جمهورية مصر العربية

معهد التخطيط القومى

سلسلة مذكرات خارجية

مذكرة خارجية رقم (١٦٢٩)

التحليل الإحصائى بإستخدام برنامج SPSS

إعداد

أ.د.ماجدة ابراهيم

د.زلفى شلبى

أ.أحمد عبد الباقى

أكتوبر ٢٠٠٥

جمهورية مصر العربية - طريق صلاح سالم - مدينة نصر - القاهرة - مكتب بريد رقم ١١٧٦٥

A.R.E Salah Salem St. Nasr City, Cairo P.O. Box:11765

الفهرس

صفحة

أولاً : مقدمة

يعتبر البرنامج الجاهز SPSS (وهو اختصار للحروف الأولى من Statistical Package of Social Sciences) من أكثر البرامج الإحصائية استخداما من قبل شريحة واسعة من الطلبة والباحثين فى المجالات التربوية والاجتماعية والفنية والهندسية والزراعية والطبية فى إجراء التحليلات الإحصائية اللازمة .

وقد بدأت شركة SPSS؛ بإعداد هذا النظام الذى كان يعمل تحت نظام تشغيل Ms-Dos وقد تم تطويره ليعمل فى بيئة نظام التشغيل Windows فى عام ١٩٩٣ متلافياً بذلك الصعوبات التى كانت تواجه العاملين فى هذا النظام فى بيئة Ms-Dos ويتميز هذا البرنامج بالمميزات التالية :

١- السهولة فى إدخال البيانات حيث يقدم لنا نافذة (Data Editor) عبارة عن Spreadsheet به صفحة لتعريف المتغيرات وصفحة أخرى لإدخال وتعديل البيانات.

٢- السهولة فى استخراج النتائج ، حيث تظهر النتائج فى نافذة (SPSS viewer) بطريقة تمكن من الاختيار والتعديل ، كما يمكن طباعة هذه النتائج بطريقة محسنة عن طريق استخدام أنواع وأحجام ومواصفات عديدة لخطوط وأشكال النتائج المطبوعة وفى شكل جداول محورية pivot Tables ذات خطوط طولية وعرضية.

٣- العرض البياني للبيانات عن طريق رسم بياني غاية فى الدقة وبأشكال وألوان متعددة (أعمدة -خطوط-دائرة ...الخ)

٤- الاستفادة من بيئة Windows بما تقدمه لنا من رموز iconsصناديق حوارية Dialog boxes تسهل من تنفيذ الأوامر.

٥- إمكانية استخدام بيانات من برامج أخرى تعمل تحت Windowsكما يمكن استدعاء البيانات من برامج الجداول الالكترونية مثل الـ Excel وجداول قواعد البيانات مثل Access كما يمكن الاستفادة من النتائج فى برامج أخرى.

٦- يقدم لنا البرنامج نافذة (Syntax Editor) يمكن عن طريقها كتابة الأوامر كما يمكن ترجمة الاختيارات من الصناديق الحوارية إلى أوامر عن طريق لصقها (Paste) فى نافذة الـ syntax وتخزينها لاستخدامها لاحقاً.

٧- وجود وسيلة مساعدة (Statistics Coach) تساعد على التحليل الاحصائى المناسب لكل نوع من أنواع البيانات المختلفة.

ولقد غير برنامج SPSS وغيرة من البرامج المناظرة الحياة لكثير من المشتغلين بالإحصاء وتحليل البيانات ومنهم الطلاب الذين يتعلمون الإحصاء والمعلمون الذين يدرسونها والباحثون الذين يطبقونها ومع ذلك فمازال هناك الكثيرون الذين يجدون فى التعامل مع البرامج الإحصائية عملية صعبة وشاقة وخالية من أى متعة

ومازالوا يواجهون بصعوبات جمة وعراقيل تمنعهم من إتقان هذه المهارات الهامة ومن هذه العراقيل:-

١- ضعف الخلفية الإحصائية اللازمة للتعامل مع البرنامج تجعل المستخدم يحارب للوصول إلى هدفه دون جدوى، لأنه غير قادر على تحقيق شئ. واستخدام برنامج SPSS أو غيره من البرامج الإحصائية دون الالمام بالأساليب الإحصائية عملية خطيرة تقود إلى نتائج خاطئة وغير سليمة . ولذلك يحتاج العمل فى برنامج SPSS إلى خلفية إحصائية قوية تمكن المستخدم من التعامل براحة ويسر مع المفاهيم الإحصائية المختلفة .

٢-يمكن لبعض الباحثين التفكير فى طرق عدة يحللون بها بياناتهم ولكن يقفون مترددين حائرين لا يدرون انسبها وأصلحها للبيانات التى لديهم، وبخاصة عندما يتعلق الأمر بمسلمات يجب استيفاءها حتى يمكن قبول النتائج . وحتى إذا استطاعوا أن يصلوا إلى القرارات السليمة والنتائج المطلوبة ، فمازال أمامهم أن يعدو تقريراً يفسرون به نتائجهم، ويكون متمشياً مع الطرق المتعارف عليها أكاديميا فى كتابة التقارير.

٣-كمية النتائج التى يعطيها البرنامج قد تكون كبيرة جداً مما يدفع الباحثين إلى التراجع أمامها وبخاصة عندما يحاولون تفسيرها بل وقد يتركون الأمر برمته ويحاولون البحث عمن يستطيع مساعدتهم أمام هذا الخضم من الطلاسم المكتوبة سلفا غير مفهومة.

٤-رغم السهولة الكبيرة لبرنامج SPSS إلا أن المستخدمين الجدد يجدونه صعباً ومعقداً للغاية . وعليهم أن يتعلموا كيفية إدخال البيانات فى محرر البيانات وحفظها واسترجاعها، بل وإجراء بعض التعديلات ولكن المثابرة والجهد المتواصل فى دراسة البرنامج وفهمه كثيراً ما يؤدى ثماره فى الحصول على مهارة من أهم المهارات وهى استخدام الحاسب الآلى فى تحليل بيانات البحوث.

ولهذه المشاكل السابقة وغيرها تم إعداد هذه المادة العلمية حتى تكون عوناً للباحثين فى مختلف التخصصات ولمساعدتهم على تخطى العقبات التى تصادفهم أثناء تحليل البيانات وأثناء تفسير النتائج ، والإصدار الذى استخدم فى أمثلة هذه المذكرة **SPSS ver 11.1.0**

الباب الأول
تشغيل برنامج SPSS والتعامل مع البيانات والملفات

١-١ : مراحل التحليل الاحصائي باستخدام برنامج SPSS على الحاسب الآلي

نقطة البداية لأى عمل نقوم به فى برنامج SPSS هى عادة محرر البيانات وهو عبارة عن جدول الكتروني يستخدم لإدخال البيانات فيه وتحديد أسماء المتغيرات . وبعد الانتهاء من هاتين العمليتين يتم الانتقال إلى خطوات تحليل البيانات التى نريدها واختبار النتائج.

الخطوة الأولى : إدخال البيانات

الخطوة الأولى بطبيعة الحال هى إدخال البيانات وأخبار SPSS ما تمثله هذه البيانات وأسهل طريقة للقيام بذلك هى استخدام محرر البيانات وفيه:-

١- تدخل البيانات فى صفوف وأعمدة محرر البيانات.

٢- تسمية المتغيرات التى تستخدم فى تحليل البيانات.

الخطوة الثانية : تحديد التحليل الاحصائى

الخطوة التالية بعد إدخال البيانات هى إصدار التعليمات لبرنامج SPSS للتيام بالعمليات الإحصائية المرغوبة . هناك طريقتان لتنفيذ هذه الخطوة :-

١- الطريقة الأولى : هى طريقة التأشيرة والضغط على زر الفأرة Point-and-click Method وفى هذه الطريقة نقوم بالتحليل الذى نريده باستخدام الفأرة لفتح قوائم مسدله ومربعات حوار واختيار مانريد منها. وهذه الطريقة أسهل الطريقتين لأنها لاتتطلب معرفة أية لغة من لغات البرمجة (Syntax) . فالبرنامج مكتوب بالفعل وجاهز ولكنه موجود فى الخلفية وغير ظاهر للعين . ولكنه رهن إشارتك فى اى وقت . وهذه الطريقة مريحة ومألوفه لمستخدمى بيئة النوافذ حيث أن الواجهة المستخدمة شبيهه بالواجهات الأخرى المستخدمة فى برامج النوافذ.

٢- الطريقة الثانية: وهى التى يمكن أن نطلق عليها الطريقة اللغوية Syntax Methad وتتضمن استخدام برنامج SPSS بالطريقة التقليدية والتى كانت مستخدمة فى الماضى عندما كان البرنامج يعمل فى بيئة Dos .

وعند استخدام هذه الطريقة نبدأ بفتح نافذة جديدة يطلق عليها محرر اللغة Syntax Editor ويكتب فيها التعليمات بلغة البرمجة الخاصة ببرنامج SPSS . وتتطلب هذه الطريقة تعلم هذه اللغة ، وان كان هذا الأمر يبدو صعباً بعض الشيء فى البداية ، إلا أن هناك مزايا عدة لاستخدام هذه الطريقة اقلها أن المستخدم يستطيع عمل أشياء بها غير متاحة فى طريقة التأشيرة والضغط .

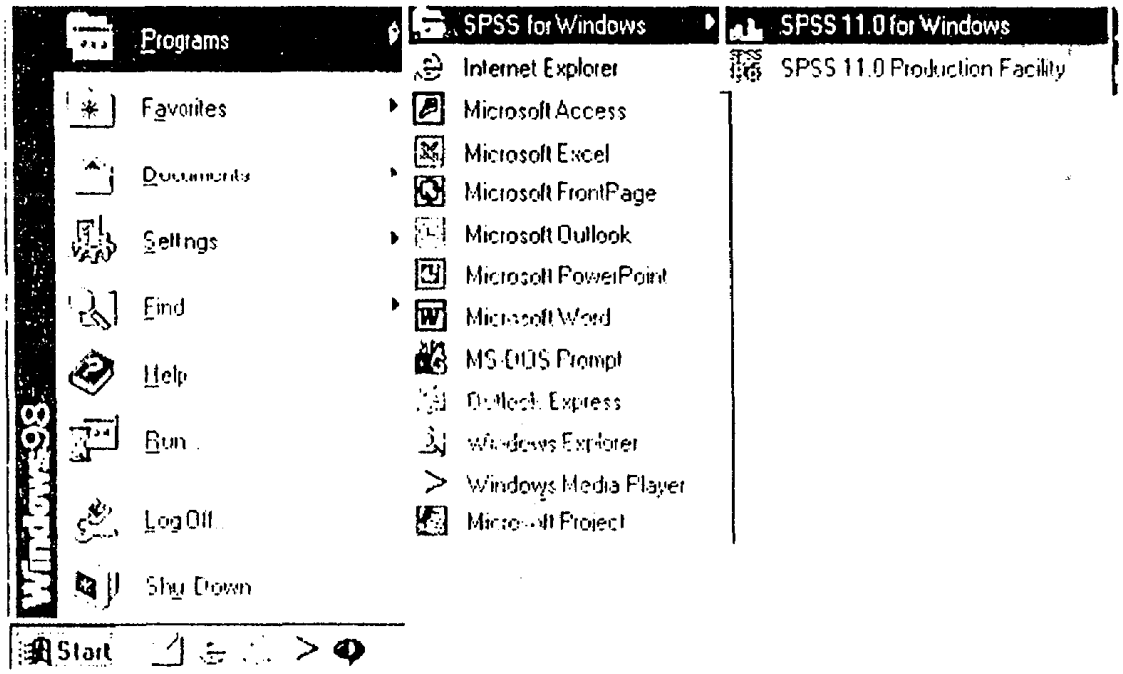
وتستخدم كلا الطريقتين عند مناقشة تحليل البيانات تحليلاً إحصائيا . ويمكن للمستخدم الاقتصار على طريقة واحدة فقط منهما، وتمكنه أن ينتقل بين الطريقتين كيفما شاء.

الخطوة الثالثة: فحص المخرجات وتعديلها:-

بعد ادخال البيانات وتحليلها باستخدام احدى الطريقتين السابقتين ، تظهر نافذة جديدة تحتوى على نتائج التحليل. ويمكن طباعة النتائج أو حفظها على القرص الصلب أو القرص المرن او القرص المضغوط للعودة اليها فى المستقبل .

١-٣ : طريقة تشغيل البرنامج

يتم تحميل النظام عن طريق نقر رمز Start ومنها يتم اختيار Programs ومنها الرمز SPSS for Windows ومنها SPSS 11.0 for Windows .

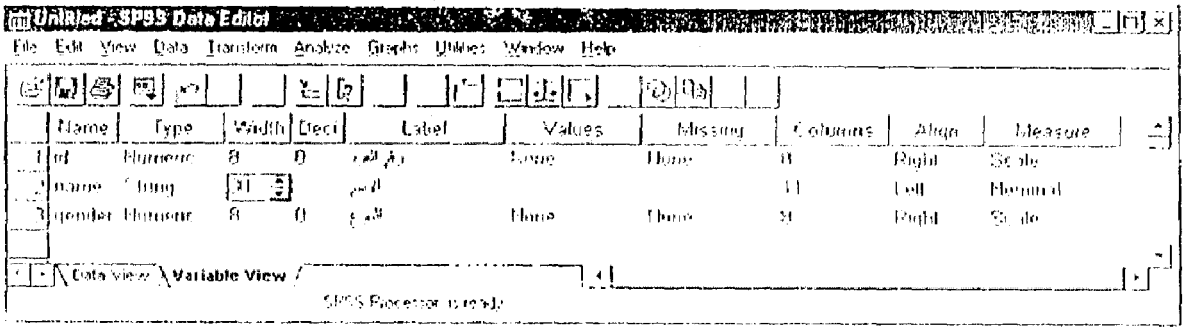


ويمكن إنشاء أيقونة خاصة بالبرنامج على سطح المكتب Desktop واستدعاء البرنامج منها مباشرة .

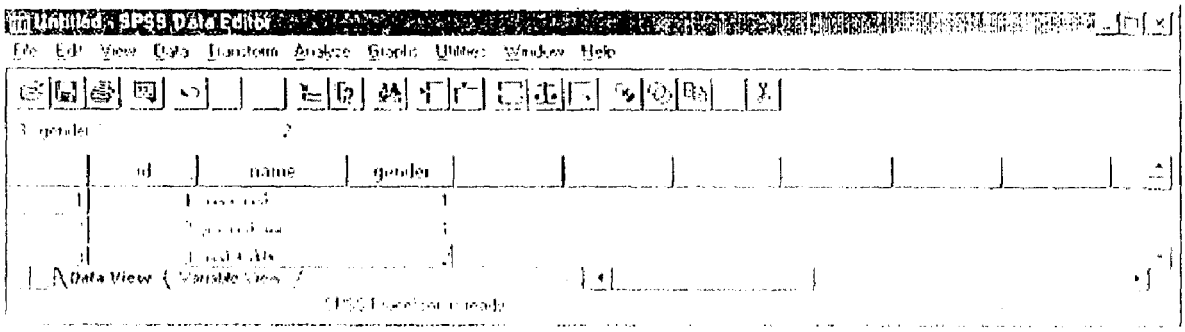
بعد تشغيل البرنامج تظهر (نافذة البيانات SPSS Data Editor) ، وهى عبارة عن نافذة على شكل صفوف وأعمدة Spreadsheet من النوع Tab Pages وهى تشتمل على صفحتين:

صفحة تعريف المتغيرات (Variable View) وصفحة إدخال البيانات (Data View) .

تختلف محتويات هذه النافذة حسب الصفحة النشطة ، ففي صفحة تعريف المتغيرات (Variable View) تعبر الصفوف عن المتغيرات حيث يعبر كل صف عن متغير ، كما تعبر الأعمدة عن مواصفات هذه المتغيرات مثل اسم المتغير Name ، نوع المتغير Type ، حجم المتغير Width ... الخ .



أما إذا كانت الصفحة النشطة هي صفحة البيانات Data View فإن الأعمدة تعبر عن المتغيرات ، حيث يعبر كل عمود عن متغير Variable مثل رقم الفرد Id ، اسم الفرد Name ، نوعه Gender ... إلخ ، وكل صف يعبر عن حالة Case أي مفردة من مفردات التحليل كبيانات فرد معين .



وتشمل هذه النافذة على شريط القوائم ويشمل مجموعة القوائم التالية :

File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Window Help

وتستخدم هذه القوائم فى تنفيذ أوامر النظام ، كما تشمل على شريط الأدوات (Toolbar) وهو يستخدم لتنفيذ الأوامر شائعة الاستخدام فى النظام .

١-٣ : التعامل مع البيانات والملفات

بعد تشغيل البرنامج ، ولإجراء التحليل الإحصائى على بيانات معينة بعد جمعها من مصادرها المختلفة ، يلزم تعريف هذه البيانات للبرنامج حتى يمكن إجراء التحليل الإحصائى عليها ، ويشمل ذلك ما يلى :

- **تعريف وتوصيف المتغيرات Define Variables**

يستخدم فى تعريف المتغيرات نافذة البيانات Data Editor ، ويشمل تعريف المتغيرات معلومات عن هذه المتغيرات ، ويمكن تعريف المتغيرات قبل أو بعد إدخال البيانات ، ولكن يفضل تعريفها قبل الإدخال .

وقبل البدء فى تعريف المتغيرات يجب التأكد أن النافذة النشطة Data Editor وأن الصفحة النشطة هى صفحة تعريف المتغيرات Variable View حيث يظهر عنوانها SPSS Data Editor ، وعنوان الملف Untitled مضاءً وهو عبارة عن Spreadsheet الصفوف تمثل المتغيرات Variables والأعمدة تمثل مواصفات هذه المتغيرات مثل النوع والحجم ... إلخ ويتم التعريف على النحو التالى :

File Edit View Data Transform Analyze Graphs Utilities Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	id	Numeric	8	2		None	None	8	Right	Scale
2	name	String	30					8	Left	Nominal
3	b_date	Date				None		8	Right	Scale
4	gender	Numeric	8	2		{1.00, ...	None	8	Right	Scale

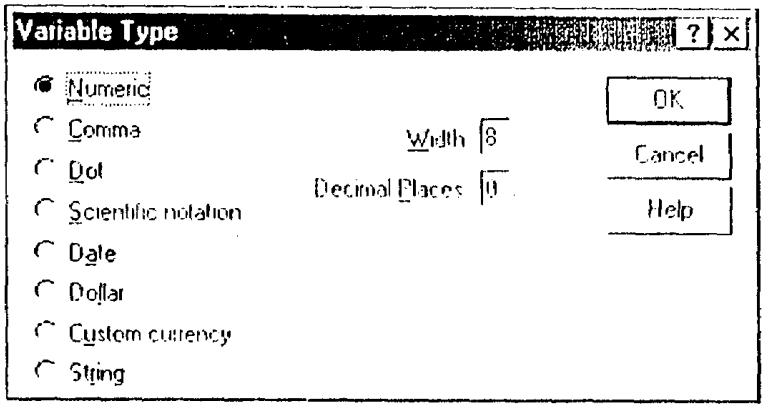
(١) اسم المتغير Variable name

أكتب اسم المتغير فى عمود Name مع ملاحظة أن اسم المتغير لا يزيد عن ٨ خانات ، ولابد أن تبدأ بحرف ، وغير مسموح بالمسافات والعلامات الخاصة مثل علامتى الجمع والطرح ، كما أن اسم المتغير لا يتكرر فى الملف الواحد ويفضل أن يعبر اسم المتغير عن البيانات الداخلة فيه فمثلاً متغير اسم الفرد يسمى Name أو النوع يسمى Gender أو تاريخ الميلاد يسمى B_date ... الخ ، أما إذا كانت البيانات عبارة عن أسئلة استبيان Questionnaire فيمكن أن تكون أسماء المتغيرات Q1,Q2,...etc بأرقام أسئلة الاستبيان للتبسيط ويمكن شرح وتوصيف هذه المتغيرات فى خانة Label .

(٢) نوع المتغير Variable Type

Type
Numeric ...

لتعريف نوع البيانات فى هذا المتغير ، أنقر النقاط الـثلاث فى خانـة Type حيث يظهر الصندوق الحوارى Variable Type التالى :

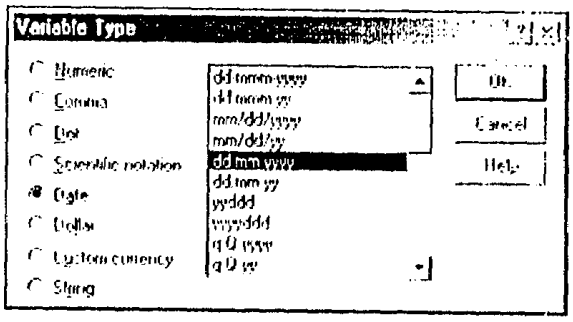


وهو يحدد الأنواع المختلفة للمتغيرات ، مع ملاحظـة أن نـوع المـتغير يتحـدد تلقائيـاً (by default) بأنه رقمى ، ويمكن تعريف نوع المتغير بأحد الأنواع التالية :

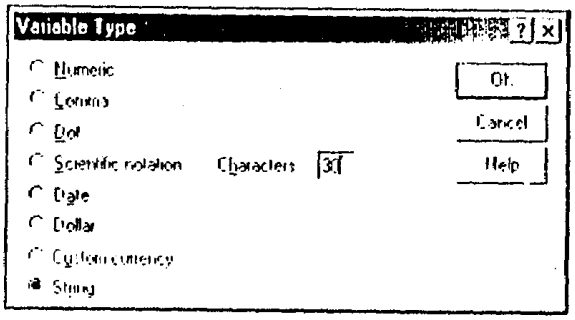
- Numeric رقمى: ويشمل التعريف العدد الكلى لخانات المتغير Width ، وعدد الخانات العشرية Decimal Places .
- Comma : متغير رقمى ، ولكن أرقامه تنفصل بـفاصلة comma بين الملايين والألوف ، وبنقطة للأرقام العشرية مثلاً الرقم 134565.15 يمثل هكذا 134,565.15
- Dot : متغير رقمى ، ولكن أرقامه تنفصل بنقطة بين الملايـين والألـوف ، وبـفاصلة comma للأرقام العشرية . مثلاً الرقم 134565.15 يمثل هكذا 134.365,15
- Scientific notation : متغير رقمى ، ولكن يظهر بالـ E Format أى الرقم مضروباً فى ١٠ مرفوعةً لأس معين (موجب فى حالة البيانات الكبيرة ، وسالب فى حالة البيانـات الصغيرة) وذلك يستخدم فى حالة تمثيل البيانات الكبيرة جداً أو الصغيرة جداً ، ويلاحظ أن البرنامج SPSS يمثل به الكثير من مخرجاته فمثلاً الرقم 14000000 يمثل 1.4 E+07 أى 1.4 مضروباً فى 10^7 والرقم 0.000023 يمثل هكذا 2.3E-5 أى 2.3 مقسوماً على 10^5

- Custom currency : متغير تظهر به رموز عملات ، يمكن اختيارها

- Dollar : متغير رقمى قيمته تشتمل على علامة الدولار .



- Date متغير يمثل التـــاريخ ، وفى حالـــة اختياره يسمح بعدة أشكال معينة للتاريخ يمكن الاختيار من بينها ، حيث تمثـــل dd اليوم ، تمثل mm الشهر ، تمثل yy السنة وكل من اليوم والشهر والسنة يظهر فى خانتين فمثلا إذا اخترنا الشكل dd.mm.yy فـــإن التاريخ ٤ أغسطس ٢٠٠٣ يسجل هكذا ٠٣/٨/٣ ويظهر بالشكل ٠٣,٠٨,٠٣ ويلاحظ أن الـ Date لا يحتاج إلى تحديد حجم المقل حيث أنه يخزن تلقائياً فى ٨ خانات .



- String مـــتغير حـــرفى: ويســـتخدم للبيانات الحرفية مثل أسماء أفراد الأسرة التى لا تجرى عليها عمليـــات حســـابية وفيه يتحدد عدد حروف هذا المتغير فى خانة Character .

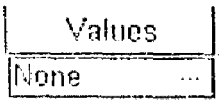
ملاحظة

الخانات Decimal , Width خاصة بالبيانات الرقمية وفيها يتم تحديد عدد خانات هـــذا المتغير ، وعدد الخانات العشرية فمثلاً الرقم ١٣٣,٨ الـــ Decimal=1 Width=5 ، ويلاحظ دائماً أن يكون حجم المتغير Width بحيث يتسع لأكبر رقم فيه وإلا فإن الرقم سيمثل بالـــــ E format ، كما يجب أيضاً أن تكون عدد الخانات العشرية حسب الدقة المطلوبة وإلا سيؤدى إلى تقريب الأرقام .

(٣) تعريف وصف المتغير Label

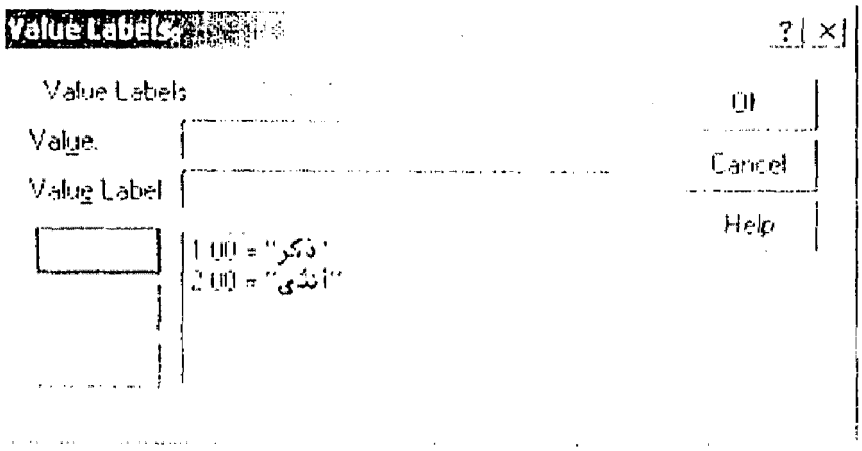
فى خانة Label اكتب وصفاً يشرح كل متغير ، باللغة العربية أو الإنجليزية حسب البيانات ، وذلك لسهولة تفسير البيانات والنتائج فمثلا فى إستمارة الاستبيان السؤال الأول يعطى اسم المتغير q1 أما الـ Label فيكون "النوع" ، السؤال الثانى يعطى اسم المتغير q2 أما الـ Label فيكون "صلة القرابة برب الأسرة"

(٤) تعريف أدلة المتغير Values



كما نعلم فإن معظم إجابات أسئلة الاستبيان تأخذ فمثلاً q_1 أى السؤال الأول وهو النوع يرمز الكود ١ إلى ذكر ، يرمز الكود ٢ إلى أنثى .

ولكى يتم تعريف هذه البيانات المكودة أنقر الثلاث نقاط فى عمود Values أمام المتغير المراد تعريفه وليكن متغير gender فيظهر الصندوق الحوارى Value Labels :



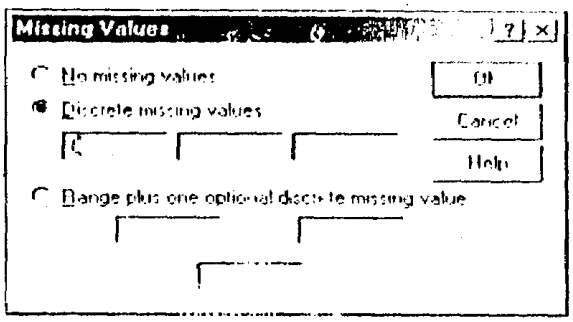
أمام خانة Value أكتب الرقم ١ ، أمام خانة Value Label: أكتب التعريف "ذكر" ثم انقر أمر Add ، أمام خانة Value أكتب الرقم ٢ أمام خانة Value Label: أكتب "أنثى" ثم أنقر أمر Add .

ملاحظة : يمكن إعادة تعريف هذه الأكواد أو إلغاؤها عن طريق الأوامر Remove , Change

(٥)تحديد القيم المفقودة Missing Values

يمكن تحديد عدة قيم للقيم المفقودة (وهى البيانات غير المبينة) والتى لم يتم الحصول عليها ، أو البيانات التى لا تنطبق .

لتعريف هذه البيانات المفقودة أنقر الثلاث نقاط فى عمود Missing فيظهر الصندوق الحوارى Missing Values الآتى :



ويوجد به الاختيارات التالية :

No Missing Value

لا يوجد قيم مفقودة

Discrete missing values القيم المفقودة لها رموز مقطعة مثل 99 , 9 وهكذا

Discrete missing values Range plus one discrete

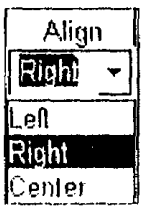
يسمح بتعريف مدى متصل مع تعريف قيمة واحدة متقطعة مثل من 100 إلى 105 ، 999 حيث تكتب 100 فى الخانة Low ، 105 فى الخانة High ، 999 فى خانة Discrete value.

ويعطى البرنامج هذه الإمكانيات فى القيم المفقودة لأن أنواعها قد تتعدد فيلزم تحديد كود لكل نوع ، فهناك قيم مفقودة Missing لأن مصادرها غير متاحة وأخرى لأن السؤال لا ينطبق Not Applicable على هذه الحالة ، والثالثة لأن الفرد امتنع عن الإجابة Not Answer .

(٦)تحديد عرض العمود (المتغير)

فى العمود Column يتم تحديد عرض العمود فى طريقة إظهاره فقط على الشاشة ، أو فى الطباعة ، أما السعة التخزينية للمتغير فتتحدد فى العمود Width.

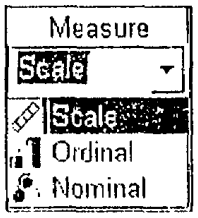
(٧)تحديد محاذاة المتغير Align



فى العمود Align يتم محاذاة المتغير إما فى يمين العمود أو إلى الوسط أو فى اليسار .

(٨)تحديد طريقة قياس المتغير Measure

حيث يحدد ثلاثة أساليب لقياس المتغير :



أ- اسمى Nominal ، حيث أن الأرقام لا تدل على كميات ، وإنما ترمز فقط إلى أكواد ، مثل متغير الحالة الزواجية (١ أعزب ، ٢ متزوج ، ٣ مطلق ، ٤ أرمل) وكمتغير النوع (١ ذكر ، ٢ أنثى) ، ويستخدم مع هذه البيانات أنواع معينة من المقاييس مثل المنوال Mode ، كا^{2} Chi Square .

ب- متغير ترتيبى Ordinal ، ويستخدم عندما تكون البيانات فى شكل ترتيبى من الأدنى إلى الأعلى مثل الحالة التعليمية (٠ أقل من السن ، ١ أمى ، ٢ يقرأ ويكتب ، ٣ إبتدائى ، ٤ إعدادى ، ٥ ثانوى ، ٦ فوق المتوسط وأقل من جامعى ، ٧ جامعى ، ٨ عليا) ، ويوجد أيضاً أنواع معينة من التحليل تناسب هذا النوع من البيانات مثل معامل ارتباط سبيرمان للرتب .

ج- رقمى Scale ، حيث أن الرقم يدل على قيمة للمتغير مثل العمر ، الوزن ، الدخل ... الخ

ملاحظة

فى حالة تشابه خصائص المتغيرات فى Type, Width, Values...etc يمكن استخدام أوامر النسخ Copy واللصق Paste لنسخ الخصائص المتشابهة.

- إدخال البيانات

كيفية إدخال البيانات عن طريق النافذة (Data Editor)

- يمكن إدخال البيانات قبل توصيف المتغيرات وفى هذه الحالة يحدد البرنامج للمتغيرات أسماء var00001, var00002, ... etc إلى أن يتم توصيف هذه المتغيرات
- المتغيرات Variables تدخل أعمدة ، والحالات تدخل صفوف .

ملاحظة

يمكن إظهار Value Labels فى الخلايا بدلاً من الأكواد عن طريق اختيار Value Labels من القائمة View فمثلاً يمكن إظهار متغير النوع ذكر ، أنثى بدلاً من ١ ، ٢ .

مثال : أدخل البيانات التالية ، وهى عبارة عن بيانات افتراضية لمجموعة من العاملين بإحدى الجهات :

رقم الفرد	النوع	العمر	الحالة التعليمية	الحالة الزواجية	المهنة
1	ذكر	48	فوق المتوسط	متزوج	أعمال فنية
2	أنثى	40	فوق المتوسط	متزوج	أعمال فنية
3	ذكر	17	ثانوى	أعزب	أعمال كتابية
4	أنثى	15	إعدادى	أعزب	وظائف معاونة
5	ذكر	56	جامعى	متزوج	أعمال تخصصية
6	أنثى	50	أمى	متزوج	وظائف معاونة
7	ذكر	22	جامعى	مطلق	أعمال تخصصية
8	أنثى	20	ثانوى	متزوج	أعمال فنية
9	أنثى	17	إعدادى	متزوج	وظائف معاونة
10	ذكر	35	فوق المتوسط	أرمل	أعمال كتابية

- يتم تسمية المتغيرات كالآتى :

رقم الفرد ID

النوع Gender

العمر Age

الحالة التعليمية Educat

الحالة الزواجية marriage

المهنة Work

- يتم تكويد المتغيرات الوصفية كالتالى :

النوع Gender

١ ذكر

٢ أنثى

الحالة التعليمية Educat

١ أمى

٢ يقرأ ويكتب

٣ مؤهل متوسط

٤ فوق المتوسط

٥ جامعى

الحالة الزواجية marriage

١ أعزب

٢ متزوج

٣ مطلق

٤ أرمل

المهنة Work

١ أعمال تخصصية

٢ أعمال فنية

٣ أعمال كتابية

٤ وظائف معاونة

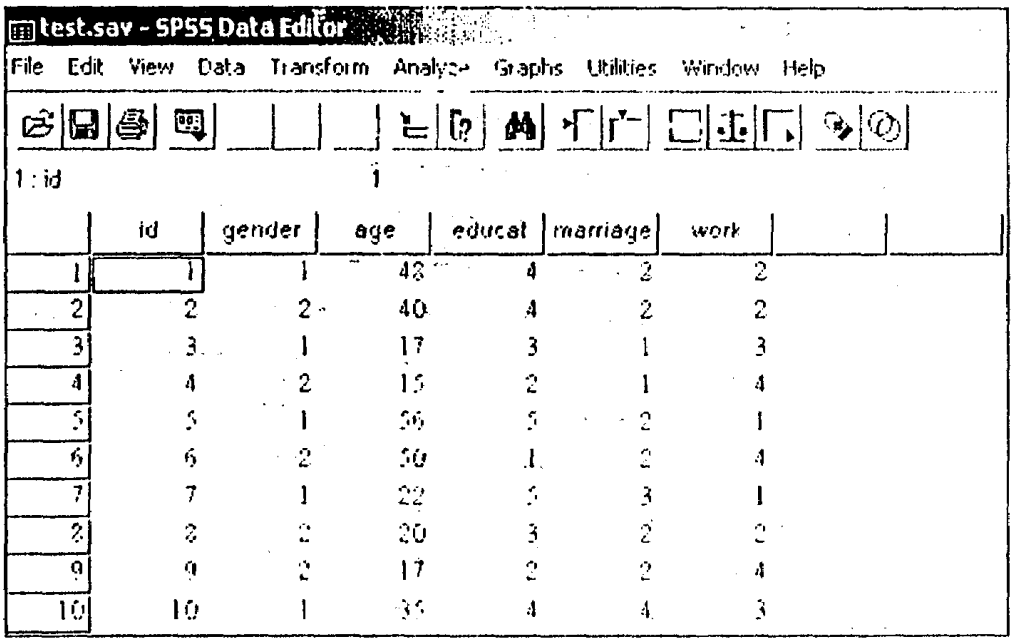
- بعد تكويد البيانات ، يظهر جدول البيانات بالشكل التالى :

ID رقم الفرد	Gender النوع	Age العمر	Educat الحالة التعليمية	marriage الحالة الزواجية	Work المهنة بالتفصيل
1	1	48	4	2	2
2	2	40	4	2	2
3	1	17	3	1	3
4	2	15	2	1	4
5	1	56	5	2	1
6	2	50	1	2	4
7	1	22	5	3	1
8	2	20	3	2	2
9	2	17	2	2	4
10	1	35	4	4	3

- يتم توصيف المتغيرات كما سبق أن تعرفنا فى صفحة Variable View بالتوصيف التالى :

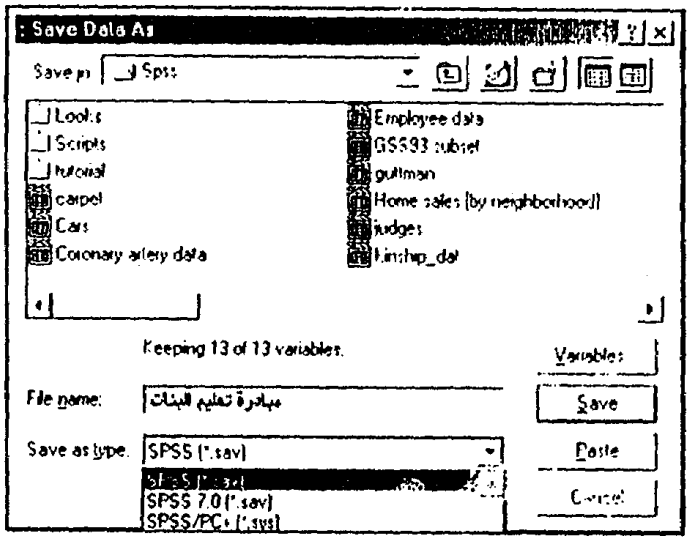
Name	Type	Width	Decimals	Label	Values	Missing	Column	Align	Measure
ID	Numeric	3	0	رقم الفرد	None	None	3	Right	Nominal
Gender	Numeric	1	0	النوع	١- ذكر ٢- انثى	None	3	Right	Nominal
Age	Numeric	3	0	العمر	None	None	3	Right	Scale
Educat	Numeric	1	0	الحالة التعليمية	١- امى ٢- يقرأ ويكتب ٣- مؤهل متوسط ٤- فوق المتوسط ٥- جامعى	None	3	Right	Scale
Marriage	Numeric	1	0	الحالة الزواجية	١-أعزب ٢-متزوج ٣-مطلق ٤-أرمل	None	2	Right	Ordinal
Work	Numeric	1	0	المهنة	١- أعمال تخصصية ٢- أعمال فنية ٣-أعمال فنية ٤- وظائف معاونة	None	2	Right	Nominal

- بعد ذلك يتم إدخال البيانات فى صفحة إدخال البيانـــات (Data View) فتظهـــر النافذة بالشكل التالى :



	id	gender	age	educat	marriage	work
1	1	1	43	4	2	2
2	2	2	40	4	2	2
3	3	1	17	3	1	3
4	4	2	15	2	1	4
5	5	1	56	5	2	1
6	6	2	50	1	2	4
7	7	1	22	3	3	1
8	8	2	20	3	2	2
9	9	2	17	2	2	4
10	10	1	35	4	4	3

حفظ البيانات



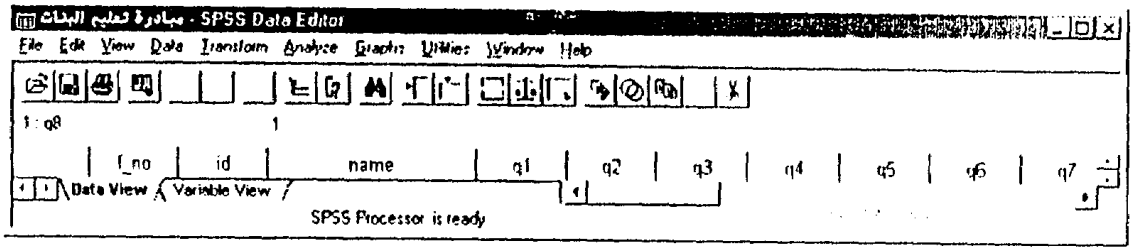
نجد أن البرنامج يخصص لهذه البيانات الاسم Untitled ، ولحفظ هذه البيانات وتخصيص اسم للملف ، نختار الأمر Save من القائمة File (أو الضغط على أيقونة Save) فيظهر الصندوق الحوارى Save Data As كما فى الشكل المقابل :

يتم كتابة اسم الملف فى خانة File Name ، ويتم تحديد وسيط التخزين ، الدليل من خانة Save in .

ملاحظات

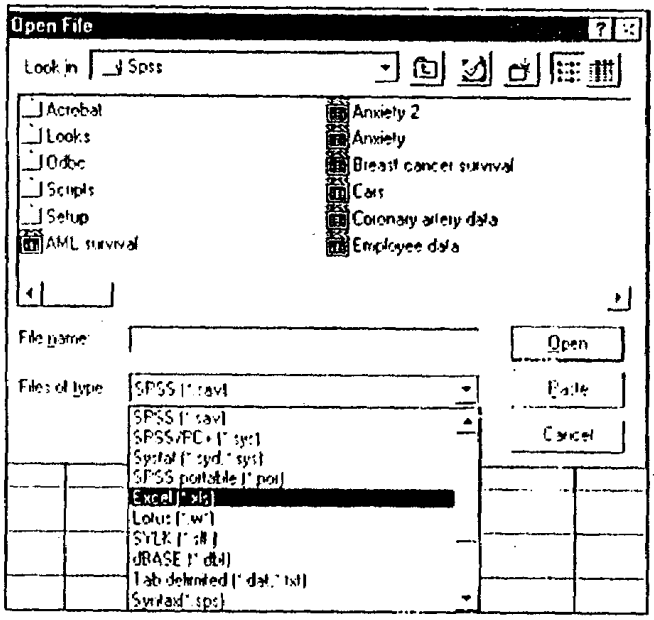
- عند الحفظ لأول مرة يفتح الصندوق الحوارى Save Data As ليسمح بكتابة اسم الملف ، ومكان حفظ الملف ، أما عند تكرار الحفظ باختيار أمر Save من القائمة File فسيتم الحفظ بدون فتح الصندوق الحوارى .
- إذا أردت أن تحتفظ بنسخة من الملف باسم آخر أو فى مكان آخر اختر أمر File Save As← سيتم فتح الصندوق الحوارى Save Data As السابق .
- البرنامج يحفظ البيانات تلقائياً فى الدليل Program Files\SPSS ، وهذا الدليل يحذف إذا تم حذف برنامج SPSS من على الجهاز مما يؤدى إلى فقد البيانات ولذلك يفضل الحفظ فى دليل آخر ينشئه مستخدم البرنامج .
- يفضل الاحتفاظ بنسخ من البيانات Data على وسيط تخزين خارجى مثل الأقراص المرنة Floppy diskettes أو الاسطوانات CD-Rom .

- ملفات بيانات البرنامج SPSS يعطى لها البرنامج الامتداد SAV ، وإذا أردنا أن نحفظ الملف بنوع آخر من أنواع البيانات كالـ Excel فيمكن اختيار هذا النوع من القائمة Save as Type فى الصندوق الحوارى Save Data As.
- بعد حفظ الملف ، يظهر اسمه فى شريط العنوان كما يلى :



التعامل مع بيانات من أنظمة أخرى

على الرغم من سهولة استخدام Data Editor كنافذة للإدخال والتعديل والتعامل مع البيانات فى نظام SPSS إلا أنه أيضاً يتيح فتح الملفات التى تم تسجيلها بأنظمة أخرى مثل dBase , Excel على النحو التالى :



١) من قائمة File اختر أمر Open ومنها اختر Data يتم فتح الصندوق الحوارى Open File.

٢) من خانة Files of type تظهر قائمة بأنواع الملفات التى يتعامل معها البرنامج مثل Excel وهى الملفات التى امتدادها xls ، وملفات قواعد البيانات التى امتدادها dbf يتم تحديد نوع الملفات المطلوب فتحها ، وذلك بعد تحديد وسيط ودليل التخزين ، ثم نختار أمر Open يتم فتح الملف فى نافذة البيانات Data Editor .

تعديل البيانات Editing Data

بعد إدخال البيانات أو نقلها من أى نظام آخر ، قد نحتاج إلى إجراء تعديلات على هذه البيانات ، وتتيح نافذة البيانات إمكانيات كبيرة فى عمليات التعديل المختلفة كالآتى :

(١) استبدال البيانات

قد يوجد خطأ فى البيانات المسجلة فى بعض الخلايا ، نختار الخلية ، ونجرى التصحيح

(٢) نسخ ونقل البيانات

يمكن قص Cut ، أو نسخ Copy البيانات ثم لصقها Paste طبقاً للحالات التالية :

- يمكن نقل أو نسخ خلية واحدة إلى خلية أو مجموعة خلايا .
- يمكن نقل أ ونسخ صف (Case) إلى صف أو مجموعة صفوف .
- يمكن نقل أو نسخ عمود (Variable) إلى عمود أو مجموعة أعمدة .

