قسم علوم الطاقة المتجددة

Statistics

Introduction of statistic [definitions, descriptive, inferential]

Date [tables, diagram],

Measure of central tendency [Mean, Median, Mode, Harmonic Mean],

Measure of Dispersion [Rang, Mean derivation, Mean absolute deviation(Mean deviation)],

The semi-inter Quartile [rang deviation, variance standard deviation],

the standardized and the standard unit,

Moment correlation (pearson' s correlation)

regression

simple linear regression

Multiple linear regession.

Introduction of statistic

definitions, descriptive, inferential

مقدمة (Introduction) :

قد نشأ الإحصاء الرياضي من تقاطع الإحصاء والرياضيات، وهو وطيد الصلة بنظريتي الاحتمالات والعينات

ولذلك حرصنا عند إعداد هذه المحاضرات لتغطية مقرر (الاحصاء الرياضي), على أن تكون موضوعاتها مبسطة ومختصرة ومتفقة مع منهاج السنة الثانية والثالثة لطلاب كلية علوم الطاقة والبيئة قسم الطاقة المتجددة.

امثلة احصائية من القران الكريم

- (لقد احصاهم و عدهم عدا)... سورة مريم, الاية (94).
- (ليعلم ان قد ابلغوا رسالات ربهم و احاط بما لديهم و احصى كل شيء عددا)... سورة الجن, الاية (28).
- (واتاكم من كل ما سألتموه وان تعدوا نعمت الله لا تحصوها)... سورة إبراهيم, الاية (28).

الاحصاء(Statistic):

الإحصاء احد فروع الرياضيات الواسعة ذات التطبيقات الكثيرة ، يهتم علم الاحصاء بجمع و تلخيص و تمثيل و ايجاد الااستنتاجات من مجموعة البيانات المتوفرة ، محاولا التغلب على المشاكل مثل عدم تجانس البيانات و تباعدها.

كل هذا يجعله ذو اهمية تطبيقية واسعة في شتى مجالات العلوم من الفيزياء إلى العلوم الاجتماعية و حتى الانسانية ، كما يلعب دورا في السياسة و الأعمال.

انواع الاحصاء:

لقد ادى التطور السريع الذي مر به علم الإحصاء الى زيادة فروعه التطبيقية في مجالات الحياة كافة بهدف وصف البيانات و المساعدة على التوصل الى استنتاجات معينة فيها مما حدا ببعض الإحصائيين و منهم على سبيل المثال لا الحصر (فيركسون) الى ان يصنفوا الإحصاء تحت قسمين رئيسين هما:

1- الإحصاء الوصفي Descriptive Statistic:

يهدف إلى وصف مجموعة من البيانات وتنظيمها وتصنيفها وتلخيصها وعرضها بطريقة واضحة في صورة جداول أو أشكال بيانية وحساب المقاييس الإحصائية المختلفة لوصف متغير ما (أو أكثر) في مجتمع ما.

و يشمل الطرق الإحصائية المستعملة في وصف مجموعة معينة من البيانات.و تتضمن هذه الطرق الإحصائية أساليب

جمع البيانات (Collection of data) في

قياسات رقمية (Numerical Measurements).

ثم تبويبها او تنظيمها (Organizing)

وتلخيصها (Summarizing)

وعرضها (Presenting)

و حساب بعض المقاييس الإحصائية المختلفة لها فالإحصاء الوصفي يختص بطرائق جمع ووصف و تلخيص البيانات و ذلك بإستخدام الجداول التكرارية والرسومات البيانية و بعض المقاييس الإحصائية.

2- الاحصاء الاستدلالى inferential Statistic:

يعرف الإحصاء الاستدلالي أو الاستنتاجي Inferential or analysis بأنه النوع الثاني من أنواع علم الإحصاء، والذي يعنى بدراسة العلاقات بين المتغيرات المتعلقة بالدراسة الإحصائية، حيث يركز على الاستنتاجات الناتجة من الحسابات الرياضية الناتجة عن الإحصاء الوصفي (النوع الأول) ويعمل على تحليلها بقصد إطلاق التنبؤات والتعميمات وفقًا لما جاء فيها، كما يتم استخدامه كوسيلة للحكم على بعض البيانات غير المرئية، حيث يتم تحليلها واستخلاص النتائج منها.

و يضم هذا القسم فرعين رئيسيين هما:

1– التقدير Estimation:

و يهتم بايجاد قيم تقديرية للاستدلال منها على القيم الحقيقيه لمصدر جمع البيانات و هذه القيم التقديرية أما ان تكون تقديرا محدودا أي عند نقطة معينة (Estimation Point)

أو تقديرا في فترة أو مدى Interval Estimation).

2– اختبار الفرضيات Test of Hypotheses:

و يتضمن اختبار الفرضيات التي توضع كتفسير اولي للظاهرة المراد دراستها للوصول الى قرار بقولها او رفضها.

فالإحصاء الاستنتاجي او الاستدلالي و ان كان يختص باستنتاج و اتخاذ القرارات المناسبة للظاهرة قيد الدراسة مع حساب درجة الثقة المصاحبة لتلك القرارات و الاستنتاجات فانه يستخدم ايضا عند دراسة ظاهرة معينة يصعب الحصول على بيانات دقيقة عنها بما يغطي جوانب الدراسة كافة و يعزى ذلك الى سعة مجتمع

البحث و ارتفاع كلفة الدراسة, او عدم تطلب نتائج ذات دقة عالية جدا او ان تكون الدراسة تشمل حالات يستحيل قياسها مما يضطر الباحثون الى اخذ عينة للبحث تراعى في اختيارها الاسس والضوابط المعتمدة في انتخاب العينات حيث تتمكن من دراسة حالتها بدرجة مقبولة من الدقة, ثم نعمم نتائج الدراسة على مجتمع البحث كما ان هذا النوع من الإحصاء يهتم باساليب التحليل المعقدة التي تستدعي المقارنة و الاستنتاج للحكم على الظواهر.

كذلك يصنف الإحصاء الى

أ- الإحصاء المعلمي:

اذا كان مصطلح (المعلمية) قد اشتق من مفهوم (معلم) (Parameter) و الذي يمثل كل قيمة من القيم التي تتعلق بخصائص المجتمع, فان الاختبار المعلمي نوع من الإحصاء ذات طرق (معلمية) (Parametric) و هي طرق تتطلب الوفاء بافتراضات معينة حول المجتمع الذي تسحب منه العينة, و من هذه الفرضيات ان يكون توزيع المجتمع طبيعيا,

فالمعلمة هي صفة او خاصية من خصائص مجتمع معين, و من امثلة هذه الخواص الوسط الحسابي و الانحراف المعياري وشكل توزيع ذلك المجتمع في صفة معينة كالطول او العمر او الذكاء او التحصيل الدراسي او ما شابه ذلك و الطرق او الاختبارات الإحصائية المعلمية هي التي تعتمد على الافتراضات الخاصة بخصائص المجتمع, و من الامثلة لهذه الطرق ما يسمى بالاختبار التائي (T, test) لعينة واحدة, واختبار

(T) للفرق بين متوسطي عينتين واختبار (T) للمشاهدات المزدوجة وتحليل التباين لمعيار واحد, للمقارنات المستقلة, تحليل الاتجاهات, تحليل التباين لمعيارين, تحليل

التباين المشترك, تحليل الارتباط و الانحدار, و الارتباط الخطي البسيط, الارتباط الجزئي, تحليل الانحدار, التحليل العاملي.

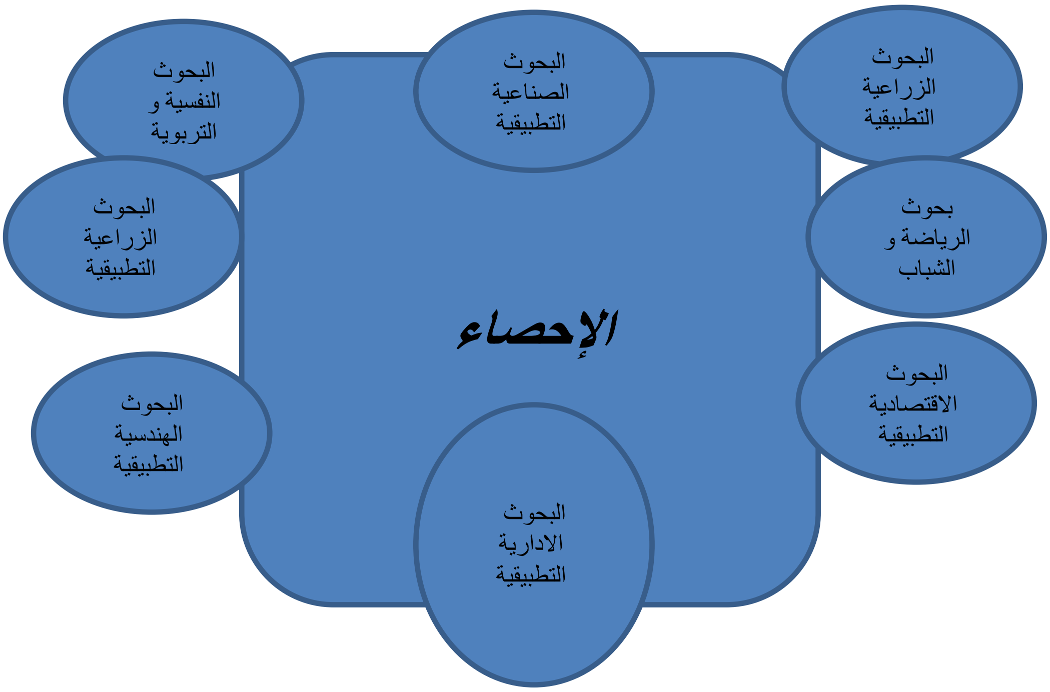
ب-الإحصاء اللامعلمي (Non Parametric Statistic):

يستخدم هذا النوع من الإحصاء طرقا لا معلمية و التي يسميها بعض الباحثين بالطرق (اللابارا مترية) كتعريف لمصطلح (Non Parametric), فالطرق الإحصائية اللامعلمية قد حظيت باهمية متميزة بسبب شيوع استخدامها في البحوث التربوية والنفسية والاجتماعية لملائمتها لطبيعة تلك البحوث والبيانات الخاصة بها من جهه, ولسهولة فهمها واستخدامها من جهة اخرى فالاختبارات اللا معلمية و ان كانت اختبارات لا تعتمد احصائية الاختبار فيها على معالم المجتمع كمعلمة المتوسط (Mean) او التباين (Variance),

فإنها لا تفترض توزيعا ما للبيانات و لهذا فهي تعرف ايضا باختبارات التوزيع الحر (Distribution-free tests) .

