مدخل إلى علم الإحصاء

Introduction to Statistics

علم الإحصاء Statistics: هو أحد فروع الرياضيات الهامّة ذات التطبيقات الواسعة سواءً في العلوم الطبيعية أو الاجتماعية أو غيرها .. بحيث يهتمّ هذا العلم بشكلٍ أساسي بجمع وتنظيم وتلخيص وعرض وتحليل وتفسير ونشر البيانات الإحصائية.

ينقسم علم الإحصاء إلى قسمين أساسيين:

١. الإحصاء الوصفي: يشمل جمع وتبويب البيانات الإحصائية. والطرق الوصفية تحتوي على توزيعات تكرارية، ورسوم بيانية، وطرق حساب مقاييس النزعة المركزية ومقاييس التشتت ومختلف القياسات الأخرى.
٢. الإحصاء الاستقرائي أو الاستدلالي: يبحث في تحليل البيانات واستقراء النتائج واتخاذ القرارات.

البيانات Data: هي عبارة عن معلومات تكون مرتّبة ومنظّمة ومخزّنة بطريقة سهلة التداول والفهم والتفسير ومدعّمة بالأرقام.

من أين نحصل على البيانات ؟

- من مشاهدة ومراقبة الظواهر الموجودة في الطبيعة.

مثل: كميّة الأمطار الهاطلة في مكان معيّن، أو درجات الحرارة، أو سرعة الرياح..

- من خلال إجراء الاختبارات أو التجارب.

مثل: تجربة كميّات مختلفة من السماد أو كميّات مختلفة من مياه الري للوصول إلى الكميّة المُثلَى التي تعطي أفضل إنتاج من نوع نباتي معيّن.

- من خلال إجراء الدراسات والأبحاث والمقابلات.

مثل: إجراء دراسة ميدانية لجمع بيانات حول الوضع الاجتماعي والاقتصادي للمزارعين في منطقة معيّنة، أو إجراء مقابلات مع العاملين في شركة معيّنة لأخذ بيانات حول ظروف العمل في الشركة ومدى رضاهم عن العمل بما يمكن أن يساهم في تحسين ظروف العمل في الشركة وبالتالي إنتاجية العمال.

تصنيف البيانات: يمكن تصنيف البيانات بشكلٍ عام إلى نوعين رئيسيين:

- البيانات النوعيّة أو الوصفية أو الكيفية Qualitive Data: وتنقسم بدورها إلى نوعين:

- البيانات الإسمية Nominal Data: تكون في صورة غير عددية أي لا يمكن قياسها وإجراء العمليات الحسابية عليها، ولكن بالإمكان عدّ أفراد كل فئة. مثل الجنس (ذكور، إناث) (١، ٢) أو لون بتلات الأزهار (أحمر ١، وردي٢، برتقالي٣) أو سؤال إجابته نعم أم لا (١، ٠).
- البيانات الترتيبية Ordinal Data: تكون في صورة غير عددية ولا يمكن إجراء العمليات الحسابية عليها والفرق بينها وبين البيانات الإسمية هي عملية المفاضلة والترتيب بين فئات أو رتب المتغير ولكن ليس شرطاً تساوي المسافات بين الرتب المتجاورة مثل المستوى التعليمي (ابتدائي ١، إعدادي ٢، ثانوي ٣، جامعي ٤).

- <u>البيانات الكميّة Quantitative Data</u>:

- البيانات الفترية Interval Data: تكون في صورة عددية ويمكن إجراء العمليات الحسابية عليها مثل المتوسّط الحسابي والانحراف المعياري وغير ذلك.. وتمتاز هذه البيانات بتساوي المسافات بين الرتب المتجاورة كما أنّه في هذه البيانات لا يوجد صفر حقيقي وإنّما صفر افتراضي لا يعني انعدام الخاصيّة فدرجة طالب تساوي صفر مثلاً لا يعني أنّه لا يعرف شيئاً في المقرر، ومن أمثلة هذه البيانات التواريخ ودرجات الحرارة.
- البيانات النسبية Rational Data: هي بيانات ذات مستوى أعلى من البيانات السابقة حيث تمتاز بالتصنيف (مثل البيانات الاسمية) والترتيب (مثل البيانات الترتيبية) والمسافات المتساوية (مثل البيانات الفترية) وخاصيّة النسبية التي تعني أنّ للصفر خاصيّة انعدام الظاهرة (مثل أنّ سرعة سيارة تساوي الصفر تعني أن السيارة متوقّفة) أي أنّ الصفر هنا صفر حقيقي.

يتمّ الحصول على البيانات الكميّة غالباً من خلال قراءات الأجهزة والأدوات العلمية المستخدمة في الأبحاث، كقياسات الأوزان والأطوال والحجوم والأعمار والزمن والسعة والمسافات ودرجات الحرارة والدخل وغيرها من أنواع البيانات الكمية .. كما يمكن أن يتم ذلك من خلال الاستفتاءات الميدانية والاستبانات والجداول الاحصائية الرسمية وغير الرسمية.

مصادر جمع البيانات الإحصائية:

- <u>المصادر الأولية (المباشرة أو الميدانية)</u>: هي البيانات التي يجمعها الباحث بنفسه من عيّنات البحث، وهو أكثر دقّة من المصدر الآخر لكنه يستهلك الكثير من الجهد والوقت والمال.
- <u>المصادر الثانوية (غير المباشرة أو التاريخية)</u>: هي البيانات التي يتمّ الحصول عليها بشكلٍ غير مباشر من جهات معيّنة كاستخدام الدراسات السابقة وهو أقلّ دقّة لكنه يوفّر الوقت والجهد والمال.

المجتمع Population: عدد لا نهائي من الأفراد والعناصر التي تتعايش مع بعضها البعض وتتميّز بخصائص ومواصفات تميّزها عن بقيّة المجتمعات حيث أنّه لكل مجتمع خصائص ومميزات تميّزه عن المجتمعات الأخرى.

العيّنة Sample: هي جزء من المجتمع يجب ألّا يقلّ عدد أفرادها عن ٢ – ١٠ % من عدد أفراد المجتمع، وهي تُؤخَذ بطريقة عشوائية.

أساليب جمع البيانات:

- أسلوب الحصر الشامل: يتمّ في هذا الأسلوب دراسة كل فرد خاضع للبحث من دون استثناءات ممّا يجعله دقيقاً جداً وواقعي لكنه يحتاج للكثير من الوقت والجهد والمال.
- أسلوب المعاينة: يتمّ في هذا الأسلوب دراسة مجموعة صغيرة مختارة بأسس علمية ثمّ تعميم النتائج على المجتمع ككل ممّا يجعله غير دقيق ولكنه يوفر الوقت والجهد والمال.

أدوات جمع البيانات:

إذا اختار الباحث طريقة جمع البيانات من المصدر المباشر، يمكنه اختيار أحد الأسلوبين التاليين:

- الأسلوب الميداني المباشر: ويتطلّب هذا الأسلوب الاتصال المباشر للباحث أو من ينوب عنه بعناصر المجتمع أو أفراد العينة، ومن الأدوات التي يستخدمها هذا الأسلوب: الاستبيان، المقابلة، الملاحظة...
- أسلوب المراسلة غير المباشر: يتطلّب هذا الأسلوب إرسال الإستمارة لعناصر المجتمع أو أفراد العينة من خلال البريد أو الهاتف أو الفاكس أو الإنترنت....

