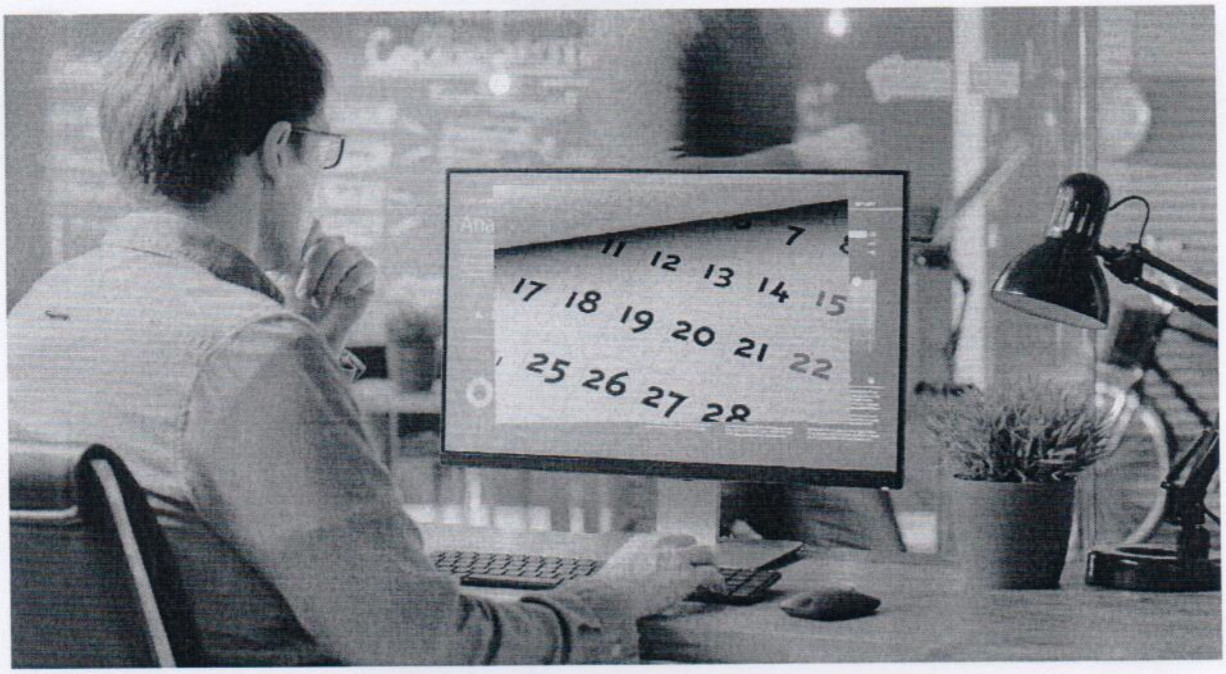
جامعة شط العرب
كلية الإدارة والاقتصاد



مبادئ الاحصاء

قسم إدارة الاعمال

المرحلة الاولى

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 كالآتي.

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 Scalcs

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 Variablel

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 Y_i 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**

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

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

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

ويرمز لمجموع حاصل ضرب قيم متغيرين y,x

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

ويرمز لحاصل ضرب مجموعتين لقيم متغيرين:

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

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

حيث C تمثل قيمة ثابتة

<u>مثال:</u> 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:

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

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

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

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

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

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

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

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

= (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:

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

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

***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

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

n: عدد المفردات

ب- طريقة Yule

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

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:

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

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

<u>الحدود الحقيقة للفئات:</u>
يمكن حساب الحدود الحقيقة بعدة طرق

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

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

او

3- إذا كانت البيانات لمتغير منفصل

الحد الأدنى الحقيقي = الحد الادنى للفئة - $\frac{1}{2}$

الحد الأعلى الحقيقي = الحد الاعلى للفئة + $\frac{1}{2}$

او

4- اذا كانت البيانات لمتغير مستمر

الحد الأدنى الحقيقي لأي فئة = الحد الادنى للفئة نفسها

الحد الاعلى الحقيقي لأي فئة = الحد الاعلى للفئة نفسها

<u>طول الفئة</u>

ويمكن حسابها بالشكل التالي:

أ- لبيانات متغير منفصل

طول الفئة =الحد الأعلى – الحد الأدنى + ١

ب – بيانات متغير متصل (مستمر) ⟵ وزن، طول، مسافة، سرعة....

طول الفئة =الحد الأعلى – الحد الأدنى

او طول الفئة =

الفرق بين الحد الأعلى لفئتين متتاليتين

الفرق بين الحد الادنى لفئتين متتاليتين

الفرق بين الحدين الحقيقي الأعلى او الأدنى لفئتين متتاليتين

الفرق بين مركزي فئتين متتاليتين

مثال: أكمل الجدول التالي

الفئات Classes	التكرار fi	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

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

وهكذا لبقية الفئات.
لاكمال الجدول كالآتي

Classes الفئات	التكرار fi	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 = $\frac{\text{تكرار تلك الفئة}}{\text{مجموع التكرارات}} = \frac{fi}{\sum fi}$

وقد يعبر عن التكرار النسبي كنسبة مئوية بالضرب في 100

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

مثال: أكمل الجدول

Classes الفئات	التكرار fi	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 = $\frac{\text{تكرار الفئة}}{\text{مجموع التكرارات}}$

$= \frac{1}{80}$

$= 0.0125$

التكرار النسبي المئوي للفئة الأولى = 0.0125×100

$= 1.25$

وهكذا لبقية الفئات

الفئات Classes	التكرار fi	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

وهكذا، 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=∑fi

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

LCF 2=∑fi-f1

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

LCF 3=∑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.

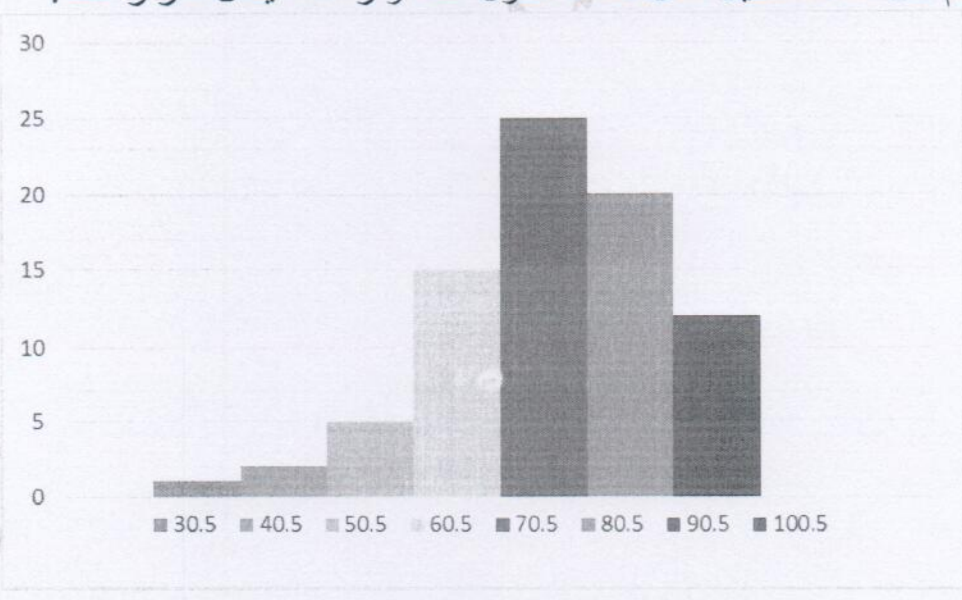
Classes الفئات	fi التكرار
31-40	1
41-50	2
51-60	5
61-70	15
71-80	25
81-90	20
91-100	12

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 متساوبة بحيث يشمل اعلى تكرار.
- يرسم لكل فئة مستطيل تمثل قاعدته طول الفئة وارتفاعه يمثل تكرار الفئة.

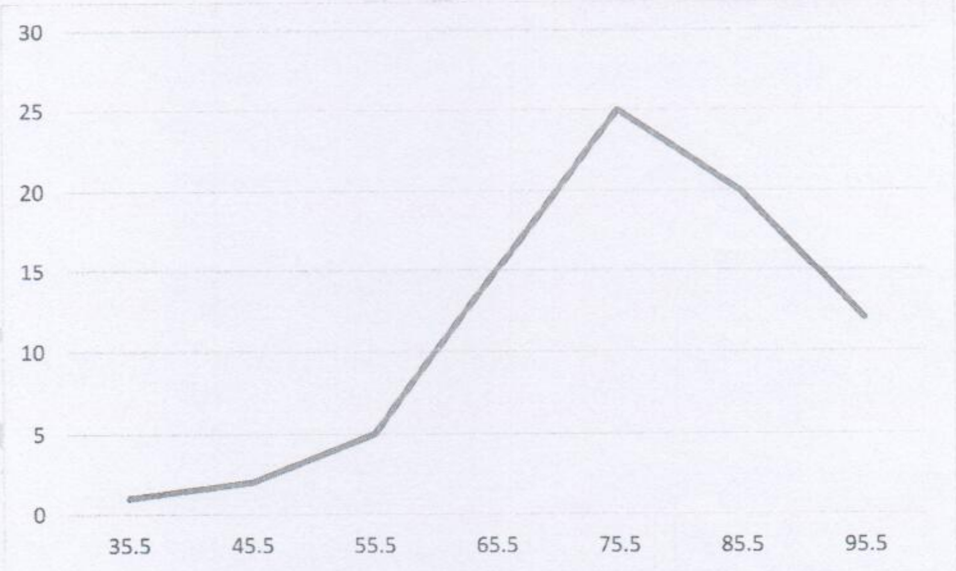


2. المضلع التكراري: Frequency Polygon

- إيجاد مراكز الفئات.

