ماده الإحصاء

قسم علوم الحياه

المرحله الثانيه

المقدمة:

تنقسم المنهجية الإحصائية إلى نوعين رئيسيين هما الإحصاء الاستدلالي (Statistical Inference) والإحصاء الوصفي(Descriptive Statistics) ، وذلك وفقًا لطبيعة البيانات المستخدمة. تشمل الإحصاءات الوصفية رسم البيانات وتصميم الجداول التكرارية وحساب المقاييس المركزية ومقاييس التشتت لتوفير وصف شامل لتلك البيانات لاتخاذ القرارات.

أحد فروع علم الإحصاء الذي يهتم بالبيانات الحيوية أو الطبية أو الصحية يُسمى الإحصاء الطبي أو الإحصاء الحيوي .(Biostatistics) وما يميز هذا العلم هو أنه يقدم وصفًا دقيقًا لمختلف الظواهر الحيوية أو العلاقات التي تربط بعض المتغيرات الحيوية مثل الجنس والعمر والطول ولون البشرة ولون العين وما إلى ذلك مع الأمراض التي تصيب الإنسان أو الحيوان على حد سواء. ومن خلال دراسة هذه العلاقات وغيرها التي تندرج ضمن هذا العلم، يتم اتخاذ بعض القرارات الصائبة.

يُعتبر علم الإحصاء علمًا قائمًا بذاته، ويُعتبر أداة لتحليل بيانات الظواهر المختلفة على أسس رياضية خاصة به. ويستخدم على نطاق واسع في دراسة مختلف العلوم الصرفة والإنسانية لرسم سياساتها وخططها المستقبلية، وللتخطيط التخطيط العلمي السليم، ويؤيده البيانات الإحصائية التي تم جمعها بالاستقصاء الإحصائي العلمي للظواهر المختلفة. ويهدف هذا الكتاب، بشكل خاص، إلى تمكين الطالب من: التعرف على مفهوم الإحصاء الحيوي، ومجالاته، والتعرف على مفهوم ومبادئ ومقاييس الإحصاءات الحيوية. لذا، يقدم هذا الكتاب المفاهيم والمبادئ الأساسية للإحصاء الحيوي، التي تعتبر ضرورية وجوهرية لبناء خلفية إحصائية أكاديمية متماسكة لدى غير المتخصصين وذوي القدرات الرياضية المحدودة.

يحتل الإحصاء مكانًا بين العلوم لما له من استعمالات واسعة للوصول إلى قرارات صائبة لوصف أو تفسير الظواهر المختلفة . استخدام الإحصاء الحيوي في مجالات متنوعة مثل البحوث السريرية والتجارب السريرية والتركيب العشوائي للتجارب والدراسات التوافقية وتحليل البيانات الحيوية.

في البحوث السريرية، يتم استخدام الإحصاء الحيوي لتصميم الدراسات وتحليل البيانات المتعلقة بالمرضى والعلاجات الطبية. يمكن استخدام تصميمات الدراسات الاستطلاعية أو التجارب السريرية العشوائية لتقييم فعالية العلاجات أو لدراسة ارتباط بين المتغيرات الحيوية المختلفة.

في التجارب السريرية، يتم تصميم دراسات لتقييم فعالية العلاجات الجديدة أو الأدوية في البشر. يتم توزيع المشاركين بشكل عشوائي إلى مجموعتين، مجموعة تتلقى العلاج الجديد ومجموعة تتلقى علاجًا وهميًا أو علاجًا قياسيًا. ثم يتم تحليل البيانات لمقارنة فعالية العلاجات.

يستخدم التركيب العشوائي للتجارب في الإحصاء الحيوي لتصميم الدراسات التجريبية. يتم توزيع المشاركين بشكل عشوائي إلى المجموعات المختلفة للتحكم في المتغيرات المعروفة وغير المعروفة التي يمكن أن تؤثر على النتائج.

يستخدم التحليل الإحصائي في الإحصاء الحيوي لتحليل البيانات المتعلقة بالمرضى والعلاجات الطبية. يمكن استخدام الاختبارات الإحصائية المختلفة لتقييم الفروق في المتغيرات بين المجموعات المختلفة ولتحديد العلاقات بين المتغيرات.

بالإضافة إلى ذلك، يتم استخدام الإحصاء الحيوي في تحليل البيانات الحيوية مثل الدراسات الوبائية والدراسات الاستطلاعية لفهم انتشار الأمراض وعوامل الخطر والعوامل المؤثرة على الصحة والسلوك الصحي.

باختصار، يلعب الإحصاء الحيوي دورًا مهمًا في فهم وتحليل البيانات الحيوية والطبية. يساهم في اتخاذ القرارات الصحية وتقييم فعالية العلاجات وفهم العوامل المؤثرة على الصحة والمرض.

عرض وتلخيص البيانات Data presentation and Summarization

بعد جمع بيانات كافية عن ظاهرة معينة لابد من تحليل هذه البيانات وبالتالي الوصول من خلال هذه البيانات إلى اتخاذ قرار سليم حولها . ولكن لا يمكننا تحليل هذه المعلومات ما لم يتم تفريغها وتبويبها وبيان ميزاتها .

إن تبويب االبانات يمكننا من التعامل بسهولة مع الكم الكبير من الأرقام وذلك من خلال نظرة سريعة . ويراعى في هذا التفريغ الإشارة إلى الأرقام الملفتة للنظر بكبرها أو بصغرها .

وهناك أمور أخرى يمكن ملاحظتها من خلال التبويب كتجمع الأرقام حول قيمة معينة ، نسبة مجموعة إلى مجموعة أخرى أو مجموعة بالنسبة للمجموع الكلي أو ما شابه، ذلك وعلى سبيل المثال لو سألنا 190 شخصاً من ركاب طائرة متجه إلى الخليج العربي عن جنسياتهم لاختلفت الإجابات باختلاف الأشخاص وبالتالي فإن الاستمارات الموزعة عليهم سوف تحمل عدة إجابات وتكون الخطوة الأولى هي إعداد جدول (أو جداول) لتفريغ هذه المعلومات (الإجابات - Data) بحيث يحوي الجدول عدة حقول بعدد جنسيات الركاب . ونبدأ بفرز الإجابات وذلك بوضع إشارة (tally) في العمود المناسب والمكان المناسب,

وهناك ثلاث طرق اساسية لعرض وتلخيص البيانات وهي :-

1- طريقة العرض الجدولي Tabular Presentation
2- طريقة العرض البياني Graphic Presentation
3- حساب المقاييس الاحصائية

1-1 جداول التوزيعات التكرارية : Frequency Distributions

تأخذ الظواهر الإحصائية المدروسة قيماً عددية كثيرة ومتكررة وفي بعض الأحيان تكون النتائج الملاحظة غير عددية ، فيمكن في هذه الحالات تحويلها إلى قيم عددية ، مثلاً يمكن تحويل " نعم " أو " لا " أو " صح " أو " خطأ " .....الخ إلى " 1 " أو " صفر " . مما يتيح بتشكيل جداول تكرارية ،.

إن تصنيف وتبويب جميع المشاهدات يعني بالضرورة ترتيب هذه البيانات تصاعديا أو تنازلياً مما يسمح لنا استخلاص صورة واضحة عن المدى "Range" الذي تتراوح فيه البيانات على عدد من الفئات "Classes" معتبرين هذه الفئات وجوها للظاهرة المدروسة حيث يتم تفريغ المعلومات على أساس هذه الفئات، ومن ثم نحدد العدد المقابل لكل فئة من هذه الفئات لنستنتج تكرارات القيم العددية ضمن فئاتها . ونسمي الجدول الذي يضم الفئات والتكرارات المقابلة لها جدول التوزيع التكراري Frequency Distribution Table .

لنفرض أننا قمنا بتقسيم قراءات ظاهرة معينة والبالغ عددها n قراءة إلى k فئة ، فإننا نسمي في هذه الحالة نسبة تكرار كل فئة على المجموع الكلي للتكرارات بالتكرار النسبي .

إذا رمزنا لتكرار الفئة i بالرمز f_i فإننا نجد أن :

$$\sum_{i=1}^{r} \frac{fi}{n} = 1$$

وهذا أمر منطقي لأن.. $\sum_{i=1}^{r} f_i = 1$

حيث ان $\sum$ هو حرف يوناني ويقرأ سيكما (segma) ويرمز هنا كما يرمز دوماً للمجموع .

مراحل إنشاء الجداول التكرارية لمجموعة من البيانات بالشكل التالي :

1- اختيار عدد الفئات وطول كل فئة وهذا يعتمد على شكل البيانات الموجوده ويجب أن تراعي ما يأتي :
نختار عدد الفئات بين 5 و 15 فئة.

2- أن كل عدد موجود من البيانات أو المشاهدات تقع في فئة واحدة فقط.
يعني بذلك أن القراءات الحدية والواقعة بين فئتين متتاليتين أن تنتمي إلى أحد الفئتين وليس لكليهما، وبالتالي لا تحوي الفئات المتتالية بيانات أو أرقام مشتركة ولهذا الغرض نقوم بإضافة نصف واحدة دقه إلى الحد الأعلى للفئة ونطرح نصف واحدة دقه من الحد الأدنى للفئة وهذا ما نسميه حدود الفئة الفعلية Class boundaries" ، في حالة الأعداد الصحيحة أو البيانات المتقطعة, بينما نضيف ونطرح 0.05 أو 0.005 حسب عدد خانات العدد .

3- نجعل الفئات ذات أطول متساوية وبالتالي تغطي مجالات متساوية . مع العلم أنه يمكن أن تكون الفئات ذات أطوال مختلفة ولكن لن نتعرض لدراسة هذا النوع من الجداول التكرارية في هذا الكتاب.
ومن الأسهل أن تختار أطوال الفئات من الشكل 5 , 10 , 1000،100،20 وذلك لتسهيل الدراسة .

اطوال الفئات اما ان تكون مغلقة أو تكون نصف مغلقة أو تكون نصف مفتوحة وذلك باختيار أقل من أو أقل أو يساوي (≤)وذلك لجعل عدد الفئات المطلوبة اقل عندما يكون عدد كبير من الفئات أصغر بكثير أو أكبر بكثير من باقي الفئات، ومع ذلك يجب الامتناع عن ذلك إن أمكن وذلك لصعوبة الحسابات الناتجة عن ذلك .

مثال : التالي يمثل درجات الطلاب في امتحان

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

كون جدول تكراراي لهذا الجدول؟

الحل

1- الفئات Classes:- وهي المجاميع التي قسمت اليها قيم المتغير وكل فئة تأخذ مدى معين من قيم المتغير ولكل فئة دائما يكون لدينا حد اعلى وحد ادنى وكذلك حدان حقيقيان حد ادنى حقيقي وحد اعلى حقيقي وتدعى بالحدود الفعلية .

الحد الادنى Lower class limit: هي اقل قيمة من قيم المتغير يسمح لها بالدخول في هذه الفئة على سبيل المثال الفئة (.10 -39) فالحد الادنى للفئة هو(10)

الحد الاعلى upper Class limit : وهو اكبر قيمة من قيم المتغير يسمح لها بالدخول في هذه الفئة مثل (10 – 39) فالحد الاعلى (39) ، ولكل فئة حدان حقيقيان حد ادنى حقيقي Lower class boundary وحد اعلى حقيقي Upper class Boundary وقد تسمى هذه الحود الحقيقية بالحدود الفعلية حيث يتم تحويل البيانات من بيانات منفصلة الى بيانات مستمرة او متصلة .

2- طول الفئة Class length or Class widthهو مقدار المدى بين حدود الفئات العليا والدنياوهذا يستحسن ان تكون اطوال الفئة متساوية وسنرمز لطول الفئة بالرمز (M) .

3- مركز الفئة Class mark :هو عبارة عن منتصف المدى بين حدي الفئة وسنرمز بالرمز (C.C)

4- تكرار الفئة Class frequency :هو عدد المشاهدات التي تقع في مدى تلك الفئة وسنرمز له بالرمز (f_r) ويكون مجموع التكرارات مساوية للعدد الكلي لقيم الظاهرة .

وسوف نوضح ما سبق شرحه بالتفصيل في الجدول التالي الذي يبين توزيع درجات طلبة كلية طب الاسنان في ماده الاحصاء

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

باختيار عدد الفئات 6 يمكن أن نحسب طول الفئة وهي هنا 5 ~ 4.833 = 6 ÷ 29

وفي مثل هذه الحالات يقرب الناتج للأعلى دوماً ، أي لأقرب عدد صحيح تالي. وبالتالي فإن حدود الفئات هي

10-14 , 15-19 , 20-24 , 25-29 , 30-34 and 35-39

لاحظ أنه يمكن اختيار عدد آخر للفئات فإذا اخترنا العدد 8 مثلاً فإن أطوال الفئات سيكون

4 ~ 3.625 = 8 ÷ 29 وسوف تكون الفئات الثمان كالتالي:

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

ملاحظه/

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

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

لاحظ في الجدول أن الحدود الفعلية للفئة الأولى 9.5-14.5 وللفئة الثانية 14.5-19.5 ..... وهكذا.

ونحصل على هذه الحدود بطرح و جمع نصف واحدة دقة للحد الادنى والحد الاعلى على الترتيب .
والمقصود بنصف واحدة الدقة هو (0.5) اذا كانت البيانات صحيحة و(0.05) اذا كانت البيانات تحتوي على عدد عشري واحد بعد الفاصلة و(0.005) اذا كانت البيانات تحتوي على عددين عشريين بعد الفاصلة .
وفي الجدول أعلاه يمثل العمود في أقصى اليمين التكرار(Frequency) عدد مرات وقوع المشاهدات في الفئات الملائمة وهو ما ندعوه تردد الفئة أو التكرار Class Frequency .
إن أصغر رقم وأكبر رقم في عمود أقصى اليسار يعبر عنه بحدود الفئات Class limits فالأرقام

10 15 20 25 30 35

هي الحدود الدنيا للفئات بينما

14 19 24 29 34 39

فتسمى الحدود العليا للفئات Lower class limits and Upper class limits على الترتيب .
"Class boundaries" أو " Real class limits "، إن حدود الفئات هي الأكثر استعمالاً أو شيوعاً .
تعتبر مراكز الفئات Class marks إحدى عناصر التوزيعات التكرارية الهامة وهي بالتعريف منتصف طول الفئة ونحصل عليه بإضافة الحد الأدنى للفئة لحدها الأعلى وقسمة الناتج على 2 أي 2 /(a+b) حيث a الحد الأدنى b الحد الأعلى . أما طول الفئة Class interval فهو الفرق بين الحد الأعلى الفعلي و الحد الأدنى الفعلي وغالباً ما يسمى مدى الفئة . نحصل على مراكز الفئات بحساب مركز الفئة الأولى ومن ثم نضيف له طول الفئة فنحصل على مركز الفئة الثانية ، نضيف طول الفئة للناتج فنحصل على مركز الفئة التالي... وهكذا . وبالعكس فإن الفرق بين مركزي فئتين متتاليتين يعطي طول الفئة .

مثال :

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

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

مثال :

احسب تكرار التوزيع النسبي percentage distribution للفترات الزمنية التي قضاها 80 طالباً في المكتبة خلال أسبوع ما والمفرغة في الجدول 2-3 .
الحل :
لقد شاهدنا في الجدول 2-3 التردد لكل فئة وبالتالي يمكن قسمة تردد كل فئة على المجموع الكلي وضرب الناتج بـــ 100 أي أن الفئة الأولى تحوي $\frac{8}{80}\times 100 = 10\%$ من البيانات وبالنسبة للفئة الثانية فتحوي $\frac{28}{80}\times 100 = 35\%$ من البيانات و هكذا، وأخيراً الفئة الأخيرة تحوي $\frac{1}{80}\times 100 = 1.25\%$ من البيانات . وهكذا يمكن تلخيص ذلك بجدول على غرار الجدول السابق يحتوي نتائج المثال الحالي والسابق بالشكل الآتي :

Class limits	Frequency	Percentage	Class marks
10-14	8	(8÷80)×100=10%	(10+14) ÷2=12
15-19	28	(28÷80)×100=35%	(15+19) ÷2=17
20-24	27	33.75 %	22
25-29	12	15 %	27
30-34	4	5 %	32
35-39	1	1.25 %	37

TOTEL	80	100%	

هذا ويمكن تعديل جدول التوزيع وتحويله إلى جدول من الشكل " أقل من أو يساوي " أو " أقل من " أو "أكثر من " وهذا ما يدعى بـ " التوزيعات التكرارية المتجمعة " Cumulative frequency distribution. ونحصل على ذلك بجمع تكرار الفئات بداية من الفئة الأولى إلى الثانية ، الناتج إلى الثالثة الناتج الى الرابعة ...وهكذا .