واذا كان سبب استعمال الاختبارات اللامعلمية يعود الى عدم توفر الفرضيات الخاصة بالاختبارات المعلمية.

شكل رقم (1) يمثل مجالات تطبيق علم الإحصاء



انواع البيانات

Date [tables, diagram]

1-البيانات النوعية Qualitative Data:

وتنقسم إلى نوعين هما

البيانات الإسمية (Nominal Data):

البيانات الإسمية تكون في صورة غير عددية أي لا يمكن قياسها وتتكون من فئات لا يتم التفاضل بينهما مثل الجنس يتكون من طبقتين ذكور ونرمز لهما الرقم (1) والإناث ونرمز لهما بالرقم (2) أو سؤال تكون إجابته "نعم " ونرمز له بالرقم (1) و" لا" ونرمز له بالرقم (0) والخطأ الذي يقع فيه الباحث هو إجراء عمليات حسابية على البيانات الاسمية.

البيانات الترتيبية (Ordinal Data):

أما البيانات الترتيبية هي أيضا تكون في صورة غير عددية ولا يمكن إجراء العمليات الحسابية عليها والفرق بينها وبين البيانات الاسمية هي عملية المفاضلة والترتيب بين طبقات المتغير مثل المستوى التعليمي ابتدائي (1)، إعدادي(2)، ثانوي (3)، جامعي فأكثر (4) في هذه الحالة المتوسط الحسابي أو أي عملية حسابية على تلك البيانات ليس لها معنى.

***تنحصر جميع التسميات السابقة بالبيانات المتقطعة اوغير مستمرة Discrete data**

2-البيانات الرقمية أو الكمية Quantitative Data

تنقسم أيضا إلى نوعين هما

البيانات الفترية (Interval Data):

البيانات الفترية تكون في صورة عددية ويمكن إجراء العمليات الحسابية عليها مثل المتوسط الحسابي والانحراف المعياري وغيرها ويمتاز هذا المقياس بتساوي المسافات بين الرتب حيث أنه يسمي أحيانا " بمقياس المسافة " ، ويستخدم هذا المقياس كثيراً في العلوم التربوية والنفسية والاجتماعية مثل قياس الذكاء وغيرها ، والجدير بالذكر أن هذا المقياس لا يعني الصفر فيه عدم وجود الخاصية فدرجة طالب تساوي صفر مثلا لا يعني أنه لا يعرف شيء في المقرر.

البيانات النسبية (Rational Data):

البيانات النسبية تكون أعلى مستوي من أنواع البيانات السابقة حيث يمتاز بالتصنيف (بيانات إسمية) و الترتيب (بيانات ترتيبية) والمسافات المتساوية (بيانات فترية) وخاصية النسبية والتي يعني إن للصفر خاصية العدم أي خاصية انعدام الظاهرة مثل سرعة سيارة تساوي صفر تعني أن السيارة متوقفة ، أو أن وزن شخص يساوي 60 كيلو جرام هو ضعف وزن شخص وزنه 30 كيلو جرام.

***تنحصر جميع التسميات السابقة بالبيانات المستمرة Continue data**

وظائف الإحصاء والفئات المهتمة بدراسته:

للإحصاء كأي علم من العلوم الاخرى وظائف ومهام يمكن استثمارها لبلوغ الأهداف المطلوبة بأيسر الطرق وأدقها وأقلها كلفة من الجهد و الوقت والموارد المتاحة الأخرى, و على هذا الأساس يمكن إجمال بعض وظائفه بما يلي:-

1. مساعدة الباحثيين على اعطاء أوصاف على جانب كبير من الدقة العلمية.
2. المساعدة على تلخيص النتائج بشكل ملائم ومفهوم.
3. مساعدة الباحثين على استخلاص النتائج العامه من النتائج الجزئية.
4. تمكن الباحث من التنبؤ بالنتائج التي يحتمل ان يحصل عليها من الظروف الخاصة.
5. ييسر على الباحث تنظيم خطوات بحثه فهو يحتاج للوسائل الاحصائية في مرحلة تصميم البحث وتخطيطه, حتى يمكنه في النهايه ان يخرج من بحثه بالنتائج التي يسعى الى تحقيقها.

<u>ألفئات التي تهتم بدراسة الإحصاء</u>

هناك فئات عديدة تهتم بدراسة الإحصاء منها

1- القائمون على البحث العلمي و منهم الباحثون في العلوم التربوية والنفسية, فهم بحاجة إلى جمع المعلومات عن حالة الأشخاص و سلوكهم و عاداتهم المختلفة.

2- الباحثون في أمور الدعاية و الإعلام و تسويق البضائع المختلفة فمن دون عمل إحصاءات عن الوضع الاجتماعي و الاقتصادي و النظر في رغبات الناس, فلن يكون بمقدورهم الحصول على تسويق ناجح للمنتوجات المختلفة.

3- المالكون لشركات التأمين التي تحتاج الى جمع معلومات وبيانات عن الحوادث المختلفة لمعرفة و امكانية ادراجها على قائمة نشاطاتها.

4- اصحاب المصانع المنتجة للملابس و الاغذية, و التي تنتج قياسات و احجاما مختلفة تتناسب و حاجات الناس و رغباتهم.

5- المؤسسات الحكومية المختلفة, مثل: وزارة الصحة و وزارة التخطيط المهتمة بحصر وجمع المعلومات عن الاشخاص في البلاد والتي على اساسها يتم وضع الخطط المناسبة من اجل تطوير البلاد في النواحي المختلفة و أغلب الوزارات الاخرى.

المجتمع الاحصائي (Statistic population):

يعرف المجتمع من الناحية الإحصائية بأنه جميع الافراد او العناصر التي تشترك في صفة واحدة او اكثر تميزه عن بقية المجتمعات, فقد يشكل جميع العاملين في المدارس التعليم الخاص في إحدى المدن مجتمعا إحصائيا او جميع الطلبة في احدى الكليات يشكلون مجتمعا إحصائيا ... فإذا حددنا المجتمع الخاص تعتبر مجتمعا إحصائيا.

العينة (sample):

تعرف العينة بانها ذلك الجزء من المجتمع الذي تتمثل فيها كافة خصائص المجتمع الاصلي و عادة تؤخذ لعدم القدرة على تناول كافة مفردات المجتمع لاعتبارات مادية و فنية و ما تتطلب دراسته و قتا و جهدا, وان الهدف من اخذ عينة من مجتمع هو قياس عناصرها او افرادها بانسبة لمتغير معين او مجموعة متغيرات و ذلك لكي تعمم النتائج المستخلصة منها عموم المجتمع الذي اختيرت منه.

انواع العينات

أ- **العينة العشوائية البسيطة:** هي عينة قائمة على الصدفة, و هي أبسط أنواع العينات رغم انها تتبع خطوات معروفة.

ب- **العينة العشوائية المنتظمة:** و فيه نختار العينة عن طريق اختيار المفردات من مسافات متساوية على القائمة بعد اعداد إطار المجتمع الاصلي

المدرج التكراري (Frequency Histogram):

المدرج التكراري هو مجموعة من الاعمدة يكون لكل عمود قاعدة مقدارها فئة محددة و يكون ارتفاعه عدد الحالات او التكرارات الخاصة بتلك الفئة و تحدد الفئات على المحور الافقي (X) في حين تحدد التكرارات على المحور العمودي (y) و كمثال

Example:

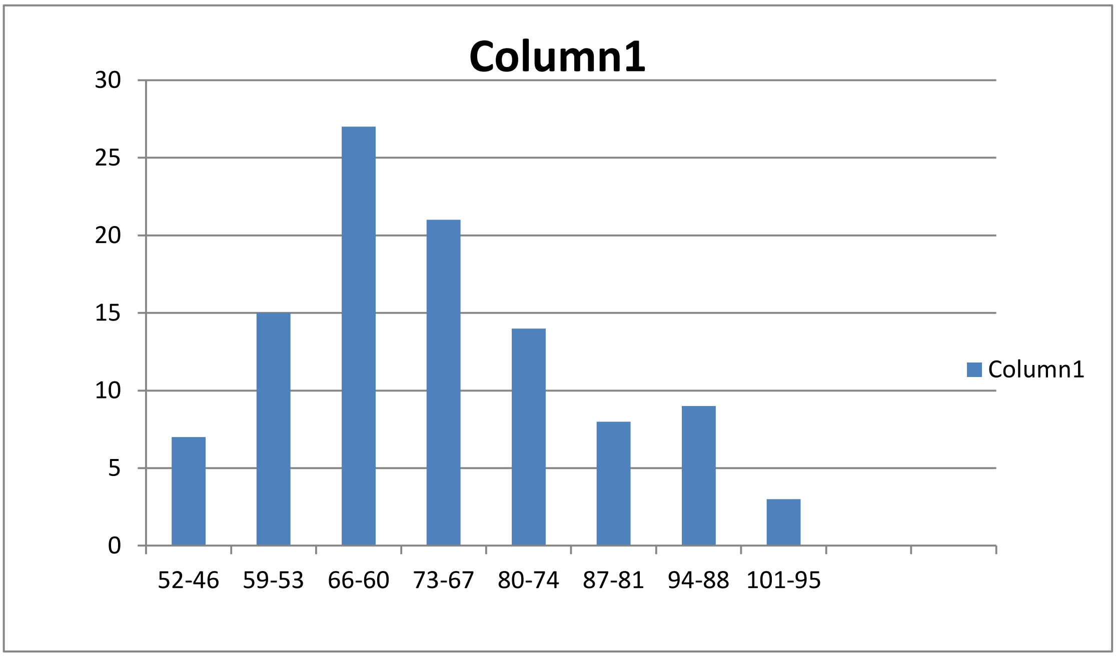
Draw a Frequency Histogram for the following data.