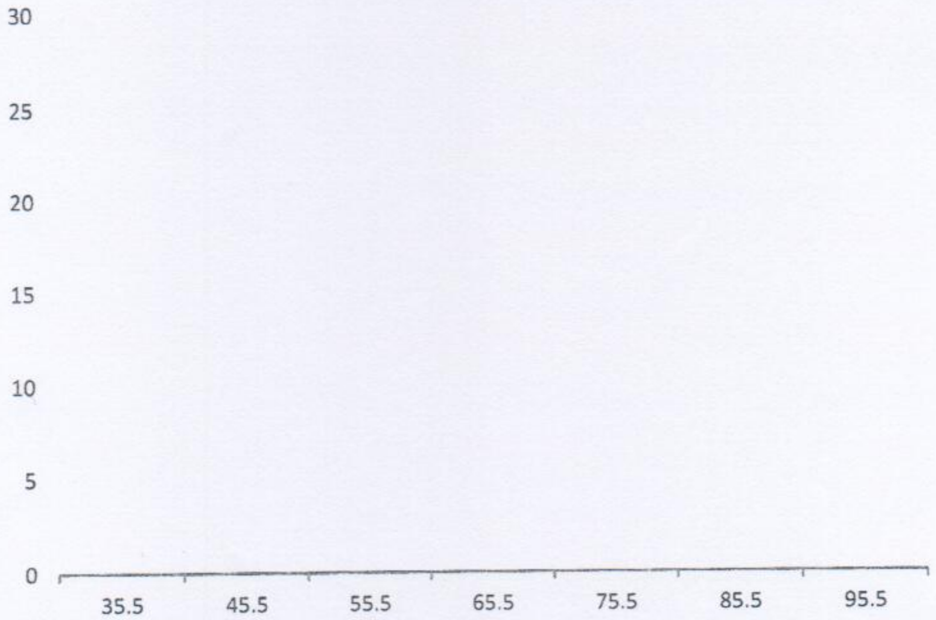
الفئات Classes	التكرار fi	مركز الفئة
31-40	1	35.5
41-50	2	45.5
51-60	5	55.5
61-70	15	65.5
71-80	25	75.5
81-90	20	85.5
91-100	12	95.5

- 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 متساويه 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 النزعة المركزية

وجدنا في الفصل السابق اهمية تنظيم البيانات الإحصائية وتلخيصها في جداول تكرارية وتمثيلها بيانيا وذلك في توضيح شكل التوزيعات التكرارية للبيانات الإحصائية وطبيعتها بشكل عام، ولكن الحاجة تدعو الى إيجاد مؤشرات تلخص البيانات بأقل قدر من التفصيل او نموذج يمثل المجموعة الإحصائية ومفرداتها او معيار تقارن بواسطته المجموعة كلها بالنسبة الى المجموعات الإحصائية الأخرى.

لقد وجد ان معظم القيم بمختلف الظواهر الطبيعية تتركز عامة في الوسط او قريبا منه، اذ يحدث في كثير من التوزيعات ان يتراكم عدد كبير من القيم حول قيمة معينة ويقل هذا التراكم تدريجيا كلما ابتعد المتغير عن هذه القيمة، هذا التجمع او التراكم حول قيمة معينة يسمى بالنزعة المركزية للتوزيع ونسبة القيمة التي يحدث حولها التراكم بمقياس النزعة المركزية، ومن اهم هذه المقاييس هي الوسط الحسابي، الوسط الهندسي، الوسط التوافقي، الوسيط، المنوال بالإضافة الى أوساط ومقاييس أخرى.

أولا: الوسط الحسابي $(\bar{X})$ The Arithmetic Mean

وهو من ابسط مقاييس النزعة المركزية واوسعها انتشارا من ناحية الاستخدام، ومن مميزاته:

- يمكن استخدامه في اغلب البيانات الإحصائية ولمختلف الظواهر.
- كونه يأخذ بنظر الاعتبار جميع القيم الإحصائية او البيانية.
- لا يحتاج لإيجاده الى تنظيم البيانات.
- دائما يكون مجموع انحرافات القيم عن وسطها الحسابي يساوي صفر.

الا ان من عيوبه انه يتأثر بالقيم الشاذة لأنه يأخذ جميع البيانات.

ويمكن إيجاد الوسط الحسابي بالصيغ التالية:

1- بالنسبة للبيانات غير المبوبة :

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

حيث أن :

xi = القيم المتغيرة.

n = عدد القيم.

مثال :

من البيانات الآتية استخرج الوسط الحسابي :

10 , 15 , 13 , 12 , 20 , 17 , 18 , 10

الحل :

$$\bar{x} = \frac{\sum xi}{n} = \frac{120}{8} = 15$$

مثال :

استخرج قيمة Z التي تجعل للقيم التالية وسطاً حسابياً مقداره (25) :

30 , 17 , 20 , 37 , Z

الحل :

n x = 25

$$\bar{x} = \frac{\sum xi}{n} \rightarrow 25 = \frac{\sum xi}{n} \rightarrow \sum xi = 125$$

$$\therefore Z = 125 - 104 = 21$$

مثال :

جد الوسط الحسابي للبيانات الآتية :

2 , 3 , 5 , 2 , 4 , 2 , 52

الحل :

$$\bar{x} = \frac{\sum xi}{n} = \frac{70}{7} = 10$$

من المثال أعلاه نلاحظ :

أ- أن قيمة الوسط الحسابي تساوي (10) مع أن معظم القيم كانت أقل من (5) وهذا ما يفسر لنا تأثير الوسط الحسابي في القيم الشاذة ، أي أنه تأثر بالقيمة الأخيرة (52).

ب- يمكن التأكد من صحة الحل عن طريق طرح القيم من الوسط الحسابي ، فإذا كان مجموعها يساوي صفر فهذا يعني أن الحل صحيح.

2- الوسط الحسابي بالنسبة للبيانات المبوبة :

ويمكن إيجاده عن طريق الصيغة الآتية :

$$\bar{x} = \frac{\sum fi\,xi}{\sum fi}$$

حيث أن :

xi = مراكز الفئات

fi = التكرارات

مثال :

للجدول التكراري الآتي استخرج الوسط الحسابي :

الفئات	التكرارات fi	مركز الفئة xi	fi xi
0 – 4	6	2	12
5 – 9	10	7	70
10 – 14	15	12	180
15 – 19	30	17	510
20 – 24	20	22	440
25 – 29	8	27	216
30 – 34	2	32	64
	91		$\sum 1492$

$$\bar{X} = \frac{\sum fi\,xi}{\sum fi} = \frac{1492}{91} = 16.395$$

ثانياً : الوسط الترجيحي (X_w) Weighted Average :

لاحظنا عند حساب الوسط الحسابي أننا تعاملنا مع كل قيم المتغير معاملة واحدة ولكن في بعض الأحيان تختلف قيم المتغير وفقاً لنسبتها في العينة أو المجتمع ، ولهذا يتم استخدام الوسط الترجيحي كوسط معبر عن هذه البيانات ، ويحسب من خلال الصيغة الآتية :

$$X_w = \frac{\sum ni\,xi}{\sum ni}$$

حيث أن :

ni = وزن المتغير

xi = قيم المتغير

مثال :

جد المعدل الطالب (س) :

ni xi	الدرجة xi	عدد الوحدات ni	المادة
210	70	3	رياضيات
250	50	5	كيمياء
240	60	4	إحصاء
160	80	2	فيزياء
90	90	1	ديمقراطية
$\sum 950$		$\sum 15$	

$X_w = \frac{\sum ni\, xi}{\sum ni} = \frac{950}{15}\ 63.33$

ثالثاً : الوسط الهندسي Geometric Mean G :

1- للبيانات غير المبوبة :

ويمكن إيجاده بالصيغة الأتية :

$$\text{Log G} = \frac{\log x1 + \log x2 + \log x3 \ldots \log xn}{n}$$

مثال :

جد الوسط الهندسي للبيانات الآتية :

12 – 45 – 50

الحل :

$$\text{Log G} = \frac{\log 12 + \log 45 + \log 50}{3}$$

$$\text{Log G} = \frac{4.43}{3} \rightarrow \text{Log G} = 1.477$$

$$G = 30$$

2- الوسط الهندسي للبيانات المبوبة :

ويحسب من الصيغة الآتية :

$$\text{Log } G = \frac{\sum fi \ Log \ xi}{\sum fi}$$

مثال :

استخدم التوزيع التكراري الذي يتضمن توزيع عدد من الطلبة في جامعة ما حسب أوزانهم بالكيلو غرام لإيجاد الوسط الهندسي :

الحل :

fi Log xi	Log xi	مركز الفئة xi	أعداد الطلبة fi	الفئات
8.891	1.778	61	5	60 – 62
32.511	1.806	64	18	63 – 65
76.696	1.826	67	42	66 – 68
49.817	1.845	70	27	69 – 71
14.906	1.863	73	8	72 – 74
$\sum = 182.822$			$\sum = 100$	

$$\text{Log } G = \text{Log } G = \frac{\sum fi \ Log \ xi}{\sum fi}$$

$$\text{Log } G = \frac{182.822}{100} = 1.8282$$

$$G = 67.3$$

ملاحظة :

1- يتضح مما سبق أن الوسط الهندسي لقيم مختلفة موجبة دائماً أصغر من الوسط الحسابي.

2- يكثر استخدام الوسط الهندسي في الأرقام القياسية للأسعار أو إيجاد متوسط لعدد من النسب أو في معدل المتغيرات في المبيعات أو السكان ... الخ.

