جامعة عـمار ثليجي الأغـواط

كلية العلوم الإنسانية والاسلامية والحضارة

قسم علوم الإعلام والاتصال

برمجيات تحليل البيانات الصحفية

SPSS

الدكتور/ جخدم موسى

مطبوعة موجهة لطلبة سنة ثانية إعلام واتصال

السداسي: الرابع

عنوان الوحدة : وحدة تعليم منهجية

المادة: برمجيات تحليل البيانات

أهداف التعليم

التعرف على المجال التطبيقي لتحليل البيانات الكمية والكيفية عن طريق برمجيات معلوماتية

المعارف المسبقة المطلوبة

مكتسبات مادة عرض وتحليل البيانات (سداسي.3)

محتوى المادة

التعريف ببعض البرمجيات المعلوماتية المحللة للبيانات الكمية والكيفية من عروض نظرية تطبيقية لنماذج منها: SPSS,

طريقة التقييم

مراقبة مستمرة، علامة الأعمال التطبيقية

المحتويات

تحليل البيانات (الإحصاء) 5
الإحصاء الوصفي: 9
الإحصاء الاستدلالي: 9
المجتمع Population: 10
الحصر الشامل Census: 10
المقاييس الإحصائية 10
أولاً: مقاييس النزعة المركزية Measures of Central Tendency 10
ثانياً: مقاييس التشتت المطلق Measures of Dispersion 11
ثالثاً: الالتواء Skewness 12
تشغيل والتعرف على البرنامج SPSS 13
نوافذ البرنامج 13
استرجاع البيانات والملفات: 15
حفظ الملف: 16
إضافة، تعديل والتحكم بالمتغيرات 16
إضافة متغير أو مشاهدة: 18
إلغاء متغير أو مشاهدة أو حالة 19
ترتيب المشاهدات حسب متغير معين Rank Cases 19
تكوين متغير جديد باستخدام معادلة 19
الأمر Compute 19
استخدام الدالة IF مع Compute 20
اختيار خلايا SELECT CASES 21
إعادة الترميز Recode 25
الإحصاء الوصفي والمدرج التكراري للبيانات 27
(1) التكرارات والمدرج التكراري Histogram and Frequencies 27
(2) الإحصاء الوصفي Descriptive Statistics 28
(3) المستكشف Explore 28
(4) جداول الاقتران CROSS TABULATION 29
الرسم البياني 31
اختبار الفرضيات Test of Hypotheses 41
1- الفرضية الإحصائية 44
2- مستوى المعنوية أو مستوى الاحتمال 46
3- دالة الاختبار الإحصائية 47
4- القيمة الاحتمالية: (Sig. or P-value) 48
خطوات اختبار الفرضيات: 50
أولاً: اختبار T في حالة اختبار فرضيات متعلقة بمتوسط واحد 51
البيانات التالية تمثل درجات عشرين طالباً في مساق ما: 52
المطلوب: اختبار الفرضية المبدئية القائلة بأن متوسط درجات الطلاب = 65 درجة. 53
ثانياً: اختبارات الفروق بين متوسطين مجتمعين مستقلين 55
ثالثاً: اختبارات الفروق بين متوسطي مجتمعين من عينات مرتبطة 58
تحليل التباين(ANOVA) Analysis of Variance 59
أولاً: تحليل التباين الأحادي One-Way ANOVA 60
ثانياً: تحليل التباين الثنائي Two-Way ANOVA 63
ثالثاً: تحليل التباين الثلاثي Three-Way ANOVA 65
الاختبارات غير المعلمية Nonparametric Tests 66
مزايا استخدام الاختبارات غيرالمعلمية: 67
عيوب استخدام الاختبارات غيرالمعلمية: 69
اختبار التوزيع الطبيعي 70

اختبار ويلكوكسن "Wilcoxon Test" 61
اختبار مان – وتني "Mann Whitney Test" 71
اختبار كروسكال – والاس "Kruskal-Wallis Test" 73
اختبار فريدمان "Friedman Test" 74
الارتباط الخطي البسيط Simple Linear Regression 75
معامل الارتباط Correlation Coefficient: 75
حساب قيمة معامل الارتباط: 75
الانحدار الخطي البسيط Simple Linear Regression 77

تحليل البيانات (الإحصاء)

قديما كان يعرف الإحصاء بأنه هو العلم الذي يهتم بأساليب جمع البيانات وتنظيمها في جداول إحصائية ثم عرضها بيانياً. ومع تطور هذا العلم في العصر الحديث يمكن تعريفه تعريفاً شاملاً بأنه العلم الذي يبحث في:

– جمع البيانات والحقائق المتعلقة بمختلف الظواهر وتسجيلها في صورة رقمية وتصنيفها وعرضها في جداول منظمة وتمثيلها بيانياً، وإيجاد المقاييس الإحصائية المناسبة.

– مقارنة الظواهر المختلفة ودراسة العلاقات والاتجاهات بينها واستخدامها في فهم حقيقة تلك الظواهر ومعرفة القوانين التي تسير تبعاً لها.

– تحليل البيانات واستخراج النتائج منها ثم اتخاذ القرارات المناسبة.

وينقسم علم الإحصاء إلى قسمين أساسيين هما:

الإحصاء الوصفي:

عبارة من مجموعة الأساليب الإحصائية التي تعنى بجمع البيانات وتنظيمها وتصنيفها وتلخيصها وعرضها بطريقة واضحة في صورة جداول أو أشكال بيانية وحساب المقاييس الإحصائية المختلفة لوصف متغير ما (أو أكثر من متغير) في مجتمع ما أو عينه منه.

الإحصاء الاستدلالي:

عبارة عن مجموعة من الأساليب الإحصائية التي تستخدم بغرض تحليل بيانات ظاهرة (أو أكثر) في مجتمع ما على أساس بيانات عينة احتمالية تسحب منه وتفسيرها للتوصل إلى التنبؤ واتخاذ القرارات المناسبة.

ويتلخص الأسلوب الإحصائي في الخطوات التالية:

1– جمع البيانات عن طريق التجربة والمشاهدة بوفرة كافية لاستخلاص النتائج منها.

2– عرض هذه البيانات بطريقة تساعد على تفهمها والاستفادة منها حيث أن البيانات الإحصائية في صورتها الأولية لا يمكن الاستفادة أو استخلاص النتائج منها وذلك في حالة وجود عدد كبير من الأرقام أو الصفات.

المجتمع:

هو مجموع كل المفردات الممكنة سواء كانت أفراداً أو أشياء أو وحدات تجريبية أو قياسات موضوع الاهتمام في الدراسة، وقد يتكون المجتمع من عدد محدود من المفردات أو أن يكون عدد مفرداته لا نهائي، كما أن المجتمع قد يكون حقيقيا أو افتراضيا.

الحصر الشامل:

هو جمع البيانات من جميع مفردات المجتمع المراد دراسته.

وفي بعض الحالات لا نتمكن من حصر كل مفردات المجتمع مثل مجتمعات الأسماك أو النباتات أو تؤدى عملية الحصول على البيانات من مفردات المجتمع إلى إهلاكها أو إتلافها وبالتالي لا يمكن جمع البيانات من كل المفردات أو قد تحتاج عملية جمع البيانات من جميع المفردات إلى وقت طويل أو جهد أو تكاليف باهظة، وفي مثل هذه الحالات يتم جمع البيانات بأخذ جزء فقط من مفردات المجتمع وهو ما يسمى بالعينة.

المقاييس الإحصائية

أولاً: مقاييس النزعة المركزية :

معظم قيم مفردات أي ظاهرة لها الرغبة في التجمع أو التمركز حول قيمة معينة تسمى القيمة المتوسطة، هذا التجمع عند هذه القيمة يسمى بالنزعة المركزية للبيانات.

أهم مقاييس النزعة المركزية:

الوسط الحسابي، الوسيط ، المنوال ،الرُبيعات، الوسط الهندسي،الوسط التوافقي.

(1) الوسط الحسابي:

الوسط الحسابي لمجموعة من القيم هو القيمة التي لو أعطيت لكل مفردة من مفردات المجموعة لكان مجموع القيم الجديدة مساويا لمجموع القيم الأصلية ويرمز له بالرمز $\bar{x}$. وستخدم الوسط الحسابي في حالة البيانات الرقمية فقط.

(2) الوسيط *Median*:

يعرف الوسيط لمجموعة من البيانات بأنه القيمة التي تقع في وسط المجموعة تماماً بعد ترتيبها تصاعدياً أو تنازلياً، أي هو القيمة التي تقسم مجموعة البيانات إلى قسمين بحيث يكون عدد القيم الأكبر منها مساوياً عدد القيم الأصغر منها ويرمز له بالرمز M_e. ويستخدم الوسيط في حالة البيانات الترتيبية.

(3) المنوال *Mode*:

يعرف المنوال لمجموعة من البيانات بأنه القيمة الأكثر شيوعاً (تكراراً) في المجموعة ويرمز له بالرمز M_o. يفضل استخدام المنوال في حالة البيانات الوصفية والترتيبية.

4- الرُبيعات *Quartiles*

يمكن تقسم المساحة تحت المضلع التكراري إلى أربعة أقسام متساوية تسمى الرُبيعات وعددها ثلاثة هي من اليسار إلى اليمين:

الرُبيع الأول (الأدنى) $\mathbf{Q_1}$: وهو القيمة التي تقسم مجموعة القراءات (بعد ترتيبها تصاعدياً) إلى قسمين بحيث يسبقها ربع البيانات ويليها ثلاثة أرباع البيانات.

الرُبيع الثاني (الوسيط) $\mathbf{Q_2}$: وهو القيمة التي تقسم مجموعة القراءات (بعد ترتيبها تصاعدياً) إلى قسمين بحيث يسبقها نصف البيانات ويليها نصف البيانات أيضاً.

الرُبيع الثالث (الأعلى) $\mathbf{Q_3}$: وهو القيمة التي تقسم مجموعة القراءات (بعد ترتيبها تصاعدياً) إلى قسمين بحيث يسبقها ثلاثة أرباع البيانات ويليها ربع البيانات.

ثانياً: مقاييس التشتت المطلق Measures of Dispersion

من أهم مقاييس التشتت المطلق: المدى، نصف المدى الرُبيعي (الانحراف الرُبيعي)، الانحراف المتوسط ، التباين والانحراف المعياري.

(1) المدى *Range*:

المدى هو أبسط مقاييس التشتت المطلق ويُعرف بأنه الفرق بين أكبر وأصغر قيمة في مجموعة البيانات ويرمز له بالرمز *R*.

(2) نصف المدى الرُبيعي (الانحراف الربيعي) *Quartile Deviation*:

يمكن التخلص من العيب الذي يسببه المدى وهو تأثره بالقيم المتطرفة وذلك بأن نستبعد الرُبع الأول من القراءات والرُبع الأخير منها ويُحسب المدى للقراءات الباقية. وتستخدم نصف المسافة بين الرُبيعيين الأدنى والأعلى كمقياس للتشتت في حالة وجود قيم متطرفة ويسمى هذا المقياس بنصف المدى الرُبيعي أو الانحراف الرُبيعي

(3) التباين والانحراف المعياري:

يعتبر التباين من أهم مقاييس التشتت المطلق ويعرف تباين مجموعة من القيم بأنه متوسط مجموع مربعات انحرافات هذه القيم عن وسطها الحسابي وبذلك فإن وحدات التباين هي مربع وحدات البيانات الأصلية. فإذا كانت وحدات القراءات الأصلية بالدينار فتكون وحدات التباين (الدينار)^2 وهكذا، ويرمز له بالرمز S^2.

والانحراف المعياري لمجموعة من البيانات هو الجذر التربيعي الموجب للتباين، وبذلك فإن وحدات الانحراف المعياري هي نفس وحدات البيانات الأصلية ويرمز له الرمز S، وغالباً يفضل استخدام الانحراف المعياري لأن مقياس التشتت المطلق يجب أن يكون له نفس وحدات القراءات الأصلية وهو متحقق في حالة الانحراف المعياري.

ثالثاً: الالتواء Skewness

الالتواء هو بعد التوزيع عن التماثل، وقد يكون هذا التوزيع متماثلاً أو ملتوياً جهة اليمين أو ملتوياً جهة اليسار.

- ففي حالة التوزيعات المتماثلة فإن الوسط الحسابي = الوسيط = المنوال
- إذا كان التوزيع ملتوياً جهة اليمين فإن:

  الوسط الحسابي > الوسيط > المنوال

ويسمى توزيع موجب الالتواء وفيه يكون الطرف الأيمن للمنحنى أطول من الأيسر.

- إذا كان التوزيع ملتوياً جهة اليسار فإن:

  الوسط الحسابي < الوسيط < المنوال

ويسمى توزيع سالب الالتواء وفيه يكون الطرف الأيسر للمنحنى أطول من الأيمن.

تشغيل والتعرف على البرنامج SPSS

يعمل البرنامج الإحصائي SPSS في بيئة النوافذ، ويتم تشغيله باختيار الأمر START من اللائحة الرئيسية PROGRAMS وبعد ذلك حدد برنامج SPSS.

نوافذ البرنامج

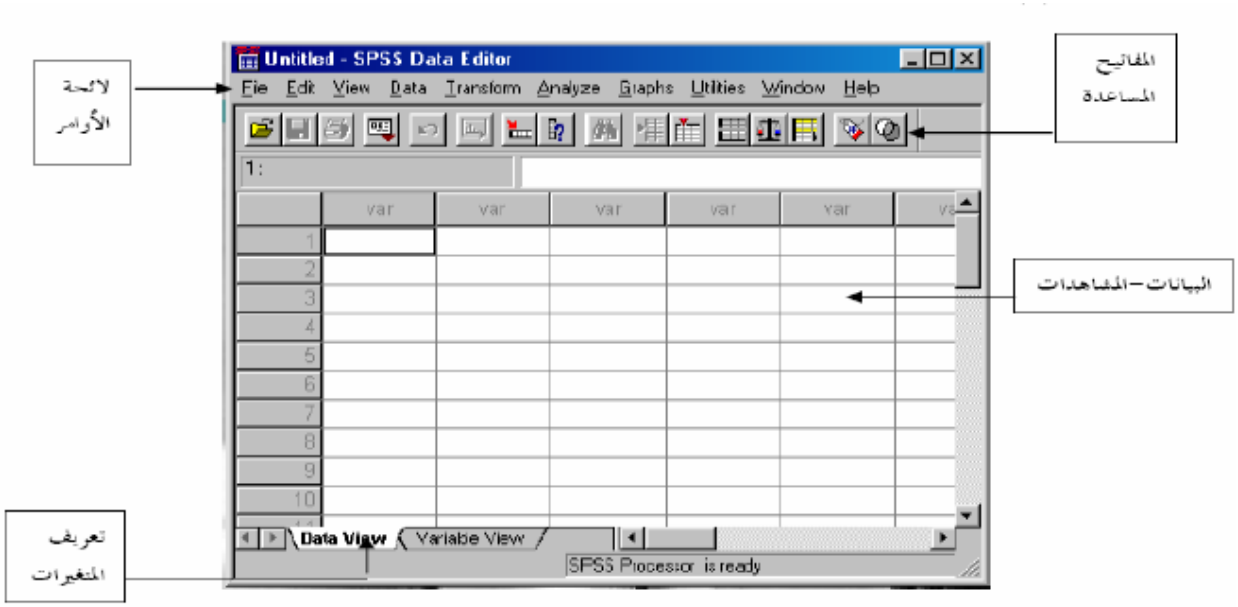
هناك عدة نوافذ للبرنامج نذكر منها ما يلي:

1- لائحة الأوامر COMMAND FUNCTIONS.

2- شاشة البيانات DATA VIEW.

3- شاشة تعريف المتغيرات VARIABLE VIEW.

4- لائحة التقارير والمخرجات OUTPUT NAVIGATOR.



1- لائحة الأوامر

وهو الجزء الخاص بالأوامر، حيث يمكن اختيار الأمر من خلال ICON لكل عملية إحصائية وتعرض النتائج في لائحة التقارير، وتشمل اللائحة على 9 أوامر رئيسة (بدون Help) يتفرع منها عدد من الأوامر الفرعية.

2- لائحة البيانات

لإضافة وإلغاء البيانات التابعة لكل متغير، حيث يتم تمثيل المتغير بعمود Column ويعطي الاسم VAR مع رقم يبدأ من 1 حتى 100,000، أما الأسطر فتمثل عدد المشاهدات لكل متغير. ويتم التحويل ما بين المشاهدات والمتغيرات بالضغط على Data View و Variable View.

