إعداد / دكتورة عوضية محمد إسماعيل

# مقدمة في مبادئ الإحصاء

**مقدمة :**

يعتبر علم الإحصاء في الوقت الحالي واحد من أهم العلوم الحديثة التي تلعب دوراً حيوياً في كثير من العلوم والدراسات المختلفة. كما يعتبر الإحصاء من أقدم العلوم حيث ظهر مع حاجة الإنسان الأولى للتعامل مع القيم والأعداد لتسيير الحياة اليومية. فالتاجر يسعى إلى حصر وحفظ البيانات المتعلقة بتجارته والمزارع يقوم دوماً بإحصاء الإنتاج والمعلومات الأخرى المتعلقة كعدد الأشجار وأوقات الحصاد والبذر وغيرها من المعلومات والبيانات ذات العلاقة.

ومع التطور الهائل في العلوم كافة في أواخر القرن العشرين تطور علم الإحصاء ليستفيد من تقنيات الحاسب الآلي بشكل يجعله العلم الأكثر تداخلاً مع العلوم الأخرى المختلفة، حيث أصبح يستخدم علم الإحصاء في العلوم التجارية وعلوم الطب والهندسة والأدب وجميع العلوم الأخرى دون استثناء. كما ساهم عصر المعلومات والانفتاح العالمي الحديث في إبراز أهمية تفعيل عملية التعامل مع البيانات بأسلوب يضمن السيطرة عليها وقراءتها، مما كان له الأثر الواضح على تطور علم الإحصاء كونه العلم الذي يحقق تلك الغاية. كما اتجهت كثير من العلوم والدراسات الأكاديمية والبحثية لاسيما التطبيقية إلى استخدام علم الإحصاء من خلال حصر بيانات مشكلة البحث والتعامل معها إحصائيا للوصول إلى فهم أفضل وحلول موضوعية.

يتم الاستفادة من علم الإحصاء في مجالات متنوعة تشمل ميادين عديدة كالصناعة والزراعة والطب والبحوث وغيرها من مجالات الإدارة والأعمال والعلم بشكل عام. ويتم تطبيق الأساليب الإحصائية في الجوانب المختلفة للصناعة كمراقبة جودة المنتجات وتسويقها والتخزين وتشغيل خطوط الإنتاج. كما يتم استخدام علم الإحصاء في المجال الطبي لدراسة الأمراض المختلفة والبحث في مسبباتها وطرق علاجها. وفي مجال الزراعة يتم بحث إحصاءات الثروة الحيوانية والنباتية ودراسة العلاقة بين أنواع الأسمدة والأساليب الزراعية المختلفة وزيادة الإنتاج. كما يتم دراسة السكان والمساكن من خلال الإحصاء الديموجرافي، حيث يتم التركيز على القوى العاملة وخصائصها والأجور والدخل والإنفاق. أما في مجال الأعمال والتجارة فان الإحصاء يلعب دوراً حيوياً يتمثل في دراسة السوق واتجاهات المستهلكين ودراسات الأسعار وكميات لإنتاج.

# الوحدة الأولى

**تعريف علم الإحصاء :**

- الإحصاء هو علم من علوم الرياضيات المهمة، يعتمد الإحصاء على مجموعة من البيانات المتوفّرة التي يقوم بتلخيصها وتمثيلها واستنتاج نتائجها بغرض التنبؤ واستنتاج والقرارات اللازمة .
- تشر كلمة إلى جميع المؤشرات المأخوذة من العينة (والعينة جزء من المجتمع محل الدراسة )
- تشر كلمة إحصاء إلى المعلومات الكمية والبيانات العددية المتعلقة بموضوع معين كتعداد السكان وبيانات القوة العاملة والإنتاج الزراعي وغيرها من الإحصاءات .

و الإحصاء يتكون من الآتي : المتغيرات والأساليب الإحصائية والبيانات.

**1- المتغيرات :**

المتغير : هو الخاصية أو الظاهرة التي تأخذ قيما مختلفة إما عددية أو نوعية

وتنقسم المتغيرات إلى نوعين :

1- **المتغير النوعي :** وهو المتغير الذي يتم التعبير عنه بصفة او تصنيف , ومن الأمثلة على المتغيرات النوعية تحديد جنس الطالب ( ذكر أو أنثى ) , مثلا المستوى الثقافي ( ممتاز , جيد جدا, جيد, مقبول )

**وتقسم المتغيرات النوعية إلى الأنواع التالية :**

أ- متغيرات اسمية حيث يتم تعريفها باسم او صفة او فئة .

ب- متغيرات ترتيبية حيث يمكن ترتيب تصنيفات هذا المتغير تصاعديا او تنازليا ( مهندسون , فنيون , محاسبون ، معلمون ) .

2- **المتغير الكمي :** وهو المتغير الذي يتم التعبير عنه بواسطة الأعداد ويقاس بوحدات معينة .

**وتقسم إلى الآتي :**

أ- متغيرات منفصلة : وهي المتغيرات التي تأخذ قيما قابلة للعد وتكون محدودة أو لانهائية ولكنها معدودة وتكون القيم أعدادا صحيحة ( عدد أفراد الأسرة , عدد الزوار إلى منطقة معينة ) .

ب- متغيرات متصلة وهي التي تكون قيمها الممكنة عبارة عن فترة أو اتحاد فترتين أو أكثر وليس هناك قفزات بين قيم المتغير المتصل ومن الأمثلة على المتغير المتصل ( درجة الحرارة , الوزن , الطول , الزمن , العمر) .

**2- الأساليب الإحصائية :** وهي نوعان **أساليب وصفية و أساليب استدلالية**

**الأساليب وصفية :**

## الإحصاء الوصفي :Descriptive statistics

هي عبارة عن مجموعة الأساليب، والطرق التي يتُم استخدامها في جمع ما يلزم من بيانات، وتنظيمها، وتلخيصها، ثم عرضها، واستكشاف كل خصائصها الأساسية، ثم تلخيص هذه البيانات في صورة مؤشرات رقمية.

**أ- مثل التكرارات و النسب المئوية و مقاييس النزعة المركزية و ومقاييس التشتت .**

**ب- العلاقة بين المتغيرات بين متغير تابع واحد ومتغير مستقل واحد ( الارتباط الخطي البسيط والانحدار الخطي البسيط )**

**ت- العلاقة بين عدة المتغيرات مستقلة و متغير تابع واحد أو عدة متغيرات تابعة و متغير مستقل واحد ( الارتباط الخطي المتعدد والانحدار الخطي المتعدد ) .**

**الأساليب الاستدلالية :**

## الإحصاء الاستدلالي: Inferential Statistics

1- هو عبارة عن مجموعة الأساليب، والطرق التي يتم أُستخدامها لتعميم نتائج العينة على كل خصائص المجتمع n -1 الذي تم سحب العينة منه

2- كما يستخدم لقياس العلاقات الموجودة بين الخصائص المتعددة للمجتمع، وكذلك التنبؤ بما سيحدث لهذه الخصائص مستقبلا. مثل التقدير ألمجالي ( فترات الثقة ) واختبارات الفروض الإحصائية .

3- البيانات :

فالبيانات هي قياسات أو مشاهدات مسجلة لظواهر محددة . إما المعلومات فهي مجموعة من الاستنتاجات والحقائق التي تستخلص من البيانات بعد معالجتها إحصائيا .

فالمعلومات جملة استنتاجات صيغت بطريقة مناسبة لاتخاذ القرار , تستند إلى جملة من البيانات بعد تحليلها .

**وظائف الإحصاء :**

جمع البيانات – وصف البيانات – الاستقراء – صنع القرار .

**جمع البيانات الإحصائية**

إنّ أهمّ مرحلة في العملية الإحصائية هي مرحلة جمع البيانات، إنّ أيّ خطأ في عملية جمع البيانات سينتج عنه إحصاء خاطئ، يجب تتبع النقاط الآتية لجمع البيانات:

- **مصادر البيانات**
  يجب أن يكون المصدر صحيح ودقيق، وهناك نوعان من مصادر البيانات:

**المصادر الأولية**: وهي البيانات التي يجمعها الباحث بنفسه من عينات البحث، كإجراء بحث عن الأسرة وجمع المعلومات من رب الأسرة، وهو أكثر دقة من المصدر الآخر ولكنها تستهلك الكثير من الوقت والجهد والمال.

- **المصادر الثانوية**: وفي هذه العملية يتم الحصول على البيانات بشكل غير مباشر من جهات معينة أو أجهزة كاستخدام النشرات والدراسات، وهو مصدر غير دقيق تماماً ولكنه يوفر الوقت والجهد والمال على عكس المصدر الأخر.
- أسلوب جمع البيانات
  تعتمد هذه العملية على الهدف من البحث وحجم عدد الأشخاص المشمولين بالبحث،
- ومن أساليب جمع البيانات:

**أولا المقابلات الشخصية:**

هو أسلوب يعتمد كلياً على التفاعل اللفظي بين الأفراد، حيث يتمركز حوارهم حول موضوع معين يدأب الباحث إلى استثارة رأي الآخر للحصول على المعلومات أو التغيرات التي تطرأ على المبحث.

**مزايا المقابلة :**

1- يمتاز بأنّه وسيلة مؤكدة للمعلومات .
2- أنها أفضل وسيلة لاختيار وتقويم الصفات الشخصية .
3- أنها ذات فائدة كبرى في تشخيص ومعالجة المشاكل الإنسانية وخاصة العاطفية فيها .
4- أنها ذات فائدة كبرى في الاستشارات .

5- أنها تزود الباحث بمعلومات إضافية كتلك المعلومات التى حصل عليها بواسطة وسائل أخرى من وسائل جمع المعلومات .
6- قد يستخدمها الباحث مع وسيلة الملاحظة للتأكد من صحة البيانات التى حصل عليها وسيلة للمراسلة .
7- تكاد تكون المقابلة الوسيلة الوحيدة لجمع البيانات فى المجتمعات الأمية .
نسبة المردود عالي عند مقارنتها بالاستبيان .

**عيوب المقابلة :**

**1- أن نجاحها يعتمد إلى حد كبير على رغبة المستجيب فى التعاون وإعطاء معلومات موثوق فيها ودقيقة .**
**2- أنها تتأثر بالحالة النفسية والعوامل الأخرى التى تؤثر على الشخص الذى يجرى المقابلة أو على المستجيب أو عليهما معا .**
**3- تعاب المقابلة بأنها متأثرة بالحالة النفسية للمستجيب .**

**ثانيا الملاحظة:**

هو **الأسلوب المعتمد كلياً على مدى انتباه الباحث والتفاته لظاهرة أو شيء ما، أما فيما يتعلق بالملاحظة العلمية فإنّها باعتبار انتباه كمنهج لظاهرة ما أو حادثة من خلال مراقبتها، ويهدف الباحث من الملاحظة في الظواهر العلمية هو تفسيرها أو اكتشاف أسبابها، والخروج بمجموعة من القوانين التي تحكمها.**

**مزايا الملاحظة :**

**1- أنها أفضل طريقة مباشرة لدراسة عدد الظواهر .**
**2- إذ أن هناك عدة جوانب للتصرفات الإنسانية لا يمكن دراستها إلا بهذه الوسيلة .**
**3- أنها لا تتطلب جهودا كبيرة مقارنة مع الطرق البديلة .**
**4- أنها تمكن الباحث من جمع البيانات تحت ظروف سلوكية مألوفة .**
**5- أنها تمكن الباحث من جمع حقائق عن السلوك في نفس وقت حصولها .**
**6- أنها لا تعتمد كثيرا على الاستنتاجات .**
**7- أنها تسمح بالحصول على معلومات وبيانات قد ليكون الأفراد قد فكروا بها .**

**أما عيوب الملاحظة : :**

1- **قد يتعمد الأفراد موضوع البحث إعطاء الباحث انطباع جيد أو غير ذلك .**
**2- من الصعب توقع حادثة عفوية بشكل مسبق لكي يكون الباحث حاضراً في نفس الوقت .**
**3- قد تعيق في بعض الحالات عوامل غير منظورة كالتقلبات في الطقس .**

**ثالثا الرجوع إلى المصادر المكتوبة:**

وتتمثل باللجوء إلى الكتب والمخطوطات والأطروحات والدوريات والمواقع الإلكترونية التي تقدم معلومات ذات علاقة بمادة البحث.