مراجعة وتدقيق البيانات:

عند إتمام عملية جمع البيانات وفق الوسيلة المناسبة لذلك يجب أن تتمّ مراجعة وتدقيق البيانات للتأكد من مطابقتها وتكاملها مع متطلّبات الدراسة فعلى سبيل المثال لو كانت الوسيلة المستخدمة في جمع البيانات هي الاستبيان عندئذٍ يتوجّب مراجعة وتدقيق الإستمارات الإحصائية التي تمّ جمعها واستبعاد ما هو غير متكامل أو واضح أو دقيق وعزل الاستمارات التي يعتقد الباحث أنّها غير مطابقة لما هو مطلوب. ومن الأمور التي يتمّ التركيز عليها في هذه العملية ما يلي:

- عدم وضوح الكتابة، في هذه الحالة يُراجَع المبحوث للاستيضاح وفي حالة التعذّر تُلغَى اجابته.
- مدى توخّي الدقّة من المبحوث في اجابته، ويُلاحَظ ذلك من خلال التعارض في الاجابة، والاجابات النمطية إلخ....
- ترك بعض الأسئلة دون اجابة.

طرائق عرض البيانات الإحصائية

Methods of graphical data view

بعد الانتهاء من جمع البيانات تُستخدَم طرق عديدة لعرضها منها البسيطة ومنها المعقّدة، هذه الطرق سنقوم بذكرها فيما يلي:

١. **الجداول البسيطة:** عند استخدام هذا النوع من الجداول يجب مراعاة ما يلي:

- رقم الجدول: حيث يُخصَّص لكل جدول رقم خاص به يختلف عن الرقم الذي يسبقه وعن الرقم الذي يليه .. مع مراعاة أن تكون أرقام هذه الجداول مرتّبة بشكل تسلسلي تصاعدي.
- عنوان الجدول: حيث يُخصَّص لكل جدول عنوان يوضِّح بشكلٍ مختصر طبيعة المعلومات المُدوَّنة في الجدول.
- وحدة القياس: يجب أن تُوضَع الوحدات التي تُقَاس بها المعلومات المدونة في الجدول سواءً أعلى أو أسفل الجدول.
- المصدر: يتمّ ذكر المصدر الذي أُخِذَت منه المعلومات ويتمّ وضعه أسفل الجدول عادةً.
- تقريب الأرقام: يجب أن تكون الأرقام الموضوعة في الجدول من طرازٍ واحد كأن تكون كلّها صحيحة إن أمكن أو مقرَّبة لرقمين بعد الفاصلة أو ثلاثة.. إلخ، وذلك حسب الحاجة ونوع المعلومات.

مثال على الجداول البسيطة:

Table (1): Development of Arable & Cultivated Lands in Syria during (2007 -2014).

الجدول (١): تطوّر الأراضي القابلة للزراعة والمُستثمَرَة في سورية خلال الفترة (٢٠٠٧ - ٢٠١٤).

Area: 1000 Hec **المساحة: ألف هيكتار**

Cultivated Lands أراضي مُستثمَرَة	Arable Lands أراضي قابلة للزراعة	Years السنوات
٥٦٨٢	6039	2007
٥٦٦٦	6024	2008
٥٦٦٤	6012	2009
٥٦٩٦	6045	2010
٥٧١٦	6068	2011
٥٧٣١	٦٠٧٩	2012
٥٧٣٣	٦٠٨٣	2013
٥٧٣٢	٦٠٨١	2014

المصدر: المجموعة الإحصائية الزراعية السنوية.

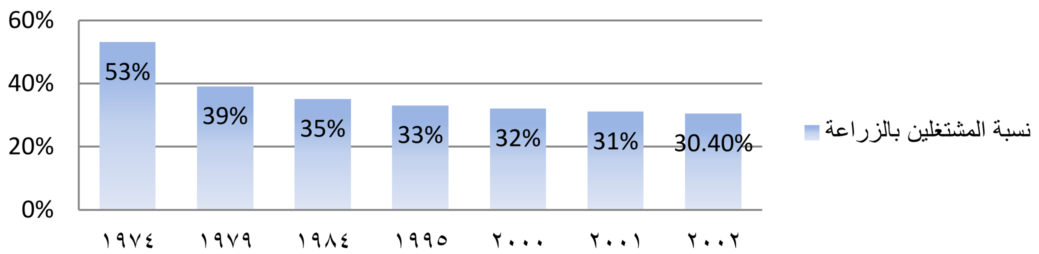
٢. **الجداول التكرارية (الفئوية أو المبوّبة):** هي جداول خاصّة لها أهمية كبيرة في الإحصاء، وتعتبر أفضل من النوع السابق من الجداول حيث تعطينا معلومات أكثر وضوحاً وتعبيراً عن البيانات المدرجة فيها. وهي تتألّف من عدد من الأسطر والأعمدة. تشمل الأعمدة ما يلي:

- العمود الأول يمثّل الحدود العليا والدنيا للفئات.
- العمود الثاني يمثّل مراكز الفئات.
- العمود الثالث يمثّل تكرارات الفئات.
- العمود الرابع يمثّل التكرارات المتجمّعة الصاعدة للفئات.
- العمود الخامس يمثّل التكرارات المتجمّعة الهابطة للفئات.
- العمود السادس يمثّل أيّة معلومات توضيحية إضافية يراها الباحث مناسبة.

٣. **الأعمدة (الأشرطة) البيانية:** هي عبارة عن أعمدة أفقية أو عمودية ذات عرض واحد وطول يختلف تبعاً للمتغيّر الذي تمثّله، وتنقسم هذه الأشرطة البيانية إلى مجموعتين:

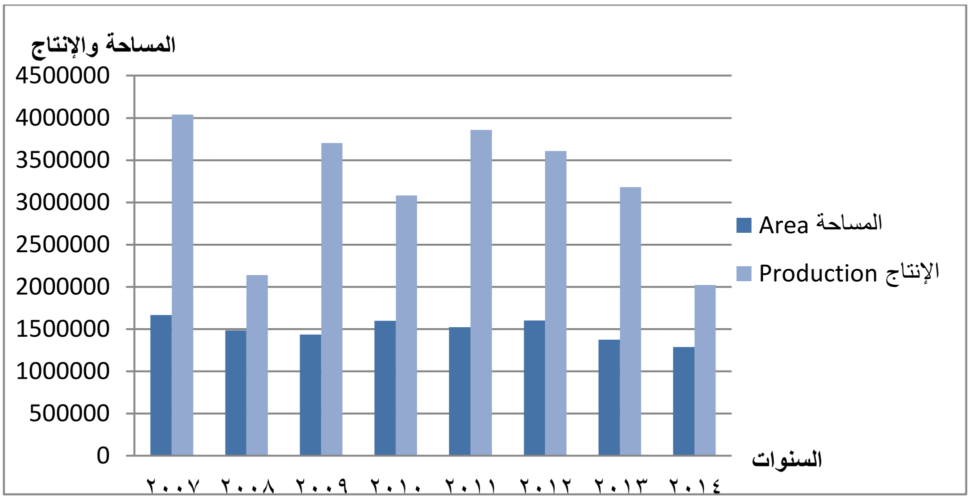
أ. الأشرطة البيانية البسيطة: هذا النوع من الأشكال البيانية يستخدم للتمثيل البياني عندما تحتوي البيانات متغير واحد مع الزمان أو المكان، كما في المثال التالي:

المخطط (١): تغيّر نسبة المشتغلين بالزراعة في سورية بين عامي ١٩٧٤ ـ ٢٠٠٢ م.



