إصدارات مدونة عيون المعرفة

http://knoweyes.blogspot.com

مقدمة في الإحصاء والاحتمالات

إعداد

المهندس / عبدالحفيظ العمري

أولاً: الإحصاء

Statistics

مقدمة في علم الإحصاء

إن علم الإحصاء هو مجموعة من الأساليب والعمليات الخاصة لمعالجة البيانات المتوفرة لدينا سواء الكمية منها أو الرقمية وللوصول لقيمة ذات دلالة للقيم موضوع البحث تتجانس معها القيم الأخرى وهذه القيمة ذات الدلالة تعرف بمقياس النزعة المركزية - Measures of Central Tendency - وهي المتوسط الحسابي (Average or Mean) والوسيط (Median) والمنوال (Mode) وغيرها. بالإضافة لمقاييس أخرى هي مقاييس التشتت (Variance measurement) وهي: المدى المطلق أو المدى (Range)، والانحراف المعياري (Standard Deviation) وغيرها.

ومدلول كلمة إحصاء (Statistics) يراه البعض مجرد بيانات وأرقام بينما يراه آخرون طرق لجمع البيانات ووصفها وعرضها بصورة مبسطة في حين يراه البعض الآخر وسيلة لاتخاذ القرار حال عدم التأكد.

معنى كلمة إحصاء

في القرآن الكريم:

ثُمَّ بَعَثْنَاهُمْ لِنَعْلَمَ أَيُّ الْحِزْبَيْنِ أَحْصَى لِمَا لَبِثُوا أَمَدًا (12) - الكهف -

إِنَّا نَحْنُ نُحْيِي الْمَوْتَى وَنَكْتُبُ مَا قَدَّمُوا وَآثَارَهُمْ وَكُلَّ شَيْءٍ أحْصَيْنَاهُ فِي إِمَامٍ مُبِينٍ (12) - ياسين -

لِيَعْلَمَ أَن قَدْ أَبْلَغُوا رِسَالَاتِ رَبِّهِمْ وَأَحَاطَ بِمَا لَدَيْهِمْ وَأَحْصَى كُلَّ شَيْءٍ عَدَدًا (28) - الجن -

وَكُلَّ شَيْءٍ أَحْصَيْنَاهُ كِتَابًا (29) - النبأ -

وَوُضِعَ الْكِتَابُ فَتَرَى الْمُجْرِمِينَ مُشْفِقِينَ مِمَّا فِيهِ وَيَقُولُونَ يَا وَيْلَتَنَا مَالِ هَذَا الْكِتَابِ لا يُغَادِرُ صَغِيرَةً وَلا كَبِيرَةً إِلا أَحْصَاهَا وَوَجَدُوا مَا عَمِلُوا حَاضِرًا وَلا يَظْلِمُ رَبُّكَ أَحَدًا (49) - الكهف -

يَوْمَ يَبْعَثُهُمُ اللَّهُ جَمِيعًا فَيُنَبِّئُهُم بِمَا عَمِلُوا أَحْصَاهُ اللَّهُ وَنَسُوهُ وَاللَّهُ عَلَى كُلِّ شَيْءٍ شَهِيدٌ (6) - المجادلة -

لَقَدْ أَحْصَاهُمْ وَعَدَّهُمْ عَدًّا (94) - مريم -

وَآتَاكُم مِّن كُلِّ مَا سَأَلْتُمُوهُ وَإِن تَعُدُّواْ نِعْمَتَ اللّهِ لاَ تُحْصُوهَا إِنَّ الإِنسَانَ لَظَلُومٌ كَفَّارٌ (34) - إبراهيم -

يستفاد من مفهوم الكلمة (إحصاء) هنا بأنها بيانات

في اللغة:

عبر العرب عن كثرة الشيء وحجمه بالحصى، ويقال حصيت أي عددت وأحصيته أي ميزته بعضه عن بعض، والحصاة بمعنى العقل، واختار الإمام ابن القيم في البدائع (164/1) أن الإحصاء على ثلاثة مراتب هي :

1- إحصاء ألفاظها وعددها

2- فهم معانيها ومدلولها .

3- دعاؤه بها

والإحصاء في الكلام : على ثلاث مراتب

1) العدد، ومنه قوله تعالى {وأحصى كل شيء عدداً} [الجن :28] - تخص أصحاب اليمين -

2) الفهم، ومنه يقال: رجل ذو حصاة أي: ذو لب وفهم، ومنه سمي العقل. - تخص السابقين -

3) الإطاقة على العمل والقوة، ومنه قوله تعالى: { علم ألن تحصوه } [المزمل : 20] أي: لن تطيقوا العمل بذلك. - للصديقين -

استخدام كلمة إحصاء:

تستخدم بصيغة الجمع أي إحصائيات (Statistics) ومرادفتها بيانات وتستخدم كمفرد إحصاء أو إحصاءه بمعنى النتيجة التي توصلنا إليها من إجراء الرياضيات على البيانات فمتوسط العمر لمجموعة من الناس يطلق عليه إحصاء أو إحصاءه.

علماً : فرع من فروع العلم، فالإحصاء مجرد بيانات تعتبر المادة الخام لهذا العلم والبيانات تجمع من مصادرها المطلوبة حسب المطلوب فتبوب وتلخص وتقييم ومن ثم يستنتج المراد منها.

الإحصاء الوصفي (Statisticts Descriptive)

منذ نهاية العصور الوسطى بدأت الحكومات بحفظ السجلات وخاصة الأراضي منها وتدوين البيانات عنها وعن السكان والأمور الأخرى وعلى الطرف الآخر بدأت الشركات والمؤسسات وخاصة شركات التأمين بإعداد جداول خاصة بالوفيات بهدف مالي بحت وكان الأمر لا يتعدى جمع البيانات وتصنيفها وعرضها وهو ما يعرف بالإحصاء ألوصفي من كونه يعطي وصفاً للبيانات المجموعة والابتعاد عن التعميم حال التعرف على مقياس لمتوسط مجموعة من الأفراد مثلاً، وعلى العموم يمكن القول عن هذا النوع من الإحصاء بأنه يختص بالطرق المختلفة لجمع البيانات وتحليها ووصفها حتى تقدم بصورة بسيطة سهلة الفهم وذات دلالة دون أي تعميم.

وسائل هذا النوع من الإحصاء العرض البياني والحساب فالعرض البياني يعني جدولة البيانات وعرضها بيانياً والحساب المقاييس الإحصائية لتلك البيانات كالنزعة المركزية من أمثلته الوسط الحسابي أو مقاييس التشتت أو عدم التجانس من أمثلته المدى ومقاييس أخرى سنتعرض لها في حينها كمعامل الارتباط .

مثال: في اختبار ما حصل أربعة طلاب على الدرجات 70 ، 80 ، 75 ، 95 فقولنا إن متوسط درجات الطلاب هو 80 درجة فالمتوسط يعطي وصفاً لدرجات الطلاب لأن الوسط الحسابي هذا لخص ووصف البيانات هذه دون تعميمها على باقي الطلاب وكذلك يمكننا عرض هذه البيانات بطرق العرض البياني والتي سيرد ذكرها لاحقاً.

الإحصاء الاستدلالي (Statistics Inferential)

هو مجموعة الطرق للتعرف على خصائص المجتمع من خلال عينة عشوائية (الإحصائية) من هذا المجتمع معتمدة طرق إحصائية محددة.

يشكل الإحصاء الاستدلالي مع الإحصاء ألوصفي فرعا علم الإحصاء الحديث وهما ضروريان لاتخاذ القرار.

الإحصاء الاستدلالي يتعامل مع التعميم والتقدير و التنبوء إلا أنه يتسم في بعض الحالات بعدم التأكد مما يدعونا لمعالجة القياس في هذه الأحوال تحت باب علم الاحتمالات مما يعطي فكرة عن الخطأ المحتمل وقوعه من الباحث حال التعميم على المجتمع المسحوب منه العينة العشوائية محل الدراسة.

وسيلتي الإحصاء الاستدلالي التقدير ، اختبار الفروض فالتقدير قيمة مثل قيمة المتوسط الحسابي في حين اختبار الفروض يعني القبول أو الرفض وسنذكر ذلك لاحقاً بالتفصيل.

مثال: ادعت مؤسسة بأنها قامت بتصنيع نوع جديد من أقراص الليزر المستخدمة في الحاسب الآلي بفاعلية 95 % بمعنى من بين كل 100 قرص نجد 95 سليمة تماماً ولتقبل قول المؤسسة هذه قمنا بأخذ 20 قرصاً وجرى استخدامها فتبين وجود قرصين غير سليمين أي 18 قرص سليم في حين المؤسسة تدعي $0.95 \times 20 = 19$سليم فنرفض ادعاء المؤسسة لأن 18> 19.

مفهوم الإحصاء (Definition of Statistics)

مجموعة من الطرق العلمية القياسية تهدف لجمع البيانات عن ظاهرة ومن ثم جدولتها وعرضها بصورة مختصرة وتقيمها للوصول لنتائج من خلال عينة صغيرة من مجتمع للتعرف على خصائص المجتمع بصورة تقريبية لخصائص العينة.

البيانات الإحصائية (Statistical Data)

المقصود هنا بالبيانات والمعلومات الإحصائية المتعلقة بالمجتمع وتختلف البيانات باختلاف الظاهرة محل الدراسة، وتختلف البيانات من حيث النوع والطبيعة حسب الظاهرة محل الدراسة فمنها البيانات السكانية والتربوية والصحية وغير ذلك مما هو موجود في المجتمع وما يتعلق بنشاطات المجتمع الرياضية والفكرية وغيرها ولكل منها طريقته الخاصة في البيانات المتوفرة في المجتمع محل الدراسة، وقد يتطلب الأمر في بعض الحالات تحويل البيانات الوصفية لكمية باستخدام أدوات القياس سواء الكمية منها أو النوعية

<u>البيانات النوعية (Qualtaive data</u>

البيانات النوعية وهي تصف ظاهرة بصورة غير رقمية كالجنس (ذكر و أنثى) والتقدير (ممتاز و جيد و ...) وقد تأخذ قيم كما هو الحال في استطلاعات الرأي فالقيمة هنا تعبر عن الشعور أو الرغبة للشخص المستجيب ويُعتقد أن هذا النوع من البيانات تساعد على حل العديد من المشاكل.

<u>البيانات الكمية (Quantative data</u>

والبيانات كما ذكرنا أما نوعية أو كمية (رقمية) وتعرف الرقمية بالبيانات المقيسة مثل الكيلوجرام للوزن والمتر للطول والدينار للسعر، وهي تعبر عن ظاهرة في المجتمع بصورة رقمية كإنتاج القطن بالطن، والبيانات الكمية تعبر عن خاصية ما في المجتمع.

<u>المجتمع (Population)</u>

المجتمع يمثل جميع المفردات الممكنة للظاهرة محل الدراسة، والمجتمع قد يكون سكان دولة ما أو أجور عمال مهنة ما في بلد ما أو العاطلين في بلد ما والمجتمع قد يكون محدوداً (يمكن عد مفرداته ولو من الناحية النظرية) أو يكون مجتمع غير محدود (لا يمكن عد مفرداته)، وتعرف خصائص المجتمع التي يمكن قياسها كمياً بمعالم المجتمع (Parameters) كمتوسط أجر المدرسيين في الدولة والمتوسط كقيمة يعبر عن القيمة الحقيقية لمعلمة المجتمع ونسبة الناجحين في امتحانات الثانوية العامة بالبحرين هي من معالم مجتمع المتقدمين لامتحانات الثانوية العامة وللمجتمع الغير محدود يستحيل الوصول للقيمة الحقيقية عند دراسة ظاهرة ما فلذا نستدل عليها بأخذ مفردات قليلة العدد من المجتمع (العينة) للاستدلال على معالم المجتمع.
والمجتمع قد يكون مستهدف أو معاين فالمجتمع المستهدف هو ذلك المجتمع الذي يعمل الباحث لاتخاذ قرار بشأنه والمعاين هو مجموعة المفردات التي يختبرها الباحث والمعروف بالعينة الآتي تعريفها.

<u>العينة (Sample)</u>

العينة هي جزء من المجتمع ودراستها وما ينجم عنها من خصائص يمكن بواسطتها الاستدلال على خواص المجتمع ككل، وإحصائية العينة أي أحد خصائصها كالمتوسط الحسابي لأفرادها وهو قياس كمي أو جزء من أفرادها للكل كنسبة الناجحين بين أفراد العينة ويعتبر الوسط الحسابي أو النسبة (الاحتمال) من خصائص العينة ومنها يستدل على معالم المجتمع المسحوب منه العينة فرسوب ثلاثة طلاب من بين 20 طالباً (العينة) أي $3 \div 50 = 0.06$ هي من خواص العينة ويمكن الاستدلال منها حال عدد مفردات المجتمع (1000 مثلاً) من الطلاب المتقدمون للامتحان السابق في العينة يستدل برسوب $1000 \times 0.06 = 60$ طالب.
العينة البسيطة هي تلك التي يكون لكل مفردة من مفرداتها نفس الاحتمال في الظهور.
العينة المنتظمة هي تلك التي يكون البعد بين مفرداتها متساوي (كل مفردة بالتي تليها أو تسبقها).
العينة الطبقية التي مفرداتها تنم عن طبقات تجريبي وغير تجريبي - مدينة وريف ، ...

المعلمة والإحصاءة

المعلمة (Parameter) وهي القياس المستنتج من بيانات المجتمع.

والإحصاءة (Statistic) وهي المقياس المستخرج من بيانات العينة وهي محل التعامل مع بياناتها في معظم الأحوال.

وعليه يكون للعينة أهمية قصوى في معرفة خصائص مجتمع أو أكثر من خلال خصائص العينة محل الدراسة وباعتبار معلمة المجتمع مجهولة فتقدر بالمناسب من خلال حسابه من بيانات العينة.

الرمز Σ

رمز المجموع Σ ويقرأ سيجما (حرف إغريقي) يختصر مجموعة من الأعداد أو الرموز في جملة رياضية واحدة فبدلاً من كتابة 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 يعبر عنها بواسطة الرمز سيجما بالصورة التالية:

$$\sum_{i=1}^{8} X_i = 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8$$

وفي حالة وجود البيانات المبوبة (عدة قيم للمتغير X) فإن كان مجموع القيم هذه يساوي 95 مثلاً نقول $\sum X = 95$ وللرمز هذا قواعده يلزم الرجوع عليها في علم الرياضيات ولكننا سنذكرها عند تعرضنا لها في حينه.

الحاسب الآلي

استخدام الحاسب الآلي في الإحصاء وغيره أصبح أمراً ملحاً وناجحاً بدرجة كبيرة لتوفير الوقت ودقة النتائج وغيرها الكثير، وهناك الكثير من استخدامات الحاسب الآلي نذكر أهمها هنا فيما يتعلق بالإحصاء بكافة أنواعه وجود برامج متخصصة في هذا المجال أهمها SPSS, MINITAB فباستخدام أياً منهم نستطيع الحصول على نتائج التحليل وغيره في وقت وجيز علاوة على الحاسب الآلي يمدنا بأمور أخرى كثيرة تساعدنا في عملية جمع البيانات أو البحث في شبكة الإنترنت عما يفيدنا في مجال البحوث.

لذا يجب التعرف على الحاسب الآلي ومن خلاله التعرف على البرامج ذات الصلة بعلم الإحصاء كبرنامج SPSS أو حتى برنامج EXCEL ضمن مجموعة Microsoft ويجب التعلم على كيفية استخراج النتائج ونحن هنا أوردنا بالتفصيل كيفية الحصول على تلك النتائج في معظم أمثلتنا وبالبرامج الثلاثة المذكورة أعلاه بما يخص موضوع المسألة المطروحة دون التطرق لتفاصيل أخرى في علوم تلك البرامج والتي لا يمكن سردها هنا بل هناك الكثير من الكتب والبرامج المتخصصة في شرح تلك البرامج فيمكن الرجوع إليها وفي الحد الأدنى يمكن استغلال وجود Help المصطحب مع تلك البرامج.

البيانات الخاطئة

تواجه الباحث مشكلة أساسية في كيفية الحصول على أفضل بيانات محل الدراسة التي يقوم بها بالإضافة لأمور أخرى (تحديد نوع المقياس المطلوب) فالمقياس المطلوب ستلزم بيانات معينة على الباحث أن يتجه إليها و إلا فستكون بياناته غير صحيحة بأن يتبع العكس أي يجمع البيانات ثم يبدأ تحليله الإحصائي ليجد نفسه ضمن بيانات يصعب الحصول منها على النتائج المرجوة من بحثه، هناك البيانات التصنيفية (تصنيف أفراد لذكور وإناث) أو بيانات رتبية (كترتيب أفراد مجموعة من حيث الطول مثلاً)، أو البيانات الفئوية والنسبية والتي يجب أن يلجأ إليها الباحث كمرشد له في انتقاءه الأسلوب الإحصائي المناسب وإلا يكون جمعه لبيانات خاطئة.

