مبـادئ الإحصــاء

(لطلاب الأنتساب)

دكتور/ على أبو الســــعود

قسم الإحصاء والرياضة والتأمين ـ جامعة كفر الشيخ

قسم التأمين ـ جامعة أم القرى

مقدمة

بادئ ذى بدء نحمد الله سبحانه وتعالى ونشكره ونتوب إليه ونستغفره ، نحمده على نعمائه الفاخرة وآلاءه الزاخرة ، وبعد ، نعيش الآن فى عصر المعلومات حيث حلبة السباق بين كافة الدول فى العلوم والتكنولوجيا بات مرهوناً بقدر المعلومات المتاحة.

ولما كانت سرعة الحصول على المعلومات تقتضى سرعة اتخاذ القرارات فى كافة المجالات حتى لا تضيع الفرصة أمام متخذى هذه القرارات لا سيما فى المجالات التجارية والاقتصادية والرياضية عامة لذا فإن الأساليب الإحصائية تساعدنا دائماً فى اتخاذ القرارات الادارية ، حيث بهذه الأساليب يتم تحليل وتفسير البيانات فضلاً عن تقييم نتائج الاختبارات والمقاييس المختلفة فى العلوم الاقتصادية والتجارية والاجتماعية والانسانية ، ومن ثم بزغت أهمية الأساليب الإحصائية فى كافة المجالات الاقتصادية والتجارية.

وفى يتكون هذا الجزء من خمسة فصول الفصل الأول منها أنصب على دراسة التعريفات الأساسية للأساليب الإحصائية ، أما الفصل الثانى اهتم بتعريف الدارس وتدريبه على كيفية جمع البيانات وتبويبها وعرضها جدولياً وبيانياً بينما الفصل الثالث تعرض لدراسة مقاييس النزعة المركزية والفصل الرابع يدرس مقاييس التشتت .

وإننا إذ نقدم هذا العمل نرجو من الله تعالى أن نكون قد وفقنا فى تقديمه بالطريقة المبسطة والتى رعيناها أثناء عرض المادة العلمية لكى تتحقق الفائدة التى نتوخاها وهى تقريب وتبسيط المفاهيم الأساسية للأساليب الإحصائية لدارسى العلوم الادارية والمالية والاقتصادية.

والله نسأل أن يوفقنا جميعا الى مافية الخير لأمتنا العربية والإسلامية والحمد لله الذى هدانا لهذا وما كنا لنهتدى لولا أن هدانا الله.

المؤلف

Dr. Ali Abo Elsooud

الفصل الأول
تعريفات أساسية للأساليب الإحصائية

■ مقدمة

رغم أن الإحصاءات كانت تستخدم من قديم الأزل لأغراض حربية وضريبية وفلكية والتى زادت أهميتها بعد نشوب الثورة الصناعية إلا أن علم الإحصاء لم يظهر بمفهومه الحديث إلا فى القرن الثامن عشر حيث كان أول من أرسى قواعده العالم كواتيله Quetelet (1796-1874).

والإحصاء بمفهومه الحديث يخدم كل الباحثين ومتخذى القرار فى كافة المجالات العلمية . فعلى سبيل المثال الباحث فى مجال الاقتصاد يمكنه اختبار سلوك المستهلك عن طريق الأساليب الإحصائية كما أن الباحث فى مجال الطب يمكنه قياس كفاءة دواء جديد عن طريق ذات الأساليب . فضلاً عن ذلك الباحث فى المجال الرياضى يمكنه قياس العلاقة بين طول الفرد (المتدرب) ووزنه وأدائه الرياضى (لياقته الرياضية).

أيضاً الباحث فى المجال الزراعى يمكنه معرفة آثار الأسمدة على المحصول الزراعى الذى تم تسميده بهذه الأسمدة.

ومن ثم فإنه يمكن القول بأن علم الإحصاء يتغلغل فى كافة المجالات والعلوم حيث لا يوجد ميدان أو مجال من مجالات البحث العلمى إلا وطرقه علم الإحصاء ولعب دوراً كبيراً فى تطوره لدرجة أن متخذى القرار فى أية مجال من المجالات لا يمكنهم الاستغناء عن الأساليب الإحصائية فى دراستهم للقرارات البديلة قبل اتخاذ قراراتهم حيث هذه الأساليب هى الصانعة لهذه القرارات

ويعتبر علم الإحصاء من أهم العلوم لذكره فى كتاب الله العظيم القرآن الكريم فى آيتين هما قول الله تعالى "وإن تعدوا نعمت الله لا تحصوها" كذلك قول الله تعالى "وأحاط بما لديهم وأحصى كل شئ عددا" صدق الله العظيم.

ومن ثم تتضح أهمية هذا العلم للعلوم الأخرى حيث يستخدم فى كافة المجالات والعلوم (الاقتصاد - الإدارة - التأمين - الزراعة - علم النفس - التربية الرياضية .... الخ).

لذا فإن القارئ العزيز يجب عليه عدم الخلط بين كلمة الإحصاءات وكلمة الإحصاء ، فمن الضرورى التمييز بينهما حيث تعنى كلمة الإحصاءات مجموعة البيانات العددية التى توصف ظاهرة معينه مثل الإحصاءات السكانية - إحصاءات الزواج - إحصاءات المواليد - إحصاءات الوفيات - الإحصاءات الزراعية ..... الخ . أما كلمة الإحصاء فتعنى مجموعة الأدوات التى تكون فى متناول الباحث أو متخذ القرار يطلق عليها الطرق الإحصائية أو الأساليب الإحصائية.

من هنا نتساءل ما هو تعريف علم الإحصاء؟

نرد ونقول إن علم الإحصاء يعرف بأنه: الأساليب التى تستخدم فى تلخيص وتصنيف وتحليل البيانات العددية لظاهرة (أو عدة ظواهر) عن طريق:

1- جمع وتنظيم وعرض البيانات الخاصة بهذه الظاهرة (أو عدة ظواهر) بطرق تساعد القارئ على فهمها فى شكل جداول- رسوم - أشكال بيانيه.

2- تحليل البيانات الخاصة بهذه الظاهرة (أو عدة ظواهر) واستخلاص نتائج منها.

3- إيجاد العلاقات بين الظواهر المختلفة (مثل حساب المتوسطات والانحرافات - معاملات الارتباط.

4- تفسير نتائج التحليل بغرض استخدامها فى اتخاذ القرارات.

ومن تعريفنا لكل من الإحصاءات وعلم الإحصاء السالفين يتبين لنا أن الإحصاءات تعتبر بمثابة المادة الخام التى يمكن معالجتها وتشكيلها باستخدام أسلوب من الأساليب الإحصائية لعلم الإحصاء لتصبح قابلة للاستخدام.

■ الأساليب الإحصائية

من الجدير بالذكر أن الأساليب الإحصائية يمكن تقسيمها إلى ثلاثة أقسام رئيسية وهى:

أولاً: الأساليب الإحصائية الوصفية Descriptive Statistics

وهذه الأساليب تتكون من مجموعة طرق تعنى بجمع البيانات لظاهرة معينة وتنظيمها وتلخيصها وعرضها بطريقة واضحة فى صورة (جداول - أشكال بيانيه - حساب مقاييس إحصائية مختلفة) بمعنى آخر أن الأساليب الإحصائية الوصفية تختص بوصف خصائص البيانات لظاهرة معينة وذلك عن طريق معرفة توزيع هذه البيانات وما إذا كانت تتمركز حول قيمة معينة أم أنها متباينة فضلاً عن ذلك فإن هذه الأساليب تختص بإيجاد العلاقة بين أى ظاهرة وظاهرة أخرى وقوة العلاقة بينهما.

وباختصار فإن الأساليب الإحصائية الوصفية ما هى إلا طرق إحصائية تستخدم فى جمع وتلخيص وعرض بيانات العينات أو المجتمعات المختلفة (باستخدام الجداول - الأشكال البيانية) ووصف الجوانب المختلفة للبيانات (مثل الوسط الحسابى والمدى) بغرض استنباط الخواص الأساسية التى تتسم بها هذه البيانات.

ثانياً : الأساليب الإحصائية التحليلية Analytical Statistics

وهذه الأساليب تتكون من الطرق الخاصة باستخلاص النتائج عن المجتمع من بيانات تم جمعها من عينة ثم اختيارها من هذا المجتمع.

بمعنى آخر أن الأساليب الإحصائية التحليلية (الإحصاء الاستدلالى) تختص باستخلاص نتائج عامة من بعض المشاهدات تم جمعها من عينة بمجتمع ما (عن طريق أسلوب المعاينة الإحصائية) وتعميمها على هذا المجتمع بأسره.

من هنا يمكن القول بأن هناك فرق ما بين الأساليب الإحصائية الوصفية (الإحصاء الو صفى) والأساليب الإحصائية التحليلية (الإحصاء التحليلى أو الاستدلالى) وهو أن الإحصاء الأخير يختص بتقدير خواص المجتمع من واقع دراسات لمجموعة بيانات لعينة أو أكثر تم سحبها من هذا المجتمع حيث يبنى هذا التقدير على أساس مجموعة من الافتراضات عن العلاقة بين العينة والمجتمع الذى سحبت منه.

مثال لذلك "عند أخذ عينة من الأفدنه الزراعية المزروعة بمحصول الأرز فى محافظة الغربية لتقدير متوسط إنتاج الفدان من الأرز بتلك المحافظة" ، أما فى الإحصاء الو صفى لا تمتد خواص العينة والمقاييس المستخلصة من بياناتها على المجتمع الذى سحبت منه.

وجدير بالذكر أن الإحصاء التحليلى يعتمد على التحليل الذى يستند بصفة رئيسية على نظرية الاحتمالات حيث يقوم الإحصائى بتقدير خواص المجتمع من واقع خواص العينة وباختصار فان الإحصاء التحليلى ما هو إلا طرق إحصائية تستخدم فى تعميم النتائج على المجتمع اعتماداً على بيانات العينات المأخوذة من هذا المجتمع.

ملحوظة هامة

أحياناً لا تمثل العينة المستخدمة المجتمع الذى سحبت منه تمثيلاً حقيقياً لذا فإنه يجدر على الباحث أن ينظر لأى معلومة تستنتج من العينة على أنها معلومة تقترب من المعلومة الفعلية التى كان سيحصل عليها إذا قام بدراسة المجتمع بأسره ...... وهنا تنبع فائدة الإحصاء التحليلى (الإحصاء الاستدلالى) حيث يمكن الباحث من تحديد الخطأ المتوقع أو نسبة الخطأ.

وخلاصة ما سلف أن الإحصاء التحليلى تتبلور أهدافه فى:

1- تقدير معالم غير معلومة عن المجتمع من خلال مشاهدة بعض المقاييس المحسوبة من عينات من داخل المجتمع.

2- اختبار فروض معينة.

ثالثاً: التنبؤ الاستدلالى Inferential Prediction

وهذا الأسلوب يقصد به استخدام المشاهدات الماضية للاستدلال بها لما يحدث للظاهرة نفسها فى المستقبل (فترة مستقبله).

على سبيل المثال: إذا كان لدينا علاقة خطية بين متغيرين X , Y ممثلة فى الدالة:

$$Y = A + B X$$

حيث أن: Y هى كمية المبيعات من سلعة معينة

X الزمن بالسنوات

A , B مقادير ثابتة

وبفرض أننا نريد التنبؤ بمبيعات السلعة فى فترة زمنية مستقبله . فإن التنبؤ هنا يقوم على استخدام العلاقة بين المتغيرين للاستدلال على قيمة المتغير Y (كمية المبيعات) فى فترة مستقبله استناداً إلى استمرار العلاقة بين المتغيرين فى المستقبل على ما كانت عليه العلاقة بين نفس المتغيرين فى الماضى.

■ المجتمع والعينة

سلف أن ذكرنا أن الإحصاء التحليلى أو الاستدلالى ينصب على تعميم نتائج العينة على المجتمع الذى سحبت منه لذا نتساءل ما هو الفرق بين العينة والمجتمع؟

نرد ونقول أن:

أولاً: المجتمع Population

المجتمع هو مجموعة set تتكون من كل المفردات أو المشاهدات محل الدراسة لظاهرة معينة والتى نرغب فى معرفة حقائق أو خصائص عنها فعلى سبيل المثال عندما نرغب فى جمع بيانات خاصية معينة عن طلاب جامعة ام القرى (أطوال الطلاب - أوزان الطلاب) فإننا نجمع البيانات من جميع وحدات المجتمع (طلاب الجامعة) المطلوب إجراء الدراسة بشأنه وهذا ما يسمى بالحصر الشامل Complete Census وتسمى مجموعة الطلاب المطلوب إجراء الدراسة عليها Population or inverse أما القيمة التى تحسب لوصف الخاصية المعينة تسمى معلمه Parameter.

بنفس المنوال يمكن أن يكون المجتمع هو مساكن الجامعة - العائلات - الطيور - الأسماك - الشركات التجارية - الشركات الصناعية ..... فكل منها يعتبر مجتمع إحصائى ، فالمجتمع ما هو إلا

مجموعة مفردات تجمعها خصائص معينة . لهذا فإن المجتمع من الناحية الإحصائية قد يكون حقيقى Real أو افتراضياً Hypothetical كما قد يكون محدود Determinant أى يمكن حصر جميع مفرداته مثل مجتمع الطلاب فى الجامعات - عدد العائلات فى محافظة الغربية ...... الخ ، أو مجتمع غير محدود Non-Determinant لا يمكن عد أو حصر جميع مفرداته مثل مجتمع الأسماك أو مجتمع النباتات (ثمار المانجو أو العنب ..... الخ).

ثانيا: العينة Sample

العينة هى جزء من مفردات المجتمع يتم اختبارها بحيث تكون ممثلة للمجتمع ككل مثل عينة من سكان ام القرى - عينة من درجات امتحانات الطلاب فى مادة معينة ولتكن الإحصاء - عينة من أطوال طلاب كلية التجارة ...... الخ وتسمى القيمة التى تحسب لوصف الخاصية المعينة لوحدات العينة تقديراً Estimator or Statistic.

ومن ثم كلما زاد حجم العينة (عدد مفرداتها) كلما اقترب التقدير Statistic المحسوب منها من قيمة المعلمة Parameter المطلوب تقديرها.

وغالباً يلجأ الباحث أو الإحصائى فى دراسته إلى العينات بدلاً من الحصر الشامل لأن الأخير قد تكون تكلفته باهظة أو يحتاج لوقت طويل ...... الخ.

وجدير بالذكر أن طريقة اختيار وحدات العينة يطلق عليها "طريقة المعاينة" Sampling Method حيث ينبغى اختيار العينة بحيث تكون ممثلة للمجتمع.

وتجدر الإشارة بأن الباحثين غالباً يستخدمون العينات بدلاً من الحصر الشامل عند إجراء بحوثهم وهذا يعزى لأسباب كثيرة أهمها توفير وقت وجهد الباحث فضلاً عن الاقتصاد فى التكاليف اللازمة لجمع البيانات وتلخيصها ، ومن ثم نتساءل هل هناك عدة أنواع للعينات؟

نرد ونقول بلا ريب هناك عدة أنواع للعينات تنحصر بإيجاز فيما يلى:

أ- العينة بالحصة Quota Sample

ب- العينة الاحتمالية Probability Sample

جـ - العينة العشوائية البسيطة Simple Random

د- العينة العشوائية المنتظمة Systematic Random Sample

هـ - العينة الطبقية Stratified Sample

و- العينة ذات المراحل المتعددة Multi-Stages Sample

وهناك عدة شروط وقواعد ينبغى اتباعها حتى لا تكون العينة متحيزة أو صغيرة ولا تمثل مجتمع البحث ، هذه القواعد تتلخص فى الآتى:

أ- أن تكون العينة صادقة ما أمكن فى تمثيل المجتمع الإحصائى المراد دراسته بالحصول على بيانات من أفراده . فمثلاً إذا كان المراد معرفة كمية حديد التسليح التى تباع خلال شهر معين فيجب عدم اختيار متاجر الأسمنت ضمن أفراد العينة (لأن العينة تشمل متاجر حديد التسليح وليس متاجر الأسمنت).

ب- ضرورة توافر قائمة بمفردات المجتمع المراد معاينته (اختيار العينة منه) حيث تساعد هذه القائمة على تحديد نوع العينة الممكن سحبها من المجتمع (عينة عشوائية - منتظمة - طبقية - متعددة المراحل ...... الخ).

ج- استخدام العينة ينبغى أن ينجم عنه ضآلة التكاليف مقارنة بالحصر الشامل.

د- إمكانية قياس دقة نتائج العينة وتقدير حدود الخطأ فيها.

فضلاً عما سلف ينبغى اختيار وحدة المعاينة قبل إجراء المعاينة ذاتها ...... هل وحدة المعاينة هى الفرد أم الأسرة ...... الخ ، كذلك يجب تحديد حجم العينة ونوعها كما يتبين لنا فى البند التالى (3-3-1) ، بما يتواءم مع الميزانيات المالية المقررة لجمع بيانات العينة ... الخ كما ينبغى تجنب التحيز حتى تكون العينة ممثلة للمجتمع تمثيلاً حقيقياً أقرب ما يكون إلى الدقة.

■ البيانات Data

ويقصد بالبيانات تلك المعلومات أو المشاهدات التى يتم جمعها وتنظيمها وتحليلها بواسطة الباحثين أو الإحصائيين. ويوجد نوعان من هذه البيانات هى:

أولا: البيانات الوصفية Qualitative Data

وتعنى البيانات التى تكون فى صورة غير كمية (غير عددية) مثل الجنس - الديانة - لون الشعر - فصيلة الدم - الحالة الاجتماعية - تقديرات الطلاب فى الجامعة.