Class	Frequency
46-52	7
53-59	15
60-66	27
67-73	21
74-80	14
81-87	8
88-94	9
95-101	3

و لكي نرسم مدرج تكراري للبيانات الخاصة بالتوزيع التكراري الموضح في الجدول فاننا نتبع الخطوات التالية:

أ- نرسم المحورين الافقي (X) و العمودي(y).

ب- نقسم المحور الافقي (X) الى اجزاء متساوية ليمثل كل جزء منها طول الفئة



Example:

Draw a Frequency Histogram for the following data.

Class	Frequency
0−7	7
7−11	15
11−19	24
19−25	2
25−32	10
32−38	39

2.الاشرطة البيانية (Bar–Charts):

و هي عبارة عن مجموعة من المستطيلات الافقية او العمودية قواعدها متساوية و تمثل الصفة التي تم على اساسها التبويب (سنة, محافظة, و هكذا) و ارتفاعاتها تمثل البيانات البيانية المقابلة لتلك الصفة و الاشرطة البيانية على نوعين هما:

1) الاشرطة البيانية المفردة

و هي اشرطة بيانية تخص صنف واحد للبيانات مثل عدد الطلبة المقبولين او تطور عدد سكان العراق حسب التعداد السكاني

Example:

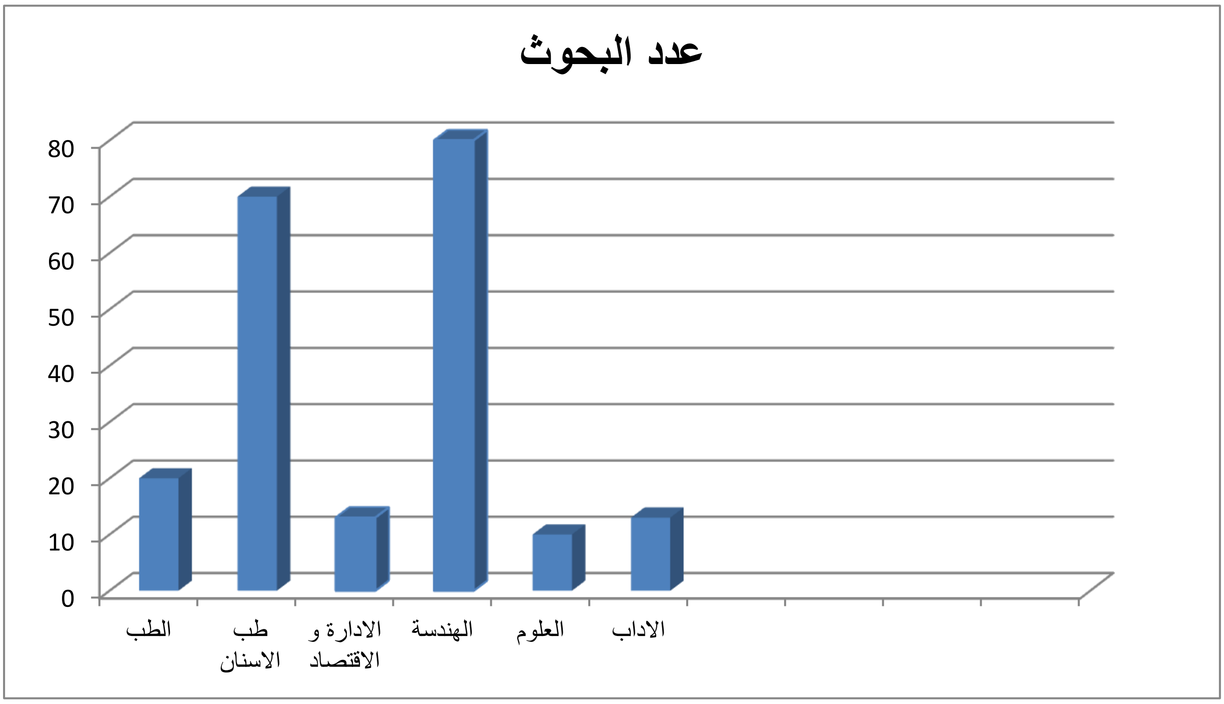
البيانات الواردة في الجدول التالي تمثل عدد البحوث العلمية المنجزة من قبل اعضاء هيئة التدريس في الجامعه المستنصرية موزعين حسب كلياتهم لعام 2006

الكليات	***عدد البحوث***
الطب	20
طب الاسنان	70
الادارة و الاقتصاد	15
الهندسة	80
العلوم	10
الاداب	12

الحل:-

يتم رسم الاشرطة البيانية بعد وضع المحور الافقي (السيني) X و الذي يمثل الكليات في هذا المثال و المحور العمودي (الصادي) و الذي يمثل عدد البحوث بعد ان يؤخذ تقسيم مناسب للمحور (Y) كما يأتي:

عدد البحوث



المضلع التكراري (Frequency polygon):

عبارة عن عدد من المستقيمات المتصلة مع بعضها على شكل سلسلة بحيث ان نقطة اتصال المستقيم بالارض تمثل مركز الفئة هذه يعني عند استخدام هذه النوع من التمثيل البياني يستوجب ايجاد مراكز الفئات.

Example:

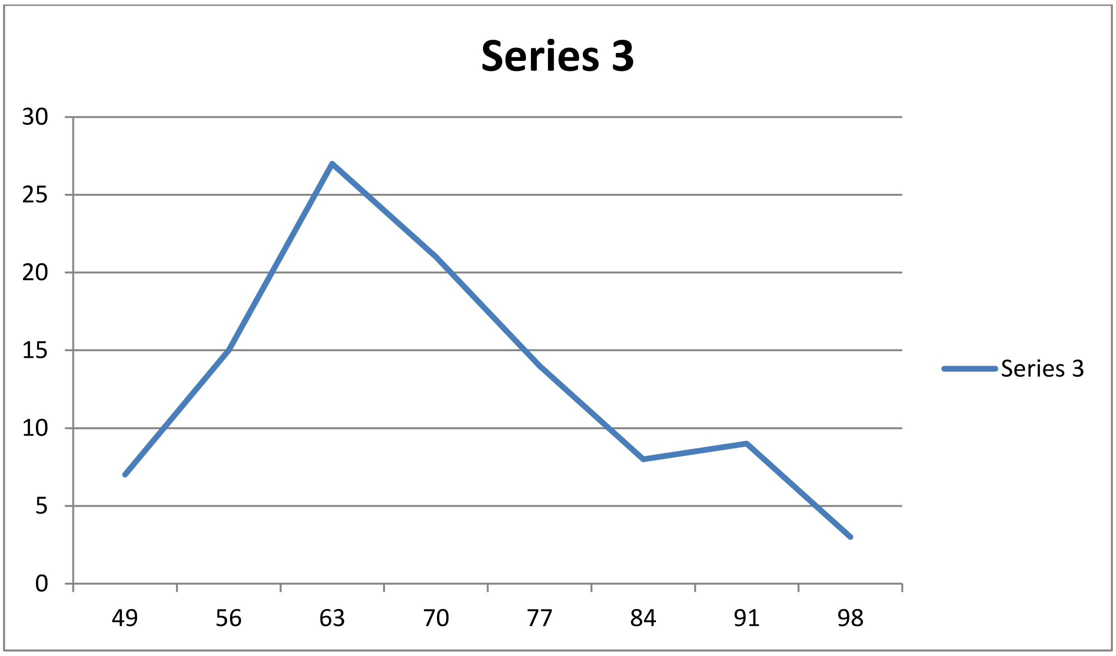
Draw a Frequency polygon for the following data.

Class	Frequency
46−52	7
53−59	15
60−66	27
67−73	21
74−80	14
81−87	8
88−94	9
95−101	3

Solution.

$$\boldsymbol{Class\ mark} = \frac{46+52}{2} = 49$$

Class	Frequency	*Class mark*
46−52	7	49
53−59	15	56
60−66	27	63
67−73	21	70
74−80	14	77
81−87	8	84
88−94	9	91
95−101	3	98



Example:

Draw a Frequency polygon for the following data.

