تعريف علم الإحصاء

يعرف علم الإحصاء بأنه ذلك الفرع من العلوم الذى يختص بالطرق العلمية لجمع وتنظيم وتلخيص وعرض وتحليل البيانات، وذلك للوصول إلى نتائج مقبوله وقرارات سليمة على ضوء هذا التحليل.

ينقسم علم الإحصاء إلى قسمين أساسين هما:

١- الإحصاء الوصفى Descriptive Statistics.

٢- الإحصاء الاستدلالى Statistical Inference.

الإحصاء الوصفي: هو عبارة عن الطرق الخاصة بتنظيم وتلخيص المعلومات والغرض من التنظيم هو المساعدة على فهم المعلومات. والطرق الوصفية تحتوى على توزيعات تكرارية (الجداول التكرارية) ورسوم بيانية وطرق حساب مقاييس النزعة المركزية ومقاييس التشتت ومختلف القياسات الأخرى.

الإحصاء الاستدلالي: هو عبارة عن الطرق العلمية التى تعمل للاستدلال عن معالم المجتمع بناء على المعلومات التى تم الحصول عليها من العينة المأخوذة منه، وذلك وفق الطرق الإحصائية المعلومة.

المجتمع الإحصائى والعينة الإحصائية

يعرف المجتمع على أنه مجموعة من الأفراد محل الدراسة والتى لها خصائص مشتركة. ويقسم المجتمع الإحصائى إلى قسمين:

١- مجتمع محدود: والذى يكون فيه عدد محدود من الأفراد مثل عدد أجهزة الكمبيوتر فى المعمل، عدد طلاب الفرقة الأولى فى كلية ما...إلخ.

٢- مجتمع غير محدود: هو المجتمع الذى يكون فيه عدد الأفراد غير منته مثل عدد النجوم فى السماء، عدد حبات القمح المحصود فى مزرعة معينة...إلخ.

فى بعض الأحيان يكون من الصعب ملاحظة بيانات جميع أفراد المجتمع لما يكلف ذلك من جهد ووقت ومال، أو قد يكون فى بعض الأحيان استحالة ذلك مثل فحص جميع دم المريض. وللتغلب على ذلك يمكن اختيار جزء من المجتمع يسمى بالعينة.

وتعرف **العينة** بأنها جزء من المجتمع والتى يتم اختيارها بحيث تمثل جميع صفات المجتمع وينفرد بها فرع خاص من علم الإحصاء يسمى نظرية العينات، وهو خارج نطاق كتابنا هذا. وقد تكون الحاجة ضرورية لأخذ العينة بديلا عن دراسة المجتمع كله، مثل أخذ عينة من دم مريض لفحصها حيث إننا لانستطيع فحص كل دم المريض لأن ذلك يؤدى إلى الوفاة.

البيانات

هى مجموعة من المشاهدات أو الملاحظات التى تؤخذ أثناء دراسة معينة، وقد تكون بيانات رقمية (كمية) مثل أطوال وأوزان مجموعة من الطلاب أو دخول مجموعة من الأسر أو بيانات غير رقمية (وصفية) مثل لون البشرة والجنس ... إلخ.

المعلمة والإحصائية

المعلمة: شيء يميز المجتمع ككل، وذلك مثل متوسط الدخل الشهرى للأسر فى دولة معينة، أو متوسط طول الطلاب فى مدرسة ما، أو نسبة المدخنين فى مجتمع معين، أو نسبة المعيب فى الإنتاج لإحدى السلع، وهكذا...

الإحصائية: هى شيء يميز العينة مثل متوسط الدخل الشهرى لعينة مكونة من ١٠٠ أسرة فى دولة ما، أو متوسط الطول لعينة مكونة من ٥٠ طالب فى مدرسة ما، وهكذا...

