الربيع الثاني Q2 : وهي القيمة التي تقسم البيانات الى قسمين متساويين أي يسبقها 50% من المشاهدات عند ترتيبها تصاعدياً وهذه القيمة هي الوسيط Median .

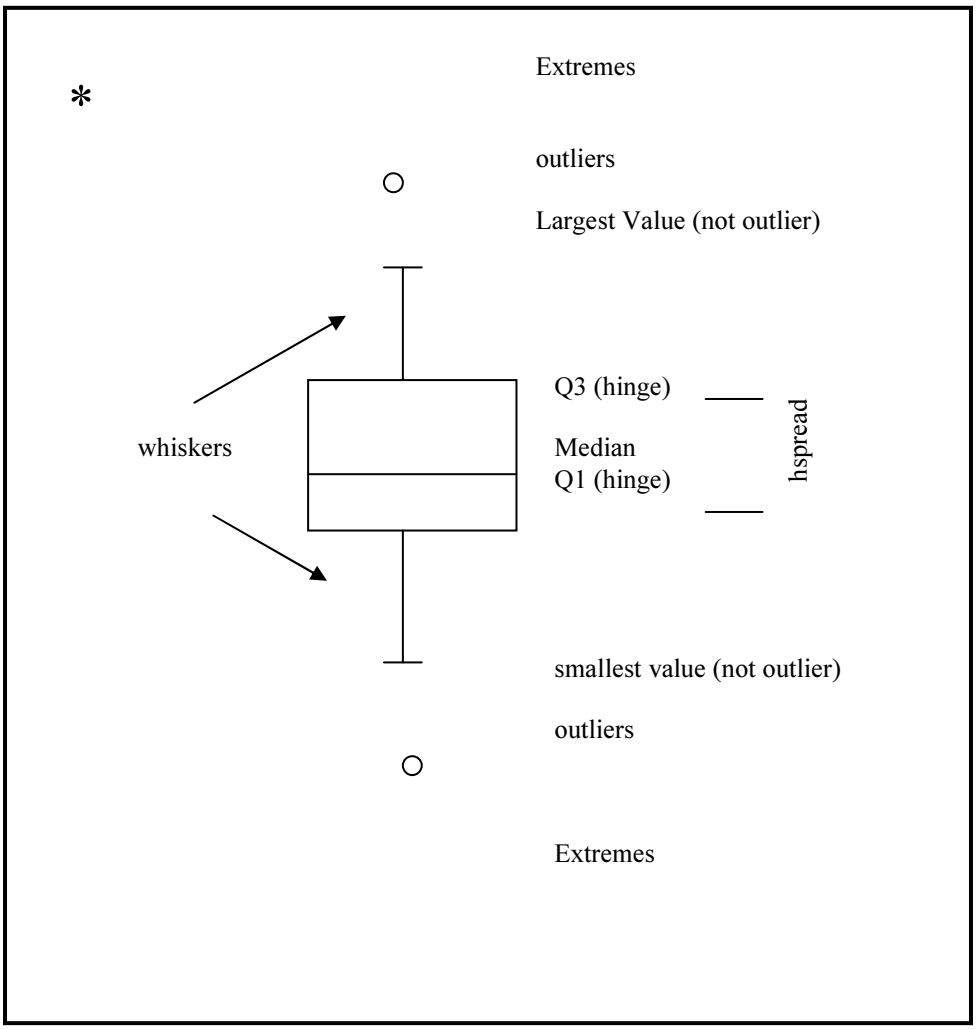
الربيع الثالث Q3 : وهي القيمة التي تسبقها 75% من مشاهدات المتغير المعني عند ترتيبها تصاعدياً .

أن Q1 و Q3 تمثلان حافتي الصندوق ويطلق عليها hinges . أما طول الصندوق فهو Q3-Q1 ويطلق عليه hspread ويعرف بالمدى الربيعي Inter Quartile Range . أما الخط الوسطي داخل الصندوق فهو الوسيط Median

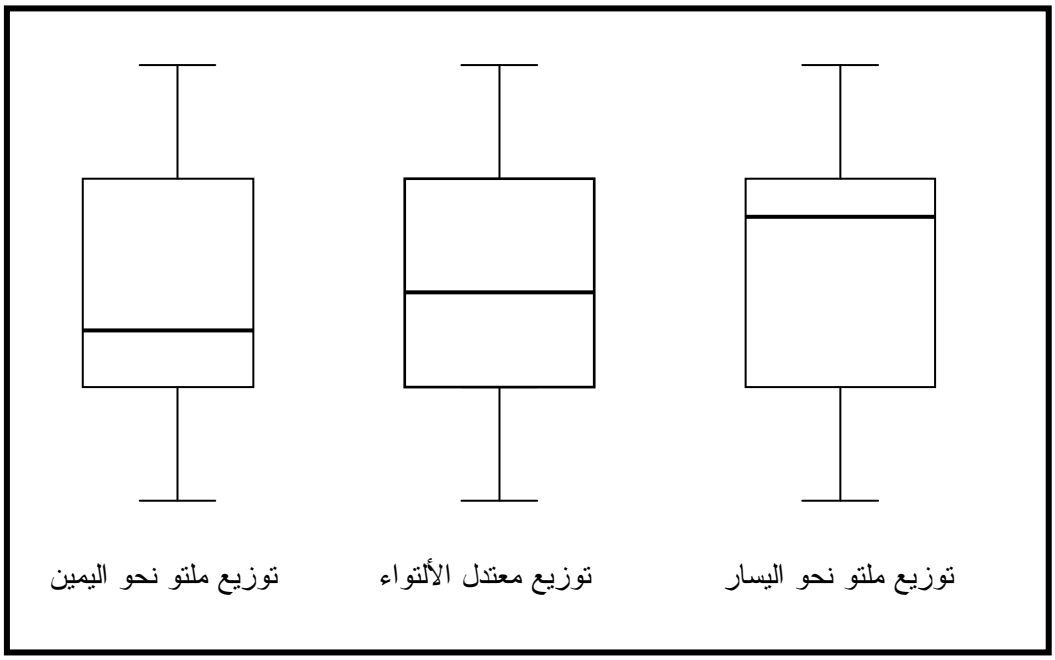
ب. الأستطالات خارج الصندوق Whiskers :وتمتد من حافتي الصندوق الى أعلى واقل قيمة غير شاذة .

ج. القيم الشاذة outliers والقيم المتطرفةExtremes :

حيث أن القيم الشاذة تبعد عن حافتي الصندوق أكثر من 1.5 من طول الصندوق أما القيم المتطرفة فتبعد عن حافتي الصندوق أكثر من 3 من طول الصندوق وكما في الشكل التالي :



أن هذا المخطط يعطي فكرة عن توزيع المشاهدات (الالتواءSkewness) فاذا لم يكن الوسيط في منتصف الصندوق فأن التوزيع ملتو أما إذا كان الوسيط أقرب الى الربيع الأول فأن التوزيع ملتو الى اليمين (موجب الالتواء) وإذا كان الوسيط أقرب الى الربع الثالث فأن التوزيع ملتو الى اليسار (سالب الالتواء) كما في الشكل التالي كما أن Whiskers تعطي فكرة عن طول ذيل التوزيع .



ويتضمن مخطط Boxpolts الخيارات التالية :

Factor Levels together : يتم عرض المخطط للمجاميع الجزئية لكل متغير معتمد .

Dependent together: يتم عرض المخطط للمتغيرات المعتمدة داخل كل مجموعة جزئية .

None: عدم عرض مخطط Boxplot .

في هذا المثال يمكن تأشير أي من الخيارين الأول أو الثاني وذلك لعدم وجود متغير تجزئة Factor Variable .

2.المخططات الوصفية Descriptive

يمكن عرض المخططات التالية :

أ-مخطط Stem-and-leaf : في هذا المخطط تتم قسمة أي رقم الى جزأين الأول Stem (الجذع) والثاني Leaf (الورقة) ويمثل Stem الجزء الأيسر وleaf الجزء الأيمن . فاذا كانت لدينا القيم التالية 5،7،12،15،16،20،21،23،30 فأننا نقسمها الى جزأين الأول Stem الذي يمثل خانة العشرات والثاني Leaf الذي يمثل خانة الآحاد وكأن المتغير قسم الى فئات طول كل منها 10 درجات ويلاحظ أن هذا المخطط يشبه مخطط histogram والفرق بينهما أن التكرارات في Histogram تمثل بأعمدة في حين تمثل بالقيم الحقيقية في حالة Stem-and-Leaf ولذلك فأنه يعكس معلومات عن طبيعة القيم الموجودة .

```
VAR1 Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     2.00        0 .  57
     3.00        1 .  256
     3.00        2 .  013
     1.00        3 .  0

 Stem width:     10.00
 Each leaf:       1 case(s)
```

ب- المدرج التكراري Histogram : لعرض المدرج التكراري .

Normality Plots with Tests.3 : عمل مخططات لاختبار التوزيع الطبيعي للبيانات .

spread vs. Level with Levene Test .4 : لاختبار تجانس التباين في تصميم التجارب .علماً أن هذا الاختيار لا يكون فعالاً (نشطاً) ألا عندما يكون هناك متغير تجزئة Factor Variable .(سيتم توضيح هـذا الخيار بالتفصيل في المثال 3) .

- عند نقر زر OK في صندوق حوار Explore تظهر النتائج التالية :

Explore

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
TALL	56	100.0%	0	.0%	56	100.0%

Descriptives

			Statistic	Std. Error
TALL	Mean		68.1607	2.2948
	95% Confidence Interval for Mean	Lower Bound	63.5618	
		Upper Bound	72.7596	
	5% Trimmed Mean		68.5635	
	Median		70.0000	
	Variance		294.901	
	Std. Deviation		17.1727	
	Minimum		32.00	
	Maximum		99.00	
	Range		67.00	
	Interquartile Range		26.2500	
	Skewness		-.314	.319
	Kurtosis		-.639	.628

الخطأالمعياري للأوساط Standard Error

Mean = 68.1607, Std.Deviation = 17.1727, Std. Error = $SD/\sqrt{n} = 17.1727/\sqrt{56} = 2.2948$

حيث أن Std. Error هو الخطأ المعياري (الانحراف المعياري للأوساط الحسابية المحسوبة من العينة) .

تكوين فترة ثقة 95% لمتوسط المجتمع μ : تستخرج كما يلي :

$$\overline{X} \mp t_{0.025,55} * Std.Error$$

حيث أن t تستخرج من جدول توزيع t لاحتمال 0.025 (المساحة الى يمين قيمة المتغير t) ودرجة حريـة n-1=55 أو يمكن استخراجها مباشرة من برنامج SPSS ودون الرجوع الى الجداول مـن خـلال الأمـر Compute → Transform ثم اختيار الدالة IDF.T(p,df) في صندوق حوار Compute Variable حيث أن p تمثل الأحتمال التجميعي 0.975 = 1-0.025 = p و 55 = df وعليه فأن = IDF.T(0.975,55) 2 = $t_{0.025,55}$ وعليه تكون حدود فترة ثقة 95% للمتوسط كما يلي :

Upper Bound = 68.1607+2*2.2948=72.75
Lower Bound = 68.1607-2*2.2948=63.57

وتكتب فترة الثقة على الصورة التالية :

$$Pr(63.57 < \mu < 72.57) = 95\%$$

حيث أن Pr تمثل الاحتمال أي أن احتمال وقوع متوسط المجتمع بين القيمتين 63.57 و72.57 يساوي 95%

ملاحظة :

يمكن إنجاز العديد من الفعاليات على مخرجات برنامج SPSS مثلاً جدول Descriptives أعلاه حيث يمكن تغيير هيئة الجدول ،عدد المراتب العشرية Decimals ، نوع الخط Font وحجمه ولون الكتابة ، الموقع Alignment ، عرض الأعمدة ..... ويتم ذلك بنقر الجدول المطلوب مرتين في SPSS Viewer (شاشة عرض المخرجات) ثم اختيار الأمر Table Properties → Format لأجراء فعاليات على الجدول ككل ولتغيير مواصفات خلية معينة في الجدول يتم بعد الوقوف على الخلية المعنية ثم اختيار الأمر Cell Properties → Format .

<u>الوسط الحسابي المشذب Trimmed Mean</u>

يستخرج بعد ترتيب القيم تصاعدياً ثم حذف 5% من القيم من الأعلى و5% من القيم من الأسفل ثم حساب المتوسط للقيم المتبقية الذي لن يتاثر بالقيم الشاذة .

في هذا المثال كان عدد قيم المتغير tall هو 56 قيمة وأن 5% من القيم يساوي 2.8 أي أنه يجب حذف 2.8 قيمة من الأعلى و 2.8 قيمة من الأسفل أما القيم المتبقية بعد الحذف فهي 50.4 =0.90*56 قيمة حيث تستخرج قيمة الوسط المشذب كما يلي :

$$TrimmedMean=\frac{68.1607*56-99-95-32-33-0.8*93-0.8*35}{50.4}=\frac{3455.5992}{50.4}=68.5635$$

حيث أن 68.1607*56 تمثل مجموع 56 قيمة وأن 99 و95 تمثلان أعلى قيمتين وأن 32 و33 تمثلان أدنى قيمتين وأن 93و35 تمثلان القيمتين العليا والدنيا التالية للقيمتين العليا والدنيا على الترتيب .وقد حصلنا على هذه القيم من مخرجات برنامج SPSS للقيم المتطرفة Extremes كما في الجدول التالي :

Extreme Values

			Case Number	Value
TALL	Highest	1	28	99.00
		2	38	95.00
		3	6	93.00
		4	16	92.00
		5	23	92.00
	Lowest	1	43	32.00
		2	41	33.00
		3	5	35.00
		4	42	37.00
		5	44	41.00

الربيعات والمئينات

تتكون الربيعات Quartiles من ثلاثة قيم Q1,Q2,Q3 وتقسم المجموعة الى أربعة أقسام متساوية (راجع مخطط Boxplot حول تعريف الربيعات)أما المئينات Percentiles فتقسم المجموعة الى مائة قسم متساوي فالمئين الخامس تسبقه 5% من البيانات عند ترتيبها تصاعدياً وتليه 95% من البيانات والمئين العاشر تسبقه 10% من البيانات عند ترتيبها تصاعدياً وتليه 90% من البيانات ... وهكذا .

علماً أن الربيع الأول Q1 يقابل المئين 25 (25th Percentile)

وأن الربيع الثاني(الوسيط) Q2 يقابل المئين 50 (50th Percentile)

وأن الربيع الثالث Q3 يقابل المئين 75 (75th Percentile)

الجدول التالي يمثل مخرجات البرنامج للمئينات :

Percentiles

	Percentiles						
	5	10	25	50	75	90	95
Weighted Average(Definitior TALL	34.7000	43.1000	55.2500	70.0000	81.5000	91.3000	93.3000
Tukey's Hinges TALL			55.5000	70.0000	81.0000		

وعليه فأن Q1= 55.2 , Q2= Median= 70 , Q3= 81.5

Inter quartile Range = 81.5 - 55.2 = 26.25

أن الطريقة المتبعة في حساب المئينات هي طريقة المتوسط الموزون Weighted Average Method فلحساب ترتيب المئين الخامس مثلاً تستعمل الصيغة التالية (n+1)*P = 57*0.05 = 2.85 أي أن ترتيب المئين الخامس هو 2.85 عند ترتيب القيم تصاعدياً .وعليه فأن قيمة المئين الخامس تقع بين القيمة التي ترتيبها 2 وهي 33 والقيمة التي ترتيبها 3 وهي 35 (من جدول Extreme Values) وعليه نحصل على قيمة المئين بأجراء استكمال خطي Interpolation بين القيمتين المستخرجتين وكما يلي :

5th Percentile = 33 * 0.15 + 35 * 0.85 = 34.7

أختبار التوزيع الطبيعي للبيانات من نسبة معامل الألتواء

يمكن أختبار التوزيع الطبيعي من ملاحظة نسبة معامل الالتواء Skewness الى الخطأ المعياري له (جدول Descriptives) وفي هذا المثال كانت النسبة كما يلي 0.98 - = 0.319 /0.314– وبما أن هذه النسبة تقع ضمن المدى (-2,2)فأذن نقبل فرضية العدم القائلة بأن المتغير Tall يتبع التوزيع الطبيعي .أما أذا كانت النسبة اكبر من 2 فهذا يعني أن التوزيع ملتو التواءاً موجباً (الى اليمين) وأذا كانت النسبة أقل من 2– فهذا يعني أن التوزيع ملتو ألتواءاً سالباً (الى اليسار) .

المخرج التالي يبين التقديرات الحصينة للوسط الحسابي M-Estimators والمحسوبة بأربعة طرق ويلاحظ عدم وجود فرق محسوس بين المتوسط المحتسب بموجب هذه الطرق والمتوسط المحتسب بالطريقة الأعتيادية للمتغير tall.

M-Estimators

	Huber's M-Estimator[a]	Tukey's Biweight[b]	Hampel's M-Estimator[c]	Andrews' Wave[d]
TALL	69.1469	69.3859	68.9754	69.3749

a. The weighting constant is 1.339.

b. The weighting constant is 4.685.

c. The weighting constants are 1.700, 3.400, and 8.500

d. The weighting constant is 1.340*pi.

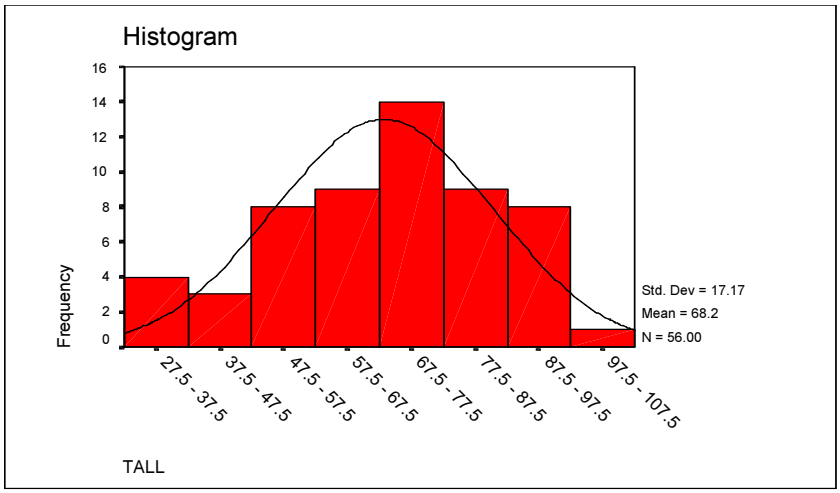
مخطط Stem-and-Leaf

المخرج التالي يمثل مخطط Stem-and-Leaf لمتغير الطول ومنه يستدل أن هذا المتغير يتبـــع التوزيـــع الطبيعي فهو يبدو متماثلاً تقريباً حول المتوسط .

```
TALL Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     4.00        3 .  2357
     5.00        4 .  14789
     7.00        5 .  0123569
     9.00        6 .  001334568
    14.00        7 .  00001122344669
     9.00        8 .  000233458
     8.00        9 .  00122359

 Stem width:     10.00
 Each leaf:       1 case(s)
```

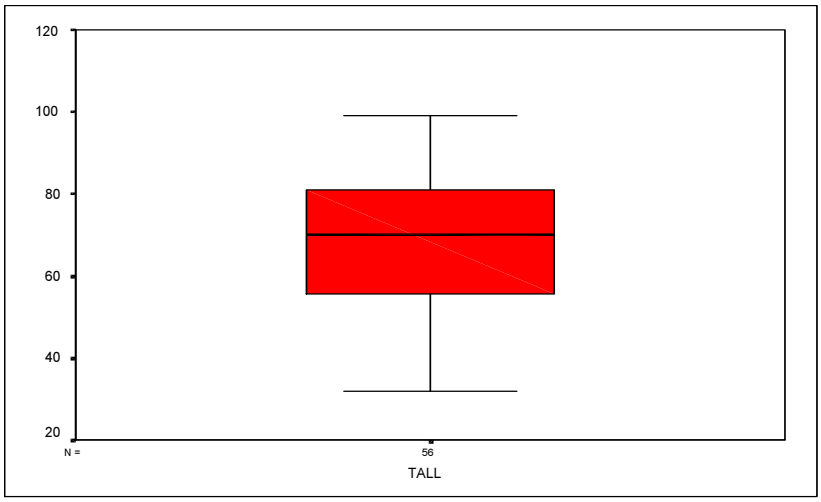
المدرج التكراري Histogram

المخطط التالي يمثل المدرج التكراري لمتغير الأطوال وقد اضيف أليه منحنى التوزيع الطبيعي الذي يبدو متماثلاً حول المتوسط (راجع فصل المخططات لمزيد من التفاصيل حول المدرج التكراري) .



مخطط Boxplots

المخرج التالي يمثل مخطط Boxplots للمتغير Tall ومنه يلاحظ عدم وجود قيم متطرفة أو شاذة وأن موقع الوسيط في وسط الصندوق تقريباً مما يشير الى أن المتغير يتوزع طبيعياً (عدم وجود التواء) .



الخيار Normality Plots with Test : يتيح هذا الخيار إمكانية اختبار التوزيع الطبيعي للبيانات باستخدام اختبار Kolmogrov-Semirnov بالإضافة الى اختبار التوزيع الطبيعي للبيانات باستخدام نوعين من المخططات هما :

- Normal Q-Q Plot
- Detrended Normal Q-Q Plot

1.اختبار Komogrov-Smirnov للتوزيع الطبيعي: ويعد من الاختبارات اللامعلمية للتوزيع الطبيعي Non-Parametric Goodness of Fit Test حيث تختبر فرضية العدم القائلة بأن مشاهدات متغير معين تتبع التوزيع الطبيعي ضد الفرضية البديلة القائلة بأن البيانات لا تتوزع طبيعياً .وقد تم الحصول على المخرج التالي لاختبار التوزيع الطبيعي لمشاهدات المتغير tall .

Tests of Normality

	Kolmogorov-Smirnov[a]		
	Statistic	df	Sig.
TALL	.096	56	.200*

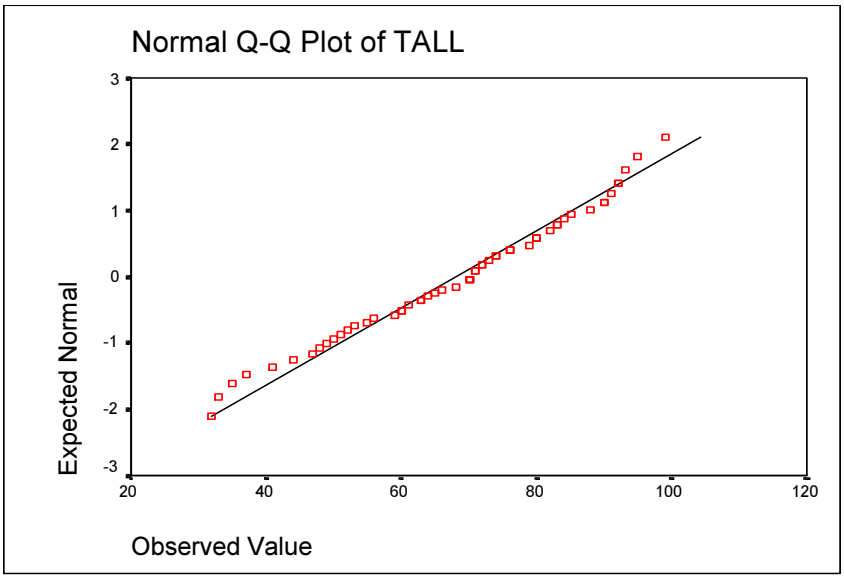
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

حيث تستخدم الإحصائية D في الاختبار $D = \sup_x \left| F_S(x) - F_T(x) \right|$ حيث أن $F_S(x)$ تمثل دالة التوزيع التجميعي للعينة وأن $F_T(x)$ تمثل دالة التوزيع النظري (التوزيع الطبيعي في هذا المثال) والتي تقارن مع القيمة النظرية لـــ D من جدول Kolmogrov بستوى دلالة معين ودرجت حرية n (حجم العينة) . نلاحظ في هذا المثال أن D= .096 وأن P-Value = 0.20>0.05 مما يدعونا الى قبول فرضية العدم بمستوى دلالة 5% أي أن توزيع الأطوال يتبع التوزيع الطبيعي .

2 .مخطط Normal Q-Q Plot

في هذا المخطط يتم رسم كل مشاهدة من البيانات الأصلية على المحور الأفقي مقابل قيم التوزيع الطبيعي القياسي المتوقعة لها Expected Z Score (راجع أمر Rank Cases حول طريقة احتساب القيم القياسية المتوقعة) كما في المخطط التالي للمتغير tall .



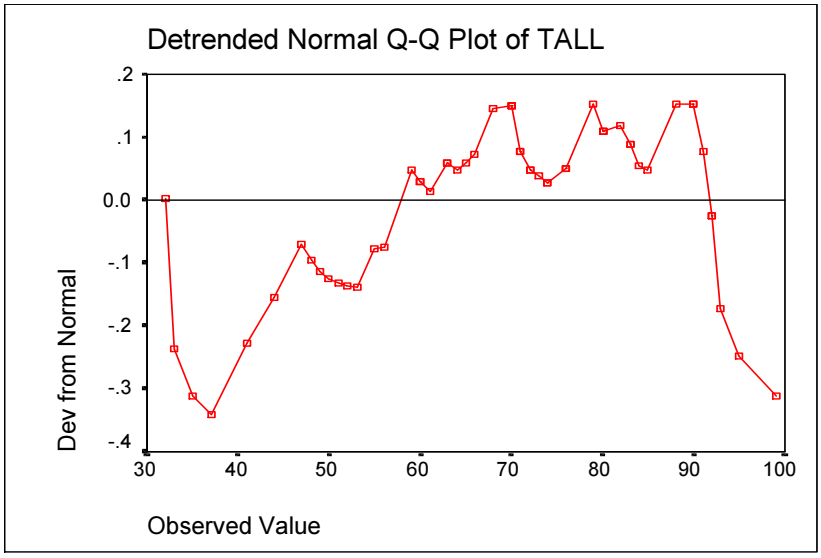
أن كل نقطة على الخط المستقيم في المخطط أعلاه تمثل بواسطة القيم المتوقعة لدرجات التوزيع الطبيعي على المحور العمودي مقابل الدرجات المعيارية للتوزيع الطبيعي للبيانات على المحور الأفقي وعلى الشكل التالي (1.5,1.5),(0,0),(-1.5,-1.5) ...وهكذا . فاذا كانت العينة مسحوبة من مجتمع يتوزع طبيعياً فأن نقاط شكل الانتشار ستقع تقريباً بمحاذاة الخط المستقيم أما اذا كانت تقع بعيدة عن الخط المستقيم فهذا يعني أن البيانات لاتتبع التوزيع الطبيعي .أن انتشار النقاط في الشكل أعلاه بمحاذاة الخط المستقيم يشير الى أن مشاهدات المتغير tall تتبع التوزيع الطبيعي .

3.مخطط Detrended Normal Q-Q Plot

كما ذكرنا بالنسبة للمخطط السابق أنه في حالة عدم انتشار المشاهدات بمحاذاة الخط المستقيم فأننا نستنتج أن البيانات لاتتبع التوزيع الطبيعي ولكن السؤال الذي يطرح الى أي مدى تكون الانحرافات عن الخط المستقيم كافية لنقرر أن المشاهدات لا تتوزع طبيعياً . أن برنامج SPSS يعرض مخطط Detrended Normal Q-Q Plot لهذا الغرض حيث يتم تمثيل المشاهدات الأصلية على المحور الأفقي أما المحور العمودي فيمثل انحرافات القيم المعيارية للمشاهدات عن قيم التوزيع الطبيعي المتوقعة لها.

فإذا كانت معظم نقاط شكل الانتشار (يمكن أن نقرر النسبة 95% أو 90%) تقع ضمن المدى (-2,2) فيمكن أن نقرر أن التوزيع طبيعي للبيانات والعكس صحيح .

في المخطط أدناه نلاحظ أن معظم الانحرافات للمتغير Tall تقع ضمن المدى (-2,2) عدا ستة نقاط تقريباً (أي ما يقارب 90% من الانحرافات) وعليه يمكن أن نقرر أن متغير الأطوال يتوزع توزيعاً طبيعياً .



تم ربط نقاط المخطط من خلال Linear Interpolation في Chart Editor لتبسيط عملية رصد الانحرافات والذي يمكن الدخول فيه بنقر المخطط بزر الماوس الأيسر مرتين .

ملاحظة : في حالة الاستنتاج بأن مشاهدات متغير ما لا تتوزع طبيعياً فأنه يمكن عمل تحويلات للبيانات Transformation لغرض جعل التوزيع طبيعي (نفس الطريقة المتبعة في المثال 3) .

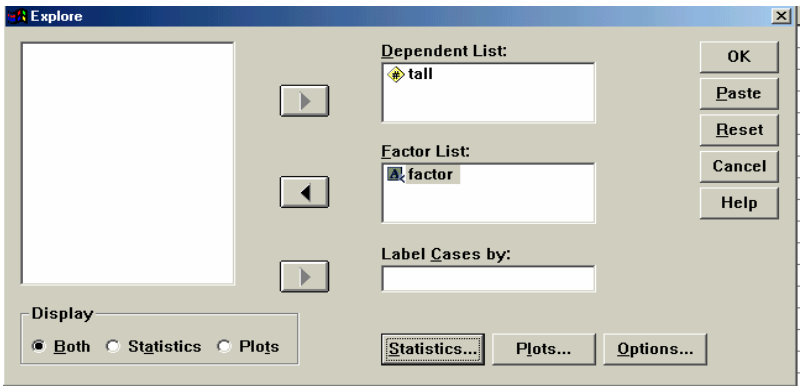
مثال 2

في المثال السابق للمتغير Tall لو اعتبرنا أن 22 حالة الأولى تمثل أطوال الفئة الأولى من البيانات وأن 34حالة الثانية تمثل أطوال الفئة الثانية من البيانات ونرغب في أجراء تحليل إحصائي لكل مجموعة على حدة في هذه الحالة يتوجب إدخال متغير تجزئة (لنسمه Factor) في قائمة Factor List في صندوق حوار Explore الذي يتم ترتيبه بالشكل التالي :

tall	factor
80	A
84	A
71	A
72	A
35	A
93	A
91	A
74	A
60	A
63	A
79	A
80	A
70	A
68	A
90	A
92	A
80	A
70	A
63	A
76	A
48	A
90	A
92	B
85	B
83	B
76	B
61	B
99	B
83	B
88	B
74	B
70	B
65	B
51	B
73	B
71	B
72	B
95	B
82	B
70	B
33	B
37	B
32	B
41	B
44	B
49	B
47	B
50	B
59	B
55	B
53	B
56	B
52	B
64	B
60	B
66	B

أما ترتيب البيانات في Data Editor فيكون كما يلي : **ترتيب البيانات في Data Editor**

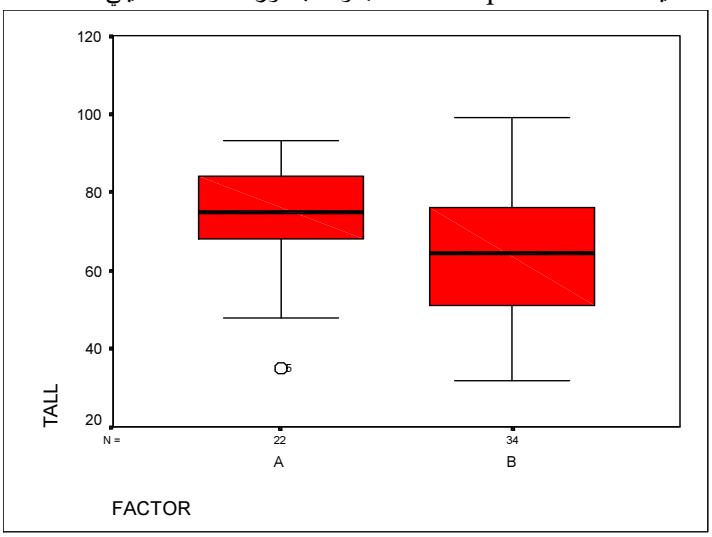
عند تنفيذ أمر Explore فأن المخرجات تعرض لكل من الفئتين A و B بصورة مستقلة مثلاً القيم المتطرفة تعرض كما يلي :



Extreme Values

FACTOR				Case Number	Value
TALL	A	Highest	1	6	93
			2	16	92
			3	7	91
			4	15	90
			5	22	90
		Lowest	1	5	35
			2	21	48
			3	9	60
			4	19	63
			5	10	63
	B	Highest	1	28	99
			2	38	95
			3	23	92
			4	30	88
			5	24	85
		Lowest	1	43	32
			2	41	33
			3	42	37
			4	44	41
			5	45	44

كما ينفذ مخطط Boxplots لكل مجموعة بصورة منفصلة كما يلي :



نلاحظ وجود قيمة شاذة Outlier في الفئةA وهي الحالة رقم 5 وقيمتها 35 حيث أنها تبعد عن الربيع الأدنى Q1=66.75 بمسافة تزيد عن 1.5 وتقل عن Ȯ من طول الصندوق18.75 ويرمز لها بالرمز في مخطط Boxplots وبجانب الرمز يعرض رقم الحالة 5 وهو الحالة الافتراضية في حالة عدم وجود Label للحالة (متغير يتم إدخاله في قائمة Label Cases by في صندوق حوار Explore) .

