمكن إتباع أسلوب تحليل التباين ، طالما أن الفرض البديل يصاحبه علامة ≠

٣/٤ : إستخدام برمجية Minitab في تحليل التباين

لاستخدام Minitab في تحليل التباين لعدد من المعالجات تقوم بالآتي (سوف نستخدم بيانات تطبيق (٤-١)) .

١ - إدخال بيانات كل عينة في عمود : C1 , C2 , C3 مثلاً .

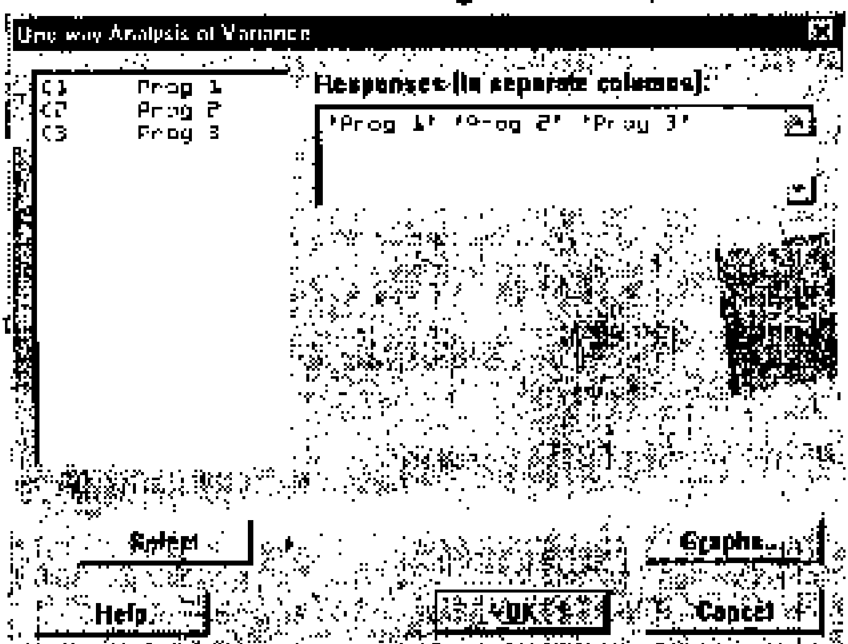
٢ - يفضل أن تعطى كل عينة إسماً مميزاً .

مثلاً عمود C1 يسمى Prog 1 ، عمود C2 يسمى Prog 2 ، عمود C3 يسمى Prog 3 .

٣ - من القائمة الرئيسية إختر :

Stat → ANOV → One - way (unstacked)

٤ - يظهر صندوق حوار : بالنقر مرتين على الأعمدة التي تم إدخال بيانات كل عينة فيها - ينتقل إسم العمود إلى مربع Response كالآتي :



وبالنقر على $\boxed{\text{OK}}$ تظهر المخرجات التالية :

One-way ANOVA: Prog 1, Prog 2, Prog 3

```
Analysis of Variance
Source     DF        SS        MS        F        P
Factor      2     336.0     168.0     2.63    0.095
Error      21    1340.0      63.8
Total      23    1676.0
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  --------+---------+---------+------
Prog 1      8    47.000     9.008        (--------*---------)
Prog 2      8    44.000     8.315    (--------*---------)
Prog 3      8    53.000     6.414               (---------*--------)
                                   --------+---------+---------+------
Pooled StDev =    7.988                  42.0      48.0      54.0
```

ومقارنة جدول تحليل التبايـن ، بالجدول الذي سبق حسابه وإعداده في تطبيق (٤-١) . نجد النتائج واحدة ، ونجد أيضاً أن : قيمة (P) هي الاحتمال المشاهد وقيمته من نتائج Minitab = ٠,٩٥ ، وحيث أن P = ٠,٩٥ > ∝ = (٠,٠٥) ∴ نقبل الفرض العدمي ، ويتفق هذا مع النتيجة التي سبق الحصول عليها .

ونلاحظ أن المخرجات تشمل أيضاً مؤشرات وصفية عن الثلاث معالجات

$ن_1 = 8 \quad \bar{س}_1 = ٤٧ \qquad ع_1 = ٩,٠٠٨$

$ن_2 = 8 \quad \bar{س}_2 = ٤٤ \qquad ع_2 = ٨,٣١٥$

$ن_3 = 8 \quad \bar{س}_3 = ٥٣ \qquad ع_3 = ٦,٤١٤$

وأيضاً : $ع_م$ Pooled StDev = $\sqrt{ع^2_م}$ = ٧,٩٨٨

ونجد أن $ع_م = \sqrt{\text{متوسط المربعات داخل المعالجات}}$ (Error) .

$$= \sqrt{٦٣,٨} = ٧,٩٨٨$$

ويوجد أيضاً في المخرجات فترات الثقة لمتوسط مجتمع كل معالجة من المعالجات بمستوى ثقة ٩٥ ٪ كالآتي :

لإنشاء فترة ثقة ، تستخدم المعادلة

$$\bar{س} \pm ت_{٠,٠٢٥} \times \frac{ع_م}{\sqrt{ن}}$$

، ت بدرجات حرية = درجات حرية الأخطاء (داخل المعالجات) = ٢١

∴ $ت_{٠,٠٢٥} = ٢,٠٨٠$ بدرجات حرية = ٢١ .

وفي Minitab ، يتم حساب فترة ثقة لمتوسط المعالجة «و» باستخدام المعادلة

$$\bar{س}_و \pm ٢,٠٨٠ \times \frac{ع_م}{\sqrt{ن_و}}$$

حيث $ع_م = \sqrt{\text{متوسط مربعات الأخطاء}}$

$$= \sqrt{٦٣,٨١} = ٧,٩٨٨$$

∴ فترة ثقة ٩٥ ٪ لمتوسط المجتمع الممثل بالعينة Prog 1 .

$$= ٤٧ \pm ٢,٠٨٠ \times \frac{٧,٩٨٨}{\sqrt{٨}}$$

= ٤٧ ± ٥,٨٧ = { ٤١,١٣ إلى ٥٢,٨٧ }

، فترة ثقة لمتوسط المجتمع الممثل بالعينة Prog 2 .

$$= ٤٤ \pm ٢,٠٨٠ \times \frac{٧,٩٨٨}{\sqrt{٨}}$$

= ٤٤ ± ٥,٨٧ = { ٣٨,١٣ إلى ٤٩,٨٧ }

، فترة ثقة لمتوسط المجتمع الممثل بالعينة Prog 3 .

$$= ٥٣ \pm ٢,٠٨٠ \times \frac{٧,٩٨٨}{\sqrt{٨}}$$

= ٥٣ ± ٥,٨٧ = { ٤٧,١٣ إلى ٥٨,٨٧ }

تمارين

١ - اعتقد باحث أن طريقة عرض البضائع ، ومدة رؤية العميل للبضائع داخل فتيرينات العرض تؤثر على الشراء .

قام الباحث بعرض أربعة فترينات بطرقة مختلفة ، وكان هدفه هو قياس الزمن الذى يقضيه العميل أمام كل فترينه ، فافترض ان طريقة العرض تؤثر على زمن الرؤية وزمن الرؤية يؤثر بطريقة طردية على الشراء

وكانت نتائج الباحث كالتالى (مقاسه بالدقيقة)

فترينه العرض

(١)	(٤)	(٤)	(٤)
١١	٨	٩	١٠
١٣	٨	١٢	١٥
١٠	١٤	١١	١٨
١٣	٩	١٠	
	١٧		

هل تعتقد أن متوسط زمن الوقوف أمام فترينه العرض تختلف من فترينه إلى أخرى ؟ إستخدام ∝ = ٥٪ .

٢- لوحظ فى أحد فروع الشركات المتعددة الجنسية ، أن العاملين بها يحاولون تغيير الفرع بعد فترة معينة . أراد مدير المستخدمين مقارنة مدة العمل قبل طلب تغيير الفرع ، فى أربعة فروع مختلفة . واعتقد أن تغيير الفرع ليس له علاقة بموقع الفرع . لذا فيعتقد أن متوسط مدة التوظف فى أى فرع لا

تختلف من فرع إلى آخر . لاحظ المدة قبل طلب تغيير الفرع وكانت النتائج كالتالي (بالشهر) .

الفرع (١)	الفرع (٢)	الفرع (٣)	الفرع (٤)
٢٨	١١	٢٠	٢٥
٨	٢٤	٩	٢٠
١٣	٨	٢	١٣
٢٠	١٧	١٥	٦

هل تتفق مع مدير إدارة المستخدمين بأن متوسط مدة التوظيف واحدة في الأربعة فروع ؟ استخدم α = ٥ ٪

٣- لاحظ مدير الإنتاج أن نسبة المعيب في إنتاج اللعبات الكهربائية تختلف من وردية إلى أخرى . فقام بعملية عد المعيب من اللعبات في كل ألف لعبة في ثلاث ورديات مختلفة . فكانت النتائج كالتالي .

الوردية (١)	الوردية (٢)	الوردية (٣)
١٥	١٨	٩
٢٠	٢٦	٢٩
١٨	٣٢	٣٥
٢١	١٤	٦

إستخدام α = ٥ ٪ هل تعتقد أن متوسط نسبة المعيب يختلف من وردية إلى أخرى ؟

٤- عادة ، ما يعمل مندوبوا المبيعات إما بمرتب ثابت أو بنسبه عمولة ثابتة او بمرتب ثابت ونسبة عمولة .

اراد مدير الشئون المالية بمؤسسة معينة مقارنة متوسط المبيعات بالثلاثة أساليب أعلاه . وهل تختلف المبيعات طبقا لطريقة التعامل المالي مع المندوب أم لا . سحب مدير الشئون المالية عينة عشوائية من مندوبي المبيعات ولاحظ أن عدد المبيعات لكل مندوب كانت كالتالي :

مرتب فقط	عمولة فقط	مرتب + عمولة
٢٨	٢٠	٢٤
٢٦	١٩	٢٣
٢٣	١٨	٢٠
٢٤	١٧	٢٠
٢٢	١٨	٢١
٢٤	١٩	٢٣
٢٥	٢٠	٢٢
٢٦	١٧	١٩
٢٢	١٩	٢٠
٢٣	١٩	٢٢

هل تتفق مع مدير الشئون المالية بأن عدد المبيعات تختلف باختلاف التعامل المالى مع المندوب ؟ إستخدم $\alpha = 5\%$. استخدم Minitab وعلق على النتائج.

٥- أراد مدير الإنتاج أن يعرف عما إذا كانت الإنتاجية مرتبطة بعدد سنوات

العمل في المصنع . وأعتقد أن الإنتاجية تزيد كلما زاد عدد سنوات العمل في المصنع . قام المدير بسحب عينة عشوائية من العاملين لأقل من عام ، والعاملون من (١-٥) أعوام والعاملين لمدة تزيد على خمسة أعوام . وتم قياس الإنتاجية في المجموعات الثلاثة وكانت كالآتي .

خمس سنوات فأكثر	من سنة - خمس سنوات	أقل من سنة
١١٠	٩٣	٨٢
٧٦	٧٣	٧٠
٩٠	٨٦	٧٨
١١٣	٩٥	٨٠
٧٨	٧٤	٧١
٨٩	٨٥	٧٧
١٠٨	٩٢	٧٩
٨٠	٧٤	٧٠
٩٠		٧٨
٩٥		٧٤

هل تعتقد أن متوسط الإنتاجية متساوي في الثلاث مجموعات ؟ استخدم Minitab وعلق على النتائج .

٦ - أرادت شركة تعمل في الوثائق التي يتم التعامل فيها في سوق المال ، معرفة اللازمة حتى سيتم إعداد الإصدار وتسليمه للمشترين . وأرادت أيضاً معرفة عما إذا كان الوقت المستغرق حتى إعداد الوثائق يختلف حسب قيمة الوثيقة أم لا .

تم سحب عينة عشوائية من سجلات الوثائق السابقة ، وكان الوقت المستغرق (بالساعة) لإعداد هذه الوثائق حسب حجم الوثيقة كالآتي :

وثائق صغيرة	وثائق متوسطه	وثائق كبيرة
٨٠	٩١	٩٦
٨٣	٩٢	٧٢
٧٨	٨١	٦٨

هل تعتقد أن متوسط الوقت المستغرق يختلف باختلاف حجم الوثيقة ؟

٧- مصنع مسحوق الغسيل الأوتوماتيك ، لديه ثلاثة ماكينات لتعبئة المسحوق فى عبوات ٥ كج . أراد المصنع شراء ماكينة إضافية ولكن مسئول الإنتاج ، رأى أنه يجب أولا معرفة هل يتساوى متوسط عدد العبوات فى الساعة الواحدة للماكينات الثلاث أم لا . وإذا إختلفت فما هى الماكينة التى تعطى متوسطا أعلى . سحب المسئول عينات عشوائية من إنتاج الثلاث ماكينات وكانت كالآتى :

الآلة (١)	الآلة (٢)	الآلة (٣)
٥٤	٥٣	٤٩
٤٩	٥٦	٥٣
٥٢	٥٧	٤٧
٥٥	٥١	٥٠
٤٨	٥٩	٥٤

(أ) هل تختلف الثلاث آلات من حيث متوسط عدد العبوات المعبأة كل ساعة ؟

(ب) قارن بين كل آلتين للوصول إلى الآلة التي تعطي متوسطاً أكبر .

(جـ) إعط النصح لمدير الإنتاج ، وأى نوع من الآلات يجب إضافة آلة رابعة منه .

٨- أعد تحليل التمرين (٧) باستخدام Minitab .

٩- التطبيق التالي يقارن بين متوسط أربعة مجموعات بياناتهم كالتالي :

المجموعة (١)	المجموعة (٢)	المجموعة (٣)	المجموعة (٤)
١٩	١٤	١١	٢٤
٢١	١٦	١٤	١٩
٢٦	١٤	٢١	٢١
٢٤	١٣	١٣	٢٦
١٨	١٧	١٦	٢٠
	١٣	١٨	

إستخدم أمر Minitab وإعط مخرجات البرنامج وعلق على النتائج .

الباب الخامس

إختبارات كـا^{٢}

Chi-Square Tests

الباب الخامس

إختبارات كا٢

Chi - Square Tests

في الأبواب السابقة ، تم عرض كيفية إختبار تساوي قيمة متوسط مجتمع واحد بقيمة معينة ، وكيفية إختبار تساوي متوسطي مجتمعين ، وكيفية إختبار تساوي متوسطات عدة مجتمعات مستقلة .

تم أيضًا عرض ، كيفية إختبار تساوي النسبة في مجتمع واحد بنسبة محددة مقدمًا ، وكيفية إختبار تساوي نسبتين من مجتمعين . واستخدمت نظرية الحد المركزي كتقريب للتوزيع الإحتمالي للإختبار الأول (ل = ل٠) ، والإختبار الثاني (ل١ − ل٢ = صفر) . وكان الشرط المطلوب هو أن يكون أحجام العينات كبيرًا (نر ≤ ٣٠ ، نر لَ̂ر ≤ ٥ ، نر (١ − لَ̂ر)) ، حيث (ر = رقم المجتمع = ١ ، ٢) .

سوف يكون هذا الباب إمتدادًا لإختبار تساوي نسبتين ، فسوف يكون الإهتمام بكيفية إختبار أكثر من نسبتين ، إما تساوي عدة نسب ، أو إختبار عما إذا كانت النسب في مجتمعات العينة تساوي قيمًا معينة . وفي هذا الباب أيضًا، نعرض كيفية إختبار عما إذا كان متغيرين مستقلين أم لا .

مثلاً هل معدلات الربح تختلف بإختلاف نوع نشاط المنشأة أم لا ؟ إذا كان معدلات الربح ونوع نشاط المنشأة مستقلين ، بمعنى عدم وجود علاقة بين المتغيرين ، نجد أن معدلات الربح لا تختلف بإختلاف نوع النشاط والعكس صحيح . وهذه الإختبارات تسمى إختبارات الإستقلال .

في جميع المواقف السابقة (تساوي النسب في عدة مجتمعات ، أو النسب في المجتمعات تساوي قيمة محددة لكل مجتمع ، أو إختبارات الإستقلال) يستخدم توزيع كـا^٢ . والخطوات التي تتبع في هذه الإختبارات ، هي نفس الخطوات السابق عرضها في الأبواب السابقة وهي :

١ - صياغة الفروض .

٢ - جمع بيانات عينة أو عدة عينات .

٣ - تحديد مستوى المعنوية ، وتحديد القيمة الحرجة باستخدام جدول توزيع كـا^٢ بالدرجات الحرية المحددة لكل إختبار ، والمنطقة الحرجة تقع دائماً على يمين القيمة الحرجة .

٤ - حساب قيمة الإختبار الإحصائي .

٥ - الوصول إلى قرار بشأن صحة الفرض العدمي ، وذلك بمقارنة قيمة الإختبار الإحصائي (الخطوة (٤)) بالمنطقة الحرجة (الخطوة (٥)) .

وإختبار ى ، وإختبار ت ، وإختبار ف اللاتي تم عرضهم في الأبواب السابقة ، لا تستخدم ولا تطبق إلا إذا كانت البيانات مستمرة Continuous ومقاسة بمقياس نسبي Ratio Scale . وهذا القياس النسبي يسمح بإجراء جميع العمليات الحسابية على المتغير ، وطبقاً لهذا المقياس أيضاً ، يكون المتغير س = ٤ له نصف قيمة س = ٨ .

ولكن ليست جميع البيانات مقاسة بهذا القياس . ففي دراستنا السابقة اعرضنا المقاييس الرمزية Nominal التي تقسم المتغير إلى مجموعات ، والترتيب ليس ذا أثر ، والمقاييس الترتيبية Ordinal والتي تقسم المتغير إلى مجموعات ،

والترتيب لـه أهمية . ويوجد المقاييس الفترية Interval ، وهي تسمح بقياس «البعد» بين كل قيمة وأخرى . ويوجـد المقاييس النسبية Ratio التي هي أعلى درجة من المقاييس الثلاثة السابقة ، وتسمح بجميع العمليات الحسابية .

وتم تقسيم المتغيرات من قبل إلى متغيرات نوعية Categorical ، وتشمل المقياسين الرمزي والترتيبي ، وإلى متغيرات كمية Quantitative وتشمل المقياسين الفترى والنسبي.

ولكن المتغيرات الكمية ، بدورها ، متغيرات متقطعة Discerte وتشمل بياناتها «أعداد» أو «تكرارات» ، ومتغيرات كمية مستمرة Continuous وبياناتها مقاسة لأي قيمة ما بين حد أدنى وحد أعلى (الأوزان ، الدخول ، الدرجات مثلاً) .

ولا تستخدم إختبارات كـا^2 على المتغيرات الكمية المستمرة ولكنها تستخدم على المتغيرات الكمية المتقطعة (تكرارات) ، وعادة تكون التكرارات مقسمة إلى مجموعات أو في جداول تكرارية ، إما جداول تكرارية مفردة (متغير واحد) أو مزدوجة (متغيرين) .

وتعتبر إختبارات كـا^2 من الإختبارات اللامعلمية Nonparametric ، وهي إختبارات لا تعتمد على توزيعات معينة ، ولا تفترض أي إفتراضات حول توزيع الظاهرة في المجتمع . بعكس إختبارات ى أو ت أو ف التي تفترض أن توزيع الظاهرة في المجتمع يتبع التوزيع الطبيعي .

يجب إذن تذكر الآتي :

١ – إختبارات كـا^2 تتم على تكرارات ، أما إختبارات ى ، ت ، ف فتتم على قراءات .

٢ - إختبارات كـا^{٢} من الاختبارات اللامعلمية .

١/٥ اختبارات كـا^{٢} : نظرة عامة

في القسمين التاليين ، سوف نعرض اختبارين من إختبارات كـا^{٢} هما .

(١) إختبار استقلال متغيرين .

(٢) إختبار تساوي نسبتين (أو أكثر) في مجتمعين (أو أكثر) .

وفي كل من هذين الإختبارين سوف نعرض :

١ - الفروض العدمية والفروض البديلة .

٢ - سحب عينة أو عدة عينات .

٣ - حساب التكرارات المتوقعة .

٤ - تحديد (α)، ودرجات الحرية وتحديد المنطقة الحرجة، وهي المنطقة ≤ (كـا$^{٢}_{\alpha}$) .

٥ - حساب قيمة إختبار كـا^{٢} :

$$\text{كا}^{٢} = \text{مجـ} \left(\frac{[\,(م_{ر} - ت_{ر})\,]^{٢}}{ت_{ر}} \right)$$

حيث : $م_{ر}$: هو التكرار المشاهد في المجموعة (أو العينة) ر .

$ت_{ر}$: التكرار المتوقع في المجموعة (أو العينة) ر

٦ - الوصول إلى قرار بشأن الفرض العدمي .

وتختلف صياغة الفروض العدمية والفروض البديلة ، بإختلاف الإختبار . أما التكرار المتوقع في الخطوة (٥) فهو التكرار المتوقع إذا كان الفرض العدمي صحيحًا .

مثلاً إذا كان الفرض العدمي أن النسبة في أحد المجتمعات ، محل الإختبار = ٢٥ ٪ وكان أحجام جميع العينات ١٠٠٠ ، نتوقع إذن أن التكرارات المصاحبة لهذه النسبة في هذا المجتمع هي = ٠,٢٥ × ١٠٠٠ = ٢٥٠ .

حساب التكرار المتوقع يختلف من إختبار إلى آخر ، ولكن بصفة عامة :

التكرار المتوقع = (الإحتمال المناسب (النسبة)) × ن

حيث : الإحتمال المناسب هو الإحتمال تحت صحة الفرض العدمي .

وباستخدام التعريف أعلاه نجد أن :

مجموع التكرارات المشاهدة = مجموع التكرارات المتوقعة .

والتوزيع الدقيق للإختبار الإحصائي هو توزيع متقطع Discrete . ولكن توزيع كـا٢ توزيع مستمر ، ولكنه يمدنا بتقريب إلى التوزيع الدقيق المتقطع ، بشرط ألا يكون التكرار المتوقع في أي مجموعة أقل من ٥ .

وفي حالـة وجود تكرارات أقـل من ٥ في أي مـجموعـة من مجمـوعات المتغير المتقطـع ، يدمج مجموعتان متجاورتان . فيتـم دمج التكرارات المشاهدة في المجموعـتين، ويتم أيضًا دمج التكـرارات المتوقعة في المجمـوعتين . وهذا الدمج صحيح، ولكنه يؤثر على درجة حرية الإختبار، كما سنرى فيما بعد .

وفي الخطوة (٤) ، نحدد مسـتوي المعنوية (∝) ، وهو حجم الخطأ من النوع الاول ، الذي يمثـل احتمال رفض الفرض العدمي وهـو فرض صحيح . وفي إختبارات كـا^2 توضع المنطقة الحرجة دائمًا على يمين التوزيع . أي أن ∝ تحدد مسـاحة الذيل الأيمن فـي التوزيع ، وهي المنـطقة الحرجة ، فـدائمًا تحدد القيمة الحرجة بالمقدار كـا$^2_{\infty}$ بدرجات الحرية المناسبة .

وفي الخطوة (٦) ، نصل إلى قرار بشـأن صحة الفرض العدمي . ويلاحظ أن الإختبار ، يقارن مجموع [مربع الفرق بـين التكرار المشـاهد والتكرار المتوقع منسوبًا إلى التكرار المتوقع] . وكلمـا زاد الفرق بين التكرار المشـاهد والتكرار المتوقع ، كلمـا أدى هذا إلى كبر قيمـة الإختبار الإحصائي . ويجـعل إحتمال رفض الفرض العدمي كبيرًا .

٢/٥ إختبارات الإستقلال Test of Independence

تهتم دوائر المال والإقتـصـاد والتجارة بمعرفة العلاقة بين مـتغيرين . مثلاً ، هل «النوع» له علاقة بمشاهدة البرامج الإخبارية ؟ هل تختلف معدلات الارباح باختلاف نوع المنشأة ؟ هل يوجد علاقة بين حجم السيارة وإستهلاكها للوقود ؟ هل يوجـد علاقة بين عمـل الزوجة وعدد الاولاد الذيـن تنجبهـم المرأة ؟ هل يوجد عـلاقة بين أسعار الـشقق والمنطقـة الـسكنيـة ؟ وهكذا . يوجد أسـاليب

إحصائية عديدة لقياس العلاقة بين متغيرين ، منها مثلاً أساليب الإنحدار والإرتباط .

ولكن أي أسلوب من هذه الأساليب يجب أن يتبع ؟ يعتمد الأسلوب على طبيعة البيانات ، فإذا كانت البيانات المتاحة عبارة عن تكرارات ، يجب إتباع أسلوب إختبار كا[2] .

وفي إختبارات الإستقلال ، يوجد متغيران ، كل متغير منهما مقسم إلى مجموعات منافية . ويُكَوّن جدول تكراري مزدوج لمجموعات المتغيرين ، فيوجد عدد من الصفوف، عددهم (ص) ، يكونوا مجموعات أحد المتغيرين ، ويوجد عدد من الأعمدة عددهم (ع) ، يكونوا مجموعات المتغير الأخر

وبالتالي يكون التكرار المشاهد في خلايا الجدول ، هو التكرار في مجموعة الصف ومجموعة العمود . وسوف نرمز له بالرمز "$م_{رو}$" حيث "ر" يمثل رقم الصف و "و" رقم العمود .

وبالمثل التكرار المتوقع في خلايا الجدول تمثل التكرار المتوقع لمجموعة الصف ومجموعة العمود . وسوف نرمز له بالرمز "$ت_{رو}$" حيث "ر" يمثل رقم الصف و "و" رقم العمود .

∴ $م_{رو}$ = التكرار المشاهد في الصف ر ، العمود و

$ت_{رو}$ = التكرار المتوقع في الصف ر و العمود و

وتكون خطوات إختبارات الإستقلال كالتالي :

١ - صياغة الفروض :

ف. : المتغيران مستقلان (غير مرتبطان)

ف١ : المتغيران غير مستقلين (مرتبطان)

٢ - تسحب عينة عشوائية ، ونكون جدول تكرارى مزدوج طبقاً لمجموعات المتغير الأول ومجموعات المتغير الثانى كما هو موضح فى الجدول التالى (مثلا) حيث عدد الصفوف · ص = ٢ وعدد الأعمدة · ع = ٣ ·

مجموعات المتغير الأول / مجموعات المتغير الثانى	(١)	(٢)	(٣)	مجموع هامشى
١	ت١١	ت٢١	ت٣١	مجموع الصف الأول
٢	ت١٢	ت٢٢	ت٣٢	مجموع الصف الثانى
مجموع هامشى	مجموع العمود الأول	مجموع العمود الثانى	مجموع العمود الثالث	ن

٣ - للحصول على التكرار المتوقع $ت_{ر و}$ ، نستخدم المعادلة :

$$ت_{ر و} = \frac{\text{مجموع الصف ر} \times \text{مجموع العمود و}}{ن}$$

وهذا التكرار المتوقع ، هو التكرار تحت فرض صحة فرض الإستقلال .
حيث أن المتغيرين يكونا مستقلين إذا كان :

$$ح (أ \cap ب) = ح (أ) \times ح (ب)$$

حيث ح (أ $\cap$ ب) هو احتمال التقاطع (الخلايا) وبتحويل هذه الإحتمالات إلى تكرارات بالضرب في ن ، نجد أن :

$$ن ح (أ \cap ب) = ن [ح (أ) \times ح(ب)]$$

$\therefore$ ح (أ) = مجموع الصف أ ÷ ن

ح (ب) = مجموع العمود (ب) ÷ ن

$$\therefore ت_{ر و} = ن \left[\frac{\text{مجموع الصف ر}}{ن} \times \frac{\text{مجموع العمود و}}{ن} \right]$$

$$\therefore ت_{ر و} = \frac{\text{مجموع الصف ر} \times \text{مجموع العمود و}}{ن}$$

٤ - تحدد ∞ ، ويكون درجة حرية الإختبار = (عدد الصفوف - ١) × (عدد الأعمدة - ١) = (ص - ١) (ع - ١) ، وتحدد قيمة كا$^2_{\infty}$ بدرجات حرية (ص - ١) (ع - ١) .

