علم الإحصاء

من المفاهيم الشائعة عن الاحصاء، ما هي الا ارقام وبيانات رقمية فقط، كأعداد السكان واعداد المواليد والوفيات، واعداد المزارع والمزارعين، واعداد الجيش واسلحته...الخ. ومن ثم ارتبط مفهوم الاحصاء بأنه العد والحصر للاشياء والتعبير عنها بارقام وهذا المفهوم المحدود لعلم الاحصاء. ولكن الاحصاء كعلم يمكن تعريفه كاللآتي.

تعريف علم الإحصاء

هو ذلك العلم الذي يعمل على استخدام الأسلوب العلمي في طرق جمع البيانات وتبويبها وتلخيصها وعرضها وتحليلها بهدف الوصول منها على استنتاجات وقرارات مناسبة.

مما سبق يمكن تصنيف علم الاحصاء الى قسمين رئيسين هما:

1. الاحصاء الوصفي: (Descriptive Statistics)

وهو ذلك الفرع من الاحصاء الذي يتناول جمع وتنظيم وتلخيص وتبويب وعرض البيانات وحساب بعض المقاييس الإحصائية المختلفة لها.

2- الاحصاء الاستنتاجي او الاستدلالي: (Statistical inference)

ويشمل الطرق الاحصائية التي تهدف إلى عمل استنتاجات او استدلالات حول المصدر الذي جمعت منه البيانات والتوقعات عن المجتمع من خلال دراسة عينة من ذلك المجتمع.

تعريف المتغير

عند دراسة صفة معينة مثل عدد الثمار في شجرة البرتقال لمجموعة معينة من اشجار البرتقال فسنجد اختلافات في عدد الثمار من شجرة الأخرى، وفي هذه الحالة يطلق على صفة عدد الثمار بالشجرة بمصطلح المتغير. وكذلك عند دراسة صفة حاصل البذور لمحصول الرز في وحدة المساحة (كغم/ دونم) لمجموعة من المزارع، فسنجد اختلافات في كمية حاصل البذور في وحدة المساحة من مزرعة لأخرى، وفي هذه الحالة نطلق على صفة الحاصل بمصطلح المتغير.

تعريف المتغير: (Variable)
هو اي ظاهرة تظهر اختلافات بين مفرداتها

أنواع المتغيرات

تنقسم المتغيرات إلى نوعين:

1 -متغيرات نوعية او وصفية: (Qualitative Variable)

هي بيانات غير رقمية، أو بيانات رقمية مرتبة في شكل مستويات أو في شكل فئات رقمية، ومن ثم تقاس البيانات الوصفية بمعيارين هما:

أ- بيانات وصفية مقاسة بمعيار اسمي: Nominal Scale

وهي بيانات غير رقمية تتكون من مجموعات: متنافية، كل مجموعة لها خصائص تميزها عن المجموعة الأخرى، كما أن هذه المجموعات لا يمكن المفاضلة بينها، ومن الأمثلة على ذلك:

- **الجنس:** متغير وصفي تقاس بياناته بمعيار اسمي "ذكر - انثي."
- **الحالة الاجتماعية:** متغير وصفي تقاس بياناته بمعيار اسمي" متزوج ، اعزب ارمل، مطلق"
- **اصناف التمور:** متغير وصفي يقاس بياناته بمعيار اسمي " برحي، خستاوي، زهدي، مكتوم".
- **الجنسية:** متغير وصفي يقاس بياناته بمعيار اسمي" عراقي غير عراقي" وهذا النوع من البيانات يمكن اعطاء مجموعاته رموز او ارقام، فمثلاً الجنسية يمكن إعطاء الجنسية عراقي"الرمز (1)، والجنسية " غير العراقي " الرمز (2)

ب- بيانات وصفية مقاسة بمعيار ترتيبي: Ordinal Scales

تتكون من مستويات، أو فئات يمكن ترتيبها تصاعدياً أو تنازلياً، ومن الأمثلة على ذلك:

- **تقديرات النجاح للطالب:** متغير وصفي تقاس بياناته بمعيار ترتيبي (مقبول، متوسط، جيد، جيد جدا، امتياز)
- **المستوى التعليمي:** متغير وصفي تقاس بياناته بمعيار ترتيبي" أمي، يقرا ويكتب، ابتدائية، متوسطة، ثانوية، جامعية، أعلى من جامعية"
- **المستوى الاقتصادي:** متغير وصفي تقاس بياناته بمعيار ترتيبي:" غني، متوسط ، فقير"

2- متغيرات كمية (عددية): Quantitative Variables

وهي تلك الظواهر او الصفات التي يمكن قياسها مباشرة بارقام عددية وتنقسم إلى:

أ- متغير متقطع (غير مستمرة): Discrete Variable

وهو المتغير الذي يأخذ أعداد صحيحة، فمثلاً إذا كان X متغير يمثل عدد أفراد الأسرة، فإنه يمكن أن يأخذ القيم 2، 3، 4، 5 ... ولا يمكن أن يأخذ X القيم 1.5، 3.25، 5.17 وكأمثلة اخرى على المتغيرات المتقطعة هي عدد الثمار على النباتات، عدد الطلبة في المدارس، عدد اشجار النخيل في محافة ما، بصورة عامة البيانات المستحصله من طريقة العد (counting) فهي بيانات لمتغير منقطع او غير مستمر.

ب - متغير متصل (مستمر): Continuous Variableا

وهو المتغير الذي لا يمكن ان ياخذ أي قيمة بين قيمتين معنيتين، وكامثلة عن المتغيرات المتصلة كمية الحاصل، الطول، الوزن، الزمن، السرعة...الخ، فإذا كان X هو متغير

الطول فمثلاً X يمكن أن تأخذ القيم 15متر، 11،3 متر، 14،75 متر، أي أن المتغير X يمكن أن ناخذ اي قيمة في فترة زمنية معينة. وبصورة عامة البيانات المستحصلة من طريقة القياس (Measurement) تعتبر المتغير مستمر أو متصل.

تعريف المجتمع والعينة

المجتمع: (Population)

المجتمع من الناحية الاحصائية يمثل جميع الأفراد (او العناصر التي تشترك في صفة متغيرة واحدة أو أكثر تميزه تمييزا تاماً عن بقية المجتمعات.

تعريف المجتمع: (Population)
هو جميع القيم لمتغير ما

ويتعلق مفهوم المجتمع بالهدف المحدد للبحث الاحصائي. فمثلاً اذا كان هدف البحث حساب عدد النخيل في العراق فعندها يكون المجتمع هو جميع مزارع النخيل في العراق بدون استثناء. وتختلف المجتمعات في احجامها (عدد مفرداتها) فبعضها صغير الحجم وبعضها كبيرة والبعض الاخر غير معروف الحجم .لذا فان المجتمعات تقسم الى:

أ- **مجتمع محدود: Finite population**: فاذا كان عدد افراد المجتمع محدود كما هي الحال في عدد اشجار النخيل في مزرعة ما، أو عدد الطلبة في كلية الزراعة.

ب- **مجتمع غير محدود: Infinite Population**: اذا كان حجم المجتمع كبير جدا ولا يمكن حصره مثلاً عدد الأسماك في الخليج العربي، عدد الحشرات على اشجار الحمضيات في محافظة معينة.

العينة: (sample)

في حالة عدم امكانية الحصول على قيم او بيانات عن المجتع لاسباب مادية أو فنية، لذا تلجأ الى اخذ عينة معينة من المجتمع بطريقة ما بحيث تمثله تمثيلا حقيقيا ، لذا تعرف العينة كالاتي:

تعريف العينة: (Sample)
هي جزء من المجتمع مأخوذة منه بطريقة عشوائية وتكون ممثلا له تمثيلا حقيقياً

يعتمد اسلوب العينات على معاينة جزء من المجتمع محل الدراسة، يتم اختياره بطريقة علمية سليمة، ودراسته ثم تعميم نتائج العينة على المجتمع، ويتميز هذا الأسلوب بالآتي:

1- تقليل الوقت والجهد.

2- تقليل التكلفة.

3- الحصول على بيانات أكثر تفصيلاً، وخاصة اذا جمعت البيانات من خلال استمارة استبيان.

4- كما أن اسلوب المعاينة يفضل في بعض الحالات التي يصعب فيها أجراء حصر شامل مثل معاينة دم المريض، اعداد الاسماك في البحر.

ولكن يعاب على هذا الأسلوب بان النتائج المستحصلة بهذا الاسلوب تكون اقل دقة من نتائج اسلوب الحصر الشامل وخاصة اذا كانت العينة المختارة لا تمثل المجتمع تمثيلاً جيداً.

الثوابت (او المعالم) والاحصاءات: (Parameters and Statistics)

يطلق على المقاييس التي تحسب من المجتمع نفسه (أي من جميع القيم) مصطلح الثوابت او المعالم، اما المقاييس المناظرة المحسوبة من العينة فتسمى الاحصاءات او التقديرات لانها لا تمثل سوى تقديرات المعالم المجتمع الذي اخذت منه العينة. وتجدر الاشارة الى ان معالم المجتمع محددة القيم (ثابتة) بينما الاحصاءات المناظرة تتغير بتغير العينة.

الرموز الاحصائية: Statistical Notation

لو اشتملت الدراسة على متغير واحد فعادة ما يرمز لهذا المتغير بحرف هجائي كبير وعادة ما يكون الحرف (X) أما اذا تناولت الدراسة متغيرين او اكثر فيخصص حرف هجائي كبير لكل منها اي (X , Y , W , ... الخ) وعادة ما يرمز لقيمة المتغير بحرفه الهجائي الكبير مع رمز دليلي لتمييز العنصر الذي تقود له تلك القيمة. فلو رمزنا على سبيل المثال للمتغير (ارتفاع نبات الحنطة بالرمز Y : فأن , , هي رموز احصائية تدل على قيمة المتغير (اي ارتفاع النبات) لكل من النباتات رقم 1,2,3 .... الخ . فلو كان ارتفاع النبات الاول هو 115 سم فأن ذلك يعني ان:

Y1=115=CM

وعادة ما يشير الرمز Yi لقيمة المتغير Y للعنصر رقم . 1 وعليه فأن الرمز الدليلي (i) يمثل رقم التسلسل لذلك العنصر، اما الرمز ∑ فأنه حرف يوناني او اغريقي يعني الجمع، فلو كان لدينا خمسة عناصر وان قيمتها للمتغير Y هي y1,y2,y3,y4,y5 فأن:

$$\sum_{i=1}^{n} yi$$

حيث ان ∑ يعني الجمع والرقمان 1 و n هما حدا المجموع

$$\sum_{i=1}^{n} yi = Y1 + Y2 + Y3 + \ldots + Yn$$

وعليه فأن هذا الرمز يقرأ كالآتي:- مجموع قيم Y مبتدأ من المشاهدة الأولى وحتى الأخيرة.

فأن مجموع المشاهدات الخمسة تكتب كما يلي:-

$$\sum_{i=1}^{n=5} yi = Y1 + Y2 + Y3 + Y4 + Y5$$

وللاختصار والسهولة قد يكتب الرمز السابق بدون ذكر حدي المجموع اي $(\sum yi)$ فقط اذا لم يكن هنالك خوف من الالتباس.

وهنالك مجموع جزئي مثل:

$$\sum_{i=3}^{5} yi$$ اي مجموع المشاهدة $5, 4, 3$

$$\sum_{i=3}^{5} yi = Y3 + Y4 + Y5$$

وان مجموع مربعات جميع المشاهدات $\sum_{i=1}^{n=5} yi^2 = Y1^2 + Y2^2 + Y3^2 + Y4^2 + Y5^2$

ويرمز لمربع مجموع المشاهدات $\left(\sum_{i=1}^{n} yi\right)^2 = (Y1 + Y2 + Y3 + \ldots + Yn)^2$

ويرمز لمجموع حاصل ضرب قيم متغيرين y,x

$$\sum xiyi = X1Y1 + X2Y2 + X3Y3 + \ldots + XnYn$$

ويرمز لحاصل ضرب مجموعتين لقيم متغيرين:

$$\left(\sum xi\right)\left(\sum yi\right) = (X1 + X2 + \ldots + Xn)(Y1 + Y2 + \ldots + Yn)$$

وان

$$\sum_{i=1}^{5}(\boldsymbol{yi-c})^{2}=(\boldsymbol{Y1-C})^{2}+(\boldsymbol{Y2-C})^{2}+\ldots+(\boldsymbol{Y5-C})^{2}$$

حيث C تمثل قيمة ثابتة

مثال: لو كانت ارتفاعات خمسة نباتات من الحنطة هي:

110 , 115 , 98 , 120 , 102 سم فأن

Y1+110, Y2=115, Y3=98, Y4=120, Y5=102

فأن قيمة المقادير التالية هي:

$$\sum_{i=1}^{5}\boldsymbol{yi}=\mathbf{Y1+Y2+\ldots+Yn}$$

$$=110+115+98+120+102=545\mathrm{cm}$$

$$\sum_{i=1}^{5}\boldsymbol{yi}^{2}=\boldsymbol{Y1}^{2}+\boldsymbol{Y2}^{2}+\ldots+\boldsymbol{Y5}^{2}$$

$$=(110)^{2}+(115)^{2}+(98)^{2}+(120)^{2}+(102)^{2}=59733$$

$$\left(\sum_{i=1}^{5}\boldsymbol{yi}\right)^{2}=(\boldsymbol{Y1+Y2+\ldots+Y5})^{2}$$

$$=(110+115+98+120+102)^{2}=(545)^{2}=297025$$

$$\sum_{i=1}^{5}\left(\boldsymbol{y_i^2-10}\right)=\left(\boldsymbol{y_1^2-10}\right)+\left(\boldsymbol{y_2^2-10}\right)+\ldots+\left(\boldsymbol{y_5^2-10}\right)$$

$$=(\mathbf{110^2-10})+(\mathbf{115^2-10})+\ldots+(\mathbf{102^2-10})$$

= (12100-10) +(13225-10) +(9604-10) +(14400-10) +(10404-10)

