ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/379046128

# التحليل الاحصائي للبيانات: مفاهيم احصائية وتطبيقات في البرنامج SPSS

**Book** · March 2024

CITATIONS
0

READS
2,838

**1 author:**

Khalid Hassan
University of Diyala
**103** PUBLICATIONS **162** CITATIONS

SEE PROFILE

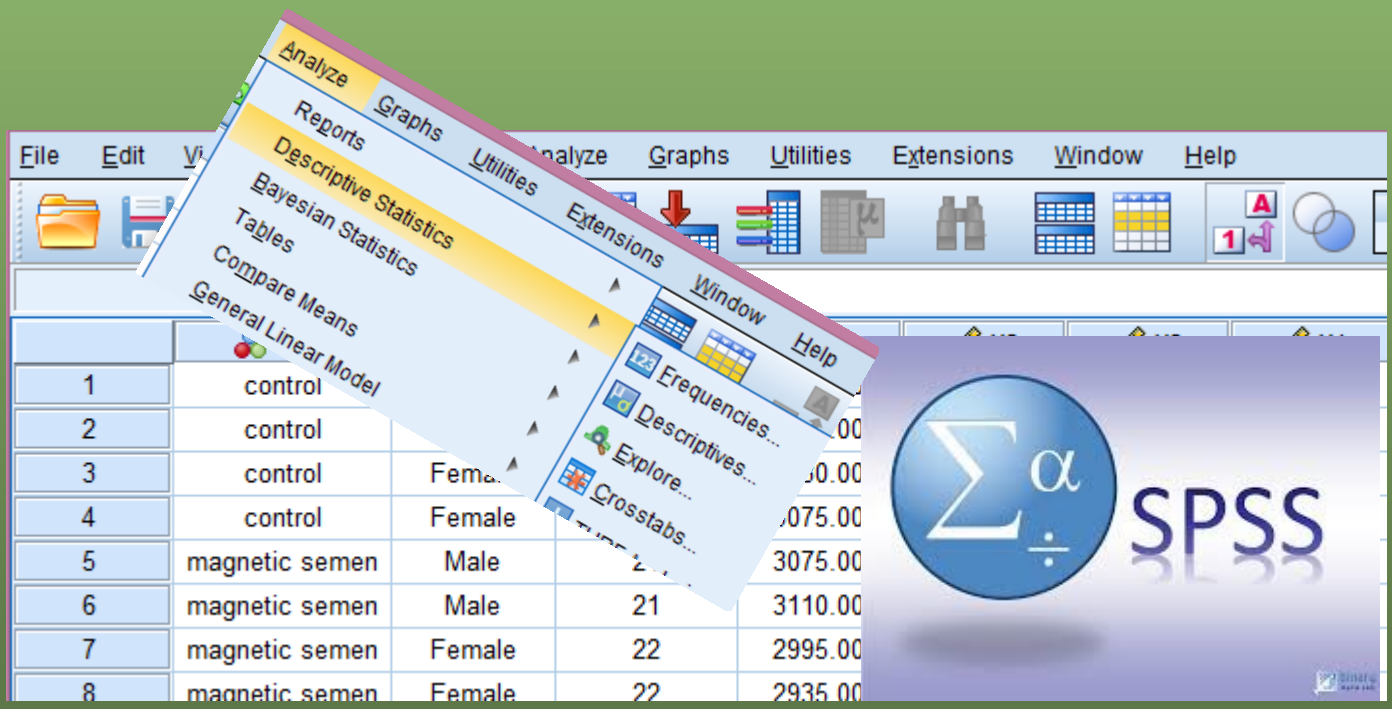
All content following this page was uploaded by Khalid Hassan on 18 March 2024.

The user has requested enhancement of the downloaded file.

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة ديالى

# التحليل الاحصائي للبيانات: مفاهيم احصائية وتطبيقات في البرنامج SPSS

الأستاذ الدكتور
خالد حامد حسن




2024

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة ديالى

# التحليل الاحصائي للبيانات: مفاهيم احصائية وتطبيقات في البرنامج SPSS

الأستاذ الدكتور
خالد حامد حسن

2024

# المحتويات

**تمهيد**

**الفصل الاول** : ..... ... 1 - 45

علم الاحصاء ، اقسام علم الاحصاء ، البيانات ، انواع البيانات ، انواع المتغيرات ، دقة القياسات ، المفاهيم الاساسية لاختبار الفرضيات ، اختبار التوزيع الطبيعي لبيانات العينة ، الكشف عن القيم المتطرفة ومعالجتها ، مجاميع السيطرة في البحوث العلمية ، العينة وحجم العينة ، الوحدة التجريبية، المكررات، الخطأ التجريبي .

**الفصل الثاني :** ..... 46 - 70

الاحصاء الوصفي ، مقاييس النزعة المركزية، مقاييس التشتت، انشاء الجداول التكرارية، انشاء الجدول التكراري للمتغيرات الاسمية ، تجزئة المتغيرات الكمي الى فئات ، انشاء الفئات باستخدام الامر Visual binning ،انشاء الجدول باستخدام الامر custom tables

**الفصل الثالث :** ..... 71 - 102

الاختبارات المعلمية والاختبارات اللامعلمية، فروض الاختبارات الاحصائية، اختبار التوزيع الطبيعي للبيانات، العشوائية، تجانس التباينات، استقلالية المشاهدات، الاختبارات المعلمية : اختبار متوسط واحد للمجتمع ، اختبار فرضية العدم باستخدام حدود الثقة ، اختبار فرضية الفرق بين متوسطين لمجتمعين ، اختبار العينات المرتبطة .

**الفصل الرابع** : ..... 103 - 149

تحليل التباين ، النماذج الخطية ، الفروض الواجب توفرها في البيانات ، النموذج الثابت والنموذج العشوائي ، التصميم العشوائي الكامل ، المقارنات المتعددة لمتوسطات المعاملات في البرنامج spss ، تصميم القطاعات العشوائية الكاملة مع تسجيل مشاهدة واحدة ، تصميم القطاعات العشوائية الكاملة مع تسجيل اكثر من مشاهدة ، تصميم المربع اللاتيني .

**الفصل الخامس :** ..... 150 - 219

التجارب العاملية ، تجربة عاملية وفق التصميم العشوائي الكامل ،تجربة عاملية وفق تصميم المربع اللاتيني، التجارب المتشعبة ، التجارب العاملية المتشعبة ، القطع المنشقة ، القطع المنشقة وفق تصميم المربع اللاتيني، القطع المنشقة-المنشقة ، تصميم القطع الشريطية .

**الفصل السادس** : ..... 220 - 250

الاختبارات اللامعلمية : اختبار الاشارة، اختبار مان وتني ، اختبار كروسكال واليس ، اختبار فريدمان ، اختبارات مربع كاي ، اختبار مربع كاي لحسن المطابقة ، اختبار مربع كاي للاستقلالية .

**الفصل السابع** : ..... 251 - 265

الارتباط ، انواع معاملات الارتباط في البرنامج الاحصائي spss ، معاملات الارتباط للمتغيرات الرتبية، معامل الارتباط سبيرمان ، معامل الارتباط كندال تاو معامل الارتباط ايتا ، معامل الارتباط فاي ، الارتباط البيسري .

**الفصل الثامن :** ..... 266 - 299

الانحدار الخطي البسيط ، الفروض الاساسية للانحدار الخطي البسيط ، الانحدار الخطي المتعدد ، الانحدار اللوجستي ، الانحدار اللاخطي .

**الفصل التاسع :** ..... 300 - 309

الجداول الاحصائية

**الفصل العاشر :** ..... 310 - 311

المصادر

## تمهيد

تداخل علم الاحصاء مع بقية العلوم الانسانية والصرفة بطريقة لا يمكن الاستغناء عن تطبيقاته، ومن الصعوبة ان نتصور طبيعة الحياة بدون وجود احد اشكال المعلومات الاحصائية المتاحة لدينا، ويساعد الاحصاء الباحثين في وصف البيانات ، تصميم التجارب واختبار العلاقات بين العوامل التي يهتم بها الباحث ، لذلك يعد الاحصاء العمود الفقري للطريقة العلمية في منهجية البحث العلمي ، مما يتطلب من العاملين في مجال البحث العلمي في مختلف الاختصاصات الالمام بأساسيات علم الاحصاء وخيارات المعالجات الاحصائية، اذ يمثل الاداة التي تساعد الباحثين في تقدير معالم المجتمع، فضلا عن اختبار الفرضيات ضمن حدود ثقة منتظمة وتوفر الثقة في امكانية تكرار النتائج التي تحققها البحوث العلمية وتحدد هامش الصدفة في ظهور النتائج ، وتعطي الطمأنينة لأصحاب رؤوس الاموال في استثمار نتائج البحوث العلمية.

تعد البرامج الاحصائية ادوات مهمة في التحليل الاحصائي للبحوث العلمية مع توفير السرعة والدقة في الانجاز وفي اتخاذ القرارات الاحصائية السليمة ، مما ساعد الباحثين في تحسين جودة البحوث العلمية في مختلف الاختصاصات فضلا عن التعامل مع كتلة البيانات الضخمة التي اصبحت متاحة للباحثين. يمتاز البرنامج الاحصائي Statistical Package for the Social Sciences (spss) بإمكانيات كبيرة في التعامل مع البيانات وينفذ طيف واسع من العمليات الاحصائية وفق نظام النوافذ الذي يتيح للباحثين التعامل مع البرنامج من خلال تدريب بسيط فضلا عن امكانية تداخل الباحثين المتخصصين في البرمجة من خلال شاشة syntax الخاصة بالبرنامج ، كذلك توافق البرنامج مع برامج مايكروسوفت اوفيس مثل برنامج اكسل Excel في ادراج البيانات وبرنامج word في نقل النتائج وتداولها وطباعتها ، وهذا يفسر الاستخدام الواسع للبرنامج عالميا. ويتطلب من مستخدم البرنامج معرفته بأساسيات علم الاحصاء ومفاهيمه للتعامل مع متطلبات التحليل الاحصائي ونتائجه، ولا يغني البرنامج عن المعرفة الاحصائية للمستخدم في توجيه مسارات البرنامج الى الهدف المطلوب من التحليل الاحصائي او الاستنتاج الصحيح من نتائج التحليل، لذلك يتطلب من العاملين في مجال البحث العلمي ان يمتلكوا المعرفة الاساسية في علم الاحصاء لتجنب تكرار الاخطاء الشائعة وتبني الافكار غير الصحيحة لغير المختصين في علم الاحصاء .

يوفر المؤلف الحالي المعلومات الضرورية للباحثين حول اسلوب استخدام البرنامج الاحصائي spss في تحليل البيانات الخاصة بالبحث العلمي لتدريب الباحثين وتطوير مهاراتهم في استخدام الامكانيات الكبيرة للحاسوب في تحسين جودة البحث العلمي الذي يتطلب من الباحث الالمام بالجوانب النظرية لاختصاصه العلمي ومنهجية البحث العلمي والمعرفة السليمة للمعالجة الاحصائية لبيانات البحث العلمي، تم اعتماد الكتاب العلمي بموجب كتاب وزارة التعليم العالي والبحث العلمي ج ع/ م.هـ/ 6 في 3 / 1 / 2024 والامر الجامعي لجامعة ديالى – الشؤون العلمية 4266 في 19 / 2 / 2024 . املين الفائدة للجميع وخدمة مسيرة البحث العلمي.

والله ولي التوفيق المؤلف

# الفصل الاول

## علم الاحصاء Statistics

هو فرع من علوم الرياضيات mathematics وهو العلم الذي يهتم بعملية جمع البيانات، تبويبها ، تلخيصها وعرضها وتحليلها للوصول الى قرارات مناسبة حول الظاهرة قيد الدراسة.

يقسم علم الاحصاء الى قسمين اساسيين:

1-الاحصاء الوصفي Descriptive statistics

يهتم الاحصاء الوصفي بتنظيم وتلخيص البيانات وعرضها بصورة مناسبة لوصف المعلومات في البيانات الاولية سواء بالعرض الجدولي او العرض البياني او كلاهما، ويشمل الاحصاء الوصفي مقاييس النزعة المركزية ، مقاييس التشتت ، معاملات الارتباط.

2-الاحصاء الاستدلالي inference statistics

يهتم هذا القسم باستخدام البيانات للحصول على استنتاجات و تنبؤات حول المجتمع الذي اخذت منه البيانات. ويتضمن العمليات الاحصائية التي توفر الفرصة لاتخاذ القرارات بصدد مصدر جمع البيانات من خلال التعامل مع العينات الممثلة للمجتمع، ويشمل هذا القسم : التقدير estimation ، اختبار الفرضيات Test of hypothesis

## البيانات Data

تمثل البيانات المادة الخام للإحصاء ، ويمكن الاشارة اليها بالأرقام numbers والتي يمكن ان تسجل عن طريق اخذ القياسات مثل تسجيل الاوزان ، درجة الحرارة وغيرها ، والنوع الاخر من البيانات قد ينتج عن طريق عملية العد count وتسجل اعداد الطلبة ، عدد الثمار وغيرها ، وقد تستخدم الارقام للترميز لمعلومات وصفية جرى تسجيلها من عناصر العينة ، وجميع هذه الاشكال تسمى البيانات.

ويجري جمع البيانات من مصادر مختلفة منها سجلات المؤسسات والشركات والمنظمات التي يجري ادامتها بصورة روتينية وتمثل مصدرا غنيا للمعلومات، كما يمكن جمع البيانات عن طريق المسح survey واخذ المعلومات مباشرة من عناصر الدراسة ، فضلا عن التجارب experiments التي تمثل مصدرا مهما لانتاج

البيانات من خلال التحكم بالظروف المحيطة بالوحدات التجريبية وخضوعها الى معاملات مسيطر عليها.

## انواع البيانات

يتوجب على الباحث معرفة مستويات القياس التي تحدد بموجبها انواع البيانات المتاحة للقياس ، اذ ان الخصائص التي تمتلكها الاعداد الحقيقية تندرج في قابلية **خاصية التفرد والتمييز** وان كل رقم يتميز بذاته عن الرقم الاخر اذ يختلف الرقم 3 عن الرقم 5 وعن الارقام الاخرى، و**خاصية الترتيب** وفيها يمكن ترتيب الارقام تصاعديا او تنازليا وفق علاقة الاصغر والاكبر، **وخاصية الاضافة** اذ ان اضافة الارقام الى بعضها ينتج عنه قيمة تعتمد على كم الارقام المضافة ، وكذلك يشمل العمليات الحسابية الاربعة، **وخاصية الصفر المطلق** والذي يشير الى انعدام وجود الظاهرة قيد الدراسة . ان امتلاك البيانات جزء من هذه الخصائص او جميعها يصنف البيانات الى عدد من فئات البيانات كما موضح في الشكل 1.

**شكل 1. خصائص الرقم الحقيقي التي تمتلكها انواع البيانات المختلفة**

ومن خلال الشكل السابق نجد ان البيانات النسبية تمتلك جميع خصائص الرقم الحقيقي بينما البيانات الاسمية تمتلك خاصية التميز والتفرد فقط.

## انواع المتغيرات

من الخطوات الاولى التي يجب دراستها في البيانات لغرض اختيار الاختبارات الاحصائية هي تحديد نوع المتغير الذي يتم التعامل معه، وتصنف المتغيرات الى

ثلاثة انواع: المتغيرات الكمية measurements variables ، المتغيرات الاسمية nominal variables و المتغيرات الرتبية ranked variables ، فالتجارب المتشابهة في فرضية العدم والفرضية البديلة يجري تحليلها بشكل مختلف تماما اعتمادا على نوع المتغير في التجربة، على سبيل المثال توفرت لدينا عينة من الدجاج مكونة من 64 ذكر و 72 انثى ، وكانت فرضية العدم هي عدم وجود فروق معنوية بين الذكور والاناث في قيم المتغير X ، ولدراسة متغير كمي X في القطيع مثل طول عظم الساق shank length بوحدات مليمتر ، وبذلك يمكن للباحث تحليل البيانات وفق اختبار T لعينتين مستقلتين او وفق تحليل التباين ANOVA . اما اذا كان المتغير الخاضع للدراسة هو مقارنة تكرار التراكيب الوراثية لصفة عاري الرقبة NaNa ، Nana ، nana وبذلك نتعامل مع متغير اسمي وفق اختبار مربع كاي chi-square test او وفق G-test for independence. اما اذا كان هدف الباحث دراسة ظاهرة سلوكية مثل عدد مرات ظاهرة نقر الريش pick feather والذي يمثل متغير رتبي ويجري التعامل معه وفق الاختبارات اللامعلمية .

1- المتغيرات الكمية measurement variables

المتغير الكمي هو متغير يمكن قياسه مثل الطول ، الوزن ، الاس الهيدروجيني (pH) ، كثافة العظم وغيرها ومن الناحية العملية تستعمل الة قياس لتحديد قيمة المتغيرات الكمية وتكون هذه المتغيرات مستمرة.