(6 – 2) *اختبار تجانس التباين* *Test of Homogeneity of Variances*

أن أحد الفرضيات التي ينبغي توفرها عند أجراء تحليل التباين ANOVA هو تجانس أو تساوي تباين المعالجات (المعاملات) ففي حالة عدم تحقق هذا الشرط فأن ذلك سوف يؤدي للتوصل الى قرارات خاطئة عند اختبار الفرضية الخاصة بتساوي تأثير المعاملات (أو تساوي أوساط المعالجات) .

يوفر برنامج SPSS اختبارا لتجانس التباين هو Levene Test حيث تختبر فرضية العدم القائلة بتجانس تباين المعالجات ضد الفرضية البديلة القائلة بعدم تجانس التباين .وغالباً عندما تكون التباينات غير متساوية فأن ذلك يؤدي الى عدم تحقق شرط التوزيع الطبيعي داخل المعالجات وهذا الشرط ضروري لاختبار الفرضيات حيث نفترض أن الخطأ العشوائي له توزيع طبيعي $\varepsilon \sim N(0,\sigma^2)$. وفي هذه الحالة يتوجب أجراء تحويلات على البيانات Transformation لضمان تجانس التباين وأن القيام بهذه التحويلات يجعل التوزيع يقترب من الطبيعي داخل كل معالجة أيضاً .

لقد أقترح Box & Cox أداة تشخيصية للوصول الى قرار فيما أذا كان سينجم عن التحويلات تجانس التباينات . تعتمد الأداة على تحديد وجود علاقة بين أوساط المعالجات وانحرافاتها المعيارية (يجب أن تكون الأوساط مستقلة عن الانحراف المعياري لضمان تجانس التباين) .حيث يتم حساب الميل من انحدار لوغاريتم الانحرافات عن المتوسط على لوغاريتم أوساط المعالجات لتكوين Power Transformation للبيانات حسب الصيغة التالية y^{1-b} حيث أن y تمثل البيانات الأصلية وأن 1-b تمثل الأسPower وهذه بعض التحويلات شائعة الاستعمال :

Transformation	**Power**	**Slop b**
Square	2	-1
None	1	0
Square Root	1\2	1\2
Logarithm	0	1
Reciprocal of Square Root	-1\2	3\2
Reciprocal	-1	2

يلاحظ أن الأس يتتراوح من –1 الى 2 فاذا كانت قيمة الميل تساوي 1 فهذا يعني أن أوساط المعالجات مستقلة عن انحرافاتها المعيارية وبالتالي لا تحتاج البيانات الى تحويل نظراً لتجانس التباين أي أن الأس يساوي 0 واذا كانت معلمة الميل مساوية الى 0.671 فأن الأس 1-b سيكون 0.329 فأننا سنختار تحويلة الجذر التربيعي للبيانات الأصلية لأن 0.329 قريبة من 1/2 أما في حالة أن الآس يساوي 1 فأننا نختار تحويلة اللوغاريتم علماً أنه لا يوجد فرق بين استعمال اللوغاريتم الطبيعي أو اللوغاريتم للأساس 10 .

الخيار Spread vs. Level with Leven Test

يتيح هذا الخيار(الذي يظهر في صندوق حوارPlots كما ورد سابقاً) إمكانية اختبار تجانس تباين المعاملات (المعالجات) باستخدام Levene Statistics إضافة الى اختبار تجانس التباين بواسطة المخطط Spread –Versus Level Plot (يمثل نفس أسلوبBox & Cox مع تحوير بسيط) حيث يتم تمثيل لوغاريتم الوسيط للمعالجات (المستوى Level)على المحور الأفقي لهذا المخطط وتمثيل لوغاريتم المدى الربيعي Inter Quartile Range على المحور العمودي (التباعد Spread) ففي حالة عدم وجود علاقة بين المستوى والتباعد فأن نقاط هذا المخطط تتجمع في خط أفقي (Slop=0) الذي يشير الى تجانس تباين المعالجات وعكس ذلك فأن تجمع النقاط حول خط اتجاه عامTrend يعني وجود علاقة وبالتالي عدم تجانس تباين المعالجات مما يستلزم أجراء تحويل على البيانات حسب جدول التحويلات المذكور أنفاً . علماً أنه بالإمكان جعل البرنامج يقوم برسم المخطط بعد التحويل لغرض التحقق من أن عملية التحويل قد أثمرت عن تجانس التباين (أي أن الميل سيكون قريب من الصفر) .

مثال 3: الجدول التالي يمثل نتائج تجربة عامليه استخدمت فيها أربعة معالجات A,B,C,D مع 12 تكرار لكل معالجة :

Treatments / Obs.	A	B	C	D
1	895	1520	43300	11000
2	540	1610	32800	8600
3	1020	1900	28800	8260
4	470	1350	34600	9830
5	428	980	27800	7600
6	620	1710	32800	9650
7	760	1930	28100	8900

8	537	1960	18900	6060
9	845	1840	31400	10200
10	1050	2410	39500	15500
11	387	1520	29000	9250
12	497	1685	22300	7900
Mean	670.75	1701.25	30775	9395.83
Std. Deviation	233.92	356.54	6688.68	2326.04

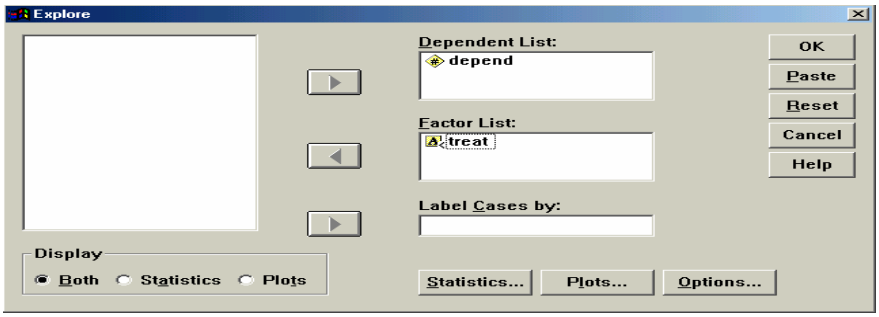
المصدر: د.محمود المشهداني وكمال علوان المشهداني ، تصميم وتحليل التجارب ، جامعة بغداد ، ص49 .

المطلوب اختبار تجانس تباين المعالجات مع تحديد التحويل المناسب للبيانات في حالة عدم التجانس.

يتم إدخال البيانات الى نافذة Data Editor لبرنامج SPSS كما هو في (الصفحة التاليـــة) حيـــث أن المتغير depend يمثل المشاهدات (المتغير المعتمد) وأن treat هو متغير تجزئة للمتغيـــر المعتمـــد ويمثـــل المعالجات A,B,C,D. لاختبار تجانس التباينات نتبع الخطوات التالية :

- من شريط القوائم نختار

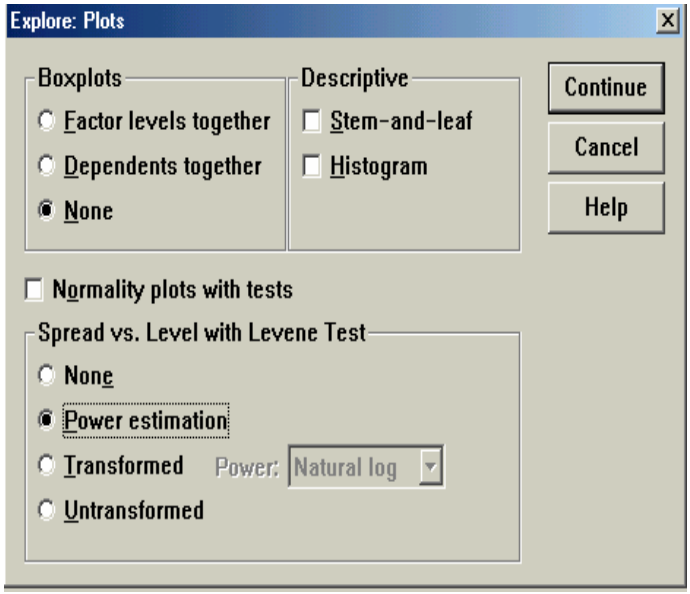
Analyze→Descriptive Statistics→ Explore حيث يظهر صندوق حوار Explore الذي نقوم بترتيبه على الشكل التالي :



- أنقر الزر Plots في صندوق حوار Explore وفي خانة spread vs. Level with levene test (الذي يكون فعالاً نظراً لوجود متغير تجزئةTreat) ثم نتبع الخطوتين الرئيسيتين التاليتين .

<u>الخطوة الأولى :</u> أختر Power Estimation حيث يظهر صندوق حوار Plots بعد ترتيبه كما يلي :

أن الهدف من هذه الخطوة هو التالي :



ترتيب البيانات في Data Editor للمثال 3

Treat	Depend
A	895
A	540
A	1020
A	470
A	428
A	620
A	760
A	537
A	845
A	1050
A	387
A	497
B	1520
B	1610
B	1900
B	1350
B	980
B	1710
B	1930
B	1960
B	1840
B	2410
B	1520
B	1685
C	43300
C	32800
C	28800
C	34600
C	27800
C	32800
C	28100
C	18900
C	31400
C	39500
C	29000
C	22300
D	4000
D	8600
D	8260
D	9830
D	7600
D	9650
D	8900
D	6060
D	10200
D	15500
D	9250
D	7900

1. الاختبار الإحصائي لتجانس التباين بواسطة إحصائية Levene .
2. اختبار العلاقة بين المستوى والانتشار من خلال الرسم البياني مع تحديد نوع التحويل الملائم للبيانات Power Transformation في حالة وجود علاقة (أي التباين غير متجانس) .

➢ أنقر زر Continue ثم .زر OK فنحصل على النتائج التالية :

1. الأختبار الأحصائي لتجانس التباين بواسطة أحصائية Levene

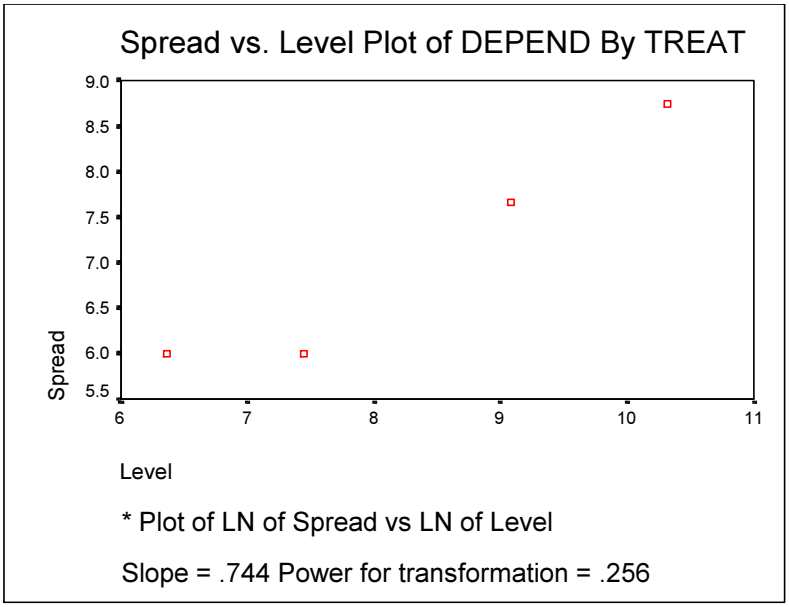
المخرج أدناه يبين نتيجة اختبار فرضية العدم (تجانس التباين) ضد الفرضية البديلة(عدم تجانس التباين) باستخدام إحصائية Levene حيث أن قيمة هذه الإحصائية أما تكون مبنية على المتوسط Mean أو الوسيط Medianأو الوسط المشذب فبالنسبة للإحصائية المبنية على المتوسط فقد بلغت 10.783 وبمعنوية عالية (p-value ≈ 0.05>.000) وعليه ترفض فرضية العدم بمستوى دلالة 5% وتؤخذ الفرضية البديلة القائلة بعدم تجانس التباينات وهي نفس النتيجة التي نتوصل إليها عند اعتماد قيمة الإحصائية المبنية على الوسيط والوسط المشذب .

Test of Homogeneity of Variance

	Levene Statistic	df1	df2	Sig.
Mean	10.783	3	44	.000
Median	10.621	3	44	.000
Median and ted df	10.621	3	15.853	.000
trimmed mean	10.779	3	44	.000

2.أختبار تجانس التباين من خلال مخطط Spread vs. Level Plot

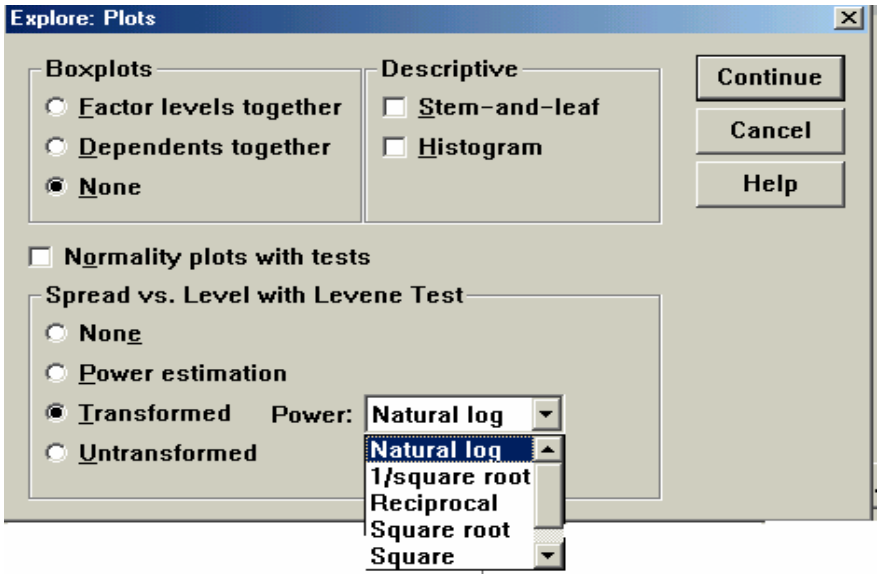
يمكن من خلال هذا المخطط تشخيص مشكلة عدم تجانس التباين بالإضافة الى تقدير الأس المناسب لتحويل البيانات في حال وجود المشكلة (الخيارPower estimation) كما في المخطط التالي :



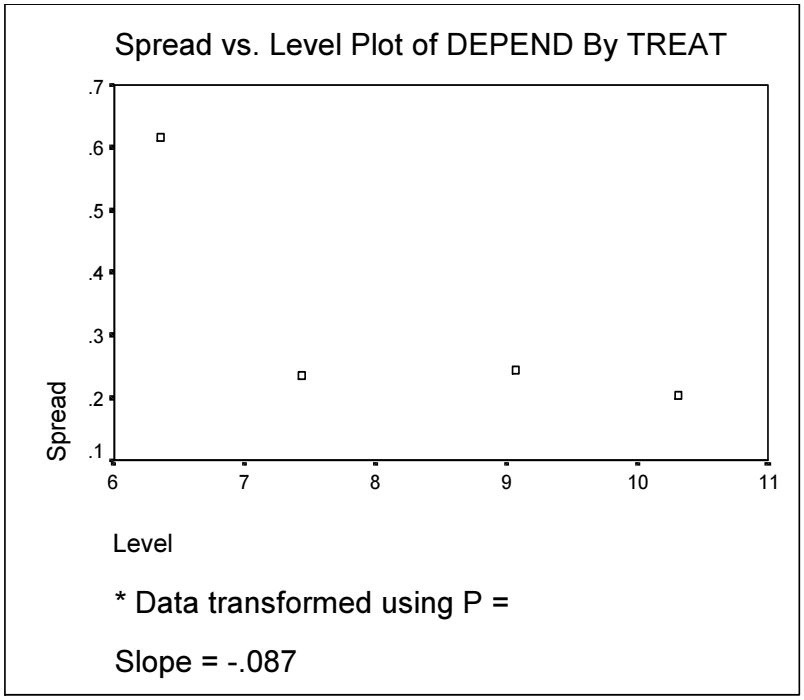
حيث يتضح من خلال الاتجاه العام لشكل الانتشار وجود علاقة بين لوغاريتم الوسيط للمعالجات A,B,C,D ولوغاريتم المدى الربيعي لها وأن الميل b=0.744 وأن الأس اللازم لتحويل البيانات هو 1-b = 0.256 وبالرجوع الى جدول التحويلات الشائعة نجد أن قيمة الأس 0.256 تقع بين الأس 0 (تحويل لوغاريتمي) وبين الأس 1/2(تحويلة الجذر التربيعي) وفي هذه الحالة يمكن تجربة كلا النوعين في التحويل واعتماد التحويل الذي يعطي نتائج افضل (ميل قريب من الصفر أو الأس يساوي واحد) ان تنفيذ ذلك يتم مباشرة بواسطة الخيار Transformed في صندوق حوار Plots (الخطوة التالية) .

الخطوة الثانية: تحويل البيانات بواسطة الخيار Transformed

نستنتج من الخطوة الأولى أن تباين المعالجات غير متجانس وعلية نقوم بتحويل البيانات لجعل التباين متجانسا ويتم ذلك بالعودة الى صندوق حوار Explore:Plot ثم نقر Transformed فتصبح دالة التحويلPower فعالة ثم أختر التحويلة Natural Log وكما في الشكل التالي :

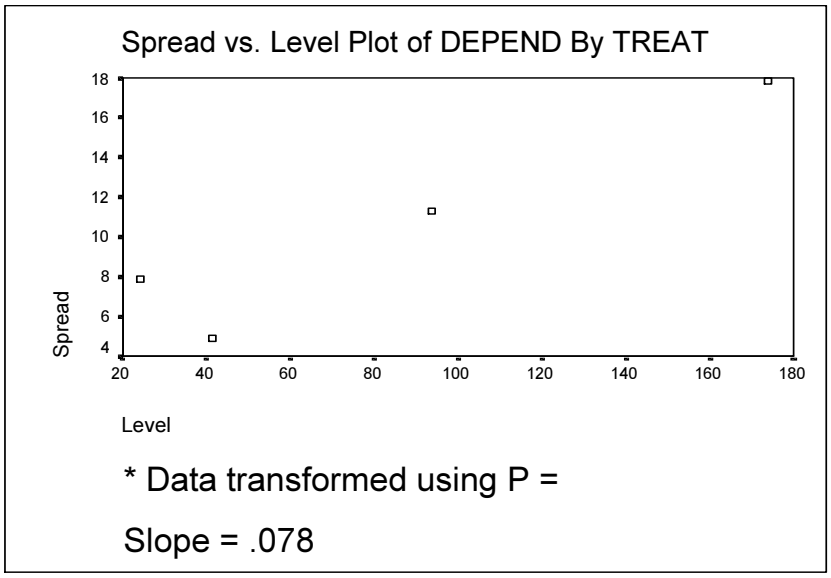


عند نقر زر Continue ثم زر OK يظهر مخطط البيانات المحولة (اللوغاريتمات كما يلي) :



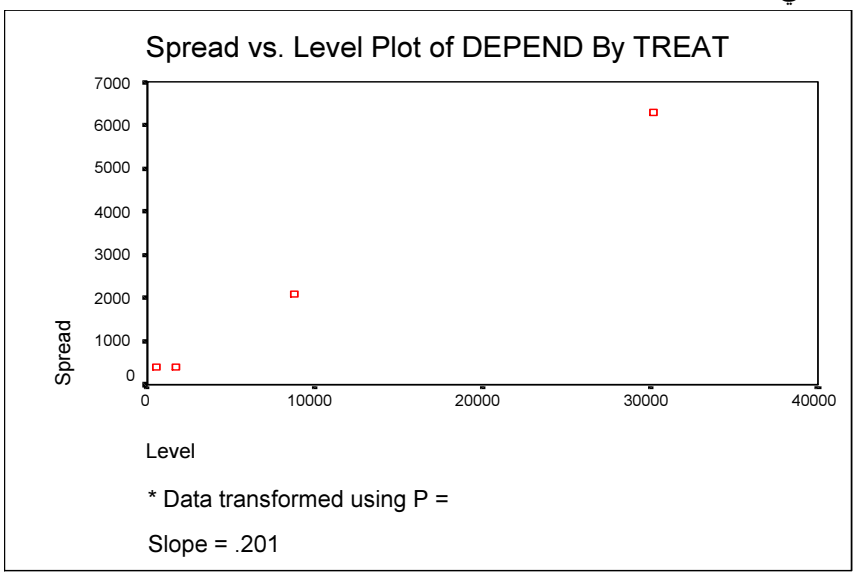
حيث يتضح عدم وجود اتجاه عام للعلاقة بين الوسيط والانتشار للبيانات المحولة وأنما تتركز النقـــاط حول خط أفقي وهذا يشير الى انتفاء مشـــكلة عـــدم تجـــانس التبـــاين للبيانـــات المحولـــة كمـــا أن قيمـــة الميل b=-0.087 وعليه فأن الأس سيكون 1.087 = 1-b وهو قريب من الواحـــد أي نســـتطيع القـــول أن البيانات المحولة لاتعاني من مشكلة عدم تجانس التباين .

أما في حالة اختيار تحويلة الجذر التربيعي نحصل على المخطط التالي :



حيث أن الميل يساوي 0.078 وعليه فأن الأس سيكون 0.922 أي أنه قريب من الواحد وهذا يعنـــي تجانس التباين للبيانات المحولة بالجذر التربيعي .وقد كانت النتائج متقاربة لكلا النوعين من التحويلات وعليه يمكن اعتماد أي منهما .

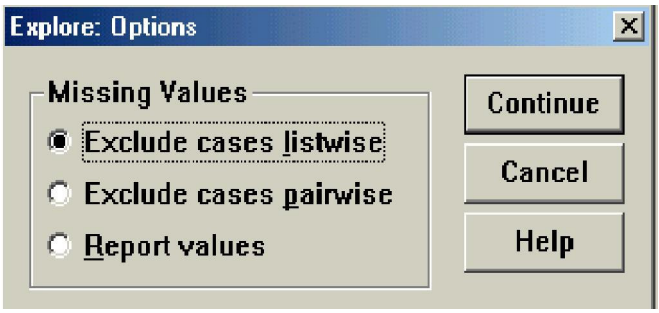
أما بالنسبة للخيار Untransformed في صندوق حوار Plots فأنه ينفذ المخطط بالاعتماد على القيم الأصلية(بدون تحويل البيانات) وذلك برسم العلاقة بين الوسيط والمدى الربيعي (IQR) بدون لوغاريتمات كما في الشكل التالي :



ويلاحظ وجود اتجاه عام للعلاقة بين الوسيط والمدى الربيعي علماً أن أن هذا المخطط لا يستعمل في تحديد نوع التحويل لأنه لا يمثل العلاقة بين لوغاريتم الوسيط ولوغاريتم المدى الربيعي للمعالجات.

(6 –3) التعامل مع القيم المفقودة

يمكن تحديد كيفية استبعاد الحالات التي تحوي قيماً مفقودة بنقر الزر Options في صندوق حوار Explore فيظهر صندوق حوار Options التالي :



حيث أن :

Exclude Cases Listwise: يستعمل لاستبعاد الحالة بكاملها (لكافة المتغيرات) أذا كانت تحتوي على قيمة مفقودة لأي من المتغير المعتمد Dependent أو متغير التجزئة Factor وهذا الخيار هو الحالة الافتراضية .

Exclude Cases Pairwise : يتم استبعاد الحالات التي لها قيم مفقودة لأي من أو كلا المتغيرات التي تستعمل في خساب أحصائية معينة .

Report Values : تعامل القيم المفقودة لمتغير التجزئة كفئة مستقلة ويتم عرض النتائج لهذه الفئة الإضافية .

لتوضيح ما سبق نفترض لدينا البيانات التالية التي أدخلت في شاشة Data Editor :

dep1	dep2	fac
1	10	1
2	11	1
.	.	1
4	13	2
5	14	.
6	.	2
7	8	2

- في البداية ننفذ الأمر Explore → Descriptive Statistics → Analyze
- في صندوق حوار Explore نتبع مايلي :
  - أنقر المتغيرين dep1 و dep2 وأدخلهما الى خانة Dependent List .
  - أنقر المتغير fac وأدخله الى خانة Factor List .
  - أنقر زر Options وأختر Exclude Cases list Wise .
- أنقر زر OK فتظهر النتيجة التالية :

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	FAC	N	Percent	N	Percent	N	Percent
DEP1	1	2	66.7%	1	33.3%	3	100.0%
	2	2	66.7%	1	33.3%	3	100.0%
DEP2	1	2	66.7%	1	33.3%	3	100.0%
	2	2	66.7%	1	33.3%	3	100.0%

Descriptives

	FAC		Statistic	Std. Error
DEP1	1	Mean	1.50	.50
			.	.
	2	Mean	5.50	1.50
			.	.
DEP2	1	Mean	10.50	.50
			.	.
	2	Mean	10.50	2.50
			.	.

لاحظ كيفية حساب المتوسطات (على سبيل المثال) في حالة Exclude Cases Listwise حيث تم استبعاد أي حالة تحتوي على قيمة مفقودة لأي من المتغيرات في قائمة Dependent List أو Factor List .فمثلاً أحتسب المتوسط للمتغير dep1 للقيمة 1 لمتغير التجزئة fac بالاعتماد على الحالتين 1 و2 وعلى الحالات 4 و7 للقيمة 2 لمتغير التجزئة أما بالنسبة لمتوسط dep2 للقيمة 1من متغير التجزئة فقد أحتسب بالاعتماد على الحالتين 1 و2 ... وهكذا .

أما في حالة الخيار Exclude Cases Pairwise فأن المتوسطات تحتسب كما يلي :

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	FAC	N	Percent	N	Percent	N	Percent
DEP1	1	2	66.7%	1	33.3%	3	100.0%
	2	3	100.0%	0	.0%	3	100.0%
DEP2	1	2	66.7%	1	33.3%	3	100.0%
	2	2	66.7%	1	33.3%	3	100.0%

Descriptives

	FAC		Statistic	Std. Error
DEP1	1	Mean	1.50	.50 . .
	2	Mean	5.67	.88 .
DEP2	1	Mean	10.50	.50 . .
	2	Mean	10.50	2.50 . .

لاحظ أن متوسط المتغير dep1 لقيمة متغير التجزئة fac قد أحتسب من الحالات 1و2 أمـــا بالنســـبة لمتوسط dep1 للقيمة 2 لمتغير التجزئة فقد أحتسب من الحالات 4 و6 و7 ونفس الشيء للمتغير dep2 .

الآن ماذا يحصل لو حذفنا المتغير fac من قائمة Factor List في صندوق حـــوار Explore مـــع الإبقاء على الخيار Exclude Cases Pairwise ؟

في هذه الحالة بتم أستبعد الحالات الحاوية على قيمة مفقودة لكل متغير وبصورة مستقلة حيث نحصل على القيم التالية للمتوسطات الحسابية .

Descriptives

		Statistic	Std. Error
DEP1	Mean	4.17	.95
DEP2	Mean	11.20	1.07

فالقيمة 4.17 هي متوسط الحالات 1و2 و4 و 5 و6 و7 للمتغيـــر dep1 والقيمـــة 11.02 هـــي متوسط الحالات 1و 2 و4 و 5 و7 للمتغير dep2 .

سنحاول الأن تجربة الحالة الأخيرة للقيم المفقودة Report Values .أرجع المتغير fac الـــى خانـــة Factor List في صندوق حوار Explore مع تأشـــير الخيـــار Report Values فـــي صـــندوق حـــوار Options . عند نقر زر OK في صندوق حوار Explore نحصل على النتائج التالية :

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	FAC	N	Percent	N	Percent	N	Percent
DEP1	. (Missing)	1	100.0%	0	.0%	1	100.0%
	1	2	66.7%	1	33.3%	3	100.0%
	2	2	66.7%	1	33.3%	3	100.0%
DEP2	. (Missing)	1	100.0%	0	.0%	1	100.0%
	1	2	66.7%	1	33.3%	3	100.0%
	2	2	66.7%	1	33.3%	3	100.0%

لاحظ أن متغير التجزئة Fac له أصلاً فئتين هما 1 و 2 وفي حالة اختيـــار Report Values فقـــد أضيفت فئة ثالثة بأسم Missing للقيم المفقودة .يتم احتساب المتوسطات (على سبيل المثال) للخيار Report Values كما يلي :

Descriptives

	FAC		Statistic	Std. Error
DEP 1	1	Mean	1.50	.50 . .
	2	Mean	5.50	1.50 . .
DEP 2	1	Mean	10.50	.50 . .
	2	Mean	10.50	2.50 . .

يلاحظ أن أستبعد القيم المفقودة و حساب المتوسط قد تم بنفس طريقة Exclude Listwise حيث حصلنا على نفس النتائج تماماً .

الفصل السابع

جداول التقاطع

Crosstabs

يستعمل الأمر Crosstabs في عمل جداول الاقتران Tables 2 × 2 (تتكون من صفين وعمـــودين) والجداول المتعددة Multiway Tables (تتكون من أكثر من صفين أو عمودين) مـــع احتســـاب مؤشـــرات الارتباط لهذه الجداول والاختبارات الإحصائية المرافقة .أن لهذا الأمر فائدة كبيرة في البحوث التطبيقية فمثلاً يسهل عملية إعداد جداول الاقتران والجداول المتعددة لأسئلة الاستبيانات الإحصائية .

مثال 1: الجدول التالي يمثل عينة حجمها 22 مشاهدة للمرضى الذين خضعوا لعلاج a والذين لم يخضـــعوا لعلاج b ويعبر عن ذلك المتغير treat أما المتغير recover فيمثل الشفاء من المرض يرمز للشفاء a1 وعدم الشفاء b2 أما المتغير gender فيمثل جنس المريض m للذكور و f للإناث وقد أدخلت البيانات في شاشـــة Data Editor كما يلي :

treat	recover	gender
a	a1	m
b	a1	m
b	b1	m
b	b1	m
a	a1	m
a	a1	m
b	b1	m
a	b1	m
b	b1	m
a	a1	m
a	a1	m
b	a1	m
a	a1	m
b	b1	m
a	a1	f
b	b1	f
b	b1	f
b	b1	f
a	a1	f
b	a1	f
a	b1	f
b	b1	f

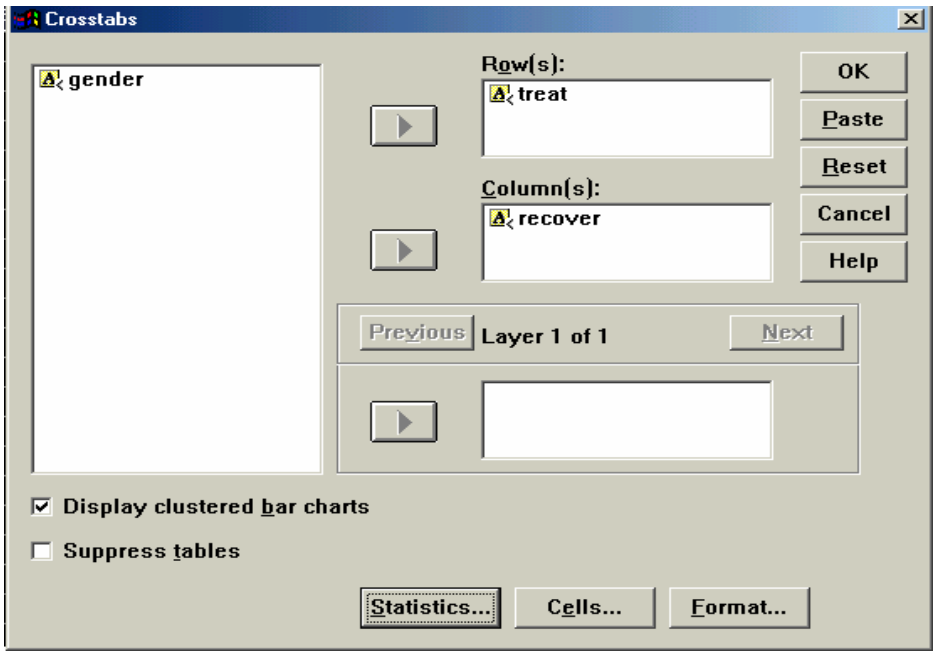
يطلب مايلي :

1. تكوين جدول اقتران للعلاقة بين المتغيرين treat و recover مع اختبار الاستقلالية بين المتغيـــرين المذكورين أي اختبار وجود علاقة معنوية بين تعاطي العلاج والشفاء من المرض .
2. تكوين جدول اقتران للعلاقة بين المتغيرين treat و recover مع اختبار الاستقلالية بين المتغيـــرين المذكورين حسب متغير الجنس gender .

