التحليل الإحصائي للبيانات

باستخدام SPSS-WIN

أ. د. إسماعيل الفقي
د. محمد قايد عبدالجواد
د. مرفت مهدي

العبيكان
Obeikan

التحليل الإحصائي للبيانات
باستخدام SPSS-WIN

إعداد

أ.د. إسماعيل الفقي

أستاذ علم النفس - كلية العلوم الاجتماعية

جامعة الإمام محمد بن سعود الإسلامية بالرياض

د. محمد قاير

أستاذ الإحصاء المشارك بكلية العلوم

جامعة الملك سعود في الرياض

د. مرفت مهدي

أستاذ الإحصاء المساعد

كلية التجارة، جامعة بنها

العبيكان
Obeikan

مع تمنياتي بمطالعة ممتعة ا. محمد عموش

ⓒ مكتبة العبيكان، ١٤٣٢هـ

فهرسة مكتبة الملك فهد الوطنية أثناء النشر

الفقي، إسماعيل

التحليل الإحصائي للبيانات باستخدام SPSS-WIN /.

محمد فايد عبدالجواد؛ مرفت مهدي؛ إسماعيل الفقي.- الرياض، ١٤٣٢هـ.

٢٣٨ص، ٢٠× ٢٨ سم

ردمك: ١ - ١٥٦ - ٥٠٣ - ٦٠٣ - ٩٧٨

١- البرمجة (حواسيب) ٢- التحليل الإحصائي - معالجة

أ. مهدي، مرفت (مؤلف مشارك) ب- الفقي، إسماعيل (مؤلف مشارك)

أ. العنوان

ديوي ٠٠٥.٣ ١٤٣٢/٤٩٥٩

ردمـــــك: ١ - ١٥٦ - ٥٠٣ - ٦٠٣ - ٩٧٨

رقم الإيداع: ١٤٣٢/٤٩٥٩

الطبعة الأولى

١٤٣٣هـ / ٢٠١٣م

الناشر العبيكان للنشر

الرياض-شارع العليا العام-جنوب برج المملكة

هاتف ٢٩٢٧٥٧٤ - ٢٩٢٧٥٨١ / فاكس ٢٩٢٧٥٨٨

ص.ب ٦٧٦٢٢ الرمز ١١٥١٧

التوزيع: مكتبة العبيكان

الرياض- العليا- تقاطع طريق الملك فهد مع العروبة

هاتف ٤١٦٠٠١٨ - ٤٦٥٤٤٢٤/ فاكس ٤٦٥٠١٢٩

ص.ب ٦٢٨٠٧ الرمز ١١٥٩٥

لا يسمح بإعادة إصدار هذا الكتاب أو نقله في أي شكل أو وسيلة، سواء أكانت إلكترونية أو ميكانيكية، بما في ذلك التصوير بالنسخ «فوتوكوبي» أو التسجيل، أو التخزين والاسترجاع، دون إذن خطي من الناشر.

مع تمنياتي بمطالعة ممتعة ا. محمد عموش

مع تمنياتي بمطالعة ممتعة

أ. محمد عموش

مع تمنياتي بمطالعة ممتعة

أ. محمد عموش

محتويات الكتاب

الموضوع — الصفحة

مقدمة 11
الفصل الأول: مقدمة للحزمة الإحصائية 13
مقدمة 13
التطور التاريخي 13
الدخول للبرنامج 15
مكونات الحزمة 16
المربع الحواري 19
أنواع الملفات في البرنامج 19
الفصل الثاني: إعداد البيانات ومعالجتها 21
مقدمة 21
تعريف المتغيرات 21
محاذاة النص 28
أنواع وحدات القياس 28
إدخال البيانات 29
عمليات مختلفة على البيانات 31
ترتيب البيانات 36
الفصل الثالث: إنشاء متغيرات جديدة 43
مقدمة 43
قائمة Transform 43
استخدام العمليات الرياضية لإنشاء متغيرات مستمرة جديدة/الأمر Compute 44
تكويد متغير لفظي إلى متغير رقمي 54
إحلال قيم مفقودة بقيم أخرى 56
الترتيب Ranking 57
إنشاء سلسلة زمنية 60

مع تمنياتي بمطالعة ممتعة — أ. محمد عموش

الفصل الرابع: العرض والتحليل الإحصائي لمتغير أحادي 63
مقدمة 63
عرض البيانات باستخدام Bar, Line, and Area 63
مقاييس النزعة المركزية Measures Of Central Tendency 69
مقاييس التشتت المطلق والنسبي 70
الفصل الخامس: العرض والتحليل الإحصائي لمتغيرين 81
مقدمة 81
الجداول المزدوجةCross Tables: 81
شكل الانتشار Scatter Plot 88
الفصل السادس: اختبارات الفروض 91
مقدمة 91
تعاريف أساسية في اختبارات الفروض الإحصائية 91
اختبار عينة واحدة 95
شروط الاختبارات للعينتين المترابطتين 104
الفصل السابع: تحليل التباين 109
مقدمة 109
تحليل التباين في اتجاه واحد (One-Way ANOVA) 111
تحليل التباين في اتجاهين (Two Way ANOVA) 122
الفصل الثامن: تحليل الارتباط 129
مقدمة 129
الارتباط الخطي البسيط Simple Correlation 129
الارتباط الجزئي Partial Correlation 137
الفصل العاشر: تحليل الانحدار 141
مقدمة 141
نموذج الانحدار الخطي البسيط Simple Regression 141
الانحدار الخطي المتعدد Multiple Regression 161
اختبار المتغيرات المستقلة الداخلة في النموذج الخطي 168
الفصل العاشر: الاختبارات اللامعلمية (اللابارامترية) 175
مقدمة 175

اختبار مربع كاي (Chi-Square test) 177
اختبار الدورة (Run Test) 189
اختبار كلوموجروف سيمنروف لعينة واحدة 192
اختبار مان – ويتني 198
اختبار كرسكال والاس (The Kruskal – Wallis) 205
الفصل الحادي عشر: التعامل مع النتائج 213
قص - نسخ -لصق جداول النتائج 213
نسخ الجداول والأشكال البيانية 215
الإظهار Expand والإخفاء Collapse وتحريك Move النتائج 215
إضافة عنوان جديد Add New Title 221
محرر الجداول Pivot Tables 223
تعديل وتنقيح الأشكال البيانية 230
المراجع العربية والأجنبية 237

مُقَدِّمَة

الحمد لله رب العالمين والصلاة والسلام على المبعوث رحمة للعالمين سيدنا محمد بن عبد الله وعلى آله وصحبه أجمعين.

أما بعد...

لما كانت المكتبة العربية في حاجة إلى مزيد من المؤلفات التي تتناول تطبيقات العلوم المختلفة في شتى المجالات، فإنه يسعدنا أن نضع بين يدي القارئ الكريم كتابًا في تطبيقات علم الإحصاء في العلوم الإنسانية والاجتماعية وتطبيقات العلوم الطبيعية، وهو كتاب (التحليل الإحصائي للبيانات باستخدام SPSS-WIN)، راجين الله -عز وجل- أن ينفع به طلاب العلم والمعرفة؛ وليكون خطوة على درب البحث العلمي في مجال التحليل الإحصائي للبيانات، ويكون عونًا لهم على كيفية التعامل مع بيانات البحوث والدراسات المختلفة؛ بداية من تنظيم هذه البيانات وإدخالها للحاسب الآلي، ثم معالجتها إحصائيًا باستخدام برنامج "SPSS"، وكيفية التعامل مع مخرجات هذه المعالجات.

وحيث إن عملية تحليل البيانات من المراحل الأساسية والمهمة في إعداد البحوث والدراسات العلمية، فإن هذا الكتاب يهدف -بعون الله تعالى- إلى مساعدة الباحثين وطلاب العلم على التعامل مع بياناتهم بفهم واعٍ، وسهولة ويسر.

ومن أجل تحقيق هذا الهدف حرصنا على تقديم المفاهيم الإحصائية المختلفة في صورة بسيطة ميسرة وواضحة، وفي خطوات متسلسلة أوضحنا كيفية التعامل مع برنامج SPSS - الإصدار الثامن عشر- وهو الصورة الحديثة من الحزمة الإحصائية لبرنامج SPSS من خلال الأمثلة التوضيحية المتعددة لكل العمليات الإحصائية التي جاءت في ثنايا هذا العمل لبيان الجوانب التطبيقية للعلوم المختلفة.

ونأمل أن يكون هذا الكتاب مكملاً لما بذله الآخرون من جهد في هذا المضمار، وأن يمثل حلقة في التواصل العلمي لتحقيق التكامل بين مختلف العلوم، وأن يحقق الاستفادة المرجوة لطلاب العلم والمعرفة في الدراسات التي تتخذ من الأساليب الإحصائية الكمية طريقًا لمعالجة البيانات.

ونتوجه بالشكر والعرفان لكل من ساهم في إخراج الكتاب على هذه الصورة، ونخص بالشكر الأستاذ الدكتور خلف سلطان، والدكتور إبراهيم البطل، والدكتور محمود الدريني، والدكتور فتحي معبد، والأستاذ عادل باسمق والأستاذ محمد الصالح، على ما قدموه لمعدي هذا الكتاب.

والله نسـأل أن ينفع بهذا الكتاب طلاب العلـم في مرحلتي البكالوريوس والدراسات العليا، والباحثين والزملاء في التخصصات العلمية المختلفة في تحليل ومعالجة البيانات.

ويسعدنـا أن نتلقى ملاحظاتكم حول هذا الكتاب، ونسأل الله تعالى التوفيق والسداد، وهو الهادي إلى سواء السبيل.

وآخر دعوانا أن الحمد لله رب العالمين

الرياض: شعبان 1432 هـ

يوليو 2011 م

الفصل الأول

مقدمة للحزمة الإحصائية *SPSS-WIN*

SPSS-WIN Package

(1-1) مقدمة :

علم الإحصاء أداة أساسية لا غنى عنها لتوصيف البيانات وتحليلها وحساب التقديرات والتنبؤات المستقبلية؛ ونظراً لكبر حجم البيانات التي يتعامل معها علم الإحصاء من جهة، واعتماده على أساليب كمية مطولة من جهة أخرى، فإن استخدام الحاسب الآلي أصبح ضرورياً لإنجاز العمليات الإحصائية اختصارا للجهد والوقت.

يعد برنامج Statistical Package for Social Science (SPSS) أقدم البرامج الإحصائية وأكثرها استخداماً من قبل شريحة واسعة من الطلبة والباحثين في مختلف التخصصات الإحصائية والطبية والهندسية والزراعية و الاجتماعية والتربوية والنفسية. ونظراً لقلة عدد من يجيد استخدام برنامج الـ SPSS بصورة وافية، إضافة إلى افتقار المكتبة العربية إلى كتب تعليمية حول هذا البرنامج، فقد كان هدفنا من خلال هذا الإصدار أن نوفر مصدراً تفصيلياً بين أيدي الباحثين في مختلف المجالات وطلاب مرحلة البكالوريوس في أقسام الإحصاء والتخصصات المختلفة الأخرى.

يتضمن هذا الكتاب تعريفاً بالجوانب الأساسية لحزمة SPSS، بهدف إكساب العديد من المهارات اللازمة لتحقيق الاستفادة القصوى من إمكانيات البرنامج المتاحة آخذين في الحسبان أن هناك عددًا كبيراً من الدارسين ليست لديهم خلفية إحصائية وافية تمكنهم من التعامل مع البرنامج بصورة صحيحة. وقد تم التعامل مع تطبيقات البرنامج من خلال أمثلة مبسطة تتيح للقارئ الانتقال إلى خطوات متقدمة بسهولة، ومعظم هذه الأمثلة مأخوذ من مصادر عربية وأجنبية معتمدة.

(1-2) التطور التاريخي:

من المعلوم أنه لحساب قيمة أي مقياس إحصائي فإننا في حاجة إلى إجراء العديد من الخطوات الرياضية. على سبيل المثال؛ لحساب قيمة الانحراف المعياري (أحد مقاييس التشتت) نقوم بإجراء العمليات الحسابية الآتية:

1. إيجاد عدد المشاهدات،

2. حساب مجموع المشاهدات،

3. قسمة قيمة ناتج العملية الحسابية رقم (2) على عدد البيانات الذي يعطي قيمة الوسط الحسابي،

4. لكل مشاهدة يتم إيجاد الانحراف عن الوسط الحسابي المحسوب في العملية الحسابية رقم (3)،

5. حساب مربع قيمة ناتج العملية الحسابية رقم (4) لكل حالة،

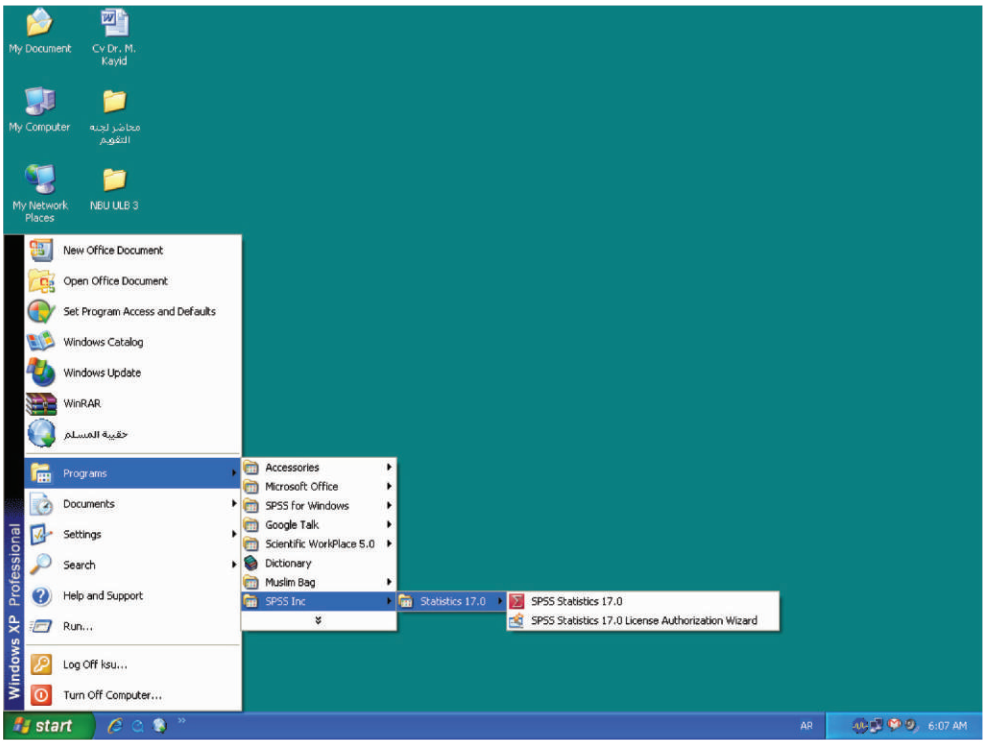
6. إيجاد مجموع مربعات قيمة ناتج العملية الحسابية رقم (5)،

7. قسمة قيمة ناتج العملية الحسابية رقم (6) على عدد المشاهدات إذا كنا نتعامل مع بيانات المجتمع أو القسمة على عدد المشاهدات مطروح منها الرقم واحد إذا كنا نتعامل مع العينة ونريد إيجاد أفضل تقدير، وبشكل عام هذه الخطوة تعطي تباين المجتمع أو أفضل تقدير له.

8. للحصول على الانحراف المعياري نوجد الجذر التربيعي لقيمة ناتج العملية الحسابية رقم (7).

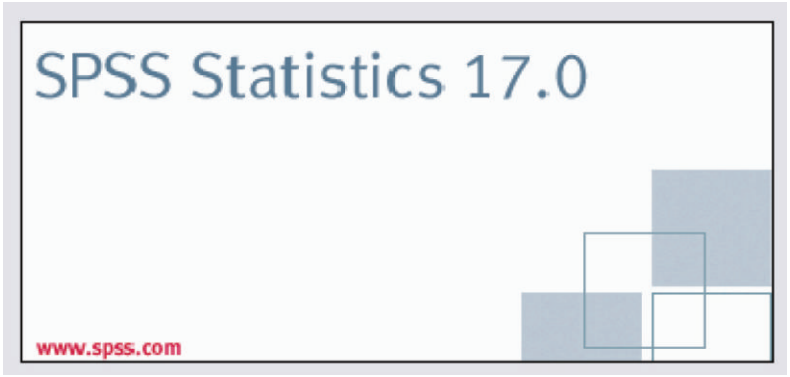
وعلى الرغم من أن الخطوات السابقة قد توحي بأن العملية بسيطة فإنه إذا كان عدد المشاهدات كبيراً فإن مقدار الجهد المطلوب لحساب الخطوات السابقة سيكون كبيراً، فضلاً عن الخطأ الذي قد يقع فيه الباحث إذا تم الحساب بشكل يدوي. ومن ثم فإن استخدام برنامج SPSS أو أي برنامج إحصائي آخر، هو البديل لتجنب مثل هذه الأخطاء، وذلك من خلال اختيار المقياس المناسب، وتحديد بيانات المتغيرات قيد الدراسة داخل البرنامج. ومنذ ظهور البرنامج عام 1968م، ومع زيادة ابتكار مقاييس إحصائية عديدة لمعالجة المشكلات الجديدة التي تظهر في العلوم كان من المناسب مواكبة البرنامج لهذا التقدم ليشمل معظم العلوم، ما جعل البرنامج أكثر انتشاراً واستخداماً. ومن ناحية توافق البرنامج مع برامج التشغيل، فقد كان التطور في ذلك مناسباً جداً، حيث كان البرنامج يعمل تحت نظام التشغيل MS-DOS، وتم تطويره ليعمل في بيئة نظام التشغيل WINDOWS في عام 1993م، متلافياً بذلك الصعوبات التي كانت تواجه مستخدمي هذا النظام في بيئة MS-DOS. وقد توالت الإصدارت لهذا النظام التي كان آخرها الإصدار السابع عشر، حيث يوفر هذا النظام مجالاً واسعاً للتحليلات الإحصائية وإعداد المخططات البيانية لتلبية حاجة المختصين والمهتمين بمجال الإحصاء. كما يوفر إمكانية تناقل البيانات مع قواعد البيانات وبرامج EXCEL و LOTUS وغيرهما من البرامج الأخرى، ومن ثم فإن هذه الإمكانيات مناسبة جداً لموضوعات الكتاب.

بعد تحميل البرنامج ننقر على Start في شريط المهام، ومن ثم النقر على Programs ثم اختيار SPSS كما في الشكل (1-1) الآتي.



شكل (1-1)

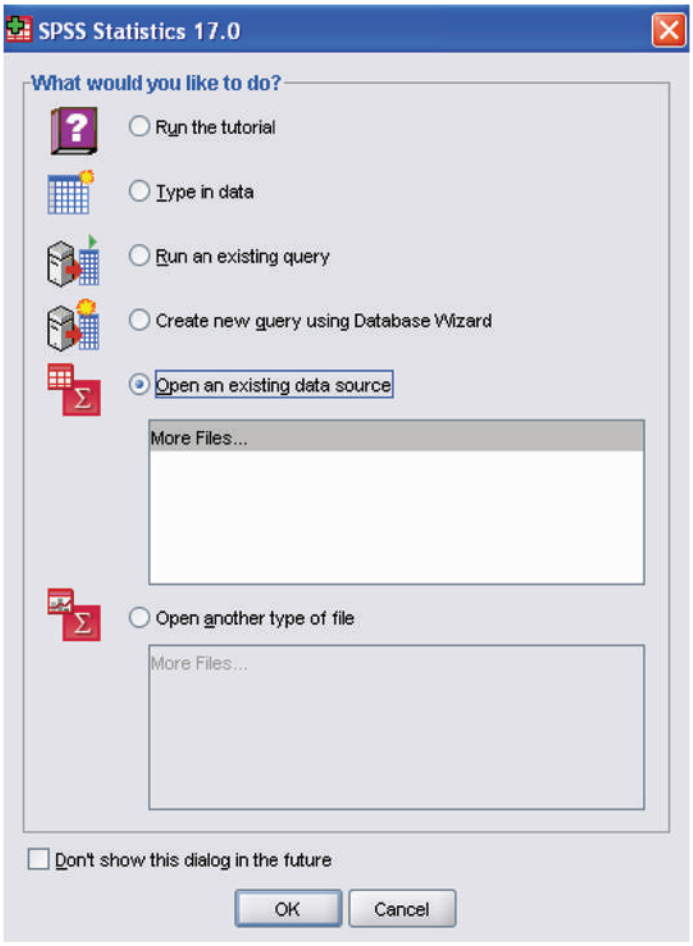
نبـدأ الدخول إلى البرنامج وذلك بظهـور الشاشة الافتتاحية للبرنامج، وهـو إعلان عن الحزمة لمدة ثوان، كما في الشكل الآتي:



شكل (1-2)

بعد ظهور الشاشة الافتتاحية تبدأ أول شاشة من شاشات الحزمة في الظهور.

ملاحظة: في هذا الإصدار وبعد ظهور الشاشة الافتتاحية تصدر شاشة اختيارية بعنوان: (What Would You Like To Do?)



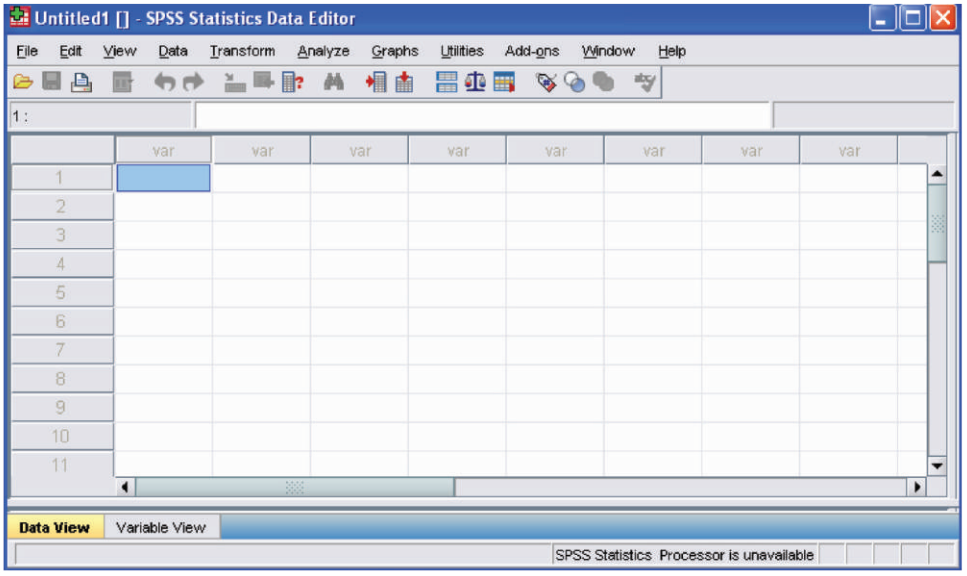
شكل (1-3)

هذه الشاشة اختيارية وبها عدة اختيارات، منها: هل يطلب المستخدم تشغيل البرنامج التعليمي، إدخال البيانات، فتح ملف؟ ... إلخ، يمكن للمستخدم اختيار ما يريد تنفيذه، وذلك بالنقر على الاختيار المطلوب، ثم النقر على Ok أو النقر على Cancel للانتقال إلى الشاشة الآتية. ويمكن للمستخدم إعطاء أمر للبرنامج بعدم إظهار هذه الشاشة مرة أخرى، وذلك بالنقر على العبارة: (Do Not Show This Dialog In The Future).

(1-4) مكونات الحزمة:

- نافذة محرر البيانات (Data Editor):

تعرض هذه النافذة محتويات ملف معين من البيانات، حيث يمكن إنشاء ملف جديد أو تحرير ملف موجود، وإن هذه النافذة تفتح تلقائيًا عند بدء تشغيل البرنامج كما في الشكل (1-4) الآتي:



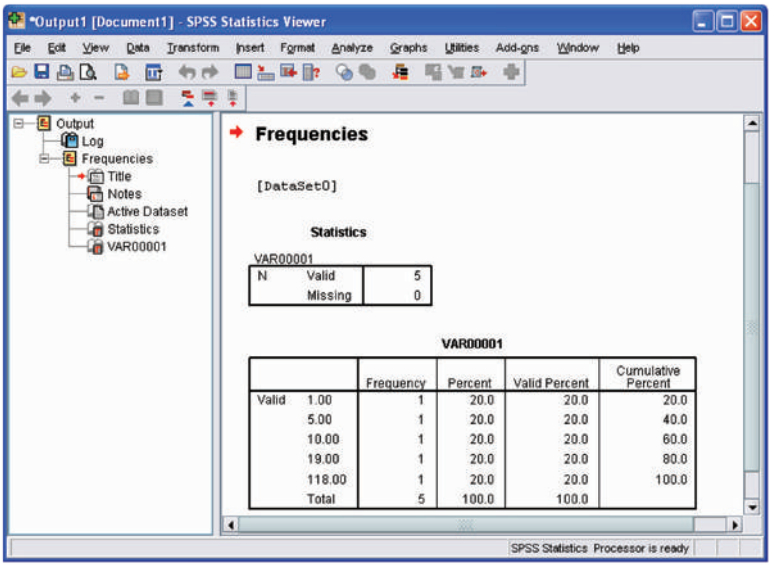
شكل (1-4)

مع العلم أن بيانات هذه الشاشة قابلة للتعديل والحفظ والطباعة للحصول على الأشكال البيانية والجداول، بالإضافة إلى العرض والتحليل الإحصائي للحصول على المقاييس والاختبارات الإحصائية المتاحة.

نافذة شاشة المخرجات:

يعرض في هذه الشاشة كل النتائج الإحصائية، حيث تنقسم إلى جزأين:

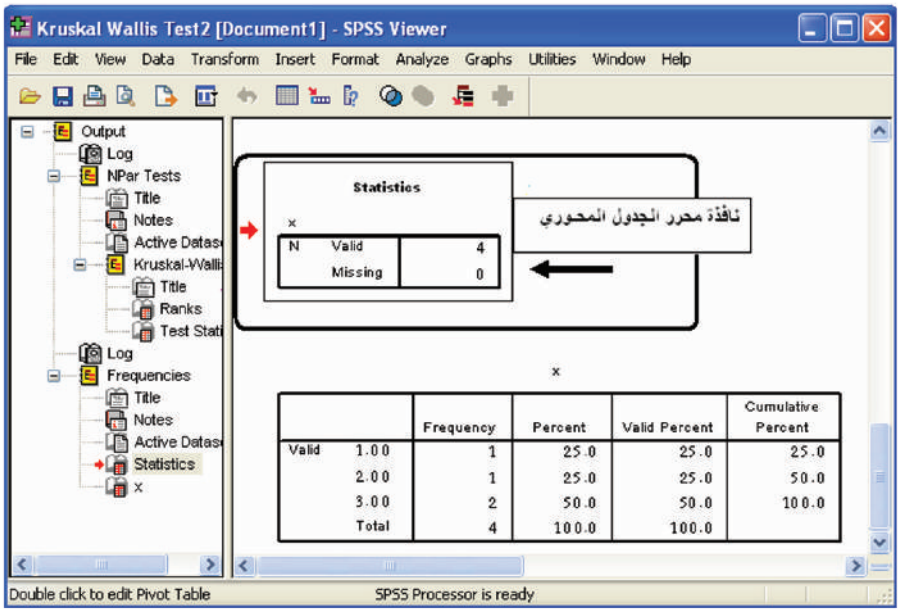
جـز ء على اليسار يحتـوي على فهرس لمحتويات الشاشة، وجز ء على اليمين لعرض محتويات الشاشة مـن جداول وأشكال بيانيـة ونتائج وتحليلات إحصائية ... إلخ . كما تحتـوي الشاشة على شريط للأدوات وشريط للأوامر كما في شاشة محرر البيانات. انظر الشكل (1 – 5) الآتي:



الشكل(1 – 5)

نافذة محرر الجدول المحوري:

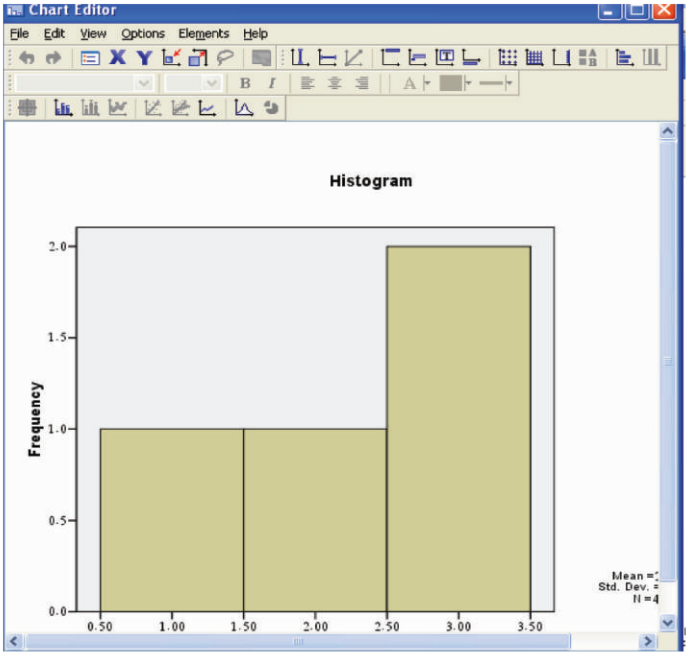
يمكننا إجراء تعديل على الجداول في صفحة النتائج، وذلك بالنقر على الزر الأيمن بالفأرة على الشكل المراد تعديله فيتم تحديد الجدول، عندئذ يمكن تعديل أو إلغاء صفوف وأعمدة من الجداول، وكذلك تغيير الألوان وإنشاء جداول متعددة الاتجاهات، انظر الشكل (1 – 6) الآتي.



شكل(1 - 6)

نافذة محرر الأشكال:

يمكن إجراء تعديل على الأشكال البيانية والرسومات بواسطة محرر الأشكال، عند فتح صفحة النتائج والنقر المزدوج بالمفتاح الأيمن للفأرة على الشكل المراد تعديله سيؤدي ذلك إلى فتح شاشة المحرر وظهور شريط الأدوات الخاص بهذه الشاشة، ويمكن العودة إلى شاشة النتائج بالنقر المزدوج على الشكل المعدل، انظر شكل (1 – 7) الآتي:



شكل (1 – 7)

كما يوجد بعض النوافذ الأخرى مثل نافذة محرر النص (Text Output Editor)، نافذة محرر القواعد (Syntax Editor)، ونافذة محرر الدور (Script Editor).

(1-5) المربع الحواري:

يتيح هذا الإصدار من البرنامج اختيار المتغيرات التي نرغب في إجراء التحليلات الإحصائية عليها عن طريق المربعات الحوارية، كما أنه يعد بديلا عن كتابة أوامر البرمجة المعقدة، عند طلب أي أمر من الحزمة SPSS، علماً بأن البرنامج يُظهر مربعاً حوارياً يحتوي على مجموعة من الخيارات، كل خيار يُمَثَّل بزر (أمر)، وللباحث اختيار ما يناسبه ليضعها موضع التنفيذ وبسرعة فائقة. يتكون مربع الحوار في برنامج SPSS من العناصر الآتية:

- قائمة متغيرات المصدر Source Variables List: وتشمل جميع المتغيرات الموجودة في الملف الحالي ذات الأنواع المسموح الاستخدام للأسلوب الإحصائي المختار.
- قائمة متغيرات الهدف Target Variables List: واحدة من القوائم التي تتضمن أسماء المتغيرات المختارة لإجراء التحليل الإحصائي عليها.
- أزرار الأوامر Command Pushbuttons: و هذه المفاتيح تقوم بإعلام البرنامج لتنفيذ عمل معين مثلاً تمشية البرنامج أو الحصول على مساعدة.

(1-6) أنواع الملفات في البرنامج:

تتوافر في برنامج SPSS عدة أنواع من الملفات، منها ما يأتي:

- ملفات البيانات Data Files:

تتكون هذه الملفات باستخدام محرر البيانات Data Editor، وهي تحتوي على البيانات التي تستخدم في التحليل الإحصائي، ويكون لهذا النوع من الملفات الامتداد SAV.

- ملفات المخرجات الإحصائية Output Files:

وهي ملفات تحتوي على مخرجات التحليل الإحصائي أو المخططات وتكون ذات امتداد SPS.

- ملفات التعليمات Syntax:

وهي الملفات التي تحوي الإجراءات الإحصائية التي تخزن على شكل أوامر وتكون ذات امتداد SPS.

الفصل الثاني

إعداد البيانات ومعالجتها

Preparation and Manipulation of Data

(2-1) مقدمة:

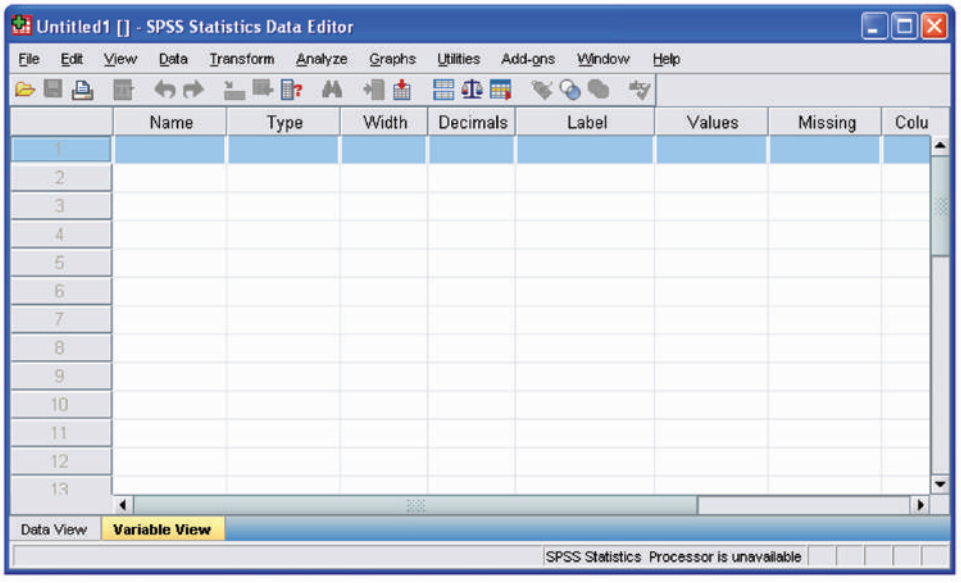
استعرضنا في الفصل السابق طريقة الدخول للبرنامج وبعض النوافذ الرئيسة بشكل عام. في هذا الفصل نناقش طريقة تعريف المتغيرات وإدخال البيانات قيد الدراسة للبرنامج تمهيداً لإجراء التحليل الإحصائي عليها.

(2-2) تعريف المتغيرات:

تعريف المتغير في برنامج SPSS يعنى أنه يجب تحديد ما يأتي:

- اختيار اسم المتغير Names Variables.
- اختيار مميز المتغير Variable label.
- اختيار قيمة المميز إن وجد Value label.
- تعريف القيم المفقودة Missing Values.
- تحديد نوع المتغير Variable Type.
- تحديد شكل عرض البيانات في الأعمدة Column format.

يتم تعريف المتغير من شاشة Variable View الظاهرة في الشكل الآتي:



شكل (2-1)

وسنقوم بشرح كل عمود من أعمدة الشاشة السابقة كما يأتي:

☜ اختيار اسم المتغير (Name):

تبـدأ عمليـة تعريف المتغير بتحديد اسم للمتغير في العمـود الأول من أعمدة Variable View الظاهر في الشكل، وذلك بالنقر عليه نقراً مزدوجاً بزر الماوس الأيمن، وكتابة اسم المتغير باستخدام لوحة المفاتيح الخاصة بجهازك.

توجد عدة شروط يجب أن تؤخذ في الاعتبار عند تحديد اسم المتغير هي:

- يجب ألا يزيد اسم المتغير على 64 حرفاً (for version 15 or 12) أو ثمانية حروف أو أرقام للنسخ السابقة.
- أن يبدأ الاسم بحرف (ليس رقماً) ولا تكون به فراغات .
- لا يسمح باستخدام الحروف الآتية كجزء من اسم المتغير (. / ! /؟ /، / *).
- لابد أن يكون الاسم وحيدًا في الملف نفسه بمعنى أن كل متغير يأخذ اسمًا لا يأخذه متغير آخر.
- لا يمكن استخدام الكلمات الآتية كاسم للمتغير With – Ge – Not – And – Gt – Or - It by – all – Ne – Eq – To – Le؛ لأن هذه الكلمات لها استخدامات معينة داخل الحزمة SPSS.
- يمكن كتابة الاسم باللغة الإنجليزية سواء بحروف كبيرة أو صغيرة.

☜ تعريف نوع المتغير (Type):

العنصـر الثاني من تعريف المتغيرات وهو العمود الثاني من نافذة Variable View، ويعد هذا العنصر في منتهـى الأهميـة، حيث تختلف أنـواع المقاييس الإحصائيـة المستخدمة على حسب نـوع المتغيرات محل الدارسة. الجدول الآتي يوضح أنواع البيانات المتاحة في برنامج SPSS:

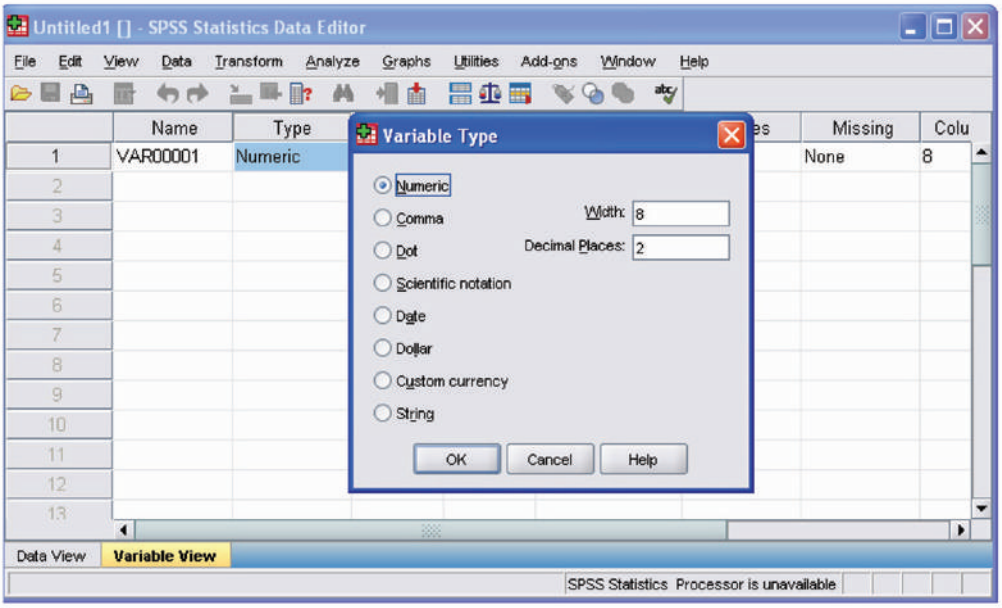
جدول (2-1)

نوع المتغير	مثال
Numeric	1000.36
Comma	1،000.005
Scientific	1*e3
Dollar	$1،000.00
String	Mervat

ونلاحظ أن البرنامج يضع بشكل تلقائي نوع المتغير Numeric.

لتغيير نوع البيانات إلى رقمي نتبع الخطوات الآتية:

مـن نافـذة Variable view نختار العمود الثاني الذي يحمل عنـوان Type، حيث نلاحظ أن بجوار نوع المتغير توجد نقاط على شكل "..." نقوم بالنقر عليها فيظهر الشكل (2-2) الآتي:

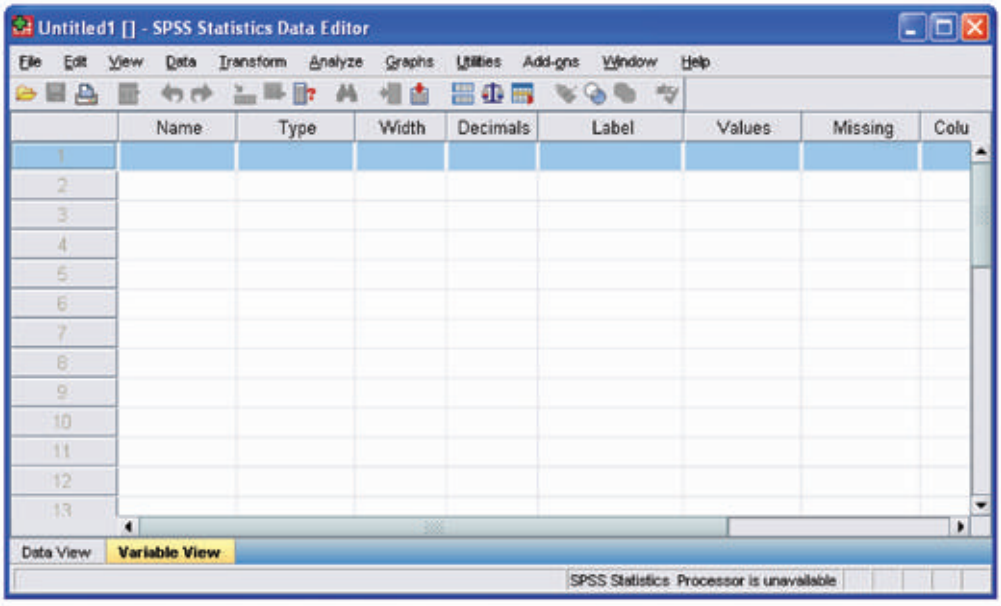


شكل (2-2)

نختـار Numeric وهـي أول خيار، ونستطيع من ذلك تحديد عرض عمـود البيانات لهذا المتغير بحيث يظهر عدد من خانات الأرقام أكبر، وهنا نختار Width 8 وعدد الخانات العشرية 2 يقوم البرنامج بتقريب الخانـات العشريـة إلى أقرب رقمـين عشريين فقط . وبعد ذلـك ننقر على Ok لتأكيـد الاختيار، ونلاحظ أننـا بعد النقر على Ok تتغير قيمـة العمود الثالث الذي يحمل عنوان Width إلى 8، وهو عدد خانات الرقم الأساسية، وتتغير أيضًا قيمة العمود الرابع الذي يحمل عنوان Decimals إلى 2 وهو عدد الخانات العشرية المتاحة.

☜ عرض المتغير (Width):

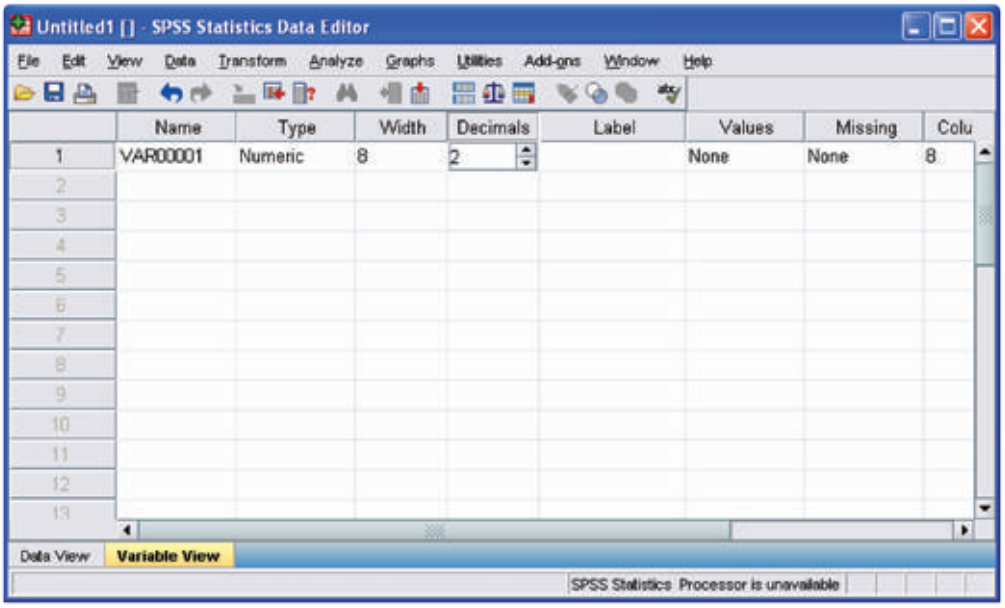
العنصـر الثالـث من عناصر تعريف المتغير هو العمود الثالث من نافذة Variable View، حيث يمكننا مـن تغيير عرض المتغير مباشرة عندما نقوم بتعريف نوع المتغير أو بالنقر مباشرة على الخلية الموجودة في عمود عرض المتغير، وصف المتغير المراد تغيير عرضه فيظهر سهم إلى أعلى وسهم إلى أسفل، وعند النقر علـى السهـم العلوي يقوم بزيـادة عرض المتغير مقدار وحـدة في كل مرة ضغط، وهكـذا بالنسبة للأسهم السفلى، حيث يقوم بتخفيض عرض المتغير بمقدار وحدة واحدة في كل مرة ضغط، والشكل (2-3) يوضح هذه العملية:



شكل (2-3)

☜ عدد الخانات العشرية للمتغير (Decimals):

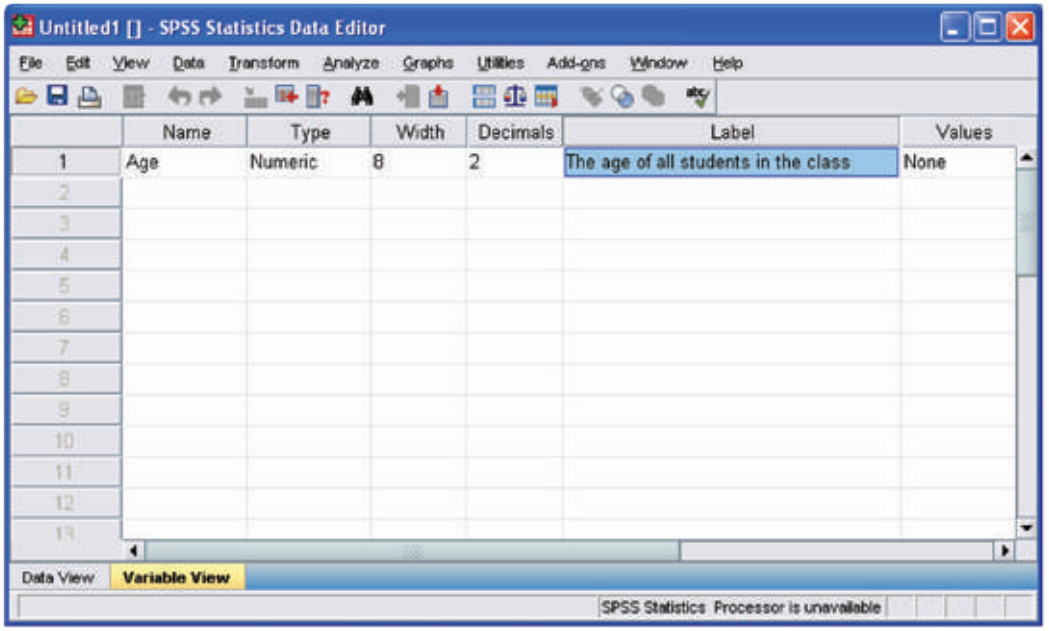
العنصر الرابع من عناصر تعريف المتغير هو العمود الرابع من Variable View، حيث يمكننا من تغيير عدد الخانات العشرية للمتغير مباشرة عندما نقوم بتعريف نوع المتغير أو بالنقر مباشرة على الخلية الموجودة في عمود عدد الخانات العشرية للمتغير (Decimals) وصف المتغير المراد تغيير عدد خانته العشرية فيظهر سهم إلى أعلى وسهم إلى أسفل، وعند النقر على السهم العلوي يقوم بزيادة عدد الخانات العشرية للمتغير بمقدار وحدة في كل مرة نقر، وهكذا بالنسبة للسهم السفلي حيث يقوم بتخفيض عدد الخانات العشرية له أيضًا بمقدار وحدة واحدة في كل مرة نقر، والشكل (2-4) يوضح هذه العملية:



شكل (2-4)

☜ مميز المتغير (Label):

العنصـر الخامس من عناصر تعريف المتغير حيث يمكن أن يأخذ المتغير عنوانا يصل عدد رموزه إلى 256 رمزًا يستعمل لوصف المتغير مع إمكانية وضع فواصل ورموز خاصة كما في الشكل (2-5) الآتي: .

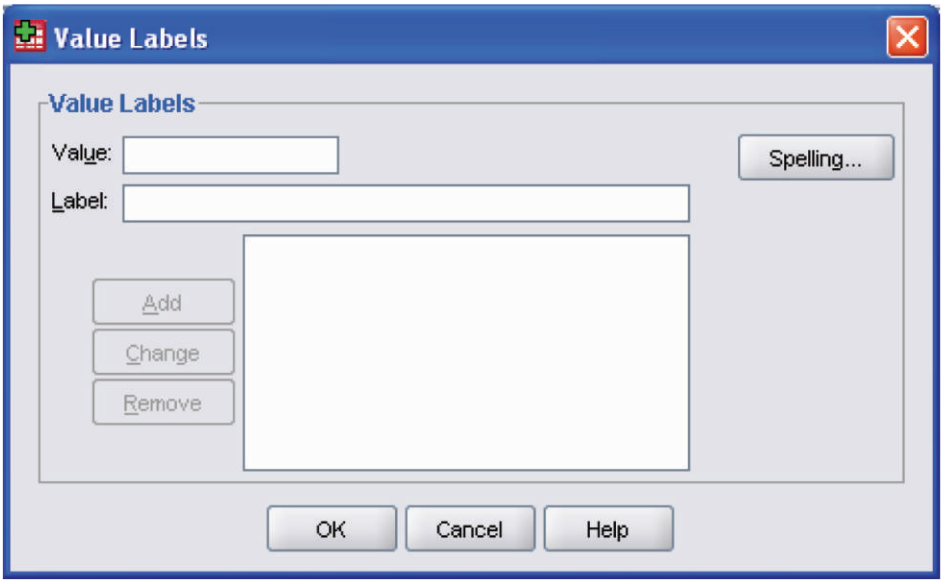


شكل (2-5)

☜ قيمة المميز (Values):

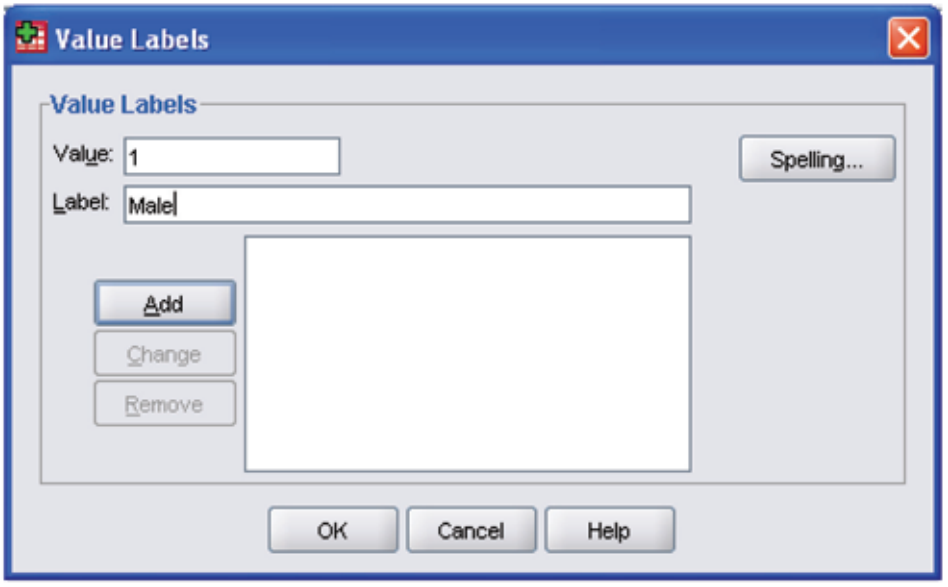
العنصر السادس من عناصر تعريف المتغير حيث نحتاج أحيانًا إلى تعيين عنوان للقيمة، كون المتغير يستعمل قيمًا عددية للتعبير عن قيم غير عددية مثلا / متغير Gender يستعمل الرقم 1 للتعبير عن الذكور، والرقم 2 للتعبير عن الإناث، ويتم تنفيذ ذلك بالبرنامج باتباع الخطوات الآتية:

مـن نافذة Variable view نختار الخلية الموجودة في العمود السادس الذي يحمل عنوان "Value"، ثم نقر النقاط التي تحمل اللون الرمادي فيظهر مربع الحوار الآتي:



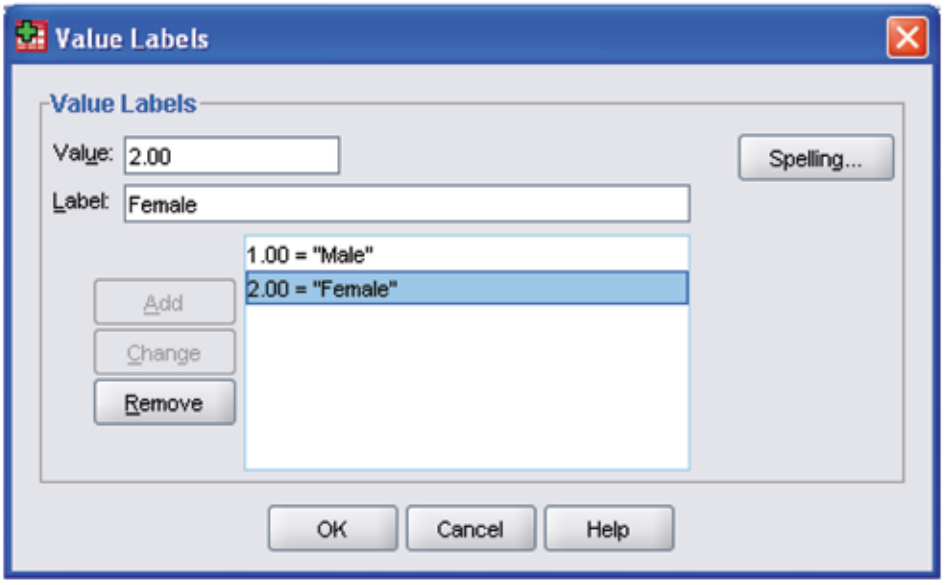
شكل (2-6)

لإضافة الكود "1" للذكر و"2" للأنثى فإننا نكتب أمام الخانة Value القيمة "1"، ونكتب أمام الخانة Label كلمة "ذكر" فتظهر الشاشة الآتية:



شكل (2-7)

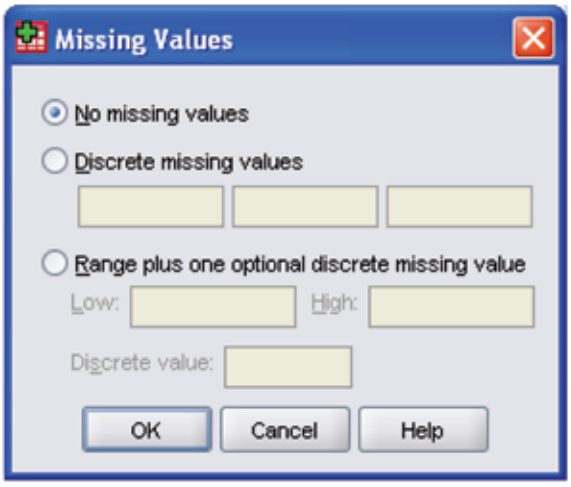
ثم ننقر Add وذلك لإضافة الكود الخاص بالأنثى ثم نكرر الخطوة السابقة ونلاحظ أننا يمكننا حذف أي كود لا نرغب فيه فقط ننقر على الكود المراد حذفه وبعدها النقر على Remove كما هو موضح في الشكل الآتي:



شكل (2-8)

☜ القيم المفقودة (Missing Values):

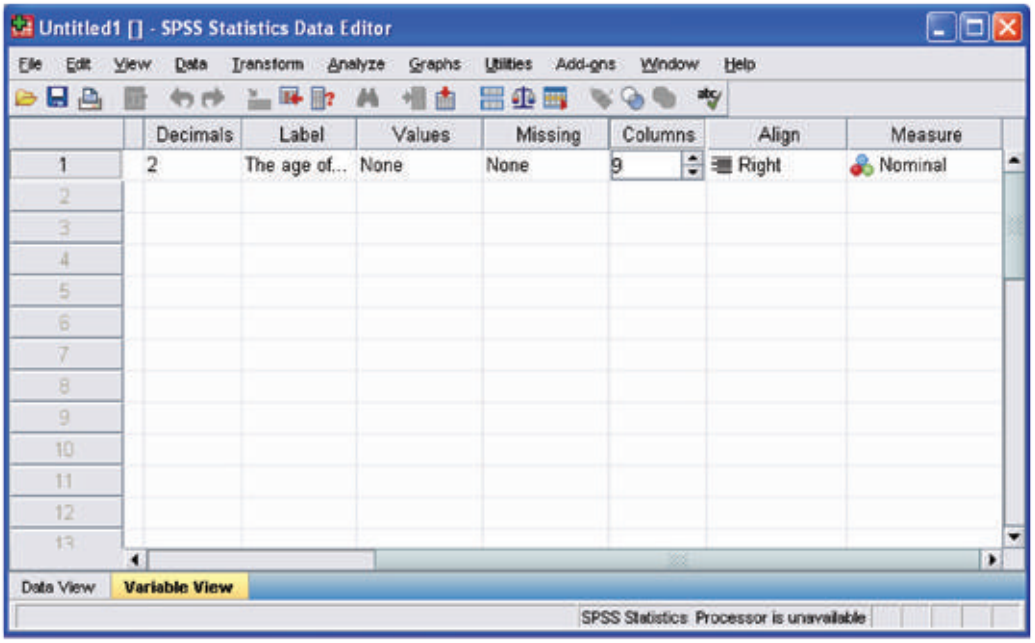
القيم المفقودة هي قيم موجودة أصلاً، ولكننا لا نرغب في إدخالها في التحليل الإحصائي كونها مثلاً قيماً شاذة أو أن نوع السؤال لا ينطبق على المستجيب.



شكل (2-9)

☜ عرض العمود (Column width):

يمكن تحديد عرض العمود لمتغير معين بالوقوف على الخلية الواقعة ضمن العمود المعنون Column في نافذة variable view، حيث يمكن زيادة أو تقليل عرض العمود بواسطة الأسهم إلى الأعلى أو الأسفل (أو كتابة عرض العمود مباشرة).



شكل (2-10)

☜ محاذاة النص (Alignment):

لضبط محاذاة النص داخل خلايا المتغير انقر الخلية التابعة لمتغير معين في ورقة variable view الواقعة ضمن العمود المعنون Align، ثم انقر السهم المتجه للأسفل لاختيار أمر مما يأتي:

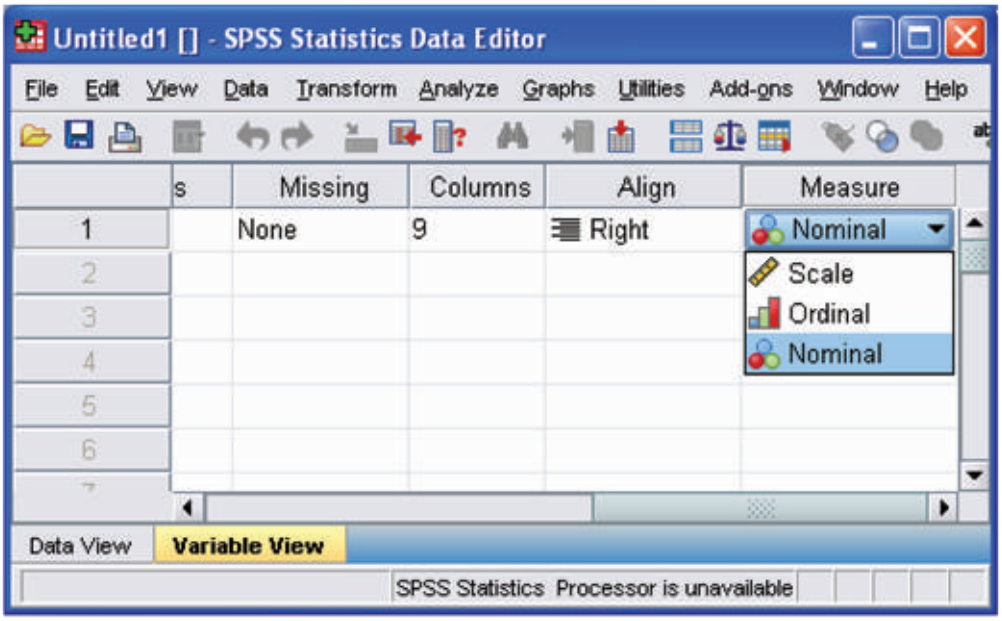
- **Left**: لمحاذاة النص إلى يسار الخلية.
- **Center**: لمحاذاة النص إلى وسط الخلية.
- **Right**: لمحاذاة النص إلى يمين الخلية.

علماً بأن المحاذاة الافتراضية هي (Right).

☜ أنواع وحدات القياس (Measures Types):

- **Nominal** : يستعمل لقياس المتغيرات الاسمية، وهي متغيرات لها عدد من الفئات دون أفضلية لإحداها على الأخرى، ولذلك لا يمكن ترتيبها تصاعديًا أو تنازليًا، كما لا يمكن إجراء العمليات الحسابية على هذا النوع من المتغيرات مثل / تقسيم المجتمع إلى ذكور وإناث.
- **Ordinal**: يستخدم لقياس المتغيرات الترتيبية، حيث إن هذا المتغير ذو عدد محدد من الفئات يمكن ترتيبها تصاعديًا أو تنازلياً، ولكن لا يمكن تحديد الفروق بينها بدقة. مثلاً تقدير طالب في الامتحان قد يكون "ممتاز، جيد جدًا، جيد ... إلخ"، ويمكن أن يكون المتغير رمزيًا أو عددياً على أنه يفضل الأخير (عددي).
- **Scale**: ويستخدم لقياس المتغيرات الكمية في فترة أو نسبة كالدخل والعمر و الأسعار ... إلخ.

ولغرض تعريف مقياس لمتغير معين انقر خلية المتغير التي تقع ضمن عمود measure في نافذة variable view حيث ستظهر ثلاثة خيارات nominal – ordinal – scale ونختار نوع المتغير كما يأتي:



شكل (2-11)

(2-3) **إدخال البيانات** (Entering Data):

الآن وصلنا إلى مرحلة إدخال البيانات استعداداً لإجراء التحليل الإحصائي.

يتم إدخال البيانات كما يأتي:

- نختار الخلية المراد إدخال قيمة المتغير فيها.
- نكتب القيمة المراد إدخالها في الخلية في مكان اسمه محرر الخلية Cell Editor موجود تحت شريط الأدوات.
- عند الانتقال للخلية الآتية بالأسهم أو النقر على الأمر Enter تكتب القيمة المدخلة تلقائياً.
- عند حدوث أي أخطاء في عملية الإدخال يمكن إجراء التصحيح كالآتي:

  ننتقل إلى الخلية المراد تصحيحها ويتم النقر عليها.

  نكتب القيمة الصحيحة في محرر الخلية.

ملاحظة: ممنوع إدخال تعبيرات حسابية (جمع / ضرب / ...).

مثال (2-1):

الجدول الآتي يمثل بعض المعلومات العامة عن مجموعة معينة من الأشخاص.

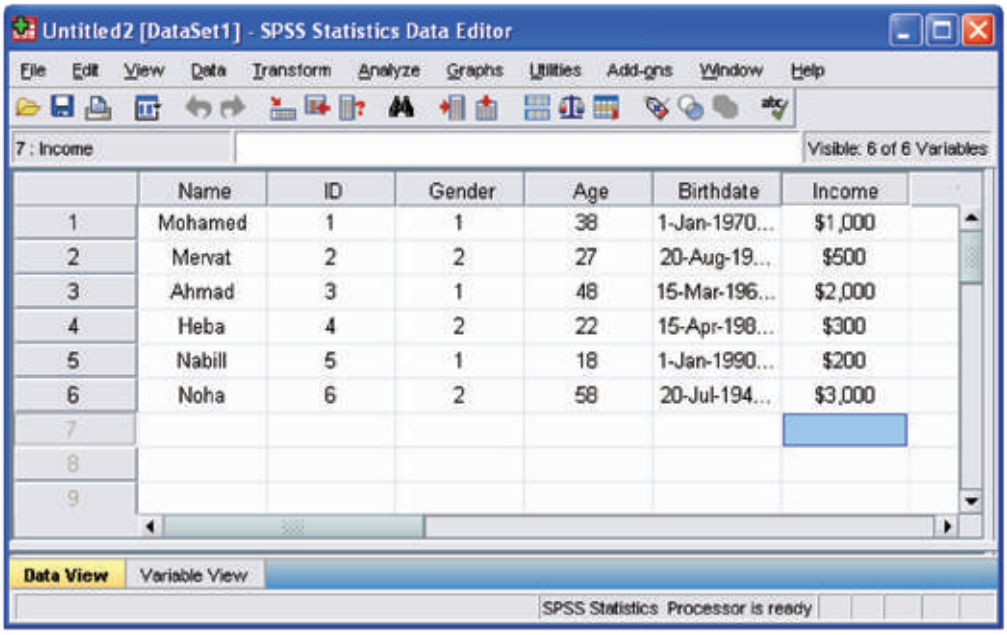
جدول (2-2)

Name	ID	Gender	Age	Birth date	Income
Mohamed	1	1	38	1/1/1970	1000
Mervat	2	2	27	20/8/1980	500
Ahmad	3	1	48	15/3/1960	2000
Heba	4	2	22	15/4/1985	300
Nabil	5	1	18	1/1/1990	200
Noha	6	2	58	20/7/1949	3000

كل عمود في المثال السابق يمثل متغيرًا من المتغيرات التي يمكن تقسيمها كما يأتي:

- المتغير الأول: متغير الاسم ID وهو متغير رمزي.
- المتغير الثاني: متغير الجنس Gender (الرقم 1 يمثل الذكور والرقم 2 يمثل الإناث).
- المتغير الثالث: متغير العمر Age وهو متغير رقمي.
- المتغير الرابع: متغير تاريخ الميلاد Birth date وهو متغير تاريخ .
- المتغير الخامس: متغير مستوى الدخل Income وهو متغير رقمي.

الشكل الآتي هو شكل نافذة الـ Data view بعد إدخال بيانات جدول (2-2):



شكل (2-12)

(4-2) عمليات مختلفة على البيانات:

☜ تحديد المتغيرات والحالات:

- ⊙ لاختيار (تحديد أو تظليل) متغير ما variable انقر الخلية التي تحتوي على اسم المتغير في أعلى العمود في ورقة Data view بزر الماوس الأيسر.
- ⊙ لاختيار حالة case بأكملها انقر الخلية الحاوية على رقم الحالة في ورقة Data view بزر الماوس الأيسر.
- ⊙ لاختيار مجموعة من المتغيرات المتجاورة: انقر الخلية الحاوية اسم المتغير الأول.
- ⊙ النقر على مفتاح shift.
- ⊙ انقر الخلية الحاوية على اسم المتغير الأخير مع استمرار النقر على مفتاح shift.
- ⊙ لاختيار مجموعة من المتغيرات المتباعدة انقر بزر الماوس الأيسر على الخلية الحاوية اسم المتغير الأول لاختياره.
- ⊙ انقر على مفتاح Ctrl مع استمرار النقر عليه (انقر الخلية الحاوية اسم المتغير الثاني لاختيارها، وهكذا بالنسبة لباقي المتغيرات الآتية).
- ⊙ بالطريقة نفسها المستخدمة لاختيار مجموعة من المتغيرات المتجاورة أو غير المتجاورة يمكن استعمالها لاختيار مجموعة من الحالات Cases المتجاورة أو غير المتجاورة.

☜ إدخال (حشر) حالة جديدة (Insert a New Case):

نستطيع حشر حالة بين حالتين موجودتين كالآتي:

- ⊙ ننتقل إلى (الصف) أسفل المكان المراد إضافة الحالة فيه ويتم النقر.
- ⊙ من القائمة Data ننقر على الأمر Insert Case فتظهر حالة جديدة خالية.
- ⊙ ندخل البيانات المراد إدخالها إلى الخلايا عن طريق محرر الخلية.

☜ إدخال (حشر) متغير جديد (Insert new variable):

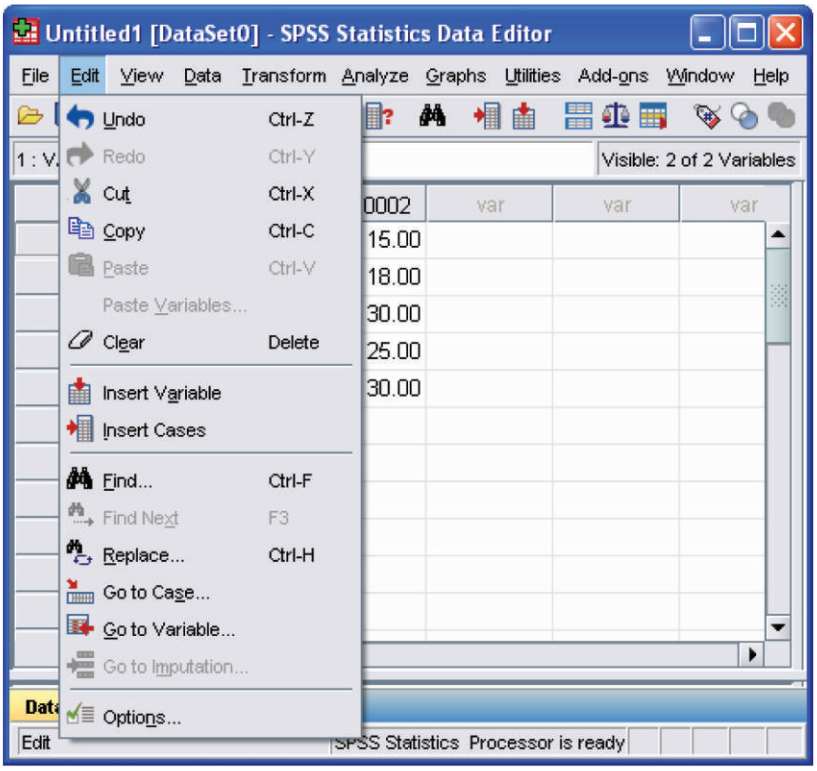
بالوقوف على المتغير المراد إضافته على يسار المتغير ثم النقر بزر الماوس الأيمن واختيار insert variable وتسمية المتغير بعد ذلك في شاشة variable view بالطريقة السابقة نفسها.

☜ حذف صف (Delete Row):

انقر اسم الحالة في ورقة Data View بزر الماوس الأيسر، من القائمة Edit انقر على Clear تتم الإزالة الفورية للحالة. ويمكن استخدام الأمر Delete بعد اختيار المتغير مباشرة.

☜ حذف عمود (Delete Column):

يتم حذف المتغير كالآتي: حدد المتغير المراد حذفه من القائمة Edit التي تتضمن القائمة الآتية:



شكل (2-13)

انقر على Clear ويمكن استخدام الأمر Delete.

☜ حذف صف (Delete Case):

بطريقة حذف المتغيرات نفسها يمكننا حذف الحالات غير المرغوب فيها كالآتي:

انقر اسم الحالة في ورقة Data View بزر الماوس الأيسر.