ولتنفيذ ذلك يتم اختيار الخلايا المطلوب نسخها أو نقلها ويتم اختيار Cut من قائمة Edit فى حالة النقل ، أو Copy فى حالة النسخ ، ثم تحديد الخلايا المراد نقل أو نسخ البيانات إليها ، واختيار أمر Paste من القائمة Edit .

(٣) إدراج أو حذف الصفوف

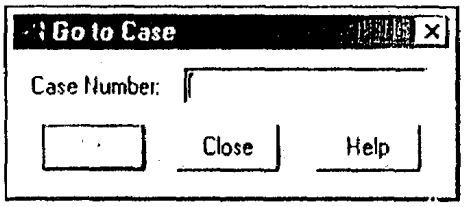
يمكن إدراج صفوف (Cases) فى أى مكان فى نافذة البيانات ، حيث يتم اختيار صف أسفل الصف المراد إدراجه ثم من قائمة Data نختار الأمر Insert case ، ولكى نحذف صف نختار الصف المراد حذفه عن طريق نقر رقم الصف ، ثم اختيار أمر Cut من قائمة Edit .

(٤) إدراج أو حذف أعمدة

بنفس الطريقة السابقة يمكن إدراج أعمدة (Variables) فى أى مكان فى نافذة البيانات حيث يتم اختيار متغير بعد المتغير المراد إدراجه ثم من قائمة Data نختار الأمر Insert Variable ، ولكى نحذف متغير معين نختار المتغير المراد حذفه عن طريق نقر اسم المتغير ، ثم اختيار أمر Cut من قائمة Edit

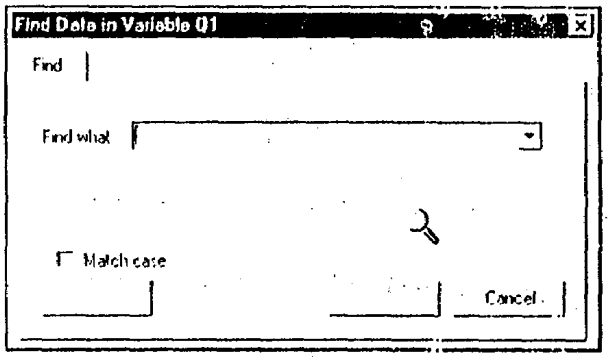
(٥) الانتقال إلى صف معين

يمكن الانتقال خلال نافذة البيانات عن طريق الـ Scrolling bar فيمكن الانتقال صف لأعلى أو لأسفل عن طريق نقر أسهم التصفح ، كما يمكن الانتقال شاشة لأعلى أو لأسفل عن طريق النقر أعلى أو أسفل مربع التصفح ، وللانتقال إلى صف (أو حالة) معينة نتبع الآتى :



من قائمة Data اختر أمر Go to Case فيظهر الصندوق الحوارى Go to Case وأمام خانة Case Number أكتب رقم الصف المطلوب .

(٦) البحث عن بيانات



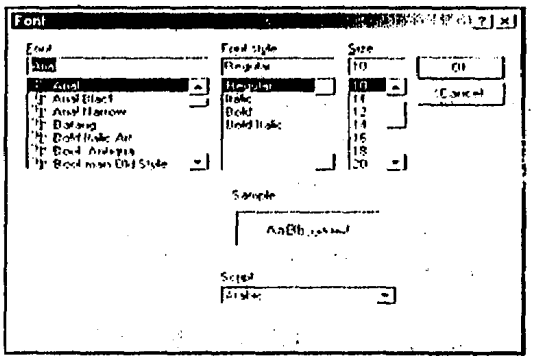
يمكن البحث عن بيانات معينة باتباع ما يلى :

يتم تحديد المتغير المراد البحث عن قيمة معينة فى بياناته ، ومن قائمة Edit اختر أمر Find

أمام خانة Find what أكتب البيان المراد البحث عنه ، يتم تحديد أول قيمة تتفق مع قيمة البحث ثم يتم تحديد القيمة الثانية ، وهكذا حتى ينتهى البحث .

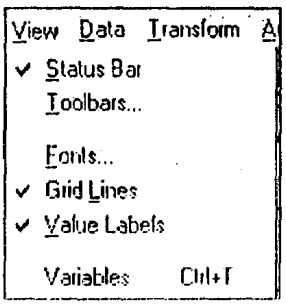
طرق العرض فى نافذة إدخال البيانات View

يمكن التحكم فى شكل البيانات المعروضة فى شاشة الإدخال ، وذلك لتحسين الشكل كالآتى



من القائمة View يتم اختيار :

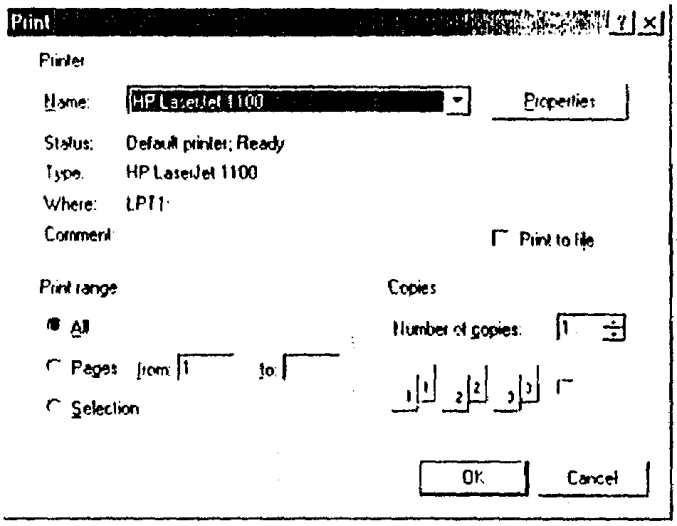
(١) Fonts وذلك للتحكم فى الخط المعروض من حيث حجمه ، ونوعه ومواصفاته . يتم اختيار Arabic من خانة script إذا كانت هناك أدلة باللغة العربية



(٢) يمكن تنشيط الاختيار Gridlines لإظهار الخطوط الطولية والعرضية ، أو عدم تنشيطه لإخفائها .

(٣)يمكن تنشيط الاختيار Value Labels لإظهار التوصيف Category للبيانات المكودة ، أو عدم تنشيطه لإخفاء التوصيف وإظهار الأكواد .

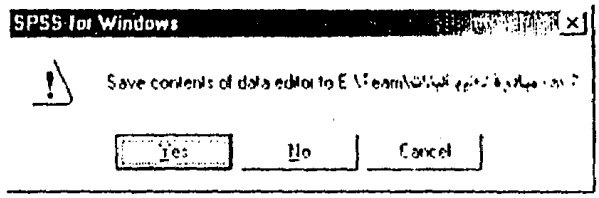
طباعة البيانات



يمكن طباعة البيانات بالكامل أو جزء منها ، أو صفحات معينة كما يمكن تحديد عدد النسخ عن طريق اختيار أمر Print من القائمة File فتظهر نافذة Print المقابلة . ويمكن تحديد مدى الطباعة من خانة Print range على النحو التالى :

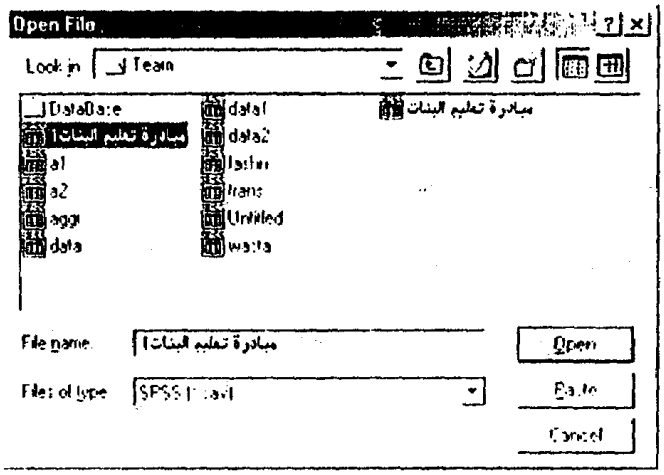
- All لطباعة كل المستند .
- Pages لطباعة صفحات متصلة ، يتم تحديد رقم صفحة البداية أمام خانة from: ورقم صفحة النهاية أمام خانة to .
- Selection لطباعة مدى تم تحديده من قبل .

الخروج من البرنامج



لكى يتم الخروج من البرنامج اختر أمر Exit من القائمة File ، تظهر رسالة تحذيرية ترشد إلى حفظ الملفات ، اختر أمر Yes للحفظ .

ملاحظة



لفتح ملف موجود اتبع الآتى :

١- بعد تشغيل البرنامج ، اختر الأمر File → Open → Data من نافذة البيانات يظهر الصندوق الحوارى Open File .

٢- حدد مكان تخزين الملف ، واسم الملف ، ثم اختر الأمر Open يتم فتح الملف فى نافذة البيانات .

١-٤ : تحويل البيانات

فى بعض الأحيان يمكن إجراء التحليلات الإحصائية على البيانات الخام ، ولكن فى الغالب نحتاج إلى إجراء بعض التحويلات على البيانات مثل إيجاد متغيرات جديدة عبارة عن علاقة بين المتغيرات الموجودة باستخدام صيغ ومعادلات معينة ، كما يمكن أن يكون هذا التحويل مقترناً بشرط أو غير مشروط ، كما نحتاج إلى اختيار حالات معينة لإجراء التحليل وليس كل الحالات .

حساب قيم متغير بناءً على دالة معينة Compute Variable :

يمكن حساب قيم متغير وإحلال القيم الجديدة محل القيم القديمة لنفس المتغير ، كما يمكن وضع نتيجة الحساب فى متغير جديد ، ولإجراء ذلك نتبع الآتى :

١) من قائمة Transform اختر أمر Compute ، يظهر الصندوق الحوارى Compute .

٢) أمام خانة Target Variable أكتب المتغير الجديد المراد وضع نتيجة الحساب فيه ، أو اسم المتغير المراد تحويله إذا أردنا أن يتم الحساب فى نفس المتغير .

٣) أكتب الدالة المراد إجراؤها (أى الطرف الأيسر للعملية الحسابية) مستعيناً بالـ Functions الموجودة فى قائمة Functions وبالتعبيرات الحسابية والمنطقية الموجودة .

٤) اختر أمر OK .

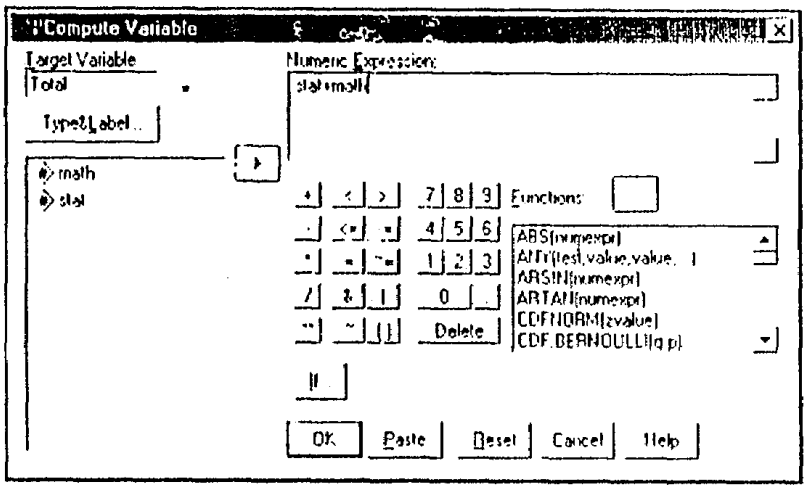
مثال : المثال التالى عبارة عن درجات عينة من الطلبة فى مادتى الرياضة والإحصاء ، ونريد أن نضيف متغيراً أخر عبارة عن مجموع المادتين :

الرياضة (Math)	٦٧	٦٣	٥٧	٥٢	٦١	٦٣	٥٠	٦٧	٦٩	٦٠
الإحصاء (Stat)	٥٨	٩٢	٦١	٧٤	٢١	٦٧	٦٨	٦٤	٦٥	٧٣
النوع (Sex)	٠	١	١	١	١	١	١	١	٠	١

١) سجل المتغيرات فى نافذة إدخال البيانات باسم sex , math , stat حيث ترمز stat لدرجة الإحصاء ، math لدرجة الرياضيات ، sex للنوع ، حيث الكود ٠ يشير إلى أنثى ، الكود ١ يشير إلى ذكر .

٢) من القائمة Transform اختر أمر Compute فيظهر الصندوق الحوارى

Compute Variable أمام خانة Target Variable أكتب اسم المتغير الجديد Total



٣) أمام خانة Numeric Expression أكتب العملية المطلوب إجراؤها

Math + Stat

٤) يمكن تحديد Type & Label للمتغير الجديد عن طريق اختيار Type & Label

٥) يمكن استخدام Function من الدوال الموجودة تحت خانة Functions

Total =sum(math , stat)

بعد اختيار أمر OK تظهر نافذة البيانات ويضاف إليها متغير جديد على النحو التالى :

	math	stat	sex	total				
1	67	58	0	125				
2	63	92	1	155				
3	57	61	1	118				
4	52	74	0	126				
5	61	21	0	82				
6	63	67	0	130				
7	50	68	0	118				
8	67	64	1	131				
9	69	65	0	134				
10	60	73	1	133				

إذا أردنا أن نحتفظ بالمتغيرات المحسوبة فى ملف البيانات فينبغى حفظ هذا الملف ، أمـــا إذا أغلقنا بدون حفظ فلن يتم حفظ هذه المتغيرات المحسوبة بالأمر Compute .

ملحوظة :

إذا تم التعديل فى البيانات ، يجب أن يعاد الحساب مرة ثانية ، حيث أنه لا يتم تحديث البيانات فى المتغيرات المحسوبة تلقائياً .

يمكن استخدام الـ Arithmetic Operators التالية :

+ للجمع (Addition) ، – للطرح (subtraction) ، * للضـــرب (Multiplication) ، /للقسمة (Division) ، ** للأس (Exponential) ، كما يمكن استخدام الأقواس للتحكم فى أولوية تنفيذ العمليات الحسابية

يراعى الأولوية التالية فى تنفيذ العمليات الحسابية :

١) فك الأقواس .

٢) رفع الأس

٣) الضرب أو القسمة أيهما أسبق

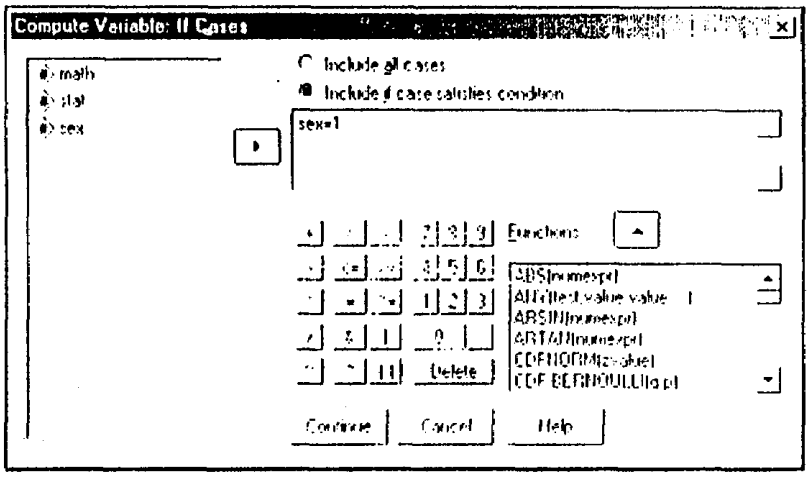
٤) الجمع أو الطرح أيهما أسبق .

التحويل المشروط للمتغيرات Compute variable if case :

التحويل المشروط للمتغيرات يسمح باختيار حالات معينة لإجراء التحويلة المطلوبـــة عليهـــا ، نتيجة شرط معين ، إذا كانت نتيجة الشرط True تجرى التحويلة ، إذا كانت نتيجة الشـــرط False لا تجرى التحويلة .

يتم تنفيذ ذلك بنقر If من الصندوق الحوارى Compute Variable .

فى المثال السابق يمكن اختيار عينة الذكور (male) فقط دون الإناث لذلك يتم تنشيط If من الصندوق الحوارى Compute



واختيار Include if cases satisfies condition ثم كتابة الشـــرط sex=1 ثم اختيـــار Continue ، يظهر هذا الشرط أمـــام خانـــة If ... فى الصـــندوق الحـــوارى Compute Variable ثم اختيار أمر OK يتم حساب الـــ total لبيانات الذكور فقط .

تستخدم الـــ Relational Operators التالية :

> أكبر من

>= أكبر من أو يساوى

< أصغر من

<= أصغر من أو يساوى

~= لايساوى

كما تستخدم الـــ Logical operators التالية :

Not ~ And & Or |

وتكون أولوية التنفيذ Not أولاً ، ثم And ، ثم Or ، وتستخدم فى حالة تعريف أكثر من شرط.

بعد تنفيذ Compute على متغير معين ، يمكن تعريف Type & Label للمتغير الذى يحتـــوى نتيجة الحساب .

الدوال Functions :

يقدم البرنامج العديد من الدوال Functions التى يمكن استخدامها فى الـــ Compute منها :

الوظيفة	اسم الدالة
القيمة المطلقة	ABS
أخذ الرقم الصحيح وبتر الكسر	TRUNC
الجذر التربيعى	SQRT
اللوغاريتم للأساس ١٠	LG10
اللوغاريتم للأساس (E)	LN
جيب الزاوية (جا)	SIN
جيب تمام الزاوية (جتا)	COS
الرقم للأساس (E)	EXP

ويراعى القيم المفقودة Missing Values فى العمليـــات الحســـابية ، لأن الـــــ Missing Values حينما تدخل فى العمليات الحسابية تكون النتيجة Missing .

إعادة تكويد المتغيرات Recoding Variables :

قد يوجد متغير ونحتاج إلى تقسيمه إلى فئات ، وذلك عن طريق إعطاء قيم جديدة لهـــذا المـــتغير تسمى Recoding ، وقد يكون ذلك التكويد بتغيير قيم المتغير نفسه إلى الأكواد الجديدة ، أو يكون التكويد داخل متغير جديد .

التكويد داخل نفس المتغير Recode into same variable :

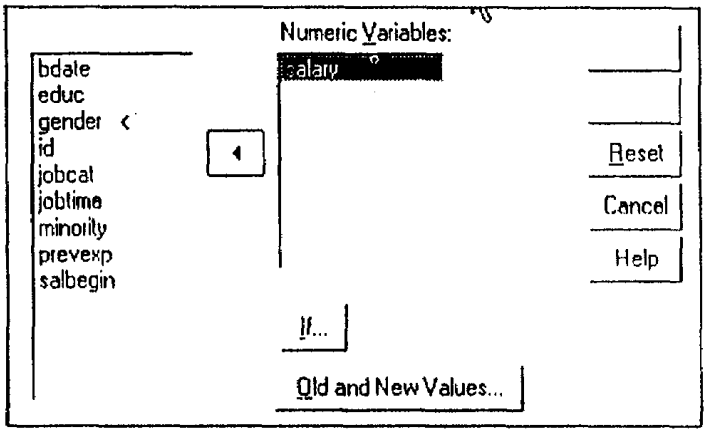
تحل الأكواد الجديدة محل القيم الفعلية :

مثال : يمكن عمل Recoding للمتغير salary (الرواتب) فى الملف Employee data.sav بتقسيمه إلى فئات .

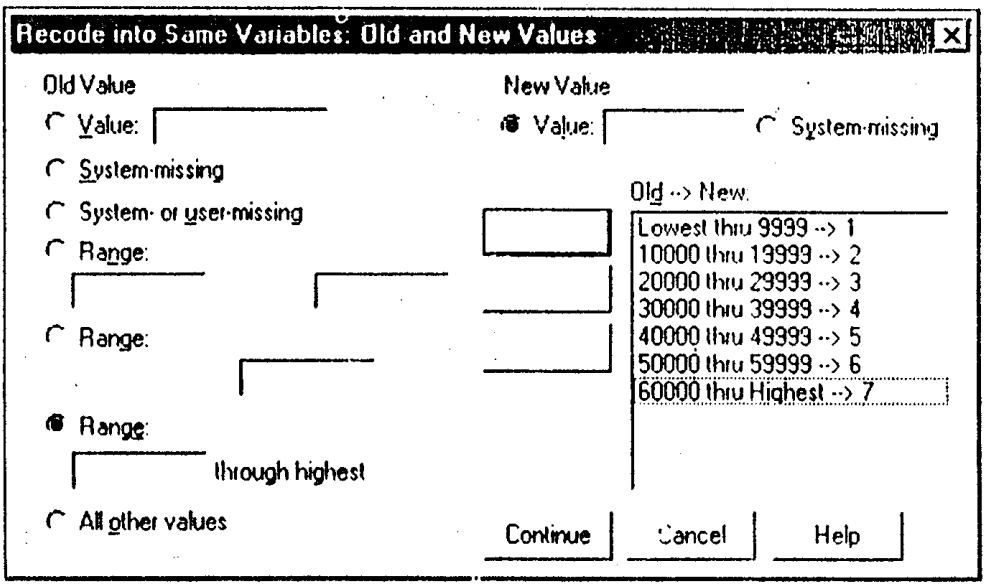
ولتكويد المتغير بتغيير القيم الفعلية يتم اتباع الآتى :

١) افتح الملف Employee data.sav

٢) من قائمة Transform اختر أمر Recode بنســـدل اختيـــاران ، اختـــر Into Same Variables يظهر الصندوق الحوارى التالى :



٢) أنقر Old and New Values يظهر الصندوق التالى :



٢) سجل القيمة الفعلية لنمتغير (أرالمدى المعين) المراد إعادة تكويده فى خانة Old Value ، ثم سجل الأكواد الجديدة فى خانة New Value ، ثم انقر أمر Add .

أثناء تسجيل قيم المتغير القديمة Old يمكن الاستعانة بالاتى :

Range : وذلك فى حالة تحديد مدة من البيانات مثل 10000 thru 19999

Lowest through : مثل Lowest thru 9999 أى من أقل قيمة إلى 9999 ، وذلك لتشمل جميع القيم الدنيا

Through highest : مثل 6000 thru Highest أى من 6000 إلى أعلى قيمة ، وذلك لتشمل جميع القيم العليا

٣) بعد الانتهاء من تسجيل الأكواد أنقر أمر Continue ، ثم اختر أمر OK يتم استبدال قيم المتغير الأصلية بالقيم المكودة ، وفى حالة الحفظ تختفى القيم الأصلية للمتغير وتفقد نهائياً .

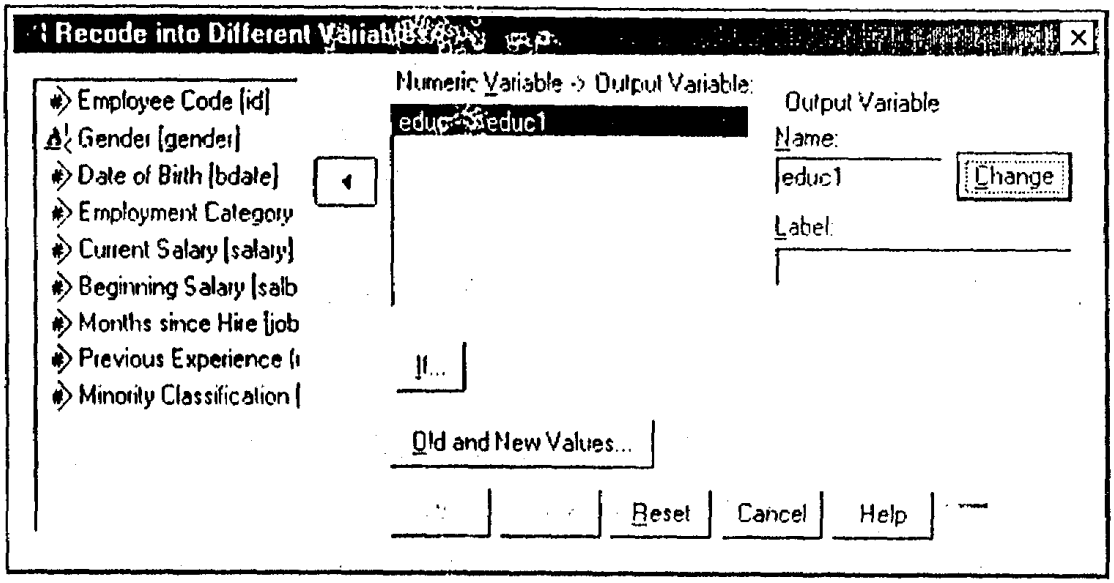
التكويد فى متغيرات جديدة Recode into Different Variables :

أحياناً نحب أن نكود المتغيرات مع الاحتفاظ بالقيم الفعلية لها ، لأن تكويد المتغيرات رغم أنه يؤدى إلى تلخيص البيانات وسهولة عرضها ، وسرعة استيعابها ، إلا أنه قد يؤثر على دقة المقاييس الإحصائية المترتبة على هذه البيانات .

ولتكويد المتغيرات داخل متغيرات جديدة نتبع الخطوات التالية :

١) افتح الملف Employee data.sav

٢) من قائمة Transform اختر أمر Recode تنسدل قائمة بها اختياران ، اختر



Into Different Variables ، يفتح الصندوق الحوارى التالى :

٢) فى خانة Input Variable اختر المتغير المراد تكويده .

٣) فى خانة Output Variable اكتب اسم المتغير الجديد الذى سيحتوى القيم المكودة ، ثم انقر أمر Change .

٤) أنقر أمر Old and New Values يفتح الصندوق الحوارى

Recode into Different Variables: Old and New Values

٥) سجل القيم الفعلية للمتغير الأصلى فى خانة Old Value ، ثم سجل القيم الجديدة (الأكواد) فى خانة New Value ، بنفس الطريقة السابقة ثم اختر أمر Ok

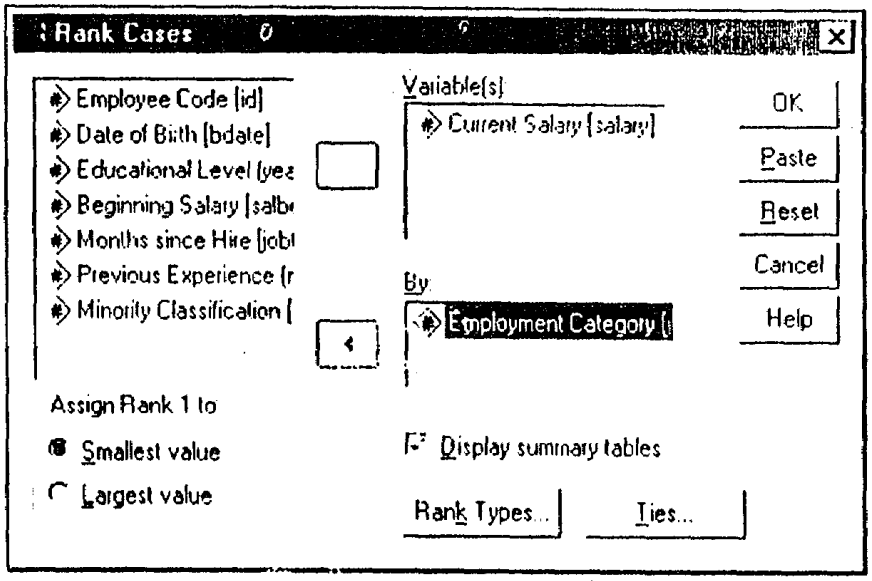
إعطاء رتب للحالات Rank Cases :

يمكن إعطاء رتب للحالات بالنسبة لمتغير معين ، البرنامج يضع هذه الرتب داخل متغيرات مناظرة للمتغيرات الأصلية ، يمكن أن تكون هذه الرتب تصاعدية أو تنازلية .

ولإعطاء رتب للحالات وفقاً لمتغير معين نتبع الآتى :

١) افتح الملف Employee data.sav

٢) من قائمة Transform اختر أمر Rank Cases ، يفتح الصندوق الحوارى

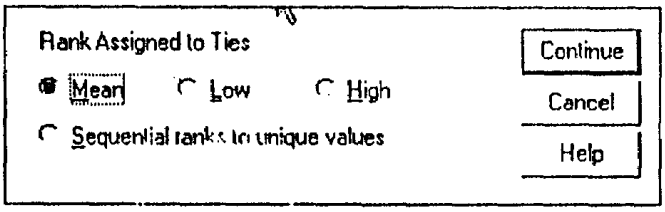


٣) أمام خانة Variable اختر المتغير المراد إعطاء رتب للحالات على أساسه ، وليكن المستغير salary .

٤) أمام خانة By اختر المتغير المراد تقسيم الرتب إلى مجموعات على أساسه وليكن Jobcat ، أى كلما تغيرت الـ Jobcat كلما بدأ الترتيب من رقم ١ .

٥) إذا أردنا أن تكون الرتب تصاعدية فإننا نختار Smallest value من خانة Assign Rank 1 to ، أما إذا أردنا أن تكون الرتب تنازلية فإننا نختار Largest Value

٦) إذا أردنا معرفة طريقة معالجة القيم المتشابهة فى الرتب ، فإننا نختار Ties يفتح الصندوق الحوارى التالى :



ولتوضيح الفرق بين هذه الأنواع فى معالجة القيم المتشابهة نعرض الجدول التالى

Value	Mean	Low	High	Sequential
١٠	١	١	١	١
١٥	٣	٢	٤	٢
١٥	٣	٢	٤	٢
١٥	٣	٢	٤	٢
١٦	٥	٥	٥	٣
٢٠	٦	٦	٦	٤

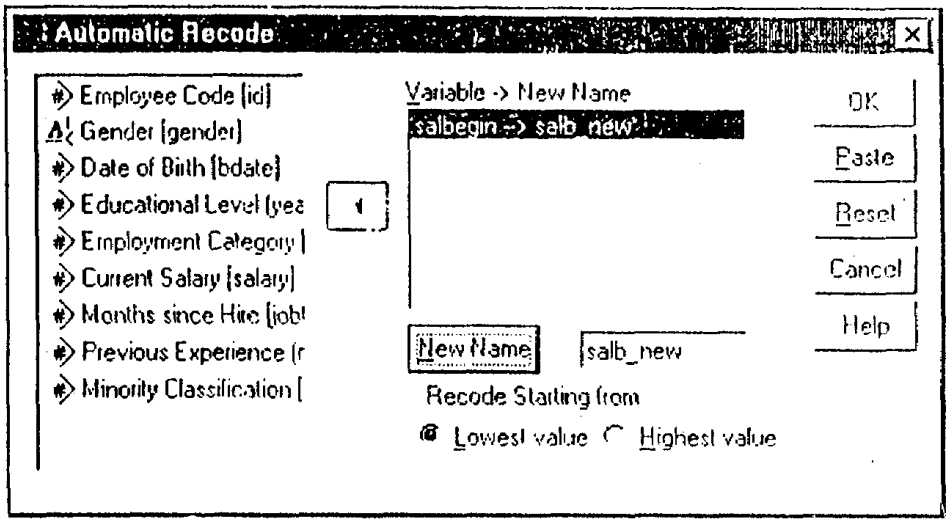
التكويد التلقائى Automatic Recode :

يقوم بترتيب البيانات وإعادة تكويدها فى متغير آخر اعتباراً من ١ فصاعداً إذا اخترنا Lowest Value ، وهو أشبه بالرتب ، ولعمل Automatic Recode لمتغير معين نتبع الآتى :

١) افتح الملف Employee data.sav

٢) من قائمة Transform اختر أمر Automatic Rcode يفتح الصندوق الحوارى

٣) أمام خانة Variable اختر المتغير المراد تكويده .



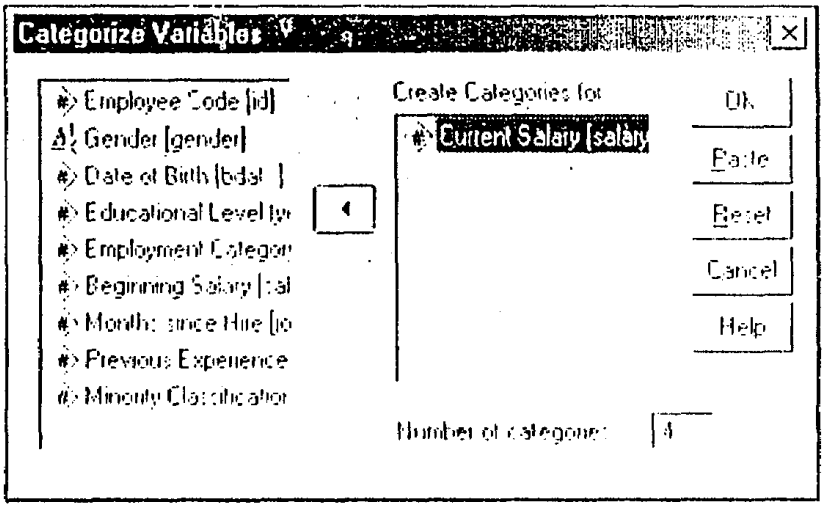
٣) أمام خانة New Name أكتب اسم المتغير الجديد ، ثم انقر أمر New Name

٣) اختر أمر Ok

تحويل متغير كمى الى فئات Categorize Variable

قد يوجد لدينا بيانات كمية مثل بيانات العمر أو الراتب ونريد تقسيمها إلى فئات Categories ، يمكن استخدام الأمر Recode السابق التعرف عليه ، كما يمكن استخدام أمر آخر وهو Categorize Variable

١- افتح الملف Employee data.sav .



٢- من قائمة Transform اختر أمر Categorize Variable

٢- اختر المتغير المراد تقسيمه إلى فئات .

٣- أمام خانة Number of categories اختر عدد الفئات المراد تقسيم المتغير إليها .

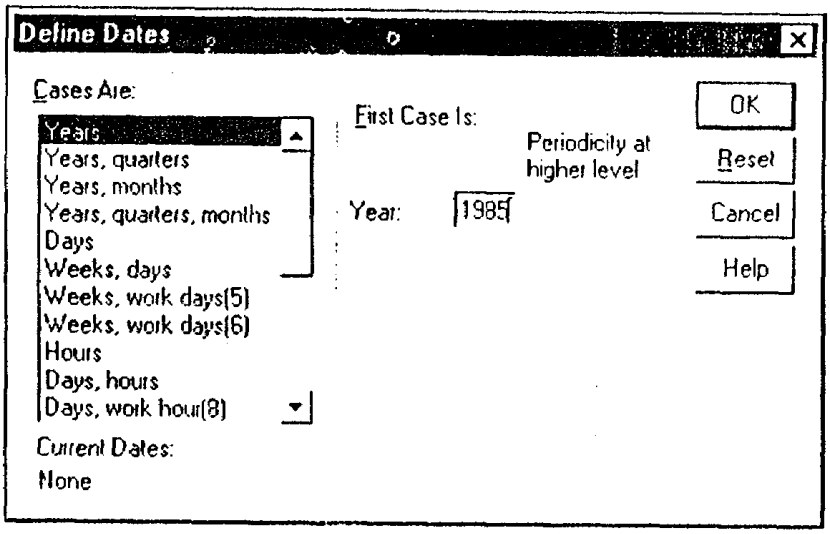
تعريف سلسلة زمنية Define Dates :

يمكن تعريف متغير على أنه سلسلة زمنية ، وذلك يفيد فى تحليل السلاسل الزمنية ، كما يفيد فى أن هناك وظائف معينة Functions يمكن إجراؤها على المتغيرات الزمنية

ولتعريف متغير على أنه سلسلة زمنية تبدأ من ١٩٨٥ نتبع الآتى :

١) افتح ملفاً جديداً .

٢) من قائمة Data اختر أمر Dates Define يظهر المربع الحوارى Define dates



٣) أمام خانة Cases Are اختر وحدة الزمن المطلوبة (Years , Weeks , Days , ... etc)

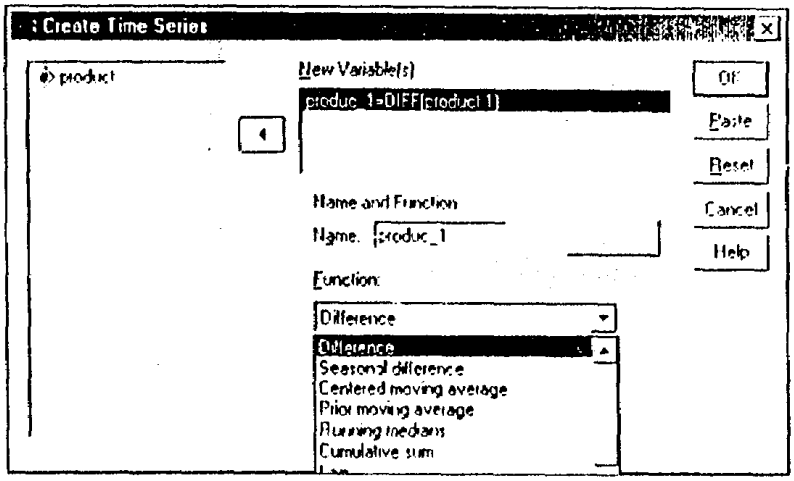
٤) أمام خانة First Case Is أكتب تاريخ البداية ، ثم أنقر أمر OK يظهر متغير جديد بملف البيانات عبارة عن سلسلة زمنية تبدأ بالرقم ١٩٨٥

إجراء عمليات على متغيرات زمنية Create Time Series :

لإيجاد متغيرات تعتمد على متغيرات فى سلسلة زمنية معينة مثل تطور قيمة الإنتاج الصناعى خلال عشر سنوات .

١) افتح الملف السابق .

٢) من قائمة Transform اختر Create Time Series يظهر الصندوق الحوارى Create Time Series



٢) اختر الـ Function من خانة Function ، يلاحظ أن هناك عدة أنواع من الـ Functions نذكر منها :

Difference : وهو الفرق بين القيمة الحالية والقيمة السابقة للمتغير الأصلى.

Lag : وهو القيمة السابقة للمتغير الأصلى (فى سنة سابقة)

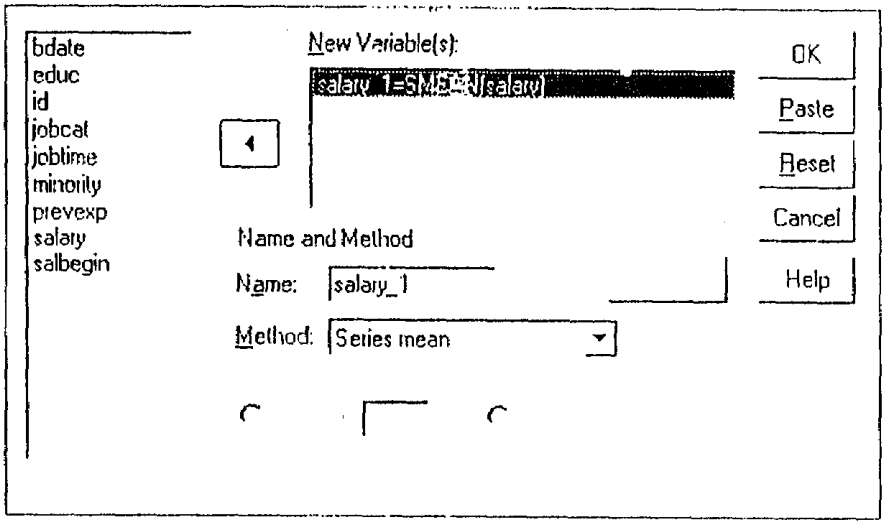
Led : وهو القيمة التالية للمتغير الأصلى (فى سنة تالية) ... وهكذا .

استبدال القيم المفقودة Repalcing Missing Values :

القيم المفقودة قد تسبب بعض المشاكل فى التحليلات الإحصائية خاصة فى تحليل السلاسل الزمنية ، لذلك قد نرغب فى استبدال هذه القيم الـ Missing باشتقاق متغيرات جديدة مع استبدال القيم المفقودة بقيم محسوبة من المتغيرات الأصلية بعدة طرق Methods ، ولتنفيذ ذلك نتبع الخطوات التالية:

١) افتح الملف Employee data.sav

٢) من قائمة Transform اختر Replace Missing Values يظهر الصندوق الحوارى التالى



٣) اختر طريقة التقدير من خانة Method .

٤) اختر المتغير المراد استبدال قيمه المفقودة .

٥) نجد أن البرنامج يحدد اسما للمتغير الجديد مشتقاً من المتغير الأصلى ، يمكن تغيير هذا الاسم .

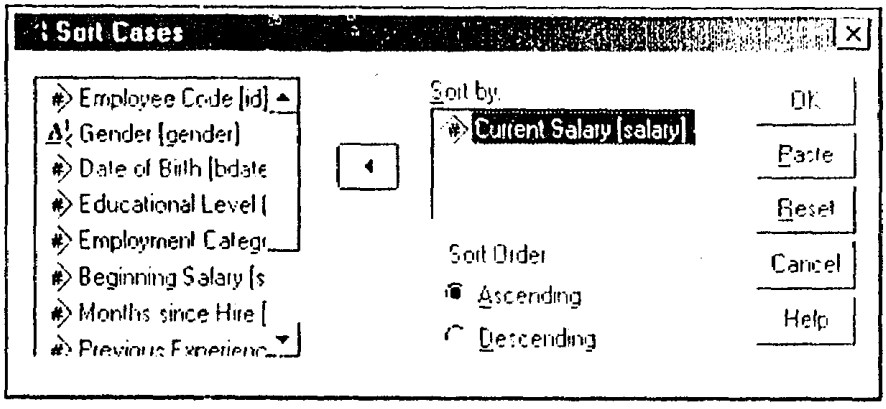
٦) اختر أمر Ok .

فرز البيانات Sort Cases :

قد نحتاج إلى فرز (ترتيب الحالات) تصاعدياً أو تنازلياً بناءً على قيم متغير معين ، أو عدة متغيرات ، ولتنفيذ ذلك نتبع الخطوات التالية :

١) افتح الملف Employee data.sav .

٢) من قائمة Data اختر أمر Sort Cases



٣) أمام خانة Sort by ، اختر المتغير أو المتغيرات المراد الترتيب على أساسها

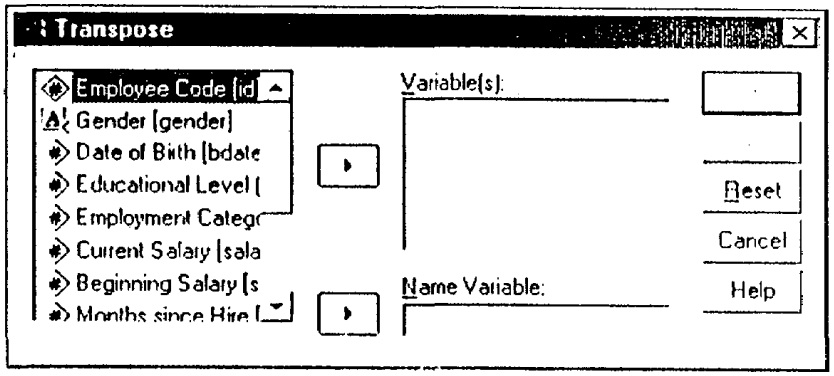
٤) أمام خانة Sort Order ، اختر نوع الترتيب (تصاعدى أو تنازلى) .

٥) اختر أمر OK يتم فرز البيانات بالترتيب المطلوب .

تدوير البيانات Transpose :

أى جعل الصفوف أعمدة ، والأعمدة صفوف .

قد تسجل البيانات بوضع لا يناسب التحليل الإحصائى ، لذلك تحتاج إلى قلب الصفوف أعمدة والأعمدة صفوف ، ولتنفيذ ذلك نتبع الآتى :



١) افتح الملف Employee data.sav

٢) من قائمة Data اختر أمر Transpose

٣) أمام خانة Variables اختر المتغيرات المراد عمل Transpose لها ، مع ملاحظة أن المتغيرات التى لايتم اختيارها سوف تفقد

٤) اختر أمر OK .

٥) يتم تدوير البيانات تصبح المتغيرات Cases وتأخذ أرقاماً مسلسلة ، وتصير الحالات متغيرات بأسماء var001,var002,... etc. ، إلى أن يتم إعادة تعريف هذه المتغيرات .

٦) يتم حفظ البيانات على ملف آخر ، حتى تحتفظ بالبيانات الأصلية قبل تدويرها

دمج الملفات Merging Data Files :

قد يكون حجم البيانات التى نجرى عليها التحليل كبيراً ولذلك يتم تجزئة البيانات إلى عدة ملفات ، ثم نحتاج إلى دمجها بعد ذلك ، ومن الممكن أن يتم تجزئة البيانات من حيث الحالات وفى هذه الحالة يكون الدمج عن طريق Add Cases ، أو عن طريق تجزئة المتغيرات وفى هذه الحالة يكون الدمج عن طريق Add Variables.