1. لتنفيذ المطلوب الأول نتبع الخطوات التالية :

➢ من شريط القوائم نختار Crosstabs ← Descriptive Statistics ← Analyze

فيظهر صندوق حوار Crosstabs الذي نرتبه كما يلي :



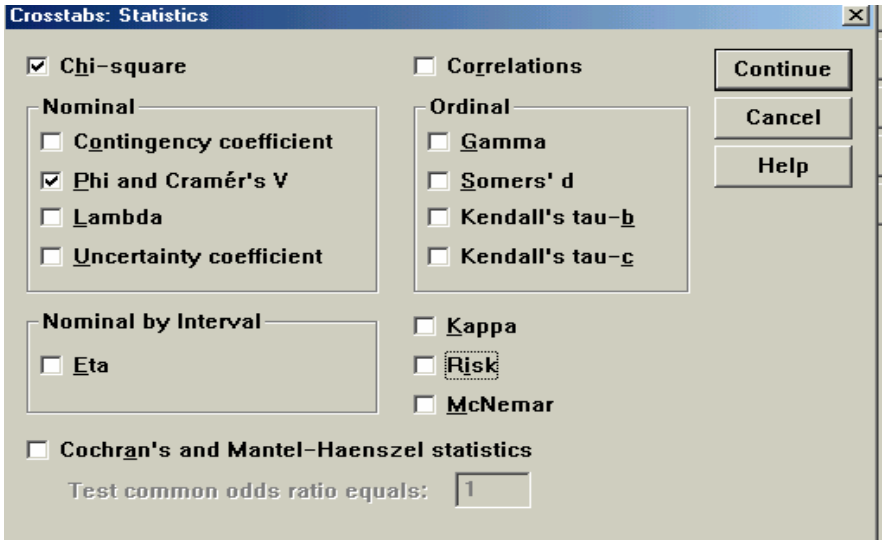
حيث أن :

Row(s) : يتضمن المتغير الذي نرغب في جعل فئاته صفوفاً في الجدول .

Column(s) : : يتضمن المتغير الذي نرغب في جعل فئاته أعمدة في الجدول .

Display Clustered bar Charts : لعرض الأشرطة البيانية عند التأشير .

Supress tables: لاخفاء جداول الاقتران والجداول المتعددة عند التأشير .



➢ عند نقر زر Statistics يظهر صندوق الحوار Statistics الذي نرتبه كما يلي :

حيث أن :

Chi-Sqare : لحساب إحصائية chi-Square وملحقاتها لاختبار استقلالية الصفوف عن الأعمدة .

Correlation : لحساب معاملي الارتباط لـــ Spearman و Pearson في أن معاً علماً أنه يجب أن تكون

قيم المتغيرات الداخلة عدديةNumeric فيتم حساب معامل ارتباط Spearman باعتبار أن كلا من الصفوف والأعمدة عبارة عن قيم ترتيبية Ordered Values (الخيارOrdinal في العمود Measure في شاشة Variable view) مثلاً الحالة العلمية (أمي ،يقرأ ، يقرأ ويكتب ...) كما يحتسب معامل ارتباط Pearson باعتبار أن كلا من الصفوف والأعمدة متغيرات كميةQuantitative Variables ويعبر عنها بالخيارScale في عمود Measure في شاشة Variable View ويشار لها أيضاً Interval.

Nominal :تتيح هذه القائمة حساب أربعة معاملات للاقتران (ارتباط) في حالة كون كلاً من الصفوف والأعمدة عوامل غير كمية بالإضافة الى عدم إمكانية ترتيب البيانات مثلاً متغير الجنس (ذكور،إناث) أو متغير المحافظة (بغداد،موصل ،بصرة) .

Ordinal : تتيح هذه القائمة حساب أربعة معاملات للاقتران في حالة كون كلا الصفوف والأعمدة متغيرات ترتيبية .

Nominal by Interval : لحساب إحصائية Eta للاقتران بين متغيرين أحدهما المعتمد يقاس ضمن فترة Interval Scale مثل متغير الدخل Income والمتغير الأخر مستقل له عدد محدود من الفئات Categories مثل الجنس gender .

يمكنك التعرف على بقية إحصائيات الصندوق بنقرها بزر الماوس الأيسر لاظهار التعليق الخاص بها .

نظراً لكون الصفوف والأعمدة في هذا المثال متغيرات اسميةNominal (لها الخيار Nominal في عمودMeasure في شاشة Variable View) فقد اكتفينا بحساب إحصائيةChi- Square لاختبار الاستقلالية مع مقياس Phi and Gramer's v للاقتران للمتغيرات الاسمية Nominal Variables.

- عند نقر زر Continue في صندوق حوار Statistics ثم ok في صندوق حوار Crosstabs نحصل على المخرجات التالية :

TREAT * RECOVER Crosstabulation

Count

		RECOVER		Total
		a1	b1	
TREAT	a	8	2	10
	b	3	9	12
Total		11	11	22

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	6.600[b]	1	.010		
Continuity Correction[a]	4.583	1	.032		
Likelihood Ratio	6.994	1	.008		
Fisher's Exact Test				.030	.015
N of Valid Cases	22				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.00.

يتم اختبار فرضية العدم القائلة باستقلالية الصفوف عن الأعمدة أي استقلالية العلاج عن الشفاء ضد الفرضية البديلة القائلة بوجود علاقة بين العلاج و الشفاء باستخدام إحصائية Chi-Square التي لها درجة حرية (r-1)(c-1) حيث ان r يمثل عدد الصفوف و c عدد الأعمدة و حسب الصيغة التالية :

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

حيث أن O_i يمثل التكرار المشاهد و E_i يمثل التكرار المتوقع وقد كانت قيمة الإحصائية 6.6 وأن قيمة p-value= 0.010 < 0.05 تدعونا الى رفض فرضية العدم بمستوى دلالة 5% أي توجد علاقة بين العلاج والشفاء من المرض .

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.548	.010
	Cramer's V	.548	.010
N of Valid Cases		22	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

أما مقياس الأقتران= Phi $= 0.548 = \sqrt{\frac{\chi^2}{n}} = \sqrt{\frac{6.6}{22}}$ فأن قيمة p-value =0.010 فتشير إلى معنوية الاقتران بمستوى دلالة 5 % ،حيث تستعمل إحصائية χ^2 لاختبار الاستقلالية بدرجة حرية(r-1)(c-1) .

<u>ملاحظة 1 :</u> يحتسب تصحيح Yates للاستمرارية Continuity Correction في حالة أن درجة الحرية تساوي واحد أي في جداول 2×2 حسب الصيغة التالية $\chi^2 = \sum \frac{(|O_i - E_i| - 1/2)^2}{E_i}$.

<u>ملاحظة 2 :</u> أن مقياس الأقتران Phi غالباً ما يستعمل مع جداول الاقتران 2*2 والحالة العامة لهذا المقياس هو ما يعرف بـ Cramer Coefficient C الذي يحتسب لجداول متعددة r*c من العلاقة التالية:

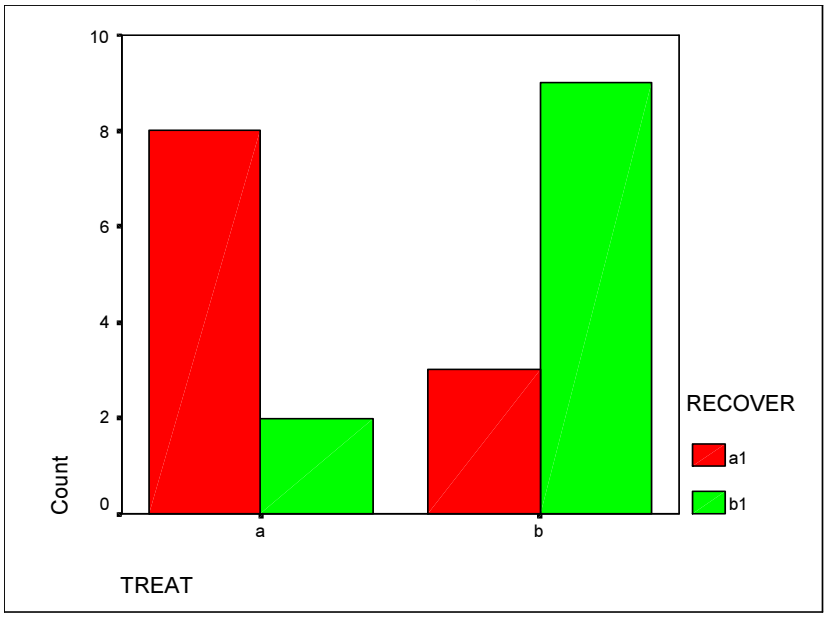
$$C = \sqrt{\frac{\chi^2}{N(L-1)}} = \sqrt{\frac{6.6}{22}} = 0.548$$

حيث أن N يمثل حجم العينة و L العدد الأصغر بين الصفوف والأعمدة في جدول التوافق حيث تستعمل إحصائية χ^2 لاختبار الاستقلالية بدرجة حرية (r-1)(c-1) ونلاحظ أن قيمة هذا المقياس مساوية تماماً لمقياس الاقتران Phi .

<u>ملاحظة 3</u> :هناك مقياس اخر مقارب يعرف بمعامل التوافق Contingency Coefficient (موجود في خانة Nominal في صندوق حوار Crosstsbs:Statistics المذكور أنفاً) ويحتسب لأي جدول توافقي r*c من العلاقة التالية:

$$C=\sqrt{\frac{\chi^2}{\chi^2+N}}=\sqrt{\frac{6.6}{6.6+22}}=0.480$$

حيث أن N يمثل حجم العينة و تستعمل إحصائية χ^2 لاختبار الاستقلالية بدرجة حرية (r-1)(c-1). حيث أن قيمة P-Value المستخدمة في اختبار الاستقلالية هي نفسها للمعاملات الثلاثة أعلاه وتساوي 0.010. أن ناتج تأشير Check Box المجاور لـ Display Clustered Bar Charts في صندوق حوار Crosstabs يعطينا المخطط التالي :



ملاحظة 4 :

1. يمكن التحكم بعرض محتويات جدول Crosstab وذلك بنقر الزر Cells في صندوق حوار Crosstabs حيث يظهر صندوق حوار Cell Display الذي يحتوي الخيارات التالية :

- Counts الذي يتضمن خيارين :

Observed : وهو الخيار الافتراضي حيث تملأ خلايا الجدول بالتكرار المشاهد O_i .

Expected : بموجب هذا الخيار تملأ خلايا الجدول بالتكرار المتوقع E_i .

- Percentages الذي يتضمن ثلاثة خيارات :

Rows :تملأ خلايا الجدول بالنسب المؤية من مجموع الصف .

Columns : تملأ خلايا الجدول بالنسب المؤية من مجموع العمود .

Total : تملأ خلايا الجدول بالنسب المؤية من المجموع الكلي .

- Residuals :ويتضمن ثلاثة خيارات :

Unstandardised : تملأ خلايا الجدول بالفرق بين التكرار المشاهد والتكرار المتوقع $Oi-Ei$.

Standardized : تملأ خلايا الجدول بالفرق بين التكرار المشاهد والتكرار المتوقع مقسوماً على الخطأ المعياري له .

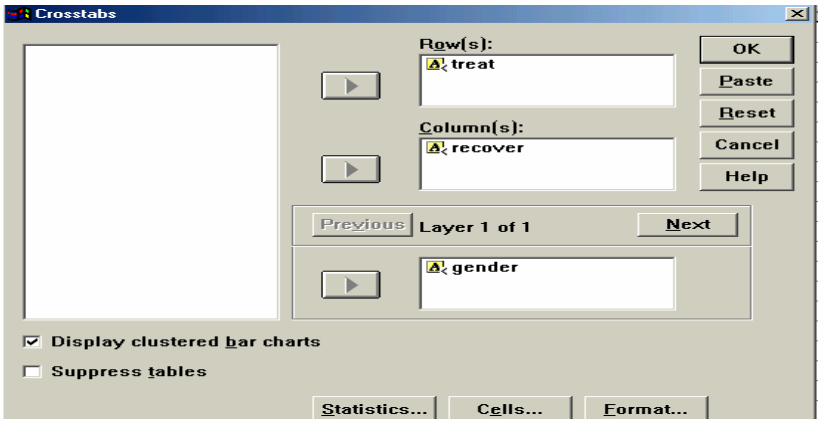
Adj.Standardised :نفس الخيار السابق معبراً عنه بوحدات الانحراف المعياري عن المتوسط .

ملاحظة 5 : يمكن ترتيب صفوف الجدول تصاعدياً أو تنازلياً بنقر الزر Format في صندوق حوار Crosstabs

2. لتنفيذ المطلوب الثاني نتبع الخطوات التالية :

➢ من شريط القوائم نختار Crosstabs → Descriptive Statistics → Analyze

فيظهر صندوق حوار Crosstabs الذي نرتبه كما يلي :



نلاحظ أنه تم إدخال المتغير gender ضمن قائمة Layer (طبقة) ويعرف بمتغير السيطرة Control Variable وهو متغير فئوي categorical Variable (له عدد محدد من القيم في هذا المثال للمتغير gender قيمتين m و f) حيث يتم تكوين الجدول Crosstabs للعلاقة بين المتغيرين treat و recover واحتساب المقاييس لكل قيمة من قيم متغير السيطرة أي أنه يتم تكوين جدولين أحدهما للذكور m والأخر للإناث f .

➢ عند نقر OK في صندوق حوار Crosstabs نحصل على النتائج التالية :

TREAT * RECOVER * GENDER Crosstabulation

Count

GENDER			RECOVER a1	RECOVER b1	Total
f	TREAT	a	2	1	3
		b	1	4	5
	Total		3	5	8
m	TREAT	a	6	1	7
		b	2	5	7
	Total		8	6	14

Chi-Square Tests

GENDER		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
f	Pearson Chi-Square	1.742[b]	1	.187		
	Continuity Correction[a]	.320	1	.572		
	Likelihood Ratio	1.762	1	.184		
	Fisher's Exact Test				.464	.286
	N of Valid Cases	8				
m	Pearson Chi-Square	4.667[c]	1	.031		
	Continuity Correction[a]	2.625	1	.105		
	Likelihood Ratio	5.004	1	.025		
	Fisher's Exact Test				.103	.051
	N of Valid Cases	14				

a. Computed only for a 2x2 table

b. 4 cells (100.0%) have expected count less than 5. The minimum expected count is 1.13.

c. 4 cells (100.0%) have expected count less than 5. The minimum expected count is 3.00.

نلاحظ من خلال قيمة p-value لإحصائية Pearson Chi-Square عدم وجود علاقة بين تعـــاطي الدواء والشفاء من المرض بالنسبة للمرضى الإناث في حين تؤكد الإحصائية على وجـــود علاقـــة بالنســـبة للمرضى الذكور ولمستوى دلالة 5% .

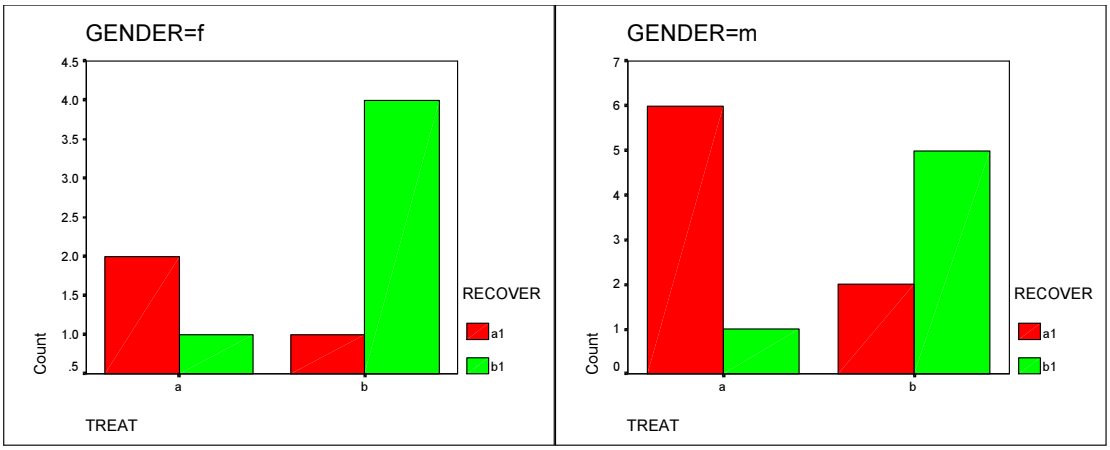
أن إحدى الفرضيات لاختبار Chi-Square في الجداول من نوع 2×2 تنص علـــى أن التكـــرار المتوقع Ei يجب أن لايقل عن 5 بأي حال لأي من خلايا الجدول وألا فأن الاختبار سوف يقود الى نتـــائج مظللة وفي هذا المثال نلاحظ أن التكرار المتوقع يقل عن 5 لكافة خلايا الجدول100% ولكل مـــن الـــذكور والإناث مما يستوجب توخي الحذر في اعتماد نتائج الاختبار .

Symmetric Measures

GENDER			Value	Approx. Sig.
f	Nominal by Nominal	Phi	.467	.187
		Cramer's V	.467	.187
	N of Valid Cases		8	
m	Nominal by Nominal	Phi	.577	.031
		Cramer's V	.577	.031
	N of Valid Cases		14	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.



الفصل الثامن

مقارنة المتوسطات

Compare Means

(8 –1) *تحليل المتوسطات Means*

يفيد هذا التحليل في حساب المتوسطات للمجاميع الجزئية subgroups لمتغير معين بالإضافة الى مؤشرات أخرى مثل (المجموع، عدد الحالات للمجموعة الجزئية ، الوسيط ، المنوال ، الانحراف المعياري ...الخ) ويمكن أيضاً أجراء تحليل التباين لمعيار واحد واختبار التأثير الخطي للمعاملات Test of Linearity واختبار الاقتران eta .

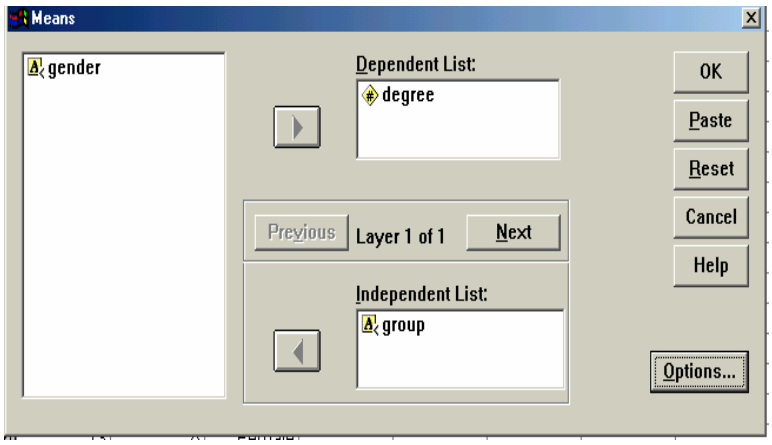
مثال 1

الجدول التالي يبين توزيع درجات 16 طالب ينتمون الى 3 مجاميع مختلفة A، B ، C وحسب الجنس Gender والتي تم إدخالها في شاشة Data Editor على الشكل التالي : :

degree	group	gender
70	A	Female
90	B	Male
88	A	Male
86	B	Male
68	C	Male
64	C	Male
76	B	Male
83	A	Female
79	B	Female
55	C	Female
97	B	Male
100	A	Male
64	C	Female
59	C	Female
90	A	Male
73	A	Female

لحساب متوسط الدرجات حسب المجاميع الفرعية A ، B ، C وعدد الحالات والانحراف المعياري لكل مجموعة نتبع الخطوات التالية :-

- من شريط القوائم أختر Means → Compare means → Analyze فيظهر صندوق حوار Means الذي نقوم بترتيبه على الشكل التالي :-

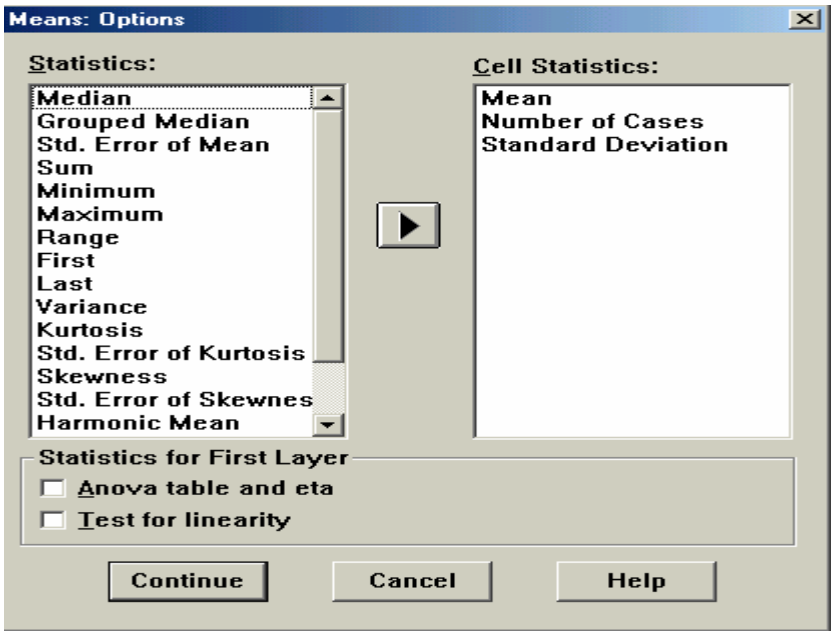


حيث أن :-

Dependent List : المتغير المعتمد أو ما يعرف بمتغير الاستجابة Response Variable نظراً لتـــأثره بالمتغير المستقل Independent Variable . أن المتغير المعتمد في هذا المثال هـــو درجات الطلبة Degree .

Independent List : وهو المتغير المستقل أو المؤثر ويمثل المعاملات Treatments في حالة تصـــميم التجارب أو تحليل التباين ANOVA . أن المتغير المستقل في هذا المثال هو المجاميع الجزئية group الذي يعمل كمتغير تجزئة لدرجات الطلبة .

- أنقر زر Options في صندوق حوار Means ثم أختر Means ، No. of Cases ، Standard Deviation في صندوق حوار Means:Options وهي المؤشرات المطلوب حسابها حسب المجاميع الجزئية حيث يظهر هذا الصندوق بعد الترتيب كما يلي:-



- أنقر زر Continue ثم OK فتظهر النتائج التالية :-

Report

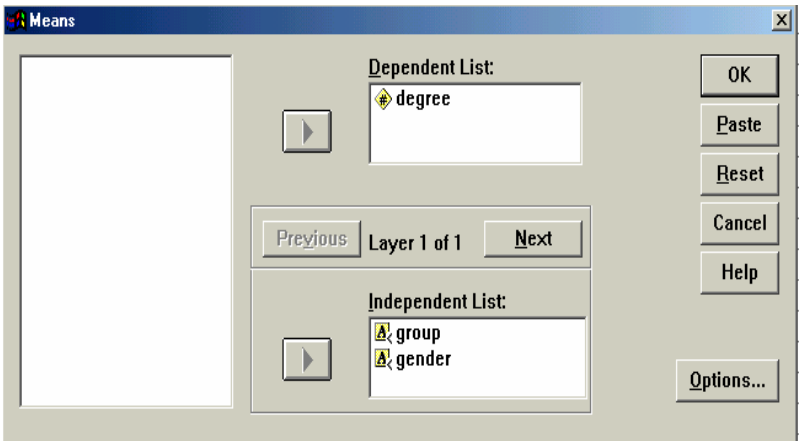
DEGREE

GROUP	Mean	N	Std. Deviation
A	84.00	6	11.19
B	85.60	5	8.44
C	62.00	5	5.05
Total	77.63	16	13.65

نلاحظ أنه تم احتساب المؤشرات المطلوبة للمتغير Degree حسب كل مجموعة جزئية.

مثال 2 :

لنفس بيانات المثال 1 يطلب حساب المتوسط وعدد الحالات والانحراف المعياري لمتغير الدرجات حسب المجاميع group وحسب الجنس gender بصورة مستقلة . لتنفيذ ذلك نتبع نفس الخطوات الواردة في المثال 1 ثم نقوم بترتيب صندوق حوار Means على الشكل التالي :



تظهر نتائج هذا الترتيب كما يلي

DEGREE * GROUP

DEGREE

GROUP	Mean	N	Std. Deviation
A	84.00	6	11.19
B	85.60	5	8.44
C	62.00	5	5.05
Total	77.63	16	13.65

DEGREE * GENDER

DEGREE

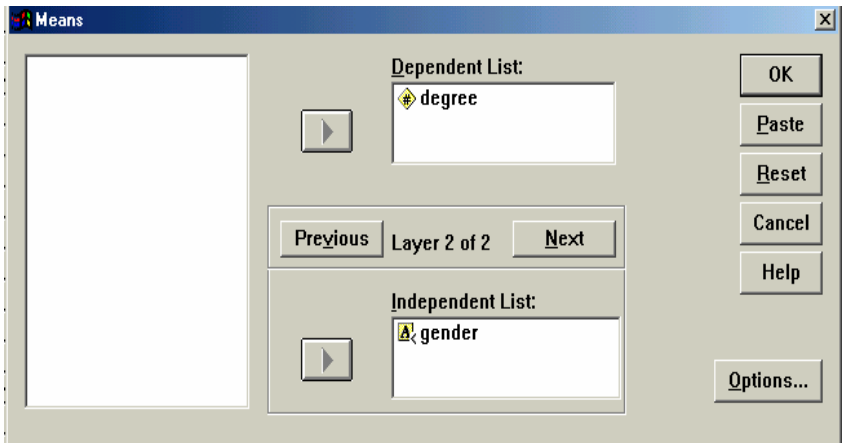
GENDER	Mean	N	Std. Deviation
Female	69.00	7	10.28
Male	84.33	9	12.43
Total	77.63	16	13.65

مثال 3

لنفس المثال 1 يطلب حساب المتوسط وعدد الحالات والانحراف المعياري للمتغير degree حسب الجنس (ذكور،إناث) في داخل كل من المجاميع الفرعية A ، B ، C . لتنفيذ ذلك نقوم بتكوين طبقتين Layers في صندوق حوار Means الطبقة الأولى تضم المتغير group والطبقة الثانية تضم المتغير gender ويتم عمل هذه الطبقات عند ظهور صندوق حوار Means حسب التسلسل التالي :

1. أنقل المتغير group الى خانة Independent List في الطبقة الأولى Layer1 .
2. أنقر زر Next أعلى خانة Independent List للانتقال الى الطبقة الثانية Layer2 .
3. أنقل المتغير gender الى خانة Independent List في الطبقة الثانية Layer2.

الشكل التالي يبين صندوق حوار Means للطبقة الثانية :



للرجوع الى الطبقة الأولى أنقر زر Previousأعلى خانة Independent List. تظهر نتائج هذا الترتيب عند نقر الزر OK كما يلي :

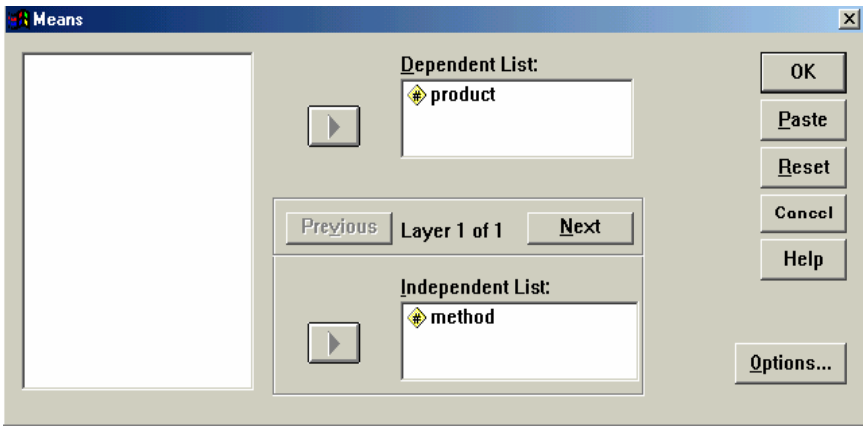
Report

DEGREE

GROUP	GENDER	Mean	N	Std. Deviation
A	Female	75.33	3	6.81
	Male	92.67	3	6.43
	Total	84.00	6	11.19
B	Female	79.00	1	.
	Male	87.25	4	8.77
	Total	85.60	5	8.44
C	Female	59.33	3	4.51
	Male	66.00	2	2.83
	Total	62.00	5	5.05
Total	Female	69.00	7	10.28
	Male	84.33	9	12.43
	Total	77.63	16	13.65

مثال 4 : (اختبار التأثير الخطي Test of Linearity)

هذا الاختبار يفيد في اكتشاف وجود اتجاه عام خطي Linear trend معنوي لأوساط المعالجات في تحليل التباين لمعيار واحد One-Way ANOVA. إذا أردنا اختبار التأثير الخطي للمعاملات للبيانات الواردة في المثال 1 من البند (1-9) نقوم بترتيب صندوق حوار Means كما يلي :



ثم نقوم بتأشير الخيار بين ANOVA and eta و Test for Linearity في صندوق حوار Means : Options الوارد أنفاً حيث تظهر النتائج كما يلي :

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
PRODUCT * METH	Between Groups	(Combined)	402.000	3	134.000	7.053	.012
		Linearity	86.400	1	86.400	4.547	.066
		Deviation from Linea	315.600	2	157.800	8.305	.011
	Within Groups		152.000	8	19.000		
	Total		554.000	11			

Measures of Association

	R	R Squared	Eta	Eta Squared
PRODUCT * METHOD	-.395	.156	.852	.726

أن جدول تحليل التباين يشير الى وجود فروق معنوية بين متوسطات المعالجات (F=7.053). يتم الحصول مجموع مربعات الانحرافات عن الخطية كما يلي :

SS Deviation From Linearity = SS. (Combined) – SS. Linearity =402-86.4=315.6

وتحتسب قيمة F لاختبار Test for Linearity كما يلي :

F=MS. Deviation From Linearity/ within Groups MS.=157.8/19=8.31

وأن قيمة P-value المقابلة وتساوي 0.011 تدعونا إلى رفض فرضية العدم القائلة بأن أوساط المعالجات تقع على خط مستقيم في حالة أجراء الاختبار بمستوى دلالة 5% أي أن الأوساط تنحرف عن الخط المستقيم (أنظر البند 8-2 حول تعريف P-Value) .

أن قيمة Eta هي مقياس للاقتران (الارتباط) Measure of association بين المتغير المعتمد والمتغير المستقل وتقع ضمن المدى 0 الى 1 حيث أن القيمة 0 تشير الى انعدام الارتباط والقيمة 1 تشير الى ارتباط تام ، أما Eta-Square فتمثل نسبة التباينات الكلية في المتغير المعتمد (product) التي يفسرها المتغير المستقل (Method) حيث أن Eta-Square = SS. between Groups / SS. Total = 402/554= 0.726.

أن قيمة R تمثل معامل الارتباط البسيط Simple Correlation بين المتغير (product) والمتغير (Method) باعتبار أن هذا الأخير يعبر عنه بكودات من 1 الى4 وفي حالة وجود أكثر من متغير مستقل فأن R يمثل معامل الارتباط المتعدد Multiple R أما R-Square فهو مقياس لجودة توفيق نموذج الانحدار الخطى Linear Regression .

(8 – 2) *أختبار T لعينة واحدة One Sample T-Test*

يفيد هذا الاختبار في اكتشاف وجود اختلاف معنوي Significant Difference لمتوسط المجتمع الذي سحبت منه العينة عن قيمة ثابتة Constant .إضافة الى إمكانية تقدير فترة ثقة لمتوسط المجتمع Confidence Interval ويستعمل هذا الاختبار للعينات الصغيرة (n<30) .

مثال

المتغير التالي Weight يمثل عينة حجمها 10 لأوزان لحم الدجاج بعمر 52 يوماً وقد تم إدخال المتغير في شاشة Data editor .

WEIGHT	1.10	.80	1.20	1.00	.90	1.20	1.10	.90	.80	1.20

يطلب مايلي

1. اختبار فرضية العدم القائلة بأن المتوسط الحسابي للمجتمع الذي سحبت منه العينة هو 1.25ضد الفرضية البديلة القائلة بأن متوسط المجتمع لا يساوي 1.25بمستوى دلالة 5% و1% .
2. تقدير فترة ثقة99% للوسط الحسابي للمجتمع .