من القائمة Edit انقر على Clear تتم الإزالة الفورية للحالة. ويمكن استخدام الأمر Delete بعد اختيار المتغير مباشرة.

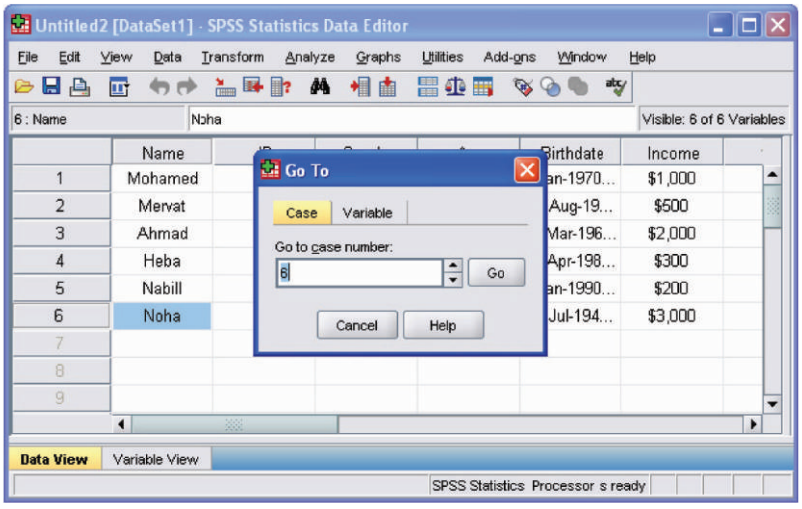
☜ نسخ وقص ولصق الحالات والمتغيرات:

لعمل نسخة من متغير معين Copy نتبع الخطوات الآتية:

- حدد اسم المتغير ثم من شريط القوائم نختار Edit، ومن ثم اختيار Copy من القائمة المدرجة، ويمكن النقر مباشرة على Ctrl+C لإجراء عملية النسخة بشكل سريع نحدد بعدها المتغير المراد النسخ فيه.
- من شريط القوائم نختار Edit ومن ثم اختيار Paste، ويمكن النقر مباشرة على Ctrl+V لإجراء عملية اللصق بشكل سريع. وعملية القص واللصق تتم بالطريقة نفسها، كما أن عملية نسخ ولصق وقص الحالات أيضا تتم بالطريقة ذاتها مع تبديل اختيار العمود (المتغير) بالصف (الحالة).

البحث عن حالات:

- انقر الأمر Go To Case من القائمة Edit فيظهر المربع الحواري (انظر الشكل) (2 – 14) الآتي:



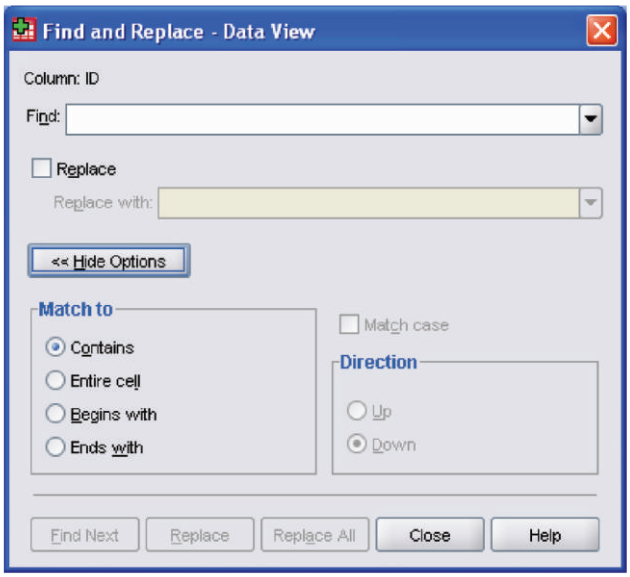
شكل (2-14)

- في المربع Case Number اكتب رقم الحالة التي تريد الانتقال إليها.
- انقر Ok فيتم الانتقال إلى الحالة المحددة.

البحث عن قيمة:

يتم كالآتي:

- ننتقل إلى قمة العمود الخاص بالمتغير.
- انقر الأمر Find من القائمة Edit فيظهر المربع الحواري الآتي (انظر الشكل 2–15):

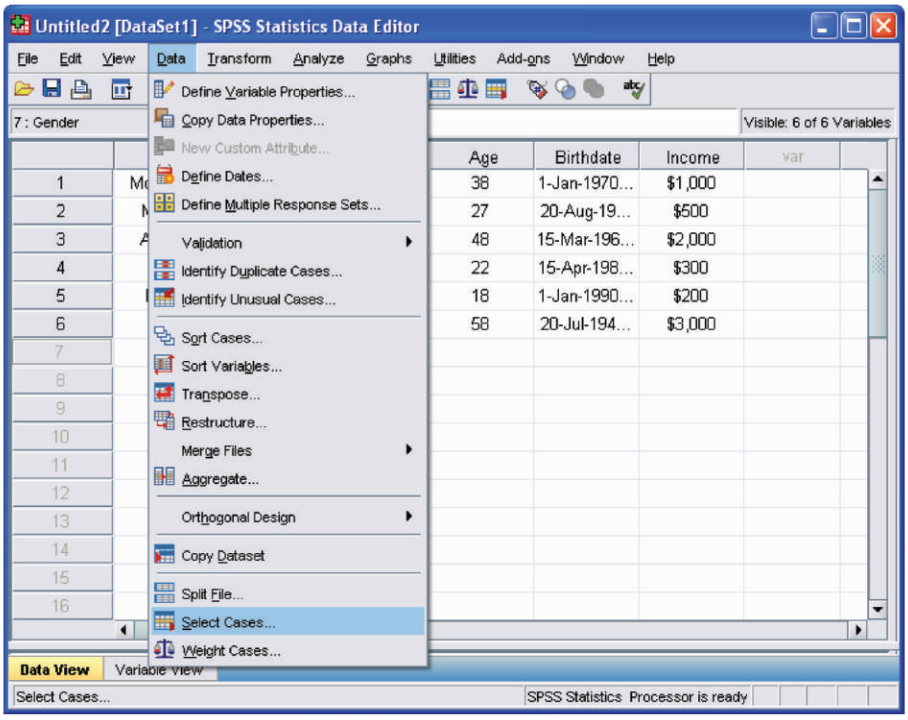


شكل (2-15)

⊙ في المربع الحواري حدد الرقم المطلوب البحث عنه في المربع Find، ونلاحظ أننا إذا أردنا البحث عن القيمة "2 " فإننا نكتب القيمة "2 " في خانة Find وبعدها النقر على Find next فنلاحظ أن البرنامج يقف على أول حالة أو متغير لديه القيمة "2 "ويحددها كما في الشكل السابق.

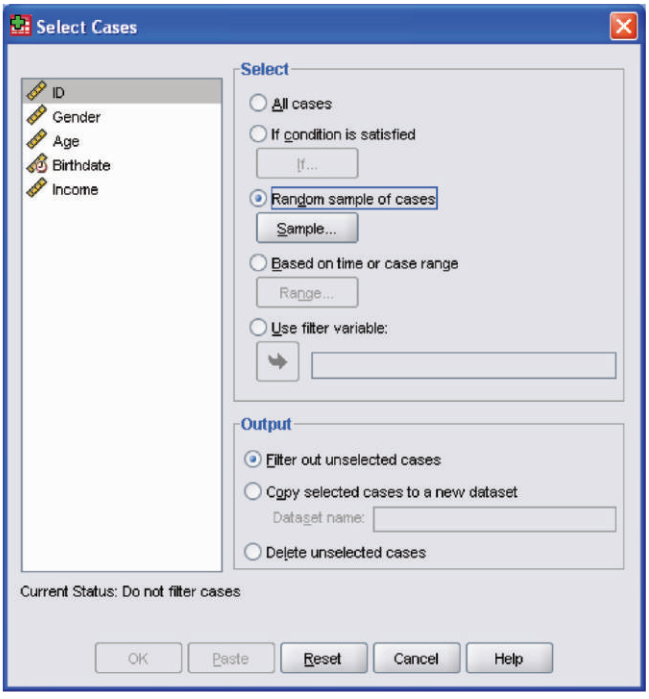
اختزال (تخفيض) حجم البيانات:

بفرض أننا لدينا مليونا بيان تم الحصول عليها من المشاهدات، ولتخفيض حجم هذه البيانات نقوم باختيار عينة عشوائية منها لإجراء التحليل الإحصائي عليها. نفرض أنه تم اختيار مائة ألف مشاهدة بطريقة العينة العشوائية البسيطة، ولضمان عدم وجود تحيز في الاختيار نقوم باختيار Select Cases من القائمة Data كما في الشكل الآتي:



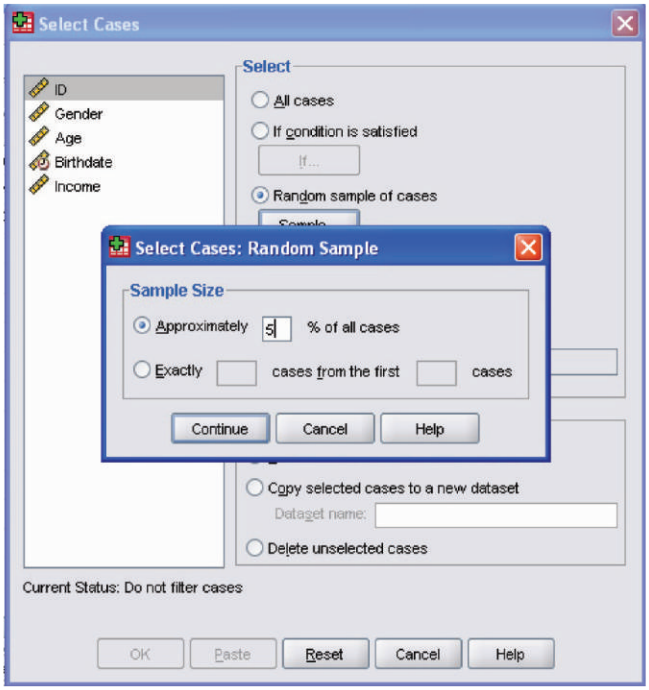
شكل (2-16)

بعدها تظهر الشاشة الآتية:



شكل (2-17)

فنقــوم باختيــار Random Sample of Cases والنقر بعدها على مربع الحوار Sample، ولأننا نختار مائة ألف من مليوني بيان فإننا نحدد حجم العينة بتقريب بنسبة 5٪ من البيانات، ومن ثم تظهر الشاشة الآتية:



شكل (2-18)

ترتيب البيانات:

نقصد بعملية ترتيب البيانات أن نرتب بيانات أحد المتغيرات (أو أكثر من متغير) في الملف، ثم نرتب بقية المتغيرات تبعا لذلك. لترتيب بيانات ملف نذكر بعض الملاحظات المهمة:

- أن عملية ترتيب بيانات الملف تتم بالنسبة لمتغير أو أكثر وليست مطلقة.
- أن عملية الترتيب تعني نقل الحالات الموجودة في الملف إلى أعلى أو إلى أسفل.
- لا يجوز نقل قيمة معينة (أو قيم) من حالة من مكانها إلى مكان آخر بل يجب نقل الحالة كلها.

مثال (2-2):

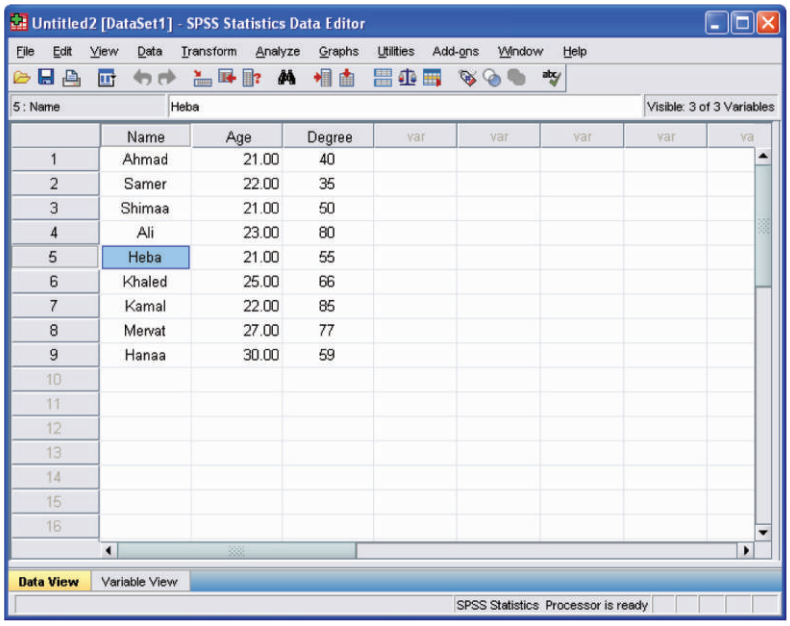
لدينا بيانات الملف الآتي لخريجي إحدى كليات العلوم:

جدول (2-2)

Name	Age	Total degree
Ahmad	21	40
Samer	22	35
Shimaa	21	50
Ali	23	80
Heba	21	55
Khaled	25	66
Kamal	22	85
Mervat	27	77
Hanaa	30	59

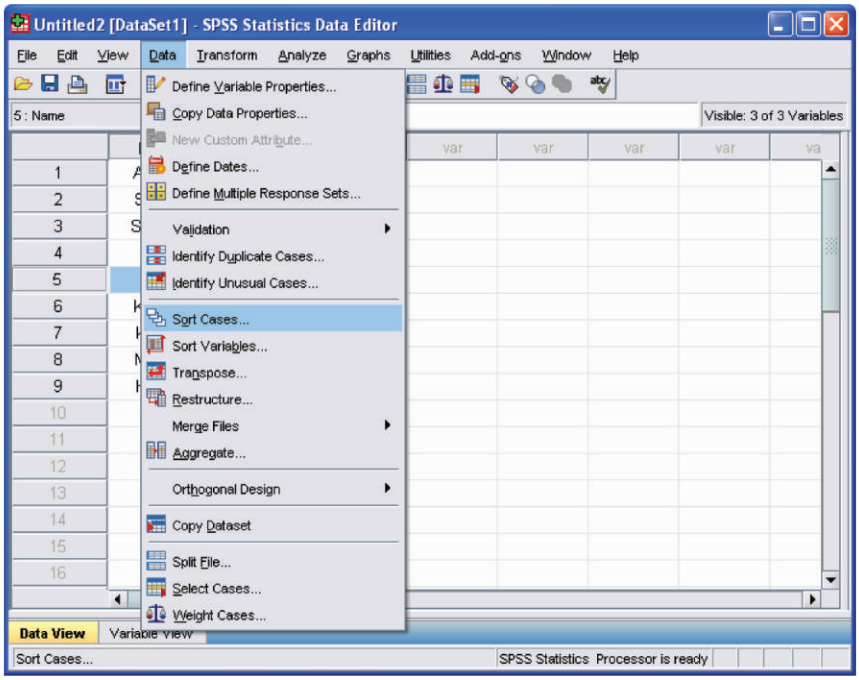
لتنفيذ ترتيب البيانات باستخدام البرنامج نتبع الخطوات الآتية:

نقوم بإدخال البيانات في البرنامج فتظهر الشاشة الآتية:



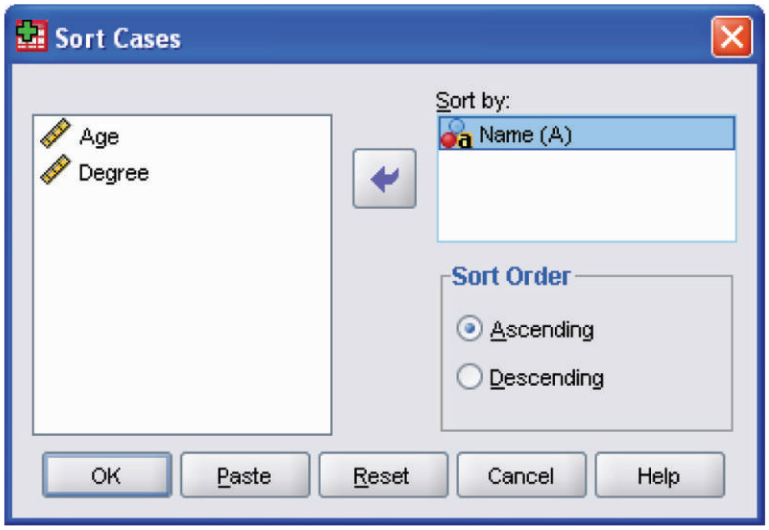
شكل (2-19)

انقر Data فتظهر الشاشة الآتية:



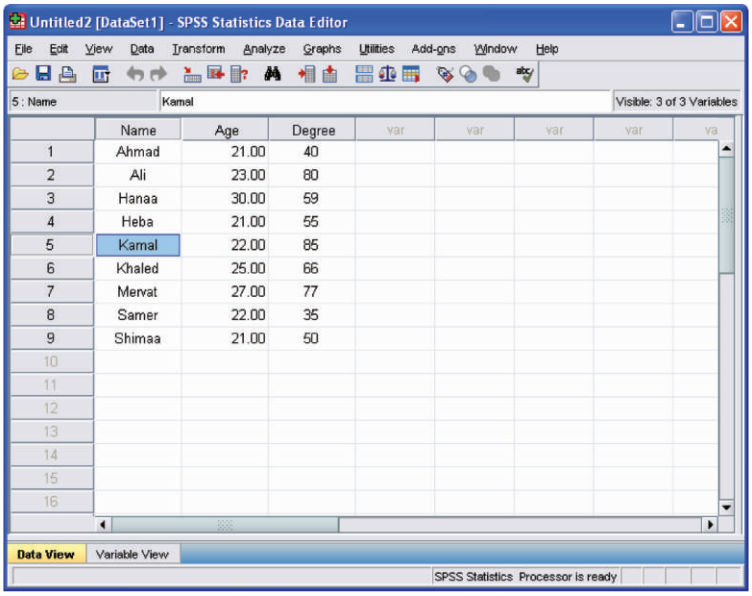
شكل (2-20)

ثم نقوم باختيار Sort Cases، يفتح الصندوق الحواري Sort Cases وبه أسماء جميع المتغيرات. نختار المتغير الذي سيتم على أساسه الترتيب وهو Name وينقل إلى المربع Sort By.



شكل(2-21)

نختـار نوع الترتيب من Sort Order هل الترتيب سيكون تصاعديًا Ascending أو تنازليًا Descending ثم ننقر على OK فتظهر الشاشة الآتية:



شكل (2-22)

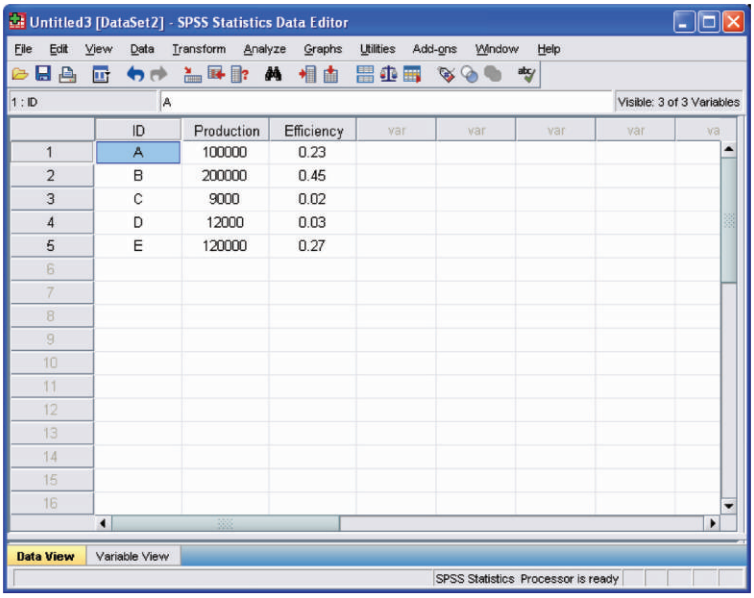
نلاحظ بعد تنفيذ الأمر السابق أن البرنامج سيقوم بترتيب البيانات تصاعديًا.

☜ تلخيص الحالات:

يستعمـل هذا الأمر لتلخيص المعلومات المتعلقـة بمجموعة من الحالات cases في حالة تجميعية واحدة وتكوين ملف تجميعي جديد، ويمكن أن نجري تلخيصًا للحالات باستخدام البرنامج كما في المثال الآتي:

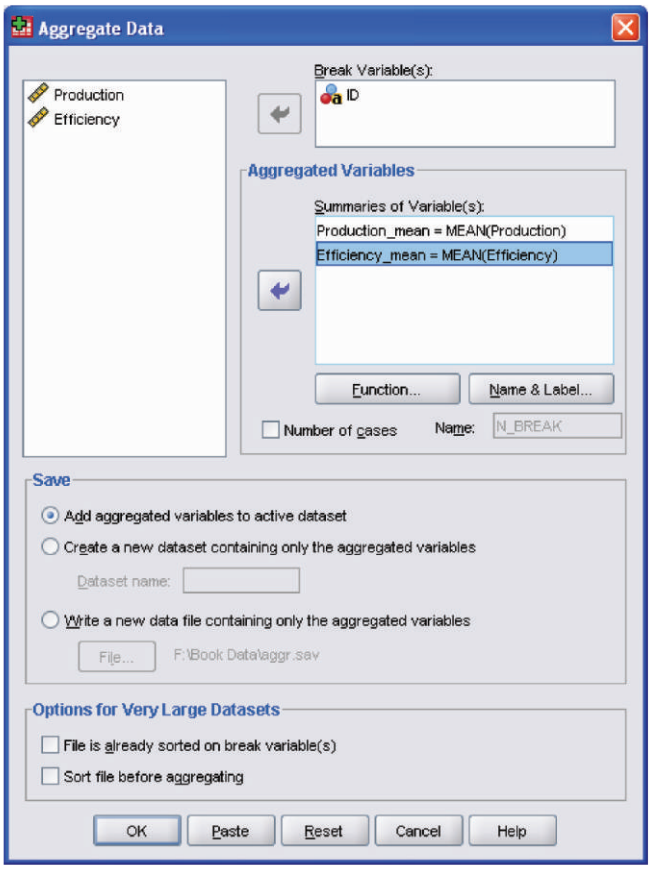
مثال (2-3):

الملف الآتي يوضح بيانًا بإنتاجية خمس قطع زراعية معينة تنتج محصول القمح ونسبة كفاءة كل قطعة زراعية كما يأتي:



شكل (2-23)

من القائمة Data انقر الأمر Aggregate يظهر الصندوق Aggregate Data الموضح في الشكل الآتي:

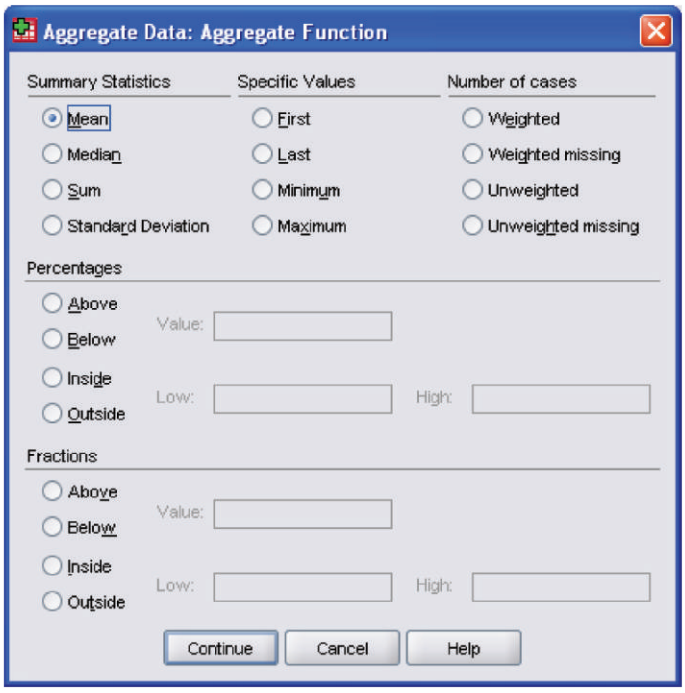


شكل (2-24)

- من قائمة المتغيرات على اليسار اختر المتغير الذي سيتم على أساسه التقسيم وانقله إلى Break Variable، ويتم النقل عن طريق زر الانتقال الموجود بجوار Break Variable من ناحية اليسار الظاهر بشكل ▶، ونلاحظ أنه عندما ننقل هذا المتغير فإنه يحذف من عمود مجموعة المتغيرات.

- قم بنقل المتغير (أو المتغيرات) المطلوب إجراء العمليات الإحصائية لها بناء على متغير التقسيم إلى Summaries of variable(s). لاحظ أننا يمكننا الرجوع عن الاختيار باستخدام الزر الآتي الموجود بجوار Summaries of variable(s) من ناحية اليسار الظاهر بشكل ◀.

- انقر على الأمر Function فيظهر الصندوق Aggregate Data: Aggregate Function، وهو يحتوي على عمليات إحصائية كثيرة، ولاحظ أنه لا يحق للمستخدم أن يختار أكثر من عملية إحصائية في الوقت نفسه (انظر الشكل 2-25).



شكل (2-25)

- انقر على الأمر Continue للعودة للصندوق الأصلي.
- انقر على الأمر Ok للتنفيذ.

لاحظ أنه تم إيجاد الوسط الحسابي كما طلبنا في الأمر Aggregate Function.

☜ ترجيح حالات:

يتيح هذا الأمر إمكانية إعطاء أوزان لحالات Cases ملف معين نظرا لاختلافها من ناحية الأهمية، وهي الوظيفة نفسها التي تقوم بها الأيقونة في شريط الأدوات.

مثال (2-4):

القيم الآتية تمثل نتائج أحد الطلاب في اختبار مادة الإحصاء علما بأن لكل اختبار وزنًا (أهمية) مختلفًا:

جدول (2-3)

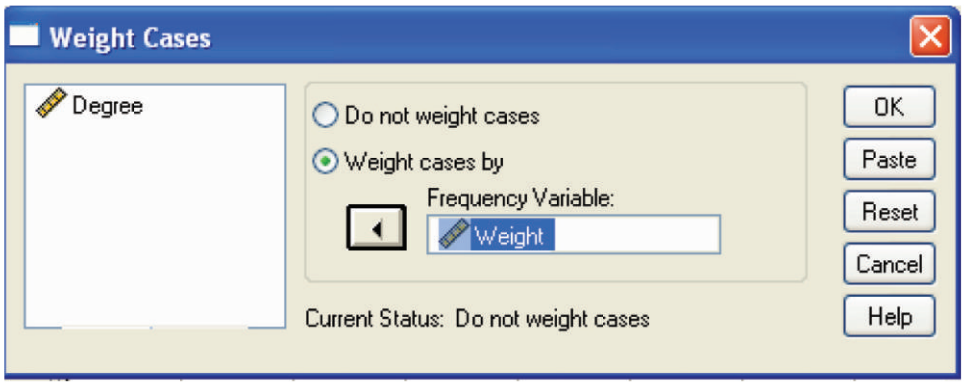
Test	Degree	Weight
A	100	50
B	60	35
C	75	5
D	55	15

المطلوب حساب الوسط الحسابي المرجح للامتحانات الأربعة؟

الحل:

لحساب الوسط الحسابي المرجح بالمتغير Weight للامتحانات الأربعة نتبع الخطوات الآتية:

- من شريط القوائم Data نختار Weight Cases الذي نقوم بترتيبه (الشكل 2-26):



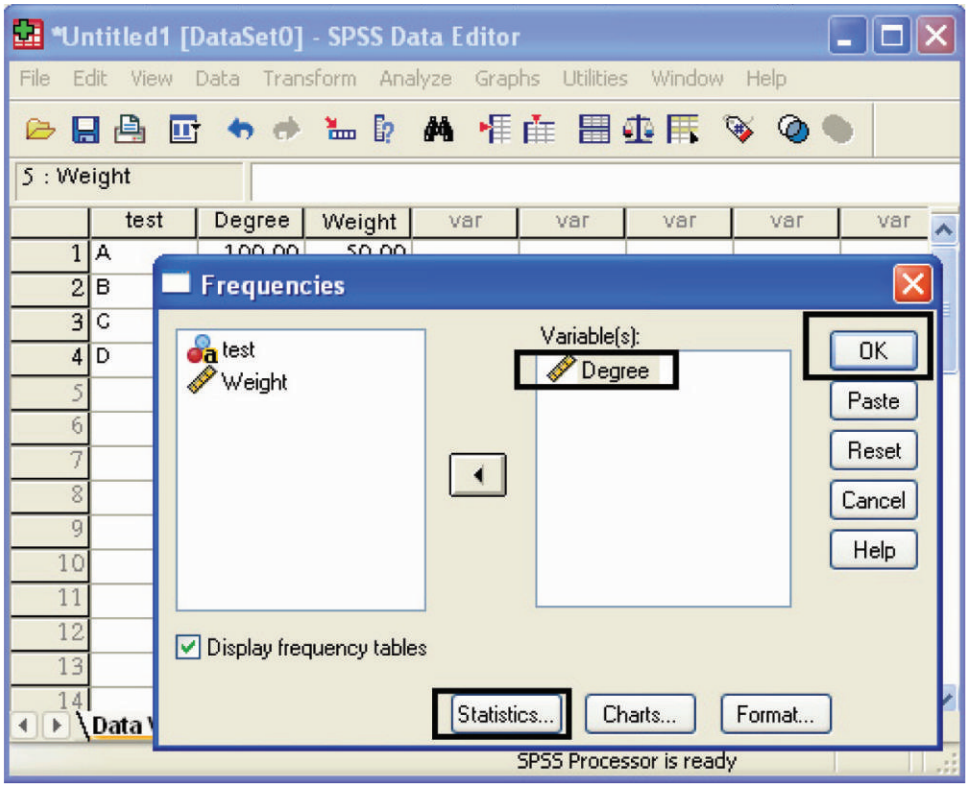
شكل (2-26)

- عند نقر زر Ok يتم وزن حالات الملف بالمتغير Weight (علما بأنه لا يلاحظ أي تغير في شاشة Data Editor).

لحساب الوسط الحسابي نختار من شريط القوائم:

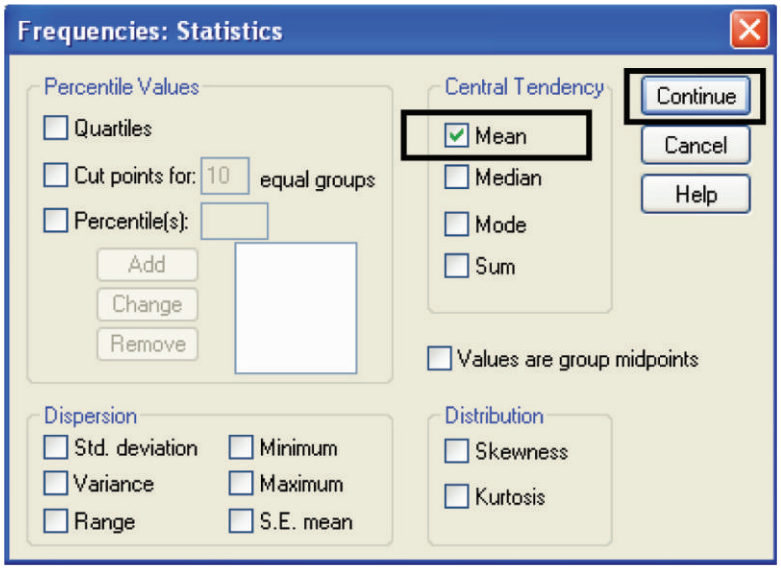
Analyze ⟵ Descriptive Statistics ⟵ Frequencies

انظر الشكل (2-27):



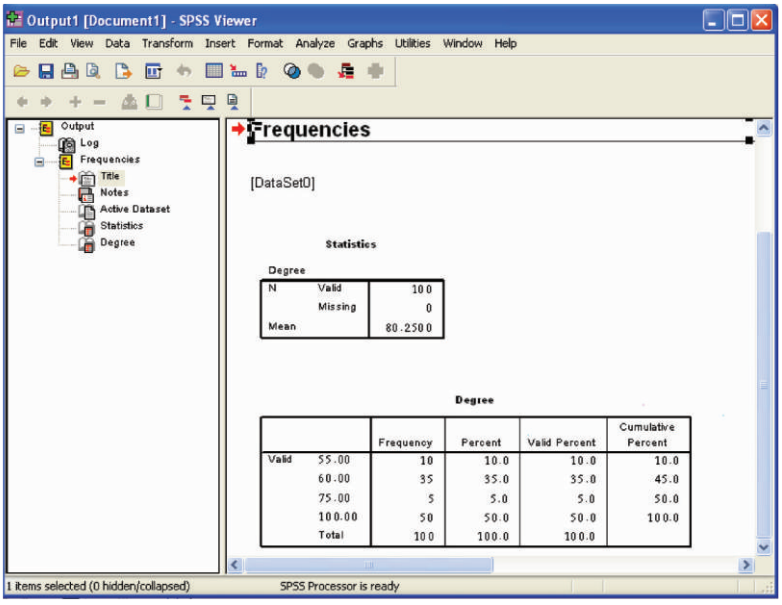
شكل (2-27)

⊙ نقـوم بتحديـد المتغير المراد حسـاب الوسط الحسابي المرجـح له، ويتم نقله عن طريـق زر النقل ▶، ويمكننـا الرجوع عن الاختيار باستخـدام رز الرجوع ◀ وباختيار Statistics نقوم بتأشير الخيار Mean كما في الشكل الآتي:



شكل (2-28)

⊙ وبنقر Continue ثم Ok تظهر شاشة المخرجات الآتية بها الوسط الحسابي المرجح:



شكل (2-29)

ملاحظة مهمة:

✓ لإلغاء ترجيح حالات الملف نقوم بتأشير الخيار Do Not Weight Cases في صندوق حوار Weight Cases.

الفصل الثالث

إنشاء متغيرات جديدة

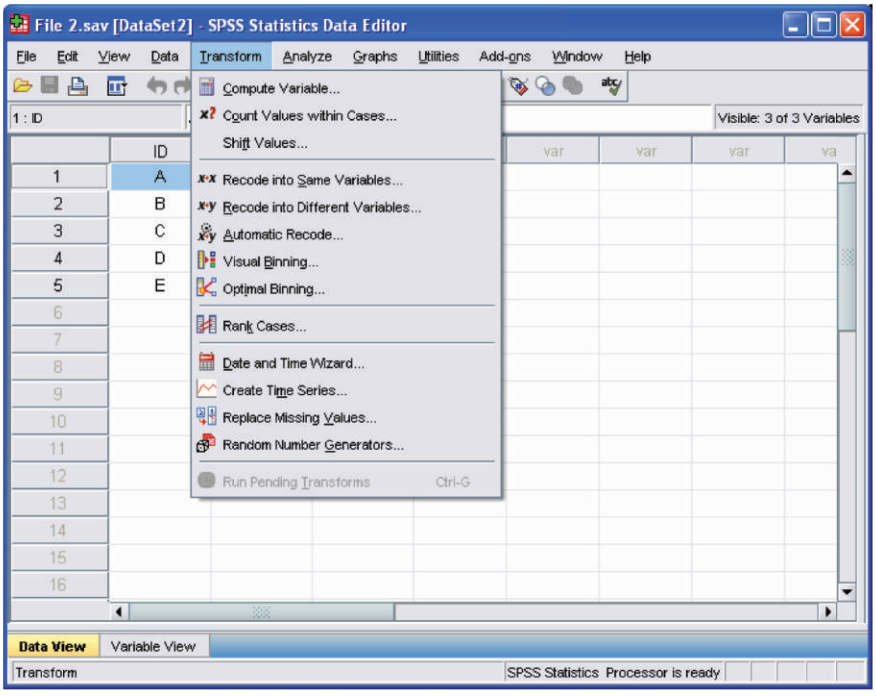
Created New Variables

(3-1) مقدمة :

يمكن إجراء تحويلات على البيانات الأصلية ما بين تجميع بيانات في فئات معينة إلى تكوين متغيرات جديدة بالاعتماد على معادلات وصيغ شرطية، فمثلا إذا كانت لدينا بيانات استهلاك ودخول أفراد مجتمع معين يمكننا حساب حجم الادخار في هذا المجتمع عن طريق إنشاء متغير جديد عبارة عن الفرق بين الدخل والاستهلاك تمهيداً لاستخدامه في حساب حجم الاستثمارات في المجتمع. كذلك البيانات التي لا تتبع التوزيع الطبيعي يمكن تحويلها إلى بيانات طبيعية. ونجد أنه في بعض الأحيان يصعب التعامل مع البيانات اللفظية، ويكون التعامل أسهل عند تكويد البيانات اللفظية وتحويلها إلى أكواد رقمية، ويتم ذلك عن طريق التحويلات.

(3-2) قائمة Transform

تحتوي هذه القائمة على الأوامر الموضحة في الشكل الآتي:



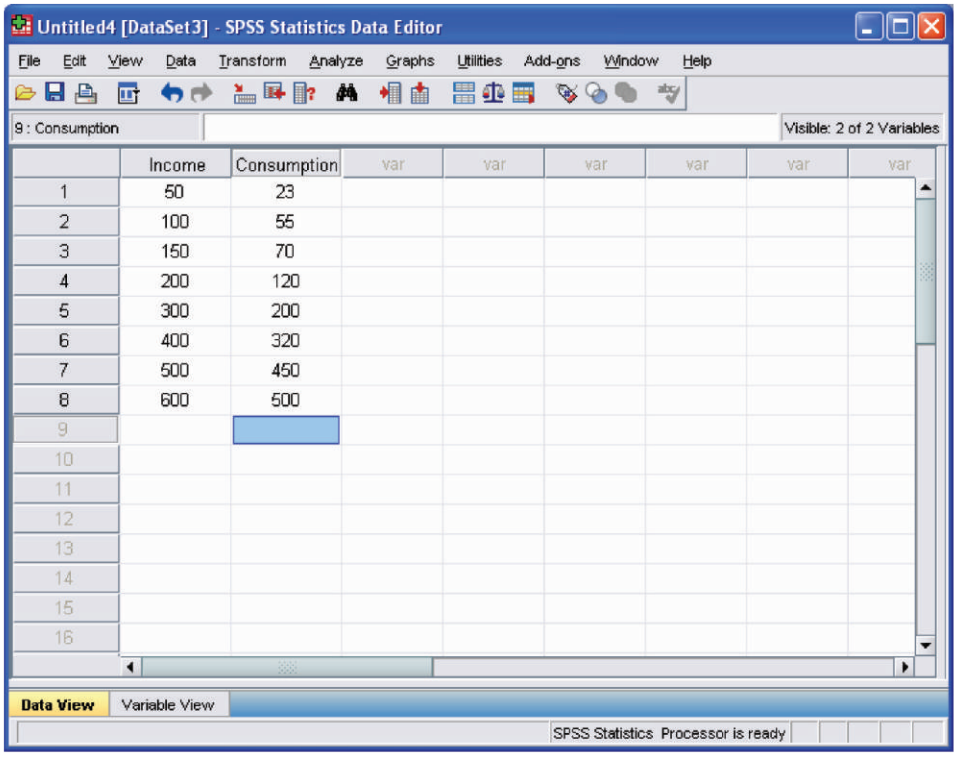
شكل (3-1)

(3-3) استخدام العمليات الرياضية لإنشاء متغيرات مستمرة جديدة / الأمر Compute:

يتيـح هذا الأمر إمكانيـة حساب متغيرات جديدة باستخدام أكثر من دالة تتضمن (دوال إحصائية، توزيعات احتمالية)، أو تحويل متغيرات موجودة إلى متغيرات أخرى.

مثال (3-1):

بفرض أننـا نريد حساب حجم الادخـار (Saving) للبيانات الموجودة التـي تمثل الدخل (Income) والاستهلاك (Consumption) في نافذة Data View الآتية:

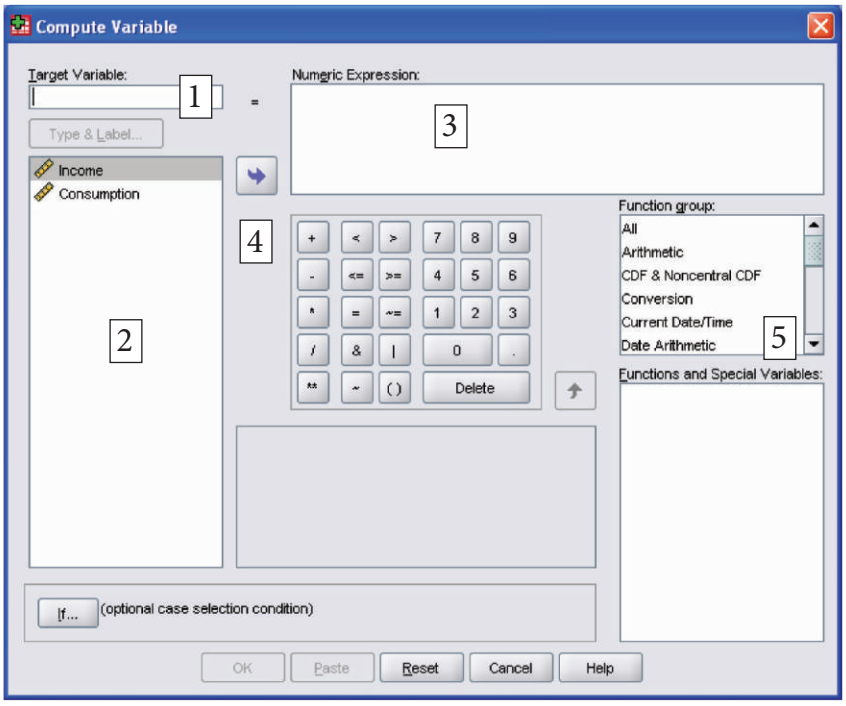


شكل (3-2)

حيث: الادخار(Saving) = الدخل (Income) – الاستهلاك (Consumption)

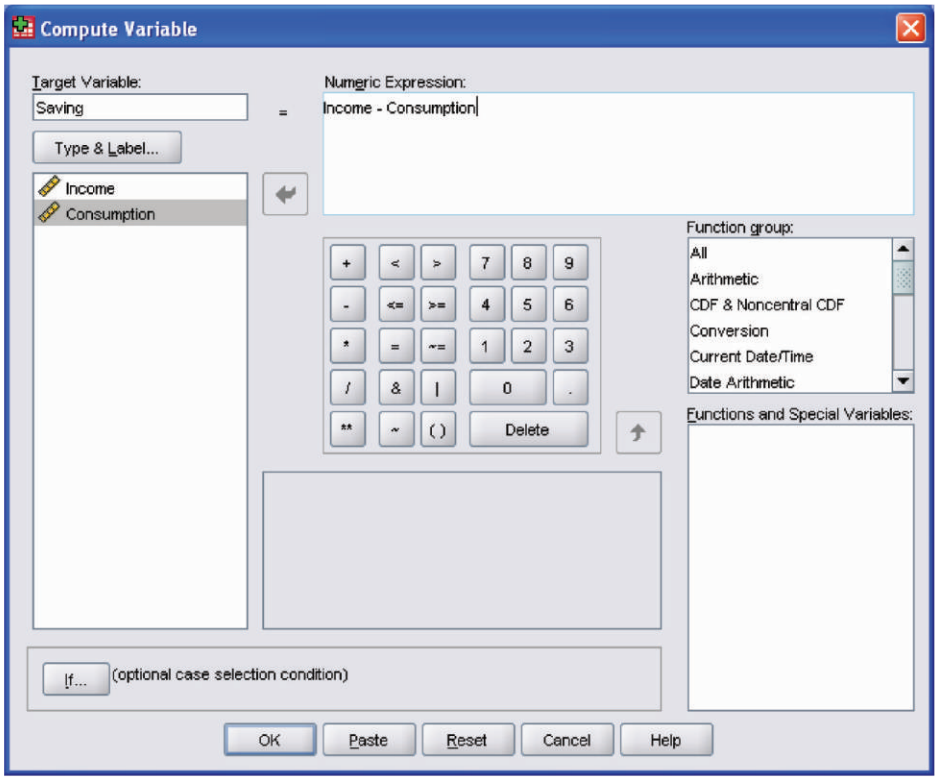
للحصـول علـى قيمة الادخـار فإننا نقوم بطـرح قيم الاستهـلاك في العمود الثاني مـن قيم الدخل بالعمود الأول عن طريق إجراء الخطوات الآتية:

من قائمة Transform نختار Compute Variable فيظهر الشكل الآتي:



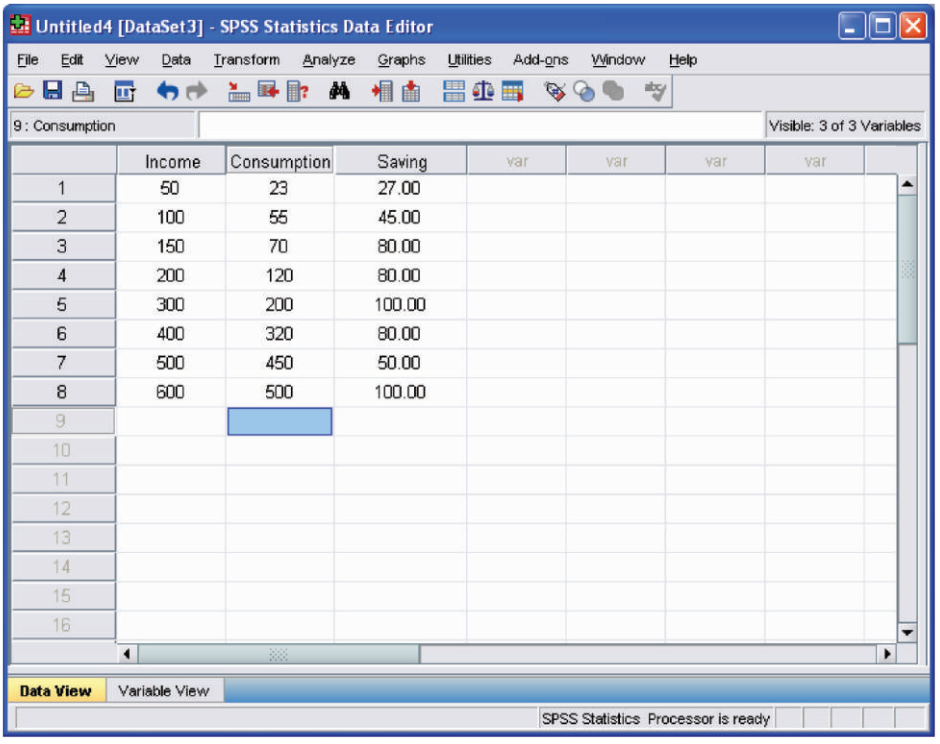
شكل (3-3)

- في المنطقة رقم "1" نضع اسم المتغير الجديد.
- في المنطقة رقم "2" تظهر المتغيرات المتاح الاختيار منها لإجراء العمليات الحسابية عليها، ونلاحظ أنه يوجد متغيران اثنان فقط هما الاستهلاك والدخل.
- في المنطقة رقم "3" وهي المنطقة التي يظهر فيها شكل العمليات الحسابية المطلوبة.
- في المنطقة رقم "4" هي منطقة تتيح لوحة مفاتيح للأرقام وبعض العمليات الحسابية البسيطة.
- في المنطقة رقم "5" هي منطقة الدول الأكثر تعقيداً.
- في هذا المثال نقوم بكتابة كلمة "Saving" في المنطقة رقم "1" ليكون اسم المتغير الجديد.
- نختار المتغير "Income" من المنطقة رقم "2" عن طريق تحديد المتغير والنقر على زر الانتقال ▶ ونلاحظ أن متغير الدخل ظهر بعد ذلك في المنطقة رقم "3" تمهيداً لإجراء العمليات الحسابية عليه.
- نختار علامة الطرح من المنطقة رقم 3، وهي بالشكل الآتي −، ونلاحظ أن بعد الاختيار تظهر تلك العلامة في المنطقة رقم "3" بعد متغير الدخل.
- نختار مرة أخرى متغير "Consumption" من المنطقة "2". فيظهر الشكل الآتي:



شكل (3-4)

ننقر على Ok فتظهر شاشة Data View كالآتي:



شكل (3-5)

☜ استخدام بعض دوال SPSS لإنشاء متغيرات جديدة:

يوفر لنا البرنامج مجموعة من الدوال ليست فقط دوال رياضية، ولكن دوال إحصائية ومنطقية أيضاً. والمثال الآتي يوضح كيف يمكننا استخدام الدوال الإحصائية في إنشاء متغيرات جديدة:

مثال (3-2):

الجدول الآتي يمثل المتغيرين X_1، X_2 تم إدخالها إلى Data Editor كما يأتي:

جدول (3-2)

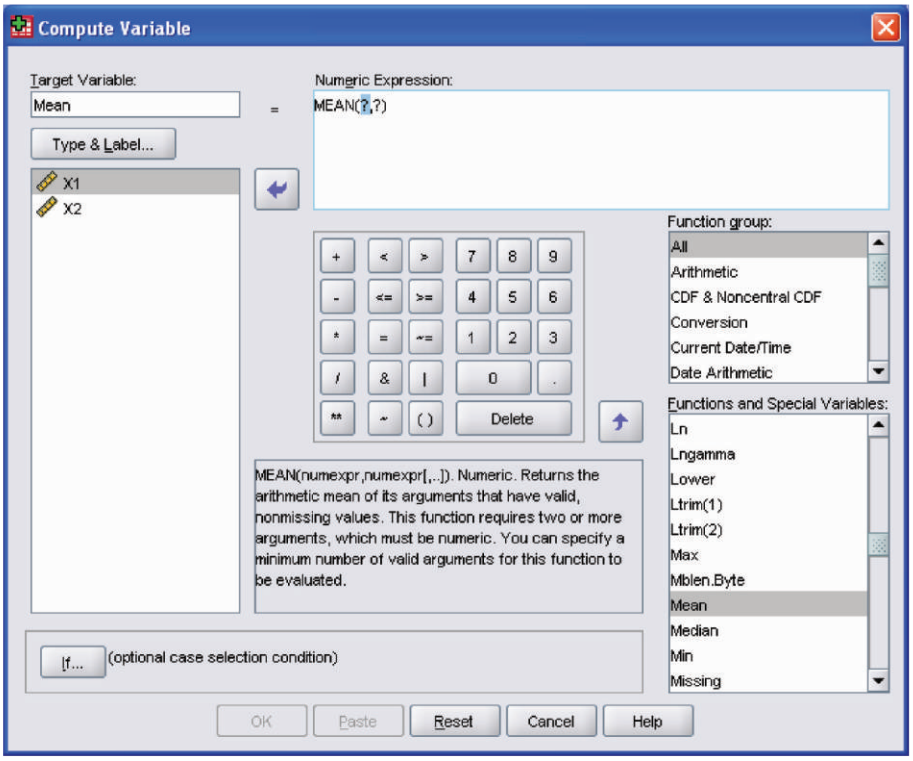
X_1	X_2
60	90
87	88
70	33
90	80
57	55
73	73
95	90
66	50
30	55
55	80
85	75
88	86
35	70

المطلوب حساب الوسط الحسابي Mean للمتغيرين X_1، X_2 عندما يكون $X_1 \geq 50$ و $X_2 \geq 50$.

الحل:

نستخدم الخطوات الآتية:

من شريط القوائم نختار Transform ثم ننقر Compute فيظهر صندوق حوار Compute Variable انظر الشكل (3-6)، ونكتب اسم المتغير الجديد، وهو الوسط الحسابي (mean) لكلا من X_1، X_2 يلي ذلك اختيار الدالة Mean من منطقة الدوال المتاحة فتظهر في منطقة العمليات الحسابية وبجوارها علامتا استفهام: العلامة الأولى متاحة لكتابة المتغير الأول، والأخرى متاحة لكتابة المتغير الثاني ويمكن كتابة أكثر من متغير مفصول بين كل متغير والآخر بفصلة:



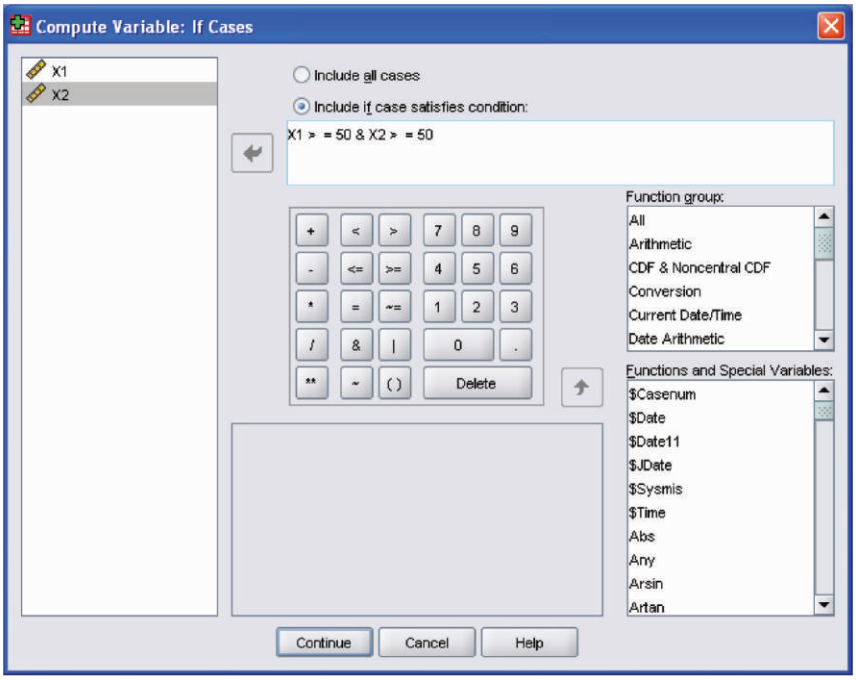
شكل (3-6)

نكتب X_1 بدلاً من علامة الاستفهام الأولى "؟" ونكتب X_2 بدلاً من علامة الاستفهام الثانية "؟".

⊙ بعد الانتهاء انقر IF في صندوق حوار شكل (3-6)، حيث لدينا شرط في المتغير الجديد، وهو أن يكون $X_1 \geq 50$ و $X_2 \geq 50$ فيظهر صندوق حوار If Cases كما بالشكل (3-7):

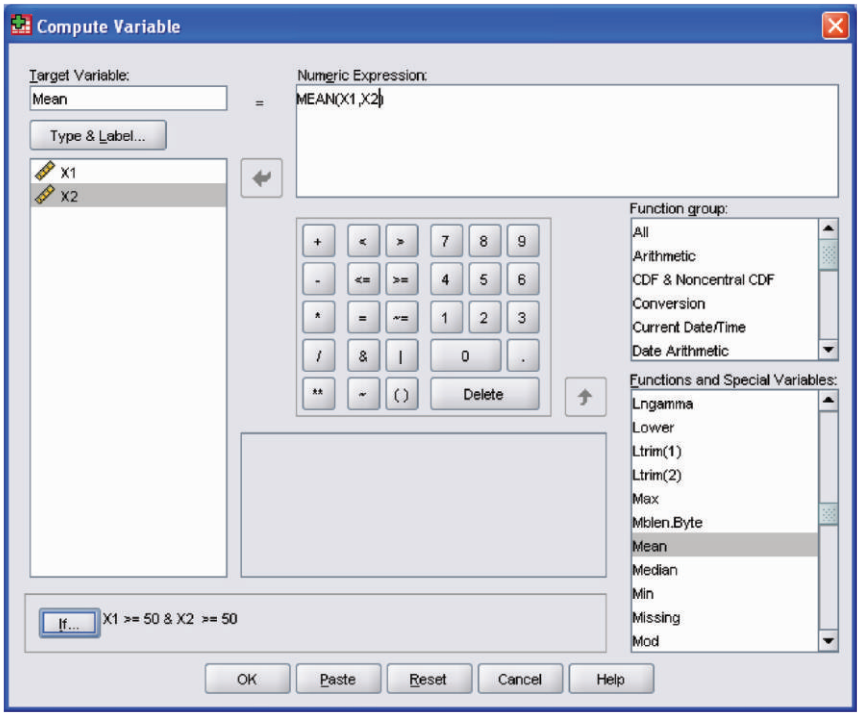
✓ لاختيار الحالات كافة انقر Include All Cases.

✓ لاختيار جزء من الحالات انقر Include If Cases Satisfies Condition. بما أننا نريد اختيار جزء من حالات نقوم بنقر الجزء الأخير. ونقوم بكتابة الشرط في شاشة الشرط كما في الشكل الآتي:



شكل (3-7)

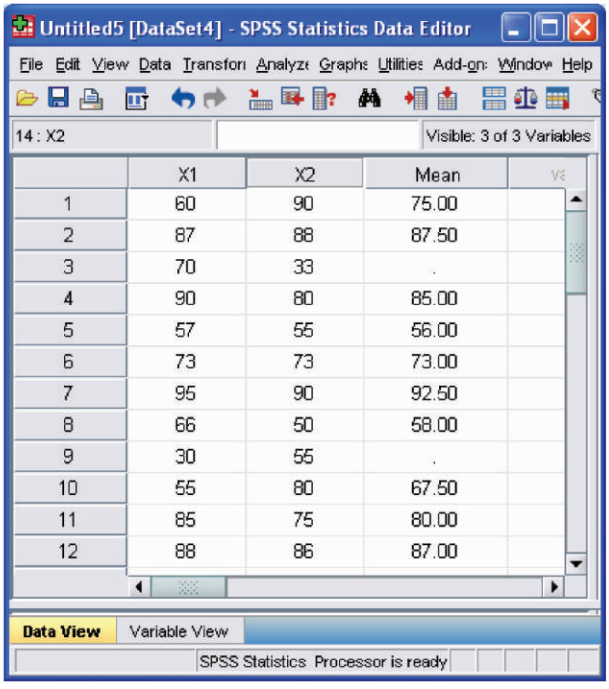
⊙ انقـر Continue في صندوق حـوار if cases للعودة إلى صندوق الحوار الأصلي، ونلاحظ أن الشرط قد تمت كتابته بجوار أيقونة If كما في الشكل الآتي:



شكل (3-8)

⊙ انقر زر ok في صندوق حوار Compute Variable.

يتم الحصول على نتائج حساب المتوسط، حيث تتم إضافة متغير آخر هو Mean إلى Data Editor كما يأتي:



شكل (3-9)

الأمر Recode:

في كثير من الأحيان نحتاج إلى عملية التكويد (أو الترميز)، وذلك بتعديل قيم البيانات وتغييرها من بيانات لفظية إلى بيانات رقمية بغرض تسهيل التعامل معها، توجد عدة حالات يمكن فيها استخدام عملية التكويد، منها دمج بيانات كثيرة لمتغير مستمر في جدول تكراري ليسهل التعامل معها. المثال الآتي يوضح طريقه تكويد البيانات المستمرة.

مثال (3-3):

المتغير العشوائي (X) يمثل قيم الأجر لخمسة عشر موظفًا:

X	100	150	200	250	300	500	270	300	350	700	350	300	550	600	850

وقد تم إدخال قيم المتغير في شاشة Data Editor، وتم تخصيص كود لكل قيمة من قيم المتغير حسب الترتيب الآتي:

جدول (3-3)

الفئة	الكود
150 فأقل	1
300-151	2
350-301	3
600-351	3
601 فأكثر	5

نرغب في ترميز المتغير Salary (الأجر) حسب الفئات المذكورة ونحفظ الرموز في متغير مختلف.

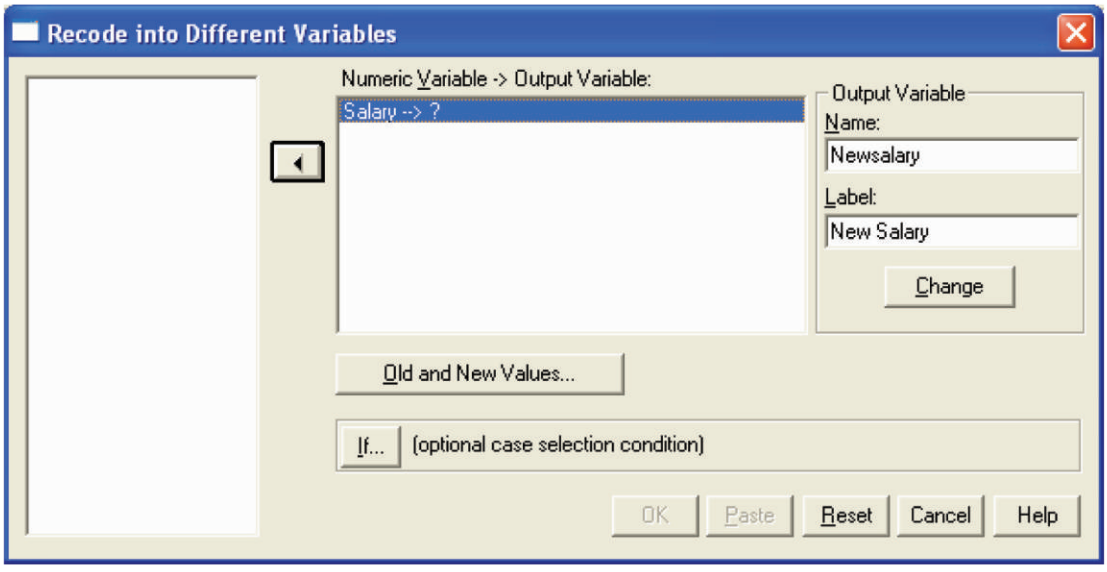
الحل:

لتنفيذ ذلك نتبع الخطوات الآتية:

- من القائمة Transforms يوجد في هذا المجال اختياران:
  - Recode into the Same Variable، وهذا يعني أن نجري العملية ونضع القيم الجديدة فوق القيم القديمة للمتغير.
  - Recode into Different Variable ويعني هذا أن نضع القيم الجديدة تحت اسم متغير جديد.

✓ عند الاختيار الأول تتم العملية تلقائياً، أما عند الثاني فسيطلب البرنامج معلومات عن المتغير الجديد.

✓ عند اختيار الأمر الثاني يظهر الصندوق الحواري Recode Into Different Variable ننقل المتغير المراد تكويده إلى Input Variable وليكن Salary (الأجر)، ثم نحدد اسم المتغير الجديد في المكان المسمى Name وليكنY انظر الصندوق الحواري الشكل (3-10):



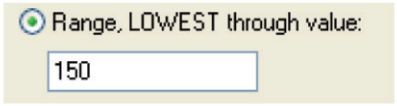
شكل (3-10)

⊙ تبـدأ عمليـة التكويد بالنقـر علـى Old and new values يظهر الصندوق الحـواري Recode Into Different Variable: Old and New Values انظر الشكل (3-11):

✓ فإذا كنا نريد تكويد قيمة

فإننا نقوم بالنقر تحت الأمر Old Values، ولكننا هنا نريد تكويد فئات لها أكثر من نوع.

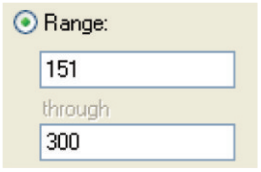
✓ أول فئة لدينا هي 150 فأقل فإننا نختار



تحـت الأمـر Old Values ثم يلي ذلـك أن ندخل الترميز "1" في المستطيـل الموجود تحت كلمة New value ويكون بالشكل الآتي:

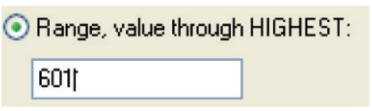
⊙ ثم ننقر على كلمة Add فتنتقل القيم المختارة إلى الصندوق New – Old

بالطريقـة نفسهـا يتم تكويـد الفئات الآتية حتـى الفئة قبل الأخـيرة كل فئة لها حد أدنـى وحد أعلى معـرف، مثـلاً "151-300" الحد الأدنى هو 151 والحد الأعلى هو 300 وهكذا باقي الفئات، في هذه الحالة نقوم بإدخال الحد الأدنى في الخانة العلوية والحد الأعلى في الخانة السفلية كما يلي:



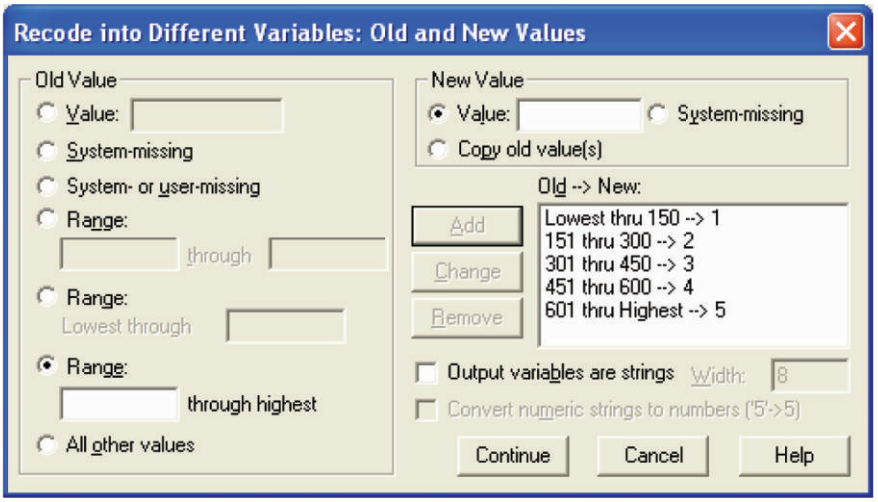
ثـم يلـي ذلك أن ندخل الترميـز "2" في المستطيل الموجود تحت كلمـة New value ثم النقر علي add لانتقال إلى إدخال فئة جديدة لتكويد.

⊙ الفئة الأخيرة لها حد أدنى وليس لها حد أعلى وتسمى الفئة المفتوحة من أسفل فنقوم بإدخال الحد الأدنى لها كما يلي:



ثم يلي ذلك أن ندخل الترميز "5" في المستطيل الموجود تحت كلمة New value.

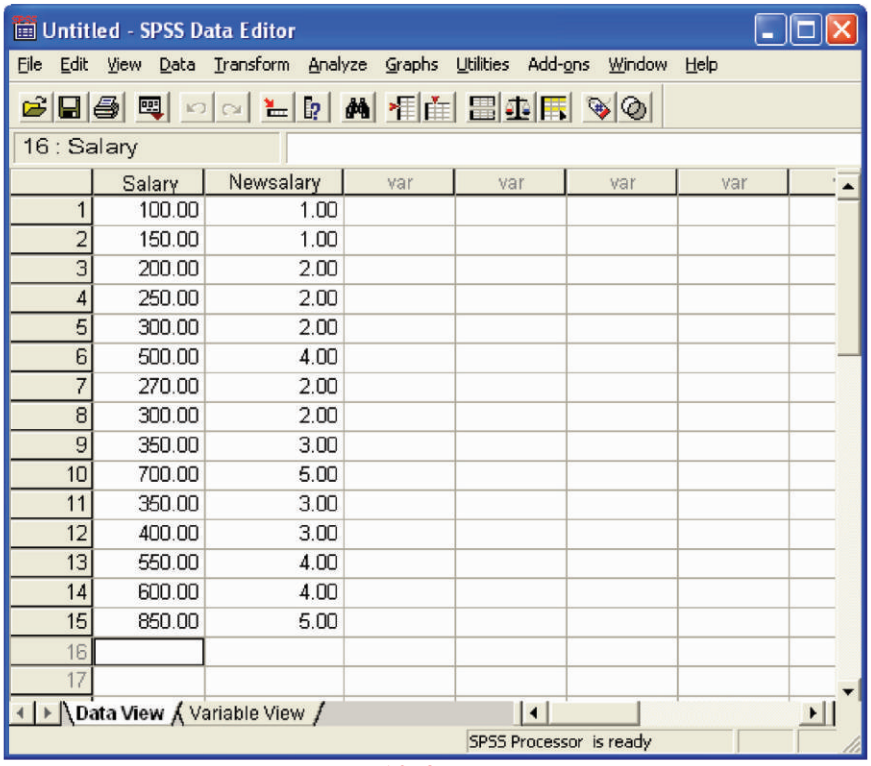
عنـد الانتهاء من عمليـة التكويد يظهر الشـكل (3-11) الآتي، في النهاية ثم نختـار الأمر Continue لنعود إلى صندوق الحوار الأصلي.



شكل (3-11)

ننقر Change ليتم التغيير من قيم المتغير إلى الترميز الجديد.

ننقـر على الأمـر Ok فيتم تغيير القيم الأصليـة إلى الترميز الجديد حسب الفئـات تحت اسم متغير جديد يرمز له بالاسم Newsalary كالآتي:



شكل (3-12)

(3-4) تكويد متغير لفظي إلى متغير رقمي:

⊙ إنشاء متغيرات وهمية وفئوية باستخدام التكويد:

يعد التكويد أداة مهمة في تحليل العلوم الاجتماعية، حيث تكون عملية المقارنة بين المجموعات المختلفة عملية مهمة في التحليل مثل المقارنة بين كفاءة الذكور و الإناث في رياضة معينة، المقارنة بين جودة التعليم العامة والخاصة في الجامعات ... إلخ. وعلى ذلك فإنه ينبغي لتحليل تلك المقارنات استخدام متغيرات وهمية ومتغيرات فئوية.

⊙ المتغيرات الاسمية (الوهمية) والمتغيرات الفئوية:

المتغيرات الاسمية (الوهمية) هي المتغيرات التي تأخذ قيمتين عادة ما تكون صفرًا وواحدًا. كل قيمة ترمز إلى فئة معينة والقيمة الأخرى ترمز إلى الفئة الثانية.

جدول (3-4)

Value	Category
0	Male
1	Female

المتغيرات الفئوية هي المتغيرات التي تحتوي على أكثر من قيمتين، كل قيمة ترمز إلى فئة معينة، ولا توجد أفضلية لفئة على الأخرى مثل أصول الجنسية في الولايات المتحدة تنقسم إلى ست فئات والتحويل من الفئات إلى القيمة موضح في الجدول الآتي:

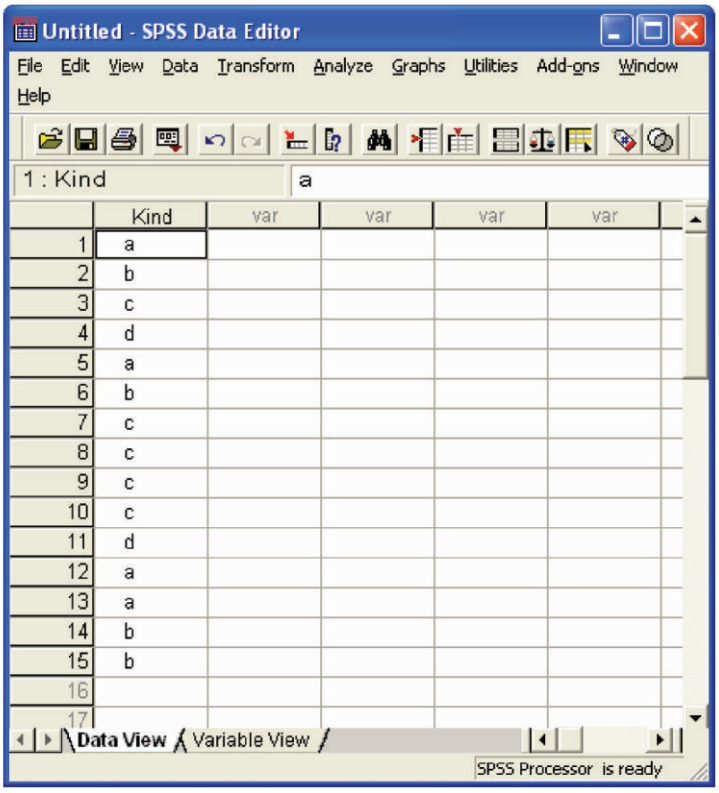
جدول (3-5)

Value	Category
0	White-American
1	African-American
2	Asian-American
3	Hispanic-American
3	Native-American
5	Other

والآن نقوم بشرح كيفية تحويل تلك المتغيرات باستخدام البرنامج.