المتغيرات التي تتطلب اجراء عملية العد counting مثل عدد المستعمرات البكتيرية في طبق بتري او عدد الفقرات في الحيوان تسمى المتغيرات المتقطعة discrete variables وهي متغيرات كمية يجري تحليلها بنفس الطريقة للمتغيرات المستمرة ، ولكن يجب الحذر من الخلط في المفاهيم فعندما نقوم بالعد لمتغير اسمي لا يمكن التعامل معه كمتغير كمي ، على سبيل المثال عدد المستعمرات البكتيرية في طبق بتري يمثل متغير كمي عندما نحسب 50 مستعمرة في الطبق الاول و 42 مستعمرة في الطبق الثاني وبذلك ينتج عن كل طبق قيمة واحدة ، اما اذا كان التعامل مع متغير اسمي مثل الطبق الواحد يحتوي مستعمرات باللون الاحمر واخرى باللون الابيض وقام الباحث بحساب عدد المستعمرات لكل لون بذلك سوف نتعامل معه كمتغير اسمي ، اذ ان كل مستعمرة تمثل نقطة بيانية منفصلة اما حمراء او بيضاء وفي هذا السياق يكون الهدف الوصف للكلمة وليس الرقم مما يجعله متغيرا اسميا.

2- المتغيرات الدائرية Circular variables

نوع خاص من المتغيرات الكمية وفيها تكون اعلى قيمة تعقبها ادنى قيمة في القياس، وتكون نقطة الصفر عشوائية، وهذا النوع من المتغيرات شائع في العلوم البيولوجية مثل الوقت في اليوم ، الوقت في السنة، فعندما نقيس السنة بالأيام يمكن ان يكون اليوم الاول للقياس هو الاول من كانون الثاني او يوم ميلادك وسيكون اليوم الذي يليه رقم 2 بينما اليوم الذي يسبقه هو نهاية السنة الذي تصله بعد انقضاء 365 يوما، وبذلك تتشكل دائرة من وحدات متساوية.

3- المتغيرات الاسمية Nominal Variables

تسمى ايضا المتغيرات الوصفية qualitative variables او المتغيرات الفئوية categorical variables وفيها تصنف المشاهدات الى عدد قليل من الفئات ، وفيها توصف المشاهدة الواحدة بكلمة وليس رقم مثالها جنس الفرد وفيها تكون المشاهدة ذكر او انثى ، والتركيب الوراثي يكون AA او Aa او aa ، والحالة الاجتماعية تكون اعزب او متزوج او ارمل وهكذا. وتمتلك بيانات المتغير الاسمي خاصية التفرد والتميز فقط من خصائص الرقم الحقيقي، وفيها تستخدم الارقام في هذا المتغير للتمييز بين المفردات مثل ارقام السيارات ، ارقام اللاعبين، الذكور نرمز لها 1 والاناث نرمز لها 2 وغيرها وهذه الارقام لا تحمل قيمة رياضية حقيقية.

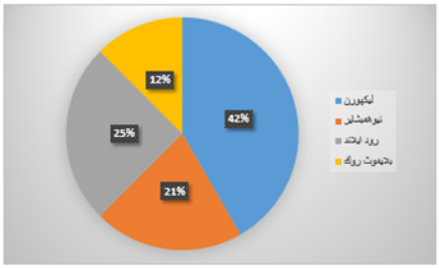
تستعمل المتغيرات الاسمية لتقسيم الافراد الى فئات ، وبذلك يمكن التعامل مع المتغيرات الاخرى لإجراء المقارنة بين هذه الفئات على سبيل المثال عند مقارنة تركيز الكلوكوز في الدم بين الذكور والاناث نقوم بتصنيف الافراد حسب متغير الجنس (اسمي) الى فئتين ذكور واناث ونستعمل المتغير الكمي (تركيز الكلوكوز) للمقارنة بين الجنسين.

ومن الممكن تحويل المتغير الكمي الى متغير اسمي وذلك بتقسيم المشاهدات الى عدد قليل من الفئات اعتمادا على مدى المتغير الكمي ، وعلى سبيل المثال اذا اجريت دراسة لتحديد العلاقة بين مستوى الكوليسترول HDL وضغط الدم ، بذلك يمكن قياس مستوى الكوليسترول في دم الافراد ( وهو متغير كمي) وتقسيم الافراد الى مجموعتين : الاولى كوليسترول منخفض ( اقل من 40 ملغم / ديسيلتر ) والثانية مجموعة طبيعية لمستوى الكوليسترول ( 40 ملغم / ديسيلتر واكثر ) وبذلك نقارن متوسط ضغط الدم بين المجموعتين باستخدام اختبار T-test. الا ان العديد من المصادر لا يؤيد هذا التوجه لان اختيار الحد الفاصل بين المجاميع قد يخضع الى

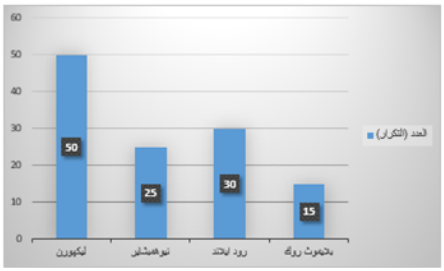
الاجتهاد او التحيز مما ينتج عنه قرارات غير صحيحة خصوصا في مجال بحوث الصحة. ويصلح حساب الاحصاء الوصفي للمتغير الاسمي ما يأتي: التكرار ، النسبة ، النسبة المئوية ، المنوال ، ومن الرسم البياني : الاعمدة و الدائرة وكما موضح في الشكل 2.

| السلالة | العدد (التكرار) | النسبة المئوية (%) |
|---|---|---|
| ليكهورن | 52 | 42 |
| نيوهمبتشاير | 25 | 21 |
| رود ايلاند | 30 | 25 |
| بلايموت روك | 15 | 12 |
| المجموع | 120 | |

التكرار والنسبة المئوية




الرسم البياني - الدائرة




الرسم البياني - الاعمدة

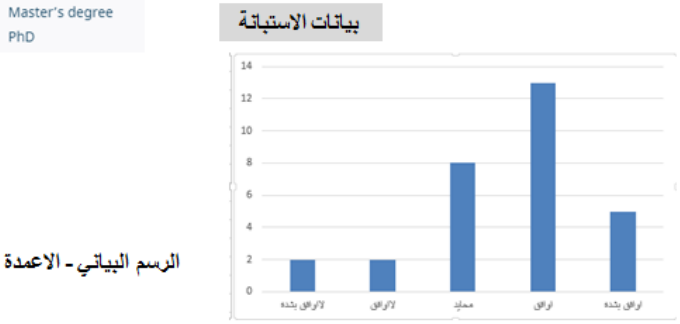
**شكل 2. مقاييس الاحصاء الوصفي المناسبة للبيانات الاسمية**

4- المتغيرات الرتبية Ranked variables

ويطلق عليها ايضا Ordinal variables وفيها يمكن ترتيب المشاهدات من الاصغر الى الاكبر او العكس (تصاعديا او تنازليا) بغض النظر عن تساوي الفروق بين مرحلة واخرى للرتب . ان هذا النوع من المقاييس يعكس فقط الحجم ولا يتضمن مسافات متساوية بين الرتب. ويمتلك هذا النوع من البيانات خاصية التميز فضلا عن خاصية الترتيب من خصائص الرقم الحقيقي، مثلا ترتيب الفائزين في سباق رياضي، شدة الاصابة (خفيفة، متوسطة ، شديدة) ، التحصيل الدراسي ، الالقاب العلمية لهيئة التدريس... وغيرها . وبذلك يكون اتفاق منطقي على ترتيبها ولكن لا يمكن مقارنتها حسابيا، ويطبق المقياس الرتبي في الاستبانة بكثرة ، مثلا الاجابة على السؤال بالاختيارات المتعددة : **موافق بشدة** – موافق – **غير موافق** – غير موافق بشدة. يستخرج للمتغير الرتبي : التكرار ، النسبة ،الوسيط، المنوال كما موضح في الشكل 3.

| Question | Options |
|---|---|
| What is your age? | • 0-18<br>• 19-34<br>• 35-49<br>• 50+ |
| What is your education level? | • Primary school<br>• High school<br>• Bachelor's degree<br>• Master's degree<br>• PhD |

| الاستجابة | التكرار | التكرار النسبي |
|---|---|---|
| لااوافق بشدة | 2 | 0.07 |
| لااوافق | 2 | 0.07 |
| محايد | 8 | 0.27 |
| اوافق | 13 | 0.43 |
| اوافق بشدة | 5 | 0.17 |
| المجموع | 30 | 1.00 |




**شكل 3. الاحصاء الوصفي للبيانات الرتبية**

وتستعمل الاختبارات اللامعلمية في العمليات الاحصائية للمتغيرات الرتبية بسبب ان هذه الاختبارات لا تشترط افتراضات محددة يجب توفرها في البيانات كما هو الحال في الاختبارات المعلمية.

5- متغيرات الفترة Interval variables

يتضمن المتغير نقاط انتقال متساوية في المقياس ، فالفرق بين درجات الحرارة 40 و 60 هو نفس الفرق بين درجتي الحرارة 60 و 80 ، ولكن هذا المقياس لا يتضمن صفر حقيقي وعندما نقول درجة الحرارة صفر لا يعني انعدام الحرارة. ويمكن استعمال العمليات الرياضية الجمع والطرح مع المتغيرات الفترية ولكن بسبب عدم وجود الصفر المطلق لا يمكننا القول ان القيمة 9 تساوي ثلاثة اضعاف القيمة 3 . ويعد صفة التاريخ من الامثلة الشائعة لتوضيح متغير الفترة فعلى سبيل المثال اذا كان لدينا اربعة وفود وصلت الينا بتاريخ الاول ، الخامس ، العاشر و الخامس عشر من شباط على التوالي ، والوحدات المتساوية للتاريخ هنا اليوم، ولا يمكن القول ان الوفد الرابع يساوي ثلاثة اضعاف الوفد الثاني مثلا، ومن الصفات الاخرى درجة الحرارة اذ ان درجة 10 مئوية ليست لا تمتلك ضعف الحرارة التي تمتلكها درجة حرارة 5 مئوية . ويستخدم هذا المتغير كثيرا في البحوث النفسية والتربوية وفيها يجري قياس الفروق بين شخصين في امتلاك الصفة طبق عليهما نفس المقياس مثل مقياس الذكاء ، التحصيل ، القلق ولا يوجد في هذا المقياس الصفر المطلق حيث ان الصفات لا تنعدم عند الافراد فلا يوجد من الافراد صفر في الذكاء ولكن في جزء الاختبار الذي مارسه.

6- متغيرات النسبة Ratio variables

يقاس هذا المتغير بوحدات متساوية فضلا عن وجود الصفر الحقيقي الذي يدل على انعدام القيمة او غيابها، ويستخدم مع هذا المقياس جميع العمليات الحسابية مثالها الطول ، الوزن ، العمر وغيرها. ويمكن للأرقام ان ترتب بطريقة معينة ينتج عنها الارقام المشتقة derived numbers ومنها التناسب ratio ، النسبة proportion ، النسبة المئوية percentage ، المعدل rate .

ويشير المصطلح ratio الى العلاقة بين رقمين ، مثلا لدينا عينة فيها 18 ذكور و 25 اناث اي ان العينة 18 : 25 ذكور واناث ويمكن كتابتها بطريقة اخرى بقسمة الرقمين على 18 وتصبح 1: 1.39 . بينما يشير المصطلح proportion الى النسبة بين الجزء والكل ، وعلى سبيل المثال وزن ذبيحة الدجاج الكاملة 1000 غم و وزن قطعية الصدر 400 غم تكون النسبة 1000/400 = 0.40 ، وعندما يكون المتغير من المتغيرات المتقطعة discrete variables عبارة عن اعداد وتكرارات فان النسبة تمثل الجزء على الكل ففي مثالنا لعينة الذكور والاناث يكون نسبة الذكور هي 18 / 43 = 0.418 . اما النسبة المئوية percentage نحصل عليها من ضرب النسبة في 100 ، ففي مثالنا السابق تكون النسبة المئوية لقطعية الصدر تساوي 40 % . ويشير المصطلح rate الى تناسب قيمة المشاهدة الى فترة زمنية معينة ويستعمل في وصف معدل النمو للجسم او التغير في التعداد السكاني خلال فترة زمنية.

ويمتلك المتغير النسبي جميع خصائص الرقم الحقيقي ويصلح حساب مقاييس الاحصاء الوصفي كاملة ( كما في الشكل 4)، فضلا عن الاختبارات المعلمية حسب توفر افتراضاتها والاختبارات اللامعلمية.

| Gender | | Salary |
|---|---|---|
| female | Mean | 562.286 |
| | N | 7 |
| | Std. Deviation | 26.3800 |
| | Std. Error of Mean | 9.9707 |
| | Sum | 3936.0 |
| | Minimum | 527.0 |
| | Maximum | 600.0 |
| | Range | 73.0 |
| | Variance | 695.905 |
| | Kurtosis | -1.001 |
| | Skewness | .134 |
| | % of Total Sum | 49.3% |
| | % of Total N | 50.0% |

**شكل 4. مقاييس الاحصاء الوصفي المناسبة للبيانات الكمية.**

يحدد نوع المتغير مقاييس الاحصاء الوصفي الذي يمكن حسابه فعلى سبيل المثال المتغير الاسمي يمكن حساب المنوال فقط ، بينما المتغير الرتبي يمكن حساب المنوال والوسيط ، اما المتغير من نوع النسبة والفترة يمكن حساب المنوال والوسيط والمتوسط الحسابي ، فضلا عن المحددات في استخدام الاختبارات الاحصائية حسب نوع المتغير.

## المتغيرات المربكة Confounding variables

لا يوجد اثنان من الكائنات الحية متطابقين تماما بسبب الاختلافات الوراثية والتطورية والعوامل البيئية، لذلك عندما نعمل على تصميم تجربة معينة لتحديد تأثير متغير X في متغير اخر مثل Y يجب علينا التفكير بدقة ومراجعة الدراسات السابقة للتأكد من وجود متغير اخر مثل Z من عدمه، يمكن ان يسهم في العلاقة بين المتغيرين قيد الدراسة X و Y ويسمى المتغير المربك والذي يشارك في اظهار النتائج خارج ارادة الباحثين ( McDonald ، 2014 ).

ولتوضيح مفهوم المتغير المربك، نفترض ان تجربة علمية استهدفت مقارنة انتاجية سلالتين من الدجاج البياض، وان البيانات التي تم الحصول عليها اشارت الى تفوق السلالة على السلالة في انتاج البيض، ان هذا الاختلاف قد يسهم فيه عوامل اخرى غير العوامل الوراثية منها اختلافات بيئية اثناء مراحل النمو الجنيني وكذلك اثناء

مرحلة النمو قبل النضج الجنسي للقطيع، وان هذه العوامل قد تعمل على تقليص او زيادة الاختلافات بين السلالتين وتسمى بالمتغيرات المربكة التي يحاول الباحث التعرف عليها لاستبعاد تأثيرها من خلال اختيار التصميم التجريبي المناسب، او اللجوء الى اساليب ادارة القطيع تضمن ازالة تأثير هذه العوامل مثل تربية الافراخ للسلالتين في نفس القاعة وظروف ادارة متشابهة قبل الوصول الى النضج الجنسي والبدء بتسجيل بيانات الدراسة لإنتاجية السلالتين.

## دقة القياسات precision of measurements

تعد القياسات الكمية من اجود انواع البيانات لأنها غير خاضعة الى التحيز فضلا عن عدم اعتمادها على خبرة القائمين بالقياس ، على سبيل المثال قياس الوزن باستعمال ميزان دقيق يعطي نتائج موثوقة لا تعتمد على طبيعة الاشخاص القائمين بالقياس بينما لصفة نوعية تعتمد على خبرة القائمين بالقياس مثل النكهة او المذاق سوف تختلف النتائج من شخص لآخر، ومع ذلك فان القياسات الكمية عندما تكون من متغيرات متقطعة تكون اكثر دقة مثلا عدد البيض في عش الطير ثمان بيضات وهو قياس لا يحتمل الاختلاف، الا ان القياسات التي تعتمد ادوات القياس مثل الوزن والطول تكون الدقة مختلفة حسب حساسية الجهاز المستخدم في القياس ، فعند استخدام شريط قياس الطول بتدريج المتر (وحدة القياس متر) فان القياسات سوف تؤخذ لأقرب قياس لوحدة المتر وبذلك فان التدريج $1 \pm 0.5$ يعتمده الشريط متر واحد وبذلك فان القياسات التي تتراوح بين 9.5 – 10.5 م يجري تسجيلها بطول 10 م ، وهذه الحالة تنطبق على الشريط الاكثر دقة مثلا شريط القياس بتدريج 0.1 م (10 سم) فان القياس يجري لاقرب 0.1 تدريج في المقياس وبذلك فان القياسات ضمن المدى 10.55 – 10.65 سوف تسجل 10.60 وهكذا لبقية المقاييس. لذلك يجب اختيار دقة القياسات المناسبة للصفة المدروسة والتي تعكس حقيقة التغير في الصفة المدروسة. وهنا لابد من الاشارة الى مصطلحين تستخدم لهذا الغرض فالميزان يكون بحساسية (دقة ) precision ضمن 0.1 ولكن عندما يكون غير موزون بشكل صحيح لا يعمل بشكل دقيق accuracy .

## المفاهيم الاساسية لاختبار الفرضيات

## Basic concepts of hypothesis test

الفرضية بصورة عامة تمثل ادعاء قد يكون صحيح وقد يكون خاطئ ويحتاج الباحث لاختباره من خلال جمع البيانات واتباع عملية احصائية تقود الى القرار بقبول الفرضية او رفضها عند مستوى احتمال محدد.