مثال:

شكل جدول توزيع متجمع " أقل من "(Less than) للبيانات الواردة في المثال الأول أعلاه والمفرغة في الجدول 2-3

الحل :

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

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

هناك جداول تكرارية معاكسة للجدول التكراري الصاعد وهي جداول تكرارية تناقصية و مشابهة للجدول أعلاه ولكنها تتناقص من المجموع .إن مجموع القيم السابقة هو 80 فتكون الفئة الثانية 72=8-80 والفئة الثالثة 44 = 28-72 وهكذا .... ويكون التكرار المتناقص للفئة الأخيرة 0. لاحظ في العمود الاخير هناك

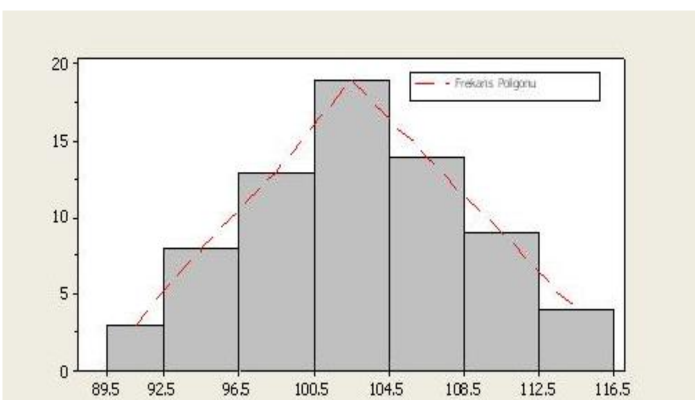
80 قراءة أكبر من 10 لكن 0 قراءة اقل من 10 ، 8 قراءات أقل من 15 ولكن 72 قراءة أكبر من 15 . وهكذا .. وهذا ينطبق على بقية الفئ

التمثيل البياني للبانات

التمثيل البياني هو طريقة لتوضيح نتائج الدراسة الإحصائية بيانيا. هناك العديد من أنواع الرسوم البيانية المختلفة, بعد الانتهاء من تشكيل جدول التوزيع التكراري بضغط العدد الكبير للمعلومات وعرضها بشكل يسهل التعامل معه في بيان القيم الأكثر تكراراً ، الأقل تكراراً ، الأكثر تطرفاً .... الخ يمكن عرض النتائج بيانياً. والعرض البياني للبيانات هو أحد طرق التي يمكن استخدامها في وصف البيانات، من حيث شكل التوزيع ومدى تمركز البيانات، وفي كثير من النواحي التطبيقية يكون العرض البياني أسهل وأسرع في وصف الظاهرة محل الدراسة، وتختلف طرق عرض البيانات بيانيا حسب نوع البيانات المبوبة في شكل جدول تكراري، وفيما يلي عرض للأشكال البيانية المختلفة.

1-المدرج التكراري Histogram : المدرج التكراري هو التمثيل البياني للجدول التكراري البسيط الخاص بالبيانات الكمية المتصلة، وهو عبارة عن أعمدة بيانية متلاصقة، حيث تمثل التكرارات على المحور الرأسي، بينما تمثل قيم المتغير (حدود الفئات) على المحور الأفقي، ويتم تمثيل كل فئة بعمود، ارتفاعه هو تكرار الفئة، وطول قاعدته هو طول الفئة .

تتمثل هذه الطريقة برسم مجموعة من المستطيلات المتلاصقة ذات عرض واحد ولكنها بأطوال مختلفة حيث يتناسب طول كل مستطيل مع تكرار الفئة التي يمثلها وتكون المستطيلات المتلاصقة المتساوية العرض (عرض المستطيل) منطبقة على المحور الأفقي ومراكز هذه القواعد منطبقة على مراكز الفئات بينما يشار إلى أطوال هذه المستطيلات (ارتفاعات هذه المستطيلات) بتكرار كل فئة على المحور الرأسي .



مع ملاحظه أن المدرج التكراري لا يمكن استخدامه عندما تكون الفئات مفتوحة ، وكذلك عندما تكون الفئات غير متساوية ففي هذه الحالة يفضل استخدام مساحة المستطيلات للدلالة على التكرار عوضاً عن ارتفاعات المستطيلات في الفئات المتساوية.

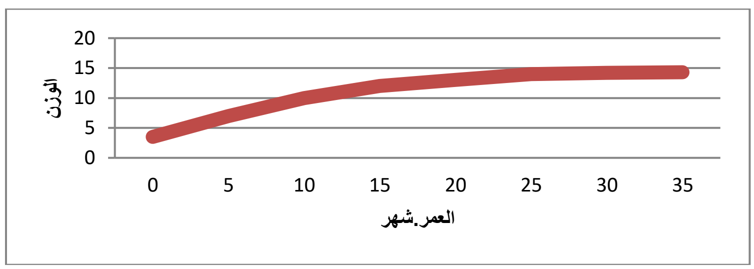
ويستخدم المدرج التكراري لتمثيل البيانات التي تم عرضها في جدول توزيع تكراري، وفيه يمثل كل مستطيل فئه من فئات التوزيع التكراري.
يتم تقسيم المحور العمودي فى المدرج التكرارى حسب التكرار (فقد نستخدم التكرار الأصلي في حالة تمثيل التوزيع التكرارى، وكذلك يمكن أن نستخدم التكرار النسبى في حالة تمثيل التوزيع التكرارى النسبى).

2- المنحى التكراري Polygon :

إذا جعلنا المضلع التكراري منحنى بدلا من خطوط منكسرة فإننا نحصل على المنحنى التكراري، ويلاحظ أنه ينبغي عدم رسم المنحنى التكراري إلا إذا كانت الفئات كثيرة العدد، وذات طول صغير وكان عدد البيانات كبيرا وكانت هذه البيانات من النوع المتصل مثلا الوقت والوزن.

خطوات رسم المنحنى التكراري:

- نرسم محورين أفقي وعمودي بحيث يمثل المحور الأفقي الفئات والمحور العمودي التكرارات.
- نقوم بتكوين فئه سابقه عند النقطة الأولى في التوزيع التكراري يقابلها تكرار أصلى يساوى الصفر.
- نقوم بإنشاء فئه لاحقه للفئه الأخيرة في التوزيع التكراري يقابلها تكرار أصلى يساوى الصفر أيضاً.



التوزيعات التكرارية المتجمعة:

التوزيعات التكرارية المتجمعة هي مفهوم إحصائي يستخدم لتوصيف تراكم القيم في توزيع معين. تسمح هذه التوزيعات بتحليل وفهم البيانات التي تحتوي على قيم متكررة وتتيح تقدير الاحتمالات المرتبطة بكل قيمة.

توزيع التكرار المتجمع (Cumulative Frequency Distribution) يوفر معلومات حول عدد المرات التي يظهر فيها قيمة معينة أو أقل في مجموعة البيانات. يتم ترتيب القيم في التوزيع بترتيب تصاعدي ويتم حساب التراكم التراكمي لكل قيمة.

التوزيعات التكرارية المتجمع (الصاعد):

يتم استخدام المنحنى المتجمع لتمثيل التوزيع التكراري المتجمع سواء بالقيم المطلقة للتكرارات أو بالتكرار النسبي. يتم توضيح وضع النقاط المتعلقة بالتكرارات عند الحد الأعلى لكل فئة في المنحنى المتجمع الصاعد، وذلك لأنه يعبر عن العدد الإجمالي لأوجه الظاهرة التي تتواجد أسفل الحد الأعلى للفئة.

على سبيل المثال، إذا كان لدينا مجموعة بيانات تتضمن درجات طلاب في امتحان، يمكننا استخدام المنحنى المتجمع لتمثيل توزيع تكرار الدرجات. سيتم تحديد قيم التكرار النسبي أو القيم المطلقة للتكرارات لكل فئة من الدرجات، ويتم رسم النقاط على المنحنى المتجمع الصاعد بناءً على القيم النسبية أو القيم المطلقة.

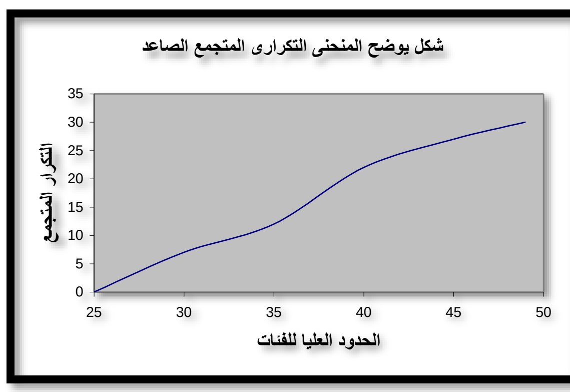
يمكن استخدام المنحنى المتجمع لتحليل البيانات بشكل أكثر تفصيلاً، ومقارنة توزيع القيم بين مجموعات مختلفة، وحساب النسب المئوية للقيم في فئات محددة، وتحديد النقاط الحدودية الهامة في التوزيع. كما يمكن استخدام المنحنى المتجمع للتنبؤ بالقيم المستقبلية أو اتخاذ قرارات استراتيجية بناءً على التحليل الإحصائي.

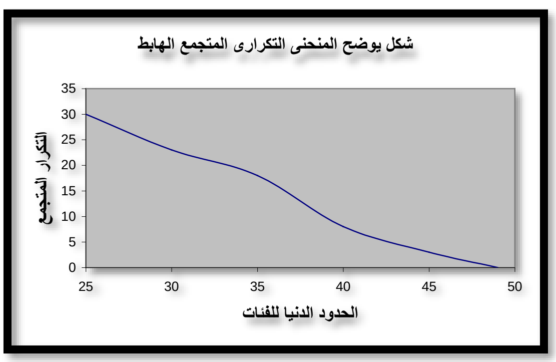
<u>المنحنى المتجمع الهابط (النازل):-</u>

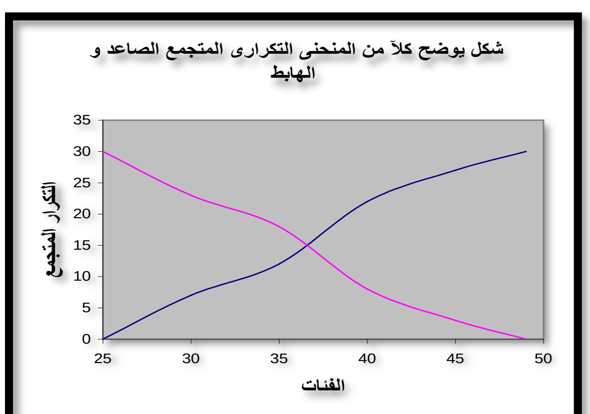
لمنحنى المتجمع الهابط أو المنحنى المتجمع النازل هو نوع آخر من المنحنيات التكرارية المتجمعة. يُستخدم هذا النوع من المنحنيات لتمثيل توزيع القيم التكرارية التي تنخفض أو تناقص بمرور الوقت أو مع زيادة القيم.

عند رسم المنحنى المتجمع الهابط، يتم ترتيب القيم في ترتيب تصاعدي ويتم حساب التراكم التراجعي لكل قيمة. يتم تمثيل النقاط على المنحنى بناءً على القيم التراكمية التنازلية، حيث يتم تحديد قيمة التكرار النسبي أو القيمة المطلقة للتكرارات لكل فئة أو قيمة. يمكن استخدام المنحنى المتجمع الهابط لتحليل الاتجاهات الزمنية أو التطورات المتناقصة في البيانات. يمكن استخدامه لدراسة النماذج السلوكية التي تنخفض مع مرور الوقت أو لتحليل تراجع الأحداث في تسلسل زمني محدد

<u>التوزيعات التكرارية المتجمعه:</u>







التمثيل الدائري: Pie chart

يمكننا رسم دائرة وتقسيمها إلى قطاعات دائرية تتناسب مساحتها (زاويتها) مع تكرار الفئة التي تمثلها. في هذه الحالة، يتم حساب مساحة القطاع الدائري المقابل لكل فئة بناءً على التكرار النسبي لتلك الفئة وتضربه في 360 درجة، والتي تمثل زوايا الدائرة.

لنقم بتوضيح ذلك باستخدام مثال:
لنفترض أن لدينا بيانات توضح توزيع الدرجات في امتحان لمجموعة من الطلاب، وتم تجميع البيانات في الجدول التالي:

الفئة	التكرار
فئة 1	8
فئة 2	12
فئة 3	6
فئة 4	10
فئه5	4

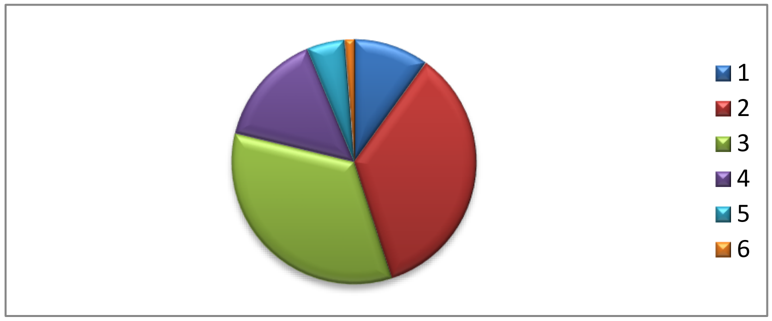
حساب التردد النسبي لكل فئة، نقوم بتقسيم تكرار الفئة على مجموع البيانات الكلي (80 في هذا المثال) ثم نضرب الناتج في 100. هذا يعطينا التردد النسبي لكل فئة.

مثلاً، لحساب التردد النسبي للفئة الأولى:
تردد الفئة الأولى = (8 / 80) * 100 = 10٪

ثم نقوم بضرب التردد النسبي بـ 360 درجة للحصول على زاوية القطاع المقابل للفئة الأولى:
زاوية القطاع الأول = 10٪ * 360 = 36 درجة

وبنفس الطريقة، يمكننا حساب زوايا القطاعات الأخرى المقابلة لكل فئة وهي على التوالي:
زاوية القطاع الثاني = (12 / 80) * 360 = 54 درجة
زاوية القطاع الثالث = (6 / 80) * 360 = 27 درجة
زاوية القطاع الرابع = (10 / 80) * 360 = 45 درجة
زاوية القطاع الخامس = (4 / 80) * 360 = 18 درجة

بهذه الطريقة، يمكننا تمثيل البيانات في شكل دائرة مقسمة إلى قطاعات دائرية، حيث يكون حجم كل قطاع متناسبًا مع تكرار الفئة المقابلة
السادسة . ترسم القطاعات ثم تلون بأشكال مختلفة للتمييز بينها . لاحظ أيضاً أن مجموع زوايا القطاعات الزاوية يساوي 360°



تمارين

1 -تمثل البيانات التالية أطوال 15 طالباً من طلاب قسم وهي معطاة بالسنتيميتر وهي

141	166	136	153	171
161	155	145	183	158
149	138	160	175	150

a-مثل هذه البيانات بطريقة المخطط النقطي

b- مثل هذه البيانات بطريقة الساق والورقة مستخدما للساق الارقام:

18 17 16 15 14 13

2 - تمثل البيانات التالية التقرير الأسبوعي في قسم الإسعاف في إحدى المستشفيات الحكومية .

كسور	24
رضوض	12
عيون	4
موت	3
غيبوبة	3
أعراض أخرى	2

مثل هذه البيانات بيانياً بطريقة الأعمدة Chart-Bar . اختر تمثيلاً بيانياً آخر لهذه البيانات ونفذه .

3 –التالي درجات 60 طالبا لماده علم الطفيليات :

36.9	24.6	34.1	25.4	20.6	29.5	33.7	27.3	21.2	28.3
32.5	15.2	27.1	35.5	29.4	21.8	27.5	28.9	24.8	29.6
28.1	22.2	33.5	29.3	32.7	28.4	21.3	25.0	21.9	37.5
25.4	22.7	29.6	17.3	29.5	26.9	34.6	30.2	29.0	26.8
23.0	24.5	29.3	23.9	36.8	28.7	33.2	23.6	34.5	31.3
29.2	34.8	37.0	38.4	31.0	24.0	28.3	18.6	23.5	26.4

وزع هذه البيانات وفق الفئات التالية :39.9- 15.0-19.9, 20.0-24.9,....35.0 ثم أنشئ الجدول التكراري و مثل ذلك بيانياً بطريقة المستطيلات histogram ، ثم بطريقة الدائرة Pie chart .

4 - أوجد التكرار التجميعي الصاعد للبيانات الواردة في المسألة السابقة

الفصل الثاني

مقاييس النزعه المركزيه والتشتت

تسعى الدراسات الإحصائية إلى كشف معالم المجتمعات المرتبطة بالظاهرة المدروسة. ويتطلب ذلك بالطبع حصر جميع البيانات الكمية المتصلة بمشكلة الدراسة والمتوفرة في المجتمع. وفي حال حصر جميع قيم المجتمع فانه يتم الحصول على المعالم المطلوبة والتي تكون بالطبع قيم ثابتة. في المقابل عندما يكون المجتمع غير محدود أو كبير جدا لدرجة تعذر عملية حصر جميع القيم الموجودة فيه فان الأسلوب الإحصائي الأمثل هنا يكمن في استخدام عينة عشوائية ممثلة للمجتمع ومن ثم الحصول على تقديرات للمعالم المطلوبة. وبحكم كون العينة عشوائية فان جميع المؤشرات الإحصائية التي يتم الحصول عليها من خلال العينة العشوائية والتي يطلق عليها بإحصائيات العينة تعتبر متغيرات عشوائية تختلف قيمها تبعا لاختلاف العينة العشوائية المستخدمة. لذا فان عملية حساب المؤشرات الإحصائية سواء المتعلقة بالنزعة المركزية أو بالتشتت ستختلف تبعا لنوع وعاء البيانات الإحصائية المستخدم، مجتمع أم عينة عشوائية.