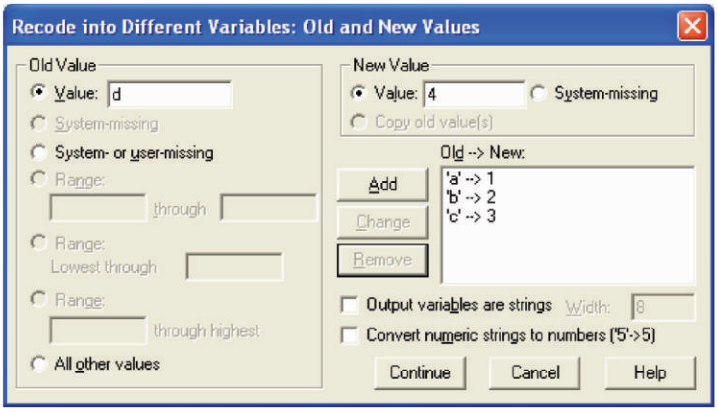
مثال (3-4):

نفترض أننا نريد تكويد المتغير الآتي إلى متغير جديد:



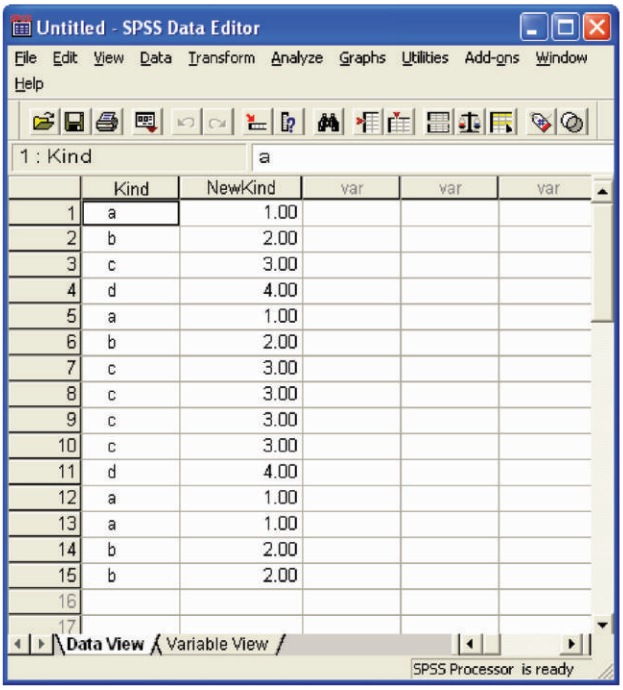
شكل (3-13)

لتنفيذ ذلك ننفذ كل الخطوات السابقة في مثال (3-3) إلى أن يظهر الصندوق الحواري Old And New Values من الأمر Old Values، نختار أول مستطيل يسارا من أعلى المسمى Value ويكتب به الحرف المراد تكويده، ثم نضع الرقم المناظر لذلك الحرف في أول مستطيل يمينا من أعلى New Value وننقر Add لتضاف إلى القائمة Old-New، نكرر هذه العملية لنحصل على الصندوق الحواري الشكل (3-14):



شكل (3-14)

ننقر Continue ونمر بالخطوات السابقة نفسها إلى أن يتم التغيير فيصبح المتغير بعد التكويد بالشكل الآتي:



شكل (3-15)

(3-5) إحلال قيم مفقودة بقيم أخرى:

إن وجـود قيـم مفقودة لبعض المتغيرات يعد أحيانًا عقبة كبيرة تواجـه تطبيق أسلوب إحصائي معين، ويجـب في هـذه الحالة تقدير القيم المفقـودة، حيث يوفر البرنامج SPSS هـذه الطريقة. بفرض أن متغير الدخل به مجموعة من القيم المفقودة، تتم عملية الاستبدال عن طريق البرنامج كما يأتي:

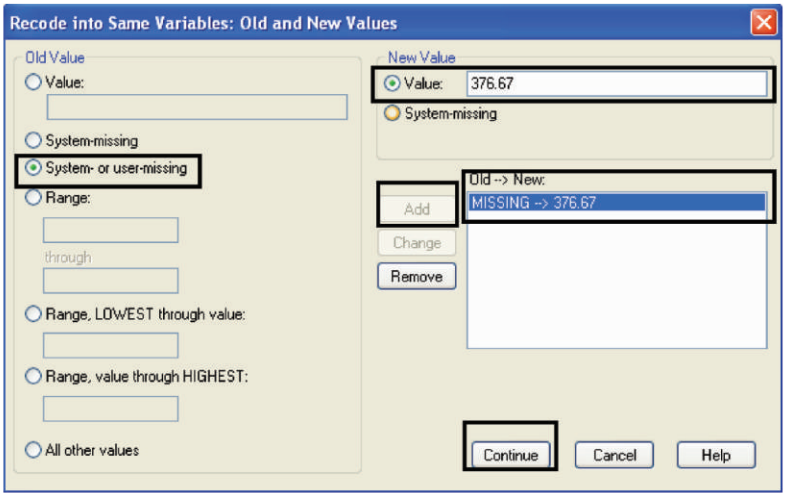
مـن القائمـة Transform انقر الأمـر Recode into same variables يظهـر الصندوق الحواري. اختر المتغير (أو المتغيرات) من قائمة المتغيرات وانقلها إلى المستطيل Variables.

- انقـر علـى Old And New Values لفتح الصندوق الحواري الآتي الخاص بعمليات التكويد والمسمى Recode into same Values: Old And New Values.
- في المربع Old Value نختار System-Or User –Missing.

في المستطيل New Value اكتب القيمة المراد إحلالها مكان القيمة المفقودة ولتكن الوسط الحسابي للبيانات مثلاً. ننقر الأمر Add فيتم الاختيار.

- انقر الأمر Continue للعودة للصندوق الأصلي.

انظر الصندوق الآتي بفرض أن الوسط الحسابي 376.67 الشكل (3-16).



شكل (3-16)

- ننقر الأمر Ok لتنفيذ العملية.

(3-5) الترتيب Ranking:

يمكن بواسطة هذا الأمر تكوين متغيرات جديدة هي عبارة عن رتب متغيرات معينة، وتكون هذه الرتب تصاعدية أو تنازلية، ويمكن إعطاء رتب لمتغير بواسطة متغيرات أخرى، حيث إننا في كثير من طرق العرض والتحليل الإحصائي نرغب في التعامل مع الرتب الخاصة بالمتغير وليس بقيم المتغير.

مثال (3-5):

بفرض أن لدينا المتغيرات الآتية: الدخل، الاستهلاك والادخار كالآتي:

جدول (3-6)

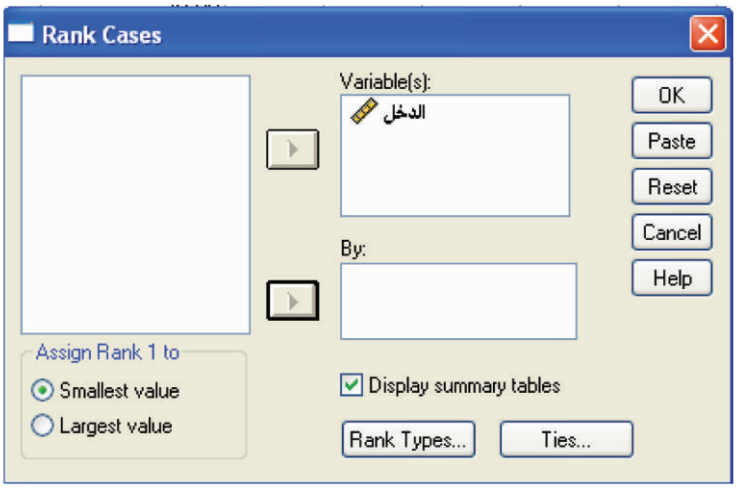
Name	Income
Ahmed	50 ,00
Mohamoud	100 ,00
Mona	150 ,00
Heba	150 ,00
Amal	200 ,00
Mohamed	300 ,00
Shamaa	300 ,00
Ezi Aldin	500 ,00
Nora	500 ,00
Abdul Tawaab	600 ,00
Amr	630 ,00
Mahdi	830 ,00

ونريد إعطاء رتب تصاعدية لهذا المتغير.

الحل:

لتنفيذ ذلك نتبع الخطوات الآتية:

من القائمة Transform نختار Rank Cases فيظهر صندوق حوار Rank Cases كما في الشكل (3-17):



شكل (3-17)

- لقد قمنا بإدخال العمود الذي نريد إعطاءه رتبة الدخل في خانة Variables

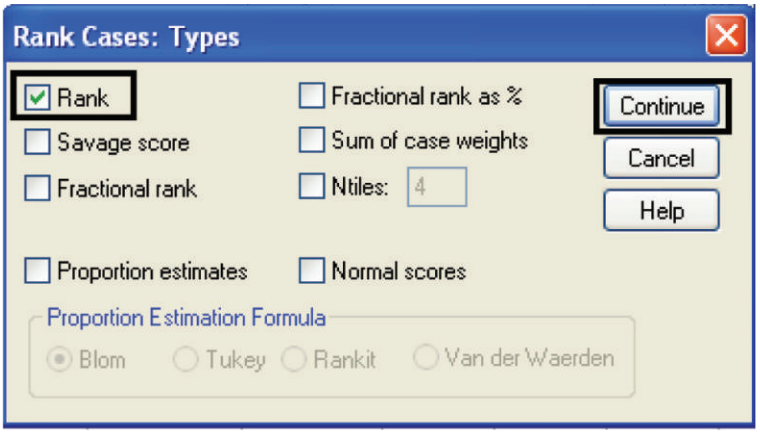
وفي خانة Assign Rank To أشرنا الخيار Smallest Value لإعطاء الرتب تصاعديًا.

- لاختيار نوع الرتبة انقر زر Rank Types في صندوق حوار Rank Cases يظهر صندوق حوار Rank Cases: Types فتظهر أنواع مختلفة من الرتب:

- Rank: وهي الرتبة البسيطة (حيث يتم إعطاء رتبة لكل قيمة من قيم المتغير تعبر عن ترتيبه ضمن المجموعة).
- Savage Scores: تعطي رتبًا لقيم المتغير بموجب التوزيع الآسي.
- Fractional Rank: وهي الرتبة الناتجة من قسمة الرتبة البسيطة لقيم المتغير على مجموع الأوزان لكل الحالات (أو عدد الحالات في حالة عدم وجود أوزان؛ أي يعد الوزن مساوياً للواحد).
- Fractional Rank as %: يتم الحصول على هذه الرتبة من حاصل ضرب الرتبة السابقة في 100.
- Sum Of Cases Weights: الرتبة تكون مساوية لكل الحالات، وتمثل مجموع الأوزان لكل الحالات (أو عدد الحالات في حالة عدم وجود أوزان).
- Ntiles: يتم إعطاء رتب بعد تقسيم قيم المتغير إلى مجاميع تعطي كل منها رتبة معينة (بعد ترتيبها تصاعدياً أو تنازلياً) فإذا اخترنا Ntiles 4 (حالة الترتيب تصاعدي) فإنه سيتم إعطاء الرتبة 1

للقيم التي ترتيبها أقل من 25%، وتعطي الرتبة 2 للقيم من 25% إلى 50%، والرتبة 3 للقيم 50% إلى 75%، والرتبة 3 لـ 75% فما فوق.

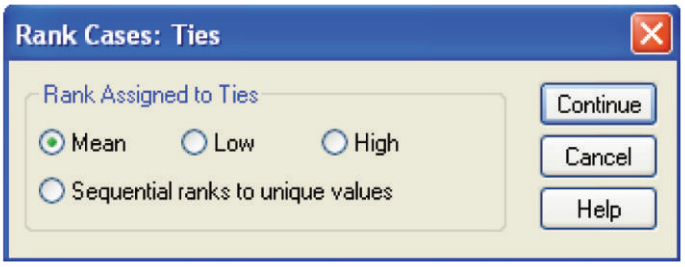
وهنا في هذا المثال يتم اختيار النوع البسيط للترتيب Rank انظر الشكل (3-18):



شكل (3-18)

ثم ننقر على Continue

النقر على الأمر Ties يعطي الصندوق الحواري الفرعي Ties:Rank Cases انظر الشكل (3-19):



شكل (3-19)

وفيـه يختار المستخـدم طريقة التعامل مع التداخلات Rank Assign to Ties هل باستخدام الوسط أم القيمة الصغرى أم الكبرى للرتب المتداخلة؟

ثم زر Ok.

فيضاف متغير جديد (متغير الرتب) باسم الدخل R إلى Data Editor كما في الجدول الآتي:

جدول (3-7)

الاسم	الدخل	الدخلR
احمد	50.00	1.000
محمود	100.00	2.000
منى	150.00	3.500
هبة	150.00	3.500
امال	200.00	5.000
محمد	300.00	6.000
شيماء	400.00	7.000
عز الدين	500.00	8.500
نورا	500.00	8.500
عبد الثواب	600.00	10.000
عمرو	630.00	11.000
مهدي	830.00	12.000

☜ إنشاء سلسلة زمنية Create Time Series:

هـي عبـارة عن قيم متغير معين خلال فترات زمنية متساويـة كالأيام أو الأشهر أو السنوات، ويمكن إجـراء بعض العمليات الإحصائية على السلسلـة الزمنية من خلال عـدة دوال إحصائية تضمن: الفروق Differences، الأوساط المتحركة Moving Averages، التأخر Lag، التقدم Lead، وغيرها.

مثال (3-6):

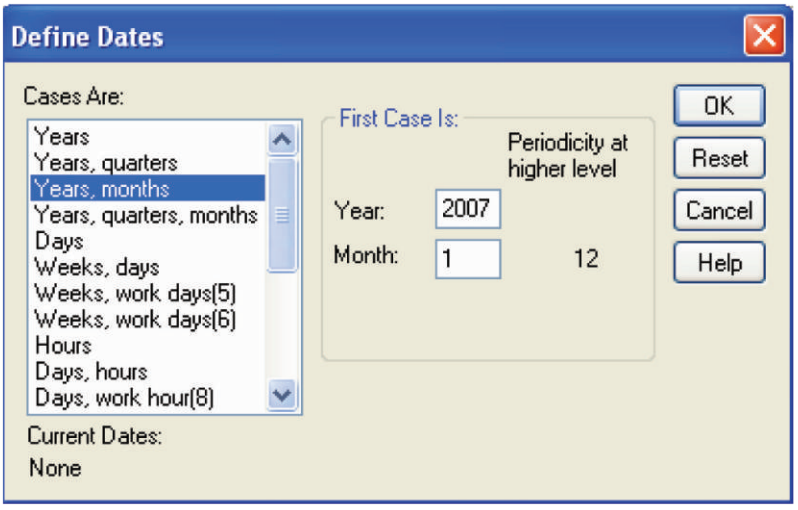
لدينـا متغير يمثل حجم المبيعـات الشهرية من أجهزة المحمول خلال 12 شهراً في مؤسسة معينة لسنة 2007 م، ونرغب في عمل فروق Differences من الدرجة الأولى لهذا المتغير:

X	273	207	255	350	382	351	268	380	309	335	335

الحل:

⊙ أولاً نقوم بإنشاء السلسلة الزمنية بإنشاء الشهور بالنسبة لسنة 2007 كالآتي:

انقـر على الأمـر Define Dates من القائمـة Data فيظهر صندوق الحـوار Define dates الشكل (3-20) الذي نقوم بترتيبه كما يأتي:



شكل (3-20)

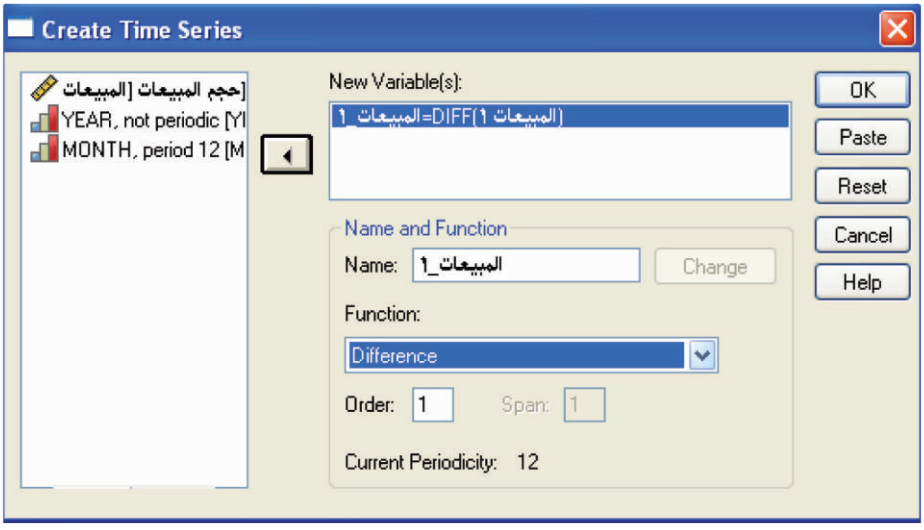
حيث اخترنا من Cases are السنة والشهور Months، Years ويمكن للباحث أن يختار ما يشاء سواء الأيام أو الساعات أو الأسابيع وغيرها.

- في قائمة First case is نقوم بتعريف التاريخ للحالة الأولى في السلسلة الزمنية: في حقل السنة ندخل سنة البداية 2007، أما في حقل الشهر ندخل شهر البداية 1.
- أما Periodicity at Higher Level فيبين دورية التاريخ، حيث يبين أكبر قيمة يمكن تزويدها للبرنامج مثلاً في المثال أعلى دورية للشهور 12 شهراً.
- عند نقر Ok في صندوق حوار Define Dates تضاف متغيرات التاريخ إلى Data Editor كما يأتي:

المبيعات	YEAR_	MONTH_	DATE_
274.00	2007	1	JAN 2007
207.00	2007	2	FEB 2007
255.00	2007	3	MAR 2007
350.00	2007	4	APR 2007
382.00	2007	5	MAY 2007
383.00	2007	6	JUN 2007
351.00	2007	7	JUL 2007
268.00	2007	8	AUG 2007
380.00	2007	9	SEP 2007
409.00	2007	10	OCT 2007
445.00	2007	11	NOV 2007
455.00	2007	12	DEC 2007

- ثانياً نقوم بعمل الفروقات Differences بأتباع الخطوات الآتية:

- انقر على الأمر Create Time Series من القائمة Transform، يظهر الصندوق الحواري Create Time Series حدد اسم المتغير الذي تعتمد عليه السلسلة وانقله إلى المستطيل New Variables انظر الشكل (3-21).



شكل (3-21)

حدد اسم المتغير الجديد في المستطيل Name تحت الأمر Name and Function وليكن "المبيعات_1".
يتم تحديد الدالة الجديدة التي سيتم على أساسها إنشاء بيانات المتغير الجديد من الأمر Function التي في المثال هي Differences، المتغير الجديد المختار هو دالة الفروق.

عند نقر Ok يضاف متغير جديد باسم المبيعات_1 إلى Data Editor كما يأتي في الشكل (3-22):

المبيعات	YEAR_	MONTH_	DATE_	المبيعات_1
274.00	2007	1	JAN 2007	.
207.00	2007	2	FEB 2007	-67.00
255.00	2007	3	MAR 2007	48.00
350.00	2007	4	APR 2007	95.00
382.00	2007	5	MAY 2007	32.00
383.00	2007	6	JUN 2007	1.00
351.00	2007	7	JUL 2007	-32.00
268.00	2007	8	AUG 2007	-83.00
380.00	2007	9	SEP 2007	112.00
409.00	2007	10	OCT 2007	29.00
445.00	2007	11	NOV 2007	36.00
455.00	2007	12	DEC 2007	10.00

شكل (3-22)

الفصل الرابع

العرض والتحليل الإحصائي لمتغير أحادي

View and Statistical Analysis of Univariate Variable

(4-1) مقدمة :

يبـدأ التحليل الإحصائـي بتحليل السمات الإحصائية لكل متغير على حده الذي نطلق عليه التحليل الأحادي Univariate Analysis، ومن خلال التحليل الإحصائي يمكننا الحصول على المعلومات الآتية:

- ⊙ معرفة كيف تتوزع قيم المفردات- توزيعًا طبيعيًا، أسيًا، ذا الحدين...،
- ⊙ حسـاب مقاييس النزعة المركزية، وهي مجموعة من المقاييس تصف المتغير مثل الوسط، الوسيط، المنوال...،
- ⊙ معرفـة تشتت المفردات عن القيم المركزية المحسوبـة في الخطوة السابقة، وتقاس عادة باستخدام التباين، الانحراف المعياري، المدى ...،
- ⊙ توضيح هل توجد قيم شاذة في المفردات من خلال مقياس معامل الاختلاف.
- ⊙ إجراء اختبارات الفروض على البيانات.

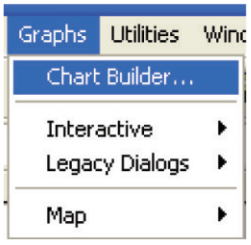
في هذا الفصل نقوم بتوضيح الطرق المختلفة لعرض وتحليل متغير أحادي، ففي المبحث الأول نقوم بتوضيـح كيف يمكن عرض البيانات في أشكال بيانية أو مخططات رسومية للمتغير. ونستكمل فيما تبقى من الفصل كيفية إيجاد المقاييس الإحصائية المناسبة للمتغير.

(4-2) عرض البيانات باستخدام

☜ الرسومات البيانية البسيطة:

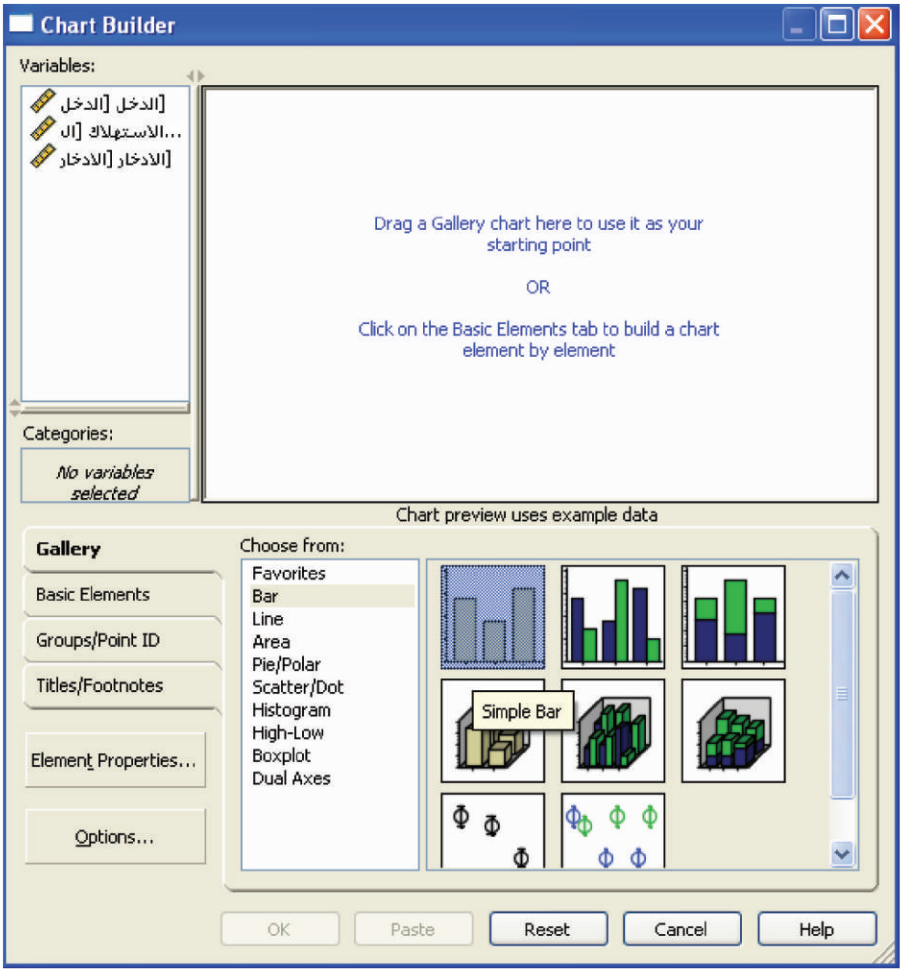
الرسومـات البيانية نستطيـع أن نستخدمها للحصول على معلومات محددة عن المتغير مثل الوسط، الوسيط، التوزيع المتراكم ولعمل رسم بياني نتبع الخطوات الآتية:

- ⊙ نختار Graphs ثم نقوم باختيار Chart Builder... من الشكل الآتي:



شكل (4-1)

⊙ فيظهـر الشـكل (4-2) فنختار منه Bar من عمود Choose from: ثم نقـوم باختيار الشكل الأول وهو Simple Bar وذلك بالنقر عليه نقراً مزدوجًا:



شكل (4-2)

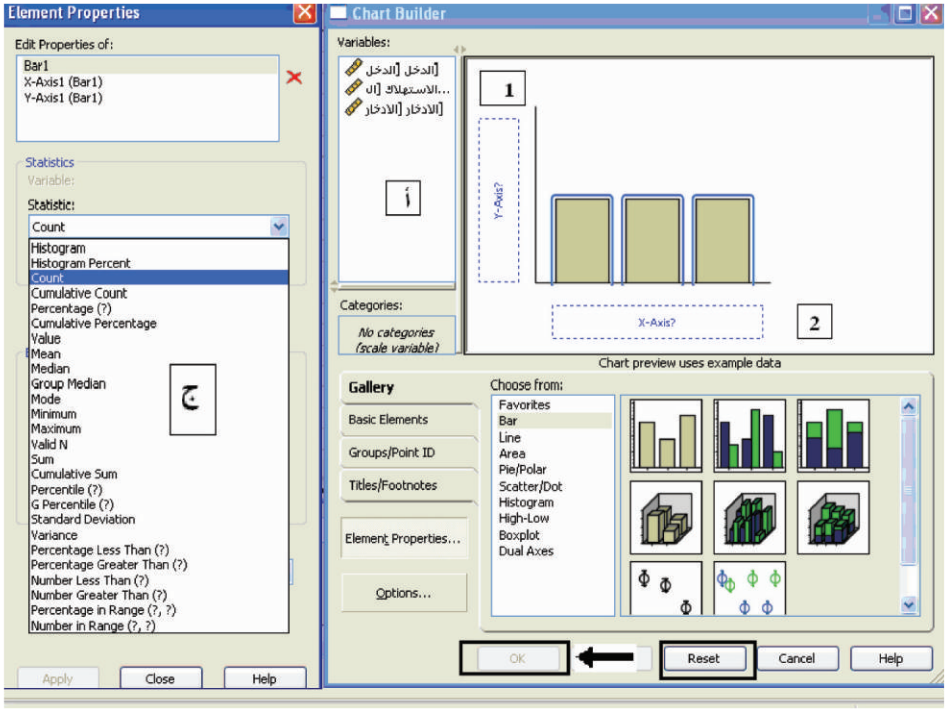
⊙ بعـد النقر على الشكل نقـراً مزدوجًا يظهر الشـكل (4-3) الآتي الذي يتكون مـن جزأين بالشكل الآتي:

أ - الجزء الأيمن هو شكل الرسم البياني والمتغيرات المتاحة للرسم.

ب - الجزء الأيسر وهو شكل خصائص المحاور المتاحة.

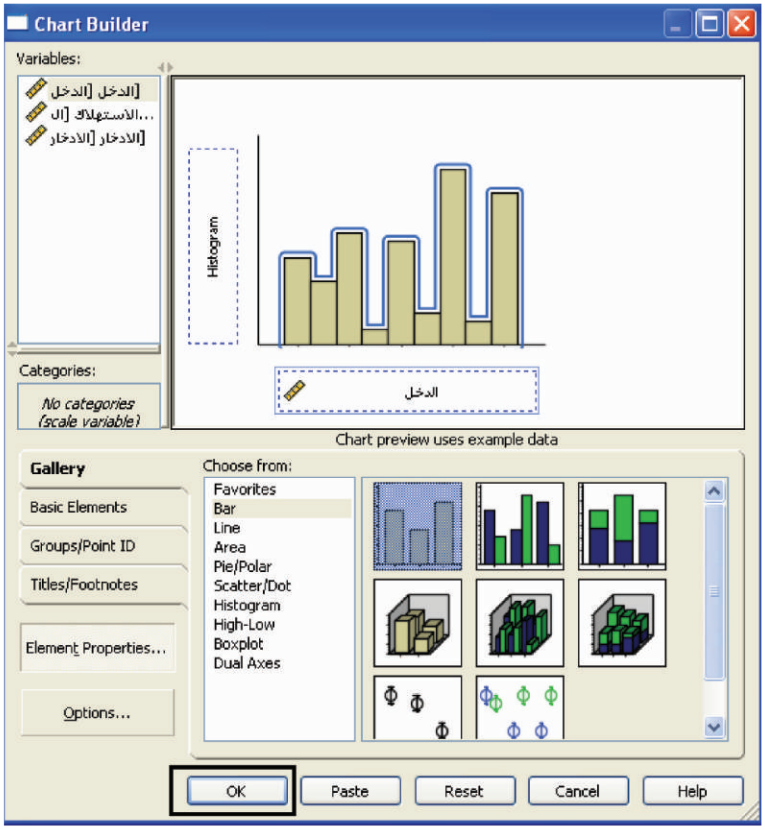
وهنا يوجد داخل شاشة الرسم ما يأتي:

- "أ" وهو عمود المتغيرات، وهنا نفترض أننا نريد رسم المتغير الدخل عن بيانات المثال (3-1).
- "1" Y-Axis? وهو محور العمليات التي نوضح عليها الشكل، وهنا نفترض أننا نريد أن نوضح الشكل التكراري للمتغير Histogram فإننا ننتقل إلى الجزء الثاني من الشكل (3-4).
- "2" وهو المحور X في الشكل البياني المراد رسمه، ونفترض أننا نريد رسم المتغير الدخل على المحورX فإننا ننقر عليه ونسحبه إلى محور X والمحدد في الشكل برقم 2.
- "ج" وهو عمود يحمل اسم Statistics، وهو عمود العمليات التي نريد إظهارها على المحور Y، ونستطيع الاختيار المتعدد لها ولنفرض أننا نريد إظهار تكرار قيم متغير الدخل فإننا نختار Histogram من (ج).



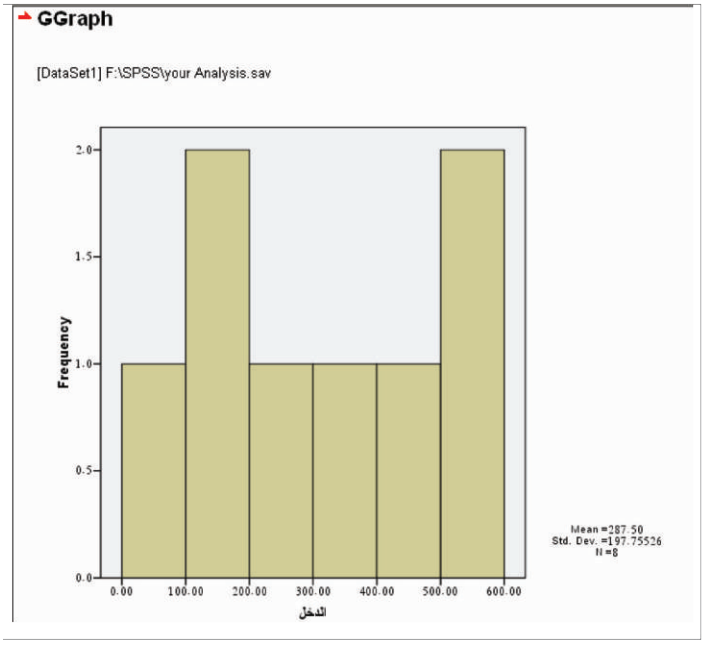
شكل (3-4)

- بعد الانتهاء من العمليات السابقة يصبح الشكل (3-4) السابق الصورة الآتية:



شكل (4-4)

ومن ثم النقر على OK

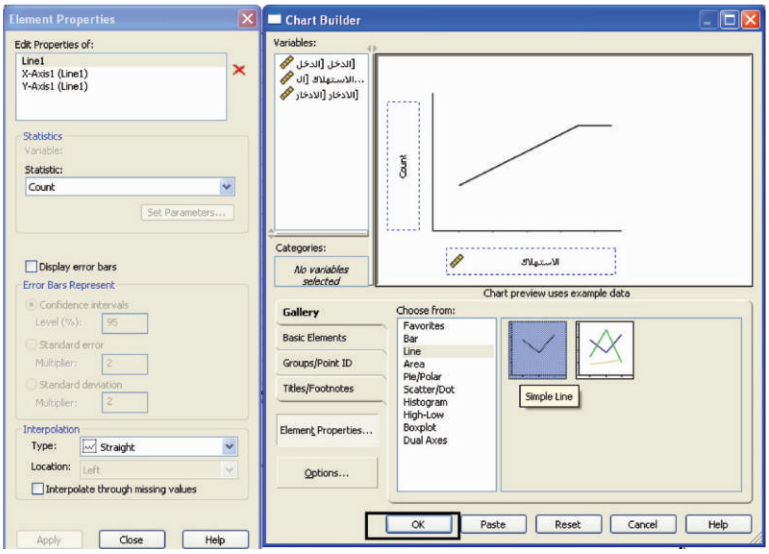


شكل (4-5)

- ونلاحظ أن الشكل السابق يوضح الرسم البياني لمتغير الدخل وبجوار الرسم بعض المقاييس الإحصائية عن المتغير مثل الوسط الحسابي Mean يساوي 287.4، الانحراف المعياري يساوي 197.74426 وعدد البيانات يساوي 8.

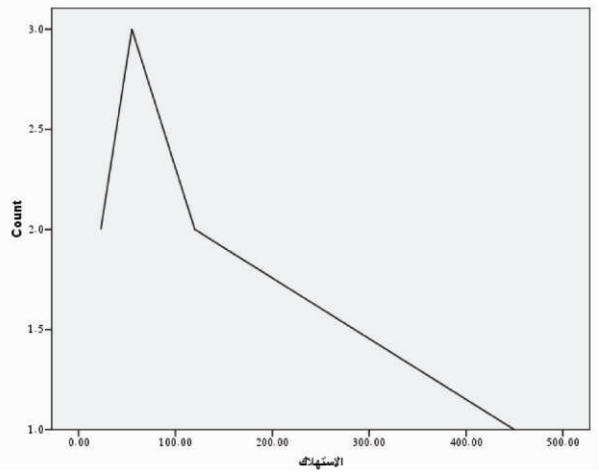
Line Graphs

إذا أردنا عمل شكل بياني من نوع line لمتغير الاستهلاك فإننا نختار Graphs ثم نقوم باختيار Chart Builder... فيظهر الشكل (4-2) فنختار منه Line من عمود Choose from: ثم نقوم باختيار الشكل الأول وهو Simple Line، وذلك بالنقر عليه نقراً مزدوجًا، ومن ثم اختيار المتغير الاستهلاك من عمود المتغيرات بطريقة اختيار الدخل السابق نفسها، وذلك بسحبها إلى عمود X فيظهر الشكل (4-6) الآتي:



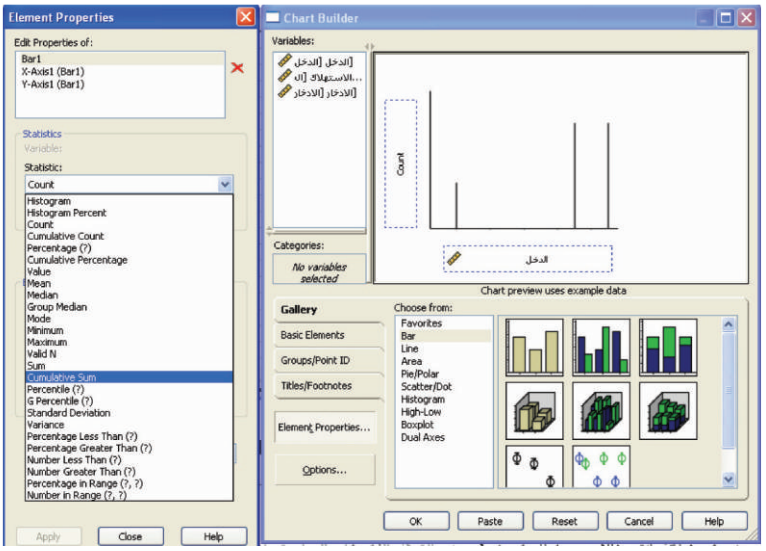
شكل (4-6)

- وبعد النقر على ok لتأكيد الاختيار فإن الشكل (4-7) يظهر في نافذة output.



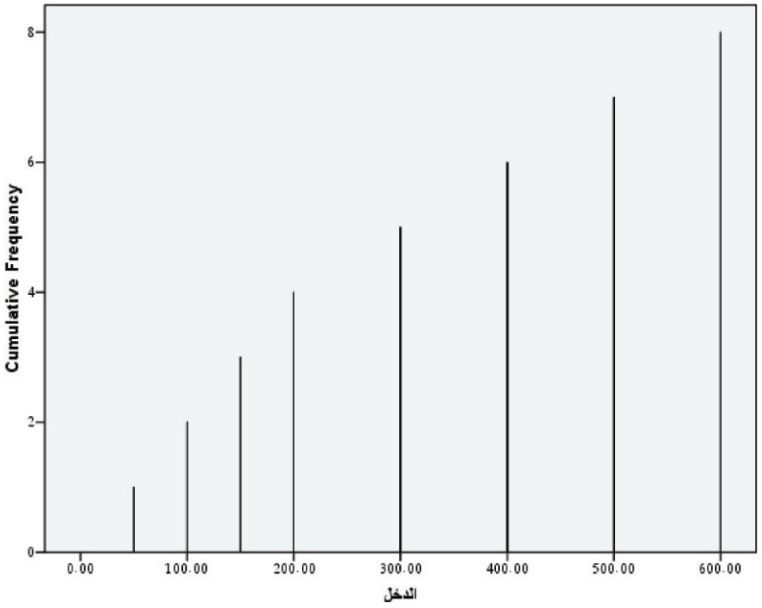
شكل (4-7)

إذا أردنـا عمـل شكل بياني من نـوع Graphs for cumulative frequency للمتغـير الدخل فإننا نختار Graphs ثـم نقـوم باختيـار Chart Builder... فيظهـر الشكل (4-2) فنختار منـه Bar من عمود Choose from: ثـم نقـوم باختيـار الشكل الأول وهـو Simple Bar وذلك بالنقر عليه نقراً مزدوجًـا، ومن ثم اختيار المتغـير الدخـل من عمود المتغيرات، وذلـك بسحبه إلى عمود X ومن الجزء الذي علـى اليسار يتم اختيار cumulative sum من عمود Statistic فيظهر الشكل (4-8) الآتي:



شكل (4-8)

وبعد ذلك النقر على Ok فيظهر الشكل (4-9) الآتي:



شكل (4-9)

الحصول على مقياس إحصائي:

في كثير من النواحي التطبيقية يكون الباحث في حاجة إلى حساب بعض المؤشرات التي يمكن الاعتماد عليها في وصف الظاهرة من حيث القيمة التي تتوسط القيم أو تنزع إليها القيم، وتسمى مقاييس النزعة المركزية، ومن حيث التعرف على مدى تجانس القيم التي يأخذها المتغير وتسمى مقاييس التشتت، وأيضًا ما إذا كانت هناك قيم شاذة أم لا من خلال مقياس يسمى معامل الاختلاف. والاعتماد على العرض البياني وحده لا يكفي، ولذا يتناول هذا الفصل أيضاً كيف يمكن وصف المتغير باستخدام المقاييس الإحصائية المناسبة، حيث توجد مجموعة رئيسة من المقاييس تعبر عن البيان الإحصائي منها:

(3-4) مقاييس النزعة المركزية:

⊙ الوسط الحسابي:

يعد الوسط الحسابي من أهم المقاييس في التحليل الإحصائي بشكل عام، إلا أنه يناسب الظواهر الكمية فقط، ونستطيع تعريفه على أنه القيمة الناتجة من خارج قسمة قيمة مجموع المشاهدات على عددها، وتحسب بشكل رياضي من العلاقة الآتية:

$$\overline{X} = \sum_{i=1}^{n} x_i / n,\ \mathrm{i} = 1, 2, \ldots \mathrm{n}$$

حيث:

$\overline{X}$ رمز للمتوسط وتقرأ "إكس بار"،

$\sum_{i=1}^{n}$ رمز للمجموع وتقرأ "سيجما من i=1 إلى n"،

n عدد المفردات محل الدراسة.

⊙ الوسيط:

الوسيط هو أحد أهم مقاييس النزعة المركزية، ويمكن أن تكون الظاهرة أو المتغير كميًا أو ترتيبيًا، ويعرف بأنه القيمة التي تقع في منتصف البيانات بعد الترتيب (تصاعديًا أو تنازليًا) في حالة البيانات المفردة.

⊙ المنوال:

وهو القيمة الأكثر تكرارًا في البيانات، وقد يكون للبيانات منوال وحيد أو أكثر من منوال، وأحياناً ينعدم المنوال لبعض البيانات.

(4-4) مقاييس التشتت المطلق والنسبي:

⊙ المدى:

ويعرف بأنه الفرق بين أكبر رقم وأصغر رقم للبيانات وهو كالآتي:

المدىR = أكبر قراءة – أصغر قراءة.

⊙ الانحراف المعياري:

يقيس تشتت البيانات حول المتوسط، وهو الجذر التربيعي الموجب للتباين σ^2 ويحسب كالآتي:

$$\sigma = \sigma^2 = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

حيث:

σ رمز للانحراف المعياري وتقرأ "سيجما".

σ^2 رمز التباين وتقرأ "سيجما تربيع".

⊙ معامل الاختلاف:

وهو من المقاييس النسبية للتشتت، ويعرف بأنه خارج قسمة الانحراف المعياري على الوسط أي

$$\gamma = \frac{\sigma}{\mu}$$

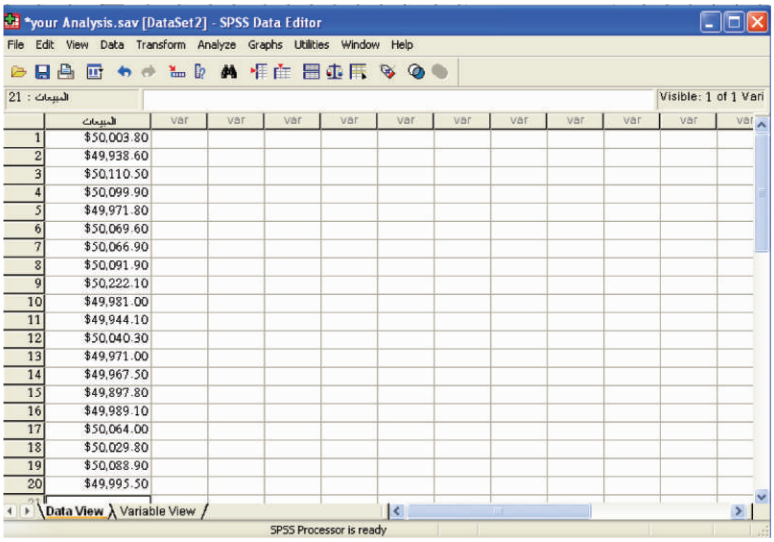
وهو من أفضل مقاييس التشتت النسبية، ويستخدم للتعرف على وجود قيم شاذة للبيانات أم لا.

الآن نستطيع إيجاد المقاييس الإحصائية السابقة وغيرها من خلال برنامجنا SPSS وما يحويه من أوامر تمكننا من تنفيذ وتحديد كل المقاييس السابقة، فضلاً عن إيجاد توزيع تكراري للبيانات لمتغير واحد أو أكثر.

☜ الأمر Frequencies:

يستعمل هذا الأمر لعرض تكرار كل قيمة لمتغير ما وحساب النزعة المركزية ومقاييس التشتت والربيعيات والمئينات مع عرض بعض المخططات البيانية. يمكن الحصول على هذا الأمر من القائمة Analyze ثم الأمر Frequencies.

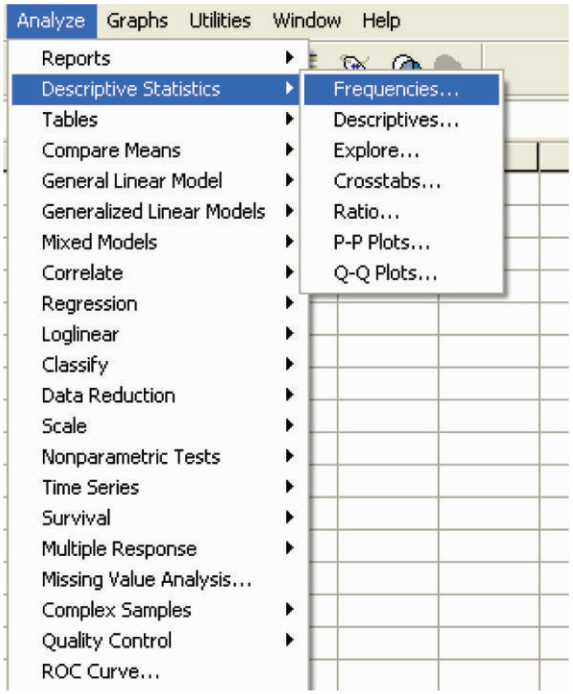
إذا افترضنا أننا لدينا بيانات عن قيمة مبيعات منتج السيارات لشركة ما وتم إدخالها في شاشة data view وكانت كالآتي:



شكل (4-10)

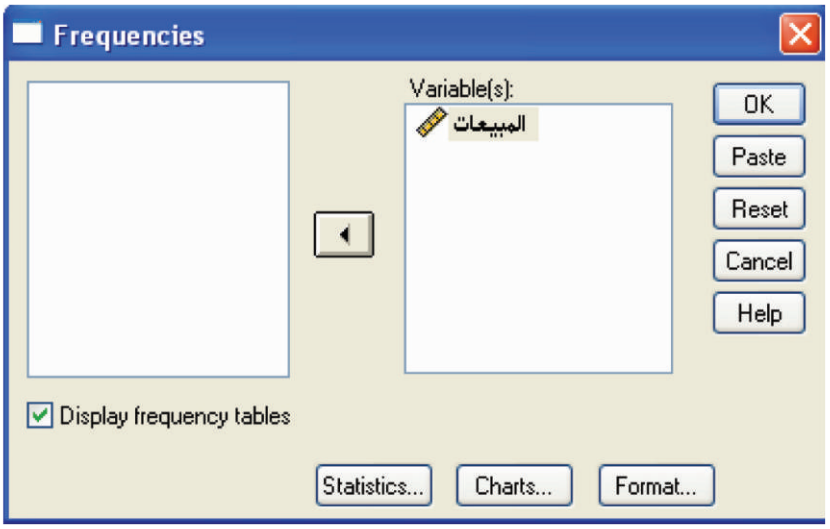
ونريد إيجاد توزيع تكراري للبيانات، وإيجاد مقاييس النزعة المركزية ومقاييس التشتت كالآتي:

من القائمة Analyze الظاهر في الشكل (4-11) الآتي:



شكل (4-11)

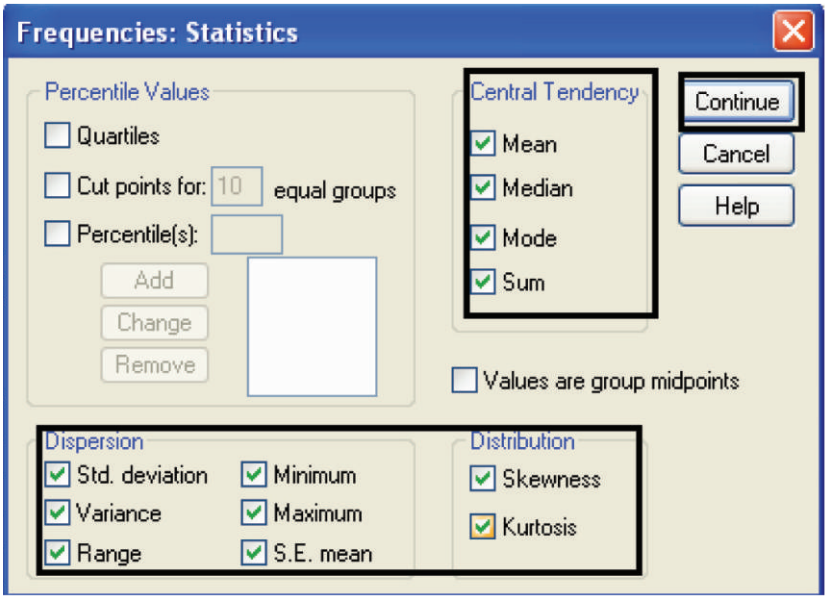
- انقـر على Descriptive statistics ثـم انقر على الأمر الفرعـي Frequencies فيظهر صندوق حوار Frequencies الذي يظهر الشكل الآتي بعد إدخال المبيعات عن طريق تحديدها ثم النقر على علامة الانتقال ▶ الذي يرمز للقيم في قائمة Variables انظر الشكل (4-12):



شكل (4-12)

حيث إن:

- Display Frequency Tables :- يعرض الجدول التكراري عند تأشيرة المربع المجاور له.
- Statistics :- يعرض بعض المؤشرات الإحصائية حسب حاجة المستفيد، عند نقره يظهر صندوق حواري Statistics وقد اخترنا المؤشرات الآتية للعرض كما يلي انظر الشكل (4-13):

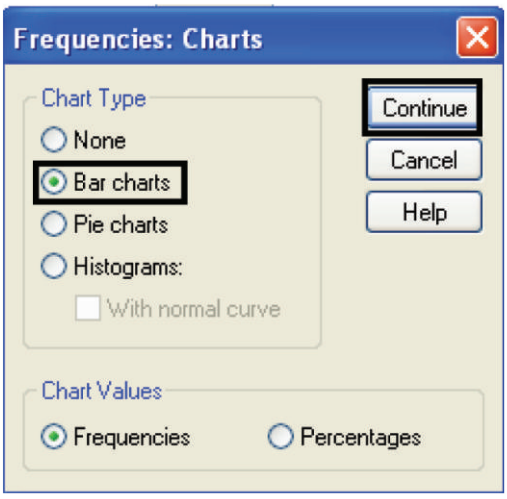


شكل (4-13)

Percentile Values :- تشمل الربيعات Quartile والمئينات Percentiles، ولتحديد المئين نكتبه في المربع المجاور لـ Percentiles ثم ننقر زر Add فيتم إضافة المئين الذي نريد إلى المستطيل أسفل Percentiles، ويستعمل الزر Change لتغيير قيمة المئين والزر Remove لإزالته.

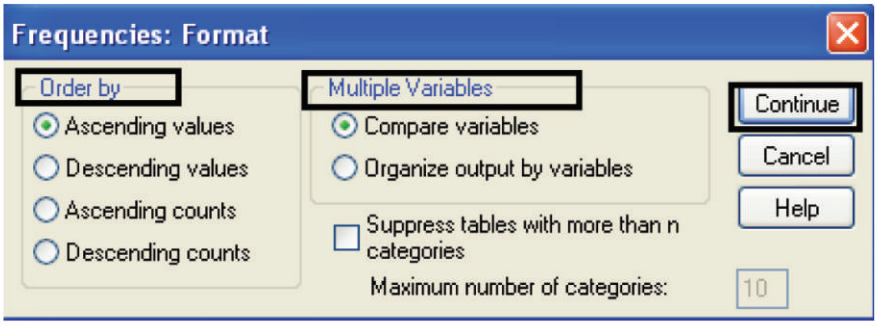
- ❑ Dispersion: لعرض مقاييس التشتت.
- ❑ Central Tendency: لعرض مقاييس النزعة المركزية.
- ⊙ وبعد تحديد ما نريد من مقاييس ننقر على continue.

❑ Chart: لعرض المخططات Bar، Pie وغيرها من المخططات ونفرض أننا نريد Bar فإننا نحددها ثم ننقر على continue كما في الشكل (4-14) الآتي:



شكل (4-14)

❑ Format: عند النقرعليه يظهر صندوق الحوار انظر الشكل (4-15):



شكل (4-15)

حيث إن:-

❑ Ordered By: لترتيب المشاهدات في الجدول التكرار تصاعديًا أو تنازليًا حسب القيم Values أو التكرارات Counts، وفي هذا المثال اخترنا الترتيب حسب القيم.

❑ Multiple Variables: يستخدم هذا الأمر في حالة وجود أكثر من متغير في القائمة Variables في صندوق حوار Frequencies، وتشمل ما يلي:

✓ Compare Variables: لعرض المؤشرات الإحصائية للمتغيرات كافة في جدول واحد.

✓ Organize Out By Variables: لعرض مؤشرات كل متغير في جدول مستقل، وفي هذا المثال لا يهم اختيار أي من هذين المتغيرين لوجود متغير واحد فقط.

✓ Suppress Tables with More Than n Categories: لإخفاء الجدول التكراري للمتغيرات التي يزيد عدد فئاتها على العدد المحدد من قبل المستفيد، ونلاحظ عند تنشيطها يتم تنشيط خانة بعنوان Maximum numbers of categories عندها نستطيع ما هو العدد المحدد الذي نريد إظهاره فقط.

⊙ عند النقر على زر Ok في صندوق حوار Frequencies تظهر النتائج الآتية:

جدول (4-1)

Statistics

المبيعات

N	Valid	20
	Missing	0
Mean		########
Std. Error of Mean		$16.97829
Median		########
Mode		########[a]
Std. Deviation		$75.92924
Variance		5765.250
Skewness		.650
Std. Error of Skewness		.512
Kurtosis		.782
Std. Error of Kurtosis		.992
Range		$324.30
Minimum		########
Maximum		########
Sum		$1000544

a. Multiple modes exist. The smallest value is shown

حيث إن:

❑ Valid: تمثل القيم الصحيحة (غير المفقودة).

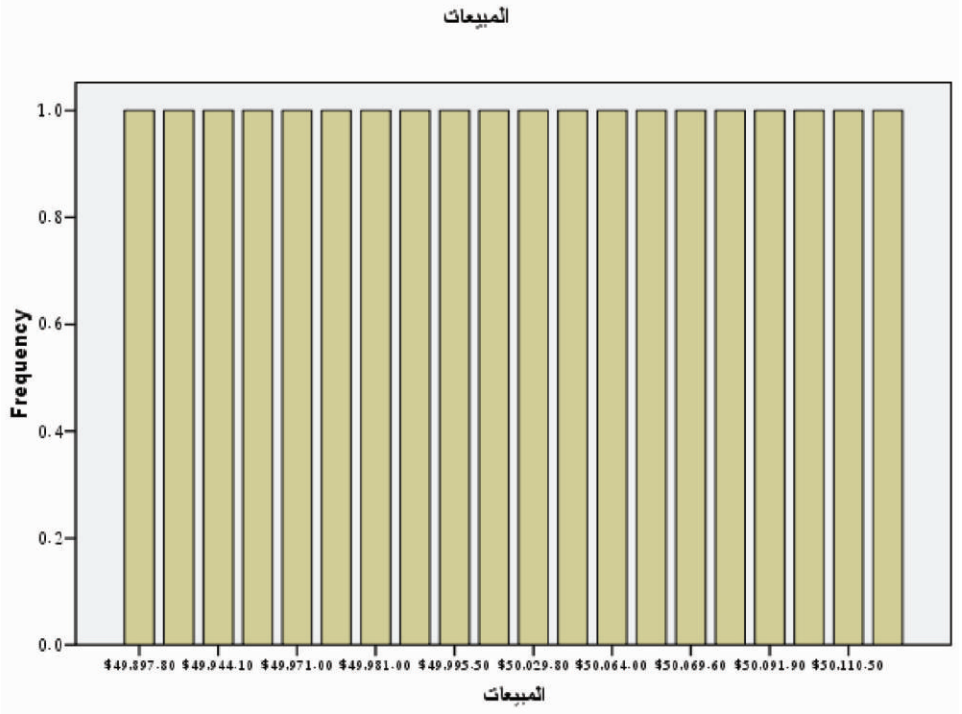
❑ Missing: تمثل القيم المفقودة.

جدول (4-2)

المبيعات

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	$49,897.80	1	5.0	5.0	5.0
	$49,938.60	1	5.0	5.0	10.0
	$49,944.10	1	5.0	5.0	15.0
	$49,967.50	1	5.0	5.0	20.0
	$49,971.00	1	5.0	5.0	25.0
	$49,971.80	1	5.0	5.0	30.0
	$49,981.00	1	5.0	5.0	35.0
	$49,989.10	1	5.0	5.0	40.0
	$49,995.50	1	5.0	5.0	45.0
	$50,003.80	1	5.0	5.0	50.0
	$50,029.80	1	5.0	5.0	55.0
	$50,040.30	1	5.0	5.0	60.0
	$50,064.00	1	5.0	5.0	65.0
	$50,066.90	1	5.0	5.0	70.0
	$50,069.60	1	5.0	5.0	75.0
	$50,088.90	1	5.0	5.0	80.0
	$50,091.90	1	5.0	5.0	85.0
	$50,099.90	1	5.0	5.0	90.0
	$50,110.50	1	5.0	5.0	95.0
	$50,222.10	1	5.0	5.0	100.0
	Total	20	100.0	100.0	

نلاحظ أن المشاهدات في الجدول قد رتبت تصاعديًا حسب قيم المتغير X.



ونلاحظ أن كل البيانات لها تكرار واحد، ولا توجد قيمة من البيانات تم تكرارها مرتين أو أكثر.

☜ الأمر Descriptive:

من القائمة Analyze يمكن الحصول على الأمر Descriptive، حيث يمكن إجراء التحليل الإحصائي للبيانات عن طريقه أيضًا، وذلك بالحصول على المقاييس الإحصائية المختلفة وعلى بعض الأشكال البيانية، هذا الأمر يسمح أيضًا بالحصول على ما يسمى القيم المعيارية، ويتم حساب القيم المعيارية وفق العلاقة:

$$Z = \frac{x_i - \overline{X}}{S}$$

حيث x_i هي القيمة المدخلة، $\overline{X}$ هي المتوسط الحسابي، S هو الانحراف المعياري، حيث تظهر القيم المعيارية عادة في نهاية ملف البيانات.

مثال (4-1):

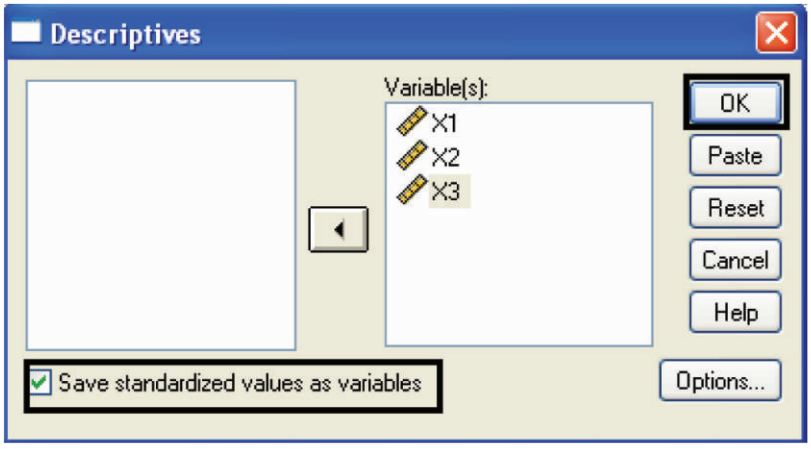
الجدول الآتي يتضمن المتغيرات X_1, X_2, X_3 التي تم إدخالها في Data editor للبرنامج:

جدول (4-3)

X_1	X_2	X_3
90	50	12
70	52	14
56	55	19
65	60	22
85	65	20
60		
69		
57		
50		
75		
62		
51		
85		

لغرض استخراج المقاييس الوصفية بالأمر Descriptive نتبع الخطوات الآتية:

- نختار من القائمة Analyze الخيار Descriptive Statistics ، ومن القائمة الفرعية الخيار Descriptive يظهر مربع الحوار الذي نقوم من خلاله بإدخال المتغيرات على الشكل انظر الشكل (4-16):

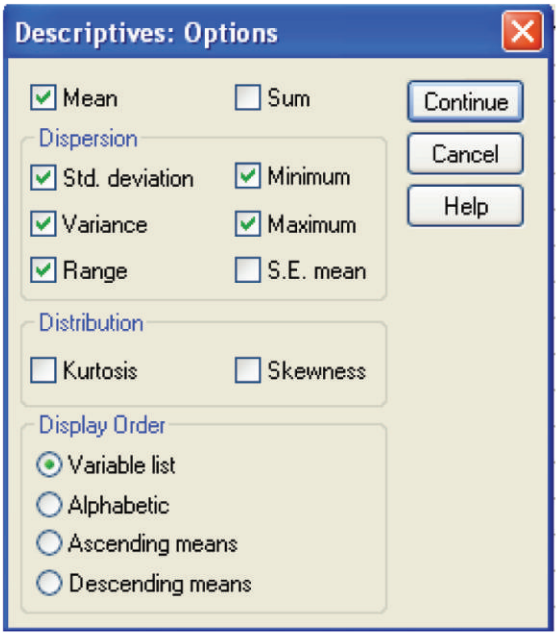


شكل (4-16)

حيث إن:

Save Standardized Values As Variables: لإضافة المتغيرات المعيارية z إلى Data editor عند التأشير عليها في مربع الحوار.

Options: عند نقره يظهر صندوق الحوار انظر الشكل (4-17):



شكل (4-17)

اختر المقاييس المطلوبة.

- ✓ Display Order: يمكننا من خلالها ترتيب عرض المقاييس الوصفية للمتغيرات، وقد اخترنا Variable List: الـذي يعرضـ المقاييس الوصفية حسب تسلسل المتغيرات الـوارد في خانة Variables في صندوق حوار Descriptive.
- ✓ ثم انقر على Continue لنعود لمربع الحوار Descriptive.
- ✓ ثم ننقر على زر Ok في صندوق حوار Descriptive يعرض المخرجات الآتية:

جدول (4-4)

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
X1	13	40.00	50.00	90.00	67.3077	13.21858	174.731
X2	5	15.00	50.00	65.00	56.4000	6.10737	37.300
X3	5	10.00	12.00	22.00	17.6000	4.03733	16.300
Valid N (listwise)	5						

حيـث تم عرض المقاييس الإحصائية المختارة حسب ترتيب المتغـيرات في خانة Variables في صندوق حوار Descriptive، كما تمت إضافة القيم المعيارية للمتغيرات إلى Data editor كما في الشكل الآتي:

X1	X2	X3	ZX1	ZX2	ZX3
90.00	50.00	12.00	1.71670	-1.04791	-1.38706
70.00	52.00	15.00	.20368	-.72044	-.64399
56.00	55.00	19.00	-.85544	-.22923	.34676
65.00	60.00	22.00	-.17458	.58945	1.08983
85.00	65.00	20.00	1.33844	1.40813	.59445
60.00	.	.	-.55284	.	.
69.00	.	.	.12802	.	.
57.00	.	.	-.77979	.	.
50.00	.	.	-1.30935	.	.
75.00	.	.	.58193	.	.
62.00	.	.	-.40153	.	.
51.00	.	.	-1.23370	.	.
85.00	.	.	1.33844	.	.

شكل (4-18)

الفصل الخامس

العرض والتحليل الإحصائي لمتغيرين

View and Statistical Analysis of Bivariate Random Variable

(5-1) مقدمة:

بعد عرض وصف متغير واحد في الفصل السابق يكون من الضروري هنا أن نوضح كيفية وصف متغيرين أو أكثر، حيث نجد أن العلاقة بين متغيرين تكون من أهم الأشياء التي لا غنى عنها في التحليل الإحصائي. ولذلك كان من المهم التعرف على كيفية وصف متغيرين وتوضيح العلاقة بينهما باستخدام أشكال الانتشار والجداول المزدوجة، بالإضافة إلى المقاييس الإحصائية المناسبة لهم.

توجد ثلاث طرق للعرض في حالة متغيرين:

- العرض الجدولي.
- العرض البياني.
- العرض باستخدام أحد المقاييس الإحصائية.

(5-2) الجداول المزدوجة:

الجدول المزدوج هو جدول ثنائي يلخص العلاقة بين متغيرين جدوليًا فيبرز الخصائص المشتركة لهما.

مثال (5-1):

الجدول الآتي يمثل توزيع الطلاب والطالبات المقبولين في كلية الآداب لعام 2007 / 2008م.

جدول (5-1)

التخصص	اللغة العربية	اللغة الإنجليزية	التاريخ	الجغرافيا	الأعلام	الاجتماع	المجموع
عدد الطلاب	74	115	254	109	201	138	891

عدد الطالبات	129	134	155	57	–	485	950
المجموع	203	249	409	155	201	523	1851

يمكن تحويل التكرارات الموجودة في الجدول إلى نسب (لكل قيم الجدول) كالآتي:

- ✓ إذا تمت قسمة جميع أرقام الجدول على مجموع التكرارات، وهو 1851 نحصل على نسبة كل خلية للمجموع الكلي.
- ✓ إذا تمت قسمة تكرارات أي صف على مجموع الصف فسنحصل على نسبة الخلية للصف.
- ✓ إذا تمت قسمة التكرار في كل خلية على مجموع العمود فسنحصل على نسبة الخلية لمجموع العمود.

مثال (5-2):

فيما يلي قائمة ببعض البيانات التي استخرجت من ملفات بعض العاملين في إحدى المصالح الحكومية، الذين يمثلون عينة عشوائية تم اختيارها من بين العاملين في تلك المصلحة الحكومية، التي تمثل مدة الخدمة للعامل والمرتب الشهري.

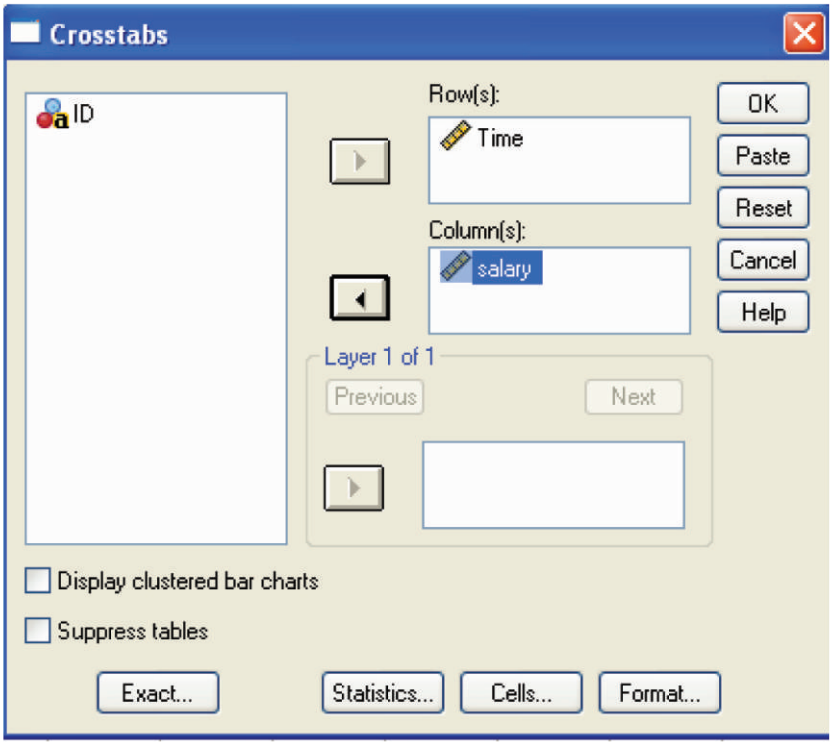
جدول (5-2)

رقم الملف	مدة الخدمة	المرتب الشهري
15/1	0.4	200
2/33	2	240
2/23	1	200
85/1	5	390
5/45	18	520
52/12	5	400
2/125	11	490
5/12	18	490
5/123	9	390
55/12	12	500
12/125	4	400
55/123	0.8	200
89/44	20	390
55/23	19	200

لإنشاء الجدول المزدوج باستخدام الحزمة SPSS لبيانات المثال نمر بالخطوات الآتية:

- ⊙ من القائمة Analyze نختار Descriptive Statistics وننقر الأمر Cross tabs.

فيظهر الصندوق انظر الشكل(5-1):

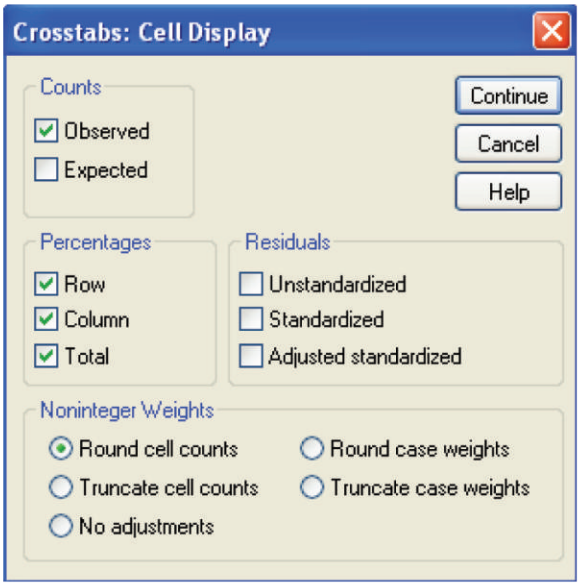


شكل (5-1)

حيث إن:

- Row(s): يتضمن المتغير الذي نرغب في جعل فئاته صفوفًا في الجدول.
- Column(s): يتضمن المتغير الذي نرغب في جعل فئاته أعمدة في الجدول.
- Display clustered bar charts: لتمثيل بيانات الجدول بالأعمدة البيانية المزدوجة.
- Suppress tables: لإخفاء الجداول المزدوجة عند التأشير.