الحل

يمكن كتابة فرضيتي العدم H_0 و البديلة H_1 كما يلي :-

$$H0: \mu = 1.25 \quad \text{or} \quad \mu - 1.25 = 0$$
$$H1: \mu \neq 1.25$$

تكون إحصائية t المستخدمة في الاختبار كما يلي :-

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

حيث أن $\bar{x}$ =1.020 يمثل الوسط الحسابي للعينة وأن S =.162 يمثل الانحراف المعياري للعينة وn=10 حجم العينة وأن μ الوسط الحسابي للمجتمع بموجب فرضية العدم ويساوي 1.25 وان هذه الإحصائية تتبع

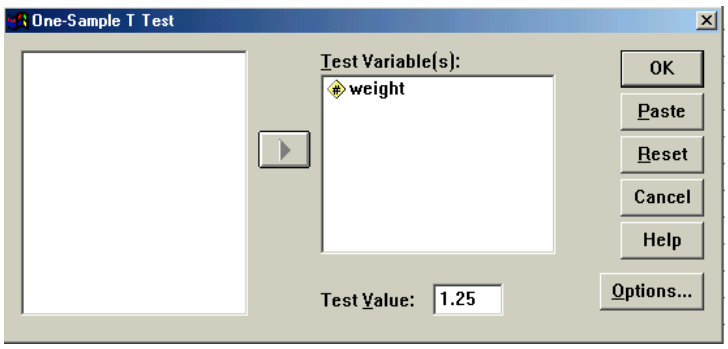
توزيع t بدرجة حرية $v = n-1 = 9$ (تحتسب إحصائيتي $\bar{x}$ و S من قبل البرنامج) .

لتنفيذ هذا الاختبار نتبع الخطوات التالية :

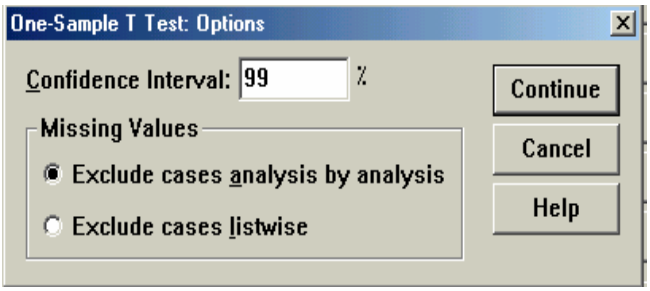
- من شريط القوائم نختار

Analyze → Compare Means → One sample T-Test

فيظهر صندوق حوار One sample t-Test الذي نقوم بترتيبه كما يلي :



- أنقر زر Options فيظهر صندوق حوار Options الذي نرتبه كما في الشكل التالي :



يمكن بواسطة الخيار Options تحديد فترة الثقة (المطلوب الثاني من المثال)حيث نقوم بكتابة الفترة 99% في حقل Confidence Interval.

أما حقل Missing Values فيستفاد منه في التعامل مع القيم المفقودة (في حالة أجراء T-Test لمجموعة متغيرات أي أن خانة Test Variables في صندوق حوار One sample T-Test تضم اكثر من متغير) وفي هذه الحالة يمكن تأشير أحد الخيارين التاليين :

Exclude Cases Analysis by Analysis : يتم استبعاد الحالات المفقودة Missing Cases للمتغير الذي يجرى عليه اختبار T في حالة احتوائه على قيم مفقودة وفي حالة وجود متغير أخر لا يحتوي قيم مفقودة فأن كافة قيم المتغير تستخدم لاحتساب إحصائية T أي أنه يمكن أن يكون حجم العينة مختلفاً للمتغيرات المضمنة في التحليل .

Exclude Cases List Wise : يتم تضمين الحالات الصحيحةValid(غير المفقودة) فقط لكافة المتغيرات فإذا كانت الحالة 6 مثلا مفقودة لأحد المتغيرات فأنه يتم استبعاد الحالة رقم 6 لبقية المتغيرات (وأن كانت غير مفقودة لهذه المتغيرات) وفي هذه الحالة يكون حجم العينة متساوياً للمتغيرات المضمنة في التحليل .

في هذا المثال لا يؤثر اختيار أي من الخيارين على النتائج النهائية للتحليل لعدم وجود قيم مفقودة أولا ولاحتواء خانة Test Variablesعلى متغير واحد ثانياً .

- عند نقر زر OK في صندوق حوار One sample T-Test تظهر النتائج التالية للتحليل :

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
WEIGHT	10	1.020	.162	5.121E-02

One-Sample Test

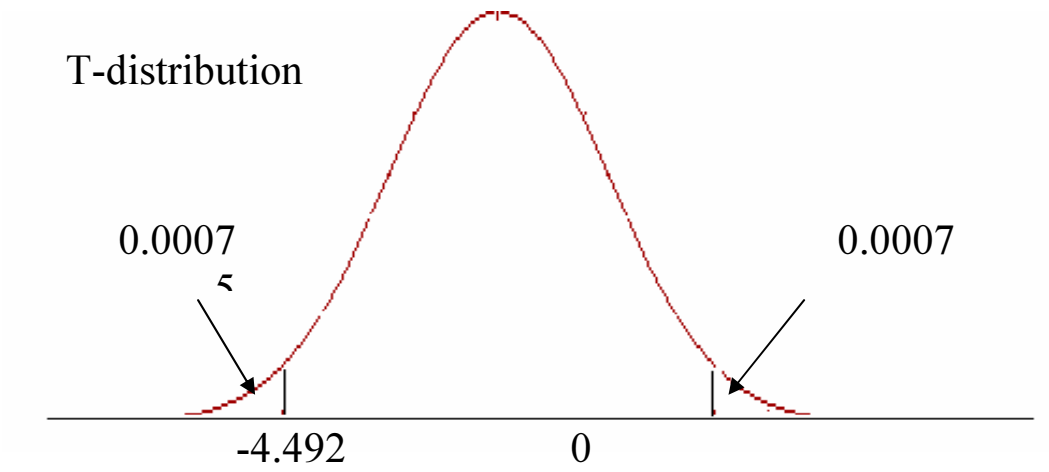
	Test Value = 1.25					
					99% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
WEIGHT	-4.492	9	.0015	-.230	-.396	-6.36E-02

حيث تحتسب إحصائية t بموجب الصيغة التالية 0.0512 /(1.020-1.25) = -4.492 t = وهـذه تعرف بـــ t المحسوبة t(cal.) ويجري مقارنتها بقيمة t الجدولية t(tab) بدرجة حرية v = 9 ولمســتوى دلالة αحيث ترفض H0 في حالة أن $|t(cal.)| \geq t(tab.), v, \alpha/2$ حيث أن قيمة t الجدولية (تســتخرج من جداول توزيع t)ولمستوى دلالة $\alpha/2$ (لان الأختبار من طرفين Two Tailed test) هي كالأتي :

$$t,9,0.025 = 2.262 \quad (\alpha = 5\%)$$
$$t,9,0.005 = 3.250 \quad (\alpha = 1\%)$$

وبما أن القيمة المطلقة لـــ t المحسوبة (4.492) أكبر من الجدولية ، أذن نرفض فرضية العدم ونأخذ الفرضية البديلة أي أن الوسط الحسابي للمجتمع يختلف معنوياً عن القيمة 1.25 لمستويي دلالة 5%و1% .

أن الطريقة الثانية المستخدمة في الاختبار تعتمد على القيمة P-value وتتميز على الطريقة الأولـــى كونها لا تحتاج لاستخدام جداول التوزيعات ويتم احتسابها مباشرة من قبل برنامج SPSS وهي القيمة Sig. في الجدول أعلاه .ويمكن تعريف P-value بأنها اقل قيمة لـــ α التي ترفض عندها فرضية العدم .حيــث نرفض فرضية العدم أذا كانت P-value أقل من α . المخطط التالي يوضح طريقة احتساب P-value .



$$\text{P-value} = \Pr(t \geq 4.492) + \Pr(t \leq -4.492) = 0.00075+0.00075=0.0015$$

Pr يمثل الاحتمال Probability

وبما أن P-value < 0.05 وأن P-value <0.01 أذن نرفض فرضية العدم لمستويي دلالة 5% 1%

أما بالنسبة لفترة ثقة 99% للفرق بين متوسط العينة ومتوسط المجتمع Mean Difference فيمكن كتابتها كما يلي :

$$\Pr(\ -0.396 < \bar{x} - \mu - 0.23 < -0.0636) = 99\%$$

أن تقدير فترة ثقة للفروق بين الأوساط قليل الاستعمال في التحليل الإحصائي وبدلاً من ذلك نقوم بتقدير فترة ثقة 99% لمتوسط المجتمع μ كما يلي :

$$\Pr(\ \bar{x} - t9{,}0.005*(S/\sqrt{n}) < \mu < \bar{x} + t9{,}0.005*(S/\sqrt{n})\ ) = 99\%$$

$$\Pr(\ 0.8536 < \mu < 1.1864) = 99\%$$

أي أحتمال أن يقع متوسط المجتمع بين القيمتين 0.8536 و1.1864 يساوي 99% وعليه نرفض فرضية العدم بمستوى دلالة 5% لأن القيمة 1.25 تقع خارج فترة الثقة وهذه تعتبر طريقة ثانية لاختبار الفرضيات .

(8- 3) *أختبار T للفرق بين متوسطي عينيتين Independent samples T-Test*

يستعمل هذا الاختبار للمقارنة بين متوسطي مجموعتين من الحالات وتستعمل إحصائية T لأجراء الاختبار .

مثال

في تجربة لمقارنة نسبة البروتين في صنفين من الحنطة A و B تم اختبار 12 نباتاً من كل صنف وقدرت نسبة البروتين فيها وكانت النتائج كالتالي :

صنف الحنطة A	صنف الحنطة B
12.5	9.4
9.4	8.4
11.7	11.6
11.3	7.2
9.9	9.7
8.7	7.0
9.6	10.4
11.5	8.2
10.3	6.9
10.6	12.7
9.6	7.3
9.7	9.2

المطلوب اختبار وجود فرق معنوي بين متوسطي نسبة البروتين في الصنفين لمستويي دلالة 5%و1% .أي اختبار الفرضية التالية :

$$H0 : \mu A = \mu B$$

$$H1 : \mu A \neq \mu B$$

لتنفيذ ذلك نتبع الخطوات التالية :

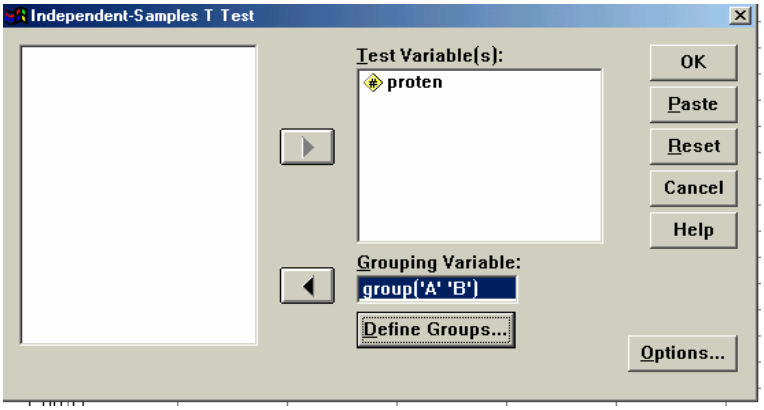
- نقوم بإدخال البيانات في Data Editor كما يظهر في الشكل المجاور حيث أن المتغير Proten يمثل نسبة البروتين والمتغير Group هو متغير تجزئة .

Proten	Group
12.50	A
9.40	A
11.70	A
11.30	A
9.90	A
8.70	A
9.60	A
11.50	A
10.30	A
10.60	A
9.60	A
9.70	A
9.40	B
8.40	B
11.60	B
7.20	B
9.70	B
7.00	B
10.40	B
8.20	B
6.90	B
12.70	B
7.30	B
9.20	B

- من شريط القوائم أختر

Analyze → Compare Means → Independent Samples T Test

فيظهر صندوق حوار Independent Samples T Test الذي نقوم ترتيبه كما يلي :



فقد قمنا بإدخال المتغير Proten في قائمة Test Variables والمتغير Group (متغير التجزئة) في قائمة Grouping Variable ثم تعريف المجاميع A و B عن طريق نقر الزر Define Groups حيث أن المتغير Group له قيمتين متميزتين A و B (هذا المتغير يسمى dichotomous Variable) أو يمكن أن يأخذ القيمتين 1و2 علماً أنه يتوجب التمييز بين الحروف الكبيرة والصغيرة لحالات متغير التجزئة .

كما أن هناك طريقة أخرى في تحديد المجاميع من خلال تحديد نقطة فصل Cut Point في صندوق حوار Define Groups (مثلاً القيمة 10) فكل الحالات التي لها قيمة أقل من 10 لمتغير التجزئة تكون مجموعة واحدة والحالات التي لها قيمة أكبر أو تساوي 10 تكون مجموعة ثانية (في هذه الحالة فأن متغير التجزئة يجب أن يكون عددياً Numeric) .

أما الزر Options في صندوق حوار Independent sample T Test فله نفس الوظيفة في صندوق حوار One Sample T Test .

ملاحظة : يمكن إدخال أكثر من متغير في قائمة Test Variables أي نقوم باختبار الفرق بين متوسطي مجموعتين لعدة متغيرات في أن معاً .

- عند نقر زر OK في صندوق حوار Independent sample T Test تظهر النتائج التالية :

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
PROTEN	A	12	10.4000	1.1314	.3266
	B	12	9.0000	1.8742	.5410

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
PROTEN	Equal variances assumed	2.776	.110	2.215	22	.037	1.400	.6320	8.936E-02	2.7106
	Equal variances not assumed			2.215	18.08	.040	1.400	.6320	7.267E-02	2.7273

نلاحظ أن الاختبار يجرى في حالتين الأولى تفترض تساوي تباين المجتمعين المأخوذة منهما العينتين أي $\sigma_A^2 = \sigma_B^2$ أما الحالة الثانية فتفترض عدم تساوي التباينات $\sigma_A^2 \neq \sigma_B^2$. ففي الحالة الأولى فأن قيمة P-Value = 0.037 <0.05 تدفعنا الى رفض فرضية العدم بمستوى دلالة 5% أي توجد فروق معنوية بين متوسطي نسبة البروتين لصنفي الحنطة أما في حالة الاختبار بمستوى دلالة 1% فأن قيمة P-Value > 0.01 ولهذا نقبل فرضية العدم بمستوى دلالة 1% .

أما اختبار Leven لتجانس التباين فأن P-value = 0.11>0.05 تدعونا الى قبول فرضية العدم القائلة بتجانس (تساوي) تباين المجتمعين المأخوذة منهما العينتين .

(8 –4) أختبار T للمشاهدات المزدوجة Paired-Samples T-Test

يستعمل هذا الاختبار لاكتشاف معنوية الفروق بين متوسطي متغيرين لمجموعة(عينة) واحدة حيث تكون مشاهدات العينة على هيئة أزواج مثلاً اختبار معنوية الفرق بين متوسط نسبة الكوليسترول قبل تعاطي عقار معين وبعده في عينة مكونة من 12 شخصاً .

مثال:

زرع صنفين(AوB) من الذرة الصفراء في عشر مناطق واستخدمت قطعتان متساويتان في كل منطقة زرعت إحداهما بالصنف A وزرعت الأخرى بالصنف B والبيانات التالية تمثل كمية المحصول في كل قطعة :

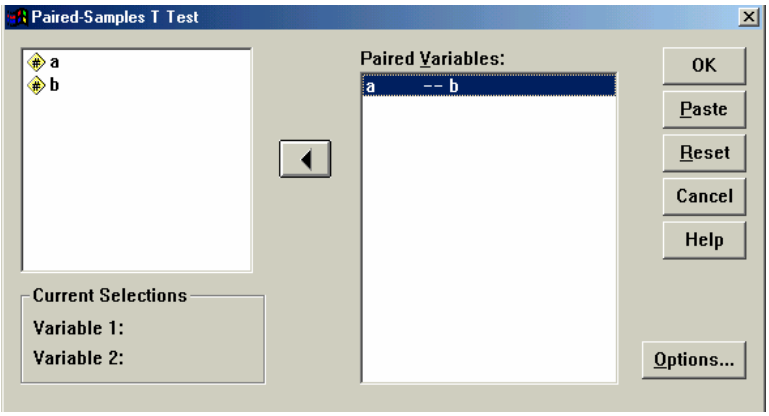
المنطقة	1	2	3	4	5	6	7	8	9	10
الصنفA	127	195	162	170	143	205	168	175	197	136
الصنفB	135	200	160	182	147	200	172	186	194	141

المطلوب اختبار الفرضية القائلة بتساوي متوسطي كميتة الإنتاج للمحصولين بمستوى دلالة 5% .

لتنفيذ هذا الاختبار نتبع الخطوات التالية :

- من شريط القوائم أختر Analyze → Compare Means → Paired Samples T Test

فيظهر صندوق حوار Paired Samples T Testالذي نقوم بترتيبه على الشكل التالي :



المتغيران a و b يمثلان مشاهدات الصنفين a و b حيث يتوجب إدخال كلا المتغيرين في نفس الوقت في قائمة Paired variables باعتبار أن المشاهدات مزدوجة حسب الخطوات التالية :

1. نقر المتغير a بزر الماوس الأيسر لاختياره .
2. أضغط مفتاح Shift مع نقر المتغير b بزر الماوس الأيسر لاختيار المتغيرين معاً .
3. أنقر الزر لنقل المتغيرين الى قائمة Paired variables

➢ أنقر زر OK للحصول على نتائج الاختبار وكما يلي :

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	A	167.80	10	26.58	8.40
	B	171.70	10	24.60	7.78

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	A & B	10	.978	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	A - B	-3.90	5.74	1.82	-8.01	.21	-2.147	9	.060

حيث تم احتساب معامل الارتباط بين المتغيرين 0.978 وهو معنوي بمستوى دلالة 5%و1% .أما بالنسبة لاختبار T فأن قيمة P-value=0.060>0.05 تدعونا الى قبول فرضية العدم $H0: \mu A = \mu B$ أي عدم وجود اختلافات معنوية بين متوسطي الإنتاج لصنفي الحنطة .

الفصل التاسع

تحليل التباين

Analysis of Variance

يقصد بتحليل التباين العمليات الرياضية الخاصة بتقسيم مجموع المربعات الكلي لمجموعة من البيانات الى مصادره المختلفة وتلخص نتائج التحليل في جدول يعرف بجدول تحليل التبـــاين ANOVA Table .أن الهدف من أجراء هذا التحليل هو اختبار فرضية تساوي متوسطات مجموعة من العينات وتعرف بالمعالجات (أو المعاملات Treatments) دفعة واحدة ولهذا فهو يعتبر توسيعاً لاختبار t الذي يستعمل لاختبار الفرضية الخاصة بتساوي متوسطي عينتين فقط .

(9 – 1) تحليل التباين لمعيار واحد One Way ANOVA

مثال1 :

استخدمت أربعة طرق صناعية لانتاج نوع معين من القماش وبثلاثة مكررات لكل طريقـــة وكانـــت نتيجة التجربة كما في الجدول التالي:

الطريقة / المكررا	1	2	3	المتوسط
الطريقة 1	55	47	48	50
الطريقة 2	55	64	64	61
الطريقة 3	55	49	52	52
الطريقة 4	50	44	41	45

المصدر : د.محمود المشهداني وكمال المشهداني ، تصميم وتحليل التجارب ، جامعـــة بغـــداد ، 1989 ، ص63 .

المطلوب مايلي :

1. أجراء تحليل التباين واختبار معنوية الفرق بين متوسطات الطرق الصناعية لمستوى دلالة 5%.
2. في حالة ظهور معنوية الفروق بين الطرق الصناعية باستخدام اختبار F يطلب مايلي :

أ. اختبار معنوية الفروق بين متوسطي كل معالجتين (طريقتين) باســتخدام طريقـــة الفـــرق المعنـــوي الأصغر .L.S.D لمستوى دلالة 5% .

ب. اختبار معنوية الفرق بين متوسط كل من الطرق 2و3و4 وبين متوسط الطريقة 1 بافتراضها الطريقة القياسية (السيطرة) باستخدام طريقة Dunnett .

1. أن الهدف من أجراء تحليل التباين هو اختبار فرضية العدم القائلة بتســـاوي متوســـطات الطـــرق ضـــد الفرضية البديلة التي تنص على عدم تساوي متوسطي طريقتين على الأقل أي اختبار الفرضية التالية :

$$H0: \mu 1 = \mu 2 = \mu 3 = \mu 4$$

$$H1: \mu 1 \neq \mu 2 \neq \mu 3 \neq \mu 4$$

حيث ان μ يمثل متوسط المجتمع للطريقة أما المتوسط الوارد في الجدول أعلاه يمثل متوسط العينـــة لكل طريقة .

لأجراء تحليل التباين نتبع الخطوات التالية :

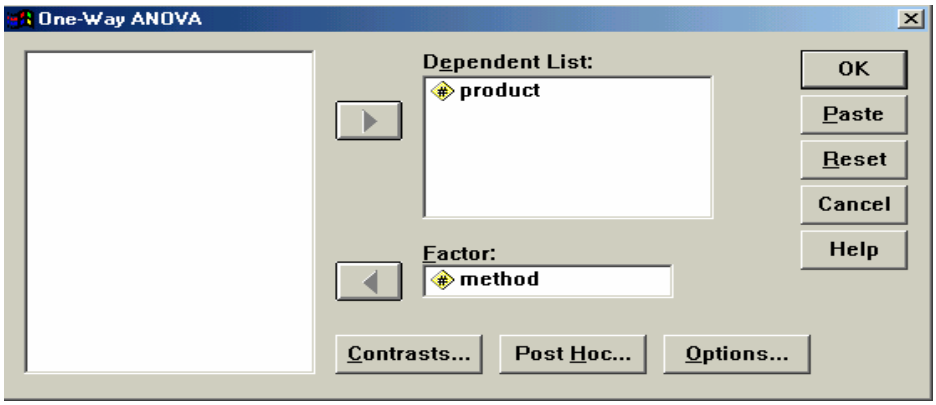
- إدخال البيانات في ورقة Data Editor على الشكل التالي :

method	product
1	55
1	47
1	48
2	55
2	64
2	64
3	55
3	49
3	52
4	50
4	44
4	41

- **من شريط القوائم أختر**

Analyze → Compare Means → One-Way NOVA

فيظهر صندوق حوار One-Way ANOVA الذي نقوم بترتيبه كما يلي :



حيث أن :

Dependent List : يمثل المتغير المعتمد ويمثل نتيجة التجربة في هذا المثال ويجب أن يكون متغيراً عددياً .علماً أنه بالامكان استخدام أكثر من متغير معتمد (المكررات) لنفس المعالجات Factor وفي هذه الحالة نحصل على عدد من جداول تحليل التباين بقدر عدد المتغيرات المضمنة في الخانة Dependent .

Factor :يمثل المتغير المستقل الذي يستعمل في تعريف الفئات ويمثل نوع الطريقة الصناعية في هذا المثال وأن هذا المتغير يجب أن يكون متغيراً عددياً Numeric.

- عند نقر زر OK يعرض البرنامج جدول تحليل التباين كالأتي :

ANOVA

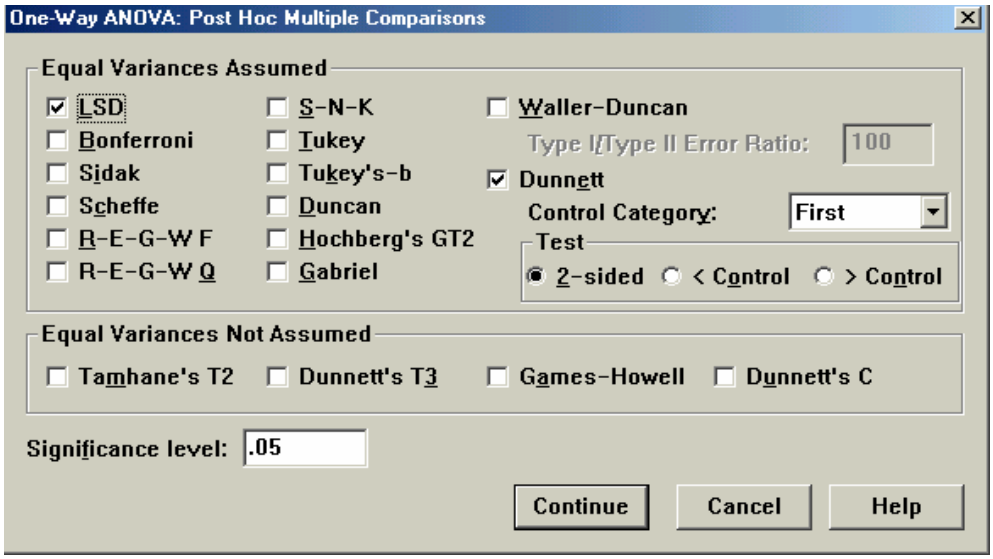
PRODUCT

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	402.000	3	134.000	7.053	.012
Within Groups	152.000	8	19.000		
Total	554.000	11			

أن قيمة P-Value=0.012 المصاحبة لإحصائية F أقل من 0.05 ولهذا نستطيع رفض فرضية العدم لمستوى دلالة 5% أي توجد فروق معنوية بين متوسطات الطرق الصناعية الأربع .

2. نظراً لوجود فروق معنوية بين متوسطات الطرق فهذا يعني عدم تساوي متوسطي طريقتين على الأقل ولاختبار معنوية الفرق لكل زوج من المعالجات نلجأ الى المقارنات المتعددة Multiple Comparisons باستخدام طريقتي .L.S.D و Dunnett بأتباع الخطوات التالية :

- أنقر زر Post Hoc في صندوق حوار One-Way ANOVA فيظهر صندوق حوار Post Hoc Multiple comparisons الذي نرتبه كما يلي :



نلاحظ وجود طرق عديدة للمقارنات المتعددة حيث قمنا بتأشير LSD و Dunnett وبالنسبة للأخيرة يتوجب تحديد مجموعة السيطرة Control Category (First , Last) وقد اخترنا First لأن الطريقة الأولى هي طريقة السيطرة التي يطلب المقارنة بها مع تحديد نوع الاختبار (من طرف واحد أو طرفين في Test. واخيراً تحديد مستوى المعنوية المطلوب 5% في Significance Level .

- عند نقر زر Continue ثم زر OK نحصل على المخرجات التالية :

Multiple Comparisons

Dependent Variable: PRODUCT

	(I) METHOD	(J) METHOD	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
LSD	1	2	-11.00*	3.56	.015	-19.21	-2.79
		3	-2.00	3.56	.590	-10.21	6.21
		4	5.00	3.56	.198	-3.21	13.21
	2	1	11.00*	3.56	.015	2.79	19.21
		3	9.00*	3.56	.035	.79	17.21
		4	16.00*	3.56	.002	7.79	24.21
	3	1	2.00	3.56	.590	-6.21	10.21
		2	-9.00*	3.56	.035	-17.21	-.79
		4	7.00	3.56	.085	-1.21	15.21
	4	1	-5.00	3.56	.198	-13.21	3.21
		2	-16.00*	3.56	.002	-24.21	-7.79
		3	-7.00	3.56	.085	-15.21	1.21
Dunnett t (2-sided) [a]	2	1	11.00*	3.56	.037	.75	21.25
	3	1	2.00	3.56	.896	-8.25	12.25
	4	1	-5.00	3.56	.409	-15.25	5.25

*. The mean difference is significant at the .05 level.

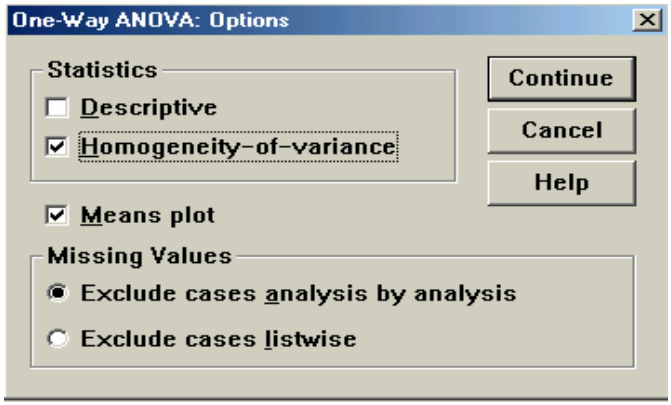
a. Dunnett t-tests treat one group as a control, and compare all other groups against it.

يلاحظ أنه باستعمال طريقة LSD فقد ظهرت فروق معنوية بمستوى دلالة5% بين متوسطات المعالجات (1و2)، (2و3)،(2و4) حيث كانت قيمة P-Value أو Sig. أقل من 0.05 مع وجود علامة * على فروق المتوسطات .

أما أختبار Dunnett فيشير الى وجود فروق معنوية بمستوى دلالة5% بين متوسطي الطريقة الأولى و الطريقة الثانية عند المقارنة بالطريقة الأولى باعتبارها مجموعة سيطرة .

ملاحظة :

عند نقر زر Options في صندوق حوار One Way ANOVA يظهر صندوق الحوار التالي :



حيث أن :

Descriptive : لعرض بعض المقاييس الوصفية مثل الوسط الحسابي ، الانحراف المعياري ....

Homogeneity-of-Variance :لاختبار تجانس تباين المعالجات (الطرق في هذا المثال) باستخدام إحصائية Levene .حيث يعتبر تجانس التباين أحد الفروض المهمة في أجراء تحليل التباين .

Means plot :لعرض تخطيط يمثل متوسط المعالجات .

Exclude Cases Analysis by Analysis :استبعاد الحالات التي تحتوي قيماً مفقودة للمتغيرات المشمولة بالتحليل فقط .

Exclude Cases Listwise :استبعاد الحالات التي تحتوي قيماً مفقودة لأي واحد من المتغيرات .

في هذا المثال لايهم تأشير أي من الخيارين لعدم أحتواء البيانات على قيم مفقودة .

- عند نقر زر Continue في صندوق الحوار أعلاه ثم زر OK في صندوق حوار One-Way ANOVA يتم عرض النتائج التالية :

Test of Homogeneity of Variances

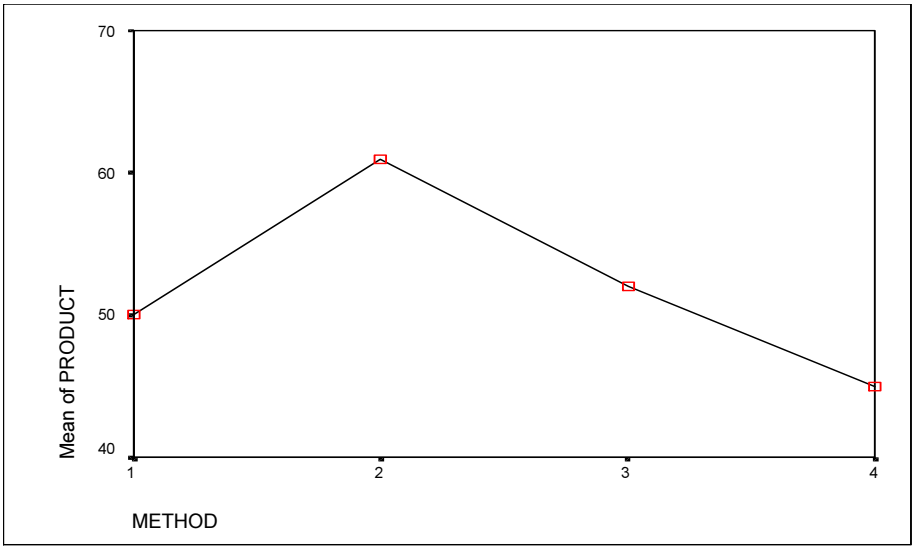
PRODUCT

Levene Statistic	df1	df2	Sig.
.667	3	8	.596

المخرج أعلاه يبين نتيجة أختبار فرضية العدم (تجانس التباين) ضد الفرضية البديلة(عدم تجانس التباين) باستخدام إحصائية Levene حيث أن قيمة P-Value =0.596>0.05 تدعونا الى قبول فرضية العدم القائلة بتجانس التباينات.(راجع الأمر Explore حول إحصائية Levene) .

المخطط التالي يمثل الوسط الحسابي لكل معالجة :

Means Plots



(9 –1–1) المقارنات المستقلة *Orthogonal Comparisons*

في معظم التجارب يهتم الباحث بأجراء مقارنة بين بعض المعالجات دون غيرها .فاذا كان لدينا معالجات (مجموعات) عددها t فأنه يمكن تكوين t-1 من المقارنات المستقلة وتعرف باسم Contrast (تقابلات) الذي تأخذ العلاقة الخطية التالية :

$$Q = \sum C_i Y_{i.} \quad \text{with} \quad \sum C_i = 0$$

حيث ان $Y_{i.}$ يمثل مجموع المعالجة i وأن C_i هي معاملات Coefficients فاذا كان لدينا مقارنتين $Q_1 = \sum C_{1i} Y_{i.}$ وأن $Q_2 = \sum C_{2i} Y_{i.}$ فأنهما تكونان مستقلتين (متعامدتين) Orthogonal إذا كان $\sum C_{1i} C_{2i} = 0$ وعليه يكون من الممكن تقسيم مجموع مربعات المعالجاتSSt الذي له درجة حرية t-1 الى t-1 من المقارنات المستقلة ولكل منها درجة حرية واحدة ويكون بالإمكان اختبار معنوية كل من هذه المقارنات على حدة .لتوضيح ما سبق نأخذ المثال التالي :

مثال 2:

أجريت تجربة لدراسة تأثير خمسة أنواع من العلائق على نمو الفراخ وسجلت بيانات عن الوزن وبأربعة مكررات لكل نوع ونظمت في الجدول التالي

نوع العليقة	الوزن Weight				المجموع $Y_{i.}$
1	46	40	42	40	168
2	51	48	47	42	188
3	36	42	44	46	168
4	42	42	45	43	172
5	35	36	37	36	144

المصدر : د.خاشع الراوي وأخرون : تحليل وتصميم التجارب الزراعية ،جامعة الموصل، ص 56 .