Class	Frequency
0−7	7
7−11	15
11−19	24
19−25	2
25−32	10
32−38	39

المنحني التكراري (Frequency curve):

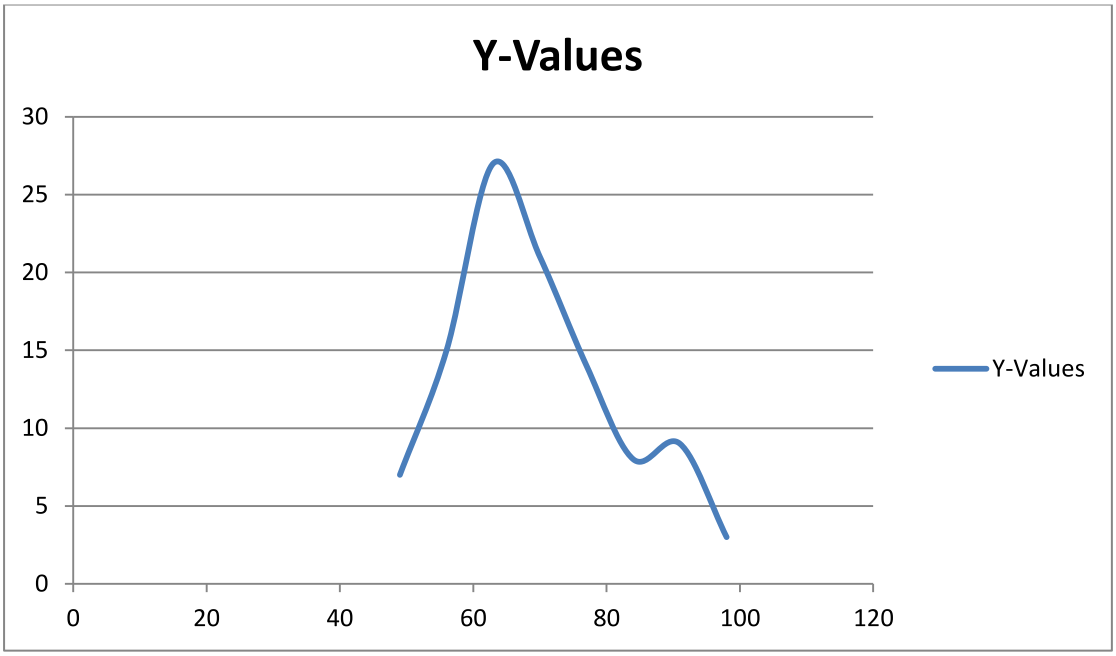
هوخط منحني يصل بين النقاط الناتجة عن مراكز الفئات والتكرارات التابعة لها ويستخدم هذه النوع من التمثيل في حالة التوزيعات المستمرة فقط.

Example:

Draw a Frequency curve for the following data.

$$\textit{Class mark} = \frac{46+52}{2} = 49$$

Class	Frequency	*Class mark*
46−52	7	49
53−59	15	56
60−66	27	63
67−73	21	70
74−80	14	77
81−87	8	84
88−94	9	91
95−101	3	98



<u>الدائرة البيانية (Pie- chart)</u>:

هي عبارة عن شكل هندسي مثل المستطيل البياني و لكن يتم هنا تمثيل البيانات بقطاعات داخل دائرة بحيث ان مجموع هذه القطاعات تمثل مساحة الدائرة الكلية و يتم تحديد زاوية القطاع وفق الصيغة التالية:

زاوية القطاع = (البيانات الجزئية/البيانات الكلية) * 360°

Example:

Draw a Pie- chart for the following data.

مصاريف مباشرة	مصاريف غير مباشرة	الاجور	المواد الخام
90	70	80	60

Solution.

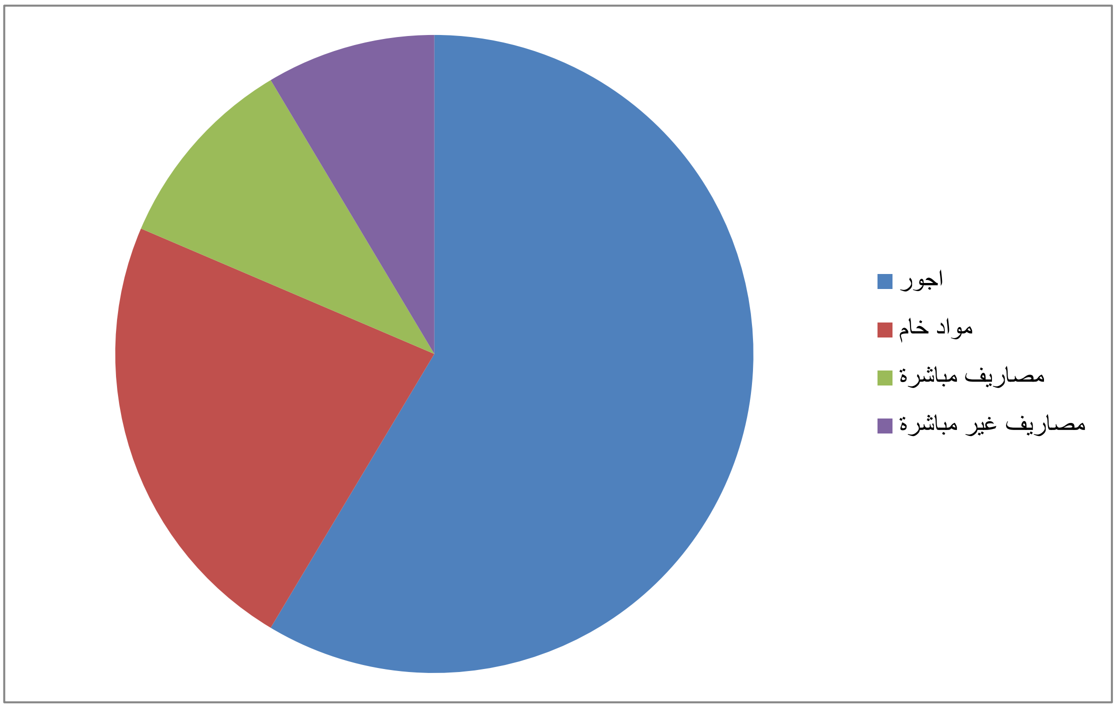
مصاريف مباشرة	مصاريف غير مباشرة	الاجور	المواد الخام	المجموع
90	**70**	**80**	**60**	**300**

زاوية قطاع المصاريف المباشرة = $(90/300) * 360^\circ = 108^\circ$

زاوية قطاع المصاريف غير المباشرة = $(70/300) * 360^\circ = 84^\circ$

زاوية قطاع الاجور = $(80/300) * 360^\circ = 96^\circ$

زاوية قطاع المواد الخام = $(60/300) * 360^\circ = 72^\circ$



Example:

Draw a Pie– chart for the following data.

مصاريف مباشرة	مصاريف غير مباشرة	الاجور	المواد الخام	مواد متفرقة
107	**45**	**70**	**69**	**59**

جدول التوزيع التكراري المتجمع

(Cumulative frequency distribution tables)

يستعمل هذا النوع من الجداول للتعرف على عدد القيم التي تقل عن قيمة معينة اوتزيد عنها وهناك نوعين من الجداول

التكرار المتجمع الصاعد

(Ascending cumulative frequency)

عملية تجميع التكرارات ابتدا من الفئة الاولى وانتهائا بالفئة الاخيرة وذلك بعد تحديد الحدود العليا تحت عبارة (Less than).

Example:

Find the Ascending cumulative frequency table.

Class	Frequency	Less than	fi
60−74	4	74	4
75−89	5	89	9
90−104	10	104	19
105−119	12	119	31
120−134	16	134	47
135−149	7	149	54
150−164	6	164	60

	$\sum 60$		

Exercises:

Find the Ascending cumulative frequency table.

Class	Frequency
46−52	7
53−59	15
60−66	27
67−73	21
74−80	14
81−87	8
88−94	9
95−101	3

التكرار المتجمع النازل

(descending cumulative frequency)

عملية تناقص التكرارات ابتدا من الفئة الاولى وانتهائا بالفئة الاخيرة وذلك بعد تحديد الحدود الدنيا للفئات تحت عبارة (mor than)

Example:

Find the descending cumulative frequency table.

Class	Frequency	mor than	fi
60−74	4	60	60
75−89	5	75	56
90−104	10	90	51
105−119	12	105	41
120−134	16	120	29
135−149	7	135	13
150−164	6	150	6
	$\sum 60$		

Exercises:

Find the descending cumulative frequency table.

Class	Frequency
46−52	7

53-59	15
60-66	27
67-73	21
74-80	14
81-87	8
88-94	9
95-101	3

مقاييس النزعة المركزية

Measure of central Tendency
[Mean, Median, Mode, Arithmetic Mean]

النزعة المركزية تعني ميل المفردات او المشاهدات نحو التمركز او التجمع حول قيمة رقمية معينة في التوزيع وبالتالي فان هذه القيمة التي تتمركز حولها البيانات تكون ممثلة لباقي القيم لوصف البيانات وإظهار الخصائص المهمة للظاهرة التي يهدف الباحث الى دراستها.

وهناك العديد من المقاييس الخاصة بقياس النزعة المركزية للبيانات حول الاحداث او الظواهر وفيما يلي عرضا لاهم المقاييس من حيث خصائصها وكيفية حسابها.