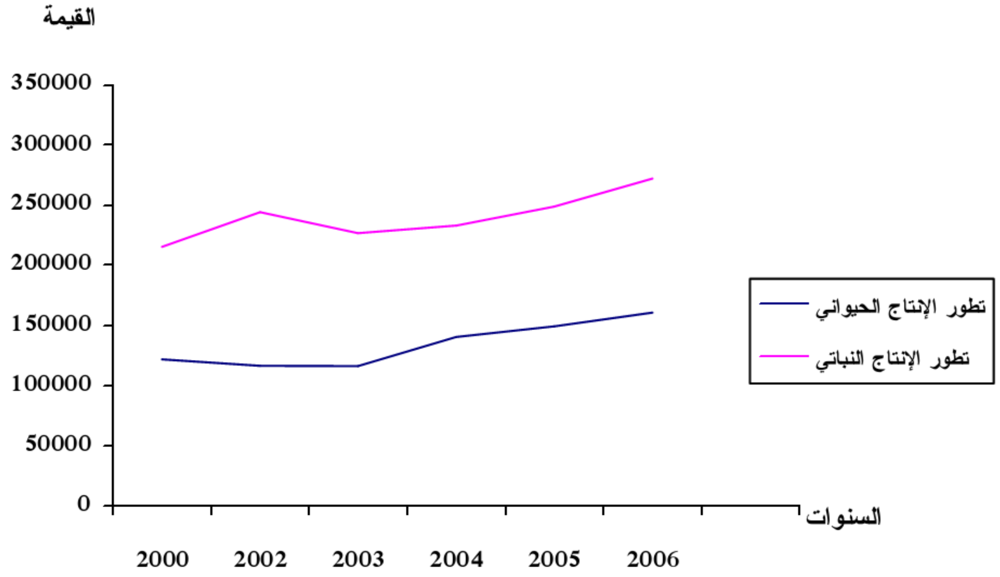
ب. الأشرطة البيانية المركّبة: هذا النوع من الأشكال البيانية يستخدم للتمثيل البياني عندما تحتوي البيانات أكثر من متغير مع الزمان أو المكان، كما في المثال التالي:

المخطط (٢): تطوّر كل من إنتاج القمح والمساحات المزروعة به في سورية خلال الفترة (٢٠٠٧ – ٢٠١٤) م.



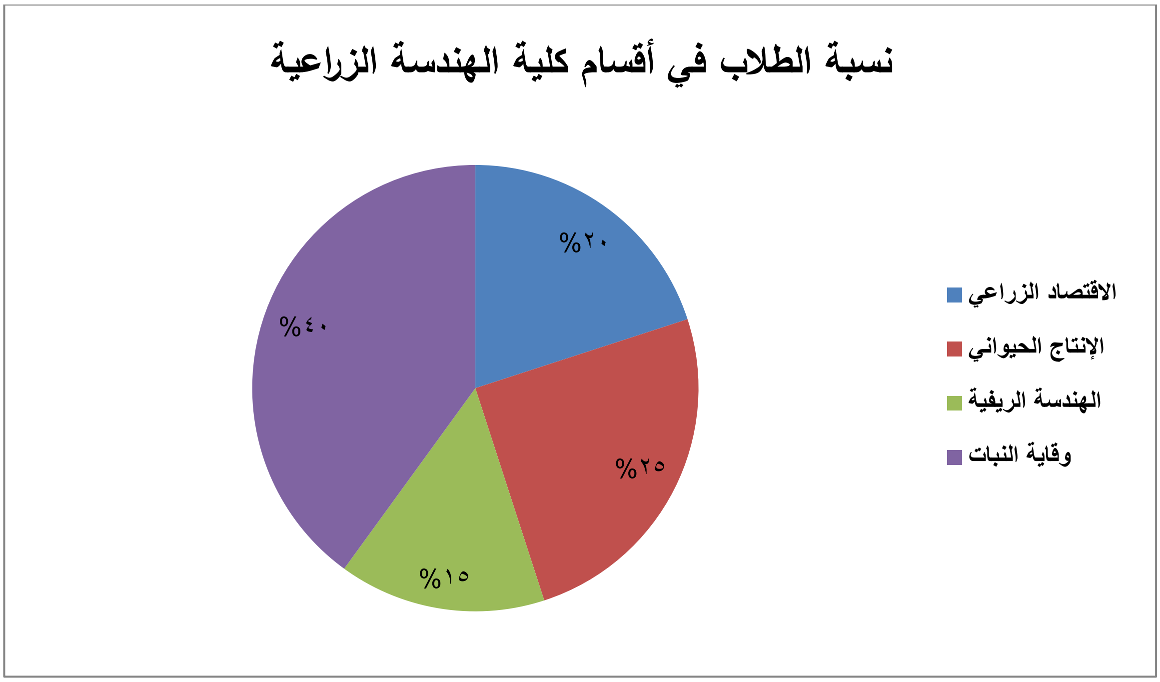
٤. **المنحنيات البيانية (المضلع والمنحني التكراري):** هو عبارة عن خط يمثّل العلاقة بين أحد المتغيرات مع متغير آخر مثل الزمن أو المكان. كما في المثال التالي:

المخطط (٣): تطوّر قيمة الإنتاج النباتي والحيواني في سورية بأسعار عام ٢٠٠٠ الثابتة/مليون ليرة سوريّة.



٥. **الرسوم البيانية القطاعية:** هي إحدى طرق العرض البياني شائعة الاستخدام، حيث يتمّ فيها تمثيل البيانات الإحصائية باستخدام دائرة تمثّل المجموع الكلي للبيانات، ومن ثمّ تُقسَم الدائرة إلى أجزاء أو شطائر تمثّل نسبة كل قيمة من القيم الإحصائية إلى المجموع الكلي للبيانات الإحصائية كنسبة مئوية.