١- الدمج عن طريق إضافة حالات Add Cases:

مثال : تم تسجيل بيانات الطلبة فى مادتى الرياضيات والإحصاء على ملف male كالتالى :

math	٥٨	٩٢	٦١	٧٤	٢١	٦٧	٦٨	٦٤	٦٥	٧٣
stat	٦٧	٦٣	٥٧	٥٢	٦١	٦٣	٥٠	٦٧	٦٩	٦٠
sex	١	١	١	١	١	١	١	١	١	١

وتم تسجيل بيانات الإناث على ملف female كالتالى :

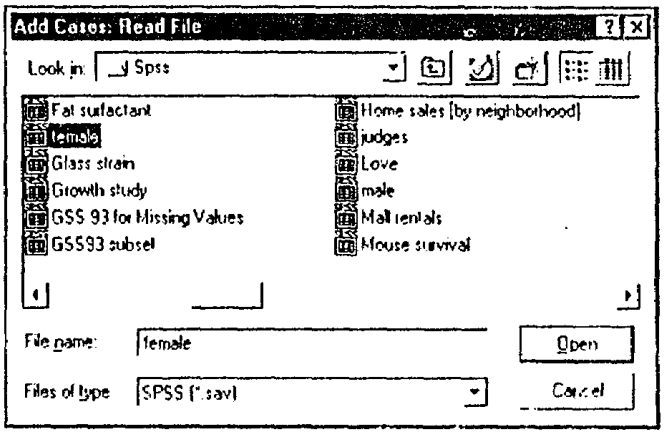
math	٣٠	٦٥	٥٥	٧٤	٥٧	٩٥	٨٠	٧١	٤٣	٥٢
stat	٦١	٧١	٥٩	٣٦	٧٠	٩٠	٥٩	٦٣	٦٧	٥٦
sex	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢

ونريد دمج الملفين معاً ، ولدمج الملفين عن طريق Add Cases يفضل أن تكون متغيرات الملف الأول هى نفس متغيرات الملف الثانى ، كما يفضل ترتيب بيانات الملفين بطريقة واحدة قبل الدمج ، ولدمج حالات الملفين نتبع الآتى :

١- افتح الملف الأول male

٢- من القائمة الرئيسية اختر Data ومنها اختر Merge Files ومنها اختر Add Cases

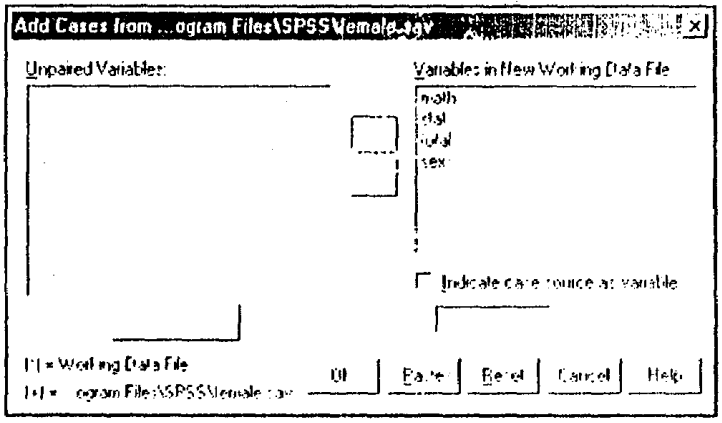
يظهر الصندوق الحوارى التالى :



٣- اختر اسم الملف المطلوب دمجه مع الملف الأول ، ثم اختر أمر Open

٤- يظهر الصندوق الحوارى وبه اسم الملف ، وبه أسماء المتغيرات فى الملف الجديد بعد الدمج فى خانة Variables in New Working Data File

٥- يمكن تنشيط الاختيار Include case source as variable لإضافة متغير جديد به رقم يعبر عن الملف (مصدر البيانات) .



٦- اختر أمر OK يتم دمج بيانات الملفين معاً

٢- دمج البيانات : Add Variables

يفضل فى هذه الحالة أن تتساوى عدد الحالات ، فإذا لم تتساوى عدد الحالات فإن النظام يضع للحالات الأقل (.) system missing حتى تتساوى مع عدد الحالات الأكبر

مثال :

١- أدخل البيانات التالية فى ملف A1 وهى عبارة عن بيانات الطول والوزن :

No	١	٢	٣	٤	٥
height	١٦٠	١٨٠	١٧٠	١٧٥	١٦٥
weight	٦٢	٩٠	٨٠	٦٧	٧٢

٢- ثم أدخل بيانات العمر فى ملف آخر A2 :

No	١	٢	٣	٤	٥
age	٣٠	٢٠	٢٢	٢٥	٢٦

ويراد دمج بيانات الملفين معاً ، ولتنفيذ ذلك نتبع الآتى :

١- افتح بيانات الملف الأول a1 .

٢- من قائمة Data اختر Merge Files ، ومنها اختر Add Variables .

٣- يفتح الصندوق الحوارى Add Variables : Read File

٤- اختر اسم الملف a2 واختر أمر Continue

٥- يظهر الصندوق الحوارى Add Variables from وبه اسم ومسار الملف الثانى a2 .

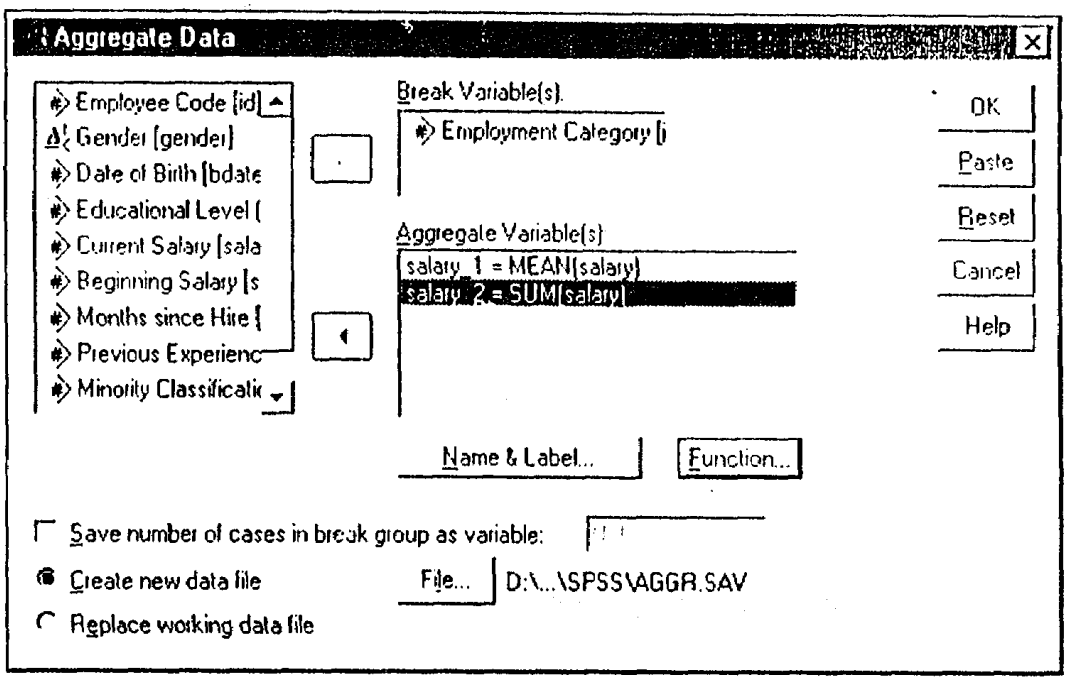
٦- اختر أمر Continue .

حساب مقاييس إحصائية للبيانات مقسمة إلى مجموعات Aggregate Data :

يمكن تقسيم البيانات إلى مجموعات (حسب متغير معين) وحساب مقاييس معينة لهذه المجموعات مثل (المتوسط ، الانحراف المعيارى ... الخ) ، ويمكن أن يكون هذا التقسيم على نفس الملف أو على ملف آخر ، البرنامج يخصص ملف Aggr.sav تلقائياً ويمكن تغييره ، ولعمل ذلك نتبع الآتى :

١- افتح الملف Employee data.sav .

٢- من قائمة Data اختر أمر Aggregate يفتح الصندوق الحوارى التالى :



٢- أمام خانة Break يتم اختيار المتغير الذى يتم تجميع البيانات على أساسه ، وكل قيم من قيم هذا المتغير ستناظر حالة واحدة من البيانات بعد التجميع .

٣- أمام خانة Aggregate Variable(s) يتم اختيار المتغيرات التى سيتم تجميعها ويخصص البرنامج لها أسماء مشتقة من المتغيرات الأصلية ، ويمكن إعادة تسمية هذه المتغيرات عن طريق اختيار أمر Name & Label .

٤- أنقر أمر Function يتم اختيار المقاييس الإحصائية المراد حسابها لهذه البيانات المجمعة .

٥- اختر أمر تظهر البيانات فى ملف Aggr.sav يمكن فتحه فيظهر على النحو التالى

jobcat	salary_1	salary_2
Clerical	27838.54	10105390
Custodial	30938.89	835350.0
Manager	63977.80	5374135

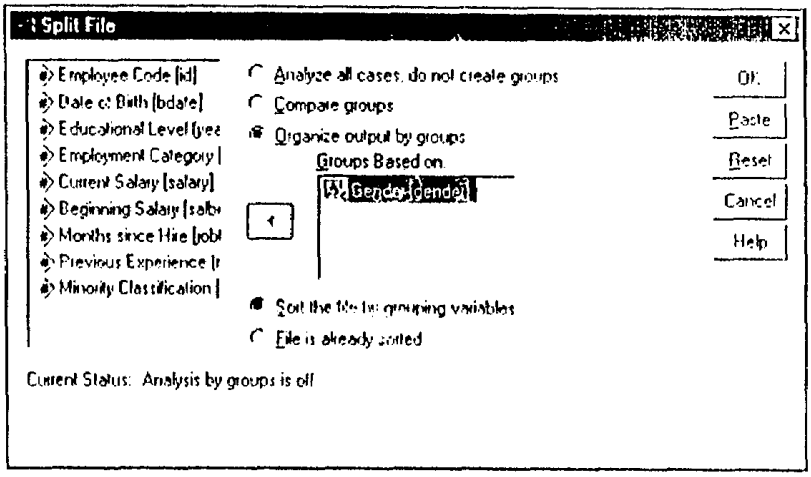
تقسيم الملفات Split Files :

يمكن تقسيم ملف البيانات إلى مجموعات منفصلة لأغراض التحليل بناءً على قيم متغير معين مثل النوع ولتنفيذ ذلك نتبع الآتى :

١) افتح الملف Employee data.sav

٢) من قائمة Data اختر أمر Split File

يظهر الصندوق الحوارى Split File .



٣) نشط خانة Organize output by groups .

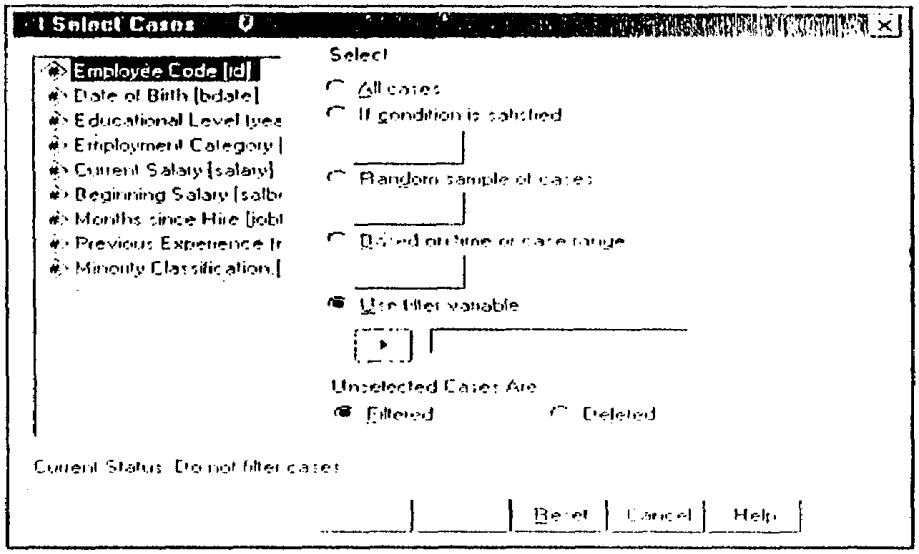
٢) اختر المتغير المراد التقسيم على أساسه أما خانة Groups based on ، ثم اختر أمر OK

اختيار حالات Select Cases:

قد نكون مهتمين بعرض وتحليل حالات معينة من البيانات ، وليس كل البيانات ولتنفيذ ذلك نتبع الآتى :

١) افتح الملف Employee data.sav

٢) من قائمة Data اختر أمر Select Cases



يظهر الصندوق الحوارى Select Cases .

ويظهر أمامه الاختيارات التالية :

١) الاختيار All Cases وذلك فى حالة اختيار كل الحالات

٢) الاختيار المشروط وفى هذه الحالة يتم تنشيط If condition is satisfied ، ونقر if وكتابة الشرط المطلوب اختيار الحالات على أساسه ، مثلاً فى حالة اختيار الذكور يتم كتابة gender= 'm' فى الصندوق الحوارى select case if

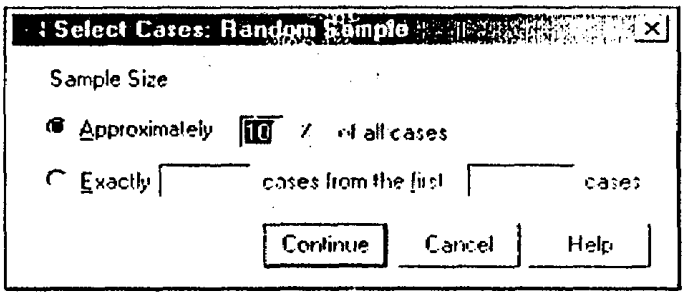
يتم اختيار بيانات الـ Male فقط أما بيانات الـ Female فيظهر بها شطب على النحو التالى

	id	gender	bdate	educ	jobcat	salary	salbegin
1	2	Male	05/23/58	16	Clerical	$40,200	$18,750
2	3	Female	07/26/29	12	Clerical	$21,450	$12,000
3	4	Female	04/15/47	8	Clerical	$21,900	$13,200
4	5	Male	02/09/55	15	Clerical	$45,000	$21,000
5	6	Male	08/22/58	15	Clerical	$32,100	$13,500
6	7	Male	04/26/56	15	Clerical	$36,000	$18,750
7	8	Female	05/06/66	12	Clerical	$21,900	$9,750
8	9	Female	01/23/46	15	Clerical	$27,900	$12,750
9	10	Female	02/13/46	12	Clerical	$24,000	$13,500
10	11	Female	02/07/50	16	Clerical	$30,300	$16,500
11	12	Male	01/11/66	8	Clerical	$20,350	$12,000

– إذا تم اختيار Filtered فيكون هذا الاختيار اختياراً مؤقتاً ، ويمكن بعد ذلك اختيار كل الحالات ، ويضيف إلى البيانات متغيراً يأخذ القيمة ١ فى حالة البيانات المختارة ، صفر فى حالة البيانات غير المختارة. إما إذا اخترنا Deleted فيكون هذا الاختيار اختياراً دائما ولا يمكننا بعد ذلك اختيار كل الحالات

اختيار عينة عشوائية :

١) لاختيار عينة عشوائية يتم اختيار sample من الصندوق الحوارى select Cases وذلك بعد تنشيط الاختيار Random sample of cases كالتالى :

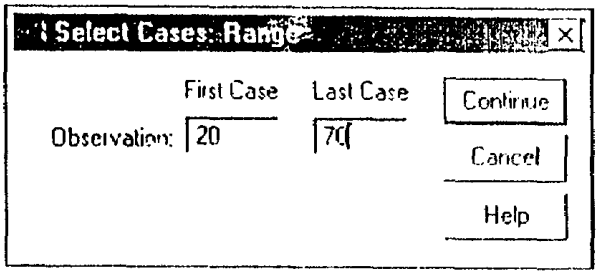


٢) يتم تحديد حجم العينة إما نسبة من البيانات ، أو عدد حالات من إجمالى الحالات .

تحديد مدى معين من البيانات

٣) لتحديد مدى معين من البيانات يتم اختيار Range وذلك بعد تنشيط المجموعة Based on time or case range من الصندوق الحوارى Select cases

وأمام خانة First Case , Last Case يتم اختيار مدى البيانات المراد التحليل على أساسه .



يلاحظ أن التحليل سيقتصر على البيانات المختارة فقط (والبيانات الغير مختارة سيظهر عليها شطب أمام رقم الحالة) فإذا أردنا أن يتم التحليل على كل البيانات فلابد من اختيار All Cases .

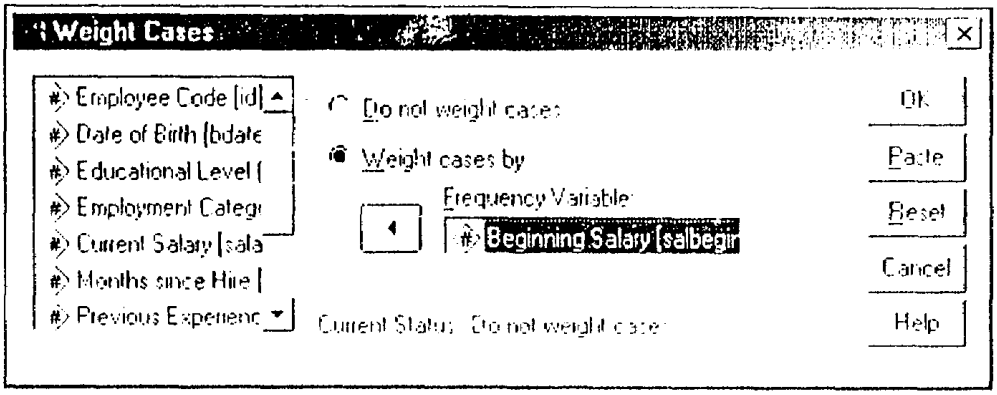
ملحوظة

إذا أردنا التعامل مع البيانات بعد ذلك بدون اختيار بيانات معينة ، فيتم اختيار All Cases من الصندوق الحوارى Select Cases

ترجيح الحالات :

فى بعض التحليلات نحتاج لترجيح البيانات خاصة إذا كانت البيانات مبوبة فى جداول تكرارية فيتم ترجيح البيانات بأوزانها أى بتكراراتها ويتم ذلك عن طريق :

١) من قائمة Data اختر أمر Weight Cases يفتح الصندوق الحوارى Weight Cases



٢) نشط الاختيار Weight Cases by ، ثم اختر المتغير المراد الترجيح به فى خانة Frequency Variable .

٣) إذا أردنا بعد ذلك التعامل مع البيانات بدون ترجيح يتم اختيار Do not weight cases

الباب الثانى
الأسلوب الإحصائى وعرض وتحليل البيانات

٢-١ : تحديد الأسلوب الاحصائى المناسب لتحليل البيانات :-

إن أصعب مرحلة تواجه الباحث فى بحثه هى مرحلة التحليل الاحصائى وذلك لكثرة تعداد الطرق الإحصائية . فالباحث وهو يستعرض مجموعة كبيرة من مختلف الطرق الإحصائية سيقف حائراً ولا يدرى أيا منها يختار مالم يكن لديه معايير ومعلومات مسبقة يستنير بها فى هذا الاختيار ، وان عدم الأخذ بهذه المعايير أو المعلومات تجعل الباحث فى النهاية يسئ استخدام الإحصاء فيستخدم بالتالي طرقا لا تتناسب مع طبيعة بيانات بحثه، ويمكن التغلب على هذه المشكلة بسهولة إذا استطاع الباحث مقدماً التعرف أيا من الطرق الإحصائية يناسب بياناته ، أى الطرق المعلمية أم الطرق اللامعلمية ؟ وفيمايلى تفصيل لكل منهما :-

أولاً : الطرق المعلمية (parametric)

الإحصاءات المعلمية هى تلك الطرق التى تتطلب الوفاء بافتراضات معينة حول المجتمع الذى تسحب منه العينة، ومن هذه الافتراضات أن تتخذ المشاهدات فى المجتمع شكل التوزيع الطبيعي على سبيل المثال لا الحصر . وقد أشتق مصطلح " معلمية " من مفهوم " معلم " الذى يعنى صفة أو خاصية من خصائص مجتمع معين ، فكل قيمة من القيم التى تتعلق بخصائص المجتمـــــع تسمى **parameter** أما تلك الخصائص المتعلقة بالعينة التى سحبت عشوائياً من المجتمع فتسمى كل منها تقدير **Estimate** أى أن القيمة المستخرجة لأى خاصية فى العينة ماهى إلا تقدير لقيمة تلك الخاصية فى المجتمع والتى على الأغلب تكون غير معروفة ، إذ أنها لو كانت معروفة لاتكون هناك حاجة لحساب تقديرها. ومن الأمثلة على هذه الخصائص الوسط الحسابى، والانحراف المعياري وشكل توزيع المجتمع .

والاختبارات الإحصائية المعلمية هى التى تعتمد على الافتراضات الخاصة بخصائص المجتمع مثل اختبار (ت) ، واختبار (ف) " وتحليل التباين الاحادى وتحليل التباين المتعدد.

بالإضافة إلى ما سبق يجب أن يكون اختيار العينة من المجتمع بصورة عشوائية ، وأن تكون إحصاءات العينة (مقاييس النزعة المركزية والتشتت) صورة مقربة للمعلمات الإحصائية للمجتمع، وكذلك تستخدم الطرق الإحصائية المعلمية لمعالجة وتحليل البيانات الكمية.

وكذلك يفترض أن يكون حجم العينة ليس صغيراً جداً لأن الحجم الصغير للعينة يؤثر على خصائص التوزيع التكرارى للعينة الصغيرة ، فتبعد بذلك عن اعتدالية التوزيع التكرارى للمجتمع.

وتكون كذلك الطرق المعلمية أكثر ملائمة لتحليل البيانات الفترية والبيانات النسبية فضلاً على أنها يمكن استخدامها (مع بعض التحفظ) لتحليل البيانات الرتبية فقط، وذلك بعد أن يتم تحويلها إلى بيانات فترية من خلال إعطاء كل رتبة درجات تتناسب مع قيمتها مع مراعاة أن تكون البيانات فى صورة رتب ذات درجات متتابعة مثل موافق بشدة ، موافق ، غير موافق ، غير موافق بشدة.

ثانيـــــاً : الطرق اللامعلمية (Non Parametic)

الإحصاءات اللامعلمية هى تلك الطرق التى تستخدم فى تحليل البيانات واختبار الفرضيات الخاصة بالبيانات الاسمية والرتبية والتى تكون من النوع المتقطع (المنفصل) عادة ويمكن استخدامها مع البيانات الفترية والنسبية بعد ان يتم تحويلها الى بيانات أسمية أو رتبية. وفى حالة الإحصاءات اللامعلمية يجب أن تكون العينة المختارة عشوائية بمعنى أن يكون هناك استقلال بين مفردات العينة. هذا فضلاً عن أن هذه الطرق لاتتقيد بالشروط الواجب توافرها لاستخدام الإحصاءات المعلمية حيث أنها تستخدم فى الحالات التى لا يكون فيها التوزيع النظرى للمجتمع الاصلى الذى اختيرت منه العينة معروفاً، وفى حالة عدم إمكانية الوفاء بافتراض أن التوزيع النظرى للمجتمع طبيعيا ويكون كذلك حجم العينة صغيراً جداً .

وتحتوى الحزمة SPSS على مجموعة متميزة من الاختبارات اللامعلمية. وسوف نستعرض هذه الاختبارات وفقاً إذا كان الاختبار يتعلق بعينة أو عينتين أو أكثر (مستقلتين أو غير مستقلتين) .

أولاً : اختبار عينة :

يوجد أربعة اختبارات لعينة واحدة الأول هى

- اختبار كا^{2}
- اختبار ذى الحدين
- اختبار العشوائية

- ثم أخذنا كولو مجروف

ويمكن إجرائها كالتالى :-

١- اختبار كا^{٢} يعد من أشهر الاختبارات اللامعلمية وهناك ثلاثة شروط لتطبيقه كا^{٢} وهى :

١- أن تكون العينة عشوائية

٢- استقلال المشاهدات.

٣- أن حجم العينة أكثر من ٣٠

وفى هذا الاختبار نختبر هل هناك فرقاً معنوياً بين التكرارات المشاهدة والتكرارات المتوقعة التى تم الحصول عليها طبقاً لقاعدة أو نظرية معينة ويكون الفرض العدمى أن التكرار النظرى يتبع التكرار المشاهد والفرض البديل أنه لايتبعه.

٢- اختبار ذى الحدين

يستخدم هذا الاختبار فى معرفة هل متغير ثنائي يتبع عشوائيا توزيع ذي الحدين أم لا؟ الفرض العدمى فى هذه الحالة أن البيانات تتبع توزيع ذي الحدين بمعلمة ٥ر أو بأي معلمة يحددها الاختبار و الفرض البديل أن البيانات لاتتبع ذي الحدين.

٣- اختبار العشوائية

يستخدم اختبار العشوائية لمتغير يقاس بمقياس اسمي، لإجراء ذلك الاختبار يحدد رقماً معيناً سوف يستخدمه برنامج SPSS فى تقسيم قيم المتغير إلى مجموعتين .

القيم الأقل منه تعامل معاملة واحدة وكذلك القيم التى تساوى أو تزيد عنه. الفرض العدمى يكون أن العينة عشوائية أو وسيط العينة يساوى الرقم المحدد.

٤- اختبار كولومجروف سيمنروف لعينة واحدة:

يهدف هذا الاختبار إلى معرفة هل البيانات المتاحة تتوزع حسب توزيع معين أم لا؟ يتولى البرنامج توفيق البيانات أولاً إلى احد التوزيعات الأربعة الآتية ثم اختبار جودة التوفيق بمعنى هل اختبار التوزيع موفقاً أم لا :

١- التوزيع الطبيعي

٢- التوزيع المنتظم

٣- التوزيع الأسى

٤- توزيع بواسون

ويمكن اعتبار هذا الاختبار من اختبارات جودة التوفيق للتوزيعات الأربعة المشار إليها.

ثانياً : اختبار عينتين غير مستقلتين :-

إذا كان لدينا عينتين مستقلتين فانه يمكن معرفة هل هناك فرقاً بينهما أم لا وذلك باستخدام مجموعة من الاختبارات اللامعلمية حيث كل اختبار يقيس الاختلاف من وجه نظر معينة. والاختبارات اللامعلمية الأربعة هي:-

١- اختبـــــار مان وتينى. Mann-whithney

٢- اختبار كولومجروف سيمنروف. Kolomogerov-Smirnov

٣- اختبار موزيس للقيم الشاذة. Moses Extreme Reactions

٤- اختبار والد للدورة. Wald-Wolfowitz

ثالثاً : اختبار عينتين مستقلتين :-

إذا كان لدينا عينتين غير مستقلتين وأردنا إجراء اختبار لامعلمى وذلك لمعرفة هل هناك اختلاف بين العينتين أم لا؟ يمكن إجراء ثلاثة اختبارات فى هذه الحالة وهى :-

١- اختبار ولكوكسن Wilcoxon Test

٢- اختبار الإشارة Sign Test

٣- اختبار ماكنمار MCNemar Test

رابعاً : اختبار أكثر من عينتين مستقلتين :-

إذا كان لدينا أكثر من عينتين مستقلتين وكانت احد الشروط اللازمة لتطبيق اختبار تحليل التباين غير مستوفاة فإن يمكن إجراء تحليل التباين لامعلمى للرتب وذلك باستخدام اختبار يطلق عليه أسم كيرسكال ويلز حيث يستخدم للفرق بين رتب أكثر من عينتين مستقلتين وهو يعتبر الصورة العامة لاختبار مان ويتنى ويجرى الاختبار تحت الفروض التالية :-

١. نفرض أن لدينا K من العينات المستقلة ذات الإحجام $n_1, n_2 .... n_\mu$ ونعنى بالاستقلال انه بين وداخل العينات .

٢. المتغيرات محل الدراسة متغيرات مستمرة وأن وحدة القياس على الأقل ترتيبية.

٣. المجتمعات المسحوبة منها العينات متطابقة فيما عدا أن مجتمع واحد على الأقل مختلف فى مقياس الموضع.

الفرض العدمى والفرض البديل كالآتى:-

الفرض العدمى : المجتمعات لها نفس الوسيط.

الفرض البديل : المجتمعات ليس لها نفس الوسيط.

خامساً : اختبار أكثر من عينتين غير مستقلتين :-

نفرض أن لدينا أكثر من عينتين غير مستقلة ونريد معرفة هل هناك فرقاً معنوياً بينهما أم لا؟

فيكون الفرض العدمى والفرض البديل كالتالي:-

الفرض العدمى : لا يوجد فرق بين العينات.

الفرض البديل : يوجد زوج واحد من العينات على الأقل الفرق بينهما معنوي.

ويوجد أكثر من اختبار فى هذه الحالة وهى :-

١- اختبار كندال.

٢- اختبار كوكران.

٣- اختبار فريدمان.

وفيما يلى عرض تلخيص للمقارنة بين الطرق المعلمية واللامعلمية

الطرق الإحصائية المعلمية	الطرق الإحصائية اللامعلمية
- يتطلب استخدام الطرق المعلمية الوفاء بافتراضات معينة حول التوزيع الاساسى للمجتمع، من حيث أن يكون توزيع المجتمع طبيعياً.	- لايتطلب استخدام الطرق اللامعلمية ايه افتراضات أو معلومات حول خصائص التوزيع الاساسى للمجتمع. (لذلك يطلق بعض الإحصائيين عليها إحصاءات التوزيعات الحرة).
- لايمكن استخدام الطرق الإحصائية المعلمية لمعالجة وتحليل البيانات فى المواقف التجريبية التى يكون فيها حجم العينة صغير جداً، لأن الحجم الصغير للعينة يؤثر على خصائص التوزيع التكرارى للعينة الصغيرة ، فتبتعد بذلك عن اعتدالية التوزيع التكرارى للمجتمع الأب.	- يمكن استخدام بعض الطرق اللامعلمية لمعالجة وتحليل البيانات فى المواقف التجريبية التى يكون فيها حجم العينة صغير جداً.
- تكون الطرق المعلمية أكثر ملائمة لتحليل البيانات الفترية والبيانات النسبية فضلاً عن أنها يمكن استخدامها (مع بعض التحفظ) لتحليل البيانات الرتبية فقط وذلك بعد أن يتم تحويلها إلى بيانات فترية من خلال إعطاء كل	- تكون الطرق اللامعلمية عادة أكثر ملاءمة للاستخدام عندما تكون البيانات الخاصة بالبحث من النوعين الأسمى والرتبى فضلاً عن أنها يمكن استخدامها أحياناً عندما تكون البيانات فترية أو نسبية ، وذلك بعد أن

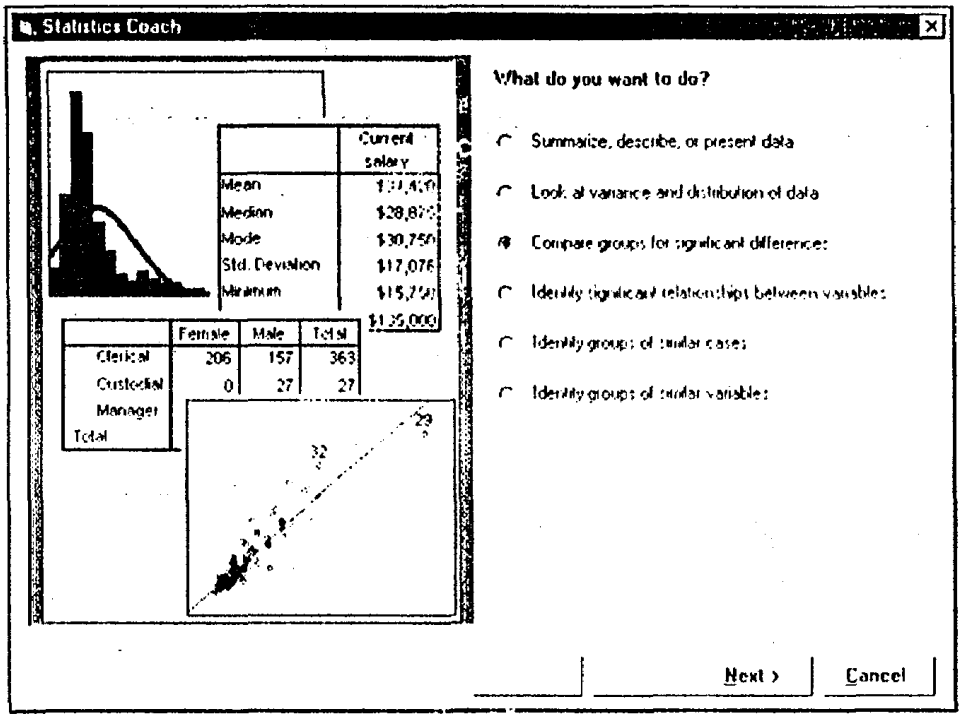
الطرق الإحصائية المعلمية	الطرق الإحصائية اللامعلمية
رتبة درجات تتناسب مع قيمتها ، مع مراعاة أن تكون البيانات فى صورة رتب ذات درجـــات متتابعة مثل : موافق بشدة، موافق، غير موافق، غير موافق بشدة.	يتم تحـــويلها إلى بيانات اسمية أو رتبية (وهذا فيه إهـــدار لجزء كـــبير من البيانات التى نحللها ، لأن الطرق الإحصائية اللامعلمية تستخدم اجزاء يسيراً من تلك البيانات) لذلك أوضح بعض الإحصائيين انه عندما نستخدم الطرق الإحصائية اللامعلمية للبيانات الفترية والنسبية فإنها لاتستخدم إلا للتقدير المبدئى، على أن يتلوه بعد ذلك استخدام الطرق المعلمية.
– تعتبر الطرق الإحصائية المعلمية بشكل عام أقوى من الطرق الإحصائية اللامعلمية، حيث أن الطرق المعلمية تميل إلى رفض الفرضية الصفرية أكثر من ميل الطرق اللامعلمية لرفض نفس الفرضية.	– بالرغم من تحرر الطرق الإحصائية اللامعلمية من الشروط والخصائص التى قـــد تعـــوق أحياناً استخدام الطرق الإحصائية المعلمية، إلا أن الطرق الإحصائية اللامعلمية بشكل عام أقل قوة من الطرق الإحصائية المعلمية
– تعتمد الطرق الإحصائية المعلمية بشكل عام على الدرجات الأصلية والتى يتم تحليلها كما هى .	– تعتمد الطرق اللامعلمية فى اغلب الأحيان على البيانات التى هى بشكل تكرارات أو رتب مما يؤدى إلى ضياع بعض المعلومات المفيدة.
– تعتبر الطرق الإحصائية المعلمية بصورة عامة أصعب فى الاستخدام من الطرق الإحصائية اللامعلمية ، وبالتالى فإن الوقت الذى يحتاجه الباحث لتحليل بياناته يكون أطول ، مما يؤدى إلى تأخـــر الحصول. عـــلى النتائج (نسبياً) والإفادة منها تطبيقياً.	– تعتبر الطرق الإحصائية اللامعلمية أفضل وسيلة للتقدير المبدئى السريع، فهى بصورة عامـــة أسهـــل استخداماً من الطـــرق المعلمية وبالتالى فإن الوقت الذى يحتاجـــه الباحـــث لتحليل بيـــاناته. يكـــون اقل، مما يؤدى إلى الإسراع فى الحصول على النتائج والإفادة منها تطبيقياً.
– تعتبر الطرق الإحصائية المعلمية أكثر شيوعاً واستخداماً لدى المختصين فى الإحصاء.	– تعتبر الطرق الإحصائية اللامعلمية أكثر شيوعاً واستخداماً لدى غير المتخصصين فى الإحصاء.
– ينبغى أن يكون اختبار العينة من المجتمع بصورة عشوائية، وأن تكون إحصاءات العينة (مقاييس النزعة المركزية والتشتت) صورة مقربة للمعلمات الإحصائية للمجتمع.	– ليس هناك ضرورة أن يكون اختيار العينة من المجتمع بصورة عشوائية.

الطرق الإحصائية المعلمية	الطرق الإحصائية اللامعلمية
– تستخدم الطرق الإحصائية المعلمية لمعالجة وتحليل البيانات الكمية.	– تستخدم الطـــرق الإحصائية اللامعلمية لمعالجة وتحليل البيانات النوعية، والتى لا يمكن عادة استخدام أى طريقة إحصائية معلمية لتحليلها.
– من الأمثلة على الطرق الإحصائية المعلمية اختبار (ز) واختبار (ت) واختبار (ف).	– من الأمثلة على الطرق الإحصائية اللامعلمية اختبـــار(كا^{2})واختبار مان ويتنى واختبـــار كروسكال واليز.

تحديد الأسلوب المناسب للتحليل الإحصائى باستخدام البرنامج الإحصائى SPSS

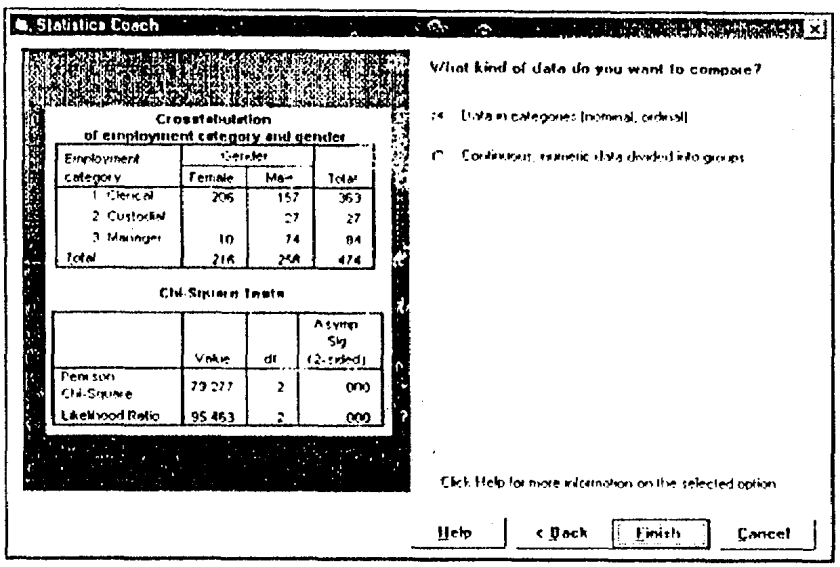
يمكن الاستعانة بالبرنامج فى تحديد الأسلوب المناسب للتحليل كالتالى :

١– من قائمة Help اختر أمر Statistics Coach يظهر المعالج التالى

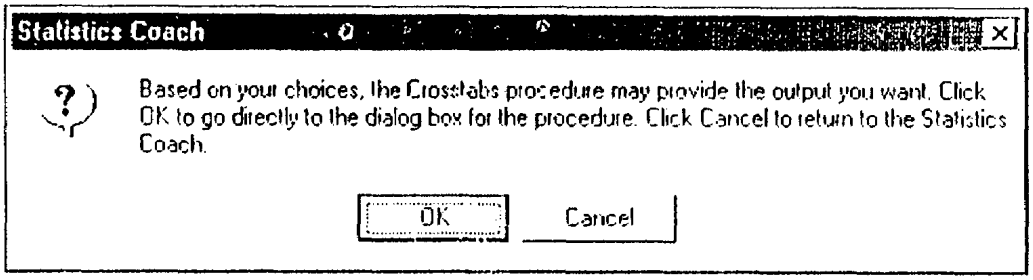


٢– اختر الموضوع الذى تريده وليكن

Compare groups for significant differences من قائمـــة الموضـــوعات ، ثم اختر أمر Next يظهر الصندوق التالى

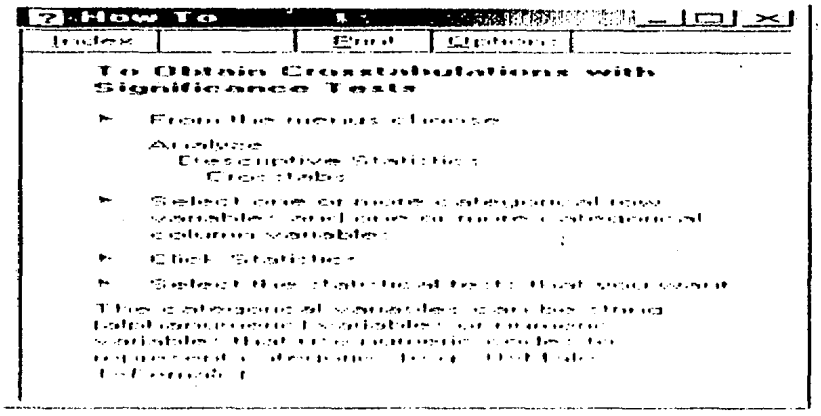


٣- يسأل هل البيانات وصفية Data in categories (nominal , ordinal) فيكون التحليل المناسب لها هو Cross tabulation , Chi square ويظهر نموذج من هذا التحليل فى خانة المعاينة



٤- وإذا كانت البيانات غير وصفية يمكن اختيـــار Continuous , Numerical Data Divided Into groups ، واختيار التحليل المناسب .

٥- اختر أمر Finish يحدد البرنامج طريقة إجراء هذا التحليل على النحو التالى



٢-٢ الرسوم البيانية باستخدام SPSS :-

يتيح البرنامج إمكانيات هائلة فى الرسوم البيانية ، حيث تعتبر وسيلة مناسبة لعرض واستيعاب البيانات لأن الرسم البيانى أكثر قدرة على توضيح المعلومات ، وقد تعرضنا لبعض الرسوم البيانية الإحصائية مثل Histogram , Boxplot وسنعرض صورة مبسطة لبعض إمكانيات البرنامج فى الرسم البيانى .

١- إنشاء الرسم البيانى Graph:

مثال :يوضح الجدول التالى درجة الحرارة العظمى ودرجة الحرارة الصغرى فى بعض المدن

المدن	القاهرة	الإسكندرية	مطروح	السويس	شرم الشيخ	الأقصر
درجة الحرارة العظمى	٣٥	٢٨	٢٨	٣٦	٣٩	٤٣
درجة الحرارة الصغرى	٢٠	٢١	٢١	٢٢	٢٦	٢٣

مثل هذه البيانات بالأعمدة البيانية

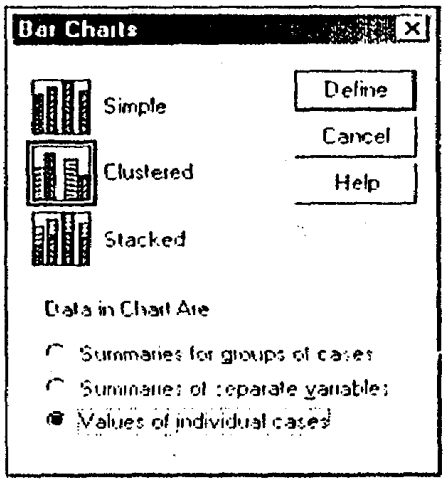
لتمثيل البيانات بالرسم البيانى نتبع الآتى :

١- سجل البيانات فى ملف بيانات كالتالى

	city	max	min
1	القاهرة	35.00	20.00
2	الاسكندرية	28.00	21.00
3	مطروح	28.00	21.00
4	السويس	36.00	22.00
5	شرم الشيخ	39.00	26.00
6	الأقصر	43.00	23.00

٢- من القائمة الرئيسية اختر أمر Graph تظهر قائمة بأنواع الرسوم البيانية المتاحة ، حدد نوع الرسم المطلوب وليكن Bar ،

٣- بعد تحديد نوع الرسم يظهر الصندوق الحوارى Bar Charts

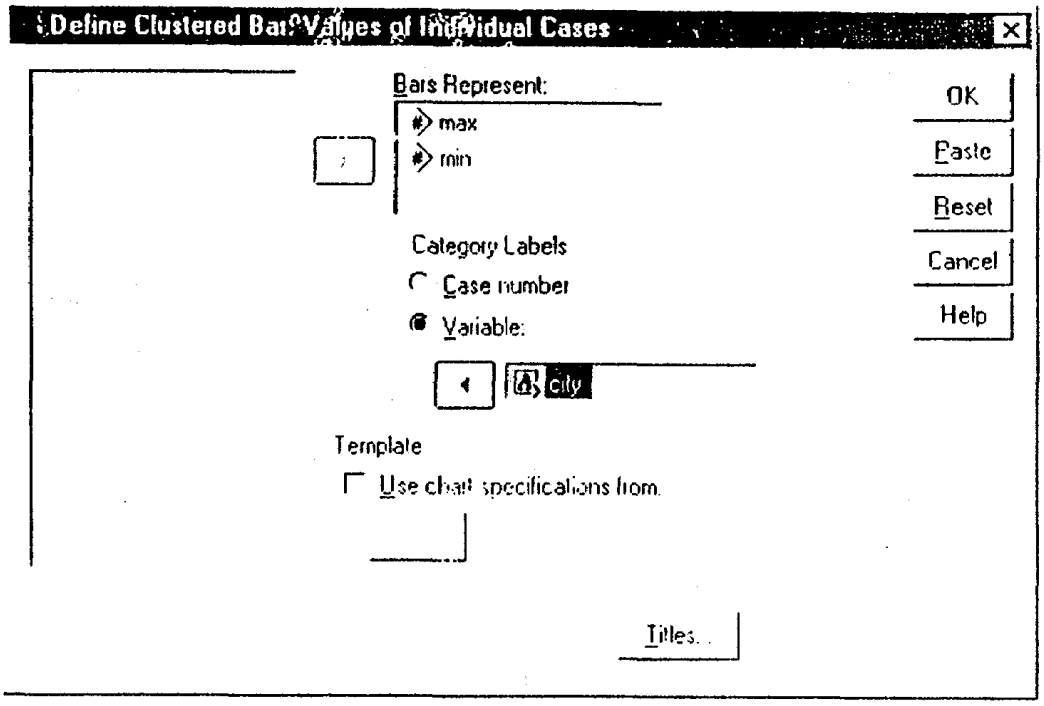


٤- من هذا الصندوق حدد نوع الأعمدة البيانية المطلوبة (Simple , Clustered, Stacked) ، اختر Clustered

٥- نجد فى خانة Data in Chart Are ثلاثة اختيارات وتعتمد هذه الاختيارات على طريقة تمثيل البيانات :

- Summaries for group of cases : فى حالة متغير أو متغيرات يراد تقسيمها إلى مجموعات
- Summaries for separate variables : فى حالة تمثيل أكثر من متغير بيانيا .
- Values of individual cases : فى حالة متغير واحد ويراد تمثيل بيانات كل مفردة على حدة .