1− الوسط الحسابي **Arithmetic Mean**:

The arithmetic mean is mean of N number $x_1, x_2, \dots, x_n$ and denoted by $\bar{X}$

In the case of ungrouped date في حالة البيانات غير المبوبة

في هذه الحالة هناك طريقتان لحساب الوسط الحسابي

أ− الطريقة المباشرة (The direct method):

اذا كان لدينا قيم $x_1, x_2, \dots, x_n$ تمثل مفردات عينة مسحوبة من مجتمع ما يتم حساب الوسط الحسابي لهذه المفردات حسب الصيغة الاتية

$$\bar{X} = \frac{x_1 + x_2 + \cdots + x_N}{N}$$

Or

$$\bar{X} = \frac{\sum_{i=1}^{N} x_i}{N}$$

ب− الطريقة المختصرة طريقة الانحرافات:

The shortcut method (method of deviations)

في حالة كون قياسات العينة اعداد كبيرة يصعب التعامل معها مما يؤدي الى اختزال هذه الاعداد الى اعداد اصغر يسهل التعامل معها

لنفرض لدينا القيم $x_1, x_2, \dots, x_n$ تمثل قياسات عينة من مجتمع بحجم N وليكن a ثابت اختياري حقيقي يؤخذ من البيانات الاصلية عندئذ يتم حساب الوسط الحسابي حسب الصيغة

$$\bar{X} = a + \frac{\sum_{i=1}^{N} d_i}{N}$$

Where

$d_i = x_i - a$

Example:

Find the Arithmetic Mean by The direct method for the following data.

75 91 55 68 82 77 63

Solution.

$$\bar{X} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{91 + 77 + 82 + 68 + 55 + 63 + 75}{7}$$
$$= \frac{511}{7} = 73$$

Example:

Find the Arithmetic Mean by The shortcut method (method of deviations) for the following data.

75 91 55 68 82 77 63

Solution.

55 63 **68** **75** **77** 82 91

نختار اي رقم متوسط مثل a=82

$\boldsymbol{x_i - a = d_i}$

63−82= −19

91−82= 9

77−82= −5

82−82= 0

68−82= −14

55−82= −27

75−82= −7

= − 63

$$\bar{X} = a + \frac{\sum_{i=1}^{N} d_i}{N} = 82 + \frac{-63}{7} = 73$$

In the case of grouped date في حالة البيانات المبوبة

هي البيانات المبوبة في جدول توزيع تكراري وهناك طريقتان لحساب الوسط الحسابي

أ- الطريقة المباشرة (The direct method):

لوكان لدينا القيم $x_1, x_2, \dots, x_n$ تمثل مراكز الفئات لتوزيع تكراري عدد فئاتة N وان $f_1, f_2, \dots, f_n$ تمثل تكرار كل فئة فان الوسط الحسابي في هذه الحالة يتم حسابه حسب الصيغة التالية.

$$\bar{X} = \frac{\sum_{i=1}^{N} f_i x_i}{\sum_{i=1}^{N} f_i}$$

ب- الطريقة المختصرة طريقة الانحرافات

The shortcut method (method of deviations)

يتم في هذه الطريقة اختيار الثابت a على انه ثابت حقيقي يؤخذ من البيانات الاصلية (ويفضل ان يكن مساي لمركز الفئة الوسطي)

فان الوسط الحسابي في هذه الحالة يتم حسابه الصيغة التالية.

$$\overline{X} = a + \frac{\sum_{i=1}^{N} f_i d_i}{\sum_{i=1}^{N} f_i}$$

Where

$$d_i = x_i - a$$

Example:

Find the Arithmetic Mean by The two methods (method of deviations) for the following data.

class	f_i
0−1	4
1−2	8
2−3	12
3−4	16
4−5	20

5–6	25
6–7	6
7–8	4
	$\sum 95$

Solution.

The direct method

$$\bar{X} = \frac{\sum_{i=1}^{N} f_i x_i}{\sum_{i=1}^{N} f_i}$$

class	f_i	x_i	$f_i x_i$
0–1	4	$\frac{0+1}{2} = 0.5$	2
1–2	8	*1.5*	12
2–3	12	*2.5*	30
3–4	16	*3.5*	56
4–5	20	*4.5*	90
5–6	25	*5.5*	173.5
6–7	6	*6.5*	39
7–8	4	*7.5*	30
	$\sum 95$		$\sum 396.5$

$$\bar{X} = \frac{\sum_{i=1}^{N} f_i x_i}{\sum_{i=1}^{N} f_i} = \frac{396.5}{95} = 4.174$$

The shortcut method (method of deviations)

$$\bar{X} = a + \frac{\sum_{i=1}^{N} f_i d_i}{\sum_{i=1}^{N} f_i}$$

Let $a = 4.5$

x_i	$x_i - a = d_i$	f_i	$f_i d_i$
$\frac{0+1}{2} = 0.5$	0.5−4.5=−4	4	−16
1.5	1.5−45=−3	8	−24
2.5	2.5−4.5=−2	12	−24
3.5	3.5−4.5=−1	16	−16
4.5	4.5−4.5=0	20	0
5.5	5.5−4.5=1	25	25
6.5	6.5−4.5=2	6	12
7.5	7.5−4.5=3	4	12
		$\sum 95$	$\sum -31$

$$\bar{X} = a + \frac{\sum_{i=1}^{N} f_i d_i}{\sum_{i=1}^{N} f_i} = 4.5 + \frac{-31}{95} = 4.174$$

Exercises:

Find the Arithmetic Mean by The two methods (method of deviations) for the following data.

Class	Frequency
60−74	4
75−89	5
90−104	10
105−119	12
120−134	16
135−149	7
150−164	6

Example:

Find the Arithmetic Mean for the following data.

x_i	y_i	z_i
3	5	8
4	7	12
4	9	13
5	11	16
6	12	18
8	15	23
8	17	25
10	18	28
48	94	142

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{n} = \frac{48}{8} = \mathbf{6}$$

$$\bar{y} = \frac{\sum_{i=1}^{N} y_i}{n} = \frac{94}{8} = \mathbf{11.7}$$

$$\bar{z} = \bar{x} + \bar{y} = 6 + 11.7 = 17.7$$

$$z_i = x_i + y_i$$

$$\bar{z} = \frac{\sum_{i=1}^{N} z_i}{n} = \frac{142}{8} = \mathbf{17.7}$$

2- الوسيط (Median):

Is the Value of the middle item when the date are arranged in an increasing or decreasing order is denoted by (Me)

الوسيط في حالة البيانات غير المبوبة

In the case of ungrouped date:

If n is odd then the median is

$$\frac{n+1}{2}$$

يعتبر الوسيط مقياس اخر من مقاييس النزعة المركزية في العلوم الصرفة يعرف الوسيط بانه النقطة او الدرجة في مجموعة من الدرجات التي تكون (50%) من الدرجات اعلى منها و (50%) من الدرجات تقع تحتها

Example:

Find the median for the following

data.

(25,21,18,8,17, 32,29, ,3,1)

Solution

1 3 8 17 18 21 25 29 32

If n is odd then the median is

$$\frac{n+1}{2} = \frac{9+1}{2} = 5$$

the median is 18

If n is even then the median is

$$\frac{n}{2}, \frac{n}{2} + 1$$

Example:

Find the median for the following data.

(25,21,18,8,17, 32,29,,3)

Solution

3 8 17 18 21 25 29 32

If n is even then the median is

$$\frac{n}{2} = \frac{8}{2} = 4$$

$$\frac{n}{2} + 1 = \frac{8}{2} + 1 = 5$$

The median is 18 and 21

الوسيط في حالة البيانات المبوبة

In the case of grouped date

اما كيفية حساب الوسيط في التوزيعات التكرارية فهي عملية اكثر تعقيدا و صعوبة حسب القانون

$$me = Lm + \frac{\frac{\sum fi}{2} - Fm - 1}{Fm} * \ C$$

Example:

Find the median for the following date.

class	f_i	Fi
50–59	8	8
60–69	10	18
70–79	16	34
80–89	14	48
90–99	10	58
100–109	5	63
110–119	2	65
	$\sum 65$	

$$\frac{\sum \boldsymbol{fi}}{\mathbf{2}} = \frac{\mathbf{65}}{\mathbf{2}} = \mathbf{32.5}$$

$$me = Lm + \frac{\frac{\sum fi}{2} - F_{m-1}}{fm} * \ C$$

$$me = 70 + \frac{32.5 - 18}{16} * \ 9 = 78.06$$

Exercises:

Find the median for the following data.

Class	Frequency
60−74	4
75−89	5
90−104	10
105−119	12
120−134	16
135−149	7
150−164	6

المنوال

(The mode)

Is the value of the observation which a occurs most frequently.

يعتبر المنوال من اسهل مقاييس النزعة المركزية التي يمكن الحصول عليها بدون اجراء عمليات حسابية معقدة سواء كانت البيانات مبوبة او غير مبوبة او كانت بشكل توزيعات تكرارية.

يعرف المنوال بانه الدرجة الاكثر شيوعا او الدرجة التي تتكرر اكثر من غيرها من الدرجات في مجموعة معينة من البيانات الاحصائية

المنوال بالنسبة للبيانات غير المبوبة

In the case of ungrouped date

Example:

Find the mode for the following data.

(6, 6, 8, 9, 11, 12, 12, 14, 14, 14, 15, 16, 17, 17, 18, 19, 19)

فسنلاحظ ان الدرجة (14) قد تكررت (4) مرات وهي اكثر الدرجات تكرارا, ولذلك فان الدرجة (14) هي المنوال

Example:

Find the mode of the following data

2 3 4 6 3 4 1 6 4 5

Solution.

$m_o = 4$

المنوال بالنسبة للبيانات غير المبوبة

The mode in the case of grouped date

$$m_o = L_o + \frac{F_{mo} - F_{mo-1}}{(F_{mo} - F_{mo-1}) + (F_{mo} - F_{mo+1})} * C$$

بحيث

المنوال m_o

الحد الادنى للفئة المنوالية $\leftarrow L_o$

تكرار الفئة المنوالية F_{mo}

طول الفئة C

تكرار الفئة التي تلي الفئة المنوالية F_{mo+1}

تكرار الفئة التي تسبق الفئة المنوالية F_{mo-1}

Example:

Find the mode for frequency table bellow.

class	f_i
5–10	2
10	8
15	16
20	17

25	23
30−35	14
	$\sum 80$

$$\frac{\sum \boldsymbol{fi}}{\mathbf{2}} = \frac{\mathbf{80}}{\mathbf{2}} = \mathbf{40}$$

$$m_o = Lo + \frac{F_{mo} - F_{mo-1}}{(F_{mo} - F_{mo-1}) + (F_{mo} - F_{mo+1})} * \ C$$

$$m_o = 25 + \frac{23 - 17}{(23 - 17) + (23 - 14)} * \ 5 = 27$$

Exercises:

Find the mode for the following data.

