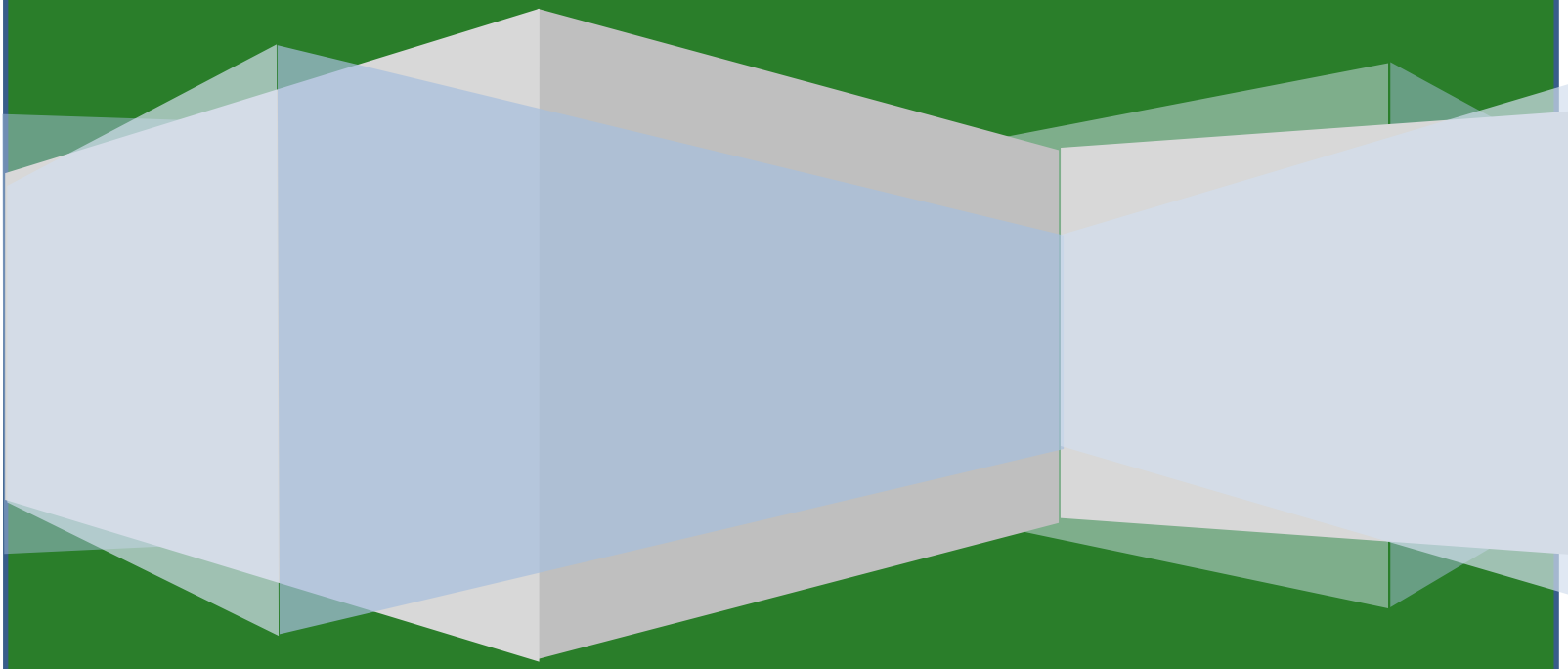
OIC ACCREDITATION CERTIFICATION PROGRAMME FOR OFFICIAL STATISTICS

DATA ANALYSIS

TEXTBOOK



ORGANISATION OF ISLAMIC COOPERATION

STATISTICAL ECONOMIC AND SOCIAL RESEARCH
AND TRAINING CENTRE FOR ISLAMIC COUNTRIES

OIC ACCREDITATION CERTIFICATION PROGRAMME FOR OFFICIAL STATISTICS

DATA ANALYSIS

TEXTBOOK

{{EMAN, BNYMFAREJ}}

ORGANISATION OF ISLAMIC COOPERATION

STATISTICAL ECONOMIC AND SOCIAL RESEARCH
AND TRAINING CENTRE FOR ISLAMIC COUNTRIES

© 2015 The Statistical, Economic and Social Research and Training Centre for Islamic Countries (SESRIC)

Kudüs Cad. No: 9, Diplomatik Site, 06450 Oran, Ankara – Turkey
Telephone +90 – 312 – 468 6172
Internet www.sesric.org
E-mail statistics@sesric.org

The material presented in this publication is copyrighted. The authors give the permission to view, copy download, and print the material presented that these materials are not going to be reused, on whatsoever condition, for commercial purposes. For permission to reproduce or reprint any part of this publication, please send a request with complete information to the Publication Department of SESRIC.

All queries on rights and licenses should be addressed to the Statistics Department, SESRIC, at the aforementioned address.

DISCLAIMER: Any views or opinions presented in this document are solely those of the author(s) and do not reflect the views of SESRIC.

ISBN: xxx-xxx-xxxx-xx-x

Cover design by Publication Department, SESRIC.

For additional information, contact Statistics Department, SESRIC.

قائمة المحتويات

قائمة المحتويات i

الاختصارات iii

شكر وتقدير iii

الفصل 1- مقدمة 1

1.1 مفهوم تحليل البيانات 1

2.1 أهمية تحليل البيانات 2

3.1 خطوات التحليل الاحصائي 3

4.1 أنواع التحليل الاحصائي 3

تمارين 4

الفصل 2- طرق بناء المؤشرات الاحصائية (التحليل الوصفي) 5

1.2 المتغيرات الاحصائية وأنواعها 6

2.2 تعريف المؤشر الاحصائي 8

3.2 طرق عرض المؤشرات 9

4.2 مقاييس النزعة المركزية Measures of Central Tendency 11

5.2 مقاييس التشتت 19

تمارين 23

الفصل 3- الاستقراء (التحليل الاستدلالي) 24

1.1.3 معامل ارتباط بيرسون 26

2.1.3 معامل ارتباط سبيرمان 26

2.3 تقدير دالة الانحدار 27

1.2.3 تقدير النماذج وتحديد السناريوهات 29

3.3 اختبار الفرضيات Test of Hypotheses 32

1.3.3 تعريف الفرضية statistical hypothesis 32

2.3.3 أنواع الفرضيات.....32
3.3.3 خطوات اختبار الفرضية الإحصائية.....34
4.3 أنواع الاختبارات الاحصائية.....40
1.4.3 الاختبارات المعلمية 40
2.4.3 الاختبارات اللامعلمية:.....40
4.3 الشروط المرجعية لتحديد نوع الاختبار المطلوب 42
تمارين.....43
الفصل 4- البرامج المستخدمة في تحليل البيانات 44
1.4 برنامج R 44
2.4 برنامج SAS.....45
3.4 برنامج STATA.....46
4.4 برنامج SPSS.....46
1.4.4 تطبيقات عملية على برنامج SPSS 46
تمارين.....64
المصطلحات.....65
قائمة القراءات 68
قائمة المصادر.....77

الاختصارات

r: Correlation Coefficient

SPSS : Statistical Package for Social Sciences

SAS: Statistical Analysis System

$\bar{x}$: Mean

S : Standard Deviation

s^2: Variance

شكر وتقدير

اشترك في اعداد هذا الكتاب دائرة الاحصاءات العامة (DoS) في عمان- الأردن ومركز الأبحاث الإحصائية والاقتصادية والاجتماعية

والتدريب للدول الاسلامية (SESRIC) تحت اعتماد منظمة المؤتمر الاسلامي للبحوث الاقتصادية والاجتماعية لمخططات البرامج القطرية (OIC) وبدعم من مجموعة بنك التنمية الاسلامي (IBA)، هذا الكتاب في تحليل البيانات يغطي موضوعات متنوعة في مجال التحليل الاحصائي للبيانات بالاعتماد على مراجع عديدة في مجال تحليل البيانات.

بداية يود المؤلف أن يعبر عن عميق شكره وامتنانه لدائرة الاحصاءات العامة – الأردن، ممثلة بالمدير العام للدائرة د. قاسم الزعبي، كما يود الكاتب أن يؤكد شكره الكبير لـ د. عبدالله درادكة لدعمه الذي لا ينقطع والذي يقدمه المؤلف بشكل مستمر.

نهاية الشكر الكبير لمركز الأبحاث الإحصائية والاقتصادية والاجتماعية والتدريب للدول الاسلامية على هذه المبادرة العلمية والتي من شأنها رفع قدرات الباحثين في المجالات الإحصائية ورفد ينابيع العلم بكل ما هو جديد ومميز.

الفصل 1

مقدمة

يقدم هذا الفصل توطئة عامة من خلال تعريف مفهوم تحليل البيانات، والوقوف على اهمية عملية التحليل للبيانات، ومعرفة الاساليب الاحصائية المختلفة في عمليات التحليل، كما وسيتم استعراض خطوات التحليل الاحصائي، اضافة على التعرف على أنواع التحليل المختلفة.

الهدف العام للفصل

- شرح وتوضيح مفهوم تحليل البيانات والعلاقات التي يقودنا التحليل الى تفسيرها.
- معرفة أنواع تحليل البيانات.
- الوصول الى خطوات واضحة لتحليل البيانات.

1.1 مفهوم تحليل البيانات

يعد تحليل البيانات تلك العملية التي من خلالها يتم تجهيز البيانات باستخدام كافة الطرق سواء الرياضية أم المنطقية وذلك من أجل الوصول إلى معلومات مفيدة يمكن اتخاذ القرارات بناء عليها. ومن هنا فعملية تحليل البينات هي عملية مهمة جداً وفي كافة المجالات، فأي تقرير يتوجب على الإنسان إعداده وكتابته أو أية تجربة يتوجب عليه القيام بها أو قرار يتوجب عليه اتخاذه أو أي شيء آخر فإنه يحتاج إلى عملية تحليل للبيانات التي تم جمعها، أما المجالات التي تستخدم فيها هذه العملية الهامة فهي أيضاً متعددة، منها مجال الإحصاءات التي تقوم بها أجهزة الدولة أو المجالات العلمية الطبية أو الفيزيائية أو الكيميائية أو الهندسية أو الحيوية، بالإضافة إلى استخدام هذه العملية في مجالات تطبيقية كالصناعة والزراعة والتجارة والاقتصاد بشكل عام، وأيضاً في السياسة وفي مجالات الخدمة الإنسانية، وربيمكن ايضا أن تستخدم في المجالات الفنية وغيرها من المجالات المختلفة الأخرى، فكل هذه المجالات هي مجالات تستخدم فيها عملية تحليل المعلومات على أوسع مدى.

تبدأ عملية التحليل للبينات بتحديد نوعية البينات التي يجب أن يتم جمعها، وهذا يعتمد على طبيعة المجال وعلى طبيعة الجهة التي ستستفيد من هذه البينات التي يتم جمعها، ثم

تبدأ بعد ذلك عملية جمع البيانات وهي ليست بالعملية السهلة، بل هي عملية تتطلب الوقت والجهد والتخطيط، حيث يجب أن يتم أولاً تحديد المصادر التي ستجمع البيانات منها، ومن ثم ينبغي ترتيبها بعد جمعها بطريقة يسهل قراءتها وإجراء التحليلات اللازمة عليها. وبعد ترتيبها نقوم بإجراء العلميات الضرورية واللازمة من أجل تحويل هذه البيانات إلى معلومات نستطيع من خلالها أن نتخذ قرارات بناء عليها وهذا ما يعرف بمعالجة البيانات، ومن ثم نقوم بعمل تحسين وتعديل على هذه البيانات بحيث نقوم بمعالجة الخطأ بالطرق العلمية المناسبة أو بتقليل نسبة الخطأ إن وجدت، وبعد ذلك نقوم بعمل التحليل لهذه البيانات التي تمت معالجتها وتنقيتها من الأخطاء، وتتضمن عملية التحليل؛ استخدام الوسائل الإحصائية الرياضية المختلفة والمتعددة. وأخيراً نقوم بعرض هذه المعلومات التي تم استنتاجها بطرق العرض المناسبة. وقد سهلت التكنولوجيا الحديثة وخاصة الحاسوب عملية معالجة البيانات، وقد أصبحت عملية معالجة البيانات وعملية تحليل البيانات باستخدام الحاسوب أمر سريع جداً وسهل ويقلل من احتمالية حدوث الأخطاء وبنسبة كبيرة جدا، وهناك العديد من البرمجيات المستعملة في هذه المجال والتي قد تكون متخصصة في بعض الأحيان حسب المجال الذي ستستعمل فيه.

ويمكن تلخيص عملية تحليل البيانات على أنها تنظيم وترتيب البيانات؛ وذلك من أجل إخراجها وإبرازها على شكل معلومات يتم استخدامها بهدف الإجابة على أسئلة معينة، وتكون مرحلة تحليل البيانات بعد جمع المعلومات وتنظيمها بشكل مرتب لتسهيل تحليلها مثل؛ وضع الإجابات في جداول لعرضها وتحليلها، ويتم تحليل البيانات بشكل عام لعدّة أغراض نذكر منها:

- يزيد من قدرة الباحث في تفسير المتغيرات التي تؤثر في ظاهرة معينة، وهذا يعتمد على اختيار الأسلوب التحليلي المناسب.
- تسمح بالوقوف على مدى جوهر تأثير المتغيرات على الظاهرة.
- تمكن الباحث من تقدير البيانات المجتمعية من واقع البيانات للعينات الاحتمالية المأخوذة من المجتمع.

2.1 أهمية تحليل البيانات

يمكن تلخيص اهمية تحليل البيانات بالنقاط التالية:

- التنبؤ لفترات زمنية ماضية أو مستقبلية واتخاذ القرار عن طريق اختيار بديل من عدة بدائل متاحة.
- امكانية التحقق من صحة أو عدم صحة الفرض الإحصائي.
- الرقابة على جودة البيانات ومطابقتها للمواصفات والمقاييس.

3.1 خطوات التحليل الاحصائي

1. إدخال البيانات: وتأتي بعد عملية جمع المعلومات، بحيث يقوم الباحث بإدخال البيانات إلى الحاسوب باستخدام بعض البرامج مثل برنامج SPSS وبرنامج Excel، وهنا لا بدّ من مراعاة الدقة عند إدخال البيانات، وتجنب الخطأء المتعمد أو غير المتعمد والفهم غير الكافي لإدخال البيانات.

2. تشغيل البيانات: وهي عبارة عن حصر وعدّ عدد الحالات لكل متغيّر أو خاصية بحيث يكون الهدف من هذه العملية:

   - تحديد التوزيع المتكرر للمتغيرات التي تخضع للتحليل.
   - عمل بعض التحليلات الإحصائية البسيطة للبيانات بشكل عام.
   - التلخيص أو الوصفية للمتوسط الحسابي والنسب المئوية.

3. تحليل البيانات وتحويلها إلى معلومات مهمة ومفيدة؛ وتتمّ من أجل استنتاج المعلومات التي تساعد في الإجابة على الأسئلة التي تم تحديدها مسبقاً، وهذه البيانات يفضل أن يكون تنفيذها وتخطيطها بشكل جماعي لتنوع الأراء للحصول على تحليل دقيق.

4. تفسير وتحويل المعلومات إلى نتائج؛ هنا تعتمد على عملية ربط الحقائق أو الأمور التي حدّدت من خلال تحليل البيانات مع المؤشرات والغرض من تحليل البيانات، مع مراعات أن المعلومات التي تم الحصول عليها وجمعت تتحول إلى أدلة للإجابة على الأسئلة التي تم طرحها.

4.1 أنواع التحليل الاحصائي

ينقسم التنحليل الاحصائي الى نوعين:

التحليل الوصفي:Descriptive Statistics

وهو مجموعة الأساليب الإحصائية التي تعنى بجمع البيانات وتنظيمها وتصنيفها وتلخيصها وعرضها بطريقة واضحة في صورة جداول أو أشكال بيانية وحساب المقاييس الإحصائية المختلفة؛ لوصف متغير ما (أو أكثر من متغير) في مجتمع ما أو عينه منه.، مثل حساب الوسط الحسابي والوسيط والمنوال وحساب الجداول التكرارية والنسب المئوية، إضافة الى استخراج مقاييس التشتت مثل المدى والانحراف المعياري، كما ان بناء الجداول المتقاطعة يعتبر وصفا جيدا للبيانات يوضح ترابط البيانات وتقاطعاتها معا.

التحليل الاستدلالي: Statistics Inferential

هو مجموعة من الأساليب الإحصائية التي تستخدم بغرض تحليل بيانات ظاهرة (أو أكثر) في مجتمع ما على أساس بيانات عينة احتمالية تسحب منه، ومن ثم تفسيرها للتوصل إلى التنبؤ واتخاذ القرارات المناسبة مثل؛ اختبار الفرضيات الاحصائية والتي يمكن اجراؤها باستخدام عدة طرق سيتم شرحها في الفصول اللاحقة، إضافة الى ايجاد معاملات الارتباط للبيانات وتقدير وحساب دالة الانحدار البسيطة والمتعددة.

تمارين

أوجد حلول التمارين التالية

1. إن معرفة نوعية البيانات التي يجب تحليها تعتبر عملية غير ضرورية في مرحلة التجهيز لتحليل البيانات، حيث تختصر مهمة التحليل في وصف البيانات وتحديد العلاقات بينها. ناقش هذه الفكرة؟
2. التحليل الوصفي والتحليل الاستدلالي هما نوعين من التحليل يعتبر كل منهما مكملا للآخر، ولتحقيق نتائج دقيقة تصف البيانات الحالية وتعطي تنبوء للمستقبل يمكن الاعتماد على هذين النوعين من التحليل معا، وضح هذا مع ذكر أمثلة على كل نوع من انواع التحليل؟

3. اذكر اهم فوائد عملية تحليل البيانات وناقشها حسب رأيك الشخصي؟

الفصل 2

طرق بناء المؤشرات الاحصائية (التحليل الوصفي)

يتناول هذا الفصل طرق عرض المؤشرات الاحصائية والتي تعطي ملخصا لخصائص المتغيرات، إلا ان طرق العرض هذه قد تكون غير دقيقة، لذا كان لا بد من التعرف على طرق بناء المؤشرات العددية من خلال دراسة المقاييس المختلفة مثل؛ مقاييس النزعة المركزية ومقاييس التشتت، اضافة الى التمييز بين الأنواع المختلفة من المتغيرات.

