اختبار تحليل التباين

يستخدم اختبار تحليل التباين لاختبار الفروق بين متوسطات ثلاث عينات فأكثر، ويشترط لاستخدامه بأن تكون البيانات تتبع التوزيع الطبيعي . يستخدم في حالة وجود متغير تابع وله متغير مستقل ولكن بمستويات متعددة.

جدول (1) المتغيرات المستقلة والتابعة في تحليل التباين الأحادي One – Way ANOVA

المتغير التابع درجة التحضر	**المتغير المستقل: المجموعات الاقتصادية (ثلاثة مستويات) Factor**				
	الدول الصناعية	الشرق الوسط	إفريقيا		
المتغير التابع الدخل الشهري	**المتغير المستقل: المستوى التعليمي أربعة مستويات Factor**				
	ابتدائي	إعدادي	ثانوي	جامعي	
المتغير التابع عدد العمال	**المتغير المستقل: (تصنيف المصانع) خمسة مستويات Factor**				
	مصانع صغيرة جداً	مصانع صغيرة	مصانع متوسطة	مصانع كبيرة	مصانع كبيرة جدا

نلاحظ في تحليل التباين الأحادي وجود متغير مستقل واحد (بثلاثة مستويات فأكثر) ومتغير تابع واحد. وفيه نختبر فرضية اختلاف الأوساط الحسابية للمتغير التابع بين المستويات الموجودة في المتغير المستقل.

منطق اختبار تحليل التباين: التباين هو مقياس للتشتت والاختلاف، فالاختبار يقيس التباين داخل كل مجموعة(عينة) With Variance ويقيس التباين بين المجموعات Between Variance

ومن شروط استخدام اختبار تحليل التباين :

1- أن تكون مفردات العينات مستقلة

2- أن يكون المتغير التابع مقاس على الأقل على المستوى الفئوي.

3- المتغير التابع موزع توزيعاً قريباً من التوزيع الطبيعي في كل مجموعة.

4- تجانس التباين بين المجموعات

السؤال الأول : في دراسة حول حجم حبيبات البرد الساقطة على إحدى المدن ،قام أحد الباحثين بتقسيم المدينة إلى أربعة أقسام حسب بعدها عن مركز المدينة ، لمعرفة أثر التلوث الهوائي في حجم حبيبات البرد ، فحصل على البيانات التالية :

المطلوب: **هل يوجد اختلاف حقيقي بين حجم حبيبات البرد في المناطق الأربع؟ وهل يوجد اختلاف بيت المنطقة A والمنطقة B ؟**

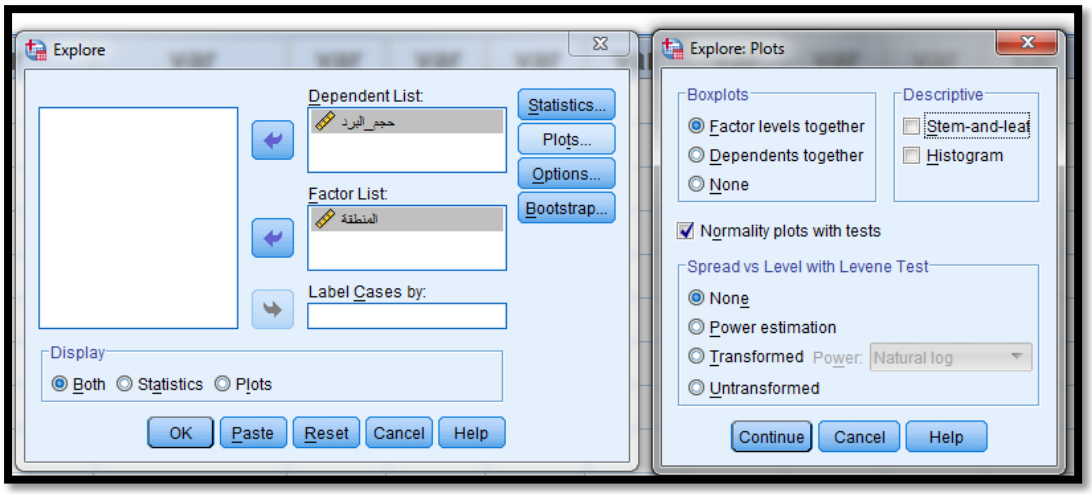
المركز A	الثانية B	الثالثة C	الرابعة D
10	8	8	11
11	10	9	12
18	15	12	10
17	12	10	12
12	10	6	8
13	11	12	9
15	14	11	8
8	7	6	6
6	6	5	3
7	5	5	4
13	12	5	12