Class	Frequency
60−74	4
75−89	5
90−104	10
105−119	12
120−134	16
135−149	7
150−164	6

Remark

There is a relation between arithmetic Mean, median and mode.

$$\bar{X} - m_o = 3(\bar{X} - m_e)$$

$$\bar{X} = \frac{3m_e - m_o}{2}$$

$$m_o = 3m_e - 2\bar{X}$$

$$m_e = \frac{1}{3}(2\bar{X} + m_o)$$

Example:

Find the arithmetic Mean, median and mode for the following data.

class	f_i	x_i	$f_i x_i$
0–1	4	$\frac{0+1}{2} = 0.5$	2
1–2	8	*1.5*	12
2–3	12	*2.5*	30
3–4	16	*3.5*	56
4–5	20	*4.5*	90
5–6	25	*5.5*	173.5
6–7	6	*6.5*	39
7–8	4	*7.5*	30
	$\sum 95$		$\sum 396.5$

$$\bar{X} = \frac{\sum_{i=1}^{N} f_i x_i}{\sum_{i=1}^{N} f_i} = \frac{396.5}{95} = 4.174$$

$$m_o = Lo + \frac{F_{mo} - F_{mo-1}}{(F_{mo} - F_{mo-1}) + (F_{mo} - F_{mo+1})} * \ C$$

$$m_o = 5 + \frac{25 - 20}{(25 - 20) + (25 - 6)} * \ 1$$

$$= 5 + \frac{5}{5 + 19} = 5.20$$

$$\boldsymbol{m_e = \frac{1}{3}(2\bar{X} + m_o)}$$

$$\boldsymbol{m_e = \frac{1}{3}(2(4.174) + 5.20) = --}$$

الوسط التوافقي

Harmonic mean

هوة مقلوب الوسط الحسابي للقيم لمقلوبات تلك القيم

Harmonic mean in the case of ungrouped date

$$H = \frac{N}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

Harmonic mean in the case of grouped date:

$$H = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} \frac{f_i}{x_i}}$$

Example:

Find the Harmonic mean of the following data.

3 2 4 6 5

Solution.

$$\boldsymbol{H = \frac{N}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{5}{\frac{1}{3}+\frac{1}{2}+\frac{1}{4}+\frac{1}{6}+\frac{1}{5}} = 3.44}$$

Example:

Find the Harmonic mean for table bellow.

class	f_i	x_i	$\boldsymbol{\frac{f_i}{x_i}}$
2−4	10	3	3.3
4	25	5	5
6	30	7	4.2
8	20	9	2.2
10−12	15	11	1.4

	$\sum 100$		$\sum 16.19$

$$H = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} \frac{f_i}{x_i}} = \frac{100}{16.19} = 6.17665$$

Exercises:

Find the Harmonic mean for frequency table bellow.

class	f_i
5−10	2
10	6
15	16
20	17
25	30
30−35	17

مقاييس التشتت

Measure of Dispersion (Variation)

ان استعمال مقاييس النزعة المركزية بمفردها في وصف الظواهر لا يمكن ان تقدم صور كاملة ومفصلة عنها وبشكل يخدم اجراءات البحث العلمي وخصوصا ما يتصل منها بطبيعة المشاهدات وكيفية توزيعها.

لذا اصبح من الضروري ان يكون هناك تكامل بين النظرة لطبيعة التبعثر في القيم ونزوعها او التمركز نحو قيمة معينة.

والتشتت يعني مدى التقارب و التباعد في قيم المشاهدات عن بعضها وان هذا التقارب والتباعد يقاس بدرجة انتشار البيانات حول الوسط او قيمة معينة.

وتسمى درجة الانتشار هذه بالتشتت او التبعثر وكلما كان التشتت او التبعثر محدودا كلما اعتبرت البيانات متجانسة وكلما توسع هذا التشتت او التبعثر كلما كانت البيانات غير متجانسة.

Measure of Dispersion (Variation)
مقاييس التشتت والاختلاف
[Rang, Mean derivation, Mean absolute deviation(Mean deviation)]

هناك انواع متعددة لمقاييس التشتت وتتباين فيما بينها من حيث الدقة ودرجة البساطة والتعقيد وجميع مقاييس التشتت قيم موجبة ولايمكن ان تكون قيمتها اقل من الصفر في استخدامها ومن اهم هذه الانواع هي:

1- مقاييس التشتت المطلق Mean absolute deviation

اي التي تكون وحداتها نفس وحدات القيم الاصلية واهمها

The Rang المدى

The range of the set number N is the different between the largest and smallest value.

يعتبر المدى ابسط المقاييس الاحصائية المستخدمة في التعرف على مدى تشتت درجات التوزيع ويعرف المدى بانه مقدار الفرق الموجود بين اعلى درجة وادنى درجة في التوزيع

فلو تم وضع الدرجات لمجموعة ما بصورة مرتبة تصاعديا من ادنى درجة الى اعلى درجة على خط مستقيم لكان المدى هو المسافة الموجودة بين الدرجتين الكبرى والصغرى فالمدى لمجموعة الدرجات

(15,90، 24، 28، 40، 65)،
هو الفرق بين اكبر درجة في المجموعة وهي (90) وادنى درجة في المجموعة وهي (15) اي ان مدى هذه المجموعة هو (90−15=75).

The rang In the case of ungrouped date

$$Rang = X_L - X_S$$

The rang In the case of grouped date

$$Rang = f_L - f_S$$

Example:

Find the rang of the following data

2 3 3 5 5 5 8 12

Solution.

$$\boldsymbol{Rang = X_L - X_S}$$

$$\boldsymbol{Rang = 12 - 2 = 10}$$

Example:

Find the rang for frequency table bellow.

class	f_i
5−10	2
10	*6*
15	*16*
20	17
25	*30*
30−35	*17*

Solution.

$$Rang = f_L - f_S$$

$$Rang = 30 - 2 = 28$$

Mean absolute deviation متوسط الانحراف المطلق

Or

Mean deviation

يعرف الانحراف المتوسط لمجموعة من الدرجات بانه متوسط الانحرافات المطلقة لتلك الدرجات عن وسطها الحسابي ويقصد بالانحراف المطلق بانه الفرق بين الدرجة و الوسط الحسابي بغض النظر عن اتجاه هذا الفرق سواء كان سالبا او موجبا.

Mean deviation In the case of ungrouped date

$$M.D = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

Example:

Find the M.D for the date **50 56 49 56 59 48**

Solution.

$$\bar{x} = \frac{\sum_{i=1}^{6} \boldsymbol{x}}{\boldsymbol{n}} = \frac{\mathbf{318}}{\mathbf{6}} = \mathbf{53}$$

$$M.D = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

$$M.D = \frac{|50-53|+|56-53|+|49-53|+|56-53|+|59-53|+|48-53|}{6}$$

$$M.D = \mathbf{4}$$

Mean deviation In the case of grouped date

$$M.D = \frac{\sum_{i=1}^{n} f_i |x_i - \bar{x}|}{\sum_{i=1}^{n} f_i}$$

Example:

Find the M.D for frequency dis table.

Solution.

$$\bar{x} = \frac{\sum_{i=1}^{6} f_i x_i}{\sum_{i=1}^{6} f_i} = \frac{\mathbf{800}}{\mathbf{200}} = \mathbf{4}$$

x_i	f_i	$f_i x_i$	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
2	18	36	2	36
3	42	126	1	42
4	80	320	0	6
5	46	230	1	46
6	10	60	2	20
7	4	28	3	12

	$\sum 200$	800		$\sum 156$

$$M.D = \frac{\sum_{i=1}^{n} f_i |x_i - \bar{x}|}{\sum_{i=1}^{n} f_i} = \frac{156}{200} = 0.78$$

Remark:

We can find M.D by mode or median

In the case of ungrouped date

$$M.D = \frac{\sum_{i=1}^{n} |x_i - me|}{n}$$

$$M.D = \frac{\sum_{i=1}^{n} |x_i - mo|}{n}$$

Exercises:

Find the M.D by mode and median.

50 56 43 56 51 48

In the case of grouped date

$$M.D = \frac{\sum_{i=1}^{n} f_i |x_i - me|}{\sum_{i=1}^{n} f_i}$$

$$M.D = \frac{\sum_{i=1}^{n} f_i |x_i - mo|}{\sum_{i=1}^{n} f_i}$$

Exercises:

Find the M.D for frequency table bellow by mode and median.

class	f_i
5−10	2
10	*6*
15	*16*
20	17
25	*30*
30−35	*17*

The semi interquartile

[rang deviation, variance standard deviation]

نصف المدى الربيعي (الانحراف الربيعي)

The semi-inter Quartile rang deviation(IQR)

احد مقاييس التشتت المطلق ويستخدم للتغلب على العيوب الموجودة في المدى وذالك لانه يستبعد القيم المتطرفة من الطرفين وذالك لانه يعتمد في حسابه على الربيع الاول Q_1 والربيع الثالث Q_3 ويمكن حسابه بالمعادلة الاتية

$$\mathbf{IQR} = \frac{\boldsymbol{Q_3 - Q_1}}{\mathbf{2}}$$

Example:

Find The semi-inter Quartile rang deviation(IQR)

4 4 10 11 15 7 14 12 6

Solution.