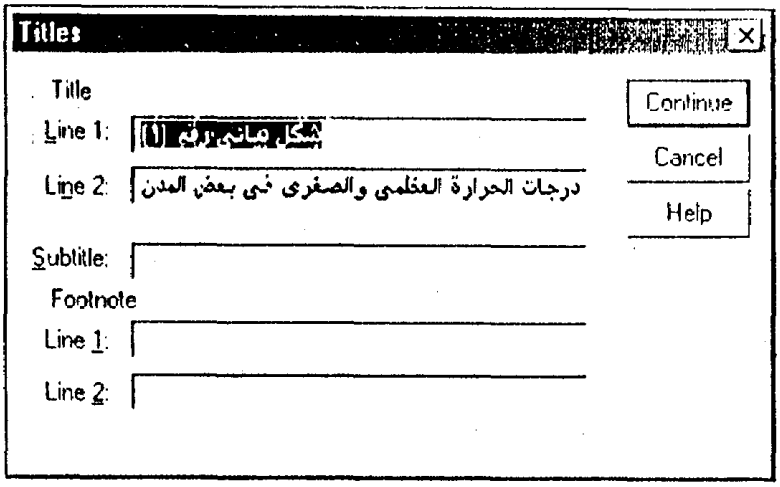
٦- حدد الاختيار المطلوب وليكن Values of individual cases ، ثم اختر أمر Define يظهر الصندوق الحوارى التالى :



٧- أمام خانة Category Labels اختر المتغير الذى يحدد عناوين الأرقام (أسماء المدن) وليكن City .

٨- أمام خانة Bar represent اختر المتغيرات المراد تمثيلها max , min

٩- انقر أمر Titles يظهر الصندوق الحوارى التالى :

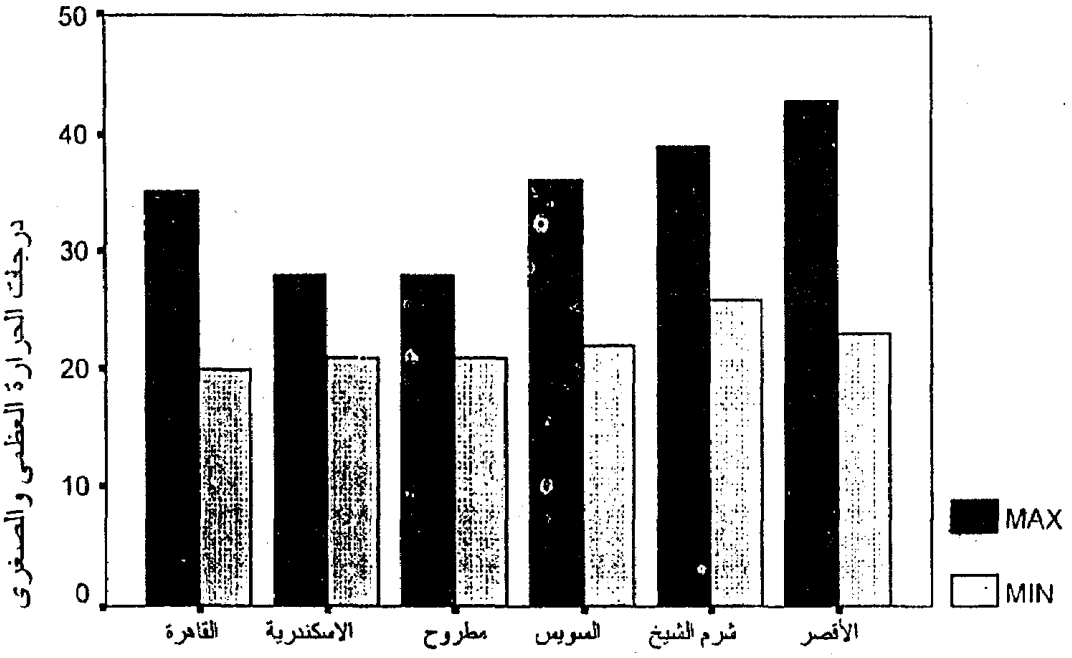


١٠- بعد كتابة البيانات المطلوبة اختر أمر **Continue** ثم أمر **OK** يظهر الرسم البيانى التالى:

Graph

شكل بيانى رقم (1)

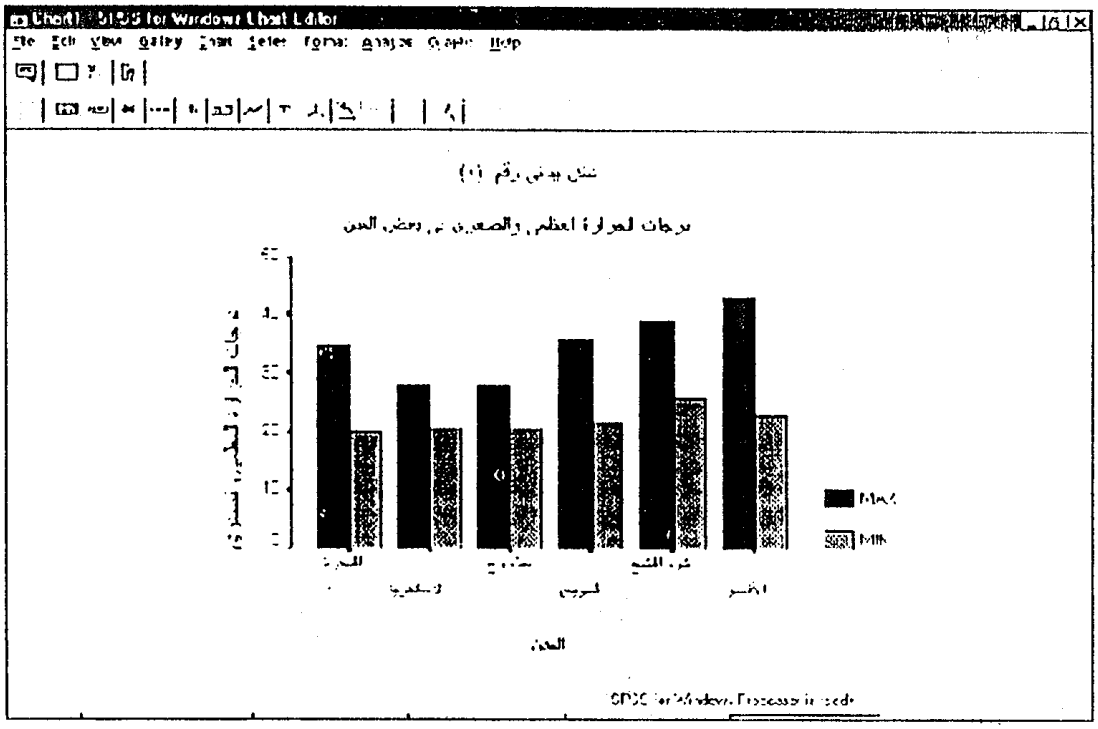
درجات الحرارة العظمى والصغرى فى بعض المدن



يلاحظ أن شكل الرسم البيانى غير مناسب وفى حاجة إلى تعديلات ، لذلك سنتعرض إلى تعـديل الرسم

تعديل الرسم Editing Chars :

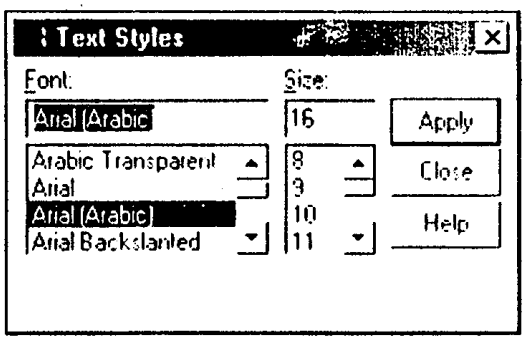
يلاحظ أن الرسم البيانى يظهر داخل نافذة (Output Navigator) ، ولتعديل الرسم نضغط Double Click فى منطقة الرسم فيظهر الرسم فى نافذة مستقلة تسمى Chart Editor تتيح لنا التعامل مع الرسم على النحو التالى



١ – تغيير حجم خط الكتابة :

يلاحظ فى الرسم السابق أن الكتابة (عناوين ، توصيف المحاور) قد لا تظهر باللغة العربية لذلك يلزم تغيير خط هذه الكتابة :

– حدد الكتابة ، ثم من قائمة Format اختر أمر Text ،حيث يمكن اختيار شكل وحجم خط الكتابة ، ثم اختر أمر Apply .

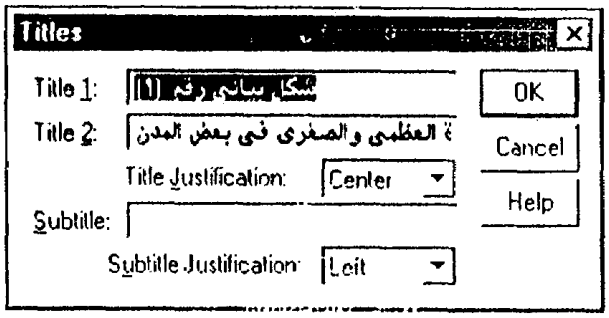


ويلاحظ أن الخط المناسب للعناوين العربية هو Arial Arabic أو Arabic Transparent

٢- محاذاة العناوين :

يلاحظ أن العناوين يتم محاذاتها نحو الشمال وأحياناً يكون من المناسب بالنسبة للعناوين أن يتم توسيطها فى الرسم ولعمل ذلك :

١- من قائمة **Chart** اختر أمر **Titles** ، وأمام خانة **Title justification** اختر أمر **Center** يتم توسيط العنوان

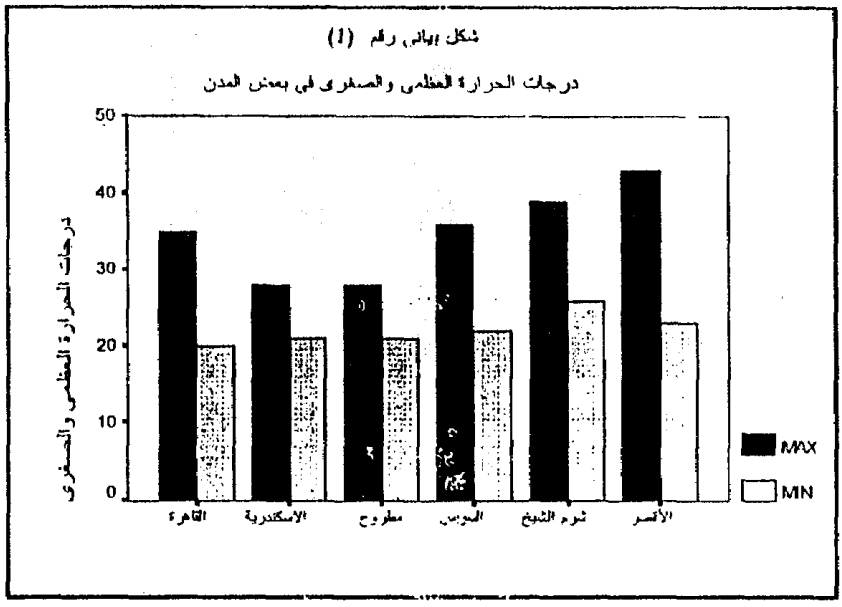


٢- بنفس الطريقة السابقة يتم محاذاة الـ **Footnote**

٣- وضع إطار خارجي للرسم :

١- من قائمة **Chart** نشط أمر **Outer frame** يتم وضع إطار خارجي للرسم

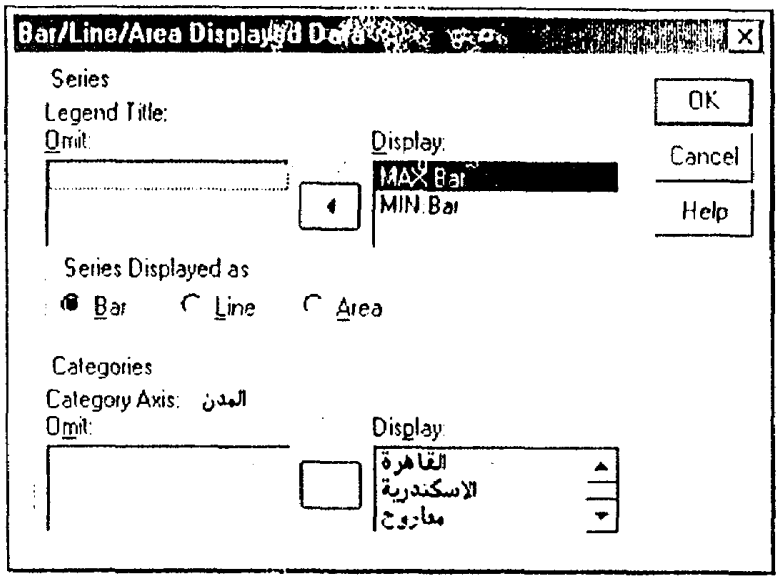
٢- بعد الاختيارات السابقة يظهر الرسم البيانى بالشكل المناسب التالى :



٤ -إظهار وإخفاء الأعمدة :

يمكن إظهار وإخفاء بعض الأعمدة على النحو التالى :

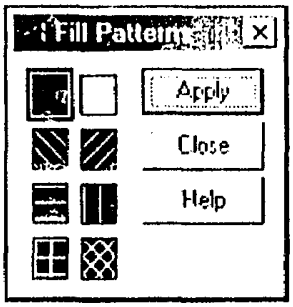
١- من قائمة **Series** اختر أمر **Display** يظهر الصندوق الحوارى التالى :



٢- يلاحظ أن الأعمدة التى تظهر فى خانة **Display** هى الأعمدة الممثلة بيانياً ، ولإخفاء بعض الأعمدة اختر العمود المراد إخفاؤه ثم انقر السهم المجاور ينتقل إلى خانة **Omit** أى يحذف من الرسم البيانى .

٥ -تهشير الأعمدة :

لتهشير الأعمدة اختر الأعمدة المراد تهشيرها ، ثم من قائمة **Format** اختر أمر **Fill Pattern** يظهر الشكل التالى :

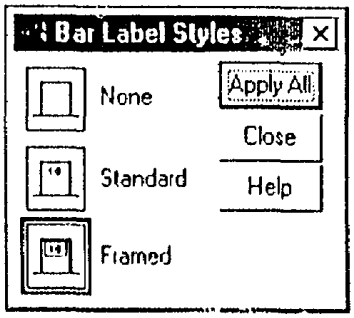


اختر الشكل المطلوب ، ثم اختر أمر **Apply**

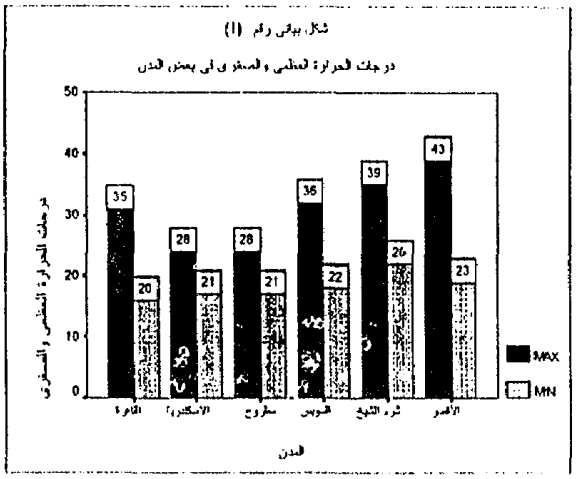
٦- إظهار الأرقام على الرسم :

أ. لإظهار أرقام البيانات الممثلة على الأعمدة ، من قائمة Format اختر أمر Bar Label Style

ب. يظهر الصندوق الحوارى التالى :

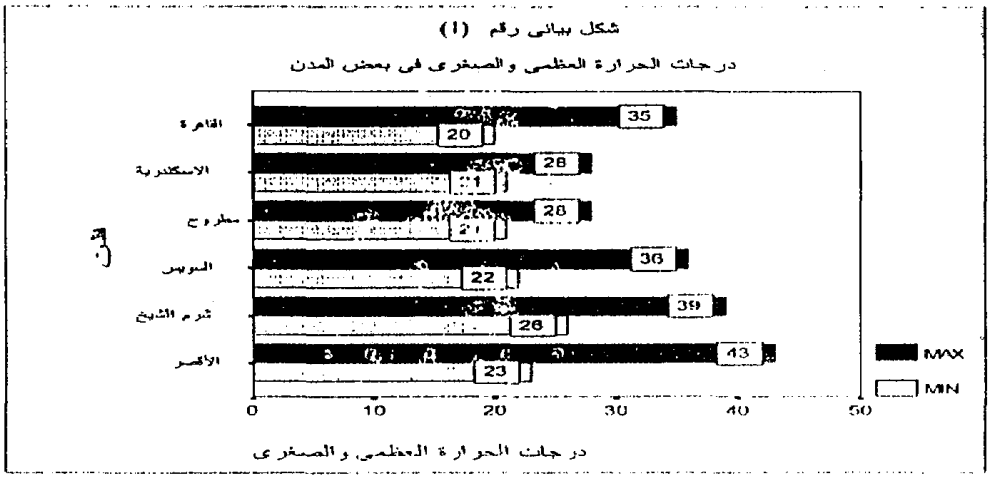


ج. اختر الشكل Framed يظهر الرسم البيانى بالشكل التالى :



٧- تدوير الرسم :

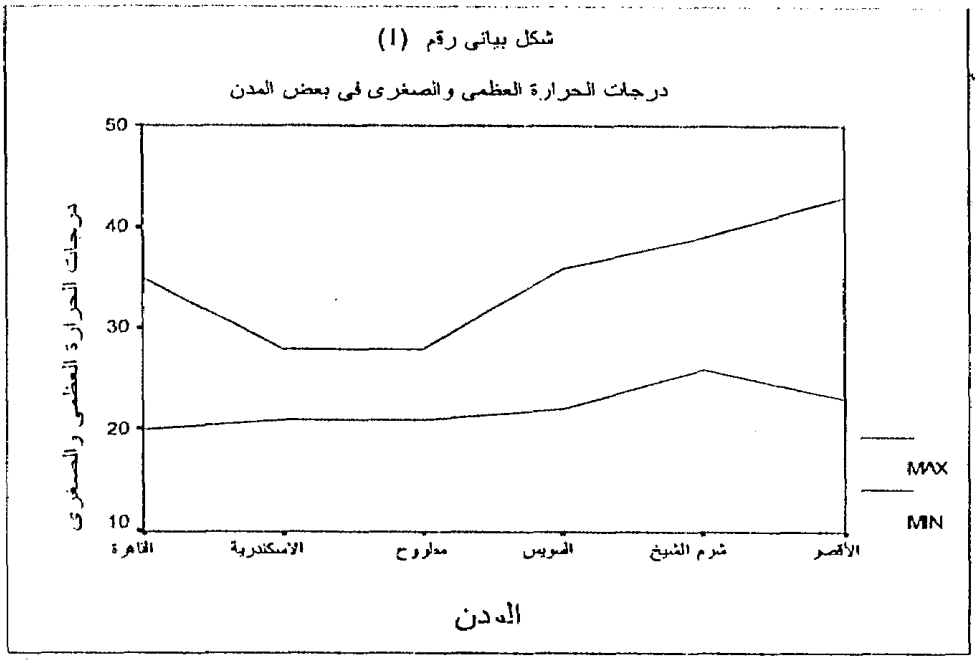
لتدوير الرسم اختر أمر Format ، ثم اختر أمر Swap Axes يظهر الرسم البيانى بالشكل التالى



٨- تعديل نوع الرسم :

- لتعديل نوع الرسم (Bar Line Area .... etc) نتبع الآتى
- من قائمة Gallery اختر النوع المطلوب وليكن Line ، ثم اختر Multiple

واختر Replace يظهر الرسم البيانى بالشكل التالى :



التمثيل الدائرى للبيانات

إذا كانت البيانات نسباً مئوية ، ومجموع هذه النسب ١٠٠% فيصلح تمثيلها بالقطاعات الدائرية

تمرين : فى إحدى السنوات وجد أن التركيب المحصولى فى جمهورية مصر العربية كالتالى

محاصيل حقلية	خضروات	فواكه	نباتات طبية وعطرية
%٧٠	%١٠	%١٢	%٨

والمراد تمثيل هذه البيانات بقطاعات دائرية

٢-٣ : الجداول التكرارية

يتطلب التحليل الاحصائى-فى مرحلته الأولى-جمع البيانات المطلوب دراستها وتأتى المرحلة التالية بعد الحصول على البيانات ، وهى تنظيم وعرض بيانات العينة بمجرد الحصول عليها. ويعتبر تبويب البيانات وعرضها من الخطوات الأساسية ، لأنه من النادر استخلاص أية نتائج مفيدة أو التعرف على أية ملامح أساسية لبيانات غير مبوبة . ويتم تبويب البيانات من خلال الجداول التكرارية.

خطوات بناء الجدول التكراري:-

أولاً : تحديد عدد الفئات : معظم الإحصائيين يفضلون أن لا يقل عدد الفئات عن ٤ وأن لا يزيد عن ١٥ (وهذه ليست قاعدة جامدة) وذلك لأن اختصار البيانات فى عدد قليل من الفئات يرافقه إضاعة الكثير من حقيقة البيانات . واختيار عدد كبير من الفئات يعنى أن هذه البيانات لم تختصر.

ثانياً : تحديد طول الفئة : لتحديد طول الفئة لابد من معرفة مدى التوزيع، والمدى هو الفرق بين أكبر مشاهدة واصغر مشاهدة ، ثم نحدد عدد الفئات ويكون

$$\text{طول الفئة} = \frac{\text{المدى}}{\text{عدد الفئات}}$$

ثالثاً : حدود الفئات : معظم الاحصائيين يميلون لأن يكون الحد الأدنى للفئة الأولى من مضاعفات طول الفئة بشرط أن يكون أقل من أو يساوى اصغر مشاهدة فى التوزيع . أما الحد الأعلى فيكون عبارة عن الحد الأدنى مضافاً إليه طول الفئة

مثال :البيانات التالية توضح نتائج ٣٥ متدرب فى دورة تدريبية لتحسين الأداء فى إدارة ما، كون التوزيع التكرارى المناسب لهذه البيانات .

٥٠، ٥٤، ٥٨، ٤٨، ٤٩، ٤٦، ٧٨، ٤٨، ٣٧، ٦٢، ٤٠، ٥٨، ٣٣، ٥٩، ٤٨، ٥٤، ٣٩، ٦٨، ٤٤، ٦١، ٤٧، ٤٥، ٤١، ٥٧، ٤٣، ٦٣، ٥٢، ٦٨، ٥١، ٧٠، ٧٣.

الحل : ١ - المدى = الفرق بين أعلى قيمة -أدنى قيمة = ٧٨-٣٣ = ٤٥

٢ - اختيار عدد الفئات وليكن ١٠

٣ - طول الفئة بالتقريب = $\frac{٤٥}{١٠}$ = ٤٫٥ ≈ ٥

٤- تعيين حدود الفئات، وحيث أن أصغر مشاهدة فى التوزيع = ٣٣ وطول الفئة = ٥

فإن الحد الأدنى للفئة يفضل أن يكون أقرب ما يكون إلى اصغر مشاهدة وعـلى ذلك فإن الحد الأدنى للفئة الأولى يكون ٣٠ ، والحد الأعلى يكون الحد الأدنى للفئـة مضاف إليه طول الفئة أى يساوى ٣٥ وعلى ذلك نحصل على الجدول التكـراري التالى :-

(التوزيع التكرارى لنتائج ٣٥ متدرب في دورة)

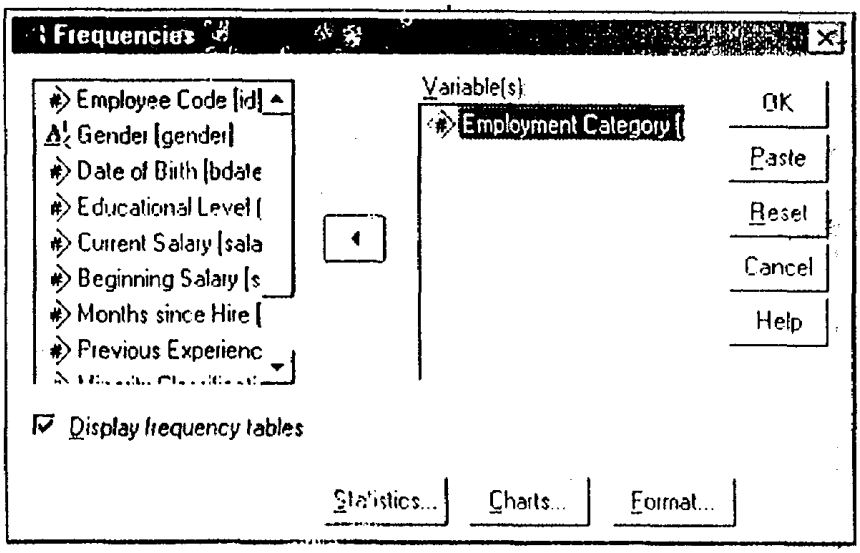
تدريبية لتحسين الأداء

الفئـات	العـلاقات	التكرارات
٣٠-	/	١
٣٥-	//	٢
٤٠-	////	٤
٤٥-	//// ////	٩
٥٠-	/ ////	٦
٥٥-	////	٥
٦٠-	///	٣
٦٥-	//	٢
٧٠-	//	٢
٧٥- أقل من ٨٠	/	١

الجداول التكرارية للمتغيرات باستخدام البرنامج SPSS

١. افتح الملف Employee Data.sav

٢. افتح قائمة Analyze ، ثم اختـر منـها Descriptive Statistics ، ومنـها اختـر Frequencies ، يظهر الصندوق الحوارى Frequencies



٢. اختر المتغير أو المتغيرات المراد عمل جدول تكرارى لها فى خانة (Variables)

٣. يمكن إظهار الجدول التكرارى بتنشيط خانة Display frequency tables

٤. اختر أمر OK تظهر النتائج التالية :

Frequencies

Statistics

Employment Category

N	Valid	474
	Missing	0

Employment Category

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Clerical	363	76.6	76.6	76.6
	Custodial	27	5.7	5.7	82.3
	Manager	84	17.7	17.7	100.0
	Total	474	100.0	100.0	

ونجد بالنتائج اسم وتوصيف المتغير ، يتبعه جدول به Value Label ، وقيم المتغير Value ، التكرار Frequency ، النسب المئوية Percent ، النسب المئوية بعد استبعاد القيم المفقودة إن وجدت Valid Percent ، ثم التكرار المتجمع Cumulative Percent .

٢-٤ : التحليل الوصفى للبيانات

يتم وصف البيانات من خلال حساب مقاييس النزعة المركزية ومقاييس التشتت وكذا مقاييس الالتواء والتفلطح أو التفرطح (التدبب)

أولاً : مقاييس النزعة المركزية وتشمل :-

١- الوسط الحسابى : (Arithmetic Mean) الوسط الحسابى أو المتوسط لمجتمع ما لمجموعة من القيم هو مجموع القيم مقسوماً على عددها.

الوسط الحسابى من بيانات غير مبوبة :-

الوسط الحسابى للمجتمع μ

$$\mu = \frac{x_1 + x_2 + \ldots + x_n}{N} = \frac{\sum x_i}{N}$$ حيث N حجم بيانات المجتمع

الوسط الحسابى للعينة $\bar{x}$:-

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum x_i}{n}$$ حيث n حجم بيانات العينة

الوسط الحسابى للعينة $\bar{x}$:-

من بيانات مبوبـــــة :-

الوسط الحسابى للمجتمع μ

$$\mu = \frac{\sum f_i m_i}{N}$$ حيث f_i = تكرار الفئة i

m_i = مركز الفئة

N = حجم بيانات المجتمع

الوسط الحسابى للعينة $\bar{x}$: يستخدم القانون السابق مع إحلال n ، مكان N فى المعادلة .

٢- الوسيط : - Median

يعرف الوسيط لمجموعة من البيانات هو قيمة المشاهدة الموجودة فى منتصف البيانات مرتبة ترتيباً تصاعدياً أو تنازلياً

أولاً : البيانات غير المبوبة:-

لايجاد الوسيط للبيانات غير المبوبة يجب ترتيبها أولاً ترتيباً تصاعدياً أو تنازلياً،فإذا كان عدد المشاهدات فردياً فإن قيمة الوسيط هى قيمة المشاهدة رقم $\frac{n+1}{2}$ الموجودة فى منتصف البيانات المرتبة.

وإذا كان عدد المشاهدات زوجياً فإن قيمة الوسيط هى متوسط المشاهدتين اللتين ترتيبهما $\frac{n}{2}$ والتى تليها $\frac{n+1}{2}$

ثانياً : البيانات المبوبة:-

يمكن إيجاد الوسيط للبيانات المبوبة فى جدول تكرارى، باستخدام المعادلة التالية:-

$$\text{Median} = \left[\frac{\frac{n}{2} - F}{f_m}\right] w + L_m$$

حيث n = مجموع المشاهدات فى العينة

F = مجموع التكرارات السابقة على الفئة الوسيطة.

f_m = تكرار الفئة الوسيطية.

L_m = الحد الأدنى للفئة الوسيطة.

w = طول الفئة.

٣- المنـــــوال The Mode

المنوال هو القيمة أو الصفة الأكثر شيوعاً (تكراراً) فى البيانات.

ويمكن أن يكون للبيانات منوال واحد فقط يسمى التوزيع فى هذه الحالة أحادى المنوال (unimodal) أو منولان (ثنائى المنوال Bimodal) أو متعدد المنوال (Multimodal)، كما يمكن أن لايوجد منوال للمشاهدات.

البيانات المبوبة:يمكن إيجاد المنوال للبيانات المبوبة فى جدول تكرارى باستخدام المعادلة التالية:-

$$\text{Mode} = L_{m_0} + \left[\frac{d1}{d_1 + d_2}\right] w$$

حيث d_1 = تكرار الفئة المنوالية –تكرار الفئة التالية (بعد) الفئة المنوالية.

d_2 = تكرار الفئة المنوالية –تكرار الفئة السابقة (قبل) الفئة المنوالية.

w = طول الفئة ، L_{m_0} = الحد الأدنى للفئة المنوالية.

ثانيا : مقاييس التشتت :

تستخدم مقاييس التشتت لوصف كيفية انتشار البيانات على جانبي القيمة المتوسطة ولا يقل ذلك أهمية عن تحديد مركز البيانات ويعرف انتشار القيم حول القيمة المتوسطة بالتشتت فإذا زاد التشتت فإن هذا يدل على قلة تجانس البيانات أما إذا اقل التشتت فهذا يدل على زيادة تجانس البيانات . وإذا كان التشتت مساوياً للصفر فهذا يعنى تطابق قيم جميع المفردات أى تجانسها الكامل وهذا بالطبع نادر الحدوث وإذا تساوى متوسطا عينتين فإن هذا لا يعنى أنهما متماثلتان فى تشتت مفرداتهما . وترجع أهمية مقاييس التشتت إلى أنه من الممكن وجود عينتين فى المفردات لهما نفس القيمة المتوسطة ولكنهما تختلفان فى التشتت . ومن أكثر الطرق فى قياس التشتت التباين والانحراف المعياري والمدى ومعامل الاختلاف .

المدى The Range :

المدى هو عبارة عن الفرق بين اكبر القيم وأصغرها . وفى حالة البيانات المبوبة فإن المدى هو الفرق بين الحد الأعلى لأكبر فئة والحد الأدنى لأصغر فئة أو هو الفرق بين مركز الفئة (لأعلى فئة) ومركز الفئة (لأقل فئة).

التباين The variance

أولاً : التباين من بيانات غير مبوبة يحسب من المعادلتين التاليتين:-

التباين للمجتمع σ^2 :-

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N} = \frac{\sum x_i^2 - N\mu^2}{N}$$

حيث μ = متوسط المجتمع

N= حجم بيانات المجتمع

التباين للعينة S^2 :

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$$

حيث $\bar{x}$ = متوسط العينة

n = حجم بيانات العينة حيث n ليست كبيرة

ثانياً : التباين من بيانات مبوبة:-

وتستخدم المعادلتين التاليتان لحساب التباين من البيانات المبوبة

تباين المجتمع $$\sigma^2 = \frac{\Sigma f_i m_i^2 - N\mu^2}{N}$$

تباين الفئة $$S^2 = \frac{\Sigma f_i m_i^2 - n\bar{x}^2}{n-1}$$

٣- الانحراف المعيارى The standard Deviation

الانحراف المعيارى لمجموعة من القيم هو الجذر التربيعى الموجب للتباين

$$S = \sqrt{S^2} \ , \ \sigma = \sqrt{\sigma^2}$$

٣- معامل الاختلاف The coefficient of variation

يستخدم معامل الاختلاف (وهو مقياس نسبى) لمقارنة تشتت مجموعتين أو اكثر من البيانات.

ويحسب معامل الاختلاف من المعادلتين التاليتين:-

معامل الاختلاف للمجتمع $$= cv = \frac{\sigma}{\mu} \times 100$$

معامل الاختلاف للعينة $$= cv = \frac{s}{\bar{x}} \times 100$$

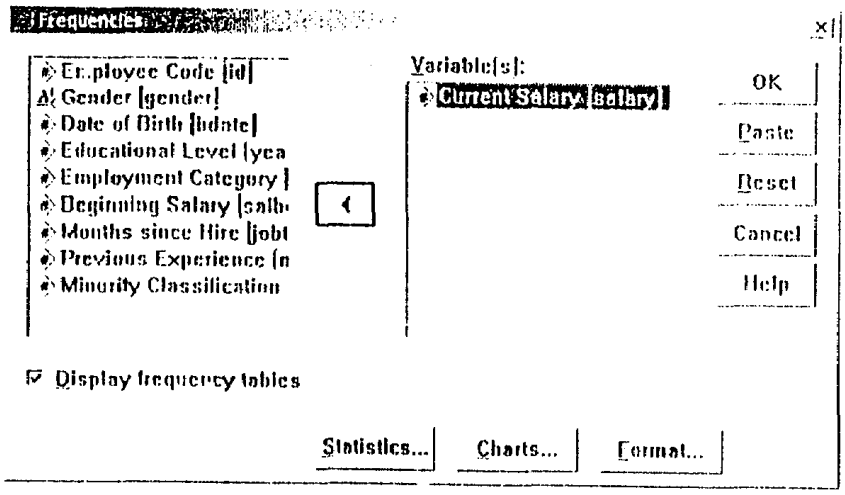
حساب المقاييس الإحصائية من برنامج SPSS

يلاحظ أن هذه المقاييس تصلح للبيانات الكمية

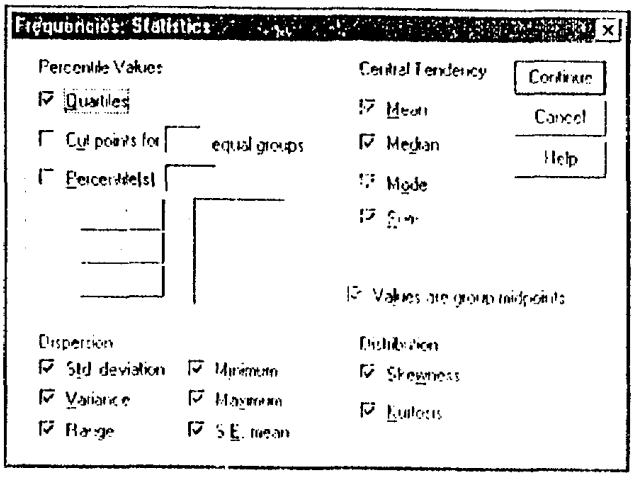
لحساب المقاييس الإحصائية نتبع الآتى :

١. افتح الملف Employee Data.sav

٢. من قائمة Analyze اختر أمر Descriptive Statistics ثم اختر Frequencies يظهر الصندوق الحوارى Frequencies



٣. اختر المتغير Salary من الصندوق الحوارى Frequencies وأنقر أمـر Statistics يظهـر الصندوق الحوارى التالى :



تشتمل هذه المقاييس الإحصائية على ما يلى :

- مقاييس النزعة المركزية Central Tendency : وتشمل الوسط الحسابى Mean ، الوسيط Median ، المنوال Mode ، المجموع Sum .
- مقاييس التشتت Dispersion : وتشمل الانحراف المعيارى Std. deviation ، التبـاين Variance ، المدى Range ، أقل قيمة للمـتغير Minimum ، أكـبر قيمـة للمـتغير Maximum ، الخطأ المعيارى للمتوسط S.E. mean
- مقاييس التوزيع Distribution : وتشمل مقياس الالتواء Skewness ، ومقياس التفرطح Kortosis .

- مقاييس المدى (الربيعات – العشيرات .....) Percentile Values : وتشمل الربيعات Quartiles ، العشيرات Cut points for 10 equal groups ، يمكن تقسيم المدى إلى أى مجموعات متساوية أخرى ، كما يمكن كتابة النسب المطلوبة فى خانة Percentile(s) .

٢. يمكن تنشيط الخانة Values are group mid points إذا كنا قد أخذنا القيم عن طريق مراكز الفئات وليست البيانات الخام .

٣- بعد اختيار المقاييس الإحصائية اختر أمر Continue تظهر النتائج التالية

Statistics

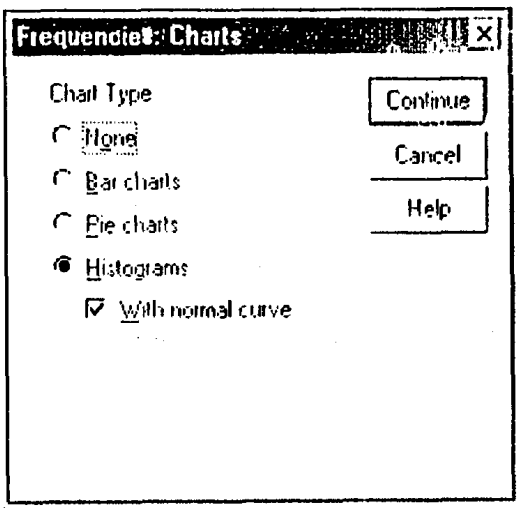
Employment Category

N	Valid	474
	Missing	0
Mean		1.41
Std. Error of Mean		3.55E-02
Median		1.28[a]
Mode		1
Std. Deviation		.77
Variance		.60
Skewness		1.456
Std. Error of Skewness		.112
Kurtosis		.268
Std. Error of Kurtosis		.224
Range		2
Minimum		1
Maximum		3
Sum		669
Percentiles	25	[b]
	50	1.28
	75	1.89

a. Calculated from grouped data.

b. The lower bound of the first interval or the upper bound of the last interval is not known. Some percentiles are undefined.

c. Percentiles are calculated from grouped data.



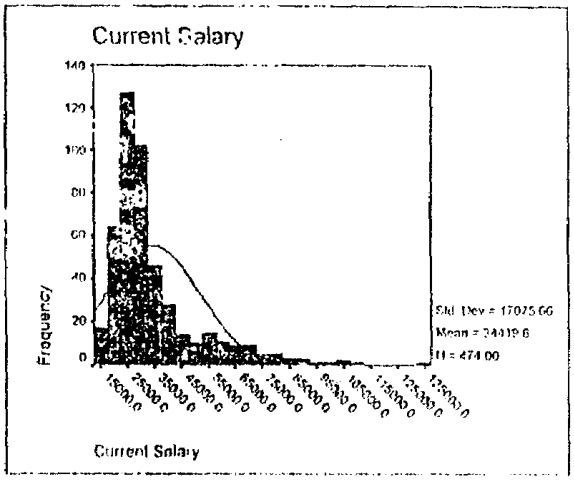
٣. لإظهار الرسوم البيانية التى تستخدم فى عرض البيانات ، أنقر خانة Charts من الصندوق الحوارى Frequencies يظهر الصندوق الحوارى التالى :

ويشتمل هذا الصندوق على مايلى :

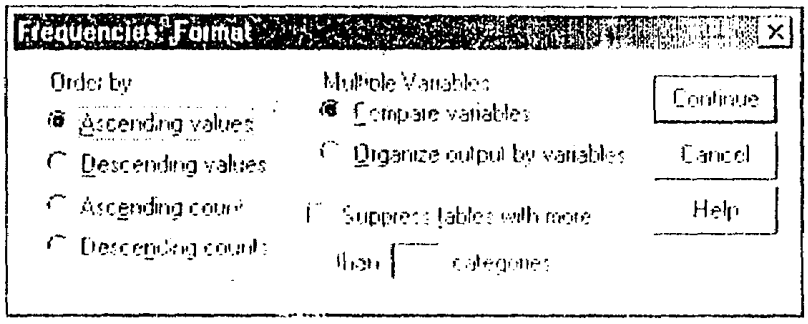
- None : فى حالة عدم إظهار أى رسوم بيانية
- الأعمدة البيانية Bar chart(s) : ويستخدم فى حالة إظهار رسم بيانى (أعمدة) للبيانات المنقطعة

- المدرج التكرارى Histogram(s) : ويستخدم فى حالة إظهار الرسم البيانى للمتغيرات المستمرة ، كما يمكن إظهار منحنى التوزيع المعتدل باختيار With normal curve .

بعد اختيار أمر Continue ، ثم OK يظهر الرسم البيانى التالى :



للتحكم فى شكل إظهار الجدول التكرارى ، انقر الأمر Format من الصندوق الحوارى Frequencies يظهر الصندوق الحوارى التالى :



ويشتمل هذا الصندوق على ما يلى :

- Order by أى ترتيب البيانات بالجدول :
- Ascending values أى بناءً على قيم المتغير تصاعدياً ، Descending values أى بناء على قيم المتغير تنازلياً ، Ascending counts أى بناءً على التكرارات تصاعدياً ، Descending counts أى بناءً على التكرارات تنازلياً .
- Suppress tables with more than 10 categories : لإهمال الجداول التى تزيد عدد القيم للمتغير بها عن ١٠ ، كما يمكن تحديد رقم آخر .

وفى حالة حساب المقاييس الإحصائية مثل المتوسط الحسابى والوسيط ...الخ لأكثر من متغير Multiple Variables يمكن :

Compare variables: أى إظهارالمقاييس الإحصائية لكل المتغيرات فى جدول

Oraganize output by variables : أى إظهار المقاييس الإحصائية لكل متغير فى جدول على حدة

حساب المقاييس الإحصائية من جدول تكرارى باستخدام برنامج SPSS :

قد لا تتوافر لدينا تفاصيل البيانات ، ولكن تعطى فى صورة جدول تكرارى ، ونريد حساب المقاييس الإحصائية لها .

مثال : لدينا الجدول التكرارى التالى :

قيم (X)	التكرارات (Fre)
٥	٢٥
٨	٦٤
٧	٤٩
١٠	١٠٠
١١	١٢١
٤	١٦
٩	٨١

لحساب المقاييس الإحصائية للمتغير نتبع الخطوات التالية :

١) يتم تسجيل بيانات المتغير الأصلى X والتكرارات fre

٢) يتم ترجيح البيانات بالمتغير Fre عن طريق اختيار الأمر weight cases من قائمة Data .