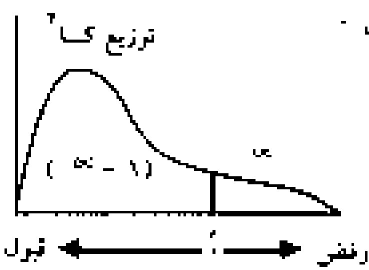
حيث ص = عدد الصفوف ، ع = عدد الأعمدة .

وإذا تم دمج صفين أو عمودين يكون (ص) عدد الصفوف في الجدول بعد الدمج ، ع عدد الأعمدة في الجدول بعد الدمج .

٥ - نحسب قيمة كــا² .

$$\text{كـا}^2 = \sum_{ر}\sum_{و} \frac{(م_{رو} - ت_{رو})^2}{ت_{رو}}$$

٦ - نصل إلى قرار بشأن صحة الفرض العدمي .



ويلاحظ أن الفرض العــدمي في إختبار كــا² هو الفرض الذي يود الباحث قبوله . فدائمًا نفترض في إختبار كـــا² أن المتغيرين مــستقلان . ويفـضل عنـد صياغة الـفرض العدمي لإختبار كــا² أن يكتب المتغيران بأسمائهم ، مثلاً نوع الــسيارة وحجم الــسيارة مستقلان، بدلاً من صياغة الفرض العدمي : المتغيران مستقلان .

هل العلاقة طردية أم عكسية

إذا تم رفض الفرض العدمي ، فإن هذا يعكس وجود علاقة بين المتغيرين ، كيف نتعـرف على إتجاه الـعلاقة ؟ وهـل هي علاقـة طردية ، بمعـنى ترتيـب المجموعات الـعليا في المتغيـر الأول يصاحبه ترتيب المجـموعة العليا في المـتغير الثاني . وهل هي علاقة عكسية ، بمعني هل تـرتيب المجموعة العليا في المتغير الأول يصاحبه ترتيب المجموعة الدنيا في المتغير الثاني ؟

للتعرف على إتجاه العلاقة ، نفحص الجدول ونقوم بالآتي :

١ – نحدد الخلايا ذات التكرار المشاهد أكبر من التكرار المتوقع .

٢ – نحدد موقع الخلايا في (١) أعلاه ، هل تشكل الخلايا أعلاه قطراً أم لا ؟ إذا لم تشكل قطراً ، فإن النتيجة هي أنه يصعب الحكم على إتجاه العلاقة بين المتغيرين .

٣ – إذا شكلت الخلايا في (١) أعلاه قطراً ، نفحص الخلايا (إذا كانت مرتفع/ مرتفع ، منخفض/ منخفض) تكون العلاقة طردية وإذا كانت خلايا القطر (مرتفع/ منخفض ، منخفض/ مرتفع) تكون العلاقة عكسية

تطبيق(٥-١)

هدف باحث إلى معرفة عما إذا كان معدلات الأرباح مرتبطة بنسب التوزيع أم لا . فكانت فروضه كالآتي :

ف. : معدلات الأرباح ونسب التوزيع مستقلة .

ف١ : معدلات الأرباح ونسب التوزيع غير مستقلة .

جمع الباحث بيانات عن ٢٠٠ منشأة ولاحظ نسب التوزيع . وقام بجدولة بيانات ٢٠٠ منشأة حسب عائد الأرباح (مرتفع - متوسط - خسارة) ، وطبقاً لنسب التوزيع - (صفر- أقل من ٢٥ ٪ ، ومن ٢٥ ٪ – أقل من ٥٠ ٪ ، ومن ٥٠ ٪ – إلى أقل من ٧٥ ٪ ، ومن ٧٥ ٪ إلى ١٠٠ ٪)، وحصل على البيانات الآتية :

عائد الربح / نسب التوزيع	صفر - ٢٢٥	٢٢٥ - ٢٥٠	٢٥٠ - ٢٧٥	٢٧٥ - ٤١٠	مجموع
مرتفع	١٠	٢٩	١٤	١١	٦٤
منخفض	١٦	٢٨	١٠	٦	٨٠
خسارة	٢٣	٢٨	٥	٠	٥٦
مجموع	٤٩	١٠٥	٢٩	١٧	٢٠٠

قام الباحث بحساب التكرار المتوقع كالتالي :

$$\text{ت}_{١١} = \frac{٤٩ \times ٦٤}{٢٠٠} = ١٥,٧ \qquad \text{ت}_{٢١} = \frac{١٠٥ \times ٦٤}{٢} = ٣٣,٦$$

$$\text{ت}_{٣١} = \frac{٢٩ \times ٦٤}{٢٠٠} = ٩,٣ \qquad \text{ت}_{٤١} = \frac{١٧ \times ٦٤}{٢٠٠} = ٥,٤$$

$$\text{ت}_{١٢} = \frac{٤٩ \times ٨٠}{٢٠٠} = ١٩,٦ \qquad \text{ت}_{٢٢} = \frac{١٠٥ \times ٨٠}{٢٠٠} = ٤٢,٥$$

$$\text{ت}_{٣٢} = \frac{٢٩ \times ٨٠}{٢٠٠} = ١١,٦ \qquad \text{ت}_{٤٢} = \frac{١٧ \times ٨٠}{٢٠٠} = ٦,٨$$

$$\text{ت}_{١٣} = \frac{٤٩ \times ٥٦}{٢٠٠} = ١٣,٧ \qquad \text{ت}_{٢٣} = \frac{١٠٥ \times ٥٦}{٢٠٠} = ٢٩,٤$$

$$\text{ت}_{٣٣} = \frac{٢٩ \times ٥٦}{٢٠٠} = ٨,١ \qquad \text{ت}_{٤٣} = \frac{١٧ \times ٥٦}{٢٠٠} = ٤,٨$$

ويمكن تلخيص التكرارات المشاهدة والتكرارات المتوقعة في جدول واحد (حيث التكرار المتوقع هو القيم بين قوسين (. . .)) كالتالي :

نسب التوزيع / عائد الربح	صفر – ٢٥٪	٢٥٪ – ٥٠٪	٥٠٪ – ٧٥٪	٧٥٪ – ١٠٠٪	مجموع
مرتفع	١٠ (١٥,٧)	٣٩ (٣٣,٦)	١٤ (٩,٣)	١١ (٥,٤)	٦٤
منخفض	١٦ (١٩,٣)	٤٨ (٤٢,٥)	١٠ (١١,٦)	٦ (٦,٨)	٨٠
خسارة	٢٣ (١٣,٧)	٢٨ (٢٩,٤)	٥ (٨,١)	صفر (٤,٨)	٥٦
مجموع	٤٩	١٠٥	٢٩	١٧	٢٠٠

ونلاحظ من الجدول أعلاه أن التكرار المتوقع في خلية الصف الثالث والعمود الرابع أي ت(٣٤) = ٤,٨ < ٥ . فقررت أن ندمج العمودين الأخيرين (وكان من الممكن عدم الدمج ، نظراً لأن ٤,٨ لا تقل قليلاً عن ٥) وبعد الدمج كان الجدول كالتالي (القيم المتوقعة بين قوسين) :

نسب التوزيع / عائد الربح	صفر – ٢٥٪	٢٥٪ – ٥٠٪	٥٠٪ أقل من ١٠٠٪
مرتفع	١٠ (١٥,٧)	٣٩ (٣٣,٦)	١٤ + ١١ = ٢٥ (٩,٣ + ٥,٤) = (١٤,٧)
منخفض	١٦ (١٩,٦)	٤٨ (٤٢,٠)	١٠ + ٦ = ١٦ (١١,٦ + ٦,٨) = (١٨,٤)
خسارة	٢٣ (١٣,٧)	٢٨ (٢٩,٨)	٥ + صفر = ٥ (٨,١ + ٤,٨) = (١٢,٩)

وتكون قيمة الإختبار الإحصائي :

$$\text{كـا}^{٢} = \frac{(١٥,٧ - ١٠)^{٢}}{١٥,٧} + \frac{(٣٣,٦ - ٢٩)^{٢}}{٣٣,٦} + \frac{(١٤,٧ - ٢٥)^{٢}}{١٤,٧}$$

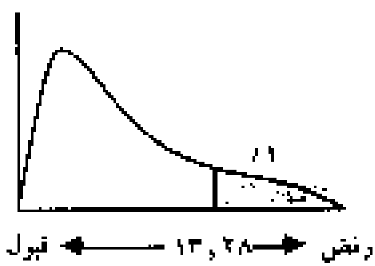
$$+ \frac{(١٩,٦ - ١٦)^{٢}}{١٩,٦} + \frac{(٤٢ - ٤٨)^{٢}}{٤٢} + \frac{(١٨,٤ - ١٦)^{٢}}{١٨,٤}$$

$$= \frac{(١٣,٧ - ٢٢)^{٢}}{١٣,٧} + \frac{(٢٩,٨ - ٢٨)^{٢}}{٢٩,٨} + \frac{(١٢,٩ - ٥)^{٢}}{١٢,٩}$$

$$= ٢٢,٩٧$$

وتكون قيمة كـا^{٢} الحرجة بمستوى معنوية (∞) .

ودرجـــــات حريـــة (ص - ١) (ع - ١) = (٣ - ١) (٣ - ١) = ٤ ، ∞ = ١ ٪ = ١٣,٢٨ .



وحيث أن كـا^{٢} المحسوبة (٢٢,٩٧) تقع داخل منطقة الرفض .

∴ نرفض الفرض العدمي . ونقبل النتيجة بأن معدل الربح ونسبة التوزيع مرتبطان .

ولمعرفة إتجاه العلاقة بين معدل الربح ونسبة التوزيع نجد أن التكرار المشاهد أكبر من التكرار المتوقع في الخلايا المحددة في الجدول التالي :

نسب التوزيع / عائد الربح	صفر - ٢٥٪	٢٥٪ - ٥٠٪	٥٠٪ - ١٠٠٪	مجموع
مرتفع	١٠ (١٥,٧)	٢٩ (٢٣,١)	٢٥ (١٤,٧)	٦٤
منخفض	١٦ (١٩,٦)	٤٨ (٤٢,٠)	١٦ (١٨,٤)	٨٠
خسارة	٢٣ (١٣,٧)	٢٨ (٢٩,٨)	٥ (١٢,٩)	٥٦
مجموع	٤٩	١٠٥	٤٦	

ونجد أن الخلايا حيث التكرار المشاهد > التكرار المتوقع تشكل قطراً وهذا القطر يشكل (مرتفع/ مرتفع) إلى (منخفض/ منخفض) .

∴ العلاقة طردية مباشرة .

تطبيق(٥-٢)

قام مكتب توظيف بتعيين مستخدمين من حملة الشهادات الجامعية العليا ، وحمله للمؤهلات المتوسطة خمس سنوات وحملة المؤهلات المتوسطة ٣ سنوات وبعد ستة أشهر من تعيينهم ، طلب من رؤسائهم تقييم كل منهم على مقياس : ممتاز ، جيد جداً ، وجيد .

وكانت نتائج التبويب المزدوج لعينة مكونة من ١٤٨ موظفاً كالآتى :

المؤهل / التقييم	ممتاز	جيد جداً	جيد	مجموع
عالى	١٢	١٤	١٨	٤٤
خمس سنوات	١٨	٢٨	١٦	٦٢
ثلاث سنوات	٢٢	١٢	٨	٤٢
مجموع	٥٢	٥٤	٤٢	١٤٨

أراد المكتب معرفة هل يوجد علاقة بين درجات التقييم والمؤهل الذى حصل عليه الموظف أم لا . قامت إدارة الأبحاث بصياغة الفروض كالآتى :

$\text{ف}_{٠}$: لا يوجد علاقة بين درجة التقييم والمؤهل (متغيران مستقلان)

$\text{ف}_{١}$: يوجد علاقة بين درجة التقييم والمؤهل (متغيران غير مستقلان)

وقام بحساب التكرارات المتوقعة كالتالى :

$$\text{ت}_{١١} = \frac{٥٢ \times ٤٤}{١٤٨} = ١٥,٥ \qquad \text{ت}_{٢١} = \frac{٥٤ \times ٤٤}{١٤٨} = ١٦,١$$

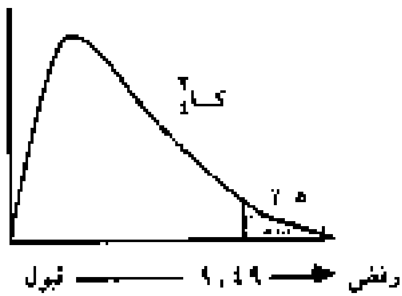
$$\text{ت}_{٣١} = \frac{٤٢ \times ٤٤}{١٤٨} = ١٢,٥ \qquad \text{ت}_{١٢} = \frac{٥٢ \times ٦٢}{١٤٨} = ٢١,٨$$

$$\text{ت}_{٢٢} = \frac{٥٤ \times ٦٢}{١٤٨} = ٢٢,٦ \qquad \text{ت}_{٣٢} = \frac{٤٢ \times ٦٢}{١٤٨} = ١٧,٦$$

$$ت_{13} = \frac{52 \times 42}{148} = 14.8 \qquad ت_{23} = \frac{54 \times 42}{148} = 15.3$$

$$ت_{33} = \frac{42 \times 42}{148} = 11.9$$

والتكرارات المتوقعة جميعها أكبر من ٥ . فإذا اختار الباحث α = ٥ ٪ ، تكون قيمة كـا٢ الحرجة بدرجات حرية (٣-١) (٣-١) = ٤ = ٩,٤٩ .



وقام الباحث بتلخيص التكرارات المشاهدة والتكرارات المتوقعة فى الجدول التالى ، وتم حساب قيمة كـا٢ المشاهدة كما هو مبين .

المؤهل / القيم	ممتاز	جيد جداً	جيد	مجموع
عالى	١٢ (١٥,٥)	١٤ (١٦,١)	١٨ (١٢,٥)	٤٤
خمس سنوات	١٨ (٢١,٨)	٢٨ (٢٢,٦)	١٦ (١٧,٦)	٦٢
ثلاث سنوات	٢٢ (١٤,٨)	١٢ (١٥,٣)	٨ (١١,٩)	٤٢
مجموع	٥٢	٥٤	٤٢	١٤٨

$$\text{كا}^{٢} = \frac{(١٥,٥ - ١٢)^{٢}}{١٥,٥} + \frac{(١٦,١ - ١٤)^{٢}}{١٦,١} + \frac{(١٢,٥ - ١٨)^{٢}}{١٢,٥}$$

$$+ \frac{(٢١,٨ - ١٨)^{٢}}{٢١,٨} + \frac{(٢٢,٦ - ٢٨)^{٢}}{٢٢,٦} + \frac{(١٧,٦ - ١٦)^{٢}}{١٧,٦}$$

$$= \frac{(١٤,٨ - ٢٢)^{٢}}{١٤,٨} + \frac{(١٥,٣ - ١٢)^{٢}}{١٥,٣} + \frac{(١١,٩ - ٨)^{٢}}{١١,٩}$$

$$= ١١,٠٧$$

وحيث أن قيمة كا^{٢} المشاهدة (١١,٠٧) أكبر من قيمة كــا^{٢} الحرجة (٩,٤٩) بمستوى معنوية ٥ ٪ ، يرفض ، إذن ، الفرض العدمي ، ونصل إلى أن التقييم والمؤهل متغيران مرتبطان .

ولمعرفة إتجاه العلاقة نجد أن التكرار المشاهد أكبر من التكرار المتوقع في الخلايا المحددة في الجدول السابق . وتشكل قطراً . ويسير الإتجاه (مرتفع/ منخفض) إلى (منخفض/ مرتفع) إذن العلاقة عكسية ، بمعنى أنه كلما ارتفع المؤهل كلما إنخفضت درجة التقييم والعكس صحيح .

٣/٥ إختبار نسبتين أو أكثر من مجتمعين أو أكثر

إذا أردت معرفة هل نسبة مبيعات التليفونات المحمولة واحدة بين محافظات الدلتا أم لا ؟ وإذا أردت معرفة هل نسبة الطلبة الذين يحصلون على تقدير عام «ممتاز» واحد في الأربعة سنوات الدراسية أم لا ؟

يمكننا إجراء إختبار تساوي النسب باستخدام إختبار كـا^2 . ونتبع الخطوات التالية :

١ - صياغة الفروض :

ف. : النسب في المجتمعات متساوية .

ف١ : النسب ليست جميعها متساوية .

٢ - تسحب عينة عشوائية من كل مجتمع من المجتمعات . وتبوب البيانات في جدول مزدوج ، حيث يمثل الصف الأول ، التكرارات في العينة التي ينطبق عليها الظاهرة ويمثل الصف الثاني : التكرارات في العينة التي لا ينطبق عليها الظاهرة . وتمثل الأعمدة : المجتمعات المختلفة .

٣ - نقوم بحساب التكرارات المتوقعة حيث :

$$ت_{ر و} = \frac{\text{مجموع الصف ر} \times \text{مجموع العمود و}}{ن}$$

٤ - نحدد مستوى المعنوية α ، درجات الحرية = (ص-١) (ع-١) = (٢-١) (ع-١) = (ع-١) ، حيث ع = عدد الأعمدة = عدد المجتمعات .

٥ - نحسب فيه الإختبار الإحصائي :

$$كا^2 = \text{مجـ مجـ} \left[\frac{(ح_{ر و} - ت_{ر و})^2}{ت_{ر و}} \right]$$

٦ – تقارن بين قيمة كـا^2 المشاهدة وقيمة كـا^2 الحرجة وتصل إلى قرار .

وفى الإختبار أعلاه ، تلاحظ الآتى :

١ – نقوم بسحب عدة عينات ، عينة من كل مجتمع ، فى حين أنه فى إختبار الإستقلال ، يوجد عينة واحدة فقط .

٢ – فرض تساوى النسب له علاقة بإختبار الإستقلال . إذا كانت الظاهرتان مستقلتان ، نستطيع القول بأن النسب متساوية فى المجموعات المختلفة وإذا اختلفت النسب ، فيوجد علاقة بين متغير الصف والمجموعات المختلفة .

تطبيق (٥-٣)

مصنع له أربعة مواقع . قام المصنع بسياسة جديدة بخصوص الغياب . وأردت أن تقيم ردود الأفعال فى الاربعة مواقع ، بالنسبة للسياسة الجديدة .

وافترضت أن نسبة ردود الأفعال واحدة في الاربعة مصانع . فكان الفرض العدمي والفرض البديل كالآتي :

ف$_0$: نسبة الإرتياح للسياسة الجديدة متساوية في الاربعة مواقع .

ف$_1$: نسبة الإرتياح للسياسة الجديدة مختلفة في الاربعة مواقع .

قام الباحث بسحب عينة عشوائية من كل موقع من المواقع ووجد الآتي :

المصنع / الشعور	(١)	(٢)	(٣)	(٤)	مجموع
الإرتياح	٤١	٢٨	٢٤	٣٤	١٢٧
عدم الإرتياح	٤٢	٢٠	١٥	٣١	١٠٨
مجموع	٨٣	٤٨	٣٩	٦٥	٢٣٥

قام الباحث بحساب التكرارات المتوقعة كالآتي :

$$\text{ت}_{\text{ر و}} = \frac{\text{مجموع الصف ر} \times \text{مجموع العمود و}}{\text{ن}}$$

$$\text{ت}_{١١} = \frac{٨٣ \times ١٢٧}{٢٣٥} = ٤٥ \qquad \text{ت}_{١٢} = \frac{٨٣ \times ١٠٨}{٢٣٥} = ٣٨$$

$$\text{ت}_{٢١} = \frac{٤٨ \times ١٢٧}{٢٣٥} = ٢٦ \qquad \text{ت}_{٢٢} = \frac{٤٨ \times ١٠٨}{٢٣٥} = ٢٢$$

$$\text{ت}_{٣١} = \frac{٣٩ \times ١٢٧}{٢٣٥} = ٢١ \qquad \text{ت}_{٣٢} = \frac{٣٩ \times ١٠٨}{٢٣٥} = ١٨$$

$$\text{ت}_{٤١} = \frac{٦٥ \times ١٢٧}{٢٣٥} = ٣٥ \qquad \text{ت}_{٤٢} = \frac{٦٥ \times ١٠٨}{٢٣٥} = ٣٠$$

إختار الباحث مستوى معنوية α = ٥ ٪ ، وحيث أن التكرارات المتوقعة جميعها أكبر من ٥ تكون درجات الحرية = (ص-١) (ع-١) = (٢-١) (٤-١) = ٣ . ∴ كـا^{٢} الحرجة = ٧,٨١ . نلخص التكرارات المشاهدة والتكرارات المتوقعة في جدول كالآتي :

الشعور / المصنع	(١)	(٢)	(٣)	(٤)	مجموع
الإرتياح	(٤٥) ٤١	(٢٦) ٢٨	(٢١) ٢٤	(٣٥) ٣٤	١٢٧
عدم الإرتياح	(٣٨) ٤٢	(٢٢) ٢٠	(١٨) ١٥	(٣٠) ٣١	١٠٨
مجموع	٨٣	٤٨	٣٩	٦٥	٢٣٥

ونحسب قيمة كـا^{٢} المشاهدة كالتالي :

$$\text{كا}^{٢} = \frac{(٤١-٤٥)^{٢}}{٤٥} + \frac{(٢٨-٢٦)^{٢}}{٢٦} + \frac{(٢٤-٢١)^{٢}}{٢١}$$

$$+ \frac{(٣٤-٣٥)^{٢}}{٣٥} + \frac{(٤٢-٣٨)^{٢}}{٣٨} + \frac{(٢٠-٢٢)^{٢}}{٢٢}$$

$$= \frac{(١٥-١٨)^{٢}}{١٨} + \frac{(٣١-٣٠)^{٢}}{٣٠} = ٢,١٠$$

وحيث أن كـا^{٢} المشاهدة (٢,١٠) < قيمة كـا^{٢} الحرجة . ∴ نقبل الفرض العدمي بتساوي نسب الارتياح نحو سياسة الغياب الجديدة في الأربعة مواقع .

٤/٥ إختبار جودة المطابقة Goodness-of- Fit Test

إختبار جودة المطابقة هو إختبار كا٢ الذي يسمح بإختبار عما إذا كانت التكرارات المشاهدة تطابق التكرارات المتوقعة المحسوبة بإفتراض توزيع معين في المجتمع أم لا ، أي التكرارات المتوقعة على أساس أن النسبة المعطاة تحت صحة الفرض العدمي صحيحة .

هذا الإختبار ، إذن ، يمكن من إختبار مدى مطابقة التوزيع المشاهد للتوزيع المتوقع مثلاً ، يفترض توزيع ت أن البيانات تتبع التوزيع الطبيعي .

وقبل إجراء الإختبار يجب نود إختبار عما إذا كان التوزيع في المجتمع يتبع التوزيع الطبيعي أم لا ؟

ويتبع إختبار جودة المطابقة الخطوات الآتية :

١ – صياغة الفروض :

ف٠ : التكرارات المشاهدة تتبع توزيعًا معينًا

ف١ : التكرارات المشاهدة لا تتبع توزيعًا معينًا

٢ – يسحب عينة عشوائية وتبوب بيانات العينة في مجموعات شاملة ومتنافية .

٣ – تحسب التكرار المتوقع كالآتي :

ت ر = ن (الاحتمال المناسب للمجموعة ر) طبقًا لتوزيع الفرض العدمي

٤ – نحدد مستوى المعنوية = ودرجات الحرية المناسب لإختبار كا^٢ حيث :

درجة الحرية = (ك - ١ - م)

حيـث : ك = عدد المجمـوعات المتنـافيـة ، م = عدد المعـلمات الـتي تم تقديرها من البيانات .

وعدد المجموعات المتنافية هى عدد المجـموعات التى استخدمت فى حـساب قيمة كـا^2 . أو عدد المجموعات بعد دمجها (إن تم دمجها) .

ونحدد قيمة كـا^2 الحرجة بمستوى معنوية α ، ودرجات الحرية المناسبة .

٥ - نحسب قيمة الإختبار الإحصائى :

$$كـا^2 = مجـ \left[\frac{(م_ر - ت_ر)^2}{ت_ر} \right]$$

٦ - نصل إلى قرار بشأن صحة الفرض العدمى .

وفى الخطوة الأولى وهى صـياغة الفروض ، قد تعطى بعـض النسب التى يعتقد صحتها فى المجموعات المختلفة . مثلاً قد يفترض أن نسبة من يسدد قيمة المبيعات نقدًا ، باستخدام كارت إئتمان ، وباستخدام شيكات ، هى : ٣٠ ٪ ، ٢٨ ٪ ، ٤٢ ٪ على التوالى . فيكون الفرض العدمى :

$ف_٠$: $ل_١$ = ٠,٣٠ ، $ل_٢$ = ٠,٢٨ ، $ل_٣$ = ٠,٤٢

والفرض البديل : $ف_١$: أحد هذه النسب غير صحيح .

ويلاحظ أن النسب شاملة حيث أن : ٠,٣٠ + ٠,٢٨ + ٠,٤٢ = ١,٠٠

١/٤/٥ إختبار مطابقة البيانات لنسب معينة في المجتمع

تكون الخطوات كالآتي :

١ - صياغة الفروض : ف. : ل١ = ل١. ، ل٢ = ل٢. ، ل ر = ل ر.

حيث : ل١. + ل٢. + ... + ل ر. = ١

والفرض البديل : ف١ : أحد هذه النسب ليست صحيحة .

٢ - تسحب عينة عشوائية وتقسم إلى مجموعات متنافية وشاملة ، وتحدد التكرارات المشاهدة في كل مجموعة .

٣ - نحسب التكرارات المتوقعة في كل مجموعة .

التكرار المتوقع : ت ر = ن × ل ر.   حيث : ل ر. = نسبة المجموعة ر تحت الفرض العدمي .

٤ - نحدد α ، درجات الحرية = (ك-١) ، حيث ك = عدد المجموعات المتنافية الشاملة التي استخدمت في حساب قيمة كـا^2 المشاهدة .

٥ - تحسب قيمة كـا^2 المشاهدة :

$$\text{كـا}^2 = \text{مجـ} \left[\frac{(\text{ع}_\text{ر} - \text{ت}_\text{ر})^2}{\text{ت}_\text{ر}} \right]$$

٦ - نقارن بين قيمة كـا^2 المشاهدة وقيمة كـا^2 الحرجة ونصل إلى قرار بشأن صحة الفرض العدمي .