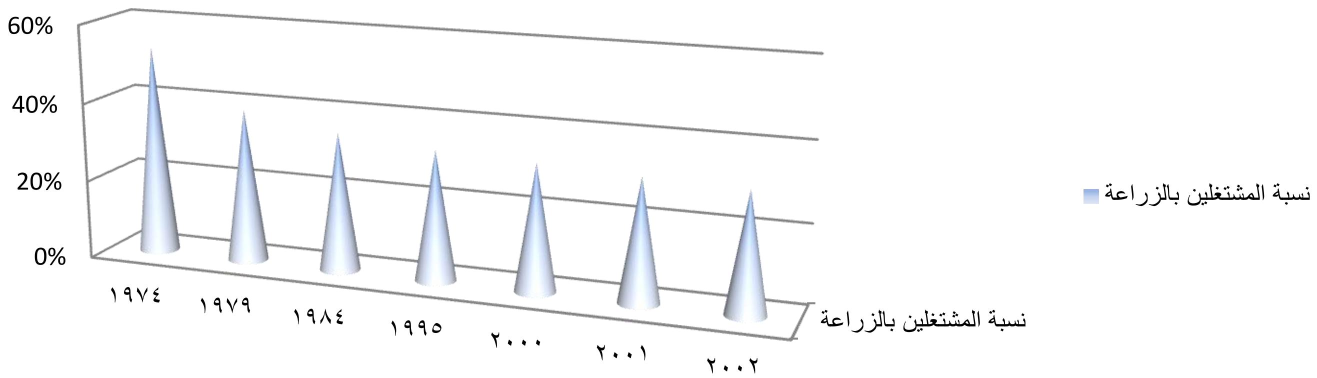
المخطط (٤): نسبة أعداد الطلاب في الأقسام المختلفة من كلية الهندسة الزراعية في إحدى السنوات إلى العدد الكلي.



٦. **الرسوم البيانية المساحية:** هي عبارة عن أشكال هندسية مثل المربع أو المستطيل وغيرهما.. بحيث تستخدم مساحة هذه الأشكال لتمثيل قيم المتغير المدروس، ويتمّ حساب مساحة الشكل الهندسي من خلال حساب طول الضلع، وذلك حسب الشكل الهندسي وفقاً للدستور الرياضي المناسب.

٧. **الرسوم البيانية الحجمية:** هي عبارة عن أشكال هندسية مثل المكعّب والاسطوانة والمخروط وغيرها.. بحيث يستخدم حجم هذه الأشكال لتمثيل قيم المتغير المدروس، ويتمّ حساب حجم الشكل الهندسي من خلال حساب طول الضلع، وذلك حسب الشكل الهندسي وفقاً للدستور الرياضي المناسب، كما في المخطط التالي:

المخطط (٥): تغيّر نسبة المشتغلين بالزراعة في سورية بين عامي ١٩٧٤ - ٢٠٠٢ م.



٨. **الخرائط البيانية:** وهي عبارة عن خرائط جغرافية من حيث المبدأ تستخدم أشكال أو رموز مختلفة للتعبير عن قيم المتغير.

مقاييس النزعة المركزية

Measures of Central Tendency

تسلك الظواهر الطبيعية سلوكاً محدداً من حيث توزّع أفرادها، حيث نلاحظ أن عدداً كبيراً من القيم يميل إلى القيمة الوسط أو ما يمكن أن نسميه المركز Central Value. هذه الظاهرة تُسمَّى **النزعة المركزية Central Tendency** وهي ميّزة تميّز كل المجتمعات الطبيعية وتعني نزوع الأفراد في كل مجتمع من المجتمعات الطبيعية إلى التجمّع حول المركز. وللتعرف أكثر على خصائص هذه الظاهرة لابدَّ من توافر معايير أو مقاييس معينة تعطينا فكرة بسيطة عن الظاهرة المدروسة. وهذه المعايير ستكون موضوع جلستنا العملية.

١. **المتوسّط الحسابي Arithmetic Main**: يُعرَّف المتوسط الحسابي بأنه عبارة عن حاصل قسمة مجموع عناصر العينة على عدد عناصر هذه العينة. ويُرمَز للمتوسّط عادةً بالرمز $\overline{X}$ وتُقرَأ x_bar. ويُحسَب المتوسط الحسابي عادةً بالطريقة الآتية:

أ. عندما تكون عناصر العينة غير مكررة، يُحسَب المتوسط الحسابي من خلال تطبيق الدستور العام:

$$\overline{X}= \frac{\sum_{i=1}^{n} Xi}{N}$$

$$\overline{X}= (X1+X2+....Xn)/N$$

مثال: لدينا العينة X = {3، ٥، ٦، ٩، ١٢} والتي تمثّل عدد الأزهار على نبات الورد الجوري بحيث تضمّ هذه العينة خمس نباتات مختارة بشكل عشوائي من إحدى الحدائق في مدينة سلمية. والمطلوب حساب المتوسط الحسابي لهذه العينة بالطريقة المباشرة.

الحل: نطبّق الدستور المذكور أعلاه فيكون لدينا:

$$\overline{X}= (3+5+6+9+12)/5= 35/5=7$$

ب. عندما يكون واحد من عناصر العينة على الأقل مكرّراً أكثر من مرة في العينة، فإن المتوسط الحسابي يحسب من خلال تطبيق الدستور الآتي:

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

٢. الوسيط Median: يعرّف الوسيط بأنّه القيمة من العينة المدروسة التي تقع في الوسط بعد ترتيب عناصر العينة ترتيباً تصاعدياً أو تنازلياً. وهذا يعني أنّ الوسيط هو تلك القيمة التي يكون عدد القيم التي يسبقها مساوٍ لعدد القيم التي تليها وذلك بعد ترتيب عناصر العينة ترتيباً تصاعدياً أو تنازلياً. ويتمّ عادةً حساب الوسيط للعينة المدروسة طبقاً لإحدى الحالتين:

- **عندما يكون عدد عناصر العينة فردياً:** وفي هذه الحالة تكون قيمة الوسيط مساوية لقيمة العنصر الواقع في المكان الوسط أي العنصر ذو الترتيب $(N+1)/2$ بعد ترتيب عناصر العينة ترتيباً تصاعدياً أو تنازلياً.

مثال: إذا كانت لدينا العينة الإحصائية الآتية:

$$\mathbf{X}= \{3, 5, 1, 9, 2, 7, 4\}$$

والمطلوب حساب الوسيط للعينة المدروسة.

نرتّب عناصر العينة ترتيباً تصاعدياً فتصبح على الشكل الآتي:

X: 1 2 3 4 5 7 9

فيكون الوسيط هو القيمة ذات الترتيب (٧+١)/٢=٤ أي القيمة ذات الترتيب الرابع بعد ترتيب عناصر العينة ترتيباً تصاعدياً.

- **عندما يكون عدد عناصر العينة زوجياً:** في هذه الحالة تكون قيمة الوسيط مساوية لمتوسّط قيمتي العنصرين ذو الترتيب N والعنصر ذو الترتيب $\frac{N}{2}+1$ بعد ترتيب عناصر العينة ترتيباً تصاعدياً أو تنازلياً.

مثال: إذا كانت لدينا العينة الإحصائية الآتية:

$$\mathbf{X}= \{3, 5, 6, 4, 2, 9\}$$

والمطلوب حساب الوسيط للعينة المدروسة.