٣) بعد ذلك نختار أمر Summarize ، ومنها Frequencies ، ونختار المقاييس الإحصـــائية بالطريقة السابقة ، تظهر لنا النتائج التالية :

Frequencies

Statistics

X

N	Valid	456
	Missing	0
Mean		9.0000
Std. Error of Mean		8.901E-02
Median		9.3702[a]
Mode		11.00
Std. Deviation		1.9008
Variance		3.6132
Skewness		-.960
Std. Error of Skewness		.114
Kurtosis		.270
Std. Error of Kurtosis		.228
Range		7.00
Minimum		4.00
Maximum		11.00
Sum		4104.00
Percentiles	25	7.8584[b]
	50	9.3702
	75	10.5158

a. Calculated from grouped data.

b. Percentiles are calculated from grouped data.

X

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	4.00	16	3.5	3.5	3.5
	5.00	25	5.5	5.5	9.0
	7.00	49	10.7	10.7	19.7
	8.00	64	14.0	14.0	33.8
	9.00	81	17.8	17.8	51.5
	10.00	100	21.9	21.9	73.5
	11.00	121	26.5	26.5	100.0
	Total	456	100.0	100.0	

التوصيف Descriptives

يفضل استخدام هذا الأمر مع المتغيرات الكمية

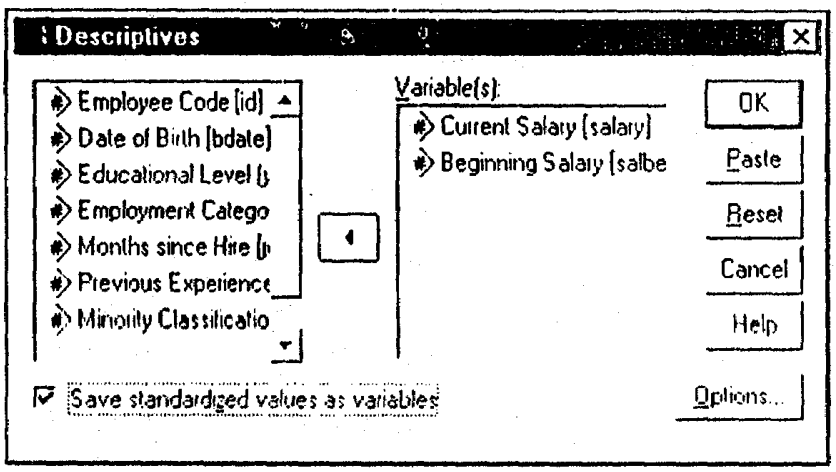
الأمر Descriptives يعطى توصيف ومقاييس إحصائية لعدة متغيرات فى جدول واحد مما يسهل عملية المقارنة ، كما يمكن حساب متغير عيارى (Z scores) يسهل عملية المقارنة إذا اختلفت وحدات القياس ، كما يمكن حساب العديد من المقاييس الإحصائية التى يتم حسابها عن طريق الـ Frequencies .

ولتوصيف Descriptives لمتغير أو لعدة متغيرات نتبع الآتى :

١. من قائمة Statistics اختر أمر Analyze ، ومنها اختر Descriptive Statistics ، يظهر الصندوق الحوارى التالى :

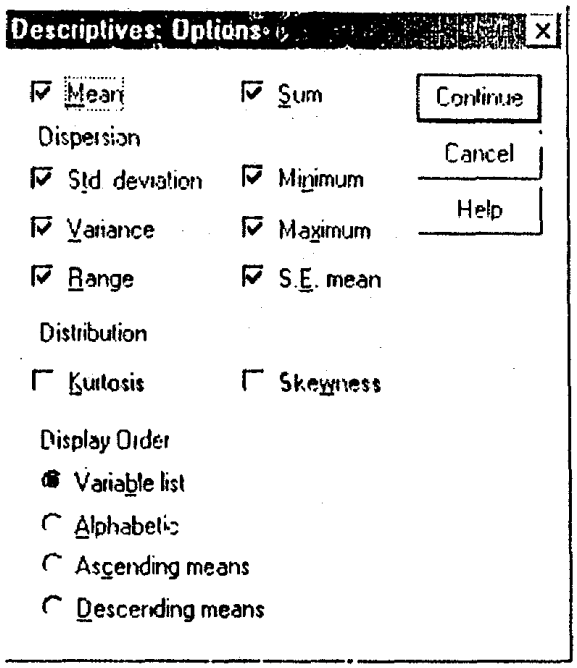
٢. أمام خانة Variable(s) يتم اختيار المتغير أو المتغيرات التى يراد توصيف Descriptive لها .

٣. إذا أردنا حساب متغير عيارى (Z scores) فإننا ننشط الخانة Save Standardized



Values as variables ، فى هذه الحالة تحسب قيم المتغير العيارى لكل متغير وتدرج بملف البيانات ، إذا أردنا الاحتفاظ بهذه المتغيرات فينبغى حفظ ملف البيانات

٤. إذا أردنا حساب المقاييس الإحصائية للمتغيرات ، فإننا ننقر الأمر Options يظهر الصندوق الحوارى التالى :



من هذا الصندوق يتم تنشيط المقاييس الإحصائية المطلوبة ، فتظهر النتائج التالية :

Descriptives

Descriptive Statistics

	Statistic			Std. Error	
	Current Salary	Beginning Salary	Valid N (listwise)	Current Salary	Beginning Salary
N	474	474	474		
Range	$119,250	$70,980			
Minimum	$15,750	$9,000			
Maximum	$135,000	$79,980			
Sum	$16,314,875	$8,065,625			
Mean	$34,419.57	$17,016.09		$784.31	$361.51
Std. Deviation	$17,075.66	$7,870.64			
Variance	291578214.453	61946944.959			
Skewness	2.125	2.853		.112	.112
Kurtosis	5.378	12.390		.224	.224

٢-٥ استكشاف البيانات Explore

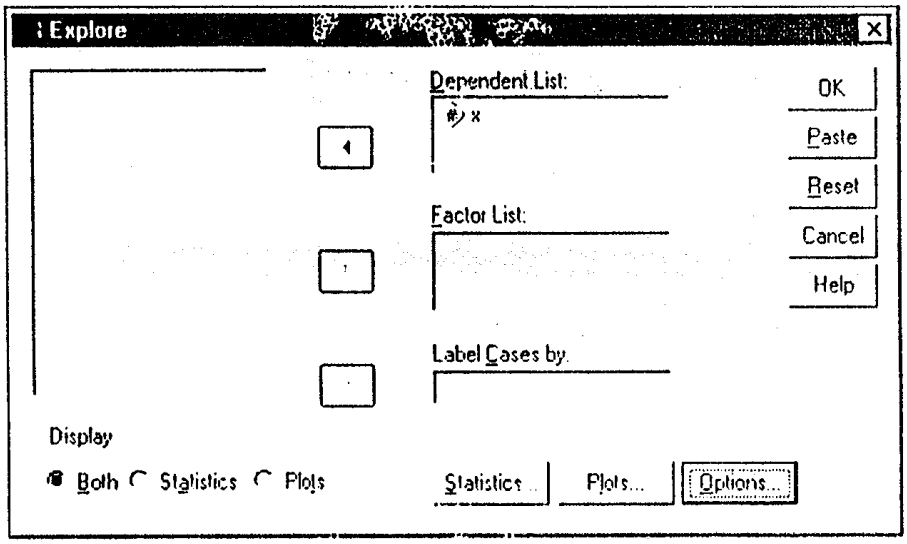
نحتاج لاختبار البيانات لمعرفة القيم الشاذة (خارج نطاق معين Outlire) ، معرفة توزيع البيانات لمعرفة إمكانية تطبيق المقاييس الإحصائية عليها أو أنها تحتاج إلى بعض التحويلات ، كما نحتاج إلى بعض الرسوم البيانية للمقارنة بين مجموعات البيانات .

مثال : فى عينة لمشاهدات عن المتغير العشوائى الذى يعبر عن شدة الزلازل التى حدثت فى أحـد المناطق مقاسة بمقياس ريختر ، اختبر هذه البيانات .

١ ، ٨,٣ ، ٣,١ ، ١,١ ، ٥,١ ، ١,٢ ، ١ ، ١,١ ، ٤ ، ٢ ، ١,٩ ، ٦,٣ ، ١,٤ ، ١,٣ ، ٣,٣ ، ٢,٢ ، ٢,١ ، ٢,١ ، ١,٤ ، ٢,٧ ، ٢,٤ ، ٣ ، ٤,١ ، ٥ ، ٢,٢ ، ١,٢ ، ٧,٧ ، ١,٥

ولاختبار البيانات نتبع الخطوات التالية :

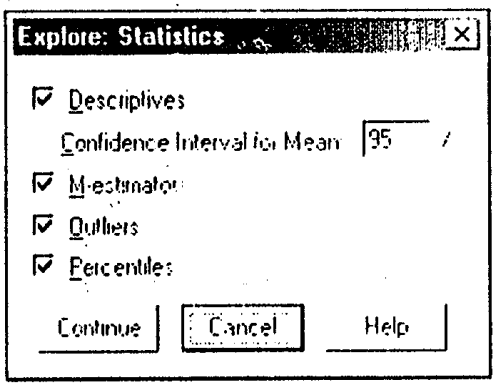
١. من قائمة Statistics اختر أمر Analyze ، ومنها اختر Descriptive statistics ومنها اختر Explore يظهر الصندوق الحوارى التالى :



٢. أمام خانة Dependent List اختر المتغير أو المتغيرات المراد اختبارها .

٣. أمام خانة Factor List اختر المتغير الذى يقسم هذه المتغيرات إلى مجموعات

٤. أمام خانة Label Cases by اختر المتغير الذى يراد ترقيم الحالات على أساسه ، ولإظهار المقاييس الإحصائية ، أنقر أمر Statistics فيظهر الصندوق الحوارى التالى :



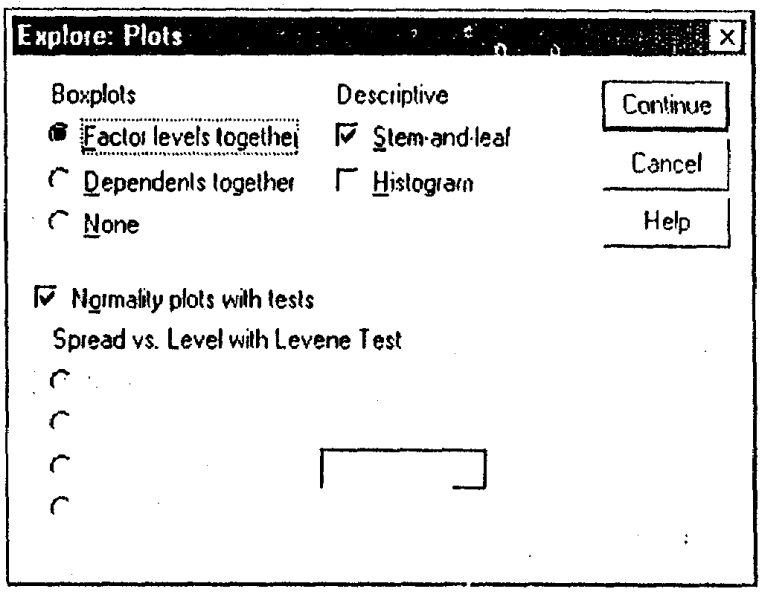
يشتمل هذا الصندوق على ما يلى :

- Descriptives : يحسب المقاييس الإحصائية السابق التعرف عليها (مقاييس النزعة المركزية – مقاييس التشتت – مقاييس التوزيع)

M-estimators حساب تقديرات الوسط الحسابي والوسيط للمجتمع من العينة ،

- Outliers : يظهر أكبر ٥ قيم ، وأصغر ٥ قيم للمتغير مع إظهار أرقام الحالات .
- Percentiles : مقاييس تقسيم البيانات إلى نسب مئوية .

ولإظهار الرسوم البيانية الخاصة باختبار البيانات Explore ، ننقر Plots من الصندوق الحوارى Explore فيظهر الصندوق الحوارى التالى



بعد هذه الاختيارات تظهر لنا النتائج التالية :

Explore

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
X	29	100.0%	0	.0%	29	100.0%

Descriptives

			Statistic	Std. Error
X	Mean		2.8069	.3668
	95% Confidence Interval for Mean	Lower Bound	2.0556	
		Upper Bound	3.5582	
	5% Trimmed Mean		2.6125	
	Median		2.1000	
	Variance		3.901	
	Std. Deviation		1.9750	
	Minimum		1.00	
	Maximum		8.30	
	Range		7.30	
	Interquartile Range		2.3000	
	Skewness		1.529	.434
	Kurtosis		1.763	.845

M-Estimators

	Huber's M-Estimator[a]	Tukey's Biweight[b]	Hampel's M-Estimator[c]	Andrews' Wave[d]
X	2.2495	2.0547	2.2537	2.0552

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.

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Weighted Average(Definition 1)	X	1.0000	1.1000	1.3500	2.1000	3.6500	6.3000	8.0000
Tukey's Hinges	X			1.4000	2.1000	3.3000		

Extreme Values

			Case Number	Value
X	Highest	1	2	8.30
		2	28	7.70
		3	13	6.30
		4	5	5.10
		5	25	5.00
	Lowest	1	1	1.00
		2	7	1.00
		3	8	1.10
		4	4	1.10
		5	27	.[a]

a. Only a partial list of cases with the value 1 are shown in the table of lower extremes.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
X	.207	29	.003	.811	29	.010**

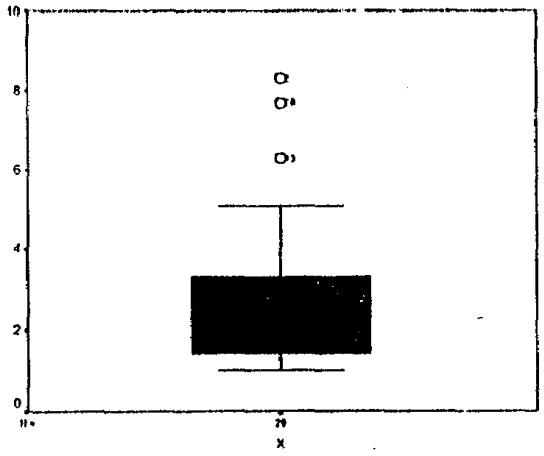
**. This is an upper bound of the true significance.

a. Lilliefors Significance Correction

X Stem-and-Leaf Plot

Frequency	Stem &	Leaf
9.00	1 .	001122344
3.00	1 .	599
6.00	2 .	011224
1.00	2 .	7
3.00	3 .	003
.00	3 .	
2.00	4 .	00
.00	4 .	
2.00	5 .	01
3.00	Extremes	(>=6.3)

Stem width: 1.00
Each leaf: 1 case(s)



الباب الثالث

اختبارات الفروض الإحصائية

تعتبر اختبارات الفروض الإحصائية **Testing statistical Hypothesis** واحدة من أهم التطبيقات التى قدمها علم الإحصاء كحل للمشاكل العلمية المختلفة بشتى فروع العلم ، باستخدام نظرية الاحتمالات وخصائص التوزيعات العينية أمكن التعرف على ما يسمى باختبارات الفروض الإحصائية ، ومن خلالها يمكن لاى شخص أن يتخذ قرار برفض أو قبول فرض معين أو مجموعة من الفروض المتعلقة بمشكلة معينة موجودة فى الحياة العامة . وقد تزايد الاهتمام باختبارات الفروض فى السنوات الأخيرة حتى أصبحت تدخل فى شتى فروع العلم (الطب–الزراعة–الفلك ، الاقتصاد، الصناعة ، علم النفس ...الخ) مما أدى تطورها تطوراً هائلاً وملحوظاً.

وتعتبر اختبارات الفروض الإحصائية أسلوبا لاتخاذ قرار بخصوص فرضين مختلفين بناء على بيانات العينة المتاحة والاختبارات الإحصائية المناسبة. والفرض الاحصائى هو عبارة عن صياغة مبدئية حول واحد أو أكثر من معالم المجتمع.

٣–١ : الخطوات الأساسية فى اختبارات الفروض الإحصائية :–

تتضمن اختبارات الفروض عدة خطوات هى :–

١– التعرف على توزيع المجتمع الأصلى

٢– صياغة الفروض

أ – فرض العدم (H_0) ويسمى أيضا الفرض الصفرى.

ب– الفرض البديل (H_1)

٣– تحديد مستوى المعنوية (α)

٤– صياغة قاعدة القرار وتتضمن :–

أ – تحديد إحصائية الاختبار $z , t , \ldots$

ب– تحديد المنطقة الحرجة والتى تتضمن منطقتى القبول والرفض.

٥– أتخاذ القرار بناء على المعلومات السابقة وذلك برفض أو عدم رفض (قبول) الفرض الصفرى (فرض العدم).

وسنتناول كل منها بالتفصيل كما يلى

أولاً : التعرف على توزيع المجتمع الأصلي :-

يجب التعرف على توزيع المجتمع الأصلي وتحديد ما إذا كان يتبع التوزيع الطبيعى أو توزيع ذى الحدين أو ما إلى ذلك .

ثانياً : صياغة الفروض :-

هناك فرضان : الفرض الأول وهو فرض العدم ويرمز له بالرمز (H_0) (أو الفرض الصفرى) والفرض الثانى وهو الفرض البديل ويرمز له بالرمز (H_1)

ويصاغ الفرض الصفرى أو فرض العدم كالتالى

$$H_0 : \mu = \mu_0$$

أى أنه صياغة مبدئية عن معلمة من معالم المجتمع المجهولة (وليكن وسط المجتمع μ)، μ_0 قيمة محددة

أما الفرض البديل μ_1 فهو صياغة مبدئية تشير إلى نفس المعلمة المجهولة لها قيمة تختلف عن القيمة التى حددها فرض العدم . أى أنه يتم قبول الفرض البديل فى حالة رفض فرض العدم.

ويأخذ الفرض البديل إحدى الصور التالية:-

1 $\quad H_1 : \mu \neq \mu_0$

وذلك فى حالة عدم توافر معلومات عن معلمة المجتمع إذ قد تكون أكبر أو أصغر .

2 $\quad H_1 : \mu < \mu_0$

3 $\quad H_1 : \mu > \mu_0$

وتستخدم هذه الصيغ حينما تكون μ_0 هى مواصفات قياسية أو قيمة يجب أن تحقق وتشك فى تحقيقها.

ثالثاً :تحديد مستوى المعنوية(α) :-

يقيس مستوى المعنوية α درجة عدم التأكد ، فلو كان لدينا فترة ثقة ٩٥% فيكون لدينا عدم تأكد مقدارة ٥% أى أن مستوى المعنوية هو احتمال رفض فرض العدم وهو صحيح ويسمى ذلك خطأ من النوع الأول كما يتضح من الجدول :

القرار	الخطأ	نوع الخطأ
رفض H_0 وهو صحيح	خطأ من النوع الأول	α
قبول H_0 وهو خطأ	خطأ من النوع الثانى	β

رابعاً : صياغة قاعدة القرار :-

أ- تحديد إحصائية الاختبار :- وتتحدد بناء على نوع التوزيع الاجمالى لمجتمع الدراسة وعلى حجم العينة n بصفة عامة .

١- فإن كان التوزيع طبيعى وكانت $n \geq 30$ والانحراف المعيارى للمجتمع σ معلوم. فتكون إحصائية الاختبار هى

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

٢- وإذا كان توزيع المجتمع طبيعيا وكانت $n < 30$ والانحراف المعيارى للمجتمع غير معلوم فتكون إحصائية باستخدام توزيع t حيث

$$t = \frac{\bar{X} - \mu}{\sigma}$$

٣- أما إذا كان توزيع المجتمع يتبع توزيع كاي تربيع $كا^2$ أو توزيع F فإن إحصائية الأختبار تكون مختلفة وسوف نتعرض له بالتفصيل فى أجزاء تالية :

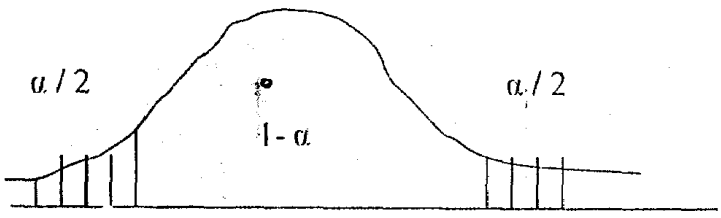
ب- تحديد المنطقة الحرجة : وهى أما منطقة رفض أو منطقى رفض فرض العدم . وهى أما تكون على جانبى المنحنى من (ذيلين) أو فى جانب واحد (من ذيل واحد) طبقاً لطبيعة الأختبار البديل .

ويعتمد تحديد المنطقة الحرجة على كل من حجم العينة (n) ومستوى المعنوية (α) وقيمة إحصائية الاختبار المستخدم وطبيعة الفرض البديل.

والإشكال التالية تبين منطقة الرفض (المنطقة المظللة) وتسمى المنطقة الحرجة ، منطقة عدم الرفض (الغير مظللة)

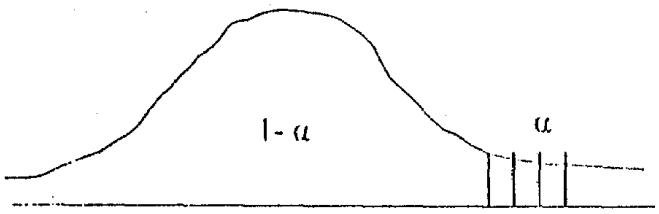
$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$



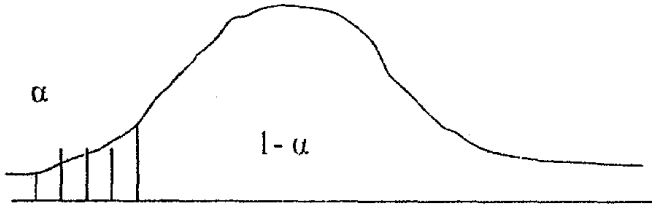
$H_0 : \mu = \mu_0$

$H_1 : \mu > \mu_0$



$H_0 : \mu = \mu_0$

$H_1 : \mu < \mu_0$



خامساً : أتخاذ القرار

توجد أربع قرارات ممكنة فى أى مشكلة من مشاكل اختبارات الفروض وهى كما يلى :-

١- رفض فرض العدم وهو صحيح ويسمى هذا بخطأ النوع الأول وتمثل (α) احتمال وقوع هذا الخطأ.

٢- عدم رفض فرض العدم وهو صحيح وهذا قـــــرار سليم واحتمال اتخاذ هذا القرار هو ($1 - \alpha$).

٣- عدم رفض العدم وهو خاطئ ويسمى هذا بخطأ من النوع الثانى وتمثل بيتا β احتمال وقوع هذا الخطأ.

٤- رفض فرض العدم وهو خاطئ قرار سليم واحتمال اتخاذه ($1 - \beta$). ويوضح الجدول التالي ماسبق:-

القرار	H_0 صحيح	H_0 خاطئ
رفض H_0	قرار خاطئ خطأ من النوع الأول (α)	قرار سليم ($1-\beta$)
عدم رفض H_0	قرار سليم ($1-\alpha$)	خطأ من النوع الثانى β

والاختبارات التى سوف نتناولها هى اختبارات معلمية تتعلق بعينة واحدة أو عينتين . وفى هذه الاختبارات الفروض تدور حول معلمة المجتمع المجهولة وهى الوسط الحسابى والنسبة والتباين.

ولإجراء الاختبارات السابقة نحتاج إلى ما يعرف باسم احصائى الاختبار **Test statistic** تحدد بمعرفة بعض النظريات الإحصائية وهى عبارة عن حقائق ونظريات تم برهانها ويمكن الرجوع إليها فى مراجع الإحصاء .

وهناك طريقتان لاتخاذ القرار فى خامساً وذلك أما برفض أو قبول الفرض العدمى ، الطريقة الأولى تعتمد على قيم محسوبة لاحصائى الاختبار وقيم جدولية له تستخرج من جداول أعدت خصيصاً لذلك الغرض ، وتوجد طريقة أخرى لرفض أو قبول الفرض العدمى لاتتطلب وجود جداول بل تعتمد على قيم تسجلها الحزمة SPSS أو أية حزمة احصائية نلخص الطريقتين كالآتى :-

<u>الطريقة الأولى لرفض أو قبول الفرض العدمى البديل</u>:-

أ- الاختبار الاحصائى قد يكون من طرف **One tail test** أو من طرفى **Two tail test** إذا كان الاختبار من طرفين توجد قيمتين حرجتين نحصل عليهما من جداول التوزيع العينى الذى يتحكم فى الاختبار أحداهما فى الذيل الأيمن والأخر فى الذيل الأيسر (طرفين) للتوزيع العينى تحت الفرض العدمى هذه القيم تحجز مساحة فى كل ذيل مقدراها $\alpha/2$ ، وإذا وقعت القيمة المحسوبة لاحصائى الاختبار من بيانات العينة بين هاتين القيمتين نقبل الفرض العدمى خلاف ذلك نقبل البديل .

ب- إذا كان الاختبار من طرف واحد (طرف أيمن) فإن منطقة الرفض تكون من جهة اليمين فنحسب القيمة الجدولية التى تحجز مسافة فى الذيل مقدارها α ، إذا كانت القيمة المحسوبة اقل من القيمة الجدولية نقبل العدمى والعكس صحيح . وإذا كانت منطقة الرفض جهة اليسار (طرف أيسر) تحسب القيم الجدولية التى تحجز مساحة فى الذيل مقدراها α. إذا كانت القيمة المحسوبة اكبر من القيمة الجدولية نقبل العدمى والعكس صحيح .

الطريقة الثانية لرفض أو قبول الفرض العدمى البديل:-

هناك طريقة أخرى للرفض والقبول استخدمت كثيراً فى الآونة الأخيرة وتعتمد على احتمال محسوب يسمى **p-value** ويرمز له فى الحزمة SPSS وبالرمز . **sig** ويعرف بأنه مستوى معنوية محسوب أو خطأ من النوع الأول محسوب ، هذه الطريقة تستخدم كثيراً فى البرامج الجاهزة ، وتتلخص هذه الطريقة فى الاتى :-

١- يجرى الاختبار بنفس الخطوات السابق الحديث عنها لانحسب قيم جدولية لاحصائى الاختبار ولكن نستخرج احتمال محسوب يرمز له بالرمز . **sig** ، تعبر هذه القيم عن مساحة معينة فى ذيل (أو ذيلين) للتوزيع العينى لاحصائى الاختبار وتسمى قيمة مستوى المعنوية المحسوب.

٢- إذا كان الاختبار من طرفين نقارن هذه القيمة بقيمة $\alpha/2$، إذا زادت القيمة المحسوبة عن $\alpha/2$ نقبل الفرض العدمى والعكس صحيح. إذا كان الاختبار من طرف واحد نقارن القيمة المحسوبة بقيمة α، إذا ما زادت عنها نقبل الفرض العدمى والعكس صحيح .

٣-٢ : إجراء اختبارات الفروض الإحصائية باستخدام برنامج SPSS

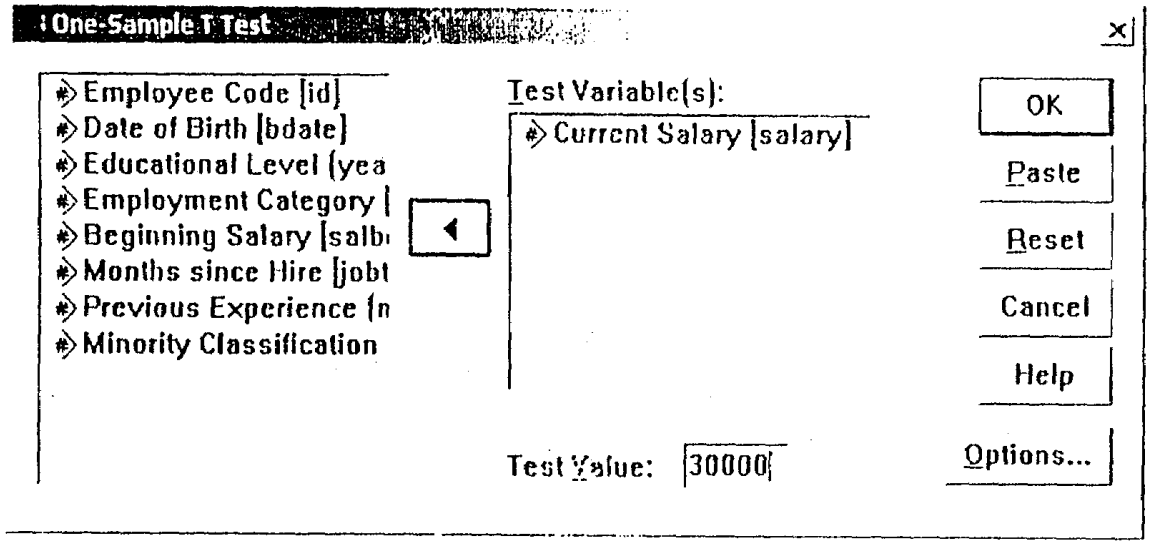
٣-٢-١ : اختبار ت (T test) : للمقارنة بين المتوسطات :

يستخدم اختبار ت (T test) للمقارنة بين متوسطى عينتين لمعرفة هل الفرق بين هـذين المتوسطين معنوياً أو غير معنوى ، ويستخدم هذا الاختبار فى البيانات الكمية ، قد تكون هاتـان العينتان مستقلتان ، أو غير مستقلتين ، أو قد تكون عينة واحدة ويراد مقارنتها بقيمة معينة .

١ – اختبارات عينة واحدة

لإجراء اختبار عينة واحدة ، أى مقارنة متوسط عينة واحدة بقيمة معينة ، فعلى سبيل المثال لاختبار أن متوسط الراتب السنوى للعاملين فى الملف Employee Data.sav يعادل ٣٠٠٠٠ دولار نتبع الآتى :

- افتح الملف Employee Data.sav
- من قائمة Analyze اختر أمر Compare Means ، ثم One Sample T Test فيظهر الصندوق الحوارى One Sample T Test



- اختر المتغير Salary وهو المتغير المراد اختبار المتوسط بالنسبة له فى خانة Test Variable (s)
- اكتب المتوسط المراد المقارنة به هو ٣٠٠٠٠ فى خانة Test Value
- اختر أمر Ok فتظهر النتائج التالية :

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
SALARY Current Salary	474	$34,419.57	$17,075.661	$784.311

One-Sample Test

	Test Value = 30000					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence terval of the Difference Lower	Upper
SALARY Current Sala	5.635	473	.000	$4,419.57	2,878.40	5,960.73

من النتائج نلاحظ ما يلى :

١– الفرض العدمى : متوسط الراتب السنوى لا يختلف معنويا عن ٣٠٠٠٠ دولار .

٢– الفرض البديل : يوجد فرق معنوى بين متوسط الراتب السنوى والقيمة ٣٠٠٠٠ دولار.

القرار المتخذ :نلاحظ أن Sig (2-tailed) أقل من مستوى المعنوية (٠,٠٥) لـذلك نـرفض الفرض العدمى ، ونقبل الفرض البديل أى أنه هناك اختلاف معنوى بين متوسط المرتبات والقيمة ٣٠٠٠٠ دولار .

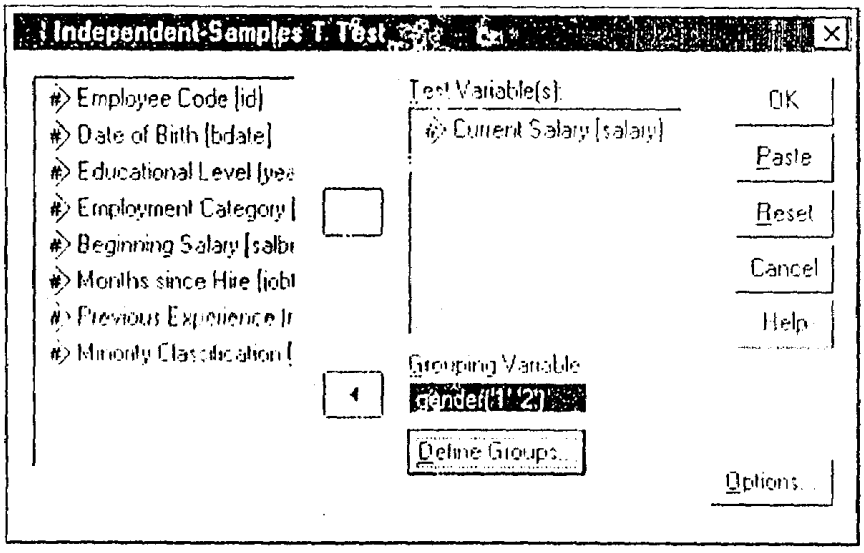
٢– اختبار عينتين مستقلتين Compare Means

المقارنة بين عينتين مستقلتين :

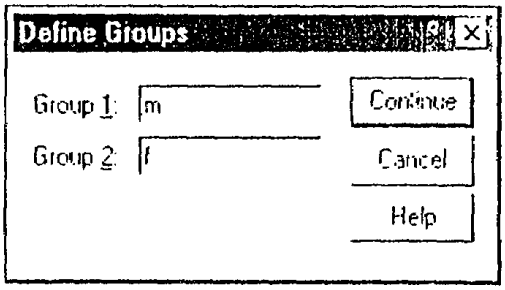
فى هذه الحالة يكون لدينا متغير تابع يتم تقسيمه إلى مجموعتين بناء على متغير آخـر مسـتقل ، ويراد معرفة معنوية الفرق بين هذين المتوسطين ، ولتنفيذ ذلك تابع الخطوات التالية :

١– افتح الملف Employee Data.sav

٢– من قائمة Analyze اختر أمر Compare Means ، ومنها اختـر Independent-samples T Test يظهر الصندوق الحوارى التالى :



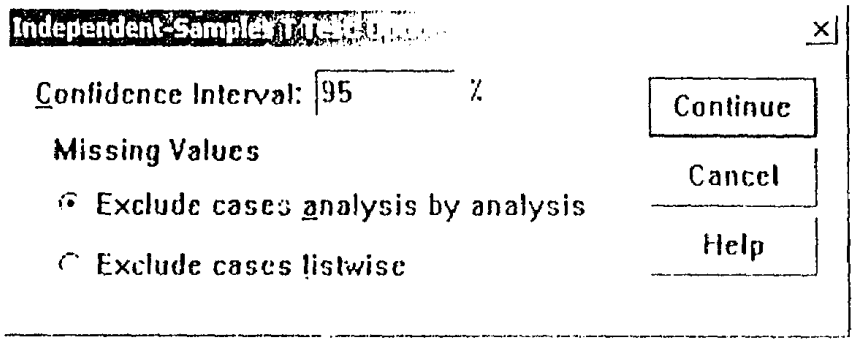
٢– أمام خانة Test Variable(s) اختر المتغير أو المتغيرات المراد المقارنة بين متوسطاتها ، وليكن salary



٣- أمام خانة Grouping Variable(s) اختر المتغير الذى يقسم المتغيرات السابقة إلى مجموعتين ، وليكن gender.

٤- اختر Define Groups لتحديد قيمتى هذا المتغير .

٥- اختر Options لتحديد درجة الثقة Confidence Interval ، إمكانية إظهار Labels ، وطريقة معاملة الـ Missing Values .



بعد تنفيذ الاختيارات السابقة تظهر النتائج التالية :

T-Test

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Current Salary	Male	258	$41,441.78	$19,499.21	$1,213.97
	Female	216	$26,031.92	$7,558.02	$514.26

Independent Samples Test

			Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variance	F		119.669	
	Sig.		.000	
t-test for Equality of Means	t		10.945	11.688
	df		472	344.262
	Sig. (2-tailed)		.000	.000
	Mean Difference		$15,409.86	$15,409.86
	Std. Error Difference		$1,407.91	$1,318.40
	95% Confidence Interv of the Difference	Lower	$12,643.32	$12,816.73
		Upper	$18,176.40	$18,003.00

نلاحظ ما يلى

البرنامج يقوم باختبار للتجانس أولاً ، ويحسب قيمة T فى الحالتين ، فى حالة التجـانس Equal variances assumed ، وفى حالة عـدم التجـانس Not equal variances assumed ، ولإجراء اختبار التجانس يصاغ الفرض العدمى والبديل ، والقرار المتخذ كالآتى

١- الفرض العدمى : العينتان متجانستان ، أى لا يوجد فرق معنوى بين تباينهما

٢- الفرض البديل : العينتان غير متجانستين .

القرار المتخذ : قيمة F = 119.669 بمستوى معنوية أقل من ٠,٠٠٠٥ ، لذا فإننا نـرفض الفرض العدمى ، ونقبل الفرض البديل ، أى أن العينتين غير متجانستين ، ولإجراء اختبـار T وهو الاختبار الأساسى للفرق بين المتوسطين ، يكون اختبارنا بناءً علـى الاختيـار Equal variance not assumed ، ويصاغ الفرض العدمى والبديل والقرار المتخذ كالتالى :

١- الفرض العدمى : ليس هناك اختلاف بين متوسطى مرتبات الذكور والإناث أى (المتوسطان متساويان) .

٢- الفرض البديل : يوجد اختلاف بين متوسطى أعمار الـذكور والإنـاث أى (المتوسـطان مختلفان) .

٣- القرار المتخذ : نلاحظ أن قيمـة T = ١١,٦٦٨ ، بـدرجات حريـة = ٣٣٤,٢٦٢ ، ومستوى المعنوية (Sig (tailed أقل من (٠,٠٥) لذلك نرفض الفرض العدمى ، ونقبـل الفرض البديل أى أنه هناك اختلاف بين متوسطى مرتبات الذكور والإناث ، أى أن متوسط مرتبات الذكور ٤١٤٤١,٧٨ أعلى من متوسط مرتبات الإناث ٢٦٠٣,٩٢ حيث يوجد فرق معنوى بين المتوسطين .

ملحوظة : فى حالة Equal variance not assumed يجرى البرنامج تحويلة على قيمة T ، وعلى درجات الحرية لتخليصهما من اثر عدم التجانس ، لذلك من المحتمل أن نجد درجات الحرية بها قيمة كسرية

٣- المقارنة بين عينتين غير مستقلتين :

قد يراد المقارنة بين عينتين مرتبطتين ، كما لو أخذنا قراءات معينة لمتغير معين قبل إجراء معالجة معينة ، وبعد إجراء هذه المعالجة ، كمتوسط المبيعات لمجموعة معينة قبل إجراء دورة تدريبية فى فن البيع ، وبعد إجراء هذه الدورة ، ويراد معرفة هل الفرق بين هذين المتوسطين معنوياً أم لا .

مثال :

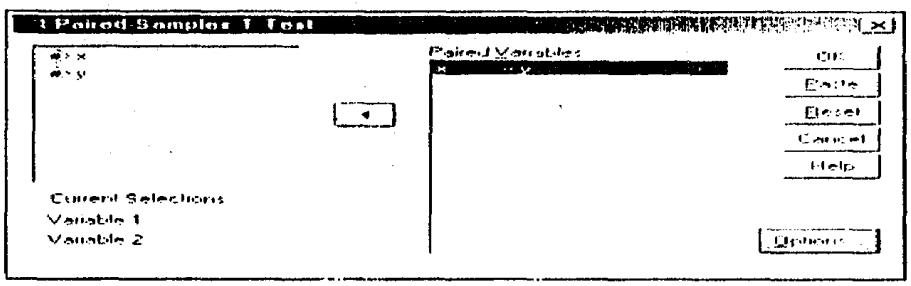
أراد طبيب أن يقرر أثر نوع معين من الدواء على تخفيض ضغط الدم فاختار عينة من ١٥ فرداً ، وأخذ قياس ضغط الدم قبل تعاطى الدواء (X) ، وبعد تعاطى الدواء (y) .

x	٧٠	٨٠	٧٢	٧٦	٧٦	٧٦	٧٢	٧٦	٧٦	٨٢	٦٤	٧٤	٩٢	٧٤	٨٤
y	٦٨	٧٢	٦٢	٧٠	٥٧	٦٦	٦٨	٥٢	٦٤	٧٢	٧٤	٦٠	٧٤	٧٢	٧٤

ولتنفيذ ذلك تابع الخطوات التالية :

١- سجل البيانات فى ملف Data ، ولتكن اسماء المتغيرات x , y .

٢- من قائمة Analyze اختر أمر Compare Means ، ومنها اختر Paired-samples T Test يظهر الصندوق الحوارى التالى :



٢- اختر المتغيرين المراد المقارنة بين متوسطيهما، وهما المتغيران X,Y ، وانقلهما إلى القائمة Paired Variables ، ثم اختر أمر OK تظهر النتائج التالية :

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	X	75.2000	15	6.7103	1.7326
	Y	67.0000	15	6.7718	1.7485

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	X & Y	15	-.270	.330

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 X - Y	8.2000	10.7451	2.7744	2.2496	14.1504	2.956	14	.010

يصاغ الفرض العدمى والبديل كالتالى :

الفرض العدمى : ليس هناك فرق معنوى بين المتوسطين (أى قبل تعاطى الدواء ، وبعد تعاطى الدواء) .

الفرض البديل : يوجد فرق معنوى بين المتوسطين

القرار المتخذ : حيث أن الـ (Sig (2-tailed أقل من ٠,٠٥ لذلك نرفض الفرض العدمى القائل بتساوى المتوسطين (قبل وبعد الدواء) ونقبل الفرض البديل القائل بأن هناك فرق فى متوسط ضغط الدم ، حيث أنه انخفض بعد الدواء إلى ٦٧ فى حين كان قبل الدواء ٧٥,٢ أى أن الدواء له أثر على تخفيض ضغط الدم

٣-٢-٢ اختبار F لتحليل التباين :

اختبار تحليل التباين يعتمد أساساً على احصائى اختبار يطلق عليه اسم F نسبة إلى توزيع احتمالى شهير يسمى **F Distribution** هذا التوزيع له تطبيقاته العديدة فى اختبار الفروض ، ويرتبط تحليل التباين بعلم تصميم التجارب الذى يستخدم بكثرة فى مجال البحوث الزراعية ومجال اختبار فعالية طرق تدريب وتدريس مختلفة أو أنواع مختلفة من الأدوية ... وهو يعتبر اختبار للمتوسطات ولكن بين أكثر من عينتين.

وبصفة عامة فى اختبار تحليل التباين يحسب تقديراً إجماليا لتباين المجتمع ثم يقسم إلى جزئين الأول يسمى التباين بين المجموعات **Between groups** والثانى يسمى التباين داخل المجموعات **Within groups** ثم بعد ذلك يتم حساب احصائى الاختبار يعتمد على النسبة بين هذين التباينين .

ويكون الفرض العدمى والبديل لهذا الاختبار كالتالى:-

الفرض العدمى : متوسطات المجتمعات المحسوبة منها العينات متساوية

الفرض البديل : يوجد زوج واحد على الأقل من المجتمعات المحسوبة منها العينات متوسطاته مختلفة معنوية .

ويشترط لاستخدام اختبار تحليل التباين توفر عدة شروط أهمها:-

١- أن تكون العينات عشوائية مستقلة ، ويتم التحقق من هذه الشروط عند سحب العينات.

٢- أن تكون العينات مسحوبة من مجتمعات موزعة توزيعاً طبيعياً.

٣- أن تكون وحدة القياس بقترة (بيانات مستمرة).

٤- أن تكون تباينات المجتمعات متساوية . ويستخدم فى ذلك عدد من الاختبارات مثل اختبار " بار تليت **Bartlet** " أو اختبار "هارتلى **Hartly** "

لإجراء اختبار تحليل التباين يكون لدينا متغير مستقلاً ، التغير الذى حدث فى ذلك المتغير يمكن أن يكون فى اتجاه واحد طبقاً لشروط معينة، والاختبار المستخدم فى هذه الحالة يسمى اختبار فى اتجاه واحد . وإذا كنا نهتم بالتغير الذى حدث على المتغير المستقل فى اتجاهين فإن الاختبار فى هذه الحالة يسمى تحليل التباين فى اتجاهين.

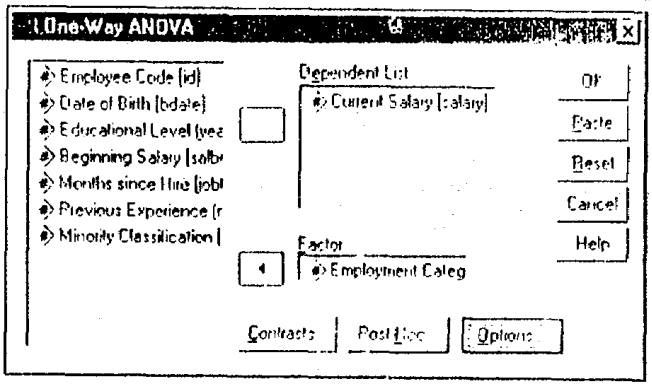
تحليل التباين فى اتجاه واحد باستخدام برنامج SPSS:

تحليل التباين فى اتجاه واحد يعنى دراسة أثر عامل واحد (متغير مستقل واحد) على المتوسطات (متغير تابع) .