الهدف العام للفصل

- التمييز بين الأنواع المختلفة من المتغيرات الاحصائية
- التعرف على طرق بناء المؤشرات الاحصائية المختلفة
- التمييز بين المقاييس الاحصائية المختلفة

1.2 المتغيرات الاحصائية وأنواعها

المتغيرات الاحصائية متعددة ومختلفة الأنواع، وقبل استعرض هذه الأنواع سنتناول بداية مفهومي البيانات والمعلومات والفرق بينهما.

- ***البيانات:*** تشير إلى الحقائق والأرقام والحروف والكلمات والإشارات التي تعبر عن فكرة ما أو موقف محدد كأرقام المبيعات وأرقام الإنتاج وأعداد العاملين .. الخ.
- ***المعلومات***: هي حقائق منظمة تفيد في اتخاذ القرارات وتشير إلى نتائج تشغيل البيانات التي تصف أحداث العمليات التي تقع في المنظمة وإخراجها بشكل له معنى للمستفيد منها مثل؛ معدل التضخم، نسبة المؤمنين صحياً ومعدل البطالة .. الخ.

الجدول التالي يوضح الفرق بين البيانات والمعلومات

البيانات	المعلومات
حقائق غير منظمة لا تفيد مباشرة في اتخاذ القرارات.	حقائق منظمة تفيد مباشرة في صنع القرارات.
بمثابة مادة خام في نظام المعلومات.	بمثابة المنتج في نظام المعلومات.
مدخلات لنظم المعلومات.	مخرجات لنظم المعلومات.

اذن هذه البيانات تنتظم على شكل متغيرات تختلف هذه المتغيرات حسب خصائص معينة تفرز أنواع مختلفة للمتغيرات نوضحها فيما يلي:

يمكن تقسيم المتغيرات إلى نوعين أساسيين:

- **متغيرات نوعية (وصفية) Qualitative Variables**

وهي التي يمكن حصرها في عدة أوجه وصفية حسب نوعها دون أن تأخذ قيما عددية ولا يمكن إجراء عمليات حسابية عليها، وتقسم إلى نوعين:

1. **بيانات نوعية إسمية (Nominal)**

تعتمد على التصنيف النوعي بغض النظر عن أهمية الترتيب. ولهذا فهي تفتقد إلى صفة الترتيب، وكمثال على ذلك متغير الجنس (ذكر، أنثى) والفرع الدراسي الذي يصنف إلى ادبي وعلمي و أقسام المستشفى وفصيلة الدم والجنسية .... الخ

2. **بيانات نوعية ترتيبية (Ordinal)**

يلعب الترتيب دوراً أساسياً في تحديد معالم الظاهرة، أي أن الترتيب مهم، ومثال ذلك الرتب العسكرية.

- **متغيرات كمية Quantitative Variables**

وهي المتغيرات التي يتم تمييزها بالدرجة أو التكرار أو الكمية، ويمكن الحصول عليها في شكل أعداد ويمكن ترتيبها، وكمثال على ذلك الرواتب الشهرية وعلامات الطلاب.

أقسام المتغيرات الكمية:

تقسم المتغيرات الكمية إلى قسمين هما:

1. **متغيرات كميّة منفصلة Discrete Variables**

يمكن عدّها وتدل القيمة "0" على عدم وجود الظاهرة وتكون محددة على مقياس معين، ومن أمثلة ذلك:

- عدد أفراد الأسرة
- عدد تلاميذ الصف
- عدد المستشفيات
- الدخل الشهري.

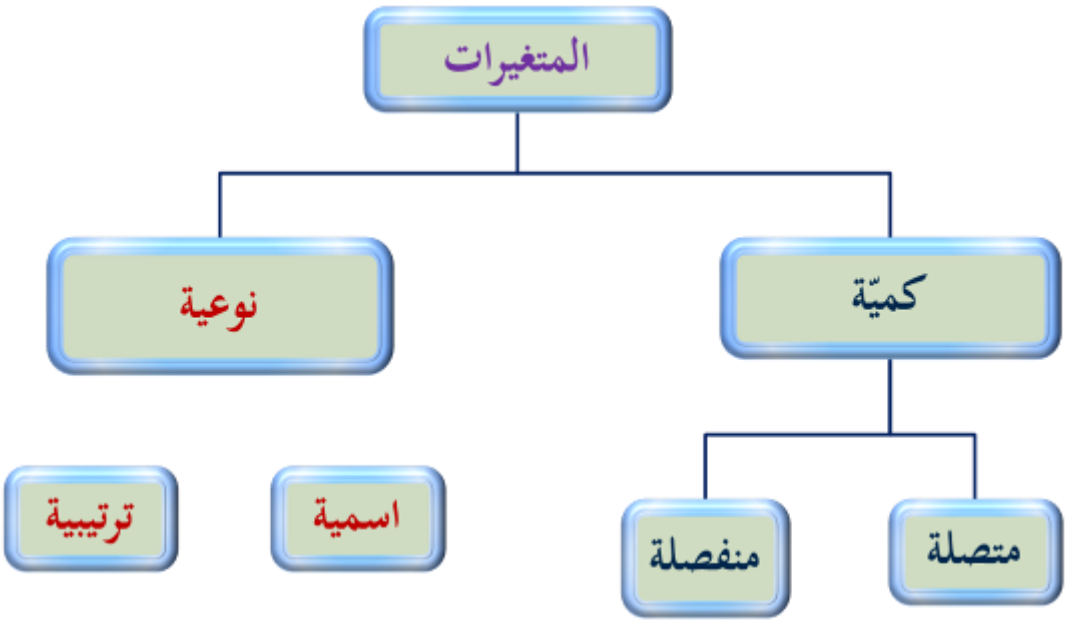
2. **متغيرات كميّة متصلة (مستمرة) Continuous Variables**

وهي بيانات رقمية تقاس بمقدار بعدها عن الصفر، أي أن للصفر دلالة على وجود الظاهرة، ولا يمكن عدّها، ويتم الحصول عليها عن طريق القياس، وتأخذ أي قيمة داخل مدى معين سواء كانت صحيحة أو كسر ومن أمثلة ذلك:

- درجة الحرارة: متغير كمي تقاس بياناته بمعيار بعدي، حيث أن درجة الحرارة "الصفر" ليس معناه انعدام الظاهرة، ولكنه يدل على وجود الظاهرة.
- درجة الطالب في الاختبار: متغير كمي يقاس بياناته بمعيار بعدي، حيث

حصول الطالب على الدرجة "0" لا يعني انعدم مستوى الطالب.

النموذج ادناه يوضح تصنيف المتغيرات وانواعها:



2.2 تعريف المؤشر الاحصائي

المؤشر: هو قيمة عددية أو كمية تعكس جوانب الظاهرة وتصف المجتمع، مثل الوسط الحسابي والانحراف المعياري. وفي الأجزاء اللاحقة من هذا الفصل سنبين وبشكل تفصيلي المؤشرات الاحصائية المختلفة.

الجدول التالي يوضح الفرق بين البيانات الاحصائية والمعلومات والمؤشرات

البيانات Data	المعلومات Information	المؤشرات Indicators
عبارة عن تجميع رقمي للاجابة عن تساؤل ما، وقد تكون هذه الأرقام في شكل	عبارة عن تجميع للبيانات في شكل له معنى او مفهوم خاص	هي تحويل البيانات الاحصائية المستخدمة من السجلات والمستندات الادارية من مادة خام الى مؤشرات لها

رقمي او اجمالي مثل الاعداد أو النسب في الجداول او الرسومات البيانية	يوضح العلاقات بين هذه البيانات لاستخدامها في تحديد المشكلة والتخطيط والمتابعة والتقييم والتقويم	جوانبها وابعادها المختلفة التي تساعد على التشخيص وتحديد المشكلات وبالتالي تساعد على التخطيط والقيام بأعمال المتابعة والتقييم والتقويم للأداء. والمؤشرات قد تكون رقم واحد او مجموعة ارقام

3.2 طرق عرض المؤشرات

من أجل إعطاء فكرة واضحة ودقيقة عن المؤشرات فأنها تعرض بهيئة رسوم بيانية وأشكال هندسية متعددة، والتالي يوضح طرق عرض المؤشرات:

1- **العرض الجدولي التكراري Frequency table** بعد جمع البيانات يتم تفريغ البيانات في جداول تصمم حسب البيانات المجمعة، ويتم عرضها وهذا يساعد الناظر على أخذ فكرة سريعة عن الظاهرة قيد الدراسة دون تعب وإجهاد، ويكتسب العرض الجدولي التكراري (Frequency table) أهمية كبرى بعد ان يقوم الباحث بتفريغ البيانات الإحصائية ضمن جداول لها ميزات رئيسية منها:

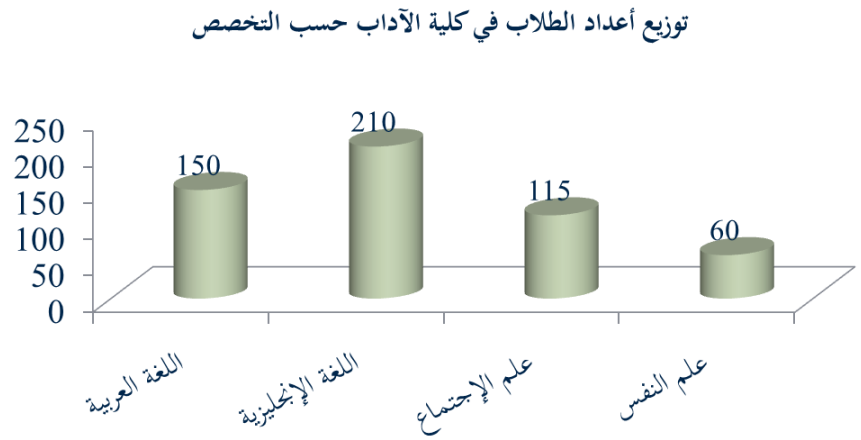
- أن يكون للجدول عنواناً كاملاً مختصرا ومعبراً عما يتضمن الجدول من بيانات
- أن يعطى لكل جدول رقم معين
- أن يضع لكل من الصفوف والأعمدة عناوين بارزة

مثال: جدول 1: التوزيع النسبي لطلبة الجامعة حسب التخصص العلمي

التخصص العلمي	عدد الطلبة	النسبة %
الكيمياء	95	17
الإحصاء	108	20
الرياضيات	65	12
الأحياء	43	8
الجيولوجيا	70	13
اللغة العربية	150	27
التاريخ	20	4
المجموع	**551**	**100**

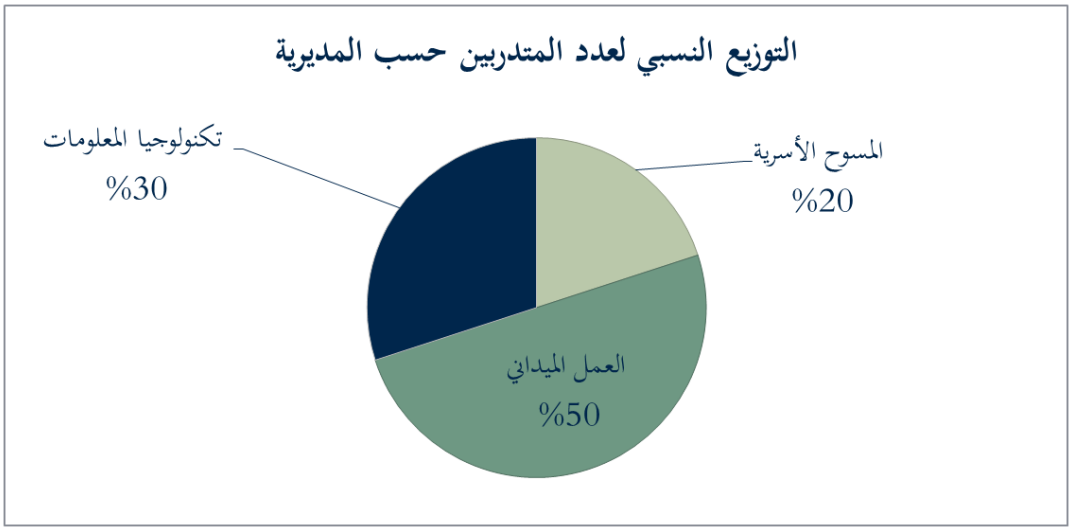
2. **العرض البياني Graphic representation** يتم تمثيل البيانات بأشكال هندسية ذات فائدة علمية، ويتم تمثيل البيانات بعدة طرق وهي

- الأعمدة أو المستطيلات Histogram

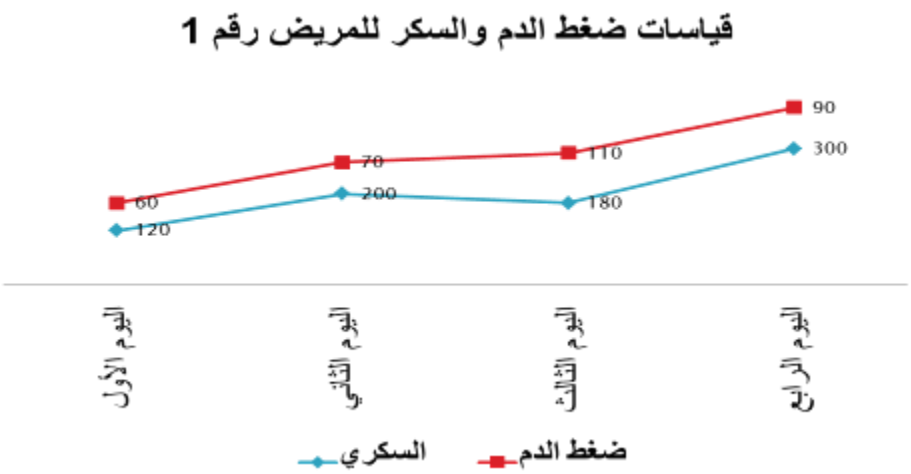


طرق

لتمثيل البيانات ذات الصفة المشتركة



- **التمثيل بالخط البياني** Line Chart، يستخدم في حال المقارنات للبيانات ضمن سلاسل زمنية معينة



4.2 مقاييس النزعة المركزية Measures of Central Tendency

في كثير من النواحي التطبيقية يكون الباحث في حاجة إلى حساب بعض المؤشرات التي يمكن الاعتماد عليها في وصف الظاهرة من حيث القيمة التي تتوسط القيم أو تنزع إليها القيم، ومن حيث التعرف على مدى تجانس القيم التي يأخذها المتغير، وأيضا ما إذا كان هناك قيم شاذة أم لا. والاعتماد على العرض البياني وحدة لا يكفى، لذا سنتناول هنا ما يسمى بمقاييس النزعة المركزية، حيث أن معظم قيم مفردات أي ظاهرة لها الرغبة في التجمع أو التمركز حول قيمة معينة تسمى القيمة المتوسطة، هذا التجمع عند هذه القيمة يسمى بالنزعة المركزية للبيانات. ومن أهم مقاييس النزعة المركزية: الوسط الحسابي، الوسيط، المنوال، الوسط التوافقي والرُبيعات.

1. الوسط الحسابي Mean

يعتبر من أهم مقاييس النزعة المركزية وأكثرها استخداما في النواحي التطبيقية، ويمكن حسابه للبيانات المبوبة وغير المبوبة ، كما يلي :

أولا: الوسط الحسابي للبيانات غير المبوبة

يعرف الوسط الحسابي بشكل عام على أنه مجموع القيم مقسوما على عددها. فإذا كان لدينا n من القيم، ويرمز لها بالرمز : $x_1, x_2, \ldots, x_n$.

فإن الوسط الحسابي لهذه القيم ، ونرمز له بالرمز $\bar{x}$ يحسب بالمعادلة التالية :

$$\text{الوسط الحسابي} = \frac{\text{مجموع القيم}}{\text{عدد القيم}}$$

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

حيث يدل الرمز Σ على المجموع .

مثـال

فيما يلي درجات 8 طلاب في مادة الإحصاء .

34 32 42 37 35 40 36 40

والمطلوب إيجاد الوسط الحسابي لدرجة الطالب في الامتحان .

الحـل

لإيجاد الوسط الحسابي للدرجات تطبق المعادلة رقم (3-1) كما يلي:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

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

أي أن الوسط الحسابي لدرجة الطالب في اختبار مادة الإحصاء يساوي 37 درجة

ثانيا: الوسط الحسابي للبيانات المبوبة

من المعلوم أن القيم الأصلية، لا يمكن معرفتها من جدول التوزيع التكراري، حيث أن هذه القيم موضوعة في شكل فئات، ولذا يتم التعبير عن كل قيمة من القيم التي تقع داخل حدود الفئة بمركز هذه الفئة ، ومن ثم يؤخذ في الاعتبار أن مركز الفئة هو القيمة التقديرية لكل مفردة تقع في هذه الفئة.

فإذا كانت k هي عدد الفئات ، وكانت $x_1, x_2, \ldots, x_k$ هي مراكز هذه الفئات،

$f_1, f_2, \ldots, f_k$ هي التكرارات ، فإن الوسط الحسابي يحسب بالمعادلة التالية:

$$\overline{x} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_k f_k}{f_1 + f_2 + \ldots + f_k} = \frac{\sum_{i-1}^{k} x_i f_i}{\sum_{i-1}^{k} f_i} \qquad (3-2)$$

مثـال

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.

الحـل

لحساب الوسط الحسابي باستخدام المعادلة رقم (3-2) يتم إتباع الخطوات التالية :

1- إيجاد مجموع التكرارات $\sum f$. 2- حساب مراكز الفئات x .
3- ضرب مركز الفئة في التكرار المناظر له $(x f)$، وحساب المجموع $\sum xf$
4- حساب الوسط الحسابي بتطبيق المعادلة رقم (3-2) .