3- لا يمكن إيجاد الوسط الهندسي إلا إذا كانت مجموعه القيم موجبة ، وأن تأثره بالقيم المتطرفة يكون أقل منه في الوسط الحسابي.

رابعاً : الوسط التوافقي Harmonic Mean H :

1- للبيانات غير المبوبة :

الوسط التوافقي هو مقلوب الوسط الحسابي :

$$\acute{X} = \frac{\sum xi}{n} \qquad H = \frac{n}{\sum \frac{1}{xi}}$$

ملاحظة : مقلوب القيم يعني واحد تقسيم القيمة = $\frac{1}{xi}$

مثال :

استخرج الوسط التوافقي للقيم التالية :

5 , 17 , 26 , 14 , 9 , 18 , 22

الحل :

xi	$\frac{1}{xi}$
5	$\frac{1}{5} = 0.2$
17	0.058
26	0.038
14	0.071
9	0.111
18	0.055
22	0.075
	$\sum = 0.578$

$\sum = 0.578$

$$H = \frac{n}{\sum \frac{1}{xi}} \rightarrow H = \frac{7}{0.578} = 12.11$$

2- الوسط التوافقي للبيانات المبوبة :

ويأخذ الصيغة التالية :

$$H = \frac{\sum fi}{\sum fi\left(\frac{1}{xi}\right)} \rightarrow H = \frac{\sum fi}{\sum \frac{fi}{xi}}$$

المثال السابق (أوزان الطلبة) :

fi / xi	مركز الفئة xi	أعداد الطلبة fi	الفئات
0.082	61	5	60 – 62
0.281	64	18	63 – 65
0.626	67	112	66 – 68
0.385	70	27	69 – 71
0.109	73	8	72 - 74
$\sum = 1.483$		100	

$$H = \frac{100}{1.483} = 67.43$$

يستخدم الوسط التوافقي في إيجاد المتوسطات للمعدلات الزمنية مثل إيجاد متوسط القراءة لمجموعة من الأفراد بدلالة عدد الكلمات في الدقيقة.

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

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

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

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

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

= 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

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

<u>مثال:</u> جد الوسيط لجدول التوزيع التكراري الآتي:

Classesالفئات	**fiالتكرار**
31-40	1
41-50	2
51-60	5
61-70	15
71-80	25
81-90	20
91-100	12

<u>الحل:</u>نجد التكرار التجميعي التصاعدي

Classesالفئات	**fiالتكرار**	**UCFالتكرار التجميعي التصاعدي**
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 و 41

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

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

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	التكرار fi
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

مقـــــــــــاييس التشتت او الاختـــــــلاف

Measures of Dispersion or Variation

مقدمـــــــــــــــة :

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

مثال: إذا كان لدينا مجموعتين من الطلاب، وكان درجات المجموعتين كالتالي:

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

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

(1) مقاييس التشتت : Dispersion Measurements

من هذه المقاييس : المدى ، والانحراف الربيعي ، والانحراف المتوسط ، والتباين ، والانحراف المعياري .

اولاً : المدى : Rang

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

المدى في حالة البيانات غير المبوبة = أكبر قيمة – أقل قيمة

$$Rang = Max - Min$$

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

المدى في حالة البيانات المبوبة = مركز الفئة الأخيرة – مركز الفئة الأولى

مثـال1 : لديك القيم التالية .

5.03 4.63 5.08 5.18 5.29 5.18 5.4 6.21 4.8

المطلوب : حساب المدى .

الحل : المدى = أكبر قيمة – أقل قيمة

أكبر قيمة = 6.21 أقل قيمة = 4.63

$$Rang = Max - Min = 6.21 - 4.63 = 1.58$$

مثال2 : الجدول التكراري التالي يبين الاجور اليومية بالدينار ل (60) عامل في احدى الشركات .

فئات الأجور	40-45	35-40	30-35	25-30	20-25	15-20
عدد العمال	3	12	18	15	9	3

المطلوب: حساب المدى .

الحل:

المدى = مركز الفئة الأخيرة – مركز الفئة الأولى

مركز الفئة الأخيرة: $\frac{(40+45)}{2} = \frac{85}{2} = 42.5$

مركز الفئة الأولى: $\frac{(15+20)}{2} = \frac{35}{2} = 17.5$

إذا $R = 42.5 - 17.5 = 25$

مزايا وعيوب المدى :

- من مزايا المدى

1- أنه بسيط و سهل الحساب.
2- يكثر استخدامه عند الإعلان عن حالات الطقس و المناخ الجوي مثل درجات الحرارة والرطوبة والضغط الجوي.
3- يستخدم في مراقبة الجودة .

- ومن عيوبه

1- أنه يعتمد على قيمتين فقط ، ولا يأخذ جميع القيم في الحسبان .
2- يتأثر بالقيم الشاذة .

ثانيا: الانحراف المتوسط Mean Deviation (MD)

هو أحد مقاييس التشتت ، ويعبر عنه بمتوسط الانحرافات المطلقة للقيم عن وسطها الحسابي ، فإذا كانت $x_1, x_2, \ldots, x_n$ هي القراءات التي تم أخذها عن ظاهرة معينة ، وكان ($\bar{x} = \sum x/n$) عبارة عن الوسط الحسابي لهذه القراءات ، فإن الانحراف المتوسط (MD) يحسب بتطبيق المعادلة التالية :

$$MD = \frac{\sum |x - \bar{x}|}{n}$$

وهذه الصيغة تستخدم في حالة البيانات غير المبوبة.

مثال1 : إذا كانت الطاقة التصديرية لخمس محطات لتحلية المياه بالمليون متر مكعب كما يلي :

4 5 2 10 7

المطلوب : أوجد قيمة الانحراف المتوسط للطاقة التصديرية .

الحل : لحساب قيمة الانحراف المتوسط يتم استخدام المعادلة :

- الوسط الحسابي :

$$\bar{x} = \frac{\sum x}{n} = \frac{28}{5} = 5.6$$

ويتم تكوين الجدول التالي :

الانحرافات المطلقة $\lvert x-5.6 \rvert$	الانحرافات $(x-\bar{x})=(x-5.6)$	الطاقة التصديرية x
1.6	4 - 5.6 = -1.6	4
0.6	5 - 5.6 = -0.6	5
3.6	2 - 5.6 = -3.6	2
4.4	10 - 5.6 = 4.4	10
1.4	7 - 5.6 = 1.4	7
11.6	**0**	**المجموع**

- إذا الانحراف المتوسط قيمته هي :

$$MD = \frac{\sum|x-\bar{x}|}{n} = \frac{11.6}{5} = 2.32$$

وفي حالة البيانات المبوبة ، يحسب الانحراف المتوسط باستخدام المعادلة التالية .

$$MD = \frac{\sum|x-\bar{x}|f}{\sum f_i}$$

حيث أن f هو تكرار الفئة و x هو مركز الفئة و $\bar{x}$ هو الوسط الحسابي .

<u>مثال2 :</u> يبين الجدول التكراري التالي توزيع 40 أسرة حسب الإنفاق الشهري بالألف ريال.

الإنفاق	2 – 5	5 - 8	8 – 11	11 – 14	14 – 17
عدد الأسرة f	1	8	13	10	8

أوجد الانحراف المتوسط .

<u>الحـــل :</u> لحساب الانحراف المتوسط ، يتم تطبيق المعادلة ، ويتبع الآتي :

- تكوين جدول لحساب مكونات المعادلة :

حدود الإنفاق	عدد الأسر f	مركز الفئة x	$x\,f$	$\lvert x-\bar{x}\rvert$	$\lvert x-\bar{x}\rvert f$
2 - 5	1	3.5	3.5	7.2	7.2
5 - 8	8	6.5	52	4.2	33.6
8 - 11	13	9.5	123.5	1.2	15.6
11 - 14	10	12.5	125	1.8	18
14 - 17	8	15.5	124	4.8	38.4
المجموع	**40**		**428**		**112.8**

نجد الوسط الحسابي هو :

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{428}{40} = 10.7$$

و نجد الانحراف المتوسط هو :

$$MD = \frac{\sum \lvert x-\bar{x}\rvert f}{\sum f} = \frac{112.8}{40} = 2.82$$

<u>مزايا وعيوب الانحراف المتوسط :</u>

- من مزايا الانحراف المتوسط أنه يأخذ كل القيم في الاعتبار، ولكن يعاب عليه ما يلي :

1-يتأثر بالقيم الشاذة .

2-يصعب التعامل معه رياضيا.