مقاييس النزعة المركزية

تهتم مقاييس النزعة المركزية بتوفير مؤشرات كمية تمثل التوجه العام لقيم المتغير الكمي المدروس، حيث يتم الحصول على مؤشر يفيد عن توجه القيم، دون الحاجة إلى التعامل مع جميع القيم المختلفة للمتغير الكمي. لذا فان مقاييس النزعة المركزية تنتج في النهاية أرقام محدودة تمثل التقديرات لتلك المقاييس وذلك بغض النظر عن عدد القيم الأصلي، سواء كان صغيراً أم كبيراً. كما أن الاستعاضة عن القيم كلها برقم واحد يفقدنا كثير من المعلومات حول البيانات، إلا أنه لا يمكن حجب الأهمية الكبيرة والدور الهام لمقاييس

النزعة المركزية في مجال الاستدلال الإحصائي، حيث تعتبر هذه المقاييس حجر الأساس ونقطة البداية لأي دراسة تحليلية إحصائية، كما يتم الانطلاق منها إلى مستويات متقدمة في عمليات التحليل ومن ثم الاستدلال حول توجه البيانات والصفات المميزة للمتغيرات الكمية التابعة لها. تتنوع استخدامات مقاييس النزعة المركزية في عمليات الاستدلال الإحصائي مما ينتج عنه تنوع في طبيعة تلك المقاييس المختلفة، لذا سيتم التطرق في هذا الفصل إلى ثلاث أنواع من مقاييس النزعة المركزية والتي تمثل الأهم، وهي كل من الوسط الحسابي والوسيط والمنوال

الوسط الحسابي:

يمثل الوسط الحسابي أو المتوسط (Arithmetic Average or Mean) مقياس النزعة المركزية الأكثر شهرة والأكثر أهمية في المقاييس المختلفة. وتمثل قيمة الوسط الحسابي القيمة التي تتمركز حولها جميع القيم المختلفة للمتغير الكمي. يمكن الحصول على القيمة الحقيقية لمتوسط متغير عشوائي في مجتمع محدود إذا تم التعامل مع كافة القيم في المجتمع. في هذه الحالة يرمز لقيمة الوسط الحسابي المحصل بالرمز μ والتي تمثل معلمة المجتمع. وبافتراض التعامل مع متغير عشوائي X لمجتمع محدود حجمه N فانه يمكن حساب قيمة الوسط الحسابي من خلال الدالة التالية (لبيانات خام أو غير مبوبة)،

$$\mu_X = \frac{\sum_{i=1}^{N} X_i}{N}$$

أما في حال التعامل مع عينة عشوائية ممثلة لمجتمع الدراسة حجمها n قراءة، فان معادلة تقدير قيمة الوسط الحسابي يتم من خلال الدالة،

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

حيث تم الإشارة رياضيا لقيمة الوسط الحسابي باستخدام الشرطة على رمز المتغير الكمي. وبالطبع تعتبر القيمة $\bar{X}$ إحصائية عينة تمثل تقدير مقبول إحصائيا لقيمة معلمة المجتمع المجهولة μ.

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

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$= \frac{x_1 + x_2 + ... + x_n}{n}$$

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

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

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

أمثلة :

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

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

إن متوسط عدد الرسائل اليومية في هذه الحالة هو :

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

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

$$\bar{x} = \frac{\sum_{i}^{k} x_i f_i}{\sum_{i}^{k} f_i} = \frac{\sum_{1}^{8} x_i f_i}{\sum_{1}^{8} f_i}$$

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

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

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

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

2- الوسيط: ***Median***

وسيط هو قيمة يتساوى عدد القيم على طرفيها بعد ترتيبها تصاعديًا. ببساطة، الوسيط هو القيمة التي تقع في المنتصف عند ترتيب القيم من الأصغر إلى الأكبر.عندما يكون لدينا عدد فردي من القيم، فإن الوسيط هو القيمة النصفية التي تقسم القيمة إلى نصفين متساويين. على سبيل المثال، إذا كانت هناك 9 قيم، فإن الوسيط سيكون القيمة التي تقع في المركز بعد ترتيب القيم، حيث ستكون هناك 4 قيم أصغر من الوسيط و4 قيم أكبر من الوسيط.عندما يكون لدينا عدد زوجي من القيم، فإن الوسيط يتم حسابه بجمع القيمتين المتواجدتين في المنتصف ومن ثم قسمهما على 2. على سبيل المثال، إذا كان هناك 8 قيم، فإن الوسيط سيكون المتوسط الحسابي للقيمتين الموجودتين في المنتصف بعد ترتيب القيم.

الوسيط يوفر مقياسًا للموضعية ويعكس القيمة التي تقسم القيمة إلى نصفين متساويين. إنه مقياس مهم في تحليل البيانات ويستخدم لتوصيف النزعة المركزية أما إذا كان عدد القيم زوجياً فالوسيط هو الوسط الحسابي لمجموع القيمتين الوسيطتين ويرمز للوسيط بالرمز $\tilde{x}$ وهو القراءة التي ترتيبها $\frac{n+1}{2}$ في حالة n عدد فردي.

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

أمثلة :

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

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

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

b) لتكن القيم 4,-5,6,-1,3,8,10,11 والتي تصبح بعد ترتيبها تصاعدياً 5,-1,3,4,6,810,11- أوجد وسيط هذه البيانات .

الحل :

بما أن $n = 8$ وعددها زوجياً فإن الوسيط هو متوسط القيمتين الوسيطيتين, أي $\frac{8+1}{2} = 4.5$ وبالتالي فهو الوسط الحسابي للقيمتين الرابعة والخامسة

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

و تستخدم هذه الطريقة إذا كان عدد البيانات صغيرا . أما إذا كان عدد البيانات كبيرا و غير مرتب في جدول تكراري فان هناك طريقة عامة سنذكرها في الفقرة القادمة.

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

ويعطى الوسيط $\tilde{x}$ في هذه الحالة بالعلاقة التالية :

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

حيث نرمز بـ a للحد الأدنى الفعلي للفئة الوسيطية ، Δ طول الفئة ، k مجموع تكرارات الفئات التي تسبق a أما j فهو تكرار الفئة الوسيطية وأخيراً $n = f_1 + f_2 + ... + f_k$. لاحظ هنا أن العلاقة السابقة هي علاقة عامة حيث الوسيط انما هو P_{50} وبالتالي فيمكن حساب أي مئوي بنفس الطريقة.

c) أوجد الوسيط لمجموعة القيم المعطاة بالجدول

التالي :

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

لاحظ عمود التكرار التجميعي الصاعد تجد أن :

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

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

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

<u>المنوال</u>

يمثل المنوال (Mode) القيمة الأكثر شيوعا من بين القيم المختلفة للمتغير العشوائي محل الدراسة. ويتم تحديد قيمة المنوال من خلال تحديد تكرار جميع القيم المختلفة للمتغير العشوائي محل الدراسة إذا كانت البيانات غير مبوبة (بيانات خام)، بينما يتم الاستعانة بقاعدة رياضية إذا كانت قيم المتغير العشوائي متوفرة في جدول تكراري (بيانات مبوبة)
بالنسبة للبيانات الخام، يتم تحديد قيمة وحيدة للمنوال إذا وجدت قيمة واحدة تكررت أكثر من باقي القيم المختلفة للمتغير العشوائي. كذلك يمكن أن يكون المنوال متمثل بأكثر من قيمة إذا كان هنالك أكثر من قيمة واحدة لها نفس التكرار الأكثر من بين جميع التكرارات المتوفرة. وفي حال عدم تكرر أي قيمة من قيم المتغير العشوائي المختلفة فانه في هذه الحالة لا يكون هنالك منوال بين قيم المتغير العشوائي.

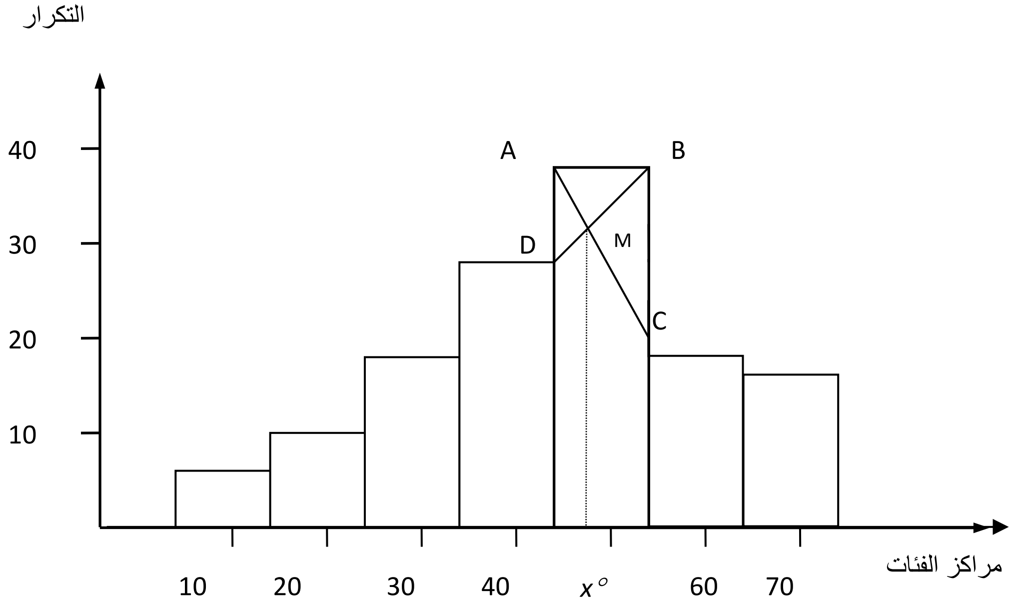
المنوال: بيانات خام (غير مبوبة)
المنوال لبيانات خام هو القيمة أو القيم الأكثر شيوعا أو تكرارا.

يعرف المنوال على أنه هو القيمة الأكثر تردداً أو الأكثر شيوعاً أو تكراراً ونرمز له بالرمز x° فمثلاً للقيم 1,1,2,3,3,4,5,5,5,6,7 منوالا واحدا وهو $x^{\circ}=5$
قد لا يكون لمجموعة من القيم منوال كما هو الحال في القيم 1,2,3,4,5,6,7 وقد يكون لمجموعة من القيم أكثر من منوال واحد كما هو الحال في القيم 1,1,1,2,2,2,3,4,5,5,6 .
و إذا كانت البيانات غير عددية (أي نوعية) كأن نسأل مثلا ستة أشخاص عن أي الألوان المحببة لهم ، قد تكون إجاباتهم كالتالي:
أزرق, أصفر, أبيض, أزرق, أبيض أو أزرق فسيكون المنوال في هذه الحالة هو اللون "الأزرق" .
أما إذا كانت البيانات مبوبة في جدول توزيع تكراري فنحسب المنوال كما يلي : نعين الفئة المنوالية وهي الفئة التي تقابل أكبر قيمة للتكرارات ثم نحسب الفرق بين تكرار الفئة المنوالية وتكرار الفئة التي قبلهاΔ_1 ثم نحسب الفرق بين تكرار الفئة المنوالية وتكرار الفئة التي بعدها وليكن Δ_2 نرمز للحد الأدنى الفعلي للفئة المنوالية بالرمز a ولطول الفئة بالرمز Δ فنحصل على المنوال كما يلي :

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

وكتطبيق مباشر على ذلك ، المنوال للقيم المعطاة بالجدول 3-1 أعلاه هو

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



وذلك وفقاً لما يلي :

1- نرسم المدرج التكراري موضحاً الفئة المنوالية والفئة التي قبلها والفئة التي تليها.

2- نوصل بين رأس الفئة المنوالية والزاويتين المجاورتين للفئة المنوالية أي BD , AC .

3- نسقط من نقطة الالتقاء M عموداً على محور السينات تمثل قيمة المنوال وهنا يمكن تعيين النقطة x^{o} بشكل تقريبي وبصورة هندسية.

هناك علاقة تقريبية بين المنوال x° والوسط $\bar{x}$ والوسيط $\tilde{x}$ من أجل التوزيعات المتناظرة نسبيا بحيث تؤمن معرفتنا لنقطتين النقطة الثالثة وذلك من العلاقة التالية :

$$x^{\circ} = \bar{x} + 3(\tilde{x} - \bar{x})$$

لنحسب مثلاً منوال البيانات الواردة في الجدول 3-1، حيث وجدنا أن $\tilde{x} = 45.9$.

أما الوسط الحسابي فمن العلاقة (3-2) نجد أن $\bar{x} = 45.1$

وبتطبيق العلاقة السابقة نجد $x^{\circ} = 45.1 + 3(45.9 - 45.1) = 47.5$

مقاييس ا لتشتت

تمثل مقاييس التشتت الجانب الآخر من المقاييس الإحصائية الأساسية بجانب مقاييس النزعة المركزية، حيث تستخدم تلك المقاييس في وصف البيانات والتعرف على خصائصها. كما تعمل مقاييس التشتت كجزئية مكملة ومهمة جدا بجانب مقاييس النزعة المركزية في عمليات الاستدلال الإحصائي المبنية على عملية التعامل مع البيانات. وينصب الاهتمام عند التعامل مع مقاييس التشتت حول قياس درجة الاختلاف بين القيم

المختلفة للمتغير الكمي المدروس، ويتم ذلك من خلال عدة مقاييس مختلفة يهتم كل واحد منها بقياس درجة الاختلاف من زاوية مختلفة. يمثل التباين والانحراف المعياري والانحراف المتوسط والانحراف الربيعي بالإضافة إلى المدى مقاييس مختلفة لقياس تشتت المتغيرات الكمية.

يتم الحصول على تصور دقيق عن خصائص المتغير الكمي في حال توفر كل من مقياس النزعة المركزية ومقياس التشتت، حيث تعطي مقاييس النزعة المركزية تصور عن تمركز القيم بينما تعطي مقاييس التشتت تصور عن درجة اختلاف تلك القيم عن بعضها البعض. لذا يمكن القول بأن الاعتماد على مقياس واحد قد لا يغني عن الآخر في عملية الاستدلال الإحصائي، حيث ينتج عنه دوما قصور في المعلومة المعتمد عليها ومن ثم عدم القدرة على قراءة البيانات إحصائيا بشكل سليم.

يستخدم المدى الممثل للفرق بين اكبر قيمة واصغر قيمة كمقياس بسيط وسطحي عن درجة تشتت قيم المتغير الكمي. ولكن لا يجب الأخذ بهذا المقياس والاعتماد عليه في العمليات الاستدلالية الإحصائية حيث انه يتأثر بشدة بالقيم المتطرفة بالإضافة إلى عدم استخدامه لباقي قيم المتغير الكمي. وكمقياس أدق يعمل في حال وجود قيم متطرفة، يمكن استخدام الانحراف الربيعي والذي يعتمد على ترتيب القيم كما هو معمول به في عملية حساب الوسيط. في المقابل عندما لا تكون مشكلة القيم المتطرفة حاضرة فانه يمكن استخدام الانحراف المتوسط أو الانحراف المعياري كمقياس للتشتت، حيث يتم استخدام كافة القيم في عملية حساب المقاييس السابقة.

يتم في الواقع حساب مقاييس التشتت للبيانات الكمية بصيغتيها الخام والمبوبة، ولكن الطريقة المتبعة في عمليات الحساب تختلف باختلاف طبيعة البيانات المدروسة. وبالطبع، كما تم الإشارة إليه سابقا، تمثل التقديرات المحصلة من البيانات الخام معلومة أدق وأكثر صحة من المعلومة المحصلة من البيانات المبوبة، لذا فانه يجب الاعتماد على البيانات الخام في حال توفرها في عملية حساب مقاييس التشتت، كما يجب قصر الاعتماد على البيانات المبوبة ليتم فقط في حالة عدم توفر الصيغة الخام للبيانات حيث أنها تعطي قيم تقريبية لا ترقى إلى دقة التقديرات المحصلة من خلال استخدام البيانات الغير مبوبة.

عند التعامل مع بيانات خام فانه يمكن أن يتم تغطية مجتمع الدراسة إذا كان المجتمع محدود الحجم وكانت الدراسة موجهة للتعامل مع المجتمع كاملا مع ما يصاحب تلك العملية من جهد ووقت وتكاليف مادية عالية. وإذا كان مجتمع الدراسة غير محدود أو كانت الدراسة لا تستطيع تحمل الوقت أو الجهد أو التكلفة العالية المطلوبة لتغطية مجتمع الدراسة، فانه يمكن الاعتماد على عينة عشوائية مسحوبة من المجتمع وممثلة له للحصول على تقديرات لمقاييس التشتت المطلوبة. تختلف طريقة حساب التباين والانحراف المعياري جوهريا عند التعامل مع مجتمع كامل عنها عند التعامل مع عينة عشوائية. لذلك فسيتم إيضاح الفرق في طريقة التقدير عند التطرق إلى كل من الانحراف المعياري والتباين لاحقا. تجدر الإشارة إلى أن جميع مقاييس التشتت هي قيم موجبة، وذلك شرط أساسي يجب توفره في جميع مقاييس التشتت. لذا فان مقاييس التشتت لا تأخذ قيم سالبة أبدا بل تكون قيمها موجبة دوما أو مساوية للصفر فقط وذلك إذا كانت جميع قيم المتغير الكمي محل الدراسة متساوية، أي انه لا يوجد تباين أو تشتت أصلا.

الإنحراف المعياري: Standard Deviation

يعتبر الانحراف المعياري والتباين من أهم مقاييس التشتت الإحصائية. ويرتبط المقياسين بعلاقة رياضية قوية، حيث يمكن دوما الحصول على المقياس الآخر في حال معرفة قيمة احدهما. يرمز للتباين بالرمز σ^2

في حال الحصول على قيمته من خلال تغطية مجتمع الدراسة، بينما يتم استخدام الرمز s^2 للدلالة على مقدر التباين المحصل من خلال بيانات عينة عشوائية مسحوبة من مجتمع الدراسة. وبأخذ الجذر التربيعي للتباين يتم الحصول على قيمة الانحراف المعياري وذلك في الحالتين، حالة المجتمع وحالة العينة

نعرف الانحراف المعياري لعينة حجمها n مسحوبة من مجتمع ما بأنه الجذر التربيعي لتباين هذه البيانات وبالتالي فإن الانحراف المعياري للبيانات $x_1 , x_2 , ... , x_n$ والتي وسطها الحسابي $\bar{x}$ هو :

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

أما في حال القيم المبوبة في جدول توزيع تكراري ذو k فئة فإن الإنحراف المعياري s يعطى بالشكل:

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

وبشكل عام فإن الانحراف المعياري هو الجذر التربيعي للتباين إن كان للعينة أو المجتمع . وعلى سبيل المثال فالإنحراف المعياري للبيانات المعطاة في المثال السابق 5,8,2,4,7,4, هو $s = \sqrt{4.8} = 1.67$ ويجب أن نذكر هنا أن الانحراف المعياري للمجتمع يرمز له بالرمز σ وتقرأ (سكما ـ segma) أما الإنحراف المعياري للعينة فهو s .