=12090+13215+9594+14390+10394=59683

$$\sum_{i=1}^{5}(yi-50)^2=(y1-50)^2+(y2-50)^2+\cdots+(y5-50)^2$$
$$=(60)^2+(65)^2+(48)^2+(70)^2+(52)^2$$
$$=15429$$

مثال: نفترض بأن قيمة المتغير Y هي كالاتي Yi=3,9,6,2

وان قيمة المتغير X هي كالاتي Xi=4,2,3,7

اوجد قيمة كل مما يأتي:

$$\sum_{i=1}^{n} yi = Y1+Y2+Y3+Y4=3+9+6+2=20$$

$$\sum_{i=2}^{3} yi = Y2+Y3=9+6=15$$

$$\sum Yi^2 = Y1^2+\ldots+Y4^2=3^2+9^2+6^2+2^2=130$$

$$\left(\sum Yi\right)^2=(Y1+Y2+Y3+Y4)^2=(3+9+6+2)^2$$
$$=(20)^2=400$$

$$\sum XiYi = X1Y1+X2Y2+\ldots+X4Y4$$
$$=(3\times 4)+(9\times 2)+(6\times 3)+(2\times 7)=62$$

$$\left(\sum Yi\right)\left(\sum Xi\right)=(X1+X2+\ldots+X4)(Y1+\ldots+Y4)$$
$$=(16)(20)=320$$

ويجب التفريق بين بعض الرموز الاحصائية مثل:

$$\sum \frac{Xi}{Yi}=\frac{X1}{Y1}+\frac{X2}{Y2}+\ldots+\frac{Xn}{Yn}$$

$$\frac{\sum \boldsymbol{Xi}}{\sum \boldsymbol{Yi}} = \frac{\boldsymbol{X1} + \boldsymbol{X2} + \ldots + \boldsymbol{Xn}}{\boldsymbol{Y1} + \boldsymbol{Y2} + \ldots + \boldsymbol{Yn}}$$

***تمرين:** اوجد قيمة كل من المقادير التالية : اذا كانت لديك القيم كما يلي:-

yi=11 , 5 , 6 , 12 , 9 , 8

xi= 10, 7, 4, 12, 13, 8

$$\sum yi \,, \sum xi \,, \sum yi^2 \,, \sum xi^2 \,, \left(\sum yi\right)^2, \left(\sum xi\right)^2, \quad \sum YiXi \,, \sum yi^2 - 5 \,, \sum (Xi-2)^2 \,, \sum \frac{Xi}{Yi} \,, \frac{\sum yi}{\sum xi}$$

العرض الجدولي والتمثيل البياني

عند جمع بيانات حول ظاهرة معينة تبوب في جداول أو يعبر عنها برسوم بيانية لإعطاء الفكرة التي تتضمنها البيانات بأسلوب سريع وبسيط.

العرض الجدولي:

جدول التوزيع التكراري: هو جدول بسيط يتكون من عمودين:

الأول: يحتوي على قيم متغير أو تقسم فيه قيم المتغير الى اقسام او مجموعات تدعى الفئات Classes

الثاني: يبين عدد مشاهدات كل قيمة من قيم المتغير او عدد المشاهدات التي تنتمي الى الفئة ويسمى التكرار Frequency

مثال: جدول توزيع تكراري يتكون من عمود للقيم وعمود للتكرارات، لمصنع تعليب ينتج علب بالأوزان التالية: 3 و4 و5 و6 كغم

التكرارات (عدد العلب المنتجة في يوم) fi	القيم (اوزان العلب بالكيلوغرام) yi
120	3
155	4
145	5
100	6

على سبيل المثال ان عدد العلب المنتجة في اليوم بوزن 3 كغم هو 120 علبة.

مثال: جدول توزيع تكراري يتكون من عمود للفئات وعمود للتكرارات لدرجات الطلبة في مادة الإحصاء.

عدد الطلبة، التكرارfi	فئات درجات الطلبة Classes
3	41-45
7	46-50
10	51-55
10	56-60
8	61-65
5	66-70
4	71-75
1	76-80

على سبيل المثال عدد الطلبة الذين حصلوا على درجات ضمن المدى 61 الى 65 هو 8 طلاب.

تبويب البيانات (انشاء جدول توزيع تكراري)

البيانات غير المبوبة: وهي البيانات الأولية او الاصلية التي جمعت ولم تبوب في جدول.

البيانات المبوبة: وهي البيانات التي بوبت ونظمت في جدول توزيع تكراري.

البيانات التالية تمثل اطوال 80 نباتا من نباتات الذرة الصفراء، المطلوب تبويب هذه البيانات في جدول توزيع تكراري.

80, 87, 98, 81, 74, 48, 79, 80, 78, 82, 93, 91, 70, 90, 80, 84, 73, 74, 81, 56, 65, 92, 70, 71, 86, 83, 93, 65, 51, 85, 68, 72, 68, 86, 43, 74, 73, 83, 90, 35, 75, 67, 72, 90, 71, 76, 92, 93, 81, 88, 91, 97, 72, 61, 80, 91, 77, 71, 59, 80, 95, 99, 70, 74, 63, 89, 67, 60, 82, 83, 63, 60, 75, 79, 88, 66, 70, 88, 76, 63

وهذه البيانات غير مبوبة، نتبع الخطوات التالية لتبويبها في جدول توزيع تكراري

1.استخراج المدى: The Range

المدى = اعلى قيمة – اقل قيمة

= 99 – 35

= 64

2.اختيار وتحديد عدد الفئات: Number of classes

وهناك عدة طرق حسابية تقريبية لايجاد عدد الفئات أهمها:

أ- طريقة Sturges

$$\text{Number of classes} = 1 + (3.3 * \log n)$$

n: عدد المفردات

ب-طريقة Yule

$$\text{Number of classes} = 2.5 * \sqrt[4]{n}$$

ولكل من الطريقتين ميزات وعيوب ولن نستعمل أيا منها هنا بل سنختار عدد الفئات اختيارا على ان لا تقل عن 5 ولا تزيد عن 15 فئة وذلك تبعا لطبيعة البيانات وعدد مفرداتها ومدى التغير فيها.

ولنفرض اننا اخترنا 7 فئات.

3.إيجاد طول الفئة: Class length

يجب ان لا يقل طول الفئة عن $\frac{\text{مدى التغير}}{\text{عدد الفئات}}$ مقربة الى اقرب عدد صحيح اكبر

طول الفئة = $\frac{\text{المدى}}{\text{عدد الفئات}}$

$= \frac{64}{7}$

$= 9.14$

لذا يستحسن ان يكون طول الفئة = 10

ويفضل ان تكون اطوال الفئات متساوية.

4.كتابة حدود الفئات: Class limits

يجب كتابة حدود الفئات بحيث ان جميع قيم المتغير تقع بين الحد الأدنى للفئة الأولى والحد الأعلى للفئة الأخيرة. يمكن ان يكون الحد الأدنى هو قيمة اقل مفردة او اقل منها بقليل، لذا من الممكن ان يكون الحد الأدنى للفئة الأولى 31 وبما ان طول الفئة = 10 لذا فان الحد الاعلى للفئة الأولى هو 40، اذآ الفئة الأولى هي من 31 الى 40 والفئة الثانية من 41 الى 50 والفئة السابعة (الأخيرة) من 91 الى 100 يجب ان يكون الحد الأدنى للفئة الأولى والحد الأعلى للفئة الأخيرة يحوي كافة قيم المتغير.

5.استخراج عدد التكرارات لكل فئة: Class frequency

يتم ذلك بتسجيل القيم الاصلية واحدة تلو الأخرى في الفئة الخاصة بها على شكل علامات أولا ثم ترجمتها الى ارقام كما مبين في الجدول ادناه:

الفئات Classes	fi التكرار بالعلامات	fi التكرار رقمآ
31-40	I	1
41-50	II	2
51-60	IIIII	5
61-70	IIIII IIIII IIIII	15
71-80	IIIII IIIII IIIII IIIII IIIII	25
81-90	IIIII IIIII IIIII IIIII	20
91-100	IIIII IIIII II	12
المجموع		80

ويجب ان يكون المجموع الكلي للتكرارات يساوي العدد الكلي لقيم المتغير (80).

الفئات: Classes: وهي المجاميع التي قسمت اليها قيم المتغير، وهي عبارة عن مدى معين من قيم المتغير.

حدود الفئة: Class Limits: لكل فئة حدان حد ادنى وحد اعلى، فمثلا في الجدول السابق الفئة الأولى حدها الأدنى هو 31 وحدها الأعلى هو 40.

مركز الفئة: وهو عبارة عن منتصف المدى بين حدي الفئة، ويمكن حسابه عن طريق العلاقة الاتية:

$$\text{مركز الفئة} = \frac{\text{الحد الادنى} + \text{الحد الاعلى}}{2}$$

الحدود الحقيقية:

الحد الأدنى الحقيقي للفئة = مركز الفئة - $\frac{1}{2}$ طول الفئة

الحد الأعلى الحقيقي للفئة = مركز الفئة + $\frac{1}{2}$ طول الفئة

طول الفئة = الحد الحقيقي الأعلى للفئة – الحد الحقيقي الأدنى للفئة

مثال: أكمل الجدول التالي

Classes الفئات	fi التكرار	مركز الفئة	الحدود الحقيقية
31-40	1		
41-50	2		
51-60	5		
61-70	15		
71-80	25		
81-90	20		
91-100	12		

الحل:

مركز الفئة = $\frac{\text{الحد الادنى+الحد الاعلى}}{2}$

مركز الفئة الأولى = $\frac{40+31}{2}$

= 35.5

مركز الفئة الثانية = $\frac{50+41}{2}$

= 45.5

وهكذا لبقية الفئات.

الحدود الحقيقية:

الحد الأدنى الحقيقي للفئة الاولى = مركز الفئة الاولى - $\frac{1}{2}$ طول الفئة

= 35.5- $\frac{1}{2}$ (10)

= 30.5

الحد الاعلى الحقيقي للفئة الاولى = مركز الفئة الاولى + $\frac{1}{2}$ طول الفئة

= 35.5 + $\frac{1}{2}$ (10)

= 40.5

الحد الأدنى الحقيقي للفئة الثانية = مركز الفئة الثانية - $\frac{1}{2}$ طول الفئة

= 45.5- $\frac{1}{2}$ (10)

= 40.5

الحد الاعلى الحقيقي للفئة الثانية = مركز الفئة الثانية + $\frac{1}{2}$ طول الفئة

$$(10)\,\frac{1}{2} + 45.5 =$$
$$50.5 =$$

و هكذا لبقية الفئات.

لاكمال الجدول كالآتي

Classes الفئات	fi التكرار	مركز الفئة	الحدود الحقيقية
31-40	1	35.5	30.5-40.5
41-50	2	45.5	40.5-50.5
51-60	5	55.5	50.5-60-5
61-70	15	65.5	60.5-70.5
71-80	25	75.5	70.5-80.5
81-90	20	85.5	80.5-90.5
91-100	12	95.5	90.5-100.5

جدول التوزيع التكراري النسبي

هو جدول يبين الأهمية النسبية لكل فئة ويحسب:

التكرار النسبي لأي فئة $= \dfrac{\text{تكرار تلك الفئة}}{\text{مجموع التكرارات}} = \dfrac{\boldsymbol{fi}}{\boldsymbol{\sum fi}}$

وقد يعبر عن التكرار النسبي كنسبة مئوية بالضرب في 100

التكرار النسبي المئوي لأي فئة $= \dfrac{\text{تكرار تلك الفئة}}{\text{مجموع التكرارات}} \times \mathbf{100}$

مثال: أكمل الجدول

Classes الفئات	fi التكرار	التكرار النسبي	التكرار النسبي المئوي
31-40	1		
41-50	2		
51-60	5		
61-70	15		
71-80	25		
81-90	20		
91-100	12		

التكرار النسبي للفئة الاولى $= \dfrac{\text{تكرار الفئة}}{\text{مجموع التكرارات}}$

$= \frac{1}{80}$

$= 0.0125$

التكرار النسبي المئوي للفئة الأولى = 0.0125×100

$= 1.25$

و هكذا لبقية الفئات

Classes الفئات	fi التكرار	التكرار النسبي	التكرار النسبي المئوي
31-40	1	0.0125	1.25
41-50	2	0.0250	2.50
51-60	5	0.0625	6.25
61-70	15	0.1875	18.75
71-80	25	0.3125	31.25
81-90	20	0.2500	25.00
91-100	12	0.1500	15.00
المجموع	80	1	100

التوزيعات المتجمعة

1. التكرار التجميعي التصاعدي: ويرمز له UCF

التكرار التجميعي التصاعدي للفئة الأولى = تكرار الفئة الأولى :

$$UCF\ 1=f1$$

التكرار التجميعي التصاعدي للفئة الثانية = تكرار الفئة الثانية + تكرار الفئة الأولى:

$$UCF\ 2=f2+f1$$

التكرار التجميعي التصاعدي للفئة الثالثة = تكرار الفئة الثالثة +تكرار الفئة الثانية +تكرار الفئة الأولى:

$$UCF\ 3=f3+f2+f1$$

و هكذا، أي ان التكرار التجميعي التصاعدي لاي فئة = تكرار تلك الفئة + تكرار ما قبلها من الفئات.