فرضية العدم Null hypothesis

وتسمى ايضا الفرضية الصفرية ويرمز لها $H_0$ وهي عبارة التي يروم الباحث اختبارها ، وبصورة عامة تفترض فرضية العدم ان تكون الاشياء موضوع الاختبار متساوية وعدم وجود فروق بينها او عدم اختلافها عن قيم نظرية متوقعة ( McDonald ، 2014 ). وعلى سبيل المثال قام احد الباحثين بقياس حجم اقدام ذكور واناث الدجاج لاستكشاف الفروق بينهما، وبذلك تكون فرضية العدم ان معدل حجم اقدام ذكور الدجاج ($\mu_1$) تساوي معدل حجم اقدام اناث الدجاج ($\mu_2$ ) وتكتب احصائيا في احدى حالاتها كما يأتي:

$H_0: \mu_1 = \mu_2$

واذا استهدف البحث حساب عدد الذكور والاناث في فقسة افراخ الدجاج فإننا نتوقع نظريا ان تكون النسبة متساوية اي النسبة 1 : 1 وتصاغ فرضية العدم بموجبها.

وأشار Daniel و Cross (2013 ) الى ان فرضية العدم يجب ان تحتوي عبارة التساوي بإحدى الصور الاتية : $=$ ، $\geq$ ، او $\leq$ وتمثل الفرضية التي يجري اختبارها.

اما الفرضية البديلة Alternative hypothesis تفترض ان الاشياء موضوع الاختبار تختلف عن بعضها البعض او تختلف عن توقعات نظرية ، ومن امثلتنا السابقة يمكننا القول ان الفرضية البديلة قد تكون ان معدل حجم اقدام الذكور لا يساوي حجم اقدام اناث الدجاج، وتكتب احصائيا كما يأتي:

$H_A: \mu_1 \neq \mu_2$

وغالبا ما نجد ان الفرضية البديلة تكون مثيرة لاهتمام الباحث ويرغب بالوصول اليها برفض فرضية العدم وقبول الفرضية البديلة لأنها تفتح نقاط واهتمامات بحثية جديدة اخرى فمثالنا اذا كان الاستنتاج ان حجم اقدام ذكور الدجاج اكبر من الاناث قد يقود ذلك الى مزيد من الاسئلة والاكتشافات البحثية في مجالات عديدة تتعلق بعلم الاحياء

التطوري، فسيولوجيا الغدد الصماء، الانتخاب الجنسي الطبيعي وغيرها بينما الوصول الى قرار قبول فرضية العدم يوقف العمل في هذا الاتجاه. وبصورة عامة فان فرضية العدم والفرضية البديلة مكملة احداهما الاخرى.

## اختبار فرضية العدم

ان الهدف الرئيس من الاختبار الاحصائي هو تحديد فيما اذا كانت البيانات المشاهدة تختلف عن تلك المتوقع الحصول عليها بموجب فرضية العدم ويتوجب على الباحث رفضها، وبصيغة اخرى ما هي احتمالية الحصول على نتائج مشابهة للنتائج التي توصل اليها بحثك؟ على سبيل المثال في سلالات الدجاج البياض تكون اناث الدجاج اكثر قيمة من الذكور ولذلك تسعى العديد من البحوث لتغيير النسبة الجنسية الى صالح الاناث . فاذا كنت بصدد اختبار معاملة معينة وحصلت على 25 من الافراخ الاناث و 23 من الافراخ الذكور فان النظرة الاولية يمكن ان تعطي انطباع ان هذه الارقام تم الحصول عليها بالصدفة ولا يوجد سبب لرفض فرضية العدم التي تشير الى النسبة المتساوية للذكور والاناث، واذا حاولت في معاملة اخرى وحصلت على 47 من الافراخ الاناث وذكر واحد فقط فان الكثير سوف يفكر ان هذه الارقام لا يمكن ان تحدث بالصدفة وقد يتخذ القرار برفض فرضية العدم وان هذه المعاملة غيرت النسبة الجنسية بصورة حقيقية قابلة للتكرار. وبعيدا عن هاتين الحالتين المتطرفتين اذا وجد الباحث ان الفقسة تكونت من 31 من الافراخ الاناث و 17 من الافراخ الذكور ، وهنا الاناث اكثر عددا من الذكور بصورة واضحة ولكن في هذه الحالة نحتاج الى حساب باسلوب احصائي احتمالية الحصول على انحراف يكون اكبر مما يحدث بسبب الصدفة .

ان هدف اختبار الفرضية الاحصائية هو تقدير احتمالية الحصول على بياناتك المشاهدة في ضوء فرضية العدم. وفي حالة ان تكون بيانات العينة تدعم فرضية العدم فإننا لا نقول ان فرضية العدم صحيحة وانما نقول ان فرضية العدم ربما تكون صحيحة (لوجود هامش محدد من الخطأ) وان بيانات العينة لم توفر ادلة كافية لرفضها، اما اذا قررنا رفض فرضية العدم فنقول ان بيانات العينة لم تتطابق مع فرضية العدم وانها تدعم فرضية اخرى هي الفرضية البديلة، التي تكون بإحدى الصور الاتية:

$H_A: \mu_1 \neq \mu_2$

$H_A: \mu_1 \leq \mu_2$

$H_A: \mu_1 \geq \mu_2$

وبصورة عامة يكون الغرض من اختبار الفرضيات مساعدة مدراء الاعمال ، المصانع، المستشفيات ... على اتخاذ القرار السليم بناء على بيانات سليمة حيث يعتمد المدير على القرار الاحصائي، فاذا تم رفض فرضية العدم فان المدير يلجأ الى الفرضية البديلة والعكس صحيح واحيانا يلجأ الى دمج بيانات اخرى لتعزيز النتائج.

وذكر Daniel و Cross (2013 ) ان الية اختبار الفرضية ولأغراض التوضيح يمكن ايجازها في عدة خطوات:

1-استكشاف طبيعة البيانات وتحديد المعالجة الاحصائية المناسبة. ومنها التعرف على نوع البيانات ومدى تأثر البيانات بالمتغيرات المربكة confounding variables والتي تتداخل في اظهار النتائج وتضلل الباحث وتقوده الى قرارات غير صحيحة، لذلك توصل الباحث الى وجود عوامل ومتغيرات غير المعاملات المدروسة تؤثر في النتائج يتطلب اختيار التصميم التجريبي المناسب لاستبعاد تأثير هذه المتغيرات.

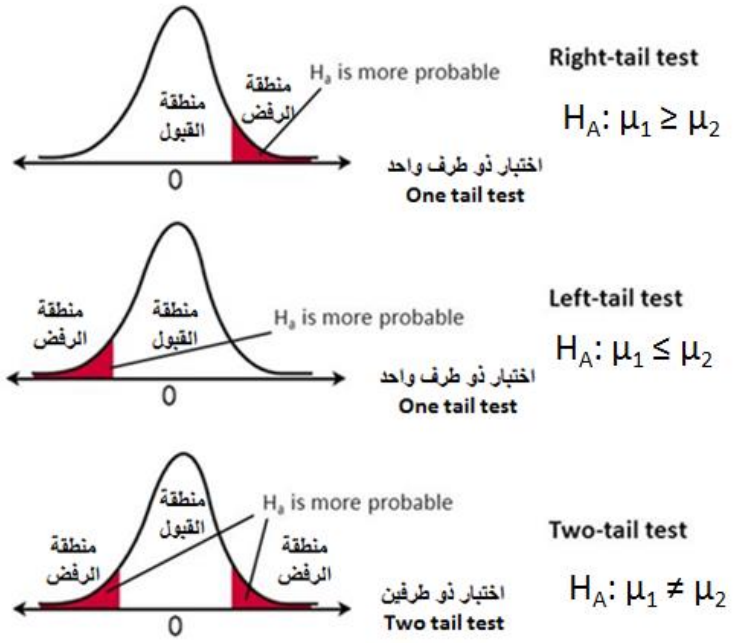
2-دراسة الافتراضات الواجب توفرها لتطبيق الاختبار الاحصائي والمعالجة الاحصائية، وتتعلق هذه الافتراضات في التوزيع الطبيعي لبيانات المجتمع، وتساوي تباينات ، وان تكون العينات مستقلة.

3-صياغة فرضية العدم والفرضية البديلة .

4- حساب المختبر الاحصائي Test statistic من بيانات العينة ، وتعتمد قيمة المختبر الاحصائي على طبيعة العينة المأخوذة من المجتمع ويستعمل المختبر الاحصائي مؤشرا يعتمد عليه باتخاذ القرار الاحصائي لرفض او قبول فرضية العدم اعتمادا على حجم حساب المختبر الاحصائي.

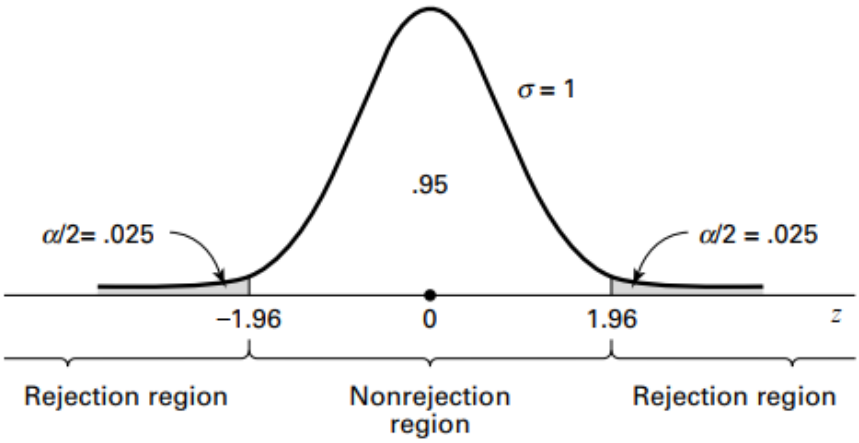
5- اتخاذ القرار الاحصائي statistical decision ، وتكون جميع نقاط المختبر الاحصائي المحتملة تقع على المحور السيني للرسم البياني لتوزيع المختبر الاحصائي والذي يمكن تقسيم المساحة تحت المنحنى الى منطقتين الاولى تسمى منطقة الرفض rejection region وفيها ترفض فرضية العدم عند وقوع قيمة المختبر الاحصائي المحسوب فيها ، اذ تقع قيمة المختبر الاحصائي في منطقة

الرفض عندما تكون احتمالية حدوثها اقل عندما تكون فرضية العدم صحيحة في حقيقتها، ومنطقة القبول او عدم الرفض nonrejection region وهي المنطقة التي تقبل فيها فرضية العدم عند وقوع قيمة المختبر الاحصائي المحسوب فيها. ويوضح الشكل 5 مناطق القبول والرفض تحت المنحنى في حالة الاختبار ذو طرف واحد one tail test والاختبار ذو طرفين two tail test وهي حالة تحققها عدم التساوي فقط دون تحديد الاتجاه في الفرضية البديلة.




**شكل 5. مناطق القبول والرفض التي تحدد القرار الاحصائي في اختبار الفرضيات.**

ويحدد مستوى المعنوية significant level (ويرمز له α ) مناطق القبول والرفض تحت المنحنى ويمكن تعريفه على انه احتمالية رفض فرضية العدم وهي صحيحة. وبذلك تتحدد مناطق القبول والرفض لفرضية العدم عند مستوى معنوية 0.05 في اختبار ذو طرفين كما في الشكل 6.




ان قيمة 1.96 الموجبة والسالبة تمثل قيمة Z الجدولية التي تكافئ مستوى المعنوية 0.025

**شكل 6. تحديد مناطق القبول والرفض في ضوء مستوى المعنوية 0.05 في حالة عدم التساوي.**

ويمكن ان يقع الباحث بالخطأ في اتخاذ القرار الاحصائي ضمن درجة احتمال معينة، والاخطاء في اتخاذ القرار الاحصائي على نوعين هما : الخطأ من النوع الاول Type I Error ويشير الى احتمالية رفض فرضية العدم وهي صحيحة في حقيقتها ويمثل مستوى المعنوية ويرمز له α والمستويات شائعة الاستعمال في علوم البيولوجي 0.05 و 0.01 ولا يمنع ذلك من اتخاذ القرار الاحصائي بمستويات اخرى، بينما يشير الخطأ من النوع الثاني الى قبول فرضية العدم وهي خاطئة في حقيقتها ويرمز له β ويمكن استعمال الخطأ من النوع الثاني في حساب قوة الاختبار الاحصائي power of the test اذ يمثل power of the test = 1- β وكما موضحة في الشكل 7.

| القرار بصدد فرضية العدم | حقيقة فرضية العدم: صحيحة | حقيقة فرضية العدم: خاطئة |
|---|---|---|
| رفض الفرضية | Type I Error ( α ) الخطأ من النوع الأول | قرار صائب |
| قبول الفرضية | قرار صائب | Type II Error ( β ) الخطأ من النوع الثاني |

Power of the test = 1- β

**شكل 7. الاخطاء المحتملة في اتخاذ القرار الاحصائي .**

وعند كتابة البحوث العلمية يجب الاشارة الى مستوى المعنوية التي تم رفض فرضية العدم او قبولها عنده، وقديما يشار الى المستوى 0.05 او 0.01 فقط بغض النظر عن مدى قرب مستوى المعنوية للمختبر الاحصائي، الا ان استخدام برامج الحاسوب وسهولة الحصول على مستوى المعنوية الذي رفضت عنده فرضية جعل من الضروري ادراجه بصورة رقمية مثلا 0.0052 ولا نقول معنوي عند مستوى 0.01 لأننا حصلنا على مستوى المعنوية الحقيقي وذلك يعزز مدى الثقة بالنتائج، ويبقى مستوى الرفض او القبول للفرضية يحدده الباحث ( Daniel و Cross ،2013 ).

ان عملية اتخاذ القرار الاحصائي برفض فرضية العدم او قبولها يعتمد على حساب قيمة المختبر الاحصائي Calculation of test statistic باستعمال بيانات العينة الخاضعة للدراسة ومقارنة القيمة المحسوبة مع مناطق القبول والرفض التي جرى تحديدها عند مستوى المعنوية المطلوب، فاذا وقعت القيمة المحسوبة للمختبر الاحصائي ضمن منطقة الرفض يكون القرار برفض فرضية العدم وفي حالة وقوع المختبر الاحصائي في منطقة القبول لا ترفض فرضية العدم، وتتخذ التوصية على هذا الأساس .

ان قيمة الاحتمالية التي يتخذ عندها القرار probability value والتي يرمز لها بصورة شائعة P-value تمثل هامش الشك في امكانية تكرار النتائج التي تم الحصول عليها من العينة المدروسة وتمثل اصغر قيمة لمستوى المعنوية $\alpha$ يجري عنده رفض فرضبة العدم ( Daniel و Cross ،2013 )، وبمعنى اخر اذا كانت P-value = 0.050 عند تكرار التجربة او المعاملة 100 مرة فإننا يمكن ارتكاب الخطأ من النوع الاول خمس مرات برفض فرضية العدم وهي صحيحة ( يمثل هامش الشك) واننا واثقون 95 % اننا اتخذنا القرار الصائب.

## اختبار التوزيع الطبيعي لبيانات العينة

يتطلب اجراء الاختبارات الاحصائية استكشاف طبيعة البيانات المتوفرة لدى الباحث من حيث نوع البيانات وتوزيعها ، اذ ان لكل اختبار احصائي عدد من الافتراضات التي يجب توفرها قبل اجراء الاختبار واعتماد نتائجه وان عدم توفر الافتراضات في مجموعة البيانات يؤدي الى نتائج وقرارات غير صحيحة ومضللة، ومن اهم الافتراضات التي تحدد اتباع الباحث للاختبارات المعلمية او الاختبارات اللامعلمية هو ملائمة توزيع البيانات للتوزيع الطبيعي والتي تحدد هل ان العينة اخذت من مجتمع يتوزع طبيعيا ام لا، ويمكن تنفيذ اختبار توزيع البيانات طبيعيا بعدة طرائق منها:

1-استخدام مقاييس الالتواء Skweness والتفلطح Kurtosis

يمكن الحصول على مقاييس الالتواء والتفلطح ضمن مقاييس الاحصاء الوصفي باستخدام البرنامج الاحصائي spss ، ويؤثر حجم العينة في حساب مقاييس الالتواء والتفلطح لذلك نتبع الاتي:

أ-عندما يكون حجم العينة اصغر من 50مشاهدة، نقوم بقسمة مقياس الالتواء او التفلطح على الخطأ القياسي للمقياس ومقارنة نتيجة القسمة مع قيمة توزيع Z المطلقة 1.96 فاذا تجاوزتها فان البيانات لا تتوزع طبيعيا.

ب-عندما يكون حجم العينة بين 50 – 300 مشاهدة ، يجري مقارنة نتيجة القسمة مع توزيع Z المطلقة 3.29 وفي حالة تجاوزها يكون التوزيع غير طبيعي.

جـ - عندما يكون حجم العينة اكبر من 300مشاهدة، يجري مقارنة النتيجة مع قيمة Z المطلقة 2.0 في حالة مقياس الالتواء والقيمة 7.0 في حالة التفلطح وفي حالة تجاوز هذه القيم فان البيانات لا تتوزع طبيعيا.

2-استخدام اختبار شابيرو Shapiro–Wilk test او اختبار كولمجروف Kolmogorov-Smirnov

يعتمد اختبار شابيرو في حالة العينات التي لا تتجاوز 50 مشاهدة ، بينما يفضل استخدام اختبار كولمجروف في حالة العينات الكبيرة. ويمكن حساب قيمة اختبار شابيرو وفق الصيغة الاتية ومقارنتها مع القيمة الجدولية للقرار بملائمة البيانات للتوزيع الطبيعي من عدمه.

$$W = \frac{\left[\sum_{i=1}^{[\frac{n}{2}]} a_i \left(x_{(n-i+1)} - x_{(i)}\right)\right]^2}{\sum_i (x_i - \bar{x})^2}$$

بينما يتوفر الاختبار في برنامج spss ويجري الاختبار ضمن مسار الاحصاء الوصفي عن طريق الخيار .Descriptive stat ضمن قائمة analyze ومنه الخيار Explore ومن مربع الحوار يتم تأشير plot ثم الخيار Normality .

