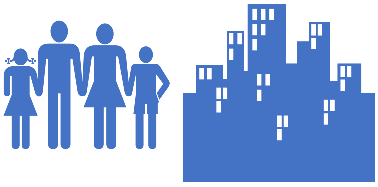
دبلوم ادارة اعمال

1444



مبادئ الاحصاء

101 كمي

جامعة الملك سعود

مركز التدريب وخدمة المجتمع

الفصل الأول

جمع البيانات الإحصائية وتنظيمها

Collecting Statistical Data and their Organizing

ينظر إلى علم الإحصاء على أنَّه أحد أهم فروع العلوم الرياضياتية خلال الخمسين سنة الأخيرة، وذلك لأنَّه يلعب دوراً مهمّاً في شتى مجالات الحياة بحيث يكاد لا يخلو أي مجال من مجالات المعرفة من استخدامه. إنَّ الخدمات الكبيرة لعلم العشوائيات دفع علماء الرياضيات قُدماً في تسخير كل قدراتهم لتطوير هذا العلم حتّى يلبي متطلبات التطور التقني المعاصر، ولذا نجد أنَّ الأستاذ الدكتور إيفو شنايدر Ivo Schneider (**أستاذ في تاريخ العلوم الطبيعية**) يذكر في هذا الصدد (**في كتابه تاريخ الاحتمالات منذ البدايات وحتى عام 1933**)، أنَّه ما من فرع من فروع الرياضيات تطوَّر خلال العقود الأخيرة مثل التطور الذي حظي به فرع العشوائيات (**الاحتمالات والإحصاء الرياضي**). في الواقع إنَّ علم العشوائيات أوسع بكثير من أن يُلَّم به في كتاب أو مجلد أو حتى عدّة مجلدات، وذلك لكثرة اختصاصاته الفرعية وموضوعاته المتنوّعة.

1-1- تعاريف ومفاهيم أساسية

1-2- تنظيم البيانات الخام وتمثيلها

1-3- التوزيعات التكرارية

1-4- التمثيلات البيانية لجداول التوزيع التكرارية

1-5- أشكال التوزيعات التكرارية

1-1- تعاريف ومفاهيم أساسية:

Basic definitions and concepts

في الواقع إنّ علم الإحصاء يتعامل بشكل أساسي مع البيانات المستحصل عليها من التجارب والظواهر، ولكيلا يكون هنا التباس في استخدم كلمة "**بيانات**" سنقدَّم تعريفها على النحو الآتي.

1-1-1- تعريف (علم الإحصاء Statistics)

إنَّ علم الإحصاء هو ذلك الفرع من الرياضيات الذي يهتمّ بجمع البيانات، وتنظيمها (**في جداول وعروضٍ بيانية مناسبة**)، ودراسة خصائصها، وتحليلها، واستقرائها، وأخيراً اتخاذ القرارات المناسبة بشأنها.

كما هو واضح من تعريف علم الإحصاء فإنَّه يمكن تجزئة هذا العلم إلى قسمين رئيسيين. الأول منهما يهتمّ بجمع البيانات وتنظيمها ودراسة خصائصها العددية (**ويُدعى الإحصاء الوصفي**)، وأمّا الآخر فإنَّه يهتمّ بتحليل البيانات واستقرائها واتخاذ القرارات المناسبة بشأنها (**ويُدعى الإحصاء الاستدلالي، وفي مراجع أخرى يذكر باسم الاستدلال الإحصائي أيضاً**)، وبناءً على ذلك يمكننا أن نقدِّم التعاريف الآتية.

1-1-2- تعريف (الإحصاء الوصفي Descriptive Statistics)

الإحصاء الوصفي هو ذلك الفرع من الإحصاء الذي يهتمّ بجمع البيانات وتنظيمها في جداول وعروض بيانية مناسبة، ودراسة خصائصها العددية.

في الواقع إنَّ هذه الجزئية من العلم يشترك فيها الاختصاصي وغير الاختصاصي أيضاً، فعلى سبيل المثال في عملية جمع البيانات الإحصائية يهتمّ المختصون بالطّرائق والأساليب (أي بالتقنيّة) التي ستجمع بها البيانات، في حين يمكن لأشـــخاص غيــر مختصين (ولكن مدربين) القيام بجمــع هذه البيانات، فعلى

سبيل المثال لو أخذنا عملية المسح السكاني الشامل في بلد ما، فإنَّ دور المختصين هو تحديد الطرائق والأساليب التي ستجمع بها البيانات، وإعداد الاستبيانات المناسبة للأهداف المطلوبة من هذه العملية. في حين يدرَّب معلّمون من المدارس على تعبئة الاستبيانات وكيفية تحصيل البيانات من الأهالي (لكيلا يحصل التّضليل الإحصائي). بعد ذلك يقوم الاختصاصي مرّةً أخرى بتنظيم هذه البيانات في جداول وعروض بيانية مناسبة (لهدف الدراسة الإحصائية) من أجل أخذ انطباع أولي عن سلوك هذه البيانات.