نرتب الاعداد تصاعديا

4 4 6 7 10 11 12 14 15

ايجاد الوسيط بين 9 اعداد هوة 10

ايجاد الوسط الحسابي عن يمين وشمال 10

4 4 6 7 10 11 12 14 15

المتوسط الحسابي عن يسار العشرة لل4 و6 هوو $5 = \frac{4+6}{2}$

المتوسط الحسابي لل14 و12 هوو $13 = \frac{14+12}{2}$

المدى الربيعي

$$\mathbf{IQR} = \frac{\boldsymbol{Q_3 - Q_1}}{\mathbf{2}}$$

$$\mathbf{IQR} = \frac{\mathbf{13 - 5}}{\mathbf{2}} = \frac{8}{2} = 4$$

Example:

Find The semi-inter Quartile rang deviation(IQR)

7 12 9 10 10 10 11 12 9 14

Solution.

ايجاد الوسيط

7 9 9 10 10 10 11 12 12 14

الوسط الحسابي 10 10 هو 10

7 9 9 10 10 10 10 11 12 12 14

المتوسط الحسابي عن يسار العشرة هو فردي 9

7 9 9 10 10 10 10 11 12 12 14

المتوسط الحسابي عن يمين العشرة هو فردي12

7 9 9 10 10 10 10 11 12 12 14

المدى الربيعي

$$\mathbf{IQR} = \frac{Q_3 - Q_1}{2}$$

$$\mathbf{IQR} = \frac{12 - 9}{2} = \frac{3}{2}$$

Exercises:

Find The semi-inter Quartile rang deviation(IQR)

4 4 5 9 1 7 8 9 6 10

الانحراف المعياري

Standard devotion

يعرف الانحراف المعياري بانه الجذر التربيعي الموجب للتباين ويستخرج بنفس الطرق والمعادلات التي تم استعراضها لاستخراج التباين ثم يؤخذ الجذر التربيعي الموجب للناتج

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

Or

If the date we devotion on n−1 if the date small $n < 30$

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

الطريقة المختصرة

the shortcut method

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

In the case of grouped date

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

Or

الطريقة المختصرة

the shortcut method

$$S = \sqrt{\frac{\sum_{i=1}^{n} f_i\,(x_i)^2 - \sum_{i=1}^{n}\frac{(f_i x_i)^2}{n}}{\sum_{i=1}^{n} f_i}}$$

Example:

Find the standard deviation for the date.

$$15 \quad 9 \quad 14 \quad 6 \quad 23 \quad 19 \quad 33$$

Solution.

$$\bar{x} = \frac{\sum_{i=1}^{7} x_i}{n} = \frac{15+9+14+6+23+19+33}{7} = 17$$

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

$$= \sqrt{\frac{(15-17)^2+(9-17)^2+(14-17)^2+(6-17)^2+(23-17)^2+(19-17)^2+(33-17)^2}{7-1}}$$

$$S = 9.07$$

Example:

Find the standard deviation by the shortcut method for the date.

x_i
4
6
5
9
8
7

Solution.

x_i	$(x_i)^2$
4	16

6	36
5	25
9	81
8	64
7	49
$\sum 39$	$\sum 271$

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

$$S = \sqrt{\frac{271 - \frac{(39)^2}{6}}{5}} = \sqrt{\frac{17.5}{5}} = \sqrt{3.5} = 1.87$$

Exercises:

Find the standard deviation for the date.

4 4 5 9 1 7 8 9 6 10

Example:

Find the standard deVotion for frequency dist table.

Solution.

$$\bar{x} = \frac{\sum_{i=1}^{6} f_i x_i}{\sum_{i=1}^{6} f_i} = \frac{1500}{100} = 15$$

x_i	f_i	$f_i x_i$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
3	6	18	144	6(144)
8	25	200	49	25(49)
13	29	377	4	29(4)
18	20	360	9	20(9)
23	10	230	2	10(2)
28	5	190	169	5(169)
33	3	99	324	3(324)
38	2	76	529	2(529)
	$\sum 100$	$\sum 1500$		$\sum 5900$

$$S = \sqrt{\frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{n} f_i}} = \sqrt{\frac{5900}{100}} = 7.68$$

ملاحظة : عندما يعطي الفئة على شكل 4−2 فان قيمة x_i مراكز فئات

Exercises:

Find the standard deVotion for frequency dist table.

f	X
4	46
1	51
2	56
2	61

2	66
9	71
5	76
10	81
4	86
8	91
3	96

Variance التباين

التباين هوة مربع الانحراف المعياري

In the case of ungrouped date

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

or

$$S^2 = \frac{\sum_{i=1}^{n}(x_i)^2 - \frac{\sum_{i=1}^{n} \quad (x_i)^2}{n}}{n-1}$$

In the case of grouped date

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

Or

$$S^2 = \frac{\sum_{i=1}^{n} f_i (x_i)^2 - \frac{\sum_{i=1}^{n} (f_i x_i)^2}{n}}{\sum_{i=1}^{n} f_i - 1}$$

التباين بطريقة الانحراف

$$S^2 = \frac{\sum_{i=1}^{n} (d_i)^2 - \frac{\sum_{i=1}^{n} (f_i d_i)^2}{n}}{n}$$

Example:

Find the Variance for frequency dist table.

Solution.

x_i	f_i	$f_i x_i$	$(x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^2$	$d_i = x_i - a$
2	15	30	10,89	163,35	2-6
4	18	72	1,69	42و30	4-6
a=6	27	162	0,49	13,23	0
8	10	80	7,29	72,9	2
10	6	60	22,09	132,54	4
	$\sum 76$	404		12	

Solution.

$$\bar{x} = \frac{\sum_{i=1}^{5} f_i x_i}{\sum_{i=1}^{5} f_i} = \frac{404}{100} = 5,3$$

$$S^2 = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{n} f_i} = \frac{412,43}{76} = 5,3$$

——————————————————————————————————————

————————————————————————

Simple linear regression
is a statistical method that allows us to summarize and study relationships between two continuous (quantitative) variables:. One variable, denoted x, is regarded as the predictor, explanatory, or independent variable.; The other variable, denoted y, is regarded as the response, outcome, or dependent variable

الانحدار الخطي البسيط هو حساب المربعات الصغري من نموذج الانحدار الخطي مع متغير تفسيري واحد .[1][2][3] وبعبارة أخرى، الانحدار الخطي البسيط هو خط مستقيم يمر بمجموعة من النقاط بطريقة تجعل مجموع مربع النقط المتبقية من النموذج (أي، المسافات الرأسية بين النقطة المتبقية والخط) أقل ما يمكن. هذا يشير الي حقيقة أن الانحدار هو واحد من أبسط الأساليب المستخدمة في مجال الإحصاء حيث أن ميل الخط يساوي العلاقة بين y و x مصححة بنسبة الانحرافات المعيارية لهذه المتغيرات. نقطة تقاطع الخط مع محور الصادات هي مركز كتلة نقاط البيانات (x, y). توجد طرق انحدار أخرى بجانب المربعات الصغري البسيطة (انظر الانحدار الخطي). علي وجه الخصوص، عندما يريد شخص أن يقوم بفعل الانحدار عن طريق العين فانه يميل عادة الي رسم خط حاد قليلا ويكون قريبا من ذلك الذي ينتج من طريقة أقل مربعات كليه. يحدث هذا لأنه طبيعي أكثر لعقل الانسان ملاحظة المسافات المتعامدة علي خط الانحدار بدلا من تلك الراسية كما يحدث في طريقة المربعات الصغري .

☑

دفتر موجدة

Moment correlation (pearson' s correlation)

في هذه الفصل ستم دراسة العلاقة بين متغيرين كمثال(x, y) باستخدام مقياس الارتباط

<u>مقاييس الارتباط:</u>

لقد لاحظنا فيما سبق ان مقاييس النزعة المركزية ومقاييس التشتت تعتبر مفيدة في وصف الدرجات لتوزيع واحد بصورة منفصلة عن مدى علاقته بالدرجات في توزيعات اخرى على متغير اخر اذ ان هناك في ميادين التربية وعلم النفس متغيرات عديدة تتطلب دراستها بشكل اخر يختلف عما تعرفنا عليه من طرق واساليب للدراسة, فقد يتطلب الموضوع دراسة علاقة توزيع الدرجات بواحد او اكثر من التوزيعات الاخرى لدرجات من نمط اخر او بمعنى اخر دراسة العلاقة الموجودة بين قيم متغيرين بدلا من دراسة خصائص متغير واحد والسؤال الذي يهمنا هو ان نجيب عن التساؤل ما مدى العلاقة الموجودة بين قيم متغيرين (X) و (Y) هل هناك علاقة بين سرعة الطالب في القراءة وبين فهمه للمعاني؟ او هل هناك علاقة بين تحصيل الطالب في العلوم او في الرياضيات وبين نسبة ذكائه لذا فقد تمكن الاحصائيون من ايجاد طرق مختلفة لمعرفة مدى العلاقة بين قيم المتغيرين دون ان تكون حاجة الى ترتيبها.