تمرين

توفرت لدينا البيانات الاتية، المطلوب اختبار ملائمتها للتوزيع الطبيعي.

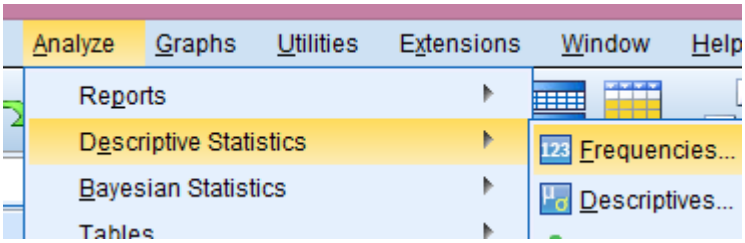
| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 66 | 51 | 50 | 59 | 54 | 79 | 56 | 68 | 63 | 49 | 61 | 56 | 48 | 63 | 57 |
| 42 | 55 | 45 | 71 | 60 | 67 | 59 | 67 | 56 | 64 | 74 | 45 | 57 | 41 | 64 |

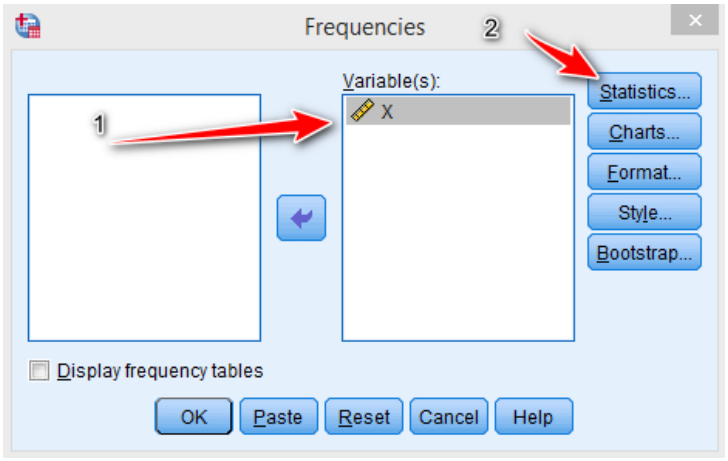
تعريف المتغير وادخال البيانات:

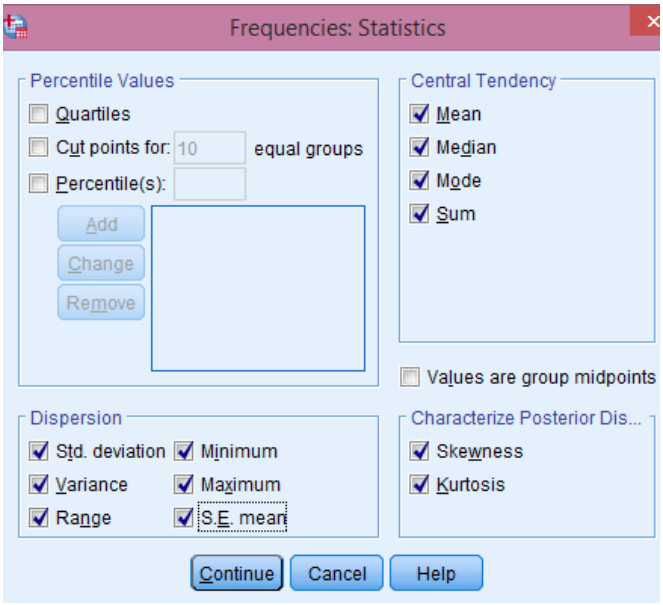
| | Name | Type | Width | Decimals | Label | Values | Missing | Columns | Align | Measure |
|---|---|---|---|---|---|---|---|---|---|---|
| | X | Numeric | 8 | 0 | X | None | None | 8 | Center | Scale |

| | X |
|---|---|
| | 57 |
| | 63 |
| | 48 |
| | 56 |
| | 61 |
| | 49 |
| | 63 |
| | 68 |
| | 56 |
| | 79 |
| | 54 |
| | 59 |
| | 50 |
| | 51 |
| | 66 |
| | 64 |
| | 41 |
| | 57 |
| | 45 |
| | 74 |

للحصول على مقاييس الالتواء والتفلطح نتبع المسار الاتي:






وبعد الضغط على استمرار نرجع الى مربع الحوار السابق ونضغط على OK ونحصل على النتائج الاتية:

**Descriptives**

| | | | Statistic | Std. Error |
|---|---|---|---|---|
| X | Mean | | 58.23 | 1.710 |
| | 95% Confidence Interval for Mean | Lower Bound | 54.74 | |
| | | Upper Bound | 61.73 | |
| | 5% Trimmed Mean | | 58.11 | |
| | Median | | 58.00 | |
| | Variance | | 87.702 | |
| | Std. Deviation | | 9.365 | |
| | Minimum | | 41 | |
| | Maximum | | 79 | |
| | Range | | 38 | |
| | Interquartile Range | | 14 | |
| | Skewness | | .061 | .427 |
| | Kurtosis | | -.319 | .833 |

بما ان حجم العينة يساوي 31 مشاهدة ، يتم قسمة مقياس الالتواء او التفلطح على الخطأ القياسي ومقارنة النتيجة مع قيمة Z المطلقة 1.96 .

Skeweness = 0.061 / 0.427 = **0.142**

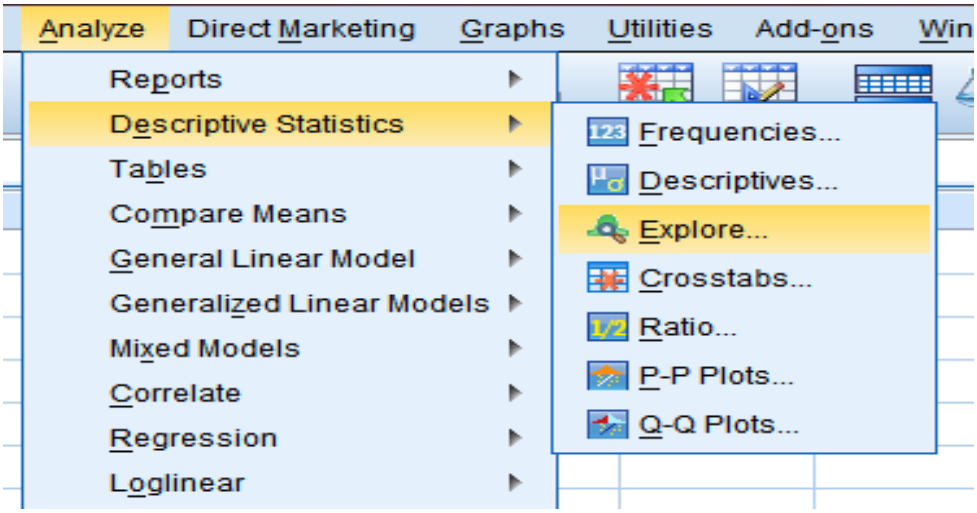
Kurtosis = - 0.319 / 0.833 = **- 0.382**

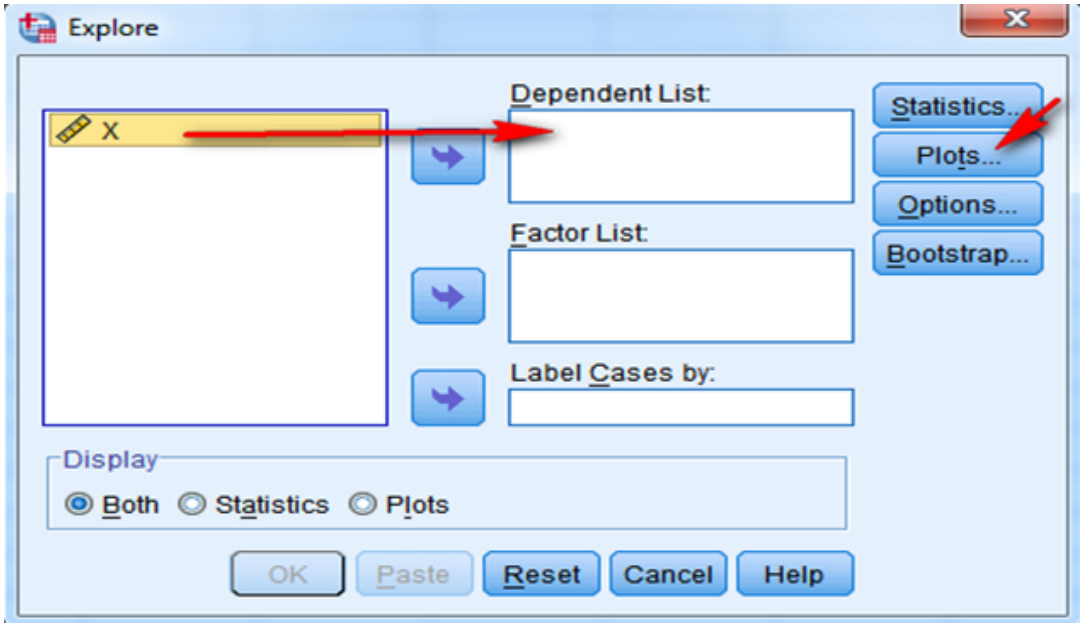
الاستنتاج : حاصل قسمة مقياس الالتواء او التفلطح على الخطأ القياسي للمقياس اقل من القيمة المطلقة لقيمة Z وهي 1.96 ، وبذلك فان البيانات تتوزع طبيعيا

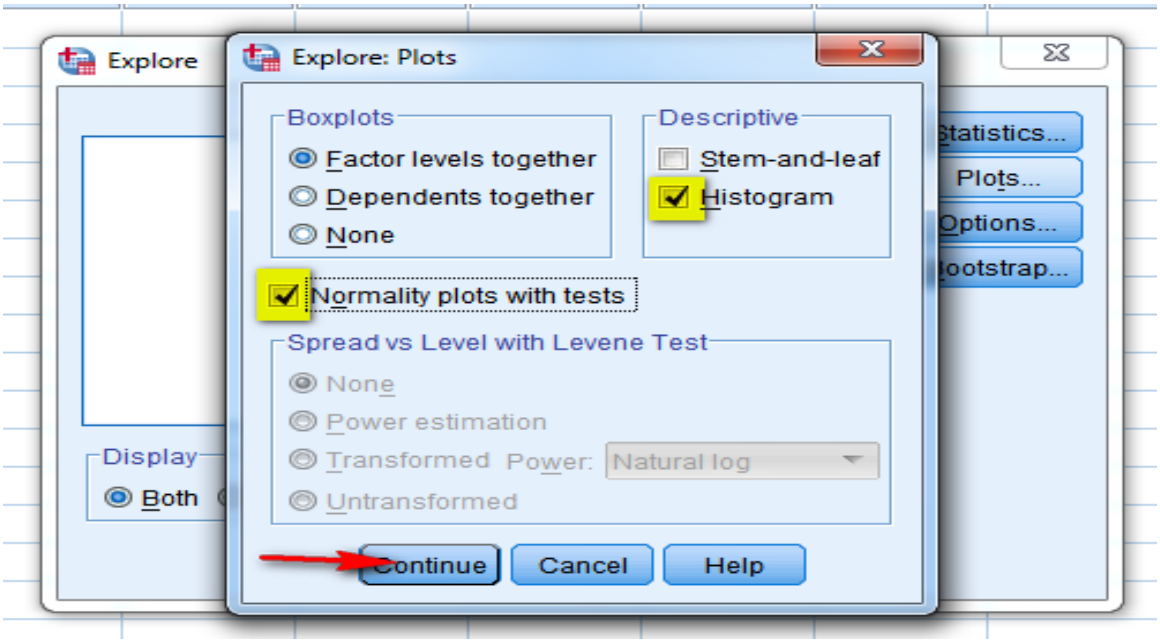
**الاختبار وفق اختبار شابيرو:**

**فرضية العدم** : توزيع البيانات لا يختلف عن التوزيع الطبيعي (عدم وجود اختلاف)

**الفرضية البديلة** : توزيع البيانات يختلف عن التوزيع الطبيعي (وجود اختلاف)





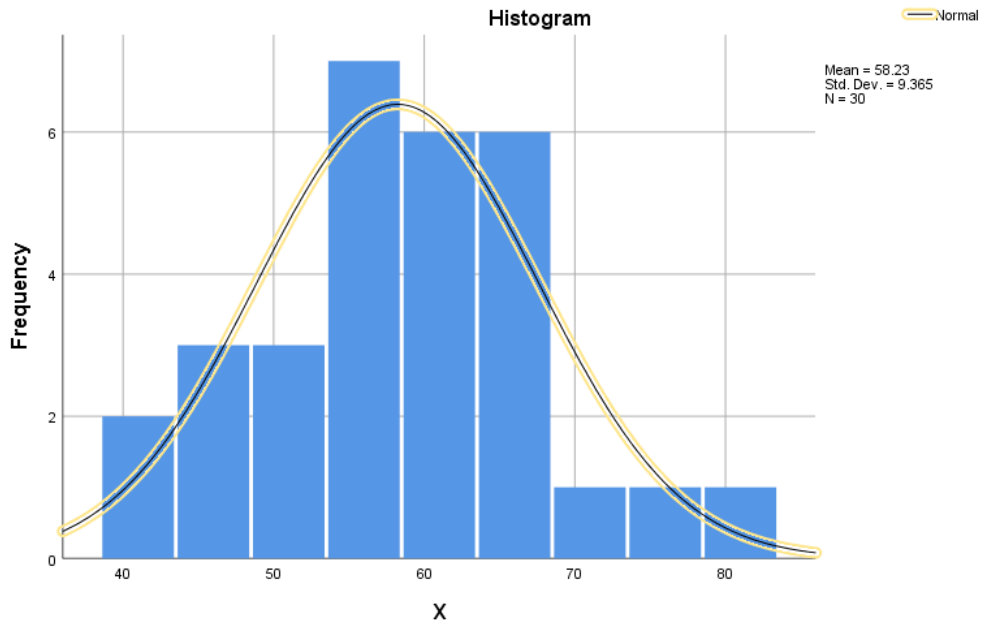

ونحصل على النتائج الاتية:

**Tests of Normality**

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| X | .072 | 30 | .200* | .986 | 30 | .956 |

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction




الاستنتاج: بما ان قيمة sig = 0.956 هي اكبر من 0.050 لذلك تقبل فرضية العدم أي ان البيانات تتوزع طبيعيا.

تمرين لمجموعة بيانات لا تتوزع طبيعيا:

توفرت لدينا العينة الاتية من القياسات، المطلوب اختبار مدى ملائمتها للتوزيع الطبيعي.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1.11 | 1.07 | 2.90 | 1.20 | 2.84 | 0.84 | 0.96 | 1.14 | 2.20 | 0.61 | 3.13 |
| 0.98 | 3.82 | 1.89 | 1.60 | 1.32 | 1.51 | 2.94 | 2.05 | 2.15 | 1.73 | 0.66 |
| 3.53 | 2.04 | 1.31 | 2.16 | 1.00 | 1.85 | 2.53 | 0.79 | 2.32 | 5.30 | 1.75 |
| 1.99 | 1.50 | 2.52 | 5.00 | 0.42 | 1.94 | 0.47 | 2.51 | 2.69 | 0.81 | 1.36 |
| 0.57 | 2.83 | 0.66 | 1.85 | 0.93 | 1.16 | 0.54 | 1.45 | 2.10 | 1.70 | 2.14 |
| 1.23 | 0.77 | 1.93 | 4.62 | 1.20 | 0.67 | 2.22 | 1.28 | 0.90 | 3.39 | 1.42 |
| 1.36 | 2.68 | 0.82 | 1.66 | 1.04 | 1.76 | 1.10 | 1.80 | 1.69 | 1.16 | 1.25 |
| | | | | | | | | 3.49 | 2.21 | 2.12 |

بما ان حجم العينة 80 مشاهدة ، يجري حساب مقاييس الالتواء والتفلطح ، وقسمة المقياس على الخطأ القياسي ومقارنة النتيجة مع قيمة Z المطلقة 3.29 وكما يأتي:

**Descriptives**

| | | | Statistic | Std. Error |
|---|---|---|---|---|
| Y | Mean | | 1.8017 | .11315 |
| | 95% Confidence Interval for Mean | Lower Bound | 1.5765 | |
| | | Upper Bound | 2.0269 | |
| | 5% Trimmed Mean | | 1.7138 | |
| | Median | | 1.6750 | |
| | Variance | | 1.024 | |
| | Std. Deviation | | 1.01206 | |
| | Minimum | | .42 | |
| | Maximum | | 5.30 | |
| | Range | | 4.88 | |
| | Interquartile Range | | 1.13 | |
| | Skewness | | 1.299 | .269 |
| | Kurtosis | | 2.049 | .532 |

Skeweness = 1.299 / 0.269 = **4.83**

Kurtosis = 2.049 / 0.532 = **3.85**

الاستنتاج : حاصل قسمة مقياس الالتواء او التفلطح على الخطأ القياسي للمقياس اكبر من القيمة المطلقة لتوزيع Z وهي 3.29، وبذلك فان البيانات لا تتوزع طبيعيا.