ثانيا : البيانات الكمية Quantitative Data

وتعنى البيانات التى تكون فى صورة عددية مثل درجات الحرارة - عدد الحوادث - قيم الدخل - عدد أفراد الأسرة - الأطوال - الأوزان - درجات الطلاب - عدد سكان المدن والقرى - عدد العمال بالمصنع ...... الخ.

وجدير بالذكر أن البيانات بنوعيها قد يجمعها الباحث من مصادرها المباشرة وفى هذه الحالة يطلق عليها "بيانات أولية" أو يجمعها من تقارير أو سجلات دورية وهى ما يطلق عليها "بيانات ثانوية" كما هو الحال فى سجلات أو نشرات الجهاز المركزى للتعبئة والإحصاء.

ثالثا: المتغيرات وأنواعها Variables and its Types

يوجد نوعان من المتغيرات العشوائية هما:

1- المتغير العشوائى المتقطع Discrete Random

يقصد بالمتغير العشوائى المتقطع المتغير الذى يأخذ قيم فى شكل قفزات بمعنى أنه يأخذ قيم تقع عند نقاط منفصلة وليست قيم بين القفزات أى أن هذا المتغير ينجم من عملية العد مثل حالة السلعة جيدة أم معيبة $X = 0, 1$ أو عدد أفراد الأسرة .... $X = 1 , 2 , 3$ أو عدد العمال فى مصنع $X = 20 , 21 , 22$ ..... أو عدد الزوجات فى عصمة رجل $X = 1 , 2 , 3 , 4$ أو عدد طلاب كلية $X = 8000 , 10000 , 15000$ ...... أو عدد حوادث السيارات ...... $X = 0 , 1 , 2 , 3$ ، لهذا فإن المتغير العشوائى المتقطع لا يمكنه أن يأخذ قيم كسرية بين 1 , 0 أو بين 3 , 2 , 1 ..... الخ ، فمثلاً لا يمكن أن يقال أن عدد أفراد الأسرة 2.5 فرد أو عدد طلاب كلية 8000.5 طالب ..... الخ أو عدد حوادث السيارات 3.5 حادثة ...... الخ.

2- المتغير العشوائى المتصل أو المستمر Continuous Random Variable

يقصد بهذا المتغير الذى يأخذ أى قيمة ممكنه فى مدى معين (عدد لا نهائى من القيم العددية بين نقطتين)، مثال لذلك وزن الطالب الذى يريد الالتحاق بكلية التربية الرياضية بالكيلوجرام قد يكون 75.45 كيلوجرام أو 75.5 كيلوجرام - كذلك طول الطالب بالسنتيمتر قد يكون 175.5 سنتيمتر أو 175.8 سنتيمتر - درجة الحرارة - سرعة السيارة بالكيلومتر - ...... الخ . لذا فإن المتغير العشوائى المتصل دائماً ينجم من عملية قياس.

■ معالم المجتمع والعينة

لما كانت الدراسات الإحصائية تتناول دراسة المتغيرات العشوائية لذا فإن هذه المتغيرات تعتمد على بعض المقاييس العددية الوصفية مثل متوسط المتغير العشوائى المحسوب من كل مفردات (مشاهدات) المجتمع أو نسبة مفردات المجتمع التى تتوفر فيها صفة معينة وتسمى هذه المقاييس بمعالم المجتمع Population Parameters .

لذا فإن معلمة المجتمع Population Parameter ما هى إلا مقياس يصف خاصية من خصائص المجتمع وهى ترافق دائماً المتغير العشوائى وتميز المجتمع عن غيره من المجتمعات حيث لكل مجتمع متوسط يطلق عليه معلمه بمقتضاها يمكن تحديد المجتمع.

ولما كان من الصعب دراسة كل مفردات المجتمع لذا فإن غالبية الدراسات تقتصر على عينة من المجتمع لإيجاد ما يسمى بالإحصاء أو التقدير Estimator or Statistic وهو مقياس يصف خاصية من خصائص العينة وتتحدد قيمته من مفردات العينة . هذا الإحصاء أو التقدير هو بالفعل متوسط العينة (معلمة العينة) أو نسبة مفردات العينة التى تتوافر فيها صفة معينة.

وتجدر الإشارة بأن معلمة العينة (الإحصاء أو التقدير أو متوسط العينة) تفيدنا بأنها مقاييس تصف العينة ذاتها فضلاً عن أنها تمكننا من عمل الاستدلال حول معالم المجتمع التى تم اختيار العينة منه.

■ البحث العلمى ومراحل البحث الإحصائى

إن البحث العلمى ليس عملية روتينية بل هو ابتكار وتقدم فكرى لذا فإن الباحث ينبغى أن يكون حاذق وقادر على اختيار وتحديد مشكلة البحث التى قد تروق له . حيث يبدأ البحث العلمى عادة بالمشاهدة أو الملاحظة Observation التى يقوم بها الباحث وتثير فى نفسيته إحساساً بمشكلة معينة خاصة بظاهرة

معينة يترتب عليه وضع فرض علمى تفسيراً للظاهرة أو المشكلة ثم محاولته الحصول على البيانات والمعلومات التى تساعده بعد استقرائها وتحليلها على إثبات صحة هذا الفرض أو تفنيده . ومن خلال تحليل النتائج يمكن للباحث الخروج أو استنباط نظرية عامة أو خاصة لها فعاليتها يمكن تعميمها بعد ذلك طالما أن هذه النظرية ذات صفة علمية.

فضلاً عن ذلك فإن طرق البحث العلمى تعتبر هى الوسيلة الفعالة لإحداث أى تطور أو تقدم منشود فى كافة مجالات العلم والمعرفة فعن طريق البحث العلمى يمكن التوصل إلى معلومات ونتائج جديدة وأساليب ونظريات حديثة تساهم فى استكمال البناء المعرفى للعلم وتقديم حلول عملية للمشاكل التطبيقية التى تواجه العاملين فى نطاق هذا العلم.

وتعتبر الأساليب الإحصائية فرض عين على كل باحث يريد أن يلتزم بأصول البحث العلمى فيما يبغى إجراؤه من أبحاث يستوى فى ذلك أن يكون الباحث (صاحب المشكلة) فى محراب العلم أو فى مجال آخر.

وهناك عدة مراحل يمر بها أى بحث إحصائى أو بمعنى آخر هناك سلسلة من المراحل ينبغى أن يسلكها البحث الإحصائى . هذه المراحل تتلخص فى الآتى:

أولا: تحديد المشكلة موضع البحث Statement of the problem

وهذه المرحلة هى فى الواقع أول خطوة ينبغى أن يخطوها الباحث فالشعور بالمشكلة وتحديدها هو نقطة البداية وأول الطريق فى البحث العلمى على وجه العموم فبدون ذلك لا يستطيع الباحث أن يحدد لنفسه فكرة واضحة وشاملة عن نوعية البيانات التى لا بد وأن يسعى إلى جمعها ومن ثم لا تبدأ المرحلة التالية لهذه المرحلة (مرحلة جمع البيانات) ، لهذا إذا أهمل الباحث هذه الخطوة وظن أنه يعرف المشكلة جيداً فى حين أنه فى الواقع لا يعرفها بالتحديد فإنه لا يستطيع أن يخطو الخطوات التالية فى البحث الإحصائى.

ثانيا : جمع البيانات الخاصة بالبحث

بعد تحديد الباحث لمشكلة البحث تحديداً دقيقاً ينبغى عليه القيام بجمع البيانات الخاصة بها غير أنه يجب عليه معرفة ماهية البيانات التى سلف أن جمعت فى نفس الموضع الذى بداخله المشكلة حتى لا يضيع الوقت والمجهود فى إعادة جمع هذه البيانات مرة أخرى.

وتجدر الإشارة بأن جمع البيانات تبدأ بمشاهدة أو ملاحظة الظواهر التى يبحثها الباحث وتسجيلها بطريقة يسهل الرجوع إليها واستنباط القوانين التى تسير تبعاً لها هذه الظواهر . وكل ما على الباحث أن

يسجل الحقائق التى يجمعها عن الظواهر فى صيغة رقمية وهى ما يطلق عليها البيانات الأولية (البيانات الخام).

ثالثا: تصنيف البيانات وعرضها جدولياً وبيانياً

بعد الانتهاء من مرحلة جمع البيانات ينبغى على الباحث القيام بتصنيف البيانات وذلك بوضع المشاهدات المتماثلة فى مجموعات بحيث تشترك المشاهدات الكائنة بكل مجموعة فى خاصية معينة تميزها عن غيرها من المشاهدات فى المجموعات الأخرى.

وعند الانتهاء من تصنيف البيانات يبدأ الباحث فى التفكير فى كيفية عرضها هل وضعها فى شكل جداول أم فى شكل رسوما بيانيه ...... الخ وأيهما أفضل لذا فيتبادر إلى ذهنه عدة أسئلة حول طرق عرض البيانات ... وعموماً فإن الرسوم البيانية ليست بدائل لجداول البيانات ولكنها تعتبر طريقة لتحليلها.

رابعا : التحليل الإحصائى للبيانات

أحياناً يكتفى الباحث (أو متخذ القرار) بالجداول أو الرسوم البيانية لتحليل البيانات إلا أنها لا تصل بالباحث إلى النتائج المرغوبة التى يزمعها ...... لهذا فإن الباحث عندما يرى أنه فى حاجة إلى قدر كبير من التحليل الإحصائى فلا بد له أن يتغلغل ويغوص فى الأساليب الإحصائية واختيار الأسلوب الإحصائى الذى يتوافق مع مشكلته والبيانات التى قام بتجميعها وتصنيفها لهذه المشكلة.

من ثم تنبع أهمية الأساليب الإحصائية المختلفة فى عمل البحث الإحصائى ... وهى الأساليب التى سنقوم بدراستها معاً أيها القارئ العزيز فى متن هذا الكتاب حيث بمقتضى هذه الأساليب يتمكن الباحث من استخلاص النتائج Deductions التى قد تكون من التعقيد بمكان إذا ما حاول الباحث وصفها باللغة العادية ، فضلاً عن ذلك فإن هذه الأساليب الإحصائية تعين الباحث على اختبار الفروض Testing of Hypotheses وتعميم نتائج البحث على مجتمع البحث.

■ أهمية دراسة الأساليب الإحصائية فى العلوم الاجتماعية والإنسانية

بلا ريب أن العاملين فى مجال العلوم الاجتماعية والإنسانية كثيرون منهم المحاسبون والاقتصاديون ورجال الإدارة الذين يعملون فى المدارس أو الجامعات أو الشركات والهيئات كذلك الباحثين الذين يتولون التدريس فى الجامعات لأجيال متتالية من الشباب والشابات ...... كلهم فى حاجة ماسة للأساليب الإحصائية لمساعدتهم فى تحليل دراساتهم والوصول إلى نتائج كمية تساعدهم على اتخاذ

القرارات . فلا يمكن لأى منهم رفع مستواه العلمى بدون الإطلاع وتنفيذ البحوث العلمية التى تعتمد على الأساليب الإحصائية.

من ثم فإن الأساليب الإحصائية باتت ذات أهمية بالغة لكافة العاملين فى المجالات المختلفة فضلاً عن الباحثين فى الجامعات والمعاهد العلمية ومراكز البحوث كما يتضح لنا من النقاط التالية.

أولاً: أهمية الأساليب الإحصائية فى معالجة وتحليل وتفسير البيانات.

لما كانت الأساليب الإحصائية مهتمة بطرق جمع ومعالجة البيانات وتحليلها وتفسير النتائج التى يتم الحصول عليها من تحليل البيانات لذا فإن لهذه الأساليب أهمية خاصة لكافة الباحثين فى المجالات المختلفة (الجامعات أو غيرها) حيث بمقتضاها يمكنهم الوصول إلى قرارات هامة فعلى سبيل المثال ما معنى أن يحصل طالب بالجامعة على 50 درجة فى مادة الإحصاء فهل هذا الرقم يدل على شئ ؟ بالطبع لا يدل على شئ إطلاقاً إلا بعد القيام بإجراءات إحصائية تتمثل فى جمع ومعالج وتحليل البيانات المرتبطة بهذه الظاهرة (درجات مادة الإحصاء) حيث منها يستطيع الباحث أن يحدد مستوى الطالب هل ممتاز أم ضعيف هل مجموعة الطلاب التى ينتمى إليها هذا الطالب مثل بقية المجموعات الأخرى أم تختلف عنها ...... وماهية حدود هذا الاختلاف ...... الخ.

أما مرحلة تفسير البيانات فبلا ريب تعتمد على الإحصاءات الوصفية أو الإحصاءات التحليلية أو كلاهما معاً ...... وهذا سوف يتضح لنا داخل متن هذا الكتاب.

ثانيا: أهمية الأساليب الإحصائية فى تفسير وفهم نتائج البحث العلمى

كما ذكرنا سالفاً أن البحث العلمى يعتمد بدرجة كبيرة على الأساليب الإحصائية - لهذا حتى يستطيع الباحث الإطلاع على كافة البحوث ونتائجها وفهمها وفهم نتائجها وتطبيقها فإنه يحتاج دائماً إلى المزيد من المهارات الأساسية وعلى رأسها دراسة الأساليب الإحصائية التى تفيده فى الإطلاع على مثل هذه البحوث وتطبيق نتائجها.

الفصل الثانى
جمع البيانات وعرضها

■ **جمع البيانات**

تنقسم طرق جمع البيانات الإحصائية فى حالة "البحث الميدانى" إلى:

1- طريقة المشاهدة (الملاحظة)

وبمقتضى هذه الطريقة يمكن للباحث جمع بياناته عن طريق تدوين مشاهداته عن مفردات العينة التى حددها كما هو الحال عند دراسة مشكلة المرور وهى الدراسة التى تعطى المعلومات اللازمة لتخطيط الطرق . فى هذه الدراسة يقوم الباحث بتدوين عدد السيارات التى تمر عند نقطة معينة والزمن الذى تمر فيه من هذه النقطة . كما يمكن للباحث استخدام بعض الأجهزة الاتوماتيكية لتسهيل عملية عد السيارات عند نقطة مرور السيارات.

وتجدر الإشارة بأن هذه الطريقة تقتضى من الباحث أن يكون على دراسة وخبره عالية بموضوع البحث فضلاً عن حسن التصرف وسرعة البديهة عند تسجيل ملاحظاته ومشاهداته لأفراد العينة.

2- طريقة الأسئلة

وبمقتضى هذه الطريقة يمكن للباحث الحصول على البيانات التى يريدها من خلال أسئلة يوجهها لمفردات العينة إما عن طريق:

- الخطابات التى ترسل لأفراد العينة وانتظار الرد على الأسئلة.
- التليفون وذلك بالاتصال التليفونى المباشر بأفراد العينة ومعرفة إجابات الأسئلة مباشرة منهم عن طريق التليفون.
- المقابلة الشخصية وذلك بلقاء كل أفراد العينة شخصياً وتدوين إجاباتهم.
- استطلاع الرأى باستخدام شبكة الانترنت

والجدير بالذكر أن أفضل طريقة من الطرق السابقة هى التى تضمن الحصول على المعلومات المطلوبة وبأقل تكلفة ووقت ممكن ، وعادة ما توضع الاسئلة فى كشف يطلق عليه صحيفة استقصاء أو استمارة استقصاء Questionnaire

■ **تبويب البيانات جدولياً**

ذكرنا سلفاً أن البيانات التى يجمعها الباحث قد تكون مفردات مجتمع البحث بالكامل (الحصر الشامل) أو مفردات عينة يتم اختيارها من هذا المجتمع إلا أن هذه البيانات فى صورتها الأولية تسمى

البيانات الخام Raw Data . ولما كانت هذه البيانات الخام الخاصة بالظواهر التى تعترضنا فى أبحاثنا ودراستنا لا توضح هذه الظواهر ولما كان علم الإحصاء (الأساليب الإحصائية) هدفه الأساسى هو التعرف على خصائص هذه الظواهر الخاصة بالبيانات الأولية (الخام) التى تم جمعها بشكل مباشر أو غير مباشر . لذا فإن التعرف على هذه الخصائص يقتضى تبويب البيانات الأولية (الخام) وتصنيفها فى شكل جداول لتسهيل تفسيرها وتحليلها.

مثال لذلك

عينة لأوزان 100 طالب يريدون الالتحاق بكلية التربية الرياضية ...... فما قيمة هذه الأرقام بصورتها الحالية (بيانات خام) فليس لها معنى أو دلالة علمية أو إحصائية حيث لا يمكننا تحليلها والحصول منها على نتائج مفيدة.

لهذا يكون من الضرورى تلخيص هذه البيانات ووضعها فى صورة تسهل عملية الاستفادة منها وذلك بتبويبها لفهمها والتوصل منها إلى نتائج ذات معنى ودلالة إحصائية وعلمية تساعد الباحثون ومتخذى القرار على اتخاذ القرارات الرشيدة السليمة.

ولما كانت البيانات تنقسم إلى بيانات وصفية Qualitative وبيانات كمية Quantitative لذا فهناك عدة طرق لتبويب هذه البيانات بنوعيها وعرضها جدولياً والتى تتضمن إعداد جداول تكرارية بصورها المختلفة والتى تنقسم إلى نوعين:

1- الجداول التكرارية البسيطة. 2- الجداول التكرارية المزدوجة.

أولا: تبويب البيانات الوصفية Qualitative Data

يقصد بالبيانات الوصفية البيانات غير الكمية أى البيانات التى تتسم بصفات معينة مثال لذلك تقديرات امتحان الطلاب فى كليات الجامعة التى تقسم إلى (مقبول - جيد - جيد جداً - ممتاز) أو تقسيم المجتمع إلى (أمى - متعلم) أو تقسيم الحالة الاجتماعية إلى (أعزب - متزوج - متزوج ويعول) ...... الخ.