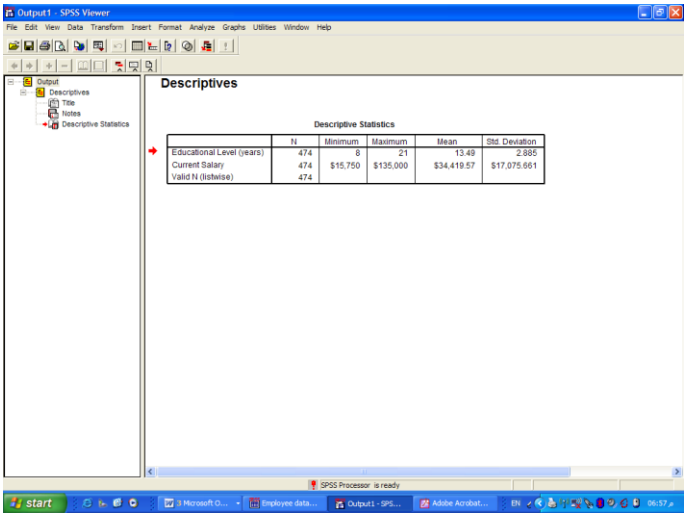
3- شاشة تعريف المتغيرات

لتعريف المتغيرات يتم الضغط على العمود مرتين DOUBLE CLICK او بالضغط على VARIABLE VIEW الموجود في أسفل الشاشة لتظهر شاشة أخرى لتعريف المتغيرات بتحديد اسم المتغير النوع، الحجم، العنوان، الترميز. ويتم الترميز بالضغط على عامود VALUES ومن ثم تحديد قيمة الرمز ووصفه مع الضغط على مفتاح ADD لإضافة الرمز.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align
1	id	Numeric	4	0	Employee Co	None	None	8	Right
2	gender	String	1	0	Gender	{f, Female}...	None	1	Left
3	bdate	Date	10	0	Date of Birth	None	None	13	Right
4	educ	Numeric	2	0	Educational L	{0, 0 (Missing	0	8	Right
5	jobcat	Numeric	1	0	Employment	{0, 0 (Missing	0	8	Right
6	salary	Dollar	8	0	Current Salar	{$0, missing}	$0	8	Right
7	salbegin	Dollar	8	0	Beginning Sal	{$0, missing}	$0	8	Right
8	jobtime	Numeric	2	0	Months since	{0, missing}..	0	8	Right
9	prevexp	Numeric	6	0	Previous Exp	{0, missing}..	None	8	Right
10	minority	Numeric	1	0	Minority Clas	{0, No}...	9	8	Right
11	new	Numeric	8	2	Salary Classif	{1.00, 15000	None	10	Right
12	educnew	String	8	0	Eductionalcle	None	None	10	Left
13									
14									
15									
16									

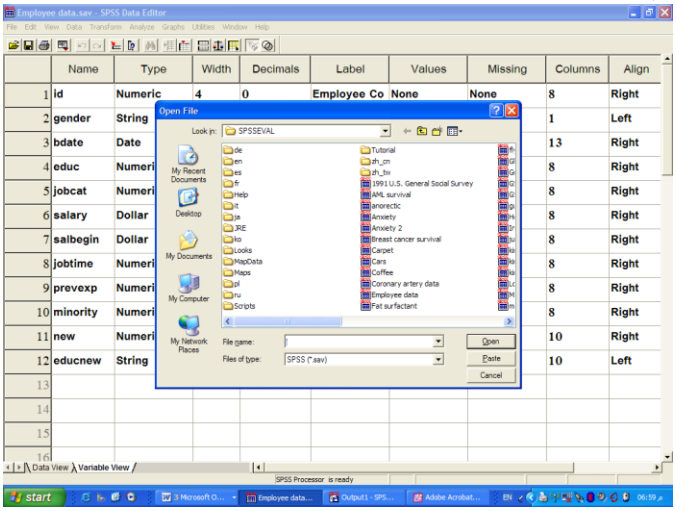
4- لائحة التقارير والنتائج:

شاشة لإظهار النتائج والتقارير، ويتم التحويل ما بين شاشة النتائج وشاشة البيانات بالضغط على الأمر WINDOW ومن ثم اختيار ملف البيانات.



استرجاع البيانات والملفات:

باختيار الأمر FILE ثم الفرعي OPEN، لا بد بعد ذلك من تحديد نوعية الملف المراد استرجاعه.



ويتم استرجاع التالي:

1- بيانات (المتغيرات) (*. SAV).

2- تقارير، والمقصود بتقارير نتائج العمليات الإحصائية التي تم عملها سابقاً (*.SPO).

وذلك بعد اختيار اسم الملف المطلوب مع التأكيد على مفتاح OPEN. وكذلك يمكن استرجاع ملفات الاكسيل (*.xls) وأنواع ملفات أخرى.

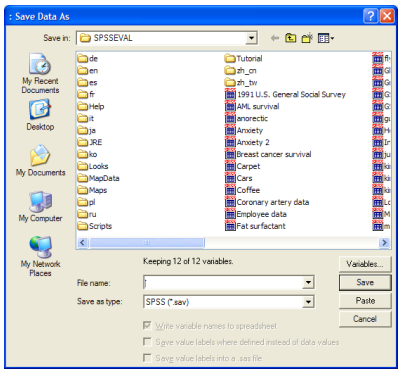
حفظ الملف:

الأمر الفرعي SAVE و SAVE AS خاصان لحفظ البيانات، حيث

1) SAVE AS يستخدم لإعطاء اسم جديد للملف مع حفظه ويمكن كما ذكر سابقاً حفظ ما يلي:

- بيان المتغيرات "DATA"
- تقارير "OUTPUT NAVIGATOR"

2) SAVE لحفظ التعديلات الجديدة التي طرأت على الملف.



إضافة، تعديل والتحكم بالمتغيرات

انتقل إلى نافذة DATA EDITOR واختر متغير غير محجوز (عمود) وأضف البيانات مع التأكيد على مفتاح ENTER أو تحرير السهم إلى أسفل (ملاحظة: . تعني MISSING أي لا توجد قيمة في هذه الخلية).

1) تعديل البيانات:

ويمكن بسهولة تعديل أي قيمة وذلك بتحريك السهم إلى الصف (الخلية) والكتابة عليها بالقيمة الجديدة.

2) تعريف المتغيرات:

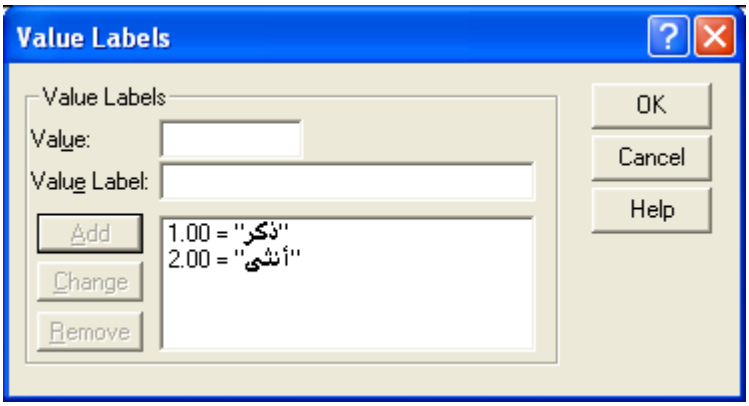
يمكن تحديد نوعية البيانات المضافة فالمتغيرات والمؤشرات الاقتصادية يمكن إضافتها كما هي، أما المتغيرات والبيانات تحدد من قبل الباحث بطريقة البدائل (ذكر أو أنثى، متعلم أو غير متعلم) ويتم تعريف المتغير بالانتقال إلى شاشة تعريف المتغيرات VARIABLE VIEW وتحديد الآتي:

- اسم المتغير، النوع، حجم المتغير، عدد النقاط العشرية.
- تحديد قيم المتغير (الترميز) في خانة VALUES.
- إدخال قيمة الرمز في خانة VALUE واسم الرمز في خانة VALUE LABEL والضغط على مفتاح ADD في كل مرة.
- بعد إجراء الخطوات السابقة يتم إضافة المتغيرات في شاشة البيانات ولإظهار القيم الكتابية المرادفة بدل القيم الرقمية وذلك بإجراء ما يلي:

I. اختر الأمر VIEW من اللائحة الرئيسة.

II. اختر الأمر الفرعي VALUE LABELS أو الضغط على المفتاح .

أنظر المربع الحواري التالي مثلاً:



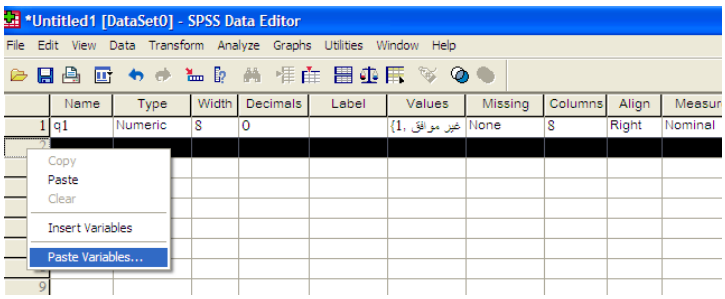
مثال:

في حالة وجود أكثر من متغير بنفس عناوين قيم البيانات ، وتكون الاختيارات: موافق بشدة، موافق، متردد، غير موافق، غير موافق على الإطلاق وبفرض أنه يوجد 10 متغيرات في مثل هذه الحالة، ولتنفيذ ذلك يمكن إتباع الخطوات التالية:

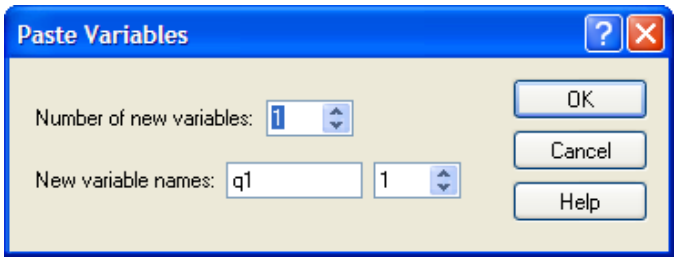
1- يتم تعريف الاختيارات السابقة كما تم شرحه في تعريف قيم المتغيرات.

2- نسخ المتغير السابق تعريفه، (EDIT, COPY) أو CTRL + C

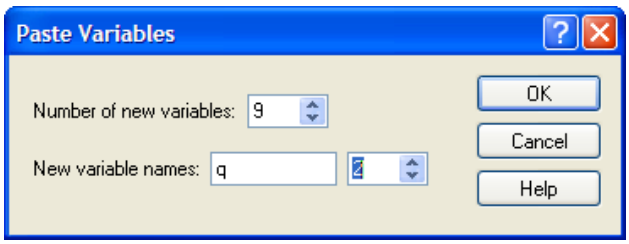
3- اختر الصف التالي للمتغير السابق بالفأرة ثم اضغط على المفتاح الأيمن للفأرة، من القائمة المنسدلة يتم اختيار PASTE VARIABLES... كما في الشكل التالي.



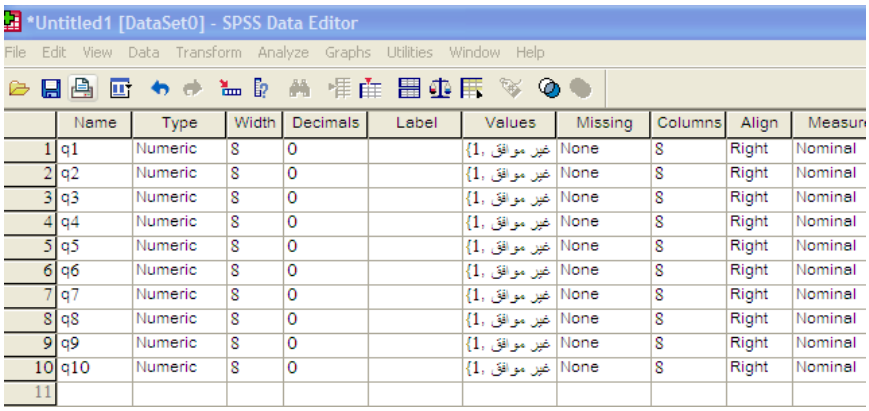
4- يظهر المربع الحواري التالي:



5- أكمل المربع الحواري السابق كما يلي:



6- اختر OK فنحصل على المطلوب كما في الشكل التالي:



إضافة متغير أو مشاهدة:

يمكن إضافة مشاهدة أو متغير جديد وذلك باستعمال الأمر الرئيسي DATA ثم:

1- الأمر الفرعي INSERT VARIABLE في حالة إضافة متغير جديد أو الضغط على مفتاح .

2- الأمر الفرعي INSERT CASE في حالة إضافة مشاهدة جديدة أو الضغط على مفتاح .

3- الأمر الفرعي SORT CASES لترتيب البيانات حسب المتغير المراد الترتيب به.

4- الأمر الفرعي GOTO CASE لتحويل المؤشر إلى مشاهدة معينة أو الضغط على مفتاح .

5- ولعرض المتغيرات المستخدمة قيد الدراسة يتم الضغط على مفتاح أو باستخدام الأمر الرئيسي UTILITIES ثم الأمر الفرعيVARIABLES .

إلغاء متغير أو مشاهدة أو حالة

ضع المؤشر في مكان المتغير المراد إلغاؤه ثم اضغط على مفتاح DEL، وفي حالة إلغاء مشاهدة ضع المؤشر على مكان الخلية (المشاهدة) ثم اضغط على مفتاح DEL. ولإلغاء حالة معينة يجب أن تضغط بالفأرة على تلك الحالة ثم اضغط على مفتاح DEL.

ترتيب المشاهدات حسب متغير معين Rank Cases

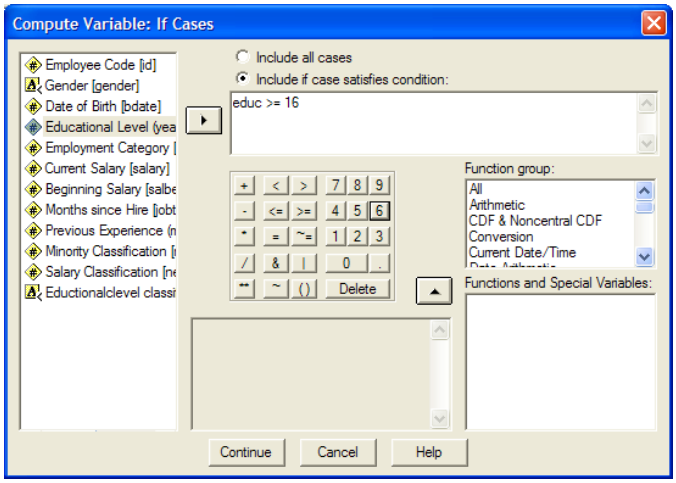
يقوم برنامج SPSS بانشاء متغير جديد يحتوي على الرقم التسلسلي لترتيب المشاهدات إما تصاعدياً أو تنازلياً، وذلك باختيار الأمر الفرعي RANK CASES من الأمر الرئيسي TRANSFORM.

تكوين متغير جديد باستخدام معادلة

الأمر Compute

أختر من اللائحة الرئيسة الأمر TRANSFORM، ثم الأمر الفرعي COMPUTE بعد ذلك حدد اسم المتغير الجديد في TARGET VARIABLE ثم كتابة المعادلة التي سوف تقوم بتكوينها باستخدام المتغيرات المعرفة مسبقاً. وبالضغط على مفتاح

لتحديد شرط تحقيق المعادلة. أنظر المربع الحواري التالي:



استخدام الدالة IF مع Compute

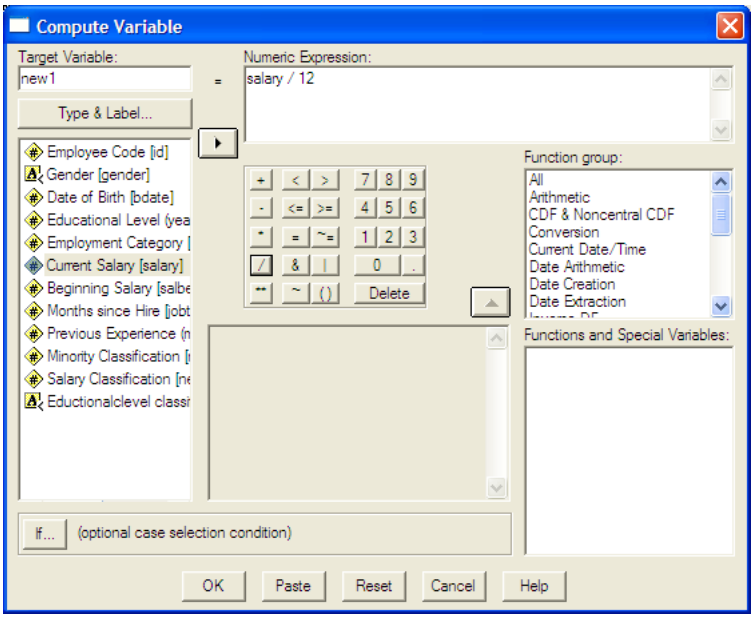
تستخدم الدالة IF في حالة إضافة شرط معين لحساب قيم متغير جديد بالنسبة لمتغير موجود مسبقاً

- فمثلاً: افتح الملف Employee Data.

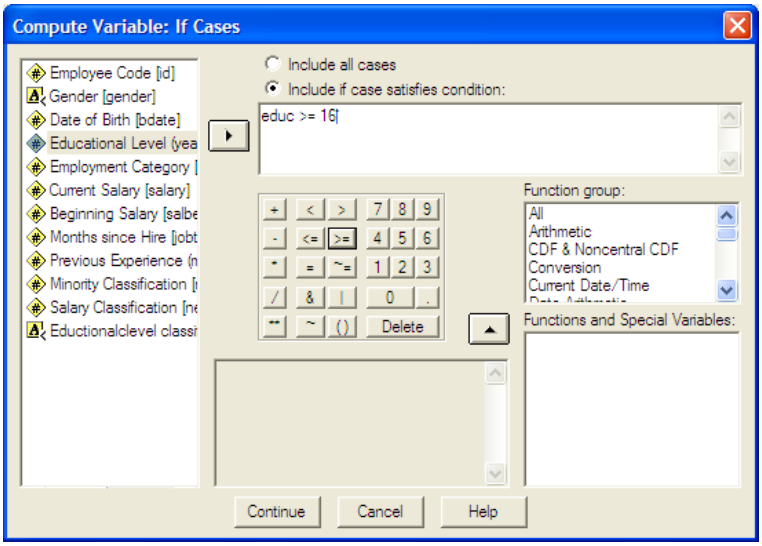
SPSS STEP BY STEP

Transform ⇒ Compute

- أكمل المربع الحواري كما يلي:



- اضغط على الاختيار If... ثم أكمل المربع الحواري كما يلي:



نلاحظ أنه تم إضافة متغير باسم new1 يشتمل على مكافأة شهر للموظفين الذين عدد سنوات تعليمهم 16 سنة فأكثر وخلايا مفقودة (بدون قيم) لباقي الموظفين.