2. التكرار التجميعي التنازلي: ويرمز له LCF

التكرار التجميعي التنازلي للفئة الأولى = مجموع التكرارات :

$$LCF\ 1=\sum fi$$

التكرار التجميعي التنازلي للفئة الثانية = مجموع التكرارات – تكرار الفئة الأولى:

$$LCF\ 2=\sum fi-f1$$

التكرار التجميعي التنازلي للفئة الثالثة = مجموع التكرارات – تكرار الفئة الثانية- تكرار الفئة الأولى:

$$LCF\ 3=\sum fi-f2-f2-f1$$

مثال: اكمل الجدول

Classes الفئات	fi التكرار	UCF التكرار التجميعي التصاعدي	LCF التكرار التجميعي التنازلي
31-40	1		
41-50	2		
51-60	5		
61-70	15		
71-80	25		
81-90	20		
91-100	12		

الحل:

التكرار التجميعي التصاعدي للفئة الأولى = تكرار الفئة الأولى

1 =

التكرار التجميعي التصاعدي للفئة الثانية = تكرار الفئة الثانية + تكرار الفئة الأولى

1+2 =

3=

التكرار التجميعي التصاعدي للفئة الثالثة= تكرار الفئة الثالثة +تكرار الفئة الثانية +تكرار الفئة الأولى

1+2+5 =

8 =

التكرار التجميعي التصاعدي للفئة الرابعة

1+2+5+15 =

23 =

وهكذا.

التكرار التجميعي التنازلي للفئة الأولى = مجموع التكرارات:

80=

التكرار التجميعي التنازلي للفئة الثانية = مجموع التكرارات – تكرار الفئة الأولى:

1-80=

79=

التكرار التجميعي التنازلي للفئة الثالثة = مجموع التكرارات – تكرار الفئة الثانية- تكرار الفئة الأولى:

1-2-80=

77=

وهكذا

ليكتمل الجدول كما يلي

Classes الفئات	fi التكرار	UCF التكرار التجميعي التصاعدي	LCF التكرار التجميعي التنازلي
31-40	1	1	80
41-50	2	3	79
51-60	5	8	77
61-70	15	23	72
71-80	25	48	57
81-90	20	68	32
91-100	12	80	12

التمثيل البياني لجدول التوزيع التكراري

لتمثيل الجدول التالي بيانيا بالمدرج التكراري والمضلع التكراري والمنحنى التكراري.

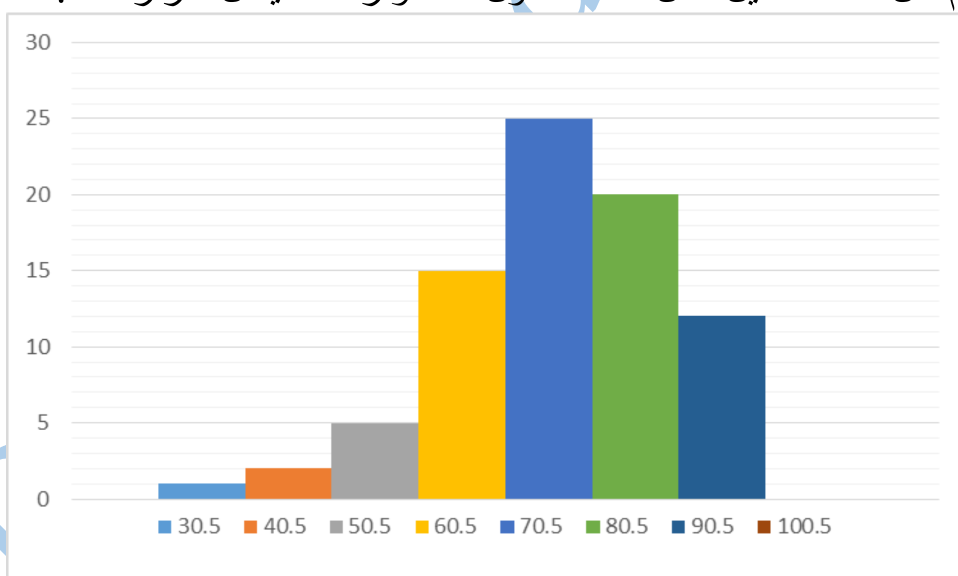
Classes الفئات	fi التكرار
31-40	1
41-50	2
51-60	5
61-70	15
71-80	25
81-90	20
91-100	12

1. المدرج التكراري: Histogram

إيجاد الحدود الحقيقية للفئات.

الفئات Classes	التكرار fi	الحدود الحقيقية
31-40	1	30.5-40.5
41-50	2	40.5-50.5
51-60	5	50.5-60-5
61-70	15	60.5-70.5
71-80	25	70.5-80.5
81-90	20	80.5-90.5
91-100	12	90.5-100.5

- رسم المحورين الافقي x والعمودي y
- تدريج المحور الافقي الى اقسام متساوية بمقياس رسم مناسب بحيث يشمل جميع الحدود الحقيقية للفئات.
- يقسم المحور العمودي الى اقسام متساوية بحيث يشمل اعلى تكرار.
- يرسم لكل فئة مستطيل تمثل قاعدته طول الفئة وارتفاعه يمثل تكرار الفئة.

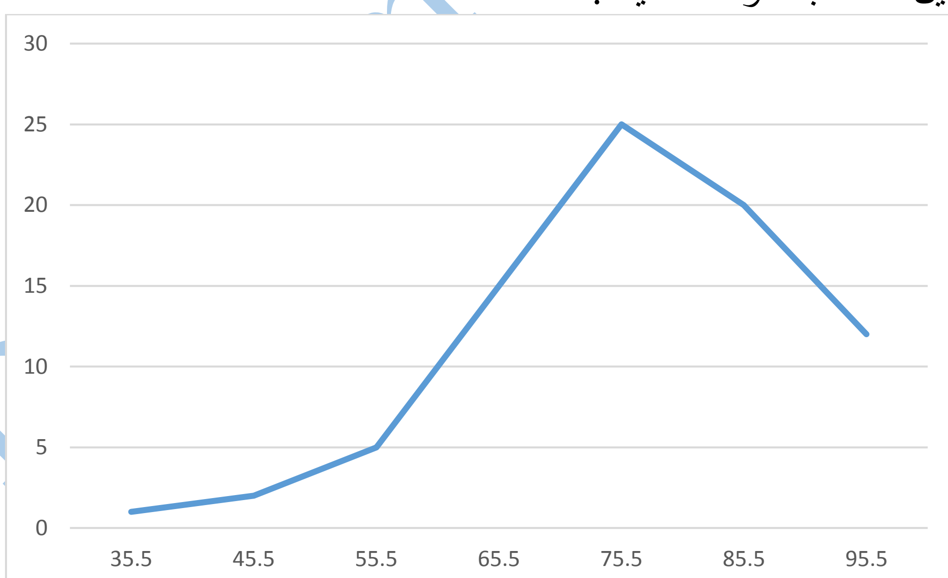


2. المضلع التكراري: Frequency Polygon

- إيجاد مراكز الفئات.

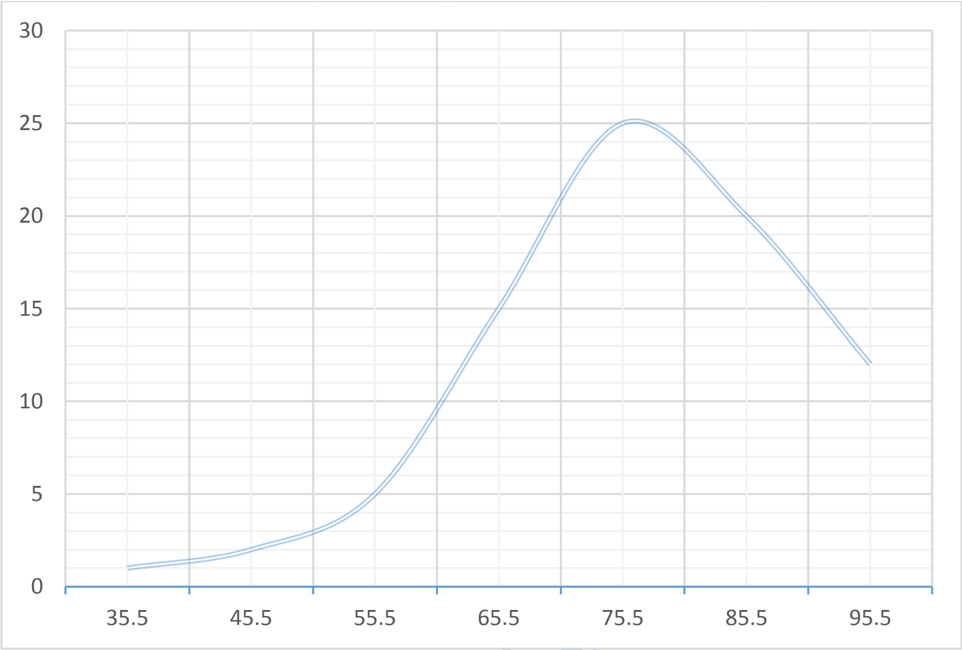
الفئات Classes	التكرار fi	مركز الفئة
31-40	1	35.5
41-50	2	45.5
51-60	5	55.5
61-70	15	65.5
71-80	25	75.5
81-90	20	85.5
91-100	12	95.5

- رسم المحورين الافقي x والعمودي y
- تدريج المحور الافقي الى اقسام متساوية بمقياس رسم مناسب بحيث يشمل جميع مراكز الفئات.
- يقسم المحور العمودي الى اقسام متساوية بحيث يشمل اعلى تكرار.
- يرسم فوق مركز كل فئة نقطة ارتفاعها بقدر تكرار تلك الفئة (نقاط التقاء مركز الفئة مع تكرارها)
- توصيل النقاط بخطوط مستقيمة.



3. المنحنى التكراري: Frequency Curve

وهو عبارة عن منحني يمر بمعظم النقاط الواقعة على مراكز الفئات والتي ارتفاعها يمثل تكرار تلك الفئة (نقاط التقاء مركز الفئة مع تكرارها)



مقاييس التمركز او التوسط
Measures of Central Tendency

هي تلك المقاييس التي تبحث في تقدير قيمة تتمركز حولها اغلبية البيانات، وان هذه القيمة المتوسطة أو المتمركزة تعبر عن او تمثل جميع بيانات المجموعة.

ومن أهم مقاييس التوسط هي:

1. الوسط الحسابي (المتوسط) The Arithmetic Mean
2. الوسيط The Median
3. المنوال The mode

وسيتم شرح كيفية حساب هذه المقاييس في حالتين:

1. حالة البيانات غير المبوبة
2. حالة البيانات المبوبة

الوسط الحسابى (المتوسط): The Arithmetic Mean

هو من اكثر مقاييس النزعة المركزية شيوعا واستخداما ويطلق عليه احيانا Arithmetic average

1. في حالة البيانات غير المبوبة:

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

للمجتمع: $\mu = \frac{\sum Xi}{N}$

للعينة: $\bar{X} = \frac{\sum xi}{n}$

مثال:

اذا علمت ان كمية المطر الساقطة سنويا على مدينة الرمادي خلال خمسة سنوات هي 380، 520، 350، 450، 400 ملم، فما هو متوسط سقوط المطر خلال هذه السنوات الخمس؟

الحل:

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

$$= \frac{\mathbf{380+520+350+450+400}}{\mathbf{5}}$$

=420 mm

2. في حالة البيانات المبوبة:

$$\bar{X} = \frac{\sum fixi}{\sum fi}$$

حيث: fi التكرارات

Xi تمثل القيم في حالة كون الجدول يتكون من قيم وتكراراتها و Xi تمثل مراكز الفئات في حالة كون الجدول يتكون من فئات وتكراراتها.

أي ان الوسط الحسابي = $\frac{\text{مجموع حاصل ضرب القيم في تكراراتها}}{\text{مجموع التكرارات}}$ في حالة جدول القيم وتكراراتها

والوسط الحسابي= $\frac{\text{مجموع حاصل ضرب مراكز الفئات في تكراراتها}}{\text{مجموع التكرارات}}$ في حالة جدول الفئات وتكراراتها

مثال: الجدول ادناه يمثل التوزيع التكراري لأوزان العلب التي ينتجها مصنع تعليب في اليوم، جد الوسط الحسابي

التكرارات (عدد العلب المنتجة في يوم) **fi**	القيم (اوزان العلب بالكيلوغرام) **Xi**
120	3
155	4
145	5
100	6

الحل:

$$\bar{X} = \frac{\sum fixi}{\sum fi}$$

نقوم بإيجاد $\sum$ fixi بإنشاء عمود لحاصل ضرب القيم في تكراراتها ثم جمعها، ثم قسمة الناتج على مجموع التكرارات ($\sum$fi)

Xi	fi	fi xi
3	120	360
4	155	620
5	145	725
6	100	600
	520	2305

$$\bar{X} = \frac{2305}{520}$$

$$= 4.43 \text{ kg}$$

مثال: جد الوسط الحسابي لأطوال نباتات الذرة الصفراء في الجدول التالي.

Classes فئات الطول	fi التكرارات
31-40	1
41-50	2
51-60	5
61-70	15
71-80	25
81-90	20
91-100	12

الحل:

$$\bar{X} = \frac{\sum fixi}{\sum fi}$$

- إيجاد مراكز الفئات
- ضرب التكرارات في مراكز الفئات وإيجاد المجموع
- قسمة الناتج على مجموع التكرارات

Classes الفئات	**fi التكرار**	**xi مركز الفئة**	**fi xi**
31-40	1	35.5	35.5
41-50	2	45.5	91.0
51-60	5	55.5	277.5
61-70	15	65.5	982.5
71-80	25	75.5	1887.5
81-90	20	85.5	1710.0
91-100	12	95.5	1146.0
	∑ fi = 80		**∑fixi= 6130**

$$\bar{X} = \frac{6130}{80}$$

$$= 76.62$$

الوسيط: The Median

هو القيمة التي تتوسط القيم بعد ترتيبها تصاعديا او تنازليا ويرمز له بالرمز Me

1. في حالة البيانات غير المبوبة

- في حالة كون عدد القيم فرديا فان مرتبة الوسيط = $\frac{n+1}{2}$ حيث n هو عدد القيم

مثال: جد مرتبة وقيمة الوسيط لمجموعة القيم التالية:

10، 9، 8، 11، 12، 16، 9

نرتب البيانات تصاعديا كالأتي:

8، 9، 9، 10، 11، 12، 16

عدد القيم 7 وهو فردي

اذاً مرتبة الوسيط = $\frac{n+1}{2}$

$= \frac{7+1}{2}$

$= 4$

مرتبة الوسيط هي الرابعة

قيمة الوسيط هي 10 لأنها تحتل المرتبة الرابعة في ترتيب القيم.