وتجدر الإشارة بأن البيانات الوصفية يتم تبويبها فى شكل جداول تكرارية كما هو مبين لنا فيما يلى:

1- الجداول التكرارية البسيطة للبيانات الوصفية

ويقصد بهذه الجداول تلك الجداول التى تلخص وتبوب البيانات الوصفية لصفة واحدة فى أبسط صورة كما هو مبين فى المثال التالى:

مثال 1

فيما يلى تقديرات 44 طالباً من كلية التجارة فى مادة المحاسبة:

A	B	C	D	D	E	B	C	A	E
C	C	C	B	A	B	C	B	B	C
C	E	D	C	C	D	B	E	A	B
D	C	C	C	B	C	B	C	C	C
B	D	C	D						

علماً بأن A ممتاز ، B جيد جداً ، C جيد ، D مقبول ، E راسب

المطلوب: إعداد جدول التوزيع التكرارى لبيانات هذه التقديرات؟

الحل

لكى نعد جدول توزيع تكرارى ينبغى عمل جدول آخر لتبويب هذه البيانات يطلق عليه جدول تفريغ لتقديرات الـ 44 طالباً هذا الجدول يقسم إلى ثلاثة أعمدة هى:

العامود الأول تدون قيمة تقديرات الطالب (A, B, C, D, E) أما العامود الثانى فهو خاص بالعلامات التى بمقتضاها يمكن معرفة عدد الطلاب فى كل تقدير من التقديرات الخمس أما العامود الثالث فهو خاص بعدد الطلاب فى كل تقدير.

ولتفريغ تقديرات الطلاب الـ 44 فى جدول التفريغ نقرأ أول تقدير وهو (A) ثم نضع علامة أمام التقدير (A) وعادة فإن هذه العلامة تكتب على شكل خط مائل (/) ثم نقرأ التقدير الثانى (C) (القراءة تتم عامود) فنضع علامة أمام التقدير (C) ...... وهكذا . وعندما تبلغ عدد العلامات خمسة فتكتب العلامة الخامسة فى ميل عكس العلامات الأربعة الأولى ليتقاطع معها ويطلق عليها "حزمة خماسية" ...... وذلك لتسهيل عملية عد العلامات كما هو مبين فى جدول التفريغ التالى:

جدول (1) تفريغ تقديرات 44 طالباً فى مادة المحاسبة بكلية التجارة

التقدير	العلامات	عدد الطلاب
A	////	4
B	~~////~~ ~~////~~ /	11
C	~~////~~ ~~////~~ ~~////~~ ///	18
D	~~////~~ //	7
E	////	4
المجموع		44

وبعد انتهاءك من التفريغ يجب عليك أن تبدأ العد لكل فئة من الفئات مع التأكد بأن المجموع الكلى يساوى المجموع الفعلى لتقديرات الـ 44 طالباً ومن جدول التفريغ السالف يمكن اشتقاق جدول التوزيع التكرارى البسيط المناظر وذلك عن طريق الاكتفاء بالعامودين الأول والثالث وحذف عامود العلامات المائلة ومن ثم فإن جدول التفريغ التكرارى البسيط لتقديرات 44 طالب هو:

جدول (2) توزيع تكرارى لتقديرات 44 طالب فى المحاسبة بكلية التجارة

التقدير	عدد الطلاب	التكرار النسبى
A	4	0.091
B	11	0.250
C	18	0.409
D	7	0.159
E	4	0.091
المجموع	44	1.00

ملحوظة هامة

العامود الأخير فى جدول (2) يمثل نسبة تكرار كل تقدير إلى العامود الكلى للطلاب.

على سبيل المثال التكرار النسبى للتقدير $\boxed{\beta = \frac{11}{44} = 0.250}$ ...... وهكذا

2- جدول تكرارى مزدوج للبيانات الوصفية

ويقصد بهذه الجداول تلك الجداول التى تلخص وتبوب البيانات الوصفية لصفتين فى وقت واحد كما هو مبين فى المثال التالى:

مثال 2

فيما يلى تقديرات 36 طالب فى مادتى الإحصاء والاقتصاد من طلاب الفرقة الثانية بكلية التجارة - جامعة كفر الشيخ

الإحصاء	الاقتصاد	الإحصاء	الاقتصاد	الإحصاء	الاقتصاد	الإحصاء	الاقتصاد
C	C	C	C	D	C	D	D
C	C	D	A	C	C	C	D
D	D	C	D	B	D	D	C
B	B	B	C	D	C	C	C
B	C	D	B	C	B	B	B
A	A	A	A	A	C	D	B
D	D	D	C	C	B	C	D
C	C	B	B	C	C	B	C
B	B	B	A	C	B	D	A

المطلوب: إعداد جدول توزيع تكرارى مزدوج؟

الحل

لإعداد جدول توزيع تكرارى مزدوج ينبغى كما ذكرنا أنفا عمل جدول تفريغ حيث يشمل هذا الجدول تقديرات مادة الإحصاء رأسياً وتقديرات مادة الاقتصاد أفقياً ثم إضافة صف أخير لإيجاد مجموع مادة الإحصاء كذلك إضافة عامود آخر لإيجاد مجموع مادة الاقتصاد.

وبعد ذلك يتم استقراء تقديرات الطالب الأول وهو الطالب الذى حصل على تقدير (D) فى الاقتصاد و (D) فى الإحصاء وبالتالى توضع علامة مائلة (/) فى العامود الخامس تحت (D) الاقتصاد وفى الصف الخامس أمام (D) فى الإحصاء وبالمثل بالنسبة للطالب الثانى توضع علامة (/) مائلة فى العامود الخامس تحت (D) فى الاقتصاد وفى الصف الرابع أمام (C) فى الإحصاء ...... وهكذا إلى أن ننتهى من 36 طالب ثم نوجد المجموع لكل صف ولكل عامود بشرط أن يكون المجموع الكلى مساوياً عدد الطلاب وهو 36 كما هو مبين فى جدول التفريغ التالى:

جدول (3) تفريغ مزدوج لتقديرات 36 طالب فى مادتى الإحصاء والاقتصاد بكلية التجارة

الاقتصاد / الإحصاء	A	B	C	D	المجموع
A	//		/		3
B	/	////	///	/	9
C		///	~~////~~ //	///	13
D	//	//	////	///	11
المجموع	5	9	15	7	36

ومن جدول التفريغ السالف يمكن اشتقاق جدول التوزيع التكرارى المزدوج عن طريق استبدال العلامات بالأرقام الممثلة لها كما هو مبين فى جدول (4) التالى:

جدول (4) توزيع تكرارى مزدوج لتقديرات 36 طالب فى مادتى الإحصاء و الاقتصاد

الاقتصاد / الإحصاء	A	B	C	D	المجموع
A	2		1		3
B	1	4	3	1	9
C		3	7	3	13
D	2	2	4	3	11
المجموع	5	9	15	7	36

ثانيا: تبويب البيانات الكمية Quantitative Data

تقسم البيانات الكمية إلى بيانات متقطعة وأخرى مستمرة . وسلف أن نوهنا أن البيانات الأولى (المتقطعة) تختص بقياس المتغيرات المتقطعة وهى التى تأخذ قيم معينة فقط فى شكل قفزات ولا تأخذ أى قيمة بين هذه القيم مثل عدد الطلاب فى كلية معينة -عدد الكتب الدراسية - عدد حوادث السيارات .. الخ فهى أعداد صحيحة لا يمكن تجزئتها أما البيانات المستمرة تختص بقياس المتغيرات المستمرة وهى التى يمكن أن تأخذ أى قيمة بين قيمتين مثل أطوال الطلاب- أوزان الطلاب- درجات امتحان الطلاب- درجات الحرارة .. الخ ، وتبوب البيانات الكمية فى شكل:

1- جداول تكرارية بسيطة 2- جداول تكرارية مزدوجة

1- الجداول التكرارية البسيطة للبيانات الكمية

وهى الجداول التى تأخذ صور عديدة تتوقف على حجم البيانات والغرض من تبويبها حيث تأخذ هذه الجداول عدة صور منها:

أ- جدول تكرارى بسيط بحسب الأعداد.

ب- جدول تكرارى بسيط بحسب الفئات المتساوية.

ج- جدول تكرارى بسيط بحسب الفئات غير المتساوية.

ويمكن عرض هذه الجداول من خلال أمثلة عملية كما هو مبين فيما يلى:

أ- جدول تكرارى بسيط بحسب الأعداد (البيانات المتقطعة)

مثال 3

فى عينة من 30 وثيقة من وثائق تأمين السيارات المؤمن عليها لدى شركة مصر للتأمين تبين أن عدد الحوادث التى وقعت لهذه السيارات كالآتى:

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

المطلوب: إعداد جدول توزيع تكرارى بسيط لهذه الحوادث؟

الحل

لإعداد جدول توزيع تكرارى بسيط ينبغى كما ذكرنا أنفا إعداد جدول تفريغ للبيانات وذلك بتحديد المدى وهو الفرق بين أكبر قيمة وأصغر قيمة ، وفى مثالنا هذا يكون المدى مساوياً (5 = 0 - 5) لذلك نكون جدول تفريغ لعدد الحوادث يتكون من ثلاثة أعمدة العامود الأول يخصص لعدد الحوادث ويحتوى على القيم من 0 إلى 5 والعامود الثانى يخصص للعلامات المائلة أما العامود الثالث فيخصص لعدد السيارات كما هو مبين فى الجدول التالى:

جدول (6) تفريغ لعدد حوادث السيارات فى 30 وثيقة تأمين سيارات

عدد الحوادث	العلامات	عدد السيارات
0	///	3
1	////	4
2	~~////~~ ~~////~~	10
3	~~////~~ ///	8
4	///	3
5	//	2
المجموع	30	30

ومن جدول التفريغ السابق يمكن إيجاد جدول توزيع تكرارى بسيط لعدد حوادث السيارات فى 30 وثيقة من وثائق تأمين السيارات وذلك بالاكتفاء بالعامودين الأول والثالث كما هو مبين فى الجدول التالى:

جدول (7) توزيع تكرارى بسيط لعدد حوادث السيارات فى 30 وثيقة تأمين سيارات

عدد الحوادث	عدد السيارات	التكرار النسبى
0	3	0.10
1	4	0.14
2	10	0.33
3	8	0.26
4	3	0.10
5	2	0.07
المجموع	30	1.00

ملحوظة

العامود الأخير عبارة عن نسبة تكرار عدد الحوادث إلى العدد الكلى لحوادث السيارات.

ب ـ جدول تكرارى بسيط بحسب الفئات المتساوية

فى المثال السابق كان المدى يتمثل فى قيم صغيرة أما إذا كان المدى الممثل للفرق بين أقل قيمة وأكبر قيمة من قيم الظاهرة محل الدراسة كبيراً فإنه من الصعب إعداد جدول تكرارى بسيط بحسب الأعداد لمثل هذه البيانات ولكن الأمر يقتضى إعداد جدول تكرارى بسيط بحسب الفئات كما يتبين لنا من المثال التالى:

مثال 4

فيما يلى أطوال عينة من 80 طالب

165	170	180	170	165	185	187	161	187	191	180	190
185	180	175	182	181	175	174	186	170	180	196	171
175	171	184	172	175	182	186	182	172	176	190	176
200	180	173	181	176	188	173	188	178	182	191	195
179	172	197	187	205	184	174	194	197	174	184	179
180	182	173	193	191	182	171	181	192	182	196	174
179	179	185	186	172	196	195	196				

المطلوب: إعداد جدول تكرارى بسيط لهذه البيانات؟

الحل

من الجدير بالذكر إذا اتبعنا الطريقة السالفة وهى إعداد جدول تكرارى بسيط بحسب الأعداد للبيانات السابقة فإننا نضطر إلى إعداد جدول ذو أعداد كثيرة حيث أن أقل طول 165 وأكبر طول 205 فيكون المدى 40 (165 - 205) وهذا يؤدى إلى أن الجدول سيحتوى على 40 قيمة وتكراراتها ...... وهذا بلا ريب عدد كبير من القيم لا يمكننا من التحليل واستخلاص النتائج.

لهذا يجب صياغة البيانات بطريقة سهلة ومبسطة وذلك بتقسيمها إلى فئات Classes وحساب عدد الطلاب فى كل فئة أى حساب التكرار لكل فئة ، غير أنه ينبغى أن تضع فى ذهنك أيها القارئ العزيز أن عدد الفئات يجب ألا يكون صغيراً حتى لا ينمحى معه معالم التوزيع (أى يفقد التوزيع كثير من تفاصيله أو معالمه) ولا يكون عدد الفئات كبيراً حتى لا تصعب العمليات الإحصائية ويفقد تلخيص

البيانات أهميته . وعموماً يمكن القول أنه يجب ألا يقل عدد الفئات عن 5 فئات وألا يزيد عن 20 فئة ...... وعلى قدر الإمكان أن تكون الفئات متساوية.

وتجدر الإشارة أن هناك صيغة تستخدم لتحديد العدد التقريبى لعدد الفئات وهى:

عدد الفئات =	1 + 3.3 log n	(حيث	n هى عدد مفردات العينة)

ولإعداد جدول تكرارى فى مثالنا هذا يجب مراعاة الآتى:

1- تحديد المدى وهو (40 = 165 - 205)

2- تحديد عدد الفئات ويتحدد بالصيغة السابقة:

عدد الفئات = $1 + 3.3 \log 80 = 7.280$

3- وحيث أن عدد الفئات لا يمكن أن يكون عدد كسرى لذا فيجب تقريبه إلى 8 أو 7 وفى مثالنا هذا سنفترض أن عدد الفئات = 8 فئات (من الممكن أن نقرب إلى 7).

4- تحديد طول الفئة ويتحدد على أساس المدى مقسوماً على عدد الفئات أى طول الفئة = $\frac{40}{8} = 5$ ، لهذا فإن طول الفئة فى مثالنا سيكون 5.

ونبدأ بالفئة الأولى بحيث تشمل أقل طول وهو 165 إلى أقل من 170 كذلك الفئة الثانية تبدأ من 170 إلى أقل من 175 ...... وهكذا ، أى أن حدود الفئات تكتب هكذا

(165-, 170-, 175-, 180-, 185-, 190-, 195-, 200-205)

5- نعد جدول تفريغ للبيانات مكون من ثلاث أعمدة العامود الأول يحتوى على الفئات للسالف ذكرها والعامود الثانى يخصص لتفريغ البيانات عن طريقة قراءة الأرقام رقماً رقماً ووضع علامة مائلة لكل رقم فى الفئة التى تشمله وعند اكتمال خمسة علامات تكتب على هيئة حزمة والعامود الثالث فيخصص لعدد الطلاب وهو عدد التكرارات الموجود أمام كل فئة حيث أن هذه التكرارات هى التى تمثل التوزيع التكرارى مع ملاحظة أن عدد التكرارات الكلية يجب أن يتساوى مع عدد مفردات الظاهرة موضوع الدراسة وهو عدد الطلاب وذلك كما هو مبين فى الجدول التالى:

جدول (8) تفريغ أطوال 80 طالب يرغبون الالتحاق بكلية التربية الرياضية

الفئات	العلامات	عدد الطلاب (التكرارات)
165-	///	3
170-	𝍸 𝍸 ///	13
175-	𝍸 𝍸 𝍸 /	16
180-	𝍸 𝍸 𝍸 ///	18
185-	𝍸 𝍸 //	12
190-	𝍸 ///	8
195-	𝍸 ///	8
200-205	//	2
المجموع	80	80

ومن جدول التفريغ السالف يمكن تكوين جدول التوزيع التكرارى البسيط كالآتى:

جدول (9) توزيع تكرارات لأطوال 80 طالب يرغبون الالتحاق بكلية التربية الرياضية

الفئات	عدد الطلاب (التكرارات)	التكرار النسبى
165-	3	0.0375
170-	13	0.1625
175-	16	0.2000
180-	18	0.2250
185-	12	0.1500
190-	8	0.1000
195-	8	0.1000
200-205	2	0.0250
المجموع	80	1.000

■ **عرض البيانات بيانيا**

يعادل الرسم البيانى الواحد لظاهرة ما ألف كلمة أو عدة صفحات عن هذه الظاهرة ، مما أعطى أهمية كبيرة للرسم البيانى ليس فقط فى الأبحاث العلمية ولكن فى الحياة العلمية وأكبر مثال لذلك تلك التقارير الاقتصادية اليومية المصحوبة برسوم بيانية التى يراها مستمع النشرات الاقتصادية فى التليفزيون يومياً حيث يمكنه من ملاحقة الخبر من نظرة واحدة.

والجدير بالذكر أن الرسوم البيانية ليست بديلاً للجداول وإنما تساندها فى سرعة عرض الأرقام وتبسيطها ، لهذا يجب قبل البدء فى أى رسم بيانى تحديد الهدف منه ونوعه وحجمه وعنوانه.

وسوف نستعرض هنا أهم طرق عرض البيانات باستخدام الرسوم البيانية سواء للبيانات غير المبوبة (البيانات الخام) أو البيانات المبوبة (التوزيعات التكرارية) كما هو مبين فيما يلى:

أولا: الرسوم البيانية فى حالة البيانات غير المبوبة (البيانات الخام)

وتتمثل أهم الرسوم البيانية فى حالة البيانات غير المبوبة فى ثلاثة أنواع وهى:

1- الدائرة البيانية Pie Chart

الدائرة عبارة عن شكل هندسى يتم تمثيل البيانات بقطاعات داخلها بحيث ان مجموع هذه القطاعات تمثل مساحة الدائرة الكلية ويتم تحديد زاوية أى قطاع وفقا للقاعدة التالية.

$$\text{زاوية القطاع} = \frac{f_i}{N} \times 360$$

حيث f_i تمثل عدد التكرارات الخاصة بكل قطاع.