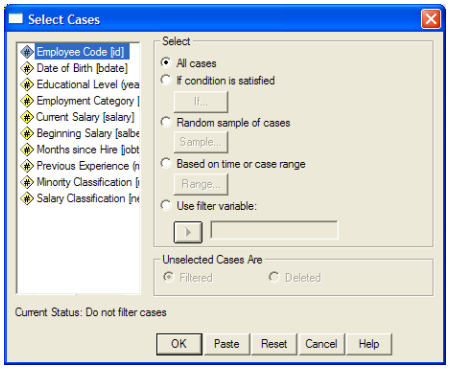
فمثلاً الموظف رقم 2: عدد سنوات التعليم الخاصة به 16 سنة وراتبه السنوي الحالي 40200$، نلاحظ أنه استحق مكافأة مقدارها 3350$ (40200/12=3350).

اختيار خلايا SELECT CASES

يستخدم هذه الأمر لاختيار الحالات التي تحقق شرط معين لاستخدامها في تحليل إحصائي خاص لبعض الحالات المطلوبة، فمثلاً إذا كان المطلوب اختيار الذكور الذين يعملون في وظيفة مدير أو اختيار عينة عشوائية ذات حجم معين.

SPSS STEP BY STEP

Data ⇒ Select Cases



توجد عدة اختيارات في المربع الحواري السابق هي:

1. All cases

يستخدم هذا الاختيار في حالة استخدام جميع الخلايا دون تحقيق شرط معين وهذا هو الاختيار المبدئي في SPSS.

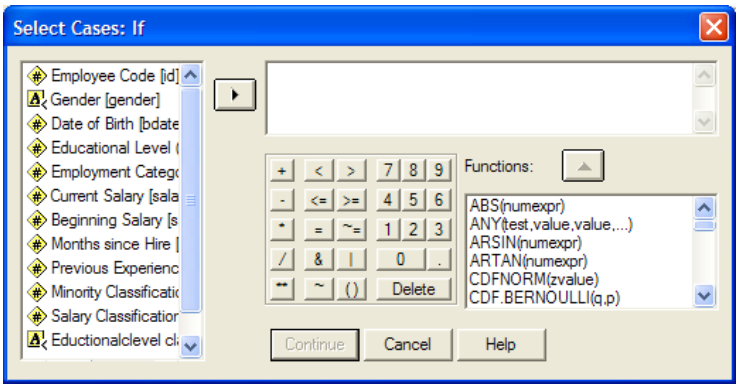
2. If condition is satisfied

يستخدم هذا الاختيار في حالة اختيار بعض الخلايا التي تحقق شرط معين، ويمكن استخدام الرموز التالية مع هذا الاختيار:

<	أصغر من	<=	أصغر من أو يساوي
>	أكبر من	>=	أكبر من أو يساوي
=	يساوي	~=	لا يساوي

يمكن استخدام الرموز المنطقية التالية مع الدالة If: " & "، and " | "، or

ولتنفيذ ذلك نشط هذا الاختيار ثم اضغط If فيظهر المربع الحواري التالي:



فمثلاً:

- لاختيار الحالات التي أقل من 18 سنة مثلاً لقيم المتغير educ نستخدم علامة أقل من " < " يمكن استخدام الشرط التالي:

educ < 18 أو educ <= 17

- لاختيار الموظفين بدون المدراء فقط يمكن استخدام العلامة لا يساوي " ~= " حيث تم تصنيف المدراء بالرقم 3 لتنفيذ ذلك استخدام الشرط التالي:

Jobcat ~= 3

- لاختيار الموظفين الذكور الذين تعلموا أكثر من 18 سنة ومدراء يمكن استخدام الشرط التالي:

Gender = "m" & educ >18 & jobcat = 3

علماً بأن المتغير Gender متغير وصفي تم تصنيفه إلى نوعين هما: m: ذكور، f: إناث، وفي حالة المتغير الوصفي يجب وضع الرمز المناسب (m, f) بين علامتي تنصيص " ".

- لاختيار الموظف الذي يعمل في وظيفة كاتب أو مدير يمكن استخدام الشرط التالي:

Jobcat = 1 | Jobcat = 3

مع ملاحظة أنه من الضروري تكرار اسم المتغير، أي أنه من الخطأ استخدام الشرط السابق على النحو التالي:

Jobcat = 1 | 3

يمكن استخدام دالة any لاختيار الموظف الذي يعمل في وظيفة كاتب أو مدير كما يلي:

any(Jobcat, 1 , 3)

- لاختيار الموظفين الذين تعلموا بين 18 سنة و 20 سنة مثلاً يمكن استخدام الشرط التالي:

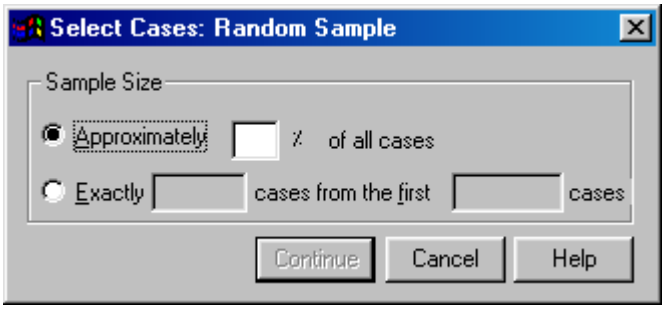
educ >=18 & educ <= 20

أو يمكن استخدام الشرط في الصورة التالية:

range (educ,18,20)

3. Random sample of cases

يستخدم هذا الاختيار في حالة اختيار عينة عشوائية بحجم معين، ولتنفيذ ذلك نشط هذا الاختيار ثم اضغط Sample فيظهر المربع الحواري التالي:



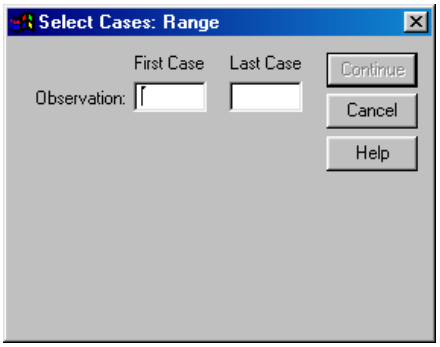
يوجد اختياران في المربع الحواري السابق هما:

Approximately: يستخدم لاختيار نسبة مئوية تقريبية من الحالات، فمثلاً يمكن اختيار 20% تقريباً من كل الخلايا.

Exactly: يستخدم لاختيار عينة عشوائية ذات حجم معين من أول عدد مناسب من الخلايا مع ملاحظة أن عدد الخلايا المطلوب اختيارها يجب أن يكون أقل من عدد الخلايا المطلوب الاختيار منها، فمثلاً يمكن اختيار 100 خلية فقط من أول 150 خلية.

4. Based on time or case range

يستخدم هذا الاختيار في حالة اختيار عينة عشوائية بحجم معين، ولتنفيذ ذلك نشط هذا الاختيار ثم اضغط Range



فيظهر المربع الحواري التالي:

لاختيار الحالات بين 20، 50 مثلاً اكتب في المربع الحواري السابق اكتب 20 في المستطيل أسفل First Case، 50 في المستطيل أسفل Last Case.

5. Use filter variable

يستخدم هذا الاختيار في حالة استخدام متغير رقمي كمتغير لتصفية الخلايا المطلوبة، وفي هذه الحالة فإن الخلايا التي قيمها لا تساوي صفراً أو ليست قيم مفقودة لمتغير التصفية سوف يتم اختيارها.

الاختيار **Filtered** أسفل Are: **Unselected Cases** يستخدم لتصفية الخلايا الغير مطلوبة مع إبقائها في ملف البيانات، أما الاختيار Deleted فيستخدم لمسح الخلايا الغير مطلوبة من ملف البيانات.

إعادة الترميز Recode

يستخدم الأمر Recode في عمليات الفرز لمجموعات مختلفة، وذلك بهدف إنشاء جداول تكرارية مختصرة ويمكن تنفيذ ذلك على نفس المتغير أو إنشاء متغير جديد وينصح بإنشاء متغير جديد لأن تنفيذ الأمر Recode على نفس المتغير يعمل على مسح قيم المتغير الأصلية التي قد تستخدم فيما بعد لأغراض تحليلية أخرى.

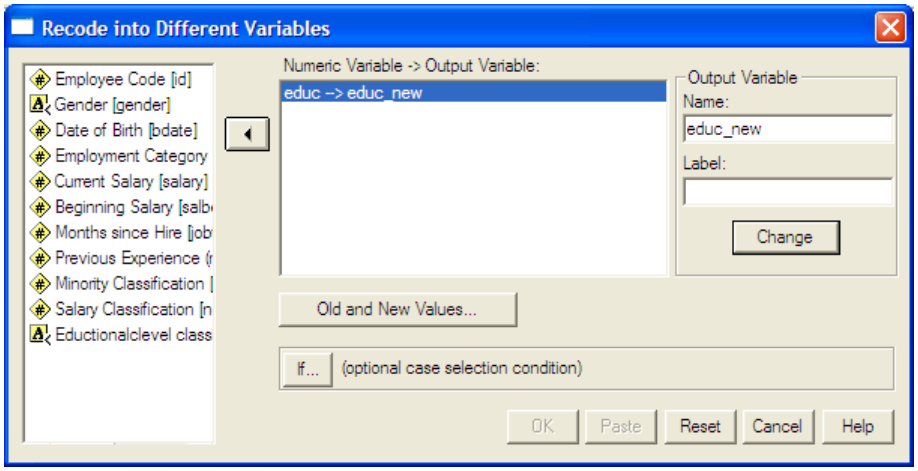
المطلوب: فرز عدد سنوات التعليم (educ) في ملف Employee data وذلك في متغير جديد باسم educ_new حسب التصنيف التالي:

مدى الدرجات	12-8	16-13	18-17	21-19
التصنيف	1	2	4	4

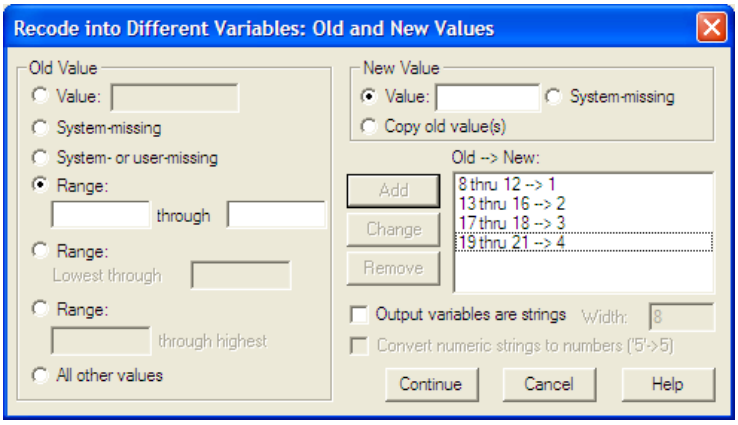
SPSS STEP BY STEP

Transform ⇒ Recode ⇒ Into Different Variables

- أكمل المربع الحواري كما يلي:



- اضغط على Old and New Values ثم أكمل المربع الحواري كما يلي:



- **المطلوب** تصنيف البيانات السابقة كما يلي:

مدى الدرجات	12-8	16-13	18-17	21-19
التصنيف	ثانوي فأقل	ليسانس	ماستر	دكتوراة

اتبع نفس الخطوات في المثال السابق مع اختيار Output variables are strings في المربع الحواري السابق مع استبدال التصنيف السابق (1،2،3،4) بالتصنيف الجديد (ثانوي فأقل، جامعي، ماجستير، دكتوراة) حيث أن التصنيف في هذه الحالة متغير وصفي.

الشكل التالي يمثل جزء من نافذة ملف البيانات بعد الانتهاء من تنفيذ الأمر.

ملاحظات:

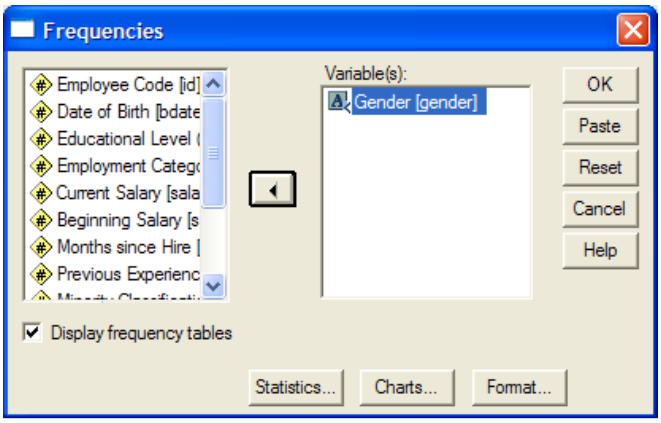
- يمكن فرز كلاً من المتغيرات الرقمية والوصفية بطريقة منفصلة، ولا يجوز فرزها معاً.
- في حالة اختيار عدة متغيرات يجب أن تكون كلها من نفس النوع (رقمية أو اسمية).
- يستخدم الاختيار IF إذا كانت هناك شروط خاصة يجب تحقيقها لعملية الفرز.
- في حالة اختيار Into Same Variable سيتم استبدال قيم المتغير الأصلية بنتائج عملية الفرز مما يعني فقدان القيم الأصلية.

الإحصاء الوصفي والمدرج التكراري للبيانات

(1) التكرارات والمدرج التكراري Histogram and Frequencies

اختر من اللائحة الرئيسة ما يلي:

- ANALYZE
- اختر الأمر DESCRIPTIVE STATISTICS.
- FREQUENCIES وتستخدم لعرض الجداول التكرارية للمتغيرات موضع الدراسة.

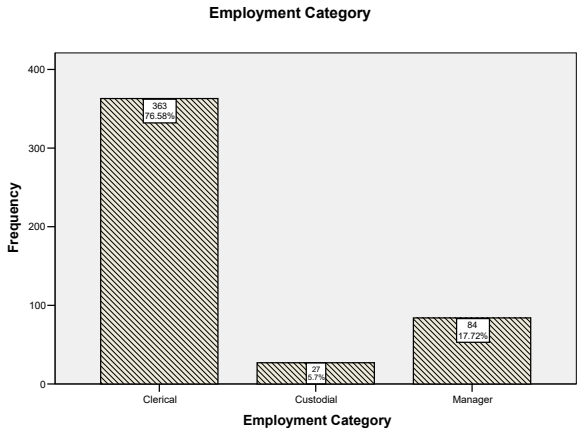


Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	216	45.6	45.6	45.6
	Male	258	54.4	54.4	100.0
	Total	474	100.0	100.0	

يمكن تحديد المطلوب إظهاره بتحديد الاختيارات بالضغط على مفتاح

والضغط على مفتاح الرسم البياني



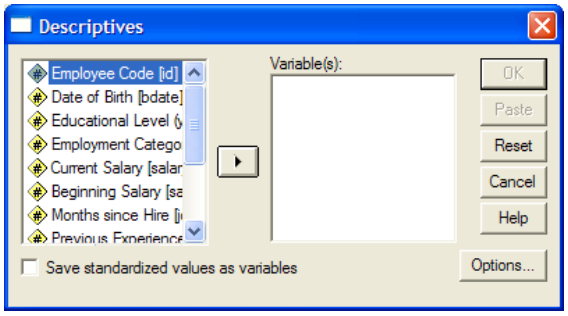
(2) الإحصاء الوصفي Descriptive Statistics

اختر من اللائحة الرئيسة ما يلي:

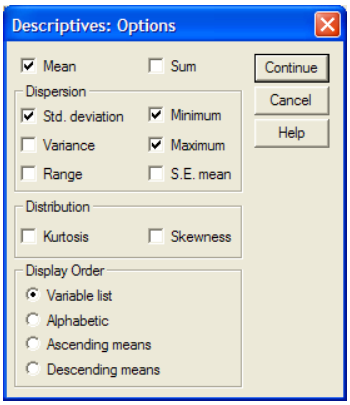
1- ANALYZE

2- اختر من الأمر DESCRIPTIVE STATISTICS

3- DESCRIPTIVES وتعني الإحصاء الوصفي



ولتحديد مخرجات الإحصاء الوصفي اختر OPTION من اللائحة الفرعية، ثم حدد ما هو المطلوب.



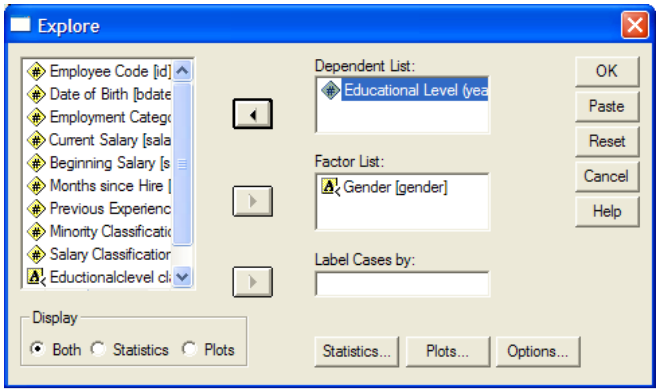
(3) المستكشف Explore

اختر من اللائحة الرئيسة ما يلي:

1- ANALYZE

2- اختر الأمر DESCRIPTIVE STATISTICS

3- EXPLORE وتعني إظهار الخصائص الإحصائية للمتغير- جميع المتغيرات كل على حدة أو حسب مجموعات ذات خصائص معينة. وذلك بكتابة المتغير "المراد إظهار صفاته الإحصائية" في خانة DEPENDENT LIST ولتحديد المجموعة يتم كتابة المتغير في خانة FACTOR LIST.