مقاييس التشتت والاختلاف | ٣- التباين Variance

التباين أكثر مقاييس التشتت استخداما وهو متوسط مربعات انحرافات القيم عن وسطها الحسابي ويرمز له (S^2)، وصيغته:-

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

حيث:

$\sum$ = مجموع

x = قيمة الظاهرة

$\bar{x}$ = متوسط الظاهرة

n = حجم العينة

- أي انه مجموع مربعات انحرافات القيم عن وسطها الحسابي مقسوما على حجم العينة ناقص واحد (درجات الحرية) ويستخدم لكي نحصل على قيمة تباين مقدرة غير متحيزة
- يستخدم التباين كثيراً في الاحصاء الاستدلالي
- يلاحظ ان وحدة قياس التباين هي مربع الانحرافات تفاديا للانحرافات المطلقة كما هو في الانحراف المتوسط، والهدف لكي لا يكون مجموع المربعات = صفر ونتخلص من هذه المشكلة بتربيع الانحرافات

مقاييس التشتت والاختلاف | ٣- التباين Variance

❖ التباين :-

صيغة اخرى للتباين بعد فك قوس مربع الانحرافات $(x-\bar{x})^2$ يصبح كالتالي:-

$$S^2 = \frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \quad ......(2)$$

حيث:

$\sum x^2$ = مجموع مربع (x)

$(\sum x)^2$ = مربع مجموع (x)

مقاييس التشتت والاختلاف | ٣- التباين Variance

❖ التباين :-

مثال:

احسب التباين للبيانات التالية: (12 , 10 , 11 , 7 , 5)

الحل:

١- اولا نحصل على الوسط الحسابي = $\bar{x} = \frac{\sum x_i}{n} = \frac{12+10+11+7+5}{5} = \frac{45}{5} = 9$

٢- ثانيا نحصل على الانحرافات عن الوسط $(x-\bar{x})$ ومربعها $(x-\bar{x})^2$ كالتالي:

X	$(x-\bar{x})=(x-9)$	$(x-\bar{x})^2=(x-9)^2$
12	3	9
10	1	1
11	2	4
7	-2	4
5	-4	16
$\sum$		34

مقاييس التشتت والاختلاف | ٣- التباين Variance

❖ التباين :-

٣- نحصل على التباين كالتالي:

X	$(x-\bar{x})=(x-9)$	$(x-\bar{x})^2=(x-9)^2$
12	3	9
10	1	1
11	2	4
7	-2	4
5	-4	16
$\sum$		34

التباين = $S^2 = \frac{\sum (x-\bar{x})^2}{n-1} = \frac{34}{4} = 8.5$

نلاحظ ان قيمة التباين (8.5) كبيرة نسبيا مقارنة بقيم الظاهرة، وهذا بسبب ان التباين هو مربع الانحرافات، ونظرا لان مقاييس التشتت لابد ان تاخذ نفس وحدات القيم الاصلية، فاننا ناخذ الجذر التربيعي وهذا يسمى بالانحراف المعياري الذي سناخذه لاحقا

مقاييس التشتت والاختلاف | ٣- التباين Variance

❖ التباين :-

باستخدام الصيغة الثانية:

$$S^2 = \frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \ldots\ldots(2)$$

X	x^2
12	144
10	100
11	121
7	49
5	25
$\sum x$ = 45	$\sum x^2$ = 439

التباين = $$S^2 = \frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) = \frac{1}{4}\left(439 - \frac{45^2}{5}\right) = \frac{1}{4}(34) = 8.5$$

مقاييس التشتت والاختلاف | ٣- التباين Variance

❖ التباين :- من البيانات المبوبة:-

التباين للبيانات المبوبة ياخذ الصيغة التالية:-

$$S^2 = \frac{\sum (x-\bar{x})^2 \times f}{\text{مج} - 1}$$

حيث:

x = مركز الفئة

$\bar{x}$ = الوسط الحسابي

f = التكرار

- أي انه مجموع حاصل ضرب مربعات انحرافات مراكز الفئات عن وسطها الحسابي مقسوما على حجم العينة ناقص واحد (درجات الحرية)

مقاييس التشتت والاختلاف | ٣- التباين Variance

❖التباين :- من البيانات المبوبة:-

مثال: احسب التباين للبيانات المبوبة التالية:-

الفئة	التكرار (f)
5 -	2
10 -	0
15 -	4
20 -	3
25 - 30	1

مقاييس التشتت والاختلاف | ٣- التباين Variance

❖التباين :- من البيانات المبوبة:-

الحل: علينا ان نحصل على البيانات التالية:-

- مركز الفئة = x
- الوسط الحسابي = $\overline{x}$
- انحرافات مراكز الفئة عن الوسط الحسابي = $(x-\overline{x})$
- مربع الانحرافات = $(x-\overline{x})^2$
- حاصل ضرب مربع الانحرافات والتكرار= $(x-\overline{x})^2 \times f$

وهذه الخطوات موضحة في الجدول التالي:-

مقاييس التشتت والاختلاف

٣- التباين Variance

❖ التباين :- من البيانات المبوبة:-

الحل:

$(x-\bar{x})^2 \times f$	$(x-\bar{x})^2$	$(x-\bar{x})$	$x \times f$	مركز الفئة x	التكرار f	الفئة
220.5	110.25	10.5-	15	7.5	2	5 -
0	30.5	- 5.5	0	12.5	0	10 -
1	0.25	- o.5	70	17.5	4	15 -
60.75	20.25	4.5	67.5	22.5	3	20 -
90.25	90.25	9.5	27.5	27.5	1	25 - 30
372.5			180		10	

الوسط الحسابي = $$\bar{x} = \frac{\sum x \times f}{\sum f} = \frac{180}{10} = 18$$

∴ التباين $$S^2 = \frac{\sum (x-\bar{x})^2 \times f}{n-1} = \frac{372.5}{9} = 41.38$$

مقاييس التشتت والاختلاف

٤- الانحراف المعياري Standard Deviation

الانحراف المعياري اهم مقاييس التشتت واكثرها استخداماً، ويرمز له (S): وهو عبارة عن الجذر التربيعي للتباين (S^2)، وصياغته:

$$\boxed{S = \sqrt{S^2} = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} \quad ....(1)}$$

حيث:

$\sum$ = مجموع S = الانحراف المعياري

x = قيمة الظاهرة S^2 = التباين

$\bar{x}$ = متوسط الظاهرة

n = عدد مفردات الظاهرة

وبالتالي فان الانحراف المعياري يقيس التشتت بنفس وحدات القيم الاصلية ، وعليه فهو يقيس متوسط انحرافات القيم عن الوسط الحسابي – او متوسط التشتت حول الوسط الحسابي

مقاييس التشتت والاختلاف | ٣- التباين Variance

❖ التباين :-

صيغة اخرى للتباين بعد فك قوس مربع الانحرافات $(x-\overline{x})^2$ يصبح كالتالي:-

$$S=\sqrt{S^2}=\sqrt{\frac{1}{n-1}\left(\sum x^2-\frac{(\sum x)^2}{n}\right)}.......(2)$$

حيث:

$\sum x^2$ = مجموع مربع (x)

$\left(\sum x\right)^2$ = مربع مجموع (x)

مقاييس التشتت والاختلاف | ٤- الانحراف المعياري Standard Deviation

❖ الانحراف المعياري (S.D):-

مثال: اوجد الانحراف المعياري للبيانات التالية: (32 , 36 , 28 , 44 , 40)

الحل:

- اولا نحصل على الوسط الحسابي = $\overline{x}=\frac{\sum x}{n}=\frac{40+44+28+36+32}{5}=36$

- ثم نحصل على التباين ومنه على الانحراف المعياري كالتالي:-

X	$(x-\overline{x})$	$(x-\overline{x})^2$
40	4	16
44	8	64
28	- 8	64
36	0	0
32	- 4	16
		160

$$S^2=\frac{\sum(x-\overline{x})^2}{n-1}=\frac{160}{4}=40$$

∴ الانحراف المعياري $S=\sqrt{S^2}=\sqrt{40}=6.32$

Standard Deviation ٤- الانحراف المعياري | مقاييس التشتت والاختلاف

❖الانحراف المعياري (S.D):-

مثال: اوجد الانحراف المعياري للبيانات التالية: (32 , 36 , 28 , 44 , 40)

الحل: باستخدام الصيغة (2)

$$S=\sqrt{S^2}=\sqrt{\frac{1}{n-1}\left(\sum x^2-\frac{(\sum x)^2}{n}\right)}.......(2)$$

- نحصل على مجموع (x)
- ونحصل على مربع (x) كالتالي:-

X	x^2
40	1600
44	1936
28	784
36	1296
32	1024
$\sum x$= 180	$\sum x^2$ = 6640

$$S=\sqrt{S^2}=\sqrt{\frac{1}{4}\left(6640-\frac{(180)^2}{5}\right)}$$

$$S=\sqrt{S^2}=\sqrt{\frac{1}{4}\left(6640-\frac{32400}{5}\right)}=\sqrt{\frac{1}{4}(6640-6480)}$$

∴ الانحراف المعياري $S==\sqrt{\frac{160}{4}}=\sqrt{40}=6.32$

Standard Deviation ٤- الانحراف المعياري | مقاييس التشتت والاختلاف

❖الانحراف المعياري (S.D):-

وتفسير ناتج الانحراف المعياري (٦،٣٢) هو ان البيانات تتشتت عن الوسط الحسابي بمعدل ٦،٣٢ وحدات

- وكلما كان الانحراف المعياري صغيرا كلما كان تشتت البيانات صغيرا أي ان القيم متقاربة من الوسط الحسابي

- وكلما كان الانحراف المعياري كبيرا كلما كان التشتت كبيرا ودل ذلك على ان القيم متباعدة عن الوسط الحسابي

مقاييس التشتت والاختلاف

٤- الانحراف المعياري Standard Deviation

الانحراف المعياري (S.D):- للبيانات المبوبة

الحل: علينا ان نوجد الاتي:

- مركز الفئة: (x)
- الوسط الحسابي: $(\overline{x})$
- الانحرافات عن الوسط الحسابي: $(x-\overline{x})$
- مربع الانحرافات: $(x-\overline{x})^2$
- ضرب مربعات الانحرافات في التكرارات : $(x-\overline{x})^2 \times f$

