



جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة الفرات الاوسط التقنية
المعهد التقني الكوفة / قسم تقنيات ادارة المواد
المرحلة الاولى

الأحصاء

Statistic



أستاذ المادة

ذوالفقار حسن علوان

2021م

1442

الاسبوع 1

المقدمة:

ينحدر مصطلح الإحصاء (Statistics) من أصل لاتيني (Status) و هو يعني الدولة أو القوة السياسية أو من أصل روماني (Statist) , و يعني الدولة أيضا ، لكونه مرتبطا بالشئون العسكرية و على وجه التحديد يعتبر الإحصاء من العلوم القديمة والتي ظهرت مع ظهور الإنسان على الأرض . لقد أحصى الإنسان القديم الأشياء التي تحيط به ونقشها بطريقته الخاصة على جدران الكهوف التي عاش فيها وكثيراً ما كان يحتفظ بأشياء تعود إلى طرائده ليتباهى بها كأن يحتفظ بأنياب أو ذبول هذه الطرائد بالإضافة لجلودها كوسيلة إحصائية لعددها ونوعيتها .

ولا يمكن للإنسان قديماً أو حديثاً الاستغناء عن مبدأ العد والإحصاء إن كان ذلك لإحصاء الأشياء العينية أو النقدية . وكلمة إحصاء من الناحية اللغوية تعنى "عد" أو "حسب" ، فأحصى الأشياء يعني عدها أو سجلها وحفظها . وقد ذكرت كلمة الإحصاء في القرآن الكريم عدة مرات ووردت في العديد من آياته وسوره . قال تعالى في كتابه العزيز " وأحطنا بما لديهم وأحصى كل شيء عددا " وقال أيضاً في سورة إبراهيم " وإن تعدوا نعمة الله لا تحصوها " ... وفي سور أخرى كثيرة . لقد احتل الإحصاء ، كعلم ، مركزاً مرموقاً بين العلوم الأخرى خلال القرنين الماضيين . وقد برز الإحصاء في وقتنا الحاضر كأحد أهم العلوم التطبيقية التي لا يمكن الاستغناء عنها في أي فرع من فروع الحياة ، وبرز ذلك جلياً بانتشار الشعب الإحصائية ، والتي تسمى في المؤسسات الإنتاجية

" السيطرة النوعية أو التحكم النوعي " (Quality control) ، في كل المصانع والمؤسسات الإنتاجية و الخدمية للدول المتقدمة . والهدف من إنشاء هذه الشعب الإحصائية هو مراقبة الإنتاج وحساب معدلاته ومراقبة تطوراتها والنسب المئوية للتوالف والمستهلكات ، وكذلك ضمن اهتمام هذه الشعب إحصائيات عن العمال والموظفين والزبائن والمستهلكين بالإضافة لمراقبة الإيرادات ورؤوس الأموال .

لقد اعتمد الإحصاء كعلم الجداول الإحصائية (Statistical tables) والوسائل البدائية قديماً أما حديثاً فتستخدم برامج متطورة وحواسيب ذات قدرة كبيرة لتسهيل التعامل مع الكم الهائل من الأعداد والأرقام الناتجة والتي تحتاج إلى تحليل ودراسة بالإضافة لظهور آلات حاسبة علمية تسهل هذه المهمة .

تعريف علم الإحصاء :

هو العلم أو مجموعة القواعد والطرق والنظريات التي تهتم بجمع البيانات وتبويبها وعرضها بيانياً ثم تحليلها وتفسيرها وإجراء المقارنات واستنتاج العلاقات بهدف استخدامها في اتخاذ القرارات المناسبة .

من هذا التعريف يمكن أن تستنتج عدد من الملاحظات وهي :

اولاً - أن المراحل الأساسية للعملية الإحصائية هي ٤ مراحل :

١- جمع البيانات.

٢- تبويب البيانات .

٣ - العرض البياني للبيانات .

٤- تحليل البيانات.

ثانيا - الهدف الأساسي من العملية الإحصائية هو تحليل البيانات وتفسيرها.

ثالثا - يمكن تطبيق عملية الإحصاء في مختلف المجالات.

رابعا - أن البيانات هي المجال الرئيسي لمراحل علم الإحصاء

الطريقة الإحصائية في البحث العلمي

- ١- تحديد المشكلة ووضع الفرضيات.
- ٢- جمع البيانات والمعلومات ذات العلاقة بالبحث.
- ٣- تصنيف وتبويب وعرض البيانات.
- ٤- حساب المؤشرات الإحصائية.
- ٥- تحليل وتفسير النتائج على ضوء فرضية البحث .
- ٦- تفسير النتائج واتخاذ القرار.

أساليب جمع البيانات

١- **أسلوب التسجيل الشامل:** يتم جمع البيانات هنا عن طريق تسجيل مفردات المجتمع (مفردة، مفردة) لذلك فإن هذا الأسلوب يتطلب وقت وجهد ويفترض عند استخدام هذا الأسلوب إن يكون المجتمع محدود مثل تعداد السكان أو حصر عدد العاملين في أجهزة الدولة ويمتاز هذا الأسلوب بكونه يتعامل مع جميع بيانات مجتمع البحث مما يؤدي للحصول على نتائج دقيقة .

أسلوب العينات : وهي عملية جمع البيانات والمعلومات عن مجموعة معينة من مفردات المجتمع مثل استفتاء بعض الأشخاص عن موضوع معين وهذا الأسلوب يحتاج إلى وقت وجهد وموارد أقل مما يحتاجه الأسلوب الأول بالإضافة إلى أنه مفيد في حالة دراسة المجتمعات الغير محدودة ولكن دقة النتائج هنا تكون أقل كونه لا يتعامل مع جميع المفردات.

الأخطاء الشائعة في جمع البيانات

- ١- **خطأ التحيز:** يحدث هذا الخطأ نتيجة جمع البيانات من مصادرها غير الرئيسية أو التحيز لمفردة معينة في المجتمع قيد الدراسة دون الأخرى لسبب أو لآخر .
- ٢- **خطأ الصدفة:** يحدث هذا الخطأ نتيجة اعتماد الباحث على معلوماته الشخصية أو جمع البيانات ناقصة .

تصنيف البيانات

هي عملية تقسيم البيانات الى مجاميع صغيرة او اصناف على اساس قاعدة معينة كاشتراكها في بعض الصفات او الخصائص كالمهنة او الحالة الاجتماعية وحسب متطلبات البحث.

عرض البيانات

التوزيعات التكرارية : هي جداول لعرض البيانات في اعمدة احدها يشمل الفئات والثاني يشمل تكرار هذه الفئات وتقسم هذه الجداول بعدة صور :

١. تبعا لإمكانية معرفة او تحديد الحدود العليا والدنيا لفئات البيانات الى :-
 - a. جداول مغلقة : وهي الجداول التي تم تحديد حدودها العليا والدنيا .
 - b. جداول مفتوحة : وهي الجداول التي يمكن ان يكون أي من حديها مفتوح سواء الأدنى او الأعلى وقد يكون سبب عدم تحديد هذه الحدود التكلفة العالية لتحديد هذه النقاط مثل تحديد العمر الافتراضي للمبات النيون وفي هذه الحالة يمكن تحديد هذه الحدود عن طريق قيم تقديرية او عن طريق خبرات سابقة او يمكن تركها غير محددة ويكون الجدول مفتوح ويكون له طرق تحليل تخص الجداول المفتوحة.

جدول مغلق	
الفئات	التكرار
0-5	10
5-10	20
10-15	10
15-20	20
20-25	10
25 -30	20

جدول مفتوح	
الفئات	التكرار
-5	10
5-10	20
10-15	10
15-20	20
20-25	10
25 -	20

٢. تبعا لانتظام الفئات الى :

- a. جداول منتظمة : وهى الجداول التي يكون فيها طول الفئة ثابت وطول الفئة هو الفرق بين الحد الاقصى والادنى للفئة
b. جداول غير منتظمة : وهى الجداول التي يكون فيها طول الفئات غير متساو.

جدول غير منتظم	
الفئات	التكرار
0-5	10
5-8	20
8-15	10
15-23	20
23-25	10
25 -30	20

جدول منتظم	
الفئات	التكرار
0-5	10
5-10	20
10-15	10
15-20	20
20-25	10
25 -30	20

تحليل البيانات :

إذا نظرنا إلى بيانات أي ظاهرة مدروسة ، نجد أن أكثر هذه المفردات تتجمع حول قيمة ما تمثل متوسط الظاهرة، ونجد أيضا بأن هناك تناقصا تدريجيا في عدد المفردات كلما بعدت هذه القيم عن المتوسط ومن الجانبين ، أن هذا التجمع حول قيمة متوسطة هو ما يسمى بالنزعة المركزية، أي ميل المفردات إلى الاقتراب والتجمع حول قيمة معينة هي القيمة المتوسطة، وكأن القيمة المتوسطة تحاول شد البيانات لكي تتراكم حولها، ولكن لا تحدث هذه الظاهرة في جميع التوزيعات التكرارية، فمن الطبيعي ألا تنجح جميع المفردات من الاقتراب من القيمة المتوسطة، فهناك بيانات تبتعد بشكل بسيط عن المتوسط، وهناك مفردات أقل تبتعد بشكل أكبر عن المتوسط، وهكذا نجد أن العدد الذي يختلف عن المتوسط اختلافا كبيرا جدا يكاد يتلاشى ، بناء على ذلك نلاحظ أن أساليب التحليل الإحصائي للبيانات هي مقاييس النزعة المركزية ومقاييس التشتت، حيث أن مقاييس النزعة المركزية هي التي بواسطتها يمكن إيجاد تلك القيمة التي تتمحور حولها كل القيم، بينما مقاييس التشتت هي التي توجد حالة الانتشار أو بعد المعطيات عن المتوسط.

الاسبوع 2

مصادر البيانات :

المصدر الأول : المصادر التاريخية للبيانات :

وتشمل الإحصاءات والنشرات الإحصائية التي تصدرها المؤسسات الحكومية والأهلية لتبين أوجه التغير والتطور في المجال الذي تختص به هذه المؤسسات . وهذه البيانات هي من مسؤولية هذه الشركة أو المؤسسة .
- ونلاحظ على هذا المصدر ٤ ملاحظات :

١. عدم توفر جميع البيانات . (لا تغطي جميع أوجه البحث)
٢. قد تكون قديمة . (لم تحدث)
٣. قد تكون غير دقيقة ، وقد تكون من جهة غير موثوق فيها (الغرض من البيانات الدعاية فقط) .
٤. قد لا تكون البيانات المنشورة تغطي جميع جوانب البحث .

المصدر الثاني : المصادر الميدانية للبيانات :

يلجأ الباحث لجمع البيانات بنزوله للميدان يبحث هو أو عن طريق الاستبيانات الخاصة بذلك يحصل منهم على البيانات مباشرة ، ويمكن أن يشمل أفراد أو أجزاء أو أن يشمل عينة من الإطار .
(يقوم هو بنفسه بجمع البيانات وتعبئه الاستبيانات)

١. المقابلة الشخصية: حيث يقوم الباحث بالاتصال المباشر بالأشخاص محل الدراسة وبهذا يستطيع الباحث ان يحقق أعلى درجات الدقة في جمع البيانات الا ان هذه الطريقة ورغم ما تمتاز به من دقة المعلومات قد تكون مكلفة ورسما في حالة العينات الكبيرة الحجم.

٢. الاستمارة الاحصائية: وفيها يقوم الباحث بتصميم استمارة تشمل العديد من الاسئلة الرئيسية والفرعية التي تحقق جميع الاهداف للبحث محل الدراسة ويراعى في هذه الاستمارة الشروط التالية:-

- a. يجب ان تكون الاسئلة واضحة وسهلة.
 - b. يجب ان تكون الاستمارة غير طويلة وذلك حتى لا يصاب الشخص بالملل والضيق.
 - c. يجب التأكيد على سرية البيانات للشخص محل الدراسة حتى لا تكون اجاباته بعيدة عن الواقع.
 - d. يجب ان تشمل الاستمارة على بعض الاسئلة المكررة في اكثر من موضوع تحقيقا لمصادقية الشخص محل الدراسة.
 - e. يجب ان تحقق الاستمارة الاهداف محل الدراسة.
- يجب على الباحث بعد اعداد الاستمارة ان يقوم باختبارها عن طريق اخذ عينة صغيرة جدا وذلك للتحقق من جميع الشروط السابقة.

أساليب البحث الميداني :

إذا قرر الباحث ان يبحث بنفسه فله أن يختار أسلوبين :

١. الحصر الشامل .
٢. العينات .

الأسلوب الأول : تعريف المجتمع الإحصائي (الحصر الشامل) :

وهو جميع المفردات التي يجمعها إطار معين ، أو مجموعة من الخصائص المشتركة العامة . (مثل عدد خلايا في شعبة) يلاحظ علىه :

- a. الشمول
- b. تنوع المجتمعات الإحصائية : (بشري – نباتي - حيواني)
- c. المجتمع الإحصائي قد يكون محدود أو غير محدود . (محدود بقاعه معينه بكلية وقد يكون غير محدود بمجتمع معين)
- d. جميع أفراد المجتمع يجمعهم إطار معين وخصائص معينة .

الأسلوب الثاني : للبحث الميداني (العينات) :

تعريف العينة: هي جزء أو شريحة من المجتمع تتضمن خصائص المجتمع الأصلي الذي نرغب في التعرف على خصائصه ويجب أن تكون تلك العينة ممثلة لجميع مفردات هذا المجتمع تمثيلاً صحيحاً .

والعينة هي جزء من المجتمع ونقوم بدراستها للتعرف على خصائص المجتمع التي سحبت منه هذه العينة - ولكي تصلح النتائج التي نحصل عليها للتعبير عن المجتمع لا بد وان تكون العينة ممثلة للمجتمع (أي جميع المفردات المراد بحثها) تمثيلاً صحيحاً .

واستخدام العينات معروف منذ القدم ونشاهد له أمثلة عديدة في الحياة العملية فالكيميائي في معمله يقوم بدراسة خواص المادة من واقع عينة من هذه المادة والطبيب يقوم بتحليل دم المريض من واقع عينة صغيرة تتكون من بضعة نقاط من دمه الخ

ويتم إتباع دراسة العينات وأسلوب المعاينة وذلك اختصاراً للوقت وتوفيراً للجهد والنفقات ولرفع مستوى العمل البحثي وجعله أكثر دقة وذلك لأن دراسة عدد قليل من المفردات أو الحالات يتيح للباحث فرصة جمع معلومات دقيقة وكثيرة عن كل مفردة أو حالة .

طرائق الحصول على البيانات الإحصائية

يتم الحصول على البيانات الإحصائية حسب الهدف من الدراسة وأسلوب التحليل إما من خلال مصادر رسمية متخصصة (سجلات رسمية وتقارير ونشرات دورية تصدرها جهات رسمية) أو من مصادر ميدانية مباشرة (قائمة علي تجربة أو قائمة علي مشاهدة).

أساليب جمع البيانات الميدانية

أولاً - طريقة المسح الشامل : حيث تعتمد هذه الطريقة على المسح و دراسة جميع أفراد المجتمع المراد دراسته ، ومما لاشك فيه أن هذه الطريقة تتميز بالدقة ولكنها شاقة ومتعبة وباهظة التكاليف .

ثانياً - طريقة العينة الإحصائية : (Statistical sample) : وتعتمد هذه الطريقة على جمع المعلومات عن طريق دراسة جزء من المجتمع الإحصائي يمثل هذا المجتمع تمثيلاً صادقاً . وللعينة الإحصائية عدة أنواع منها :

- العينة الطبقية : حيث يقسم المجتمع إلى طبقات معينة وتحسب عينات كل طبقة من الطبقات .
 - العينة المنتظمة : تسحب بشكل منتظم من المجتمع حيث يؤخذ مضاعفات الرقم 10 مثلاً أو 20.
 - العينة العشوائية : تسحب بشكل عشوائي وتخلط عناصر المجتمع في كل مرة يتم فيها السحب .
- ومن الشروط الهامة التي يجب أن تتميز بها العينة هي الموضوعية والشمولية والاستقلالية وعدم التمييز .

ويجب أن نذكر هنا بأنه عند جمع المعلومات ميدانياً أو عند اختيار عينات ضمن الشروط الواردة أعلاه ، يجب أعداد استمارات خاصة بالبحث وتتضمن أسئلة متعلقة بمهمة البحث حيث يطلب من أعضاء المجتمع الإجابة عن هذه الأسئلة باختصار . ويراعى في وضع أسئلة البحث السهولة والإيجاز بعيداً عن الشرح والتفصيل ، ويفضل إن أمكن أن يجاب عن هذه الأسئلة بنعم أو لا أو الإجابة برقم مثل

الإجابة على عدد أفراد الأسرة أو متوسط الدخل الشهري ، الطول ، الوزن ... و تسمى هذه النشرات استبيان
وسائل جمع البيانات
- الاستمارة الاحصائية
- المقابلة الشخصية

مزاياه :

١. توفير الوقت والجهد .
٢. يوفر عليك عملة الإتلاف .وتعمل على جزء معين .

عيوبه :

١. عدم دقة النتائج (قد يكون لديك ٣ آلاف طالب وأنت تعمل البحث على 100 طالب، فالمعلومات هنا قد لا تكون دقيقة)
٢. قد لا تكون نهائية.

مثال على الاستبيان:

بسم الله الرحمن الرحيم

عزيزي الطالب بهدف هذا الاستبيان إلى معرفة آرائكم حول توفير خدمة الإنترنت للاستخدام الحر، فيرجى التعاون بالرد على هذا الاستبيان لما فيه مصلحة الجميع.

الرقم	الأسئلة	نعم	لا	قليل
1	هل لديك علم بالإنترنت ؟			
2	هل تود تعلم الإنترنت ؟			
3	هل تستخدم الإنترنت دائماً ؟			
4	هل تعتبر استخدام الإنترنت مفيداً ؟			
5	هل تعتبر استخدام الإنترنت إهداراً للوقت ؟			
6	هل تؤيد توفير الإنترنت للاستخدام الحر؟			
7	هل لديك وقت فراغ لاستخدام الإنترنت ؟			
8	هل سيساعدك في البحث عن المعلومات ؟			
9	هل تفتح مواقع غير مفيدة ؟			
10	هل تؤيد وجود مشرف ؟			

الاسبوع 3

عرض البيانات :

تتوقف طريقة عرض البيانات على نوع هذه البيانات وعلى الحقائق المطلوب إبرازها. وهناك طريقتان أساسيتان لعرض وتبويب البيانات الإحصائية وهما :

أولاً : العرض الجدولي للبيانات الإحصائية :

بعد عملية تبويب وتعيين الصفات التي تميز المفردات ، ترصد النتائج في جداول مناسبة توضح الشكل النهائي للمجموعات المميزة وتسمى هذه العملية التي يتم تجميع البيانات في مجموعات مميزة ومتجانسة بعملية التصنيف وتصنف البيانات الإحصائية بوجه عام وفقاً لإحدى القواعد التالية :

- ١- تصنيف جغرافي
- ٢- تصنيف تاريخي أو زمني .
- ٣- تصنيف نوعي أو وصفي .
- ٤- تصنيف كمي .

ويمكن التمييز بين مجموعة أشكال من الجداول الإحصائية نذكرها فيما يلي :

تبويب البيانات الخام في جدول تكراري بسيط :

والمقصود بالجدول البسيط هو ذلك الجدول الذي يتم وضع قيم الدرجات فيه مرتبة ترتيباً تصاعدياً في عموده الأول أما العمود الثاني فيسمى بعمود التكرار ويرصد فيه عدد مرات تكرار كل درجة أو حدث .

مثال : البيانات التالية هي درجات حصل عليها عشرون طالباً في مادة الإحصاء الاجتماعي بالفرقة الأولى قسم الاجتماع في امتحان نهاية العام :

12	11	15	14	12	10	15	13	12	10
14	10	13	12	15	13	12	10	12	15

والمطلوب تبويب هذه البيانات في جدول توزيع تكراري بسيط ؟

الحل :

يتم ترتيب البيانات دون تكرار تصاعدياً ثم وضع هذه البيانات في العمود الأول من الجدول وتسمى (x) ثم وضع عدد مرات التكرار باستخدام العلامات في العمود الثاني أما العمود الثالث فيمثل التكرار ويرمز له بالرمز (f) .

العلامات	التكرار f	x
////	4	10
/	1	11
/ ///	6	12
///	3	13
//	2	14
////	4	15
sum	20	

مثال :

البيانات التالية هي تقديرات 20 طالباً في مادة الإحصاء بالفرقة الأولى لقسم الاجتماع في العام الجامعي 2005 / 2006 والمطلوب هو وضع هذه البيانات في جدول بسيط ؟

جيد جداً	جيد	مقبول	جيد جداً	جيد	مقبول	جيد	جيد	مقبول	جيد
مقبول	جيد	جيد	ممتاز	جيد	مقبول	جيد جداً	ممتاز	جيد	ممتاز

الحل :

التقدير	التكرار f
مقبول	5
جيد	9
جيد جداً	3
ممتاز	3
المجموع	20

تبويب البيانات في جدول تكراري ذو فئات :

قبل التعرض إلى إعداد هذا الجدول سنقوم أولاً بالتعرف على معنى الفئات وطرق كتابتها .

المقصود بالفئات :

الفئة هي مجموعة من البيانات متشابهة إلى حد كبير جداً في الصفات ، وفي حالة زيادة عدد البيانات الخام التي يتم الحصول عليها من الاستبيان لا يمكن استخدام الجداول البسيطة في التعبير عن هذه الحالات وإلا سنحتاج إلى مئات الصفحات ، وإنما يتم تقسيم البيانات إلى مجموعات متقاربة ومتشابهة في الصفات تسمى فئات .

طرق كتابة الفئات : يوجد عدة طرق لكتابة الفئات هي :

الطريقة الأولى: نذكر كلا من الحد الأدنى والحد الأعلى للفئة كما بالجدول التالي :

X	التكرار f
20 - 10	5
30 - 20	20
40 - 30	50
50 - 40	25

وتتطرق الفئة الأولى مثلاً (من 20 إلى 30) وليس (20 شرطة 30) وهذه الطريقة معيبة لأن نهاية الفئة الأولى هي نفسها بداية الفئة الثانية وهكذا وفي هذه الحالة لا نعرف إلى أي فئة ينتمي هذا الرقم .

الطريقة الثانية :

نذكر كلا من الحد الأدنى والحد الأعلى للفئة ولكن نقوم بترك فاصل مقدراه الواحد الصحيح بين نهاية الفئة الأولى وبداية الفئة الثانية وهكذا كما بالجدول التالي .

X	التكرار f
19 -10	5
29 -20	20
39 -30	50
49 -40	25

ويعاب على هذه الطريقة أنها لا تصلح في حالة البيانات التي تحتوى على كسور .

الطريقة الثالثة :

نذكر الحد الأدنى فقط للفئة ونضع بعده شرطة وتنطق الفئة الأولى مثلاً (10 إلى أقل من 20) وهذه الطريقة تصلح لكافة الظواهر.

X	التكرار f
-10	5
-20	20
-30	50
-40	25

الطريقة الرابعة :

نذكر الحد الأعلى فقط للفئة ونضع قبله شرطة وتنطق الفئة الأولى مثلاً (أكثر من صفر الى ٢٠) وهذه الطريقة تصلح لكافة الظواهر أيضاً ولكنها أقل شيوعاً .

X	التكرار f
20 -	5
30 -	20
40 -	50
50 -	25

خطوات بناء جدول التوزيع التكراري ذو الفئات :

١- حساب المدى = أكبر قيمة - أصغر قيمة

٢- حساب عدد الفئات $3.3 \cdot \log(n) + 1$

$$N.C = 2.5 \cdot \sqrt[4]{n}$$

حيث n هي حجم العينة، و إذا كان عدد الفئات عددا كسريا يتم تقريبه إلى أقرب عدد صحيح يليه مباشرة.

٣- حساب طول الفئة = المدى / عدد الفئات $\text{Range} = \text{Maximum} - \text{Minimum}$

٤- اختيار بداية الفئة الأولى أي الحد الأدنى لها مساوى لأقل قيمة موجودة بالبيانات أو أقل بقليل منها فمثلاً تكون من الأرقام الصفرية لتسهيل الحسابات بعد ذلك .

٥. تعيين الحد الأدنى للفئة الأولى على أن يكون هذا الحد أصغر من أو يساوي أصغر قيمة في البيانات.

٦. تعيين الحد الأدنى للفئة الثانية كالتالي:

الحد الأدنى للفئة الثانية = الحد الأدنى للفئة الأولى + طول الفئة

٧. تعيين الحدود الدنيا و الحدود العليا لباقي الفئات بنفس الطريقة.

ملاحظة:

الحد الأعلى للفئة الأولى هو نفسه الحد الأدنى للفئة الثانية و الحد الأعلى للفئة الثانية هو نفسه الحد الأدنى للفئة الثالثة و هكذا.....

حجم العينة يساوي دائما مجموع التكرارات أي أن $\Sigma f = n$
٨- بناء الجدول ووضع العلامات التي تمثل التكرار .

مثال: قام باحث بجمع بيانات تمثل درجات اختبار مادة الحاسب الآلي لخمسين طالبا من طلاب المرحلة الثانية من الثانوية العامة في الجدول التالي :

50	26	48	88	46	55	39	63	59	36	25	57	42	51	55	70
25	35	30	40	20	55	64	54	49	45	65	53	63	47	60	45
42	65	61	45	53	52	78	52	41	42	75	55	82	39	65	33
														58	64

والمطلوب هو إعداد جدول توزيع تكراري ذو فئات للجدول السابق؟

الحل:

- المدى = أكبر قيمة - أصغر قيمة = $88 - 20 = 68$
- عدد الفئات =

$$\text{عدد الفئات} = 2.5\sqrt{n}$$

$$\text{عدد الفئات} = 2.5\sqrt{50} = 6.65 = 7$$

- نقرب عدد الفئات لأقرب رقم صحيح فتكون عدد الفئات = 7
- طول الفئة = المدى / عدد الفئات = $68 / 7 = 9.7$
- نقرب طول الفئة لأقرب رقم صحيح فتصبح طول الفئة = 10
- نختار بداية الفئة الأولى وهو أصغر رقم = 20
نبدأ في بناء الجدول كالتالي :

التكرار f	العلامات	X
4	////	20 -
6	/-////	30 -
12	//-////-////	40 -
14	////-////-////	50 -
9	////-////	60 -
3	///	70 -
2	//	80 - 90
50	المجموع	

تبويب البيانات في الجدول التكراري المتجمع الصاعد :

ويقصد بالتكرار المتجمع الصاعد هو تجميع تكرار كل فئة على جميع التكرارات السابقة لها بحيث يكون مجموع التكرار التصاعدي للفئة الأخيرة مساوي لمجموع التكرارات .

مثال: من نفس بيانات المثال السابق كون جدول التكرار المتجمع الصاعد.

الحل :

بنفس الخطوات السابقة نكون جدول التوزيع التكراري ذو الفئات ومنه نكون جدول التوزيع التكراري المتجمع الصاعد كالتالي :

التكرار المتجمع الصاعد f	X
صفر	أقل من 20
4	أقل من 30
10	أقل من 40
22	أقل من 50
36	أقل من 60
45	أقل من 70
48	أقل من 80
50	أقل من 90

تبويب البيانات في الجدول التكراري المتجمع الهابط :

ويقصد بالتكرار المتجمع الهابط هو تجميع تكرار كل فئة على جميع التكرارات التالية لها بحيث يكون مجموع التكرار التنازلي للفئة الأولى مساوياً لمجموع التكرارات .

مثال : من نفس بيانات المثال السابق كون جدول التكرار المتجمع الهابط

الحل :

بنفس الخطوات السابقة نكون جدول التوزيع التكراري ذو الفئات ومنه نكون جدول التوزيع التكراري المتجمع الصاعد كالتالي :

التكرار المتجمع الهابط f	حدود الفئات
50	20 فأكثر
46	30 فأكثر
40	40 فأكثر
28	50 فأكثر
14	60 فأكثر
5	70 فأكثر
2	80 فأكثر
0	90 فأكثر

التوزيع التكراري المزدوج

وهو الجدول الذي يربط بين متغيرين في نفس الوقت وكل متغير منهم له فئاته فيتم بناؤه بإتباع عدة خطوات هي :

- ١- تحديد المتغيرين
- ٢- تحديد المتغير المستقل والمتغير التابع
- ٣- تحديد فئات كل من المتغيرين
- ٤- تكوين الجدول بحيث يحتل المتغير المستقل أعلى الجدول أي يكون أفقياً أما المتغير التابع فيحتل الجزء الأسفل أي يكون عمودياً.
- ٥- وضع العلامات التي تمثل التكرار.
- ٦- إعادة كتابة الجدول بالأرقام .

مثال :

الجدول التالي يوضح البيانات التي حصل باحث في دراسة بين النوع و مشاهدة البرامج التعليمية لمجموعة من طلاب الصف الثالث الثانوي على النحو التالي :

النوع	مشاهدة البرامج	النوع	مشاهدة البرامج
ذكر	يشاهد	ذكر	لا يشاهد
ذكر	يشاهد	أنثى	لا يشاهد
أنثى	يشاهد	أنثى	لا يشاهد
ذكر	لا يشاهد	أنثى	يشاهد
أنثى	يشاهد	ذكر	يشاهد
أنثى	لا يشاهد	ذكر	يشاهد
أنثى	لا يشاهد	ذكر	لا يشاهد
ذكر	لا يشاهد	ذكر	لا يشاهد
ذكر	يشاهد	أنثى	يشاهد
أنثى	لا يشاهد	أنثى	لا يشاهد

والمطلوب تكوين الجدول المزدوج للعلاقة بين المتغيرين (النوع ومشاهدة البرامج التعليمية) ؟
الحل :

- ١- المتغيرين (النوع – مشاهدة البرامج التعليمية)
- ٢- المتغير المستقل هو النوع والمتغير التابع هو مشاهدة البرامج التعليمية .
- ٣- فئات المتغير النوع هي (ذكور – إناث) فئات المتغير مشاهدة البرامج التعليمية (يشاهد – لا يشاهد)
- ٤- تكوين الجدول بحيث يحتل المتغير المستقل أعلى الجدول أى يكون أفقياً أما المتغير التابع فيحتل الجزء الأسفل أى يكون عمودياً .

كالتالي :

النوع	ذكور	إناث
مشاهدة البرامج التعليمية		
يشاهد		
لا يشاهد		

وضع العلامات .

النوع	ذكور	إناث
مشاهدة البرامج التعليمية		
يشاهد	////	////
لا يشاهد	////	///

إعادة كتابة الجدول بالأرقام .

النوع	ذكور	إناث	المجموع
مشاهدة البرامج التعليمية	5	4	9
يشاهد	5	6	11
لا يشاهد	10	10	20
المجموع			

الاسبوع 4

العرض البياني للبيانات الإحصائية

يعتبر العرض البياني للبيانات الإحصائية بمثابة تلخيص للبيانات الإحصائية في شكل يسهل منه استيعاب خصائص موضوع بحث الدراسة ، وتختلف طرق عرض البيانات المبوبة عن البيانات الغير مبوبة ، وسنتعرض لكل منها بالتفصيل فيما يلي :-

أولاً : العرض البياني للبيانات الغير مبوبة :

والمقصود بالبيانات الغير مبوبة تلك البيانات المفردة أي لا يوجد بها فئات وهناك عدة طرق لعرض البيانات الغير مبوبة .

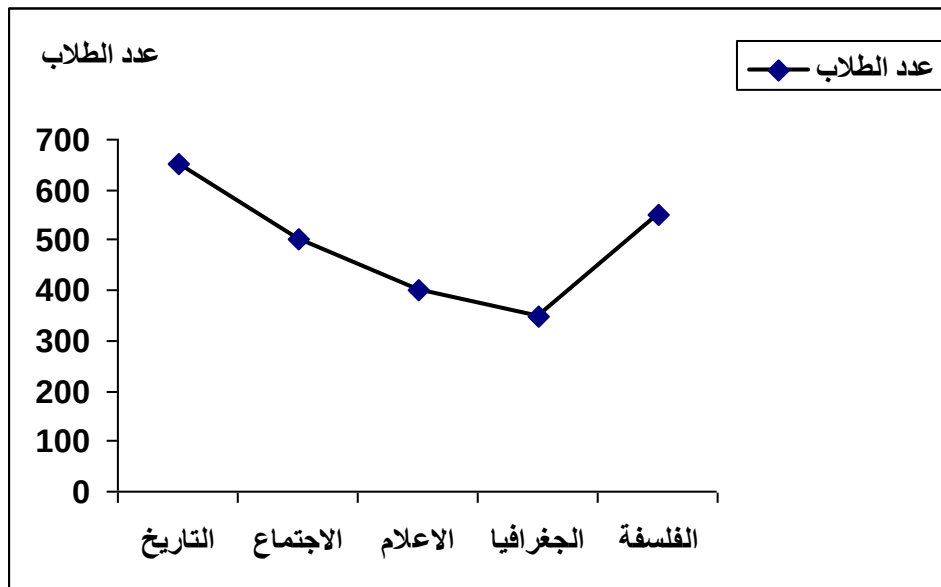
- طريقة الخط البياني المنكسر :

وفي هذه الطريقة يمثل محور السينات المتغير أما محور الصادات يمثل قيمة المتغير ويتم توقيع نقاط بين كل قيمة من قيم المتغير على محور السينات والقيمة المقابلة على محور الصادات ثم يتم توصيل تلك النقاط بخط منكسر باستخدام المسطرة .

مثال :

الجدول التالي يوضح أعداد الطلاب ببعض أقسام كلية الآداب جامعة بغداد والمطلوب عرض هذه البيانات باستخدام طريقة الخط البياني المنكسر؟

القسم	التاريخ	الاجتماع	الإعلام	الجغرافيا	الفلسفة
عدد الطلاب	650	500	400	350	550

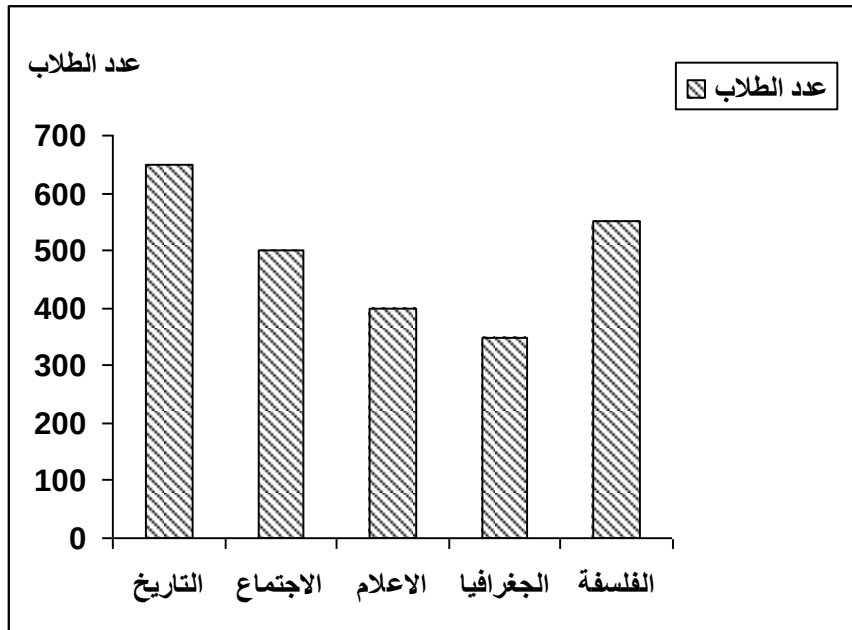
- طريقة الأعمدة البيانية البسيطة :

وفي هذه الطريقة يمثل محور السينات قيم المتغير أما محور الصادات يمثل القيمة المقابلة لقيمة المتغير ويتم رسم عمود حول المتغير وارتفاعه يمثل قيمة المتغير .

مثال :

الجدول التالي يوضح أعداد الطلاب ببعض أقسام كلية الآداب جامعة بغداد والمطلوب عرض هذه البيانات باستخدام طريقة الأعمدة البيانية البسيطة ؟

القسم	التاريخ	الاجتماع	الإعلام	الجغرافيا	الفلسفة
عدد الطلاب	650	500	400	350	550



- طريقة الدائرة البيانية :

وفى هذه الطريقة يتم رسم دائرة ثم نحسب زاوية قطاع كل قيمة على حدة ونقوم برسم تلك الزاوية داخل الدائرة حتى تنتهى الدائرة.
ونحسب زاوية قطاع الجزء من العلاقة :

$$\text{زاوية قطاع الجزء} = \frac{\text{التكرار الفعلي للجزء}}{\text{مجموع التكرارات}} \times 360$$

مثال :

الجدول التالى يوضح أعداد الطلاب ببعض أقسام كلية الآداب جامعة المنصورة والمطلوب عرض هذه البيانات باستخدام طريقة الدائرة البيانية ؟

القسم	التاريخ	الاجتماع	الإعلام	الجغرافيا	الفلسفة
عدد الطلاب	650	500	400	350	550

الحل :

نحسب مجموع التكرارات = 550 + 350 + 400 + 500 + 650

مجموع التكرارات = 2450

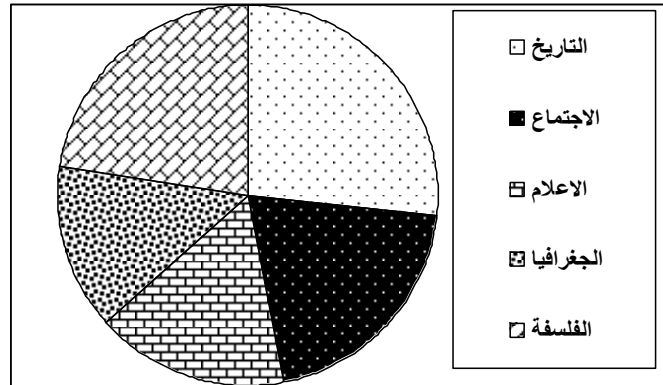
$$\text{زاوية قطاع التاريخ} = 360^\circ \times \frac{650}{2450} = 95.5^\circ$$

$$\text{زاوية قطاع الاجتماع} = 360^\circ \times \frac{500}{2450} = 73.5^\circ$$

$$58.7^\circ = 360^\circ * \frac{400}{2450} = \text{زاوية قطاع الاعلام}$$

$$51.4^\circ = 360^\circ * \frac{350}{2450} = \text{زاوية قطاع الجغرافية}$$

$$80.8^\circ = 360^\circ * \frac{550}{2450} = \text{زاوية قطاع الفلسفة}$$



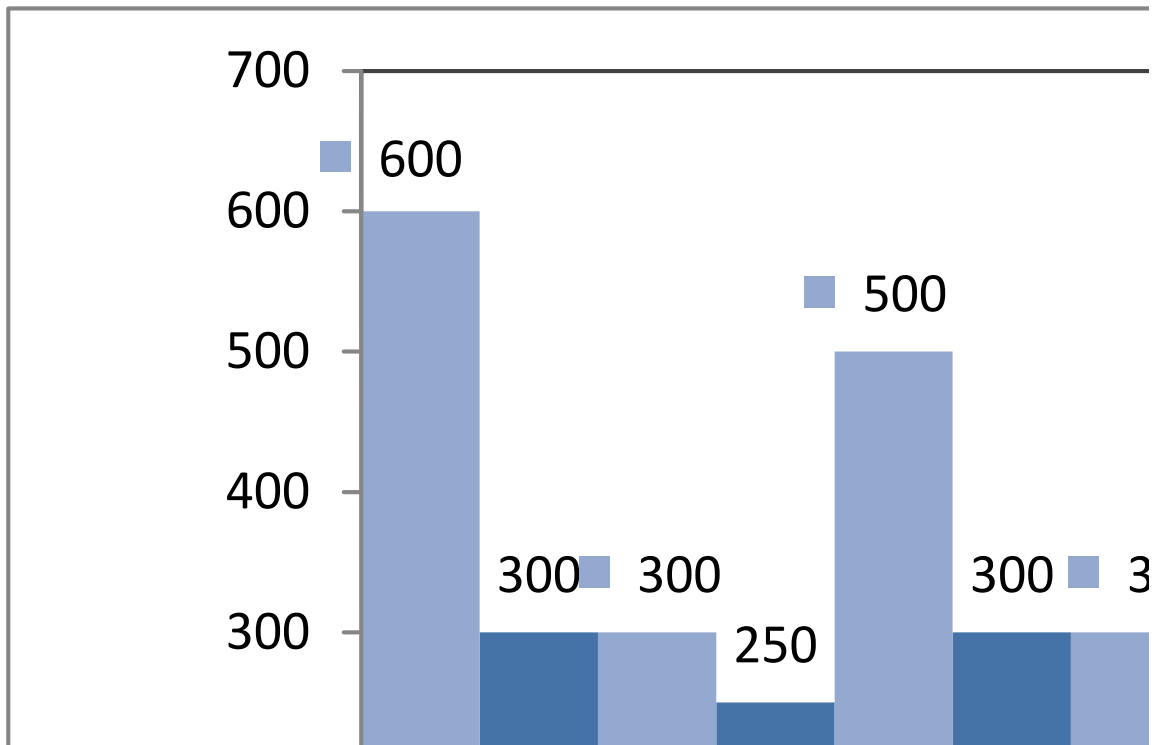
- طريقة الأعمدة البيانية المتلاصقة :

تسمى هذه الطريقة أيضا بطريقة الأعمدة البيانية المتجاورة وهي تشبه طريقة العمدة البيانية البسيطة ولكن يتم رسم عدد من الأعمدة متلاصقة يمثل كل منهم احد قيم المتغير .

مثال :

الجدول التالي يوضح أعداد الطلاب ببعض أقسام كلية الآداب جامعة بغداد والمطلوب عرض هذه البيانات باستخدام طريقة الأعمدة البيانية المتلاصقة ؟

القسم	التاريخ	الاجتماع	الإعلام	الجغرافيا	الفلسفة
طالب	300	250	300	250	300
طالبة	200	300	500	300	600

الحل :

الاسبوع 5

العرض البياني للبيانات المبوبة :

والمقصود بالبيانات المبوبة تلك البيانات المقسمة إلى فئات وهناك عدة طرق لعرض البيانات المبوبة .

(١) المدرج التكراري :

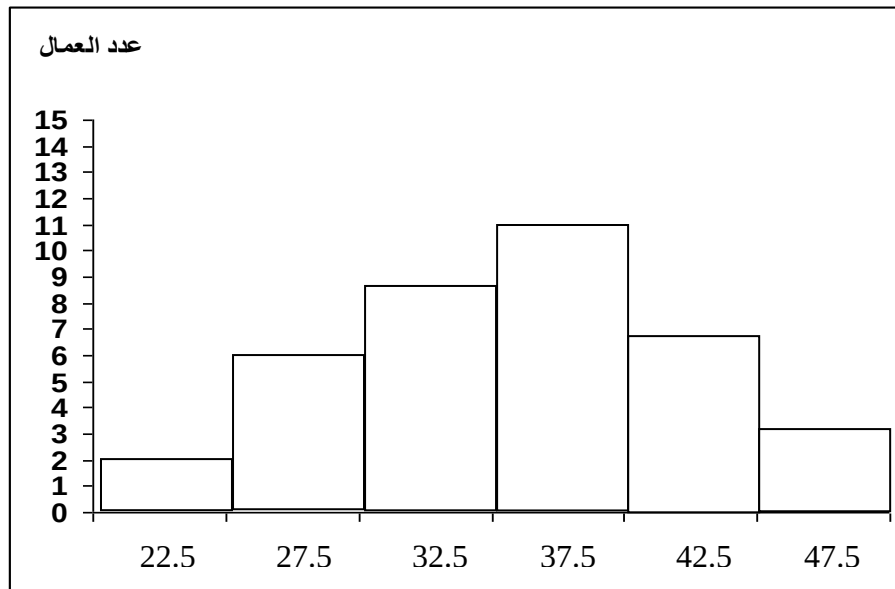
أحد طرق عرض البيانات المبوبة حيث يتم تخصيص عمود لكل فئة وتكرارها ، بحيث يكون طول الفئة هي قاعدة العمود والتكرار هو ارتفاع العمود ، ويفضل ترك فراغ كاف قبل الفئة الأولى وفراغ آخر بعد الفئة الأخيرة ، أما بالنسبة لمنتصف العمود فيكون هو مركز الفئة .