المتغيرات

هو مقدار له خصائص رقمية (كمية) وغير رقمية (وصفية) تتغير قيمته من عنصر إلى آخر من عناصر المجتمع أو العينة. فمثلا إذا رغبنا فى دراسة ظاهرة مثل الطول أو الوزن أو لون البشرة أو لون العين فإن قراءة المفردات لمتغير الطول أو اللون أو الذكاء تكون بيانات كمية (رقمية) Quantitative وظاهرة الجنس أو لون الشعر أو لون العين تأخذ قيما وصفية (غير رقمية) Qualitative.

(مقاييس النزعة المركزية)

الوسط الحسابى (المتوسط)

المتوسط أو الوسط الحسابى يعتبر من أهم مقاييس النزعة المركزية والأكث ر اس تخداما ف ى الإحصاء والحياة العملية، إذ يستخدم عادة فى الكثير من المقارنات بين الظواهر المختلفة. ولو أسندت قيمة المتوسط لكل مشاهدة فإن مجموع هذه القيم الجديدة يكون مساويا لمجموع الم شاهدات الأص لية (أنظر مثال (١)) ويعرف كالتالي:

إذا كان لدينا مجموعة من المشاهدات للمتغير x وهى $(x_1, x_2, ..., x_n)$ فإن الوس ط الح سابى يساوى حاصل جمع المشاهدات أو البيانات مقسوما على عددها ويرمز له بالرمز $\bar{x}$ وعليه فإن:

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

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

الوسيط

عند ترتيب البيانات (المشاهدات) ترتيبا تصاعديا أو تنازليا فإن الوسيط يكون البيان (المشاهدة) التى يقع ٥٠% من البيانات قبلها فى الترتيب و ٥٠% من البيانات بعدها فى الترتيب. فإذا كان عدد البيانات فرديا فإن الوسيط يكون المشاهدة التى تقع فى المنتصف، وإذا كان عدد البيانات زوجيا فإن الوسيط هو متوسط المشاهدتين اللتين تقعان فى المنتصف.

المنوال

يعرف المنوال على أنه القيمه الأكثر شيوعا (تكرارا) فى مجموعة البيانات. وقد يكون لمجموعة البيانات منوال واحد ولذلك يطلق عليها وحيدة المنوال، أو يكون لها أكثر من منوال وتسمى متعددة المنوال. وقد لا يكون لمجموعة البيانات أى منوال وبذلك تسمى عديمة المنوال.

العلاقة بين الوسط الحسابى والوسيط والمنوال

توجد علاقة تجريبية بين المقاييس الثلاثة (الوسط الحسابى – الوسيط – المنوال) وذلك فى حالة التوزيعات التكرارية أحادية المنوال وغير المتماثلة والمتماثلة وذات الالتواء البسيط. وتعط ى ه ذه العلاقة من خلال المعادلة التالية:

الوسط الحسابى – المنوال = ٣ (الوسط الحسابى – الوسيط)

وقد وجد أن الوسيط تقع قيمته بين قيمتى الوسط الحسابى والمنوال.

وفى حالة التوزيعات التكرارية المتماثلة الوحيدة المنوال فإن قيمة الوسط الحسابى تكون مساوية لقيمة الوسيط تكون مساوية لقيمة المنوال أى أن:

الوسط الحسابى = الوسيط = المنوال

الوسط الهندسى

الوسط الهندسى .G.M لمجموعة من القيم $x_1, x_2, ..., x_n$ هو الجذر النونى لحاصل ضرب ه ذه القيم: $G.M. = \sqrt[n]{x_1 x_2 ... x_n}$، يمتاز الوسط الهندسى عن الوسط الحسابى بأنه أقل تأثرا بالقيم الشاذة فى البيانات لأنه معلوم رياضيا بأن الوسط الهندسى لمجموعة من القيم أقل من وسطها الحسابي، وع ادة يحسب الوسط الهندسى باستخدام اللوغاريتمات كالتالي:

$$\text{Log G.M.} = \frac{1}{n}\left(\sum_{i=1}^{n} Log(x_i)\right)$$

الوسط التوافقي

يستخدم الوسط التوافقى عندما يكون مقلوب المتغير له دلاله كأن يعين نسبة بين متغيرين مرتبطين مثل السرعة بالنسبة للزمن. والوسط التوافقى H لمجموعة من القيم $x_1, x_2, ..., x_n$ هو مقلوب الوسط الحسابى لمقلوبات هذه القيم، أى أن:

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

ومن الناحية العملية يكون كالتالي:

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

(مقاييس التشتت)

المدى

يعرف المدى بأنه الفرق بين أكبر قراءة وأصغر قراءة فى مجموعة القراءات أى أن:

نصف المدى الربيعي

من أهم عيوب المدى أنه يتأثر بالقيم الشاذة وبالتالى فهو لا يعطى صورة صادقة عن طبيعة البيانات. لذلك دعت الحاجة إلى إيجاد مقياس آخر يتم من خلاله التخلص من تأثير القيم الشاذة وهو ما يسمى بنصف المدى الربيعي. ويعرف كما يلي:

إذا كانت لدينا مجموعة من البيانات عددها n قيمة فيتم ترتيب القيم تصاعديا وتقسم إلى أربعة أقسام متساوية كما هو موضح على الخط الأفقى التالي:

$$Qn \_\_\_\_\_\_\_\_\_\_ Q_2 \_\_\_\_\_\_\_\_\_\_ Qn$$

تسمى القيمة التى يسبقها ربع البيانات بالربيع الأول ويرمز له بالرمز Q_1 ورتبته $\frac{n}{4}$، وتسمى القيمة التى يسبقها ثلاثة أرباع البيانات بالربيع الثالث ويرمز له بالرمز Q_3 ورتبته $\frac{3n}{4}$ كما يسمى المقدار الناتج من الفرق بين Q_1، Q_3 بالمدى الربيعى وهو يمثل النصف الأوسط للقيم، ويؤخذ نصف هذا المدى مقياسا للتشتت ويسمى بنصف المدى الربيعى ويرمز له بالرمز Q ويعطى من خلال العلاقة:

$$Q=\frac{Q_3-Q_1}{2},$$

ويلاحظ أن Q_2 هو الربيع الثانى وهو القيمة التى يسبقها نصف البيانات ورتبته $\frac{n}{2}$ أى أن Q_2 هو الوسيط الذى سبق شرحه فى مقاييس النزعة المركزية

الانحراف المتوسط

يعرف الانحراف المتوسط بأنه متوسط الانحرافات المطلقة للبيانات عن وسـ طها الحـ سابى $\bar{x}$ ويرمز له بالرمز .M.D ويعرف رياضيا كالتالي:

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

الانحراف المعيارى

من الصعب التعامل رياضيا (تحليليا) مع الانحراف المتوسط، ولذلك دعت الحاجة إلى استخدام مقياس للتشتت بنفس قوة الانحراف المتوسط، ولكى يكون من السهل التعامل معه تحليليًا، وبما أن الفكرة هى التخلص من الإشارات للانحرافات فإن تربيع الانحرافات يخلصنا من الإشارة. ولهذا فإن الانحراف المعيارى يعرف عن طريق التباين والذى يعرف على أنه متوسط مربع انحرافات القيم عن وسطها الحسابى ويرمز له بالرمز σ^2، والجذر التربيعى للتباين ينتج عنه مقياس من أهم وأدق مقاييس التشتت وهو ما يسمى بالانحراف المعيارى ويرمز له بالرمز σ، وسوف نتناول طريقة حسابه فى حالة البيانات المباشرة والبيانات المبوبة كما يلي:

الانحراف المعيارى فى حالة البيانات المباشرة

إذا كان لدينا مجموعة من البيانات من مجتمع إحصائى عدد مفرداته N على الصورة $X_1, X_2, ..., X_N$ ومتوسط هذه البيانات $\bar{X}$ فإن مربعات انحرافات هذه القيم عن وسطها الحسابى يكون على الصورة:

ويعرف التباين σ^2 كالتالي:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2 ,$$

والانحراف المعيارى هو:

$$\sigma = \sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{N}(X_i - \bar{X})^2} ,$$