فئات الوزن (C)	التكرارات f	مراكز الفئات x	$x f$
32-34	4	(32+34)÷2=33	4×33=132
35-37	7	36	7×36=252

38-40	13	39	13×39=507
41-43	10	42	10×42=420
44-46	5	45	5×45=225
47-49	1	48	1×48=48
المجموع	40		1584

إذا الوسط الحسابي لوزن التلميذ هو :

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

أي أن متوسط وزن التلميذ يساوي 39.6 k.g

2- الوسيط Median

هو أحد مقاييس النزعة المركزية، والذي يأخذ في الاعتبار رتب القيم ، ويعرف الوسيط بأنه القيمة التي يقل عنها نصف عدد القيم $(n/2)$ ، ويزيد عنها النصف الآخر $(n/2)$ ، أي أن 50% من القيم أقل منه، 50% من القيم أعلى منه. وفيما يلي كيفية حساب الوسيط في حالة البيانات غير المبوبة، والبيانات المبوبة.

أولا: الوسيط للبيانات غير المبوبة

لبيان كيف يمكن حساب الوسيط للبيانات غير المبوبة ، نتبع الخطوات التالية:

- ترتب القيم تصاعديا .
- تحديد رتبة الوسيط، وهي : رتبة الوسيط $= \left(\frac{n+1}{2}\right)$
- إذا كان عدد القيم (n) فردي فإن الوسيط هو:

$$\text{الوسيط} = \text{القيمة رقم} \left(\frac{n+1}{2}\right)$$

- إذا كان عدد القيم (n) زوجي، فإن الوسيط يقع بين القيمة رقم $(n/2)$، والقيمة رقم $((n/2)+1)$، ومن ثم يحسب الوسيط بتطبيق المعادلة التالي:

$$\text{الوسيط} = \frac{\text{القيمة رقم} \left(\frac{n}{2}\right) + \text{القيمة رقم} \left(\frac{n}{2}+1\right)}{2}$$

مثـال

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 :

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

والمطلوب حساب وسيط الإنتاج لكل نوع من السماد المستخدم، ثم قارن بينها.

الحـل

أولا : حساب وسيط الإنتاج للنوع الأول (a)

- ترتيب القيم تصاعديا :

قيمة الوسيط

الإنتاج	1.2	1.5	2	2.3	2.75	3	3.25
الرتبة	1	2	2	4	5	6	7

رتبة الوسيط

- عدد القيم فردى $(n = 7)$

- إذا رتبة الوسيط هي: $((n+1)/2=(7+1)/2=4)$.
- ويكون الوسيط هو القيمة رقم 4 ، أي أن وسيط الإنتاج للنوع a هو:

$Me_a = 2.3$ طن / هكتار

ثانيا : حساب وسيط الإنتاج للنوع الثاني (b) :

- ترتيب القيم تصاعديا .

- عدد القيم زوجي $(n=10)$ إذا
- رتبة الوسيط هي : $((n+1)/2=(10+1)/2=5.5)$.
- الوسيط = الوسط الحسابي للقيمتين الواقعتين في المنتصف (رقم 5 ، 6) .

$$Me_b = \frac{2.5+3}{2} = 2.75$$

وبمقارنة النوعين من السماد ، نجد أن وسيط إنتاجية النوع (a) أقل من وسيط إنتاجية النوع (b)، أي أن : $Me_b > Me_a$.

ثانيا: الوسيط للبيانات المبوبة

لحساب الوسيط من بيانات مبوبة في جدول توزيع تكراري ، يتم إتباع الخطوات التالية .

- تكوين الجدول التكراري المتجمع الصاعد .
- تحديد رتبة الوسيط : $\left(\frac{n}{2}\right)=\left(\frac{\sum f}{2}\right)$
- ويحسب الوسيط ، بتطبيق المعادلة

$$Me = L_k + \frac{h_k}{f_k}\left[\frac{\sum f_i}{2} - f'_{k-1}\right]$$

حيث أن :

L_k : هي الحد الأدنى للفئة الوسيطية المقابلة لأكبر تكرار.

h_k : طول الفئة الوسيطية

f_k : تكرار الفئة الوسيطية.

f'_{k-1} : التكرار المتجمع الصاعد السابق للفئة الوسيطية

3. المنوال Mode

يعرف المنوال بأنه القيمة الأكثر شيوعا أو تكرارا ، ويكثر استخدامه في حالة البيانات الوصفية ، لمعرفة النمط (المستوى) الشائع، ويمكن حسابه للبيانات المبوبة وغير المبوبة كما يلي:

أولا: حساب المنوال في حالة البيانات غير المبوبة

المنوال (Mod) = القيمة (المستوى) الأكثر تكرارا

ثانيا: حساب المنوال في حالة البيانات المبوبة.

$$Mod = L_k + \frac{(f_k - f_{k-1})}{(f_k - f_{k-1}) + (f_k - f_{k+1})} * h_k$$

حيث أن :

Lk : الحد الأدنى لفئة المنوال المقابلة لأكبر تكرار.

f_k : تكرار الفئة المنوالية

f_{k-1} : التكرار السابق للفئة المنوالية

f_{k+1} : التكرار اللاحق للفئة المنوالية

h_k : طول فئة المنوال

4- الوسط التوافقي Harmonic mean

هو مقلوب الوسط الحسابي لمقلوب القيم و يتم حسابه وفق الصيغة التالية:

أولا: حساب الوسط التوافقي في حالة البيانات غير المبوبة

$$\overline{H} = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}}$$

مثال: أوجد الوسط التوافقي للبيانات التالية:

2 , 5 , 3 , 4 , 7 , 8 , 8

$$\sum \frac{1}{X_i} = \frac{1}{2} + \frac{1}{5} + \frac{1}{3} + \frac{1}{4} + \frac{1}{7} + \frac{1}{8} + \frac{1}{8} = 1.68$$

$$\therefore \overline{H} = \frac{7}{1.68} \; 4.17$$

ثانيا: حساب الوسط التوافقي في حالة البيانات المبوبة

الوسط التوافقي لمراكز الفئات X1, X2,… ,Xn والمرجح بالتكرارات المناظرة

f1,f2,…,fn يكون

$$\overline{H} = \frac{\Sigma f_i}{\Sigma \frac{f_i}{X_i}}$$

5- الرُبيعات Quartiles

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

الرُبيع الأول (الأدنى :Q1 (وهو القيمة التي تقسم مجموعة القراءات (بعد ترتيبها تصاعدياً) إلى قسمين بحيث يسبقها ربع البيانات ويليها ثلاثة أرباع البيانات.

الرُبيع الثاني (الوسيط :Q2 (وهو القيمة التي تقسم مجموعة القراءات (بعد ترتيبها تصاعدياً) إلى قسمين بحيث يسبقها نصف البيانات ويليها نصف البيانات أيضاً.

الرُبيع الثالث (الأعلى :Q3 (وهو القيمة التي تقسم مجموعة القراءات (بعد ترتيبها تصاعدياً) إلى قسمين بحيث يسبقها ثلاثة أرباع البيانات ويليها ربع البيانات.

قانون عام للربيعات والمئينات والعشيرات

$$i = \frac{P}{100}(n+1)$$

حيث i: رتبة المئين
P: قيمة المئين
n: عدد الحلات

مثال :اوجد الربيع الأول للبيانات التالية المرتبة تصاديا والبالغ عددها 9
مرتبة تصاعدياً :22 21 18 17 16 16 13 12 11

(n = 9)

$Q1 = 25^{th}$ الربيع الأول $\frac{25}{100}(9+1) = 2.5$

اذن رتبة المئين 25 اي الربيع الاول = 2.5

الربيع الأول = 12.5

5.2 مقاييس التشتت

عند مقارنة مجموعتين من البيانات ، يمكن استخدام شكل التوزيع التكراري، أو المنحنى التكراري ، وكذلك بعض مقاييس النزعة المركزية ، مثل الوسط الحسابي والوسيط ، والمنوال ، والإحصاءات الترتيبية ، ولكن استخدام هذه الطرق وحدها لا يكفي عند المقارنة ، فقد يكون مقياس النزعة المركزية للمجموعتين متساوي ، وربما يوجد اختلاف كبير بين المجموعتين من حيث مدى تقارب وتباعد البيانات من بعضها البعض ، أو مدى تباعد أو تقارب القيم عن مقياس النزعة المركزية .

ومثال على ذلك ، إذا كان لدينا مجموعتين من الطلاب ، وكان درجات المجموعتين كالتالي:

المجموعة الأولى	**63**	**70**	**78**	**81**	**85**	**67**	**88**
المجموعة الثانية	**73**	**78**	**77**	**78**	**75**	**74**	**77**

لو قمنا بحساب الوسط الحسابي لكل مجموعة ، نجد أن الوسط الحسابي لكل منهما يساوي **76** درجة ، ومع ذلك درجات المجموعة الثانية أكثر تجانسا من درجات المجموعة الأولى . من أجل ذلك لجأ الإحصائيون إلى استخدام مقاييس أخرى لقياس مدى تجانس البيانات، أو مدى انتشار البيانات حول مقياس النزعة المركزية، ويمكن استخدامها في المقارنة بين مجموعتين أو أكثر من البيانات، ومن هذه المقاييس ، مقاييس التشتت، والالتواء ، و التفرطح ، وسوف نركز هنا على مقاييس التشتت .

من هذه المقاييس: المدى ، الانحراف المتوسط ، والانحراف المعياري ، والتباين

1. **المدى** Rang: هو أبسط مقاييس التشتت ، ويحسب المدى في حالة البيانات غير المبوبة بتطبيق المعادلة التالية .

$$\mathbf{R = X_L - X_s}$$

وأما المدى في حالة البيانات المبوبة له أكثر من صيغة، ومنها المعادلة التالية

المدى = (الحد الأعلى للفئة الأخيرة) – (الحد الأدنى للفئة الأولى)

<u>مثـال</u>

الجدول التكراري التالي يبين توزيع **60** مزرعة حسب المساحة المنزرعة بالذرة بالألف دونم

المساحة	15-20	21-26	27-32	33-38	39-44	45-50
عدد	3	9	15	18	12	3

						المزارع

والمطلوب حساب المدى للمساحة المنزرعة بالذرة .

الحـل

المدى = (الحد الأعلى للفئة الأخيرة) – (الحد الأدنى للفئة الأولى)

المدى = 50 – 15 = 35 دونم **R = 50 – 15 = 35**

2. التباين Variance

هو أحد مقاييس التشتت ، وأكثرها استخداما في النواحي التطبيقية ، ويعبر عن متوسط مربعات انحرافات القيم عن وسطها الحسابي. ويحسب التباين من بيانات العينة كتقدير لتباين المجتمع ، فإذا كانت قراءات عينة عشوائية حجمها ***n*** هي ، $x_1, x_2, ..., x_n$ ، فإن تباين العينة ويرمز له بالرمز $\mathbf{s^2}$ هو:

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

حيث أن $\bar{x}$ هو الوسط الحسابي لقراءات العينة ، أي أن : $\bar{x} = \sum x/n$ ، وتباين العينة

3. الانحراف المعياري Standard Deviation

عند استخدام التباين كمقياس من مقاييس التشتت، نجد أنه يعتمد علي مجموع مربعات الانحرافات، ومن ثم لا يتمشى هذا المقياس مع وحدات قياس المتغير محل الدراسة ، من أجل ذلك لجأ الإحصائيين إلى مقياس منطقي يأخذ في الاعتبار الجذر ألتربيعي للتباين ، لكي يناسب وحدات قياس المتغير، وهذا المقياس هو الانحراف المعياري.

إذا الانحراف المعياري ، هو الجذر ألتربيعي الموجب للتباين ، أي أن:

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

4. مقاييس التشتت النسبية Relative measures of dispersion

قد يتطلب الأمر في بعض الأحيان إجراء مقارنة بين تشتت مجموعتين أو أكثر من القيم المختلفة من حيث الوسط الحسابي أو أن قيم مفردات كل مجموعة مقاسه بوحدات قياس تختلف عن الأخر . و عندئذ فان مقاييس التشتت أيا كان سوف لن يكون نافعا لوحده في إجراء مقارنات من هذا النوع ، إنما يستوجب الأمر إيجاد مقياس تشتت أخر أكثر ملائمة لهذه الحالات ، هذا النوع من المقاييس تسمى بمقاييس التشتت النسبي وهي:

1- معامل التشتت المستند إلى الانحراف المتوسط **Mean deviation of dispersion coefficient**

افرض ان **M.D(A)** تمثل الانحراف المتوسط المحتسبة على أساس نقطة اختيارية **(A)** وقد تكون (الوسط الحسابي أو الوسيط أو المنوال) عندئذ يعرف معامل التشتت للمتوسط على نحو الأتي:

$$C.D(A)=\frac{M.D(A)}{A}$$

2- معامل التشتت المستند إلى الانحراف المعياري (معامل الاختلاف) **Coefficient of Variation (C.V)**

افرض ان $\bar{X}$ يمثل الوسط الحسابي لمجموعة من القيم ، و أ، (**S**) يمثل الانحراف المعياري لها . عندئذ يعرف معامل الاختلاف و الذي نرمز له بالرمز **(C.V)** على النحو الأتي:

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

إن معامل الاختلاف يعتبر بحق أفضل معاملات التشتت الأنفة الذكر كونه يعتمد على أفضل مقاييس النزعة المركزية و أفضل مقاييس التشتت . إن هذا المعامل يوضح نسبة حصة كل وحدة من وحدات الوسط الحسابي و الانحراف

المعياري و عليه عند إجراء مقارنة بين قيم مجموعتين ثم مقارنة معامل الاختلاف للمجموعة الأولى مع معامل الاختلاف للمجموعة الثانية ، عندئذ يقال عن المجموعة بأنها أكثر تجانسا إذا كان معامل الاختلاف اقل من الأخر.

تمارين

اوجد حلول التمارين التالية

1. وضح الفرق بين البيانات والمعلومات مبينا ارتباطهما بمفهوم المؤشرات.
2. أجب ب (صح) أو (خطأ) امام كل من العبارات التالية:

-( ) المدى هو عبارة عن مقياس لحساب عدد الفئات.
-( ) الوسط الحسابي يحدد لنا مركز التوزيع الطبيعي.
- ( ) يعرف المنوال بأنه القيمة التي تقسم البيانات إلى قسمين متساويين .
-( ) متغير عدد علامات طلاب الصف يعتبر متغير نوعي أسمي.
-( ) يعرف الوسط الحسابي بأنه مجموع القيم على عددها.
-( ) تعتبر طريقة عرض البيانات pie-chart أفضل الطرق لمقارنة متغيرات عبر تسلسل زمني معين.

3. احسب متوسط أعمار الطلاب للبيانات التالية

فئات العمر	5-6	7-8	9-10	11-12	13-14
عدد الطلاب	2	5	8	4	1

4. أوجد الوسيط لدرجات الطلاب التالية

72, 60, 72, 40, 80, 63

5. احسب المدى، والتباين والانحراف المعياري للبيانات التالية:

6,3,5,5,9,4,6,7,1,2,4,8

6. ما هو منوال البيانات التالية :

70،70،70،70،7،70

الفصل 3

الاستقراء (التحليل الاستدلالي)

لاحظنا في الفصول السابقة الطرق والأساليب المختلفة في جمع البيانات وتصنيف وتبويب البيانات، كذلك عملية بناء بعض المؤشرات بلاعتماد على المقاييس المختلفة كالمتوسطات ومقاييس التشتت، إن الأساليب استندت على البيانات المجمعة عن متغير واحد فقط. وفي أحوال كثيرة يحتاج الباحث دراسة متغيرين أو أكثر في آن واحد لبيان طبيعة ونوع العلاقة التي ترتبط بها هذه المتغيرات، وعليه فان هذا الفصل سوف يخصص لدراسة مقاييس أخرى تحدد درجة ونوع وشكل العلاقة بين متغيرين أو أكثر.

الهدف العام للفصل

- القدرة على تحديد معاملات الارتباط الصحيحة حسب أنوع المتغيرات
- تقدير النماذج وبناء السيناريوهات
- القدرة على تحديد نوع الاختبار المطلوب
- التمييز بين الاختبارات المعلمية واللامعلمية

1.3 معاملات الارتباط Correlation Coefficient

إن مفهوم الارتباط يقترن بحالة وجود متغيرين أو أكثر، وترتبط هذه المتغيرات مع بعضها البعض بعلاقات خطية معينة، على سبيل المثال العلاقة بين طول الشخص و وزنه، العلاقة بين نسبة الشفاء من مرض معين وكمية الجرعة من الدواء المخصص للمريض وعمر المريض. فإذا كان التغير في إحدى المتغيرات يؤثر في تغير متغير أخر أو مجموعة متغيرات أخرى عندها يقال أن هذه المتغيرات مرتبطة فيما بينها وإذا كان المتغيرين المرتبطين (أو مجموعة من المتغيرات المترابطة) يتغيران (تتغير) بنفس الاتجاه أي زيادة أو نقصان في

إحداهما تؤدي إلى الزيادة أو النقصان في الأخر (الأخرى) عندئذ يقال أن الارتباط ما بينهما هو ارتباط موجب.

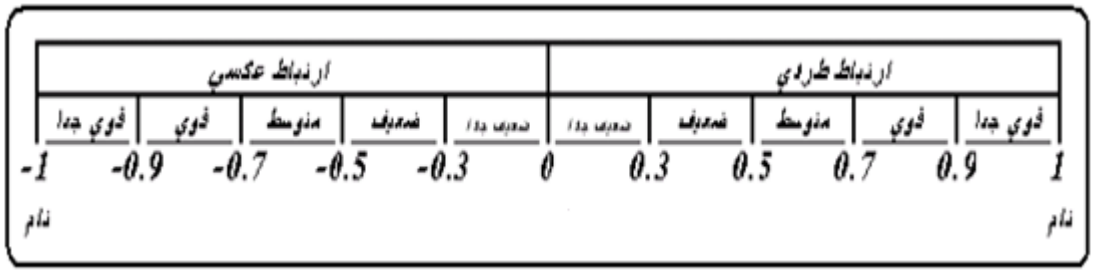
و يقال أن الارتباط بين المتغيرين أو أكثر هو ارتباط تام إذا كان التغير في إحداهما متناسب مع التغير في الأخر، وعلى سبيل المثال إن الارتباط بين درجة الحرارة المئوية ودرجة الحرارة الفهرنهايتية هو ارتباط تام باعتبار إن التغير في الأول متناسب مع التغير في الثاني.