Descriptives

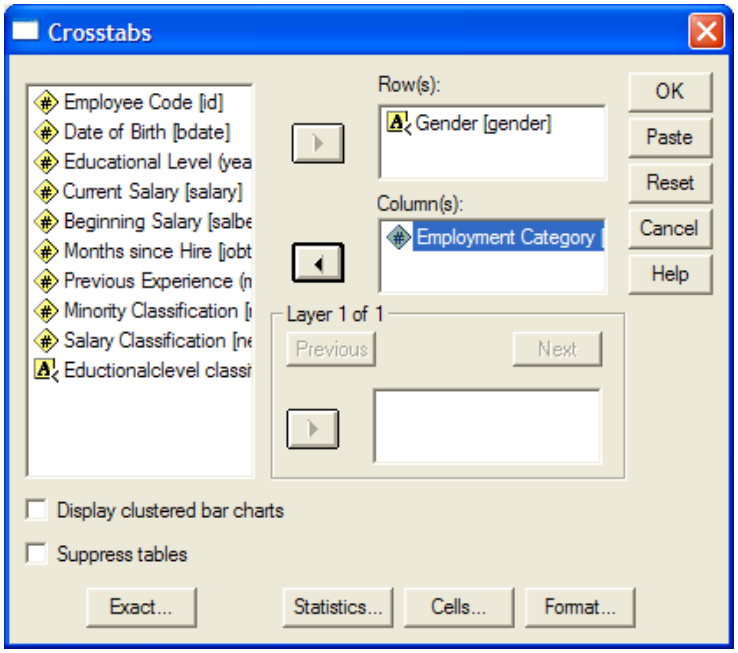
	Gender			Statistic	Std. Error
Educational Level (years)	Female	Mean		12.37	.158
		95% Confidence Interval for Mean	Lower Bound	12.06	
			Upper Bound	12.68	
		5% Trimmed Mean		12.41	
		Median		12.00	
		Variance		5.378	
		Std. Deviation		2.319	
		Minimum		8	
		Maximum		17	
		Range		9	
		Interquartile Range		3	
		Skewness		-.250	.166
		Kurtosis		-.207	.330
	Male	Mean		14.43	.185
		95% Confidence Interval for Mean	Lower Bound	14.06	
			Upper Bound	14.80	
		5% Trimmed Mean		14.52	
		Median		15.00	
		Variance		8.876	
		Std. Deviation		2.979	
		Minimum		8	
		Maximum		21	
		Range		13	
		Interquartile Range		4	
		Skewness		-.455	.152
		Kurtosis		-.044	.302

(4) جداول الاقتران CROSS TABULATION

اختر من اللائحة الرئيسة ما يلي:

1- ANALYZE ثم اختر الأمر DESCRIPTIVE STATISTICS.

2- CROSSTABS، تستخدم إحصائية CHI-SQAURE في جداول الاقتران لمعرفة مدى استقلالية المتغيرات عن بعضها البعض.



Gender * Employment Category Crosstabulation

			Employment Category			
			Clerical	Custodial	Manager	Total
Gender	Female	Count	206	0	10	216
		% within Gender	95.4%	.0%	4.6%	100.0%
		% within Employment Category	56.7%	.0%	11.9%	45.6%
		% of Total	43.5%	.0%	2.1%	45.6%
	Male	Count	157	27	74	258
		% within Gender	60.9%	10.5%	28.7%	100.0%
		% within Employment Category	43.3%	100.0%	88.1%	54.4%
		% of Total	33.1%	5.7%	15.6%	54.4%
Total		Count	363	27	84	474
		% within Gender	76.6%	5.7%	17.7%	100.0%
		% within Employment Category	100.0%	100.0%	100.0%	100.0%
		% of Total	76.6%	5.7%	17.7%	100.0%

Chi-Square Tests

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

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

الرسم البياني

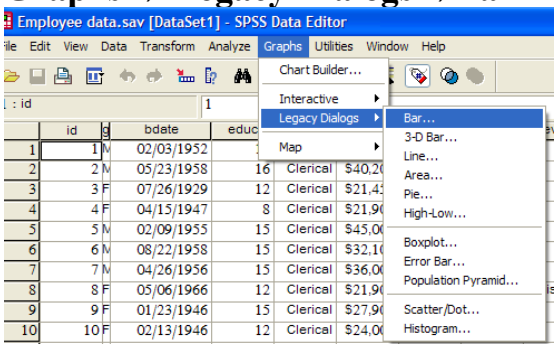
يمكن تمثيل المتغيرات بالرسم البياني وذلك لتحليلها وتفسيرها، ويتفرع من الأمر الرئيسي GRAPHS العديد من الأوامر المتعددة بأشكال الرسم البياني ولكل أمر فرعي اختيارات معينة حسب رغبة الباحث، على سبيل المثال الاختيار BAR وتعني تمثيل البيانات بالأعمدة البيانية البسيطة والمزدوجة.

بعد تحديد الرسم البياني واختيار المتغيرات تظهر النتائج في نافذة خاصة للرسم البياني، حيث يمكن إضافة وتعديل العناوين بالضغط على الرسم البياني مرتين بالماوس.

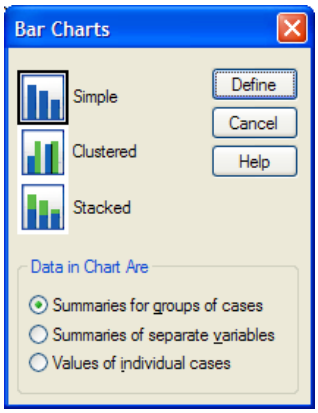
افتح ملف البيانات Employee data

SPSS STEP BY STEP

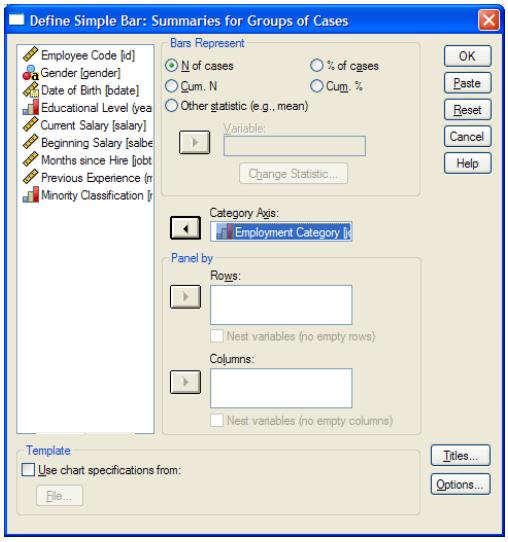
Graphs ⇒ Legacy Dialogs ⇒ Bar



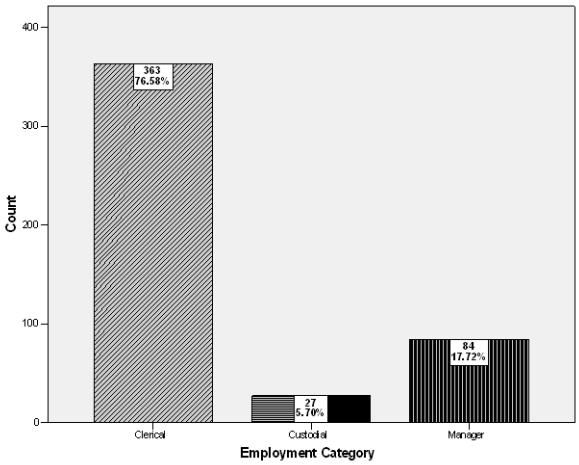
اختر Simple ، Summaries for groups of cases كما هو موضح في المربع الحواري التالي:



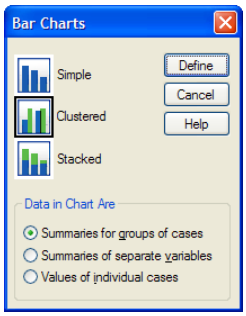
أكمل المربع الحواري كما يلي:



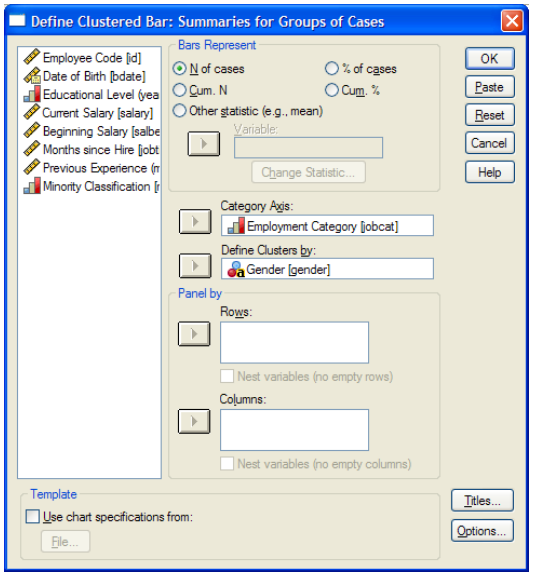
فنحصل على الرسم البياني التالي بعد تنسيقه



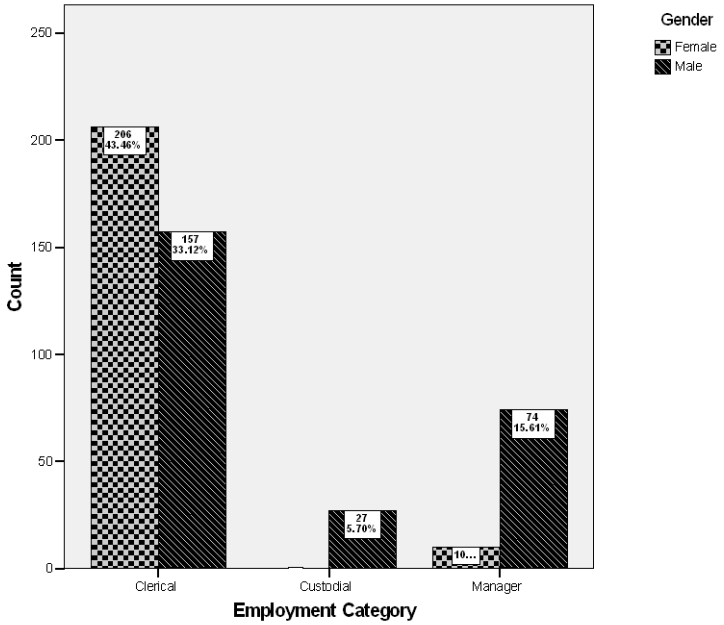
اختر Clustered، Summaries for groups of cases كما هو موضح في المربع الحواري التالي:



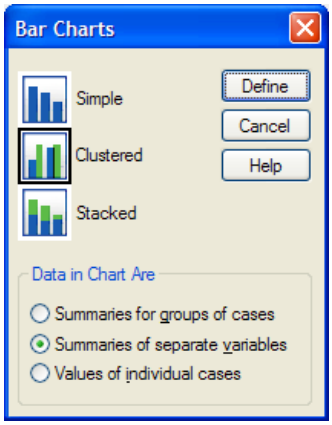
أكمل المربع الحواري كما يلي:



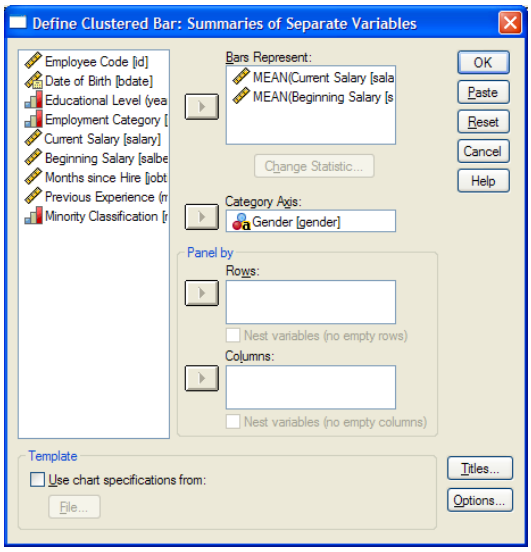
فنحصل على الرسم البياني التالي بعد تنسيقه



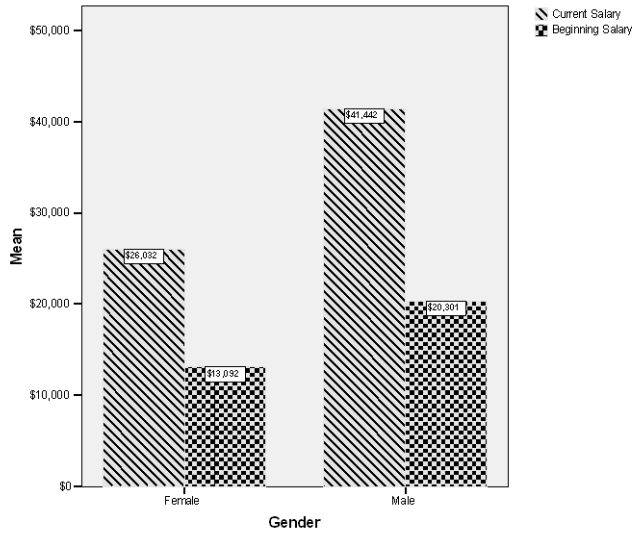
اختر Clustered، Summaries for separate variables كما هو موضح في المربع الحواري التالي:



أكمل المربع الحواري كما يلي:



فنحصل على الرسم البياني التالي بعد تنسيقه

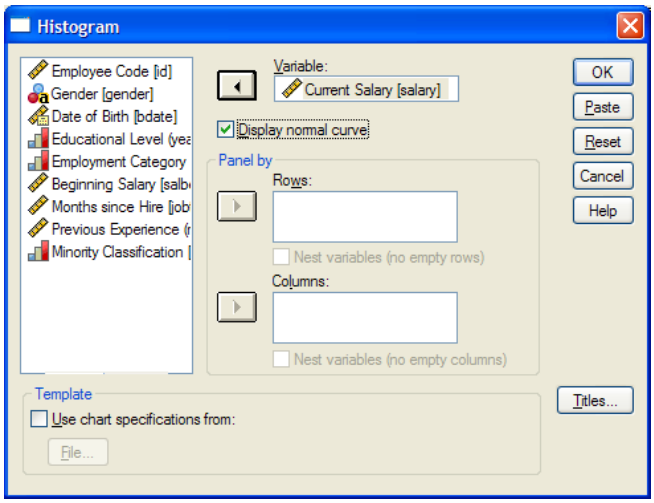


المدرج التكراري Histogram

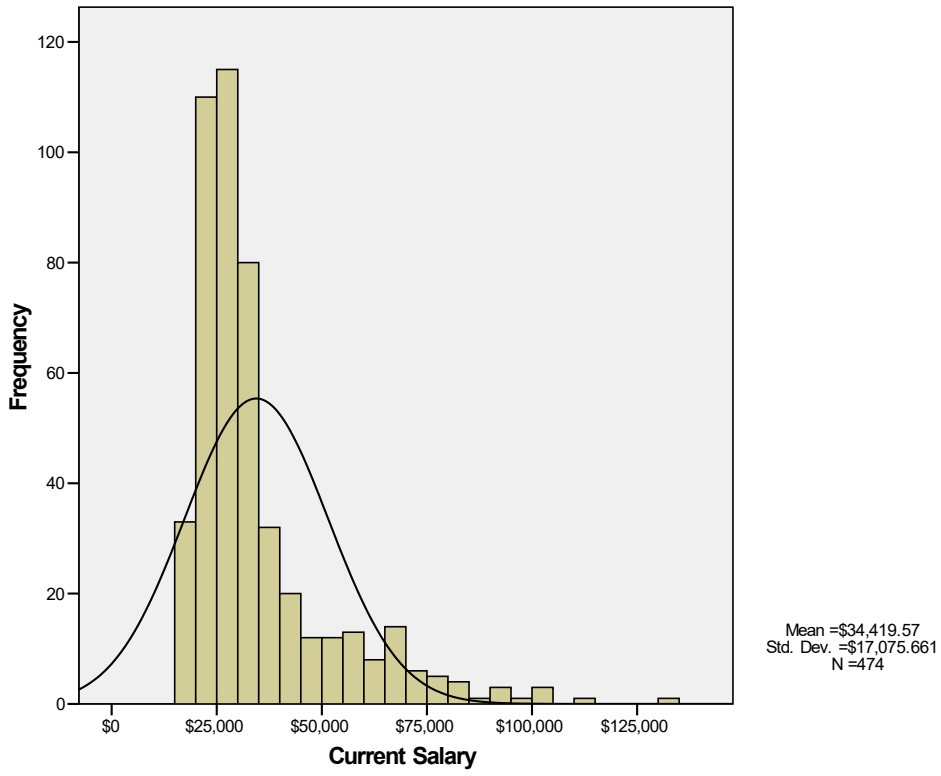
SPSS STEP BY STEP

Graphs⇒ Legacy Dialogs⇒Histogram

أكمل المربع الحواري كما يلي:



فنحصل على الرسم البياني التالي



اختبار الفرضيات Test of Hypotheses

يعتبر موضوع اختبار الفرضيات الإحصائية من أهم الموضوعات في مجال اتخاذ القرارات وسنبدأ بذكر **بعض المصطلحات الهامة في هذا المجال.**

1- الفرضية الإحصائية

هي عبارة عن ادعاء قد يكون صحيحاً أو خطأ حول معلمة أو أكثر لمجتمع أو لمجموعة من المجتمعات.