نرتّب عناصر العينة ترتيباً تنازلياً فتصبح على الشكل الآتي:

X: 9 6 5 4 3 2

العنصر ذو الترتيب N/٢ هو الرقم ٤ أمّا العنصر ذو الترتيب $\frac{N}{2}+1$ هو الرقم ٥ وبالتالي المتوسط الحسابي لكلا القيمتين هو ٤.٥ =٢/(٥+٤) إذاً الوسيط للعينة المدروسة هو ٤.٥

٣. **المنوال (القمّة) Mode:** يُعرَّف المنوال بأنه القيمة الأكثر تكراراً من بين عناصر العينة أو المجتمع المدروس. وتجدر الملاحظة هنا أنّه في الحالة الخاصة التي تحصل عندما يكون لدينا عينة قيد الدراسة تحوي مجموعة من العناصر غير المتكررة فإنّ المنوال في هذه الحالة يعتبر غير موجود.

لحساب المنوال للعينة المدروسة في عينة غير موزعة في جدول تكراري:

مثال: لدينا عينة قيد الدراسة تحوي مجموعة من العناصر هي التالية:

$$\mathbf{X} = \{3, 5, 6, 11, 9, 11, 12, 21\}$$

والمطلوب حساب المنوال للعينة المذكورة.

الحل: إذا دققنا النظر في عناصر العينة قيد الدراسة نجد أن المنوال هو القيمة ١١ لأنّها القيمة الأكثر تكراراً بين عناصر العينة المدروسة.

مثال: لدينا عينة قيد الدراسة تحوي مجموعة من العناصر هي التالية:

$$\mathbf{X} = \{3, 5, 6, 9, 11, 12, 21\}$$

والمطلوب حساب المنوال للعينة المذكورة.

الحل: إذا دققنا النظر في عناصر العينة قيد الدراسة نستنتج عدم وجود منوال بسبب عدم وجود عنصر متكرر أكثر من مرة وبالتالي نقول أن هذه العينة لا تحوي منوال.

مقاييس التشتّت والاختلاف

Measures of Dispersion and Variation

قمنا في الجلسة العملية السابقة بدراسة مقاييس النزعة المركزية ووجدنا أنها تعطينا فكرة جيدة ووافية عن عناصر العينة المدروسة ومدى نزعتها للتجمع حول المتوسط، لكن في حالة تساوي قيمة المتوسط لعينتين بالرغم من اختلافهما من حيث قيم العناصر نحتاج لكي نفرق بين العينتين إلى وجود مؤشرات خاصة لاستخدامها في حال تساوي قيمة المتوسط لعينتين وأكثر هذه المؤشرات استخداماً هي مقاييس التشتت والاختلاف.

١. **المدى Range**: هو الفرق بين أكبر وأصغر قيمة في العينة المدروسة، وفي جدول التوزيع التكراري يتمّ حساب المدى على أنه الفرق بين الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة الأولى.

Range = Max (X) – Min (X)

٢. **التباين Variance**: يعتبر التباين أهم مقياس من مقاييس التشتت والمستخدمة على نطاق واسع جداً خاصة في تحليل البيانات التجريبية للاختبارات والتجارب الزراعية، وهو يُعطَى بالعلاقة التالية:

$$S^2 = \frac{\sum_{i=1}^{n}(Xi-\bar{X})^2}{N-1}$$

كما يمكن حساب التباين Variance بطريقة أخرى يطلق عليها طريقة تربيع القيم، وهي كما يلي:

$$S^2 = \frac{\sum_{i=1}^{n} Xi^2 - \frac{\left(\sum_{i=1}^{n} Xi\right)^2}{N}}{N-1}$$

ومن الجدير بالذكر أن الطريقة الأخيرة (طريقة تربيع القيم) تستخدم عندما تكون قيم العينة قيم كسرية.

أمّا إذا أردنا حساب التباين في جدول توزيع تكراري:

$$S^2 = \frac{\sum_{i=1}^{n} Fi(Xi-\overline{X})^2}{N-1}$$

حيث: Xi: هو مركز الفئة.

Fi: هو تكرار الفئة.

N: هو مجموع التكرارات.

مثال: أوجد التشتّت أو التباين والانحراف المعياري للعينة التالية باستخدام الطريقة المناسبة:

$$X= \{5, 6, 7, 1, 3, 2\}$$

الحل: يمكن الاستعانة بإحدى العلاقتين السابقتين آنفتي الذكر، ولكن من أجل ذلك علينا إنشاء الجدول المبسّط التالي:

X^2	$(Xi-\overline{X})^2$	$(Xi-\overline{X})$	X
٢٥	١	١	٥
٣٦	٤	٢	٦
٤٩	٩	٣	٧
١	٩	٣ -	١
٩	١	١ -	٣
٤	٤	٢ -	٢
١٢٤	٢٨	٠	٢٤

الآن يمكننا تطبيق إحدى العلاقتين السابقتين للحصول على التباين كما يلي:

$$S^2 = \frac{28}{6-1} = 5.6$$

أمّا ٣. **الانحراف المعياري** للعينة فهو سهل الحساب باعتباره الجذر الموجب للتباين، فإذاً هو يساوي:

$$S = +\sqrt{S^2} = + \sqrt{5.6} = 2.37$$

من الجدير بالذكر أنه يمكن الملاحظة بسهولة أن الفرق بسيط بين قانون تباين العينة وقانون تباين المجتمع.

عند تساوي المتوسطات يستخدم عادةً الانحراف المعياري كأساس للمقارنة بين عينتين أو أكثر شرط أن تكون العينات مأخوذة من مجتمع واحد، وبالتالي فالعينة ذات الانحراف المعياري الأقل هي التي تكون أكثر استقراراً من غيرها وعناصرها أقل تبعثراً وأكثر تجانساً.

أمَّا إذا لم يتحقق الشرط المذكور آنفاً أي عندما لا تكون العينات مأخوذة من مجتمع واحد فإنّ الانحراف المعياري لا يمكن اتخاذه كأساس للمقارنة بين عينتين أو أكثر.

٤. عامل الاختلاف Coefficient of Variation: ويُستخدَم عادةً مقياس آخر يطلق عليه يُرمَز له بِ C. V وهو يعرف بأنّه أحد مقاييس التشتت أو التباين الهامة والذي يستخدم بشكل عام للمقارنة بين عينتين أو أكثر عندما لا تكون العينات المأخوذة من مجتمع معين بل من مجتمعات مختلفة، وهو أي عامل الاختلاف يُعطَى بالعلاقة:

$$\mathbf{C.\ V} = \frac{S}{\bar{X}} * 100$$

C. V: عامل الاختلاف.

S: الانحراف المعياري.

$\bar{X}$: متوسّط العينة.

من الجدير بالذكر أنّه كلما كانت قيمة عامل الاختلاف C. V أصغر كلما دلّ ذلك على ثبات العينة وتجانس أفرادها وكلما كانت عناصرها أقلّ تبعثراً مقارنةً بالعينة الأخرى الأكبر بقيمة عامل الاختلاف.

الارتباط Correlation

الارتباط لغوياً يعني التلازم، ويمكن تعريفه: بأنّه تلازم صفتين تلازماً كمياً بحيث أن تغيّر الصفة الأولى يؤدي إلى تغير الصفة الثانية، ويكون التغير بالطبع بالزيادة أو النقصان. ولا يعدُّ وجود ارتباط بين المتغيرين مؤشّراً على أن أحدهما يحدث نتيجةً لتأثير المتغير الآخر، ولا يحدث إلّا في وجوده أو بسببه، حيث يمكن أن تكون هناك عوامل تؤثّر على المتغيرين معاً فتجعل التغير الذي يحدث في أحدهما كأنه ناتج من التغير الذي حدث في المتغير الثاني.