يطلب مايلي :

1. أجراء تحليل التباين لاختبار معنوية الفرق بين متوسطات المعالجات .
2. اختبار معنوية المقارنات الأربعة المستقلة t-1 التالية :

$$Q_1 = 4Y_{1.} - Y_{2.} - Y_{3.} - Y_{4.} - Y_{5.} = 0$$

$$Q_2 = Y_{2.} + Y_{3.} - Y_{4.} - Y_{5.} = 0$$
$$Q_3 = Y_{2.} - Y_{3.} = 0$$
$$Q_4 = Y_{4.} - Y_{5.} = 0$$

في البداية يتم إدخال البيانات الى ورقة Data Editor كما يلي :

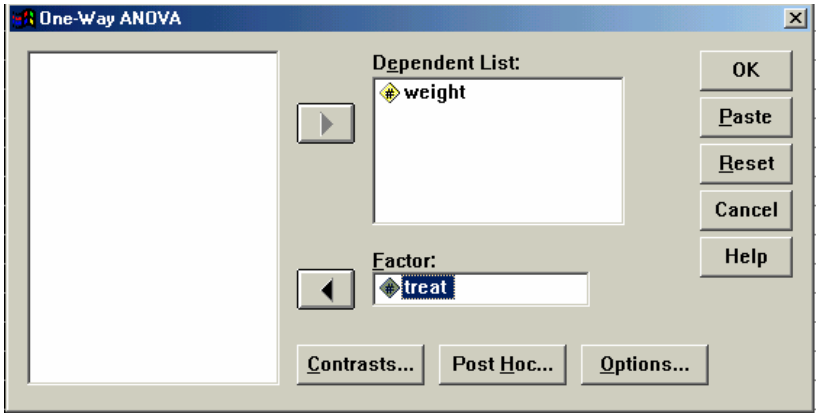
treat	weight
1	46
1	40
1	42
1	40
2	51
2	48
2	47
2	42
3	36
3	42
3	44
3	46
4	42
4	42
4	45
4	43
5	35
5	36
5	37
5	36

لتنفيذ المطلوبين الأول والثاني نتبع الخطوات التالية :

- **من القوائم أختر**

Analyze → Compare Means → One-Way ANOVA يظهر صندوق

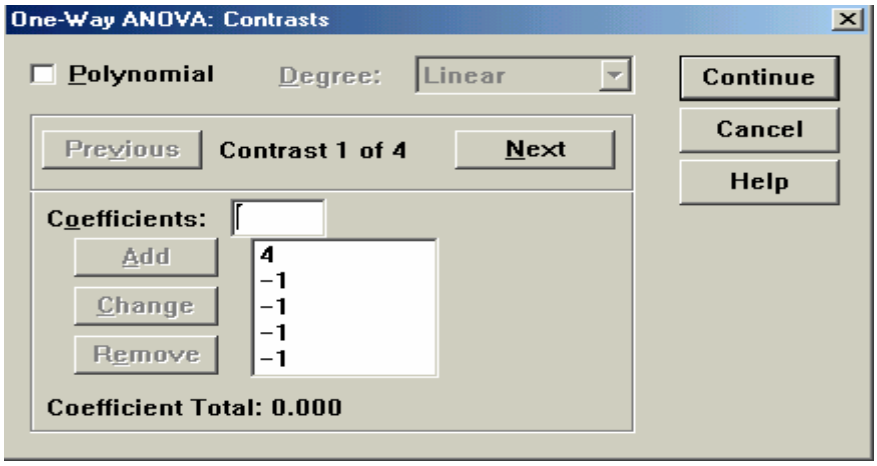
حوار One-Way ANOVA الذي نرتبه كما يلي :



- أنقر زر Contrasts في صندوق حوار One-Way ANOVA فيظهر صندوق حوار Contrasts حيث نقوم بإدخال المقارنات واحدة تلو الأخرى فبالنسبة للمقارنة الأولى يتم إدخال معاملاتها كما يلي :

- أنقر المربع المجاور لكلمة Coefficients ثم أكتب الرقم 4 ثم أنقر زر Add فيتم إضافة المعامل4 الى المستطيل في الأسفل .
- أكتب الرقم 1- في المربع المجاور لكلمة Coefficients ثم أنقر الزر Add فيتم إضافة المعامل -1 الى المستطيل في الأسفل .
- أكتب الرقم 1- في المربع المجاور لكلمة Coefficients ثم أنقر الزر Add فيتم إضافة المعامل -1 الى المستطيل في الأسفل .
- أكتب الرقم 1- في المربع المجاور لكلمة Coefficients ثم أنقر الزر Add فيتم إضافة المعامل -1 الى المستطيل في الأسفل .
- أكتب الرقم 1- في المربع المجاور لكلمة Coefficients ثم أنقر الزر Add فيتم إضافة المعامل -1 الى المستطيل في الأسفل .

وبهذا نكون قد أدخلنا كافة معاملات المقارنة الأولى .لإدخال معاملات المقارنة الثانية أنقر زر Next ثم أبدأ بإدخال المعاملات بنفس طريقة إدخال معاملات المقارنة الأولى .وهكذا لبقية المقارنات . تظهر معاملات المقارنة الأولى مثلاً في صندوق حوار Contrasts كما يلي :



للانتقال الى المقارنة التالية أنقر زر Next وللعودة الى مقارنة سابقة أنقر الزر Previous.

- أنقر زر Continue ثم زر OK في صندوق حوار One-Way ANOVA فيقوم البرنـــامج بعرض المخرج التالي :

ANOVA

WEIGHT

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	248.000	4	62.000	7.154	.002
Within Groups	130.000	15	8.667		
Total	378.000	19			

أن اختبار F يشير الى معنوية الفروق بين متوسطات المعالجات (العلائق) بمستوى دلالة 5%و1% .

Contrast Coefficients

Contrast	TREAT				
	1	2	3	4	5
1	4	-1	-1	-1	-1
2	0	1	1	-1	-1
3	0	1	-1	0	0
4	0	0	0	1	-1

الجدول أعلاه يبين معاملات المقارنات الأربعة .أما الجدول التالي فيبين اختبارt لمعنويـــة المقارنـــات ومنه يتضح أن المقارنة الأولى غير معنوية(Q1=0) بمستوى دلالة 5% لأن P-Value =1 > 0.05 أما بقية المقارنات فهي معنوية (تختلف جوهرياً عن الصفر) بمستوى دلالة 5% لأن P-Value < 0.05 فـــي حالـــة افتراض تساوي التباينات .مع العلم أن قيمة المقارنة Value of Contrast تستخرج كما يلي :

$$\text{Value of Contrast} = \sum C_i Y_{i.} / r$$

r : يمثل عدد المشاهدات في كل مجموعة ويساوي 4 .

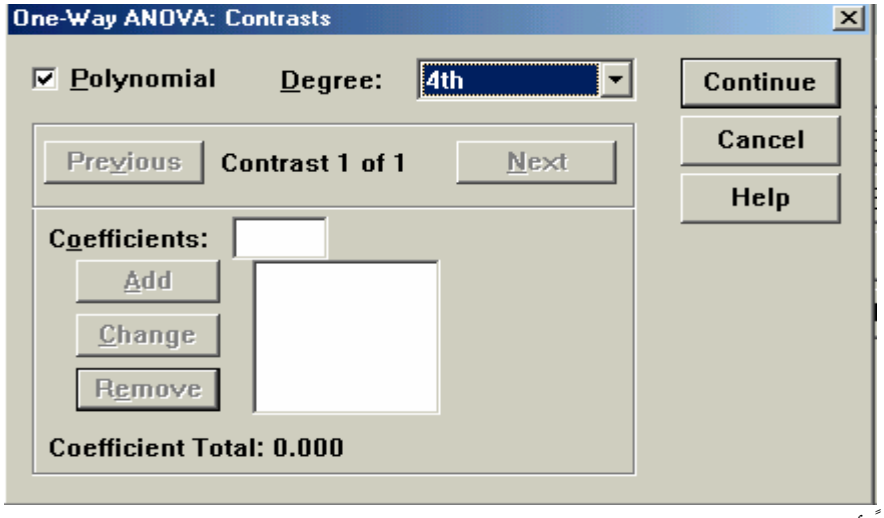
Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
WEIGHT	Assume equal varianc	1	.00	6.58	.000	15	1.000
		2	10.00	2.94	3.397	15	.004
		3	5.00	2.08	2.402	15	.030
		4	7.00	2.08	3.363	15	.004
	Does not assume equa variances	1	.00	6.39	.000	4.727	1.000
		2	10.00	2.97	3.365	6.823	.012
		3	5.00	2.86	1.750	5.880	.132
		4	7.00	.82	8.573	4.800	.000

(9-1-2) تحليل الاتجاهات *Trend Analysis*

يمكن تجزئة مجموع مربعات المعالجات الى مركبات خطية ، من الدرجة الثانية ، تكعيبية ...حيث أن درجة متعدد الحدودPolynomial تساوي t-1 حيث أن t يمثل عدد المعالجات ويستفاد من هذا التحليل في التجارب التي تكون المعالجات فيها عبارة عن مستويات لعامل كمي مثلاً استخدام كميات مختلفة من سماد معين ويكون الهدف هو اختبار معنوية المركبات المختلفة لغرض تكوين فكرة عن استجابة الصفة المدروسة لمستوى المعالجة وبالتالي تقدير المستوى الأمثل من العامل المدروس .

لتحليل مجموع مربعات المعالجات للمثال السابق فأن درجة متعدد الحدود تساوي 4=1-5 أي أنه بالإمكان الحصول على المركبة الخطية ، من الدرجة الثانية ، تكعيبية ، من الدرجة الرابعة .حيث نقوم بتحوير صندوق حوار Contrasts كما يلي :



علماً أن هذه المركبات تكون مستقلة عن بعضها ولها نفس صفات المقارنات Contrasts. لاحظ أننا لانحتاج الى كتابة المعاملات Coefficients بعد أن اخترنا Polynomial وحددنا درجة متعدد الحدود 4th.عند نقر زر Continue ثم زر OK في صندوق حوار One-way ANOVA نحصل على النتيجة التالية :

ANOVA

WEIGHT

			Sum of Squares	df	Mean Square	F	Sig.
Between Groups	(Combined)		248.000	4	62.000	7.154	.002
	Linear Term	Contrast	102.400	1	102.400	11.815	.004
		Deviation	145.600	3	48.533	5.600	.009
	Quadratic Term	Contrast	92.571	1	92.571	10.681	.005
		Deviation	53.029	2	26.514	3.059	.077
	Cubic Term	Contrast	1.600	1	1.600	.185	.674
		Deviation	51.429	1	51.429	5.934	.028
	4th-order Terı	Contrast	51.429	1	51.429	5.934	.028
Within Groups			130.000	15	8.667		
Total			378.000	19			

لاحظ أنه تم توزيع مجموع المربعات بين المعالجات Between Groups على المركبات الأربعة مع اعطاء درجة حرية واحدة لكل مركبة كما أن جميع المركبات ظهرت معنوية عند اختبارها بمستوى دلالة 5% عدا المركبة التكعيبية Cubic term .

(2-9) تحليل التباين لمعيارين Two Way ANOVA

يستفاد من تحليل التباين لمعيارين في اختبار تساوي متوسطات معالجات العامل الأول إضافة الى اختبار تساوي متوسطات العامل الثاني و هذا النموذج يقارب تصميم القطاعات الكاملة العشوائية RCB Design في تصميم التجارب .

مثال3 :

في تجربة قطاعات كاملة عشوائية تم الحصول على الجدول التالي الذي يبين الفرق في سمك الأطار(المتغير المعتمد) بعد قطع مسافة معينة وقد استخدمت أربعة أنواع من الإطارات (العامل الأول) واستخدمت أربعة سيارات (القطاعات وسنعتبرها العامل الثاني) وقد كانت البيانات المختزلة (المبسطة) كما يلي :

D	C	B	A	نوع الإطار / السيارة
0	-1	1	4	1
-2	-1	1	1	2
-2	-3	0	0	3
-4	-4	-5	0	4

المصدر: شارلز هيكس – ترجمة قيس سبع خماس ، المفاهيم الأساسية في تصميم التجارب ،الجامعة المستنصرية ،1984، ص83 .

يرغب الباحث في تكوين جدول تحليل التباين لمعيارين لاختبار معنوية الفرق بين متوسطات أنواع الإطارات Tyre وكذلك معنوية الفرق بين أنواع السياراتcar وبمستوى دلالة 5% .

في البداية نقوم بإدخال البيانات قي Data Editor وكما يلي :

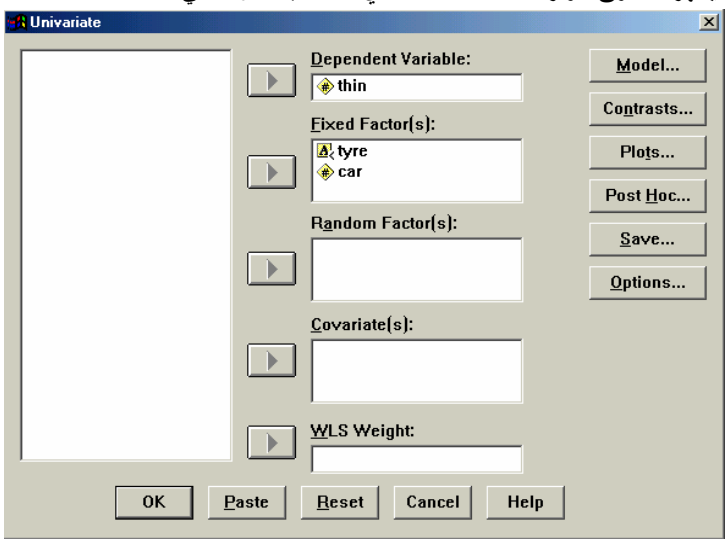
car	tyre	thin
A	1	4
A	2	1
A	3	0
A	4	0
B	1	1
B	2	1
B	3	0
B	4	-5
C	1	-1
C	2	-1
C	3	-3
C	4	-4
D	1	0
D	2	-2
D	3	-2
D	4	-4

لتكوين جدول تحليل التباين نتبع الخطوات التالية :

- من شريط القوائم نختار :

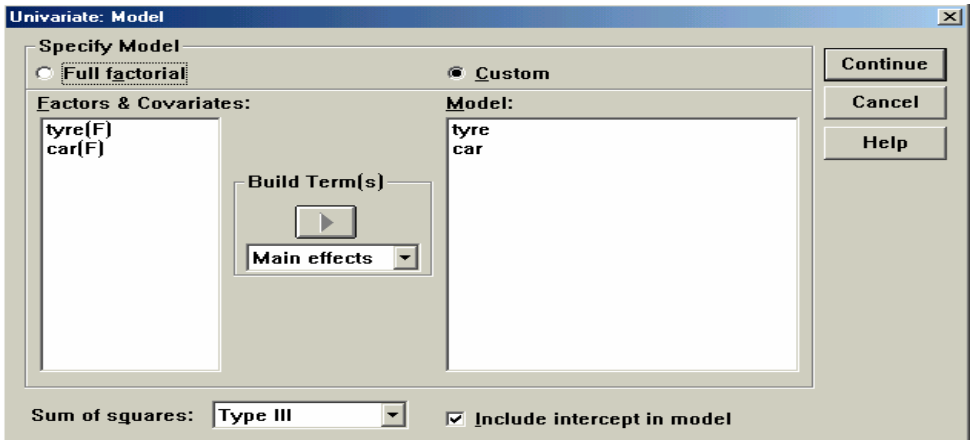
Analyze→ General Linear Model → Univariate

فيظهر صندوق حوار Univariate الذي ننظمه بالشكل التالي :



أدخلنا العاملين في Fixed Factors على اعتبار أن كافة معالجات العامل قد ضمنت في التجربة أما في حالة أخذ عينة من معالجات العامل فأننا سنستعمل خانة Random Factors. لاحظ أن العامل الثابت Fixed Factor يمكن أن يكون متغيراً رمزياً على العكس منه في حالة One-Way ANOVA حيث يجب أن يكون Factor متغيراً عددياً .

- أنقر زر Model فيظهر صندوق حوار Model ، أختر Custom بدلاً من Full Factorial وذلك لأننا لا نرغب في ظهور التفاعل Interaction في جدول تحليل التباين (car*tyre) لعدم وجود درجات حرية كافية للخطأ التجريبي حيث يظهر صندوق حوار Model بعد ترتيبه كما يلي :



حيث أن تأشير Include Intercept in Model يعمل على تضمين الحد الثابت في النموذج الخطي العام باعتباره نموذج انحدار .

أما خانة Build Terms فتستعمل لتعيين نوع التأثيرات Effects التي يراد إظهارها في جدول تحليل التباين حيث قمنا بنقل المتغيرين tyre و car من خانةFactors & Covariates الى خانة Model بنقر

كل منهما بزر الأيسر ثم نقر زر ▶ لتظهر كتأثيرات رئيسية Main Effects (بعد التأكد مـــن أن خانة Build Terms تتضمن الخيار Main effects).

➢ عند نقر زر Continue ثم OK في صندوق حوار Univariate نحصل على المخرج التالي :

Between-Subjects Factors

		N
TYRE	A	4
	B	4
	C	4
	D	4
CAR	1	4
	2	4
	3	4
	4	4

Tests of Between-Subjects Effects

Dependent Variable: THIN

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	69.375[a]	6	11.563	9.000	.002
Intercept	14.063	1	14.063	10.946	.009
TYRE	30.688	3	10.229	7.962	.007
CAR	38.688	3	12.896	10.038	.003
Error	11.563	9	1.285		
Total	95.000	16			
Corrected Total	80.938	15			

a. R Squared = .857 (Adjusted R Squared = .762)

لاحظ أن :

SS. Corrected Model = SS.TYRE + SS. CAR
SS. Corrected Total = SS. Total –SS.Intercept

أن هذا الجدول يخص النموذج الخطي العام GLM الذي هو نموذج انحدار ويمكن تبسيط هذا الجدول (راجع فصل الجداول المحورية) ليكون جدول تحليل تباين بمعيارينTwo-Way ANOVA كما يلي :

Tests of Between-Subjects Effects

Dependent Variable: THIN

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
TYRE	30.688	3	10.229	7.962	.007
CAR	38.688	3	12.896	10.038	.003
Error	11.563	9	1.285		
Corrected Total	80.938	15			

ومنه يتضح معنوية الفروق بين الإطارات tyer وكذلك بين السيارات car بمستوى دلالة 5% .

<u>ملاحظات :</u>

1. لإظهار التفاعل بين tyre وcar (علماً أنه لن نحصل على قيمة F لهذا المثـــال لعـــدم كفايـــة درجات الحرية) نتبع الخطوات التالية:

- أنقر السهم المتجه نحو الأسفل في خانة Build Terms وأختر Interaction في صـــندوق حـــوار Model.
- أختر المتغيرين tyre و car من الخانةFactors & Covariates في نفس الوقت (بنقر المتغير tyre ثم أضغط مفتاح Shift مع نقر المتغير car) .
- أنقر زر ▶ لنقل المتغيرين المذكورين الى خانة Model .

علماً أن نفس الشيء بمكن أن ينجز باختيار Full Factorial في صندوق حوار Model .

2. يمكن استعمال النموذج الخطي العام GLM في أجراء تحليل التباين لمعيار واحد .

<u>مثال 4 : (تحليل التباين لمعيارين مع تسجيل أكثر من مشاهدة لكل وحدة تجريبية)</u>

الجدول التالي يمثل المبيعات الأسبوعية لسلسلة من المتاجر حسب موقـــع الـــرف Shelf Location (العامل الأول) و حجم المتجر Store Size (العامل الثاني) وقد أخذت مشـــاهدتين لكـــل مـــن التوافيـــق الممكنة :

حجم المتجر Size	موقع الرف Shelf Location			
	A	B	C	D
Small	45 50	56 63	65 71	48 53
Medium	57 65	69 78	73 80	60 57
Large	70 78	75 82	82 89	71 75

يطلب تكوين جدول تحليل التباين لمعيارين مع اختبار معنوية الفروق بين متوسطات معالجات عامـــل الحجم Size وبين معالجات الموقع Location واختبار معنوية التفاعل Size*Location بمستوى دلالـــة 5% مع توضيح ذلك بيانياً .

في البداية نقوم بإدخال البيانات في Data Editor وكما يلي :

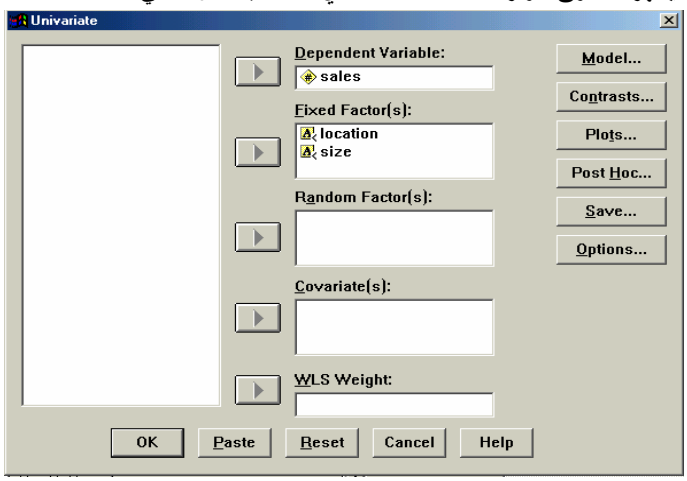
location	size	sales
A	Small	45
A	Small	50
A	Medium	57
A	Medium	65
A	Large	70
A	Large	78
B	Small	56
B	Small	63
B	Medium	69
B	Medium	78
B	Large	75
B	Large	82
C	Small	65
C	Small	71
C	Medium	73
C	Medium	80
C	Large	82
C	Large	89
D	Small	48
D	Small	53
D	Medium	60
D	Medium	57
D	Large	71
D	Large	75

يتم تنفيذ المطلوب حسب الخطوات التالية :

- **من شريط القوائم نختار :**

Analyze →General Linear Model → Univariate

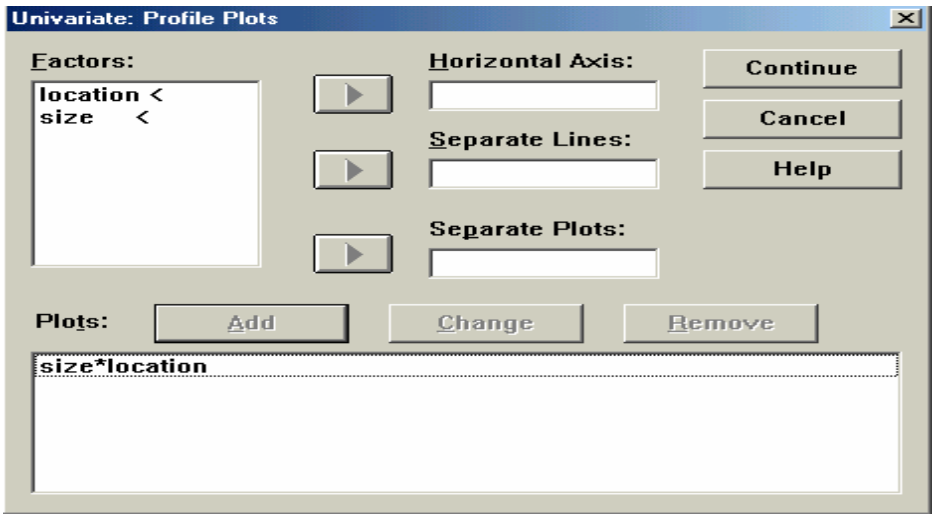
فيظهر صندوق حوار Univariate الذي ننظمه بالشكل التالي :



- أنقر زر Model وأختر Full Factorial لأننا نرغب في الحصول على التفاعل Interaction لوجود عدد كافي من درجات الحرية للخطأ العشوائي .

(إذا كنت ترغب في اختيار جزء من التأثيرات أنقر Custom وأختر التأثيرات المطلوبة) علماً أن Full Factorial هو الخيار الافتراضي Default .

- أنقر زر Plots لأعداد تخطيط تمثل فيه فئات المتغير size على المحور الأفقي تقابله سلاسل المتوسطات الأربع للمتغير Location فيظهر صندوق حوار Profile Plots أو Interaction Plot الذي يستفاد منه لمقارنة المتوسطات الحدية للمتغير المعتمد في النموذج و ينظم كما يلي :

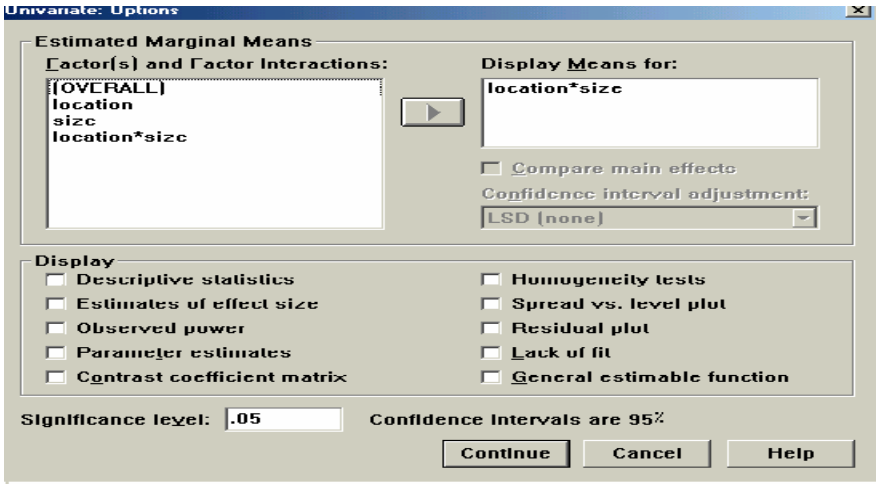


يتم إدخال المتغيرات في الصندوق أعلاه كما يلي :

- أنقل المتغير size من خانة Factors الى خانة Horizontal Axis.
- أنقل المتغير Location الى خانة Separate Lines .
- ستلاحظ أن الزر Add أصبح فعالاً أنقره فيضاف التفاعل size*location الى خانة Plot في الأسفل .

أما قائمة Separate Plots فيستفاد منها في عمل تخطيط عند كل مستوى من مستويات عامل ثالث .

- أنقر زر Continue للرجوع الى صندوق حوارUnivariate .
- أنقر زر Options لعرض المتوسطات الحدية Estimated Marginal Means للتفاعل location*size حيث ينظم صندوق حوار Options بالشكل التالي :



➢ عند نقر زر Continue ثم OK في صندوق حوار Univariate نحصل على النتائج التالية .

Tests of Between-Subjects Effects

Dependent Variable: SALES

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3019.333[a]	11	274.485	12.767	.000
Intercept	108272.667	1	108272.667	5035.938	.000
LOCATION	1102.333	3	367.444	17.090	.000
SIZE	1828.083	2	914.042	42.514	.000
LOCATION * SIZE	88.917	6	14.819	.689	.663
Error	258.000	12	21.500		
Total	111550.000	24			
Corrected Total	3277.333	23			

a. R Squared = .921 (Adjusted R Squared = .849)

أن هذا الجدول يمثل نتائج الانحدار للنموذج الخطي العام وبالإمكان تبسيطه ليكون متلائماً مع نمـــوذج تحليل التباين وكما يلي :

Tests of Between-Subjects Effects

Dependent Variable: SALES

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
LOCATION	1102.333	3	367.444	17.090	.000
SIZE	1828.083	2	914.042	42.514	.000
LOCATION * SIZE	88.917	6	14.819	.689	.663
Error	258.000	12	21.500		
Corrected Total	3277.333	23			

أن اختبار F يشير إلى وجود فروقات معنوية بين معالجات العامل Location وكذلك بين معالجـــات العامل size بينما لم تظهر معنوية حد التفاعل بمستوى دلالة 5% لأن0.05 > p-Value = 0.663 .

Estimated Marginal Means

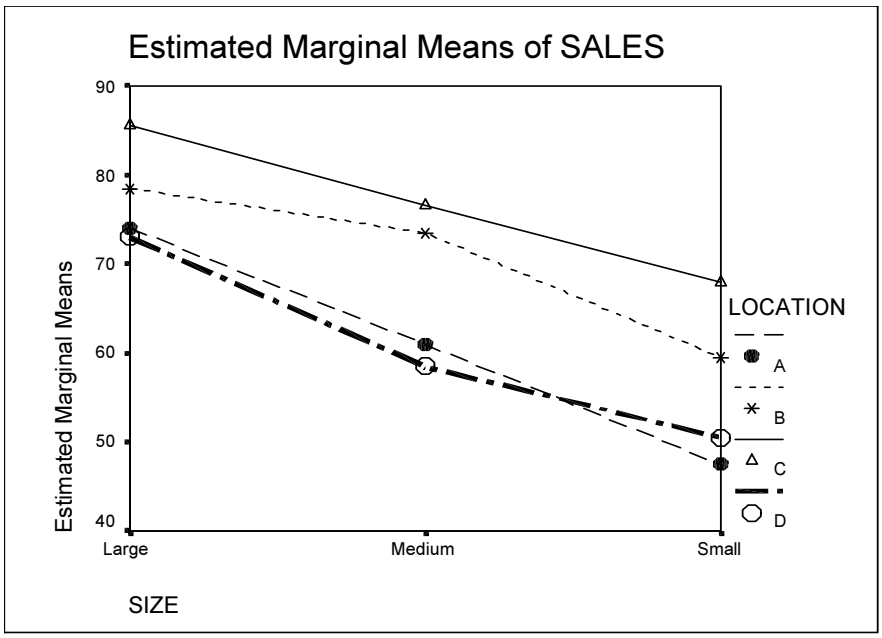
LOCATION * SIZE

Dependent Variable: SALES

LOCATION	SIZE	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
A	Large	74.000	3.279	66.856	81.144
	Medium	61.000	3.279	53.856	68.144
	Small	47.500	3.279	40.356	54.644
B	Large	78.500	3.279	71.356	85.644
	Medium	73.500	3.279	66.356	80.644
	Small	59.500	3.279	52.356	66.644
C	Large	85.500	3.279	78.356	92.644
	Medium	76.500	3.279	69.356	83.644
	Small	68.000	3.279	60.856	75.144
D	Large	73.000	3.279	65.856	80.144
	Medium	58.500	3.279	51.356	65.644
	Small	50.500	3.279	43.356	57.644

المخطط التالي يبين درجة التفاعل بين العـــاملين Location و Size حيـــث يلاحـــظ أن الخطـــوط متوازية تقريباً مما يشير الى أن الفروق في المواقع لا تتغير بتغير حجم المتجر كما أن الخطـــوط المتوازيـــة تؤشر عدم وجود تفاعل بين العاملين size و location وهذا يدعم ما توصلنا اليه بتطبيق اختبار F .

Profile Plots



(9 –3) *تحليل التباين المشترك* *Covariance Analysis*

من المشاكل التي تواجه تحليل التباين أو تصميم التجـــارب أن هنـــاك عوامـــل مســـتقلة (مصـــاحبة) Covariates يرمز لها بالرمز X التي تؤثر على الظاهرة المراد قياس تأثير المعالجات عليها ويرمز لهـــا بالرمز Y وهو المتغير المعتمد Dependent Variable مثلاً عند قياس تأثير عدد من الأغذية (المعالجات) على زيادة وزن الحيوان Y فأن وزن الحيوان X في بداية التجربة يؤثر على مدى استجابته للغذاء وبالتالي فأن الباحث في نهاية التجربة لا يستطيع معرفة فيما إذا كانت الاختلافات بين الأوزان النهائية للحيوان (فـــي نهاية التجربة) ناتجة عن أثر الغذاء أم عن أثر الوزن في بداية التجربة .لغرض التخلص من أثـــر المتغيـــر المصاحب X نستخدم تحليل التباين المشترك وهو أسلوب يجمع بين تحليل التباين والانحدار .

مثال 5 :

أجريت تجربة على تغذية الأفراخ لمقارنة تأثير أربعة علائق (المعالجات) مختلفة علـــى زيـــادة وزن الأفراخ وأستخدمت ستة مشاهدات لكل معالجة .الجدول التالي يبين الزيادة في وزن الأفراخ Y فـــي نهايـــة التجربة كما يبين وزن الأفراخ X عند بداية التجربة .

المعالجات treat	المتغير	Observations المشاهدات					
T1	X Y	30 165	27 170	20 130	21 156	33 167	29 151
T2	X Y	24 180	31 169	20 171	26 161	20 180	25 170
T3	X Y	34 156	32 189	35 138	35 190	30 160	29 172
T4	X Y	41 201	32 173	30 200	35 193	28 142	36 189

يطلب اختبار معنوية الفروق بين متوسطات المعالجات بعد إزالة أثر الوزن في بداية التجربة X (أسلوب تحليل التباين المشترك) .