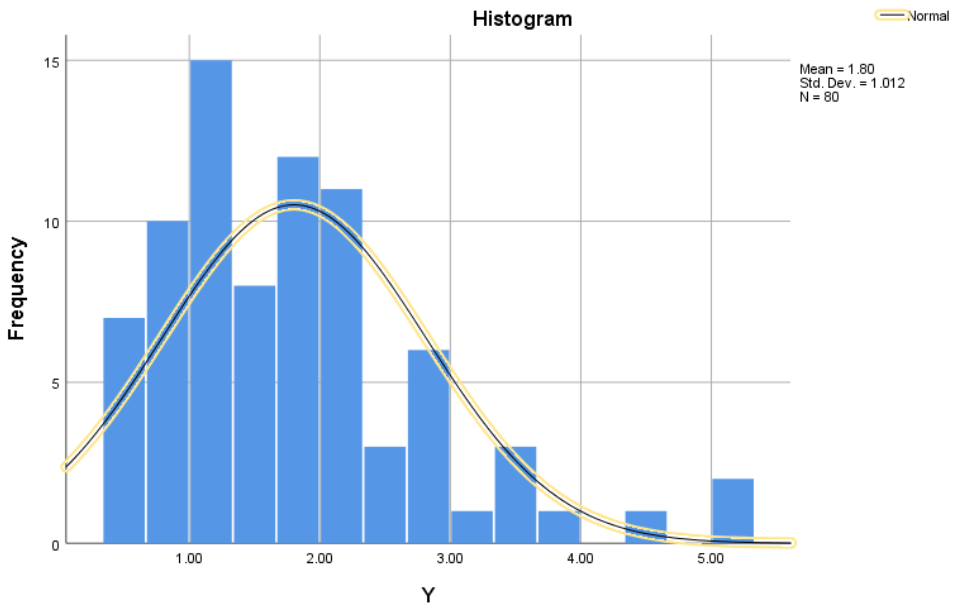
اختبار شابيرو: يتم الحصول على نتيجة الاختبار وفق المسار الذي تم وصفه سابقا والحصول على النتيجة الاتية:

**Tests of Normality**

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| Y | .115 | 80 | .011 | .904 | 80 | .000 |

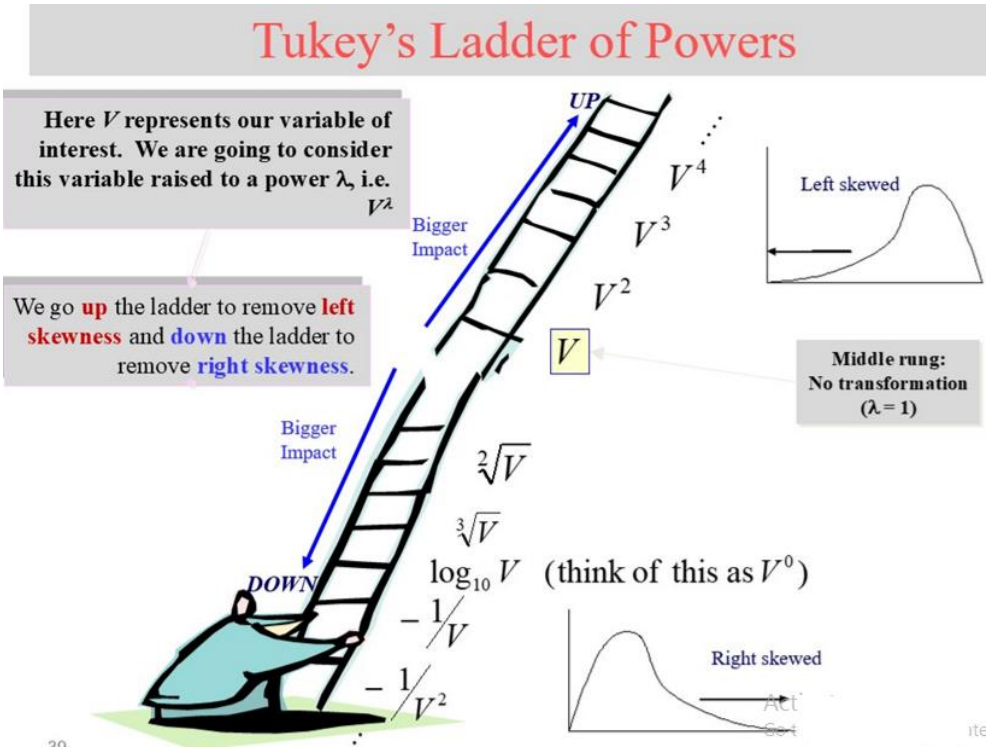
a. Lilliefors Significance Correction

يظهر اختبار شابيرو ( 0.000 ) وجود فرق عالي المعنوية مقارنة بالتوزيع الطبيعي مما يدل على ان البيانات لا تتوزع طبيعيا .




ويظهر الرسم البياني وجود التواء موجب للبيانات لامتداد ذيل منحنى البيانات الى اليمين.

ان معالجة البيانات التي لا تتوزع طبيعيا يمكن ان يتم بإجراء التحويلات ومنها التحويل الى اللوغاريتم، الجذر التربيعي، رفع البيانات الى قوة معينة .... وغيرها، ويوضح الشكل 8 التحويلات المقترحة حسب طبيعة الالتواء للبيانات.




**شكل 8 . التحويلات المقترحة للبيانات التي لا تتوزع طبيعيا.**

وقد اقترح Verma و Abdel-Salam(2019) استخدام التحويلات الاتية:

1-التحويل الى لوغاريتم القيمة log(x) عند وجود الالتواء الموجب في البيانات.

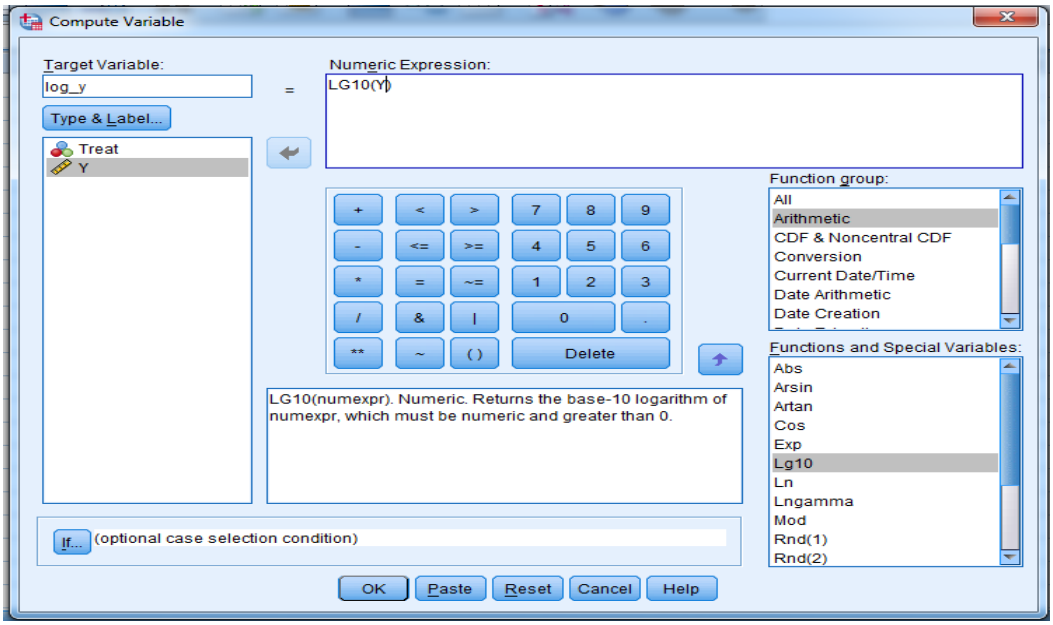
2- التحويل الى القوة الثانية (تربيع القيمة) $x^2$ عند وجود الالتواء السالب في البيانات.

3-التحويل الى الجذر التربيعي عند توزيع البيانات وفق التوزيع البواسوني Poisson distribution ( 0 ، 1 ، 2 ، ... ).

4-استخدام التحويل الزاوي arcsin(x) عندما تكون البيانات بصيغة النسب.

ويجري اختبار توزيع القيم المحولة ومدى ملائمتها للتوزيع الطبيعي، وكما يأتي:

من قائمة التحويلات Transformations نختار compute ونقوم باستحداث متغير جديد وتسمية جديدة ونختار التحويل الى اللوغاريتم الطبيعي:




ونحصل على عمود جديد يمثل التحويل الى اللوغاريتم لكل قيمة من القيم الاصلية وكما يأتي:

| Y | log_y |
|---|---|
| 3.13 | .50 |
| .61 | -.22- |
| 2.20 | .34 |
| 1.14 | .06 |
| .96 | -.02- |
| .84 | -.08- |
| 2.84 | .45 |
| 1.20 | .08 |
| 2.90 | .46 |
| 1.07 | .03 |
| 1.11 | .05 |
| 1.25 | .10 |
| 1.16 | .06 |

ونقوم باختبار توزيع القيم المحولة بنفس المسار الذي تم وصفه سابقا ونحصل على النتيجة الاتية باستخدام مقاييس الالتواء والتفلطح:

**Descriptives**

| | | | Statistic | Std. Error |
|---|---|---|---|---|
| log_Y | Mean | | .1913 | .02704 |
| | 95% Confidence Interval for Mean | Lower Bound | .1375 | |
| | | Upper Bound | .2451 | |
| | 5% Trimmed Mean | | .1924 | |
| | Median | | .2240 | |
| | Variance | | .058 | |
| | Std. Deviation | | .24183 | |
| | Minimum | | -.38 | |
| | Maximum | | .72 | |
| | Range | | 1.10 | |
| | Interquartile Range | | .31 | |
| | Skewness | | -.130 | .269 |
| | Kurtosis | | -.321 | .532 |

Skeweness = -0.130 / 0.269 = **0.483**

Kurtosis = -0.321 / 0.532 = **0.603**

الاستنتاج : حاصل قسمة مقياس الالتواء او التفلطح على الخطأ القياسي للمقياس اقل من القيمة المطلقة لقيمة Z وهي 3.29، وبذلك فان البيانات المحولة تتوزع طبيعيا

والحصول على نتيجة اختبار شابيرو كما يأتي:

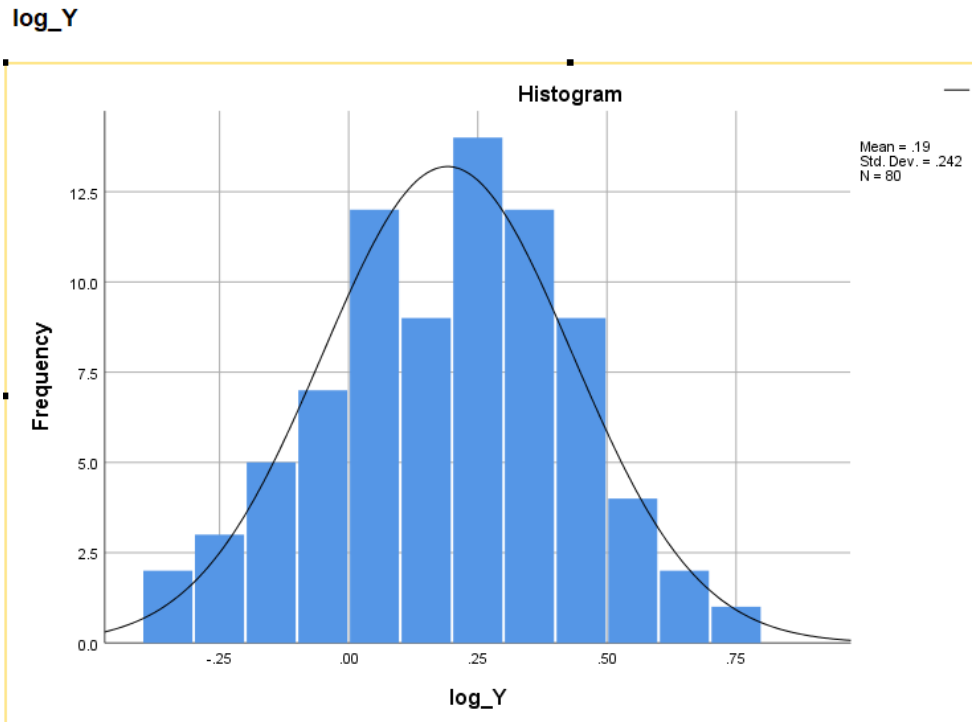
**Tests of Normality**

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| log_Y | .060 | 80 | .200* | .992 | 80 | .905 |

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

يظهر اختبار شابيرو ( sig= 0.905 ) مما يشير الى عدم وجود فرق معنوي عن التوزيع الطبيعي و يدل على ان البيانات تحولت الى التوزيع الطبيعي بسبب تحويلها الى الصيغة اللوغاريتمية.

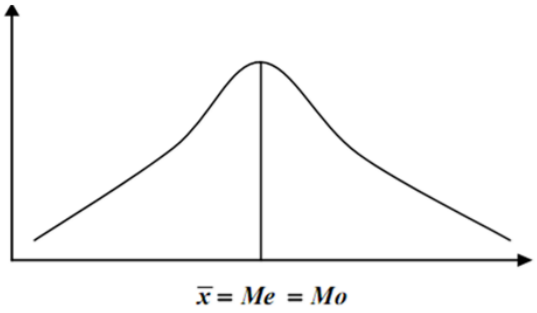



ويظهر الرسم البياني اقتراب البيانات من التوزيع الطبيعي ، وان الانحراف عن الطبيعي غير معنوي.

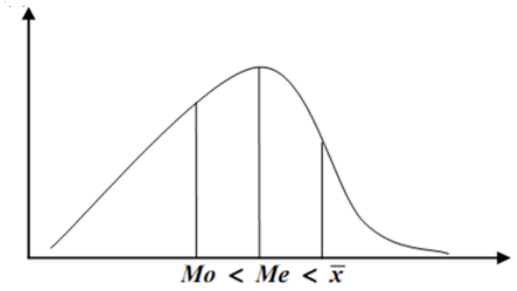
ويشير McDonald (2014) الى ان الواجب اعادة القيم والمتوسطات الى طبيعتها عند عرضها في الجداول او الاشكال في الاحصاء الوصفي وعدم استعمال القيم المحولة في الاحصاء الوصفي ويقتصر استعمالها في الاختبارات الاحصائية.

**تأثير ملائمة بيانات العينة للتوزيع الطبيعي في صلاحية مقاييس التمركز لتمثيل العينة**

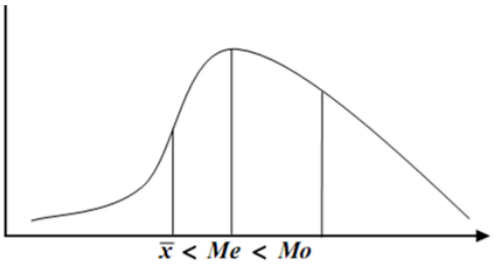
يؤثر ملائمة توزيع بيانات العينة للتوزيع الطبيعي من عدمه على صلاحية استخدام مقاييس التمركز لتمثيل بيانات العينة، ففي حالة التوزيع الطبيعي نجد ان مقاييس النزعة المركزية الثلاثة تكون بنفس القيمة كما موضحة في الشكل البياني ادناه بينما انحراف توزيع البيانات عن التوزيع الطبيعي يؤدي الى اختلاف في قيم المقاييس الثلاثة ومدى صلاحية تمثيلها لبيانات العينة.




في حالة التوزيع الطبيعي تتساوى قيم مقاييس النزعة المركزية




في حالة الالتواء الى اليمين يكون المتوسط فيه اكبر من الوسيط و المنوال .




في حالة الالتواء الى اليسار يكون المتوسط فيه اصغر من الوسيط و المنوال .

تمرين

توفرت لدينا البيانات الاتية والتي تتوزع توزيعا طبيعيا، نلاحظ ان قيم مقاييس التمركز للعينة تكون متساوية (المتوسط = الوسيط = المنوال = 16).