- في حالة كون عدد القيم زوجيا ففي هذه الحالة توجد مرتبتان للوسيط وتحسبان كالاتي:

المرتبة الاولى = $\frac{n}{2}$ و**المرتبة الوسطية الثانية =** $\frac{n}{2}+1$، اما كيفية استخراج قيمة الوسيط في حالة كون عدد البيانات زوجيا فهي كالأتي:

الوسيط = $\frac{\text{القيمة الاولى} + \text{القيمة الثانية}}{2}$

مثال: ما هو الوسيط لمجموعة البيانات التالية:

(10، 14، 7، 5، 20، 22،7، 12)

الحل:

نرتب القيم تصاعديا كالاتي:

5، 7، 7، 10، 12، 14، 20، 22

المرتبة الوسطية الاولى = $\frac{n}{2}$

$= \frac{8}{2}$

$= 4$

المرتبة الوسطية الثانية = $\frac{n}{2}+1$

$= \frac{8}{2}+1$

$= 5$

القيمة عند المرتبة 4 = 10

القيمة عند المرتبة 5 = 12

$$\text{الوسيط} = \frac{\text{القيمة الاولى} + \text{القيمة الثانية}}{2}$$

$$= \frac{10+12}{2}$$

= 11 قيمة الوسيط

2. في حالة البيانات المبوبة

- **اذا كانت مبوبة حسب القيم وتكراراتها**

مثال:

جد قيمة الوسيط لجدول التوزيع التكراري الاتي:

القيم Xi	التكرارات Fi
3	2
4	5
5	4
6	8
	19

الحل: نجد التكرار التجميعي التصاعدي

Xi	fi	التكرار التجميعي التصاعدي
3	2	2
4	5	7
5	4	11
6	8	19
	19	

نجد مرتبة الوسيط

$$\text{مرتبة الوسيط} = \frac{n+1}{2}$$

$$= \frac{19+1}{2}$$

= 10

قيمة الوسيط هي القيمة التي تكرارها التجميعي التصاعدي يشمل مرتبة الوسيط، فنجد ان 5 تكرارها التجميعي التصاعدي 11 وهو اول او اقل تكرار تجميعي تصاعدي يشمل مرتبة الوسيط (10)

اذاً قيمة الوسيط = 5

اذا كانت البيانات مبوبة حسب الفئة وتكرارها فان قيمة الوسيط تحدد حسب المعادلة التالية:

$$\mathbf{Me = L + \frac{(U - L)}{K} \times (nm - A)}$$

حيث ان:

L= قيمة الحد الادنى لفئة الوسيط

U= الحد الاعلى لفئة الوسيط

K= تكرار فئة الوسيط

nm= مرتبة الوسيط

A= التكرار التجميعي التصاعدي للفئة السابقة لفئة الوسيط

مثال: جد الوسيط لجدول التوزيع التكراري الآتي:

الفئات Classes	التكرار fi
31-40	1
41-50	2
51-60	5
61-70	15
71-80	25
81-90	20
91-100	12

الحل: نجد التكرار التجميعي التصاعدي

الفئات Classes	التكرار fi	التكرار التجميعي التصاعدي UCF
31-40	1	1
41-50	2	3
51-60	5	8
61-70	15	23
71-80	25	48
81-90	20	68
91-100	12	80

نجد مرتبة الوسيط: بما عدد القيم (80) زوجي، اذاً توجد مرتبتان للوسيط، ولإيجادهما:

مرتبة الوسيط الأولى = $\frac{80}{2}$

= 40

مرتبة الوسيط الثانية = $\frac{80}{2}+1$

= 41

نحدد فئة الوسيط، وهي الفئة التي تكرارها التجميعي التصاعدي يشمل مرتبتي الوسيط (40 و41)

فنجد ان الفئة الخامسة هي فئة الوسيط لان تكرارها التجميعي التصاعدي 48 يشمل مرتبتي الوسيط، يعني انه اقل تكرار تجميعي تصاعدي يشمل مرتبتي الوسيط.

نجد الوسيط عند المرتبتين 40 و41

$$Me = 71 + \frac{(80-71)}{25} \times (40-23)$$
$$= 71 + 0.36 \times 17$$
$$= 77.12$$

$$Me = 71 + \frac{(80-71)}{25} \times (41-23)$$
$$= 71 + 0.36 \times 18$$
$$= 77.48$$

قيمة الوسيط هو الوسط الحسابي لقيمتي الوسيط

$$Me = \frac{77.12+77.48}{2}$$
$$= 77.3$$

المنـــوال: The Mode

هو القيمة الاكثر شيوعا او تكرار بين مجموعة القيم ويرمز له Mo.

1. في حالة البيانات غير المبوبة:

مثال: جد قيمة المنوال للبيانات التالية والتي تمثل اطوال 10 اشجار

(15، 14، 14، 15، 12، 12، 14، 16، 14، 13)

الحل//

يبدو ان الطول 14 هو الاكثر شيوعا لذا فان قيمة المنوال هي 14

2. في حالة البيانات المبوبة:

- في حالة كون البيانات مبوبة حسب القيمة وتكرارها:

مثال: جد قيمة المنوال للبيانات التالية التي تمثل أوزان 20 حملا عند الولادة

الوزن kg	عدد الحملات (التكرارات)
2	3
2.5	7
4	7
4.5	3

القيمتان 2.5 و4 كغم هما الأعلى تكرارا ومتساويتان في التكرار لذا فأن كل منهما تمثل قيمة مستقلة للمنوال وعليه فأن قيمتي المنوال هي 2.5 و4 وهذا يعني ان الاوزان الشائعة للحملان قيد الدراسة هي 2.5 و4.

- أما اذا كانت البيانات مبوبة حسب الفئات وتكرارها فان قيمه المنوال تحدد وفق المعادلة التالية:

$$\mathrm{Mo} = \mathrm{L} + \frac{\mathrm{K2} - \mathrm{K1}}{(\mathrm{K2} - \mathrm{K1}) + (\mathrm{K2} - \mathrm{K3})} \times (\mathrm{U} - \mathrm{L})$$

حيث ان:

L= الحد الادنى للفئة المنوالية

K1= تكرار الفئة السابقة للفئة المنوالية

K2= تكرار الفئة المنوالية

K3= تكرار الفئة اللاحقة للفئة المنوالية

U= قيمة الحد الاعلى للفئة المنوالية

مثال:

جد المنوال لجدول التوزيع التكراري التالي:

Classes الفئات	fi التكرار
31-40	1
41-50	2
51-60	5
61-70	15
71-80	25
81-90	20
91-100	12

الحل: تحديد الفئة المنوالية، وهي الفئة الأعلى تكراراً.

الفئة الخامسة هي الفئة المنوالية لانها الأعلى تكراراً (25).

نطبق القانون

$$\mathrm{Mo} = \mathrm{L} + \frac{\mathrm{K2} - \mathrm{K1}}{(\mathrm{K2} - \mathrm{K1}) + (\mathrm{K2} - \mathrm{K3})} \times (\mathrm{U} - \mathrm{L})$$

$$Mo = 71 + \frac{25 - 15}{(25 - 15) + (25 - 20)} \times (80 - 71)$$

$$Mo = 71 + \frac{10}{(10) + (5)} \times (9)$$

$$= 77$$

لاحظ ان الناتج يقع ضمن الفئة المنوالية

مقاييس التشتت أو الاختلاف
Measures of Dispersion or Variation

مقاييس التشتت او الاختلاف هي مؤشرات لمدى التباعد او التقارب بين قيم مشاهدات متغير ما، وتضم مقاييس التشتت:

أولا: مقاييس التشتت المطلق:

وهي التي تكون وحداتها نفس وحدات القيم الاصلية وأهمها:

1. المدى: The Range
2. الانحراف المتوسط: The Mean Deviation
3. التباين والانحراف القياسي: The Variance and The Standard Deviation

ثانيا: مقاييس التشتت النسبي:

وهي المقاييس الخالية من وحدات القياس وأهمها:

معامل الاختلاف: Coefficient of Variation

المدى: The Range

المدى هو الفرق بين اعلى قيمة واقل قيمة ويرمز له R

$$R = y_{max} - y_{min}$$

مثال: جد المدى لكل من المجموعات التالية:

a = 5,7,2,1, 9, 6
b= 98, 104, 102, 100, 103, 99
c = 3, 2,6,4,5,210

الحل:

(a) $R = 9-1= 8$

(B) $R= 104 -98 = 6$

(c) $R = 210 -2 = 208$

ان المدى أحيانا يكون مضللاً ولا يعطي فكرة واضحة عن طبيعة البيانات لأنه يعتمد على القيمتين الصغرى والكبرى اللتين كثيراً ما تكون شاذة.

الانحراف المتوسط: The Mean Deviation

يعرف الانحراف المتوسط بانه معدل الانحرافات المطلقة عن وسطها الحسابي، ويرمز له .M.D اذا هو يمثل معدل تشتت القيم عن وسطها الحسابي ويحسب:

$$\text{للعينة}\ \ M.D. = \frac{\sum|xi-\bar{x}|}{n}$$

حيث ان:

$\bar{X}$= متوسط القيم

Xi= القيم

n= عدد القيم

$|xi-\bar{x}|$ = مجموع الانحرافات المطلقة عن المتوسط الحسابي، اي دون الاخذ بنظر الاعتبار الاشارة السالبة او الموجبة وانما تكون جميع القيم موجبة ويسمى بالفرق المطلق.

وتجدر الاشارة الى ان مجموع انحرافات القيم عن وسطها الحسابي = 0 دائماً، لذلك تؤخذ القيمة المطلقة لحساب الانحراف المتوسط.

$$\text{للمجتمع}\ \ \mathbf{M.D.} = \frac{\sum|xi-\bar{u}|}{N}$$

مثال: جد الانحراف المتوسط للبيانات التالية، والذي يمثل أوزان 5 دجاجات.

Xi= 2, 1, 2.5, 1.5, 3 kg

1- أستخراج المتوسط الحسابي $\bar{X}$ للقيم

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

$$= \frac{2+1+2+2.5+1.5+3}{5}$$

$$= \frac{10}{5}$$

$$= 2$$

تسلسل القيم	Xi	الفرق بين القيمة والمتوسط الحسابي $Xi - \bar{X}$	الفرق المطلق $\lvert xi - \bar{x}\rvert$
1	3	3-2 =1	1
2	1.5	1.5 -2 = -0.5	0.5
3	2.5	2.5-2 = 0.5	0.5
4	1	1-2 = -1	1
5	2	2-2 = 0	0
المجموع	**10**	**0**	**3**

$$M.D. = \frac{\sum |xi - \bar{u}|}{N}$$
$$= \frac{3}{5}$$
$$= 0.6 \text{ kg}$$

و هذا يعني ان معدل انحراف الوزن عن المتوسط = 0.6 كغم

Variance :التباين

هو من اهم مقاييس التشتت و يعبر عن مجموع مربعات انحرافات القيم عن وسطها الحسابي ويرمز له (δ^2) للمجتمع و (S^2) للعينة

$$\boldsymbol{\delta^2 = \frac{\sum(xi - \bar{\mu})^2}{N}}$$

μ= المتوسط الحسابي للمجتمع

N= حجم المجتمع

ويمكن كتابة القانون اعلاه بصيغة اخرى:

$$\boldsymbol{\delta^2 = \frac{\sum xi^2 - \frac{(\sum xi)^2}{N}}{N}}$$

أما تباين العينة

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

$\bar{x}$ = المتوسط الحسابي للعينة

n = حجم العينة

ويمكن ان تكتب المعادلة بالصيغة التالية:

$$\boldsymbol{S^2 = \frac{\sum xi^2 - \frac{(\sum xi)^2}{n}}{n - 1}}$$

مثال: جد قيمة التباين للقيم التالية التي تمثل وزن اللحم الصافي لعينة من ستة رؤوس من الغنم.

35, 34, 40, 38, 37, 32 kg

الحل:

- إيجاد مجموع القيم
- إيجاد مجموع مربعات القيم

Xi	Xi²
35	1225
34	1156
40	1600
38	1444
37	1369
32	1024
$\sum Xi = 216$	$\sum Xi^2 = 7818$

$$s^2 = \frac{\sum xi^2 - \frac{(\sum xi)^2}{n}}{n-1}$$
$$= \frac{7818 - \frac{(216)2}{6}}{6-1}$$
$$= \frac{7818 - \frac{46656}{6}}{6-1}$$
$$= \frac{42}{5} =$$
$$= 8.4\ Kg^2$$

وهذا يعني ان متوسط تباين كل قيمة عن المتوسط الحسابي للمجتمع هو 8.4 كغم^2 وبما ان هذه الوحدات غير متداولة في الحياة العامة او غير مألوفة (كغم^2) لهذا يمكن التعبير عن التشتت بوحدات قياس اعتيادية وذلك عن طريق استخدام مقياس تشتت يطلق علية بالانحراف القياسي او المعياري.

Standard Deviation:الانحراف القياسي

الانحراف القياسي (S) هو عبارة عن الجذر التربيعي للتباين (S^2) وعلية فان الانحراف القياسي يمكن حسابة من المعادلات الآتية بالنسبة للمجتمع او العينة:

$$\text{للمجتمع}\ \delta = \sqrt{\frac{\sum x^2 - \frac{(\sum xi)^2}{N}}{N}}$$

$$\text{للعينة}\ S = \sqrt{\frac{\sum x^2 - \frac{(\sum xi)^2}{n}}{n-1}}$$

مثال: جد قيمة الانحراف المعياري للبيانات المعطاة في المثال السابق؟

$$S = \sqrt{\frac{\sum x^2 - \frac{(\sum xi)^2}{n}}{n-1}} = \sqrt{\frac{7818 - \frac{(216)^2}{6}}{5}} = \sqrt{8.4} = 2.89 \text{ kg}$$

ويبدو واضحا ان الانحراف المعياري يأخذ بنظر الاعتبار جميع القيم دون استثناء ويعطي التشتت مقاساً بوحدات قياس متداولة ومتعارف عليها. ولهذه الاسباب وغيرها فان الانحراف المعياري يعتبر أكثر مقاييس التشتت شيوعاً سواء بصيغته المباشرة او عن طريق المقاييس المشتقة منه كمعامل الاختلاف.