في البداية نرتب بيانات الجدول في ورقة Data Editor بالشكل التالي :

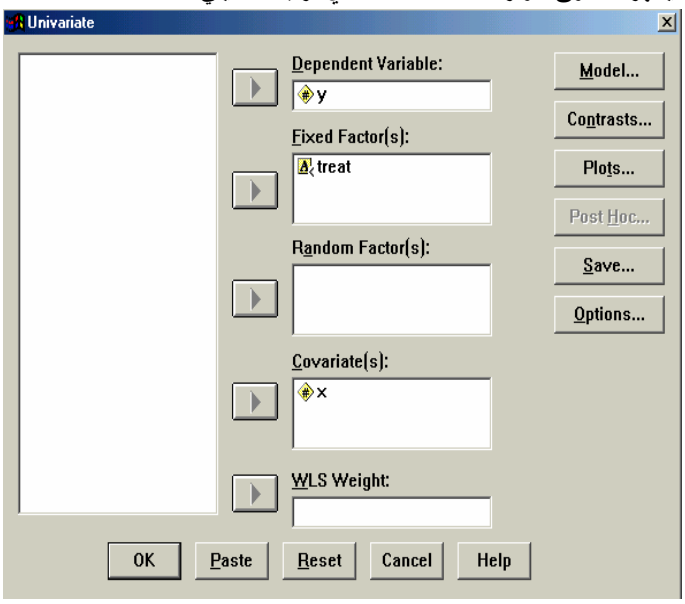
treat	X	Y
T1	30	165
T1	27	170
T1	20	130
T1	21	156
T1	33	167
T1	29	151
T2	24	180
T2	31	169
T2	20	171
T2	26	161
T2	20	180
T2	25	170
T3	34	156
T3	32	189
T3	35	138
T3	35	190
T3	30	160
T3	29	172
T4	41	201
T4	32	173
T4	30	200
T4	35	193
T4	28	142
T4	36	189

لتنفيذ أسلوب تحليل التباين المشترك نتبع الخطوات التالية :

- **من القوائم اختر**

Analyze→ General Linear Model→ Univariate

فيظهر صندوق حوار Univariate الذي نرتبه كما يلي :



- **عند نقر زر OK نحصل على النتيجة التالية:**

Tests of Between-Subjects Effects

Dependent Variable: Y

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2845.956[a]	4	711.489	2.572	.071
Intercept	6938.602	1	6938.602	25.087	.000
X	682.831	1	682.831	2.469	.133
TREAT	1609.595	3	536.532	1.940	.157
Error	5255.002	19	276.579		
Total	699323.000	24			
Corrected Total	8100.958	23			

a. R Squared = .351 (Adjusted R Squared = .215)

وقد بسطنا الجدول أعلاه ليكون في صورة تحليل التباين المشترك وكما يلي :

Tests of Between-Subjects Effects

Dependent Variable: Y

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
	a				
TREAT	1609.595	3	536.532	1.940	.157
Error	5255.002	19	276.579		
Total+Error	6864.597	22			

a. R Squared = .351 (Adjusted R Squared = .215)

يلاحظ أن قيمة $0.157 > 0.05$ =P-Value مما يدعونا الى قبول فرضية العدم بمستوى دلالة 5% التي تنص على تساوي متوسطات الأنواع الأربعة من العلائق بعد إزالة أثر المتغير المصاحب X .

الفصل العاشر

تحليل الارتباط و الانحدار

Correlation & Regression Analysis

(10 – 1) الارتباط *Correlation*

تسمى العلاقة بين ظاهرتين بالارتباط Correlation مثلاً العلاقة بين الدخل والاستهلاك فمن البديهي أن زيادة دخل الفرد يؤدي الى زيادة استهلاكه من السلع والخدمات (علاقة طردية) كما أن ارتفاع سعر سلعة ما يؤدي الى تدني الطلب عليها (علاقة عكسية) علماً أن الارتباط قد يكون خطياً Linear أو غير خطي Non Linear .أن المقياس المستخدم الذي يقيس درجة الارتباط يعرف بمعامل الارتباط Correlation Coefficients ويرمز له r وتتراوح قيمته بين –1 الى 1 $(-1 \le r \le 1)$.

(10 – 2) الارتباط الخطي البسيط *Simple Linear Correlation*

يحتسب معامل الارتباط الخطي البسيط بافتراض وجود علاقة خطية بين أثنين من المتغيرات فقط مع العلم أن الحصول على قيمة صغيرة (قريبة من الصفر) لهذا المعامل لا يعني عدم وجود علاقة بين المتغيرين فقد توجد علاقة من الدرجة الثانية (ارتباط غير خطي) .

مثال 1:

البيانات التالية تمثل الدرجات التي حققها 10 طلاب في اختبار اللغة Lang واختبار الرياضيات Math وتظهر في ورقة Data Editor لبرنامج SPSS كما يلي :

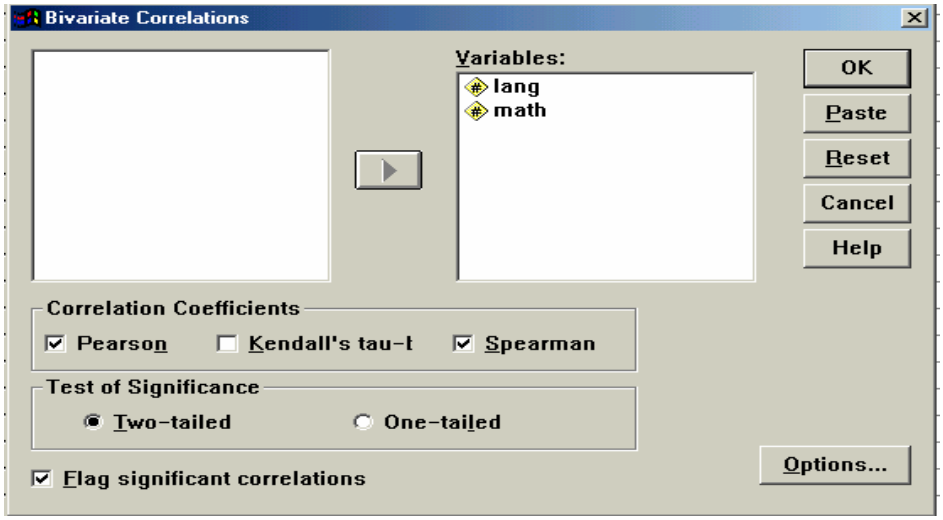
Lang	Math
60	56
68	60
60	64
74	82
80	76
84	72
80	74
72	66
62	64
82	86

المطلوب مايلي :

1. أوجد معامل الارتباط الخطي البسيط لـ Pearson ، معامل ارتباط الرتب لـ Spearman .
2. أختبر معنوية معامل الارتباط بمستوى دلالة 5% .

لتنفيذ المطاليب أعلاه نتبع الخطوات التالية :

- من شريط القوائم أختر Analyze → Correlate→ Bivariate فيظهر صندوق حوار Bivariate Correlation الذي نرتبه كما يلي :



يتضمن حقل Correlation Coefficients الخيارات التالية :

Pearson : لاستخراج معامل الارتباط الخطي البسيط للمتغيرات الكمية .

Kendall's Tau: لاستخراج معامل الارتباط بالطرق اللا معلمية وباستعمال الرتب Ranks في حالة الرغبة في تقدير معامل الارتباط بصورة تقريبية .كما أنه من الضروري استخدام هذا المعامل في حالة كون إحدى الظاهرتين أوكلاهما غير مقاستين (ليست متغيرات كمية) ولكن يمكن إعطائها رتباً تصاعدية أو تنازلية وفي هذه الحالة يتم إدخال الرتب المتناظرة للظاهرتين بدلاً من القيم الأصلية .

Spearman : هو معامل ارتباط للرتب كما هو الحال بالنسبة لمعامل Kendall .

وقد أشرنا الخيارين Pearson و Spearman .

في حقل Test significance أشرنا الخيار Two Tailed لاختبار الفرضية من طرفين . لاختبار الفرضية من طرف واحد نؤشر One Tailed .

Flag Significance Correlation : لتعليم الارتباطات المعنوية (إظهار علامة Star) .

- عند نقر زر OK تظهر النتائج التالية :

Correlations

		LANG	MATH
LANG	Pearson Correlation	1.000	.776**
	Sig. (2-tailed)	.	.008
	N	10	10
MATH	Pearson Correlation	.776**	1.000
	Sig. (2-tailed)	.008	.
	N	10	10

**. Correlation is significant at the 0.01 level

أن قيمة معامل الارتباط الخطي لبيرسون بين LANG و MATH هي r = 0.776 لاختبار الفرضية التالية (أختبار من طرفين) :

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

نستخدم إحصائية T التي تتبع توزيع T بدرجة حرية n-k حيث أن n يمثل حجم المجتمع و k عـــدد المتغيرات .

$$T = r\frac{\sqrt{n-k}}{\sqrt{1-r^2}} = 0.776\frac{\sqrt{10-2}}{\sqrt{1-0.776^2}} = 3.48$$

يمكنك الحصول على P- Value من خلال الأمر Compute Transform ثم استعمال الدالة التجميعية لتوزيع T وكمـــا يلـــي $(1-CDF.T(3.48,8))*2 = 0.008$.أن قيمة P-Value=0.008<0.05 تشير الى أن الارتبـــاط بـــين درجة امتحان اللغة وامتحان الرياضيات يختلف معنوياً عن الصفر بمستوى دلالة 5% (الاخـــتلاف معنـــوي بمستوى دلالة 1% أيضاً) .

المخرج التالي يمثل معامل ارتباط الرتب لسبيرمان :

Nonparametric Correlations

Correlations

			LANG	MATH
Spearman's rho	LANG	Correlation Coefficient	1.000	.786**
		Sig. (2-tailed)	.	.007
		N	10	10
	MATH	Correlation Coefficient	.786**	1.000
		Sig. (2-tailed)	.007	.
		N	10	10

**. Correlation is significant at the .01 level (2-tailed).

لاختبار فرضية العدم $H_0 : \rho = 0$ ضد الفرضـــية البديلـــة $H_1 : \rho \neq 0$ لمعامـــل Spearman تستخدم نفس إحصائية T لأختبار Pearson أعلاه .أن قيمة P-Value المرافقة لمعامل أرتباط Spearman تشير الى معنوية هذا المعامل بمستوى دلالة 5% وكذلك 1% .

ملاحظة :

لاختبار فرضية معامل الارتباط من طرف واحد نؤشر الخيار One Tailed في صـــندوق حـــوار Bivariate Correlations وهنا توجد حالتين للأختبار :

1. أذا كانت قيمة معامل الارتباط المقدرة موجبة يقوم البرنامج باختبار الفرضية التالية :

$$H_0 : \rho = 0$$

$$H_1 : \rho \succ 0$$

مثلاً كانت قيمة معامل أرتباط بيرسون موجبة 0.766 وتظهر نتيجة اختبار الفرضية أعلاه كما يلي :

Correlations

		LANG	MATH
LANG	Pearson Correlation	1.000	.776**
	Sig. (1-tailed)	.	.004
	N	10	10
MATH	Pearson Correlation	.776**	1.000
	Sig. (1-tailed)	.004	.
	N	10	10

**. Correlation is significant at the 0.01 level

حيث يتم احتساب إحصائية T نفسها في حالة الاختبار من طرفين وتساوي 3.48 و يمكن الحصول على P-Value المرافقة من خلال الأمر Compute → Transform ثم استعمال الدالة التجميعية لتوزيع T وكما يلي $(1-CDF.T(3.48,8)) = 0.004$ لاحظ أن المقدار لا يضرب في 2 لأن الاختبار من طرف واحد .أن قيمة P-Value=0.004<0.05 تشير الى أن الارتباط بين درجة امتحان اللغة وامتحان الرياضيات أكبر من الصفر بمستوى دلالة 5% (الاختلاف معنوي بمستوى دلالة 1% أيضاً) .

2. أذا كانت قيمة معامل الارتباط المقدرة سالبة يقوم البرنامج باختبار الفرضية التالية :

$$H_0 : \rho = 0$$

$$H_1 : \rho \prec 0$$

(10 – 3) الارتباط الجزئي Partial Correlation

يقيس معامل الارتباط الجزئي قوة العلاقة بين متغيرين بثبوت متغير ثالث . مثلاً قد نحصل على قيمة عالية لمعامل الارتباط البسيط للعلاقة بين أسعار اللحوم البيضاء واللحوم الحمراء فقد لاتوجد علاقة فعلية بين المتغيرين ولكن كلا المتغيرين يتأثر بعامل ثالث هو المستوى العام للأسعار فاذا استبعدنا المستوى العام للأسعار (أو تثبيته) عند قياس العلاقة بين أسعار اللحوم البيضاء والحمراء فسيتم الحصول على قيمة أقل لمعامل الارتباط وهذا يعرف بالارتباط الجزئي .علماً أنه يمكن استبعاد أي عدد من المتغيرات عند قياس العلاقة بين ظاهرتين .

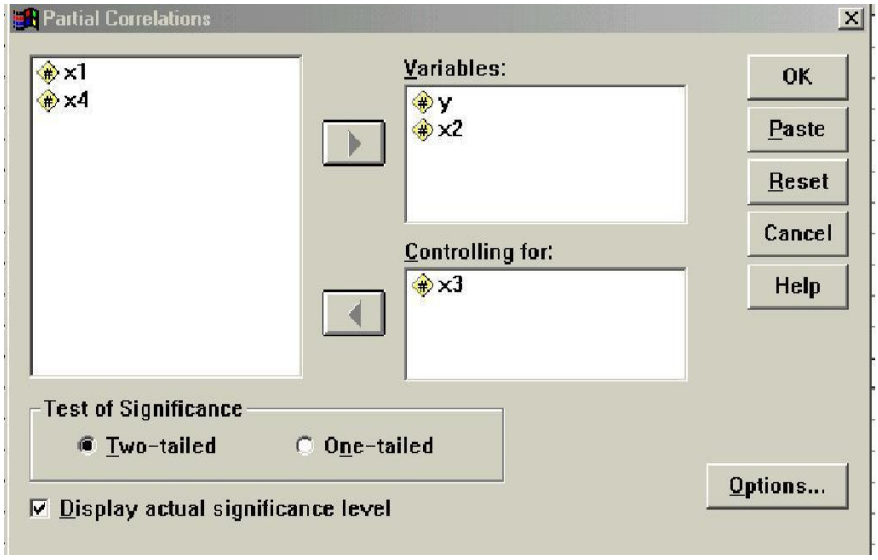
مثال 2:

للبيانات الواردة في المثال 6 من البند (10 – 4 – 3) والتي يبلغ حجم العينة 13 يطلب ما يلي :

1. حساب معامل الارتباط الجزئي بين Y و X_2 بثبات X_3 .
2. حساب معامل الارتباط الجزئي بين Y و X_2 بثبات X_3 , X_4 .
3. أختبر معنوية معاملات الارتباط (من طرفين) بمستوى دلالة 5%.

الحل:

- من شريط القوائم أختر Analyze→ Correlate → Partial فيظهر صندوق حوار Partial Correlation الذي نرتبه كما يلي للمطلوب الأول :



في خانة Variables يتم إدخال المتغيرات التي يراد حساب معامل الارتباط الجزئي لها ، وفي خانة Controlling for يتم إدخال المتغير (المتغيرات) الذي يراد استبعاد أثره .

- عند نقر زر Ok نحصل على النتيجة التالية .

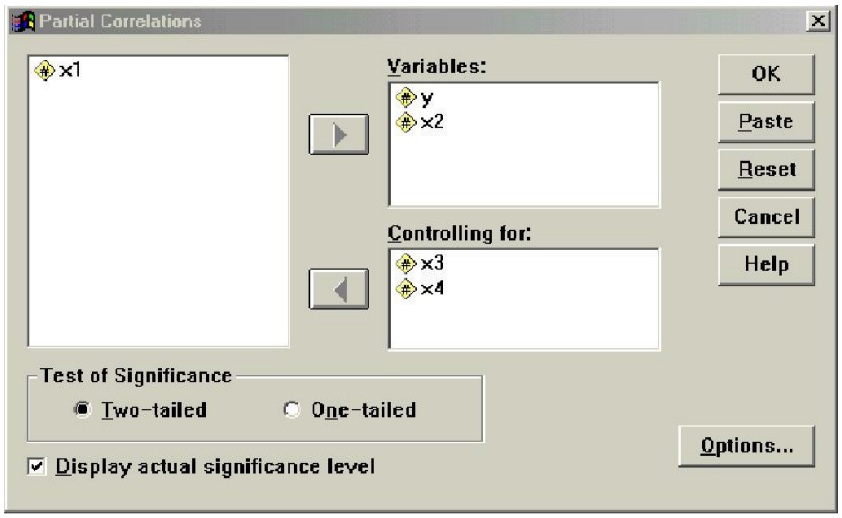
```
--- PARTIAL CORRELATION COEFFICIENTS ---

Controlling for..      X3

                     Y          X2

Y              1.0000       .2851
              (    0)     (   10)
              P= .        P= .369

X2              .2851      1.0000
              (   10)     (    0)
              P= .369     P= .

(Coefficient / (D.F.) / 2-tailed Significance)

" . " is printed if a coefficient cannot be computed
```

حيث أن $r_{yx2.x3}=0.285$ لاختبار معنوية معامل الارتباط الجزئي (ذو طـرفين) تسـتعمل نفـس إحصائية T لاختبار معامل الارتباط البسيط .لاحظ أن الرقم 10 داخل قوسين والذي يظهر في الجدول أعلاه لا يمثل حجم العينة وأنما درجات الحرية اللازمة لاختبار T حيث df = n – k = 13-3 = 10 وأن T قـد احتسبت كما يلي :

$$T = r\frac{\sqrt{n-k}}{\sqrt{1-r^2}} = (0.2851)\frac{\sqrt{13-3}}{\sqrt{1-0.2851^2}} = 0.941$$

وأن قيمة 0.05 > P-Value=0.369 ولهذا فأن معلمة المجتمع ρ لا تختلف معنوياً عـــن الصـــفر بمستوى دلالة 5% .

لحساب معامل الارتباط الجزئي بين Y و X_2 بثبات X_3 , X_4 (المطلـــوب الثـــاني) نكـــرر نفـــس الخطوات السابقة ونرتب صندوق حوار Partial Correlations على الشكل التالي :



وتظهر النتائج كما يلي :

```
--- PARTIAL CORRELATION COEFFICIENTS ---

Controlling for..     X3          X4

                      Y           X2

Y                  1.0000       .2338
                  (    0)     (    9)
                  P= .        P= .489

X2                  .2338      1.0000
                  (    9)     (    0)
                  P= .489     P= .

(Coefficient / (D.F.) / 2-tailed Significance)

" . " is printed if a coefficient cannot be computed
```

ومنه يظهر أن $r_{yx2.x3x4}$ لا يختلف معنوياً عن الصفر بمستوى دلالة 5% .

ملاحظة :

يمكن أستخراج معامل الارتباط البسيط لـــ Pearson بنقـــر زر Options فـــي صـــندوق حـــوار Partial Correlations ثم تأشير الخيار Zero Order Correlations .

(10 – 4) تحليل الانحدار Regression Analysis

أن نموذج الانحدار يعبر عن علاقة بين متغير معتمد Dependent Variable وبين واحد أو اكثر من المتغيرات المستقلة Independent Variables أو Regressors فاذا أحتوى النموذج على متغير مستقل واحد فيعرف بنموذج الانحدار البسيط Simple Regression model واذا أحتوى أكثر من متغير مستقل فهو نموذج الانحدار المتعدد Multiple regression Model كما أن النموذج قد يكون خطياً Linear Model او غير خطي Non Linear Model .

(10 –4–1) نموذج الأنحدار الخطي البسيط

يأخذ الصيغة العامة التالية : $Y = B_0 + B_1 X + e$

حيث أن :

Y : المتغير المعتمد

X :المتغير المستقل

B_0 : الحد الثابت أو معلمة تقاطع خط الانحدار مع المحور الصادي Intersection Parameter .

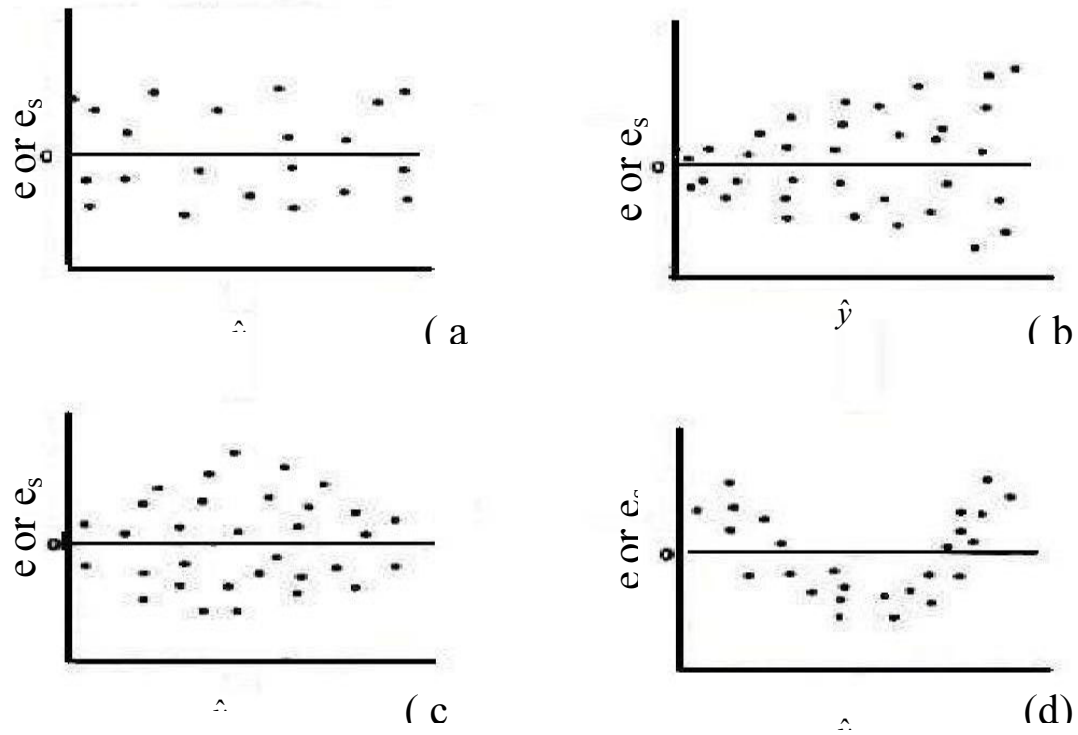
$B1$:معلمة الميل Slop Parameter $B_1 = \frac{\Delta Y}{\Delta X}$

e: الخطأ العشوائي وهو الفرق بين القيمة الحقيقيةY والقيمة التقديرية $\hat{Y}$ويعرف بالمتبقي residual حيث أن $e = Y - \hat{Y}$.

أن الطريقة المتبعة غالباً في تقدير معالم الانحدار B_0 و $B1$ هي طريقة المربعات الصغرى Least Squares Method (OLS) علماً أن نموذج الانحدار الخطي البسيط يجب أن يحقق الفرضيات التالية :

1. وجود علاقة خطية بين Y وX .
2. أن الأخطاء العشوائية تتوزع بمتوسط مساوي للصفر .
3. أن الأخطاء العشوائية لها تباين ثابت يساوي σ^2 (فرضية تجانس تباين الخطأ العشوائيHomoscedasticity) .
4. الأخطاء تتوزع طبيعياً وهذا الشرط ليس ضرورياً لتقدير المعالم بطريقة OLS ولكنه ضروري لاختبار الفرضيات المتعلقة بمعاملات الانحدار B_0 و $B1$.
5. عدم وجود ارتباط ذاتي Autocorrelation بين الأخطاء العشوائية .

بمكن التحقق من توفر فرضيات النموذج الخطي البسيط من خلال تخطيط Scatter plots بتمثيل $\hat{y}$ على المحور الأفقي(أو x) يقابله الخطأ العشوائي e أو الأخطاء المعيارية Standardized Residuals يرمز لها e_s على المحور الرأ سي كما هو واضح في الشكل التالي .



(a) توفر فرضيات التحليل (عدم وجود مشكلة) .

(b) زيادة تباين الخطأ العشوائي بزيادة y .

(c) زيادة وتناقص في تباين الخطأ العشوائي (مشكلة عدم تجانس تباين الخطأ العشوائي) .

(d) عدم ملائمة العلاقة الخطية (يتوجب استعمال نماذج أخرى مثلاً نموذج من الدرجة الثانية) .

<u>مثال 3 :</u>

البيانات التالية تمثل العمر X وضغط الدم Y (ملم زئبق) لعينة مكونة من 10 أشخاص وقد تم إدخال البيانات في ورقة Data Editor وكما يلي :

Obs.	X	Y
1	35	112
2	40	128
3	38	130
4	44	138
5	67	158
6	64	162
7	59	140
8	69	175
9	25	125
10	50	142

يطلب مايلي :

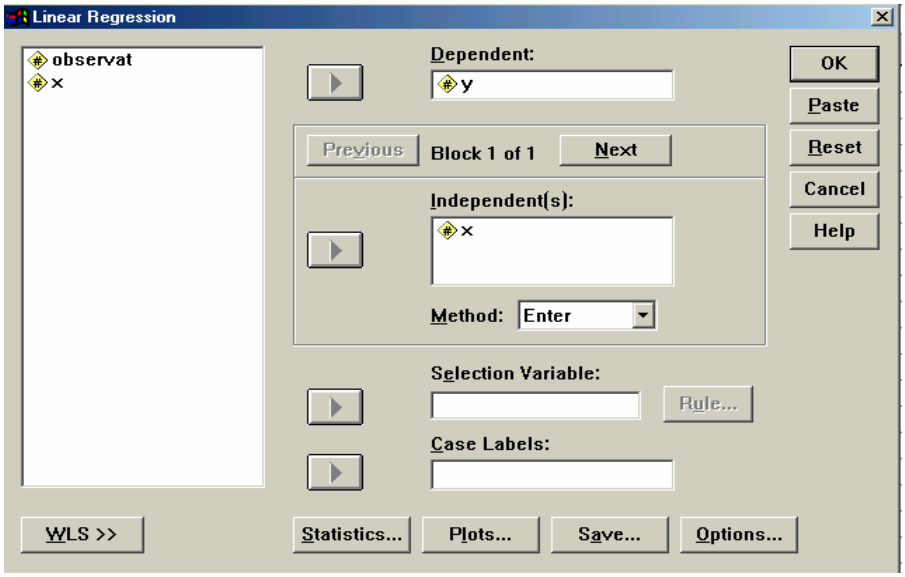
1. استخراج معادلة انحدار Y/X مفترضاً العلاقة الخطية واختبار معنوية معالم النموذج .
2. أستخرج فترة ثقة 95% لكل من معلمتي الانحدار B_0 و B_1 .

3. استخراج جدول تحليل التباين ANOVA .
4. أختبر جودة توفيق النموذج الخطي (باستعمال معامل التحديد R^2) مع تحليل الأخطاء العشوائية بالرسم البياني .
5. أختبر التوزيع الطبيعي للأخطاء العشوائية بيانياً .

الحل :

لتنفيذ المطاليب أعلاه نتبع الخطوات التالية :

- من شريط القوائم اختر Linear Regression → Analyze فيظهر صندوق حوار Linear Regression الذي نرتبه كما يلي :



حيث أن :

Dependent: يمثل المتغير المعتمد .

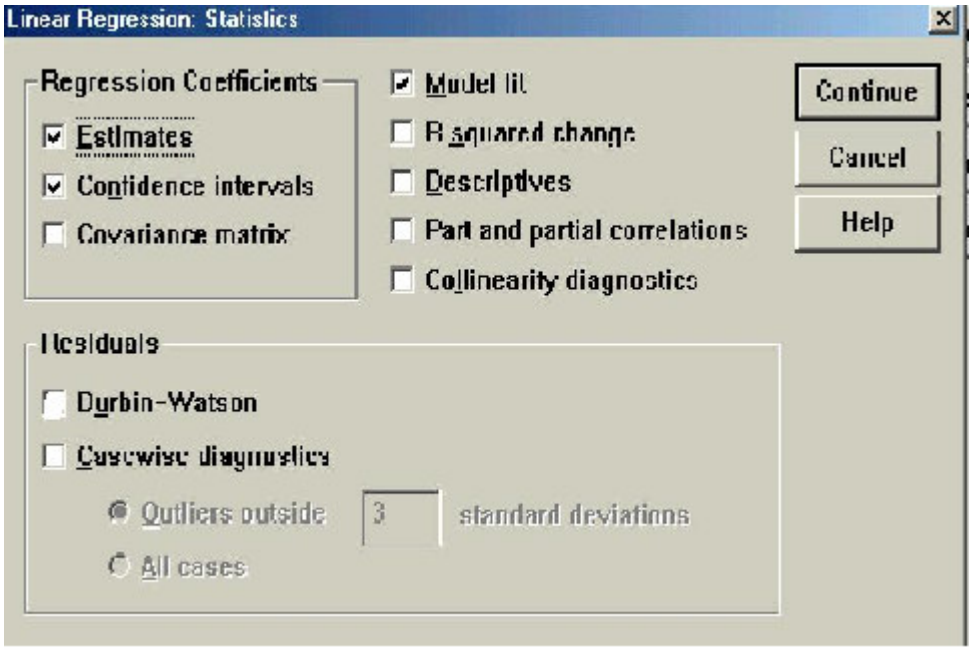
Independent : المتغير (أو المتغيرات المستقلة). يمكن إدخال اكثر من مجموعة من المتغيرات المستقلة كل مجموعة تدخل ضمن Block له رقم تسلسلي ويمكن الانتقال من Block الى أخر بالزرين Next و Previous. فاذا كان لدينا نموذجين لأحدهما متغير مستقل X والنموذج الأخر له متغير مستقل Z مع متغير معتمد واحد هو Y لكلا النموذجين في هذه الحالة يتم إدخال المتغير X في Block1 والمتغير Z في Block2 .

Method : نوع الطريقة المستخدمة في الانحدار (الطريقة الاعتيادية هي Enter) .

Selection Variable: يستعمل في تحديد التحليل لمجموعة معينة من الحالات التي لها قيمة معينة لمتغير الاختيار (مثلاً اقتصار نموذج الانحدار على الحالات التي تكون فيها قيمة المتغير Observat أكبر من 5) يتم التحديد بواسطة الزر Rule .

Case Lebels : متغير تستخدم قيمه كعناوين لنقاط شكل الأنتشار Scatterplots .

- أنقر زر Statistics فيظهر صندوق حوار Statistics الذي نرتبه كما يلي :



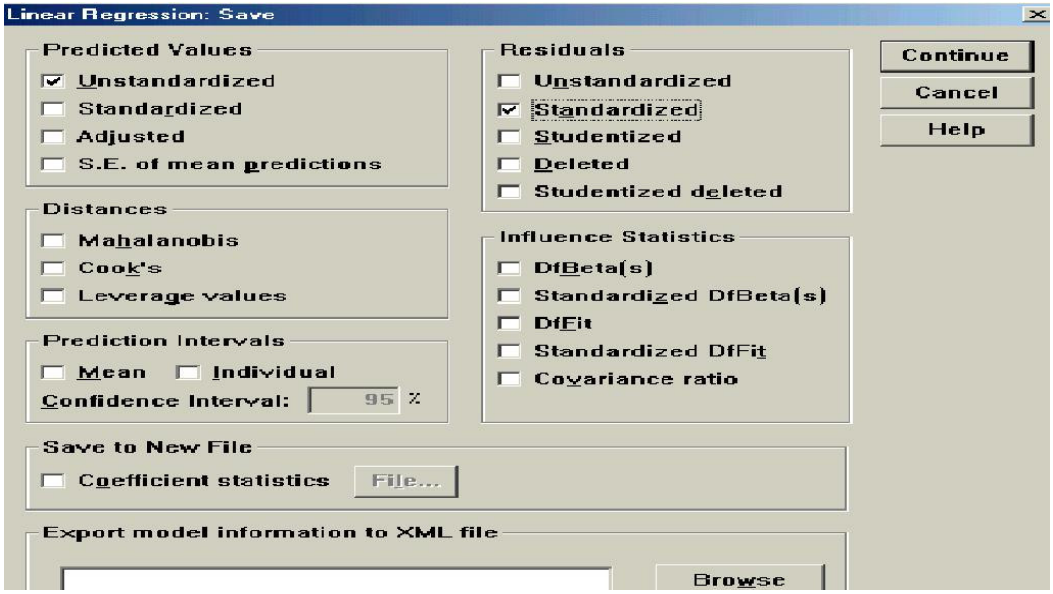
وقد تم تأشير الخيارات التالية :

Estimate :لتقدير معالم نموذج الانحدار واختبارات t المرافقة.