مثال : اعرض لهذا الجدول بيانياً باستخدام المدرج التكراري ؟

فئات العمر x	-20	-25	-30	-35	-40	-45
عدد العمال f	2	6	9	11	7	3

الحل :

فئات العمر x	عدد العمال f	مركز الفئة
-20	2	22.5
-25	6	27.5
-30	9	32.5
-35	11	37.5
-40	7	42.5
-45	3	47.5

(٢) المضلع التكراري :

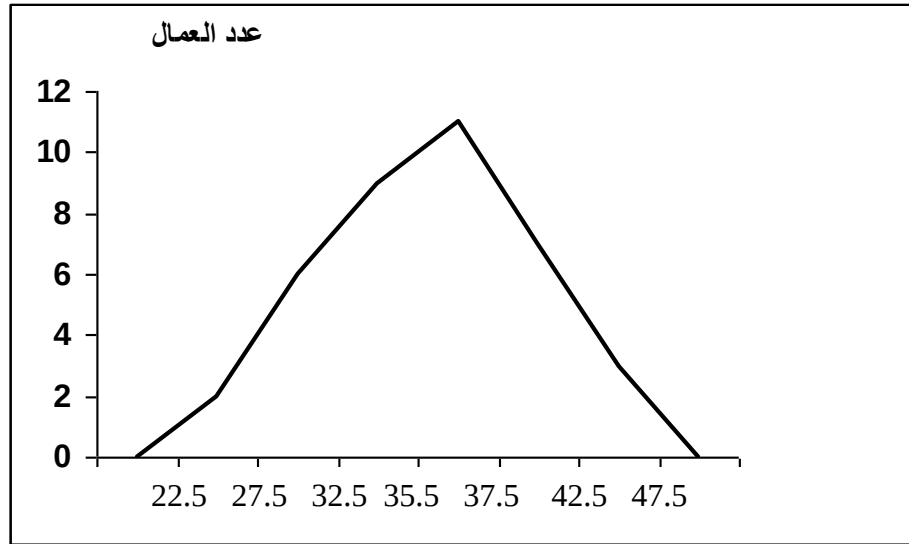
تخصص لكل فئة وتكرارها نقطة ، بحيث يكون الإحداثي السيني لها هو مركز الفئة بينما الإحداثي الصادي لها هو التكرار ، نفترض فئة سابقة للفئة الأولى وفئة لاحقة للفئة الأخيرة وتكرار كل منهما صفر ، ثم نوصل كل نقطتين متتاليتين بخط مستقيم بالمسطرة .

ملاحظة : مساحة الشكل تحت المدرج التكراري = مساحة الشكل تحت المضلع التكراري .

مثال : اعرض لهذا الجدول بيانياً باستخدام المضلع التكراري ؟

فئات العمر x	-45	-40	-35	-30	-25	-20
عدد العمال f	3	7	11	9	6	2

الحل :



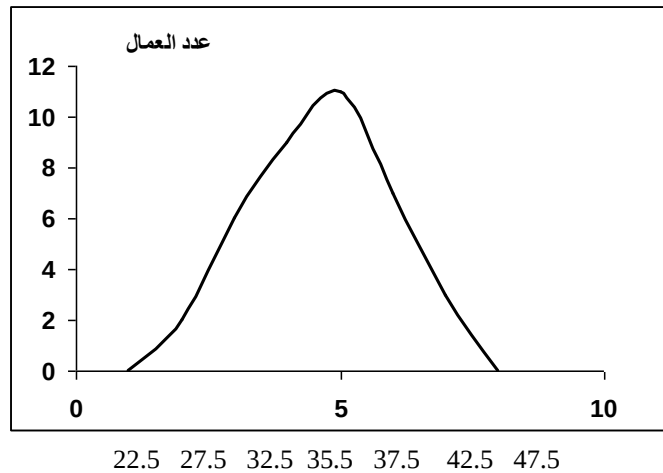
(٣) المنحنى التكراري :

بعد رصد النقاط كما في الطريقة السابقة نوصل كل نقطتين متتاليتين بمنحنى باليد .

مثال : اعرض لهذا الجدول بيانياً باستخدام المنحنى التكراري ؟

فئات العمر x	-45	-40	-35	-30	-25	-20
عدد العمال f	3	7	11	9	6	2

الحل :



التمثيل البياني للبيانات الكمية المتصلة (المستمرة):-

- المدرج التكراري: وهو يمثل العلاقة بين الفئات (المحور الأفقي) و التكرارات (المحور العمودي).
- المنحنى التكراري: وهو يمثل العلاقة بين مراكز الفئات (المحور الأفقي) و التكرارات (المحور العمودي).
- المضلع التكراري: وهو يمثل العلاقة بين مراكز الفئات (المحور الأفقي) و التكرارات (المحور العمودي).

ملاحظة:

- مركز الفئة $x_i = ((\text{الحد الأدنى للفئة} + \text{الحد الأعلى لها}) / 2)$ حيث يتم حسابه لكل فئة على حدى.
- في كل من المضلع والمنحنى التكراري يتم إضافة فئة قبل الفئة الأولى و أخرى بعد الفئة الأخيرة على الرسم فقط تكرار كل منهما يساوي صفر و تمثل بخط متقطع.

مثال: فيما يلي درجات 30 طالبا في اختبار مقرر الاحصاء:

37 , 15 , 7 , 17 , 34
41 , 23 , 26 , 46 , 24
57 , 44 , 37 , 27 , 53
22 , 28 , 44 , 28 , 20
32 , 34 , 60 , 52 , 39
32 , 30 , 5 , 31 , 40

المطلوب: تصنيف هذه البيانات في جدول تكراري مكون من 6 فئات و تمثيلها بيانيا.

مثال: البيانات التالية تمثل عدد أيام الغياب 30 طالب في الصف الأول المتوسط بإحدى المدارس في الأسبوع الأول من شهر رمضان:

1	2	3	5	2	2	1	0	0	0	0	1	3	2	0
3	3	3	2	2	3	2	3	1	0	2	1	1	1	1

الجدول التكراري:-

عدد أيام الغياب	العلامات	التكرار
0		6
1		8
2		8
3		7
4		0
5		1
المجموع		30

مثال: البيانات التالية تمثل درجات 50 طالب في مقرر الإحصاء (لأقرب درجة صحيحة).

32	30	34	29	28	25	26	26	27	25
30	32	31	30	34	31	31	30	33	34
35	39	36	37	38	39	35	33	31	32
38	42	37	44	36	40	35	38	42	37
40	33	48	44	49	40	45	41	39	43

لتكوين الجدول التكراري نتبع الخطوات التالية:

١. نحدد عدد الفئات المناسب و يكون عادة بين 5 - 15 فئة تبعاً لعدد البيانات.

في مثالنا نختار عدد الفئات = 5

٢. نحدد مدى البيانات = أكبر قيمة - أصغر قيمة.

$$\text{المدى} = 49 - 25 = 24$$

٣. نحدد طول الفئة $\Delta = \frac{\text{المدى}}{\text{الفئات عدد}}$

$$\Delta = \frac{24}{5} = 4.8 \approx 5$$

٤. نحدد الحد الأدنى لأول فئة و يكون $\geq$ أصغر قيمة للبيانات.

الحد الأدنى = 25

الحد الأدنى الفعلي = $25 - 0.5 = 24.5$

٥. نحدد الحد الأعلى الفعلي لأول فئة وذلك بإضافة طول الفئة للحد الأدنى الفعلي

الحد الأعلى الفعلي = $24.5 + 5 = 29.5$

الحد الأعلى = $29.5 - 0.5 = 29$

٦. نحدد الحدود الدنيا و العليا لكل فئة بإضافة طول الفئة لكل حد ، و نعين الحدود الفعلية بإضافة طول الفئة أيضاً لكل حد فعلي.

حدود الفئة	الحدود الفعلية للفئة
25 - 29	24.5 - 29.5
30 - 34	29.5 - 34.5
35 - 39	34.5 - 39.5
40 - 44	39.5 - 44.5
45 - 49	44.5 - 49.5

٧. نعين مراكز الفئات x_i باستخدام القانون:

$$x_i = \frac{\text{الحد الأدنى} + \text{الحد الأعلى}}{2}$$

أو يمكن إيجاد مركز الفئة الأولى فقط ثم إضافة طول الفئة لإيجاد مركز الفئة التالية.

$$x_1 = \frac{29 + 25}{2} = 27$$

$$x_2 = 27 + 5 = 32$$

٨. نكون الجدول التكراري.

حدود الفئة	الحدود الفعلية للفئة	مركز الفئة	العلامات	التكرار f_i
25 - 29	24.5 – 29.5	27		7
30 – 34	29.5 – 34.5	32		18
35 – 39	34.5 – 39.5	37		13
40 - 44	39.5 – 44.5	42		9
45 - 49	44.5 – 49.5	47		3
				$\sum f_i = 50$

سؤال: ما هو عدد الطلاب الذين درجاتهم أقل من أو تساوي 39؟
 $38 = 13 + 18 + 7$

وهو ما يعرف بالتكرار المتجمع للفئة:
 التكرار المتجمع للفئة: مجموع تكرارات جميع القيم التي تكون أقل من أو تساوي الحد الأعلى الفعلي للفئة.
 لإنشاء الجدول التوزيع التكراري المتجمع:-

حدود الفئة	الحدود الفعلية للفئة	التكرار	أقل من الحد الأعلى الفعلي للفئة	التكرار المتجمع
25 - 29	24.5 – 29.5	7	أقل من 29.5	7
30 – 34	29.5 – 34.5	18	أقل من 34.5	25
35 – 39	34.5 – 39.5	13	أقل من 39.5	38
40 - 44	39.5 – 44.5	9	أقل من 44.5	47
45 - 49	44.5 – 49.5	3	أقل من 49.5	50
		$\sum f = 50$		

بنفس الطريقة يمكن الحصول على التوزيع التكراري المتجمع النسبي و التوزيع التكراري المتجمع المئوي و لكن بالتجميع المتتالي للتكرار النسبي و المئوي للفئات على الترتيب.

الجدول التكراري المتجمع الصاعد:- وهو الجدول الذي يتم فيه حساب التكرارات بصورة تصاعدية و يتم انشاؤه عن طريق عمودين الاول به الحدود العليا للفئات والثاني باسم التكرار المتجمع الصاعد وهو يستخرج من العمودين الرئيسيين في الجدول الأصلي مع ملاحظة :-

- ان التكرار المتجمع الصاعد يبدأ بصفر وينتهي بالمجموع الكلي للتكرارات
- عدد فئاته اكبر بفئة من فئات الجدول الأصلي
- يمكن ان يشتمل هذا الجدول على أي نوع من البيانات سواء الوصفية او الكمية المتصلة أو المنفصلة

الجدول التكراري المتجمع الصاعد	
الحدود العليا للفئات	التكرار المتجمع الصاعد
LESS THAN 0	0
LESS THAN 10	5
LESS THAN 20	13
LESS THAN 30	16
LESS THAN OR EQUAL 40	20

الجدول الأصلي	
الفئات (الدرجة)	التكرار (عدد الطلاب)
10 - 0	5
20 - 10	8
30 - 20	3
40 - 30	4
المجموع	20

الجدول التكراري المتجمع الهابط:- وهو الجدول الذي يتم فيه حساب التكرارات بصورة تنازلية و يتم انشاؤه عن طريق عمودين الاول به الحدود العليا للفئات والثاني باسم التكرار المتجمع الهابط وهو يستخرج من العمودين الرئيسيين في الجدول الأصلي مع ملاحظة :-

- ان التكرار المتجمع الصاعد يبدأ بالمجموع الكلي للتكرارات وينتهي بالصفر
- عدد فئاته اكبر بفئة من فئات الجدول الأصلي
- يمكن ان يشتمل هذا الجدول على أي نوع من البيانات سواء الوصفية او الكمية المتصلة أو المنفصلة

الجدول التكراري المتجمع الهابط	
الحدود الدنيا للفئات	التكرار المتجمع الهابط
0 OR MORE	20
10 OR MORE	15
20 OR MORE	7
30 OR MORE	4
MORE THAN 40	0

الجدول الأصلي	
الفئات (الدرجة)	التكرار (عدد الطلاب)
10 - 0	5
20 - 10	8
30 - 20	3
40 - 30	4
المجموع	20

مثال : فيما يلي بيانات عن المستوى التعليمي لعينة من 50 فرد.

متوسط	يقرأ ويكتب	ثانوي	متوسط	ثانوي	أعلى من جامعي	متوسط	ابتدائي
يقرأ ويكتب	متوسط	ثانوي	ثانوي	متوسط	ثانوي	ابتدائي	متوسط
ابتدائي	ثانوي	يقرأ ويكتب	جامعي	ثانوي	ابتدائي	يقرأ ويكتب	ثانوي
متوسط	ابتدائي	متوسط	ثانوي	ابتدائي	متوسط	جامعي	متوسط
ثانوي	متوسط	ابتدائي	ثانوي	يقرأ ويكتب	ابتدائي	ثانوي	ابتدائي
جامعي	ثانوي	جامعي	ابتدائي	جامعي	أعلى من جامعي	ثانوي	ثانوي
متوسط	يقرأ ويكتب						

والمطلوب: ١- اعرض البيانات في شكل جدول تكراري.

الحل

١- عرض البيانات في شكل جدول تكراري:
المستوى التعليمي (يقرأ ويكتب- ابتدائي- متوسط- ثانوي- جامعي- أعلى من جامعي) متغير وصفي ترتيبى، ويمكن عرض البيانات أعلاه في شكل جدول تكراري بإتباع الآتي:

• تكوين جدول تفريغ البيانات: الإشكال البيانية أو العرض الهندسي للبيانات

من أجل إعطاء فكرة واضحة و دقيقة و سريعة عن البيانات المبوبة فأنها تعرض بهيئة رسوم بيانية و أشكال هندسية متعددة و بشكل عام سوف نخصص المحور الأفقي (X- axis) لفئات المتغير بينما نخصص المحور العمودي (Y- axis) للتكرارات.

و دائما نبدأ تقسيم المحورين من الصفر.

الحل

١- عرض البيانات في شكل جدول تكراري:
المستوى التعليمي (يقرأ ويكتب- ابتدائي- متوسط- ثانوي- جامعي- أعلى من جامعي) متغير وصفي ترتيبى، ويمكن عرض البيانات أعلاه في شكل جدول تكراري بإتباع الآتي:

• تكوين جدول تفريغ البيانات:

جدول تفريغ البيانات

عدد الأفراد (التكرارات)	العلامات الإحصائية	المستوى التعليمي
6		يقرأ ويكتب
10		ابتدائي
12		متوسط
15		ثانوي
5		جامعي
2		أعلى من جامعي
50		Sum

- تكوين الجدول التكراري:
التوزيع التكراري لعينة حجمها 50 فرد حسب المستوى التعليمي

التوزيع التكراري النسبي	عدد الأفراد (التكرارات) (f)	المستوى التعليمي
0.12	6	يقرأ ويكتب
0.20	10	ابتدائي
0.24	12	متوسط
0.30	15	ثانوي
0.10	5	جامعي
0.04	2	أعلى من جامعي
1.00	50	Sum

مثال : فيما يلي بيانات درجات 70 طالب في الاختبار النهائي لمقرر مادة الإحصاء التطبيقي.

56	65	70	65	55	60	66	70	75	56
60	70	61	67	61	71	67	62	71	66
68	72	57	68	72	69	57	71	69	75
72	62	67	73	58	63	66	73	63	65
58	73	74	76	74	80	81	60	74	58
76	82	77	83	77	85	91	78	94	72
79	64	57	79	55	87	64	88	78	62

المطلوب:

- 1- كون التوزيع التكراري لدرجات الطلاب.
- 2- ما هو نسبة الطلاب الحاصلين على درجة ما بين 70 إلى أقل من 80؟
- 3- ما هو نسبة الطلاب الحاصلين على درجة أقل من 70 درجة؟
- 4- ما هو نسبة الطلاب الحاصلين على درجة 80 أو أكثر؟

الحل

- 1- تكوين التوزيع التكراري:
درجة الطالب في الاختبار متغير كمي مستمر، ولكي يتم تبويب البيانات في شكل جدول تكراري، يتم اتباع الآتي:

- (حساب المدى Range R)

$$\text{Range} = \text{Maximum} - \text{Minimum}$$

$$R = 94 - 55 = 39$$

معادلة 1

- (تحديد عدد الفئات C:Classes)

تحدد عدد الفئات وفقاً لاعتبارات منها: رأي الباحث، والهدف من البحث، وحجم البيانات، ويرى كثيراً من الباحثين أن أفضل عدد للفئات يجب أن يتراوح بين 5 إلى 15، بفرض أن عدد الفئات هو 8 فئات، أي أن: (C=8).

- (حساب طول الفئة L Length):

$$L = \frac{\text{Range}}{\text{Classes}} = \frac{R}{C} = \frac{39}{8} = 4.875 \approx 5$$

معادلة 2

- تحديد الفئات:

- الفئة تبدأ بقيمة تسمى الحد الأدنى، وتنتهي بقيمة تسمى الحد الأعلى، ومن ثم نجد أن :
- الحد الأدنى للفئة الأولى هو أقل قراءة (درجة) أي أن الحد الأدنى للفئة الأولى = 55
الحد الأعلى للفئة الأولى = الحد الأدنى + طول الفئة = $55 + L = 55 + 5 = 60$
إذا الفئة الأولى هي: " 55 to less than 60 " وتقرأ " من 55 إلى أقل من 60 "
 - الحد الأدنى للفئة الثانية = الحد الأعلى للفئة الأولى = 60
الحد الأعلى للفئة الثانية = الحد الأدنى للفئة + طول الفئة = $60 + 5 = 65$
إذا الفئة الثانية هي: " 60 to less than 65 " وتقرأ " من 60 إلى أقل من 65 "
 - وبنفس الطريقة يتم تكوين حدود الفئات الأخرى، وهي:
- الفئة الثالثة : 65 to less than 70 الفئة الرابعة : 70 to less than 75
الفئة الرابعة : 75 to less than 80 الفئة السادسة : 80 to less than 85
الفئة الخامسة : 85 to less than 90 الفئة الثامنة : 90 to less than 95

ويمكن كتابة الفئات بأشكال مختلفة كما هو مبين بجدول تفريغ البيانات:

- تكوين جدول تفريغ البيانات:

جدول تفريغ البيانات

الدرجة			العلامات الإحصائية	عدد الطلاب (التكرارات)
فئات	فئات	فئات		
55 to less than 60	55 – 60	55-	<del> </del> <del> </del>	10
60 to less than 65	60 – 65	60-	<del> </del> <del> </del>	12
65 to less than 70	65 – 70	65-	<del> </del> <del> </del> /	13
70 to less than 75	70 – 75	70-	<del> </del> <del> </del> <del> </del> /	16
75 to less than 80	75 – 80	75-	<del> </del> <del> </del>	10
80 to less than 85	80 – 85	80-		4
85 to less than 90	85 – 90	85-		3
90 to less than 95	90 - 95	90-95		2
Sum				70

- تكوين الجدول التكراري:
التوزيع التكراري لعدد 70 طالب حسب درجاتهم في اختبار مقرر الإحصاء

فئات الدرجة	عدد الطلاب (التكرارات) (f)
55 – 60	10
60 – 65	12
65 – 70	13
70 – 75	16
75 – 80	10
80 – 85	4
85 – 90	3
90 – 95	2
Sum	70

مثال: فيما يلي التوزيع التكراري لأوزان عينة من الدواجن بالجرام، حجمها 100 اختيرت من أحد المزارع بعد 45 يوم.

الوزن	600-	620-	640-	660-	680-	700-720	Sum
عدد الدجاج	10	15	20	25	20	10	100

والمطلوب:

- ١- ما هو طول الفئة؟
- ٢- ارسم المدرج التكراري.

الحل

١- طول الفئة (L)

طول الفئة = الحد الأعلى للفئة – الحد الأدنى للفئة

$$L = upper - Lower$$

معادلة 4

$$L = 620 - 600 = 640 - 620 = \dots = 720 - 700 = 20$$

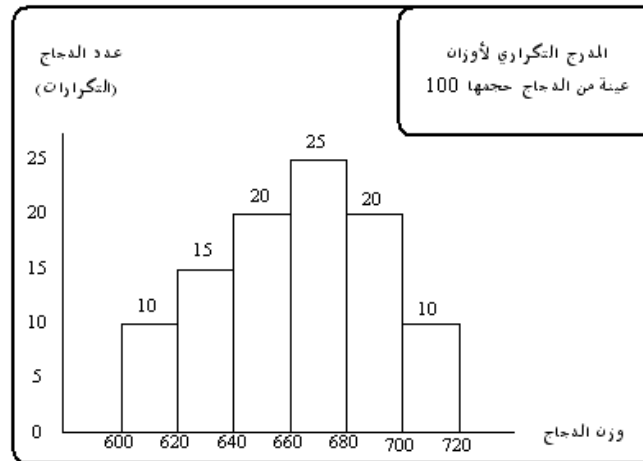
إذا طول الفئة = 20

١- رسم المدرج التكراري.

لرسم المدرج التكراري يتم إتباع الخطوات التالية:

- رسم محوران متعامدان، الرأسى ويمثل التكرارات، الأفقى ويمثل الأوزان.
 - كل فئة تمثل بعمود ارتفاعه هو تكرار الفئة، وطول قاعدته هو طول الفئة.
 - كل عمود يبدأ من حيث انتهى به عمود الفئة السابقة.
- والشكل ادناه يبين المدرج التكراري لأوزان الدجاج.

المدرج التكراري لأوزان عينة من الدجاج حجمها 100 دجاجة



٢. المضلع التكراري

هو تمثيل بياني أيضا للجدول التكراري البسيط، حيث تمثل التكرارات على المحور الرأسي، ومراكز الفئات على المحور الأفقي، ثم التوصيل بين الإحداثيات بخطوط منكسرة، وبعد ذلك يتم توصيل طرفي المضلع بالمحور الأفقي. ومركز الفئة هي القيمة التي تقع في منتصف الفئة، وتحسب بتطبيق المعادلة التالية:

$$\text{مركز الفئة} = \frac{\text{الحد الأدنى للفئة} + \text{الحد الأعلى للفئة}}{2}$$

$$\text{Midpoint} = \frac{\text{Lower} + \text{Upper}}{2}$$

معادلة 5

ونظرا لعدم معرفة القيم الفعلية لتكرار كل فئة، يعتبر مركز الفئة هو التقدير المناسب لقيمة كل مفردة من مفردات الفئة.

مثال: استخدم بيانات الجدول التكراري في المثال السابق لرسم المضلع التكراري.

الحل

لرسم المضلع التكراري يتبع الآتي:

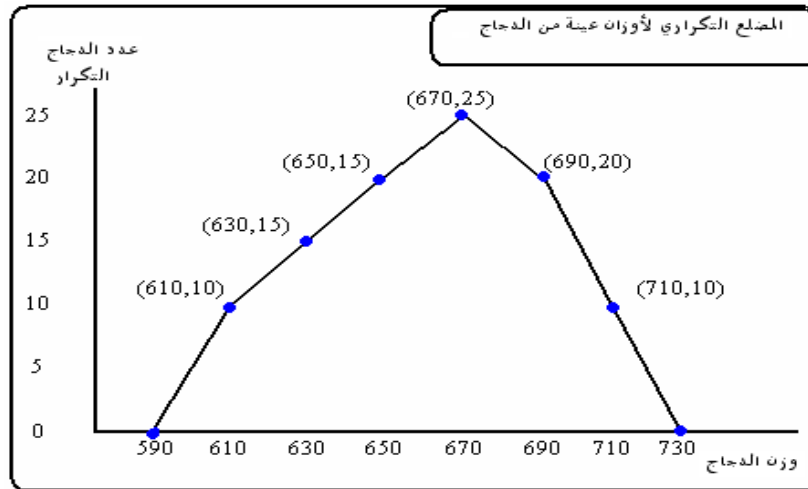
- حساب مراكز الفئات بتطبيق المعادلة رقم ٥

الوزن	عدد الدجاج (التكرار) (f)	مركز الفئة (x)
600-	10	$(600+620)/2= 610$
620-	15	$(620+640)/2=630$
640-	20	650
660-	25	670
680-	20	690
700-720	10	$(700+720)/2=710$
Sum	100	

- نقط الإحداثيات هي :

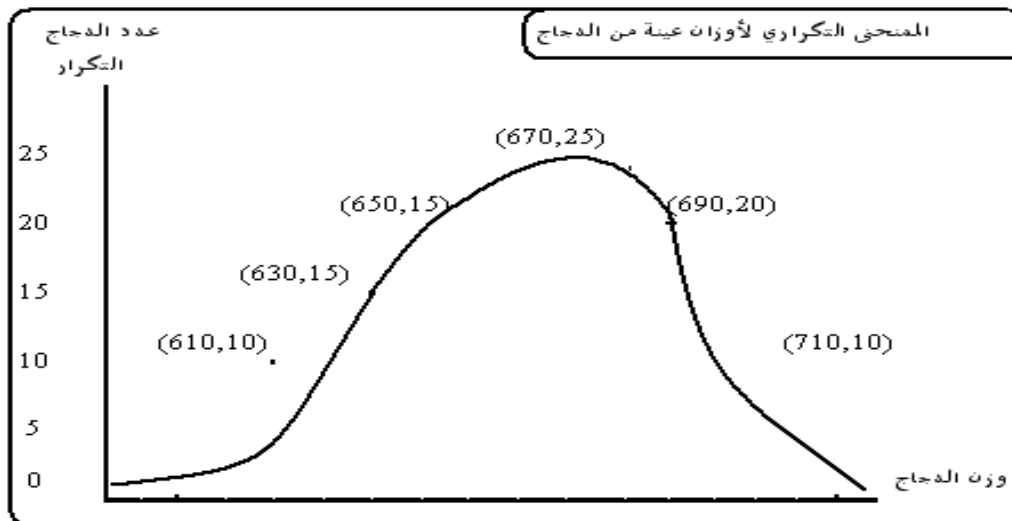
مركز الفئة (x)	590	610	630	650	670	690	710	730
التكرار (f)	0	10	15	20	25	20	10	0

المضلع التكراري لأوزان عينة من الدجاج حجمها 100 دجاجة



المنحنى التكراري : بإتباع نفس الخطوات السابقة في رسم المضلع يمكن رسم المنحنى التكراري، ولكن يتم تمهيد الخطوط المنكسرة في شكل منحنى بحيث يمر بأكثر عدد من النقاط، وفي المثال السابق يمكن رسم المنحنى التكراري، والشكل ادناه يبين هذا الشكل.

المنحنى التكراري لأوزان عينة من الدجاج حجمها 100 دجاجة



التوزيعات التكرارية المتجمعة: في كثير من الأحيان قد يحتاج الباحث إلى معرفة عدد المشاهدات التي تقل عن قيمة معينة أو تزيد عن قيمة معينة، ومن ثم يلجأ الباحث إلى تكوين جداول تجميعية صاعدة أو هابطة، وفيما يلي بيان كيفية تكوين كل نوع من هذين النوعين على حدة:

التوزيع التكراري المتجمع الصاعد:

لتكوين الجدول التكراري المتجمع الصاعد، يتم حساب مجموع التكرارات (عدد القيم) التي تقل عن كل حد من حدود الفئات.

مثال : الجدول التكراري التالي يبين توزيع 40 بقرة في مزرعة حسب كمية الألبان التي تنتجها البقرة في اليوم باللتر.

كمية الألبان	18-	22-	26-	30-	34-38	Sum
عدد الأبقار	4	9	15	8	4	40

والمطلوب:

١- كون جدول التوزيع التكراري المتجمع الصاعد.

الحل

١- التوزيع التكراري المتجمع الصاعد.

توزيع تكراري متجمع صاعد

التوزيع التكراري

كمية الإنتاج باللتر	عدد الأبقار
18-	4
22-	9
26-	15
30-	8
34-38	4
Sum	40

أقل من	تكرار متجمع صاعد
أقل من 18	0
أقل من 22	4
أقل من 26	13
أقل من 30	28
أقل من 34	36
أقل من 38	40

التوزيع التكراري المتجمع الهابط (النازل):

لتكوين الجدول التكراري المتجمع النازل، يتم حساب مجموع التكرارات (عدد القيم) التي تساوي أو تزيد عن كل حد من حدود الفئات.

مثال : الجدول التكراري التالي يبين توزيع 40 بقرة في مزرعة حسب كمية الألبان التي تنتجها البقرة في اليوم باللتر.

كمية الألبان	18-	22-	26-	30-	34-38	Sum
عدد الأبقار	4	9	15	8	4	40

١- كون التوزيع التكراري المتجمع النازل.

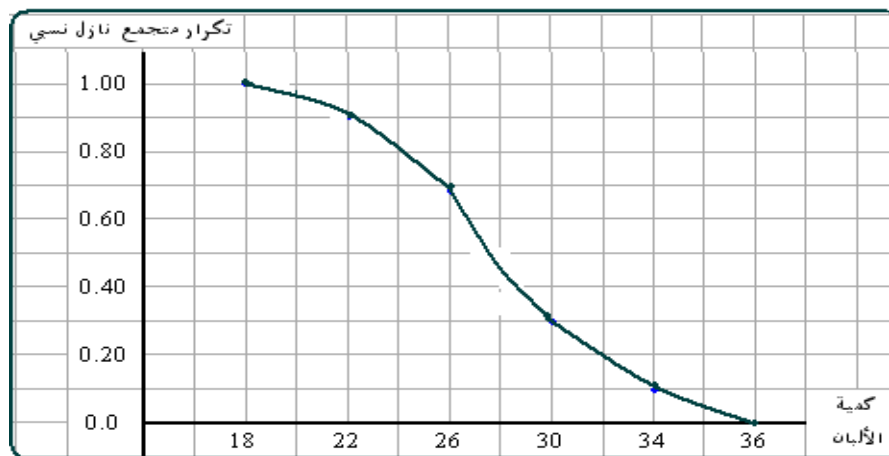
الحل:

١- تكوين التوزيع التكراري المتجمع النازل.

التوزيع التكراري	
كمية الإنتاج باللتر	عدد الأبقار
18-	4
22-	9
26-	15
30-	8
34-38	4
Sum	40

توزيع تكراري متجمع نازل	
أكثر من أو يساوي	تكرار متجمع نازل
أكثر من أو يساوي 18	40
أكثر من أو يساوي 22	36
أكثر من أو يساوي 26	27
أكثر من أو يساوي 30	12
أكثر من أو يساوي 34	4
أكثر من أو يساوي 38	0

رسم المنحنى التكراري المتجمع النازل.

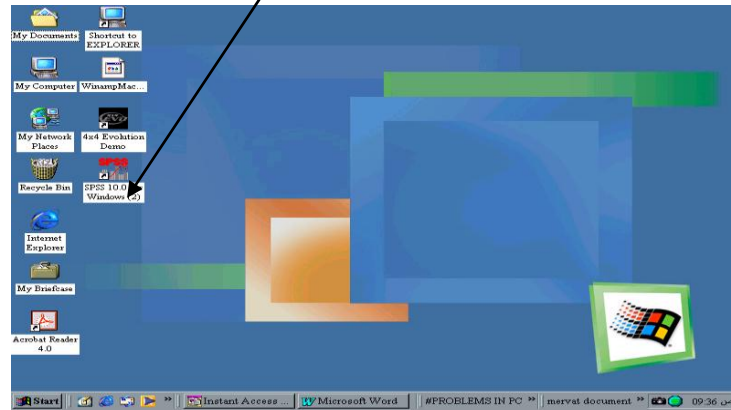


الاسبوع 6 مقدمة تشغيل برنامج SPSS

❖ تشغيل برنامج SPSSWIN

يعتبر برنامج SPSS أداة هامة ومتقدمة لإجراء التحليل الإحصائي اللازم لتحليل بيانات الأبحاث العلمية في المجالات المختلفة، وبرنامج (SPSSWIN) هو اختصار لمجموعة التحليلات الإحصائية للعلوم الاجتماعية من خلال برنامج النافذة (Statistical Package for Social Science under Windows)

ويبدأ تشغيل البرنامج من أيقونة برنامج (SPSSWIN) التي تظهر ضمن برنامج النوافذ عند بدء تشغيله (انظر إلي أيقونة SPSS في نافذة تشغيل Windows Millennium)

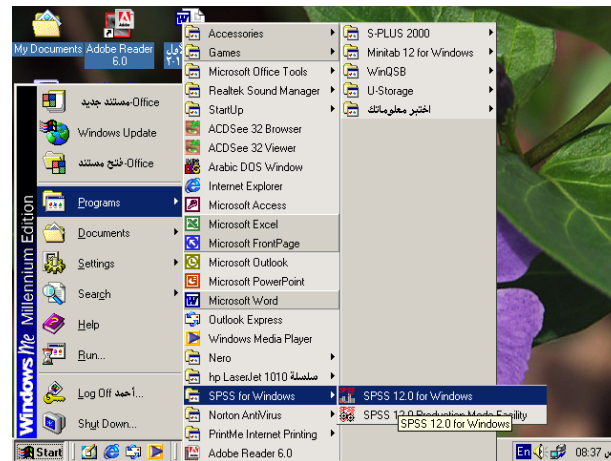


ويمكن الدخول إلى برنامج SPSS من قائمة Start

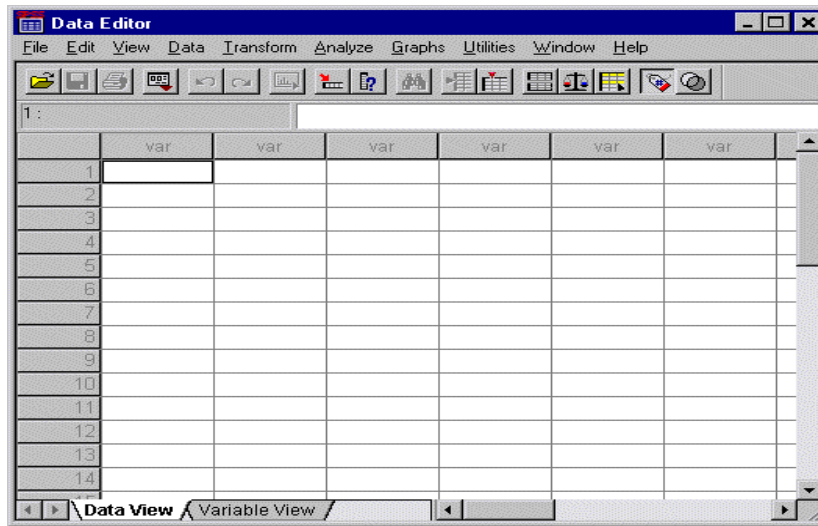
للدخول إلى برنامج SPSS نتبع الخطوات التالية:

- من قائمة Start اختر

Programs → SPSS for Widows → SPSS12 for Windows



فيتم فتح البرنامج وتظهر الشاشة الافتتاحية للبرنامج وتحتوي على شاشتين مهمتين الأولى **Variable view** يتم فيها كتابة المتغيرات (الرمز . النوع (كمى أو وصفى) . عدد الارقام العشرية....)



وفي شاشة Data view يتم إدخال البيانات

عناصر شريط الأوامر

١- File (ملف)

يتناول أساسا تنفيذ عمليات فتح الملفات وحفظها وإغلاقها وطبعها والخروج من البرنامج. والقائمة المنسدلة منه تحتوي علي الأوامر التالية

New فتح ملف جديد للبيانات.

Open فتح ملف موجود مسبقا وعند اختياره يتم تحديد اسمه في صندوق الحوار.

Close أقفال النافذة الأمامية.

Save حفظ الملف المفتوح على الشاشة بنفس اسمه.

Save As حفظ الملف المفتوح على الشاشة باسم آخر.

Display Data Information عرض معلومات عن البيانات.

Apply Data Dictionary تطبيق قاموس البيانات.

Print طبع الملف المعروض على النافذة الأمامية.

Stop SPSS Processor إيقاف معالج الحقيبة الإحصائية.

Exit خروج من الملف أو البرنامج.

٢- Edit (تحرير)

يتناول أساسا تنفيذ عمليات تحرير أو تعديل محتويات الملفات من خلال النسخ والصق والبحث. وتحتوي القائمة المنسدلة منه علي الأوامر التالية

Undo التراجع.

Cut قطع أو إزالة جزء من الملف.

Copy نسخ جزء من الملف.

Paste لصق أو إدماج جزء من الملف.

Clear إزالة جزء من الملف بعد تظليله.

Find البحث عن ملف ما.

Options الاختيارات.

٣ - Data (بيانات)

يتناول أساسا تنفيذ عمليات تعريف أسماء المتغيرات ونوعيتها وأوزان ترجيحها وتحديد تواريخ السلاسل الزمنية وفرز الحالات وإضافة حالات جديدة واختيار حالات معينة بشروط معينة ودمج وانقسام ملفات البيانات. وتحتوي القائمة المنسدلة منه علي الأوامر التالية

Define Variable يستخدم في تسمية المتغير.

Define Dates تعريف التواريخ الزمنية خصوصا للسلاسل الزمنية.

Templates تكرار النماذج المحددة مسبقا لتعريف خصائص متغيرات متشابهة.

Insert Variable إدخال عمود متغير جديد في ملف البيانات.

Insert Case إدخال صف جديد في ملف البيانات.

Go to Case الانتقال إلى حالة معينة في ملف البيانات.

Sort Cases ترتيب الحالات (الصفوف) في ملف البيانات.

Transpose تحويل الصفوف لأعمدة أو الأعمدة إلى صفوف في ملف البيانات.

Merge Files دمج الملفات.

Aggregate دمج الحالات.

Split File انقسام الملف.

Select Cases اختيار حالات معينة للتشغيل إذا ما توافر شرط ما.

Weight Cases تخصيص أوزان للحالات.

٤ - Transform (التحويل)

يتناول أساسا تنفيذ عمليات تحويل المتغيرات وإيجاد قيم المتغيرات الجديدة من المتغيرات الحالية وترتيب الحالات وإنشاء السلاسل الزمنية والتعامل مع البيانات المفقودة.

Compute لإيجاد قيمة متغير جديد بدلالة علاقة رياضية تربط بين أكثر من متغير قائم.

Random Number Seed نقطة بداية توليد أرقام العشوائية.

Count إجراء العد لبيانات محددة.

Recode إعادة إيجاد قيم متغير قائم بشكل جديد.

Rank Cases إعادة ترتيب حالات بشكل محدد.

Automatic Recode إعادة ترتيب الحالات بشكل تلقائي.

Create Time Series إنشاء السلاسل الزمنية.

Replace Missing Values إحلال قيم مكان بيانات غير متوافرة.

Run Pending Transform تشغيل مع إيقاف تحويل البيانات.

٥- Analyze (التحليل الإحصائي)

تعتبر هذه القائمة من أهم القوائم حيث تشتمل على مختلف الأساليب الإحصائية التي يقوم البرنامج بتنفيذها. ويعرض الشكل التالي قائمة **Analyze**، وتتمثل هذه القائمة فيما يلي :-

التقارير Reports

الإحصائيات الوصفية Descriptive Statistics .

مقارنة الأوساط الحسابية Compare Means وتشمل.

الأوساط الحسابية Means .

اختبار (t) لعينة واحدة One- Sample (T) Test.

اختبار (t) لعينات المستقلة Independent Sample (T) Test.

اختبار (t) لأزواج من العينات Paired Samples (T) Test.

تحليل التباين في اتجاه واحد One Way ANOVA.

النموذج الخطي العام General linear model

الارتباط Correlate.

التثائي (بين متغيرين) Bivariate .

ارتباط جزئي Partial.

وأساليب إحصائية أخرى

٦- Graphs (الأشكال البيانية)

يتناول أساسا تنفيذ عمليات الرسوم البيانية للبيانات بمختلف أنواعها مثل الأعمدة Bar، الخط Line، الدائرة

Pie، الرسم الصندوقي Box Plot

وشكل الانتشار Scatter والمدرج التكراري Histogram.

٧- Utilities (مساعدات)

يتناول أساسا مساعدات مكملة وتعديلات بشأن تعديل المخرجات وإظهار شبكات نافذة البيانات وغيرها. مثل

المتغيرات Variables، معلومات ملف File Info، وتعريف المجموعات Define Sets

٨- Window (إطار)

يتناول أساسا تنفيذ عمليات أساليب عرض نوافذ البرنامج. مثل تصغير النافذة Minimize All Windows.

٩ - Help (مساعدة)

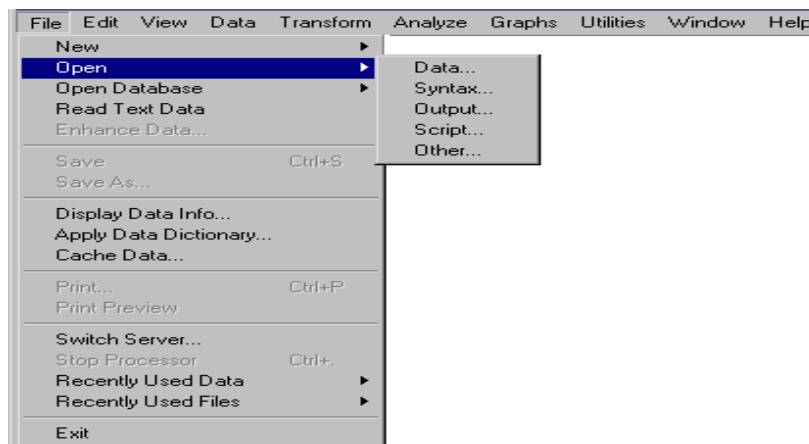
ويتضمن إرشادات المعاونة (المساعدات) التي قد يطلبها مستخدم البرنامج عند تشغيل البرنامج.