N تمثل عدد التكرارات الكلية.

مثال 7

إذا علمت أن إنفاق إحدى الدول على المنتخبات الرياضية المختلفة خلال فترة الاعداد للمشاركة فى الدورة الاوليمبية جاء كما هو مبين فى الجدول الآتى:

جدول (16) بيان الإنفاق الحكومى على المنتخبات المشاركة فى الدورة الاولمبية (مليون دولار)

اسم المنتخب	منتخب كرة القدم	منتخب كرة السلة	منتخب كرة الطائرة	المجموع
الانفاق	5	2	3	10

المطلوب: عرض هذه البيانات باستخدام الدائرة البيانية؟

الحل

ونظرا لان مجموع الزوايا داخل الدائرة تساوى 360^{o} درجة ومن واقع الجدول المعطى فى المثال يجب تقسيم الدائرة الى ثلاثة قطاعات الاول يمثل منتخب كرة القدم والثانى يمثل منتخب كرة السلة والثالث يمثل منتخب كرة الطائرة على الوجه التالى.

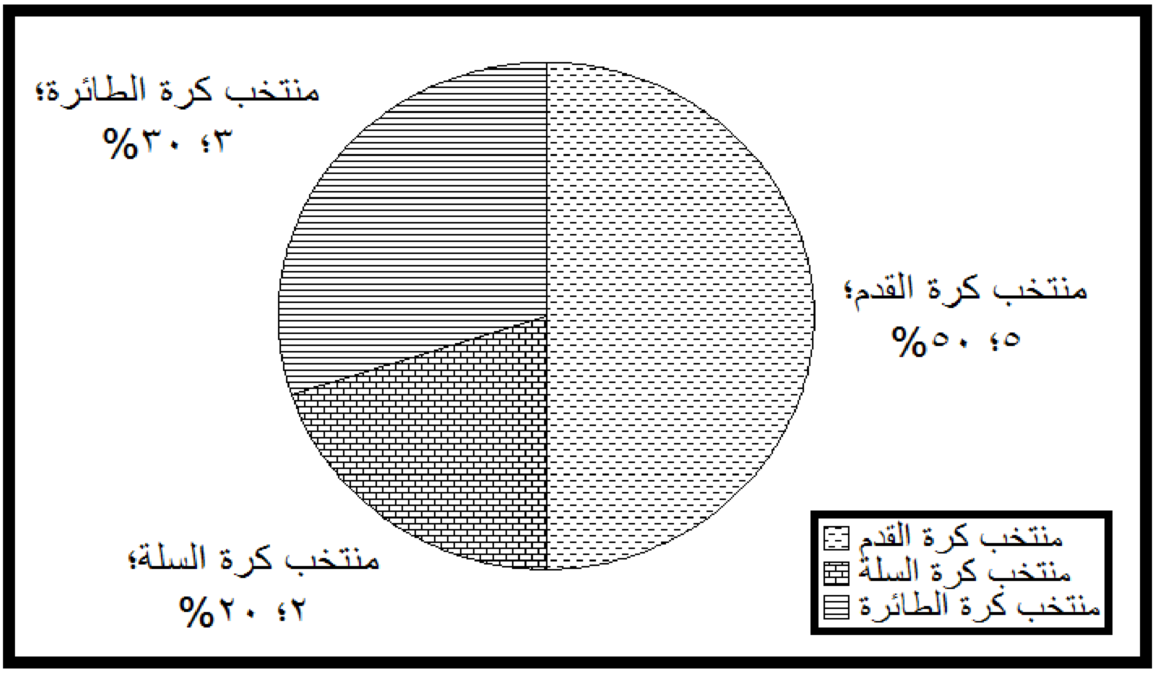
زاوية منتخب كرة القدم $= \frac{5}{10} \times 360 = 180^{o}$

زاوية منتخب كرة السلة $= \frac{2}{10} \times 360 = 72^{o}$

زاوية منتخب كرة الطائرة $= \frac{3}{10} \times 360 = 90^{o}$

وبتقسيم الدائرة التالية وفقا لدرجات الزوايا المحسوبة لكل قطاع نجدها على الشكل التالى.

شكل (2) الإنفاق على المنتخبات الرياضية بالمليون دولار



2- الأعمدة البيانية Bar Charts

من الجدير بالذكر أنه إذا كان لدى الباحث متغير وصفى أو متغير كمى متقطع Discrete فإن أنسب رسم بيانى لتمثيل هذا المتغير هو الأعمدة البيانية ، حيث أنها طريقة سهلة بمقتضاها يتم رسم أعمدة بيانية تتواءم أطوالها مع الصفات أو القيم التى تمثلها ، هذه الأعمدة ينبغى أن تكون لها قواعد متساوية فضلاً عن المسافات بينها يجب أن تكون متساوية . ويجب أن تكون الأعمدة مرتبة ترتيباً تنازلياً حتى يسهل إجراء المقارنة بينها وحتى يكون الرسم أكثر إيضاحا.

وتجدر الإشارة أن الأعمدة يمكن عرضها متباعدة ولكن يفضل أن يبدأ مقياس الرسم على المحور الرأسى من الصفر حتى لا تعطى الأطوال النسبية للأعمدة انطباعاً خطئا عند مقارنة الأعمدة ببعضها البعض (إذا بدأنا بقيمة غير الصفر).

علاوة على ذلك تستخدم الأعمدة البيانية عند مقارنة ظاهرتين أو أكثر وذلك برسم الأعمدة متلاصقة حتى تسهل عملية المقارنة بين الظواهر كما يتضح لنا من الأمثلة الآتية:

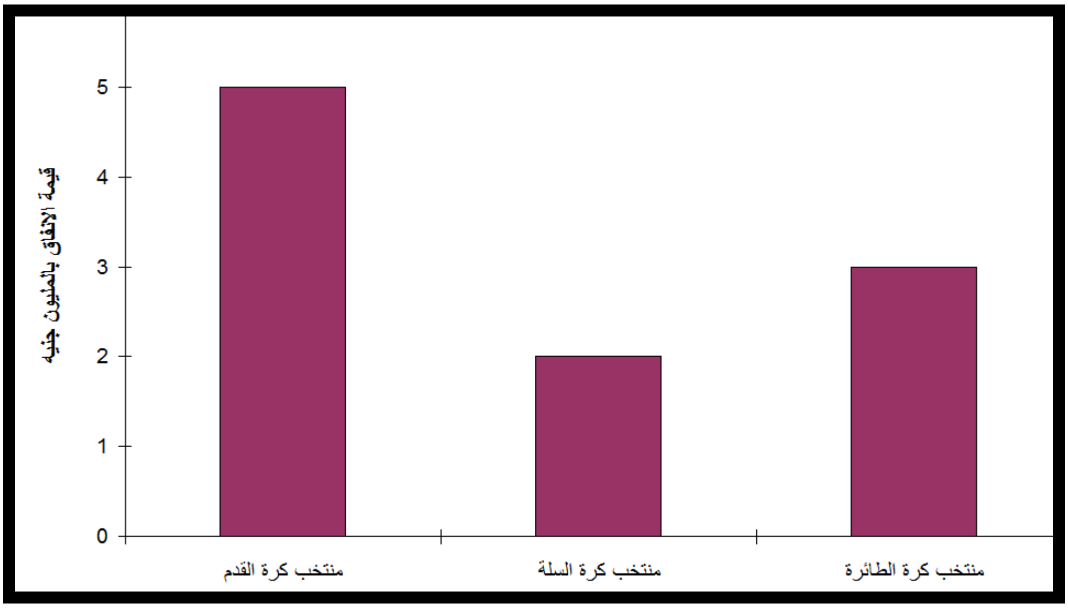
مثال

أعرض البيانات الواردة فى المثال السابق باستخدام الأعمدة البيانية؟

الحل

الأعمدة البيانية عبارة عن اشرطة بيانية لعرض متغير واحد ويخصص عمود لكل قطاع يحدد ارتفاعه طبقا لعدد التكرارت الخاصة به، ونعرض هنا بيانات المثال السابق باستخدام الاعمدة، والشكل التالى يوضح هذا.

شكل (2) الإنفاق على المنتخبات الرياضية بالمليون دولار



3- الخطوط البيانية Chart Lines

تعتبر الخطوط البيانية من أسهل الطرق لتمثيل البيانات حيث تستخدم لتمثيل العلاقة بين متغيرين أو أكثر بشرط أن يكون أحد هذه المتغيرات هو الزمن والذى يمكن تمثيلة على المحور الأفقى (كما هو الحال فى بيانات السلاسل الزمنية) أما المتغيرات الأخرى فتمثل على المحور الرأسى كما يتبين لنا من المثال التالى:

مثال

يوضح الجدول التالى اعداد الطلاب والطالبات الذين اجتازوا بنجاح أختبار مقرر مبادئ الاحصاء على مدار الخمس سنوات الاخيرة.

جدول (55) اعداد الطلاب والطالبات الذين اجتازوا بنجاح أختبار مقرر مبادئ الاحصاء

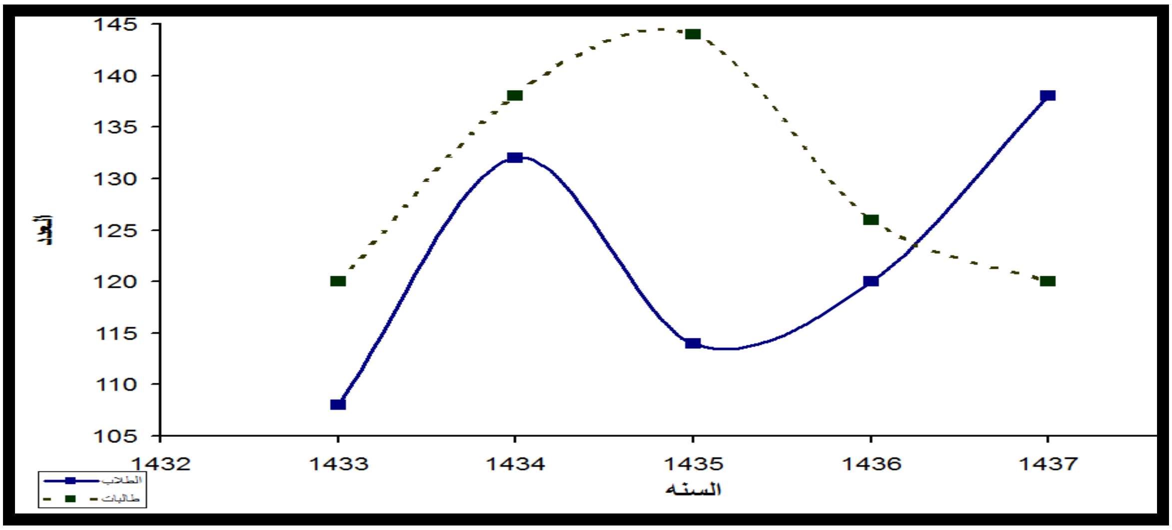
السنه	طلاب	طالبات
1433	108	120
1434	132	138
1435	114	144
1436	120	126
1437	138	120

أعرض البيانات الواردة فى الجدول السابق باستخدام الخطوط البيانية؟

الحل

حيث أن السنوات تمثل متغير الزمن لذلك فيتم تمثيلها على المحور الأفقى أما عدد الطلاب وعدد الطالبات فيتم تمثيلها على المحور الرأسى كما هو مبين لنا فى الشكل البيانى التالى:

شكل (6) الخطوط البيانية للاتجاه العام للطلاب والطالبات **الخاصة بإجتياز أختبار مقرر مبادئ الاحصاء**



ثانيا: الرسوم البيانية فى حالة البيانات المبوبة

يقصد بالبيانات المبوبة تلك البيانات التى تم تصنيفها وتبويبها فى شكل جداول تكرارية تحتوى على فئات Classes وتكرارات FrequEncies ، وهناك أربعة أنواع رئيسية من الرسوم البيانية لتمثيل هذه البيانات المبوبة تتمثل فى:

1- المدرج التكرارى Histogram

يتم استخدام المدرج التكرارى فى حالة إذا كان لدينا بيانات تمثل متغير كمى متصل والتى تقسم قيَّمة عادة إلى فئات Classes ، لهذا فإنه يتم اللجوء إلى المدرج التكرارى فى حالة التوزيعات التكرارية ذات الفئات حيث تندرج فكرة المدرج التكرارى فى تمثيل التوزيع التكرارى بمستطيلات تتواءم مساحاتها مع قيم التكرارات بحيث تكون قاعدة كل مستطيل هى طول الفئة.

وتجدر الإشارة أن ارتفاعات المستطيلات تساوى تكرارات الفئات التى ترسم فوقها لهذا بعد رسم كافة المستطيلات الممثلة للتكرارات للفئات المختلفة فى التوزيع التكرارى نحصل على ما يطلق عليه المدرج التكرارى Histogram .

مثال

الجدول التالى يوضح توزيع أعمار عينة من 20 شخص ، والمطلوب عرضها باستخدام المدرج التكرارى ؟

فئات السن	العدد
30-	1
40-	2
50-	5
60-	6
70-	4
80-90	2
المجموع	20

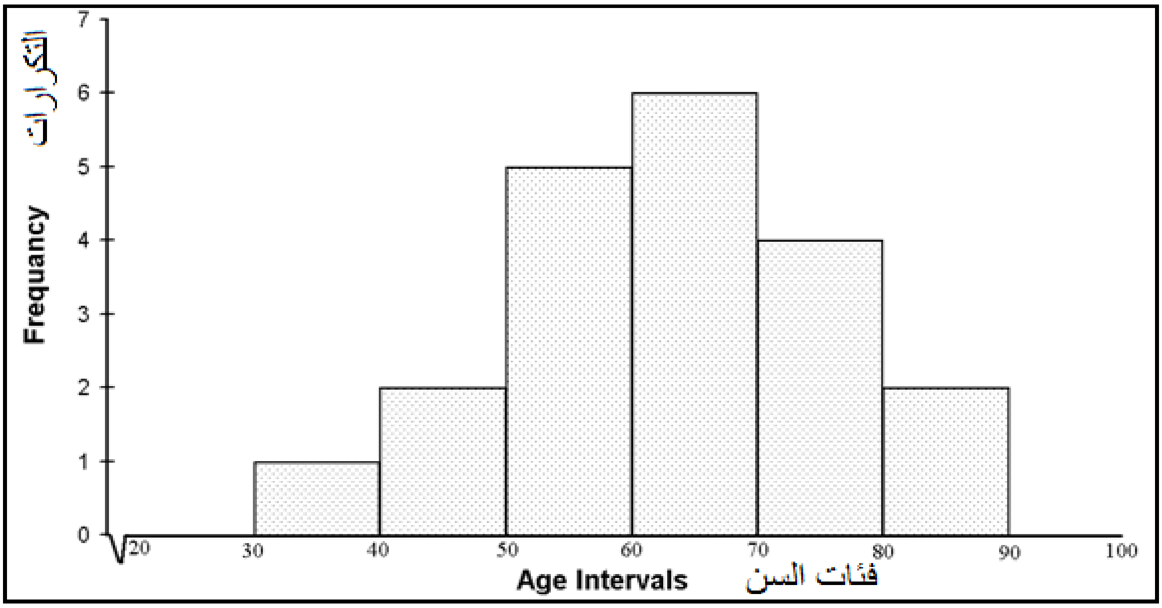
الحل

حيث أن الفئات متساوية لذا فيمكن رسم المدرج التكرارى بإتباع الخطوات الآتية:

1- يقسم المحور الأفقى إلى أجزاء متساوية وتمثل عليه الفئات حيث نبدأ بالحد الأدنى لفئة أصغر من الفئة الأولى ثم تالية الحدود الدنيا لباقى الفئات وينتهى تمثيل الفئات بالحد الأعلى للفئة الأخيرة.

2- يرسم على كل فئة مستطيلاً رأسياً قاعدته هى طول الفئة وارتفاعه قيمة التكرارات الخاصة بالفئة وذلك كما هو مبين بالشكل التالى.

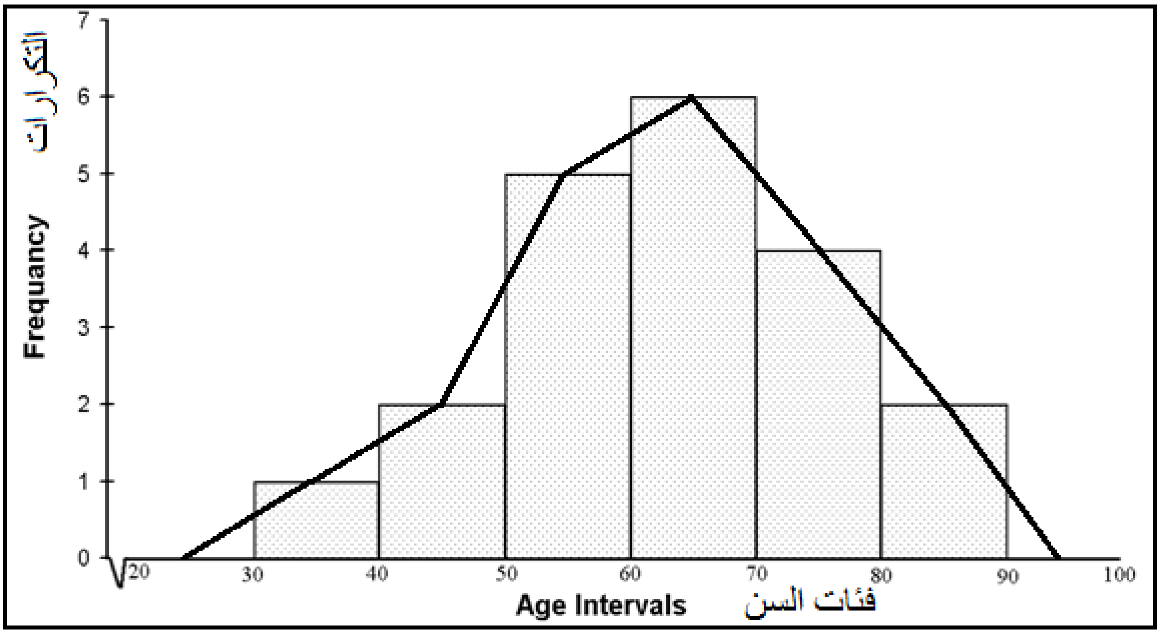
شكل (7) المدرج التكرارى لأعمار 20 شخص



2- المضلع التكرارى Frequency PolYgon

تجدر الإشارة بأنة يمكن رسم المضلع التكرارى على نفس المدرج التكرارى عن طريق توصيل منتصف القاعدة العليا لكل مستطيل من مستطيلات المدرج التكرارى بخطوط مستقيمة كما هو مبين فى الشكل التالى.