1-1-3- تعريف (الإحصاء الاستدلالي Inferential Statistics)

الإحصاء الاستدلالي هو ذلك الفرع من علم الإحصاء الذي يهتمّ بتحليل البيانات واستقرائها (تعميم نتائج العينات على المجتمع الإحصائي) ومن ثمَّ اتخاذ القرارات المناسبة بشأنها.

إنَّ العمل في هذا القسم من الإحصاء يقع على عاتق الاختصاصيين حصراً، وذلك لأنَّها تتطلب مهارات علميَّة على مستوى أكاديمي، حيث يبدؤون بتحليل البيانات ودراسة توزيعاتها وتعيين مَعْلماتها ومن ثمَّ البحث عن مقدِّرات مناسبة لهذه المعالم وبعد ذلك دراسة سلوك تلك المقدِّرات و.... وأخيراً اتخاذ القرارات المناسبة بشأن المجتمع قيد الدراسة. بالطبع سوف لن نتناول هذه الدراسة في كتابنا هذا لأنَّها خارج إطار خطته.

ونستعمل الإحصاء الوصفي لدراسة مجتمع صغير كطلاب الصف في المدرسة, بينما نستعمل الإحصاء الاستدلالي لدراسة المجتمع الكبير كالمدن والقرى.

في هذا الفصل سوف نتناول أول وأبسط جزئية من علم الإحصاء ألا وهي "**البيانات الإحصائية جمعها وتنظيمها**" كما ورد في عنوان هذا الفصل، وهذه الدراسة تتطلب العديد من المفاهيم الإحصائية وأولها هو الآتي.

1-1-4- تعريف (البيانات Data) البيانات هي قياسات أو ملحوظات (أو مشاهدات) تهدف إلى غرض معيّن في مجتمع إحصائي (سنأتي على تعريفه بعد قليل) مُحدّد، ويتمّ تدوينها كنتيجة لعمليّة إنتاجيّة (مثل كميّات القمح الناتجة عن الزّراعة في عام أو أعوام متتالية في بلد ما) أو لتجربة معمليّة (مثل معرفة الألوان النّاتجة عن تحليل ضوء الشمس) أو لمراقبة (ملاحظة) كائنات موجودة في الطبيعة (مثل أعداد النجوم في مَجرّة في الفضاء الكوني) أو ....

من المعتاد لدى دراسة أي علم من العلوم التّطرق أولاً إلى تعريف هذا العلم، وذلك لأنَّ أي علم من العلوم يبدأ بتعريفه الذي يعدّ بمثابة حجر الأساس لذلك العلم. لذلك لا بدَّ لنا أولاً أن نتعرف على ما تعنيه عبارة "**علم الإحصاء**". في الحقيقة يمكن تعريف علم الإحصاء على النحو الآتي.

1-1-5- المجتمعات الإحصائية Populations

في الواقع إنَّ الدّراسة الإحصائية لأيّة مسألة تنطلق ممّا يُعرف باسم "**المجتمع الإحصائي**" الذي يكوِّن الأساس للبيانات التي ستخضع للدراسة، ولذلك لابدّ من تقديم تعريف واضح لمعنى المجتمع الإحصائي كي لا يكون هناك أي لبس في ذهن القارئ لهذا المفهوم المهمّ.

1-1-5-1- تعريف (المجتمع الإحصائي Population)

المجتمع الإحصائي هو أي تجمُّع لأشياء تَجمَع بينها صفة مشتركة واحدة على الأقل لتكون محل دراسة لهدف محدَّد أي هو المجتمع الهدف الذي تجرى عليه الدراسة لحل مشكلة معينة.

1-1-5-2- ملاحظات

1- سنستخدم كلمة "**مجتمع**" عوضاً عن "**مجتمع إحصائي**" على سبيل الاختصار والتبسيط، وإذا ما كُتبت بين الحين والآخر فإنَّ ذلك من باب التذكير بها فقط.

2- تُدعى مكوّنات المجتمع **عناصر** أو **أفراداً**.

3- إنَّ عدد عناصر المجتمع يُدعى **حجم المجتمع**، ولذلك فمن الممكن أن يكون حجم المجتمع:

- **محدوداً**، وفي هذه الحالة يكون عدد عناصر المجتمع منتهياً ويُرمز له بعدد طبيعي، وقد درجت العادة على استخدام أحرف لاتينية كبيرة من قبيل N ، M و... للدلالة على حجم المجتمع، فعلى سبيل المثال من أجل دراسة إحصائية ما على طلاب جامعة الملك سعود يمكن النظر إلى طلاب هذه الجامعة على أنَّه مجتمع محدود.