فتح ملف البيانات:

لفتح ملف بيانات

- اختر من القوائم

File → Open → Data



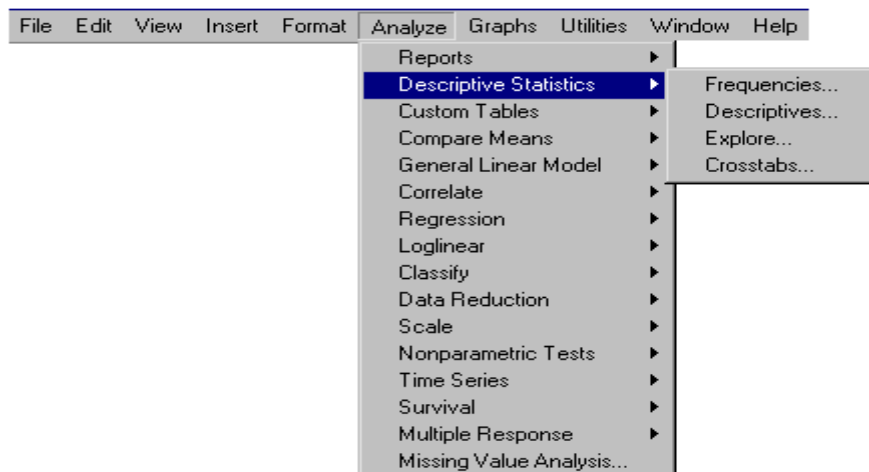
وهنا تظهر ملفات البيانات الخاصة بالبرنامج SPSS وامتدادها ".sav"، ولعرض ملفات ذات امتداد آخر يمكنك تغيير الامتداد الذي سيقوم البرنامج بعرض الملفات التي تحمله باستخدام Files of Type وذلك بالنقر علي مفتاح  فتظهر أنواع الامتدادات التي يمكن لبرنامج SPSS فتحها

- اختر ملف بيانات وأضغط "Open"

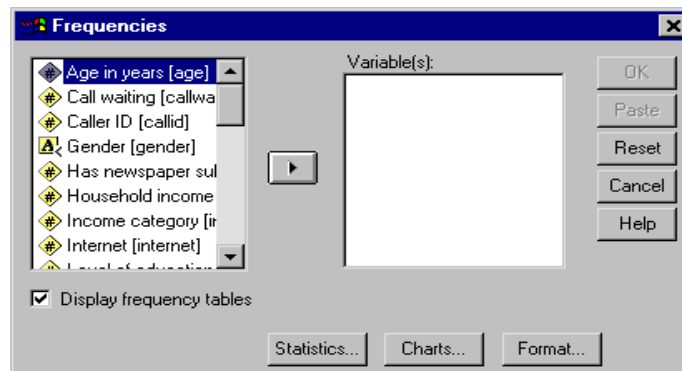
❖ إنشاء جدول تكراري بسيط:

اختر من القوائم

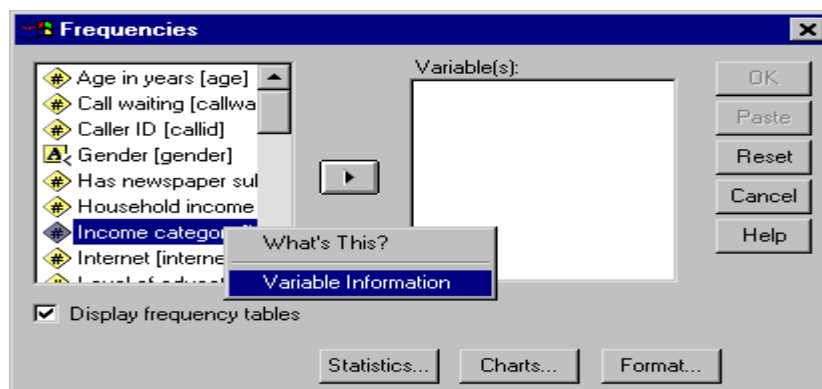
Analyze → Descriptive Statistics → Frequencies



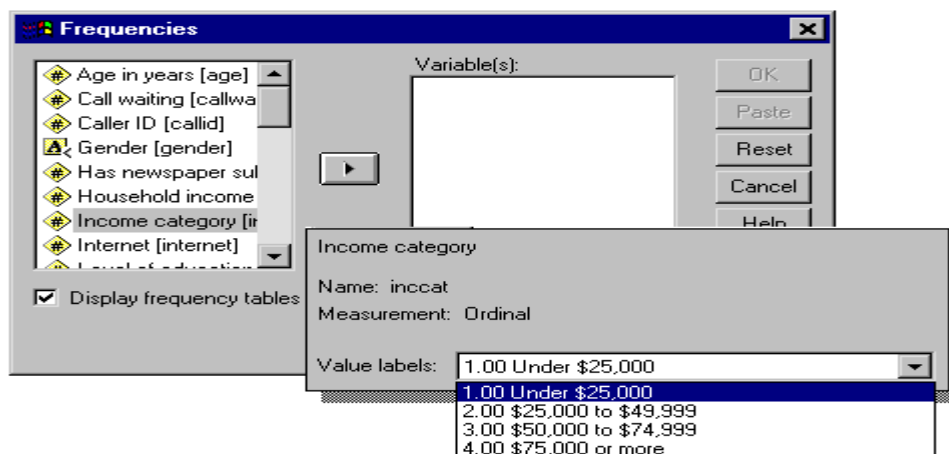
يفتح صندوق حوار Frequencies تظهر معلومات كاملة أكثر عن كل متغير، عندما يقف المؤشر عليه تظهر أسماء المتغيرات في الأقواس المربعة مثل متغير inccat وعنوان المتغير Income category وفي حالة عدم وجود عنوان، يظهر أسم المتغير فقط في قائمة الصندوق.



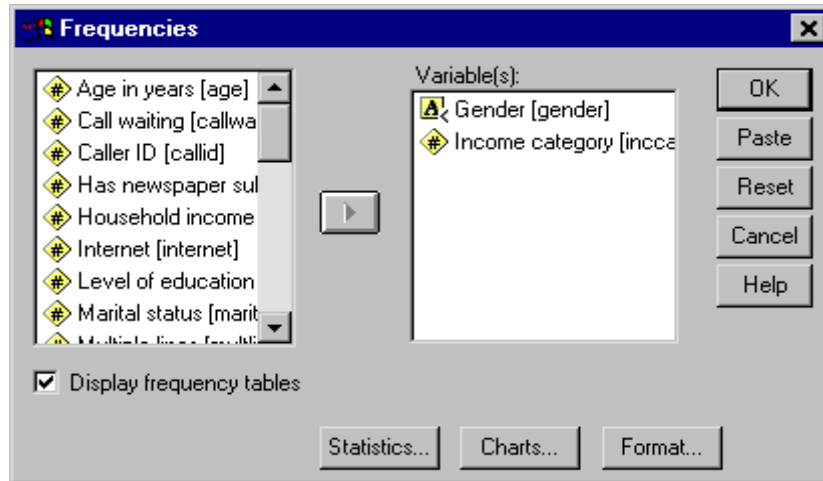
في صندوق الحوار اختر المتغيرات التي ترغب في تحليلها من قائمة المصدر علي اليسار ، وحرك المتغيرات لقائمة المتغيرات Variable(s) علي اليمين ، يختفي مفتاح Ok وذلك حتى يظهر علي الأقل متغير واحد في قائمة المتغيرات. ويمكن الحصول على معلومات عن المتغير بالضغط على الزر الأيمن للفارة (الماوس) عند income category [inccat] والضغط بالزر الأيسر للماوس على Variable Information



تظهر المعلومات التالية عن المتغير



- انقر على السهم المتجه إلى أسفل ▾ في قائمة Value Labels تظهر عناوين كل القيم.
- انقر على متغير النوع Gender [gender] في قائمة المصدر ، وأضغط على السهم المتجه إلى اليمين ▶ لتحريك المتغير إلى قائمة المتغيرات
 - أضغط على الدخل income category [inccat] في قائمة المصدر واضغط على زر السهم الأيمن مرة ثانية لإدخال المتغير في قائمة المتغيرات.



- تشير علامة العدد (#) أن المتغير عددي numeric
- يشير الحرف (A) أن المتغير حرف أبجدي string وقد يحتوى المتغير على كل من الحروف والأعداد
- أضغط على Ok لتنفيذ أمر عمل جدول تكراري وتظهر النتائج في نافذة العرض Viewer window

يمكن الوصول إلى التوزيع التكراري لأي متغير بسرعة في نافذة Viewer بالضغط على هذا المتغير في القائمة التي على اليسار ، على سبيل المثال يمكن الوصول إلى التوزيع التكراري للدخل بالنقر على income category ، وهذا الأمر يأخذك مباشرة إلى جدول التوزيع التكراري لفئات الدخل ، فيعرض الجدول العدد والنسبة المئوية في كل فئة من فئات الدخل.

		Frequency	Percent	Valid Percent	Percent
Valid	Female	493	49.3	49.3	49.3
	Male	507	50.7	50.7	100.0
	Total	1000	100.0	100.0	

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Under \$25	264	26.4	26.4	26.4
	\$25 - \$49	382	38.2	38.2	64.6
	\$50 - \$74	148	14.8	14.8	79.4
	\$75+	206	20.6	20.6	100.0
	Total	1000	100.0	100.0	

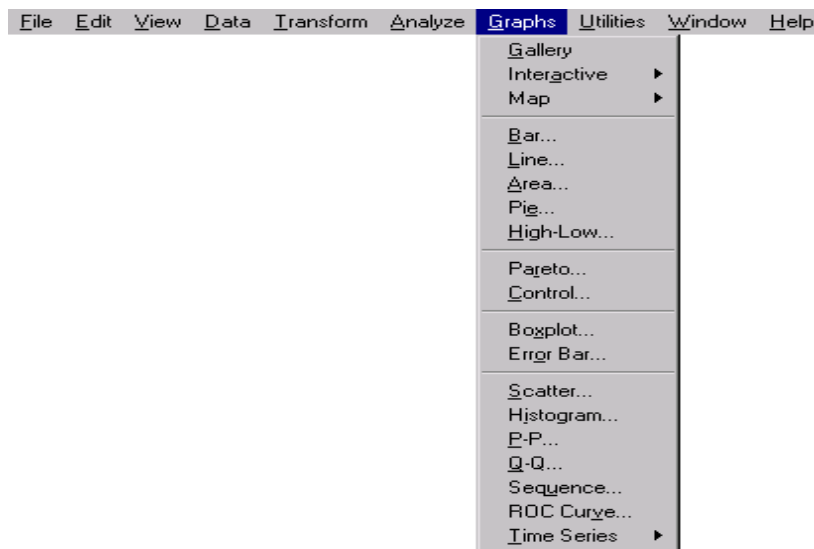
❖ قائمة الرسومات:

مع أن معظم الأساليب الإحصائية يمكن من خلال نوافذها إيجاد الأشكال البيانية. يمكنك أيضاً رسم الأشكال البيانية من خلال قائمة الرسومات.

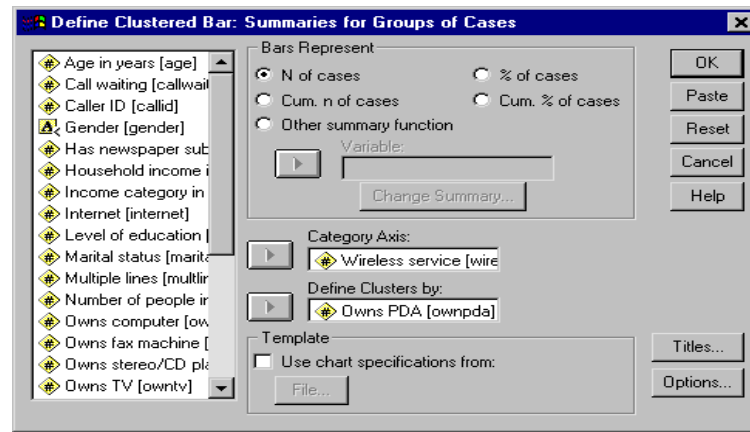
كمثال: يمكنك عرض العلاقة بين خدمة التليفون اللاسلكي ومالك PDA بيانياً باتباع الخطوات التالية.
اختر من القوائم:

Graphs → Bar

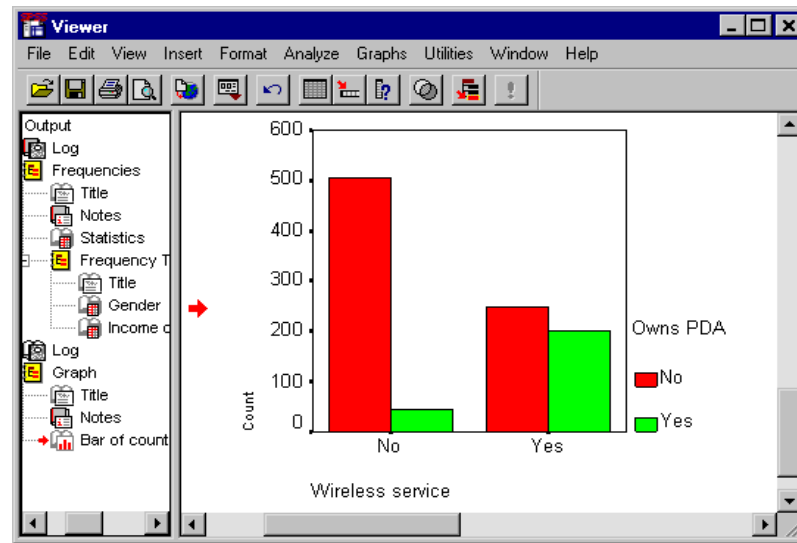
أضغط علي الأعمدة المتلاصقة clustered ثم اضغط Define ، تظهر قائمة المتغيرات



- اختر متغير [wireless] Wireless service متغير خدمة اللاسلكي كمتغير لمحور الفئات (المجموعات) Category Axis
- اختر متغير [ownpda] Owns PDA مالكي جهاز PDA لتعريف الأعمدة المتلاصقة Define clusters by

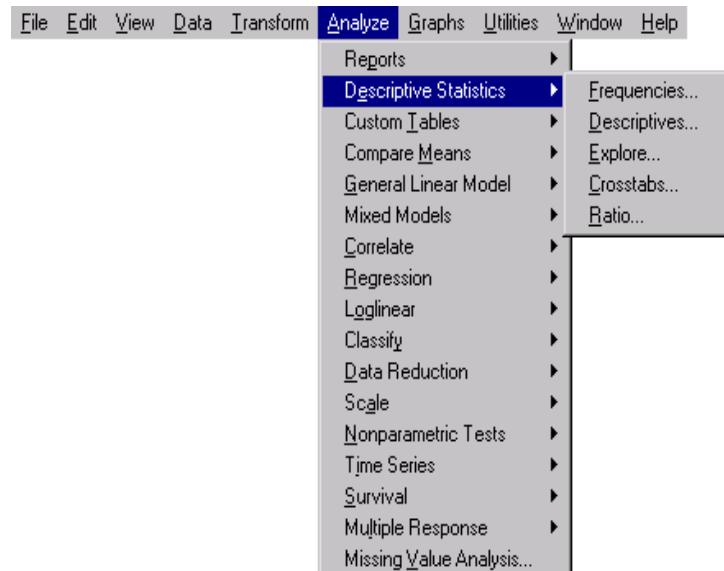


أضغط على Ok يظهر لك الشكل البياني في نافذة Viewer.

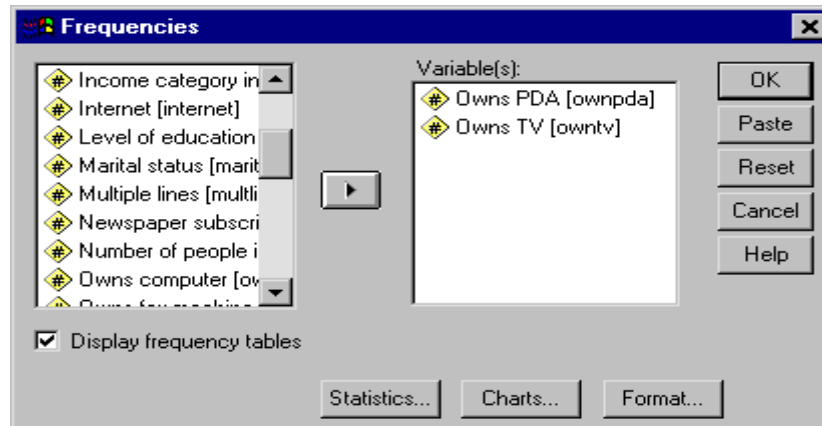


يتم أسلوب التكرارات Frequencies Procedure الباحث بالجدول التكرارية التي تعرض كل من العدد ونسبة الحالات لكل قيمة مشاهدة من المتغيرات ويتم ذلك بإتباع الخطوات التالية
اختر من القوائم

Analyze → Descriptive Statistics → Frequencies



يظهر صندوق حوار التكرارات frequencies



- اختر متغير Owns PDA (ownpda) ومتغير Owns TV (owntv) وحركهم لقائمة المتغيرات
- اختر Ok لتنفيذ الإجراء فتظهر لنا الجداول التكرارية في نافذة العرض Viewer Window

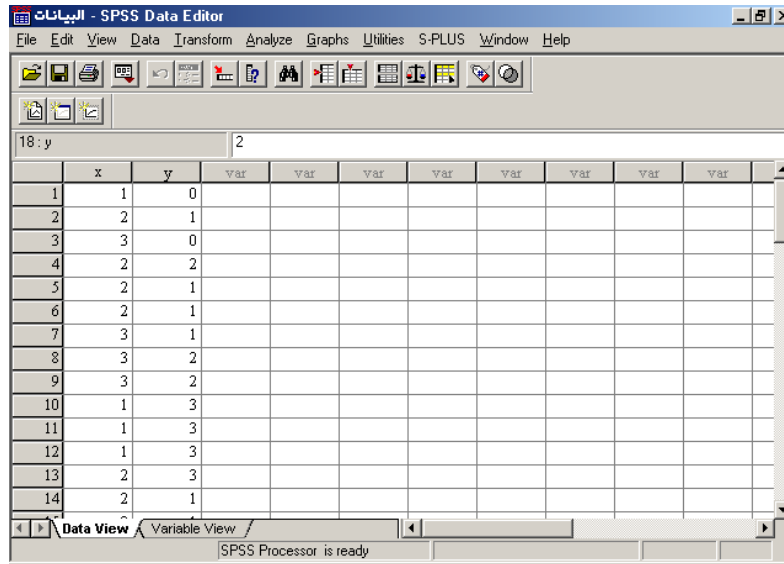
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid No	5077	79.3	79.3	79.3
Yes	1323	20.7	20.7	100.0
Total	6400	100.0	100.0	

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid No	54	.8	.8	.8
Yes	6346	99.2	99.2	100.0
Total	6400	100.0	100.0	

توضح الجداول التكرارية أن حوالي ٢١% فقط من الأشخاص عندهم PDAS ، لكن في الغالب كل واحد تقريباً عنده TV وذلك بنسبة (٩٩.٢%) وهذا التوضيح له أهمية كبيرة ومع ذلك قد يكون ذات أهمية لإظهار أكثر المجموعة الصغيرة من الأشخاص الذين لا يملكون TV

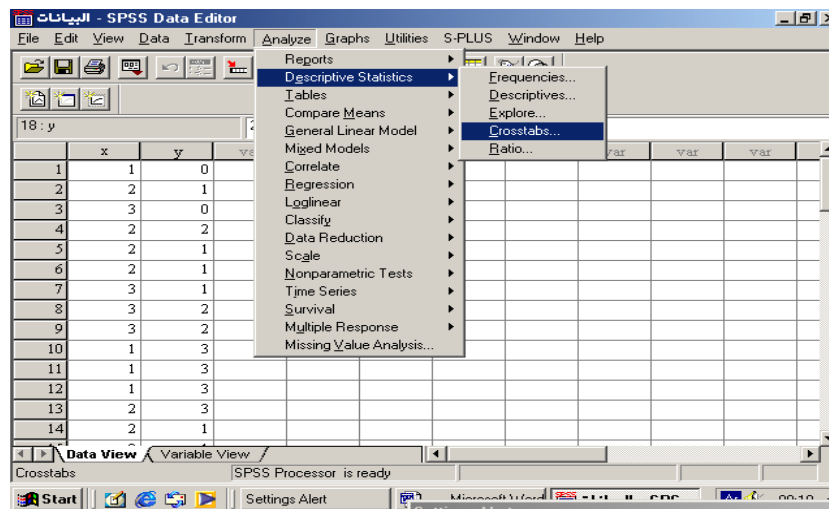
. الجداول المزدوجة:

في حالة وجود متغيرين مثل المؤهل وعدد مرات الغياب فإنه يفضل استخدام الجداول التكرارية المزدوجة كما يلي:

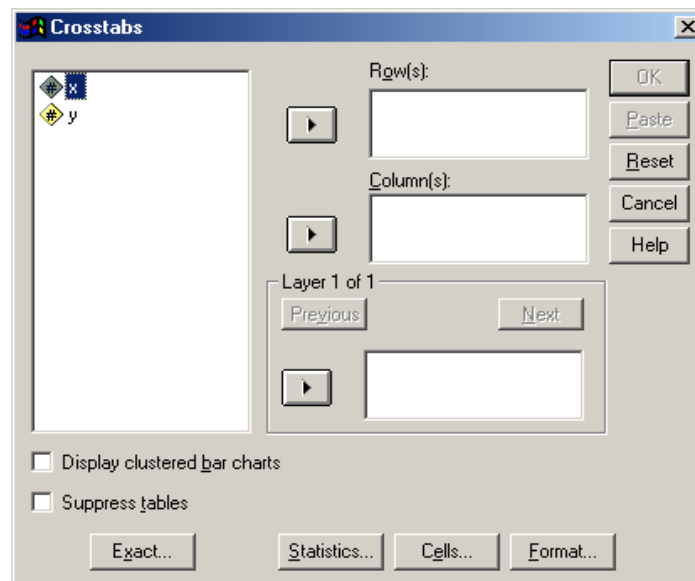


ويتم ذلك بإتباع الخطوات التالية
اختر من القوائم

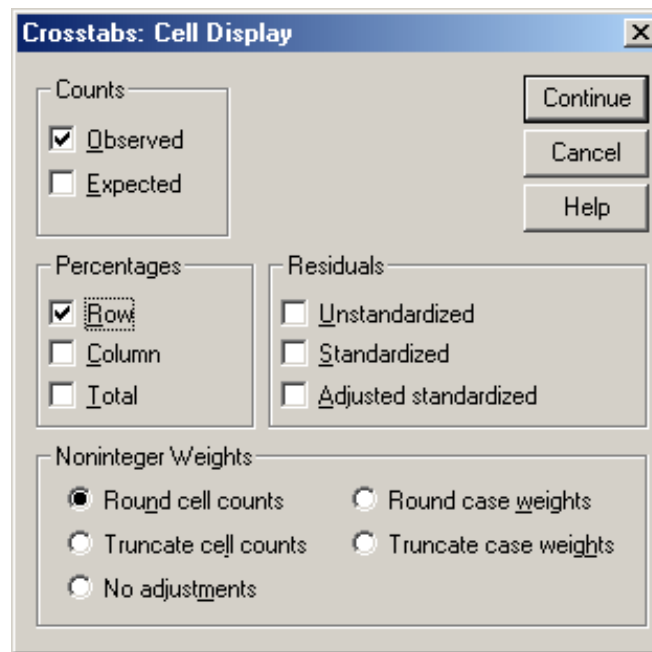
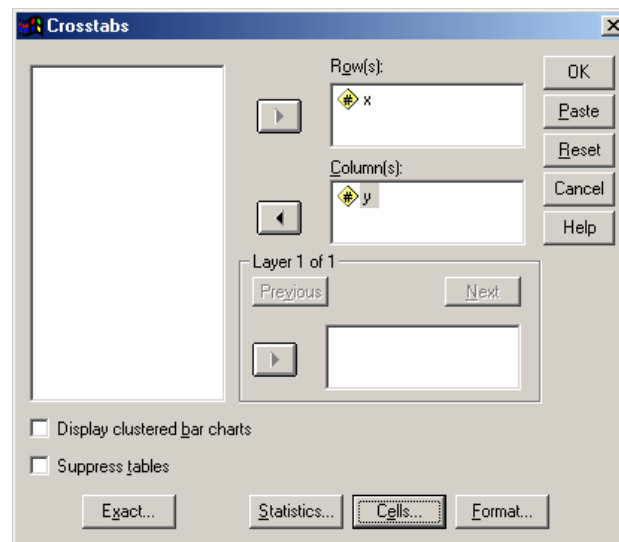
Analyze → Descriptive Statistics →



فتظهر الشاشة



نحدد المتغيرات المؤهل x في الصف وعدد مرات الغياب y في العمود بالضغط على cells يتم تحديد النسبة المئوية للصف أو للعمود



وبالضغط على Continue نرجع للشاشة السابقة ونضغط على OK نحصل على Output

Output1 - SPSS Viewer

File Edit View Data Transform Insert Format Analyze S-PLUS Graphs Utilities Window Help

Input

- Crosstabs
- Notes
- Crosstabs
- Title
- Notes
- x * y Crosstabulation

x * y Crosstabulation

			y			
			0	1	2	3
x	1	Count	1	1	4	3
		% within x	11.1%	11.1%	44.4%	33.3%
	2	Count	2	8	2	1
		% within x	15.4%	61.5%	15.4%	7.7%
	3	Count	3	3	2	0
		% within x	37.5%	37.5%	25.0%	.0%
Total		Count	6	12	8	4
		% within x	20.0%	40.0%	26.7%	13.3%

SPSS Processor is ready

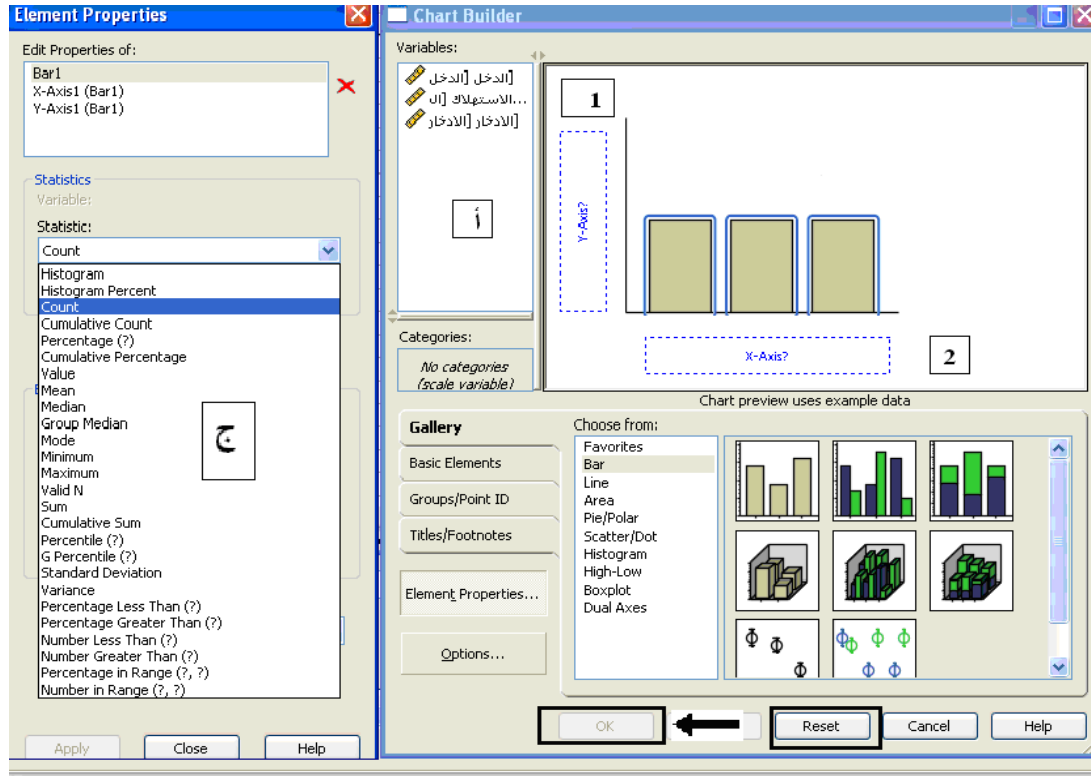
مثال: لدينا المتغير Salary (X) (الأجر) يأخذ القيم التالية:

100, 150, 200, 250, 300, 500, 270, 300, 350, 700, 350, 400, 550, 600, 850
وقد تم ادخال قيم المتغير في شاشة Data Editor. وتم تخصيص كود لكل قيمة من قيم المتغير حسب الترتيب التالي:

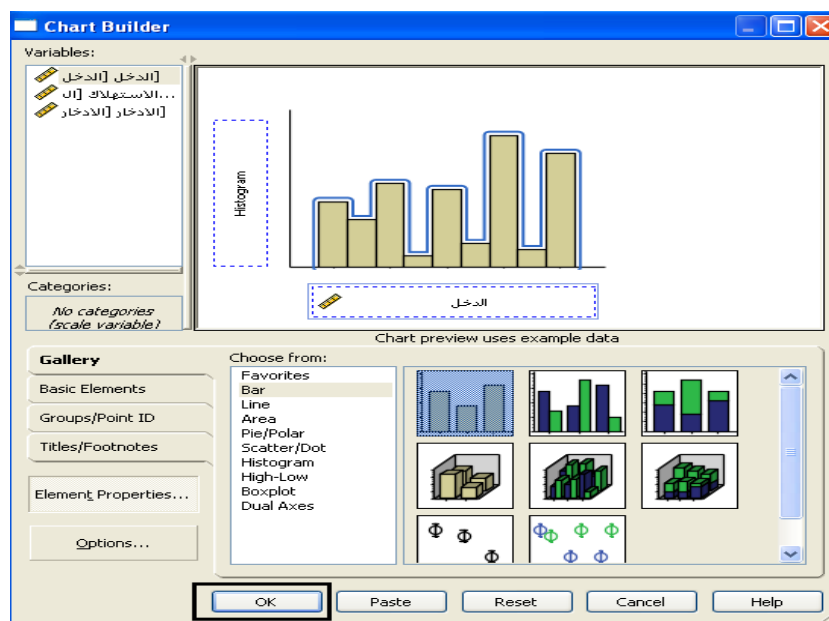
الكود	الفئة
1	150 فأقل
2	151-300
3	301-450
4	451-600
5	601 فأكثر

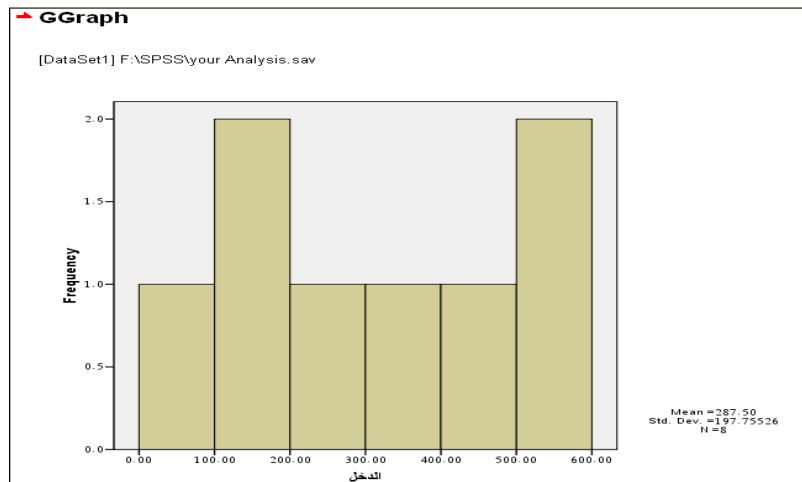
نرغب في ترميز المتغير Salary حسب الفئات المذكورة وخرن الرموز في متغير مختلف.
⊕ "2" وهو المحور X في الشكل البياني المراد رسمه ونفترض أننا نريد رسم المتغير الدخل على المحور X فإننا ننقر عليه ونسحبه إلى محور X والمحدد في الشكل برقم 2.

- ⊕ "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 ادناه.
- ⊕ "ج" وهو عمود يحمل اسم Statistics وهو عمود العمليات التي نريد اظهارها على المحور Y ونستطيع الاختيار المتعدد لها ولنفرض أننا نريد اظهار تكرار قيم متغير الدخل فإننا نختار Histogram من (ج).



- ⊕ بعد الإنتهاء من العمليات السابقة يصبح الشكل السابق على الشكل التالي ومن ثم النقر على OK.

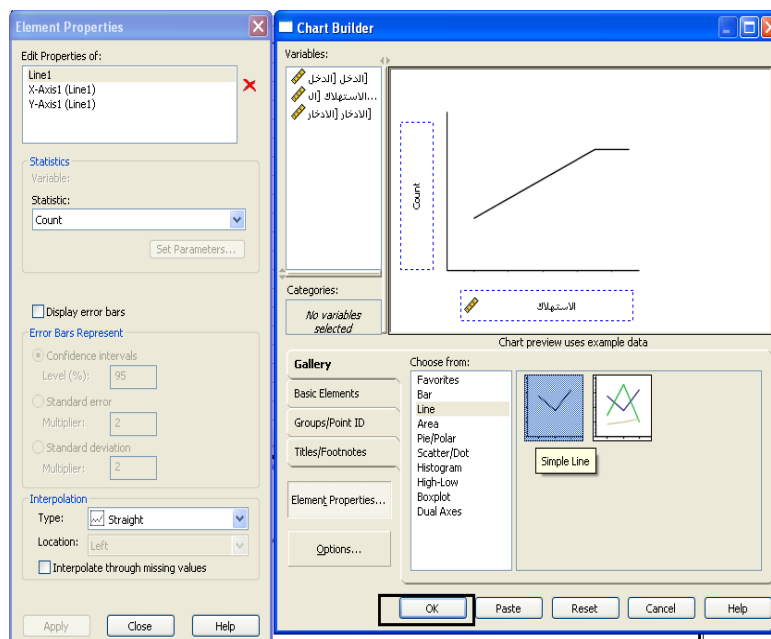




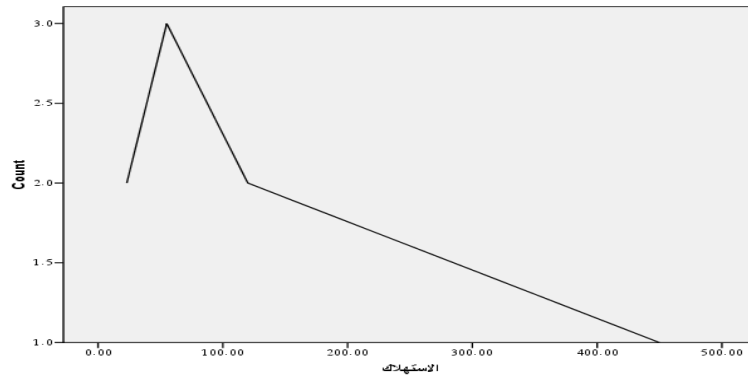
ونلاحظ إن الشكل السابق يوضح الرسم البياني لمتغير الدخل وبجوار الرسم بعض المقاييس الإحصائية عن المتغير مثل الوسط الحسابي mean يساوي 287.5، الانحراف المعياري يساوي 197.75526 وعدد البيانات يساوي 8.

Line Graphs

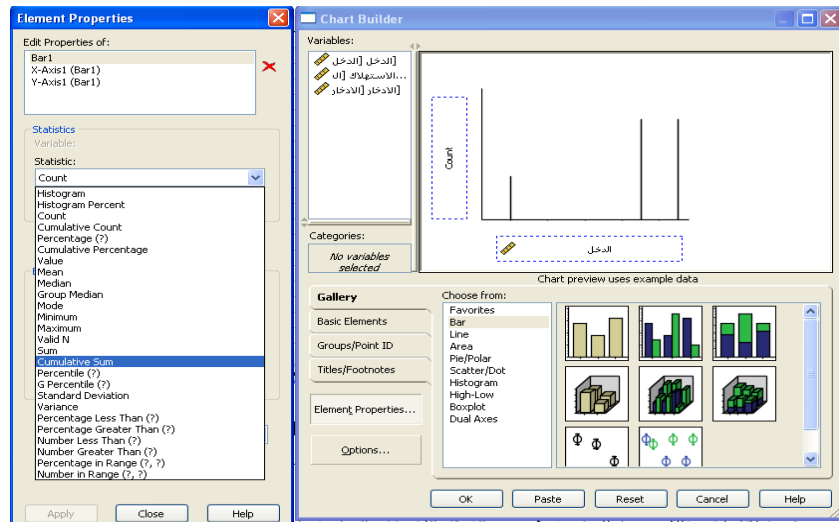
إذا أردنا عمل شكل بياني من نوع line للمتغير الاستهلاك فإننا نختار Graphs ثم نقوم باختيار Chart Builder... فيظهر الشكل فنختار منه Line من عمود Choose from: ثم نقوم باختيار الشكل الأول وهو Simple Line وذلك بالنقر عليه نقراً مزدوجاً ومن ثم اختيار المتغير الاستهلاك من عمود المتغيرات بنفس طريقة اختيار الدخل السابق وذلك بسحبها إلى عمود X فيظهر الشكل الآتي:



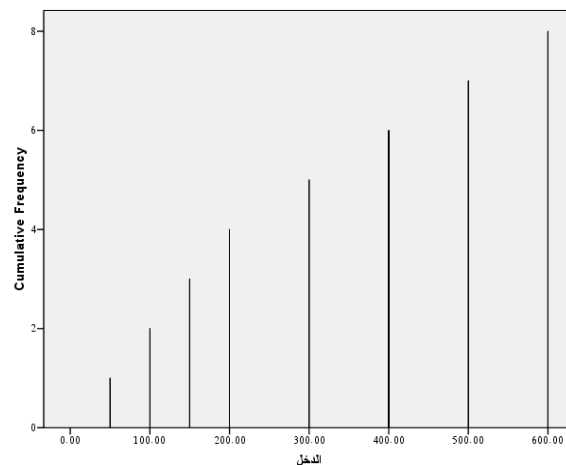
وبعد النقر علي ok لتأكيد الاختيار فإننا الشكل الآتي يظهر في نافذة output.



إذا أردنا عمل شكل بياني من نوع Graphs for cumulative frequency للمتغير الدخل فإننا نختار Graphs ثم نقوم باختيار Chart Builder... فيظهر الشكل فنختار منه Bar من عمود Choose from: ثم نقوم باختيار الشكل الأول وهو Simple Bar وذلك بالنقر عليه نقراً مزدوجاً ومن ثم اختيار المتغير الدخل من عمود المتغيرات وذلك بسحبه إلى عمود X ومن الجزء الذي على اليسار يتم الاختيار فيظهر الشكل الآتي واختيار cumulative sum من عمود Statistic:



وبعد ذلك النقر علي Ok فيظهر الشكل الآتي :



الاسبوع 7 - 9

مقاييس النزعة المركزية والتشتت: Measures of Central tendency and Dispersion

نسعى من وراء دراستنا لظاهرة إحصائية معينة إلى استنتاج صفة مميزة ما أو أكثر من صفة فعند دراستنا لدرجات طلاب السنة التمهيدية في مادة الرياضيات نستنتج أن معظم الدرجات تميل إلى التركز حول نقطة أو درجة معينة وهي درجة النجاح بينما نجد أن الدرجات المتطرفة الأخرى من الدرجات المدروسة ، قليلة نوعاً ما وهذه الدرجات ، المتطرفة ، هي الدرجات العالية جداً أو المنخفضة جداً أو كلاهما . نسمي هذا الميل بالتركز أو بالنزعة المركزية ، ونسمي القيم التي تتركز حولها القيم الأخرى مقاييس النزعة المركزية Measures of Central location ولهذه المقاييس أهمية كبيرة في علم الإحصاء فهي تعطينا فكرة عامة عن قيم الظاهرة المدروسة وبالتالي مقارنة مجموعتين أو أكثر . هناك عدة مقاييس لهذا الغرض منها الوسط الحسابي أو المتوسط (المعدل) Mean الوسيط Median والمنوال Mode .

١. الوسط الحسابي أو المتوسط: Mean :

إن الوسط الحسابي (المعدل) لمجموعة البيانات $x_1, x_2, \dots, x_n$ هو مجموع هذه البيانات مقسوماً على عددها ويرمز لذلك بالرمز $\bar{x}$ ويعطى بالعلاقة التالية :

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i \\ &= \frac{x_1 + x_2 + \dots + x_n}{n}\end{aligned}\quad (1)$$

أما إذا كانت البيانات معطاة بجدول توزيع تكراري ذو k فئة فإن الوسط الحسابي يعطى بالعلاقة :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k x_i f_i \quad (2)$$

حيث رمزنا بـ f_i ، x_i لتكرار الفئة i th ومركزها على الترتيب وكذلك $n = \sum_{i=1}^k f_i$

أمثلة :

(a) عدد الرسائل اليومية التي تلقاها مكتب تجاري خلال ثمانية أيام عمل هي كالاتي :

2 , 0 , 3 , 15 , 4 , 13 , 5 , 6

إن متوسط عدد الرسائل اليومية في هذه الحالة هو :

$$\bar{x} = \frac{1}{8} [2 + 0 + 3 + 15 + 4 + 13 + 5 + 6] = \frac{48}{8} = 6$$

(b) إن الوسط الحسابي أو المتوسط للبيانات الواردة في جدول التوزيع التكراري ، يحسب كالتالي :

x_i	21	24	27	30	33	36	39
f_i	3	7	23	37	24	5	1

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{\sum_{i=1}^k f_i} = \frac{\sum_{i=1}^8 x_i f_i}{\sum_{i=1}^8 f_i} = \frac{1}{100} [21(3) + 24(7) + 27(23) + 30(37) + 33(24) + 36(5) + 39(1)] = 29.73$$

الوسط الحسابي يتمتع بعدة خواص منها أن مجموع انحرافات قيم مجموعة بيانات عن وسطها الحسابي يساوي الصفر أي :

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

وذلك بالنسبة للقيم الافرادية. أما بالنسبة للقيم المعطاة في جدول توزيع تكراري فهي :

$$\sum_{i=1}^k (x_i - \bar{x}) f_i = 0$$

مثال

فيما يلي درجات 8 طلاب في مادة الإحصاء .

40	36	40	35	37	42	32	34
----	----	----	----	----	----	----	----

والمطلوب إيجاد الوسط الحسابي لدرجة الطالب في الامتحان .

الحل

لإيجاد الوسط الحسابي للدرجات تطبق المعادلة رقم (1) كما يلي:

$$\begin{aligned} \bar{x} &= \frac{x_1 + x_2 + \dots + x_n}{n} \\ &= \frac{34 + 32 + 42 + 37 + 35 + 40 + 36 + 40}{8} = \frac{296}{8} = 37 \end{aligned}$$

أي أن الوسط الحسابي لدرجة الطالب في اختبار مادة الإحصاء يساوي ٣٧ درجة
مثال الجدول التالي يعرض توزيع 40 تلميذ حسب أوزانهم .

فئات الوزن	32-34	34-36	36-38	38-40	40-42	42-44
عدد التلاميذ	4	7	13	10	5	1

والمطلوب إيجاد الوسط الحسابي.

الحل

لحساب الوسط الحسابي باستخدام المعادلة رقم (2) يتم إتباع الخطوات التالية :

١- إيجاد مجموع التكرارات $\sum f$. ٢- حساب مراكز الفئات x .

٣- ضرب مركز الفئة في التكرار المناظر له $(x f)$ ، وحساب المجموع $\sum xf$

٤- حساب الوسط الحسابي بتطبيق المعادلة رقم (2) .

فئات الوزن (C)	التكرارات f	مراكز الفئات x	$x \cdot f$
32-34	4	$(32+34) \div 2 = 33$	$4 \times 33 = 132$
35-37	7	36	$7 \times 36 = 252$
38-40	13	39	$13 \times 39 = 507$
41-43	10	42	$10 \times 42 = 420$
44-46	5	45	$5 \times 45 = 225$
47-49	1	48	$1 \times 48 = 48$
المجموع	40		1584

إذا الوسط الحسابي لوزن التلميذ هو :

$$\bar{x} = \frac{\sum_{i=1}^6 x_i f_i}{\sum_{i=1}^6 f_i} = \frac{1584}{40} = 39.6 \text{ k.g}$$

أي أن متوسط وزن التلميذ يساوي 39.6 kg
مثال: الجدول التالي يمثل الأجر الأسبوعي للعامل بالدولار في مائتين محل بمنطقة بغداد الجديدة :-

الأجر الأسبوعي بالدولار	5 -	15 -	25 -	35 -	55 -45	المجموع
عدد المحلات	30	20	60	50	40	200

المطلوب حساب متوسط الأجر الأسبوعي للعامل.
يمكن الاستفادة من خصائص الوسط الحسابي في حل المثال بثلاث طرق لتحسين سرعة الحساب الحل المطول:

الفئات	التكرار	مراكز الفئات	$F \cdot X$
5 -	30	$\frac{15 + 5}{2} = 10$	$30 \times 10 = 300$
15 -	20	$\frac{25 + 15}{2} = 20$	$20 \times 20 = 400$
25 -	60	$\frac{35 + 25}{2} = 30$	$60 \times 30 = 1800$
35 -	50	$\frac{45 + 35}{2} = 40$	$50 \times 40 = 2000$
45 -	40	$\frac{55 + 45}{2} = 50$	$40 \times 50 = 2000$
	Sum 200		6500

$$\bar{x} = \frac{\sum FX}{\sum F} = \frac{6500}{200} = 32.5$$

أي أن متوسط الأجر الأسبوعي للعامل هو 32.5 دولار.