مثال : فى بيانات employee data.sav نريد أن نعرف هل هناك فرق بين متوسط الرواتب salary بين مجموعات الوظائف jobcat ، ولتنفيذ ذلك تابع الخطوات التالية :

١- افتح الملف Employee Data.sav .

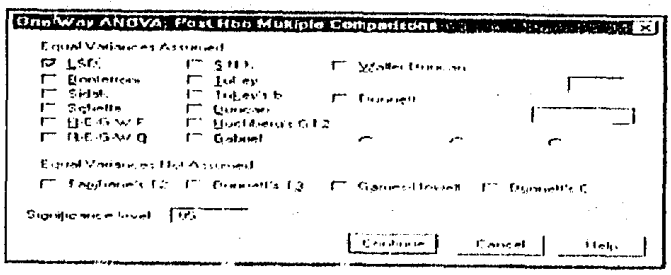
٢- من قائمة Analyze اختر أمر Compare Means ومنها اختر أمر One-way ANOVA يظهر الصندوق الحوارى التالى :



٢- أمام خانة Dependent list اختر المتغير التابع salary .

٣- أمام خانة Factor اختر المتغير الذى يقسم المتغير التابع إلى مجموعات jobcat .

٤- أنقر أمر Post Hoc لتحديد الاختبارات الإحصائية التى تظهر الأزواج من المتوسطات التى بينها اختلاف ، وأشهر هذه الاختبارات LSD .



٦- اختر أمر Continue لنرجع إلى الصندوق الحوارى الأصلى ، ثم اختر أمر OK تظهر النتائج التالية :

Oneway

ANOVA

Current Salary

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	89438483926	2	44719241963.0	434.481	.000
Within Groups	48478011510	471	102925714.459		
Total	137916495436	473			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: Current Salary

LSD

(I) Employment Cate	(J) Employment Cate	Mean Difference (I-J)	td. Error	Sig.	5% Confidence Interva	
					Lower Bound	Upper Bound
Clerical	Custodial	3,100.35	,023.76	.126	$7,077.06	$876.37
	Manager	6,139.26*	,228.35	.000	38,552.99	33,725.53
Custodial	Clerical	3,100.35	,023.76	.126	-$876.37	$7,077.06
	Manager	3,038.91*	,244.41	.000	37,449.20	28,628.62
Manager	Clerical	6,139.26*	,228.35	.000	33,725.53	38,552.99
	Custodial	3,038.91*	,244.41	.000	28,628.62	37,449.20

The mean difference is significant at the .05 level.

تحليل النتائج

١- الفرض العدمى : المتوسطات متساوية بين مختلف الوظائف

٢- الفرض البديل : يوجد زوج واحد على الأقل بين أنواع الوظائف مختلف فى المتوسط .

٣- القرار المتخذ من تحليل التباين اتضح أننا نرفض الفرض العدمى القائل بأن المتوسطات بين الرواتب salary حسب طبيعة الوظيفة jobcat متساوية ، ونقبل الفرض البديل القائل بأن هناك فرق بين المتوسطات ، حيث أن قيمة F 484.481 ، ومستوى المعنوية أقل من ٠,٠٠٠٥

٤- من تحليل المقارنات المتعددة Multiple Comparisons يوجد فرق معنوى بين متوسط راتب الـ Manager و كل من الـ Clerical والـ Custodial ولا يوجد فرق معنوى بين الـ Clerical والـ Custodial فى متوسط الراتب

تحليل التباين فى اتجاهين أو أكثر :

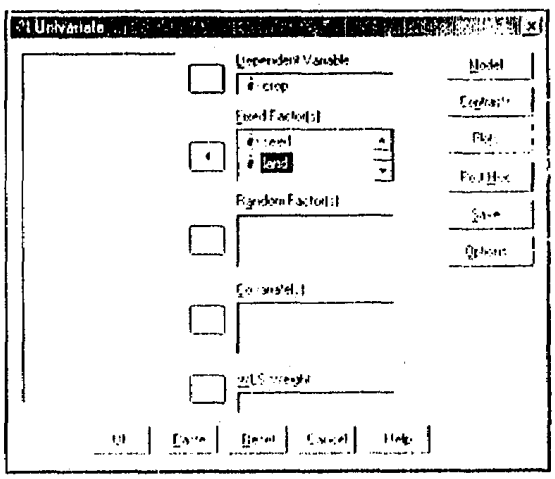
فى حالة تحليل التباين فى اتجاهين أو أكثر سنقارن الفرق بين المجموعات لمتغير تابع (كمى) بعد تقسيم الحالات طبقاً لخصائص متغيرين مستقلين أو أكثر فى نفس الوقت (متغيرات وصفية) .

مثال : أجريت تجربة لاختبار تأثير اختلاف التربة (٣ قطع من الأرض) ، واختلاف نوع القمح (٤ سلالات) على محصول الحبوب ، وكان ناتج المحصول كالتالى (بالطن)

قطعة الأرض	السلالة			
	(١)	(٢)	(٣)	(٤)
(١)	١١	٢٤	١٩	٢٣
(٢)	١٢	٢٥	١٨	١٥
(٣)	١٠	٢١	١٧	٢٧

يتم تسجيل البيانات على ملف ، وليكن Anova2 كالتالى :

land	seed	crop
١	١	١١
١	٢	٢٤
١	٣	١٩
١	٤	٢٣
٢	١	١٢
٢	٢	٢٥
٢	٣	١٨
٢	٤	١٥
٣	١	١٠
٣	٢	٢١
٣	٣	١٧
٣	٤	٢٧



لاختبار أثر عاملى (التربة ، وسلالة القمح) على الإنتاج نتبع الآتى :

١ - افتح الملف Anova2.sav .

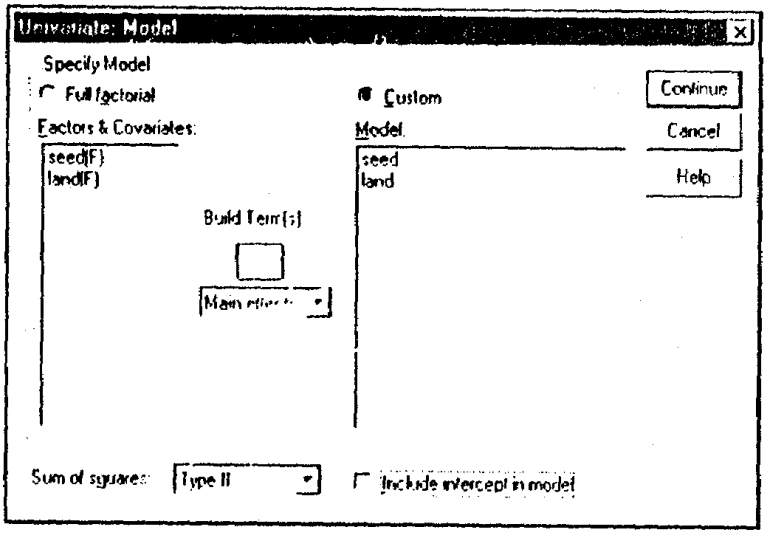
٢- من قائمة Analyze اختر أمر General Linear Model ، ومنها اختر Univariate يظهر الصندوق الحوارى Univariate :

٣- أمام خانة Factors Fixed اختر المتغيرات المستقلة land , seed

٤- أمام Dependent Variable اختر المتغير الكمى التابع Crop

٥- اختر Model فيظهر الصندوق الحوارى Model ، أمام خانة Build Term(s) اختر Main Effects ، وأمام خانة Model اختر المتغيرين seed, land واحداً تلو الآخر .

٦- أمام خانة Sum of squares اختر Type II ، ولا تنشط خانة Include intercept in model .



٢- اختر أمر Continue فنرجع إلى الصندوق الحوارى الأصلى ، ثم اختر أمر OK فتظهر النتائج التالية :

Univariate Analysis of Variance

Between-Subjects Factors

		N
SEED	1.00	3
	2.00	3
	3.00	3
	4.00	3
LAND	1.00	4
	2.00	4
	3.00	4

Tests of Between-Subjects Effects

Dependent Variable: CROP

Source	Type II Sum of Squares	df	Mean Square	F	Sig.
Model	4383.167[a]	6	730.528	54.225	.000
SEED	269.667	3	89.889	6.672	.024
LAND	6.500	2	3.250	.241	.793
Error	80.833	6	13.472		
Total	4464.000	12			

a. R Squared = .982 (Adjusted R Squared = .964)

تصاغ الفروض العدمية والبديلة كما يلى

الفروض العدمية

١- لا يوجد أثر لسلالة البذور على المحصول

٢- لا يوجد أثر لنوع قطعة الأرض

الفروض البديلة

١- يوجد أثر لسلالة البذور على المحصول

٢- يوجد أثر لنوع قطعة الأرض

القرار المتخذ

١- قيمة F بالنسبة لسلالة البذور 6.672 ومستوى المعنوية 0.024 ، لذلك فإننا نرفض الفرض العدمى ، ونقبل الفرض البديل ، أى يوجد تأثير للبذور على كمية المحصول .

٢- قيمة F بالنسبة لقطع الأرض 0.241 ، ومستوى المعنوية 0.793 ، لذلك فإننا نقبل الفرض العدمى ، أى لا تأثير لقطع الأرض على كمية المحصول .

٣-٢-٣ اختبار كا^2 Chi – Square Test

يعتبر أختبار كا^{2}من الاختبارات التى قد تستخدم كاختبار معلمى واختبار لامعلمى. فى حالة اختبار الفروض حول تباين المجتمع فإننا نطبق اختبار كا^2، وفى هذه الحالة يكون الاختبار معلمى. ويوجد اختبارين معلمين يستخدم فيهما الاحصائى كا^2، أولهما اختبار جودة التوفيق لمتغير واحد غير مقيد وثانيهم اختبار الاستقلال بين المتغيرات الغير مقيدة ، ويبنى الاختبار على مقارنة التكرارات المشاهدة والتكرارات المتوقعه لمعرفة هل هناك فرقاً معنوياً بينهما أم لا ؟

والشروط اللازمة لتطبيق اختبار كا^2 :-

١-عشوائية العينة **Randam sampling** : العينة يجب أن تكون مختارة عشوائياً من مجتمع بإرجاع، أى أن مفردات العينة الواحدة مستقلة بعضها عن البعض .

٢-استقلال المشاهدات **independence of observatios** فى هذا نشترط أن كل مشاهدة مستخدمة فى الاختبار يجب أن تكون مأخوذة من مصدر مستقل عن المشاهدة الأخرى.

٣-حجم المشاهدات المتوقعة **Size of Expected Frequencies** :حجم العينة المستخدم فى التحليل يفضل أن يكون اكبر من ٣٠، عندما يكون حجم العينة صغيراً ، وعدد الخلايا الداخلة فى الاختبار اقل من عشرة خلايا فان اقل تكرار متوقع مطلوب للاختبار هو خمسة تكرارات للخلية الواحدة.

٤-حجم التكرارات المشاهدة **Size of observed Frequencies** : التكرارات المشاهدة ممكن أن تكون صفر أو أقل من خمسة تكرارات فى الخلية فلا شروط عليها.

ملاحظة : فى حالة عدم تحقيق بعض هذه الشروط يمكن إجراء عمليات ضم للتكرارات المشاهده والمتوقعة بقدر المستطاع.

١- اختبار كا^2 لجودة التوفيق Chi –square Goodness-of –fit Test فى اختبار جودة التوفيق يكون لدينا تكرارا مشاهد تم جمعه مسبقاً ولدينا نظرية إحصائية (أو فرض معين) يستخدم فى حساب التكرار المتوقع المناظر. ويكون الفرض العدمى والبديل بصفة عامة يكون كالآتي :-

الفرض العدمى : لا يوجد فرق بين التكرار المشاهد والمتوقع

الفرض البديل : يوجد فرق بين التكرار المشاهد والمتوقع.

٢- اختبار مربع كا^2 للاستقلال Chi-square Test for independence

نحتاج فى حالات كثيرة إلى التعرف عما إذا كانت هناك علاقة بين صفتين من صفات مجتمع ما . مثلاً قد نحتاج لمعرفة هل توجد علاقة بين مستوى الدخل والمستوى التعليمى ؟ أو هل توجد علاقة بين لون العينتين ولون الشعر فى مجتمع ما؟ للإجابة على مثل هذه الأسئلة يجب أن نختار عينة عشوائية من المجتمع ثم تصنف مشاهدات العينة حسب مستويات كل صفة من الصفتين ووضعها فى جدول يسمى جدول التوافق Contingency Table كما يلى :-

	B_1	B_2	.....	B_j	....	B_s	
A_1	O_{11}	O_{12}	.....	O_{1j}	....	O_{1s}	u_1
A_2	O_{21}	O_{22}	.....	O_{2j}	....	O_{2s}	u_2
A_i	O_{i1}	O_{i2}	.....	O_{ij}	....	O_{is}	u_i
A_r	O_{r1}	O_{r2}	.....	O_{rj}	....	O_{rs}	u_r
المجموع	v_1	v_2	.....	v_j	.....	v_s	n

حيث مستويات الصفة A هى $A_1, A_2 \ldots\ldots A_r$

مستويات الصفة B هى $B_1, B_2 \ldots\ldots B_s$

وترمز O_{ij} إلى عدد المشاهدات التى يتوفر فيها المستويان A_i من الصفة A , B_j من الصفة B

ويرمز u_i إلى عدد المشاهدات التى يتوفر فيها المستوى A_i من الصفة A

ترمز v_i إلى عدد المشاهدات التى يتوفر فيها المستوى B_j من الصفة B

ويكون فرض العدم والبديل كمايلى :-

فرض العدم H_o : الصفتان A ، B مستقلتان.

الفرض البديل H_1 : الصفتان A ، B غير مستقلتين .

إجراء اختبار كا٢ باستخدام البرنامج الإحصائى SPSS

العرض الإحصائى لمتغيرين

الجداول التكرارية المزدوجة Crosstabs

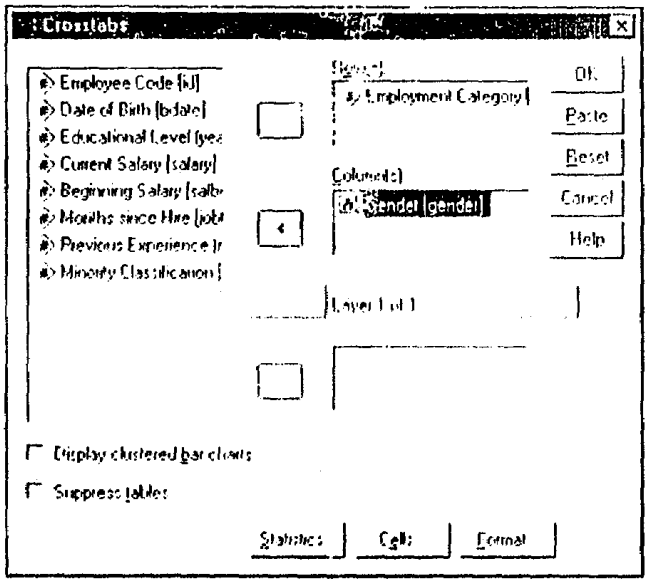
تمثل العلاقة بين متغيرين ، عن طريق تمثيل هذه العلاقة فى صورة متغير صفى ، ومتغير عمودى (Crosstab) ، كما يعطى عدة مقاييس إحصائية لقياس العلاقة بين هذين المتغيرين ، طبيعة المتغيرات (اسمية ، ترتيبية ، رقمية) هى التى تحدد المقياس المناسب لقياس هذه العلاقة ، ولكى يتم تمثيل البيانات المستمرة Continuous ينبغى إعادة تكويدها وتحويلها إلى فئات .

ولعمل جدول تكرارى مزدوج نتبع الآتى :

١. افتح الملف Employee Data.sav

٢. من قائمة Statistics اختر أمر Analyze ، ومنها اختر Descriptive Statistics ومنها اختر Crosstabs.

يظهر الصندوق الحوارى التالى :

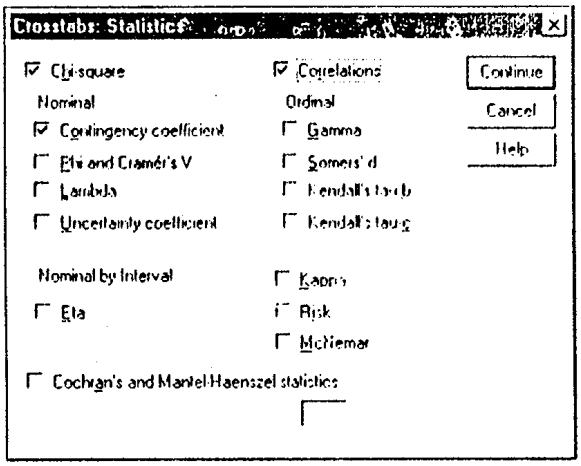


٢. اختر المتغيرات المراد تمثيلها كصفوف أمام خانة Row(s) ، وليكن Jobcat

٣. اختر المتغيرات المراد تمثيلها كأعمدة أمام خانة Column(s) ، وليكن gender

٤. اختر الـ Control Variables أمام خانة Layer ، ويلاحظ أنه سيتم إنشاء جدول تكرارى مزدوج لكل قيم من قيم الـ Control Variable

٥. لحساب المقاييس الإحصائية المناسبة ، اختر أمر Statistics من الصندوق الحوارى Crosstabs ، واختر أمر Chi-Square , Correlation , Contingency Coefficient



٦. بالنسبة للجداول ٢×٢ تسمى جداول الاقتران ، الجداول أعلى من ذلك C×R بمعنى أى عدد من الصفوف وأى عدد من الأعمدة تسمى جداول التوافق .

٧. بالنسبة لجداول الاقتران ٢×٢ يستخدم لقياس معنوية العلاقة Chi-Square حيث يحسب Person chi-squre , Likelihood chi-square , Fisher's exact test and Yate's corrected chi-square (continuity correction) .

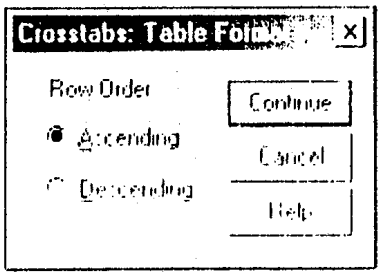
٨. بالنسبة للبيانات المقاسة بمقياس اسمى Nominal Data أى البيانات الوصفية مثل النوع أو الديانة توجد بالإضافة إلى Chi-Square مقاييس معامل التوافق Contingency coefficient ، .... الخ .

٩. بالنسبة للبيانات المقاسة بمقياس ترتيبى Ordinal بالإضافة إلى Chi-square يمكن اختيار Gamma , Somer'd ،

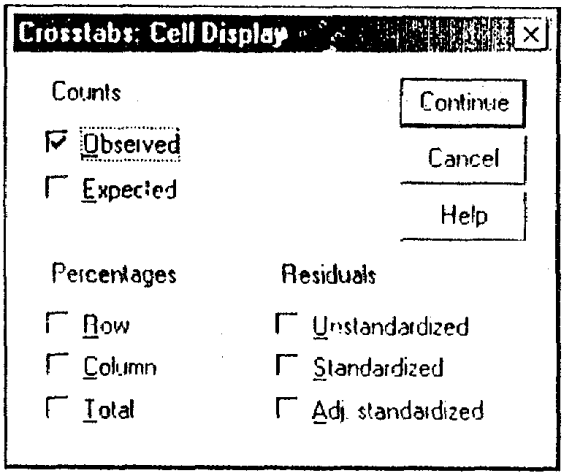
١٠. حينما يكون أحد المتغيرين وصفيا Category والآخر رقمى Numerical (Numeric by Interval) يمكن اختيار Eta , Kappa

١١. فى المتغيرات الرقمية يمكن اختيار معامل الارتباط الخطى Correlation ، وفى جداول التوافق عندما يكون عدد الصفوف يساوى عدد الأعمدة نختار Cappa

١٢. يمكن اختيار Format للتحكم فى شكل المخرجات كالتالى :



١٣. عادة تظهر فى الخلايا القيم المشاهدة ، ولكن باختيار الأمر cell يظهر الصندوق الحوارى التالى :



ومنه يمكن إظهار فى خلايا الجدول البيانات التالية :

(١) أعداد Counts :

- القيم المشاهدة Observed
- القيم المتوقعة Expected

(٢) البواقى Residual : وهو الفرق بين القيم المشاهدة والمتوقعة :

- غير قياسى Unstandardized.
- قياسى. Standardized
- قياسى معدل Adj-standardized

(٣) نسب مئوية Percentages

- Row : نسبة لإجمالى الصف .
- Column : نسبة لإجمالى العمود .
- Total : نسبة للإجمالى العام .

بعد اختيار أمر Continue ، فنرجع إلى الصندوق الحوارى الأصلى ، فنختار أمر OK فتظهر لنا النتائج التالية :

Crosstabs

Warnings

CORR statistics are available for numeric data only.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Gender * Employment Category	474	100.0%	0	.0%	474	100.0%

Gender * Employment Category Crosstabulation

			Employment Category			
			Clerical	Custodial	Manager	Total
Gender	Female	Count	206		10	216
		% within Employment Category	56.7%		11.9%	45.6%
	Male	Count	157	27	74	258
		% within Employment Category	43.3%	100.0%	88.1%	54.4%
Total		Count	363	27	84	474
		% within Employment Category	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	79.277[a]	2	.000
Likelihood Ratio	95.463	2	.000
N of Valid Cases	474		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 12.30.

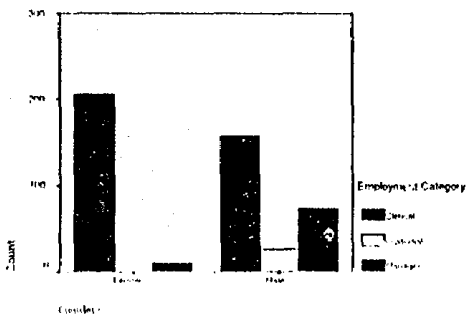
Symmetric Measures[c]

		Value	Approx. Sig.
Nominal by Nominal	Contingency Coefficient	.379	.000
N of Valid Cases		474	

a. Not assuming the null hypothesis.

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

c. Correlation statistics are available for numeric data only.



تحليل البيانات من جدول تكرارى مزدوج :

قد تعطى البيانات فى جدول تكرارى مزدوج ، ونريد حساب المقاييس الإحصائية المناسبة :

مثال :لدينا التوزيع المشترك لخاصيتى الوفاة والتدخين لعينة من كبار السن ممثلة فى الجدول التالى :

الوفاة	التدخين		المجموع
	يدخن	لا يدخن	
١ توفى	٥٤	١١٧	١٧١
٢ حى	٣٤٨	٩٥٠	١٢٩٨
المجموع	٤٠٢	١٠٦٧	١٤٦٩

والمطلوب اختبار الفرض القائل بأن هناك علاقة بين التدخين والوفاة ، تمثل البيانات بالطريقة التالية :

x1	x2	fre
1	1	54
1	2	117
2	1	348
2	2	950

	x1	x2	x
1	توفى	يدخن	54.00
2	توفى	لايدخن	7.00
3	حي	يدخن	348.00
4	حي	لايدخن	950.00

وبعد تخصيص Value Label للبيانات وتسجيلها تظهر فى نافذة Data Editor بالشكل

١) نقوم بترجيح البيانات بالمتغير fre عن طريق اختيار weight cases من قائمة Data .

٢) نقوم بإجراء Crosstab ونختار X1 كمتغير صفى ، X2 كمتغير عمودى ونختار المقاييس الإحصائية المناسبة فتظهر لنا النتائج التالية :

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
X1 * X2	1359	100.0%	0	.0%	1359	100.0%

X1 * X2 Crosstabulation

			X2		Total
			يدخن	لايدخن	
X1	توفى	Count	54	7	61
		% within X2	13.4%	.7%	4.5%
	حى	Count	348	950	1298
		% within X2	86.6%	99 3%	95.5%
Total		Count	402	957	1359
		% within X2	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	106.526[b]	1	.000		
Continuity Correction[a]	103.584	1	.000		
Likelihood Ratio	97.860	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	106.447	1	.000		
N of Valid Cases	1359				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 18.04.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Contingency Coefficient	.270	.000
N of Valid Cases		1359	

a. Not assuming the null hypothesis.

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

الباب الرابع

الارتباط والانحدار

٤-١ الارتباط

لقياس العلاقة بين متغيرين ، ويستخدم معامل الارتباط الخطى لبيرسون لقياس العلاقة الخطية بين متغيرين (كميين) وتتراوح قيمته بين -١ ، ١ ، وكلما اقترب من الواحد كلما دل ذلك على قوة العلاقة ، وكلما اقترب من الصفر دل ذلك على ضعف العلاقة ، كما أن إشارة المعامل (موجبة أو سالبة) تدل على اتجاه العلاقة ، فإذا كانت الإشارة موجبة كانت العلاقة طردية ، وإذا كانت الإشارة سالبة كانت العلاقة عكسية .

هناك عدة معاملات للارتباط ، ففى حالة العلاقة بين متغيرين ترتيبيين يستخدم معامل كندال تـــاو B ، سبيرمان للرتب .

وللحصول على معامل الارتباط باستخدام البرنامج الإحصائى SPSS نتبع الآتى :

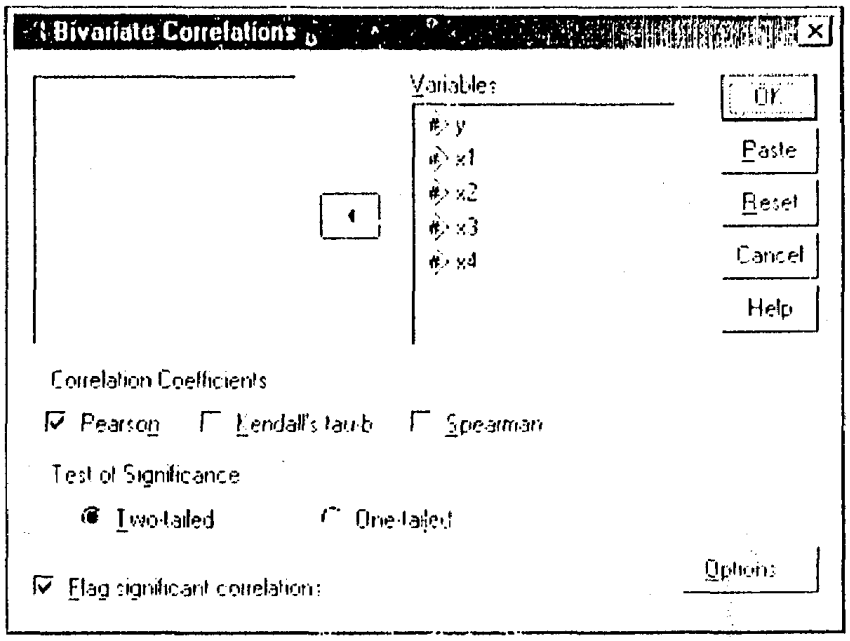
إذا أتيحت لك البيانات التالية ، أوجد معامل الارتباط الخطى بين المتغيرات

الأجـــور الزراعية X4	العمالة الزراعية بالألف X3	المســـاحة الزراعية X2	الاستثمارات الزراعيـــة الثابتة x1	الناتج المحلى الزراعى الثابت y	السنوات Year
٢٦٧,٤٠	٤٢١٨,٠٠	٥,٧٩	٨٣,٤٠	٩٩٦,٩٠	١٩٧٥
٢٦٢,١٠	٤٠٦٨,٠٠	٥,٨٦	٩٨,٠٠	١٠٦٣,٧٠	١٩٧٦
٢٨٧,٧٠	٤١٠٤,٠٠	٥,٨٥	١١٤,٥٠	١٦١٧,٨٠	١٩٧٧
٣١٦,٢٠	٤١٣٥,٠٠	٥,٨٦	١٤٤,١٠	١٥١٥,٧٠	١٩٧٨
٣٤٧,٨٠	٤١٦٥,٠٠	٥,٨٦	١٨٧,٢٠	١٦٩٠,٨٠	١٩٧٩
٤٤٠,٣٠	٤٢٠٠,٠٠	٥,٨٧	٢٠٤,٤٠	١٧٠٧,٢٠	١٩٨٠
٩٧١,١٠	٤٢٢٥,٠٠	٥,٨٨	٤٢٣,٦٠	١٨٨٢,١٠	١٩٨١
١٠٥٥,٦	٤٢٦٤,٠٠	٥,٩٢	٤١٦,٣٠	٣٤٥٥,٦٠	١٩٨٢
١١٤١,١	٤٣٠٤,٠٠	٥,٨٤	٤٠٩,٨٠	٤٢٠٤,٩٠	١٩٨٣
١٢٢٨,٩	٤٣٤٤,٠٠	٥,٨٣	٤٥٢,٨٠	٤٠٦١,١٠	١٩٨٤
١٢٧١,٩	٤٣٨٤,٠٠	٥,٩٧	٤٩٥,٨٠	٤٠١٨,٧٠	١٩٨٥
١٥٨٦,١	٤٣٣٠,٠٠	٥,٩٧	٧٢٣,٥٠	٤٥٤٠,٠٠	١٩٨٦

السنوات Year	الناتج المحلى الزراعى الثابت y	الاستثمارات الزراعية الثابتة x1	المساحة الزراعية X2	العمالة الزراعية بالألف X3	الأجور الزراعية X4
١٩٨٧	٥٦١٠,٠٠	١٥١٥,٢٠	٦,٠٠	٤٣٨١,٠٠	١٨٥٥,٩
١٩٨٨	٥٧٨١,٠٠	١٦٦٧,٩٠	٦,١٨	٤٤٣٢,٠٠	٢١٠٥,٤
١٩٨٩	٥٩٣٥,٠٠	١٨٢١,٣٠	٦,٢٧	٤٤٧٨,٠٠	٢٣٨٢,٤
١٩٩٠	٦١٣٠,٠٠	١٩٠٥,٢٠	٦,٩٢	٤٥٣٣,٠٠	٢٦٩٨,٠
١٩٩١	٧٩٣٩,٠٠	١٩٩٥,٥٠	٧,٠٢	٤٥٨٥,٠٠	٣٠١٢,٠
١٩٩٢	٨١٣٦,٠٠	٢٢٨٨,٥٠	٧,١٢	٤٦٢٤,٠٠	٣٤٢٣,٠
١٩٩٣	٨٤٤٨,٠٠	٢٧٠٦,١٠	٧,١٧	٤٦٨٢,٠٠	٣٨٩٥,٠
١٩٩٤	٨٦٩٣,٠٠	٣٣٨٧,٦٠	٧,١٨	٤٧٤٤,٠٠	٤٣٦٩,٠
١٩٩٥	٨٩٦٠,٠٠	٣٧٢٩,٧٠	٧,١٨	٤٨١٢,٠٠	٤٩٤٦,٠

١- سجل البيانات فى ملف وليكن Corr.sav

٢- من قائمة Analyze اختر أمر Correlate ، ومنها اختر Bivariate يظهر الصندوق الحوارى التالى :



٣- اختر المتغيرات المراد إيجاد معاملات الارتباط بينها ولتكن x4 , x3 , x2 , x1 , y .

٤- إذا كانت المتغيرات كمية اختر Person من خانة معاملات الارتباط ، أما إذا كانت البيانات ترتيبية اختر Kendall's tau-b , spearman

٥- اختر أمر OK تظهر النتائج التالية :

Correlations

Correlations

		Y	X1	X2	X3	X4
Y	Pearson Correlation	1.000	.943**	.898**	.971**	.968**
	Sig. (2-tailed)	.	.000	.000	.000	.000
	N	21	21	21	21	21
X1	Pearson Correlation	.943**	1.000	.925**	.961**	.986**
	Sig. (2-tailed)	.000	.	.000	.000	.000
	N	21	21	21	21	21
X2	Pearson Correlation	.898**	.925**	1.000	.917**	.936**
	Sig. (2-tailed)	.000	.000	.	.000	.000
	N	21	21	21	21	21
X3	Pearson Correlation	.971**	.961**	.917**	1.000	.984**
	Sig. (2-tailed)	.000	.000	.000	.	.000
	N	21	21	21	21	21
X4	Pearson Correlation	.968**	.986**	.936**	.984**	1.000
	Sig. (2-tailed)	.000	.000	.000	.000	.
	N	21	21	21	21	21

**. Correlation is significant at the 0.01 level (2-tailed).

نلاحظ ما يلى :

- مصفوفة الارتباط مصفوفة متماثلة ، حيث أن القيم على القطر = ١ ، حيث أنها تمثل (ارتباط المتغير بنفسه) ، والقيم أعلى القطر = القيم اسفل القطر ، حيث أن ارتباط المتغير x1 بالمتغير x2 هو نفسه ارتباط المتغير x2 بالمتغير x1 .
- يوجد بكل خلية من خلايا الجدول ثلاثة قيم ، القيمة الأولى تمثل قيمة معامل الارتباط (Person Correlation) ، القيمة الثانية تمثل معنوية معامل الارتباط Sig. (2-tailed) ، إذا كانت هذه القيمة اقل من ٠,٠٥ كان الارتباط معنوياً ، أما إذا كانت أكبر من ٠,٠٥ كان الارتباط غير معنوى ، القيمة الثالثة تمثل عدد القيم (N) .
- يظهر البرنامج الرمز (*) بجوار معامل الارتباط إذا كان معنوياً عند ٠,٠٥ ، إما إذا كان معنوياً عند ٠,٠١ فإنه يظهر الرمز (**)

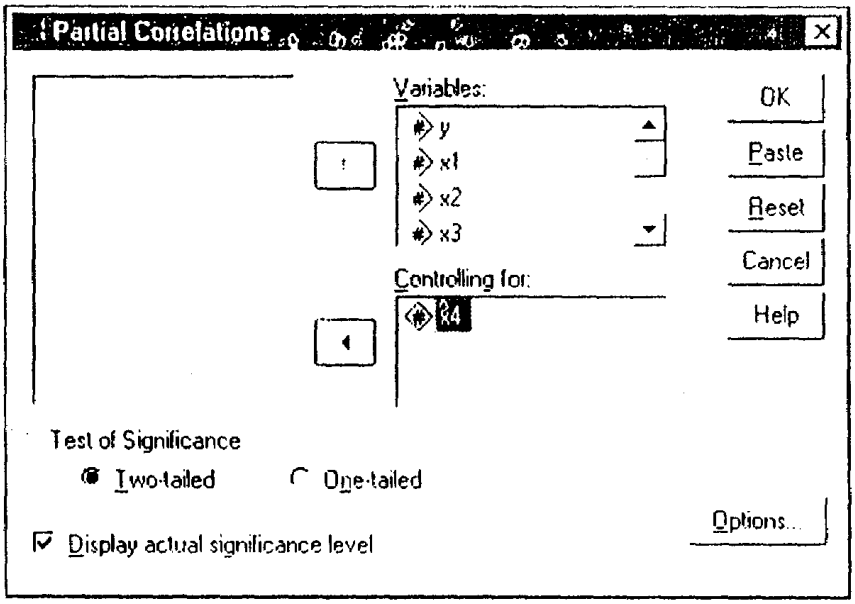
الارتباط الجزئى Partial Correlation :

يقيس معامل الارتباط الجزئى العلاقة بين متغيرين مع وجود متغير ثالث يؤثر عليهما ، ولإيجاد معامل الارتباط الجزئى (أى بعد عزل تأثير المتغير المؤثر على كل من المتغيرين) نتبع الآتى :

١- افتح الملف السابق corr.sav .

٢- من قائمة Analyze اختر أمر Correlate ، ومنها اختر امر Partial يظهر الصندوق الحوارى التالى :

٢- أمام خانة variables اختر المتغيرات المراد ايجاد معامل الارتباط الجزئى لها ، ولتكن y, x1, x2, x3



٣- أمام خانة controlling for اختر المتغيرات التى تؤثر على المتغيرات المرتبطة ، وليكن X4

٤- اختر أمر OK تظهر النتائج التالية :