Confidence Interval : لتقدير فترة ثقة 95% لكل من معلمتي الانحدار.

Model Fit :لعرض R^2 و ANOVA .

- أنقر زر Plots فيظهر صندوق حوار Plots أشر الخيـــار Normal Probability Plot لاختبـــار التوزيع الطبيعي للأخطاء العشوائية المعيارية (المطلوب الخامس) .
- أنقر زر Save فيظهر صندوق حوار Save الذي نرتبه كما يلي :



لاحـــظ أننـــا اخترنـــا Unstandardized Predicted Values أي $\hat{y}$ وكـــذلك Standardized Residuals أي e_s سيتم استعمال هذين المتغيرين لرسم Scatterplots للشطر الثاني من المطلوب الرابـــع ،علماً أنه سيتم إضافة هذين المتغيرين(عند تمشية البرنامج) الى ورقة Data Editor الى جانب متغيرات X و Y و Observat حيث يضاف المتغير Unstandardized Predicted Values بأسم Pre_1 ويضـــاف متغير Standardized Residuals بأسم Zre_1 .

يمكن الاختيار بين تضمين الحد الثابت في النموذج او حذفه من خلال خيارات الـــزر Options فـــي صـــندوق حـــوار Linear Regression.

- عند نقر زر OK في صندوق حوار Linear Regression نحصل على النتائج التالية :

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	85.044	9.970		8.530	.000027	62.052	108.036
	X	1.140	.195	.900	5.846	.00038	.690	1.589

a. Dependent Variable: Y

من خلال الجدول أعلاه يمكن كتابة النموذج كما يلي :

$$\hat{y} = 85.044 + 1.140x$$
$$(9.97) \quad (0.195)$$

أن معلمة الميل تشير الى أن زيادة العمر سنة واحدة يؤدي الى زيادة ضغط الدم بمقـــدار 1.140 ملم زئبق .الأرقام داخل الأقواس تمثل الخطأ المعياري للمعلمة المقابلة .

يستعمل اختبار T لاختبار الفرضية التالية لمعلمة الميل B_1 :

فرضية العدم $H_0 : B_1 = 0$

الفرضية البديلة $H_1 : B_1 \neq 0$

كما يستعمل اختبار T لاختبار الفرضية التالية لمعلمة التقاطع (الحد الثابت) B_0 :

فرضية العدم $H_0 : B0 = 0$

الفرضية البديلة $H_1 : B_0 \neq 0$

نستخدم قيمة P-Value المرافقة لإحصائية T للمعلمة في الاختبار كما يلي :

أذا كانت P-Value < 0.05 نرفض فرضية العدم بمستوى دلالة 5% .

أذا كانت P-Value < 0.01 نرفض فرضية العدم بمستوى دلالة 1% .

عكس هذا نقبل فرضية العدم .

أن P-value لمعلمة الميل تساوي0.00038 وهي أقل مـــن 0.01 و أن P-value لمعلمـــة الحـــد الثابت تساوي0.000027 وهي أقل من 0.01 ولهذا نرفض فرضيتا العدم لكل مـــن المعلمتـــين أي أن كـــلا المعلمتين تختلف جوهرياً عن الصفر أن ظهور معلمة الميل معنوية يعكس أهمية متغير العمر في النموذج .

أما معلمة Standardized Coefficient ويشار لها بـــ Beta فهي معلمة الميل للنمـــوذج المقـــدر باستعمال القيم المعيارية $(X-\overline{X})/S$ لكل من المتغير المستقل والمعتمد بدل القيم الأصلية ولا يحتوي هذا النموذج على معلمة تقاطع B_0 وكما يلي $\hat{y}^* = \beta x^*$ حيث أن $\hat{y}^*$ و x^* متغيرات معيارية .

يمكن كتابة فترة ثقة 95% للحد الثابت كما يلي :

$$\Pr(62.052 \le B_0 \le 108.036) = 95\%$$

حيث أن Pr بمثل الاحتمال كما يمكن كتابة فترة ثقة 95% لمعلمة الميل كما يلي :

$$\Pr(.690 \le B_0 \le 1.589) = 95\%$$

الجدول التالي يعرف بجدول تحليل التباين ANOVA ويشتمل علـــى إحصـــائية F لاختبـــار نفـــس الفرضية الخاصة بمعلمة الميل B_1 وهذا الاختبار مكـــافئ تمامـــاً لاختبـــار T لمعلمـــة الميـــل (لاحـــظ أن قيمة P-Value متساوية لكلا الأختبارين) علماً أن $F=T^2$.

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2661.050	1	2661.050	34.174	.00038[a]
	Residual	622.950	8	77.869		
	Total	3284.000	9			

a. Predictors: (Constant), X

b. Dependent Variable: Y

الجدول التالي يتضمن أهم مؤشـــر لنمـــوذج الانحـــدار وهـــو معامـــل التحديـــد Coefficient Of Determination ويرمز له R^2 ويعتبر مقياساً لجودة توفيق النموذج .ويحتسب من جدول تحليل التباين كما يلي :

$$R^2 = \frac{ExplainedVariations}{Total\ \text{Variations}} = \frac{SSR}{SST} = \frac{2661.05}{3284} = 0.81 \qquad 0 \le R^2 \le 1$$

وتفسير ذلك أن 81% من التباينات (الانحرافات الكلية في قيم المتغير Y) تفسرها العلاقة الخطية أي نموذج الانحدار وأن 19% من التباينات ترجع الى عوامل عشوائية كأن تكون هناك متغيرات مهمة لم تضمن في النموذج .وعلى العموم كلما اقتربت قيمة R^2 من 100% دل ذلك على جودة توفيق النمـــوذج .هـــذا وأن $r = \sqrt{R^2}$ حيث أن r معامل الارتباط الخطي البسيط لبيرسون وأن إشارة r هي نفس إشارة معلمة الميل .

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.900[a]	.810	.787	8.82

a. Predictors: (Constant), X

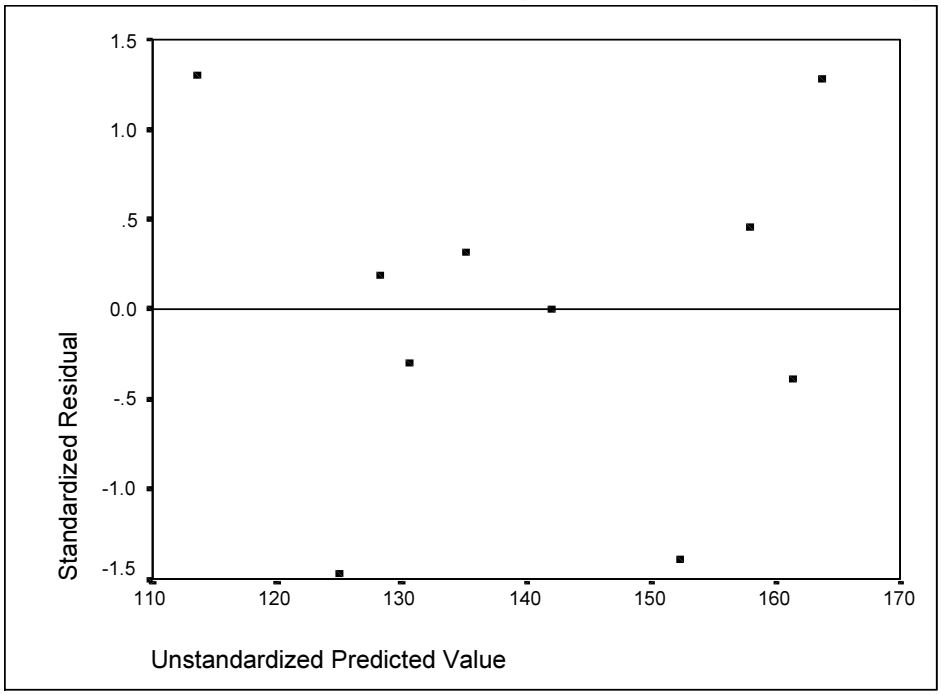
يتصف معامل التحديد بأنه لو أضيف متغير مستقل للنموذج فأن قيمته سترتفع حتى لو لم تكن هنـــاك أهمية للمتغير المستقل في النموذج حيث أن إضافة متغير مستقل الى نموذج الانحدار يؤدي الى زيـــادة R^2

بسبب زيادة مجموع المربعات العائدة للانحدار SSR مع ثبات مجموع المربعات الكلية SST ولهذا يتم احتساب معامل التحديد المصحح Adjusted R Square الذي يأخذ بالاعتبار النقصان الحاصل في درجات الحرية وقيمته دائماً أقل من قيمة معامل التحديد (غير المصحح) في هذا المثال 79% ولهذا يمكن القول أن النموذج جيد التوفيق.

أما الخطأ المعياري للتقدير Standard Error of Estimate فيقيس تشتت القيم المشاهدة عن خط الانحدار وأن الحصول على قيمة صغيرة لهذا المؤشر يعني صغر الأخطاء العشوائية وبالتالي جودة تمثيل خط الانحدار لنقاط شكل الانتشار .

لتحليل الأخطاء العشوائية بيانياً نكون شكل الانتشار Scatterplots بتمثيل القيم التقديرية $\hat{y}$ على المحور الأفقي والأخطاء المعيارية e_S على المحور العمودي وكما يلي (راجع فصل المخططات البيانية) :

- من شريط القوائم اختر Simple Scatter → Graphs → فيظهر صندوق حوار Scatterplots .
- أنقر المتغير Zre_1 لإدخاله الى خانة Y في صندوق حوار Scatter .
- أنقر المتغير Pre_1 لإدخاله الى خانة X .
- أنقر زر OK فيظهر المخطط في شاشة SPSS Viewer .
- أنقر المخطط مرتين للانتقال الى شاشة SPSS Chart Editor .
- من شريط القوائم أختر Reference Line → Chart لإضافة Reference Line الى المحور الرأسي حول الصفر .فيظهر المخطط كما يلي :

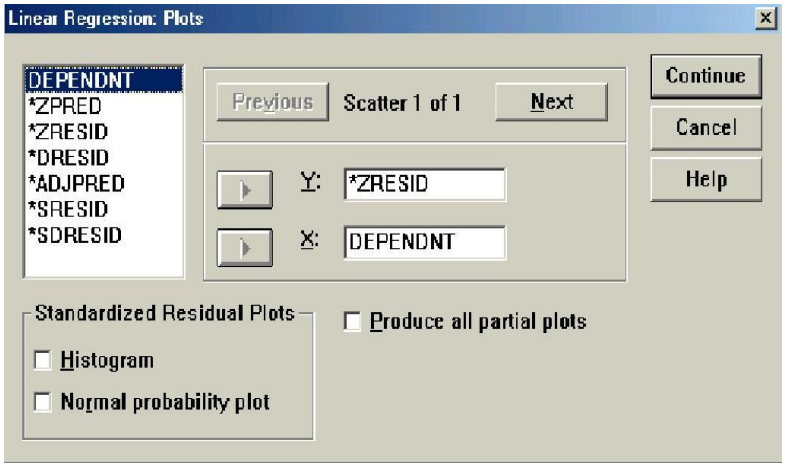


نلاحظ أن النقاط تتوزع بشكل شريط أفقي متساوي حول الصفر مما يدل على توفر فرضيات التحليل بصورة عامة حيث لا يعاني النموذج من مشكلة عدم تجانس تباين الخطأ العشوائي ولا توجد حاجة لاستخدام علاقة من درجات أعلى .

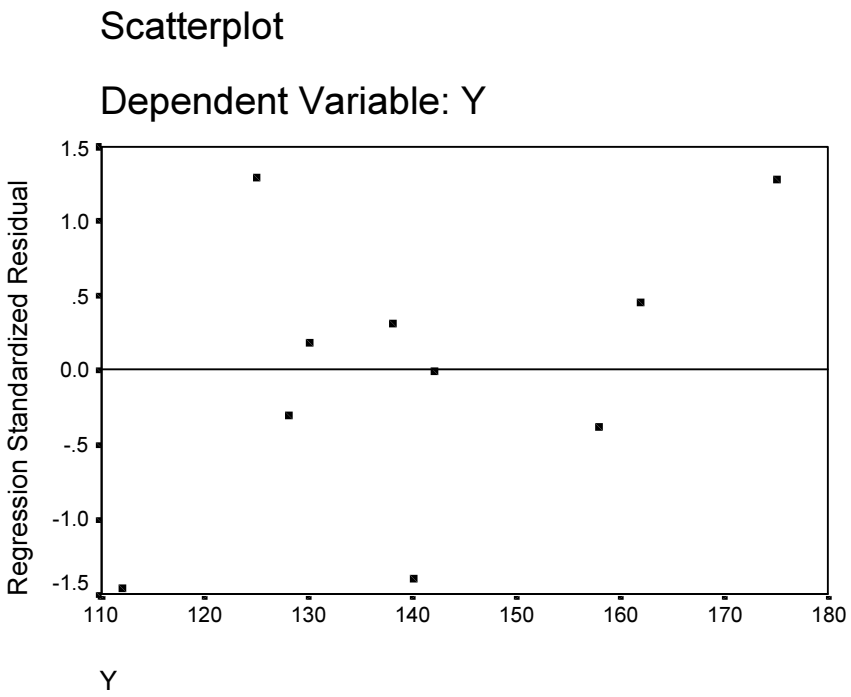
ملاحظة :

يمكن تحليل الأخطاء العشوائية بيانياً بطريقة مشابهة بتمثيل القيم الحقيقية للمتغير المعتمد y على المحور الأفقي والأخطاء المعيارية e_S على المحور العمودي حيث يمكن التوصل الى مخطط مقارب للمخطط السابق وكما يلي :

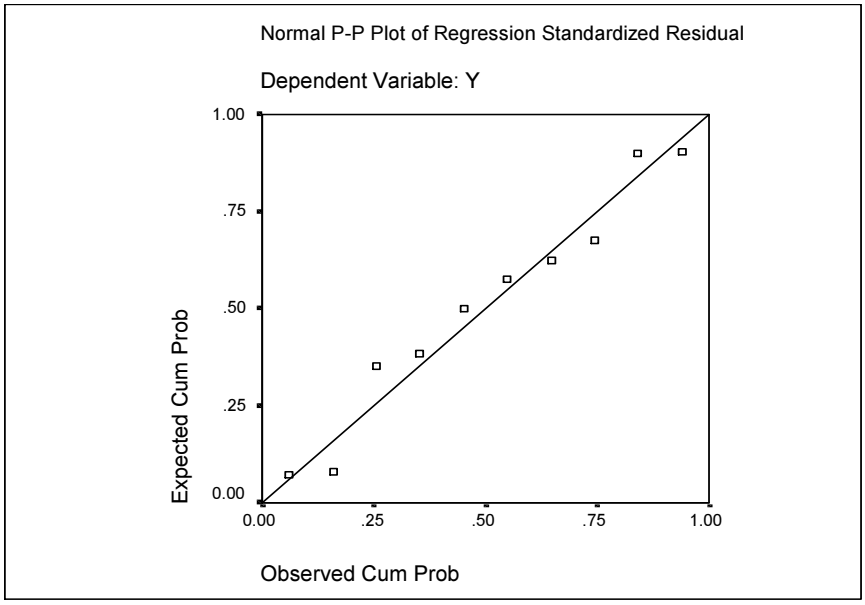
- أنقر زر Plots في صندوق حوار Linear Regression فيظهر صندوق حوار Plots الذي نرتبه كما يلي :



لقد قمنا بنقل المتغير DEPENDENT أي المتغير المعتمد من قائمة المتغيرات في جهة اليسار (هذه المتغيرات تحتسب تلقائياً) الى خانة الإحداثي الأفقي X كما نقلنا متغير الأخطاء (البواقي) المعيارية ZRESID الى خانة الإحداثي الرأسي Y وعند نقر زر Continue في هذا الصندوق وزر OK في صندوق Linear Regression يتم عرض مخطط لشكل الانتشار التالي والذي يقودنا الى نفس الاستنتاج الذي توصلنا اليه من خلال المخطط السابق:



أن اختبار التوزيع الطبيعي للأخطاء العشوائية (المطلوب الخامس) يمكن أن يتم بطريقتين : الأولى من خلال رصد الأخطاء المعيارية فاذا وقعت 95% من الأخطاء ضمن المدى (-2,2) فـأن الأخطـاء تتـوزع طبيعياً ، من المخطط الأخير نلاحظ أن الأخطاء المعيارية لا تتعدى المدى (-1.5,1.5) ومنه نستدل على أن الأخطاء تتوزع طبيعيا .أما الطريقة الثانية فتتمثل في عرض مخطـط Normal probability Plot الـذي أشرناه في صندوق حوار Plot حيث يعرض المخطط التالي عند تنفيذ البرنامج :



نلاحظ أن معظم النقاط تقريباً تتجمع قرب الخط المستقيم وهذا يدل على التوزيع الطبيعي للأخطـاء العشوائية .

(10 – 4 – 2) *طريقة المربعات الصغرى الموزونة* *Weighted Least Squares Method*

تستخدم هذه الطريقة في حالة عدم تحقق فرضية تجانس تباين الخطأ العشوائيHomoscedasticity وغالباً ما تظهر هذه المشكلة في بيانات المقطع العرضي Cross-Section Data ونادراً ما تظهر في بيانات السلاسل الزمنية .

فقد لوحظ مثلاً في بحوث ميزانية الأسرة أن تباين الاستهلاك C يتزايد مع تزايد الـدخل المتـاح Y وعليه فأن تباين الخطأ لا يكون ثابتاً وفي الغالب يكون متناسباً مع مربع الدخل Y حيث تكون صيغة تبـاين الخطأ العشوائي كما يلي $\operatorname{var} e = \sigma^2 Y^2$ فاذا اعتبرنا المقدار $W = 1/y^2$ يمثل الوزن Weight فاذا كان نموذج الانحدار يأخذ الصيغة التالية :

$$C = \beta_0 + \beta_1 Y + e$$

يكون بالإمكان تثبيت تباين الخطأ العشـوائي بتـرجيح طرفـي المعادلـة بالمقـدار $\sqrt{W} = 1/Y$ حيث نقوم بتقدير النموذج التالي :

$$\frac{C}{Y} = \frac{B_0}{Y} + B_1 + \frac{e}{Y}$$

أن تبـاين حـد الخطـأ العشـوائي للنمـوذج الأخيـر يحتسـب كمـا يلـي $\operatorname{var}\frac{e}{Y} = \frac{1}{Y^2}\operatorname{var} e = \frac{1}{Y^2}\sigma^2 Y^2 = \sigma^2$ أي اننا حصلنا على تباين ثابت للخطأ العشـوائي .ويجـب

ملاحظة أن $B1$ قد أصبحت حداً ثابتاً في النموذج الجديد وأن B_0 قد أصبحت معلمة الميل .مع العلـــم أن

برنامج SPSS يقوم بحساب المعالم من النموذج الأخير ثم تستخدم هذه المعالم في النموذج الأصلي لاستخراج البواقي وجدول تحليل التباين وفي حساب R^2 , SEE وغيرها من التقديرات . أن هذه الطريقـــة تعـــرف أيضاً بطريقة المربعات الصغرى المعممة GLS .

<u>مثال 4 : (على طريقة المربعات الصغرى الموزونة)</u>

الجدول التالي يمثل الاستهلاك c والدخل المتاح y لعدد من الأسر عددها 30 أسرة [2].وقد تـــم أدخـــال المتغيرين المذكورين الى شاشة Data editor لبرنامج SPSS كما يلي (المتغير w أحتسب فيما بعد) :

c	y	w
10600	12000	6.9444E-09
10800	12000	6.9444E-09
11100	12000	6.9444E-09
11400	13000	5.9172E-09
11700	13000	5.9172E-09
12100	13000	5.9172E-09
12300	14000	5.1020E-09
12600	14000	5.1020E-09
13200	14000	5.1020E-09
13000	15000	4.4444E-09
13300	15000	4.4444E-09
13600	15000	4.4444E-09
13800	16000	3.9063E-09
14000	16000	3.9063E-09
14200	16000	3.9063E-09
14400	17000	3.4602E-09
14900	17000	3.4602E-09
15300	17000	3.4602E-09
15000	18000	3.0864E-09
15700	18000	3.0864E-09
16400	18000	3.0864E-09
15900	19000	2.7701E-09
16500	19000	2.7701E-09
16900	19000	2.7701E-09
16900	20000	2.5000E-09
17500	20000	2.5000E-09
18100	20000	2.5000E-09
17200	21000	2.2676E-09
17800	21000	2.2676E-09
18500	21000	2.2676E-09

يطلب ما يلي :

1.استخراج معادلة انحدار C على Y بطريقة المربعات الصغرى الاعتيادية OLS ثم اختبار تجانس تبـــاين الخطأ العشوائي من الرسم البياني .

2. بافتراض أن تباين الخطأ العشوائي غير متجانس ويرتبط مع Y بالعلاقة التالية $\operatorname{var} e = \sigma^2 Y^2$ أستخدم طريقة المربعات الصغرى الموزونة في تقدير نموذج الانحدار .

[2] دومنيك سالفاتور ، الإحصاء والاقتصاد القياسي ، ملخصات شوم، دار ماكروهيل ،1982 ، ص . 217 .

الحل :

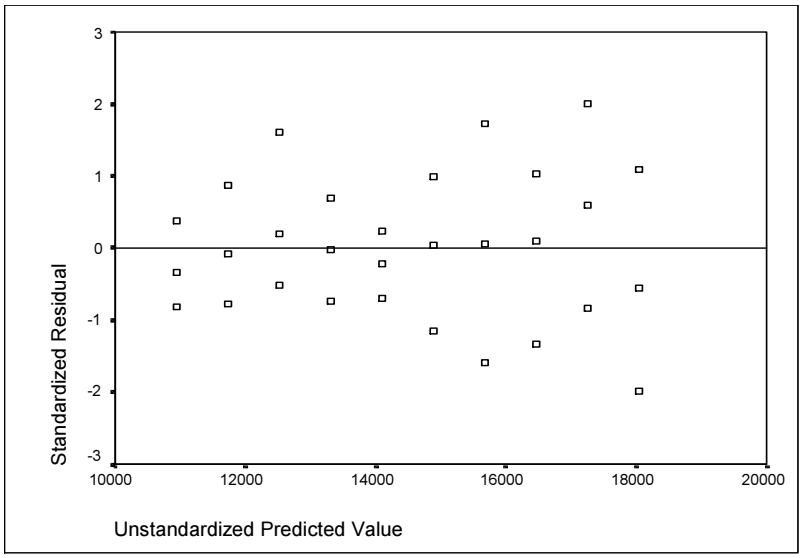
1. يمكن تقدير نموذج الانحدار الخطي البسيط بطريقة OLS بنفس الخطوات المتبعة في المثال السابق وقد تم التوصل الى النموذج التالي :

$$R^2 = 0.97 \qquad \hat{C} = 1408.0 + 0.788Y_d$$
$$(449.6) \quad (0.27)$$

الأرقام داخل الأقواس تمثل الخطأ المعياري للمعالم .لاختبار وجود مشكلة عدم تجانس تباين الخطأ العشوائي بيانياً نقوم بتمثيل القيم التنبؤية للمتغير المعتمد $\hat{C}$ على المحور الأفقي لشكل الانتشار والأخطاء المعيارية على المحور العمودي Standardized Residuals ، لتنفيذ ذلك نتبع الخطوات التالية:

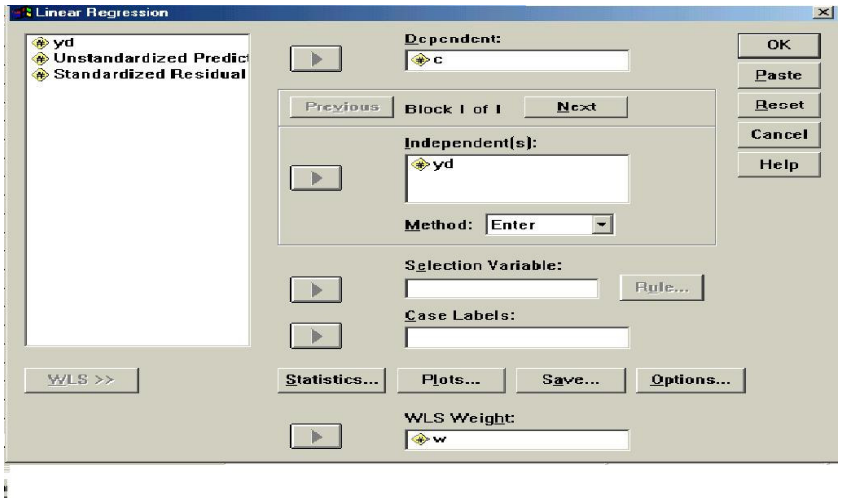
- أختر Analyze → Regression → Linear من شريط القوائم لفتح صندوق Linear Regression ثم أنقر زر Save في هذا الصندوق لفتح صندوق Save في هذا الصندوق أشر الخيار Unstandardized في الحقل Predicted Values لكي يعرض البرنامج القيم التقديرية للمتغير المعتمد $\hat{C}$ في شاشة Data editor (تظهر في الشاشة بأسم Pre-1) ثم أشر الخيار Standardized في الحقل Residuals لكي يعرض البرنامج الأخطاء المعيارية (تظهر بأسم Zre-1) .
- من شريط القوائم اختر Graphs→ Scatter → Simple فيظهر صندوق حوار ScatterPlots يتم ترتيب هذا الصندوق كما يلي :
  - أنقل المتغير Zre-1 الى حقل Y-axis .
  - أنقل المتغير Pre-1 الى حقل X-Axis .
  - أنقر زر OK .

فيظهر المخطط التالي بعد إضافة Reference Line إليه وكما يلي :



نلاحظ أن نقاط شكل الانتشار لا تتوزع بانتظام حول الصفر وأن هناك اتجاها عاماً في زيادة تباين الخطأ العشوائي كلما زادت $\hat{Y}$ مما يشير الى وجود مشكلة عدم تجانس تباين الخطأ ويتوجب معالجتها .

2. بافتراض أن تباين الخطأ يرتبط مع المتغير المستقل Y بالعلاقة التالية $\operatorname{var} e = \sigma^2 Y^2$ فلكي نطبق طريقة المربعات الصغرى الموزونة نحدد الوزن وهو مقلوب التباين $W = 1/y^2$ ونقوم بحساب هذا المتغير بالأمر Compute → Transform وإضافته الى شاشة Data Editor كما في الجدول أعلاه ثم نطبق نفس خطوات أجراء الانحدار الخطي حيث يظهر صندوق حوار Linear Regression بعد ترتيبه كما يلي :



لاحظ أننا قمنا بإدخال متغير الوزن W في خانة WLS Weight بعد نقر زر WLS .

- عند نقر زر OK يظهر المخرج التالي :

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1421.278	395.496		3.594	.001
	YD	.792	.025	.986	31.511	.000

a. Dependent Variable: C

b. Weighted Least Squares Regression - Weighted by W

ويمكن كتابة المعادلة المقدرة بطريقة المربعات الصغرى الموزونة كما يلي :

$$\hat{C} = 1421.278 + 0.792 Y_d \qquad R^2 = 0.97$$
$$(395.496) \quad (0.25)$$

حاول أن ترسم شكل الانتشار لعلاقة القيم التقديرية بالأخطاء المعيارية للتأكد من أن البيانات تتوزع بشكل شريط أفقي حول الصفر وبالتالي ثبات تباين الخطأ العشوائي.

(10 -4 -3) نموذج الأنحدار الخطي المتعدد

يأخذ النموذج الخطي العام الصيغة التالية :

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + e$$

حيث أن

β_0 : يمثل الحد الثابت .

$\beta_1, \beta_2, ..., \beta_k$: معاملات الانحدار الجزئية Partial regression Coefficients أو الميول الجزئية .

e : الخطأ العشوائي .

حيث يكون عدد معالم النموذج الخطي العام هو P = K+1 (K يمثل عدد المتغيرات المستقلة في النموذج) .

أن فرضيات النموذج الخطي المتعدد هي نفسها فرضيات النموذج البسيط يضاف الى ذلك فرضية عدم وجود ارتباط خطي متعدد بين المتغيرات المستقلة (Multicollinearity) .

مثال 5 :

الجدول التالي يتضمن دخل الفرد y (ألف دولار) مع نسبة القوة العاملة في الزراعة x_1 (نسبة مئوية) ومتوسط سنوات التعليم للسكان x_2 لـ 15 دولة في سنة 1981[3].وقد أدخلت البيانات في شاشة Data Editor كما في الجدول التالي :

y	x_1	x_2
6	9	8
8	10	13
8	8	11
7	7	10
7	10	12
12	4	16
9	5	10
8	5	10
9	6	12
10	8	14
10	7	12
11	4	16
9	9	14
10	5	10
11	8	12

المطلوب :

1. حساب معادلة انحدار yعلى x_1 وx_2 وتفسير النتائج .
2. تكوين جدول تحليل التباين ANOVA واختبار معنوية المعاملات .
3. اختبار مشكلة الارتباط الذاتي المتسلسل باستخدام إحصائية DW .

لتنفيذ المطاليب المذكورة نتبع الخطوات التالية :

- من شريط القوائم اختر Analyze → Regression → Linear فيظهر صندوق حوار Linear Regression الذي يرتب كالتالي :
  - أنقر المتغير Y وادخله الى خانة Dependent .
  - أشر المتغيرين X1,X2 وانقلهما الى خانة Independent .
  - في خانة Method تأكد أن نوع طريقة الانحدار هي الطريقة الاعتيادية Enter .

[3] دومينيك سالفاتور ، الإحصاء والاقتصاد القياسي ، دار ماكروهيل ،1982 ، ص 172 .

➢ أنقر Statistics في صندوق حوار Linear Regression فيظهر صندوق حوار Statistics حيث نقوم بتأشير كل مما يلي :

Estimate : لتقدير معالم النموذج والإحصاءات المرافقة .

Model Fit : لتقدير R^2 و ANOVA.

Durbin - Watson : لحساب إحصائية DW .

➢ عند نقر زر OK في صندوق حوار Linear regression يتم عرض النتائج التالية :

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.833[a]	.693	.642	1.01	.946

a. Predictors: (Constant), X2, X1

b. Dependent Variable: Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6.203	1.862		3.331	.006
	X1	-.376	.133	-.461	-2.834	.015
	X2	.453	.120	.615	3.786	.003

a. Dependent Variable: Y

يمكن معادلة كتابة نموذج الانحدار المتعدد كما يلي :

$$\hat{y} = 6.203 - 0.376X_1 + 0.453X_2 \qquad \overline{R}^2 = 0.64$$
$$(1.862)\ (0.133) \quad (0.120)$$

أن معلمة المتغير X1 تشير الى أن زيادة نسبة القوى العاملة في الزراعة بمقدار 1% من أجمالي القوى العاملة يؤدي الى نقصان في دخل الفرد بمقدار 376 دولار بافتراض ثبات المتغير X_2 كما أن زيادة متوسط سنوات التعليم سنة واحدة يؤدي الى زيادة دخل الفرد بمقدار 453 دولار بافتراض ثبات المتغير X_1 ، كما أن قيمة $\overline{R}^2$ تشير إلى جودة توفيق متوسطة للنموذج .

أن الغرض من حساب جدول تحليل التباين هو تحليل مجموع مربعات الانحرافات الكلية SST لقيم المتغير المعتمد $\sum(y-\overline{y})^2$ الى مجموع المربعات العائدة للانحدار SSR ومجموع مربعات الخطأ SSE .كما يتم احتساب إحصائية F التي يستفاد منها في اختبار الفرضية التالية :

$$H_0 : \beta_1 = \beta_2 = 0$$

$$H_1 : \beta_1 \neq \beta_2 \neq 0$$

لاحظ أن اختبار F لا يشمل معلمة التقاطع B_0 .لقد تم الحصول على جدول تحليل التباين التالي :

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	(SSR)27.728	(p-1) 2	(MSR) 13.864	13.557	.001[a]
	Residual	(SSE)12.272	(n-p) 12	(MSE) 1.023	F =MSR/MSE	
	Total	(SST)40	(n-1) 14			

a. Predictors: (Constant), X2, X1

b. Dependent Variable: Y

حيث أن P = 3 ويمثل عدد المعالم و n يمثل حجم العينة .أن قيمة P-Value= 0.001< 0.05 تدعونا الى رفض فرضية العدم بمستوى دلالة 5% أي أن الانحدار معنوي أو ان المتغيرين المستقلين مجتمعين لهما تأثير معنوي على الانحدار أو أن واحدة على الأقل من معلمتي الانحدار B_1 وB_2 تختلف معنوياً عن الصفر .

لمعرفة تأثير كل متغير مستقل على المتغير المعتمد بصورة انفرادية نلجأ الى اختبار t (جدول Coefficients في الأعلى) ومنه يتضح معنوية معلمتي الميل لكل من X_1 وX2 بمستوى دلالة 5% وعليه فكلا المتغيرين مؤثرين ويوصى بإبقائهما في نموذج الانحدار .