الطريقة المختصرة:- وهنا نطرح وسطا فرضيا (مقدار ثابت) من مراكز الفئات ثم نعيد إضافته إلى الوسط الحسابي بعد حسابه من مراكز الفئات (المعدلة) .

وذلك للتغلب على الصعوبات في الحساب عندما تكون مراكز الفئات أرقام كبيرة أو كسرية .

الفئات	التكرار f	مراكز الفئات X	مراكز الفئات المعدلة $X_1 = x_i - 30$	$F * X_1$
5 -	30	10	-20	$30 * (-20) = -600$
15 -	20	20	-10	$20 * (-10) = -200$
25 -	60	30	0	$60 * 0 = 0$
35 -	50	40	10	$50 * 10 = 500$
45 -	40	50	20	$40 * 20 = 800$
	Sum 200			500

$$\bar{x} = \frac{\sum FX_1}{\sum F} + 30 = \frac{500}{200} + 30 = 32.5$$

الطريقة الأكثر اختصارا: وهنا نقسم مراكز الفئات المعدلة (سابقا) على مقدار ثابت ثم نعيد ضربه في الوسط الحسابي بعد حسابه من مراكز الفئات (النهائية).

الفئات	التكرار f	مراكز الفئات X	مراكز الفئات المعدلة $X_1 = x - 30$	مراكز الفئات النهائية $X_2 = x_1 / 10$	$F * X_2$
5 -	30	10	-20	-2	$30 * (-20) = -600$
10	20	20	-10	-1	$20 * (-10) = -200$
20	60	30	0	0	$60 * 0 = 0$
30	50	40	10	1	$50 * 10 = 500$
40	40	50	20	2	$40 * 20 = 800$
	Sum 200				500

$$\bar{x} = \frac{\sum FX_2}{\sum F} * 10 + 30 = \frac{500}{200} * 10 + 30 = 32.5$$

عموما إذا كان الجدول التكراري منتظما (أطوال الفئات متساوية) فانه يمكن وضع صفر أمام أي فئة، ووضع الأرقام 1-، 2-، 3-... أمام الفئات السابقة لهذه الفئة ، ووضع الأرقام 1، 2، ... أمام الفئات التالية لها .

مثال: فيما يلي درجات 8 طلاب في مادة إحصاء تطبيقي .

34 32 42 37 35 40 36 40

والمطلوب إيجاد الوسط الحسابي لدرجة الطالب في الامتحان .

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$$= \frac{34 + 32 + 42 + 37 + 35 + 40 + 36 + 40}{8} = \frac{296}{8} = 37$$

مثال: الجدول التالي يعرض توزيع 40 تلميذ حسب أوزانهم .

فئات الوزن	32-34	34-36	36-38	38-40	40-42	42-44
عدد التلاميذ	4	7	13	10	5	1

الحل:-

فئات الوزن (C)	التكرار f	مراكز الفئة x	$x * f$
32-34	4	$(32+34)/2=33$	$4 \times 33=132$
34-36	7	35	$7 \times 35=245$
36-38	13	37	481
38-40	10	39	390
40-42	5	41	205
42-44	1	43	43
المجموع	40		1496

$$\bar{X} = \frac{\sum_{i=1}^6 x_i f_i}{\sum_{i=1}^6 f_i} = \frac{1496}{40} = 37.4 \text{ k.g}$$

اهم خصائص الوسط الحسابي :-

• الوسط الحسابي للمقدار الثابت يساوى الثابت نفسه .

$$\bar{X} = \frac{a + a + \dots + a}{n} = \frac{na}{n} = a$$

$$\bar{x} = \frac{63+63+63+63+63}{5} = \frac{315}{5} = 63 \text{ k.g}$$

- مجموع انحرافات القيم عن وسطها الحسابي يساوي صفر.

$$\sum (x - \bar{x}) = 0$$

$$\bar{x} = \frac{34+32+42+37+35+40+36+40}{8} = \frac{296}{8} = 37$$

x	34	32	42	37	35	40	36	40	296
(x-37)	34-37	32-37	42-37	37-37	35-37	40-37	36-37	40-37	0
	-3	-5	5	0	-2	3	-1	3	

- إذا أضيف مقدار ثابت إلى كل قيمة من القيم ، فإن الوسط الحسابي للقيم المعدلة (بعد الإضافة) يساوي الوسط الحسابي للقيم الأصلية (قبل الإضافة) مضافا إليها هذا المقدار الثابت.

x	34	32	42	37	35	40	36	40	296
(x-37)	34-37	32-37	42-37	37-37	35-37	40-37	36-37	40-37	0
	-3	-5	5	0	-2	3	-1	3	

X	34	32	42	37	35	40	36	40	296
Y=x+5	34+5	32+5	42+5	37+5	35+5	40+5	36+5	40+5	336
	39	37	47	42	40	45	41	45	

$$\bar{y} = \frac{\sum y}{n} = \frac{336}{8} = 42 \rightarrow (\bar{x} + 5 = 37 + 5 = 42)$$

- الوسط الحسابي يتأثر بالعمليات الحسابية الأربع (الجمع-الطرح-الضرب-القسمة)
- الوسط الحسابي يتأثر بالقيم الشاذة "المتطرفة"

مزاي و عيوب الوسط الحسابي:-

يتميز الوسط الحسابي بالمزايا التالية :

- أنه سهل الحساب .
- يأخذ في الاعتبار كل القيم .
- أنه أكثر المقاييس استخداماً وفهماً .

ومن عيوبه .

- أنه يتأثر بالقيم الشاذة والمتطرفة .
- يصعب حسابه في حالة البيانات الوصفية .
- يصعب حسابه في حالة الجداول التكرارية المفتوحة .

2- الوسيط: Median

عند ترتيبنا لقيم ظاهرة إحصائية ما تصاعدياً أو تنازلياً يمكننا أن نعين وضعية إحدى هذه القيم ولتكن x_j ، وقولنا أن ربع قيم المجتمع هي أعلى من هذه القيمة وثلاثة أرباع القيم أقل منها أو أن ثلثها أقل من هذه القيمة x_j وثلثها أكبر منها . نسمي مثل هذه المقاييس مقاييس وضعية ، حيث يعتبر الوسيط أحد أهم هذه المقاييس . وبالتعريف : الوسيط هو القيمة التي يتساوى على طرفيها عدد القيم بعد ترتيبها تصاعدياً بحيث تكون كل قيمة من القيم التي تسبقه أصغر منه وكل قيمة من القيم التي تليه أكبر منه . أما إذا كانت القيم مرتبة تنازلياً فتكون القيم التي تسبقه أكبر والتي تليه أصغر . فإذا كان عدد هذه القيم فردياً عددها n (حيث n عدد فردي) فالوسيط هو القيمة النصفية التي تقسم هذه القيم ، أما إذا كان عدد القيم زوجياً فالوسيط هو الوسط الحسابي لمجموع القيمتين الوسطيتين ويرمز للوسيط بالرمز $\tilde{x}$ وهو القراءة التي ترتيبها $\frac{n+1}{2}$ في حالة n عدد فردي.

أما إذا كان n عدداً زوجياً فالوسيط هو متوسط القراءتين $\frac{n}{2}$ و $\frac{n}{2}+1$

أمثلة :

(a) القيم 3,-2,5,6,1,2,10 والتي تصبح بعد ترتيبها تصاعدياً -2,1,2,3,5,6,10- أوجد الوسيط لهذه البيانات .

الحل : بما أن $n = 7$ والوسيط هو القيمة النصفية وبالتالي : $\frac{7+1}{2} = 4$

أي أن القيمة المطلوبة هي الرابعة وبالتالي فإن $\tilde{x} = 3$

(b) لتكن القيم 4,-5,6,-1,3,8,10,11 والتي تصبح بعد ترتيبها تصاعدياً -5,-1,3,4,6,8,10,11- أوجد وسيط هذه البيانات .

الحل :

بما أن $n = 8$ وعددها زوجياً فإن الوسيط هو متوسط القيمتين الوسيطتين، أي $\frac{8+1}{2} = 4.5$ وبالتالي فهو الوسط الحسابي

للقيمتين الرابعة والخامسة

$$\tilde{x} = \frac{4+6}{2} = 5$$

و تستخدم هذه الطريقة إذا كان عدد البيانات صغيراً . أما إذا كان عدد البيانات كبيراً و غير مرتب في جدول تكراري فإن هناك طريقة عامة سنذكرها في الفقرة القادمة.

أما إذا كانت القيم معطاة في جدول توزيع تكراري وموزعة توزيعاً عادلاً ضمن فئات متساوية الطول فإن الفئة الوسطية هي أول فئة يزيد تكرارها المتجمع الصاعد عن $\frac{n}{2}$ أو يساويه حيث n مجموع التكرارات ويعطى الوسيط $\tilde{x}$ في هذه الحالة

بالعلاقة التالية :

$$\tilde{x} = a + \frac{\frac{n}{2} - k}{j} \cdot \Delta$$

حيث نرسم بـ a للحد الأدنى الفعلي للفئة الوسطية ، Δ طول الفئة ، k مجموع تكرارات الفئات التي تسبق a أما z فهو تكرار الفئة الوسطية وأخيراً $n = f_1 + f_2 + \dots + f_k$. لاحظ هنا أن العلاقة السابقة هي علاقة عامة حيث الوسيط إنما هو P_{50} وبالتالي فيمكن حساب أي منوي بنفس الطريقة .
(c) أوجد الوسيط لمجموعة القيم المعطاة بالجدول التالي :

التكرار التجميعي الصاعد	مركز الفئات x_i	التكرار f_i	حدود الفئات	الفئات
3	10	3	5-14	1
8	20	5	15-24	2
20	30	12	25-34	3
45	40	25	35-44	4
80	50	35	45-54	5
93	60	13	55-64	6
100	70	7	65-75	7

لاحظ عمود التكرار التجميعي الصاعد تجد أن :

$$\frac{n}{2} = \frac{100}{2} = 50$$

وبالتالي فالفئة الخامسة تحتوي على الوسيط وحدها الأدنى 45 وبالتالي فإن حدها الأدنى الفعلي هو 44.5 . فبتطبيق العلاقة السابقة (3) نجد أن وسيط البيانات هو

$$\tilde{x} = 44.5 + \frac{50 - 45}{35} 10 = 45.9$$

مثال: احسب الوسيط للقيم 6،5،4،3،112
الترتيب التصاعدي للقيم

$$(3)_1 \& (4)_2 \& (5)_3 \& (6)_4 \& (112)_5$$

$$n = 5 [odd]$$

$$Median Order = \frac{5+1}{2} = 3$$

مثال: احسب الوسيط للقيم

$$(-3)1 \& (-1)2 \& (3)3 \& (6)4 \& (7)5 \& (8)6$$

$$n = 6 [even]$$

$$Median Orders = \frac{6}{2} \& \frac{6}{2} + 1 = 3 \& 4$$

$$Median = \frac{3+6}{2} = 4.5$$

ثانيا : في حالة البيانات المبوبة:-

الوسيط هو القيمة المقابلة لنصف مجموع التكرارات لذلك

$$\frac{\sum F}{2} = \text{رتبة الوسيط}$$

يجب حساب الوسيط من أحد الجدولين التكراريين المتجمعين الصاعد أو الهابط *في حالة الحساب من الجدول التكراري المتجمع الصاعد (بعد تكوينه).
ويحسب الوسيط ، بتطبيق المعادلة

$$\text{الوسيط} = \frac{\text{الحد الأدنى لفئة الوسيط} + \text{طول فئة الوسيط} \left(\text{رتبة الوسيط} - \text{التكرار المتجمع الصاعد السابق لفئة الوسيط} \right)}{\text{التكرار الأصلي لفئة الوسيط}}$$

تكرار متجمع صاعد سابق f_1	الحد الأدنى لفئة الوسيط (A)
رتبة الوسيط $(n/2)$	الوسيط Med
تكرار متجمع صاعد لاحق f_2	الحد الأعلى لفئة الوسيط

$$Med = A + \left\{ \left[\frac{\left(\frac{n}{2} \right) - f_1}{f_2 - f_1} \right] \bullet L \right\}$$

مثال الوسيط للبيانات المبوبة:-

فيما سبق توزيع 50 عجل متوسط الحجم ، حسب احتياجاته اليومية من الغذاء الجاف بالكيلوجرام والمطلوب حساب الوسيط : أ - حسابيا ب- بيانيا

فئات الاحتياجات اليومية	1.5 -	4.5 -	7.5 -	10.5 -	13.5 – 16.5
عدد العجول f	4	12	19	10	5

$$\frac{n}{2} = \frac{\sum f}{2} = \frac{50}{2} = 25 \quad \text{رتبة الوسيط :}$$

أقل من	تكرار متجمع صاعد	
1.5	0	
4.5	4	
A 7.5	f_1 16	
Med (الوسط)	25	رتبة الوسط
10.5	f_2 35	
13.5	45	
16.5	50	

$$A = 7.5, f_1 = 16, f_2 = 35, L = 10.5 - 7.5 = 3$$

$$Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 7.5 + \frac{25 - 16}{35 - 16} \times 3$$

$$= 7.5 + \frac{9}{19} \times 3 = 7.5 + \frac{27}{19} = 7.5 + 1.421 = 8.921 \text{ k.g}$$



مثال: الجدول التالي يمثل الأجر الأسبوعي للعامل بالدولار في مائتين محل بمنطقة بغداد الجديدة :-

الأجر الأسبوعي بالدولار	5 -	15 -	25 -	35 -	55 - 45	المجموع
عدد المحلات	30	20	60	50	40	200

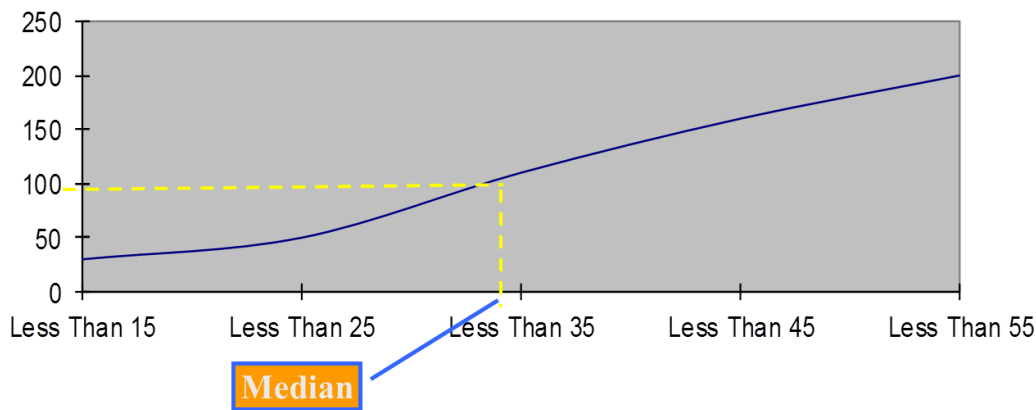
المطلوب حساب متوسط الأجر الأسبوعي للعامل باستخدام الوسيط

الترتيب التجميعي الصاعد	الحدود العليا للفئات	الترتيب f_i	حدود الفئات	الفئات
30	Less than 15	30	5-15	1
50	Less than 25	20	15-25	2
110	Less than 35	60	25-35	3
160	Less than 45	50	35-45	4
200	Less than 55	40	45-55	5

$$100 = \frac{200}{2} = \frac{\sum F}{2} = \text{رتبة الوسيط}$$

$$\begin{aligned} \text{Median} &= 25 + 10 \left(\frac{100 - 50}{60} \right) \\ &= 25 + 10 \left(\frac{50}{60} \right) \\ &= 33.33 \$ \end{aligned}$$

المنحنى المتجمع الصاعد



مثال: تم تقسيم قطعة أرض زراعية إلى 17 وحدة تجريبية متشابهة ، وتم زراعتها بمحصول القمح ، وتم استخدام نوعين من التسميد هما : النوع (a) وجرب على 7 وحدات تجريبية ، والنوع (b) وجرب على 10 وحدات تجريبية ، وبعد انتهاء الموسم الزراعي ، تم تسجيل إنتاجية الوحدة بالطن / هكتار ، وكانت على النحو التالي : والمطلوب حساب وسيط الإنتاج لكل نوع من السماد المستخدم، ثم قارن بينهما.

النوع (a)	1.2	2.75	3.25	2	3	2.3	1.5			
النوع (b)	4.5	1.8	3.5	3.75	2	2.5	1.5	4	2.5	3

$$(n=7) \quad , \quad ((n+1)/2 = (7+1)/2 = 4) \quad , \quad Med_a = 2.3$$

	قيمة الوسيط						
الإنجاز	1.2	1.5	2	2.3	2.75	3	3.25
الرتبة	1	2	2	4	5	6	7
	رتبة الوسيط						

$$(n=10) \quad , \quad ((n+1)/2 = (10+1)/2 = 5.5) \quad , \quad Med_b = \frac{2.5+3}{2} = 2.75$$

	$\frac{2.5+3}{2} =$ قيمة الوسيط									
الإنجاز	1.5	1.8	2	2.5	2.5	3	3.5	3.75	4	4.5
الرتبة	1	2	3	4	5	6	7	8	9	10
	رتبة الوسيط									

يفضل استخدام الوسيط في حالة التعامل مع:-

- البيانات التي تكثر بها القيم الشاذة .
- الجداول التكرارية المفتوحة من أحد طرفيها أو من كليهما .
- التوزيعات التكرارية غير المتساوية في طول الفئات .

مزايا وعيوب الوسيط:-

من مزايا الوسيط

١. لا يتأثر بالقيم الشاذة أو المتطرفة .

٢. كما أنه سهل في الحساب .

٣. مجموع قيم الانحرافات المطلقة عن الوسيط أقل من مجموع الانحرافات المطلقة عن أي قيم أخرى . أي أن :

$$\sum |x - Med| \leq \sum |x - a| , \quad a \neq Med$$

ومن عيوب الوسيط:

١. أنه لا يأخذ عند حسابه كل القيم في الاعتبار، فهو يعتمد على قيمة أو قيمتين فقط .

٢. يصعب حسابه في حالة البيانات الوصفية المقاسة بمقياس اسمي nominal .

3- المنوال Mode :

يعرف المنوال على أنه هو القيمة الأكثر تردداً أو الأكثر شيوعاً أو تكراراً ونرمز له بالرمز x° فمثلاً للقيم 1,1,2,3,3,4,5,5,5,6,7 منوال واحد وهو $x^\circ = 5$ قد لا يكون لمجموعة من القيم منوال كما هو الحال في القيم 1,2,3,4,5,6,7 وقد يكون لمجموعة من القيم أكثر من منوال واحد كما هو الحال في القيم 1,1,1,2,2,2,3,4,5,5,6 .
وإذا كانت البيانات غير عددية (أي نوعية) كأن نسال مثلاً ستة أشخاص عن أي الألوان المحببة لهم ، قد تكون إجاباتهم كالتالي:

أزرق، أصفر، أبيض، أزرق، أبيض أو أزرق فسيكون المنوال في هذه الحالة هو اللون "الأزرق" .
أما إذا كانت البيانات مبوبة في جدول توزيع تكراري فنحسب المنوال كما يلي : نعين الفئة المنوالية وهي الفئة التي تقابل أكبر قيمة للتكرارات ثم نحسب الفرق بين تكرار الفئة المنوالية وتكرار الفئة التي قبلها Δ_1 ثم نحسب الفرق بين تكرار الفئة المنوالية وتكرار الفئة التي بعدها وليكن Δ_2 نرمز للحد الأدنى الفعلي للفئة المنوالية بالرمز a ولطول الفئة بالرمز Δ فنحصل على المنوال كما يلي :

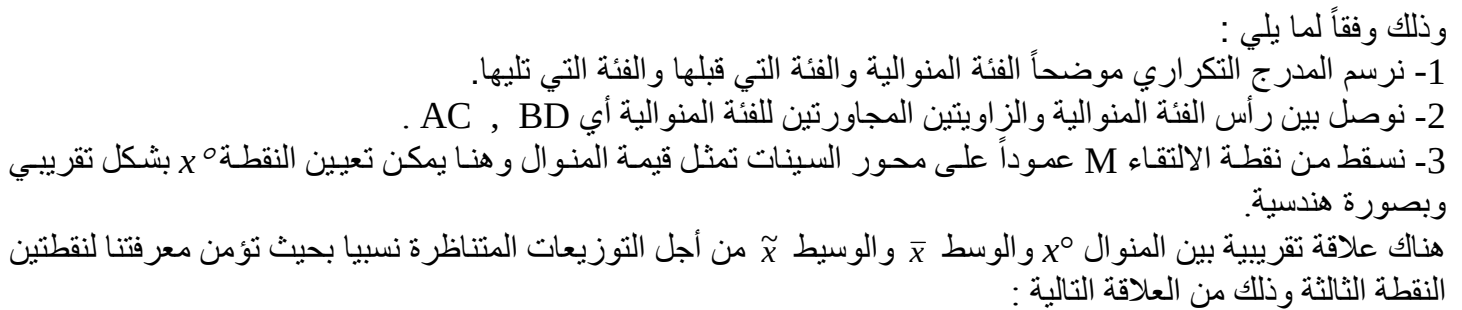
$$x^\circ = a + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot \Delta$$

وكتطبيق مباشر على ذلك ، المنوال للقيم المعطاة بالجدول التالي

التكرار التجميعي الصاعد	مركز الفئات x_i	التكرار f_i	حدود الفئات	الفئات
3	10	3	5-14	1
8	20	5	15-24	2
20	30	12	25-34	3
45	40	25	35-44	4
80	50	35	45-54	5
93	60	13	55-64	6
100	70	7	65-75	7

$$x^\circ = 44.5 + \frac{35 - 25}{10 + 22} \times 10 = 47.63$$

لاحظ أنه يمكن حساب المنوال بيانياً (هندسياً) وذلك بعد رسم المدرج التكراري للبيانات المبوبة وعلى سبيل المثال لنحسب المنوال للبيانات المعطاة بالجدول .



لنحسب مثلاً منوال البيانات الواردة في الجدول 3-1، حيث وجدنا أن $\bar{X} = 45.9$
 أما الوسط الحسابي فمن العلاقة (2) نجد أن $\bar{x} = 45.1$
 وتطبيق العلاقة السابقة نجد $x^{\circ} = 45.1 + 3(45.9 - 45.1) = 47.5$
 لاحظ الفرق بين المنوال هنا والمنوال في العلاقة (5) .

(a) لتكن لدينا المجموعة التالية من القراءات والتي تمثل أطوال عينة من 48 سمكة مقدرة بالملم أجريت في أحد المختبرات :

231	234	214	216	210	221	223	215
210	209	207	206	204	205	203	216
218	219	217	212	216	221	226	222
210	206	203	209	230	225	215	235
230	240	190	208	195	201	211	313
213	210	205	190	270	220	250	210

61

2- احسب متوسط طول السمكة الواحدة ثم احسب التشتت والانحراف المعياري .

الحل :

لاحظ أن أكبر قيمة وأصغر قيمة هما 313 , 190 على الترتيب ويكون المدى :

$$R = 313 - 190 = 123$$

لنقسم هذه القراءات إلى 6 فئات متساوية فيكون طول الفئة الواحدة

$$123/6 = 20.5 \sim 21 .$$

ويصبح جدول التوزيع التكراري من الشكل :

رقم الفئة	حدود الفئات	مراكز الفئات x_i	التكرار f_i
1	190-210	200	20
2	211-231	221	21
3	232-252	242	4
4	253-273	263	1
5	274-294	284	1
6	295-315	305	1
			48

أما المتوسط فممن (2) نجد :

$$\bar{x} = \frac{1}{48} [200(20) + 221(21) + 242(4) + 263(1) + 305(1) + 248(1)] = 217.93mm$$

(b) أوجد الوسط الحسابي للبيانات التالية :

19.1	22.5	19.9	21.9	22.8	22.0	23.0	23.2	20.5	4.6
21.4	21.1	20.9	20.8	19.8	22.2	22.6	21.7	19.4	21.3

الحل :

يمكن إيجاد النتائج باستخدام الآلة الحاسبة وبالتالي :

$$\bar{x} = \frac{427.7}{20} = 21.38$$

(d) أوجد المتوسط (الوسط الحسابي) للبيانات الواردة في جدول التوزيع التكراري التالي :

الفئات Classes	التكرار Frequency
5.0-12.9	3
9.0-12.9	10
13.0-16.9	14
17.0-20.9	25
21.0-24.9	17
25.0-28.9	9
29.0-32.9	2

الحل :

يمكن الاستعانة بجدول آخر يسهل عملية الحسابات مع العلم أنه يمكن إيجاد الناتج دون اللجوء إلى هذا الجدول .

x_i	f_i	$x_i f_i$
6.95	3	20.85
10.95	10	109.50
14.95	14	209.30
18.95	25	473.75
22.95	17	390.15
26.95	9	242.55
30.95	2	61.90
المجموع	80	1508

وبالتالي نحصل على المتوسط وهو

$$\bar{x} = \frac{1}{80} \sum_{i=1}^7 x_i f_i = \frac{1508}{80} = 18.85$$

مثال: فيما يلي توزيع 30 أسرة حسب الإنفاق الاستهلاكي الشهري لها بالآلاف الدنانير .

فئات الإنفاق	2 -	5 -	8 -	11 -	14 - 17
عدد الأسر	4	7	10	5	4

والمطلوب حساب منوال الإنفاق الشهري للأسرة، باستخدام طريقة الفروق .

$$d_1 = (10 - 7) = 3 \quad d_2 = (10 - 5) = 5$$

التكرارات	الفئات
4	2 -
7	5 -
10	8 -
5	11 -
4	14 - 17

فئة المتوال $A = 8$

$d_1 = 10 - 7 = 3$

أكبر تكرار

$d_2 = 10 - 5 = 5$

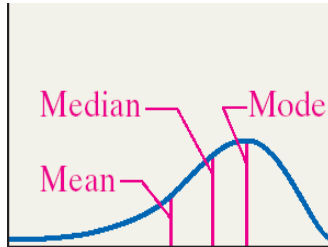
$$Mod = A + \frac{d_1}{d_1 + d_2} \times L$$

$$= 8 + \frac{3}{3+5} \times 3 = 8 + 1.125 = 9.125$$

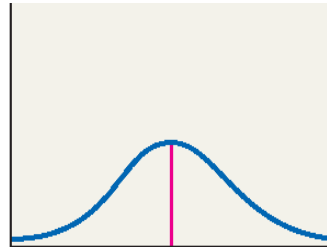
استخدام مقاييس النزعة المركزية في تحديد شكل توزيع البيانات

خلاصة: العلاقة بين المتوسطات الثلاث و شكل التوزيع

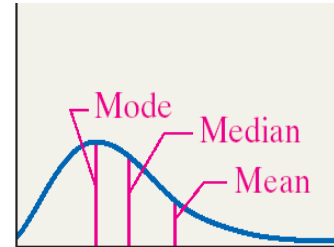
الشكل متناظر فان الوسط = الوسيط = المنوال
 الشكل ملتوي نحو اليسار فان الوسط أصغر من الوسيط و الوسيط أصغر من المنوال
 الشكل ملتوي نحو اليمين فان الوسط أكبر من الوسيط و الوسيط أكبر من المنوال



(a) Skewed Left
Mean < Median



Mean = Median = Mode
(b) Symmetric
Mean = Median



(c) Skewed Right
Mean > Median

التوزيع ملتوي نحو اليسار
الوسط أصغر من الوسيط
الوسيط أصغر من المنوال

التوزيع متناظر
الوسط = الوسيط = المنوال

التوزيع ملتوي نحو اليمين
الوسط أكبر من الوسيط
الوسيط أكبر من المنوال

العلاقة بين الوسط والوسيط والمنوال

$$\text{المنوال} = 3 * \text{الوسيط} - 2 * \text{الوسط}$$

مثال: إذا علمت أن قيمة الوسط = 5 وقيمة الوسيط = 10 احسب قيمة المنوال .
الحل:

$$\text{المنوال} = 3 * \text{الوسيط} - 2 * \text{الوسط}$$

$$\text{المنوال} = 3 * 10 - 2 * 5$$

$$\text{المنوال} = 30 - 10 = 20$$

تحديد التواء التوزيع مباشرة من مقاييس النزعة المركزية:

مثال: إذا علمت أن قيمة الوسط = 5 وقيمة الوسيط = 10 احسب قيمة المنوال ، ثم حدد نوع التواء التوزيع .
الحل:

$$\text{المنوال} = 3 * \text{الوسيط} - 2 * \text{الوسط}$$

$$\text{المنوال} = 3 * 10 - 2 * 5$$

$$\text{المنوال} = 30 - 10 = 20$$

نلاحظ أن

$$\text{الوسط} > \text{الوسيط} > \text{المنوال}$$

التوزيع ملتوي التواء سالب .

تمارين :

1- البيانات التالية تمثل أعمار 15 شخص من الذين التحقوا بدورة تمرير:

33, 24, 19, 39, 48, 45, 26, 35, 38, 23, 34, 29, 37, 25, 33

I - أوجد متوسط أعمار هؤلاء الأشخاص ، ما هو مدى هذه الأعمار

2 - البيانات التالية تمثل عدد الساعات التي قضاها عشرة أشخاص في مراجعة مادة الإحصاء استعدادا للامتحان ، 7 , 14 ,

22 , 20 , 19 , 25 , 11 , 20 , 13 , 24

a- أوجد متوسط عدد الساعات ثم أوجد الوسيط لهذه البيانات

3- يتألف صف من 35 طالب وطالبة كانت درجاتهم في اختبار الإحصاء مرتبة تصاعدياً كالتالي :

28	35	41	47	51	54	56
58	58	60	61	62	62	64
64	66	67	68	69	70	70
70	72	73	74	75	76	77
78	80	81	83	88	90	94

a- أوجد الوسيط لهذه الدرجات ومداها.

4 - أخذت أطوال 100 جندي لا قرب سم ووزعت هذه الأطوال في جدول بالشكل التالي :

الطول لأقرب سم	التكرار	التكرار التجميعي الصاعد
1.61m أو أقل	3	3
1.63m أو أقل	7	10
1.65m أو أقل	13	23
1.67m أو أقل	20	43
1.69m أو أقل	25	68
1.71m أو أقل	16	84
1.73m أو أقل	9	93
1.75m أو أقل	4	97
1.77m أو أقل	1	98
1.79m أو أقل	2	100

والمطلوب :

a- اشر إلى الفئة المنوالية وعين حداها الفعليين ثم أوجد المنوال لهذه البيانات .

b- اشر إلى الفئة الوسيطة (الفئة التي تحوي الوسيط) ثم احسب وسيط هذه البيانات.

c- كم جندي من هذه المجموعة أطول من 161cm .

5- الجدول التالي يحتوي على عدد الأهداف التي سجلت في الدوري الكروي في الموسم الماضي:

عدد الأهداف	0	1	2	3	4	5	6	7
التكرار	2	1	5	2	13	18	9	2

المطلوب :

a - أوجد التكرار التجميعي الصاعد و مثله بيانياً ثم استنتج الوسيط.

6-البيانات التالية تمثل الزمن بالدقيقة الذي يقضيه 13 طالباً بانتظار حافلة الجامعة

10 15 , 10 , 2 , 17 , 6 , 8 , 3 , 10 , 2 , 9 , 13 , 1

a- أوجد الوسط الحسابي والوسيط.

7- أحسب المتوسط للبيانات التالية إذا أمكن مع ذكر السبب

A-

التكرار	الفئات
6	30 - 39
8	40 - 49
14	50 - 59
7	60 - 69
2	70 - 79

B-

التكرار	الفئات
4	أقل من 60
10	60 - 69
12	70 - 74
7	أكثر من 79

C -

التكرار	الفئات
50	أقل أو يساوي 50
40	51-60
21	61-70
10	71-80

8 -لدينا البيانات التالية:

24 , 48, 27, 18, 39, 27,30, 31, 29, 12, 18

I – احسب الوسيط والوسط الحسابي

9 - بلغ متوسط الدخل الشهري لـ 140 من موظفي أحد البنوك الوطنية 1200 دولار بينما بلغ متوسط الدخل الشهري لـ 260 موظفاً آخر 450 دولار. ما هو متوسط الدخل الشهري لهؤلاء الـ 400 موظف .

10 – الجدول التالي يبين الأخطاء المطبعية على الصفحة الأولى لـ 150 إصدار من جريدة الحقيقة.

عدد الاخطاء	16-14	19-17	22-20	25-23	28-26	31-29	34-32
التكرار	5	10	49	39	24	17	6

والمطلوب :

- I - مثل هذه البيانات بيانياً بالطريقة التي تختارها مناسبة لاستنتاج المنوال بيانياً .
 II - أوجد التكرار التجميعي الصاعد ثم مثله بيانياً والوسيط .
 تنسيق البيانات :

إن أول خطوة تجاه فهم مسألة ما، هو جمع المعلومات الكافية عنها أولاً وبعد جمع هذه المعلومات والبيانات العددية نبدأ بفرزها وتنسيقها . وعلى سبيل المثال إذا سألنا 100 شخص من مالكي السيارات الخاصة عن أطوال المسافات التي يقطعها كل منهم يومياً سنجد إجابات هؤلاء الأشخاص معطاة (بالكيلومتر) كالتالي :

44.0	35.4	28.4	37.0	46.0	35.4	19.4	20.4	56.4	43.2
36.2	38.4	49.2	31.8	86.4	12.6	27.4	14.0	39.4	39.4
15.8	28.8	38.0	44.0	38.4	74.0	23.0	11.4	39.8	30.2
29.2	40.6	49.6	30.4	12.2	123.8	24.0	47.0	32.4	39.2
35.2	56.4	31.0	45.0	90.2	100.0	39.0	37.0	49.4	28.2
12.6	27.0	47.8	52.6	41.0	40.0	28.0	23.6	37.6	37.8
30.0	45.8	18.0	41.0	22.6	24.2	89.6	90.4	43.0	29.8
56.2	24.8	12.6	53.6	125.4	16.2	39.0	40.8	33.6	39.4
45.6	37.4	18.0	50.6	103.4	52.4	20.2	64.6	22.2	60.0
42.2	42.0	16.2	108.2	44.0	42.6	39.4	37.6	41.4	40.4

والسؤال الآن ماذا يمكن أن نفعل بهذه المعلومات وما هي الفائدة منها أو بالأحرى كيف يمكن الاستفادة منها ؟ البعض يهتم بالنظر إلى القيم المتطرفة والبعض الآخر يحسب المتوسط أو القيم المتكررة . وفي هذه المجموعة من المعلومات يمكن ملاحظة أن أقصر مسافة هي 11.4 km بينما أطول مسافة هي 125.4 km وهنا يمكن أن نفكر بإعادة كتابة هذه البيانات مرتبة ترتيباً تنازلياً أو تصاعدياً، وهذا ما يحتاج إلى وقت أطول وإضافة عملية حسابية تتناسب صعوبتها مع جمع البيانات المراد التعامل معها ، وحتى بعد ترتيب هذه البيانات تصاعدياً أو تنازلياً لا يمكن الحصول منها على الكثير من المعلومات .

مثال :

تمثل القراءات التالية الفترات الزمنية التي يقضيها 80 طالباً من طلاب كليات الجامعة في المكتبة أسبوعياً .
 شكل جدول توزيع تكراري لهذه البيانات .

24	15	22	28	17	12	20	16	23	16
22	15	23	18	11	21	17	30	16	29
19	39	19	18	14	20	28	18	29	24
20	22	34	12	29	25	17	23	20	18
24	14	32	27	18	21	15	19	17	16
21	19	23	26	24	13	23	15	25	22
25	16	18	23	20	10	22	15	18	16
15	26	17	20	21	19	20	21	27	31

الحل :

بما أن أصغر قراءة هي 10 وأكبر قراءة 39 فيمكن حساب المدى وهو الفرق بين أكبر قراءة وأصغر قراءة وهو
 $R = 39 - 10 = 29$

باختيار عدد الفئات 6 يمكن أن نحسب طول الفئة وهي هنا $5 \sim 4.833 = 29 \div 6$
 وفي مثل هذه الحالات يقرب الناتج للأعلى دوماً ، أي لأقرب عدد صحيح تالي. وبالتالي فإن حدود الفئات هي
 10-14 , 15-19 , 20-24 , 25-29 , 30-34 and 35-39

لاحظ أنه يمكن اختيار عدد آخر للفئات فإذا اخترنا العدد 8 مثلاً فإن أطوال الفئات سيكون
 $4 \sim 3.625 = 29 \div 8$ وسوف تكون الفئات الثمان كالتالي:

10-13 , 14-17 , 18-21 , 22-25 , 26-29 , 30-33 , 34-37 , 38-41

وهنا يمكن أن نذكر القارئ بأن الرقم 41 غير موجود في البيانات أصلاً ووجوده هنا لا يؤثر على البيانات الأصلية وإنما وضعناه لكي تكون أطوال الفئات متساوية كما ويمكن أن نبدأ بعدد أصغر من الرقم 10 الذي هو أصغر قراءة موجودة ولنفس السبب . لقد ذكرنا هنا إمكانييتين للتوضيح فقط وسنتابع شرح المثال على ست فئات والسبب في ذلك التفاصيل التي ذكرت سابقاً وهي أن أطوال الفئات يمكن أن تكون 5 , 10 , 100 ، ثم نصنف هذه البيانات وفق الفئات الست كما يأتي:

تكرار Frequency	إشارة tally	ساعة
8	IIII III	10-14
28	IIII IIIII IIIII III	15-19
27	IIII IIIII IIIII IIII II	20-24
12	IIII IIII II	25-29
4	IIII	30-34
1	I	35-39

لاحظ في الجدول أن الحدود الفعلية للفئة الأولى 9.5-14.5 وللجنة الثانية 14.5-19.5 وهكذا . ونحصل على هذه الحدود بطرح و جمع نصف واحدة دقة للحد الأدنى والحد الأعلى على الترتيب . والمقصود بنصف واحدة الدقة هو (0.5) اذا كانت البيانات صحيحة و(0.05) اذا كانت البيانات تحتوي على عدد عشري واحد بعد الفاصلة و(0.005) اذا كانت البيانات تحتوي على عددين عشريين بعد الفاصلة . وفي الجدول أعلاه يمثل العمود في أقصى اليمين التكرار (Frequency) عدد مرات وقوع القراءات في الفئات الملائمة وهو ما ندعوه تردد الفئة أو التكرار Class Frequency . إن أصغر رقم وأكبر رقم في عمود أقصى اليسار " الساعات hours " يعبر عنه بحدود الفئات Class limits فالأرقام

10	15	20	25	30	35
----	----	----	----	----	----

هي الحدود الدنيا للفئات بينما

14	19	24	29	34	39
----	----	----	----	----	----

فتسمى الحدود العليا للفئات Lower class limits and Upper class limits على الترتيب . في هذا المثال قربت فترات الدراسة لأقرب ساعة وهذا يعني أن الفئة الأولى تغطي الفترة من 9.5 إلى 14.5 والفترة الثانية من 14.5 إلى 19.5 ساعة وهكذا ... تدعى هذه الأرقام حدود الفئات الفعلية "Class boundaries" أو "Real class limits" ، إن حدود الفئات هي الأكثر استعمالاً أو شيوعاً . تعتبر مراكز الفئات Class marks إحدى عناصر التوزيعات التكرارية الهامة وهي بالتعريف منتصف طول الفئة ونحصل عليه بإضافة الحد الأدنى للفئة لحدها الأعلى وقسمة الناتج على 2 أي $(a+b)/2$ حيث a الحد الأدنى b الحد الأعلى . أما طول الفئة Class interval فهو الفرق بين الحد الأعلى الفعلي و الحد الأدنى الفعلي وغالباً ما يسمى مدى الفئة . نحصل على مراكز الفئات بحساب مركز الفئة الأولى ومن ثم نضيف له طول الفئة فنحصل على مركز الفئة الثانية ، نضيف طول الفئة للناتج فنحصل على مركز الفئة التالي... وهكذا . وبالعكس فإن الفرق بين مركزي فئتين متتاليتين يعطي طول الفئة .

مثال :

أوجد مراكز الفئات وأطوال الفئات في المثال السابق .
 إن مراكز الفئات Class mark هي على الترتيب $(10+14)/2 = 12$ ثم $(15+19)/2 = 17$... وهكذا وأخيراً $(35+39)/2 = 37$. ولحساب طول الفئة لاحظ هنا أن $17-12=5$.

قد نكون بحاجة إلى معيار آخر ويدعى بتكرار التوزيع النسبي المؤي Percentage-distribution ويحسب بقسمة تكرار كل فئة على المجموع العام وضرب الناتج بـ 100. أو معياراً آخر ويدعى التكرار النسبي والذي يحسب بقسمة التكرار لكل فئة على المجموع العام (Relative frequency).

مثال :

احسب تكرار التوزيع النسبي percentage distribution للفترة الزمنية التي قضاها 80 طالباً في المكتبة خلال أسبوع ما والمفرغة في الجدول ٣-٢ .

الحل :

لقد شاهدنا في الجدول 2-3 التردد لكل فئة وبالتالي يمكن قسمة تردد كل فئة على المجموع الكلي وضرب الناتج بـ 100 أي أن الفئة الأولى تحوي $10\% = \frac{8}{80} \times 100$ من البيانات وبالنسبة للفئة الثانية فتحتوي $35\% = \frac{28}{80} \times 100$ من البيانات و.....

هكذا، وأخيراً الفئة الأخيرة تحوي $1.25\% = \frac{1}{80} \times 100$ من البيانات . وهكذا يمكن تلخيص ذلك بجدول على غرار الجدول السابق يحتوي نتائج المثال الحالي والسابق بالشكل الآتي :

مراكز الفئات Class marks	التكرار Frequency	التكرار Percentage	الفئات " ساعة " Class limits
$(10+14) \div 2 = 12$	8	$(8 \div 80) \times 100 = 10\%$	10-14
$(15+19) \div 2 = 17$	28	$(28 \div 80) \times 100 = 35\%$	15-19
22	27	33.75 %	20-24
27	12	15 %	25-29
32	4	5 %	30-34
37	1	1.25 %	35-39
	80	100%	المجموع

هذا ويمكن تعديل جدول التوزيع وتحويله إلى جدول من الشكل " أقل من أو يساوي " أو " أقل من " أو " أكثر من " وهذا ما يدعى بـ " التوزيعات التكرارية المتجمعة "

Cumulative frequency distribution . ونحصل على ذلك بجمع تكرار الفئات بداية من الفئة الأولى إلى الثانية ، الناتج إلى الثالثة الناتج إلى الرابعة ... وهكذا .

مثال :

شكل جدول توزيع متجمع " أقل من " (Less than) للبيانات الواردة في المثال الأول أعلاه والمفرغة في الجدول أعلاه

الحل :

لا توجد قيم أقل من 10 ، 8 قيم أقل من 15 ، $8+28=36$ قيمة أقل من 20 وهكذا . فيمكن بالتالي ترتيب النتائج في جدول تجميعي " صاعد " كما يلي :

الفئات	التكرار Frequency	التكرار التجميعي الصاعد cumulative Frequency	التكرار التجميعي الهابط Reverse Cumulative Freq.
أقل (أكبر أو يساوي) من 10	0	0	80
أقل (أكبر أو يساوي) من 15	8	0+8=8	80-8=72
أقل (أكبر أو يساوي) من 20	28	8+28=36	72-28=44
أقل (أكبر أو يساوي) من 25	27	36+27=63	44-27=17
أقل (أكبر أو يساوي) من 30	12	63+12=75	17-12=5
أقل (أكبر أو يساوي) من 35	4	75+4=79	5-4=1
أقل (أكبر أو يساوي) من 40	1	79+1=80	1-1=0

هناك جداول تكرارية معاكسة للجدول التكراري الصاعد وهي جداول تكرارية تناقصية و مشابهة للجدول أعلاه ولكنها تتناقص من المجموع. إن مجموع القيم السابقة هو 80 فتكون الفئة الثانية $80-8=72$ والفئة الثالثة $72-28=44$ وهكذا ويكون التكرار المتناقص للفئة الأخيرة 0. لاحظ في العمود الأخير هناك 80 قراءة أكبر من 10 لكن 0 قراءة أقل من 10 ، 8 قراءات أقل من 15 ولكن 72 قراءة أكبر من 15. وهكذا .. وهذا ينطبق على بقية الفئات. وقد وضعنا ذلك في الجدول أعلاه في العمود الأول إلى اليمين. إن التمثيل البياني للتوزيع الصاعد والهابط على محور واحد سيلعب دوراً هاماً في دراسة الإحصاء والذي سنتعرض له لاحقاً.