من تعريف الارتباط نلاحظ وجود عدة احتمالات:

١. زيادة الصفة أو الظاهرة الأولى تقود إلى زيادة الصفة أو الظاهرة الثانية، وبالتالي نقول أن لدينا علاقة ارتباط طردية بين الظاهرتين وأن الارتباط موجب أي أن زيادة الصفة الأولى أدّت إلى زيادة الصفة الثانية.

مثال: زيادة كمية السماد المضافة تؤدي إلى زيادة الإنتاجية وبالتالي لدينا علاقة طردية والارتباط موجب.

أو: زيادة كمية مياه الري تؤدي إلى زيادة الإنتاجية وبالتالي لدينا علاقة طردية وارتباط موجب.

٢. زيادة الصفة أو الظاهرة الأولى تقود إلى نقصان الصفة أو الظاهرة الثانية، وبالتالي نقول أن لدينا علاقة ارتباط عكسية بين الظاهرتين وأن الارتباط سالب أي أن زيادة الصفة الأولى أدّت إلى نقصان الصفة الثانية.

مثال: زيادة الإصابة بالأمراض والحشرات تؤدي إلى نقصان الإنتاجية وبالتي لدينا علاقة عكسية والارتباط سالب.

٣. نقصان الصفة أو الظاهرة الأولى تقود إلى نقصان الصفة أو الظاهرة الثانية، وبالتالي نقول أن لدينا علاقة ارتباط طردية بين الظاهرتين وأن الارتباط موجب أي أن نقصان الصفة الأولى أدّت إلى نقصان الصفة الثانية.

مثال: نقصان كميّة البذار تؤدي إلى نقصان الإنتاجية وبالتالي لدينا علاقة طردية والارتباط موجب.

أو: نقصان كمية الأعلاف المقدمة للحيوانات الزراعية تؤدي إلى نقصان الإنتاجية وبالتالي لدينا علاقة طردية والارتباط موجب.

٤. نقصان الصفة أو الظاهرة الأولى تقود إلى زيادة الصفة أو الظاهرة الثانية، وبالتالي نقول أن لدينا علاقة ارتباط عكسية بين الظاهرتين وأن الارتباط سالب أي أن نقصان الصفة الأولى أدّت إلى زيادة الصفة الثانية.

مثال: نقصان تكاليف الإنتاج تؤدي إلى زيادة الربح وبالتالي لدينا علاقة عكسية والارتباط سالب.

كل ما تكلّمنا عنه سابقاً يدعى الارتباط البسيط Simple Correlation حيث ندرس العلاقة بين متغيرين فقط، وهو ما سنركّز عليه في دراستنا لأهميته وبساطته مقارنةً بأنواع الارتباط الأخرى.

أنواع الارتباط: تشمل ما يلي:

١. **الارتباط البسيط:** يستخدم الارتباط البسيط لدراسة العلاقة بين متغيرين احتماليين فقط.
٢. **ارتباط الرتب:** يستخدم ارتباط الرتب لدراسة العلاقة بين المتغيرات الاحتمالية عندما تكون بيانات هذه العوامل مرتّبة أو يمكن ترتيبها في رتب معينة.
٣. **الارتباط الجزئي:** يستخدم الارتباط الجزئي لدراسة العلاقة بين المتغيرات الاحتمالية عندما يكون لدينا عدة متغيرات ونرغب بحساب الارتباط بين متغيرين محددين علماً بأن الارتباط مع بقية المتغيرات معلومة.
٤. **الارتباط المتعدد:** يستخدم الارتباط المتعدد لدراسة العلاقة بين المتغيرات الاحتمالية عندما يكون لدينا عدة متغيرات ونرغب بحساب الارتباط بين متغير محدد مع بقية المتغيرات.

تختلف العلاقة بين متغيرين من حيث قوّتها، فإذا كان تغير أحد هذين المتغيرين يعتمد كلياً على تغير الآخر نقول أن الارتباط بينهما كامل Perfect Correlation مثل العلاقة بين مساحة الدائرة ونصف قطرها، أمّا إذا كان الارتباط بين المتغيرات غير كامل أي أن تغير أحدهما لا يعتمد كلياً على تغير الآخر، فنقول بأن الارتباط هو ارتباط غير تام مثل العلاقة بين وزن الفرد وطوله، وبين التحصيل العلمي وعدد ساعات الدراسة.

معامل الارتباط r:

معامل الارتباط هو قيمة عددية تقيس شدة الارتباط ويُرمَز له بالرمز r إلّا إذا أراد الباحث أن يرمز لمعامل الارتباط برمز آخر غير r. ندرس معامل الارتباط لكي نحكم على الارتباط هل هو جيد أو سيء أو قوي أو ضعيف.

$$r \in [-1, +1]$$

فإذا كان: **r = +1** فإنّ العلاقة طردية لأن القيمة موجبة والارتباط موجب تام. أي أنّ الظاهرة الأولى عند زيادتها سوف تزيد الظاهرة الثانية وبنفس النسبة.

أمّا إذا كان: **r = -1** فإنّ العلاقة عكسية لأن القيمة سالبة والارتباط سالب تام. أي أنّ الظاهرة الأولى عند نقصانها سوف تزيد الظاهرة الثانية وبنفس النسبة. إنّ هاتين الحالتين السابقتين نادرتا الحدوث.

إذا كان: **r = 0** فإنّ العلاقة معدومة أي أن الارتباط معدوم ولا توجد علاقة بين المتغيرين المدروسين.

تتحدّد نوعية الارتباط بالاعتماد على الجدول (٢) الآتي:

الجدول (٢): أنواع الارتباط اعتماداً على قيم معامل الارتباط.