وتتباين اساليب وطرق استخراج معاملات الارتباط تبعا لنوع المتغيرات واساليب قياسها. فاذا كان احد المتغيرين فئويا والاخر نسبيا او العكس او كان كلاهما فئويان او نسبيان فيستخدم معامل الارتباط بيرسون اما اذا كان المتغيران رتيبان معامل الارتباط سبيرمان للرتب ... وهكذا

<u>انواع معاملات الارتباط :</u>

ان مقاييس معامل الارتباط يقسم الى عدة انواع من حيث عدد المتغيرات ونوع القيم التي تتضمنها اذا كانت عددية او مرتبة (نوعية) وهذه الانواع هي:

1-**معامل الارتباط الخطي الوسيط** :

ويقيس الارتباط بين متغيرين اثنين ذات قيم عددية.

2-**معامل الارتباط الخطي العددي**:

ويقيس الارتباط عندما يكون عدد المتغيرات اكثر من اثنين و ذات قيم عددية.

3-**معامل الارتباط الجزئي**:

ويقيس العلاقة بين متغيرين اثنين فقط من بين عدة متغيرات على فرض ان تأثير بقية المتغيرات الاخرى تبقى ثابتة.

4-**معامل الارتباط للرتب**:

ويقيس الارتباط بين متغيرين ليس باستخدام قيمها الاصلية وانما بين رتب هذه القيم.

1-***معامل ارتباط بيرسون***:

من الامور التي يهتم بها الباحثون في مجال التربية وعلم النفس وفي غيرها من المجالات للتعرف على مدى العلاقة الموجودة بين قيم متغيرين مستمرين سواء كان كلاهما من النوع الفئوي او النسبي او كان احدهما نسبيا والاخر فئويا او بالعكس يحتاج الباحث الى طريقة لاستخراج معامل الارتباط بين المتغيرين.

وفيما يلي نص قانون بيرسون لإيجاد معامل الارتباط:

$$\mathrm{R} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

حيث ان R = معامل الارتباط

N= عدد الافراد

X= الدرجة في الاولى

Y= الدرجة الثانية.

<u>مثال:</u>

قام باحث باختيار عينة من (12) طالب وطبق عليهم اختباري للتحصيل احدهما في الرياضيات والثاني في اللغة العربية وبعد تصحيح اوراق الاجابة كانت الدرجات لكل طالب كالاتي:

XY	Y^2	X^2	درجته في اللغة العربية (Y)	درجته في الرياضيات (X)	تسلسل الطالب
60	36	100	6	10	1
28	16	49	4	7	2
84	49	144	7	12	3
96	64	144	8	12	4
90	100	81	10	9	5
112	49	256	7	16	6
120	100	144	10	12	7
270	225	324	15	18	8
40	25	64	5	8	9
72	36	144	6	12	10
154	121	196	11	14	11
208	169	256	13	16	12

المجموع	**146**	**102**	**1902**	**990**	**1334**

وبالتعويض بهذه القيم في القانون السابق نجد ان:

$$\mathrm{R} = \frac{12(1334) - (146)(102)}{\sqrt{[12(1902) - (146)^2][12(990) - (102)^2]}} = 0{,}784$$

2-<u>معامل الارتباط سبيرمان:</u>

اشتق قانون معامل الارتباط للرتب من قبل العالم سبيرمان براون لمعالجة حلات خاصة تعتمد على رتب القيم او المشاهدات بدلا من استخدام القيم العددية الاصلية التي يجري معالجتها باستخدام معامل ارتباط بيرسون وكثيرا ما يستعمل هذا المعامل في التعامل مع البيانات الوصفية التي يستحيل فيها استخراج معامل الارتباط بطريقة بيرسون الذي يعتمد على القيم العددية, و عليه يلجأ البعض الى حساب معامل الارتباط للرتب وذلك لسهولة حسابه من ناحية او لتعذر التعامل مع القيم الاصلية بدقة كافية من اخرى وبالأخص عندما تكون حساباتها طويلة ومعقدة وكما يشيع استخدام هذا المعامل عندما يكون عدد ازواج البيانات للمتغيرين قليلة نسبيا اي لا تزيد عن ثلاثين زوجا لهذا يكون اهتمام الباحث بالرتب اكثر من اهتمامه بقيمها الحقيقية.

ويتم استخراج معامل الارتباط للرتب (سبيرمان) بموجب القانون الاتي:

$$R = 1 - \frac{6\sum D^2}{N(N^2-1)}$$

حيث ان

R = معامل الارتباط للرتب

$\sum D^2$ = مجموع مربعات الفروق بين الرتب المتناظرة للمتغيرين.

<u>مثال:</u>

اراد باحث معرفة العلاقة الرتباطية بين مادتي الخط العربي (X) وبين نشاطه بتوجيه الاسئلة والمناقشة في الصف (7) لطلبة السادس فأجرى اختبار على عينة مؤلفة من (14) فرد من الطلاب في احدى المدارس الاساسية ثم قام باملاء قطعة من كتاب المطالعة المقررة لصف اخر, فاعطى اوراق الاجابة الى عدد من المعلمين كمحكمين وطلب منهم ترتيب الاوراق بحسب جودة الخط بحيث تعطى الرتبة (1) للطالب ذو الاجابة ذات الخط الافضل و الرتبة (2) للطالب الذي يليه وهكذا حتى الرتبة (14) التي تعطى للطالب الاقل جودة في خطه عن بقية زملائه ثم طلب من مجموعة من المعلمين تصنيف الطلاب حسب نشاطهم باعطاء الرتبة (1) للطالب الاكثر نشاطا والمرتبة (14) للطالب الاقل نشاطا في توجيه الاسئلة و المناقشة في الصف فكانت النتائج كالاتي:

ت	الترتيب في الخط (X)	الترتيب في النشاط (Y)	(X−Y)	$(X-Y)^2$ D^2
1	14	12	2	4
2	12	11	1	1
3	10	8	2	4
4	6	3	3	9
5	7	9	2−	4
6	4	4	صفر	صفر
7	1	1	صفر	صفر
8	3	5	2−	4
9	2	2	صفر	صفر
10	9	10	1−	1
11	5	6	1−	1
12	8	7	1	1

4	**2−**	**13**	**11**	**13**
1	**1−**	**14**	**13**	**14**
34				المجموع

نطبق القانون اعلاه

$$R = 1 - \frac{6\sum D^2}{N(N^2-1)}$$

$$R = 1 - \frac{6(34)}{14(195-1)} = .,93$$

Linear correlation

Is it linear relation between two variable and it is denoted by r where $-1 \le r \le 1$

$$S_x = \sqrt{\frac{\sum(x_i - \overline{x})^2}{n}}$$

$$S_y = \sqrt{\frac{\sum(y_i - \overline{y})^2}{n}}$$

$$S_{xy} = cov(x, y) = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{n}$$

$$r = \frac{S_{xy}}{S_x . S_y} = \frac{cov(x.y)}{S_x . S_y}$$

$$\boldsymbol{r} = \frac{(\boldsymbol{x_i} - \overline{\boldsymbol{x}})(\boldsymbol{y_i} - \overline{\boldsymbol{y}})}{\sqrt{\sum(\boldsymbol{x_i} - \overline{\boldsymbol{x}})^2 \sum(\boldsymbol{y_i} - \overline{\boldsymbol{y}})^2}}$$

$$\boldsymbol{r} = \frac{\sum \boldsymbol{x_i}\,\boldsymbol{y_i} - \boldsymbol{n}\overline{\boldsymbol{x}}\overline{\boldsymbol{y}}}{\sqrt{(\sum \boldsymbol{x_i}^2 - \boldsymbol{n}\overline{\boldsymbol{x}}^2)(\sum \boldsymbol{y_i}^2 - \boldsymbol{n}\overline{\boldsymbol{y}}^2)}}$$

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

This is pearson' s coefficient of simple correlation

*If $r = -1$ this relation is called inverse complete correlation

*If $r = 0$ there is not correlation

*If $-\boldsymbol{1} < \boldsymbol{r} < \boldsymbol{0}$ this correlation relation is called weaker inverse not complete

*If $-\frac{\boldsymbol{1}}{\boldsymbol{2}} < \boldsymbol{r} < \frac{\boldsymbol{1}}{\boldsymbol{2}}$ this correlation relation is called weaker

*If $\frac{\boldsymbol{1}}{\boldsymbol{2}} < \boldsymbol{r} < \boldsymbol{1}$ this correlation relation is very strong.

Example: Find the pearson' s coefficient of simple correlation between x, y.

x_i	y_i
4	3
5	4
3	2
6	5
7	7
8	7
9	6
4	4
5	3
12	9
63	50

Solution.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{63}{10} = 6{,}3$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{50}{10} = 50$$

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \overline{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
4	3					
5	4					
3	2					

6	5					
7	7					
8	7					
9	6					
4	4					
5	3					
12	9					
63	50			51	68,1	44

$$r = \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} = \frac{51}{54,74} = 0,932$$

الانحدار Regression

The regression theory is to describe the linear relation as a mathematical equation.

في بعض الاحيان يتطلب الامر معرفة مدى تاثير احد هذين المتغيرين على الاخر وحساب مقدار هذه التغيير الذي يحصل على احدهما عندما يزداد المتغير الاخر وكذالك تقدير قيمة المتغير الاول عند قيمة محددة للمتغير الثاني

نظرية الانحدار تمكننا من معرفة هذه الامور من خلال معادلة الانحدار التي تتكون من هذين المتغيرين والذي يدعى

المتغير الاول بالمتغير التابع dependent variable

اما المتغير الثاني فيدعى المتغير المستقل Independent variable

اذا كان لدينا المتغيرين x ,y حيث ان

y هو المتغير التابع

x هو المتغير المستقل

فان معادلة انحدار المتغير y على المتغير x يرمز له بالرمز $\frac{y}{x}$

معادلة انحدار $\frac{y}{x}$ **هي**

$$y_i = \alpha + \beta x_i + e_i$$

حيث

simple linear regression

Multiple linear regession.