تطبيق (٥-٥)

قام مركز أبحاث بإجراء دراسة عن رأى طلبة الجـامعات في تعاطى العقاقير المهدئة في أيام الإمتحانات .

كان الإعتقاد الـسائد هو أن ٧ ٪ من الطلبة يؤيدون التعاطى ، ١٨ ٪ لا يؤيدون التعاطى ، ٦٥ ٪ يؤيدون التعاطى تحت إشراف طبيب ، ١٠ ٪ يمتنعون عن الإجابة .

قام الباحث بصياغة فروضه كالآتى :

ف. : ل١ = ٠,٠٧ ، ل٢ = ٠,١٨ ، ل٣ = ٠,٦٥ ، ل٤ = ٠,١٠

ف١ : أحد هذه النـسب ليس صحيحاً .

قام الباحث بسحب عينة عشـوائية ، حجمها ٥٠٠ طالبًا وطـالبة من طلبة الجامعات وأعـطى لهم محاضرات عن أثـر التعاطى وأضراره ، ثم طلـب منهم الإدلاء برأيهم في التعاطى فكانت الآراء كالتالى :

٣٩ مؤيدون   ٩٩ معارضون

٣٣٦ مؤيدون تحت إشراف طبيب   ٢٦ ممتنعون

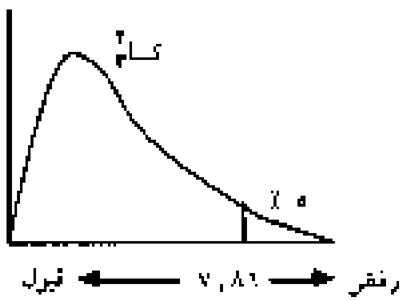
قام البـاحث بحساب الـتكرارات المتوقـعة طبقًا للإعـتقاد الـسائـد (الفرض العدمى) وكانت كالآتى :

التكـرار المتوقع لمجمـوعة المؤيـديـن = ٥٠٠ × ٠,٠٧ = ٣٥

التكـرار المتوقع لمجمـوعة المعارضـين = ٥٠٠ × ٠,١٨ = ٩٠

التكرار المتوقع لمجمـوعة المؤيدين بشرط = ٥٠٠ × ٠,٦٥ = ٣٢٥

التكرار المتوقع لمجموعة الممتنـعين = ٥٠٠ × ٠,١٠ = ٥٠



حـدد الباحـث α = ٥٪ ،
وحيث أن التكرار المتوقع فـي جميع الخلايا > ٥ ، إذن ، درجات الحرية = (ك − ١) = (٤−١) = ٣

∴ قيمة كـا² د.ح بدرجات حرية ٣ = ٧,٨١ .

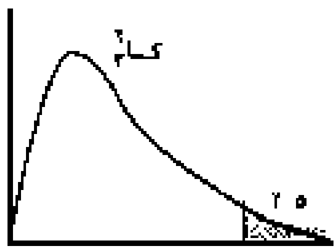
ولخص الباحث بياناته كالتالي :

التكرارات	الأداء				مجموع
	مؤيدون	معارضون	مؤيدون بشرط	ممتنعون	
المشاهدة	٣٩	٩٩	٣٣٦	٢٦	٥٠٠
المتوقعة	(٣٥)	(٩٠)	(٣٢٥)	(٥٠)	٥٠٠

وقام بحساب قيمة كـا^{٢} المشاهدة كالتالي :

$$كا^{٢} = \frac{(٣٥ - ٣٩)^{٢}}{٣٥} + \frac{(٩٠ - ٩٩)^{٢}}{٩٠} + \frac{(٣٢٥ - ٣٣٦)^{٢}}{٣٢٥}$$

$$+ \frac{(٥٠ - ٢٦)^{٢}}{٥٠} = ١٠,٨٣$$



وحيـث أن قيمـة كـــا^{٢} المشـاهدة (١٣,٢٤٩) > قيـمة كـــا^{٢} الحرجـة (٧,٨١) .

∴ نرفض الـفرض العدمي ، ونسـتطيع القـول أن أحد النسـب المعطاة تحت الفرض العدمي غير صحيحة .

تماريــــن

١ - فى كل من هذه الإختبارات ، أوجد قيمة كـا^2 الحرجة :

(أ) درجات الحرية ٢٣ ∝ = ٥ ٪ .

(ب) درجات الحرية ٨ ∝ = ١٠ ٪ .

(جـ) درجات الحرية ١٦ ∝ = ١٠ ٪ .

(د) درجات الحرية ٣ ∝ = ٥ ٪ .

٢ - أراد مدير تسويق معرفة هل توجد علاقة بين وسيلة الإعلان وقيمة مبيعات الأدوات الهندسية خلال يومين ، قام باحث بإدارة التسويق بتجميع بيانات عن وسيلة الإعلان عن السلعة والمبيعات . وبوب بياناته فى الجدول التالى :

وسيلة الإعلان / الإنفاق	الصحف	التليفزيون	المجلات	إذاعة الشرق الأوسط	لافتات	مجموع
أقل من ٢٥ جنيه	١٨	٣٢	١٧	٤٨	٥٣	١٦٨
٢٥ -	٢٠	٣٨	٢٤	٤٠	٣٩	١٦١
٥٠ -	٣٢	٤٣	٣٠	٣٢	٣٠	١٦٧
١٠٠ فأكثر	٣٠	٤١	٢٠	١٦	١٠	١١٢
مجموع	١٠٠	١٥٤	٩١	١٣٦	١٣٢	٦٠٨

(أ) إستخدم ∝ = ٥ ٪ ، هل تعتقد أن وسيلة الإعلان وقيمة المبيعات متغيران مستقلان ؟

(ب) بفرض أنك وجدت أن المتغيرين مرتبطان ، ما إتجاه العلاقة ؟

٣ - أرادت شركة أن تعرف هل يوجد علاقة بين الإتفاق المالى مع مندوب المبيعات (مرتب ثابت ، عمولة + مرتب ، عمولة فقط) وإنتاجية المندوب (منخفضة - متوسطة - مرتفعة) . وقام بتجميع بيانات عن ٢٣٥ مندوب وبوبت كالتالى :

الإنتاجية	الإتفاق المالى			مجموع
	مرتب	مرتب + عمولة	عمولة	
منخفضة	٣٣	٢٨	١١	٧٢
متوسطة	٢٤	٤٩	١٨	٩١
مرتفعة	٤٠	٣٣	٩	٨٢
مجموع	٩٧	١١٠	٣٨	٢٣٥

إستخدم α = ١٠ ٪ ، هل يوجد علاقة بين الإتفاق المالى مع المندوب وإنتاجية المندوب ؟ إن وجدت علاقة ، هل هى طردية أم عكسية .

٤ – الجدول التالي يعطى قيمة الإنفاق اليومي على الإعلانات وقيمة المبيعات اليومية لسلعة الإعلان .

الإنفاق اليومي على الإعلان	قيمة المبيعات (بالألف جنيه)			مجموع
	٥ –	١٠ –	١٥ – ٢٠	
٢٠٠٠ – أقل من ٥٠٠٠	١٠	٨	٧	٢٦
٥٠٠٠ – أقل من ١٠٠٠٠	٩	١٢	١٠	٣١
١٠٠٠٠ فأكثر	٦	١١	١٧	٣٤
مجموع	٢٦	٣١	٣٤	٩١

هل تعتقد وجود علاقة بين قيمة الإنفاق اليومي على الإعلانات وقيمة مبيعات سلعة الإعلان ؟ ∝ = ١٠ ٪ .

٥ – اهتمت وكالة تسويق بمعرفة رأى السيدات العاملات ، فى أحد منتجاتها وعما إذا كان رأى السيدات العاملات يختلف عن رأى السيدات غير العاملات . إختارت الوكالة عينة عشوائية من السيدات (عاملات وغير عاملات) وسئلن عن رأيهن فى المنتج (ممتاز – متوسط – عادى) . ولخصت البيانات فى الجدول التالى :

رأيهـــن	سيدات عاملات	سيدات غير عاملات
ممتــاز	٦٢	٣٤
متوســط	٨٤	٤٤
عــادي	٢٤	٢٢

هل تعتقد أن كون السيدة عاملة أو غيـر عاملة له علاقة برأيهن فى المنتج ؟
استخدم ∝ = ١٠ ٪ .

٦ - الجدول التالي يعطى عدد الأعطال فى أربعة ماكينات في ثلاث ورديات .

الماكينة / الوردية	(أ)	(ب)	(جـ)	(د)
الأولــى	١٠	٦	١٣	١٣
الثانيـة	١٠	١٢	١٩	٢١
الثالثـة	١٣	١٠	١٣	١٨

هل تعتقد وجود علاقة بين الوردية والأعطال ؟

٧ - أردت أن تعرف عما إذا كان سعر منتج معين مرتبط بدخول الأسر أم لا . وأشـــارت دراسات السوق أن : ٣٦ من ٨٠ أسرة دخلها منخفض ، ١٩ أسرة من ٥٠ أسرة دخلها متوسط ، أسرتين من ٢٠ أسرة ذو دخل مرتفع يستخدمون هذا المنتج . إستخدم ∝ = ٥ ٪ هل تعتقد أن نسبة من يستخدم هذا المنتج واحدة بين المستويات الثلاثة للدخول .

٨ - أردت أن تعرف عما إذا كانت القناة التليفزيونيـة الثانية مفضلة بنفس النسبة بين شباب ثلاثـة مدن ، سحبت عـينة عشـوائية حجمــها ١٠٠ شاب من شباب القاهرة والأسكـندرية وطنطا . ووجد أن عدد من يفضلــــوا القناة الثانية ٢٣ ، ٣٢ ، ١٨ لـثلاث مدن علـى التـوالـى . هـــل تعتقد أن نسبة الشباب الذي يتابع القناة الـثانية واحدة في الثلاثة مدن ؟ إستخدم ∝ = ٥ ٪ .

٩ - شركة متخصصة في الشحن ، أرادت معرفة هل نسبة المفقودات واحدة إذا تم إستــخدام ثلاث طرق للشـحن أم لا . سحبت الشركة عـينة عشوائية حجمها ٣٠٠ طرد تم شحنهم بالطريقة الأولى ووجد أن ١١ طرد منهم قد فقــدوا . وسحبت عينـة عشوائـية حجمـها ٢٥٠ طـرداً تم شحنـتهم بالطريقة الثانية ، ووجــد أن ٩ منهم قد فقدوا . وسحبت عينـة عشوائية حجــمها ٢٠٠ طرداً تــم شــحنهم بالطريقة الثالثة ووجد أن ١٥ منهم قد فقدوا .

هل تعتقد أن نسبة المفقودات واحدة بين الثلاث طرق ؟

١٠- رمي زهرة طاولة ١٢٠ مرة وأعطي النتائج التالية :

النتيجة	١	٢	٣	٤	٥	٦
التكرار	١٨	٢٣	١٦	٢١	١٨	٢٤

إختبر فرض تساوي النسب في المجموعات الست أعلاه . ∝ = ٥ ٪ .

١١- في تجربة كثيرات الحدود ، تحتوي علي ثلاثة مجموعات . سحبت عينة عشوائية حجمها ٢٠٠ مفردة وأعطت النتائج الآتية .

النتيجة	١	٢	٣
التكرار	٤٨	٦٩	٨٣

هل هناك شواهد بأن نسب هذه المجموعات في المجتمع هي :

ل١ = ٠,٢٥ ، ل٢ = ٠,٢٥ ، ل٣ = ٠,٥٠

١٢- سحبت عينة عشوائية من طلبة كلية معينة ، وكانت نتائج العينة كالتالي :

النتيجة	الأولى	الثانية	الثالثة	الرابعة
التكرار	١٥	٣٠	٢٥	٣٠

هل تعتقد أن نسبة الطلبة في الكلية متساوية في الأربعة فرق ؟ ∝ = ٥ ٪

١٣- إهتمت شركة تأمين بدراسة نسبة العملاء الذين يلغون بوالص التأمين مع الشركة ، وإعتقدت الشركة أن نسبة الإلغاء متساوية بغض النظر عن قيمة البوليصة . سحبت الشركة عينة من ملفات البوالص الملغاة . وقامت بتبويب البيانات كالآتي :

قيمة البوليصة بالألف جنيه	١٠	١٥	٢٠	٢٥	٣٠
عدد الإلغاءات	٣١	٣٩	٦٧	٥٤	٥٩

هل تؤيد الشركة فى أن نسبة الإلغاءات واحدة ، بغض النظر عن قيمة الوثيقة ؟

١٤- من المعروف أن سداد قيم المبيعات تتم بالنسب الآتية : ٢٨ ٪ دفع نقدى، ١٨ ٪ باستخدام كارت إئتمان البنك (أ) ، ١٧ ٪ باستخدام كارت إئتمان البنك (ب) ، ٣٧ ٪ باستخدام الشيكات .

رأى مدير المبيعات فى أحد الفروع أنه إذا كانت النسب منافية للنسب أعلاه، فيجب إعادة النظر فى عدد المستخدمين (الكاشير) وإعادة تدريب بعضهم. تم سحب عينة عشوائية من المبيعات ، ووجد أن السداد تم كالآتي :

طريقة السداد	نقداً	كارت البنك (أ)	كارت البنك (ب)	شيك
عدد العمليات	١١٨	٧٠	٨٢	١٤٧

هل تعتقد أن هذا الفرع فى حاجة لإعادة النظر في عدد المستخدمين وإعادة تدريب بعضهم ؟ إستخدم ∝ = ٥ ٪ .

١٥ - الجدول التالي يعطى حصص مبيعات ستة أنواع سيارات في السوق المصرى .

نوع السيارة	١	٢	٣	٤	٥	٦
حصة السوق	٪ ٣٤,٥	٪ ٤١,٦	٪ ٩,٣	٪ ٩,٢	٪ ٨,٣	٪ ١٧,١

تم سحب عينة حجمها ٢٠٠٠ سيارة حديثة مباعة ووجد أن مبيعات كل نوع من الأنواع الست أعلاه كانت كالآتي :

نوع السيارة	١	٢	٣	٤	٥	٦
عدد المبيعات	٧٦٥	٤٤٦	١٧٥	١٨٧	١٧٨	٢٩٩

هل تتفق المبيعات الفعلية مع حصة السوق ؟

الجزء الثانى

الاستدلال الإحصائي في الانحدار والارتباط
الخطي البسيط

الباب السابع
تحليل الانحدار الخطي المتعدد

تأليف
د/ محمـــود الدرينـــى
مدرس الإحصاء

الباب السابع

الاستدلال الإحصائي لأسلوب الانحدار والارتباط الخطي البسيط

١/٦ مقدمة

تحليل الانحدار ، هو أسلوب إحصائي يستخدم في دراسة وتحليل العلاقة بين متغير تابع كمي ، ومتغير أو أكثر من المتغيرات المستقلة ، ومن الأمثلة المألوفة لتطبيقه ، ما يلي :

١- دراسة وتحليل العلاقة بين الكمية المطلوبة من سلعة معينة كمتغير تابع ، ومجموعة المتغيرات المستقلة المحددة لسلوكه مثل : سعر السلعة ، وسعر السلعة البديلة ، ودخل المستهلك .

٢- دراسة وتحليل العلاقة بين الإنفاق العائلي كمتغير تابع ، والدخل العائلي كمتغير مستقل .

٣- دراسة العلاقة بين كمية الإنتاج من منتج معين كمتغير تابع ، وحجم المواد الخام ، وعدد العمال كمتغيرين مستقلين .

٤- يستخدم هذا الأسلوب عند دراسة أثر الإنفاق على الدعاية كمتغير مستقل على حجم المبيعات كمتغير تابع .

٥- دراسة وتحليل العلاقة بين حجم المبيعات كمتغير مستقل ، يؤثر على المرتب الذي يتقاضاه مندوب المبيعات كمتغير تابع .

وهكذا ، يستخدم تحليل الانحدار في كثير من النواحي التطبيقية لتحقيق أغراض معينة منها: (أ) تقدير معالم النموذج الرياضي المفترض لتمثيل العلاقة بين المتغير التابع ، ومجموعة المتغيرات المستقلة، باستخدام بيانات عينة ، للتحقق من صحة

التابع ، ومجموعة المتغيرات المستقلة ، باستخدام بيانات عينة ، للتحقق من صحة النظرية التي بمقتضاها تم تكوين هذا النموذج . (ب) إذا تبين أن النموذج المفترض مناسب لتمثيل البيانات ، يمكن استخدامه لغرض التنبؤ بقيمة المتغير التابع من قيمة المتغير المستقل ؛ ولذا يطلق على المتغير التابع أحيانا ، بالمتغير المتنبأ به ، وعلى المتغيرات المستقلة بالمتغيرات المفسرة ، أو المتغيرات المتنبأ منها .

وأبسط صور تحليل الانحدار تلك الذي يشمل متغير مستقل واحد ، ويسمى بالانحدار البسيط ، حيث يهتم هذا التحليل بدراسة وتحليل العلاقة بين متغير تابع ومتغير مستقل واحد . ووصلا لما سبق دراسته في العام الماضي ، حول شكل نموذج الانحدار الخطي البسيط ، وكيفية تقدير معالم النموذج ، وهو ما يطلق عليه التقدير بنقطة ، سوف يهتم هذا الفصل بالاستدلال الإحصائي للانحدار الخطي البسيط ، ويشمل اختبارات الفروض المتعلقة بمناسبة نموذج الانحدار في تمثيل البيانات ، ثم تقدير فترة ثقة لمعامل الانحدار ، واختبار معنويته .

كما يتناول هذا الفصل الاستدلال الإحصائي للارتباط الخطي البسيط ، فقد درس الطالب كيف يمكنه حساب معامل الارتباط من خلال عينة من أزواج القيم (س ، ص) ، ومن ثم نعرض في هذا الفصل اختبار معنوية معامل الارتباط في المجتمع .

٦/٢ الاستدلال الإحصائي للانحدار الخطي البسيط

يهتم تحليل الانحدار الخطي البسيط ، بدراسة وتحليل أثر متغير مستقل واحد ، على متغير تابع كمي ، وخاصة عندما تمثل العلاقة بين هذين المتغيرين بمعادلة خط مستقيم تأخذ الصورة : ص = أ + ب س ، ويلاحظ أن الطرف الأيمن

الانحدار بالانحدار " الخطي " ، وسمي "بسيط" لأن النموذج يحتوي على متغير مستقل واحد هو س .

وهناك عدد من المشاكل الإحصائية ، التي تعترض تطبيق تحليل الانحدار ، وأهمها مشكلة تحديد المنحنى الذي يمثل البيانات تمثيلا جيدا ، ويقصد بذلك تحديد أفضل نموذج انحدار ، يمثل العلاقة بين المتغير التابع والمتغير المستقل تمثيلا جيدا . ويمكن حل هذه المشكلة ، باتباع إحدى الطريقتين التاليتين :

<u>الطريقة الأولى</u> : يتم عرض أزواج قيم مشاهدات العينة ($س_1$ ، $ص_1$) ، ($س_2$ ، $ص_2$) ، ..... ، ($س_ن$ ، $ص_ن$) ، في شكل بياني يسمى بنقط الانتشار ، ومن خلال الشكل الناتج يمكن تحديد أفضل نموذج يمثل هذه البيانات ، هل هو نموذج يأخذ شكل معادلة خط مستقيم ، أو نموذج يأخذ شكل معادلة غير خطية كمعادلة من الدرجة الثانية ، أو معادلة لوغاريتمية ، أو معادلة أسية . وهكذا . ومثال على ذلك يبين الشكل (٦-١) عدد من الأشكال لنقط الانتشار .

شكل (٦)

أشكال بيانية لنقط الانتشار

معادلة من الدرجة الثانية	معادلة لوغاريتمية	معادلة أسية	معادلة خط مستقيم
ص = أ + $ب_1$ س + $ب_2$ $س^2$	ص = أ + ب لو (س)	ص = أ $هـ^{ب س}$	ص = أ + ب س

<u>الطريقة الثانية</u> : وهي شائعة الاستخدام ، وتتضمن عدة خطوات منطقية هي :

١ - افتراض أن النموذج الأفضل لتمثيل البيانات يأخذ شكل معادلة خط مستقيم :

ص = أ + ب س كما في مبين في شكل (٦-١) (أ) .

٢ - يتم تقدير معالم النموذج المفترض ، أي حساب الثابت أ ، ومعامل الانحدار ب ، باستخدام بيانات عينة عشوائية من أزواج القيم : (س، ص) .

٣- تستخدم بعض اختبارات الفروض الإحصائية ، لاختبار صحة ، أو مناسبة النموذج المفترض في الخطوة رقم (١) في تمثيل البيانات .

٤- إذا تبين من الاختبار أن النموذج مناسب لتمثيل العلاقة بين المتغير التابع ص ، المتغير المستقل س ، يكون الخط المستقيم هو أفضل توفيق للبيانات ، وإذا أثبت الاختبار عدم مناسبة نموذج الخط المستقيم ، يتم افتراض نموذج آخر لتمثيل البيانات كنموذج يأخذ معادلة لوغاريتمية ، أو أسية ، أو من الدرجة الثانية ، أو غير ذلك ، وتكرر الخطوات من (١) إلى (٤) ، حتى يتم التوصل إلى نموذج صالح لوصف العلاقة بين المتغير التابع ص ، المتغير المستقل س ؛ ومن ثم يمكن استخدامه في التنبؤ .

وسوف نهتم في هذا الفصل بتحليل نموذج الانحدار الخطي البسيط ، الذي يأخذ شكل معادلة خط مستقيم .

٦/٢/١ شكل نموذج الانحدار الخطي البسيط

إذا كان ص هو المتغير التابع ، س هو المتغير المستقل ، فإن المشاهدة التابعة رقم ر ، يمكن تمثيلها بنموذج انحدار خطي بسيط يأخذ الصورة التالية:

$$ص_ر = \alpha + \beta س_ر + د_ر \qquad (٦-١)$$

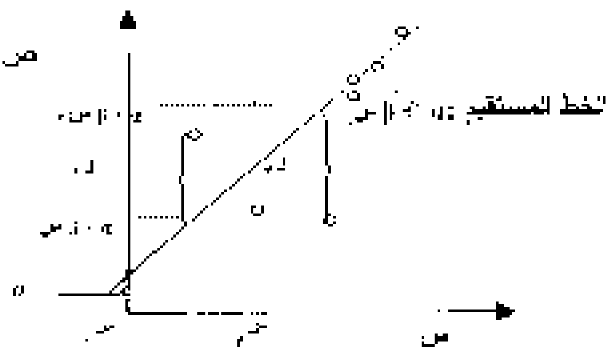
حيث أن :

α ، β هما معاملي الانحدار في المجتمع ، تعبر α عن الجزء المقطوع من محور الصادات عند تمثيل معادلة الخط المستقيم $ص_ر = \alpha + \beta س_ر$ ، أي أنها تمثل قيمة

ص عندما تكون قيمة س مساوية للصفر ، β هي ميل الخط المستقيم ، وتمثل مقدار التغير في المتغير التابع ص عندما تتغير قيمة س بمقدار وحدة واحدة ، أي أنها تعبر عن أثر المتغير المستقل س على المتغير التابع ص .

د : يمثل الخطأ العشوائي ، وهو الفرق بين القيمة الفعلية ص ، وقيمة ص الواقعة على الخط المستقيم أي أن : د = ص − (α + β س) ، كما هو مبين في الشكل (٦-٢) .

شكل (٦-٢)



ويوجد عدد كبير جدا من القيم المختلفة للمعاملين (α ، β) ومن ثم يكون لدينا عدد كبير من الخطوط المستقيمة (α + β س) ، وأفضل هذه الخطوط هو الذي يمر بأكبر عدد من نقط الانتشار ، بمعنى آخر أن أفضل هذه الخطوط هو الذي ينتج عنه أصغر مجموع مربعات أخطاء عشوائية مجـ د^2 = مجـ(ص - (α + β س))^2 ، وللوصول إلى هذا الخط المستقيم ، يجب توافر الشروط التالية :

١ - قيم المتغير المستقل س قيم معطاة **Fixed** ، وعند كل قيمة من هذه القيم ، فإن المتغير التابع ص متغير عشوائي له توزيع طبيعي ، توقعه م ص/س ، وتباينه

$\sigma^2_{\text{ص|س}}$ ، وهو ثابت من مشاهدة لأخرى ، ويطلق على $\mu_{\text{ص|س}}$ ، $\sigma^2_{\text{ص|س}}$ (توقع وتباين المتغير التابع ص عند القيمة المحددة لـ س) .

٢ - أن قيم المتغير التابع ص ، بالإضافة إلى الفرض السابق مستقلة إحصائيا ، فمثلا عند دراسة العلاقة بين الإنفاق العائلي كمتغير تابع ، والدخل العائلي كمتغير مستقل ، نجد أن إنفاق كل لأسرة لا يتأثر بإنفاق أي أسرة من الأسر الأخرى .

٣ - أن التوقع الشرطي $\mu_{\text{ص|س}}$ يأخذ شكل معادلة خط مستقيم هي : $\mu_{\text{ص|س}} = \alpha + \beta\text{س}$ ، وعند كل قيمة من قيم س ، فإن القيمة المتوقعة لـ ص تقع على هذا الخط كما هو مبين بالشكل (٦-٢) .

٤ - يعبر الخطأ العشوائي د عن الفرق بين القيمة الفعلية ص ، والقيمة المتوقعة له $\mu_{\text{ص|س}}$ ، ويفترض أن له عند كل قيمة من قيم س توزيع طبيعي ، القيمة المتوقع له صفر ، وتباينه $\sigma^2_{\text{ص|س}}$.

٥ - أن الخطأ العشوائي د مستقل إحصائيا عن المتغير المستقل س .

٦/٢/٢ استخدام طريقة المربعات الصغرى في الحصول على أفضل توفيق

تعتبر طريقة المربعات الصغرى The Least- Squares Method أحد طرق التقدير الإحصائي التي تستخدم في تقدير معالم نموذج الانحدار الخطي ، ويفرض أن لدينا عينة من أزواج القيم (س_1 ، ص_1)، (س_2 ، ص_2) ، ......، ($\text{س}_\text{ن}$ ، $\text{ص}_\text{ن}$) ، وأن (أ ، ب) هما مقدري المربعات الصغرى للمعاملين (α ، β) على التوالي في النموذج (٦-١) ، فإن أ ، ب هما اللذان يجعلان مجموع مربعات

الأخطاء العشوائية مجـ $ه_ر^2$ = مجـ$(ص_ر - (\alpha + \beta س_ر))^2$ أصغر ما يمكن ، ويحسبان من بيانات العينة باستخدام المعادلات التالية .

$$ب = \frac{\sum_{ر=1}^{ن} (س_ر - \bar{س})(ص_ر - \bar{ص})}{\sum_{ر=1}^{ن} (س_ر - \bar{س})^2} \qquad (٦-٢)$$

ويمكن تبسيط المعادلة أعلاه في صورة أخرى لتسهيل العمليات الحسابية كما يلي :

$$ب = \frac{ن \text{ مجـ } س ص - \text{ مجـ } س \text{ مجـ } ص}{ن \text{ مجـ } س^2 - (\text{مجـ } س)^2} \qquad (٦-٣)$$

وأما الثابت أ ، فيحسب بالمعادلة التالية :

$$أ = \bar{ص} - ب \bar{س} \qquad (٦-٤)$$

حيث أن $\bar{ص}$ هو الوسط الحسابي في العينة للمتغير التابع ص ، $\bar{س}$ هو الوسط الحسابي في العينة للمتغير المستقل س .