⊙ انقر على Cells فنحصل على الصندوق انظر الشكل (5-2):



شكل (5-2)

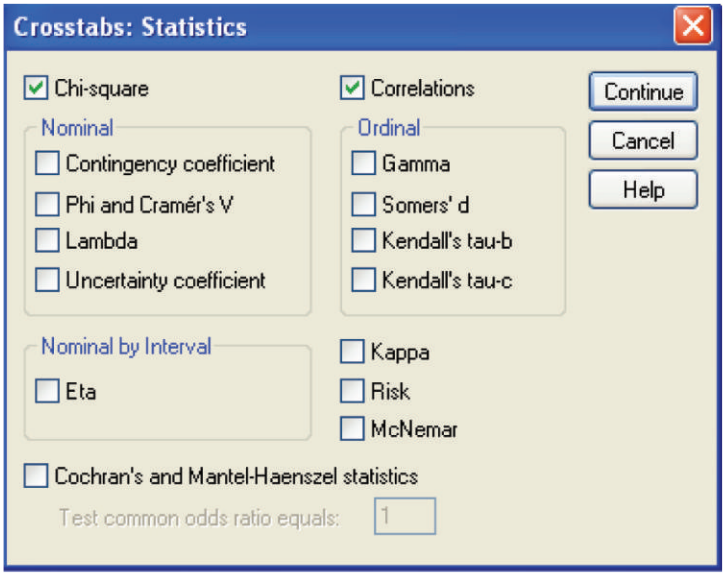
والذي نلاحظ احتواءه على العناصر الآتية:

- Observed: التكرار المشاهد للخلية.
- Expected: التكرار المتوقع للخلية (يستخدم لحساب اختبار الاستقلال).
- Residuals: للحصول على البواقي.

ننقر على continue للعودة للصندوق الحواري الأصلي.

ثم ننقر على الأمر Statistics يظهر الصندوق Cross tabs: Statistics.

انظر الشكل(5-3):



شكل (5-3)

والذي نلاحظ احتواءه على العناصر الآتية:

- Chi-square: لحساب استقلالية الصفوف عن الأعمدة.
- Correlation: يستخدم لحساب معاملي الارتباط Spearman وPearson يحسب Spearman عندما تكون قيم المتغيرات الداخلة عددية Numeric، وباعتبار أن كلا من الصفوف والأعمدة عبارة عن قيم ترتيبية.
- ويحسب Pearson باعتبار أن كلا من الصفوف والأعمدة متغيرات كمية ويعبر عنها بالخيار Scale.
- Nominal: تتيح هذه القائمة حساب أربعة معاملات للارتباط في حالة كون كل من الصفوف والأعمدة عوامل غير كمية، بالإضافة إلى عدم إمكانية ترتيب البيانات مثل متغير الجنس (ذكور/ إناث).
- Ordinal: تتيح هذه القائمة حساب أربعة معاملات للارتباط في حالة كون كل من الصفوف والأعمدة متغيرات ترتيبية.
- Nominal by Interval: لحساب إحصائية Eta للاقتران بين متغيرين أحدهما المعتمد يقاس ضمن فترة interval scale مثل متغير الدخل والآخر مستقل له عدد محدود من الفئات مثل الجنس.

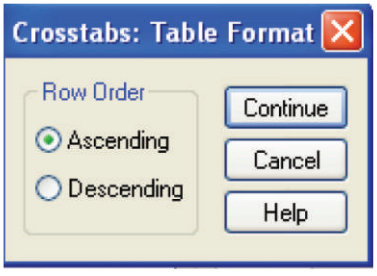
نظرا لكون الصفوف والأعمدة في هذا المثال متغيرات كمية Numeric فقد قمنا بحساب إحصائية Chi-square لاختبار الاستقلالية مع حساب قيمة معامل الارتباط بين المتغيرين.

ننقر على continue للعودة للصندوق الأصلي.

- ثم ننقر على Ok للتنفيذ.

ملاحظة:

- يمكن ترتيب صفوف الجدول تصاعديًا أو تنازليًا بنقر الزر Format في صندوق حوار Cross tabs كما في الشكل الآتي:



شكل (5-4)

بعد تحديـد نوعية الترتيب، هنا تم اختيار الترتيب التصاعدي، نقوم بالنقر على Continuous لنرجع إلى القائمة الظاهرة في الشكل (5-1) وننقر على OK.

النتائج التي تم الحصول عليها في شاشة النتائج:

جدول (5-3)

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Time *salary	14	100.0%	0	.0%	14	100.0%

الجدول السابق به البيانات المتاحة والمفقودة.

جدول (5-4)

Time * salary Crosstabulation

			salary							Total
			200.00	240.00	390.00	400.00	490.00	500.00	520.00	
Time	.40	Count	1	0	0	0	0	0	0	1
		% within Time	100.0%	.0%	.0%	.0%	.0%	.0%	.0%	100.0%
		% within salary	25.0%	.0%	.0%	.0%	.0%	.0%	.0%	7.1%
		% of Total	7.1%	.0%	.0%	.0%	.0%	.0%	.0%	7.1%
	.80	Count	1	0	0	0	0	0	0	1
		% within Time	100.0%	.0%	.0%	.0%	.0%	.0%	.0%	100.0%
		% within salary	25.0%	.0%	.0%	.0%	.0%	.0%	.0%	7.1%
		% of Total	7.1%	.0%	.0%	.0%	.0%	.0%	.0%	7.1%
	1.00	Count	1	0	0	0	0	0	0	1
		% within Time	100.0%	.0%	.0%	.0%	.0%	.0%	.0%	100.0%
		% within salary	25.0%	.0%	.0%	.0%	.0%	.0%	.0%	7.1%
		% of Total	7.1%	.0%	.0%	.0%	.0%	.0%	.0%	7.1%
	2.00	Count	0	1	0	0	0	0	0	1
		% within Time	.0%	100.0%	.0%	.0%	.0%	.0%	.0%	100.0%
		% within salary	.0%	100.0%	.0%	.0%	.0%	.0%	.0%	7.1%
		% of Total	.0%	7.1%	.0%	.0%	.0%	.0%	.0%	7.1%
	4.00	Count	0	0	0	1	0	0	0	1
		% within Time	.0%	.0%	.0%	100.0%	.0%	.0%	.0%	100.0%
		% within salary	.0%	.0%	.0%	50.0%	.0%	.0%	.0%	7.1%
		% of Total	.0%	.0%	.0%	7.1%	.0%	.0%	.0%	7.1%
	5.00	Count	0	0	1	0	0	0	0	1
		% within Time	.0%	.0%	100.0%	.0%	.0%	.0%	.0%	100.0%
		% within salary	.0%	.0%	33.3%	.0%	.0%	.0%	.0%	7.1%
		% of Total	.0%	.0%	7.1%	.0%	.0%	.0%	.0%	7.1%
	6.00	Count	0	0	0	1	0	0	0	1
		% within Time	.0%	.0%	.0%	100.0%	.0%	.0%	.0%	100.0%
		% within salary	.0%	.0%	.0%	50.0%	.0%	.0%	.0%	7.1%
		% of Total	.0%	.0%	.0%	7.1%	.0%	.0%	.0%	7.1%

	9.00	Count	0	0	1	0	0	0	0	1
		% within Time	.0%	.0%	100.0%	.0%	.0%	.0%	.0%	100.0%
		% within salary	.0%	.0%	33.3%	.0%	.0%	.0%	.0%	7.1%
		% of Total	.0%	.0%	7.1%	.0%	.0%	.0%	.0%	7.1%
	11.00	Count	0	0	0	0	1	0	0	1
		% within Time	.0%	.0%	.0%	.0%	100.0%	.0%	.0%	100.0%
		% within salary	.0%	.0%	.0%	.0%	50.0%	.0%	.0%	7.1%
		% of Total	.0%	.0%	.0%	.0%	7.1%	.0%	.0%	7.1%
	12.00	Count	0	0	0	0	0	1	0	1
		% within Time	.0%	.0%	.0%	.0%	.0%	100.0%	.0%	100.0%
		% within salary	.0%	.0%	.0%	.0%	.0%	100.0%	.0%	7.1%
		% of Total	.0%	.0%	.0%	.0%	.0%	7.1%	.0%	7.1%
	18.00	Count	0	0	0	0	1	0	1	2
		% within Time	.0%	.0%	.0%	.0%	50.0%	.0%	50.0%	100.0%
		% within salary	.0%	.0%	.0%	.0%	50.0%	.0%	100.0%	14.3%
		% of Total	.0%	.0%	.0%	.0%	7.1%	.0%	7.1%	14.3%
	19.00	Count	1	0	0	0	0	0	0	1
		% within Time	100.0%	.0%	.0%	.0%	.0%	.0%	.0%	100.0%
		% within salary	25.0%	.0%	.0%	.0%	.0%	.0%	.0%	7.1%
		% of Total	7.1%	.0%	.0%	.0%	.0%	.0%	.0%	7.1%
	20.00	Count	0	0	1	0	0	0	0	1
		% within Time	.0%	.0%	100.0%	.0%	.0%	.0%	.0%	100.0%
		% within salary	.0%	.0%	33.3%	.0%	.0%	.0%	.0%	7.1%
		% of Total	.0%	.0%	7.1%	.0%	.0%	.0%	.0%	7.1%
Total		Count	4	1	3	2	2	1	1	14
		% within Time	28.6%	7.1%	21.4%	14.3%	14.3%	7.1%	7.1%	100.0%
		% within salary	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	28.6%	7.1%	21.4%	14.3%	14.3%	7.1%	7.1%	100.0%

الجدول السابق هو الجدول التكراري المزدوج للمتغيرين الأعمدة لـ Recover والصفوف لـ .Treat

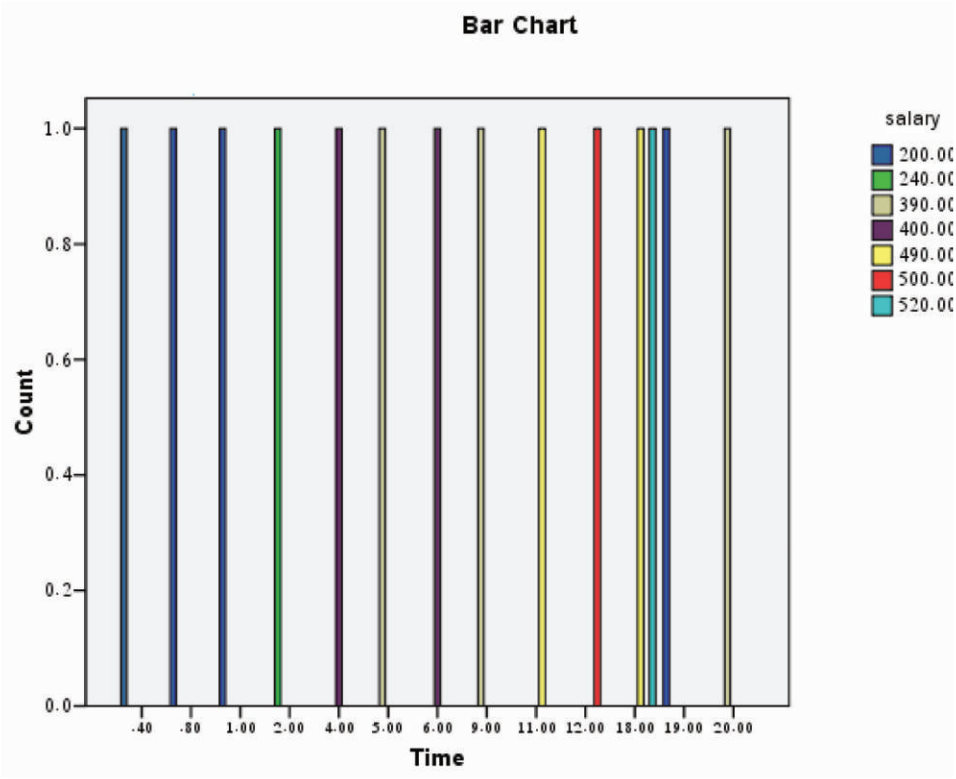
- Count: هو التكرار المشاهد.
- within treat %: النسبة للصف.
- Within recover %: النسبة للعمود. ثم للكلي total of %.

جدول (5-5)

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	73.500[a]	72	.429
Likelihood Ratio	47.894	72	.987
Linear-by-Linear Association	3.758	1	.053
N of Valid Cases	14		

a. 91 cells (100.0%) have expected count less than 5. The minimum expected count is .07.



شكل (5-5)

✓ تم الحصول على رسمة الأعمدة البيانية المزدوجة للمتغيرين معا، وهي أعمدة متلاصقة كل عمود مخصص لبيانات متغير معين وله لون مختلف، حيث إن تكرار مدة الخدمة واحد.

(5-3) شكل الانتشار:

شكل الانتشار ما هو إلا تمثيل للبيانات المزدوجة بنقاط على محورين أحدهما المحور السيني ويمثل عليه المتغير المستقل Independent والآخر المحور الصادي ويمثل عليه المتغير التابع Dependent. وبالنظر إلى شكل الانتشار نستطيع تحديد نوع العلاقة ودرجتها تقريبًا طبقًا للضوابط الآتية:

- إذا تجمعت النقاط حول خط يصنع زاوية حادة أو منفرجة مع المحور الأفقي تكون العلاقة خطية، وإذا تجمعت البيانات حول أي شكل آخر غير الخط تكون العلاقة غير خطية.
- إذا كانت العلاقة خطية واقتربت النقاط حول خط وهمي يمكن تخيله وتحديد ما إذا كانت العلاقة قوية والعكس صحيح.
- إذا كانت العلاقة خطية (أم غير خطية) ووقعت النقاط في شكل الانتشار جميعها فوق الخط كانت العلاقة تامة، وهذه أشد صورة من صور الارتباط.

مثال (5-3):

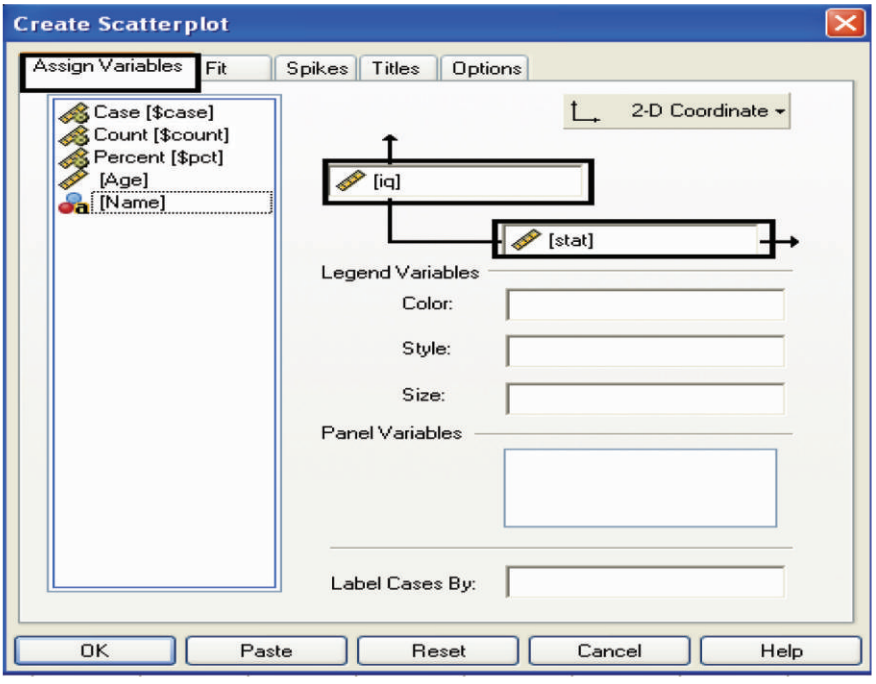
نفرض أن لدينا بيانات خاصة بنتائج 25 طالبًا في اختبار للذكاء، ودرجة مقرر الإحصاء مقررين يرمز لهما بالرمزين IQ ، Stat

جدول (5-6)

الاسم	العمر	STAT	IQ
Ahmad	21 ,00	40 ,00	50 ,00
Hanna	30 ,00	59 ,00	70 ,00
Heba	21 ,00	55 ,00	50 ,00
Kamala	22 ,00	85 ,00	90 ,00
Mervat	27 ,00	77 ,00	80 ,00
Mohamed	23 ,00	80 ,00	83 ,00
Samer	22 ,00	35 ,00	50 ,00
Satar	25 ,00	55 ,00	50 ,00
Shimaa	21 ,00	50 ,00	80 ,00

باستخدام البرنامج يمكن رسم شكل الانتشار كالآتي:

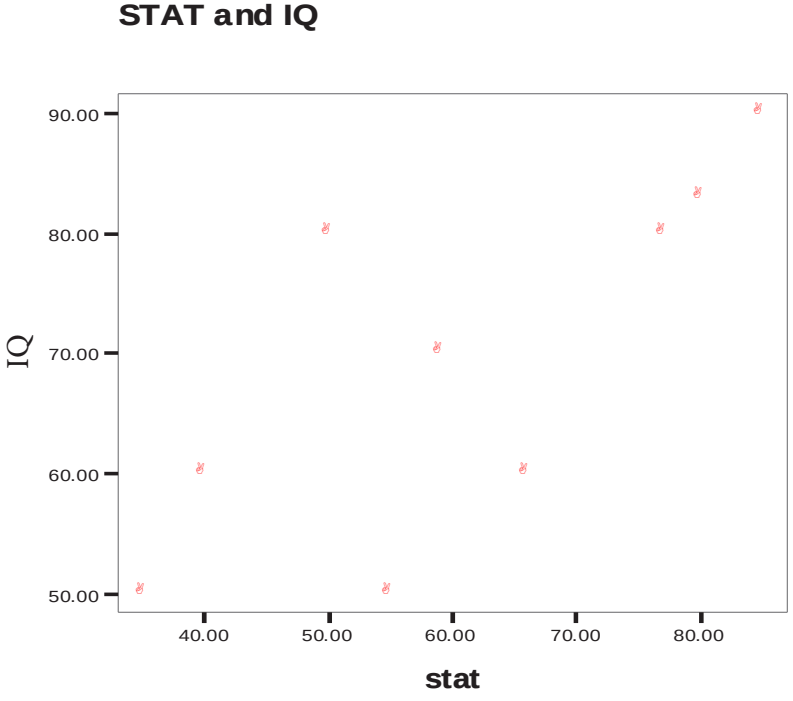
✓ من القائمة Graphs انقر الأمر interactive ثم انقر Scatter plot يفتح الصندوق Scatter plot انظر الشكل (5-6).



شكل (5-6)

انقـر علـى الأمر Assign Variables وذلك لإدخال المتغيرات التـي نرغب في إيجاد شكل الانتشار لها، وليكن المتغير iq المستطيل الخاص بالمحور الرأسي Y Axis ، والمتغير stat إلى مستطيل المحور الأفقي X Axis كما هو واضح في الشكل السابق.

- انقـر علـى الأمر Title لكتابـة عنوان الشكل المرسوم ثـم انقر continue للعـودة للصندوق الأصلي وليكن نريده أن يكون STAT and IQ.
- انقر على الأمر options لتحدد موقفك عن كيفية معالجة القيم المفقودة.
- انقر على الأمر continue لتعود للصندوق الحواري الأصلي.
- انقر على Ok لتنفيذ الأمر.



شكل (5-7)

يتضـح من شكل الانتشـار أن البيانات متجمعة حول خط يصنع زاوية حـادة مع المحور الأفقي، أي أن العلاقة بين المتغيرين يمكن أن تكون خطية طردية، وكذلك العلاقة غير تامة لأن النقاط جميعها لم تتجمع فوق خط معين، لا يمكن معرفة قيمة معامل الارتباط من الشكل، ولكن يمكن تصور اقترابه من رقم معين. وسوف نتعرض في الفصول القادمة لكيفية حساب معامل الارتباط للبيانات المختلفة.

الفصل السادس

اختبارات الفروض

Test Hypothesis

(6-1) مقدمة :

يحتـاج الباحث في مرحلة ما من بحثه إلى اتخاذ قرار بناءً على معلومات محسوبة من عينة، وطبيعي أن يتخذ هذا القرار بشيء من الحكمة وبأقل قدر ممكن من المخاطر، إذ إن هذا القرار تترتب عليه نفقات قد تكون طائلة، ومن ثم لابد أن يكون لها ما يبررها.

وبصفة عامة فإن خطوات المنهج العلمي في البحث هي:

- تحديد المشكلة تحديداً دقيقًا، حيث يعد ذلك أهم الخطوات على الإطلاق.
- وضـع الفروض، وهـي مرحلة الربط بين هـذه المعلومات، وذلك لمعرفة الأسبـاب الحقيقية وليست الظاهرية للمشكلة.
- الفروض: Hypothesis: إحـدى ضـرورات الحياة العلميـة، وهي عبارة عن حلـول مقترحة لعلاج أسبـاب مشكلة تحت الدراسة. وتنشأ الفروض، أي الحلول المقترحة كنتيجة لملاحظات الباحث وما حصل عليه من معلومات بخصوص تلك المشكلة.
- اختبـار صحـة الفروض، حيث يتـم اختبار صحـة الفرض بالعمـل التجريبي وأخـذ الملاحظات، وباستخـدام أدوات التحليـل المختلفة فتستبعد الفروض عديمة الأثـر وتستبقى الفروض التي ثبتت قدرتها على التأثير في أسباب المشكلة وعلاجها.

(6-2) تعاريف أساسية في اختبارات الفروض الإحصائية

الآتي هو مجموعة من المصطلحات التي تستخدم في اختبارات الفروض الإحصائية:

- الفرض H_A (Hypothesis):

هو تقرير أو تأكيد عن التوزيع الاحتمالي و قد يكون هذا التقرير أو التأكيد متعلقًا بأحد معالم التوزيع الاحتمالي أو بأحد المؤشرات التي تعكس خاصية معينة من خواص التوزيع الاحتمالي.

- الفرض العدمي (Null Hypothesis):
- فرضـ العـدم (الفرض الصفري) نرمز لـه بالرمز H_0، وهو فرض حول معلمـة المجتمع التي يتم إجـراء الاختبـار حولها باستخـدام بيانات تم الحصول عليهـا من العينة، وهـذا الفرض نرفضه عندمـا تتوافر لدينا دلائل على عدم صحتـه، وتعني كلمة Null أنه لا يوجد فرق بين معلمة المجتمع وقيمة إحصاء العينة.

⊙ الفرض البديل H_A (Alternative Hypothesis):

يرمـز لهذا الفرض بالرمز H_A ويضعه الباحـث كبديل عن فرض العدم، ونقبله عندما نرفض فرض العدم بوصفه ليس صحيحاً بناءً على المعلومات التي تم الحصول عليها من العينة.

⊙ اختبار الفرض الإحصائي (Test of A Statistical Hypothesis):

يعرف اختبار الفرض الإحصائي بأنه القاعدة التي يتم اتخاذها بناءً على دراسة مفردات العينة التي تـؤدي إلى قـرار معين بشأن قبول أو رفضـ فرض العدم، ويلاحظ أن القرار الذي يتـم اتخاذه يتعلق عادة بفرض العدم أي H_0، فإما أن يكون القرار هو رفض H_0 (وفي هذه الحالة يعني ذلك قبول H_A) أو قبول H_0 (وفي هذه الحالة يعني ذلك رفض H_A).

⊙ الخطأ من النوع الأول والخطأ من النوع الثاني (Type 1 error and Type 2 error)

يعتمـد القـرار الذي يتم اتخاذه بشأن فرض العدم على النتيجة التـي تعطيها العينة العشوائية، وهنا قـد تعطـي العينة معلومات أو نتيجة تؤدي إلى رفض فرض العدم برغم كونه في الحقيقة صحيحاً، وفي هذه الحالـة يكـون القرار خاطئًا، أو قـد تعطي العينة معلومات في صالح فرض العـدم فيتم قبوله رغم كونه في الحقيقة غير صحيح، وهذا قرار خاطئ آخر.

وحيث إن القرار الذي يتم اتخاذه بشأن فرض العدم يعتمد على المعلومات التي يتم الحصول عليها من العينة العشوائية، فإن الحالات الممكنة للقرار الذي يتخذه بشأن فرض العدم هي كما يلي:

القـــرار		على مستوى المجتمع	
		H_0 صحيح	H_0 غير صحيح
على مستوى العينة	قبول H_0	القرار صحيح	خطأ من النوع الثاني
	رفض H_0	خطأ من النوع الأول	القرار صحيح

⊙ أنواع اختبارات الفروض:

عندما نقبل فرض العدم فإننا نقبله بمستوى دقة معين فقد تكون 0,90 أو 0,95 أو 0,99 أو غير ذلك، وتسمـى مستويـات الثقة (Significance Levels)، أي توجد نسبة خطأ معين في قبولنا للفرض العدم ونرمز لها بالرمز α ويسمى مستوى المعنوية، أي إذا كان مستوى الثقة ($1-\alpha = 0.95$) فإن مستوى المعنوية α تساوي 0,05، وهي عبارة عن مساحة منطقة تحت منحنى التوزيع تمثل منطقة الرفض، وتكون إما على صورة ذيل واحد جهة اليمين أو اليسار أو ذيلين متساويين في المساحة، واحد جهة اليمين والثاني جهة اليسار.

⊙ **اختبار الفروض في جانب واحد:**

هـو اختبـار يضع الفرض البديل للمعلمة المجتمع في صورة أكبر أو أصغر من قيمة فرض العدم لتلك المعلمـة، فتكون منطقة الرفض إما من اليمـين إذا كان الفرض البديل أكبر، أو من اليسار إذا كان الفرض البديل أقل.

⊙ **تعريف اختبار الفروض في جانبين (ذيلين):**

هو اختبار يكون فيه الفرض البديل عن معلمة المجتمع في صورة لا تساوي (≠) قيمة فرض العدم لتلك المعلمة.

⊙ **خطوات اختبارات الفروض الإحصائية لحجم ثابت للعينة:**

عند إجراء اختبارات الفروض الإحصائية عملياً تكون هناك خطوات محددة لإجراء الاختبار، حيث تكـون لازمـة وبالترتيب المحدد لكل منها لإجراء الاختبار واتخاذ القـرار بطريقة سليمة. والخطوات الآتية تحدد كيفية إجراء اختبارات الفروض الإحصائية التي تعتمد على حجم ثابت للعينة:

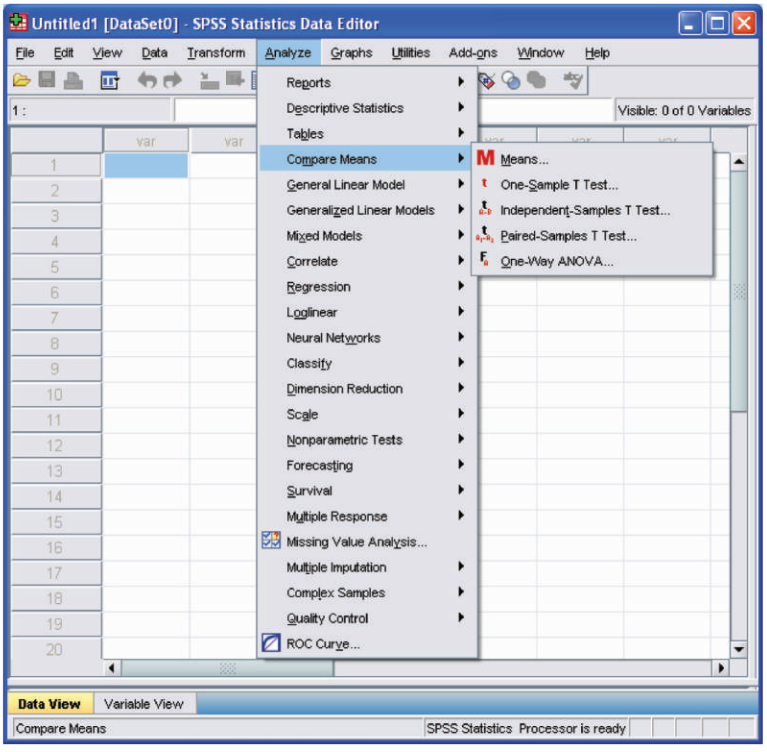
- صياغـة كل من الفرض العدم H_0 والفرض البديل H_A حسب المشكلة المراد إجراء اختبارات الفروض لها.
- تحديـد مستوى المعنوية للاختبار بما يناسب تكلفة الوقـوع في الخطأ من النوع الأول أو الخطأ من النوع الثاني.
- تحديـد دالـة الاختبار وهي إحصائيـة، أي أنها تعتمد فقط على مشاهـدات العينة أو الفرض العـدم، وتختـار بحيث يكون توزيعها الاحتمـالي معلومًا تمامًا تحت فرض العـدم، أي أن دالة الاختبار ما هي إلا دالة في مشاهدات العينة توزيعها الاحتمالي معلوم.
- تكويـن قاعدة القرار، وتعتمـد على كل من الفرض البديل ومستـوى المعنوية، ونقصد بقاعدة القـرار هي القاعدة التـي يبنى عليها اتخاذ القـرار، إما بقبول فرض العـدم (إذا وقع خارج المنطقـة الحرجـة أي منطقة الرفـض للاختبـار)، أو يتم اتخـاذ قرار برفـض فرض العدم (إذ وقعت داخل المنطقة الحرجة أي منطقة الرفض).
- حساب قيمة دالة الاختبار.
- اتخـاذ القرار: وهنا توجد طريقتان لاتخاذ قرار في الاختبارات الإحصائية، وذلك إما برفض أو قبول الفرض العدمي.

⊙ **طريقة رفض أو قبول الفرض العدمي:**

⊙ تعتمد طريقة قبول أو رفض الفرض العدمي على احتمال محسوب يسمى P-Value ويرمز له في الحزمة SPSS بالرمز (Sig.)، وتتلخص هذه الطريقة في الآتي:

- يجرى الاختبار بالخطوات نفسها السابق الحديث عنها، و نستخدم احتمالاً محسوبًا يرمز له بالرمز Sig.، تعبر هذه القيمة عن مساحة معينة في ذيل (أو ذيلين) للتوزيع العيني لإحصائي الاختبار، وتسمى قيمة مستوى المعنوية المحسوب.
- إذا كان الاختبار من طرفين تقارن هذه القيمة بقيمة $\alpha / 2$، إذا زادت القيمة المحسوبة عن ($\alpha / 2$) نقبل الفرض العدمي والعكس صحيح. إذا كان الاختبار من طرف واحد نقارن القيمة المحسوبة بقيمة α، إذا زادت عنها نقبل الفرض العدمي والعكس صحيح.

⊙ تظهر أوامر اختبارات معلمية بالنقر على Analyze في شريط الأوامر كما في الشكل الآتي:



شكل (6-1)

(3-6) اختبار عينة واحدة

لإجراء اختبار متوسط عينة نفرض أن متوسط المجتمع (المعلمة) مجهول ونريد اختبار بعض الفروض التي تدور حول المعلمة المجهولة، تحديد إحصائي الاختبار سيتوقف على هل تباين المجتمع المسحوبة منه العينة معروف (فرض نظري لا يمكن تحقيقه) أم غير معروف؟ إذا كان تباين المجتمع معروفا فإن التوزيع العيني الذي يتحكم في عمليات الرفض والقبول هو التوزيع الطبيعي، إذا لم يكن تباين المجتمع معروفًا يكون البديل الوحيد للتوزيع الطبيعي هو توزيع t.

⊙ وتوزيع T هو أحد التوزيعات المتماثلة ويشبه كثيرًا التوزيع الطبيعي، يؤول هذا التوزيع إلى التوزيع الطبيعي عندما يزيد حجم العينة (أكبر من 30 مفردة)، معنى ذلك أنه في الحالات التي يطبق فيها إحصائي الاختبار t يستبدل بالتوزيع الطبيعي عندما يزيد حجم العينة على 30 مفردة. اختبار t وهو اختبار معلمي يستخدم هذا الاختبار لفحص فرض يتعلق بالوسط الحسابي.

◀ الفروض الإحصائية لاختبار T:

فرض العدم: $H_0: \mu = \hat{\mu}$

الفرض البديل: $H_A\ \mu \neq \hat{\mu}$ أو $H_A: \mu > \hat{\mu}$ أو $H_A: \mu < \hat{\mu}$

حيث: (μ) هي متوسط المجتمع وتقرأ "ميو".

$(\hat{\mu})$ هي قيمة متوسط العينة وتقرأ "ميو هات".

◀ شروط استخدام توزيع t:

✓ يجب أن يتبع توزيع البيانات التوزيع الطبيعي، ويعوض عن هذا الشرط بزيادة حجم العينة إلى أكثر من 30 مشاهدة.

✓ يجب أن تكون العينة عشوائية أي لا تعتمد مشاهداتها على بعضها.

◀ إحصاءة t المستخدمة في الاختبار:

$$t = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}}$$

حيث: $\overline{X}$ متوسط العينة وتقرأ "إكس بار"

S الانحراف المعياري للعينة،

n حجم العينة.

◀ الفرق بين اختبار t واختبار Z:

في حالة العينات الصغيرة (n<30) وتباين المجتمع مجهول يتم استخدام اختبار t، وفي حالة العينات الكبيرة ($n \geq 30$) يتم استخدام اختبار Z. أي أن توزيع t يؤول إلى توزيع Z عندما يزيد حجم العينة على 30 مشاهدة.

مثال (6-1):

قررت لجنة فحص طلبية من أوراق الطباعة من النوع 80 جرامًا للتأكد من أن المتوسط الحسابي لعدد الأوراق في الرزمة الواحدة لا يساوي 500 ورقة، أخذت عينة عشوائية من الطلبية من عشر رزم (10 رزم) وتم وزن كل رزمة فكانت النتائج كالآتي:

جدول (6-1)

الرزمة	1	2	3	4	5	6	6	8	9	10
الوزن بالكليو جرام	3,9	4,1	3,8	3,6	4	4,2	3,9	3,9	4	3,6

المطلوب ما يلي:

هل ترفض اللجنة الطلبية بناء على هذه المعلومات عند مستوى معنوية 0,05 و0,01.

الحل:

إذا كانت الرزمة تحتوي فعلاً على 500 ورقة ووزن الورقة الواحدة 80 جرامًا فإن وزن الرزمة يجب أن يساوي 4 كجم (500×80 جرامًا = 4000 جرام)، وبفرض أن وزن الرزمة يتوزع كالتوزيع الطبيعي فإن قرار اللجنة يجب أن يكون مبنياً على الاختبار الإحصائي الآتي:

⊙ الفروض الإحصائية:

H_0 : متوسط وزن الرزمة = المتوسط المفترض 4 كجم.

H_A : متوسط وزن الرزمة لا يساوي المتوسط المفترض (الاختبار من طرفين).

$$H_0 : \mu = 4\ km$$
$$H_A : \mu \neq 4\ km$$

إحصاءة الاختبار هي:

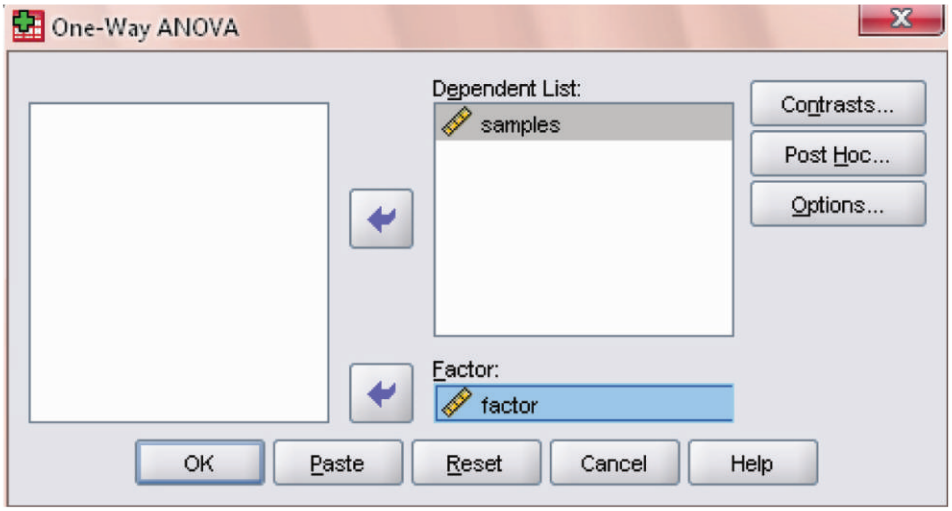
$$t = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}}$$

(لأن تباين المجتمع مجهول وحجم العينة أقل من 30)

حيث $\overline{X}$ يمثل الوسط الحسابي للعينة، وS يمثل الانحراف المعياري للعينة، n حجم العينة، و μ الوسط الحسابي للمجتمع بموجب فرضية العدم ويساوي 4، وهذه الإحصاءة تتبع توزيع T بدرجة حرية $\upsilon = n - 1$.

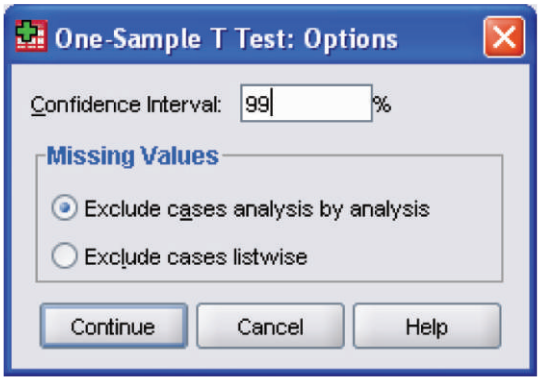
لتنفيذ الاختبار نتبع الخطوات الآتية:

← من شريط القوائم نختار Analyze، ومن قائمة الأوامر نختار Compare Means ، ونختار الأمر الفرعي One Sample t-test فيظهر صندوق حوار One Sample T-test انظر الشكل (6-2):



شكل (6-2)

← انقر Options فيظهر صندوق حوار Options الذي نرتبه انظر الشكل (6-3):



شكل (6-3)

يمكن بواسطة Options تحديد فترة الثقة، حيث نقوم بكتابة الفترة 99% في حقل Confidence Interval، وذلك لحساب 1% مستوى معنوية فيترتب عليه درجة ثقة مقدارها 99%.

أما Missing Values فيستفاد منه في التعامل مع القيم المفقودة.

ننقر Ok في صندوق حوار One sample T-test فتظهر النتائج الآتية:

جدول (6-2)

One-Sample Statistics

Std. Error Mean	Std. Deviation	Mean	N	
.05121	.16193	3.9200	10	Weight

جدول (6-3)

One-Sample Test

	Test Value = 4					
	t	df	Sig. (2-tailed)	Mean Difference	99% Confidence Interval of the Difference	
					Lower	Upper
Weight	-1.562	9	.153	-.08000	-.2464	.0864

القرار الإحصائي:

حيث إن sig.>0.025 و sig>0.005 إذن نقبل الفرض العدم لمستوى معنوية 0,05 و0,01 .

✓ ملحوظة:

في المثال السابق حيث إن الاختبار من طرفين فقد تمت مقارنة sig. بقيمة $\frac{\alpha}{2}$.

(6-4) اختبار الفروض حول الفرق بين متوسطي مجتمعين مستقلتين:

في كثير من التطبيقات يكون الباحثون مهتمين بعمل اختبارات لفرضيات بحثية متعلقة بالفرق بين متوسطي مجتمعين مستقلين، وذلك للمقارنة بين المجتمعين. فعلى سبيل المثال، قد يكون الباحث مهتمًا بمعرفة ما إذا كان دخل الأسرة التي تسكن المدينة أكبر من دخل الأسرة التي تقطن القرية. يمكن للباحث التحقق من هذا التساؤل من خلال مقارنة متوسط دخول الأسر التي تسكن المدن ومتوسط دخل الأسر التي تقطن القرى.

الفروض الإحصائية: يمكن أن تأخذ الفروض الإحصائية أحد الأشكال الآتية:

$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \geq \mu_2$	$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \leq \mu_2$	$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$

عند اختيار Independent-Sample T- Test تجب مراعاة الشروط الآتية:

- ✓ العينتان مستقلتان.
- ✓ البيانات في كل عينة تتبع التوزيع الطبيعي.
- ✓ تم اختيار العينتين عشوائياً.
- ✓ ليس من الضروري أن تتساوى أحجام العينتين.

المجتمعان المسحوبة منهما العينتان متجانسان (تباين المجتمع المسحوب منه العينة الأولى يساوي تباين المجتمع المسحوب منه العينة الثانية)، وللتأكد من التجانس يجب إجراء اختبار سابق لاختبار T يسمى اختبار التجانس، ويكون الفرض العدمي والبديل للاختبار كالآتي:

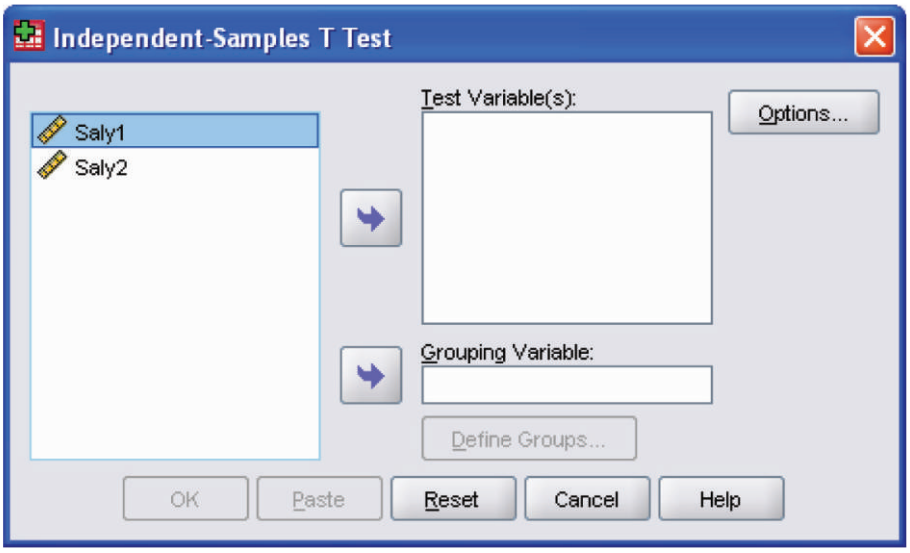
- ⊙ الفرض العدمي هناك تجانس (تباين المجتمع الأول يساوي تباين المجتمع الثاني).
- ⊙ الفرض البديل لا يوجد تجانس (تباين المجتمع الأول لا يساوي تباين المجتمع الثاني)، الذي يأخذ الشكل الآتي:

$$H_o: \sigma_1^2 = \sigma_2^2 \quad H_A: \sigma_1^2 = \sigma_2^2$$

ملحوظة:

- ✓ نستمر في إجراء اختبار T إذا تم قبول فرض العدمي (هناك تجانس).
- ✓ لا يجوز استخدام اختبار ويستبدل باختبار آخر شبيه باختبار T إذا تم قبول الفرض البديل (عدم وجود التجانس).
- ✓ البرنامج يجري اختبار التجانس ويعطي نتائج اختبار T إذا كان هناك تجانس ويعطي أيضًا نتائج لاختبار الشبيه باختبار T إذا لم يكن هناك تجانس، وعلى المستخدم اختيار النتيجة الملائمة له.
- ✓ اختبار التجانس يتم بواسطة اختبار آخر يسمى F-test.

وعند تفعيل هذا الخيار يظهر الشكل الآتي:



شكل (6-4)

المثال الآتي يوضح طريقة تطبيق هذا الاختبار من خلال البرنامج:

مثال(6-2):

أُخـذت عينتـان عشوائيتان من مجموعة متشابهة من الأطفال، وأعطي أطفال العينة الأولى غذاء (X) وأعطـي أطفـال العينة الثانية غذاء آخر (Y)، وكانت الزيـادة في أوزان الأطفال بالكيلو جرام في العينتين بعد مدة معينة كما في الجدول الآتي:

جدول (6-4)

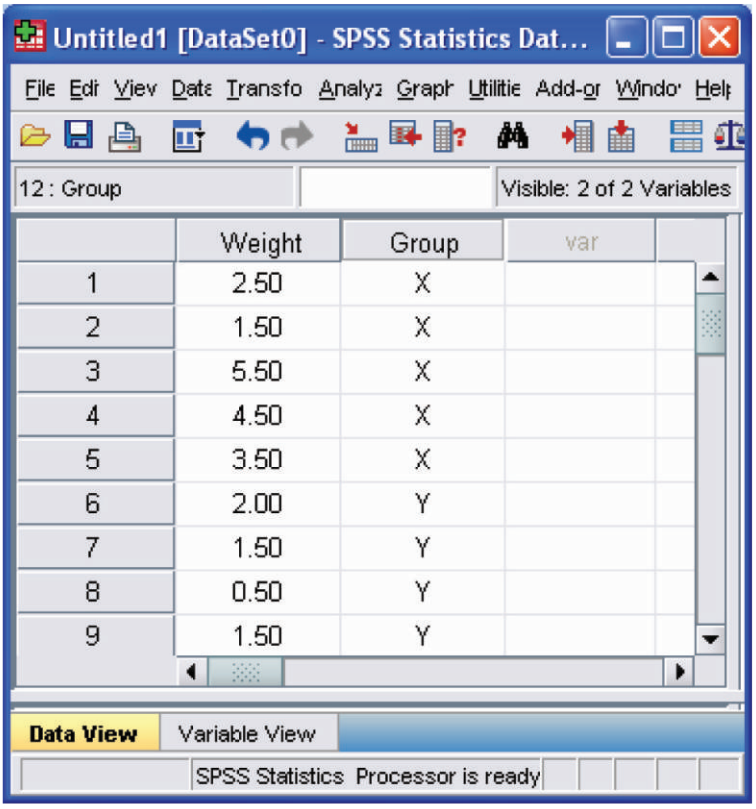
العينة الأولى (X)	2,5	1,5	5,5	4,5	3,5	--
العينة الثانية (Y)	2	1,5	0,5	1,5	1	2,5

اختبر فرض عدم وجود فرق بين أثر الغذائين "X" و"Y" في متوسط زيادة وزن الأطفال عند مستوى معنوية 0,05.

الحل:

لإجراء هذا الاختبار نتبع الخطوات الآتية:

يجـب إدخـال البيانـات بطريقة معينـة في شاشة المحـرر، وهي الطريقـة التي يتعامل بهـا البرنامج كالآتي:

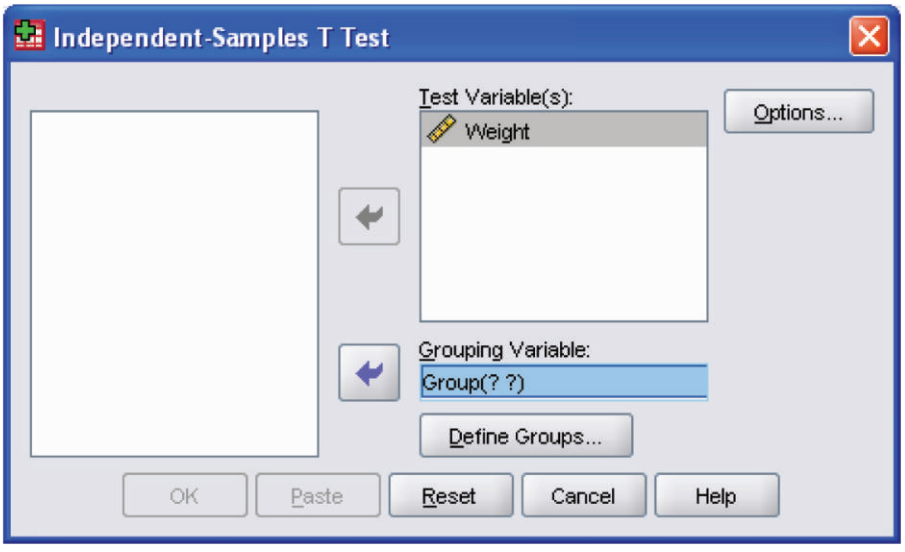


شكل (6-5)

ويتم ذلك كما يلي:

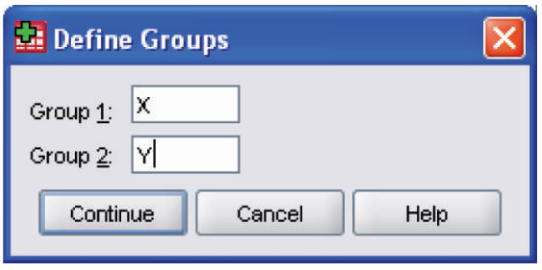
← مـن القائمة Analyze ننقـر الأمر Compare Means ثم ننقر الأمر الفرعي Independent sample test ليفتح صندوق حوار Independent sample T-test

← الذي نقوم بترتيبه كالآتي انظر الشكل (6-6):



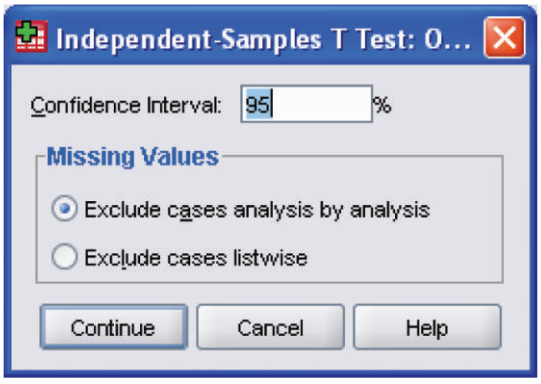
شكل (6-6)

قمنا بنقل المتغير Weight إلى test variable والمتغير Group (متغير التجزئة) في قائمة Grouping variables، ثم تعريف المجاميع X وY عن طريق الزر Define Groups كالآتي انظر الشكل (6-7):



شكل (6-7)

يمكن بواسطة النقر على Options تحديد فترة الثقة، حيث نقوم بكتابة الفترة المطلوبة في حقل Confidence Interval وهنا فترة ثقة مقدرها 95٪ عن طريق الشكل الآتي:



شكل (6-8)

أما Missing values فيستفاد منه في التعامل مع القيم المفقودة.

انقر على Continue للعودة إلى الصندوق الأصلي.

انقر على الأمر Ok للتنفيذ.

تم الحصول على النتائج الآتية:

الجدول الآتي يوضح حجم المجموعة والوسط والانحراف المعياري والخطأ المعياري لكل مجموعة.

جدول (6-5)

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Weight	X	5	3.5000	1.58114	.70711
	Y	6	1.5000	.70711	.28868

الجدول الآتي به اختباران، الأول اختبار التجانس والثاني اختبار t وهو كالآتي:

جدول (6-6)

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Weight	Equal variances assumed	3.165	.109	2.803	9	.021	2.00000	.71362	.38567	3.61433
	Equal variances not assumed			2.619	5.326	.044	2.00000	.76376	.07229	3.92771

ويوضح الجدول من اليسار إلى اليمين الآتي:

- العمـود الأول يسـارا به اسم المتغـير Weight والعمود الثاني والثالث يسار لإجـراء اختبار التجانس كالآتي:

- فرض العدم: هناك تجانس Equal Variances.
- الفرض البديل: هناك عدم تجانس Equal Variances not assumed.
- قيمـة مستوى المعنوية المحسوبـة Sig.=0.109، وهي تزيد على مستوى المعنوية المحدد 0.05 ومن ثم نقبل الفرض العدمي القائل بأن هناك تجانسًا ونرفض الفرض البديل.
- العمود الرابع والخامس لإجراء إما اختبار t أو الاختبار الشبيه باختبار t، وحيث إننا قبلنا أن هناك تجانسًا نتعامل مع اختبار t والنتائج في الصف الأول من الجدول ونهمل الثاني؛ لأنه خاص بالاختبار الشبيه باختبار t.
- من الصف الأول (Equal variances assumed) نجد نتائج اختبار t، قيمة مستوى المعنوية المحسوبة هي Sig.=0.021 وهي أقل من 0.025 (0.05 مقسومة على 2 لأن الاختبار من طرفين)، إذن نرفض الفرض العدمـي القائل إنه يوجد فرق بين أثر الغذائـين X وY في متوسط زيادة وزن الأطفال عند مستوى معنوية 0.05.

(3-6) اختبـار الفروض حول الفرق بـين متوسطي عينتين غـير مستقلتين (العينـات المزدوجة) Paired Sample Test:

(4-6) استخدمنا في الاختبار السابق اختبار الفرق بين عينتين مستقلتين، بمعنى أن مفردات العينة الأولى ($x_1, x_2, \ldots, x_r$) مستقلـة عن مفـردات العينة الثانيـة ($y_1, y_2, \ldots, y_r$). ولكن قد يحدث في الحياة العمليـة أن المشاهـدات تكون على شـكل أزواج مرتبة من القيـم $[(x_r, y_r), r = 1,2,\ldots,n]$، حيث x_1 تمثـل قيمة المشاهدة للعينـة الأولى مأخوذة من المفردة رقم 1 ، و y_1 تمثل قيمة المشاهدة للعينة الثانية المأخوذة من المفردة رقم 1 وبصفة عامة فإن x_r تمثل قيمة المشاهدة العينة الأولى المأخوذة عـن المفـردة رقمr، بينما y_r تمثل قيمة المشاهدة للعينة الثانية المأخـوذة عن المفردة رقم r أي أن n مـن المفـردات وكل مفردة مسجل لها مشاهدتـان (أو قيمتان أو قياسيتـان) أولهما للعينة الأولى، وثانيهما للعينة الثانية. مثل تسجيل قياسات ضغط الدم لكل شخص في عينة مكونة من 10 أشخاص مرة قبل تعاطي دواء معين ومرة أخرى بعد ذلك الدواء بعد أسبوعين أو ما يطلق عليه في الدراسات التجريبيـة الاختبـار القبلي (Pre test) والاختبار البعـدي (Post test) على العينة نفسها. وبصفة عامـة إذا كان لدينـا n من أزواج المشاهدات المرتبـة $(x_r, y_r), r = 1,2,\ldots\ldots,n$ مأخوذة على n من المفـردات فإنه يمكن اعتبار أن: $x_1, x_2, \ldots\ldots, x_N$ عينة مسحوبة من مجتمع متوسطه μ_1 وأن $y_1, y_2, \ldots\ldots, y_N$ عينة مسحوبة من مجتمع متوسطه μ_2.

يستعمـل هذا الاختبار لاكتشاف معنويـة (دلالة) الفرق بين متوسطي متغيرين غير مستقلين لمجموعة (عينة) واحدة، معنى ذلك أن لدينا عينة واحدة، وكل مفردة في العينة تعطي قراءتين، القراءة الأولى تمثل العينة الأولى، والقراءة الثانية تمثل العينة الثانية.

يعتمـد الاختبار على إيجاد الفـرق بين القراءات المتناظرة في العينتين (نرمـز إلى مجموع الفرق بين العينتين بالرمز d).

وتكون صياغة الفروض كالآتي:

فرض العدم $H_0: d = 0$

الفرض البديل $H_A: d \neq 0$

شروط الاختبارات للعينتين المترابطتين:

الشروط الأساسية لاختبار t.

الفرق d بين العينتين يتبع التوزيع الطبيعي، وذلك إذا كان ($n < 30$).

أما إذا كان ($n \geq 30$) فلا نحتاج لهذا الشرط.

المثال الآتي يوضح كيفية تطبيق هذا الاختبار باستخدام البرنامج:

مثال(6-3):

عقـدت دورة تدريبيـة لمجموعة من الموظفـين، ولمعرفة مدى استفادة المتدربين مـن الدورة أجري لهم اختباران الأول (القبلي) في بداية الدورة، والثاني (البعدي) في نهايتها. وفيما يلي بيان بنتيجة الاختبارين لعينة عشوائية مكونة من تسعة من المتدربين:

جدول (6-7)

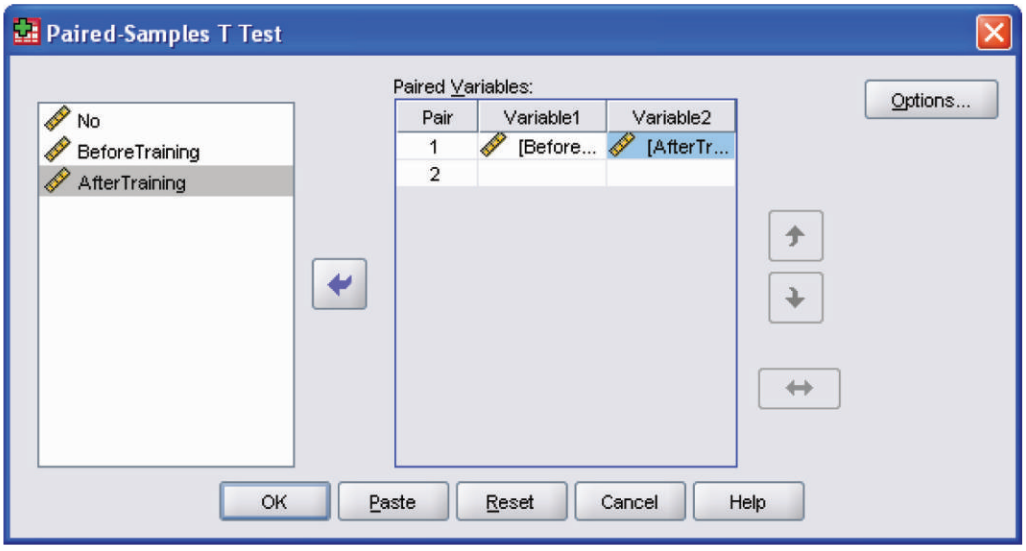
رقم المتدرب	1	2	3	4	5	6	6	8	9
الدرجة قبل التدريب	6	9	4	6	6	8	6	11	6
الدرجة بعد التدريب	15	18	16	18	14	16	18	20	19

اختبر فرض أن الدورة التدريبية ليس لها أثر في رفع مستوى المتدربين عند مستوى معنوية 5 ٪ .

الحل:

لتنفيذ الاختبار نتبع الخطوات الآتية:

مـن القائمة Analyze انقـر الأمر Compare Means ثم انقر الأمر الفرعـي Paired – Sample T test فيظهر صندوق حوار Paired – Sample T test الذي نقوم بترتيبه انظر الشكل (6-9):



شكل (6-9)

المتغيرات قبل التدريب وبعد التدريب تمثل مشاهدات النوعين قبل التدريب وبعده، حيث يجب إدخال كلا المتغيرين في الوقت نفسه في قائمة Paired variables.

يمكن بواسطة النقر على Options تحديد فترة الثقة، حيث نقوم بكتابة الفترة المطلوبة في حقل Confidence Interval.

أما Missing values فيستفاد منه في التعامل مع القيم المفقودة.

انقر على Continue للعودة إلى الصندوق الأصلي.

انقر على الأمر Ok للتنفيذ.

تم الحصول على النتائج الآتية:

- الجدول الآتي يحتوي على بيانات إحصائية عن المتغيرين، وتتضمن الوسط وعدد الحالات والانحراف المعياري والانحراف المعياري للوسط.

جدول (6-8)

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	BeforeTraining	7.11	9	2.028	.676
	AfterTraining	17.11	9	1.965	.655

الجدول الآتي يحسب معامل الارتباط بين البيانات قبل وبعد، ويعطي اختبارًا لمعامل الارتباط، حيث يكون الفرض العدمي الارتباط يساوي صفرًا، والفرض البديل الارتباط لا يساوي صفرًا. معامل الارتباط موجب وهو 0.499 أما قيمة Sig.=0.172 وهي أكبر من 0.05 فنقبل فرض العدم القائل بأن الارتباط غير معنوي.

جدول (6-9)

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	BeforeTraining & AfterTraining	9	.499	.172

جدول (6-10)

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	BeforeTraining AfterTraining	-10.000	2.000	.667	-11.537	-8.463	-15.000	8	.000

الجـدول السابـق يعطي اختبار t للفرق بـين عينتين غير مستقلتين، نلاحـظ أن Sig.=.000 آخر عمود وهي أقل من 0.025 إذن نرفض الفرض العدمي القائل إنه لا يوجد فرق بين العينتين، ونقبل الفرض البديل القائل إن هناك فرق درجات حدث نتيجة التدريب.

الفصل السابع

تحليل التباين

Analyzes of Variance ANOVA

(7-1) مقدمة :

في دراستنا لاختبارات الفروض في الفصل السابق، كانت الدارسة قاصرة على مجتمع واحد أو مجتمعين على الأكثر. وشملت تلك الدارسة اختبار فروض عن متوسط مجتمع واحد واختبارات فروض بخصوص تساوي متوسطي مجتمعين. ومن الطبيعي تعميم هذا الأسلوب إلى حالة وجود m من المجتمعات ($m \geq 2$)، ويكون المطلوب منا مقارنة متوسطاتها. فمثلا إذا كان المطلوب إجراء مقارنة بين متوسطات إنتاجية الفدان لخمسة أصناف مختلفة من القطن، رمزنا لمتوسطات إنتاج الفدان من هذه الأصناف الخمسة بالرموز: $\mu_1 = \mu_2 = ... = \mu_5$ يكون فرض عدم وجود فرق بين هذه المتوسطات (أي فرض تساوي متوسطات هذا الأصناف الخمسة) هو:

$$H_0: \mu_1 = \mu_2 = ... = \mu_k \qquad (1-7)$$

ويكون الفرض البديل H_A هو أن فرض العدم ليس صحيحا، أي أن H_A هو فرض عدم تساوي متوسطين علي الأقل من هذة المتوسطات، ويكون:

H_A: يوجد اثنان على الأقل من المتوسطات $\mu_1, \mu_2, ..., \mu_5$ غير متساويين (2-7)

ويستخدم أسلوب تحليل التباين في اختبار فرض تساوي متوسطات عدة مجتمعات ضد فرض عدم تساويها، فمثلاً يكون فرض العدم والفرض البديل كما في (7-1)، (7-2)، ويتضح مدى أهمية أسلوب تحليل التباين في اختبار فرض عدم مثل H_0 في (7-1) ضد فرض بديل مثل H_A بملاحظة أنه لإجراء مثل هذا الاختبار باتباع الأسلوب السابق عرضه (باستخدام اختبار t) لمقارنة متوسطي كل مجموعتين على حده، فإن هذا سوف يستلزم إجراء عشرة اختبارات من ذلك النوع، ولاشك في أن هذا سوف يحتاج إلى كثير من الوقت والجهد. ومن الواضح أنه كلما زاد عدد المجتمعات المطلوب مقارنة متوسطاتها زادت صعوبة إجراء المقارنات المطلوبة باتباع أسلوب مقارنة كل متوسطين على حده. فمثلاً إذا كانت مقارنة متوسطات عشرة مجتمعات مطلوبة فإن ذلك يستلزم إجراء 45 اختبارًا، كل منها لمقارنة متوسطين فقط باستخدام اختبار t. ويستخدم أسلوب تحليل التباين لمقارنة متوسطات عدة مجتمعات في العديد من البحوث مثل مقارنة متوسطات إنتاج

عـدة أصناف من محصول معين، مقارنة متوسطات عدة أيام لازمة للشفاء من مرض معين لعدة أنواع من الأدوية، مقارنة عدة طرق من طرق التدريس، مقارنة عدة طرق لإنتاج سلعة معينة، مقارنة متوسط الدخل الشهري للأسرة بعدة محافظات وهكذا.

وفي تحليـل التباين يستخدم تعبير معالجـات *Treatments* ليعبر عن التصنيفات المختلفة المستخدمة والمطلـوب مقارنـة متوسطاتهـا، فقد تكـون المعالجات أنواعًـا مختلفة من الأدويـة أو قد تكـون أنواعًا من الأسمـدة أو طرقًا مختلفة للتدريـب أو محافظات مختلفة، ومطلوب إجراء مقارنات بين متوسطات ظاهرة ما باستخدام عينة عشوائية من كل منها.

وباختصار يقصد بتحليل التباين العمليات الرياضية الخاصة بتقسيم مجموع المربعات الكلي لمجموعة من البيانات إلى مصادره المختلفة، وتلخص نتائج التحليل في جدول يعرف بجدول تحليل التباين *ANOVA Table*. إن الهـدف مـن إجراء هذا التحليـل هو اختبار فرضيـة تساوي متوسطات مجموعـة من العينات وتعرف بالمعالجات أو المعاملات *Treatment* دفعة واحدة، ولهذا فهو يعد توسيعاً لاختبار *t*.

☜ **اختبار إف:** Test-F

اختبار تحليل التباين يعتمد على إحصائي اختبار يطلق عليه اسم توزيع F نسبة إلى التوزيع الاحتمالي F Distribution.

ولاختبار مساواة متوسطات المجموعات يتم تقسيم التباين الكلي للمتغير التابع إلى مركبتين الأولى معروفة المصـدر وتسمى بين المجموعات (Between Group) ومصدرها الفرو ق بين متوسطات المجموعات، فإذا كان هـذا الجزء كبيرًا فإن متوسطات المجموعـات غير متساوية، والثانية داخل المجموعات (Within Group) وهي الجزء غير معروف المصدر الذي يسمى في بعض الأحيان الباقي Residuals أو الخطأ Error.

✓ متى نرفض الفرضية التي تقول: إن متوسطات المجموعات متساوية؟

نرفض هـذه الفرضية إذا كانت نسبـة التباين بين المجموعات (معروف المصـدر) إلى التباين داخل المجموعـات (غـير معروف المصدر) كبيرة، وهـذه النسبة تسمى قيمة F، فإذا كانـت قيمة F كبيرة نسبيًا فـإن متوسطـات المتغير التابـع للمجموعات غير متساوية، ولكن إلى أي حد تعـد قيمة F كبيرة حتى نرفض الفرضية التي تقول أن متوسطات المجموعات متساوية؟

✓ متى نقول إن قيمة F كبيرة نسبيًا؟

إذا كانـت المساحـة فوقها (مستـوى الدلالة Sig) أقل من المستوى المقبول لدينـا α التي غالبًا تساوي 0,05، فإذا كانت قيمة Sig أقل من α فإن متوسطات المجموعات غير متساوية، وإذا كانت قيمة Sig. أكبر من $\alpha = 0.05$ فإن متوسطات المجموعات متساوية.

(7-2) تحليل التباين في اتجاه واحد (One-Way ANOVA):

يقصد بتحليل التباين العمليات الرياضية الخاصة بتقسيم مجموع المربعات الكلي لمجموعة من البيانات إلى مصادره المختلفة وتلخيص نتائج التحليل في جدول يعرف بجدول تحليل التباين ANOVA Table. إن الهدف من إجراء هذا التحليل هو اختبار فرضية تساوي متوسطات مجموعة من العينات، وتعرف بالمعالجات أو المعاملات Treatment دفعة واحدة.

متى يستخدم تحليل التباين في اتجاه واحد؟

يستخدم إذا كان لدينا عدد من العينات ونريد أن نختبر هل هذه العينات مسحوبة من مجتمعات متوسطاتها متساوية أم لا؟ بمعنى آخر هل هذه العينات مسحوبة من المجتمع نفسه أم لا؟ يمكن اختبار ذلك الفرض باستخدام اختبار تحليل التباين. أي أن

الفروض الإحصائية:

الفرض العدم

$$H_0: \mu_1 = \mu_2 = ... = \mu_k$$

ضد الفرض البديل

يوجد متوسطان على الأقل غير متساويين: H_A

يمكن تلخيص خطوات الاختبار كالآتي:

- يحسب مجموع مشاهدات كل عينة وكذلك متوسطها، نفرض أن $\sum_{i=1}^{N} X_{ir}$ ترمز لمجموع مشاهدات العينة r، $\overline{X}_r = (\sum_{i=1}^{N} X_{ir}) / N$ ترمز إلى وسطها الحسابي، حيث إن $r = 1,...,k$ وعدد البيانات هي N فيمكن تنظيم حساب مجموع مشاهدات كل عينة ومتوسطها باستخدام جدول البيانات الآتي:

جدول (7-1)

المشاهدات	المجتمع				
	1	2	3	...	k
1	X_{11}	X_{12}	X_{13}	...	X_{1k}
2	X_{21}	X_{22}	X_{23}	...	X_{2k}
3	X_{31}	X_{32}	X_{33}	...	X_{3k}
: :	: :	: :	: :	... ...	: :
N_j	$X_{N_1 1}$	$X_{N_2 2}$	$X_{N_3 3}$	...	$X_{N_k k}$
$\sum$	$\sum_{i=1}^{N_1} X_{i1}$	$\sum_{i=1}^{N_2} X_{i2}$	$\sum_{i=1}^{N_3} X_{i3}$	...	$\sum_{i=1}^{N_k} X_{ik}$
$\overline{X}$	$\overline{X}_1$	$\overline{X}_2$	$\overline{X}_3$	...	$\overline{X}_k$

يحسب المجموع الكلي لمشاهدات جميع العينات ونرمز له بالرمز $\sum$ والمتوسط العام $\overline{X}$ كالآتي:

$$\sum = \sum_{i=1}^{N_1} X_{i1} + \sum_{i=1}^{N_2} X_{i2} + \sum_{i=1}^{N_3} X_{i3} + \dots + \sum_{i=1}^{N_k} X_{ik}$$

وأيضا $\overline{X} = \frac{\sum}{N}$ حيث إن $N = N_1 + N_2 + N_3 + \dots + N_k$.

يحسب مجموع المربعات الكلي Total sum of squares كالآتي، وهو الذي نرمز له بالرمز SST:

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{N_i} X^2_{ji} - \frac{\sum^2}{N}$$

$$SST = X_{11}^2 + X_{21}^2 + \dots + X_{N_1 1}^2 + X_{12}^2 + X_{22}^2 + \dots + X_{N_2 2}^2 + \dots + X_{N_k k}^2 - \frac{\sum^2}{N}$$

(3-7)

أي أن مجموع المربعات الكلي نحصل عليه بإيجاد مجموع مربعات مشاهدات العينات مطروحًا منه خارج قسمة مربع المجموع الكلي لمشاهدات جميع العينات على N.

ويمكن حساب مجموع المربعات الكلي بطريقة أخرى وهي كالآتي:

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{N_i} \left(X_{ji} - \overline{X}\right)^2$$

أي أن مجموع المربعات الكلي هو مجموع مربعات انحرافات كل مشاهدة من مشاهدات جميع العينات المأخوذة عن المتوسط العام لجميع هذه المشاهدات $\overline{X}$.

- يحسب مجموع مربعات بين المعالجات Sum of squares Between treatments

مجموع مربعات بين المعالجات وهو الذي نرمز له بالرمز SSR:

$$SSR = \frac{\left(\sum_{i=1}^{N_1} X_{i1}\right)^2}{N_1} + \frac{\left(\sum_{i=1}^{N_2} X_{i2}\right)^2}{N_2} + \ldots + \frac{\left(\sum_{i=1}^{N_k} X_{ik}\right)^2}{N_k} - \frac{(\sum)^2}{N}$$

(4-7)

ويمكن حساب مجموع مربعات بين المعالجات بطريقة أخرى كالآتي:

$$SSR = \sum_{i=1}^{K} N_i \left(\overline{X}_i - \overline{X}\right)^2$$

$$SSR = N_1\left(\overline{X}_1 - \overline{X}\right)^2 + N_2\left(\overline{X}_2 - \overline{X}\right)^2 + \ldots + N_k\left(\overline{X}_k - \overline{X}\right)^2$$

- يحسب مجموع مربعات الأخطاء (مربعات داخل المعالجات)

Sum of Square within treatments or sum of square of Error

يحسب بالفرق بين مجموع المربعات الكلي ومجموع المربعات بين المعالجات الذي تم حسابها في الخطوتين السابقتين، والذي نرمز له بالرمز SSE. ويمكن حسابه بطريقة أخرى وهي كما يلي:

$$SSE = \sum_{i=1}^{k} \sum_{j=1}^{N_i} \left(X_{ji} - \overline{X}_i\right)^2$$

(5-7)

أي أن مجموع مربعات داخل المعالجات يعبر عن التشتت داخل العينات أي عن تباعد مشاهدات كل عينة عن وسطها الحسابي.