- **غير محدود**: وفي هذه الحالة يكون عدد عناصر المجتمع غير منتهٍ، ولذلك لا يستخدم رمزاً للدلالة على حجم المجتمع في هذه الحالة، فعلى سبيل المثال من أجل دراسة إحصائية ما على سلوك الأعداد الأوليّة (**مثل عشوائيتها وتوزيعها**) فإنَّه يمكن النظر إلى مجموعة الأعداد الأولية على أنَّها مجتمع غير محدود.

4- من الأمور المهمّة هنا هي أن ندرك أنَّ المجتمع ليس بالضرورة أن يكون مجتمعاً بشرياً أو حتى مجتمعاً لأحياء، إذ إنَّه من الممكن أن يكون جماداً أو أي شيء آخر أيضاً، والمثالان الآتيان يوضّحان لنا ذلك.

- لدى تحديد نسبة خام النّحاس في فلز معدني في منجم معيّن، حيث يمكن أن تتواجد أنواع عديدة من مركّبات النحاس في هذا المنجم، ولكنَّها جميعاً تحوي على معدن النحاس، ولذلك فلز النحاس في هذا المنجم يكوِّن مجتمعاً.

- في دراسة لتحديد الحالة الفنيّة للطائرات السّفريّة في المملكة العربية السعودية، حيث يوجد أنواع عديدة من الطائرات السفرية، ولكنَّها جميعاً تتصف بأنَّها طائرات سفرية، ولذلك الطائرات السفرية الموجودة في المملكة العربية السعودية تكوِّن مجتمعاً.

1-1-6- العيّنات Samples

قد تكون عملية إخضاع جميع عناصر المجتمع للبحث والدراسة شاقة، وأكثر من ذلك قد تكون في كثير من الحالات غير ممكنة أيضاً، فعلى سبيل المثال:

- لو أرادت هيئة الرّقابة على الأدوية التحقُّق من مكوّنات عقار دوائي معيّن معبأ في كبسولات، فعندئذٍ من غير المجدي أن تقوم هذه الهيئة بإخضاع كل إنتاج المصنع (ومن ثمَّ إتلاف كافة الإنتاج) للتحليل المخبري من أجل التثبت من أنَّ المنتج مُحققاً للمواصفات المقدَّمة من قِبل المصنع.

- في عملية تحليل الدم لمريض فمن غير المعقول ولا المقبول أخذ كل دم المريض (ومن ثمَّ قتل المريض) لتحليله من أجل الكشف على أسباب مرضه.

بالطبع هناك عوامل أخرى قد تضطرُّنا إلى عدم التعامل مع عناصر المجتمع كلّه لأسباب أخرى منها الاقتصاديّة والزمنيّة أيضاً، ولذلك يلجأ المرء في مثل هذه الحالات إلى أخذ جزء من المجتمع لدراسته، وهذا العمل يندرج تحت مفهوم العينة والذي يقدِّمه لنا التعريف الآتي.

1-1-6-1- تعريف (العيّنة Sample)

تُعرَّف العيّنة على أنَّها جزء من مجتمع إحصائي يتمّ اختياره (أو سحبه) بطريقة مناسبة بحيث يُنظر إلى هذا الجزء كممثِّل جيد للمجتمع وغالبا ما تكون العينة عشوائية.

أي ان العينة جزء من المجتمع الاحصائي التي تجرى عليه الدراسة الفعلية وتعمم النتيجة الدراسة على المجتمع الاحصائي.

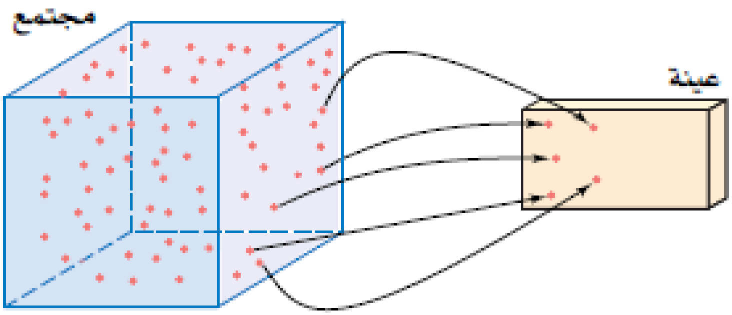
1-1-6-2- ملاحظات

1- سنستخدم كلمة عيّنة عوضاً عن كلمة عيّنة إحصائيّة على سبيل الاختصار والتبسيط.

2- نشير إلى أنَّ عدد عناصر العيّنة يجب أن يكون منتهياً.

3- يُطلق على عدد عناصر العيّنة اسم "حجم العيّنة"، وقد درجت العادة على استخدام أحرف لاتينية صغيرة من قبيل n، m و... للدلالة على حجم العيّنة.