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 16 | 16 | 16 | 21 | 11 | 10 | 22 | 7 | 25 | 15 | 17 | 14 | 18 | 12 | 20 |

**Statistics**

X

| | | |
|---|---|---|
| N | Valid | 15 |
| | Missing | 0 |
| Mean | | 16.00 |
| Median | | 16.00 |
| Mode | | 16 |

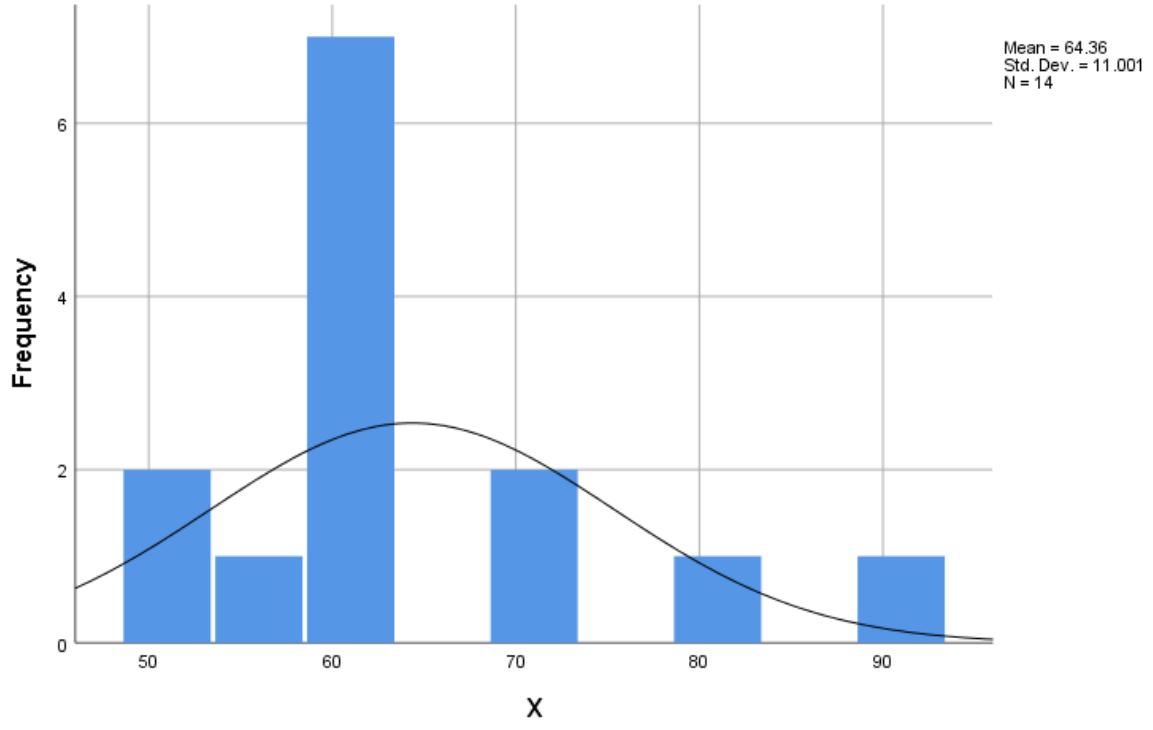
وقد اشار Verma و Abdel-Salam(2019) الى ان عدم توزيع بيانات العينة طبيعيا قد يجعل استخدام المتوسط غير صحيحا في تمثيل بيانات العينة ويكون استخدام الوسيط ملائما ويعكس صورة البيانات بشكل افضل، كما موضح في التمرين الاتي.

تمرين

توفرت البيانات الاتية

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $X$ | 51 | 52 | 55 | 59 | 60 | 61 | 61 | 62 | 62 | 63 | 72 | 71 | 82 | 90 |

وتم اختبار ملائمة توزيع العينة للتوزيع الطبيعي وحساب مقاييس النزعة المركزية كما في ادناه:



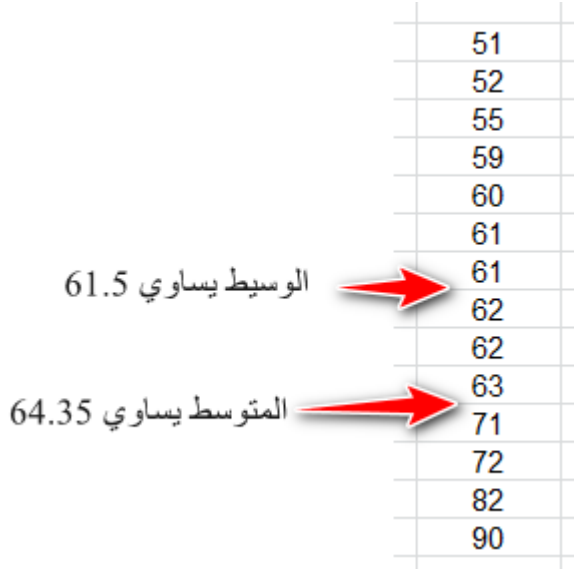
**Statistics**

X

| | | |
|---|---|---|
| N | Valid | 14 |
| | Missing | 0 |
| Mean | | 64.36 |
| Median | | 61.50 |
| Mode | | 61[a] |

a. Multiple modes exist. The smallest value is shown

نلاحظ ان بيانات العينة اظهرت التواء موجب (الى اليمين) ولا تتوزع طبيعيا، وعند ترتيب البيانات تصاعديا او تنازليا لا تمثل قيمة المتوسط (64.36) مركز القيم ولا تتوسط القيم ، بينما يكون الوسيط (61.50) هو الاكثر تمثيلا لتمركز القيم حوله كما موضح في الشكل ادناه ، ويأتي ذلك نتيجة عدم توزيع البيانات طبيعيا. لذلك ينصح باستخدام الوسيط في البيانات التي لا تتوزع طبيعيا بدلا عن المتوسط الحسابي الشائع استخدامه في العرض الجدولي للبيانات التي تتوزع طبيعيا.




## الكشف عن القيم المتطرفة ومعالجتها Outlier values

القيم المتطرفة هي مشاهدات ضمن مجموعة البيانات تنحرف كثيرا في قيمتها خارج مدى بقية المشاهدات سواء كانت للأعلى او الادنى . واشار Chawsheen و Latif (2006) الى ضرورة تحديد ومعالجة القيم المتطرفة لأنها تؤثر على جودة البيانات ونتائج التحليل الاحصائي. كما ذكر Montgomery ( 1982) ان مصدر القيم المتطرفة قد تكون القياسات او التسجيل الخاطئ للبيانات مثل ادراج القياس 5.8

وكتابته 58 او نتيجة اعطال في اجهزة القياس او ان تكون ضمن القيم البيولوجية الشاذة في المجتمع، ولذلك يجب معالجة القيم المتطرفة بإعادة القياس او استبعادها من البيانات. كما ان صغر حجم العينة قد ينتج عنه القيم المتطرفة نتيجة تباين العينات او نتيجة الشذوذ الوراثي لاحد الوحدات التجريبية. وقد تنتج القيم المتطرفة ضمن البيانات بشكل طبيعي مثل اصابة بعض حيوانات التجربة بالأمراض التي تؤثر على الاداء الانتاجي بشكل كبير يجعل بياناتها متطرفة عن الحالة الطبيعية لبقية الحيوانات، ولذلك ينصح تدوين الملاحظات بدقة عن الظروف المصاحبة للتجربة العلمية لتساعد في تشخيص اسباب ظهور القيم المتطرفة.

اشار Kwak و Kim (2017) الى ان البيانات التي لا تتوزع طبيعيا لا تحتاج الى تحديد القيم المتطرفة بينما يكون تحديدها ومعالجتها مهما في الاختبارات الاحصائية التي تفترض ان البيانات تتوزع طبيعيا.

تؤثر القيم المتطرفة على مقاييس الاحصاء الوصفي وبالتالي تعطي فكرة وتصور غير حقيقي عن المجتمع الذي اخذت منه العينة، فاذا توفرت لدينا البيانات في ادناه:

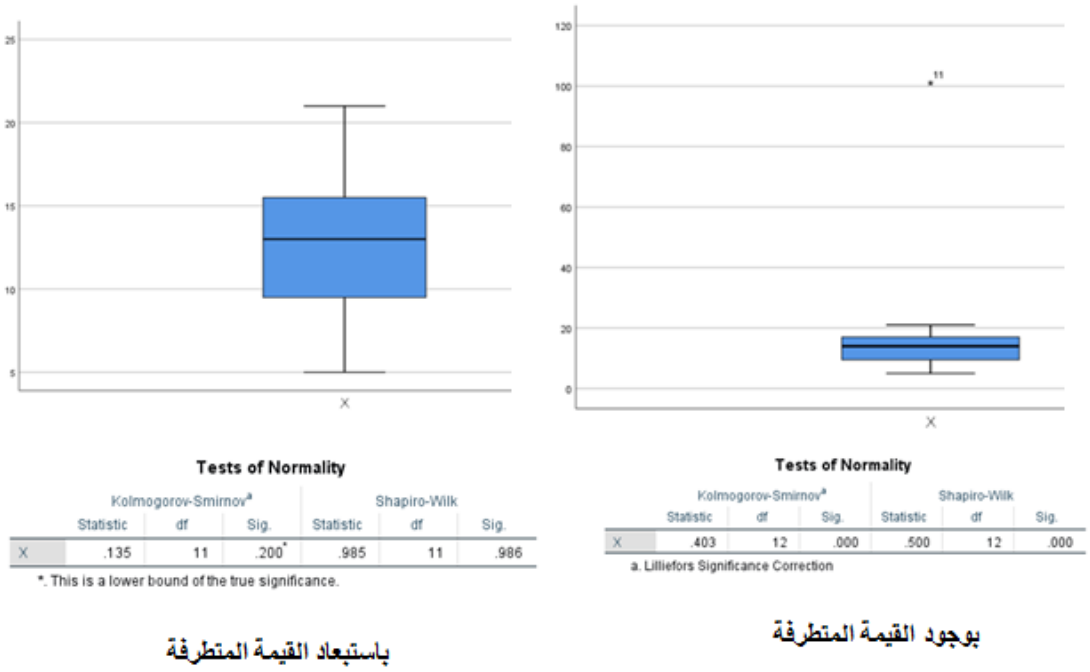
[15، **101**، 18، 7، 13، 16، 11، 21، 5، 15، 10، 9].

ومن خلال التدقيق نجد ان القيمة 101 هي قيمة متطرفة ، ولمعرفة تأثيرها في مقاييس التمركز والتشتت تم حساب المقاييس بوجود القيمة المتطرفة ، وفي حالة استبعادها من مجموعة البيانات.

| with outlier | without outlier |
|---|---|
| Mean: 20.08 | Mean: 12.72 |
| Median: 14.0 | Median: 13.0 |
| Mode: 15 | Mode: 15 |
| Variance: 614.74 | Variance: 21.28 |
| Std dev: 24.79 | Std dev: 4.61 |

نجد من خلال حساب المقاييس ان المتوسط والانحراف المعياري هي الاكثر تأثرا .

كما يؤثر وجود القيمة المتطرفة على الرسم البياني وبالأخص الرسم الصندوقي واختبار توزيع البيانات مثل اختبار شابيرو كما موضح في الشكل 9 .



**Tests of Normality**

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| X | .135 | 11 | .200* | .985 | 11 | .986 |

*. This is a lower bound of the true significance.

باستبعاد القيمة المتطرفة

**Tests of Normality**

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| X | .403 | 12 | .000 | .500 | 12 | .000 |

a. Lilliefors Significance Correction

بوجود القيمة المتطرفة

**شكل 9 . تأثير القيمة المتطرفة على الرسم الصندوقي وعلى نتيجة اختبار التوزيع الطبيعي للبيانات.**

كما تؤثر القيم الشاذة بصورة كبيرة في حساب معامل الارتباط، كما في المثال الاتي:

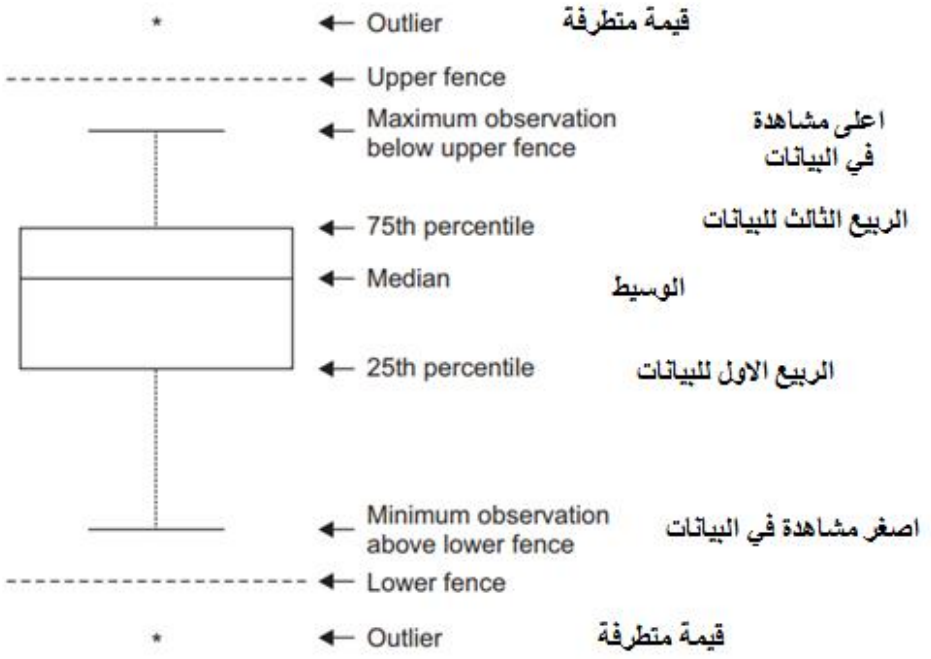
| عدم وجود قيمة متطرفة | | | وجود قيمة متطرفة | |
|---|---|---|---|---|
| **Y** | **X** | | **Y** | **X** |
| 2 | 1 | | 2 | 1 |
| 4 | 3 | | 4 | 3 |
| 6 | 5 | | 6 | 5 |
| 8 | 7 | | 8 | 7 |
| 12 | 9 | | 12 | 9 |
| | | | 9 | 22 |
| **correlation** | **0.99** | | **correlation** | **0.63** |

## طرائق تحديد القيم المتطرفة

اشار Kwak و Kim (2017) الى عدد من الطرائق التي يمكن بواسطتها تحديد القيم المتطرفة ويمكن اجمالها:

1- الرسم البياني الصندوقي Boxplot

احدى الوسائل المهمة في العرض البياني للقيم ويعتمد على الوسيط ومديات التقسيم الرباعي للقيم quartile range التي تكون اقل حساسية للقيم المتطرفة، وتعتمد الطريقة في اعتبار القيم متطرفة عند وقوعها خارج الحدود العليا او الدنيا للرسم الصندوقي كما موضح في الشكل 10 .




**Fences are usually found with the following formulas:**

1. Upper fence = Q3 + (1.5 * IQR)
2. Lower fence = Q1 — (1.5 * IQR).

**compute IQR=Q3-Q1**

**شكل 10. الرسم البياني الصندوقي ومواقع القيم المتطرفة في الرسم الصندوقي.**

ويجري ترتيب القيم تصاعديا وتحديد النقطة الفاصلة بين القيم المعتمدة والقيم المتطرفة وفق الصيغ الاتية التي ذكرها Tukey (1977):

1-calculate the 1st and 3rd quartiles(Q1, Q3)

2- compute IQR=Q3-Q1

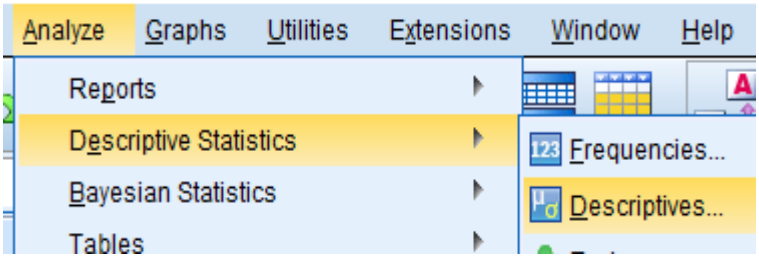
3- compute lower bound = (Q1–1.5*IQR), upper bound = (Q3+1.5*IQR)

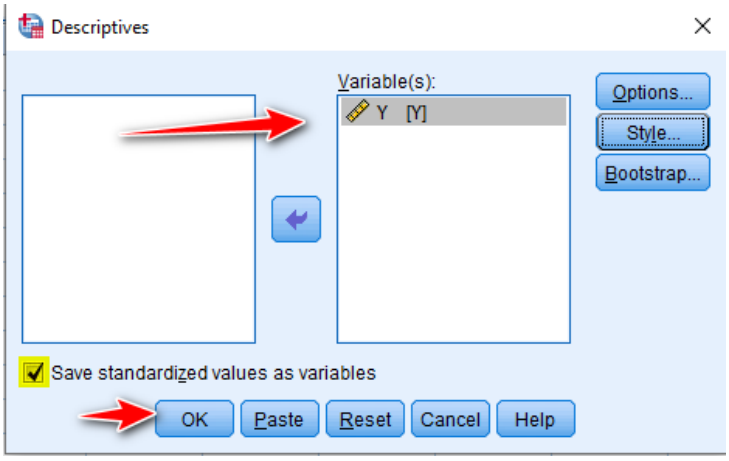
**2- التحويل الى قيم Z ( Z-score )**

عند تحويل المشاهدات الى قيم Z ، فان القيمة التي تقع بعد ثلاثة انحرافات قياسية تعد قيم متطرفة (الشكل 11) ، الا ان هذه الطريقة يعدها Kwak و Kim (2017) ضعيفة لانها تعتمد على مقاييس المتوسط والانحراف المعياري وهي مقاييس احصائية تتأثر بالقيم المتطرفة. وتنفذ الطريقة بتحويل جميع القيم الى قيم Z باستخدام الصيغة الاتية:

Z-score= (Xi-mean)/std.