- يحسب متوسط مربعات بين المعالجات $\sigma^2_{Between}$ وذلك بقسمة مجموع مربعات بين المعالجات على (عدد المعالجات-1) أي على k-1، ويسمى k-1 بدرجات حرية بين المعالجات أي أن:

$$MSR = (k-1)/SSR = \sigma^2_{Between} \qquad (6-7)$$

- يحسب متوسط مربعات داخل المعالجات، وذلك بقسمة مجموع مربعات داخل المعالجات على (عدد المشاهدات الكلية – عدد المعالجات) أي على N-K، ويسمى N-K بدرجات حرية داخل المعالجات أي أن:

$$MSE = (N-K)/SSE = \sigma^2_{Between} \qquad (7-7)$$

- يحسب خارج قسمة متوسط مربعات بين المعالجات على متوسط مربعات داخل المعالجات، وإذا رمزنا بالرمز F للنسبة الناتجة فإن:

$$F = \frac{\sigma^2_{Betwen}}{\sigma^2_{within}} \qquad (8-7)$$

ويكون لها توزيع F بدرجات حرية k–1، N-k. وتعد F المحسوبة هي إحصاءة الاختبار.

- إذا كانت F المحسوبة أكبر من $F_{(\alpha,k-1,n-k)}$ يرفض الفرض العدم H_0 و فيما عدا ذلك يقبل فرض العدم H_0.

والجدول الآتي الذي يعرض ملخص الخطوات السابقة يسمى جدول تحليل التباين analysis of variance table، وتكون له الصيغة العامة المبينة في الجدول الآتي:

جدول (2-8)

مصدر الاختلاف	مجموع المربعات SS	درجات الحرية d.f	متوسط المربعات MS	F
بين المعاملات	SSR	K-1	MSR	$F\ calc = \frac{MSR}{MSE}$
الخطأ	SSE	$\sum_{i=1}^{K} N_i - k$	MSE	
المجموع	SST	$\sum_{i=1}^{K} N_i - 1$		

والآن نتناول كيف يمكننا تنفيذ هذا باستخدام البرنامج ما يعمل على إتاحة استخدام ذلك مع توفير الوقت والجهد في حالة زيادة عدد بيانات التحليل الإحصائي؛ وذلك من خلال أمثلة تطبيقية متنوعة:

مثال(7–1):

استخدمت ثلاث طرق تعليمية مختلفة في تدريس مادة الحساب لثلاث عينات متشابهة من الطلاب، وكانت درجاتهم في الامتحان النهائي كما في الجدول الآتي:

جدول (7-3)

العينة الأولى	8	7,5	7	5,5
العينة الثانية	9	10	8	9
العينة الثالثة	6	5	5,5	3,5

اختبر ما إذا كانت هناك فروق معنوية بين متوسطات درجات الطلبة في الطرق الثلاث عند مستوى معنوية 0,05.

الحل:

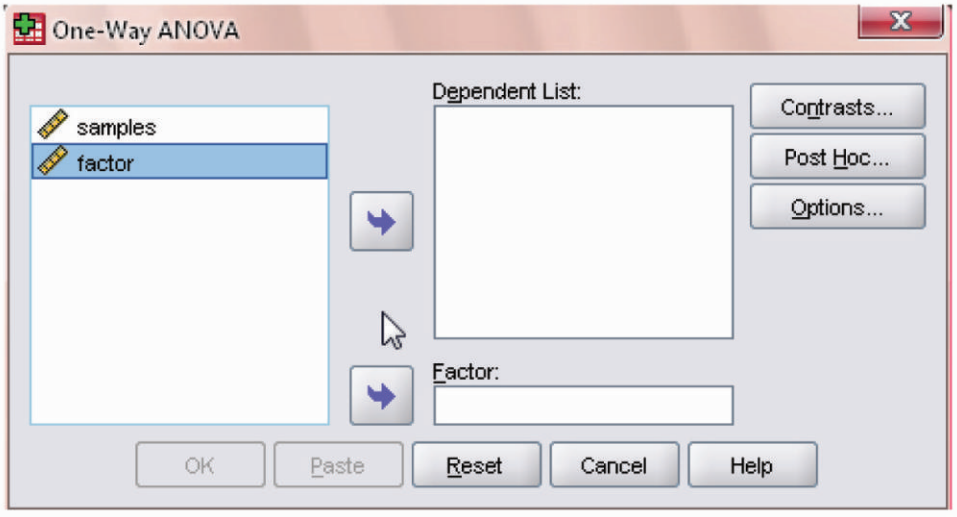
في البداية يتم إدخال البيانات كلها في البرنامج في عمود واحد مع إضافة عمود ثان يشير إلى رقم العينة فيكون شاشة data editor وقد سميناه factor نضع فيه قيمة 1 لرمز أن القيمة الموجودة في العمود samples هي قيمة العينة الأولى، ونضع القيمة 2 لرمز أن القيمة الموجودة في العمود samples هي قيمة العينة الثانية وهكذا، ونضع القيمة 3 لرمز أن القيمة الموجودة في العمود samples هي قيمة العينة الثالثة وتكون البيانات كما في الشكل الآتي:

samples	factor
8.00	1
7.50	1
7.00	1
5.50	1
9.00	2
10.00	2
8.00	2
9.00	2
6.00	3
5.00	3
5.50	3
3.50	3

وبعد إدخال البيانات نتبع الخطوات الآتية:

- اختيار Compare means.
- اختيار One-Way ANOVA.

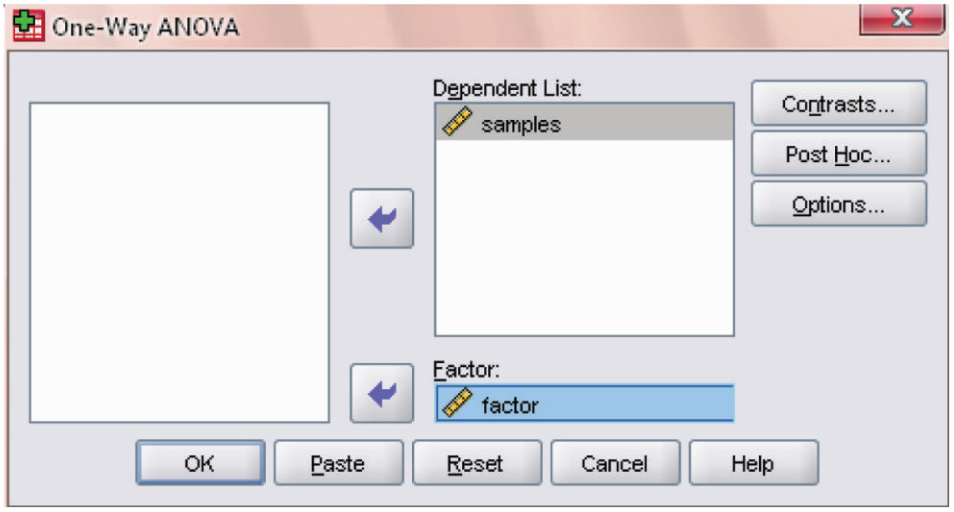
فيظهر الشكل الآتي:



شكل (7-1)

ننقل المتغيرات المراد حساب الفروق فيها إلى Dependent List، ومتغير تصنيف المشاهدات في Factor.

فيظهر الشكل الآتي:



شكل (7-2)

ننقر على OK تظهر النتائج كما في الجدول الآتي:

جدول (7-4)

ANOVA

samples

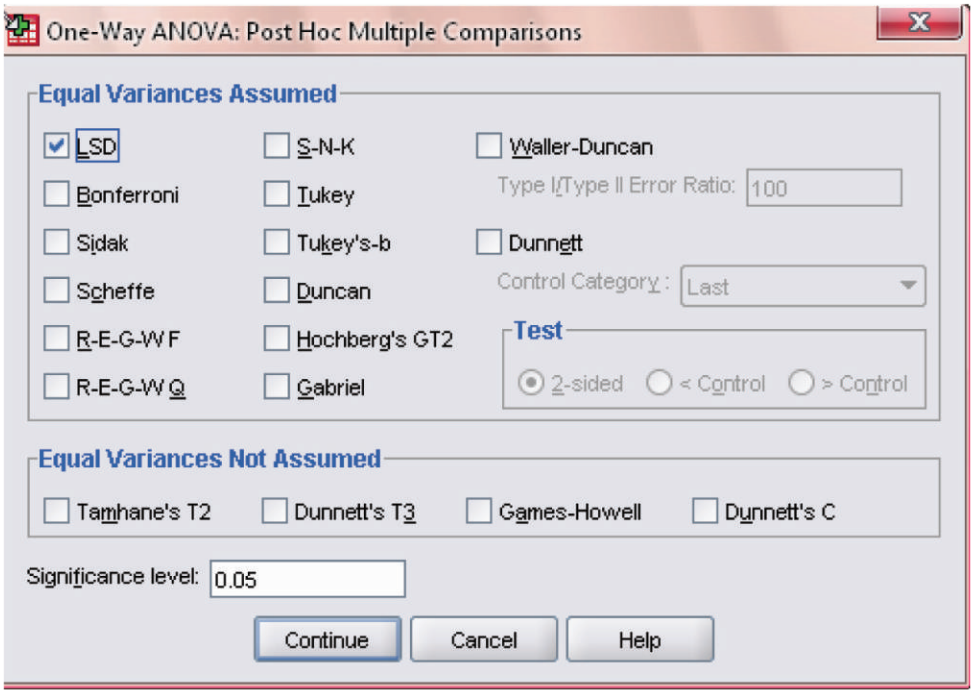
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	32.000	2	16.000	16.000	.001
Within Groups	9.000	9	1.000		
Total	41.000	11			

القرار الإحصائي:

يتضح من الجدول أن P-value = 0.001 وهي أقل من 0.05 عند F = 16.00 وهذا يدل على وجود فروق بين الطرق الثلاث لتدريس المهارات الحسابية.

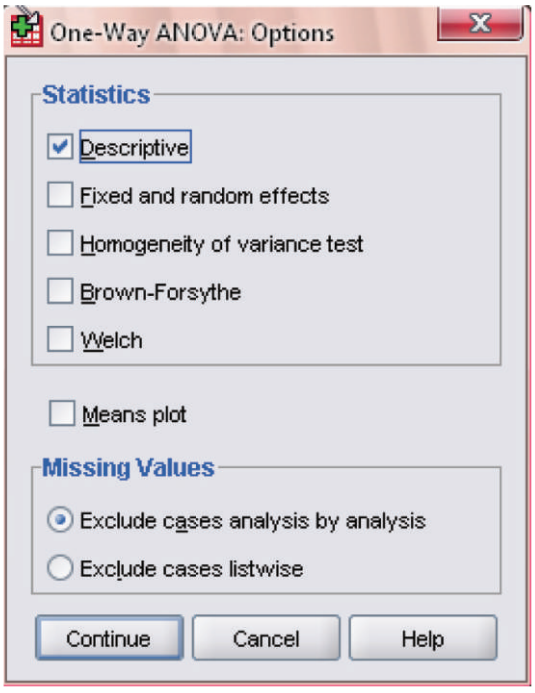
نظراً لوجود فروق معنوية بين متوسطات الطرق فهذا يعني عدم تساوي متوسطي طريقتين على الأقل، ولاختبار معنوية الفروق لكل زوج من المعالجات نلجأ إلى المقارنات المتعددة باستخدام طريقتي L.S.D وباتباع الخطوات الآتية:

باستخدام الخطوات السابقة حتى الشكل رقم (2-7) فنقوم بالنقر على post Hoc المحددة في الشكل (2-7) فيظهر الشكل (3-7):



شكل (3-7)

و نلاحظ وجود طرق عديدة للمقارنات المتعددة عندما يكون التباين لكل عينة متساويًا، حيث قمنا بتأشير LSD، وتوجد طرق للمقارنة عندما يكون التباين غير متساو، وأخيراً تحديد مستوى المعنوية المطلوب 0.05 في Significance Level، عند نقر زر Continue، كما نلاحظ أنه يوجد زر في الشاشة الرئيسة شاشة (2-7) يطلق عليها Options وعند تفعيل هذا الاختيار نتمكن من حساب الإحصائيات الوصفية للقيم الفعلية أو القيم المفقودة كما في الشكل الآتي:



شكل (7-4)

ثم زر Ok نحصل على المخرجات الآتية:

جدول (7-5)

Multiple Comparisons

SAMPLES

LSD

(I) FACTOI	(J) FACTOI	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interv	
					Lower Bound	Upper Bound
1	2	-2.0000*	.707	.020	-3.5996	-.4004
	3	2.0000*	.707	.020	.4004	3.5996
2	1	2.0000*	.707	.020	.4004	3.5996
	3	4.0000*	.707	.000	2.4004	5.5996
3	1	-2.0000*	.707	.020	-3.5996	-.4004
	2	-4.0000*	.707	.000	-5.5996	-2.4004

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

Post Hoc Tests

نلاحظ أنه باستعمال طريقة LSD فقد ظهرت فروق معنوية بمستوى معنوية 0.05 بين متوسطات المعالجات (1و2) – (1و3) و (2و3)، حيث كانت قيمة p-value أو sig. أقل من 0.05 مع وجود علامة (*) على فروق المتوسطات ووجود تلك العلامة علامة وجود فروق بين العينات.

مثال (7-2):

لمقارنة أربعة أنواع من الأدوية a، b، c، d لعلاج مرض معين اختيرت أربع عينات عشوائية كل منها مكون من خمسة أشخاص من المصابين بذلك المرض، وأعطيت كل عينة نوعًا من الأدوية، وكان عدد أيام العلاج اللازمة حتى الشفاء للأشخاص في العينات الأربع كما في الجدول الآتي:

جدول (7-6)

العينة الأولى (دواء a):	7	3	4	6	5
العينة الثانية (دواء b):	6	5	9	8	7
العينة الثالثة (دواء c):	3	2	5	6	4
العينة الرابعة (دواء d):	8	6	3	7	6

اختبر فرض عدم وجود فرق معنوي بين متوسطات عدد الأيام اللازمة حتى الشفاء للأدوية الأربعة a، b، c، d وذلك عند مستوى معنوية 0,01 ,0.

الحل:

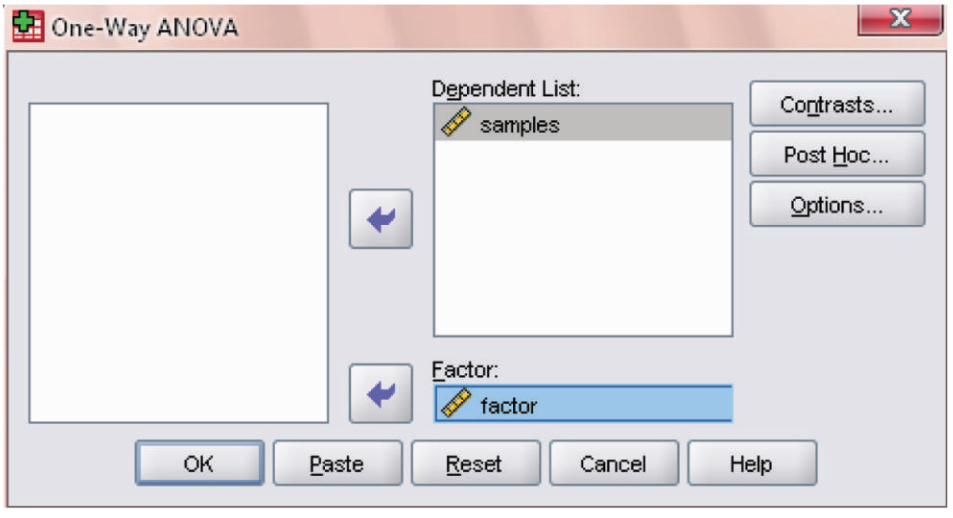
في البداية يتم إدخال البيانات كلها في البرنامج في عمود واحد مع إضافة عمود ثان يشير إلى رقم العينة فيكون شاشة data editor وقد سميناه factor نضع فيه قيمة 1 لرمز أن القيمة الموجودة في العمود samples هي قيمة العينة قيمة الدواء الأولى a، ونضع القيمة 2 لرمز أن القيمة الموجودة في العمود samples هي قيمة الدواء الثاني b وهكذا، ونضع القيمة 3 لرمز أن القيمة الموجودة في العمود samples هي قيمة الدواء الثالث c وهكذا، ونضع القيمة 4 لرمز أن القيمة الموجودة في العمود samples هي قيمة الدواء الرابع d وتكون البيانات كما في الشكل الآتي:

samples	factor
7.00	1
3.00	1
4.00	1
6.00	1
5.00	1
6.00	2
5.00	2
9.00	2
8.00	2
7.00	2
3.00	3
2.00	3
5.00	3
6.00	3
4.00	3
8.00	4
6.00	4
3.00	4
7.00	4
6.00	4

وبعد إدخال البيانات نتبع الخطوات الآتية:

- اختيار Compare means.
- اختيار One-Way ANOVA.
- ننقل المتغيرات المراد حساب الفروق فيها إلى Dependent List، ومتغير تصنيف المشاهدات في Factor.

فيظهر الشكل الآتي:



شكل (5-7)

ننقر على Ok تظهر النتائج كما في الجدول الآتي:

جدول (7-7)

ANOVA
Sample

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	25.000	3	8.333	3.030	.060
Within Groups	44.000	16	2.750		
Total	69.000	19			

القرار الإحصائي:

يتضح من الجدول أن p-value = 0.06 وهي أكبر من 0.01 عند F = 3.03، وهذا يدل على عدم وجود فروق بين الأنواع الأربعة من الأدوية، ويؤكد ذلك جدول تحليل الفروق الآتية:

Post Hoc Tests

جدول (7-8)

Multiple Comparisons
Sample
LSD

(I) Factor	(J) Factor	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1.00	2.00	-2.00000-	1.04881	.075	-4.2234-	.2234
	3.00	1.00000	1.04881	.355	-1.2234-	3.2234
	4.00	-1.00000-	1.04881	.355	-3.2234-	1.2234
2.00	1.00	2.00000	1.04881	.075	-.2234-	4.2234
	3.00	3.00000*	1.04881	.011	.7766	5.2234
	4.00	1.00000	1.04881	.355	-1.2234-	3.2234
3.00	1.00	-1.00000-	1.04881	.355	-3.2234-	1.2234
	2.00	-3.00000-*	1.04881	.011	-5.2234-	-.7766-
	4.00	-2.00000-	1.04881	.075	-4.2234-	.2234
4.00	1.00	1.00000	1.04881	.355	-1.2234-	3.2234
	2.00	-1.00000-	1.04881	.355	-3.2234-	1.2234
	3.00	2.00000	1.04881	.075	-.2234-	4.2234

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

نلاحظ أنه باستعمال طريقة LSD فقد لم تظهر فروق معنوية بمستوى معنوية (دلالة) 0.01 بين متوسطات المعالجات كافة، حيث كانت قيمة P-Value أو Sig. أكبر من 0.05 مع عدم وجود علامة (*) عند فروق المتوسطات.

(3-7) تحليل التباين في اتجاهين (Two Way ANOVA):

إن تحليل التباين في اتجاه واحد يستخدم لدراسة أثر عامل واحد فقط على متغير ما، ولكن عندما نريد دراسة أثر عاملين أو أكثر في هذه الحالة يمكننا استخدام تحليل التباين في اتجاهين (أو ما يسمى تحليل التباين الثنائي).

فتحليل التباين الثنائي Two Way ANOVA يمكن استخدامه لدراسة أثر متغيرين عاملين نسمي أحدهما الصفوف (المعالجات)، والآخر الأعمدة (القطاعات)، ويمكن وصف البيانات تحت هذه التجارب كالآتي:

جدول (7-9)

الصفوف (Rows)		الأعمدة (Columns) القطاعات					
		2	1	...	j	...	c
معالجات	1	X_{1c}	...	X_{1j}	...	X_{12}	X_{11}
	2	X_{2c}	...	X_{2j}	...	X_{22}	X_{21}
	.	.	...	.	...	.	.
	.	.	...	.	...	.	.
	i	X_{ic}	...	X_{ij}	...	X_{i2}	X_{i1}
	.	.	...	.	...	.	.
	.	.	...	.	...	.	.
	.	.	...	.	...	.	.
	r	.	...	.	...	.	.
		X_{rc}	....	X_{rj}	...	X_{r2}	X_{r1}

يمكن صياغة الفروض الإحصائية في هذه الحالة كالآتي:

(أ) للمعالجات:

$$H_0^{(1)}: \mu_1 = \mu_2 = \ldots = \mu_r$$

ضد الفرض البديل

يوجد متوسطان على الأقل غير متساويين: $H_A^{(1)}$

(ب) للقطاعات:

$$H_0^{(2)}: \mu_1 = \mu_2 = \ldots = \mu_c$$

ضد الفرض البديل

يوجد متوسطان على الأقل غير متساويين: $H_A^{(2)}$

مثال (7-3)

إذا كانت لدينا أربعة أنواع من الأسمدة A،B،C،D استخدمت كمعالجات لأربعة قطاعات مختلفة، حيث كانت النتائج كما في الجدول الآتي:

القطاعات/ المعالجات	القطاع الأول	القطاع الثاني	القطاع الثالث	القطاع الرابع	المجموع
A	9.3	9.4	9.6	10	38.3
B	9.4	9.3	9.8	9.9	38.4
C	9.2	9.4	9.5	9.7	37.8
D	9.7	9.6	10	10.2	39.5
المجموع	**37.6**	**37.7**	**38.9**	**39.8**	**154**

استخدم تحليل التباين في اتجاهين لمعرفة ما إذا كان لها تأثير مختلف في زيادة إنتاج القمح أم لا؟

الحـــل:

أولا: الفروض الإحصائية بالنسبة للمعالجات:

H_0: لا توجد فروق بين متوسطات المعالجات (الصفوف).

H_1: توجد فروق بين متوسطات المعالجات على الأقل لاثنين منهم.

ثانيا: الفروض الإحصائية بالنسبة للقطاعات:

H_0: لا توجد فروق بين متوسطات القطاعات (الأعمدة).

H_1: توجد فروق بين متوسطات القطاعات على الأقل لاثنين منهم.

ندخل البيانات كما في الشكل الآتي:

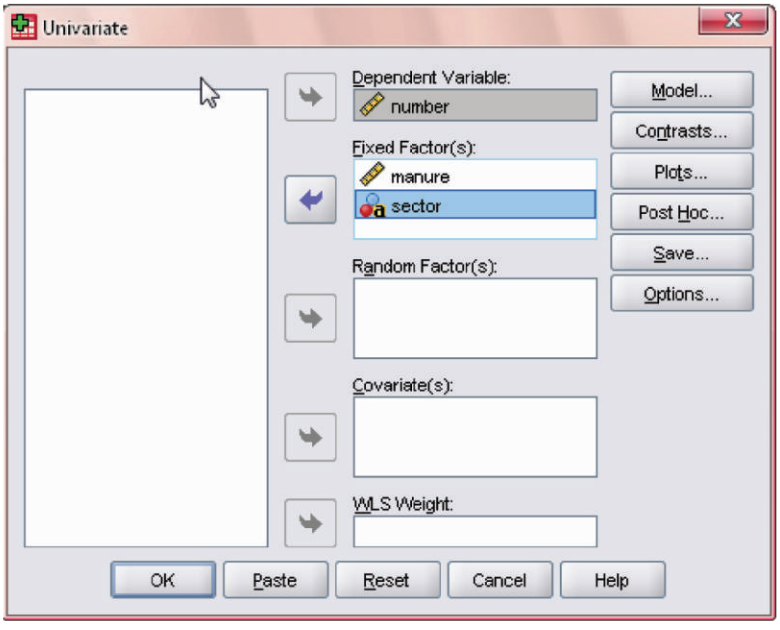
manure	sector	number
1.00	A	9.30
1.00	B	9.40
1.00	C	9.20
1.00	D	9.70
2.00	A	9.40
2.00	B	9.30
2.00	C	9.40
2.00	D	9.60
3.00	A	9.60
3.00	B	9.80
3.00	C	9.50
3.00	D	10.00
4.00	A	10.00
4.00	B	9.90
4.00	C	9.70
4.00	D	10.20

شكل (7-7)

ثم نتبع:

Analyze → General linear Model → Univariate

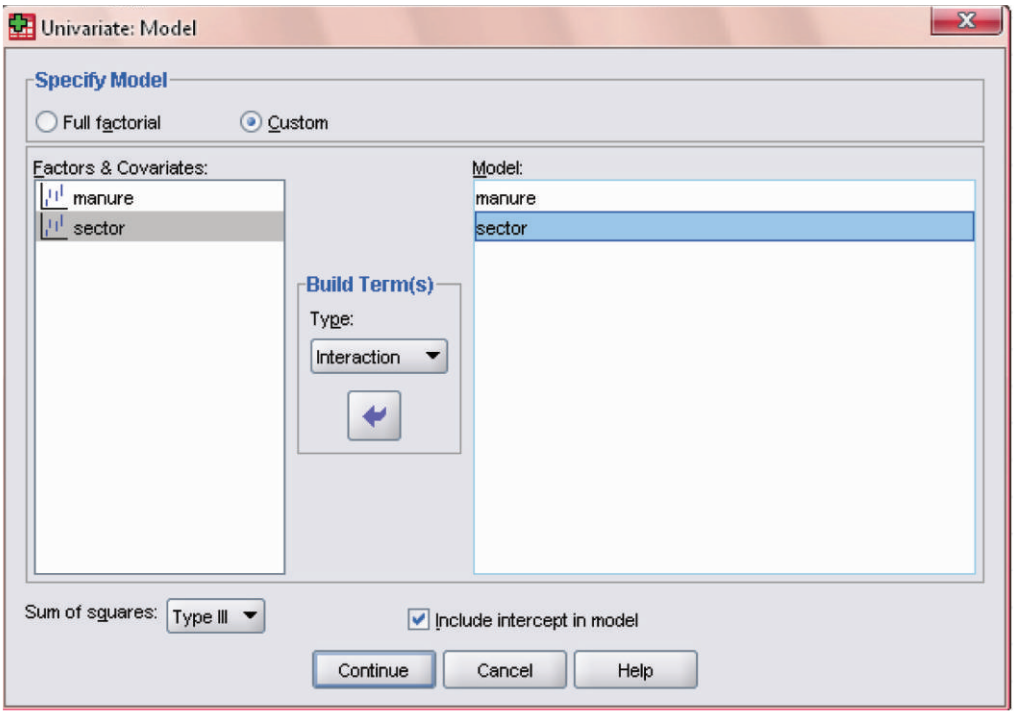
ليظهر الشكل الآتي:



شكل (7-8)

- أدخلنا العاملين في Fixed Factors على اعتبار أن جميع معالجات العامل قد ضمنت في التجربة، أما في حالة أخذ عينة من معالجات العامل فإننا سنستعمل خانة Random Factor.

- انقر زر Model فيظهر صندوق حوار Model، اختر Custom بدلا Full Factorial، وذلك لأننا لا نرغب في ظهور التفاعل Interaction في جدول تحليل التباين (Manure*Sector) لعدم وجود درجات حرية كافية للخطأ التجريبي، حيث يظهر صندوق حوار Model بعد ترتيبه كما في الشكل (7-9):



شكل (7-9)

حيث إن تأشير Include Intercept In Model يعمل على تضمين الحد الثابت في النموذج الخطي العام بوصفه نموذج انحدار.

أما خانة Build Terms فتستعمل لتعيين نوع التأثيرات Effects التي يراد إظهارها في جدول تحليل التباين، حيث قمنا بنقل المتغيرين manure و sector من خانة Factors & Covariates إلى خانة Model، لتظهر كتأثيرات رئيسة Main Effect (بعد التأكد من أن خانة Build Effects تتضمن الخيار Main effect ثم انقر على Continue ثم Ok في صندوق حوار Univariate.

ثم تحصل على المخرجات الآتية كما في الشكل الآتي:

جدول (7-10)

Between-Subjects Factors

		N
manure	1	4
	2	4
	3	4
	4	4
sector	A	4
	B	4
	C	4
	D	4

Tests of Between-Subjects Effects

Dependent Variable:number

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.385[a]	3	.128	1.702	.220
Intercept	1482.250	1	1482.250	19654.144	.000
manure	.000	0	.	.	.
sector	.000	0	.	.	.
Error	.905	12	.075		
Total	1483.540	16			
Corrected Total	1.290	15			

a. R Squared = .298 (Adjusted R Squared = .123)

الجـدول الأول يعطي بيانًا بعدد الحالات والمفقود، الجدول الثاني يبين مجموع المربعات بين المعالجات = 0.385 ومجموع المربعات بين القطاعات = 0.825 مجموع مربعات الخطأ = 0.08 ومجموع المربعات الكلي = 1.29.

⊙ **القرار الإحصائي للمعالجات:**

وأيضا تتضح معنوية الفرق بين الأسمدة وكذلك بين القطاعات بمستوى معنوية 0.05،

حيث إن p-value=0.0 عند صف الأسمدة manure وهي أقل من 0.05 أي نرفض H_o ونقبل H_A، ونقول إنه توجد فروق بين متوسطات المعالجات.

⊙ **القرار الإحصائي للقطاعات:**

وأيضا p-value=0.001 عند صف القطاعات sector وهي أقل من 0.05 نرفض H_o ونقبل H_A، ونقول إنه توجد فروق بين متوسطات القطاعات.

الفصل الثامن

تحليل الارتباط

Correlation Analysis

(8-1) مقدمة :

تعرضنا في الفصول السابقة لدراسة ظاهرة واحدة كالاستهلاك، الدخل، المبيعات، العمر، والوزن وبدأنا ببيان كيفية عرض البيانات التي جمعناها في جدول أو أشكال بيانية مختلفة، واستطردنا بعد ذلك إلى كيفية تلخيص تلك البيانات من خلال مقاييس النزعة والتشتت لمجموعة واحدة أو مجموعات مختلفة، وكيفية تقدير تلك المقادير، وكيفية إجراء اختبارات الفروض عليها، والآن نبحث كيفية إيجاد العلاقات الرياضية التي تربط المتغيرات ببعضها، وما مقدار هذا الارتباط من خلال ما يسمى مقاييس الارتباط المختلفة التي سوف نتعرض لها في هذا الفصل، كما نناقش في الفصل الآتي كيف يمكن التنبؤ بأحد المتغيرات لقيمة محددة للمتغير الآخر وهي ما تسمى معادلة الانحدار.

نستطيع القول إن تحليل الارتباط والانحدار هو أداة إحصائية نستفيد منها في تحديد العلاقة بين متغيرين أو أكثر للتنبؤ بأحد المتغيرات استنادا إلى قيم المتغير أو المتغيرات الأخرى. فمثلا إذا علمنا العلاقة بين مصروفات الدعاية والمبيعات، فيمكننا الاستفادة من تحليل الارتباط للتنبؤ بالمبيعات حالما تتوافر لنا قيمة نفقات الدعاية.

وسوف نتناول في هذا الفصل قضية الارتباط بين متغيرين سواء كانا كميين أو غير ذلك، بينما في الفصل القادم نتناول كيفية التنبؤ بأحد المتغيرات في ضوء متغير آخر أو أكثر من خلال دارسة الانحدار.

(8-2) الارتباط الخطي البسيط Simple Correlation

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

ومن الأمثلة على ذلك:

- الإنفاق، والدخل العائلي.
- سعر السلعة، والكمية المطلوبة منها.
- الفترة الزمنية لتخزين الخبز، وعمق طراوة الخبز.
- تقديرات الطلاب في مقرر الإحصاء، وتقديراتهم في مقرر الرياضيات.
- كميات السماد المستخدمة، وكمية الإنتاج من محصول معين تم تسميده بهذا النوع من السماد.
- عدد مرات ممارسة نوع معين من الرياضة البدنية، ومستوى الكوليسترول في الدم.
- وزن الجسم، وضغط الدم.

يحسب معامل الارتباط الخطي البسيط بافتراض وجود علاقة خطية بين اثنين من المتغيرات فقط مع العلم أن الحصول على قيمة صغيرة (قريبة من الصفر) لهذا المعامل لا يعني عدم وجود علاقة بين المتغيرين، فقد توجد علاقة من الدرجة الثانية (ارتباط غير خطي). ويختلف نوع المقاييس الذي نستخدمه في حساب معامل الارتباط طبقًا لنوع البيانات، وسوف يجري حسابه في حالة البيانات الكمية، والبيانات الوصفية المقاسة بمعيار ترتيبي.

الغرض من تحليل الارتباط الخطي البسيط

الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين، ويرمز له في حالة المجتمع بالرمز ρ (وتقرأ "رُو")، وفي حالة العينة بالرمز r، وحيث إننا في كثير من النواحي التطبيقية نتعامل مع بيانات عينة مسحوبة من المجتمع، سوف نهتم بحساب معامل الارتباط في العينة r كتقدير لمعامل الارتباط في المجتمع، ومن التحديد السابق للغرض من معامل الارتباط، نجد أنه يركز على نقطتين هما:

- نوع العلاقة:. وتأخذ ثلاثة أنواع حسب إشارة معامل الارتباط كما يلي:

✓ إذا كانت إشارة معامل الارتباط سالبة ($r < 0$) توجد علاقة عكسية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبها انخفاض في المتغير الثاني، والعكس.

✓ إذا كانت إشارة معامل الارتباط موجبة ($r > 0$) توجد علاقة طردية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين تصاحبها زيادة في المتغير الثاني، والعكس.

✓ إذا كانت قيمة معامل الارتباط صفرًا ($r = 0$) دل ذلك على انعدام العلاقة بين المتغيرين.

قوة العلاقة :-

يمكـن الحكـم علـى قوة العلاقة من حيث درجة قربها أو بعدها عـن [1±]، حيث إن قيمة معامل الارتباط تقع في المدى [$1 \le r \le 1-$].

معامل الارتباط الخطي البسيط "لبيرسون" Karl Pearson

في حالـة جمـع بيانات عن متغيرين كميين $(x \ , \ y)$، يمكن قياس الارتباط بينهما، باستخدام طريقة "بيرسون" Pearson، ومن الأمثلة على ذلك: قياس العلاقة بين الوزن والطول، والعلاقة بين الإنتاج والتكلفة، والعلاقة بين الإنفاق الاستهلاكي والدخل، والعلاقة بين الدرجة التي حصل عليها الطالب وعدد ساعات الاستذكار، وهكذا الأمثلة على ذلك كثيرة.

إذا افترضنا أن لدينا عينة مكونة من n من أزواج المشاهدات $(x_1 \ , \ y_1), (x_2 \ , \ y_2), \ldots (x_n \ , \ y_n)$ لمتغيرين X وY، فإن معامل الارتباط لبيرسون يعرف بأنه:

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sqrt{\left[\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right]\left[\sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}\right]}}$$

(1-8)

فـإذا نظرنـا إلى المعادلة السابقة عرفنا مقدار المعاناة عند حسـاب مقدار معامل الارتباط، فضلاً عـن حجـم التحليل والحسابات إذا زاد حجم العينة، والآن نتعرف كيف يمكن تطبيق ذلك بمنتهى السهولة باستخدام البرنامج وذلك من خلال المثال القادم:

مثال(8-1):

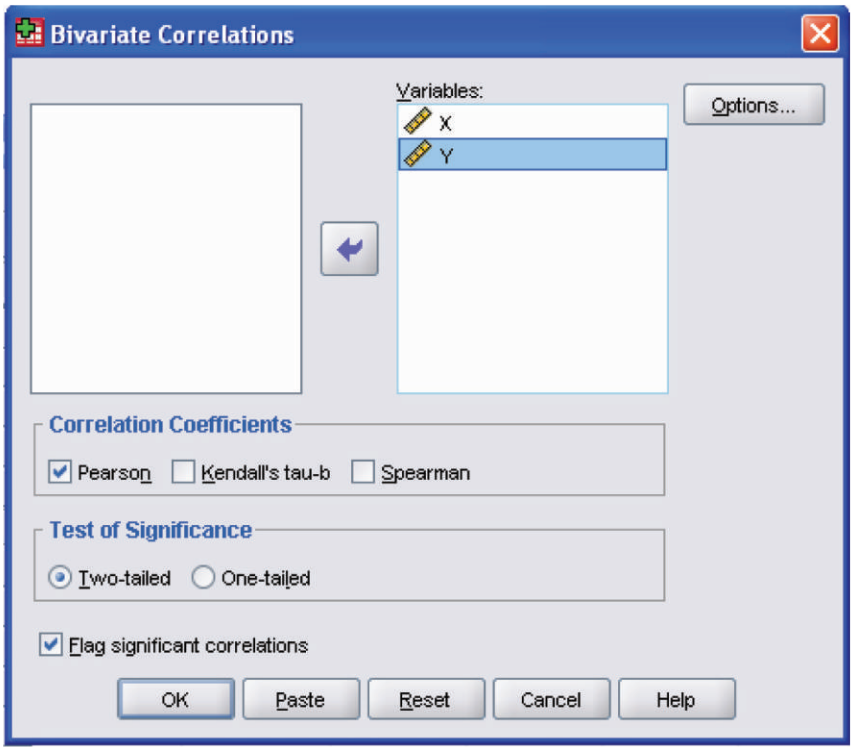
احسـب معامـل الارتباط المناسب بين درجـة الطالب في الرياضيات ودرجتـه في الإحصاء باستخدام البيانات الآتية لعينة من خمسة طلاب:

جدول (8-1)

(x) درجة الرياضة	15	19	11	3	7
(y) درجة الإحصاء	18	20	16	14	12

الحـل:

من شريط قوائم نختار Analyze ثم نختار Correlate ثم نختار Bivariate فيظهر الشكل الآتي:



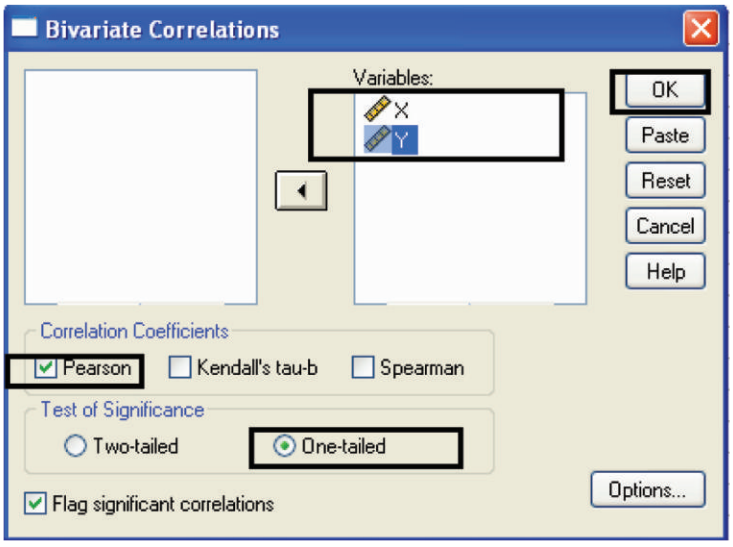
شكل (8-1)

ونرى أن الشكل السابق يتضمن مجموعة من الأجزاء:

- عمود المتغيرات المتاحة وهو العمود الأول من جهة اليسار، ومنه يتم نقل المتغيرات إلى العمود الذي بجواره لنبدأ حساب معامل الارتباط وهنا نقوم بنقل X، Y.
- معامـل الارتبـاط (*Correlation Coefficients*): وهو يتيح لنـا تحديد نوع معامل الارتباط الذي نريده ونلاحظ أنه متوافر لدينا 3 اختيارات:
  - ✓ Pearson: وهو معامل ارتباط الظواهر الكمية سواء كانت مبوبة أو غير مبوبة.
  - ✓ Kendall's tau: وهـو معامـل ارتبـاط الظواهر الترتيبية، ويفضل للعينـات الصغيرة والبيانات غير المبوبة.
  - ✓ Spearman: وهـو أيضًا معامل الارتباط بين الظواهر الكمية والترتيبية، ويفضل للعينات الصغيرة، ويستخدم في حالة البيانات غير المبوبة. وهنا نقوم بتحديد معامل بيرسون.
  - ✓ Test of Siginificance: ولأننا نستخدم بيانات عينة فإننا نحتاج إلى التأكد من قيمة معامل الارتباط فنحتاج إلى اختبار هذه القيمة، وهنا نختار إما الاختبار من جهة أو جهتين.

- *Flag Significant Correlation*: هنـا تستخـدم لتعليم الارتباط بعلامة نجمة (star) في حالة الارتباط المعنوي (الدال) أي الارتباط الحقيقي الذي يؤكده الاختبار.

وبعد الاختيار السابق يظهر الشكل (8-3) في الشكل الآتي:



شكل (8-2)

- ثم ننقر على ok فيظهر الجدول الآتي:

جدول (8-2)

Correlations

		X	Y
X	Pearson Correlation	1	.900*
	Sig. (2-tailed)		.037
	N	5	5
Y	Pearson Correlation	.900*	1
	Sig. (2-tailed)	.037	
	N	5	5

*. Correlation is significant at the 0.05 level (2-tailed).

- ونجـد أن معامـل الارتباط بين درجة الطالب في الإحصاء ودرجتـه في الرياضيات مقدارها 0.9، وهذا يعني أن الارتباط طردي وقوي جدا ولاختبار الفرض الصفري:

$$H_0: \rho = 0$$

ضد الفرض البديل:

$$H_A: \rho \neq 0$$

ونلاحظ أن الفرض البديل لا يساوي صفرًا، وهذا ما دل عليه اختيارنا السابق، حيث إن الاختبار ذو اتجاهين، فإذا افترضنا أننا نريد اختبار ذلك بمستوى معنوية مقدارها 0.05: α فإن قيمة: $0.019 > \alpha = 0.05 = \text{p-value}$ دل ذلك على أن الارتباط بين درجة الرياضيات والإحصاء ارتباط يختلف عن الصفر بمستوى معنوي مقداره 0.05، أي إننا نقبل الفرض البديل القائل: إن الارتباط لا يساوي صفرًا.

معامل ارتباط الرتب لـ(سبيرمان) Spearman

إذا كانت الظاهرة محل الدراسة تحتوي على متغيرين وصفيين ترتيبيين، ومثال على ذلك قياس العلاقة بين تقديرات الطلبة في مادتين، أو العلاقة بين درجة تفضيل المستهلك لسلعة معينة، ومستوى الدخل، فإنه يمكن استخدام طريقة أخرى في حساب معامل ارتباط يعتمد على رتب مستويات المتغيرين كبديل للقيم الأصلية، ويطلق على هذا المعامل "معامل ارتباط سبيرمان" Spearman.

مثال (8-2):

فيما يلي بيانات درجة تفضيل ومستوى دخل لعشرة مستهلكين لسلعة ما كعينة، والمطلوب معرفة الارتباط بين مستوى الدخل ودرجة تفضيل المستهلكين للسلعة، وما مدلول ذلك بالنسبة لجميع المستهلكين:

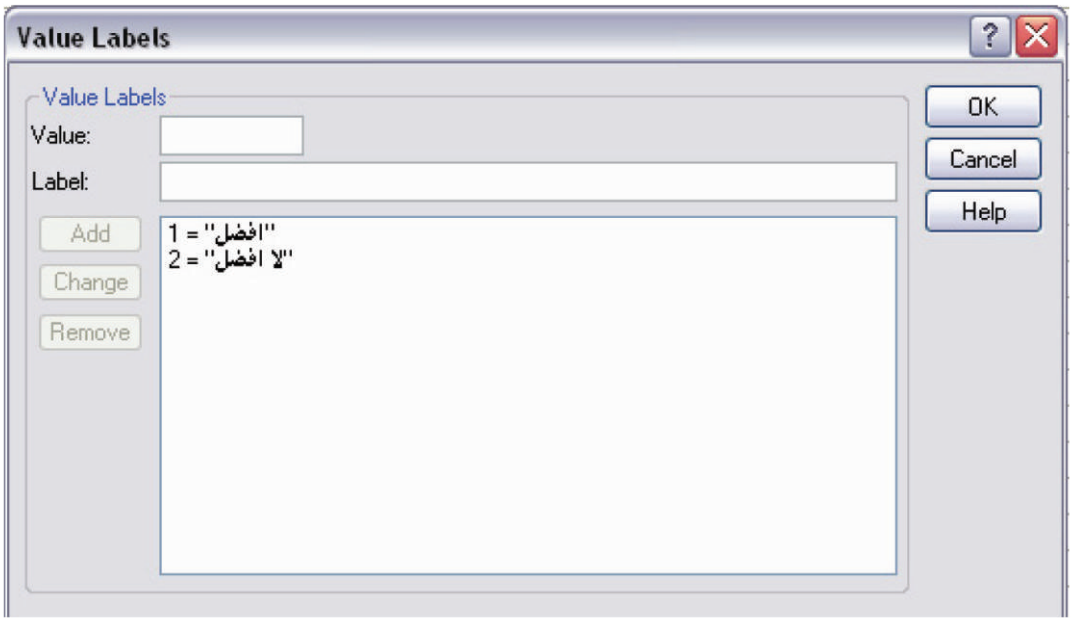
جدول (8-3)

مستوى التفضيل	أفضل	لا أفضل	أفضل	لا أفضل	لا أفضل	لا أفضل	أفضل	لا أفضل	أفضل	أفضل
مقدار الدخل الشهري	100	200	250	400	130	270	700	500	330	200

الحل:

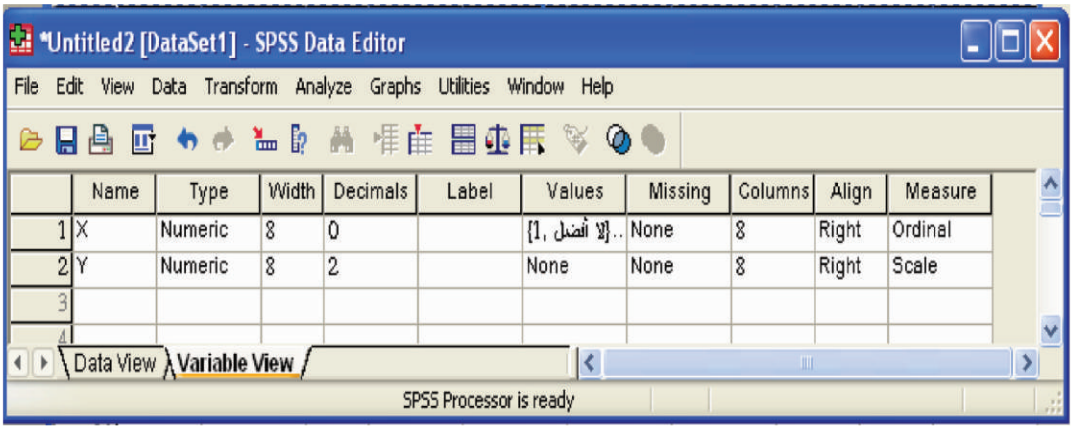
خلال البرنامج نقوم بإدخال البيانات على فرض أن "أفضل" نرمز لها بالرمز 1 و"لا أفضل" برمز 2.

ونقوم بتعريف المتغير الترتيبي x كما يلي في محرر المتغيرات نقوم بتعريف قيم المتغير في عمود values كما يلي:



شكل (8-3)

ثم في عمود measure نعرف المتغير X بأنه متغير ترتيبي ordinal، ونعرف المتغير Y بأنه متغير كمي عادي فيظهر المحرر كالآتي:



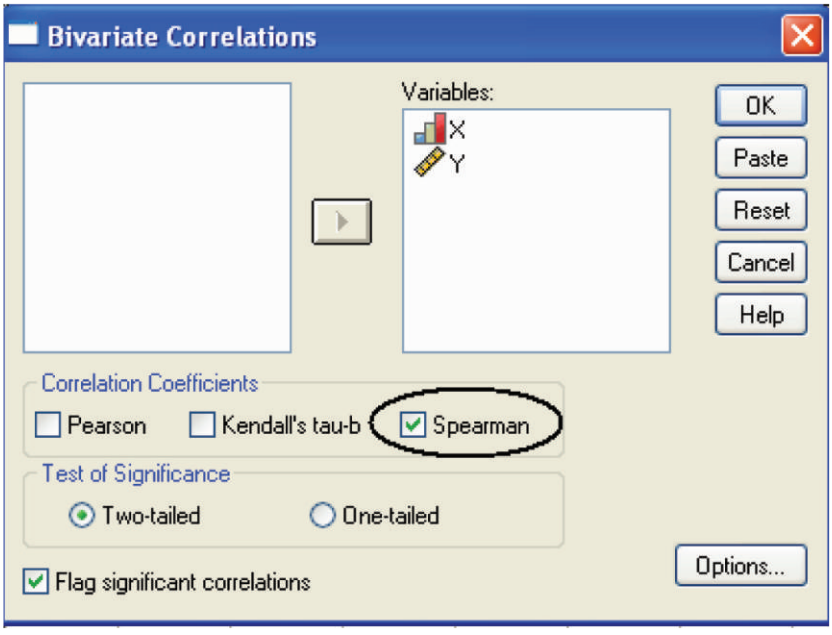
شكل (8-4)

ثم شاشة محرر البيانات نقوم بإدخال البيانات السابقة فتظهر في الشكل الآتي:

	X	Y
1	1	100.00
2	2	200.00
3	1	250.00
4	2	400.00
5	2	130.00
6	2	270.00
7	1	700.00
8	2	500.00
9	1	330.00
10	1	200.00

شكل (8-5)

من شريط قوائم نختار Analyze ثم نختار Correlate ثم نختار Bivariate فيظهر الشكل الآتي:



شكل (8-6)

ونلاحظ أن المتغير X معرف في عمود المتغير المستخدم في التحليل (العمود الأول من اليمين)، وأننا اخترنا Spearman كمعامل للارتباط، والاختيارات السابقة نفسها في مقياس Pearson ثم ننقر على ok فيظهر الجدول الآتي:

جدول (8-4)

Correlations

Control Variables			x	z
y	x	Correlation	1.000	.693
		Significance (2-tailed)	.	.057
		df	0	6
	z	Correlation	.693	1.000
		Significance (2-tailed)	.057	.
		df	6	0

ونجد أن معامل الارتباط بين درجة تفضيل المستهلكين لسلعة ومستوى دخولهم مقداره 0.070، وهذا يعني أن الارتباط طردي وضعيف ولاختبار الفرض القائل إن:

$$H_0: \rho = 0$$

ضد الفرض البديل القائل إن:

$$H_A: \rho \neq 0$$

ونلاحظ أن الفرض البديل معامل الارتباط لا يساوي صفرًا، وهذا ما دل عليه اختيارنا السابق أن الاختبار ذو اتجاهين، فإذا افترضنا أننا نريد اختبار ذلك بمستوى معنوية مقدارها $\alpha = 0.05$، فإن قيمة P-eulav $= 0.05 = \alpha$ < 0.424 دل ذلك على أن الارتباط بين تفضيل المستهلكين لسلعة ومستوى دخولهم ارتباط يساوي الصفر بمستوى معنوي مقداره 0.05، أي إننا نقبل فرض العدم القائل: إنَّ الارتباط يساوي صفرًا في كامل بيانات المجتمع.

(8 - 3) الارتباط الجزئي Partial Correlation

يقيس معامل الارتباط الجزئي قوة العلاقة بين متغيرين بثبوت متغير ثالث أو أكثر. مثلا قد نحصل على قيمة عالية لمعامل الارتباط البسيط للعلاقة بين أسعار اللحوم البيضاء واللحوم الحمراء، فقد لا توجد علاقة فعلية بين المتغيرين، ولكن كلا المتغيرين يتأثر بمعامل ثالث هو المستوى العام للأسعار، فإذا استبعدنا المستوى العام للأسعار (أو تثبيته) عند قياس العلاقة بين أسعار اللحوم البيضاء والحمراء، فسيتم الحصول على قيمة أقل لمعامل الارتباط، وهذا يعرف بالارتباط الجزئي . علما بأنه يمكن استبعاد أي عدد من المتغيرات عند قياس العلاقة بين ظاهرتين.

والآتي مثال يوضح كيفية تطبيق ذلك من خلال البرنامج:

مثـال (8-3):

البيانـات الآتية تعطي الدخل الشهري لمجموعـة أسر (X) والإنفاق الشهري على الطعام (Y) وحجم مدخراتها الشهرية (Z) وذلك في عينة من تسع أسر:

جدول (8-5)

الأسرة	1	2	3	4	5	6	7	8	9
x	40	65	90	60	75	80	120	45	65
Y	22	28	40	20	32	32	50	20	35
Z	6	10	8	25	12	7	10	5	7

المطلوب:

1 – حساب معامل الارتباط الجزئي لـx وy بثوت z.

2 – حساب معامل الارتباط الجزئي لـx وz بثوت y.

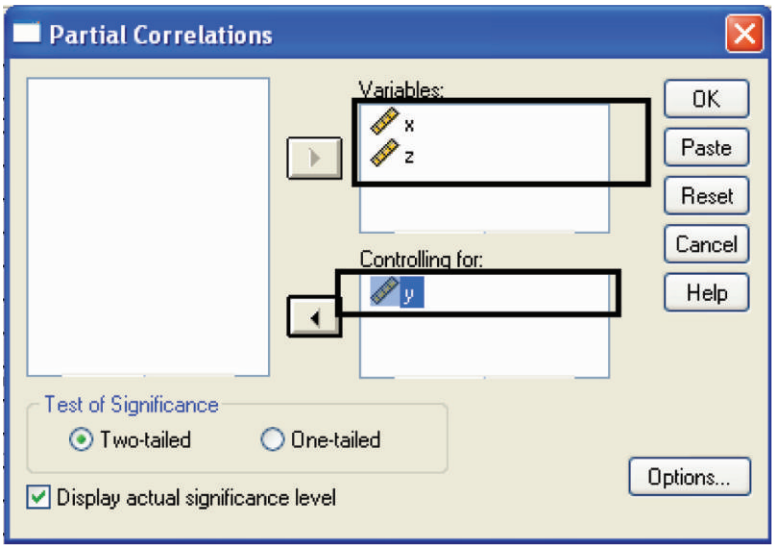
3 – حساب معامل الارتباط الجزئي لـz وy بثوت x.

4 – اختبار معنوية (دلالة) الارتباط بمستوى معنوية قدره 0,05.

الحل:

أولا: حساب معامل الارتباط الجزئي لـx وy بثوت z.

مـن شريـط قوائم نختار Analyze ثم نختـار Correlate ثم نختار Partial فيظهـر مربع حوار معامل الارتباط الجزئي الآتي في الشكل الآتي:



شكل (8-7)

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

وعند النقر على زر ok نحصل على النتيجة الآتية:

جدول (8-6)

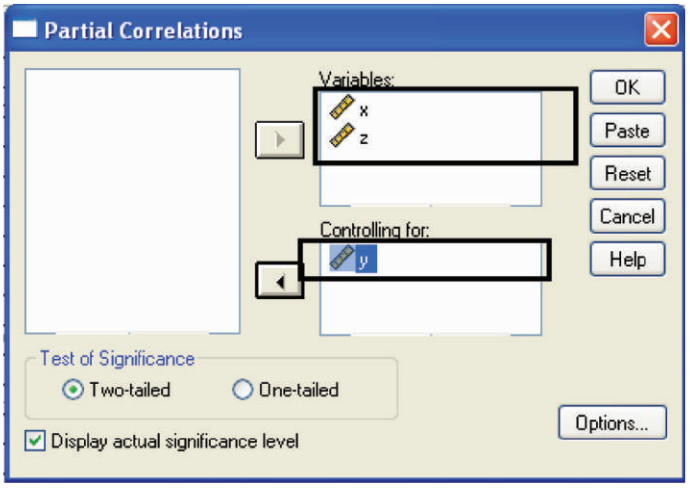
Correlations

Control Variables			z	y
x	z	Correlation	1.000	-.710
		Significance (2-tailed)	.	.048
		df	0	6
	y	Correlation	-.710	1.000
		Significance (2-tailed)	.048	.
		df	6	0

نلاحظ أن قيمة معامل الارتباط من الشكل السابق تساوي 0,965، أي إنه توجد علاقة طردية وقوية بين الدخل والإنفاق، كما أن قيمة p-value=0 وهي أقل من قيمة مستوى المعنوية مقسومًا على 2=0,025، حيث إننا نختبر فرض عدم ذا اتجاهين، ولذلك فإننا نقبل الفرض البديل القائل إن الارتباط بين المتغيرات في المجتمع لا يساوي صفرًا.

ثانياً: حساب معامل الارتباط الجزئي لـx و z بثوت y.

من شريط قوائم نختار Analyze ثم نختار Correlate ثم نختار Partial فيظهر مربع حوار معامل الارتباط الجزئي الآتي في الشكل الآتي:



شكل (8-8)

يتم إدخال المتغيرات (x،z) في خانة variables، وفي خانة Controlling for يتم إدخال المتغير (y).

عند النقر على زر ok نحصل على النتيجة الآتية:

جدول (8-7)

Correlations

Control Variables			x	z
y	x	Correlation	1.000	.693
		Significance (2-tailed)	.	.057
		df	0	6
	z	Correlation	.693	1.000
		Significance (2-tailed)	.057	.
		df	6	0

أن قيمة معامل الارتباط من الشكل السابق تساوي 0 ,693 أي إنه توجد علاقة طردية بين الدخل الادخار، كما أن قيمة p-value=0 ,057 وهي أكبر من قيمة مستوى المعنوية مقسومًا على 2، حيث إننا نختبر فرض عدم ذا اتجاهين، ولذلك فإننا نقبل الفرض العدم القائل إن الارتباط بين المتغيرات في المجتمع يساوي صفرًا.

ثالثاً: حساب معامل الارتباط الجزئي لـz و y بثوت x.

كما فعلنا في أولا وثانيًا من خلال إدخال وتعريف المتغيرات نفعل في تلك الخطوة وتظهر نتيجة معامل الارتباط كآلاتي:

جدول (8-8)

Correlations

Control Variables			z	y
x	z	Correlation	1.000	-.710
		Significance (2-tailed)	.	.048
		df	0	6
	y	Correlation	-.710	1.000
		Significance (2-tailed)	.048	.
		df	6	0

فنجد أن قيمة معامل الارتباط بين الاستهلاك والادخار تساوي 0 ,710 -، وهذا يعني أنه توجد علاقة بين الاستهلاك والادخار عكسية وقوية، ونجد أيضا أن قيمة p-value=0 ,048 وهي أكبر من قيمة مستوى المعنوية مقسومًا على 2=0 ,025، حيث إننا نختبر فرض عدم ذا اتجاهين، ولذلك فإننا نقبل الفرض العدم القائل إن الارتباط بين المتغيرات في المجتمع يساوي صفرًا في ضوء بيانات العينة.

الفصل التاسع

تحليل الانحدار
Regression Analysis

(9-1) مقدمة :

إن نموذج الانحدار يعبر عن علاقة بين متغير تابع وبين واحد أو أكثر من المتغيرات المستقلة، فإذا احتوى النموذج على متغير مستقل واحد فيعرف بنموذج الانحدار البسيط Regression Model Simple أو ما يسمى Ordinary Least Squares، وإذا احتوى على أكثر من متغير مستقل فهو نموذج الانحدار المتعدد، كما أن النموذج قد يكون خطياً أو غير خطي.

(9-2) نموذج الانحدار الخطي البسيط Simple Regression

إن الغرض من استخدام أسلوب تحليل الانحدار الخطي البسيط، هو دراسة وتحليل أثر متغير كمي على متغير كمي آخر، ومن الأمثلة على ذلك ما يلي:

- دراسة أثر كمية السماد في إنتاجية القمح.
- دراسة أثر الإنتاج في التكلفة.
- دراسة أثر كمية البروتين التي يتناولها الأبقار في الزيادة في الوزن.
- أثر الدخل في الإنفاق الاستهلاكي.
- دراسة أثر الذكاء في التحصيل الدراسي.

وهكذا هناك أمثلة في كثير من النواحي الاقتصادية، والزراعية، والتجارية، والعلوم السلوكية، وغيرها من المجالات الأخرى.

في تحليل الانحدار البسيط، نجد أن الباحث يهتم بدراسة أثر أحد المتغيرين ويسمى بالمتغير المستقل أو المتنبأ منه، على المتغير الثاني ويسمى بالمتغير التابع أو المتنبأ به، ومن ثم يمكن عرض نموذج الانحدار الخطي في شكل معادلة خطية من الدرجة الأولى، تعكس المتغير التابع كدالة في المتغير المستقل كما يلي:

$$y = \beta_0 + \beta_1 \chi + e \qquad (1-9)$$

حيث إن:

- ✓ y : هو المتغير التابع (الذي يتأثر).
- ✓ x : هو المتغير المستقل (الذي يؤثر).

✓ β_0 : هو الجزء المقطوع من المحور الرأسي y، وهو يعكس قيمة المتغير التابع في حالة انعدام قيمة المتغير المستقل χ، أي في حالة $\chi = 0$.

✓ β_1 : ميل الخط المستقيم $(\beta_0 + \beta_1\chi)$، ويعكس مقدار التغير في y إذا تغيرت بوحدة واحدة.

✓ e : هو الخطأ العشوائي، الذي يعبر عن الفرق بين القيمة الفعلية، والقيمة المقدرة $\hat{y} = \beta_0 + \beta_1\chi$، أي إن $e = y-(\beta_0 + \beta_1\chi)$.

تقدير نموذج الانحدار الخطي البسيط

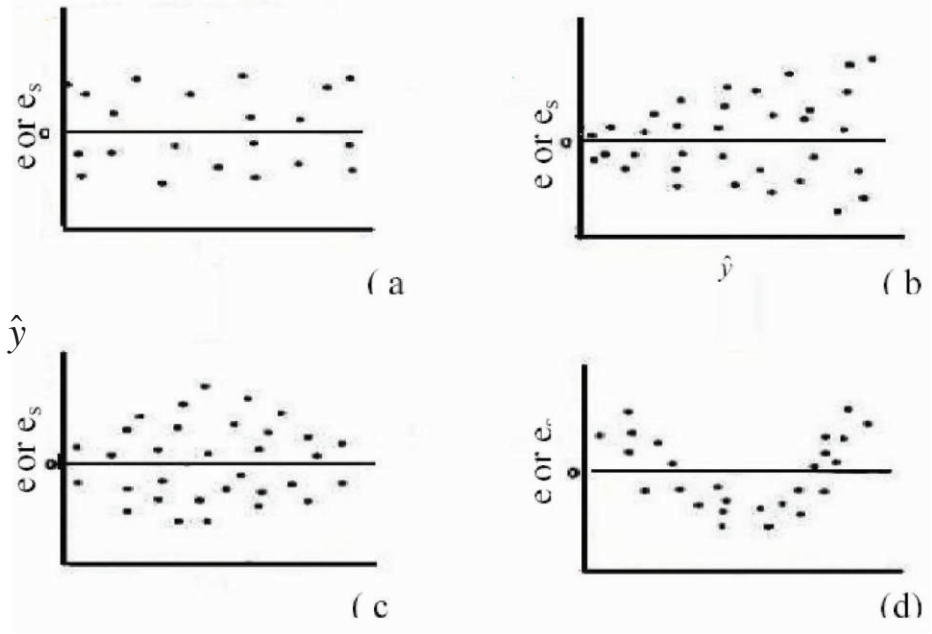
يمكن تقدير معاملات الانحدار $(\beta_0$ و $\beta_1)$ في النموذج (9-1) باستخدام طريقة المربعات الصغرى، وهذا التقدير هو الذي يجعل مجموع مربعات الأخطاء العشوائية $\sum e^2 = \sum (y-(\beta_0 + \beta_1\chi))^2$ أقل ما يمكن، وتكون القيمة المقدرة للمتغير التابع هو: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1\chi$ ، ويطلق على هذا التقدير" تقدير معادلة انحدار y على x.

علماً بأن نموذج الانحدار الخطي البسيط يجب أن يحقق مجموعة الفروض الآتية:

- وجود علاقة خطية بين X وY.
- أن الأخطاء العشوائية تتوزع بمتوسط مساوٍ للصفر.
- أن الأخطاء العشوائية لها تباين ثابت يساوي σ^2 (فرضية تجانس تباين الخطأ العشوائية Ho-moscedasticity).
- الأخطاء تتوزع طبيعياً وهذا الشرط ليس ضرورياً لتقدير المعالم بطريقة OLS، ولكنه ضروري لاختبار الفرضيات المعلقة بمعاملات الانحدار β_0 و β_1.

عدم وجود ارتباط ذاتي Autocorrelation بين الأخطاء العشوائية.

ويمكن التحقق من توافر فرضيات النموذج الخطي البسيط من خلال تخطيط Scatter plots بتمثيل $\hat{y}$ على المحور الأفقي أو (X) يقابله الخطأ العشوائي e أو الأخطاء المعيارية Standardized Residual التي يرمز لها e_s على المحور الرأسي، وذلك كما هو موضح في الشكل الآتي:



(a) توافر فروض التحليل جميعها (عدم وجود مشكلة).

(b) زيادة تباين الخطأ العشوائي بزيادة $\hat{y}$.

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

(d) عدم ملاءمة العلاقة الخطية (يجب استعمال نماذج أخرى مثلا نموذج الدرجة الثانية).

وفيما يلي نبين كيف يمكننا اختبار هذه الشروط وكيفية تقدير معادلة الانحدار وتحديد الأخطاء باستعمال البرنامج:

تطبيقات:

مثال (9-1)

فيما يلي بيانات عن كمية البروتين اليومي بالجرام التي يحتاج إليها العجل الرضيع، ومقدار الزيادة في وزن العجل بالكليو جرام، وذلك لعينة من العجول الرضيعة حجمها 10.

جدول (9-1)

كمية البروتين X	10	11	14	15	20	25	46	50	59	70
الزيادة في الوزن Y	10	10	12	12	13	13	19	15	16	20

والمطلوب:

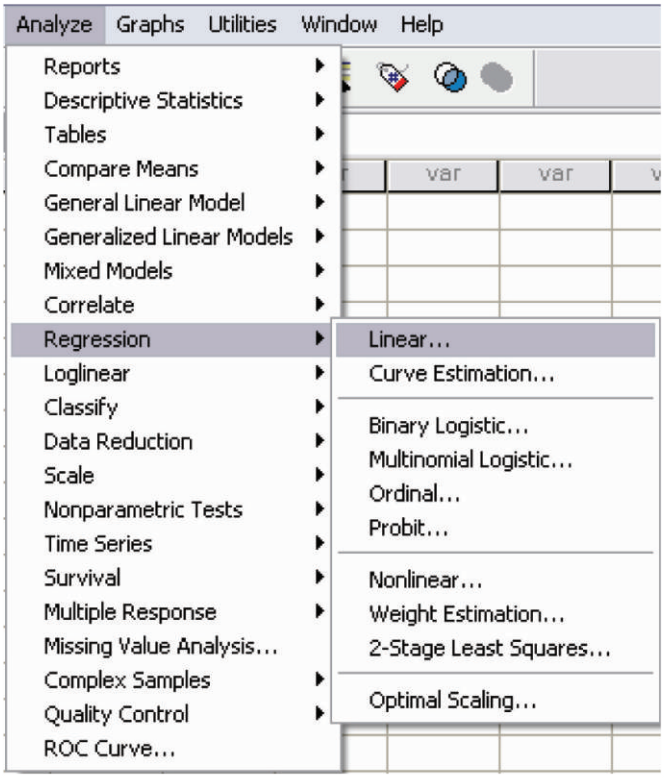
- **قدر معادلة انحدار الوزن** y على كمية البروتين x.

- فسر معادلة الانحدار.
- ما مقدار الزيادة في الوزن عند إعطاء العجل 50 جرامًا من البروتين؟
- وما مقدار الخطأ العشوائي؟
- استخراج فترة ثقة 95% لكل من معلمتي الانحدار β_0 و β_1 .
- استخراج جدول تحليل التباين ANOVA للمعاملات.
- اختبر جودة توفيق النموذج الخطي (باستعمال معامل التحديد R^2) مع تحليل الأخطاء العشوائية بالرسم البياني.
- اختبر التوزيع الطبيعي للأخطاء العشوائية بيانياً.

الحل:

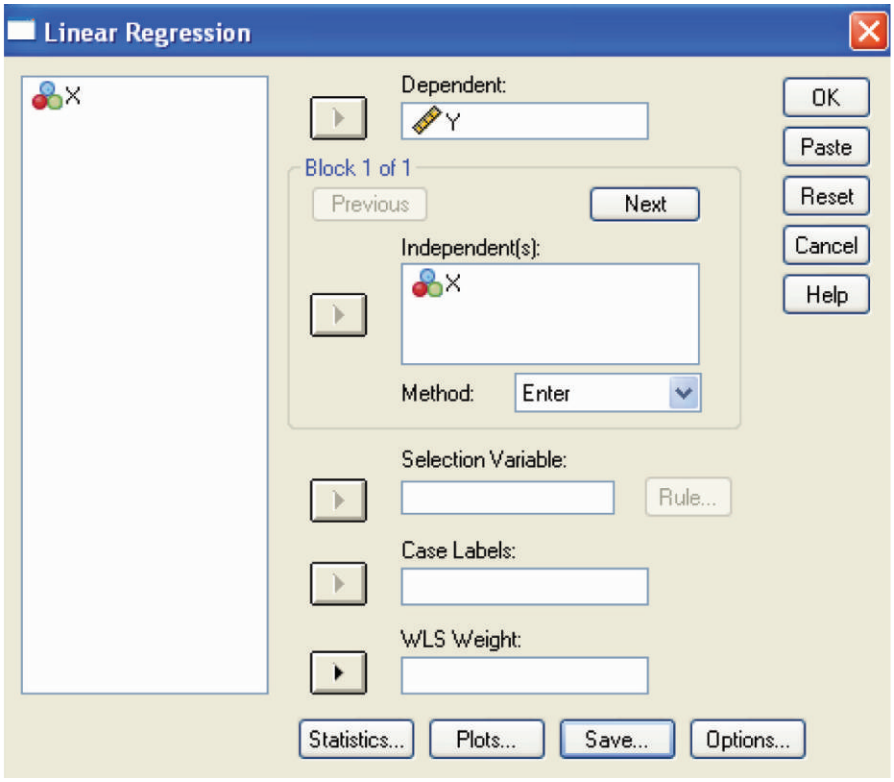
لتنفيذ المطلوب السابق باستخدام البرنامج نتبع الخطوات الآتية:

Analyze ثم ⇐ Regression ⇐ Linear كما في الشكل (9-1):



شكل (9-1)

فيظهر صندوق الحوار الآتي:



شكل (9-2)

ونجد أن الشكل (9-2) يتكون من:

Dependent: يمثل المتغير التابع.