تأخذ العلاقة بين المتغيرات ثلاث أنواع حسب إشارة معامل الارتباط كما يلي:

1- إذا كانت إشارة معامل الارتباط سالبة ($r < 0$) توجد علاقة عكسية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه انخفاض في المتغير الثاني، والعكس.
2- إذا كانت إشارة معامل الارتباط موجبة ($r > 0$)توجد علاقة طردية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه زيادة في المتغير الثاني، والعكس .
3- إذا كان معامل الارتباط قيمته صفرا ($r = 0$) دل ذلك على انعدام العلاقة بين المتغيرين.

اما من حيث قوة العلاقة، فيمكن 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



يمكن حساب قيمة معامل الارتباط بعدة طرق مختلفة تبعاً لنوع البيانات.

الارتباط بين المتغيرات الرقمية: معامل بيرسون للارتباط.

الارتباط بين المتغيرات الترتيبية: معامل سبيرمان للرتب

الارتباط بين المتغيرات الوصفية: مربع كاي Chi-Square.

1.1.3 معامل ارتباط بيرسون

يستخدم معامل الارتباط بيرسون لقياس التغير الذي يطرأ على المتغير Y عندما تتغير قيم X أو العكس، ويستخدم عادة في حالة البيانات الكمية، فإذا كان لدينا ازواج المشاهدات التالية:

$$(x_1,y_1),(x_2,y_2),\ldots..,(x_n,y_n)$$

فإن معامل الارتباط r يحسب من خلال العلاقة:

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

2.1.3 معامل ارتباط سبيرمان

في حال كانت كل من X,Y متغيرين من النوع الوصفي والبيانات المستحصل عليها من X, Y على أساس عينة عشوائية n هي بهيئة صفات غير قابلة للقياس الكمي .

وعلى فرض أن **(Xi , Yi , i=1,2,3,…..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)، هنا يمكن اعطار رتب لهذه البيانات بعد ترتيبها، وفق المعادلة التالية ويسمى هذا المعامل معامل ارتباط سبيرمان

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

حيث r_{xy} : معامل ارتباط سبيرمان، n تمثل عدد ازواج القيم (x,y)، d هي الفرق بين القيم (x,y).

2.3 تقدير دالة الانحدار

الارتباط الخطي البسيط Simple Linear Regression

يعرف الانحدار الخطي البسيط بأنه عملية تقدير العلاقة الخطية لمتغيرين أحدهما متغير يقاس دون خطأ ويسمى متغير مستقل Independent variable والآخر يأخذ قيماً تعتمد على قيمة المتغير المستقل ويسمى التابع Dependent variable

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

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

- المتغير التابع :Dependent (Response) Variable هو المتغير الذي يقيس نتيجة دراسة ما، وعادة يرمز له بالرمز Y
- المتغير المستقل :Independent (Explanatory) Variable هو المتغير الذي يُفسِّر أو يسبب التغيرات في المتغير التابع، أي هو الذي يؤثر في تقدير قيمة المتغير التابع، وعادة يرمز له بالرمز X فمثلاً عدد أيام الغياب X و درجة الطالب في الإحصاء Y ، العُمر X والإصابة بضغط الدم Y

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

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

إن مفهوم الانحدار الخطي البسيط يقترن بمفهوم الارتباط الخطي البسيط. و يهدف الانحدار الخطي البسيط إلى تقدير قيم عددية لمعالم النموذج، أي تقدير قيم عددية لكل من **(a,b).**

ومعادلة خط الانحدار البسيط هي كالأتي:

Y = a + b X + e

إذا X تغيرت بمقدار وحدة واحدة فإن Y ستتغير بمقدار

Y, X هما متغيرين بينهما علاقة قوية طردية أو عكسية.

يعرف **a** على انه ثابت الانحدار (الجزء المقطوع من محور **y**) ويحسب من خلال العلاقة

$$a = \bar{y} - b\bar{x}$$

كذلك يعرف **b** بمعامل انحدار **y** على **x** ويحسب من خلال العلاقة التالية

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

من المعادلة السابقة، يبين لنا أن المتغير **Y** هو متغير **تابع** تعتمد قيمته على المتغير **المستقل X**.

e مقدار الخطأ

مثال يوضح أحد السيناريوهات الممكن بناؤها بين متغيري الوزن والطول:

Height = 10 + 3 * Weight

المتغير **Height** هو المتغير التابع **Dependent**

المتغير **Weight** هو المتغير المستقل **Independent**

الثابت **(a)** هو **10**

الثابت **(b)** هو **3**

1.2.3 تقدير النماذج وتحديد السناريوهات

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

والإنحدار الخطي المتعدد هو عبارة عن إيجاد معادلة رياضية تعبر عن العلاقة بين متغيرين وتستعمل لتقدير قيم سابقة ولتنبؤ قيم مستقبلية، وهو عبارة أيضاً عن إنحدار للمتغير التابع (**Y**) على العديد من المتغيرات المستقلة **X1 , X2 , XK...** لذا فهو يستخدم في التنبؤ بتغيرات المتغير التابع الذي يؤثر فيه عدة متغيرات مستقلة أي تعتمد فكرته على العلاقات الدلالية التي تستخدم ما يعرف بشكل التشتت أو االانتشار ، فبإمكاننا التنبؤ بالمستوى الرقمي في فعالية رمي المطرقة على سبيل المثال إعتمادا على دراسة حالات أخرى للرامي كالعمر الزمني والعمر التدريبي والمهارة والمواصفات الجسمية وغيرها .

إن الإنحدار الخطي المتعدد ليس مجرد أسلوب واحد وإنما مجموعة من الأساليب التي يمكن أستخدامها لمعرفة العلاقة بين متغير تابع مستمر وعدد من المتغيرات المستقلة التي عادةً ما تكون مستمرة)

والمعادلة الخطية في الإنحدار الخطي المتعدد هي :

$$Y = a + b1X1 + b2X2 + \ldots\ldots\ldots + e$$

حيث أن **Y** = المتغير التابع

a = قيمة ثابتة **Constant** أو **Intercept**

b1 = ميل ألإنحدار **y** على المتغير المستقل الأول

b2 = ميل ألإنحدار **y** على المتغير المستقل الثاني

$X1$ = المتغير المستقل الأول

$X2$ = المتغير المستقل الثاني

ويمكن استخدام الإنحدار الخطي المتعدد في حالة توافر الشروط التالية :

1. أن تكون العلاقة خطية بين المتغيرات المستقلة والمتغير التابع .
2. أن تكون البيانات موزعة توزيعاً طبيعياً للمتغيرات المستقلة والمتغير التابع .
3. يجب أن تكون قيم المتغير التابع من المستوى الترتيبي على الأقل .

بعد الحصول على نتائج معادلة الانحدار يجب علينا أن نبين هل أن هذه المعاملات مقبولة من الناحية الإحصائية أي معنوية احصائياً مع التنويه بأن المعنوية تكون لكل معامل على حدة .

ولتقدير معادلة إنحدار في حال وجود أكثر من متغيرين يجب التعرف على بعض النماذج التي يمكن من خلالها بناء سيناريوهات مختلفة والتي من خلال تطبيقها قد نحتاج الى جراء بعض المعالجات الاحصائية مثل

- أخذ اللوغاريتم الطبيعي للمتغير التابع (**GDP**).
- إضافة/ حذف متغير مستقل.

حيث أن هذه المعالجات تمكننا من الحصول على نموذج مثالي لتقدير الدالة

ونورد بعض المثلة على سيناريوهات احصائية لعدد من النماذج:

1- تقدير دالة الناتج المحلي الإجمالي:

$$GDP = C + I + G + (X-M) + e$$

Where:

GDP: Gross Domestic Product الناتج المحلي الإجمالي

C: Private Consumption الإستهلاك الخاص

I: Investment الإستثمار

G: Government Expenditure الإنفاق الحكومي

(X-M): Export minus Import (Net Foreign Trade) الصادرات ناقص المستوردات

e: error term مقدار الخطأ

2. تقدير دالة العائد على التعليم:

$$Y = edu + no_edu + age + G + exp + Ms + e$$

Where:

Y: الدخل من العمل

edu: المستوى التعليمي

no_edu: عدد سنوات التعليم

age: العمر

G: النوع الاجتماعي

Exp: عدد سنوات الخبرة

Ms: الحالة الزواجية

e: error term مقدار الخطأ

3. تقدير دائرة عرض العمل:

$$WH = edu + no_edu + age + G + exp + Ms + Prof + Wage + e$$

Where:

عدد ساعات العمل :WH

المستوى التعليمي :edu

عدد سنوات التعليم :no_edu

العمر :age

النوع الاجتماعي :G

عدد سنوات الخبرة :Exp

الحالة الزواجية :Ms

المهنة :Prof

الأجر أو الدخل من العمل :Wage

مقدار الخطأ e: error term

بعد الحصول على نتائج معادلة الانحدار يجب علينا ان نبين هل ان هذه المعاملات مقبولة من الناحية الاحصائية (معنوية احصائيا) مع التنويه ان المعنوية تكون لكل معامل على حدة ، لكي نحكم على معنوية معاملات الانحدار نستعين باختبار T او احصائية t ، ومستوى الاحتمالية المقابل لها ، وبرنامج SPSS يقوم تلقائيا باستخراج اختبار t ومستوى الاحتمالية المقابل لها، حيث سنوضح هذا الموضوع في الفصل اللاحق.

3.3 اختبار الفرضيات Test of Hypotheses

1.3.3 تعريف الفرضية statistical hypothesis

هي عبارة عن إدعاء او تخمين معين حول معلمة من معالم المجتمع ويكون المطلوب اختبار صحة هذا الادعاء أو التخمين.

2.3.3 أنواع الفرضيات

هناك نوعين من الفروض :

- فرض العدم (null hypothesis) ويرمز له بالرمز H_o ويصاغ في صورة عدم وجود فرق أو عدم وجود علاقة أو عدم وجود تغير – مثال : في مثال أعمار الطلاب وطالبات الجامعة فإن فرض العدم هو

H_o : نفترض عدم وجود اختلاف بين متوسطي اعمار الطلاب والطالبات

- الفرض البديل (alternative hypothesis) ويرمز له بالرمز H_1 وهو الفرض الذي يجب أن يكون صحيحا اذا كان فرض العدم غير صحيح – مثال : في مثال أعمار الطلاب وطالبات الجامعة فإن الفرض البديل هو

H_1: يوجد اختلاف حقيقي وليس ظاهري بين متوسط اعمار الطلاب والطالبات .

إن القرار الذي سوف نتخذه بناء على الاختبار الإحصائي لا يمكن اعتباره صحيح 100 % فهناك مقدار من الخطأ لأن المعلومات التي نتخذ قرارنا بناء عليها بيانات مأخوذة من عينة وليس من المجتمع الأصلي

في اختبار فرض معين، فإن مقدار ثقتنا في القرار المتخذ بالرفض أو القبول يسمى بدرجة الثقة ويرمز له بالرمز $(1 - \alpha)$ كما وأن مقدار عدم الثقة أو مقدار الخطأ يسمى بمستوى المعنوية ويرمز له بالرمز α

وعادة يحدد الباحث مستوى المعنوية أو درجة الثقة قبل البدء في عملية الاختبار.

عند اختبار فرض العدم H_o ضد الفرض البديل H_1 نجد أننا امام احدى الحالات الاربع الاتية :

	H0 صحيح	H0 خطأ
قبول H0	قرار سليم	خطأ من النوع الثاني
رفض H0	خطأ من النوع الاول	قرار سليم

1) أن يكون فرض العدم صحيحا ويكون القرار بقبوله ....وهذا قرار سليم

2) أن يكون فرض العدم صحيحا ويكون القرار برفضه ..... . وهذا قرار خاطئ

(الخطأ من النوع الأول : رفض H_o عندما يكون H_o صحيحا ويرمز لحجم هذا الخطأ بالرمز α)

3) أن يكون فرض العدم خطأ ويكون القرار برفضه وهذا قرار سليم

4) أن يكون فرض العدم خطأ ويكون القرار بقبوله ..وهذا قرار خاطئ (الخطأ من النوع الثاني : قبول H_o عندما يكون H_o خاطئ ويرمز لحجم هذا الخطأ بالرمز β)

- احتمال الوقوع في الخطأ من النوع الأول يسمى مستوى المعنوية ويرمزله بالرمز α أي ان α = احتمال رفض فرض العدم H_o عندما يكون صحيح = مستوى المعنوية

- احتمال الوقوع في خطأ من النوع الثاني يرمز له بالرمز β أي أن

β = احتمال قبول فرض العدم H_o عندما يكون خطأ

3.3.3 خطوات اختبار الفرضية الإحصائية

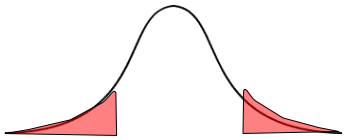
لإجراء الاختبار الإحصائي فإننا نتبع الخطوات التالية :

1- صياغة فرض العدم Ho

$Ho : \mu = \mu 0$

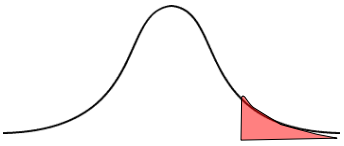
والفرض البديل هو احد الحالات التالية :

$H1: \mu \neq \mu 0$ -1



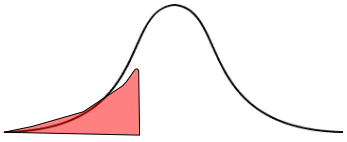
(اختبار من طرفين)

$H1:\mu > \mu 0$ -2



(اختبار من طرف واحد ،الجهة اليمنى)

$H1:\mu < \mu 0$ -3



(اختبار من طرف واحد ،الجهة اليسرى)

2- تحديد قيمة احصاءة الاختبار(قيمة Z المحسوبة):

حيث أن هذا الاحصاءة يتبع تقريبا توزيعا طبيعيا قياسياً

$$Z_C = \frac{\overline{X} - \mu_O}{\frac{\sigma}{\sqrt{n}}}$$

3- تحديد القيمة الجدولية و تحدد على حسب نوع الاختبار وقيمة α :

نوع الاختبار	مستوى المعنوية α	درجه الثقة $(1-\alpha)$	الدرجة المعيارية
اختبار من طرفين	$5\%=0.05$	$95\%=0.95$	$= Z_{\alpha/2} \pm 1.96$
	$1\%=0.01$	$99\%=0.99$	$= Z_{\alpha/2} \pm 2.58$

نوع الاختبار	مستوى المعنوية α	درجه الثقة $(1-\alpha)$	القيمة الجدولية (القيمة الحرجة)
اختبار من طرف واحد (الجهة اليمنى)	$5\%=0.05$	$95\%=0.95$	$=1.64 \alpha\ Z$
	$1\%=0.01$	$99\%=0.99$	$=2.33 \alpha\ Z$
اختبار من طرف واحد(الجهةاليسرى)	$5\%=0.05$	$95\%=0.95$	$= -1.64 \alpha\ Z$
	$1\%=0.01$	$99\%=0.99$	$= -2.33 \alpha\ Z$

4- اتخاذ القرار:

نتخذ القرار بناءاً على قيمة احصاءة الاختبار

نرفض Ho إذا وقعت قيمة احصاءة الاختبار في منطقة الرفض

لا نرفض Ho إذا وقعت قيمة احصاءة الاختبار في منطقة القبول

إذا كان الاختبار من طرفين : نقبل فرض العدم إذا تحققت المعادلة التالية :

$$-Z_{\alpha/2} < Z_C < Z_{\alpha/2}$$

نرفض فرض العدم إذا تحققت إحدى المعادلتين : $Z_C > Z_{\alpha/2}$ $Z_C < -Z_{\alpha/2}$

إذا كان الاختبار من طرف واحد الجهة اليمنى :

نقبل فرض العدم إذا تحققت المعادلة : $Z_C < Z_\alpha$

نرفض فرض العدم إذا تحققت المعادلة : $Z_C > Z_\alpha$

إذا كان الاختبار من طرف واحد الجهة اليسرى :

نقبل فرض العدم إذا تحققت المعادلة : $Z_C > -Z_\alpha$

نرفض فرض العدم إذا تحققت المعادلة : $Z_C < -Z_\alpha$

مثال (1):

شركة متخصصة في صناعة لعب الأطفال تعاقدت لشراء نوع جديد من الخيوط الصناعية، يدعي صانع هذه الخيوط أن متوسط قوة تحمل الخيط 15 كجم بانحراف معياري نصف كجم .

ولاختبار صحة ادعاء الصانع أخذت عينة عشوائية من 50 خيطاً وتم اختبارها فوجد أن متوسط قوة التحمل في العينة 14.8كجم . فهل يمكننا تأييد ادعاء المدير؟ استخدم مستوى معنوية 5 % .

الحل:

$$n=50 \qquad \mu_o=15\ kg$$
$$\overline{X}=14.8\ kg \quad \sigma=0.5\ kg$$

1- صياغة الفرض الإحصائي:

الفرض العدم : عدم وجود فرق معنوي بين متوسط الخيوط القديمة والحديثة $H_o:\mu=15$

الفرض البديل : وجود فرق معنوي بين متوسط الخيوط القديمة والحديثة $H_1:\mu\neq15$

حيث μ هي متوسط قوة تحمل الخيط .

2- إيجاد قيمة إحصاء الاختبار:

لان σ مجهولة والعينة كبيرة فانه يمكن استخدام S بدلاً من σ :

$$Z_C=\frac{\overline{X}-\mu_O}{\frac{\sigma}{\sqrt{n}}}=\frac{14.8-15}{\frac{0.5}{\sqrt{50}}}=-2.83$$

3- تحديد القيمة الجدولية :

$$Z_{\alpha/2}=\pm1.96$$

4- اتخاذ القرار:

بما أن القيمة المحسوبة وقعت في منطقة الرفض، فإن القرار هو: رفض فرض العدم

أي أن الادعاء غير صحيح وأن هناك فرق معنوي بين المتوسط الحقيقي والمتوسط المدعى.