مصادر البيانات

إن عملية جمع البيانات من مصادرها التاريخية أو الوثائقية كحصيلة لنشاط العديد من المؤسسات والشركات والوزارات وغيرها أو تلك المؤلفات المتوفرة في المكتبات وغيرها تضم العديد من المعطيات الإحصائية والتي يجب الرجوع إليها من قبل الباحث وهي على نوعين:

1) أصلية وهي من الجهة التي تقوم بجمعها كالتعداد السكاني.

2) ثانوية وهي تلك الجهات التي تقوم بنشر البيانات بعد تسلمها من جهتها الأصلية.

كما أن المواقع الميدانية مصدراً لجمع البيانات عن طريق الاستمارات أو التعداد أو أخذ عينة من المجتمع الإحصائي ممثلة لكافة خصائص المجتمع، وهناك عدة طرق للقيام بجمع البيانات.

طريقة المشاهدة كمعرفة حركة المرور في نقطة معينة وتسجيل البيانات منها.

طريقة الاستبيان بطرح أسئلة يتم الإجابة عليها على أن تكون تلك الأسئلة تتناول موضوع معين كمجانية التعليم أو طبيعة السكن.

طريقة الالتقاء المباشر بين الباحث مع المبحوثين شخصياً للحصول على البيانات المطلوبة مع ضرورة شرح المطلوب للمبحوث للحصول على أفضل الإجابات.

طريقة الهاتف حال توفر الهاتف بصورة غالبية المجتمع محل الدراسة.

لا مانع اليوم من استخدام طرق أخرى كالبريد الإلكتروني أو نشر المطلوب عبر شبكة الإنترنت وطلب الإجابة عليه من قبل عينة من المجتمع.

المصدر الميداني للبيانات:

يعتبر أهم مصدر لجمع البيانات حيث يتولى الباحث مهمة الجمع للبيانات بنفسه وهي عملية ليست سهلة وتخضع لشروط وقيود توضع من قبل الباحث وألا حصلنا على بيانات مضللة وغير ممثلة للمجتمع محل الدراسة وتتضمن الخطوات التالية:

1) معرفة نوع المجتمع محل الدراسة من كونه بشري أو بشري ومستوى التعليم ومتوسط السن والمواصلات والدخل وما إلى ذلك.

2) وحدة المعاينة فالباحث يجب أن يحدد بدقة وحدة المعاينة فدراسة الزواج مثلاً فعليه اختيار الأسرة كوحدة معاينة ودراسة النجاح للطالب فالوحدة هنا هي المدرسة مثلاً

3) حجم العينة الممثلة للمجتمع محل الدراسة وأسلوب جمع البيانات عن العينة التي يجب أن تكون ممثلة للمجتمع.

4) تحديد النقاط الشاملة للبحث وتصميم صحيفة بحث والطريقة المتبعة في جمع البيانات كما ذكرنا سابقاً من طرق.

5) اختيار مناسب لجامعي البيانات بحيث يكونوا ضمن التخصص محل الدراسة فالباحث إن هدفه طبي فيفضل أن يكون جامعي البيانات ذات دراية بالطب.

6) كيفية جمع البيانات مباشرة أو غير مباشرة والمباشرة هي الأهم حيث تكون عن طريق المشاهدة أو البريد أو التليفون أو المقابلات الشخصية.

7) تصميم استمارة البحث من حيث الشكل والمضمون ونوع الأسئلة وكيفية صياغتها ووضوحها وغير ذلك

المصادر الإحصائية للبيانات:

إن عملية جمع البيانات من مصادرها التاريخية أو الوثائقية كحصيلة لنشاط العديد من المؤسسات والشركات والوزارات وغيرها أو تلك المؤلفات المتوفرة في المكتبات وغيرها تضم العديد من المعطيات الإحصائية والتي يجب الرجوع إليها من قبل الباحث وهي على نوعين:

1) أصلية وهي من الجهة التي تقوم بجمعها كالتعداد السكاني.

2) ثانوية وهي تلك الجهات التي تقوم بنشر البيانات بعد تسلمها من جهتها الأصلية.

كما أن المواقع الميدانية مصدراً لجمع البيانات عن طريق الاستمارات أو التعداد أو أخذ عينة من المجتمع الإحصائي ممثلة لكافة خصائص المجتمع، وهناك عدة طرق للقيام بجمع البيانات.

طريقة المشاهدة كمعرفة حركة المرور في نقطة معينة وتسجيل البيانات منها.

طريقة الاستبيان بطرح أسئلة يتم الإجابة عليها على أن تكون تلك الأسئلة تتناول موضوع معين كمجانية التعليم أو طبيعة السكن.

طريقة الالتقاء المباشر بين الباحث مع المبحوثين شخصياً للحصول على البيانات المطلوبة مع ضرورة شرح المطلوب للمبحوث للحصول على أفضل الإجابات.

طريقة الهاتف حال توفر الهاتف بصورة غالبية المجتمع محل الدراسة

لا مانع اليوم من استخدام طرق أخرى كالبريد الإلكتروني أو نشر المطلوب عبر شبكة الإنترنت وطلب الإجابة عليه من قبل عينة من المجتمع

<u>المجتمع والعينة:</u>

المجتمع يشمل كافة وحدات الظاهرة محل الدراسة فهو أي المجتمع يمثل حجم المجموع والمقصود بالمجموع عدد من تقع عليهم الدراسة فطلبة الجامعة وعددهم 900 محل الدراسة فإن حجم المجتمع هنا يساوي 900 وقد يكون الحجم غير محدد كمجتمع الأسماك ولذا نقوم بأخذ عينة من المجتمع بصورة عشوائية أو حتى غير عشوائية على أن تكون ممثلة لخصائص المجتمع كافة.

<u>البيانات المبوبة:</u>

عند الحصول على البيانات والقيام بجدولتها بهدف عرضها بصورة مختصرة حيث توجد عدة طرق لتمثيلها كالمدرج التكراري والمضلع التكراري والمنحنى التكراري والساق والورقة وفي حالة البيانات النوعية كالجنس (ذكر - أنثى) تمثل بطريقة الأعمدة البيانية ولكل طرقه الخاصة به.

<u>التوزيع التكراري (Frequency Distribution)</u>

هو مجموعة من البيانات التي توضع بشكل منظم في جدول بهدف تلخيص تلك البيانات للوصول بسهولة اتخاذ القرار بإجراء ما والجدول هذا يشمل على عدد من الفئات المتساوية يقابل كل منها التكرار المناسب من البيانات حيث يتم حصر كل البيانات في الجدول والمعروف بجدول التوزيع التكراري.

<u>التوزيع التكراري البسيط:(Simple Frequency Distribution) لبيانات كبيرة نسبياً</u>

تبوب البيانات على شكل فئات تكرارية مع تحديد عدد المشاهدات لكل من هذه الفئات ويعرف عدد المشاهدات هنا بالتكرار فإذا أخذنا مجموعة البيانات التالية لأعمار (بالسنة) لثلاثين مريضاً لمراجعتهم المستشفى:

13	27	12	13	17	12
20	22	18	27	22	18
21	20	18	16	14	13
12	21	20	23	22	27
26	25	14	16	17	18

الخطوات اللازمة لتكوين جدول التوزيع التكراري:

أولا: عدد الفئات - Intervals -

يجب ألا يكون عدد الفئات قليل لنفقد بعض البيانات أو كبيرة لنفقد صفة الاختصار وخلق فئات خالية من التكرار ولذا نهتم بمدى الاختلاف بين البيانات للوصول لعدد الفئات والذي يتراوح بين 5 ، 20 فئة ويمكن تطبيق القاعدة حيث ن=30

عدد الفئات = 1 + 3.322 لـو$_{10}$(ن) = 1 + 3.322 × 1.4771 = 1 + 4.9 = 5.9 ≈ 6

ثانياً: طول الفئة (مدى الفئة) - Interval Width -

بقسمة الفرق بين أكبر وأصغر قيمة في البيانات مضافاً إليه واحد على 6 أو عدد الفئات المقترحة أي:

طول الفئة = (أكبر قيمة – أصغر قيمة + 1) ÷ 6 = (27 – 12 + 1) ÷ 10 = 16 ÷ 6 = 2.67

نقربها إلى 3 مع ملاحظة إمكانية تجاهل إضافة العدد واحد عندما نجد توسع في طول الفئة وفي حالتنا هنا لا يؤثر لكون التقريببين2.67 (16 ÷ 6) أو 2.5 (15 ÷ 6) هو 3

ثالثاً حدود الفئات - Interval Limits -

الحد الأدنى للفئة الأولى يتمثل بأصغر قيمة في البيانات (12) والحد الأعلى للفئة الأولى يتحدد بقيمة: الحد الأدنى + طول الفئة – 1 = 12 + 3 – 1 = 14

جدول التفريغ - توزيع التكرارات لكل فئة - Frequency Table

نكون جدول يعرف بجدول التفريغ ويتكون من ثلاثة أعمدة للفئات والعلامات والتكرار وصفوف بعدد الفئات مع زيادة صفين يختصان بالمجموع والعنوان ومن ثم نحذف عمود العلامات للحصول على جدول التوزيع التكراري والجدولين هما:

الفئات	العلامات	التكرار
12 – 14	///// ///	8
15 – 17	////	4
18 – 20	///// //	7
21 – 23	///// /	6
24 – 26	//	2

الفئات	التكرار
12 – 14	8
15 – 17	4
18 – 20	7
21 – 23	6
24 – 26	2

27 – 29	3
المجموع	30

27 – 29	///	3
المجموع		30

الفئات المفتوحة:- Open Intervals -

للقيم الشاذة، فإن كانت من أعلى أي وجود قيمة مثل 70 في البيانات السابقة فنضيف فئة جديدة هي 30 فأكثر تعرف بالفئة المفتوحة وإن وجدت فئة شاذة من أسفل مثل القيمة 2 فنضيف فئة من أسفل أقل من 11 وتعرف بالفئة المفتوحة وتعتبر القيم 2 ، 70 قيم متطرفة لباقي القيم أو متباينة جداً مع القيم الأخرى.

الحدود الحقيقية (الفعلية) للفئات:- Exact Interval Limits -

من الملاحظ بان الفئات غير متصلة ببعضها البعض أي أن هناك بعض القيم بين فئة وأخرى ولذا نحدد الحد الأعلى والأدنى الفعليين للفئة من:

الحد الأعلى الفعلي للفئة المطلوبة = (الحد الأعلى للفئة المطلوبة + الحد الأدنى للفئة اللاحقة) ÷ 2

الحد الأدنى الفعلي للفئة المطلوبة = (الحد الأعلى للفئة السابقة + الحد الأدنى للفئة المطلوبة) ÷ 2

مراكز الفئات:- Mid Intervals -

مركز الفئة هو القيمة التي تتوسط الفئة وهي الوسط الحسابي لحديها الأعلى والأدنى أو حديها الفعليين، ومن الملاحظ أننا نستخدم مركز الفئة لحساب مجموع القيم وذلك من خلال ضرب مركز كل فئة في تكرارها ويلاحظ من الجدول التالي أن مجموع حاصل ضرب مركز الفئة × التكرار يساوي تقريباً مجموع القيم الأصلية على افتراض أن نصيب كل مَن في التكرار يأخذ القيمة الممثلة بمركز الفئة.

الجدول الآتي يبين ما سبق ذكره أعلاه

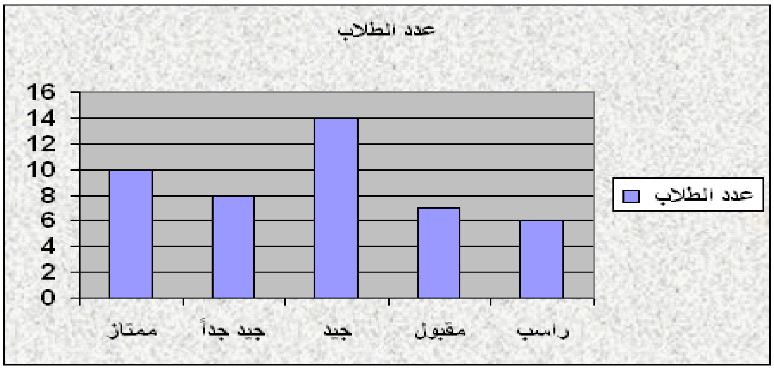
الفئات	الحدود العليا الفعلية للفئات	الحدود الدنيا الفعلية للفئات	مركز الفئة	التكرار	مركز الفئة × التكرار
12 – 14	(14 +15) ÷ 2 = 14.5	(12 + 11) ÷ 2 = 11.5	(12 + 14) ÷ 2 = 13	8	104
15 – 17	(17 + 18) ÷ 2 = 17.5	(14 + 15) ÷ 2 = 15.5	(15 + 17) ÷ 2 = 16	4	64
18 – 20	20.5	18.5	19	7	133
21 – 23	23.5	21.5	22	6	132
24 – 26	26.5	24.5	25	2	50
27 – 29	29.5	27.5	28	3	84
المجموع				30	567

الأعمدة البيانية: Bar Charts

ليكن لدينا الجدول التالي لتقدير 45 طالب في أحد المواد الدراسية

التقدير	عدد الطلاب
ممتاز	10
جيد جداً	8
جيد	14
مقبول	7
راسب	6
المجموع	45

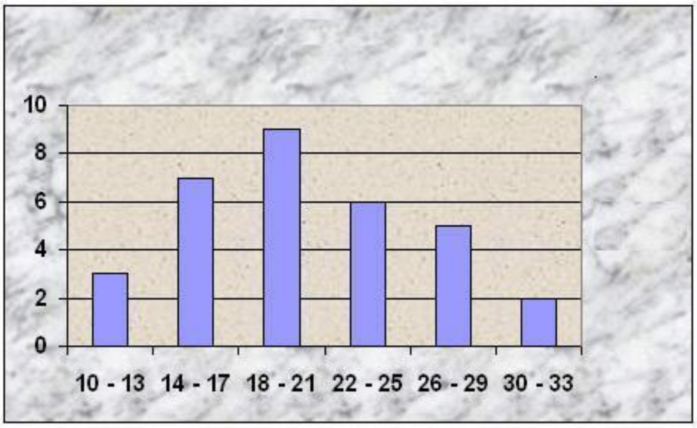
نأخذ محورين متعامدين (الأفقي والرأسي أو السيني والصادي) ونختار أبعد متساوية على المحور الأفقي بطول 2 سم نرسم مستطيل بعرض 1 سم وطول يساوي التكرار لكل من التقديرات المبينة فالتقدير ممتاز يمثله المستطيل الأول من جهة اليسار بارتفاع (طول ضلع المستطيل) 10 وطول القاعدة على المحور الأفقي بطول 1 سم لنحصل على الرسم البياني التالي والممثل للتوزيع التكراري المبين بالجدول أعلاه.

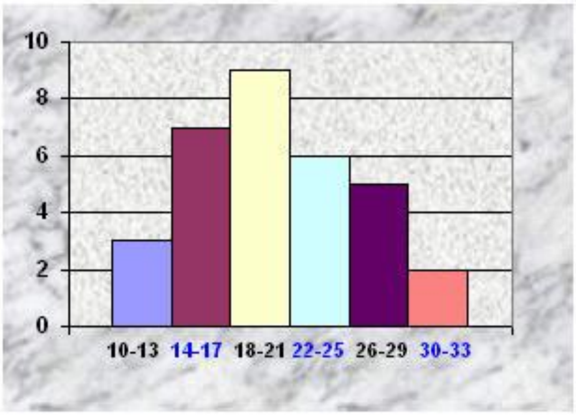


<u>المدرج التكراري: **Frequency histogram**</u>

يتم إدراج الفئات أو مراكزها أو الحدود الفعلية للفئات (الأفضل) على المحور الأفقي ويتم وضع التكرارات على المحور الرأسي (الخاصة بالظاهرة محل الدراسة أو المتغير) إلا أن الأعمدة قد تكون متلاصقة أو غير متلاصقة والجدول التالي يبين درجات 32 طالب في مادة الإحصاء.