أمثلة :

١- البيانات التالية تمثل أعمار 40 شخصاً من موظفي إحدى شركات النقل البحري

49	22	35	21	35	52	31	21	41	51
35	20	35	60	40	39	35	31	30	45
44	48	30	32	26	27	28	27	30	19
42	30	31	41	42	41	50	60	59	45

وزع هذه البيانات في سبع فئات واستنتج التكرار التجميعي الصاعد

الحل :

لنحسب أولاً المدى Range . والمدى هو الفرق بين أكبر وأصغر قيمة وهو بالتالي $R = 60 - 19 = 41$ أما طول الفئة الواحدة فهو $41 \div 7 = 5.857 \sim 6$ حيث قربنا الناتج للعدد الصحيح التالي . بعد تفريغ البيانات نحصل على جدول التوزيع التكراري التالي : ١

الفئة	حدود الفئات	العلامات	التكرار	التكرار النسبي	مراكز الفئات	التكرار الصاعد
1	19-14	III	5	0.125	21.5	5
2	25-30	III III	8	0.200	27.5	13
3	31-36	III III	9	0.225	33.5	22
4	37-42	III II	7	0.175	39.5	29
5	43-48	III	4	0.100	45.5	33
6	49-54	III	4	0.100	51.5	37
7	55-60	III	3	0.075	57.5	40
	المجموع		40	1.000		

٢- تمثل البيانات التالية مبيعات إحدى شركات السيارات مقدرة بآلاف الدنانير ومقربة لأقرب دينار. وذلك خلال 100 يوم عمل

25	33	27	29	24	31	30	29	37	31
32	23	29	32	22	34	37	26	31	28
28	31	33	21	32	30	28	29	25	30
26	29	29	24	29	32	32	32	32	30
30	33	27	30	26	28	28	27	33	33
21	31	28	31	34	31	31	27	28	30
36	29	32	28	26	28	30	34	33	31
27	33	23	32	30	33	33	29	30	30
32	28	30	31	26	27	30	31	39	32
32	31	25	31	30	25	28	30	33	28

وزع هذه البيانات في جدول توزيع تكراري يحتوي على سبع فئات ثم أوجد:

- 1- التكرار الصاعد ثم مثله بيانياً
- 2- مثل الناتج بيانياً بطريقة الأعمدة البيانية Histogram .
- 3- مثل الناتج بيانياً بطريقة الدائرة Pie chart .
- 4- أوجد المضلع التكراري للبيانات السابقة Frequency polygon

الحل :

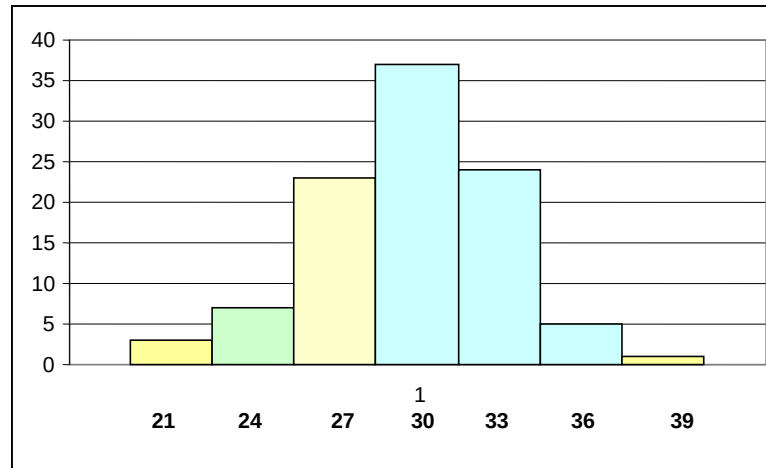
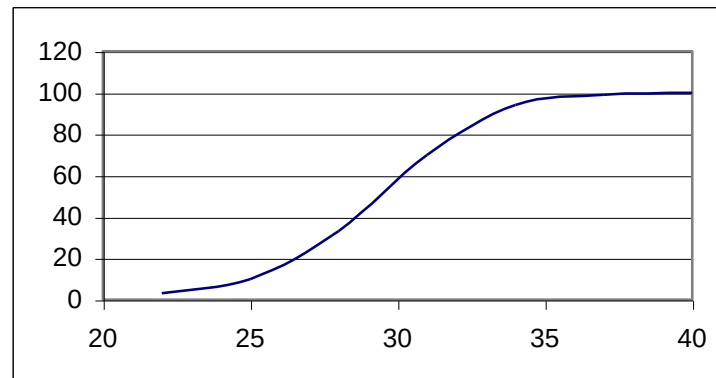
لابد أولاً من تفريغ البيانات أعلاه ولهذه الغاية نحسب مدى هذه البيانات وهو

$$R = 39 - 21 = 18$$

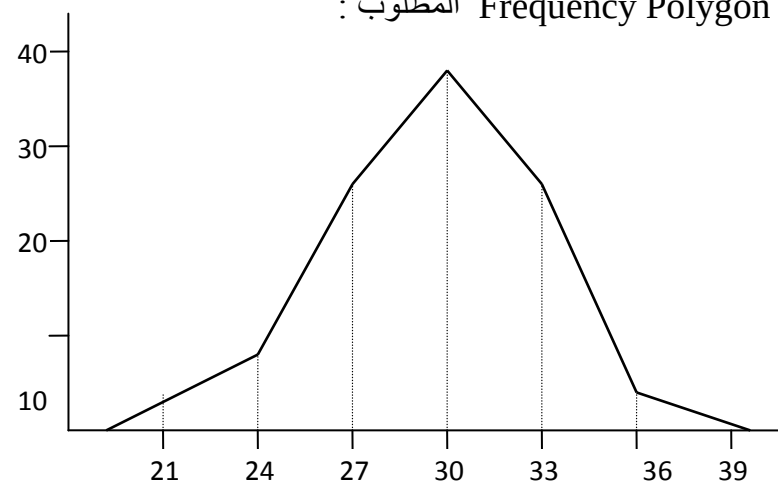
أما طول الفئة الواحدة فهو $3 \sim 2.57 = 18 \div 7$ حيث قسمنا على العدد 7 لأننا نريد تقسيم البيانات إلى سبع فئات ونقربه إلى العدد الصحيح التالي فيكون طول الفئة الواحدة 3 ويصبح الجدول المطلوب بعد تفريغ البيانات كما هو مبين أدناه.

التكرار المتجمع الصاعد	مراكز الفئات	التكرار النسبي	التكرار	حدود الفئة	الفئة
3	21	0.03	3	20-22	1
10	24	0.07	7	23-25	2
33	27	0.23	23	26-28	3
70	30	0.37	37	29-31	4
94	33	0.24	24	32-34	5
99	36	0.05	5	35-37	6
100	39	0.01	1	38-40	7
		1	100	المجموع	

وكنتيجة مباشرة من الجدول أعلاه يمكن ملاحظة منحنى التكرار الصاعد والأعمدة البيانية وهي موضحة أدناه في الاشكال أدناه



أخيراً ولأن المضلع التكراري هو خط منكسر يصل بين النقاط التي فواصلها مراكز الفئات و التكرار لهذه الفئات ومغلقة من الطرفين بحيث يتصل مع محور الفواصل ذلك لأن التكرار معدوماً مثل النقطة الأولى وما بعد النقطة الأخيرة فإن الشكل ادناه يوضح المضلع التكراري Frequency Polygon المطلوب :

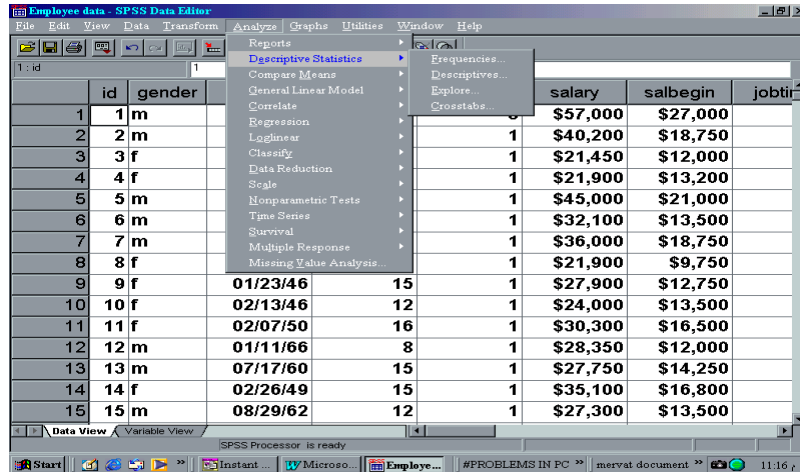


المقاييس الإحصائية من برنامج SPSSWIN

ولتنفيذ الإحصاءات الوصفية في برنامج (SPSSWIN) تجرى الخطوات التالية:

أولاً: النقر بالفأرة على الأوامر التالية من صف القوائم المنسدلة:

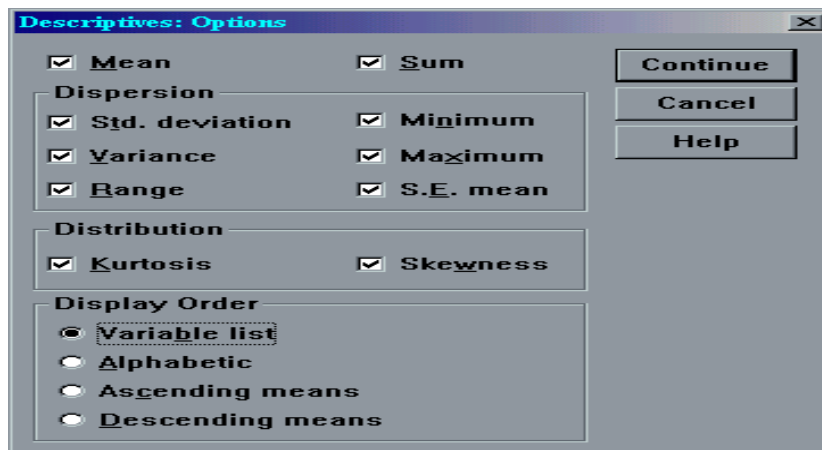
Analyze → Descriptive statistics → Descriptive



يظهر صندوق حوار الذي عنوان نافذته Descriptive ، ندخل كل المتغيرات المطلوب إيجاد المقاييس الإحصائية لها ، ويوضح إدخال متغير مستوى التعليم Educational level ، والمرتب الحالي Current salary ، الموجودين في ملف Employee data.

صندوق حوار المتغيرات المطلوب إيجاد المقاييس الإحصائية لها

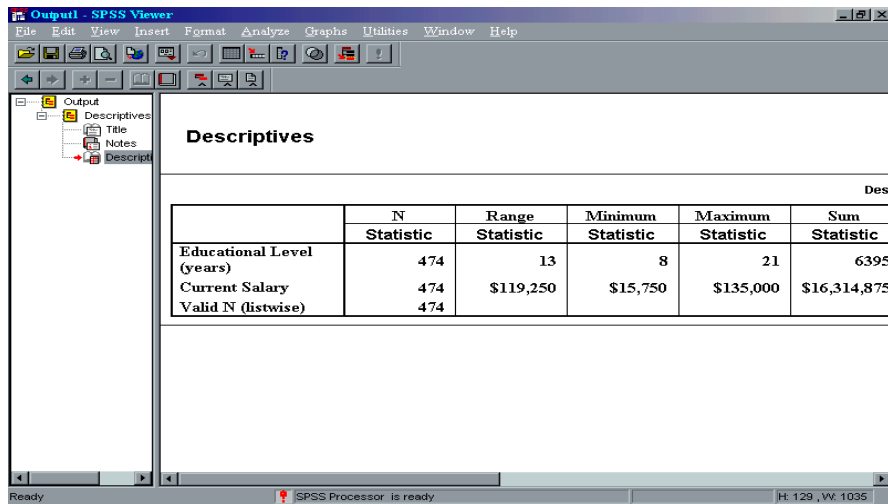
وبالضغط علي مفتاح الاختيارات Options يظهر صندوق حوار به مقاييس إحصائية نختار منها ما نريده من مقاييس وصفية.



المقاييس الإحصائية هي:

☐ Mean	الوسط الحسابي	☐ Sum	المجموع
Dispersion	التشتت		
☐ Std. Deviation	الانحراف المعياري	☐ Minimum	الحد الأدنى
☐ Variance	التباين	☐ Maximum	الحد الأعلى
☐ Range	المدى	☐ S.E. mean	الخطأ المعياري للوسط الحسابي
Distribution	التوزيع		
☐ Kurtosis	التفرطح	☐ Skewness	الالتواء

ننقر بالفأرة أمام المقاييس الإحصائية المطلوبة فتظهر علامة × في الصندوق ☐ الذي بجانبه ثم بالضغط على زر الاستمرار (Continue) يقل صندوق الحوار. أخيراً يضغط على زر (Ok) في صندوق الحوار (Descriptive) فتظهر النتائج في نافذة المخرجات (Output).



The screenshot shows the SPSS Output window with the 'Descriptives' table. The table lists statistics for 'Educational Level (years)' and 'Current Salary'. The 'Valid N (listwise)' is 474 for both.

	N Statistic	Range Statistic	Minimum Statistic	Maximum Statistic	Sum Statistic
Educational Level (years)	474	13	8	21	6395
Current Salary	474	\$119,250	\$15,750	\$135,000	\$16,314,875
Valid N (listwise)	474				

الاسبوع 10-12

مقاييس التشتت Measures of Dispersion

تناولنا في محاضرات سابقة المتوسطات أو مقاييس النزعة المركزية والتي تعبر عن المستوى العام للظاهرة التربوية أو السياسية أو الاجتماعية محل البحث. ولكن ترى هل هذا يعتبر كافياً لوصف البيانات وتحليلها كميًا لنصل بالتالي إلى فهم أكثر وضوحاً للظاهرة محل الدراسة؟ للإجابة على هذا السؤال نسوق المثال التالي:

مثال ١:

نفترض أن لدينا شريحتين من شرائح المجتمع تعيشان في منطقتين مختلفتين وكانت دخولهم الأسبوعية (بالدولار الأمريكي) هي كما يلي:

المجموعة الأولى A:

70	75	71	75	74	76	73	78
----	----	----	----	----	----	----	----

المجموعة الأولى B:

99	56	80	100	29	70	65	93
----	----	----	-----	----	----	----	----

وبحساب الوسط الحسابي للشريحتين المذكورتين في كل من المجموعتين - حسب ما أوضحناه في الفصل السابق - فإن الوسط الحسابي لدخول المجموعة الأولى:

$$\bar{X}_A = \frac{70 + 75 + 71 + 75 + 74 + 76 + 73 + 78}{8}$$

$$\bar{X}_A = \frac{592}{8} = 74 \text{ دولاراً}$$

أي أن الوسط الحسابي لدخل المجموعة A هو 74 دولاراً. وكذلك بالنسبة للمجموعة B فإن الوسط الحسابي لدخولها هو:

$$\bar{X}_B = \frac{99 + 56 + 80 + 100 + 29 + 70 + 65 + 93}{8}$$

$$\bar{X}_B = \frac{592}{8} = 74 \text{ دولاراً}$$

والوسط الحسابي لدخل المجموعة B هو أيضاً 74 دولاراً. ومعنى هذا أن المستوى العام لدخل الأفراد في المجموعتين واحد ويساوي 74 دولاراً. ولكن بإمعان النظر في دخول المجموعة الأولى نجد أنها متجانسة إلى حد كبير أي أنها قريبة جداً من بعضها أو من الوسط الحسابي والذي يساوي 74 دولاراً. وهنا نقول أن تشتت الدخول قليل أو صغير. بينما دخول المجموعة الثانية غير متجانسة فهي بعيدة عن بعضها، أو عن الوسط الحسابي بشكل كبير. وهنا نقول أن تشتت دخول المجموعة الثانية كبير فهي أقل تجانساً (أو أكثر تشتتاً) من المجموعة الأولى.

(الفرق بين أكبر دخل وأصغر دخل في المجموعة الأولى: $78 - 70 = 8$ دولارات فقط، بينما الفرق بين أكبر دخل وأصغر دخل في المجموعة الثانية: $100 - 29 = 71$ دولاراً)

نستنتج من هذا المثال بأن المتوسط ليس كافياً لتوصيف البيانات أو تحليلها كميًا. فها هو المتوسط واحد في المجموعتين ورغم ذلك فإن البيانات تختلف تماماً في مدى تشتتها (أو تجانسها). وهنا تبرز الحاجة إلى مقاييس كمية أو إحصائية ليقاس مدى تشتت البيانات ولتعطي للباحث بالتالي صورة أكثر وضوحاً وصدقاً

للمظاهرة السياسية محل الدراسة. وفيما يلي نتناول بعض مقاييس التشتت التي تخدم الغرض من هذا الكتاب وهي :

المدى، والتباين والانحراف المعياري، ومعامل اختلاف

المدى The Range

يعرّف المدى بأنه " الفرق بين أكبر قيمة وأصغر قيمة ".

فمدى المجموعة A في المثال رقم (1) هو : $78 - 70 = 8$

ومدى المجموعة B في نفس المثال هو : $100 - 29 = 71$

وواضح تماماً أن مدى المجموعة الثانية أكبر بكثير جداً من مدى المجموعة الأولى مما يعني أن المجموعة الثانية أكثر تشتتاً من المجموعة الأولى.

ومن تعريف المدى يتضح أنه مقياس بسيط جداً وأن كل المطلوب معرفته هو أكبر قيمة وأصغر قيمة فقط. ومن هذا التعريف نخرج بالملاحظات التالية :

- ١ - أن المدى يعتمد في حسابه على قيمتين فقط (أكبر قيمة وأصغر قيمة) وأنه يهمل بالتالي باقي القيم.
- ٢ - أنه لا يقيس تشتت البيانات عن متوسطها، فهو لا يشير - من قريب أو من بعيد - إلى متوسط البيانات (أو مركزها).

٣ - أنه حساس جداً لأي قيمة شاذة أو متطرفة والمثال التالي يوضح هذه النقطة بشيء من التفصيل.
مثال (٢) : إذا كانت أعمار أعضاء السلطة التشريعية في بلد ما هي :

70 75 73 74 20 78 72

فإن المدى في هذه الحالة هو : $78 - 20 = 58$

وهنا نلاحظ وجود قيمة شاذة بالنسبة لباقي القيم وهي 20 وهي أصغر قيمة. وإذا أهملت هذه القيمة (أو لم تكن موجودة أصلاً) لكان المدى :

$$78 - 70 = 8$$

وهذا يعني أن وجود قيمة شاذة (20) رفعت قيمة المدى من 8 سنوات إلى 58 سنة. وهذا يوضح مدى حساسية هذا المقياس للقيم الشاذة (أو المتطرفة).

ولكل هذه الأسباب - أو الملاحظات - فإن كثيراً من الإحصائيين والباحثين لا يعتمدون كثيراً على المدى كمقياس للتشتت. ويستخدم فقط إذا كان المطلوب فكرة سريعة أو عامة (وليست دقيقة) عن مدى تشتت البيانات.

مثال: باستخدام المدى قارن بين تشتت المجموعات التالية

إذا كان لدينا المجموعات التالية من البيانات :

المجموعة الأولى : 6 6 6 6 6

المجموعة الثانية : 8 7 6 5 4

المجموعة الثالثة : 13 10 6 1 0

الحل :

$$\text{مدى المجموعة الأولى} = 6 - 6 = 0$$

$$\text{مدى المجموعة الثانية} = 8 - 4 = 4$$

$$\text{مدى المجموعة الثالثة} = 13 - 0 = 13$$

بما أن مدى المجموعة الأولى يساوى صفر، فيعنى ذلك أنه لا يوجد اختلاف أو تشتت بين قيم هذه المجموعة، أي أن كل القيم داخل هذه المجموعة متساوية. ونلاحظ أن مدى المجموعة الثانية أصغر من مدى المجموعة الثالثة، وهذا يعنى أن تشتت القيم داخل المجموعة الثانية أقل من تشتت القيم داخل المجموعة الثالثة.

مثال:

قارن بين تشتت درجات مادة الإحصاء لثلاث مجموعات من الطلبة:

درجات المجموعة الأولى : 80 , 45 , 70 , 30 , 62 , 50

درجات المجموعة الثانية : 35 , 77 , 82 , 65 , 55 , 70 , 0 , 20

درجات المجموعة الثالثة : 75 , 65 , 63 , 60 , 50 , 85 , 49 , 54

الحل :

$$\text{مدى المجموعة الأولى} = 80 - 30 = 50 \text{ درجة .}$$

$$\text{مدى المجموعة الثانية} = 82 - 0 = 82 \text{ درجة .}$$

$$\text{مدى المجموعة الثالثة} = 85 - 49 = 36 \text{ درجة .}$$

نلاحظ أن المجموعة الثالثة أقلهن تشتتاً ثم يليها المجموعة الأولى ثم يليها المجموعة الثانية .

مثال:

احسب المدى للبيانات التالية التي توضح أوزان 100 طالب بالكلية جرامات :

الوزن (بالكيلو جرام)	عدد الطلبة
60 إلى أقل من 63	5
63 إلى أقل من 66	15
66 إلى أقل من 69	40
69 إلى أقل من 72	28
72 إلى أقل من 75	12

الحل :

أكبر قيمة = الحد الأعلى للفئة الأخيرة = 75 كيلو جرام .

أصغر قيمة = الحد الأدنى للفئة الأولى = 60 كيلو جرام .

المدى هو : $R = 75 - 60 = 15$ كيلو جرام .

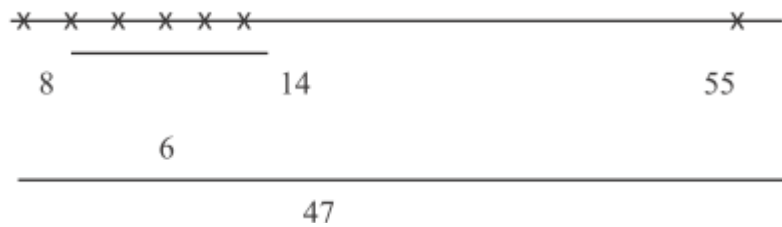
خواص المدى :

1. المدى مقياس سهل في حسابه وبسيط في مفهومه ودلالته .
2. اهتم بقيمتين فقط في البيانات ويهمل بقية القيم .
3. يعتمد في حسابه على القيمة الكبرى والقيمة الصغرى فقط ولذلك يعتبر مقياساً مضللاً . لأنه عندما تكون القيمة الكبرى أو القيمة الصغرى أو كليهما قيم شاذة ففي هذه الحالة يكون المدى كبيراً بينما قيم المجموعة تكون غير متباعدة . فمثلاً إذا كان لدينا البيانات التالية :

55 , 14 , 11 , 9 , 8 , 12 , 10

فنجد أن : المدى هو $R = 55 - 8 = 47$

فالمدى كبير، ويشير إلى وجود تشتت كبير في المجموعة في حين أن القيم متقاربة . لذلك فهو مقياس مضلل . وسبب ذلك هو اعتماده على القيم المتطرفة . فلو حذفنا القيمة المتطرفة وهي 55 فنجد أن قيمة المدى تساوى $14 - 8 = 6$ وهي قيمة صغيرة وواقعية . وشكل يوضح ذلك .



4. لا نستطيع حساب المدى في حالة التوزيعات التكرارية المفتوحة وذلك لأنه في هذه الحالة يكون الحد الأدنى للفئة الأولى أو الحد الأعلى للفئة الأخيرة أو كلاهما مجهولاً .

مثال :

احسب المدى للبيانات التالية :

95 – 200 – 250 – 300 – 110 – 90 – 150 – 100 – 350 – 80

الحل :

نرتب القيم أولاً : (80-90-95-100-110-150-200-250-300-350)

المدى = $350 - 80 = 270$

حساب المدى من البيانات المبوبة

المدى = الحد الأعلى للفئة الأخيرة – الحد الأدنى للفئة الأولى

مثال:

احسب المدى للجدول التالى :

الفئات	16-	20-	24-	28-	32-36
عدد المبحوثين	10	15	40	20	15

الحل :

المدى = الحد الأعلى للفئة الأخيرة – الحد الأدنى للفئة الأولى

المدى = $36 - 16 = 20$

التباين The Variance :

يعتبر التباين أحد مقاييس التشتت المهمة لأنه من ناحية يأخذ جميع القيم في الاعتبار عند حسابه، ومن ناحية أخرى لأنه يقيس التشتت عن الوسط الحسابي للقيم، هذا بالإضافة إلى أنه تسهل معالجته رياضياً، وأنه يدخل في تكوين عدد من المقاييس والاختبارات الإحصائية المهمة.

والفكرة الأساسية للتباين هي حساب انحرافات جميع القيم عن وسطها الحسابي (أي حساب الفرق بين كل قيمة والوسط الحسابي)، وسوف نجد أن بعض القيم أكبر من الوسط فتكون الفروق (أو الانحرافات) بالموجب، والبعض الآخر أصغر من الوسط فتكون الفروق (أو الانحرافات) بالسالب. ودائماً يكون مجموع هذه الانحرافات مساوياً للصفر. ويكون الحل هنا إما إهمال الإشارات السالبة أو تربيع هذه الانحرافات. وإهمال الإشارات السالبة ليس له مبرر رياضي، فيكون الحل هو تربيع هذه الانحرافات، ثم نحسب متوسط الانحرافات المربعة فنحصل على التباين. أي أن التباين يعرف كما يلي:

التباين : هو متوسط مربعات انحرافات القيم عن وسطها الحسابي. والمثال التالي يوضح كيفية حساب التباين.

مثال (٣) : أحسب تباين دخول الشريحة الاجتماعية التالية : مثال رقم (1) :

70 75 71 75 74 76 76 78

الحل :

١ - نحسب أولاً الوسط الحسابي للدخول كما يلي :

$$\bar{X} = \frac{592}{8} = 74$$

٢ - نحسب انحرافات القيم عن الوسط الحسابي (أي الفروق بينها وبين الوسط الحسابي) كما يلي :

- 4 , + 1 , - 3 , + 1 , 0 , + 2 , - 1 , + 4

(لاحظ أن مجموع الانحرافات يساوي صفر)

٣ - نربع هذه الانحرافات كما يلي :

16 1 9 1 0 4 1 16

٤ - التباين يساوي متوسط هذه المربعات، أي يساوي :

$$\frac{16+1+9+1+0+4+1+16}{8} = \frac{48}{8} = 6$$

الانحراف المعياري : Standard Deviation :

نلاحظ أن تمييز التباين سيكون " دخول مربعة " وبصفة عامة " وحدات مربعة " لأنه يتم تربيع الانحرافات (أو الفروق). لذلك فإنه يؤخذ الجذر التربيعي الموجب للتباين فنحصل على ما يسمى " الانحراف المعياري " والذي سيكون تمييزه هو نفس الوحدات. ففي هذا المثال يكون الانحراف المعياري للدخول هو :

$$\sqrt{6} = 2.449$$

دولاراً

والانحراف المعياري - عادة - يستخدم بدلاً من التباين كأهم مقياس للتشتت. فالتباين ما هو إلا مربع الانحراف المعياري، والانحراف المعياري ما هو إلا الجذر التربيعي الموجب للتباين.

تعريف الانحراف المعياري : الانحراف المعياري هو. الجذر التربيعي الموجب للوسط الحسابي لمربعات انحرافات القيم عن وسطها الحسابي. أي هو الجذر التربيعي الموجب للتباين. وعادة يرمز للانحراف المعياري للعينة بالرمز S ولتباين العينة بالرمز S² وللانحراف المعياري للمجتمع بالرمز اللاتيني σ (سيجما) ولتباين المجتمع بالرمز σ².

ويمكن التعبير عن كل من التباين والانحراف المعياري بالرموز كما يلي :
إذا فرضنا أن قيم العينة هي :

$X_1, X_2, \dots, X_n$ (أي عددها n) فإن الخطوات تكون كما يلي :
١ - حساب الوسط الحسابي أي حساب $\bar{X}$ حيث :

$$\bar{X} = \frac{\sum X}{n}$$

٢ - حساب إنحرافات القيم عن الوسط الحسابي، أي :

$$X_1 - \bar{X}, X_2 - \bar{X}, \dots, X_n - \bar{X}$$

٣ - تربيع هذه الانحرافات، أي :

$$(X_1 - \bar{X})^2, (X_2 - \bar{X})^2, \dots, (X_n - \bar{X})^2$$

٤ - التباين هو الوسط الحسابي لهذه المربعات، أي :

$$S^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n}$$

والتي يمكن كتابته (باختصار البسط) كما يلي :
تباين العينة

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

٥ - الانحراف المعياري هو الجذر التربيعي الموجب للتباين، أي :
الانحراف المعياري للعينة

$$S = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}}$$

وبالطريقة نفسها إذا كانت لدينا قيم

المجتمع الإحصائي. فإذا كانت قيم

المجتمع هي : $X_1, X_2, \dots, X_N$ أي عددها N فإن قانوني التباين والانحراف المعياري للمجتمع -
بافتراض أن الوسط الحسابي للمجتمع هو μ (ميو) - يمكن كتابتها كما يلي:
تباين المجتمع

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

الانحراف المعياري للمجتمع

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

ملاحظة مهمة :

يمكن حساب كل من التباين والانحراف المعياري بطريقة مختصرة - وتعطي النتائج نفسها بطبيعة الحال -
كما يلي :

* تباين العينة بالطريقة المختصرة :

$$S^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n} \right)^2$$

* الانحراف المعياري للعينة بالطريقة المختصرة :

$$S = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n} \right)^2}$$

وفي الطريقة المختصرة فإن كل المطلوب معرفته لحساب التباين أو الانحراف المعياري هو $\sum x$ (أي مجموع القيم)، $\sum x^2$ (أي مجموع مربعات القيم) ثم التعويض في القانون.

مثال (٤) :

أحسب التباين والانحراف المعياري بالطريقة المختصرة لبيانات المثال رقم (٣).

الحل :

يمكن تنظيم الحل في الجدول التالي :

X	X^2
70	4900
75	5625
71	5041
75	5625
74	5476
76	5776
73	5329
78	6084
$\sum x = 592$	$\sum x^2 = 43856$

ويكون التباين (حيث $n = 8$) هو :

$$\begin{aligned} S^2 &= \frac{\sum x^2}{n} - \left(\frac{\sum x}{n} \right)^2 \\ &= \frac{43856}{8} - \left(\frac{592}{8} \right)^2 \\ &= 5482 - 5476 \\ S^2 &= 6 \end{aligned}$$

وبالتالي فإن الانحراف المعياري هو : $S = \sqrt{6} = 2.449$ وهي النتائج نفسها التي حصلنا عليها سابقاً ولكن بطريقة أبسط.

بعض خصائص الانحراف المعياري :

١ - قيمة الانحراف المعياري دائماً موجبة أو أكبر من أو تساوي صفر. فأقل قيمة تساوي الصفر (وذلك عندما تكون جميع القيم متساوية، وفي هذه الحالة لا توجد فروق أو إنحرافات بينها وبين الوسط الحسابي وبالتالي لا يوجد أي تشتت بين القيم، وبالتالي فإن قيمة الانحراف المعياري في حالة تساوي جميع القيم تساوي الصفر).

٢ - كلما كان التشتت كبيراً حول الوسط كلما كان الانحراف المعياري كبيراً، والعكس صحيح.

٣ - إذا أضفنا وطرحنا مقدراً ثابتاً من كل القيم فإن قيمة الانحراف المعياري (أو التباين) لا تتغير (أي لا تتأثر قيمة الانحراف المعياري بالطرح أو الجمع). ولتوضيح هذه الخاصية نأخذ المثال التالي :

مثال (٥) : حل المثال السابق رقم ٤ بعد طرح 70 (على سبيل المثال من كل القيم لاحظ أن هذا سوف يسهل الحسابات).

الحل :

المربعات (بعد الطرح)	القيمة بعد طرح (70)
X^2	X
0	0
25	5
1	1
25	5
16	4
36	6
9	3
64	8
$\Sigma X^2 = 176$	$\Sigma X = 176$

ويكون التباين

$$S^2 = \frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n} \right)^2$$

$$= \frac{176}{8} - \left(\frac{32}{8} \right)^2$$

$$= 22 - 16$$

$$= 6$$

والانحراف المعياري $S = \sqrt{6} = 2.449$ وهي النتائج نفسها مع ملاحظة أن العمليات الحسابية أسهل في هذه الحالة. وتجدر الإشارة إلى أنه لو طرحنا أي قيمة أخرى سنحصل على النتائج نفسها.

٤ - إذا ضربنا كل قيمة في مقدار ثابت ثم حسبنا الانحراف المعياري للقيم الجديدة فإنه يجب القسمة على هذا المقدار الثابت. وإذا قسمنا كل قيمة على مقدار ثابت ثم حسبنا الانحراف المعياري للقيم الجديدة فإنه يجب الضرب في هذا المقدار الثابت.

معامل الاختلاف : The Coefficient of Variation

لمقارنة تشتت مجموعتين (أو أكثر) من البيانات وكانت البيانات تختلف في مستواها العام (أي في أوساطها الحسابية) و / أو تختلف في وحدات القياس (مثلاً مقارنة بيانات الدخل حيث تقاس بالريال ببيانات العمر حيث تقاس بالسنوات) فإن المقارنة لا تتم مباشرة بمقارنة الانحراف المعياري لكل منهما بل تتم من خلال مقياس آخر هو "معامل الاختلاف" أو ما يسمى أحياناً بمقياس التشتت النسبي حيث ينسب الانحراف المعياري لكل مجموعة إلى وسطها الحسابي والضرب في 100 فنحصل على مقياس نسبي أو مؤوي (وبدون تمييز) أي تتم المقارنة بحساب معامل الاختلاف لكل منهما، والمجموعة التي لها معامل اختلاف أكبر تكون أكبر تشتتاً والعكس صحيح أي أن :

$$\text{معامل الاختلاف} = \frac{\text{الانحراف المعياري}}{\text{الوسط الحسابي}} \times 100$$

وإذا رمزنا لمعامل الاختلاف بالرمز C.V فإن :

$$C.V = \frac{S}{\bar{X}} \times 100$$

مثال (٦) إذا كان الوسط الحسابي والانحراف المعياري لدخول عينة من العمال بالدولار هو :
 $X_1 = 1500, S_1 = 152$
 وكان الوسط الحسابي والانحراف المعياري لأعمارهم (بالسنوات) هو :
 $X_2 = 42, S_2 = 9.2$

فأيهما أكثر تشتتاً الدخل أم العمر ؟

الحل :

لمقارنة التشتت نحسب معامل الاختلاف لكل من الدخل والعمر كما يلي :

١- معامل اختلاف الدخل :

$$c.v_1 = \frac{S_1}{\bar{X}_1} \times 100 = \frac{152}{1500} \times 100 = 10.13\%$$

٢- معامل اختلاف العمر :

$$c.v_2 = \frac{S_2}{\bar{X}_2} \times 100 = \frac{9.2}{42} \times 100 = 21.9\%$$

٣- بما أن معامل اختلاف العمر أكبر من معامل اختلاف الدخل فإن بيانات العمر تكون أكثر تشتتاً من بيانات الدخل.

حساب التباين والانحراف المعياري للبيانات المبوبة :

إذا كانت البيانات مبوبة على شكل قيم وتكرارات، فإن التباين والانحراف المعياري بالطريقة المختصرة يمكن حسابهما كما يلي :

التباين في حالة البيانات المبوبة :

$$s^2 = \frac{\sum x^2 f}{\sum f} - \left(\frac{\sum xf}{\sum f} \right)^2$$

الانحراف المعياري للبيانات المبوبة

$$s = \sqrt{\frac{\sum x^2 f}{\sum f} - \left(\frac{\sum xf}{\sum f} \right)^2}$$

حيث X تمثل القيم، f تمثل التكرارات. وإذا كانت البيانات على شكل فئات تحسب مراكز هذه الفئات ويرمز لها بالرمز X ، ويتم حساب $\sum X^2 f$ ، $\sum Xf$

مثال (٧) : الجدول التالي يمثل توزيع مجموعة من الناخبين حسب أعمارهم وفي وقت معين، والمطلوب حساب الانحراف المعياري لأعمار الناخبين ؟

الأعمار X	عدد الناخبين f
20	3
25	4
30	5
35	2
المجموع	14

الحل :

ينظم الحل كما في الجدول التالي حيث يتم حساب كل من Xf ، $X^2 f$ للحصول على $\sum Xf$ ، $\sum X^2 f$

الأعمار X	أعداد الناخبين f	حاصل ضرب Xf	حاصل ضرب $X^2 f$
20	3	60	1200
25	4	100	2500
30	5	150	4500
35	2	70	2450
المجموع	$\sum f = 14$	$\sum Xf = 380$	$\sum X^2 f = 10650$

ويحسب الانحراف المعياري باستخدام الصيغة التالية :

$$s = \sqrt{\frac{\sum x^2 f}{\sum f} - \left(\frac{\sum xf}{\sum f}\right)^2}$$

$$= \sqrt{\frac{10650}{14} - \left(\frac{380}{14}\right)^2}$$

$$= \sqrt{760.7 - 736.6}$$

$$= \sqrt{24.1} = 4.9$$

أي أن الانحراف المعياري لأعمار الناخبين يساوي 4.9 سنة.
مثال (٨) شامل على التشتت :

البيانات في الجدول الخاص بالمثل الشامل رقم (٦) على المتوسطات التي تمثل أهم الحروب التي شهدتها العالم خلال الفترة من عام 1945، وحتى عام 1980م. احسب المدى، الانحراف المعياري، ومعامل الاختلاف لعدد الحروب خلال الفترة ؟
الحل :

يمكن تلخيص بيانات الجدول رقم (١) في الجدول التالي :

التكرارات f	الحروب X
4	0
8	1
7	2
9	3
6	4
2	6
36	المجموع

والجدول يوضح أن هناك أربع سنوات لم تحدث فيها حروب، وثمان سنوات حدثت في كل منها حرب واحدة، وسبع سنوات حدثت في كل منها حربان، وهكذا....
أولاً: المدى

المدى هو الفرق بين أكبر قيمة وأصغر قيمة. وأكبر قيمة هنا تساوي 6 حروب وأصغر قيمة تساوي صفر (لا توجد حروب)، وبالتالي فالمدى هو :
(6 - 0 = 6) أي أن المدى يساوي 6 حروب.

ثانياً: الانحراف المعياري :

يمكن تلخيص الحسابات كما في الجدول التالي :

الحروب X	التكرارات f	حاصل ضرب $X \cdot f$	حاصل ضرب $X^2 f$
0	4	0	0
1	8	8	8
2	7	14	28
3	9	27	81
4	6	24	96
6	2	12	72
المجموع	$\Sigma f = 36$	$\Sigma xf = 85$	$\Sigma x^2 f = 285$

$$S = \sqrt{\frac{\Sigma x^2 f}{\Sigma f} - \left(\frac{\Sigma xf}{\Sigma f}\right)^2}$$

$$= \sqrt{\frac{285}{36} - \left(\frac{85}{36}\right)^2}$$

$$= \sqrt{7.917 - 5.569}$$

$$= \sqrt{2.348} = 1.53$$

أي أن الانحراف المعياري لعدد الحروب خلال تلك الفترة يساوي 1.53 حرباً
(لاحظ أن التباين يساوي 2.348 فهو القيمة قبل استخراج الجذر)

معامل الاختلاف :

معامل الاختلاف يساوي خارج قسمة الانحراف المعياري على الوسط الحسابي مضروباً في (100) أي
أن :

$$CV = \frac{S}{\bar{X}} \times 100$$

١- الوسط الحسابي (كما حسبناه سابقاً) :

$$\bar{X} = \frac{\Sigma xf}{\Sigma f} = \frac{85}{36} = 2.36$$

٢- الانحراف المعياري (كما حسبناه أيضاً)

$$S = 1.53$$

٣- معامل الاختلاف هو :

$$CV = \frac{1.53}{2.36} \times 100$$

$$= 64.8 \%$$

مثال: احسب التباين للبيانات التالية التي تمثل درجات 8 طلبة:

6	5	5	8	9	10	6	7
---	---	---	---	---	----	---	---

الحل :

$$\bar{X} = \sum X_i \div n = 56 \div 8 = 7$$

الوسيط الحسابي هو :

القيم (X_i)	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
7	0	0
6	- 1	1
10	3	9
9	2	4
8	1	1
5	- 2	4
5	- 2	4
6	- 1	1
$\sum (X_i - \bar{X})^2$		24

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{n - 1} = \frac{24}{8-1} = \frac{24}{7} = 3.43$$

التباين هو

مثال: احسب التباين للبيانات التالية التي تمثل توزيع 200 عامل على حسب الوقت الذي يقضونه في العمل في المصنع شهرياً:

الوقت الإضافي (بالساعات)	عدد العاملين
0 إلى أقل من 10	20
10 إلى أقل من 20	80
20 إلى أقل من 30	50
30 إلى أقل من 40	40
40 إلى أقل من 50	10
المجموع	200

الحل:

الفئة	التكرار f_i	مركز الفئة (X_i)	$X_i f_i$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$(X_i - \bar{X})^2 f_i$
0 إلى أقل من 10	20	5	100	- 17	289	5780
10 إلى أقل من 20	80	15	1200	- 7	49	3920
20 إلى أقل من 30	50	25	1250	3	9	450
30 إلى أقل من 40	40	35	1400	13	169	6760
40 إلى أقل من 50	10	45	450	23	529	5290
المجموع	200		4400			22200

$$S^2 = \frac{\sum f_i (X_i - \bar{X})^2}{\sum f_i - 1} \quad \text{التباين هو :}$$

$$S^2 = \frac{22200}{200-1} = \frac{22200}{199} = 111.56 \text{ (ساعة)}^2$$

ونلاحظ أن التباين وحداته هي مربع وحدات القياس الأصلية . وكثيراً ما تكون غير ذات معنى فمثلاً في هذا المثال التباين = 111.56 ساعة تربيع. فهذا ساعة تربيع ليس لها أي معنى . وهذا هو عيب التباين . ولذلك نرجع الوحدات إلى أصلها بأخذ الجذر التربيعي للتباين . ويسمى المقياس الجديد بالانحراف المعياري .