تقبل الفرضية في حالة أن بيانات العينة تساند النظرية، وترفض عندما تكون بيانات العينة على النقيض منها، وفي حالة عدم رفضنا للفرضية الإحصائية فإن هذا ناتج عن عدم وجود أدلة كافية لرفضها من بيانات العينة ولذلك فإن عدم رفضنا لهذه الفرضية لا يعني بالضرورة أنها صحيحة، أما إذا رفضنا الفرضية بناء على المعلومات الموجودة في بيانات العينة فهذا يعني أن الفرضية خاطئة، ولذلك فإن الباحث يحاول أن يضع الفرضية بشكل يأمل أن يرفضها، فمثلاً إذا أراد الباحث أن يثبت بأن طريقة جديدة من طرق التدريس أحسن من غيرها فإنه يضع فرضية تقول بعدم وجود فرق بين طرق التدريس.

إن الفرضية التي يأمل الباحث أن يرفضها تسمى بفرضية العدم (الفرضية المبدئية) ويرمز لها بالرمز H_0، ورفضنا لهذه الفرضية يؤدى إلى قبول فرضية بديلة عنها تسمى الفرضية البديلة ويرمز لها بالرمز H_1.

2- مستوى المعنوية أو مستوى الاحتمال

وهي درجة الاحتمال الذي نرفض به فرضية العدم H_0 عندما تكون صحيحة أو هو احتمال الوقوع في الخطأ من النوع الأول ويرمز له بالرمز α، وهي يحددها الباحث لنفسه منذ البداية وفي معظم العلوم التطبيقية نختار α مساوية 1% أو 5 % على الأكثر.

3- دالة الاختبار الإحصائية

عبارة عن متغير عشوائي له توزيع احتمالي معلوم وتصف الدالة الإحصائية العلاقة بين القيم النظرية للمجتمع والقيم المحسوبة من العينة.

4- القيمة الاحتمالية: (Sig. or P-value)

احتمال الحصول على قيمة أكبر من أو تساوي (أقل من أو تساوي) إحصائية الاختبار المحسوبة من بيانات العينة أخذاً في الاعتبار توزيع إحصائية الاختبار بافتراض صحة فرض العدم H_0 وطبيعة الفرض البديل H_1. ويتم استخدام القيمة

الاحتمالية لاتخاذ قرار حيال فرض العدم.

خطوات اختبار الفرضيات:

(1) تحديد نوع توزيع المجتمع

يجب تحديد ما إذا كان المتغير العشوائي الذي يتم دراسته يتبع التوزيع الطبيعي أم توزيع بواسون أم توزيع ذو الحدين أم غيره من التوزيعات الاحتمالية المتصلة أو المنفصلة، معظم التوزيعات الاحتمالية يكون توزيعها مشابهاً للتوزيع الطبيعي خاصة إذا كان حجم العينة كبيراً.

هناك نوعان من الطرق الإحصائية التي تستخدم في اختبار الفرضيات:

(أ) الاختبارات المعلمية: وتستخدم في حالة البيانات الرقمية التي توزيعها يتبع التوزيع الطبيعي.

(ب) الاختبارات غير المعلمية: وتستخدم في حالة البيانات الرقمية التي توزيعها لا يتبع التوزيع الطبيعي طبيعي، وكذلك في حالتي البيانات الترتيبية والوصفية.

2- صياغة فرضيتا العدم والبديلة

مثلاً: عند اختبار أن متوسط المجتمع μ يساوى قيمة معينة μ_0 مقابل الفرضية القائلة بأن μ لا يساوى μ_0، فإن فرضية العدم H_0 والفرضية البديلة H_1 تكون على النحو التالي:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

3- اختيار مستوى المعنوية α

4- اختيار دالة الاختبار الإحصائية المناسبة

5- جمع البيانات من العينة وحساب قيمة دالة الاختبار الإحصائية

6- اتخاذ القرارات

نرفض H_0 ونقبل H_1 إذا كانت قيمة الاحتمال (Sig. or P-value) أقل من أو تساوي مستوى المعنوية (α)، أما إذا كانت قيمة الاحتمال أكبر من α فلا يمكن رفض H_0.

وبرنامج SPSS يعطي Sig. 2-tailed فبالتالي نرفض فرضية العدم H_0 عندما تكون $P-Value(Sig.) < \alpha$.

أولاً: اختبار T في حالة اختبار فرضيات متعلقة بمتوسط واحد

إذا كان المطلوب اختبار فرضية العدم $H_0 : \mu = \mu_0$ على مستوى دلالة α مقابل

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

2- $H_1 : \mu > \mu_0$

3- $H_1 : \mu < \mu_0$

مثال (1)

البيانات التالية تمثل درجات عشرين طالباً في مساق ما:

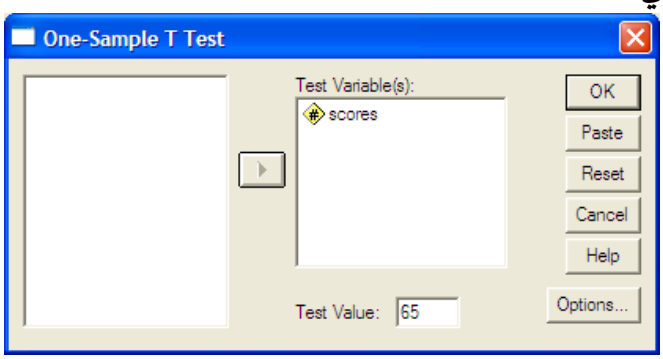
65, 72, 68, 82, 45, 92, 87, 85, 90, 60, 48, 60, 68, 72, 79, 68, 73, 69, 78, 84

المطلوب: اختبار الفرضية المبدئية القائلة بأن متوسط درجات الطلاب = 65 درجة.

SPSS STEP BY STEP

Analyze ⇒ Compare Means ⇒ One-Sample T Test

أكمل المربع الحواري كما يلي:



نتائج الاختبار

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
scores	20	72.25	12.867	2.877

One-Sample Test

	Test Value = 65					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
scores	2.520	19	.021	7.250	1.23	13.27

من النتائج السابقة يمكن استنتاج ما يلي:

t = 2.52، Sig.(2-tailed)=0.021، وهي أقل من 0.05 (مستوى المعنوية) فبالتالي نرفض الفرضية المبدئية القائلة بأن متوسط درجات الطلاب في الرياضيات نساوي 65 درجة، ونستنتج أن درجات الطلاب لا تساوي (تختلف عن) 65.

يمكن اختبار الفرضية البديلة القائلة بأن متوسط درجات الطلاب أكبر من 65.

حيث أن نتيجة الوسط الحسابي للعينة تتوافق مع الفرضية البديلة (متوسط درجات الطلاب أكبر من 65 درجة) فبالتالي نستنتج أن متوسط درجات الطلاب أكبر من 65 درجة.

ثانياً: اختبارات الفروق بين متوسطين مجتمعين مستقلين

في هذه الحالة نأخذ عينة عشوائية من توزيع طبيعي $N(\mu_1, \sigma_1^2)$ ، وعينة عشوائية أيضاً من توزيع طبيعي $N(\mu_2, \sigma_2^2)$ ومستقل عن التوزيع الأول، وتكون $\sigma_1^2 = \sigma_2^2$ ولكنهما مجهولتان.

إذا كان المطلوب اختبار فرضية العدم $H_1 : \mu_1 - \mu_2 = 0$ على مستوى دلالة α مقابل

$H_1 : \mu_1 - \mu_2 \neq 0$ (1)

$H_1 : \mu_1 - \mu_2 > 0$ (2)

$H_1 : \mu_1 - \mu_2 < 0$ (3)

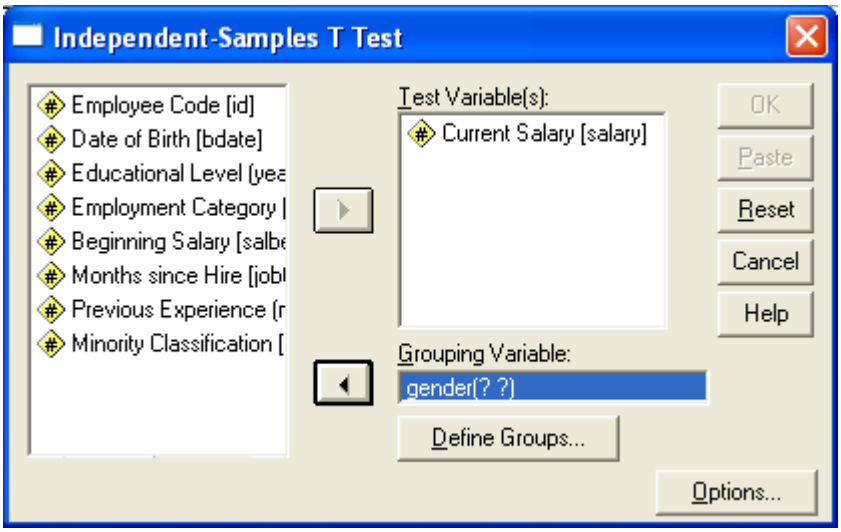
مثال (2)

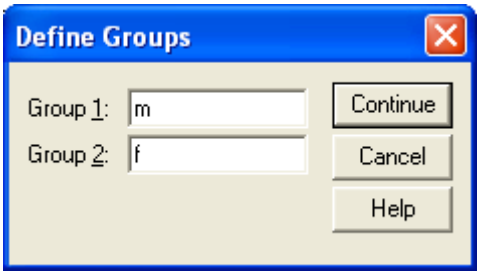
مستخدماً الملف employee. المطلوب اختبار ما إذا كان هناك فرق معنوي بين متوسط الراتب الحالي السنوي للموظفين (salary) يعزى إلى متغير الجنس (gender) مستخدماً مستوى معنوية $\alpha = 0.05$.

SPSS STEP BY STEP

Analyze ⇒ Compare Means ⇒ Independent- Samples T Test

أكمل المربع الحواري كما يلي:





نتيجة الاختبار

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Current Salary	Male	258	$41,441.78	$19,499.214	$1,213.968
	Female	216	$26,031.92	$7,558.021	$514.258

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
Current Salary	Equal variances assumed	119.669	.000	10.945	472	.000	$15409.86	$1,407.906	$12,643.322	$18,176.401
	Equal variances not assumed			11.688	344.262	.000	$15409.86	$1,318.400	$12,816.728	$18,002.996

من النتائج السابقة يمكن استنتاج ما يلي:

تباينيا المجتمعين غير متساويين حسب اختبار ليفين (Levene's Test)، حيث Sig. = 0.000. حيث أن قيمة t=11.688، Sig. = 0.000 فبالتالي نرفض فرضية العدم القائلة بأنه لا يوجد فرق معنوي بين متوسطي الراتب الحالي السنوي للذكور والإناث على أساس مستوى معنوية 5%.

95% فترة الثقة للفرق بين متوسطي المجتمعين هي: (18003.00 , 12816.73). ونجد أن الصفر لا ينتمي إلى الفترة السابقة مما يؤكد أنه يوجد فرق معنوي بين متوسطي الراتب الحالي السنوي للذكور والإناث، وهي نفس النتيجة التي حصلنا عليها في حالة استخدام اختبار t.

يمكن اختبار الفرضية البديلة القائلة بأن متوسط الراتب الحالي السنوي للذكور أكبر منه للإناث.

حيث أن نتيجة الوسط الحسابي للفرق بين متوسطي الذكور والإناث موجباً (15409.88) يتوافق مع الفرضية البديلة بالتالي نستنتج أن متوسط الراتب الحالي السنوي للذكور أكبر منه للإناث.

ثالثاً: اختبارات الفروق بين متوسطي مجتمعين من عينات مرتبطة

في هذه الحالة تكون البيانات مزدوجة، أي أن العينتين مرتبطتان حيث أن البيانات تكون على شكل أزواج وبالتالي فإن حجم العينتين لابد أن يكون متساوياً.

مثال (3)

البيانات التالية تمثل نتائج تجربة أجريت على عشرين شخصاً لاختبار مدى فعالية حضور المحاضرات الوزن، حيث تم القياس قبل البدء في تطبيق هذا النظام، وبعد اتباع هذا النظام الخاص لمدة ثلاثة شهور.

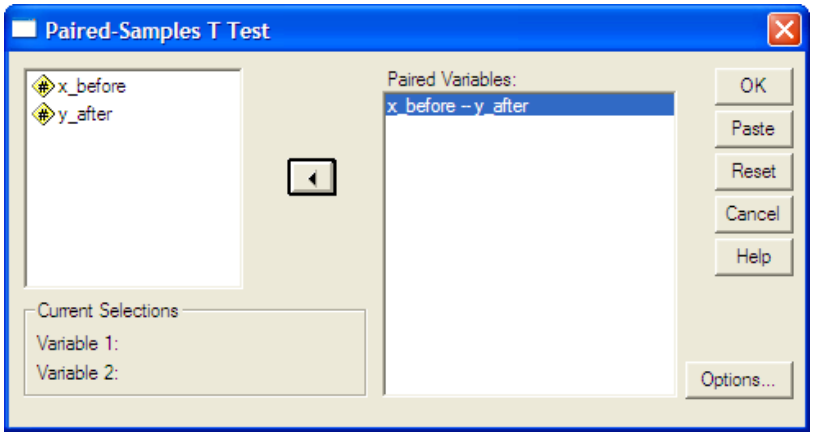
92	103	120	89	93	107	94	90	110	96	**Before**
84	95	103	76	85	104	87	85	96	90	**After**
123	111	90	95	123	105	110	86	94	86	**Before**
107	102	83	89	109	95	102	80	84	78	**After**

المطلوب: هل تستطيع أن تستنتج أن نظام الحضور كان فعالاً مستخدماً مستوى دلالة $\alpha = 0.05$؟

SPSS STEP BY STEP

Analyze ⇒ Compare Means ⇒ Paired- Samples T Test

أكمل المربع الحواري كما يلي:



Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	x_before	100.8500	20	12.11035	2.70796
	y_after	91.7000	20	10.13644	2.26658

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 x_before & y_after	20	.957	.000

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1 x_before - y_after	9.15000	3.78744	.84690	7.37742	10.92258	10.804	19	.000

من النتائج السابقة يمكن استنتاج ما يلي:

يوجد ارتباط طردي قوي بين الحضور قبل قبل وبعد حيث أن $R = 0.957$.

$t = 10.804$، Sig. (2 tailed) $= 0.000$ وبالتالي نرفض فرضية العدم القائلة بأنه لا يوجد فرق بين متوسطي الحضور قبل وبعد ، ونستنتج أنه يوجد فرق معنوي بين متوسطي الحضور.

يمكن اختبار الفرضية البديلة القائلة بأن متوسط الحضور حيث أن نتيجة الوسط الحسابي للفرق بين متوسطي الحضور موجباً (9.15) يتوافق مع الفرضية البديلة فبالتالي نستنتج أن متوسط الحضور عند مستوى دلالة $\alpha = 0.05$.

تحليل التباين(Analysis of Variance (ANOVA

في هذه الحالة يكون الاهتمام مركزاً على دراسة تأثير عامل واحد له عدد من المستويات المختلفة وعند كل مستوى تكرر التجربة عدد من المرات، فمثلاً إذا أردنا اختبار ما إذا كانت هناك فروق بين ثلاثة أساليب لتدريس مقياس تحليل البيانات مثلاً، ويكون المطلوب بحث ما إذا كانت هذه الأساليب لها تأثيرات متساوية في درجة تحصيل الطالب مع ملاحظة أن وجود اختلاف بين درجات الطلاب قد يرجع إلى عدة عوامل أخرى منها الفروق الفردية وعدد ساعات الدراسة وعدد أفراد الأسرة مثلاً أو غيرها من العوامل الأخرى.

أولاً: تحليل التباين الأحادي One-Way ANOVA

أسلوب تحليل التباين يعطي نتائج جيدة إذا تحققت الشروط التالية:

1- المتغيرات (قيمة مفردات الظاهرة) مستقلة ولها توزيع طبيعي بنفس قيمة التباين.

2- مجموعة البيانات في المستويات المختلفة تشكل عينات عشوائية مستقلة ولها تباين مشترك σ^2

فإذا لم تتحقق هذه الشروط يمكن استخدام الاختبارات غير المعلمية

تحت الفروض السابقة، فإن الاختلاف الكلي المشاهد في مجموعة البيانات ينقسم إلى مركبتين الأولى نتيجة العامل والثانية للخطأ التجريبي.

ويكون المطلوب في تحليل التباين الأحادي اختبار الفرضية المبدئية H_0 أنه لا يوجد فروق بين متوسطات المجتمعات على مستوى دلالة α.

بفرض أن العامل المراد دراسته له r من المستويات المستقلة فيكون المطلوب اختبار الفرضية المبدئية (فرضية العدم):

$H_0 : \mu_1 = \mu_2 = ... = \mu_r$ أي أنه لا يوجد فروق بين متوسطات المجتمعات.

مقابل الفرضية البديلة:

H_a : أي أنه يوجد فروق بين متوسطات المجتمعات.

يوجد متوسطين على الأقل من أوساط المجتمعات غير متساويين

عند رفض فرضية العدم والتي تنص على تساوي المتوسطات وقبول الفرضية البديلة أنه يوجد اثنين أو أكثر من المتوسطات غير المتساوية، ونريد اختبار أي من هذه المتوسطات متساوٍ أو غير متساوٍ، وللإجابة على هذا التساؤل سنعرض عدة اختبارات.

لتنفيذ ذلك عملياً اضغط Post - Hoc في نافذة One-Way ANOVA.

مثال (4)

يمثل الجدول التالي درجات مجموعة من الطلبة تم تدريسهم مقياس تحليل البيانات العامة بثلاثة أساليب مختلفة:

M_1 , M_2 , M_3

M_3	M_2	M_1
48	64	70
94	45	83
83	56	87
84	50	78
80	71	
87		
90		

المطلوب:

1- إدخال البيانات السابقة في متغير اسمه (marks).