ملاحظات هامة :

(1)-إن السبب في القسمة على (n-1) عوضاً عن n لأن هناك (n-1) إنحرافاً مستقلاً من الشكل $x_i - \bar{x}$. ولأن مجموع هذه الانحرافات يساوي الصفر دوماً فإن أيا منها يساوي مجموع كل البقية بإشارة سالبة و أي منها يعطى بدلالة مجموع القيم الأخرى وبإشارة معاكسة . ولتوضيح هذه الفكرة تصور أن لدينا ثلاث بيانات x_1 , x_2 , x_3 ولدينا $\bar{x} = \frac{1}{3}(x_1 + x_2 + x_3)$.

نعلم أن $(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) = 0$

وبالتالي يمكن التعبير عن أي منهم وليكن الأول بـ $x_1 - \bar{x} = -[(x_2 - \bar{x}) + (x_3 - \bar{x})]$

أو $x_2 - \bar{x} = -[(x_1 - \bar{x}) + (x_3 - \bar{x})]$

أو $x_3 - \bar{x} = -[(x_1 - \bar{x}) + (x_2 - \bar{x})]$

أي يمكن تمثيل أي انحراف بدلالة الاثنين الآخرين وحيث لدينا n = 3 فسنقسم على 2=1-3 .

(2) -عندما تكون البيانات كبيرة وغالباً ما تكون كذلك, يمكن استخدام علاقة بديلة عن علاقتي التباين , وتسمى العلاقتان البديلتان بالعلاقتين الحسابيتين حيث يمكن حساب التباين (التشتت) منهما بسهولة .

$$s^2 = \frac{1}{n-1}[\sum_{i-1}^{n} x_i{}^2 - n\bar{x}^2] = \frac{n\sum_{i=1}^{n} x_i{}^2 - (\sum_{i=1}^{n} x_i)^2}{n(n-1)}$$

$$s^2 = \frac{1}{n-1}[\ \sum_{1}^{k}\ x_i^2\, f_i - n\bar{x}^2] = \frac{n\sum_{i=1}^{k} x_i^2\, f_i - (\sum_{i=1}^{k} x_i\, f_i)^2}{n(n-1)}$$

وبالطبع العلاقة الأولى للقيم المفردة أما العلاقة الثانية فهي للبيانات المبوبة في جداول توزيع تكرارية ذو k فئة .

يقيس الانحراف المعياري والتباين كمية التباعد الحاصلة في مجموعة بيانات وهذا التباعد يعتمد في الدرجة الأولى على وحدة القياس .

فلمقارنة التباين في عدة مجموعات من البيانات غالباً مايستخدم التباين النسبي relative variation لهذا الغرض أو معامل التباين coefficient of variation الذي يعطي الإنحراف المعياري كنسبة مئوية للمتوسط أي :

$$v = \frac{s}{\bar{x}} 100\ \%$$

حيث $\bar{x}$، s , هما المتوسط و الإنحراف المعياري على الترتيب لمجموعة من البيانات المراد دراستها .

مثال :

حُسب متوسط قياسات أطول مجموعة من المسامير أنتجت في مصنع معين باستخدام قياس معين فكان هذا المتوسط 3. 92mm و بإنحراف معياري 0.015mm .

وحسبت أطول مجموعة أخرى من البراغي أنتجت في مصنع آخر وذلك باستخدام مقياس مختلف فكان المتوسط والإنحراف المعياري لهذه المجموعة 0.008cm,1.54cm والمطلوب:

أي من المقياسين أكثر دقة نسبياً ؟ .

الحل :

بالنسبة للمقياس الأول $v = (\ 0.015\ /\ 3.92)\ .100\% =\ 0.38\ \%$

بالنسبة للمقياس الأخر $v =\ (\ 0.008\ /\ 1.54\ )\ .100\% =\ 0.52\ \%$

إذن القياسات التي أجريت بالنسبة للمقياس الأول أكثر دقة نسبياً .

<u>المدى : Range</u>

المدى هو الفرق بين أكبر قيمة وأصغر قيمة في مجموعة ما، فمدى المجموعة الأولى 10 = 5 - 15 .
بينما مدى المجموعة الثانية 18 = 1-19 لاحظ أن المجموعة الثانية أكثر تشتتاً من المجموعة الأولى .أما المدى لقيم معطاة في جدول توزيع تكراري فيحسب من الفرق بين الحد الأعلى للفئة العليا والحد الأدنى للفئة الدنيا . ففي الجدول (3-5) المدى هو 70 = 5 - 75 .

7-3 الانحراف المتوسط: Mean deviation

يعتبر الانحراف عن المتوسط أو (الانحراف المتوسط) أحد مقاييس التشتت .

ويعرف الانحراف المتوسط بأنه متوسط الفروق للبيانات عن وسطها الحسابي بقيمها المطلقة فإذا كانت لدينا مجموعة البيانات

$$x_1, x_2, ..., x_n$$

وسطها الحسابي $\bar{x}$ فإنه يمكن حساب الانحراف المتوسط بالعلاقة التالية :

$$M_D = \frac{1}{n} \sum_{i=1}^{n} \left| x_i - \bar{x} \right|$$

أما من أجل قيم معطاة في جدول توزيع تكراري فإن الإنحراف المتوسط يعطى بالعلاقة التالية:

$$M_D = \frac{1}{n} \sum_{i=1}^{k} \left| x_i - \bar{x} \right| f_i$$

حيث :

f_i تكرار الفئة ، و $n=\sum_{i=1}^{k} f_i$

لنوجد على سبيل المثال الانحراف المتوسط لمجموعة البيانات التالية :

5 , 6 , 8 , 10 , 12 , 14 , 15

إن المتوسط لهذه القيم $\bar{x}=10$ فيكون الانحراف المتوسط

$$M_D=\frac{1}{7}[|5-10|+|6-10|+|8-10|+|10-10|+|12-10|+|14-10|+|15-10|]=22/7$$

وبنفس الطريقة نجد أن الانحراف المتوسط للمجموعة الأخرى

1,2,5,10,15,18,19 هو 44/7

لاحظ من هذين المثالين أنه كلما كان الانحراف المتوسط كبيراً كلما كان التباعد بين القيم كبيراً وكلما كان صغيراً كانت القيم متقاربة .

8-3 التباين (التشتت): Variance

يعرف التباين بأنه الوسط الحسابي لمربعات فروقات البيانات عن وسطها الحسابي . ففي مجتمع ما إذا كان هذا المجتمع يتألف من n عنصر وكان وسطه الحسابي معطى وهو يساوي $\bar{x}$ فإن التباين (التشتت) σ^2 (ويقرأ σ^2 سكما مربع) للمجتمع يعطى بالشكل التالي :

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

إذا كانت القيم معطاة بشكل مفرد (غير مجدول) . أما إذا كانت القيم مبوّبة ووسطها الحسابي $\bar{x}$ أي معطاة في جدول توزيع تكراري ذو k فئة ولدينا $\bar{x}$ معلومة فإن σ^2 يعطى بالشكل :

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

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

أما تباين العينة فإنه يعطى وبصورة مشابهة تماماً للطريقة أعلاه بالشكل التالي :

إذا كانت لدينا عينة حجمها n مسحوبة من مجتمع ما بحيث أن متوسط العينة $\bar{x}$ معطى فإن تباين العينة الذي يرمز له بالرمز s^2 يعطى بالشكل :

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

إذا كانت القيم x_i معطاة بشكل غير مجدول (مفرد) .

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

حيث x_i ، f_i هما مركز الفئة وتكرارها على الترتيب .

مثال :

أوجد تباين العينة الممثلة بالبيانات 5,8,4,7,4,2

الحل : إن الوسط الحسابي لهذه البيانات هو $\bar{x}=5$ ويكون التباين :

$$s^2=\frac{1}{5}[(5-5)^2+(8-5)^2+(4-5)^2+(7-5)^2+(4-5)^2+(2-5)^2]=\frac{24}{5}=4.8$$

معامل الاختلاف(C.V):

في الإحصاء، معامل الاختلاف يشير عادة إلى قياس تشتت البيانات أو التباين في مجموعة من القيم. يتم استخدام معاملات الاختلاف لتقدير مدى انتشار البيانات حول قيمة مركزية مثل المتوسط.

هناك عدة معاملات الاختلاف المستخدمة في الإحصاء، ومن بينها:

1-الانحراف المعياري :(Standard Deviation) يقيس درجة التشتت أو التباين العام للبيانات حول المتوسط. إنه يحسب من خلال حساب جذر التباين المربع للمتغير.

2-الانحراف المتوسط المطلق :(Mean Absolute Deviation) يقيس متوسط قيم الاختلاف بين كل قيمة في المجموعة والمتوسط. يتم ذلك عن طريق حساب متوسط القيم المطلقة للفروق.

3-الانحراف المتوسط المربع :(Mean Squared Deviation) يقيس متوسط مربعات الاختلاف بين كل قيمة في المجموعة والمتوسط. يتم ذلك عن طريق حساب متوسط القيم المربعة للفروق.

4-نسبة الانحراف المعياري :(Coefficient of Variation) يستخدم لقياس التباين النسبي بين المتوسط والانحراف المعياري. يتم حسابه عن طريق قسمة الانحراف المعياري على المتوسط وضرب النتيجة في 100 للحصول على النسبة المئوية.

تستخدم هذه المعاملات في تحليل البيانات الإحصائية لفهم وتقييم المدى الذي يتراوح فيه البيانات وتفاوتها عن قيمة مرجعية معينة.، ويتم حساب معامل الاختلاف من خلال قيم كل من المتوسط الحسابي والانحراف المعياري كما يلي:

$$CV_p = \frac{\sigma_X}{\mu_X} \times 100$$

حيث يمثل القيمة الحقيقة لمعامل الاختلاف المبني على قيم معالم المجتمع. أما إذا لم تتوفر قيم معالم المجتمع وتم الحصول على تقديرات لها بواسطة قيم عينة عشوائية ممثلة لمجتمع الدراسة، فانه يمكن تقدير قيمة معامل الاختلاف من خلال العلاقة،

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

مع الإشارة إلى أن معامل الاختلاف يكون بدون تمييز (وحدة قياس) حيث انه نسبة مئوية.

الفصل الثالث

نظرية الاحتمال

نظرية الاحتمال هي فرع من الرياضيات يهتم بدراسة الظواهر والأحداث التي تحدث بشكل عشوائي أو غير مؤكد. تركز هذه النظرية على تحليل وقياس الاحتمالات وتطبيقها في توقع النتائج المحتملة للأحداث.

تعتمد نظرية الاحتمال على مفهوم الفضاء العينة (Sample Space) والأحداث .(Events) الفضاء العينة يمثل مجموعة جميع النتائج المحتملة للظاهرة المدروسة، في حين تعتبر الأحداث مجموعات من النتائج الممكنة.

تستخدم نظرية الاحتمال مجموعة من المفاهيم والأدوات لتحليل الاحتمالات، ومن بينها:

1-الاحتمال: يمثل قياسًا لتقدير مدى حدوث حدث معين. يتراوح قيم الاحتمال بين 0 و1، حيث يكون 0 يشير إلى عدم حدوث الحدث و1 يشير إلى حدوثه بالتأكيد.

2-القواعد الأساسية للمجموعة: توفر قواعد لحساب الاحتمالات باستخدام العمليات الرياضية مثل الجمع والضرب والطرح.

3-الأحداث المستقلة: تحدث عندما لا يكون حدث واحد مرتبطًا بحدث آخر، بمعنى أن حدوث أحد الأحداث لا يؤثر على احتمالية حدوث الآخر.

4-التوزيعات الاحتمالية: تصف توزيع الاحتمال لمجموعة محددة من النتائج الممكنة، وتستخدم لتحديد الاحتمالات المتوقعة للأحداث في سياق معين.

نظرية الاحتمال لها تطبيقات واسعة في مجالات مختلفة مثل الإحصاءات، والعلوم الطبيعية، والهندسة، والاقتصاد، والمالية، والعلوم الاجتماعية، حيث يمكن استخدامها لتحليل البيانات واتخاذ القرارات الاستراتيجية بناءً على المعرفة المتاحة حول الاحتمالات المحتملة...

التجربة العشوائية :(RANDOM SAMPLING)

التجربة العشوائية هي تجربة يمكن أن تنتج نتائج مختلفة عند تكرارها في ظروف مماثلة. في هذه التجربة، لا يمكن التنبؤ بالنتيجة النهائية بدقة مطلقة، وذلك بسبب وجود عناصر العشوائية أو غير المؤكدة في العملية.

تعتبر القمة المثالية للتجربة العشوائية هي رمي النرد. عند رمي النرد، يكون لديك عدة نتائج محتملة، وهي الأعداد الموجودة على وجوه النرد (مثل 1، 2، 3، 4، 5، 6). لا يمكن التنبؤ بالنتيجة المحددة للرمي قبل حدوثه، ويعتمد النتيجة على العوامل العشوائية مثل طريقة الرمي وزاوية سقوط النرد والسطح الذي يستقر عليه النرد.

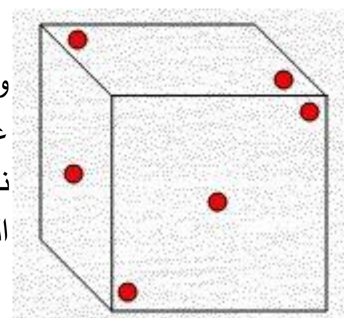
تشمل أمثلة أخرى على التجارب العشوائية تقدير الاحتمالات في القمار، رمي العملة، سحب البطاقات من العبوة في اللعبة، وقياس الوقت الذي يستغرقه حدث معين للحدوث في ظروف معينة.

تستخدم نظرية الاحتمال لتحليل وفهم التجارب العشوائية، وتقدير الاحتمالات المحتملة للنتائج المختلفة. يتم استخدام النماذج الرياضية والإحصائية لوصف وتحليل الظواهر العشوائية وتوقع النتائج المحتملة<u>فضاء النواتج :(Sample Space)</u>

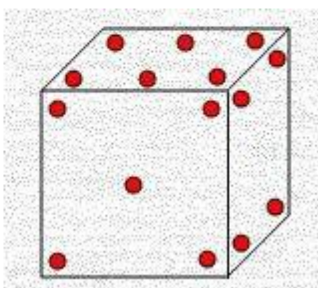
تعرف المجموعة {1، 2، 3، 4، 5، 6} في مثالنا السابق للتجربة العشوائية بفضاء النواتج أو قضاء الإمكانيات أو فضاء العينة(Sample Space)

فضاء العينة لتجربة إلقاء قطعة نقود مرة واحدة T } ، {H أو تمثل بشكل فن مستطيل أو دائرة بالداخل العناصر الخاصة بالتجربة العشوائية

الأحداث: Events



الحدث هو مجموعة جزئية من فضاء العينة وعدد الأحداث تخضع للصيغة $2^{ن}$ حيث ن عدد عناصر فضاء العينة واحتمال وقوع الحدث A هو نسبة عدد حالات وقوعه بالفعل بالنسبة لكل الحالات الممكنة لوقوعه أي أن : P(A) = N ÷ M حيث M عدد حالات وقوع A بالفعل



، N عدد الحالات الممكنة فاحتمال ظهور عدد فردي عند إلقاء حجر النرد مرة واحدة هو 0.5 لأن الأعداد الفردية ثلاثة (1، 3، 5) والتي تحقق المطلوب (عدد فردي) وكل الأعداد ستة (1، 2، 3، 4، 5، 6) فالاحتمال 3 ÷ 6 = 0.5 ، الشكل المقابل لحجر النرد أو الزار أو الزهرة

الحدث البسيط :(Simple event) وهو الحدث المكون من عنصر واحد مثل {1 {في تجربة إلقاء حجر النرد.

الحدث المركب :(Compound event) الحدث المكون من أكثر من عنصر مثل {2، 4، 6} حدث العدد زوجي في تجربة إلقاء حجر النرد.

الحدث المستحيل :الحدث الذي لا يحوي أي عنصر كحدث ظهور العدد 7 في تجربة إلقاء حجر النرد.

الحدث المؤكد :الحدث الذي يضم كافة عناصر الفضاء كحدث ظهور عدد أقل من 7 في تجربة إلقاء حجر النرد.

الحدثان المتنافيان :(Mutually Exclusive events) الحدثان اللذان لا يشتركا في أي عنصر وتقاطعهم المجموعة الخالية أي A ∩ B = f مثل {2}، {3}، وتعرف بالأحداث غير المتصلة.

الأحداث المنتظمة :(dependent events) المتساوية في احتمالاتها. ففي تجربة إلقاء حجر النرد مرة واحدة يكون1:6 = P(1) = P(2) = P(3) =P(4) = P(5) = P(6) :

الأحداث الشاملة :(Exhaustive events) إذا كان S فضاء عينة ما فإن الأحداث A, B, C شاملة إذا تحقق الشروط الثلاثة الآتية:

(1) متنافية فيما بينها أي A ∩ B = f :و A ∩ C = f وC ∩ B = f

(2) أياً منها ليست خالية أي A ≠ f و B ≠ f وC ≠ f

(3) إتحادها يساوي S أيA ᴜ B ᴜ C = S

الأحداث المكملة:(Complementary events)

_الحدثان اللذان اتحادهم يساوي فضاء العينة بمعنىA حدث فإن\`A الحدث المكمل حيثA ᴜ\`A = S

الحدثان المستقلان :(Independent events) اللذان لا يتأثر أي منهم بالآخر (وقع أحدهم لا يؤثر أو يتأثر بوقوع أو عدم وقوع الآخر.