معامل الاختلاف: Coefficient of Variation

أن درجة تشتت مفردات مجموعة معينة قد تختلف عن درجة تشتت مجموعة أخرى، وقد يكون هذا الاختلاف كبيراً او صغيراً. وبناء على ذلك، يستخدم معامل الاختلاف كوسيلة لمقارنة درجات التشتت بين مجموعات مختلفة.

ويرمز لمعامل الاختلاف بالرمز C. V. ويحسب وفق المعادلة الاتية:

$$\mathbf{C.V. = \frac{S_x}{\bar{x}} \times 100}$$

S_x هو الانحراف القياسي

$\bar{x}$ هو الوسط الحسابي

مثال: أجريت تجربة لدراسة طول النبات (سم) والحاصل (كغم دونم) لمحصول الذرة الصفراء، وتم الحصول على النتائج المبينة في الجدول ادناه:

	الطول (سم)	كمية الحاصل (كغم)
الوسط الحسابي	200	800
الانحراف القياسي	16	36

قارن بين تشتت الطول والحاصل (ايهما أكثر تشتتاً)

الحل:

$$C.V. = \frac{S_x}{\bar{x}} \times 100$$

بالنسبة للطول $C.V. = \frac{16}{200} \times 100$

$$= 8\ \%$$

بالنسبة للحاصل $C.V. = \frac{36}{800} \times 100$

$$= 4.5\ \%$$

نستنتج ان الطول كان أكثر تشتتاً

نلاحظ انه لو قارنا التشتت بمقياس الانحراف القياسي لكان التشتت أكبر في الحاصل.

Correlation الارتباط

الارتباط البسيط: Simple Correlation

وهو مقياس لتحديد طبيعة العلاقة وقوتها بين متغيرين مستقلين مثل x و y، وتسمى العلاقة طردية او موجبة اذا كانت زيادة قيم أحد المتغيرين يصحبها زيادة قيم المتغير الاخر، واذا اخذت القيم اتجاهين متعاكسين تسمى العلاقة عكسية او سالبة.

تقدير معامل الارتباط: Estimation of Correlation Coefficient

ان معامل الارتباط والذي يرمز له بالرمز r لعينة عشوائية حجمها n لمتغيرين (x و y)

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

يحسب من المعادلة التالية

$$r = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)\left(\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right)}}$$

تتراوح قيمة معامل الارتباط (r) بين +1 و -1 أو أي قيمة بينهما ($-1 \leq r \leq 1$)، وهاتين القيمتين هما درجتا الارتباط التام، الطردي التام (+1) والعكسي التام (-1)، وتجدر الإشارة الى ان معامل الارتباط بين متغيرين هو مقياس وصفي للترابط الخطي بينهما، وعندما تكون $r = 0$ فان ذلك يعني عدم وجود ترابط خطي بينهما وليس بالضرورة عدم وجود علاقة بينهما.

مثال: احسب معامل الارتباط للبيانات التالية والتي تمثل طول وعرض الورقة لنبات ما.

عرض الورقة (x)	13	19	13	18	14	17	14	17	15	16
طول الورقة (y)	15	22	13	20	13	20	15	19	15	18

الحل: لأيجاد معامل الارتباط (r) حسب المعادلة السابقة نرتب البيانات في جدول كما يلي:

x_i	y_i	x_iy_i	x_i^2	y_i^2
13	15	195	169	225
19	22	418	361	484
13	13	169	169	169
18	20	360	324	400
14	13	182	196	169
17	20	340	289	400
14	15	210	196	225
17	19	323	289	361
15	15	225	225	225
16	18	288	256	324
$\sum x_i = 156$	$\sum y_i = 170$	$\sum x_iy_i = 2710$	$\sum x_i^2 = 2474$	$\sum y_i^2 = 2982$

$$r = \frac{\sum x_iy_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)\left(\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right)}}$$

$$r = \frac{2710 - \frac{(156)(170)}{10}}{\sqrt{\left(2474 - \frac{(156)^2}{10}\right)\left(2982 - \frac{(170)^2}{10}\right)}}$$

$$= 0.95$$

نوع العلاقة: طردية (موجبة)

اختبار معنوية الارتباط:

ان الحكم على المعنوية الإحصائية لمعامل الارتباط يعتمد على اختبارات إحصائية محددة، وذلك بمقارنة القيمة المطلقة لـ r المحسوبة مع قيمة r الجدولية عند درجة حرية n-2، حيث n = عدد ازواج المشاهدات، وعند مستوى الاحتمالية المطلوب، فاذا كانت قيمة r المحسوبة اكبر من او تساوي r الجدولية فان هذا يعني ان الارتباط معنوي اما اذا كانت قيمة r المحسوبة اقل من r الجدولية فان هذا يعني ان الارتباط غير معنوي ففي المثال السابق نستخرج قيمة r الجدولية من جدول r عند درجة حرية 8 ومستوى احتمالية 0.05 فنجدها تساوي 0.63

الاستنتاج: بما أن قيمة r المحسوبة (0.95) هي اكبر من قيمة r الجدولية (0.63) نستنتج ان الارتباط معنوي.

df\α	0.2	0.1	0.05	0.02	0.01	0.001
1	0.951057	0.987688	0.996917	0.999507	0.999877	0.999999
2	0.800000	0.900000	0.950000	0.980000	0.990000	0.999000
3	0.687049	0.805384	0.878339	0.934333	0.958735	0.991139
4	0.608400	0.729299	0.811401	0.882194	0.917200	0.974068
5	0.550863	0.669439	0.754492	0.832874	0.874526	0.950883
6	0.506727	0.621489	0.706734	0.788720	0.834342	0.924904
7	0.471589	0.582206	0.666384	0.749776	0.797681	0.898260
8	0.442796	0.549357	0.631897	0.715459	0.764592	0.872115
9	0.418662	0.521404	0.602069	0.685095	0.734786	0.847047
10	0.398062	0.497265	0.575983	0.658070	0.707888	0.823305
11	0.380216	0.476156	0.552943	0.633863	0.683528	0.800962
12	0.364562	0.457500	0.532413	0.612047	0.661376	0.779998
13	0.350688	0.440861	0.513977	0.592270	0.641145	0.760351
14	0.338282	0.425902	0.497309	0.574245	0.622591	0.741934
15	0.327101	0.412360	0.482146	0.557737	0.605506	0.724657
16	0.316958	0.400027	0.468277	0.542548	0.589714	0.708429
17	0.307702	0.388733	0.455531	0.528517	0.575067	0.693163
18	0.299210	0.378341	0.443763	0.515505	0.561435	0.678781
19	0.291384	0.368737	0.432858	0.503397	0.548711	0.665208
20	0.284140	0.359827	0.422714	0.492094	0.536800	0.652378
21	0.277411	0.351531	0.413247	0.481512	0.525620	0.640230
22	0.271137	0.343783	0.404386	0.471579	0.515101	0.628710
23	0.265270	0.336524	0.396070	0.462231	0.505182	0.617768
24	0.259768	0.329705	0.388244	0.453413	0.495808	0.607360
25	0.254594	0.323283	0.380863	0.445078	0.486932	0.597446
26	0.249717	0.317223	0.373886	0.437184	0.478511	0.587988
27	0.245110	0.311490	0.367278	0.429693	0.470509	0.578956
28	0.240749	0.306057	0.361007	0.422572	0.462892	0.570317
29	0.236612	0.300898	0.355046	0.415792	0.455631	0.562047
30	0.232681	0.295991	0.349370	0.409327	0.448699	0.554119

df\α	0.2	0.1	0.05	0.02	0.01	0.001
35	0.215598	0.274611	0.324573	0.380976	0.418211	0.51889
40	0.201796	0.257278	0.304396	0.357787	0.393174	0.48957
45	0.190345	0.242859	0.287563	0.338367	0.372142	0.46467
50	0.180644	0.230620	0.273243	0.321796	0.354153	0.44320
60	0.164997	0.210832	0.250035	0.294846	0.324818	0.40786
70	0.152818	0.195394	0.231883	0.273695	0.301734	0.37979
80	0.142990	0.182916	0.217185	0.256525	0.282958	0.35681
90	0.134844	0.172558	0.204968	0.242227	0.267298	0.33754
100	0.127947	0.163782	0.194604	0.230079	0.253979	0.32109
125	0.114477	0.146617	0.174308	0.206245	0.227807	0.28860
150	0.104525	0.133919	0.159273	0.188552	0.208349	0.26431
175	0.096787	0.124036	0.147558	0.174749	0.193153	0.24528
200	0.090546	0.116060	0.138098	0.163592	0.180860	0.22984
250	0.081000	0.103852	0.123607	0.146483	0.161994	0.20607
300	0.073951	0.094831	0.112891	0.133819	0.148019	0.18843
350	0.068470	0.087814	0.104552	0.123957	0.137131	0.17465
400	0.064052	0.082155	0.097824	0.115997	0.128339	0.16352
450	0.060391	0.077466	0.092248	0.109397	0.121046	0.15427
500	0.057294	0.073497	0.087528	0.103808	0.114870	0.14643
600	0.052305	0.067103	0.079920	0.094798	0.104911	0.13378
700	0.048427	0.062132	0.074004	0.087789	0.097161	0.12393
800	0.045301	0.058123	0.069234	0.082135	0.090909	0.11598
900	0.042711	0.054802	0.065281	0.077450	0.085727	0.10938
1000	0.040520	0.051993	0.061935	0.073484	0.081340	0.10380
1500	0.033086	0.042458	0.050582	0.060022	0.066445	0.08482
2000	0.028654	0.036772	0.043811	0.051990	0.057557	0.07348
3000	0.023397	0.030027	0.035775	0.042457	0.047006	0.06002
4000	0.020262	0.026005	0.030984	0.036773	0.040713	0.05199
5000	0.018123	0.023260	0.027714	0.032892	0.036417	0.04651

جدول (r)

α: مستوى الاحتمالية

df: درجات الحرية

مثال: جد قيمة معامل الارتباط وبين نوع الارتباط مع اختبار معنوية الارتباط للمتغيرين x و y

X	6	1	8	10	2
y	5	7	4	3	9

الحل: لإيجاد معامل الارتباط (r) نرتب البيانات في جدول كما يلي:

x_i	y_i	x_iy_i	x_i^2	y_i^2
6	5	30	36	25
1	7	7	1	49
8	4	32	64	16
10	3	30	100	9
2	9	18	4	81
$\sum x_i=27$	$\sum y_i=28$	$\sum x_iy_i=117$	$\sum x_i^2=205$	$\sum y_i^2=180$

$$r = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sqrt{(\sum x_i^2 - \frac{(\sum x_i)^2}{n})(\sum y_i^2 - \frac{(\sum y_i)^2}{n})}}$$

$$r = \frac{117 - \frac{(27)(28)}{5}}{\sqrt{(205 - \frac{(27)^2}{5})(180 - \frac{(28)^2}{5})}}$$

$$r = \frac{-34.2}{\sqrt{(59.2)(23.2)}}$$

$$r = \frac{-34.2}{37.05}$$

$$r = -0.92$$

العلاقة عكسية (سالبة)

لاختبار معنوية الارتباط نستخرج قيمة r الجدولية من جدول r عند درجة حرية 3 (n-2) ومستوى احتمال 0.05 فنجدها تساوي 0.87

بما أن قيمة r المحسوبة (0.92-) قيمتها المطلقة (0.92) أكبر من قيمة r الجدولية (0.87) نستنتج ان الارتباط معنوي على مستوى احتمال 0.05. اما على مستوى احتمال 0.01 نجد ان قيمة r الجدولية تساوي 0.95، بما أن قيمة r المحسوبة (0.92-) قيمتها المطلقة (0.92) اقل من قيمة r الجدولية (0.95) نستنتج ان الارتباط غير معنوي على مستوى احتمال 0.01.

الانحدار Regression

الانحدار الخطي البسيط: Simple Linear Regression

كان اهتمامنا في في المحاضرات السابقه منصب حول قضايا الاحصاء الاستنتاجي التي تعود الى توزيع ذو متغير واحد ورمزنا لهذا المتغير بالرمز (y)

اما الان فتحول اهتمامنا الى قضايا تخص توزيع ذو متغيرين وسنرمز لهذين المتغيرين بالرمز (x) و(y) فمثلاً قد يكون المتغير (x) هو عدد نباتات القطن في وحده المساحه، بينما المتغير (y) هو كميه المحصول الناتج او قد يكون (x) هو درجات الحرارة، بينما المتغير (y) هو الكميه المذابة في 100 غرام من الماء من ماده كيميائيه معينه او قد يكون (x) هو المعدل الفصلي للطلبة، بينما (y) هو الدرجات النهائيه لهم لمادة الاحصاء ومن ذلك يتضح بأن كل فرد من افراد العينه له قياسان ان احدهما للمتغير (x) والاخر للمتغير (y) فمثلاً لكل طالب درجتان هما المعدل الفصلي (x) ودرجته النهائية (x)

ان الغاية الرئيسيه من دراسه توزيع ذو متغيرين هي:

أ. لتحديد العلاقه الحقيقيه بين (x) و(y) ووضعهما بشكل معادله بحيث يمكن التنبؤ منها عن (y) بدلاله (x)(وهذا ما يسمى بالانحدار Regression)

ب. لقياس درجة العلاقة بين المتغيرين (وهذا مايسمى correlation او بمعنى اخر قياس مدى التلازم والقرابةبين متغيرين مستقلين.