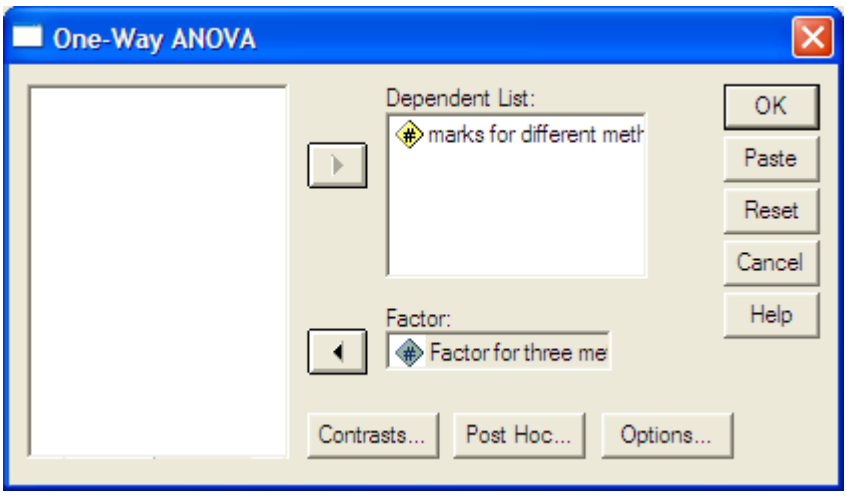
2- إنشاء متغير جديد اسمه (factor) له ثلاثة قيم، (1) تمثل الأسلوب الأول، (2) تمثل الأسلوب الثاني و (3) تمثل الأسلوب الثالث.

3- هل هناك فرقاً بين أساليب التدريس الثلاثة مستخدماً مستوى دلالة $\alpha = 0.05$ ؟

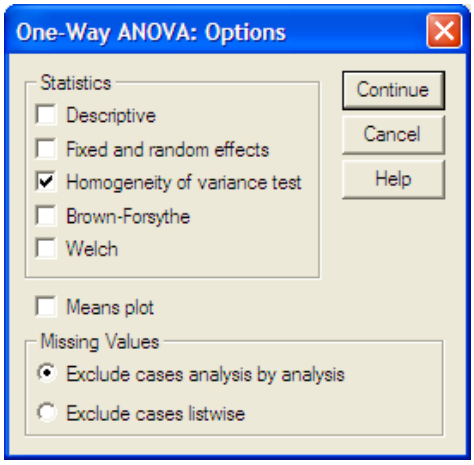
الحل العملي:

SPSS STEP BY STEP

Analyze ⇒ Compare Means ⇒ One-Way ANOVA



انقر بالفأرة على Options ثم أكمل المربع الحواري كما يلي:



ANOVA

marks for different methods

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1849.093	2	924.546	6.044	.014
Within Groups	1988.657	13	152.974		
Total	3837.750	15			

Test of Homogeneity of Variances

marks for different methods

Levene Statistic	df1	df2	Sig.
.322	2	13	.730

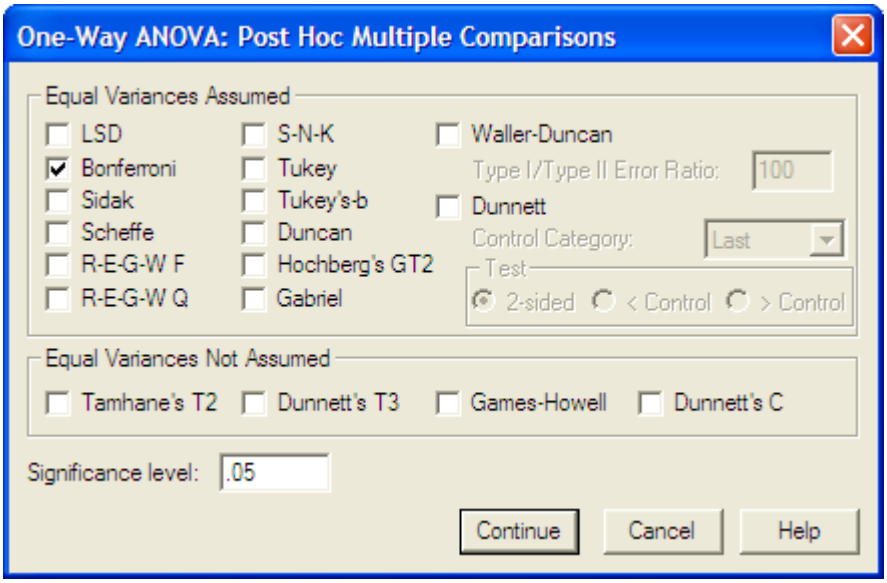
من النتائج السابقة نستنتج ما يلي:

قيمة إحصاء ليفين = 0.322، Sig. = 0.73 وهذا يدل على تجانس تباين طرق التدريس.

F = 6.044، Sig. = 0.014 وبالتالي نرفض الفرضية المبدئية والتي تنص على أنه لا يوجد فروق بين متوسطات طرق التدريس الثلاثة ونستنتج أن هناك فرقاً بين أساليب التدريس المختلفة، أي أنه يوجد دليل كافٍ على أن متوسطات أساليب التدريس المختلفة ليست كلها متساوية، وذلك باستخدام مستوى دلالة $\alpha = 0.05$

عند رفض فرضية العدم والتي تنص على تساوي المتوسطات وقبول الفرضية البديلة أنه يوجد اثنين أو أكثر من المتوسطات غير المتساوية، ونريد اختبار أي من هذه المتوسطات متساوٍ أو غير متساوٍ، وللإجابة على هذا التساؤل سنعرض عدة اختبارات.

لتنفيذ ذلك عملياً اضغط Post – Hoc في نافذة One-Way ANOVA ثم أكمل المربع الحواري كما يلي:



توجد عدة اختبارات في حالة تحقق شرط تجانس التباين من عدمه.

حيث أن شرط تجانس تباين مستويات أساليب التدريس متحقق فيمكن اختيار اختبار بونفيروني (Bonferroni) أو شفييه (Scheffe) وذلك في حالة تساوي أو عدم تساوي حجوم العينات.

Multiple Comparisons

Dependent Variable: marks for different methods

Bonferroni

(I) Factor for three methods	(J) Factor for three methods	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Method_1	Method_2	22.30000	8.29687	.056	-.4827	45.0827
	Method_3	-1.35714	7.75221	1.000	-22.6442	19.9300
Method_2	Method_1	-22.30000	8.29687	.056	-45.0827	.4827
	Method_3	-23.65714*	7.24211	.018	-43.5435	-3.7708
Method_3	Method_1	1.35714	7.75221	1.000	-19.9300	22.6442
	Method_2	23.65714*	7.24211	.018	3.7708	43.5435

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

من النتائج السابقة يمكن استنتاج ما يلي:

يوجد فرق معنوي بين متوسطي أسلوبي التدريس الثاني والثالث وذلك لأن Sig. =0.018 وهي أقل من مستوى الدلالة $\alpha = 0.05$.

درجات الطلاب باستخدام الأسلوب الثالث أفضل من درجات الطلاب باستخدام الأسلوب الثاني، وذلك لأن الفرق بين وسطيهما موجباً (23.66).

ثانياً: تحليل التباين الثنائي Two-Way ANOVA

مثال (1):

يمثل الجدول التالي عدد الوحدات المنتجة في الأسبوع وذلك لعشرة عمال باستخدام ثلاثة أنواع مختلفة من الالالات

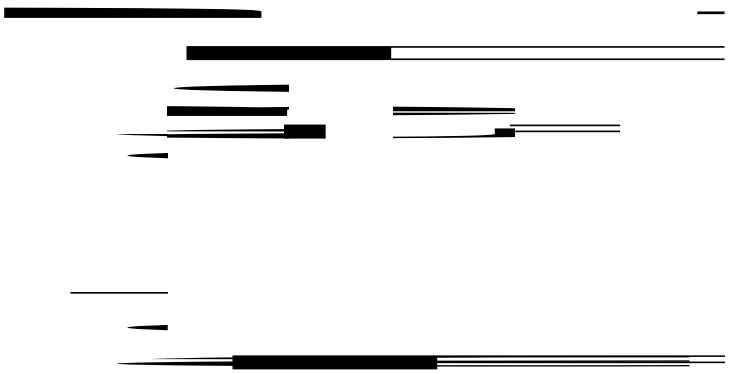
العامل	1	2	3	4	5	6	7	8	9	10
1	90	80	70	90	60	76	68	94	83	74
2	82	92	60	70	50	65	79	82	68	80
3	76	65	82	80	90	92	86	71	93	68

المطلوب اختبار:

أ) ما إذا كان العمال متساويين في الإنتاج.

ب) ما إذا كانت الآلات متساوية في الإنتاج مستخدماً مستوى دلالة $\alpha = 0.05$

الحل العملي:



من النتائج السابقة يمكن استنتاج ما يلي:

أولاً: اختبار ما إذا كان العمال متساويين في الإنتاج

فرضية العدم: $H_0 : \mu_1 = \mu_2 = ... = \mu_{10}$

حيث أن: F = 0.562 (صغيرة)، Sig. = 0.81 أكبر من $\alpha = 0.05$ (مستوى الدلالة) فبالتالي لا يمكن رفض H_0 ونستنتج أن إنتاج العمال متساوٍ.

ثانياً: اختبار ما إذا كانت الآلات متساوية في الإنتاج

فرضية العدم: $H_0 : \mu_1 = \mu_2 = \mu_3$

حيث أن: $F = 1.040$ (صغيرة)، $Sig. = 0.374$ أكبر من $\alpha = 0.05$ (مستوى الدلالة) فبالتالي لا يمكن رفض H_0 ونستنتج أن الماكينات متساوية في الإنتاج.

مثال (2)

يمثل الجدول التالي زيادة وزن الأطفال (مقدراً بالكيلوجرام) باستخدام ثلاثة أنواع مختلفة من الفيتامينات، وأربعة أنواع من الغذاء الخاص خلال ستة أشهر.

الغذاء الخاص	B_1	B_2	B_3
A_1	2, 1.5	1.8, 2.2	2.3, 1.6
A_2	2.3, 2.6	2.3, 1.5	1.7, 2.1
A_3	1.5, 2	2.1, 1.8	2.3, 1.7
A_4	2.1, 1.8	1.5, 2.1	1.9, 1.5

المطلوب:

مستخدماً مستخدماً مستوى الدلالة $\alpha = 0.05$ المطلوب:

أ) هل توجد فروق ذات دلالة إحصائية بين متوسطات الفيتامينات ؟

ب) هل توجد فروق ذات دلالة إحصائية بين متوسطات أنواع الغذاء ؟

ج) هل يوجد تفاعل بين نوع الفيتامين، ونوع الغذاء الخاص

الحل العملي:

من النتائج السابقة يمكن استنتاج ما يلي:

أولاً: اختبار ما إذا كانت هناك فروق بين متوسطات الفيتامينات الثلاثة

فرضية العدم: $H_0 : \mu_1 = \mu_2 = \mu_3$

حيث أن: F = 0.126 ، Sig. = 0.883 أكبر من $\alpha = 0.05$ (مستوى الدلالة) فبالتالي لا يمكن رفض H_0 ونستنتج أن متوسطات الفيتامينات الثلاثة متساوية التأثير على زيادة وزن الأطفال.

ثانياً: اختبار ما إذا كان هناك فروق بين متوسطات أنواع الغذاء الخاص الأربعة

فرضية العدم: $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

حيث أن: F = 0.589 ، Sig. = 0.634 أكبر من $\alpha = 0.05$ (مستوى الدلالة) فبالتالي لا يمكن رفض H_0 ونستنتج أن متوسطات أنواع الغذاء الأربعة متساوية التأثير على زيادة وزن الأطفال.

ثالثاً: اختبار ما إذا كان هناك تفاعل بين نوع الفيتامين ونوع الغذاء الخاص

فرضية العدم: لا يوجد تفاعل بينهما

حيث أن: F = 0.741 (صغيرة)، Sig. = 0.627 أكبر من $\alpha = 0.05$ (مستوى الدلالة) فبالتالي لا يمكن رفض H_0 ونستنتج أنه لا يوجد تفاعل بين نوع الفيتامين ونوع الغذاء الخاص، أي أنهما مستقلان.

في حالة رفض فرضية العدم ولمقارنة متوسطات كل عامل على حده لمعرفة أي منهم أكثر تأثيراً اضغط Post Hoc... ثم اختر Bonferroni كما في حالة تحليل التباين الأحادي.

ثالثاً: تحليل التباين الثلاثي Three-Way ANOVA

يستخدم تحليل التباين الثلاثي في حالة تجارب يؤثر عليها ثلاثة عوامل **A**,**B**,**C** مثلاً.

هناك سبعة اختبارات في حالة تحليل التباين الثلاثي مع وجود تفاعل بين العوامل الثلاثة وهي:

1. اختبار الفرضية H_1: لا يوجد فروق بين متوسطات مستويات العامل الأول **A**.
2. اختبار الفرضية H_2: لا يوجد فروق بين متوسطات مستويات العامل الثاني **B**.
3. اختبار الفرضية H_3: لا يوجد فروق بين متوسطات مستويات العامل الثالث **C**.
4. اختبار الفرضية H_4: لا يوجد تفاعل بين العاملين الأول والثاني **A**,**B**.
5. اختبار الفرضية H_5: لا يوجد تفاعل بين العاملين الأول والثالث **A**,**C** .
6. اختبار الفرضية H_6: لا يوجد تفاعل بين العاملين الثاني والثالث **B**,**C**.
7. اختبار الفرضية H_7: لا يوجد تفاعل بين العوامل الثلاثة **A**,**B**,**C**.

مثال (5):

عند إنتاج مادة معينة. كان هناك ثلاثة عوامل مهمة وهى: **A**: تأثير المهندس (هناك ثلاثة مهندسين)**B** : المادة المساعدة على إنتاج المادة المطلوبة (هناك ثلاثة أنواع من المواد المساعدة) **C**: زمن التعبئة بعد الإنتاج (هناك فترتان 15 دقيقة و 20 دقيقة). يمثل الجدول التالي نتائج تجربة أجريت لهذا الغرض.

		B_1	B_2	B_3
C_1	A_1	10.7, 10.8, 11.3	10.3, 10.2, 10.5	11.2, 11.6, 12.0
	A_2	11.4, 11.8, 11.5	10.2, 10.9, 10.5	10.7, 10.5, 10.2
	A_3	13.6, 14.1, 14.5	12.0, 11.5, 11.6	11.1, 11.0, 11.5
C_2	A_1	10.9, 12.1, 11.5	10.5, 11.1, 10.3	12.2, 11.0, 11.7
	A_2	9.8, 11.3, 10.9	12.6, 7.5, 9.9	10.8, 10.2, 11.5
	A_3	10.7, 11.7, 12.7	10.2, 11.5, 10.9	11.9, 11.6, 12.2

المطلوب:

كوِّن جدول تحليل التباين الثلاثي ثم فسُّر النتائج الكاملة التي يمكن الحصول عليها منه

Univariate Analysis of Variance

Warnings

Post hoc tests are not performed for قبعلا نمز because there are fewer than three groups.

التعليق:

لا يمكن استخدام اختبارات **Post Hoc** لمتغير "زمن التعبئة" لأنه يتكون من مجموعتين فقط ويستخدم في هذه الحالة اختبار **T** في حالة العينات المستقلة كما تم شرحه سابقاً (لمزيد من التفاصيل أنظر الباب الخامس).

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

Dependent Variable: قيرعم قدام جلاا

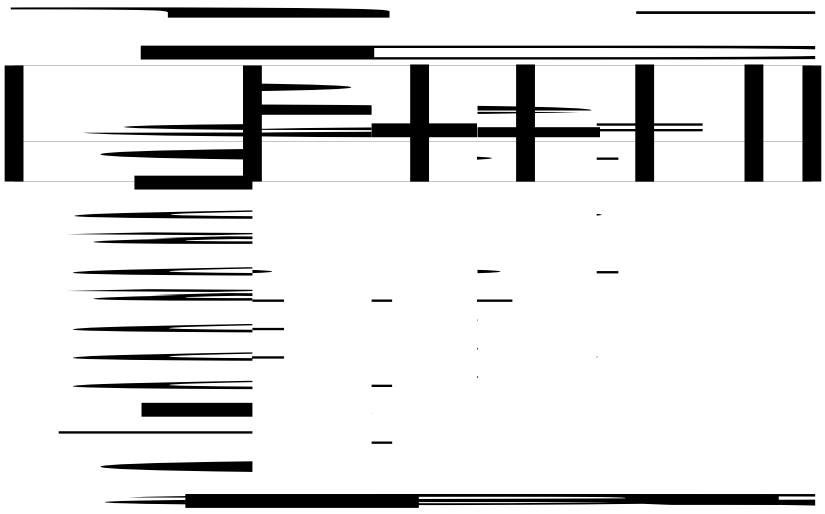
F	df1	df2	Sig.
2.281	17	36	.019

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

a. Design: Intercept+سنجه لها+نمزلا+قدالها+سنجه لها+ * سنجه لها+نمزلا+قدالها+سنجه لها+المادة+المهندس * الزمن+المادة * الزمن+المهندس * المادة * الزمن

التعليق:

قيمة إحصاء ليفين = **2.281**، **Sig. = 0.019** وهذا يدل على عدم تجانس العوامل الثلاثة.