مثال: احسب التباين للبيانات التالية التي تمثل درجات 8 طلبة:

6	5	5	8	9	10	6	7
---	---	---	---	---	----	---	---

الحل :

$$S^2 = \frac{1}{n-1} \left(\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right) \quad \text{التباين هو :}$$

x_i	6	5	5	8	9	10	6	7
x_i^2	36	25	25	64	81	100	36	49

إذن :

$$S^2 = \frac{1}{8-1} \left[416 - \frac{(56)^2}{8} \right] \quad \text{التباين هو :}$$

$$= \frac{1}{7} [416 - 392] = \frac{24}{7} = 3.43$$

من ذلك تكون قيمة الانحراف المعياري هي :

$$\sqrt{3.43} = 1.85$$

مثال: احسب التباين والانحراف المعياري للبيانات التالية التي تمثل توزيع 200 عامل على حسب الوقت الذي يقضونه في العمل في المصنع شهريا:

عدد العاملين	الوقت الإضافي (بالساعات)
20	0 إلى أقل من 10
80	10 إلى أقل من 20
50	20 إلى أقل من 30
40	30 إلى أقل من 40
10	40 إلى أقل من 50
200	المجموع

$$S^2 = \frac{1}{\sum f_i - 1} \left[\sum X_i^2 f_i - \frac{(\sum X_i f_i)^2}{\sum f_i} \right] \quad \text{الحل :}$$

$X_i^2 f_i$	X_i^2	$X_i f_i$	مركز الفئة X_i	التكرار f_i	(الفئة)
500	25	100	5	20	0 إلى أقل من 10
18000	225	1200	15	80	10 إلى أقل من 20
31250	625	1250	25	50	20 إلى أقل من 30
49000	1225	1400	35	40	30 إلى أقل من 40
20250	2025	450	45	10	40 إلى أقل من 50
119000		4400		200	المجموع

$$S^2 = \frac{1}{200-1} \left[119000 - \frac{(4400)^2}{200} \right] \quad \text{التباين هو :}$$

$$= \frac{1}{199} (119000 - 96800) = 111.56$$

ومن ذلك نجد أن قيمة الانحراف المعياري هي :

$$S = \sqrt{\text{التباين}} = \sqrt{111.56} = 10.56$$

خواص الانحراف المعياري :

1. الانحراف المعياري هو أهم مقاييس التشتت وأكثرها استخداماً
2. تدخل في حسابه جميع القيم المشاهدة دون إهمال أي قيمة .
3. يتميز بقابليته للمعالجات الجبرية .
4. لا يمكن حسابه للتوزيعات التكرارية المفتوحة .

مثال:

قارن بين التشتت أطوال الطلبة وأوزانهم . وذلك باستخدام البيانات التالية .
التي تمثل أطوال وأوزان 100 طالب .

الأوزان (كجم)	عدد الطلبة
60 إلى أقل 63	6
63 إلى أقل 66	16
66 إلى أقل 69	40
69 إلى أقل 72	28
72 إلى أقل 75	10

الأطول بالسنتيمتر	عدد الطلبة
110 إلى أقل من 120	12
120 إلى أقل من 130	15
130 إلى أقل من 140	25
140 إلى أقل من 150	30
150 إلى أقل من 160	10
160 إلى أقل من 170	8

الحل :

نحسب معامل الاختلاف لكل من التوزيعين ثم نقوم بالمقارنة . ولحساب معامل الاختلاف لكل توزيع يلزمنا حساب الوسط الحسابي والانحراف المعياري لكل منهما . وذلك كما يلي :

أولاً : توزيع الأطوال :

نوضح العمليات الحسابية التي تلزمنا لحساب الوسط الحسابي والانحراف المعياري في الجدول التالي :

(الفئة)	التكرار f_i	مركز الفئة X_i	$X_i f_i$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
110 إلى أقل من 120	12	115	1380	-23.5	6627
120 إلى أقل من 130	15	125	1875	-13.5	2733.75
130 إلى أقل من 140	25	135	3375	-3.5	306.25
140 إلى أقل من 150	30	145	4350	6.5	1267.5
150 إلى أقل من 160	10	155	1550	16.5	2722.5
160 إلى أقل من 170	8	165	1320	26.5	5618
المجموع	100		13850		19275

الوسط الحسابي هو : $\bar{X} = \sum X_i f_i \div \sum f_i = 138.50$

قيمة التباين هي :

$$S^2 = \frac{1}{\sum f_i - 1} \left[\sum f_i (X_i - \bar{X})^2 \right]$$

$$= \frac{1}{100-1} (19275) = 194.7$$

وتكون قيمة الانحراف المعياري هي : $S = \sqrt{194.7} = 13.93$

معامل الاختلاف = $100 \times \frac{\text{الانحراف المعياري}}{\text{الوسط الحسابي}}$

$$\%10.02 = 100 \times \frac{13.93}{138.5} = \text{معامل الاختلاف للأطول}$$

ثانيا : توزيع الأوزان :

الجدول التالي يوضح لنا العمليات الحسابية اللازمة للحصول على الوسط الحسابي والانحراف المعياري للأوزان :

(الفئة)	التكرار f_i	مركز الفئة X_i	$X_i f_i$	$X_i - \bar{X}$	$(X_i - \bar{X})^2 f_i$
60 إلى أقل من 63	6	61.5	369	-6.6	261.36
63 إلى أقل من 66	16	64.5	1032	-3.6	207.36
66 إلى أقل من 69	40	67.5	2700	-0.6	14.4
69 إلى أقل من 72	28	70.5	1974	2.4	161.28
72 إلى أقل من 75	10	73.5	735	5.4	291.6
المجموع	100		6810		936

قيمة الوسط الحسابي هو : $X = \sum X_i f_i \div \sum f_i = 68.1$

قيمة التباين هي : $S^2 = \frac{1}{\sum f_i - 1} \sum f_i (X_i - X)^2$

$$= \frac{936}{99} = 9.45$$

وتكون قيمة الانحراف المعياري هي : $S = \sqrt{9.45} = 3.07$

$$100 \times \frac{\text{الانحراف المعياري}}{\text{الوسط الحسابي}} = \text{معامل الاختلاف}$$

$$\text{معامل الاختلاف للأوزان} = 100 \times \frac{3.06}{68.1} = 4.49\%$$

بما أن معامل الاختلاف للأطوال أكبر من معامل الاختلاف للأوزان ، إذن تشتت الأطوال أكبر من تشتت الأوزان .

المتغير المعياري

يستخدم بين التوزيعات المختلفة ، يستخدم لدراسة ومقارنة التشتت بين مجموعتين أو أكثر ، يقيس الانحرافات عن المتوسط الحسابي .

$$Z = \frac{X - \bar{X}}{\sigma} \quad \text{مثال: وقانونه هو}$$

المادة	الدرجة X	الوسط الحسابي	الانحراف المعياري
الرياضيات	85	80	3
الإحصاء	80	70	4

$$Z = \frac{85 - 80}{3} = \frac{5}{3} = 1.67 \quad \text{المتغير المعياري لمادة الرياضيات}$$

$$Z = \frac{80 - 70}{4} = \frac{10}{4} = 2.5 \quad \text{المتغير المعياري لمادة الإحصاء}$$

المتغير المعياري للإحصاء أكبر من المتغير المعياري للرياضيات ومنه نستنتج أن الطالب أكثر استيعاباً في مادة الإحصاء .

مثال

إذا فرضنا أن درجة طالب ما في الرياضة 60 ودرجته في مادة الإحصاء 71 . وكان توزيع درجات الرياضة له وسط حسابي يساوي 55 وانحراف معياري 4 . وتوزيع درجات الإحصاء وسطه الحسابي 68 وانحرافه المعياري 5 . فقارن بين مستواه في المادتين؟

الحل :

لإجراء المقارنة يجب تحويل كل قيمة إلى قيمتها المعيارية ثم نقارن بين القيمتين المعياريتين .

$$\begin{aligned} Z &= \frac{60-55}{4} = 1.25 \\ &= \frac{71-68}{5} = 0.60 \end{aligned} \quad \text{القيمة المعيارية للإحصاء}$$

نلاحظ أن القيمة المعيارية للرياضة أكبر من القيمة المعيارية للإحصاء وبالتالي فمستوى الطالب في الرياضة أعلى من مستواه في الإحصاء .

تمارين:

1. عرف خاصية التشتت . مع ذكر أهم مقاييسها .

2. إذا كان لدينا مجموعتين من البيانات :

المجموعة الأولى : 25 12 14 16 23 15 21

المجموعة الثانية : 56 82 70 28 60 64 42 30

أ. لكل مجموعة من هاتين المجموعتين . احسب ما يلي :

المدى . الانحراف الربيعي . متوسط الانحرافات المطلقة . التباين .
الانحراف المعياري .

ب. قارن بين تشتت المجموعتين باستخدام معامل الاختلاف .

3. يوضح الجدول التالي التوزيع التكراري لدرجات 100 طالب :

الدرجة	19-0	39-20	59-40	79-60	99-80
عدد الطلبة	5	15	35	40	5

فاحسب :

أ. المدى .

ب. الانحراف الربيعي .

ج. متوسط الانحرافات المطلقة .

د. التباين .

هـ. الانحراف المعياري .

و. معامل الاختلاف .

4. يوضح جدول التوزيع التكراري التالي توزيع 40 عامل على حسب عدد الأيام الغائبين فيها خلال سنة .

عدد أيام الغياب	عدد العاملين
2 – 0	10
5 – 3	15
8 – 6	8
11 – 9	5
14 – 12	2

احسب :

أ. المدى .

ب. الانحراف الربيعي .

ج. متوسط الانحرافات المطلقة .

د. التباين .

هـ. الانحراف المعياري .

و. معامل الاختلاف .

5. الجدول التالي يبين أوزان 50 طالب بالكيلو جرامات :

الوزن (بالكيلوجرام)	عدد الطلبة
55 إلى أقل من 65	8
65 إلى أقل من 75	10
75 إلى أقل من 85	20
85 إلى أقل من 95	8
95 إلى أقل من 105	4

فاحسب :

أ. المدى .

ب. الانحراف الربيعي .

ج. متوسط الانحرافات المطلقة .

د. التباين .

هـ. الانحراف المعياري .

و. معامل الاختلاف .

6. يوضح جدول التوزيع التكراري التالي توزيع 40 طالب على حسب عدد الساعات التي يقضيها الطالب في المذاكرة شهرياً .

عدد الطلبة	ساعات المذاكرة
3	24 إلى أقل من 40
5	40 إلى أقل من 56
10	56 إلى أقل من 72
12	72 إلى أقل من 88
5	88 إلى أقل من 104
5	104 إلى أقل من 120

فاحسب .

أ. المدى .

ب. الانحراف الربيعي .

ج. متوسط الانحرافات المطلقة .

د. التباين .

هـ. الانحراف المعياري .

7. إذا كانت القيم التالية مرتبة ترتيباً تصاعدياً :

$$X_1, 22, 25, 30, X_5$$

وعلمت أن المدى = 18 . والوسط الحسابي = 27 . فما قيمة X_5, X_1 .

8. علل ما يأتي :

- لا نستطيع حساب المدى للجداول التكرارية المفتوحة .
- يعتبر المدى مقياساً مضللاً في حالة وجود قيم متطرفة .
- نهمل إشارة الانحرافات عند حساب متوسط الانحرافات المطلقة .
- التباين لا يمكن أن يكون مقداراً سالباً .
- الانحراف المعياري أكثر استعمالاً في التحليل الإحصائي من متوسط الانحرافات المطلقة .
- لا نستطيع مقارنة تشتت الأطوال والأوزان باستخدام الانحراف المعياري .

9. إذا علمت أن معامل الاختلاف = 25% وأن التباين ($S^2 = 25$) . وذلك للبيانات التالية :

$$10, 18, 25, X, 15$$

فاحسب القيمة المجهولة X .

10. إذا علمت أن لتوزيع ما ، معامل الاختلاف = 20% . وأن القيمة الأصلية

$$X = 40$$

القيمة المعيارية التي تقابل هذه القيمة هي ($Z = 3$) . فما

قيمة الوسط الحسابي والانحراف المعياري لهذا التوزيع .

11. حول كل قيمة من القيم التالية إلى قيمتها المعيارية (Z) . ثم احسب الوسط

الحسابي والتباين للقيم المعيارية .

$$\text{القيم } (X) : 9, 3, 8, 6, 4$$

12. توضح البيانات التالية التوزيعين التكراريين لمائة عامل على حسب مرتباتهم وعلى حسب عدد أطفالهم . فقارن بين تشتت هذين التوزيعين .

عدد الأطفال	عدد العاملين
0 – 2	15
3 – 5	25
6 – 8	35
9 – 11	20
12 – 14	5

المرتبات (بالدينار)	عدد العاملين
100 إلى أقل من 150	10
150 إلى أقل من 200	20
200 إلى أقل من 250	30
250 إلى أقل من 300	25
300 إلى أقل من 350	15

13. إذا علمت أن درجة طالب ما في المحاسبة 72 ودرجته في الاقتصاد 69 . ودرجته في مادة الإدارة 75 . فإذا علمت أن توزيع درجات المحاسبة له وسط حسابي يساوي 60 وانحراف معياري 3 . وتوزيع درجات الاقتصاد وسطه الحسابي 68 وانحرافه المعياري 4 . وتوزيع مادة الإدارة وسطه الحسابي 70 وانحرافه المعياري 2 فقارن بين مستواه في هذه المواد .

14. مقاييس التشتت في هذا الفصل معروضة بحيث كل مقياس يغطي عيب أو أكثر من عيوب المقياس الذي قلبه فاذا ذكر .

- عيوب المدى التي يغطيها الانحراف الربيعي .
- عيوب الانحراف الربيعي التي يغطيها متوسط الانحرافات المطلقة .
- عيوب متوسط الانحرافات المطلقة التي يغطيها التباين.
- عيوب التباين التي يغطيها الانحراف المعياري .

⊕ الأمر Descriptive:

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

$$Z = \frac{X - \mu}{S}$$

حيث X هي القيمة المدخلة، μ هي المتوسط الحسابي للمتغير X وتقرأ (ميو)، S هو الانحراف المعياري للمتغير X والقيم المعيارية تظهر عادة في نهاية ملف البيانات.

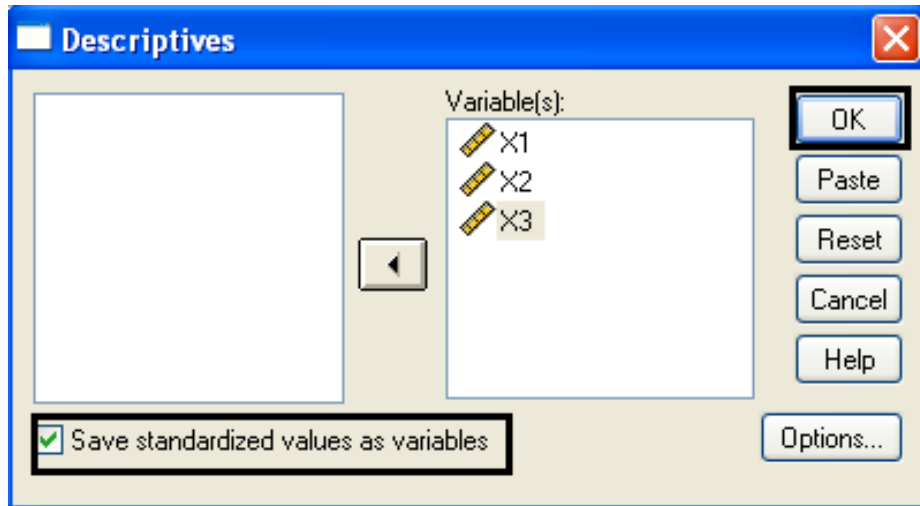
مثال (٥-١):

الجدول التالي يتضمن المتغيرات X_1 ، X_2 ، X_3 والتي تم إدخالها في Data editor للبرنامج:

X_1	X_2	X_3
90	50	12
70	52	15
56	55	19
65	60	22
85	65	20
60		
69		
57		
50		
75		
62		
51		
85		

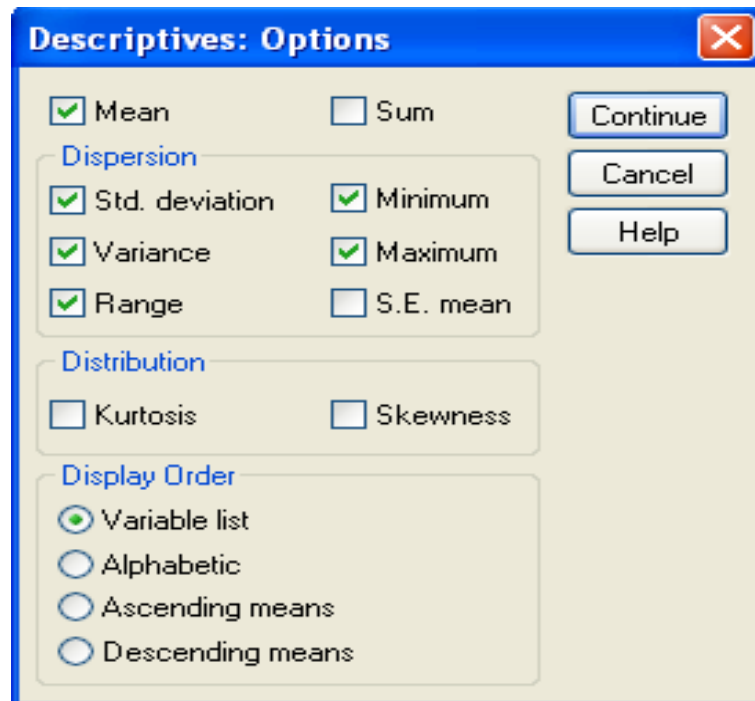
لغرض استخراج المقاييس الوصفية بالأمر Descriptive نتبع الخطوات التالية :

⊕ 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 الاتي:



حيث أن :

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



اختر المقاييس المطلوبة.

✓ Display Order : يمكننا من خلالها ترتيب عرض المقاييس الوصفية للمتغيرات، وقد اخترنا Variable List: الذي يعرض المقاييس الوصفية حسب تسلسل المتغيرات الوارد 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 التالية:

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						

حيث تم عرض المقاييس الإحصائية المختارة حسب ترتيب المتغيرات في خانة Variables في صندوق حوار Descriptive، كما تم إضافة القيم المعيارية للمتغيرات إلى Data editor كما في الشكل التالي:

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	.	.

الاسبوع 13 - 14

الارتباط والانحدار : Correlation and Regression

معامل الارتباط Auto Correlation

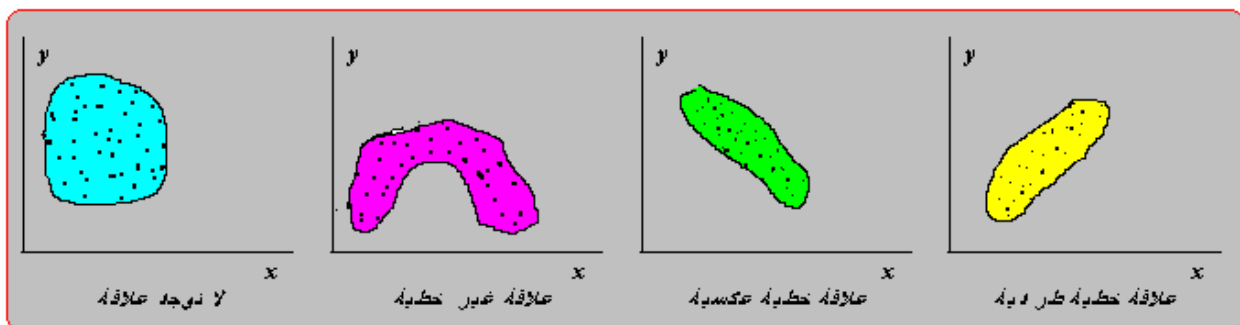
لاحظنا في الفصول السابقة الطرق و الأساليب المختلفة في جمع البيانات و تصنيف و تبويب البيانات ، كذلك عملية استخراج بعض المقاييس التي تعطي فكرة أكثر و وضوحا عن تلك البيانات كالمتوسطات و مقاييس التشتت ، إن هذه الطرق و الأساليب استندت على البيانات المجمعة عن متغير واحد فقط سواء كانت هذه البيانات مبوبة في توزيع تكراري أم غير ذلك . وفي أحوال كثير يواجه الباحث حالات تتطلب دراسة متغيرين أو أكثر في آن واحد لبيان طبيعة و نوع العلاقة التي تربط بها هذه المتغيرات ، عليه فان هذا الفصل سوف يخصص لدراسة مقاييس أخرى تحدد درجة و نوع وشكل العلاقة بين متغيرين أو أكثر.

الارتباط الخطي Linear Correlation

أن مفهوم الارتباط الخطي يقتصر بحالة وجود متغيرين أو أكثر ترتبط مع البعض بعلاقات خطية معينة على سبيل المثال العلاقة بين طول الشخص و وزنه ، العلاقة بين حصول الطالب المتخرج من الكلية و معدل درجاته في الثانوية و مستوى المعاشي لأسرته ، العلاقة بين نسبة الشفاء من مرض معين و كمية الجرعة من الدواء المخصص للمريض و عمر المريض. فإذا كان التغير في إحدى المتغيرات يؤثر في تغير متغير آخر أو مجموعة متغيرات أخرى عندئذ يقال أن هذه المتغيرات مرتبطة فيما بينها و إذا كان المتغيرين المرتبطين (أو مجموعة من المتغيرات المترابطة) يتغيران (تتغير) بنفس الاتجاه أي زيادة أو نقصان في إحدهما تؤدي إلى الزيادة أو النقصان في الآخر (الأخرى) عندئذ يقال أن الارتباط ما بينهما هو ارتباط موجب ، على سبيل المثال زيادة طول الشخص يتوقع أن يقابلها زيادة في وزنه. انخفاض في دخل الفرد يتوقع عنه انخفاض في إنفاقه على بعض السلع. أما إذا كان المتغيرين المرتبطين (أ، مجموعة من المتغيرات المرتبطة) يتغيران (تتغير) باتجاه معاكس أي زيادة أ، نقصان في إحدهما تؤدي إلى نقصان (زيادة) في الآخر (الأخرى)، عندئذ يقال أن الارتباط ما بينهما هو ارتباط سالب على سبيل المثال زيادة سعر الوحدة من سلعة معينة يتوقع أن يؤدي إلى انخفاض في الطلب على تلك السلعة ، انخفاض في درجات الحرارة يتوقع أن ينجم عنه زيادة الطلب على الوقود.

و يقال أن الارتباط بين المتغيرين أو أكثر هو ارتباط تام perfect إذا كان التغير في إحدهما متناسب مع التغير في الآخر على سبيل المثال إن الارتباط بين درجة الحرارة المئوية و درجة الحرارة الفهرنهايتية هو ارتباط تام باعتبار إن التغير في الأول متناسب مع التغير في الثاني.

$$F = (9/5) C + 32$$



الارتباط الخطي البسيط Simple Correlation

إذا كان الغرض من التحليل هو تحديد نوع وقوة العلاقة بين متغيرين ، يستخدم تحليل الارتباط ، وأما إذا كان الغرض هو دراسة وتحليل أثر أحد المتغيرين على الآخر ، يستخدم تحليل الانحدار، وفي هذا الفصل يتم عرض أسلوب تحليل الارتباط الخطي البسيط، أي في حالة افتراض أن العلاقة بين المتغيرين تأخذ الشكل الخطي .

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 (رو)، وفي حالة العينة بالرمز r ، وحيث أننا 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 ثلاث 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 يصاحبه انخفاض في المتغير الثاني، والعكس.

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

إذا كان معامل الارتباط قيمته صفراً ($r = 0$) دل ذلك على انعدام العلاقة بين المتغيرين.

قوة العلاقة:- ويمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن (± 1)، حيث أن قيمة معامل الارتباط تقع في المدى ($-1 < r < 1$)، وقد صنف بعض الإحصائيين درجات لقوة العلاقة يمكن تمثيلها على الشكل التالي

درجات قوة معامل الارتباط

ارتباط عكسي					ارتباط طردي					
قوي جدا	قوي	متوسط	ضعيف	ضعيف جدا	ضعيف جدا	ضعيف	متوسط	قوي	قوي جدا	
-1	-0.9	-0.7	-0.5	-0.3	0	0.3	0.5	0.7	0.9	1
نام					نام					

معامل الارتباط الخطي البسيط Simple Linear Correlation Coefficient

يعرف معامل الارتباط الخطي البسيط بأنه القيمة العددية للعلاقة الخطية بين متغيرين و تحسب من القانون الاتي:

$$r_{xy} = \frac{n \sum_{i=1}^n X_i Y_i - (\sum X_i \sum Y_i)}{\sqrt{\left\{ (n \sum X_i^2 - (\sum X_i)^2) (n \sum Y_i^2 - (\sum Y_i)^2) \right\}}}$$

مثال : احسب معامل الارتباط الخطي البسيط للبيانات التالية :

عدد العمال X	الإنتاج Y
1	6
2	11
3	15
4	19
5	21
15	72

الحل : يجب تطبيق القانون التالي

$$r = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{n \sum X^2 - (\sum X)^2} \sqrt{n \sum Y^2 - (\sum Y)^2}}$$

ولتطبيق القانون يجب إيجاد قيم كل من

قيمة $N =$ هي عدد القيم وهي (5)

قيمة $X =$ قيمة Y وقيمة X^2 وقيمة Y^2 وقيمة XY وطريق الحل حسب الجدول التالي

عدد العمال X	الإنتاج Y	X^2	Y^2	XY
1	6	1	36	6
2	11	4	121	22
3	15	9	225	45
4	19	16	361	76
5	21	25	441	105
15	72	55	1184	254

أيضاً نوجد $(\sum Y)^2$ وهو $72 \times 72 = 5184$

أيضاً نوجد $(\sum X)^2$ وهو $15 \times 15 = 225$

بعد إيجاد القيم نطبق القانون دون تغيير

$$r = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{n \sum X^2 - (\sum X)^2} \sqrt{n \sum Y^2 - (\sum Y)^2}}$$

$$r = \frac{5(254) - (15 \times 72)}{\sqrt{5(55) - (15)^2} \sqrt{5(1184) - (72)^2}}$$

$$r = \frac{1270 - 1080}{\sqrt{275 - 225} \sqrt{5920 - 5184}}$$

$$r = \frac{190}{\sqrt{50 \times 736}} = \frac{190}{\sqrt{36800}} = \frac{190}{191.83} = .99$$

وبهذه النتيجة أصبح حساب معامل الارتباط الخطي هو (0.99) ويعتبر ارتباط طردي قوي (لأن الإشارة موجبة) ، وإذا كانت الإشارة سالبة فإن النتيجة هي ارتباط عكسي .
ولكن إذا كانت النتيجة بإشارة موجب تصبح ارتباط طردي وإذا كانت النتيجة بإشارة السالب تكون ارتباط عكسي



ملاحظة : يجب أن تكون إجابة معامل الارتباط أن لا تتجاوز (+1) صحيح ، فإذا وجدنا قيمة أكبر من (1) فنقول الناتج خاطئ. ، فيجب أن تنحصر القيمة بين $-1 \leq r \leq +1$.
إذا كانت العلاقة (0) فنقول عنها معدومة ، وإذا كانت (1) فنقول علاقة تامة .

بعض خصائص معامل الارتباط.

١. معامل الارتباط بين X, Y هو نفسه بين Y, X .
٢. تنحصر قيمة معامل الارتباط بين $-1, +1$ وإذا كانت قيمة معامل الارتباط بالموجب فإن الارتباط يكون طردياً وإذا كانت قيمة معامل الارتباط بالسالب فن الارتباط يكون عكسياً.
٣. قيمة معامل الارتباط بين العامل المتغير ونفسه هي (1) .
٤. لا يتأثر معامل الارتباط بأي من العمليات الحسابية أو الجبرية التي تجري على X, Y أي لا يتأثر بالطرح والجمع أو الضرب والقسمة.
٥. إذا كان المتغيران X, Y مستقلين فإن قيمة معامل الارتباط صفراً ولكن العكس غير صحيح.

مثال:

52	53	50	49	48	X
31	39	33	29	28	Y

الحل :

باستخدام معامل الارتباط بالصيغة المباشرة . نكون الجدول التالي :

Y^2	X^2	$Y X$	Y	X	ر.م
784	2304	1344	28	48	1
841	2401	1421	29	49	2
1089	2500	1650	33	33	3
841	2601	1475	29	29	4
961	2704	1612	31	31	5
4516	12510	7506	150	150	المجموع

$$r = \frac{n \sum XY - \sum X \sum Y}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}}$$

$$r = \frac{5(7506) - (150)(250)}{\sqrt{[5(12510) - (250)^2][5(4516) - (150)^2]}}$$

$$r = \frac{30}{\sqrt{(50)(80)}} = \frac{30}{63.2453} = 0.47$$

وهذا يعنى أن الارتباط طردي ضعيف

مثال:

أوجد معامل الارتباط بين المتغيرين X , Y مبينا نوعه ودرجته حسب بيانات

الجدول :

5	4	3	2	1	X
2	3	1	1	3	Y

الحل :

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

لتحديد معامل الارتباط نكون الجدول التالي :

Y^2	X^2	$Y X$	Y	X	ر.م
9	1	3	3	1	1
1	4	2	1	2	2
1	9	3	1	3	3
9	16	12	3	4	4
4	25	10	2	5	5
24	55	30	10	15	المجموع

$$r = \frac{5(30) - (15)(10)}{\sqrt{[5(55) - (15)^2][5(24) - (10)^2]}}$$

$$r = \frac{150 - 150}{\sqrt{(275 - 225)(120 - 100)}} = \frac{0}{\sqrt{(50)(25)}} = \frac{0}{\sqrt{1000}} = 0$$

وهذا يعنى الارتباط منعدم بين المتغيرين .

⊕ معامل الارتباط الخطي البسيط " لبيرسون " Karl Pearson

في حالة جمع بيانات عن متغيرين كميين (y, x) ، يمكن قياس الارتباط بينهما، باستخدام طريقة "لبيرسون" $Pearson$ ، 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 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ لمتغيرين X و Y ، فإن معامل الارتباط لبيرسون يعرف على أنه:

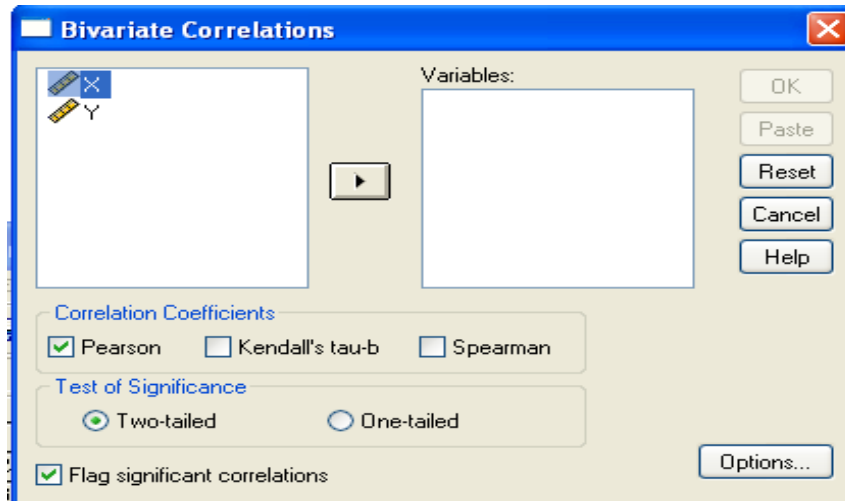
$$r = \frac{\sum_{i=1}^n X_i Y_i - \frac{\sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n}}{\sqrt{\left[\sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i \right)^2}{n} \right] \left[\sum_{i=1}^n Y_i^2 - \frac{\left(\sum_{i=1}^n Y_i \right)^2}{n} \right]}}$$

فإذا نظرنا إلى المعادلة السابقة عرفنا مقدار المعاناة عند حساب مقدار معامل الارتباط ناهيك عن حجم التحليل والحسابات إذا زادت حجم العينة و الآن نتعرف كيف يمكن تطبيق ذلك بمنتهى السهولة باستخدام البرنامج وذلك من خلال المثال القادم:

مثال: احسب معامل الارتباط المناسب بين درجة الطالب في الرياضيات ودرجته في الإحصاء باستخدام البيانات الآتية لعينة من خمسة طلاب:

درجة الرياضة (x)	١٥	١٩	١١	٣	٧
درجة الإحصاء (y)	١٨	٢٠	١٦	١٤	١٢

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



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 Y, X .

⊕ Correlation Coefficients: وهو يتيح لنا تحديد نوع معامل الارتباط الذي نريده ونلاحظ أنه متوفر لدينا ٣ اختيارات:

✓ Pearson: وهو معامل ارتباط الظواهر الكمية سواء كانت مبوبة أو غير مبوبة.

✓ Kendall's tau: وهو معامل ارتباط الظواهر الترتيبية ويفضل للعينات الصغيرة والبيانات الغير مبوبة.

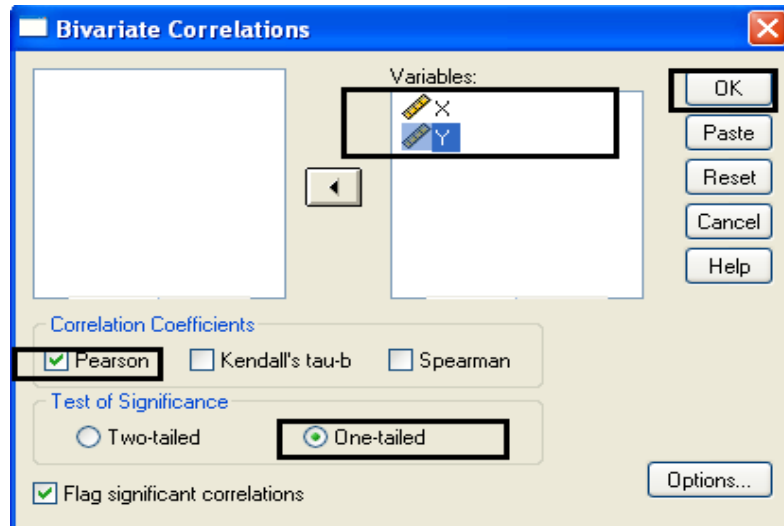
✓ Spearman: وهو أيضا معامل الارتباط بين الظواهر الكمية والترتيبية ويفضل للعينات الصغيرة ويستخدم في حالة البيانات الغير مبوبة.

✓ وهنا نقوم بتحديد معامل بيرسون.

✓ Test of Significance: 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: هنا نستخدم لتعليم الارتباط بعلامة نجمة (star) في حالة الارتباط المعنوي (الدال) أي الارتباط الحقيقي والذي يؤكد الاختبار.

وبعد الاختيار السابق يظهر الشكل الآتي:



⊕ ثم ننقر على ok فيظهر الجدول الآتي:

Correlations		X	Y
X	Pearson Correlation	1	.900*
	Sig. (1-tailed)		.019
	N	5	5
Y	Pearson Correlation	.900*	1
	Sig. (1-tailed)	.019	
	N	5	5

*. Correlation is significant at the 0.05 level (1-tailed)

⊕ ونجد أن معامل الارتباط بين درجة الطالب في الإحصاء ودرجته 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 صفر وهذا ما دل عليه اختيارنا السابق أن الاختبار ذو اتجاهين فإذا افترضنا أننا نريد اختبار ذلك بمستوى معنوية مقدارها $\alpha = 0.05$ فإن قيمة $p\text{-value} = 0.019 < \alpha = 0.05$ دل ذلك على أن الارتباط بين درجة الرياضيات و الإحصاء ارتباط يختلف عن الصفر بمستوى معنوي مقداره 0.05 أي أننا نقبل الفرض البديل القائل بأن الارتباط لا يساوي صفر.

الاسبوع 15 - 17

معامل الارتباط الرتب لسبيرمان Spear mans rank correlation Coefficient

افرض إن X, Y متغيرين من النوع الوصفي و افرض إن البيانات المستحصل عليها من X, Y على أساس عينة عشوائية من المفردات قوامها n هي بهيئة صفات غير قابلة للقياس الكمي .

افرض إن $(X_i, Y_i, i=1, 2, 3, \dots, n)$ ممكنة الترتيب تصاعديا أو تنازليا وفق معيار معين يمتاز به كل متغير (مثلا تقديرات درجات مجموعة من الطلبة يمكن ترتيبها تصاعديا على أساس معيار الأقل إلى اعلي درجة أو العكس) عندئذ استنادا لهذا الترتيب و لكل متغير يمكن تخصيص قيم سلسلة الإعداد الطبيعية $(1, 2, 3, \dots, n)$ لصفات الترتيب بحيث إن كل صفة يخصص لها احد إعداد هذه السلسلة ، في حالة عدم تكرار أية صفة منها (وقد يكون التخصيص تنازلي) و سوف نوضح حالة التكرار بعض الصفات و أسلوب علاج ذلك في أمثلة. و القانون المستخدم لإيجاد الارتباط لمثل هذه الحالات:

$$r_{xy} = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad , \quad d_i = X_i - Y_i$$

مثال:

إذا كانت تقديرات 5 طلبة في مادتين كما هو مبين أدناه . فاحسب معامل الارتباط بين هاتين الظاهرتين .

الطالب	1	2	3	4	5
تقدير المادة الأولى :	جيد	ممتاز	جيد جداً	مقبول	ضعيف
تقدير المادة الثانية :	مقبول	جيد جداً	جيد	ممتاز	ضعيف

الحل :

لاحظ أن البيانات وصفية وبالتالي معامل الارتباط المناسب لها هو معامل ارتباط الرتب على النحو التالي :

الطالب	تقدير المادة الأولى	تقدير المادة الثانية	رتب المادة الأولى	رتب المادة الثانية	الفروق d	مربعات الفروق
1	جيد	مقبول	3	2	1	$1^2=1$
2	ممتاز	جيد جداً	5	4	1	$1^2=1$
3	جيد جداً	جيد	4	3	1	$1^2=1$
4	مقبول	ممتاز	2	5	-3	$-3^2=1$
5	ضعيف	ضعيف	1	1	0	$0^2=0$
المجموع			15	15	0	12

$$r = 1 - \frac{6\sum d^2}{n(n^2 - 1)} \quad \text{وعليه :}$$

$$r = 1 - \frac{6(12)}{5(25-1)} = 1 - \frac{6(12)}{5(24)} = 1 - \frac{3}{5} = \frac{2}{5} = 0.4$$

وهو ارتباط طردي ليس قويا"

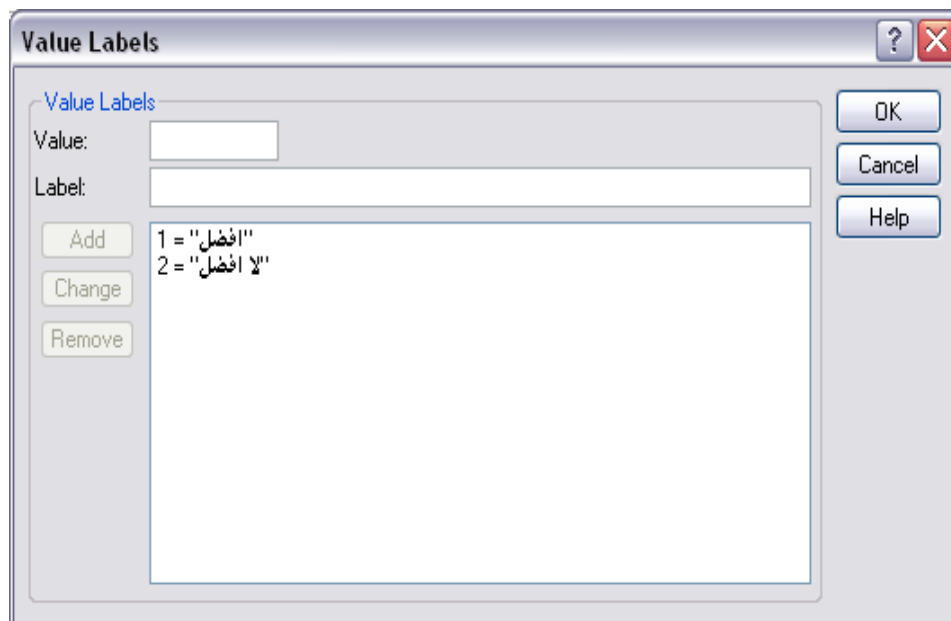
مثال:

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:

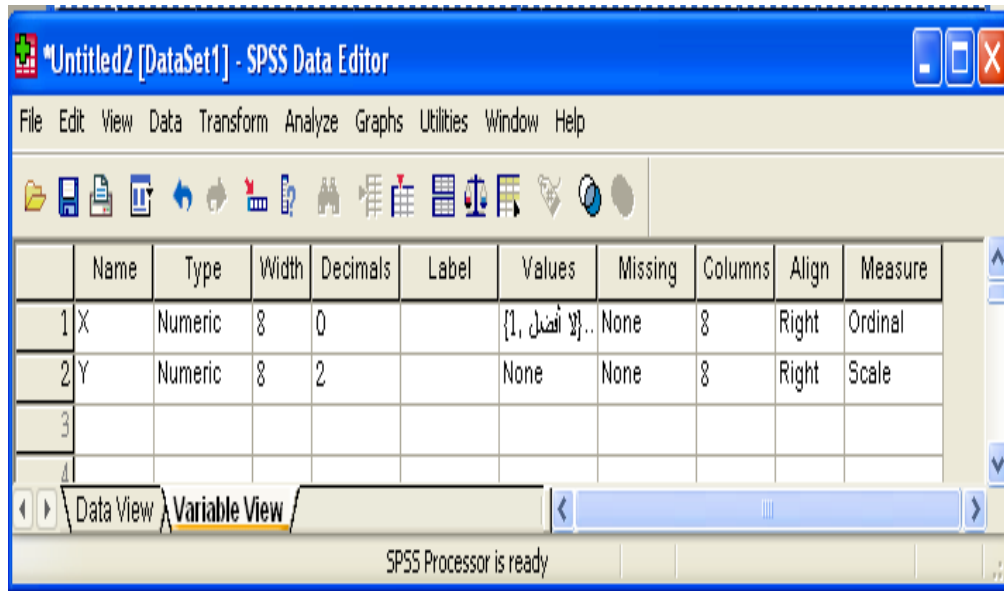
مستوي التفضيل	أفضل	لا أفضل	أفضل	لا أفضل	لا أفضل	لا أفضل	لا أفضل	أفضل	لا أفضل	أفضل
مقدار الدخل الشهري	100	20000	25000	50000	50000	50000	50000	50000	50000	50000

الحل:

خلال البرنامج نقوم بإدخال البيانات على الفرض أن "أفضل" نرسم 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:



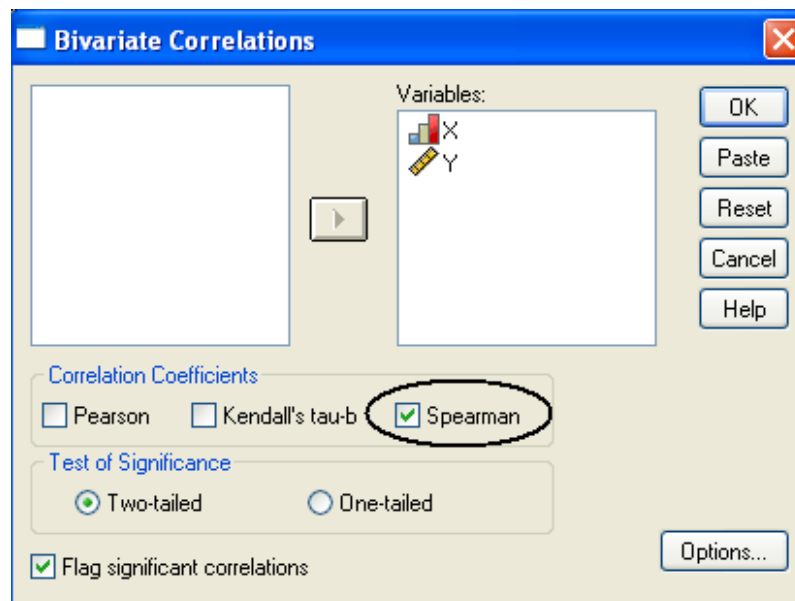
ثم في عمود measure نعرف المتغير X على أنه متغير ترتيبي ordinal ونعرف المتغير Y على أنه متغير كمي عادي فيظهر المحرر كالاتي:



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

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



ونلاحظ أن المتغير X معرف في عمود المتغير المستخدمة في التحليل (العمود الأول من اليمين) وأننا اختارنا spearman كمعامل للارتباط ونفس الاختيارات السابقة في مقياس Pearson ثم ننقر على ok فيظهر الجدول الآتي:

Correlations

		X	Y
Spearman's rho	X	1.000	.070
	Correlation Coefficient		
	Sig. (2-tailed)	.	.848
	N	10	10
Y	Correlation Coefficient	.070	1.000
	Sig. (2-tailed)	.848	.
	N	10	10

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

$$H_0 : \rho = 0$$

ضد الفرض البديل القائل :

$$H_A : \rho \neq 0.$$

ونلاحظ أن الفرض البديل معامل الارتباط لا يساوي صفر وهذا ما دل عليه اختيارنا السابق أن الاختبار ذو اتجاهين فإذا افترضنا أننا نريد اختبار ذلك بمستوى معنوية مقداره $\alpha = 0.05$ فإن قيمة p-value $\alpha = 0.05 = 0.424 >$ دل ذلك على أن الارتباط بين تفضيل المستهلكين لسلعة ومستوى دخولهم ارتباط يساوي الصفر بمستوى معنوي مقداره 0.05 أي أننا نقبل الفرض العدم القائل الارتباط يساوي صفر في كامل بيانات المجتمع.