أما فى حالة العينة التى حجمها n المأخوذة مـ ن المجتمع فإن الانحراف المعيارى فى هذه الحالة يرمز له بالرمز S والتبـ اين S^2 ويعـ رف بقـ سمة مجموع مربعات الانحرافات على $(n-1)$ ويكتب كما يلي

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

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

مقاييس الالتواء

لقد سبق وتحدثنا عن طرق عرض البيانات جدوليا وبيانيا ثم مقاييس النزعة المركزية ومقاييس التشتت ولم نتعرض لأى مقياس يوضح درجة التواء المنحنى التكراري. ويقصد بكلمة التواء هو بعد المنحنى عن التماثل، ويكون منحنى التوزيع التكرارى ملتويا نحو اليمين إذا كانت القيم المتطرفة نحو اليمين تؤثر على الوسط الحسابى وتتجه به نحو اليمين، وبذلك يكون الوسط الحسابى أكبر من الوسيط. أما إذا كان التوزيع ملتويا نحو اليسار فإن الوسط الحسابى يكون أصغر من الوسيط.

مقياس الالتواء له أشكال مختلفة نذكر ثلاثة منها وهي:

1- $r = \dfrac{3(\bar{x} - Med)}{s}$,

2- $r = \dfrac{\bar{x} - Mod}{s}$,

مقاييس التفرطح

هو مقياس يقيس درجة علو أو انخفاض أى منحنى توزيع تكرارى بالنسبة للمنحنـ ى الطبيعـ ى وهو منحنى متماثل حول الرأس يمر بالمتوسط.

ويعرف معامل التفرطح k كما يلي:

$$k = \frac{m_4}{s^4} - 3$$

حيث:

$$m_4 = \frac{\sum (x - \bar{x})^4}{n},$$

الارتباط والانحدار الخطي البسيط

الارتباط

يوفر تحليل الارتباط وسيلة للاستدلال على قوة العلاقة بين متغيرين أو اكثر، أي أن الارتباط هو مقياس للدرجة التي تتغير فيها قيم المتغير بأسلوب منتظم. وهو يعتبر مؤشر كمي لتحديد درجة الاعتماد على متغير أو اكثر في التنبؤ بقيم متغير آخر. من المهم معرفة ما يمكن أن يوفره التحليل الارتباط وبنفس الأهمية يتوجب معرفة مالا يمكن أن يوفره هذا النوع من التحليل. فتحليل الارتباط لا يقدم أية معلومات للتنبؤ بقيم متغير ما، كما انه لا يوفر أي مؤشر فيما لو كانت العلاقة بين المتغيرات سببية، انما يستطيع التحليل تحديد فقط فيما لو كان درجة التباين المشترك ذات دلالة. ولذا تعرف العلاقة بين الظاهرتين او متغيرين بالارتباط. وقد يكون الارتباط طرديا بمعنى ان تتغير الظاهرتين في نفس الاتجاه بحيث اذا زادت احدى الظاهرتين تميل الثانية الى الزيادة وبالعكس. وقد يكون الارتباط عكسيا بمعنى ان تتغير الظاهرتان في اتجاهين متضادين بحيث اذا زادت احدى الظاهرتين تميل الثانية الى النقصان وبالعكس.

يلاحظ ان قيمة معامل الارتباط هي قيمة عددية نسبية تنحصر بين 1+ و1- ولا تكون هذه القيمة 1+ و1- الا اذا كان الارتباط تاما.

أنواع الارتباط:

يقسم الارتباط الى عدة أنواع، وذلك بحسب نوع المتغير المراد قياسه، اذ هناك متغيرات كمية مقاسة، وهناك متغيرات أخرى نوعية، قياسها لا يعتمد على كميات عددية.