التعليق:

1. **أثر المهندس على كمية إنتاج المادة:**

F=11.645، Sig.=.000 أصغر من 0.05 (مستوى الدلالة) بالتالي فإنه يوجد تأثير للمهندس على كمية إنتاج المادة.

2. **أثر نوع المادة المساعدة على كمية إنتاج المادة:**

F=8.48، Sig.=.001 أصغر من 0.05 (مستوى الدلالة) بالتالي فإنه يوجد تأثير لنوع المادة المساعدة على كمية إنتاج المادة.

3. **أثر زمن التعبئة على كمية إنتاج المادة:**

F=1.974، Sig.=.169 أكبر من 0.05 (مستوى الدلالة) بالتالي فإنه لا يوجد تأثير لزمن التعبئة على كمية إنتاج المادة.

4. **التفاعل بين المهندس والمادة المساعدة:**

F=1.988، Sig.=.117 أكبر من 0.05 (مستوى الدلالة) بالتالي فإنه لا يوجد تفاعل بين المهندس والمادة المساعدة.

5. **التفاعل بين المهندس وزمن التعبئة**

F=2.427، Sig.=.103 **أكبر من** 0.05 **(مستوى الدلالة) بالتالي فإنه لا يوجد تفاعل بين المهندس وزمن التعبئة.**

6. التفاعل بين المادة المساعدة وزمن التعبئة:

F=3.026، Sig.=.061 **أكبر من** 0.05 **(مستوى الدلالة) بالتالي فإنه لا يوجد تفاعل بين المادة المساعدة وزمن التعبئة.**

7. التفاعل بين المهندس والمادة المساعدة وزمن التعبئة:

F=2.043، Sig.=.103 أكبر من **0.05** (مستوى الدلالة) بالتالي فإنه لا يوجد تفاعل بين المهندس والمادة المساعدة وزمن التعبئة.

التعليق:

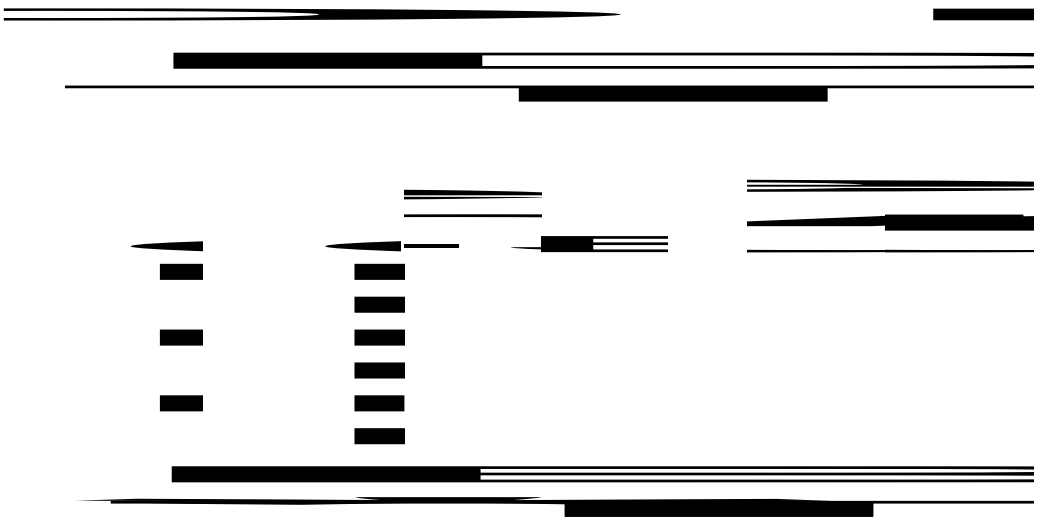
تم استخدام اختبار Tamhane لمقارنة متوسطات كل عامل على حده.

- لا يوجد فرق معنوي بين متوسطي تأثير المهندسين الأول والثاني على كمية إنتاج المادة لأن Sig.=.402 أكبر من 0.05 (مستوى الدلالة).
- يوجد فرق معنوي بين متوسطي تأثير المهندسين الأول والثالث على كمية إنتاج المادة لأن Sig.=.048 أصغر من **0.05** (مستوى الدلالة) وحيث أن $\mu_1 - \mu_3 = -.8$ بالتالي فإن تأثير المهندس الثالث أكبر من تأثير المهندس الأول على كمية إنتاج المادة.

- يوجد فرق معنوي بين متوسطي تأثير المهندسين الثاني والثالث على كمية إنتاج المادة لأن Sig.=.007 أصغر من **0.05** (مستوى الدلالة) وحيث أن $\mu_2 - \mu_3 = -1.2278$ بالتالي فإن تأثير المهندس الثالث أكبر من تأثير المهندس الثاني على كمية إنتاج المادة.

من هذه النتائج يمكن القول بأن المهندس الثالث له تأثير أكبر من المهندسين الأول والثاني على كمية إنتاج المادة.

تأثير المادة المساعدة



التعليق:

- يوجد فرق معنوي بين متوسطي تأثير المادة المساعدة الأولى والثانية على كمية إنتاج المادة لأن Sig.=.030 أصغر من **0.05** (مستوى الدلالة) وحيث أن $\mu_1 - \mu_2 = 1.0611$ بالتالي فإن تأثير المادة المساعدة الأولى أكبر من تأثير المادة المساعدة الثانية على كمية إنتاج المادة.

- لا يوجد فرق معنوي بين متوسطي تأثير المادة المساعدة الأولى والثالثة على كمية إنتاج المادة لأن Sig.=.427 أكبر من 05. (مستوى الدلالة)
- لا يوجد فرق معنوي بين متوسطي تأثير المادة المساعدة الثانية والثالثة على كمية إنتاج المادة لأن Sig.=.151 أكبر من 05. (مستوى الدلالة)

من هذه النتائج يمكن القول بأن المادة المساعدة الأولى لها تأثير أكبر من المادة المساعدة الثاني والثالثة على كمية إنتاج المادة.

الاختبارات غير المعلمية Nonparametric Tests

في بعض الحالات قد لا تتوافر في المجتمع موضع الدراسة أن يكون توزيع هذا المجتمع له توزيع طبيعي أو يقترب منه، لذلك فإن استخدام الاختبارات المعلمية في مثل هذه الحالات قد يؤدي إلى نتائج غير دقيقة، كذلك يفترض أن تكون بيانات الظاهرة موضع الدراسة دقيقة، ولكن في بعض الأحيان يتعذر أخذ قياسات عددية دقيقة على بعض الظواهر، لذلك فإننا نستخدم طرق غير معلمية لا تعتمد على شروط معينة تتعلق بتوزيع المجتمع ولا تحتاج إلى قياسات دقيقة.

مزايا استخدام الاختبارات غيرالمعلمية:

1. سهولة العمليات الحسابية المستخدمة.
2. لا تحتاج إلى شروط كثيرة لذلك فإن إمكانية إساءة استعمالها قليلة جداً.
3. تستخدم عندما لا تتحقق الشروط اللازمة لتطبيق الاختبارات المعلمية مثل أن يكون توزيع المجتمع طبيعياً.
4. تستخدم في حالة صعوبة الحصول على بيانات دقيقة.
5. لا يتطلب استخدامها معرفة دقيقة في مجال الرياضيات أو الإحصاء.
6. لا تشترط استخدامها أن يكون حجم العينات كبيراً، لذلك فإن عملية جمع البيانات في هذه الحالة توفر الوقت والمجهود والتكلفة.

عيوب استخدام الاختبارات غيرالمعلمية:

1. تستخدم أحياناً في الحالات التي يجب استخدام الاختبارات المعلمية وذلك لسهولة استخدامها.
2. صعوبة الحصول على توزيع دوال الاختبار المستخدمة في هذه الاختبارات.

يمكن استخدام الاختبارات غيرالمعلمية في الحالات التالية:

1. للحصول على قرار سريع.
2. إذا كانت البيانات المتوفرة عن ظاهرة ما لا تتفق مع الاختبارات المعلمية.
3. إذا كانت الشروط المطلوب توافرها في الاختبار المعلمي غير متحققة.

سنعرض فيما يلي استخدام برنامج SPSS في الاختبارات غيرالمعلمية التالية:

1. استخدام اختبار كولمجروف – سمرنوف "One-Sample Kolmogorov-Smirnov Test" لمعرفة ما إذا كانت البيانات تتبع التوزيع الطبيعي.
2. اختبار الإشارة "Sign Test"لاختبار فرضيات حول متوسط مجتمع واحد.
3. اختبار ويلكوكسن "Wilcoxon Test" لاختبار فرضيات حول مقارنة متوسطي مجتمعين في حالة العينات المرتبطة.
4. اختبار مان – وتني "Mann Whitney Test" لاختبار الفرضيات حول الفرق بين متوسطي مجتمعين في حالة العينات المستقلة.
5. اختبار كروسكال – والاس "Kruskal-Wallis Test"لاختبار فرضيات لمقارنة متوسطات عدة مجتمعات مستقلة (تحليل التباين في حالة العينات المستقلة).
6. اختبار فريدمان "Friedman Test"الذي يعالج موضوع تحليل التباين في حالة المشاهدات المتكررة (Repeated Measures)أو العينات المرتبطة .

اختبار التوزيع الطبيعي: كولمجروف– سمرنوف

مثال(1):

تمثل البيانات التالية كمية الإنتاج بالطن لسلعة ما في الأسبوع في أحد المصانع:

74 83 94 68 76 60 90 70 80 90 80 68 82 79 65 50 70
60 92 82 68 93 71 86 92 90 80 82 65 76

المطلوب: استخدم اختبار كولمجروف – سمرنوف لمعرفة أن البيانات السابقة لها توزيع طبيعي أم لا مستخدماً مستوى دلالة $\alpha = .05$.

الحل العملي:

One-Sample Kolmogorov-Smirnov Test

		QUANTITY
N		30
Normal Parameters [a,b]	Mean	77.2000
	Std. Deviation	11.3058
Most Extreme Differences	Absolute	.105
	Positive	.075
	Negative	-.105
Kolmogorov-Smirnov Z		.573
Asymp. Sig. (2-tailed)		.898

a. Test distribution is Normal.

b. Calculated from data.

من الجدول السابق: Sig.=.898 لذلك لا يمكن رفض فرضية العدم القائلة بأن كمية الإنتاج لها توزيع طبيعي وذلك على مستوى دلالة $\alpha = .05$.

اختبار ويلكوكسن "Wilcoxon Test"

مثال(2):

لمعرفة تأثير إشارة ضوئية جديدة، تم الحصول على البيانات التالية التي تمثل عدد الحوادث في 12 مفترق خطر خلال أربعة أسابيع قبل وبعد تركيب الإشارة الضوئية:

(2,1), (3,2), (2,0), (1,3), (2,1), (6,3), (5,3), (4,1), (5,2), (3,2), (2,3), (4,2).

اختبر الفرضية المبدئية القائلة أنه لا يوجد تأثير للإشارة الضوئية الجديدة مستخدماً مستوى دلالة $\alpha = .05$.

الحل العملي:

من الجدول السابق Sig.=.039 لذلك نرفض فرضية العدم القائلة بأنه لا يوجد تأثير للإشارة الضوئية الجديدة مستخدماً مستوى دلالة $\alpha = .05$، ونستنتج أنه يوجد تأثير للإشارة الضوئية الجديدة، بمعنى أن معدل عدد الحوادث قد تناقص بعد تركيب الإشارة الضوئية الجديدة.

اختبار مان – وتني "Mann Whitney Test"

مثال(3):

قامت إحدى الشركات بتدريب بعض عمالها على العمل على آلات جديدة وردت إلى مصانع الشركة، واستخدمت برنامجين للتدريب، البرنامج الأول محاضرات نظرية لمدة أسبوعين ومن ثم القيام بالتدريب العملي، والبرنامج الثاني محاضرات نظرية تتبعها تطبيقات عملية في نفس اليوم ولمدة أسبوعين. وكان الزمن اللازم للمتدربين لاكتساب المهارات المطلوبة مقدرة بالأيام كما يلي:

البرنامج الأول	البرنامج الثاني
40	29
44	27
33	32
26	25
31	27
29	28
34	31
31	23
38	37
33	28
42	22
35	31
	24

هل تستطيع أن تستنتج أن البرنامج الثاني أكثر فاعلية من البرنامج الأول على مستوى دلالة $\alpha = .05$؟

الحل العملي:

Ranks

	FACTOR	N	Mean Rank	Sum of Ranks
TIME	first	12	17.54	210.50
	second	13	8.81	114.50
	Total	25		

من النتائج السابقة: Sig.=.002 لذلك نرفض فرضية العدم القائلة بأنه لا يوجد فرق بين فاعلية البرنامجين، ونستنتج بأنه يوجد فرق بين فاعلية البرنامجين، أو يمكن القول بأن البرنامج الثاني أكثر فاعلية من البرنامج الأول على مستوى دلالة $\alpha = .05$.

اختبار كروسكال – والاس "Kruskal–Wallis Test"

مثال(4):

البيانات التالية تمثل درجات طلاب مساق تحليل البيانات باستخدام ثلاثة أساليب مختلفة:

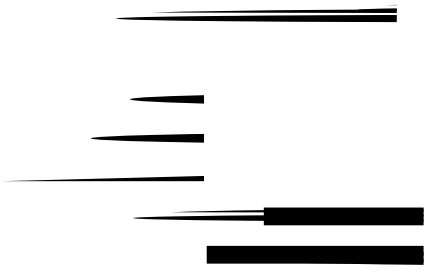
الأسلوب الأول	الأسلوب الثاني	الأسلوب الثالث
86	82	75
81	66	78
84	69	61
71	72	69
81	67	75
88	68	
79	77	
77		

هل تعطي هذه البيانات دليلاً كافياً على وجود فروق معنوية بين متوسطات درجات الطلاب باستخدام الأساليب الثلاثة السابقة على مستوى دلالة $\alpha = .05$؟

الحل العملي:

Ranks

	FACTOR	N	Mean Rank
MARK	Method_1	8	15.06
	Method_2	7	7.29
	Method_3	5	7.70
	Total	20	



من النتائج السابقة $\chi^2 = 7.968$، Sig.=.019 لذلك نرفض فرضية العدم القائلة بأنه لا يوجد فرق بين درجات الطلاب باستخدام الأساليب الثلاثة ، ونستنتج بأنه يوجد فرق بين درجات الطلاب في مساق تحليل البيانات باستخدام الأساليب الثلاثة على مستوى دلالة $\alpha = .05$.

اختبار فريدمان "Friedman Test"

مثال(5):

البيانات التالية تمثل زمن الشفاء (مقدراً بالأيام) من مرض معين عند تناول المرضى أربعة أنواع مختلفة من الأدوية.

النوع الأول	النوع الثاني	النوع الثالث	النوع الرابع
10	7	11	13
8	13	6	10
7	15	11	9
11	11	9	14
9	12	8	11
7	8	7	12
8	14	5	10
11	10	10	13

هل تعطي هذه البيانات دليلاً كافياً على وجود فروق معنوية بين متوسطات الزمن باستخدام أنواع الأدوية الأربعة على مستوى دلالة $\alpha = .05$؟

الحل العملي:

Ranks

	Mean Rank
TYPE_1	2.00
TYPE_2	3.00
TYPE_3	1.63
TYPE_4	3.38

Test Statistics[a]

N	8
Chi-Square	10.130
df	3
Asymp. Sig.	.017

a. Friedman Test

من النتائج السابقة: $\chi^2 = 10.13$، Sig.=.017 لذلك نرفض فرضية العدم القائلة بأنه لا يوجد فرقاً بين أنواع الأدوية الأربعة، ونستنتج بأنه يوجد فرق بين تلك الأنواع السابقة على مستوى دلالة $\alpha = .05$.

الارتباط الخطي البسيط Simple Linear Regression

في معظم التطبيقات العملية نجد أن هناك علاقة بين متغيرين (أو أكثر)، فمثلاً نجد أن هناك علاقة وارتباط بين درجة الطالب وعدد ساعات الدراسة. يوجد نوعان من المتغيرات هما:

المتغير التابع Dependent (Response) Variable: هو المتغير الذي يقيس نتيجة دراسة ما، وعادة يرمز له بالرمز Y.

المتغير المستقل Independent (Explanatory) Variable:

هو المتغير الذي يُفسِّر أو يسبب التغيرات في المتغير التابع، أي هو الذي يؤثر في تقدير قيمة المتغير التابع، وعادة يرمز له بالرمز X. فمثلاً عدد أيام الغياب X و درجة الطالب في الإحصاء Y، العُمر X والإصابة بضغط الدم Y.

في بعض التطبيقات العملية يكون لدينا أكثر من متغيرين تحت الدراسة، فمثلاً قد توجد علاقة خطية بين ضغط الدم وكل من العُمر والوزن، ويسمى الارتباط في هذه الحالة الارتباط الخطي المتعدد.

عند دراسة العلاقة بين متغيرين X, Y فإن شكل الانتشار Scatter plot يمكن أن يوضح طبيعة هذه العلاقة، وتكون العلاقة بين X, Y قوية جداً إذا وقعت معظم نقاط شكل الانتشار على منحنى أو خط مستقيم، وتكون ضعيفة كلما تناثرت نقاط شكل الانتشار حول منحنى أو خط مستقيم يمر بتلك النقاط.