المطلوب: **هل يوجد اختلاف حقيقي بين حجم حبيبات البرد في المناطق الأربع؟ وهل يوجد اختلاف بيت المنطقة A والمنطقة B ؟**

	حجم_البرد	المنطقة
1	10	المركز
2	11	المركز
3	18	المركز
4	17	المركز
5	12	المركز
6	13	المركز
7	15	المركز
8	8	المركز
9	6	المركز
10	7	المركز
11	13	المركز
12	8	الثاني
13	10	الثاني
14	15	الثاني
15	12	الثاني

تختبر إذا ما كانت البيانات تتبع التوزيع الطبيعي

فرضية العدم : البيانات تتبع التوزيع الطبيعي **الفرضية البديلة** : البيانات لا تتبع التوزيع الطبيعي



Tests of Normality							
	المنطقة	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	Df	Sig.
الحجم	A	.109	11	.200*	.967	11	0.851
	B	.136	11	.200*	.966	11	0.848
	C	.223	11	.131	.868	11	0.073
	D	.149	11	.200*	.902	11	0.196

بما أن قيمة مستوى الدلالة في العينات الأربع أكبر من 0.05 ، لذا نقبل فرضية العدم القائلة **بأن البيانات تتبع التوزيع الطبيعي**، لذلك سنستخدم **اختبار تحليل التباين**

2- نختبر إذا ما كانت العينات متجانسة التباين :

فرضية العدم : يوجد تجانس للتباين بين العينات الأربع

الفرضية البديلة : لا يوجد تجانس للتباين بين العينات الأربع

Test of Homogeneity of Variance					
		Levene Statistic	df1	df2	Sig.
البرد_حجم	Based on Mean	0.306	3	40	0.821

النتيجة : بما أن قيمة اختبار LEVNE = 0.306 ومستوى الدلالة= 0.821، وهي أكبر من 0.05 ، لذلك نقبل فرضية العدم القائلة بوجود تجانس للتباين.

لذلك نستخدم اختبار تحليل التباين ، واختبار LS'D

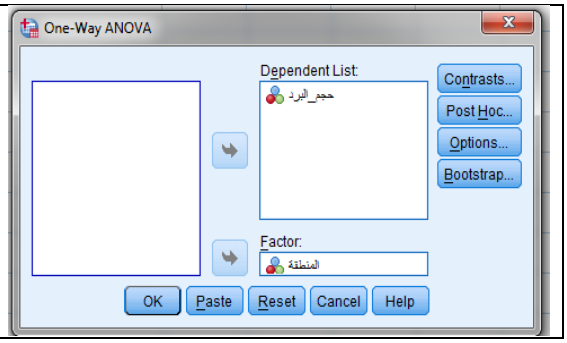
	N	Mean	Std. Deviation
المركز	11	11.82	3.920
الثاني	11	10.00	3.225
الثالث	11	8.09	2.844
الرابع	11	8.64	3.202
Total	44	9.64	3.518

من الجدول المقابل : نجد أن الفرق بين متوسطات أحجام الحبيبات مختلف ، فهل الاختلاف بين هذه المتوسطات اختلاف حقيقي ؟؟

فرضية العدم : لا يوجد اختلاف حقيقي بين متوسطات العينات الأربع ، أي أن المتوسطات الأربعة متساوية

الفرضية البديلة : يوجد اختلاف حقيقي بين متوسطات العينات الأربع ، المتوسطات الأربعة غير متساوية

NOVA					
	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	91.091	3	30.364	2.754	.055
Within Groups	441.091	40	11.027		
Total	532.182	43			



نجد أن قيمة اختبار تحليل التباين = 2.754 و مستوى الدلالة = 0.055 و هي أكبر من 0.05

لذلك لا نستطيع أن نرفض فرضية العدم ولا نقبل الفرضية البديلة

القرار : الفرق بين متوسط العينات فرق ليس له دلالة احصائية .