ويمكن تحويل القيم الى Z باستخدام البرنامج spss عن طريق المسار:

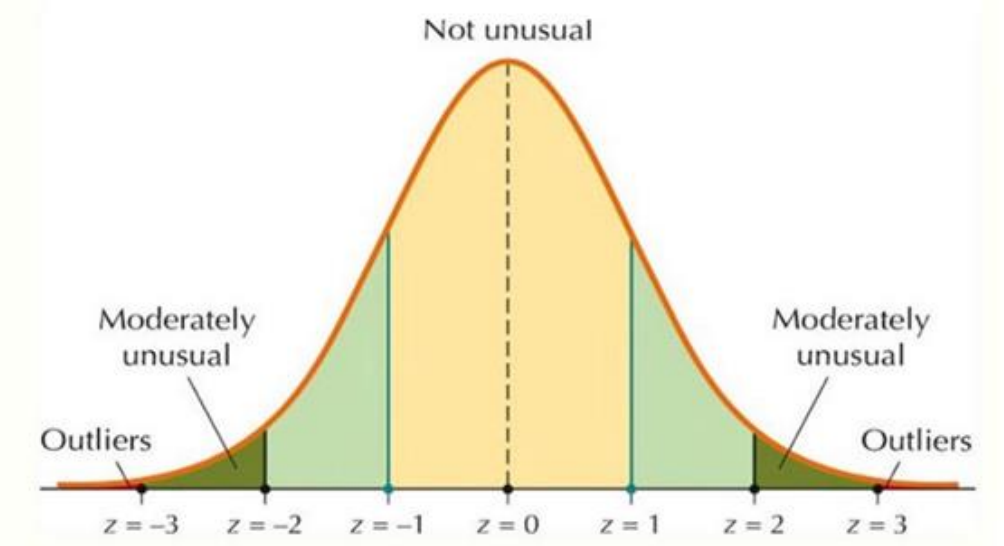


ونحصل على متغير جديد في صفحة ادخال البيانات:

| | Y | ZY |
|---|---|---|
| 1 | 3.13 | 1.31243 |
| 2 | .61 | -1.17756 |
| 3 | 2.20 | .39351 |
| 4 | 1.14 | -.65387 |
| 5 | .96 | -.83172 |
| 6 | .84 | -.95030 |
| 28 | 1.75 | -.05113 |
| 29 | 5.30 | 3.45659 |
| 30 | 2.32 | .51208 |
| 49 | .42 | -1.36529 |
| 50 | 5.00 | 3.16016 |
| 78 | 4.62 | 2.78468 |
| 79 | 1.93 | .12672 |

بعد تحويل القيم الى Z Score ، تعد قيم Z التي تزيد عن 3 قيما متطرفة .



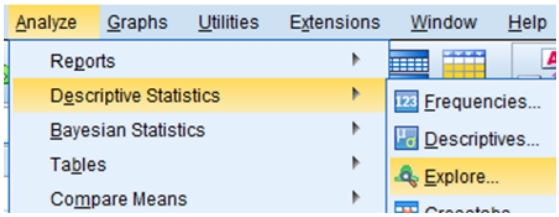

**شكل 11. توزيع قيم Z ، وموقع القيم المتطرفة في التوزيع.**

ولمعالجة القيم المتطرفة، هناك العديد من الطرائق المتبعة منها: استبعاد القيم المتطرفة دون تعويضها ويعتبرها بعض الاحصائيين بانها معالجة قاسية، ويميل الكثير الى استبدالها باعلى القيم او اقلها (حسب موقع القيمة المتطرفة) والتي لاتتجاوز حدود الرسم الصندوقي للبيانات.

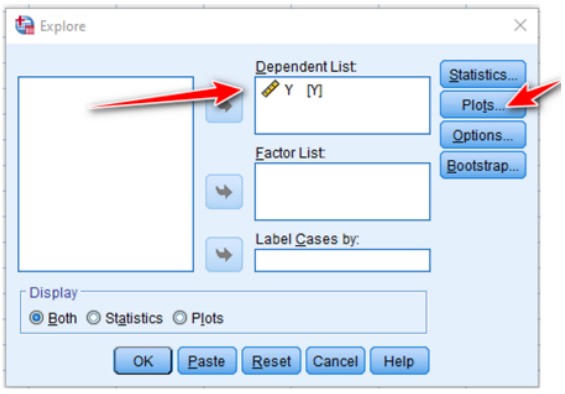
تمرين لتحديد القيم المتطرفة مع تطبيق في البرنامج spss

توفرت لدينا المشاهدات الاتية ، المطلوب تحديد القيم المتطرفة ضمن مجموعة البيانات.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 79 | 56 | 68 | 63 | 49 | 61 | 56 | 48 | 63 | 57 |
| 54 | 56 | 64 | 74 | 45 | 110 | 41 | 64 | 66 | 51 |
| 50 | 59 | 11 | 55 | 45 | 71 | 60 | 67 | 59 | 67 |

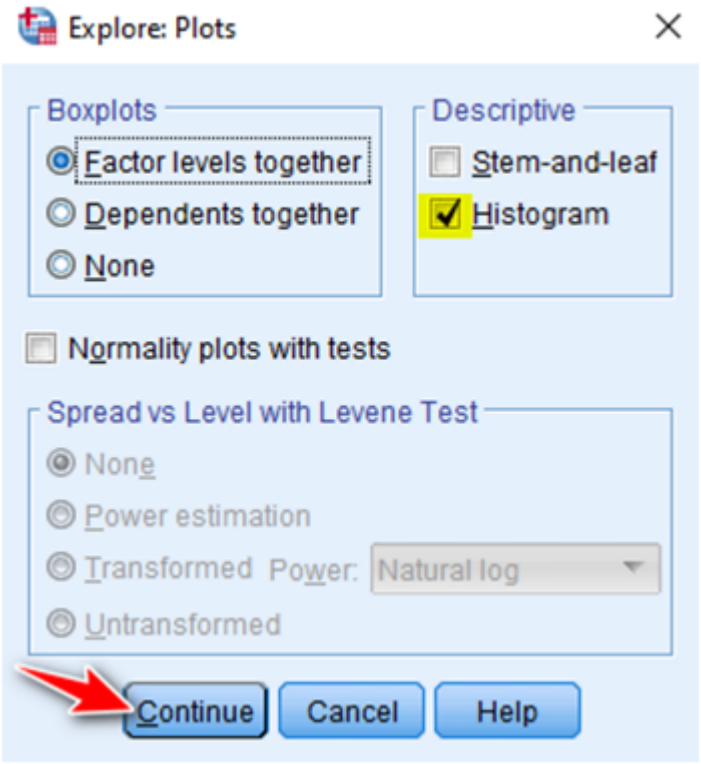



1

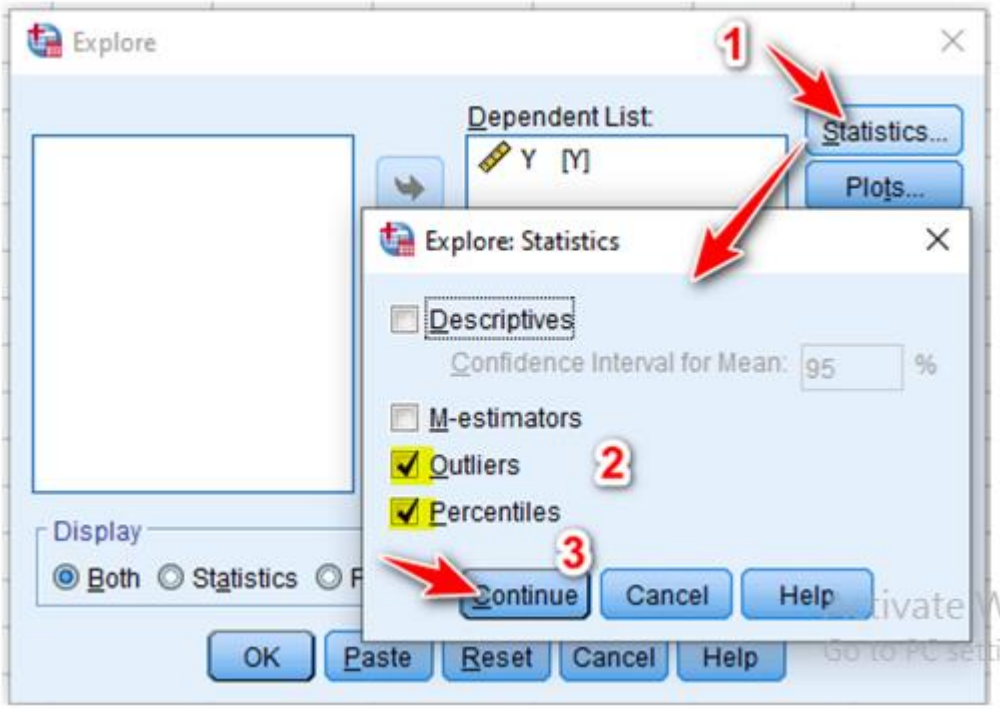



2

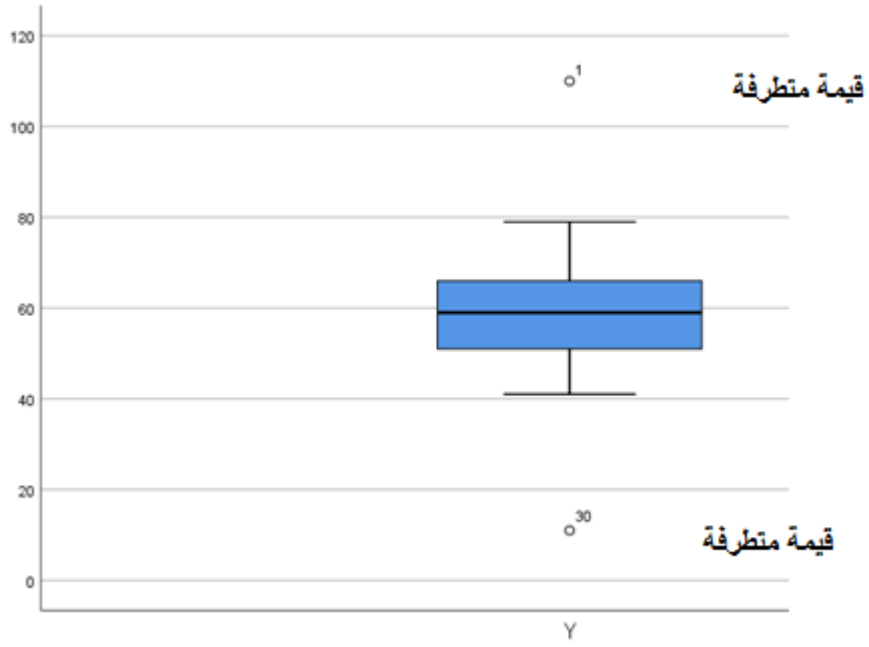
يظهر لدينا مربع الحوار لطلب الرسم الصندوقي




وبعدها نعود الى مربع الحوار explore ونضغط على خيار statistics ونحصل على مربع الحوار الاتي:




وبعد الضغط على مفتاح OK من مربع الحوار explore نحصل على النتائج الاتية:



| | | Percentiles | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Percentiles | | | | | | |
| | | 5 | 10 | 25 | 50 | 75 | 90 | 95 |
| Weighted Average (Definition 1) | Y | 27.50 | 45.00 | 50.75 | 59.00 | 66.25 | 73.70 | 92.95 |
| Tukey's Hinges | Y | | | 51.00 | 59.00 | 66.00 | | |

Q1 Q3

compute IQR=Q3-Q1 IQR=66.25 -50.75= **15.5**

lower bound = (Q1–1.5*IQR) **lower bound = (50.75–1.5*15.5) = 27.5**

upper bound = (Q3+1.5*IQR) upper bound = (66.25+1.5*15.5)= **89.5**

**Extreme Values**

| | | | Case Number | Value |
|---|---|---|---|---|
| Y | Highest | 1 | 1 | 110 |
| | | 2 | 10 | 79 |
| | | 3 | 20 | 74 |
| | | 4 | 27 | 71 |
| | | 5 | 8 | 68 |
| | Lowest | 1 | 30 | 11 |
| | | 2 | 17 | 41 |
| | | 3 | 28 | 45 |
| | | 4 | 19 | 45 |
| | | 5 | 3 | 48 |

## مجاميع السيطرة في البحوث العلمية، انواعها واهميتها

توجد العديد من المتغيرات سواء كانت وراثية أو بيئية يمكن لها ان تؤثر على نتائج تجارب الحيوانات ، لذلك يكون ضروريا استخدام مجموعة من الحيوانات تسمى مجموعة السيطرة Controls group للتقليل من تأثير هذه المتغيرات الخارجية على الاستنتاجات و القرارات التي يتخذها الباحث او لتنظيم الوجود المحتمل للمتغيرات غير المرغوبة في التجربة ، بصورة عامة كل تجربة علمية يجب ان تتضمن مجموعة او مجاميع سيطرة من الحيوانات يجري مقارنتها مباشرة مع المجاميع التجريبية الاخرى من الحيوانات . وهناك انواع عديدة من مجاميع السيطرة هي :

1- السيطرة الموجبة Positive controls

عند استخدام مجموعة السيطرة الموجبة يكون التغير في المجاميع التجريبية متوقعة وبذلك تعمل مجموعة السيطرة الموجبة كمجموعة قياسية لتحديد الاختلافات في شدة التأثير بين المجاميع التجريبية ، وكمثال للسيطرة الموجبة هو تقديم السموم إلى الحيوان و الذي يؤدي إلى تغيرات فسيولوجية أو إصابات ، وفي هذه الحالة فان المعاملات الجديدة يمكن استخدامها في المجاميع التجريبية التي تناولت السموم لتحديد إمكانية منع حدوث هذه التغيرات أو الشفاء منها .

2- السيطرة السالبة Negative controls

في حالة استخدام مجموعة السيطرة السالبة فإننا لا نتوقع حدوث تغييرات عن الحالة الطبيعية ، ففي حالة المثال السابق في السيطرة الموجبة تكون مجموعة السيطرة السالبة هي مجموعة حيوانات لم تخضع للمعاملة مع السموم و الغرض من هذه المجموعة هو للتأكد من عدم وجود متغير غير معروف يؤثر عكسيا على الحيوانات التجريبية او لتحديد وجود تذبذب في الظروف البيئية لم تؤخذ في الاعتبار عند تصميم التجربة و بالتالي قد تؤدي الى استنتاجات خاطئة .

3- السيطرة الزائفة Sham controls

تمثل مجموعة من الحيوانات يطبق عليها كافة الإجراءات المستخدمة مع المعاملات التجريبية باستثناء استخدام المادة التي يجري اختبارها في التجربة ، مثال لهذه المجموعة هو دراسة تأثير زراعة عضو بطريقة جراحية في التجويف البطني للحيوان على الاستجابة المناعية للحيوان ، ففي هذه الحالة يتم زراعة العضو في حيوانات المجاميع التجريبية بطريقة جراحية بينما مجموعة الحيوانات في السيطرة الزائفة يجري عليها نفس الإجراءات الجراحية و فتح التجويف البطني بنفس الطريقة للحيوانات المعاملة و لكن بدون زراعة العضو المستخدم للاختبار .

4- السيطرة الناقلة ( العربة ) Vehicle controls

يجري استخدام هذه السيطرة في الدراسات التي يستخدم فيها مادة معينة ( محلول ملحي ، زيت معدني .....) كواسطة لتكوين محلول من المادة المختبرة ، وفي هذه الحالة تستخدم المادة الناقلة فقط دون المادة المختبرة مع حيوانات مجموعة السيطرة وبنفس الطريقة المتبعة مع مجاميع الحيوانات التجريبية في المعاملات ، وعند مقارنة هذه المجموعة للسيطرة مع مجموعة السيطرة غير المعاملة ( السالبة ) Untreated

control سوف توضح لنا السيطرة الناقلة فيما إذا كانت هناك تأثيرات غير محسوبة للمادة الناقلة بمفردها .

5 - السيطرة المقارنة Comparative controls

السيطرة المقارنة تكون عادة سيطرة موجبة مع معاملة معروفة و تستخدم لغرض إجراء مقارنة مع معاملة أخرى مختلفة تماما ، على سبيل المثال استخدام نظام للعلاج الكيمياوي في الحيوانات المصابة بالسرطان لمقارنة نظام العلاج الكيمياوي المعتمد حاليا مع نظام جديد لعلاج السرطان في هذه الحيوانات .

## العينة وحجم العينة

يقصد بالعينة sample هي جزء من المجتمع، تؤخذ في اغلب الحالات عشوائيا لتكون ممثلة للمجتمع وغير متحيزة لغرض الحصول على نتائج يمكن تعميمها على المجمع الذي اخذت منه.

توفر العينة العديد من الفوائد والمزايا منها:

1-تستخدم العينة للاستدلال عن قيم معالم المجتمع من خلال المقاييس الاحصائية للعينة.

2-توفير التكلفة والوقت والجهد.

3-يمكن ان نحصل من مفردات العينة على معلومات اكثر تفصيلا مقارنة بالمجتمع.

4-حتمية استخدام العينة في بعض الحالات التي تتلف او تهلك فيها مفردات العينة مثل المعلبات ، البيض ، صفات نوعية الذبيحة .. وغيرها.

5-لايمكن دراسة جميع المفردات في المجتمع غير المحدود مثل نوع معين من السمك في البحر او دراسة فلكية للنجوم او الكواكب وهي غير معلومة العدد.

تستخدم العينة العشوائية في تقدير معالم المجتمع ومنها متوسط المجتمع ونسبة صفة معينة في المجتمع فضلا عن التباين ، ويؤثر حجم العينة العشوائية في كفاءة تقدير المعلمات، اذ كلما كبر حجم العينة العشوائية زادت كفاءة تقدير المعلمات وانخفضت تقديرات التباين للتقديرات وزيادة في دقة القرارات الاحصائية والعكس صحيح في حالة العينات الصغيرة جدا اذ قد ينتج عنها الاخفاق في الوصول الى اهداف الدراسة، اذ ان حجم العينة الكبير يقلل خطأ العينة العشوائي ويساعد في انجاز تحليل

البيانات بدقة اكبر، الا ان زيادة حجم العينة يصاحبه ارتفاع في تكاليف الدراسة وزيادة الجهد والوقت المطلوب لإنجاز البحث، لذلك يجري تحديد الحد الادنى لحجم العينة المسموح باختياره بما يضمن دقة تقدير المعلمات. ان الغاية من اختيار عدد من العناصر في العينة هو تقليل الخطأ العشوائي للمعاينة والناتج عن الاختلافات الذاتية لعناصر المجتمع على سبيل المثال الاختلافات الوراثية بين الافراد في المجتمع والتداخل بين البيئة والوراثة التي لا يمكن السيطرة عليها او التنبؤ بتأثيرها من قبل الباحث عند اختيار العينة وانما تمثل انحرافات عشوائية تضاف الى الخطأ العشوائي.