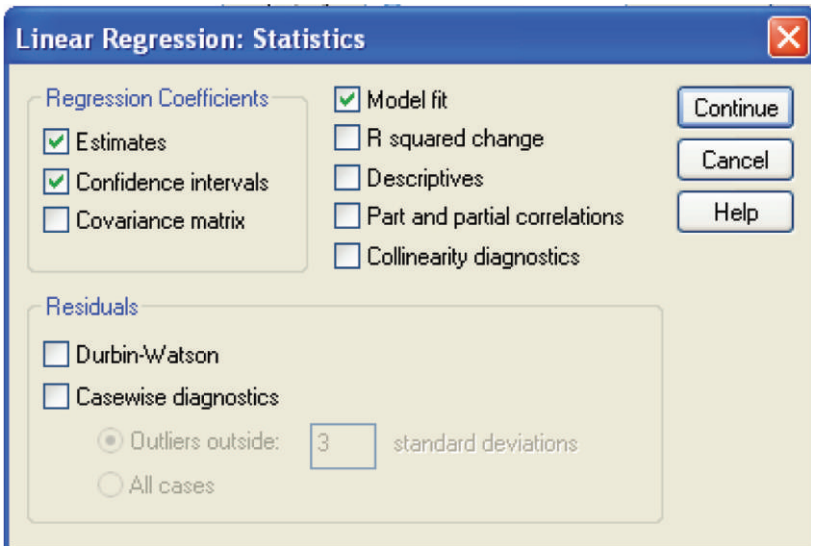
Independent: وهو يمثل المتغير المستقل ويمكننا إدخال مجموعة من المتغيرات المستقلة هنا كما سوف نرى في مبحث الانحدار المتعدد، ولأننا نريد معادلة انحدار لـ y على X فإننا نعني أن y متغير تابع وX متغير مستقل.

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

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

Case Lebels: متغير يستخدم قيمة كعناوين لنقاط شكل الانتشار Scatterplots.

Statistics: عند النقر على هذا الزر يظهر مربع الحوار الآتي:



شكل (9-3)

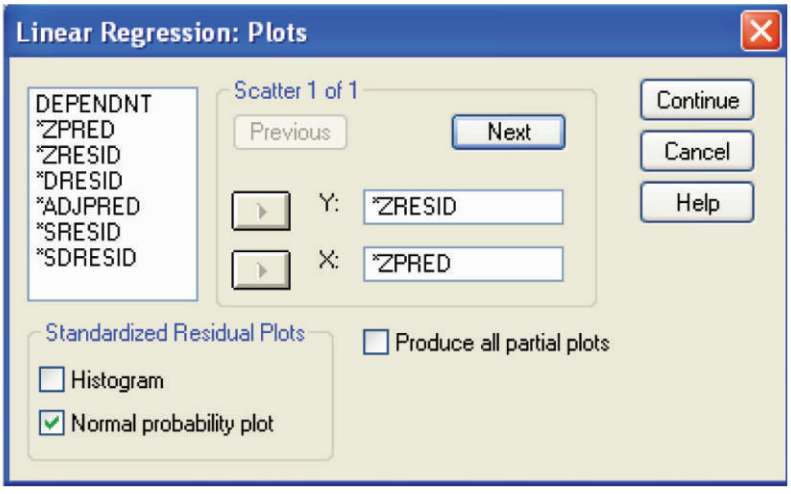
وقد تم تأشير الخيارات الآتية:

✓ Estimate: لتقدير معالم نموذج الانحدار واختبارات t المرافقة.

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

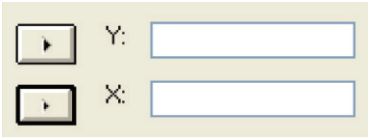
✓ Model Fit: لعرض R^2 و ANOVA.

وبعـد الانتهاء النقر على زر Continue. وهناك زر آخر في شكل (9-2) بعنوان Plots وعند النقر عليه يظهر مربع الحوار الآتي:



شكل (9-4)

ونلاحظ أننا اخترنا Normal probability plot: لاختبار التوزيع الطبيعي للأخطاء العشوائية المعيارية (المطلوب ن). كما أننا نجد أعلى شكل (9-4) الشكل الآتي:



في هـذا الشـكل يتم إدخـال المتغيرات التي نـود ان تظهر في شـكل تحليل الأخطاء في المحـور السيني والمحور الصادي والمتغيرات موجودة في الجانب الأيمن من الشكل السابق وتوجد عدة خيارات:

1- (*ZPRED) Standardized predicted values: القيمة المعيارية للقيمة **المتوقعة** $\hat{y}$.

2- (*ZRESID) Standardized residuals: القيمة المعيارية للأخطاء.

3- (*DRESID) Deleted residuals: حذف الأخطاء.

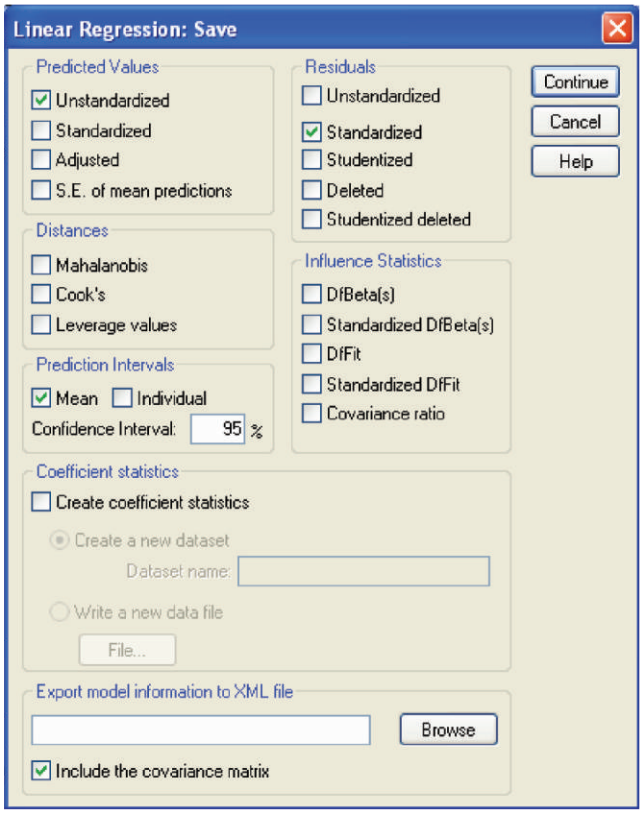
4- (*ADJPRED) Adjusted predicted values: القيمة المعيارية المعدلة لقيمة متوقعة $\hat{y}$.

5- (*SRESID) Studentized residuals: جعل الأخطاء تتبع توزيع t.

6- (*SDRESID) Studentized deleted residuals: الأخطاء المحذوفة التي تتبع توزيع t.

وفي تلك الحالة يتم تخصص كل من: ZPRED لمحور السينات، ZRESID لمحور الصادات وبعد الانتهاء يتم النقر على زر Continues.

وهناك رز آخر في شكل (9-2) بعنوان Save وعند النقر عليه يظهر مربع الحوار الآتي:



شكل (9-5)

ونلاحظ أننا اخترنا Unstandardized predicted Values أي $\hat{y}$ وكذلك Standardized Residual أي e_s سيتم استعمال هذين المتغيرين لرسم Scatterplots الجزء الثاني من المطلوب م، علماً بأنه سيتم إضافة هذين المتغيرين إلى ورقة Data Editor إلى جانب متغيرات X و Yو Observat حيث يضاف المتغير Unstandardized predicted Values باسم Pre_1 ويضاف Standardized Residual باسم Zre_1. وبعد الانتهاء ينقر زر Continues.

و في النهاية ننقر على Ok في الشكل (2-10) لإظهار النتائج كالآتي:

جدول (2-9)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	9.436	.864		10.926	.000	7.444	11.427
	X	.143	.023	.913	6.325	.000	.091	.195

a. Dependent Variable: Y

من خلال الجدول أعلاه يمكن كتابة النموذج كما يلي:

$$\hat{y} = 9.436 + 0.143\,\chi$$

وتشير المعادلة السابقة إلى أن زيادة البروتين مقدار جرام تعمل على زيادة الوزن بمقدار 0.143 جرام، بينما توجد زيادة ليست متعلقة بالبرويتن مقدارها 09.436 لكل جرام في الوزن، وعليه فإن زيادة البروتين بمقدار 50 جرامًا تعمل على أن يكون

$$\hat{y} = 9.436 + 0.143\,(50)$$
$$= 16.436$$

ومن ثم مقدار الخطأ العشوائي يساوي:

$$\hat{e}_{\chi=50} = y_{\chi=50} - \hat{y}_{\chi=50}$$
$$= 15 - 16.436$$
$$= -1.43$$

ويستعمل اختبار t لاختبار الفرضية الآتية لمعلمة الميل β_1:

$$H_0 : \beta_1 = 0$$
$$H_A : \beta_1 \neq 0$$

ويستعمل اختبار t لاختبار الفرضية الآتية لمعلمة الميل β_0:

$$H_0 : \beta_0 = 0$$
$$H_A : \beta_0 \neq 0$$

نستخدم قيمة *P-value* المرافقة لإحصائية t للمعلمة في الاختبار كما يلي:

إذا كانت *P-value<0.05* نرفض فرضية العدم بمستوى 5%.

إذا كانت *P-value<0.01* نقبل فرضية العدم بمستوى 1%.

عكس هذا نقبل فرضية العدم.

ونلاحظ أن *P-value* لمعلمة الميل تساوي صفرًا، وهي أقل من 0.01، إن *P-value* لمعلمة الحد الثابت تساوي صفرًا، وهي أقل من 0.01، ولهذا نرفض فرضي العدم لكل من المعلمتين، أي إن كلا من المعلمتين تختلف جوهرياً عن الصفر. إن ظهور معلمة الميل معنوية يعكس أهمية متغير البروتين في التأثير على الوزن في النموذج.

ويمكن كتابة فترة الثقة للحد الثابت، وذلك باستعمال نتائج الشكل (10-6) كما يلي:

$$\Pr[7.444 \leq \beta_0 \leq = 11.427] = 95\%$$

حيث إن Pr تمثل الاحتمال كما يمكن كتابة فترة ثقة 95% لمعلمة الميل كما يلي:

$$\Pr[0.091 \leq \beta_0 \leq = 0.195] = 95\%$$

الجدول الآتي يعرف بجدول تحليل التباين ANOVA، ويشتمل على إحصائية F لاختبار الفرضية الخاصة بمعلمة الميل β_1 نفسها، وهذا الاختبار مكافئ تماماً لاختبار t لمعلمة الميل (لاحظ أن قيمة P-value متساوية لكلا الاختبارين) علما بأن $F = t^2$.

جدول (9-3)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	90.000	1	90.000	40.001	.000[a]
	Residual	18.000	8	2.250		
	Total	108.000	9			

a. Predictors: (Constant), X

b. Dependent Variable: Y

الجدول الآتي يتضمن أهم مؤشر لنموذج الانحدار وهو معامل التحديد Coefficient of Determination ويرمز له R^2 ويعد مقياساً لجودة توفيق النموذج:

جدول (9-4)

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.913[a]	.833	.813	1.49999

a. Predictors: (Constant), X

b. Dependent Variable: Y

ويحتسب من جدول تحليل التباين كما يلي:

$$R^2 = \frac{\text{Explained Variations}}{\text{Total Variations}}$$

$$= \frac{SSR}{SST}$$

$$= \frac{90}{108.00}$$

$$= 0.833333 \quad , 0 \le R^2 \le 1$$

وتفسير ذلك أن 83.333% من التباينات (الانحرافات الكلية من قيم المتغير y) تفسرها العلاقة الخطية أي نموذج الانحدار، وأن 15.76% من التباينات ترجع إلى عوامل عشوائية كأن تكون هناك متغيرات مهمة لم يتضمنها النموذج. وعلى العموم كلما اقتربت قيمة R^2 من 100% دل ذلك على جودة توفيق النموذج. هذا، أن $r = \sqrt{R^2} = 0.91287$، حيث إن r معامل الارتباط الخطي البسيط لبيرسون، وإن إشارة r هي إشارة معلمة الميل نفسها، أي إن الإشارة موجبة، وهذا يعني أن العلاقة بين مقدار البروتين والوزن علاقة طردية وقوية.

لاحظ أن معامل التحديد لو أضيف متغير مستقل للنموذج فإن قيمته سترتفع حتى لو لم تكن هناك أهمية للمتغير المستقل في النموذج، حيث إن إضافة متغير مستقل إلى نموذج الانحدار تؤدي إلى زيادة R^2 بسبب زيادة مجموع المربعات العائدة للانحدار SSR مع ثبات مجموع المربعات الكلية SST، ولهذا يتم احتساب معامل التحديد المصحح Adjusted R Square الذي يأخذ في الحسبان النقصان الحاصل في درجات الحرية، وقيمته دائما أقل من قيمة معامل التحديد (غير المصحح) وفي هذا المثال 81.3% ولهذا يمكن القول إن النموذج جيد التوفيق.

أمـا الخطـأ المعيـاري للتقديـر Standard Error of Estimate فيقيس تشتت القيـم المشاهدة عن خط الانحدار، وأن الحصول على قيمة صغيرة لهذا المؤشر يعني صغر الأخطاء العشوائية، ومن ثم جودة تمثيل خط الانحدار لنقاط شكل الانتشار.

لتحليل الأخطاء العشوائية بيانياً نكون شكل الانتشار Scatterplots بتمثيل القيم التقديرية $\hat{y}$ على المحور الأفقي والأخطاء المعيارية على المحور العمودي أو الرأسي كما يلي:

ففي شاشة Data Editor نجد أنه كما سبق أن أوضحنا أنه سوف تتم إضافة المتغير Unstandardized predicted Values باسم Pre_1 ويضاف Standardized Residual باسم Zre_1 كما يلي:

	X	Y	PRE_1	ZRE_1
1	10	10.00	10.86212	-.57475
2	11	10.00	11.00475	-.66984
3	14	12.00	11.43264	.37824
4	15	12.00	11.57527	.28315
5	20	13.00	12.28843	.47439
6	25	13.00	13.00158	-.00105
7	46	19.00	15.99684	2.00212
8	50	15.00	16.56736	-1.04491
9	59	16.00	17.85104	-1.23403
10	70	20.00	19.41998	.38668

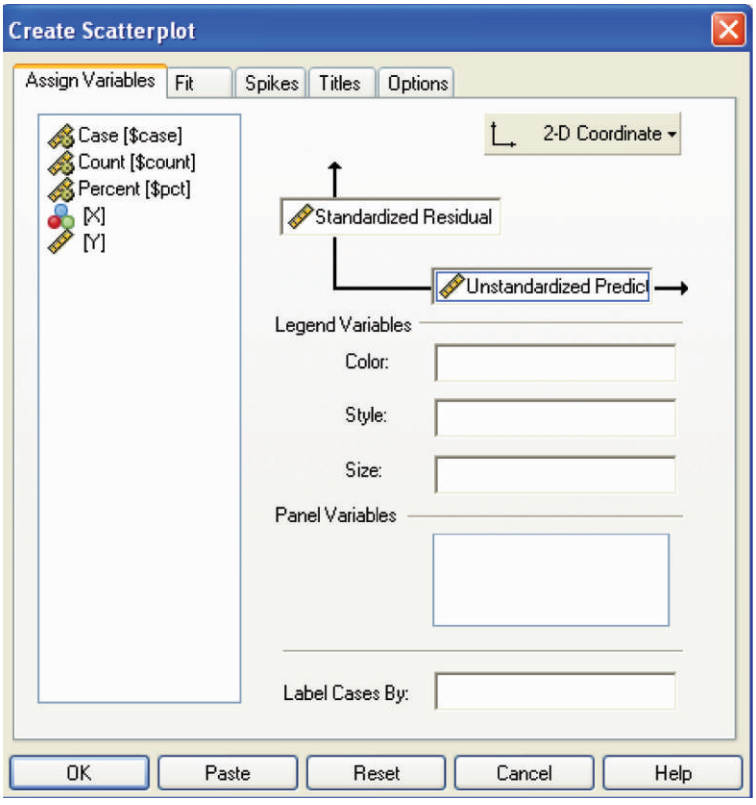
شكل (9-6)

ثـم بعد ذلـك من شريـط القوائـم نختـار Simple ⇐ Scatter ⇐ Interactive ⇐ Graphs فيظهر صندوق حوار بعنوان Create Scatterplots:

فنقوم بالنقر على Zre_1 لإدخاله إلى الخانة Y في صندوق الحوار وبالمثل

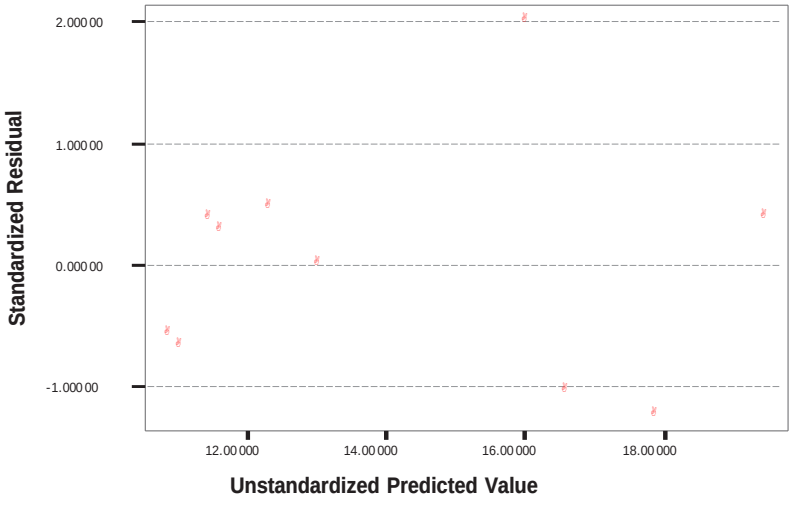
فنقوم بالنقر على Pre_1 لإدخاله إلى الخانة X في صندوق الحوار وبالمثل

فيظهر مربع الحوار Create Scatterplots كما يلي:



شكل (9-7)

فننقر على Ok فيظهر المخطط الآتي:



شكل (9-8)

نلاحظ أن النقاط تتوزع بشكل شريط أفقي متساو حول الصفر، ما يدل علي توافر فرضيات التحليل بصورة عامة، حيث لا يعاني النموذج مشكلة عدم تجانس تباين الخطأ العشوائي، ولا توجد حاجة لاستخدام علاقة من درجات أعلى.

مثال (9-2)

لدينـا شركة تقوم بصناعة قطع غيار معينة على دفعات شهريـة، وقد أعطتنا الشركة البيانات لعشر دفعات في الجدول الآتي:

جدول (9-5)

دورة الإنتاج	حجم الدفعة	ساعات العمل
1	30	73
2	20	50
3	60	128
4	80	170
5	40	87
6	50	108
7	60	135
8	30	69
9	70	148
10	60	132

تريـد الشركة معرفة نموذج الانحدار الخطي الخاص لهـا لكي تتنبأ بأحجام الدفعات المستقبلية بناء على ساعات العمل عند مستوى معنوية 0.05.

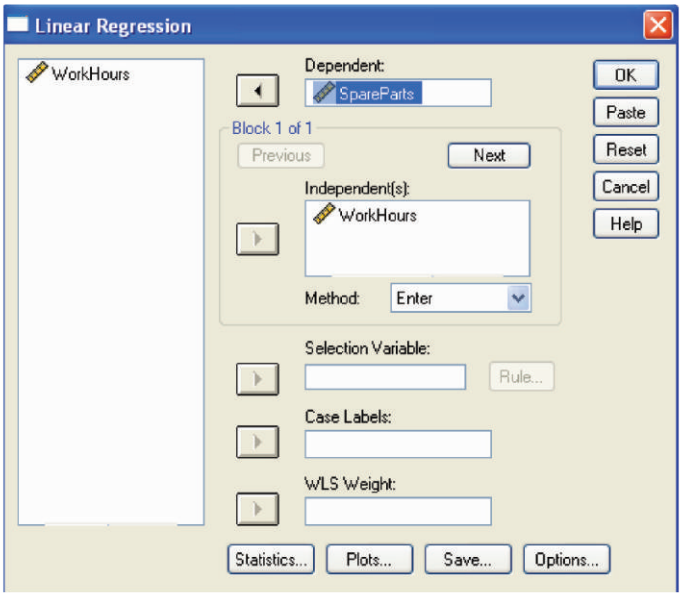
الحـل:

ندخل البيانات في عمودين أحدهما لحجم الدفعة من قطع الغيار Spare Parts والآخر لساعات العمل على كل دفعة Work Hours كما يأتي:

	SpareParts	WorkHours
1	30.00	73.00
2	20.00	50.00
3	60.00	128.00
4	80.00	170.00
5	40.00	87.00
6	50.00	108.00
7	60.00	135.00
8	30.00	69.00
9	70.00	148.00
10	60.00	132.00

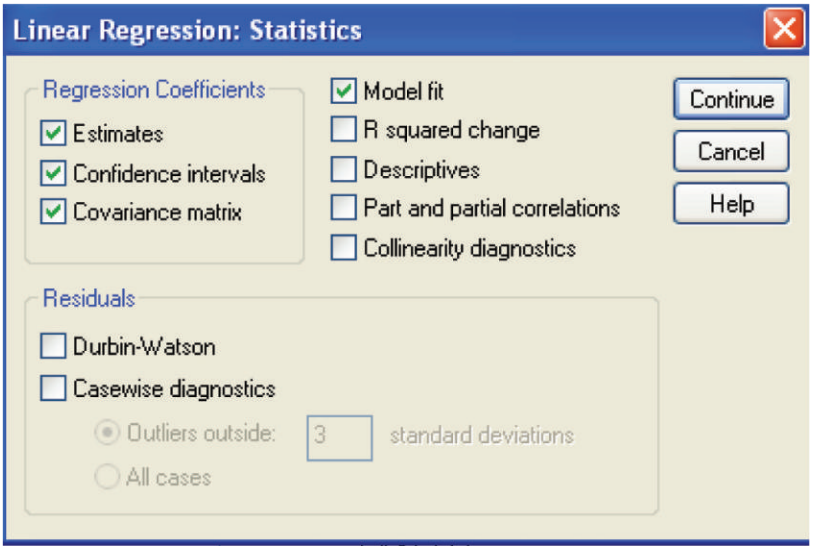
شكل (9-9)

نقـوم بتفعيـل نافذة الانحدار ونختار المتغير المستقل هو Work Hours، أما التابع فهو Spare Parts كما في الشكل الآتي:



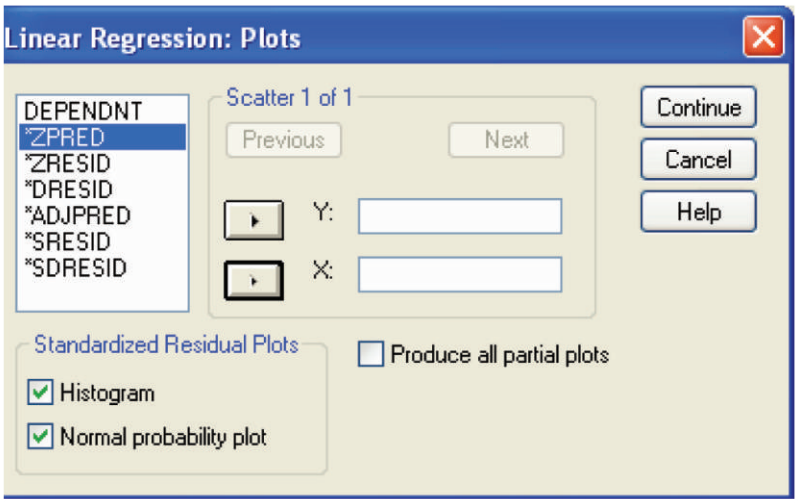
شكل (9-10)

نختـار الطريقـة المعيارية وهـي Enter في قائمة الخيـارات Method؛ لأنها الطريقـة الخاصة بتقدير معاملات الانحدار المجهولة للمتغيرات المستقلة، بعد ذلك نقوم باختيار الأمر الفرعي Statistics ونتأكد من اختيار كل من Model fit، Covariance matrix، Confidence intervals، Estimates، كما في الشكل الآتي:



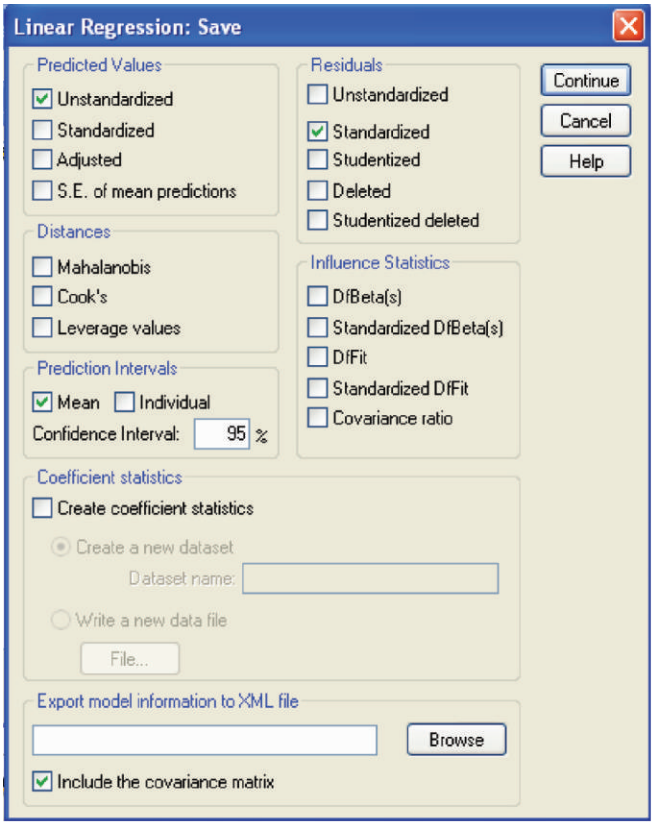
شكل (9-11)

ثـم نعود إلى مربع الحوار السابق بالنقر علـى Continue، ونذهب إلى الأمر الفرعي Plots ونتأكد من اختيار Histogram وNormal probability plot كما يأتي:



شكل (9-12)

ثـم نعود إلى مربع الحـوار السابق بالنقر على Continue ونذهب إلى الأمـر الفرعي Save ونتأكد من الاختيارات السابقة في المثال السابق كما يلي:



شكل (9-13)

ثم نعود إلى مربع الحوار السابق بالنقر على Continue ثم ننقر OK فتظهر لدينا النتائج الآتية:

يظهـر لدينا جدول يبـين الطريقة المستخدمة في الانحـدار والمتغير التابع وهـو Spare Parts والمتغيرات المستقلة الداخلة في نموذج الانحدار الخطي، وفي حالتنا يوجد متغير واحد فقط وهو Work Hours كما يلي:

جدول (9-6)

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Age, Weight[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: BloodPresure

الجدول الآتي يمثل بعض المقاييس التي تهمنا لمعرفة هل المتغيرات المستقلة لها تأثير كبير في النموذج أم لا هو:

جدول (9-7)

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.998[a]	.996	.995	1.36630

a. Predictors: (Constant), WorkHours

b. Dependent Variable: SpareParts

يظهر فيه معامل التحديد R وله قيمة عالية جدًا ولدينا أيضًا R Square المعدلة، و قيمته 0,996 أي إن المتغير المستقل يقوم بتفسير ما يقارب 99% من النموذج. ولدينا خطأ التقدير، وهو هنا خطأ صغير قيمته 1,36، وكلما اقتربت قيمته من الصفر كان أفضل.

جدول تحليل التباين:

جدول (9-8)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3385.066	1	3385.066	1813.333	.000[a]
	Residual	14.934	8	1.867		
	Total	3400.000	9			

a. Predictors: (Constant), WorkHours

b. Dependent Variable: SpareParts

من هذا الجدول نستطيع اختبار معنوية الانحدار وذلك باستخدام الاختبار الآتي:

الفروض الإحصائية:

H_0: الانحدار غير معنوي.

H_1: الانحدار معنوي.

خطوات الاختبار:

نقارن P-value = .000 بمستوى المعنوية 0.05 نلاحظ أن القيمة P-value أقل من مستوى المعنوية.

القرار الإحصائي:

نقوم برفض الفرض العدمي، وهذا معناه أن الانحدار معنوي (دال).

كما يمكننا معرفة مجموع المربعات الآتية ودرجات حرياتها:

- مجموع مربعات البواقي SSE = 14.934 بدرجة حرية 8
- مجموع مربعات الانحدار SSR = 3358.066 بدرجة حرية 1
- مجموع المربعات الكلي SST = 4300.00 بدرجة حرية 9

وأيضًا هناك متوسطات المربعات الآتية:

- MSE = 1.867 متوسط مربع الخطأ.
- MSR =3385.066 متوسط مربع الانحدار

جدول المعالم المقدرة:

جدول (9-9)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-4.758	1.357		-3.508	.008	-7.887	-1.630
	WorkHours	.498	.012	.998	42.583	.000	.471	.525

a. Dependent Variable: SpareParts

من هذا الجـدول نستطيع الحصول على تقديرات لقيم معالم الانحدار المجهولة، وبذلك نحصل على معادلة الانحدار المقدرة الآتية:

Spare Parts = -4.758 + 0.498Workhours

حيث:

$$\hat{\beta}_0 = -4.758$$
$$\hat{\beta}_1 = 0.498$$

مـن معادلة الانحدار المقدرة بإمكاننا التنبؤ بقطع الغيار التي سنحصل عليها بناء على ساعات العمل فمثلا إذا كانت ساعات العمل هي 70 ساعة فيكون الإنتاج المتوقع لقطع الغيار كالآتي:

Spare Parts = -4.758 + .498 (70)

=30.102

أي تقريبًـا سيكـون الإنتاج 30 قطعة غيار، كما يمكن من هذا الجـدول معرفة الخطأ المعياري لكل من التقديـرات و فترات الثقة لكل تقدير. وأيضًا يمكن إجـراء اختبارات المعنوية لكل من هذه التقديرات على حدة.

فلنختبر معنوية β_1 عند مستوى معنوية 0.05:

الفروض الإحصائية:

$$H_0 : \beta_1 = 0$$
$$H_A : \beta_1 \neq 0$$

خطوات الاختبار:

نقارن P-value = .008 بمستوى المعنوية 0.05 نلاحظ أن القيمة P-value أقل من مستوى المعنوية.

القرار الإحصائي:

نقوم برفض الفرض العدمي وهذا معناه أن β_0 معنوي.

جدول البواقي:

جدول (9-10)

Residuals Statistics(a)

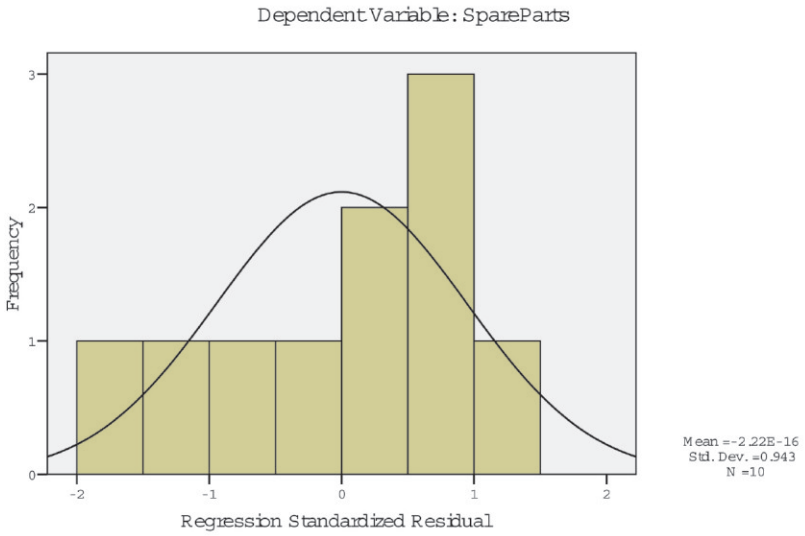
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	20.1318	79.8682	50.0000	19.39377	10
Std. Predicted Value	1.540 -	1.540	000.	1.000	10
Standard Error of Predicted Value	433.	824.	597.	137.	10
Adjusted Predicted Value	20.2070	79.7930	50.0073	19.35863	10
Residual	-2.44510	1.44949	00000.	1.28816	10
Std. Residual	-1.790	1.061	000.	943.	10
Stud. Residual	-1.936	1.143	003.-	1.026	10
Deleted Residual	-2.86228	1.68296	00726.-	1.52759	10
Stud. Deleted Residual	-2.485	1.169	067.-	1.153	10
Mahal. Distance	003.	2.372	900.	852.	10
Cook's Distance	004.	320.	089.	103.	10
Centered Leverage Value	000.	264.	100.	095.	10

a Dependent Variable: Spare Parts

الأشكال البيانية:

- المدرج التكراري: ويستخدم لمعرفة إذا كانت البيانات تتوزع توزيعًا طبيعيًا أم لا، والمدرج التكراري الخاص بالنموذج هو:

Histogram



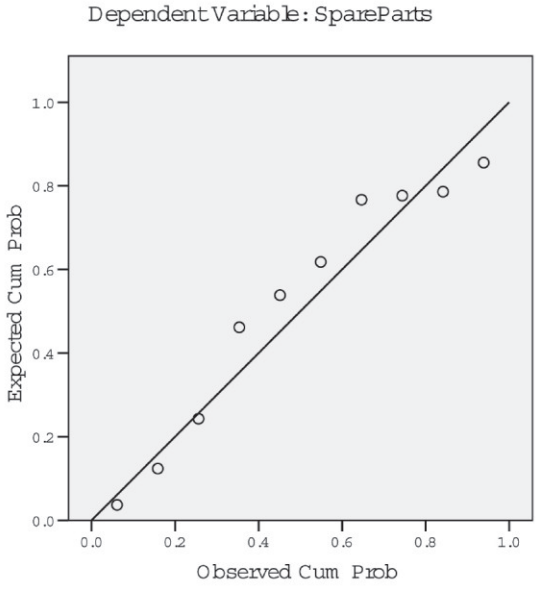
شكل (10-14)

نلاحظ أن شكله قريب من شكل التوزيع الطبيعي، وهذا معناه أن البيانات تتبع التوزيع الطبيعي ونستطيع تطبيق تحليل الانحدار عليها.

رسم منحى الانحدار:

نشاهد أن النقاط تتوزع بشكل متساو، حيث إن النقاط الأعلى تساوي النقاط الأدنى من الخط، ولذلك فإن فروض النموذج محققة فيه.

Normal P-Pot of Regression Standardized Residual



شكل (9-15)

(3-9) الانحدار الخطي المتعدد Multiple Linear Regrassion

تحليل الانحدار المتعدد هو من بين الأدوات الإحصائية كافة، الأداة المستخدمة على أوسع نطاق، الفرق الأساسي بينه وبين الانحدار الخطي البسيط هو أنه هنا لدينا متغير تابع واحد ومتغيرات مستقلة متعددة لها تأثيرات مختلفة في هذا المتغير التابع، ويكون شكل النموذج الخطي العام الذي يحتوي على عدد p-1 متغيرات مستقلة هو:

$$Y_1 = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

حيث p هو عدد المعالم المجهولة في النموذج، ونلاحظ هنا أنه إذا وضعنا p=2 نحصل على النموذج الخطي البسيط.

تدعى المعالم $\beta_1, \beta_2, \ldots, \beta_{p-1}$ بمعالم انحدار جزئية Partial regression Coefficients أو الميول الجزئية؛ لأن كل واحد منها يعكس تأثيرًا جزئيًا على المتغير التابع للمتغير المستقل الخاص به عندما يكون المتغير المستقل الآخر ثابتًا.

إن فروض النموذج الخطي المتعدد هي فروض النموذج البسيط نفسها يضاف إلى ذلك فرض عدم وجود ارتباط خطي متعدد بين المتغيرات المستقلة (Multicollinearity).

مثال (9-3)

نريد دراسة العلاقة بين ضغط دم الأطفال الرضع وعمرهم بالأيام ووزنهم عند الولادة مقاسًا بالأونصة عن طريق نموذج انحدار خطي متعدد لستة عشر رضيعًا باستخدام البيانات الآتية:

جدول (9-11)

عدد	الوزن عند الولادة	العمر بالأيام	ضغط الدم
1	135	3	89
2	120	4	90
3	100	3	83
4	105	2	77
5	130	4	92
6	125	5	98
7	125	2	82
8	105	3	85
9	120	5	96
10	90	4	95
11	120	2	80
12	95	3	79
13	120	3	86
14	150	4	97
15	160	3	92
16	125	3	88

حيث نريد تقدير معاملات الانحدار لهذا النموذج، أي إيجاد نموذج الانحدار المقدر عند مستوى معنوية 0.05.

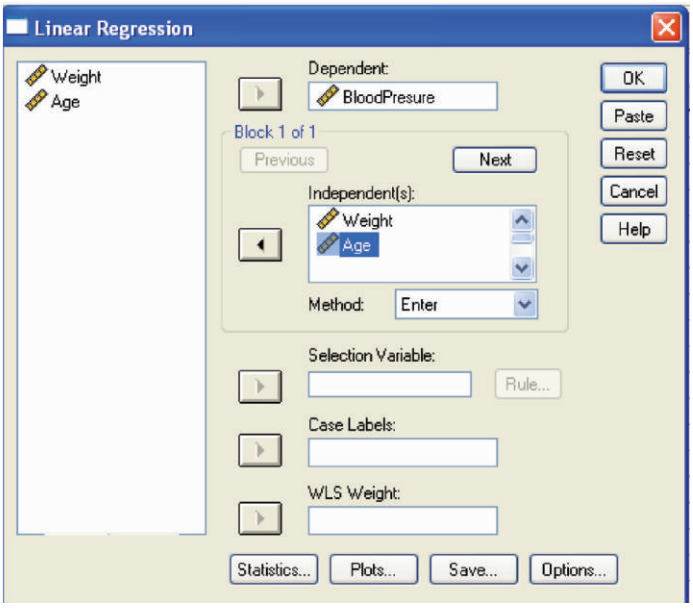
الحل:

نقوم بإدخال البيانات في ثلاثة أعمدة أحدها الوزن Weight، العمر Age ، وضغط الدم Blood Pressure كما يلي:

	Weight	Age	BloodPressure
1	135.00	3.00	89.00
2	120.00	4.00	90.00
3	100.00	3.00	83.00
4	105.00	2.00	77.00
5	130.00	4.00	92.00
6	125.00	5.00	98.00
7	125.00	2.00	82.00
8	105.00	3.00	85.00
9	120.00	5.00	96.00
10	90.00	4.00	95.00
11	120.00	2.00	80.00
12	95.00	3.00	79.00
13	120.00	3.00	86.00
14	150.00	4.00	97.00
15	160.00	3.00	92.00
16	125.00	3.00	88.00

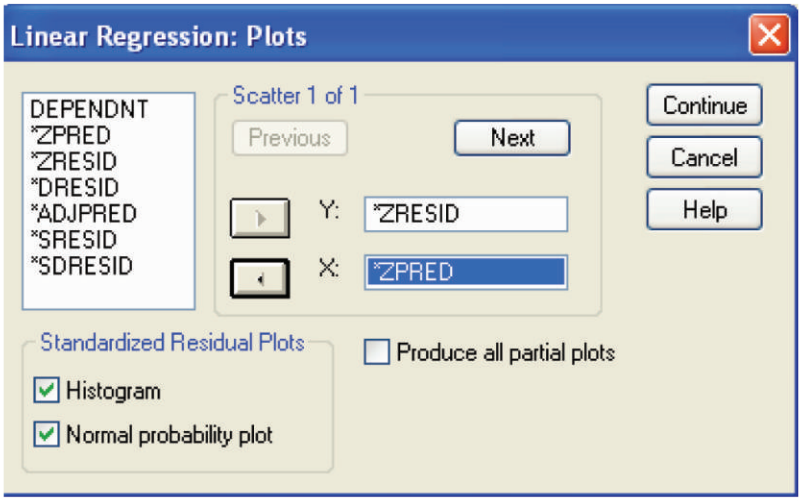
شكل (9-16)

نقوم بتفعيل نافذة الانحدار شكل (9-17) ونختار المتغيرات المستقلة هي Weight، Age، أما المتغير التابع فهو Blood Pressure كما يلي:



شكل (9-17)

بعـد ذلك نقـوم باختيار الأمر الفرعي Statistics و نتأكد من اختيـار كل من Estimates، Confidence intervals، Covariance matrix، Model fit، Casewise diagnostics كمـا في الانحـدار الخطي البسيط، ثم نعـود إلى مربـع الحوار السابق بالنقر علـى Continue ونذهب إلى الأمر الفرعـي Plots ونتأكد من اختيار Histogram وNormal probability plot وأيضًـا نضـع *ZRESID في المستطيـل Y ، ونضـع *ZPRED في المستطيل X كما يلي:



شكل (9-18)

ثم نعود إلى مربع الحوار السابق بالنقر على Continue ثم ننقر OK فتظهر لدينا النتائج الآتية: يظهر لدينا جدول يبـين الطريقة المستخدمة في الانحدار والمتغير التابع وهوBlood Pressure والمتغيرات المستقلة الداخلة في نموذج الانحدار الخطي وفي حالتنا يوجد متغيران وهما Age، Weight كما يلي:

جدول (9-12)

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Age, Weight[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: BloodPresure

بعض المقاييس التي تم حسابها للنموذج:

جدول (9-13)

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.939[a]	.881	.863	2.47917

a. Predictors: (Constant), Age, Weight

b. Dependent Variable: BloodPresure

لدينا المقياس R وكما نلاحظ فقيمته عالية وأيضا لدينا المقياس R Square الذي قيمته تقريبًا .88 وقيمته المعدلة 0.863، أي إن المتغيرين المستقلين في هذا النموذج لهما تأثير بمقدار 86,3% في المتغير التابع، كما أنه لدينا الخطأ المعياري للتقدير في النموذج وهو 2.48 تقريبًا.

جدول تحليل التباين:

جدول (9-14)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	591.036	2	295.518	48.081	.000[a]
	Residual	79.902	13	6.146		
	Total	670.938	15			

a. Predictors: (Constant), Age, Weight

b. Dependent Variable: BloodPresure

وهو يظهر لنا قيم مجاميع المربعات ودرجات الحرية الخاصة بكل مجموع مربعات، بالإضافة إلى متوسطي مربعات الخطأ والبواقي، كما نلاحظ أنه عند إجراء اختبار المعنوية عند مستوى 0,05 نجد أن نموذج الانحدار المتعدد معنوي من آخر قيمة في الجدول.

جدول المعالم المقدرة:

جدول (9-15)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	53.450	4.532		11.794	.000	43.660	63.241
	Weight	.126	.034	.352	3.657	.003	.051	.200
	Age	5.888	.680	.833	8.656	.000	4.418	7.357

a. Dependent Variable: BloodPresure

يختلف هذا الجدول عن جدول المعالم المقدرة الخاص بالانحدار الخطي البسيط بوجود أكثر من معلمتين فهنا لدينا ثلاثة معالم لنموذج الانحدار المتعدد وتكون معادلة الانحدار المقدرة:

Blood Pressure = 53.45 +0 .126Weight + 5.888Age

حيث:

$$\hat{\beta}_0 = 53.45$$

$$\hat{\beta}_1 = 0.126$$

$$\hat{\beta}_2 = 5.888$$

يمكننا أيضًا القيام باختبارات المعنوية لهذه المعالم كل على حده بمقارنة آخر قيمة في الجدول بمستوى المعنوية فنجد أن جميع المعالم معنوية.

جدول الارتباطات والتغايرات:

Coefficient Correlations **(a)**

جدول (9-16)

Model			Age	Weight
1	Correlations	Age	1.000	-.107
		Weight	107.-	1.000
	Covariances	Age	463.	-.002
		Weight	002.-	001.

a Dependent Variable: BloodPressure

يظهر لنا هذا الجدول قيم الارتباطات والتغايرات بين المتغيرات Age, Weight.

جدول البواقي:

Residuals Statistics **(a)**

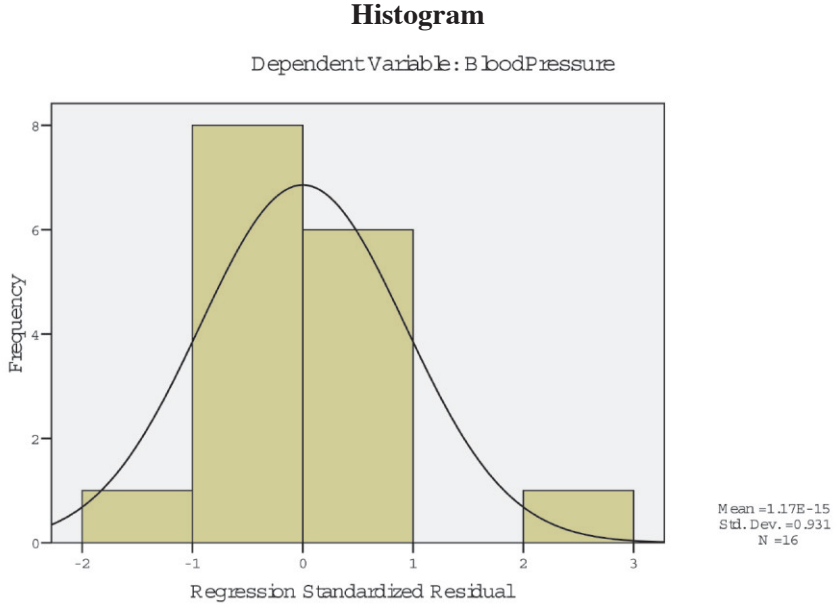
جدول (9-17)

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	78.4119	98.5867	88.0625	6.27713	16
Std. Predicted Value	-1.537	1.677	000.	1.000	16
Standard Error of Predicted Value	655.	1.532	1.043	261.	16
Adjusted Predicted Value	78.8120	98.8089	87.9800	6.31927	16
Residual	-4.04376	6.69644	00000.	2.30798	16
Std. Residual	-1.631	2.701	000.	931.	16
Stud. Residual	-1.808	3.208	014.	1.074	16
Deleted Residual	-4.96979	9.44826	08249.	3.08420	16
Stud. Deleted Residual	-2.008	6.756	223.	1.871	16
Mahal. Distance	109.	4.793	1.875	1.358	16
Cook's Distance	000.	1.410	123.	349.	16
Centered Leverage Value	007.	320.	125.	091.	16

a Dependent Variable: BloodPressure

الأشكال البيانية:

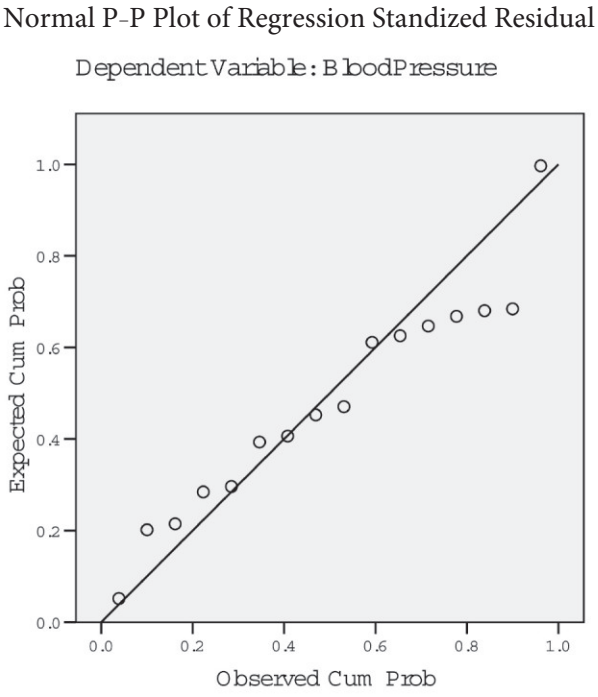
رسم المدرج التكراري:



شكل (9-19)

نستطيع منه القول إن البيانات تتوزع وفق التوزيع الطبيعي.

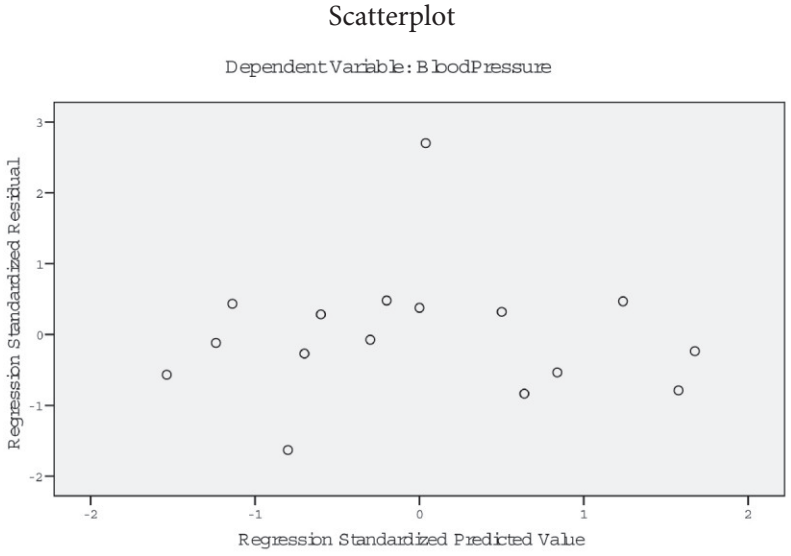
رسم منحنى الانحدار:



شكل (9-20)

هنـا أيضاً البيانات لها تقريبًا نمط انحـدار واضح على الرغم من وجود بعض القيم الشاذة عن خط الانحدار شذوذًا كبيرًا.

رسم انتشار البواقي:



شكل (9-21)

مـن هـذا الشكل نستطيع أن نقول إنـه لا يوجد نمط معين لانتشار البواقي ضد القيـم المتوقعة، أي إن الانتشار عشوائي، وهذا يتفق مع شرط الخطية المطلوب في الانحدار سواء كان بسيطًا أم متعددًا.

(4-9) اختيار المتغيرات المستقلة الداخلة في النموذج الخطي:

المعلومـات النظريـة التي ستستخدم لاختيـار المتغيرات المستقلة قد يمكن تعيينهـا بسهولة في بعض الميادين، وغالبًا ما يمكن القيـام بتجارب تحت سيطرة المجرب تزوده ببيانات يمكـن على أساسها تقدير معـالم الانحـدار واختبار الشكل النظري لدالة الانحـدار. لكن في العديد من الميادين الأخرى مثل العلوم الاجتماعيـة، السلوكية، الصحية، والإدارية، يندر نسبيًا وجود نمـاذج قابلة للاستخدام، ولمزيد من تعقيد الأمـور قد تنطوي النماذج النظرية المتوافرة على متغيرات مستقلة غير قابلة للقياس مباشرة مثل الدخول المستقبليـة للأسـرة في السنوات العشر القادمة، وتحت هذه الظروف يضطـر الباحثون إلى توقع متغيرات مستقلة يتصورون أنها يمكن أن تكون على صلة بالمتغير التابع المراد دراسته، ومن الواضح أن هذه المتغيرات المستقلة قد تكون مجموعة كبيرة وليس بالضرورة أن يكون لها تأثير في المتغير المراد دراسته فيلجأ الباحثون إلى عدة طرق يمكن أن يقوموا من خلالها بتحديد المتغيرات التي لها تأثير في المتغير المدروس حتى يتمكنوا مـن التوصل إلى نمـوذج انحدار نهائي يحتوي علـى أفضل المتغيرات التي لها تأثـير في النموذج، وسنقوم بدراسة إحدى هذه الطرق.

- الانحدار باستخدام الانحدار التدريجي.

تقوم طريقة الانحدار التدريجي التي تسمى أيضًا انحدار الخطوة فخطوة إلى الأمام، بالإضافة إلى تقدير معالم نموذج الانحدار المقدر باختيار أفضل نماذج انحدار يمكن التوصل إليها من متغيرات مستقلة عـدة مع متغـير تابع، حيث إنه ليـس بالضرورة أن تكون لجميـع المتغيرات المستقلة التـي اختارها الباحث التأثـير نفسـه في المتغير التابع، وهذا يـؤدي إلى أنه يمكـن ترشيح عدة نماذج انحـدار بمتغيرات مستقلة مختلفة وللمتغير التابع نفسه على أنها تكون نماذج جيدة.

يقـوم روتـين الانحـدار التدريجـي أولا بتوفيق نمـوذج انحدار خطـي بسيط لكل مـن المتغيرات X المرشحة وعددها p-1، حيث نقوم في كل نموذج انحدار من النماذج التي حصلنا عليها باختبار ما إذا كان الميل مساويًا للصفر أم لا، حيث نستخدم الإحصاءة الآتية:

$$F_k^* = \frac{MSR\ (X_k)}{MSE\ (X_k)}\ ,\ k = 1,2,...,\ P\text{-}1$$

حيـث إن $MSR\ (X_k)$ يقيس الانخفاض في التغير الكلي للمتغير التابع المصاحب لاستخدام المتغير Xk، والمتغـير Xi الـذي له أعلى قيمة في الإحصـاءة يكون هو المتغير المرشح لأول إضافـة في النموذج، حيث إذا تجاوزت قيمة F* مستوى محددًا سلفا يضاف المتغير X، وفيما عدا ذلك ينتهي البرنامج معتبرا أنه لا يوجد

أي متغير X مفيد بما يكفي لدخول نموذج الانحدار. لفرض أن المتغير X7 هو المتغير الذي تم اختياره في أول محاولة. سيقوم البرنامج بعدها بتوفيق جميع نماذج الانحدار المتضمنة لمتغيرين مستقلين أحدهما X7، ولكل نموذج نقوم باختبار ما إذا كان الميل مساويًا للصفر مستخدمين الإحصاءة:

$$F_k^* = \frac{MSR\ (X_k\ /\ X_7)}{MSR\ (X_k\ /\ X_7)}$$

فإذا تجاوزت قيمة F^* مستوى $F_{(1-a,1,n-p)}$ الذي نحدده سلفا يضاف المتغير الثاني، وفيما عدا ذلك ينتهي البرنامج باختيار المتغير X_7 فقط.

لنفرض أن X3 تمت إضافته في الخطوة السابقة عندئذ سيقوم روتين الانحدار التدريجي باختبار ما إذا كان ينبغي حذف أي من المتغيرات المستقلة الموجودة في النموذج، وبناءً عليه إما أن تحذف متغيرات أو يظل النموذج بالمتغيرات التي قد اختيرت سابقًا، ثم نرجع مرة أخرى إلى إضافة متغيرات مستقلة في النموذج ثم إلى حذف المتغيرات كما في الخطوات السابقة إلى أن نصل إلى مرحلة لا يمكننا فيها إضافة أو حذف متغير مستقل من النموذج، وعندها تنتهي عملية البحث عن النموذج، ويكون النموذج الموجود لدينا هو نموذج الانحدار المقدر بأفضل المتغيرات المستقلة.

مثال (9-4)

نريد معرفة المتغيرات التي لها تأثير في المصروفات المعيشية للأسرة قمنا باختيار 30 أسرة وسجلنا قيم المتغيرات الآتية لها:

جدول (9-18)

X4i	X3i	X2i	X1i	Yi	i
5	7	2	6	6,5	1
6	13	3	16	11,2	2
8	19	4	15	11,2	3
8	13	4	14	10,5	4
9	9,6	2	10	9,3	5
5	8	3	10	7,2	6
7	15	4	9	13,4	7
7	5	3	6	5	8
6	12	4	7	11,6	9
7	14	4	12	11,2	10
5	7	3	4	6,2	11
11	19	6	0	12	12

9	16	6	14	11,3	13
8	11	4	5	9,2	14
9	6	4	2	5,5	15
7	7,6	3	4	6	16
8	25	1	16	12,5	17
6	10	2	4	9,8	18
5	6,5	3	5	6,1	19
11	15,1	6	15	14,3	20
10	18	6	15	45,5	21
6	11	1	2	10,8	22
5	5,6	1	4	4,5	23
8	8,5	2	3	6,7	24
5	6,3	1	2	4,5	25
5	14	3	5	9,8	26
7	4,6	1	5	4	27
7	7,5	1	3	5,5	28
7	11,2	2	8	10	29
5	9,5	0	7	8,5	30

حيث:

Yi: المصروفات المعيشية (ألف جنيه مصري).

X1i: مستوى تعليم رب الأسرة (عدد سنوات دراسته).

X2i: عدد الأطفال.

X3i: دخل الأسرة (ألف جنيه مصري).

X4i: عدد أفراد الأسرة.

نريد إيجاد نموذج الانحدار الخطي الخاص بالمصروفات المعيشية للأسر عند مستوى معنوية 0.05.

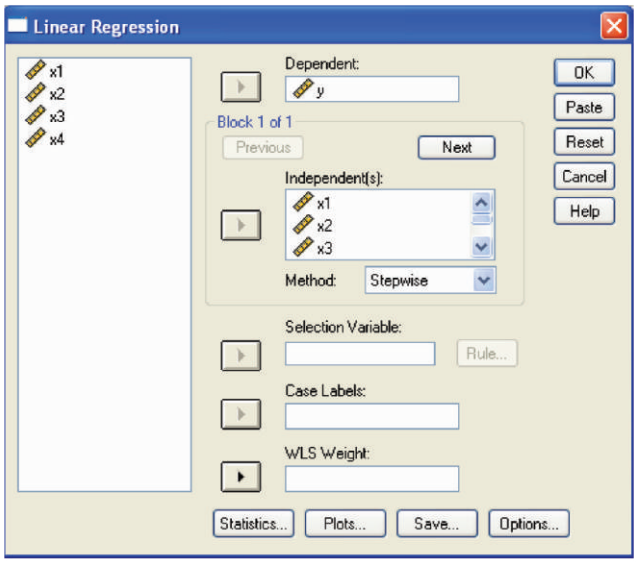
الحل:

ندخل البيانات في SPSS البرنامج:

	y	x1	x2	x3	x4
1	6.50	6.00	2.00	7.00	5.00
2	11.20	16.00	3.00	13.00	6.00
3	11.20	15.00	4.00	19.00	8.00
4	10.50	14.00	4.00	13.00	8.00
5	9.30	10.00	2.00	9.60	9.00
6	7.20	10.00	3.00	8.00	5.00
7	13.40	9.00	4.00	15.00	7.00
8	5.00	6.00	3.00	5.00	7.00
9	11.60	7.00	4.00	12.00	6.00
10	11.20	12.00	4.00	14.00	7.00
11	6.20	4.00	3.00	7.00	5.00
12	12.00	.00	6.00	19.00	11.00
13	11.30	14.00	6.00	16.00	9.00
14	9.20	5.00	4.00	11.00	8.00

شكل (9-22)

نقـوم بتفعيـل نافذة الانحـدار ونختار المتغـير التابع y والمتغـيرات المستقلة هــي x1 x2 x3 x4 ونقوم باختيار الطريقة Stepwise كما يلي:



شكل (9-23)

بالنسبــة للخيــارات Statistics، Plots، Save نقــوم بما قمنا به في الانحدار المتعــدد ثم ننقر ok فينتج لدينا:

جدول المتغيرات الداخلة في النموذج:

جدول (9-19)

Model	Variables Entered	Variables Removed	Method
1	x3	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050، Probability-of-F-to-remove >= .100).

Variables Entered/Removed (a)

a Dependent Variable :y

يتضح لنا من هذا الجدول أنه تم اختيار المتغير X3 فقط في النموذج، أي إن دخل الأسرة هو المتغير الوحيد المؤثر في المصروفات المعيشية من بين المغيرات التي قمنا باختيارها.

بعض مقاييس النموذج:

جدول (9-20)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.570(a)	.325	.301	6.11488

a Predictors: (Constant)، x3

b Dependent Variable: y

نلاحظ من قيمة R Square المعدلة أن المتغير X3 يفسر تقريباً 30% من Y.

جدول تحليل التباين:

ANOVA (b)

جدول (9-21)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	505.129	1	505.129	13.509	.001(a)
	Residual	1046.970	28	37.392		
	Total	1552.099	29			

a Predictors: (Constant)، x3

b Dependent Variable: y

من هذا الجدول نستطيع اختبار معنوية الانحدار، وذلك باستخدام الاختبار الآتي:

الفروض الإحصائية:

H0: الانحدار غير معنوي (غير دال).

H1: الانحدار معنوي (دال).

جدول (9-22)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	.553	2.801		.197	.845	-5.184	6.290
	x3	.845	.230	.570	3.675	.001	.374	1.317

خطوات الاختبار:

نقارن P-value = .001 بمستوى المعنوية 0.05، نلاحظ أن القيمة P-value أقل من مستوى المعنوية.

القرار الإحصائي:

نقوم برفض الفرض العدمي، وهذا معناه أن الانحدار معنوي.

جدول المعاملات:

Coefficients **(a)**

a Dependent Variable: y

من جدول المعاملات نجد أن معادلة الانحدار المقدرة هي:

$$Y=0.553 + 0.845X_3$$

حيث تكون قيمة $\hat{\beta}_0 = 0.553$, $\hat{\beta}_3 = 0.845$ فلنختبر معنوية β_3 عند مستوى معنوية 0.05:

الفروض الإحصائية:

فرض العدم هو: $H_0 : \beta_3 = 0$

ضد الفرض البديل: $H_A : \beta_3 \neq 0$

خطوات الاختبار:

نقارن *P-value =.001* بمستوى المعنوية *0.05* نلاحظ أن القيمة *P-value* أقل من مستوى المعنوية.

القرار الإحصائي:

نقوم برفض الفرض العدمي وهذا معناه أن β_3 معنوي.

جدول المتغيرات المستبعدة من النموذج:

Excluded Variables **(b)**

جدول (9-23)

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	x1	.221(a)	1.120	.273	.211	.614
	x2	.312(a)	1.859	.074	.337	.786
	x4	.218(a)	1.211	.237	.227	.729

a Predictors in the Model: (Constant)، 3x

b Dependent Variable: y

جدول البواقي:

Residuals Statistics **(a)**

جدول (9-24)

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	4.4418	21.6882	9.9933	4.17352	30
Std. Predicted Value	-1.330	2.802	.000	1.000	30
Standard Error of Predicted Value	1.116	3.372	1.514	.457	30
Adjusted Predicted Value	4.4878	25.7033	10.0856	4.57232	30
Residual	-9.18822	29.72967	00000.	6.00853	30
Std. Residual	-1.503	4.862	000.	.983	30
Stud. Residual	-1.801	5.123	007.-	1.048	30
Deleted Residual	-13.20330	33.01107	09227.-	6.84940	30
Stud. Deleted Residual	-1.881	20.104	.491	3.727	30
Mahal. Distance	.000	7.852	.967	1.485	30
Cook's Distance	.000	1.448	.076	.290	30
Centered Leverage Value	.000	.271	.033	.051	30

a Dependent Variable: y

الفصل العاشر

الاختبارات اللامعلمية (اللابارامترية)

Non-parametric Tests

(10-1) مقدمة:

في معظم الأساليب التي تكلمنا عنها في الاختبارات المعلمية (البارامترية) نجد أنها مبنية على الفرضية التي تقول إن العينة أو العينات العشوائية التي تم اختيارها للدراسة من مجتمع طبيعي، وغالبًا ما تكون هذه الأساليب غير دقيقة إلى حد ما عندما يكون مجتمع العينة غير طبيعي، وحيث إن بعض المجتمعات لا تفي بالشروط المطلوبة لتطبيق تلك الأساليب، دعت الحاجة للبحث عن أساليب أخرى لا يتطلب تطبيقها مثل ذلك الشرط. هذه الأساليب يطلق عليها تسمية الأساليب اللامعلمية (اللابارامترية)؛ لأنه وكما لوحظ في الفصل الأول كان اهتمامنا يركز على معلمة (بارامتر) أو أكثر من معلمات (بارامترات) المجتمع الإحصائي (المتوسط، التباين، النسبة،...إلخ) علاوة على ذلك، وكما أشرنا في الفصل الأول لكي نصل إلى استنتاج إحصائي تجب معرفة صيغة التوزيع الاحتمالي للمجتمع التي تم اختيار العينة منها.

وهناك نوعان من الأساليب الإحصائية تتم معاملتها على أنها أساليب لا معلمية وهما:

أساليب لا معلمية بما تعنيه الكلمة، وهي أساليب تختبر الفرضيات التي لا تتضمن أي نص يتعلق بمعلمات المجتمع الإحصائي، أما الأساليب الأخرى فهي أساليب التوزيعات الحرة، وهي الأساليب التي لا تضع أي افتراضات على مجتمع العينة، وبصرف النظر عن التمييز بين هذين الأسلوبين فإن كلاهما ستتم معاملتها على أنهما أساليب لا معلمية، هذه الأساليب يتم تطبيقها على سبيل المثال لا الحصر في الحالات الآتية:

1 - إذا كانت الفرضية المطلوب اختبارها لا تتضمن معلمة المجتمع.

2 - البيانات مقاسة بمقياس أضعف من المقاييس المطلوبة لتطبيق الأساليب المعلمية مثل (المقياس الاسمي، المقياس الترتيبي، مقياس الفترة، المقياس النسبي).

3 - إن لم تتوافر الشروط المطلوبة لتطبيق الأساليب المعلمية.

بعض مزايا الاختبارات اللامعلمية (اللابارامترية) ما يلي:

- مهما كان شكل التوزيع المأخوذ منه العينة فإن الاختبار اللامعلمي الذي له مستوى معنوية (دلالة) معين يكون له هذا المستوى فعلاً بشرط أن تكون العينة قد اختيرت عشوائياً، كما يشترط أيضاً في بعض الحالات استمرار التوزيع.
- الإحصاءات اللامعلمية هي الأسلوب الوحيد الممكن استخدامه في حالة العينات الصغيرة جداً إلا إذا كان توزيع المجتمع معروفاً تماماً.
- يمكن استخدامها أحيانًا للعينات التي تحتوي على مشاهدات من عدة مجتمعات متفاوتة.
- تصلح لتحليل البيانات التي تكون على صورة رتب دون الحاجة إلى معرفة التوزيع في المجتمع الأصلي للبيانات.
- تستخدم في حالة كون البيانات تتضمن إحدى صيغتي التفضيل مثلاً سليم أو معيب، حيث السليم تكون له إشارة موجبة والمعيب إشارة سالبة، وهنا لا تصلح الطرق التقليدية.
- يمكن تطبيقها عندما تكون البيانات مقاسة بمقياس ضعيف.
- تعتمد على افتراضات قليلة، ومن ثم فرصة تطبيقها خطأ ستكون صغيرة.
- الحسابات الضرورية للأساليب اللامعلمية عادة ما تكون سهلة ويمكن إنجازها بسرعة.
- سهولة فهمها وطريقة حسابها تجعلها مناسبة جدًا للباحثين الذين ليست لهم خلفية علمية جيدة في الرياضيات والإحصاء.

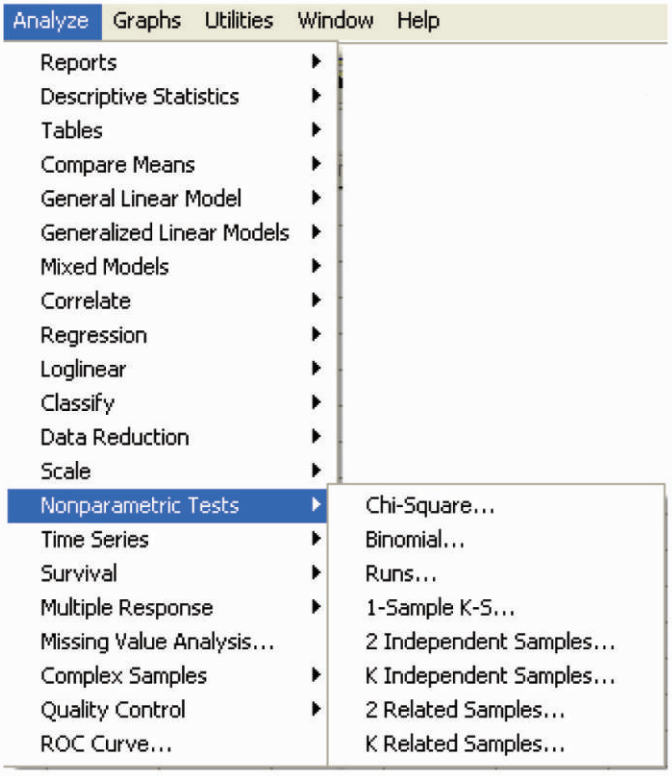
بعض عيوب الاختبارات اللامعلمية (اللابارامترية):

- نتيجة لسهولة حسابها، في بعض الأحيان يتم تطبيقها في مسائل يكون من الأفضل تطبيق أساليب معلمية عليها، ما يتسبب في ضياع المعلومات.
- في حالة العينات الكبيرة يؤدي استخدامها إلى جهد أكبر من الأساليب التقليدية.
- في حالة تحليل بيانات من توزيع طبيعي فإن استخدام الاختبارات اللامعلمية يعد فقدًا للبيانات، وتقاس درجة الفقد بكفاءة الاختبار اللامعلمية.

وتنقسم الاختبارات اللامعلمية حسب عدد العينات عند إجراء الاختبار إلى:

- حالة عينة واحدة One sample case.
- حالة عينتين Two samples case (وهنا يوجد اختلاف بين المقياس الذي يعتمد على: أ- العينتين مستقلتين ب- العينتين غير مستقلتين).
- حالة عدد العينات K التي ربما تفترض استقلالا للعينات أو ارتباطها.

وتظهر أوامر الاختبارات اللامعلمية بالنقر على Analyze في شريط الأوامر فيظهر الشكل الآتي:



شكل (10-1)

ومن الشكل السابق نستطيع اختيار نوع الاختبار الملائم لنوع المشكلة لدينا، وسوف نتعرف فيما يلي على تلك الأنواع من الاختبارات وكيفية معالجتها للمشكلات المختلفة:

(10-2) اختبار مربع كاي (Chi-Square test):

إن من أشهر وأقدم اختبارات جودة المطابقة هو اختبار مربع كاي لجودة المطابقة، الذي اقترحه بيرسون (1900م)، ويستخدم هذا الاختبار لتحديد ما إذا كانت التكرارات المشاهدة في جدول توزيع تكراري بسيط (لظاهرة واحدة) تتبع توزيعًا احتماليًا معينًا مثل:

- اختبار أن عدد الحوادث التي تقع في ميدان معين لها توزيع بواسون (وهو أحد التوزيعات الاحتمالية المتقطعة شائعة الاستخدام في كثير من التطبيقات)،
- اختبار أن درجات الطلبة في أحد الامتحانات لها توزيع طبيعي، اختبار أن متوسط أطوال القطع التي تنتجها إحدى الآلات لها توزيع طبيعي.

في كل هذه الحالات وأمثالها يقوم الاختبار أساسًا على مقارنة التكرارات المشاهدة بالتكرارات المتوقعة التي تحسب باستخدام ذلك التوزيع الاحتمالي المعين (المطلوب اختبار ما إذا كانت البيانات تتبعه أم لا)، ومن الفروق بين التكرارات المشاهدة والمتوقعة تحسب x^2 بتطبيق الصيغة الآتية:

$$x^2 = \sum \frac{(observed - expected)^2}{expected}$$

Observed: القيم المشاهدة.

Expected: القيم المتوقعة.

وإذا كان الجدول التكراري به m من الخلايا تكون x^2 المحسوبة لها توزيع x^2 بدرجات حرية يساوي عدد الخلايا مطروحًا منها 1 أي $m-1$ ومستوى معنوية مقداره α وعليه إذا كان

$$x^2 = calcuate > x^2 = (\alpha, m-1)$$

يرفض فرض العدم عند مستوى معنوية α، حيث فرض العدم هنا هو فرض أن التكرارات المشاهدة بالجدول تتبع التوزيع الاحتمالي المعين.

هذا الاختبار يشبه اختبارات كاي للاستقلالية والتجانس من حيث كون إحصاءة الاختبار تنتج من مقارنة التكرارات المشاهدة ولكن أوجه تطبيقها مختلف تمامًا.