Partial Corr

```
- - - P A R T I A L   C O R R E L A T I O N   C O E F F I C I E N T S - -
-

Controlling for..    X4

                    Y          X1          X2          X3

Y             1.0000      -.3054      -.0959       .4090
              (     0)    (    18)    (    18)    (    18)
              P= .        P= .190     P= .688     P= .073

X1            -.3054      1.0000       .0312      -.3451
              (    18)    (     0)    (    18)    (    18)
              P= .190     P= .        P= .896     P= .136

X2            -.0959       .0312      1.0000      -.0787
              (    18)    (    18)    (     0)    (    18)
```

```
                P= .688     P= .896     P= .        P= .741

X3               .4090       -.3451      -.0787      1.0000
                (   18)     (   18)     (   18)     (    0)
                P= .073     P= .136     P= .741     P=

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

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

نلاحظ من النتائج ما يلى :

- النتائج السابقة تمثل مصفوفة الارتباط بين المتغيرات x3 , x2 , x1 , y بعد عزل تأثير المستغير x4 ، كما يشير الرقم بين القوسين إلى درجات الحرية (عدد المشاهدات – ٣) ، وتشير الـ P إلى مستوى المعنوية .
- تتميز بنفس خصائص مصفوفة الارتباط السابق التعرف عليها من حيث أن معاملات الارتباط على القطر تساوى الواحد الصحيح ، والمعاملات أعلى القطر تساوى المعاملات أسفل القطر .

٤-٢ الانحدار الخطى البسيط

يمثل تقدير العلاقة بين متغيرين أهمية كبيرة فى مجال العلوم الاجتماعية وخاصة المجال الاقتصادي حيث تظهر الحاجة إلى قياس كمى بين بعض المتغيرات كالعلاقة بين الدخل الممكن التصرف فيه والاستهلاك مثلا وذلك لتحديد شكل العلاقة ومدى قوتها وتقدير المعاملات الخاصة بها وقياس درجة معنوياتها الإحصائية .لذلك فان تحليل الانحدار يستخدم لاختبار العلاقة بين متغيرين أحدهم تابع أو مفسر وليكن (y) والآخر مستقل أو مفسر (x) ، كما يمكن استخدامه فى التنبؤ بقيم المتغير التابع عند حدوث تغير فى المتغير المستقل،

ولتحديد شكل العلاقة بين المتغيرين المعنيين غالبا ما نبدأ بالتعرف على شكل الانتشار للقيم الخاصة بهم وما إذا كان يتقارب مع معادلة الخط المستقيم أو بتعبير رياضي على الصورة .

$$y_i = b_o + b_1 x_i$$

وحيث أنه فى العلوم الاجتماعية من الصعب أن تقع قيم y بالدقة على الخط المستقيم لذلك فان الشكل السابق يعدل ليحتوى على الأخطاء العشوائية والذي يجعل المعادلة السابقة على الصورة

$$y_i = b_o + b_1 x_i + u_i$$

حيث يلاحظ أن الأخطاء العشوائية u_i تعبر عن الجزء غير المنتظم فى العلاقة نتيجة السلوك البشرى ، كما يمكن أن يعبر عن الأخطاء الناتجة عن حذف بعض المتغيرات فى المعادلة ، أو أخطاء القياس والتجميع فى العلاقة ، كما قد يرجع الى تبسيط العلاقة واختزالها على صورة خط مستقيم.

ويخضع هذا المتغير العشوائى لمجموعة من الافتراضات

١. المتغير العشوائى له توزيع إحتمالى اى أنه عند كل نقطة من نقط المتغير x يوجد توزيع اجمالى للقيم المتناظرة للمتغير y وبالتالي للمتغير u وهذه التوزيعات تتركز حول الخط المراد قياسه .

٢. له توزيع طبيعي والوسط الحسابى لقيم الأخطاء يساوى صفر ومعنى هذا أن هناك قيما مختلفة للأخطاء بعضها اكبر من الصفر وبعضها اقل ، لكن متوسطها يساوى صفر

٣. التباين بالنسبة للأخطاء العشوائية حول متوسطها والذي يساوى صفر ثابت عند كل القيم الممكنة للمتغير x بمعنى آخر أن انتشار الأخطاء حول متوسطها ثابت لا يختلف باختلاف القيم المناظرة للمتغير x وليكن σ^2

٤. قيم الأخطاء العشوائية غير مرتبطة فيما بينها، أى أن القيمة التى يأخذها المتغير u عند مستوى معين للمتغير x لا تتأثر بالقيمة التى يأخذها عند مستوى آخر للمتغير x

٥. المتغير المستقل x غير مرتبط بالأخطاء العشوائية، أى أنه ليس هناك علاقة بين تغير x وتغير الأخطاء العشوائية.

ولتوفيق أفضل خط مستقيم لأزواج القراءات الخاصة بالمتغيرين y و x يمكن استخدام طريقة المربعات الصغرى حيث أنها تتضمن تقليل مجموع مربعات الانحرافات عن خط الانحدار أو بصيغة رياضية

$$\text{Min} \quad \Sigma \left(y_i - \hat{y}_i \right)^2$$

حيث y_i هى لقيم الحقيقية ، أما لقيم $\hat{y}_i$ هى القيم على خط الانحدار أو بمعنى آخر فان

$$y_i - \hat{y}_i = e_i$$

حيث e_i تقدير لقيم الأخطاء العشوائية

ومجموع مربعات الأخطاء اقل ما يمكن تعنى

$$\text{Min} \quad \Sigma e_i^2 \quad = \text{Min} \, \Sigma \left(y_i - \hat{y} \right)^2$$

أى أن

$$\Sigma e_i^2 \quad = \Sigma \left(y_i - \hat{y}_i \right)^2 \quad = \Sigma \left(y_i - b_o - b_1 x \right)^2$$

وعلى ذلك فإننا نبحث عن قيم b_0 , b_1 التى تجعل هذه البواقي اقل ما يمكن.

ويمكن استخراج هذه القيم بالتفاضل جزئيا بالنسبة لـ b_0 , b_1

تنتج المعادلات الأساسية التالية

$$\Sigma y_i \quad = \quad n b_o \quad + \quad b_1 \Sigma x$$

$$\Sigma x_i y_i = b_0 \Sigma x + b_1 \Sigma x^2$$

حيث n عدد القراءات

من هذه المعادلات يمكن استنتاج كل من قيمة b_0 , b_1 حيث

$$b_0 = \bar{y} - b_1 \bar{x}$$

حيث $\bar{y} = (\Sigma y)/n$, $\bar{x} = (\Sigma x)/n$, وتمثل الأوساط الحسابية لكل من y,x

$$\hat{b}_1 = \frac{\Sigma xy - n\bar{x}\bar{y}}{\Sigma x_i^2 - n\bar{x}^2} = \frac{\Sigma (x - \bar{x})(y - \bar{y})}{\Sigma (x - \bar{x})^2}$$

حيث أن القيم $(x - \bar{x})$, $(y - \bar{y})$ هى قيم المتغيرات مقاسة من أوساطها الحسابية.

بتقدير قيم b_0 , b_1 يكون قد تم تقدير معالم معادلة الانحدار ويتبقى ضرورة تقدير معنوية معادلة الانحدار ومعالمها.

وتتميز طريقة المربعات الصغرى بأنها تقديرات خطية غير متحيزة وكفأة أى لها أصغر تباين ، كافية

➢ تقدير معنوية معادلة الانحدار.

لتقدير معنوية هذه المعادلة يتطلب تكوين جدول تحليل التباين والمتعارف عليه باسم ANOVA لعلاقة الانحدار وبالتالى اختبار القوة التفسيرية لهذا الخط.

يلاحظ أن يمكن تقسيم الفرق بين القيمة الحقيقية y ووسطها الحسابى $\bar{y}$ إلى جزئين.

١- جزء يرجع إلى الأخطاء العشوائية وهو يمثل الفرق بين القيم الحقيقية والقيم المقدرة $(y-\hat{y})$

٢- جزء آخر يرجع إلى اختلاف القيم المقدرة عن وسطها $(\hat{y} - \bar{y})$ وعلى ذلك فانه يمكن كتابة (*)

$$\Sigma (y - \bar{y})^2 = \Sigma (y-\hat{y})^2 + \Sigma (\hat{y} - \bar{y})^2$$

$$SST = \Sigma e^2 + SSR$$

$$SST = SSE + SSR$$

ويكون جدول تحليل التباين كما يلى.

(*) هناك إثبات رياضي يمكن الرجوع إليه لمن يرغب فى أى من المراجع الإحصائية.

جدول تحليل التباين

فى حالة الانحدار الخطى البسيط

مصدر التغير	درجات الحرية	مجموع المربعات	متوسط مجموع المربعات	$F_{1, n-2}$ المحسوبة
SSR الانحدار	1	$SSR=\Sigma(\hat{y}-\bar{y})^2$	$\frac{SSR}{1} = M_{xR}$	M_{xR}/σ^2_E
البواقى (الانحراف SSE عن خط الانحدار)	n-2	$SSE=\Sigma(y_i-\hat{y})^2$	$\frac{SSE}{n-2} = \sigma^2_E$	
الكلى	n-1	$SST=\Sigma(y_i-\bar{y})^2$		

لاختبار القدرة التفسيرية للعلاقة الخطية تستخدم قيمة F المحسوبة بدرجات حرية البسط 1 والمقام (n-2) وتقارن بقيمتها الجدولية عند مستوى معنوية وليكن α ويمكن عدم رفض فرض العدم (قبول لفرض العدم) وهو أن لا يوجد مقدرة تفسيرية للخط المقدر وذلك إذ كانت قيمة F المحسوبة اقل من القيمة الجدولية.

كذلك فان يمكن حساب الانحراف المعيارى لهذا النموذج

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$$

اختبار معنوية المعالم المقدرة

تم تقدير معالم النموذج ومن ثم من الضرورة اختبار المعنوية الإحصائية لهذه المعالم ، لذلك يجب تقدير تباين هذه المعالم.

كما يمكن استخدام خاصية التوزيع الطبيعى أو توزيع t حسب حجم العينة المستخدم فى التقدير ففى حالة العينات الصغيرة أى عدد القراءات اقل من ٣٠ مفردة يستخدم توزيع t

➢ تباين المعالم المقدرة

لو فرضنا أن $S^2_{b_1}$ تباين المعلمة $\hat{b}_1$ أى ميل خط الانحدار فإنها تساوى

$$S^2_{b_1} = \frac{\hat{\sigma}_{E2}}{\Sigma(x-\bar{x})^2} = \frac{\hat{\sigma}_{E2}}{\Sigma x^2 - n\bar{x}^2}$$

ذلك لأن

$$\hat{b}_1 = \frac{\Sigma(x - \bar{x}) (y - \bar{y})}{\Sigma(\bar{x} - x)^2}$$

ومنها نجد

$$\hat{b}_1 = \frac{\Sigma(x - \bar{x})y}{\Sigma(x - \bar{x})^2} - \frac{\Sigma(x - \bar{x}) y}{\Sigma(x - \bar{x})^2}$$

$$\hat{b}_1 = \frac{\Sigma(x- \bar{x}) (b_0 + b_1 x+u)}{\Sigma(x - \bar{x})^2}$$

ويكون تباين $\hat{b}_1$ مساويا لـــــ

$$\hat{v} (b_1) = \frac{v (u) \Sigma(x^2)}{n\Sigma (x - \bar{x})^2}$$

وتستخدم القيمة $\hat{\sigma}_E^2$ السابقة بيانها بالجدول كتقدير لتباين u

كذلك فان تقدير تباين $\hat{b}_0$ يكون

$$\hat{v}(b_0) = v (u) \frac{\Sigma x^2}{n\Sigma(x - \bar{x})^2}$$

باستخراج قيم التباينات يمكن إجراء الاختبار الاحصائى اللازم عن قيم هذه المعلمات فتكون :

نفرض أن b_1 = صفر أى أن الفرض العدمى

$$H_0 \Rightarrow b_1 = o$$

وأن الفرض البديل $H_1 \Rightarrow b_1 \neq o$

وتكون

$$t_{\alpha,(n-2)} = \frac{\hat{b}_1 - b_1}{\sqrt{\hat{v}(b_1)}}$$

وبنفس الطريقة يمكن اختبار المعنوية الإحصائية للمعلمة $\hat{b}_0$ حيث تكون

$$t_{\alpha,n-2} = \frac{b_0 - \hat{b}_0}{\sqrt{\hat{v}(b_0)}}$$

وذلك على اعتبار أن

$$H_o \Rightarrow b_o = o \, , H_1 \Rightarrow b_o \neq o$$

ولاحظ أن الغرض من تحليل الانحدار هو قبول الفرض البديل أى أن كل من b_o, b_1 لايساوى الصفر ، وفى هذا النطاق نستخدم اختبار الذيلين.

معامل التحديد R^2

معامل التحديد عبارة عن النسبة بين مجموع مربعات الانحدار ومجموع المربعات الكلى ويستخدم لمعرفة القوة التفسيرية للنموذج المقدر أى نسبة التغير الكلى للمتغير y بسبب علاقة الانحدار بين x, y ومن الجدول قد لاحظنا أن

$$SST = SSR + SSE$$

$$1 = \frac{SSR}{SST} + \frac{SSE}{SST}$$

$$= \frac{\Sigma(y - \hat{y})^2}{\Sigma(y - \bar{y})^2} + \frac{\Sigma e_i^2}{\Sigma(y - \bar{y})^2}$$

$$1 = \frac{\Sigma e_i^2}{\Sigma(y - \bar{y})^2} = R^2$$

من هذا نجد أن قيمة R^2 تقع بين صفر ، 1 أى أن $o \leq R^2 \leq 1$

R^2 = صفر عندما تكون الأخطاء كبيرة وتساوى مجموع المربعات الكلية أى عندما تقع نقط العينة على خط مستقيم $y = \bar{y}$. أو على دائرة ولا توجد علاقة بين المتغيرين x , y

$R^2 = 1$ عندما تقع كل نقط العينة على خط الانحدار المقدر .

الانحدار الخطى البسيط Linear Regression باستخدام البرنامج الاحصائى SPSS

يمكن توضيح ذلك من خلال المثال التالى :

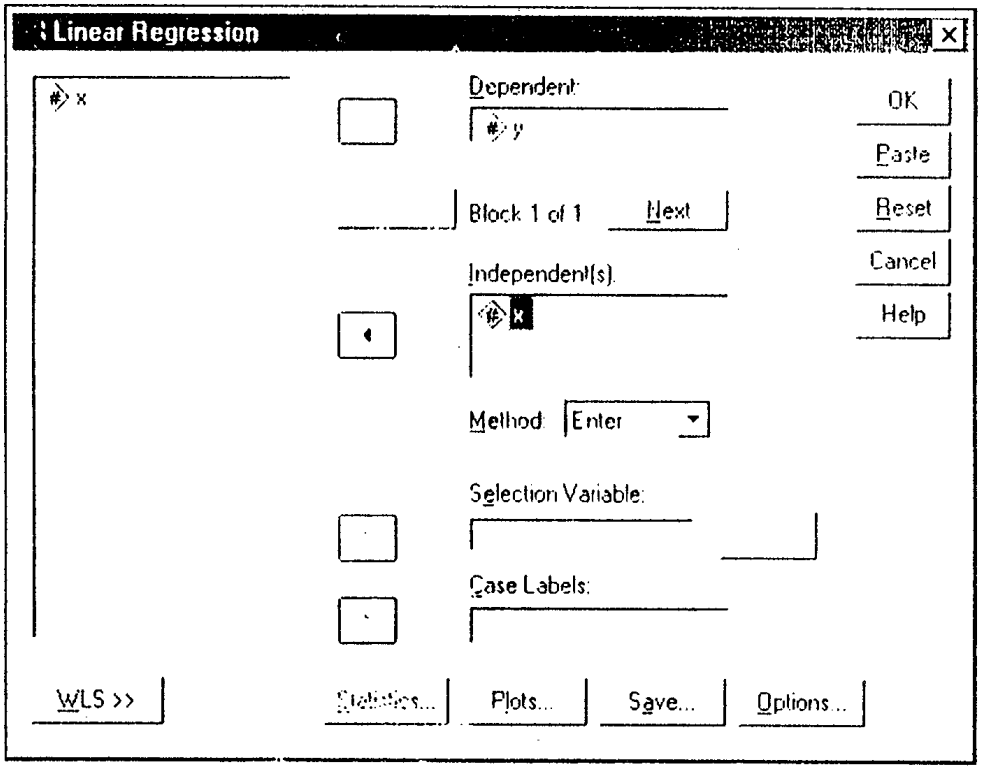
مثال : فيما يلى بيانات الناتج المحلى الزراعى الثابت (y) ، الاستثمارات الزراعية الثابتة (x) ، والمطلوب إيجاد نموذج للعلاقة الخطية بين المتغيرين

السنوات	الناتج المحلى الزراعى الثابت	الاستثمارات الزراعية
١٩٧٥	٩٩٦,٩٠	٨٣,٤٠
١٩٧٦	١٠٦٣,٧٠	٩٨,٠٠
١٩٧٧	١٦١٧,٨٠	١١٤,٥٠
١٩٧٨	١٥١٥,٧٠	١٤٤,١٠
١٩٧٩	١٦٩٠,٨٠	١٨٧,٢٠
١٩٨٠	١٧٠٧,٢٠	٢٠٤,٤٠
١٩٨١	١٨٨٢,١٠	٤٢٣,٦٠
١٩٨٢	٣٤٥٥,٦٠	٤١٦,٣٠
١٩٨٣	٤٢٠٤,٩٠	٤٠٩,٨٠
١٩٨٤	٤٠٦١,١٠	٤٥٢,٨٠
١٩٨٥	٤٠١٨,٧٠	٤٩٥,٨٠
١٩٨٦	٤٥٤٠,٠٠	٧٢٣,٥٠
١٩٨٧	٥٦١٠,٠٠	١٥١٥,٢٠
١٩٨٨	٥٧٨١,٠٠	١٦٦٧,٩٠
١٩٨٩	٥٩٣٥,٠٠	١٨٢١,٣٠
١٩٩٠	٦١٣٠,٠٠	١٩٠٥,٢٠
١٩٩١	٧٩٣٩,٠٠	١٩٩٥,٥٠
١٩٩٢	٨١٣٦,٠٠	٢٢٨٨,٥٠
١٩٩٣	٨٤٤٨,٠٠	٢٧٠٦,١٠
١٩٩٤	٨٦٩٣,٠٠	٣٣٨٧,٦٠
١٩٩٥	٨٩٦٠,٠٠	٣٧٢٩,٧٠

ولإيجاد ذلك النموذج الخطى نتبع الآتى :

١- من قائمة Analyze اختر أمر Regression ، ومنها اختر أمر Linear

يظهر الصندوق الحوارى التالى :

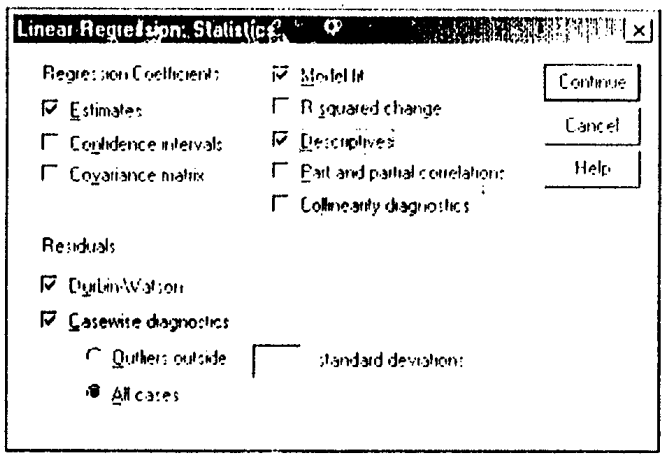


٢- أمام خانة **Dependent** اختر المتغير التابع **Y**

٣- أمام خانة **Independent(s)** اختر المتغير المستقل **X1**

٤- أمام خانة **Method** اختر **Enter** .

٥- يمكن اختيار **Statistics** لإظهار بعض المقاييس الإحصائية الأخرى كالتالى :



٦- اختر أمر **OK** تظهر النتائج التالية :

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
Y	4589.8333	2747.4347	21
X	1179.5429	1147.9951	21

Correlations

		Y	X
Pearson Correlation	Y	1.000	.943
	X	.943	1.000
Sig. (1-tailed)	Y	.	.000
	X	.000	.
N	Y	21	21
	X	21	21

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	X[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Y

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.943[a]	.888	.883	941.3460	.552

a. Predictors: (Constant), X

b. Dependent Variable: Y

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.3E+08	1	1.3E+08	151.367	.000[a]
	Residual	1.7E+07	19	886132.2		
	Total	1.5E+08	20			

a. Predictors: (Constant), X

b. Dependent Variable: Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1928.961	298.232		6.467	.000
	X	2.256	.183	.943	12.303	.000

a. Dependent Variable: Y

Casewise Diagnostics[a]

Case Number	Std. Residual	Y	Predicted Value	Residual
1	-1.190	996.90	2117.0992	-1120.20
2	-1.154	1063.70	2150.0346	-1086.33
3	-.605	1617.80	2187.2561	-569.4561
4	-.784	1515.70	2254.0293	-738.3293
5	-.702	1690.80	2351.2564	-660.4564
6	-.725	1707.20	2390.0571	-682.8571
7	-1.065	1882.10	2884.5394	-1002.44
8	.624	3455.60	2868.0717	587.5283
9	1.436	4204.90	2853.4087	1351.4913
10	1.180	4061.10	2950.4103	1110.6897
11	1.032	4018.70	3047.4118	971.2882
12	1.040	4540.00	3561.0689	978.9311
13	.279	5610.00	5347.0256	262.9744
14	.095	5781.00	5691.4939	89.5061
15	-.109	5935.00	6037.5414	-102.5414
16	-.103	6130.00	6226.8072	-96.8072
17	1.602	7939.00	6430.5105	1508.4895
18	1.110	8136.00	7091.4746	1044.5254
19	.440	8448.00	8033.5177	414.4823
20	-.933	8693.00	9570.8796	-877.8796
21	-1.469	8960.00	10342.6060	-1382.61

a. Dependent Variable: Y

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	2117.0991	10342.61	4589.8333	2589.7050	21
Residual	-1382.61	1508.4895	6.496E-14	917.5105	21
Std. Predicted Value	-.955	2.221	.000	1.000	21
Std. Residual	-1.469	1.602	.000	.975	21

a. Dependent Variable: Y

نلاحظ من النتائج ما يلى :

يظهر البرنامج معامل الارتباط (R) وهو ٩٤. وهذا يعنى أن هناك ارتباط قوى وطردى بين المتغير التابع (Y) والمتغير المستقل (x1) ، ويحسب (R square) وهو مربع معامل الارتباط ويسمى معامل التحديد وهى نسبة التغيرات فى المتغير التابع التى يشرحها المتغير المستقل .

كما يظهر البرنامج جدول تحليل التباين والذى يختبر معنوية النموذج ككل ، ومنه نجد أن قيمة Signif F قيمة صغيرة جدا ، وهذا يعنى أننا نرفض الفرض العدمى القائل بأن معاملات الانحدار = صفر ، أى أن نموذج الانحدار معنوى .

كما يظهر البرنامج معاملات الانحدار b , a (constant) واختبارات معنوية هذه المعاملات (T) و (T Sig)

كما يظهر بيانات عن القيم الفعلية والقيم المتوقعة من معادلة الانحدار Predicated ، كما يظهر بيانات الفرق بين القيم الفعلية والمتوقعة Residual

توفيق المنحنيات Curve Fitting :

فى حالة العلاقة بين متغيرين قد لا نستطيع أن نحدد شكل العلاقة بين المتغيرين (Linear,Cubic,Power,....etc) ونريد معرفة أفضل نموذج لتمثيل هذه العلاقة ، لذلك يتيح البرنامج عدة أشكال للعلاقة بين المتغيرين ، ويعطى لنا الفرصة لاختيار أفضل النماذج ، وهذه النماذج هى :

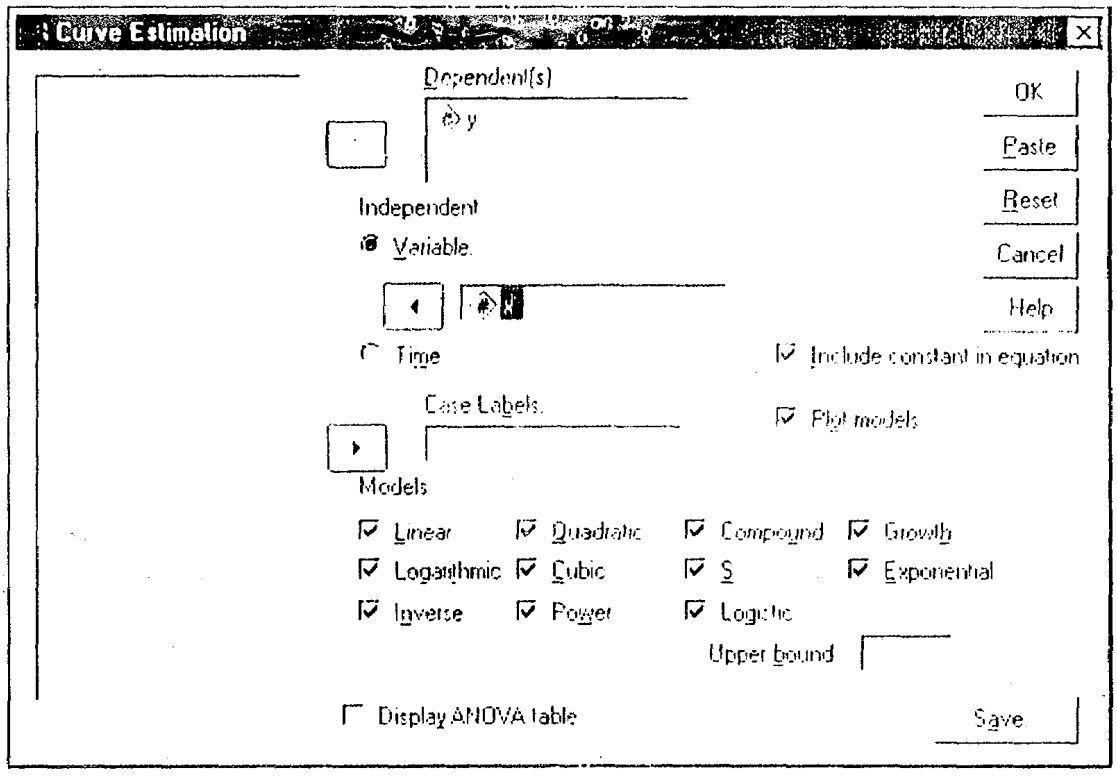
النموذج	شكل العلاقة
Linear	$Y = b_0 + b_1 * X$
Logarithmic	$Y = b_0 + b_1 * ln(X)$
Inverse	$Y = b_0 + b_1 / X$
Quadratic	$Y = b_0 + b_1 * X + b_2 * X**2$
Cubic	$Y = b_0 + b_1 * X + b_2 * X**2 + b_3 * X**3$
Power	$Y = b_0 * X**(b1)$
Compound	$Y = b_0 * b_1**X$
S	$Y = e**(b_0 + b_1 / X)$

النموذج	شكل العلاقة
Logistic	$Y = 1/(1/u + b_0 * b_1 **X)$
Growth	$Y = e**(b_0 + b_1 * X)$
Exponential	$Y = b_0 + e**(b_1 * X)$

يمكن أن تكون العلاقة بين متغير تابع (y) ومتغير مستقل (x) ، ويمكن أيضاً أن تكون بيانات المتغير التابع (y) عبارة عن سلسلة زمنية ونريد إيجاد علاقة بين المتغير (y) والزمن (t)

فى المثال السابق نريد توفيق أفضل نموذج للعلاقة بين الناتج المحلى الزراعى الثابت (y) والاستثمارات الزراعية الثابتة (x1) ، ولإيجاد ذلك نتبع الآتى :

١- من قائمة **Statistics** اختر أمر **Regression** ومنها اختر أمر **Curve Estimation** يظهر



الصندوق الحوارى التالى :

٢- أمام خانة **Dependent(s)** اختر المتغير التابع (y)

٣- أمام خانة **Independent** يوجد اختياران :

_ **Variable** فى حالة ما إذا كانت العلاقة بين المتغير التابع (y) ومتغير آخر (x) .

Time _ فى حالة ما إذا كان المتغير التابع (y) عبارة عن سلسلة زمنية ، ونريد إيجاد علاقة بينه وبين الزمن T .

٤- إذا أردنا ظهور الثابت (a) فى المعادلة فإننا ننشط الاختيار **Include constant in equation** أما إذا أهملنا هذه الخانة فإن خط الانحدار يمر بنقطة الأصل (**origin**) .

٥- اختر النماذج المطلوبة من خانة **Models** ، وأشكال هذه النماذج موضحة بالجدول السابق .

٦- إذا أردنا عرض جدول تحليل التباين نشط الاختيار **Display ANOVA table**

٧- اختر أمر **OK** تظهر النتائج التالية :

```
Curve Fit
MODEL:  MOD_1.
_


Independent:  X
```

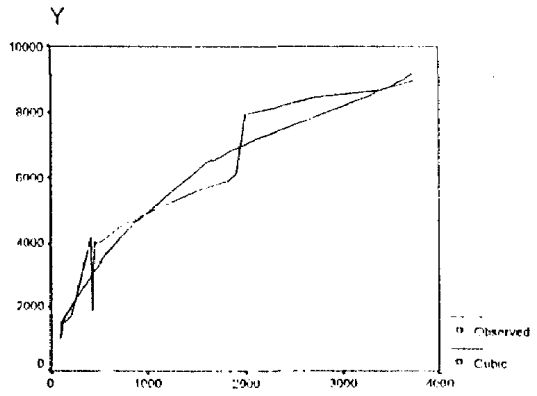
Dependent	Mth	Rsq	d.f.	F	Sigf	Upper bound	b0	b1	b2	b3
Y	LIN	.888	19	151.37	.000		1928.96	2.2559		
Y	LOG	.927	19	240.93	.000		-9080.8	2116.11		
Y	INV	.657	19	36.34	.000		6530.27	-625782		
Y	QUA	.935	18	128.99	.000		1291.28	3.9497	-.0005	
Y	CUB	.940	17	88.16	.000		940.267	5.6007	-.0017	2.1E-07
Y	COM	.721	19	49.18	.000		1940.86	1.0005		
Y	POW	.934	19	268.05	.000		94.5684	.5667		
Y	S	.835	19	96.48	.000		8.7946	-188.35		
Y	GRO	.721	19	49.18	.000		7.5709	.0005		
Y	EXP	.721	19	49.18	.000		1940.86	.0005		
Y	LGS	.721	19	49.18	.000		.0005	.9995		

واضح من البيانات أن أفضل نموذج لتمثيل البيانات هو النموذج **QUB** حيث أن لـه أعلـى **Rsq** أى مربع معامل الارتباط **0.940**

بعد ذلك يتم اختيار ذلك النموذج ، وتحديد معاملاته على النحو التالى :

```
Curve Fit
MODEL:  MOD_2.
_


Dependent variable.. Y                    Method.. CUBIC

Listwise Deletion of Missing Data
```

```
Multiple R              .96933
R Square                .93961
Adjusted R Square       .92895
Standard Error        732.34120

            Analysis of Variance:

                 DF    Sum of Squares       Mean Square

Regression        3      141850449.4        47283483.1
Residuals        17        9117501.9          536323.6

F =       88.16222       Signif F =  .0000

---------------------- Variables in the Equation ----------------------

Variable                     B         SE B       Beta          T  Sig T

X                     5.600722     1.498362   2.340220      3.738  .0016
X**2                  -.001700      .001027  -2.446629     -1.655  .1162
X**3          2.12499876E-07   1.8223E-07   1.068165      1.166  .2597
(Constant)       940.266994   419.118963                    2.243  .0385

The following new variables are being created:

  Name          Label

  FIT_1         Fit for Y with X from CURVEFIT, MOD_2 CUBIC
  ERR_1         Error for Y with X from CURVEFIT, MOD_2 CUBIC
```



يتم تمثيل القيم المتوقعة **Predicated** والبواقى **Residual** فى ملـف البيانـات باسـم **FITT_1, ERR_1** على النحو التالى

Y	X1	FITT_1	ERR_1
٩٩٦,٩٠	٨٣,٤٠	١٣٩٥,٦٧	٣٩٨,٧٧-
١٠٦٣,٧٠	٩٨,٠٠	١٤٧٣,٠١	٤٠٩,٣١-
١٦١٧,٨٠	١١٤,٥٠	١٥٥٩,٥٨	٥٨,٢٢
١٥١٥,٧٠	١٤٤,١٠	١٧١٢,٦٧	١٩٦,٩٧-

Y	X1	FITT_1	ERR_1
١٦٩٠,٨٠	١٨٧,٢٠	١٩٣٠,٥٤	٢٣٩,٧٤-
١٧٠٧,٢٠	٢٠٤,٤٠	٢٠١٥,٨٤	٣٠٨,٦٤-
١٨٨٢,١٠	٤٢٣,٦٠	٣٠٢٣,٨٤	١١٤١,٧٤-
٣٤٥٥,٦٠	٤١٦,٣٠	٢٩٩٢,٥٦	٤٦٣,٠٤
٤٢٠٤,٩٠	٤٠٩,٨٠	٢٩٦٤,٥٧	١٢٤٠,٣٣
٤٠٦١,١٠	٤٥٢,٨٠	٣١٤٧,٤٥	٩١٣,٦٥
٤٠١٨,٧٠	٤٩٥,٨٠	٣٣٢٥.١١	٦٩٣,٥٩
٤٥٤٠,٠٠	٧٢٣,٥٠	٤١٨٢,٩٩	٣٥٧,٠١
٥٦١٠,٠٠	١٥١٥,٢٠	٦٢٦٢,٧٦	٦٥٢,٧٦-
٥٧٨١,٠٠	١٦٦٧,٩٠	٦٥٣٨,٤٥	٧٥٧,٤٥-
٥٩٣٥,٠٠	١٨٢١,٣٠	٦٧٨٥,٥٢	٨٥٠,٥٢-
٦١٣٠,٠٠	١٩٠٥,٢٠	٦٩٠٩,٦٣	٧٧٩,٦٣-
٧٩٣٩,٠٠	١٩٩٥,٥٠	٧٠٣٥,٥٨	٩٠٣,٤٢
٨١٣٦,٠٠	٢٢٨٨,٥٠	٧٤٠١,٠٧	٧٣٤,٩٣
٨٤٤٨,٠٠	٢٧٠٦,١٠	٧٨٥٨,٣٠	٥٨٩,٧٠
٨٦٩٣,٠٠	٣٣٨٧,٦٠	٨٦٦٥,٢٩	٢٧,٧١
٨٩٦٠,٠٠	٣٧٢٩,٧٠	٩٢٠٦,٠٦	٢٤٦,٠٦-

٤-٣ الانحدار الخطى المتعدد

يوضح الانحدار الخطى المتعدد العلاقة بين متغير واحد تابع وأكثر من متغير مستقل، وهناك العديد من الأمثلة فى هذا الصدد كالعلاقة بين الكمية المطلوبة من سلعة معينة وتأثرها بسعر السلعة وأسعار السلع الأخرى والدخل كمتغيرات مستقلة ، والعلاقة الخطية تعنى أن التغير فى أحد المتغيرات المستقلة بوحدة واحدة يؤدى إلى تغير ثابت بالمتغير التابع. كما ذكر سابقا فإن نموذج الانحدار الخطى فى تقدير العلاقات وخاصة الاجتماعية يتطلب إدخال ما يسمى بالخطأ العشوائي حيث أن الأفراد لا يتصرفون بنفس الطريقة أو أن تفضيلاتهم متماثلة. وبذلك يمكن كتابة النموذج فى هذه الحالة على الصورة.

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + ...+ u_i$$

حيث u_i تشير إلى الأخطاء العشوائية ويخضع لنفس الافتراضات السابق ذكرها بالانحدار الخطى البسيط.

لتقدير معالم الدالة السابقة... b_0 , b_1 , b_2 فيمكن تطبيق طريقة المربعات الصغرى العادية إلا أنه يلاحظ إضافة شرط هام فى هذه الحالة وهو .

- استقلال المتغيرات $x_1 , x_2 , x_3 \ldots$ عن بعضها البعض أي لا يرتبط أى زوجين منها.

نبحث عن b_1 , b_0 التى تجعل مجموع مربعات الأخطاء أقل ما يمكن

أى

$$\text{Min } \Sigma (y - \hat{y})^2$$

$$\Sigma e_i^2 = \text{Min } \Sigma (y - \hat{y})^2$$

$$= \text{Min } \Sigma(y - b_0 - b_1 x_1 - b_2 x_2 \ \ldots)^2$$

لذلك فيتم التفاضل الجزئى بالنسبة لكل معلمة وينتج لدينا مجموعة المعادلات التالية:-

$$\Sigma y = b_0 n + b_1 \Sigma x_1 + b_2 \Sigma x_2 + b_3 \Sigma x_3 + \ldots)$$

$$\Sigma x_1 y = b_0 \Sigma x_1 + b_1 \Sigma x^2_1 + b_2 \Sigma x_1 x_2 + b_3 \Sigma x_1 x_3 + \ldots)$$

$$\Sigma x_2 y = b_0 \Sigma x_2 + b_1 \Sigma x_2 x_1 + b_2 \Sigma x^2_2 + b_3 \Sigma x_3 x_2 + \ldots)$$

$$\Sigma x_n y = b_0 \Sigma x_n + b_1 \Sigma x_n x_1 + b_2 \Sigma x_n x_2 + b_n \Sigma x^2_n + \ldots)$$

ويمكن حل هذه المجوعة من المعادلات إما بأسلوب المحددات أو المصفوفات وسنستخدم هنا أسلوب المصفوفات حيث يمكن التعبير عنها بالشكل التالى :-

$$M_{xy} = (X' X) B$$

حيث المصفوفة (X X) عبارة عن مصفوفة مربعة تتكون عناصرها من حواصل ضرب القراءات للمتغيرات أو مربعاتها أى أن

$$X X = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ x_{11} & x_{12} & x_{13} & \ldots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \ldots & x_{2n} \\ x_{n1} & x_{n2} & x_{n3} & \ldots & x_{no} \end{bmatrix} \begin{bmatrix} 1 & x_{1i} & x_{21} & \ldots & x_{n1} \\ 1 & x_{12} & x_{22} & & x_{n2} \\ 1 & x_{13} & x_{23} & & \\ & x_{1n} & x_{2n} & & x_{nn} \end{bmatrix} = \begin{bmatrix} n & \Sigma x_{1i} & \Sigma x_{2i} \ldots \Sigma x_{ni} \\ \Sigma x_{1i} & \Sigma x^2_i & \Sigma x_1 x_2 \ \ldots \\ & \Sigma x^2_2 & \\ & & \Sigma x^2_{nn} \end{bmatrix}$$

وبالتالى فإن تقدير المعالم B يكون

$$\hat{B} = (x' x)^{-1} x' y$$

وبالتالى يمكن الحصول على متجه المعالم ، كما يمكن استنتاج تباين المعالم حيث

$$V(\hat{B}) = (x' x)^{-1} \hat{\sigma}^2_E$$

حيث $(x' x)^{-1}$ هو مقلوب المصفوفة $(x' x)$

$\hat{\sigma}^2_E$ هو تقدير تباين الأخطاء ، والذى يمكن من اختبار المعنوية الإحصائية لهذه المعالم حيث يلاحظ أن جميع العناصر على القطر الرئيسى لهذه المصفوفة تمثل تباين التقديرات $\hat{b}_0$, $\hat{b}_1$, $\hat{b}_2$ ..... والعناصر الأخرى على جانب القطر الرئيسى تمثل تغاير هذه التقديرات مع بعضها البعض وهذه البيانات تعطى ضمن نتائج برنامج SPSS

ويمكن كتابة جدول تحليل التباين فى صورته العامة كالتالى:

جدول تحليل التباين

فى حالة الانحدار الخطى المتعدد

مصدر التغير	درجات الحرية	مجموع المربعات	متوسط مجموع المربعات	المحسوبة
الانحدار	k	$SSR=\Sigma(\hat{y}-\bar{y})$	$\frac{SSR}{k} = M_{xR}$	$(M_{xR})\ \hat{\sigma}^2_E$
البواقى (الانحراف عن خط الانحدار)	n-k-1	$SSE=\Sigma(y-\hat{y})^2$	$\frac{SSE}{n-k-1} = \hat{\sigma}^2_E$	
الكلى	n-1	$SST=\Sigma(y-\bar{y})^2$		

حيث تستخدم قيمة F بدرجات حرية k , (n –k-1) لاختبار قوة العلاقة بين المتغيرات المستقلة والمتغير التابع اى المعنوية الكلية للانحدار ، كما يمكن أن يختبر R^2 معامل التحديد وذلك بحساب قيمة F وبدرجات حرية (n-k-1), k حيث

$$F_{k,(n-k-1)} = \frac{R^2 / k}{(n-k-1)/ (1-R^2)}$$

وتقارن F المحسوبة بنظيرتها المجدولة وبدرجات الحرية للبسط K والمقام(n-k-1)ومستوى المعنوية المطلوب ٥% أو ١% . أى بدرجات ثقة ٩٥% ، ٩٩% على التوالى

كما يمكن اختبار المعنوية الإحصائية للمعالم المقدرة باستخدام توزيع t وبدرجات حرية (n-k-1)

فلو فرض أننا نريد اختبار المعنوية الإحصائية للمعلمة b_i

فإننا نفترض فرض العدم أى $H_o : \hat{b_i} = o$

مقابل الفرض البديل أى $H_o : \hat{b_i} \neq o$

$$t_\alpha = \frac{\hat{b_i}}{\sqrt{var(\hat{b_i})}}$$

كما يمكن اختبار الفرض الخاص بالقيمة الحقيقة b_i وإنشاء فترة ثقة لهذه القيمة أى

$$t_\alpha = \frac{\hat{b_i} - b_i}{\sqrt{var(\hat{b_i})}}$$

أى

$$t_\alpha \sqrt{var(\hat{b_i})} + \hat{b_i} > b_i > \hat{b_i} - t_\alpha \sqrt{var(\hat{b_i})}$$

كما يمكن اختبار ما إذا كان زوجين من هذه المعالم يساوى بعضهم البعض فان نفترض :

فرض العـــــــدم $H_o : b_2 = b_3$

مقابل الغرض البديل $\hat{H_1} : b_2 \neq b_3$

وتكون t_α مساوية للمقدار

$$t_\alpha = \frac{\hat{b_2} - \hat{b_3}}{\sqrt{\hat{v}(b_2)+\hat{v}(b_3) -2cov\ (\hat{b_2},\hat{b_3})}}$$

حيث $cov\ (\hat{b_2}, \hat{b_3})$ استخراجه من مصفوفة التباين سالفة الذكر

يلاحظ أن إضافة متغيرات مستقلة من شأنه أن يرفع قيمة المربعات الخاصة بالانحدار وبالتالى فان قيمة R^2 تزيد وإذا اخذ فى الاعتبار نقص عدد درجات الحرية مع إضافة متغيرات إضافية مستقلة فيمكن حساب R^2 المعدلة أو $\bar{R}^2$ من الصيغة

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k}$$

حيث n هى عدد القراءات, k عدد المعالم المقدرة

ويمكن استنتاج عندما

$k = 1$ فإن $(n-1)/(n-k) = 1$ ويكون $\bar{R}^2 = R^2$

عندما

$k > 1$ فإن $(n-1)/(n-k) > 1$ فتكون $\bar{R}^2 < R^2$

وعندما تكون n كبيرة لقيمة معينةلـ(k) فإن المقدار $(n-1)/(n-k)$ يقترب من الواحد ولن تختلف كثيرا $\bar{R}^2$ عن R^2 أما عندما تكون n صغيرة وتكون k كبيرة بالنسبة إلى nفان $\bar{R}^2$ تكون اصغر كثيرا من R^2 وقد تكون $\bar{R}^2$ (سالبة بالرغم من أن $0 \geq R^2 \geq 1$

صور أخرى لبعض دوال تحليل الانحدار

قد تكون العلاقة بين المتغير المستقل والمتغير التابع غير خطية إلا أنه فى بعض الأحوال يمكن تحويل مثل هذه العلاقات إلى علاقات خطية، وبهذا يمكن استخدام أسلوب المربعات الصغرى**OLS** (مع تحقق الشروط) لتقدير معالمها ، ومن هذه الدوال الأمثلة التالية .

1- $y = b_0 \, x^{b1} \, e^u \Rightarrow y^* = b^*_0 + b_1 x^* + u$

حيث $y^* = \ln y$ ، $x^* = \ln x$ ، u الأخطاء العشوائية. وهى دالة لوغاريتمية من الطرفين

2- $\ln y = b_0 + b_1 x + u \Rightarrow y^* = b_0 + b_1 x + u$

وهى دالة نصف لوغاريتمية

3- $y = b_0 + \frac{b_1}{x} + u \Rightarrow y = b_0 + b_1 z + u$

حيث $z = \frac{1}{x}$

4- $y = b_0 + b_1 x + b_2 x^2 + u \Rightarrow y = b_0 + b_1 x + bx^2 + u$

حيث $w = x^2$

الانحدار المتعدد Multiple Regression من خلال البرنامج الاحصائى SPSS

ويمكن توضيح ذلك من خلال المثال التالى :

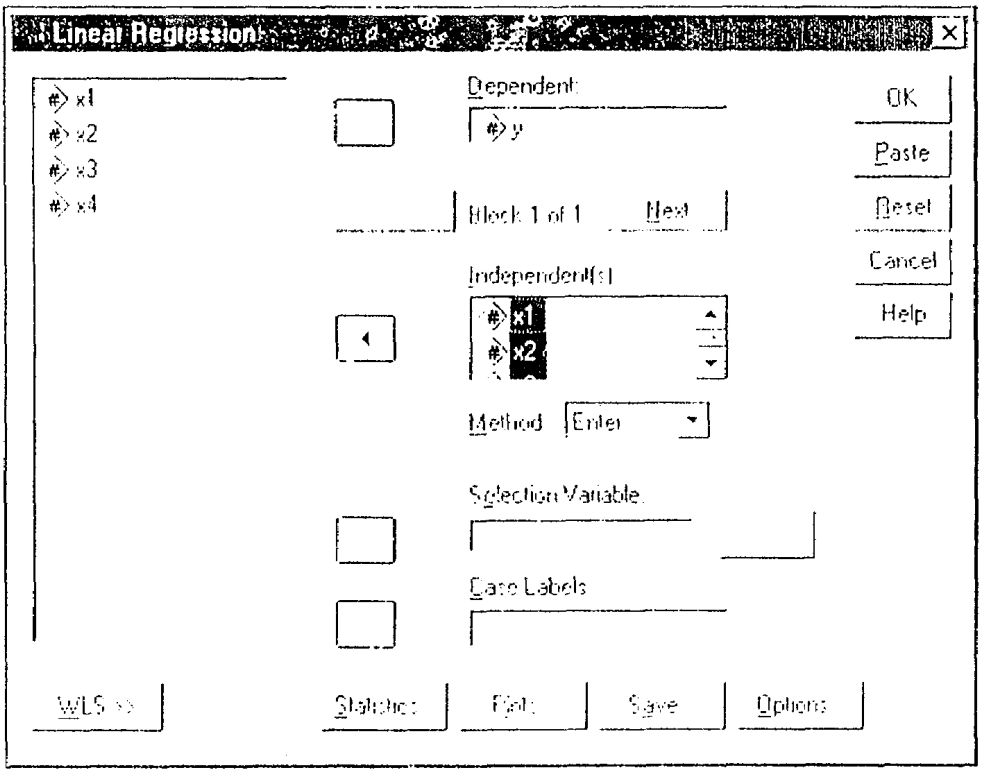
مثال : البيانات التالية تمثل العلاقة بين الناتج المحلى الإجمالى الثابت (y) ، الاستثمارات الزراعية الثابتة (x1) ، المساحة الزراعية (x2) ، العمالة الزراعية بالألف (x3) ، الأجور الزراعية (x4) ، وذلك فى سلسلة زمنية من سنة ١٩٧٥ غلى سنة ١٩٩٥ ، ونريد تحديد شكل للعلاقة بين المتغير التابع (y) والمتغيرات الزراعية (x1,x2,x3,x4) .

السنوات Year	الناتج المحلى الزراعى الثابت y	الاستثمارات الزراعية الثابتة x1	المساحة الزراعية X2	العمالة الزراعية بالألف X3	الأجور الزراعية X4
١٩٧٥	٩٩٦,٩٠	٨٣,٤٠	٥,٧٩	٤٢١٨,٠٠	٢٦٧,٤٠
١٩٧٦	١٠٦٣,٧٠	٩٨,٠٠	٥,٨٦	٤٠٦٨,٠٠	٢٦٢,١٠
١٩٧٧	١٦١٧,٨٠	١١٤,٥٠	٥,٨٥	٤١٠٤,٠٠	٢٨٧,٧٠
١٩٧٨	١٥١٥,٧٠	١٤٤,١٠	٥,٨٦	٤١٣٥,٠٠	٣١٦,٢٠
١٩٧٩	١٦٩٠,٨٠	١٨٧,٢٠	٥,٨٦	٤١٦٥,٠٠	٣٤٧,٨٠
١٩٨٠	١٧٠٧,٢٠	٢٠٤,٤٠	٥,٨٧	٤٢٠٠,٠٠	٤٤٠,٣٠
١٩٨١	١٨٨٢,١٠	٤٢٣,٦٠	٥,٨٨	٤٢٢٥,٠٠	٩٧١,١٠
١٩٨٢	٣٤٥٥,٦٠	٤١٦,٣٠	٥,٩٢	٤٢٦٤,٠٠	١٠٥٥,٦
١٩٨٣	٤٢٠٤,٩٠	٤٠٩,٨٠	٥,٨٤	٤٣٠٤,٠٠	١١٤١,١
١٩٨٤	٤٠٦١,١٠	٤٥٢,٨٠	٥,٨٣	٤٣٤٤,٠٠	١٢٢٨,٩
١٩٨٥	٤٠١٨,٧٠	٤٩٥,٨٠	٥,٩٧	٤٣٨٤,٠٠	١٢٧١,٩
١٩٨٦	٤٥٤٠,٠٠	٧٢٣,٥٠	٥,٩٧	٤٣٣٠,٠٠	١٥٨٦,١
١٩٨٧	٥٦١٠,٠٠	١٥١٥,٢٠	٦,٠٠	٤٣٨١,٠٠	١٨٥٥,٩
١٩٨٨	٥٧٨١,٠٠	١٦٦٧,٩٠	٦,١٨	٤٤٣٢,٠٠	٢١٠٥,٤
١٩٨٩	٥٩٣٥,٠٠	١٨٢١,٣٠	٦,٢٧	٤٤٧٨,٠٠	٢٣٨٢,٤

السنوات Year	الناتج المحلى الزراعى الثابت y	الاستثمارات الزراعيـــة الثابتة x1	المســـاحة الزراعية X2	العمالة الزراعية بالألف X3	الأجـــور الزراعية X4
١٩٩٠	٦١٣٠,٠٠	١٩٠٥,٢٠	٦,٩٢	٤٥٣٣,٠٠	٢٦٩٨,٠
١٩٩١	٧٩٣٩,٠٠	١٩٩٥,٥٠	٧.٠٢	٤٥٨٥,٠٠	٣٠١٢,٠
١٩٩٢	٨١٣٦,٠٠	٢٢٨٨,٥٠	٧,١٢	٤٦٢٤,٠٠	٣٤٢٣,٠
١٩٩٣	٨٤٤٨,٠٠	٢٧٠٦,١٠	٧,١٧	٤٦٨٢,٠٠	٣٨٩٥,٠
١٩٩٤	٨٦٩٣,٠٠	٣٣٨٧,٦٠	٧,١٨	٤٧٤٤,٠٠	٤٣٦٩,٠
١٩٩٥	٨٩٦٠,٠٠	٣٧٢٩,٧٠	٧,١٨	٤٨١٢,٠٠	٤٩٤٦,٠

ولإجراء ذلك نتبع الآتى :

١- من قائمة **Analyze** اختر أمر **Regression** ومنها اختر أمر **Linear** يظهر الصندوق الحـوارى التالى :



٢- اختر المتغير التابع (y) فى خانة **Dependent**

٣- اختر المتغيرات المستقلة (x1,x2,x3,x4) فى خانة **Independent(s)**

٤- اختر Enter من خانة Method ، ثم اختر أمر OK تظهر النتائج التالية :

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	X4[a], X2, X3, X1		Enter

a. All requested variables entered.

b. Dependent Variable: Y

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.975[a]	.950	.938	685.2465	1.024

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

b. Dependent Variable: Y

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.4E+08	4	3.6E+07	76.377	.000[a]
	Residual	7513004	16	469562.7		
	Total	1.5E+08	20			

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

b. Dependent Variable: Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		Sig.
		B	Std. Error	Beta		
1	(Constant)	-22902.9	18529.063		-1.236	.234
	X1	-.677	.865	283	-.783	.445
	X2	-222.844	788.384	045	-.283	.781
	X3	6.138	4.310	.483	1.424	.174
	X4	1.549	1.129	.814	1.372	.189

a. Dependent Variable: Y

Casewise Diagnostics[a]

Case Number	Std. Residual	Y	Predicted Value	Residual
1	-1.544	996.90	2054.6376	-1057.74
2	-.053	1063.70	1100.2525	-36.5525
3	.388	1617.80	1351.9403	265.8597
4	-.071	1515.70	1564.1097	-48.4097
5	-.113	1690.80	1768.0372	-77.2372
6	-.591	1707.20	2112.2900	-405.0900
7	-1.540	1682.10	2737.4455	-1055.35
8	.222	3455.60	3303.7497	151.8503
9	.731	4204.90	3703.9410	500.9590
10	.004	4061.10	4058.6028	2.4972
11	-.426	4018.70	4310.4391	-291.7391
12	.333	4540.00	4311.6402	228.3598
13	1.619	5610.00	4500.2437	1109.7563
14	1.058	5781.00	5056.3427	724.6573
15	.425	5935.00	5643.9338	291.0662
16	-.203	6130.00	6268.7946	-138.7946
17	1.383	7939.00	6990.9968	948.0032
18	.714	8136.00	7646.5125	489.4875
19	.012	8448.00	8439.9776	8.0224
20	-.581	8693.00	8891.4431	-398.4431
21	-1.767	8960.00	10171.1696	-1211.17

a. Dependent Variable: Y

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1100.2526	10171.17	4589.8333	2678.1985	21
Residual	-1211.17	1109.7563	-5.37E-12	612.9031	21
Std. Predicted Value	-1.303	2.084	.000	1.000	21
Std. Residual	-1.767	1.619	.000	.894	21

a. Dependent Variable: Y

يلاحظ من النتائج السابقة عدم معنوية معاملات الانحدار على الرغم من ارتفاع قيمة معامل التحديد فى معادلة الانحدار ، وبرجع ذلك غالباً إلى وجود ارتباط بين المتغيرات المفسرة (المستقلة) مما يعنى عدم ملاءمة طريقة المربعات الصغرى العادية لتقدير تلك المعامل ، وسوف نناقش هذه المشكلة فيما بعد .

الباب الخامس

مشاكل إسقاط الفروض عند تقدير معالم نموذج الانحدار

عرضنا فيما سبق لأسلوب تقدير معلمات نموذج الانحدار الخطى بطريقة المربعات الصغرى العادية وما تنطوى عليه من فروض أساسية إلا أنه إذا لم يتحقق أحد هذه الشروط سوف يؤدى إلى عدة مشاكل، وفيما يلى بعض هذه المشاكل وإمكانية التغلب عليها :-

٥-١ الارتباط الخطى بين المتغيرات المستقلة Multicollinearity

تنشأ مشكلة الارتباط الخطى عندما يكون واحد على الأقل من المتغيرات المستقلة مرتبط بعلاقة خطية مع متغير أو متغيرات أخرى وغالبا ما يكون هذا الارتباط غير تام ، مما يصعب معه تقدير معالم المتغير بمعزل عن المتغيرات الأخرى ويمكن التعرف على مشكلة الارتباط الخطى من خلال :-

١. إرتفاع قيمة معامل التحديد R^2 مصاحب بتقديرات غير معنوية لمعظم المعالم . وعلى الرغم من حساسية ذلك إلا أن لها عدم ميزة (عيب) يتمثل فى أنها شديدة جداً بمعنى أن الارتباط الخطى يعتبر ضار فقط عندما يكون كل تأثيرات المتغيرات المفسرة على المتغير y غير محددة disentangled .

٢. إرتفاع قيمة معامل الارتباط بين أزواج المتغيرات، ٠,٨ فأكثر .

٣. كبر تباينات المعالم المقدرة ومن ثم اتساع فترات الثقة الخاصة بها وهناك عدة طرق لاختبار هذه المشكلـــــة مثل اختبار كلاين Klein test، واختبار فارارجلوبر Farrar Glauber Test ويمكن التغلب على مشكلة وجود الارتباط الخطى للمتغيرات المستقلة بمجموعة من الحلول التى يمكن أن تخفف من تأثير هذا الارتباط :-

١- استخدام عدد أكبر من القراءات أى تكبير حجم العينة لأن الارتباط الخطى مشكلة عينة.

٢- استخدام معلومات مسبقة فى حالة توافرها فمثلاً إذا كنا بصدد تقدير معالم دالة الاستهلاك

$$y = b_0 + b_1 x_1 + b_2 x_2$$

حيث y قيمة الاستهلاك ، x_1 = تمثل الدخل ، x_2 تمثل الثروة ، وإذا كان هناك علاقة بين x_1، x_2 بحيث أن $b_2 = c\, b_1$ حيث c نسبة ثابتة معينة، فى هذه الحالة

٣- يمكن استخدام بيانات مركبة من سلاسل زمنية ومقطعية فلو فرض أننا بصدد تقدير دالة الطلب على سلعة معينة

$$y = b_0 + b_1 x_1 + b_2 x_2$$

حيث y الكمية المباعة، x_1 متوسط السعر، x_2 الدخل وإذا كان هناك إرتباط جوهرى بين x_1 ، x_2 ، فيمكن التغلب على هذه المشكلة عن طريق تقدير العلاقة بين الكمية المباعة والدخل من بيانات مقطعية [على مستوى المحافظات-الهيئات ...الخ] حيث يكون السعر فى هذه الحالة ثابت ثم تستخدم المعلمة الخاصة بالدخل b_2 فى تقدير b_1 من بيانات السلسلة الزمنية.

٤- يمكن إسقاط بعض المتغيرات التفسيرية ذات الارتباط الخطى، خاصة إذا كانت قليلة الأهمية في التأثير على الظاهرة موضوع البحث ، ففي المثال الموجود بالنقطة رقم (٢) يمكن إسقاط الثروة من العلاقة -إلا أن ذلك قد يؤدى فى بعض الأحيان إلى مشاكل خاصة بالنموذج وما قد يتبعه من مشاكل الارتباط الذاتى.

٥- تحويل المتغيرات قد يساعد على التخفيف من مشكلة الارتباط الخطى وذلك باستخدام الفروق الأولى للمتغيرات.

٦- هناك طرق أخرى للتغلب على مشكلة الارتباط الخطى وذلك باستخدام أسلوب تحليل العوامل **Factor analysis** وأسلوب المكونات الرئيسية **Principal Component** وغيرها من الطرق.

٧- برنامج SPSS يمكن من استخدام أسلوب الانحدار المتدرج فى هذا الشأن ، جدير بالذكر أن نموذج الانحدار الخطى الذى يتصف بمشكلة الارتباط الخطى يعتبر مقبولا إذا ماكان الهدف منه التنبؤ.

معالجة الازدواج الخطى باستخدام البرنامج الإحصائى SPSS (الانحدار المتدرج **Stepwise**) :

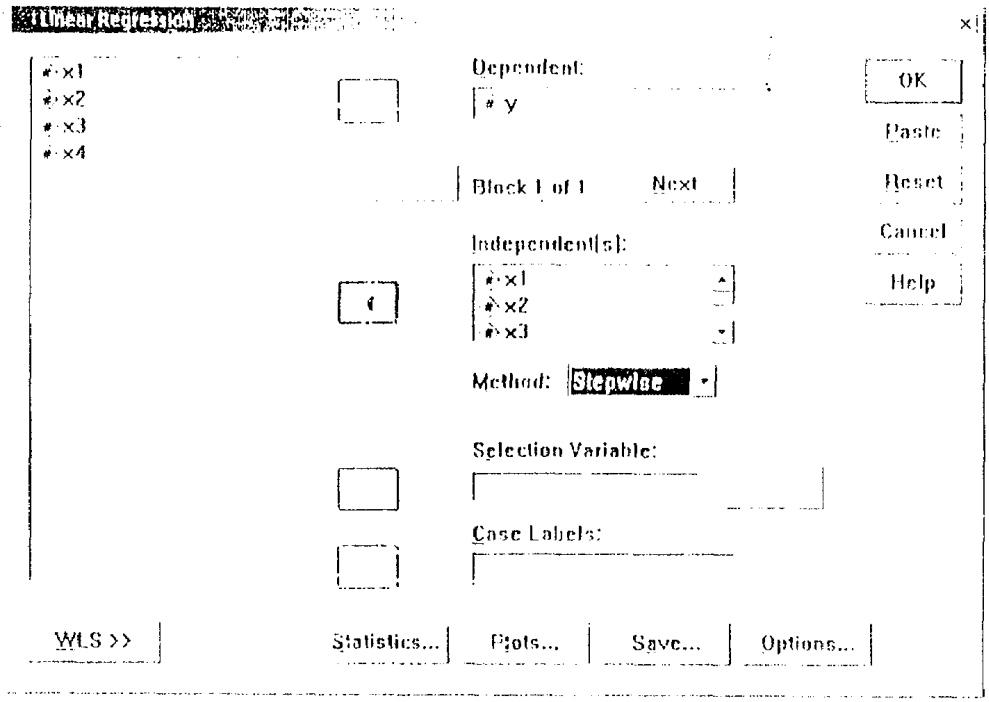
يبنى الانحدار المتعدد على عدة فروض أحدها ألا يكون هناك علاقة بين المتغيرات المستقلة ، ولكن فى الواقع العملى قد يوجد ارتباط بين المتغيرات المستقلة ، ولحل هذه المشكلة نستخدم الانحدار التدريجى **Stepwise** حيث يتم حساب معاملات الارتباط لجميع المتغيرات الداخلة فى النموذج ، ثم يتم اختيار أكثر المتغيرات المستقلة تأثيراً على معامل التحديد وعلى اختبار F ، ثم يضيف متغيراً مستقلاً آخر يلى المتغير الأول فى الارتباط بالمتغير التابع وهكذا مع الأخذ فى الاعتبار عدم وجود علاقة بين المتغيرات المستقلة الداخلة فى النموذج

ولتنفيذ ذلك على تابع الخطوات التالية :

١- سجل البيانات التالية على ملف وليكن step.sav

الأجور الزراعية X4	العمالة الزراعية بالألف X3	المساحة الزراعية X2	الاستثمارات الزراعية الثابتة x1	الناتج المحلى الزراعى الثابت y	السنوات Year
٢٦٧,٤٠	٤٢١٨,٠٠	٥,٧٩	٨٣,٤٠	٩٩٦,٩٠	١٩٧٥
٢٦٢,١٠	٤٠٦٨,٠٠	٥,٨٦	٩٨,٠٠	١٠٦٣,٧٠	١٩٧٦
٢٨٧,٧٠	٤١٠٤,٠٠	٥,٨٥	١١٤,٥٠	١٦١٧,٨٠	١٩٧٧
٣١٦,٢٠	٤١٣٥,٠٠	٥,٨٦	١٤٤,١٠	١٥١٥,٧٠	١٩٧٨
٣٤٧,٨٠	٤١٦٥,٠٠	٥,٨٦	١٨٧,٧٠	١٦٩٠,٨٠	١٩٧٩
٤٤٠,٣٠	٤٢٠٠,٠٠	٥,٨٧	٢٠٤,٤٠	١٧٠٧,٢٠	١٩٨٠
٩٧١,١٠	٤٢٢٥,٠٠	٥,٨٨	٤٢٣,٦٠	١٨٨٢,١٠	١٩٨١
١٠٥٥,٦	٤٢٦٤,٠٠	٥,٩٢	٤١٦,٣٠	٣٤٥٥,٦٠	١٩٨٢
١١٤١,١	٤٣٠٤,٠٠	٥,٨٤	٤٠٩,٨٠	٤٢٠٤,٩٠	١٩٨٣
١٢٢٨,٩	٤٣٤٤,٠٠	٥,٨٣	٤٥٢,٨٠	٤٠٦١,١٠	١٩٨٤
١٢٧١,٩	٤٣٨٤,٠٠	٥,٩٧	٤٩٥,٨٠	٤٠١٨,٧٠	١٩٨٥
١٥٨٦,١	٤٣٣٠,٠٠	٥,٩٧	٧٢٣,٥٠	٤٥٤٠,٠٠	١٩٨٦
١٨٥٥,٩	٤٣٨١,٠٠	٦,٠٠	١٥١٥,٢٠	٥٦١٠,٠٠	١٩٨٧
٢١٠٥,٤	٤٤٣٢,٠٠	٦,١٨	١٦٦٧,٩٠	٥٧٨١,٠٠	١٩٨٨
٢٣٨٢,٤	٤٤٧٨,٠٠	٦,٢٧	١٨٢١,٣٠	٥٩٣٥,٠٠	١٩٨٩
٢٦٩٨,٠	٤٥٣٣,٠٠	٦,٩٢	١٩٠٥,٢٠	٦١٣٠,٠٠	١٩٩٠
٣٠١٢,٠	٤٥٨٥,٠٠	٧,٠٢	١٩٩٥,٥٠	٧٩٣٩,٠٠	١٩٩١
٣٤٢٣,٠	٤٦٢٤,٠٠	٧,١٢	٢٢٨٨,٥٠	٨١٣٦,٠٠	١٩٩٢
٣٨٩٥,٠	٤٦٨٢,٠٠	٧,١٧	٢٧٠٦,١٠	٨٤٤٨,٠٠	١٩٩٣
٤٣٦٩,٠	٤٧٤٤,٠٠	٧,١٨	٣٣٨٧,٦٠	٨٦٩٣,٠٠	١٩٩٤
٤٩٤٦,٠	٤٨١٢,٠٠	٧,١٨	٣٧٢٩,٧٠	٨٩٦٠,٠٠	١٩٩٥

٢- من قائمة Analyze اختر أمر Regression ، ومنه اختر أمر Linear يظهر الصندوق الحوارى Linear Regression



١- أمام خانة **Dependent** اختر المتغير التابع Y

٢- أمام خانة **Independent(s)** اختر المتغيرات المستقلة ، ولتكن **x1 , x2 , x3 , x4**

٣- أمام خانة **Method** اختر **Stepwise** ، ثم اختر أمر **Ok** ، فتظهر النتائج التالية :

Regression

Variables Entered/Removed[b]

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

a. Dependent Variable: Y

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.971[a]	.943	.940	571.7121	1.230

a. Predictors: (Constant), X3

b. Dependent Variable: Y

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.4E+08	1	1.4E+08	315.594	.000[a]
	Residual	8572745	19	451197.1		
	Total	1.5E+08	20			

a. Predictors: (Constant), X3

b. Dependent Variable: Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-49538.3	3050.428		-16.240	.000
	X3	12.354	.695	.971	17.765	.000

a. Dependent Variable: Y

Excluded Variables[b]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	X1	.121[a]	.601	.555	.140	7.641E-02
	X2	.049[a]	.348	.732	.082	.159
	X4	.396[a]	1.298	.211	.293	3.092E-02

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

b. Dependent Variable: Y

Casewise Diagnostics[a]

Case Number	Std. Residual	Y	Predicted Value	Residual
1	-2.341	996.90	2569.7039	-1572.80
2	.517	1063.70	716.6440	347.0560
3	.679	1617.80	1161.3784	456.4216
4	-.043	1515.70	1544.3441	-28.6441
5	-.334	1690.80	1914.9561	-224.1561
6	-.953	1707.20	2347.3367	-640.1367
7	-1.152	1882.10	2656.1800	-774.0800
8	.473	3455.60	3137.9756	317.6244
9	.853	4204.90	3632.1249	572.7751
10	-.097	4061.10	4126.2742	-65.1742
11	-.896	4018.70	4620.4235	-601.7235
12	.873	4540.00	3953.3220	586.6780
13	1.528	5610.00	4583.3623	1026.6377
14	.845	5781.00	5213.4027	567.5973
15	.228	5935.00	5781.6744	153.3256
16	-.493	6130.00	6461.1297	-331.1297
17	1.244	7939.00	7103.5238	835.4762
18	.820	8136.00	7585.3193	550.6807
19	.218	8448.00	8301.8358	146.1642
20	-.558	8693.00	9067.7673	-374.7673
21	-1.411	8960.00	9907.8211	-947.8211

a. Dependent Variable: Y

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	716.6440	9907.8213	4589.8333	2668.2879	21
Residual	-1572.80	1026.6377	-4.31E-12	654.7039	21
Std. Predicted Value	-1.452	1.993	.000	1.000	21
Std. Residual	-2.341	1.528	.000	.975	21

a. Dependent Variable: Y

يلاحظ أن البرنامج اكتفى بالمتغير X3 نظراً لوجود ارتباط خطى بين المتغيرات المستقلة ، مما يعنى أن الناتج المحلى الزراعى الثابت يتأثر بالدرجة الأولى بالعملة الزراعية X3

٥-٢ الارتباط الذاتي للأخطاء Autocorrelation

استقلال الأخطاء عن بعضها من الافتراضات الهامة فى نموذج الانحدار الخطى أى أن

$$E\ (u_i\ \ u_j) = o \qquad\qquad i \neq j$$

وعدم تحقق هذا الشرط يعنى وجود ما يسمى بالارتباط الذاتي للأخطاء أو الارتباط السلسلى Serial correlation ، بمعنى أن قيمة u فى فترة معينة مرتبطة بقيمها فى فترات أخرى ، ويعرف الارتباط الذاتى للأخطاء من الدرجة الأولى بأنه ارتباط الأخطاء بالفترة التى تليها.

هذا وقد يكون الارتباط الذاتي سالبا عندما تعبر الأخطاء المتتالية إشاراتها كثيرا ، ولكن عندما تكون لعدة بواقى متتالية لها نفس الإشارة فيكون الارتباط الذاتي موجباً.

ويؤدى إسقاط فرض استقلال الأخطاء أو ظهور مشكلة الارتباط الذاتي إلى تقليل كفاءة المعالم المقدرة ، فمثلاً إذا وجد أن هناك ارتباط سلسلى موجب فإن الانحرافات المعيارية للمعالم المقدرة فى هذه الحالة تكون أقل من المفروض أن تكون عليه مما يترتب عليه مبالغة فى الدقة وفى المعنوية الإحصائية لهذه المعالم ، كما يؤدى بالتبعية إلى ارتفاع قيمة معامل التحديد R^2 . كما يؤدى هذا الارتباط فى الأخطاء إلى أن التنبؤات المحسوبة بطريقة المربعات الصغرى يكون لها تباين اكبر من التنبؤات التى يمكن الحصول عليها بطريقة أخرى كطريقة المربعات الصغرى العامة .

ويجدر الإشارة فى هذا المقام أن الارتباط الذاتي لا يؤثر على تحيز المعالم المقدرة.

➢ الكشف عن الارتباط الذاتي

يمكن استنتاج وجود ارتباط سلسلى وذلك بتتبع نمط تغير الاتجاه العام للأخطاء المقدرة e_t من معادلة الانحدار ، كما يمكن ايجاد معامل الارتباط بين السلسلتين e_t e_{t-1} بالصيغة التالية

$$Re_t\, e_{t-1} = \frac{\Sigma e_t\, e_{t-1}}{\sqrt{\Sigma e_t^2\ \Sigma e_{t-1}^2}}$$

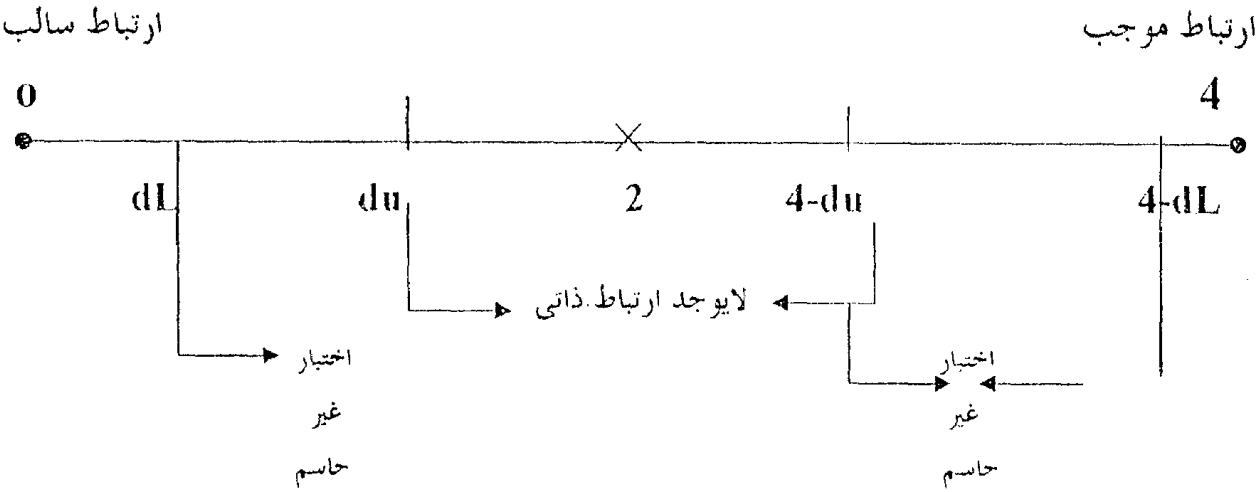
وقد يكون هذه الطرق غير عالية الدقة ومن ثم لابد من التأكد أن عنصر الارتباط الذاتى لايرجع إلى عامل الصدفة ، ولذلك لابد من إجراء اختبار المعنوية ، وأهم هذه الاختبارات هو اختبار ديربن–واتسون Durbin-Watson test ، حيث يفترض فى هذا الشأن أن الارتباط السلسلى يأخذ النمط البسيط ويرمز له بالرمز DW أو وهو على الصورة .

$$D = D\,W = \frac{\sum_{t=1}^{T} (\hat{e}_t - e_{t-1})^2}{\sum e^2_t}$$

ويأخذ هذا الاختبار القيم بين (0 , 4) وذلك لقيم معامل الارتباط المختلفة حيث

$$DW \approx 2 \quad \Rightarrow \quad \rho = 0$$
$$DW \approx 0 \quad \Rightarrow \quad \rho = 1$$
$$DW \approx 4 \quad \Rightarrow \quad \rho = -1$$

ويوضح الشكل التالى قيم d التى تشير إلى وجود أو غياب ارتباط ذاتى من الدرجة الأولى سالب أو موجب ، كذلك يوضح المناطق التى تجعل هذا الاختبار غير حاسم



حيث dL, du هى القيمة العليا والدنيا فى جداول ديربن واتسون وهى مجدولة طبقا لدرجات الحرية ومستوى المعنوية ١% أو ٥% وبالتالي كلما اقتربت قيمة d من 2 فإن هذا يعنى عدم وجود إرتباط ذاتي ويظهر هذا الاختبار فى برنامج الحاسب SPSS

معالجة مشكلة الارتباط الذاتى

١- عند وجود الارتباط يمكن تطبيق طريقة المربعات الصغرى العادية بصورة معدلة ، ويكون نموذج الانحدار على الصورة

$$y - \rho y_{t-1} = b_0 (1-\rho) + b_1 (x_t - \rho x_{t-1}) + (e_t - \rho e_{t-1})$$

حيث ρ معروفة ونجد فى هذه الحالة أن الأخطاء $(e_t - \rho e_{t-1})$ لا ترتبط ويمكن تقدير المعالم بالطريقة المعتادة.

أما إذا كانت ρ غير معروفة فإنها تقدر باستخدام المعادلة

$$\hat{e}_t = \hat{\rho} e_{t-1} + v_t$$

ثم يتم استخدامها فى نموذج الانحدار لتقدير المعالم الخاصة به.

يلاحظ أن تحويل العلاقة الأصلية إلى علاقة يمكن قياسها بأسلوب المربعات الصغرى العادية تؤدى إلى فقد المشاهدة الأولى ويمكن تعويض هذه المشاهدة عن طريق التعويض التالى ،

$$y_1 = \sqrt{1-\rho^2} \qquad x_1 = \sqrt{1-\rho^2}$$

٢- يمكن افتراض أن قيمة $\rho = 1$ أى وجود ارتباط ذاتى موجب تام وتكون العلاقة على الصورة

$$y_t - y_{t-1} = b_1(x_t - x_{t-1}) + (e_t \; e_{t-1})$$

أى يمكن إستخدام الفروق الأولى للمتغيرات الداخلة فى نموذج الانحدار لتقدير معالم الدالة باستخدام أسلوب المربعات الصغرى.

٣- إذ كان الارتباط السلسلى يرجع إلى حذف بعض المتغيرات التفسيرية (*) ، فإن الحل يكون بادراج هذا المتغير المحذوف صراحة فى النموذج ، كما يمكن أن يرجع الارتباط الذاتي إلى خطأ فى صياغة نموذج الانحدار ذاته كأن يكون من الدرجة الثانية مثلاً.

٤- يمكن التغلب على مشكلة الارتباط الذاتى بإختيار قيم متتالية لمعامل ارتباط الخطأ ρ والتى قد تأخذ قيما تتراوح بين صفر ، ١ وتستخدم فى تقدير النموذج المعتمد على ρ ويقارن مجموع مربعات الأخطاء الناتجة من كل حالة وإختيار قيمة ρ التى تؤدى إلى أصغر قيمة لمجموع هذه المربعات.

الارتباط الذاتى بين البواقى ، وكيفية معالجته باستخدام البرنامج الإحصائى SPSS

مثال : الجدول التالى يتضمن متغيرين **Y** كمتغير تابع ، **X1** كمتغير مستقل

Years	Y	X1
1962	188	460
1963	192	486
1964	200	561
1965	191	553
1966	232	682
1967	280	1010
1968	297	793
1969	306	840

(*) يمكن التعرف على ذلك من خلال تقدير انحدار البواقى على كل متغير من المتغيرات التفسيرية المحذوفة.

Years	Y	X1
1970	396	851
1971	420	1117
1972	227	1102
1973	439	1262
1974	564	3532
1975	759	3711
1976	1030	4281
1977	1368	4563
1978	1474	4977
1979	1671	7597
1980	2196	8757
1981	2445	8875
1982	3287	7612
1983	3179	7789
1984	2780	7893
1985	2774	7322
1986	2575	7164

والمطلوب

١- حساب معادلة انحدار Y على X1

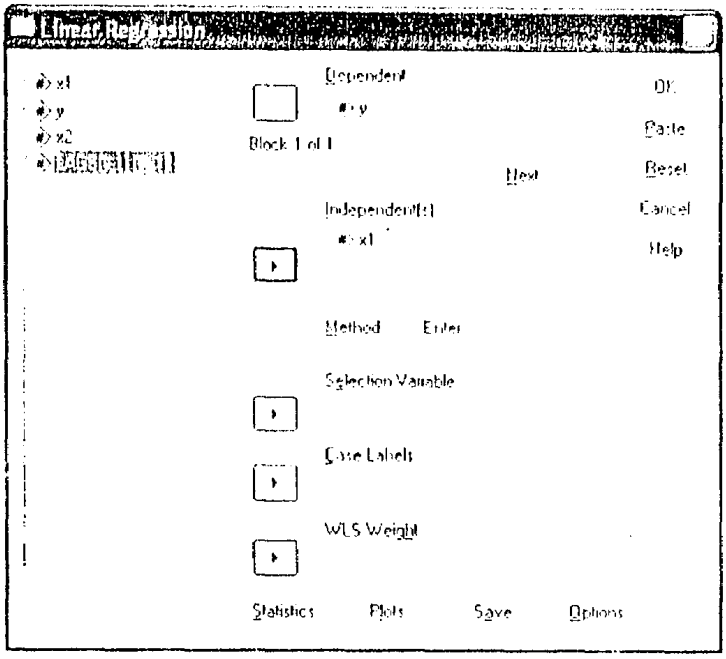
٢- اختبار مشكلة الارتباط الذاتى باستخدام إحصائية DW

٣- معالجة مشكلة الارتباط الذاتى فى حالة وجودها

لتنفيذ ذلك باستخدام البرنامج الإحصائى SPSS تابع الخطوات التالية

١- سجل البيانات فى Data Editor و احفظ الملف .

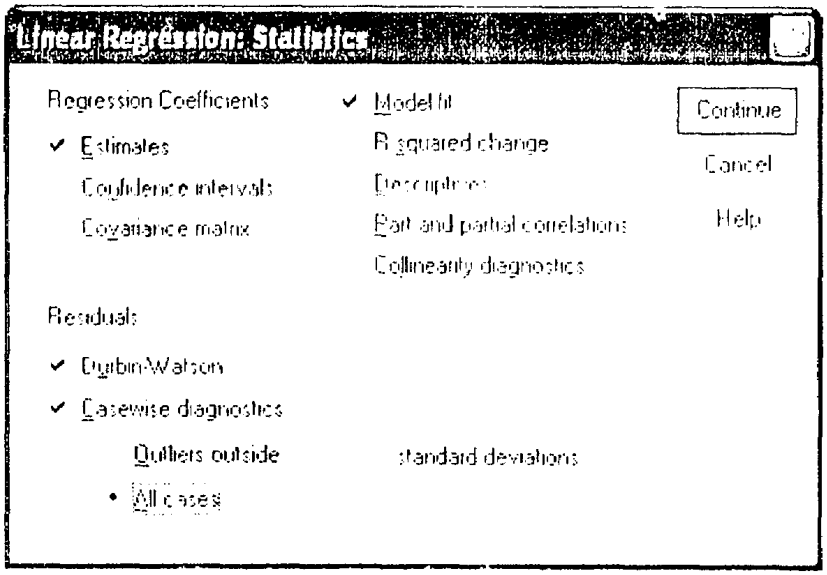
٢- من شريط القوائم اختر Analyze ، ومنها اختر أمر Regression ، ومنها Linear Regression فيظهر الصندوق الحوارى Linear Regression



٣- انقل المتغير Y إلى خانة Dependent

٤- انقل المتغيرين X1 إلى خانة Independent ، وفى خانة Method اختر Enter

٥- اختر أمر Statistics فيظهر الصندوق الحوارى الفرعى Statistics



٦- قم بتنشيط الخيارات التالية

- Model fit : وذلك لتقدير ANOVA, R^2
- Estimates ، لتقدير معالم النموذج
- Durbin-Watson وذلك لحساب إحصائية DW

- Casewise diagnostics ، وذلك لحساب القيم المتوقعة ، والبواقى .

٧- بعد اختيار امر Ok من الصندوق الفرعى ، والصندوق الرئيسى Linear Regression تظهر النتائج التالية

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	x1[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: y

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.941[a]	.886	.881	378.876	.865

a. Predictors: (Constant), x1

b. Dependent Variable: y

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	25555771	1	25555771.07	178.030	.000[a]
	Residual	3301587	23	143547.258		
	Total	28857358	24			

a. Predictors: (Constant), x1

b. Dependent Variable: y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-30.437	118.133		-.258	.799
	x1	.322	.024	941	13.343	.000

a. Dependent Variable: y

Casewise Diagnostics[a]

Case Number	Std. Residual	y	Predicted Value	Residual
1	.185	188	117.83	70.167
2	.174	192	126.21	65.787
3	.131	200	150.39	49.613
4	.114	191	147.81	43.191
5	.112	232	189.39	42.611
6	-.040	280	295.11	-15.112
7	.190	297	225.17	71.833
8	.173	306	240.32	65.684
9	.402	396	243.86	152.138
10	.239	420	329.60	90.399
11	-.258	227	324.77	-97.766
12	.165	439	376.34	62.662
13	-1.436	564	1108.02	-544.017
14	-1.073	759	1165.71	-406.714
15	-.843	1030	1349.44	-319.439
16	-.191	1368	1440.34	-72.335
17	-.263	1474	1573.78	-99.778
18	-1.972	1671	2418.27	-747.271
19	-1.574	2196	2792.17	-596.169
20	-1.017	2445	2830.20	-385.204
21	2.280	3287	2423.11	863.894
22	1.845	3179	2480.16	698.842
23	.703	2780	2513.68	266.320
24	1.173	2774	2329.63	444.368
25	.782	2575	2278.70	296.296

a. Dependent Variable: y

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	117.83	2830.20	1178.80	1031.903	25
Residual	-747.271	863.894	.000	370.899	25
Std. Predicted Value	-1.028	1.600	.000	1.000	25
Std. Residual	-1.972	2.280	.000	.979	25

a. Dependent Variable: y

تفسير النتائج

يمكن كتابة معادلة الانحدار كما يلي :

$$Y = -30.437 + 0.332\ X1$$

اختبار معنوية النموذج

يلاحظ من جدول تحليل التباين ANOVA أن قيمة F = ١٧٨,٠٣ بمستوى معنوية أقل من ٠,٠٠٠٥ مما يدل على النموذج معنوى

اختبار معنوية المعاملات المقدرة

- قيمة T بالنسبة للثابت -٠,٢٥٨ بمستوى معنوية ٠,٧٩٩ مما يدل على أن الثابت غير معنوى
- قيمة T بالنسبة للمتغير X1 = ١٣,٣٤٣ بمستوى معنوية أقل من ٠,٠٠٠٥ مما يدل على أن المتغير X1 معنوى

القدرة التفسيرية للنموذج

بلغت قيمة R2 = ٠,٨٨٦ مما يدل على ان المتغيرات المستقلة تفسر حوالى ٨٨٧% من التغيرات التى تحدث فى المتغير التابع .

اختبار وجود الارتباط الذاتى للأخطاء العشوائية

بلغت قيمة ديربن واتسون ٠,٨٦٥ ، وبالكشف فى جداول ديربن واتسون بمتغير واحد K=1 ودرجات حرية الخطأ ٢٣ نجد ان

$d_L = 1.018$ $d_U = 1.187$

وحيث أن $0 < DW < d_L$

لذا يوجد ارتباط ذاتى ، ويمكن حسابه كما يلى

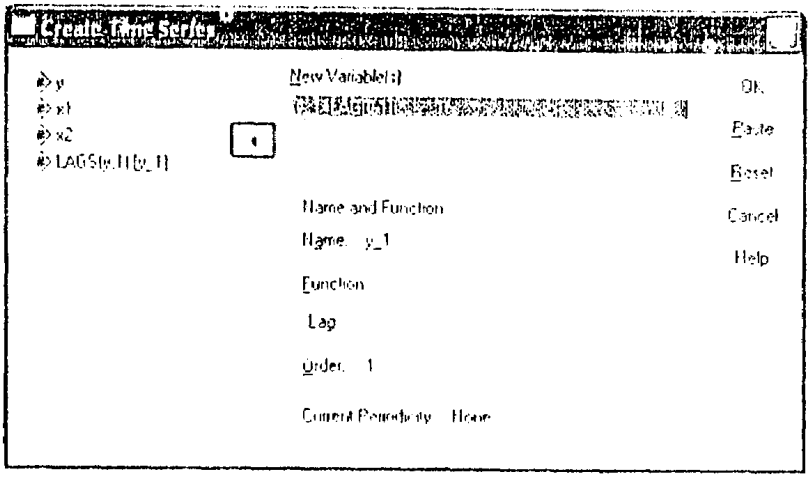
معامل الارتباط الذاتى

$\rho = 1 - (DW/2) = 1 - (0.865/2) = 0.568$

علاج مشكلة الارتباط الذاتى

يمكن علاج مشكلة الارتباط الذاتى عن طرق اختيار المتغير التابع بفترة إبطاء ، ضمن المتغيرات المستقلة ، ولتنفيذ ذلك تابع الخطوات التالية

١- من القائمة Transform اختر امر Create Time Series فيظهر الصندوق الحوارى Create Time Series



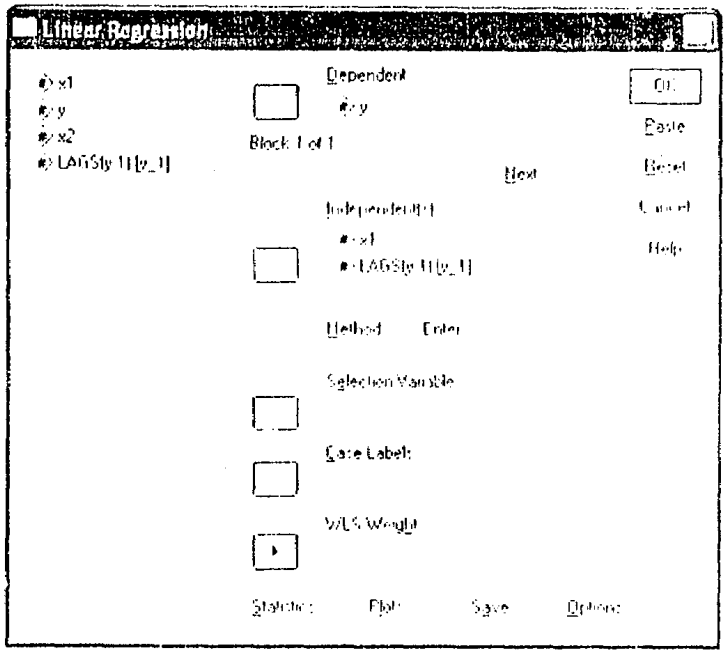
٢- أنقل المتغير Y إلى قائمة New Variable(s)

٣- من خانة Name يظهر اسم المتغير الجديد Y_1 ، ويمكن إعادة تسميته

٤- من خانة Function اختر Lag ، ومن خانة Order اكتب فترة الإبطاء ١

٥- اختر أمر Ok يتم إضافة المتغير الجديد Y_1 إلى المتغيرات فى نافذة Data Editor

٦- نفذ الانحدار المتعدد بالطريقة السابق ذكرها بعد اختيار المتغير التابع Dependent = Y ، والمتغيرات المستقلة X1 , Y_1



٧- بعد إظهار نتائج الانحدار يمكن النظر إلى إحصائية DW حيث نجدها = ٢,٢٣٦

Model Summary[c]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.974[a]	.950	.947	252.710	
2	.985[b]	.971	.968	197.601	2.236

a. Predictors: (Constant), y_1 LAGS(y,1)

b. Predictors: (Constant), y_1 LAGS(y,1), x1

c. Dependent Variable: y

ومنه نجد أن $2 < dw < (4-du)$

أى $2 < 2.236 < 2.723$ مما يدل على عدم وجود ارتباط ذاتى

٥-٣ وجود أخطاء القياس والمشاهدة فى المتغيرات التفسيرية.

تشترط طريقة المربعات الصغرى فى التقدير عدم وجود أخطاء مشاهدة أو قياس بالمتغيرات التفسيرية ، وهو شرط يضمن إستقلال المتغيرات التفسيرية عن عنصر الخطأ العشوائي . وترجع أخطاء القياس إلى أن المتغيرات المستخدمة غالبا ما تكون تجميعية أو أن تستخدم متغيرات ليست هى المتغيرات المفروض إدخالها فى النموذج كاستخدام الناتج القومى الاجمالى كمتغير محل الدخل الممكن التصرف فيه عند تقدير دالة الاستهلاك ، واستخدام الأرقام القياسية للأسعار لتحويل القيم النقدية إلى قيم حقيقية ، ومعروف أن هذه الأرقام عرضه للأخطاء . ولا يوجد اختبارات معينة للكشف عن وجود أخطاء فى المتغيرات ولكن يمكن أن تعطى النظرية الاقتصادية أو المعرفة بالطريقة التى جمعت بها البيانات إشارة إلى مدى خطورة المشكلة .

ويترتب على وجود هذه الأخطاء واستخدام أسلوب المربعات الصغرى العادية فى التقدير تحيز التقديرات.

علاج أخطاء القياس

يمكن علاج هذه الأخطاء بأسلوبين

١- أسلوب الانحدار المرجح **Weighted Regression**

٢- المتغيرات المساعدة

يلاحظ شيوع عدم ثبات التباين فى البيانات المقطعية ففي الإنفاق الاستهلاكي مثلا نجد أن تقلباته فى فئات الدخل العليا أعلى من تقلبات فئات الدخل المنخفضة .

ويؤدى عدم ثبات التباين إلى عدم صلاحية استخدام تباينات المعالم المقدرة بأسلوب المربعات الصغرى لإجراء اختبارات المعنوية الإحصائية وتحديد فترات الثقة لهذه التقديرات ،وذلك بالرغم من عدم تحيزها.

ويمكن اختبار عدم ثبات التباين بمعامل ارتباط الرتب سبيرمان وذلك عن طريق تقدير علاقة الانحدار بالطريقة العادية وإيجاد قيم البواقى e التى هى تقدير للأخطاء u ثم ترتب مع قيم e ترتيبا تصاعديا أو تنازليا (مع إهمال الإشارة) معامل ارتباط الرتب من المعادلة

$$r_{e,x} = 1 - \frac{6 \sum D_i^2}{N(N^2 - 1)}$$

حيث D_i هى الفرق بين رتب القيم المتناظرة من x , e كلما كبرت قيمة r كلما دل ذلك على عدم ثبات التباين لعنصر الخطأ .

علاج عدم ثبات التباين

يمكن معالجة مشكلة عدم ثبات التباين بتعديل النموذج بحيث يصبح عنصر الخطأ له تباين ثابت.

فالعلاقة

$$y_i = b_o + b_1 x_i + u_i$$

يمكن أن تكون

$$y_i / x_i = b_1 + b_o (1/x_i) + (u_i/x_i)$$

وهذه العلاقة تسمح بأن يكون عنصر الخطأ العشوائى له تباين ثابت حيث

إذا كان $Var(u_i) = c\ x^2_i$

$$v\left(\frac{u_i}{x_i}\right) = \frac{c}{x^2_i} x^2_i = c$$

فان تباين $u^* = \frac{u_i}{x_i}$ يكون ثابت

وبالتالى يمكن إستخدام المربعات الصغرى العادية لتقدير معالم الدالة والتى سوف تتسم بعدم التميز وأصغر تباين .

٢ – قد تكون هناك معلومات عن قيمة تباين الخطأ من معلومات سابقة كأن يكون

$$Var(e_i) = \sigma^2_i$$

فى هذه الحالة يمكن إستخدام هذا التباين فى تعديل علاقة الانحدار وتكون على الصورة

$$y_i / \sigma_i = b_o + \sigma_i + b_1 x_1 / \sigma_i + e_i / \sigma_i$$

وتقدر المعالم بطريقة المربعات الصغرى العادية.

المراجع

أولا : المراجع العربية

١- دومينك سالفاتور (ترجمة د. سعدية منتصر) الإحصاء والاقتصاد القياسى سلسلة ملخصات شوم، دار ماكجروهيل للنشر-١٩٨٢.

٢- ربيع ذكى عامر، تحليل الانحدار أساليبه وتطبيقاته العملية باستخدام البرنامج الجاهز + SPSS/pc - جامعة الكويت ١٩٨٩م.

٣- رجاء محمد أبو علام ، التحليل الاحصائى للبيانات باستخدام برنامج SPSS القاهرة - دار النشر للجامعات ٢٠٠٣م.

٤- سمير كامل عاشور ، سامية أبو الفتوح ، العرض والتحليل الاحصائى باستخدام + SPSS/pc معهد الدراسات والبحوث الإحصائية ١٩٩٣م.

٥- سمير كامل عاشور ، سامية أبو الفتوح ، العرض والتحليل الاحصائى باستخدام SPSSwin الجزء الأول المدخل والأساسيات-معهد الدراسات والبحوث الإحصائية ٢٠٠٣م.

٦- سعد زغلول بشر، دليلك إلى البرنامج الاحصائى SPSS الإصدار العاشر Version 10 - المعهد العربى للتدريب والبحوث الإحصائية بغداد ٢٠٠٣م.

٧- عبد الحميد عبد اللطيف : استخدام الحاسب الآلى فى مجال العلوم الاجتماعية (استخدام برنامج SPSS من خلال windows) جامعة عين شمس ٢٠٠١م.

٨- عبدالله بن عمر النجار، استخدام حزمة البرامج الإحصائية (SPSS) فى تحليل البيانات مؤسسة شبكة البيانات -الرياض -٢٠٠٣م.

٩- عبد القادر محمد عبد القادر : طرق قياس العلاقات الاقتصادية ، مع تطبيقات الحاسب الألكترونى، ١٩٩٠.

١٠- عبد الرحيم عبد الحميد الساعاتى ، مدحت فهمى صالح ، مبادئ التحليل الإحصائى للاقتصاد والإدارة باستخدام برنامج EXCEL ، ٢٠٠٤ ، دار جدة .

١١- لنكون شاو- تعريب د.عبد المنعم حامد عزام-الإحصاء فى الإدارة -دار المريخ للنشر ١٩٩٦م.

١٢- محمد الدسوقى حبيب ، جلال الصياد ، مقدمة فى الطرق الإحصائية، ١٩٩٧.

١٣- مجيد علي حسن. عفاف عبد الجبار سعيد . الاقتصاد القياسي النظرية والتطبيق-دار وائل للنشر والتوزيع عمان الأردن ١٩٩٨م.

ثانيا : المراجع بالغة الأجنبية

1- **Damodar Gujarati, Basic Econometries, 3rd ed., Mc Graw-Hall. Inc., 1995.**

2- **Sheridan J.coa Kes, Lyndall G.steed, SPSS Analysis without Anguish, version 10.0 for windows, 2000.**