مثال(2) :

إذا كان من المعروف أن جسم الإنسان البالغ يحتاج يوميا في المتوسط 800 ميللجرام من الكالسيوم لكي يقوم بوظائفه خير قام.ويعتقد احد علماء التغذية أن الأفراد ذوي الدخل المنخفض لا يستطيعون تحقيق هذا المتوسط, ولاختبار ذلك تم اختيار عينة من 50 شخصا بالغا من بين ذوي الدخل المنخفض فكان متوسط ما يتناوله من كالسيوم يوميا هو 755.3 ميللجرام والانحراف المعياري هو 239.3 ميللجرام .فهل تدل هذه النتائج على أن متوسط ما يتناوله الأشخاص البالغون من ذوي الدخل المنخفض من كالسيوم يقل عن 800 ميللجرام؟ استخدم مستوى معنوية 0.05

الحل :

1- صياغة الفرض الإحصائي:

فرض العدم هو

$$Ho: \mu = 800$$

والفرض البديل

$$H1: \mu < 800$$

حيث μ هي متوسط ما يتناوله الإنسان البالغ ذوي الدخل المنخفض من الكالسيوم .

2- إيجاد قيمة إحصاء الاختبار:

لان σ مجهولة والعينة كبيرة فانه يمكن استخدام s=239.3 بدلا منها .وبالتعويض نجد ان قيمة احصاءة الاختبار هي

$$Z= \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{755.3-800}{239.3/\sqrt{50}} = -1.32$$

3- تحديد القيمة الجدولية :

ونلاحظ هنا أن الاختبار ذو جانب (طرف) واحد هو الجانب (الطرف) الأيسر وحيث أن مستوى المعنوية $\alpha=0.05$ فانه من جدول التوزيع الطبيعي القياسي نجد أن القيمة الحرجة هي :

$$= -1.64 \alpha\ Z$$

4- اتخاذ القرار:

بما أن قيمة الاحصاءة -1.32 اكبر من القيمة الحرجة 1.64- وهي تقع في منطقة القبول وبالتالي فإننا لانرفض فرض العدم H_0 وهو أن متوسط ما يتناوله الإنسان البالغ ذو الدخل المنخفض من الكالسيوم يساوي 800 ميللجرام .

مثال (3) :

في عينة عشوائية مكونة من تسجيل 100 حالة وفاة في قرية معينة تبين أن متوسط العمر في العينة 67.5 عاما والانحراف المعياري 8 أعوام. فهل هذا يوضح أن متوسط العمر في هذه القرية اكبر من 65 عاما ؟

استخدم مستوى معنوية 5 %.

الحل:

نفرض أن μ متوسط العمر في هذه القرية .

1- صياغة الفرض الإحصائي:

$$H_0 : \mu = 65$$

$$H_1 : \mu > 65$$

2- إيجاد قيمة إحصاء الاختبار:

$$Z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{67.5 - 65}{8/\sqrt{100}} = 3.125$$

3- تحديد القيمة الجدولية :

ونلاحظ هنا أن الاختبار ذو جانب (طرف) واحد هو الجانب (الطرف) الأيمن وحيث أن مستوى المعنوية $\alpha = 0.05$ فانه من جدول التوزيع الطبيعي القياسي نجد أن القيمة الحرجة هي :

$$Z_{\alpha} = 1.64$$

4- اتخاذ القرار:

نجد أن قيمة Z المحسوبة 3.125 اكبر من القيمة الجدولية 1.64 لهذا فان Z المحسوبة تقع في منطقة الرفض لهذا فان القرار هو رفض H_0

ونستنتج من ذلك أن متوسط العمر في هذه القرية اكبر من 65 عاما .

وفي الفصل التالي سنتناول بعض المثلة التطبيقية على اختبار الفرضيات من خلال التعامل مع برنامج spss.

4.3 أنواع الاختبارات الاحصائية

هناك نوعان من الطرق الإحصائية التي تستخدم في اختبار الفرضيات:

1.4.3 الاختبارات المعلمية: وهي الإختبارات الإحصائية التي تستخدم في التحقق من صحة الفروضيات المتعلقة بعينات لها الصفات التالية:

- اعتدالية التوزيع في البيانات
- أن يكون حجم العينة كبير وتم اختيارها عشوائياً
- من أمثلته : اختبارات (ت)، الارتباط الخطى، تحليل التباين ANOVA

فيما يلي أنواع الاختبارات المعلمية

1. One – Sample T-Test يستخدم هذا الاختبار لفحص فرضية تتعلق بالوسط الحسابي
2. Independent – Sample T-Test يستخدم لفحص فرضية متعلقة بمساواة متوسط متغير ما لعينتين مستقلتين
3. Paired Sample T-Test يستخدم لفحص فرضية متعلقة بمساواة متوسط متغيرين لعينة واحدة
4. One Way ANOVA يستخدم لاختبار معنوية الفرق بين المتوسطات لعدة عينات بمقارنة واحدة

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

الطبيعي طبيعي، وكذلك في حالتي البيانات الترتيبية والوصفية.، وفي حالة حجم العينة كان صغيرا.

مزايا استخدام الاختبارات اللامعلمية:

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

عيوب استخدام الاختبارات اللامعلمية:

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

فيما يلي أنواع الاختبارات اللامعلمية:

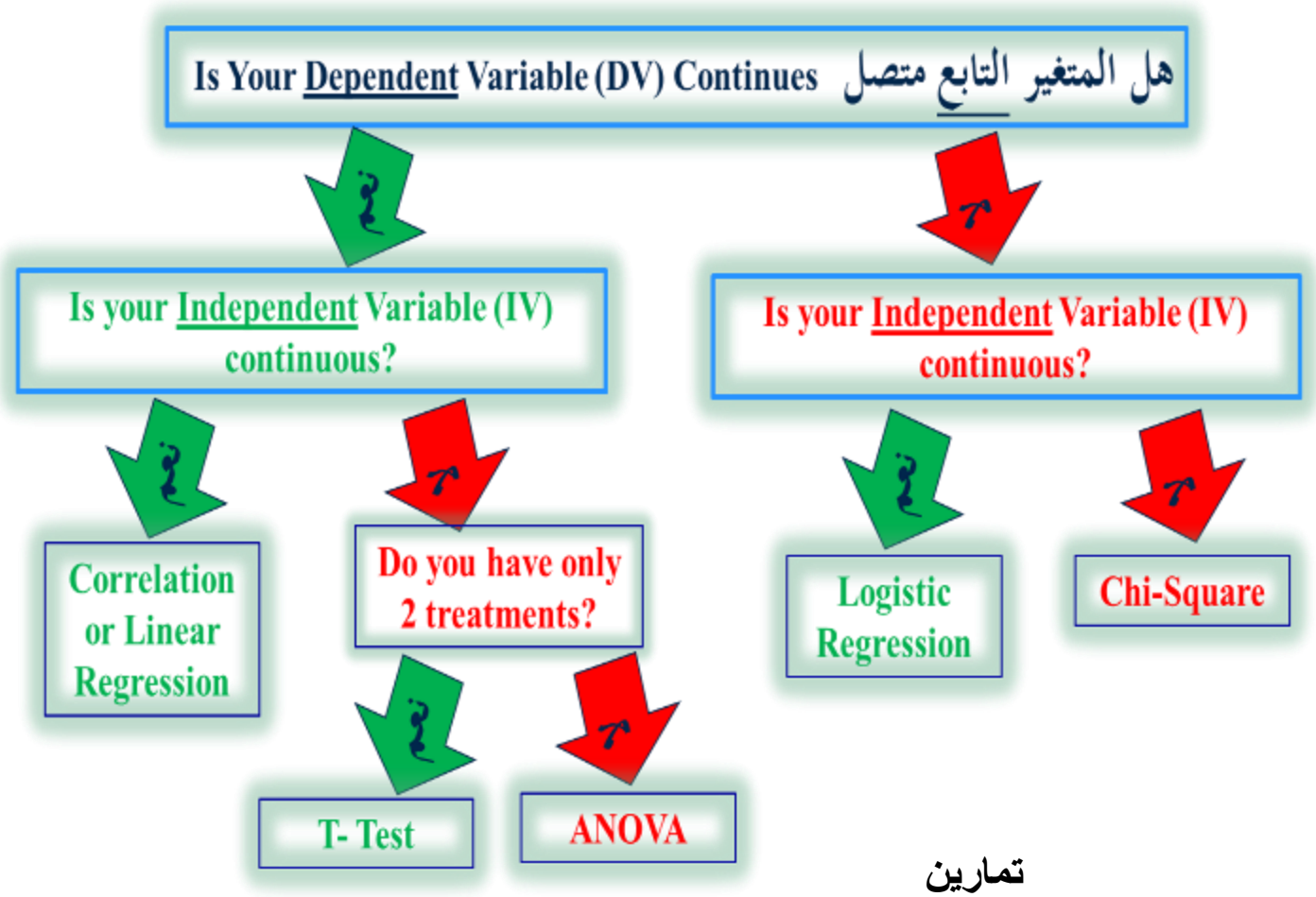
- استخدام اختبار كولمجروف – سمرنوف"One-Sample Kolmogorov-Smirnov Test" لمعرفة ما إذا كانت البيانات تتبع التوزيع الطبيعي.
- اختبار ويلكوكسن "Wilcoxon Test" لاختبار فرضيات حول مقارنة متوسطي مجتمعين في حالة العينات المرتبطة.
- اختبار مان – وتني "Mann Whitney Test" لاختبار الفرضيات حول الفرق بين متوسطي مجتمعين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

الجدول التالي يوضح الاختلاف بين استخدام الاختبارات المعلمية والاختبارات اللامعلمية

الاختبارات المعلمية parametric tests	الاختبارات اللامعلمية nonparametric tests
تصلح للعينات الكبيرة بشكل اساسي	تصلح للعينات الكبيرة والصغيرة (مثلا اذا كان حجم العينة 6 او اقل فلا بديل عن استخدام الاختبارات اللامعلمية
يشترط توفر معلومات عن توزيع المجتمع	لا يشترط افتراضات او معلومات حول توزيع المجتمع
تستخدم في التوزيعات المقيدة بالاعتدالية	تستخدم في حالة التوزيعات الحرة غير المقيدة
تناسب البيانات الفترية interval والنسبية ratio	تناسب البيانات الاسمية nominal والرتبية ordinal كما يمكن استخدامها في حالة البيانات الفترية والنسبية
تشترط تجانس التباين في المجتمعات التي تسحب منها العينات	لا تشترط تجانس التباين
تعتبر أكثر قوة في رفض الفرضية الصفرية عندما تكون خاطئة عند توافر الشروط المطلوبة للاختبارات المعلمية	تعتبر اقل قوة وتزداد قوة الاختبار اللامعلمي بزيادة حجم العينة
تستخدم جميع المعلومات في العينة	لا تستخدم جميع المعلومات في العينة حيث أن الدرجات الخام يتم تحويلها إلى رتب ranks أو إشارات signs

4.3 الشروط المرجعية لتحديد نوع الاختبار المطلوب

لتحديد نوع الاختبار المناسب للبيانات المتوفرة على الباحث أن يكون ملما بالنواع المختلفة للمتغيرات قيد الدراسة، كما عليه أن يكون قادرا على تحديد المتغيرات المستقلة من المتغيرات التابعة، والنموذج أدناه يعطي توجيها واضحا لتحديد نوع الاختبار بعد تحديد أنواع المتغيرات وتصنيفها بين تابع ومستقل.



تمارين

أوجد حلول التارين التالية:

1. هل هناك علاقة تربط بين علامات الطلاب بين مادتي الاحصاء والرياضيات الموضحة في الجدول التالي، ما هو معامل الارتباط المناسب في هذه الحالة؟

الإحصاء x	13	9	19	15	11	8	16	11
الرياضيات y	15	7	17	15	10	9	14	10

2. أوجد معامل ارتباط الرتب لتقديرات الطلاب في كل من مادتي الاحصاء والرياضيات كما هو موضح في الجدول التالي:

الرياضيات x	A	C	C	C	B	D
الإحصاء y	B	B	D	C	A	E

3. أوجد معادلة خط الانحدار البسيط للبيانات في التمرين 1
4. اخذت عينة من 46 طالب من احدىالمدارس فوجدأن متوسط الطول هو 155 سم فإذا كان الانحراف المعياري للمجتمع يساوي 5 سماختبر الفرض عند مستوى الدلالة $\alpha = 0,05$ القائل :

$$H_0 : \mu = 160$$
$$H_a : \mu \neq 160$$

اذا علمت ان القيمة الجدولية Z 0.025 = 1.96

الفصل 4

البرامج المستخدمة في تحليل البيانات

هذا الفصل سيتناول وبصورة سريعة وموجزة أهم البرامج الحديثة المستخدمة في التحليل الاحصائي، مثل برمجية R، وبرنامج SAS، وبرنامج STATA، ونهاية الفصل سيتم عرض برمجية SPSS والتي تعد الأكثر رواجا واستخداما، كما وسيتم عرض بعض التطبيقات لأهم ما ورد في الفصول السابقة على عمليات التحليل الاحصائي.

الهدف العام للفصل:

- التعرف على اهم البرامج المستخدمة في التحليل الحصائي
- اجراء تطبيقات عملية على برنامج SPSS

1.4 برنامج R

كانت بداية R أكثر من عادية عندما حاول "Ross Ihaka" من جامعة أوكلاند بنيوزيلاندا إعادة كتابة بعض الدوال الإحصائية بلغة S ولغةS هذه هي أولى اللغات المختصة بالحوسبة الإحصائية في سبعينات القرن الماضي كبديل لفورتران .

"GNU R"أو المعروفة اختصارًا ب R هي لغة وبيئة تطوي (Language and development environment) متخصصة في تحليل وتمثيل البيانات والحوسبة الإحصائية . وتتكونR من حزمة رئيسية "core" يمكن توسيعها بواسطة حزم أخرى موجودة بمستودعات المشروع الرئيسي ."CRAN" يبلغ عدد هذه الحزم لحد الآن

2456 .وهي تقدم مكتبات للطرق الإحصائية الأساسية والمقدمة كالإحصاء الوصفي، والاختبارات الإحصائية، وتخطيط التجارب، وتحليل الارتباطات الخطية، واللاخطية، وتحليل المتتاليات الزمنية، والتحليل متعدد المتغيرات، وتحليل الإشارات. في السنوات العشر الأخيرة، تجاوزت R الدوائر الأكاديمية إلى قطاعات تكنولوجية متعددة حيث تقوم باستعمالها شركات كبرى بما في ذلك غوغل ومايكروسوفت وموقع فيسبوك لتحليل البيانات.

2.4 برنامج SAS

ان برنامج SAS هو مختصر لـــ (Statistical Analysis System) اذ نفذ وطور من قبل مجموعة من الباحثين في معهد SAS في الولايات المتحدة الامريكية ، وبيعت اول نسخ له عام 1976 ويجرى عليه تطويرمستمر حتى وصلت آخرنسخة معدلة منه الى version 9.2 .

يستعمل هذا البرنامج لاجراء معظم التحليلات الاحصائية ولجميع التصاميم (التصميم العشوائي الكامل ، تصميم القطاعات العشوائية الكاملة ، المربع اللاتيني فضلا عن تصاميم التجارب العاملية للانواع الثلاث المذكورة). كما يمكن استعماله لاجراء مختلف الاختبارات الاحصائية المعلمية واللامعلمية مثل T-Test (Least significance Differences) ncan واختبار Tukey واختبار مربع كاي واختبار ولكاكسن وماكنيمار وتقدير حدود الثقة، كما يستعمل ايضا لتقديرمعامل الارتباط والانحدار البسيط والمتعدد والانحدار اللوجستيكي وتقدير نسبة الخطورة (Ratio Risk) ونسبة الارجحية (OddsRatio) واختبارات التشخيص فضلا عن امكانية الحصول على رسوم بيانية.كما يمكن اعتماده لغرض تقدير المكافيء الوراثي والارتباطات الوراثية والبيئية والمظهرية علاوة على تقدير قيم BLUP الخاصة بدراسات التقييم الوراثي للحيوانات.

3.4 برنامج STATA

STATA هو برنامج تحليل البيانات التفاعلية التي تعمل على مجموعة متنوعة من المنصات، ويمكن استخدامه في كل التحليلات الإحصائية البسيطة والمعقدة. يستخدم STATAواجهة يمكن التحكم بها بالفأرة وبكتابة الأوامر أيضًا، مما يجعله سهل الاستخدام. كما يسهل إنشاء الرسوم البيانية ومخططات البيانات والنتائج باستخدام برنامج STATA

يتركز التحليل في STATA حول أربع نوافذ: نافذة الأوامر، نافذة المراجعة، نافذة النتيجة، ونافذة المتغيرات. يتم إدخال أوامر التحليل في نافذة الأوامر حيث تسجل نافذة المراجعة تلك الأوامر. تسرد نافذة المتغيرات كل المتغيرات التي تتوفر في البيانات الحالية وتعرضها مع تسمياتها، أما نافذة النتائج فهي المكان الذي تظهر فيه النتائج.

4.4 برنامج SPSS

يعتبر برنامج التحليل الإحصائي SPSS أحد البرامج الإحصائية التي لاقت شيوعاً في استخدامها من قبل الباحثين للقيام بالتحليلات الإحصائية، ويستخدم البرنامج في كثير من المجالات العلمية والتي تشمل على سبيل المثال، العلوم الإدارية والاجتماعية والهندسية والزراعية. وكلمة SPSS هي اختصار للمسمى الكامل للبرنامج وهو: "Statistical Package for Social Sciences" والتي تعني " البرنامج الإحصائي للعلوم الاجتماعية".