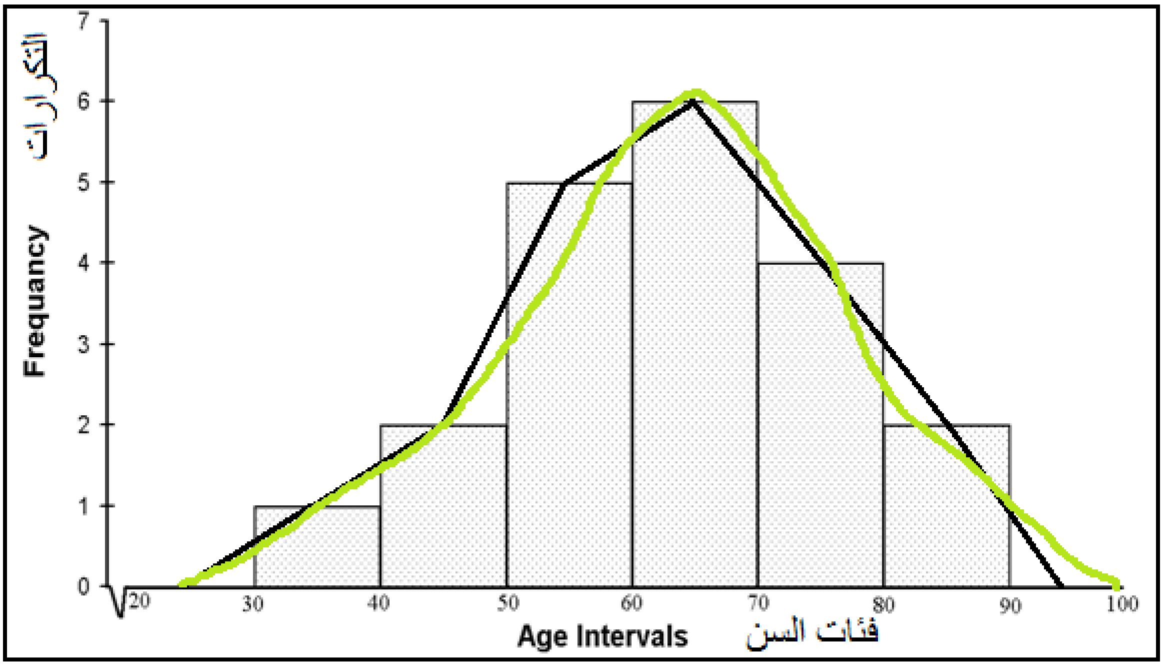
شكل (8) المدرج التكرارى والمضلع التكرارى لأعمار 20 شخص



3- المنحنى التكرارى Frequency Curve

يقترب شكل المضلع التكرارى أكثر فأكثر من شكل المنحنى إذا قمت بتمهيد الخطوط المنكسرة للمضلع التكرارى باليد فيتحول المضلع التكرارى فى النهاية إلى منحنى تكرارى مكا يتضح من الشكل التالى.

شكل (9) المدرج والمضلع والمنحنى التكرارى لأعمار 20 شخص



الفصل الثالث
مقاييس النزعة المركزية

■ **مقدمة**

كما نوهنا سالفاً أن البيانات تميل إلى التجمع أو التمركز حول قيمة معينة ، لذا فإن هذه القيمة يمكن استخدامها لتمثيل هذه المجموعة من البيانات ، والمقاييس المستخدمة فى التعرف على هذه القيمة المركزية يطلق عليها مقاييس النزعة المركزية Measures of Central Tendency أو المتوسطات Averages ، وهذه المقاييس ينبغى أن يتوافر فى كل منها عدة صفات لكى يكون مقياس جيد وهى:

أ- أن يتم حساب المقياس باستخدام كل المشاهدات (البيانات).

ب- أن يكون المقياس سهل الحساب والفهم.

ج- أن لا يتأثر المقياس بوجود القيم المتطرفة والشاذة (الأرقام الكبيرة جداً أو الصغيرة جداً).

د- أن تتوفر فى المقاييس القابلية للتعامل الجبرى.

ومن أهم مقاييس النزعة المركزية (المتوسطات) التى سيتم دراستها:

- المتوسط الحسابى Arithmetic Mean
- المتوسط الحسابى المرجح Weighted Arithmetic Mean
- الوسيط Median
- المنوال Mode
- المتوسط الهندسى Geometric Mean
- المتوسط التوافقى Harmonic Mean

وتجدر الإشارة بأن الصفات السالف ذكرها قد لا تتوفر كلها فى مقياس واحد ولكن كل مقياس يفضل استخدامه فى حالات دون حالات أخرى.

فضلاً عن ذلك فإننا نفترض فى كل المقاييس أن هناك متغير واحد (X) تم جمع بياناته من مجموعة من المفردات أو المشاهدات عددها (n) مفردة ويرمز لها بالرموز $x_1, x_2, x_3 ..... x_n$ تستخدم لحساب مقاييس النزعة المركزية كما هو مبين فيما يلى:

أولا: المتوسط الحسابى Arithmetic Mean

يعتبر المتوسط الحسابى أو ما يطلق عليه أحياناً الوسط الحسابى أكثر استخداماً بالمقارنة بالمتوسطات الأخرى (الوسيط ، المنوال ..... الخ) نتيجة لسهولة حسابه.

لذا يمكن للقارئ العزيز معرفة كيفية حسابه فى حالتى البيانات غير المبوبة والبيانات المبوبة كما يتضح لنا فيما يلى:

أ- حساب الوسط الحسابى فى حالة البيانات غير المبوبة:

من المعروف أن البيانات غير المبوبة هى البيانات الخام (المفردات - المشاهدات) فى صورتها الأولية لهذا يمكن حساب الوسط الحسابى بصورة رياضية كالآتى:

بفرض أننا جمعنا بيانات عن متغير (x) وكانت عدد المفردات أو المشاهدات أو القيم الخاصة بهذا المتغير $x_1, x_2, x_3 \ldots x_n$ وأن الوسط الحسابى لهذه المفردات يرمز له بالرمز ($\bar{x}$) فإننا نحصل على الوسط الحسابى لهذه القيم (المفردات) عن طريق قسمة مجموع هذه القيم (المشاهدات) على عددها (n) كالآتى:

$$\text{الوسط الحسابى } (\bar{x}) = \frac{\text{مجموع القيم}}{\text{عددها}} = \frac{\sum x}{n}$$

أى أن:

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

$$\boxed{\bar{X} = \frac{1}{n}\sum_{i=1}^{n} Xi = \frac{1}{n}\sum Xi} \qquad (1)$$

حيث: ($\sum$x) ترمز إلى مجموع المفردات أو المشاهدات

مثال 1

بفرض أن هناك عينة من 15 طالباً أطوالهم عند الالتحاق بكلية الشرطة كالآتى:

170	169	191	195	182	183	186	190
172	181	179	175	175	185	184	

المطلوب: حساب الوسط الحسابى لطول الطالب فى هذه العينة؟

الحل

$$\because \quad \bar{X} = \frac{1}{n}\sum Xi$$

$$\text{الوسط الحسابى } (\bar{X}) = \frac{170+164+191+195..........+185+184}{15}$$

$$(\bar{X}) = \frac{2715}{15} = 181.133 \text{ سنتيمتر}$$

مثال 2

أوجد الوسط الحسابى لتوزيع درجات 100 طالب فى امتحان مادة الحاسب الآلى ، إذا علمت أن التوزيع التكرارى لهؤلاء الطلاب على الوجه الآتى:

المجموع	85-90	80-	75-	70-	65-	60-	55-	50-	45-	40-	فئات الدرجات(X)
100	1	2	3	6	14	27	19	18	8	2	عدد الطلاب (f)

الحل

لإيجاد الوسط الحسابى ($\bar{x}$) لدرجات 100 طالب نستخدم القانون التالى:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{m} f_i x_i$$

ثم نتبع الخطوات التالية:

1- نوجد مراكز الفئات وهى (X_i) باعتبار أن التكرارات فى أى فئة تقع فى منتصف هذه الفئة أى فى مركزها حيث يمكن إيجاد مركز الفئة بالمعادلة الآتية:

$$\text{مركز الفئة} = \frac{\text{الحد الادنى للفئة + الحد الاعلى للفئة}}{2}$$

فمثلاً: مركز الفئة 40-45 هو 42.5 وهكذا .. لبقية الفئات

2- نضرب مركز كل فئة × تكرار نفس الفئة فنحصل على ($f_i x_i$).

3- نجمع كل القيم التى حصلنا عليها فى الخطوة (2) السابقة فنحصل على ($\sum \mathbf{x_i f_i}$).

4- نقسم المجموع الكلى للتوزيع التكرارى الذى حصلنا عليه فى الخطوة (3) السابقة وهو ($\sum \mathbf{x_i f_i}$) على مجموع التكرارات (n) فنحصل على الوسط الحسابى ($\bar{X}$) كما هو مبين لنا من الجدول التالى:

جدول حساب الوسط الحسابى لتوزيع درجات 100 طالب فى مادة الحاسب الآلى

الفئات	التكرارات (f)	مراكز الفئات (x)	Fx
40-	2	42.5	85
45-	8	47.5	380
50-	18	52.5	945
55-	19	57.5	1092.5
60-	27	62.5	1687.5
65-	14	67.5	945
70-	6	72.5	435
75-	3	77.5	232.5
80-	2	82.5	165
85-	1	87.5	87.5
المجموع	100		6055

$$\therefore \text{الوسط الحسابى } (\bar{\mathbf{X}}) = \frac{\sum_{i=1}^{m} f_i x_i}{n} = \frac{6055}{100} = 60.55 \text{ درجة}$$

ملحوظة هامة

للوسط الحسابى خصائص شتى أهمها المجموع الجبرى لانحرافات القيم عن وسطها الحسابى=صفر كما أنه يتمتع بالعديد من المزايا وهى:

1- سهولة الحساب

2- أنه يأخذ فى الاعتبار جميع القيم أو المفردات عند حسابه.

إلا أنه له عيوب وهى:

1- يتأثر بالقيم الشاذة أو المتطرفة.

2- لا يمكن حسابه فى حالة التوزيعات التكرارية المفتوحة من أحد طرفيها أو من الطرفين معاً.

لهذا يفضل استخدام الوسيط بدلاً من الوسط الحسابى فى حالة وجود مثل هذه القيم المتطرفة Outliers Values.

ثانيا: الوسيط Median

يعرف الوسيط بأنه القيمة (أو المفردة أو المشاهدة) التى تقسم قيم المتغير (عدد المشاهدات) إلى جزءين بحيث يكون عدد القيم التى أقل منها يساوى عدد القيم التى أكبر منها.

بمعنى آخر الوسيط هو القيمة التى يزيد عنها نصف المشاهدات ويقل عنها النصف الآخر، لهذا يتبين لنا أن قيمة الوسيط تتحدد بموضعه ولا تتحدد على أساس إضافة القيم لبعضها فضلاً عن ذلك فإن الوسيط لا يعطى قيم مضللة فى حالة وجود قيم قليلة متطرفة ، ويمكن حساب الوسيط فى حالتى البيانات غير المبوبة والبيانات المبوبة كما يتضح لنا فيما يلى:

1- حساب الوسيط فى حالة البيانات غير المبوبة

بفرض أن عدد مشاهدات المتغير هى ($x_1, x_2, x3, ... x_n$) فإنه لإيجاد أو حساب الوسيط والذى يرمز له بالرمز (M) ينبغى أولاً ترتيب البيانات (المشاهدات) تصاعدياً أو تنازلياً مع التفرقة بين عدد المشاهدات أو القيم هل هى فردى أم زوجى

<u>فإذا كان:</u> عدد القيم (المفردات) فردى فإن

رتبة الوسيط (أى موقعه) = $\frac{n+1}{2}$

وقيمة الوسيط = $M = X_{\left(\frac{n+1}{2}\right)}$

<u>أما إذا كان:</u> عدد القيم (المفردات) زوجى فإن

رتبة الوسيط (موقعه) = $\mathbf{\frac{n}{2}, \frac{n}{2}+1}$

وقيمة الوسيط = $\mathbf{M = \left(X_{\frac{n}{2}} + X_{\frac{n}{2}+1}\right)\frac{1}{2}}$

مثال 3

بفرض أن صاحب أحد محلات الأحذية أخذ عينة من أعمار عملائه سواء الإناث أو الذكور وكانت أعمارهم بالسنوات هى:

أعمار الذكور (y)					أعمار الإناث (X)							
38	70	25	35	18	31	35	50	18	42	32	40	12

المطلوب: حساب الوسيط لأعمار عملاء هذا المحل من الإناث والذكور لأخذها فى الاعتبار عند شراء طلبيات جديدة؟

الحل

1- نرتب بيانات الأعمار (قيم الأعمار) تصاعدياً أو تنازلياً لكى يمكن إيجاد الوسيط كالآتى:

أعمار الذكور (y)					أعمار الإناث (X)							
70	38	35	25	18	50	42	40	35	32	31	18	12

الوسيط= Y3 الوسيط = X4 - X5

2- تحدد موقع الوسيط (رتبة الوسيط) لكل من الذكور والإناث كالتالى:

- موقع الوسيط (ترتيب الوسيط) بالنسبة للذكور= $\frac{n+1}{2} = \frac{\mathbf{5+1}}{\mathbf{2}} = 3$

(لأن عدد القيم أو المفردات فردى حيث n=5)

∴ قيمة الوسيط بالنسبة للذكور هى القيمة Y3 = 35 سنة وهى القيمة الثالثة بعد ترتيب المشاهدات.

- أما موقع الوسيط (ترتيب الوسيط بالنسبة للإناث) = $\frac{n}{2}, \frac{n}{2}+1 = \frac{8}{2}, \frac{8}{2}+1 = 4\ ,\ 5$

(لأن عدد القيم أو المفردات زوجى حيث n = 8)

∴ قيمة الوسيط هى متوسط القيمتين الرابعة والخامسة (Y4, X5) بعد ترتيب المشاهدات تصاعدياً.

أى أن قيمة الوسيط بالنسبة للإناث = $\frac{32+35}{2} = 33.5$ سنة

2- حساب الوسيط فى حالة البيانات المبوبة (التوزيعات التكرارية)

لحساب الوسيط فى حالة التوزيعات التكرارية ينبغى إتباع الخطوات الآتية:

1- نكون الجدول التكرارى المتجمع الصاعد (أو النازل) للمفردات أو المشاهدات.

2- نحدد موقع الوسيط (ترتيب الوسيط) وذلك بقسمة مجموع التكرارات على 2 أى أن موقع الوسيط (ترتيب الوسيط) = $\frac{\mathbf{n}}{\mathbf{2}} = \frac{\sum \mathbf{f_i}}{\mathbf{z}}$

3- نحدد الفئة الوسيطة وهى الفئة التى تقع فيها الوسيط.

4- نحسب الوسيط (M) بتطبيق القانون التالى

$$\boxed{\mathbf{M = L + \frac{\left(\frac{n}{2} - C_1\right)}{C_2 - C_1} * h}} \qquad (5)$$

حيث أن:

L هى الحد الأدنى للفئة الوسيطة

h هى طول الفئة الوسيطة

C1 هى التكرار المتجمع الصاعد السابق لموقع الوسيط (السابق للفئة الوسيطة)

C2 هى التكرار المتجمع الصاعد اللاحق لموقع الوسيط (اللاحق للفئة الوسيطة)

مثال 4

احسب الوسيط للتوزيع التكرارى لدرجات 100 طالب فى امتحان مادة الحاسب الآلى المبين فى المثال (2) .