إن الغرض من استخدام أسلوب تحليل الانحدار الخطي البسيط، هو دراسة وتحليل اثر متغير كمي على متغير كمي أخر، ومن الأمثلة على ذلك ما يلي:

- دراسة أثر كمية السماد على إنتاجية الدونم
- دراسة أثر الإنتاج على التكلفة
- دراسة أثر كمية البروتين التي يتناولها الأبقار على الزيادة في الوزن
- أثر الدخل على الإنفاق الاستهلاكي

وهكذا هناك أمثلة في كثير من النواحي الاقتصادية، والزراعية، والتجارية.

في تحليل الانحدار البسيط، نجد أن الباحث يهتم بدراسة أثر أحد المتغيرين ويسمى بالمتغير المستقل او المتنبا منه، على المتغير الثاني ويسمى بالمتغير التابع أو المتنبأ به، ومن ثم يمكن عرض نموذج الانحدار الخطي في شكل معادلة خطية من الدرجة الأولى، تعكس المتغير التابع كدالة في المتغير المستقل كما يلي:

$$\bar{y}_x = \alpha + \beta x$$

حيث أن:

y هو المتغير التابع الذي يتأثر

x هو المتغير المستقل الذي يؤثر

α هو الجزء المقطوع من المحور الصادي y ، وهو يعكس قيمة المتغير التابع في حالة انعدام قيمة المتغير المستقل x ، أي في حالة $(x = 0)$.

β ميل خط المستقيم، ويعكس مقدار التغير في y اذا تغيرت x بوحدة واحدة

ان المعادلة اعلاه تسمى معادلة خط الانحدار للمجتمع وان α و β هي ثوابت

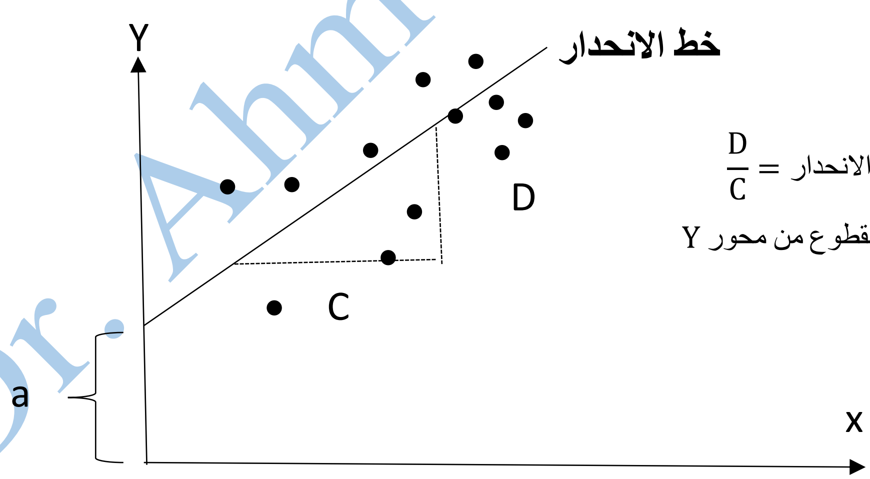
يعرف الثابت α بانه معدل قيمة y عندما x تساوي صفر وتسمى نقطة تقاطع خط الانحدار مع المحور الصادي (y- Intercept)

يعرف الثابت β بميل خط الانحدار للمجتمع (ويسمى معامل انحدار y على x بأنه معدل التغيير في y عندما تتغير قيمة x قيمة واحدة وعند تقدير الميل من العينة يرمز له b

وبذلك نمثل معادلة خط الانحدار للعينة بالمعادلة التالية

$$y = a + bx$$

ومن المعادلة اعلاه يمكن التنبؤ بقيمة y من قيم x عند معرفة الثوابت a و b



$b =$ ميل خط الانحدار $= \frac{D}{C}$

$a =$ الجزء المقطوع من محور Y

كيفية حساب معامل الانحدار b

يحسب معامل الانحدار او ما يسمى ايضا ميل خط الانحدار b بالمعادلة التالية:

$$\mathbf{b} = \frac{\sum(\mathbf{X_i} - \overline{\mathbf{X}})(\mathbf{Y_i} - \overline{\mathbf{Y}})}{\sum(\mathbf{X_i} - \overline{\mathbf{X}})^2}$$

او يمكن استعمال المعادلة المختصرة التالية

$$\mathbf{b} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

كيفية حساب معامل تقاطع خط الانحدار مع المحور الصادي a:

يمكن حساب معامل تقاطع خط الانحدار مع المحور الصادي a بالمعادلة التالية:

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

استخدامات تحليل الانحدار:

يستخدم تحليل الانحدار في المجالات التالية:

1. وصف العلاقة الكمية بين المتغير التابع والمتغير المستقل.
2. التنبؤ بقيم المتغير التابع عند مستويات محددة للمتغير المستقل.
3. السيطرة على المتغير التابع عن طريق التحكم بمستوى المتغير المستقل الذي يؤثر فيه.

الدرجة النهائية y	الدرجة الفصلية x
85	65
74	50
76	55
90	65
85	55
87	70
94	65
98	70
81	55
91	70
76	50
74	55

x_i	y_i	x_i^2	y_i^2	x_iy_i
65	85	4225	7225	5525
50	74	2500	5476	3700
55	76	3025	5776	4180
65	90	4225	8100	5850
55	85	3025	7225	4675
70	87	4900	7569	6090
65	94	4225	8836	6110
70	98	4900	9604	6860
55	81	3025	6561	4455
70	91	4900	8281	6370
50	76	2500	5776	3800
55	74	3025	5476	4070
$\sum x_i= 725$	$\sum y_i= 1011$	$\sum x_i^2= 44475$	$\sum y_i^2= 85905$	$\sum x_iy_i= 61685$

$$\mathbf{b} = \frac{\sum \mathbf{x_i y_i} - \frac{(\sum \mathbf{x_i})(\sum \mathbf{y_i})}{\mathbf{n}}}{\sum \mathbf{X_i}^2 - \frac{(\sum \mathbf{x_i})^2}{\mathbf{n}}}$$

$$b = \frac{61685 - \frac{(725)(1011)}{12}}{4475 - \frac{(725)^2}{12}} = 0.897$$

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

= 84.250 - (0.897) (60.417) =30.056

$$\bar{y}_x = \alpha + bx$$

وبالتعويض بمعادلة خط الانحدار نحصل على:

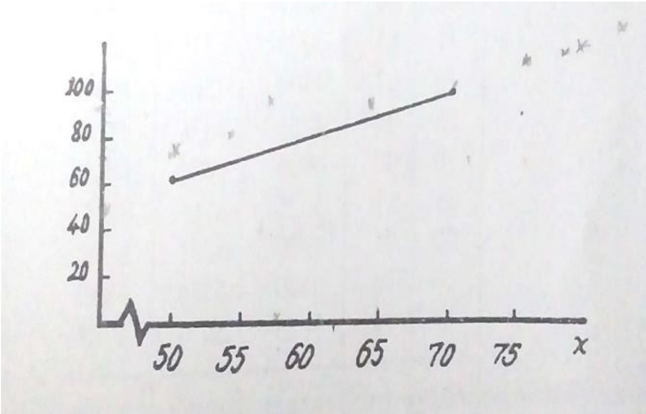
$$\bar{y}_x = 30.05 + 0.897x$$

والان لو عوضنا عن اي قيمتين من قيم x المعطاة بالسؤال ولتكونا 50 و 70 فأن

$$\bar{y}_{50} = 30.05 + 0.897(50) = 74.9$$

$$\bar{y}_{70} = 30.056 + 0.897(70) = 92.8$$

والخط المستقيم الذي يمر بهاتين النقطتين هو خط الانحدار



هذا بالتعويض في هذه المعادلة عن كل قيمة من قيم x فأنة يمكن حساب القيم التقديرية لـ y اي $\bar{y}_x$ او $(\hat{y})$ والتي تقع جميعها على خط الانحدار البسيط وهذه القيم موضحة في الجدول التالي:

x_i	y_i	$(\hat{y})$ او $\bar{y}_x$
65	85	88.4
50	74	74.9
55	76	79.4
65	90	88.4
55	85	79.4
70	87	92.8
65	94	88.4
70	98	92.8
55	81	79.4
70	91	92.8
50	76	74.4
55	74	79.4

هذا ومن خواص معادلة الانحدار الخطي البسيط هو:

1. ان قيمة النقطة $(\bar{y}_1, \bar{x})$ تقع على خط الانحدار.
2. ان خط الانحدار يمر من جميع قيم $(x_i, y_{i.....})$
3. ان مجموع الانحرافات عن خط الانحدار يساوي صفراً اي $\sum(X_i - \bar{X}_x)^2 = 0$
4. وان مجموع مربعات الانحرافات عن خط الانحدار هي اقل ما يمكن اي

$$\sum(X_i - \bar{X}_x)^2 = \text{minimum}$$

هذا ويمكن استخدام معادلة خط الانحدار للتنبؤ عن قيمة **y** لقيمة معينة من **x** غير موجودة في العينة

فمثلاً الدرجة النهائية المتوقعة لطالب معدلة الفصلي 73 هي

$$\bar{y}_{50} = 30.05 + 0.897(73) = 95.5$$

مبادئ نظرية الاحتمال: Elementary probability theory

ان نظرية الاحتمال تلعب دورا هاما في نظريات وتطبيقات علم الإحصاء، ونظرية الاحتمال تعنى بدراسة التجارب العشوائية.

مصطلحات وتعاريف:

1-التجربة العشوائية: The random experiment:

هي التجربة التي لا يمكن معرفة نتيجتها لخضوعها لقوانين الاحتمال، أن رمي زار الطاولة هي تجربة عشوائية، لان النتائج الممكنة لهذه التجربة تخضع لقوانين الاحتمال، وإذا اراد كيمياوي ان يقدر كمية الزيت في بذور القطن فان العينة التي سيبني تقديره عليها والنتائج التي سيحصل عليها ستخضع لقوانين الاحتمال ولهذا فان هذه التجربة عشوائية.

2-فضاء العينة Sample space:

فضاء العينة هو مجموعة من النقاط تمثل جميع النتائج الممكنة لتجربة ما، حيث أن كل نتيجة Out come تمثل بنقطة Point او عنصر Element في فضاء العينة، فعند رمي قطعة نقود متزنة عشوائياً ولمرة واحدة فان فضاء العينة يكون من نتيجتين ممكنتان: H وT اي صورة وكتابة.

3-الحادث (الحدث) The event:

هو نقطة او عدة نقاط في فضاء العينة ويرمز له (E)، فالحصول على الصورة (H) من رمي قطعة نقود مرة واحدة يسمى حدثا وهو يتكون من نقطة واحدة من مجموعة نقاط فضاء العينة [H, T]

4-الحوادث المتنافية (المستبعدة): Mutually exclusive) events:

يقال عن الحدثين، E_2E_1 انهما متنافيان (مستبعدان أي ان أحدهما يبعد الاخر) إذا استبعد حدوثهما معنا، فمثلاً عند رمي قطعة نقود فمن المستحيل الحصول على صورة وكتابة في نفس الوقت.

5-الحوادث المستقلة: Independent events:

هي الحوادث التي إذا وقع أحدها لا يمنع او يؤثر على وقوع الاحداث الاخرى فمثلاً عند رمي قطعتي نقود فالحصول على الصورة من القطعة الأولى لايؤثر في نتيجة القطعة الثانية.

6-الحوادث غير المستقلة: Non independent events:

هي الحوادث التي إذا وقع أحدها يؤثر في وقوع الأحداث الأخرى. فمثلاً في حالة صندوق به كرات بيضاء وسوداء فعند سحب كرتان على التوالي بحيث لا تعاد الكرة الأولى فان نتيجة السحبة الثانية تتاثر بنتيجة السحبة الاولى لذا فالحادثان غير مستقلين.

7-الحالات الممكنة: Possible cases:

هي الحالات المختلفة التي يمكن أن تظهر في تجربة معينة، فعند رمي قطعة النقود فعدد الحالات الممكنة هنا حالتين (صورة او كتابة) وعند رمي زهرة النرد فعدد الحالات هي (6) حالات.

8-الحالات المواتية:

هي الحالات التي تحقق ظهور الحادث المراد دراسته وتسمى ايضا بحالات النجاح، فمثلاً عند رمي زهر الطاولة فاذا كان الحادث هوالحصول على عدد زوجي فالحالات التي تتحقق ظهور هذا الحادث هي الحصول على (2،4،6) فهذه الحالات الثلاثة تسمى الحالات المواتية.

9-الحالات المتماثلة:

هي الحالات المتكافئة والمتساوية في اماكنية حدوثها عند رمي قطعة النقود فان الظروف المهيئة للحصول على أي وجه (صورة، كتابة) تكون متكافئة فيقال بان الحالتين اللتان تنتج عن تجربة رمي قطعة النقود حالتان متماثلتان.

10 -مضروب n (n Factorial)

مضروب n (ويرمز له n!) ويعرف بانه:

n! = n (n-1) (n-2) ...1

فمثلا مضروب5! هو.

n! = 5 (4) (3) (2) (1) = 120

11-التباديل: Permutation

ويقصد بالتباديل بأنها عدد طرق الاختيار المرتب التي يمكن تكوينها من عدد أشياء بأخذها كلها أو بعضها ويرمز لها ب npr أي تباديل r من n وقانونه هو

$$npr = \frac{n!}{(n-r)!}$$

مثال: إذا كان لدينا اربعة حروف هي (ABCD) ونريد ان نعمل تشكيلة مكونة من حرفين فقط من الاربعة حروف اعلاه، فما هو عدد الطرق التي يمكن اختيارها لهذين الحرفين؟

الحل:

$$4P2 = \frac{4!}{(4-2)!} = \frac{4 \times 3 \times 2 \times 1}{2 \times 1} = 12$$

أي أن عدد الطرق = 12 وهي:

AB, AC, AD, BC, BD, CD, BA, CA, DA, CB, DB, DC

حيث إن كلا منهما يمثل ترتيبا مختلفة للحروف، فهنا مثلا ترتيب الحرفين بطريقة AB يختلف عن طريقة الترتيب BA وهكذا في ترتيب بقية الحروف.