كما في الجدول التالي:-

مقاييس التشتت والاختلاف

٤- الانحراف المعياري Standard Deviation

❖الانحراف المعياري (S.D):- للبيانات المبوبة

فئات الاجور	عدد العمال (f)	(x)	$(x.f)$	$(x-\overline{x})$	$(x-\overline{x})^2$	$(x-\overline{x})^2 \times f$
3.50 – 3.59	1	3.55	3.55	- 0.39	0.15	0.15
3.60 -3.69	2	3.65	7.30	- 0.29	0.08	0.17
3.70 – 3.79	2	3.75	7.50	- 0.19	0.04	0.07
3.80 – 3.89	4	3.85	15.40	- 0.09	0.1	0.03
3.90 – 3.99	5	3.95	19.75	- 0.01	0.00	0.00
4.00 – 4.09	6	4.05	24.30	0.11	0.01	0.06
4.10 – 4.19	3	4.15	12.45	0.21	0.04	0.13
4.20 – 4.29	2	4.25	8.50	0.31	0.09	0.19
	f= n = 25		98.45			0.8

الوسط الحسابي: $\overline{x} = \dfrac{\sum x.f}{\sum f} = \dfrac{98.45}{25} = 3.95$

مقاييس التشتت والاختلاف

٤- الانحراف المعياري Standard Deviation

❖الانحراف المعياري (S.D):- للبيانات المبوبة

فئات الاجور	عدد العمال (f)	(x)	$(x.f)$	$(x-\bar{x})$	$(x-\bar{x})^2$	$(x-\bar{x})^2 \times f$
3.50 – 3.59	1	3.55	3.55	- 0.39	0.15	0.15
3.60 -3.69	2	3.65	7.30	- 0.29	0.08	0.17
3.70 – 3.79	2	3.75	7.50	- 0.19	0.04	0.07
3.80 – 3.89	4	3.85	15.40	- 0.09	0.1	0.03
3.90 – 3.99	5	3.95	19.75	- 0.01	0.00	0.00
4.00 – 4.09	6	4.05	24.30	0.11	0.01	0.06
4.10 – 4.19	3	4.15	12.45	0.21	0.04	0.13
4.20 – 4.29	2	4.25	8.50	0.31	0.09	0.19
	f= n = 25		98.45			0.8

∴ الانحراف المعياري:

$$s = \sqrt{\frac{\sum (x-\bar{x})^2 \times f}{\cancel{f}-1}} = \sqrt{\frac{0.8}{24}} = \sqrt{0.033} = 0.0011$$

مقاييس التشتت والاختلاف

٥- معامل الاختلاف Coefficient of Variation

❖معامل الاختلاف (C.V):-

هو مقياس يقيس درجة التشتت النسبي، ويتم حسابه من خلال نسبة تشتت القيم الى متوسطها

ونحصل عليه من خلال قسمة الانحراف المعياري على الوسط الحسابي كالتالي:-

معامل الاختلاف = $\boxed{C.V = \frac{S}{\bar{x}} \times 100}$

حيث:

الانحراف المعياري: S

الوسط الحسابي: $\bar{x}$

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

	الطول	كمية الحاصل
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

الأرتـــــــــــبـاط

(دراسة العلاقة بين متغيرين)

مقدمـــــــــة :

تناولنا في الفصول السابقة دراسة متغير واحد أو ظاهرة واحدة من حيث قياس وحساب متوسط هذه الظاهرة وكذلك حساب مقياس لتشتتها . ولكن في الحياة العملية كثيراً ما يحتاج الباحث لدراسة العلاقة بين ظاهرتين (أو متغيرين) لمعرفة مدى الارتباط بينهما ونوع هذا الارتباط. فقد يريد الباحث معرفة ما إذا كان هناك علاقة بين التدخين والإصابة بمرض في الرئة، أو بين درجة تعليم الشخص ومستوى دخله . أو بين الحالة التعليمية والحالة الاجتماعية للناخب . وكما نرى فإنه يمكن أن نذكر الكثير بين الأمثلة في مختلف المجالات بل قد يرغب الباحث في دراسة العلاقة بين أكثر من متغيرين في وقت واحد . فمثلاً قد يريد الباحث معرفة تأثير درجة التعليم ومستوى الدخل وحجم الأسرة على درجة الوعي السياسي للشخص . في هذا المثال يريد الباحث معرفة تأثير ثلاثة متغيرات على متغير رابع . وفي هذا السنة سوف نركز على دراسة العلاقة بين **متغيرين اثنين فقط** وهو ما يعرف **بالارتباط " البسيط "** Simple Correlation . بينما الحالات التي تتناول الدراسة فيها **أكثر من متغيرين** تعرف **بالارتباط المتعدد** Multiple Correlation وهي – كما ذكرنا – خارج نطاق دراستنا لهذه السنة .

أنواع العلاقة بين المتغيرين :

إذا كان المتغيران يتغيران معاً في الاتجاه نفسه بمعنى أنه إذا زاد أو نقص أحدهما ، زاد أو نقص الأخر ، فإن العلاقة بينهما تكون طردية والارتباط بينهما يكون موجباً . مثال ذلك العلاقة بين زيادة حجم الطبقة الوسطى في المجتمع وزيادة الاستقرار السياسي .

وإذا كان المتغيران يتغيران معاً ولكن في عكس الاتجاه بمعنى أنه إذا زاد أحدهما نقص الأخر، أو إذا نقص أحدهما زاد الأخر، فإن العلاقة بينهما تكون عكسية والارتباط بينهما يكون سالباً . مثال العلاقة بين تدني مستوى الفرد التعليمي ودرجة الوعي الاجتماعي .

وتختلف العلاقات بين الظواهر من حيث القوة . فقد تكون العلاقة قوية جداً (أو حتى تامة) ، وقد تكون متوسطة ، أو ضعيفة ، أو منعدمة تماماً . وفي هذا الفصل سوف نتناول بالتفصيل كيفية حساب الارتباط بين متغيرين سواء كان المتغيران كميين أو وصفيين (ترتيبيين أو اسميين)، أو أحدهما كمياً والأخر وصفياً.

شكل الانتشار : Scatter Diagram

هناك وسيلة مبدئية يعرف الباحث من خلالها نوع الارتباط بين المتغيرين وما إذا كان الارتباط قوياً وضعيفاً أو منعدماً ، وما إذا كانت العلاقة خطية أو غير خطية ، موجبة أو سالبة . هذه الوسيلة هي " شكل الانتشار " .

والمقصود بشكل الانتشار هو تمثيل قيم الظاهرتين بيانياً على المحورين ، المتغير الأول على المحور الأفقي ، والمتغير الثاني Y على المحور الرأسي ، حيث يتم تمثيل كل زوج Pair من القيم بنقطة ، فنحصل على شكل يمثل كيفية انتشار القيم على المستوى ، وهو الذي يسمى شكل الانتشار . وطريقة انتشار القيم تدل على وجود أو عدم وجود علاقة بين المتغيرين ومدى قوتها ونوعها . فإذا كانت تتوزع بشكل منتظم دل ذلك على وجود علاقة (يمكن استنتاجها) ، أما إذا كانت النقط مبعثرة ولا تنتشر حسب نظام معين دل ذلك على عدم وجود علاقة بين المتغيرين أو أن العلاقة بينهما ضعيفة . والأشكال التالية تظهر بعض أشكال الانتشار المعروفة :

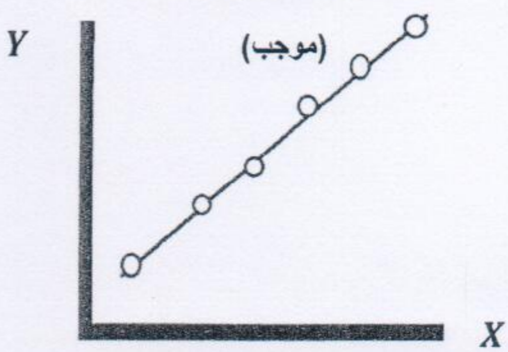
<u>الشكل الأول :</u>

إذا وقعت جميع النقاط على خط مستقيم ، دل ذلك على أن العلاقة بينهما خطية وأنها ثابتة أو تامة . وهذه تمثل أقوى أنواع الارتباط بين المتغيرين " ارتباط تام " . فإذا كانت العلاقة طردية فإن " الارتباط طردي تام " كما في الشكل الأول (أ) . ومثاله العلاقة بين الكمية المشتراة من سلعة والمبلغ المدفوع لشراء هذه الكمية .

أما إذا كانت العلاقة عكسية (وجميع النقاط تقع على خط مستقيم واحد) فإن " الارتباط عكسي تام " كما في الشكل الأول (ب) . ومثال على ذلك العلاقة بين السرعة والزمن .

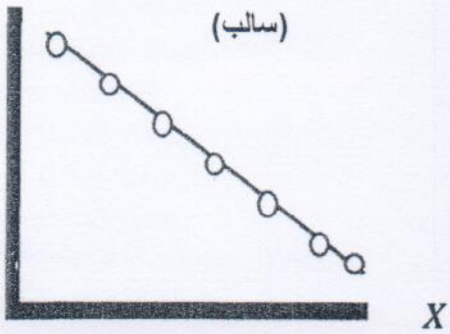
الشكل الأول (أ)

ارتباط طردي تام



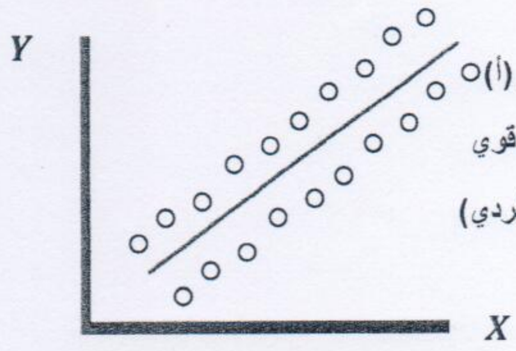
الشكل الأول (ب)

ارتباط عكسي تام

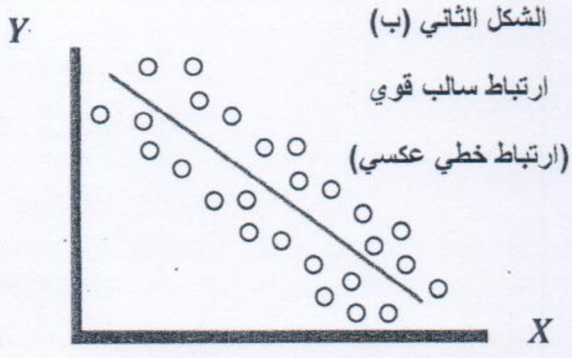


الشكل الثاني :

أما إذا كانت النقاط تأخذ شكل خط مستقيم ولكن لا تقع جميعها على الخط قيل أن العلاقة خطية (موجبة أو سالبة) كما في الشكل الثاني أ ، ب.