الاسبوع 18 - 21

تحليل الانحدار Regression Analysis

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 Model Simple Regression أو ما يسمى Ordinary Least Squares و إذا احتوى على أكثر من متغير مستقل فهو نموذج الانحدار المتعدد كما أن النموذج قد يكون خطياً أو غير خطي .

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 على إنتاجية القمح.
- ⊕ دراسة أثر الإنتاج على التكلفة.
- ⊕ دراسة أثر كمية البروتين التي يتناولها الأبقار على الزيادة في الوزن.
- ⊕ أثر الدخل على الإنفاق الاستهلاكي.
- ⊕ دراسة أثر الذكاء على 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:

$$y = \beta_0 + \beta_1 x + e$$

حيث أن:

✓ 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 x ، أي في حالة $x=0$

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

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

⊕ تقدير نموذج الانحدار الخطي البسيط

يمكن تقدير معاملات الانحدار (β_1, β_0) في النموذج باستخدام طريقة المربعات الصغرى، وهذا التقدير هو الذي يجعل مجموع مربعات الأخطاء العشوائية $\sum e^2 = \sum (y - (\beta_0 + \beta_1 x))^2$ أقل ما يمكن، ويحسب هذا التقدير بالمعادلة التالية:

$$\hat{\beta}_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

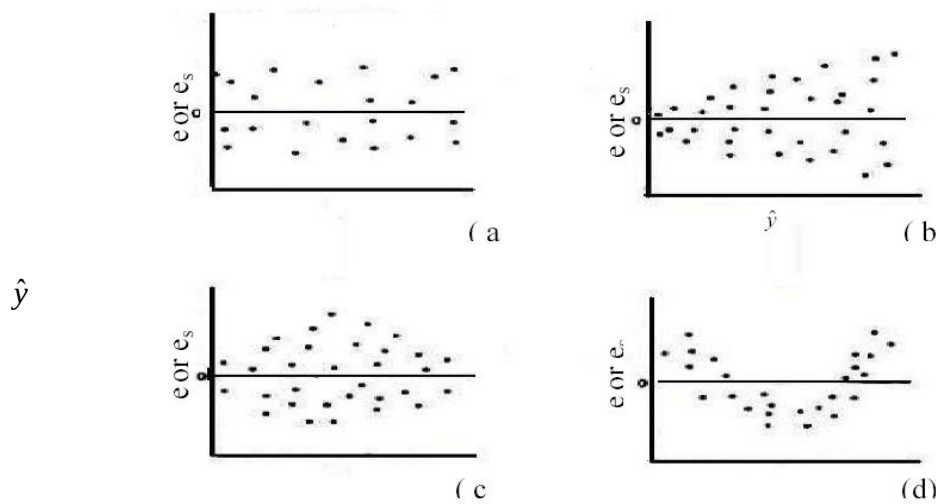
حيث أن $\bar{x}$ هو الوسط الحسابي لقيم x ، $\bar{y}$ هو الوسط الحسابي لقيم y ، وتكون القيمة المقدرة للمتغير التابع هو: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ ، ويطلق على هذا التقدير "تقدير معادلة انحدار y على x ".
علماً بأن نموذج الانحدار الخطي البسيط يجب أن يحقق مجموعة الفروض الآتية:

⊕ وجود علاقة خطية بين X و Y .
⊕ أن الأخطاء العشوائية تتوزع بمتوسط مساوي للصفر.
⊕ أن الأخطاء العشوائية لها تباين ثابت يساوي σ^2 (فرضية تجانس تباين الخطأ العشوائية (Homoscedasticity).

⊕ الأخطاء تتوزع طبيعياً وهذا الشرط ليس ضرورياً لتقدير المعامل بطريقة OLS ولكنه ضروري لاختبار الفرضيات المتعلقة بمعاملات الانحدار β_0 و β_1 .

عدم وجود ارتباط ذاتي Autocorrelation بين الأخطاء العشوائية.

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



- (a) توفر 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 ملائمة العلاقة الخطية (يتوجب استعمال نماذج أخرى مثلا نموذج الدرجة الثانية).
 مثال:

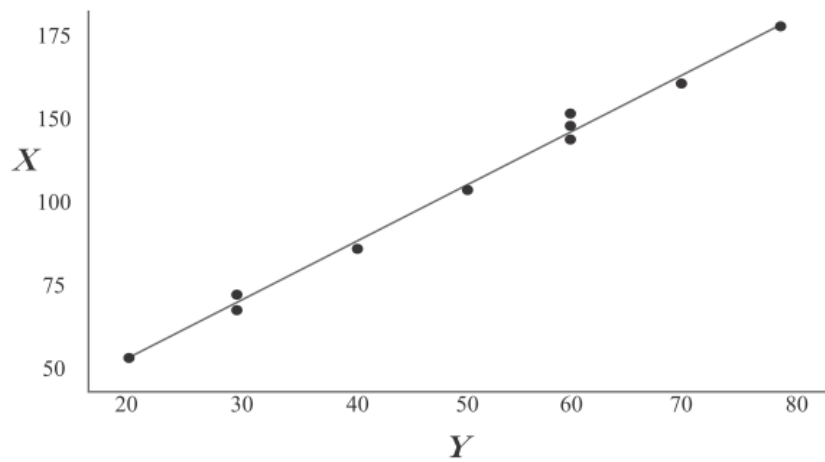
البيانات التالية تمثل الدخل والإنفاق الشهري بالدينار لعشر أسر . والمطلوب :

1. تحديد العلاقة المناسبة من خلال الشكل الانتشاري .
2. تقدير معادلة انحدار الإنفاق (Y) على الدخل (X) المقدرة باستخدام طريقة المربعات الصغرى .
3. تحديد (تقدير) الإنفاق عندما يكون الدخل 100 دينار .
4. تحديد معامل الارتباط بين الدخل والإنفاق .

132	148	69	135	108	87	170	128	50	73	الدخل (X)
60	70	30	60	50	40	80	60	20	30	الإنفاق (Y)

الحل :

لتحديد العلاقة المناسبة بين المتغيرين X , Y نرسم الشكل الانتشاري كما يلي :



1. من الشكل نلاحظ أن العلاقة الخطية $\hat{Y} = \hat{a}X + \hat{b}$ هي العلاقة بين المتغيرين X , Y .

2. لتحديد معاملات العلاقة الخطية $\hat{a}$, $\hat{b}$ نتبع الآتي :

ر.م	الدخل X	الإنفاق Y	YX	X^2	Y^2
1	73	30	2190	5329	900
2	50	20	1000	2500	400
3	128	60	7680	16384	3600
4	170	80	13600	28900	6400
5	87	40	3480	7569	1600
6	108	50	5400	11664	2500
7	135	60	8100	18225	3600
8	69	30	2070	4761	900
9	148	70	10360	21904	4900
10	132	60	7920	17424	3600
المجموع	1100	500	61800	134660	28400

$$\hat{a} = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \sum X^2}$$

$$\hat{a} = \frac{10(61800) - (500)(1100)}{10(134600) - (1100)^2}$$

$$\hat{a} = \frac{618000 - 550000}{134600 - 1210000} = \frac{68000}{13600} = 0.4978$$

لا حظ أن $\hat{a}$ هي معامل معادلة الانحدار المقدرة . أي أنه كلما تغيرت X (الدخل) بدینار واحد تغيرت Y (الإنفاق) بمقدار 0.4978 دينار .

$$\hat{b} = \bar{Y} - \hat{a}\bar{X}$$

$$\bar{X} = \frac{\sum X}{n} = \frac{1100}{10} = 110 \quad , \quad \bar{Y} = \frac{\sum Y}{n} = \frac{500}{10} = 50$$

$$\hat{b} = 50 - 0.4978(110) = 50 - 54.758 = -4.758$$

وحيث أن b هي الجزء المقطوع في معادلة الانحدار المقدرة أي أن قيمة Y (الإنفاق) المقدرة = -4.758 عندما يكون الدخل (X) = صفر . وبالتالي فإن معادلة انحدار Y على هي :

$$\hat{Y} = 0.4978X - 4.758$$

3. الإنفاق عندما يكون الحل $(X) = 110$ دنانير

$$\hat{Y} = 0.4978(100) - 4.758 = 50$$

4. معامل الارتباط

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

$$r = \frac{(10)(61800) - (500)(1100)}{\sqrt{[10(134660) - (1100)^2][10(28400) - (500)^2]}}$$

$$r = \frac{618000 - 550000}{\sqrt{(1346600 - 1210000)(284000 - 250000)}}$$

$$= \frac{68000}{\sqrt{(136600)(34000)}} = \frac{68000}{68149.834} = 0.9979$$

معامل الارتباط ر قريب من الواحد وفي هذه الحالة الارتباط طردي قوي وهذا دليل على قوة العلاقة الخطية بين الدخل (X) والإنفاق (Y) .

السلاسل الزمنية Time Series

تعريف السلسلة الزمنية:

السلسلة الزمنية هي مجموعة مشاهدات مرتبة حسب الزمن (وغالبا ما تكون الفترات الزمنية متساوية ومتعاقبة وتختلف هذه الفترات حسب طبيعة الظاهرة).
استخدامات السلاسل الزمنية:

والسلاسل الزمنية لها استخدامات هامة في مجالات عديدة منها الإحصاءات الاقتصادية والتجارية والسكانية. كما تستخدم نماذج السلاسل الزمنية عادة في التنبؤ بقيم متغير ما. إذا كان المتغير المراد دراسته غير معروفة محدثاته ، ولا العوامل التي تؤثر فيه، كما يستخدم في حالة كون المتغير يخضع لتوقعات المتعاملين معه والتي تنعكس في المستقبل بناء على ما حدث في الماضي.

○ فعلى سبيل المثال ، إذا أخذنا متغير مثل الناتج المحلي الإجمالي خلال سنوات معينة وكان هدفنا تقديره. و حيث أن المتغير كغيره من المتغيرات يتحرك مع الزمن ، فإننا نتوقع أن يأخذ عدة أشكال عند رسمه كأن يكون على شكل خطي ، أو شكل غير خطي .

○ كما توجد متغيرات أخرى مثل التنبؤ بحالة الجو، ودرجات الحرارة ، تدفق المياه ، واستهلاك الوقود ، حالة السوق والأسعار، كمية النقود المتداولة ، ودائع البنوك ، المبيعات الأرباح الخ.....

○ كل هذه المتغيرات يمكن أن تأخذ أشكال سلاسل زمنية إذا لاحظنا تطورها حسب الزمن .

○ وتوجد عدة عوامل تؤثر في تطوره حسب الزمن تؤدي إلى زيادته (دفعة إلى أعلى)، أو نقصانه

(جذبه إلى أسفل) ومن أهم هذه العوامل ما يطلق عليها (عناصر) مكونات السلسلة الزمنية .
مكونات السلسلة الزمنية:

- ١- الاتجاه العام General Trend: يقصد به ميل الظاهرة نحو الزيادة أو النقصان خلال فترة طويلة من الزمن
 - ٢- التغيرات الموسمية Seasonal Variations: وهى تغيرات تحدث للظاهرة خلال السنة بسبب اختلاف طبيعة مواسم السنة نفسها.
 - ٣- التغيرات الدورية Cyclical Variations: وهى تغيرات تحدث للسلسلة كل عدة سنوات بحيث تكرر السلسلة نفسها على فترات دورية منتظمة. مثل (دورات النمو والانكماش في الاقتصاد العالمي)
 - ٤- التغيرات العشوائية Variations Random: هي تغيرات تحدث بصفة غير منتظمة وبسبب عوامل فجائية. (مثل الزلازل- الفيضانات - الحرائق - إفلاس البنوك)
- طرق قياس الاتجاه العام للسلسلة الزمنية:
- يتطلب تحليل السلاسل الزمنية عادة تحليل المكونات الأربعة للسلسلة الزمنية التي سبق ذكرها وغالبا ما تخلو السلاسل الزمنية من التغيرات الموسمية ، والتغيرات الدورية ، والتغيرات العشوائية لذلك سنكتفي بتعيين الاتجاه العام الخطى للسلسلة الزمنية كما يلي:
- باعتبار إن السلسلة الزمنية تحتوى على متغيرين أحدهما مستقل X (ويمثل الزمن بالسنوات، بالشهور وهكذا ..)، والآخر تابع Y (ويمثل قيم الظاهرة محل الدراسة) ، ويتم تقدير معادلة الاتجاه العام الخطى على الصورة:

$$Y = B_0 + B_1x + \psi$$

حيث B_0 : مقدار ثابت وهو الجزء المقطوع من المحور الرأسى محور y ،
 B_1 : هو ميل محور y بالنسبة للزمن،
 ψ : حد الخطأ العشوائى.

حيث:

$$B_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$B_0 = \frac{\sum y}{n} - B_1 \frac{\sum x}{n}$$

عدد أزواج قيم المتغيرين n

مثال: الجدول التالي يمثل سلسلة زمنية لتكلفة إنتاج إحدى السلع (بالآلف دينار)

السنة	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
التكلفة	15	16	18	18	20	21	22	24	28	27

والمطلوب:

- ١- تقدير معادلة الاتجاه العام للسلسلة الزمنية.
- ٢- تقدير القيم الاتجاهية لعامي 2002 ، 2006 ثم تخلص قيم العامين السابقين من أثر الاتجاه العام.
- ٣- تقدير التكلفة لعامي 2010 ، 2020 .

الحل

١- تقدير معادلة الاتجاه العام:

كما يلي:

باعتبار أن سنة الأساس

هي سنة 2000 يتم تكوين

الجدول التالي:

السنة	Y	X	X.Y	X ²
2000	15	0	0	0
2001	16	1	16	1
2002	17	2	34	4
2003	18	3	54	9
2004	20	4	80	16
2005	21	5	105	25
2006	22	6	132	36
2007	24	7	168	49
2008	28	8	224	64
2009	27	9	234	81
Σ	208	45	1056	285

حساب B_0 ، B_1 كما يلي

$$B_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} = \frac{10 \cdot 1056 - 45 \cdot 208}{10 \cdot 245 - (45)^2} = 1.45$$

$$B_0 = \frac{\sum y}{n} - B_1 \frac{\sum x}{n} = \frac{208}{10} - 1.45 \cdot \frac{45}{10} = 14.3$$

بالتعويض في المعادلة:

$$Y = B_0 + B_1 x$$

وبالتالي فإن معادلة الاتجاه العام المقدرة هي :

$$Y = 14.3 + 1.45x$$

- تقدير القيم الاتجاهية لعامي 2002 ، 2006.

بالتعويض في معادلة الاتجاه بقيم X المقابلة لعامي 2002 ، 2006 ثم تخلص القيم من أثر الاتجاه العام بالمعادلة :

$$\text{القيمة الاتجاهية} = \frac{\text{القيمة الفعلية}}{\text{القيمة مخرصة من أثر الاتجاه العام}} \times 100$$

(أ) القيمة الاتجاهية لعام 2002.

$$Y = 14.3 + 1.45 (2) = 17.2$$

■ تخليص قيمة عام 2002 من أثر الاتجاه العام.

$$\text{القيمة مخرصة من أثر الاتجاه العام} = \frac{17.2}{17} \times 100 = 101.18\%$$

(ب) القيمة الاتجاهية لعام 2006.

$$Y = 14.3 + 1.45 (6) = 23$$

■ تخليص قيمة عام 2006 من أثر الاتجاه العام.

$$\text{القيمة مخرصة من أثر الاتجاه العام} = \frac{23}{22} \times 100 = 104.54\%$$

٣- تقدير التكلفة لعامي 2010 ، 2020 .

أولا - تقدير التكلفة لعامي 2010.

حساب قيمة x لعام 2010

$$x = 2010 - 2000 = 10 = \text{سنة الأساس} - \text{السنة المراد تقديرها}$$

بالتعويض في المعادلة عن x=10

$$Y = 14.3 + 1.45 (10) = 28.8$$

إذن التكلفة المقدرة لعام 2010 = 28.8 ألف دينار.

ثانيا - تقدير التكلفة لعامي 2020.

حساب قيمة x لعام 2010

$$x = 2020 - 2000 = 20 = \text{سنة الأساس} - \text{السنة المراد تقديرها}$$

بالتعويض في المعادلة عن x=20

$$Y = 14.3 + 1.45 (20) = 43.3$$

إذن التكلفة المقدرة لعام 2020 = 43.3 ألف دينار

مثال يوضح معادلة خط الاتجاه العام .

السنوات	قيمة الانتاج	X
1423	3	1
1424	5	2
1425	4	3
1426	7	4
1427	5	5
1428	8	6

كم الإنتاج المتوقع للعام ١٤٣٢ هـ

حسب المعادلة التالية

$$Y = B_0 + B_1X$$

السنوات	قيمة الإنتاج y	X (نفترضها من عندنا)
1423	3	1
1424	5	2
1425	4	3
1426	7	4
1427	5	5
1428	8	6
1429		7
1430		8
1431		9
1432		10

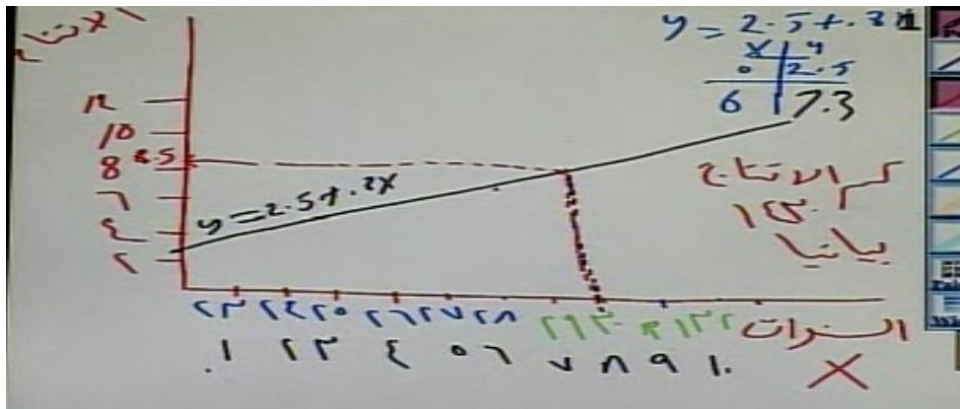
قيمة x باللون الأصفر هي قيمة افتراضية وأخذناها بالتسلسل للقيم السابقة

حسب المعادلة : $y = 2.5 + 0.8X$ وقيمة X في ١٤٣٢ هي ١٠

$$2.5 + 0.8(10) =$$

$$= 2.5 + 8 = 10.5$$

10.5 الإنتاج المتوقع للعام 1432.



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

تطبيقات:

مثال : فيما يلي بيانات عن كمية البروتين اليومي بالجرام التي يحتاجها العجل الرضيع، ومقدار الزيادة في وزن العجل بالكيلو جرام، وذلك لعينة من العجول الرضيعة حجمها 10.

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

والمطلوب :

⊕ قدر معادلة انحدار الوزن y على كمية البروتين x.

⊕ فسر معادلة الانحدار.

⊕ ما هو مقدار الزيادة في الوزن عند إعطاء العجل 50 جرام من البروتين ؟

⊕ وما هو مقدار الخطأ العشوائي؟

⊕ استخراج فترة ثقة 95% لكل من معلمتي الانحدار β_0 و β_1 .

⊕ استخراج جدول تحليل التباين ANOVA للمعاملات.

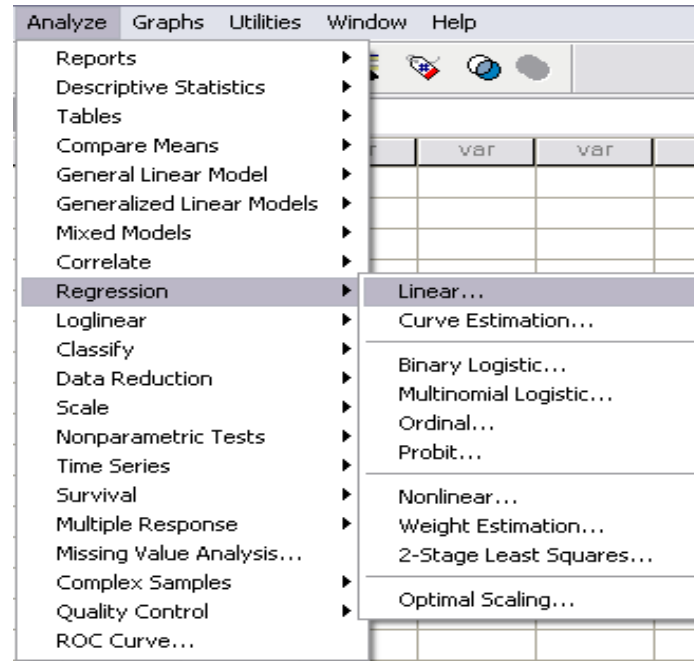
⊕ اختبر جودة توفيق النموذج الخطي (باستعمال معامل التحديد R^2) مع تحليل الأخطاء العشوائية بالرسم البياني.

⊕ اختبر التوزيع الطبيعي للأخطاء العشوائية بيانياً.

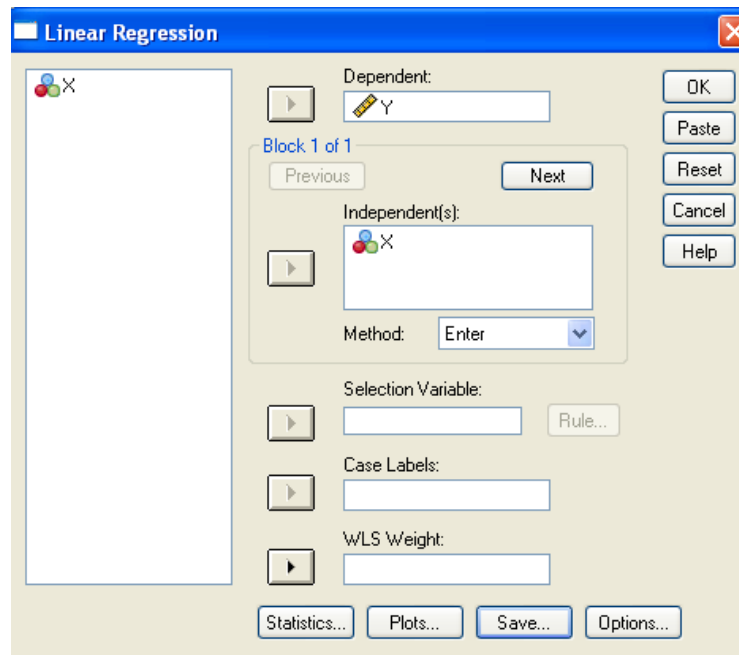
الحل:

لتنفيذ المطلوب السابق باستخدام البرنامج نتبع الخطوات التالية:

Analyze ثم Linear < Regression < Linear كما في الشكل التالي:



فيظهر صندوق الحوار التالي:



الشكل (1)

ونجد أن الشكل التالي يتكون من:

Dependent: يمثل المتغير التابع.

Independent: وهو يمثل المتغير المستقل ويمكننا إدخال مجموعة من المتغيرات المستقلة هنا كما

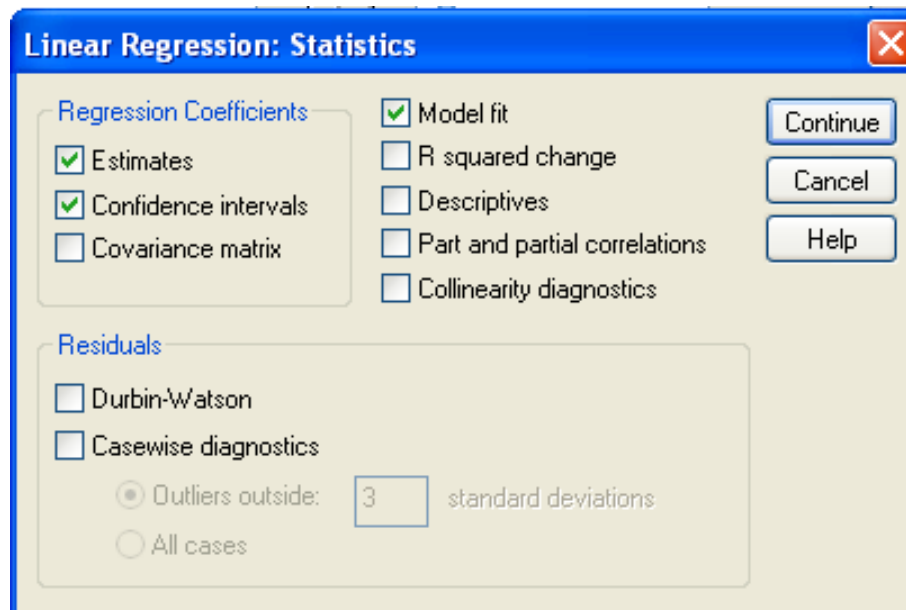
سوف نرى في مبحث الانحدار المتعدد ولأننا نريد معادلة انحدار ل y على X فإننا نعني أن y متغير تابع X متغير مستقل .

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.

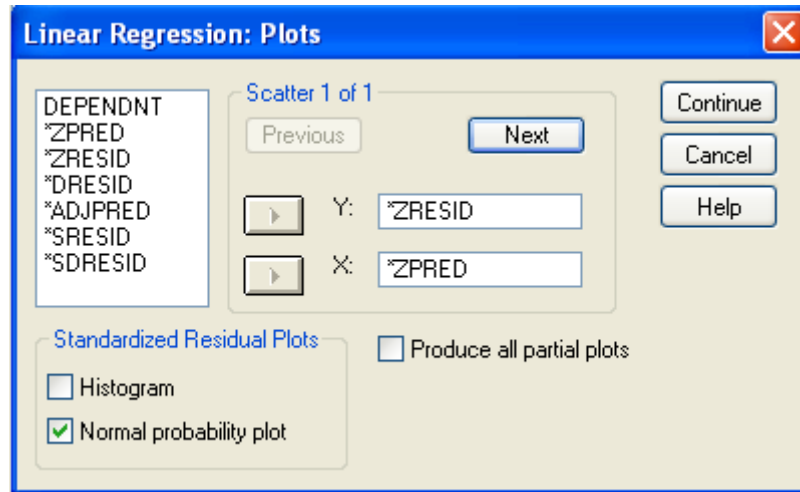
Case Labels: متغير يستخدم قيمة كعناوين لنقاط شكل الانتشار Scatterplots.

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



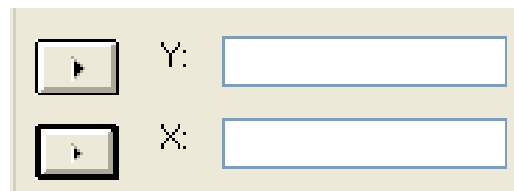
الشكل (2)

- وقد تم تأشير الخيارات التالية:
- ✓ Estimate: لتقدير معالم نموذج الانحدار واختبارات t المرافقة .
 - ✓ Confidence Interval: لتقدير فترة ثقة 95% لكل من معلمتي الانحدار.
 - ✓ Model Fit: لعرض R^2 و ANOVA .
 - ✓ وبعد الانتهاء النقر على زر Continue. وهناك رز آخر في شكل (1) بعنوان Plots وعند النقر عليه يظهر مربع الحوار الآتي:



الشكل (3)

ونلاحظ أننا اخترنا Normal probability plot: لاختبار التوزيع الطبيعي للأخطاء العشوائية المعيارية . كما أننا نجد أعلى شكل (3) الشكل الآتي:



في هذا الشكل يتم ادخال المتغيرات التي نود أن تظهر في شكل تحليل الأخطاء في المحور السيني والمحور الصادي والمتغيرات موجود في الجانب الأيمن من الشكل السابق ويوجد عدة خيارات:

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

Standardized residuals (*ZRESID): القيمة المعيارية للأخطاء.

Deleted residuals (*DRESID): حذف الأخطاء.

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

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

Studentized deleted residuals (*SDRESID): الأخطاء المحذوفة

والتي تتبع توزيع t .

وفي في تلك الحالة يتم تخصيص كلا من ZPRED لمحور السينات، ZRESID لمحور الصادات وبعد الانتهاء يتم النقر على زر Continues.

وهناك زر آخر في شكل (1) بعنوان Save وعند النقر عليه يظهر مربع الحوار الآتي:

الشكل (4)

ونلاحظ أننا اختارنا Unstandardized predicted Values أي $\hat{y}$ وكذلك Standardized Residual أي e_i سيتم استعمال هذين المتغيرين لرسم Scatterplots الجزء الثاني من المطلوب م ، علماً أنه سيتم إضافة هذين المتغيرين إلى ورقة Data Editor إلى جانب متغيرات X و Y و Observat حيث يضاف المتغير Unstandardized predicted Values باسم Pre_1 ويضاف Standardized Residual باسم Zre_1. وبعد الانتهاء النقر على زر Continues. وفي النهاية النقر على Ok في الشكل (1) لإظهار النتائج كالآتي:

Coefficients ^a								
		Unstandardized Coefficients		Standardized Coefficients			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

الشكل (5)

من خلال الجدول أعلاه يمكن كتابة النموذج كما يلي:

$$\hat{y} = 9.436 + 0.143x$$

وتشير المعادلة السابقة بأن زيادة البروتين مقدار جرام يعمل على زيادة الوزن بمقدار 0.143 جرام بينما توجد زيادة ليست متعلقة بالبروتين مقدارها 9.436. لكل جرام في الوزن وعليه فإن زيادة البروتين بمقدار 50 جرام يعمل على أن يكون

$$\begin{aligned}\hat{y} &= 9.436 + 0.143(50) \\ &= 16.436\end{aligned}$$

ومن ثم مقدار الخطأ العشوائي يساوي:

$$\begin{aligned}\hat{e}_{x=50} &= y_{x=50} - \hat{y}_{x=50} \\ &= 15 - 16.43 \\ &= -1.43\end{aligned}$$

ويستعمل اختبار t لاختبار الفرضية التالية لمعلمة الميل β_1 :

$$H_0 : \beta_1 = 0$$

$$H_A : \beta_1 \neq 0$$

ويستعمل اختبار t لاختبار الفرضية التالية 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 $P\text{-value} < 0.05$ نرفض فرضية العدم بمستوى 5%.

إذا كانت $P\text{-value} < 0.01$ نقبل فرضية العدم بمستوى 1%.

عكس هذا نقبل فرضية العدم.

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

ويمكن كتابة فترة الثقة للحد الثابت وذلك باستعمال نتائج الشكل (5) كما يلي:

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

حيث أن $\Pr$ تمثل الاحتمال كما يمكن كتابة فترة ثقة 95% لمعلمة الميل كما يلي:

$$\Pr[0.091 \leq \beta_1 \leq 0.195] = 95\%$$

الجدول التالي يعرف بجدول تحليل التباين ANOVA و يشتمل على إحصائية F لاختبار نفس الفرضية الخاصة بمعلمة الميل β_1 وهذا الاختبار مكافئ تماماً لاختبار t لمعلمة الميل (لاحظ أن قيمة P-value متساوية لكلا الاختبارين) علماً أن $F = t^2$.

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

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:

$$\begin{aligned}
 R^2 &= \frac{\text{Explained Variations}}{\text{Total Variations}} \\
 &= \frac{SSR}{SST} \\
 &= \frac{90}{108.00} \\
 &= 0.833333 \quad , \quad 0 \leq R^2 \leq 1
 \end{aligned}$$

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 تتضمنها 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 أن r معامل الارتباط الخطي البسيط لبيرسون وأن إشارة r هي نفس إشارة

معلمة الميل أي أن الإشارة موجبة وهذا يعني أن العلاقة بين مقدار البروتين والوزن علاقة طردية وقوية. يتصف معامل التحديد بأنه 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 يؤدي 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 الاعتبار النقصان الحاصل في درجات الحرية وقيمته دائماً أقل من قيمة معامل التحديد (الغير المصحح) وفي هذا المثال 81.3% ولهذا يمكن القول أن النموذج جيد التوفيق.

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

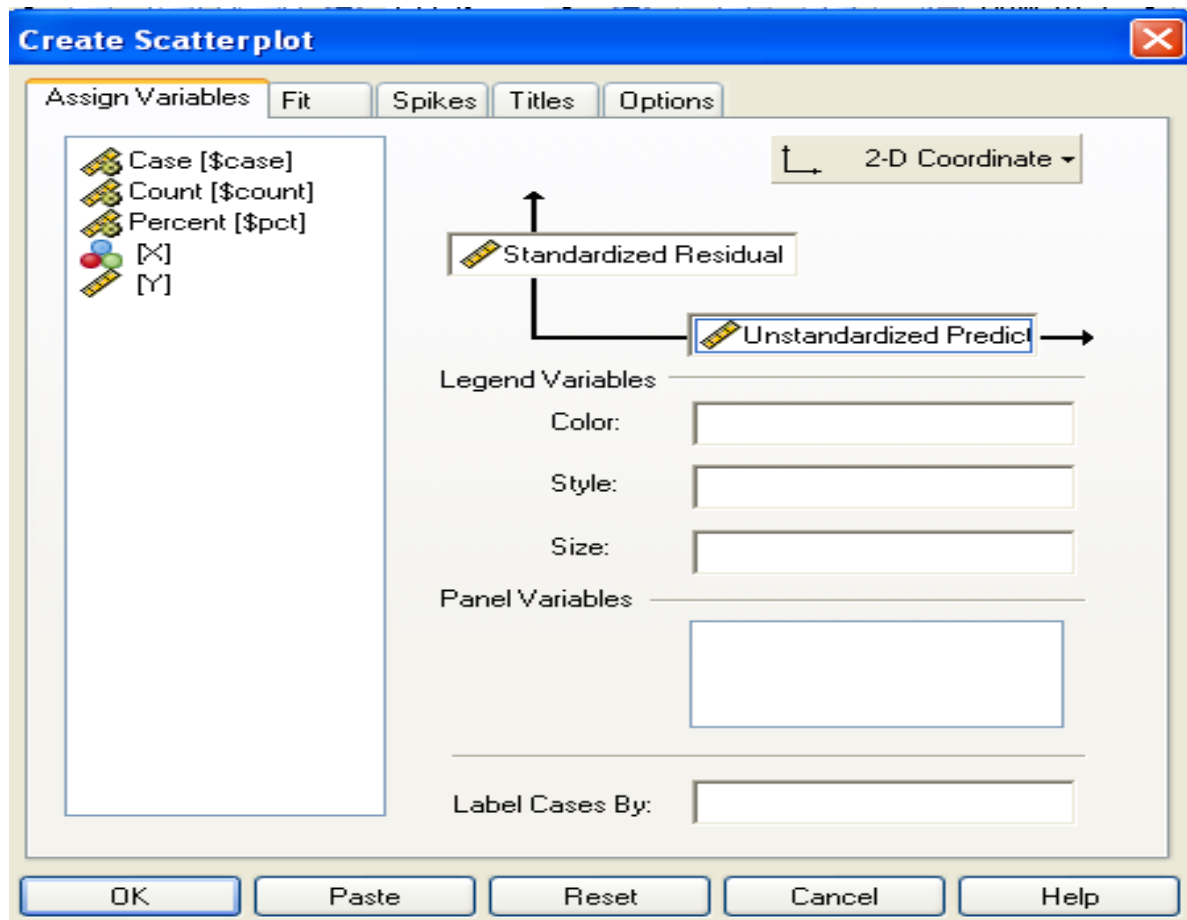
لتحليل الأخطاء العشوائية بيانياً نكون شكل الانتشار Scatterplots بتمثيل القيم التقديرية $\hat{y}$ على المحور الأفقي و الأخطاء المعيارية على المحور العمودي أو الراسي وكما يلي :

ففي شاشة Data Editor نجد أنه كما سبق أن أوضحنا أنه سوف يتم 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

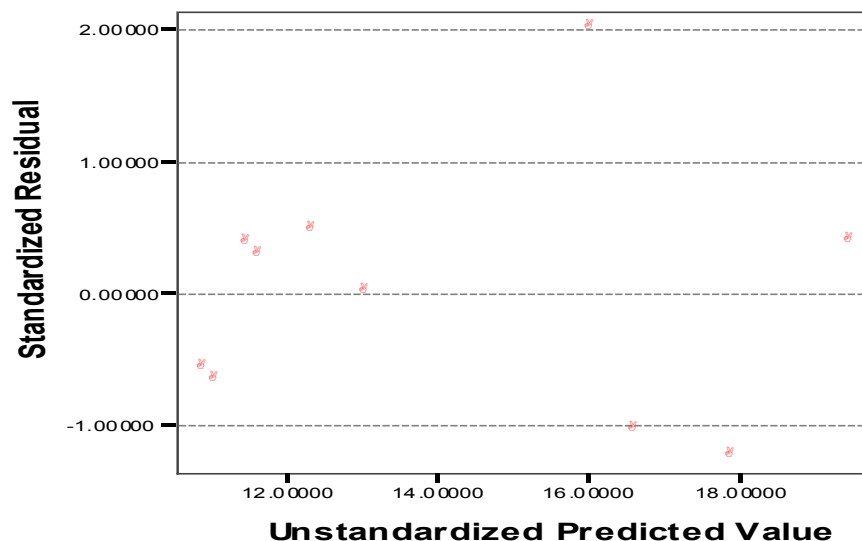
ثم بعد ذلك من شريط القوائم نختار Graphs < Interactive < Simple Scatter فيظهر صندوق حوار بعنوان Create Scatterplots :

فنقوم بنقر على Zre_1 لادخاله إلى الخانة Y في صندوق الحوار وبالمثل فنقوم بنقر على Pre_1 لادخاله إلى الخانة X في صندوق الحوار وبالمثل فيظهر مربع الحوار Create Scatterplots كما يلي:



الشكل (6)

فننقر على Ok فيظهر المخطط الآتي:



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

الاسبوع 22 - 24

الأرقام القياسية INDEX NUMBER

تعريف الرقم القياسي هو : هو رقم نسبي يستخدم لقياس التغير النسبي في أسعار أو كميات أو قيم ظاهرة ما أو عدة ظواهر من فترة زمنية لأخرى أو من مكان لأخر .

استخدامات الرقم القياسي:

- ١- يستخدم الرقم القياسي في دراسة الظواهر الاقتصادية مثل (أسعار السلع – الكميات المستهلكة منها – الصادرات – الواردات الخ.....)
- ٢- يمكن استخدام الرقم القياسي أيضا في دراسة الظواهر الاجتماعية والتربوية.
- ٣- كما تستخدم الأرقام القياسية للتعرف على الأحوال الاقتصادية ، بمقارنة الأرقام القياسية للأسعار بغيرها من الأرقام القياسية مثل أرقام الإنتاج ، والرقم القياسي لنفقة المعيشة وهكذا.
- ٤- يمكن استخدام الأرقام القياسية للتعرف على الاتجاه العام والتغيرات الموسمية لسلاسل الأرقام بعد تركيبها على مر السنين مثل : سلسلة أرقام الصادرات والواردات لسلع مختلفة.

تحديد سنة الأساس:

توجد عدد من العوامل يجب مراعاتها عند اختيار سنة الأساس:

عند حساب الرقم القياسي يجب اختيار سنة الأساس بحيث تكون متميزة بالآتي:

- ١- الاستقرار والبعد عن أي ظروف شاذة مثل : الأزمات الاقتصادية، والحروب ، والأوبئة.....
 - ٢- أن تكون قريبة نسبيا من فترة المقارنة حتى لا تختلف الظروف بين فترتي أساس والمقارنة وبالتالي يفقد الرقم القياسي أهميته في التعبير عن التغير في الظاهرة.
- تعريف الرقم القياسي للأسعار :
- يعرف الرقم القياسي للأسعار بأنه رقم نسبي يقيس التغير في أسعار سلعة واحدة أو أكثر بين سنتي الأساس والمقارنة.

الرموز المستخدمة عند حساب الأرقام القياسية للأسعار :

P_0 سعر السلعة في سنة الأساس

P_1 سعر السلعة في سنة المقارنة

Q_0 كمية السلعة في سنة الأساس

Q_1 كمية السلعة في سنة المقارنة

، سنة الأساس هي السنة الصغرى ، سنة المقارنة هي السنة الكبرى .

دلالة الرقم القياسي:

- ١- إذا كان الرقم القياسي اقل من 100 فذلك يدل على النقصان في الأسعار بنسبة الفرق بين الرقم القياسي الأقل من 100% و 100%
 - ٢- إذا كان الرقم 100 فذلك يدل على الزيادة في الأسعار بنسبة الزيادة في الرقم القياسي عن 100% .
- القياسي اكبر من

الأرقام القياسية للأسعار:

١- الرقم القياسي التجميع البسيط للأسعار:

$$I_s = \frac{\sum P_1}{\sum P_0} \times 100$$

٢- الرقم القياسي التجميعي للأسعار المرجح بكميات سنة الأساس (رقم لاسبير)

$$I_r = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times 100$$

٣- الرقم القياسي التجميعي للأسعار المرجح بكميات سنة المقارنة (رقم باش)

$$I_p = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} \times 100$$

٤- الرقم القياسي الأمثل للأسعار (رقم فيشر)

$$I_f = \sqrt{I_r \cdot I_p} = \sqrt{\text{باش} \times \text{لاسيبير}}$$

مثال :

يبين الجدول التالي الأسعار بالريالات و الكميات حسب المستهلكة من أربعة سلع استهلاكية للعامين 2000 م ، 2005 م

السلع	عام 2000		عام 2005	
	السعر	الكمية	السعر	الكمية
A	5	150	10	100
B	8	200	8	220
C	6	80	15	100
D	7	60	21	90

المطلوب حساب ما يلي :

١- الرقم القياسي التجميع البسيط للأسعار:

٢- الرقم القياسي التجميعي للأسعار المرجح بكميات سنة الأساس رقم لاسبير.

٣- الرقم القياسي التجميعي للأسعار المرجح بكميات سنة المقارنة رقم باش .

٤- الرقم القياسي الأمثل للأسعار رقم فيشر.

لحساب الأرقام المطلوبة يتم تكوين الجدول التالي :

باعتبار أن سنة الأساس هي سنة 2000 ، وسنة المقارنة هي سنة 2005

السلع	P0	Q0	P1	Q1	P0Q0	P0Q1	P1Q0	P1Q1
A	5	150	10	100	750	500	1500	1000
B	8	200	8	220	1600	1760	1600	1760
C	6	80	15	100	480	600	1200	1500
D	7	60	21	90	420	630	1260	1890
المجموع Σ	26	-	54	-	3250	3490	5560	6150

١- الرقم القياسي التجميع البسيط للأسعار:

$$I_s = \frac{\sum P_1}{\sum P_0} \times 100 = \frac{54}{26} \times 100 = 207.69\%$$

و هذا يدل أن المستوى العام لأسعار السلع الأربعة قد ارتفع في سنة 2005 بنسبة 107.69% و ذلك مقارنة بأسعار سنة 2000 .
٢ - الرقم القياسي التجميعي للأسعار المرجح بكميات سنة الأساس (رقم لاسبير) :

$$I_r = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times 100 = \frac{5560}{3250} \times 100 = 171.1\%$$

و هذا يدل أن المستوى العام لأسعار السلع الأربعة قد ارتفع في سنة 2005 بنسبة 71.1% و ذلك مقارنة بأسعار سنة 2000 .