ومن ثم يكون أ هو التقدير بنقطة للثابت في المجتمع α ، ب هو التقدير بنقطة لمعامل الانحدار في المجتمع β ، ويطلق على أ ، ب بمقدري المربعات الصغرى .

ومن ثم يكتب نموذج الإنحدار المقدر في صورة خط مستقيم كما يلي :

$$\hat{ص}_ر = أ + ب س_ر \qquad (٦-٥)$$

وتحت توافر الشروط السابق عرضها ، يتصف مقدري المربعات الصغرى (أ ، ب) بخصائص المقدر الجيد ، وهي :

١- أ ، ب مقدران غير متحيزين للمعلمتين α ، β على الترتيب .

٢- انهما مقدران متسقان .

٣- انهما مقدران كافيان .

٤- لهما أقل تباين .

ويطلق على هذه الخصائص اسم " المقدر الخطي غير المتحيز ذو الأقل تباين " .

وتحت توافر الشروط السابقة أيضا ، فان المقدران أ ، ب كل منهما يتبع توزيع طبيعي ، بتوقع ، وتباين كما يلي :

$$\text{ت}(\text{أ}) = \alpha \quad , \quad \sigma^2_{\text{أ}} = \sigma^2_{\text{ص.س}} \left[\frac{1}{\text{ن}} + \frac{\bar{\text{س}}^2}{(\text{ن}-1)\,\text{ع}^2_{\text{س}}} \right]$$

$$\text{ت}(\text{ب}) = \beta \quad , \quad \sigma^2_{\text{ب}} = \sigma^2_{\text{ص.س}} \left[\frac{1}{(\text{ن}-1)\,\text{ع}^2_{\text{س}}} \right]$$

حيث أن $\sigma^2_{\text{ص.س}}$ هو تباين الخطأ العشوائي ، $\text{ع}^2_{\text{س}}$ هو تباين العينة للمتغير المستقل س ، ويحسب باستخدام الصيغة التالية :

$$\text{ع}^2_{\text{س}} = \frac{\text{مجـ س}^2 - \text{ن}\,\bar{\text{س}}^2}{(\text{ن}-1)}$$

وحيث أن تباين الخطأ العشوائي $\sigma^2_{\text{ص.س}}$ غير معلوم ، مما يصعب حساب $\sigma^2_{\text{أ}}$ ، $\sigma^2_{\text{ب}}$ ، وقد درس الطالب في العام الماضي كيف يمكن حساب تقدير المربعات الصغرى لتباين الخطأ العشوائي من بيانات العينة ، ويرمز له بالرمز $\text{ع}^2_{\text{ص.س}}$ ، وتذكرا لما سبق ، فإنه يحسب بالمعادلة التالية :

$$\text{ع}^2_{\text{ص.س}} = \text{مجـ}\,(\text{ص} - \hat{\text{ص}})^2 \quad (\text{ن}-2) \qquad (٦-٦)$$

وبالحصول على $\text{ع}^2_{\text{ص.س}}$ يمكن حساب تبايني مقدري أ ، ب ، وهما $\text{ع}^2_{\text{أ}}$ ، $\text{ع}^2_{\text{ب}}$ ، وباخذ الجذر التربيعي لكل منها يمكن الحصول على الخطأ المعياري لهما .

مثال (٦-١)

البيانات التالية تمثل الإنفاق الاستهلاكي والدخل العائلي بالألف جنيه في السنة لعينة من ١٠ أسر.

رقم المفردة	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠
الاستهلاك (ص)	٢	٣	٤	٤	٥	٥	٦	٧	٧	٨
الدخل (س)	٣	٤	٤	٦	٧	٦	٨	٩	٩	١١

بفرض أن العلاقة بين الاستهلاك والدخل يمكن تمثيلها في شكل خط مستقيم.

المطلوب:

١- إيجاد تقدير المربعات الصغرى لمعاملي الانحدار (α، β).

٢- اكتب شكل معادلة الانحدار المقدرة، وفسر هذه المعادلة.

٣- حساب الخطأ المعياري للتقدير، وكذلك الخطأ المعياري للمقدر أ، والمقدر ب.

الحل: فيما يلي حساب المجموع للحصول على تقديرات معاملات الانحدار.

رقم المفردة	ص	س	س ص	س²	ص²
١	٢	٣	٦	٩	٤
٢	٣	٤	١٢	١٦	٩
٣	٤	٤	١٦	١٦	١٦
٤	٤	٦	٢٤	٣٦	١٦
٥	٥	٧	٣٥	٤٩	٢٥
٦	٥	٦	٣٠	٣٦	٢٥
٧	٦	٨	٤٨	٦٤	٣٦
٨	٧	٩	٦٣	٨١	٤٩
٩	٧	٩	٦٣	٨١	٤٩
١٠	٨	١١	٨٨	١٢١	٦٤
المجموع	٥١ مجـ ص	٦٧ مجـ س	٣٨٥ مجـ س ص	٥٠٩ مجـ س²	٢٩٣ مجـ ص²

١- إيجاد تقدير المربعات الصغرى للمعاملين (α ، β) :

تقدير β هو ب ، ويحسب بتطبيق المعادلة رقم (٦-٣)

$$\text{ب} = \frac{\text{ن مجـ س ص} - \text{مجـ س مجـ ص}}{\text{ن مجـ س}^{٢} - (\text{مجـ س})^{٢}}$$

$$= \frac{(١٠)(٣٨٥) - (٦٧)(٥١)}{(١٠)(٥٠٩) - (٦٧)^{٢}}$$

$$= \frac{٤٣٣}{٦٠١} = ٠,٧٢$$

وتقدير α هو أ ويحسب بتطبيق المعادلة رقم (٦-٤)

$$\text{أ} = \bar{\text{ص}} - \text{ب}\,\bar{\text{س}}$$

$$= \frac{٥١}{١٠} - (٠,٧٢)\frac{٦٧}{١٠}$$

$$= ٥,١ - ٤,٨٢٤ = ٠,٢٧٦$$

٢ - كتابة شكل معادلة الانحدار المقدرة ، وتفسيرها .

$$\hat{\text{ص}} = ٠,٢٧٦ + ٠,٧٢\ \text{س}$$

ومن المعادلة أعلاه ، نجد أن قيمة الإنفاق العائلي المتنبأ به $\hat{\text{ص}}$ عندما يكون الدخل س قيمته صفرا هو ٠,٢٧٦ ألف جنيه ، ويسمى بالإنفاق الثابت ، بينما يدل المعامل (ب=٠,٧٢) على أنه إذا تغير الدخل س بمقدار وحدة واحدة ، تغير الإنفاق في نفس الاتجاه بمقدار ٠,٧٢ ألف جنيه . فمثلا إذا زاد الدخل من ٣ إلى ٤ ألف جنيه في السنة ، يمكن حساب التغير الذي يطرأ على الإنفاق العائلي كما يلي :

عند س=٣ ، $\hat{\text{ص}} \mid_{\text{س}=٣} = ٠,٢٧٦ + ٠,٧٢(٣) = ٢,٤٣٦$ ألف جنيه .

عند س = ٤ ، ص̂ | س = ٤ = ٠,٢٧٦ + ٠,٧٢ (٤) = ٣,١٥٦ ألف جنيه

مقدار الزيادة في الإنفاق = ص̂ | س = ٤ - ص̂ | س = ٣

= ٣,١٥٦ - ٢,٤٣٦ = ٠,٧٢ ـــــــ قيمة (ب)

أي أن معامل الانحدار ب ، يمثل مقدار التغير في ص إذا تغيرت س بمقدار وحدة واحدة ، وهذا المعدل ثابت ، ويبين هذا المعامل أيضا نوع العلاقة بين المتغيرين ص ، س (طردية أو عكسية) .

٣ - حساب الخطأ المعياري للتقدير (ع ص س) .

$$ع_{ص س} = \sqrt{\frac{مجـ (ص - \hat{ص})^2}{ن - ٢}}$$

ولحساب هذا الخطأ يجب حساب أخطاء التقدير : د = (ص - ص̂) عند كل قيمة من قيم س المعطاة في جدول رقم (٦-١) كما يلي :

جدول رقم (٦ - ٢)

ص	س	ص̂ = ٠,٢٧٦ + ٠,٧٢ س	(ص - ص̂)	(ص - ص̂)²
٣	٣	٠,٢٧٦ + ٠,٧٢ (٣) = ٢,٤٣٦	٠,٥٦٤	٠,٣١٨٠٩٦
٣	٤	٠,٢٧٦ + ٠,٧٢ (٤) = ٣,١٥٦	-٠,١٥٦	٠,٠٢٤٣٣٦
٤	٤	٠,٢٧٦ + ٠,٧٢ (٤) = ٣,١٥٦	٠,٨٤٤	٠,٧١٢٣٣٦
٤	٦	٠,٢٧٦ + ٠,٧٢ (٦) = ٤,٥٩٦	-٠,٥٩٦	٠,٣٥٥٢١٦
٥	٧	٠,٢٧٦ + ٠,٧٢ (٧) = ٥,٣١٦	-٠,٣١٦	٠,٠٩٩٨٥٦
٥	٦	٠,٢٧٦ + ٠,٧٢ (٦) = ٤,٥٩٦	٠,٤٠٤	٠,١٦٣٢١٦
٦	٨	٠,٢٧٦ + ٠,٧٢ (٨) = ٦,٠٣٦	-٠,٠٣٦	٠,٠٠١٢٩٦
٧	٩	٠,٢٧٦ + ٠,٧٢ (٩) = ٦,٧٥٦	٠,٢٤٤	٠,٠٥٩٥٣٦
٧	٩	٠,٢٧٦ + ٠,٧٢ (٩) = ٦,٧٥٦	٠,٢٤٤	٠,٠٥٩٥٣٦
٨	١١	٠,٢٧٦ + ٠,٧٢ (١١) = ٨,١٩٦	-٠,١٩٦	٠,٠٣٨٤١٦
٥١	٦٧	٥١	صفر	١,٧٠٣٨٣٦
			مجـ (ص - ص̂)	مجـ (ص - ص̂)²

ومن الجدول اعلاه نجد ان الخطأ المعياري للتقدير هو :

$$ع_{ص.س} = \sqrt{\frac{مجـ(ص - \hat{ص})^2}{ن - ٢}} = \sqrt{\frac{١,٧٠٣٨٢٦}{(١٠-٢)}}$$

$$= \sqrt{٠,٢١٢٩٧٨٢٥} = ٠,٤٦١$$

وتدل النتيجة السابقة أيضا أن تقدير الانحراف المعياري للمتغير التابع ص عند كل قيمة من قيم المتغير المستقل س قيمته = ٠,٤٦١ . فهو يساوي الخطأ المعياري للتقدير ع ص.س ، كما هو مبين في شروط الانحدار السبق عرضها .

حساب الخطأ المعياري للمقدرين (أ ، ب) .

أولا الخطأ المعياري للمقدر أ ($\hat{أ}$)

$$ع_{أ} = ع_{ص.س}\sqrt{\frac{١}{ن} + \frac{\bar{س}^2}{(ن-١)ع^2_{س}}}$$

$$ع^2_{س} = \frac{مجـ س^2 - ن\,\bar{س}^2}{(ن-١)}$$

$$= \frac{٥٠٩ - (١٠)(٦,٧)^2}{(١٠-١)} = \frac{٦٠,١}{٩} = ٦,٦٧٨$$

إذا

$$ع_{أ} = ٠,٤٦١\sqrt{\frac{١}{١٠} + \frac{(٦,٧)^2}{(٦,٦٧٨)(١٠-١)}} \quad ٠,٤٢٤٢٦$$

ثانيا : الخطأ المعياري للمقدر ب (ع ب)

$$ع_{ب} = ع_{ص.س} \sqrt{\frac{١}{(ن-١)ع^{٢}_{س}}}$$

$$= ٠.٤٦٦ \sqrt{\frac{١}{(١٠-١)(٦.٦٧٨)}} = ٠.٠٥٩٤٦$$

٣/٢/٦ مؤشرات جودة النموذج واختبار صلاحيته

أحد الأغراض الرئيسية من تحليل الانحدار ، استخدام نموذج الانحدار في التنبؤ ، وتتوقف نتائج التنبؤ على جودة النموذج . وهناك بعض المقاييس الاحصائية التي يمكن الاستدلال منها على جودة النموذج من ناحية ، وبعض الاختبارات الإحصائية التي يستدل منها على صلاحية او عدم صلاحية النموذج في تمثيل البيانات من ناحية أخرى .

ويكون التنبؤ تام إذا كانت كل قيمة من القيم الفعلية واقعة على الخط المستقيم الموفق ، أي عندما $ص_{ر} = \hat{ص}_{ر}$ ، عند جميع قيم (ر = ١، ٢، ....، ن) ويستدل من ذلك على ان الخطأ العشوائي $د_{ر}$ ، وهو يمثل انحراف القيمة الفعلية عن القيمة المتنبأ بها تساوي صفرا عند كل قيمة من قيم المتغير المستقل $س_{ر}$. ومعنى ذلك يكون التنبؤ تام ، إذا كان مجموع مربعات الأخطاء العشوائية ونرمز له بالرمز " م . م . خ " مساوية للصفر .

وعلى العكس من ذلك ، التنبؤ يكون غير جيد إذا كان مجموع مربعات الأخطاء العشوائية كبير . وهناك عاملان يساهمان في زيادة قيمة مجموع مربعات

الأخطاء (م . م . خ) هما :

١ ـ مقدار انحرافات القيم الفعلية عن الخط المستقيم الموفق ، حيث أن :

م . م . خ = مجـ (ص − ص̂)^{٢}

فكلما كان انحرافات القيم الفعلية عن الخط المستقيم المقدر كبير ، زاد مجموع مربعات الأخطاء ، ويستدل من ذلك على أن استخدام النموذج المقدر سوف ينتج عنه تنبؤ غير جيد .

٢ ـ إن قيمة مجموع مربعات الأخطاء العشوائية تتأثر بانحرافات القيم الفعلية عن وسطها الحسابي ، بمعنى أنها تتأثر بقيمة تباين القيم الفعلية ع^{٢}ص ، حيث أن هناك علاقة بين تباين الأخطاء العشوائية ع^{٢}ص.س ، وتباين القيم الفعلية ع^{٢}ص يمكن استنتاجها كما يأتي :

م . م . خ = مجـ (ص − ص̂)^{٢}

= مجـ (ص − (أ + ب س))^{٢} = مجـ (ص − (صَ − ب سَ + ب س))^{٢}

= مجـ ((ص − صَ) − ب (س − سَ))^{٢}

= مجـ (ص − صَ)^{٢} − ٢ ب مجـ (س − سَ)(ص − صَ) + ب^{٢} مجـ (س − سَ)^{٢}

ومن المعادلة (٦ ـ ٢) نجد أن :

مجـ (س − سَ)(ص − صَ) = ب مجـ (س − سَ)^{٢}

إذا : م . م . خ = مجـ (ص − صَ)^{٢} − ٢ ب^{٢} مجـ (س − سَ)^{٢} + ب^{٢} مجـ (س − سَ)^{٢}

= مجـ (ص − صَ)^{٢} − ب^{٢} مجـ (س − سَ)^{٢}

= (ن − ١) ع^{٢}ص − ب^{٢} (ن − ١) ع^{٢}س = (ن − ١) (ع^{٢}ص − ب^{٢} ع^{٢}س)

وبقسمة الطرفين على (ن − ٢) نجد أن :

$$\frac{\text{م . م . خ}}{(\text{ن} - ٢)} = \frac{(\text{ن} - ١)}{(\text{ن} - ٢)} (\text{ع}^{٢}_{\text{ص}} - \text{ب}^{٢} \text{ع}^{٢}_{\text{س}})$$

$$\text{ع}^2_{\text{ص.س}} = \frac{(\text{ن}-1)}{(\text{ن}-2)} \left(\text{ع}^2_{\text{ص}} - \text{ب}^2\ \text{ع}^2_{\text{س}}\right) \qquad (٦-٧)$$

أي أن تباين الخطأ العشوائي ، والذي يعكس انحرافات القيم الفعلية عن القيم المتنبأ بها يزداد بزيادة تباين القيم الفعلية ، والذي يعكس انحرافات القيم عن وسطها الحسابي .

ومن ثم يعتبر تباين الخطأ العشوائي أحد المؤشرات الإحصائية التي يستدل منها على جودة النموذج ، وفي المثال السابق نجد أن :

$$\text{ب} = ٠.٧٢ \quad ، \quad \text{ع}^2_{\text{س}} = ٦.٦٧٨$$

$$\text{ع}^2_{\text{ص}} = \frac{\text{مجـ ص}^2 - \text{ن}\ \bar{\text{ص}}^2}{(\text{ن}-1)} = \frac{٢٩٣ - (١٠)(٥.١)^2}{(١٠-١)} = ٣.٦٥٦$$

إذاً :

$$\text{ع}^2_{\text{ص.س}} = \frac{(١٠-١)}{(١٠-٢)} \left(٣.٦٥٦ - (٠.٧٢)^2 (٦.٦٧٨)\right) = ٠.٢١٨٣٩$$

ويلاحظ أن تباين الخطأ العشوائي صغير ، بالنسبة إلى تباين القيم الفعلية ، مما يدل على أن النموذج المقدر هو تمثيل جيد للبيانات .

هناك مؤشر آخر ، يمكن الاستدلال منه على جودة النموذج ، يعتمد هذا المقياس على نسبة مساهمة النموذج المقدر في تفسير الاختلافات في المتغير التابع ص ، ويمكن عرضه على النحو التالي :

الاختلاف في المتغير التابع ص ، ويعبر عنه بانحرافات القيم عن وسطها الحسابي : $(\text{ص} - \bar{\text{ص}})$ يمكن تقسيمه إلى جزأين هما :

الاختلاف بسبب الانحدار (ص̂ - ص̄) ، وهو الجزء المفسر للاختلاف في المتغير التابع ، ومرجعه إلى المتغير المستقل س ، وهو يعبر عن انحرافات القيم المقدرة عن وسطها الحسابي .

الاختلاف بسبب الأخطاء العشوائية (ص - ص̂) ، وهو الجزء الغير مفسر للاختلاف في المتغير التابع ، ومرجعه الخطأ العشوائي ، ويعبر أيضا عن انحرافات القيم عن وسطها الحسابي وقيمته صفراً .

أي أن :

الاختلاف الكلي = الاختلاف بسبب الانحدار + الاختلاف بسبب الأخطاء

(ص - ص̄) = (ص̂ - ص̄) + (ص - ص̂)

فإذا رمزنا إلى مجموع مربعات الاختلافات الكلية بالرمز " م . م . ك " ، ومجموع مربعات الاختلافات بسبب الانحدار بالرمز " م . م . ح " ، وسبق أن رمزنا لمجموع مربعات الاختلاف بسبب الأخطاء بالرمز " م . م . خ " ، فإنه يمكن إثبات أن :

$$\left.\begin{array}{c} \text{م . م . ك} = \text{م . م . ح} + \text{م . م . خ} \\ \text{مجـ } (\text{ص} - \bar{\text{ص}})^2 = \text{مجـ } (\hat{\text{ص}} - \bar{\text{ص}})^2 + \text{مجـ } (\text{ص} - \hat{\text{ص}})^2 \\ \text{مجموع المربعات الكلي} = \text{مجموع مربعات الانحدار} + \text{مجموع مربعات الأخطاء} \end{array}\right\} \quad (٨-٦)$$

فإذا كان الاختلاف بسبب الانحدار كبير ، دل ذلك على جودة النموذج ، ويعني ذلك أن المتغير المستقل س يساهم بجزء أكبر في تفسير الاختلاف الكلي . وهذا الجزء يمكن حساب نسبته المئوية ، ويطلق عليه اسم " معامل التحديد " ، أو مربع معامل الارتباط R Square ، ونرمز له بالرمز $ر^2$.

أي أن :

$$\text{معامل التحديد } (ر^2) = \frac{\text{م . م . ح}}{\text{م . م . ك}} = \frac{\text{مجـ}(\hat{ص} - \bar{ص})^2}{\text{مجـ}(ص - \bar{ص})^2} \qquad (٦-٩)$$

ويدل معامل التحديد ($ر^2$) على نسبة الجزء المفسر للاختلافات الكلية ، ويرجع إلى المتغير المستقل س . وأما النسبة الباقية (١ - $ر^2$) تمثل نسبة الجزء الغير مفسر ، ويرجع إلى الأخطاء العشوائية .

يعتمد حساب معامل التحديد ($ر^2$) على مجموع المربعات . ويمكن تبسيط العمليات الحسابية المستخدمة في حساب هذه المجاميع على النحو التالي :

مجموع المربعات الكلي (م . م . ك) = مجـ (ص - $\bar{ص}$)^2

= (ن - ١) $ع^2_ص$ = مجـ $ص^2$ - ن $\bar{ص}^2$

مجموع مربعات الانحدار (م . م . ح) = مجـ ($\hat{ص}$ - $\bar{ص}$)^2

= $ب^2$ (ن - ١) $ع^2_س$ = $ب^2$ (مجـ $س^2$ - ن $\bar{س}^2$)

مجموع مربعات الأخطاء (م . م . خ) = م . م . ك - م . م . ح

ففي المثال السابق نجد أن :

ب = ٠,٧٢ ، $ع^2_س$ = ٣,٦٥٦ ، $ع^2_ص$ = ٦,٦٧٨

مجموع المربعات الكلي (م . م . ك) = مجـ (ص - $\bar{ص}$)^2 = (ن - ١) $ع^2_ص$

= (١٠ - ١) (٣,٦٥٦) = ٣٢,٩٠٤

مجموع مربعات الانحدار (م . م . ح) = مجـ ($\hat{ص}$ - $\bar{ص}$)^2 = $ب^2$ (ن - ١) $ع^2_س$

= $(٠,٧٢)^2$ (١٠ - ١) (٦,٦٧٨) = ٣١,١٥٧

مجموع مربعات الأخطاء (م . م . خ) = م . م . ك - م . م . ح = ٣٢,٩٠٤ - ٣١,١٥٧ = ١,٧٤٧

$$\text{معامل التحديد } (ر^2) = \frac{\text{م . م . ح}}{\text{م . م . ك}} = \frac{٣١,١٥٧}{٣٢,٩٠٤} = ٠,٩٤٧$$

أي أن الدخل العائلي (س) يفسر ٩٤,٧٪ من الاختلافات في المتغير التابع ، والنسبة الغير مفسرة قيمتها ٥,٣٪ وترجع إلى الأخطاء العشوائية . وحيث أن قيمة معامل التحديد(٠,٩٤٧) كبيرة ، يستدل منه على أن النموذج المقدر جيد .

اختبار صلاحية النموذج

يمكن استخدام بعض الاختبارات الإحصائية في اختبار صلاحية النموذج في تمثيل البيانات تمثيلا جيدا ، ومن ثم استخدامه في التنبؤ . فمن المعلوم أن مجموع المربعات الكلي (م . م . ك) أمكن تقسيمه إلى جزأين هما مجموع مربعات الانحدار (م . م . ح) ، ومجموع مربعات الأخطاء العشوائية (م . م . خ) ، وكل مجموع منها يمثل مجموع مربعات لانحرافات القيم عن وسطه الحسابي ؛ وإذا قسم كل مجموع على درجات حريته ، يمكن الحصول على التباين الخاص به ، أو ما يسمى بمتوسط المربعات ، وذلك كما يلي .

١ – التباين الكلي (متوسط المربعات الكلي) ، وهو ($ع^2_{ص}$)

درجات حريته = (ن – ١)

متوسط المربعات الكلي($ع^2_{ص}$) = مجـ (ص – $\bar{ص}$) ^{2} ÷ (ن - ١)

أي أن متوسط المربعات الكلي هو تباين المتغير التابع ص .

٢ - تباين الانحدار (متوسط مربعات الانحدار) ، ويرمز له بالرمز ($ع^2_{ح}$)

درجات حريته = عدد المتغيرات المستقلة في النموذج = ١

متوسط مربعات الانحدار ($ع^2_{ح}$) = مجـ ($\hat{ص}$ – $\bar{ص}$) ^{2} ÷ ١

٣ - تباين الخطأ العشوائي (متوسط مربعات الأخطاء) ، وهو ($ع^2_{ص.س}$)

درجات حريته = (ن - عدد المتغيرات المستقلة - ١) = (ن - ١ - ١) = (ن - ٢)

متوسط مربعات الأخطاء (ع^{٢} ص.س) = مجـ (ص - صَ)^{٢} ÷ (ن - ٢)

ومن ثم ، نجد أن النسبة :

$$\text{ف} = \frac{\text{ع}^{٢}_{\text{م}}}{\text{ع}^{٢}_{\text{ص.س}}} = \frac{\text{متوسط مربعات الانحدار}}{\text{متوسط مربعات الأخطاء}} \qquad (٦-١٠)$$

تتبع توزيع ف بدرجات حرية بسط = ١ ، ودرجات حرية مقام = (ن-٢) ، ومن ثم يمكن استخدام هذه النسبة في اختبار صلاحية النموذج ، وفيما يلي خطوات إجراء الاختبار :

١ - صياغة الفرض العدمي والفرض البديل وهما :

ف. : النموذج غير مناسب . ف١ : النموذج مناسب .

٢ - تحديد التوزيع الاحتمالي المستخدم :

التوزيع الاحتمالي المستخدم في الاختبار هو توزيع ف بدرجات حرية بسط = ١ ، ودرجات حرية مقام = (ن - ٢) .

٣ - القيمة الحرجة ، ومناطق الرفض والقبول :

تستخرج القيمة الحرجة من جدول توزيع عند مستوى معنوية α ، ودرجات حرية (١ ، ن-٢) ، وهي ف(α ، ١ ، (ن - ٢)) ومنطقة الرفض هي المساحة أسفل منحنى توزيع ف على يمين هذه القيمة كما هو مبين بالشكل التالي :

ف(α ، ١ ، (ن - ٢))

٤ – الاحصاء المستخدم في الاختبار :

الاحصاء المستخدم في الاختبار هو النسبة ف العينية في المعادلة (٦ - ١٠)

٥ – اتخاذ القرار :

ويتوقف على مقارنة النسبة المحسوبة ف بالقيمة الجدولية ، أو حسب موقع قيمة ف المحسوبة ، إذا كانت ف المحسوبة في الخطوة ٤ تقع في منطقة الرفض ، أي على يمين القيمة الجدولية ، رفض الفرض العدمي ، وقبل الفرض البديل ، ويستدل من ذلك على أن النموذج (الخط المستقيم الموفق للبيانات) مناسب في تمثيل العلاقة بين المتغير التابع ص ، والمتغير المستقل س ، ومن ثم فهو صالح لاستخدامه في التنبؤ .