شروط تطبيق الاختبار:

✓ تتضمن البيانات عينة عشوائية بها n من المفردات المستقلة عن بعضها بعضًا تم اختيارها من مجتمع x، ويمكن وضع هذه البيانات في جدول توافقي كما يلي:

جدول (10-1)

الصنف	1 2 3 .... i ..... r	المجموع
التكرار المشاهد	O_1 O_2 O_3 ...O_i ...O_r	n

حيث O_1 تمثل عدد المفردات التي تقع في الصنف i، حيث $i = 1,2,3,....,r$ مع ملاحظة أنه من الممكن أن يكون التصنيف نوعيًا أو كميًا، فمثلا من الممكن تصنيف مجموعة من الأشخاص حسب الجنس (ذكور، إناث) وكذلك من الممكن التصنيف بالعمر...إلخ.

✓ وحدة القياس على الأقل اسمية (nominal).

الفروض الإحصائية:

إذا رمزنـا لدالة التوزيـع غير المعروفة لمجتمع X بالرمز F(x) ولدالة التوزيـع الفرضية بالرمز $F_0(x)$، وهـي محددة بالكامل عدا أنَّه من الممكن أن تكون المعلمة غـير معروفة، ويجب تقديرها من بيانات العينة، فإنه يمكن صياغة الفرضيات الإحصائية كما يلي:

$H_0: F(x) = F_0(x)$ لجميع قيم x

$H_A: F(x) \neq F_0(x)$ على الأقل لقيمة واحدة من قيم x.

إحصاءة الاختبار:

حيث إن هناك احتمالاً بأن تقع أي مفردة يتم اختيارها من المجتمع بأي صنف من التصنيفات المختلفة، ومـن ثم يمكن الرمز لهذه الاحتمالات بالرمـز $P_1, P_2, \ldots, P_r$ على التوالي، وذلك لأنه يوجد r صنف، وعليه في حالة H_0 يمكن حساب التكرار المتوقع بكل صنف كما يلي:

$$T = \sum_{i=1}^{r} \frac{(O_i - E_i)^2}{E_i} \quad \& \quad Ei = np_i$$

حيث:

O_i: القيم المشاهدة،

E_i: القيم المتوقعة،

P_i: قيمة الاحتمال،

n: مجموع المشاهدات.

القرار الإحصائي:

نرفض H_o إذا كان $T > x^2_{\alpha, r-1}$

مثال (10 - 1)

أخـذت عينـة عشوائية مكونة من 400 أسرة مـن الأسر التي لكل منها ثلاثة أطفـال أو أقل، فوجد أن التوزيع التكراري لتلك الأسر حسب عدد الأطفال الذكور كالآتي:

جدول (10-2)

عدد الأطفال الذكور X	0	1	2	3
عدد الأسر	48	147	153	52

اختبر فرض أن عدد الأطفال الذكور بكل أسرة لديها ثلاثة أطفال له توزيع ذو الحدين بنجاح 0,05 = θ وعدد المحاولات $N = 3$، وذلك عند مستوى معنوية مقدارها 0,05.

الحل:

أولا طريقة إدخال البيانات في البرنامج:

نقـوم بتصميم متغيرين أحدهما للمتغير الذي يصف عدد الأطفـال في الأسرة، والآخر للمتغير الذي يصف التكرار المشاهد.

Number	Observe
00.	48.0000
1.00	147.000
2.00	153.000
3.00	152.000

شكل (10-2)

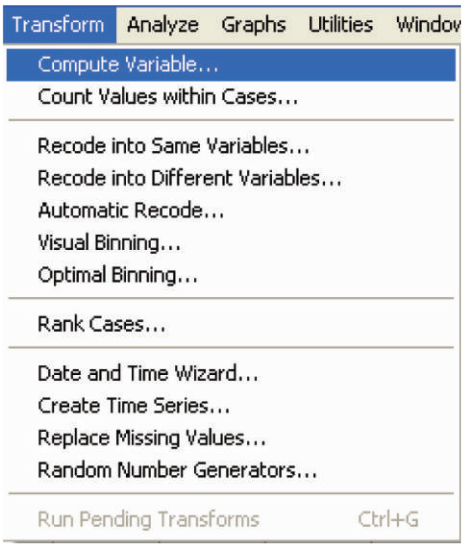
ثانياً: نقوم بحساب التكرار المتوقع:

ونلاحـظ أننـا نريد حساب التكرار المتوقع مـن توزيع ذي الحدين، ولذلك يتـم أولا حساب الاحتمالات التي نرمز لها بالرمز pi، ومن ثم حساب التكرار المتوقع كالآتي:

$$Ei = npi$$

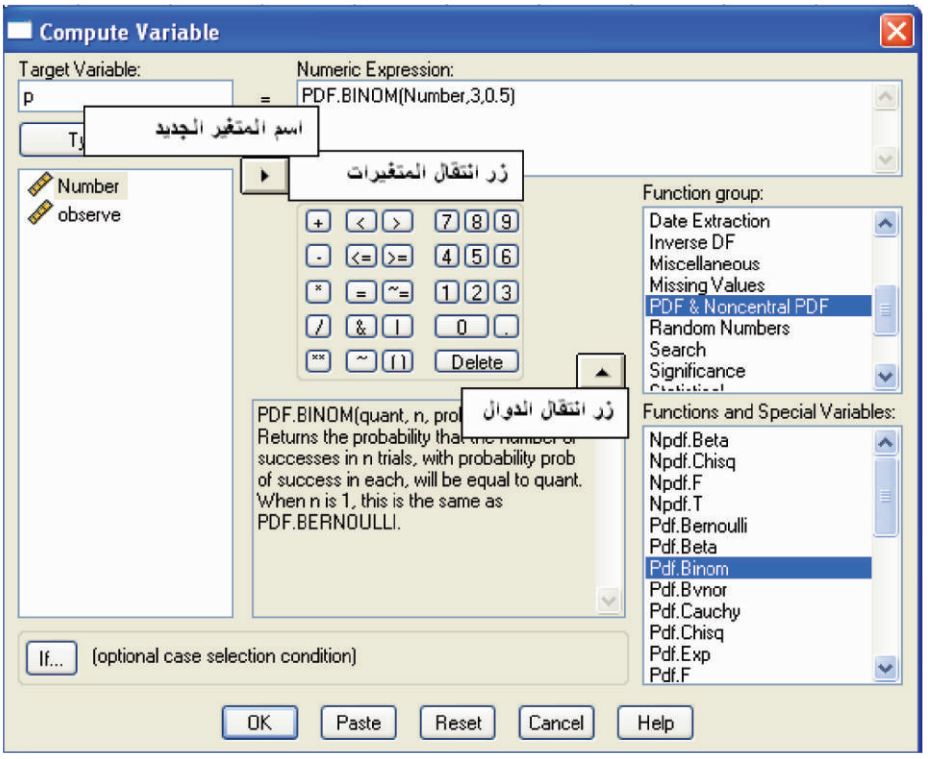
$$= 400\ pi$$

فمن قائمة Transform نقوم باختيار Compute Variable.... ومن ثم يظهر الشكل الآتي:



شكل (10-3)

ثم نقوم باختيار PDF and Noncentral PDF من عمود Function group وبعدها تظهر قائمة بدوال فرعية في عمود أسفل العمود السابق ونقوم باختيار pdf Binom، وهذا يعني دالة الاحتمال لذي الحدين أي pi فتضغط علي الدالة مرتين متتاليتين أو النقر على زر الانتقال يظهر الشكل الآتي:



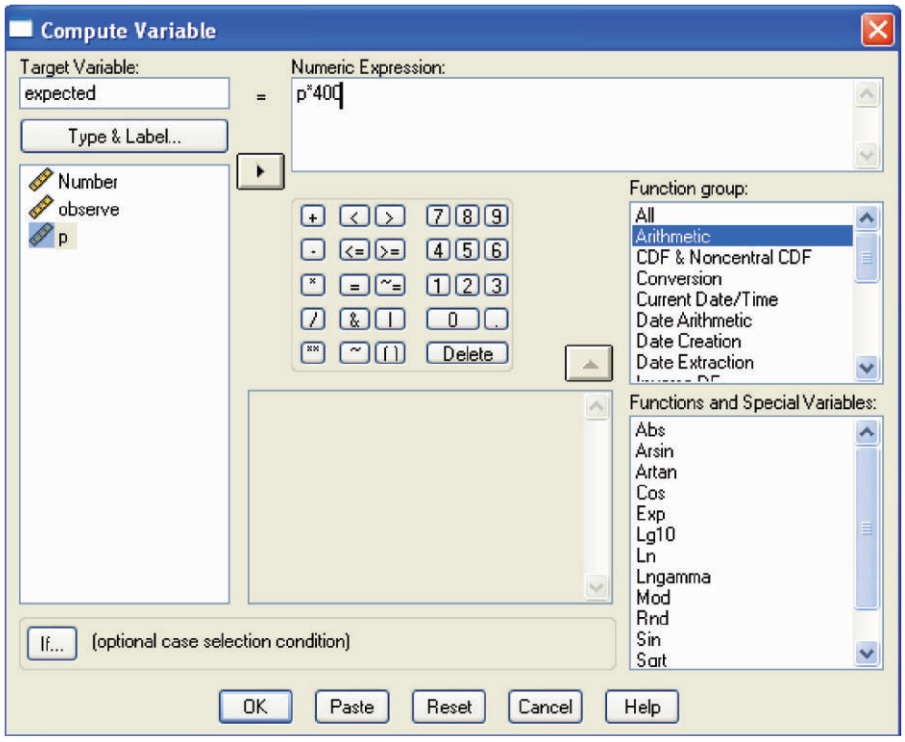
شكل (10-4)

ففي علامة الاستفهام الأولى نقوم بإدخال المتغير Number من زر إدخال المتغيرات، بعد ذلك نقوم بإدخال عدد مرات المحاولة، وهي في التمرين 3، وأخيرا في علامة الاستفهام الثالثة والأخيرة نقوم بإدخال احتمال الحدوث أو احتمال النجاح وهو في التمرين 5, 0. ونقوم تحت عنوان Target variable بإدخال اسم المتغير الجديد، وهنا هو p الذي يرمز لدالة الاحتمال، ومن ثم النقر على ok فتظهر شاشة Data View كالآتي:

جدول (10-3)

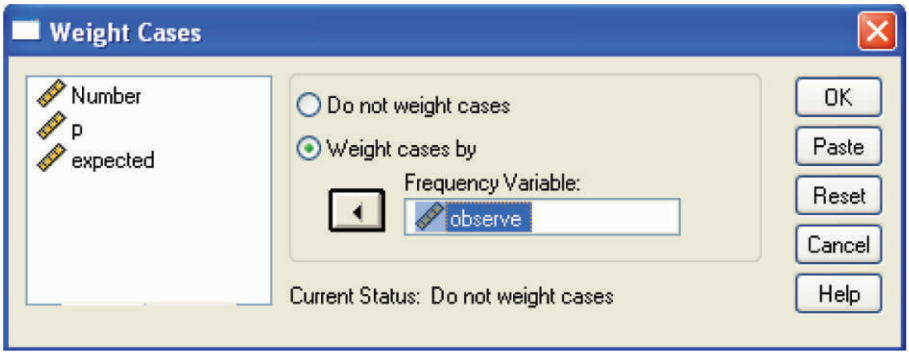
Number	Observe	P
00.	48.0000	.13
1.00	147.000	.38
2.00	153.000	.38
3.00	152.000	.13

ومرة أخرى نقوم بحساب $np_1 = 400\, p_i$ بالطريقة السابقة ونسمي المتغير الجديد expected كما في الشكل الآتي:



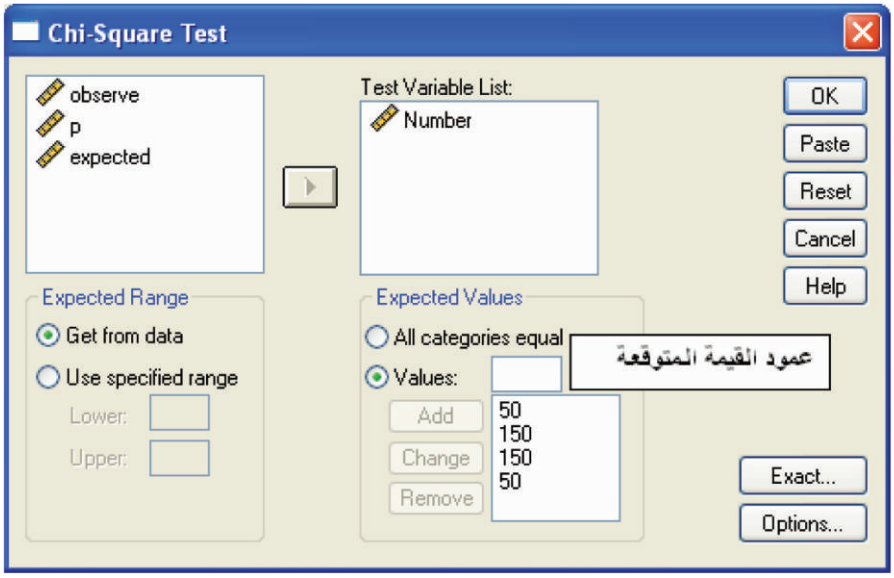
شكل (10-5)

بعـد ذلك ننتقـل إلى الخطوة الأخيرة، حيث نقوم بتعريف أن العمود observed هو عمود تكرار العمود Number، وذلـك مـن خلال اختيار قائمة Data من شريط القوائم، ومن ثم اختيار Weight cases ثم نقوم بقياس العمود الثاني كما في الشكل الآتي:



شكل (10-6)

وبعد ذلك نقوم بفتح Analysis ثم Nonparametric ثم Chi square فيظهر الشكل الآتي:



شكل (10-7)

فنقـوم بإدخال المتغير Number في عمـود test variable list ونقوم بإدخال القيم المتوقعة المحسوبة في العمود الأخير في صفحة Data viewer على التوالي كما في الشكل الآتي، حيث نقوم بتنشيط الخيار values مـن عمـود Expected values ، وبعد ذلك نقوم بإدخال القيمة الأولى ثم ننقـر على add ثم الثانية وهكذا، وأخيرًا نقوم بالنقر علي ok في الشكل (10-7) وتخرج النتائج الآتية:

جدول (10-4)

Number

	Observed N	Expected N	Residual
.00	48	50.0	-2.0
1.00	147	150.0	-3.0
2.00	153	150.0	3.0
3.00	52	50.0	2.0
Total	400		

جدول (10-5)

Test Statistics

	Number
Chi-Square[a]	.280
df	3
Asymp. Sig.	.964

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 50.0.

ويوضح الجدول (10-3) لمربع كاي القيمة المتوقعة لكل خلية.

القرار الإحصائي:

يوضح الجدول (11-4) قيمة الاختبار = 0.28 أيضا p-value =.964، وهي أكبر من مستوى المعنوية 0.05 إذا نقبل H_0 بأن عدد الأطفال الذكور بكل أسرة لديها ثلاثة أطفال له توزيع ذو الحدين بنجاح $\theta = 0.05$، وذلك عند مستوى معنوية مقدارها 0.05.

مثال (10-2)

اختيرت عينة عشوائية مكونة من 500 طالب من طلبة الفرقة الثالثة بإحدى كليات التجارة، وتم تصنيفهم حسب التخصص ونتيجة الامتحان في الإحصاء، فكان التصنيف كما يلي في الجدول الآتي:

جدول (10-6)

	ناجح P	راسب F	Total
إدارة الأعمال	240	60	300
اقتصاد	120	80	200
Total	360	140	500

اختبر ما إذا كانت هناك علاقة بين التخصص ونتيجة الامتحان في الإحصاء عند مستوى معنوية 0.05.

الحل:

الفروض الإحصائية:

الفرض العدمي: لا توجد علاقة بين التخصص ونتيجة الإحصاء.

الفرض البديل: توجد علاقة بين التخصص ونتيجة الإحصاء.

أولا طريقة إدخال البيانات في البرنامج:

نقوم بإدخال رقم الصف الأول والعمود الأول ثم الصف الأول العمود الثاني ثم الصف الثاني العمود الأول ثم الصف الثاني العمود الثاني، وبعد ذلك القيم تكون في عمود مستقل، ونعرف العمود الأول x حيث يشير إلى التخصص فيأخذ التخصص إدارة الأعمال (1) والتخصص الاقتصاد (2) والعمود الثاني y يشير إلى النتيجة في التخصص فتأخذ (1) في حالة النجاح، وتأخذ (2) في حالة الرسوب انظر فصول أساسيات العرض والتحليل الإحصائي باستخدام البرنامج فيظهر الشكل الآتي:

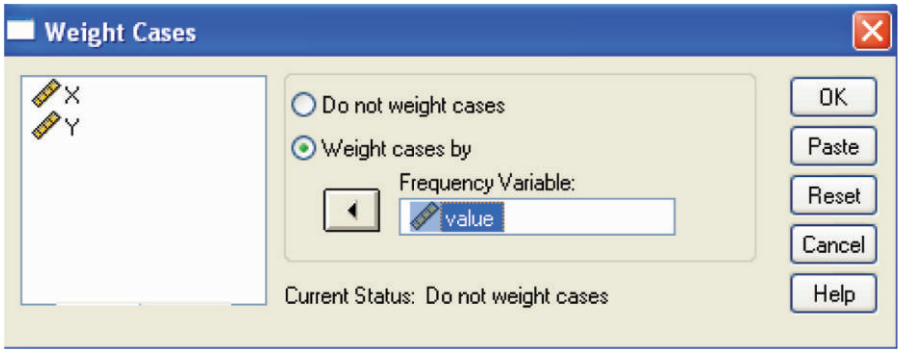
X	Y	Value
1	1	240.00
1	2	60.00
2	1	120.00
2	2	80.00

شكل (10-8)

ثم نقوم بقياس القيم في العمود الثالث بالطريقة الآتية:

Data ← Weight cases

فيظهر الشكل الآتية:

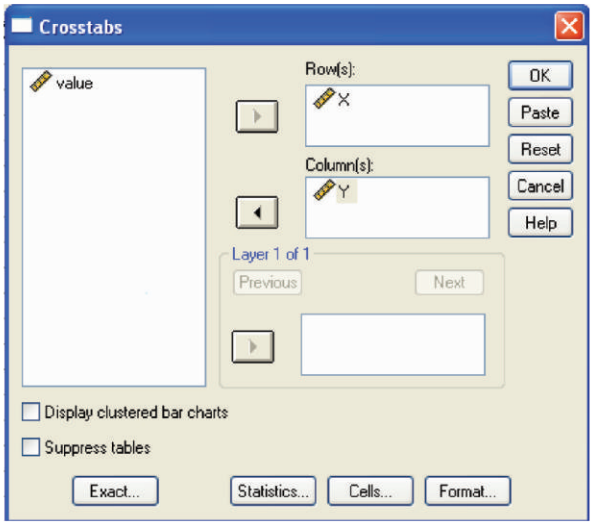


شكل (10-9)

ثم نقوم بالآتي:

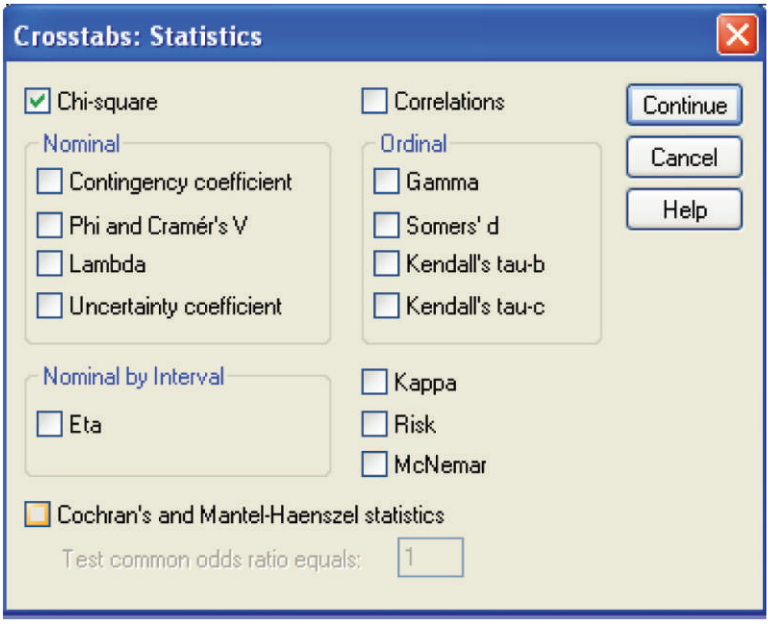
Analyze .. ← Descriptive Statistics ← Cross tabs

ثم ندخل العمود الأول في Rows والعمود الثاني في column كما يلي:



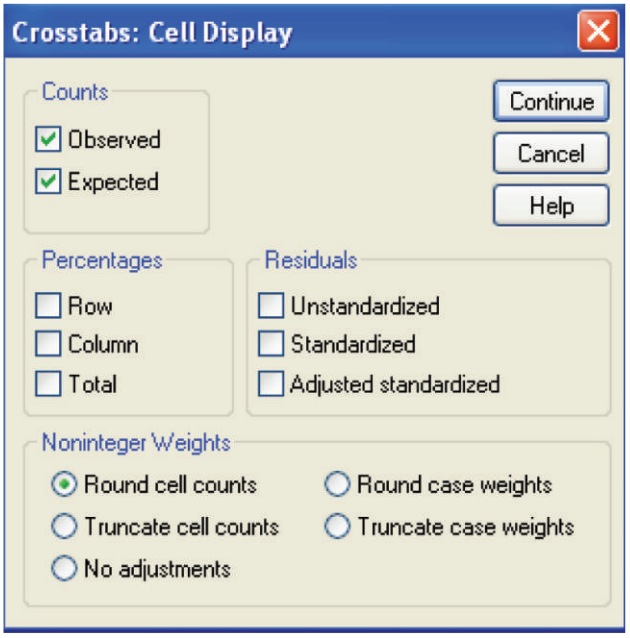
شكل (10-10)

ثـم نقـوم بالنقر على Statistics، ونقوم باختيار Chi-square كمـا في الشكل الآتي، ومن ثم النقر على continue لتنشيط:



شكل (10-11)

ثم نقوم بالنقر على Cells ثم نقوم باختيار القيم المتوقعة لكل خلية كما في الشكل الآتي:



شكل (10-12)

وبعد ذلك ننقر على Continue ثم OK في الشكل (10-11) و(10-12) فنحصل على النتائج الآتية:

جدول (10-7)

X * Y Crosstabulation

			Y		Total
			1	2	
X	1	Count	240	60	300
		Expected Count	216.0	84.0	300.0
	2	Count	120	80	200
		Expected Count	144.0	56.0	200.0
Total		Count	360	140	500
		Expected Count	360.0	140.0	500.0

جدول (10-8)

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	23.810[b]	1	.000		
Continuity Correctiðr	22.828	1	.000		
Likelihood Ratio	23.507	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	23.762	1	.000		
N of Valid Cases	500				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 56 00.

ويوضح جدول (10-7) القيمة المتوقعة لكل خلية.

القرار الإحصائي:

يوضح جدول (10-8) قيمة الاختبار = 23.810 وأيضًا .000 = p-value وهي أقل من مستوى المعنوية 0.05 إذا نرفض H0، ومن ثم توجد علاقة بين التخصص ونتيجة الإحصاء .

مثال (10-3)

طرحت مائة فرشة أسنان جديدة على مائة رجل ومائة سيدة لاستخدامها ثم إبداء آرائهم هل يفضلون استخدامها أم لا، 32 من الرجال و 26 من السيدات أجابوا بأنهم لا يفضلون استخدام الفرشاة الجديدة. هل هذا يدل على وجود فرق في التفضيل بين الرجال والسيدات عند مستوى معنوية 0.05؟ حيث كانت البيانات كما في الجدول الآتي:

جدول (10-9)

	يفضل	لا يفضل	Total
رجل	68	32	100
سيدة	74	26	100
Total	142	58	200

الحل:

الفروض الإحصائية:

فرض العدم: لا يوجد فرق في التفضيل بين الرجال والسيدات.

الرفض البديل: يوجد فرق في التفضيل بين الرجال والسيدات.

بالطريقة السابقة نفسها نحصل على النتائج الآتية:

جدول (10-10)

VAR00001 * VAR00002 Crosstabulation

			VAR00002		
			like	don't like	Total
VAR00001	man	Count	68	32	100
		Expected Count	71.0	29.0	100.0
	woman	Count	74	26	100
		Expected Count	71.0	29.0	100.0
Total		Count	142	58	200
		Expected Count	142.0	58.0	200.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.874[b]	1	.350		
Continuity Correction [a]	.607	1	.436		
Likelihood Ratio	.875	1	.349		
Fisher's Exact Test				.436	.218
Linear-by-Linear Association	.870	1	.351		
N of Valid Cases	200				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 29.00.

القرار الإحصائي:

في جدول (10-10) نلاحظ قيمة اختبار مربع كاي = 0.874، ونلاحظ أيضا قيمة p-value=0.35، وهي أكبر من 0.05 إذا نقبل H0 ونقول إنه لا يوجد فرق في التفضيل بين الرجال والسيدات في تفضيلهم للمنتج.

(10-3) اختبار الدورة (Run Test):

إن معظـم الأساليب الإحصائيــة التي تستخدم لدراسة ظاهرة معينة بمجتمـع ما، تفترض أن مفردات العينة التي يتم اختيارها من ذلك المجتمع ستكون عشوائية حتى يكون للاستنتاج الإحصائي معنى، فإذا وجد شـك في عدم صحة هـذا الافتراض يجب أن يكون لدينا أسلوب علمي للتحقق من ذلك، فمثلا عند استخدام أساليـب مراقبة الجودة نقوم برسم خرائط للتحكـم ودراسة عدد الوحدات المعيبة في الإنتاج، وللقيام بذلك عادة ما تؤخذ عينات من الإنتاج بشكل دوري ومعرفة عدد الوحدات المعيبة، ومن ثم قد يكون هذا العدد أكبر أو أصغر من العدد المسموح به، وأن الهدف من وراء ذلك هو معرفة ما إذا كان هذا العدد من الوحدات المعيبة بالإنتاج التي ظهرت بالعينات المختارة يحدث بشكل عشوائي؛ لأنه إذا لم يكن عشوائيًا فهو مؤشر على ضعف التحكم في الإنتاج، ولقد اقترح أسلوب لاختبار العشوائية في مثل هذه الحالة يطلق عليه اسم اختبار الدورة.

إن الأساليـب أو الطرائـق التي تستخدم لدراسة العشوائيــة بظاهرة معينة تعتمد أساسا على طبيعة وعـدد الـدورات الموجودة في بيانات تلك الظاهرة، وتعرف الدورة على أنها متتابعة من العناصر المتشابهة، تسبق وتلحق بعناصر من نوع آخر، وعدد العناصر داخل كل الدورة يطلق عليه طول الدورة. وغالبا ما نشك في عشوائية السلسلة التي تمثل ظاهرة معينة خاصة إذا كان عدد الدورات قليلا أو كثيرا.

شروط تطبيق الاختبار:

تتضمن البيانات متتابعة من المفردات مرتبة على حسب حدوثها، ويمكن تصنيفها إلى نوعين منفصلين فقـط . ولنفرضـ أن n تمثل حجم العينة وn1 تمثل عدد مفردات أو مشاهـدات النوع الأول، وn2 تمثل عدد مفردات أو مشاهدات النوع الثاني.

الفروض الإحصائية:

H0: نتائج حدوث النوع الأول والثاني من المفردات عشوائية.

H1: نتائج حدوث النوعين غير عشوائية.

مثال (10-4)

نفرض أن لدينا البيانات الآتية، وهي خاصة بعدد الوحدات المعيبة خلال 20 يومًا:

جدول (10-11)

الأيام	عدد الوحدات	الأيام	عدد الوحدات
1	8	11	15
2	5	12	12
3	9	13	9
4	14	14	8
5	9	15	7

6	16	3	6
5	17	5	7
3	18	10	8
2	19	9	9
1	20	3	10

والمطلوب معرفة هل هذه البيانات عشوائية أم لا، وذلك عند مستوى معنوية (دلالة) 0.05؟

الحل:

الفروض الإحصائية:

الفرض العدمي: البيانات عشوائية.

الفرض البديل: البيانات غير عشوائية.

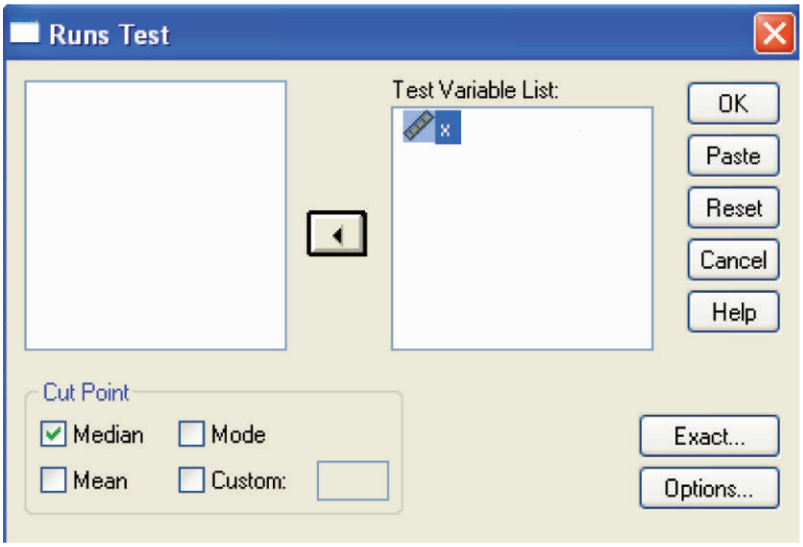
نقوم بإدخال البيانات في عمود متغير تمت تسميته X كما هو موضح في الجدول الآتي:

جدول (10-12)

	X
1	1.00
2	2.00
3	3.00
4	5.00
5	6.00
6	7.00
7	8.00
8	9.00
9	12.00
10	15.00
11	3.00
12	9.00
13	10.00
14	15.00
15	3.00
16	9.00
17	14.00
18	9.00
19	5.00
20	8.00

ثم نقوم بإجراء اختبار الدورة كما يلي:

Analyze → Nonparametric Tests → Runs…



شكل (10-13)

ثم ننقر على OK فتكون نتيجة الاختبار كالآتي:

جدول (10-13)

Runs Test

	x
Test Value[a]	8.00
Cases < Test Value	9
Cases >= Test Value	11
Total Cases	20
Number of Runs	8
Z	-1.114
Asymp. Sig. (2-tailed)	.265

a. Median

من جدول المخرجات السابقة يتضح لنا أن قيم الوسيط والقيم الأكبر و الأصغر منها وعدد القيم الكلية وإحصائي الاختبار وهو Z.

القرار الإحصائي:

ونلاحظ أن p-value=.265 وهي أكبر من 0.05 يعني أننا نقبل الفرض الذي يقول إن البيانات عشوائية.

(10-4) اختبار كلوموجروف سيمنروف لعينة واحدة (Sample K-S.1):

اقترح العالم الروسي كلوموجروف في سنة 1933 اختبار جودة المطابقة في حالة عينة واحدة، وفي سنة 1939 اقترح العالم الروسي سيمنروف اختبار جودة المطابقة في حالة بيانات تتعلق بعينتين، وبسبب وجود التشابه ما بين الاختبارين فقد أطلق على الاختبار الأول اسم كلوموجروف – سيمنروف لعينة واحدة وعلى الثاني اختبار كلوموجروف – سيمنروف لعينتين.

يعتمد هذا الاختبار على توزيعين احتماليين هما التوزيع الاحتمالي التراكمي النظري والتوزيع الاحتمالي التراكمي التجريبي، فعند اختيار عينة عشوائية من مجتمع بتوزيع F(x) غير معروف، حيث $F(x) = P(X \le x)$ فإن الهدف هو تحديد ما إذا كانت $F(x) = F_0(x)$ لجميع قيم x، حيث $F_0(x)$ تمثل دالة التوزيع التراكمي الفرضية، ولتحقيق هذا الهدف فإن اختبار كلوموجروف – سيمنروف لعينة واحدة ينظر إلى التقارب ما بين $F_0(x)$ و S(x) حيث S(x) تمثل دالة التوزيع التراكمي التجريبي، فإذا كان هذا التقارب ضعيفا فإنه يعني عدم صحة الافتراض القائل: بأن $F(x) = F_0(x)$ وخلاف ذلك الافتراض صحيحا.

يقوم هذا الاختبار بتحديد هل تتبع البيانات الداخلة توزيعًا معينًا أم لا، يستخدم الاختبار باختبار أربعة أنواع من التوزيعات وهي:

- التوزيع الطبيعي.
- التوزيع المنتظم.
- التوزيع الآسي.
- توزيع بواسون.

شروط الاختبار:

تتألف البيانات من عينة عشوائية $X_1, \ldots, X_n$ عدد مفرداتها يساوي n من مجتمع دالة توزيعه غير معروفة ونرمز لها بالرمز F(x).

الفروض الإحصائية:

$H_0 : F(x) = F_0(x)$ لجميع قيم x.

$H_1 : F(x) = F_0(x)$ على الأقل لقيمة واحدة من قيم x.

مثال (10 - 5)

البيانات الآتية هي 30 مفردة لمتغير يمثل عدد ساعات تشغيل آلة معينة X تم إدخالها في الجدول الآتي:

جدول (10-14)

المشاهدة	عدد الساعات	المشاهدة	عدد الساعات
1	7	16	1
2	4	17	2
3	4	18	1
4	4	19	2
5	3	20	4
6	22	21	22
7	4	22	9
8	0.4	23	6
9	9	24	8
10	0.5	25	6
11	5	26	4
12	7	27	7
13	2	28	2
14	3	29	0.47
15	8	30	7

المطلوب اختبار هل هذه البيانات تتبع التوزيع الأسي (وهو أحد التوزيعات الإحصائية المتصلة المهمة) أم لا عند مستوى معنوية 0.05؟

الحل:

الفروض الإحصائية:

الفرض العدمي: البيانات تتبع التوزيع الأسي.

الفرض البديل: البيانات لا تتبع التوزيع الأسي.

نقوم بإدخال البيانات على النحو الآتي:

جدول (10-15)

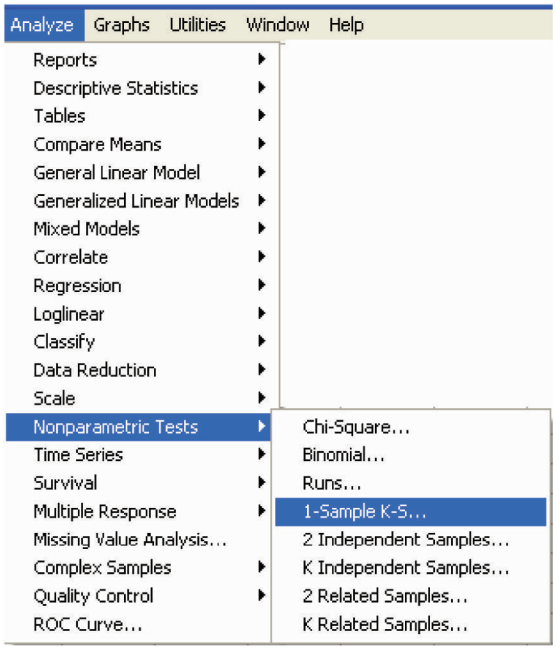
	X
1	7.00
2	4.00
3	3.00
4	4.00
5	9.00
6	5.00
7	2.00
8	8.00

9	2.00
10	2.00
11	22.00
12	6.00
13	6.00
14	7.00
15	.40
16	4.00
17	4.00
18	22.00
19	.40
20	.50
21	7.00
22	3.00
23	1.00
24	1.00
25	4.00
26	9.00
27	8.00
28	4.00
29	2.00
30	4.00

ثم نقوم بإجراء الاختبار كما يلي:

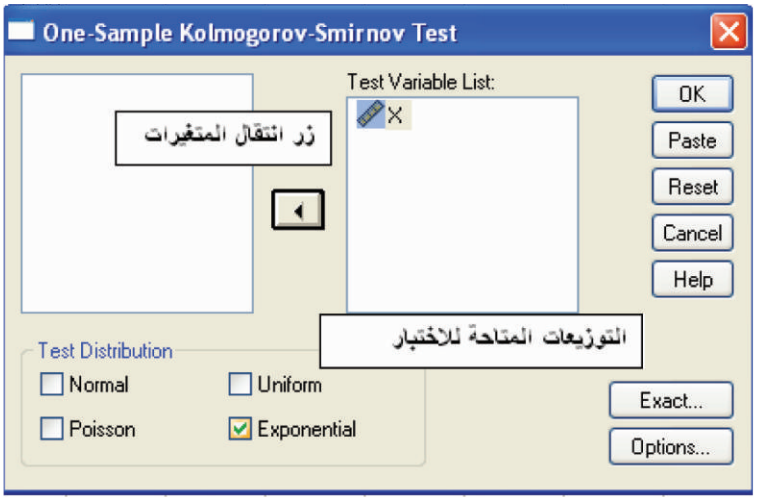
Analyze → Nonparametric Tests → 1 sample K-S…

كما في الشكل الآتي:



شكل (10 - 14)

فتظهر لدينا الشاشة الآتية:



شكل (10 - 15)

بعد النقر على OK تكون المخرجات كالآتي:

جدول (10 - 16)

One-Sample Kolmogorov-Smirnov Test

		X
N		30
Exponential parameter[a,b]	Mean	5.3767
Most Extreme Differences	Absolute	.158
	Positive	.121
	Negative	-.158
Kolmogorov-Smirnov Z		.866
Asymp. Sig. (2-tailed)		.441

a. Test Distribution is Exponential.

b. Calculated from data.

المخرجات السابقة بها الوسط وبعض المقاييس وقيمة دالة الاختبار وغيرها.

القرار الإحصائي:

نلاحظ أن p-value=.441 وهي أكبر من 0.05 إذا نقبل الفرض العدمي ونقول إن البيانات تتبع التوزيع الأسي بمتوسط 5.3767.

مثال (10-6)

البيانات الآتية هي 20 مشاهدة لمتغير يمثل عدد المنتجات التي تم بيعها في إحدى الشركات يوميًا، والتي نرمز لها بالرمز Y تم إدخالها في الجدول الآتي:

جدول (10 - 17)

المشاهدة	عدد المنتجات	المشاهدة	عدد المنتجات
1	2	11	4
2	14	12	16
3	3	13	9
4	16	14	20
5	4	15	8
6	18	16	13
7	11	17	2
8	17	18	14
9	5	19	11
10	19	20	8

المطلوب اختبار هل هذه البيانات تتبع التوزيع الطبيعي أم لا عند مستوى معنوية (دلالة) 0.05؟

الحل:

الفروض الإحصائية

الفرض العدمي: البيانات تتبع التوزيع الطبيعي.

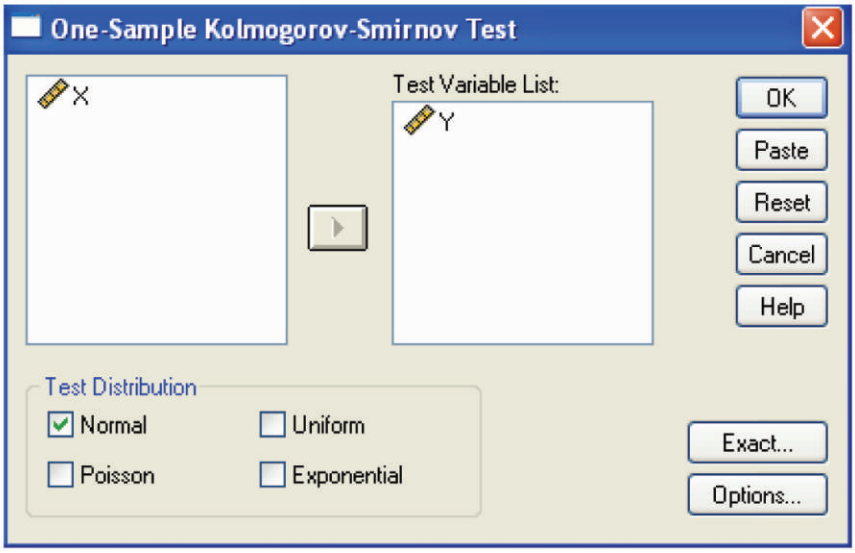
الفرض البديل: البيانات لا تتبع التوزيع الطبيعي.

نقوم بإدخال البيانات كالآتي:

جدول (10 - 18)

Y
2.00
3.00
4.00
11.00
5.00
4.00
9.00
8.00
2.00
11.00
14.00
16.00
18.00
17.00
19.00
16.00
20.00
13.00
14.00
8.00

ثم نقوم بإجراء الاختبار باختيار التوزيع الطبيعي كالآتي:



شكل (10 - 16)

بعد النقر على OK تكون المخرجات كالآتي:

جدول (10 - 19)

One-Sample Kolmogorov-Smirnov Test

		Y
N		20
Normal Parameters[a,b]	Mean	10.7000
	Std. Deviation	5.99210
Most Extreme Differences	Absolute	.129
	Positive	.129
	Negative	-.112
Kolmogorov-Smirnov Z		.578
Asymp. Sig. (2-tailed)		.892

a. Test distribution is Normal.

b. Calculated from data.

المخرجات السابقة بها الوسط والانحراف المعياري وغيرهما.

القرار الإحصائي:

نلاحظ أن p-value=0.892 وهي أكبر من 0.05 إذا نقبل الفرض العدمي ونقول إن البيانات تتبع التوزيع الطبيعي بمتوسط 10.7 وانحراف معياري 5.9921.

(10-5) اختبار مان – ويتني (The Mann – Whitney test "U")

يستخدم هذا الاختبار لاختبار الفرضية H_0 التي تهدف إلى معرفة مدى تطابق مجتمعين من حيث معلمتي الموقع (المتوسط أو الوسيط)، وذلك على أساس اختيار عينتين عشوائيتين منهما على أن تكون بيانات العينتين من نوع ترتيبي، أي إنه يساعد على الإجابة عن الأسئلة التي من النوع "هل أحد المجتمعين يبدو أنه يعطي قيمًا أكبر من المجتمع الآخر؟ "أو" هل وسيطا المجتمعين متساويان؟". ويعد هذا الاختبار من أقوى الاختبارات اللامعلمية المستخدمة لهذا الغرض، ويستخدم هذا الاختبار رتب المفردات نفسها، ويفضل استخدام الرتب للأسباب الآتية:

- إذا كانت الأعداد المعطاة للمفردات لا معنى لها في حَدِّ ذاتها ولكن يكون لها معنى في حالة مقارنتها بالترتيب مع الأعداد الأخرى فقط؛ أي إن الأعداد لا تحتوي على معلومات أكثر مما تحتويه الرتب، وهذا من طبيعة البيانات التي من نوع ترتيبي.
- حتى إذا كان لهذه الأعداد معنى ولكن دالة التوزيع لا تتبع التوزيع الطبيعي، فإن نظرية الاحتمالات عادة لا تكون في متناولها عندما تكون إحصاءة الاختبار تعتمد على البيانات الحقيقية، علاوة على ذلك فإن نظرية الاحتمالات المبنية على الرتب تُعَدُّ نسبيًا سهلة ولا تعتمد على التوزيع في كثير من الحالات.
- إن الكفاءة النسبية لاختبار مان–ويتني ليست بسيئة مقارنة باختبار t المألوف، وعليه يفضل استخدامه للأسباب المذكورة أعلاه.

شروط تطبيق الاختبار:

1- تتضمن البيانات عينة عشوائية من المفردات $X_1 \ldots X_m$ من المجتمع "1" بدالة توزيع F، وهذا يعني طبعا استقلالية المشاهدات، عينة عشوائية $Y_1 \ldots Y_n$ أخرى من المجتمع "2" بدالة توزيع G، وذلك يعني استقلالية البيانات عن بعضها داخل العينة الواحدة.

2- العينتان المستقلتان عن بعضهما بعضًا.

3- وحدة القياس على الأقل ترتيبي.

4- إذا وجد اختلاف بين دوال توزيع المجتمعين، فإن الاختلاف سيكون في موقع التوزيع، أي إنه إذا كانت $F(x) \neq G(x)$ حيث F هي دالة التوزيع الأول، G هي دالة التوزيع الثاني فإن:

$$F(x) = G(x+\Delta)$$

أو

$$G(x) = F(x+\Delta)$$

وهذا يعني أن

$$Y^d = X + \Delta$$

d حيث تشير إلى أنه من التوزيع نفسه، ونجد أيضًا أن المعلمة Δ تشير إلى Location shift، ومن ثم فإن $\Delta = E(Y) - E(X)$.

الفروض الإحصائية:

$$H_0 : (E(X) = E(Y))$$

$$H_A : (E(X) \neq E(Y))$$

أو

$$H_0 : \Delta = 0$$

ضد الفرض البديل

$$H_A : \Delta \neq 0$$

إحصاءة الاختبار:

$$T = \sum_{i=1}^{m} R(X_i)$$

حيث:

$\sum_{i=1}^{m} R(X_i)$: مجموع الرتب لعينة المجتمع الأول.

m: حجم العينة الأولى و n: حجم العينة الثانية.

- **القرار الإحصائي:**

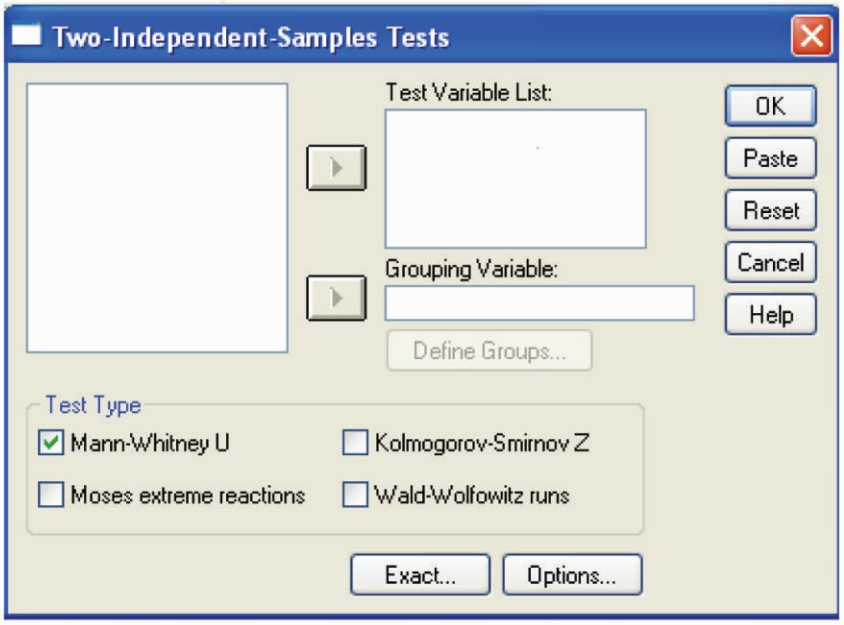
نرفض H_0 إذا كان

$$T \geq \omega_{\frac{\alpha}{2}} \quad or \quad T \leq n(m+n+1) - \omega_{\frac{\alpha}{2}}$$

ونلاحظ أن $\omega_{\frac{\alpha}{2}}$ يتم استخراجها من جدول التحليل اللامعلمي.

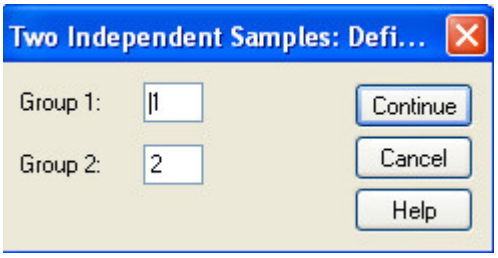
نقوم بتفعيل اختبار مان-ويتني كالآتي:

Analyze → Nonparametric Tests → 2 Independent Samples → Mann – Whitney U



شكل (10 - 17)

حيث يجب على الباحث أن يدخل قيم المتغير إلى test variable list وتصنيف المتغير في Grouping Variable، حيث يعني أن "1" أن تلك القيمة تخص العينة الأولى، والقيمة "2" تعني أن القيمة تخص العينة الثانية كما في الشكل الآتي:



شكل (10 - 18)

مثال (10-7)

في دارسة لمقارنة بين متوسطي نسبة الملوحة في مياه نهر دجلة ومياه نهر النيل أخذت عينة على مدار 10 أيام لمياه دجلة وعلى مدار 5 أيام لمياه النيل وسجلت النسب كما يلي:

مياه النيل العينة الأولى:

جدول (10 - 20)

0.3	0.15	0.13	0.1	0.5

العينة الثانية عينة مياه نهر دجلة:

جدول (10 - 21)

.170	.250	.530	.90	0.32	.150	0.33	0.45	0.4	0.6

المطلـوب اختبار أن نسبة الملوحة في الميـاه نهر دجلة هي نفسها نسبة الملوحة في نهر النيل عند مستوى معنوية 0.05 . مع العلم أن النسب مسحوبة من مجتمعين مستقلين؟

الحل:

الفروض الإحصائية:

H0: لا يوجد فرق في نسبة الملوحة بين المياه.

H1: يوجد فرق في نسبة الملوحة في النوعين.

نقوم بإدخال قيم المتغيرين في العمود ونسميه X وقيم المتغيرين في العمود الثاني الأول، وفي هذا المثال يأخذ 1 و 2 و نسميه Group كما يلي:

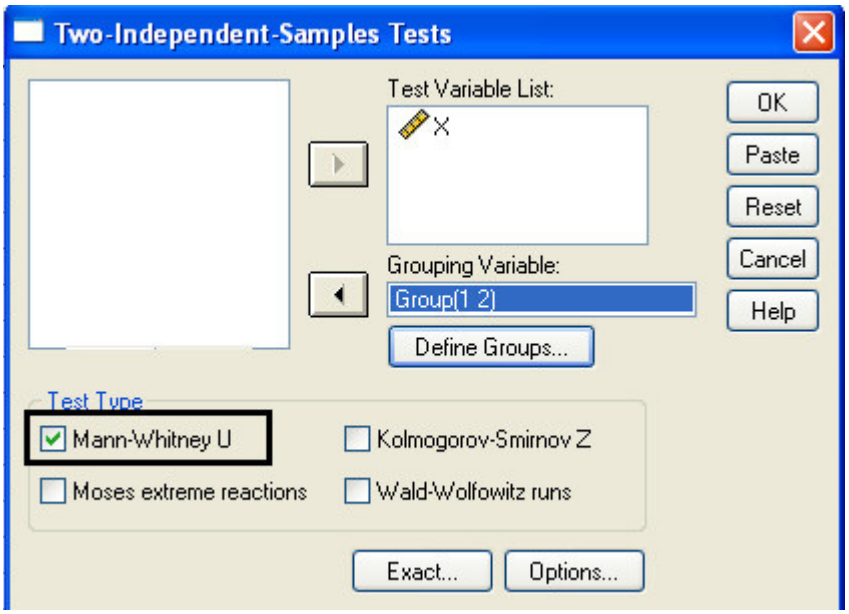
جدول (10 - 22)

	X	Group
1	50.	1
2	10.	1
3	13.	1
4	15.	1
5	30.	1
6	60.	2
7	40.	2
8	45.	2
9	33.	2
10	15.	2
11	32.	2
12	90.	2
13	53.	2
14	25.	2
15	.17	2

ثم نقوم بإجراء اختبار عينتين مستقلتين مان ويتني:

Analyze → Nonparametric Tests →2 Independent Samples → Mann – Whitney U

ونضـع X في test variable list و Group في Grouping variable بعد تعريفها كما ذكرنا سابقا،كما في الشكل الآتي:



شكل (10 - 19)

بعد النقر على OK تخرج لنا النتائج الآتية:

جدول (10 - 23)

Ranks

	Group	N	Mean Rank	Sum of Ranks
X	1	5	5.10	25.50
	2	10	9.45	94.50
	Total	15		

جدول (10 - 24)

	X
Mann-Whitney U	10.500
Wilcoxon W	25.500
Z	-1.777
Asymp. Sig. (2-tailed)	.075
Exact Sig. [2*(1-tailed Sig.)]	.075[a]

a. Not corrected for ties.

b. Grouping Variable: Group

الجدول الأول شكل (10-30) يحتوي على اسم المتغير وعن متوسط الرتب لكل متغير وغيرها.

القرار الإحصائي:

الجـدول (11-24) يوضح بيانات إحصـاءة الاختبار، وهو اختبار مان ويتني ونلاحظ أيضا أن قيمة p-value = 0.075، وهـي أكـبر من $\frac{\alpha}{2} = 0.075$ إذًا نقبل الفرض العدمي ونقول: إنه لا يوجد فرق في نسبة الملوحة بين النهرين.

مثال (10-8)

في تجربة لمقارنة متوسط الإنتاج اليومي من اللبن لنوعين من الأبقار أخذت عينة بحجم أربع بقرات من النوع A وعينة بحجم خمس بقرات من النوع B، وسجل الإنتاج اليومي كالآتي:

جدول (10 - 25)

A	18	12	10	15	–
B	22	21	19	20	16

هل يمكن أن نستنتج أن الإنتاج اليومي لكل من النوعين متساوٍ عند مستوى معنوية 0.05؟

الحل:

الفروض الإحصائية:

H0: لا يوجد اختلاف في الإنتاج اليومي لكلا النوعين.

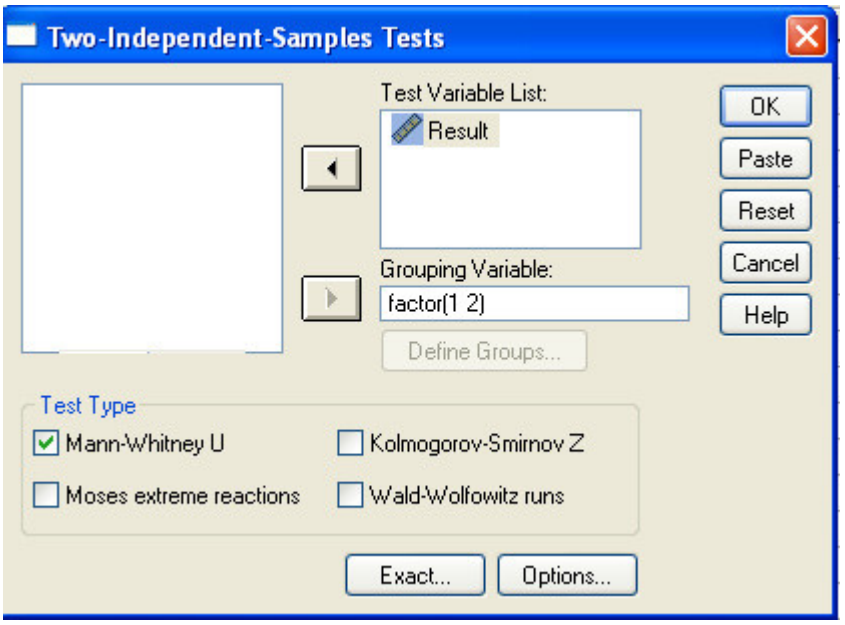
H_1: يوجد اختلاف في الإنتاج اليومي لكلا النوعين.

نقوم بإدخال البيانات كما في السابق وكما هو موضح في الشكل الآتي:

جدول (10 - 26)

	factor	Result
1	1.00	18.00
2	1.00	12.00
3	1.00	10.00
4	1.00	15.00
5	2.00	22.00
6	2.00	21.00
7	2.00	19.00
8	2.00	20.00
9	2.00	16.00

ثم نقوم بإدخال كما في السابق وكما هو موضح في الشكل الآتي:



شكل (10 - 20)

بعد النقر على OK تخرج لنا النتائج الآتية:

جدول (10 - 27)

Ranks

	factor	N	Mean Rank	Sum of Ranks
Result	1.00	4	2.75	11.00
	2.00	5	6.80	34.00
	Total	9		

Test Statistics [b]

	Result
Mann-Whitney U	1.000
Wilcoxon W	11.000
Z	-2.205
Asymp. Sig. (2-tailed)	.027
Exact Sig. [2*(1-tailed Sig.)]	.032[a]

a. Not corrected for ties.

b. Grouping Variable: factor

من المخرجات نلاحظ مجموع ومتوسط الرتب لكلا المتغيرين، وهذا في الجدول الأول (10-27).

القرار الإحصائي:

الجـدول الثـاني نلاحظ أن p-value= 0.027 وهي أكبر من $\frac{\alpha}{2}$ = .025 (لأن الاختبار من طرفين) أي نقبل الفرض العدمي ونقول: إنَّه لا يوجد اختلاف في الإنتاج اليومي لكلا النوعين.

(10-6) اختبار كرسكال والاس (The Kruskal – Wallis):

يستخدم لاختبار الفرضية التي تهدف إلى معرفة ما إذا كانت عدة عينات قيد الدراسة قد تم اختيارها من مجتمعات بدوال توزيع متطابقة، علاوة على ذلك أنه يستخدم أكبر قدر ممكن من المعلومات التي بالعينات مقارنة باختبارات أخرى تستخدم للغرض نفسه، وعليه فهو أكثر قوة ويفضل استخدامه خاصة عندما تكون وحدة قياس البيانات على الأقل ترتيبي.

شروط تطبيق الاختبار:

- تحتوي البيانات على مفردات k عينة عشوائية حجم كل عينة منهم n_1 .... n_k على التوالي.
- المفردات مستقلة عن بعضها بعضًا خلال العينة ومن عينة إلى أخرى.
- وحدة القياس على الأقل ترتيبي.
- جميع دوال التوزيعات لجميع العينات $F_1,F_2,...,F_k$ ترتبط بالعلاقة الآتية:

$$Fj(t) = F(t-\tau_i), \quad -\infty < t < \infty$$

حيث إن $j = 1,2,...,k$ والرمز τ يقرأ "تاو".

الفروض الإحصائية:

H_0: دوال التوزيع لجميع المجتمعات متطابقة أو

$$H_0 = [\tau_1,...,\tau_k \text{ all equal}]$$

H_A: ليس جميع دوال التوزيع متطابقة أو

$$H_A = [\tau_1,...,\tau_k \text{ not all equal}]$$

إحصاءة الاختبار:

$$T = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3\,(n+1)$$

حيث:

$R_i = \sum R(X_{ij})$: مجموع الرتب المعطاة لمفردات العينة i

n: مجموع عدد المفردات في جميع العينات.

القرار الإحصائي:

إذا كان حجم العينات صغيرًا فإننا

نرفض H_0 إذا كان $T \geq \omega_{1-\alpha}$

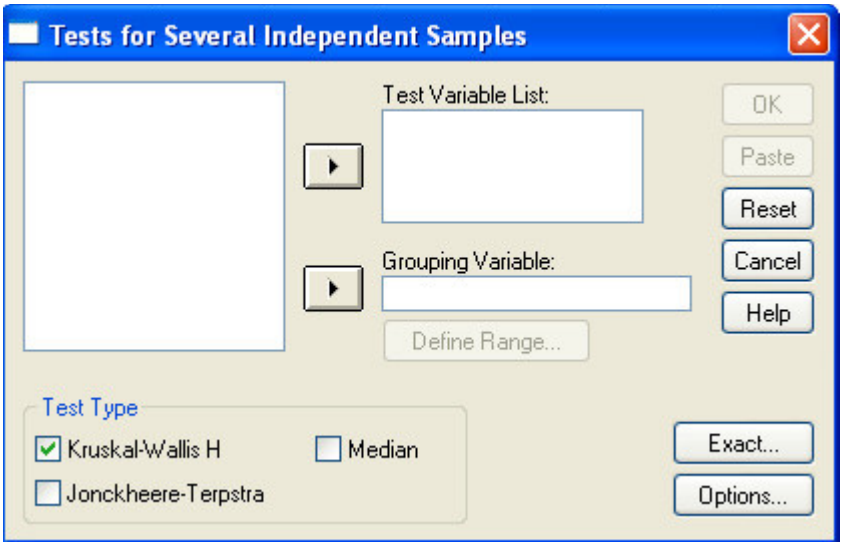
حيث تستخرج تلك القيم من جداول كرسكال والأس لتحليل اللامعلمي. انظر (Holiander and Wolfe 1998).

و إذا كان حجم العينات كبيرًا فإننا

نرفض H_o إذا كان $T \geq x_{\alpha,k-1}$

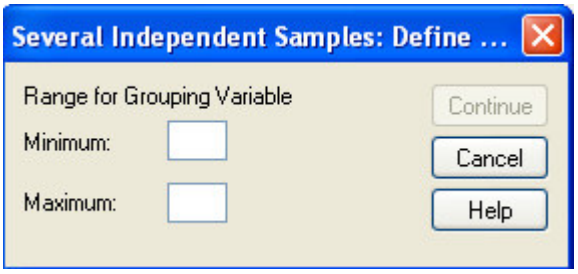
نقوم بتفعيل اختبار كرسكال والأس كالآتي:

Analyze ← Nonparametric Test ← k independent Sample ← Kruskal – Wallis



شكل (10 - 21)

حيث نقوم بإدخال عمود قيم المتغيرات في خانة تسمى Test Variable List وإدخال عمود تصنيف القيم "Group" في خانة Grouping Variable ثم نقوم بتعريفها بالنقر على مربع Define Range، حيث يدخل الباحث رقم أول مجموعة دائما تكون واحد في خانة Minimum ورقم آخر مجموعة، وهي عدد العينات في خانة Maximum كما في الشكل الآتي:



شكل (10 - 22)

مثال (10-9)

تم اختبار ثلاثة أنواع من الأغذية على الأطفال، تم اختبار الطعام A على عينة من 5 أطفال، تم اختبار الطعام B على عينة من 6 أطفال، تم اختبار الطعام C على عينة من 5 أطفال، وتم قياس أوزانهم بعد 7 أسابيع وكانت النتائج كالآتي:

جدول (10 - 28)

Food Type	أوزان الأطفال					
A	11.2	12,1	10,9	11,3	12	–
B	12,6	10,8	11,3	11	1,2	10,7
C	11,3	11,9	12,4	10,6	12	–

اختبر ما إذا كان متوسط أوزان الأطفال الذين تناولوا الأطعمة A، B، C متساوية عند مستوى معنوية 0,05.

الحل:

الفروض الإحصائية:

H_0: جميع الأطعمة متساوية التأثير في وزن الأطفال.

H_A: على الأقل يوجد نوع من أنواع الأطعمة ذو تأثير متميز في الوزن.

أولا: نقوم بإدخال البيانات في عمود تكون فيه قيم المتغيرات الثلاث، وعمود يكون فيه تصنيف القيم كما يلي:

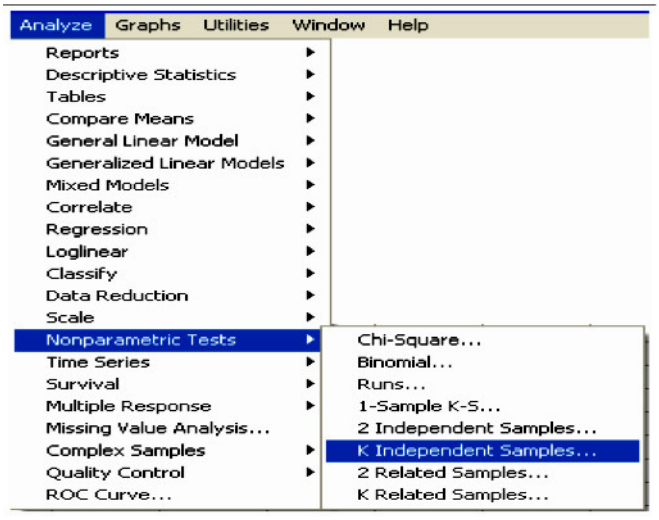
جدول (10-29)

	Vaule	Group
1	11.30	1
2	11.90	1
3	12.40	1
4	10.60	1
5	12.00	1
6	12.60	2
7	10.80	2
8	11.30	2
9	11.00	2
10	12.00	2
11	10.70	2
12	11.20	3
13	12.10	3
14	10.90	3
15	11.30	3
16	12.00	3

ثم نقوم بالآتي:

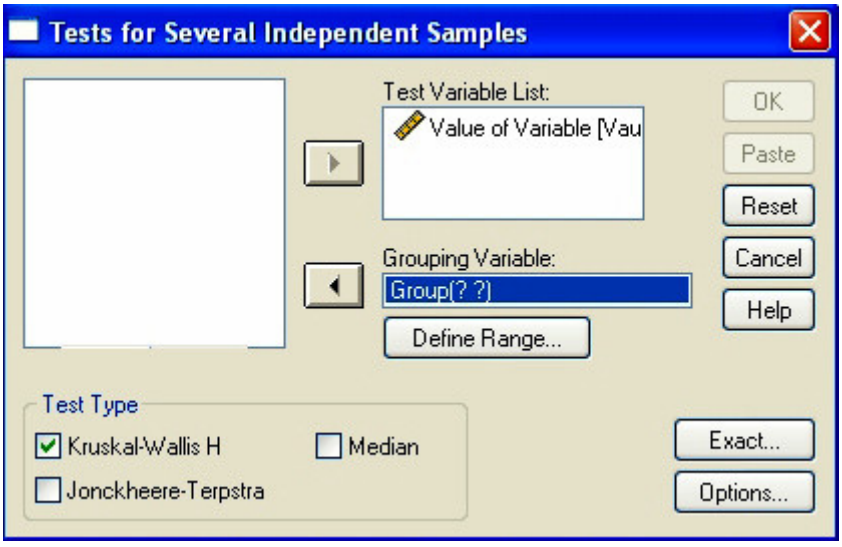
- من شريط القوائم نختار Analyze.
- ثم نختار Nonparametric Test.
- لاختبار اختبار أكثر من عينتين مستقلتين نختار k independent Sample

كما في الشكل الآتي:



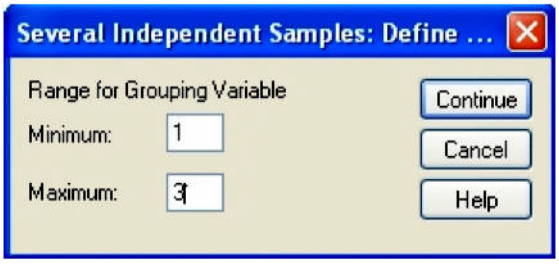
شكل (10 - 23)

- في الشكل (10-24) نقوم باختيار Kruskall – Wallis.
- نقوم بإدخال عمود Value في مربع Test Variable List
- وGroup في Grouping Variable كما في الشكل الآتي:



شكل (10-24)

- ونعرف المجموعة كما في الشكل الآتية:



شكل (10-25)

- ثم بعد النقر على OK تخرج لنا النتائج الآتية:

جدول (10 - 30)

Ranks

	Group	N	Mean Rank
Value of Variable	1	5	9.20
	2	6	7.67
	3	5	8.80
	Total	16	

جدول (10 - 31)

Test Statistics[a,b]

	Value of Variable
Chi-Square	.315
df	2
Asymp. Sig.	.854

a. Kruskal Wallis Test
b. Grouping Variable: Group

نلاحظ من الجدول (10 – 30) عدد العينة لكل مجموعة وأيضا متوسط الرتب لكل مجموعة.

والجدول (10 – 31) هو أهم جدول لدينا لاحتوائه على معلومات الاختبار.

القرار الإحصائي:

نلاحظ من الجدول الثاني أن p-value= 0.854 وهو أكبر من مستوى المعنوية 0.05 إذا نقبل الفرض العدمي ونقول إن تأثير الثلاثة أنواع من الأطعمة متساوية على الوزن.

مثال (10-10)

لمعرفة مدى تأثير ثلاث طرق إنتاج A،B، C لإنتاج منتج ما، تم عمل اختبار لتلك الطرق وكانت نتائج عدد الوحدات المنتجة في كل طريقة على مدار 5 أيام مختلفة كالآتي:

جدول (10-32)

الطريقة A	13	15	25	15	17
الطريقة B	27	13	12	11	23
الطريقة C	14	16	18	19	20

اختبر صحة الفرض القائل: إنَّه على الأقل توجد طريقة من الطرق الثلاثة مؤثرة في حجم الإنتاج عند مستوى معنوية (دلالة) 0.01.

الحـــل:

الفروض الإحصائية:

H_0: الثلاث طرق متساوية التأثير في الإنتاج.

H_A: على الأقل توجد طريقة واحدة لها تأثير مميز في حجم الإنتاج.

أولاً: نقوم بإدخال البيانات عمود يكون فيه قيم المتغيرات الثلاثة وعمود يكون فيه تصنيف القيم كما في الشكل الآتي:

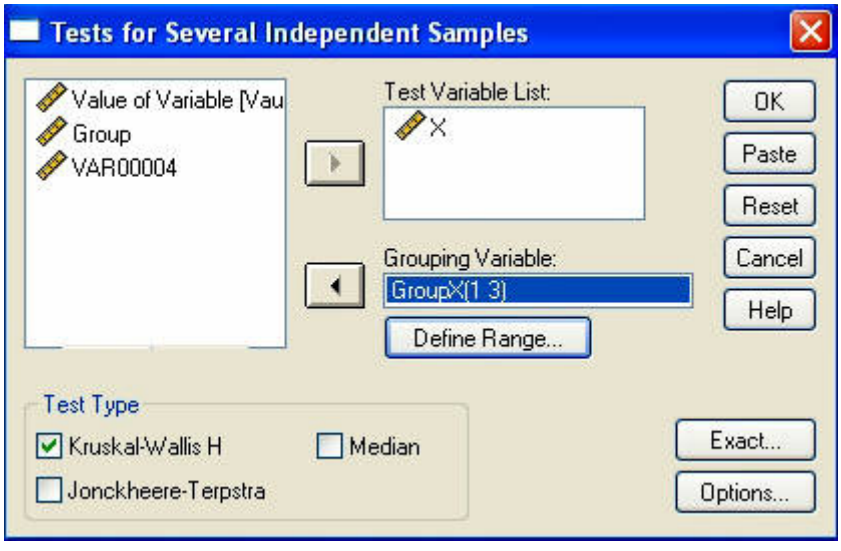
جدول (10 - 33)

Vaule	Group
13.00	1.00
15.00	1.00
25.00	1.00
15.00	1.00
17.00	1.00
27.00	2.00
13.00	2.00
12.00	2.00
11.00	2.00
23.00	2.00
14.00	3.00
16.00	3.00
18.00	3.00
19.00	3.00
20.00	3.00

نقوم بالآتي:

Analyze ← Nonparametric Test ← k independent Sample ← Kruskal – Wallis

ثم نقوم بإجراء اختبار عدة عينات مستقلة بإدخال X إلى test variable list و GroupX إلى Grouping variable كما في الشكل الآتي:



شكل (10-26)

ثم بعد النقر على OK تخرج لنا النتائج الآتية:

جدول (10-34)

Ranks

	GroupX	N	Mean Rank
X	1.00	5	7.90
	2.00	5	6.90
	3.00	5	9.20
	Total	15	

جدول (10-35)

Test Statistics[a,b]

	X
Chi-Square	.667
df	2
Asymp. Sig.	.716

a. Kruskal Wallis Test
b. Grouping Variable: GroupX

القرار الإحصائي:

نلاحظ من الجدول الثاني أن p-value=.716 وهو أكبر من مستوى المعنوية 0.01 إذًا نقبل الفرض العدمي ونقول: إنَّ الثلاث طرق إنتاج لها تأثير متساوٍ في الإنتاج.