٣ - الرقم القياسي التجميعي للأسعار المرجح بكميات سنة المقارنة (رقم باش):

$$I_p = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} \times 100 = \frac{6150}{3490} \times 100 = 176.2\%$$

و هذا يدل أن المستوى العام لأسعار السلع الأربعة قد ارتفع في سنة 2005 بنسبة 76.2% و ذلك مقارنة بأسعار سنة 2000 .

٤- الرقم القياسي الأمتل للأسعار (رقم فيشر)

$$I_f = \sqrt{171.1 \times 176.2} = 173.6\%$$

و هذا يدل أن المستوى العام لأسعار السلع الأربعة قد ارتفع في سنة 2005 بنسبة 73.6% و ذلك مقارنة بأسعار سنة 2000 .

بعض الملاحظات على الحل :

- لاحظنا أن مستوى الأسعار قد زاد لان الأرقام جميعها اكبر من 100% و لكن لو كانت الأرقام أقل من 100% سوف يكون هناك انخفاض في الأسعار .
- مثلاً لو كان الرقم القياس 75% معنى ذلك أن الأسعار انخفضت بمعدل 25% أي بمقدار الفرق بين 100% و 75% .

- يمكن حساب الأرقام القياسية السابقة للكميات بنفس طريقة الأرقام القياسية للأسعار.
- سمى رقم فيشر بالرقم القياسي الأمثل لأنه يتغلب على عيوب الأرقام القياسية الأخرى كما أنه يحقق اختبارات الأرقام القياسية من حيث الانعكاس في الزمن، والانعكاس في المعامل.
- يتم تكوين الرقم القياسي لنفقة المعيشة Standard of Living لقياس التغير في الأسعار التي يدفعها المستهلك النهائي لمجموعة من السلع والخدمات على مر الزمن. وذلك لأن الرقم القياسي للأسعار لا يوضح مدى التغير في نفقة المعيشة للفئات المختلفة من المجتمع.

مثال على ذلك :

المقارنة 1428	الأساس 1420
40	30

$$I = \frac{40}{30} \times 100 = 133\%$$

نوجد الرقم القياسي بتطبيق القانون التالي =

الرقم القياسي التجميعي البسيط :

$$I = \frac{\sum P_1}{\sum P_0} \times 100$$

وقانونه كالتالي :

مثال:

أسعار 1428	أسعار 1420
رمز سنة الأساس P_1	رمز سنة الأساس P_0
15	5
35	25
40	35
90	65

$$I = \frac{90}{65} \times 100 = 138\%$$

$$I = \frac{\sum P_1}{\sum P_0} \times 100 = \text{نطبق القانون التالي}$$

الرقم القياسي النسبي البسيط :

$$I = \frac{1}{n} \sum \left(\frac{p_1}{p_0} \right) \times 100$$

= قانون الرقم النسبي البسيط

$$I = \frac{1}{3} (3 + 1.4 + 1.14) \times 100$$

$$I = \frac{1}{3} \left(\frac{15}{5} + \frac{35}{25} + \frac{40}{35} \right) \times 100$$

= ولو طبقناه على المثال السابق

$$I = \frac{1}{3} (5.45) \times 100 = 1.84 \times 100 = 184$$

الأرقام المرجحة :

الرقم القياسي التجميعي المرجح (الاسبير) :

$$I = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times 100$$

وقانونه كالتالي :

مثال:

أسعار	أسعار	كميات	كميات	الوزن المرجح للأساس	الوزن المرجح للمقارنة	$P_1 \times Q_0$	$P_0 \times Q_0$
1420	1428	1420	1428	Q_0	Q_1		
5	15	20	60	$20 \div 120 = 0.17$	$60 \div 200 = 0.3$	2.55	0.85
25	35	40	90	$40 \div 120 = 0.33$	$90 \div 200 = 0.45$	11.55	8.25
35	40	60	50	$60 \div 120 = 0.5$	$50 \div 200 = 0.23$	20	17.5
65	90	120	200			34.1	26.6

$$I = \frac{34.1}{26.6} \times 100 = 1.28 \times 100 = 128$$

=

$$I = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times 100$$

نطبق القانون التالي

$$\sum \left(\frac{P_1}{P_0} \right) Q_0 \times 100 =$$

الرقم القياسي النسبي المرجح (لاسبير) : قانونه كالتالي :

أسعار 1420	أسعار 1428	كميات 1420	كميات 1428	الوزن المرجح للأساس	الوزن المرجح للمقارنة	$P_1 \times Q_0$	$P_0 \times Q_0$	$\frac{P_1}{P_0}$	$\left(\frac{P_1}{P_0} \right) Q_0$
p_0	p_1			Q_0	Q_1				
5	15	20	60	0.17	0.3	2.55	0.85	3	0.51
25	35	40	90	0.33	0.45	11.55	8.25	1.4	0.46
35	40	60	50	0.5	0.25	20	17.5	1.14	0.57
65	90	120	200			34.1	26.6	5.54	1.54

$$\sum \left(\frac{P_1}{P_0} \right) Q_0 \times 100 =$$

ثم نقوم بتطبيق القانون التالي :

$$1.54 \times 100 = 154$$

$$I = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} \times 100$$

الرقم القياسي التجميعي المرجح (باش) : قانونه هو :

مثال :

أسعار 1420	أسعار 1428	كميات 1420	كميات 1428	الوزن المرجح للأساس	الوزن المرجح للمقارنة	$P_1 \times Q_1$	$P_0 \times Q_1$	$\frac{P_1}{P_0}$
p_0	p_1			Q_0	Q_1			
5	15	20	60	0.17	0.3	4.5	1.5	3
25	35	40	90	0.33	0.45	15.75	11.25	1.4

1.14	8.75	10	0.25	0.5	50	60	40	35
5.54	21.5	30.25			200	120	90	65

نعوض في القانون (باش)

$$I = \frac{30.25}{21.5} \times 100 = 1.82 \times 100 = 182 \quad I = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} \times 100$$

$$\sum \left(\frac{P_1}{P_0} \right) Q_1 \times 100 =$$

الرقم القياسي النسبي المرجح (باش) : قانونه هو :

$\left(\frac{P_1}{P_0} \right) Q_1$	$\frac{P_1}{P_0}$	$P_0 \times Q_1$	$P_1 \times Q_1$	الوزن المرجح للمقارنة	الوزن المرجح للأساس	كميات 1428	كميات 1420	أسعار 1428	أسعار 1420
				Q_1	Q_0			p_1	p_0
0.9	3	1.5	4.5	0.3	0.17	60	20	15	5
0.63	1.4	11.25	15.75	0.45	0.33	90	40	35	25
0.28	1.14	8.75	10	0.25	0.5	50	60	40	35
1.82	5.54	21.5	30.25			200	120	90	65

$$1.82 \times 100 = 182 = \sum \left(\frac{P_1}{P_0} \right) Q_1 \times 100 =$$

نعوض في القانون :

الاسبوع 25 - 26

اختبار ت T Test

- إحصاء الاختبار في هذه الحالة له التوزيع الاحتمالي والمسمى توزيع T (T Distribution) لذا يسمى هذا الاختبار باختبار ت T ويستخدم اختبار ت عندما يكون المجتمع (أو المجتمعات) لها التوزيع الطبيعي والتباين (التباينات) غير معلوم وأيضا حجم العينة (العينات) أقل من 30 لكن إذا زاد حجم العينة عن 30 سوف يكون الإحصاء الاختبار التوزيع الطبيعي المعياري وهو اختبار معلمي.
- اختبار T يتوقف على عدة شروط يجب توافرها قبل إجراء الاختبار وهي
1. يجب ان تكون وحدة القياس المقاسة بها البيانات بفترة على الأقل وهذا يعنى ان البيانات يجب ان تكون لمتغيرات مستمرة. فإذا كانت وحدة القياس اسميه او ترتيبيه فلا يطبق الاختبار
 2. العينة يجب ان تكون مختارة عشوائيا من المجتمع
 3. بيانات العينة او العينات الداخلة في الاختبار يجب ان تتوزع حسب التوزيع الطبيعي
- الشرط الأول والثاني يجب ان يتحقق الباحث منهما لكن الشرط الثالث يمكن التحقق منه عن طريق الأمر Frequencies أو الأمر Explore كما تم سابقا
- وسوف نهتم هنا بالشرح والتوضيح لاستخدامات اختبارات T والذي يستخدم في عدد من الاختبارات منها:
1. هل متوسط عينه واحده يساوى متوسط مجتمع.
 2. اختبار متوسط عينتين مستقلتين Independent Samples ويكون المطلوب اختبار هل العينتين مسحوبتان من نفس المجتمع ام لا؟
 3. اختبار الفرق بين متوسطي عينتين غير مستقلتين.

الاختبارات الاحصائية لعينة واحدة One Sample Test

بفرض ان لدينا عينة من 20 شخص وتم قياس وزن كل شخص بالكيلوجرام وكانت لدينا البيانات التالية:

100 130 55 85 75 70 80 90 100 165 70 90 100 60 90 110 120 80 75 65

بفرض اننا نريد اختبار الفرض العدمي: متوسط الوزن في المجتمع المسحوب منه العينة يساوى 100 كجم

الفرض البديل: متوسط المجتمع لا يساوى 100 كجم

الحل:

أولاً: بوضع هذه البيانات في ملف وليكن weight.sav

ثانياً: يجب اختبار هل هذه البيانات تتبع التوزيع الطبيعي ام لا وذلك باستخدام الأمر Frequencies أو Explore كما سبق

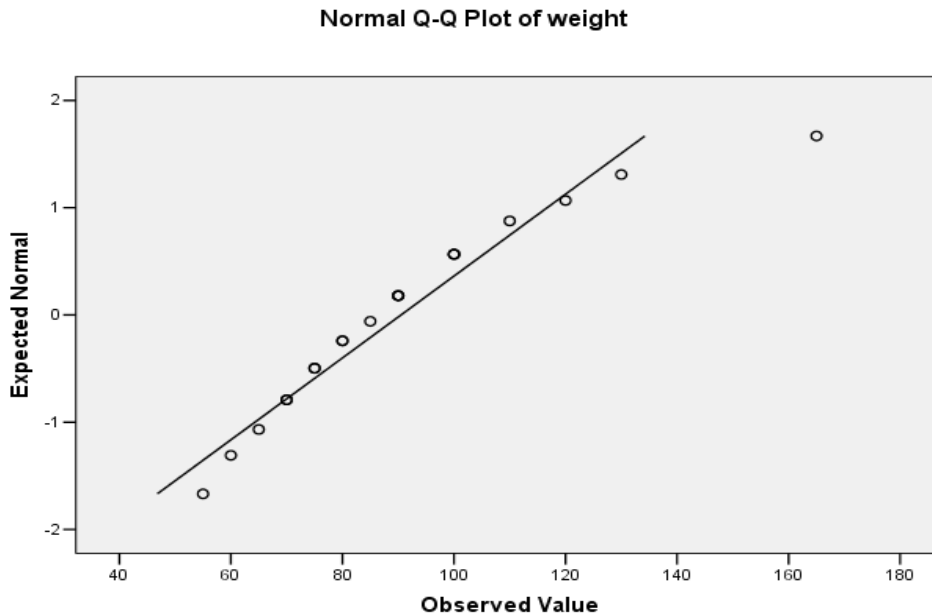
من اختبار Test of Normality نجد ان قيمة Sig. اكبر من 0.05 لذا سوف نقبل فرض العدم وهو ان البيانات لها التوزيع الطبيعي

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
weight	.158	20	.200*	.913	20	.072

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

a. Lilliefors Significance Correction



من الرسم البياني نجد ان البيانات تتركز حول الخط المستقيم لذا فأنها تتوزع تبعا للتوزيع الطبيعي

ثالثا: بما ان حجم العينة 20 أي اقل من 30 لذا فان احصاء الاختبار هو T بمعنى اننا سوف نستخدم اختبار T وسنجرى الاختبار كالتالي
١ - الفروض الاحصائية

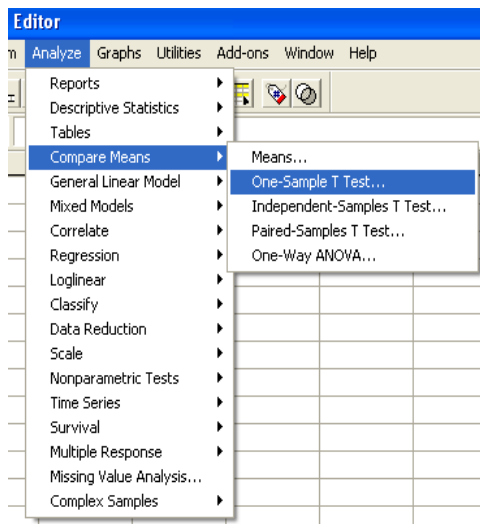
فرض العدم Null Hypothesis سيكون له الشكل التالي:

الفرض البديل Alternative Hypothesis سيكون له الشكل التالي: $H_0: \mu \neq 100$

$$T = \frac{\bar{X} - \mu_0}{S / \sqrt{n}}$$

احصاء الاختبار هو

وبتعيين الوسط الحسابي والانحراف المعياري للعينة يمكن حساب قيمة احصاء الاختبار T ومن جدول توزيع T يمكن حساب القيمة الحرجة $t_{\alpha/2} = t(n-1, \alpha/2)$ فاذا كانت |T| أكبر من القيمة الحرجة نرفض فرض العدم ونقبل الفرض البديل لكن اذا كانت أقل سوف نقبل فرض العدم ونرفض الفرض البديل



ويمكن اجراء هذا الاختبار عن طريق حزمة SPSS بسهولة كما يلي:

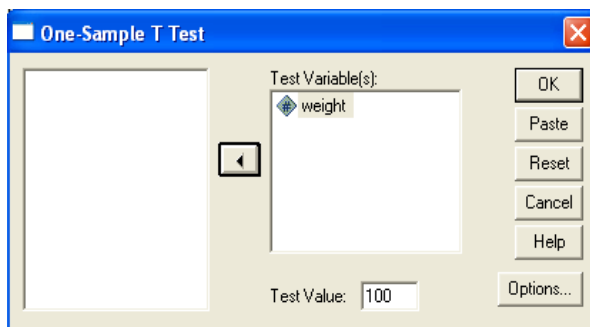
١- بتخزين البيانات في ملف Weight.sav

٢- من قائمة Analyze نختار Compare Means

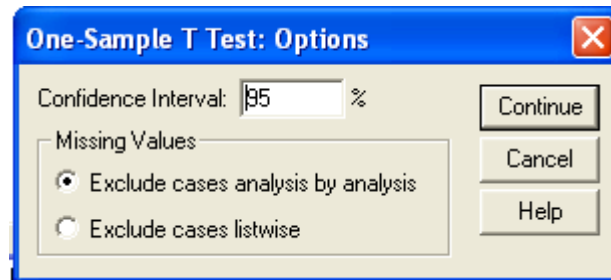
٣- من القائمة الفرعية نختار One Sample T Test

٤- تظهر شاشة جديده بعنوان One-Sample T test ننقل المتغير Weight لقائمة Test Variable(s)

٥- في المستطيل Test Value: نكتب قيمة فرض العدم وهي 100



نختار الأمر Options تظهر شاشة جديده بعنوان One-Sample T Test: Options لاحظ انه في خانة Confidence interval القيمة 95% بمعنى ان مستوى المعنوية $\alpha=0.05$ ويمكن تغييرها ثم نختار Continue نعود للشاشة السابقة ثم نضغط Ok



تظهر النتائج التالية

الجدول الأول:

بمعنوان One-Sample Statistics وعرض عدد حالات المتغير والوسط الحسابي والانحراف المعياري وخطأ التقدير للوسط الحسابي.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
weight	20	90.5000	26.20265	5.85909

One-Sample Test

	Test Value = 100					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
weight	-1.621	19	.121	-9.50000	-21.7632	2.7632

الجدول الثاني: بعنوان One-Sample Test وعرض نتيجة الاختبار حيث يحتوى الجدول على

١. فرض العدم اعلى الجدول Test Value = 100

٢. واسم المتغير

٣. قيمه احصاء الاختبار T = -1.621

٤. درجة الحرية n-1 = 19

٥. القيمة الاحتمالية P-value وهى Sig. = 0.121

٦. الفرق بين وسط العينة وقيمة الفرض العدمي Mean Difference

95% فترة ثقة لوسط المجتمع المسحوب منه العينة

من الجدول الثاني يمكن اتخاذ قرار بناء على قيمة Sig.=0.121 حيث ان الاختبار ذو طرفين فأنا سوف نقارنها بالقيمة $\alpha/2 = 0.025$ وهى اكبر لذا سوف نقبل فرض العدم.

يمكن للمستخدم ان يستخدم القيمة الموجبة لإحصاء الاختبار |T| ومقارنتها بالقيمة الحرجة والتي نحصل عليها من جدول T عند مستوى معنويه $\alpha/2 = 0.025$ ودرجة حرية n-1 = 19 ومن الملاحظ ان

$$t_{\alpha/2} = t(0.025, 19) = 2.262, \quad |T| = 1.621$$

وبالتالي فان

$$|T| < t_{\alpha/2}$$

لذا سوف نقبل فرض العدم وهو ان وسط المجتمع المسحوب منه العينة يساوى 100 كجم ويفضل دائما عند التعامل مع الحزمة ان نستخدم Sig. لاتخاذ القرار وليست احصاء الاختبار

اختبار عينتين مستقلتين

بفرض ان لدينا عينتين مستقلتين ونهتم بمتغير معين فى كلا العينتين ونرغب فى اختبار ان متوسطى المجتمعين المسحوب منهما العينتين لهما نفس الوسط الحسابى ام لا لذا سوف تصاغ الفروض الاحصائية كالتالى:

١. اختبار ذو طرفين $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 \neq \mu_2$

٢. اختبار ذو طرف واحد

$$\begin{aligned} H_0 : \mu_1 = \mu_2 & , & H_1 : \mu_1 < (>) \mu_2 \\ \text{or } H_0 : \mu_1 < \mu_2 & , & H_1 : \mu_1 \geq \mu_2 \\ \text{or } H_0 : \mu_1 > \mu_2 & , & H_1 : \mu_1 \leq \mu_2 \end{aligned}$$

ولإجراء هذا الاختبار يجب توافر بعض الشروط

الشروط اللازمة لتوافرها

١. حجم العينات المسحوبة أقل من 30 لإمكانية استخدام اختبار T لكن إذا كانت أكبر من 30 سوف نستخدم Z
 ٢. يجب أن تكون العينات مستقلة
 ٣. يجب أن تكون المجتمعات المسحوب منها العينات متجانسة
 ٤. يجب أن تكون المجتمعات لها التوزيع الطبيعي
- ❖ الشرطين الأول والثاني يتأكد منهما الباحث لكن الشرطين الثالث والرابع يمكن التأكد منهما باستخدام حزمة SPSS

اختبار طبيعية البيانات تم إجرائه سابقا باستخدام الأمر Explore والأمر Frequencies اختبار التجانس ويعني أن تباين المجتمعين متساوي وستكون الفروض الإحصائية لها الشكل التالي

إذا تم قبول فرض العدم فهذا يعني أن هناك تجانس لذا سوف نستمر في الاختبار أما إذا تم رفض فرض العدم فيعني عدم وجود تجانس لذا لا يجوز إجراء الاختبار.

تمرين:

بفرض أن لدينا عينة من الطلاب وتم تسجيل درجاتهم في مقرر الإحصاء كالتالي

Sample 1	20	17	10	13	15	14	14		
Sample 2	19	15	17	10	3	8	19	10	16

والمطلوب معرفة هل هناك فرق بين مستوى التحصيل للمجموعتين أم لا؟

الحل:

١- من الواضح أن حجم العينات أقل من 30

٢- العينات مستقلة

٣- يجب اختبار هل المجتمعات لها التوزيع الطبيعي أم لا كما يلي

سنقوم بإدخال البيانات في متغيرين أحدهما يسمى Factor والآخر Data حيث تحتوي data على القراءات في العينة الأولى والمتغير factor يحتوي الرقم 1 إذا كانت القيمة في المتغير data من العينة الأولى والرقم 2 إذا كانت القيمة من العينة الثانية، كالتالي

	Factor	Data
1	1.00	20.00
2	1.00	17.00
3	1.00	10.00
4	1.00	13.00
5	1.00	15.00
6	1.00	14.00
7	1.00	14.00
8	2.00	19.00
9	2.00	15.00
10	2.00	17.00
11	2.00	10.00
12	2.00	3.00
13	2.00	8.00
14	2.00	19.00
15	2.00	10.00
16	2.00	16.00
17		

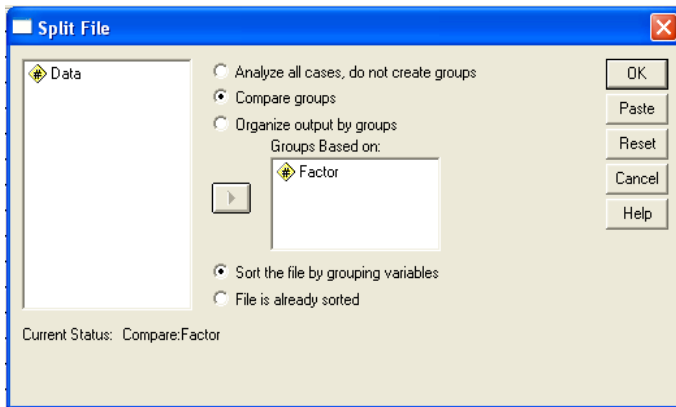
لأجراء اختبار الطبيعية على البيانات data سوف نقوم بالخطوتين التاليتين:

الخطوة الاولى:

١- من قائمة Data نختار Split File

٢- تظهر الشاشة التالية نختار Compare groups وننقل المتغير Factor لخانة Groups Based on

٣- نختار Ok سيتم تقسيم البيانات لمجموعتين تبعا للمتغير Factor وهو تقسيم غير ظاهري

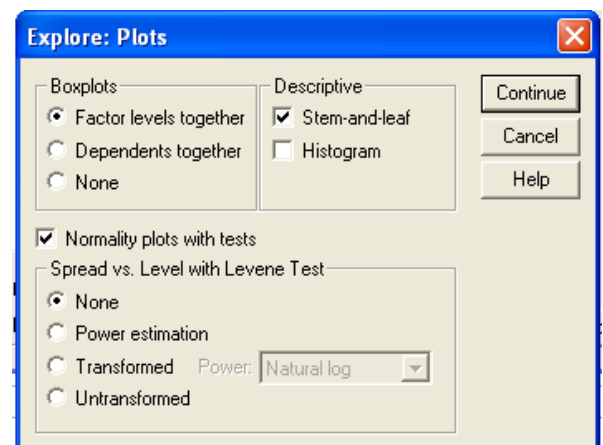
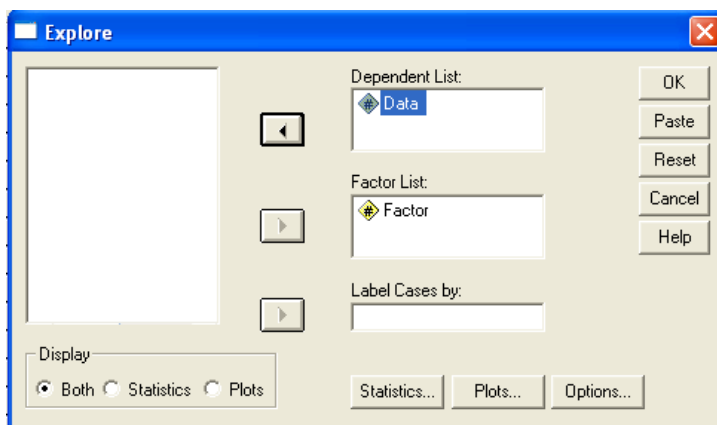


الخطوة الثانية:

١- من قائمة Analyze نختار Descriptive Statistics ثم نضغط Explore

٢- تظهر الشاشة التالية ننقل المتغير Data لخانة Dependent List وننقل المتغير Factor لخانة Factor List

٣- نختار Plots ثم نحدد الاختيار Normality plots with tests سوف نختار Continue نعود للشاشة السابقة ثم نختار Ok تظهر النتائج التالية



Tests of Normality

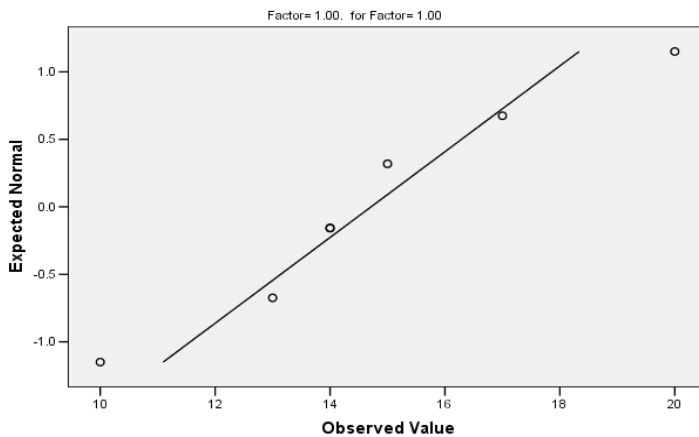
			Factor	
			1.00	2.00
Statistic	Kolmogorov-Smirnov ^a	Data	.178	.197
	Shapiro-Wilk	Data	.968	.915
df	Kolmogorov-Smirnov ^a	Data	7	9
	Shapiro-Wilk	Data	7	9
Sig.	Kolmogorov-Smirnov ^a	Data	.200*	.200*
	Shapiro-Wilk	Data	.883	.354

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

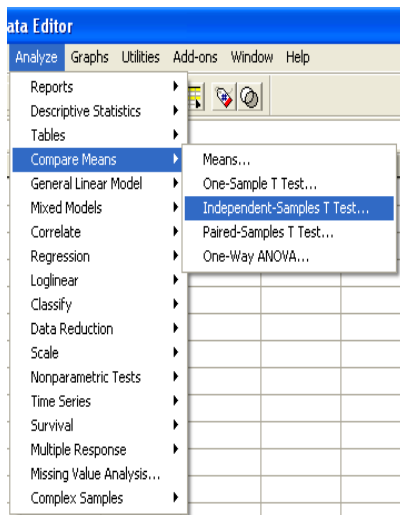
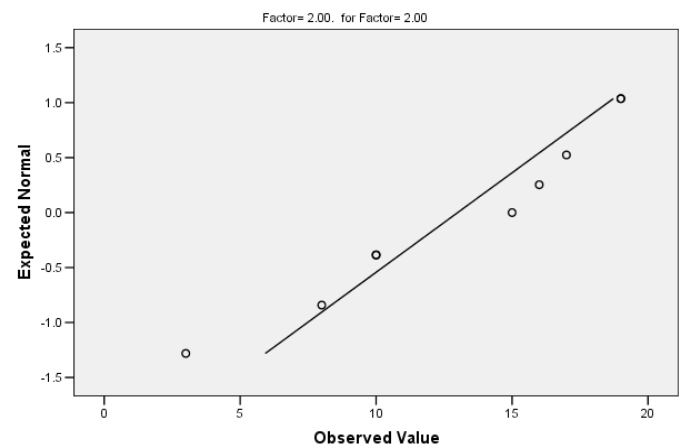
a. Lilliefors Significance Correction

نجد من جدول Tests of Normality أن قيمة Sig. في جميع الحالات اكبر من 0.05 لذا سوف نقبل فرض العدم القائل أن البيانات لها التوزيع الطبيعي ويتضح ذلك أيضا من الرسم البياني التالي

Normal Q-Q Plot of Data



Normal Q-Q Plot of Data



يجب إلغاء التقسيم لبيانات الملف indp_samples قبل إجراء أى عملية أخرى على البيانات.

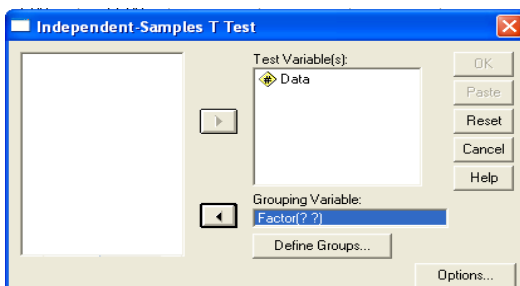
٤- هل المجتمعات المختار منها العينات متجانسة؟

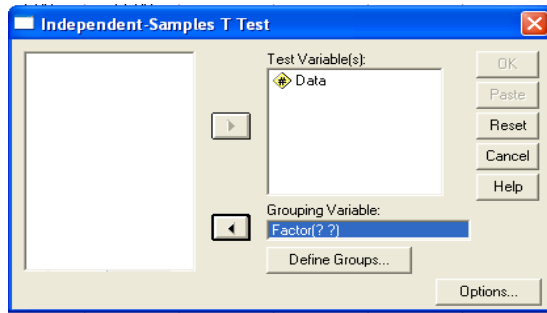
١- من قائمة Analyze نختار Compare Means

٢- من القائمة المنسدلة نختار Independent Sample T Test

٣- ننقل المتغير Data لخاصة Test variable(s)

٤- ننقل المتغير Factor لخاصة Grouping Variable:

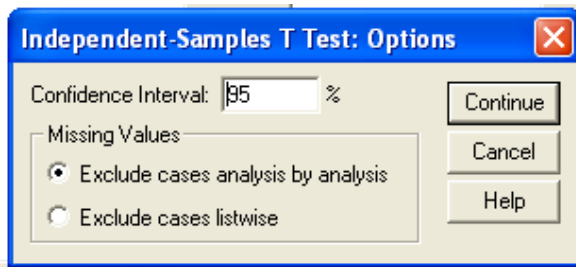
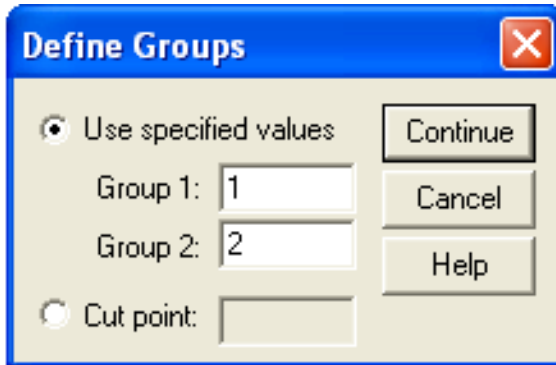




٥- نضغط على Define Groups لتحديد المجموعات

٦- تظهر شاشة جديدة تكتب الرقم 1 المميز للمجموعة الأولى

والرقم 2 المميز للمجموعة الثانية ثم نختار Continue



٧- بالضغط على Option سوف تظهر شاشة جديدة نحدد فيها قيمة فترة الثقة $100\%(1-\alpha)$ وستكون باستمرار 95% أي أن مستوى المعنوية $\alpha = 0.05$ وايضا كيفية التعامل مع القيم المفقودة.

٨- بالضغط على Continue ثم Ok تظهر النتائج التالية

الجدول الأول:

Group Statistics					
	Factor	N	Mean	Std. Deviation	Std. Error Mean
Data	1.00	7	14.7143	3.14718	1.18952
	2.00	9	13.0000	5.52268	1.84089

يحتوى على حجم العينات والوسط والانحراف المعياري والخطأ المعياري لكل عينة

الجدول الثاني: يحتوى على اختباري التجانس واختبار T

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
Data	Equal variances assumed	4.481	.053	.731	14	.477	1.71429	2.34614	-3.31768	6.74626
	Equal variances not assumed			.782	13.043	.448	1.71429	2.19177	-3.01915	6.44772

١. العمود الأول يحتوى اسم المتغير Data

٢. العمود الثاني والثالث يسارا لأجراء اختبار التجانس وحيث أن قيمة $\text{Sig.} = 0.053$ فهي أكبر من 0.05 لذا سوف نقبل فرض العدم وهو تجانس المجتمعين

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
Data	Equal variances assumed	4.481	.053	.731	14	.477	1.71429	2.34614	-3.31768	6.74626
	Equal variances not assumed			.782	13.043	.448	1.71429	2.19177	-3.01915	6.44772

٣. العمود الرابع والخامس والسادس لأجراء اختبار T وحيث ان المجتمعات متجانسه سوف نهتم بالصف الأول ومن العمود السادس Sig. = 0.477 وهى اكبر من 0.025 لذا سوف نقبل فرض العدم وهو ان وسطى المجتمعين متساوي أي لا يوجد فرق بين مستوى الطلاب في المجموعتين.
٤. الأعمدة الاخيرة تقدم فترة الثقة للفرق بين متوسطي المجتمعين.

اختبار عينتين غير مستقلتين

يستخدم ذلك الاختبار عندما يكون لدينا عينتين غير مستقلتين بمعنى ان لدينا عينة واحده ولكل مفردة من مفردات العينة قراءتين، القراءة الاولى تمثل العينة الأولى والقراءة الثانية تمثل العينة الثانية.

ويجب قبل اجراء الاختبار التحقق من الشروط التالية

- ١- العينات غير مستقله
 - ٢- عدد المفردات اقل من 30
 - ٣- العينات لها التوزيع الطبيعي
- اذا كان عدد المفردات اكبر من 30 فأننا لن نهتم بالشروط الثالث.

تمرين:

بفرض ان لدينا عينه مكونه من 10 اشخاص تم قياس ضغط الدم لكل شخص في العينة ثم بعد فترة زمنية تم قياس ضغط الدم مره ثانية لنفس الأشخاص وكانت البيانات كالتالي

Id	1	2	3	4	5	6	7	8	9	10
Before	130	140	150	130	140	145	135	110	120	150
After	120	140	130	150	130	110	110	120	110	140

هل يوجد فرق معنوي بين ضغط الدم قبل وبعد تلك الفترة الزمنية.

الحل: للإجابة على هذا السؤال يجب اختبار الفروض الاحصائية التالية

$$H_0 : \mu_d = 0, H_1 : \mu_d \neq 0, \mu_d = \mu_{Before} - \mu_{After}$$

وبإدخال البيانات في متغيرين Before للقراءات قبل الفترة الزمنية والمتغير After للقراءات بعد الفترة الزمنية. ثم التحقق من بعض الشروط

	Before	After
1	130.00	120.00
2	140.00	140.00
3	150.00	130.00
4	130.00	150.00
5	140.00	130.00
6	145.00	110.00
7	135.00	110.00
8	110.00	120.00
9	120.00	110.00
10	150.00	140.00
11		

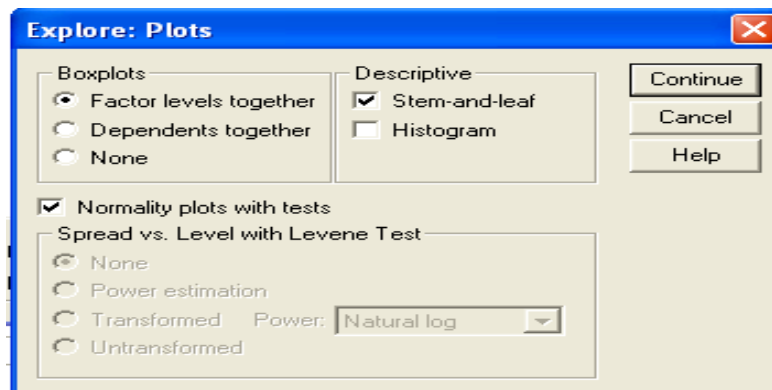
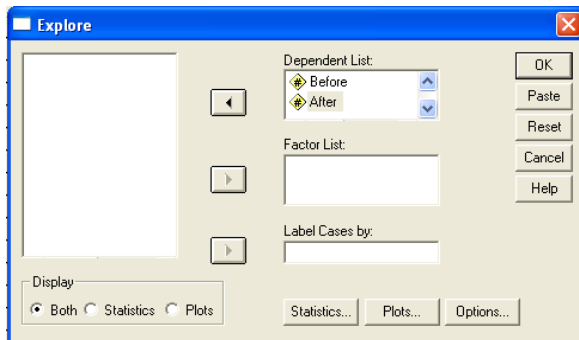
١- القراءات قبل الفترة الزمنية تعتبر عينه من مجتمع القراءات قبل الفترة والقراءات بعد تلك الفترة هي عينه اخرى من مجتمع القراءات بعد الفترة الزمنية وتلك العينات غير مستقله لأنها لنفس الأشخاص.

٢- حجم العينات 10 أقل من 30

٣- يمكن اختبار هل العينات لها التوزيع الطبيعي ام لا؟

من قائمة Analyze نختار Descriptive Statistics ثم نختار Explore

ننقل المتغيرات Before وايضا After لكانه Dependent List



وبالضغط على Plots وبتحديد الاختيار Normality plots with tests سوف تظهر النتائج التالية

Tests of Normality

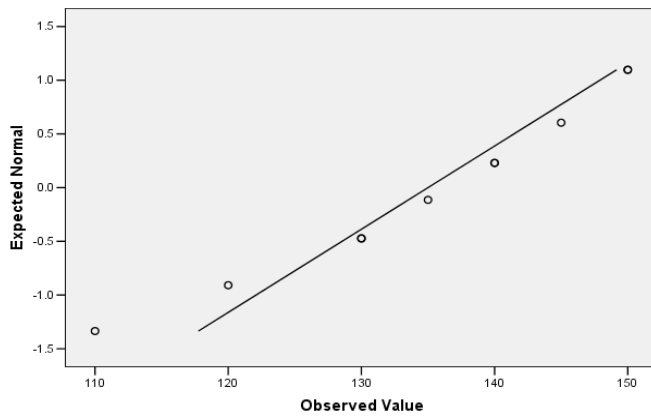
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Before	.151	10	.200*	.936	10	.505
After	.168	10	.200*	.908	10	.268

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

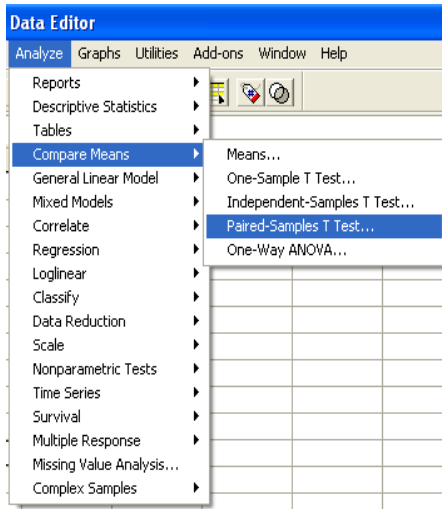
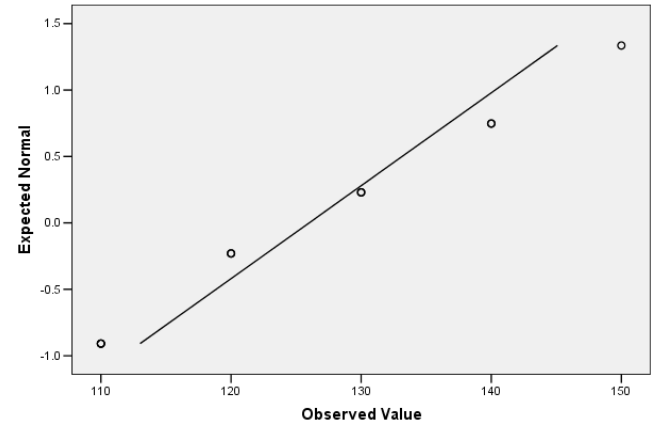
a. Lilliefors Significance Correction

جدول بعنوان Tests of Normality ومنه نجد ان البيانات لها التوزيع الطبيعي ونحصل على نفس النتيجة من الرسم البياني التالي

Normal Q-Q Plot of Before

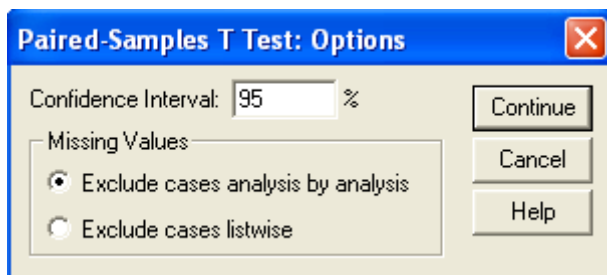
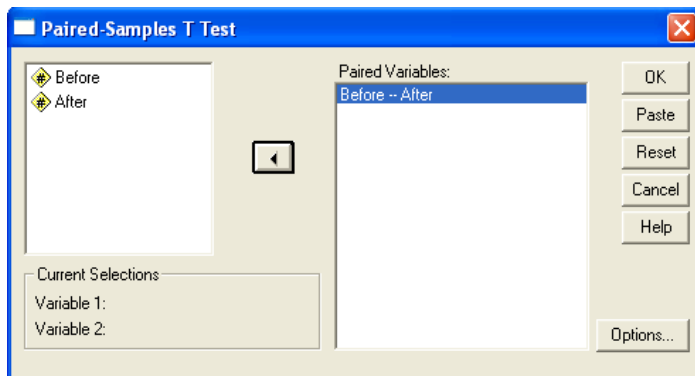


Normal Q-Q Plot of After



سبق نستطيع اجراء اختبار T كما يلي:

١. من قائمة Analyze نختار Compare Means
 ٢. من القائمة المنسدلة نختار Paired Samples t Test
 ٣. تظهر شاشة جديدة بعنوان Paired Samples T Test
- ننقل المتغيرين After, Before مع لقائمة Variables



٤. نضغط على الاختيار Option تظهر شاشة نحدد فيها مستوى المعنوية وكيفية التعامل مع القيم المفقودة.
٥. بالضغط على Continue نعود للشاشة السابقة نختار Ok تظهر النتائج

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Before	135.0000	10	12.90994	4.08248
	After	126.0000	10	14.29841	4.52155

الجدول الأول: Paired Samples Statistics بعنوان

يحتوى على الوسط الحسابى وعدد القيم والانحراف المعياري والخطأ المعياري لكل عينه

الجدول الثاني: Paired Samples Correlations بعنوان

يحتوى على عدد القيم ومعامل الارتباط بين المتغيرين وايضا قيمة $\text{Sig.} = 0.398$ لاختبار معنوية معامل الارتباط وقيمتها اكبر من 0.05 لذا سوف نقبل فرض العدم وهو عدم وجود ارتباط بين عناصر المجتمعين (لا يوجد معنوية للارتباط).

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 Before & After	10	.301	.398

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 Before - After	9.00000	16.12452	5.09902	-2.53478	20.53478	1.765	9	.111

الجدول الثالث: Paired Samples Test بعنوان

- يحتوى العمود الأول على اسم المتغير الجديد وهو الفرق بين القراءات قبل وبعد الفترة الزمنية ($d = \text{Before} - \text{After}$)
- يحتوى العمود الثاني على قيمة الوسط الحسابي للفروق بين القراءات
- العمود الثالث يحتوى على الانحراف المعياري للفروق بين القراءات قبل وبعد الفترة الزمنية

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 Before - After	9.00000	16.12452	5.09902	-2.53478	20.53478	1.765	9	.111

٤. العمود الرابع يحتوى الخطأ المعياري للفروق
٥. العمود الخامس يحتوى على فترة 95% ثقه للفرق بين متوسطي المجتمعين
٦. العمود السادس يحتوى على قيمة احصاء الاختبار T ودرجة الحرية في العمود السابع ($n-1=9$)
٧. العمود الاخير يحتوى على $\text{Sig.} = 0.111$ لاختبار T وهى اكبر من $\alpha/2=0.025$ لذا سوف نقبل فرض العدم : انه متوسطي المجتمعين متساوي ونرفض الفرض البديل: متوسطي المجتمعين مختلف.