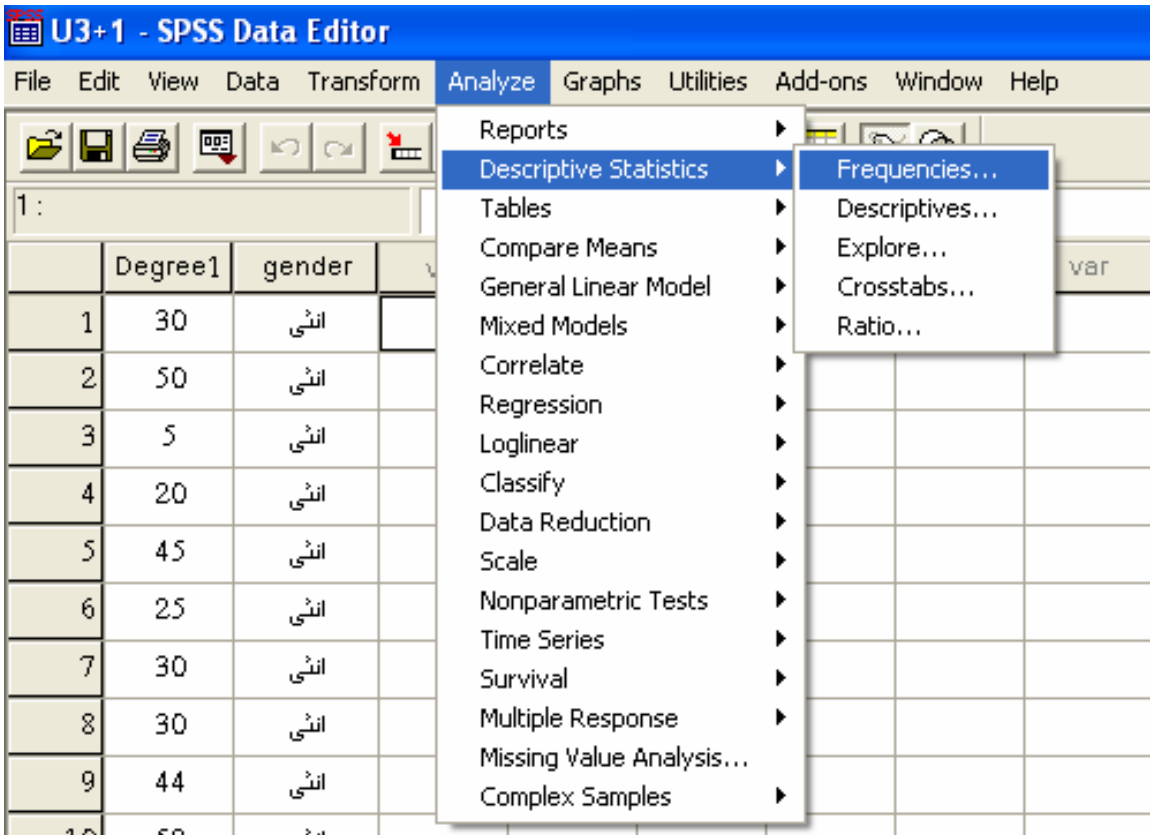
هناك عدة إصدارات من برنامج SPSS، بدأت شركة SPSS إعداد هذا البرنامج الذي كان يعمل تحت نظام التشغيل MS_DOS وقد تم تطويره ليعمل ضمن نظام التشغيل Windows في عام 1993 متلافياً بذلك الصعوبات التي كانت تواجه مستخدمي هذا البرنامج ضمن نظام التشغيل MS_DOS.

توالت الإصدارات من 6، ثم 7، 7.5، 9، 10، 11.5، 12، 13، 14، 15، 16، 17 ،18، 19 ، 20 وكان آخر إصدار هو الإصدار 22.

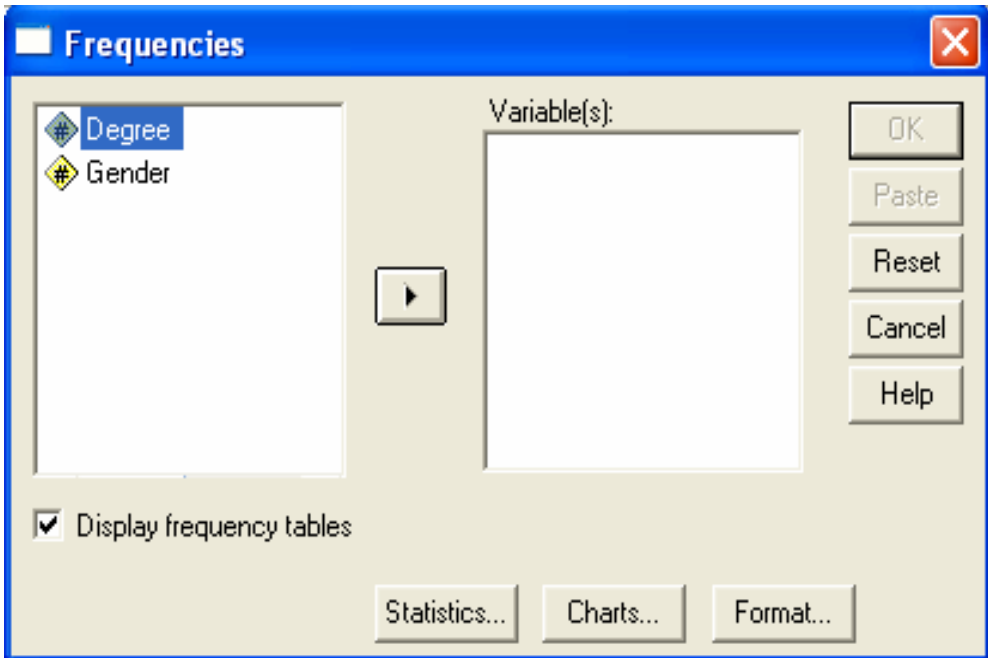
1.4.4 تطبيقات عملية على برنامج SPSS

- ✓ **لحساب مقاييس النزعة المركزية من خلال البرنامج**

من قائمة التحليل Analyze اختار القائمة الفرعية للإحصاءات الوصفية Descriptive Statistic ثم عليك ان تختار أمر التكراراتFrequency كما في الشكل التالي

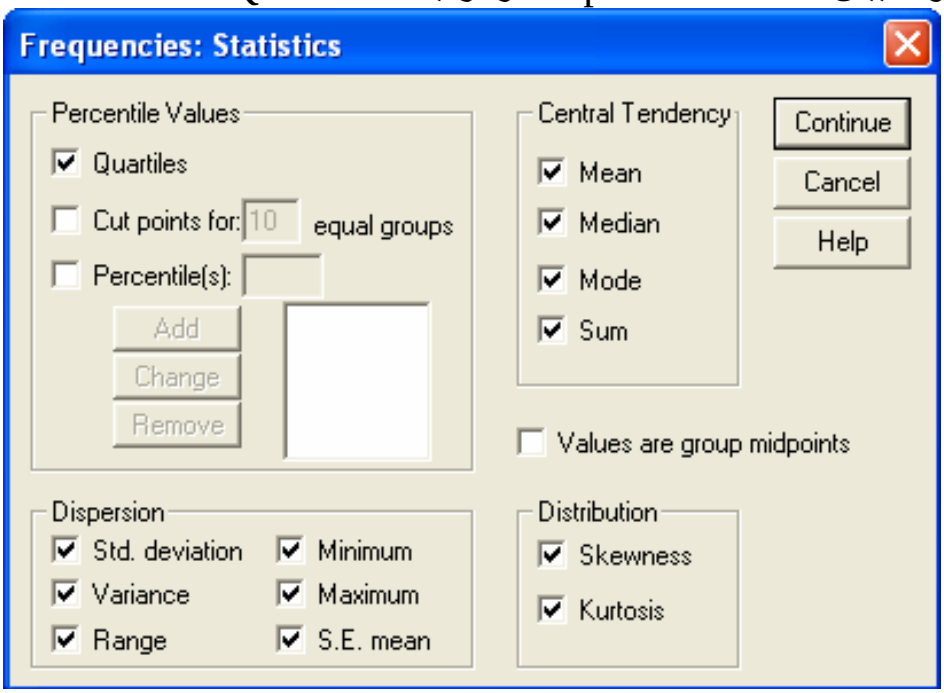


يظهر المربع الحواري التالي



ومن ثم ننقل المتغيرDegree لخانة المتغيرات Variable وذلك من خلال الضغط بالماوس على المتغير

أو من خلال الضغط مرتين بالماوس على المتغيرDegree
ثم بالضغط على زرStatistics يظهر المربع الحواري التالي ونختار منه حساب المقاييس الاحصائية التي تقع تحت مقاييس النزعة المركزية Central Tendency ومقاييس التشتت Dispersion والربيعات Quartile



✓ **خطوات اختبار الفرضيات في برنامج SPSS**

- تحديد نوع توزيع المجتمع، يجب تحديد ما إذا كان المتغير العشوائي الذي يتم دراسته يتبع التوزيع الطبيعي أم لا.
- صياغة فرضيتا العدم (الصفرية) والبديلة

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

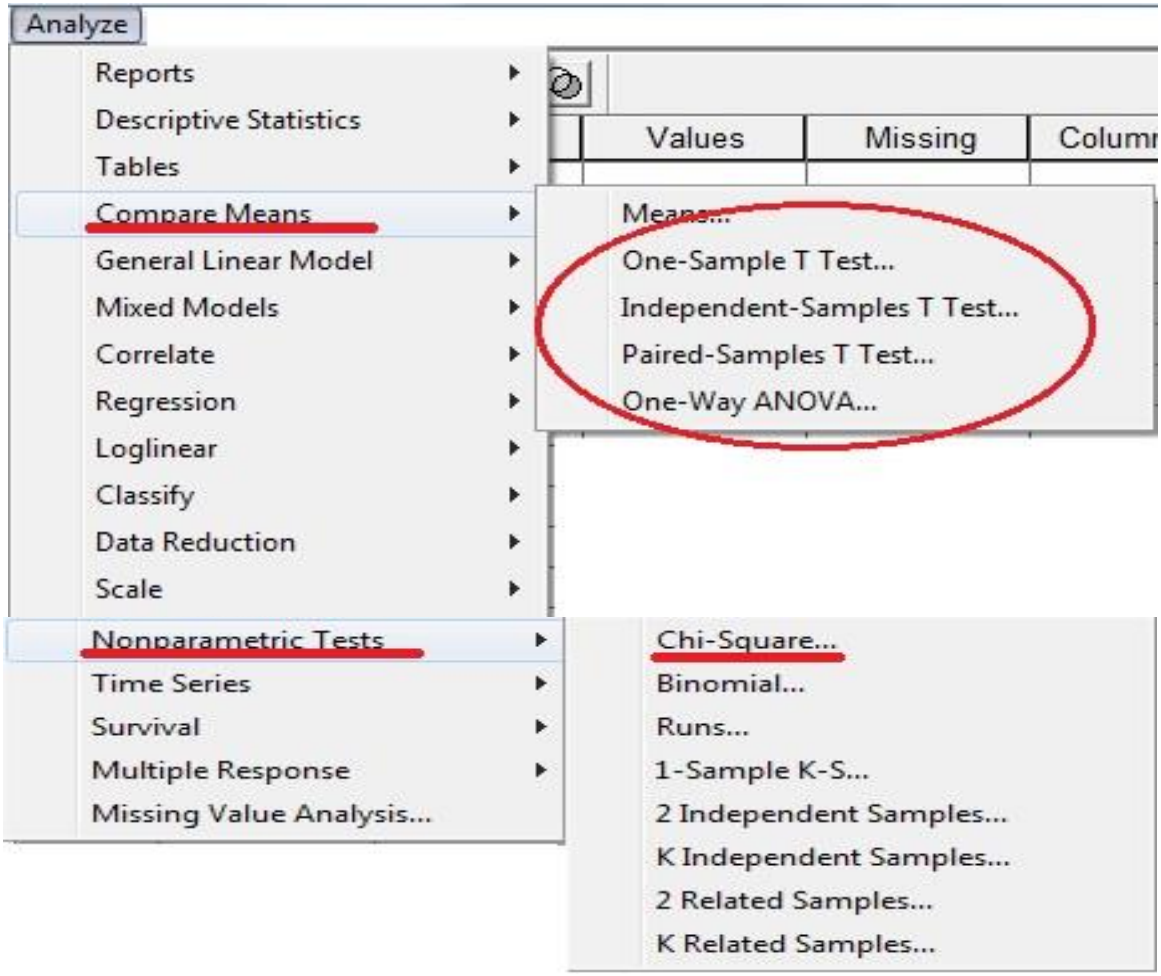
$$H_0 : \mu = \mu_0$$
$$H_a : \mu \neq \mu_0$$

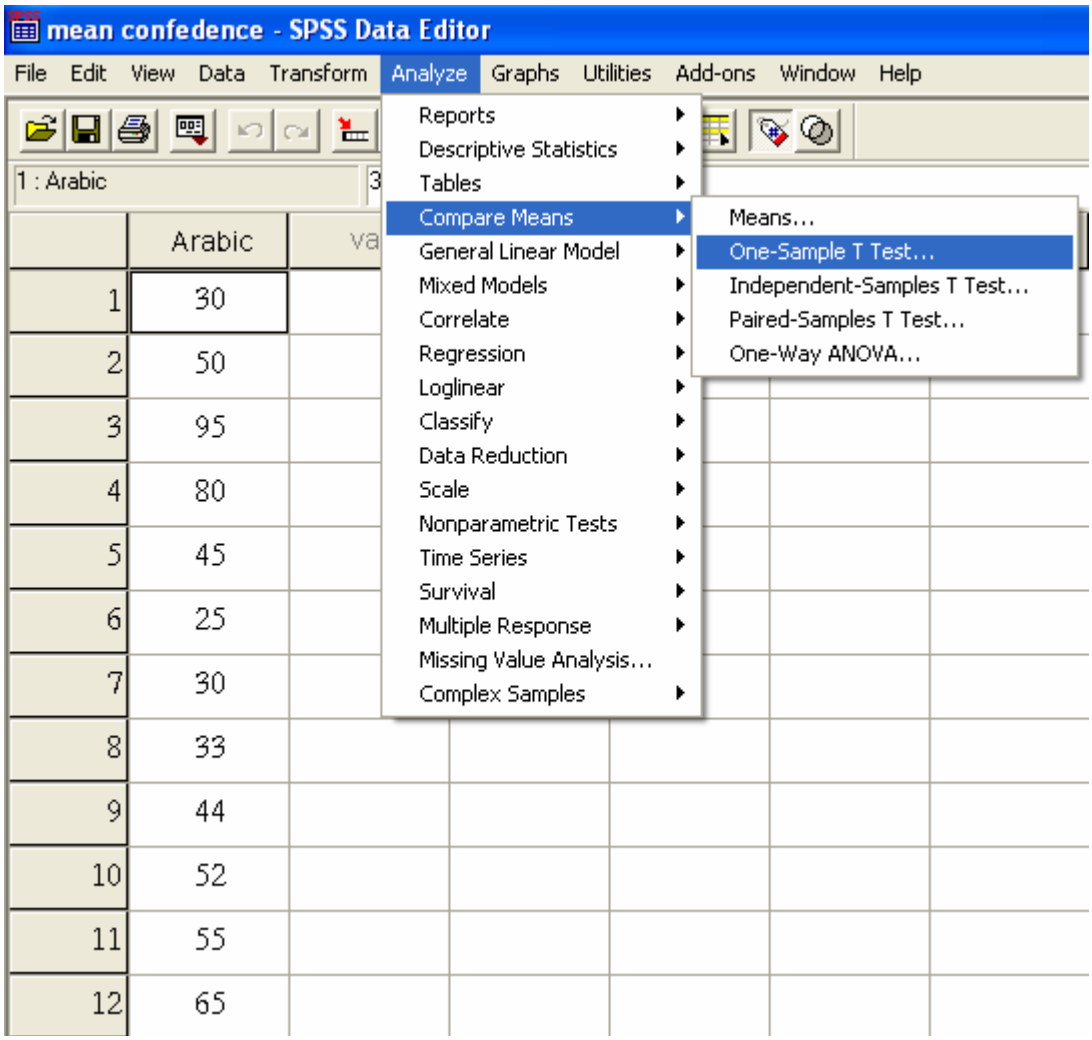
- تحديد مستوى المعنوية α
- اختيار دالة الاختبار الإحصائية المناسبة
- جمع البيانات من العينة وحساب قيمة دالة الاختبار الإحصائية
- اتخاذ القرارات

إذا كانت قيمة P-values (Sig.) أقل من قيمة α هذا يعني أننا نرفض الفرضية العدمية ونقبل الفرضية البديلة.

دائما قيمة α الإفتراضية في SPSS تساوي 0.05 أي أن اختبارات الثقة تجري عند مستوى ثقة 95%.

الشكل التالي يوضح كيفية الوصول للاختبارات في برنامج SPSS





التالي يوضح النتائج التي يتم استخراجها باستخدام حزمة SPSS

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Arabic	30	63.40	21.021	3.838

One-Sample Test

	Test Value = 68					
	t	df	Sig. (2-tailed)	Mean Difference	90% Confidence Interval of the Difference	
					Lower	Upper
Arabic	-1.199	29	.240	-4.600	-11.12	1.92

يتكون ناتج الاختبار من جدولين، الجدول الأول يعطي بعض المعلومات الوصفية عن المتغير
كالعدد والوسط الحسابي والانحراف المعياري، اما الجدول الثاني يعطي ناتج اختبار T لوسط عينة واحدة.