أظهرت نتيجة اختبار التباين الأحادي أن الفروق بين المتوسطات الحسابية بين المجموعات الأربعة، فهل كل المجموعات الفروق بينها غير دالة إحصائياً أم أن هناك مجموعات دالة إحصائياً :

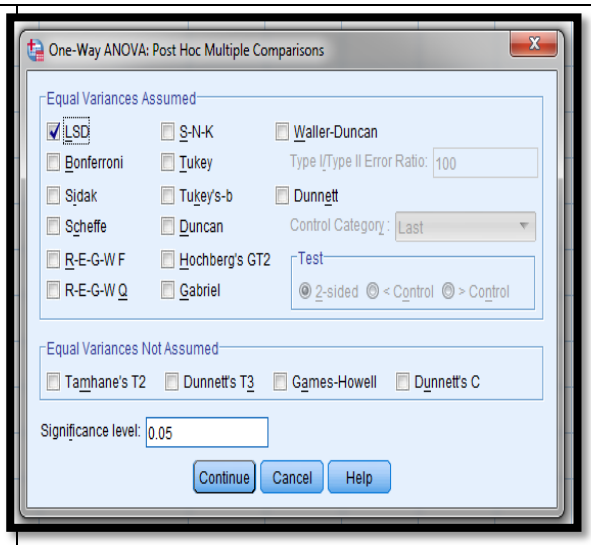
نجري اختبار Post Hoc لمعرفة الفرق بين العينات الأربع بالشكل التالي :

Multiple Comparisons LSD

Dependent Variable: البرد_حجم

(I) المنطقة	(J) المنطقة	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
المركز	الثاني	1.818	1.416	0.207	-1.04-	4.68
	الثالث	3.727*	1.416	0.012	0.87	6.59
	الرابع	3.182*	1.416	0.030	0.32	6.04
الثاني	المركز	-1.818-	1.416	0.207	-4.68-	1.04
	الثالث	1.909	1.416	0.185	-0.95-	4.77
	الرابع	1.364	1.416	0.341	-1.50-	4.23
الثالث	المركز	-3.727-*	1.416	0.012	-6.59-	-.87-
	الثاني	-1.909-	1.416	0.185	-4.77-	.95
	الرابع	-.545-	1.416	0.702	-3.41-	2.32
الرابع	المركز	-3.182-*	1.416	0.030	-6.04-	-.32-
	الثاني	-1.364-	1.416	0.341	-4.23-	1.50
	الثالث	.545	1.416	0.702	-2.32-	3.41

*. The mean differece is significant at the 0.05 level.



من الجدول: الفرق بين متوسط حبيبات البرد :

* **دال إحصائياً** بين مركز المدينة والمستوى الثالث (0.012) والمستوى الرابع (0.03)

* **غير دال إحصائياً** للفروق بين باقي المستويات

السؤال الثاني : أخذت ثلاث عينات من مزارعي الفراولة حول كمية الأسمدة التي تضاف للتربة في محافظة شمال غزة

رقم المنطقة التي أخذت منها العينة	1	2	3	4	5	6	7	8	9	10	11
العينة " 1 "(تربة رملية)	1000	1200	1500	700	600	800	500	900	1500	1800	2000
العينة " 2 "(تربة رملية طينية)	800	752	1533	1200	1000	900	800	750	850	654	123
العينة "3 " (تربة طينية)	700	800	500	300	200	150	100	400	300	450	600

المطلوب: **هل يوجد اختلاف حقيقي بين متوسطات العينات الثلاث ؟ وهل يوجد اختلاف حقيقي بين العينة الأولى والثانية؟**

خطوات الحل :

أولاً / نختبر إذا ما كانت البيانات تتبع توزيع طبيعي أم لا .

فرضية العدم : البيانات لا تتبع التوزيع الطبيعي **الفرضية البديلة** : البيانات تتبع التوزيع الطبيعي

بعد إدخال البيانات لبرنامج SPSS نحصل على النتيجة التالية :

Tests of Normality							
	العينة	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
كميةالسماد	العينة1	.152	11	.200*	.939	11	0.508
	العينة2	.204	11	.200*	.918	11	0.300
	العينة3	.139	11	.200*	.964	11	0.817

بما أن مستوى الدلالة في اختبار شابيرو في العينات الثلاث (0.508 ، 0.3 ، 0.817) أكبر من 0.05 ، فإننا نقبل الفرضية المبدئية القائلة **بأن البيانات تتبع التوزيع الطبيعي** .