حالة وجود المجموعات المتشابهة:

لندرس الان عدد الطرق المختلفة التي يمكننا فيها ترتيب حروف كلمة (باب) قد تبدو لأول وهلة ان عدد هذه الطرق هو 6 =!3 الا ان تكرار الحرف (ب) مرتين يمنع استعمال هذه القاعدة لانه لا يمكن التميز بين (ب) الموجودة في أول الكلمة و(ب) الموجودة في اخر الكلمة، وفي الحقيقة أن عدد الطرق هو (3) وليس (6) وهي ب ا ب ، اب ب ، ب ب 1. فعدد الحروف 3، اثنان منها متشابهة لذا فان عدد الطرق الممكن بها ترتيب حروف كلمة (باب) هو:

$$\frac{3!}{2!\,1!} = 3$$

ويمكن القول بانه إذا كانت لدينا من الاشياء، ومنهما m متشابهة فان عدد طرق ترتيب هذه الاشياء على خط هي: $\frac{n!}{m!}$

فاذا كانت هناك m_1 من الاشياء وm_2 من الاشياء المتشابهة الاخرى فان عدد الطرق الممكنة الترتيب هذه الاشياء على خط هو: $\frac{n!}{m_1!m_2!}$

فمثلاً عدد الطرق الممكنة لترتيب حروف كلمة (السلسلة) هو:

$$\frac{7!}{2!\,3!\,1!\,1!} = 420$$

وبالإمكان الان تعميم هذه القاعدة فنقول إن التباديل في حالة وجود مجموعات متشابهة هي:

$$\frac{n!}{m_1!\,m_2!\,m_3!}$$

علماً بان: $m_1 + m_2 + m_3 + \ldots\ldots = n$

مثال: ما هو عدد الترتيبات التي يمكن تكوينها من أحرف كلمة (Statistics).

الحل:

عدد الأحرف الكلية = 10

وان الحرف s تكرر 3 مرات

وان الحرف t تكرر 3 مرات

وان الحرف a تكرر مرة واحدة

وان الحرف ز تكرر مرتان

وان الحرف c تكرر مرة واحدة

لذا فان:

$$\frac{n!}{m_1!\,m_2!\,m_3!\,m_4!\,m_5} = \frac{10!}{3!\,3!\,1!\,2!\,1!} = 50400$$

12-التوافيق: Combinations

يقصد بالتوافيق بانها عدد طرق الاختبار غير المرتب التي يمكن تكوينها من عدة أشياء بأخذها

كلها أو بعضها ويرمز له بـ (nCr) او $\binom{n}{r}$

وقانونه هو:

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$

أي أن الترتيب في حالة التوافيق غير مهم، على العكس من التباديل، ففي المثال السابق عن ترتيب الحروف (A,B,C,D) كان يهمنا أي حرف ياتي اولا وايهما ياتي ثانية حيث كل حالة تعتبر صورة مغايرة للاخرى، اما في التوافيق فلا يهم أي حرف ياتي اولا وايهما ثانية فكلا الحالتين تعتبر نتيجة واحدة.

لاحظ بأن:

$$\binom{n}{r} = \binom{n}{n-r}$$

مثال: ماعدد طرق الاختيار التي يمكن الحصول عليها لاختيار لجنة مؤلفة من 5 اشخاص من مجموع 9 اشخاص؟

الحل: لاحظ بان ترتيب الأشخاص هنا غير ضروري لان اختيار عمر قبل حيدر او العكس هي نتيجة واحدة.

$$\binom{9}{5} = \frac{9!}{5!\,(9-5)!} = \frac{9!}{5!\,4!} = 126$$

الاحتمال: Probability

من المفاهيم الاساسية في الأحتمالات مفهوم المحاولة، وتعرف بأنها العملية التي تجرى ويتوقع أن ينتج عنها شيء، ومثال ذلك ولادة مولود، رمي قطعة نقود، السحبة، النجاح، الفشل، ألخ ..... والمحاولات نوعان:

أ-نوع ينتج عنه نتيجتين فقط لا ثالث بينهما وتسمى محاولات برنولي:

ومثال ذلك رمي قطعة النقود، ومن شروط محولات برنولي ان تكون النتيجة واضحة ومستقلة وتنفي الحدث الاخر وان يكون هناك احتمال بحصول الحدث.

ب -محاولات اخرى تكون نتائجها أكثر من اثنين:

لنفرض ان حدثا معينا (E) يمكن أن يتحقق (او يحدث او ينجح في n من الحالات (وتسمى الحالات المواتية (favorable cases من مجموعة N من الحالات الممكنة All possible cases وعلى فرض ان جميع الحالات المكنة هذه متساوية الحدوث Equally likely cases فان درجة احتمال ظهور الحدث E_i ويرمز له $P(E_i)$ هي:

$$P(E_i) = \frac{\text{عدد الحالات المؤاتية للحدث}}{\text{عدد الحالات الممكنة}} = \frac{n}{N}$$

هذا وان احتمال عدم ظهور هذا الحدث (أي فشله) ويرمز له بـ P(Ɇ)

هو:

P(Ɇ) = 1- P(E)

ويسمى P(Ɇ) باحتمال الحادث المكمل او الاحتمال العكسي.

<u>**مثال1:**</u> لو رميت قطعة نقود متزنة بصورة عشوائية مرة واحدة فقط فما هو احتمال حصول الحدث (H) أي الصورة؟

<u>**الحل:**</u>

$P(H)=n/N =1/2$

لان الحالات الممكنة (N) = 2 وهي اما الصورة او الكتابة.

<u>**مثال2:**</u> رميت قطعة نقود متزنة 3 مرات وبصورة عشوائية فما هو احتمال الحصول على الصورة مرتين.

<u>**الحل:**</u> عدد الحالات الممكنة = 8 حالات (الان 2×2×2 = 8) وهي:

HHH, HHT, HTH, THH, HTT, THT, TTH, TTT

اما الحالات المواتية = 3 حالات وهي:

(HHT, HTH, THH)

لذا فاحتمال الحصول على الصورة مرتين هما:

$$P(2H) = \frac{3}{8}$$

بعض خواص الاحتمالات:

1) إذا رمزنا لاحتمال حدوث حادث بالرمز $p(E)$ واحتمال عدم حدوث هذا الحادث بالرمز $P(E)$ فإن: $P(E_2) + P(E_2) = 1$

أن درجة احتمال أي حادث تتراوح بين الصفر والواحد اي $0 < P(E) < 1$ أي انه إذا ظهر الحدث المطلوب فان احتماله = 1 وعندها يسمى الحدث (E) بانه اكيد، وذا لم يظهر فاحتماله = صفر وعندها يسمى الحدث (E) بانه مستحيل.

3) اذا كانت

$E_1E_2,\ldots..E_n$ هي عناصر او نقاط فضاء العينة فان مجموع درجات احتمالاتها =1 أي أن:

$$P(E_1) + P(E_2) +\ldots+ P(E_n) = 1$$

$$\sum P(E_i) = 1$$

مثال: صندوق يحتوي على (6) كرات حمراء و(4) بيضاء و(5) صفراء فاذا سحبت منه كرة عشوائية فاحتمال أن تكون (1) حمراء (2) بيضاء (3) صفراء.

الحل:

حمراء: $P(R) = \frac{6}{15}$

بيضاء: $P(w) = \frac{4}{15}$

صفراء : $P(Y) = \frac{5}{15}$

$$P(R) + P(W) + P(Y) = 1$$

Tests of Hypotheses :اختبار الفرضيات

يعد اختبار الفرضيات الإحصائية من أهم المواضيع في مجال اتخاذ القرارات

مصطلحات ضرورية:

1. **الفرضية الاحصائية:** هي عبارة عن أدعاء أو تصريح قد يكون صائبا أو خاطئا حول معلمة أو أكثر لمجتمع أو لمجموعة من المجتمعات.

تؤخذ العينة من المجتمع وتدرس وتستخدم جميع المعلومات المتحصل عليها للوصول الى قرار بقبول أو رفض الفرضية الاحصائية. في حالة كون بيانات العينة تساند النظرية فأن الفرضية تقبل، أما أذا كانت البينات تناقض النظرية ففي هذه الحالة ترفض الفرضية. تجدر الاشارة هنا الى ان قبولنا الفرضية الاحصائية هو ناتج عن عدم وجود أدلة كافيه لرفضها من بيانات العينة ولذلك فأن قبولنا لهذه الفرضية لا يعني بالضرورة كونها صحيحه أما أذا رفضنا الفرضية بناءً على المعلومات الموجودة في بيانات العينة فأن ذلك يعني بأن الفرضية خاطئة. لذلك فأن الباحث أو الاحصائي يحاول دائما أن يضع الفرضية بشكل يأمل أن يرفضها، فمثلا أذا أراد الباحث أن يقارن صنفا جديدا من الحنطة مع الصنف المحلي فأنه يضع فرضية مفادها بأنه لا يوجد فرق معنوي بين الصنفين.

أن الفرضية التي يضعها الباحث على امل أن يرفضها تدعى بفرضية العدم Null Hypothesis ويرمز لها بـH_0، ورفضنا لفرضية العدم يقودنا الى قبول فرضية بديلة عنها هذه الفرضية تدعى الفرضية البديلة ويرمز لها بـ H_1

2. الاخطاء التي يقع فيها الباحث

أن طريقة اتخاذنا القرارات قد يقودنا الى الوقوع في نوعين من الخطأ هما

- خطأ من النوع الاول: يقع الباحث في الخطأ من النوع الاول أذا رفض فرضية العدم عندما تكون هي الفرضية الصحيحة.
- خطأ من النوع الثاني: يقع الباحث في الخطأ من النوع الثاني أذا قبل فرضية العدم عندما تكون هي الفرضية الخاطئة

كما هو موضح في الجدول التالي:

الحالة الحقيقية / القرار	H_0 صحيحه	H_1 خاطئة
قبول H_0	قرار صائب	خطاء من النوع الثاني
رفض H_0	خطاء من النوع الاول	قرار صائب

3. مستوى المعنوية: Level of Significant

أو مستوى الاحتمال Probability level

أو حجم الاختبار Size of the test

يعرف **مستوى المعنوية** بأنه درجة الاحتمال الذي نرفض به فرضية العدم H_0 عندما تكون هي الصحيحة أو بعبارة أخرى هو احتمال الوقوع في الخطاء من النوع الاول ويرمز لها بـ α أي:

$$\alpha = P\,(\text{Type I error})$$
$$= P(\text{Reject } H_0 \mid H_0 \text{ is true})$$

فاذا أستخدمنا مستوى معنوية 5% ورفضنا فرضية العدم H_0 نقول بأن هناك فرق معنوي، أما أذا رفضت فرضية العدم بمستوى معنوية 1% فيقال بأن هناك فرق معنوي جدا وفي الحالة الاولى فهذا يعني بأنه لو تكررت التجربة 100 مره فأن 5 منها تؤيد صحة فرضية العدم و 95 لا تطابق فرضية العدم.

منطقة الرفض: هي تلك المنطقة التي أذا وقعت قيمة المختبر الإحصائي داخلها تسبب في رفض فرضية العدم H_0.
وتحدد منطقة الرفض بعد تعيين مستوى المعنوية
تسمى المنطقة تحت المنحني غير منطقة الرفض بمنطقة القبول التي أذا وقعت قيمة المختبر الاحصائي فيها يسبب في قبول فرضية العدم H_0.
المختبر الإحصائي: هو عبارة عن متغير عشوائي له توزيع احتمالي معلوم ويصف المختبر الاحصائي العلاقة بين القيم النظرية في المجتمع والقيم المعلومة أو القيم المحسوبة في المجتمع وعادة تقارن قيمة المختبر الاحصائي المحسوب من العينة مع قيمته المستخرجة من توزيعه الاحتمالي (الموجود في جداول خاصة) ومنها تتخذ القرارات برفض أو قبول فرضية العدم H_0.

خطوات اختبار الفرضية:

1- تحديد نوع توزيع المجتمع أي هل هو توزيع طبيعي أو نوع من التوزيعات الاخرى أو توزيع معتدل. أن البحث يوضح التوزيع حول المتغير العشوائي هل هو توزيع طبيعي أو توزيع ذي حدين علما بأن معظم الظواهر يكون توزيعها مشابه للتوزيع الطبيعي.
2- صياغة فرضية العدم والبديلة، تصاغ الفرضيات بلغة معالم المجتمع فمثلا عندما يكون الاختبار عن المتوسط الحسابي بأنه يساوي قيمة معينة، فرضاً أن معدل درجات الطلبة 70 إذن فرضية العدم H_0: $\mu=70$ والفرضية البديلة $\mu \neq 70$
3- اختبار مستوى المعنوية، يكون اختيار مستوى المعنوية مسبقاً من قبل الباحث وفي البحوث العلمية يأخذ مستوى المعنوية 5% (معنوي) أو 1% (معنوي جداً) وبعد ذلك تحدد منطقة القبول
4- المختبر الإحصائي، حيث يتم اختيار المختبر الإحصائي الذي سيكون قاعدة لاختبار الفرضيات، والمختبرات الإحصائية منها Z او T أو X^2 أو اختبار F.
5- جمع البيانات من العينة وحساب المختبر الاحصائي بحسب صيغة حساب معينه نذكرها لاحقاً.
6- اتخاذ القرارات أذا وقعت قيمة المختبر الإحصائي في منطقة الرفض ترفض فرضية العدم وتقبل البديلة ويقال بأن هنالك فرق معنوي بين القيم النظرية للمجتمع والقيم المحسوبة من العينة وإذا وقعت قيمة المختبر في منطقة القبول تقبل فرضية العدم ويقال بأنه لا يوجد فرق معنوي بين نتائج العينة والقيم النظرية للمجتمع.