مثال (٦ - ٢)

في المثال (٦ - ١) : هل النموذج المقدر صالح في التنبؤ ؟ ، α = ٠,٠٥ .

الحل :

يتطلب اختبار صلاحية النموذج حساب متوسط مربعات الانحدار ، ومتوسط مربعات الأخطاء ، لحساب الاحصاء المستخدم في الاختبار ف ، ولحساب متوسط المربعات ، يجب اولا حساب مجموع المربعات ، وقد سبق حساب هذه المجاميع :

مجموع المربعات الكلي (ج . م . ك) = ٣٢,٩٠٤

مجموع مربعات الانحدار (م . ج . ح) = ٣١,١٥٧

مجموع مربعات الأخطاء (م . ج . خ) = ١,٧٤٧

ومن ثم يحسب متوسط المربعات كما يلي :

متوسط مربعات الانحدار ($ع^{٢}_{ح}$) = مجـ ($\hat{ص}$ – $\bar{ص}$) ^{٢} ÷ ١

= ٣١,١٥٧ ÷ ١ = ٣١,١٥٧

متوسط مربعات الأخطاء (ع² ص س) = مج (ص − صَ)² ÷ (ن − ٢)

= ١,٧٤٧ ÷ (١٠−٢) = ٠,٢١٨٣٧٥

ف = متوسط مربعات الانحدار / متوسط مربعات الأخطاء = ٣١,١٥٧ / ٠,٢١٨٣٧٥ = ١٤٢,٧

وفيما يلي خطوات الاختبار :

١ ـ الفرض العدمي ، الفرض البديل :

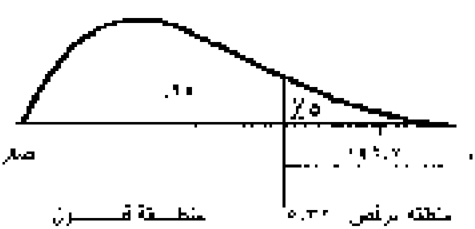
ف٠ : النموذج غير مناسب في تمثيل البيانات .

ف١ : النموذج مناسب في تمثيل البيانات .

٢ ـ التوزيع المستخدم هو توزيع ف بدرجات حرية بسط = ١ ، ودرجات حرية مقام = ٨

٣ ـ مناطق الرفض والقبول .

منطقة الرفض هي المساحة أسفل المنحنى على يمين ف (٠,٠٥ ، ١ ، ٨) = ٥,٣٢



٤ ـ الإحصاء المستخدم في الاختبار :

ف = ١٤٢,٧

٥ ـ القرار : بما أن ف المحسوبة (١٤٢,٧) تقع في منطقة الرفض ، فهي أكبر من القيمة الحرجة (٥,٣٢) لذا يرفض الفرض العدمي ف٠ ، ويقبل الفرض البديل ف١ . ويستدل من ذلك على أن النموذج الذي يأخذ شكل الخط المستقيم ، هو الأفضل

تمثل للعلاقة بين الانفاق كمتغير تابع ، والدخل كمتغير مفسر .

جدول تحليل التباين

يمكن تلخيص النتائج المتعلقة بقياس جودة النموذج ، واختبار صلاحيته في جدول تحليل تباين يأخذ الشكل الآتي :

جدول (٦-١)

جدول تحليل التباين

المصدر	درجات الحرية	مجموع المربعات	متوسط المربعات	ف
الانحدار (س)	١	م . م . ح = مجـ (ص^ - ص̄)^2	ع$^2_{ح}$ = م . م . ح / ١	ع$^2_{ح}$ / ع$^2_{ص.س}$
الخطأ	ن − ٢	م . م . خ = مجـ (ص - ص^)^2	ع$^2_{ص.س}$ = م . م . خ / (ن - ٢)	
الكلي	ن − ١	م . م . ك = مجـ (ص - ص̄)^2		

ومن هذا الجدول يمكن الحصول على مؤشرات جودة النموذج ، وكذلك الإحصاء المستخدم في اختبار صلاحية النموذج في التنبؤ، وذلك على النحو التالي :

الخطأ المعياري للتقدير (ع$_{ص.س}$) = $\sqrt{\text{متوسط مربعات الأخطاء}}$

معامل التحديد (ر^2) = م . م . خ ÷ م . م . ك

الإحصاء المستخدم في الاختبار (ف) = ع$^2_{ح}$ ÷ ع$^2_{ص.س}$

وبالتطبيق على بيانات مثال (٦-١) ، يأخذ الجدول الشكل التالي :

جدول تحليل التباين

المصدر	درجات الحرية	مجموع المربعات	متوسط المربعات	ف
الانحدار (س)	١	م.م.ح = ٣١.١٥٧	$ع^2_ح = \frac{٣١.١٥٧}{١} = ٣١.١٥٧$	$\frac{٣١.١٥٧}{٠.٢١٨٣٧٥} = ١٤٢.٧$
الخطأ	٨	م.م.خ = ١.٧٤٧	$ع^2_{ص.س} = \frac{١.٧٤٧}{٨} = ٠.٢١٨٣٧٥$	
الكلي	٩	م.م.ك = ٣٢.٩٠٤		

٤/٣/٦ اختبار معنوية معامل الانحدار في المجتمع (β)

يقصد باختبار معنوية معامل الانحدار في المجتمع (β) ، اختبار أن هذا المعامل يختلف عن الصفر ، ويستدل منه على أن أثر المتغير المستقل س على المتغير التابع ص معنوي ، وأن وجوده في النموذج يحسن في قدرته التنبؤية . ومن هنا يأخذ الفرض العدمي ، والفرض البديل الصورة التالية :

$ف_٠ : β = صفر$ $\qquad\qquad$ $ف_١ : β ≠ صفر$

ولاختبار هذا الفرض يمكن استخدام توزيع معاينة مقدر (β) ، وهو (ب) ، وكما سبق ، فإن المقدر ب له توزيع طبيعي توقعه β ، وخطأه المعياري هو ($ع_ب$) ، ومن ثم يأخذ إحصاء الاختبار الصورة التالية :

$$ت = \frac{ب - β}{ع_ب}$$

وهذا الإحصاء له توزيع ت بدرجات حرية = (ن – ٢) ، وهي مساوية

لدرجات حرية الخطأ العشوائي الموجودة في جدول تحليل التباين .

وتحت صحة الفرض العدمي ف. : β = صفر يأخذ الاحصاء ت الصورة التالية :

$$ت = \frac{ب}{ع_ب} \qquad (٦-١١)$$

وفيما يلي خطوات اجراء الاختبار :

١ – صياغة الفرض العدمي والفرض البديل :

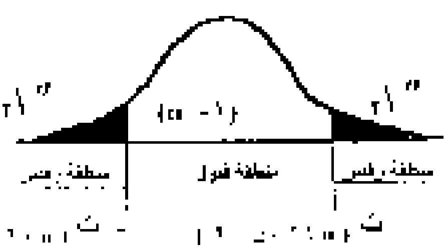
ف. : β = صفر ف١ : $\beta \neq$ صفر

٢ – التوزيع المستخدم في الاختبار هو توزيع ت .

٣ – وتحديد مناطق الرفض والقبول .

تستخرج القيمة الحرجة من جدول توزيع ت عند درجات حرية (ن – ٢) ، ومستوى معنوية (α) ، وهي ± ت لأن الاختبار في الاتجاهين .

والشكل التالي يبين مناطق الرفض والقبول .



٤ – الإحصاء المستخدم في الاختبار هو ت :

$$ت = \frac{ب}{ع_ب}$$

٥ – القرار : إذا وقعت قيمة الإحصاء ت داخل منطقة الرفض ، يقبل الفرض البديل ف١ : $\beta \neq$ صفر ، ويدل ذلك على أن أثر المتغير المستقل على المتغير التابع معنوي

مثال (٦ - ٢)

من بيانات مثال (٦ - ١) اختبر معنوية معامل الانحدار في المجتمع (α = ٥٪) ، ثم قدر له فترة ثقة ٩٥٪ .

الحل :

١ - صياغة الفرض العدمي والفرض البديل :

ف٠ : β = صفر   ف١ : β ≠ صفر

٢ - التوزيع المستخدم في الاختبار هو توزيع ت .

٣ - تحديد مناطق الرفض والقبول .

تستخرج القيمة الحرجة من جدول توزيع ت عند درجات حرية (٨) ، ومستوى معنوية (٠,٠٥ ÷ ٢ = ٠,٠٢٥) وهي : ت(٨ ، ٠,٠٢٥) = ± ٢,٣٠٦

٤ - الإحصاء المستخدم في الاختبار هو ت :

$$ت = \frac{ب}{ع_ب} = \frac{٠,٧٢}{٠,٠٥٩٤٦} = ١٢,١٠٩$$

٥ - القرار : تقع قيمة إحصاء الاختبار (١٢,١٠٩) في منطقة الرفض اليمنى ، لذا يرفض الفرض العدمي ، ويقبل الفرض البديل . ويستدل من ذلك على أن المتغير المستقل له أثر معنوي ، وأنه يحسن القدرة التنبوية للنموذج .

ولتقدير فترة ثقة ، يتبع نفس الأسلوب المتبع في تقدير فترة ثقة للوسط الحسابي في المجتمع ، وتكون فترة ثقة ٩٥٪ لمعامل الانحدار في المجتمع(β) هو :

التقدير بنقطة ± (الحد الأقصى للخطأ المسموح به)

ب ± (ع ب) (ت (٠,٠٢٥ ، ن-٢)) = ب ± (ع ب) (ت (٠,٠٢٥ ، ٨))

= ٠,٧٢ ± (٠,٠٥٩٤٦) (٢,٣٠٦) = (٠,٧٢ ± ٠,١٣٧١١)

= (٠,٥٨٢٨٩ ، ٠,٨٥٧١١)

أي أننا على ثقة ٩٥٪ بأن معامل الانحدار في المجتمع β يتراوح بين (٠,٥٨٢ ، ٠,٨٥٧) .

مثال (٦-٤)

الجدول التالي يبين كمية الإنتاج من الخيوط المغزولة بالكيلو (ص) في الوردية في أحد مصانع النسيج ، وسنوات الخبرة (س) لعينة من ١٢ عامل .

العامل	١	٢	٣	٤	٥	٦	٧	٨	٩	١	١١	١٢
ص	٥	٥	٥	٦	٧	٩	٨	١٠	١٠	١١	١١	٩
س	٤	٤	٥	٣	٦	٦	٤	٧	٩	٩	٨	٧

والمطلوب :

١- استخدم طريقة المربعات الصغرى لتقدير خط انحدار ص / س ، بافتراض أن العلاقة بين ص ، س تأخذ شكل خط مستقيم .

٢- كون جدول تحليل التباين ، ثم اختبر صلاحية النموذج ، α = ٠,٠١

٣- احسب معامل التحديد ، وما هو مدلوله ؟

٤- اختبر معنوية معامل الانحدار في المجتمع (β) ، α = ٠,٠٥

الحل :

نبدأ بتكوين جدول لحساب المجاميع ، كما يلي .

العامل رقم	كمية الإنتاج (ص)	عدد سنوات الخبرة (س)	س ص	س^{٢}	ص^{٢}
١	٥	٤	٢٠	١٦	٢٥
٢	٥	٤	٢٠	١٦	٢٥
٣	٥	٥	٢٥	٢٥	٢٥
٤	٦	٣	١٨	٩	٣٦
٥	٧	٦	٤٢	٣٦	٤٩
٦	٩	٦	٥٤	٣٦	٨١
٧	٨	٤	٣٢	١٦	٦٤
٨	١٠	٧	٧٠	٤٩	١٠٠
٩	١٠	٩	٩٠	٨١	١٠٠
١٠	١١	٩	٩٩	٨١	١٢١
١١	١١	٨	٨٨	٦٤	١٢١
١٢	٩	٧	٦٣	٤٩	٨١
المجموع	٩٦	٧٢	٦٢١	٤٧٨	٨٢٨
	مجـ ص	مجـ س	مجـ س ص	مجـ س^{٢}	مجـ ص^{٢}

١ـ تقدير خط انحدار ص / س :

تقدير β هو ب ، ويحسب بتطبيق المعادلة التالية :

$$\text{ب} = \frac{\text{ن مجـ س ص} - \text{مجـ س مجـ ص}}{\text{ن مجـ س}^{٢} - (\text{مجـ س})^{٢}} = \frac{(١٢)(٦٢١) - (٧٢)(٩٦)}{(١٢)(٤٧٨) - (٧٢)^{٢}}$$

$$= \frac{٥٤٠}{٥٥٢} = ٠,٩٧٨٣$$

وتقدير α هو أ ويحسب بتطبيق المعادلة التالية :

$$\text{أ} = \bar{\text{ص}} - \text{ب}\,\bar{\text{س}} = (٩٦ \div ١٢) - ٠,٩٧٨٣\,(٧٢ \div ١٢) = ٢,١٣٠٢$$

معادلة انحدار ص على س : ص̂ = ٢,١٣٠٢ + ٠,٩٧٨٣ س

وتدل المعادلة السابقة على الآتي :

عندما تكون سنوات الخبرة مساوية للصفر ، يكون انتاج العامل في الوردية ٢,١٣ كيلو جرام ، كما أن كل زيادة في سنوات الخبرة سنة واحدة ، تزيد كمية الإنتاج بمقدار ٠,٩٨٧ كيلو جرام .

٢ - تكوين جدول تحليل التباين :

حساب مجموع المربعات ، ومتوسط المربعات ، والإحصاء (ف) :

مجموع المربعات الكلي (م . م . ك) = مجـ ص² - ن ص̄² = ٨٢٨ - ١٢ (٨)² = ٦٠

درجات الحرية = (ن - ١) = (١٢ - ١) = ١١

مجموع المربعات الانحدار (م . م . ح) = ب² (مجـ س² - ن س̄²)

= (٠,٩٧٨٣)² (٤٧٨ - ١٢ (٦)²) = ٤٤,٠٢٥٢٦

درجات الحرية = عدد المتغيرات المستقلة = ١

متوسط مربعات الانحدار (ع²ح) = (م . م . ح) ÷ ١

= ٤٤,٠٢٥٢٦ ÷ ١ = ٤٤,٠٢٥٢٦

مجموع مربعات الأخطاء (م . م . خ) = (م . م . ك) - (م . م . ح)

= ٦٠ - ٤٤,٠٢٥٢٦ = ١٥,٩٧٤٧٤

درجات الحرية = (ن - عدد المتغيرات المستقلة - ١) = (ن - ٢) = (١٢ - ٢) = ١٠

متوسط مربعات الأخطاء (ع²ص.س) = (م . م . خ) ÷ (ن - ٢)

= ١٥,٩٧٤٧٤ ÷ ١٠ = ١,٥٩٧٤٧٤

احصاء الاختبار (ف) = (ع²ح) ÷ (ع²ص.س)

= ٤٤,٠٢٥٢٦ ÷ ١,٥٩٧٤٧٤ = ٢٧,٥٥٩

جدول تحليل التباين

المصدر	درجات الحرية	مجموع المربعات	متوسط المربعات	ف
الانحدار	١	٤٤,٠٢٥٢٦	٤٤,٠٢٥٢٦	٢٧,٥٥٩
الخطأ	١٠	١٥,٩٧٤٧٤	١,٥٩٧٤٧٤	
المجموع	١١	٦٠		

اختبار صلاحية النموذج :

- الفرض العدمي والفرض البديل :

$ف_0$: النموذج غير مناسب $ف_1$: النموذج مناسب

- التوزيع المستخدم توزيع ف بدرجات حرية بسط =١ ، ودرجات حرية مقام = ١٠

- مناطق الرفض والقبول :

(α=٠,٠١) ، منطقة الرفض هي ف > ($ف_{(٠,٠١ ، ١ ، ١٠)}$ = ١٠,٠٤)

- الإحصاء المستخدم في الاختبار : ف = ٢٧,٥٥٩

- القرار : بما أن ف المحسوبة (٢٧,٥٥٩) أكبر من القيمة الحرجة(١٠,٠٤) لذا يرفض الفرض العدمي $ف_0$ ، ويقبل الفرض البديل $ف_1$ ، ويدل ذلك على أن خط الانحدار المقدر ، هو أفضل تمثيل للعلاقة بين كمية الإنتاج كمتغير تابع، وعدد سنوات الخبرة كمتغير مستقل .

٣ – حساب معامل التحديد :

معامل التحديد (ر^2) = (م . م . ح) ÷ (م . م . ك) = ٤٤,٠٢٥٩٦ ÷ ٦٠ = ٠,٧٣٤

أي أن عدد سنوات الخبرة كمتغير مستقل ، يفسر ٧٣,٤٪ من الاختلافات الكلية في المتغير التابع ، وأن النسبة الغير مفسرة (٢٦,٦ ٪) ترجع لأخطاء عشوائية أو لأسباب أخرى .

٤ – اختبار معنوية معامل الانحدار في المجتمع (β)

- الفرض العدمي والفرض البديل :

ف. : β = صفر   ف$_1$: β ≠ صفر

- التوزيع المستخدم في الاختبار هو توزيع ت بدرجات حرية (ن - ٢) = ١٠

- تحديد مناطق الرفض والقبول (α = ٠,٠٥)

منطقتي الرفض هما : ت < ت$_{(٠,٠٢٥ ، ١٠)}$ = -٢,٢٢٨ ، ت > ٢,٢٢٨

- الإحصاء المستخدم في الاختبار هو ت :

$$ت = \frac{ب}{ع_ب}$$

$$ع_ب = ع_{ص.س}\sqrt{\frac{1}{(ن-1)ع_س^2}} = ع_{ص.س}\sqrt{\frac{1}{مجـ\,س^2 - ن\,\bar{س}^2}}$$

$$= ١,٠٥٩٧٤٧\sqrt{\frac{1}{(١٢,٤٧٨)(٦)^2}} = ٠,١٨٦٣٥$$

$$\text{إذاً :}\quad \text{ت} = \frac{٠,٩٧٨٣}{٠,١٨٦٣٥} = ٥,٢٥$$

ـ القرار : بما أن قيمة إحصاء الاختبار (٥,٢٥) > ٢,٢٢٨ ، لذا يرفض الفرض العدمي ، ويستدل من ذلك على أن الخبرة لها أثر معنوي على كمية الإنتاج .

٣/٦ الاستدلال الإحصائي للارتباط الخطي البسيط

يعكس معامل الارتباط نوع وقوة العلاقة بين متغيرين ، كما يتصف هذا المعامل ببعض الخصائص الرياضية التي تربطه بخط الانحدار المستقيم .

١/٣/٦ خصائص معامل الارتباط

يحسب معامل الارتباط من بيانات عينة من أزواج القيم بالمعادلة التالية :

$$\text{معامل الارتباط (ر)} = \frac{\text{مجـ}(\text{س} - \bar{\text{س}})(\text{ص} - \bar{\text{ص}})}{\sqrt{\text{مجـ}(\text{س} - \bar{\text{س}})^2 \;\; \text{مجـ}(\text{ص} - \bar{\text{ص}})^2}} \qquad (١٢\text{-}٦)$$

ويمكن تبسيط المعادلة السابقة لتسهيل العمليات الحسابية ، حيث تكتب معادلة الارتباط كما يلي :

$$\text{ر} = \frac{\text{ن مجـ س ص} - \text{مجـ س مجـ ص}}{\sqrt{\left(\text{ن مجـ س}^2 - (\text{مجـ س})^2\right)\left(\text{ن مجـ ص}^2 - (\text{مجـ ص})^2\right)}} \qquad (١٣\text{-}٦)$$

كما تأخذ العلاقة بين معامل الارتباط (ر) ، ومعامل الانحدار (ب) الصورة التالية:

$$ر = ب \times \frac{ع_س}{ع_ص} \qquad (14.6)$$

كما أن معامل الارتباط الخطي البسيط هو الجذر التربيعي لمعامل التحديد، وتحدد إشارته حسب إشارة معامل الانحدار ب . أي أن

$ر = \pm \sqrt{}$ معامل التحديد ($ر^2$)   ويأخذ اشارة معامل الانحدار (ب)

ويتصف معامل الارتباط بالخصائص الرياضية التالية :

١ . تتراوح قيمة معامل الارتباط بين +١ ، -١ أي أن : $-1 \le ر \le 1$.

٢ . معامل الارتباط يأخذ قيمة عددية لا تعبر عن وحدات قياس .

٣ . قيمة معامل الارتباط تكون (موجب - صفر - سالب) حسب قيمة معامل الانحدار (موجب - صفر - سالب) . ويظهر ذلك من العلاقة بين (ر ، ب) في المعادلة (٦ . ١٤)

يقيس معامل الارتباط العلاقة بين المتغيرين س ، ص . ويفترض أن المتغير س متغير عشوائي . لا يأخذ قيم ثابتة ومحددة ، على عكس افتراض تحليل الانحدار حول قيم س . وتفسر العلاقة بين س ، ص على النحو التالي :

١ - إذا كان معامل الارتباط سالب دل ذلك على وجود علاقة عكسية ، وتزداد العلاقة قوة كلما اقترب ر من -١ .

٢ - إذا كان معامل الارتباط موجب دل ذلك على وجود علاقة طردية ، وتزداد العلاقة قوة كلما اقترب ر من +١ .

٣ - إذا كان معامل الارتباط يقترب من الصفر ، دل ذلك على أن العلاقة بين س ، ص علاقة ضعيفة .

إذا كان معامل الارتباط (ر) = ± ١ دل ذلك على وجود ارتباط تام .

٤- ومعنى ذلك ان التنبؤ يكون تام ، وأن جميع قيم القيم الفعلية ص تقع على خط الانحدار .

٥- إذا كان معامل الارتباط (ر) = صفر دل ذلك على انعدام العلاقة .

وفيما يلي شكل نوع العلاقة بين المتغير (س ، ص) ، وقوتها .

عكسية تامة	عكسية قوية	عكسية ضعيفة	طردية ضعيفة	طردية قوية	طردية تامة

-١ ◄ ر ► صفر ◄ ر ► $+١$

مثال (٦-٦)

من بيانات مثال (٦-٥) ، احسب معامل الارتباط ، وما هو مدلوله ؟

الحل معامل الارتباط هو :

$$\text{معامل الارتباط (ر)} = \frac{\text{ن مجـ س ص} - \text{مجـ س مجـ ص}}{\sqrt{\left(\text{ن مجـ س}^2 - (\text{مجـ س})^2\right)\left(\text{ن مجـ ص}^2 - (\text{مجـ ص})^2\right)}}$$

من بيانات مثال (٦-٥) تم ايجاد المجاميع المطلوبة لحساب هذا المعامل ، حيث ان :

ن = ١٢ مجـ س = ٧٢ مجـ ص = ٩٦ مجـ س ص = ٦٢١

مجـ س^2 = ٤٧٨ مجـ ص^2 = ٨٢٨

$$\text{معامل الارتباط (ر)} = \frac{(١٢)(٦٢١) - (٧٢)(٩٦)}{\sqrt{\left((١٢)(٤٧٨) - (٧٢)^2\right)\left((١٢)(٨٢٨) - (٩٦)^2\right)}}$$

$$= \frac{٥٤٠}{\sqrt{(٥٥٢)(٧٢٠)}} = \frac{٥٤٠}{\sqrt{٣٩٧٤٤٠}} = ٠.٨٥٧$$

يدل معامل الارتباط على وجود علاقة طردية قوية بين سنوات الخبرة (س) وكمية الإنتاج (ص) .

٦/٣/٢ اختبار معنوية معامل الارتباط في المجتمع (ρ)

يقصد باختبار معنوية معامل الارتباط في المجتمع ، ويرمز له بالرمز ρ (رو) ، اختبار أن معامل الارتباط في المجتمع ، الذى سحبت منه عينة من أزواج القيم (س ، ص) وحسب لها معامل الارتباط ر يساوى صفرا ، ومن ثم يأخذ الفرض العدمي والفرض البديل الصورة التالية :

ف. : ρ = صفر ف، : $\rho \neq$ صفر

والإحصاء المستخدم في الاختبار هو :

$$ت = \frac{ر\sqrt{(ن - ٢)}}{\sqrt{(١ - ر^{٢})}} \qquad (٦-١٥)$$

وهذا الإحصاء يتبع توزيع ت بدرجات حرية (ن - ٢) ، ولاختبار فرض معنوية معامل الارتباط في المجتمع يتبع نفس الخطوات المتبعة في اختبار معنوية معامل الانحدار وهي :

١ – صياغة الفرض العدمي والفرض البديل

ف. : ρ = صفر ف، : $\rho \neq$ صفر

٢ – التوزيع المستخدم هو توزيع ت بدرجات حرية (ن-٢)

٣ – مناطق الرفض والقبول : الاختبار في اتجاهين . توجد منطقتين لرفض الفرض العدمي هما : ت > ت (α ؛ ٢ ، ن – ١) ، ت < - ت (α ١ ، ن – ٢)

وتستخرج القيمة الحرجة من جدول توزيع ت عند درجات حرية (ن ـ ٢) ، ومستوى معنوية ($\alpha/2$) ، وهي $\pm ت_{(\alpha/2 ،\ ن-٢)}$ ، والشكل التالي يبين مناطق الرفض والقبول .

٤ ـ الإحصاء المستخدم في الاختبار هو ت :

$$ت = \frac{ر\sqrt{(ن - ٢)}}{\sqrt{(١ - ر^٢)}}$$

٥ القرار : إذا وقعت قيمة الإحصاء ت داخل منطقة الرفض ، يقبل الفرض البديل $ف_١$: $\rho \neq$ صفر ، ويدل ذلك على أن العلاقة بين (س ، ص) علاقة معنوية .

مثال (٦-٧)

في مثال (٦-٥) السابق استخدم معامل الارتباط في العينة (ر = ٠,٨٥٧) لاختبار معنوية معامل الارتباط في المجتمع ، α = ٠,٠٥ .

<u>الحل :</u>

ن = ١٢ ، ر = ٠,٨٥٧

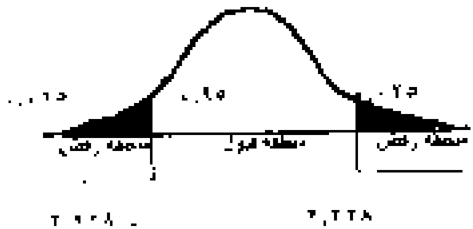
ـ الفرض العدمي والفرض البديل :

$ف_٠$: ρ = صفر $ف_١$: $\rho \neq$ صفر

ـ التوزيع المستخدم في الاختبار هو توزيع ت بدرجات حرية (ن ـ ٢) = ١٠

- تحديد مناطق الرفض والقبول (α = ٠,٠٥)

منطقتي الرفض هما : ت > ت$_{(٠,٠٢٥ ، ١٠)}$ = ٢,٢٢٨ ، ت < - ٢,٢٢٨



- الإحصاء المستخدم في الاختبار هو ت :

$$ت = \sqrt{\frac{ر^2 (ن - ٢)}{(١ - ر^2)}} = \sqrt{\frac{(٠,٨٥٧)^2 (١٢ - ٢)}{(١ - (٠,٨٥٧)^2)}} = ٥,٢٥٩$$

- القرار : تقع قيمة إحصاء الاختبار (٥,٢٥٩) في منطقة الرفض اليمنى ، لذا يرفض الفرض العدمي ، ويقبل الفرض البديل ، ويستدل من ذلك على معنوية معامل الارتباط بين سنوات الخبرة ، كمية الإنتاج .