معامل الارتباط Correlation Coefficient:

هو مقياس لدرجة العلاقة بين المتغيرين X, Y ويرمز له بالرمز r، ويحقق معامل الارتباط الخطي المتباينة:

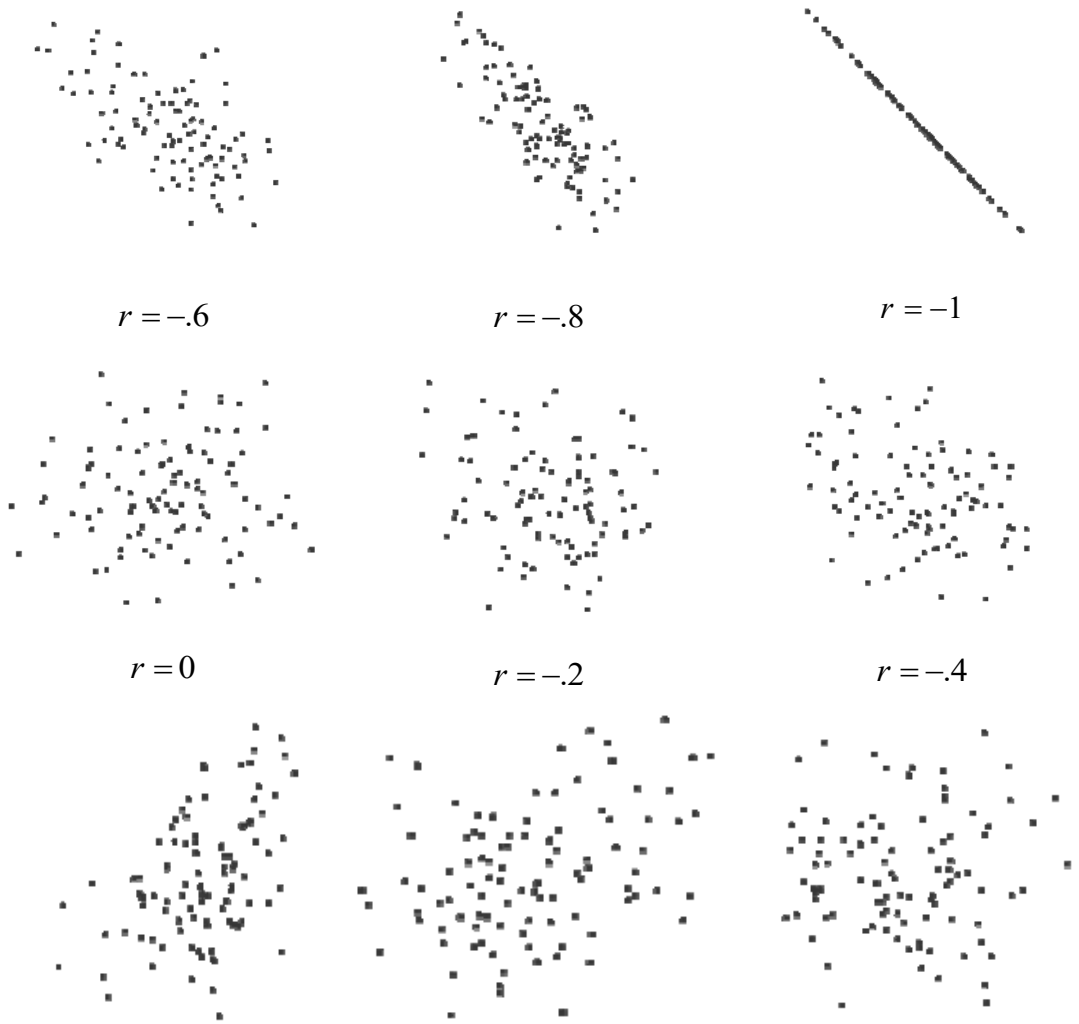
$$-1 \le r \le 1$$

أي أن قيمة معامل الارتباط محصورة بين 1−، 1+ وتدل قيمته على درجة العلاقة بين المتغيرين أو المتغيرات موضع الدراسة من حيث أنها قوية، متوسطة، أو ضعيفة، وأما الإشارة فإنها تصف نوعية العلاقة هل هي عكسية أم طردية، فالإشارة السالبة تدل على وجود علاقة عكسية أما الموجبة فتدل على وجود علاقة طردية بين المتغيرين موضع الدراسة.

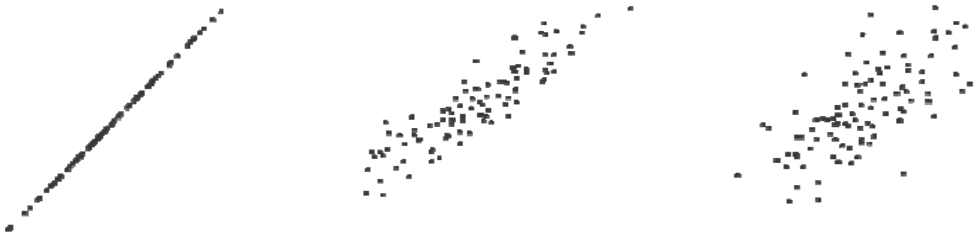
- إذا كانت قيمة معامل الارتباط مساوية للواحد الصحيح فهذا يدل على أن الارتباط بين المتغيرين ارتباطاً طردياً تاماً، أما إذا كانت قيمته مساوية لـ 1− فهذا يدل على أن الارتباط بين المتغيرين ارتباطاً عكسياً تاماً.
- إذا كانت قيمة معامل الارتباط مساوية للصفر(r = 0) فهذا يدل على عدم وجود ارتباط خطي بين المتغيرين موضع الدراسة، بمعنى أنه إذا عرفنا اتجاه تغير أحد المتغيرين استحال علينا تحديد أو معرفة اتجاه المتغير الآخر.

- أما إذا ابتعدت بعض نقاط شكل الانتشار عن الخط المستقيم فإن الارتباط يكون غير تاماً، وتزداد قوة الارتباط كلما اقتربت قيمة r من القيمة 1+ أو القيمة 1−. فمثلاً الطول والوزن لمجموعة من الأشخاص قد يوجد بينها ارتباطاً طردياً ولكن ليس ارتباطاً تاماً. العلاقة بين X, Y تكون:
  - طردية ضعيفة عندما $0 < r < \frac{1}{2}$.
  - طردية متوسطة عندما $\frac{1}{2} \le r < \frac{3}{4}$.
  - طردية قوية عندما $\frac{3}{4} \le r < 1$
  - عكسية ضعيفة عندما $-\frac{1}{2} < r < 0$
  - عكسية متوسطة عندما $-\frac{3}{4} < r \le -\frac{1}{2}$
  - عكسية قوية عندما $-1 < r \le -\frac{3}{4}$

برسم لوحة الانتشار لقيم مختارة من معاملات الارتباط الخطي يمكن الحصول على أحد الأشكال التالية:



$r = .5$ $r = .3$ $r = .1$



$r = .9$ $r = .7$

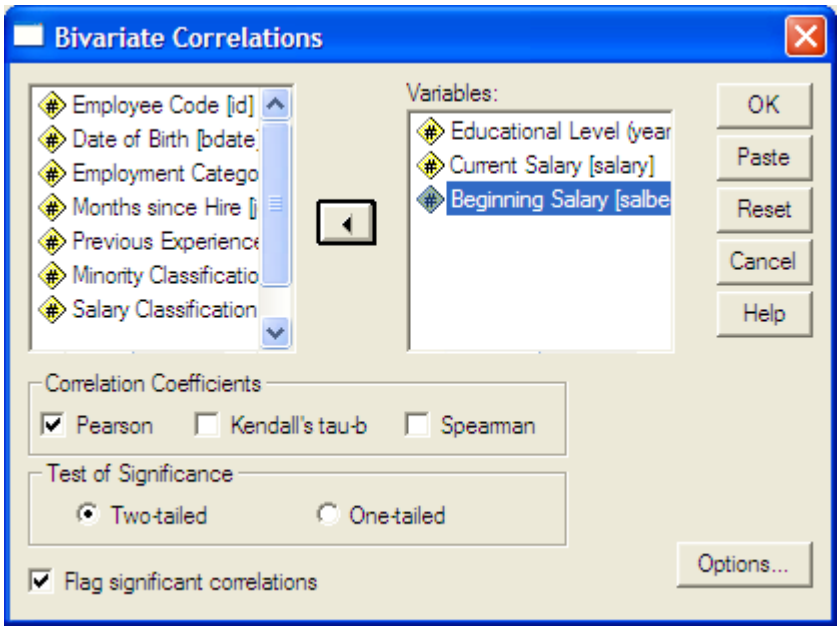
$r = 1$

حساب قيمة معامل الارتباط:

SPSS STEP BY STEP

Analyze ⇒ Correlate ⇒ Bivariate

أكمل المربع الحواري كما يلي:



من النتائج السابقة يمكن استنتاج ما يلي::

قيمة معامل الارتباط بين educ (Educational Level), salary (Current Salary) تساوي 661.

وهذا يدل على أن الارتباط بينهما طردي، Sig.=.000 مما يدل على وجود ارتباط معنوي بين المتغيرين.

قيمة معامل الارتباط بين educ (Educational Level), salbegin (Beginning Salary) تساوي 633. وهذا يدل على أن الارتباط بينهما طردي، Sig.=.000 مما يدل على وجود ارتباط معنوي بين المتغيرين.

قيمة معامل الارتباط بين salbegin (Beginning Salary), salary (Current Salary) تساوي 880.

وهذا يدل على أن الارتباط بينهما طردي، Sig.=.000 مما يدل على وجود ارتباط معنوي بين المتغيرين.

مثال (6)

فيما يلي تقديرات عشرة من طلاب في امتحان مادتي تحليل البيانات واقتصاديات وسائل الإعلام:

وسائل الإعلام	راسب	جيد	مقبول	جيد جداً	مقبول	مقبول	جيد	جيد جداً	جيد	جيد
تحليل البيانات	مقبول	جيد جداً	جيد	ممتاز	راسب	جيد	جيد جداً	ممتاز	جيد	راسب

المطلوب: احسب معامل الارتباط بين تقديرات المادتين.

بعد ادخال البيانات واتباع خطوات المثال السابق اختر Spearman فنحصل على النتائج التالية.

من النتائج السابقة يمكن استنتاج ما يلي:

معامل سبيرمان للرتب = 0.718 فبالتالي يوجد ارتباط طردي بين تقديرات الطلاب في المادتين، وذلك على أساس معامل سبيرمان للرتب. Sig.=.019 مما يدل على وجود ارتباط معنوي بين تقديرات الطلبة في مادتي اقتصاديات وسائل الإعلام وتحليل البيانات

الانحدار الخطي البسيط Simple Linear Regression

الانحدار هو دراسة للتوزيع المشترك لمتغيرين أحدهما متغير يقاس دون خطأ ويسمى متغير مستقل Independent variable ويرمز له بالرمز x والآخر يأخذ قيماً تعتمد على قيمة المتغير المستقل ويسمى التابع Dependent variable ويرمز له بالرمز y.

الهدف من دراسة الانحدار هو إيجاد دالة العلاقة بين المتغيرين المستقل والتابع والتي تساعد في تفسير التغير الذي قد يطرأ على المتغير التابع (y) تبعاً لتغير في قيم المتغير المستقل (x).

مثال (7)

لدراسة العلاقة بين الدخل والاستهلاك بالدنانير في مدينة غزة، أخذت عينة مكونة من عشرة أسر فأعطت النتائج التالية:

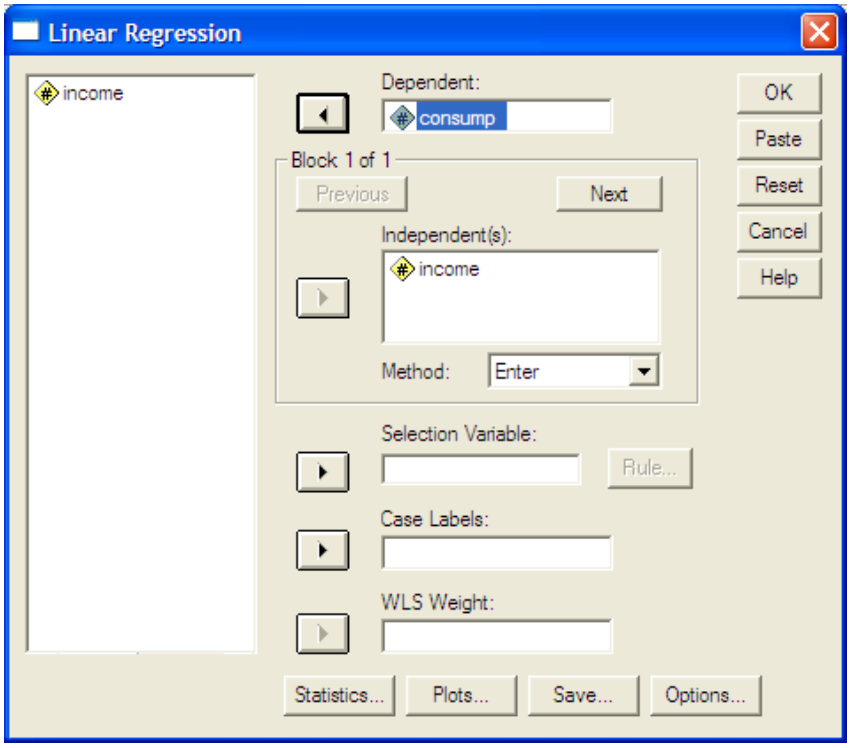
الدخل	300	350	500	600	900	1000	900	1200	1050	250
الاستهلاك	280	340	500	550	800	750	850	1050	1000	250

المطلوب: إيجاد نموذج انحدار الاستهلاك على الدخل.

SPSS STEP BY STEP

Analyze ⇒ Regression ⇒ Linear

أكمل المربع الحواري كما يلي:



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.982[a]	.965	.960	58.60899

a. Predictors: (Constant), income

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	751329.9	1	751329.889	218.727	.000[a]
	Residual	27480.111	8	3435.014		
	Total	778810.0	9			

a. Predictors: (Constant), income

b. Dependent Variable: consump

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	48.229	43.913		1.098	.304
	income	.835	.056	.982	14.789	.000

a. Dependent Variable: consump

من النتائج السابقة يمكن استنتاج ما يلي:

1- نموذج انحدار الاستهلاك على الدخل هو:

Consump. = 48.229 + 0.835 * Income

2- معامل الارتباط بين الدخل والاستهلاك = 0.982 وهو يدل على وجود ارتباط طردي قوي بينهما،($Sig. = 0.000 < \alpha = 0.05$)

3- معامل التحديد $R^2 = 0.965$ ، ومعامل التحديد المعدَّل= 0.96، الخطأ المعياري للتقدير = 58.6190.

تفسير قيمة معامل التحديد:

96.5% من تغير قيمة الاستهلاك (المتغير التابع) يمكن أن يفسر باستخدام العلاقة الخطية بين الدخل والاستهلاك والنسبة المتبقية 3.5% ترجع إلى عوامل أخرى تؤثر على قيمة الاستهلاك.

4- F= 218.727 ، $Sig. = 0.000$ وهذا يدل على وجود علاقة معنوية بين الدخل والاستهلاك وأن نموذج الانحدار السابق جيد.

5- $t_{b_0} = 1.098$ ، Sig.=0.304 وهذا يدل على أن نموذج انحدار الاستهلاك على الدخل يمر بنقطة الأصل.

6- $t_{b_1} = 14.789$ ، Sig.=0.000 وهذا يدل على أن الدخل متغير مؤثر في تقدير قيمة الاستهلاك ويجب أن يكون ضمن نموذج خط الانحدار.

المراجع

المراجع باللغة العربية:

- أحمد عامر، الإحصاءI أو الإحصاءII – دار الغرب للنشر والتوزيع، الجزائر، سنة 2020.
- جيلالي جلاطو، " الإحصاء مع تمارين ومسائل محلولة "، الطبعة الثامنة، ديوان المطبوعات الجامعية، ، الجزائر ، سنة 2010 .
- شرف الدين خليل، الإحصاء الوصفي " – شبكة الأبحاث والدراسات الاقتصادية، من الموقع الإلكتروني: www.rr4ee.net ، 2022/05/02.
- عبد الرزاق عزوز، " الكامل في الإحصاء، – دروس مفصلة، تمارين ومسائل مع الحلول"– جزئيين، ديوان المطبوعات الجامعية، ، الجزائر ، سنة 2018 .
- محمد حسين محمد رشيد، "مبادئ الإحصاء والاحتمالات ومعالجتها– باستخدام برنامج SPSS "، دار الصفاء للنشر والتوزيع– عمان، سنة 2017.
- محمد شفيق ياسين أنور اللحام، " مبادئ الإحصاء والاحتمال "، ديوان المطبوعات الجامعية، الجزائر، سنة 2020.
- موسى محمد أماني، " التحليل الإحصائي للبيانات" – مركز تطوير الدراسات العليا والبحوث كلية الهندسة –جامعة القاهرة، من الموقع الإلكتروني:

www.capscu.com , www.Pathways.cu.edu.eg ، 2020/08/01.

المراجع باللغة الأجنبية

- Grais.B « Statistique descriptive », Dumod, 2éme édition, Année :2022.
- Anderson-Sweeney-Williams « Statistique pour l'économie et la gestion », Traduction de la 5e édition américaine par Claire Borenberge, édition de boeck université, France, 2019.
- Bernard Verlant et Geneviéve Saint-Pierre « Statistique et Probabilités –Manuel de cours » , LMD,DUEA, BTS, Editions Berti, Année : 2010.
- Célyne Laliberté , « Probabilités et statistique », Distribué Par ERPI,- Québec.Année :2009.
- G . SAPORTA , Probabilités Analyse de Données Et Statistique – France,Année:2013
- Gilbert.N, « Staistique », édition H.R.W, etéé Montréal, Année :2018.
- Khaldi Khaled , « FAC Méthodes Statistiques et Probabilités » Editions CASBAH, Année : 2000.
- walder masiéri, « Cours Eco-Gestion , Statistique et Calcul des Probabilités », édition campus dalloz, France, Année : 2022.