النتيجة : يجب استخدام اختبار تحليل التباين

ثانياً / اختبار تحليل التباين

فرضية العدم : لا يوجد فرق بين متوسطات العينات الثلاث **الفرضية البديلة :** يوجد فرق حقيقي بين متوسطات العينات الثلاث

ANOVA					
	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	2954123.879	2	1477061.939	10.438	0.0000
Within Groups	4245070.364	30	141502.345		
Total	7199194.242	32			

و بما أن قيمة اختبار تحليل التباين = 10.438 و مستوى الدلالة = 0.000 ، لذلك نرفض فرضية العدم ونقبل الفرضية البديلة أي أن يوجد اختلاف حقيقي بين متوسطات العينات الثلاث عند مستوى دلالة 0.01

ثالثاً / اختبار Bost Hoc : لمعرفة الاختلاف بين متوسطات كل عينتين

أ- معرفة تجانس التباين للعينات الثلاث:

Test of Homogeneity of Variances			
كمية الانتاج من الفراولة			
Levene Statistic	df1	df2	Sig.
3.918	2	30	0.031

فرضية العدم: يوجد تجانس للتباين بين العينات الثلاث

الفرضية البديلة: لا يوجد تجانس للتباين بين العينات الثلاث

نجد أن مستوى الدلالة 0.031 وهو أقل من 0.05، لذلك نرفض فرضية العدم ونقبل الفرضية البديلة ، أي أن البيانات للعينات الثلاث ليس بها تجانس للتباين.

القرار : بما أن بيانات العينات الثلاث غير متجانسة التباين، فإننا نستخدم من خيارات Bost Hoc ، الخيار عدم تجانس التباين ونختار الخيار: Tamhane :

Dependent Variable: الكمية لانتاجية الفراولة						
		Mean			95% Confidence Interval	
(I) التربة_نوع	(J) التربة_نوع	Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
رملية	طينية رملية	285.273	184.168	0.362	-199.91-	770.46
	طينية	727.273*	166.229	0.002	276.55	1177.99
طينية رملية	رملية	-285.273-	184.168	0.362	-770.46-	199.91
	طينية	442.000*	125.031	0.007	111.51	772.49
طينية	رملية	-727.273-*	166.229	0.002	-1177.99-	-276.55-
	طينية رملية	-442.000-*	125.031	0.007	-772.49-	-111.51-

بتطبيق اختبار Tamhane نجد أن:

* دالة إحصائياً الفروق للإنتاجية بين التربة الرملية والتربة الطينية ، وبين التربة الطينية الرملية والتربة الطينية .

* غير دالة إحصائياً: الفرق بين عينة التربة الرملية والتربة الطينية الرملية.

ثانياً / تحليل التباين الثنائي

يوجد في اختبار تحليل التباين الثنائي (Two – Way ANOVA) متغيران مستقلان ومتغير تابع واحد، مثلاً في دراسة تأثير الجنس والمستوى التعليمي على عدد ساعات العمل الأسبوعية، فالمتغير المستقل الأول الجنس (له مستويان: ذكور وإناث)، بينما المتغير المستقل الثاني المستوى التعليمي له أربعة مستويات: (ثانوية عامة، دبلوم، جامعة، شهادة عليا)، ويوجد متغير تابع واحد هو متغير عدد ساعات العمل الأسبوعية، ويختبر تحليل التباين الثنائي فرضية اختلاف المتوسطات الحسابية للمجموعات المتكونة من تقاطع مستويات المتغيرين المستقلين.

جدول (2) الخلايا أو المجموعات في اختبار تحليل التباين الثنائي

المستوى التعليمي	ثانوية عامة	دبلوم	جامعة	شهادة عليا
ذكر				
أنثى				

لماذا دراسة تأثير المتغيرين المستقلين معاً؟

يسمح تحليل التباين الثنائي بدراسة تأثير المتغيرين المستقلين معاً بدلاً من تأثير كل منهما بشكل منفرد، ودراسة تأثير المتغيرين المستقلين معاً هو حقيقة أن تأثيرهما معاً قد يختلف عن تأثير كل منهما على حده. فوجود متغيرين معاً يحدث اختلافاً في التأثير، وهذا التأثير المشترك هو ما يعرف بالتفاعل Interaction، ويدرس تحليل التباين الثنائي التفاعل (التأثير المشترك)، كما يدرس التأثير الخاص بكل متغير على حده، ويجب أن يركز الباحث على دراسة التأثير المشترك ومدى تأثير تفاعلهما على المتغير التابع، إن اختبار تحليل التباين الثنائي يوفر الجهد والتكاليف والوقت مقارنة مع إجراء عدة اختبارات تحليل تباين أحادية.

ماذا نعني بالتفاعل بين المتغيرين Interaction المستقلين ؟

نعني بالتفاعل بين المتغيرين المستقلين أن تأثير أي منهما يعتمد على وجود أو غياب المتغير الآخر، أو أن أحد المتغيرين يعتمد على مستوى المتغير الثاني، أو أن تأثير المتغيرين معاً يختلف (يقل أو يزيد) عن تأثيرهما وهما منفصلين.

