جامعة قاصدي مرباح ورقلة

كلية العلوم الاقتصادية والعلوم التجارية وعلوم التسيير

محاضرات في الإحصاء الوصفي

إحصاء 1

مطبوعة موجهة لطلبة السنة الأولى جذع مشترك

د. صياغ أحمد رمزي

2015

الإحصاء الوصفي

Descriptive Statistics

صمم هذا المساق للتمكن من اكتساب المفاهيم الأساسية للإحصاء الوصفي، فهو يقدم بطريقة بسيطة كيفية وصف البيانات العديدة والمتنوعة عن ظاهرة معينة بوضوح وجيز. حيث يتناول أساليب جمع وفرز البيانات، ووصفها، وتبويبها، وعرضها وتلخصها في شكل جداول ورسومات بيانية وكذا من خلال استعراض عدد من المقاييس الأساسية الموضعية والشكلية.

الهدف من هذا المساق هو تطوير قدرات الطلبة في إكساب مهارة اختيار الأساليب المناسبة لوصف البيانات وتحليل الاستبيانات. بالتعرف على الطرق والأساليب الإحصائية التي تساعدهم في مجال تخصصهم على اتخاذ القرارات المناسبة. من خلال تعريف بأنواع البيانات وكيفية ترتيبها وهذا بعرض مبسط للأساليب الإحصائية مع التركيز على المفهوم وكيفية التطبيق.

يركز المحتوى على شرح وتوضيح الإحصاء الوصفي من خلال:

- مفاهيم عامة – مفهوم الإحصاء – المجتمع والعينة والفرد – مصادر البيانات الإحصائية – طبيعة البيانات الإحصائية.
- عرض البيانات الإحصائية – بناء الجداول وأنواعها (إيجاد الفئة، التكرارات، مركز الفئة، التكرار المجمع...)
- التمثيل البياني حسب نوع المتغير
- مقاييس النزعة المركزية الموضعية – الوسط الحسابي، الوسيط، المنوال...
- مقاييس التشتت – نصف المدى الربيعي – الإنحراف المعياري
- الأشكال: الشكل المتماثل – الإلتواء – التفلطح والتذبذب

هذا الكتاب مخصص لطلاب المراحل الأولى في الدراسات العليا تخصص علوم اجتماعية وانسانية، ولجميع فروع الإدارة والتصرف. كما يعد الطلاب لسهولة فهم واستيعاب مقررات التحليل الإحصائي والاحتمالات والاقتصاد القياسي.

الدكتور: صياغ أحمد رمزي
كلية العلوم الاقتصادية والتجارية وعلوم التسيير
جامعة قاصدي مرباح - ورقلة

المحتويات

مفاهيم عامة

General Concepts

تمهيد:

يرجع منشأ الإحصاء Statistics إلى عهد بعيد في تاريخ المدنية الإنسانية والمجتمعات البشرية المنظمة، وقد اهتم قديما الفراعنة المصريين والصينيين والإغريق بإحصاءات تخص مجتمعاتهم وكانت الحاجة ملحّة للحصول على معلومات رقمية أو وصفية حول السكان ومقدار الثروة الزراعية والمعدنية جُمعت للاهتداء بها في تصريف أمور البلد ورسم سياستها. والمفهوم السائد عن الإحصاء هو تلك الأرقام والبيانات التي تقوم الدول والهيئات أو بعض الوكالات بجمعها ومعالجتها لتناسب أغراضا معينة كتلك التي تعنى بتعداد السكان أو تلك التي تهدف الى رصد المواليد والوفيات.

ويعتبر الإحصاء اليوم بقسميه النظري والتطبيقي فرعا مهماً من فروع العلم والمعرفة لأنه يدرس بشكل أساسي الناحية الكمية للظواهر بإستخدام الطرق والمبادئ الإحصائية المناسبة. فهو يدرس الظاهرة حسب المكان وعلاقتها بالظواهر الأخرى، كما يدرس تطور هذه الظواهر حسب الزمان والتنبؤ بحجمها في المستقبل آخذا بعين الاعتبار العوامل التي تؤثر على هذه الظواهر في الماضي وتغير هذه العوامل أو تغير تأثيرها في المستقبل الذي لا غنى عنه لمعرفة حقيقة الظواهر والتخطيط لها.

يستخدم الإحصاء بكثافة في عدة مواقع وتطبيقات ويرجع هذا الاستخدام الى وجود التباين والإختلاف في البيانات المتعلقة بكل حالة مدروسة. فمثلا في أبحاث التسويق نجد أن إجابات الناس عن مجموعة من الأسئلة تتباين وتختلف، وكذلك تتباين كميات أحد المنتجات الزراعية المنتج في حقول مختلفة. هذا التباين والاختلاف يقود الى عدد من الأسئلة المشروعة -إذا كان الناس أو الأشياء تتباين وتختلف في الصفات والمقاييس وفي درجة التأثر بعوامل أو مؤثرات معينة " فما هي الصفة أو المقياس أو درجة الاستجابة التي تكون ممثلة للكل"؟ وكذا ما مدى هذا التباين والاختلاف؟

1- تعريف علم الإحصاء

يمكن تعريف علم الإحصاء بأنه ذلك العلم القائم بذاته الذي يهتم بالبحث في الطرق العلمية لجمع البيانات الخاصة بمختلف الظواهر وتبويبها وتحليلها بهدف التقرير عنها والتحقق منها أو التنبؤ بها، سعيا للوصول إلى نتائج تساعد في اتخاذ القرارات المناسبة. ويعتبر الإحصاء كأداة من أدوات البحث العلمي بمجموعة من الطرائق التي تهدف إلى تجميع معطيات تخص ظواهر تتعلق بمجموعة أفراد من مجتمع ما، وتحليل هذه المعطيات وتفسيرها الذي يعتمد على التحليل الكمي للظواهر واستخدامها في تفهم حقيقة الظواهر والوصول إلى ما تحتويه الظواهر من معان واتجاهات ومعرفة القوانين التي تخضع لها، لتكون أساسا لاستخلاص الاستنتاجات المختلفة التي تساعد على وضع النظريات أو اتخاذ القرارات السليمة.

ويشمل مجال الإحصاءات أساليب نظرية يتم تطبيقها من خلال فرعين للإحصاء التطبيقي (Applied statistics)وهما:

-**إحصاء وصفي : (Descriptive statistic)** الذي يشمل الطرق المستخدمة في وصف وتنظيم وعرض البيانات، وتتمثل هذه الطرق في التوزيعات التكرارية ومقاييس النزعة المركزية والتشتت والرسومات البيانية.

-**إحصاء استدلالي :(Inferential statistic)** والذي يشمل استخلاص النتائج من بيانات محدودة (العينة) وتعميمها على كل البيانات (المجتمع) حيز الدراسة. وهذا يمكن الاستدلال عن معالم المجتمع الإحصائي بطرق مختلفة، ومن تطبيقاته التقدير والإختبارات والتنبؤ وهذا باستخدام النتائج أو المشاهدات السابقة لتجربة ما في تقدير قيم المشاهدات التي قد تحدث مستقبلا.

2- مصادر جمع البيانات الإحصائية

2-1 جمع البيانات (Data collection):

يقصد بجمع البيانات الحصول على معلومات رقمية أو وصفية تتصف بالصحة والدقة عن ظاهرة معينة من مصدر معين في فترة زمنية محدودة، فالبيانات الإحصائية لا تجمع لذاتها ولكن لخدمة هدف معين أو حل مشكلة معينة، ولدارسة أي مشكلة لا بد أن تتوفر عنها بيانات تفصيلية في صورة رقمية تساعد في تحديد حجم هذه المشكلة.

إن مصداقية نتائج أي تحليل إحصائي تكون مبنية على مدى مصداقية ودقة البيانات المستخدمة. وهنالك عدة طرق للحصول على البيانات الإحصائية والتي يمكن ان تتوفر في قسمين من المصادر وهي:

2-2 المصادر التاريخية:

وتصنف باحتوائها للبيانات الثانوية (Secondary data) التي قد تم جمعها في تواريخ سابقة أحيانا بصورة دورية لتخدم أغراضا معينة. وهي البيانات التي يمكن التوفر عليها في الإحصاءات الرسمية والبيانات المنشورة، التي يتولى جمعها وتصنيفها ونشرها دوائر حكومية متخصصة، وكذا الوكالات والهيئات الإحصائية المختصة والبيانات المتوفرة أيضا عبر شبكات الإنترنت. ونشير أن هذه المصادر توفر مشقة جمع البيانات من الميدان وما يترتب عليه من جهد وتكاليف مادية. ونعيب عنها أنها قد تكون قديمة وغير متجددة وقد لا تفي تماما لغرض البحث كما قد يكون بها بعض التحيز التي هي مساوئ تعيق من الاستفادة من البيانات بصورة كاملة.

2-3 المصادر الميدانية أو الاصلية:

والتي يلجا الباحث إليها غالبا حين لا تتوفر بيانات من المصادر التاريخية حيث نتحصل من خلالها على البيانات الأولية (Primary Data) التي يتم تجميعها من مصادرها الأصلية مناسبة لغرض معين

ومخصص ومحدد مسبقا. ومعنى ذلك أن يتصل الباحث بموضوع بحثه مباشرة، فهي تضع الباحث وجها لوجه أمام الظواهر التي يدرسها. فالحصول على بيانات أولية يتطلب إنتهاج طرق الملاحظة أو التجربة أو المسح (Survey) وعلى العموم يتم جمع البيانات من الميدان بإتباع أحد الأسلوبين التاليين:

أولاً: أسلوب الحصر الشامل:

والذي يعتمد على جمع البيانات عن جميع مفردات المجتمع محل البحث، ويتم استخدام هذا الأسلوب في الحالات التالية:

1- البيانات المطلوبة بصفة فردية من مفردات المجتمع وعلى حدى؛
2- الحصول على نتائج البحث على مستوى عال من الدقة؛
3- عدم تجانس مفردات المجتمع وإذا ما كان المجتمع صغير نسبياً.

ثانياً: أسلوب العينات:

ويعتمد على جمع البيانات من مجموعة مختارة من مفردات المجتمع المراد دراسته، ثم يتم دراسة صفات هذه العينة من المفردات التي اختيرت، بطريقة معينة لتمثل أحسن تمثيل المجتمع محل الدراسة. ويتم تعميم النتائج التي يحصل عليها الباحث من بيانات العينة على المجتمع بأكمله. وهناك كثير من الأبحاث والدراسات التي تتم بواسطة هذا الأسلوب باعتباره من الأساليب الأكثر شيوعا في الدراسات الإحصائية الميدانية. ويستخدم أسلوب العينات كذلك عندما يكون المجتمع لا نهائي. إلا أن هناك أخطاء يتعرض لها الباحث عند استخدام أسلوب العينات منها خطأ الصدفة Random errorالذي يحدث نتيجة عدم التجانس في مفردات المجتمع أو لحجم العينة و خطأ التحيز Bias error الذي ينشأ نتيجة لعوامل إنسانية بحتة بسبب سوء اختيار العينة.

وفي ظل هذا السياق فإن التعامل مع العينات لا يكمن فقط في جمع بيانات في حد ذاتها بل أيضا في معرفة كيفية الحصول عليها وكيفية اختيار العينة بالطريقة المناسبة التي تبنى على اعتبارات معينة، مثل طبيعة التباين والاختلاف بين مفردات المجتمع المراد دراسته، وكذلك التكاليف التي يتحملها الباحث. وعموما في عملية اختيار العينات (Sampling methods) يمكن للباحث تبني طرق الاختيار العشوائي (Random sample) المبني على الاحتمالات. ونذكر من أهمها:

- العينة العشوائية البسيطة (Simple random sample)
- العينة الطبقية (Stratified random sample)
- العينة المرحلية (Multi-stage sample)
- العينة المنتظمة (Systematic random sample)
- العينة العنقودية (Cluster or area sample)
- المعاينة المتتابعة (Sequential sampling)

كما للباحث حرية اختيار العينة الغير الاحتمالية (Non-Probabilist samples) التي تعبر عن وجهة نظره دون التقيد بتحديد إطارا للعينة ويتم اختيار مفرداتها بإستخدام الطرق الذاتية أو حسب قناعة الباحث ومعرفته بالمجتمع. ومن أهم هذه العينات:

- عينة الحصص (Quota sample)
- العينة القائمة على الحكم الشخصي (Judgement sample)
- العينة الميسرة (Convenience sample)

فبمجرد أن يتم جمع وتحليل وعرض البيانات الأولية فإنها تصبح بيانات ثانوية قابلة للمعالجة ونستطيع ابتدأ من هنا المرور إلى مرحلة التحليل والتلخيص وصف وترتيب البيانات.

3- المفاهيم الإحصائية (Statistical concepts)

يتميز علم الإحصاء مثل العلوم الأخرى بأنه مليء بالتعاريف والمصطلحات الجوهرية والتي يستلزم فهم معانيها. فمن المصطلحات الأساسية نُعرِف المفاهيم التالية:

3-1- المجتمع (Population): يعرف المجتمع الإحصائي على أنه مجموعة أو تجمع كل المفردات الممكنة موضع الدراسة، سوى كانت هذه الوحدات أفراد أو أشياء أو قياسات والتي تشترك في صفات أو خصائص وسمات محددة جغرافية أو سياسية أو اجتماعية أو اقتصادية أو غيرها. وقد يكون المجتمع الإحصائي محدودا حين نستطيع حصر عدد أفراده أو غير محدود خلفا لذلك.

3-2- العينة (Sample): وهي جزء من المجتمع يتم اختيارها بطريقة معينة لتمثيل المجتمع والاستدلال على خواصه، لذلك يراعى أن تكون العينة ممثلة للمجتمع تمثيلا صادقا. فيتضح جوهريا انه لا يمكننا الوصول إلى أي نتائج مقنعة لأي دراسة إحصائية ما لم يتم تحديد نطاق المجتمع المدروس وضرورة اختيار العينة بطريقة مناسبة حتى تأتي نتائج الدراسة مقنعة وذات مصداقية. ويمكن الحصول على العينة باستخدام أحد أساليب المعاينة (Sampling) الذي سبق التطرق لها سابقا في طرق اختيار العينات.

3-3- الصفة أو الخاصية (Characteristic): ويقصد به مقدار أو سمة يمكن ملاحظتها على الفرد والتي قد تأخذ قيما أو كيفيات أو حالات مختلفة لكل عنصر في المجتمع المدروس. مثل: دخل الأسر في مدينة ما أنواع السيارات في حضيرة ما أو مستوى المبيعات اليومية لأحد محلات التسوق خلال شهر معين...إلخ. فالبيانات الإحصائية(Data) تمثل القيم الفردية لتلك الصفات المدروسة.

3-4- المتغير (Variable) : يعرف كمقدار لقياس خاصية أو ميزة لعناصر المجتمع أو العينة. ويكمن جوهر الإحصاء في دراسة شكل وسلوك تغير الصفة المدروسة عبر أفراد المجتمع المدروس. ومن المتعارف علية أن المتغير يرمز له بأحد الأحرف X أو Y وأن القيمة التي يأخذها المتغير يرمز لها بالأحرف x_i أو y_i ، بالنسبة للفرد i.

3-5- أنواع البيانات والمتغيرات (Types of data and variables) : إن أنواع البيانات تنتسب إلى أنواع المتغيرات التي تدرس خاصية أو ميزة لعناصر المجتمع أو العينة. فكل البيانات التي نتعامل معها هي عبارة عن معلومات تتنوع حسب طبيعة المتغير المدروس صحيحة ودقيقة تجمع من مصادر محددة ، وبطريقة سليمة. تصنف أنواع البيانات أو المتغيرات إلى قسمين:

3-5-1- المتغيرات النوعية (Qualitative or Categorical variables) : وهي بيانات وصفية، تشمل الظاهرات التي لا تخضع للقياسات الكمية، ولا يعبر عنها بصورة عددية. تنقسم المتغيرات النوعية إلى قسمين:

متغيرات نوعية تصنيفية (Nominal variables) أو غير مرتبة كما تعرف أحياناً بالمتغيرات الإسمية، لأنها تصنف البيانات حسب الصفات، سواء أكانت ذلك من حيث النوع أو إلى مجموعات إسمية غير عددية مختلفة لا تتأثر بأي ترتيب منطقي. مثلا للون المفضل (ابيض-احمر-أزرق) أو الحالة العائلية (أعزب–متزوج–مطلق–أرمل) أو أنواع السيارات...إلخ.

متغيرات نوعية مرتبة (Ordinal variables) وهي تصنف البيانات من حيث الدرجة، وفي هذه الحالة نجد أن القيم التي يأخذها المتغير تتبع ترتيب وتسلسل منطقي معد مسبقا أو متفق عليه. مثلا مستوى الطلاب يمكن تصنيفه إلى ممتاز، جيد، مقبول أو فاشل. حيث تعطي كل صفة من هذه الصفات رتبة خاصة، نقيس بها خصائص الظاهرة.

3-5-2- المتغيرات الكمية (Quantitative variable): وهي بيانات رقمية، تشمل الظاهرات القابلة للقياسات الكمية، ويمكن التعبير عنها بصورة عددية تعكس القيم الفعلية للظاهرات. تنقسم المتغيرات الكمية إلى قسمين:

المتغيرات الكمية المنفصلة أو المتقطعة (Discret variables) : وهو متغير يأخذ قيما يمكن عدها وحصرها بالعد وتنتسب للأعداد الطبيعية. مثلا عدد حوادث المرور أو عدد أفراد الأسرة... إلخ.

المتغيرات الكمية المتصلة أو المستمرة (Continuous variables) : وهي التي يمكن قياسها بمعايير وحدوية والتي تأخذ أي قيمة يمكن تمثيلها في مدى معين من الأعداد الحقيقية. فمثلا دخل أسرة أو كميات الأمطار أو وزن المنتوج...إلخ.

في غالب الأحيان يتمثل المتغير الكمي المتقطع في القيم التي يمكن تعدادها وحصرها بعد عدد الوحدات الموجودة بحيث يكون معيار العد غير قابل للتجزئة أو التقسيم. أما المتغير الكمي المتصل يتمثل في القيم التي يمكن قياسها بواسطة معيار وحدوي متفق عليه قابل للتجزئة العشرية (مثل الكلوغرام أو المتر ...إلخ).

تنظيم وعرض البيانات

Organization and Presentation of Data

تمهيد

يعتبر تنظيم وعرض البيانات الإحصائية أول مرحلة للتحليل الإحصائي وتتقيد طريقة تنظيم وعرض البيانات على نوع هذه البيانات وعلى الحقائق المطلوب إبرازها منها. وعليه يمكن تنظيم وعرض البيانات إما عن طريق تصميم التوزيعات أو الجداول التكرارية أو باستعمال الرسوم البيانية. الهدف منها تقديم البيانات (المجمعة مسبقا سواء من المصادر التاريخية أو من المصادر الميدانية) بطريقة مبسطة ومختصرة وهذا ليسهل فهمها واستنتاج بعض النتائج الأولية منها تساعد على تحليلها إحصائياً للتعرف علها ووصفها ومقارنتها بغيرها من الظواهر، والخروج ببعض المدلولات الإحصائية عن مجتمع الدراسة. وكذا الحصول على ملخصات واضحة وموجزة، ولكن الاحتفاظ الكثير من المعلومات الواردة في البيانات الأصلية، واستخدام التقنيات موضوعية وتفادي إعطاء صورة مشوهة للواقع.

1- التوزيعات والجداول التكرارية (Frequency distribution)

عندما يقوم الباحث بتدوين وتسجيل المعلومات المتوفرة عن كل مفردة من مفردات العينة أو المجتمع في دراسة ما فإنه يتحصل على سلسلة عشوائية من البيانات الغير مرتبة. تعرف هذه السلسلة، قبل ترتيبها أو تقسيمها إلى فئات، بالبيانات الخام .(Raw data) أو البيانات الغير مبوبة (Unorganized data). عادتا ما يكون عدد الأفراد كبير وعليه من الصعب قراءة وتفسير سلسلة البيانات الخام. لذا لا بد من تلخيصها. للقيام بذلك، علينا أولا القيام بفرز مسطح، وعد الكيفيات أو القيم، والتي سوف تكون أساس لبناء الجداول والرسوم البيانية.

فالهدف من استخدام الجداول هو تلخيص البيانات وتسهيل فهما ودراستها. لذا يجب أن يكون الجدول المسمى **بجدول التوزيع التكراري** (frequency distribution table) بسيط وغير غامض بقدر الإمكان. يتم فيه وضع البيانات وذلك بحصر الصفات والحالات التي شملها المتغير في البيانات وإيجاد عدد المفردات المناظر لتلك الصفات.

ولكون البيانات الخام التي تخص الظواهر عديدة ومتنوعة يتم اختيار طريقة التمثيل من قبل نوع من سلسلة الإحصائية (مفردة أو مزدوجة) وكذا نوع المتغير (النوعي، منفصلة الكمي أو مستمر) لتفي رغبات وأغراض الدراسة. حيث يمكن التمييز بين شكلين رئيسيين من الجداول الإحصائية، وهما الجدول البسيط: الذي يشمل بعد واحد للمتغير، والجدول المزدوج أو المتقطع الذي يبوب متغير واحد أو متغيرين أو أكثر حسب ظاهرة المدروسة. فحسب طبيعة المتغير كمي أو نوعي يمكن تصميم الجداول التكرارية التالية:

1-1- الجداول التكرارية للبيانات النوعية (Frequency distribution table for Qualitative or Categorical Data)

التوزيع التكراري لبيانات نوعية يتم برصد كل الفئات أو المجموعات (Categories) النوعية الممكنة في عمود ورصد عدد المفردات التي تتبع لكل فئة من تلك الفئات في عمود آخر وتدعى: **التكرار المطلق** n_i (Absolute Frequency) والذي يدل على عدد المرات التي ظهرت بها صفة معينة للمتغير في مجمل البيانات. مع مراعاة مجموع الأعداد الموجودة في عمود التكرارات يساوي عدد مفردات العينة أو مجموع التكرارات. .

$$N=\sum_{i=1} n_i = n_1 + n_2 + \cdots + n_k$$

المتغير		التكرار المطلق (n_i)
الصفة 1 : x_1		n_1
الصفة 2 : x_2		n_2
...		...
الصفة k : x_k		n_k
المجموع $\sum$		N

i. التوزيع التكراري النسبي والتوزيعات المئوية (Relative frequency and percentage distributions)

التكرار النسبي f_i لأي فئة أو صفة في الجدول التكراري هو حاصل قسمة تكرار تلك الفئة علي مجموع التكرارات الكلي. أي :

$f_i = \dfrac{n_i}{N}$ مع مراعاة مجموع التكرارات النسبية يساوي 1

$$\sum_i f_i = \sum_i \left(\frac{n_i}{N}\right) = \frac{n_1}{N}+\frac{n_2}{N}+\cdots+\frac{n_k}{N} = \frac{n_1+n_2+\cdots+n_k}{N} = 1$$

ويمثل التكرار النسبي المقدار الكسرى الذي تأخذه كل فئة في الجدول التكراري من المجموع الكلي للتكرارات. والقائمة التي تحتوي على كل التكرارات النسبية لكل الفئات في الجدول التكراري هي التوزيع التكراري النسبي. و لأغراض عملية يمكن حساب النسبة المئوية لأي فئة في الجدول التكراري والتي يتم الحصول عليها بضرب التكرار النسبي لتلك الفئة في 100، مع مراعاة مجموع التكرارات المئوية يساوي 100 %.

ii. التوزيع التكراري المتجمع الصاعد (Cumulative frequency distribution)

يكون الغرض في بعض الأحيان هو معرفة عدد أو نسبة المفردات التي تم تصنيفها في أكثر من فئة واحدة أو أقل من حد معين في جدول التوزيع التكراري. هذا يمكن باستخدام جدول التوزيع التكراري المتجمع الصاعد المطلق أو النسبي. التكرار المتجمع الصاعد لأي فئة هو مجموع تكرارات المفردات التي قيمها أقل من الحد الأعلي لتلك الفئة. التكرار المتجمع الصاعد النسبي لأي فئة هو نسبة تكرارات المفردات التي قيمها أقل من الحد الأعلي لتلك الفئة الى المجموع الكلي للتكرارات. ويحسب التكرار المجمع الصاعد لفئة معينة بتجميع تكرار الفئة وتكرارات القائمة السابقة وهو: $N_k \uparrow = \sum_{i=1}^{k} n_i$

وبالنسبة للتكرار النسبي فهو: $F_k \uparrow = \sum_{i=1}^{k} f_i$.

وقد يكون أحيأنا الاهتمام منصبا على عدد القيم أو النسبة التى تكون أكبر من أو تساوى قيمة معينة،لذا يحسب التكرارالمجمع النازل لفئة معينة وهذا بفصل تكرارها المجمع الصاعد عن مجموع التكرارات، وهو: $N_k \downarrow = N - N_k \uparrow = N - \sum_{i=1}^{k} n_i$ وبالنسبة للتكرار النازل النسبي فهو: $F_k \downarrow = 1 - F_k \uparrow = 1 - \sum_{i=1}^{k} f_i$

.

<u>مثال:</u>

سئل 20 شاب عن رأيهم في أداء مقدم البرنامج التلفزيونية الجديد. و كان على كل شاب أن يختار واحد من خمسة آراء هي: (1-ممتاز)، (2-فوق المتوسط)، (3-متوسط)، (4-تحت المتوسط)، و (5-ضعيف). فكان الإجابات على النحو التالي:

فوق المتوسط - تحت المتوسط - متوسط - متوسط - ممتاز - فوق المتوسط – ضعيف - ممتاز - فوق المتوسط - تحت المتوسط

ممتاز - فوق المتوسط - ممتاز - تحت المتوسط - ممتاز - متوسط - تحت المتوسط - ضعيف - فوق المتوسط - ممتاز

المطلوب تكوين جدول تكراري يشمل على التكرار، التكرار النسبي والمئوي والتكرار المتجمع الصاعد والتكرار المتجمع الصاعد النسبي وكذا حساب النسبة المئوية للشباب الذين كان رأيهم في أن الأداء كان أعلى من درجة متوسط (أي فوق المتوسط + ممتاز).

جدول التوزيع التكراري المطلق والنسبي والتكراري المتجمع الصاعد والتكراري المتجمع الصاعد النسبي لرأي الشباب في أداء المقدم الجديد البرنامج التلفزيوني.

الفئة	التكرار المطلق n_i	التكرار النسبي f_i	النسبة المئوية	التكرار المتجمع الصاعد $Ni\uparrow$	التكرار المتجمع الصاعد النسبي $Fi\uparrow$
1 (ممتاز)	6	6/20 = 0,30	0,30×100 = **30%**	6	0,30
2 (فوق المتوسط)	5	5/20 = 0,25	0,25×100 = **25%**	11	0,55
3 (متوسط)	3	3/20 = 0,15	0,15×100 = **15%**	14	0,70
4 (تحت المتوسط)	4	4/20 = 0,20	0,20×100 = **20%**	18	0,90
5 (ضعيف)	2	2/20 = 0,10	0,10×100 = **10%**	20	1,00
المجموع	**20**	**1,00**	**%100**		

iii. الجدول التكراري المزدوج أو جدول الاقتران (Contingency Table)

وهي تتناول التوزيع التكراري المشترك لمتغيرين joint distribution لأن المتغيران يقترنان فيه في توزيع مشترك. من هذا الجدول نستطيع استخراج الجدولين يعطى كل منهما ما يسمى بالتوزيع الهامشي للمتغير وهو يعنى التوزيع التكراري للمتغير الأول بصرف النظر عن المتغير الثاني.

1-2- الجداول التكرارية للبيانات الكمية (Frequency Distribution Table for Quantitative Data)

يتم تصميم جدول التوزيع التكراري للبيانات كمية حسب طريقتين:

i. رصد كل القيم الممكنة للمتغير وعدد مرات ظهورها بالنسبة للقيم الفردية للمتغير (غالبا في حالة المتغير الكمي المنفصل بعدد قليل من القيم)؛

ii. أو برصد الفئات أو المجموعات المكونة من القيم الممكنة للمتغير وعدد القيم التي تتبع لكل فئة أو مجموعة بالنسبة للقيم المتغير المجمعة في فئات (غالبا في حالة المتغير الكمي المستمر أو حالة المتغير الكمي المنفصل بعدد الكبير من القيم).

1-2-1- جدول التوزيع التكراري للقيم المفردة (Ungrouped frequency distribution table)

يتم بترتيب البيانات في جدول تكراري يعرض القيم المختلفة التي وردت في العينة وعدد المرات (التكرار) التي وردت فيها كل قيمة في تلك العينة وتستعمل غالبا حين تكون القيم التي يأخذها المتغير

قليلة ومنفصلة. حينها يستطيع الجدول أن نظهر كل القيم بصورة فردية في العمود الأول وفي العمود الثاني التكرار المقابل لكل قيمة.

قيم المتغير X	التكرار المطلق (n_i)
x_1	n_1
x_2	n_2
...	...
x_k	n_k
المجموع $\sum$	N

مثال:

أراد صاحب مكتبة تبيع الكتب الدراسية لطلاب الجامعة أن يحصر عدد الكتب التي يشتريها الطلاب في الفصل الأول من السنة الدراسية. قام صاحب المحل باختيار عينة عشوائية من 12 طالب وطالبة وسئل كل واحد منهم عن عدد الكتب التي اشتراها في الفصل الأول وكانت الإجابات كما يلي:

4	5	4	4	3	2
2	3	1	0	3	3

غرض صاحب المكتبة الأساسي هو تقدير القوة الشرائية للطالب الجامعي حتى يتمكن من توفير كميات معقولة من الكتب في مكتبته. ولكن يجب أن نضع في عين الاعتبار بأن هذه العينة صغيرة الحجم (12 مفردة) وقد لا تعكس صورة كاملة عن قوة الطلاب الشرائية داخل الجامعة وكذلك بفرضية تكافؤ النية الشرائية بين الذكور والإناث. وبالرغم من ذلك قد نخلص الى نتائج مفيدة. بالنظر الى البيانات في صورتها الحالية لا يمكننا أن نقول أي شيء عن قوة الطالب الجامعي الشرائية للكتب سوى أن معظم الطلاب يشترون أقل من 5 كتب. لكي تكون هذه البيانات الخام أكثر فائدة يجب أن يتم تنظيمها. ونلاحظ أن المتغير الذي ورد في هذه العينة هو عدد الكتب التي يشتريها الطالب في الفصل الأول وهو متغير كمي متقطع.

جدول توزيع تكراري للكتب المشترات من قبل الطلبة

عدد الكتب	التكرار المطلق (ni)
0	1
1	1
2	2
3	4
4	3
5	1
المجموع	**12**

من الملاحظ أنه بمجرد أن توضع البيانات الخام في جدول تكراري يصبح من السهل ملاحظة الوتيرة التي تظهر بها قيم المتغير (عدد الكتب). فمن السهل مثلا تحديد ما إذا كان هنالك عدد كبير من الطلاب لم يشتري أي كتاب أو اشتري أكثر من 4 كتب. أيضا نستطيع أن نحدد القوة الشرائية بالتقريب للطالب الجامعي بصوره عامة مثلا نلاحظ أن 4 طلاب من بين 12 طالب اشتروا أكثر من 3 كتب وكذا ربع (¼) الطلاب اقتنوا اقل من 3 كتب خلال الفصل الأول من السنة الدراسية.

1-2-2- جدول التوزيع التكراري للفئات (Grouped frequency distribution table)

حين تشتمل البيانات على عدد كبير من قيم متغير عددي (سواء كانت هذه القيم كمية متقطعة أو كمية متصلة)، يفضل تجميع هذه القيم في فئات. حيث يتم وضع القيم في فئات أو فترات منتظمة (متساوية الطول) أو مجموعات (Groups or Classes) حتى يسهل عرضها بصورة واضحة. ويراعى هنا ألا يكون عدد هذه الفئات كبيراً فتنتفي الحكمة أو الفائدة من عملية التجميع، وألا يكون عددها صغيراً فتضيع معالم التوزيع ويفقد الكثير من تفاصيله. ويلاحظ أن هذا الأسلوب في تكوين الفئات للقياسات المقربة يسمح بأن يكون لكل عدد في البيانات المعطاة مكان وحيد في إحدى الفئات. ولتصميم جدول تكراري يحتوي على عدد مناسب من الفئات يمكن اتباع الخطوات التالية:

i. **اختيار عدد الفئات** (Number of classes) : غالبا ما يتراوح عدد الفئات في التوزيع التكراري بين 5 الي 20 و يعتمد هذا على عدد المفردات في العينة. كما يستحسن الاستعانة بقاعدة ستورج (Sturge) أو قاعدة يول (Yule) لتحديد عدد الفئات مع التنبيه أنه يمكن إختيار عدد مناسب من الفئات دون التقيد بهاتين القاعدتين،

قاعدة ستورج (Sturge) $$k=1+(3{,}3\times log_{10}(N))$$

قاعدة يول (Yule) $$k=2{,}5\times\sqrt[4]{N}$$

حيث k يمثل عدد الفئات و N يمثل عدد البيانات أو المفردات الواردة في العينة.

ii. **تحديد عرض الفئة** (Class width) : يفضل أن يكون هنالك عرض (a) موحد ومنتظم لكل الفئات. لإيجاد عرض الفئة يجب في البداية أن نحسب المدى (E) الذي هو الفرق بين أصغر وأكبر قيمتين في البيانات أي ($E=(X_{Max}-X_{Min})$). حينها يتم إيجاد عرض الفئة بقسمة المدى على عدد الفئات المعتمد. مع مراعاة تقريب الناتج بالزيادة الى رقم صحيح مناسب.

$$a=\frac{E}{k}=\frac{\text{المدى}}{\text{عدد الفئات}}=\text{عرض الفئة}$$

iii. **تحديد الحد الأدنى للفئة الأولي** (Lower boundary of the first class): أي رقم مناسب أقل من أو يساوي أصغر قيمة في البيانات يمكن استعماله كحد أدنى للفئة الأولي. $X_{Min}\geq X_{1inf}$

مثال شامل:

أقيمت دراسة ميدانية كان الغرض منها تقصي عادة تدخين السجائر للعاملين في إحدى االمصانع. تم اختيار عينة عشوائية مكونة من 30 شخص يدخنون السجائر و سئل كل واحد منهم عن عدد قطع السجائر التي يستهلكها خلال فترة أسبوع فكانت النتائج كما يلي:

193	46	29	218	120	25	62	206	94	190
159	221	124	86	119	164	147	7	156	63
128	233	84	122	176	128	167	133	193	79

الغرض من ترتيب هذه البيانات في جدول تكراري هو الحصول علي فكرة عن توزيع القيم لهذا المتغير الكمي المتقطع (عدد قطع السجائر المستهلة خلال أسبوع). إذا وضعنا هذه البيانات في جدول تكراري بدون أن نكَون فئات للقيم كما في المثال (1) أعلاه سوف لن نجني أي فائدة تذكر لان معظم القيم الممكنة قد تظهر مرة واحدة فقط أو لا تظهر بالمرة. لكي نبسط هذه البيانات و نتمكن من إلقاء بعض الضوء علي التكوين الهيكلي للمفردات الواردة فإننا نحتاج الي وضع هذه البيانات في فئات .(Classes)

الحل:

i. باستخدام قانون استرج (Sturge) فإن العدد المناسب من الفئات هو $k=1+(3{,}3\times log_{10}(N))=1+3{,}3\times log_{10}(30)=5{,}95\cong 6$

ii. بالنظر الي البيانات نجد أن اصغر قيمة هي $X_{Min}=7$ وأكبر قيمة هي $X_{Max}=233$ ، إذن المدى $E=(X_{Max}-X_{Min})=233-7=226$ وعليه نقترح العرض المناسب للفئة $a=E/_k=226/_6=37{,}7\cong 38$

iii. بما أن أصغر قيمة في البيانات تساوي $X_{Min}=7$ يمكن استعمال أي قيمة أصغر أو تساوي 7 كحد أدنى للفئة الأولي. لنأخذ 6 كحد أدني للفئة الأولي.

إذن الفئات الممكنة للتوزيع التكراري تكون بصورة غير غامضة كما يلي:

] 234 ; 196] ;] 196 ; 158];] 158 ; 120];] 120 ; 82];] 82 ; 44];] 44 ; 6]

ما يؤخذ على عرض البيانات في جدول تكراري للفئات هو أننا سوف نفقد معلومات مفيدة فيما يتعلق بالمفردات. فمثلا لا يمكننا أن نحدد بالضبط من هو العامل الذي يستهلك 29 قطعة من السجائر في الأسبوع، كل ما نعرفه أن هنالك ثلاثة عمال يستهلكون بين 6 الى 44 قطعة سجائر في الأسبوع.

الجدول أدناه يبين التوزيع التكراري والتوزيع التكراري النسبي والتوزيع المئوي لقطع السجائر المستهلكة من المثال السابق.

جدول توزيع تكراري ونسبي ومئوي لعدد قطع التكرار المتجمع الصاعد والتكرار المتجمع الصاعد النسبي السجائر المستهلكة خلال الأسبوع

فئات إستهلاك السجائر	التكرار المطلق n_i	التكرار النسبي f_i	النسبة المئوية	التكرار المتجمع **الصاعد** النسبي $Fi\uparrow$	التكرار المتجمع **الصاعد** المطلق $Ni\uparrow$	التكرار المتجمع **النازل المطلق** $Ni\downarrow$
]6 ; 44]	3	3/30 = **0,100**	0,100×100 = **10,0%**	0,10	3	27
]44 ; 82]	4	4/30 = **0,133**	0,133×100 = **13,3%**	0,23	7	23
]82 ; 120]	4	4/30 = **0,133**	0,133×100 = **13,3%**	0,37	11	19
]120 ; 158]	8	8/30 = **0,267**	0,267×100 = **26,7%**	0,63	19	11
]158 ; 196]	7	7/30 = **0,233**	0,233×100 = **23,3%**	0,87	26	4
]196 ; 234]	4	4/30 = **0,133**	0,133×100 = **13,3%**	1,00	30	0
المجموع	**30**	**1**	**100%**			

2- التمثيل البياني للتوزيعات التكرارية

الرسوم البيانية (Diagrams) تعطي انطباعا أفضل وتبرز بنظرة سريعة الخصائص الرئيسية للبيانات وتلقي الضوء بصورة واضحة على شكل توزيع البيانات بعد تنظيمها لأنه قد نجد صعوبة في بعض الأحيان في قراءة الجداول التكرارية. لذا نتناول أهم طرق تمثيل البيانات على أساس أنها أكثر الطرق شيوعا. والأشكال الآتية تعرض أشهر هذه الصور وهي شكل:

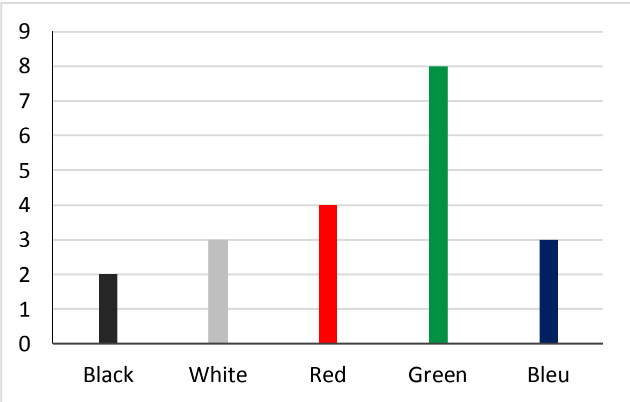
- ◊ أعمدة Bar Charts
- ◊ قطاعات دائرية Pie Charts
- ◊ مدرجات ومضلعات تكرارية Polygones and Histograms
- ◊ منحنيات تكرارية Curves Charts

ويختلف استعمال هذه الأشكال حسب طبيعة المتغير والتوزيع المرغوب تمثيله.

1-2- شكل الأعمدة البسيطة (Simple bar chart)

وهو الرسم بياني الملائم لتوزيع متغير كمي منفصل أو نوعي. هو عبارة عن رسم بياني يتكون من عدد من الأعمدة بحيث تمثل الفئات أفقيا على محور x-axis و تمثل التكرارات راسيا علي محور y-axis . يرسم عمود واحد لكل فئة بحيث يكون ارتفاع كل عمود يمثل التكرار أو التكرار النسبي أو التكرار المئوي المرتبط بكل فئة في الجدول التكراري. مع ملاحظة أن يكون عرض جميع الأعمدة متساوي كما بالإمكان استعمال اشكال تمثيلية بدلا من الاعمدة أو ألوان مختلفة للمتغير.

شكل الأعمدة للتوزيع التكراري للون المفضل للأشخاص

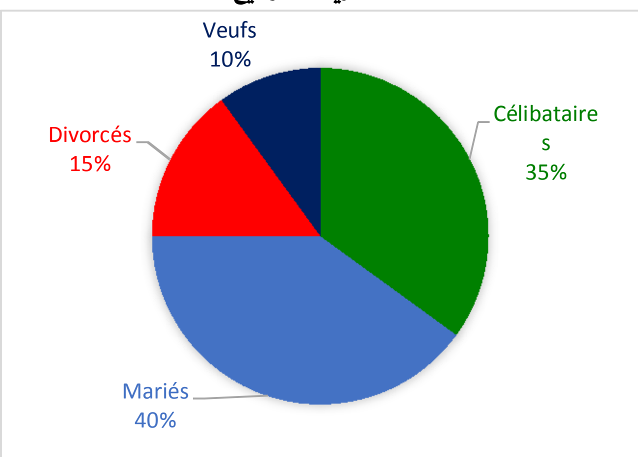


اللون المفضل للأشخاص	التكرار (n_i)
Black	2
White	3
Red	4
Green	8
Blue	3
المجموع	**20**

2-2- شكل الدائرة أو القطاعات الدائرية (Pie Charts)

الرسم البياني لتمثيل توزيع متغير نوعي أو كمي منفصل. هو عبارة عن قرص كاملا يتم تقسيمه الى أجزاء تتناسب مع التكرار النسبي أو المئوي لكل فئة من الفئات المكونة للعينة أو المجتمع. نتحصل على مقدار الزاوية المطلوبة للجزء لتمثيل كل فئة بضرب التكرار النسبي لكل فئة في 360.

شكل القطاعات الدائرية للتوزيع الحالة العائلية



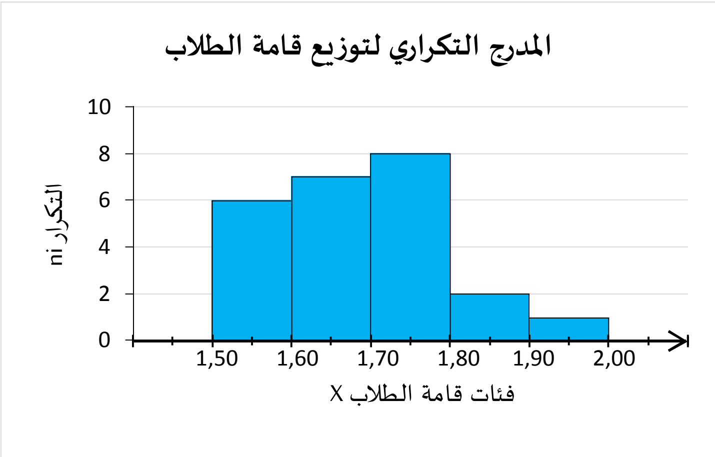
الحالة العائلية	التكرار النسبي f_i	مقدار الزاوية $f_i \times 360$
Célibataires	0,35	0,35x360 = **126°**
Mariés	0,40	0,40x360 = **144°**
Divorcés	0,15	0,15x360 = **54°**
Veufs	0,10	0,10x360 = **36°**
المجموع	**1**	**360°**

يمكن أن نشير على أن هناك عدة اشكال بيانية مخصصة من خلالها نستطيع تحويل الجداول الرقمية الى رسوم بيانية أكثر ملائمة للظواهر المدروسة نذكر منها الاعمدة البسيطة والاعمدة المجزأة التي تستعمل في حالة دراسة أكثر من متغير في والمقارنة بين المتغيرات. والخط البياني الذي يمثل اتجاه تطور البيانات في حالة الظواهر الزمنية.

2-3- المدرج التكراري (Histogram)

المدرج التكراري هو الرسم البياني المكرس للتوزيع التكراري الخاص بمتغير كمي مستمر وهو عبارة على رسم على محورين متعامدين أحدهما افقي يمثل الفئات والثاني رأسي يمثل التكرار ويتألف من عدد من المستطيلات المتلاصقة قواعدها طول فئات التوزيع وارتفاعاتها تتناسب مع التكرارات المناظرة لها.

فئات القامة الطلاب (م)	التكرار (n_i)
] 1,60 ; 1,50]	6
] 1,70 ; 1,60]	7
] 1,80 ; 1,70]	8
] 1,90 ; 1,80]	2
] 2,00 ; 1,90]	1



★المدرج التكراري في حالة الفئات الغير منتظمة:

في حالة الجداول التكرارية ذات فئات غير منتظمة تكون ارتفاعات مستطيلات المدرج التكراري غير متناسبة مع التكرار، لذلك يجب تعديل التكرار قبل رسم المدرج التكراري للفئات غير المتساوية حتى يصبح التكرار المعدل متناسبا مع ارتفاع المستطيل الخاص بالفئة غير منتظمة الطول ويكون بهذا المدرج التكراري صحيحا. وهناك طريقتان لتعديل التكرارات.

الطريقة الأولى: هي أن نقسم التكرار الأصلي لكل فئة على طولها فنحصل على تكرار معدل لجميع الفئات أي:

$$\text{التكرار المعدل} = \frac{\text{تكرار الفئة}}{\text{طول الفئة}} = \frac{n_i}{a_i}$$

الطريقة الثانية: فهي تتم بتعديل تكرار الفئات غير المنتظمة فقط، ويترك التكرار للفئات المنتظمة الباقية كما هو ويعدل التكرار للفئة غير المنتظمة بالعلاقة التالية:

$$\text{التكرار المعدل} = \frac{\text{تكرار الفئة الغير منتظمة} \times \text{طول الفئة منتظمة}}{\text{طول الفئة الغير منتظمة}} = \frac{a \times n_i}{a_i}$$

4-2- المضلع التكراري (Frequency polygon)

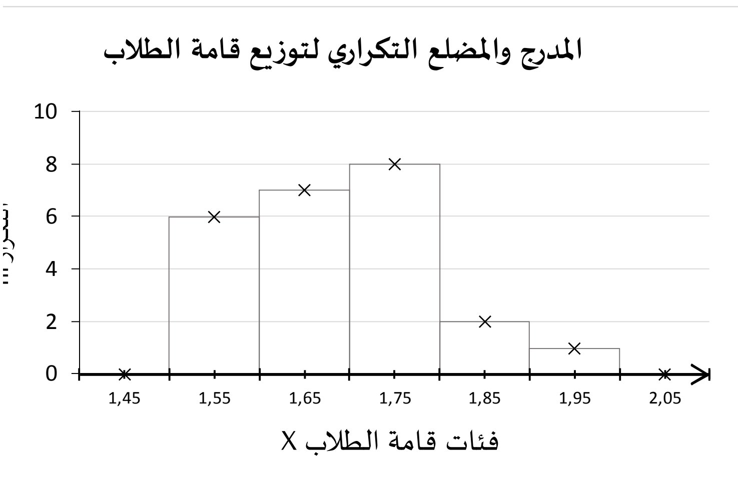
الغرض من المضلعات والمدرجات التكرارية هو إظهار ومقارنة خواص توزيع المتغيرات. حيث يتم رسم المضلع التكراري بتوصيل منتصف (مراكز) الفئات عند قمة كل مستطيل من المستطيلات المدرج التكراري (Histogram) عن طريق خطوط مستقيمة ثم يغلق المضلع من الطرفين وذلك بتصور وجود فئة إضافية في بداية التوزيع وفئة إضافية في آخره حيث التكرار في كل منهما يساوي صفر. ولرسم مضلع تكراري يجب:

أ. تحديد مراكز الفئات التي تساوي كل منها حاصل جمع الحد الأدنى والحد الأعلى لكل فئة مقسوما على 2، أو بمعنى آخر حساب متوسط الحد الأعلى والادنى لكل فئة $i \leftarrow \left]x_{inf} \quad ; \quad x_{sup}\right] \quad \therefore \quad x_i = \frac{(x_{sup} + x_{inf})}{2}$. ويؤخذ مركز الفئة بمعنى أننا نعتبر أنها قيمة تنوب جميع القيم التي بداخل الفئة،

ب. إيصال جميع مراكز الفئات ببعضها بواسطة خطوط مستقيمة عند قمم مستطيلات المدرج التكراري

ج. وأخيراً نغلق المضلع مع محور الفئات وهذا بوضع نقطة على محور الفئات يسار الحد الأدنى للفئة الأولى على بعد يساوي نصف طول الفئة، ثم نصل هذه النقطة بخط مستقيم بالنقطة المجاورة عند منتصف قمة مستطيل الفئة الأولى. وبطريقة مماثلة نضع نقطة على محور الفئات يمين الحد الأعلى للفئة الأخيرة تبعد مسافة نصف طول الفئة ثم نوصلها بالنقطة التي عند منتصف قمة مستطيل الفئة الأخيرة.

فئات القامة الطلاب (م)	مراكز الفئات (x_i)	التكرار (n_i)
] 1,60 ; 1,50]	1,55	6
] 1,70 ; 1,60]	1,65	7
] 1,80 ; 1,70]	1,75	8
] 1,90 ; 1,80]	1,85	2
] 2,00 ; 1,90]	1,95	1

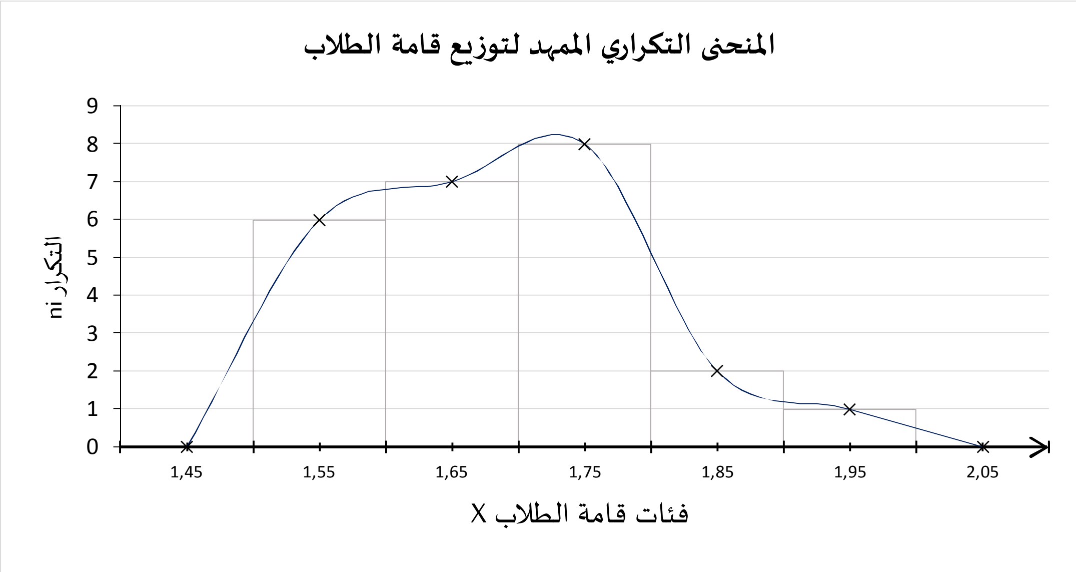


كما يحيل المضلع التكراري إلى ما يسمى بالمنحنى التكراري الممهد frequency curve وهو منحنى ناعم يمهد يدويا ماراً ببعض النقط وقريباً من البعض الآخر، لأن الهدف من رسمه هو محاولة استكشاف

الاتجاه العام لتوزيع المتغير في المجتمع الذى أخذت منه العينة ومن الواضح أن في عملية التمهيد كلما كبر حجم العينة وصغرت أطوال الفئات كلما اقترب المضلع التكراري من المنحنى التكراري.

★المنحنى التكراري الممهد (Smooth Frequency Curve)

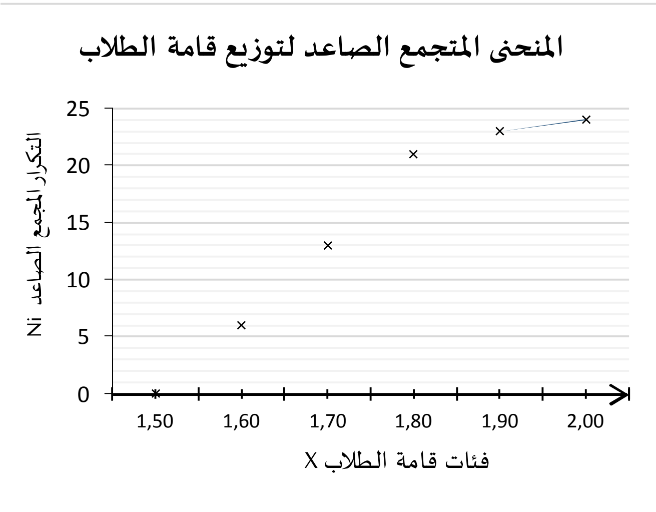
يتم رسم المنحنى التكراري الممهد بتمهيد المضلع التكراري ومحاولة المرور بأغلب نقاطه كي يأخذ شكلاً انسيابيًا.



5-2- المنحني المتجمع الصاعد (Ogive)

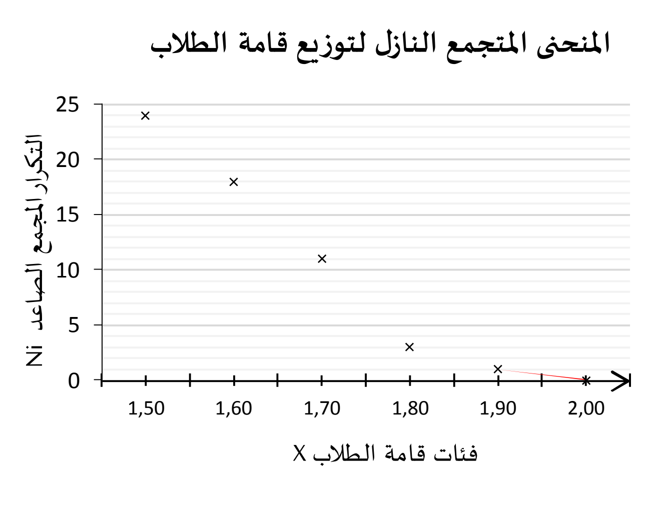
المنحنى المتجمع الصاعد (Ogive) هو منحني يمثل التوزيع المتجمع الصاعد (Cumulative frequency distribution). يتم ذلك أولا بتحديد التكرار المتجمع الصاعد أو النسبي المتجمع الصاعد على المحور الرأسي (y-axis) وثانيا برسم نقطة تمثل تقاطع الحد الأعلى للفئة على المحور الأفقي (x-axis) مع تكرارها المتجمع الصاعد على المحور الرأسي. أي أن الإحداثيات الأفقية للنقط هي الحدود العليا للفئات والإحداثيات الراسية هي التكرارات المتجمعة المطلقة أو النسبية أو المئوية المناظرة.

فئات القامة الطلاب (م)	التكرار (n_i)	التكرار المتجمع الصاعد $N_i^{\uparrow}$
] 1,60 ; 1,50]	6	6
] 1,70 ; 1,60]	7	13
] 1,80 ; 1,70]	8	21
] 1,90 ; 1,80]	2	23
] 2,00 ; 1,90]	1	24



2-6- المنحنى المتجمع النازل

فئات القامة الطلاب (م)	التكرار (n_i)	التكرار المتجمع النازل $N_i \downarrow$
] 1,60 ; 1,50]	6	24
] 1,70 ; 1,60]	7	18
] 1,80 ; 1,70]	8	11
] 1,90 ; 1,80]	2	3
] 2,00 ; 1,90]	1	1



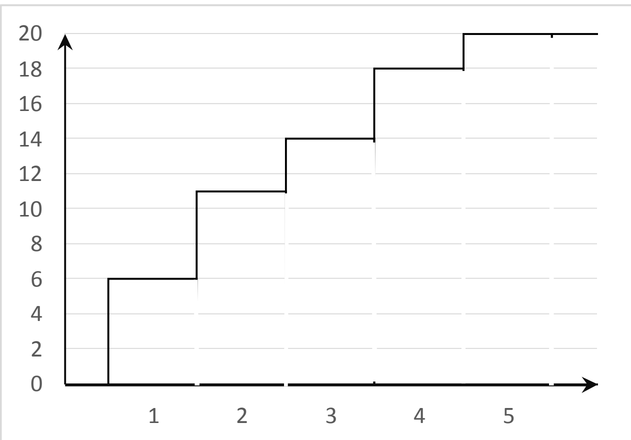
ومن المنحنيات التكرارات المتجمعة نستطيع الإجابة هندسيا إجابة تقريبية وبعمليات حسابية بسيطة عن بعض الأسئلة التي تتمثل في معرفة عدد أو نسبة المفردات التي لها قيم أقل من حد ما أو معرفة قيمة المتغير التي تمثل تكرار متجمع معين. كما نتمكن بنفس الطريقة أن نوجد تقريبا ما يسمى بالربيعات والمئينات وهي أعداد تستخدم في وصف التوزيعات والتي سوف نتطرق لها لاحقا في المقاييس الموضعية.

2-7- المنحنى السلمي (Ladder curve)

يستعمل ها النوع من الرسوم البيانية لعرض التكرار المتجمع الصاعد أو النسبي حين يتعلق الأمر بمتغير كمي منفصل (متقطع) مرتب في جدول تكراري لقيم فردية. وهو عبارة على منحنى بشكل سلم أين تعبر كل خطوة منه على التكرار التراكمي المقابل لقيمة المتغير.

جدول التوزيع التكراري المطلق والنسبي والتكراري المتجمع الصاعد والتكراري المتجمع الصاعد النسبي لرأي الشباب في أداء المقدم الجديد البرنامج التلفزيوني.

الفئة	التكرار المطلق n_i	التكرار المتجمع الصاعد $N_i\uparrow$
1 (ممتاز)	6	6
2 (فوق المتوسط)	5	11
3 (متوسط)	3	14
4 (تحت المتوسط)	4	18
5 (ضعيف)	2	20
المجموع	**20**	



تلخيص البيانات

The Data Summarizing

تمهيد

رأينا سابقا كيفية تنظيم وعرض البيانات الإحصائية في جداول تكرارية أو رسوم بيانية، بهدف الحصول جعلها أكثر طواعية للدراسة والتحليل وقد يكشف عن بعض الصفات البارزة أو الخصائص العامة تساعد على فهمها وتحليلها إحصائياً للتعرف عليها ووصفها ومقارنتها بغيرها من الظواهر. وقد يكون هذا العمل غير كافي وغير دقيق للخروج ببعض المدلولات الإحصائية عن مجتمع الدراسة، وعليه فإنه من الأفضل اللجوء إلى طرق القياس الكمي، حيث يستخدم الباحث الطريقة الرياضية في القياس. لذلك يجب أن يكون لدينا مقاييس عددية أكثر دقة تصف وتلخص لنا هذه البيانات التي تفسر توزيع لمتغير مدون في جدول تكراري. ولهذا نتعرض إلى نوع مهم من المقاييس الإحصائية تمكننا من وصف المتغير موضوعياً من حيث عدة جوانب أهمها:

1) النزعة المركزي--- Central Tendency

2) التشتت--- Dispersion

3) الالتواء--- Skewness

4) التفرطح--- Kurtosis

حيث نهتم بأهم مقاييس النزعة المركزية والتي تدرس تمركز توزيع المشاهدات من حولها والتي بالإمكان أن تلخصها وتنوب عنها. تليها دراسة مقاييس التشتت التي تستطلع عن مدى انحراف قيم المتغير بالنسبة الى بعضها وبالنسبة للقيم الوسطى وأخيرا نتطرق للمقاييس الشكلية تكميلا للدراسة الوصفية للبيانات والتي تأخذ بعدا مظهري لوصف توزيعاتها.

تعريف رمز التجميع $\sum$ والجدا $\prod$

يستحسن بداية التطرق لبعض العلاقات التي تستعمل رمز التجميع والجداء والتي قد تكون مفيدة في إثبات بعض الخصائص لبعض المقاييس المختلفة . كما تستخدم هذه الخصائص لتبسيط أو نشر عبارات تشمل رمز التجميع. وعليه إذا كان للمتغيرين x و y على التوالي المشاهدات $(x_1, x_2, \dots, x_n)$ والقيم $(y_1, y_2, \dots, y_k)$ على التوالي ومقدارين ثابتين حقيقيين a و b فإنه يمكن التعبير عن بعض أشكال حاصل جمع وجداء هذه المشاهدات كما يلي:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n \quad ; \quad \sum_{i=1}^{n} b = b + b + \cdots + b = n \times b \quad ; \quad \sum_{i=1}^{n} a x_i = a x_1 + a x_2 + \cdots + a x_n = a(x_1 + x_2 + \cdots + x_n) = a \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} (x_i \pm y_i) = (x_1 \pm y_1) + (x_2 \pm y_2) + \cdots + (x_n \pm y_n) = (x_1 + x_2 + \cdots + x_n) \pm (y_1 + y_2 + \cdots + y_n) = \sum_{i=1}^{n} x_i \pm \sum_{i=1}^{n} y_i$$

$$\prod_{i=1}^{n} x_i = x_1 \times x_2 \times \cdots \times x_n \quad ; \quad \prod_{i=1}^{n} b = b \times b \times \cdots \times b = b^n \quad ; \quad \prod_{i=1}^{n} a x_i = a x_1 \times a x_2 \times \cdots \times a x_n = a^n \prod_{i=1}^{n} x_i$$

1 - مقاييس النزعة المركزية MEASURES OF CENTRAL TENDENCY

وهو نوع مهم من المقاييس الإحصائية يسمى مقاييس الموضع أو المتوسطات. وهي مقاييس عددية تعين موقع التوزيع، وهي مهمة في حالة المقارنة بين التوزيعات المختلفة بوجه عام. وتكون فائدتها أكثر في حالة التوزيعات المتشابهة في طبيعتها وأشكالها ولكنها مختلفة في مواقعها. ونظرا الأن مثل هذه القيمة تميل إلى الوقوع في المركز داخل مجموعة البيانات لذلك لقبت هذه القيمة بالقيمة المتوسطة أو مقياس النزعة المركزية، والتي يمكن استخدامها لتمثيل مجموعة من البيانات، معتبرين بوجود عدة أسس لتحديد القيمة المتوسطة. ولما كان هذا المقياس يختلف في تركيبه بحسب طبيعة البيانات والهدف من دراستها فقد نجد عدة صور لهذه القيمة أهمها وأكثرها شيوعا الوسط الحسابي، والوسيط، والمنوال، والوسط الهندسي، والوسط التوافقي، ولكل منهم مزايا وعيوب وتعتمد على البيانات وعلى الهدف من الدراسة.

1-1- الوسط الحسابي $\overline{X}$ (Arithmetic Mean)

يعتبر الوسط الحسابي من أهم مقاييس النزعة المركزية والأكثر استخداما في الحياة العملية خاصة في المقارنة بين الظواهر المختلفة.

i. **طرق حساب الوسط الحسابي:**

★ **حالة البيانات غير المبوبة:** يتم حسابه كما يلي

$$\text{الوسط الحسابي} = \frac{\text{مجموع القيم}}{\text{عدد القيم}}$$

وتكون الصيغة الرياضية للوسط هي:

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

مثال: عينة ذات الحجم $n = 4$ قيم المتغير: 2 4 6 8 الوسط الحسابي

$$\overline{X} = \frac{\sum_{i=1}^{4} x_i}{n} = \frac{2+4+6+8}{4} = \frac{20}{4} = 5$$

★ **حالة البيانات المبوبة:** يتم حساب الوسط الحسابي لبيانات مبوبة بشكل جداول تكرارية على شكل فئات على أنه مجموع حاصل ضرب مركز كل فئة في التكرار المناظر له مقسوما على مجموع التكرارات، نذكر بأنه يتم إيجاد مراكز الفئات بحساب متوسط الحد الأعلى والادنى لكل فئة أي:لكل

فئة $i \leftarrow \left] x_{inf} \quad ; \quad x_{sup} \right[\quad \therefore \quad x_i = \frac{\left(x_{sup} + x_{inf}\right)}{2}$

وسوى استعملنا التكرار المطلق أو النسبي نعبر الوسط الحسابي بالصيغ التالية:

باستعمال التكرار المطلق

$$\overline{X} = \frac{\sum_{i=1}^{k} n_i x_i}{N} = \frac{n_1 x_1 + n_2 x_2 + \cdots + n_k x_k}{N}$$

أو باستعمال التكرار النسبي

$$\overline{X} = \sum_{i=1}^{k} f_i x_i = f_1 x_1 + f_2 x_2 + \cdots + f_k x_k$$

نرمز هنا لـ:

x_i = مركز الفئة i؛

n_i = التكرار المطلق للفئة i ؛ f_i= التكرار النسبي للفئة i ؛

N = مجموع التكرارات؛

k = عدد الفئات.

مثال (2): حساب متوسط أعمار البيانات التالية لتلاميذ النادي الموسيقي:

الفئات الأعمار	مراكز الفئات (x_i)	التكرار (n_i)	$n_i \times x_i$
] 6,5 ; 4,5]	5,5	2	11
] 8,5 ; 6,5]	7,5	5	37,5
] 10,5 ; 8,5]	9,5	8	76
] 12,5 ; 10,5]	11,5	4	46
] 14,5 ; 12,5]	13,5	1	13,5
	المجموع	**20**	**184**

الوسط الحسابي

$$\overline{X} = \frac{\sum_{i=1}^{5} n_i x_i}{N} = \frac{184}{20} = 9{,}2$$ سنة

مزايا الوسط الحسابي

السهل حساب ويتناسب للحسابات الجبرية

يأخذ في الاعتبار جميع القيم محل الدراسة

يلبي مبدأ المربعات الصغرى:

-أصغر خطأ

- موثوق حتى مع وجود n صغير

عيوب الوسط الحسابي

يتأثر بالقيم الشاذة (المتطرفة)

يشوه التوزيعات غير متجانسة

يصعب حسابه في حالة الجداول التكرارية المفتوحة

ii. بعض خصائص الوسط الحسابي:

سوف نعرض بعض خصائص الوسط الحسابي وذلك في حالة البيانات المباشرة أو البيانات المبوبة.

الخاصية الأولى :المجموع الجبري لانحرافات القيم عن وسطها الحسابي يساوى صفر.

$$\frac{\sum_{i=1}^{k} n_i (x_i - \overline{X})}{N} = 0$$

الإثبات:

$$\because \sum_{i=1}^{k} n_i = N \quad \cdots \quad (1)$$

$$\therefore \frac{\sum_{i=1}^{k} n_i(x_i - \bar{X})}{N} = \frac{\sum_{i=1}^{k} n_i x_i - \sum_{i=1}^{k} n_i \bar{X}}{N} = \frac{\sum_{i=1}^{k} n_i x_i}{N} - \frac{\sum_{i=1}^{k} n_i \bar{X}}{N} = \bar{X} - \frac{\bar{X}\left(\sum_{i=1}^{k} n_i\right)}{N} = \bar{X} - \bar{X} = 0 \quad \cdots \quad (2)$$

الخاصية 2: إذا كان للمتغير x المشاهدات $(x_1, x_2, \dots, x_k)$ فإذا أضفنا أو طرحنا من القيم الأصلية للمتغير مقدار ثابت b فإن الانحرافات الناتجة $(y_1, y_2, \dots, y_k)$ حيث: $y_i = x_i \pm b$

$i = 1, 2, \dots, k$

تعطى المتوسط كالتالي: $\bar{y} = \bar{x} \pm b$

الاثبات:

$$\equiv \quad y_i = x_i \pm b \qquad i = 1, 2, \dots, k$$

$$\Rightarrow \sum_{i=1}^{k} y_i = \sum_{i=1}^{k} (x_i \pm b) = \sum_{i=1}^{k} x_i \pm \sum_{i=1}^{k} b = \sum_{i=1}^{k} x_i \pm nb$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{k} y_i = \frac{1}{n}\sum_{i=1}^{k} x_i \pm b \qquad \Rightarrow \bar{y} = \bar{x} \pm b$$

الخاصية 3: إذا كان للمتغير x المشاهدات $(x_1, x_2, \dots, x_k)$ فإذا ضربنا هذه المشاهدات في مقدار ثابت حقيقي a للحصول على القيم $(y_1, y_2, \dots, y_k)$ حيث: $y_i = ax_i \cdots i = 1, 2, \dots, k$.

فإن متوسط القيم الجديدة $\bar{y}$ يساوى المتوسط $\bar{\ }$ مضروبا في a أي أن:

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

ومن الخاصيتين 2 و 3 نستنتج أن الوسط الحسابي له ميزة الاحتفاظ بالأشكال الخطية أي أن:

$$\therefore y_i = ax_i \pm b \quad \Rightarrow \bar{y} = a\bar{x} \pm b$$

الخاصية 4: إذا كان للمتغير x المشاهدات $(x_1, x_2, \dots, x_n)$ فإن مجموع مربعات انحرافات المشاهدات عن أي قيمة حقيقية c يكون أكبر أو يساوى مجموع مربعات انحرافات المشاهدات عن وسطها الحسابي، أي أن:

$$\sum_{i=1}^{n} (x_i - c)^2 \quad \geq \quad \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad , \bar{x} \neq c$$

الاثبات:

علما أن $$\sum_{i=1}^{n} (\bar{x} - c)^2 = (\bar{x} - c)^2 + (\bar{x} - c)^2 + \cdots + (\bar{x} - c)^2 = n \times (\bar{x} - c)^2$$

وكذلك من الخاصية الأولى لدينا $$\sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

وعليه

$$\sum_{i=1}^{n}\left(x_i-\bar{x}+\bar{x}-c\right)^2=\sum_{i=1}^{n}\left[\left(x_i-\bar{x}\right)+\left(\bar{x}-c\right)\right]^2=\sum_{i=1}^{n}\left[\left(x_i-\bar{x}\right)^2+2\left(x_i-\bar{x}\right)\left(\bar{x}-c\right)+\left(\bar{x}-c\right)^2\right]$$

فإن

$$\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(\bar{x}-c\right)+\sum_{i=1}^{n}\left(\bar{x}-c\right)^2=\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2+2\left(\bar{x}-c\right)\sum_{i=1}^{n}\left(x_i-\bar{x}\right)+\left(\bar{x}-c\right)^2\times n$$

إذا

$$\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2+n\times\left(\bar{x}-c\right)^2 \quad\Rightarrow\quad \sum_{i=1}^{n}\left(x_i-c\right)^2\geq\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2$$

يعني هذا أن الوسط الحسابي هو أقرب قيمة للبيانات من أي قيمة أخرى مهما كانت وهذا ما يجعل الوسط الحسابي القيمة الأكثر تمثيل ونيابة لمجمل الشاهدات.

1-2- الوسيط Me (Median)

وهو القيمة التي تقع في الوسط، وذلك بعد ترتيب القيم تصاعدياً أو تنازلياً وهو بذلك القيمة التي يكون عدد القيم الأصغر منها مساوياً لعدد القيم الأكبر منها.

طرق حساب الوسيط:

★ **حالة البيانات غير المبوبة:** يتم اتباع الخطوات التلية:

1. ترتيب قيم المتغير تصاعدياً

2. أ. إذ كان حجم العينة n عدد فردي (أي $n = 2k+1$).

الوسيط هو قيمة المتغير في الترتيب $(n+1)/2$ أي $Me = x_{\left(\frac{n+1}{2}\right)}$ أي المشاهدة التي تقع في المنتصف

ب. إذ كان حجم العينة n عدد زوجي (أي $n = 2k$).

الوسيط هو متوسط قيم المتغير في الرتبة $(n/2)$ و الرتبة $(n/2)+1$ وعليه

$Me = \dfrac{x_{(n/2)}+x_{(n/2)+1}}{2}$ أي هو متوسط المشاهدتين اللتين تقعان في المنتصف.

مثال: عينة من القيم

$n = 5$ القيم: 2 4 6 8 10 $\Leftarrow$ $Me = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{5+1}{2}\right)} = x_{(3)} = 6$

$n = 4$ القيم: 2 4 6 8

$$\Leftarrow \quad Me=\frac{x_{(n/2)}+x_{(n/2)+1}}{2}=\frac{x_{(4/2)}+x_{(4/2)+1}}{2}=\frac{x_{(2)}+x_{(3)}}{2}=\frac{4+6}{2}=5$$

★ **حالة البيانات المبوبة:** نستطيع حساب الوسيط حسابيا واجاده بيانيا .

i. **الوسيط حسابيا:** يتم اتباع الخطوات التالية لحساب الوسيط حسابيا

1 - نكون الجدول التكرارات المتجمع الصاعد.

2 - نوجد رتبة الوسيط $(n/2)$

3 - نحدد مكان الوسيط والتوقف عند أول فئة تضم تكرار صاعد أكبر أو يساوي نصف حجم العينة $(N/2)$ حيث يم تعينها بالفئة الوسيطة أين نحدد حدها الأدنى ونرمز له بالرمز L_{infMe}، وطولها c، ثم يعطى الوسيط بالعلاقة:

باستعمال التكرار المطلق
$$Me = L_{inf\,Me} + \frac{\left(\frac{N}{2} - N^{\uparrow}_{Me-1}\right)}{n_{Me}} \times a$$

أو باستعمال التكرار النسبي
$$Me = L_{inf\,Me} + \frac{\left(0{,}5 - F^{\uparrow}_{Me-1}\right)}{f_{Me}} \times a$$

أين يمثل: $N^{\uparrow}_{Me-1}$ التكرار المتجمع الصاعد (يمثل $F^{\uparrow}_{Me-1}$ التكرار النسبي المتجمع الصاعد) السابق للفئة الوسيطية؛

n_{Me} تكرار (f_{Me} التكرار النسبي) الفئة الوسيطية؛

N حجم العينة.

مثال : حساب وسيط لأعمار لتلاميذ النادي في المثال (2) السابق:

الفئات الأعمار	مراكز الفئات x_i	التكرار n_i	التكرار المتجمع الصاعد $N_i^{\uparrow}$
] 6,5 ; 4,5]	5,5	2	2
] 8,5 ; 6,5]	7,5	5	7
] 10,5 ; 8,5]	9,5	8	15
] 12,5 ; 10,5]	11,5	4	19
] 14,5 ; 12,5]	13,5	1	20
المجموع		**20**	

نحسب $(N/2)$ ونلاحظ أن أول فئة تضم تكرار صاعد أكبر أو يساوي $(N/2=20/2=10)$ حيث يتم تعينها الفئة الوسيطة وهي الفئة] 10,5 ; 8,5] وعليه فيكون:

$L_{infMe}=8{,}5$

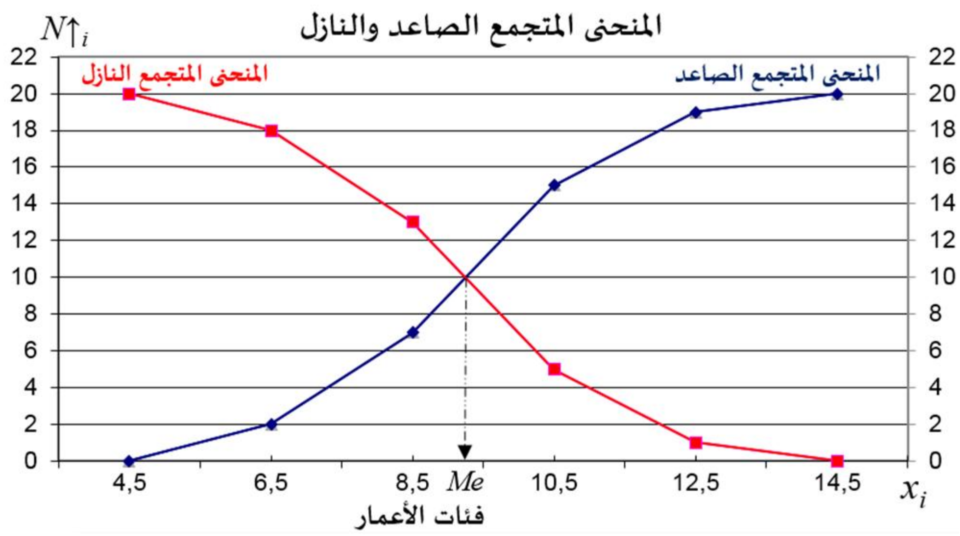
$N^{\uparrow}_{Me-1} = 7$

$n_{Me}=8$

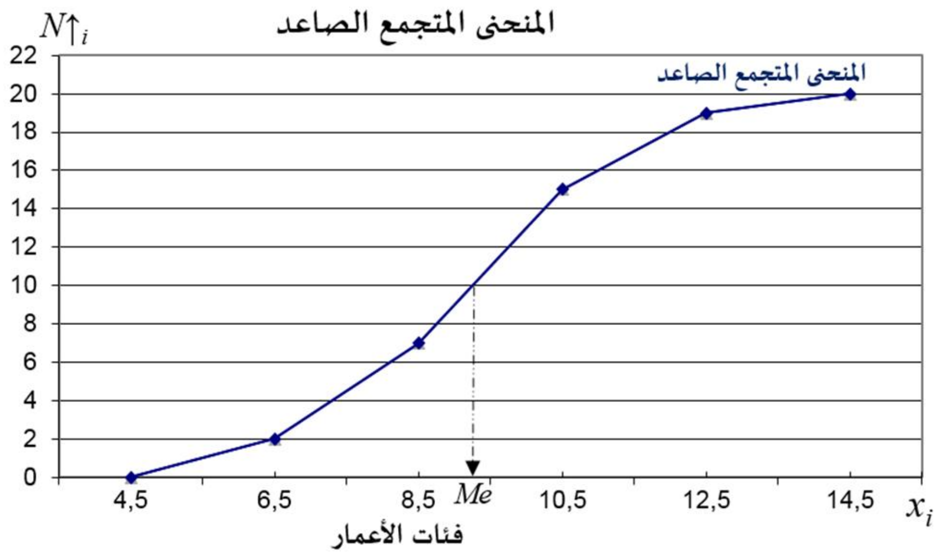
$a=10{,}5-8{,}5=2$

وبتطبيق القانون نحصل على الوسيط: $Me = L_{inf\,Me} + \frac{\left(\frac{N}{2} - N^{\uparrow}_{Me-1}\right)}{n_{Me}} \times a = 8{,}5 + \frac{(10-7)}{8} \times 2 = 9{,}25$ سنة

ii. **الوسيط بيانيا:** يمكن إيجاد الوسيط بيانيا والذي يمثل على المحور الأفقي مكان إسقاط نقطة تقاطع المنحنين الصاعد والنازل معا في رسم واحد. أو من المنحنى المتجمع الصاعد أو المنحنى المتجمع النازل كل على حده حيث على مستوى منتصف التكرارات $(N/2)$ نرسم خط أفقي حتى يمس المنحنى المتجمع ثم نسقط عمود على المحور الأفقي لنحصل على الوسيط مباشرة.



الشكل.1



الشكل.2

مزايا الوسيط	عيوب الوسيط
لا يتأثر بالقيم المتطرفة	لا يتناسب بشكل جيد للحساب
لا يعتمد على اختيار طول الفئة	لا يأخذ بعين الاعتبار جميع البيانات
صالح كذلك للمتغيرات الدورية والوصفية المرتبة	يتطلب التوزيع العادل للبيانات

1-3- المنوال Mo (Mode)

يعرف المنوال على أنه القيمة الأكثر شيوعا (تكرارا) أي أنه قيمة المفردة التي تتكرر أكثر من غيرها في مجموعة البيانات. وقد يكون لمجموعة البيانات منوال واحد أو يكون لها أكثر من منوال وتسمى متعددة المنوال، وقد لا يكون أي منوال وبذلك تسمى عديمة المنوال.

طرق حساب المنوال

i. المنوال حسابيا:

★ **حالة البيانات الغير مبوبة:** يحسب المنوال باستخدام التعريف مباشرة.

مثال : احسب المنوال من البيانات التالية: *2, 6, 9, 4, 6, 10, 6*

يوجد لهذه البيانات منوال واحد وهو القيمة *6* لأنها تكررت ثلاث مرات أكثر من غيرها.

★ **حالة البيانات المبوبة:** نتبع في حساب المنوال الخطوات التالية:

1 - نحدد فئة المنوال؛ وهي الفئة ذات أكبر تكرار.

2 - نستخدم العلاقة التالية: $Mo = L_{inf\ Mo} + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) \times a$

- L_{infMo} الحد الأدنى لفئة المنوال
- Δ_1 الفرق بين أكبر تكرار والسابق له
- Δ_2 الفرق بين أكبر تكرار واللاحق له
- a طول فئة المنوال.

مثال : حساب قيمة المنوال لأعمار تلاميذ النادي:

الفئات الأعمار	مراكز الفئات x_i	التكرار n_i
] 6,5 ; 4,5]	5,5	2
] 8,5 ; 6,5]	7,5	5
] 10,5 ; 8,5]	9,5	8
] 12,5 ; 10,5]	11,5	4
] 14,5 ; 12,5]	13,5	1
المجموع		**20**

نحدد الفئة ذات أكبر تكرار وهي فئة ذات التكرار يساوي 8 وهي الفئة] 10,5 ; 8,5] حيث يتم تعينها بفئة المنوال وعليه فيكون:

$L_{infMo}=8,5$

$\Delta_1 = 8 - 5 = 3$

$\Delta_2 = 8 - 4 = 4$

$a=10,5-8,5=2$

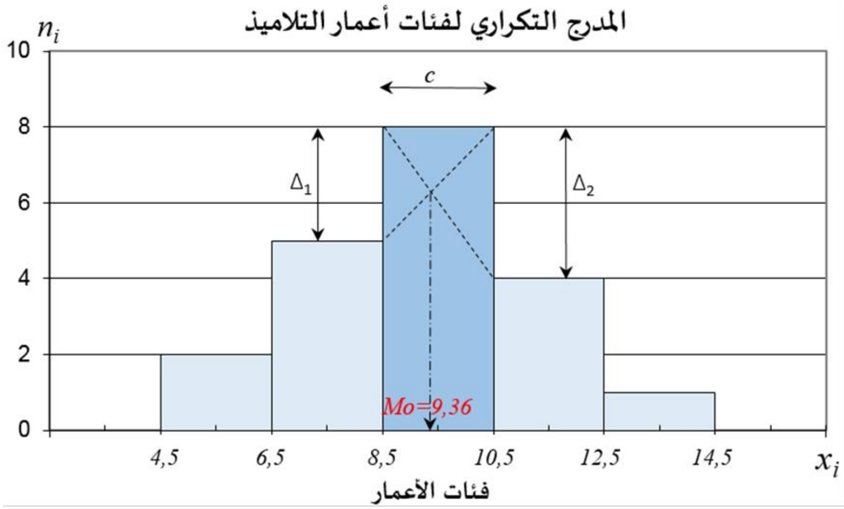
وبالتعويض في القانون ينتج أن المنوال يساوي:

$$Mo = L_{inf\ Mo} + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) \times a = 8,5 + \left(\frac{3}{3+4} \right) \times 2 = 8,5 + 0,86 = 9,36$$ سنة

ii. **المنوال بيانيا:** لإجاد المنوال بيانيا يمكن أن نتبع الخطوات التالية:

1- نرسم المدرج التكراري، ثم نحدد أعلى مستطيل والذي يمثل فئة المنوال (الفئة التي تقابل أكبر تكرار) وكذا مستطيل يمثل الفئة السابقة وآخر يمثل الفئة اللاحقة لها.

2- نصل الرأس الأيمن العلوي لمستطيل فئة المنوال بالرأس الأيمن العلوي للمستطيل السابق له، وكذلك الرأس الأيسر العلوي لمستطيل فئة المنوال بالرأس الأيسر للفئة اللاحقة وعند نقطة التقاطع نسقط خطاً عمودياً على المحور الأفقي للفئات، فتكون نقطة التقاطع هي قيمة المنوال.



ملاحظة:

في حالة عدم تساوي طول الفئات (أي الجداول التكرارية ذات فئات مختلفة الطول) يجب حساب الوسيط والمنوال باستخدام التكرارات المصححة التي تطرقنا إليها في الجزء الخاص بتنظيم وعرض البياني للبيانات.

مزايا المنوال	عيوب المنوال
سهل الحساب ولا يتأثر بالقيم الشاذة	لا يأخذ بعين الاعتبار جميع البيانات
يمكن حسابه للمتغيرات الوصفية والتوزيعات المفتوحة	قد يكون لبعض البيانات أكثر من منوال وبذلك لا يمكن تحديد قيمة وحيدة للمنوال.

★ العلاقة بين الوسط الحسابي والوسيط والمنوال

في التوزيع التكراري الاعتدالي وحيد المنوال تتحدد علاقة تجريبية بين مقاييس النزعة المركزية (الوسط الحسابي – الوسيط – المنوال) حيث تنطبق على بعضها وتتساوى جميعها. حيث: $Mo = Me = \overline{X}$ أي **الوسط الحسابي = الوسيط = المنوال**. كما تستعمل هذه العلاقة أيضا للتحقق من التماثل التام للتوزيع.

كما توجد علاقة اعتبارية تدعى العلاقة التجريبية لبيرسون Pearson empirical relationship بين مقاييس النزعة المركزية (الوسط الحسابي – الوسيط – المنوال) تتحقق في حالة المنحنيات التكرارية وحيدة المنوال وبسيطة الالتواء وهي: $(\overline{X} - Mo) = 3\times(\overline{X} - Me)$ من ما يؤدي بنا بمعنى آخر أن: **المنوال=3×الوسيط - 2×الوسط الحسابي** أي $Mo = (3\times Me) - (2\times \overline{X})$. ويظهر من خلال هذه العلاقة عن إمكانية حساب قيمة أي مقياس بدلالة المقاييس في حالة التوزيعات وضعيفة التماثل.

4-1- الوسط الهندسي X_G (Geometric Mean)

يمتاز الوسط الهندسي عن الوسط الحسابي بأنه أقل تأثرا بالقيم الشاذة في البيانات لأنه معلوم رياضيا وعليه فهو أقل حساسية من المتوسط الحسابي. وبالتالي، يعطي، تقديرا أفضل للنزعة المركزية في حالة السلاسل ذات توزيع طويل المدى. فهو يليق غالبا في حالت الظواهر ذات المنطق المضاعف وذات النمو الأسي. فمثلا يستخدم بكثرة في الاقتصاد القياسي والحسابات المالية لحساب متوسط معدل التغير ولإيجاد معدلات نسبية أو حساب متوسط الربحية السنوية للاستثمار على المدى الطويل. كما يستحيل حسابه بوجود قيم سالبة بين البيانات وفي هذه الحالة يستوجب القيام بحساب واستبدالها بمضاعفاتها العشرية المعادلة لها (مثلا بالنسبة لمعدل يساوي % 8.2- تعدل بمضاعفها العشري 8.2% - 100% = % 91.8-).

ويعرف رياضيا الوسط الهندسي X_G لمجموعة من القيم $(x_1, x_2, \ldots, x_n)$ على انه الجذر النوني لحاصل ضرب هذه القيم أي بطريقة مباشرة $\sqrt[n]{x_1 \times x_2 \cdots \times x_n} = X_G$. وعادة يحسب الوسط الهندسي باستخدام اللوغاريتمات كالتالي:

$$ln(X_G) = \frac{1}{n}\sum_{i=1}^{n} ln(x_i) = \frac{1}{n}\left(ln(x_1) + ln(x_2) + \cdots + ln(x_n)\right)$$

ويتم حسابه بالنسبة للبيانات المبوبة في جداول تكرارية بالصيغة:

$$X_G = \left(\prod_{i=1}^{k} (x_i)^n\right)^{\frac{1}{N}} = \sqrt[N]{x_1^{n_1} \times x_2^{n_2} \times \cdots \times x_k^{n_k}}$$

أو باستخدام الصيغة اللوغاريتمية

$$ln(X_G) = \frac{1}{N}\sum_{i=1}^{k}\left(n_i \times ln(x_i)\right) = \frac{1}{N}\left[n_1\, ln(x_1) + n_2\, ln(x_2) + \cdots + n_k\, ln(x_k)\right]$$

أين $N = \sum_{i=1}^{k} n_i$

مثال :

★حساب الوسط الهندسي للبيانات المباشرة:

$3, 5, 6, 6, 7, 10, 12$ وعلية فإن الوسط الهندسي يساوي

$$X_G = \sqrt[7]{3 \times 5 \times 6 \times 6 \times 7 \times 10 \times 12} = 6{,}43$$

★حساب الوسط الهندسي لبيانات مبوبة:

فئات الأعمار	مراكز الفئات x_i	التكرار n_i	لوغاريتمات القيم $Ln(x_i)$	لوغاريتمات القيم $n_i \times Ln(x_i)$
] 6,5 ; 4,5]	5,5	2	1,7047	3,4095
] 8,5 ; 6,5]	7,5	5	2,0149	10,0745
] 10,5 ; 8,5]	9,5	8	2,2513	18,0103
] 12,5 ; 10,5]	11,5	4	2,4423	9,7694
] 14,5 ; 12,5]	13,5	1	2,6027	2,6027
	المجموع	**20**		**43,8664**

نحسب بالصيغة اللوغاريتمية

$$ln(X_G) = \frac{1}{N}\sum_{i=1}^{k}(n_i \times ln(x_i))$$

$$ln(X_G) = \frac{1}{20} \times 43{,}8664 = 2{,}1933$$

$$\Rightarrow X_G = e^{ln(X_G)} = e^{2{,}1933}$$

$$\Rightarrow X_G = 8{,}9649$$

1-5- الوسط التوافقي X_H (Harmonic Mean)

يستخدم الوسط التوافقي عندما يكون مقلوب المتغير له دلاله كأن يعين نسبة بين متغيرين مرتبطين مثل السرعة بالنسبة للزمن. والاستعمال الأكثر شيوعا للوسط التوافقي يكمن في حالات معينة من حساب متوسطات النسب أو النسب المتوسطة وهو أكثر ملائمة للبيانات الكسرية بشكل الأرقام القياسية والاستدلالية التي يؤول بسط كسرها للتساوي. وفي بعض الأحيان النسب المتعلقة بمجال التمويل التي تمثل، نسب سعر للعائد والسعر/الربحية وكذا حسابات متوسط سرعات والتسارع والمسافات.

ويعرف والوسط التوافقي X_H لمجموعة من القيم $(x_1, x_2, \dots, x_n)$ بأنه مقلوب الوسط الحسابي لمقلوبات هذه القيم. أي أن:

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

والتي نستخلص منها صيغة الوسط التوافقي كالتالي:

$$X_H = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$$

وبالنسبة للبيانات المبوبة في جداول تكرارية يتم حسابه بالصيغة:

$$X_H = \frac{N}{\sum_{i=1}^{k}\frac{n_i}{x_i}} = \frac{N}{\frac{n_1}{x_1} + \frac{n_2}{x_2} + \cdots + \frac{n_k}{x_k}}$$

أين $N = \sum_{i=1}^{k} n_i$

مثال :

★ حساب الوسط التوافقي للبيانات المباشرة السابقة:

وعلية فإن الوسط التوافقي يساوي $3,\ 5,\ 6,\ 6,\ 7,\ 10,\ 12$

$$X_H = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{7}{\frac{1}{3}+\frac{1}{5}+\frac{1}{6}+\frac{1}{6}+\frac{1}{7}+\frac{1}{10}+\frac{1}{12}} = \frac{7}{\frac{501}{420}} = \frac{7\times 420}{501} = 5{,}87$$

★ حساب الوسط التوافقي لبيانات مبوبة:

فئات الأعمار	مراكز الفئات x_i	التكرار n_i	مقلوب القيم n_i/x_i
] 6,5 ; 4,5]	5,5	2	0,3636
] 8,5 ; 6,5]	7,5	5	0,6667
] 10,5 ; 8,5]	9,5	8	0,8421
] 12,5 ; 10,5]	11,5	4	0,3478
] 14,5 ; 12,5]	13,5	1	0,0741
المجموع		20	2,2943

نحسب بصيغة التوافق

$$X_H = \frac{N}{\sum_{i=1}^{k}\frac{n_i}{x_i}}$$

$$\Rightarrow X_H = \frac{20}{\sum_{i=1}^{5}\frac{n_i}{x_i}} = \frac{20}{2{,}2943}$$

$$\Rightarrow X_H = 8{,}7172 \approx 8{,}7$$

1-6- الوسط التربيعي X_Q (Quadratic Mean)

يشتق الوسط التربيعي من الأشكال التربيعية كالعزوم المستعملة بكثرة في مبرهنات نظرية الإحصاء الرياضي. ويعرف الوسط التربيعي X_Q لمجموعة من القيم $(x_1, x_2, \ldots, x_n)$ بأنه الجدر التربيعي لوسط حسابي مربعات هذه القيم. أي أن:

$$X_Q = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{\,2}\right)} = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}}$$

والتي يتم صياغتها في حالة البيانات المبوبة كالتالي:

$$X_Q = \sqrt{\frac{\sum_{i=1}^{k}\left(n_i \times (X_i^{\,2})\right)}{N}} = \sqrt{\frac{n_1x_1^2 + n_2x_2^2 + \cdots + n_kx_k^2}{N}}$$

أين $N = \sum_{i=1}^{k} n_i$.

مثال :

★ حساب الوسط التربيعي للبيانات المباشرة: *3, 5, 6, 6, 7, 10, 12*

وعلية فإن الوسط التربيعي يساوي

$$X_Q = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^{\,2}} = \sqrt{\frac{(3^2)+(5^2)+(6^2)+(6^2)+(7^2)+(10^2)+(12^2)}{7}} = \sqrt{\frac{399}{7}} = \sqrt{57} = 7{,}5498$$

★ حساب الوسط التربيعي لبيانات مبوبة:

فئة الأعمار	مراكز الفئات x_i	التكرار n_i	مربع القيم x^2_i	مربع القيم المرجح $n_i \times x^2_i$
] 6,5 ; 4,5]	5,5	2	30,25	60,5
] 8,5 ; 6,5]	7,5	5	56,25	281,25
] 10,5 ; 8,5]	9,5	8	90,25	722
] 12,5 ; 10,5]	11,5	4	132,25	529
] 14,5 ; 12,5]	13,5	1	182,25	182,25
المجموع		20		1775

نحسب الوسط التربيعي حسب الصيغة

$$X_Q = \sqrt{\frac{\sum_{i=1}^{k}\left(n_i \times (X_i^{\,2})\right)}{N}}$$

$$\Rightarrow X_Q = \sqrt{\frac{1775}{20}} = \sqrt{88{,}75} = 9{,}4207 \approx 9{,}4$$

ملاحظة هامة:

عادة ما تكون قيم المتوسطات المذكورة قريبة من بعضها البعض حيث يمكن في جميع الحالات التأكد من العلاقة التجريبية:

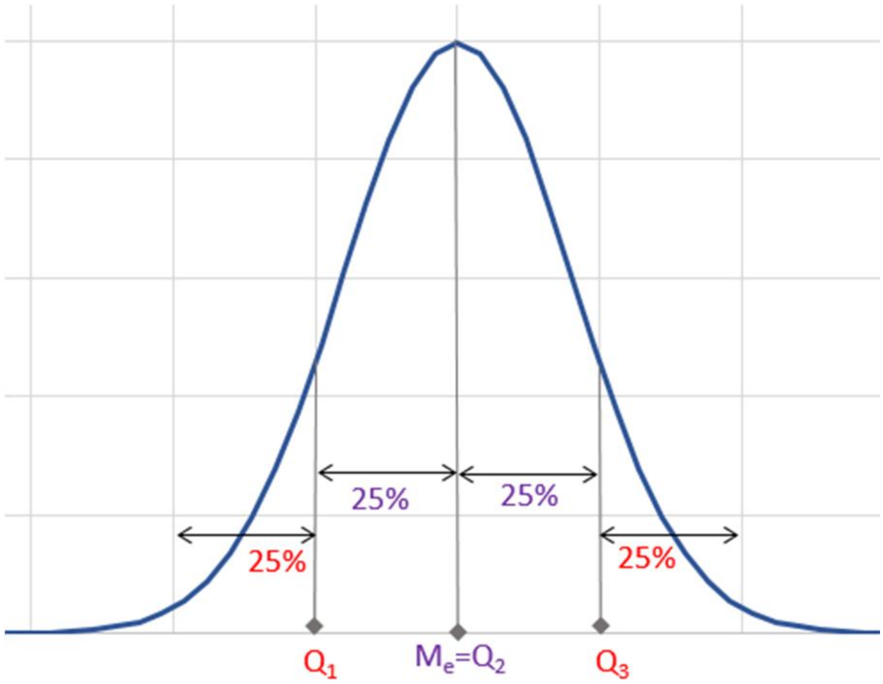
$$X_H \le X_G \le \overline{X} \le X_Q$$

1-7- الربيعات والعشيرات والمئينات (Quartiles, Deciles & Centiles)

بتعميم فكرة الوسيط الذي يكون في منتصف العينة والتي يقسمها إلى مجموعتين متساويتين في العدد أي القيمة التي تحدد 50 % من المشاهدات قبلها وبعدها. بنفس المنطق نستطيع تقسيم البيانات بعد ترتيبها إلى أربعة أجزاء متساوية يتحدد بين كل منها 25 % أي ربع المشاهدات ويرمز لها ب Q_3, Q_2, Q_1 حيث يسمى الربيع الأول Q_1 و Q_2 الربيع الثاني (الوسيط) و Q_3 يسمى الربيع الثالث. حيث بإمكاننا الحصول عل 50 % من المفردات وحصرهم بين قيم الربيعين Q_1 و Q_3. وعلى نفس المنوال يمكن إيجاد القيم التي تقسم البيانات بعد ترتيبها إلى عشرة أقسام متساوية يتحدد بين كل منها 10 % أي عشر المشاهدات ونرمز لها بالرمز $D_1, D_2, ..., D_9$. حيث D_1 يسمى العشير الأول وهو يمثل القيمة التي يسبقها عشر المشاهدات أي 10 % ويليها إذا 90 % من المشاهدات، ويسمى العشير الثاني D_2 وهو يمثل القيمة

التي يسبقها 20 % من المشاهدات ويليها 80 % منها وهكذا. وبنفس المنطق الحصول على المئينات التي تقسم البيانات إلى مائة قسم ذات حجم 1 % لكل منها، ويرمز لها ب $C_1, C_2, ..., C_{99}$.

يمكننا أن نبين منطق هذه المؤشرات في الرسم التالي:



ويعطى قانون حساب الربيعات والعشيرات والمئينات في الحالة العامة مثل قانون الوسيط على التوالي:

صيغة الربيعات مما يؤدي $M_e=Q_2$	$k=1,2,3$ $\therefore$	$Q_k = L_{inf\,Q_k} + \dfrac{\left(\frac{k\times N}{4} - N^{\uparrow}_{Q_{k-1}}\right)}{n_{Q_k}} \times a$
صيغة العشيرات مما يؤدي $M_e=D_5$	$k=1,2,\cdots,9$ $\therefore$	$D_k = L_{inf\,D_k} + \dfrac{\left(\frac{k\times N}{10} - N^{\uparrow}_{D_{k-1}}\right)}{n_{D_k}} \times a$
صيغة المئينات مما يؤدي $M_e=C_{50}$	$k=1,2,\cdots\cdots,99$ $\therefore$	$C_k = L_{inf\,D_k} + \dfrac{\left(\frac{k\times N}{100} - N^{\uparrow}_{C_{k-1}}\right)}{n_{C_k}} \times a$

مثال : في المثال السابق جد:

أ-القيمتين التي تتوسط سن التلاميذ والتي تضم 50 % من مجموع التلاميذ؛

ب – تحت أي سن نستحصل على 80 % من مجموع التلاميذ.

التكرار المتجمع الصاعد $N_i^{\uparrow}$	التكرار n_i	مراكز الفئات x_i	فئات الأعمار
2	2	5,5	] 6,5 ; 4,5]
7	5	7,5	] 8,5 ; 6,5]
15	8	9,5	] 10,5 ; 8,5]
19	4	11,5	] 12,5 ; 10,5]
20	1	13,5	] 14,5 ; 12,5]
	20	المجموع	

الربيع الأول $Q_1 = L_{inf\,Q_1} + \dfrac{\left(\frac{N}{4} - N_{Q1-1}^{\uparrow}\right)}{n_{Q_1}} \times a$

الربيع الثالث $Q_3 = L_{inf\,Q_3} + \dfrac{\left(\frac{3N}{4} - N_{Q3-1}^{\uparrow}\right)}{n_{Q_3}} \times a$

العشير الثامن $D_8 = L_{inf\,D_8} + \dfrac{\left(\frac{8\times N}{10} - N_{D_{8-1}}^{\uparrow}\right)}{n_{D_8}} \times a$

أ -القيمتين التي تتوسط مجموع المفردات والتي ينحصر بينها 50 % من المفردات هي قيم الربيعين Q_1 و Q_3.، وعليه فإن الجواب على السؤال يكون بتحديد هاتين القيمتين فقط.

نحدد الربيع الأول Q_1 بالوصول للربع الأول من التكرارات ($N/4=20/4=5$). ونلاحظ أن أول فئة تضم تكرار صاعد أكبر أو يساوي 5 هي الفئة التي تضم الربيع الأول Q_1 وهي الفئة] 6,5 ; 8,5] وعليه فيكون: $N_{Q1-1}^{\uparrow}=2$ $L_{infQ1}=6,5$
$n_{Q1}=5$ $a=8,5-6,5=2$

$$Q_1 = L_{inf\,Q_1} + \frac{\left(\frac{N}{4} - N_{Q1-1}^{\uparrow}\right)}{n_{Q_1}} \times a = 6,5 + \frac{(5-2)}{5} \times 2 = 6,5 + \frac{6}{5} = 7,7 = Q_1$$

نحدد الربيع الثالث Q_3 بالوصول ل ¾ من التكرارات المجمعة ($3xN/4=60/4=15$). ونلاحظ أن أول فئة تضم تكرار صاعد أكبر أو يساوي 15 هي الفئة التي تضم الربيع الثالث Q_3 وهي الفئة] 8,5 ; 10,5] وعليه فيكون: $N_{Q3-1}^{\uparrow}=7$
$L_{infQ3}=8,5$ $n_{Q3}=8$ $a=10,5-8,5=2$

$$Q_3 = L_{inf\,Q_3} + \frac{\left(\frac{3N}{4} - N_{Q3-1}^{\uparrow}\right)}{n_{Q_3}} \times a = 8,5 + \frac{(15-7)}{8} \times 2 = 8,5 + \frac{16}{8} = 10,5 = Q_3$$

ينحصر ال 50 % التي تتوسط مجموع التلاميذ بين السن 7,7 و 10,5 سنوات.

ب –نستطيع القول إن 80 % من التلاميذ لهم سن أقل أو يساوي القيمة التي نجمع فيها الثمان أعشار من مجموع التلاميذ وعليه نكتفي بحساب العشير الثامن D_8 لمعرفة هذا السن:

نحدد العشير الثامن D_8 بالوصول الى 80 % من التكرارات المجمعة التي يقدر عددها ($8xN/10=160/10=16$).

ونلاحظ أن أول فئة تضم تكرار صاعد أكبر أو يساوي 16 هي الفئة التي تضم العشير الثامن D_8 وهي الفئة ; 12,5] 10,5] وعليه فيكون: $N_{D8-1}^{\uparrow}=15$ $L_{infD8}=10,5$ $n_{D8}=4$ $a=12,5-10,5=2$

$$D_8 = L_{inf\,D_8} + \frac{\left(\frac{8\times N}{10} - N_{D_{8-1}}^{\uparrow}\right)}{n_{D_8}} \times a = 10,5 + \frac{(16-15)}{4} \times 2 = 10,5 + \frac{2}{4} = 11 = D_8$$

يقع 80 % من التلاميذ تحت عمر يساوي 11 سنة.

2 - مقاييس التشتت MEASURES OF DISPERSION

يقصد بالتشتت في أي مجموعة من القيم التباعد بين مفرداتها أو الاختلاف بينها، كما يدرس مدى تقارب البيانات عن بعضها البعض وبالنسبة الى قيمة وسطها. فكلما كانت قريبة من بعضها البعض أي قريبة من الوسط الحسابي تكون البيانات متجانسة، والعكس كلما كانت متباعدة تكون البيانات متباعدة أو مشتتة. وعلى ذلك يمكننا أن نتخذ مقدار التشتت كدليل على تجمع القيم أو على تفرقها وتباعدها عن بعضها. وهكذا يكون لدينا مقياس لمقدار تجانس المجموعات الإحصائية أو عدم تجانسها. يقاس التشتت بعدة مقاييس، ومن أشهرها: المدى – الانحراف الربيعي (نصف المدى الربيعي) – الانحراف المتوسط – التباين والانحراف المعياري – معامل الاختلاف.

2-1- المدى (Range)

يعرف المدى بأنه الفرق بين أكبر قيمة وأصغر قيمة في مجموعة البيانات. فهو بشكل عام سهل الحساب والفهم ويعطى فكرة سريعة عن طبيعة امتداد البيانات، هذا ما يجعله يستخدم كثيرا في مراقبة الجودة ومختصرات الأحوال الجوية. عموما يتم حسابه بالصيغة:

$$\text{المدى } R = X_{Max} - X_{Min}$$

فهو مقياس غير ثابت يعتمد في حسابه على القيمتين المتطرفة فقط من البيانات.

2-2- نصف المدى الربيعي (Interquartile Range)

نعرف على القيم التي تقسم البيانات إلى أربعة أقسام متساوية بالربيعيات والتي رمزت ب Q_3, Q_2, Q_1.

يستخدم أحيانا المقدار الذي يمثل النصف الأوسط للقيم $|Q_3 - Q_1|$ المسمى بالانحراف الربيعي (quartile deviation) كمقياس تشتت اصح من المدى. وبشكل أكثر دقة ويؤخذ نصف هذا المدى مقياسا للتشتت ويسمى بنصف المدى الربيعي (interquartile range) ويرمز له بالرمز Q هو متوسط الفرق بين الربيعيين الثالث والأول ويحسب من الصيغة الرياضية:

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

يتم تطبيق هذا القانون لإيجاد قيمة نصف المدى الربيعي، مباشرة بعد ترتب البيانات في حالة البيانات الغير مبوبة أو بتكوين جدول التكرار المتجمع الصاعد بالنسبة للبيانات مبوبة وتحديد قيم Q_1 و Q_3.

3-2- الانحراف المتوسط AD (Average Deviation)

يعرف الانحراف المتوسط بأنه متوسط الانحرافات المطلقة للبيانات عن وسطها الحسابي $\overline{}$ ، أي انه القيمة المطلقة (الناتج الموجب) لمجموع انحرافات القيم عن وسطها الحسابي مقسوما على عددها.
ويعرف رياضيا بالقانون التالي:

- بالنسبة للبيانات المباشرة $AD = \frac{\sum_{i=1}^{n}|x_i - \overline{X}|}{n} = \frac{|x_1 - \overline{X}| + |x_2 - \overline{X}| + \cdots + |x_n - \overline{X}|}{n}$ ،
- أو بالنسبة للبيانات المبوبة $AD = \frac{\sum_{i=1}^{k} n_i |x_i - \overline{X}|}{N} = \frac{(n_i \times |x_i - \overline{X}|) + (n_2 \times |x_2 - \overline{X}|) + \cdots + (n_k \times |x_k - \overline{X}|)}{N}$

استعملت القيم المطلقة للانحرافات لأن مجموع انحرافات القيم عن وسطها الحسابي يساوى صفر (من خصائص الوسط الحسابي). كما يعرف الانحراف المتوسط، أحيانا، باستخدام الوسيط بدلا من الوسط الحسابي أو أي من المتوسطات الأخرى. والانحراف 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.

4-2- التباين والانحراف المعياري σ (Variance and Standard-Deviation)

يعرف الانحراف المعياري على أنه أدق مقاييس التشتت وأكثرها استخداما، ويعرف بأنه الجذر التربيعي لمتوسط مربعات انحرافات (فروق) القيم عن وسطها الحسابي، ويرمز له بالرمز σ. وهو يعرف غالبا عن طريق التباين σ^2 حيث يفضل عند حساب الانحراف المعياري أن يحسب التباين ويأخذ الجذر التربيعي للنتيجة النهائية لنحصل على المعادلة العامة:

$$\text{الانحراف المعياري} = \sqrt{\text{التباين}}$$

i. الانحراف المعياري:

ومن الملاحظ أن التباين يتعامل مع مربع الانحراف عن الوسط أي يقاس بالوحدات المربعة وليس بوحدات المتغير وهذا يعطي قياس غير ذو معنى لذا يفضل إرجاع ذلك (بأخذ الجذر التربيعي) للمعنى المقبول. لذا نقول ببساطة بما إن الانحراف المعياري هو الجذر التربيعي للتباين، فهو يقاس بنفس وحدات المتغير محل ظاهرة الدراسة. فهو أفضل مقاييس التشتت وأشهرها استخداماً بالرغم من صعوبة حسابه في حال كبر حجم العينة.

ii. طريقة حساب الانحراف المعياري:

★ **حالة البيانات غير المبوبة:** إذا كان للمتغير x المشاهدات $(x_1, x_2, ..., x_n)$ ذات المتوسط الحسابي $\overline{X}$.

يمكن تعريف التباين σ^2 بالقانون الآتي:
$$\sigma^2 = \frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}{n}$$

ويصاغ حينها الانحراف المعياري σ كالتالي:
$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}{n}}$$

ونظرا لخصائص هذه الأشكال التربيعية، يمكننا تبسيط صيغة التباين σ^2 كالتالي :

$$\sigma^2 = \frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}{n} = \frac{\sum_{i=1}^{n}\left(x_i^2 - 2\overline{X}x_i + \overline{X}^2\right)}{n} = \frac{\sum_{i=1}^{n} x_i^2 - 2\overline{X}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}\overline{X}^2}{n}$$

علما أن $\overline{X}$ ثابت

وكذا أن $\sum_{i=1}^{n} x_i = n\overline{X}$ ،

$$\therefore \qquad \sigma^2 = \frac{\sum_{i=1}^{n} x_i^2 - \left(2\times\overline{X}\times n\times\overline{X}\right) + n\times\overline{X}^2}{n} = \frac{\sum_{i=1}^{n} x_i^2 - \left(2n\times\overline{X}^2\right) + n\times\overline{X}^2}{n}$$

فإذا:

$$\Rightarrow \sigma^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\times\overline{X}^2}{n} = \frac{\sum_{i=1}^{n} x_i^2}{n} - \overline{X}^2$$

مما يمكننا من الحصول على الصيغة المبسطة للانحراف المعياري σ حيث يصبح بالشكل:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \overline{X}^2}$$

حيث : $\sum_{i=1}^{n} x_i^2$: مجموع مربعات قيم مفردات العينة، n: حجم العينة

تحذير:

$$\sum_{i=1}^{k} x_i^2 \neq \left(\sum_{i=1}^{k} x_i\right)^2$$

مثال: احسب التباين والانحراف المعياري للأجور اليومية بالدولار US$ للعينة التالية المتكونة من خمس عمال: 50 70 80 90 60 .

الحل:

$$\overline{X}=\frac{\sum_{i=1}^{5}x_i}{5}=\frac{60+90+80+70+50}{5}=\frac{350}{5}=70\$$$

$$\frac{\sum_{i=1}^{5}(x_i)^2}{5}=\frac{60^2+90^2+80^2+70^2+50^2}{5}=5100$$

تكون قيم التباين والانحراف المعياري للأجور على التوالي:

$$\sigma^2=\frac{\sum_{i=1}^{n}x_i^2}{n}-\overline{X}^2=5100-4900=200 \quad\Rightarrow\quad \sigma=\sqrt{200}=14{,}14\$$$

★ **حالة البيانات المبوبة:** في حالة البيانات المنظمة في جدول تكراري يمكننا حساب التباين والانحراف المعياري باستخدام إحدى الصيغ الآتية:

$$\sigma^2=\frac{\sum_{i=1}^{k}n_i(x_i^2)}{N}-\overline{X}^2 \quad\Leftrightarrow\quad \sigma^2=\frac{\sum_{i=1}^{k}n_i\left(x_i-\overline{X}\right)^2}{N}$$

وكما نعرف أن الانحراف المعياري هو جدر التباين. فإن:

$$\sigma=\sqrt{\frac{\sum_{i=1}^{k}n_i(x_i^2)}{N}-\overline{X}^2} \quad\Leftrightarrow\quad \sigma=\sqrt{\frac{\sum_{i=1}^{k}n_i\left(x_i-\overline{X}\right)^2}{N}}$$

أين: n_i: التكرار

x_i: مراكز الفئات

N: مجموع التكرارات $N=\sum_{i=1}^{k}n_i$

$\overline{X}$: الوسط الحسابي.

مثال (2): حساب الانحراف المعياري لأعمار التلاميذ للبيانات التالية:

فئة الأعمار	مراكز الفئات (x_i)	التكرار (n_i)	$n_i \times x_i$	x^2_i	$n_i \times x^2_i$
] 6,5 ; 4,5]	5,5	2	11	30,25	60,5
] 8,5 ; 6,5]	7,5	5	37,5	56,25	281,25
] 10,5 ; 8,5]	9,5	8	76	90,25	722
] 12,5 ; 10,5]	11,5	4	46	132,25	529
] 14,5 ; 12,5]	13,5	1	13,5	182,25	182,25
	المجموع	**20**	**184**		**1775**

الوسط الحسابي

$$\overline{X}=\frac{184}{20}=9{,}2 \Rightarrow \overline{X}^2=84{,}64$$

الانحراف المعياري

$$\sigma=\sqrt{\frac{\sum_{i=1}^{k}n_i(x_i^2)}{N}-\overline{X}^2}=\sqrt{\frac{1775}{20}-84{,}64}$$

$$\Rightarrow \sigma=\sqrt{4{,}11}=2{,}0273$$

iii. خصائص الانحراف المعياري

سوف نتعرض لبعض خصائص التباين والتي تنتسب أيضا للانحراف المعياري وذلك في حالة البيانات المباشرة أو البيانات المبوبة.

الخاصية الأولى المجموع الجبري لمربعات انحرافات القيم عن وسطها الحسابي $\overline{X}$ تكون أصغر من مجموع مربعات انحرافات القيم عن أي وسط فرضي أخر a حيث $c\neq\overline{X}$. :

$$\sum_{i=1}^{k} n_i(x_i - c)^2 > \sum_{i=1}^{k} n_i(x_i - \overline{X})^2 \qquad \forall c \neq \overline{X}$$

الإثبات:

$$\sum_{i=1}^{k} n_i(x_i - c)^2 = \sum_{i=1}^{k} n_i(x_i - \overline{X} + \overline{X} - c)^2 = \sum_{i=1}^{k} n_i\left[(x_i - \overline{X}) + (\overline{X} - c)\right]^2$$

$$= \sum_{i=1}^{k} n_i(x_i - \overline{X})^2 + 2(\overline{X} - c)\sum_{i=1}^{k} n_i(x_i - \overline{X}) + \sum_{i=1}^{k} n_i(\overline{X} - c)^2$$

$$= \sum_{i=1}^{k} n_i(x_i - \overline{X})^2 + 2(\overline{X} - c)\sum_{i=1}^{k} n_i(x_i - \overline{X}) + (\overline{X} - c)^2\sum_{i=1}^{k} n_i$$

بما أن $\sum_{i=1}^{k} n_i(x_i - \overline{X}) = 0$

و $\sum_{i=1}^{k} n_i = N$

$$\Rightarrow \sum_{i=1}^{k} n_i(x_i - c)^2 = \sum_{i=1}^{k} n_i(x_i - \overline{X})^2 + N(\overline{X} - c)^2$$

وعلما ان $N(\overline{X} - c)^2 > 0$ فإن

$$\Rightarrow \sum_{i=1}^{k} n_i(x_i - c)^2 > \sum_{i=1}^{k} n_i(x_i - \overline{X})^2$$

الخاصية 2: إذا كان للمتغير x المشاهدات $(x_1, x_2, ..., x_k)$ فإذا أضفنا أو طرحنا من القيم الأصلية للمتغير مقدار ثابت b فإن الانحرافات الناتجة $(y_1, y_2, ..., y_k)$ حيث: $y_i = x_i \pm b$ $\quad i = 1, 2, ..., k$

تعطى التباين كالتالي: $y = x \pm b \quad \Leftarrow \quad \sigma_y^2 = \sigma_x^2$

ويمكن أن تستخدم هذه الخاصية في تبسيط البيانات، وخاصة عندما تكون قيمتها كبيرة.

الخاصية 3: إذا كان للمتغير x المشاهدات $(x_1, x_2, ..., x_k)$ فإذا ضربنا هذه المشاهدات في مقدار ثابت حقيقي a للحصول على القيم $(y_1, y_2, ..., y_k)$ حيث: $y_i = ax_i$ $\ldots i = 1, 2, ..., k$.

فإن تباين القيم الجديدة σ_y^2 يساوى تباين σ_x^2 مضروبا في a^2 أي أن:

$$\sigma_y^2 = a^2 \times \sigma_x^2 \quad \Leftarrow \quad y = ax$$

أي أن تباين للقيم الجديدة في حالة الضرب يساوى جداء مربع المقدار الثابت مع تباين القيم الأصلية.

نستنتج من الخاصيتين 2 و 3 أن التباين والانحراف المعياري لا يحتفظان بالأشكال الخطية لتوزيعات المتغيرات أي أن:

$$\therefore \forall\, y_i = ax_i \pm b \quad \Rightarrow \sigma_y^2 = a^2 \times \sigma_x^2$$

الخاصية 4: نستطيع التأكد بسهولة أن الانحراف المعياري لمجموعة من البيانات أكبر من الانحراف المتوسط لها.

$$\frac{\sum_{i=1}^{k} n_i\left|x_i-\overline{X}\right|}{N} \leq \frac{\sum_{i=1}^{k} n_i\left(x_i-\overline{X}\right)^2}{N}$$

2-5- معامل الاختلاف CV (Coefficient of variation)

هو معامل نسبي يستخدم للمقارنة بين تشتت ظاهرتين أو أكثر، فكلما كان معامل الاختلاف أكبر كلما ازدادت درجت التشتت. وهو يستعمل غالبا في البيانات ذات وحدات القياس المختلفة لأنه يعبر بقيم مأوية، ويستعان في حسابه بالعلاقة التالية:

$$cv_x = \frac{\sigma_x}{\overline{X}} \times 100$$

أين : σ_x : الانحراف المعياري

$\overline{X}$: الوسط الحسابي

تنبيه: في حالة العينات الصغيرة الحجم ($N \leq 20$) يستحسن تطبيق تصحيح لمعامل الاختلاف ليصبح:

$$CV_x^* = \left(1+\frac{1}{4N}\right)CV_x$$

مثال:

حساب معامل الاختلاف للأجور اليومية بالدولار US$ للعينة التالية المتكونة من خمس عمال: 50 70 80 90 60 .

الحل:

قيم المتوسط الحسابي والانحراف المعياري للأجور على التوالي: $\overline{X} = 70\$$ $\therefore$ $\sigma = 14{,}14\$$

بما ان $N = 5$ نستخدم معامل الاختلاف المصحح

وعليه يكون معامل الاختلاف للأجور يساوي:

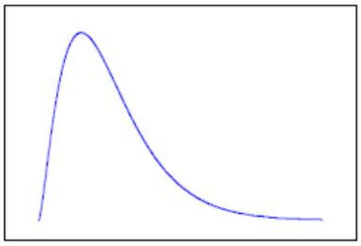
$$cv_x = \frac{\sigma_x}{\overline{X}} \times 100 = \frac{14{,}14}{70} \times 100 = 20{,}2\%$$

$$cv_x^* = \left(1+\frac{1}{4N}\right) \times 20{,}2\% = \left(1+\frac{1}{4\times 5}\right) \times 0{,}202 = 21{,}21\%$$

3 - مقاييس الالتواء MEASURES OF SKEWNESS

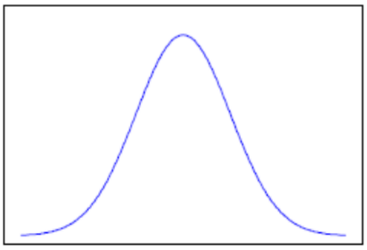
ويقصد بالالتواء بعد المنحنى التكراري عن التماثل، والتماثل يعني كون قسمة المنحنى التكراري عموديا من المنتصف تعطي قسمين منطبقين تماما. والعكس أي عدم التماثل يعني عدم تطابق القسمين ويكون حينها التوزيع ملتوي إما يمينا او يسارا. وشكليا يكون منحنى التوزيع التكراري ملتويا نحو اليمين إذا كانت القيم المتطرفة نحو اليمين تؤثر على الوسط الحسابي وتتجه به نحو اليمين، وبذلك يكون الوسط الحسابي أكبر من الوسيط. أما إذا كان التوزيع ملتويا نحو اليسار فإن الوسط الحسابي يكون أصغر من الوسيط. ونستطيع توضيح هذه الحالات كما يلي:

- توزيع ملتو لليمين: $Mo < Me < \overline{X}$
- توزيع متماثل: $Mo = Me = \overline{X}$
- توزيع ملتو لليسار: $Mo > Me > \overline{X}$



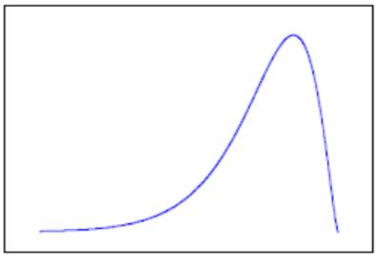
التوزيع غير متماثل وملتو من جهة اليمين

$Mo < Me < \overline{X}$



التوزيع متماثل

$\overline{X} = Me = Mo$



التوزيع غير متماثل وملتو من جهة اليسار

$\overline{X} < Me < Mo$

وغالبا ما يتيح تحليل شكل الاعمدة -او المدرج التكراري -من إدراك وتوضيح طبيعة تماثل او عدم تماثل توزيع ما. كما نستعين بمقاييس الالتواء والتي تقيس درجة التواء المنحنى التكراري وعدم التوازن في توزيع. والهدف هو مقارنة بين أشكال التوزيعات، وتكون هذه المقارنات ذات دلالة إذا إجرية بنفس المعاملات حتى ولو تم تطبيقها على توزيعات مختلفة، أي عند مقارنة التواء توزيعات مختلفة يجب استخدام نفس المعامل. ولمقاييس الالتواء أشكال مختلفة سوى كانت بسيطة أو مركبة، نذكر منها:

3-1- المعامل الربيعي للالتواء أو معامل يول وكيندال (Yule & Kendall coefficient): ويدعى أيضا معامل الاتواء الربيعي. يقيس معامل Yule التواء التوزيع حيث يأخذ بعين الاعتبار المواقع النسبية للربيعات بالنسبة للوسيط. وهو يصاغ بالشكل التالي:

$$y_k = \frac{Q_1 + Q_3 - 2Me}{Q_3 - Q_1}$$ أو بشكل مكافئ $$y_k = \frac{(Q_3 - Me) - (Me - Q_1)}{Q_3 - Q_1}$$

فبالنسبة لتوزيع متماثل تكون المسافات $(Q_3 - Me) = (Me - Q_1)$ وعليه يكون $y_k = 0$

ويكون التوزيع ملتو نحو اليسار حين $(Q_3 - Me) < (Me - Q_1)$ وعليه $y_k < 0$

ويكون التوزيع ملتو نحو اليمين حين $(Q_3 - Me) > (Me - Q_1)$ وعليه $y_k > 0$

ولقد عرض كارل بيرسون Karl Pearson طريقتين بسيطتين لإيجاد الالتواء في العينات ويتلخصان في المعاملين التاليين:

2-3- معامل بيرسون لالتواء الأول (Pearson's Coefficient of Skewness #1): يقيس معامل بيرسون تماثل منحنى توزيعي بالمقارنة بين قيم الوسط الحسابي والمنوال ويحسب بدلالة الصيغة الآتية:

$$S_{k1} = \frac{\overline{X} - Mo}{\sigma}$$

فبالنسبة لتمام التماثل تكون $Mo = Me = \overline{X}$ و عليه $S_k = 0$

ويكون التوزيع ملتو لليسار حين $\overline{X} < Mo$ وعليه $S_k > 0$

ويكون التوزيع ملتو لليمين حين $\overline{X} > Mo$ وعليه $S_k < 0$

3-3- معامل بيرسون لالتواء الثاني (Pearson's Coefficient of Skewness #2): والذي يحسب بدلالة الوسيط والوسط الحسابي وكذا الانحراف المعياري كما يلي:

$$S_{K2} = \frac{3(\overline{X} - Me)}{\sigma}$$

يجب التنبيه أنه كون هذه المعاملات اللامعلمية، بسيطة وسريعة الحساب تهتم أكثر بوجود واتجاه الالتواء ويمكن استخدمها إلا في حالة التوزيعات الأحادية المنوال ذات الشكل الطبيعي نسبيا والمتباعدة قليلا على التماثل. كما نشير أنه في التعاريف الحدية أغلب المقاييس الشكلية أصبحت أكثر دقة وتستخدم طريقة العزوم. لذا يستحب التطرق لهذه الأشكال التي تستخدم في كثير من المعاملات الإحصائية.

★ **العزوم والعزوم المركزية:** يعرف العزم $\aleph_{a,k}$ ذات الرتبة k وبالنسبة لـ: c بالشكل التالي:

$$\aleph_{c,k} = \frac{\sum n_i (x_i - c)^k}{N}$$

أين: $c \in \Re$ (ثابت حقيقي) و $k \in \aleph$ (عدد طبيعي)

و حين يكون $\overline{X} = c$ ننسب للعزم $\aleph_{c,k}$ تسمية عزم مركزي μ_k ذات الرتبة k ويأخذ الشكل التالي:

نلاحظ سريعا حسب قيم c و k أن:

الوسط الحسابي $\aleph_{0,1} = \frac{\sum n_i (x_i)}{N} = \overline{X}$

مربع الوسط التربيعي $\aleph_{0,2} = \frac{\sum n_i (x_i)^2}{N} = X_Q^2$

$$\aleph_{\overline{X},k}=\mu_k=\frac{\sum n_i(x_i-\overline{X})^k}{N}$$

التباين $\aleph_{\overline{X},2}=\frac{\sum n_i(x_i-\overline{X})^2}{N}=\sigma^2$

متوسط الانحرافات $\aleph_{\overline{X},1}=\frac{\sum n_i(x_i-\overline{X})}{N}=0$

4-3- معامل فيشر للالتواء γ_1 (Fisher Skewness coefficient): وهو من أدق المقاييس الشكلية وأكثرها استعمالا، حيث يوضح شدة واتجاه الالتواء وهو يستند الى حساب مسبق لـ μ_3 العزم المركزي من الرتبة 3. ويعرف معامل فيشر للالتواء بالصيغة:

$$\gamma_1=\frac{\mu_3}{\sigma^3}$$

أين:

$\mu_3=\frac{1}{N}\sum_{i=1}^{k}n_i(x_i-\overline{X})^3$ العزم المركزي من الرتبة 3.

و σ : الانحراف المعياري.

ويكون التوزيع

- غير متماثل وملتو من جهة اليمين إذا $\gamma_1>0 \leftarrow$
- متماثل تماما إذا $\gamma_1=0 \leftarrow$
- غير متماثل وملتو من جهة اليسار إذا $\gamma_1<0 \leftarrow$

5-3- معامل بيرسون للالتواء β_1 (Pearson's moment coefficient of skewness): وهو مقياس شكلي مستعمل غالبا في الكشف عن الالتواء، ويعرف بالصيغة:

$$\beta_1=\frac{\mu_3^2}{\sigma^6}=\left(\frac{\mu_3}{\sigma^3}\right)^2=\lambda_1^2$$

أين:

$\mu_3^2=\left(\frac{1}{N}\sum_{i=1}^{k}n_i(x_i-\overline{X})^3\right)^2$: مربع العزم المركزي من الرتبة 3.

و σ : الانحراف المعياري.

ويكون التوزيع

- غير متماثل وملتو إذا $\beta_1\neq 0 \leftarrow$
- متماثل تماما إذا $\beta_1=0 \leftarrow$

4 - مقاييس التفرطح MEASURES OF KURTOSIS

هو مقياس يقيس درجة علو أو انخفاض منحنى توزيع تكراري بالنسبة للمنحنى الطبيعي المتماثل. وهو مقياس للتشتت حول القيم الوسطى. كلما كان التوزيع كبير كلما كان المنحنى ممتد plate. فدور معامل الالتواء هو الكشف عن تماثل المنحنى التكراري، فقد يكون المنحنى متماثلا الا انه غير اعتدالي لأنه مدبب او مفرطح ولذا فنحن في حاجة الى مقياس اخر لمعرفة ما إذا كان المنحنى مدببا او مفرطحا ام لا وهذا المقياس هو معمل التفرطح والاشكال التالية توضح هذه الفكرة فالمنحنيات الثلاثة التالية متماثلة الا انها غير اعتدالية. اما المنحنى الاعتدالي المعياري فهو متماثل وليس مدببا او مفرطحا.

يمكننا تحديد تفرطح التوزيع بقيم معامل بيرسون β_2 أو معامل فيشر γ_2 للتفرطح، المرتكزين على العزم المركزي من الرتبة 4، أي μ_4 والذي يصاغ كما يلي:

$$\mu_4 = \frac{1}{N}\sum_{i=1}^{k} n_i (x_i - \overline{X})^4$$

وعليه يمكن استخدام صيغ المعاملات التالية لحساب التفرطح:

i. **معامل بيرسون للتفرطح** β_2 (Pearson Kurtosis coefficient): وهو القيمة المستوحاة من الشكل التالي:

$$\beta_2 = \frac{\mu_4}{\sigma^4} = \frac{\mu_4}{\mu_2^2}$$

ii. **معامل فيشر للتفرطح** γ_2 (Fisher Kurtosis coefficient): وهو القيمة المستخرجة من الحساب التالي:

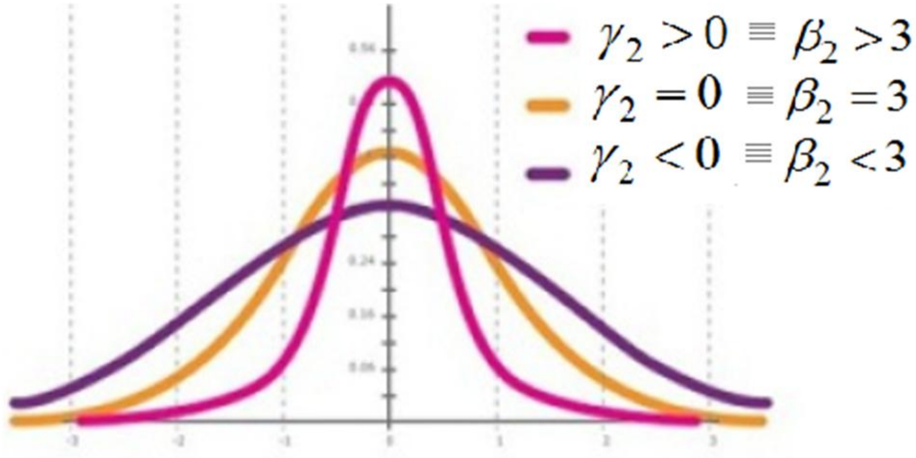
$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 \qquad \text{وبشكل موافق} \qquad \gamma_2 = \beta_2 - 3$$

فكلما كان التوزيع منكمش كلما كانت قيم المعاملات كبيرة وكلما كان التوزيع المشاهد ممتد ومنحل كلما كانت قيم المعاملات صغيرة. ونقول عموما أنه :

إذا كان $\beta_2 = 3$ أو $\gamma_2 = 0$ يكون منحنى التوزيع معتدلا (Mesokurtic distribution) أي قريب من المنحنى الطبيعي؛

إذا كان $\beta_2 > 3$ أو $\gamma_2 > 0$ يكون منحنى التوزيع مدببا (Leptokurtic distribution) معظم القيم بالقرب من الوسط الحسابي والذيلين؛

إذا كان $\beta_2 < 3$ أو $\gamma_2 < 0$ يكون منحنى التوزيع منبسطا (مفرطحا) (Platykurtic distribution) معظم القيم بعيدة عن الوسط والذيلين.

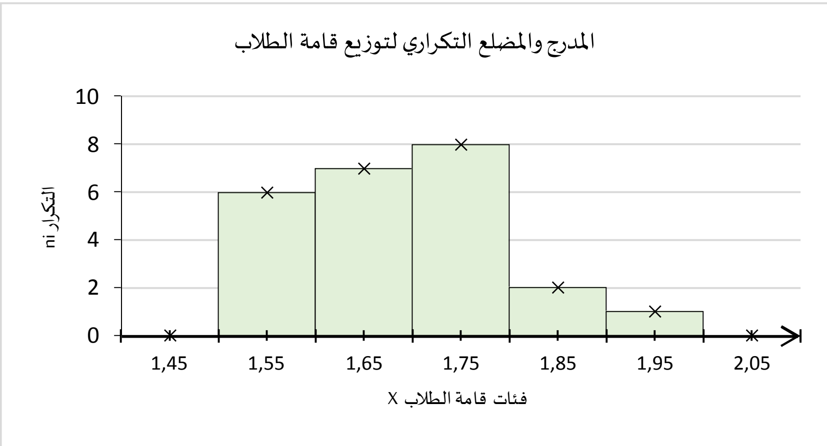

$\gamma_2 > 0 \equiv \beta_2 > 3$
$\gamma_2 = 0 \equiv \beta_2 = 3$
$\gamma_2 < 0 \equiv \beta_2 < 3$

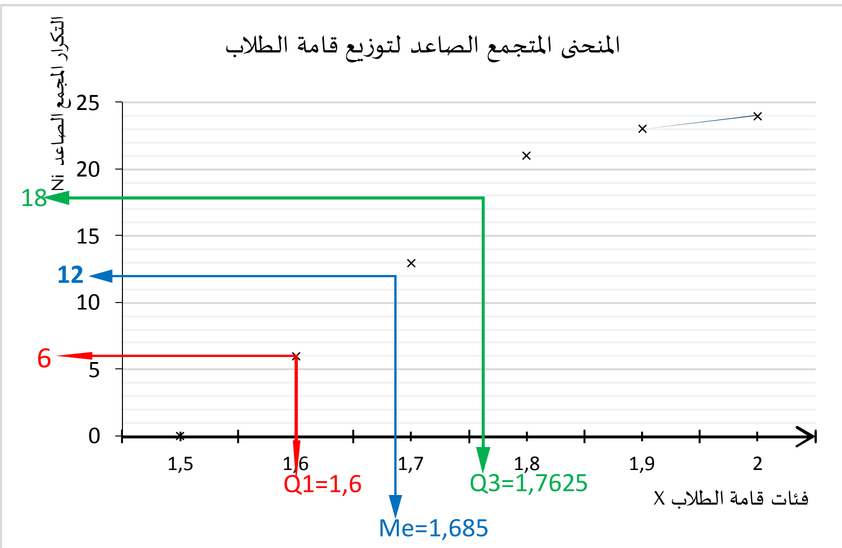
مثال شامل: توزيع قامة (بالمتر) لمجموعة مكون من 24 طالب

فئة القامة	التكرار (n_i)	التكرار المتجمع الصاعد $N_i^{\uparrow}$	مراكز الفئات (x_i)	$n_i \times x_i$	$\sum_{i=1}^{5} n_i(x_i-\overline{X})^2$	$\sum_{i=1}^{5} n_i(x_i-\overline{X})^3$	$\sum_{i=1}^{5} n_i(x_i-\overline{X})^4$
[1.50 ; 1.60 [	6	6	1,55	9,3	0,1134	-0,0156	0,0021
[1.60 ; 1.70 [	7	13	1,65	11,55	0,0098	-0,0004	0,0000
[1.70 ; 1.80 [	8	21	1,75	14	0,0313	0,0020	0,0003
[1.80 ; 1.90 [	2	23	1,85	3,7	0,0528	0,0086	0,0160
[1.90 ; 2.00 [	1	24	1,95	1,95	0,0689	0,0181	0,1140
$\sum$	**24**			**40,5**	**0,2763**	**0,0127**	**0,1325**
$\frac{\sum}{N}$				*1,6875*	*0,0115*	*0,0005*	*0,0055*

احسب المدى ونصف المدى الربيعي والانحراف المتوسط والانحراف المعياري

احسب مقاييس الالتواء والتفرطح من البيانات في المسألة





N: مجموع التكرارات وهو يمثل أيضا حجم العينة المدروسة N=24.

لنحسب المقاييس:

الوسط الحسابي Arithmetic Mean $\overline{X}=\frac{1}{N}\sum_i n_i x_i=\frac{1}{24}\sum_{i=1}^{5} n_i x_i=\frac{40,5}{24}=1,6875m;$
الوسيط Median $Me=L_{inf\,Me}+\frac{\left(\frac{N}{2}-N^{\uparrow}_{Me-1}\right)}{n_{Me}}\times a=1,6+\frac{(12-6)}{7}\times 0,1=1,685m;$
المنوال Mode $Mo=L_{inf\,Mo}+\left(\frac{\Delta_1}{\Delta_1+\Delta_2}\right)\times a=1,7+\left(\frac{1}{1+6}\right)\times 0,1=1,7143m;$
الربيع الأول بما أن $\frac{N}{4}=\frac{24}{4}=6$ فإن $Q_1=1,60m;$
الربيع الثالث $Q_3=L_{inf\,Q_3}+\frac{\left(\frac{3N}{4}-N^{\uparrow}_{Q3-1}\right)}{n_{Q_3}}\times a=1,7+\frac{(18-13)}{8}\times 0,1=1,7625m;$
العزم المركزي من الرتبة 2 أو التباين Variance $\mu_2=\sigma^2=\frac{1}{N}\sum_i n_i\left(x_i-\overline{X}\right)^2=\frac{0,2761}{24}=0,0115m^2;$
وعليه فإن الانحراف المعياري Standard-Deviation $\sigma=\sqrt{\mu_2}=\sqrt{\sigma^2}=\sqrt{0,0115}=0,1072m;$
العزم المركزي من الرتبة 3 $\mu_3=\frac{1}{N}\sum_i n_i\left(x_i-\overline{X}\right)^3=\frac{0,0125}{24}=0,0005m^3;$
معامل الاختلاف Coefficient of variation $cv_x=\frac{\sigma_x}{\overline{X}}\times 100=\frac{0,1072}{1,6875}\times 100=6,35\%;$
المدى الربيعي (interquartile range) $(Q_3-Q_1)=1,7625-1,6=0,1625\ m;$
الانحراف الربيعي (quartile deviation) $Q=\frac{(Q_3-Q_1)}{2}=\frac{0,1625}{2}=0,08125m;$
المعامل الربيعي للالتواء أو معامل ليول وكيندال (Yule & Kendall coefficient)

$$y_k = \frac{(Q_3 - Me) - (Me - Q_1)}{Q_3 - Q_1} = \frac{(1{,}7625 - 1{,}685) - (1{,}685 - 1{,}60)}{1{,}7625 - 1{,}60} = (-0{,}0461);$$
أول معامل بيرسون للالتواء (Pearson coefficient) $$S_k = \frac{\overline{X} - Mo}{\sigma} = \frac{1{,}6875 - 1{,}7143}{0{,}1072} = (-0{,}25);$$
ثاني معامل بيرسون للالتواء (Pearson Skewness coefficient) $$\beta_1 = \frac{\mu_3^2}{\sigma^6} = \left(\frac{\mu_3}{\sigma^3}\right)^2 = \frac{(0{,}0005)^2}{(0{,}00123)^2} = 0{,}1823;$$
معامل فيشر للالتواء (Fisher Skewness coefficient) $$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{(0{,}0005)}{(0{,}1072)^3} = 0{,}4275;$$
معامل بيرسون للتفرطح (Pearson Kurtosis coefficient): $$\beta_2 = \frac{\mu_4}{\sigma^4} = \frac{(0{,}0055)}{(0{,}1072)^4} = 41{,}6643$$

تتفق الأربع معاملات الالتواء على أن قامة الطلاب ذات توزيع ملتوي طفيفا نحو اليمين.

تمارين للمراجعة العامة:

التمرين 1:

أذكر بصفة وجيزة أهم أنواع المتغيرات الإحصائية وأصنافها مع إعطاء مثال حي لكل منها.

الحل 1:

أهم أنواع المتغيرات الإحصائية وأصنافها مع مثال حي لكل منها.

نوع المتغير	أصناف المتغيرات	الأمثلة
المتغير الكمي	المتغيرات الكمية المتقطعة (المنفصلة)	عدد الطلاب في الجامعة
	المتغيرات الكمية المتصلة	دخل الأسرة
المتغير النوعي	متغيرات نوعية غير مرتبة	أنواع السيارات
	المتغيرات النوعية المرتبة	مستوى الطلاب الأكاديمي

التمرين 2:

لتكن السلسلة الإحصائية المكونة من 60 قيمة لمستوى الكريات الحمراء في الدم (غ/ل) المقاسة لدى عينة من الأشخاص. وقد تم توزيع نتائج التحاليل من خلال الجدول التالي:

112	105	110	112	118	119	120	120	172	165
128	125	126	127	130	132	133	134	172	163
138	135	138	138	138	141	142	144	176	164
148	145	146	148	148	149	150	150	179	168
153	150	151	151	153	153	170	154	164	166
156	154	155	154	158	160	160	160	156	168

1- حدد المجتمع الإحصائي والمتغير الإحصائي المدروس وطبيعته.

2- رتب السلسلة من خلال جدول تكراري للتكرارات المطلقة.

- بعد تجميع السلسلة من خلال 08 فئات وهي:

[104;114[, [114;124[, [124;134[, [134;144[, [144;154[, [154;164[, [164;174[, [174;184[

3- أكتب الجدول التكراري للتكرارات المطلقة والنسبية للسلسلة المبوبة

4- أحسب التكرار المجمع الصاعد النسبي

5- أحسب التكرار المجمع النازل النسبي

6- أرسم المدرج التكراري للتكرارات المطلقة

7- أرسم المضلع التكراري للتكرارات المطلقة

8- أرسم المنحنى التكراري المجمع الصاعد للتكرارات النسبية

9- أرسم المنحنى التكراري المجمع النازل للتكرارات النسبية

10- أحسب قيم المنوال، الوسيط، والمتوسط الحسابي.

الحل 2:

1- المجتمع الإحصائي: مجموع الأشخاص قيد الدراسة

- **المتغير الإحصائي المدروس** : مستوى الكريات الحمراء في الدم وهو متغير كمي متصل وحدة قياسه (غ/ل).

2- ترتيب السلسلة من خلال جدول تكراري للتكرارات المطلقة:

3- الجدول التكراري للتكرارات المطلقة والنسبية للسلسلة المبوبة

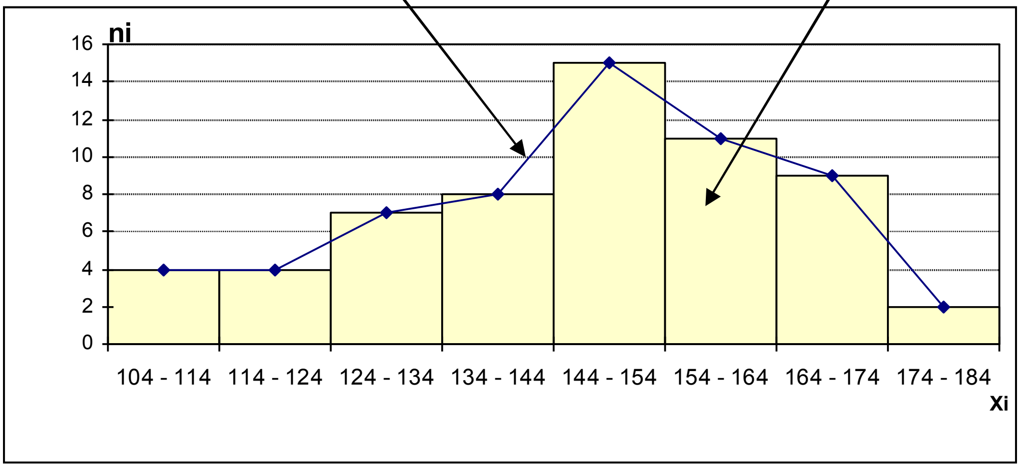
4- التكرار المجمع الصاعد النسبي

5- التكرار المجمع النازل النسبي

الفئات	مراكز الفئات Xi	التكرار المطلق n_i	التكرار النسبي fi	التكرار المجمع الصاعد النسبي Fi↑	التكرار المجمع النازل النسبي Fi↓	n_iX_i
]114 104]	109	4	0,067	0 0,067	1 0,933	436
]124 114]	119	4	0,067	0,133	0,867	476
]134 124]	129	7	0,117	0,250	0,750	903
]144 134]	139	8	0,133	0,383	0,617	1112
*]154 144]	149	15	0,250	0,633	0,367	2235
]164 154]	159	11	0,183	0,817	0,183	1749
]174 164]	169	9	0,150	0,967	0,033	1521
]184 174]	179	2	0,033	1,000	0,000	358
	المجموع	**60**	**1**			**8790**

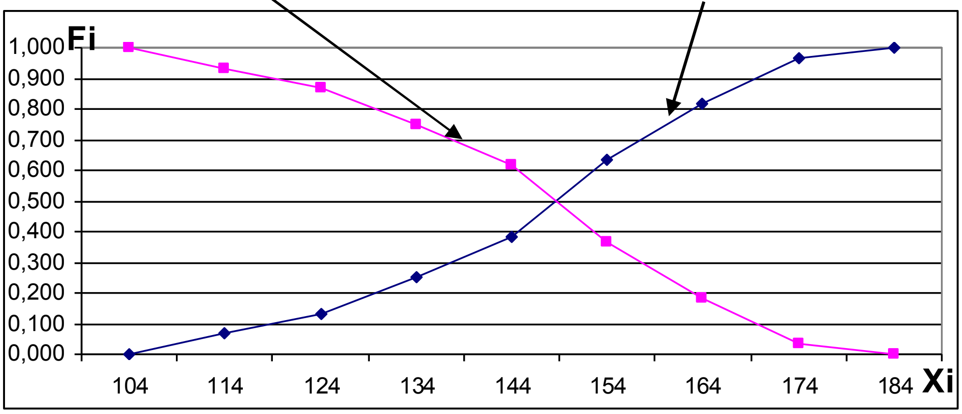
6- رسم المدرج التكراري للتكرارات المطلقة

7- رسم المضلع التكراري للتكرارات المطلقة



8- رسم المنحنى التكراري المجمع الصاعد للتكرارات النسبية

9- رسم المنحنى التكراري المجمع النازل للتكرارات النسبية



10- حساب قيم المنوال، الوسيط، والمتوسط الحسابي.

المنوال Mode :

$$Mo = c \times \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) + L_{Mo} = 10 \times \left(\frac{7}{7+4} \right) + 144 = 150{,}364$$ غ/ل

Δ₁=15–8=7   **Δ₂** =15–11=4   **c**=154-144=10   الحد الأدنى للفئة المنوالية L_{Mo} = 144

الوسيط Médiane:

$$Me = c \times \left(\frac{\left(\frac{1}{2}\right) - F\uparrow_{Me-1}}{f_{Me}} \right) + L_{Med} = 10 \times \left(\frac{0{,}5 - 0{,}383}{0{,}250} \right) + 144 = 148{,}680$$ غ/ل

ت.م.ص.ن للفئة الماقبل وسيطة $F_{Me-1}\uparrow$=0,383   **ت.ن** للفئة الوسيطة = 0,250   **c**=154-144=10

الحد الأدنى للفئة الوسيطة L_{Me}=144

المتوسط الحسابي Moyenne Arithmétique:

$$\overline{X} = \frac{\sum_{i=1}^{8} ni(Xi)}{N} = \frac{\sum_{i=1}^{8} ni(Xi)}{60} = \frac{8790}{60} = 146{,}500$$ غ/ل

التمرين 3:

برهن على أن إذا كانت **X1, X2,....,Xp** سلسلة إحصائية ذات التكرارات **n1, n2, ..., np** المذكورة على الترتيب .

حيث يمثل $\overline{X}$ متوسطها الحسابي و **a** ثابت حقيقي فإن:

$$\sum ni\left(Xi - \overline{X}\right)^2 \leq \sum ni\left(Xi - a\right)^2$$

الحل 3:

البرهان على أن إذا كانت **X1, X2,....,Xp** سلسلة إحصائية ذات التكرارات **n1, n2, ..., np** المذكورة على الترتيب .

حيث يمثل $\overline{X}$ متوسطها الحسابي و **a** ثابت حقيقي فإن:

$$\sum ni\left(Xi - \overline{X}\right)^2 \leq \sum ni\left(Xi - a\right)^2$$

$$\sum ni\left(Xi - \overline{X}\right)^2 = \sum ni\left(Xi - a + a - \overline{X}\right)^2 = \sum ni\left[\left(Xi - a\right) - \left(\overline{X} - a\right)\right]^2$$

$$\sum ni\left[\left(Xi - a\right) - \left(\overline{X} - a\right)\right]^2 = \sum ni\left[\left(Xi - a\right)^2 - 2\left(Xi - a\right)\left(\overline{X} - a\right) + \left(\overline{X} - a\right)^2\right]$$

$$= \sum ni\left(Xi - a\right)^2 - 2\left(\overline{X} - a\right)\sum ni\left(Xi - a\right) + \sum ni\left(\overline{X} - a\right)^2 = \sum ni\left(Xi - a\right)^2 - 2\left(\overline{X} - a\right)\sum\left(niXi - nia\right) + \sum ni\left(\overline{X} - a\right)$$

$$= \sum ni\left(Xi - a\right)^2 - 2\left(\overline{X} - a\right)\left[\left(\underbrace{\sum niXi}_{\mathbf{N\overline{X}}}\right) - \left(a\underbrace{\sum ni}_{\mathbf{N}}\right)\right] + \left(\overline{X} - a\right)^2\underbrace{\sum ni}_{\mathbf{N}}$$

$$= \sum ni\left(Xi - a\right)^2 - 2\left(\overline{X} - a\right)\left(N\overline{X} - aN\right) + \left(\overline{X} - a\right)^2 N = \sum ni\left(Xi - a\right)^2 - 2N\left(\overline{X} - a\right)\left(\overline{X} - a\right) + N\left(\overline{X} - a\right)^2$$

$$= \sum ni(Xi-a)^2 - 2N(\overline{X}-a)(\overline{X}-a) + N(\overline{X}-a)^2 = \sum ni(Xi-a)^2 - 2N(\overline{X}-a)^2 + N(\overline{X}-a)^2$$

$$\Rightarrow \sum ni(Xi-\overline{X})^2 = \sum ni(Xi-a)^2 - N(\overline{X}-a)^2$$

$$\Rightarrow \sum ni(Xi-\overline{X})^2 + N(\overline{X}-a)^2 = \sum ni(Xi-a)^2$$

بما أن $\forall N(\overline{X}-a)^2 \geq 0$ بالتالي:

$$\Rightarrow \sum ni(Xi-\overline{X})^2 \leq \sum ni(Xi-a)^2$$

التمرين 4:

برهن على أن إذا كانت **X1, X2,....,Xp** سلسلة إحصائية. حيث يمثل $\overline{X}$ متوسطها الحسابي فإن:

$$\sum(\boldsymbol{Xi}-\overline{\boldsymbol{X}})=0$$

الحل 4: برهان على $\sum(\boldsymbol{Xi}-\overline{\boldsymbol{X}})=0$

بما أن $\overline{\boldsymbol{X}} = \frac{\sum \boldsymbol{Xi}}{\boldsymbol{N}} \Rightarrow \boldsymbol{N}\overline{\boldsymbol{X}} = \sum \boldsymbol{Xi}$ فإن :

$$\sum(\boldsymbol{Xi}-\overline{\boldsymbol{X}}) = \sum \boldsymbol{Xi} - \sum \overline{\boldsymbol{X}} = \sum \boldsymbol{Xi} - \boldsymbol{N}\overline{\boldsymbol{X}} = \boldsymbol{N}\overline{\boldsymbol{X}} - \boldsymbol{N}\overline{\boldsymbol{X}} = 0$$