F	X
3	10 - 13
7	14 - 17
9	18 - 21
6	22 - 25
5	26 - 29
2	30 - 33
32	∑

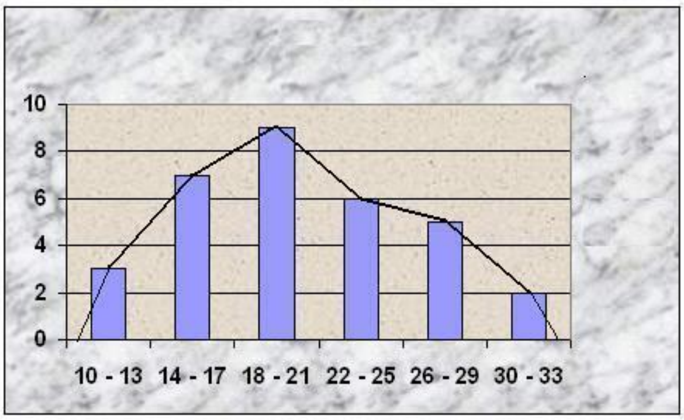


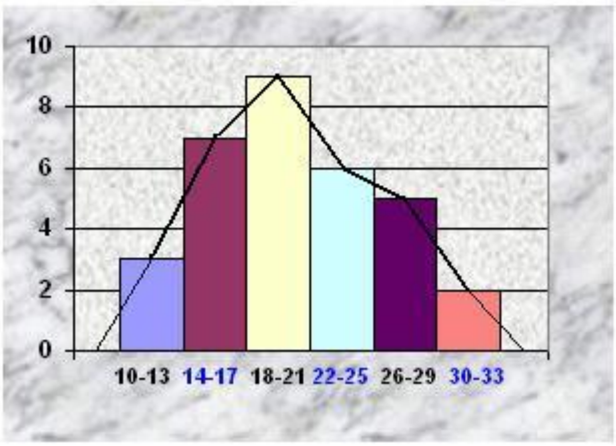


<u>المضلع التكراري</u>

فكرة المضلع التكراري نفس فكرة المدرج التكراري ونعين النقاط على المحورين، مركز الفئة مع تكرارها أو من المدرج التكراري بتوصيل نقط منتصفات القواعد العليا للمستطيلات المكونة للمدرج التكراري وكما هو مبين بالشكل التالي:

X	مركز الفئة ($\bar{X}$)	F
10 - 13	11.5	3
14 - 17	15.5	7
18 - 21	19.5	9
22 - 25	23.5	6
26 - 29	27.5	5
30 - 33	31.5	2
	∑	32

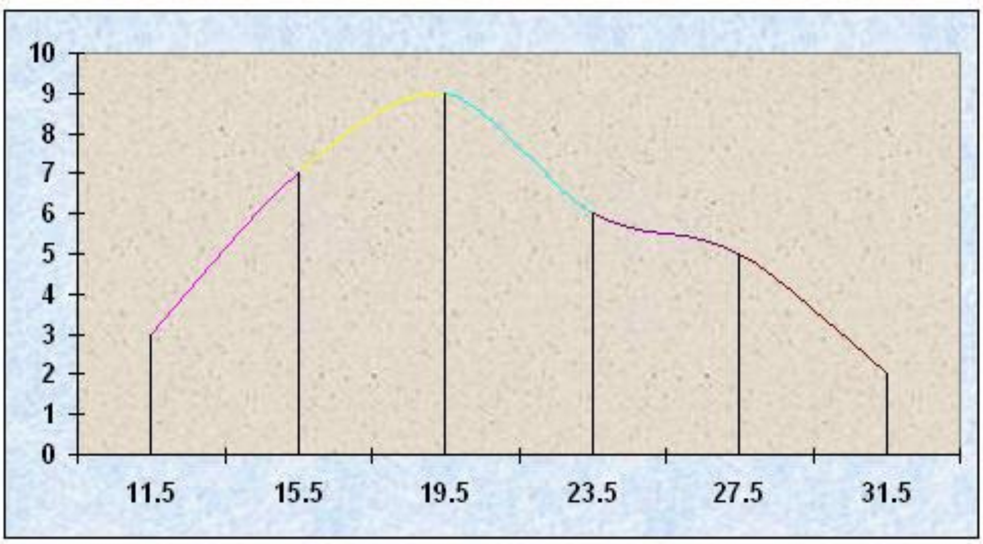




<u>المنحنى التكراري</u>

فكرة المنحنى التكراري نفس فكرة المضلع التكراري ونعين النقاط على المحورين، مركز الفئة مع تكرارها أو من المدرج التكراري بتوصيل نقط منتصفات القواعد العليا للمستطيلات المكونة للمدرج التكراري ورسم منحنى يمر بأكبر عدد ممكن من النقاط وبتوازن خلال باقي النقاط وكما هو مبين بالشكل التالي:

F	مركز الفئة (‾X)	X
3	11.5	10 - 13
7	15.5	14 - 17
9	19.5	18 - 21
6	23.5	22 - 25
5	27.5	26 - 29
2	31.5	30 - 33
32		∑



التكرار التراكمي أو المتجمع

التكرار التراكمي أو المتجمع هو تحديد لكل فئة بمجموع التكرارات السابقة لها (متجمع صاعد) أي السابقة للحد الأدنى في الفئة أو مجموع التكرارات اللاحقة لها (المتجمع الهابط أو النازل) أي اللاحقة للحد العلى للفئة وهو يختص بالبيانات القابلة للجمع (العددية) والمرتبة تنازلياً أو تصاعدياً والمثال التالي ومن خلال الجدول يوضح ذلك.

التكرار التراكمي (cf)	التكرار (f)	الفئات
8	8	12 – 14
12	4	15 – 17
19	7	18 – 20
25	6	21 – 23
27	2	24 – 26
30	3	27 – 29
	30	المجموع

مجموع التكرار دون نهاية الفئة الأول 12- 14(الحد العلى هنا 14) هو 8 بمعنى مجموع التكرارات التي أقل من 14

مجموع التكرار دون نهاية الفئة الثانية حدها الأعلى 17 هو 8 + 4 = 12

مجموع التكرار دون نهاية الفئة الثالثة حدها الأعلى 20 هو 8 + 4 + 7 = 19

مجموع التكرار دون نهاية الفئة الثانية حدها الأعلى 23 هو 8 + 4 + 7 + 6 = 25

مجموع التكرار دون نهاية الفئة الثانية حدها الأعلى 26 هو 8 + 4 +7 + 6 + 2 = 27

مجموع التكرار دون نهاية الفئة الثانية حدها الأعلى 29 هو 8 + 4 + 7 + 6 + 2 + 3 = 30 (مجموع التكرارات)

<u>أسئلة وأجوبة (1)</u>

(1 ما المقصود بالمتغير الكمي) ؟

المتغير الذي يعبر عنه بالمقدار مثل العمر، الدرجة لطالب، أيام الغياب، ... يمكن أن يأخذ صفة الترتيب

(2 ما المقصود بالمتغير النوعي)

الذي يصنف الأشياء كالجنس (ذكور وإناث) والتخصص (علمي - أدبي)، المرحلة (ابتدائي - إعدادي - ...)، مساحيق الغسيل (تايد، اومو، ...)، ... ويفقد صفة الترتيب

(3 ما هو المتغير المستقل والمتغير التابع؟

المستقل: الذي يؤثر في النتائج أو الذي يتسبب فيها ويمكن التحكم به (تغييره) ويعرف بالمتغير التجريبي كطرق التدريس أو طرق العلاج والنتائج المترتبة على المتغير المستقل تعرف بالمتغير التابع فطرق التدريس كمتغير مستقل تؤثر في مستوى التحصيل الذي يعتبر متغير تابع (ناتج التجربة لطرق التدريس)، وكذلك نتائج العلاج على المريض تعتبر متغير تابع للمتغير المستقل طرق العلاج.

(4 قارن بين العينة الاحتمالية والغير احتمالية.

	العينة الاحتمالية	العينة غير الاحتمالية
المجتمع	معروف	غير معروف
السجلات الرسمية	موثقة	غير موثقة
القوانين	تخضع لقوانين الرياضيات	تخضع لعمليات حسابية
النوع	عشوائية	قصدية

(5 ما الفرق بين المجتمعين الإحصائيين المعروف والغير معروف؟

المجتمع المعروف ما كانت عناصره من السجلات الرسمية كالمعلمين مثلاً في حين الغير معروف عناصره ليست في السجلات الرسمية كمدمني المخدرات.

(6 ما الفرق بين المتغير المتصل والمتغير الوثاب؟

المتغير المتصل: ما كان يأخذ أي قيمة في مداه مثل العمر يبدأ بالصفر ويزداد بصورة مستمرة في حين المتغير الوثاب أو المتقطع ما كان يأخذ قيماً ثابتة في مداه كعدد طلاب صفوف الصف الأول الثانوي في المدرسة وعدد أفراد الأسرة.

(7 ما هي مستويات القياس الأربع وصنفها في جدول؟

الجواب مبين بالجدول الآتي:

مستوى القياس	الخصائص	أمثلة	الإجراء الممكن
نسبي	الصفر عدد حقيقي له مفهوم	الدخل	العمليات الحسابية والمقاييس الإحصائية
فئوي	الصفر افتراضي لا يعني العدم	درجات الحرارة	العمليات الحسابية فقط على المتغير الواحد
رتبي	دلالة الأرقام للترتيب فقط	ترتيب طلبة صف كتحصيل دراسي	لا تجرى العمليات الحسابية
أسمي	الأرقام هنا للتصنيف أو الترتيب	ذكر (1)، الأنثى (0)، الرقم السكاني	لا تجرى العمليات الحسابية

8) اذكر أنواع العينات.

العينة الاحتمالية غير الاحتمالية كما مبين بالجدولين الآتيين :

العينة الاحتمالية

أنواع العينة	باللغة الإنجليزية	الغرض
العشوائية البسيطة	Simple Random Sample	احتمالات أفراد العينة متساو - تستخدم جداول الأرقام العشوائية لاختيار العينة
المنتظمة	Systematic Sample	ثبات البعد بين مفرداتها - تتعين من N مفردات المجتمع ويحدد البعد K حيث K $\leq$ N/n
الطبقية	Stratified Sample	تتعين من طبقات المجتمع (الريف - المدينة - (... عشوائياً لضمان التمثيل للطبقات
العنقودية	Cluster Sample	تنتج من تقسيم المجتمع لمجموعات جزئية يعرف كل منها بالعنقود
المتعددة المراحل	Multistage Sample	تتكون من عدة مراحل حيث تصمم باستخدام أي من العينات السابقة

العينة غير الاحتمالية

أنواع العينة	باللغة الإنجليزية	الغرض
المناسبة	Convenient Sample	سهولة الحصول عليها كقربها من مكان إقامة الباحث

تعتمد البيانات السابقة باعتقاد الباحث بتحقيقها تمثيل جيد للمجتمع	**Purposive Sample**	المتعمدة
تستخدم في دراسات التسويق وذلك بالتحكم في مفرداتها كالتقسيم العمري لمفرداتها	**Quota Sample**	الأنصبة

9) ما هي أوجه الاختلاف بين العينات الاحتمالية وغير الاحتمالية؟

وجه الخلاف كما مبين بالجدول الآتي:

العينة غير الاحتمالية	العينة الاحتمالية
لا يوجد إطار سحب	وجود إطار (المجتمع المعاين) لسحب مفرداتها
لا يتحقق هنا	احتمال مفرداتها معلوماً ولا يكون مستحيل (الاحتمال ≠ الصفر)
لا تطبق أساليب الاستدلال الإحصائي	يمكن تطبيق أسليب الاستدلال الإحصائي

<u>أسئلة وأجوبة (2)</u>

(1) فيما يلي درجات 26 طالباً في الصف الأول الثانوي في مادة الرياضيات

30 24 23 12 11 22 25 12 14 13 14 25 30

27 28 29 30 20 20 19 14 25 12 16 17 18

والمطلوب:

1) كون الجدول التكراري المناسب

2) ما هي نسبة الطلاب التي تزيد درجاتهم عن 20درجة

الحــل :

التكرار	العلامات	الفئات
8	///// ///	**15 >**
4	////	**15 - 20**
5	/////	**20 - 25**
6	///// /	**25 - 30**
3	///	**30 ≤**
26		**المجموع**

أو بدون العلامات

Interval (X)	Frequency (F)	Relative Frequency (RF)	RF %
Less Than 15	8	8 ÷ 26 = 0.3077	(8 ÷ 26) × 100 = 30.77%
15 - 20	4	0.1538	15.38%
20 - 25	5	0.1923	19.23%
25 - 30	6	0.2308	23.08%
More Than 30	3	0.1154	11.54%
Total	26	1	100.00%

نسبة الطلاب التي تزيد درجاتهم عن عشرين درجة هي:

53.85% = (14 ÷ 26) × 100

(2) إذا كان لدينا درجات مجموعتين من الطلاب في اختبار ما هي:

المجموعة الضابطة: 17 16 11 10 15 18 16 14 16 12

المجموعة التجريبية: 11 20 15 14 10 13 17 18 10 10 11 13

المطلوب إنشاء جدول تكراري مزدوج.

الحــل :

الفئات	العلامات		التكرار	
	الضابطة	التجريبية	الضابطة	التجريبية
10 - 15	/////	///// /	5	6
16 - 21	/////	///// //	5	7

أو

الفئات	التكرار	
	الضابطة	التجريبية
10 - 15	5	6
16 - 21	5	7

(3) الجدول التالي يبين استهلاك السجاير وقيمتها لمجموعة من الأفراد عددها 15والمطلوب إنشاء جدول تكراري مزدوج.

الاستهلاك	8	7	12	10	9	5	6	2	10	15	11	8	6	9	4
القيمة	5	4	7	6	5	3	4	1	9	13	7	5	3	11	2

الحــل:

نوجد أطوال الفئات وعددها

بالنسبة للاستهلاك: المدى = 15 14 = 1 + 1 = 2 – ، عدد الفئات = 1 + 3.22لو(15) = 4.91 ≈ 5، طول الفئة = 14 ÷ 5 = 2.8 ≈ 3

بالنسبة للقيمـــــة: المدى13 = 1 + 1 – 13 = ، عدد الفئات = 1 + 3.22لو(15) ≈ 5 4.91 = ()، طول الفئة = 13 ÷ 5 = 2.6 ≈ 3

نكون الجدول المزدوج حيث الاستهلاك عموديا والقيمة أفقياً كالتالي:

الجدول التكراري المزدوج

فئات القيمة → فئات الاستهلاك↓	1 - 3	4 - 6	7 - 9	10 - 12	13 - 15	المجموع
2 - 4	//					2
5 - 7	//	//				4
8 - 10		////	/	/		6
11- 13			//			2
14- 16					/	1
المجموع	4	6	3	1	1	15

يمكن إعادة كتابة الجدول السابق بحذف العلامات كما يلي:

الجدول التكراري المزدوج

فئات القيمة → فئات الاستهلاك↓	1 - 3	4 - 6	7 - 9	10 - 12	13 - 15	المجموع
2 - 4	2					2
5 - 7	2	2				4
8 - 10		4	1	1		6
11- 13			2			2

14- 16					1	1
المجموع	4	6	3	1	1	15

(4) أكتب ما تعرفه عن الفئة 15ـ 19؟ ، وكذلك عن الحدود الفعلية للفئة؟

أولاً: إذا كانت الفئة لأعداد صحيحة كعدد السكان أو الوفيات

- **الأعداد بدء من 15 وصولاً إلى 19**
- **15هو الحد الأدنى للفئة ، 19 الحد الأعلى للفئة**
- **طول الفئة 19 –15 هو 5 " 15، 16، 17، 18، 19 "**
- **الحد الفعلي الأدنى لها = 14.5 ، الحد الفعلي الأعلى لها = 19.5**
- **مركز الفئة = (15 + 19) ÷ 2 = 34 ÷ 2 = 17 أو نضيف نصف الفرق بين حديها (19 – 15) ÷ 2 = 4 ÷ 2 = 2 إلى حدها الأول (15)**
- **تسبقها الفئة 10 – 14 وتليها الفئة 20 – 24**

ثانياً: إذا كانت الفئة لأعداد كالعمر والطول والوزن

- **كل الأعداد (صحيحة وكسرية) بدء من العدد 15 إلى ما قبل العدد 19**
- **طول الفئة = 19 – 15 = 4**
- **مركز الفئة 17 = 2 ÷ 34 = 2 ÷ (19 + 15) = أو نضيف نصف الفرق بين حديها 2 = 2 ÷ 4 = 2 ÷ (15 –19)إلى حدها الأول (15)**
- **تسبقها الفئة 11 - 15 وتليها الفئة 19- 23**