P(A ∩ B) = P(B) × P(A)قاعدة الضرب للاحتمالات للإحداث المستقلة
يمكن تعميم هذه القاعدة لأكثر من حدثين
P(A ∩ B ∩ C ∩ ... ∩ Z) = P(A) × P(B) × P(C)×... × P(Z)

حدثان وقوع أحدهما يؤثر في وقوع الآخر مثل سحب ورقة من أوراق اللعب دون إرجاع مما يؤدي لتأثير سحب ورقة جديدة لنقص الفرصة بنقص عدد الأوراق (من 52 إلى 51(
فالحدثان A, B نكتب حدث وقوع A بشرط وقوع B بالصورة A / B ويكون:

$$p(B) = \frac{p(A \cap B)}{P(A/B)}$$

OR

P(A ∩ B) = P(B) × P(A / B)

لاحظ أن العلامة (/) ليست علامة القسمة بل علامة شرط وقوع ما يليها من أحداث
sP(A / B)وهو احتمال وقوع الحدث A بشرط وقوع الحدث B ، قد ترد عبارة أخرى تفيد الشرط كالقول علماً بأن... ,

وفي حالة الحدثان مستقلان أي لا يؤثر وقوع أحدهما على الآخر (when A and B are independent events) يصبح القانون:

$P(A \cap B) = P(B) \times P(A)$

<u>*مثال:*</u>

لدينا صندوق يحتوي على 14 كرة، منها 8 حمراء و6 زرقاء. نقوم بسحب كرتين من الصندوق ونريد حساب احتمالية أن تكون الكرتين الأولى زرقاء والثانية حمراء.

لنقم بحساب الاحتمالية. الكرتان يمكن أن تكون مرتبتين بـ 14 × 13 طريقة ممكنة، حيث يتم سحب الكرتين بدون إرجاع.

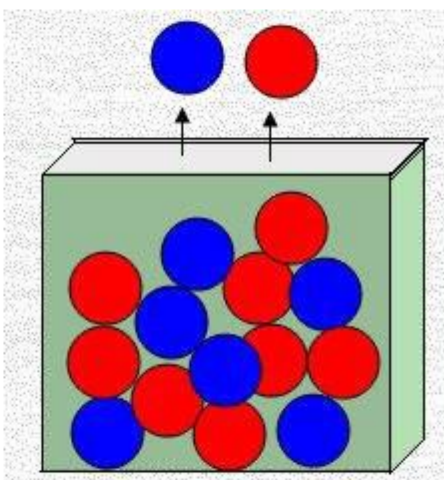
الآن دعونا نحسب عدد الطرق التي يمكن أن تكون فيها الكرة الأولى زرقاء والثانية حمراء. يمكننا سحب كرة زرقاء في السحب الأول بـ 6 طرق ممكنة، وبعد ذلك يمكننا سحب كرة حمراء في السحب الثاني بـ 8 طرق ممكنة (لأن لدينا 8 كرات حمراء متبقية). لذلك، عدد الطرق التي يمكن أن تكون فيها الكرتان الأولى زرقاء والثانية حمراء هو 6 × 8 = 48.

بالتالي، الاحتمالية المطلوبة هي:

احتمالية (الكرتتان الأولى زرقاء والثانية حمراء) = (عدد الطرق التي يمكن أن تكون فيها الكرتتان الأولى زرقاء والثانية حمراء) / (عدد الطرق الممكنة لاختيار الكرتتين)

$= 48 / (14 \times 13)$

≈ 0.276



لذلك، احتمالية أن تكون الكرتتان الأولى زرقاء والثانية حمراء هي حوالي 0.276 أو 27.6%

الحل:

ليكن A = حدث سحب كرة حمراء اللون

وليكن B = حدث سحب كرة زرقاء اللون

فالمطلوب هو sP(A / B)حيث A السحبة الثانية ، B السحبة الأولى.

$P(A \cap B) = P(B) \times P(A / B)$

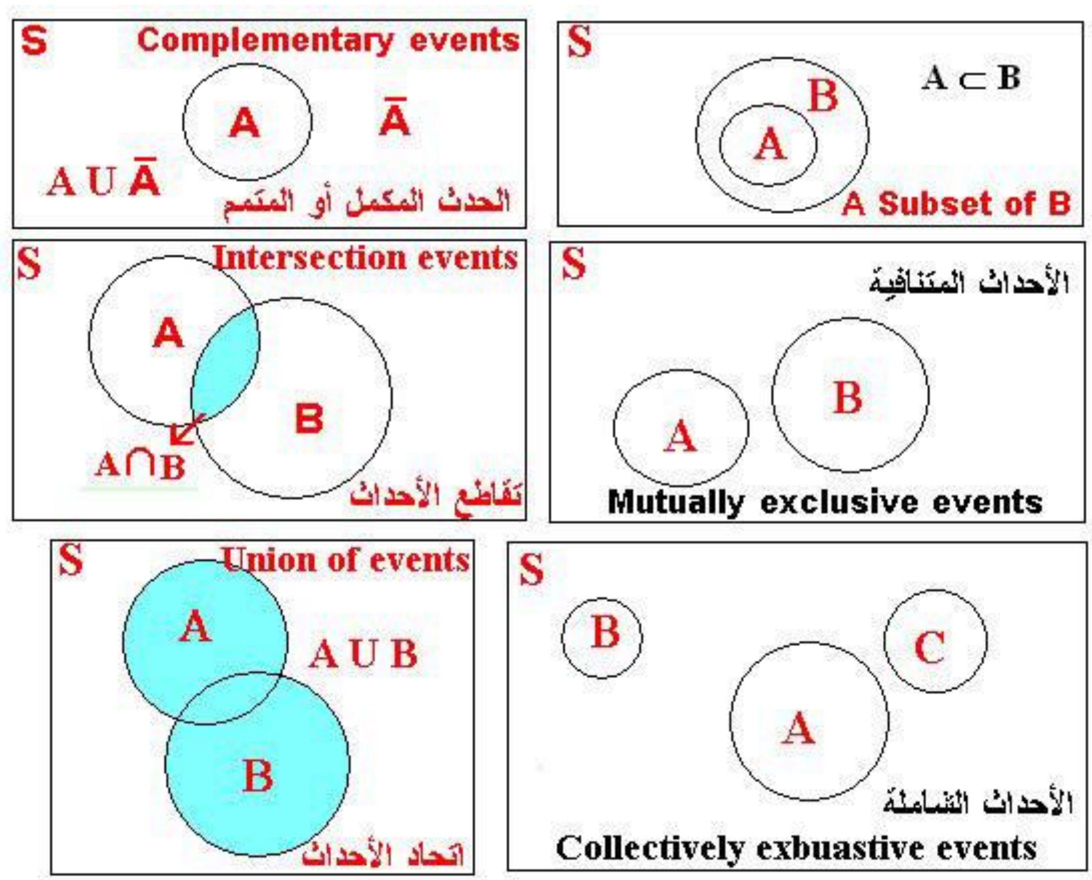
$$p(A \cap B) = \frac{8}{14} \times \frac{6}{13} = \frac{24}{91} = 0.26$$

لاحظ سحب كرتان نفس اللون = = (8÷14)×(7÷13) + (6÷14)×(5÷13) = 0.4725

لاحظ سحب كرتان مختلفتان في اللون = = 0.2637 + 0.2637 = 0.5274

لاحظ مجموع الاحتمالان السابقان 0.4725 + 0.5274 = 0.9999 =1

الأشكال التالية أشكال(فن) تبين ما سبق من أحداث بصورة مبسطة:



<u>قواعد الاحتمال</u>

1) إذا كان A حدث من S أي أنَّ A مجموعة جزئية من S فإن:

A يعبر عن احتمال وقوع الحدث P(A) الرمز

$$P(A) = \frac{M}{N} = \frac{\text{Number of events classifiable as A}}{\text{Total number of possible events}}$$

$$P(A) = \frac{M}{N} = \frac{\text{عدد حالات وقوع الحدث A بالفعل}}{\text{كل الحالات التي يمكن وقوعها}}$$

y.

$0 < P(A) < 1$, $P(S) = 1$,

2) لحدثان المتكاملان (المتتامان) S?A?$A \cup$ يكون. $P(A) + P(\bar{A})=1$

3) مجموع احتمالات الأحداث الشاملة يساوي الواحد الصحيح لأن اتحادها يساويS

4) الحدثان المتنافيان A, B أي تقاطعهم فإن:

$$P(A \cup B) = P(A) + P(B) \quad , \quad P(A \cap B) = 0$$

5) إذا كان A, B حدثان غير متنافيين (متصلين) أو احتمال وقوع أحدهم على الأقل فإن:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

عملية الطرح هنا للاحتمال $P(A \cap B)$sلتكراره مرتين عند حساب الاحتمال للجزء المشترك بين A, Bحيث يحسب مرة مع A وأخرى معB

يمكن تعميم القاعدة السابقة لأكثر من حدثين متصلين كالتالي:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C)$$

6) عدد الأحداث في فضاء النواتج (S)للتجربة العشوائية هو$s2^n$ حيث n عدد عناصر الفضاء (S) فعدد أحداث تجربة إلقاء حجر النرد مرة واحدة هو $(64 = 2^6)$ حدثاً بما فيهم الحدثان المستحيل ϕ والمؤكد $S = \{1, 2, 3, 4, 5, 6\}$

أمثلـــة

(1)في تجربة إلقاء قطعة نقود وحجر النرد ولمرة واحدة أكتب فضاء النواتج.S

الحل :قطعة النقود لها عنصران H, T صورة وكتابة ، وحجر النرد له 6 عناصر هي العداد من 1 إلى 6 وعليه يكون عدد عناصر فضاء التجربة $= 3 \times 2 = 12$ هي:

$S = \{(H,1), (H, 2), (H, 3), (H, 4), (H, 5), (H, 6), T,1), (T, 2), (T, 3), (T, 4), (T, 5), (T, 6) \}$

ويمكن كتابتها اختصاراً بالصورة:

$S = \{(H,1), (H, 2), \ldots , (T, 6) \}$

(2)سحبت كرة واحدة فقط من كيس يحوي 10 كرات متماثلة تماماً ألوانها 3 حمراء ، 2 سوداء ، 5 صفراء فما احتمال أن تكون الكرة المسحوبة حمراء

الحل: عدد الكرات التي تحقق المطلوب (حمراء اللون) هو 3 وعدد الكرات التي يمكن أن تسحب يساوي 10 وبافتراض أن A هو حدث الكرة حمراء فيكون المطلوب:

$$P(A) = M \div N = 3 \div 10 = 0.3$$

(3)إذا كان احتمال وفاة شخص هو 0.05 فما احتمال أن يعيش؟

الحل: واضح أن الاحتمال المطلوب هو الحدث المتمم للاحتمال المعطى أي أن مجموعهم يساوي الواحد الصحيح وبفرض أن:

A : حدث أن يعيش الرجل

$= 1 - 0.05 = 0.95$

(4)بين إن كانت الأحداث الآتية شاملة (دالة احتمال) حيث احتمالاتها 0.1، 0.3، 0.6 مع العلم بأنها متنافية فيما بينها

الحل: حتى تكون شاملة يجب أن يكون مجموعها يساوي الواحد الصحيح وبجمعها نجد أن: $0.1 + 0.3 + 0.6 = 1$ فالأحداث شاملة.

(5)بين إن كانت الأحداث الأربع الآتية شاملة (دالة احتمال) حيث احتمالاتها 0.0، 0.1، 0.3، 0.6

الحل: حتى تكون شاملة يجب أن لا يكون أياً منها لا يساوي ? ولكن وجود الاحتمال المساوي للصفر يعني الحدث = فالأحداث غير شاملة.

(6)إذا كان احتمال النجاح في مادة الرياضيات هو i0.45 واحتمال النجاح في مادة الإحصاء هو i0.65 واحتمال النجاح في المادتين معاً هو i0.37 أوجد احتمال النجاح في أحد المادتين على الأقل.

الحل: بتطبيق صيغة الاحتمالات للحوادث المتصلة بفرض أنَّ:

A : احتمال النجاح في مادة الرياضيات

B : احتمال النجاح في مادة الإحصاء

$A \cap B$: احتمال النجاح في المادتين معاً

فأنَّ:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= 0.45 + 0.65 - 0.37$$
$$= 0.73$$

الفصل الرابع

اختبار الفرضيات

عملية اختبار الفرضيات تشتمل على افتراض أو ادعاء يتم اختباره. في البداية، يتم افتراض عدم صحة الادعاء، ومن ثم يتم استخدام بيانات الدراسة لإثبات العكس، أي إثبات صحة الادعاء. يعزز هذا النهج قوة اختبار الفرضية من خلال التلافي عن التحيز وعدم الدقة. فعندما تكون جودة الدراسة وجمع البيانات ضعيفة، فإن النتائج المتحققة تصب في صالح الادعاء المعاكس. وبالتالي، لا يمكن قبول الادعاء إلا إذا كانت هناك دلائل إحصائية قوية تدعمه.
هذا النهج يشبه سياسة التحقيق في القضايا الجنائية، حيث يتم افتراض براءة المتهم حتى تثبت إدانته. وبالتالي، يتم تبني العكس وعدم قبول الادعاء بأن المتهم مذنب إلا إذا كانت هناك أدلة قوية تشير إلى ذلك. إذا كانت الأدلة ضعيفة أو أن المتهم يعتبر بريئًا، فإننا لا نقبل الادعاء ولا نرفض الافتراض بأن المتهم بريء من الأصل.
في عملية اختبار الفرضيات، تلعب العينة العشوائية والبيانات المستخلصة منها دور الأدلة المستخدمة لإثبات إدانة المتهم في القضايا الجنائية. وبالتالي، يمكن استخدام بيانات عينة عشوائية وتحليلها لإثبات صحة الادعاء أو عدمه. وبالطبع، يتم قبول الادعاء إذا كانت الأدلة المستندة إلى العينة العشوائية وبياناتها تشير بقوة إلى صحة الادعاء. أما إذا كانت البيانات ضعيفة أو إذا كان الادعاء غير صحيح من الأصل، فإننا لا نقبل الادعاء ولا نرفض الافتراض بأن الادعاء غير صحيح..
مفاهيم مهمة :

هناك بعض المفاهيم المتعلقة باختبارات الفروض لابد من معرفتها:

الفرض الاحصائي statistical hypothesis

هو عبارة عن إدعاء او تخمين معين حول معلمة من معالم المجتمع ويكون المطلوب اختبار صحة هذا الادعاء أو التخمين ...هناك نوعين من الفروض :
فرض العدم (null hypothesis) ويرمز له بالرمز H_0 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_0 : 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 .

مستوى المعنوية α ودرجة الثقة $(1 - \alpha)$:

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.
أخطاء اختبار الفرضيات وأنواعها
يرتبط الاستدلال الإحصائي بالتعامل مع المجاهيل، ومن ثم لا يمكن الجزم أبدا بان النتائج المحصلة صحيحة تماما. عندما توجد فرضية قائمة على ادعاء فان القرار النهائي يكون أما قبول الفرضية أو رفضها. وبحكم احتواء عملية اختبار الفرضيات على فرضيتان شاملتان ومتضادتان (فرضية العدم والفرضية البديلة) لذا فان القرار المتعلق بإحداها يمثل القرار المعاكس للفرضية الأخرى. فقبول فرضية العدم يعني رفض الفرضية البديلة، والعكس صحيح.

يوجد نوعان من الأخطاء التي يحتمل حدوثها في عملية اختبار الفرضيات. لتوضيح هذين النوعين نفترض أننا على علم بالوضع الحقيقي أو القيمة الحقيقية لمعلمة المجتمع الواقع عليها الاختبار، وأننا نرغب في إجراء الاختبار دون استخدام تلك المعلومة، أي على افتراض أننا نجهل قيمة المعلمة الحقيقية. في هذه الحالة نصبح أمام إحدى خيارين إما أن نفشل في رفض فرضية العدم (لكون اختبار الفرضيات قائم في الأصل على تبني
عند اختبار فرض العدم H_o ضد الفرض البديل H_1 نجد أننا امام احدى الحالات الاربع الاتية :

نتيجة الاختبار	الوضع الحقيقي	
	الادعاء غير صحيح H_o صحيحة	الادعاء صحيح رفض H_o وإقرار H_1
عدم رفض H_o	قرار صائب	قرار خاطئ خطاء من النوع الثاني
رفض H_o وقبول H_1	قرار خاطئ خطاء من النوع الأول	قرار صائب

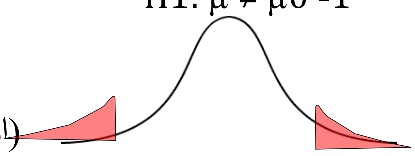
1) أن يكون فرض العدم صحيحا ويكون القرار بقبوله ....وهذا قرار سليم
2) أن يكون فرض العدم صحيحا ويكون القرار برفضه .... وهذا قرار خاطئ
(الخطأ من النوع الأول : رفض H_o عندما يكون H_o صحيحا ويرمز لحجم هذا الخطأ بالرمز α)
3) أن يكون فرض العدم خطأ ويكون القرار برفضه وهذا قرار سليم
4) أن يكون فرض العدم خطأ ويكون 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

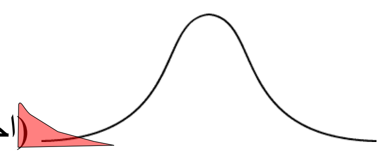
خطوات اختبار الفرض الإحصائي حول متوسط المجتمع لعينة كبيرة

لإجراء الاختبار الإحصائي فإننا نتبع الخطوات التالية:
صياغة فرض العدم Ho
$Ho : \mu = \mu0$
والفرض البديل هو احد الحالات التالية:
1- $H1: \mu \neq \mu0$



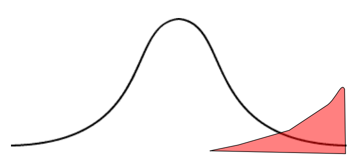
(اختبار من طرفين)

2- $H1: \mu > \mu0$



(اختبار من طرف واحد)

3- $H1: \mu < \mu0$



(اختبار من طرف واحد)

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

حيث أن هذا الإحصاء 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	$5\% = 0.05$	$95\% = 0.95$	$\pm 1.96 = Z_{\alpha/2}$
	$1\% = 0.01$	$99\% = 0.99$	$\pm 2.58 = Z_{\alpha/2}$

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

4- اتخاذ القرار:

نتخذ القرار بناءاً على قيمة إحصاءه الاختبار

نرفض H0 إذا وقعت قيمة إحصاءه الاختبار في منطقة الرفض

لا نرفض H0 إذا وقعت قيمة إحصاءه 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 ABC المتخصصة في صناعة القطن الطبي تعاقدت لشراء نوع جديد من الخيوط الصناعية. يدعي الصانع "XYZ" لهذه الخيوط أن متوسط قوة تحمل الخيط هو 15 كجم بانحراف معياري نصف كجم.