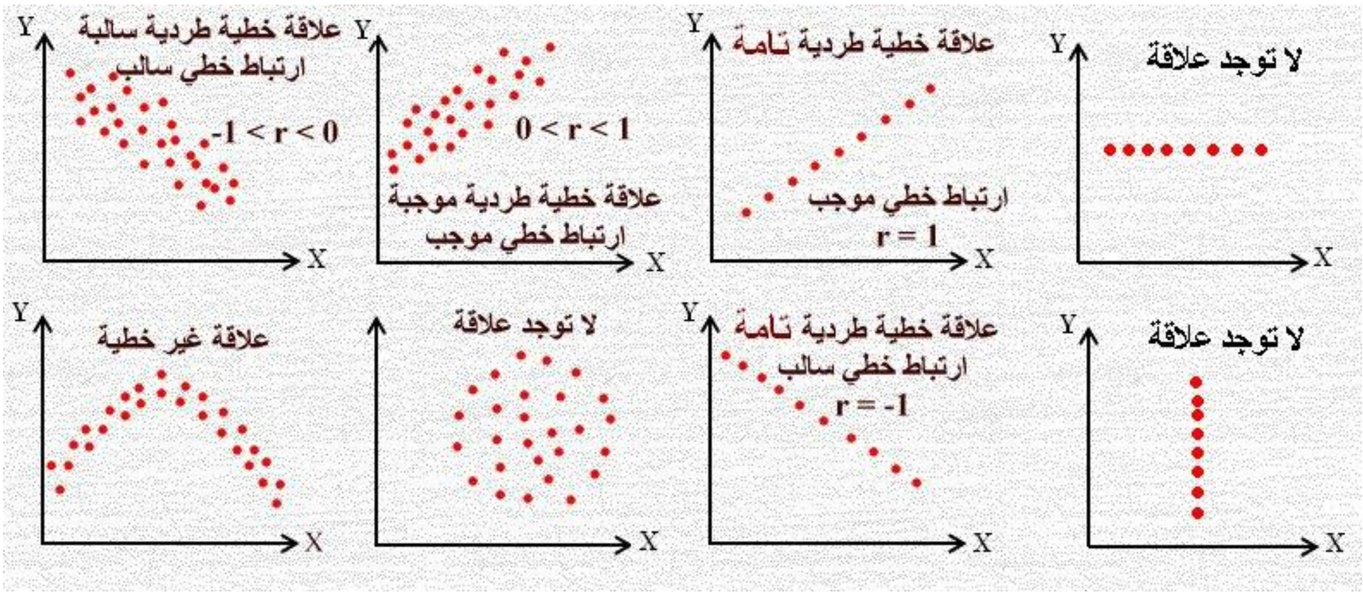
قيمة معامل الارتباط	نوع الارتباط
+1	ارتباط طردي تام
]1,7]	ارتباط طردي قوي
]7,4]	ارتباط طردي متوسط
]4,0[	ارتباط طردي ضعيف
0	ارتباط معدوم
]0,−0.4[	ارتباط عكسي ضعيف
[−0.4,−0.7[	ارتباط عكسي متوسط
[−0.7,−1[	ارتباط عكسي قوي
−1	ارتباط عكسي تام

الجدير بالذكر أن معامل الارتباط للظواهر الطبيعية بوجهٍ عام والزراعية بوجهٍ خاص لا يمكن أن تصل قيمته إلى الواحد بالقيمة المطلقة.

يجدر بالذكر أيضاً أن الارتباط البسيط قد يكون خطياً أو غير خطي، ويكون خطياً عندما يقترب شكل الانتشار الذي يمثّل قيم المتغيرين محلّ البحث من الخط المستقيم، وتكون قيمة معامل الارتباط أكبر كلما اقترب شكل الانتشار من الخط المستقيم.

يمكننا أن نحدّد شكل الارتباط بين ظاهرتين أو صفتين $(y،x)$ اعتماداً على الرسم البياني في المستوي، كما هو مبيّن في المخطط (٦) الآتي:

المخطط (٦): أشكال الارتباط المختلفة بين ظاهرتين أو صفتين $(y،x)$ اعتماداً على الرسم البياني في المستوي.



فإذا كان اتجاه الشكل الذي يحوي النقاط باتجاه الربع الأول وللأعلى بشكلٍ تام نقول أنّه لدينا ارتباط موجب حيث تزداد قيمة $(y،x)$ وكلّما ضاق المجال كلّما كان الارتباط أقوى.

أمّا إذا كان اتجاه الشكل الذي يحوي النقاط باتجاه الربع الثاني وللأعلى بشكلٍ تام نقول أنّه لدينا ارتباط سالب حيث تتناقص قيمة $(y،x)$ وكلّما ضاق المجال كلّما كان الارتباط أقوى.

حساب معامل الارتباط:

مهما تكن x و y فإنّ r يُحسَب من العلاقة الآتية:

$$r = \frac{\sum_{i=1}^{n}(xi - \bar{x}) * (yi - \bar{y})}{\sqrt{\sum_{i=1}^{n}(xi - \bar{x})^2 * \sum_{i=1}^{n}(yi - \bar{y})^2}}$$

تساؤل: ماذا نستفيد من معامل الارتباط؟

يفيدنا معامل الارتباط في حساب النسب المئوية التي تؤثّر بها إحدى الظاهرتين على الأخرى، وذلك عن طريق تربيع معامل الارتباط.

يُرمز لمربع معامل الارتباط بـ B. حيث أنّ: $B = r^2$ ويُسمَّى المعامل B **بمعامل التحديد.**

الانحدار Regression

معادلة الانحدار أو نموذج الانحدار يمكن تعريفه بأنّه علاقة رياضية تربط بين متحولين أو متغيرين أحدهما متغير تابع يكون في الطرف الأيسر من المعادلة والآخر يدعى متغير مستقل يكون في الطرف الأيمن من المعادلة.

يُعدُّ الانحدار أحد أهم الأساليب الإحصائية ويختصّ بقياس العلاقة بين متغير يسمّى المتغير التابع ومتغير آخر أو مجموعة متغيرات تسمّى المتغيرات المستقلة.

يجدر بالذكر أن نماذج الانحدار تُقسَم إلى:

أ. **انحدار خطي بسيط:** (والذي سنقوم بالتركيز عليه لأهميته وسهولة تقديره).

$$y = a + bx$$

ب. **انحدار خطي متعدد:**

$$y = a + b_1x_1 + b_2x_2 + b_3x_3$$

ج. **انحدار غير خطي بسيط:**

$$y = a + bx^2$$

د. **انحدار غير خطي متعدد:**

$$y = a + b_1x_1 + b_2{x_2}^2 + b_3{x_3}^3$$

عندما أريد دراسة ظاهرة أو صفة معينة (متغير تابع) كإنتاج نبات معيّن، بحيث يتأثر هذا الإنتاج بمجموعة من العوامل أو المتغيّرات المستقلّة المتفاوتة في تأثيرها على الإنتاج مثل كمية مياه الري، أو كمية السماد أو عدد العمال أو كمية الإصابات الحشرية والأمراض النباتية...إلخ. نقوم من أجل تسهيل الدراسة وتبسيطها بدراسة العلاقة بين الصفة المدروسة (المتغيّر التابع) وأهم العوامل المذكورة مثلاً (المتغيّر المستقل) وليكن كمية مياه الري على سبيل المثال. وبالتالي يمكن تمثيل ما سبق بالعلاقة الآتية:

$$y = a + bx$$

حيث تمثّل y كميّة الإنتاج من نبات معين، أمّا x فتمثّل كمية المياه المستخدمة في الري.

الانحدار الخطّي البسيط Simple Linear Regression:

هو أبسط أنواع الانحدار، والشكل العام لمعادلته (علاقة الانحدار الخطي): $y = a + bx$

y: الانحدار الخطي البسيط أو العامل أو المتغير. x : عامل مستقل.

a معامل ثابت (ثابت الانحدار): هو رقم ثابت، وهو الجزء المقطوع الذي يختلف حسب الدالّة.

مثلاً: في دالة الإنتاج: يمثّل الحدّ الأدنى من الإنتاج عندما تكون كمية المياه أو السماد المضاف تساوي الصفر.

في دالة العرض: يمثّل الحدّ الأدنى من الكمية المعروضة لو كان السعر صفراً.