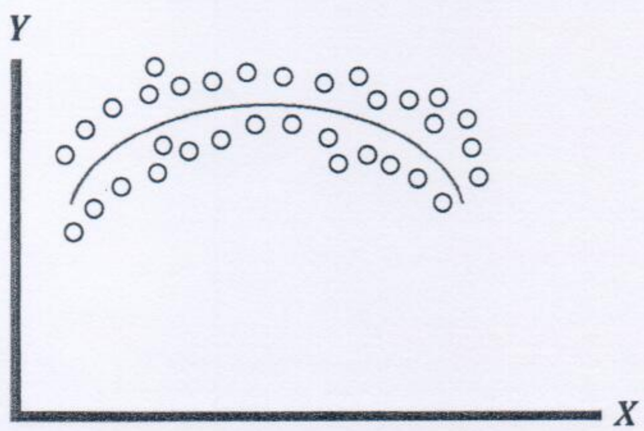
الشكل الثاني (أ)
ارتباط موجب قوي
(ارتباط خطي طردي)



الشكل الثاني (ب)
ارتباط سالب قوي
(ارتباط خطي عكسي)

الشكل الثالث :

وإذا كانت العلاقة تأخذ شكل منحنى فإن الارتباط لا يكون خطياً "ارتباط غير خطي " Non Linear Correlation كما في الشكل الثالث :

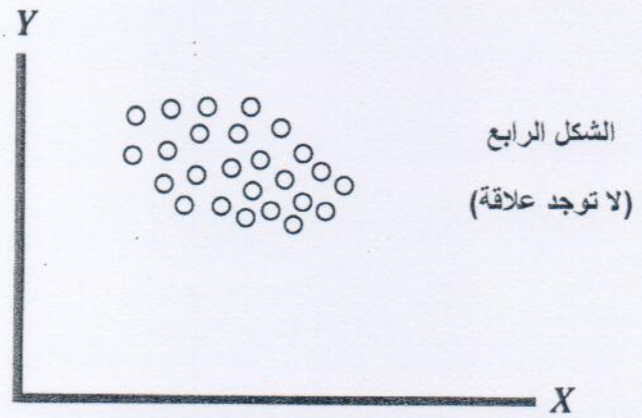


الشكل الثالث

(ارتباط غير خطى)

الشكل الرابع :

أما إذا كانت النقاط تتبعثر بدون نظام معين فإن ذلك يدل على عدم وجود علاقة بين المتغيرين (أو أن العلاقة بينهما ضعيفة جداً) كالعلاقة مثلاً بين دخل الشخص وطوله كما في الشكل الرابع :



معامل الارتباط : Correlation Coefficient

يقاس الارتباط بين متغرين بمقياس إحصائي يسمى " معامل الارتباط " ويعكس هذا المقياس درجة أو قوة العلاقة بين المتغيرين واتجاه هذه العلاقة . وتنحصر قيمة معامل الارتباط بين + 1، - 1. فإذا كانت قيمة معامل الارتباط تساوي + 1 فمعنى ذلك أن الارتباط بين المتغيرين طردي تام ، وهو أقوى أنواع الارتباط الطردي بين متغيرين . وإذا كانت قيمة معامل الارتباط تساوي - 1 فمعنى ذلك أن الارتباط بين المتغيرين عكسي تام ، وهو أقوى أنواع الارتباط العكسي بين متغيرين . وإذا كانت قيمة معامل الارتباط تساوي صفر ، فمعنى ذلك أنه لا يوجد ارتباط بين المتغيرين . وكلما اقتربت قيمة معامل الارتباط من + 1 أو - 1 كلما كان الارتباط قوياً ، وكلما اقترب من الصفر كلما كان الارتباط ضعيفاً .

معامل بيرسون للارتباط الخطي البسيط :

يفترض بيرسون Pearson أن المتغيرين كميان، وأن العلاقة بينهما خطية (أي تأخذ شكل خط مستقيم، أنظر الشكل الثاني من أشكال الانتشار) .

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
$\sum x_i= 156$	$\sum y_i= 170$	$\sum x_iy_i= 2710$	$\sum x_i^2= 2474$	$\sum y_i^2= 2982$

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

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

$$= 0.95$$

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	0.00
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

ملاحظات :

1- قوة معامل الارتباط تعتبر مؤشر مهم على وجود علاقة بين المتغيرين إلا أنه لا يمكن الاعتماد عليها على قوة معامل الارتباط فقط ، وإنما يجب أن يكون معنوياً وذلك من خلال مقارنة r المحتسبة مع r الجدولية في جدول القيم الحرية ، فإذا كانت r المحتسبة أكبر من r الجدولية كان الارتباط معنوي والعكس صحيح.

2- تختلف درجة قوة معامل الارتباط باختلاف العلوم ، فعلى سبيل المثال في العلوم الطبية لا يمكن اعتماد معامل ارتباط أقل من 90% ، أما في العلوم الزراعية فيمكن اعتماد 60% إلا أنه بصورة عامة إذا زادت قيمة r عن 70% يعتبر هناك ارتباط قوي بين المتغيرين.

ثانياً : معامل ارتباط الرتب (سبيرمان) :

وهو معامل ارتباط ثنائي يصلح في المتغيرات الكمية والنوعية وهو أقل من معامل ارتباط بيرسون ، ويمكن إيجاد معامل ارتباط سبيرمان حسب الخطوات الآتية :

1- إعطاء رموز رقمية للبيانات النوعية (الرتب) لكل من X و Y.

2- نستخرج الفرق بين رتب X ورتب Y بعمود جديد هو d.

3- نقوم بتربيع الفرق بعمود آخر هو d^2.

4- تطبيق قانون سبيرمان لإيجاد معامل الارتباط.

$$rs = 1 - \frac{6\sum d2}{n(n2-1)}$$

مثال :

كانت تقديرات مجموعة من الطلبة في مادتي التحليلات X والإحصاء Y كما يأتي :

∴ المطلوب : إيجاد معامل الارتباط بينهما والتعليق عليه.

تحليلات X	Y	رتب X	رتب Y	d	d^2
امتياز	جيد جداً	1	2	-1	1
جيد جداً	امتياز	2	1	1	1
جيد	متوسط	3,5	4,5	-1	1
جيد	جيد	3,5	3	0,5	0,25
متوسط	متوسط	5	4	0,5	0,25
					3,5

$$rs = 1 - \frac{6(3,5)}{5(25-1)} \rightarrow 1 - \frac{21}{5(24)}$$

$$rs = 1 - \frac{21}{120} \rightarrow 1 - 0,175 \rightarrow rs = 0,82$$

التعليق :

1- ان قيمة rs هي 0,82 وهذا يعني أن هناك علاقة ارتباط قوية بين تحصيل الطالب في التحليلات وتحصيله في الإحصاء.

مما سبق نستطيع إجمال بعضاً من الملاحظات فيما يلي :

1 – مجموع الفروق بين الرتب يساوي صفر .

2 – أن قيمة معامل ارتباط الرتب تنحصر بين – 1، + 1 فإذا كانت الرتبة رقم 1 للمتغير الأول تناظرها الرتبة 1 للمتغير الثاني، والرتبة 2 للمتغير الأول تناظرها الرتبة رقم 2 للمتغير الثاني، وهكذا.. فإن معامل ارتباط الرتب يساوي + 1 (ارتباط طردي تام بين الرتب) . وإذا كانت الرتبة رقم 1 (أقل رتبة) للمتغير الأول تناظرها أعلى رتبة للمتغير الثاني وهكذا.. فإن معامل ارتباط الرتب يساوي – 1 (ارتباط عكسي تام بين الرتب) .

3 – كذلك نلاحظ أن مجموع الرتب لكل من المتغيرين تساوي $\frac{n(n+1)}{2}$.

مثال2 : البيانات التالية تمثل إجابات عينة من سبعة أشخاص حول برامج الضمان الاجتماعي ، ومدى ملاءمتها لحاجات الناس .

السؤال الأول	جيدة	مقبولة	ممتازة	جيدة	جيدة جداً	مقبولة	جيدة
السؤال الثاني	جيدة جداً	مقبولة	جيدة جداً	جيدة	جيدة	جيدة	ممتازة

المطلوب : حساب معامل سبيرمان لارتباط الرتب بين هذين السؤالين ؟

الحل : تنظم الحل في الجدول التالي مع ملاحظة ما يلي :

1 – بالنسبة للسؤال الأول ، فإن التقدير الأعلى سيحصل على الرتبة رقم 1 والأقل منه مباشرة سيحصل على الرتبة رقم 2 وهكذا.. أي أن الترتيب تنازلي . ونكرر العمل نفسه مع السؤال الثاني .