الحدود الفعلية للفئات

لكل فئة حد أدنى (بدايتها) وحد أعلى (نهايتها) وقد يكونا أعداد صحيحة (ننقص 0.5) أو كسرية (ننقص 0.05 لرقم عشري واحد ، 0.005 لرقمين عشريين وهكذا)

الأعداد الصحية كما سبق في الفئة السابقة 15 ـ 19 ، الحد الفعلي الأدنى = الحد الأدنى للفئة- 0.5 ، الحد الفعلي الأعلى = الحد الأعلى للفئة + 0.5

الأعداد العشرية:

للفئة 12.1 ـ 15.1 يكون الحد الفعلي الأدنى = 12.1 – 0.05 = 12.05 والحد الفعلي الأعلى = +15.1 0.05 + .15.1 (15.15 = 0.05 إضافة الرقم 5 أمام العدد15.1)

للفئة 12.45 ـ 15.45 يكون الحد الفعلي الأدنى = 12.45 – 0.005 = 12.445 والحد الفعلي الأعلى = 15.45 + 0.005 = 15.455

للفئة 12.415 ـ 15.415 يكون الحد الفعلي الأدنى = 12.415 – 0.0005 = 12.4145 والحد الفعلي الأعلى = 15.415 + 0.0005 = 15.4155

لاحظ: الحد الفعلي الأعلى ينتج من إضافة الرقم 5 أمام العدد (يمينه) في حين الحد الفعلي الأدنى ينتج بطرح 1 من الجزء العشري ثم وضع 5 على يمين العدد كالتالي:

Interval	Lower Limits	Upper Limits	Exact Lower Limits	Exact Upper Limits
10.1 – 12.1	10.1	12.1	1 – 1 =0 , 10.05	12.15
16.4 – 20.4	16.4	20.4	4 – 1 = 3 , 16.35	20.45
32.24 – 35.24	32.24	35.24	24 – 1 = 23 , 32.235	35.245
0.18 – 0.22	0.18	0.22	18 – 1 = 17 , 0.175	0.225

للحد الفعلي الأدنى نأخذ الجزء العشري كعدد صحيح ونضربه × 1 0ونطرح من الناتج 5 ونستبدله بالجزء العشري الأصلي

فمثلاً 0.18 نأخذ 18 نضربها × 10 فنحصل على 180نطرح 5 نحصل على 175 نجعلها بدل 18 في 0.18 فيكون الحد الفعلي الأدنى = 0.175

ومثلاً 16.4 نأخذ 4 نضربها × 10 فنحصل على 40 نطرح 5نحصل على 35 نجعلها بدل 4 في 16.4 فيكون الحد الفعلي الأدنى = 16.35

ومثلاً 32.24 نأخذ 24 نضربها × 10 فنحصل على 240نطرح 5 نحصل على 235 نجعلها بدل 24 في 32.24 فيكون الحد الفعلي الأدنى = 32.235

(5) الجدول التالي يبين درجات إحدى المدارس القسم العلمي في الثانوية العامة حيث المجموع الكلي للدرجات 500

Interval (X)	Less than 50	–50	–100	–150	200–	250–	300–	350–	400–	More than 450	Total
Frequency (F)	22	14	32	11	9	4	42	38	12	16	200

والمطلوب:

1) **الجدول التكراري المتجمع الصاعد والمتجمع الصاعد النسبي.**
2) **عدد الطلاب الذين حصلوا على 70% (350 درجة) فأكثر.**
3) **حدد عدد الناجحين باعتبار الحد الأدنى لدرجة النجاح 60% (300 درجة).**
4) **حدد الدرجة التي لم يتجاوزها 40% من الطلاب.**
5) **التمثيل البياني لكل من الجدول التكراري (Histogram) والمضلع التكراري (Polygon) والمنحى التكراري (Curve).**
6) **التمثيل البياني لكل من والتكرار المتجمع الصاعد .(CF)**

الحـــل:

(1)

Cumulative Frequency (CF) التكرار المتجمع الصاعد	الفئات (X)
22	Less than 50
36	Less than 100
68	Less than 150
79	Less than 200
88	Less than 250
92	Less than 300
134	Less than 350
172	Less than 400
184	Less than 450
200	More than 450

(2) عدد الطلاب الذين حصلوا على 70% (350 درجة) فأكثر 66 = طالب (مقابل 350– في التكرار الصاعد 134وباقي 200 هو 66 من 200 .(134 –

(3) عدد الناجحين باعتبار الحد الأدنى لدرجة النجاح 60% (300 درجة) = 108 حيث 60% أي 300 درجة يقابلها 134ما قبلها راسب .(92 – 200)

(4) الدرجة التي لم يتجاوزها 40% من الطلاب هي 200

تنويه :من الممكن الحصول على النتائج (لكل من 2 ، 3 ، 4) السابقة من الجدول الأصلي كالتالي:

Total	More than 450	400–	350–	300–	250–	200–	150–	100–	50–	Less than 50	Interval (X)
200	16	12	38	42	4	9	11	32	14	22	Frequency (F)

(2) عدد الطلاب الذين حصلوا على 70% (350 درجة) فأكثر =38+ 12 + 16=66

(3) عدد الناجحين باعتبار الحد الأدنى لدرجة النجاح 60%أي 500 × 0.60أي 300 هو 42 + 38 + 12 + 16= 108 (مجموع الأعداد الأربعة الأخيرة في الجدول).

(4) الدرجة التي لم يتجاوزها 40% من الطلاب أي 0.40 × 200 = 80 طالب هي 200 (نجمع من البداية حتى ≥ 80 أي 79 أي ما قبل 200 درجة.

***

مقاييس النزعة المركزية

الوسط الحسابي أو المتوسط الحسابي أو المعدل (The Arithmetic Mean)-

أولاً: معطيات غير مبوبة (Ungrouped Data)

يعتبر الوسط الحسابي من مقاييس النزعة المركزية (الوسط - الوسيط - المنوال - ...) ويعتبر الوسط الحسابي أكثر المقاييس استخداماً ويعرف كالآتي:

الوسط الحسابي لمجموعة من القيم عددها ن هو القيمة التي لو حلت محل كل قيمة في مجموعة القيم لكان مجموع القيم الجديدة مساوياً لمجموع القيم الأصلية.

وببساطة الوسط الحسابي هو مجموع القيم (س$_1$، س$_2$ ، س$_3$ ، ... ، س$_ن$) مقسوماً على عددها (ن) ويرمز له بالرمز س أو X، وتقرأ سين بار ، أكس بار ونحن هنا بصدد العينة وفي حالة المجتمع يرمز للوسط الحسابي بالرمز μ ويقرأ ميو وعدد القيم N أي أنَّ:

للمجتمع

$$\mu = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{N}$$

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

للعينة

$$\overline{X} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

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

وفي حال وجود أكثر من مجموعة فيكون الوسط الحسابي هو:

$$\overline{X} = \frac{n_1 \overline{x}_1 + n_2 \overline{x}_2 + n_3 \overline{x}_3 + \ldots + n_n \overline{x}_n}{n_1 + n_2 + n_3 + \ldots + n_n}$$

حيث أن x_1 هو المتوسط الحسابي لمجموعة القيم التي عددها n_1

حيث أن x_2 هو المتوسط الحسابي لمجموعة القيم التي عددها n_2

حيث أن x_3 هو المتوسط الحسابي لمجموعة القيم التي عددها n_3

....

....

....

حيث أن x_n هو المتوسط الحسابي لمجموعة القيم التي عددها n_n

ثانياً: معطيات البيانات المبوبة (Grouped Data)

الوسط الحسابي من البيانات المبوبة أي الموجودة ضمن جدول تكراري، فإذا كان لدينا مجموعة من القيم المكررة والملخصة في جدول تكراري بسيط حسب القيمة مع تكرارها فالوسط الحسابي يساوي مجموع حاصل ضرب هذه القيم في تكرارها مقسوماً على مجموع التكرارات، ونستخدم القانون التالي لحساب الوسط الحسابي للبيانات المبوبة (الجدول التكراري)

$$\bar{x} = \frac{\sum f x}{\sum f}$$

حيث x مركز الفئة ، f التكرار

مثال: احسب الوسط الحسابي للبيانات المبينة في الجدول التكراري الآتي لدرجات 30 طالب في مادة الإحصاء:

الفئات	التكرار (f)
12 – 14	8
15 – 17	4
18 – 20	7
21 – 23	6
24 – 26	2
27 – 29	3
المجموع	30

الحل:

من الواضح في الجدول بأن الفئة 12 - 14 تكرارها يساوي 8 أي أن ممن حصلوا على درجات في هذه الفئة عددهم 8 فإذا فرضنا أن هؤلاء حصل كل منهم على الدرجة الواقعة في مركز الفئة وهي (12 + 14) ÷ 2 = 13 فيكون هؤلاء الثمانية قد حصلوا على درجات مجموعها 8 × 13 = 104 وعليه يكون الباقي والخطوات هي:

1) نستخرج مراكز الفئات بإضافة عمود جديد للجدول

2) نضرب مركز كل فئة × التكرار المقابل لها

3) نحصل على الجدول الآتي:

الفئات	التكرار (f)	مركز الفئة (x)	$f \times x$
12 – 14	8	13	104
15 – 17	4	16	64
18 – 20	7	19	133
21 – 23	6	22	132
24 – 26	2	25	50
27 – 29	3	28	84
المجموع	30		567

4) نحسب الوسط الحسابي من القانون:

$$\bar{x} = \frac{\sum f x}{\sum f}$$

$$= \frac{567}{30} = 18.9$$

<u>خواص المتوسط الحسابي</u>

- **يعتمد على جميع القيم والمشاهدات محل الدراسة**
- **هو نقطة اتزان المشاهدات**
- **انحراف الدرجة عن المتوسط يساوي بعدها عنه.**
- **المجموع الجبري لانحرافات القيم عن وسطها الحسابي يساوي صفراً**
- **مربع الانحرافات اقل ما يمكن عن الوسط**
- **إضافة قيمة ثابتة إلى الدرجات أو طرحها منها أو ضربها فيه أو قسمتها عليه مما يجعل المتوسط الحسابي يزداد أو يقل بقيمة ثابتة.**
- **المتوسط الحسابي يتأثر بالدرجات القريبة منه تأثرا قليلا ، بينما يتأثر بالدرجات البعيدة عنه تأثرا كبيرا .**
- **اقل مقاييس النزعة المركزية تأثرا بالتقلبات العينية**
- **يتأثر بالقيم المتطرفة والقيم الشاذة لذا لا يصلح للتوزيعات الملتوية بمعنى أن الوسط الحسابي للقيم 20، 30 ، 40 ،50 ، 120 هو 260 ÷ 5 = 52 أكبر من غالبية القيم الموجودة بسبب القيمة 120 الشاذة عن باقي القيم الأخرى المتقاربة فيما بينها وهذا يقودنا لعدم صلاحية الوسط الحسابي لتوزيعات تكرارية شديدة الالتواء.**

<u>الحالات التي لا يصلح المتوسط الحسابي فيها :</u>

1. **عندما تكون قيم التوزيع متجمعة في طرف واحد أكثر من الطرف الأخر.**
2. **إذا كانت هناك قيم شاذة تتخلل التوزيع.**
3. **إذا كانت البيانات مبوبة في فئات وكان التوزيع مفتوحا في أحد طرفيه.**
4. **إذا كانت الفئات متباعدة نسبيا.**
5. **لا يمكن إيجاده من الرسم البياني كالوسيط والمنوال.**
6. **لا يمكن إيجاده من التوزيع التكراري المفتوح إلا في حال تقدير مراكز الفئات المفتوحة.**
7. **عدم صلاحية الوسط الحسابي لتوزيعات تكرارية شديدة الالتواء.**

مثال (1): أوجد المتوسط الحسابي لمجموعة القيم 3، 5، 8، 11، 3

الحــــــل:

مجموع القيم = 3 + 5 + 8 + 11 + 3 = 30

عدد القيم = 5

الوسط الحسابي = 30 ÷ 5

= 6

مثال (2): إذا كان الوسط الحسابي لمجموعة القيم 2، 5، 9، 13، س ، 6 يساوي 7 فما قيمة س؟

الحــــــل: مجموع القيم = 2+ 5 +9 +13 + س + 6 = 35 + س

عدد القيم = 6

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

$$\bar{X} = \frac{d\ 2 + 5 + 9 + 13 + x + 6}{6}$$

$$7 = \frac{d\ \ 35 + x}{6}$$

$$42 = 35 + x$$

$$x = 42 - 35$$

$$x = 7$$

(1) باستخدام البيانات الواردة في الجدول التالي لدرجات 60 طالب في مادة الإحصاء احسب الوسط الحسابي

Total	70 - 79	60 - 69	50 - 59	40 - 49	30 - 39	20 - 29	10 - 19	Intervals
60	7	9	14	12	8	6	4	Frequency

الحــل:

<u>باستخدام مراكز الفئات:(Mid Interval)</u>

نكون جدول تكراري يضم مراكز الفئات وآخر يشمل F × X بالشكل التالي:

F × X	مركز الفئة (X)	التكرار(F)	الفئات
58	14.5	4	i10 - 19
147	24.5	6	i20 - 29
276	34.5	8	i30 - 39
534	44.5	12	i40 - 49
763	54.5	14	i50 - 59
580.5	64.5	9	i60 - 69
521.5	74.5	7	i70 - 79

المجموع	60		2880

الوسط الحسابي = 2880 ÷ 60

= 48

***

الوسيط Median

الوسيط هو أحد مقاييس النزعة المركزية

يعرف الوسيط بالقيمة المشاهدة التي تتوسط البيانات بعد ترتيبها تصاعدياً أو تنازلياً. وعليه تكون عدد القيم (المفردات) قبله مساوي لعدد القيم بعده وله حالات ثلاث:

1) عدد المفردات فردي يساوي n مثلاً فيكون ترتيب الوسيط ½(n + 1) وقيمة الوسيط للمفردات 3، 5 ، 9 ، 11 ، 13 هي 9 (المفردات مرتبة تصاعدياً).

2) عدد المفردات زوجي فتوجد قيمين في الوسط ترتيبهم 2/n ، 1+2/n والوسط الحسابي لقيمهم هو الوسيط فالمفردات 3 ، 5 ، 9 ، 11 ، 13 ، 16 فالوسيط لهذه المفردات هو الوسط الحسابي للقيمتين (ترتيبهم 6÷2 = 3 ، (6÷2) +1=4) أي 9، 11 هو (9+11)÷2 = 10 واضح أن العدد 10 قبله 3 مفردات وبعده 3 مفردات.

3) حساب الوسيط من جدول تكراري (البيانات المبوبة) من جدول التكرار المتجمع الصاعد أو النازل أو الرسم البياني لكلاهم أو أحدهم وسنورد هنا تفصيلاً كاملاً لذلك.

حساب الوسيط من البيانات المبوبة:

سبق وأن ذكرنا كيفية تكوين الجدول التكراري المتجمع (المتراكم) الصاعد وسنرمز للوسيط بالرمز MD وللتكرار المتجمع الصاعد السابق للفئة الوسيطية CF_{k-1} ولتكرار الفئة الوسيطية بالرمز FR_{xk} وللحد الأدنى للفئة الوسيطية بالرمز L_k والوسيط من الجدول التكراري المتجمع يكون (بافتراض التوزيع منتظم داخل الفئة الوسيطية:

$$MD = L_k + \frac{\frac{n}{2} - CF_{k-1}}{FR_{xm}} \times H$$

الوسيط = الحد الأدنى الفعلي للفئة الوسيطية +

$$\frac{\text{ترتيب الوسيط – التكرار المتجمع للفئة السابقة للفئة الوسيطية}}{\text{تكرار الفئة الوسيطية}^{*}} \times \text{طول الفئة} \quad .... (1)$$

* تكرار الفئة الوسيطية = التكرار المقابل للحد الأعلى للفئة الوسيطية – التكرار المقابل للحد الأدنى للفئة الوسيطية

مثال:

الجدول التكراري الآتي يبين بيانات أعمار 30iمريض لمراجعتهم المستشفى والمطلوب حساب الوسيط بكل الطرق الممكنة.