الفصل الحادي عشر

التعامل مع النتائج

Working with Output

(11 - 1) قص - نسخ - لصق جداول النتائج

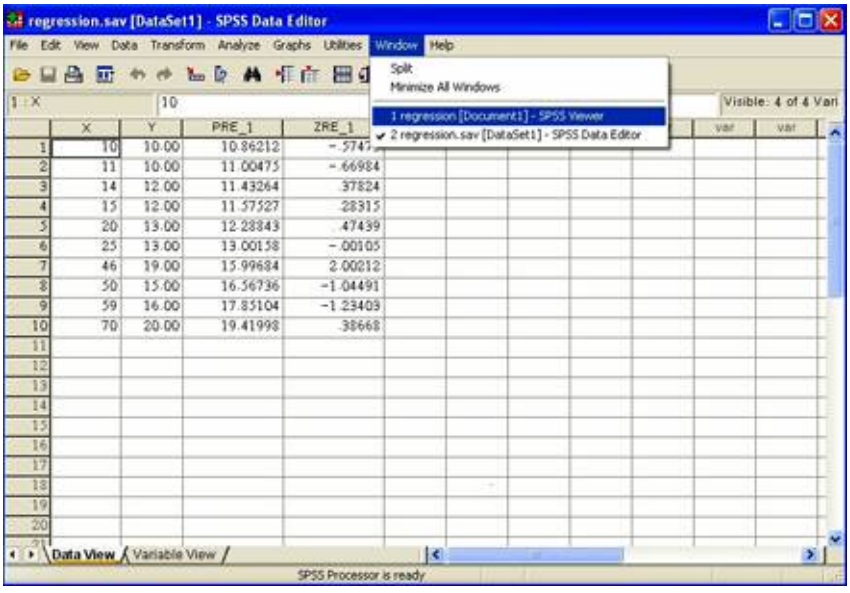
طبيعـي جدًا أننا بعد إجراء التحليل الإحصائـي فإنه سوف تكون لدينا نتائج ونريد نسخ تلك النتائج إلى ملفـات تحريـر النصوص، مثل الملفـات المصممة من خلال برنامـج الوورد. وهنا يتيـح البرنامج لنا مجموعـة مـن الخيارات مثل نسـخ وقص ولصق الجدول والأشـكال البيانية إلى ملفـات تحرير النصوص المختلفة لدينا:

تطبيقات:

قص الجداول والأشكال البيانية:

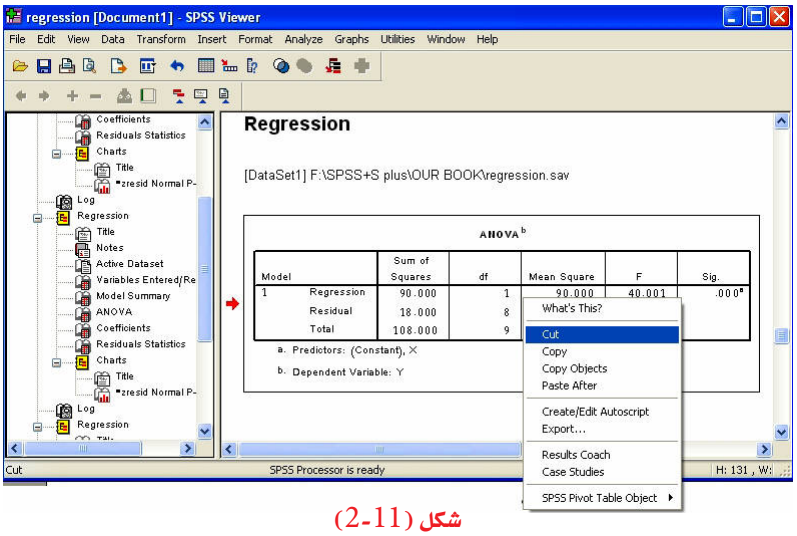
مثال (11-1):

ولتحرير النتائج في مثال رقم (9-1) نقوم بفتح SPSS viewer وذلك من خلال النقر على Windows ثم اختيار SPSS viewer مثل الشكل الآتي:

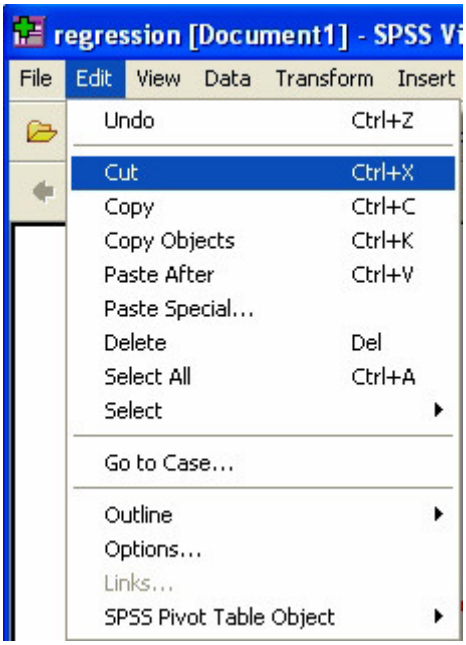


شكل (11-1)

ومن ثم ظهور شكل محرر النتائج في الشكل (11-2)، وإذا أردنا قص جدول تحليل التباين ANOVA فإننـا ننقر بزر الماوس الأيسـر عليها لتحديد الجدول، وبعد ذلك ننقر بزر الفارة الأيمن، ومن ثم الاختيار Cut أو بالنقر على edit من شريط القوائم، ومن ثم اختيار cut كما في الشكل (11-3):



شكل (11-2)



شكل (11-3)

وبعد ذلك نذهب إلى ملف تحرير النصوص الذي نستخدمه ثم النقر على قائمة edit من شريط القوائم في برنامـج التحريـر، ومن ثم اختيار paste أو لصق على حسب نوع نسخة البرنامج عربي أو إنجليزي، وكما فعلنا في الجداول يمكننا أيضًا قص الأشكال البيانية بالطريقة نفسها، ويمكننا تلخيص الخطوات كالآتي:

- فتح صفحة النتائج أو محرر النتائج.
- في صفحة محرر النتائج انقر على الجدول المراد قصه.
- انقر على الزر الأيمن للفأرة في الجدول المراد تنسدل قائمة من الأوامر انقر على الأمر Cut.
- انتقل إلى الصفحة المراد لصق الجدول فيها، وفي المكان المراد لصق الجدول فيه انقر على الزر الأيمن ثم انقر على الأمر Pastee .

(11-2) نسخ الجداول والأشكال البيانية:

هنا نقوم بتابع الخطوات الآتية:

- فتح صفحة محرر النتائج.
- في صفحة النتائج انقر على الجدول المراد نسخه.
- انقر على الزر الأيمن للفأرة تنسدل قائمة من الأوامر يظهر بها خياران متعلقان بالنسخ Copy أو Copy object، فإذا أردنا نسخ جدول بمحرره، أي نسخه مع إمكانية التعديل في جدول النتائج نختار Copy، وإذا أردنا نسخ الجدول كصورة نختار Copy objects، والخيارات تلك ظاهرة في الشكل (11-2).
- انتقل إلى الصفحة المراد نسخ الجدول إليها، وفي المكان المراد نقل الجدول إليه انقر على الزر الأيمن ثم انقر على الأمر Pastee.
- لاحظ أن الشكل المنقول غير منسق وغيابي كان نوع الخط هو Arial من الحجم 10 يمكن باستخدام محرر Word تغيير نوع الخط مثلا / Time New Roman من الحجم 12 إذا تم اختيار الأمر Copy، وإذا اخترنا Copy objects فلا يمكننا إجراء التعديل عليها فهي كصورة.

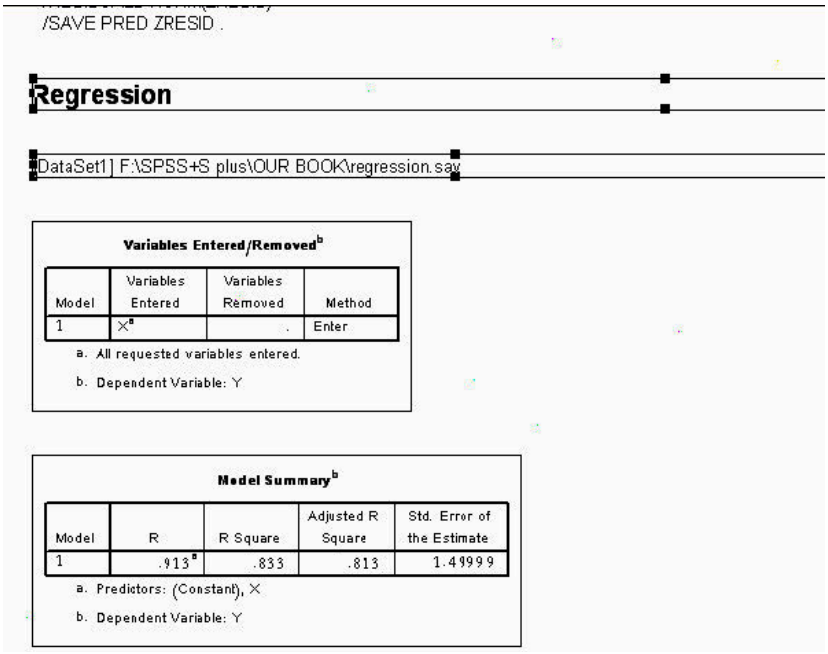
الآن نريد أن نتعامل مع المخرجات وهي الجداول بالإظهار والاختصار وإخفاء وإظهار بعض النتائج.

(11-3) الإظهار Expand والإخفاء Collapse وتحريك Move النتائج:

- **إظهار وإخفاء النتائج:**

إذا أردنا أن نخفي نتائج مثال (9-1) نقوم بالآتي:

- انتقل إلى محرر النتائج.
- نقوم بتحديد صفحة النتائج من خلال النقر على edit، ومن ثم اختيار Select all سوف تضاء كل محتويات الصفحة (يظهر إطار أسود لكل محتوى الصفحة انظر الشكل (11 - 4):


/SAVE PRED ZRESID .

Regression

[DataSet1] F:\SPSS+S plus\OUR BOOK\regression.sav

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]

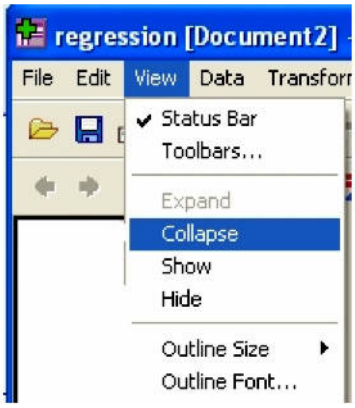
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.913[a]	.833	.813	1.49999

a. Predictors: (Constant), X

b. Dependent Variable: Y

شكل (11 - 4)

من القائمة View انقر الأمر Collapse ستختفي كل نتائج الصفحة عقب النقر انظر الآتي (انظر الشكل 11 – 5):



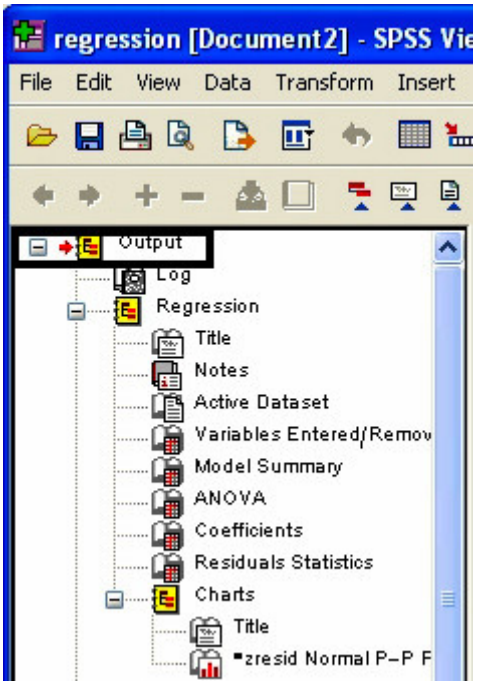
شكل (11 - 5)

لإعادة إظهار المحتويات نكرر ما أجريناه في الخطوة السابقة مع النقر على الأمر Expand.

ملاحظات:

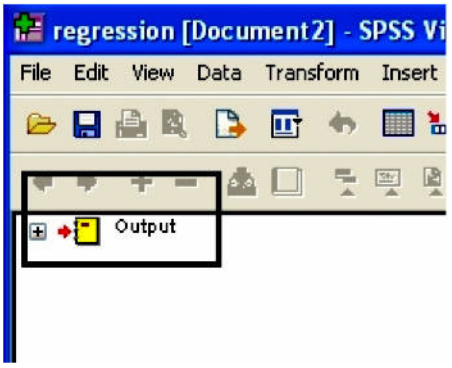
يمكـن حذف محتويات الشاشة بالنقر على الأمر Delete من القائمة Edit ثم إعادتها مرة أخرى بالأمر Undo من القائمة نفسها.

يمكن الإخفاء بالنقر على المربع الموجود بداخله علامة الطرح في الجزء الأيسر يسار كلمة Output المشار إليه في الشكل الآتي:



شكل (11 - 6)

والإظهار بإعادة النقر مرة أخرى على المربع نفسه، ونلاحظ أن علامة الطرح تحولت إلى علامة جمع كما في الشكل الآتي:



شكل (11 - 7)

إخفاء جزء من عنوان Sub-Item Titled لنتائج موجودة في صفحة النتائج كالآتي:

انقر على الجدول أو الشكل المراد إخفاؤه في محرر النتائج، فمثلا في المثال (9-1) إذا أردنا إخفاء الجدول الذي بعنوان Variables Entered/Removed نقوم بتحديده، وذلك من خلال النقر عليه فيظهر الشكل الآتي، ونلاحظ أن أي جدول يتم تحديده يظهر محدد بإطار أسود وبجواره من ناحية اليسار سهم أحمر اللون:

Variables Entered/Removed[b]

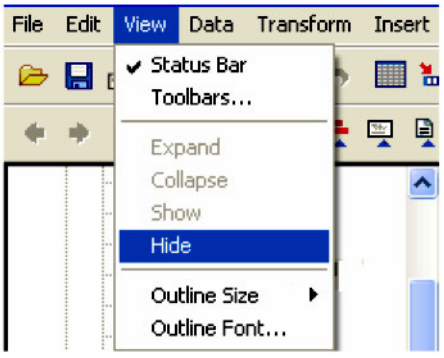
Model	Variables Entered	Variables Removed	Method
1	X[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Y

شكل (11 - 8)

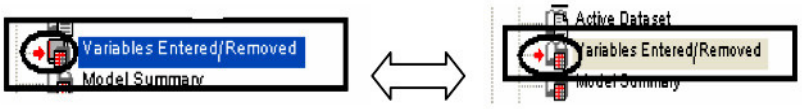
من القائمة View انقر الأمر Hide سيختفي الجزء الآتي بمجرد النقر عليه كما في الشكل الآتي:



شكل (11 - 9)

لإظهار الجزء الذي اختفى انقر على الأمر Show من القائمة View ستظهر المحتويات مرة أخرى، ونلاحظ أننا عند الإخفاء في أول مرة تكون Show غير منشطة، وعندما نريد إظهار ما أردنا إخفاءه تكون Hide هي غير المنشطة بمعنى غير متاحة الاستخدام.

ملاحظة: يمكن الإخفاء والإظهار باستخدام القوائم Menu System، ويتم ذلك بالنقر المزدوج على أيقونة الكتاب المجاورة للعنوان في الفهرس كما في الشكل الآتي، فنلاحظ أنه عندما يكون الجدول متاحًا أو ظاهرًا، فإن الكتاب يكون مفتوحًا، وعندما يكون الجدول أو الشكل غير متاح فبالنقر نقرًا مزدوجًا يصبح الكتاب مغلقًا والنقر عليه مرة أخرى يصبح مفتوحًا، ومن ثم إظهار الشكل أو الجدول والعكس صحيح:

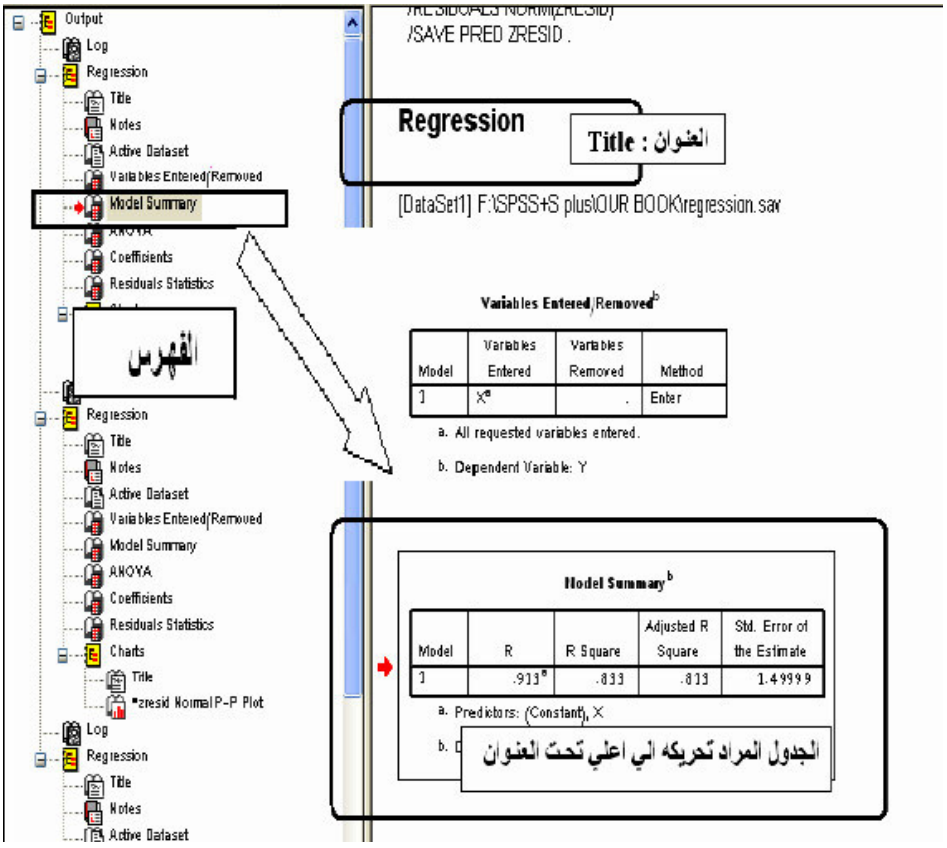


- **تحريك نتائج إلى مكان آخر:**

في صفحة النتائج يمكن تحريك Move، أي محتوى من محتويات النتائج من مكانه إلى مكان آخر.

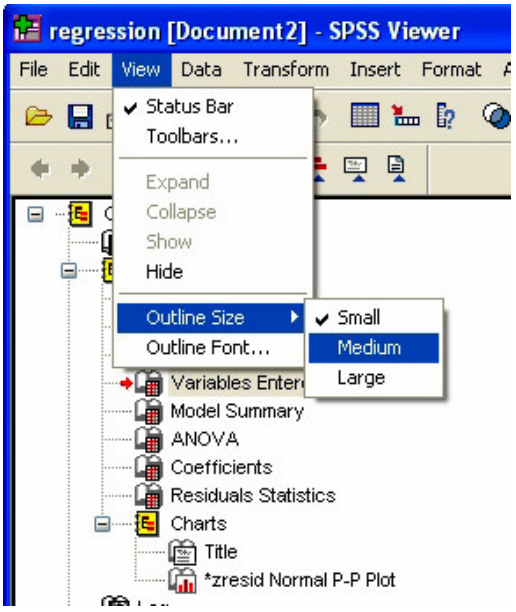
نفرض أننا نريد تحريك الجدول الذي بعنوان model summary إلى أعلى أسفل العنوان regression في المثال (10-1) فإننا نقوم باتباع الخطوات الآتية:

- انقر على الجدول المراد تحريكه في شاشة النتائج ونلاحظ أنه سيظهر السهم الأحمر والإطار الأسود وبذالك يكون الجدول نشطًا.
- سيضاء عنوان الجدول وهو المراد نقله في الفهرس يسارًا.
- في الفهرس يسار الشاشة وباستخدام الزر الأيسر للفارة انقر على model summary، واستمر في النقر مع تحريك الاسم في الفهرس إلى المكان المراد نقله إليه، ونفرض أننا نريد نقله إلى أسفل العنوان Regression مباشراً.
- عند الوقوف إلى المكان الذي نريد أن نترك الماوس فيتحرك الجدول إلى المكان الذي حددناه. انظر الشكل (11-10).



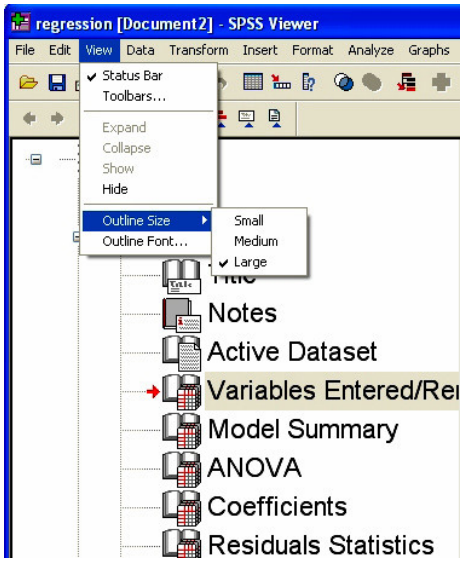
شكل (11 - 10)

- تغيير حجم وشكل الخط في الفهرس Change The Size and Font:
  - يمكن للمستخدم تغيير نوع الخط وحجمه في الفهرس بسهولة بالخطوات الآتية:
  - من القائمة View انقر على الأمر Outline Size.
  - فتظهر قائمة يتم الاختيار منها كالآتي:



شكل (11 - 11)

- فنجد من الشكل (11-11) أننا لدينا حجمًا صغيرًا للخط في الفهرس، وعرفنا ذلك من وجود علامة √ small، وأن الحجم فعلاً في الفهرس صغير، وهذا واضح من حجم الخط، وإذا أردنا أن نغيره إلى كبير فإننا ننقر على Large في الشكل (11-11)، ومن ثم يصبح حجم الخط كبيرًا كما في الشكل الآتي:

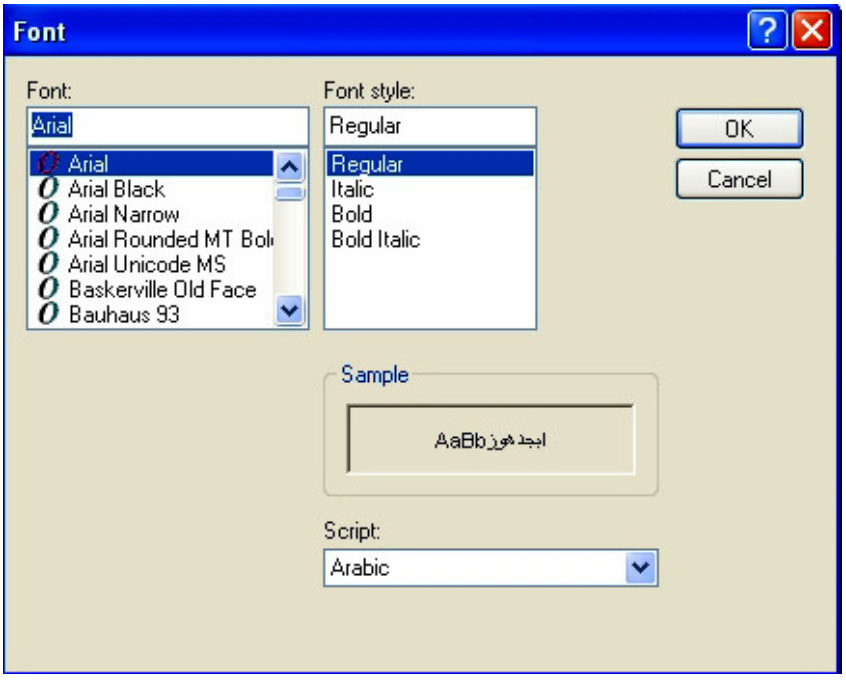


شكل (11 - 12)

كما أننا إذا أردنا أن نغير حجم الخط إلى متوسط فإننا ننقر على الأمر Medium سيتغير شكل الخط في الفهرس من الكبير إلى متوسط.

ونجد أيضًا أننا يمكننا تغيير نوع الخط في الفهرس كما يلي:

- من القائمة View انقر الأمر Outline Font سيُفتح صندوق به العديد من الأنواع انقر على النوع الذي ترغبه من الخطوط. انظر الشكل (11 – 13):



شكل (11 - 13)

- وبعد الانتهاء من الاختيار يتم النقر على الأمر Ok سيتغير نوع الخط بمجرد التنفيذ.

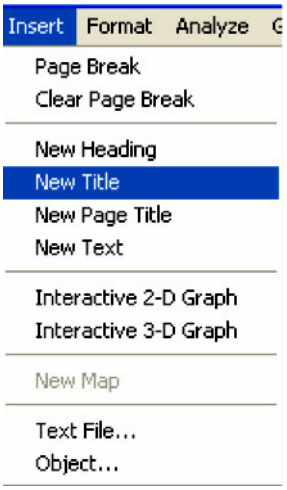
(11-4) إضافة عنوان جديد Add New Title:

مثال (11–2):

في المثال (9–1) نفرض أننا نريد إضافة عنوان أسفل العنوان Regression بعنوان جديد يسمى «نموذج الانحدار المتعدد» فإننا نتبع الخطوات الآتية:

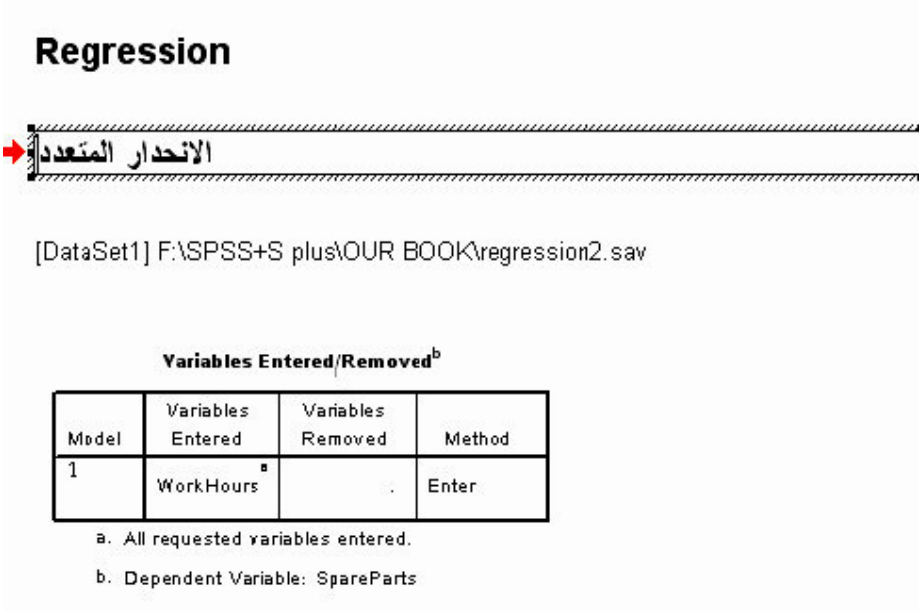
انقر على الشكل المراد إضافة عنوان أسفله حتى يضاء وهنا في المثال وهو العنوان Regression.

من القائمة Insert انقر على الأمر New Title نتيجة تنفيذ الأمر سينشأ البرنامج مستطيلا أسفل الشكل كما في الشكل الآتي:



شكل (11 - 14)

- انقر نقرًا مزدوجًا على المستطيل الجديد.
- اكتب العنوان المراد إدخاله للصفحة في المستطيل وهو «الانحدار المتعدد» فيظهر الشكل الآتي:



Regression

الانحدار المتعدد

[DataSet1] F:\SPSS+S plus\OUR BOOK\regression2.sav

Variables Entered/Removed[b]

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

a. All requested variables entered.

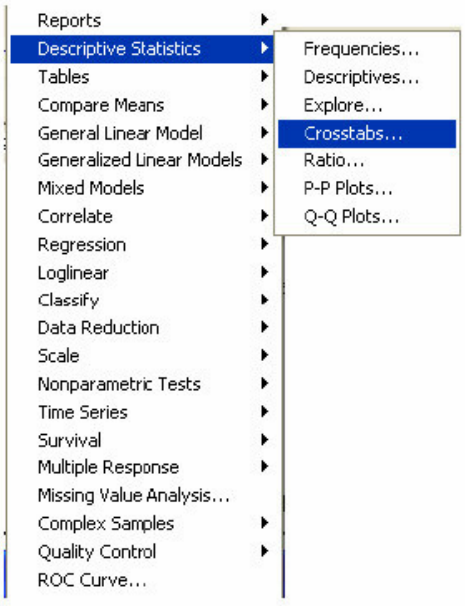
b. Dependent Variable: SpareParts

شكل (11 - 15)

- يمكن إزالة العنوان القديم وتحريك العنوان الجديد في مكان العنوان القديم.

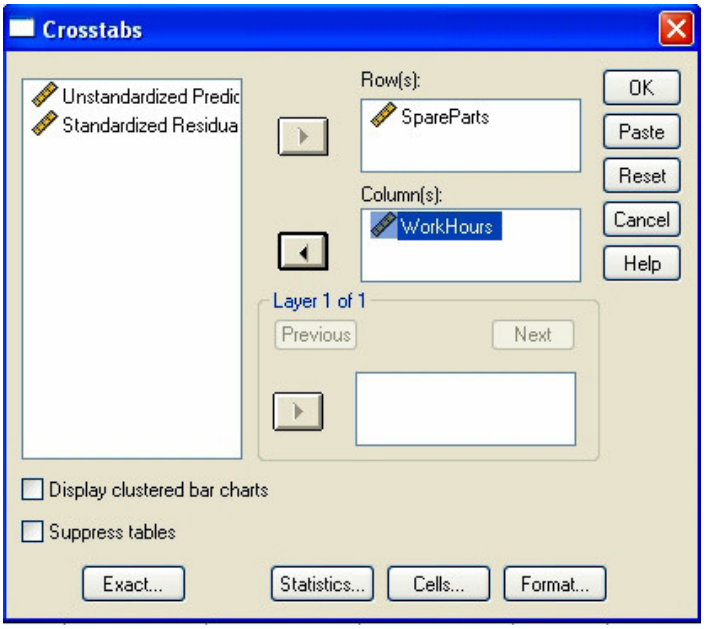
(11-5) محرر الجداول Pivot Tables :

عنـد فتح صفحة البيانات المثال (11-2) ثم من القائمـة Analyze ثم Descriptive Statistics والأمر Cross Tab من الشكل الآتي:



شكل (11 - 16)

بعـد ذلك يظهر مربع الحوار الآتي، ونقوم بإدخـال المتغير الذي نريد أن يظهر في الصف ونفرض أنه SpareParts ونفرض أن المتغير المراد إدخاله في الأعمدة هو WorkHours:



شكل (11 - 17)

فنحصل على جدول التوزيع المزدوج نتيجة تنفيذ الأمر الآتي:

جدول به المتاح والمفقود من البيانات.

جدول (11 - 1)

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
SpareParts * WorkHour	10	100.0%	0	.0%	10	100.0%

ب- جدول التوزيع الثنائي المطلوب.

Spareparts* WarkHours Crosstabulation

Count

		WorkHours										Total
		50.00	69.00	73.00	87.00	108.00	128.00	132.00	135.00	148.00	170.00	
SpareParts	20.00	1	0	0	0	0	0	0	0	0	0	1
	30.00	0	1	1	0	0	0	0	0	0	0	2
	40.00	0	0	0	1	0	0	0	0	0	0	1
	50.00	0	0	0	0	1	0	0	0	0	0	1
	60.00	0	0	0	0	0	1	1	1	0	0	3
	70.00	0	0	0	0	0	0	0	0	1	0	1
	80.00	0	0	0	0	0	0	0	0	0	1	1
Total		1	1	1	1	1	1	1	1	1	1	10

لتعديل جدول معين أو بمعنى آخر لإعداد الجدول ليستجيب إلى أي تعديلات يطلبها المستخدم نمر بالخطوات الآتية:

1 - ننقر نقرًا مزدوجًا على عنوان الجدول المراد التعديل فيه، وهو (SpareParts * WorkHours Crosstabulation) .

هذا النقر سيجعل الجدول نشطًا (انظر الشكل 11-18):

Spareparts* WorkHours Crosstabulation

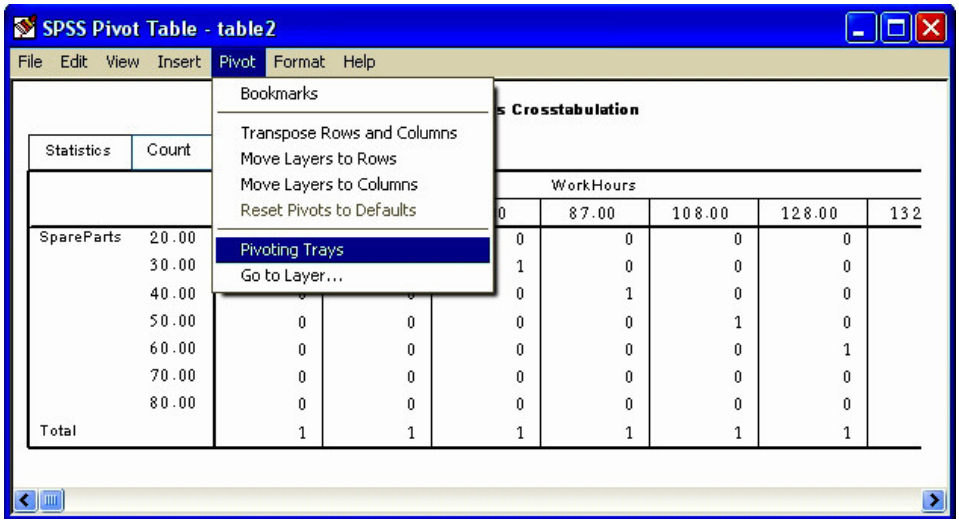
Count

		WorkHours										Total
		50.00	69.00	73.00	87.00	108.00	128.00	132.00	135.00	148.00	170.00	
SpareParts	20.00	1	0	0	0	0	0	0	0	0	0	1
	30.00	0	1	1	0	0	0	0	0	0	0	2
	40.00	0	0	0	1	0	0	0	0	0	0	1
	50.00	0	0	0	0	1	0	0	0	0	0	1
	60.00	0	0	0	0	0	1	1	1	0	0	3
	70.00	0	0	0	0	0	0	0	0	1	0	1
	80.00	0	0	0	0	0	0	0	0	0	1	1
Total		1	1	1	1	1	1	1	1	1	1	10

شكل (11 - 18)

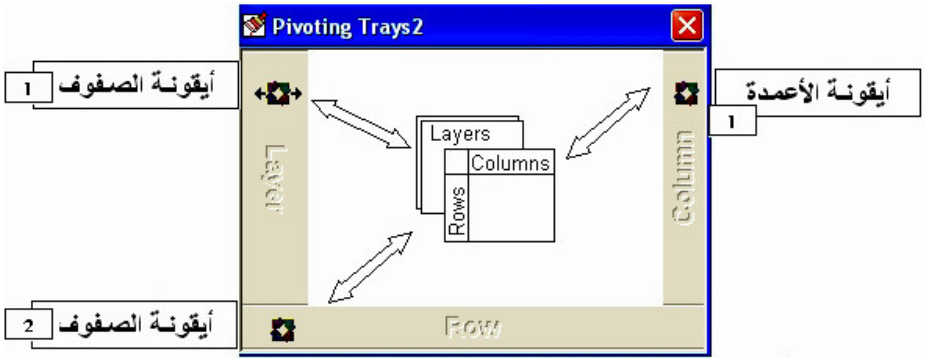
أ-لتبديل الصفوف بالأعمدة To Transpose Rows And Columns:

بعد تنشيط الجدول ومن القائمة Pivot انقر الأمر Pivot Trays كما يلي:



شكل (11 - 19)

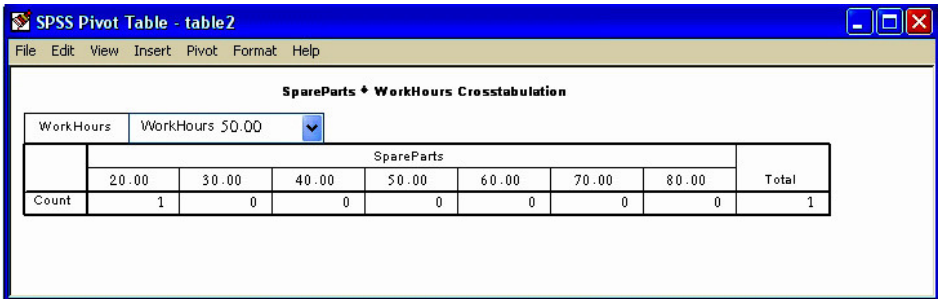
لفتح الصندوق الخاص بذلك ويسمى Pivot Trays1 (انظر الشكل 11-20):



شكل (11 - 20)

يوجد في الصندوق الحواري Pivot Trays 1 ثلاث أيقونات، اثنان على يسار الصندوق بجوار كلمة Row وواحدة بجوار كلمة Column، الأيقونة التي على اليمن تمثل المتغير (SpareParts)، والأيقونات الأفقية أحدهما تمثل WorkHours.

2 - بالفأرة اسحب الأيقونة الرأسية بجوار كلمة Column إلى الصف الأفقي، واسحب أيقونة الصفوف الأفقية بجوار Row إلى الأعمدة.

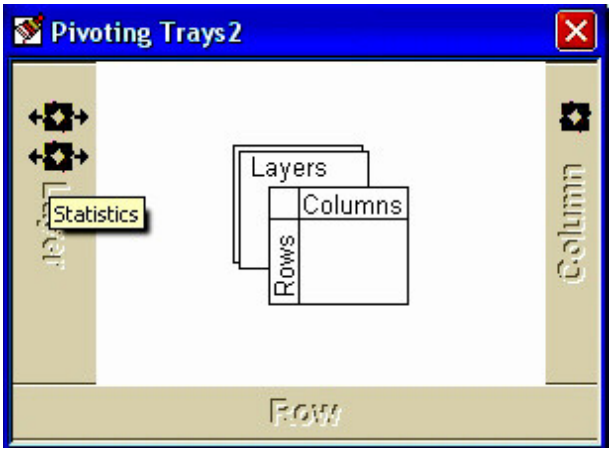


شكل (11 - 21)

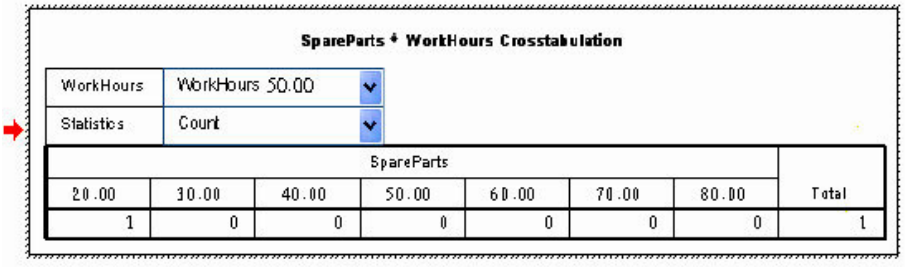
ب- لتغيير ترتيب العرض في الجدول المحور To Change The Display Order In The Pivot Table:

تتم هذه العملية كالآتي:

- تأكد أن الجدول المحور تم تنشيطه وظهر على الشاشة الصندوق Pivoting Tray1.
- في الصندوق الحواري Pivoting Tray1 انقر الأيقونة Statistics واسحبها إلى يمين شريط الصف ستلاحظ أن ترتيب العناصر داخل الجدول تغير وأصبح مفهوما كالآتي (انظر الشكل11-22):



شكل (11 - 22)



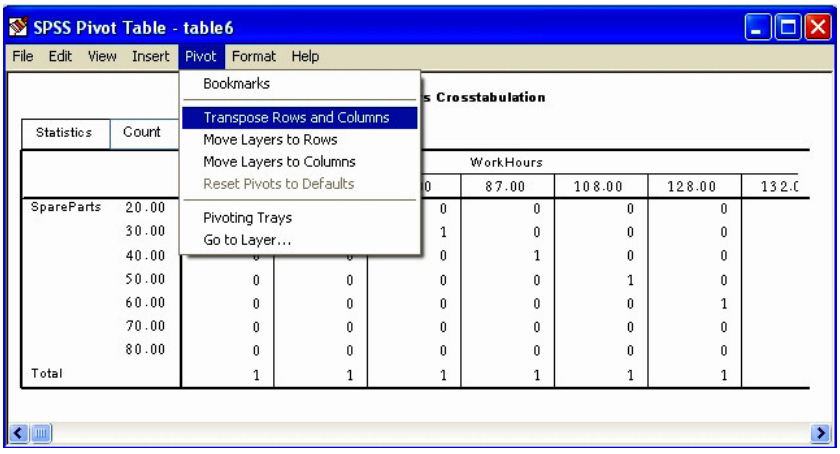
شكل (11 - 23)

في الشكل (11-23) نلاحظ أن الجدول يكون تحت فئة واحدة وهي WorkHours 50.00، وبالنقر على السهم تظهر الفئات جميعًا وباختيارها يظهر SpareParts بفئة WorkHours الجديدة.

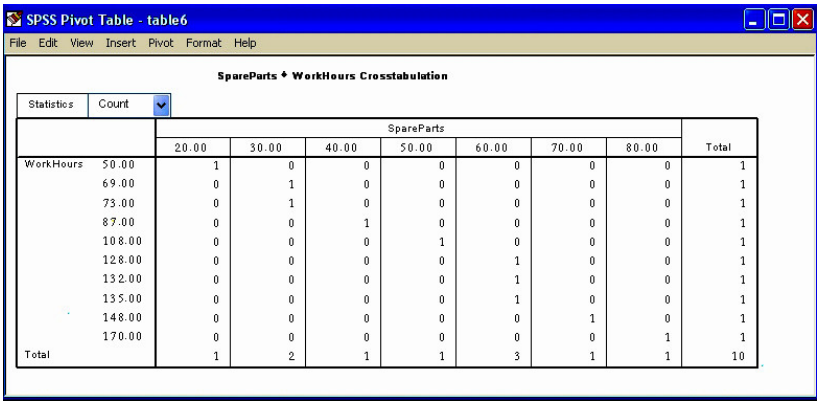
جـ- لتحريك صف أو عمود To Move Rows & Column:

لتحريك صف أو عمود في الجدول تتم العملية كالآتي:

- نقوم بتنشيط شكل رقم (11-21) بالخطوات السابقة نفسها أي بالنقر على الجدول المراد نقل أعمدته وتبديلها.
- من Pivot نختار Transpose Rows and Colums كما في الشكل (11-24)، ومن ثم يتم تحويل الصف عمود والعمود إلى صف كما في الشكل (11-25).



شكل (11 - 24)

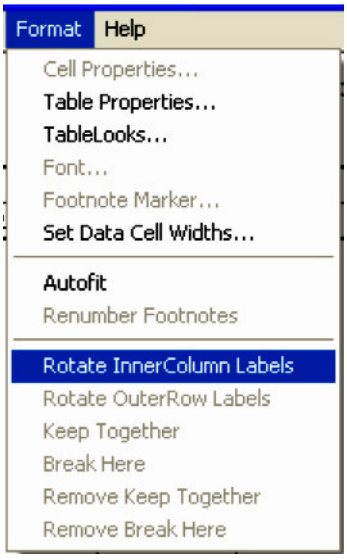


شكل (11 - 25)

د- لتدوير مميز صف أو عمود To Rotate Row Or Column Label:

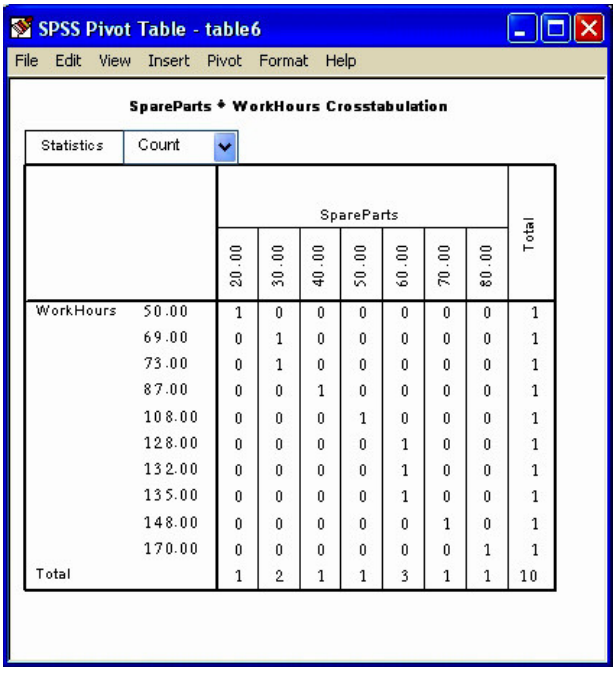
- نشط الجدول المحور.

- من القائمة Format انقر الأمر Rotate Inner- Column Labels كما في الشكل الآتي:



شكل (11 ـ 26)

- ستلاحظ أن مميز الأعمدة تم تدويره ويمكن استخدام الأمر نفسه للصفوف انظر الجدول الآتي.



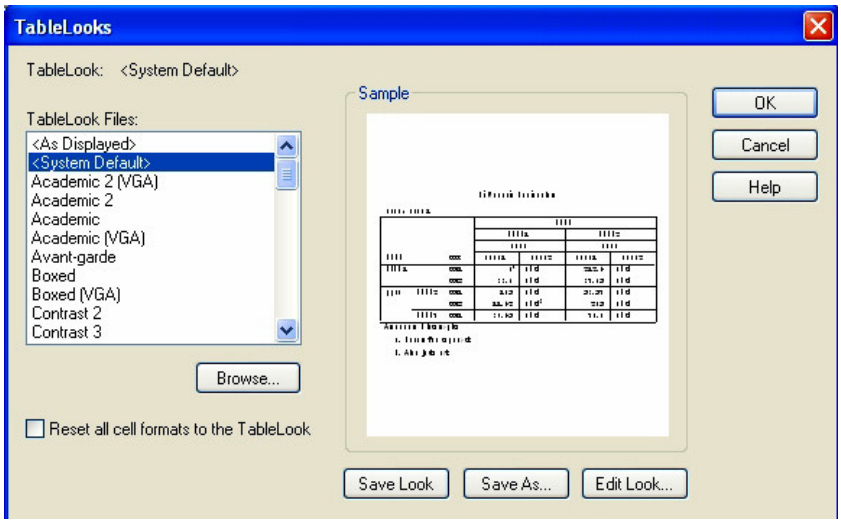
شكل (11 ـ 27)

هـ – تغيير مظهر الجدول Change The Appearance Of Tables:

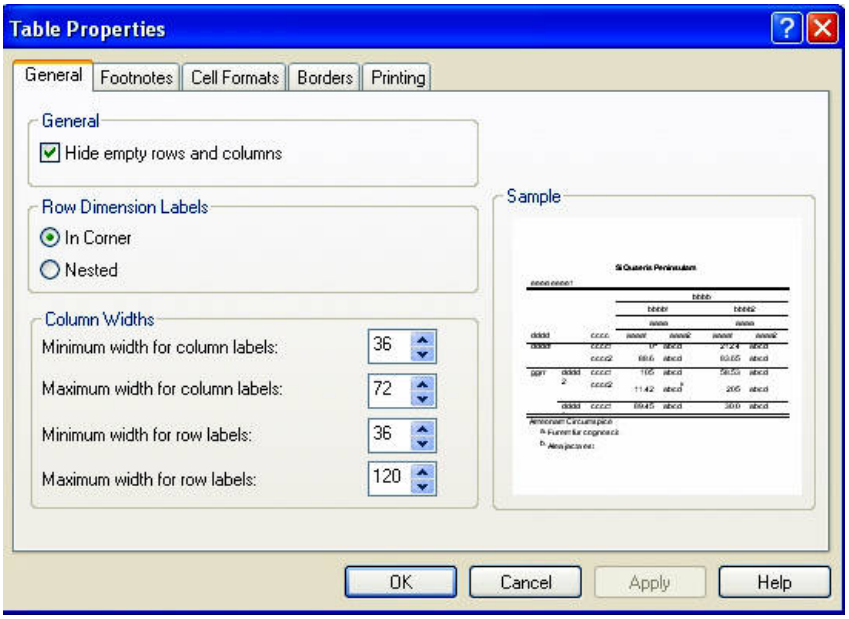
يمكن تغيير مظهر الجدول بتطبيق النظر للجدول Table Look، ولتطبيق النظر للجدول نجري الآتي:

- نشط الجدول المحور بالطرق المعروفة.
- من القائمة Format انقر الأمر Table Look.

- اختر المظهر الملائم الذي ترغب فيه من قائمة الملفات.
- لتطبيق الشكل المختار على الجدول انقر الأمر Ok انظر الشكل (11-28).
- **ملاحظة:**
- يمكـن تغيير خصائص الجدول خلال قائمة Format باستخـدام الخيارات Table Properties، ومن ثم تغير خصائص Cell والحدود والتظليل وأشياء أخرى كثيرة مرتبطة بالتنسيق (انظر الشكل 11-29).



شكل (11 - 28)



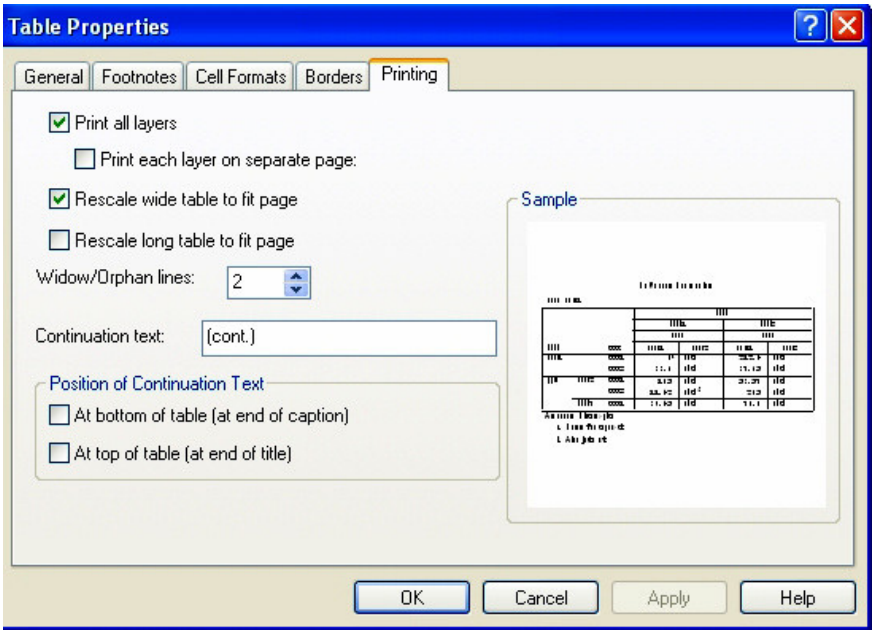
شكل (11 - 29)

و- طباعة الجداول المحور Printing Pivot Table:

لطباعة تقسيم مخفي من الجدول وتحجيم الجدول ليلائم حجم الصفحة نمر بالخطوات الآتية:

نشط الجدول بالطرق المعروفة.

- من القائمة Format انقر الأمر Table Properties لفتح الصندوق Table Properties الخاص بالعملية الموجودة بالصفحة الآتية.
- انقر على الأمر Printing فيظهر مربع حوار.
- بالفأرة اختر الصندوق Print All Layers وستلاحظ أن هناك صندوقاً، يطلب هل الطباعة تتم كل تقسيم في صفحة مستقلة.
- لتحجيم الجدول انقر Rescale Wide Table To Fit Page.
- انقر على الأمر Ok للتنفيذ. انظر الشكل (11-30).



شكل (11-30)

(11-6) تعديل وتنقيح الأشكال البيانية Modifying And Enhancing Charts Presentation:

قد يرغب المستخدم في إجراء تعديلات على الشكل البياني بتغيير الألوان أو نوع الخط أو إضافة معلومات مهمة لبحثها مثل:

1 - إضافة عنوان آخر غير العنوان الغيابي.

2 - تعديل الأعمدة البيانية لتكون ثلاثية الأبعاد 3-D أو تغيير المسافات بين الأعمدة وخلافه.

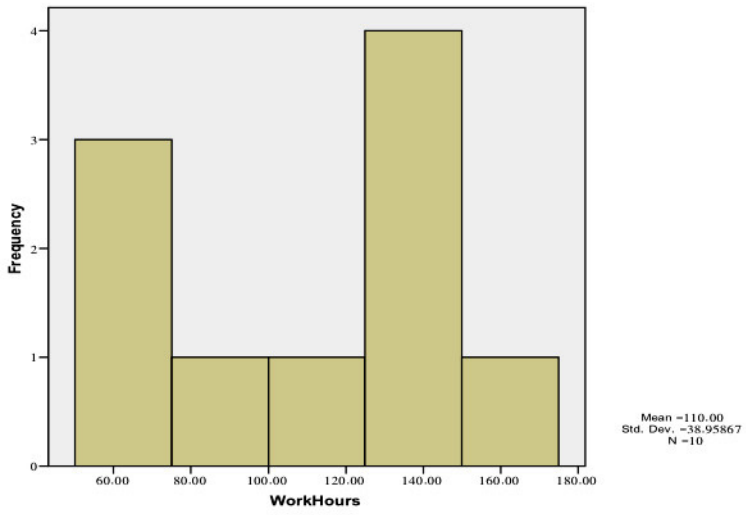
3 - حذف أو إضافة عناوين جديدة أو التغيير بطريقة تبرز خصائص الشكل البياني بطريقة أحسن.

إضافة مقاييس إحصائية.

نفرض أن لدينا العمود البياني من المثال (11-2) وهو عمود ساعات العمل:

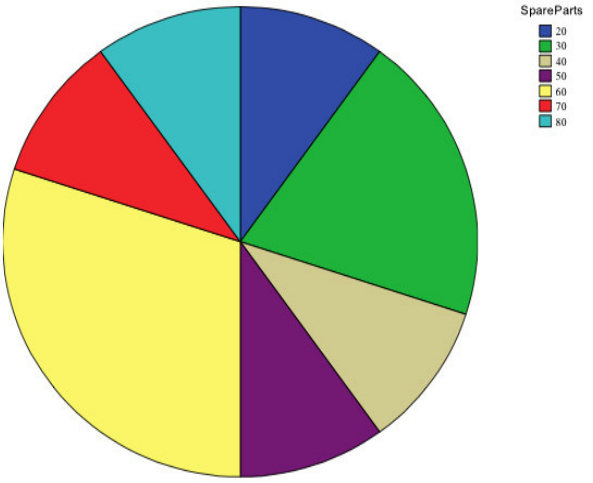
WrkHours
73.00
50.00
127.00
170.00
87.00
108.00
135.00
69.00
148.00
132.00

ونفترض أننا رسمنا المدرج التكراري لساعات العمل كما يلي ونلاحظ أننا رسمناه بطريقة رسم الأشكال نفسها. انظر الفصل الثاني:



شكل (11-31)

وتم الحصول أيضًا على الدائرة التي تمثل المتغير (SpareParts)، ونلاحظ أننا غيرنا نوع المتغير إلى ترتيبي ordinal لإنشاء دائرة منها وهي كالآتي (انظر شكل 11 - 32):

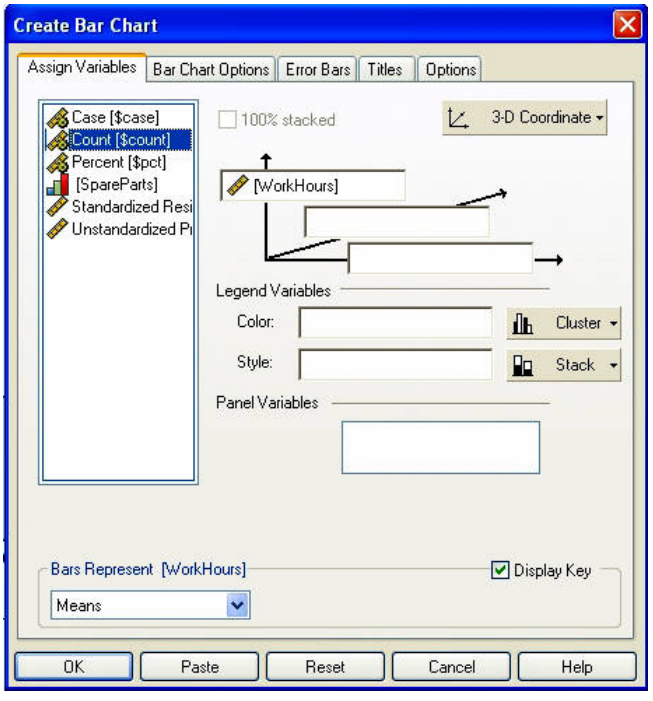


شكل (11-32)

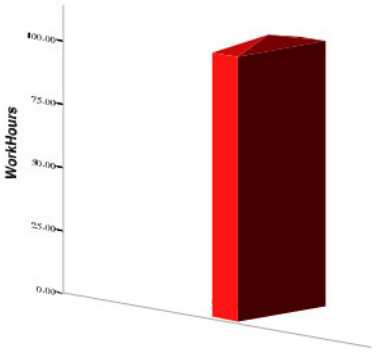
أ- تغيير أبعاد الأعمدة إلى البعد الثلاثي 3-D:

يمكن إضافة بعد ثالث للرسم من خلال محرر الشكل كالآتي:

- يتم الحصول على الشكل البياني المراد تعديله.
- انقر على الشكل المراد تعديله.
- من القائمة Graphs انقر على Interaction ثم Bar.
- انقر على الأمر 3-D Effect ستلاحظ أن أبعاد الأعمدة أصبحت ثلاثية انظر الصندوق الفرعي الآتي (انظر الشكل 11 – 33) وكذلك الأعمدة بعد التغيير. انظر الشكل(11-34):



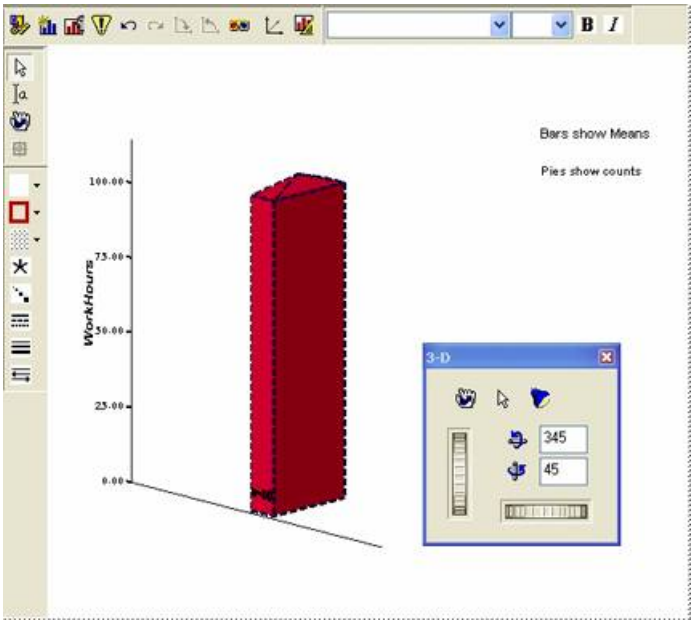
شكل (11 - 33)



شكل (11 ـ 34)

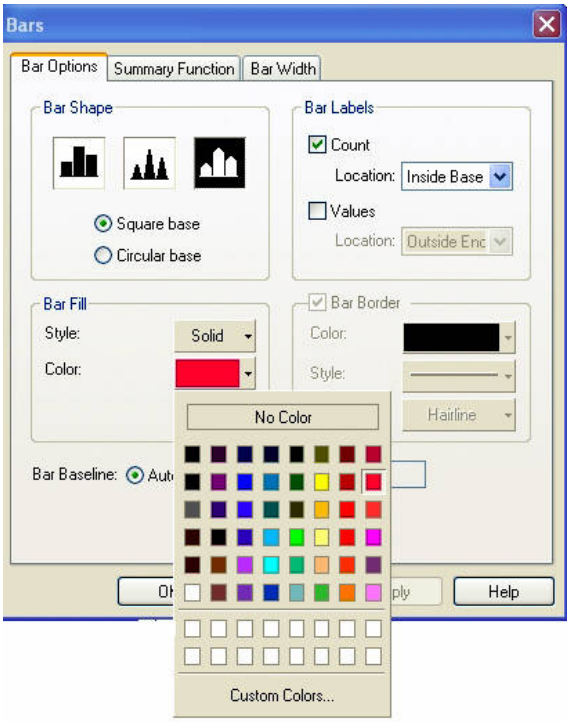
ب – تعديل خصائص الشكل Chart Attributes:

- انقر نقرًا مزدوجًا على أي مكان في الشكل سيفتح الصندوق الخاص بمحرر الشكل.
- عند النقر المزدوج على أي عنوان في الشكل سيفتح الصندوق الذي يسمح لك بضبط النص وإجراء مـا تريده من الشكل، يمكن للمستخدم تغيير العنـوان (يمينًا أو يسارًا أو في الوسط) وكذلك تغيير المميز. (انظر الشكل 11 – 35).

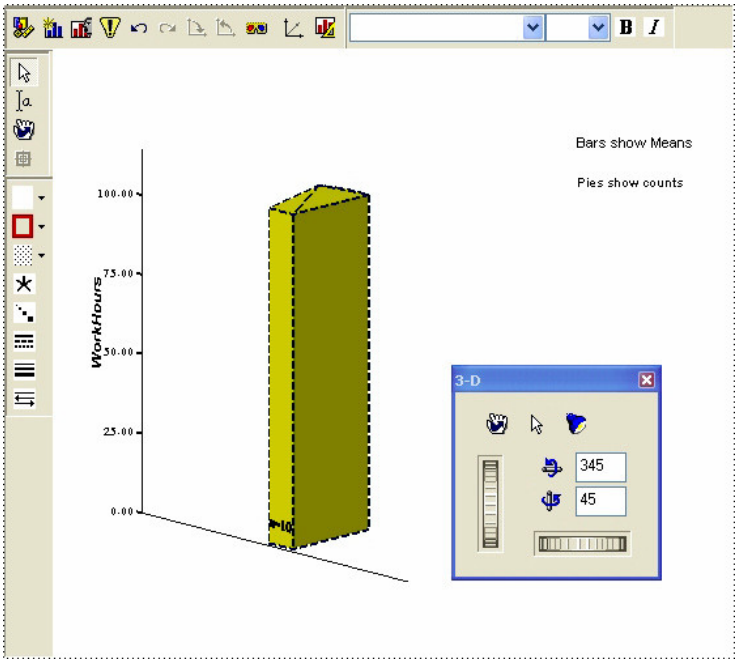


شكل (11 ـ 35)

- وننقر على أي خط أو رسمة في الشكل (11–35) مرتين متتاليتين يظهر مربع حوار خاص بخصائصه مـن تغير عرضه وتغير لونه وهكذا، فإذا افترضنا أننـا نريد لون تغير الشكل من اللون الأحمر إلى الأصفر، فإننا ننقر على الشكل الذي باللون الأحمر مرتين متتاليتين فيظهر الشكل الآتي:



شكل (11 - 36)



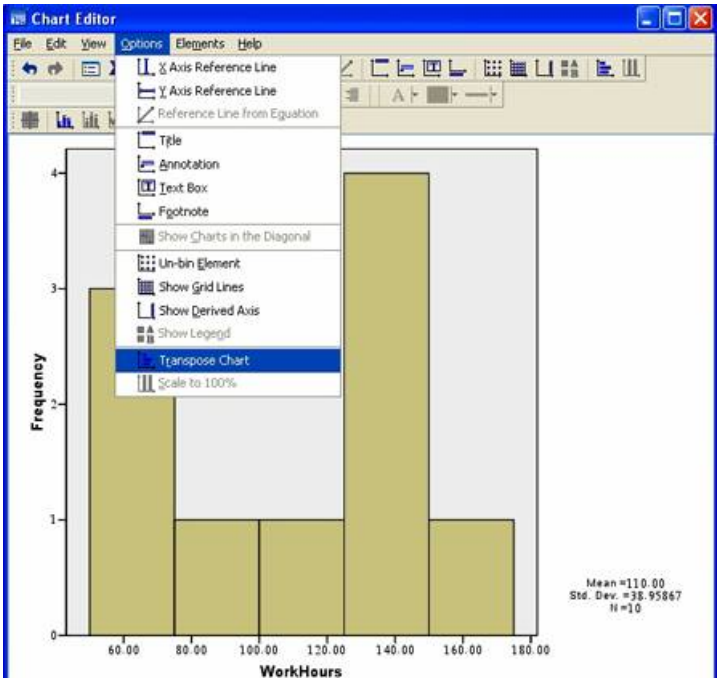
شكل (11 - 37)

ملاحظة :

- ✓ تذكـر دائمًـا عند تغيير مواصفات الشـكل أن تنقر على الأمر ok لتنفيذ الأمـر أو cancel لإغلاق الصندوق.

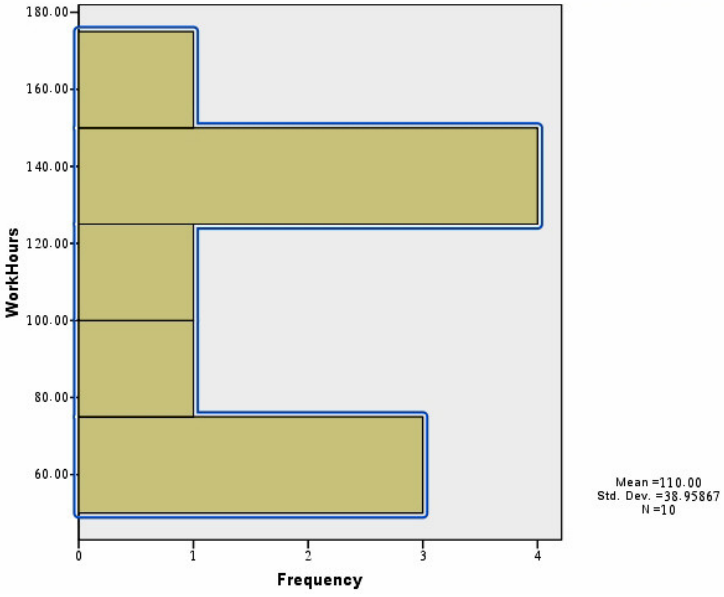
✓ د- لتبادل المتغيرات على الشكل To Transpose Variables In A Chart:

✓ ننقر على الشكل مرتين متتاليتين وسيظهر مربع حوار منه يتم اختيار option:



شكل (11 - 38)

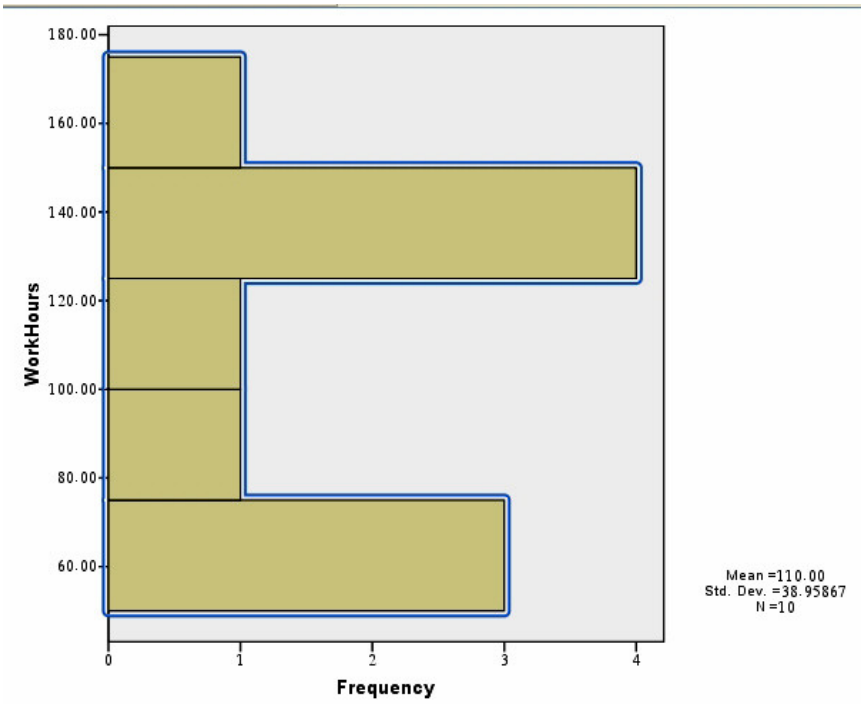
• مـن القائمـة Options انقر على الأمر Transpose فيصبح المتغير الموجـود على المحور الأفقي على المحور الرأسي والعكس. (انظر الشكل 11 – 39):



شكل (11 - 39)

هـ - لإدراج عنوان جديد للشكل To Insert A New Chart Title:

- من القائمة Options انقر على Title فيتم تنشيط عنوان أعلى الشكل لكي يتيح لنا كتابة عنوان به.
- يمكـن عن طريق لوحـة المفاتيح إضافة العنوان الجديد كما في الشكل الآتي إذا أضفنا عنوانًا على الشكل (11-19) بعنوان «المدرج التكراري» فيصبح الشكل (11-31) كالآتي:



المراجع

المراجع العربية:

- **الهلباوي، عبد الله توفيق، الإحصاء التطبيقي، مكتبة عين شمس**، 2000.
- **بشير، سعد زغلول**، دليلك إلى البرنامج الإحصائي SPSS، الإصدار العاشر، الجهاز المركزي للإحصاء جمهورية العراق.
- **عاشور، سمير كامل وسالم، سامية أبو الفتوح**، العرض والتحليل الإحصائي باستخدام SPSSWIN 2003.
- **هندي، محمود محمد إبراهيم وسلمان، خلف سلمان سلطان**، مفاهيم لطرق التحليل الإحصائي، مكتبة الرشد، الرياض، 1425.
- **سلطان، عبد الله علي حسن وحسين، علي الهفوفي**، أساسيات العرض والتحليل الإحصائي باستخدام SPSS-WIN (الجزء الأول). مشروع تخرج بجامعة الملك سعود، كلية العلوم، قسم الإحصاء وبحوث العمليات، 1427 هـ.
- **ثامر، محمد عباس منشي ودهام، ناصر دهام الدهام**، أساسيات العرض والتحليل الإحصائي باستخدام SPSS-WIN (الجزء الثاني)، مشروع تخرج بجامعة الملك سعود، كلية العلوم، قسم الإحصاء وبحوث العمليات، 1428 هـ.

المراجع الأجنبية:

- Hollander، M. & Wolf، D. A. (1998). Nonparametric Statistical Methods، Second Edition، Wiley، New York.
- Gupta، V. (1999). SPSS for Beginners. VJBooks Inc.