2 – عند حصول إجابتين أو أكثر على التقدير نفسه نعطي لكل إجابة مبدئياً رتبة **كما لو كانوا مختلفين** ثم نحسب متوسط هذه الرتب ، وهذا المتوسط هو الذي يعطى لكل إجابة .

3 – ثم نحسب الفروق بين رتب السؤالين ونرمز لها بالرمز d ثم نربع هذه الفروق فنحصل على d^2 ونعوض في القانون عن $\sum d^2$ مع ملاحظة أن $n = 7$.

السؤال الأول X	السؤال الثاني Y	رتب X	رتب Y	الفرق بين الرتب d	مربعات الفرق d^2
جيدة	جيدة جداً	4	2.5	1.5	2.25
مقبولة	مقبولة	6.5	7	− 0.05	0.25
ممتازة	جيدة جداً	1	2.5	− 1.5	2.25
جيدة	جيدة	4	5	− 1.0	1.00
جيدة جداً	جيدة	2	5	− 3.0	9.00
مقبولة	جيدة	6.5	5	1.5	2.25
جيدة	ممتازة	4	1	3.0	9.00
المجموع				**Zero**	**26.0**

$$r_s = 1 - \frac{6(\sum d^2)}{n(n^2-1)} = 1 - \frac{6(26)}{7(49-1)} = 1 - \frac{156}{336} = 1 - 0.46 = 0.54$$

<u>مثال3 :</u> البيانات التالية تمثل أعداد الساعات التي ذاكرها عشرة طلاب والدرجات التي حصلوا عليها في امتحان أحد المقررات :

عدد الساعات X	10	6	12	14	11	6	19	16	3	9
الدرجات y	60	48	83	76	74	58	98	89	37	69

أحسب معامل سبيرمان لارتباط الرتب .

<u>الحل :</u> كما في المثال السابق ننظم الحل في الجدول التالي مع ملاحظة أن n = 10

عدد الساعات X	الدرجات Y	رتب X	رتب Y	الفروق d	d^2
10	60	6	7	− 1	1.00
6	48	8.5	9	− 0.5	0.05
12	83	4	3	1	1.00
14	76	3	4	− 1	1.00
11	74	5	5	0	0
6	58	8.5	8	0.5	0.25
19	98	1	1	0	0
16	89	2	2	0	0
3	37	10	10	0	0
9	69	7	6	1	1.00
المجموع				**Zero**	**4.50**

وبالتعويض في القانون حيث $\sum d^2 = 4.5$ ، $n = 10$ نحصل على :

$$r_s = 1 - \frac{6(\sum d^2)}{n(n^2-1)} = 1 - \frac{6(4.5)}{10(1-99)} = 1 - \frac{27}{990} = 1 - 0.027$$

$$r_s = 0.973$$

مما يعني أننا أمام علاقة طردية قوية بين المتغيرين . فكلما زادت عدد الساعات التي يدرسها الطالب في هذا المثال ، كلما زادت درجاته في الامتحانات قوة وذلك بما نسبته % 97 .

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) قوس (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

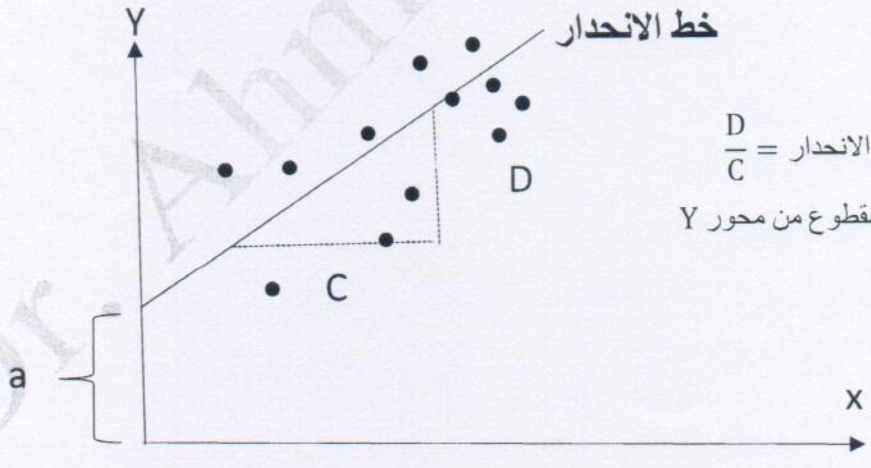
يعرف الثابت α بانه معدل قيمة y عندما x تساوي صفر وتسمى نقطة تقاطع خط الانحدار مع المحور الصادي (y- Intercept)

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



$b = \text{ميل خط الانحدار} = \frac{D}{C}$

$a = $ الجزء المقطوع من محور Y

كيفية حساب معامل الانحدار b

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

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

او يمكن استعمال المعادلة المختصرة التالية

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

كيفية حساب معامل تقاطع خط الانحدار مع المحور الصادي a:

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

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

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

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
$\sum x_i = 725$	$\sum y_i = 1011$	$\sum x_i^2 = 44475$	$\sum y_i^2 = 85905$	$\sum x_iy_i = 61685$

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

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

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

$= 84.250 - (0.897)(60.417) = 30.056$

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

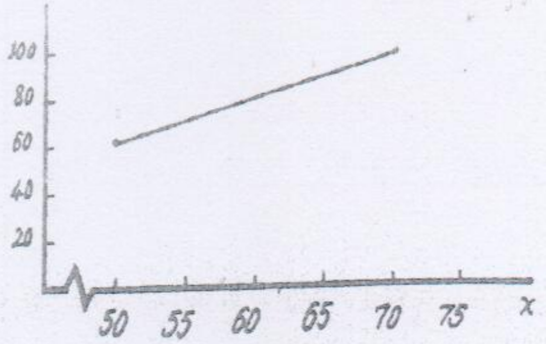
وبالتعويض بمعادلة خط الانحدار نحصل على:

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

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

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

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



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:

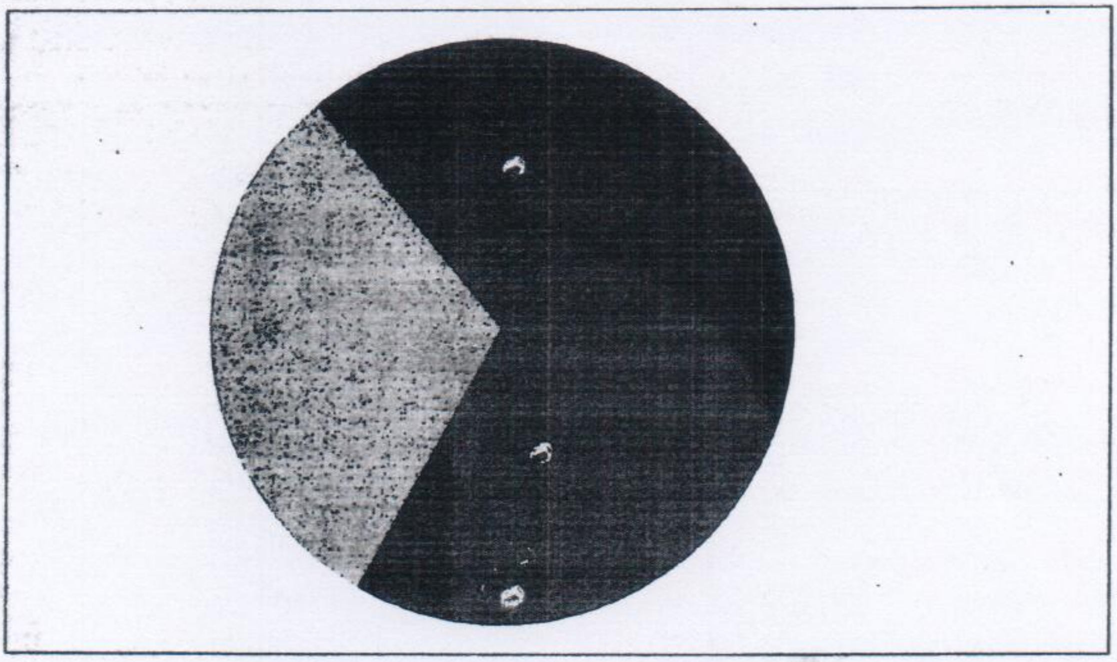
1. ان قيمة النقطة $(\bar{y}_1, \bar{x})$ تقع على خط الانحدار.
2. ان خط الانحدار يمر من جميع قيم $(x_i, y_{i.....})$
3. ان مجموع الانحرافات عن خط الانحدار يساوي صفراً اي $\sum(X_i - \bar{X}_x)^2 = 0$
4. وان مجموع مربعات الانحرافات عن خط الانحدار هي اقل ما يمكن اي

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

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

$$\hat{y} = 30.05 + 0.897(73) = 95.5$$



الانحدار الخطي المتعدد Multiple Linear Regression :

تناولنا في الانحدار الخطي البسيط العلاقة بين المتغير التابع (Y) ومتغير مستقل واحد هو (X) ، ولكن في الواقع أن نجد متغير تابع يعتمد على متغير مستقل واحد وأغلب المتغيرات في جميع العلوم ومنها الهندسية تعتمد على عدة متغيرات أخرى ، فالطلب مثلاً يعتمد على سعر السلعة ودخل المستهلك وأسعار السلع البديلة والمكملة ... الخ ، وان إنتاجية الموظف لا يعتمد على متغير التخصص فقط بل تعتمد على مستوى المهارة أو الخبرة أو التدريب ... الخ ، فلابد من معرفة أثر كل متغير مستقل في المتغير التابع وهذه المهمة يقوم بها الانحدار المتعدد بافتراض أن العلاقة خطية بين المتغيرات أو هي بيسا كذلك.