الحل

لحساب الوسيط يتبع الخطوات السالفة كالآتى:

1- نكون الجدول التكرارى المتجمع الصاعد كما يلى:

جدول لبيان حساب الوسيط لدرجات 100 طالب فى مادة الحاسب الآلى

الفئات	**التكرارات**	**الحدود العليا للفئات**	**التكرار المتجمع الصاعد**		**الفئات**	**التكرارات**	**الحدود الدنيا للفئات**	**التكرار المتجمع النازل**
40-	**2**	**أقل من-40**	**0**		**40-**	**2**	**40 فأكثر**	**100**
45-	**8**	**أقل من-45**	**2**		**45-**	**8**	**45 فأكثر**	**98**
50-	**18**	**أقل من-50**	**10**		**50-**	**18**	**50 فأكثر**	**90**
55-	**19**	**أقل من-55**	**28**		**55-**	**19**	**55 فأكثر**	**72**
60-	**27**	**أقل من-60**	**47**		**60-**	**27**	**60 فأكثر**	**53 n**
65-	**14**	**أقل من-65**	**74**		**65-**	**14**	**65 فأكثر**	**26 2**
70-	**6**	**أقل من-70**	**88**		**70-**	**6**	**70 فأكثر**	**12**
75-	**3**	**أقل من-75**	**94**		**75-**	**3**	**75 فأكثر**	**6**
80-	**2**	**أقل من-80**	**97**		**80-**	**2**	**80 فأكثر**	**3**
85-90	**1**	**أقل من-85**	**99**		**85-90**	**1**	**85 فأكثر**	**1**
		أقل من 90	**100**				**90 فأكثر**	**0**
	100					**100**		

2- نحدد موقع الوسيط (ترتيب الوسيط) = $\frac{100}{2} = 50$

3- من الجدول السابق جدول (3) يتضح من الجدول المتجمع الصاعد والجدول المتجمع النازل أن فئة الوسيط (الفئة الوسيطة) هى (60-65) وبالتالى فإن:

$L = 60 , h = 5$

ومن الجدول المتجمع الصاعد نجد أن:

$C_1 = 47 , C_2 = 74$

$$\therefore \text{ قيمة الوسيط } (M) = L + \frac{\left(\frac{n}{2} - C_1\right)}{C_2 - C_1} X\, h$$

$$= 60 + \frac{(50-47)}{74-47} X\, 5 = 60.55$$

وبالتالى فإن قيمة الوسيط $(M) = L + \frac{\left(\frac{n}{2} - C_1\right)}{C_2 - C_1} X\ h$

$$60.55 = 60 + \frac{(50-53)}{26-53} X\ 5 =$$

ثالثا: المنوال Mode

لا ريب عند استقرائنا لكلمة الوسط الحسابى أو الوسيط لمجموعة من الأرقام يتطرق إلى أذهاننا بسرعة أن هناك رقم من بين هذه الأرقام يمثلها إما متوسط الأرقام ذاتها (الوسط الحسابى) أو رقم يقع فى أوسطها (الوسيط) أما كلمة منوال قد تكون صعبة الإدراك منذ اللحظة الأولى لذا نتساءل ما معنى كلمة منوال؟

نرد ونقول أن كلمة المنوال تعنى عند استقرائها الإجابات أو الأرقام الأكثر شيوعاً ولعل خير مثال إذا سألت مجموعة من طلاب كلية معينة عن: ما هو الدكتور (المحاضر) الأكثر التزاماً فى مواعيد محاضراته والأكثر شرحاً وفهماً؟ فإن الإجابات الأكثر شيوعاً (أو عدداً) عن دكتور معين يطلق عليها المنوال.

من ثم فكلمة المنوال تعنى إجابات كثيرة من نفس الأشخاص عن سؤال معين مثل ما هى أفضل سيارة لدى المصريين أو ماهية الأرقام التى تتكرر أكثر من غيرها من الأرقام.

مثال لذلك الرقم 3 يتكرر 3 مرات فى الأرقام .. 3, 5, 7, 3, 3, 9, 11, 3

مما سلف يمكن تعريف المنوال بأنه القيمة الأكثر شيوعاً أو تكراراً من بين القيم (المشاهدات) ، وبمعنى آخر القيمة التى تتكرر أكثر من غيرها ، ويمكن حساب المنوال فى حالتى البيانات غير المبوبة والبيانات المبوبة كما هو مبين لنا فيما يلى:

1- حساب المنوال فى حالة البيانات غير المبوبة

إذا كان لدينا عدة حالات لبعض القيم أو المشاهدات (بيانات أو أرقام غير مبوبة) فأى منها المنوال؟

الحالة الأولى: 5, 6, 7, 7, 8, 9, 3

الحالة الثانية: 3, 11, 13, 4, 4, 6, 8, 9, 7, 4

الحالة الثالثة: 8, 9, 11, 15

الحالة الرابعة: 11, 10, 9, 3, 3, 6, 5, 7, 7, 15

للإجابة على السؤال السابق نرد ونقول:

- فى الحالة الأولى تعتبر القيمة 7 هى قيمة المنوال لأنها الأكثر تكراراً حيث تكررت هذه القيمة مرتين.

- أما فى الحالة الثانية فإن القيمة 4 هى قيمة المنوال لأنها الأكثر تكراراً حيث تكررت ثلاث مرات.
- بينما فى الحالة الثالثة لا يوجد منوال لأن ليس هناك قيمة تتكرر.
- وفى الحالة الرابعة هناك قيمتان للمنوال وهما القيمة 3 والقيمة 7 حيث يتكرران أكثر من مرة.

لذا يطلق على توزيع القيم فى هذه الحالة الرابعة Bimodal أى للتوزيع منوالان لأن هناك قيمتان تكررت بنفس العدد بين مجموعة كبيرة من القيم بعكس الحالة الأولى يطلق عليها Uni-modal لأن هناك قيمة واحدة للمنوال.

2- حساب المنوال فى حالة البيانات المبوبة (التوزيعات التكرارية)

الجدير بالذكر أنه فى حالة البيانات المبوبة تذوب القيم داخل الفئات لذا لا يكون لدينا قيم منوالية ولكن يكون لدينا فئة منوالية .... لذا نتساءل ماهية الفئة المنوالية؟

نرد ونقول أن: الفئة المنوالية هى "الفئة التى تناظر أكبر تكرار أو بمعنى آخر الفئة ذات الأكبر تكرار".

ويتم حساب المنوال بأكثر من طريقة (طريقة الروافع - طريقة الفروق - طريقة الرسم) الا أننا سنستخدم الطريقتين الأكثر شيوعاً وهما طريقة الفروق وطريقة الرسم وذلك كالآتى:

1- إيجاد المنوال Mode بطريقة الفروق (طريقة المنوال الخام)

تجدر الاشارة بأن هذه الطريقة يطلق عليها (طريقة الفروق) لأنها تعتمد على الفروق بين تكرار الفئة المنوالية والتكرارات السابقة واللاحقة ، وتتلخص الخطوات اللازمة لحساب المنوال بهذه الطريقة فى خطوتين هما:

1-نحدد الفئة المنوالية وهى الفئة التى تناظر أكبر تكرار.

2-نطبق القانون التالى لحساب المنوال والذى يرمز له بالرمز (D)

$$\boxed{\mathbf{D = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} * h}} \quad (7)$$

حيث أن:

L هى الحد الأدنى للفئة المنوالية (بداية الفئة المنوالية)

h هى طول فئة المنوال

Δ_1 هى الفرق بين أكبر تكرار .F والتكرار السابق له F_1

أى أن Δ_1 = تكرار الفئة المنوالية ـ التكرار السابق له

Δ_2 هى الفرق بين أكبر تكرار .F والتكرار اللاحق له F_2

أى أن Δ_2 = تكرار الفئة المنوالية ـ التكرار اللاحق له

مثال 5

احسب قيمة المنوال بطريقة الفروق فى التوزيع التكرارى لدرجات 100 طالب فى مادة الحاسب الآلى فى المثال (2)؟

الحل

بالنظر إلى التوزيع التكرارى المبين فى المثال (2) يتبين لنا أن الفئة المنوالية هى الفئة (-60) وهى الفئة التى تناظر أكبر تكرار وهو 27.

$$\mathbf{D = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} * h}$$

لذا فإنه بالنظر للتوزيع التكرارى المبين فى المثال (2) يمكن تحديد كل من:

$\Delta_1 = 27 - 19 = 8$

$\Delta_2 = 27 - 14 = 13$

$L = 60 , h = 5$

$$D = 60 + \frac{8}{8+13} * 5 = 61.904$$ درجة

ملحوظة هامة

بعد دراستنا للمنوال ومعرفة طرق حسابه ينبغى أن يضع القارئ العزيز فى ذهنه أن المنوال له مزاياه وعيوبه فى نفس الوقت كما يلى:

مزايا المنوال: من أهم مزايا المنوال الآتى:

1- أنـه المقيـاس الإحصـائى الوحيد مـن مقـاييس النزعـة المركزيـة الـذى يستخدم فـى حالـة البيانات الوصفية مثل نوع الغسالة المفضلة لربة المنزل (زانوسى - جنرال اليكترك) أو تقديرات النجاح للطلاب فى الكليـة (مقبول - جيد - جيد جدا - ممتاز) أو النوع (ذكر - أنثى) ...... الخ.

2- انـه لا يتـأثر بـالقيم المتطرفـة كالوسـط الحسـابى مـثلاً فضـلاً عـن انـه لا يتـأثر بالفئـات المفتوحة.

3- سهل الحساب والفهم ولا يحتاج إلى خلفية رياضية متقدمة.

4- أحياناً يكون المنوال أفضل المقاييس الإحصائية جميعها لأنـه أكثر منطقيـة عنها (الوسط الحسابى - الوسيط) مثال لذلك الدراسة التى بمقتضاها يتم دراسة جدوى افتتاح فرع جديد لمحـل بيـع أحذيـة بمدينـة طنطـا (Addidas مـثلاً) الغـرض منهـا التركيـز علـى أكثـر المقاسات طلباً من جمهور المستهلكين من خلال استقراء مبيعات المحلات الكائنـة فـى طنطا فى هذا المثال يعتبر المنوال بلا ريب أفضل المقاييس الإحصائية على الإطلاق لأنه يمكننـا مـن معرفـة المقاسـات الأكثـر طلبـاً مـن غيرهـا مـن خـلال استقراء أرقـام مبيعـات المحلات المختلفة.

ورغم هذه المزايا التى يتسم بها المنوال فانه يشوبه بعض العيوب أهمها:

1- له طرق كثيرة لحسابه جميعها تعتبر بمثابة طرق تقريبية غالباً لا تتساوى نتائجها.

2- أن حسابه يعتمد فقط على الفئة المنواليـة وهـى جزء مـن التوزيـع التكـرارى لـذا فهنـاك جـزء كبيـر مـن التوزيـع يـتم إهمالـه فضـلاً عـن انـه إذا كانـت الفئـة المنواليـة هـى الفئـة الأولى أو الأخيرة فى التوزيع فإن قيمة المنوال تصبح غير منطقية بل مضللة.

3- يعتبر أقل أهمية بالمقارنـة بالمقـاييس الإحصـائية الأخـرى (الوسط الحسـابى - الوسيط ...... الخ) حيث من النادر استخدامه فى المجالات العلمية.

4- أحياناً يوجد أكثر من منوال فى المفردات أو المشاهدات كذلك قد لا يوجد أطلاقا منوال لا سيما إذا تكررت كل مشاهدة من مشاهدات البيانات مرة واحدة فقط.

الخلاصة

تقييم مقارن لمقاييس النزعة المركزية (الوسط الحسابى - الوسيط - المنوال)

بعد دراستنا لكل من مقاييس المتوسطات (النزعة المركزية) الثلاثة (الوسط الحسابى - الوسيط - المنوال) وحسابها من توزيع واحد وهو التوزيع التكرارى لدرجات 100 طالب فى مادة الحاسب الآلى نتساءل أى متوسط من هذه المتوسطات أفضل عن الآخر وماهية العلاقات بينهما وهل يمكن إيجاد أحدهما من الآخر ...... الخ ...... من الأسئلة التى قد تدور فى ذهن الكثير من الدارسين والباحثين؟

نرد ونقول بالنظر إلى قيم المتوسطات الثلاثة المحسوبة لنفس التوزيع التكرارى سالف الذكر نجد أن:

الوسط الحسابى = 60.55 ، الوسيط = 60.55 ، المنوال = 61.90

أى أن قيمتى الوسط الحسابى والوسيط متساويتان أما قيمة المنوال فهى أكبر منهما إلى حد ما الا أن هذه القيم قد لا تكون كذلك فى كل الأمثلة حيث تتوقف قيمة كل منهما بناءً على شكل التوزيع التكرارى ، بمعنى أنه:

1- إذا كان التوزيع التكرارى توزيعاً متماثلاً Normal Distribution

فإن قيم المتوسطات الثلاثة تلتحم مع بعضها أى تتساوى مع بعضها بمعنى أن:

المتوسط الحسابى=الوسيط=المنوال أى ($\bar{X}$ =M=D) أنظر شكل (4-3أ)

2- إذا كان التوزيع التكرارى ملتوياً جهة اليمين فإن الوسط الحسابى يكون أكبر من المنوال.

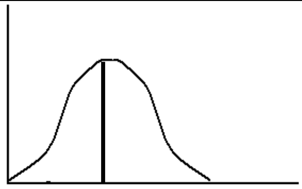
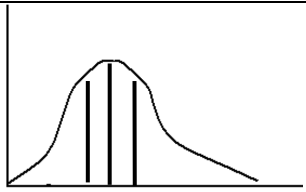
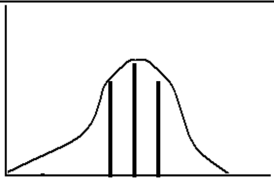
أى أن $\bar{X} > D$ أنظر شكل (4-3ب).

3- إذا كان التوزيع التكرارى ملتوياً جهة اليسار فإن الوسط الحسابى يكون أقل من المنوال.

أى أن $\bar{X} > D$ أنظر شكل (4-3جـ).

ملحوظة هامة

فى الحالتين (ب ، جـ) دائماً الوسيط يقع بين الوسط الحسابى والمنوال أى أن (M) تقع بين ($\bar{X}$, D).

منحنى متماثل	منحنى ملتوى جهة اليمين	منحنى ملتوى جهة اليسار
		
$\bar{X} = M = D$ شكل (4-3أ)	$D < M < \bar{X}$ شكل (4-3ب)	$\bar{X} < M < D$ شكل (4-3جـ)

شكل (4-3) العلاقة بين الوسط الحسابى والوسيط والمنوال $(\bar{X}, M, D)$

من الشكل السالف (4-3) يتضح أن قيمة المنوال تقابل دائماً قمة المنحى وإذا التوى المنحنى يميناً أو يساراً انتقلت قيمة المنوال هى الأخرى يميناً أو يساراً ، أما الوسط الحسابى فدائماً يأخذ مكانه فى الجانب المعاكس للمنوال . وعموماً قد لوحظ أنه كلما كان التوزيع قريباً من التماثل أو ملتوياً التواء غير شديد فإن $\bar{X} - D \approx 3(\bar{X} - M)$ ومنها نجد أن:

المنوال = ثلاثة أمثال الوسيط - ضعف الوسط الحسابى.

أما الإجابة على السؤال أى هذه المتوسطات أفضل وأجدى من الأخرى لتمثيل الظاهرة محل الدراسة؟

نرد ونقول: "أن ذلك يتوقف على طبيعة التوزيع التكرارى" بمعنى أن:

1- إذا كان التوزيع متماثل فكل المتوسطات (الوسط الحسابى - الوسيط - المنوال) لها نفس الأهمية.

2- إذا كانت التوزيع ملتوياً فيفضل الوسيط أو المنوال عن الوسط الحسابى لا سيما إذا كانت هناك قيم متطرفة.

3- إذا كان التوزيع التكرارى مفتوح من أحد طرفيه أو كليهما حيث لا يمكن حساب الوسط الحسابى لتعذر حساب مراكز الفئات المفتوحة فإن الوسيط أو المنوال يكونا أفضل من الوسط الحسابى.

4- إذا كان المنوال يقع فى بداية التوزيع أو نهايته كما هو الحال فى التوزيعات التكرارية التى يطلق علي منحنياتها بعد رسمها بالمنحنيات اللامية أو اللامية المقلوبة فإنه يفضل الوسط الحسابى أو الوسيط عن المنوال حيث لا يعبر المنوال عن التوزيع التكرارى.

خلاصة القول

أن لكل من المتوسطات الثلاثة (الوسط الحسابى - الوسيط - المنوال) مزاياه وعيوبه لذا إذا كان هناك شك أو ريبة فى مدى الاعتماد على أحدهما دون الآخر فيفضل حساب اثنين منها أو الثلاثة معاً ثم محاولة اختبار أحدهما لتمثيل الظاهرة محل الدراسة لا سيما بعد معرفة شكل التوزيع التكرارى لهذه الظاهرة من رسم المنحنى التكرارى لها.

الفصل الرابع
مقاييس التشتت

■ مقدمة

بعد دراستنا لمقاييس النزعة المركزية (المتوسطات) وعرفنا أن هذه المقاييس تحاول تلخيص البيانات (القيم) فى رقم واحد يطلق عليها المتوسط (الوسط الحسابى - الوسيط - المنوال .... الخ) كمقياس يصف هذه البيانات حيث تميل هذه البيانات (القيم) نحو قيمة معينة تسمى المتوسط . قد يثار سؤال وهو هل المتوسط الذى حصلنا عليه لمجموعة من البيانات (القيم) كافى لوصف هذه البيانات؟

الإجابة بالطبع لا لأن هذا المتوسط غير كافى لتحديد صفات وشكل التوزيع الذى تتبعه هذه البيانات (القيم) لأنه يبين فقط مدى تمركز البيانات حوله ولكن لا نعرف منه هل هذه البيانات قريبة من بعضها البعض أم بعيدة عن بعضها البعض حيث هناك بيانات قريبة من هذا المتوسط وأخرى بعيده عنه ...... لذا فلا بد أن يكون هناك مقياس آخر يسانده يبين مدى التفاف البيانات أو بعثرتها حول هذا المتوسط.

ولكى يتضح لك أيها القارئ العزيز بأن المتوسط لمجموعة من البيانات غير كافى لوصف هذه البيانات تعال معنا نوضح لك هذا بالمثال التوضيحى التالى:

مثال توضيحى لبيان متوسط مجموعة من البيانات غير كافى لوصفها

بفرض أن لدينا عينتان من طلاب كلية التجارة - جامعة كفر الشيخ كل عينة مكونه من عشرة طلاب وسألنا كل واحد منهم عن الدرجة التى حصل عليها فى مادة المحاسبة وكانت إجاباتهم كما هى مبينه فى الجدول التالى:

جدول بيان درجات عينتين من طلاب التجارة فى مادة المحاسبة

البيان	العينة الأولى من الطلاب					البيان	العينة الثانية من الطلاب				
الدرجات	65	73	71	75	77	الدرجات	42	58	67	62	54
	77	77	67	70	68		77	98	85	100	77
الوسط الحسابى الوسيط المنوال	$\bar{x}=72$ M=72 D=77					الوسط الحسابى الوسيط المنوال	$\bar{x}=72$ M=72 D=77				

وباستقرائنا الجدول (6) السالف يتضح لنا ما يلى:

1- أنه من النظرة الأولى أن درجات مجموعتى الطلاب متماثلة لأن قيم كل من المتوسط الحسابى ($\bar{x}=72$) والوسيط ($M = 72$) والمنوال ($D=77$) متساوية للمجموعتين.

2- إذا نظرنا مرة أخرى نظرة ثاقبة لدرجات مجموعتى الطلاب يتضح لنا أن هناك فرق بين درجات المجموعتين حيث تنتشر درجات المجموعة الثانية من الطلاب فى مدى أوسع أى من (100-42) بينما درجات المجموعة الأولى تقترب من بعضها البعض حيث المدى الذى تنتشر فيه ليس كبير (77-65) ، ومن ثم نستنتج أن درجات مجموعة الطلاب الثانية أكثر بعداً من بعضها البعض أى أكثر تشتتاً.

ومن ثم فإن المتوسطات (الوسط الحسابى - الوسيط - المنوال) أو بمعنى آخر مقاييس النزعة المركزية لا تظهر هذا التشتت أو التغير بين البيانات (المشاهدات أو المفردات) وبالتالى لا تكفى وحدها لوصف مجموعة من البيانات (القيم) بل ينبغى أن يصاحبها مقياس آخر من مقاييس التشتت Measures of Dispersion.

ونستخلص مما سلف أنه لا بد عند دراسة توزيع مجموعة من البيانات (القيم) لأى ظاهرة من الظواهر الا تقتصر دراستنا فى وصف هذه الظاهرة على أحد المتوسطات فحسب بل ينبغى أيضاً قياس التشتت ، فضلاً عن قياس كل من الالتواء Skewness والتفرطح Kurtosis وبهذا يصبح الوصف للظاهرة وصفاً تاماً لا يشوبه شئ .

وأخيراً نتساءل ما هو التشتت؟ وماهية أهم مقاييس التشتت؟

نرد ونقول: التشتت هو: "انحراف القيم (البيانات) أو بعدها عن بعضها البعض"

أما مقاييس التشتت فهى تنحصر فيما يلى:

- المدى The Range
- الانحراف الربيعى Quartile Deviation
- الانحراف المتوسط The Average Deviation
- الانحراف المعيارى والتباين The Standard Deviation and Variance
- معامل الاختلاف The Coefficient of Variations

وسوف نتناول مقاييس التشتت كما هو مبين لنا فيما يلى:

اولا: المدى The Range

يعتبر المدى من أبسط مقاييس التشتت وأسهلها فى الحساب حيث يرمز له بالرمز (R) ويمكن حسابه فى حالتى البيانات غير المبوبة والبيانات المبوبة كما يتبين لنا فيما يلى:

1- المدى فى حالة البيانات غير المبوبة

يعرف المدى بأنه الفرق بين أكبر قيمة (مشاهدة أو مفردة) وأصغر قيمة أى أن:

المدى = أكبر مشاهدة - أصغر مشاهدة

Range=The Largest Observation - The Smallest Observation (12)

على سبيل المثال نجد أن المدى (R) لدرجات مجموعتى الطلاب هو:

المدى (R) لدرجات مجموعة الطلاب الأولى:

$R_1 = 77 - 65 = 12$ درجة

المدى (R) لدرجات مجموعة الطلاب الثانية:

$R_2 = 100 - 42 = 58$ درجة

وهنالك يتضح أن $R_2>R_1$ وهذا يعنى أن مشاهدات المجموعة الثانية (درجات الطلاب) أكثر تشتتاً من مثيلاتها فى المجموعة الأولى بمعنى آخر أن درجات المجموعة الأولى (مشاهدات) أكثر تجانساً من مثيلاتها للمجموعة الثانية.

2- المدى فى حالة البيانات المبوبة (التوزيعات التكرارية)

يعرف المدى بأنه الفرق بين الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة الأولى أى أن:

المدى = الحد الأعلى للفئة الأخيرة - الحد الأدنى للفئة الأولى

على سبيل المثال: باستخدام بيانات التوزيع التكرارى لدرجات 100 طالب فى امتحان مادة الحساب الآلى (أنظر مثال 2) يمكن حساب المدى كالآتى:

المدى = $R = 40 - 90 = 50$ درجة

مزايا المدى: للمدى عدة مزايا يتسم بها وهى:

1- أنه مقياس سهل وبسيط فى الحساب.

2- يعتبر مقياس جيد للتشتت فقط إذا كانت عدد المشاهدات صغير لهذا يستخدم كثيراً لاسيما فى التطبيقات العملية التى تتسم بصغر عدد مشاهداتها.

مثال لذلك: عمليات الرقابة الإحصائية على جودة الإنتاج والذى فيها يحدد مدى القبول لمواصفات الإنتاج من خلال فحص عينات لا تربوا عن أربعة أو خمسة مشاهدات ، كذلك يستخدم المدى فى التطبيقات الطبية مثال لذلك المدى المقبول لكيفية أداء الجسم لوظائفه المختلفة مثل عدد نبضات القلب أو درجة الحرارة أو ضغط الدم والذى يعتبر الشخص مريضاً إذا تعدى هذا المدى.

فى ذات الوقت هناك بعض العيوب التى تشوب المدى تتبلور فيما يلى:

- أنه يهتم فقط بالمشاهدتين الأولى والأخيرة دون بقية المشاهدات الأخرى لهذا يتأثر بالقيم الشاذة (المشاهدات الصغير جداً أو الكبير جداً).
- لا يجوز استخدامه لمقارنة تشتت مجموعتين من المشاهدات (القيم) الا إذا كانت عدد المشاهدات فى كل مجموعة متساوى مع المجموعة الأخرى لأنه يزداد بزيادة عدد المشاهدات.
- تختلف قيمته من عينة لأخرى داخل نفس المجتمع بعكس المقاييس الأخرى.
- يصعب تقدير قيمته من الجداول التكرارية المفتوحة.

ثانيا: الانحراف الربيعى أو نصف المدى الربيعى Semi-Interquartile Range

لما كان من عيوب المدى أنه يتأثر بالقيم الشاذة أو المتطرفة لذا فإن أفضل طريقة للتخلص من هذا العيب هى حذف أعلى جزء من المشاهدات (الربع الأخير) وأصغر جزء من المشاهدات (الربع الأول) بعد ترتيبها تصاعدياً أو تنازلياً ثم نحسب الفرق بين قيمة الربيع الأول والربيع الثالث فنحصل على ما يسمى بالمدى الربيعى ـ لذا نتساءل ما هو نصف المدى الربيعى؟

نرد ونقول "هو متوسط الفرق بين الربيع الثالث Q3 والربيع الأول Q1"

وبالتالى فإن نصف المدى الربيعى والذى يرمز له بالرمز Q هو:

$$\boxed{Q = \frac{Q_3 - Q_1}{2}} \qquad (13)$$

ويمكن حساب نصف المدى الربيعى فى حالتى البيانات غير المبوبة والبيانات المبوبة كما يتبين لنا فيما يلى:

1- حساب نصف المدى الربيعى فى حالة البيانات غير المبوبة

مثال 1

فيما يلى درجات عينة قدرها 15 طالب فى اختبار مقرر مبادئ الإحصاء:

7.25	8.5	9.5	10	10.25	11	12.5	12.75
13.25	14	15	16.5	16.75	18	19.5	

المطلوب: حساب كل من:

أ-المدى.

ب-نصف المدى الربيعى كمقياس للتشتت لدرجات هؤلاء الطلاب.

الحل

نرتب الدرجات إذا لم تكن مرتبة ترتيباً تصاعدياً

أ- المدى (R) = 19.5 - 7.25 = 12.25

ب- لكى نحسب نصف المدى الربيعى فلا بد أولاً من تحديد رتبة الربيع الأدنى والربيع الأعلى كالتالى:

ترتيب أو موقع الربيع الأدنى = $\mathbf{\frac{1(15+1)}{4} = 4}$ (الرقم الرابع)

قيمة الربيع الأدنى (Q_1)= 10 (الرقم الرابع)

ترتيب أو موقع الربيع الأعلى = $\mathbf{\frac{3(15+1)}{4} = 12}$ (الرقم الثانى عشر)

قيمة الربيع الأدنى (Q3)= 16.5 (الرقم الثانى عشر)

نصف المدى الربيعى (Q) = $\mathbf{\frac{Q_3 - Q_1}{2}}$

$$= \mathbf{\frac{16.5-10}{2} = 3.25}$$

بمقارنة المدى بنصف المدى الربيعى نلاحظ أن هناك فرق شاسع بينهما فى قياس التشتت كما أن نصف المدى الربيعى ما زال يعتمد عند حسابه على مفردتين فقط كما هو الحال فى المدى رغم أنه تم حذف بعض المفردات أو المشاهدات (الربع الأول والربع الأخير) من المشاهدات.

2- حساب نصف المدى الربيعى فى حالة البيانات المبوبة (التوزيعات التكرارية)

مثال 2

بفرض أنه تم اختيار عينة قدرها 72 ورقة اجابة اختبار مقرر اللغة الانجليزية وكان توزيع درجاتهم كما هو مبين فى التوزيع التكرارى التالى:

فئات الدرجات	-4	-5	-6	-7	-8	9-10	المجموع
عدد الطلاب	5	8	14	20	17	8	72

المطلوب: حساب نصف المدى الربيعى كمقياس للتشتت لدرجات هؤلاء الطلاب؟

الحل

لكى يمكن حساب نصف المدى الربيعى لا بد من حساب كل من الربيع الأول (Q_1) والربيع الثالث (Q_3) وذلك بتكوين الجدول التكرارى المتجمع الصاعد التالى:

جدول لبيان الجدول التكرارى المتجمع الصاعد لحساب الربيع الأدنى والربيع الأعلى

الفئات	التكرارات	الحدود العليا للفئات	التكرار المتجمع الصاعد
4-	5	4 أقل من	0
5-	8	5 أقل من	5
6-	14	6 أقل من	13 n
7-	20	7 أقل من	27 4
8-	17	8 أقل من	47 3n
9-10	8	9 أقل من	64 4
		10 أقل من	72

موقع أو ترتيب الربيع الأول Q_1 (الأدنى) $= \frac{in}{4} = \frac{72}{4} = 18$

قيمة الربيع الأول

$$Q_1 = L + \frac{\left(\frac{n}{4} - C_1\right)}{C_2 - C_1} x h$$

$$Q_1 = 6 + \frac{(18-13)}{(27-13)} x 1 = 6.357$$

موقع الربيع الثالث (Q3) $= \frac{3n}{4} = \frac{3x72}{4} = 54$

قيمة الربيع الثالث

$$Q_3 = L + \frac{\left(\frac{3n}{4} - C_1\right)}{C_2 - C_1} x h$$

$$Q_1 = 8 + \frac{(54-47)}{(64-47)} x1 = 8.412$$

وبالتالى نجد أن نصف المدى الربيعى (A) كما يلى:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{8.412 - 6.357}{2} = 1.0275$$

مما سلف نستخلص أنه رغم أن نصف المدى الربيعى يخلصنا من بعض العيوب التى تشوب المدى حيث لا يتأثر بالقيم المتطرفة فضلاً عن إمكانية حسابه فى حالة التوزيعات التكرارية المفتوحة الا أنه ما زال يشوبه بعض العيوب من أهمها:

- أنه فى حالة البيانات غير المبوبة يعتمد على مفردتين فقط لاحتساب قيمته لذا فهو مقياس غير حساس للتشتت.
- تقتضى عملية حسابه جهد كبير على وجه الخصوص فى حالة البيانات المبوبة حيث لا بد من حساب (Q_1, Q_3) بدقة كافية حيث يعتمد عليها فى حسابه.

ثالثا: الانحراف المتوسط The Average Deviation

بعد دراستنا لمقياسين من مقاييس التشتت (المدى - نصف المدى الربيعى) تبين لنا أن المدى يعتمد فى حسابه على قيمتين فقط من قيم المشاهدات (المفردات) وهما أكبر مشاهده وأصغر مشاهده فى حين نصف المدى الربيعى (الانحراف الربيعى) يعتمد هو الآخر على كل من قيمة الربيع الأول والربيع الثالث للمشاهدات ...... دون بقية المشاهدات . لذا أضحى من الأهمية وجود مقاييس أخرى للتشتت تعتمد فى حسابها على جميع المشاهدات أو القيم من أهم هذه المقاييس الانحراف المتوسط والذى يتبلور تعريفه فى:

الانحراف المتوسط هو

"متوسط انحرافات القيم عن أحد المتوسطات (الوسط الحسابى - الوسيط)".

بمعنى آخر هو

"متوسط مجموع انحرافات القيم عن (الوسط الحسابى - الوسيط) بغض النظر عن الاشارة".

تجدر الاشارة بأن الانحراف المتوسط يأخذ فى الحسبان حجم الانحراف فقط عن القيم بغض النظر عن إشارتها ، ويعزى السبب فى إهمال الاشارة إلى أن مجموع الانحرافات كما علمنا مسبقاً يساوى صفر.

لهذا إذا لم تهمل الاشارة فإن قيمة الانحراف المتوسط ستساوى صفر باستمرار ، وبالتالى لا يمكن اتخاذ هذه الانحرافات كمقياس للتشتت.

والسؤال الذى يدور فى الذهن الآن هو: هل هذه الانحرافات هى انحرافات القيم (المشاهدات) عن الوسط الحسابى أم الوسيط؟

نرد ونقول ليست ثمة مشكلة إذا حسبت انحرافات القيم عن الوسط الحسابى أو الوسيط الا أن استخدام الوسط الحسابى أكثر شيوعاً من الوسيط ، وفى ضوء ذلك يمكن حساب الانحراف المتوسط فى حالتى البيانات غير المبوبة والبيانات المبوبة كما يتبين لنا فيما يلى؟

1-حساب الانحراف المتوسط فى حالة البيانات غير المبوبة

لما كان من خصائص الوسط الحسابى أن مجموع انحرافات القيم عن وسطها الحسابى = صفر

أى أن $\sum(x-\bar{x})=0$ نتيجة لوجود بعض الانحرافات الموجبة والبعض الآخر سالبة لذا فلا بد من إيجاد طريقة تمنع الانحرافات الموجبة والانحرافات السالبة من أن تحذف بعضها البعض وبمعنى آخر

إيجاد طريقة بمقتضاها يتم حساب الانحراف بغض النظر عن إشارته ويحدث ذلك بإهمال الاشارة السالبة أى إيجاد القيمة المطلقة للانحرافات التى تعرف بالمعادلة التالية:

$$|X| = \begin{cases} X & X \geq 0 \\ X & X < 0 \end{cases}$$

وبفرض أن البيانات (المشاهدات) المطلوب حساب الانحراف المتوسط لها هى X_1, X_2, X_3 .... X_n ومتوسط هذه البيانات (المشاهدات) هو ($\bar{X}$) فإن قيمة الانحراف المتوسط الذى يرمز له بالرمز (AD) يمكن حسابه من المعادلة التالية:

$$\boxed{AD = \frac{\sum \left| X - \bar{X} \right|}{n}} \qquad (14)$$

حيث أن: $\left| X - \bar{X} \right|$ هى انحرافات البيانات (القيم) عن الوسط الحسابى ($\bar{X}$)

لهذا ينبغى حساب الوسط الحسابى أولاً ثم إيجاد انحرافات القيم عن الوسط الحسابى ثم تطبيق القانون السابق.

مثال 3

فيما يلى درجات ثمانية طلاب فى مادة رياضيات الأعمال :

56, 83, 67, 49, 80, 60, 90, 75

المطلوب إيجاد الانحراف المتوسط لهذه الدرجات

الحل

$$\bar{X} = \frac{56 + 83 + 67 + 49 + 80 + 60 + 90 + 75}{8} = 70$$

ولكى يمكن إيجاد الانحرافات والقيم المطلقة يجب تكوين الجدول التالى:

جدول لايجاد الانحراف المتوسط لدرجات 8 طلاب فى مادة رياضيات الأعمال

X	56	83	67	49	80	60	90	75	المجموع
$X-\bar{X}$	-14	+13	-3	-21	+10	-10	+20	+5	0
$\left\|x-\bar{x}\right\|$	14	13	3	21	10	10	20	5	96

$$\because\ AD = \frac{\sum \left|(X-\bar{X})\right|}{n}$$

$$\therefore\ AD = \frac{96}{8} = 12$$

ب-حساب الانحراف المتوسط فى حالة البيانات المبوبة (التوزيعات التكرارية)

لكى يمكن حساب الانحراف المتوسط فى حالة البيانات المبوبة (جداول التوزيع التكرارى) ينبغى إتباع الخطوات التالية:

1- تحديد مراكز الفئات.

2- إيجاد الوسط الحسابى ($\bar{X}$) بإحدى الطرق المعروفة سلفاً.

3- إيجاد الانحرافات عن الوسط الحسابى ($X-\bar{X}$).

4- إيجاد الانحرافات المطلقة $\left|X-\bar{X}\right|$.

5- نضرب الانحرافات المطلقة فى التكرار المناظر ثم نوجد مجموع حاصل الضرب.