الجدول الأول One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Arabic	30	63.40	21.021	3.838

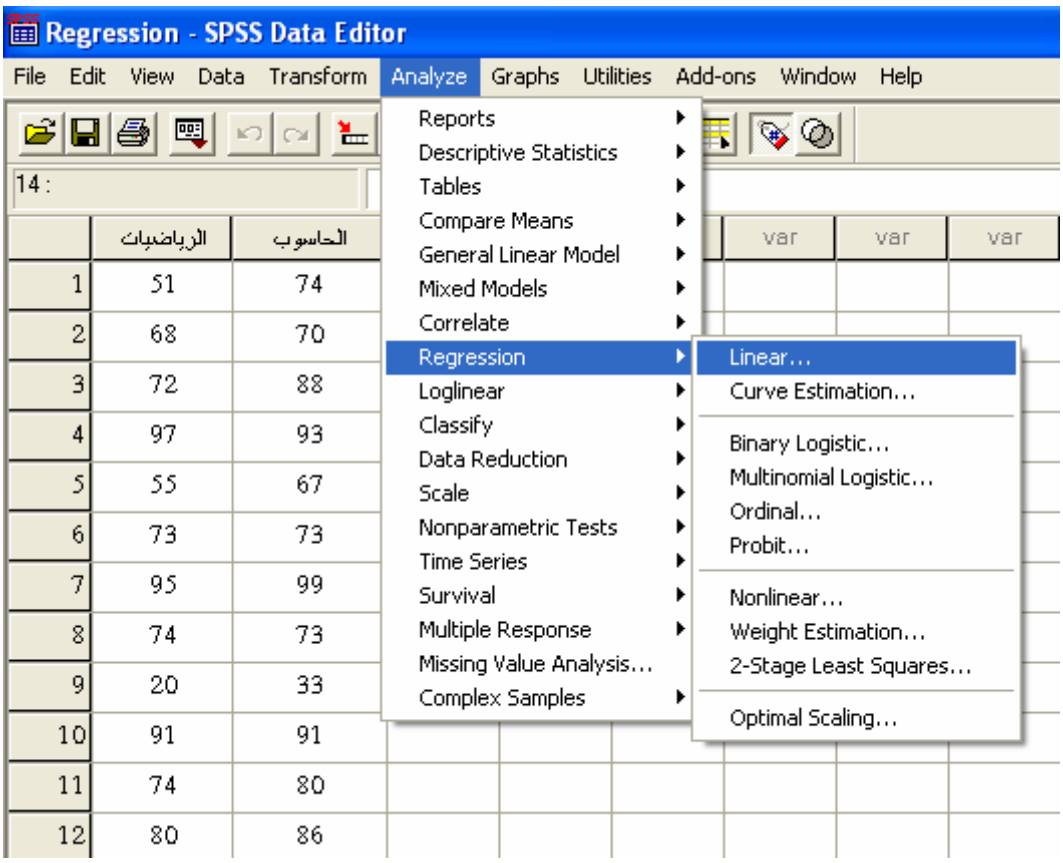
الجدول الثاني One-Sample Test

	Test Value = 68					
	t	df	Sig. (2-tailed)	Mean Difference	90% Confidence Interval of the Difference	
					Lower	Upper
Arabic	-1.199	29	.240	-4.600	-11.12	1.92

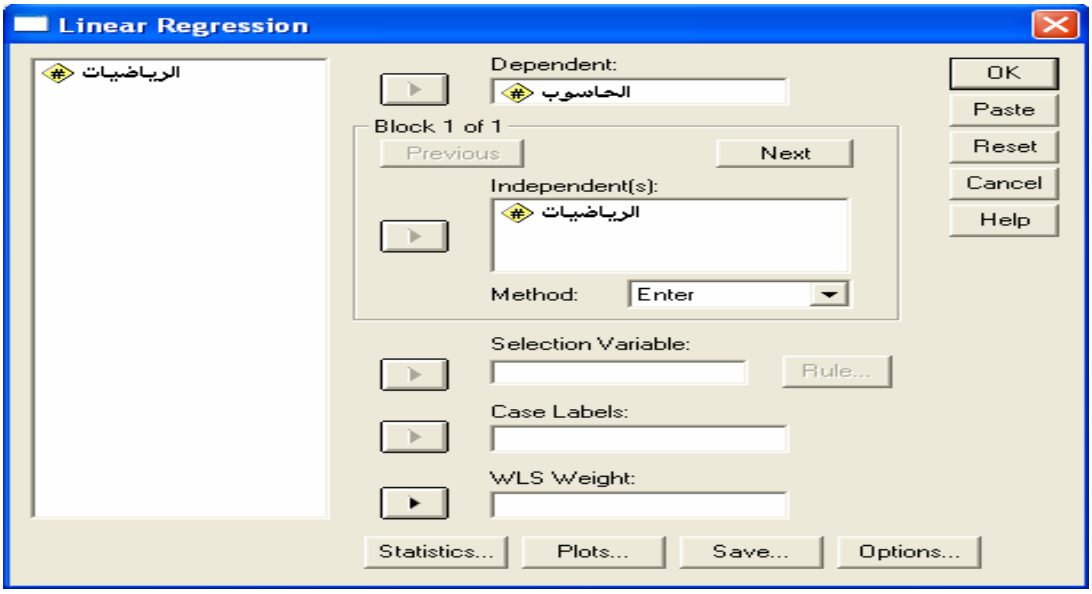
يبين أن القيمة التي تم الاختبار بناءًا عليها هي 68 وأن قيمة اختبار T لوسط مجتمع هي- 1.199بدرجة حرية مقدارها29 وان الفرق بين وسط العينة والقيمة المفحوصة هو-4.6
وبناء على قيمة Sig أو ما تسمى p-value والتي يبين الجدول انها 0.24 وهي اكبر من قيمة α =0.05 أي اننا نقبل فرضية العدم.

- ✓ **تقدير دالة الانحدار باستخدام البرنامج**

نتبع الخطوات التي يوضحها الشكل ادناه للوصول الى نوافذ الادخال الخاصة بدالة الانحدار



ثم نحدد المتغير المستقل والمتغير التابع حسب المثال في الشكل



نتائج تحليل الانحدار الخطي

Regression

Variables Entered/Removed[b]

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

a. All requested variables entered.

b. Dependent Variable: الحاسوب

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.935[a]	.874	.862	6.411

a. Predictors: (Constant), الرياضيات

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2861.180	1	2861.180	69.603	.000[a]
	Residual	411.070	10	41.107		
	Total	3272.250	11			

a. Predictors: (Constant), الرياضيات

b. Dependent Variable: الحاسوب

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	23.783	6.671		3.565	.005
	الرياضيات	.755	.090	.935	8.343	.000

a. Dependent Variable: الحاسوب

الجدول الاول
وهو يبين أن طريقة المربعات الصغرى هي المتبعة في تحليل الانحدار الخطي وأن المتغير
المستقل هو الرياضيات وأن المتغير التابع هو الحاسوب.
الجدول الثاني

يبين أن الارتباط قوي جدا بلغت قيمته 0.935 وأن مدى الدقة في المتغير التابع بلغت 87%

معامل الارتباط

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.935[a]	.874	.862	6.411

a. Predictors: (Constant), الرياضيات

الجدول الثالث

ANOVA[b]

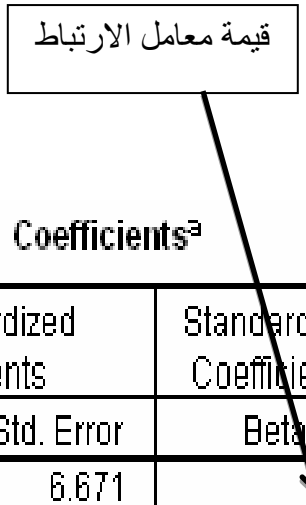
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2861.180	1	2861.180	69.603	.000[a]
	Residual	411.070	10	41.107		
	Total	3272.250	11			

a. Predictors: (Constant), الرياضيات

b. Dependent Variable: الحاسوب

الجدول أعلاه هو جدول تحليل تباين خط الانحدار حيث يدرس مدى ملائمة خطة الانحدار للبيانات وفرضيته الصفرية

الجدول الرابع



Coefficients[a]

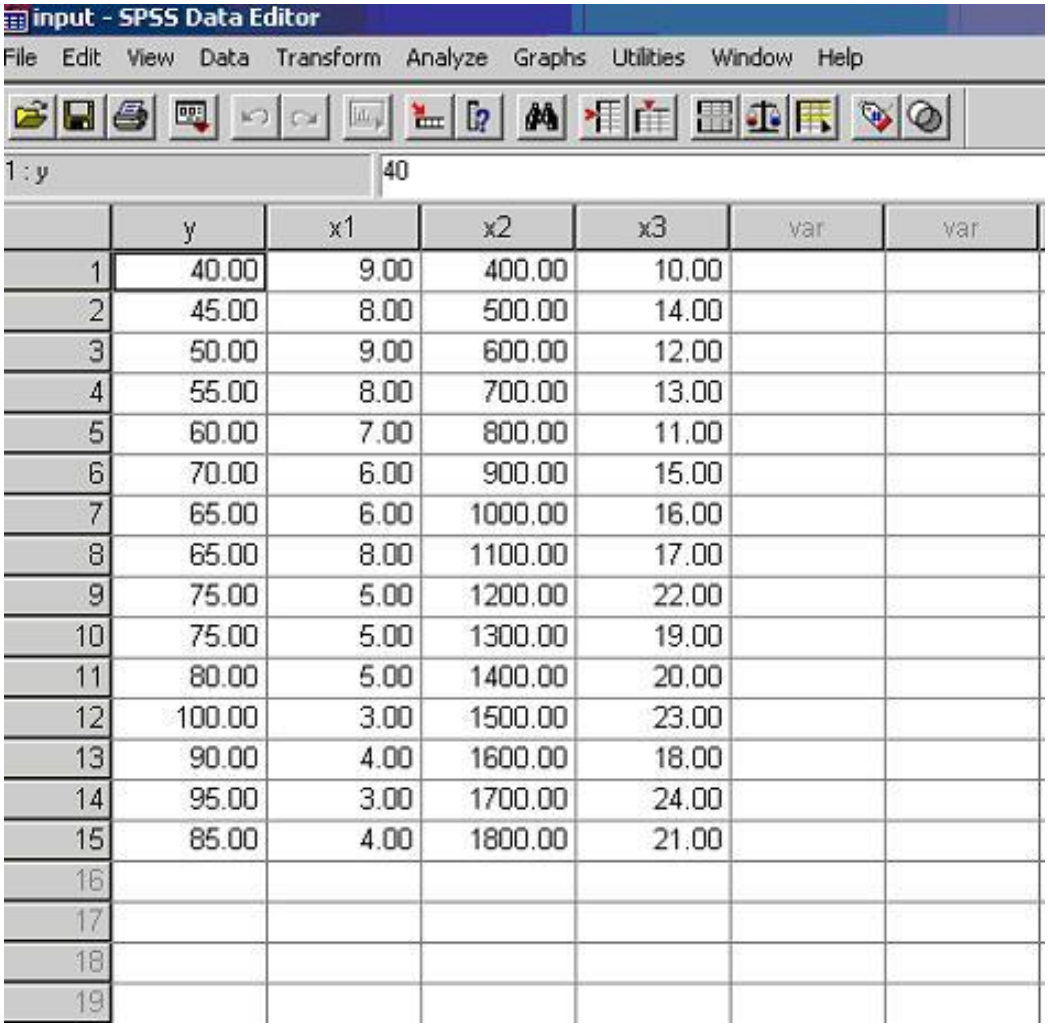
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	23.783	6.671		3.565	.005
	الرياضيات	.755	.090	.935	8.343	.000

a. Dependent Variable: الحاسوب

اذن دالة الانحدرار هي $Y= 23.683+0.755x$

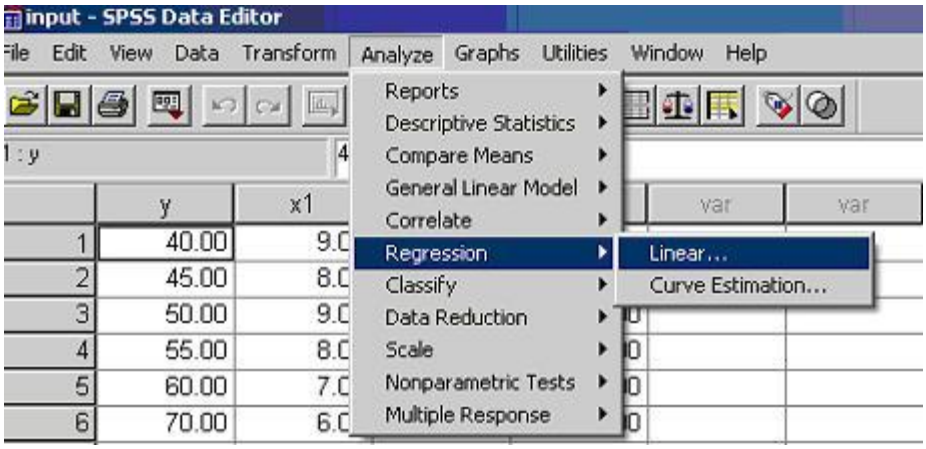
✓ **تطبيق الإنحدار الخطي المتعدد على البرنامج Multiple Linear Regression**

يوضح المثال في الجدول التالي الطريقة لإيجاد معادلة خط الإنحدار المتعدد للبيانات المعلقة بأختبار (15) طالباً إذ تمثل هذه البيانات العلاقة بين (أختبار الوثب العريض من الثبات Y) والعوامل المؤثرة عليها وهي (أختبار الشد لأعلى على العقلة X1) و (أختبار السعة الحيوية X2) و (أختبار الوثب العمودي من الثبات X3) ، وبإستخدام البرنامج الاحصائي SPSS سيتم الحصول على نتائج تقدير معادلة الانحدار الخطي المتعدد وكما يلي :

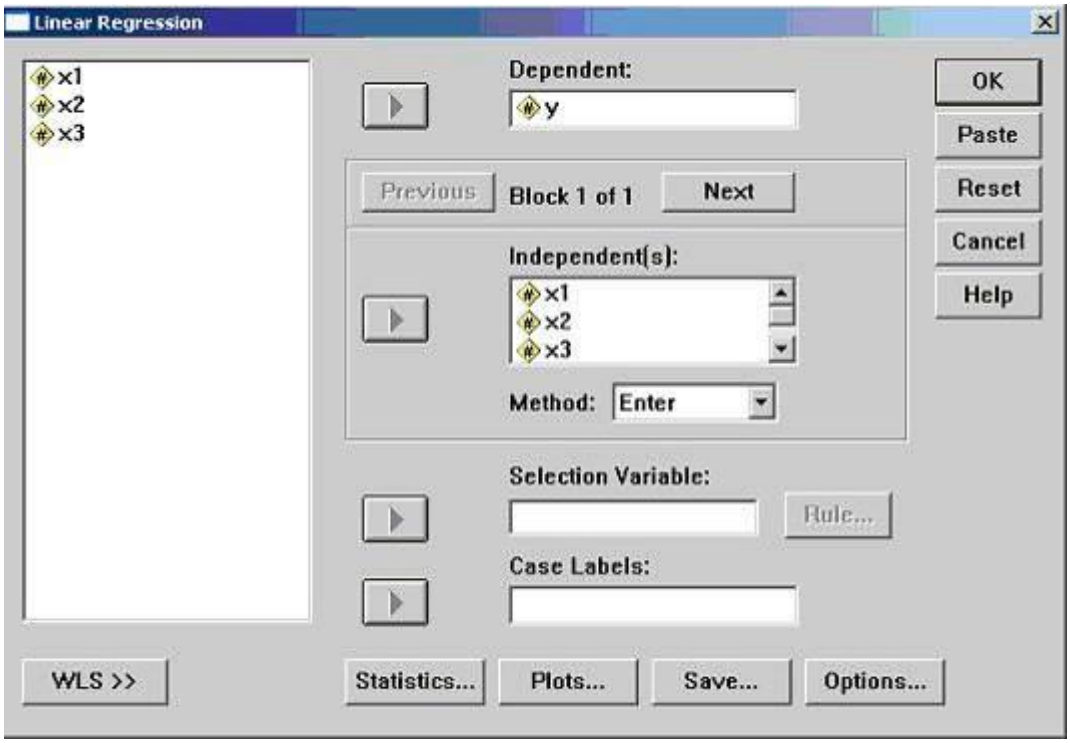


	y	x1	x2	x3	var	var
1	40.00	9.00	400.00	10.00		
2	45.00	8.00	500.00	14.00		
3	50.00	9.00	600.00	12.00		
4	55.00	8.00	700.00	13.00		
5	60.00	7.00	800.00	11.00		
6	70.00	6.00	900.00	15.00		
7	65.00	6.00	1000.00	16.00		
8	65.00	8.00	1100.00	17.00		
9	75.00	5.00	1200.00	22.00		
10	75.00	5.00	1300.00	19.00		
11	80.00	5.00	1400.00	20.00		
12	100.00	3.00	1500.00	23.00		
13	90.00	4.00	1600.00	18.00		
14	95.00	3.00	1700.00	24.00		
15	85.00	4.00	1800.00	21.00		
16						
17						
18						
19						

نذهب إلى قائمة analyze ونختار منها الأمر Regression ومن القائمة الفرعية نختار Linear ، كما في الشكل الاتي :



من نافذة تحليل الإنحدار نقوم بتحديد المتغير التابع (Y) وننقله إلى خانة المتغير التابع ثم نحدد المتغيرات المستقلة وننقلها إلى خانة المتغيرات المستقلة ثم ننقر OK كما في الشكل الآتي :



سوف نحصل على شاشة المخرجات الآتية :

Regression

Variables Entered/Removed[b]

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

a. All requested variables entered.

b. Dependent Variable: Y

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.975[a]	.951	.938	4.52761

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

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4374.508	3	1458.169	71.133	.000[a]
	Residual	225.492	11	20.499		
	Total	4600.000	14			

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

b. Dependent Variable: Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	79.106	19.782		3.999	.002
	X1	-4.928	1.611	-.563	-3.059	.011
	X2	1.590E-02	.007	.392	2.146	.055
	X3	.175	.637	.043	.275	.789

a. Dependent Variable: Y

نستنتج من المخرجات أعلاه ما يلي :

- من الجدول الأول نلاحظ بأن طريقة الإنحدار المستخدمة وهي طريقة Enter حيث يتبين ان البرنامج قام بادخال جميع المتغيرات المستقلة في معادلة الإنحدار الخطي المتعدد .
- من الجدول الثاني نلاحظ بأن قيم معامل الارتباط الثلاثة وهي معامل الارتباط البسيط R قد بلغ (0.975) بينما بلغ معامل التحديد R2 (0.951) في حين كان معامل التحديد المصحح -R2 (0.938) مما يعني بأن المتغيرات المستقلة التفسيرية (أختبار الشد لأعلى على العقلة ، أختبار السعة الحيوية ، أختبار الوثب العمودي من الثبات) استطاعت ان تفسر (0.94) من التغيرات الحاصلة في (أختبار الوثب العريض من الثبات) المطلوبة والباقي (0.06) يعزى إلى عوامل اخرى .
- كما يلاحظ في الجدول الثالث بأنه يتضمن قيم تحليل التباين والذي يمكن المعرفة من خلاله على القوة التفسيرية للنموذج ككل عن طريق إحصائية F وكما يلاحظ من جدول تحليل التباين المعنوية العالية لأختبار F ($P < 0.0001$) مما يؤكد القوة التفسيرية العالية لنموذج الإنحدار الخطي المتعدد من الناحية الإحصائية .
- أما في الجدول الرابع والأخير فيلاحظ قيمة الثابت ومعاملات الإنحدار ودلالتها الإحصائية للمتغيرات المستقلة على المتغير التابع ويمكن تلخيص هذه الجدول بالشكل الآتي :