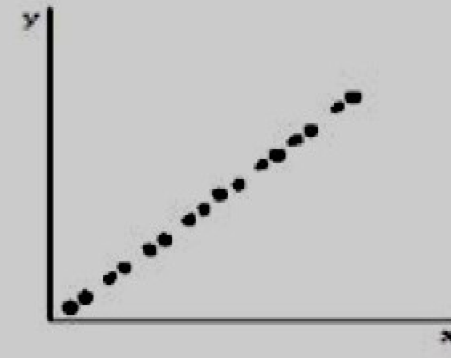
معامل الارتباط للظواهر للمتغيرات الكمية:

ويشمل دراسة العلاقة فيما بين الظواهر المقاسة، وهي الظواهر القابلة للقياس الكمي او العددي. وهذا يشمل جميع الظواهر التي يمكن التعبير عنها بصورة رقمية كالطول والدخل وكمية الإنتاج وغير ذلك من الظواهر التي يمكن التعبير عنها رقميا. ويقسم الى عدة انواع

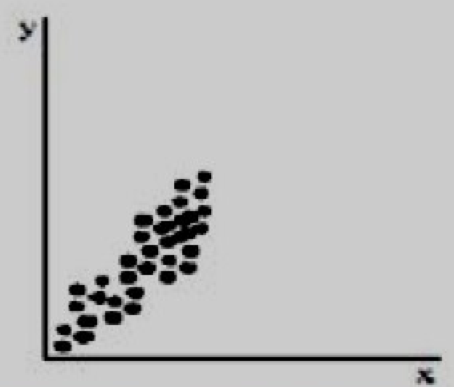
معامل الارتباط البسيط (معامل بيرسون)

وهو معامل ارتباط يحدد مقدار او حجم العلاقة واتجاهها بين متغيرين اثنين، وذلك في الحالات او الظواهر التي تقتصر فيها الدراسة على متغيرين. مثال قد يكون من المطلوب التعرف على حجم العلاقة واتجاهها بين اطوال مجموعة من الأشخاص واوزانهم. او قد يكون الهدف مثلا التعرف على حجم واتجاه العلاقة بين مقدار الدخل الشهري وحجم الانفاق الشهري للأسر في مجتمع ما.

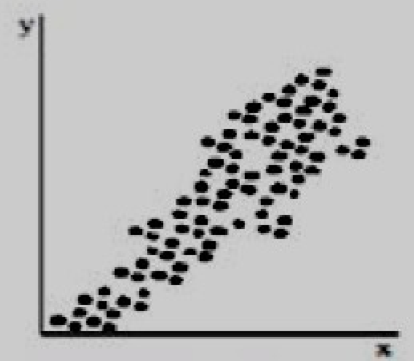
للارتباط عدة أنواع يمكن التعرف عليها من خلال كل من مقدار معامل الارتباط ومن خلال اتجاه العلاقة بين المتغيرين بالاعتماد على لوحة انتشار البيانات.