**رابعا الاستبيان:**

وسيلة فعالة في استقطاب البيانات وجمعها بواسطة الاستمارة التي يتم تعبئتها بواسطة المستجيب، وتمتاز بإمكانية الحصول على المعلومات من عدد هائل من الأشخاص المتباعدين جغرافياً خلال فترة زمنية وجيزة.

**الهدف** الأساسي من الاستبيان هو التحقق من صحة أو خطأ الفرض الذي افترضه الباحث لمشكلة الدراسة .
**أولا أنواع الاستبيان :**

**1- الاستبيان المفتوح الأسئلة:**
**وهو الذي يترك فيه الباحث فراغا للمبحوث بعد كل سؤال حتى يكتب رأيه وما يناسبه وفقا للإرشادات التي يحددها الباحث ويتم الاعتماد على هذا النوع من الأسئلة إذا أراد الباحث الحصول على بيانات يصعب تصنيفها أو تحتاج في تصنيفها إلى مجموعة كبيرة من الفئات .**

**وفى هذا النوع يواجه الباحث بعض الصعوبات في النتائج من حيث الترتيب والتلخيص .**
**و ومن إيجابيات هذا النوع من الأسئلة تكمن في عدم تدخل الباحث في إجابات المبحوث .**
**2- الاستبيان المغلق الأسئلة :**
**وهو الذي تتطلب أسئلته إجابة واحدة لكل سؤال مثلا بنعم أو لا أو موافق وغير موافق وأوافق تماماً ..... الخ .**
**ويحتاج هذا النوع إلى الدقة العالية في التصميم لأنه إذا كانت الإجابة هي اختيار نعم أو لا فيجب أن لا يكون هناك احتمال آخر لأن هناك استجابات لها احتمالات أخرى وهى تسبب حيرة للمبحوث وقد يسهم ذلك في عدم الوصول إلى الهدف .**
**وللخروج من المشكلة يترك الباحث خيار ثالث ويكتب عليه أخرى ومن ايجابيات هذا النوع من الأسئلة أنه سهل التصنيف والجدولة مما يسهل تحليلها. وأيضا عدم حاجتها إلى وقت طويل أو جهد أو تفكير من جانب المبحوث .**
**3 - الاستبيان المختلط:**
وهو أكثر الأنواع انتشارا حيث يلجأ إليه الباحثون من أجل تلافى عيوب النوعين السابقين والجمع بين مميزاتهم . الجمع بين الأسئلة المفتوحة والمغلقة.

**ثانياً : خطوات استخدام الاستبيان**
**1- إقناع الباحث بأن الاستبيان هو الوسيلة المناسبة لدراسته .**
**2- تحديد نوع المعلومات المراد الحصول عليها وتحديد نوع المبحوثين وهذا يأتى من نوع المشكلة وطبيعة الدراسة .**
**3- تقسيم المشكلة إلى عناصرها الرئيسية والفرعية حتى تكون لديه صورة واضحة عن نوع الأسئلة المراد صياغتها .**

4- **صياغة الأسئلة بصورة واضحة لوضوح الرؤية من تحليل جوانب المشكلة وتقسيمها إلى مكوناتها الرئيسية والفرعية وتعتبر هذه الخطوة من أهم الخطوات .**

**يجب على الباحث مراعاة عدم الغموض وسهولة الألفاظ المختارة وأن تكون الكلمات ذات معنى محدد وأيضاً تكون متماسكة وأن تكون التعليمات واضحة**

1- **وأن يبتعد عن التكرار والأسئلة الايجابية**
2- **الابتعاد عن الأسئلة التي تحتاج إلى جهد وتفكير عميق .**

**ثالثاً: المراجعة الطباعة والتوزيع :**
**مراجعة الباحث أسئلته حتى يزيل الغموض وأيضا اكتشاف الأخطاء والتزام الموضوعية .**
**كما يجب عرضها على متخصصين وأيضا من الممكن أن يقوم بتجربتها قبل أن يطبقها حتى يستكشف الأخطاء.**

**توزيع الاستبيان على المبحوثين أو عينة الدراسة سواء مباشرة أو عن طريق البريد وهنا يجب أن يراعى الباحث وجود غلاف خارجي يشتمل على التعليمات وأيضا على البيانات الأساسية كما يجب أن يهتم الباحث بوقت التوزيع والذي يجب أن يكون مناسب للمبحوثين .**
**رابعاً: تجميع البيانات :**
**عند توزيع الاستبيان هناك مشاكل قد تواجه الباحث عند جمع البيانات من المبحوثين وخصوصا إذا كان الباحث استعان بأشخاص آخرين لمساعدته على التوزيع ولكنها أقل جهداً من توزيع الاستبيان عبر البريد وخصوصا إننا لم نصل إلى درجة من الوعي في كيفية التعامل مع الاستبيان فقد لا يرجع الباحث إلا عدد قليل جدا هذا إذا رجع إليه استمارات عن طريق البريد.**

**خامساً: تصنيف البيانات وتبويبها :**
**تتعلق بسؤال معين لابد أن نحرص على أن يكون التصنيف متفقا مع الهدف الذي نريد الوصول إليه وهذا معناه أن نحدد الوسيلة التي تمكننا من جمع الإجابات ذات الخاصية الواحدة دون الخلط بينها وهذا يقتضى ضرورة وضع علامة مميزة لكل إجابة أو ظاهرة .**
**سادساً: المميزات والعيوب**
**مزايا الاستمارة :**
**1- لا تحتاج إلى عدد كبير من الباحثين لتنفيذه وجمع بياناته .**

**2- يوفر للمبحوث حرية الاختيار والإجابة سواء من حيث الوقت الذى يختاره كل منهم أو من حيث اختيار الكلمات التي يرونها معبرة عن استجاباته .**
**3- يمكن للمبحوث التعبير بحرية عن الجوانب الخاصة .**
**4- يمكن من جمع المعلومات من مجتمع تنتشر أفراد عينته على مساحات جغرافية واسعة دون تكاليف مرتفعة .**

**عيوب الاستمارة :**
**1- لا يصلح في المجتمعات التي تنتشر فيه الأمية وينخفض الوعي .**
**2- الفروق الفردية في فهم الأسئلة .**

**خامسا أسلوب الحصر أو المسح الشامل** :
الحصر الشامل : في هذه الطريقة يتم جمع البيانات عن جميع وحدات المجتمع محل الدراسة .وذلك حسب الحالات التالية

1- إذا كان الهدف من الدراسة جمع البيانات عن جميع وحدات المجتمع وبصفة فردية مثل التعداد السكاني .
2- وفي حالة عدم وجود كشوفات يمكن سحب عينة منها .
3- للحصول على مستوى عال من دقة والسلامة كما في حالة فحص أنابيب الغاز .
4- إذا مفردات المجتمع محل الدراسة غير متجانسة .
5- إذا المجتمع محل الدراسة صغيرا نسبيا .

**وقد يتعذر على الباحث استخدام هذه الطريقة للأسباب التالية :**

1 صعوبة الوصول إلى كل عنصر من عناصر المجتمع الإحصائي
2 الحاجة إلى وقت طويل وجهد كبير
3 النفقات والتكاليف الباهظة
4 عدم إمكانية دراسة المجتمع ففي بعض الدراسات يتسبب جمع البيانات من إتلاف أو تلوث عناصر المجتمع .

**سادسا طريقة العينات :**

يتم فيها استخدام أساليب الاستقراء وهذه الأساليب لها شروط ومتطلبات يجب مراعاتها عند جمع البيانات وهي طريقة المعاينة المناسبة وحجم العينة المناسب والقياس المناسب للمتغيرات . ويتوفر فيها الصدق والثبات وتحديد ذلك يتم باستخدام الأساليب الإحصائية المختلفة .

واستخدام العينات الإحصائية أصبح أمرا حتميا يفرضه المنطق والاعتبارات الاقتصادية والعلمية فله مميزات وحالات لاستغني عنها وهي .

1- تقليل التكاليف والإمكانيات .
2- السرعة في إظهار النتائج .
3- دقة البيانات والمعلومات التي تم جمعها .
4- الفحص قد يكون صعوبة واستحالة فحص المجتمع بالكامل .وذلك إما بسبب كبر حجم المجتمع اوعدم إمكانية تحديد المجتمع .
5- الفحص قد يكون متلفا للوحدات محل الدراسة مثل فحص الأطعمة والمفرقعات وغيرها .
6- الفحص قد يكون مؤذيا للوحدات مثل فحص دم المريض .
7- البيانات والتسجيلات التاريخية قد تكون غير مكتملة .

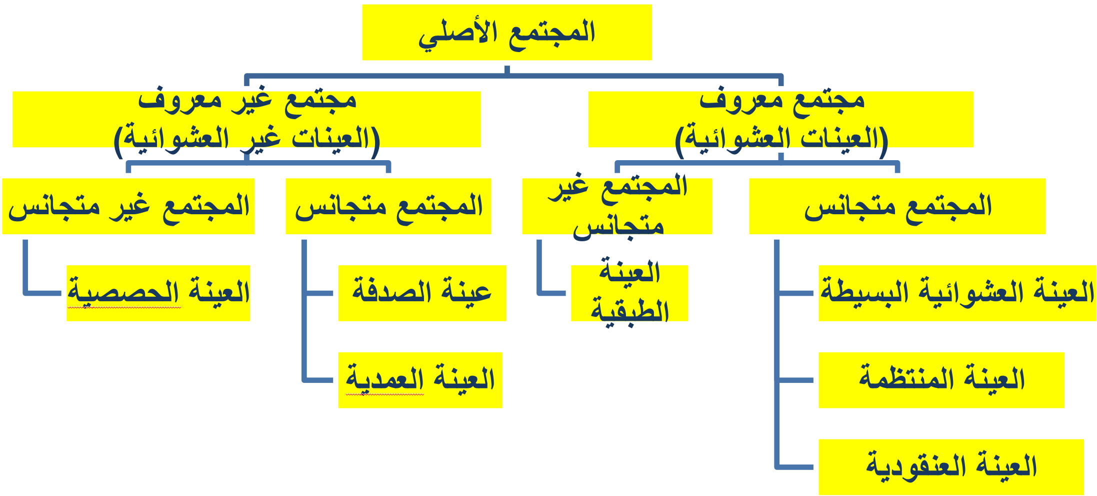
****يتّم العينات تحديد الفرق بين العينة والمجتمع المأخوذة منه، أسلوب المعاينة يتوقف على عدة عوامل منها: تحديد حجم العينة. نوع العينة. اختيار مفردات العينة.
******* **تعريف العينة و المجتمع وأسلوب المعاينة .**

**المجتمع الإحصائي :**

**العينة :**

**أسلوب المعاينة :**

**أنواع العينات :**




1- **العينات الاحتمالية أو العشوائية** ، وهي عينات تختار عشوائية من المجتمع لضمان عدم التحيز، ومن أنواعها؛ العينة العشوائية البسيطة، والعشوائية الطبقية، والعشوائية المنتظمة، والعشوائية متعددة المراحل.

2- **العينات غير الاحتمالية أو غير العشوائية** ، يتم اختيار العينات بطرق مدروسة غير عشوائية بما يحقق الهدف من الإحصاء، ومن أنواعها؛ العينة العمدية والعينة الحصصية.

**العينات الاحتمالية أو العشوائية:**

العينة العشوائية هي العينة التي يتم اختيار عناصرها اعتمادا على احتمال حدوثها وهي جزء من المجتمع قيد الدراسة وتمثل المجتمع تمثيلا صادقا وخالية من التحيز الذاتي للباحث .

ويمكن تصنيف العينات العشوائية على النحو التالي:

1- **العينة العشوائية البسيطة :** وهي العينة التي يتم اختيارها بحيث تكون فرص اختيار جميع المجموعات الجزئية من الحجم نفسه متساوية.

وتستخدم هذه العينة عندما يكون جميع عناصر المجتمع معروفة وعند وجود التجانس بين أفراد المجتمع وليس لأي منها أفضلية على غيرها في تمثيل المجتمع .ويتم اختيارها أما بطريقة القرعة أو بطريقة جداول الأرقام العشوائية

2- **العينة العشوائية المنتظمة :** وهي العينة التي يتم اختيار عناصرها عشوائيا بحيث يفصل بين العناصر المختارة فترات أو مسافات منتظمة .

وتستخدم هذه الطريقة عندما يكون المجتمع منتهيا ومتجانسا حيث إن الباحث يختار مفرداتها وفق نظام معين وذلك وفق فترات منتظمة بين الوحدات المختارة عشوائيا إي أن الباحث يختار رقما ثابتا بين كل عنصرين .

ويتم تحديد المسافة بين العنصرين من خلال القانون التالي :

المسافة المنتظمة = حجم المجتمع الكلي / حجم العينة

3- **العينة الطبقية :** تستخدم هذه العينة عندما يكون المجتمع مقسما إلى طبقات ( فئات ) مثل مجتمع العاملين في فندق حيث يتكون من إداريين عمال فنيين طهاة ........ الخ .

وتتلخص طريقة اختيارها بان نأخذ عددا من أفراد كل طبقة يتناسب مع حجم تلك الطبقة بشكل عشوائي حسب القانون التالي :

حجم العينة من كل فئة = عدد أفراد الفئة / عدد أفراد المجتمع * حجم العينة المطلوبة

4- **العينة العنقودية :** هي عينة عشوائية بسيطة يكون العنصر فيها عنقود ( مجموعة) يتم اختيارها بعد تقسيم المجتمع إلى عناقيد مختلفة شاملة لجميع عناصر المجتمع , ومن ثم نقوم بسحب عينة عشوائية من هذه العناقيد .

5- **العينة غير العشوائية ( غير الاحتمالية )**

وهي العينة التي يتم اختيار عناصرها دون الاعتماد على احتمال وقوعها بحيث تتدخل فيها رغبة الباحث وإمكاناته الشخصية .

وتستخدم العينة العشوائية إذا كان المجتمع غير متجانس أو مفرداته غير معروفة بدقة , وطبيعة هذا النوع من العينات غير احتمالي إي لاعتمد على الاحتمالات ولذا يلجا الباحث فيه إلى معايير معينة يضعها حسب قناعاته ورغبته الشخصية ,ويمكن تصنيف العينات غير الاحتمالية إلى ما يلي :

6- **العينة الحصصية :** هي العينة التي يتم اختيار عناصرها من بين مفردات المجتمع بعد تقسيمه إلى حصص أو فئات بحيث تمثل جميع الفئات بالأهمية نفسها .
7- **العينة العمدية أو القصدية :** هي العينة التي يعتمد الباحث اختيار مفردتها على أساس أنها تحقق أغراض بحثه لكونه يعرف أن هذه المفردات المختارة تمثل المجتمع تمثيلا صادقا بناء على خبراته أو دراسات سابقة
8- **عينة الصدفة :** هي العينة التي يتم اختيار مفرداتها من الأفراد الذين يقابلهم الباحث عن طريق الصدفة
9- **العينة التطوعية** : هي العينة التي تتألف من مجموعة من الأشخاص المتطوعين بالمشاركة في تجربة معينة

**ثالثا الاستقراء :**

الاستقراء هو دراسة جزء من المجتمع الإحصائي وهذا الجزء يسمى عينة بحيث تعمم نتائج الدراسة على هذا الجزء من المجتمع محل الدراسة .

**رابعا اتخاذ القرار :**

اتخاذ القرار يرتبط بوجود هدف يراد تحقيقه باختيار أحدى البدائل المتاحة في الدراسة .
الإحصاء والعلوم الأخرى :
الإحصاء يهتم بدراسة النواحي الكمية لمختلف الظواهر التي تخضع للدراسة لذلك توجد نقاط مشتركة بين علم الإحصاء والعلوم الأخرى مثلا الاقتصاد وعلم الوراثة –علم الاجتماع – والعلوم الزراعية والطبيعية وغيرها من العلوم الأخرى .

**علاقة الإحصاء بالاقتصاد :**

أصبحت الدراسات إحصائية التطبيقية مؤشرا هاما في تحليل النظريات الاقتصادية ومدى لاءمتها للواقع .فالأسلوب الكمي في التحليل الاقتصادي هو استخدام البيانات العددية والمعطيات الإحصائية في الدراسات الاقتصادية . فيتم التعبير عن مختلف القوانين الاقتصادية بمعادلات رياضية توضح العلاقة بين مختلف المتغيرات الاقتصادية التي تخضع للصفة العشوائية .

الإحصاء والسكان :

للإحصاء علاقة وثيقة بعلم السكان لأنه يعبر عن كثير من المؤشرات السكانية كالولادات والوفيات والهجرة .

الإحصاء والرياضيات :

يستخدم الإحصاء بفرعيه النظري والتطبيقي القوانين الرياضية بشكل واسع في المجالات المتعلقة بالمجتمع . لقد أدى تطور علم الإحصاء إلى ظهور فروع عديدة منها الإحصاء التطبيقي , نظرية الاحتمالات , الإحصاء الرياضي , الانحدار والارتباط .....الخ .

**مراحل البحث العلمي :**

لدراسة أي ظاهرة لابد من جمع بيانات معينة لتحليل الظاهرة .لذلك علينا إتباع أهم خطوات البحث الإحصائي وهي :

1- تحديد الهدف الذي يرمي إليه البحث .
2- تحديد المجتمع المراد دراسته .
3- تحديد مصادر البيانات .
4- جمع هذه البيانات .
5- تصنيف البيانات وعرضها .
6- تحليل البيانات .
7- استخلاص النتائج وتفسيرها واتخاذ القرار المناسب لحل المشكلة محل الدراسة .

**طرق عرض البيانات :**

**العرض البياني و العرض ألجدولي**

**تبويب البيانات :**

بعد جمع البيانات يتم ترتيبها وتنظيمها في صور وأشكال يسهل فهمها واستيعابها .

**عرض البيانات :**

1- العرض ألجدولي :
2- يهدف العرض ألجدولي إلى إعطاء فكرة سريعة لتوزيع الظاهرة محل الدراسة .

**خطوات تكوين الجدول التكراري كالآتي :**

1- تحديد الفئات :
إذا كان عدد البيانات N فان عدد الفئات M يحسب بالقانون التالي

$$M=1+3.3\log N$$

2- تحديد مدى الفئات :
وهو عبارة عن الفرق المطلق بين اكبر مفردة واصغر مفردة في المجموعة .
3- تحديد طول كل فئة
4- تحديد حدود الفئات .

نطبق المثال في التالي

مثال لدينا علامات ثمانين طالبا في مادة الاحصاء المطلوب تكوين جدول تكراري لهذه البيانات

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 22 | 32 | 44 | 50 | 54 | 57 | 65 | 70 | 77 | 85 |
| 23 | 33 | 44 | 50 | 54 | 59 | 66 | 72 | 78 | 86 |
| 26 | 36 | 45 | 51 | 54 | 60 | 67 | 74 | 79 | 87 |
| 27 | 37 | 45 | 52 | 55 | 61 | 68 | 74 | 79 | 88 |
| 27 | 38 | 46 | 52 | 55 | 61 | 68 | 75 | 80 | 92 |
| 29 | 39 | 47 | 52 | 55 | 62 | 68 | 76 | 81 | 93 |
| 30 | 40 | 49 | 53 | 55 | 63 | 69 | 76 | 83 | 97 |
| 31 | 44 | 49 | 53 | 57 | 64 | 69 | 77 | 84 | 98 |

الحل

أولا نحدد عدد الفئات M وذلك وفقا للقانون

$M= 1 + 3.3 \log N$

حيث N تمثل عدد البيانات

وهنا عدد البيانات 80

فان

$M= 1 + 3.3 \log 80 \quad = 7.820 \approx 8$

ثانيا تحديد مدى الفئات وهي اكبر مفردة – اقل مفردة ويرمز له بالرمز R

$R = M-m =98- 22 = 76$

:M: أكبر قيمة

m : اصغر قيمة

ثالثا تحديد طول الفئة .ويرمز له بالرمز I or C

$I= (M-m)/1+ 3.3\log N$

$I= (98- 22)/7.820 = 10. 38$

تحديد حدود الفئة

يجب تحديد حدود الفئات وذلك بإظهار الفئة الدنيا والفئة العليا نبدأ من 20 ونضيف 10 وهي طول الفئة تكون الفئات كما يلي

| التكرارات f | بيانات التفريغ | الفئات |
|---|---|---|
| 6 | ~~////~~ / | 20 -30 |
| 8 | ~~////~~ /// | 30-40 |
| 10 | ~~////~~ ~~////~~ | 40-50 |
| 18 | ~~////~~ ~~////~~ ~~////~~ /// | 50-60 |
| 14 | ~~////~~ ~~////~~ //// | 60-70 |
| 12 | ~~////~~ ~~////~~ // | 70-80 |
| 8 | ~~////~~ /// | 80-90 |
| 4 | //// | 90-100 |

10- الصور:

11- تشكل الصور الطريقة الأكثر استخداماً وتأثيراً في عرض البيانات؛ نظراً لما تتركه من تأثير عميق في نفس المتلقّي، فيستمتع فيها بشكل كبير خلال إظهاره للتفاعل مع ما يستعرض أمامه من بيانات، وتتسم هذه الطريقة بالقدرة العالية على ترسيخ البيانات في ذهن المتلقي لأطول فترة ممكنة، والإنسان بطبيعته يرحّب بفكرة تقديم المعلومات له بواسطة الصور.

**12- الأعمدة:**

13- تعتبر هذه الطريقة بمثابة وسيلة تسهيل وتبسيط لقراءة البيانات والاطلاع عليها من قبل المتلقي عند إجراء مقارنات بين مجموعة من القيم المتفاوتة فيما بينها بأسلوب سهل جداً، تساهم هذه الطريقة أيضاً بتحفيز المتلقي على اتخاذ القرار بشكل أسرع وأسهل.

**14- الأشكال الهندسية:**

15- يُلجأ إلى استعراض البيانات بواسطة الأشكال الهندسية لتمثيل البيانات والأرقام المختلفة بأبسط الطرق، حيث يمكن استغلال الشكل المثلث في وظيفة عرض البيانات، والدائرة، حيث تتناسب المساحات مع الأرقام بشكل كبير.

**16- الرسوم البيانية:**

ترتكز هذه الطريقة على الخطوط وانحناءاتها في تمثيل البيانات، لتكون قادرة على رصد أبعاد العلاقات المتفاوتة بين الظواهر، فمن الممكن أن تُمثل بواسطة خطوط متصلة أو متقطعة.

**الجداول والتوزيعات التكرارية :**

**أعداد جدول التوزيع التكراري التجميعي الصاعد:**

| الفئات | التكرارات f | $cf_i$ |
|---|---|---|
| 20 -30 | 6 | 6 |
| 30-40 | 8 | 14 |
| 40-50 | 10 | 24 |
| 50-60 | 18 | 42 |
| 60-70 | 14 | 56 |
| 70-80 | 12 | 68 |
| 80-90 | 8 | 76 |
| 90-100 | 4 | 80 |

**أعداد جدول التوزيع التكراري التجميعي الهابط :**

| الفئات | التكرارات f | $cf_i$ |
|---|---|---|
| 20 -30 | 6 | 80 |
| 30-40 | 8 | 70 |
| 40-50 | 10 | 66 |
| 50-60 | 18 | 56 |
| 60-70 | 14 | 38 |
| 70-80 | 12 | 24 |
| 80-90 | 8 | 12 |
| 90-100 | 4 | 4 |

**المدرج التكراري :**

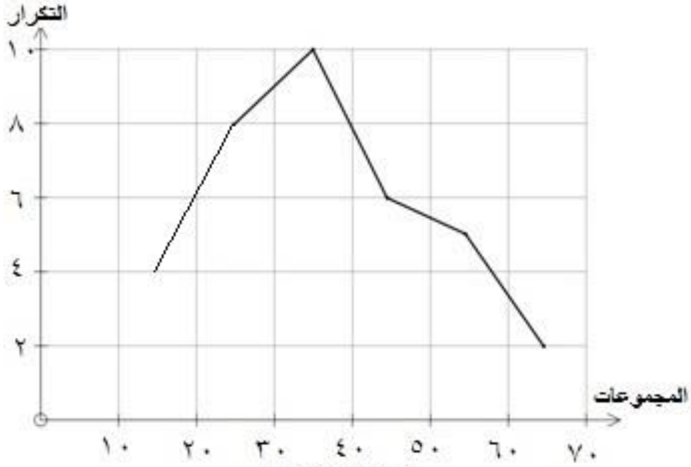
عبارة عن مستطيلات متلاصقة قواعدها عبارة عن الفئات وارتفاعاتها عبارة عن التكرارات المناظرة .بالتطبيق في المثال التالي

المطلوب رسم المدرج التكراري لبيانات درجات او علامات الطلاب في مادة الاحصاء .

| الفئات | التكرارات f |
|---|---|
| 20 -30 | 6 |
| 30-40 | 8 |
| 40-50 | 10 |
| 50-60 | 18 |
| 60-70 | 14 |
| 70-80 | 12 |
| 80-90 | 8 |
| 90-100 | 4 |

**المضلع التكراري :**

المضلع التكراري عبارة عن منحنى ممهد بيمر بكل النقاط قواعدها مراكز الفئات وارتفاعها عبارة عن تكراراتها . كما في الشكل التالي .




المضلع التكراري

التمثيل البياني بواسطة الدائرة :

يستخدم التمثيل البياني بواسطة الدائرة بالإضافة إلى الأعمدة البيانية لإظهار نسب التوزيع لمقومات ظاهرة معينة وفي هذه الحالة تقسم الدائرة إلى قطاعات تتناسب مع حجم الوحدات المدروسة وفق الخطوات التالية .

1- تحويل قيم الظاهرة إلى نسب مئوية .
2- حساب المساحة المقابلة لكل قيمة من قيم الظاهرة تساوي P وان قيمة الزاوية المركزية في الدائرة هي $360^0$ .
3- فيكون قيم الظاهرة النسبية تعادل قيم $360^0$P تقابل S ومنه

P= A/sum*%

حيث S الزاوية المركزية التابعة لهذه القيمة .

مثال :

توزيع الحجاج القادمون لموسم الحج حسب طريقة الوصول لعام 1414- 1408

| المجموع | جوا | بحرا | برا | |
|---|---|---|---|---|
| 762 | 516 | 42 | 204 | 1408 |
| 997 | 839 | 40 | 118 | 1414 |

المطلوب تمثيل هذه البيانات بواسطة دائرتين مقسمتين إلى قطاعات .

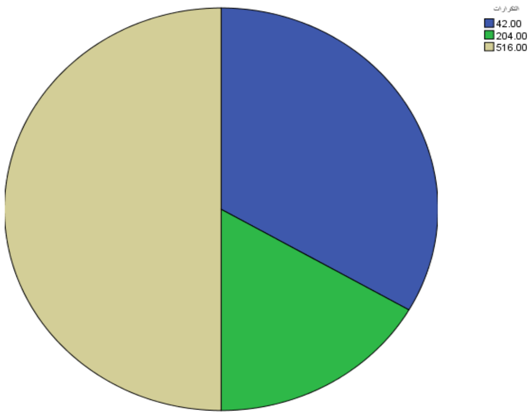
لتمثيل هذه البيانات بواسطة الدائرة نقوم بتحويل هذه الأرقام إلى نسب مئوية ثم نحسب قيمة الزاوية المقابلة لكل نسبة مئوية .يحسب وفقا للقانون

P= A/sum*%*360 حيث A تمثل أي مفردة مثلا تساوي 240او 42و....... 839

$$P = \frac{204}{762} * 360 = 96.4$$

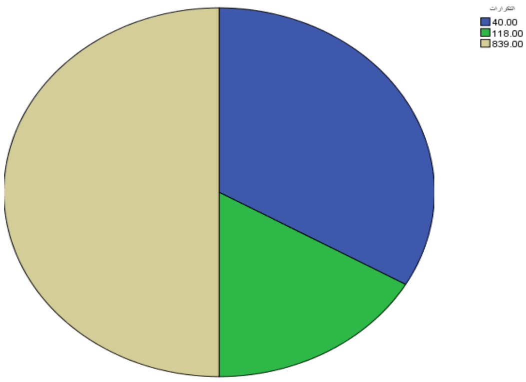
وبهذه الطريقة نحسب بقية القيم والجدول التالي يمثل حساب بقية القيم .

| جوا | بحرا | برا | |
|---|---|---|---|
| 243.7 | 19.8 | 96.4 | 1408 |




رسم دائري لعام 1408

| جوا | بحرا | برا | العام |
|---|---|---|---|
| 302.9 | 14.44 | 42.6 | 1414 |




رسم يوضح التمثيل الدائري لسنة 141

## الوحدة الثانية

## مقاييس النـزعة المركزية Measures of Central location

### مصطلحات الوحدة الثانية :

- مقاييس النزعة المركزية: والتي يقصد بها المقاييس التي تتمركز معظم البيانات حولها، كما تعبر أيضا عن القيم المثلى التي يتم توزيع معظم البيانات بالقرب منها.

أولا **الوسط الحسابي: Mean أو** المتوسط الحسابي : Mean Arithmetic

لبيانات غير مبوبة :

و هو عبارة عن حاصل ما جمع من البيانات مقسوم على عدد هذه البيانات.

فإن الوسط الحسابي( المعدل ) لمجموعة البيانات $x_1, x_2, ...x_n$ هو مجموع هذه البيانات مقسوماً على عددها ويرمز لذلك بالرمز $\overline{x}$ ويعطى بالعلاقة التالية :

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

$$= \frac{x_1 + x_2 + ... + x_n}{n} \qquad (1-3)$$

**مثال ( 1 ) :**

اوجدي الوسط الحسابي للبيانات التالية

2,5,8,6,2,1

$$\bar{x} = \frac{\sum_{i=1}^{n} X}{n} = \frac{2+5+8+6+2+1}{6} = \frac{24}{6} = 4$$

**مثال ( 2 ) :**

عدد الرسائل اليومية التي تلقاها مكتب تجاري خلال ثمانية أيام عمل هي كالآتى :

2 , 0 , 3 , 15 , 4 , 13 , 5 , 6

إن متوسط عدد الرسائل اليومية في هذه الحالة هو :

$$\overline{x} = \frac{1}{8}[2+0+3+15+4+13+5+6] = \frac{48}{8} = 6$$

## طريقة الترميز :

إذا كانت البيانات معطاة بجدول توزيع تكراري ذو k فئة فإن الوسط الحسابي يعطي بالعلاقة :

$$\overline{x} = \frac{1}{\sum_{i=1}^{k} f_i}\sum_{i=1}^{k} x_i f_i \qquad (2-3)$$

حيث رمزنا بـ $f_i$ ، $x_i$ لتكرار الفئة $i$ ومركزها على الترتيب وكذلك $\sum_{i=1}^{k} f_i$ مجموع التكرارات

**مثال ( 1 ) :**

من علامات الطلاب في الاحصاء اوجد الوسط الحسابي

| الفئات | التكرارات f |
|---|---|
| 20 -30 | 6 |
| 30-40 | 8 |
| 40-50 | 10 |
| 50-60 | 18 |
| 60-70 | 14 |
| 70-80 | 12 |
| 80-90 | 8 |
| 90-100 | 4 |
| مجموع | 80 |

الحل

$$\bar{x} = \frac{\sum_{i=1}^{n} fx}{\sum_{i=1}^{n} f}$$

نوجد مراكز الفئات $X = \frac{\text{الفئة الاولى + الفئة الثانية}}{2}$

مثلا $\frac{20+30}{2} = \frac{50}{2} = 25$

| الفئات | التكرارات f | X | fX |
|---|---|---|---|
| 20 -30 | 6 | 25 | 150 |
| 30-40 | 8 | 35 | 280 |
| 40-50 | 10 | 45 | 450 |
| 50-60 | 18 | 55 | 990 |
| 60-70 | 14 | 65 | 910 |
| 70-80 | 12 | 75 | 900 |
| 80-90 | 8 | 85 | 680 |
| 90-100 | 4 | 95 | 380 |
| المجموع | 80 | | 4740 |

$$\bar{x} = \frac{\sum_{i=1}^{n} fx}{\sum_{i=1}^{n} f} = \frac{4740}{80} = 59.25$$

**مثال ( 2 ) :**

## : احسب المتوسط الحسابى من جدول الفئات التالى :

| الفئات | -10 14 | -15 19 | -20 24 | -25 29 | -30 34 | -35 39 | -40 44 | -45 49 | 54-50 |
|---|---|---|---|---|---|---|---|---|---|
| التكرار | 3 | 3 | 4 | 7 | 12 | 8 | 5 | 5 | 3 |

**الحل:**

$$\bar{x} = \frac{\sum_{i=1}^{n} fx}{\sum_{i=1}^{n} f}$$

**يتم حساب منتصف كل فئة .**

| الفئات | التكرارات $f_i$ | $X_i$ | $F_i X_i$ |
|---|---|---|---|
| **14 – 10** | **3** | **12** | **36** |

نوجد مراكز الفئات X = $\frac{\text{الفئة الاولى} + \text{الفئة الثانية}}{2}$

**وتتضح هذه الخطوات فى الجدول التالى:**

| | | | |
|---|---|---|---|
| 19 – 15 | 3 | 17 | 51 |
| 24 – 20 | 4 | 22 | 88 |
| 29 – 25 | 7 | 27 | 189 |
| 34 – 30 | 12 | 32 | 384 |
| 39 – 35 | 8 | 37 | 396 |
| 44 – 40 | 5 | 42 | 210 |
| 49 – 45 | 5 | 47 | 235 |
| 54 – 50 | 3 | 52 | 156 |
| مجموع التكرارات | 50 | | 1645 |

**ثم نعوض فى المعادلة الموضحة سابقا لنحصل على المتوسط من الجدول التكراري للفئات .**

$$\bar{x} = \frac{\sum_{i=1}^{n} fx}{\sum_{i=1}^{n} f} = \frac{1645}{50} = 32.9$$

**<u>أهمية المتوسط:</u>**

**1- إعطاء صورة عن المستوى العام للمجموعة ، لأنه يمثل الدرجة التي يتجمع حولها معظم أفراد المجموعة .**

**2- المقارنة بين المجموعات.**

**3- تحديد مستوى كل فرد في المجموعة ، وذلك لأن مستوى الفرد يتحدد بانحراف درجته عن متوسط المجموع**

**فمثلاً إذا كان متوسط الذكاء = 100% فإن الفرد الذي يحصل على 120 يكون ذكائه فوق المتوسط.أما الفرد الذي يحصل على 89 يكون ذكائه تحت المتوسط**

**4- يستخدم المتوسط مع كثير من الأساليب الإحصائية الأخرى في عمليات التحليل الإحصائي .**

## الخواص الإحصائية للمتوسط:

**1-** من خصائص الوسط الحسابي أن مجموع انحرافات القيم عن الوسط الحسابي يساوي صفر .

$$\sum_{i=1}^{n}(\mathrm{X}-\bar{\mathrm{x}})=0$$

الإثبات

بإدخال علامة الجمع داخل القوس

$$\sum_{i=1}^{n}\mathrm{X_i}-\mathrm{n}\bar{\mathrm{X}}=0$$

بما أن $\bar{\mathrm{x}}=\frac{\sum_{i=1}^{n}\mathrm{X}}{\mathrm{n}}$ بضرب الطرفين في n نحصل على الآتي

$$\mathrm{n}\bar{\mathrm{X}}=\sum_{\mathrm{i}=1}^{\mathrm{n}}X$$

بالتعويض نحصل على

$$\mathrm{n}\bar{\mathrm{X}}-\mathrm{n}\bar{\mathrm{X}}=0$$

إذا مجموع انحرافات القيم عن الوسط الحسابي يساوي صفر

$$\sum_{i=1}^{n}(\mathrm{X}-\bar{\mathrm{x}})=0$$

**2 - يتأثر المتوسط المتطرفة أي بالقيم القريبة من بعض ، ويتأثر بالقيم البعيدة عن تأثرًا كبيرًا ( وهذه الخاصية من أهم عيوب المتوسط) .**

**مثال:**

**الدرجات : 8، 9، 10 ، 11 ، 12 ، 13 متوسطها = 10.5**

**الدرجات : 7 ، 9 ، 10 ، 11 ، 13 ، 35 متوسطها = 14.17**

**الدرجات : 7 ، 9 ، 10 ، 11 ، 13 ، 1 متوسطها = 8.5**

**3- يتأثر المتوسط بعدد الدرجات ويميل إلى الاستقرار كلما كان العدد كبيرًا**

**فإذا كان لدينا 100 درجة فإن زيادة درجة واحد على الـ 100 يكون مقدار الزيادة 0.01 أما لو كان لدينا 1000 درجة فزيادة درجة أخرى تعنى واحد من ألف.**

**فالمتوسط يتأثر فى حالة الأعداد الصغيرة.**

**4- إذا كان لدينا مجموعتين من الدرجات س ، ص فإن:**

**متوسط مجموع درجات المجموعتين**

## ثالثا: الوسط الحسابي المرجح:

**في بعض الأحيان يكون لكل قيمة من قيم المتغير أهمية نسبية تسمى أوزن ، أو ترجيحات ،وعدم أخذ هذه الأوزان في الاعتبار عند حساب الوسط الحسابي ، تكون القيمة المعبرة عن الوسط الحسابي غير دقيقة ، فمثلا لو أخذنا خمسة طلاب ، وسجلنا درجات هؤلاء الطلاب في مقرر الإحصاء التطبيقي ، وعدد ساعات الاستذكار في الأسبوع**

| مسلسل | 1 | 2 | 3 | 4 | 5 | sum |
|---|---|---|---|---|---|---|
| $x$ ( الدرجة) | 23 | 40 | 36 | 28 | 46 | 173 |
| $w$ ( عدد ساعات الاستذكار ) | 1 | 3 | 3 | 2 | 4 | |

نجد أن الوسط الحسابي غير المرجح للدرجة الحاصل عليها الطالب هي :

$$\bar{x}=\frac{\sum x}{n}=\frac{23+40+36+28+46}{5}=\frac{173}{5}=346$$

وإذا أردنا أن نحسب الوسط الحسابي للدرجات $x$ المرجحة بعدد ساعات الاستذكار $w$ ، يتم تطبيق المعادلة التالية :

$$(\overline{w})=\frac{\sum xw}{\sum w}=\frac{23\times1+40\times3+36\times3+28\times2+46\times4}{1+3+3+2+4}$$

$$=\frac{23+120+108+56+184}{13}=\frac{491}{13}=37.769$$

وهذا الوسط المرجح أكثر دقة من الوسط الحسابي غير المرجح .

إذا الوسط الحسابي المرجح $(\overline{w})$ يحسب بتطبيق المعادلة التالية :

$$\boxed{(\overline{w})=\frac{\sum xw}{\sum w}}$$

**مزايا وعيوب الوسط الحسابي :**

**يتميز الوسط الحسابي بالمزايا التالية :**

**ـ أنه سهل الحساب . ٠**

**ـ يأخذ في الاعتبار كل القيم . ٠**

**ـ أنه أكثر المقاييس استخداما وفهما . ٠**

**ومن عيوبه :**

**ـ أنه يتأثر بالقيم الشاذة والمتطرفة . ٠**

**ـ يصعب حسابه في حالة البيانات الوصفية . ٠**

**ـ يصعب حسابه في حالة الجداول التكرارية المفتوحة .**

**المتوسط الكلى = متوسط درجات المجموعة الأولى + متوسط درجات المجموعة الثانية .**

**5 ـ إذا كان لدينا مجموعتين من الدرجات فإن :**

**متوسط الفرق بين درجات المجموعتين = متوسط المجموعة الأولى – متوسط المجموعة الثانية**

**بشرط: تساوى عدد درجات المجموعتين.**

## **ثانياً :** الوسيط :Median

- وهو عبارة عن القيمة العددية التي تقوم بتقسم البيانات لقسمين متساويين وذلك بعد ترتيبها تصاعديا أو تنازليا.

**أو هو الدرجة التي تقسم توزيع الدرجات إلى قسمين متساويين من حيث العدد بحيث يكون النصف الأول للدرجات يساوى النصف الثاني للدرجات**

**فالوسيط هو نقطة التوسط في أي توزيع بحيث يصبح عدد القيم التي تعلوه مساويا لعدد القيم التي دونه . وإذا كان عدد القيم صغيرا فإنه في الإمكان إيجاد الوسيط بترتيب القيم تصاعديا أو تنازليا .**


## **طرق حساب الوسيط :**

**يختلف حساب الوسيط من الدرجات المبوبة الغير أو الخام باختلاف عدد الدرجات، فعدد الدرجات إما أن يكون عددًا فرديا أو عددًا زوجيا، لذلك نجد أن لدينا طريقتين لحساب الوسيط من الدرجات الخام.**

**<u>أ ـ حساب الوسيط إذا كان عدد الدرجات فرديا</u>**

- **ترتب الدرجات تصاعديا أو تنازليا.**
- **يتم حساب موقع الوسيط (ترتيب أو رتبة أو مكان الوسيط بين الدرجات المرتبة). من المعادلة البسيطة التالية :**

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

**مثال (1 ) احسبي الوسيط للبيانات التالية .**

5,6,8,4,3

أولا نرتب البيانات تصاعديا 3,4,5,6,8

ثانيا ترتيب البيانات

$$\frac{n+1}{2}=\frac{6}{2}=3$$
$$\therefore \text{الوسيط}=5$$

**<u>ب ـ حساب الوسيط إذا كان عدد الدرجات زوجيا:</u>**

**فى حالة عدد الدرجات n زوجيا فإنه توجد قيمتان للوسيط هما:**

**الأولى ترتيبها يساوى** $\frac{n}{2}$

**والثانية ترتيبها يساوى** $\frac{n}{2}+1$

**فالوسيط الحقيقي هو متوسط الوسيطان**

مثال جد قيمة الوسيط للبيانات التالية

2,8,4,6,5,3,7,0

الحل أولا نرتب البيانات تصاعديا

0,2,3,4,5,6,7,8

**ترتيب الوسيط الأول يساوى** $\frac{n}{2} = \frac{8}{2} = 4$

**ترتيب الوسيط الثاني يساوى** $\frac{n}{2}+1 = \frac{8}{2}+1 = 5$

∴ قيمة الاولالوسيط 5و الوسيط الثاني 4

قيمة الوسيط الحقيقي $4.5 = \frac{5+4}{2}$

**<u>2 ـ حساب الوسيط من الجدول التكرارى للفئات :</u>**

**إذا كان عدد الحالات أو عدد القيم كبيرًا وموضوعة في شكل توزيع تكراري فإننا نتبع الخطوات التالية لحساب الوسيط :**

**1 ـ نكون الجدول التكراري المتجمع الصاعد أو الهابط .**

**2- نحدد <u>ترتيب</u> أو موقع الوسيط وهو يساوى .**

**مجموع التكرارات مقسوما على العدد 2 .**

**3- نحدد فئة الوسيط التي يقع ترتيب الوسيط داخل حديها الأدنى والأعلى وذلك بالاستعانة بالجدول التكراري المتجمع الصاعد أو الهابط حيث نبحث عن القيمة التي تناظر ترتيب أو موقع الوسيط ثم نحسب قيمة الوسيط من المعادلة التالية:**

**4 ـ قيمة الوسيط في حالة التكرار المتجمع الصاعد من القانون التالي .**

**ويحسب بالعلاقة الآتية :**

$$\text{Median} = L_1 + \frac{(\frac{n}{2} - cf_1)i}{cf_2 - cf_1}$$

**حيث :**

**$L_1$ : مبدأ الفئة الوسيطية .**

**$\frac{n}{2}$ : ترتيب الوسيط وهوأول ما يحسب .**

**$cf_i$ : التكرار المتجمع الصاعد السابق لفئة الوسيط .**

**$cf_2$ :التكرار المتجمع الصاعد اللاحق لفئة الوسيط .**

**I : طول الفئة الوسيطية .**

مثال :

من بيانات علامات الطلاب في مادة الإحصاء اوجد قيمة الوسيط

| الفئات | التكرارات f | | $cf_i$ التكرار المتجمع الصاعد |
|---|---|---|---|
| 20 -30 | 6 | اقل من 30 | 6 |
| 30-40 | 8 | اقل من 40 | 14 |
| 40-50 | 10 | اقل من 50 | 24<br>$40 \xrightarrow{\text{الوسيط ترتيب}}$ |
| 50-60 | 18 | اقل من 60 | 42 |
| 60-70 | 14 | اقل من 70 | 56 |
| 70-80 | 12 | اقل من 80 | 68 |
| 80-90 | 8 | اقل من 90 | 76 |
| 90-100 | 4 | اقل من 100 | 80 |
| المجموع | 80 | | |

$$\frac{80}{2} = 40$$

نحدد أين يقع الـ 40 في عمود $cf_i$

نلاحظ يقع بين 24 و42

يسمى $cf_1$ 24 و $cf_2$ 42

50: $L_1$

10: I

$$\therefore \text{Median} = 50 + \frac{(40-24)10}{42-24}$$

$$= 58.89$$

## ثالثا الربيعيات :

الربيعيات ثلاثة الربيع الأول ويرمز له بالرمز $Q_1$ *والربيع الثاني ويرمز له بالرمز* $Q_2$ *وهو الوسيط والربيع الثالث* $Q_3$ .

إيجاد الربيع الأول لبيانات غير مبوبة .

**أولا نحسب ترتيب الربيعيات في حالة البيانات الخام أو غير مبوبة كالأتي :**

| الربيع / البيانات | $Q_1$ | $Q_2$ | $Q_3$ |
|---|---|---|---|
| غير مبوبة | $\frac{n+1}{4}$ | $\frac{2(n+1)}{4}$ | $\frac{3(n+1)}{4}$ |
| مبوبة | $\frac{n}{4}$ | $\frac{2n}{4}$ | $\frac{3n}{4}$ |

**ثانيا نحدد موقع الربيع بعد ترتيبها تصاعديا أو تنازليا**

**الربيع الأول :**

**لإيجاد قيمة الربيعيات نتبع نفس خطوات الوسيط .**

**مثال اوجد الربيع الأول للبيانات التالية 9,5,2,0,6,4,8**

***الحل***

***أولا نرتب البيانات تصاعديا* 0,2,4,5,6,8,9**

***ثانيا ترتيب الربيع* $Q_1$**

$$\frac{n+1}{4} = \frac{7+1}{4} = \frac{8}{4} = 2$$

**إذا قيمة الربيع الأول هو $Q_1 = 2$**

**أما الربيع الثاني فهو الوسيط .**

**مثال اوجد الربيع الثالث للبيانات السابقة**

***أولا نرتب البيانات تصاعديا* 0,2,4,5,6,8,9**

***ثانيا ترتيب الربيع*** **$Q_3$**

$$\frac{3(n+1)}{4}=\frac{3(7+1)}{4}=\frac{24}{4}=6$$

**اذا قيمة $Q_3=8$**

**<u>حساب الربيعيات لبيانات مبوبة أو بيانات ذات فئات :</u>**

**لإيجاد الربيعيات لبيانات مبوبة نحتاج إلى التكرار المتجمع الصاعد أو الهابط وفي هذه الحالة ترتيب الربيع يساوي مجموع التكرارات مقسوما على 4 في حالة $Q_1$. ويحسب بالعلاقة التالية :**

$$Q_1=L_1+\frac{(\frac{n}{4}-cf_1)i}{cf_2-cf_1}$$

**حيث $L_1$ مبدأ الفئة الربيعية .**

**$\frac{n}{4}$ ترتيب الربيع الأولوهو أول ما يحسب .**

**$Cf_1$ التكرار المتجمع الصاعد السابق لفئة الربيع الأول .**

**$cf_2$التكرار المتجمع الصاعد اللاحق لفئة الربيع الأول .**

**I : طول الفئة الربيعية .**

**إما الربيع الثالث يحسب بالعلاقة التالية:**

$$Q_3=L_1+\frac{(\frac{3n}{4}-cf_1)i}{cf_2-cf_1}$$

**حيث $L_1$ مبدأ الفئة الربيعية .**

**$\frac{3n}{4}$ ترتيب الربيع الثالثوهو أول ما يحسب .**

**$Cf_1$ التكرار المتجمع الصاعد السابق لفئة الربيع الثالث .**

**$cf_2$التكرار المتجمع الصاعد اللاحق لفئة الربيع الثالث .**

**I : طول الفئة الربيعية .**

**مثال بالتطبيق على بيانات علامات الطلاب اوجد $Q_1$و$Q_3$**

| الفئات | التكرارات f | | $cf_i$ |
|---|---|---|---|
| 20 -30 | 6 | اقل من 30 | 6 |
| 30-40 | 8 | اقل من 40 | 14<br>الوسيط ترتيب<br>20 ⟶ |
| 40-50 | 10 | اقل من 50 | 24 |
| 50-60 | 18 | اقل من 60 | 42 |

| | | | |
|---|---|---|---|
| 56<br>60 $\xrightarrow{\text{ترتيب الوسيط}}$ | اقل من 70 | 14 | 60-70 |
| 68 | اقل من 80 | 12 | 70-80 |
| 76 | اقل من 90 | 8 | 80-90 |
| 80 | اقل من 100 | 4 | 90-100 |
| | | 80 | المجموع |

**الحل**

**نحسب ترتيب** $Q_1 \frac{n}{4} = \frac{80}{4} = 20$

$$Q_1 = L_1 + \frac{(\frac{n}{4} - cf_1)i}{cf_2 - cf_1}$$

$$Q_1 = 40 + \frac{(20 - 14)10}{24-14}$$

$$Q_1 = 46$$

**نحسب ترتيب** $Q_3$

$$Q_3 = L_1 + \frac{(\frac{3n}{4} - cf_1)i}{cf_2 - cf_1}$$

$$\frac{3n}{4} = \frac{3*80}{4} = 60$$

$$Q_3 = L_1 + \frac{(\frac{3n}{4} - cf_1)i}{cf_2 - cf_1}$$

$$Q_3 = 70 + \frac{(60 - 56)10}{68-56}$$

$$Q_3 = 73.3$$

**العشيرات ( Deciles ):**

**مفردها العشير وهو ناتج التقسيم العشري ، 0.1iهو العشر الأدنى ($D_1$) من القيم أي ما قبله عشر القيم وبعده تسعة أعشار القيم (0.9i) وتحسب قيمته كحساب الوسيط.**

**المئينات (Percentile):**

**المئين الأول هو التقسيم 1iمن 100iورمزه $C_1$ أي $C_1 = 0.01$ من قيم المعطيات والمئين الستين $C_{60}$ يقع عند 0.60iوتحسب قيمته كحساب الوسيط.**

## رابعا المنوال Mode :

يعرف المنوال على أنه هو القيمة الأكثر تردداً أو الأكثر شيوعاً أو تكراراً ونرمز له بالرمز $x^{\circ}$ فمثلاً للقيم 1,1,2,3,3,4,5,5,5,6,7 منوالا واحدا وهو $x^{\circ} = 5$

قد لا يكون لمجموعة من القيم منوال كما هو الحال في القيم 1,2,3,4,5,6,7 وقد يكون لمجموعة من القيم أكثر من منوال واحد كما هو الحال في القيم 1,1,1,2,2,2,3,4,5,5,6 .

و إذا كانت البيانات غير عددية (أي نوعية ) كأن نسأل مثلا ستة أشخاص عن أي الألوان المحببة لهم ، قد تكون إجاباتهم كالتالي:

أزرق, أصفر, أبيض, أزرق, أبيض أو أزرق فسيكون المنوال في هذه الحالة هو اللون "الأزرق" .

**مثال (1)**

*فإذا كان لدينا القيم التالية اوجدي المنوال .*

4,5,8.9,12,8,12, 15,8

*الحل المنوال هو العدد* 8

***المنوال لبيانات مبوبة :***

إذا كانت البيانات مبوبة في جدول توزيع تكراري فنحسب المنوال كما يلي : نعين الفئة المنوالية وهي الفئة التي تقابل أكبر قيمة للتكرارات ثم نحسب الفرق بين تكرار الفئة المنوالية وتكرار الفئة التي قبلها$d_1$ ثم نحسب الفرق بين تكرار الفئة المنوالية وتكرار الفئة التي بعدها وليكن $d_2$ نرمز للحد الأدنى الفعلي للفئة المنوالية بالرمز $L_1$ ولطول الفئة بالرمز i فنحصل على المنوال وفقا لطريقة *الفروق لبيرسون تعتبر طريقة بيرسون أدق الطرق لحساب المنوال وتحسب بالعلاقة التالية* :

.

$$mode = l_1 + \frac{d_1 * I}{d_{1+\ d_2}}$$

حيث :

$L_1$ : *مبدأ المنوالية*

I : *طول الفئة المنوالية*

$D_1$ : *الفرق بين تكراري الفئة المنوالية والفئة السابقة*

$D_2$ : *الفرق بين تكراري الفئة المنوالية والفئة اللاحقة عليها*

*مثال :*

*من بيانات علامات الطلاب في مادة الإحصاء اوجد المنوال*

| التكرارات f | الفئات |
|---|---|
| 6 | 20 -30 |
| 8 | 30-40 |
| 10 | 40-50 |
| 18<br>التكرار المنوالي → | 50-60 |
| 14 | 60-70 |
| 12 | 70-80 |
| 8 | 80-90 |
| 4 | 90-100 |
| 80 | المجموع |

$L_1 = 50$

*التكرار المنوالي* = 18

$D_1 = 18-10=8$

$D_2 = 18-14=4$

*طول الفئة المتوالية*10=

$$mode = l_1 + \frac{d_1 * I}{d_{1+}\ d_2}$$

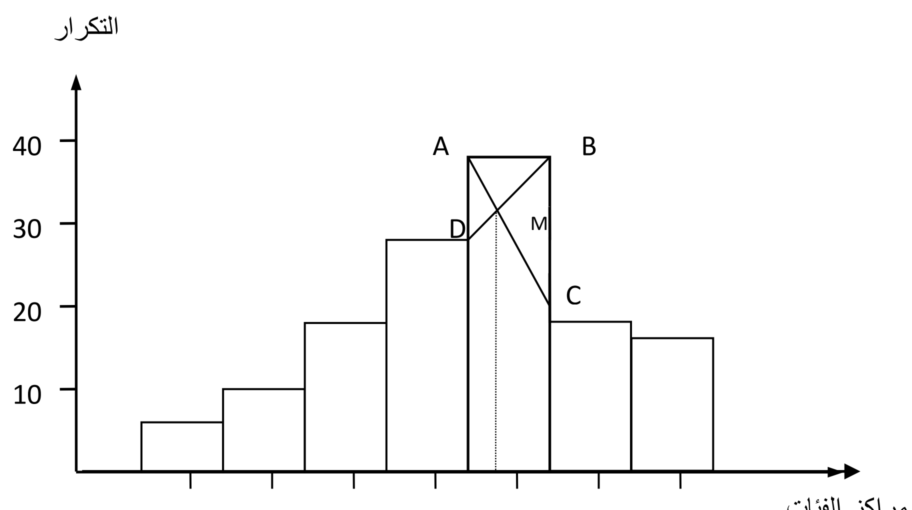
$$mode = 50 + \frac{8 * 10}{8 + 4}$$

=50 + 6.67 =56.67

## ***تحديد المنوال هندسيا :***

حساب المنوال بيانياً ( هندسياً ) وذلك بعد رسم المدرج التكراري للبيانات المبوبة وعلى سبيل المثال لنحسب المنوال للبيانات **.**

*باستخدام المدرج التكراري نحدد الفئة المتوالية وهو المقابل لأكبر تكرار برسم خط من المستطيل الذي يمثل الفئة المنوالية مع التكرار السابق واللاحق والسابق ونقطة تقاطع الخطين يمثل المنوال على المحور الأفقي كما في هو موضح في الشكل التالي :*



وذلك وفقاً لما يلي :

1- نرسم المدرج التكراري موضحاً الفئة المنوالية والفئة التي قبلها والفئة التي تليها.

2- نوصل بين رأس الفئة المنوالية والزاويتين المجاورتين للفئة المنوالية أي AC , BD .

3- نسقط من نقطة الالتقاء M عموداً على محور السينات تمثل قيمة المنوال وهنا يمكن تعيين النقطة $x^{o}$ بشكل تقريبي وبصورة هندسية.

وذلك وفقاً لما يلي :

1- نرسم المدرج التكراري موضحاً الفئة المنوالية والفئة التي قبلها والفئة التي تليها.

2- نوصل بين رأس الفئة المنوالية والزاويتين المجاورتين للفئة المنوالية أي AC , BD .

3- نسقط من نقطة الالتقاء M عموداً على محور السينات تمثل قيمة المنوال وهنا يمكن تعيين النقطة $x^{o}$ بشكل تقريبي وبصورة هندسية.

**هناك علاقة تقريبية بين المنوال $x^{o}$ والوسط $\bar{x}$ والوسيط $\tilde{x}$ من أجل التوزيعات المتناظرة نسبيا بحيث تؤمن معرفتنا لنقطتين النقطة الثالثة وذلك من العلاقة التالية :**

$$x^{\circ} = \bar{x} + 3(\tilde{x} - \bar{x})$$

**لنحسب مثلاً منوال البيانات الواردة في الجدول 3-1، حيث وجدنا أن $\tilde{x} = 45.9$ .**

**أما الوسط الحسابي فمن العلاقة ( 3-2 ) نجد أن** $\bar{x} = 45.1$

**وبتطبيق العلاقة السابقة نجد** $x^{\circ} = 45.1 + 3(45.9 - 45.1) = 47.5$

**هناك علاقة تقريبية بين المنوال $x^{o}$ والوسط $\bar{x}$ والوسيط $\tilde{x}$ من أجل التوزيعات المتناظرة نسبيا بحيث تؤمن معرفتنا لنقطتين النقطة الثالثة وذلك من العلاقة التالية :**

***عندما يكون التوزيع متماثلا فان الوسط الحسابي يساوي الوسيط ويساوي المنوال.***

***علاقة بين الوسط والوسيط والمنوال***

***المنوال- الوسط = 3 (الوسط –الوسيط ) .***

**أي المنوال = الوسط +3( الوسيط – الوسط )**

$$x^{\circ} = \bar{x} + 3(\tilde{x} - \bar{x})$$

**لنحسب مثلاً منوال البيانات الواردة في الجدول 3-1، حيث وجدنا أن $\tilde{x} = 45.9$ .**

**أما الوسط الحسابي فمن العلاقة ( 3-2 ) نجد أن** $\bar{x} = 45.1$

**وبتطبيق العلاقة السابقة نجد** $x^{\circ} = 45.1 + 3(45.9 - 45.1) = 47.5$

## خامسا الوسط الهندسي :

يعرف الوسط الهندسي لمجموعة من القيم $X_1, X_2, \ldots, X_n$

بأنه يساوي حاصل ضرب مجذور إلى قوة عددها N حيث

N : تمثل عدد البيانات . ويرمز له بالرمز G .

**الوسط الهندسي لبيانات غير مبوبة يحسب وفقا للقانون**

$$G = \sqrt[N]{X1, X2, \ldots \ldots, Xn}$$

مثال اوجدي الوسط الهندسي G للبيانات التالية 2,4,7,8,5.3

الحل

$$G = \sqrt[N]{X1, X2, \ldots \ldots, Xn}$$

$$G = \sqrt[6]{2 * 4 * 7 * 8 * 5 * 3}$$

$$G = \sqrt[6]{6720}$$

إما إذا كانت البيانات كبيرة فان

$$G = 10^{\frac{\sum logX}{n}}$$

مثال اوجدي الوسط الهندسي G للبيانات التالية 10,100,1000

الحل أولا نوجد لوغر ثم البيانات

| | | | |
|---|---|---|---|
| $X_i$ | 10 | 100 | 1000 |
| Log $X_i$ | 1 | 2 | 3 |

Log X = 1+2+3=6

$$G = 10^{6/3}$$

$$G = 10^{2}$$

G =100

**الوسط الهندسي لبيانات مبوبة أو بيانات ذات فئات يحسب بالقانون التالي :**

$$G = \sqrt[\Sigma f]{\mathrm{x}_1^{\mathrm{f}} * \mathrm{x}_2^{\mathrm{f}}, \ldots \ldots * x_{\mathrm{n}}^{\mathrm{f}}}$$

حيث $x_i$ عبارة عن مراكز الفئات المناظرة للتكرارات $f_i$

مثال : من مثال علامات الطلاب في مادة الاحصاء اوجد الوسط الهندسي لهذه البيانات

| $X_i$ | التكرارات f | الفئات |
|---|---|---|
| 25 | 6 | 20 -30 |
| 35 | 8 | 30-40 |
| 45 | 10 | 40-50 |
| 55 | 18 | 50-60 |
| 65 | 14 | 60-70 |
| 75 | 12 | 70-80 |
| 85 | 8 | 80-90 |
| 95 | 4 | 90-100 |
| | 80 | المجموع |

الحل

$$G = \sqrt[\Sigma f]{\mathrm{x}_1^{\mathrm{f}} * \mathrm{x}_2^{\mathrm{f}}, \ldots \ldots * x_{\mathrm{n}}^{\mathrm{f}}}$$

$$G = \sqrt[80]{25^6 * 35^8 * 45^{10} * 55^{18} * 65^{14} * 75^{12} * 85^8 * 95^4}$$

## سادسا الوسط التوافقي :

يعرف الوسط التوافقي لسلسلة $X_1, X_2, \ldots, X_n$ بأنه مقلوب الوسط الحسابي لمقلوبات تلك القيم ويرمز له بالرمزH. ويعبر عن ذلك رياضيا بالعلاقة التالية :

$$\mathrm{H} = \frac{\mathrm{n}}{\frac{1}{\mathrm{X}_1} + \frac{1}{X_2} + \cdots + \frac{1}{X_n}}$$

يستخدم الوسط التوافقي عندما يراد حساب العدد في الوحدة النقدية أو السرعة في الوحدة الزمنية .

مثال احسب الوسط التوافقي للبيانات التالية 2,4,5,8,10

الحل

$$\mathrm{H} = \frac{\mathrm{n}}{\frac{1}{\mathrm{X}_1} + \frac{1}{X_2} + \cdots + \frac{1}{X_n}}$$

$$\mathrm{H} = \frac{5}{\frac{1}{2} + \frac{1}{4} + \frac{1}{5} + \frac{1}{8} + \frac{1}{10}}$$

$$\mathrm{H} = \frac{5}{\frac{47}{40}}$$

$$H = \frac{5 * 40}{47}$$

H=4.26

**الوسط التوافقي لبيانات مبوبة أو ذات فئات يحسب وفقا للقانون التالي :**

$$\mathrm{H} = \frac{\sum f}{\frac{\mathrm{f}_1}{\mathrm{X}_1} + \frac{f_2}{X_2} + \cdots + \frac{f_n}{X_n}}$$

حيث $X_i$ تمثل مراكز الفئات .

مثال من بيانات علامات الطلاب في مادة الاحصاء اوجدي الوسط التوافقي .

$$\mathrm{H} = \frac{\sum f}{\frac{\mathrm{f}_1}{\mathrm{X}_1} + \frac{f_2}{X_2} + \cdots + \frac{f_n}{X_n}}$$

| الفئات | التكرارات f | X |
|---|---|---|
| 20 -30 | 6 | 25 |
| 30-40 | 8 | 35 |
| 40-50 | 10 | 45 |
| 50-60 | 18 | 55 |
| 60-70 | 14 | 65 |
| 70-80 | 12 | 75 |
| 80-90 | 8 | 85 |
| 90-100 | 4 | 95 |
| المجموع | 80 | |

الحل

$$\mathrm{H} = \frac{80}{\frac{6}{25} + \frac{8}{35} + \frac{10}{45} + \frac{18}{55} + \frac{14}{65} + \frac{12}{75} + \frac{8}{85} + \frac{4}{95}}$$

**تدريبات:**

1- **الوسط الحسابي هو أحد مقاييس**
   أ- **النزعة المركزية ب– التشتت ج- التواء**
2- **لعدد من القيم يعرف بأنه مجموع القيم مقسوماً على عددها على أنه ؟**
   أ- **الوسط الحسابي لهذه القيم ب- الانحراف المتوسط للقيم**
   **جـ - تباين القيم**

**الوسيط لمجموعة من القيم المرتبة ترتيبا تصاعديا أو تنازليا هو؟**
**أ – القيمة التي تقسم مجموعة القيم إلى مجموعتين متساويتين في العدد**
**ب- القيمة الأكثر تكرارا جـ - متوسط اكبر قيمة وأقل قيمة**
**4 – اذا كان لدينا القيم *4,9,8,5,4***
**الوسط الحسابي يساوي أ- *8* ب - *5* جـ - *6***
**الوسيط يساوي أ- *8* ب - *5* جـ - *6***
**المنوال يساوي أ- *8* ب - *5* جـ - غير موجود**
**4- إذا كان لدينا *9, 3 ,2 ,8 ,4 ,16***

**الوسط الحسابي يساوي أ- *8* ب - *5* جـ-*7***
**الوسيط يساوي أ- *8* بغير موجود - جـ - *6***
**المنوال يساوي أ- *8* ب - *5* جـ - غير موجود**

**تدريب :**

**اثبت أن انحرافات 7 ، 5 ، 3 ، 4 ، 6 عن الحسابي = مجموع الدرجات متوسطها صفر**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| س | 6 | 7 | 9 | 7 | 5 | 7 | 9 | 8 | 5 | 2 |
| ص | 5 | 8 | 10 | 4 | 3 | 7 | 10 | 7 | 6 | 9 |

**تدريب :**

**من الجدول التالى احسب المتوسط الكلى للدرجات س وص وكذلك الفرق بين متوسطى درجات المجموعتين**

## مقاييس التشتت

هي التي تعبر عن مدى تباعد القيم وتقاربها.

تختلف مقاييس النزعة المركزية من ناحية التفسير عن مقاييس التشتت. ويمكن الاستدلال مباشرة عن القيمة التي تتمركز حولها جميع القيم المشمولة في دراسة ما من خلال حساب إحدى مقاييس النزعة المركزية، بينما لا يمكن تفسير القيمة الوحيدة المحصلة من خلال حساب إحدى مقاييس التشتت. يتم استخدام مقاييس التشتت في الأصل في عمليات المقارنة بين مجموعتين من البيانات، ففي حال توفر مجموعة أخرى من البيانات يمكن حساب مقاييس التشتت للمجموعتين ومن ثم الحكم على المجموعة التي لها مقياس تشتت اكبر في القيمة بأنها المجموعة الأكثر تشتتا. لتقريب الصورة، افترض أننا نتعامل مع مجموعتين من القيم لهما نفس وحدة القياس (درجة مثلا) في كل مجموعة ثلاثة قيم. تضم المجموعة الأولى (مجموعة A) القيم 12 و 48 و 90 بينما تضم المجموعة الثانية (المجموعة B) القيم 45 و 52 و 53، عند إيجاد الوسط الحسابي للمجموعتين يتبين أن لهما نفس المتوسط،

$$\mu_A = \mu_B = 50$$

مما يشير للوهلة الأولى بان المجموعتين لهما نفس الصفات الإحصائية، بينما عند حساب إحدى مقاييس التشتت، التباين مثلا،

$$\sigma_A^2 = 1016 \quad \& \quad \sigma_B^2 = 12.7$$

يتبين الفرق الكبير جدا بين المجموعتين. فالمجموعة A لها تشتت اكبر بكثير من المجموعة B، مما يشير إلى أن قيم المجموعة A تتباعد عن بعضها بشكل كبير مقارنة بتباعد قيم المجموعة B. لذلك، فان مقاييس النزعة المركزية ومقاييس التشتت تمثلان عند اجتماعهما أداة قوية تعطي تصور واضح عن صفات القيم إحصائيا.

**أنواع مقاييس التشتت وأهميتها**

يجب العناية بمقاييس التشتت، والتأكد من احتسابها بطريقة صحيحة لأن هذا يؤثر على النتائج الإحصائية التي تنتج عنها، وهذه المقاييس هي:

- **المدى:** يتسم هذا النوع من مقاييس التشتت بالبساطة وسهولة احتساب قيمته، لكنه يعد من أقل هذه المقاييس من حيث الدقة، ويمكن استخدامه في حالة وجود مشاهدات فردية عشوائية، أو في حال وجود مجموعات تكرارية، ويكون ذلك على الشكل التالي:

- في المشاهدات الفردية: يمثل المدى الفرق بين أكبر مشاهدة موجود، وأصغر مشاهدة ضمن هذه المشاهدات جميعها ويعبر عنه رياضيًا بـ:
المدى = قيمة أكبر مشاهدة – قيمة أقل مشاهدة

في التوزيعات التكرارية: يمثل الفرق بين الحد الفعلي الأعلى للفئة العليا و الحد الفعلي الأدنى للفئة الدنيا، ويعبر عنه رياضيا بـ

المدى = الحد الفعلي الأعلى للفئة العليا – الحد الفعلي الأدنى للفئة الدنيا .

مثال من البيانات التالية اوجد المدى *3, 24,9 ,8,10,23*

المدى*2 = 3-124*

مثال من بيانات علامات الطلاب اوجد المدى لهذه البيانات

*100- 20 = 80*

الانحراف الربيعي :

يمثل نصف الفرق بين الربيع الثالث والربيع الأول.

من مثال علامات الطلاب اوجدي الانحراف الربيعي

لبيانات مبوبة

$\frac{Q3-Q1}{2} = 73.3 - 46 = 27.3$

مثال لبيانات غير مبوبة

*5,3,7,0,12,4,2,1,8,7,4*

المطلوب الانحراف الربيعي

اولانرتب البيانات تصاعديا *0,1,2,3,4,4,5,7,7,8,12*

ثانيا ترتيب الربيع الأول

$\frac{n+1}{4} = 3 = \frac{11+1}{4}$

إذا الربيع الأول هو *2*

لإيجاد الربيع الثالث

اولانرتب البيانات تصاعديا *0,1,2,3,4,4,5,7,7,8,12*

ثانيا ترتيب الربيع الأول

$\frac{3(n+1)}{4} = 9 = \frac{3(11+1)}{4}$

إذا الربيع الثالث هو *7*

## 3 الانحراف المتوسط: Mean deviation

يعتبر الانحراف عن المتوسط أو ( الانحراف المتوسط ) أحد مقاييس التشتت .

ويعرف الانحراف المتوسط بأنه متوسط الفروق للبيانات عن وسطها الحسابي بقيمها المطلقة فإذا كانت لدينا مجموعة لبيانات غير مبوبة

$$x_1, x_2, \ldots, x_n$$

وسطها الحسابي $\bar{x}$ فإنه يحسب الانحراف المتوسط للقيم $X_1, X_2, \ldots, X_n$ بالصيغة

بالعلاقة التالية :

$$M_D = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \bar{x}\right| \qquad (6\text{-}3)$$

$MD = \frac{\sum_{i=1}^{n}|X_i - \overline{x}|}{N}$ أو

مثال اوجد الانحراف المتوسط للبيانات التالية

*4,7,2,8,9*

أولا نوجد الوسط الحسابي للبيانات

$$MD = \frac{\sum_{i=1}^{n}|X_i\ - \overline{x}|}{N}$$

$$\overline{x} = \frac{\sum_{i=1}^{n} X_i}{n}\ \overline{x} = \frac{4+7+2+8+9}{5} = 6$$

$$MD = \frac{|4-6| + |7-6| + |2-6| + |8-6| + |(9-6)|}{5}$$

$$\therefore\ MD = \frac{12}{5} = 2.4$$

إيجاد الانحراف المتوسط لبيانات مبوبة يحسب وفقا للقانون

$$MD = \frac{\sum_{i=1}^{n} f_i|X_i\ - \overline{x}|}{\sum_{i=1}^{n} f_i}$$

مثال اوجد الانحراف المتوسط لبيانات علامات الطلاب في مادة الاحصاء

| الفئات | التكرارات $f$ | $X_i$ | $F_i X_i$ | $\lvert Xi\ - \bar{X}\rvert$ | $\sum f\ \lvert Xi - \bar{X}\rvert$ |
|---|---|---|---|---|---|
| *20 -30* | *6* | *25* | *150* | $\lvert 25-59.25\rvert = -34.25$ | *6*34.25* |
| *30-40* | *8* | *35* | *280* | $\lvert 35-59.25\rvert = -24.25$ | *8*24.25* |
| *40-50* | *10* | *45* | *450* | $\lvert 45-59.25\rvert = -14.25$ | *10*14.25* |
| *50-60* | *18* | *55* | *990* | $\lvert 55-59.25\rvert - 4.25$ | *18*4.25* |
| *60-70* | *14* | *65* | *910* | $\lvert 65-59.25\rvert = 5.75$ | *14*5.75* |
| *70-80* | *12* | *75* | *900* | $\lvert 75-59.25\rvert = 15.75$ | *12*15.75* |
| *80-90* | *8* | *85* | *680* | $\lvert 85-59.25\rvert = 25.75$ | *8*25.75* |
| *90-100* | *4* | *95* | *380* | *=35.75* $\lvert 95-59.25\rvert$ | *4*35.75* |
| المجموع | *80* | | *4740* | | *1236.9* |

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

= *59.25*

*MD*= $\frac{1236.9}{80} = 15.46$

## الانحراف المعياري Stander deviations :

يعتبر الانحراف المعياري والتباين من أهم مقاييس التشتت الإحصائية. ويرتبط المقياسين بعلاقة رياضية قوية، حيث يمكن دوما الحصول على المقياس الآخر في حال معرفة قيمة احدهما. يرمز للتباين بالرمز $\sigma^2$ في حال الحصول على قيمته من خلال تغطية مجتمع الدراسة، بينما يتم استخدام الرمز $S^2$ للدلالة على مقدر التباين المحصل من خلال بيانات عينة عشوائية مسحوبة من مجتمع الدراسة. وبأخذ الجذر التربيعي للتباين يتم الحصول على قيمة الانحراف المعياري وذلك في الحالتين، حالة المجتمع وحالة العينة،

- $\sigma = \sqrt{\sigma^2} \qquad \text{or} \qquad S = \sqrt{S^2}$
- وبحكم العلاقة الرياضية القوية بين كل من التباين والانحراف المعياري فإنه يمكن اعتبارهما وجهين لعملة واحدة لهما نفس الأهمية.
- يعتمد الانحراف المعياري والتباين على فكرة تربيع الفروق بين قيم المتغير الكمي $X$ ووسطها الحسابي، وفي حال التعامل مع مجتمع بحجم $N$ يتم الحصول على الانحراف المعياري من خلال الدالة التالية،
- $$\sigma_X = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \mu_X)^2}{N}}$$
- حيث يمكن تبسيطها رياضيا لتصبح،
- $$\sigma_X = \sqrt{\frac{\sum_{i=1}^{N} X_i^2}{N} - \mu_X^2} \qquad (26)$$
- أما في حالة التعامل مع عينة عشوائية بحجم $n$ مسحوبة من مجتمع الدراسة وممثلة له، فان صيغة حساب الانحراف المعياري هي،
- $$S_X = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$
- وبالتبسيط الرياضي تصبح،

$$S_X = \sqrt{\frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1}} \qquad (27)$$

**ملاحظات هامة :**

إن السبب في القسمة على ( n-1 ) عوضاً عن n لأن هناك (n-1 ) إنحرافاً مستقلاً من الشكل $x_i - \bar{x}$ . ولأن مجموع هذه الانحرافات يساوي الصفر دوماً فإن أيا منها يساوي مجموع كل البقية بإشارة سالبة و أي منها يعطى بدلالة مجموع القيم الأخرى وبإشارة معاكسة . ولتوضيح هذه الفكرة تصور أن لدينا ثلاث بيانات $x_1$ , $x_2$ , $x_3$ ولدينا $\bar{x} = \frac{1}{3}(x_1 + x_2 + x_3)$ .

نعلم أن $(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) = 0$

وبالتالي يمكن التعبير عن أي منهم وليكن الأول بـ $x_1 - \bar{x} = -[(x_2 - \bar{x}) + (x_3 - \bar{x})]$

أو $x_2 - \bar{x} = -[(x_1 - \bar{x}) + (x_3 - \bar{x})]$

**بعض خصائص الإنحراف المعياري :**

1 – قيمة الانحراف المعياري دائماً موجبة أو أكبر من أو تساوي صفر. فأقل قيمة تساوي الصفر (وذلك عندما تكون جميع القيم متساوية، وفي هذه الحالة لا توجد فروق أو إنحرافات بينها وبين الوسط الحسابي وبالتالي لا يوجد أي تشتت بين القيم، وبالتالي فإن قيمة الإنحراف المعياري في حالة تساوي جميع القيم تساوي الصفر).

2 – كلما كان التشتت كبيراً حول الوسط كلما كان الانحراف المعياري كبيراً، والعكس صحيح.

3 – إذا أضفنا وطرحنا مقداراً ثابتاً من كل القيم فإن قيمة الانحراف المعياري (أو التباين) لا تتغير (أي لا تتأثر قيمة الانحراف المعياري بالطرح أو الجمع).

- *الانحراف المعياري لمجموعة من القيم هو الجذر التربيعي الموجب لمتوسط مجموع انحرافات القيم عن وسطها الحسابي.*

## التباين ( التشتت ): Variance

يعرف التباين بأنه الوسط الحسابي لمربعات فروقات البيانات عن وسطها الحسابي . ففي مجتمع ما إذا كان هذا المجتمع يتألف من n عنصر وكان وسطه الحسابي معطى وهو يساوي $\bar{x}$ فإن التباين ( التشتت ) $\sigma^2$ ( ويقرأ $\sigma^2$ سجما تربيع ) للمجتمع يعطى بالشكل التالي :

- $$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad (7\text{-}3)$$

*مربع الانحراف المعياري فيسمى .*

***الانحراف المعياري لبيانات غير مبوبة :***

-
- $$\sigma\text{x} = \sqrt{\frac{\sum(X_i-\bar{X})^2}{N}}$$
- *مثال اوجد الانحراف المعياري للبيانات التالية*
- 4,8,6,3,1,2
- *الحل*
- $$\sigma\text{x} = \sqrt{\frac{\sum(X_i-\bar{X})^2}{N}}$$
- *أولا نحسب الوسط الحسابي*
- $=\bar{\text{x}} = \frac{\sum_{i=1}^{n}\text{X}}{\text{n}}$

$$\bar{\text{x}} = \frac{4+8+6+3+1+2}{6}$$

-
- $\bar{\text{x}} = 4$
- $$\sigma\text{x} = \sqrt{\frac{(4-4)^2+(8-4)^2+(6-4)^2+(3-4)^2+(1-4)^2+(2-4)^2}{6}}$$
-
- $$\sigma\text{x} = \sqrt{\frac{16+4+1+9+4}{6}}$$
- $\sigma\text{x} = \sqrt{5.6667}$
- *من ثم اوجد التباين للمثال السابق يرمز للتباين بالرمز* $\sigma_{\text{x}}^2$

- وهو عبارة عن مربع الانحراف المعياري
- $\therefore \sigma_x^2 = 5.667$

## معامل الاختلاف : The Coefficien of Variation

لمقارنة تشتت مجموعتين (أو أكثر) من البيانات وكانت البيانات تختلف في مستواها العام (أي في أوساطها الحسابية) و / أو تختلف في وحدات القياس (مثلاً مقارنة بيانات الدخل حيث تقاس بالريال ببيانات العمر حيث تقاس بالسنوات) فإن المقارنة لا تتم مباشرة بمقارنة الانحراف المعياري لكل منهما بل تتم من خلال مقياس آخر هو "معامل الاختلاف" أو ما يسمى أحياناً مقياس التشتت النسبي حيث ينسب الانحراف المعياري لكل مجموعة إلى وسطها الحسابي والضرب في 100 فنحصل على مقياس نسبي أو مئوي (وبدون تمييز) أي تتم المقارنة بحساب معامل الاختلاف لكل منهما، والمجموعة التي لها معامل اختلاف أكبر تكون أكبر تشتتاً والعكس صحيح أي أن :

معامل الاختلاف = ______ × 100

وإذا رمزنا لمعامل الاختلاف بالرمز C. V فإن :

الانحراف المعياري

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

الوسط الحسابي

مثال (6) إذا كان الوسط الحسابي والانحراف المعياري لدخول عينة من الناخبين بالريالات هو : $\bar{X}_1 = 1500,\ S_1 = 152$

وكان الوسط الحسابي والانحراف المعياري لأعمارهم (بالسنوات) هو :

$\bar{X}_2 = 42,\ S_2 = 9.2$

فأيهما أكثر تشتتاً الدخل أم العمر ؟

الإرتباط: هو من الأساليب الإحصائية التي تهدف لتعيين قوة وطبيعة العلاقة بين متغيرين، أو أكثر.

Plot Scatterالرسم الإنتشاري.

الفروض: هي عبارة عن تخمين ذكي من الباحث.

**بعض المصطلحات الأخرى بالعربي والإنجليزي:**

- النزعة المركزيةBar ChartS
- مركز الفئةClass Limits
- التوزيع التكراري النسبيRelative Frequency Distribution
- عينة عشوائيةRandom Sample
- أرقام قياسيةIndex numbers