للتحقق من صحة ادعاء الشركةXYZ ، تم أخذ عينة عشوائية من 50 خيطًا وتم اختبارها. وتبين أن متوسط قوة التحمل في العينة هو 14.8 كجم. الآن نحتاج لتحديد ما إذا كان بإمكاننا تأييد ادعاء الشركة XYZ بمستوى معنوية 5%.

لحساب ذلك، يمكننا استخدام اختبار الفرضيات بافتراض أن الادعاء الذي أدلى به الصانع (أي أن متوسط قوة التحمل هو 15 كجم) صحيح، ومن ثم نقارنه بالبيانات التي حصلنا عليها من العينة.

نقوم بحساب الاختبار الإحصائي باستخدام الاختبار .t حيث يكون الفرض الأول (H_0) أن متوسط قوة التحمل في العينة هو 15 كجم، والفرض البديل (H_1) هو أن متوسط قوة التحمل في العينة أقل من 15 كجم.

نقوم بحساب الاختبار الإحصائي باستخدام الصيغة التالية:

t = (عينة المتوسط - الفرضية المفترضة) / (الانحراف المعياري / √عدد العينة)

حيث:

عينة المتوسط = 14.8 كجم

الفرضية المفترضة = 15 كجم

عدد العينة = 50

بعد حساب القيمة، يمكننا استخدام جدول توزيع الاختبار t لتحديد قيمة الاختبار النهائية. يعتمد القرار على المستوى المعنوي المحدد (5% في هذه الحالة).

إذا كانت قيمة الاختبار النهائية تقع في منطقة الرفض (أي القيم التي تكون أقل من القيمة المعيارية المحددة في الجدول)، فإننا نرفض الفرضية المفترضة ونتجه نحو قبول ادعاء الشركة.XYZ

إذا قمت بحساب القيمة وتحديد القرار، يرجى توفير القيمة النهائية لكي أتمكن من مساعدتك في تحديد ما إذا كان بإمكاننا تأييد ادعاء الشركة XYZ أم لا..

الحل:

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

1- صياغة الفرض الإحصائي:

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

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 :

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

5- 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) :

للتحقق مما إذا كان متوسط ما يتناوله الأشخاص البالغون ذوو الدخل المنخفض من الكالسيوم يقل عن 800 ميلليجرام، يمكننا استخدام اختبار الفرضيات باستخدام الاختبار t.

الفرضية المفترضة :(H_0) متوسط ما يتناوله الأشخاص البالغون ذوو الدخل المنخفض من الكالسيوم هو 800 ميلليجرام.
الفرضية البديلة :(H_1) متوسط ما يتناوله الأشخاص البالغون ذوو الدخل المنخفض من الكالسيوم أقل من 800 ميلليجرام.

نقوم بحساب الاختبار الإحصائي باستخدام الاختبار t. حيث يتم استخدام الصيغة التالية:

t = (عينة المتوسط - الفرضية المفترضة) / (الانحراف المعياري / √عدد العينة)

في هذه الحالة،
عينة المتوسط = 755.3 ميليجرام
الفرضية المفترضة = 800 ميليجرام
الانحراف المعياري = 239.3 ميليجرام
عدد العينة = 50

بعد حساب القيمة، يمكننا استخدام جدول توزيع الاختبار t لتحديد قيمة الاختبار النهائية. يعتمد القرار على المستوى المعنوي المحدد (5% في هذه الحالة).

إذا كانت قيمة الاختبار النهائية تقع في منطقة الرفض (أي القيم التي تكون أصغر من القيمة المعيارية المحددة في الجدول)، فإن النتائج تدل على أن متوسط ما يتناوله الأشخاص البالغون ذوو الدخل المنخفض من الكالسيوم يقل عن 800 ميليجرام.

إذا قمت بحساب القيمة وتحديد القرار، يرجى توفير القيمة النهائية لكي أتمكن من مساعدتك في تقييم النتائج وإعطاء الاستنتاج النهائي. 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) :

لتحديد ما إذا كان متوسط العمر في القرية أكبر من 65 عامًا، يمكننا استخدام اختبار الفرضيات باستخدام الاختبار t.
الفرضية المفترضة (H_0): متوسط العمر في القرية هو 65 عامًا.
الفرضية البديلة (H_1): متوسط العمر في القرية أكبر من 65 عامًا.
نقوم بحساب الاختبار الإحصائي باستخدام الاختبار t. حيث يتم استخدام الصيغة التالية:

t = (عينة المتوسط - الفرضية المفترضة) / (الانحراف المعياري / √عدد العينة)

في هذه الحالة،
عينة المتوسط = 67.5 عامًا
الفرضية المفترضة = 65 عامًا
الانحراف المعياري = 8 أعوام
عدد العينة = 100

بعد حساب القيمة، يمكننا استخدام جدول توزيع الاختبار t لتحديد قيمة الاختبار النهائية. يعتمد القرار على المستوى المعنوي المحدد (مثلاً 5% أو 1%).

إذا كانت قيمة الاختبار النهائية تقع في منطقة الرفض (أي القيم التي تكون أكبر من القيمة المعيارية المحددة في الجدول)، فإن النتائج تدل على أن متوسط العمر في القرية أكبر من 65 عامًا.

يرجى توفير القيمة النهائية لكي أتمكن من مساعدتك في تحديد النتائج وإعطاء الاستنتاج النهائي.5.
الحل:
نفرض أن μ متوسط العمر في هذه القرية .
1- صياغة الفرض الإحصائي:

H_0 : μ65 =

H_1: μ65 >

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 α=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 .

الفصل الخامس

التحليل الإحصائي هو عملية تجهيز وتحليل البيانات المرتبطة بالبحث العلمي بطرق رياضية ومنطقية. يهدف التحليل الإحصائي إلى استخلاص النتائج وفهم العلاقات والصفات التي تميز مجتمع ما. يتم ذلك من خلال دراسة عينة محدودة من المجتمع وتعميم النتائج على المجتمع الكلي.

تاريخيًا، يعود الأصل الأول للتحليل الإحصائي إلى العصور القديمة، حيث كان اليونانيون والفراعنة والرومان يعرفون هذا العلم. وظهر التحليل الإحصائي في أوروبا خلال القرون الوسطى، خلال فترة النظام الإقطاعي، حيث استخدمت الحكومات هذا النظام لإحصاء السكان وتحديد الممتلكات وفرض الضرائب. ومع مرور الوقت، تطور التحليل الإحصائي وزادت أهميته، وأصبح بإمكانه التنبؤ بتطور السكان والاستهلاك وغيرها من القضايا الاجتماعية والاقتصادية.

مع تطور التكنولوجيا وظهور الحواسيب، أصبح من الممكن إجراء التحليل الإحصائي بسهولة باستخدام برامج مخصصة. يمكن للأفراد ملء البيانات والاعتماد على الحاسوب لتنفيذ التحليل الإحصائي وعرض النتائج بشكل سريع ودقيق.

باختصار، يعد التحليل الإحصائي عملية مهمة في البحث العلمي وفي فهم العلاقات والصفات التي تميز المجتمعات. يساعدنا التحليل الإحصائي على استخلاص المعاني الجديدة واتخاذ القرارات الأكثر فهمًا وموضوعية بناءً على البيانات المتاحة
اهميه التحليل الاحصائي

ما هي أهمية التحليل الإحصائي؟

تحمل التحليل الإحصائي أهمية كبيرة في العديد من المجالات والمجتمعات، وذلك للأسباب التالي

1-صنع القرارات الأفضل: يساعد التحليل الإحصائي على جمع وتحليل البيانات الكمية والكيفية المتعلقة بمشكلة أو قضية معينة. وبناءً على النتائج والتحليلات، يمكن اتخاذ قرارات مستنيرة وموضوعية تعتمد على أدلة قوية. يساعد التحليل الإحصائي في تقييم الخيارات المتاحة وتحديد السيناريوهات المحتملة وتقدير النتائج المتوقعة لكل سيناريو.

2-فهم العلاقات والاتجاهات: يمكن للتحليل الإحصائي كشف العلاقات والترابطات بين المتغيرات المختلفة وتحليل الاتجاهات والأنماط الموجودة في البيانات. يمكن أن يساعد في كشف العوامل المؤثرة وتحليل تأثيرها على الظواهر المختلفة. هذا يمكن أن يوجه البحوث والدراسات لفهم الظواهر المعقدة وتفسيرها بطريقة أفضل.

3-التنبؤ والتخطيط: يمكن للتحليل الإحصائي أن يساعد في التنبؤ بالأحداث المستقبلية وتقدير المخاطر والفرص المحتملة. يمكن استخدام النماذج الإحصائية لتوقع سلوك الأنظمة والأحداث والمتغيرات المختلفة. يمكن أيضًا استخدامه للتخطيط واتخاذ القرارات الاستراتيجية في المؤسسات والمنظمات.

4-دعم الأبحاث العلمية: يعد التحليل الإحصائي أداة قوية في مجال البحث العلمي. يمكن استخدامه لاختبار الفروض والفرضيات وتحليل البيانات المتعلقة بالدراسات العلمية. يمكن أن يساعد في استخلاص النتائج وإثبات العلاقات وتوضيح النتائج بطريقة إحصائية قوية.

5-دعم اتخاذ القرارات في المؤسسات والأعمال التجارية: يمكن للتحليل الإحصائي أن يوفر رؤى قيمة للشركات والمؤسسات في مجالات مثل التسويق وإدارة الموارد البشرية وتحسين الجودة واتخاذ القرارات المالية. يمكن استخدامه لتحليل سلوك العملاء وتوجيه الحملات التسويقية وتقدير الوقت والموارد لدي معلومات حول شركة سيغا الحديثة وأحدث إصداراتها حتى تاريخ قطع المعلومات في سبتمبر 2021. يُرجى ملاحظة أن المعلومات التي تتعلق بالشركات والإصدارات الحديثة قد تتغير بسرعة، ويفضل البحث عبر المصادر الرسمية أو المواقع الإخبارية المعتمدة للحصول على أحدث المعلومات حول شركة سيغا وإصداراتها الحديثة

الاختبار الاحصائي

بعد اختيار الباحث لنوع الاختبار الإحصائي المناسب، يقوم باتباع مجموعة من الخطوات الأساسية لإجراء التحليل الإحصائي. هنا هي الخطوات الأساسية:

تجميع البيانات: يقوم الباحث بجمع البيانات المتعلقة بالمتغيرات التي يرغب في دراستها. يمكن أن تكون البيانات مستقاة من مصادر مختلفة مثل الاستبيانات أو السجلات أو التجارب.

التنظيف والتحقق من البيانات: يتم فحص البيانات المجمعة للتأكد من كمالها وصحتها. يتضمن ذلك التحقق من اكتمال البيانات، التعامل مع القيم المفقودة أو البيانات غير الصحيحة، وفحص القيم القصوى والقيم الشاذة.

وصف البيانات: يتم وصف البيانات المجمعة بواسطة مقاييس إحصائية مثل المتوسط والانحراف المعياري والتوزيعات الاحتمالية. يساعد هذا الوصف على فهم طبيعة المتغيرات وتحليل البيانات بشكل أولي.

اختبار الفرضيات: يتم وضع فرضيات إحصائية حول العلاقات المفترضة بين المتغيرات. يتم استخدام الاختبارات الإحصائية لتقييم مدى توافق البيانات المجمعة مع هذه الفرضيات. يتم قبول أو رفض الفرضيات بناءً على قيم الاختبارات الإحصائية.

تفسير النتائج: يتم تفسير النتائج الإحصائية وتحليلها بناءً على الفرضيات المقبولة. يتم التركيز على النتائج الإحصائية الهامة والمعنوية من الناحية الإحصائية. يتم توظيف المفاهيم النظرية والمعرفية لتفسير العلاقات المكتشفة بين المتغيرات.

التقديم والتوصيات: يتم كتابة تقرير أو ورقة بحثية تلخص النتائج والتفسيرات والاستنتاجات المستخلصة من التحليل الإحصائي. يمكن أن تشمل التقارير التوصيات للممارسة أو البحوث المستقبلية المحتملة.

يجب أن يتم تنفيذ هذه الخطوات بعناية ودقة لضمان الحصول على نتائج صحيحة وموثوقة من التحليل الإحصائي. قد تتطلب الخطوات المحددة والتفاصيل الدقيقة للتحليل الإحصائي تخصصًا محددا بعض الإرشادات العامة حول خطوات التحليل الإحصائي، ومن المهم أن يتم استشارة كتب أو مراجع إحصائية متخصصة للحصول على مزيد من التفاصيل والإرشادات المحددة وفقًا للطريقة الإحصائية المحددة ونوع البيانات والفرضيات المراد اختبارها
.الخطوات:

:إليك خطوات أساسية لإجراء اختبار إحصائي

1-تحديد الفرضيات: يتم تحديد الفرضية العامة والفرضيات البديلة المتعلقة بالعلاقة المراد اختبارها بين المتغيرات. تُصيغ الفرضية العامة كفرضية صفر (H0) وتُصيغ الفرضيات البديلة كفرضيات بديلة.(H1)

2-تحديد مستوى الأهمية: يتم تحديد مستوى الأهمية (مستوى الاعتبار) الذي يستخدم لتحديد متى يتم رفض الفرضية الصفر. يشير مستوى الأهمية إلى الاحتمالية المقبولة لرفض الفرضية الصفر عن طريق الخطأ، ويعبر عادة عنه بالقيمة الباعثة للألفة(α) ، مثل 0.05 أو 0.01.

3-اختيار الاختبار الإحصائي: يتم اختيار الاختبار الإحصائي المناسب بناءً على النوع والطبيعة وتوزيع البيانات المتاحة والفرضيات المراد اختبارها. قد يتضمن ذلك اختبار تقارب المتوسطات) مثل اختبار (t أو اختبار الفروق بين النسب المئوية) مثل اختبار الـ (chi-square أو اختبار الارتباط بين المتغيرات (مثل اختبار الارتباط الخطي).

4-جمع البيانات والتحليل: يتم جمع البيانات المناسبة وإدخالها في البرنامج الإحصائي المناسب. يتم تنفيذ الاختبار الإحصائي المحدد وحساب القيمة الإحصائية الناتجة.

5-اتخاذ القرار: يتم استخدام القيمة الإحصائية المحسوبة لاتخاذ القرار بشأن قبول أو رفض الفرضية الصفر. إذا كانت القيمة الإحصائية أقل من قيمة الحد النقطي (الحد الحرج) المحددة مسبقًا، يتم رفض الفرضية الصفر واعتماد الفرضية البديلة.

6-تفسير النتائج: يتم تفسير النتائج الإحصائية بشكل منطقي ومعنوي. يجب أن يتم توصيل النتائج بالفرضيات المراد اختبارها وتفسير العلاقة المكتشفة بين المتغيرات بناءً على النتائج.

7-التقرير: يتم كتابة تقرير يلخص النتائج والتفسيرات والاستنتاجات المستخلصة من التحليل الإحصائي. يمكن أن يشمل التقرير تفاصيل البيانات والطريقة المستخخطوات الاختبار الإحصائي تشمل ما يلي:

8-تحديد الفرضيات: تبدأ العملية بتحديد الفرضية العامة (الفرضية الصفر) والفرضية البديلة. تصف الفرضية الصفر عدم وجود علاقة أو تأثير بين المتغيرات المدروسة، في حين تصف الفرضية البديلة وجود علاقة أو تأثير.

9-تحديد مستوى الأهمية: يتم تحديد مستوى الأهمية(significance level) ، ويمثل عادة بالرمز .α 9-يعبر مستوى الأهمية عن الاحتمالية المقبولة لرفض الفرضية الصفر عن طريق الخطأ. المستوى الشائع للأهمية هو 0.05 أو 5٪.

10-اختيار الاختبار الإحصائي: يتم اختيار الاختبار الإحصائي المناسب بناءً على النوع والطبيعة للبيانات والفرضيات المراد اختبارها. بعض الأمثلة على الاختبارات الإحصائية المشتركة تشمل اختبارt ، واختبار الفروق بين المتوسطات، واختبار الارتباط الخطي.

11-جمع البيانات وإجراء التحليل: يتم جمع البيانات المناسبة للدراسة وإدخالها في البرنامج الإحصائي المناسب. يتم تنفيذ الاختبار الإحصائي وحساب القيمة الإحصائية المرتبطة به.

12-تفسير النتائج: يتم تفسير النتائج الإحصائية بشكل منطقي ومعنوي. يتم تحليل القيمة الإحصائية المحسوبة ومقارنتها بقيمة الحد الحرج (الحد النقطي) المحددة مسبقًا. إذا كانت القيمة الإحصائية أقل من الحد الحرج، يتم رفض الفرضية الصفر واعتماد الفرضية البديلة.

13-الاستنتاجات والتوصيات: يتم استخلاص الاستنتاجات من النتائج الإحصائية وتقييم أهميتها وتأثيرها العملي. يتم توصية بالإجراءات المستقبلية أو الاتجاهات البحثية الإضافية بناءً على النتائج.