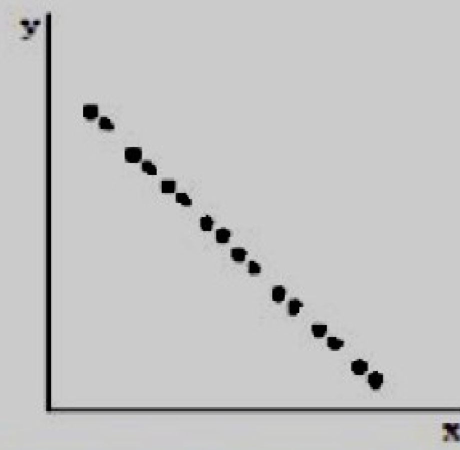
ارتباط طردي تام



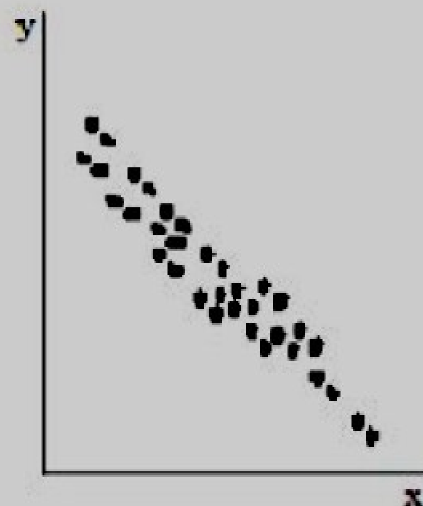
ارتباط طردي قوي



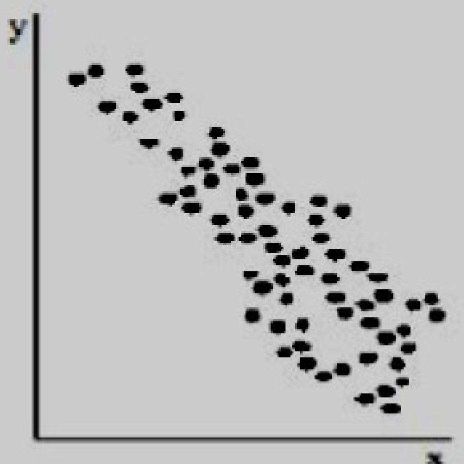
ارتباط طردي



ارتباط عكسي تام



ارتباط عكسي قوي



ارتباط عكسي

نوع علاقة الارتباط	قيمة معالم الارتباط
ارتباط طردي تام	+1
ارتباط طردي قوي	من 0.70 الى 0.99
ارتباط طردي متوسط	من 0.50 الى 0.69
ارتباط طردي ضعيف	من 0.01 الى 0.49
لا يوجد ارتباط خطي	0

وكذلك الحال وبنفس المستوى تكون علاقة الارتباط عكسية في حالة كانت إشارة معامل الارتباط سالبة.

فإن معامل الارتباط r لبيرسون يعطى من خلال العلاقة:

$$r = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{(\sum x^2 - \frac{(\sum x)^2}{n})(\sum y^2 - \frac{(\sum y)^2}{n})}},$$

ويكون لمعامل الارتباط (r) الخصائص التالية:

١- قيمته تساوى صفرا عندما تكون الظاهرتان مستقلتان تماما.

٢- قيمته مقدار موجب عندما يكون الارتباط بين المتغيرين طرديًا. ويكون قويا عندما يكون المق دار الموجب قريبا من الواحد الصحيح، وضعيفا عندما يكون المقدار الموجب قريبا من الصفر.

٣- قيمته مقدار سالب عندما يكون الارتباط بين المتغيرين عكسيا. ويكون قويا عندما يكون المق دار السالب قريبًا من (−١)، وضعيفاً عندما يكون المقدار السالب قريبًا من الصفر.

خط الانحدار البسيط

لقد سبق لنا دراسة العلاقة بين متغيرين (x, y) وإيجاد معامل الارتباط بينهما بعدة طرق وذلك لقياس قوة الارتباط وإتجاه العلاقة بينهما (طردية – عكسية) كما فى معاملى ارتباط بيرسون وسبيرمان ومدى قوة العلاقة كما فى حالة معاملى الاقتران والتوافق. وفيما يلى نبحث عن إيجاد معادلة رياضية تمثل أفضل توفيق لخط مستقيم يعبر عن البيانات فى شكلها الخطي. والغرض من إيجاد معادلة خط الانحدار هو التنبؤ بقيمة المتغير التابع لقيمة محددة من قيم المتغير المستقل. وتسمى العلاقة بين المتغير المستقل x والمتغير التابع y بمعادلة خط الانحدار البسيط. وعليه فإذا كان x متغيرًا مستقلاً، y متغيرًا تابعًا فإن المعادلة التى نحصل عليها تسمى بمعادلة خط انحدار y على x وهى على الصورة التالية:

$$y = a + bx,$$

حيث يعرف a على أنه ثابت الانحدار (الجزء المقطوع من محور y) ويحسب من خلال العلاقة:

$$a = \bar{y} - b\bar{x},$$

كذلك يعرف b بمعامل انحدار y على x ويحسب من خلال العلاقة التالية:

كذلك يعرف b بمعامل انحدار y على x ويحسب من خلال العلاقة التالية:

$$b = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}},$$