عند اختبار وجود الارتباط الذاتي للأخطاء العشوائية بمستوى دلالة 5% وبدرجة حرية n =15 و p = 3 فان إحصائية DW المحسوبة تشير الى وجود ارتباط ذاتي موجب في الأخطاء العشوائية .

<u>مثال 6: (اختبار وجود مشكلة التعدد الخطي / مع اختيار أفضل نموذج انحدار)</u>

الجدول التالي[4] يتضمن المتغير المعتمد Y والمتغيرات المستقلة X_1,X_2,X_3 والتي أدخلت في ورقة Data Editor كما يلي :

Y	X1	X2	X3	X4
43	5	3	18	12
63	9	5	27	9
71	10	7	34	11
61	8	4	24	10
81	11	6	33	6
44	12	5	22	8
58	9	4	28	9
71	7	7	32	7
72	8	5	23	8
67	13	8	20	5
64	4	5	21	4
69	10	9	36	10
68	11	10	30	11

1. أوجد معادلة الانحدار الخطي المتعدد .
2. أختبر وجود مشكلة التعدد الخطي Multicollinearity بين المتغيرات المستقلة .
3. أوجد أفضل معادلة انحدار بأسلوب Stepwise Regression .
1. لتنفيذ المطلوبين الأول والثاني من المثال نتبع الخطوات التالية :

[4] د.أموري هادي كاظم ، ومحمد مناجد الدليمي ، مقدمة في تحليل الانحدار الخطي،جامعة بغداد ،1988،ص277.

- من شريط القوائم اختر Analyze → Regression → Linear فيظهر صـــندوق حـــوار Linear Regression الذي يرتب كالتالي :
  - أشر المتغير Y وادخله الى خانة Dependent .
  - أشر المتغيرات X1,X2,X3 وانقلها الى خانة Independent .
  - في خانة Method تأكد أن نوع طريقة الانحدار هي الطريقة الاعتيادية Enter .
- أنقر زر Statistics في صندوق حوار Linear Regression فيظهر صندوق حـــوار Statistics حيث نقوم بتأشير كل مما يلي

Estimate : لتقدير معالم النموذج والإحصاءات المرافقة .

Model Fit : لتقدير R^2 و ANOVA.

Part & Partial Correlations : لتقدير معاملات الارتباط .

Collinearity Diagnostics : لتشخيص مشكلة التعدد الخطي .

- عند نقر زر OK في صندوق حوار Linear regression يتم عرض النتائج التالية :

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.810[a]	.656	.484	7.72

a. Predictors: (Constant), X4, X1, X3, X2

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	47.06	13.0		3.611	.007					
	X1	-.430	1.010	-.104	-.425	.682	.210	-.149	-.088	.713	1.403
	X2	1.199	1.486	.232	.807	.443	.540	.274	.167	.520	1.925
	X3	1.153	.476	.631	2.423	.042	.634	.651	.502	.633	1.580
	X4	-2.038	.950	-.462	-2.1	.064	-.331	-.604	-.445	.928	1.077

a. Dependent Variable: Y

يمكن كتابة معادلة نموذج الانحدار المتعدد كما يلي :

$$\hat{y} = 47.06 - 0.43X_1 + 1.199X_2 + 1.153X_3 - 2.038X_4 \qquad \bar{R}^2 = 0.48$$

$$(13) \quad (1.01) \quad (1.486) \quad (0.476) \quad (0.950)$$

أن معلمة المتغير X1 تشير الى أن زيادة قدرها وحدة واحدة في قيم المتغير ينشـــأ عنهـــا نقصـــان المتغير المعتمد بمقدار 0.430 وحدة بافتراض ثبات المتغيرين X_2 و X_3. وبنفس الطريقة يتم تفسير بقيـــة معالم النموذج .

يلاحظ ومن خلال قيمة P-value المرافقة لإحصائية t أن معلمة المتغير X_3 فقط قد ظهرت معنوية بمستوى دلالة 5% (بالإضافة الى معلمة الحد الثابت) .كما أن قيمة معامل التحديد المصحح المنخفضة نسبياً تشير الى أن النموذج لا يعبر عن العلاقة الخطية بصورة جيدة .

لقد تم أستخراج ثلاثة أنواع من الارتباطات وكما يلي :

Zero-order Correlation : معامل الارتباط البسيط لبيرسون بين المتغير المستقل والمتغير المعتمد .

Partial Correlation : معامل الارتباط الجزئي بين المتغير المعتمد والمتغير المستقل (بثبات المتغيرات المستقلة الأخرى) .

Part Correlation: معامل الارتباط الجزء بين المتغير المعتمد والمتغير المستقل باستبعاد أثر المتغيرات المستقلة عن المتغير المستقل فقط .

أن نموذج الانحدار المتعدد يعتمد الارتباطات الجزئية بين كل من المتغيرات المستقلة والمتغير المعتمد ونلاحظ أن X_3 له أعلى معامل ارتباط جزئي مع المتغير المعتمد وعليه يكون له أعلى قيمة لإحصائية t يليه X_4 في حين يعتمد نموذج الانحدار البسيط معامل الارتباط البسيط بين المتغير المستقل والمتغير المعتمد .

لغرض تشخيص مشكلة الارتباط الخطي المتعدد يتم في البداية حساب المعامل Tolerance لكل من المتغيرات المستقلة حيث ان Tolerance = 1- $R^2_{X_i.others}$ حيث أن $R^2_{X_i.others}$ يمثل مربع معامل الارتباط المتعدد بين المتغير المستقل i وبقية المتغيرات المستقلة .ثم يستخرج معامل VIF لكل متغير مستقل (Variance Inflation Factor) حيث ان $VIF = \frac{1}{Tolerance}$. يعتبر هذا المعامل مقياساً لتأثير الارتباط بين المتغيرات المستقلة على زيادة تباين معلمة المتغير المستقل (تتميز مشكلة التعدد الخطي بارتفاع تباين معالم النموذج وبالتالي عدم ظهور المعلمة معنوية نتيجة انخفاض قيمة إحصائية t بالرغم من أن المتغير قد يكون مهماً في النموذج) . أن الحصول على قيمة لمعامل VIF لأحد المتغيرات المستقلة تزيد عن 5 أو 10 تشير الى أن تقدير المعلمة المرافقة يتأثر بمشكلة التعدد الخطي ,أن قيمة VIF لمتغيرات النموذج تشير الى عدم تأثر أي منها بمشكلة التعدد الخطي ,

تستخدم الجذور المميزة لمصفوفة $X'X$ في تشخيص مشكلة التعدد الخطي (X هي مصفوفة المتغيرات المستقلة) ففي حالة وجود عدة جذور مميزة قريبة من الصفر فهذا دليل على مشكلة التعدد الخطي .لاختبار وجود ارتباط خطي بين المتغيرات المستقلة يستعمل ما يعرف بدليل الحالة Condition Index وهو عبارة عن الجذر التربيعي لحاصل قسمة أكبر جذر مميز على كل من الجذور المميزة مثلاً يحتسب هذا الدليل للحد الثابت كما يلي :

$$Condition\ \text{Index} = \sqrt{\frac{4.813}{4.813}} = 1$$

وبالنسبة للمتغير X_1 يحتسب كما يلي :

$$condition\ \text{Index} = \sqrt{\frac{4.813}{0.09868}} = 7.020$$

فاذا زادت قيمة الدليل عن 15 فهذا مؤشر على إمكانية وجود مشكلة التعدد الخطي .أما إذا زادت عن 30 فهذا مؤشر على خطورة المشكلة .في الجدول التالي نلاحظ أن أكبر قيمة للدليل هي 16 أي إمكانية وجود المشكلة .

المقياس الأخر للمشكلة هو ما يعرف بـــ Variance Proportion وهو يمثل نســبة تبــاين التقــدير المفسر بواسطة المكون الأساسي Principle Component المرافق لكل جذر مميز حيــث تعتبــر مشــكلة التعدد الخطي مؤثرة إذا كان المكون الأساسي المرافق لدليل حالة Condition Index مرتفع نسبياً يســاهم بصورة أساسية في تباين أثنين أو أكثر من المتغيرات المستقلة .في هذا المثال المكــون الأساســي الخــامس يساهم بصورة كبيرة في تباين المتغير X_3 فقط ولهذا لا تعتبر مشكلة التعدد الخطي مؤثرة بشكل كبير علــى البيانات وكما هو واضح في الجدول التالي :

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions (Constant)	X1	X2	X3	X4
1	1	4.813	1.000	.00	.00	.00	.00	.00
	2	9.768E-02	7.020	.01	.06	.17	.00	.35
	3	4.345E-02	10.525	.01	.71	.29	.08	.01
	4	2.907E-02	12.868	.30	.05	.26	.22	.63
	5	1.669E-02	16.983	.68	.18	.28	.70	.01

a. Dependent Variable: Y

لتنفيذ لمطلوب الثالث من المثال نتبع الخطوات الاعتيادية للانحدار المتعدد وكما يلي :

- من شريط القوائم اختر Analyze → Regression → Linear فيظهر صندوق حــوار Linear Regression الذي يرتب كالتالي :
  - أشر المتغير Y وادخله الى خانة Dependent .
  - أشر المتغيرات X1,X2,X3 وانقلها الى خانة Independent .
  - في خانة Method أنقر السهم المتجه للأسفل فتظهر الخيارات التالية لاختيار أفضــل نمــوذج انحدار :

1. Enter :إدخال كافة المتغيرات المستقلة الى نموذج الانحدار (الأسلوب الاعتيادي لتنفيذ الانحدار) .
2. Stepwise : إدخال المتغيرات واحداً بعد الأخر بخطوات متسلســلة إلــى النمــوذج مــع اســتبعاد المتغيرات التي تصبح غير مؤثرة بوجود بقية المتغيرات .
3. Remove : استبعاد المتغيرات غير المهمة بخطوة واحدة من النموذج .
4. Backward :استبعاد المتغيرات غير المؤثرة واحداً بعد الأخر بخطوات متسلسلة.
5. Forward : إدخال المتغيرات واحداً بعد الأخر الى النموذج ولا يتم استبعاد المتغيرات التي تصــبح غير مؤثرة فيما بعد من النموذج .

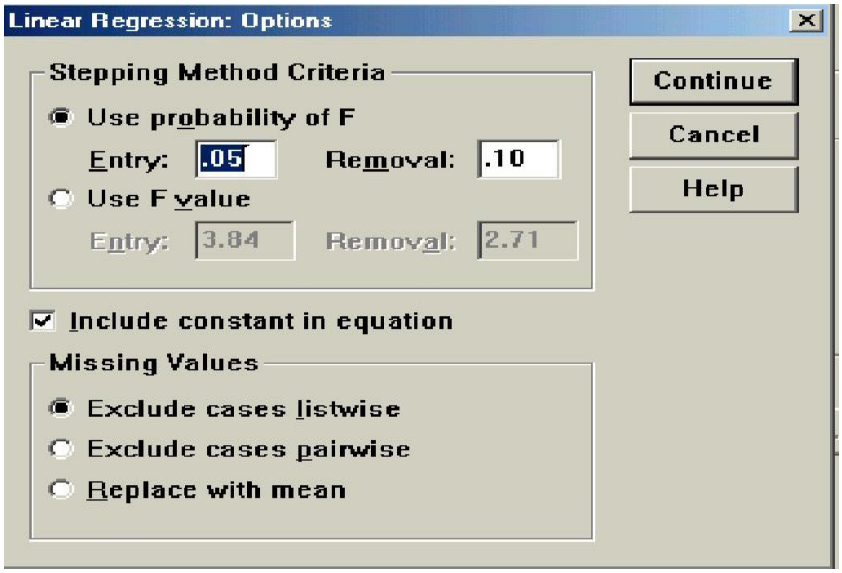
- أختر نوع طريقة الانحدار Stepwise (تعتبر أفضل الطرق) .

➢ أنقر زر Statistics في صندوق حوار Linear Regression فيظهر صندوق حـــوار Statistics حيث نقوم بتأشير كل مما يلي

Estimate : لتقدير معالم النموذج والإحصاءات المرافقة .

Model Fit : لتقدير R^2 و ANOVA .

➢ أنقر زر Options في صندوق حوار Linear Regression فيظهر صندوق حوار Options وكما يلي :



تحتاج طريقة Stepwise الى تحديد مستوى المعنوية أو قيمة F التي يتم بموجبها إدخال واســتبعاد المتغيرات من النموذج .في حقل Stepping Method criteria وذلك بموجب الخيار ين التاليين :

- Use Probability of F :تحديد مستوى المعنوية الذي سيستخدم فـــي كـــل خطـــوة لإدخـــال واستبعاد المتغيرات وان مستوى الدلالة لإدخال المتغيرات يجب أن يكون أقل من مستوى دلالة اســتبعاد المتغيرات أو مساويا له . لتضمين متغيرات اكثر في النموذج زد من قيمة Entry . لاستبعاد عدد أكبـــر من المتغيرات من النموذج قلل من قيمة Removal .
- Use F Value : تحديد قيمة F الذي ستستخدم في كل خطوة لإدخال واستبعاد المتغيـــرات وان قيمة F لإدخال المتغيرات يجب أن تكون أعلى من قيمة F لاستبعاد المتغيرات .لتضمين متغيرات اكثـــر في النموذج قلل من قيمة Entry . لاستبعاد عدد أكبر من المتغيـــرات مـــن النمـــوذج زد مـــن قيمـــة Removal .

أن الاختبار المستخدم في أدخال واستبعاد المتغيرات هو اختبار F الجزئـــي Partial F Test الـــذي يستخدم في أختبار معنوية جزء من معالم النموذج ويستخدم هنا في اختبار معنوية معلمة واحدة فقط لمتغيـــر واحد لاختبار معنوية مجموع المربعات التي يضيفها المتغير المستقل الى النموذج لكي نتوصل الى قرار بشأن استبعاده أو بقائه في النموذج ، علما أن هذا الاختبار F يكون مكافئاً تماماً لاختبار t في حالة استعمال اختبار F لاختبار معنوية معلمة واحدة فقط .

لقد اعتمدنا القيم الافتراضـــية للخيـــار الأول Use Probability of F أي 0.05 للإدخـــال و0.10 للاستبعاد .

➢ عند نقر زر OK في صندوق حوار Linear regression يتم عرض النتائج التالية :

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	X3	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	X4	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: Y

بموجب طريقة Stepwise يتم أخال المتغيرات الأربعة واحداً بعد الأخر الى النموذج علماً أن المتغير الداخل عرضة للاستبعاد في الخطوات اللاحقة إذا ثبتت عدم معنويته بوجود المتغيرات الأخرى .

لقد كان X_3 أول المتغيرات الداخلة الى النموذج لأن له أكبر معامل ارتباط بسيط مع المتغير المعتمد وبالتالي أكبر قيمة لإحصائية t . من الجدول اللاحق Coefficients نلاحظ أن قيمة P-Value المرافقة لإحصائية t تساوي 0.02 وهي أقل من 0.05 (مستوى الدلالة للإدخال Entry) ولهذا يسمح بإدخال X_3 الى النموذج (لاحظ أن أختبار t الذي أستعملناه هنا مكافئ تماماً لأختبار F الجزئي) وعليه يصبح النموذج في الخطوة الأولى كما يلي :

$$\hat{y} = 33.007 + 1.158X_3$$

في الخطوة الثانية يتم إدخال المتغير الذي له أعلى معامل ارتباط جزئي مع المتغير المعتمد بثبات المتغير X_3 وهو المتغير X_4 ولكن يجب أولاً التأكد من معنوية المتغير بحساب إحصائية T له (من الجدول اللاحق P-Value المرافقة لإحصائية t تساوي 0.033 وهي أقل من 0.05(مستوى الدلالة للإدخال Entry) ولهذا يسمح بإدخال X_4 الى النموذج ليصبح على الشكل التالي :

$$\hat{y} = 46.152 + 1.345X_3 - 2.147X_4$$

في النموذج أعلاه نجد القيمة الأقل لإحصائية t للمتغيرين X_3 و X_4 أي القيمة الأكبر لـ P-Value المرافقة لإحصائية t وقد كانت للمتغير X_4 و تساوي 0.033 وهي أقل من 0.10 (مستوى الدلالة للاستبعاد Removal) ولهذا يبقى كلا المتغيرين في النموذج .

أن النموذج أعلاه هو النموذج النهائي ويمثل أفضل نموذج انحدار ولا يتضمن المتغيرين X_1 و X_2 حيث لم تظهر معنوية المتغير الذي له أعلى ارتباط بالمتغير المعتمد بمستوى دلالة 0.05 .

الجدول التالي يبين النماذج التي تم اختبارها لحين الوصول الى النموذج النهائي .

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	33.007	11.648		2.834	.016
	X3	1.158	.426	.634	2.720	.020
2	(Constant)	46.152	11.011		4.191	.002
	X3	1.345	.360	.737	3.735	.004
	X4	-2.147	.871	-.486	-2.466	.033

a. Dependent Variable: Y

(1) النموذج الأول (2) النموذج الثاني (النهائي)

ملاحظة : في حالة اختيار Use F Value في Stepping Method Criteria فأن المقارنة تـتم بـين t^2 والقيمة المحددة لإحصائية F للإدخال والاستبعاد فعند إدخال متغير إذا كانت t^2 >F (Enter) نسمح بـدخول المتغير الى النموذج .عند إخراج متغير إذا كانت t^2 >F(Remove) نسمح ببقاء المتغير في النموذج .

الجدول التالي يبين المتغيرات التي تم استبعادها في كل من النموذجين الأول والثاني :

Excluded Variables[c]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	X1	.027[a]	.108	.916	.034	.915
	X2	.267[a]	.941	.369	.285	.682
	X4	-.486[a]	-2.466	.033	-.615	.955
2	X1	-.012[b]	-.058	.955	-.019	.909
	X2	.175[b]	.721	.489	.234	.663

a. Predictors in the Model: (Constant), X3

b. Predictors in the Model: (Constant), X3, X4

c. Dependent Variable: Y

حيث أن Beta in تمثل المعلمة المعيارية للمتغير فيما لو أدخل الى النموذج في الخطوة اللاحقة .

الفصل الحادي عشر

التحليل العاملي

Factor Analysis

(11 –1) التحليل العاملي

تهدف طرق التحليل العاملي الى أيجاد مجموعة من العوامل Factors التي تكون مسؤولة عن توليد الاختلافات Variations في مجموعة مكونة من عدد كبير من متغيرات الاستجابة Response Variables حيث يمكن التعبير عن المتغيرات المشاهدة كدالة في عدد من العوامل المستترة Factors وغالباً ما يعبر عن متغيرات الاستجابة كتركيب خطي Linear Compounds من العوامل المستترة .حيث تكون العلاقة بين المتغيرات داخل العامل الواحد أقوى من العلاقة مع المتغيرات في عوامل أخرى .

أن التحليل العاملي يساعد على فهم تركيب مصفوفة الارتباط أو التباين المشترك من خلال عدد قليل من العوامل .

(11–2) طريقة المكونات الأساسية Principal Components Method

أن طريقة المكونات الأساسية هي واحدة من أهم طرق التحليل العاملي وتأتي في مقدمة الطرق لبساطتها .

أن المكون الأساسي (أو العامل) هو عبارة عن تركيب خطي من متغيرات الاستجابة Response Variables .باعتبار أن لدينا p من متغيرات الاستجابة فأن المكون الأساسي الأول يعبر عنه كما يلي :

$$Z_1 = a_{11}X_1 + a_{21}X_2 + ... + a_{p1}X_p$$

حيث أن aij تمثل تشبعات Loadings متغيرات الاستجابة بالعامل الأول .أما المكون الأساسي الثاني فيعبر عنه كما يلي :

$$Z_2 = a_{12}X_1 + a_{22}X_2 + ... + a_{p2}X_p$$

أن المكون الأول له أعظم تباين Variance (يفسر أكبر نسبة من هيكل التباينات لمتغيرات الاستجابة) يليه المكون الأساسي الثاني .....وهكذا .وأن هذه المكونات تكون متعامدة فيما بينها Orthogonal ويمكن حساب المكونات بطريقتين :

1. استعمال مصفوفة التباين المشترك Variance-Covariance Matrix لمتغيرات الاستجابة وفي هذه الحالة فأن المتغيرات تكون مقاسه بالانحرافات عن الوسط الحسابي $X - \overline{X}$.

2. استعمال مصفوفة الارتباطات Correlation Matrix لمتغيرات الاستجابة وفي هذه الحالة تستعمل المتغيرات المعياريةStandardized Variables ويكون ذلك ضرورياً في حالة اختلاف وحدات القياس لمتغيرات الاستجابة .

مثال 1 :

البيانات التالية تمثل بعض المتغيرات الاجتماعيـــة لمســـتوى المعيشـــة حســـب المنـــاطق regions (المحافظات) في العراق لسنة 1977 والتي تظهر كما يلي بعد إدخالها في شاشـــة Data Editor لبرنـــامج SPSS :

Region	gdp	literacy	higheduc	doctors	hospbed
Dohok	17.2	30.1	1.09	17	139
Nineveh	24.0	44.2	1.85	14	172
Arbil	22.2	35.2	1.18	13	163
Sulayman	16.2	33.5	1.01	10	115
Ta'meem	32.3	49.4	1.85	18	143
Salah AL-Deen	98.4	37.9	1.32	15	80
Diala	23.1	44.1	1.93	10	153
Anbar	22.7	44.3	1.58	16	144
Baghdad	75.0	61.6	4.04	36	280
Wasit	19.5	36.7	1.11	28	199
Babylon	22.8	44.1	1.82	18	145
Kerbala	21.5	47.7	1.53	24	173
Najaf	18.7	46.2	1.59	27	190
Qadisia	21.0	35.2	.95	9	195
Muthana	21.3	33.5	.84	18	178
Thi-Qar	18.1	33.8	.73	12	144
Maysan	20.4	34.4	.90	11	301
Basrah	19.0	53.6	2.24	25	219

حيث أن المتغير gdp يمثل معدل نصيب الفرد من خدمات التنمية الاجتماعية والمتغير literacy يمثل النسبة المئوية للمتعلمين من الكبار والمتغير higheduc يمثل النسبة المئوية للحاصلين علـــى شـــهادة عليـــا والمتغير doctors يمثل عدد الأطباء لكل 100000 من السكان والمتغيـــر hospbed يمثـــل عـــدد أســـرة المستشفيات لكل 100000 من السكان .

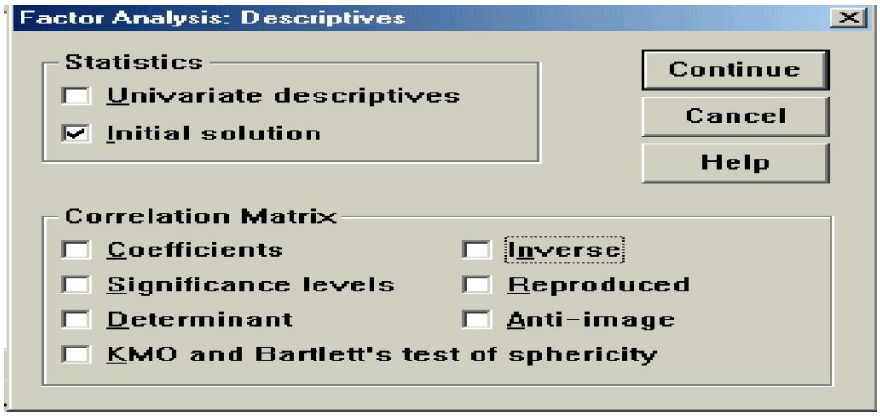
يطلب تحليل هيكل الارتباطات للمتغيرات المذكورة باستخدام طريقة المكونات الأساسية .

لتنفيذ ذلك نتبع الخطوات التالية :

- من شريط القوائم اختر Factor → Data Reduction → Analyze فيظهر صندوق حوار Factor Analysis الذي نرتبه كما يلي :

لاحظ أن المتغير Region لا يدخل في التحليل كونه متغير رمزي .

- عند نقر زر Descriptives يظهر صندوق الحوار التالي :



والذي يتضمن قسمين أساسيين :

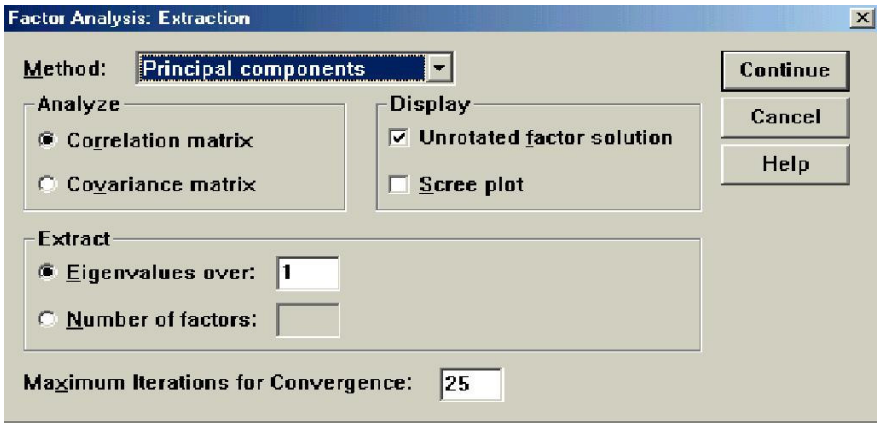
1. Statistics ويتضمن ما يلي :

univariate descriptives : ويعرض بعض الإحصائيات البسيطة للمتغيرات مثـل Mean ، Standard Deviation ...

Initial solution : ويعرض القيم الأولية للاشتراكيات Communalities ،الجذور الكامنـة (أو الكامنـة) Eigen Values والنسبة المئوية للتباين المفسر .

2. Correlation Matrix : يعرض مصفوفة معاملات الارتباطات ،مسـتوى المعنويـة ، المحـددة ومعكوس مصفوفة الارتباطات Inverse.

- عند نقر الزر Extraction في صندوق حوار Factor Analysis يظهر صندوق الحوار التالي الذي نرتبه كما يلي:



ويتضمن الصندوق التفاصيل التالية :

Method :لاختيار الطريقة المطلوبة في التحليل في هذا المثال اخترنا طريقة المكونات الأساسية (يتضـمن البرنامج سبعة طرق من طرق التحليل العاملي) .

Analyse : ويتضمن ما يلي :

Correlation Matrix : يتم تحليل مصفوفة الارتباطات للمتغيرات المدروسة ويكون ذلك ضرورياً في حالة اختلاف وحدات القياس للمتغيرات المشمولة .

covariance Matrix : : يتم تحليل مصفوفة التباين والتباين المشترك للمتغيرات المدروسة ويمكن اعتماد ذلك في حالة أن المتغيرات المدروسة لها نفس وحدات القياس .

Extraction :ويتضمن أسلوبين لاستخلاص المكونات (أو العوامل) وكما يلي :

Over Eigenvalues :يتم استخلاص المكونات الأساسية التي لها جذور كامنة أو تعرف أيضاً (بالجذور المميزة) تزيد عن قيمة معينة يحددها المستفيد (على الأغلب تساوي واحد) .

Number of Factors : يتم استخلاص عدد معين من المكونات(العوامل) يحدد عددها من قبل المستفيد.

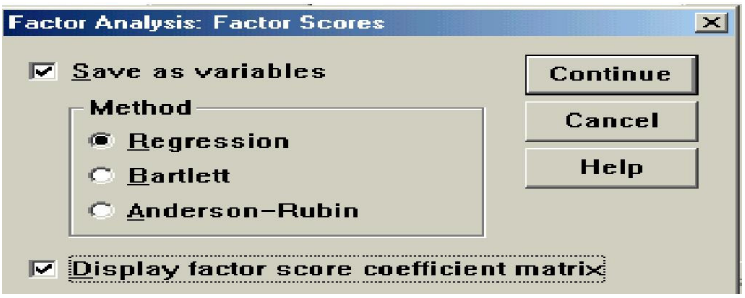
Maximum Iteration for Convergence : تحديد الحد الأعلى لعدد خطوات الخوارزمية اللازمة للوصول الى حل .

Display :يتضمن مايلي :

Unrotated factor solution يتم عرض تشبعات المتغيرات بالعوامل غير المدورة Factor Matrix وفي هذا المثال Component Matrix لأننا نستعمل طريقة المكونات الأساسية .

Scree Plot : يتم عرض مخطط يمثل المحور الأفقي رقم المكون أم المحور الصادي فيمثل الجذور الكامنة Eigen Values .

- عند نقر زر Rotation في صندوق حوار Factor Analysis يظهر صندوق حوار Rotation ويحتوي على خمسة طرق لتدوير المحاور حيث أن تدوير المحاور هي طريقة هندسية الغرض منها جعل التشبعات(Loadings) الكبيرة أكبر والتشبعات الصغيرة أصغر مما هي عليه قبل التدوير.كما يمكن أن تقلل من التشبعات السالبة وتزيد من التشبعات الصفرية في الحالات التي لايكون هناك تفسير منطقي للإشارة السالبة للتشبع .ويوفر البرنامج خمسة طرق للتدوير وهي (Varimax ، Direct Oblimin ، Quartimax ، Equamax ، Promax) . في هذا المثال أشرنا الاختيار None أي عدم تدوير المحاور .
- عند نقر زر Scores في صندوق حوار Factor Analysis يظهر صندوق حوار Factor Scores وكما يلي :



يتضمن الصندوق الفقرات التالية :

Save as Variables : عند تأشير هذا الخيار يقوم البرنامج بحساب العوامل Factor Scores (في هذا المثال المكونات الأساسية) وتضاف هذه المكونات الى يمين المتغيرات الموجودة في شاشة Data Editor علماً أن العامل Factor Score أو المكون الأساسي في هذا المثال هو عبارة عن تركيب خطي من المتغيرات المعيارية Z Scores ويمكن حساب العامل (المكون الأساسي) رقم i (عدد العوامل الكلي يكون بقدر عدد المتغيرات ويساوي p) بموجب المعادلة التالية :

$$F_i = \sum_{j=1}^{p} w_{ij} z_j \qquad (j = 1,2,...p)$$

حيث أن :

z : تمثل المتغيرات المعيارية ، w :تمثل الأوزان وهي معاملات العوامل Factor Scores Coefficients ونحصل عليها عند تأشير الخيار التالي :

Display Factor Score Coefficient Matrix علماً أن البرنامج يعرض العوامل المستخلصة فقط ومعاملاتها .

ملاحظة : يتم حساب العوامل بثلاثة طرق (Anderson-Rubin ، Bartlett، Regression) أما بالنسبة للمكونات فأن الطرق الثلاثة تعطي نفس النتائج عند تأشيرها حيث أن هناك متجه وحيد للمعاملات .

- يمكن بواسطة الزر Option في صندوق حوار Factor Analysis تنظيم استبعاد الحالات الحاوية على قيم مفقودة وكذلك ترتيب التشبعات في مصفوفة المكونات Components Matrix (التي سترد لاحقاً) حسب الحجم وعدم إظهار المعاملات التي تقل قيمتها المطلقة عن قيمة تحدد من قبل المستفيد .
- عند نقر زر OK في صندوق حوار Factor Analysis نحصل على النتائج التالية :

Communalities

	Initial	Extraction
GDP	1.000	.797
LITIRACY	1.000	.843
HIGHEDUC	1.000	.896
DOCTORS	1.000	.704
HOSPBED	1.000	.772

Extraction Method: Principal Component Analysis.

الجدول أعلاه يمثل القيم الأولية والمستخلصة للاشتراكيات Communalities حيث أن القيم الأولية للاشتراكيات تؤخذ مساوية الى الواحد في طريقة المكونات الأساسية في حالة اعتماد مصفوفة الارتباطات وتؤخذ الاشتراكيات مساوية لتباين كل متغير في حالة اعتماد مصفوفة التباينات أما بقية الطرق فتستعمل R^2 الناتج من انحدار كافة المتغيرات على متغير معين كتقدير لاشتراكية المتغير .

أن القيمة المستخلصة لاشتراكية المتغير GDP مثلاً تشير الى أن 0.797 من التباينات في قيم المتغير GDP تفسرها العوامل المشتركة(تم استخلاص عاملين) أن قيمة الاشتراكية تتراوح من 0 الى 1 وهي تعبر عن مربع معامل الارتباط المتعدد Square Multiple Correlation للمتغيرGDP مع المكونات(العوامل) وبصورة عامة نلاحظ أن العوامل المشتركة تفسر نسبة عالية من تباين المتغيرات حيث أن اقل نسبة هي 0.704 لمتغير Doctors .في حالة الحصول على قيمة صغيرة لاشتراكية أحد المتغيرات فهذا يشير الى عدم أهمية المتغير ويوصى باستبعاده من التحليل .

الجدول التالي يبين الجذور الكامنة لمصفوفة الارتباطات (تباين المكونات) ومجموعها يساوي رتبة المصفوفة ويساوي 5 بقدر عدد المتغيرات حيث أن المكون الرئيسي الأول له أكبر جذر كامن(أو تباين المكون) ويساوي 2.900 ويفسر 58% من التباينات الكلية لمتغيرات التنمية الاجتماعية حيث أن :