افتراضات اختبار تحليل التباين الثنائي:

1- استقلال المفردات

2- المتغير التابع مقاس على الأقل على المستوى الفئوي

3- المتغير التابع موزع توزيعاً قريباً من التوزيع الطبيعي قس كل مجموعة.

4- تجانس التباين بين المجموعات.

مثال :

ما هو تأثير المستوى الدراسي لطالب الجامعة (سنة أولى ، ثانية ، ثالثة، رابعة) وجنسه ذكور أو إناث على عدد ساعات المذاكرة الأسبوعية؟

مسلسل	الجنس	المستوى الدراسي	عدد ساعات المذاكرة	مسلسل	الجنس	المستوى الدراسي	عدد ساعات المذاكرة
1-	ذكر	سنة أولى	15	16-	أنثى	سنة ثالثة	37
2-	أنثى	سنة ثانية	35	17-	ذكر	سنة أولى	18
3-	ذكر	سنة أولى	22	18-	ذكر	سنة ثانية	30
4-	ذكر	سنة رابعة	35	19-	أنثى	سنة أولى	15
5-	أنثى	سنة ثالثة	33	20-	أنثى	سنة رابعة	40
6-	أنثى	سنة ثانية	35	21-	ذكر	سنة أولى	20
7-	ذكر	سنة أولى	28	22-	ذكر	سنة ثانية	25
8-	ذكر	سنة ثالثة	30	23-	ذكر	سنة أولى	28
9-	أنثى	سنة رابعة	41	24-	ذكر	سنة ثالثة	29
10-	أنثى	سنة ثالثة	35	25-	أنثى	سنة أولى	28
11-	ذكر	سنة أولى	25	26-	ذكر	سنة ثانية	22
12-	ذكر	سنة ثانية	25	27-	أنثى	سنة أولى	15
13-	ذكر	سنة أولى	20	28-	ذكر	سنة رابعة	32
14-	أنثى	سنة رابعة	39	29-	أنثى	سنة رابعة	39
15-	أنثى	سنة ثالثة	36	30-	ذكر	سنة رابعة	34

خطوات الحل :

أولاً / فرضيات الاختبار:

الفرضية الصفرية (المبدئية): لا يوجد تفاعل بين المستوى الدراسي وجنس الطالب في التأثير على معدل ساعات مذاكرة الطالب الأسبوعية، أي أن تأثير المستوى الدراسي على معدل المذاكرة الاسبوعية لا يتأثر بكون الطالب ذكراً أو أنثى.

الفرضية الصفرية الفرعية :

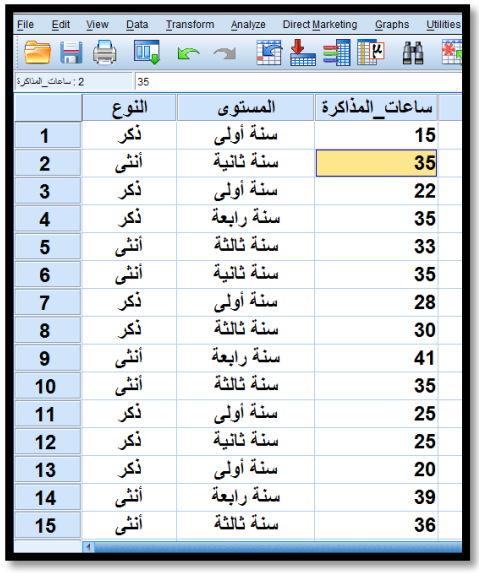
- لا يوجد اختلاف في معدل المذاكرة الاسبوعية لطالب الجامعة حسب مستواه الدراسي .
- لا يوجد اختلاف في معدل المذاكرة الاسبوعية لطالب الجامعة بين الذكور والإناث.

الفرضية البديلة(فرضية البحث): يوجد تفاعل بين المستوى الدراسي وجنس الطالب في التأثير على معدل ساعات مذاكرة الطالب الأسبوعية، أي أن تأثير المستوى الدراسي على معدل المذاكرة الاسبوعية يختلف بكون الطالب ذكراً أو أنثى.