الفئات	12 – 14	15 – 17	18 – 20	21 – 23	24 – 26	27 – 29	Total
التكرار	8	4	7	6	2	3	30

الحل:

الفئات	الحدود الفعلية للفئات	التكرار (f)	التكرار التراكمي الصاعد (cf_u)	التكرار التراكمي النازل (cf_d)	$cf_u + cf_d$
.... – 12	.... – 11.5	0	0	30	30
12 – 14	11.5 – 14.5	8	8	22	30
15 – 17	14.5 – 17.5	4	12	18	30
18 – 20	17.5 – 20.5	7	19	11	30
21 – 23	20.5 – 23.5	6	25	5	30
24 – 26	23.5 – 26.5	2	27	3	30
27 – 29	26.5 – 29.5	3	30	0	30
Total		30			

- لاحظ أن الصفر في الخانة الأولى في تكرار المتجمع الصاعد ناتج من القيم التي أقل من 11.5 والقيمة 30 في تكرار المتجمع النازل ناتجة من القيم التي أكثر من 11.5وفي الغالب لا تذكر تلك القيم بل نبدأ من أول أعلى حد فعلي (14.5).
- من ترتيب الوسيط (2/n) وهنا الترتيب 15 نبحث عنه في التكرار المتجمع الصاعد (أو النازل) فنجده
- موجود في قيم التكرار المتجمع الصاعد وليكن الترتيب 25 مثلاً فإن الوسيط القيمة المقابلة للقيمة 25 في الحدود الفعلية للفئات أي الوسيط = 23.5
- إذا كان ترتيب الوسيط 12 مثلاً فقيمة الوسيط 17.5 (الحد العلوي الفعلي المقابل للقيمة 12)
- إذا كان ترتيب الوسيط 19 مثلاً فقيمة الوسيط 20.5 (الحد العلوي الفعلي المقابل للقيمة 19)
- في حالة عدم وجوده كحالنا هن 15 فقيمة الوسيط تقع بين 17.5 ، 20.5 وسنبين ذلك أدناه ويكون $17.5 < MD < 20.5$
- قيم التكرار في عمود المتجمع الصاعد تنتج من قولنا أقل من 11.5 لا يوجد أو صف لم تدون في الجدول لعدم الأهمية وأقل من 14.5 يوجد 8 وأقل من 17.5 يوجد 8+4 وأقل من 20.5 يوجد 8+4+7 وأقل من 23.5 يوجد 8+4+7+6 وهكذا ...
- وقيم التكرار في عمود المتجمع النازل تنتج من قولنا أكثر من 14.5 يوجد 30–8=22 وأكثر من 17.5 يوجد 30–(98+4) وأكثر من 20.5 يوجد 30–(8+4+7) وأكثر من 23.5 يوجد 30–(8+4+7+6) وهكذا ... (يمكن الحصول عليه من المتجمع الصاعد بتكملة كل عدد إلى 30 مثل 8 تكمل إلى 22، 12 إلى 18، 19 إلى 11 وهكذا ...
- الوسيط كما هو مبين يهتم بالقيم التي تقع حول ترتيبه (15 هنا) أي القيم التي تتوسط البيانات على عكس الوسط الحسابي الذي يأخذ بالاعتبار كل القيم.
- وجود قيم شاذة أو متطرفة يجعل التوزيع غير متماثل فتختلف قيم الوسيط عن الوسط فيكون الوسط > الوسيط للقيم الشاذة الكبيرة والعكس الوسط < الوسيط حال وجود قيم شاذة صغيرة، وسنوضح ذلك لاحقاً عند دراسة المنوال.

<u>الحل باستخدام القانون (1) أعلاه</u>

يجب أن نعلم بأن الفئات تمثل على المحور السيني والتكرار يمثل على المحور الصادي كما سبق ذكر ذلك فقيمة ترتيب الوسيط (15هنا) تمثل على المحور الصادي (محور التكرار) يقابلها قيمة على المحور السينات ضمن فئتها التي تقع أسفل النقطة المقابلة للترتيب والواقعة على منحنى التكرار المتجمع ومن الجدول وحيث 15 هو ترتيب الوسيط وفي عمود التكرار المتجمع (أعلاه) نجد 15 تقع بين 12 ، 19 يقابلهم 17.5 ، 20.5 كما مبين في الجدول التالي.

ترتيب الوسيط = 30 ÷ 2 = 15 (مجموع التكرار ÷ 2) فالوسيط يقع بين القيمتين 12 ، 19 أي الفئة الوسطية هي 17.5 – 20.5 أي

17.5	12
Median = 17.5 + ...	15
20.5	19

فبتطبيق القانون أعلاه (1) نجد أن:

$$\text{الوسيط} = 17.5 + \frac{15-12}{19-12} \times 3 = 17.5 + 1.29 = 18.79$$

<u>باستخدام النسبة والتناسب</u>

وبملاحظة الجدول أعلاه نجد أن (الفئات) 20.5 – 17.5 = 3 (طول الفئة) قابلة مسافة (تجاوزاً) 19 – 12 = 7 أي لكل 3 في الفئات نحصل على 7 هنا أي لكل جزء في التكرار يقابله 3 ÷ 7 من الفئة ولكننا نريد 15 التي بينها وبين 12 مسافة 3 (أجزاء من 7) فلها من الفئات قيمة = 3 × (3 ÷ 7) = 9 ÷ 7 = 1.29 نضيفها على 17.5 بداية الفئة 17.5 + 1.29 = 18.79 وهي قيمة الوسيط . ماذا لو قمنا بالعكس أي قلنا 20.5 (في الفئة) يقابلها 19 (التكرار المتجمع) التي بينها وبين 15 مسافة 4 فيكون لها في الفئة قيمة (أي 4) تساوي 4 × (3 ÷ 7) = 1.71 تنقص من 20.5 أي 20.5 – 1.71= 18.79 وهو نفس الناتج ويمكن تبسيط ذلك كالتالي

15 تقع بين 12 ، 19 فنقول 7 (19– 12) لها 3 (20.5 – 17.5) وعددنا 15 يبعد عن 12 مسافة 3 فكم يقابله ؟ أي

3 ← 7

؟ ← 3

؟ = 3 × 3 ÷ 7= 1.29 ، يضاف لبداية الفئة 17.5 فنحصل على الوسيط = 17.5 + 1.29 = 18.79

وللتكرار المتجمع الهابط :

15 تقع بين 12 ، 19 فنقول 7 (19– 12) لها 3 (20.5 – 17.5) وعددنا 15 يبعد عن 19 مسافة 4 فكم يقابله ؟ أي

3 ← 7

؟ ← 4

؟ = 3 × 4 ÷ 7= 1.71 ، يطرح من نهاية الفئة الوسيطية 20.5 فنحصل على الوسيط = 20.5 – 1.71 = 18.79

<u>باستخدام الرسم البياني</u>

يمكن ذلك برسم محورين أفقي للفئات ورأسي للتكرار المتجمع - ثم نعين النقاط في المستوى (14.5، 8)،(17.5، 12)،... (29.5 ، 30)،

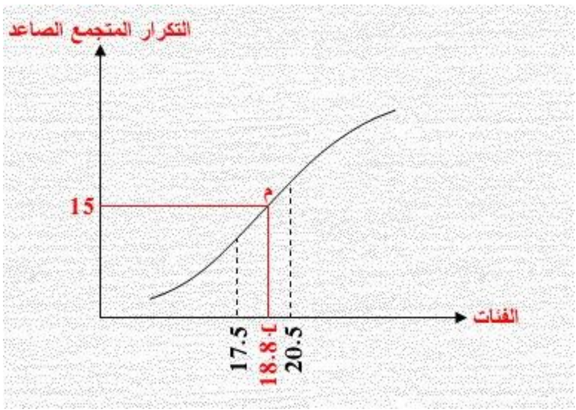
نصل بين النقاط باليد نحصل على المنحنى التكراري المتجمع الصاعد كما مبين بالشكل.

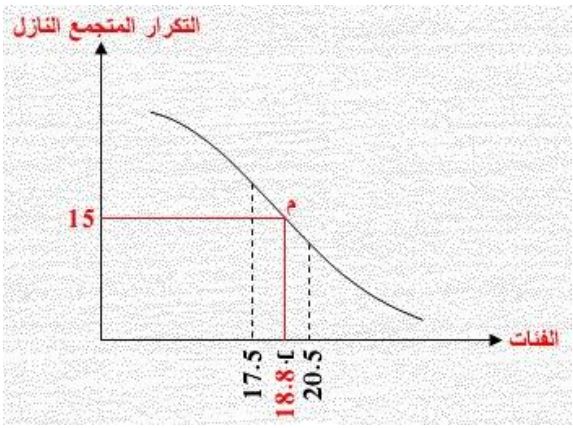
من العدد 15 (ترتيب الوسيط) نرسم مستقيماً أفقياً يقطع المنحنى في نقطة م مثلاً.

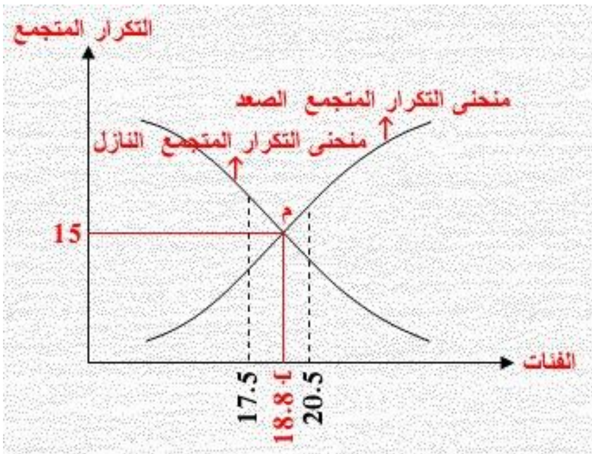
من نقطة م نسقط عموداً على المحور الأفقي يلاقيه في نقطة ولتكن ب العدد عند النقطة ب هو قيمة الوسيط والقيمة هنا هي 18.8

يمكن تكرار ذلك مع المنحنى التكرار المتجمع الهابط ونصل لنفس الناتج لاحظ الشكل.

رسم المنحنيين الصاعد والنازل ومن نقطة تقاطعهم (م) نسقط العمود على المحور الأفقي (محور الفئات) لنحصل على قيمة الوسيط 18.8 (لاحظ الشكل).







<u>مميزات الوسيط</u>

أ) عدم تأثره بالقيم الشاذة.

ب) تعتمد قيمته على التكرارات.

ج) يوجد في كل البيانات التي تتصف بالترتيب

<u>عيوب الوسيط</u>

ا) يهتم بالقيم الوسطى للبيانات فقط

ب) لا يصلح لإعطاء فكرة عن النزعة المركزية للبيانات في فئات متباعدة

ج) لا يسهل التعامل معه في التحاليل الإحصائية في دراسة ظاهرة ما في عدة عينات.

(3) كيف يحسب الوسيط بيانياً؟

أ) تكوين جدول تكراري متجمع صاعد أو هابط أو كلاهما

ب) نحسب ترتيب الوسيط ويساوي نصف مجموع التكرار

ج) نحدد قيمة الترتيب على محور التكرار

د) نرسم خط أفقي من قيمة الترتيب (الواقعة على محور التكرار) ليلاقي المنحنى الصاعد أو النازل في نقطة (في المنحنيين النازل والصاعد هي نقطة تقاطعهم)

هـ) من النقطة على المنحنى نسقط عموداً على الخط الأفقي (الفئات) والقيمة هنا هي قيمة الوسيط.

***

المنوال Mode

يعرف المنوال لمجموعة من القيم بأنه القيمة الأكثر شيوعاً بينها، وقد يوجد أكثر من منوال لمجموعة القيم.

البيانات غير المبوبة:

مجموعة القيم 12 ، 13، 8 ، 9 ، 12 ، 12، 23 لها منوال واحد هو القيمة 12

مجموعة القيم 12 ، 13، 8 ، 8 ، 12 ، 12، 23 يوجد أكثر من منوال 12، 8

يمكن استخدام المنوال للقيم الكمية والنوعية ولكن يكثر استخدامه مع المعطيات النوعية.

في حال وجود أكثر من منوال (متجاورة) فمتوسط هذه المنوالات يعتبر منوال التوزيع حال وجود تلك القيمة في التوزيع.

يسمى التوزيع أُحادي المنوال إذا وجد منوال واحد، وثنائي المنوال إذا وجد منوالان، وهكذا.

إذا وجد أكثر من منوال فالأكثر تكرار بينهم يدعى المنوال الرئيسي (Major Mode) والأخرى تعرف بالمنوال الفرعي أو الثانوي (Minor Mode)

المنوال ببساطة يمثل بقيمة مسقط أعلى نقطة في المنحنى التكراري على محور الفئات.

المنوال لا يتأثر بالقيم المتطرفة وهو يمثل غالبية القيم.

تقل قيمته كمتوسط للقيم حال تباعد القيم عن بعضها البعض.

في حالة البيانات المبوبة:

توجد أكثر من طريقة لحساب المنوال

الطريقة الحسابية نوضحها من خلال جدول البيانات التالي فإن الفئة المنوالية (التي تضم المنوال) تقابل أكبر تكرار وهنا أما نستخدم قانون خاص بالمنوال أو نقوم بعملية حسابية وسنبين هذا من خلال المثال التالي:

جدول التوزيع التكراري الآتي يبين درجات 50 طالب في مادة الإحصاء والمطلوب حساب المنوال.

الفئات	12 – 14	15 – 17	18 – 20	21 – 23	24 – 26	27 – 29
التكرار	11	10	15	7	5	2

الحل:

بالصيغة الرياضية الآتية:

إذا رمزنا لطول الفئة بالرمز I وللفرق بين تكرار الفئة المنوالية و تكرار الفئة السابقة للفئة المنوالية بالرمز F_n والفرق بين تكرار الفئة المنوالية وتكرار الفئة اللاحقة للفئة المنوالية بالرمز F_b ولبداية الفئة المنوالية بالرمز L فإن المنوال يحسب من الصيغة الآتية:

$$\text{Mode} = L + \frac{F_n}{F_b + F_n} \times I$$

بتطبيق الصيغة هذه على المثال السابق نجد أن

$$\text{Mode} = 18 + \frac{5}{5+8} x3$$

$$\text{Mode} = 18 + 1.54$$

$$\text{Mode} = 19.54$$

<u>الطريقة البيانية:</u>

من الجدول التكراري كما مبين بالشكل حيث م نقطة تقاطع المستقيان الواصلان من بداية الفئة المنوالية لبداية الفئة اللاحقة ، من نهاية الفئة المنوالية لنهاية الفئة السابقة، ومسقط م على المحور الأفقي يعطي قيمة المنوال.

<u>ملاحظة</u>

إذا كانت الفئة المنوالية أول الأعمدة أو آخرها ،فنعتمد متوسط العمود(مركز الفئة المنوالية) هي المنوال

مثال : الجدول التكراري الآتي، أوجد قيمة المنوال بطرقه المختلفة.

Interval	20 –	24 –	28 –	32 –	36 –	40 –	44 –	Total
Frequency	8	10	15	34	14	11	8	100

الحل :

باستخدام القانون السابق

$$\text{Mode} = L + \frac{F_n}{F_b + F_n} \times I$$

$$\text{المنوال} = 32 + \frac{19}{19 + 20} \times 4$$

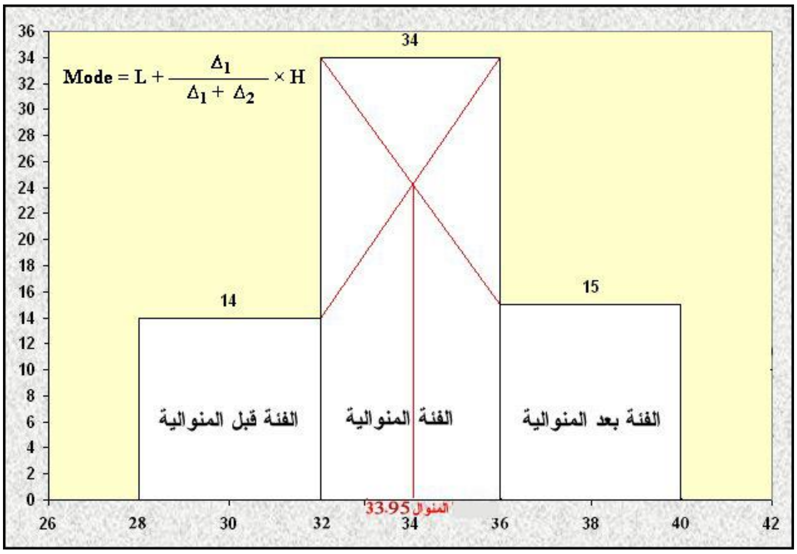
$$\text{المنوال} = 32 + \frac{76}{39}$$

$$\text{المنوال} = 32 + 1.95$$

$$\text{المنوال} = 33.95$$

باستخدام جزء من المدرج التكراري:

يمكن الاكتفاء بتمثيل الفئات المنوالية وما قبلها وبعدها (مدرج تكراري) كما مبين بالشكل التالي:



***

مقاييس التشتت (Measures of Variation)

المقدمة:-

مقاييس النزعة المركزية الوسط والوسيط والمنوال قيم تتمركز أقرب ما يمكن للقيم المشاهدة وقد بينا بيانياً بأن هذه القيم تمثل بنقط على محور الفئات وكان من أهمها الوسط الحسابي حيث بينا أن مجموع انحرافات المشاهدات عنه يساوي صفر ولكننا لو أخذنا مجموعتان أو أكثر فقد يكون لها وسط حسابي بنفس القيمة مع اختلاف تشتتها (انتشارها) حول الوسط وعن بعضها أو العكس تتشابه في درجة التشتت وتختلف في وسطها فلهذا لا بد من دراسة بُعد القيم عن متوسطها وهو ما يعرف بالتشتت

فمثلاً:

القيم لإنتاج سلعة ما 42 ، 35 ، 45 ، 56 ، 32 وسطها الحسابي 42 القيم لإنتاج نفس السلعة من مصنع آخر 30 ، 40 ، 60 ، 20 ، 60 وسطها الحسابي 42 المجموعة الأولى أكثر تجانساً من المجموعة الثانية بالرغم من تساوي الوسطين.