يؤثر في اختيار حجم العينة المناسب عدد من العوامل منها:

1-هدف الدراسة، اذ يتم اختيار حجم العينة الصغير في الدراسات الاستكشافية او الدراسات ذات الاهمية المنخفضة، بينما يتم اختيار حجم العينة الكبير للدراسات التي تتعلق بوصف معالم المجتمع او الدراسات التي يعتمد عليها في صياغة معادلات التنبؤ، او تقييم ظاهرة نادرة الحدوث ومهمة في المجتمع. كما ان الدراسات التجريبية تميل الى استخدام حجم عينة اصغر من العينات المستخدمة في الدراسات المسحية.

2-مراعاة الجوانب الاخلاقية في اختيار العينة، اذ تؤدي زيادة حجم العينة عن المستوى الضروري لضمان دقة النتائج الى زيادة الاعباء غير المبررة على المشاركين في انجاز الدراسة، فضلا عن مراعاة الجوانب الانسانية في اختيار الحد الادنى المطلوب لدقة النتائج في بعض البحوث التي تستوجب اجراء معاملات تسبب المعاناة او الاذى للحيوانات التجريبية او الهدر في الموارد.

3-طبيعة مجتمع الدراسة يؤثر في حجم العينة التي يجري تحديدها للدراسة، اذ ان المجتمع المتجانس في الصفة ذات الاهتمام يشجع على اختيار عينة عشوائية صغيرة بينما زيادة التباين للصفة في المجتمع يقتضي تحديد حجم عينة كبير، كما ان توزيع عناصر المجتمع وانتشارها المكاني وما ينتج عنه من تباين في الصفة المدروسة يؤثر في اسلوب اخذ العينة مثل العينة الطبقية او العينة العنقودية والتي تنعكس على زيادة حجم العينة لتمثيل المجتمع.

4-الموارد المتاحة للدراسة، وتكون الموارد محددة وضاغطة على الباحث لتقنين اهداف البحث وحجم العينة اللازمة لتحقيق اهداف الدراسة، اذ ان حجم العينة يجري تحديده بواسطة قسمة التمويل المتاح لجمع البيانات وتحليلها على متوسط تكلفة جمع البيانات وتحليلها لكل عنصر في العينة.

5-ان استخدام عينة صغيرة جدا وبحجم لا يكفي للكشف عن التأثيرات المعنوية التي يكون لها اهمية عملية، بينما زيادة حجم العينة بصورة كبيرة قد يؤدي الى ظهور فروق صغيرة جدا ذات دلالة احصائية ، وفي حقيقتها لا يكون لها اهمية عملية او اقتصادية، وهذا يؤكد ضرورة اجراء دراسات الجدوى الاقتصادية للبحوث التطبيقية التي تظهر اهمية احصائية قبل اعتمادها في الاستثمار الاقتصادي.

اشار Thompson (2012 ) الى الصيغة الرياضية الاتية لتقدير حجم العينة المستخدمة في تقدير متوسط المجتمع . ففي حالة المجتمع غير المحدود infinite population .

$$n_0 = \frac{z^2\sigma^2}{d^2}$$

حيث يمثل Z التوزيع الطبيعي المعياري وفي حالة الفرضيات ذات الاتجاهين Two tail test فان قيمة Z تساوي 1.96 (وفي حالة الفرضيات باتجاه واحد one tail test قيمة Z تساوي 1.645 )، بينما σ يمثل الانحراف المعياري للمجتمع للصفة المدروسة ، ويمثل d خطأ المعاينة اي بمعنى اخر خطأ في اختيار مفردات العينة والشائع استخدامه في الدراسات البيولوجية 0.05 والذي يوفر مستوى الثقة 0.95 في اختيار العينة.

وهنا تظهر امام الباحثين مشكلة تقدير تباين الصفة في المجتمع وفيها يلجأ الباحثون الى الدراسات السابقة او الدراسة الاستكشافية Exploratory study او الرائدة لتقدير التباين واستخدامه في الصيغة الرياضية ( Dell واخرون ، 2002)، الا ان الدراسات الاستكشافية تكون غير ملائمة لبعض الصفات مثل القابلية على الحياة، التوائم او الولادة المفردة حيث ان قياس هذه الصفات تحتاج الى فترة طويلة، لذلك يلجأ الباحث الى السجلات في المحطة للحصول على التقدير.

ففي حالة توفر المعطيات لمفردات الصيغة الرياضية : $Z= 1.96$ : $\sigma = 0.7$ : $d= 0.05$

يكون تقدير حجم العينة في حالة المجتمع غير المحدود يساوي 753 وحدة معاينة . اما اذا توفرت لدى الباحث المعلومات حول حجم المجتمع ( N = 2000) فيتم تعديل الصيغة الرياضية السابقة لتلائم المجتمع المحدود finite population ليكون حجم العينة n = 700 وفق الصيغة الاتية :

$$n = \frac{1}{1/n_0 + 1/N}$$

اما استخدام مؤشر النسبة الحقيقية للمجتمع في تقدير حجم العينة فيتم وفق الصيغة الرياضية الاتية في المجتمع غير المحدود :

$$n_0 = \frac{z^2 p(1-p)}{d^2}$$

وتمثل p احتمال ظهور الصفة بينما احتمالية عدم ظهور الصفة هي 1-p ، ويتم اختيار احتمالية الظهور 0.50 تماشيا مع العشوائية وعدم التحيز الى احدى الحالتين الظهور وعدم الظهور.

فاذا توفرت لدينا المعطيات للصيغة الرياضية : Z= 1.96 : p = 0.50 : d = 0.05

وتقدر حجم العينة 385 وحدة معاينة . وفي حالة توفر المعلومات لدى الباحث عن حجم المجتمع الحقيقي يتم تعديل الصيغة الرياضية :

$$n = \frac{1}{1/n_0 + 1/N}$$

ففي حالة حجم المجتمع ( N ) يساوي 3000 عنصر يكون حجم العينة وفق الصيغة الرياضية المعدلة تساوي 341 وحدة معاينة.

## الوحدة التجريبية Experimental unit

تمثل الوحدة التجريبية اصغر وحدة تطبق عليها المعاملة في التجربة العلمية، وقد تكون حيوان، نبات، ورقة، سلعة .. وغيرها . عندما يجري استخدام مجموعة من الحيوانات في اكنان pens او في الحلبة paddock و تأخذ الحيوانات في الكن نفس المعاملة ففي هذه الحالة يكون مجموعة الحيوانات في الكن وحدة تجريبية ، وتعد هذه الممارسة ضرورية في بعض الصفات مثل نسبة الاصابة، نسبة الهلاكات وغيرها. ويجب ملاحظة مدى التجانس بين الوحدات التجريبية لكونها تمثل المفتاح لاختيار التصميم التجريبي المناسب.

## المكررات Replicates

يقصد بالمكررات هو اعادة تطبيق المعاملة على الوحدة (الوحدات) التجريبية لأكثر من مرة واحدة، اذ من الصعوبة الكبيرة الحصول على التماثل في الاجراءات التي تتضمنها المعاملة وبالتالي يساهم الفشل في تكرار تطبيق المعاملة بنفس الدقة الى تباين في الحصول على النتائج ويضاف الى تباين الخطأ التجريبي، ومما يزيد من اهمية المكررات هو ان غالبية البحوث التطبيقية تسعى الى تعميم النتائج لتطبيقها من قبل شريحة واسعة وبالتالي سوف تتفاوت مهارة ودقة القائمين بتطبيق المعاملات. ان النتائج غير القابلة للتكرار لا يمكن اعتمادها لأنها ظهرت بالصدفة، وان تكرار ظهور النتائج مع تكرار تطبيق المعاملة او الدراسة يؤدي الى اكتساب النتائج المصداقية والموثوقية، وبنفس الوقت فان اثبات الدراسة عدم تكرار النتائج لمعاملة معينة يعتبر مهما لأنه يحفز الباحثين لتطبيق اتجاهات جديدة ومبتكرة للبحث للوصول الى نتائج قابلة للتكرار. ويجب علينا التمييز بين حجم العينة والمكررات في البحوث العلمية والغرض لكل منهما، فعلى سبيل المثال استهدفت تجربة تحديد تأثير الاجهاد الناتج عن نقل الافراخ من المفقس الى قاعة التربية فان نقل 100 فرخ او 5000 فرخ في عجلة النقل تمثل مكرر واحد لان المعاملة (نقل الافراخ) طبقت لمرة واحدة وان اختلاف عدد الافراخ هو اختلاف في حجم العينة (تستهدف خفض خطأ العينة العشوائي بسبب الاختلافات الوراثية الفردية بين الافراخ)، اما نقل الافراخ بأربع سيارات مختلفة ومن مناطق مختلفة تمثل مكررات لان المعاملة اعيد تطبيقها لأربع مرات وبالتأكيد سيكون اختلافات تتراوح من البسيطة الى الكبيرة في تطبيق المعاملة، وهذا يضمن اتساع صلاحية النتائج للتجربة . كما يمكن ان يكون المكررات اعادة تطبيق التجربة كاملة في مناطق مختلفة او اوقات مختلفة. ويجري تقدير عدد المكررات في التجربة لاختبار فرق معنوي مرغوب من قبل الباحث وفق الصيغة التي ذكرها Morris (2002) .

$$n = 2 \times t^2 \times \frac{S^2}{d^2}$$

ويمثل n عدد المكررات ، t القيمة الجدولية لتوزيع t عند مستوى معنوية محدد ، $S^2$ تباين العينة ، d الفرق المعنوي المرغوب اختباره. يجب ان تكون وحدات القياس للصيغة الرياضية متشابهة مثلا لا يجوز ان يكون وحدة قياس انتاج الحليب كغم / بقرة في حالة الانحراف المعياري وتكون باوند / بقرة في حالة قياس الفرق المعنوي المرغوب d وانما تكون بنفس وحدة القياس. ويمكن للباحث تجاوز مشكلة اختلاف

وحدات القياس بالاعتماد على معامل الاختلاف C.V.% بدلا عن التباين لتصبح الصيغة :

$$n = 2 \times t^2 \times \frac{(C.V.\%)^2}{d^2\%}$$

واشار الراوي وخلف الله (1980) الى امكانية تطبيق الصيغة بالاعتماد على نتائج التجارب السابقة لتصبح الصيغة :

$$n = 2 \times t^2 \times \frac{mse}{d^2}$$

يمثل :

t: قيمة t الجدولية عند مستوى معنوية محدد و درجات حرية ملائمة لتباين الخطأ .

mse : تباين الخطأ .

n : عدد المكررات لكل معاملة .

إذا رمزنا d لمقدار الاختلاف الذي نرغب لتحديده كفرق معنوي في التجربة التي يجري تصميمها ، ثم إن علينا إيجاد قيمة n التي ينتج عنها القيمة d والتي تساوي قيمة اقل فرق معنوي عند إكمال التجربة .

ان تقدير قيمة اقل فرق معنوي مرغوب d لغرض استخدامه في الصيغة يمكن ان يعتمد على الدراسات السابقة او على دراسة الجدوى الاقتصادية ويكون الهدف الاقتصادي مهما لأغلب التجارب الزراعية لذلك تؤخذ المنفعة الاقتصادية بنظر الاعتبار عند تصميم التجربة و التي يمكن ان تعطي فهم افضل للحصول على استجابة مؤثرة في الكلفة لمعاملات جديدة ، مثلا كلفة الادوية التي تستخدم في حيوانات المزرعة حسب سعر السوق الحالي و يتطلب ذلك الحصول على انتاج اضافي يبرر استخدام هذا الدواء و على ضوء ذلك يجري تحديد الفرق المعنوي المطلوب .

## الخطأ التجريبي Experimental error

الخطأ العشوائي الناتج عن الاختلافات الطبيعية بين الوحدات التجريبية التي عوملت بنفس المعاملة . وهناك العديد من المصادر التي ينشأ عنها الخطأ التجريبي

يجب محاولة التحكم فيها لتقليل قيمة الخطأ التجريبي لزيادة دقة التجربة والقرارات الاحصائية الناتجة عنها.

1-الاختلافات الذاتية بين الوحدات التجريبية الناتجة عن الاختلافات الوراثية و التداخل بين البيئة والوراثة، وهي اختلافات لا يمكن تجاوزها بشكل تام ولكن يمكن تقليص تأثيرها من خلال اختيار حجم العينة المناسب الذي يضمن تنوع التراكيب الوراثية التي تشترك في التجربة وتقترب في متوسطها من متوسط المجتمع.

2-الاختلافات الناتجة عن الفشل في تكرار تطبيق المعاملات بنفس الدقة، وهي حالة طبيعية اذ لا يمكن لنفس الشخص ان يطابق اجراء المعاملة عند التكرار مثل تجانس العليقة عند تكوين وخلط العليقة لعدة مرات، مع الاخذ بنظر الاعتبار ان المعاملة قد تطبق من اشخاص مختلفين ويمتلكون دقة ومهارة مختلفة، ان استخدام عدد مناسب من المكررات للمعاملة يمكن ان يسجل متوسط اداء مقبول، وهنا يجب التمييز بين حجم العينة وعدد المكررات والغرض لكل منها.

3-الاختلافات الناتجة عن الاخطاء الفنية، وتنتج هذه الاخطاء بسبب الاختلافات في مهارة وكفاءة الاشخاص القائمين بالقياس والتعامل مع البيانات او مدى دقة الاجهزة المستخدمة في قياس الصفات المدروسة، ويمكن تقليل هذا النوع الاخطاء عن طريق الدقة في تداول البيانات واختيار الاشخاص المدربين والحاصلين على مهارة في التعامل مع الاجهزة العلمية.

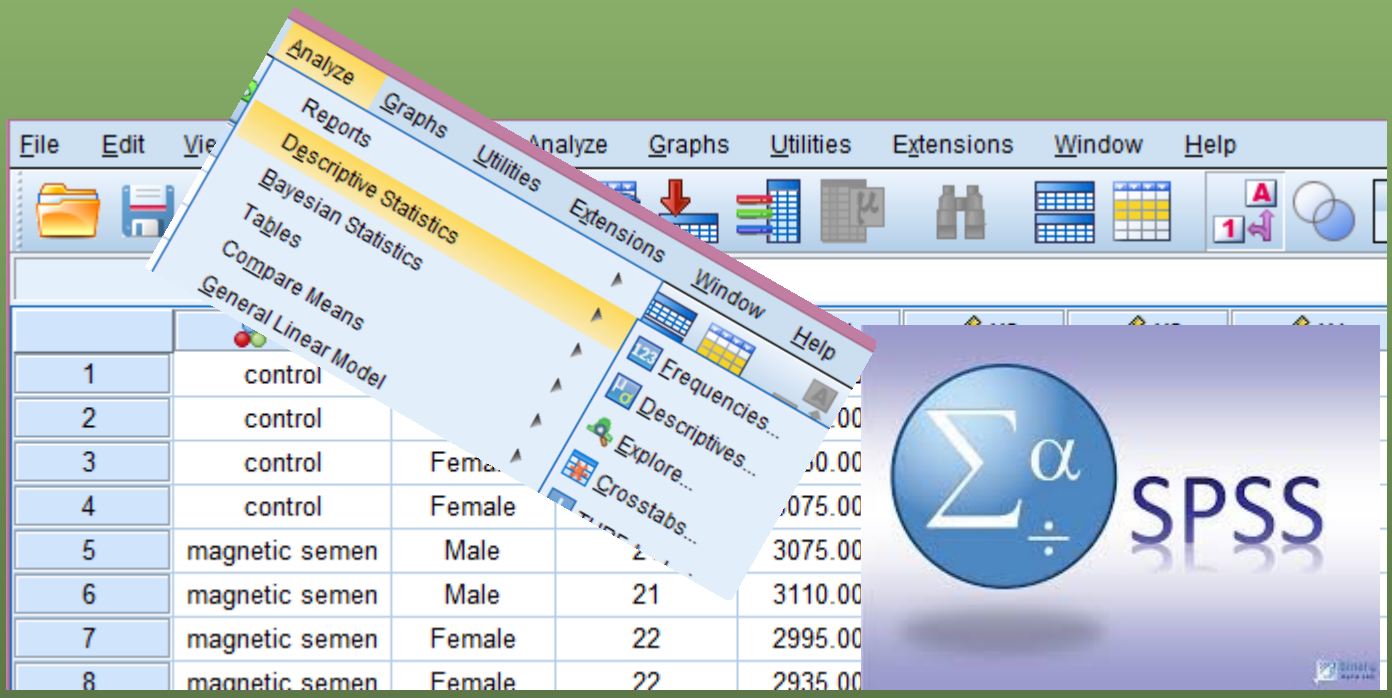
**Republic of Iraq**

**Ministry of Higher Education & Scientific Research**

**University of Diyala**

# Statistical Analysis : Statistical Concepts and Applications in SPSS

Professor Dr.
Khalid Hamid Hassan




**2024**

View publication stats