٤/٦ اختبار معنوية إشارة معاملي الانحدار والارتباط

يمكن اختبار معنوية إشارة كلا المعاملين ، وفي هذه الحالة يكون الاختبار في اتجاه واحد ، ويجري هذا الاختبار كما يلي :

بالنسبة لمعامل الانحدار β : يأخذ الفرض العدمي والبديل الشكل التالي :

ف$_{٠}$: β = صفر ف$_{١}$: β > صفر

ويدل الفرض البديل على أن المتغير المستقل س له أثر معنوي موجب على المتغير التابع ص ، وفي هذه الحالة تقع منطقة الرفض جهة اليمين ، وتحدد بالقيمة

الجدولية ([illegible] ، ن - ٢) . وإذا كان المطلوب اختبار أن المتغير المستقل له أثر عكسي على المتغير التابع يمكن صياغة الفرض في الصورة التالية :

ف. : β = صفر   ف١ : β < صفر

وفي هذه الحالة تقع منطقة الرفض جهة اليسار ، وتحدد بالقيمة [illegible] ت [illegible] ([illegible] ، ن - ٢) .

والاختبار الإحصائي المستخدم في الاختبار هو ت المبين بالمعادلة (٦ - [illegible])

(ت = ب ÷ ع(ب)) .

بالنسبة لمعامل الارتباط : يمكن اختبار ما إذا كانت العلاقة بين س ، ص علاقة موجبة ، أو سالبة ، عن طريق إعادة صياغة الفروض كما يلي :

تصاغ الفروض في حالة اختبار معنوية إيجابية العلاقة كما يلي :

ف. : ρ = صفر   ف١ : ρ > صفر

وفي هذه الحالة تقع منطقة الرفض جهة اليمين ، وتحدد بالقيمة [illegible] ت [illegible] ([illegible] ، ن - ٢) .

وتصاغ الفروض في حالة اختبار معنوية سالبية العلاقة كما يلي :

ف. : ρ = صفر   ف١ : ρ < صفر

وفي هذه الحالة تقع منطقة الرفض جهة اليسار ، وتحدد بالقيمة [illegible] ت [illegible] ([illegible] ، ن - ٢) .

والاختبار الإحصائي المستخدم في الاختبار هو ت المبين بالمعادلة (٦ - ١٥)

(ت = ر √(ن - ٢) ÷ √(١ - ر٢)) .

مثال (٦ - ٨) :

من بيانات مثال (٦ - ٤) هل نؤيد أن : β > صفر [illegible]

الحل :

من بيانات المثال (٦-٤) تم الحصول على إحصاء الاختبار ت :

$$ت = \frac{ب}{ع_ب} = ٥,٢٥$$

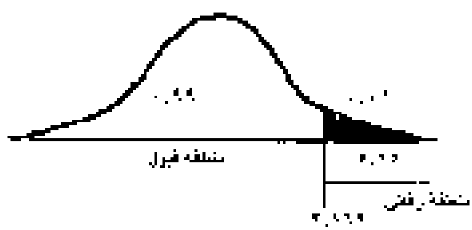
وفيما يلي خطوات الاختبار :

ـ الفرض العدمي والفرض البديل :

ف. : β = صفر   ف١ : β > صفر

ـ التوزيع المستخدم هو توزيع ت بدرجات حرية (ن-٢) =(١٢-٢) = ١٠

ـ مناطق الرفض والقبول : تقع منطقة الرفض جهة اليمين ، في المنطقة ت>ت (ع ، ن - ٢) = ت(٠,٠١ ، ١٠) ، أي المنطقة المحددة بـ ت > ٣.١٦٩ .



ـ إحصاء الاختبار : ت = ٥.٢٥

ـ القرار : بما أن قيمة إحصاء الاختبار (٥,٢٥) أكبر من القيمة الجدولية (٣.١٦٩) أي تقع في منطقة الرفض ، لذا لايمكن قبول الفرض العدمي ، ويدل ذلك على أن سنوات الخبرة لها أثر معنوي موجب على الكمية المنتجة .

مثال (٦-٩)

استخدم النتائج التالية ، وأجب عما هو مطلوب :

ن = ٢٥   صَ = ٤٣,٣٨٦ ـ ٢,٦٩٤س   عص=٦,٥٩٨٥   عس=١,٦٨٣٣

والمطلوب : ١- فسر معادلة الانحدار .

٢- احسب الخطأ المعياري للتقدير .

٣- هل تؤيد أن العلاقة بين س ، ص علاقة عكسية معنوية ؟ α = ٠,٠٥ .

الحل :

١- تفسير معادلة الانحدار .

الثابت (أ = ٣,٣٨٦) : يدل على أن قيمة ص عندما س = صفر هي ٤٣,٣٨٦

معامل الانحدار في العينة (ب = -٣,٦٩٤) : تدل إشارة هذا المعامل على أن المتغير المستقل س له أثر عكسي على المتغير التابع ص ، وتدل قيمة هذا المعامل على أنه كلما حدث زيادة في المتغير المستقل س بمقدار وحدة ، يترتب عليه انخفاض قيمة ص بمقدار ٣,٦٩٤ .

٢ - حساب الخطأ المعياري للتقدير .

هناك صيغ رياضية كثيرة لحساب هذا الخطأ ، ومنها الصيغة التالية :

$$ع_{ص.س} = \sqrt{\frac{(ن - ١)}{(ن - ٢)} \{ ع^٢_ص - ب^٢ ع^٢_س \}}$$

$$= \sqrt{\frac{(٢٥ - ١)}{(٢٥ - ٢)} \{ (٦,٥٩٨٥)^٢ - (٣,٦٩٤-)^٢ (١,٦٨٣٣)^٢ \}}$$

$$= ٢,٢٥٥٥$$

٣ - اختبار وجود علاقة عكسية معنوية بين س ، ص .

لاختبار هذا الفرض يتم حساب معامل الارتباط باستخدام الصيغة التالية :

$$ر = ب \{ ع_س \div ع_ص \} = (٣,٦٩٤-) \{ ١,٦٨٣٣ \div ٦,٥٩٨٥ \} = -٠,٩٤٢٠$$

وفيما يلي خطوات اختبار الفرض :

ـ الفرض العدمي والفرض البديل :

$$ف_٠ : \rho = \text{صفر} \qquad ف_١ : \rho \neq \text{صفر}$$

ـ التوزيع المستخدم هو توزيع ت بدرجات حرية (ن ـ ٢) = (٢٥ ـ ٢) = ٢٣

ـ مناطق الرفض والقبول : تقع منطقة الرفض جهة اليسار ، في المنطقة ت < ـ ٢,٠٦٩

ـ إحصاء الاختبار هو :

$$ت = \frac{ر\sqrt{ن - ٢}}{\sqrt{١ - ر^{٢}}} = \frac{(-٠,٩٤٢٦)\sqrt{٢٥ - ٢}}{\sqrt{١ - (-٠,٩٤٢٦)^{٢}}} =$$

$$= -١٣,٤٦٦$$

ـ القرار : تقع قيمة احصاء الاختبار (ـ ١٣,٤٦٦) في منطقة الرفض ، لذا يرفض الفرض العدمي ويقبل الفرض البديل ، ويستدل من ذلك على وجود علاقة عكسية معنوية بين س ، ص .

أوجه التشابه بين اختبار معنوية معامل الانحدار ومعامل الارتباط

في تحليل الانحدار والارتباط الخطي البسيط ، يعتبر اختبار معنوية كلا المعاملان وجهان لعملة واحدة ، فهناك تشابه بين اختبار معنوية معامل الانحدار β ، ومعامل الارتباط ρ ، وهذا التشابه نابع من العلاقة بين المعاملين الموضحة

بالمعادلة (٦-١٤) ، ومن ثم يعكس اختبار معنوية معامل الارتباط في المجتمع اختبار معنوية معامل الانحدار في المجتمع . اختبار أن (ρ = صفر) . هو أيضا اختبار ان (β = صفر) ، كما أن الإحصاء المستخدم في اختبار كلا المعاملين يتبع توزيع ت بدرجات حرية (ن - ٢) ، وبإجراء بعض العمليات الرياضية من خلال العلاقة بينهما ، نستطيع أن نثبت ان قيمة إحصاء الاختبار ، في الحالتين متساو .

٦/٥ تطبيقات على الانحدار والارتباط الخطي البسيط باستخدام الحاسب الآلي

يمكن استخدام برمجية Minitab في الحصول على نتائج تطبيق الانحدار والارتباط ، ويمكن تلخيص خطوات التطبيق في الخطوات التالية :

اولا : بالنسبة لتحليل الانحدار : بالتطبيق على بيانات مثال (٦-٤) ، يتبع الخطوات التالية

١- ادخال بيانات المتغير التابع ص في العمود C1 وتسميه Quantity وترمز الى الكمية المنتجة ، وبيانات المتغير المستقل س في العمود الثاني C2 وتسميه Exper وترمز الى سنوات الخبرة .

٢- اختيار قائمة " Stat " من شريط القوائم

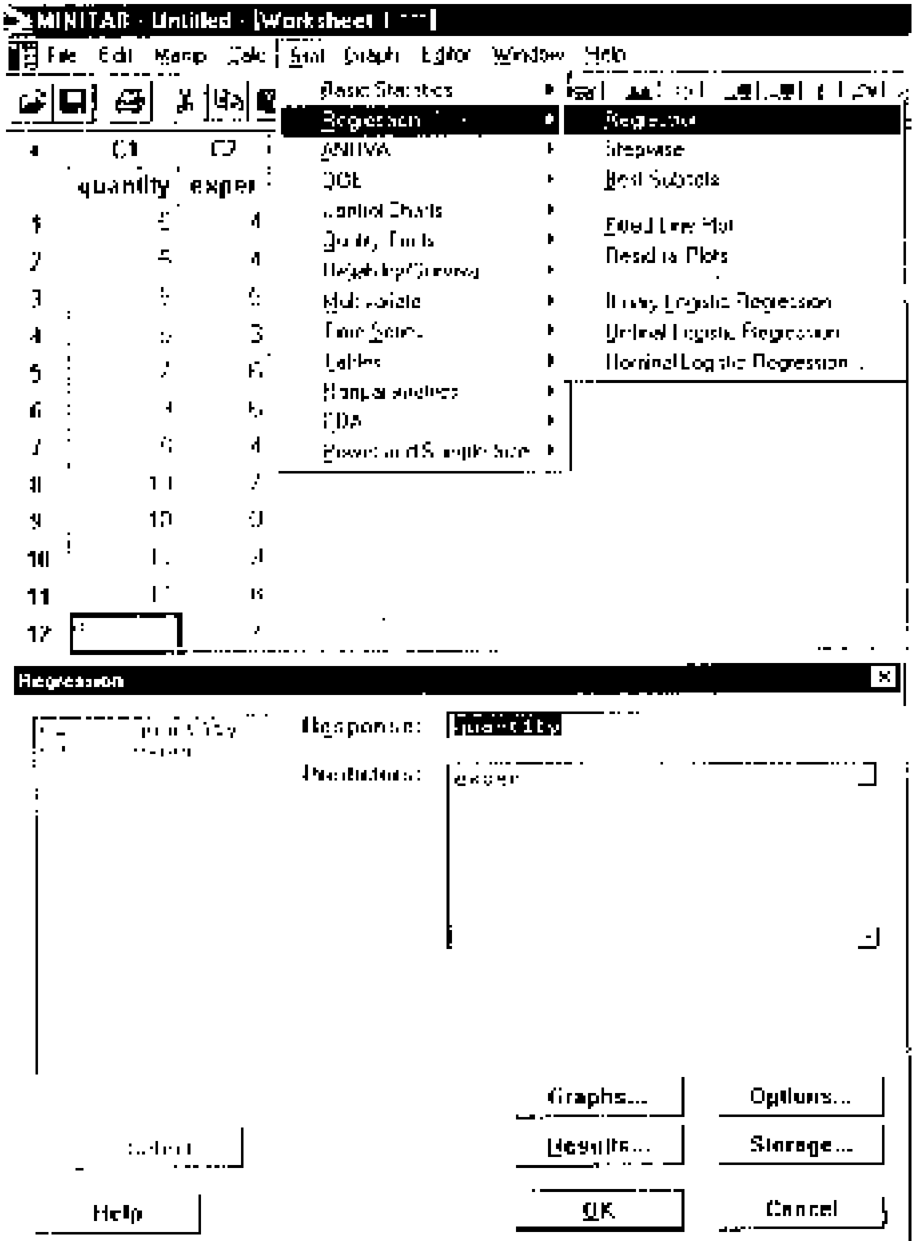
٣- اختيار قائمة " Regression " من قائمة " Stat " ، ثم النقر بالفارة على اول اختيار وهو " Regression ".

٤- إدخال المتغير التابع Quantity في المستطيل المسمى Response ، والمتغير المستقل Exper في المربع المسمى Predictors .

٥- النقر بالفارة على نافذة " OK " .

وهذه الخطوات عبر عنها بالشكل (٦-٢) التالي :

شكل (٦-٣)


MINITAB - Untitled - [Worksheet 1 ***]
File Edit Manip Calc Stat Graph Editor Window Help
Basic Statistics
Regression
ANOVA
DOE
Control Charts
Quality Tools
Multivariate
Time Series
Tables
EDA
Regression
Stepwise
Best Subsets
Fitted Line Plot
Residual Plots
Binary Logistic Regression
Ordinal Logistic Regression
Nominal Logistic Regression
C1
C2
quantity
exper
Regression
Response:
Predictors:
Graphs...
Options...
Results...
Storage...
Select
Help
OK
Cancel

وبتطبيق الخطوات الأربعة السابقة تظهر النتائج التالية :

```
Regression Analysis: quantity versus exper
The regression equation is
quantity = 2.13 + 0.978 exper

Predictor         Coef      SE Coef          T          P
Constant         2.130        1.176       1.81      0.100
exper           0.9783       0.1864       5.25      0.000

S = 1.264       R-Sq = 73.4%     R-Sq(adj) = 70.7%

Analysis of Variance

Source           DF          SS          MS          F          P
Regression        1      44.022      44.022      27.55      0.000
Residual Error   10      15.978       1.598
Total            11      60.000
```

وفيما يلي تعريف للنتائج السابقة :

١ - الجزء الأول : معادلة الانحدار المقدرة هي :

```
The regression equation is
quantity = 2.13 + 0.978 exper
```

المتغير التابع ص (quantity) ويعبر عن الكمية المنتجة .

المتغير المستقل س (exper) ويعبر عن عدد سنوات الخبرة .

(أ = ٢,١٣ ، ب = ٠,٩٧٨) وهي نفس النتائج السابق الحصول عليها .

شكل معادلة الانحدار : $\hat{\text{الكمية}}$ = ٢,١٣ + ٠,٩٧٨ الخبرة .

٢ - الجزء الثاني : نتائج اختبار معنوية معاملي الانحدار :

```
Predictor         Coef      SE Coef          T          P
Constant         2.130        1.176       1.81      0.100
exper           0.9783       0.1864       5.25      0.0004
```

ـ العمود الأول (Predictor) : يشمل أسماء المتغيرات المستقلة (المتنبأ منها) ،

وفي هذا التطبيق متغير واحد مستقل هو الخبرة (exper) .

ـ العمود الثاني (coef) : هي اختصار لكلمة المعاملات coefficients ، ويشمل هذا العمود تقديرات المعاملات حيث أن أ = ٢,١٢ ، ب = ٠,٩٧٨٣ .

ـ العمود الثالث (SE Coef) : الخطأ المعياري للمعاملات ، حيث أن :

الخطأ المعياري للثابت أ هو : $ع_أ$ = ١,٩٧٦ .

الخطأ المعياري للمعامل ب هو : $ع_ب$ = ٠,١٨٦٤ .

ـ العمود الرابع (T) : قيمة إحصاء الاختبار ت المستخدم في اختبار معنوية معاملات الانحدار ، وتحسب بقسمة العمود الثاني على العمود الثالث : أي أن : (SE Coef ÷ coef = T) ، وبالنسبة لإحصاء الاختبار المستخدم في اختبار معنوية المعامل β هو : ت = ٠,٩٧٨٣ ÷ ٠,١٨٦٤ = ٥,٢٥ ، وهي نفس النتيجة السابق الحصول عليها .

ـ العمود الخامس (p) : هو قيمة الاحتمال المشاهد ، وتحسب باستخدام التوزيع ت ، حيث أن :

p = ٢ ح(ت > ت$_م$) ، ت$_م$ = القيمة المطلقة لإحصاء الاختبار ت)

فبالنسبة لقيمة الاحتمال المشاهد الخاص بمعامل الانحدار ب ، نجد أن :

p = ٢ ح(ت > ٥,٢٥) = ٠,٠٠٠٤

وهذه القيمة الاحتمالية تبين أنه يمكن رفض الفرض العدمي (ف$_٠$: β = صفر) ، وقبول الفرض البديل (ف$_١$: β ≠ صفر) عند أي مستوى معنوية أكبر من ٠,٠٠٠٤ ، فإذا كان مستوى المعنوية المحدد (α = ٠,٠٥) ، في هذه الحالة نرفض الفرض العدمي ، ونقبل الفرض البديل ، لأن قيمة α أكبر من قيمة p .

٣ ـ الجزء الثالث : مؤشرات جودة النموذج :

S = 1.264   R-Sq = 73.4%   R-Sq(adj) = 70.7%

ـ الخطأ المعياري للتقدير (S) ، أي أن : $ع_{ص.س}$ = S = ١,٢٦٤

- معامل التحديد (R-Sq) ويرمز له ر² (R-Sq) = ٧٣,٤٪ وتدل على أن الخبرة تفسر ٧٣,٤٪ من الاختلافات في المتغير التابع، وأن النسبة المكملة ٢٦,٦٪ ترجع لأسباب أخرى.

- معامل التحديد المعدل بدرجات الحرية (R-Sq(adj)) : ويحسب بالمعادلة التالية :

$$ر^2_{معدل} = ١ - \frac{(ن - ١)}{(ن - ٢)} (١ - ر^2) = ١ - \frac{١١}{١٠} (٠,٢٦٦) = ٠,٧٠٧$$

وهذا المعامل يمكن استخدامه في المقارنة بين عدد من النماذج إذا كان عدد متغيراتها المستقلة غير متساو.

٤ - الجزء الرابع : جدول تحليل التباين (Analysis of Variance)

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	44.022	44.022	27.55	0.0002
Residual Error	10	15.978	1.598		
Total	11	60.000			

- العمود الأول Source (مصدر الاختلاف) : ويتكون من Regression (الانحدار)، Error (الخطأ)، Total (الكلي).

- العمود الثاني DF (درجات الحرية) : ويتكون من :

درجات حرية الانحدار = عدد المتغيرات المستقلة = ١

درجات حرية الأخطاء = (ن - عدد المتغيرات المستقلة - ١) = (ن - ١ - ١) = ١٠

درجات حرية الكلي = (ن - ١) = ١١

- العمود الثالث SS (مجموع المربعات الكلي)، حيث أن :

مجموع مربعات الانحدار SSR (م . م . ح) = ٤٤,٠٢٢

مجموع مربعات الأخطاء SSE (م . م . خ) = ١٥,٩٧٨

مجموع المربعات الكلي SST (م . م . ك) = (م . م . ح) + (م . م . خ)

= ٤٤,٠٢٢ + ١٥,٩٧٨ = ٦٠

ـ العمود الرابع MS (متوسط المربعات) : وهو ناتج من قسمة العمود الثالث على العمود الثاني ، أي أن : MS = SS ÷ DF

متوسط مربعات الانحدار MSR ($ع^2_ح$) = ٤٤,٠٢٢ ÷ ١ = ٤٤,٠٢٢

متوسط مربعات الأخطاء MSE ($ع^2_{ص س}$) = ١٥,٩٧٨ ÷ ١٠ = ١,٥٩٧٨

ـ العمود الخامس F (إحصاء الاختبار ف): وهو الإحصاء المستخدم في اختبار صلاحية النموذج . وهو ناتج من :

F (ف) = MSR ($ع^2_ح$) ÷ MSE ($ع^2_{ص س}$) = ٤٤,٠٢٢ ÷ ١,٥٩٨ = ٢٧,٥٥

ـ العمود الخامس p (الاحتمال المشاهد) : وهي قيمة احتمالية تحسب باستخدام التوزيع الاحتمالي ف ، وتستخدم المعادلة :

P = ح ($ف_{(١ ، ١٠)}$ > قيمة الإحصاء ف) = ح ($ف_{(١ ، ١٠)}$ > ٢٧,٥٥) = ٠,٠٠٠٤

، فإذا كان المطلوب اختبار صلاحية النموذج عند مستوى معنوية α = ٠,٠٥ ، يتخذ القرار كما يلي : بما أن (α = ٠,٠٥) أكبر من قيمة الاحتمال المشاهد (p = ٠,٠٠٠٤) ترفض الفرض العدمي ($ف_٠$: النموذج غير مناسب) ، ونقبل الفرض البديل ($ف_١$: النموذج مناسب) .

ثانيا : بالنسبة لتحليل الارتباط : بالتطبيق أيضا على بيانات مثال (٦-٤) ، يتبع الآتي :

١- إدخال بيانات المتغير ص في العمود C1 وتسميه Quantity وبيانات المتغير س في العمود الثاني C2 وتسميه Exper .

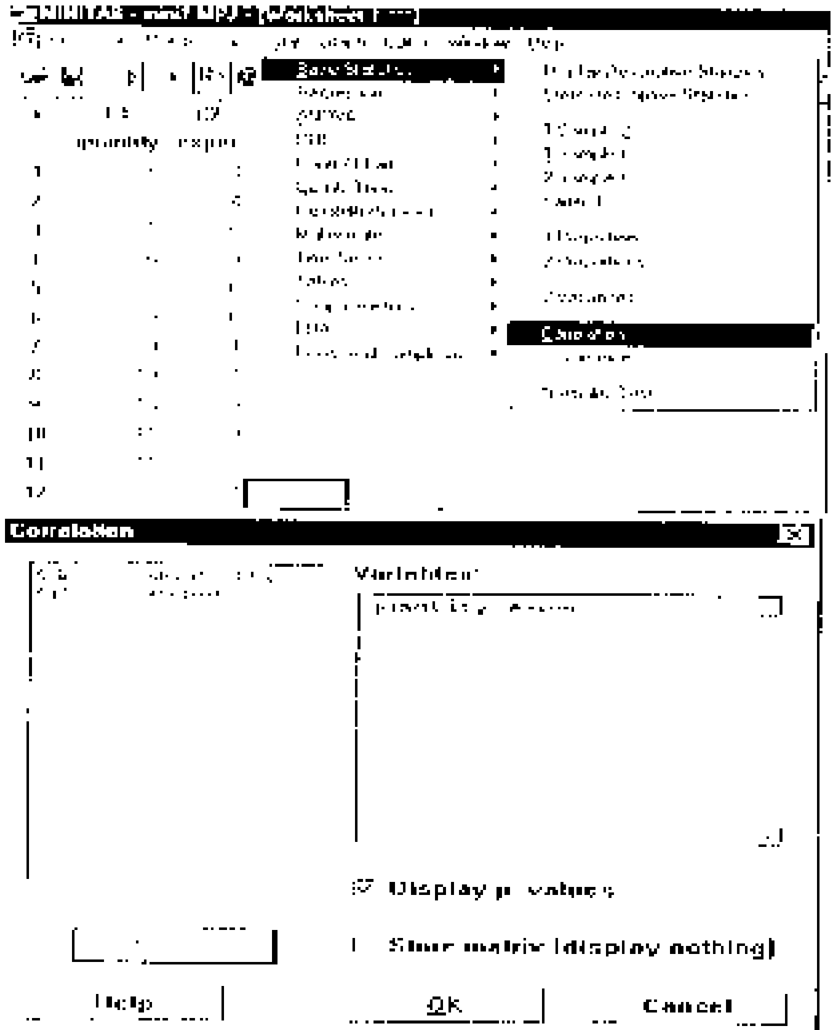
٢- اختيار قائمة " Stat " من شريط القوائم .

٣- اختيار قائمة "Basic Statistics" ، ثم اختيار كلمة Correlations

٤- إدخال المتغيرين (quantity and exper) في المستطيل المسمى باسم Correlations ، ثم النقر على كلمة Display p- value .

٥- النقر بالفارة على كلمة موافق "OK" ، كما في الشكل التالي :

شكل (٦-٤)



وبمجرد إتمام الخطوات السابقة ، والنقر على كلمة "OK" تظهر النتائج التالية :

```
Correlations: quantity; exper
Pearson correlation of quantity and exper = 0.857
P-Value = 0.0004
```

تعني كلمة : Pearson correlation of quantity and exper = 0.857

ان معامل ارتباط بيرسون (ر) بين الكمية quantity والخبرة exper = ٠.٨٥٧ ، وتدل هذه القيمة على وجود ارتباط طردي قوي بين الكمية ، والخبرة .

وتعبر كلمة (P-Value = 0.0004) عن قيمة الاحتمال المشاهد ، وهي قيمة محسوبة باستخدام التوزيع الاحتمالي ت ، وعن طريق هذه القيمة يمكن اتخاذ قرار بشأن اختبار فرض معنوية معامل الارتباط في المجتمع ρ .

فاذا كان مستوى المعنوية هو (α = ٠.٠٥) ، نجد ان قيمة (α = ٠.٠٥) أكبر من قيمة الاحتمال المشاهد (p = ٠.٠٠٠٤) لذا نرفض الفرض العدمي ف. : ρ = صفر ، ونقبل الفرض البديل ف١ : ρ ≠ صفر ، ونستدل من ذلك على ان العلاقة بين الكمية والخبرة علاقة ذات دلالة معنوية .

تمارين

١ – سحبت عينة عشوائية من أزواج القيم (س ، ص)

س	٤	٥	٣	٦	١٠
ص	٤	٦	٥	٧	٧

أ – قدر خط الانحدار ص : س ، ثم فسر المعادلة .

ب - احسب معامل التحديد ، وما هو مدلوله ؟

جـ - كون جدول تحليل التباين واختبر صلاحية النموذج ، (α = ٠٫٠٥) .

٢ – فيما يلي مشاهدات عينة عشوائية .

س	٥	٣	٦	٣	٤	٤	٦	٨
ص	١٣	١٥	٧	١٢	١٣	١١	٩	٥

أ – احسب معامل الارتباط ، وكذلك معامل التحديد ، وما مدلول كل منهما .

ب – احسب الخطأ المعياري للتقدير .

٣ – فيما يلي بيانات عن ضغط الدم (ص) ، السن (س) لعينة عشوائية من ١٥ فرد .