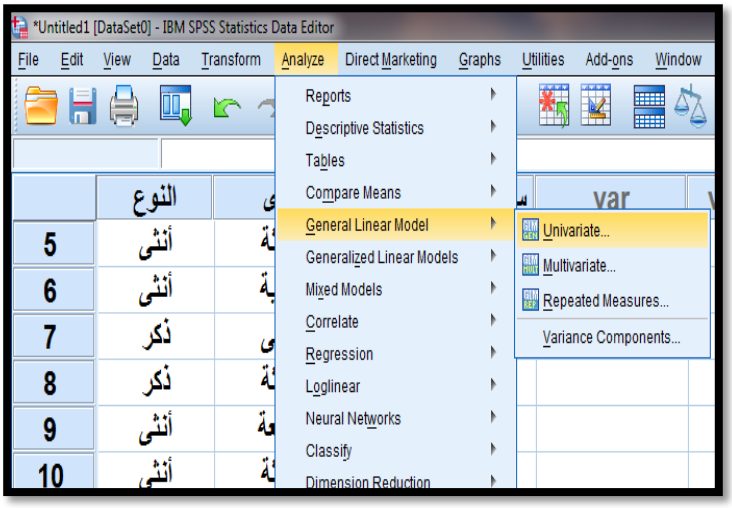
الفرضية البديلة الفرعية الخاصتان بالمتغيرين المستقلين فهما:

- يوجد اختلاف في معدل المذاكرة الاسبوعية لطالب الجامعة حسب مستواه الدراسي.
- يوجد اختلاف في معدل المذاكرة الاسبوعية لطالب الجامعة بين الذكور والإناث.

ثانياً / ندخل البيانات في الحاسوب كما هو في الشكل :

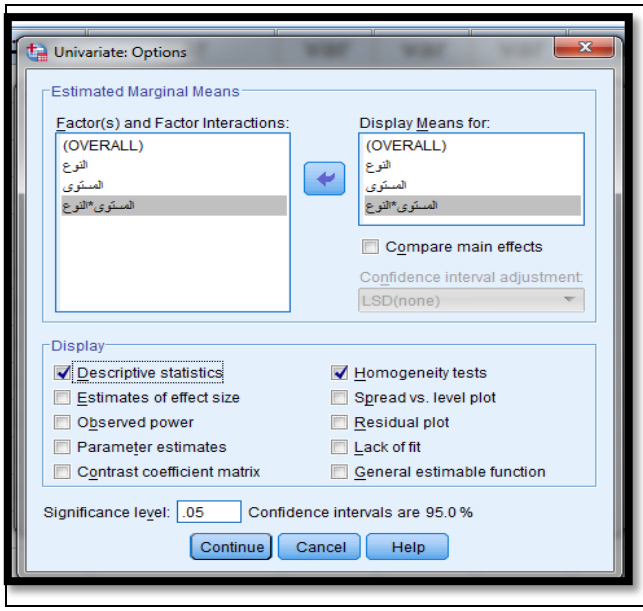


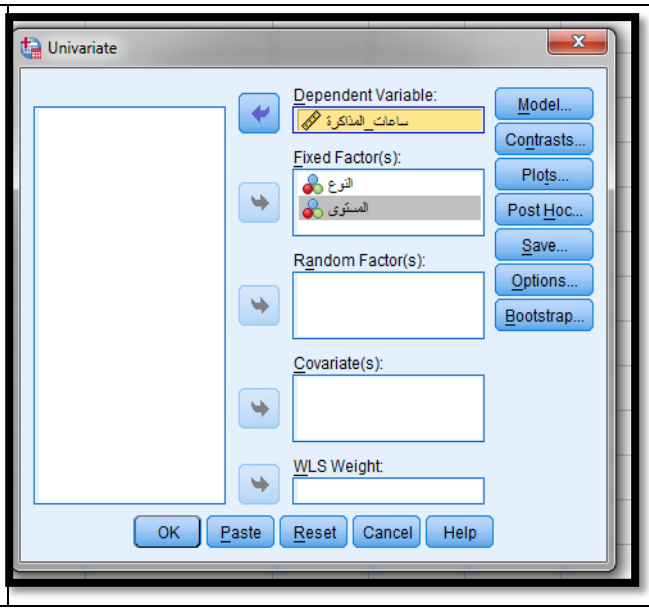
	النوع	المستوى	ساعات_المذاكرة
1	ذكر	سنة أولى	15
2	أنثى	سنة ثانية	35
3	ذكر	سنة أولى	22
4	ذكر	سنة رابعة	35
5	أنثى	سنة ثالثة	33
6	أنثى	سنة ثانية	35
7	ذكر	سنة أولى	28
8	ذكر	سنة ثالثة	30
9	أنثى	سنة رابعة	41
10	أنثى	سنة ثالثة	35
11	ذكر	سنة أولى	25
12	ذكر	سنة ثانية	25
13	ذكر	سنة أولى	20
14	أنثى	سنة رابعة	39
15	أنثى	سنة ثالثة	36