التمرين 5:

تحصلنا بصياغة المعطيات الخاصة بدرجات الحرارة المؤوية (CELSIUS) على **متوسط يساوي 18° وتباين بقيمة 25**. ماهو مقدار هذه القيم (المتوسط والتباين) إذا .تم تحويلها إلى نظام فهرنهيت (FAHRENHEIT) علما أن صيغة التحويل بين النظامين هي :

$$\boldsymbol{F} = \frac{9}{3} \times \boldsymbol{C} + 32$$

الحل 5:

إذا كان X يعبر على درجات الحرارة بمقياس (CELSIUS) فإن $\overline{\boldsymbol{X}} = 18° \therefore \therefore \boldsymbol{V_X} = 25$ وعلما أن $\boldsymbol{F}_{\text{FAHRENHEIT}} = \frac{9}{3} \times \boldsymbol{C}_{\boldsymbol{CELSIUS}} + 32$ بتحويل المتغير يكون مقدار قيم المتوسط والتباين في نظام فهرنهيت (FAHRENHEIT) إذا إفترضنا أن: $\boldsymbol{Y} = \frac{9}{3} \times \boldsymbol{X} + 32$ <==

$$\overline{Y} = \frac{9}{3} \times 18 + 32 = 86° \therefore \therefore V_Y = \left(\frac{9}{3}\right)^2 \times 25 = 9 \times 25 = 225$$

التمرين 6:

في السوق المالي المغربي أعطت نتائج مبيعات الأسهم مصنفة من خلال فئات تدل على المبالغ الإجمالي الأسهم المشترات من طرف 50 متعامل إقتصادي الجدول التالي:

فئات مبالغ الأسهم المشترات (مليون درهم)	عدد المتعاملين
]60 - 50]	02
]70 - 60]	05
]80 - 70]	15
]90 - 80]	11
]100 - 90]	11
]110 - 100]	06

1- حدد المجتمع الإحصائي والمتغير الإحصائي المدروس وطبيعته

2 - أحسب المتوسط الحسابي $\overline{X}$ والإنحراف المعياري σ

3 - أحسب الوسيط Me و المنوال Mo

4- بين أي قيمتين لمبالغ الأسهم نتحصل على الـ 50% التي تتوسط مجموع المتعاملين

<u>**الحل 6:**</u>

في السوق المالي المغربي أعطت نتائج مبيعات الأسهم مصنفة من خلال فئات تدل على المبالغ الإجمالي الأسهم المشترات من طرف 50 متعامل إقتصادي الجدول التالي:

فئات مبالغ الأسهم المشترات (مليون درهم)	Xi	عدد المتعاملين ni	ni.Xi	3025	ni.Xi²	التكرار المجمع الصاعد Ni↑
]60 - 50]	55	02	110	4225	6050	02
]70 - 60]	65	05	325	5625	21125	07
]80 - 70]	75	15	1125	7225	84375	22
]90 - 80]	85	11	935	9025	79475	33
]100 - 90]	95	11	1045	13225	99275	44
]110 - 100]	115	06	690	3025	79350	50
]110 - 100]		**50**	**4230**		**369650**	

1- المجتمع الإحصائي: **المتعاملين الإقتصادين في السوق المالي المغربي** المتغير الإحصائي المدروس وطبيعته: **مبلغ مبيعات الأسهم الإجمالي كمي متصل**

2- المتوسط الحسابي: $\overline{X} = \frac{\sum niXi}{N} = \frac{4230}{50} = 84{,}6$ مليون درهم

الإنحراف المعياري: $\sigma = \sqrt{\frac{\sum niXi^2}{N} - \overline{X}^2} = \sqrt{\frac{369650}{50} - (84{,}6)^2} = 15{,}36$ مليون درهم

3- الوسيط : $\boxed{Me = L_{Med} + \left(\frac{(50) - N\uparrow_{Me-1}}{n_{Me}}\right) \times c =}$ **82,73** م درهم و المنوال $\boxed{Mo = L_{Mod} + \left(\frac{\Delta 1}{\Delta 1 + \Delta 2}\right) \times c =}$ **77,14** م درهم

4- نتحصل على الـ 50% التي تتوسط مجموع المتعاملين بين قيمتين لمبالغ الأسهم التي تناسب الربيع الأول Q1 والثالث Q3 : أي

$\boxed{Q1 = L_{Q1} + \left(\frac{(25) - N\uparrow_{Q1-1}}{n_{Q1}}\right) \times c =}$ **73,61** م درهم و $\boxed{Q3 = L_{Q3} + \left(\frac{(75) - N\uparrow_{Q3-1}}{n_{Q3}}\right) \times c =}$ **94,09** م درهم

التمرين -7 –

يبين الجدول التالي تطور نسبة تغير مستوى الادخار خلال الفترة 1998-2001 في مجموع دول أوروبا الشرقية

1998	1999	2000	2001
-0,3%	+0,8%	+2,4%	+1,9%

أ - أحسب المعدل السنوي لنسبة الزيادة للفترة 1998-2001 باستعمال المتوسطات الهندسية

ب - إذا قدرنا أن المعدل السنوي لنسبة الزيادة يساوي% 1,5 + في للفترة 1998-2002، أستنتج نسبة التغير 2002

الحل -7

السنة	1998	1999	2000	2001
نسبة التغير TA	-0,3%	+0,8%	+2,4%	+1,9%
مؤشر التغير I	99,7%	100,8%	102,4%	101,9%

1)-

$I_{1998\text{-}2001}=(0{,}997\text{x}1{,}008\text{x}1{,}024\text{x}1{,}019)^{1/4} = (1{,}0486)^{1/4}=101{,}19\% \quad =>$

$TAM_{1998\text{-}2001} = 1{,}19\%$ المعدل السنوي لنسبة التغير

2)- $TAM_{1998\text{-}2002} = 1{,}5\% \;=> I_{1998\text{-}2002}=101{,}5\%=1{,}015=(1{,}0486 \text{ x } \boldsymbol{A})^{1/5}$

$$=> \boldsymbol{A} = \frac{(1{,}015)^5}{1{,}0486} = 1{,}0274 = 102{,}74\% \quad => \quad TA_{2002}=2{,}74\%$$

التمرين -8 –

في مؤسسة اقتصادية متوسط أجور المستخدمين هو 14.000 دج. الاجر المتوسط للعمال الذكور هو 15.000 دج و الاجر المتوسط للعاملات يبلغ 13.500 دج . ما هي نسبة العاملات في هذه المؤسسة .

الحل -8 –

متوسط أجر الرجال	X_{hommes}	15.000 DA	% hommes	A %	**نسبة الرجال**
متوسط أجر النساء	X_{femmes}	13.500 DA	% femmes	(1-A)%	**نسبة النساء**

متوسط أجر العمال X global = $A\times X_{hommes} + (1\text{-}A) \times X_{femmes}$ = 14 000 DA

X global = A×15.000 + (1-A) × 13.500= 14 000 DA

□ A × (15.000-13.500) +13.500 = 14.000 □ A × (1.500) = 500

□ A = 500/1500 = 0,3333 = 33,33%

نسبة النساء □ **1- A = 1- 0,3333 = 66,67%**

التمرين -9

لتوزيع الأجر السنوي النقابي لشركة كبيرة الخصائص التالية :

المتوسط الحسابي $\overline{X}$ = 45.000 $, الوسيط Me = 43.500 $, الانحراف المعياري δ = 1.400 $

قدمت إدارة الشركة لهيئة النقابة عرضين لزيادة الأجور:

1- زيادة في الأجر تقدر بـ: 8 % لكل عامل،

2- منح 900 $ لكل عامل مع زيادة تقدر بـ: 6 % في الأجر

- أي من العرضين يناسب الأحسن أغلب العمال؟ علل

الحل -9

خصائص توزيع الأجر السنوي النقابي :

المتوسط الحساب $\overline{X}$ =45.000 $ الوسيط Me = 43.500$ الانحراف المعياري δ = 1.400$

العرضين لزيادة الأجور:

1- زيادة في الأجر تقدر بـ: 8% لكل عامل,

$\overline{X}$ 48.600$

Me= $46.980

δ = 1.512$

2- منح 900 $ لكل عامل مع زيادة تقدر بـ: 6 % في الأجر

$\overline{X}$ 48.600$

Me= $47.010 *****

δ = 1.484$

بسبب كبر الوسيط يظهر أن الإقتراح الثاني أفضل من الأول بالنسبة لأغلب العمال

التمرين -10 –

تم تصنيف عمال مصنع وقف أجرة ساعة العمل وتحصلنا على الجدول التالي:

أجرة ساعة العمل	عدد العمال
] 21,5 - 15,5]	03
] 27,5 - 21,5]	10
] 33,5 - 27,5]	19
] 39,5 - 33,5]	28
] 45,5 - 39,5]	21
] 51,5 - 45,5]	14
] 57,5 - 51,5]	05
	100

1- أحسب التكرار المجمع الصاعد لتصنيف عمال المصنع

2- عين مراكز الفئات Xi

لنعتبر أن :

$$\sum_{i=1}^{7} n_i x_i = 3746 \qquad \sum_{i=1}^{7} n_i (x_i - \overline{x})^2 = 7395{,}84$$

3. احسب الوسط الحسابي والتباين.

4. احسب معامل الاختلاف

5. أدرس إلتواء هذا التوزيع.

الحل -10 –

أجرة ساعة العمل	عدد العمال *ni*	Xi	Ni↑	
] 21,5 - 15,5]	3	18,5	**3**	
] 27,5 - 21,5]	10	24,5	**13**	
] 33,5 - 27,5]	19	30,5	**32**	
] 39,5 - 33,5]	28	36,5	**60**	الفئة الوسطية****
] 45,5 - 39,5]	21	42,5	**81**	
] 51,5 - 45,5]	14	48,5	**95**	
] 57,5 - 51,5]	5	54,5	**100**	
المجموع	**100**			

حساب الوسط الحسابي والتباين

$$\overline{X} = \frac{\sum_{i=1}^{7} n_i x_i}{N} = \frac{3746}{100} = 37{,}46$$

$$V = \frac{\sum_{i=1}^{7} n_i (x_i - \overline{x})^2}{N} = \frac{7395{,}84}{100} = 73{,}9584$$

حساب معامل الإختلاف

$$CV = \frac{\delta}{\overline{X}}100 = \frac{\sqrt{73{,}9584}}{37{,}46}100 = 22{,}96\ldots\%$$

دراسة إلتواء التوزيع

حساب الوسيط

$$Me = L_{\inf Me} + \frac{\left(\frac{N}{2} - N\uparrow_{Me-1}\right)}{n_{Me}} \times c = 33{,}5 + \frac{50-32}{28} \times 6 = 37{,}357$$

$$\Delta = \frac{3(\overline{X} - Me)}{\delta} = 0{,}036$$

إلتواء طفيف وموجب (نحو اليمين)

التمرين -11 –

الجدول التالي يوضح توزيع (200) موظف بإحدى الوزارات حسب أعمارهم بالسنة.

فئات العمر	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60
عدد الموظفين	10	17	24	43	34	30	23	19

1. احسب الوسط الحسابي والانحراف المعياري ووسيط العمر.
2. بين أي قيمتي سن تتوسط نسبة 50% من الموظفين.
3. ما هو إذا السن الأدنى لـ 25 % من الموظفين الأكبر سنا.
4. هل توزيع أعمار الموظفين متماثل؟ دلل على إجابتك بحساب معامل الالتواء.

الحل -11

من خلال البيانات المتوفرة نستطيع بناء الجدول التالي

Class	ni	Xi	ni x Xi	(Xi)²	ni x (Xi)²	Ni↑	
20-25	10	22,5	225	506,25	5 062,50	10	
25-30	17	27,5	467,5	756,25	12 856,25	27	
30-35	24	32,5	780	1 056,25	25 350,00	51	Q_1
35-40	43	37,5	1612,5	1 406,25	60 468,75	94	
40-45	34	42,5	1445	1 806,25	61 412,50	128	
45-50	30	47,5	1425	2 256,25	67 687,50	158	Q_3
50-55	23	52,5	1207,5	2 756,25	63 393,75	181	
55-60	19	57,5	1092,5	3 306,25	62 818,75	200	
SUM	**200**		**8255**		**359050**		

1. حساب الوسط الحسابي والانحراف المعياري ووسيط العمر.

$$\bar{x} = \frac{\sum n_i x_i}{N} = \frac{8255}{200} = 41{,}275 \qquad \delta x = \sqrt{\frac{\sum n_i x_i^2}{N} - \bar{x}^2} = \sqrt{\frac{359050}{200} - 1703{,}626} = \sqrt{91{,}624} = 9{,}572$$

السن الوسيط للعمال يعني السن الذي يصل إليه نصف العمال أي 100 عامل وعليه الفئة الوسيطة هي : 40 – 45

$$Me = L_{Med} + \left(\frac{(N/2) - N\uparrow_{Me-1}}{n_{Me}}\right)\times c = 40 + \left(\frac{100-94}{34}\right)\times 5 = 40{,}882$$ سنة

2. يتوسط سن ال 50% من الموظفين بين قيم الربيع الأول والثالث أي بين:

$$Q_1 = L_{Q1} + \left(\frac{(N/4) - N\uparrow_{Q1-1}}{n_{Q1}}\right)\times c = 30 + \left(\frac{50-27}{24}\right)\times 5 = 34{,}792$$ سنة و

$$Q_3 = L_{Q3} + \left(\frac{(3N/4) - N\uparrow_{Q3-1}}{n_{Q3}}\right)\times c = 45 + \left(\frac{150-128}{30}\right)\times 5 = 48{,}667$$ سنة

أي أن 50% من العمال تتوسط عموما قيمتي 35 و 49 سنة وعليه فإن الفئة الوسطي من العمال مسنة نوعا ما

3. السن الأدنى لـ 25 % من الموظفين الأكبر سنا يقع على قيمة الربيع الثالث Q_3 أي: 48,667 سنة يعني هذا أن 25 % من العمال يفوق عمرهم هذا السن

4. دراسة تماثل توزيع أعمار : بحساب معامل بيرسون *Pearson* للالتواء $\gamma = \frac{3(x - Me)}{\delta} = \frac{3(41{,}275 - 40{,}882)}{9{,}572} = 0{,}123$ حيث يظهر أن توزيع الأعمار ملتويا بشكل ضعيف يمينا وهو قريب من التماثل التام.

التمرين -12- –

أجرية دراسة تسويقية على مستوى أوروبا حول استهلاك منتوج جديد من المصبرات وقد استخلص نتائج الدراسة التي أقيمة حول مساحتين كبيرتين للبيع (supermarché), الاولى متواجدة بفرنسا (supermarché1) والثانية في بريطانيا (supermarché2) وقد تم تدوين النتائج في تقرير أين نجد الجدول التالي:

جدول: بالإحصائيات المبينة لكميات المنتوج المباعة شهريا

الزبائن	المتوسط	الانحراف المعياري
Supermarché 1	7,3	2,6
Supermarché 2	9,5	2,4

ملاحظة: الكميات المسجلة بالنسبة للسوق 1 معبرة بالكغ بينما بالنسبة للسوق 2 فهي معبرة بالرطل.

ماهي الاستنتاجات التي يمكننا انساىها حول تشتت المبيعات الأسواق بفرنسا (supermarché1) وبريطانيا (supermarché2)

الحل -12

لمقارنة بين تشتت المبيعات الأسواق المدروسة نحسب معامل التغير CV وبالرجوع إلى صيغة المعامل تتحصل على القيم التالية:

بالنسبة الأسواق بفرنسا (supermarché1) : $cv1 = \frac{\delta}{\overline{X}}\times 100 = \frac{2{,}6}{7{,}3}\times 100 = 0{,}35\times 100 = 35\%$

بالنسبة الأسواق بريطانيا (supermarché2) : $cv2 = \frac{\delta}{\overline{X}}\times 100 = \frac{2{,}4}{9{,}5}\times 100 = 0{,}25\times 100 = 25\%$

يظهر أن تشتت مبيعات المنتوج الجديد أقوي في الاسواق الأسواق بفرنسا منها بالنسبة الأسواق في بريطانيا الذي يظهر أكثر إستقرار.

التمرين -13- –

في ولاية من ولايات الوطن تحصلنا من تقرير المصالح الفلاحية حول حجم المستثمرات الزراعية على النتائج التالية:

الحجم (هكتار)	0 - 10	10 - 30	30 - 50	50 - 100	100 - 200
العدد المستثمرات	30	80	60	20	10

المطلوب:

- 1 - تحديد المجتمع الإحصائي والمتغير الإحصائي المدروس وطبيعته
- 2 - حساب التكرارات النسبية
- 3 - حساب التكرار النسبي المجمع الصاعد
- 4 - حساب القيم الوسطى للوسيط Me والوسط الحسابي $\overline{X}$ والانحراف المعياري δ
- 5 – ماهو حجم الـ : 10 % من المستثمرات الاكبر من مجموع مستثمرات الولاية (مستعينا بالعشير التاسع D_9)
- 6 – إذا علمنا أن المتوسط الوطني لحجم المستثمرات يساوي 50 هكتار. هل يمكننا القول أن هذه الولاية على العموم ذات مستثمرات صغيرة

الحل -13

- 1 - المجتمع الإحصائي **مجموع المستثمرات الفلاحية** المتغير الإحصائي **حجم المستثمرة الفلاحية** (هكتار)

طبيعته **كمي متصل**

الحجم (هكتار)	العدد n_i	التكرار النسبي f_i	التكرار النسبي المجمع الصاعد $F_i\uparrow$	التكرار المجمع $N_i\uparrow$	X_i	$n_i.X_i$	$n_i.X_i^2$
0 - 10	30	**0,15**	**0,15**	30	5	150	750
10 - 30	80	**0,40**	**0,55**	110	20	1600	32000
30 - 50	60	**0,30**	**0,85**	170	40	2400	96000
50 - 100	20	**0,10**	**0,95**	190	75	1500	112500
100 – 200	10	**0,05**	**1,00**	200	150	1500	225000
	200	**1**				**7150**	**466250**

- 4 - حساب القيم ***الوسيط*** Me ***Médiane:***

يتواجد الوسيط في الفئة التي نجمع بها الـ: 50% من التكرارت أي الفئة 10 – 30 ومنه يساوي الوسيط:

$$Me = L_{Med} + \left(\frac{\left(\frac{1}{2}\right) - F\uparrow_{Me-1}}{f_{Me}}\right) \times c = 10 + \left(\frac{0,5-0,15}{0,40}\right) \times 20 = 27,50$$

الوسيط Me=27,5 هكتار

الوسط الحسابي $\overline{X}$ Moyenne Arithmétique:

$$\overline{X} = \frac{\sum ni(Xi)}{N} = \frac{30\times5+80\times20+60\times40+20\times75+10\times150}{200} = \frac{7150}{200} = 35,75$$

الوسط الحسابي $\overline{X}$=35,75 هكتار

الانحراف المعياري δ écart-type: $\delta = \sqrt{\frac{\sum niXi^2}{N} - \overline{X}^2} = \sqrt{\frac{466250}{200} - (35,75)^2} = 32,5$

الإنحراف المعياري δ= 32,5 هكتار

- 5 – نتحصل على الـ 10% من المستثمرات الأكبر من مجموع مستثمرات الولاية بين قيمتين لحجم المستثمرات الزراعية التي تنحصربين العشرية التاسعة D9 والعاشرة D10 : أي

$$D9 = \left(\frac{\left(\frac{9}{10}\right) - F\uparrow_{D9-1}}{f_{D9}}\right) \times c + L_{D9} = \left(\frac{0{,}9-0{,}85}{0{,}1}\right) \times 50 + 50 = 75$$ **D_9 =75 هـ** يفوق حجم الـ 10% من المستثمرات الكبرى قيمة 75 هكتار

- 6 – إذا علمنا أن المتوسط الوطني لحجم المستثمرات يساوي 50 هكتار. هل يمكننا القول أن هذه الولاية على العموم ذات مستثمرات صغيرة

إذا علمنا أن المتوسط الوطني لحجم المستثمرات يساوي 50 هكتارو متوسط الولاية يساوي 35,75 هكتار. نرى واضحا أن متوسط الولاية أقل من المتوسط الوطني لحجم المستثمرات. ويبين جدول التوزيعات التكرارية أن 170 مستثمرة من الـ 200 لهم حجم أقل من المتوسط الوطني (50هكتار) إذا **% 85 من المستثمرات لها حجم أقل من المتوسط الوطني**

المراجع

Bibliographie

1. أ.د. أماني موسى محمد،التحليل الإحصائي للبيانات، مركز تطوير الدراسات العليا والبحوث- في العلوم الهندسية، كلية الهندسة- جامعة القاهرة. 2007
2. د. إبراهيم عبد الله، مبادئ الإحصاء، قسم الإحصاء، كلية الإدارة والاقتصاد، جامعة الإمارات العربية المتحدة. 1998
3. د. جلال الصياد، أ.د. عبد الحميد ربيع، أ. عادل سمرة، الإحصاء لطلاب الدراسات الاقتصادية والإدارية، دار حافظ للنشر والتوزيع. المعتمدة من طرف، قسم الإحصاء، جامعة الملك عبد العزيز، 1429/1428 هـ
4. مقدمة في الإحصاء، 161 إحص، المؤسسة العامة للتعليم الفني والتدريب المهني، الإدارة العامة لتصميم وتطوير المناهج.
5. Alain Baccini, Statistique Descriptive, Publication du laboratoire de statistique et probabilité, Université Paul Sabatier, Toulouse III.
6. Etienne Bressoud & Jean-Claude Kahane/Statistique descriptive. Applications avec Excel et calculatrices / Ed. Pearson.
7. Gilbert Saporta, Probabilité, Analyse des Données et Statistique. Ed. Technip. Paris. 2006.
8. Spiegel, Murray R. Theory and problems of Probability and Statistcs, Schaum's Outline Series, Mc Graw-Hill. 1998