14-التقرير: يجب كتابة تقرير يشمل وصفًا للدراسة والفرضيات والمنهج المستخدم والنتائج الإحصائية والتفسيرات والاستنتاجات. يمكن أن يتضمن التقرير أيضًا الجداول والرسوم البيانية المناسبة لتوضيح النتائجالمشكلات التي تتعلق وترتبط بعينات الدراسة:
صفة عامة، هناك العديد من البرامج المتاحة للتحليل الإحصائي. وفيما يلي بعض البرامج الشائعة والمهمة في هذا المجال:

:(SPSS (Statistical Package for the Social Sciencesيعتبر SPSS واحدًا من أبرز برامج التحليل الإحصائي والأكثر استخدامًا في العلوم الاجتماعية. يتيح SPSS إجراء تحليلات إحصائية مختلفة بما في ذلك الاستدلال الإحصائي والتحليل الاستكشافي وإنتاج الرسوم البيانية.

SAS (Statistical Analysis System): يُعتَبَر SAS أحد أنظمة التحليل الإحصائي الرئيسية المستخدمة في البحوث العلمية والصناعة. يتميز SAS بقوته في التعامل مع مجموعات بيانات كبيرة وتنفيذ تحليلات متقدمة.

R: هو لغة برمجة مفتوحة المصدر وبيئة تطوير تستخدم للتحليل الإحصائي والرسوم البيانية. يوفر R مجموعة واسعة من الحزم والوظائف الإحصائية التي تمكن الباحثين من إجراء التحليلات والنمذجة الإحصائية المختلفة.

Excel: على الرغم من أن Excel هو برنامج جداول بيانات أساسي، إلا أنه يحتوي على بعض الوظائف الإحصائية المدمجة التي تسمح بإجراء تحليلات بسيطة مثل الاحتساب الإحصائية الأساسية وإنشاء الرسوم البيانية.

STATA: يُستخدَم STATA في العديد من المجالات مثل الاقتصاد والعلوم الاجتماعية وعلم النفس. يقدم STATA مجموعة قوية من الأدوات الإحصائية والرسوم البيانية ويدعم التحليل الإحصائي المتقدم.من هذه البرامج:

SPSS برنامج:

ويعني برنامج الحزم الإحصائية للتحليل للعلوم الاجتماعية، ويعد هذا البرنامج الرائد في مجال التحليل الإحصائي والأكثراستخداما .وظهر هذا البرنامج في العام 1968 ، وكان الهدف الأساسي من إنشائه تحليل البيانات في العلوم الاجتماعية، وبعد أنالأمريكية بشراء هذا البرنامج ومن IMB أثبت هذا البرنامج نجاحه تم استخدامه في باقي العلوم .وفي العام 2009 قامت شركة:IMB SPSS Statistics ثم قامت بتغيير اسمه حتى أصبحSPSS

مميزات برنامج

•سهولة استخدامه حيث كل ما على الباحث فعله وضع البيانات التي يريد إجراء التحليل الإحصائي لها في الأماكن المحددة داخلالبرنامج، ومن ثم ترك المهمة للبرنامج الذي سيقوم بإجراء التحليل الإحصائي لهذه البيانات وإعطاء نتائج دقيقة بسرعة كبيرة،بالإضافة إلى شموليته حيث يشمل عددا كبيرا من العلوم المختلفة.وبيانات برنامج ، • Excelويتميز هذا البرنامج بقدرته على التعامل مع مختلف البيانات كبيانات برنامج لتحليل الإحصائي البرنامج رقم واحد لدى الشركات الكبرى في العالم، وذلك نظرا للمعلومات التي SPSSويعد برنامج يقدمها عن سوق العمل بالإضافة إلى إتاحته إمكانية التنبؤ بالمستقبل .

SASبرنامج للتحليل الإحصائي:

ويعد من أهم برامج التحليل الإحصائي، وقام بتصميم هذا البرنامج الأستاذان جيمس غودنايت وجون سول، وذلك من أجل استخدامه في وزارة الزراعة الأمريكية .وبعد أن أثبت هذا البرنامج أهمية كبيرة في التحليل الإحصائي قاما بإنشاء(The Statistical Analysis System) SAS والتي تعني نظام التحليل الإحصائيSAS .

مميزات برنامج:

•تعد مرونة هذا البرنامج وسهولة استخدامه أحد أهم الأسباب التي ساهمت بانتشاره بشكل كبير في العالم.
•لهذا البرنامج دور كبير في التنبؤ بالمستقبل، وبالتالي مساعدة الشركات على الاطلاع على سوق العمل بشكل كبير، كما أنالنتائج التي يقدمها هذا البرنامج تعد دقيقة للغاية ويمكن الوثوق بها.

EVIEWEبرنامج للتحليل الإحصائي:

ويعد هذا البرنامج الأكثر استخداما بين الاقتصاديين، والبرنامج الأول في الاقتصاد .ويساعد هذا البرنامج في تقديم معلوماتعن الانحدار الاقتصادي، وبالتالي تبدأ الشركات بوضع الحلول لمعالجة هذا الانحدار .ويوجد لهذا البرنامج نظام حماية مميزحيث يجب على الشخص أن يقوم بتسجيل هويته على الشبكة وذلك لكي يتمكن من الدخول إلى هذا البرنامج .وتسعى الشركةالمصممة لهذا البرنامج إلى تحسينه بشكل مستمر، وذلك من خلال إصدار تحديثات جديدة لها، ويستطيع الباحث الدخول إلى
موقع الشركة وتحميل التحديث بشكل مجاني.

MINITABبرنامج

ويعد هذا البرامج من أهم البرامج بالنسبة للمبتدئين في مجال التحليل الإحصائي، وذلك نظرا لبساطته وسهولة العمل عليه .ومنأبرز المميزات التي يقدمها هذا البرنامج بالإضافة إلى قيامه بالتحليل الإحصائي، فإنه يقوم بتحليل النتائج التي تظهر، ومن ثم تفسيرها .و قبل أن يبدأ الباحث القيام بالتحليل الإحصائي باستخدام هذا البرنامج عليه أن يقوم بإتقان العمل عليه، وذلك لكييتجنب الوقوع في الخطأ__.

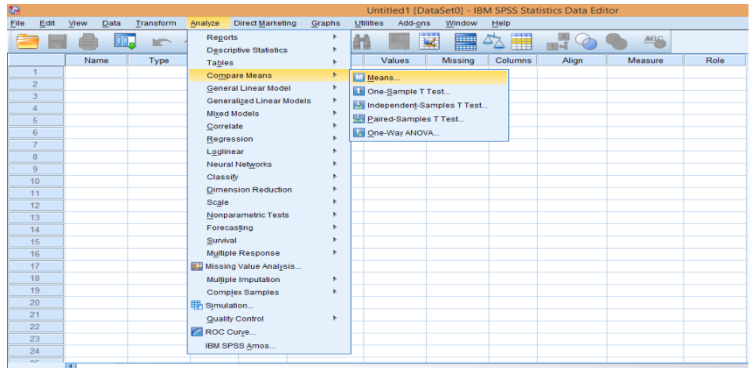
اختبار الفرضبات الاحصائيه باستخدام البرامج الاحصائيه

<u>والآن اختبار الفرضيات المختلفة</u>

اختبار مقارنة المتوسطات (Comparing Mean)

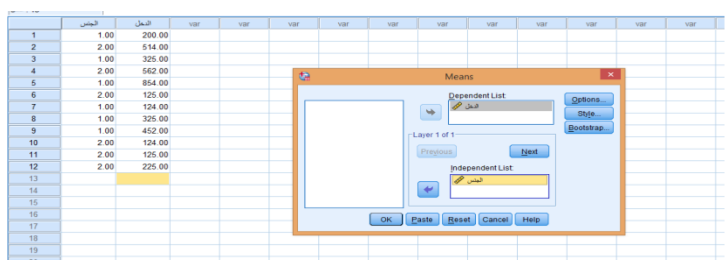
مثال : المطلوب حساب المتوسطات الحسابية لدخل النساء والرجال.

1. نختار من Analyze ← Compare Means ← Means كما تلاحظ بالشكل التالي:

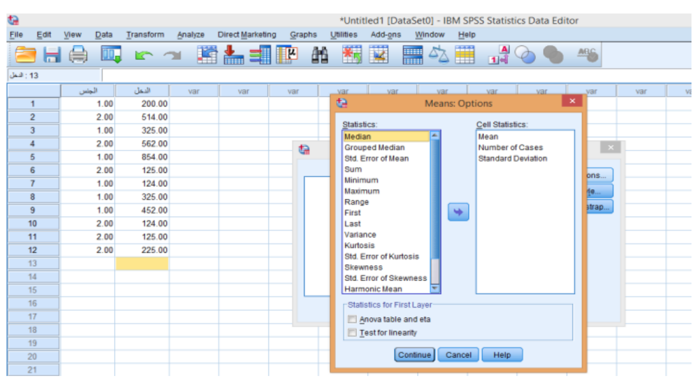


سيظهر مربع الحوار التالي:

2. انقل المتغير المعتمد إلى المستطيل Dependent List والمتغير المستقل إلى المستطيل Independent List: .



3. اضغط Options يظهر مربع الحوار التالي:



4. اختر الإحصاءات اللازمة من المستطيل Statistics وانقلها إلى المستطيل Cell Statistics ، واضغط على المربع بجانب Anova table and eta ، ثم اضغط Continue سنعود إلى مربع الحوار الأصلي

5. اضغط موافق تظهر النتائج التالية:

Means

الجدول التالي يعطي تقريرا لأعداد المشاهدات والنسب المئوية

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
الدخل * الجنس	12	100.0%	0	0.0%	12	100.0%

Report

الدخل

الجنس	Mean	N	Std. Deviation	Kurtosis	Skewness	Std. Error of Mean
ذكر	380.0000	6	258.39737	2.522	1.459	105.49028
انثى	279.1667	6	204.78615	-1.809-	.851	83.60360
Total	329.5833	12	228.44035	1.024	1.163	65.94505

الجدول التالي يعطي المقاييس الإحصائية المطلوبة حسب كل طبقة في المجتمع والسطر الأخير يعطي المقاييس الإحصائية لأفراد المجتمع بكامله ولاحظ الخلاف بين متوسط دخل كل من الذكور والإناث .

✓ *الجدول التالي هو تحليل التباين للمقارنة بين متوسطات دخل الذكور والإناث* وله دلالة إحصائية

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
Current Salary * Gender	Between Groups	(Combined)	2.79E+10	1	2.792E+10	119.798	.000
	Within Groups		1.10E+11	472	233046531		
	Total		1.38E+11	473			

عند مستوى معنوية $\alpha = 0.05$ لان قيمة Sig. = 0 في العمود الأخير من الجدول.

Measures of Association

	Eta	Eta Squared
Current Salary * Gender	.450	.202

✓ *الجدول التالي يبين مقياس إيتا لقياس العلاقة بين الراتب والجنس وهي متوسطة*

<u>اختبار شكل التوزيع</u>

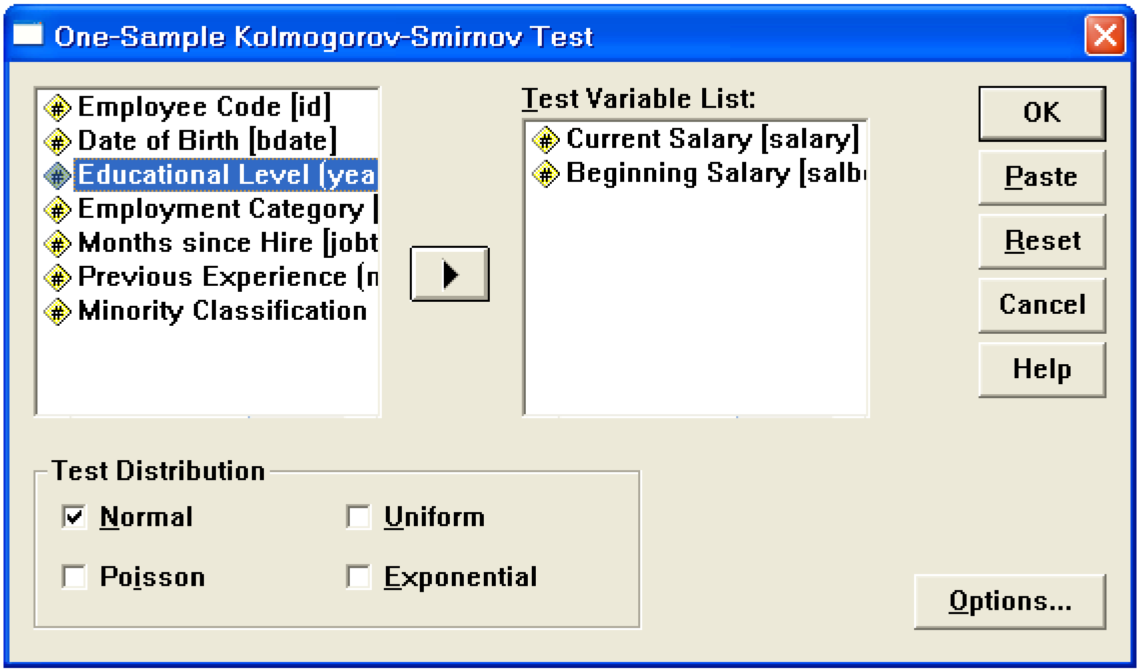
قبل الشروع في تطبيق الاختبارات المختلفة يجب الشروع في طبيعة البيانات هل تتبع التوزيع الطبيعي أم لا فإذا كانت تتبع التوزيع الطبيعي فان الاختبارات المعلمية سوف تستخدم وتطبق ، أما إذا كانت البيانات لا تتبع التوزيع الطبيعي فان الاختبارات غير المعلمية سوف تستخدم.

ولمعرفة نوع التوزيع نستخدم اختبار كولمجروف-سمنروف Kolmogrove-Smirov

مثال : اختبر الفرضية التالية: *" بيانات الرواتب في بداية العمل والرواتب الحالية تتبع التوزيع الطبيعي بمستوى معنوية 0.05"* .

لاختبار هذه الفرضية نقوم بالخطوات التالية:

1. من Analyze اختر Nonparametric Tests ومن القائمة الفرعية اختر 1-Sample K-S يظهر مربع الحوار التالي:



2. انقل المتغير salary والمتغير salbegin إلى المربع Test Variable List ، وتأكد أن المربع بجانب Normal موجود به إشارة " ✓".

3. اضغط Ok تظهر النتائج التالية:

NPar Tests

One-Sample Kolmogorov-Smirnov Test

		Current Salary	Beginning Salary
N		474	474
Normal Parameters[a,b]	Mean	$34,419.57	$17,016.09
	Std. Deviation	$17,075.662	$7,870.638
Most Extreme Differences	Absolute	.208	.252
	Positive	.208	.252
	Negative	-.143	-.170
Kolmogorov-Smirnov Z		4.525	5.484
Asymp. Sig. (2-tailed)		.000	.000

a. Test distribution is Normal.

b. Calculated from data.

من الجدول السابق ينتج أن Sig. = 0.0 لكل من المتغيرين وهي اقل من 0.05 ، لذلك نرفض الفرضية المبدئية التي تقول أن البيانات تتبع التوزيع الطبيعي ، ونقبل الفرضية البديلة التي تقول أن البيانات لا تخضع للتوزيع الطبيعي.

اختبارات T (T-Test)

✓ اختبار T للعينة الواحدة (One Sample T-Test)

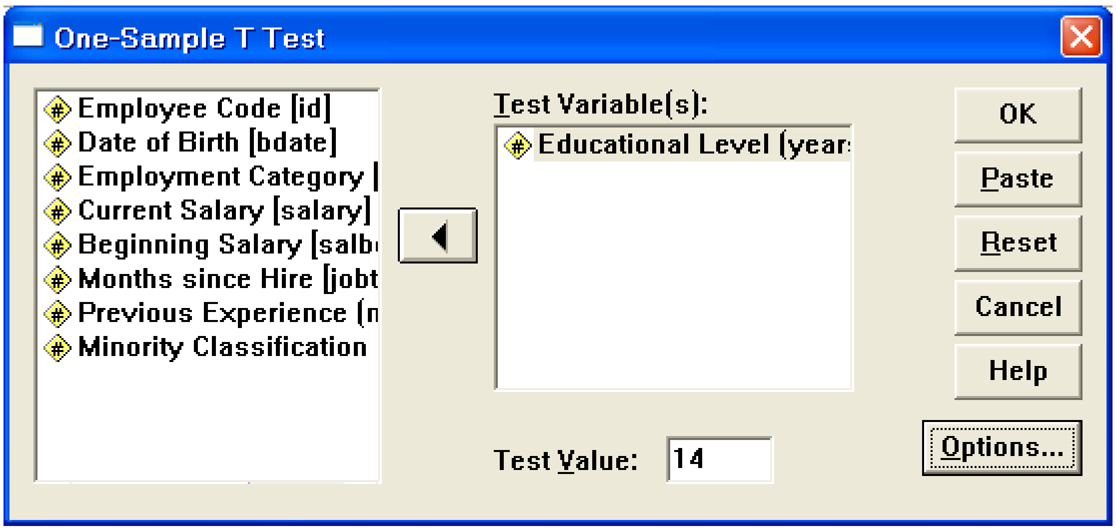
يستخدم هذا الاختبار لفحص فرضية تتعلق بالوسط الحسابي، ويجب تحقق الشرطين التاليين:

1. يجب أن يتبع توزيع المتغير التوزيع الطبيعي، ويستعاض عن هذا الشرط بزيادة حجم العينة إلى اكثر من 30 مفردة.
2. يجب أن تكون العينة عشوائية أي لا تعتمد مفرداتها على بعضها

مثال: *اختبر الفرضية القائلة بان " مستوي تعليم الموظفين يساوي 14 سنة"*

لاختبار هذه الفرضية نتبع الخطوات التالية:

نختار من القائمة Analyzes نختار Compare Mean ومن القائمة الفرعية نختار One Sample T Test يظهر مربع الحوار التالي:



2. انقل المتغير Educ في المربع Test Variable(s) وفي المربع Test Value اكتب العدد 14 ثم اضغط Ok تظهر النتائج التالية:

T-Test

الجدول التالي يبين المتوسط الحسابي للعينة 13.49 وكذلك الفرق بين متوسط العينة والقيمة المفروضة وتساوي 0.51- والانحراف المعياري وعدد أفراد العينة في جدول One-Sample Test يتبين أن .Sig

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Educational Level (years)	474	13.49	2.885	.133

= 0.00 وهي اقل من 0.05 ، لذلك نرفض الفرضية المبدئية أي أن متوسط تعليم الموظفين لا يساوي 14 سنة ، والسؤال هنا هل متوسط تعليم الموظفين في مجتمع الموظفين اكبر أم اصغر من 14 سنة وللإجابة على هذا السؤال نجد أن قيمة t = -3.837 أي سالبة دليل على أن متوسط المجتمع يقل عن 14 سنة.