نافذة Univariate

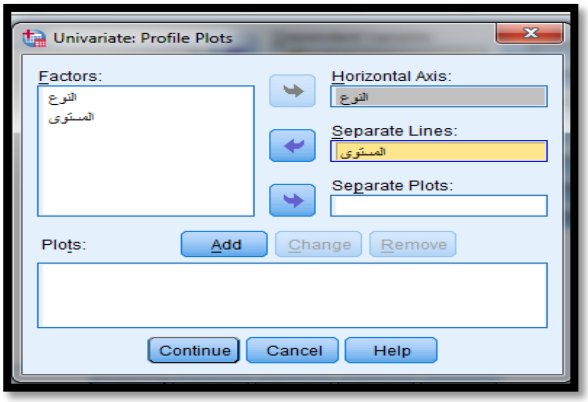
نافذة Univariate : Options





نافذة Univariate : Profile Plots

الجدول الأول



Between-Subjects Factors

		Value Label	N
النوع	1	ذكر	17
	2	أنثى	13
المستوى	1	أولى سنة	11
	2	ثانية سنة	6
	3	ثالثة سنة	6
	4	رابعة سنة	7

يحتوى الجدول الأول على عدد المفردات (الطلبة) في كل خلية أو مجموعة حسب المتغيرات المستقلة في المثال، فنلاحظ أن هناك 17 من الذكور ، و 13 من الإناث، و 11 طالباً من مستوى أول وهكذا

الجدول الثاني :

Descriptive Statistics

Dependent Variable: المذاكرة_ساعات

النوع	المستوى	Mean	Std. Deviation	N
ذكر	أولى سنة	22.00	4.690	8
	ثانية سنة	25.50	3.317	4
	ثالثة سنة	29.50	.707	2
	رابعة سنة	33.67	1.528	3
	Total	25.76	5.707	17
أنثى	أولى سنة	19.33	7.506	3
	ثانية سنة	35.00	.000	2
	ثالثة سنة	35.25	1.708	4
	رابعة سنة	39.75	.957	4
	Total	32.92	8.636	13
Total	أولى سنة	21.27	5.312	11
	ثانية سنة	28.67	5.538	6
	ثالثة سنة	33.33	3.266	6
	رابعة سنة	37.14	3.436	7
	Total	28.87	7.864	30

في الجدول الأعلى: نجد نتائج التحليل تحتوي على عدد الطلبة في كل خلية ناتجة من تقاطع المتغيرين المستقلين: الجنس وفيه مستويان، والمستوى الدراسي وفيه أربعة مستويات، كما يحتوي على الانحراف المعياري والمتوسط الحسابي لعدد ساعات المذاكرة لكل مجموعة

Levene's Test of Equality of Error Variances[a]

Dependent Variable: المذاكرة_ساعات

F	df1	df2	Sig.
3.929	7	22	**0.006**

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + النوع + المستوى + النوع * المستوى

الجدول السابق فيه اختبار **Levene's** لتجانس التباين في الخلايا :

فرضية العدم : يوجد تجانس للتباين في الخلايا

الفرضية البديلة : لا يوجد تجانس للتباين في الخلايا

وبما أن قيمة مستوى الدلالة 0.006 وهي أقل من 0.05 لذلك نرفض فرضية العدم ونقبل الفرضية البديلة القائلة بعدم وجود تجانس للتباين في الخلايا

جدول الاختبار الرئيسي للتباين الثنائي :