اختبار الفرضيات الخاصة بالمتوسط الحسابي الحقيقي للمجتمع μ:

1- اختبار Z:

شروط استخدامه

- حجم العينة العشوائية يجب أن يكون اكبر من 30 ($n \geq 30$)
- توزيع المجتمع يجب أن يكون طبيعيا أو مقارب منه
- تباين المجتمع σ^2 معلوم

وهو من الاختبارات التي تستخدم في حالة متوسط حسابي واحد ويحسب المختبر الاحصائي Z من المعادلة:

$$Z=\frac{\bar{Y}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$$

حيث:
$\bar{Y}$: الوسط الحسابي للعينة
μ_0: الوسط الحسابي للمجتمع
σ: الانحراف القياسي للمجتمع، ويمكن أن يستعاض عنه بالانحراف القياسي للعينة (S)

n: حجم العينة
بعد اختيار مستوى المعنوية فأن منطقة الرفض ومنطقة القبول ستحدد
لنفرض بأن Y يتوزع توزيعا طبيعيا بمتوسط μ وتباين σ^2 أي: $Y \sim N(\mu_1$ $\sigma^2)$
وكانت الرغبة باختبار فرضية العدم ضد الفرضية البديلة

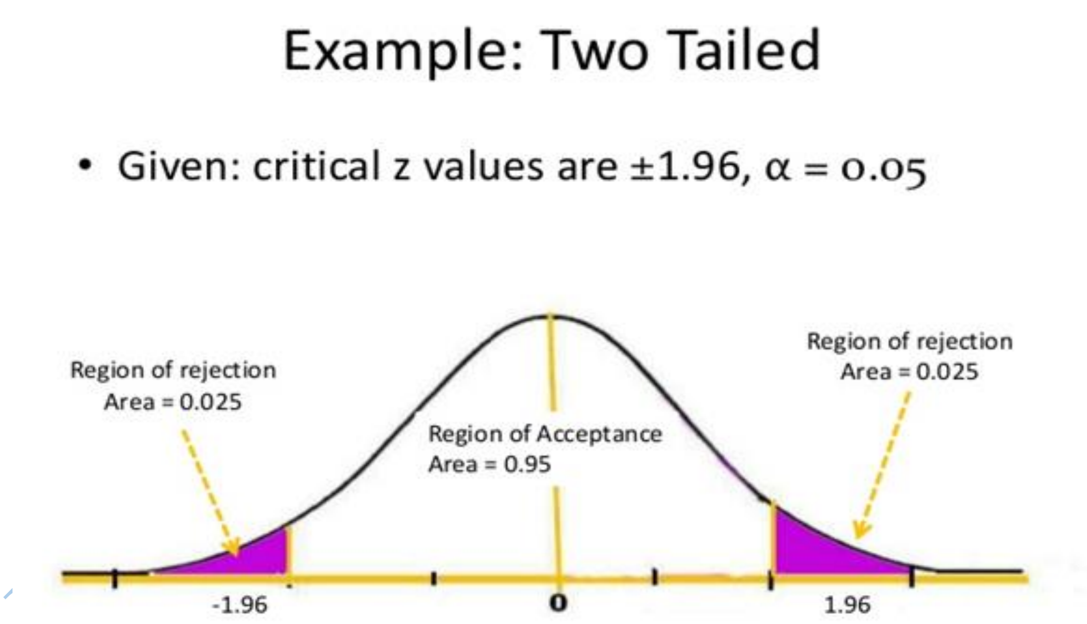
$$H_0 \quad \mu = \mu_0$$
$$H_1 \quad \mu \neq \mu_0$$

فاذا أخترنا $\alpha = .05$ كمستوى معنوية فمنطقة الرفض ستكون في هذه الحالة عن يمين ويسار القيمة μ_0، أي ذات طرفين لأن الفرضية البديلة $H_1: \ \mu \neq \mu_0$ يعني بأنه في حالة رفض فرضية العدم فأن μ قد تكون أقل أو أكثر من μ_0
ومنطقة الرفض هي

$$\mu_0 \pm Z_{\alpha/2}\, \sigma_{\bar{Y}}$$

حيث أن : $Z_{\alpha/2}$ هو المتغير القياسي الذي يقطع المنحني بـ $_{\alpha/2}$ في كلا الطرفين من التوزيع الطبيعي
وبما أن $\alpha = 0.05$ أذاً

$$Z_{\alpha/2} = Z_{.025} = 1.96$$



جدول مناطق الرفض عند اختبار H_0 : $\mu = \mu_0$ ضد ثلاث حالات من الفرضية البديلة H_1 وان:

عندما يكون حجم العينة كبير و σ^2 معلوم: $Y \sim N(\mu_1\ \sigma^2_y)$

حالات الاختبار	اتخاذ القرارات		
	$\alpha=\alpha$ ترفض H_0 إذا كانت قيمة	$\alpha=0.05$ ترفض H_0 اذا كانت قيمة	$\alpha=0.01$ ترفض H_0 اذا كانت قيمة
$H_0 : \mu=\mu_0$ $H_1: \mu\neq\mu_0$	$Z \geq Z_{\alpha/2}$ and $Z \leq -Z_{\alpha/2}$ or $\lvert Z\rvert \geq Z_{\alpha/2}$	$Z \geq 1.96$ and $Z \leq -1.96$ or $\lvert Z\rvert \geq 1.96$	$Z \geq 2.58$ and $Z \leq -2.58$ or $\lvert Z\rvert \geq$ 2.58
$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$Z \geq Z_\alpha$	$Z \geq 1.65$	$Z \geq 2.33$
$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$Z \leq Z_\alpha$	$Z \leq -1.65$	$Z \leq -2.33$

علما ان Z في هذه الحالة هو

$$\mathbf{Z}= \frac{\bar{Y}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$$

مثال: ينتج معمل لتعليب الزيت علبا متوسط وزنها 15 كغم بانحراف قياسي 0.5 كغم، أجري اختبار للتأكد من أن المعمل لازال ينتج عند المستوى المطلوب وبذلك أخذت عينة من 50 علبة ووجد أن متوسط وزنها 14.8 كغم فاذا كان وزن العلب متغير عشوائي يتوزع توزيعا طبيعيا فهل تدل العينة على أن أنتاج المعمل لازال 15 كغم، على مستوى معنوية $\alpha = 0.01$.

$$H_0: \mu=15$$
$$H_1: \mu\neq15$$

منطقة الرفض:

$$Z > 2.58$$
$$Z < -2.58$$

$$\mu_0=15$$
$$\bar{Y}= 14.8$$
$$\sigma= 0.5$$
$$n= 50$$

$$Z= \frac{\bar{Y}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$Z = \frac{14.8-15}{\frac{0.5}{\sqrt{50}}}$$

$$Z = \mathbf{-2.83}$$

القرار: بما أن قيمة Z المحسوبة 2.83 – هي أقل من 2.58 – أي أن قيمة Z المحسوبة تقع في منطقة الرفض لذا نرفض فرضية العدم وتقبل الفرضية البديلة أي أن المعمل لا ينتج علبا أوزانها 15 كغم.

او بما أن قيمة Z المحسوبة المطلقة 2.83 هي أكبر من 2.58 أي أن قيمة Z المحسوبة تقع في منطقة الرفض لذا نرفض فرضية العدم وتقبل الفرضية البديلة أي أن المعمل لا ينتج علبا أوزانها 15 كغم.

مثال: اذا كان معدل انتاج احد أصناف الحنطة هو 1600 كغم هـ$^{-1}$. وادعى أحد الباحثين انه استنبط سلالة من هذا الصنف يعطي انتاجا أكبر. ولاختبار صحة ادعاء الباحث اخذت عينة عشوائية مؤلفة من 81 نباتا ووجد ان متوسط انتاجها 1630 كغم هـ$^{-1}$ بانحراف قياسي =15 كغم. هل نتائج العينة تؤيد ادعاء الباحث عند مستوى معنوية = 0.01؟

الحل:

فرضية العدم $H0: \mu=1600$

الفرضية البديلة $H1: \mu>1600$

مستوى المعنوية $\alpha= 0.01$

منطقة الرفض $Z \geq 2.33$

المختبر الاحصائي $n=81$ $\bar{Y}=1630$ $\sigma=S=15$

$$Z= \frac{\bar{Y}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$Z= \frac{1630-1600}{\frac{15}{\sqrt{81}}}$$

$= 18$

القرار: بما ان قيمة Z المحسوبة (18) هي اكبر من قيمة Z الجدولية (2.33) لذا نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1

أي ان ادعاء الباحث كان صحيحا.

- **اختبار Z لمتوسطين حسابيين**

1- وضع فرضية العدم H_0

$$H_0 : \mu_1 - \mu_2 = d$$

علما أن d تمثل الفرق الافتراضي بين المتوسطين

2- تحديد الفرضية البديلة وتكون الفرضية البديلة أحد الاشكال الثلاثة الاتية

$H_1: \mu_1 - \mu_2 \neq d$

$H_1: \mu_1 - \mu_2 > d$

$H_1: \mu_1 - \mu_2 < d$

3- تحديد مستوى المعنوية α أما ان يكون 5% او 1%

4- تحديد منطقة الرفض:

ترفض فرضية العدم وتقبل الفرضية البديلة كما يلي

- في حالة كون الفرضية البديلة

$$H_1: \mu_1 - \mu_2 \neq d$$

تكون منطقة الرفض $|Z| \geq Z_{\alpha/2}$ (القيمة المطلقة لـ Z المحسوبة)

ترفض فرضية العدم وتقبل البديلة

- في حالة كون الفرضية البديلة

$$H_1: \mu_1 - \mu_2 > d$$

ترفض فرضية العدم وتقبل الفرضية البديلة عندما تكون قيمه Z المحسوبة أكبر من الجدولية

- في حالة كون الفرضية البديلة

$$H_1: \mu_1 - \mu_2 < d$$

ترفض فرضية العدم وتقبل الفرضية البديلة عندما تكون قيمه Z المحسوبة أقل من الجدولية

علما أن قيمة المختبر الاحصائي يحسب بالصيغة الرياضية:

$$Z = \frac{\bar{Y}1 - \bar{Y}2 - d}{\sqrt{\frac{S1^2}{n1} + \frac{S2^2}{n2}}}$$

مثال: في امتحان لمادة الاحصاء لمجموعتين من الطلبة، المجموعة الاولى من 50 طالبة متوسط درجاتهن في الامتحان 82 بانحراف قياسي (S) 8، والمجموعة الثانية من 75 طالب كان متوسط درجاتهم 76 بانحراف قياسي (S) 6 ، فهل يوجد فرق معنوي بين مستوى الطالبات والطلاب في مستوى معنوية 5%

الحل:

فرضية العدم $H_0 : \mu_1 - \mu_2 = 0$

الفرضية البديلة $H_1 : \mu_1 - \mu_2 \neq 0$

مستوى المعنوية 5% اذاً منطقة الرفض هي $|Z| \geq Z_{\alpha/2}$

$|Z| \geq 1.96$

$$Z = \frac{82 - 76 - 0}{\sqrt{\frac{64}{50} + \frac{36}{75}}}$$

$$Z = \frac{6}{\sqrt{1.28 + 0.48}}$$

$$= 4.52$$

القرار: بما أن قيمة Z المطلقة المحسوبة (4.52) أكبر من Z الجدولية (1.96) لذا ترفض فرضية العدم (H0) وتقبل البديلة (H1) أي يوجد فرق معنوي بين درجات الطلاب والطالبات تحت مستوى احتمال 5%.

2- اختبار توزيع t المعتدل: t-Distribution

وهو من التوزيعات الاحتمالية المستمرة المهمة والمستخدمة مع العينات صغيرة الحجم (أقل من 30 وأكبر من 4) وتحدد قيمة t المحسوبة وفق المعادلة التالية ونقارنها مع t الجدولية بدرجات حرية n-1

$$t = \frac{\bar{Y} - \mu}{\frac{S}{\sqrt{n}}}$$

مثال: لو أريد تقدير المتوسط الحسابي الحقيقي لوزن الدجاجة لمجتمع معين من الدجاج على أساس اختيار عينة عشوائية ممثلة لهذا المجتمع مكون من عشر دجاجات وكانت أوزان الدجاجات هي:

1.2, 0.8, 0.9, 1.1, 1.2 0.9, 1.0, 1.2, 0.8, 1.1 كغم، بانحراف قياسي (S) = 0.16،

فان اختبار فرضية العدم القائلة بأن المتوسط الحسابي الحقيقي لوزن الدجاجة لمجتمع معين من الدجاج هو 1.25 كغم يتم كما يلي:

$H_0 : \mu_x = 1.25$

$H_1 : \mu_x \neq 1.25$

المتوسط الحسابي للعينة = 1.02 كغم

$$t = \frac{\bar{Y} - \mu}{\frac{s}{\sqrt{n}}}$$

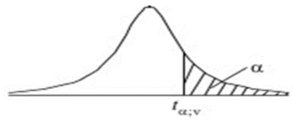
$$t = \frac{1.02 - 1.25}{\frac{0.16}{\sqrt{10}}}$$

$$= -4.510$$

القرار: بما أن قيمة t المحسوبة المطلقة (4.510) تزيد على قيمة t الجدولية (2.821) بمستوى معنوية 1% ودرجات حرية 9 لذا فأننا نرفض فرضية العدم بمستوى معنوية 1%.

Table of the Student's *t*-distribution

The table gives the values of $t_{\alpha;\nu}$ where
$\Pr(T_\nu > t_{\alpha;\nu}) = \alpha$, with ν degrees of freedom



ν \ α	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	6.314	12.076	31.821	63.657	318.310	636.620
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

المصادر:

1. الراوي، خاشع محمود. 1989. المدخل الى الإحصاء. وزارة التعليم العالي والبحث العلمي. جامعة الموصل.
2. طبيه، احمد عبدالسميع. 2007. مبادئ الإحصاء، عمان. دار البداية. ر.أ: (2007/6/1709) www.daralbedayah.com
3. David, M. Lane. Introduction to Statistics. Online Edition.
4. Internet