مقاييس النزعة المركزية غير كافية لوصف البيانات من حيث تفاوت البيانات عن وسطها (تشتتها) فالحاجة استدعت مقاييس أخرى تعرف بمقاييس التشتت أو مقاييس الانتشار كالمدى ونصف المدى الربيعي والانحراف المتوسط والتباين والانحراف المعياري ومعامل الاختلاف وأخرى.

الهدف من دراسة الإحصاء هو وجود هذا التشتت بين القيم فلو اتفق بأن أوزان الإنتاج لسلعة واحدة متساوية في المواصفات أو كان كل الناس لهم نفس الطول أو لهم نفس فصيلة الدم أو ... لما دعت الحاجة لوجود هذا العلم فالتفاوت بين المخلوقات في خصائصها أو التفاوت في التحصيل الدراسي وما إلى ذلك دعتنا لإيجاد مقاييس التشتت.

المـدى Range:-

المدى لمجموعة من القيم هو الفرق بين أكبر وأصغر قيمة فيها.

المدى لمجموعة القيم 23 ، 34 ، 33 ، 12 ، 40 هو 40 – 12 = 28 ومن الواضح أن المدى يهتم فقط بتلك القيمتين ولا يتأثر بالقيم الأخرى ويعتبر من أبسط مقاييس التشتت ولا يعتبر مقياس مهم للتشتت، وكلما صغرت قيمة المدى قل تشتت المجموعة فمجموعة القيم 17 ، 24 ، 13 ، 33 ، 22 المدى المطلق لها 33 – 13 = 20 أقل تشتت من المجموعة السابقة التي مداها 28 ويصح القول بأن المجموعة الأولى (المدى 28) أكثر تشتت من المجموعة الثانية (المدى 20)

والمدى لدرجات الحرارة في الاسكيمو لخمسة أيام متتالية كانت – 23 ، –17 ، – 15 ، – 20 ، – 9 فالمدى = – 9 – (–23) = – 9 + 23 = 14

في حالة البيانات المبوبة فيكون:

المدى = قيمة الحد الفعلي للفئة العليا – قيمة الحد الفعلي للفئة الدنيا

فالجدول التالي يبين درجات 30 طالب في مادة الرياضيات والمطلوب المدى لهذه الدرجات

الفئات	24 – 28	29 – 33	34 – 38	39 – 43	44 – 48	48 – 52
التكرار	3	4	7	6	8	2

الحد الأعلى للفئة العليا = 52

الحد الأدنى للفئة الدنيا = 24

المدى = 52 – 24 = 28

مما أوردناه نجد أن:

سهولة حساب المدى.

لا يمكن حسابه من توزيع تكراري مفتوح.

يتأثر بالقيمتين المتطرفتين (الكبرى والصغرى).

لا يتفق مع المفهوم القائل بأن زيادة مفردات العينة لظاهرة ما يؤدي لزيادة التجانس لتوزيع الظاهرة، فالقيم 45 ، 47 ، 49، 44، 43 ، 70 فالمدى 27 وبحذف القيمة الأخيرة يكون المدى 6 والقيم هنا متجانسة وغير متباينة ولكن وجود القيمة 70 عكس الصورة بعدم التجانس بين أفراد المجموعة.

بالرغم من عيوب المدى فإنه شائع الاستخدام وخاصة في درجات الحرارة اليومية وفي الإنتاج لأنه في الغالب الوحدات المنتجة متساوية فيقل تأثير حجم العينة على المدى، وبورصة الأوراق المالية (أسعار الأسهم المتداولة في اليوم)

ذو علاقة مع الانحراف المعياري (سيدرس لاحقاً) كتأكيد على صحة الانحراف المعياري حيث أن الانحراف المعياري لا يزيد أو لا يقل عن سبعة أمثال المدى فإن تحقق ذلك يعني صحة القيمة المحسوبة وإلا احتمال الخطأ في القيمة المحسوبة للانحراف المعياري.

قد يعطي نتيجة خاطئة للمقارنة بين مجموعتين مختلفتان في الحجم.

تمارين

1) إذا كانت البيانات 3i ، 2 ، 7 ، 4 ، 5 ، 1 فما المدى ؟

المدى = أكبر قيمة – أصغر فيمه = 7i – 1 = 6

2) أوجد المدى للبيانات المبينة في الجدول الآتي:

X	10 – 19	20 – 29	30 – 39	40 – 49	50 – 59	60 – 69	70 – 79
Y	2	6	4	10	4	3	8

المدى = الحد الأعلى للفئة العليا – الحد الأدنى للفئة الدنيا

= 79i – 10

= 69i

4) هل يمكن حساب المدى في جدول البيانات إذا احتوى على فترات مفتوحة

لا

5) هل يمكن استخدام المدى كمؤشر إضافي؟

نعم ، في حال تسوى المتوسطات. فإذا تقدم مجموعة من الأفراد لشغل وظيفة تتطلب الإجابة عل أربعة اختبارات كل منها من 100 درجة وكانت النتائج للأفراد الأربع مبينة بالجدول التالي حيث في حالة تساوي المتوسطات نأخذ أقلهم مدى لقلة التشتت:

Name	D1	D2	D3	D4	Mean	Range
Mohammed	82	90	72	68	78	8
Ali	88	80	84	60	78	28
Jamal	78	84	70	80	78	14
Mansoor	80	86	66	80	78	20

فتكون الوظيفة من نصيب Mohammed لكون المدى بين درجاته يساوي 8iوهو الأقل عن مدى الآخرين.

6) ما عيوب المدى؟

عيوبه تأثره بالبيانات الشاذة أو المتطرفة إلا أنه حسابه مباشر للتشتت.

7) المدى للمجموعتين A و B هو 33i ، 30على الترتيب فأي المجموعتين الأكثر تجانساً؟

الأقل مدى (تشتت) هي الأكثر تجانساً أي المجموعة الثانية والتي مداها 30i

8) الإنتاج اليومي لعينة من 60iفرداً يبنه الجدول الآتي والمطلوب حساب المنوال.

Interval	11 – 20	21 – 30	31 – 40	41 – 50	51 – 60	61 – 70	71 – 80
Employ No.	3	6	21	17	3	3	7

المدى = الحد الأعلى للفئة العليا – الحد الأدنى للفئة الدنيا

= 11 – 80i

= 69i

***

الانحراف المتوسط: Mean Deviation M.D Or D_m Deviation The Average

الانحراف المتوسط أو متوسط الانحرافات هو ناتج مجموع القيمة المطلقة (الناتج الموجب) لانحرافات القيم عن وسطها الحسابي مقسوما على عددها، ويرمز له بالرمز D_m

$$D_m = \left(\sum | X_i - \bar{X} | \right) / n \quad , i = 1, 2, 3, 4, \dots, n$$

تنبيه: القيمة المطلقة للعدد X تكتب | X | وتقرأ مقياس X أو مطلق X وإن |5|i = 5 ، | –5 | = 5

مثلاً: أوجد الانحراف المتوسط للقيم 14i ، 16 ، 10 ، 8 ، 2

مجموع قيم المشاهدات = 14 i + 16 + 10 + 8 + 2 = 50

الوسط الحسابي = 50 i ÷ 5 = 10

الانحرافات عن الوسط الحسابي هي |i|10– 14| ، |10 – 16| ، |10 – 10| ، |10 – 8| ، |10 – 2| أي 4i ، 6 ، 0 ، 2 ، 8

مجموع الانحرافات = 4i+ 6 + 0 + 8 + 2 = 20

الانحراف المتوسط = 20i ÷ 5 = 4

المعطيات المبوبة:

نحسب الوسط الحسابي بعد ضرب مركز كل فئة في تكرارها ثم حساب الانحرافات عن الوسط واستخدام القانون

$$i = 1, 2, 3, 4, \dots, n \ ، \quad D_m = \sum f_i \,|\, X_i - \bar{X} \,| \,/\, \sum f_i$$

مثال:

احسب الانحراف المتوسط من جدول التوزيع التكراري الآتي والذي يبين درجات 30i طالب في امتحان ما.

الحل:

الفئات	التكرار fi	مركز الفئة (X_i)	$f_i X_i$	$\| X_i - \bar{X} \|$	$f_i * \| X_i - \bar{X} \|$
12 – 14	3	13	39	| 13 – 18.7 | = 5.7	17.1
15 – 17	8	16	128	| 16 – 18.7 | = 2.7	21.6
18 – 20	10	19	190	| 19 – 18.7 | = 0.3	3.0
21 – 23	7	22	154	| 22 – 18.7 | = 3.3	23.1
24 – 26	2	25	50	| 25 – 18.7 | = 6.3	12.6
Total	30		561		77.4

الوسط الحسابي = i(39 + 128 + 190 + 154 + 50) ÷ 30 = 18.7 ونحسب الانحراف المتوسط D_m من القانون أعلاه أي:

$$D_m = (\sum f_i \,|\, X_i - \bar{X} \,|) \,/\, \sum f_i = 77.4 / 30 = 2.58$$

تنويه:

الانحراف المتوسط أكثر دقة من المدى والانحراف الربيعي لشموله كل القيم ولكنه محدود الاستخدام لتأثره بالقيم الشاذة وتجاهله الإشارة السالبة، كما يمكن حساب الفرق عن طريق الوسيط بدلاً من الوسط الحسابي حيث يكون المجموع أصغر ما يمكن إلا أن الحساب عن طريق الوسط الحسابي هو الأكثر شيوعاً، ومع ذلك تظل أهمية هذا المقياس محدودة.

تمارين

1) ما مزايا وعيوب الانحراف المتوسط ؟

يمتاز بالدقة عن كل من المدى والانحراف الربيعي لشموله كل القيم

لكنه: قليل الاستخدام في التحليل الإحصائي

يتأثر بالقيم الشاذة

إهماله الإشارة السالبة

2) احسب الانحراف المتوسط للبيانات المبينة بالجدول الآتي:

Interval	2 – 4	5 – 7	8 – 10	11 – 13	14 – 16
Frequency	7	9	25	30	10

نكون الجدول الآتي:

Interval (X)	Frequency (F_i)	X_i	$F_i X_i$	$\lvert X - \bar{X} \rvert$	$F_i \lvert X - \bar{X} \rvert$
2 – 4	7	3	21	7	49
5 – 7	9	6	54	4	36
8 – 10	25	9	225	1	25
11 – 13	30	12	360	2	60
14 – 16	10	15	150	5	50
Total	81		810		220
$\bar{X} = \sum(F_i X_i) / \sum F_i = 810 / 81 = 10$					

$D_m = (\sum f_i \mid X_i - X \mid) / \sum f_i$

$= 220 / 81$

$= 2.716$

***

التباين والانحراف المعياري

حساب الانحراف المعياري من البيانات المبوبة

سبق أن ذكرنا بأن مجموع انحرافات القيم عن وسطها الحسابي يساوي صفر وقولنا المجموع يساوي الصفر يعني وجود فروق سالبة وأخرى موجبة وللتخلص من الفروق السالبة قمنا بأخذ الانحراف المطلق أي بضرب الفرق السالب بسالب 1 وعرفنا ذلك بالانحراف المتوسط وتوجد طريقة أخرى للتخلص من الفروق السالبة هذه وذلك بتربيعها لتصبح موجبة ومجموع مربعات الانحرافات للقيم عن وسطها الوسط الحسابي يعرف بالتباين (Variance) في حين الجذر التربيعي لهذا المجموع (مجموع مربعات الانحرافات) يعرف بالانحراف المعياري (Standard Deviation) ، فالتباين أحد مقاييس التشتت.

التباين:

و مقياس لاختلاف البيانات وتشتتها، وهو متوسط مربعات انحرافات القيم عن وسطها الحسابي، ويرمز له بالرمز S^2 ويحسب من الصيغة الرياضية الآتية:

$$S^2 = [\sum (x_i - \bar{X})^2] / n \qquad , i = 1, 2, 3, ..., n$$

ويمكن القسمة على $n - 1$ في حالة العينة وهو ما يعرف بالقيم الحرة أو درجات الحرية حيث القيمة المتبقية من n يكمل انحرافها عن الوسط الحسابي للصفر لأن مجموع انحرافات القيم عن وسطها يساوي الصفر.

$$S^2 = [\sum (x_i - \bar{X})^2] / (n - 1) \qquad , i = 1, 2, 3, ..., n$$

حيث S^2 تباين العينة.

التباين يتعامل مع مربع الانحراف عن الوسط وهذا يعطي قياس غير ذو معنى مثل مربع الكيلوجرام أو مربع الدينار ولذا يفضل إرجاع ذلك (بأخذ الجذر التربيعي) للمعنى المقبول مثل الكيلوجرام والدينار وما إلى ذلك من وحدات وهذا الجذر التربيعي هو الانحراف المعياري لعينة ما.

<u>الانحراف المعياري:</u>

ببساطة نقول إن الانحراف المعياري هو الجذر التربيعي للتباين، ومن الملاحظ أن التباين يقاس بالوحدات المربعة وليس بوحدات المتغير والانحراف المعياري يقاس بنفس وحدات المتغير محل ظاهرة الدراسة.

الانحراف المعياري هو أفضل مقاييس التشتت وأشهرها استخداماً بالرغم من صعوبة حساباته حال كبر حجم العينة ولكن الحاسب الآلي سهل هذه الصعوبة.