Tests of Between-Subjects Effects

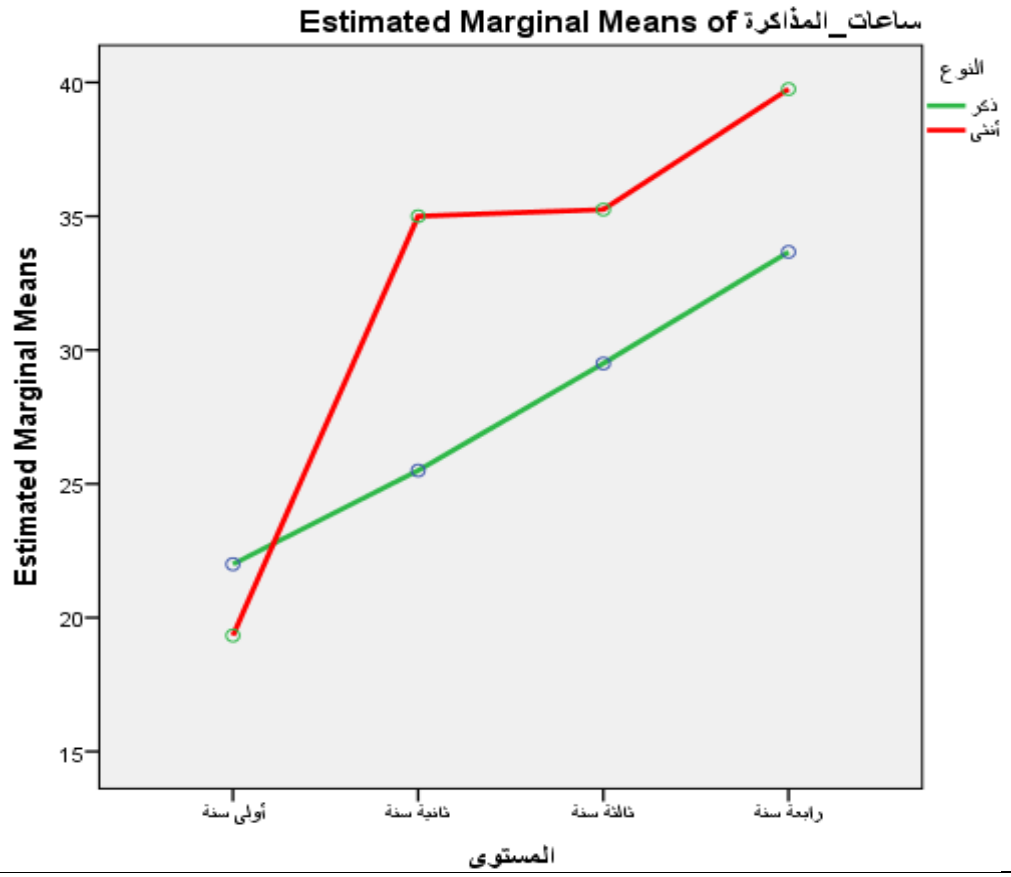
Dependent Variable: المذاكرة_ساعات

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1477.133[a]	7	211.019	14.676	.000
Intercept	22662.295	1	22662.295	1576.092	.000
النوع	137.093	1	137.093	9.534	.005
المستوى	1081.708	3	360.569	25.076	.000
المستوى * النوع	148.566	3	49.522	**3.444**	**0.034**
Error	316.333	22	14.379		
Total	26792.000	30			
Corrected Total	1793.467	29			

a. R Squared = .824 (Adjusted R Squared = .767)

نلاحظ نتيجة التفاعل بين المتغيرين المستقلين ، فنجد أن النتيجة المقابلة للسطر النوع * المستوى هي نتيجة التفاعل بين المتغيرين ، وهي نتيجة دالة إحصائياً حيث قيمة مستوى الدلالة لاختبار F تساوي 0.034 وهي أقل من 0.05 ، وهذا يعني أن التفاعل بين الجنس والمستوى التعليمي في تأثيرهما على معدل ساعات مذاكرة الطالب الجامعي دالة إحصائياً .

نلاحظ من الشكل البياني أن عدد ساعات المذاكرة عند الإناث أكبر من عند الذكور ما عدا المستوى الأول فنجد أن ساعات المذاكرة عند النوعين تكاد تكون متقاربة .



ولمعرفة أي المستويات التي يوجد بها هذا الفرق بدقة، نستخدم اختبار Post Hoc حسب LCD

(I) المستوى	(J) المستوى	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
أولى سنة	ثانية سنة	-7.39-*	1.924	**.001**	-11.39-	-3.40-
	ثالثة سنة	-12.06-*	1.924	**.000**	-16.05-	-8.07-
	رابعة سنة	-15.87-*	1.833	**.000**	-19.67-	-12.07-
ثانية سنة	أولى سنة	7.39*	1.924	.001	3.40	11.39
	ثالثة سنة	-4.67-*	2.189	.044	-9.21-	-.13-
	رابعة سنة	-8.48-*	2.110	.001	-12.85-	-4.10-
ثالثة سنة	أولى سنة	12.06*	1.924	.000	8.07	16.05
	ثانية سنة	4.67*	2.189	.044	.13	9.21
	رابعة سنة	-3.81-	2.110	.085	-8.18-	.57
رابعة سنة	أولى سنة	15.87*	1.833	.000	12.07	19.67
	ثانية سنة	8.48*	2.110	.001	4.10	12.85
	ثالثة سنة	3.81	2.110	.085	-.57-	8.18

نلاحظ أن مستوى الدلالة في جميع المستويات أكبر من 0.05 ، لذلك يمكن القول أن التباين بين المجموعات غير دال احصائياً .

انتهت المحاضرة