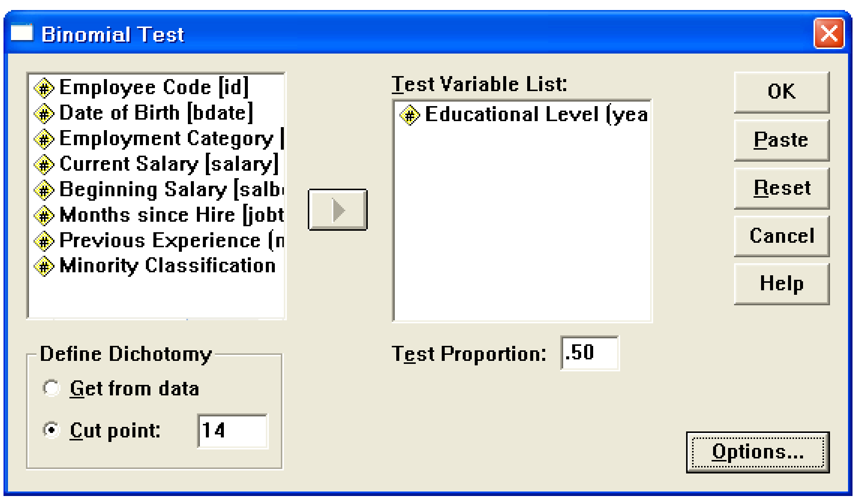
One-Sample Test

	Test Value = 14					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower	Upper
Educational Level (years)	-3.837	473	.000	-.51	-.77	-.25

اختبار الإشارة SIGN TEST (اختبار غير معلمي)

إذا كانت البيانات لا تخضع للتوزيع الطبيعي فيمكن اختبار الفرضية السابقة باستخدام الاختبارات الغير معلميه مثل اختبار الإشارة Sign Test نقوم باتباع الخطوات التالية:

1. من القائمة Analyze نختار الاختبار Parametric Tests ومن القائمة الفرعية نختار Binomial فيظهر المربع التالي:



2. ادخل المتغير educ إلى المربع Test Variable List واكتب 14 في المستطيل المقابل لـ Cut point اسفل Define Dichotomy ثم اضغط Ok تظهر النتائج التالية:

NPar Tests

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Asymp. Sig. (2-tailed)
Educational Level (years)	Group 1	<= 14	249	.53	.50	.291[a]
	Group 2	> 14	225	.47		
	Total		474	1.00		

a. Based on Z Approximation.

من الجدول السابق نجد أن Asymp. Sig. (2-tailed) = 0.291 وهي اكبر من0.05 لذلك نقبل الفرضية المبدئية التي تقول أن متوسط سنوات التعليم تساوي 14 سنة.

لاحظ اختلاف النتيجة في الاختبارين مع ملاحظة أيضا أن نتائج الاختبارات المعلمية تكون أدق من نتائج الاختبارات غير المعلمية وذلك لان الاختبارات الغير معلمية تعتمد على رتب مفردات العينة وليس القيمة الحقيقية لها.

✓ اختبار T للعينات المرتبطة Paired Sample T-Test

يستخدم هذا الاختبار في فحص الفرضيات المتعلقة بمساواة متوسط متغيرين لعينتين غير مستقلتين .

وتكتب الفرضية المبدئية والبديلة بالطريقة التالية:

الفرضية المبدئية: $H_0 : \mu_1 = \mu_2$

الفرضية البديلة: $H_a : \mu_1 \neq \mu_2$

حيث أن μ_1 متوسط العينة الأولى و μ_2 متوسط العينة الثانية

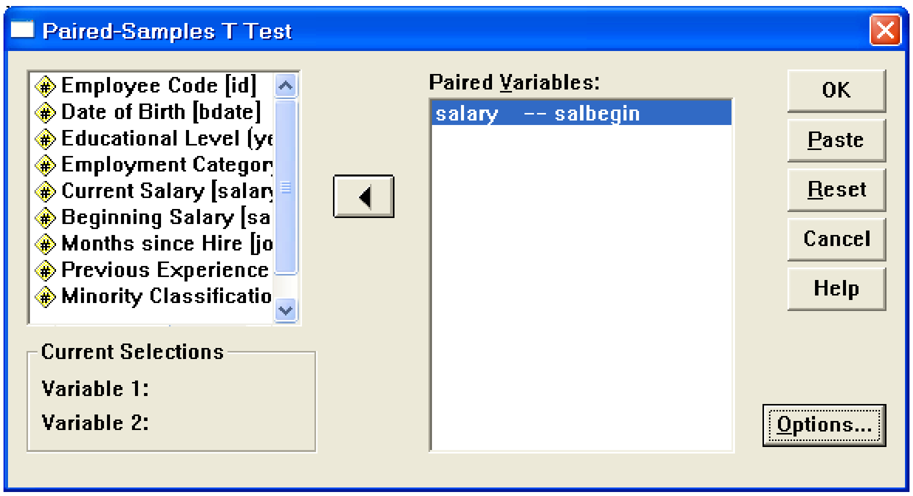
شروط استخدام الاختبار:

1. يجب أن يتبع توزيع الفرق بين المتغيرين طبيعيا، ويستعاض عن هذا الشرط بزيادة حجم العينة إلى اكثر من 30 مفردة.
2. يجب أن تكون العينة عشوائية ، ويجب أن تكون قيم الفرق بين المتغيرين مستقلة عن بعضهما البعض.

مثال: اختبر الفرضية التالية. *" لا يوجد فرق بين متوسط رواتب الموظفين في بداية العمل ومتوسط رواتب الموظفين الحالية "*

ولفحص هذه الفرضية نتبع الخطوات التالية:

1. من القائمة Analyzes نختار Compare Mean ومن القائمة الفرعية نختار Paired Sample T Test يظهر مربع الحوار التالي:



3. ننقل المتغيرين salary, وsalbegin معا إلى المستطيل Paired Variables ثم اضغط Ok تظهر النتائج التالية:

T-Test

✓ *الجدول التالي يبين بعض المقاييس الإحصائية*

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Current Salary	$34,419.57	474	$17,075.661	$784.311
	Beginning Salary	$17,016.09	474	$7,870.638	$361.510

✓ *الجدول التالي يبين معامل الارتباط بين المتغيرين وهو ارتباط قوي وقيمته 0.88*

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Current Salary & Beginning Salary	474	.880	.000

✓ *الجدول التالي يبين قيمة Sig. (2- tailed) = 0.00 وهي أقل من 0.05 وهذا دليل كاف لرفض الفرضية المبدئية ، أي أن هناك فرقا بين متوسط رواتب الموظفين في بداية العمل وفي الوقت الحالي.*

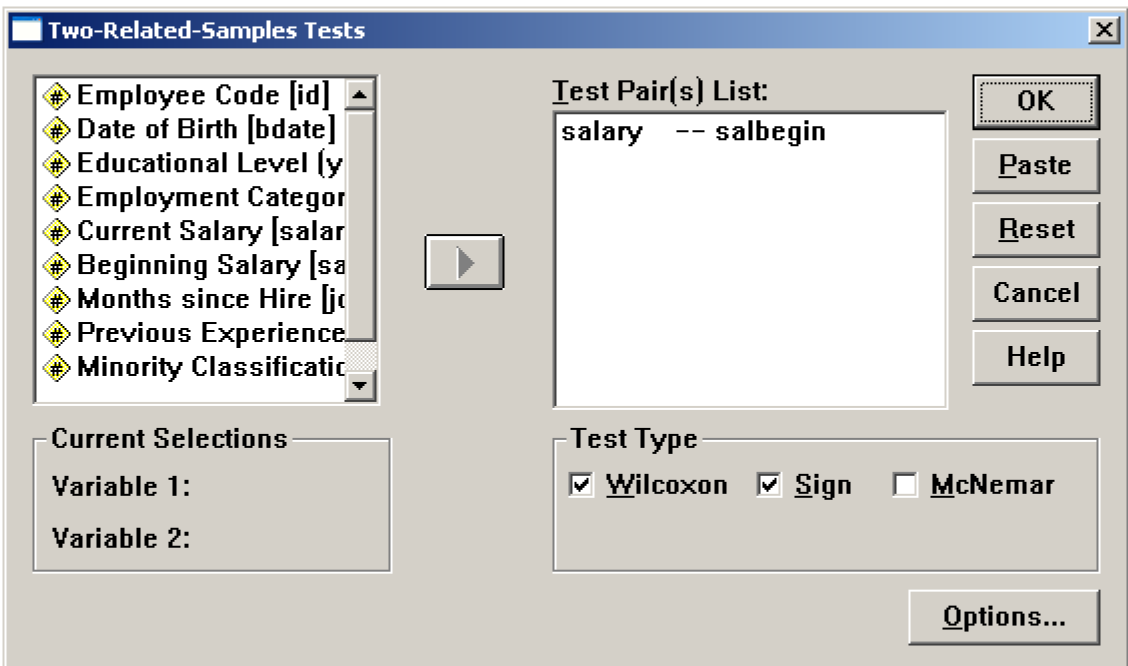
Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper			
Pair 1	Current Salary - Beginning Salary	$17,403.48	$10814.62	$496.732	$16,427.41	$18,379.56	35.036	473	.000

✓ اختبار غير معلمي لمقارنة وسطي مجتمعين في حالة العينات المرتبطة 2 Related Samples

من الممكن أن تكون البيانات لا تخضع للتوزيع الطبيعي، لذلك نلجأ إلى الاختبارات الغير معلمية ، ولفحص الفرضية في المثال السابق باستخدام الاختبارات الغير معلمية نتبع الخطوات التالية:

1. من Analyze اختر الخيار Nonparametric tests ومن القائمة الفرعية اختر 2 related samples يظهر مربع الحوار التالي:



2. ادخل المتغيرين salary و salbegin إلى المستطيل أسفل Test Pair(s) List ، اختر مربع Wilcoxon و Sign، ثم اضغط Ok . تظهر النتائج التالية

NPar Tests

Ranks

		N	Mean Rank	Sum of Ranks
Beginning Salary - Current Salary	Negative Ranks	474[a]	237.50	112575.00
	Positive Ranks	0[b]	.00	.00
	Ties	0[c]		
	Total	474		

a. Beginning Salary < Current Salary

b. Beginning Salary > Current Salary

c. Current Salary = Beginning Salary

Wilcoxon Signed Ranks Test

Test Statistics[b]

	Beginning Salary - Current Salary
Z	-18.865[a]
Asymp. Sig. (2-tailed)	.000

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

من الجدول السابق Sig. = 0.0 لذلك نرفض الفرضية المبدئية ونقبل البديلة أي أنه يوجد اختلاف بين متوسط الراتب الحالي والراتب في بداية العمل.

Sign Test

Frequencies

		N
Beginning Salary - Current Salary	Negative Differences[a]	474
	Positive Differences[b]	0
	Ties[c]	0
	Total	474

a. Beginning Salary < Current Salary

b. Beginning Salary > Current Salary

c. Current Salary = Beginning Salary

Test Statistics[a]

	Beginning Salary - Current Salary
Z	-21.726
Asymp. Sig. (2-tailed)	.000

a. Sign Test

كذلك من اختبار Sign Test نجد أن Sig.= 0.0 أي نرفض الفرضية المبدئية ونقبل البديلة

✓ <u>اختبار T للعينات المستقلة Independent sample T test</u>

هو فحص فرضية متعلقة بمساواة متوسط متغير ما لعينتين مستقلتين، وله شكلان الأول في حالة افتراض أن تباين العينتين متساو، والآخر في حالة افتراض أن تباين العينتين غير متساو.

ولاستخدام هذا المتغير يجب أن يكون لكل مفردة من مفردات العينة قيمة على متغيرين الأول يسمى متغير التجميع (Grouping Variable or Factor) وهو المتغير الذي يقسم العينة الكلية إلى عينتين جزئيتين غير متداخلتين مثل متغير الجنس الذي يقسم العينة إلى عينة ذكور وعينة إناث. والثاني يسمى متغير الاختبار (Test Variable) أو المتغير التابع، وهو متغير كمي مثل الراتب والهدف من هذا الاختبار هو فحص ما إذا كان متوسط الاختبار لفئة متغير التجميع الأولى (الذكور) مساوية لمتوسط متغير الاختبار لدى الفئة الثانية (الإناث) من متغير التجميع.

- <u>شروط اختبار T للعينات المستقلة</u>

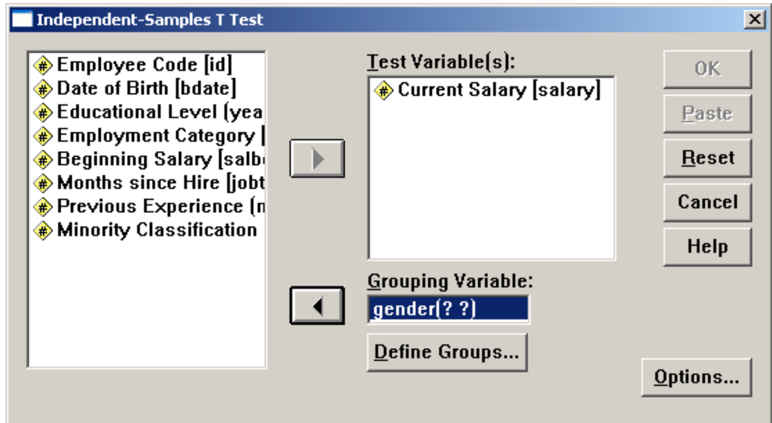
لضمان دقة نتائج اختبار T يجب أن تتوافر الشروط الثلاثة التالية:

1. يجب أن يكون متغير الاختبار طبيعيا في كل فئة من فئات متغير التجميع
2. يجب أن يكون تباين متغير الاختبار متساويا في كلا فئتي متغير التجميع، وإذا لم يتحقق هذا الشرط فان نتيجة اختبار T غير دقيقة، وفي هذه الحالة يمكن حساب قيمة تقديرية للإحصائي T لا يشترط لها مساواة التباين للعينتين.
3. يجب أن تكون العينة عشوائية، ويجب أن تكون قيم متغير الاختبار مستقلة عن بعضها.

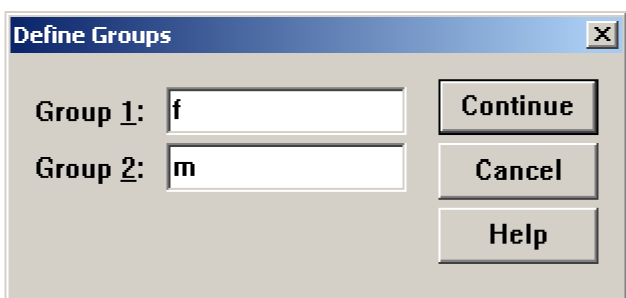
مثال: اختبر الفرضية القائلة *"لا يوجد فرق بين متوسط رواتب الذكور ومتوسط رواتب الإناث"*

ولاختبار هذه الفرضية نتبع الخطوات التالية:

1. من القائمة Analyze اختر Compare Means ثم من القائمة الفرعية اختر Independent Sample T Test فيظهر مربع الحوار التالي:



2. ادخل المتغير Salary إلى المستطيل Test Variable(s) والمتغير gender إلى المستطيل Grouping Variable ، ثم اضغط على Define Groups فيظهر مربع الحوار التالي:



3. ادخل f داخل مستطيل Group 1 وادخل m داخل مستطيل Group 2. ثم اضغط Continue سنعود لمربع الحوار الرئيسي.
4. اضغط Ok ستظهر نتائج الاختبار كالتالي:

Group Statistics

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

T-Test

5. من اختبار (Leven,s test) فقد تم حساب F= 9.669 ومستوى دلالتها Sig = 0.0 وهذا يبين أن تباين العينتين غير متساو ونستخدم اختبار T في حالة عدم تساو يتباين العينتين ونحسب قيمة t= 1.688 ومستوى دلالتها

Sig=0.0 وبذلك نرفض الفرضية المبدئية ونقبل البديلة أي أن متوسطي رواتب العينتين غير متساويين.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Current Salary	Equal variances assumed	119.669	.000	-10.945	472	.000	-$15,409.86	$1,407.906	-$18176.40	-$12643.32
	Equal variances not assumed			-11.688	344.262	.000	-$15,409.86	$1,318.400	-$18003.00	-$12816.73

✓ <u>استخدام الاختبارات الغير معلمية في حالة العينات الغير مرتبطة</u>

<u>اختبار مان-وتني(Mann-Whitney test (U- Test</u>

من المناسب استخدام اختبار مان وتني عند اختبار فرضية تنص على عدم وجود فرق بين متوسطي مجتمعين ما موضع الدراسة وذلك في حالة عدم التأكد من أن توزيع العينتين طبيعيا وكذلك تباين المجتمعين متساويين، أو أن تكون البيانات المأخوذة من العينتين غير دقيقة أو تعتمد على ترتيب عناصر العينتين من حيث القيمة.

مثال: اختبر الفرضية القائلة *"لا يوجد خلاف بين رواتب كل من الكتاب والحراس"*

لاختبار هذه الفرضية نتبع الخطوات التالية: