وزارة التعليـــم العالـــي والبحـــث العلمـــي
جامعة امحمد بوقرة – بومرداس
كلية العلوم الاقتصادية، التجارية وعلوم التسيير

سلسلة محاضرات في الإحصاء الوصفي

موجهة لطلبة الليسانس في ميدان العلوم الاقتصادية والتجارية وعلوم التسيير

مقدمة من طرف:
الدكتورة: **نبيلة عرقوب**

بسم الله الرحمن الرحيم

الإهــــــداء

إلى طلبتي الأعزاء؛

وإلى كل من يؤمن بالعلم تقدما ورُقيّا.

أهدي هذه السلسلة من المحاضرات

المقدمـــة العامـــة

بسم الله الرحمـــان الرحيم

لا شك **عزيزي الطالب** أنك **تحاول دائما** تسعى لكتابة محاضراتك والاستعداد لاجتياز امتحانات الدورات المختلفة حسب ما يمليه عليك أستاذ المادة.

وعلى هذا الأساس، **وتسهيلا لك،** رأينا أن نضع بين يديك هذه المطبوعة التي تحتوي وتشتمل على **البرنامج المفصل** لمادة **الإحصاء الوصفي (أي إحصاء 1)** والذي بات لزاما على الطالب إدراكه في مرحلته الأولى من التحاقه بالجامعة راجين أن نكون قد وفرنا عنك تعب الكتابة المستمرة للدروس.

لهذا، دعنا نقدم لك التلخيص التالي:

علم الإحصاء هو أحد علوم الرياضيات البحتة، فهو أداة للواقع التجريبي لكل المجالات المختلفة، حيث يعتمد الإحصاء على البيانات أو المشاهدات التي تظل القاعدة الأساسية في هذا العلم من أجل تجسيد الظواهر في صورة علاقات رياضية.

فعلى الطالب معرفة **ما هو مصدر البيانات؟ كيف يتم تجميعها؟ كيف يتم استعمالها؟ ما هو الهدف من استحقاق النتائج بعد تطبيق المعايير أو المقاييس الوصفية؟**

لهذا رأينا أن تحتوي مطبوعتنا على الفصول الأساسية والتفصيلية التالية:

- **الفصل الأول**: **مدخل إلى الإحصاء،** أين يتعرف الطالب على أهم الألفاظ الإحصائية المستخدمة في هذا المجال؛
- **الفصل الثاني**: **العرض الجدولي والبياني،** والذي يتطرق الطالب من خلاله إلى تلخيص البيانات وتجميعها ثمّ تبويبها أي عرضها في جداول مختلفة أهمها جدول التوزيع التكراري للمتغيرات الكمية، وبعد ذلك رسمها بيانيا للتعرف على مختلف الأشكال والرسومات حسب تنوع المتغيرات؛
- **الفصل الثالث**: والخاص **بمقاييس النزعة المركزية،** رأينا من خلاله أن نبيّن للطالب أنه بعد تجميع البيانات وعرضها في جداول مختلفة، ثمّ رسمها بيانيا، تأتي مرحلة أخرى وهي وصف هذه البيانات وذلك من خلال مجموعة من المقاييس الوصفية أهمها مقاييس النزعة المركزية، حيث سنتطرق من خلالها إلى المتوسط الحسابي، الوسيط والمنوال، والعلاقة التي تربط بين المقاييس الثلاثة، دون أن ننسى المقاييس الموضعية الأخرى (منها الربيعيات)؛
- **الفصل الرابع**: وهو تكملة للفصل السابق، أين يتعرف الطالب من خلاله على المقاييس التي تصف حالة **تباعد** أو **تشتت قيم المتغير الإحصائي** فيما بينها مقارنة بقيمتها المركزية.

وفي الأخير، نود أن نشير إلى أن هدفنا هو **ترقية العمل البيداغوجي** من أجل **إصلاح التعليم** في جامعتنا ومجتمعنا، **والله** من وراء القصد.

أستاذة المادة:

د/ نبيلة عرقوب

الفصـل الأول

مدخل إلى الإحصاء

1- تعريف:

علم الإحصاء هو أحد العلوم الرياضيات التطبيقية والذي يختص بتطبيق أكثر الطرق في جمع وتحليل وتفسير البيانات بغرض استنتاج الحقائق منها على عبارات مبنية على الاحتمال. إذن علم الإحصاء يختص بدراسة البيانات الإحصائية أي جمعها وتحليلها بهدف الاستدلال واتخاذ القرارات في ظل عدم التأكد وذلك في مجالات الاقتصاد، الاجتماع .... وعلم الإحصاء من العلوم الحديثة نسبيا والتي ازدهرت في أوائل القرن العشرين، فهو يعتبر أداة هامة لجميع العلوم البحثة أو العلوم التطبيقية.

وينقسم الإحصاء إلى قسمين:

1-1- الإحصاء الوصفي:

هو العلم الذي يهتم بتلخيص وجمع البيانات عن ظاهرة معينة وعرضها في صورة جداول إحصائية أو رسومات بيانية تمهيدا لوصفها من خلال استخدام أدوات إحصائية مختلفة مثل مقاييس النزعة المركزية، مقاييس التشتت، مقاييس التمركز، الالتواء وغيرها.

من خلال هذا التعريف، يمكن أن نميز بين المفاهيم الأساسية والمختلفة للإحصاء كما يلي :

- أن علم الإحصاء الوصفي هو العلم الذي يسعى إلى تلخيص وجمع البيانات إضافة إلى عرضها.
- وعلم الإحصاء التحليلي الذي يختص في تحليل البيانات بهدف الاستدلال واتخاذ القرارات في حالة عدم التأكد.
- أما الطريقة الإحصائية، فهي مجموعة الطرق العلمية لجمع البيانات وتبويبها، عرضها ووصفها وذلك بإمكانية التعبير عن الظواهر تعبيرا رقميا.

1-2- الإحصاء الرياضي - الاستدلالي:

هو الذي يهتم بتحليل البيانات و الاستدلال عليها باستخدام نظرية الاحتمالات، ونظريات محددة كنظرية التقدير – التنبؤ– (Estimation) ونظرية اختبارات الفروض (Teste des Hypothèses) وغيرها.

2- وظائف علم الإحصاء :

تتلخص وظائف علم الإحصاء أو ما نسميها بخطوات المنهج الإحصائي في النقاط التالية:

- تحديد المشكلة عن ظاهرة معينة المرغوب دراستها بالإضافة إلى تحديد نوع البيانات اللازمة لدراستها وطرق تحليلها.
- جمع وتنظيم البيانات، و يقصد بها تحديد طرق جمع البيانات سواء حصر كامل أو أخذ بيانات بطرق معينة، ويقصد به أسلوب العينات مع الأخذ بعين الاعتبار طبيعة المجتمع المدروس وطبيعة البيانات المطلوبة.

- تلخيص البيانات وتبويبها ثم عرضها في شكل جداول إحصائية أو رسومات بيانية أو أشكال هندسية مختلفة، وهي مرحلة تساعد فيما بعد على تحليل هذه البيانات.
- وصف البيانات بعد تلخيصها في مجموعة من المقاييس الوصفية التي تصف توزيع الظاهرة موضوع البحث بطريقة مختصرة، وقياس درجة تباين أو عدم تجانس توزيع بيانات هذه الظاهرة.
- التفسير الإحصائي: بعد الانتهاء من مرحلة وصف البيانات، تكون أمام الباحث نتائج رقمية محددة. وبعد التأكد من صحتها يتعين عليه تفسير هذه النتائج بحكم خبرته وتجربته تتوافق مع طبيعة التحليل الإحصائي.

3- المتغيرات الإحصائية وأنواعها:

تعرف البيانات الإحصائية بأنها مجموعة الحقائق والمعلومات اللازمة المتعلقة بظاهرة ما موضوع الدراسة، والمتغيرات تعبر عن قيمة هذه البيانات فهي تنقسم إلى قسمين:

3-1- متغيرات كيفية (وصفية):

هي المتغيرات التي تكون وحداتها غير قابلة للقياس، مثل: اللون، الجنس، المستوى التعليمي، الحالة المدنية، الجنسية ...الخ. ويطلق عليها كذلك اسم المتغيرات النوعية.

3-2- متغيرات كمية (رقمية):

هي المتغيرات القابلة للقياس مثل العمر، الطول، الوزن، الكميات .... الخ. وتعتبر هذه البيانات الأساس والمادة الخام (الأولية) لعلم الإحصاء.

والمتغيرات الكمية نوعان : متغيرات منقطعة ومتغيرات مستمرة.

1)- متغيرات منقطعة (بيانات منفصلة):

هي تلك المتغيرات التي تأخذ قيما ثابتة في شكل أعداد صحيحة لا يمكن تجزئتها. مثل عدد الطلبة، عدد السكان وعدد القطع المنتجة...

2)- متغيرات مستمرة (بيانات متصلة):

وهي تلك المتغيرات التي وحدات قياسها تقبل التجزئة، حيث تكون القيم ضمن مجال معين، أي أنها قد تكون أرقاما كسرية وعشرية غير صحيحة. مثل علامات الطلبة، الأطوال، أجور العمال وغيرها.

ويمكن تقسيم البيانات الإحصائية من وجهة نظر أخرى إلى قسمين كما يلي:

أ)- البيانات الأولية (الخام): هي تلك البيانات الأولية التي يتم الحصول عليها في بداية دراسة موضوع البحث، وتسمى بيانات خام نسبة إلى المعادن في شكلها الخام في البداية قبل تحويلها إلى أشكال معينة، ويمكن أن تكون كيفية أو كمية.

ب)- البيانات المبوبة: وهي بيانات مرتبة ومنظمة في جداول إحصائية يتم الحصول عليها من بيانات أولية بعد تلخيصها، جمعها وترتيبها ثم عرضها بالطرق الإحصائية. ولا تكون هذه البيانات إلا رقمية باعتبارها الأساس في العلم والتحليل الإحصائيين.

4- مصادر البيانات الإحصائية:

يمكن تقسيم مصادر البيانات إلى قسمين أساسيين:

4-1- المصادر المباشرة (الأولية):

وهي الحصول على البيانات حول ظاهرة مدروسة من مصادر أولية أو أصلية وذلك عن طريق جمعها مفردا بالاتصال مع جهات مختلفة من المجتمع. وتعد البيانات في شكلها الأصلي أفضل من المصادر الثانوية لأنها تحتوي على تفسيرات وتوضيحات عن حالة مجتمع مدروس بالإضافة إلى أنه يمكن حدوث أخطاء نتيجة نقل البيانات أو تفسيرها.

4-2- المصادر غير المباشرة (الثانوية):

في كثير من الأحيان، يجد الباحث نفسه أمام صعوبة لجمع البيانات، لذا يلجأ إلى جداول غير مباشرة أي التي تم جمعها مسبقا عن ظاهرة ما -اقتصادية كانت أو اجتماعية، صحية أو ثقافية، علمية أو تربوية- وذلك لاستحالة الحصول عليها أو لسريتها أو لعدم كفايتها لإجراء الدراسة المطلوبة. مثل دائرة الإحصاءات العامة ودائرة الأحوال المدنية، تقارير البنك المركزي، الوزارات والمؤسسات المختلفة وغيرها.

5- طرق جمع البيانات:

بعد تحديد مصدر اقتناء البيانات الإحصائية، يتم جمعها وفق أساليب إحصائية معينة نلخصها في طريقتين كالتالي:

5-1- طريقة المسح (الحصر) الشامل:

هي طريقة يتم من خلالها جمع البيانات عن ظاهرة مدروسة من جميع وحدات المجتمع الإحصائي سواء كان محدودا أو مجاله واسعا. فهي تستلزم توافر إمكانيات مادية وبشرية أكبر نسبيا من طريقة العينات، كما تعد أحسن الطرق لأنها تعطي نتائج دقيقة ومفصلة.

5-2- طريقة العينات (أو المعاينة):

في بعض الحالات يستحيل الاعتماد على طريقة المسح الشامل لكون المجتمعات لا يمكن حصرها، لذا نلجأ إلى طريقة أخرى وهي العينات.

وتعرف العينة بأنها جزء من المجتمع يتم اختيارها بطرق معينة، فهي تسمح بإجراء دراسة أكثر تفصيلا وبالتالي أكثر دقة إذا ما استخدمت على أسس علمية سليمة. فباستخدام نظرية الاحتمالات، يمكن التحكم في خطأ المعاينة مما يزيد من دقة النتائج الممكن تعميمها.

6- المفاهيم الأساسية للإحصاء :

هناك بعض المفاهيم التي يجب على كل باحث أن يعرفها ولا تستعمل في مجال غير الإحصاء، وهي:

6-1- المجتمع الإحصائي:

هو مجموعة المشاهدات التي تخص ظاهرة معينة، فهو يحتوي على كافة مفردات الظاهرة المدروسة. والمجتمع إما أن يكون من النوع المحدود، أو من النوع غير المحدود :

6-1-1- المجتمع المحدود:

هو المجتمع الذي يمكن حصر جميع أفراده. مثال ذلك مجموعة من أطفال مدينة ما، أو عدد الطلبة في جامعة معينة.

6-1-2- المجتمع غير المحدود:

وهو المجتمع الذي يصعب حصر جميع وحدات الظاهرة، مثلا البكتيريا الموجودة في كمية من اللبن أو عدد النقاط التي يتكون منها الخط المستقيم.

6-2- الوحدة الإحصائية:

هي تلك الوحدة التي يتكون منها المجتمع الإحصائي.

مثلا: عند دراسة علامات الطلبة في مادة الإحصاء في قسم علوم التسيير يجب التمييز بين:

- المجتمع الإحصائي: مجموع الطلبة على مستوى الكلية
- الوحدة الإحصائية: الطلبة (ذكور – إناث)

6-3- الظاهرة الإحصائية:

هي التعبير عن المتغير محل الدراسة، وتسمى بالخاصية المدروسة. مثلا علامات الطلبة في امتحان، عدد الأهداف المسجلة في الدور الكروي، تسجيل المواليد حسب أعمار الأمهات.

6-4- العينة:

عند استحالة أخذ المجتمع الإحصائي ككل، يمكن أن نلجأ إلى اتخاذ جزء منه يسمى بأسلوب العينات.

6-4-1- تعريف:

هي جزء من المجتمع الإحصائي تؤخذ منه بطريقة ما بحيث تكون ممثلة له تمثيلا صحيحا، وتدرس بقصد التعرف على خصائص المجتمع المأخوذة منه. والمعاينة الإحصائية هي استخدام الطرق والنظريات الإحصائية العلمية في أخذ العينة ودراستها. ومن شروط أسلوب المعاينة ما يلي:

- أن تكون العينة كبيرة بدرجة كافية لكي يتم إهمال القيم المتطرفة على المتوسط، فكلما زاد عدد بيانات العينة كلما زادت درجة الثقة في النتائج المتحصل عليها.
- يجب أن يكون اختيار العينة اختيارا عشوائيا، بمعنى أن كل مفردة من المجتمع يكون لها نفس الفرصة في أن تختار لتكوين العينة.

وعن الأسباب التي تتطلب استخدام أسلوب المعاينة فهي:

- وفرة التكاليف المادية والوقت والجهد.
- في حالة ما إذا كانت طبيعة مجتمع غير محدود نظرا لصعوبة واستحالة حصر جميع أفراده، فهذا يتطلب استخدام العينة.
- هو أسلوب عملي، لأنه في بعض الحالات يؤدي فحص المفردات إلى تدميرها. مثلا عند القيام بعملية المسح الشامل لدم المريض، فهذا يعني سحب كل دمه من أجل إجراء بعض الفحوصات مما يؤدي إلى موت المريض، ففي هذه الحالة يجب من أخذ عينة من دم المريض وفحصها.

وأما عن نتائج العينة، فهي تتطابق مع نتائج المسح الشامل بالإضافة إلى أنها تعطي نتائج سريعة نتيجة لسرعة الحصول على البيانات وسرعة تحليلها.

6-4-2- أنواع العينات:

هناك عدة أنواع للعينات يتم اختيارها حسب العوامل التالية: طبيعة المجتمع الإحصائي، الظاهرة المدروسة، التجانس بين مفردات المجتمع والاستخدامات المتوقعة للنتائج.

تنقسم العينات إلى نوعين:

1)- النوع الأول: العينات العمدية أو الغرضية:

وهي العينات التي يتم سحبها بطريقة غير عشوائية ولكن حسب رغبة الباحث في انتقاء البيانات الإحصائية وجمعها، ويراد استخدامها للحصول على تقديرات تقريبية لظاهرة معينة في وقت سريع. كما يتم استخدامها في الأبحاث المتعلقة بالاقتصاد أو بإدارة الأعمال.

2)- النوع الثاني: **العينات العشوائية**، وهناك من يطلق عليها اسم العينات الاحتمالية. ونعني بالعشوائية الاختيار العشوائي وإعطاء نفس الفرصة لجميع مفردات المجتمع للظهور في العينة.

الفصـل الثانـي

العرض الجدولي والبيانـي

المقدمة:

يعتبر جمع البيانات من أهم المراحل التي يلجأ إليها الباحث الإحصائي وذلك من أجل استعمالها في دراسات مختلفة ومتعددة، لكن استعمالها في حين الحصول عليها يعد أمرا مستحيلا، لكن بعد عملية التصنيف والجدولة يمكن تطبيق المؤشرات الإحصائية المختلفة عليها بغرض وصفها أو تحليلها.

ويتم عرض البيانات بطريقتين مختلفتين وهي: الجداول الإحصائية والرسومات البيانية.

أولا: العرض الجدولي للإحصائيات

وهو عرض البيانات بعد تجميعها وتصنيفها في جداول تلخص الظاهرة المدروسة، لكي يتسنى للباحث تطبيق مختلف المقاييس لوصف هذه البيانات.

1- الجدول الإحصائي:

يظهر الجدول معلومات أو بيانات فردية، أي عن كل وحدة على حدة بينما الجدول الإحصائي يظهر المعلومات بعد تجميعها وتصنيفها في فئات معينة قد تكون فئات زمنية أو وصفية أو كمية، وبهذا عند عرض البيانات في الجداول الإحصائية تفقد قيمتها و معالمها الشخصية وتصبح مجرد رقم معين في فئة معينة.

كما يقصد بالجدول الإحصائي ترتيب البيانات الإحصائية في عمودين أو أكثر يقصد بها إبراز أهمية تلك البيانات وتسهيل المقارنة بينها. فالأول يتضمن صفة معينة من صفات الظاهرة الإحصائية المدروسة، أما الآخر فهو يضم عدد البيانات التي تحتويها هذه الصفة.

2- عناصر الجدول الإحصائي وشروطه:

عند تصميم أي جدول إحصائي يجب مراعاة القواعد التالية:

2-1- عنوان الجدول:

يجب أن يظهر في أعلى الجدول، كما يجب أن يكون واضحا بالإضافة إلى تحديد القياسات، أي الوحدة المستخدمة (مثلا: الدينار، آلاف الدنانير، الطن، الكلغ، .....).

مثلا: تسجيل الطلبة الجدد بجامعة بومرداس

2-2- عناوين الأعمدة:

كل عمود في الجدول يجب أن يكون له عنوان يوضح الأرقام والبيانات المدونة فيه توضيحا كاملا وبإيجاز. مثلا: عدد السكان.

2-3- عناوين الأسطر:

إن وصف الأرقام لكل سطر يكتب في العمود (1) من الجهة اليمنى من الجدول، فإذا كانت عبارة عن فترات زمنية يجب أن نبدأ بالفترات الأقدم ثم نتسلسل زمنيا حتى نصل إلى أحدث فترة.

2-4- المجاميع:

يجب أن يظهر الجدول المجاميع الخاصة بالأرقام الموجودة فيه كلما كان ذلك أمرا ممكنا كما يجب التحقق من صحة هذه المجاميع.

2-5- النسب المئوية:

إذا كانت الأرقام في عمود ما عبارة عن نسب مئوية يجب أن نشير ذلك في عنوان العمود بالرمز المئوي (%).

مثال (1): لدينا الجدول التالي يبين توزيع التلاميذ في مدرسة ما حسب الطور الدراسي:

جدول رقم (1): جدول إحصائي للطور الدراسي للتلاميذ

الطور	التكرار المطلق	التكرار النسبي	التكرار المئوي
الأول	220	0.32	%32
الثاني	150	0.21	%21
الثالث	330	0.47	%47
المجموع	700	1.00	%100

ولإيجاد النسبة المئوية المقابلة لكل رقم من التكرار، نقسم هذا الأخير على مجموع التكرارت مضروبا في 100.

مثلا: الطور الأول = 0.32 × 100= 32%

يسمى هذا الجدول بالجدول الإحصائي البسيط كونه لا يستدعي الجهد الكبير لإنشائه ولا العمليات المعقدة.

3- جدول التوزيع التكراري:

إن الغرض من عمل جدول التوزيع التكراري هو ترتيب وتنظيم البيانات حتى نتمكن من استخلاص النتائج والحقائق من هذه البيانات وبالتالي عرضها في جدول يسمى "**جدول التوزيع التكراري**".

وهو جدول يتكون أساسا من عمودين، الأول خاص بالمتغير المدروس أو الفئات، أما الثاني فهو خاص بالتكرارات ويستخرج من هذين العمودين أعمدة إضافية تستعمل في وصف البيانات الإحصائية وتحليلها. ويتم تكوين هذا الجدول حسب طبيعة البيانات الخاصة بالمتغير الإحصائي المعني بالدراسة.

3-1- جدول التوزيع التكراري للمتغير الكمي المنقطع:

سبق وأننا قمنا بتعريف المتغير الكمي المنقطع بأنه ذلك المتغير الذي لا يمكن تجزئة قيمه حيث يأخذ قيما صحيحة، لذلك يظهر في الجدول الإحصائي بصفة متسلسلة ومرتبة من أصغر قيمة إلى أكبرها وهذا في العمود الأول. أما العمود الثاني، فهو يمثل عدد المرات التي تكررت بها كل متغيرة.

مثال (2): يبين الجدول التالي توزيع علامات 100 طالب في مادة المحاسبة:

جدول رقم (2) توزيع علامات الطلبة

العلامات	عدد الطلبة
1	6
5	4
9	17
13	35
15	15
18	20
20	3
المجموع	100

3-2- جدول التوزيع التكراري للمتغير الكمي المستمر:

في حالة ما إذا كان مدى البيانات صغيرا، فإننا نقوم ببناء توزيع تكراري لحالة كمية منقطعة. أما في حالة ما إذا كان عدد البيانات الإحصائية كبيرا، فإننا نقوم بتقسيمها إلى مجموعات متجانسة تقريبا وتسمى "**بالفئات**" بحيث يحدد عدد التكرارات في كل فئة.

ويمكن تلخيص خطوات جدول التوزيع التكراري فيما يلي:

3-2-1- المدى العام:

يعرف المدى بأنه الفرق بين أكبر قيمة وأصغر قيمة في البيانات الإحصائية.

المدى العام (أو المطلق) = أكبر قيمة – أصغر قيمة

ويرمز له بالرمز:

$$E = X_{max} - X_{min}$$

$$E = X_n - X_1$$

حيث: X_n هي أكبر قيمة، و X_1 هي أصغر قيمة

3-2-2- الفئة:

هي عبارة عن حدين لمجال الدراسة وهما الحد الأدنى والحد الأعلى حيث توجد ضمنه تكرارات.

- الحد الأدنى: هو أصغر قيمة تتضمنها الفئة ونرمز له بالرمز (A)
- الحد الأعلى: وهو أكبر قيمة تتضمنها الفئة ويرمز له بالرمز (B)

ويجب أن نشير إلى أنه لا يوجد عددا محددا للفئات، وإنما يبقى تحديد عددها حسب تقدير الباحث مراعيا في ذلك طبيعة البيانات.

3-2-3- مركز الفئة:

هو منتصف الفئة، يلخصها أو يعبر عنها، ويلجأ إليها لتسهيل تطبيق المؤشرات الإحصائية وإجراء العمليات الحسابية، فهو يساوي إلى:

$$\text{مركز الفئة} = \frac{\text{الحد الأدنى} + \text{الحد الأعلى}}{2}$$

ويرمز له بالرمز " X_i " حيث: $X_i = \dfrac{A+B}{2}$

ونشير إلى أنه في حالة جداول مفتوحة (أي تلك الجداول التي لا تحتوي على حد أدنى للفئة (1) أو الحد الأعلى للفئة الأخيرة) فإنه لا يمكن تحديد وحساب مركز الفئة.

3-2-4- طول الفئة:

وهو عبارة عن الفرق بين الحد الأعلى للفئة والحد الأدنى لنفس الفئة ويرمز له بالرمز "K". وبالاعتماد على معادلة ستورجس التي= 1 + 3.322 لغ (n) يمكن تحديد طول الفئة كما يلي:

$$K = \frac{E}{1+3.322\log(n)}$$

حيث E : هو المدى العام، n : عدد البيانات، $\log(n)$: اللوغاريتم العشري لعدد القيم.

نعين حدود الفئات كما يلي:

- **الفئة (1):** نأخذ أصغر قيمة في البيانات باعتبارها الحد الأدنى ثم نضيف لها طول الفئة فنحصل على الحد الأعلى.
- **الفئة (2):** الحد الأعلى للفئة السابقة يصبح حد أدنى لهذه الفئة إذ نضيف لها طول الفئة فنحصل على الحد الأعلى ... وهكذا إلى أن نصل إلى أكبر قيمة في المعطيات الإحصائية.

ومن الأفضل أن يكون طول الفئة عددا صحيحا، وفي حالة ما إذا كان عددا عشريا، فإننا نقوم بتقريب القيمة إلى عدد صحيح.

ملاحظة: إذا كان لدينا مركز الفئة وطول الفئة وطلب منا إيجاد حدود الفئات، فيمكن الاعتماد على العلاقتين التاليتين:

– الحد الأدنى = مركز الفئة – طول الفئة

– الحد الأعلى = مركز الفئة + طول الفئة

وهكذا لكل الفئات المتبقية التي تُكوّن الجدول الإحصائي التكراري.

مثال (3): تم تسجيل طول قامة 50 طالب لأحد الأفواج من قسم علوم التسيير كما يلي:

170	169	172	169	174	172	167	173	180	160
163	171	177	171	176	175	170	173	168	165
170	175	183	174	183	162	171	174	174	168
173	182	161	175	172	167	173	177	164	170
179	174	167	176	176	170	172	181	166	172

المطلوب إيجاد طول الفئة لهذه البيانات باستعمال قاعدة ستورجس ؟

$$K = \frac{E}{1 + 3.322\log(n)}$$

$$E = X_n - X_1 = 183 - 160 = 23$$

$$K = \frac{E}{1 + 3.322\log(n)} = 3.46 \cong 3$$

جدول رقم (3): جدول توزيع تكراري لأطوال قامات الطلبة

الفئات	التكرارات
]163 - 160]	3
]166 - 163]	3
]169 - 166]	6
]172 - 169]	10
]175 - 172]	14
]178 - 175]	8
]181 - 178]	2
]184 - 181]	4
المجموع	50

3-2-5- التكرار النسبي والتكرار المئوي: (Fréquence Relative et Pourcentage)

وهو النسبة بين التكرار المطلق (العادي) لكل فئة (n_i) ومجموع التكرارات ($\sum n_i$). ويلاحظ أن مجموع التكرارات النسبية = الواحد الصحيح.

$$\text{التكرار النسبي} = \frac{\text{التكرار المطلق للفئة}}{\text{مجموع التكرارات}}$$

أما التكرار النسبي المئوي، فهو التكرار النسبي مضروبا في 100، وكما هو معلوم، فإن مجموع التكرارت المئوية = 100%

$$\text{التكرار النسبي المئوي} = \frac{\text{التكرار المطلق للفئة}}{\text{مجموع التكرارات}} \times 100$$

3-2-6- التكرارت التجميعية: (Fréquence Cumulée)

يحتاج الباحث في بعض الأحيان إلى معرفة عدد المتغيرات التي تزيد أو تساوي عن قيمة معينة أو تكون أصغر منها وذلك قصد المقارنة والتفسير. وللحصول على هذه المعلومات نلجأ إلى تجميع التكرارات بنوعيها الصاعدة والنازلة.

1)- التكرار التجميعي (أو المتجمع) الصاعد: (Fréquence Cumulée Croissante)

التكرار التجميعي الصاعد لأي فئة هو عبارة عن تكرار هذه الفئة مضافا إليه مجموع تكرارات الفئات السابقة. ويجب أن يكون تكرار الفئة الأخيرة مساويا لمجموع التكرارات.

2)- التكرار التجميعي (أو المتجمع) النازل: (Fréquence Cumulée Décroissante)

التكرار التجميعي النازل لأي فئة هو عبارة عن مجموع التكرارات مطروحا منه تكرارات الفئات السابقة. ويجب أن يكون التكرار المتجمع النازل للفئة الأخيرة مساويا لتكرارها الأخير المطلق.

3)- التكرار التجميعي (المتجمع) النسبي الصاعد: (Fréquence Cumulée Relative Croissante)

يحسب بقسمة التكرار المتجمع الصاعد لكل فئة على مجموع التكرارات في 100.

4)- التكرار التجميعي (المتجمع) النسبي النازل:

(Fréquence Cumulée Relative Décroissante)

ويحسب بقسمة التكرار المتجمع النازل لكل فئة على مجموع التكرارات في 100.

مثال (4): بالرجوع إلى المثال رقم (3) السابق، أحسب التكرارات التجميعية الصاعدة والنازلة لكل من التكرارات المطلقة، النسبية والمئوية.

جدول رقم (4): جدول توزيع تكراري لأطوال قامات الطلبة

الفئات	التكرار	ت ن	%	ت ت ص	ت ت ن	ت ن ص	ت ن ن	ت م ص	ت م ن
160 163	3	0.06	6	3	50	0.06	1	6	100
163 166	3	0.06	6	6	47	0.12	0.94	12	94
166 169	6	0.12	12	12	44	0.24	0.88	24	88
169 172	10	0.20	20	22	38	0.44	0.76	44	76
172 175	14	0.28	28	36	28	0.78	0.56	78	56
175 178	8	0.16	16	44	14	0.88	0.28	88	28
178 181	2	0.04	4	46	6	0.92	0.12	92	12
181 184	4	0.08	8	50	4	1	0.08	100	8
المجموع	50	1	100	–	–	–	–	–	–

4- الجداول غير المنتظمة:

الجداول المنتظمة هي تلك الجداول التي تكون أطوال الفئات فيها متساوية كما في الجدول رقم (4). أما الجداول غير المنتظمة، فهي تلك الجداول التي تكون أطوال الفئات فيها غير متساوية ويلجأ إلى مثل هذه الجداول في حالة وجود بعض الفئات بها تكرارات قليلة أو منعدمة. لذا، يفضل تبويب هذه الظواهر في فئات غير متساوية. ولا توجد قاعدة عامة هنا يعتمد عليها الباحث في تحديد أو افتراض عدد الفئات، وإنما هو راجع لحجم البيانات وتجانسها بالإضافة إلى النتيجة المراد الوصول إليها.

كما يمكن أن نقوم بتحويل التوزيع التكراري غير المنتظم إلى توزيع منتظم وذلك بتعديل التكرارات المقابلة لكل فئة حسب العلاقة التالية:

$$\text{التكرار المعدل} = \frac{\text{التكرار المطلق للفئة}}{\text{طول الفئة}}$$

مثال (5): فيما يلي جدول تكراري لحجم الودائع في أحد البنوك بآلاف الدنانير لعينة من عملاء هذا البنك:

جدول رقم (5) يبين توزيع تكراري لحجم الودائع

فئة الودائع	عدد العملاء	التكرار المعدل
]50 - 10]	150	3.75
]100 - 50]	300	6
]200 - 100]	500	5
]400 - 200]	100	0.50
]600 - 400]	50	0.25
]1000 - 600]	50	0.125
المجموع	1150	/

5- الجداول المفتوحة:

هي تلك الجداول التي تكون فيها حدود الفئات غير معلومة، سواء كانت من الحد الأدنى من الفئة الأولى في الجدول، أو من الحد الأعلى من الفئة الأخيرة، أو من الجهتين في آن واحد أي الحد الأدنى من الفئة الأولى والحد الأعلى من الفئة الأخيرة كما هو مبين في الجدول التالي:

جدول رقم (6) يبين حالات التوزيع

الحالة (1)	الحالة (2)	الحالة (3)	الحالة (4)
[10 - 30[	أقل من 30	[10 - 30[	أقل من 30
[30 - 60[	[30 - 60[	[30 - 60[	[30 - 60[
[60 - 90[	[60 - 90[	[60 - 90[	[60 - 90[
[90 - 120[	[90 - 120[	[90 - 120[	[90 - 120[
[120 - 150[	[120 - 150[	120 فأكثر	120 فأكثر

- الحالة (1): هي حالة جدول مغلق.
- الحالة (2): وهي حالة جدول مفتوح من اليمين، أي من الحد الأدنى من الفئة الأولى.
- الحالة (3): جدول مفتوح من اليسار، أي من الحد الأعلى من الفئة الأخيرة.
- الحالة (4): وهي الحالة الأخيرة أين تبين أن الجدول مفتوح من الطرفين.

تتميز الجداول التكرارية المغلقة بإمكانية إيجاد مراكز فئتها وبالتالي تطبيق بعض المؤشرات الإحصائية عليها وذلك لبساطة إجراء العمليات الحسابية مثل إيجاد الوسط الحسابي، المدى، الانحراف المتوسط والمعياري وغيرها. أما الجداول المفتوحة فلا يمكن تطبيق عليها هذه المؤشرات وإنما تتطلب مقاييس بديلة مثل الوسيط والمدى الربيعي.

ثانيا: العرض البياني للإحصائيات

يعد البيان أكثر توضيحا للصفات المميزة للتوزيعات، لهذا يعتبر وسيلة أخرى لتلخيص وعرض البيانات الإحصائية. كما أنه يساعد في إجراء بعض التحليلات الإحصائية ومعرفة شكل التوزيع الإحصائي. ففي هذه الرسوم يمثل المحور الأفقي قيم الظاهرة المدروسة، أما المحور العمودي فتوضع به التكرارات بأنواعها المختلفة: المطلقة، النسبية، المئوية أو التجميعية.

وتختلف العروض البيانية من متغير إلى آخر حسب اختلاف طبيعة البيانات، أي إذا كانت البيانات كيفية أو كمية، منقطعة أو مستمرة، أو إذا كانت مبوبة أو غير مبوبة.

1- العرض في حالة بيانات غير مبوبة:

ونميز ثلاثة عروض: المستطيلات بأنواعها الثلاثة: البيانية، المزدوجة والمجزأة (أو المركبة)، والدائرة النسبية.

1-1- المستطيلات:

هي عبارة عن مستطيلات ترسم بنفس المسافات، حيث تكون أطوالها مقابلة مع التكرارات المكونة للظاهرة المدروسة.

وتوجد عدة أشكال لعرض البيانات عن طريق المستطيلات نقدمها فيما يلي:

1-1-1- المستطيلات البيانية:

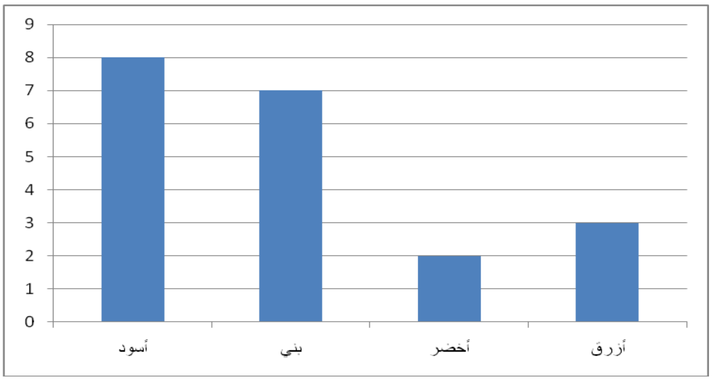
تستخدم في حالة وجود ظاهرة إحصائية واحدة.

مثال (7): تشير المعلومات التالية إلى لون العيون لعدد معين من الأشخاص:

أسود، أزرق، بني، أسود، بني، بني، بني، أخضر، أسود، بني، بني، أسود، بني، بني، أخضر، أسود، أزرق، أزرق، أسود، أسود، أسود.

المطلوب: الرسم بواسطة المستطيلات البيانية؟

شكل رقم (1): المستطيلات البيانية للون العيون



1-1-2- المستطيلات المزدوجة:

تستخدم في حالة وجود ظاهرتين أو أكثر، ويتم التمثيل هنا بعمودين متلاصقين أو أكثر مثل تقسيم الطلبة إلى ذكور وإناث، ودراسة توزيعهم الجغرافي على المناطق، أو دراسة مستواهم التعليمي...

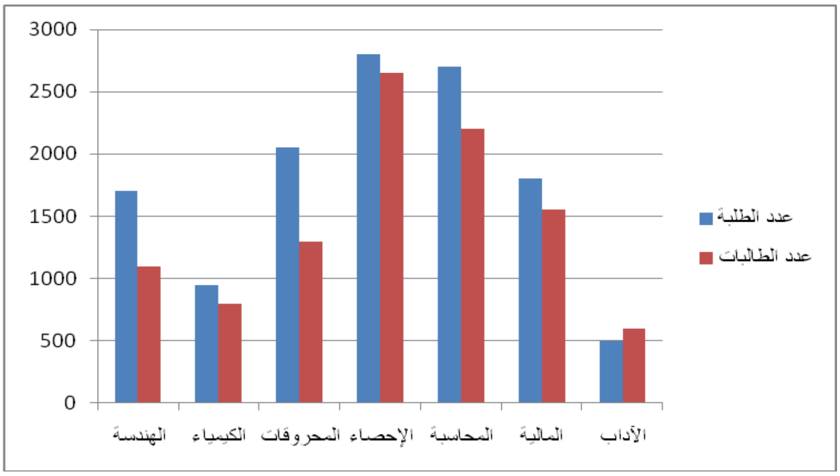
مثال (8): يمثل الجدول التالي توزيع عدد الطلبة والطالبات على الأقسام المختلفة بأحد الجامعات.

المطلوب: التمثيل بواسطة المستطيلات المزدوجة ؟

جدول رقم (7): توزيع الطلبة والطالبات على الأقسام

القسم	عدد الطلبة	عدد الطالبات
– الهندسة	1700	1100
– الكيمياء	950	800
– المحروقات	2050	1300
– الإحصاء	2800	2650
– المحاسبة	2700	2200
– المالية	1800	1550
– الآداب	500	600
المجموع	12500	10200

شكل رقم (2): المستطيلات المزدوجة لتوزيع الطلبة حسب الجنس

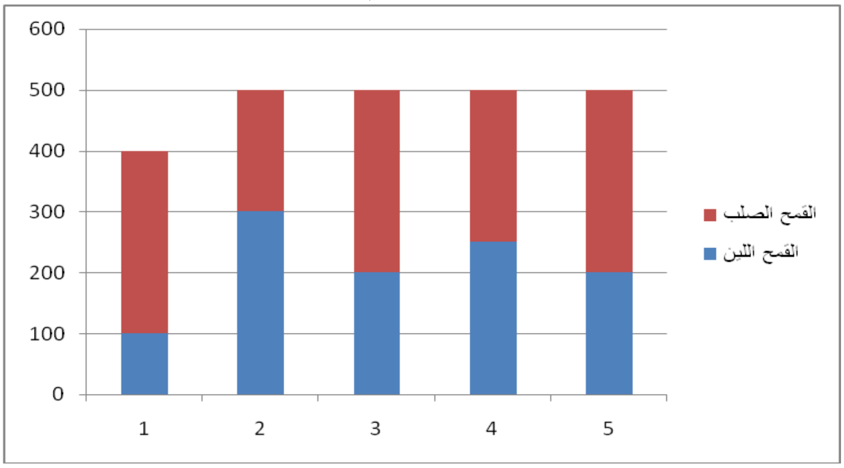


1-1-3- المستطيلات المركبة:

تستخدم في حالة ما إذا كانت هناك ظاهرة تتكون من عدة أجزاء من نوعيات مختلفة. وهو رسم الظاهرة في عمود واحد بحيث يمثل ارتفاع العمود مجموع قيم الأجزاء كما هو مبين في المثال التالي.

مثال (9): لدينا المعطيات التالية حول المحصول الزراعي المتكون من القمح الصلب والقمح اللين موزع حسب السنوات:

شكل رقم (3): تمثيل المحصول الزراعي بواسطة المستطيلات المركبة



1-2- العرض الدائري:

تستخدم الدائرة النسبية لعرض المتغيرات الكيفية، ويتم رسم هذا الشكل بوضع دائرة ذات قطر مناسب تقسم إلى أقسام جزئية بحيث تتناسب مساحة كل جزء إلى المجموع الكلي وذلك طبقا للعلاقة التالية:

$$\text{مساحة (زاوية) الجزء} = \frac{\text{قيمة الجزء}}{\text{مجموع الأجزاء}} \times 100$$

وعليه، تقسم الزاوية المركزية والتي تساوي 360^{o} تبعا لعدد الأجزاء التي تتكون منها الظاهرة، ويتم تلوين كل جزء بلون مختلف لتمييزه عن باقي الأجزاء الأخرى.

مثال (10): يبين الجدول التالي منشآت القطاع الخاص موزعة على مناطق مختلفة كما يلي:

جدول رقم (8): توزيع المنشآت حسب المناطق

المنطقة	الشمالية	الشرقية	الجنوبية	الغربية	المجموع
عدد المنشآت	24600	17200	10000	20200	72000

المطلوب: أعرض هذه البيانات بواسطة الدائرة؟

الحل: المجموع الكلي لعدد المنشآت هو 72000:

زاوية كل منطقة هي كما يلي:

1- المنطقة الشمالية = $\frac{24600}{72000} \times 360^{\circ} = 123^{\circ}$

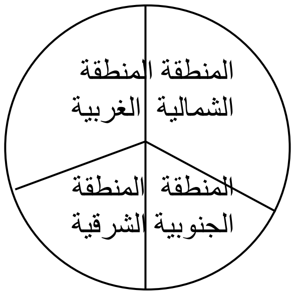
2- المنطقة الشرقية = $\frac{17200}{72000} \times 360^{\circ} = 86^{\circ}$

3- المنطقة الجنوبية = $\frac{10000}{72000} \times 360^{\circ} = 50^{\circ}$

4- المنطقة الغربية = $\frac{20200}{72000} \times 360^{\circ} = 101^{\circ}$

مجموع الزوايا $= 360^{\circ}$

شكل رقم (4): العرض الدائري لعدد المنشآت



2- العرض في حالة التوزيعات التكرارية:

تختلف طريقة عرض البيانات حسب طبيعتها، أي إذا كانت كيفية أو كمية، مبوبة أو غير مبوبة. فقد عرضنا مختلف الأشكال في حالة البيانات غير المبوبة والوصفية، وفيما يلي سوف نعرض جميع الرسومات البيانية لحالة المتغيرات الكمية، وتنقسم هذه الأخيرة بدورها إلى قسمين: متغيرات كمية منقطعة ومتغيرات كمية مستمرة.

وما نميزه هنا، هو أننا نقوم بوضح حدود الفئات الدنيا والعليا أو مراكز الفئات على المحور الأفقي ونضع بالمقابل التكرارات المطلقة أو التجميعية بنوعيها على المحور العمودي.

وسنعرض فيما يلي الأشكال المختلفة لكل من المتغير الكمي المنقطع والمتغير الكمي المستمر.

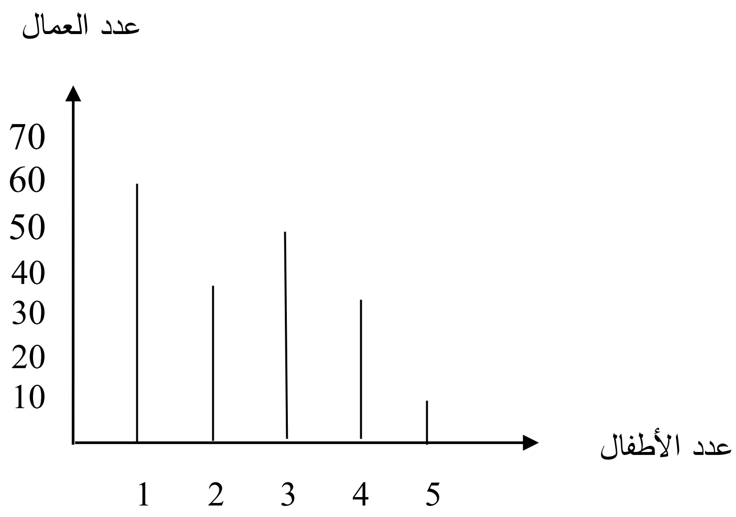
2-1- العروض البيانية في حالة المتغير الكمي المنقطع:

2-1-1- شكل الأعمدة:

يستعمل في حالة عرض المتغير الكمي المنقطع وتسمى بالأعمدة البسيطة.

مثال (11): تشير البيانات التالية إلى توزيع عدد العمال حسب عدد الأطفال لديهم:

شكل رقم (5): الأعمدة البسيطة لتوزيع العمال على عدد الأطفال

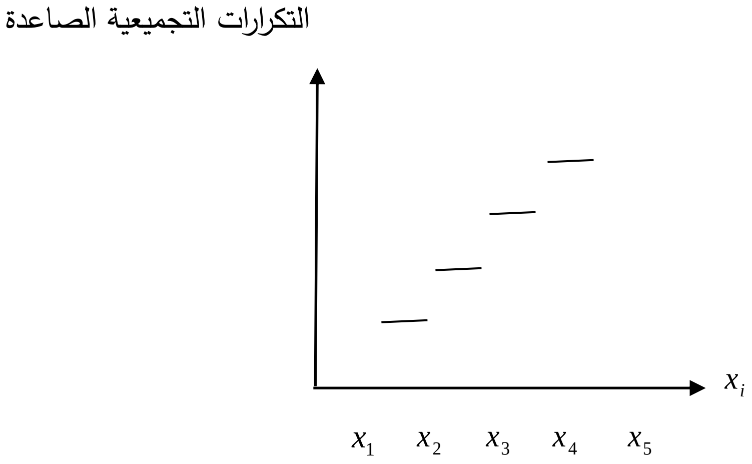


2-1-2- تمثيل التكرارات التجميعية للمتغير الكمي المنقطع:

1)- التكرارات التجميعية الصاعدة:

يعبر عنها بيانيا بواسطة قطع مستقيمة متصاعدة حسب قيم التكرارات التجميعية الصاعدة للظاهرة المدروسة.

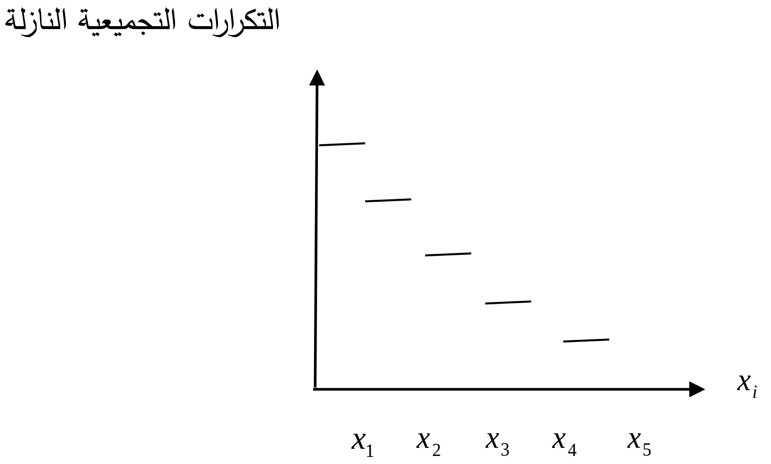
شكل رقم (6): تمثيل التكرارات التجميعية الصاعدة



2)- التكرارات التجميعية النازلة:

ويعبر عنها بواسطة قطع مستقيمة متنازلة حسب التكرارات التجميعية النازلة لقيم الظاهرة المدروسة.

شكل رقم (7): تمثيل التكرارات التجميعية النازلة



2-2- العروض البيانية في حالة المتغير الكمي المستمر:

ونميز هنا العديد من الأشكال البيانية وذلك حسب طبيعة هذه المتغيرات، وسوف نعرض كل من: المدرج التكراري، المضلع التكراري، المنحنى التكراري، بالإضافة إلى منحنى التكرارات التجميعية الصاعدة والنازلة.

2-2-1- المدرج التكراري:

هو عبارة عن مستطيلات متلاصقة فيما بينها تتناسب أطوال كل منها وتكرار كل فئة. ولرسم هذا المدرج، نضع حدود الفئات الدنيا والعليا، أو مراكز الفئات على المحور الأفقي والتكرارات – المطلقة أو النسبية – على المحور العمودي.

ويجب أن نشير إلى أن هذا النوع من الرسومات لا يتناسب عندما تكون الجداول مفتوحة أو غير منتظمة، حيث يجب غلق هذه الجداول أو تعديل تكراراتها. ومن هنا، فإن طريقة رسم المدرج التكراري تختلف في حالة ما إذا كانت لدينا فئات متساوية الأطوال عن غيرها إذا كانت عكس ذلك.

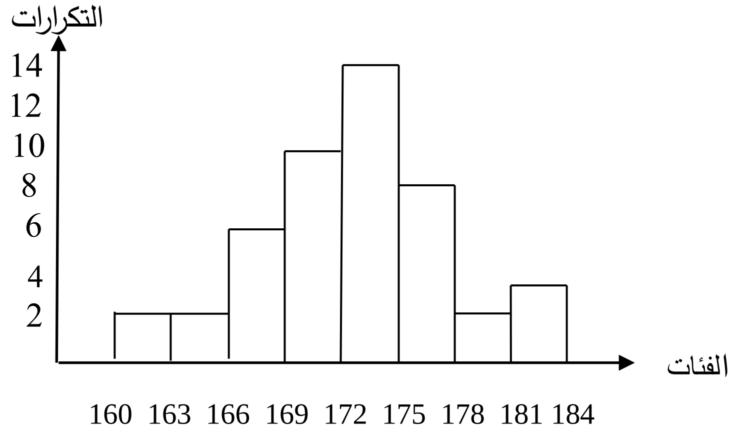
إذن، طريقة رسم المدرج التكراري تنقسم إلى قسمين:

1)- التوزيع المنتظم:

ويعني عندما تكون أطوال الفئات متساوية، حيث يتم رسم المدرج التكراري مباشرة.

مثال (12): بالرجوع إلى المثال رقم (3) السابق، أرسم المدرج التكراري لتوزيع أطوال قامات الطلبة.

شكل رقم (8): المدرج التكراري لأطوال الطلبة



2)- التوزيع غير المنتظم:

في هذه الحالة تكون الفئات غير متساوية الأطوال، وقبل عرض المدرج يجب من تعديل التكرارات وفق العلاقة التالية:

$$\text{التكرار المعدل} = \frac{\text{التكرار المطلق للفئة}}{\text{طول الفئة المقابل}}$$

ونرمز له بالرمز n_i^* كما يلي:

$$n_i^* = \frac{n_i}{k}$$

حيث: n_i هو التكرار المطلق

k هو طول الفئة

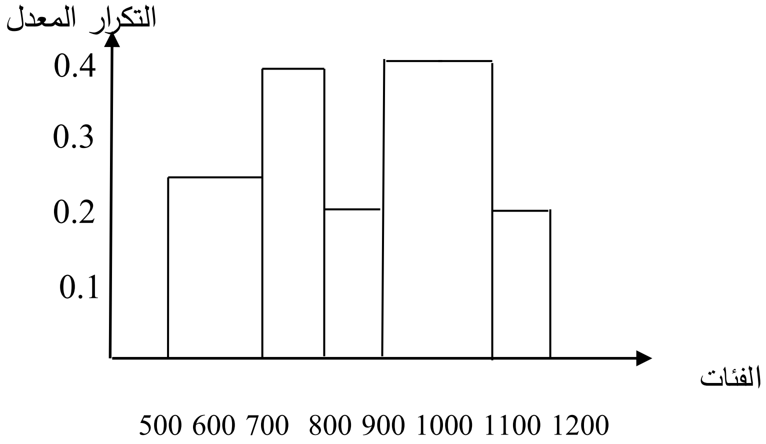
مثال (13): لدينا توزيع الأجر الشهري على العمال في مؤسسة ما كما يلي:

جدول رقم (9): توزيع الأجر الشهري على العمال

فئات الأجور	التكرار المطلق (n_i)	طول الفئة	التكرار المعدل (n_i^*)
700 - 500	50	200	0.25
800 - 700	40	100	0.40
1000 - 800	30	150	0.20
1100 - 1000	60	150	0.40
1200 - 1100	20	100	0.20
المجموع	200	-	-

المطلوب: إعداد المدرج التكراري؟

شكل رقم (9): المدرج التكراري لأجور العمال



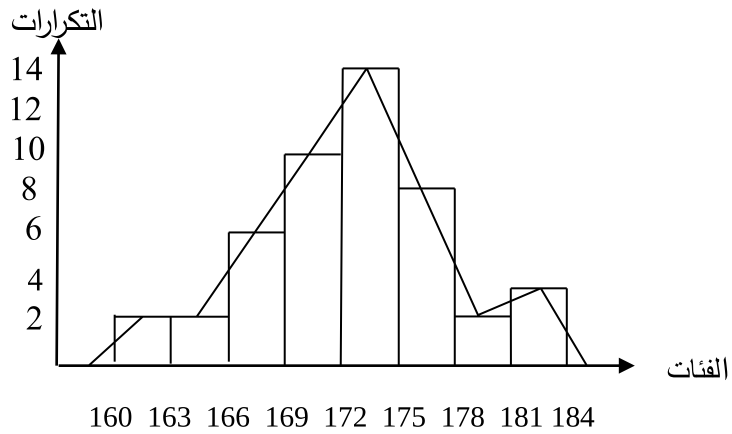
2-2-2- المضلع التكراري:

يمكن الحصول على المضلع التكراري من المدرج لنفس الشكل وذلك بتقسيم قمم المستطيلات إلى قسمين متساويين والتي تناظر مراكز الفئات حيث نقوم بتوصيل منتصفات هذه القمم بخطوط مستقيمة فيكون الشكل الناتج هو المضلع التكراري.

ويتم غلق هذا المضلع بافتراض مركز الفئة الأولى ومركز يلي مركز الفئة الأخيرة وأن تكرار كل منهما صفر.

مثال (14): نأخذ المثال رقم (12) ونرسم المضلع التكراري لتوزيع أطوال قامات الطلبة على المدرج التكراري.

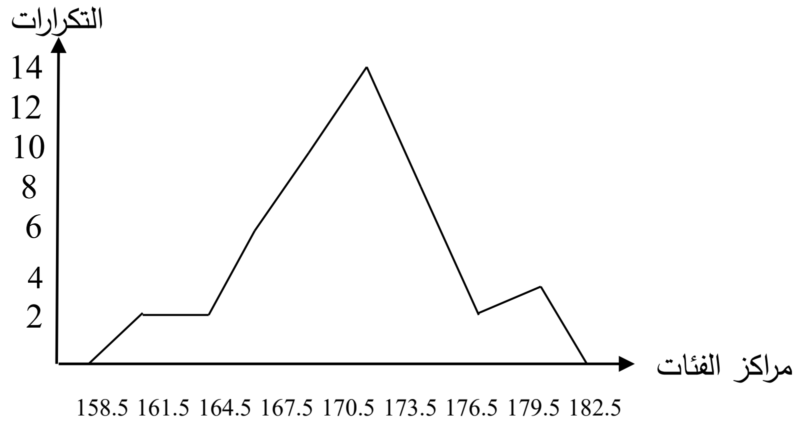
شكل رقم (10): المضلع التكراري لأطوال الطلبة



كما يمكن الحصول على المضلع خارج المدرج التكراري بحيث يخصص المحور الأفقي لمراكز الفئات والمحور العمودي للتكرارات.

مثال (15): بأخذ المثال السابق، أرسم المضلع التكراري خارج المدرج التكراري.

شكل رقم (11): المضلع التكراري خارج المدرج لأطوال الطلبة

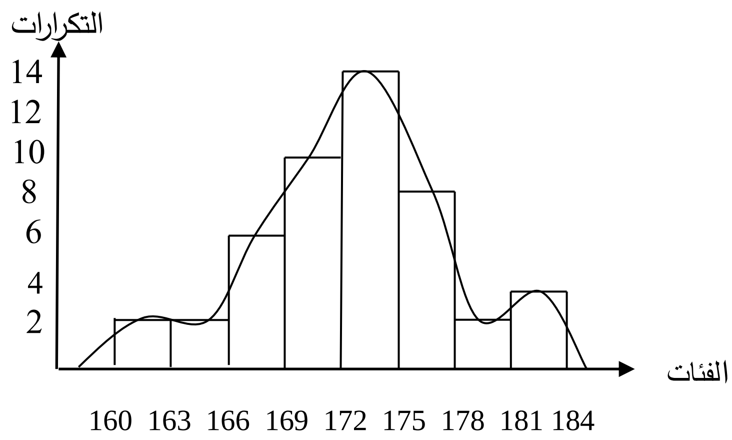


2-2-3- المنحنى التكراري:

يرسم بنفس الطريقة التي يرسم بها المضلع التكراري، وعوض أن نصل بين النقاط بقطع مستقيمة فإننا نقوم بتوصيلها باليد ويكون الشكل الناتج هو المنحنى التكراري.

مثال (16): أرسم المنحنى التكراري لأطوال قامات الطلبة ؟

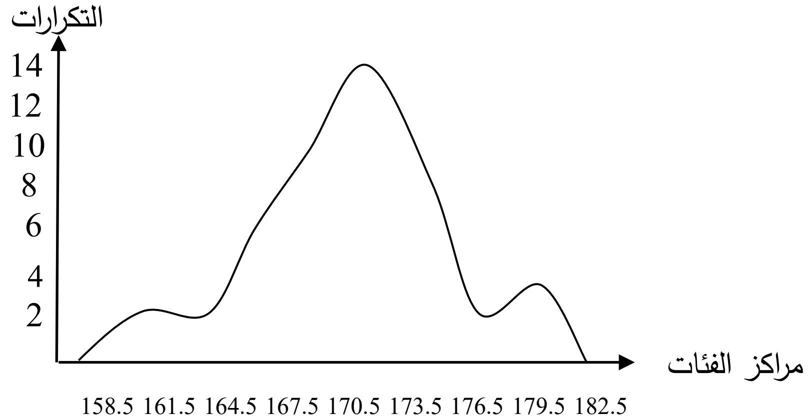
شكل رقم (12): المنحنى التكراري لأطوال الطلبة



وكما هو الحال بالنسبة للمضلع التكراري، فإن المنحنى التكراري يرسم أيضا خارج المدرج وتؤخذ مراكز الفئات في المحور الأفقي والتكرارات في المحور العمودي.

مثال (17): أرسم المنحنى التكراري لأطوال قامات الطلبة خارج المدرج التكراري ؟

شكل رقم (13): المنحنى التكراري خارج المدرج لأطوال الطلبة



2-2-4- تمثيل التكرارات التجميعية للمتغير الكمي المستمر:

يمكن تمثيل التكرارات التجميعية الصاعدة والتكرارات التجميعية النازلة بواسطة المنحنى كما يلي:

1)- تمثيل التكرارات التجميعية الصاعدة:

تمثل التكرارات التجميعية الصاعدة بمنحنى يسمى منحنى التكرار التجميعي الصاعد، ولرسمه نضع الفئات –أو مراكزها– على المحور الأفقي والتكرارات التجميعية الصاعدة على المحور العمودي، فنأخذ بعين الاعتبار الحدود العليا للفئات ونصلها مع التكرار المتجمع الصاعد الذي يقابلها وهكذا حتى نحصل على منحنى يتصاعد نحو الأعلى.

2)- تمثيل التكرارات التجميعية النازلة:

تمثل التكرارات التجميعية النازلة بمنحنى يسمى منحنى التكرار التجميعي النازل، ولرسمه نأخذ مراكز الفئات أو الحدود الدنيا للفئات وهذا على المحور الأفقي. بالمقابل نضع التكرارات التجميعية النازلة على المحور العمودي، ونحصل على منحنى متنازل نحو الأسفل.

كما يمكن أن نمثل كل من التكرارات التجميعية الصاعدة والتكرارات التجميعية النازلة في نفس الشكل فنحصل على تقاطع المنحنيين. والشيء الملاحظ هو أن هذين المنحنيين يتقاطعان في نقطة مقابلة لقيمة نصف مجموع التكرارات تسمى "الوسيط".

مثال (18): لدينا المثال حول أطوال قامات الطلبة كما يلي:

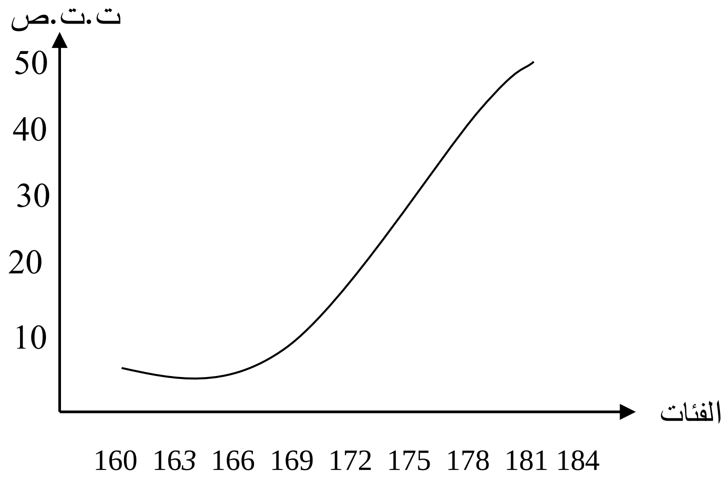
جدول رقم (10): جدول توزيع تكراري لأطوال قامات الطلبة

الفئات	التكرار	FCC	FCD
163 160	3	3	50
166 163	3	6	47
169 166	6	12	44
172 169	10	22	38
175 172	14	36	28
178 175	8	44	14
181 178	2	46	6
184 181	4	50	4
المجموع	50	–	–

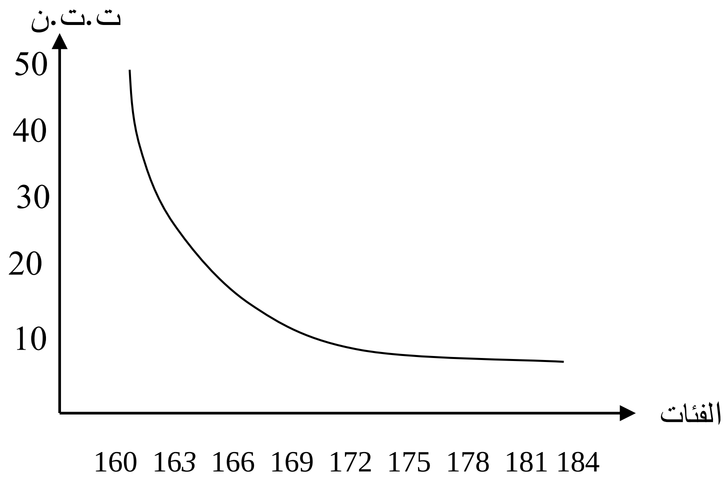
المطلوب: أرسم منحنى التكرار التجميعي الصاعد ومنحنى التكرار التجميعي النازل، ثم مثّل المنحنيين الصاعد والنازل في نفس البيان ؟

أولا: رسم منحنى التكرار التجميعي الصاعد والنازل:

شكل رقم (14): منحنى التكرار التجميعي الصاعد

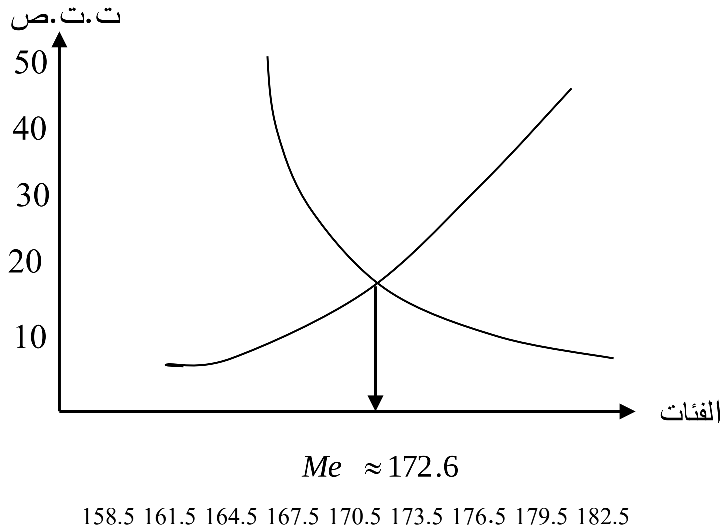


شكل رقم (15): منحنى التكرار التجميعي النازل



ثانيا: رسم منحنى التكرار التجميعي الصاعد ومنحنى التكرار التجميعي النازل:

شكل رقم (16): منحنى تقاطع التكرار التجميعي الصاعد والنازل



الفصـل الثالث

مقاييس النزعة المركزيـة

المقدمة:

بعد تجميع البيانات وتصنيفها وتمثيلها بيانيا، ننتقل إلى وصفها عن طريق إبراز الخصائص الأساسية لها والتي يمكن التعبير عنها بمقاييس محددة. ومن أهم هذه المقاييس هي تلك التي تحدد قيمة معينة من قيم المتغير المدروس وتكون هذه القيمة ممثلة للقيم الأخرى. تسمى هذه المقاييس بمقاييس النزعة المركزية وتسمى أحيانا بمقاييس التوسط أو التمركز.

ويمكن تعريف النزعة المركزية بأنها الرغبة في التمركز حول نقطة أو قيمة واحدة تسمى القيمة المتوسطة، وإن أهم مقاييسها هي:

- المتوسط (الوسط) الحسابي؛
- الوسيط والمقاييس الشبيهة بالوسيط؛
- المنوال؛

وسنتطرق إلى هذه المقاييس من خلال التعرف على كيفية استخراجها في حالة ما إذا كانت بيانات غير مبوبة أو بيانات مبوبة – توزيعات تكرارية – وتحديد أهم خصائصها ومجالات استخدامها بالإضافة إلى تحديد العلاقات التي تربط بين هذه المقاييس.

1- المتوسط الحسابي: (Moyenne Arithmétique **)**

يعتبر المتوسط الحسابي للقيم المختلفة التي يأخذها متغير ما هو القيمة الممثلة لجميع القيم التي حسب لها.

1-1- حالة البيانات غير المبوبة (المتوسط الحسابي البسيط):

يعرف المتوسط الحسابي لمجموعة من القيم على أنه حاصل قسمة مجموع هذه القيم على عددها، وهي الطريقة المباشرة. ويعتبر من أهم مقاييس التوسط وأكثرها استخداما.

فإذا كان لدينا قيم المشاهدات x_1 ، x_2 ، .... ، x_n فإن الوسط الحسابي لهذه القيم $\overline{X}$ هو:

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

وباستخدام مز المجموع نكتب:

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

حيث: i يتراوح من 1 إلى n

X_i هي قيم الظاهرة المدروسة

n هو عدد قيم الظاهرة

مثال (1): إذا كانت لدينا السلسلة العددية التالية: 33 22 17 13 8 5

فإن الوسط الحسابي البسيط لهذه السلسلة هو:

1-2- المتوسط الحسابي للتوزيعات التكرارية (المتوسط الحسابي المرجح):

يمكن إيجاد المتوسط الحسابي المرجح كما يلي:

نحدد الوسط الحسابي بتقسيم مجموع حاصل ضرب قيم المتغير الإحصائي بترجيحاتها، والترجيح هنا يمثل التكرار (n_i) ويدعى بالمتوسط الحسابي المرجح. فإذا كانت قيم المتغير الإحصائي في التوزيع التكراري تعطى على شكل مجالات أي فئات، فإن مركز الفئة يعد مقياس وصفي لكل تكرار.

وبالتعبير عن هذه العلاقة رياضيا نكتب:

$$\overline{X} = \frac{(x_1 \times x_0) + (x_2 \times x_0) + \ldots + (x_n \times x_0)}{n_1 + n_2 + \ldots + n_n}$$

وباستخدام رمز المجموع نجد:

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

مثال (3): لدينا نقاط امتحان مادة الإحصاء لـ 100 طالب موزعين حسب العلامات التي حصلوا عليها:

جدول رقم (11): جدول توزيع تكراري لعلامات الطلبة

العلامات (X_i)	عدد الطلبة (n_i)	$n_i X_i$
4	2	8
5	8	40
6	13	78
7	35	245
8	21	168
9	16	144
10	5	50
المجموع	100	733

المطلوب: إيجاد الوسط الحسابي المرجح لهذه القيم ؟

$$\overline{X} = \frac{\sum_{i=1}^{n} n_i X_i}{\sum_{i=1}^{n} n_i} = \frac{733}{100} = 7.33\,pts$$

1-3- خصائص المتوسط الحسابي:

- مجموع انحرافات القيم عن الوسط الحسابي يساوي إلى الصفر:

$$\sum\left(X_i-\overline{X}\right)=0$$
$$\sum\left(X_i-\overline{X}\right)=\sum X_i-\sum\overline{X}=\sum X_i-n\overline{X}=0$$

بقسمة الطرفين على n نجد:

$$\frac{\sum X_i}{n}-\frac{n\overline{X}}{n}=\frac{\sum X_i}{n}-\overline{X}=\overline{X}-\overline{X}=0$$

- يتأثر المتوسط الحسابي بالقيم المتطرفة (الشاذة)، في هذه الحالة يتم اللجوء إلى مقياس آخر.

مثلا: أوجد المتوسط الحسابي للقيم التالية: 10 13 21 18 1600

$$\overline{X}=\frac{\sum_{i=1}^{n}X_i}{n}=\frac{1662}{5}=332.4$$

فقيمة 332.4 بعيدة جدا عن قيم المتغير الإحصائي نتيجة للقيمة المتطرفة التي تساوي 1600؛

- إذا أضفنا (أو قمنا بضرب) مقدارا ثابتا لكل قيمة من قيم مجموعة من الأعداد، فإن الوسط الحسابي الجديد يساوي الوسط الحسابي للقيم الأصلية مضافا إليه (مضروبا بـ) نفس المقدار الثابت؛

- يستعمل الوسط الحسابي في دراسة المتغير الكمي أي القابل للقياس؛

- يعتبر من أكثر مقاييس النزعة المركزية استخداما، ويتميز بأخذه جميع المفردات بعين الاعتبار عند حسابه؛

- مجموع مربع انحرافات القيم عن متوسطها الحسابي يكون أقل ما يمكن، ويعبر عنه بالعلاقة الرياضية التالية:

$$\sum\left(X_i-\overline{X}\right)^2\rightarrow Min$$

2- الوسيط: (Médiane)

يعرف الوسيط على أنه تلك القيمة الأوسطية لمجموعة من القيم مرتبة ترتيبا تصاعديا، بمعنى آخر هو تلك القيمة التي تكوّن نسبة 50% من القيم أصغر منها و50% من القيم أكبر منها.

2-1- الوسيط للبيانات غير المبوبة:

توجد حالتان لحساب الوسيط:

- **الحالة (1):** إذا كان عدد القيم فرديا، نقوم أولا بترتيب هذه القيم تصاعديا ثمّ نجد رتبة الوسيط من العلاقة:

رتبة الوسيط = $\frac{n+1}{2}$، والوسيط هو تلك القيمة المقابلة لهذه الرتبة.

مثال (5): إذا كانت نقاط امتحان لـ 5 طلبة كما يلي: 6 19 13 16 10

نرتب ترتيبا تصاعديا: 6 10 13 16 19

والعلامة الوسيطية هي تلك القيمة ذات الرتبة $\frac{n+1}{2}$ أي $\frac{5+1}{2}=3$ والتي تقابل العلامة 13 ونرمز للوسيط بالرمز "**Me**"، أي نقطة Me = 13

- **الحالة (2):** إذا كان عدد القيم زوجيا، نجد رتبة الوسيط من العلاقة:

رتبة الوسيط = $\frac{n}{2}$ و $\frac{n}{2}+1$، والوسيط هو تلك القيمة المقابلة لهذه الرتبة.

مثال (6): بإضافة علامة سادسة ولتكن 20، فإن n عدد زوجي

رتبة الوسيط هي: $\frac{n}{2}=\frac{6}{2}=3$، و $\frac{n}{2}+1=\frac{6}{2}+1=4$. ومنه، فإن الوسيط يساوي $Me_1 = 13$ و $Me_2 = 16$

فقيمة الوسيط هي منتصف القيمتين المقابلة للرتبة 3 و4 أي:

$$Me = \frac{13+16}{2} = 14.5\,pts$$

2-2- الوسيط للبيانات ذات التوزيع التكراري:

لإيجاد الوسيط من التوزيع التكراري نتبع الخطوات التالية:

- نقوم بحساب التكرار التجميعي الصاعد؛
- نحدد رتبة الوسيط بتقسيم مجموع التكرارات على 2؛
- تحديد الفئة التي تحتوي على قيمة الوسيط، وهي تلك الفئة المقابلة للتكرار التجميعي الصاعد الذي يساوي رتبة الوسيط أو الأكبر منه مباشرة وتسمى بالفئة الوسيطية؛
- حساب الوسيط بتطبيق العلاقة الإحصائية التالية:

$$Me = A + \frac{\frac{\sum n_i}{2} - F\uparrow(n-1)}{n_{iMe}} \times K$$

حيث: A : الحد الأدنى للفئة الوسيطية

$\frac{\sum n_i}{2}$: رتبة الوسيط

$F\uparrow(n-1)$: التكرار التجميعي الصاعد السابق عن الفئة الوسيطية

n_{iMe} : التكرار المطلق المقابل للفئة الوسيطية

K : طول الفئة الوسيطية

مثال (7): يمثل الجدول التالي علامات مادة الإحصاء لعينة متكونة من 100 طالب.

المطلوب: إيجاد الوسيط؟

جدول رقم (13): جدول توزيع تكراري لعلامات الإحصاء

الفئات	التكرارات (n_i)	ت.ت.ص ($F\uparrow$)
6 - 8	3	3
8 - 10	10	13
10 - 12	26	39
12 - 14	33	72
14 - 16	14	86
16 - 18	8	94
18 - 20	6	100
المجموع	100	---

1)- إيجاد رتبة الوسيط: $\frac{\sum n_i}{2} = \frac{100}{2} = 50$

2)- تحديد الفئة الوسيطية: (وهي الفئة المقابلة للقيمة 50 أو أكثرها في التكرار التجميعي الصاعد).

الفئة الوسيطية هي: [12– 14[

3)- حساب الوسيط باستخدام العلاقة:

$$Me = A + \frac{\frac{\sum n_i}{2} - F\uparrow(n-1)}{n_{iMe}} \times K$$

$$Me = 12 + \frac{50-39}{33} \times 2 = 12.66\,pts$$

- ويمكن حساب الوسيط كما يلي:

بما أن الرتبة محصورة في الفئتين: الفئة الأولى: [10– 12[⇐ تقابل ت.ت.ص 39

الفئة الثانية: [12– 14[⇐ تقابل ت.ت.ص 72

$$Me = A + \frac{\frac{\sum n_i}{2} - F\uparrow(n-1)}{F\uparrow(n+1) - F\uparrow(n-1)} \times K$$

بتطبيق نفس المثال السابق نجد قيمة الوسيط تساوي:

$$Me = 12 + \frac{50-39}{72-39} \times 2 = 12.66\,pts$$

مثال (8): أوجد الوسيط للبيانات التالية:

جدول رقم (14): جدول توزيع تكراري للبيانات

المتغير (X_i)	التكرارات (n_i)	ت.ت.ص
4	2	2
5	8	10
6	13	23
$\boxed{7}$	35	58
8	21	79
9	16	95
10	5	100
المجموع	100	---

1)- رتبة الوسيط: $\frac{\sum n_i}{2} = \frac{100}{2} = 50$

2)- إيجاد قيمة الوسيط: الوسيط في حالة متغير كمي منقطع هو قيمة المتغير الإحصائي (X_i) المقابلة للتكرار التجميعي الصاعد الذي يحتوي على رتبة الوسيط، أي: $Me = 7$

2-3- إيجاد الوسيط بيانيا:

يمكن إيجاد الوسيط بالطريقة البيانية من منحنى التكرار التجميعي الصاعد أو من منحنى التكرار التجميعي النازل وذلك مهما كان المتغير الإحصائي المدروس منقطعا أو مستمرا.

وإن خطوات إيجاد هذه القيمة هي:

- رسم محورين متعامدين، يمثل المحور الأفقي فئات الظاهرة المدروسة أو مراكز الفئات، والمحور العمودي يمثل مختلف التكرارات التجميعية؛

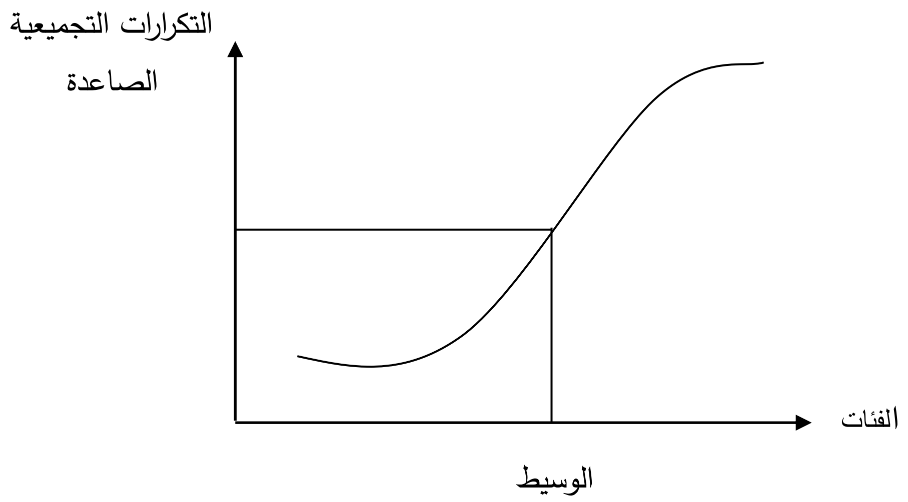
- نجد رتبة الوسيط التي تساوي $\frac{\text{مجموع التكرارات}}{2}$

- نمثل الحد الأعلى للفئة الأولى بنقطة مقابلة للتكرار التجميعي الصاعد، وهكذا لباقي الفئات، ثمّ نرسم المنحنى الذي يمر بهذه النقاط فنحصل على منحنى التكرار التجميعي الصاعد؛

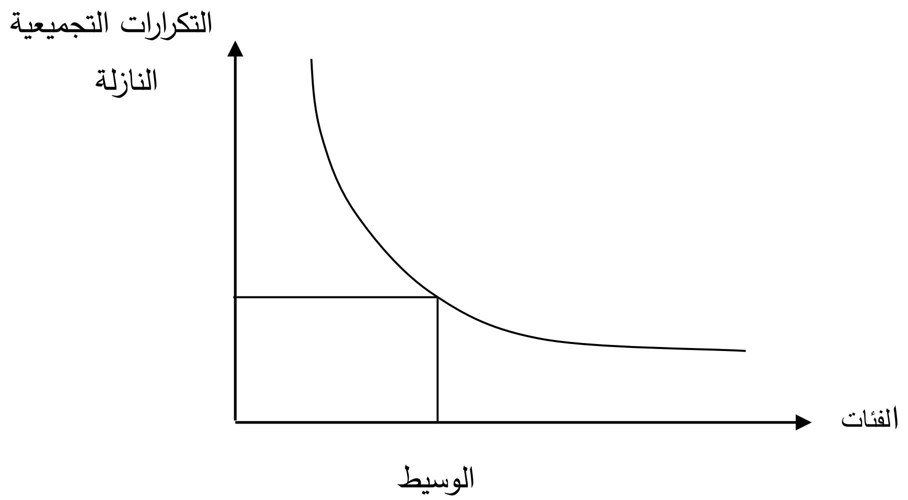
- أو نمثل الحد الأدنى للفئة الأولى بنقطة مقابلة للتكرار التجميعي النازل، ونفس الشيء لباقي الفئات، ثمّ نرسم منحنى يمر بهذه النقاط فنحصل على منحنى التكرار التجميعي النازل؛

- نعين رتبة الوسيط في المحور العمودي من البيان، نمدد بسطر نحو منحنى التكرار التجميعي (سواء كان صاعدا أو نازلا) ثمّ نسقط على المحور الأفقي اتجاه قيم الظاهرة، فتكون النقطة المتحصل عليها هي قيمة الوسيط حسب الأشكال البيانية التالية:

شكل رقم (17): منحنى التكرار التجميعي الصاعد لإيجاد الوسيط

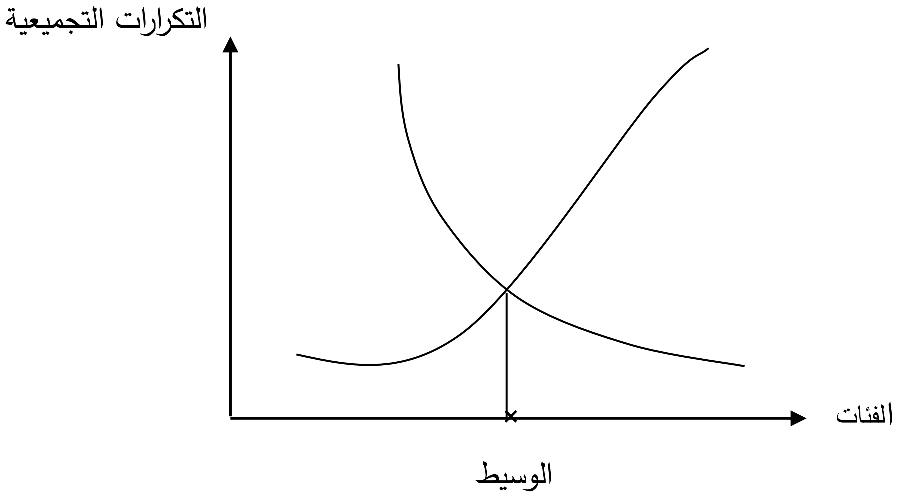


شكل رقم (18): منحنى التكرار التجميعي النازل لإيجاد الوسيط



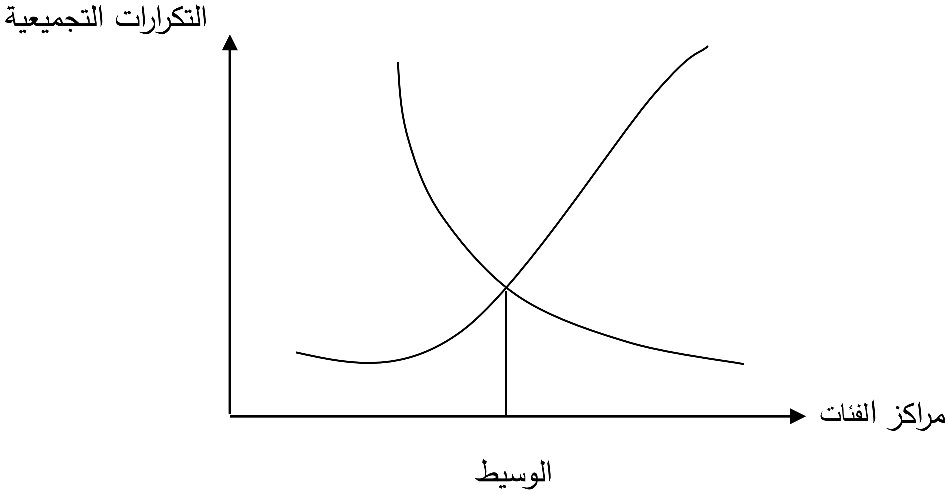
ومن الأحسن رسم المنحنيين في رسم واحد حيث يعطينا نقطة تقاطع بين منحنى التكرار التجميعي الصاعد ومنحنى التكرار التجميعي النازل، وبعد إسقاط عمود على المحور الأفقي تعطينا قيمة الوسيط كما هو مبين في الشكل التالي:

شكل رقم (19): الوسيط عن طريق تقاطع المنحنيين الصاعد والنازل



كما يمكن استعمال مراكز الفئات عوض حدود الفئات، حيث يعين كل مركز فئة للقيمة المقابلة للتكرار التجميعي الصاعد والتكرار التجميعي النازل في نفس الوقت.

شكل رقم (20): إيجاد الوسيط باستعمال مراكز الفئات

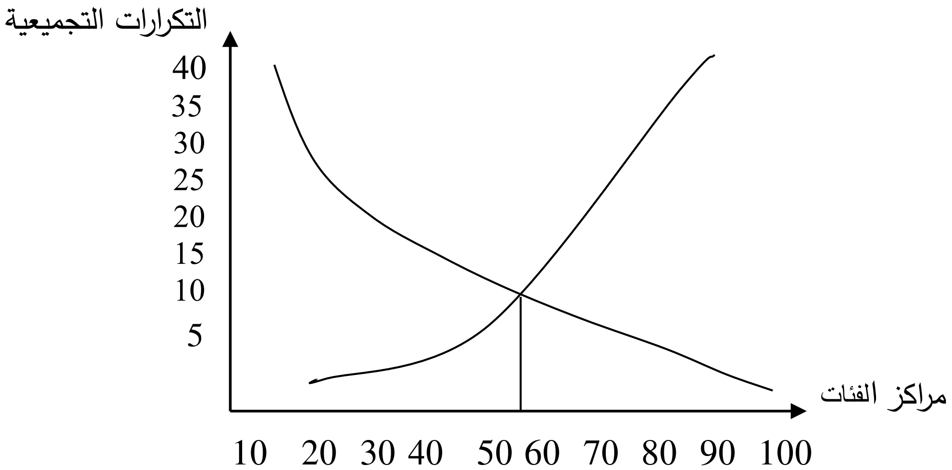


مثال (9): أوجد الوسيط بالطريقة البيانية لأجور العمال (الوحدة: 10^{2} دج):

جدول رقم (15): جدول توزيع تكراري لفئات الأجور

فئات الأجور	التكرارات (n_i)	ت.ت.ص $(F\uparrow)$	ت.ت.ن $(F\downarrow)$
10 - 20	1	1	41
20 - 30	3	4	40
30 - 40	5	9	37
40 - 50	7	16	32
50 - 60	9	25	25
60 - 70	7	32	16
70 - 80	5	37	9
80 - 90	3	40	4
90 - 100	1	41	1
المجموع	41	---	---

شكل رقم (20): إيجاد الوسيط باستعمال مراكز الفئات



$Me \approx 55$

2-4- خصائص الوسيط:

- لا تدخل جميع مفردات الظاهرة عند حساب قيمة الوسيط، لذا يعتبر الوسيط مقياسا مناسبا في التوزيعات الشاذة، أي أنه لا يتأثر بالقيم المتطرفة؛
- يمكن إيجاده للتوزيعات المفتوحة؛
- ويمكن إيجاده أيضا بيانيا على العكس من المتوسط الحسابي؛
- يتأثر الوسيط بعدد البيانات للمتغير الإحصائي؛
- لا يمكن حساب الوسيط إذا كان عدد المفردات قليلا، وبعضها متكررا.

3– المنوال: (Mode)

يعرف المنوال على أنه تلك القيمة الأكثر تكرارا أو الأكثر شيوعا في التوزيع الإحصائي، فهو لا يعتمد على العمليات الحسابية ولا يعتمد على ترتيب البيانات في استخراجه بل يتوقف على توفر تكرارات مختلفة لكل قيمة من قيم المجتمع الإحصائي، فإن لم تكن هناك تكرارات فلا وجود للمنوال.

3–1– المنوال للبيانات غير المبوبة:

يمكن أن يكون للبيانات منوال واحد، ويمكن أن يكون لها منوالين، كما يمكن أن لا يكون لها منوال.

مثال (10): – إذا كانت لدينا البيانات التالية: 67 75 84 53 95 79

لا يوجد منوال، نقول سلسلة عديمة المنوال

– إذا كانت لدينا السلسلة: 19 15 10 27 19 16 12

المنوال هو 19

– وإذا كانت السلسلة: 2 4 5 5 65 7 8 8

يوجد منوالان وهما 5 و8، نقول سلسلة ثنائية المنوال (Bimodal)

3–2– المنوال للتوزيعات التكرارية:

المنوال في هذه الحالة يوجد داخل الفئة ذات أكبر تكرار، ويستخدم في التوزيعات المفتوحة. وهناك عدة طرق لإيجاد المنوال، منها الطريقة الجبرية (الحسابية) حيث نستعمل طريقة الفروق لـ كارل بيرسون (Karl Pearson) بالإضافة إلى الطريقة الهندسية أي عن طريق البيان.

3–2–1– طريقة الفروق لـ (Karl Pearson):

إن إيجاد المنوال يختلف إذا ما كانت لدينا توزيعات منتظمة – أي أطوال فئات متساوية – أو توزيعات غير منتظمة – أي حالة فئات غير متساوية الأطوال.

1)– إيجاد المنوال في حالة توزيعات منتظمة:

تتلخص خطوات هذه الطريقة لإيجاد المنوال فيما يلي:

– نحدد الفئة المنوالية التي تقابل أكبر تكرار عندما يكون طول الفئة ثابتا؛

– نجد الفرق بين تكرار الفئة المنوالية والفئة السابقة عنها، وليكن (Δ_1)؛

– ثم نجد الفرق بين تكرار الفئة المنوالية والفئة اللاحقة لها، وليكن (Δ_2)؛

– نجد المنوال بتطبيق العلاقة التالية:

$$Mo = A + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times K$$

حيث: A : هو الحد الأدنى للفئة المنوالية

Δ_1 و Δ_2 : هي الفروق

K : هو طول الفئة المنوالية

مثال (11): أوجد المنوال للمثال (7) حول توزيع الطلبة على علامات الامتحان التي حصلوا عليها في مادة الإحصاء ؟

– لدينا الفئة المنوالية [12– 14] والتي تقابل أكبر تكرار الذي يساوي 33

– نحدد Δ_1، حيث: $\Delta_1 = 33 - 26 = 7$

– نحدد Δ_2، حيث: $\Delta_2 = 33 - 14 = 19$

– نجد المنوال من العلاقة كما يلي: $Mo = 12 + \frac{7}{7+19} \times 2 = 12.52\,pts$

2)– إيجاد المنوال في حالة توزيعات غير منتظمة:

نطبق نفس الخطوات السابقة، إلا أن الفئة المنوالية في هذه الحالة تكون تلك الفئة المقابلة لأكبر تكرار معدل، وقد لا تكون الفئة المنوالية نفسها قبل وبعد تعديل التكرارات.

$$\text{التكرار المعدل} = \frac{\text{التكرار المطلق}}{\text{طول الفئة}}$$

مثال (12): أوجد المنوال للجدول التالي حول علامات الطلبة:

جدول رقم (16): جدول توزيع تكراري لعلامات الطلبة

الفئات	التكرارات (n_i)	التكرار المعدل
0 - 3	15	5
3 - 8	35	7
8 - 10	45	22.5
10 - 12	28	14
12 - 14	20	10
14 - 16	12	6
16 - 20	4	1
المجموع	100	---

الفئة المنوالية المقابلة لأكبر تكرار معدل هي [8– 10[.

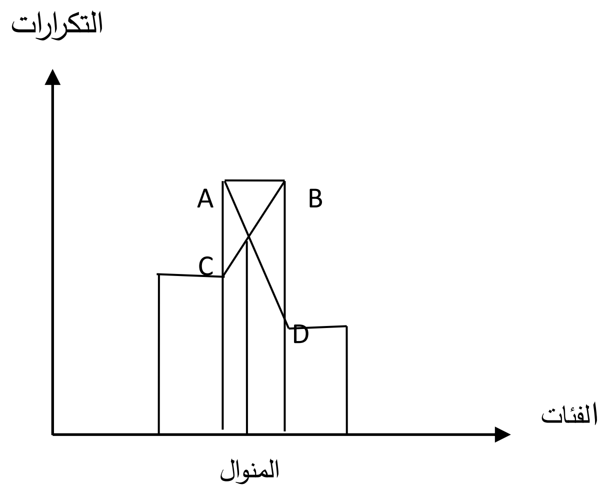
بتطبيق العلاقة نجد المنوال يساوي إلى:

$$Mo = 8 + \frac{(22.5-7)}{(22.5-7)+(22.5-14)} \times 2 = 9.29\,pts$$

3-2-2- الطريقة البيانية:

يمكن إيجاد قيمة المنوال وذلك برسم المدرج التكراري للجزء الذي يناظر الفئة المنوالية الفئة السابقة عنها واللاحقة لها، بحيث نصل النقطة (A) والتي تمثل الحد الأدنى للفئة المنوالية مع النقطة (D) التي تمثل الحد الأدنى للفئة اللاحقة لها، ثمّ نصل النقطة (B) وهي تمثل الحد الأعلى للفئة المنوالية مع النقطة (C) التي تمثل الحد الأعلى للفئة السابقة عنها. نلاحظ أن هناك نقطة تقاطع، نسقط على المحور العمودي الذي يمثل الفئات عندها نحصل على قيمة المنوال كما هو مبين في الشكل البياني التالي:

شكل رقم (21): المدرج التكراري لأطوال الطلبة



3-2-3- خصائص المنوال:

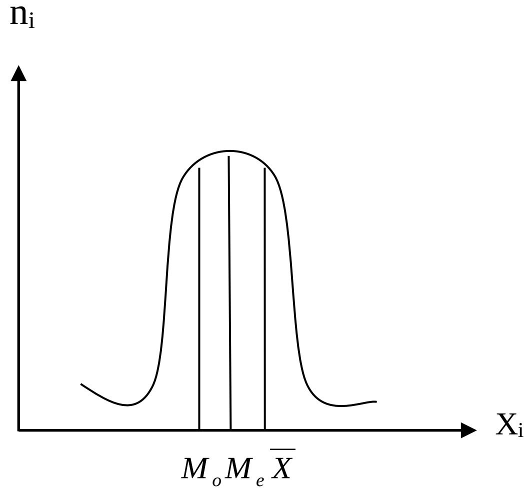
- إذا قمنا بضرب قيمة المنوال في عدد مفردات الظاهرة موضوع القياس، فلا يعطي ناتج ما سبق المجموع الأصلي للتوزيع كما هو الحال بالنسبة للمتوسط الحسابي؛
- لا يمثل المنوال القيمة الوسطى في التوزيع؛
- عند حساب قيمته، لا تدخل كل مفردات التوزيع للظاهرة كما هو الحال بالنسبة للمتوسط. لذا فهو لا يتأثر بالقيم الشاذة أو المتطرفة؛
- يمكن حسابه من الجداول التكرارية المفتوحة؛
- تتأثر قيمة المنوال بحجم العينة وبطول فئة التوزيع؛
- إذا كان للتوزيع أكثر من قيمة واحدة ذات أكبر تكرار، فهذا يعني أن هناك أكثر من منوال؛
- إذا كانت الفئة المنوالية في بداية أو في نهاية جدول التوزيع التكراري، يصعب إيجاد قيمة المنوال لهذه البيانات.

4- العلاقة بين المتوسط الحسابي، الوسيط والمنوال:

إذا كان التوزيع التكراري متماثلا، فإن المنحنى البياني الممثل له يكون معتدلا أي متماثلا حيث يمر بنقطة النهاية العظمى للتوزيع ويقسمه إلى جزئين متطابقين تماما. وإذا كان غير متماثلا، يكون المنحنى البياني الممثل له في هذه الحالة غير معتدلا ويطلق عليه منحنى ملتوي.

- إذا كانت المقاييس الثلاثة متساوية، أي أن المتوسط= الوسيط= المنوال، نقول أن التوزيع التكراري متماثل (متناظر) ويأخذ الرسم الشكل التالي:

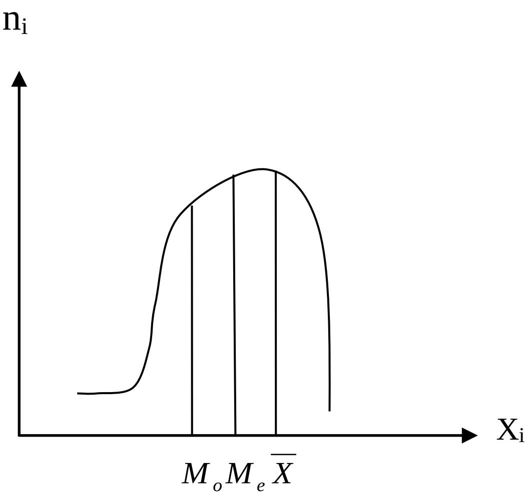
شكل رقم (22): منحنى متماثل



$$\overline{X} = M_e = M_o$$

- إذا تركزت البيانات في الجهة اليمنى، فهذا يعني أن التوزيع غير متناظر من اليمين (التواء موجب) وتكون العلاقة بين المقاييس الثلاثة كما يلي: المتوسط > الوسيط > المنوال، ويأخذ الرسم الشكل التالي:

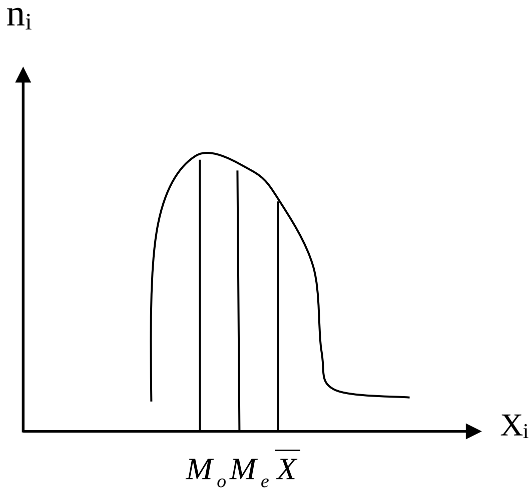
شكل رقم (23): منحنى غير متماثل (عدم التناظر من اليمين)



$$\overline{X} \rangle M_e \rangle M_o$$

– وإذا تحققت العلاقة بين المقاييس: المتوسط < الوسيط < المنوال فهذا يعني أن البيانات تركزت في الجهة اليسرى، ونطلق على هذا التوزيع بأنه غير متناظر من اليسار (التواء سالب) طبقا للرسم البياني التالي:

شكل رقم (24): منحنى غير متماثل (عدم التناظر من اليسار)



$$\overline{X} \langle M_e \langle M_o$$

في حالة التوزيعات القريبة إلى التناظر، يمكن إيجاد علاقة تربط بين المقاييس الثلاثة كما يلي:

$$\overline{X} - M_o = 3\left(\overline{X} - M_e\right)$$

ومنه، فإن:

$$\overline{X} = \frac{3M_e - M_o}{2}$$

و

$$M_e = \frac{2\overline{X} + M_o}{3}$$

وبصفة عامة، تعتبر هذه المقاييس الثلاثة أكثر المقاييس استخداما بالرغم من اختلاف دقتها، استخداماتها وانتشارها، وهذا لا يعني أنها المتوسطات الوحيدة في سلسلة مقاييس النزعة المركزية بل هناك عدد آخر من المتوسطات الشائعة الاستخدام وهذا حسب الظاهرة المدروسة.

5– الربيعيات– المقاييس الشبيهة بالوسيط:

(Les Quartiles- Mesures Quasi- Médiane)

تسمى بمقاييس شبيهة بالوسيط لتشابهها معه في التعريف وطريقة الحساب. ولحساب قيمة أي مقياس منها، تُستعمل العلاقة التي يحسب بها الوسيط مع تغيير رتبة المقياس.

وإن مفهوم الربيعيات هو تقسيم منحنى التوزيع التكراري إلى أربعة أجزاء متساوية، ومن هنا توجد ثلاثة ربيعيات: الربيعي الأول (ويسمى بالأدنى)، الربيعي الثاني (هو الوسيط) والربيعي الثالث (ويسمى بالأعلى).

5-1- الربيعي الأول (الأدنى): (Quartile Inferieur)

هو تلك القيمة التي تقسم المجتمع الإحصائي إلى قسمين، يحتوي القسم الأول على نسبة 25% من الوحدات الإحصائية، والقسم الثاني على نسبة 75% من هذه الوحدات.

5-1-1- حالة البيانات غير المبوبة:

- إذا كان n عدد البيانات فرديا، فإن رتبة الربيعي الأول هي: $\frac{n+1}{4}$
- وإذا كان n عددا زوجيا، فإن رتبة الربيعي الأول هي $\frac{n}{4}$ و $\frac{n}{4}+1$

مثال (20): لتكن السلسلة الإحصائية التالية:

14 37 28 13 5 3 42

- نرتب القيم ترتيبا تصاعديا: 3 **<u>5</u>** 13 14 28 37 42
- الرتبة: عدد فردي = $\frac{n+1}{4} = \frac{7+1}{4} = 2$
- قيمة الربيعي الأول هي القيمة المقابلة للرتبة 2 بعد ترتيب القيم، أي $Q_1 = 5$

مثال (21): إذا أضفنا قيمة 50 للبيانات الإحصائية السابقة، يصبح n عدد زوجي:

50 42 37 28 14 13 5 3

الرتبة هي:

$$\left.\begin{array}{l} \frac{n}{4} = \frac{8}{4} = 2 \\ \frac{n}{4}+1 = \frac{8}{4}+1 = 3 \end{array}\right\} \quad \begin{array}{l} \Rightarrow Q_1 = 5 \\ \Rightarrow Q_1 = 13 \end{array}$$

قيمة الربيعي الأول هي القيمة المتوسطة المقابلة للرتبتين أي:

$$Q_1 = \frac{5+13}{2} = \frac{18}{2} = 9$$

5-1-2- حالة التوزيعات التكرارية:

هي نفس خطوات إيجاد الوسيط، نحسب التكرار التجميعي الصاعد، نجد الرتبة والتي تساوي $\frac{\sum n_i}{4}$ ثم نقارنها في التكرار التجميعي الصاعد المساوٍ لها أو الأكبر منها مباشرة ونجد الفئة الربيعية الأولى المقابلة. وفي الأخير، نجد قيمة الربيعي الأدنى من العلاقة:

$$Q_1 = A + \frac{\frac{\sum n_i}{4} - F\uparrow(n-1)}{n_{i\,Q_1}} \times K$$

5-2- الربيعي الثاني:

هو تلك القيمة التي تقع عند الربع الثاني من التوزيع الإحصائي، ولهذا فهي تقابل قيمة الوسيط.

رتبة الربيعي الثاني هي: $\frac{\sum n_i}{2} = \frac{2\sum n_i}{4}$ ، ومنه فإن: $Q_2 = Me$

5-3- الربيعي الثالث (الأعلى): (Quartile Supérieur)

هو تلك القيمة التي تقسم المجتمع الإحصائي إلى قسمين، يحتوي القسم الأول على نسبة 75% من الوحدات الإحصائية، والقسم الثاني على نسبة 25% من هذه الوحدات.