6- نطبق المعادلة التالية لحساب الانحراف المتوسط.

$$AD = \frac{\sum_{i=1}^{i=m} F_i \left|(X-\bar{X})\right|}{\sum_{i=1}^{m} F}$$

حيث أن:

F_i هى التكرارات

$\sum_{i-1}^{i=m} F$ هى مجموع التكرارات

مثال 15

باستخدام التوزيع التكرارى لدرجات 100 طالب فى مادة الحاسب الآلى المبين فى المثال (2) احسب الانحراف المتوسط مع تفسير معناه؟

الحل

باستخدام الخطوات سالفة الذكر يمكن حساب الانحراف المتوسط وذلك بعد تكوين الجدول التالى:

الفئات	التكرارات (F)	مراكز الفئات (x)	$(x-\bar{x})$	$\lvert x-\bar{x}\rvert$	$F\lvert x-\bar{x}\rvert$
40-	2	42.5	-18.05	18.05	36.10
45-	8	47.5	-13.05	13.05	104.40
50-	18	52.5	-8.05	8.05	144.90
55-	19	57.5	-3.05	3.05	57.95
60-	27	62.5	1.95	1.95	52.65
65-	14	67.5	6.95	6.95	97.30
70-	6	72.5	11.95	11.95	71.7
75-	3	77.5	16.95	16.95	50.85
80-	2	82.5	21.95	21.95	43.90
85-90	1	87.5	26.95	26.95	26.95
المجموع	100				686.7

ويوضح الجدول (8) تطبيق الخطوات السابقة مع ملاحظة أن الوسط الحسابى ($\bar{X}=60.55$) قد تم حسابه فى مثال (3) ، وبتطبيق القانون التالى يمكن حساب الانحراف المتوسط:

$$AD = \frac{\sum_{i=1}^{i=m} F_i \left| (X - \bar{X}) \right|}{\sum_{i=1}^{m} F}$$

$$AD = \frac{686.7}{100} = 6.867 \text{ درجة}$$

وهـذا يعنـى أن الانحـراف المتوسـط عـن الوسـط الحسـابى لـدرجات الطـلاب ($\bar{X}=60.55$) يسـاوى 6.567 درجـة بمعنـى آخـر أن متوسـط البعـد (التشـتت) بـين درجـات الطـلاب والمتوسط الحسابى يساوى 6.867 درجة.

ومن المثال السابق يمكننا استنتاج أهم مزايا الانحراف المتوسط وهى:

1- أنه مقياس للتشتت يتم حسابه بمقتضى القيم (المشاهدات) الخاصة بالظاهرة موضوع الدراسة.

2- سهولة حسابه وتفسيره على البيانات.

ورغم مزايا الانحراف المتوسط الا أنه يعانى من بعض القصور أهمها:

1- إهمال الاشارة للانحرافات دون الاستناد على منطق علمى رياضى يبرز ذلك.

2- لا يمكن حسابه فى حالة التوزيعات التكرارية المفتوحة.

وخلاصة القول أن الانحراف المتوسط كمقياس للتشتت قليل الاستخدام فقلما يستخدم فـى عمليات التحليل الإحصائى لأنـه أقل دقـة بالمقارنـة مع بعض المقـاييس الأخرى للتشتت والتـى أهمهـا الانحـراف المعيارى كما سيتضح لنا فيما بعد.

رابعا: الانحراف المعيارى والتباين The Standard Deviation and Variance

ذكرنا سالفاً أن الانحراف المتوسط كأحد مقاييس التشتت هو متوسط مجموع انحرافات القيم (المشاهدات) عن وسطها الحسابى بغض النظر عن الإشارات السالبة للانحرافات الا أن إهمال إشارة هذه الانحرافات لا يستند إلى مبرر منطقى رياضى لذا فكان لا بد من وجود مقياس آخر للتشتت يتخلص من الإشارات السالبة للانحرافات (الفروق بين القيم والمتوسط الحسابى) بطريقة رياضية هذا المقياس يطلق عليه التباين Variance .. لذا نتساءل ما هو التباين؟

نرد ونقول أن التباين يعتبر من أهم مقاييس التشتت وأكثرها انتشاراً بمقتضاه يتم التخلص من الإشارات السالبة فى الانحرافات عن طريق تربيع جميع الانحرافات ثم إيجاد وسطها الحسابى.

بمعنى آخر التباين هو "متوسط مجموع مربع انحرافات القيم عن وسطها الحسابى".

ومما سلف يتبين أن التباين يعتمد فى حسابه على كل القيم (المشاهدات) وبالتحديد على انحرافات القيم (المشاهدات) عن وسطها الحسابى ، فاذا كانت انحرافات القيم عن وسطها الحسابى هى:

$$(X_1-\bar{X}),\ (X_2-\bar{X}),\ (X_3-\bar{X})\ ....\ (X_n-\bar{X})$$

ومربعات هذه الانحرافات عن الوسط الحسابى هو:

$$(X_1-\bar{X})^2,\ (X_2-\bar{X})^2,\ (X_3-\bar{X})^2\ ....\ (X_n-\bar{X})^2$$

فإن متوسط مجموع مربعات انحرافات القيم عن وسطها الحسابى يسمى التباين Variance ويرمز له بالرمز (σ^2) أى أن:

$$\boldsymbol{\sigma^2 = \frac{1}{n}\left\{(X_1-\bar{X})^2 + (X_2-\bar{X})^2 .... + (X_n-\bar{X})^2\right\}}$$

$$\boxed{\sigma^2 = \frac{1}{n}\sum(X_i-\bar{X})^2} \quad (16)$$

حيث أن:

n هى عدد المشاهدات

X_i هى المشاهدات (القيم)

$\bar{X}$ الوسط الحسابى

وبإيجاد الجذر التربيع للتباين نحصل على ما يسمى الانحراف المعيارى ويرمز له بالرمز (σ) أى أن

$$\boxed{\sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}}} \qquad (17)$$

ويعتبر الانحراف المعيارى أهم مقاييس التشتت بل أكثرها استخداماً وهذا يعزو لسهولة حسابه ، فضلاً عن ذلك فهو يتميز عن التباين بأنه يكون من نفس نوع الوحدات الموجودة فى مجموعة البيانات نفسها.

ويمكن حساب التباين والانحراف المعيارى فى حالة البيانات المبوبة والبيانات غير المبوبة كما يتبين لنا فيما يلى.

1- حساب التباين والانحراف المعيارى فى حالة البيانات غير المبوبة

باستخدام القانون الأساسى للتباين

بفرض أن المتغير (X) يأخذ القيم $X_1, X_2, X_3 \ldots\ldots\ldots X_n$ وأن الوسط الحسابى لهذا المتغير هو ($\bar{X}$) فانه وفقاً لتعريف التباين السالف الذكر يمكن إيجاد قيمته من المعادلة التالية:

$$\sigma^2 = \frac{1}{n}\sum (X - \bar{X})^2$$

مثال4

بفرض أنه تم اختيار عينة عشوائية مكونة من عشرة طلاب من طلاب الفرقة الأولى بكلية التجارة - جامعة كفر الشيخ وسؤالهم عن درجاتهم فى الامتحان العملى لمادة الإحصاء (الدرجة من 20) فكانت إجاباتهم كالتالى:

15, 13, 9, 10, 12, 14, 16, 8, 13, 20

المطلوب: حساب كل من التباين والانحراف المعيارى؟

الحل

لايجاد التباين والانحراف المعيارى ينبغى إتباع الخطوات الآتية:

1- إيجاد الوسط الحسابى للدرجات.

2- إيجاد انحرافات القيم (الدرجات) عن وسطها الحسابى أى$(X - \bar{X})$

3- تربيع كل انحراف من الانحرافات السابقة ثم إيجاد مجموعها أى $\sum(X-\bar{X})^2$

4- نطبق القانون $\sigma^2 = \frac{1}{n}\sum(X-\bar{X})^2$ لايجاد التباين.

5- نوجد الجذر التربيع للتباين فنحصل على الانحراف المعيارى كما يلى:

الوسط الحسابى $(\bar{X}) = \frac{15+13+9+10+12+14+16+8+13+20}{10} = \frac{130}{10} = 13$

لايجاد انحرافات القيم عن وسطها الحسابى وتربيعها ومجموع تربيعها نكون الجدول التالى:

جدول (9) جدول لبيان إيجاد التباين لدرجات 10 طلاب فى مادة الإحصاء بتطبيق القانون الأساسى للتباين

X	15	13	9	10	12	14	16	8	13	20	المجموع 130
$X-\bar{X}$	+2	0	-4	-3	-1	+1	+3	-5	0	+7	المجموع 0
$(X-\bar{X})^2$	4	0	16	9	1	1	9	25	0	49	المجموع 114

وبتطبيق القانون نجد أن التباين:

$$\sigma^2 = \frac{1}{n}\sum(X-\bar{X})^2$$

$$\sigma^2 = \frac{114}{10} = 11.4$$

الانحراف المعيارى هو:

$$\sigma = \sqrt{11.4} = 3.376$$

مثال 5

باستخدام التوزيع التكرارى لدرجات 100 طالب فى مادة الحاسب الآلى المبين فى المثال (2) احسب كل من التباين والانحراف المعيارى؟

الحل

لايجاد التباين والانحراف المعيارى للتوزيع التكرارى ينبغى إتباع الخطوات الآتية:

1- تحديد مراكز الفئات (X).

2- نوجد الوسط الحسابى $(\bar{X})$

3- نوجد الانحرافات عن الوسط الحسابى أى $(X_i - \bar{X})$ ثم نربع هذه الانحرافات أى $(X_i - \bar{X})^2$.

4- نضرب مربعات الانحرافات فى التكرارات المناظرة أى $F(X_i - \bar{X})^2$ ثم نوجد المجموع.

5- نطبق القانون السابق.

جدول حساب التباين والانحراف المعيارى

الفئات	التكرارات (F)	مراكز الفئات (x)	$(x_i - \bar{x})$	$(x - \bar{x})^2$	$F(x - \bar{x})^2$
40-	2	42.5	-18.05	325.80	651.6
45-	8	47.5	-13.05	170.30	1362.4
50-	18	52.5	-8.05	64.80	1166.4
55-	19	57.5	-3.05	9.30	176.7
60-	27	62.5	1.45	3.80	102.6
65-	14	67.5	6.95	48.30	676.2
70-	6	72.5	11.95	142.80	856.8
75-	3	77.5	16.95	287.30	861.9
80-	2	82.5	21.95	481.80	963.6
85-90	1	87.5	26.95	726.30	726.30
المجموع	100				7544.5

وحيث أن الوسط الحسابى ($\bar{X}$) = 60.55 سلف حسابه كما هو مبين بمثال (2) لذا يمكن إيجاد التباين (σ^2) كالآتى:

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

$$\sigma^2 = \frac{1}{100}(7544.5) = 75.445$$

الانحراف المعيارى هو

$$\sigma = \sqrt{\sigma^2} = \sqrt{75.445} = 8.686 \quad \text{درجة}$$

ملحوظة هامة

2- لا يمكن حساب الانحراف المعيارى فى حالة التوزيعات التكرارية ذات الفئات المفتوحة لصعوبة إيجاد مراكز الفئات.

3- عند حساب التباين أو الانحراف المعيارى لعينة من البيانات (المشاهدات) ينبغى استبدال (n - 1) بدلاً من (n) فى كل القوانين سالفة الذكر سواء فى حالة البيانات المبوبة أو غير المبوبة.

الخلاصة

أن الانحراف المعيارى يعتبر أكثر مقاييس التشتت استخداماً لما يتسم به من خصائص أهمها أن حسابه يعتمد على كافة القيم (المشاهدات) لهذا نرى من الأفضل للباحثين والدارسين استخدامه إذا كان لديهم تحليلات إحصائية أخرى تعتمد عليه.

■ مقاييس التشتت النسبية

إذا كان لدينا بيانات لتوزيعين تكراريين لظاهرتين مختلفتين أو أكثر أو إذا كان لدينا بيانات لتوزيعين تكراريين لظاهرة واحدة على فترتين من الزمن أو أكثر نتساءل هل يمكن مقارنة التشتت لبيانات هذين التوزيعين التكراريين بأحد مقاييس التشتت السالف ذكرها؟

نرد ونقول بالطبع لا لأن هناك عقبات ينبغى التخلص منها وهى:

أ- اختلاف الأوساط الحسابية لمجموعات البيانات للظاهرتين المختلفتين (أو للظاهرة الواحدة فى فترتين زمنيتين مختلفتين).

ب- اختلاف وحدات القياس لمجموعات البيانات المختلفة بمعنى أن البيانات قد تكون مقيسة بوحدات مختلفة للمجموعات المختلفة كأن تكون المجموعة الأولى مقيسة بالسنتيمتر والمجموعة الثانية مقيسة بالكيلوجرام والمجموعة الثالثة مقيسة بالسنة ...... وهكذا.

مثال6

عند مقارنة تشتت أطوال مجموعة من الطلاب يرغبون فى الالتحاق بكليات التجارة بتشتت أوزانهم ...... فهاتين مجموعتين من القيم (البيانات) الأولى مقيسة بالسنتيمتر والثانية مقيسة بالكيلوجرام ...... لهذا لا يمكن أجراء المقارنة بينهما باستخدام مقاييس التشتت السالف ذكرها ...... وحتى إذا كانت وحدات القياس واحدة ولكن كان هناك اختلاف بين المتوسط الحسابى للمجموعة الأولى ومثيلة بالنسبة للمجموعة الثانية أو كان هناك اختلاف فى حجم البيانات فلا يمكننا استخدام مقاييس التشتت السابق دراستها فى المقارنة.

لذا يجب استخدام معامل الاختلاف Coefficient of variation ($C.V$) للمقارنة بينهم والتى يكون معامل اختلافها أكبر تكون الأكثر تشتتا ويحسب كالتالى.

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

مثال 7

بفرض أن هناك مجموعه من الاشخاص عددها 130 شخص ، تم حصر اوزانهم واطوالهم فى فى التوزيعين التكراريين التاليين:-

التوزيع التكرارى للأوزان (X)

فئات الأوزان	50-	55-	60-	65-	70-	75-	80-	85-90	المجموع
عدد الطلاب	6	14	16	30	24	20	15	5	130

التوزيع التكرارى للأطوال (X)

فئات الأطوال	150-	155-	160-	165-	170-	175-	180-	185-190	المجموع
عدد الطلاب	7	13	18	28	26	18	12	8	130

المطلوب: حدد اى التوزيعين أكثر تشتتا الاوزان أم الاطوال.

الحل

$$\because \quad C.V = \frac{s}{\bar{x}} \times 100$$

لهذا يجب حسب كل من الانحراف المعيارى والوسط الحسابى لمجموعتى الأوزان والأطوال بعد تكوين الجدولين الآتيين:

جدول حساب تشتت الاوزان

الفئات	التكرارات (F)	مراكز الفئات (X)	FX	FX^2
50-	6	52.5	315	16537.5
55-	14	57.5	805	46287.5
60-	16	62.5	1000	62500.0
65-	30	67.5	2025	136687.5
70-	24	72.5	1740	126150.0
75-	20	77.5	1550	120125.0
80-	15	82.5	1237.5	102093.75
85-90	5	87.5	437.5	38281.25
المجموع	130		9110	648662.5

$$\bar{X} = \frac{\sum F_i X_i}{n} = \frac{9110}{130} = 70.077$$ كيلوجرام

$$\sigma^2 = \frac{1}{n}\left[\sum FX^2 - \frac{\sum (FX)^2}{n}\right]$$

$$\sigma^2 = \frac{1}{130}\left[648662.5 - \frac{(9110)^2}{130}\right] = 78.936$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{78.936} = 8.885$$ كيلوجرام

معامل الاختلاف (CV):

$$\mathbf{C.V = \frac{\sigma}{\bar{X}} = \frac{8.885}{70.077} x100 = 12.678\%}$$

جدول حساب تشتت الاطوال

الفئات	التكرارات (F)	مراكز الفئات (X)	FX	FX^2
150-	7	152.5	1067.5	162793.8
155-	13	157.5	2047.5	322481.3
160-	18	162.5	2925	475312.5
165-	28	167.5	4690	785575
170-	26	172.5	4485	773662.5
175-	18	177.5	3195	567112.5
180-	12	182.5	2190	399675
185-190	8	187.5	1500	281250
المجموع	130		22100	3767863

$$\overline{X} = \frac{\sum F_i X_i}{n} = \frac{2210}{130} = 170 \quad \text{سنتيمتر}$$

$$\sigma^2 = \frac{1}{n}\left[\sum FX^2 - \frac{\sum (FX)^2}{n}\right]$$

$$\sigma^2 = \frac{1}{130}\left[3767863 - \frac{(22100)^2}{130}\right] = 83.557$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{83.557} = 9.140 \quad \text{سنتيمتر}$$

معامل الاختلاف (CV):

$$C.V = \frac{\sigma}{\overline{X}} = \frac{9.140}{170} x100 = 5.377\%$$

وبمقارنة درجة التشتت المطلق بين الطول والوزن فيمكننا من أول وهلة بالنظر إلى الانحراف المعيارى القول بأن درجة التشتت فى الطول أكبر من مثيلاتها فى الوزن ولكننا إذا أخذنا فى الاعتبار أن درجة التشتت هذه مقاسه بالنسبة لمتوسطين مختلفين فضلاً عن أن وحدات القياس هى الأخرى مختلفة فإن النتيجة قد تكون مختلفة ...... وهذا بالفعل ما أثبته مقياس التشتت النسبى وهو معامل الاختلاف Coefficient of Variation حيث بمقارنة معامل الاختلاف للوزن (12.678) بمعامل الاختلاف للطول (5.377) نجد أن التشتت النسبى للوزن أكبر من التشتت النسبى للطول رغم أن الانحراف المعيارى للوزن أقل من الانحراف المعيارى للطول.