مجـ ص = ٢١٩٠   مجـ س = ٧٨٠

$ع_ص$ = ١٥٫٧١٦٢ $ع_س$ = ٧٫٠٩٠٨   ر = ٠٫٧٦٣

والمطلوب :

أ – قدر معادلة انحدار ضغط الدم على السن ، وما مدلولها ؟

ب – اختبر الفرض : β_1 > صفر ، (α = ٠٫٠١)

جـ - ما هي نسبة الجزء الغير مفسر للاختلافات الكلية ؟

٤ – النتائج التالية تم الحصول عليها باستخدام برمجية **Minitab**

Regression Analysis: cons. versus incom

```
The regression equation is
cons. = ---- + 0.923 incom
```

Predictor	Coef	SE Coef	T	P
Constant	17.26	52.81	0.33	0.748
incom.	0.92278	0.06754	----	0.000

```
S = ----        R-Sq = ----        R-Sq(adj) = 90.7%
```

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1644538		186.69	0.000
Residual Error	--	159573	8810		
Total	19	----			

Correlations: cons.; incom

```
Pearson correlation of cons. and incom = -----
P-Value = 0.000
```

والمطلوب :

أ - ما اسم المتغير التابع والمتغير المستقل .

ب - أكمل القيم المفقودة في كل جزء من الأجزاء السابقة .

جـ - فسر معادلة الانحدار .

د - هل تؤيد أن النموذج مناسب ؟ ، α = ٠,٠١

هـ - اختبر الفرض β_1 = صفر ، α = ٠,٠١

و - ما هو نسبة الجزء المفسر للاختلافات في المتغير التابع ؟

الباب السابع

تحليل الانحدار الخطى المتعدد

٧/١ مقدمــة

٧/٢ نموذج الانحدار الخطى المتعدد

٧/٣ مؤشرات جودة النموذج

٧/٤ تحليل التباين واختبار صلاحية النموذج

٧/٥ اختبار معنوية معاملات الانحدار

تحليل الانحدار الخطى المتعدد

٧/١ مقدمـــة

اهتم تحليل الانحدار الخطى البسيط بدراسة وتحليل أثر متغير مستقل واحد على المتغير التابع ، ومن المعلوم أن الاختلاف الكلى فى المتغير التابع ينقسم الى جزأين ، الجزء المفسر ومرجعه الى المتغير المستقل س ، والجزء غير المفسر ومرجعه للأخطاء العشوائية ، وكلما كان الجزء المفسر كبيرا ، كلما دل ذلك على جودة النموذج ، ومناسبته فى تمثيل العلاقة بين المتغير التابع والمتغير المستقل ، ومن أجل ذلك يمكن التحسين فى قدرة النموذج على التنبؤ بإضافة متغيرات أخرى مستقلة لها علاقة بالمتغير التابع ، فكلما أضيف متغير مستقل إلى النموذج كلما زاد نسبة الجزء المفسر ، وفى المقابل انخفض الجزء غير المفسر بمقدار هذه الزيادة . من ثم تهتم الدراسة فى هذا الباب بتحليل الانحدار الخطى المتعدد ، والغرض من هذا التحليل دراسة وتحليل أثر عدة متغيرات مستقلة على متغير تابع .

ففى المثال (٦-٤) فى الباب السادس نجد أن هناك متغيرات أخرى خلاف سنوات الخبرة تؤثر على الكمية التى ينتجها العامل من الخيوط المغزولة ، مثل عدد الدورات التدريبية التى تلقاها العامل ، وكثافة المسكن الذى يعيش فيه العامل ، وإضافة هذين المتغيرين إلى متغير سنوات الخبرة ، يؤدى الى زيادة نسبة الجزء المفسر ، وفى المقابل انخفاض الجزء غير المفسر ، ويترتب على تلك تحسين فى قدرة النموذج التنبؤية .

فى هذا المثال وجد أن معادلة خط الانحدار تأخذ الصورة التالية :

$$\hat{\text{الكمية}} = ٢٫١٣٠٢ + ٠٫٩٧٦٣ \ \text{الخبرة}$$

وإذا أضفنا كثافة المسكن الذي يعيش فيه العامل كمتغير مفسر ، وكانت بيانات هذا المتغير بالاضافة إلى البيانات السابقة كالتالي .

العامل	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠	١١	١٢
الكمية (ص)	٥	٥	٥	٦	٧	٩	٨	١٠	١٠	١١	١١	٩
الخبرة (س₁)	٤	٤	٥	٣	٦	٦	٤	٧	٩	٩	٨	٧
كثافة المسكن (س٢)	٢	٢	٢,٥	١	٣	٢,٥	٢	٣	١	١	١,٥	١,٥

وبفرض أن معادلة الانحدار المتعدد التي تصف العلاقة بين الكمية كمتغير تابع ، والخبرة ، كثافة المسكن) كمتغيرين مفسرين ثم تقديرها ، وتأخذ الصورة التالية .

$$\hat{\text{الكمية}} = ٤,٨٤ + ٠,٨٣٥ \text{ الخبرة } \quad ٠,٨٢٣ \text{ كثافة المسكن}$$

وإذا استخدمنا كلا المعادلتين ، الانحدار الخطي البسيط ، والانحدار الخطي المتعدد في حساب خطأ التقدير (د = الكمية − الكمية^) نستطيع الحصول على النتائج التالية .

العامل رقم	كمية الانتاج (ص)	عدد سنوات الخبرة (س)	كثافة المسكن	خطأ التقدير د: انحدار بسيط	خطأ التقدير د: انحدار متعدد
١	٥	٤	٢,٠	١,٠٤٣٤٨-	٠,٧١٢٦٨-
٢	٥	٤	٢,٠	١,٠٤٣٤٨-	٠,٧١٢٦٨-
٣	٥	٥	٢,٥	٢,٠٣١٧٤-	١,٩٥٩١٥-
٤	٦	٣	١,٠	٠,٩٣٤٧٨	٠,٣٠٠٠٠
٥	٧	٦	٣,٠	١,٠٠٠٠٠-	٠,٣٨٣١٠-
٦	٩	٦	٢,٥	١,٠٠٠٠٠	١,٢٠٩٦٣
٧	٨	٤	٢,٠	١,٩٥٦٥٢	١,٤٦٤٧٩
٨	١٠	٧	٣,٠	١,٠٢١٧٤	١,٧٨١٤٩
٩	١٠	٩	١,٠	٠,٨٣١٧٨-	٠,٧١١٢٧-
١٠	١١	٩	١,٠	٠,٠٦٥٢٢	٠,٥٤٣٨٠-
١١	١١	٨	١,٥	١,٠٤٣٤٨	٠,٧١٢٦٨
١٢	٩	٧	١,٥	٠,٠٢١٧٤	٠,٤٩٣١١-

يلاحظ أن خطأ التقدير د والذي يمثل الانحرافات عن خط الانحدار ، في حالة الانحدار الخطي البسيط أكبر من خطأ التقدير في حالة الانحدار الخطي المتعدد ، ولو قمنا بإيجاد مجموع مربعات الأخطاء العشوائية في الحالتين ، نجد أن :

في حالة الانحدار الخطي البسيط : SSE (م . م . خ) = ١٥,٩٧٨

في حالة الانحدار الخطي المتعدد : SSE (م . م . خ) = ١٣,٣٦٨

أي أن مجموع مربعات الأخطاء قد انخفض من ١٥,٩٧٨ إلى ١٣,٣٦٨ ، ويترتب على تلك الانخفاض نسبة الجزء غير المفسر بمقدار :

[(١٥,٩٧٨ – ١٣,٣٦٨) ÷ SST (م . م . ك)] = ٢,٦١٠ ÷ ٦٠ = ٠,٠٤٣٥

وفي المقابل زادت نسبة الجزء المفسر بهذا المقدار ، فبعد أن كان معامل التحديد مقداره ٠,٧٣١ في حالة الانحدار البسيط ، أصبح قيمته (٠,٧٣١ + ٠,٠٤٣٥) = ٠,٧٧٧ تقريبا ويدل هذا المعامل : أن الخبرة وكثافة المسكن يفسران حوالي ٧٧,٧ ٪ من الاختلافات في الكمية المنتجة ، وأن النسبة المكملة (٢٢,٣ ٪) ترجع لأخطاء عشوائية .

وهكذا كلما أضيف متغير مفسر إلى المتغيرات المفسرة في النموذج كلما زاد نسبة الجزء المفسر للاختلافات الكلية على حساب الجزء غير المفسر .

٢/٧ نموذج الانحدار الخطي المتعدد

يفرض أن ص هو المتغير التابع ، وأن لدينا م من المتغيرات المفسرة ، هي : $س_1$ ، $س_2$ ، ... ، $س_م$ فإن نموذج الانحدار الخطي المتعدد في المجتمع يأخذ الصورة التالية :

$$ص = \alpha + \beta_1 س_1 + \beta_2 س_2 + \dots + \beta_م س_م + د \qquad (٧-١)$$

وبسحب عينة عشوائية حجمها ن من مجموعات القيم (ص ، $س_1$ ، $س_2$ ، $س_م$)

يمكن الحصول على تقدير المربعات الصغرى لمعاملات انحدار المجتمع ، وتنتج بإيجاد الحل الجبري لمجموعة المعادلات الخطية التالية :

مجـ ص = أ ن + ب$_1$ مجـ س$_1$ + ب$_2$ مجـ س$_2$ + ... + ب$_م$ مجـ س$_م$

مجـ س$_1$ ص = أ مجـ س$_1$ + ب$_1$ مجـ س$_1^2$ + ب$_2$ مجـ س$_1$ س$_2$ + ... + ب$_م$ مجـ س$_1$ س$_م$

مجـ س$_2$ ص = أ مجـ س$_2$ + ب$_1$ مجـ س$_1$ س$_2$ + ب$_2$ مجـ س$_2^2$ + ... + ب$_م$ مجـ س$_2$ س$_م$

....

....

....

مجـ س$_م$ ص = أ مجـ س$_م$ + ب$_1$ مجـ س$_1$ س$_م$ + ب$_2$ مجـ س$_2$ س$_م$ + ... + ب$_م$ مجـ س$_م^2$

حيث أن أ هو التقدير بنقطة للثابت α ، ب$_ر$ هو التقدير بنقطة لمعامل الانحدار β$_ر$ ، (ر = ١ ، ٢ ، ... ، م) .

و يستخدم البرنامج الإحصائي **Minitab** في حل مجموعة المعادلات أعلاه وخاصة إذا كان عددها كثيرا ، وللحصول على كل النتائج الممكنة بتحليل الانحدار السابق عرضها في الباب السادس .

وبالحصول على التقديرات (أ ، ب$_1$ ، ب$_2$ ، ... ، ب$_م$) ، يمكن كتابة معادلة الانحدار المقدرة على الصورة التالية :

صَ = أ + ب$_1$ س$_1$ + ب$_2$ س$_2$ + ... + ب$_م$ س$_م$ (٧-٣)

شروط تحليل الانحدار المتعدد

للحصول على مقدرات جيدة يتوافر فيها خصائص المقدر الخطي غير المتحيز ذو الأقل تباين ، يجب توافر الشروط التالية .

١ - العلاقة بين المتغير التابع ومجموعة المتغيرات المستقلة علاقة خطية .

٢ - المتغير التابع متغير مستمر ، أو على الأقل يقاس بمعيار بعدي .

٣ - الأخطاء العشوائية هـ = (ص – صَ) تتبع توزيع طبيعي متوسطة

صفراً ، وتباينه σ^2 ص.س١ ... س٣ وهو ثابت من مشاهدة إلى أخرى .

٤ – أن المشاهدات التابعة مستقلة إحصائياً.

تطبيق على الانحدار المتعدد باستخدام الحاسب الآلي

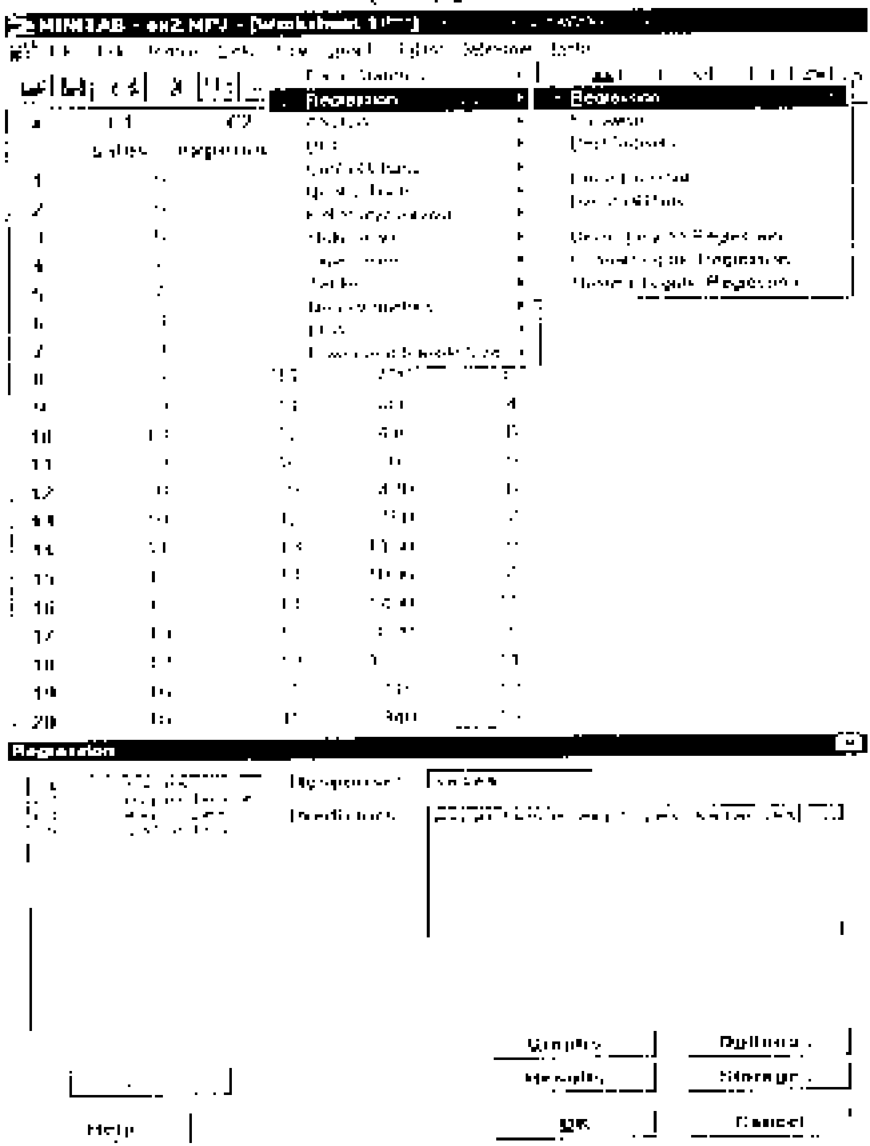
شركة لها عدة فروع ، أراد قطاع البحوث في هذه الشركة أن يدرس العلاقة بين حجم مبيعات الفرع بالمليون جنيه كمتغير تابع ، ومجموعة من المتغيرات المستقلة هي : عدد سنوات خبرة مدير الفرع ، عدد العاملين بالفرع ، وجملة المرتبات المدفوعة بالمليون . قام القطاع بسحب عينة عشوائية حجمها ٢٠ فرع ، وحصل على البيانات التالية خلال فترة زمنية معينة .

الفرع	حجم المبيعات (ص)	سنوات الخبرة (س١)	عدد العاملين (س٢)	جملة المرتبات المدفوعة (س٣)
١	٦	٣	٢٤٠	٢
٢	٩	٥	١٠٠	٥
٣	٦	٥	٩٠٠	٤
٤	٧	٩	٩٥٠	٥
٥	٧	٨	٣٠٠	٥
٦	٨	٩	٢٨٠	٥
٧	٨	١٠	٤٠٠	٦
٨	٩	١١	٤٦٠	٦
٩	٩	١١	٤٨٠	٤
١٠	١٠	١٢	٤٢٠	٦
١١	٩	١٠	٩٠٠	٨
١٢	٨	٩	٤٨٠	٦
١٣	١٠	١٢	٧٥٠	٧
١٤	١١	١٣	١٦٠٠	٩
١٥	١٣	١٤	١٠٠٠	٨
١٦	١٥	١٤	١٤٠٠	١١
١٧	١٠	١٢	٨٠٠	٩
١٨	١٣	١٣	١٢٨٠	١١
١٩	١٦	١٥	٧٠٥	١٣
٢٠	١٨	١٥	٩١٠	١٣

والمطلوب إيجاد معادلة للتنبؤ ، ونتائج تحليل الانحدار باستخدام برمجية Minitab .

الحل : يتم إتباع نفس الخطوات السابق عرضها في الباب السادس . والشكل (٧-١) يبين هذه الخطوات :

شكل (٧-١)



وبمجرد إتمام الخطوات الموضحة في الشكل (٧-١) والنقر على كلمة موافق " OK " تظهر النتائج المبينة بالشكل التالي :

شكل (٧-٢)

Regression Analysis: sales versus experience; employs; salaries   جزء (١)

The regression equation is   جزء (٢)
sales = 0.957 + 0.399 experience - 0.00050 employs + 0.719 salaries

جزء (٣)

Predictor	Coef	SE Coef	T	P
Constant	0.9573	0.9165	0.99	0.337
experience	0.3987	0.1461	2.63	0.016
employs	-0.000505	0.001295	-0.39	0.702
salaries	0.7191	0.1828	3.93	0.001

S = 1.194   R-Sq = 89.6%   R-Sq(adj) = 87.7%   جزء (٤)

Analysis of Variance   جزء (٥)

Source	DF	SS	MS	F	P
Regression	3	196.973	65.658	46.02	0.000
Residual Error	16	22.827	1.427		
Total	19	219.800			

ونلاحظ الآتي :

جزء (١) : Regression Analysis: sales versus experience; employs; salaries

عنوان التحليل Regression Analysis " تحليل الانحدار " مبينا فيه المتغير التابع sales " حجم المبيعات " ، ومجموعة المتغيرات المفسرة وهي : experience " الخبرة " ، employs " عدد الموظفين " ، salaries "جملة المرتبات المدفوعة "

جزء (٢) : The regression equation is

sales = 0.957 + 0.399 experience - 0.00050 employs + 0.719 salaries

تمثل معادلة الانحدار (أو معادلة التنبؤ)

ونكتب المعادلة كالتالي :

المبيعات^ = ٠.٩٠٧ + ٠.٣٩٩ (الخبرة) − ٠.٠٠٠٥ (عدد العاملين) + ٠.٧١٩ (جملة المرتبات)

وتفسير المعادلة كما يلي :

ــ الثابت (أ = ٠.٩٠٧) : يدل على أن حجم المبيعات قيمته ٠.٩٠٧ مليون جنيه عندما تكون قيمة كل من عدد سنوات الخبرة ، عدد العاملين بالفرع ، جملة المرتبات المدفوعة مساوية للصفر ، وبمعنى آخر حجم المبيعات الثابتة = ٠.٩٠٧ .

ــ معامل الانحدار (ب الخبرة = ٠.٣٩٩) : هذا المعامل موجب ، يدل على أنه كلما زاد عدد سنوات خبرة المدير سنة واحدة أدى ذلك الى زيادة حجم المبيعات بمقدار ٠.٣٩٩ مليون جنيه بفرض ثبات المتغيرين الآخرين (عدد العاملين ، جملة المرتبات) .

ــ معامل الانحدار (ب عدد العاملين = −٠.٠٠٠٥) : يدل هذا المعامل على أن حجم المبيعات ينخفض بمقدار ٠.٠٠٠٥ مليون جنيه اذا زاد عدد العمال بمقدار عامل في الفرع بافتراض ثبات المتغيرين الآخرين (عدد سنوات الخبرة ، جملة المرتبات) .

ــ معامل الانحدار (ب المرتبات = ٠.٧١٩) : يدل على ارتباط حجم المبيعات بجملة المرتبات بعلاقة طردية ، وتشير إلى أن حجم المبيعات يزيد بمقدار ٠.٧١٩ مليون جنيه ، إذا زادت جملة المرتبات المدفوعة بمقدار مليون جنيه ، بفرض ثبات المتغيرين الآخرين (عدد سنوات الخبرة ، وعدد العاملين) .

وبالنسبة للفرع الأول على سبيل المثال ، نجد أن المبيعات المتنبأ بها هي :

المبيعات^ = ٠.٩٠٧ + ٠.٣٩٩ (٣) − ٠.٠٠٠٥ (٢٤٠) + ٠.٧١٩ (٣) = ٤.١٤١ مليون جنيه

وأما حجم المبيعات الفعلية = ٦ مليون جنيه ، ومن ثم تكون قيمة خطأ التقدير د ، وتعبر عن انحراف القيمة الفعلية عن خط الانحدار هي :

د = حجم المبيعات الفعلية − المبيعات^ = ٦ − ٤.١٤١ = ١.٨٥٩ مليون جنيه

وهكذا يمكن حساب أخطاء التقدير ، وملاحظة شكل انحرافات القيم الفعلية عن خط

الانحدار ، كمؤشر للحكم على جودة النموذج .

٣/٧ مؤشرات جودة النموذج

يمكن الحصول على بعض المقاييس كمؤشرات تدل على مدى كفاءة النموذج ، ومنها الخطأ المعياري S ($ع_{ص.١٢...م}$) ، والذي يقيس انحرافات القيم الفعلية عن القيم المتنبأ بها (ص – $\hat{ص}$) . والمقياس الثاني هو معامل التحديد R-sq ($ر^{٢}$) ويبين نسبة الجزء المفسر الذي مرجعه إلى المتغيرات المفسرة .

الخطأ المعياري المتعدد S ($ع_{ص.١٢...م}$)

يرمز للخطأ المعياري المتعدد بالرمز $ع_{ص.١٢...م}$ ، حيث م هي عدد المتغيرات المفسرة في النموذج ، ويحسب بالمعادلة التالية .

$$ع_{ص.١٢...م} = \sqrt{\frac{\text{مجـ}(ص - \hat{ص})^{٢}}{(ن - م - ١)}} \qquad (٣-٧)$$

ويعبر البسط مجـ $(ص - \hat{ص})^{٢}$ عن مجموع مربعات الأخطاء العشوائية SSE (م . م . خ) ، ويعبر المقام (ن – م – ١) عن درجات حرية الأخطاء ، فمثلا في حالة الانحدار البسيط كان عدد المتغيرات المستقلة م = ١ ، وجدنا أن الخطأ المعياري $ع_{ص.١}$ عبارة عن :

$$ع_{ص.١} = \sqrt{\frac{\text{مجـ}(ص - \hat{ص})^{٢}}{(ن - ١ - ١)}} = \sqrt{\frac{\text{مجـ}(ص - \hat{ص})^{٢}}{(ن - ٢)}}$$

وفي التطبيق السابق ، كان عدد المتغيرات المستقلة م = ٣ ، ومن ثم يحسب الخطأ المعياري S بتطبيق المعادلة (٧ ـ ٣) كما يلي :

$$ع_{ص...} = \sqrt{\frac{\text{مجـ}(ص - \hat{ص})^2}{(ن - ٣ - ١)}} = \sqrt{\frac{\text{مجـ}(ص - \hat{ص})^2}{(٢٠ - ٤)}}$$

$$SSE\ (م . م . خ) = \text{مجـ}(ص - \hat{ص})^2 = ٢٢,٨٢٧$$

ونستخرج هذه القيمة من جزء (٥) Analysis of Variance (جدول تحليل التباين) في شكل (٧-٢) . ومن ثم نجد أن للخطأ المعياري قيمته هي:

$$ع_{ص...} = \sqrt{\frac{٢٢,٨٢٧}{(١٦)}} = ١,١٩٤ \ ...\ S$$

وهذه القيمة تظهر في جزء (t) في الشكل (٧-٢) . وتعكس هذه القيمة انحرافات قيم المبيعات الفعلية عن القيم المتنبأ بها وهذه القيمة صغيرة ، ويستدل منها على جودة النموذج .

معامل التحديد R-sq (ر ٢)

هو المقياس الثاني الذي يستدل منه على جودة النموذج ، ويحسب معامل التحديد بنفس المعادلة التي سبق استخدامها في الانحدار الخطي البسيط ، وهي :

$$ر^2 = \frac{\text{مجموع مربعات الانحدار } SSR\ (م . م . ح)}{\text{مجموع المربعات الكلي } SST\ (م . م . ك)} \qquad (٧-٤)$$

وفي التطبيق السابق ، يلاحظ أن :

مجموع المربعات الكلي SST (م.م.ك) = مجـ ص^{2} – ن $\bar{ص}^{2}$ = ٢١٩٫٨

مجموع مربعات الانحدار SSR (م.م.ح) = مجـ ($\hat{ص}$ – $\bar{ص}$)^{2} = ١٩٦٫٩٧٣

وتظهر هاتين القيمتين في جزء (٥) في الشكل (٧ – ٢) ، ومن ثم يكون معامل التحديد هو :

$$ر^{2} = \frac{١٩٦٫٩٧٣}{٢١٩٫٨} = ٠٫٨٩٦$$

وتظهر هذه القيمة في جزء (٤) في الشكل (٧-٢) ، وتدل على أن عدد سنوات خبرة المدير ، وعدد العاملين ، وجملة المرتبات المدفوعة تفسر ٨٩٫٦٪ من الاختلافات الكلية في حجم المبيعات ، وأن النسبة المكملة ١٠٫٤٪ ترجع لأسباب أخرى .

٧/٤ تحليل التباين واختبار صلاحية النموذج

في الشكل (٧-٢) يظهر جدول تحليل التباين في جزء (٦) ، وبشكل عام إذا علم عدد المتغيرات المفسرة (م) ، وحسب مجموع المربعات SS (م.م) يمكن صياغة جدول تحليل التباين كما يلي :

جدول تحليل التباين

Source المصدر	DF درجات الحرية	SS مجموع المربعات	MS متوسط المربعات	F ف
الانحدار	م	SSR (م.م.ح)	MSR (م.م.ح) = (م.م.ح ÷ م)	$\frac{MSR}{MSE}$
الخطأ	ن – م – ١	SSE (م.م.خ)	MSE (م.م.خ) = (م.م.خ ÷ (ن – م – ١))	
الكلي	ن – ١	SST (م.م.ك)		

وفي التطبيق السابق يمكن إعادة صياغة الجدول باللغة العربية كالتالي :

جدول تحليل التباين

Source المصدر	DF درجات الحرية	SS مجموع المربعات	MS متوسط المربعات	F ف
الانحدار	٣	SSR (م . ج . ع) ١٩٦,٩٧٣	MSR ($ع^2_ع$) = ٦٥,٦٥٨	٦٥,٦٥٨
الخطأ	١٦	SSE (م . م . ع) ٢٢,٨٢٧	MSE ($ع^2_{م ر}$ ...) = ١,٤٢٧	١,٤٢٧
الكلي	١٩	SST (م . م . ك) ٢١٩,٨		= ٤٦,٠٢

ومن الجدول أعلاه يمكن حساب المقاييس المستخدمة كمؤشرات لجودة النموذج ، كما سبق بيان ذلك ، ومن الاغراض الرئيسية أيضا لتحليل التباين استخدام إحصاء الاختبار F (ف) في اختبار صلاحية النموذج ، حيث يأخذ الفرض العدمي والبديل الصورة التالية :

ف. : النموذج غير مناسب   ف١ : النموذج مناسب

ويمكن صياغة الفروض بشكل آخر ، كما يلي :

ف. : $\beta_1 = \beta_2 = ... = \beta_م$ = صفر   ف١ : ليس كل المعاملات = صفر

ويشير الفرض العدمي (ف.) الى ان معاملات الانحدار كلها معا تساوي صفر ، ويشير الفرض البديل البديل (ف١) إلى ان أحد المعاملات على الأقل يختلف عن الصفر . والاحصاء المستخدم في الاختبار هو F (ف) ، ويعبر عنه بالمعادلة :

$$ف = MSR(ع^2_ع) \div MSE(ع^2_{م ر}...) \qquad (٥-٧)$$

ويتبع هذا الإحصاء توزيع ف بدرجات حرية بسط : م ومقام (ن - م - ١) ، ويتخذ قرار بشأن الفرض العدمي والبديل من خلال مقارنة قيمة هذا الإحصاء بالقيمة

الجدولية ، أو من خلال مقارنة مستوى المعنوية المحدد α بقيمة الاحتمال المشاهد p الذي يظهر في جدول تحليل التباين الناتج من استخدام برمجية Minitab .