في دالة الطلب: يمثّل الحد الأعلى من الكمية المطلوبة لو كان السعر صفراً.

b: معامل الانحدار (المعلمة الانحدارية): وتمثّل ميل الخط المستقيم الممثّل للعلاقة، وتمثّل المعامل الحدّي.

وتشير لمقدار التغيّر في المتغير التابع y نتيجة التغيّر بمقدار وحدة واحدة من المتغير المستقل x ، فكلما تغيّر العامل x بمقدار وحدة كاملة يتغيّر العامل y بمقدار وحدة كاملة، وقد يكون التغيّر إمّا بالزيادة أو بالنقصان.

عندما b **موجب** هذا يعني أن العامل المستقل يؤثر إيجاباً على العامل التابع.

عندما b **سالب** هذا يعني أن العامل المستقل يؤثر سلباً عل العامل التابع.

عند تتغيّر قيمة x بمقدار وحدة كاملة فإنّ قيمة y تتغير بمقدار b والذي هو معامل الانحدار.

علماً بأنّ:

$$a \in]-\infty, +\infty[$$

$$b \in]-\infty, +\infty[$$

أي أنّ a و b تأخذان جميع القيم السالبة والموجبة والكسرية والصحيحة.

قوانين معامل الانحدار b:

$$b = \frac{\sum_{i=1}^{n}(xi - \bar{x}) * (yi - \bar{y})}{\sum_{i=1}^{n}(xi - \bar{x})^2}$$

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

نستنتج من القانون الثاني أنّه لا يمكن حساب قيمة a إلّا بعد حساب قيمة b.

وبالتالي هناك ما ندعوه **تحليل الانحدار Regression Analysis**: هو كل طريقة إحصائية يتمّ فيها التنبّؤ بمتوسط متغير عشوائي أو عدّة متغيرات عشوائية اعتماداً على قيم وقياسات متغيرات عشوائية أخرى.

إنّ تحليل الانحدار هو أكثر من عملية ملائمة منحنى (أي اختيار المنحنى الأكثر ملائمة لمجموعة نقاط بيانية مُعطَاة) فهو يتضمّن ملائمة نموذج باستخدام مكوّنات حتمية واعتباطية (المكونات الحتمية تُدعَى المتنبئات أمّا المكونات الاعتباطية تُدعَى الخطأ).

الفرق بين الانحدار والارتباط:

من الارتباط نحصل على مؤشّر يصف العلاقة الخطية بين متغيرين ومدى قوّتها، أمّا في الانحدار يمكننا التنبّؤ بالعلاقة بين أكثر من متغيرين، ويمكن استخدام قيمة المتغير المستقل x لتوقع قيمة المتغير التابع y.

الارتباط يقيس مدى ارتباط متغيرين ولا يناسب خط، أمّا الانحدار الخطّي يجد الخط الأمثل الذي يتوقّع y من x.

مع الارتباط لا يهمّ التفكير في أي المتغيرين ندعوه x وأي المتغيرين ندعوه y، حيث ستحصل على معامل الارتباط نفسه إذا قمت بمبادلة الإثنين، أمّا في الانحدار فذلك يهمّ، حيث ستحصل على خط مختلف إذا بدّلت الإثنين.

كما أن الانحدار يفترض وجود علاقة سببية بين المتغيرين ويوضّح أيهما التابع وأيهما المستقل، أمّا الارتباط فيحدّد درجة اقتران المتغيرين دون أن يحدّد وجود أي علاقة تابعية أي أيهما تابع وأيهما مستقل.

ينحصر التشابه في الإشارة، حيث يتّفق معاملا الانحدار والارتباط بالإشارة فإذا كان أحدهما سالب فالآخر حتماً سالب، وإذ ساوى معامل الانحدار الصفر يكون معامل الارتباط مساوياً للصفر كذلك.

مثال: لدينا المعلومات التالية التي تبيّن عمر النبات مقدراً بالأسبوع وطول النبات مقدراً بالسنتيمتر.

والمطلوب: حساب كل من معامل الارتباط، ومعامل التحديد، وتشكيل معادلة الانحدار التي تعيّن العلاقة بين طول النبات وعمر النبات:

xi	yi	$xi - \bar{x}$	$yi - \bar{y}$	$(xi - \bar{x}) * (\mathrm{yi} - \bar{\mathrm{y}})$	$(xi - \bar{x})^2$	$(yi - \bar{y})^2$
١	٥	٣-	١٩-	٥٧	٩	٣٦١
٢	١٣	٢-	١١-	٢٢	٤	١٢١
٣	١٦	١-	٨-	٨	١	٦٤
٤	٢٣	٠	١-	٠	٠	١
٥	٣٣	١	٩	٩	١	٨١
٦	٣٨	٢	١٤	٢٨	٤	١٩٦
٧	٤٠	٣	١٦	٤٨	٩	٢٥٦
٢٨	١٦٨	٠	٠	١٧٢	٢٨	١٠٨٠
٤	٢٤					

حساب معامل الارتباط:

$$r = \frac{\sum_{i=1}^{n}(xi - \bar{x}) * (yi - \bar{y})}{\sqrt{\sum_{i=1}^{n}(xi - \bar{x})^2 * \sum_{i=1}^{n}(yi - \bar{y})^2}}$$

$$r = \frac{172}{\sqrt{28 * 1080}} = 0.99$$

نستنتج أنّه لدينا علاقة طردية بين عمر النبات وطول النبات ولدينا ارتباط موجب قوي حيث أنّ زيادة عمر النبات تؤثّر وبشكل قوي على زيادة طول النبات.

في المثال السابق لدينا:

$$\mathrm{B} = r^2 = 0.99^2 = 0.98$$

نستنتج من ذلك أنّ:

٩٨% من أسباب زيادة طول النبات تعود للزيادة في عمر النبات.

أمّا ٢% من أسباب زيادة طول النبات تعود إلى أسباب أخرى غير مدروسة.

نقوم بتشكيل معادلة الانحدار التي تعيّن العلاقة بين الطول والعمر، كما يلي:

$$b = \frac{\sum_{i=1}^{n}(xi - \bar{x}) * (yi - \bar{y})}{\sum_{i=1}^{n}(xi - \bar{x})^2}$$

$$b = \frac{172}{28} = 6.14$$

بعد أن قمنا بحساب قيمة b نقوم بحساب قيمة a ، كما يلي:

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

$$a = 24 - 6.14 * 4 = -0.57$$

بعد استنتاج (a و b) نشكّل معادلة الانحدار:

$$y = -0.57 + 6.14x$$

b: موجب، وهذا يعني أنّ العامل المستقل x يؤثّر إيجاباً على العامل y أي أنّ زيادة العمر تؤدي إلى زيادة الطول.

كلما زاد العمر بمقدار وحدة كاملة يزداد طول النبات بمقدار وحدة كاملة والتي هي b=6.14 أو بكلام آخر نقول أنّ طول النبات يزداد أسبوعياً بمقدار ٦.١٤ سم.

تساؤل: كم يزداد طول النبات إذا كان عمر النبات ٤ أيام ؟

يزداد بمقدار يساوي $6.14 * \frac{4}{7} \approx 3.51$ سم