تستخدم الصيغ الرياضية السابقة لحساب الانحراف المعياري سواء S للعينة أو σ للمجتمع

<u>حساب الانحراف المعياري:</u>

أولاً: بيانات غير مبوبة

<u>مثال:</u>

احسب كلاً من التباين والانحراف المعياري للقيم i 12 ، 15 ، 11 ، 17 ، 18 ، 20 ، 19

<u>الحل:</u>

نكون جدول المعلومات التالي:

$(X_i - \bar{X})^2$	$(X_i - \bar{X})$	X_i
16	12 – 16 = –4	12
1	15 – 16 = – 1	15
25	11 – 16 = – 5	11
1	17 – 16 = 1	17
4	18 – 16 = 2	18
16	20 – 16 = 4	20
9	19 – 16 = 3	19
$\sum (X_i - \bar{X})^2 = 72$	$\bar{X} = 112/7 = 16$	$\sum X_i = 112$

نحسب التباين من القانون أعلاه:

$$S^2 = [\sum (x_i - \bar{X})^2] / (n)$$

$$= 72 / 7$$

$$= 10.3$$

الانحراف المعياري يساوي الجذر التربيعي للتباين أي:

$$S.D = 3.20$$

حل آخر باستخدام القيم دون الوسط الحسابي بعد وضع القانون أعلاه في صورة جديدة كما يأتي:

$$S^2 = [\sum (x_i - \bar{X})^2] / (n - 1)$$

$$S^2 = [\sum (X_i^2) - (\sum X_i)^2 / n] / (n) \;...\; (1)$$

Or

$$S^2 = [(\sum X_i^2) - n\bar{X}^2] / (n) \;...\; (2)$$

الجدول الآتي هو تعديل للجدول أعلاه:

X_i^2	X_i
144	12
225	15
121	11
289	17
324	18
400	20
361	19
$\sum X_i^2 =$ 1864	$\sum X_i = 112$

بتطبيق هذه الصيغة رقم (1):

$$S^2 = [\sum (X_i^2) - (\sum X_i)^2 / n] / (n)$$

$$S^2 = [1864 - (112)^2 / 7] / (7)$$

$$S^2 = [1864 - 1792] / 7$$

$$S^2 = 72 / 7$$

التباين $S^2 = 10.3$

الانحراف المعياري هو الجذر التربيعي للتباين أي:

S.D = 3. 20

Or

بتطبيق هذه الصيغة رقم (2):

$$S^2 = [(\sum X_i^2) - n\bar{X}^2] / (n)$$

$$S^2 = [1864 - 7(16)^2] / (7)$$

$$S^2 = [1864 - 1792] / 7$$

$S^2 = 72 / 7$

التباين $S^2 = 10.3$

الانحراف المعياري هو الجذر التربيعي للتباين أي:

$S.D = 3.20$

<u>حساب الانحراف المعياري من البيانات المبوبة:</u>

سنبين ذلك من خلال المثال التالي وكما ورد في المثال السابق من طريقتين إحداهم باستخدام الوسط الحسابي (الطريقة المطولة) والأخرى بدون الوسط الحسابي (الطريقة المختصرة) وبالتالي سيكون لدينا الصيغ الرياضية الآتية للتباين والانحراف المعياري للبيانات المبوبة:

$i = 1,2,3,\dots,n$

$$S^2 = \frac{\sum f_i X_i^2 - \frac{(\sum f_i X_i)^2}{n}}{n-1} \quad , \quad \sigma = \sqrt{\frac{\sum f_i X_i^2 - \frac{(\sum f_i X_i)^2}{n}}{n-1}} \quad \dots(1)$$

$$S^2 = \frac{\sum f_i (X_i - \bar{X})^2}{\sum f_i} \quad , \quad \sigma = \sqrt{\frac{\sum f_i (X_i - \bar{X})^2}{\sum f_i}} \quad \dots(2)$$

$$S^2 = \frac{\sum f_i X_i^2}{\sum f_i} - \left[\frac{\sum f_i X_i}{\sum f_i}\right]^2 \quad , \quad \sigma = \sqrt{\frac{\sum f_i X_i^2}{\sum f_i} - \left[\frac{\sum f_i X_i}{\sum f_i}\right]^2} \quad \dots(3)$$

$$\sigma = \sqrt{\frac{\sum f_i D^2}{n} - \left(\frac{\sum f_i D_i}{n}\right)^2} \times I$$

حيث I طول الفئة ، D الانحراف عن الوسط الفرضي وهو القيمة التي في مركز الفئة التي تقابل أكبر تكرار ونتائج الانحرافات أعداد صحيحة

<u>مثال:</u>

احسب التشتت باستخدام الانحراف من جدول التوزيع التكراري الآتي والذي يبين درجات 30 طالب في امتحان ما.

الفئات	12 – 14	15 – 17	18 – 20	21 – 23	24 – 26	Total
التكرار	3	8	10	7	2	30

<u>الحل:</u>

نكون الجدول الشامل للبيانات المطلوبة للصيغ الرياضية الخاصة بالانحراف المعياري و باستخدام الصيغة (1) أعلاه:

نجد أن:

$$\sigma = \sqrt{\frac{\sum f_i X_i^2 - \frac{\sum (f_i X_i)^2}{n}}{n-1}}$$

$$\sigma = \sqrt{\frac{10803 - \frac{(561)^2}{30}}{30-1}}$$

$$\sigma = \sqrt{\frac{10803 - 10490.7}{29}}$$

$$\sigma = \sqrt{\frac{312.3}{29}}$$

$$\sigma = \sqrt{10.769}$$

$$\sigma = 3.28$$

$f_i X_i^2$	X_i^2	$f_i X_i$	X_i	f_i	الفئات
507	169	39	13	3	12 – 14
2048	256	128	16	8	15 – 17
3610	361	190	19	10	18 – 20
3388	484	154	22	7	21 – 23
1250	625	50	25	2	24 – 26
10803	1895	561		30	Total

الانحراف المعياري = 3.28

تنبيه:

312.3 ÷ 30 = 10.41 ويكون الانحراف المعياري = 3.23 كما هو مطابق للحلين الآخرين أدناه

تمارين:

(1) أذكر علاقة الانحراف المعياري بالتباين.

قيمة الانحراف المعياري (σ للعينة أو μ للمجتمع) تساوي الجذر التربيعي للتباين.(S^2)

(2) احسب التباين والانحراف المعياري للقيمi3 : ، 5 ، 7 ، 11 ، 13 ، 18 ، 21

- **نحسب الوسط الحسابي لمجموعة القيم بجمعها ومن ثم قسمتها على عددها أي (20 + 18 + 13 + 11 + 7 + 5 + 3)i :**

÷ 7 = 77 ÷ 7 = 11

- **نحسب فروق القيم عن وسطها وهي11 – 3 i : ، 11 – 5 ، 11 – 7 ، 11 – 11 ، 11 – 13 ، 11 – 18 ، 11 – 20**

(الانحرافات عن المتوسط الحسابي.)

i– 8 ، – 6 ، – 4 ، 0 ، 2 ، 7 ، 9

- **نربع الفروق في الخطوة السابقة أي 64 i ، 36 ، 16 ، 0 ، 4 ، 49 ، 81**
- **نجمع مربعات الفروق السابقة أي 250 = 81 + 49 + 4 + 0 + 16 + 36 + 64 i :**
- **نحسب متوسط مربعات الفروق (التباين) 35.7143 = 7 ÷ i250=**
- **نحسب الجذر التربيعي للمتوسط في الخطوة السابقة وهو الانحراف المعياري أي 5.9761 = μ**

(3) أذكر قانون حساب الانحراف المعياري للبيانات المبوبة ثم احسب الانحراف المعياري لدرجات 30 طالب مبينة في الجدول الآتي:

Total	21 – 23	18 – 20	15 – 17	12 – 14	9 – 11	6 – 8	3 – 5	Interval
30	1	3	4	8	6	5	3	Frequency

نكون الجدول التالي:

FX^2	X^2	FX	Mid Interval (X)	Frequency (F)	Interval
48	16	12	4	3	3 – 5
245	49	35	7	5	6 – 8
600	100	60	10	6	9 – 11
1352	169	102	13	8	12 – 14
1024	256	64	16	4	15 – 17
1083	361	57	19	3	18 – 20
484	484	22	22	1	21 – 23
4836		354		30	Total

نطبق القانون الرياضي السابق:

$$\sigma = \sqrt{\frac{\sum FX^2 - \frac{(\sum FX)^2}{n}}{n-1}} = \sqrt{\frac{4836 - \frac{(354)^2}{30}}{30-1}} = \sqrt{\frac{4836 - 4177.2}{29}} = \sqrt{\frac{658.8}{29}} = \sqrt{22.717} = 4.766$$

يلاحظ أن التباين يساوي S^2 =i22.717 كما يلاحظ أنَّ S هو الانحراف المعياري(σ)

***

الارتباط والانحدار الخطي Linear Correlation and Regression

الارتباط :

هو علاقة بين متغيرين يمثل كل متغير ظاهرة معينة فإن تغيرت إحدى الظاهرتين في اتجاه معين فالثانية تتغير في اتجاه الأولى أو في اتجاه معاكس للأولى.

والتغير للظاهرتين في نفس الاتجاه بمعنى الزيادة في الأولى يقابله زيادة في الثانية أو العكس نقص في الأولى يقابله نقص في الثانية فالعلاقة تكون طردية أو متزايدة (موجبة) وإن كان الزيادة في الأولى يقابله نقص في الثانية أو العكس النقص في الظاهرة الأولى يقابله زيادة في الثانية فنقول أن الارتباط عكسي أو متناقص (سالباً).

الارتباط يكون تاماً بين المتغيرين إذا عرفنا قيمة أحد المتغيرين نعرف قيمة المتغير الآخر تماماً.

تقاس قوة الارتباط بين متغيرين بمعامل الارتباط ولاستقصاء تأثير أحد المتغيرين على الآخر فنوجد علاقة جبرية تعرف بمعادلة الانحدار (سندرسها لاحقاً) وهي معادلة خط مستقيم (علاقة خطية) بين المتغيرين x، y مثلاً بالشكل $y = a x + b$ حيث a , b ثابتان يعبر a عن الاتجاه (الميل) و b الجزء المقطوع من محور الصادات (1 –) وإن كانت a موجبة كانت العلاقة طردية وإن كانت سالبة فإنَّ العلاقة عكسية والارتباط هنا تاماً لأن النقاط (x , y) يمر بها خط مستقيم واحد فمثلاً العلاقة $y = 2 x – 1$ تحدد قيم y بمعرفة قيم x وبتمثيلها بيانياً نجد أن:

1	2	3	X
1	3	5	Y

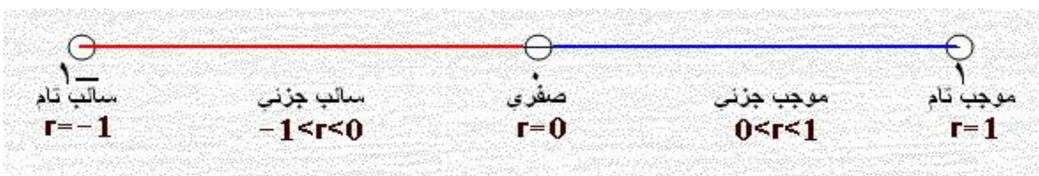
يقاس الارتباط بين متغيرين بمقياس كمي يعرف بمعامل الارتباط Correlation Coefficient

معامل الارتباط قد يساوي الصفر (منعدم) بين المتغيرين لعدم وجود علاقة بينهم فمعرفة قيمة أحدهم لا يعني معرفة قيمة الآخر فالمتغيرين الطول ودرجة الامتحان مثلاً وإن تمثيلها بيانياً يظهر مجموعة من النقاط المبعثرة.

إن معرفة العلاقة بين المتغيرات من حيث النوع والقوة والاتجاه يعتبر هدفاً مهماً من أهداف البحث العلمي.

العلاقة المبينة في الجدول أعلاه طردية حيث تزداد قيم y بزيادة قيم x وهي علاقة خطية (Linear).

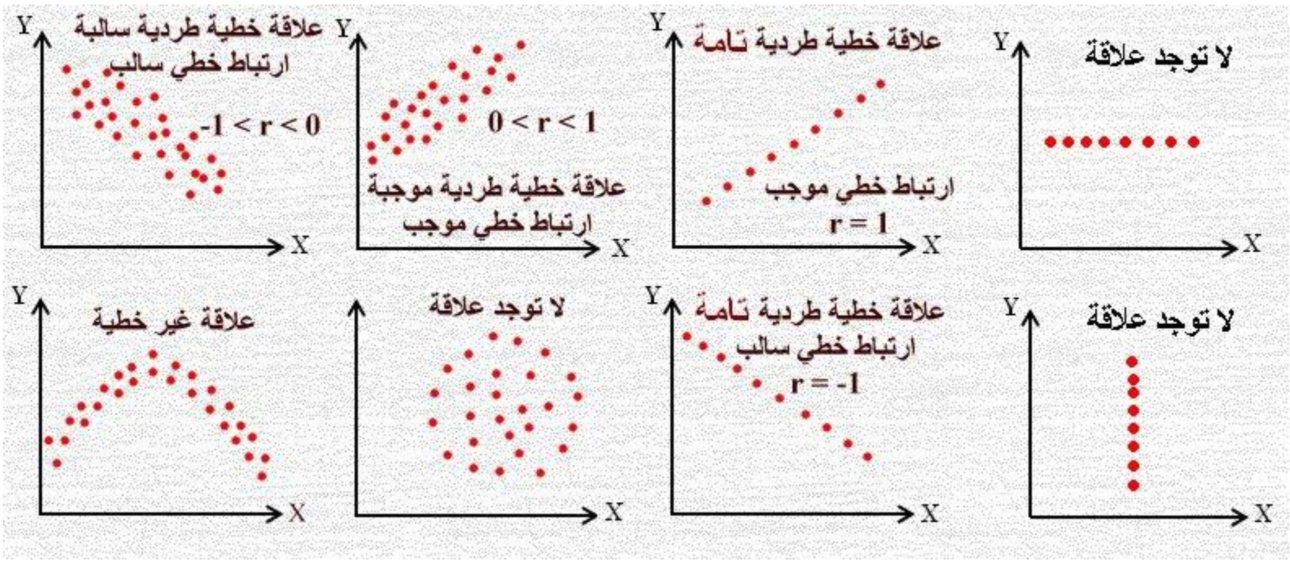
وتكون درجة العلاقة قوية عندما تكون قيمة معامل الارتباط $\pm$ 1 وتكون العلاقة هنا تامة في حين تضعف درجة العلاقة كلما اقتربت القيمة من الصفر، ونوضح قيمة معامل الارتباط بالشكل الآتي:



الشكل ألانتشاري

إذا أخذنا (س ، ص) كقيم متناظرة لمتغيرين وقمنا بتمثيلها في مستوى الإحداثيات وحصلنا على الأشكال التالية:

فلكل قيمة للمتغير x توجد قيمة تقابلها للمتغير y وإن الأزواج المرتبة (x , y) تكون مجتمع ذو بعدين ويعرف الزوج المرتب بمتغير عشوائي ذو بعدين وللمجتمع ذو البعدين نطرح سؤالاً عل توجد علاقة بين المتغيرين؟ وإن وجدت فكيف نعبر عنها بمعادلة؟ فوجود المعادلة يعني معرفة أحد المتغيرين من معرفة الآخر فالمتغير الأول يعرف بالمتغير المستقل في حين الآخر يعرف بالمتغير التابع، الشكل المرفق هنا يعرف بلوحة الانتشار وكل نقطة هنا تمثل زوج مرتب بالصورة (x , y).



معامل الارتباط الخطي Linear Correlation / بيرسون

بملاحظة المتغير العشوائي ذي البعدين (x , y) بوجود ارتباط أو علاقة بين x , y فإن الهدف من دراسة الارتباط هو قياس قوة الارتباط الخطي بين المتغيرين في حين معامل الارتباط الخطي (linear Coefficient) مقياس لقوة العلاقة الخطية بين x ، y ويقيس مدى تغير y حال زيادة قيمة x فهل y تزداد بزيادة x (ارتباط موجب) أو تنقص بزيادتها (ارتباط سالب) أو لا تتأثر بزيادة x (لا يوجد ارتباط).

معامل الارتباط لمجموعة n من الأزواج المرتبة (x_1 ، y_1) ، (x_2 ، y_2) ، ... ، (x_n ، y_n) هو:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)S_x S_y} \quad \text{Or}$$

$$r = \frac{n(\sum X_i Y_i) - (\sum X_i)(\sum Y_i)}{\sqrt{n(\sum X_i^2) - (\sum X_i)^2}\,\sqrt{n(\sum Y_i^2) - (\sum Y_i)^2}} \quad \text{Or}$$

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sqrt{\sum X_i^2 - n\bar{X}^2}\,\sqrt{\sum Y_i^2 - n\bar{Y}^2}} \quad \text{Or}$$

مع ملاحظة: 1) $\bar{X}$ الوسط الحسابي للبيانات x_1 ، x_2 ، ... ، x_n و $\bar{Y}$ الوسط الحسابي للبيانات y_1 ، y_2 ، ... ، y_n

2) S_x الانحراف المعياري للبيانات x_1 ، x_2 ، ... ، x_n و S_y الانحراف المعياري للبيانات y_1 ، y_2 ، ... ، y_n

3) r معامل الارتباط يعرف بمعامل ارتباط بيرسون (التتابعي أو العزومي) للارتباط (نسبة للعالم كارل بيرسون).

4) يشترط عند حساب معامل الارتباط لبيرسون أن يكون التوزيع لكلا المتغيرين اعتدالي وأن تكون العينة عشوائية وقيم الفرد لا تعتمد على قيم فرد آخر (استقلالية أفراد العينة).