سنقوم بدراسة موضوع الانحدار الخطي المتعدد وباستخدام متغير تابع ومتغيرين مستقلين ، أما إذا زاد عدد المتغيرات المستقلة عن أثنين فيكون من الصعب إنجاز الانحدار بطريقة يدوية وإنما لابد من الاعتماد على البرامج الإحصائية في ذلك (SPSS) مثلاً.

ونأخذ معادلة الانحدار الخطي المتعدد الصيغة الآتية :

$$Y = \propto + B_1 X_1 + B_2 X_2 + B_3 X_3 \ldots B_n X_n$$

حيث تشير :

Y : إلى القيمة المحسوبة للمتغير التابع Y.

$\propto$: الحد المستقل أو هي قيمة Y عندما تكون كل من X_1 و X_2 تساوي صفر.

B_1 : وهي معامل X_1 وتمثل المتغير الناجم في Y نتيجة تغير القيمة بمقدار وحدة واحدة بافتراض ثبات قيمة X_2.

B_2 : وهي معامل X_2 وتمثل المتغير الناجم في Y نتيجة تغير X_2 بوحدة واحدة بافتراض ثبات قيمة X_1.

مثال :

توفرت لديك البيانات التالية عن سعر السلعة X_1 ودخل المستهلك X_2 والكمية المطلوبة Y ، المطلوب تقدير معادلة الانحدار الخطي المتعدد ومن ثم شرح المعادلة المقدرة :

X_1	X_2	Y	x_1	x_2	y	x_1^2	x_2^2	y^2	x_1x_2	x_1y	x_2y
5	2	3	2	-4	-5	4	16	25	-8	-10	20
4	3	4	1	-3	-4	1	9	16	-3	4	12
3	5	6	0	-1	-2	0	1	4	0	0	2
2	8	12	-1	2	4	1	4	16	-2	-4	8
1	12	15	-2	6	7	4	36	49	-12	-14	42
15	30	40	0	0	0	10	66	110	-25	-32	84

$X_1 = (XX), (XY)$

$XX = \begin{pmatrix} \sum x_1^2 & \sum x_{1x_2} \\ \sum x_{1x_2} & \sum x_2^2 \end{pmatrix}$

$XX = \begin{pmatrix} 10 & -25 \\ -25 & 66 \end{pmatrix}$

منكوس المصفوفة $= \frac{1}{D}(C)$

$D = 10(66) - (-25)(-25) = 660 - 625 = 35$

$C = \begin{pmatrix} 66 & 25 \\ 25 & 10 \end{pmatrix}$

$B_1 = \frac{1}{35} \frac{B_1}{B_2} \begin{pmatrix} 66 & 25 \\ 25 & 10 \end{pmatrix} \begin{pmatrix} -32 \\ 84 \end{pmatrix} \frac{\sum x_1 y}{\sum x_2 y}$

$B_1 = \frac{66(-32)+25(84)}{35} = \frac{-12}{35} = -0,34$

$B_2 = \frac{25(-32)+10(84)}{35} = \frac{40}{35} = 1,14$

$\propto = Y - B_1 X_1 - B_2 X_2 = 8 - (-0.34)(3) - 1,14(6)$

$\propto = 8 + 1,02 - 6,84 = 2,18$

$Y = 2,18 - 0,34 X_1 + 1,14 X_2$

شرح المعادلة :

1- قيمة $\propto = 2,18$ الحد الثابت وتعني أن الكمية المطلوبة تساوي 2,18 عندما يكون كل من سعر السلعة ودخل المستهلك يساوي صفر.

-2 قيمة B_1 = -0,34- حيث تدل الإشارة السالبة على العلاقة العكسية بين السعر والكمية المطلوبة وهذا يعني إذا ازداد السعر بمقدار وحدة واحدة فأن الكمية المطلوبة سوف تنخفض بمقدار 0,34 (مع استبعاد أثر X_2).

-3 قيمة B_2 = 1,14 حيث تدل الإشارة الموجبة على العلاقة الطردية بين دخل المستهلك والكمية المطلوبة ، وهذا يعني إذا ازداد دخل المستهلك بمقدار وحدة واحدة (مع استبعاد أثر X_1) فأن الكمية المطلوبة سوف تزداد بمقدار 1,14.

من المثال السابق جد القوة التفسيرية للمعادلة المقدرة مع التعليق :

$$R^2 = \frac{B_1 \sum x_1 y + B_2 \sum x_2 y}{\sum y^2}$$

$$R^2 = \frac{-0,34\,(-32) + 1,14\,(84)}{110}$$

$$R^2 = \frac{10,88 + 95,76}{110} = 0.97$$

وهذا يعني أن المعادلة المقدرة تستطيع أن تفسر 97% من التغيرات التي تحدث في الكمية المطلوبة بإعادتها إلى تغيرات في كل من السعر ودخل المستهلك و 3% تعود إلى عوامل عشوائية أخرى غير داخلة في المعادلة.

معامل التحديد المصحح $\bar{R}^2$ Adjusted R Square :

يأخذ $\bar{R}^2$ بنظر الاعتبار عدد المتغيرات المستقلة الداخلة في النموذجين ، أما في نموذج الانحدار المتعدد فتكون :

$$\bar{R}^2 = 1 - \frac{n-1}{n-k}\,(1 - R^2)$$

حيث أن k هو عدد المتغيرات المستقلة.

مثال :

افترض بأن لدينا معادلة انحدار خطي بسيط لقيمة من 20 مشاهد لثلاث متغيرات (متغيرين مستقلين ومتغير تابع) وأن معامل التحديد R^2 لها كان 0,85 أوجد قيمة $\bar{R}^2$:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k}\,(1 - R^2)$$

$$\bar{R}^2 = 1 - \frac{20-1}{20-1}\,(1 - 0,85) \qquad \bar{R}^2 = 0,85$$

مثال :

نفترض أن لدينا معادلة انحدار خطي متعدد لعينة من 20 مشاهد لثلاث متغيرات (متغيرين مستقلين ومتغير تابع) وأن معامل التحديد R^2 لها كان 0,85 أوجد قيمة $\bar{R}^2$.

$$\bar{R}^2 = 1 - \frac{20-1}{20-1}\,(1 - 0,85)$$

$$\bar{R}^2 = 1 - \frac{19}{18}\,(0,15) = 0,841$$

وهذا يدل على أن معامل التحديد المصحح أقل من معامل التحديد في حالة الانحدار المتعدد.

اختبار t لمعنوية المعالم المقدرة t test :

ويعتبر من أهم وأكثر الاختبارات الشائعة في تحليل الانحدار والذي يطلق عليه اختبار المعنوية الأساسي والذي عن طريقه يمكن الإجابة عن التساؤل التالي :

هل هناك تأثير معنوي للمتغير المستقل X على المتغير التابع Y.

ويمكن الإجابة عن هذا التساؤل بعد استخراج قيمة t المحتسبة وإخضاعها إلى اختبار الفرضيات وإجراء عملية المقارنة بين قيمة t المحتسبة وقيمة t الجدولية والتي يتم استخراجها من جدول القيم الحرجة عند مستوى معنوية معين ودرجة حرية $n - 2$ وهنا نحصل على نتيجتين يتم اختيار منها :

1- حالة رفض H_0 : إذا كانت قيمة t المحتسبة أكبر من قيمة t الجدولية فأنه سيتم رفض فرضية العدم H_0 وقبول الفرضية البديلة وهذا يعني معنوية المعلمة المقدرة وبالتالي هناك تأثير معنوي للمتغير المستقل X على المتغير التابع Y.

2- حالة قبول H_0 : إذا كانت قيمة t المحتسبة أكبر من قيمة t الجدولية فأنه سيتم قبول فرضية العدم ورفض الفرضية البديلة وهذا يعني عدم معنوية المعلمة المقدرة وبالتالي لا يوجد هناك تأثير معنوي للمتغير X على المتغير Y.

كيفية إيجاد قيمة t في الانحدار الخطي المتعدد :

$$tBi = \frac{Bi}{SBi}$$

يجري هذا الاختبار حسب الخطوات التالية :

1- استخراج الخطأ : $\sum e^2 = \sum y^2 - B_1\sum x_1 y + B_{2\sum x_2 y}$

2- استخراج تباين الخطأ : $Se^2 = \frac{\sum ei^2}{n-k-1}$

3- استخراج تباين $\bar{B}_i$: $SB^2_i = \frac{Se^2}{\sum x_i^2}$

4- استخراج الخطأ المعياري : $SB_i = \sqrt{SB\,i^2}$

5- احتساب t : $tB_i = \frac{Bi}{SB_i}$

وبالعودة إلى مثالنا السابق (أثر السعر والدخل على الكمية المطلوبة) يمكن اختبار معنوية المعالم المقدرة :

1- معنوية B_1 :

$$\sum e^2 = 110 - 106{,}64 = 3{,}36$$

$$Se^2 = \frac{3{,}36}{5-2-1} = 1{,}68$$

$$SB_1^2 = \frac{1{,}68}{10} = 0{,}168$$

$SB_1 = \sqrt{0{,}168} = 0{,}409$

$tB_1 = \frac{-0{,}34}{0{,}409} = -0{,}831$

الإشارة السالبة تهمل.

2- معنوية B_2 :

$\sum e^2 = 110 - 106{,}64 = 3{,}36$

$Se^2 = \frac{3{,}36}{5-2-1} = 1{,}68$

$SB_2^2 = \frac{1{,}68}{66} = 0{,}025$

$SB_2 = \sqrt{0{,}025} = 0{,}159$

$tB_2 = \frac{1{,}14}{0{,}159} = 7{,}16$