5-3-1- حالة البيانات غير المبوبة:

- إذا كان n عدد البيانات فرديا، فإن رتبة الربيعي الثالث هي: $\frac{3(n+1)}{4}$
- وإذا كان n عددا زوجيا، فإن رتبة الربيعي الثالث هي $\frac{3n}{4}$ و $\frac{3n}{4}+1$

مثال (20): بالرجوع إلى المثال رقم (20) السابق:

n عدد فردي: الرتبة $= \frac{3(n+1)}{4} = \frac{3\times 8}{4} = 6$

$$Q_3 = 37$$

n عدد زوجي: الرتبة هي:

$$\left.\begin{array}{l} \frac{3n}{4} = \frac{3\times 8}{4} = 6 \\ \frac{3n}{4}+1 = \frac{3\times 8}{4}+1 = 7 \end{array}\right\} \begin{array}{l} \Rightarrow Q_3 = 37 \\ \Rightarrow Q_3 = 42 \end{array}$$

قيمة الربيعي الأعلى هي القيمة المتوسطة المقابلة للرتبتين أي:

$$Q_3 = \frac{37+42}{2} = 39.5$$

5-3-2- حالة التوزيعات التكرارية:

باتباع نفس خطوات إيجاد الربيعي الأول، نجد الربيعي الثالث باستعمال العلاقة:

$$Q_3 = A + \frac{\frac{3\sum n_i}{4} - F\uparrow(n-1)}{n_{iQ_3}} \times K$$

مثال (21): تمثل البيانات التالية توزيع العمال حسب فئات الأجور:

جدول رقم (20): جدول توزيع تكراري لأجور العمال

الفئات	عدد العمال	ت.ت.ص
6000 - 5000	20	20
7000 - 6000	45	65
8000 - 7000	35	100
9000 - 8000	15	115
10000 - 9000	25	140
المجموع	140	---

المطلوب: أوجد الربيعي الأول (الأدنى)، الربيعي الثاني (الوسيط) والربيعي الثالث (الأعلى) ؟

1- إيجاد الربيعي الأول (الأدنى):
$$Q_1 = A + \frac{\frac{\sum n_i}{4} - F\uparrow(n-1)}{n_{i\,Q_1}} \times K$$

الرتبة هي: $\frac{\sum n_i}{4} = \frac{140}{4} = 35$ $\Leftarrow$ الفئة الربيعية الأولى هي: [6000− 7000[

$$Q_1 = 6000 + \frac{35-20}{45} \times 1000 = 6333.33DA$$

2- إيجاد الربيعي الثاني (الوسيط):
$$Q_2 = Me = A + \frac{\frac{\sum n_i}{2} - F\uparrow(n-1)}{n_{i\,Me}} \times K$$

الرتبة هي: $\frac{\sum n_i}{4} = \frac{140}{4} = 35$ $\Leftarrow$ الفئة الربيعية الأولى هي: [7000− 8000[

$$Q_2 = Me = 7000 + \frac{70-65}{35} \times 1000 = 7142.85DA$$

3- إيجاد الربيعي الثالث (الأعلى):
$$Q_3 = A + \frac{\frac{3\sum n_i}{4} - F\uparrow(n-1)}{n_{i\,Q_3}} \times K$$

الرتبة هي: $\frac{3\sum n_i}{4} = \frac{3\times 140}{4} = 105$ $\Leftarrow$ الفئة الربيعية الثالثة هي: [8000− 9000[

$$Q_3 = 8000 + \frac{105-100}{15} \times 1000 = 8333.33DA$$

الفصـل الرابع

مقاييس التشتت

المقدمة:

يقصد بالتشتت تباعد القيم عن بعضها أو عدم تجانسها، ولهذا تظهر مقاييس التشتت اختلاف قيم العينة المدروسة فيما بينها أو مدى قربها أو بعدها من أحد مقاييس النزعة المركزية وعادة ما يكون المتوسط الحسابي.

1- المدى العام: (Etendue)

هو الفرق بين أكبر قيمة في البيانات الإحصائية وأصغرها، ويعتبر من أبسط مقاييس التشتت وأسهلها حسابا. وفي التوزيعات التكرارية، يحسب من مراكز الفئات ولا يمكن إيجاده من الجداول المفتوحة وبالتالي فهو يتأثر بالقيم المتطرفة.

$$E = X_{\max} - X_{\min}$$

ويتميز المدى العام بأنه يستعمل في المقارنة بين توزيعين إحصائيين أو أكثر.

مثال (1): لدينا مجموعتان من البيانات الإحصائية التالية:

المجموعة (1): 200 190 178 165 140

المجموعة (2): 220 199 187 144 100

$$\begin{cases} E_1 = 200 - 140 = 60 \\ E_2 = 220 - 100 = 120 \end{cases}$$

$E_2 \rangle E_1$ ومنه، فإن التوزيع الثاني أكثر تشتتا من التوزيع الأول، ويعني أن عناصر المجموعة (2) أكثر تباعدا (أقل تجانسا) فيما بينها مقارنة بالمجموعة (1).

2- المدى الربيعي: (Intervalle Interquartile)

في بعض الحالات، يؤدي استخدام المدى العام إلى تفسيرات خاطئة خاصة إذا توفرت هناك قيم شاذة (متطرفة). ومن أجل التخلص من تأثير هذه القيم المتطرفة، نستعمل المجال ما بين الربيعيات يسمى بالمدى الربيعي. ويعرف بأنه الفرق بين الربيعي الثالث أو الأعلى (Q_3) والربيعي الأول أو الأدنى (Q_1)، ويعطى بالعلاقة التالية:

$$I_Q = Q_3 - Q_1$$

ويتميز المدى الربيعي بالخصائص التالية:

- أنه سهل الحساب؛
- يمكن إيجاده للبيانات غير المبوبة؛
- يستعمل في التوزيعات التكرارية المغلقة والمفتوحة، ويعتبر أحسن مقياس في التوزيعات المفتوحة.

3- النسبة بين المدى العام والمدى الربيعي (قياس التشتت):

يبين هذا المقياس تشتت 50% من الوحدات الإحصائية المركزية (أي التي تقع في مركز التوزيع) حول الوسيط. وتعطى العلاقة كما يلي:

$$IR = \frac{Q_3 - Q_1}{E} \times 100$$

وفي التوزيعات المفتوحة:

$$IR = \frac{Q_3 - Q_1}{M_e} \times 100$$

- فإذا كان $IR = 50\%$ ، يكون التوزيع الإحصائي متماثلا أو متناظرا؛
- أما إذا كان $IR \rangle 50\%$ ، فإن التوزيع قوي التشتت بالنسبة للقيمة المركزية؛
- وإذا كان $IR \langle 50\%$ ، فإن التوزيع ضعيف أو قليل التشتت بالنسبة للقيمة المركزية (الوسيط).

مثال (3): لو أخذنا نفس المثال السابق حول أجور العمال، قس تشتت هذا التوزيع:

$$\begin{cases} I_Q = 200 \\ E = 600 - 300 = 300 \end{cases}$$

$$IR = \frac{200}{300} \times 100 = 66.67\%$$

ومنه، فإن التشتت قوي. أو نقول أن توزيع أجور العمال قوي التشتت.

4- التباين والانحراف المعياري:

4-1- التباين: (Variance)

إن انحراف القيم عن المتوسط الحسابي يعكس صورة جيدة لدرجة تشتت المعطيات (البيانات)، وإن تربيع هذه الانحرافات وجمعها يعطينا التباين.

- حالة البيانات غير المبوبة

$$V(X) = \frac{\sum \left(X_i - \overline{X}\right)^2}{n}$$

- حالة التوزيع التكراري:

$$V(X) = \frac{\sum n_i \left(X_i - \overline{X}\right)^2}{\sum n_i}$$

تسمى هذه العلاقة بعلاقة التعريف، وانطلاقا من هذه الأخيرة يمكن إيجاد علاقة أخرى تسمى بالعلاقة الموسعة كما يلي:

- بيانات غير مبوبة:

$$V(X)=\frac{\sum\left(X_i-\overline{X}\right)^2}{n}=\frac{\sum\left(X_i^2-2X_i\overline{X}+\overline{X}^2\right)}{n}$$

$$=\frac{\sum X_i^2}{n}-2\overline{X}\frac{\sum X_i}{n}+\frac{\sum \overline{X}^2}{n}$$

$$=\frac{\sum X_i^2}{n}-2\overline{X}\,\overline{X}+\overline{X}^2$$

$$=\frac{\sum X_i^2}{n}-2\overline{X}^2+\overline{X}^2=\frac{\sum X_i^2}{n}-\overline{X}^2$$

- توزيعات تكرارية:

$$V(X)=\frac{\sum n_i X_i^2}{\sum n_i}-\overline{X}^2$$

نتيجة: نلاحظ أن التباين هو عبارة عن الفرق بين المتوسط التربيعي ومربع المتوسط الحسابي.

4-2- الانحراف المعياري: (L'écart type)

هو الجذر التربيعي لمتوسط مربعات انحرافات القيم عن متوسطها الحسابي، وبذلك يكون الجذر التربيعي للتباين.

- بيانات غير مبوبة

$$\delta_{(X)}=\sqrt{\frac{\sum\left(X_i-\overline{X}\right)^2}{n}}=\sqrt{\frac{\sum X_i^2}{n}-\overline{X}^2}$$

- توزيعات تكرارية:

$$\delta_{(X)}=\sqrt{\frac{\sum n_i\left(X_i-\overline{X}\right)^2}{\sum n_i}}=\sqrt{\frac{\sum n_i X_i^2}{\sum n_i}-\overline{X}^2}$$

مثال (6): أوجد التباين والانحراف المعياري لعناصر السلسلة التالية:

1 2 4 5 7 8 9 10 11 13

1)- إيجاد التباين:

$$\overline{X}=\frac{70}{10}=7$$

$$V(X)=\frac{\sum\left(X_i-\overline{X}\right)^2}{n}=\frac{1+4+\ldots+169}{10}=\frac{140}{10}=14$$

2)- إيجاد الانحراف المعياري:

$$\delta_{(X)}=\sqrt{V(X)}=\sqrt{\frac{\sum\left(X_i-\overline{X}\right)^2}{n}}=\sqrt{14}=3.74$$

مثال (7): إذا كانت لدينا أطوال الطلبة ملخصة في شكل فئات كما يلي، أحسب التباين والانحراف المعياري:

جدول رقم (25): جدول توزيع تكراري لأطوال الطلبة

$n_i X_i^2$	X_i^2	$n_i X_i$	X_i	n_i	الفئات
78246.75	26082.25	484.5	161.5	3	163 - 160
81180.75	27060.25	493.5	164.5	3	166 - 163
168337.50	28056.25	1005	167.5	6	169 - 166
290702.50	29070.25	1705	170.5	10	172 - 169
391329.25	30102.25	2255.5	173.5	13	175 - 172
249218.00	31152.25	1412	176.5	8	178 - 175
1259014.75	---	7355.5		43	المجموع

$$\overline{X} = \frac{\sum n_i X_i}{\sum n_i} = \frac{7355.5}{43} = 171.06$$

$$V(X) = \frac{\sum n_i X_i^2}{\sum n_i} - \overline{X}^2 = \frac{1259014.75}{43} - (171.06)^2$$
$$= 29279.41 - 29261..52 = 17.89$$
$$\delta_{(X)} = \sqrt{\frac{\sum n_i X_i^2}{\sum n_i} - \overline{X}^2} = \sqrt{17.89} = 4.22$$

وعن استخداماته، فإن الانحراف المعياري يستعمل في مجالات كثيرة نظرا لأهميته في التحليل الإحصائي حيث أنه يتميز ببعض الخصائص والأسس الرياضية بالإضافة إلى استعماله في تعميم نتائجه من العينة على المجتمع

5- الاختلاف النسبي (معامل الاختلاف): (Coefficient de Variation)

يستعمل في مقارنة مجتمعات إحصائية مختلفة من حيث درجة تجانسها خاصة إذا اختلفت متوسطاتها الحسابية حيث يؤدي الانحراف المعياري – عندما تكون متوسطات البعض منها كبيرة – في المقارنة إلى استنتاجات خاطئة. لذا يحول المقياس المراد مقارنته إلى مقياس نسبي للتشتت باستخدام الحجم النسبي للاختلاف، ويعرف باسم "**معامل الاختلاف**".

يستخدم معامل الاختلاف (CV) لقياس درجة التشتت النسبي لمجموعة ما أو توزيع معين، ويمكن إيجاده طبقا للعلاقة التالية:

$$CV = \frac{\delta_{(X)}}{\overline{X}} \times 100$$

وفي التوزيعات المفتوحة: $CV = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$ ويسمى "**بمعامل التغاير**".

ويستخدم هذا المقياس أيضا للمقارنة بين مجموعتين أو أكثر تختلف وحدات قياسها بعد أن يتعذر علينا المقارنة بينها باستخدام مقاييس التشتت المطلق.

مثال (8): إذا كان لدينا نوعان من المطاط A وB حيث:

	A	B
$\overline{X}$	27000	29000
$\delta_{(X)}$	5200	2500

إن درجة الاختلاف بين A وB كنسبة هي:

$$CV_A = \frac{5200}{27000} \times 100 = 0.1925 \times 100 = 19.25\%$$

$$CV_B = \frac{2500}{29000} \times 100 = 0.0862 \times 100 = 8.62\%$$

$CV_B \langle CV_A$ أي أن النوع الثاني (A) أحسن من النوع الأول (B).

الخاتمة العامـــة

الخاتمة العامة:

نرى أنه وبعد هذا التقديم نكون قد ألممنا بجميع الجوانب الأساسية للدروس المتنوعة التي نقدمها للطالب خلال مرحلته الأولى من الالتحاق بالجامعة، حيث بيّنا مدى أهمية دراسة البيانات الإحصائية من خلال الفصل الأول لهذه المطبوعة.

ثمّ عمدنا إلى تلخيص هذه البيانات وعرضها في جداول تدعى بالجداول التكرارية، وعلى هذا قدمنا الفصل الثاني الذي بيّنا من خلاله الجداول المختلفة حسب اختلاف طبيعة المتغيرات. وفي الجزء الثاني من هذا الفصل، قدمنا مختلف الرسومات والأشكال البيانية التي تُعرض بها أنواع المتغيرات.

بعد هذا تطرقنا إلى كيفية استخدام الأدوات الإحصائية لوصف هذه البيانات، أهمها مقاييس النزعة المركزية والتي تتلخص في المتوسط، الوسيط والمنوال، ويتبين هذا في الفصل الثالث من هذه المطبوعة.

أما الفصل الرابع، فقد احتوى الصنف الثاني من أدوات الوصف الإحصائي، والذي يتمثل في مقاييس التشتت التي تدرس مدى تباعد القيم المدروسة فيما بينها مقارنة بقيمتها المركزية عادة ما تكون المتوسط الحسابي.