مثال1:

أوجد معامل الارتباط بين دخل تسعة أسر (X) والإنفاق (Y) اليومي بالدينار والمبينة في الجدول الآتي:

X	6	8	7	14	11	12	8	9	10
Y	4	8	6	10	9	11	8	7	8

عوامل التحكم في معامل ارتباط بيرسون:

- أن تقع نقاط الأزواج (x , y) على خط مستقيم أو تكون قريبة جداً منه حتى تحقق صفة أن العلاقة خطية (y = ax + b) ويمكن ملاحظة ذلك من شكل الانتشار. إن لم تكن العلاقة خطية فستخدم معامل آخر.
- مقدار التباين فالعلاقة طردية بين الزيادة في التباين ومعامل الارتباط.
- دقة معامل الارتباط تتأثر بحجم العينة.
- شكل التوزيع وتماثله للمتغيرين يزيد من قيمة معامل الارتباط فإن كان شكلا التوزيع متماثلين فيكون r = ± 1 وإن كانا الالتواء في نفس الاتجاه كان r = 1 وإن كان الالتواء في اتجاهين متضادين (احدهم التواءه موجب والآخر سالب) كان r = – 1

من خصائص معامل الارتباط عدم اعتماده على القيم نفسها بل على تباعدها عن بعضها، لا تتغير قيمة معامل الارتباط بالعمليات الحسابية الأربع الجمع والطرح والقسمة والضرب مع عدد ثابت بالنسبة لقيم x , y .

معامل ارتباط الرتب: (Correlation Conefficient Rank)

هذا المعامل يعرف بمعامل ارتباط سبيرمان (Spearman) أو معامل ارتباط الرتب (رتب القيم الأصلية وليس القيم) ولذا تختلف قيمته عن قيمة معامل بيرسون (للقيم الأصلية وليس لرتبها) وهو أقل دقة من معامل ارتباط بيرسون ويتعامل مع البيانات الرقمية وغير الرقمية للترتيب مثل جيد، جيد جدا، ... ويرمز له بالرمز r_s وهو ضمن الإحصاءات غير المعلمية ذات التوزيع الحر وقيمته موجبة أقل أو تساوي الواحد الصحيح وتحسب قيمته من الصيغة الرياضية علماً بأن:

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

حيث d الفرق بين رتبه حسب المتغير الأول x ورتبه حسب المتغير الثاني y (الفرق بين رتب القيم لكل زوج من البيانات) وفي حالة التساوي يأخذ المتوسط الحسابي (فإذا كانت لقيمتين متساويتين الرتبتين 7 ، 8 فيأخذ متوسط 7 ، 8 وتصبح الرتب لكل

منها 7.5 بدل عن 7 ، 8) ، n عدد الأزواج للقيم فإذا كان لدينا مجموعة من الأفراد وجرى ترتيبهم حسب صفتين لكل فرد من المجموعة x , y فإن $y_i - x_i = d_i$.

مثال:

تقدم عشرة طلاب لامتحان المرحلة الثانوية وكانت معدلات نتائجهم حسب الصف والمدرسة كالتالي والمطلوب حساب معامل سبيرمان للارتباط.

معدل الطالب في الصف (X)	73	80	70	66	89	71	65	88	92	74
مدل الطالب في المدرسة (Y)	69	78	66	70	92	64	55	90	88	72

الحل:

نكون جدول نبين فيه رتب كل من X (المعدل في الصف) و X (المعدل في المدرسة) والفرق d ومربع الفرق d^2 كالتالي:

X	Y	Rank X	Rank Y	d	d^2
73	69	6	7	– 1	1
80	78	4	4	0	0
70	66	8	8	0	0
66	70	9	6	3	9
89	92	2	1	1	1
71	64	7	9	– 2	4
65	55	10	10	0	0
88	90	3	2	1	1
92	88	1	3	– 2	4
74	72	5	5	0	0

بتطبيق القانون أعلاه:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

$$r_s = 1 - \frac{6 \times 20}{10(100 - 1)}$$

$$r_s = 1 - \frac{120}{990}$$

$$r_s = 1 - 0.121$$

$$r_s = 0.879$$

دلالة معامل الارتباط:

اختبار مدى المعنوية r_s (القيمة متوسطة وليست صفر أو ±1) وعندما تكون حجم العينة أكبر من وأقل من 30 (صغيرة) نقارنها مع المحسوبة من الجدول عند $\alpha/2$ وعندما تكون حجم العينة أكبر أو يساوي 30 فنوجد قيمة Z ونقارنها مع الجدولية حيث قيمة Z = قيمة معامل ارتباط الرتب مضروباً في الجذر التربيعي للعدد $n - 1$.

باعتبار أن المجتمع ذا البعدين X, Y والمأخوذ منه العينة من الأزواج المرتبة وبفرض أن ρ معامل ارتباط المجتمع فيكون r تقديراً للمعامل ρ. ولا بد من افتراض أن $\rho = 0$ لنحصل على اقتران احتمال(r) حسب النظرية:

إن جميع العينات ذات حجم n والممكنة مأخوذة من مجتمع ذي بعدين ويخضع للتوزيع المعتدل ومعامل ارتباطه $\rho = 0$ ، وأن r يعبر عن معاملات ارتباطات تلك العينات فإن:

$$t = \frac{r}{\sqrt{(1 - r^2)(n - 2)}}$$

مثال آخر: نفس المثال السابق مع البيانات التالية:

معدل الطالب في الصف (X)	73	80	70	66	88	71	65	88	92	74
مدل الطالب في المدرسة (Y)	64	78	64	70	92	64	55	90	88	72

الحل:

معدل الطالب في الصف (X)	73	80	70	66	88	71	65	88	92	74
مدل الطالب في المدرسة (Y)	64	78	64	70	92	64	55	90	88	72

نكون جدول نبين فيه رتب كل منX المعدل في الصف و X (المعدل في المدرسة) والفرق d ومربع الفرق d^2 كالتالي:

X	Y	Rank X	Rank Y	d	d^2
73	64	6	8	– 2	4
80	78	4	4	0	0
70	64	8	8	0	0
66	70	9	6	3	9
88	92	2.5	1	1.5	2.25
71	64	7	8	– 1	1
65	55	10	10	0	0
88	90	2.5	2	0.5	0.25
92	88	1	3	– 2	4

74	**72**	**5**	**5**	**0**	**0**
					Total = 20.5

بتطبيق القانون:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

$$r_s = 1 - \frac{6 \times 20.5}{10(100 - 1)}$$

$$r_s = 1 - \frac{123}{990}$$

$$r_s = 1 - 0.124$$

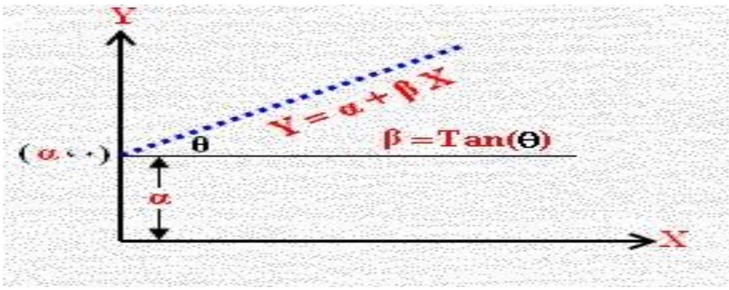
$$r_s = 0.876$$

***

الانحدار الخطي البسيط (Regression)

مقدمــة:

تهدف دراسة الانحدار التنبوء بقيمة متغير (Y) بمعرفة متغير آخر (X) ويعرف المتغير الأول بالمتغير التابع (dependent) ويرمز له Y ويقاس دون خطأ في حين يعرف المتغير الآخر بالمتغير المستقل (Independent) ويرمز له X فإذا أعطينا قيمة ما (أي قيمة تنتمي لمجموعة الأعداد الحقيقية) للمتغير X في المعادلة $Y = \beta + \alpha X$ فنحصل علي قيمة مناظرة للمتغير Y فهنا قيمة Y تتحدد بمعرفة قيمة X فلذا المتغير X عرف بالمتغير المستقيل في حين Y تتعين قيمتها تبعاً لقيمة X لذا عرفت Y بالمتغير التابع (أي تبعاً لقيمة X)، كما أن الانحدار هنا بسيط لوجود متغيرين فقط تابع ومستقل، وسنتحدث لاحقاً عن الانحدار المتعدد بوجود متغير تابع واحد فقط مع وجود متغيرين مستقلين أو أكثر، وعند ذكر كلمة الخط نعني بها خط الانحدار.



والانحدار يعني بالبحث عن هذه المعادلة أو العلاقة بين المتغيرين Xالمستقل ، Yالتابع أو المعتمد كما أن المعادلة $Y = \alpha + \beta X$ تحوي α , β وهما قيمتان ثابتتان حيث β تبين ميل الخط المستقيم ($Y = \alpha + \beta X$) الزاوية التي يصنعها المستقيم مع الاتجاه الموجب لمحور السينات (وإن إشارة α تعني:

- أن تكون β موجبة فإن علاقة Y بالمتغير المستقل X علاقة طردية موجبة (تزداد قيم Y بزيادة قيم X المناظرة لها أو العكس أي تنقص بنقصانها).
- أن تكون β موجبة فإن الخط $Y = \alpha + \beta X$ يصنع زاوية حادة مع محور السينات الموجب كما مبين بالشكل.
- أن تكون β = صفراً فتنعدم العلاقة الخطيبة (لا توجد علاقة) وأن قيمة Y ثابتة.($y = \alpha$)
- أن تكون $\beta = \infty$ فتنعدم العلاقة الخطيبة (لا توجد علاقة) كما في الشكل.
- تعرف β بميل الانحدار
- أن تكون β سالبة فإن العلاقة عكسية سالبة (تزداد قيم Y بنقص قيم X المناظرة لها أو العكس.(
- في حين أن α تبين قيمة الجزء المقطوع من محور الصادات (الرأسي) بالمستقيم $Y = \alpha + \beta X$
- الخط $Y = \alpha + \beta X$ يمر بالنقطة (0 ، α) أو أن الخط $Y = \alpha + \beta X$ يمر بالزوج (0 ، α)
- تعرف α بثابت الانحدار
- عند استخدام عينه n من الأزواج من مجتمع ذو بعدين فنكتب العلاقة بحروف صغيرة $Y = a + b X$ ويكون:
- a تقديراً لـ α و b تقديراً لـ β.

وتسمى المعادلة $Y = a + b X$ بمعادلة انحدار Y على X في حين $X = a + b Y$ معادلة X على X سواء للعينة أو المجتمع ولمعرفة المعادلة يجب معرفة كل من a , b ويجدر هنا القول إن اهتمامنا يكون على العلاقة نفسها ببين Y , X وليس على سبب وجود العلاقة أو الظروف المحيطة بها فهناك علاقة سببية بين كمية السماد وكمية الناتج الزراعي للقمح مثلاً ولكننا لا يمكن أن نجزم بوجود علاقة سببية لعدد المساكن وميزانية الدولة.

تقرأ معادلة خط الانحدار بأن Y دالة في X .

تمرين

الجدول الآتي يبين إنتاج محصول الذرة Y من المساحة المزروعة به X. أوجد معادلة انحدار Y على X

المنطقة	X المساحة المزروعة بالهكتار	Y إنتاج الذرة بالآف الكيلوجرام
1	50	140
2	200	500
3	110	400
4	80	300
5	120	356
6	74.5	240.5
7	88.9	200.6
8	5.7	33.5
9	11	69.8
10	3.2	18.7

الحل:

نكون جدول جديد للحصول على البيانات اللازمة لحساب قيمتي a , b

المنطقة	X	Y	X^2	XY
1	50	140	2500	7000
2	200	500	40000	100000
3	110	400	12100	44000
4	80	300	6400	24000
5	120	356	14400	42720

17917.25	5550.25	240.5	74.5	**6**
17833.34	7903.21	200.6	88.9	**7**
190.95	32.49	33.5	5.7	**8**
767.8	121	69.8	11	**9**
59.84	10.24	18.7	3.2	**10**
254489.2	**89017.19**	**2259.1**	**743.3**	**Total**

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}$$

$$= \frac{10 \times 254489.2 - 743.3 \times 2259.1}{10 \times 89017.19 - (743.3)^2}$$

$$= \frac{865702.97}{337.677}$$

$$= 2.5637$$

$$a = \frac{\sum Y - b\sum X}{n}$$

$$= \frac{2259.1 - 2.564 \times 743.3}{10}$$

$$= 35.35$$

المعادلة المطلوبة: Y = 2.564 X+35.35

***

تمرين 2

الجدول يبين بيانات عن متوسط سعر البترول ومعدلات النمو الاقتصادي في دولة الجزائر خلال السنوات من 1980i إلى 1989iم والمطلوب إيجاد معادلة خط الانحدار للبيانات أي $Y = a + bX$ ومدى جودتها و التنبوء بالنمو الاقتصادي عندما يكون سعر برميل البترول 20$ ثم عندما يصبح سعر البرميل 40$

السنة	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
سعر البترول دولار	35,19	39,54	35,50	30,60	29,67	29,11	14,18	18,72	16,26	18,41
معدل النمو الاقتصادي	0,9	3,6	4,0	5,6	4,1	5,2	1	-1,1	-1,8	-2,9

الحـل:

نكون جدول جديد للحصول على البيانات اللازمة لحساب قيمتي a , b ووجود Y^2 لحساب معامل الارتباط لمعرفة معامل التحديد للتبوء عن النمو الاقتصادي للسنة عند السعر الجديد للبرميل 10i أو 40i.

XY	Y^2	X^2	Y	X	Year
31.671	0.81	1238.336	0.9	35,19	1980
142.344	12.96	1563.412	3.6	39,54	1981
142	16	1260.25	4.0	35,50	1982
171.36	31.36	936.36	5.6	30,60	1983
121.647	16.81	880.3089	4.1	29,67	1984
151.372	27.04	847.3921	5.2	29,11	1985
14.18	1.0	201.0724	1.0	14,18	1986
– 20.592	1.21	350.4384	–1.1	18,72	1987
– 29.268	3.24	264.3876	–1.8	16.26	1988
– 53.389	8.41	338.928	– 2.9	18,41	1989

671.325	**118.84**	**7880.885**	**18.6**	**267.18**	**Total**

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}$$

$$b = \frac{10 \times 671.325 - 267.18 \times 18.6}{10 \times 7880.885 - (267.18)^2}$$

$$b = \frac{1743.702}{7423.698}$$

b = 0.235

$$a = \frac{\sum Y - b\sum X}{n}$$

$$a = \frac{18.6 - 0.2349 \times 267.18}{10}$$

a = – 4.416

The Equation is: Y = – 4.416 + 0.235X

نوجد الآن معامل ارتباط بيرسون

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{n\sum X^2 - (\sum X)^2}\sqrt{n\sum Y^2 - (\sum Y)^2}}$$

$$= \frac{10 \times 671.325 - 267.18 \times 18.6}{\sqrt{10 \times 7880.885 - (267.18)^2}\sqrt{10 \times 118.84 - (18.6)^2}}$$

$$= \frac{1743.702}{\sqrt{7423.698}\sqrt{842.44}}$$

$$= \frac{1743.702}{\sqrt{7423.698}\sqrt{842.44}}$$

$$= 0.697$$

هذه القيمة تعني وجود علاقة قوية نسبياً بين متوسط السعر ومعدل النمو ومن حيث القيمة موجبة فالعلاقة طردية.

نوجد معامل التحديد (مربع معامل بيرسون) ويساوي $(0.697)^2 = 0.49$ وهذا يعني أن 49% من معدل تغير النمو الاقتصادي في الجزائر يعتمد على سعر برميل البترول في حين 51% $(1 - 0.49)$ ترجع لعوامل أخرى غير سعر البرميل ولم تدخل تلك العوامل ضمن معادلة خط الانحدار التي حصلنا عليها.

إن القيم الناتجة سواء لمعامل الارتباط الذي يجب أن يكو 9% أو أكثر وكذلك معامل التحديد 80% أو أكثر وهو ما يقودنا للقول بأن معادلة الانحدار ليست ذات جودة عالية مما يستدعي إدخال عوامل أخرى بجانب سعر برميل البترول.

نعوض في المعادلة التي حصلنا عليها $Y = -4.416 + 0.235X$ عن القيم 20 ، 40 نجد أن:

التنبؤ بمعدل النمو الاقتصادي عند السعر \$20 : $Y = -4.416 + 0.235 \times 20 = 0.284$ وهو معدل النمو الاقتصادي لتلك السنة التي يكون فيها سعر البرميل 20%

التنبؤ بمعدل النمو الاقتصادي عند السعر \$40 : $Y = -4.416 + 0.235 \times 40 = 4.984$ وهو معدل النمو الاقتصادي لتلك السنة التي يكون فيها سعر البرميل 40%

***

ملاحظة :

الاحتمالات في ملف منفصل لاحقا....

مع تحيات

مدونة عيون المعرفة

http://knoweyes.blogspot.com