المتغير التابع	المتغيرات المستقلة			
Y	B الحد الثابت غير المعياري	X1	X2	X3
قيمة المعامل	79.106	4.928 -	0.015	0.175
قيم اختبار T	93.99	-3.059	2.146	0.275
المعنوية	0.002	0.01	0.055	0.789

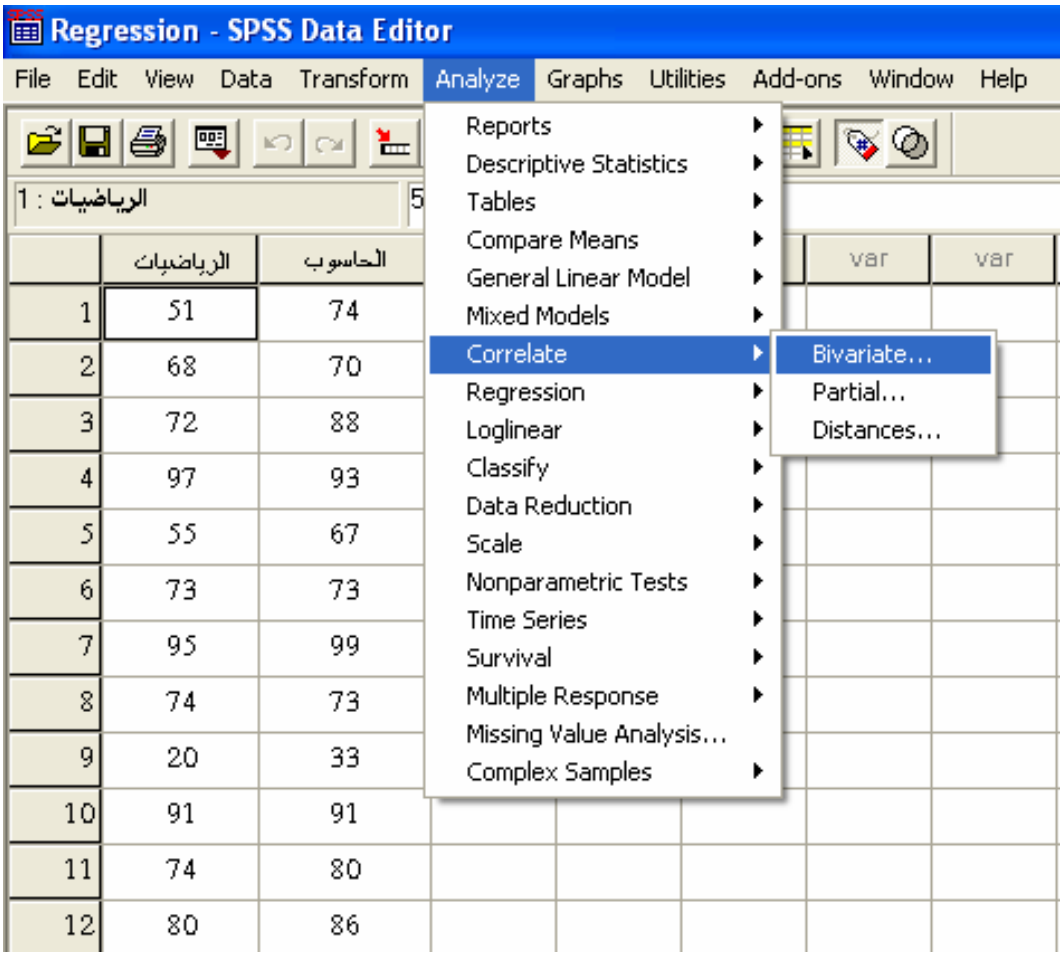
من الجدول نستنتج ان المتغيرات المستقلة (أختبار الشد لأعلى على العقلة) كان معنوياً من الناحية الإحصائية وحسب أختبار t (عند مستوى معنوية $P \leq 0.05$) ، في حين كاد (أختبار السعة الحيوية) أن يكون معنوياً (عند مستوى معنوية $P \leq 0.05$) إلا أن المتغير المستقل (أختبار الوثب العمودي من الثبات) لم يكن ذو تأثيراً معنوياً في نموذج الإنحدار المتعدد وحسب أختبار t ، ومن الجدول الرابع والأخير يمكن التوصل إلى معادلات الإنحدار باستخدام Beta غير المعيارية (الحد الثابت) وكما يلي :

- إن معادلة خط إنحدار (الوثب العريض من الثبات) على أختبارات (الشد لأعلى على العقلة والسعة الحيوية والوثب العمودي من الثبات) هي :

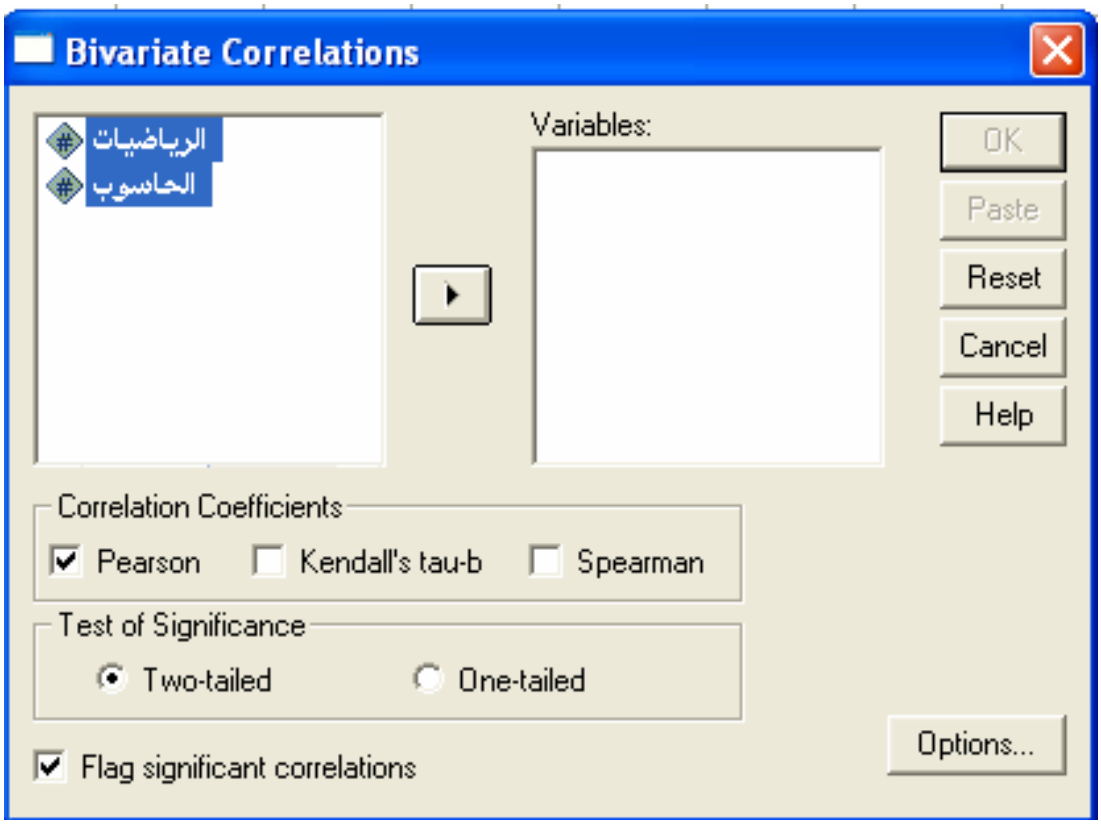
الوثب العريض من الثبات = 79.106 + 4.928 × الشد لأعلى على العقلة + 0.015 × السعة الحيوية + 0.175 × الوثب العمودي من الثبات .

✓ **تطبيقات معاملات الارتباط (بيرسون، سبيرمان) على البرنامج**

من قائمة التحليل Analyze نختار القائمة الفرعية للإرتباط Correlation ومن ثم نختار Bivariate كما يوضحه الشكل التالي



ثم يظهر مربع الحوار التالي والذي من خلاله يمكن ان نحدد معامل الرتباط المناسب للبيانات



نتائج معملات الارتباط

Correlations

معامل الارتباط

Correlations

		الرياضيات	الحاسوب
الرياضيات	Pearson Correlation	1	.935**
	Sig. (2-tailed)	.	.000
	N	12	12
الحاسوب	Pearson Correlation	.935**	1
	Sig. (2-tailed)	.000	.
	N	12	12

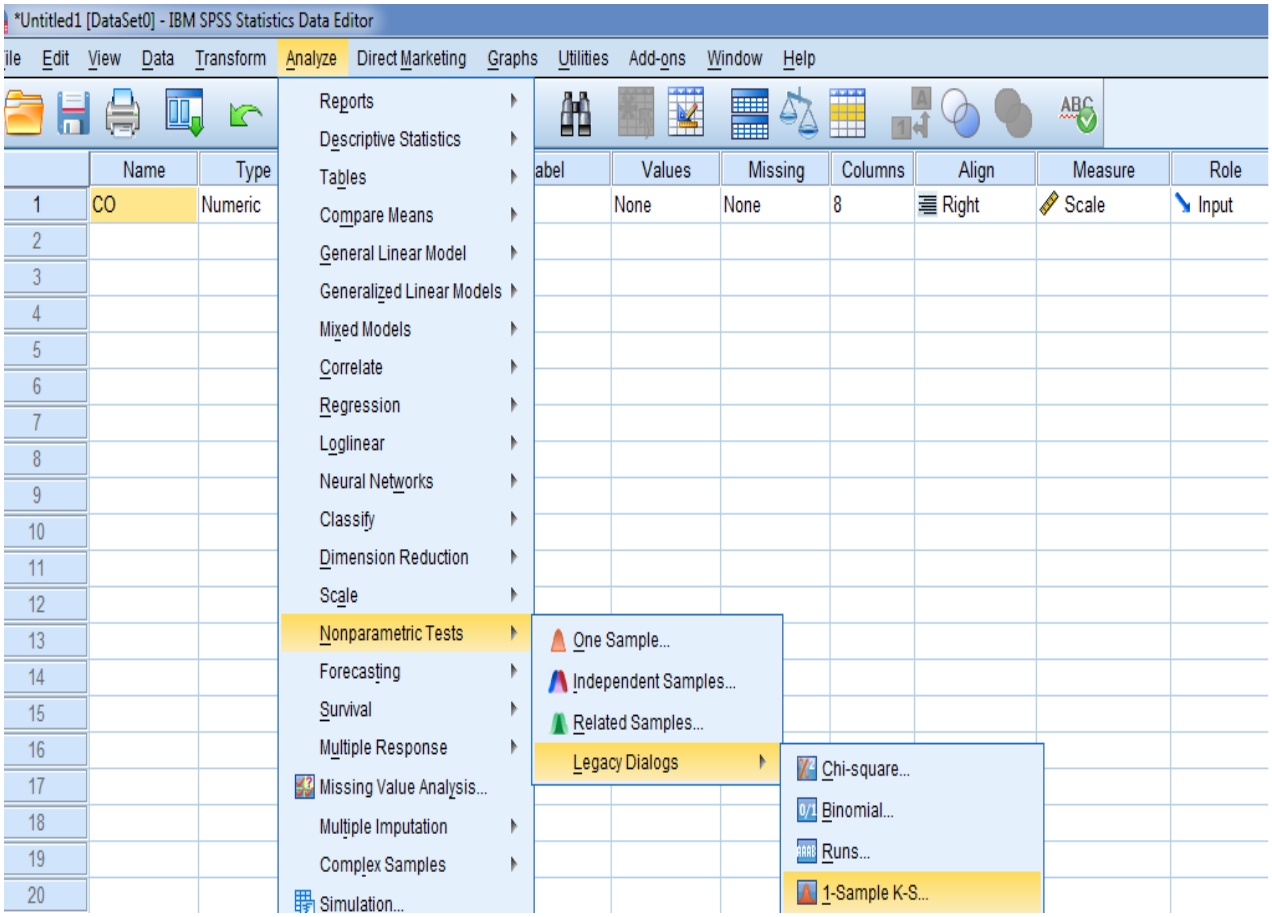
**. Correlation is significant at the 0.01 level (2-tailed).

✓ **تطبيقات الاختبارات اللامعلمية على برنامج SPSS**

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

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

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



ثم يظهر مربع حوار نحدد من خلاله نوع التوزيع المراد اختباره وهو هنا التوزيع الطبيعي
لتظهر النتائج التالية

One-Sample Kolmogorov-Smirnov Test

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

a. Test distribution is Normal.

b. Calculated from data.

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

تمارين

اوجد حلول التمارين التالية:

1. اذا كان عمر الطلاب في الجامعات الأردنية هو 19 سنة، قم بصياغة فرضية لاختبار صحة هذا الإدعاء وحدد نوع الاختبار الملطوب.
2. لدراسة العلاقة بين الدخل والاستهلاك بالدنانير في مدينة ما، أخذت عينة مكونة من عشرة أسر فأعطت النتائج التالية:

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

المطلوب: إيجاد نموذج انحدار الاستهلاك على الدخل باستخدام برنامج SPSS

3. فيما يلي تقديرات عشرة من طلاب في امتحان مادتي الرياضيات والإحصاء:

الرياضيات	راسب	جيد	مقبول	جيـــد جداً	مقبول	مقبول	جيد	جيد	جيد جداً	جيد
الإحصاء	مقبول	جيد جداً	جيد	ممتاز	راسب	جيد	جيد جداً	جيد	ممتاز	راسب

المطلوب: احسب معامل الارتباط سبيرمان بين تقديرات المادتين باستخدام برنامج SPSS.

4. هل تتوزع بيانات الدخل في التمرين 2 أعلاه؟ اختبر هذا الادعاء باستخدام اختبار one-sample Kolmoyorov-Smirnov Test

المصطلحات

علم الإحصاء :

هو العلم الذي يبحث في جمع البيانات والتقنيات المختلفة لتنظيم وتصنيف وعرض هذه البيانات وتلخيصها في صورة مؤشرات رقمية لوصف وقياس خصائصها الأساسية، وتحليلها بغرض اتخاذ القرارات المناسبة.

المجتمع Population :

هو المجموعة الكلية لمفردات الدراسة سواء كانت أفراد أو أشياء، واستخلاص خصائص هذا المجتمع هو الهدف النهائي للدراسة الإحصائية .

العينة Sample :

هي مجموعة جزئية من مفردات المجتمع محل الدراسة يتم اختيارها بحيث تكون ممثلة للمجتمع تمثيل صحيح.

المؤشر هو عنصر من البيانات الإحصائية يمثل بيانات إحصائية في وقت وزمان محددين إضافة لخصائص أخرى

الوسط الحسابي Arithmetic Mean أوAverage

الوسط الحسابي لمجموعة من القيم هو القيمة التي لو أعطيت لكل مفردة من مفردات المجموعة لكان مجموع القيم الجديدة مساويا لمجموع القيم الأصلية. ويستخدم الوسط الحسابي في حالة البيانات الرقمية فقط.

الوسيط:Median

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

المنوال:Mode

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

الفرضية الإحصائية

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

مستوى المعنوية أو مستوى الاحتمال

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

دالة الاختبار الإحصائية

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

القيمة الاحتمالية: (Sig. or P-value)

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

بيانات نوعية:
وهي البيانات التي لا يمكن قياسها رقميا (عدديا) وتقسم البيانات النوعية إلى بيانات نوعية أسمية، و بيانات نوعية ترتيبية

بيانات كمية:
وهي البيانات التي يمكن قياسها رقميا وتأخذ دائما قيما عددية صحيحة أو كسرية حسب ظروف الحالة حيث تقسم البيانات الكمية إلى بيانات كمية منقطعة و بيانات كمية مستمرة(المتصلة)

المؤشر الإحصائي:
هو عنصر من البيانات الإحصائية يمثل بيانات إحصائية في وقت وزمان محددين إضافة لخصائص أخرى

قائمة القراءات

- بري، عدنان ماجد عبدالرحمن(1998)، **اساسيات طرق التحليل الاحصائي** ، جامعة الإمام محمد بن سعود الإسلامية, المملكة العربية السعودية.
- هيكل، عبد العزيز فهمي (1992)، **طرق التحليل الإحصائي**، دار النهضة العربية للطباعة والنشر والتوزيع.
- العقيلي، صالح أرشيد (**1998**)، **التحليل الاحصائي باستخدام البرنامج - spss**، دار الشروق للنشر والتوزيع, الأردن.
- النعيمي، أسوان محمد طيب **(2009)، معالجة البيانات غير التامة وتقديرها بطريقة انحدار المركبات الرئيسية**، جامعة الموصل
- أحمد، صفاء سامي **(2008)، بحث المعاينة والمسوح الهاتفية، المؤتمر الدولي الثاني لاستطلاعات الراي العام**، الجهاز المركزي للتعبئة العام والإحصاء، مصر
- Blaisdell, E. (2001). *Statistics in Practice*. Saunders College Publishers.

– Neeter, J, Waserman, Whitmare, (1993): Applied Statistics. 4th Edition, Louise Richardson.

قائمة المراجع

– فاضل، سمير سليم (2002)، **تصميم التجارب والتحليل الإحصائي** ، دار شموع الثقافة.

– صبري، عزام (2003)، **التحليل الاحصائي بين النظرية والتطبيق** - spss ، عالم الكتب الحديث.

-محمد، اماني موسى (2007)، **التحليل الاحصائي للبيانات، مركز تطويرالدراسات العليا والبحوث**، القاهرة

-نشوان، عماد (2005)، **الدليل العلمي لمقرر الاحصاء التطبيقي**، جامعة القاهرة.

– *Amany Mousa: Cairo (2005), Statistical Data Analysis, Center for Advancement of Postgraduate Studies and Research, Faculty of Engineering, Cairo University*.

– P.S. Mann. (2003). *Introductory Statistics*. John Wiley & Sons, 5th edition.

– Keller, G and Waracck, B (2001): Statistics for Management and Economics 6th Edition Duxbury.