مثال (٧-١)

في التطبيق السابق ، هل تؤيد رأي قطاع البحوث في مناسبة النموذج في تمثيل العلاقة بين حجم المبيعات كمتغير تابع ، و عدد سنوات خبرة مدير الفرع ، وعدد العاملين بالفرع ، وجملة المرتبات المدفوعة كمتغيرات مستقلة ، (α=٠٫٠٥) .

الحل : لاختبار هذا الفرض ، نستخدم النتائج في الشكل (٧ - ٢) جزء (٥) ، ويتبع الخطوات التالية :

ـ ف. : $\beta_1 = \beta_2 = \beta_3 =$ صفر   ف١ : ليس كل المعاملات = صفر

ـ التوزيع الاحتمالي المستخدم هو توزيع ف بدرجات حرية بسط (م = ٣) ، ودرجات حرية مقام (ن − م − ١ = ١٦) .

ـ احصاء الاختبار :

ف = MSR(ع١) ÷ MSE(ع٢)

= ٦٥٫٦٥٨ ÷ ١٫٤٢٧ = ٤٦٫٠٢

ـ الاحتمال المشاهد :

P = ح(ف٣،١٦ > ٤٦٫٠٢) = صفر

ـ القرار : بما أن قيمة مستوى المعنوية (α=٠٫٠٥) أكبر من قيمة الاحتمال المشاهد (p = صفر) لذا يرفض الفرض العدمي ، ويقبل الفرض البديل ، ويستدل من ذلك على أن أحد المتغيرات الثلاثة على الأقل له أثر معنوي على المتغير التابع ، أي أن النموذج المقدر نموذج مناسب في التنبؤ .

٧/٥ اختبار معنوية معاملات الانحدار

الغرض من هذا الاختبار، معرفة ما إذا كان للمتغير المفسر $س_ر$ ، (ر = ١ ، ٢ ، ... ، م) أثر جوهري على المتغير التابع ص أم لا . فبالنسبة لاختبار معنوية معامل الانحدار في المجتمع $\beta_{س_ر}$ ، يمكن صياغة الفرض العدمي ، والفرض البديل على الصورة التالية :

$$ف_٠ : \beta_{س_ر} = صفر \qquad ف_١ : \beta_{س_ر} \neq صفر$$

ويشير الفرض العدمي ($ف_٠$) إلى أن أثر المتغير المستقل $س_ر$ على المتغير التابع ص غير معنوي ، وبمعنى آخر أن إدخال المتغير المستقل $س_ر$ في النموذج الذي يشمل باقي المتغيرات المفسرة ($س_١$ ، $س_٢$ ، ... ، $س_{ر-١}$ ، $س_{ر+١}$ ، ... ، $س_م$) لا يحسن في قدرة النموذج التنبوية ، بينما ينص الفرض البديل على أن إضافة المتغير المستقل $س_ر$ يحسن في القدرة التنبوية للنموذج .

والإحصاء المستخدم في الاختبار هو T (ت) ، ويحسب بالمعادلة التالية :

$$T(ت) = ب_{س_ر} \div ع_{ب_ر}$$

ويتبع الإحصاء T (ت) أعلاه توزيع ت بدرجات حرية (ن - م - ١) . وبمقارنة قيمة مستوى المعنوية المحدد ، بقيمة الاحتمال المشاهد يمكن اتخاذ قرار بشأن الفرض العدمي ، والفرض البديل .

مثال (٧-٢)

في التطبيق السابق ، هل تؤيد رأي قطاع البحوث في أن عدد سنوات خبرة

مدير الفرع لها أثر معنوي على حجم المبيعات ؟ ، α = ٠,٠٥ .

الحل: في شكل (٧-٢) جزء (٤) ، يظهر النتائج التالية :

تقديرات المعاملات Coef ، وهي :

أ = -٠,٩٠٧٣ ، ب الخبرة = ٠,٣٩٨٧ ، ب عدد العاملين = -٠,٠٠٠٠٥ ، ب المرتبات = ٠,٧١٩٩

الأخطاء المعيارية للتقديرات SE Coef وهي :

ع أ = ٠,٩٦١٥ ، ع ب(الخبرة) = ٠,١٤٨١ ، ع ب(العاملين) = ٠,٠٠٠٠١٣ ، ع ب(المرتبات) = ٠,١٨٢٨

قيم الإحصاء T(ت) الناتجة بتطبيق المعادلة (٧ - ٦) وهي :

ت أ = -٠,٩٠٧٣ ÷ ٠,٩٦١٥ = -٠,٩٤ ، ت ب(الخبرة) = ٠,٣٩٨٧ ÷ ٠,١٤٨١ = ٢,٦٩

ت ب(العاملين) = -٠,٠٠٠٠٥ ÷ ٠,٠٠٠٠١٣ = -٠,٣٩ ، ت ب(المرتبات) = ٠,٧١٩٩ ÷ ٠,١٨٢٨ = ٣,٩٣

قيم الاحتمال المشاهد .

ولاختبار الفرض السابق يتبع الآتي :

ـ ف٠ : β خبرة = صفر ف١ : β خبرة ≠ صفر

ـ التوزيع المستخدم هو توزيع ت بدرجات حرية (ن - م - ١ = ١٦)

ـ احصاء الاختبار :

ت ب(الخبرة) = ٠,٣٩٨٧ ÷ ٠,١٤٨١ = ٢,٦٩

ـ الاحتمال المشاهد :

P = ٢ح(ت(١٦) > ٢,٦٩) = ٠,٠١٦

ـ القرار : بما أن مستوى المعنوية (α = ٠,٠٥) أكبر من قيمة الاحتمال المشاهد (p = ٠,٠١٦) لذا يرفض الفرض العدمي ويقبل الفرض البديل . ويدل ذلك على ان خبرة المدير لها أثر معنوي على حجم المبيعات ، وبمعنى اخر أن إدخال خبرة المدير كمتغير مفسر إلى النموذج الذي يشمل المتغيرين الاخرين (عدد العاملين ، وجملة المرتبات) يحسن في قدرة النموذج التنبؤية .

ملاحظة : يمكن اختبار معنوية الاشارة لمعامل الانحدار ، من خلال استخدام القيم الجدولية ، أو باستخدام قيم الاحتمال المشاهد كما يلي :

بالنسبة للقيمة الجدولية : يتم الكشف عند (ت) ، وتقارن قيمة الإحصاء T (ت) بالقيمة الجدولية ، ويتخذ القرار .

بالنسبة لقيمة الاحتمال المشاهد p ، تحسب كما يلي :

P = ح (ت > القيمة المطلقة لإحصاء الاختبار)

= (p المأخوذة من نتائج برنامج Minitab ÷ ٢)

مثال (٧-٣)

في التطبيق السابق اختبر الفرض : β > صفر ، α = ٠,٠١

الحل : لاختبار هذا الفرض يتبع الآتي :

ف : β = صفر ف : β > صفر

التوزيع المستخدم توزيع ت بدرجات حرية ١٦

إحصاء الاختبار T (ت) = ٢,٩٣٠

الاحتمال المشاهد = (P ÷ ٢) = ٠,٠٠٠١ ÷ ٢ = ٠,٠٠٠٥

القرار : بما أن مستوى المعنوية (α = ٠,٠١) أكبر من قيمة الاحتمال المشاهد (٠,٠٠٠٥) لذا يرفض الفرض العدمي ويقبل الفرض البديل . ويدل ذلك على أن جملة المرتبات لها أثر معنوي طردي على حجم المبيعات .

مثال (٧-٤)

انشيء فترة ثقة ٩٥٪ لمعامل الانحدار β

الحل : بنفس الطرق المتبعة في إنشاء فترة ثقة يمكن إنشاء فترة ثقة للمعامل β كما يلي :

فترة الثقة هي : ب ± ت × ع

$\alpha = ٠,٠٥$ ، ن - م - ١ = ١٦ إذا ت$_{(٠,٠٢٥ ، ١٦)}$ = ٢,١٢

فترة الثقة هي :

$\hat{\beta} \pm$ ت$_{(٠,٠٢٥ ، ١٦)}$ ع$_{\hat{\beta}}$ = (٠,٣٩٨٧ ± ٢,١٢ (٠,١٤٨١))

(٠,٣٩٨٧ ± ٠,٣١٣٩٧٢) = (٠,٠٨٤٧ ، ٠,٧١٣٧)

أي أن المعامل β يتراوح بين (٠,٠٨٥ ، ٠,٧١٣) باحتمال ٠,٩٥ .

ويمكن اختبار معنوية معامل الانحدار عن طريق فترة الثقة ، فمن الملاحظ أن قيمة المعامل تحت صحة الفرض العدمي ف. : β = صفر تقع خارج فترة الثقة لذا يرفض الفرض العدمي ف. ويقبل الفرض البديل ف١ : $\beta \neq$ صفر .

مثال (٧-٥)

النتائج التالية تم الحصول عليها باستخدام برنامج Minitab .

Regression Analysis: y versus x1; x2

جزء (١)

```
The regression equation is
y = 23.1 ---- x1 + 0.531 x2
```

جزء (٢)

Predictor	Coef	SE Coef	T	P
Constant	------	2.205	10.49	0.000
x1	-1.9224	0.2352	-8.17	0.000
x2	0.5305	0.1103	----	0.000

جزء (٣)

```
S = ----        R-Sq = ----        R-Sq(adj) = 91.9%
```

جزء (٤)

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	--	788.63	------	---	0.000
Residual Error	--	63.61	2.89		
Total	24	------			

والمطلوب :

أ ـ أكمل النتائج المفقودة في كل جزء باللغة العربية .

ب ـ ما هي نسبة الاختلافات الغير مفسرة ؟

جـ ـ اختبر الفرض ف. : $\beta_1 = \beta_2 =$ صفر ، α = ٠,٠١

د ـ اختبر الفرض : $\beta_2 \neq$ صفر ، α = ٠,٠٥ .

الحل :

أ ـ إكمال النتائج المفقودة .

ـ جزء (١) : y = 23.1 ---- x1 + 0.531 x2

القيمة المفقودة هي معامل انحدار ب١ ، يوجد في جزء (٢) العمود الثاني أي أن

ب١ = ـ١,٩٢٢ ومن ثم تكتب المعادلة باللغة العربية كما يلي :

ص̂ = ٢٣,١ ـ ١,٩٢٢ س١ + ٠,٥٣١ س٢ .

ـ جزء (٢)

Predictor	Coef	SE Coef	T	P
Constant	------	2.295	10.09	0.000
x1	1.9224	0.2357	8.17	0.000
x2	0.5305	0.1103	----	0.002

العمود الثاني Coef المعاملات :

من معادلة الانحدار في جزء (١) نجد أن : الثابت أ = ٢٣,١

العمود الثالث T (احصاء الاختبار ت) :

ت س٢ = ب س٢ ÷ ع ب٢ = ٠,٥٣٠٥ ÷ ٠,١١٠٣ = ٤,٨١

ـ جزء (٣) S = ---- R-Sq = --- R-Sq(adj) = 91.9%

من جدول تحليل التباين نجد أن S (الخطأ المعياري للتقدير) :

(ع ص . س) = $\sqrt{\text{متوسط مربعات الأخطاء}}$ = $\sqrt{٢,٨٩}$ = ١,٧

من جدول تحليل التباين نجد أن R-sq (معامل التحديد) هو :

ر^{٢} = SSR(م.م.ح) ÷ SST(م.م.ك)

SST(م.م.ك) = SSR(م.م.ح) + SSE(م.م.خ) = ٧٨٨,٦٣ + ٦٣,٦١ = ٨٥٢,٢٤

ر^{٢} = ٧٨٨,٦٣ ÷ ٨٥٢,٢٤ = ٠,٩٢٥ ⟸ ٩٢,٥٪

— جزء (٤) جدول تحليل التباين :

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	--	788.63		------	0.000
Residual Error		63.61	2.89		
Total	24	--- -			

عمود DF (درجات الحرية) نجد أن :

درجات حرية الانحدار = عدد المتغيرات المستقلة = م = ٢

درجات حرية الأخطاء = (ن – م – ١) – (٢٤ – ٢) = ٢٢

عمود SS (مجموع المربعات) نجد أن :

SST(م.م.ك) = ٨٥٢,٢٤ سبق الحصول عليها .

عمود MS (متوسط المربعات) نجد ان :

MSR (ع$^{٢}_{ح}$) = SSR(م.م.ح) ÷ م = ٧٨٨,٦٣ ÷ ٢ = ٣٩٤,٣١

عمود F (إحصاء الاختبار ف) نجد أن :

F (ف) = ع$^{٢}_{ح}$ ÷ ع$^{٢}_{م}$ = ٣٩٤,٣٢ ÷ ٢,٨٩ = ١٣٦,٤٤

ب – نسبة الاختلافات الغير مفسرة = ١ – ر^{٢} = ١ – ٠,٩٢٥ = ٠,٠٧٥ ⟸ ٧,٥٪

جـ - اختبار الفرض ف. : β_1 = β_2 = صفر

— ف. : β_1 = β_2 = صفر ف$_1$: β_1 ≠ β_2 ≠ صفر

— التوزيع المستخدم هو توزيع ف بدرجات حرية بسط ، ومقام (٢ ، ٢٢)

– إحصاءة الاختبار : ف = ١٣٦,٤٤

– مستوى المعنوية ، α = ٠,٠١ ، الاحتمال المشاهد p = صفر

– القرار : بما أن مستوى المعنوية(α = ٠,٠١) أكبر من الاحتمال المشاهد (p = صفر) لذا يرفض الفرض العدمي ويقبل الفرض البديل ، ويدل ذلك على أن النموذج المقدر أفضل تمثيل للبيانات .

د – اختبار الفرض : β_2 ≠ صفر

– $ف_0$: β_2 = صفر $ف_1$: β_2 ≠ صفر

– التوزيع المستخدم هو توزيع ت بدرجات حرية (ن–م–١)= ٢٢

– إحصاءة الاختبار : T (ت) = ٨,١٧٠

– مستوى المعنوية (α = ٠,٠٥) ، الاحتمال المشاهد (p = صفر)

– القرار : بما أن مستوى المعنوية أكبر من الاحتمال المشاهد لذا يرفض الفرض العدمي ، ويقبل الفرض البديل ، ويدل ذلك على أن المتغير المستقل س، له أثر معنوي على المتغير التابع ص .

تمارين

١ ـ في دراسة قام بها طبيب لمعرفة العلاقة بين مقدار النقص في الوزن Amount بالكيلو جرام ، والفترة الزمنية Period بالشهور لتعاطي عقار معين يستخدم في التخسيس ، جرب الطبيب هذا العقار وحصل على البيانات التالية :

مقدار النقص	١	١	٢	٣	٤	٤	٥	٦	٦	٧
الفترة الزمنية	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠

واستخدم برنامج Minitab وحصل على النتائج التالية :

Regression Analysis: Amount versus Periode

The regression equation is

Amount = - 0.067 + 0.739 Period

Predictor	Coef	SE Coef	T	P
Constant	-0.0667	0.2287	-0.29	0.778
Period	0.73939	0.03667	20.16	0.000

S = 0.3349 R-Sq = 98.1% R-Sq(adj) = 97.8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	45.103	45.103	407.27	0.000
Residual Error	8	0.897	0.112		
Total	9	46.000			

والمطلوب :

١ - استخدم البيانات أعلاه للتحقق من النتائج التي تم الحصول عليها .

٢ - اكتب معادلة التنبؤ وفسر معناها .

٣ - احسب معامل الارتباط ، وما مدلوله؟ ثم اختبر الفرض : ρ = صفر ، [illegible]

٤ - اختبر صلاحية النموذج ، [illegible]

٢ ـ لدراسة وتحليل أثر الدخل العائلي (income) بالألف جنيه ، وحجم الأسرة (size) على الإنفاق الاستهلاكي للأسرة (expen.) بالألف جنيه خلال السنة

- تم جمع بيانات عن ١٥ أسرة ، وتم الحصول على النتائج التالية :

Regression Analysis: expen versus income; size

The regression equation is
expen = - 1.07 + 0.644 income + 0.560 size

Predictor	Coef	SE Coef	T	P
Constant	-1.0740	0.7320	-1.47	0.169
income	0.6440	0.1958	3.29	0.006
size	0.5800	0.2855	2.03	0.065

S = 0.9141 R-Sq = 86.2% R-Sq(adj) = 83.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	62.461	31.231	37.37	0.000
Residual Error	12	10.028	0.836		
Total	14	72.489			

والمطلوب : أ – صياغة نموذج التنبؤ باللغة العربية ، ثم فسر مدلوله .

ب – ما هو مدلول معامل التحديد ؟ ثم اختبر صلاحية النموذج ، α = ٠,٠١

جـ - هل تؤيد أن للدخل أثر إيجابي على الإنفاق ؟ α = ٠,٠١

د – هل إضافة حجم الأسرة يحسن في القدرة التنبؤية للنموذج ؟ α = ٠,٠١

٣ - تم الحصول على النتائج التالية باستخدام برنامج Minitab .

Regression Analysis: y versus x1; x2, x3; x4

Predictor	Coef	SE Coef	T	P
Constant	-1.412	1.575	-0.90	0.384
x1	0.3935	0.1436	2.74	0.015
x2	0.6323	0.2003	3.16	0.007
x3	0.00133	0.08142	0.02	0.983
x4	0.9163	0.4800	1.91	0.076

Analysis of Variance

Source	DF	SS	MS	F	P
Regression		115.895	----	-----	0.000
Residual Error	--	------	0.709		
Total	19	------			

والمطلوب : أ – ما هي حجم العينة ؟ وما هي عدد المتغيرات المفسرة في النموذج ؟

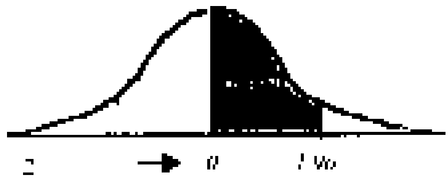
ب- اكتب معادلة التنبؤ باللغة العربية ، ثم فسر مدلولها .

جـ - أكمل جدول تحليل التباين ، ثم احسب المقاييس الدالة على جودة النموذج .

ملحق

الجداول الإحصائية

جدول التوزيع الطبيعي



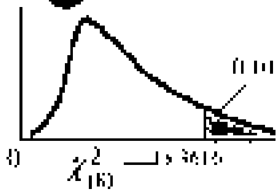
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

جدول توزيع ت

Probability

Degrees of freedom	0.40	0.25	0.10	0.05	0.025	0.01	0.005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576

جدول توزيع كا²



Degrees of freedom	probability							
	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010
1	0.0002	0.0010	0.0039	0.0158	2.7055	3.8415	5.0239	6.6349
2	0.0201	0.0506	0.1026	0.2107	4.6052	5.9915	7.3778	9.2103
3	0.1148	0.2158	0.3518	0.5844	6.2514	7.8147	9.3484	11.3449
4	0.2971	0.4844	0.7107	1.0636	7.7794	9.4877	11.1433	13.2767
5	0.5543	0.8312	1.1455	1.6103	9.2364	11.0705	12.8325	15.0863
6	0.8721	1.2373	1.6354	2.2041	10.6446	12.5916	14.4494	16.8119
7	1.2390	1.6899	2.1673	2.8331	12.0170	14.0671	16.0128	18.4753
8	1.6465	2.1797	2.7326	3.4895	13.3616	15.5073	17.5345	20.0902
9	2.0879	2.7004	3.3251	4.1682	14.6837	16.9190	19.0228	21.6660
10	2.5582	3.2470	3.9403	4.8652	15.9872	18.3070	20.4832	23.2093
11	3.0535	3.8158	4.5748	5.5778	17.2750	19.6751	21.9200	24.7250
12	3.5706	4.4038	5.2260	6.3038	18.5493	21.0261	23.3367	26.2170
13	4.1069	5.0087	5.8919	7.0415	19.8119	22.3620	24.7356	27.6882
14	4.6604	5.6287	6.5706	7.7895	21.0641	23.6848	26.1189	29.1412
15	5.2294	6.2621	7.2609	8.5468	22.3071	24.9958	27.4884	30.5779
16	5.8122	6.9077	7.9616	9.3122	23.5418	26.2962	28.8454	31.9999
17	6.4078	7.5642	8.6718	10.0852	24.7690	27.5871	30.1910	33.4087
18	7.0149	8.2308	9.3905	10.8649	25.9894	28.8693	31.5264	34.8053
19	7.6327	8.9065	10.1170	11.6509	27.2036	30.1435	32.8523	36.1909
20	8.2604	9.5908	10.8508	12.4426	28.4120	31.4104	34.1696	37.5662
21	8.8972	10.2829	11.5913	13.2396	29.6151	32.6706	35.4789	38.9322
22	9.5425	10.9823	12.3380	14.0415	30.8133	33.9244	36.7807	40.2894
23	10.1957	11.6885	13.0905	14.8480	32.0069	35.1725	38.0756	41.6384
24	10.8564	12.4011	13.8484	15.6587	33.1962	36.4150	39.3641	42.9798
25	11.5240	13.1197	14.6114	16.4734	34.3816	37.6525	40.6465	44.3141
26	12.1981	13.8439	15.3792	17.2919	35.5632	38.8852	41.9232	45.6417
27	12.8785	14.5734	16.1514	18.1139	36.7412	40.1133	43.1945	46.9629
28	13.5647	15.3079	16.9279	18.9392	37.9159	41.3372	44.4607	48.2782
29	14.2565	16.0471	17.7084	19.7677	39.0875	42.5569	45.7222	49.5879
30	14.9535	16.7908	18.4926	20.5992	40.2560	43.7729	46.9792	50.8922
40	22.1643	24.4331	26.5093	29.0505	51.8050	55.7585	59.3417	63.6907
50	29.7067	32.3574	34.7642	37.6886	63.1671	67.5048	71.4202	76.1539
60	37.4848	40.4817	43.1879	46.4589	74.3970	79.0819	83.2976	88.3794
70	45.4418	48.7576	51.7393	55.3290	85.5271	90.5312	95.0231	100.4250
80	53.5400	57.1532	60.3915	64.2778	96.5782	101.8790	106.6290	112.3290
90	61.7541	65.6466	69.1260	73.2912	107.5650	113.1450	118.1360	124.1160
100	70.0648	74.2219	77.9295	82.3581	118.498	124.342	129.561	135.807

جدول توزيع ف

	1	2														
1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
21	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
22	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
120	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

جدول توزيع ف ، α = 1%

	1	[illegible]	3	4	5	6	7	8	9	10	12	15	20	24
1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	6106	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	28.7	28.	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	27.1	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	16.0	15.5	[illegible]	[illegible]	[illegible]	[illegible]	14.5	14.4	14.2	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	11.4	11.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	9.89	9.72	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	8.75	8.47	[illegible]	[illegible]	7.98	7.87	7.72	[illegible]	7.40	[illegible]
[illegible]	[illegible]	[illegible]	8.45	[illegible]	[illegible]	[illegible]	[illegible]	6.84	6.72	6.62	6.47	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	7.01	[illegible]	[illegible]	[illegible]	6.03	5.91	5.81	5.67	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	6.42	6.06	5.80	[illegible]	[illegible]	[illegible]	5.26	5.11	[illegible]	4.81	4.73
[illegible]	[illegible]	[illegible]	[illegible]	5.99	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4.85	4.71	4.56	4.41	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4.63	4.54	4.40	4.25	4.10	[illegible]
12	[illegible]	[illegible]	[illegible]	5.41	5.06	[illegible]	[illegible]	[illegible]	4.39	4.30	4.16	4.01	3.86	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4.86	[illegible]	[illegible]	[illegible]	4.19	4.10	3.96	3.82	3.66	3.59
14	[illegible]	[illegible]	[illegible]	5.04	4.69	4.46	4.28	[illegible]	4.03	3.94	3.80	3.66	3.51	[illegible]
15	[illegible]	[illegible]	[illegible]	4.89	4.56	4.32	[illegible]	[illegible]	3.89	3.80	3.67	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	4.77	4.44	4.20	[illegible]	[illegible]	[illegible]	[illegible]	3.55	3.41	[illegible]	[illegible]
17	[illegible]	[illegible]	[illegible]	4.67	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.46	[illegible]	[illegible]	[illegible]
18	[illegible]	[illegible]	[illegible]	4.58	[illegible]	4.01	[illegible]	[illegible]	[illegible]	[illegible]	3.37	[illegible]	[illegible]	[illegible]
19	[illegible]	[illegible]	[illegible]	[illegible]	4.17	3.94	[illegible]	[illegible]	[illegible]	[illegible]	3.30	[illegible]	[illegible]	[illegible]
20	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.23	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	4.87	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.17	[illegible]	2.88	[illegible]
22	[illegible]	[illegible]	4.82	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.26	[illegible]	2.98	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.21	3.07	2.93	2.78	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	4.22	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.17	3.03	2.89	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	4.18	[illegible]	[illegible]	3.46	[illegible]	3.22	3.13	2.99	2.85	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	4.64	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.18	3.09	2.96	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.15	3.06	2.93	2.78	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.12	3.03	2.90	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.00	2.87	[illegible]	[illegible]	2.49
30	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2.98	2.84	[illegible]	[illegible]	[illegible]
40	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2.89	2.80	2.66	2.52	[illegible]	[illegible]
60	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2.82	[illegible]	2.63	2.50	[illegible]	2.20	[illegible]
120	[illegible]	[illegible]	[illegible]	3.48	3.17	[illegible]	[illegible]	2.66	2.56	2.47	2.34	2.19	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2.80	[illegible]	2.51	2.41	2.32	2.18	[illegible]	1.88	[illegible]

المراجـــع

١- سهير فهمي حجازي : " الطرق الإحصائية في التطبيقات التجارية " ، الطبعة الأولى - المكتبة الأكاديمية ، ١٩٩٩ .

2- Mann Prem S. "Statistics for Business An Economics " . John Wiley & Sons, Inc.,1995 .

3- Mansfield Edwin "Statistics for Business and Economics ", W.W.Norton & Company , 1980 .

4- Mason R.D., Lind D.A., & Marchal W. G. "Statistical Techniques in Business and Economics ",10th ed., McGraw-Hill, 1999 .

5- Morse, Lawrence C. "Statistics for Business and Economics " . Harper Collins College Publisher, 1993 .

6- Wonnacott T.H. and Wonnacott R.J. "Introductory Statistics for Business And Economics", 2nd ed ., John Wiley & Sons Inc., 1977 .

جدول المحتويات

الموضوع	الصفحة