يوضّحان مفهوم المجتمع والعيِّنة.[1-1-b] و[1-1-a]الشكلان الآتيان



الشكل [1-1-a]

الشكل [1-1-b]

1-1-6-3- تصنيف العيّنات Classification of Samples:

مما تقدّم يتبيَّن لنا أنَّ استخدام مفهوم العيّنة في الدراسة (أو البحث) إنَّما هو وسيلة لتعميم ما تمَّ التوصل إليه من نتائج على المجتمع، وذلك لاعتقادنا أنَّ هذه العيِّنة هي مُمثِّل مقبول للمجتمع. هذا وتصنَّف العينات في نوعين رئيسيين هما:

1- **عيّنات عشوائيّة**، ويتميَّز هذا النوع من العيّنات بأنَّ عناصرها تسحب من المجتمع بطرائق عشوائيَّة، ومن أهمّ أنواعها وأكثرها انتشاراً **العيّنات العشوائية البسيطة**، وهذا النوع من العينات يتميَّز بأنَّ لجميع عناصرها النصيب نفسه في الاختيار (**أو السحب أو الانتقاء**) مع أي عيّنة عشوائية أخرى بذات الحجم وممكنة التشكيل من المجتمع نفسه، وهذا يوافق الحالة التي يكون فيها جميع عناصر المجتمع مستقلَّة بعضها عن البعض الآخر ولها النصيب نفسه في الظهور (**أو الاختيار**).

2- **عيّنات عمديَّة** (**أو قصديَّة**)، ويتميَّز هذا النوع من العيّنات بأنَّ عناصرها تُنتقى من المجتمع وفقاً لرأي الباحث وخبرته، ولهذا السبب فإنَّه من النادر استخدام هذا النوع من العينات لأنَّه من الممكن أن يتحيّز الباحث في عملية الانتقاء.

تعريف الاستبانة بأنه نموذج يحتوي على مجموعة من الأسئلة التي تم وضعها من قبل الشخص الذي يريد عمل بحث عن موضوع معين ويريد تجميع معلومات وبيانات معينة ليتم جمعها وتحليلها وتسجيل النتائج. ومن الممكن أن تكون الأسئلة مفتوحة أو تحتمل الصواب والخطأ فقط أو من الممكن أن تكون الإجابة عليها بنقاط محددة ويتم تحديد الأسئلة من قبل الباحث.

1-1-7- المتغيّرات Variables

لقد لاحظنا أنَّه لدى أية دراسة إحصائية نكون أمام هدفٍ مُحدَّد المعَالم نأمل الوصول إليه، ومن أجل ذلك كنّا نقوم بتطبيق أداة معيَّنة على أفراد العينة أو المجتمع للحصول على البيانات التي نرغب بها من أجل الوصول إلى قرار بشأن دراستنا الإحصائية. في الواقع إنَّ هذه الأدوات التي ذكرناها تندرج تحت أحد المفاهيم المهمّة في الرياضيات (**ألا وهي التطبيقات**)، وتُقدَّم في الإحصاء تحت مسمّى المتغيِّرات التي يقدِّمها لنا التعريف الآتي.

1-1-7-1- تعريف (المتغيّر Variable)

يُعرَّف المتغيّر على أنَّه دالّة مجالها العيّنة أو المجتمع نفسه (**حسب طبيعة الدراسة الإحصائية**)، وأمَّا مجالها المقابل فهو مجموعة ذات طبيعة ما (**يمكن لها أن تكون أعداداً أو رموزاً أو مسمّيات**)، ويستخدم لقياس خاصيّة معيّنة لعناصر العيّنة أو المجتمع.

1-1-7-2- ملاحظة:

من التّعريف السابق يتضِّح لنا أنَّ القياسات أو المشاهدات التي يمكن أن تنتج عن متغيّر قد تكون قيماً عدديّةً أو أحرفاً أو رموزاً أو...، وبناءً على ذلك يمكننا تصنيف المتغيّرات في نوعين رئيسيين هما:

1- **المتغيّر الكميّ** Quantitative Variable، وهو متغيّر تكون قيمه أعداداً حقيقية تنتج عن التساؤل بـ "**كم**". إنَّ البيانات التي تنتج عن هذا المتغيّر تُدعى **بيانات كميَّة** Quantitative Data، فعلى سبيل المثال:

أ- المتغيّر الذي يرصد أعداد الطلاب في الجامعات السعودية تكون قيمه أعداداً طبيعيةً ينتج عن التساؤل بـ "كم"، فلو كان عدد طلاب جامعة الملك سعود يساوي 50000 طالب، فعندئذٍ تنتج هذه القيمة عن السؤال: كم عدد طلاب جامعة الملك سعود؟

ب- المتغيّر الذي يرصد أعداد الأطفال لدى الأسر في مدينة الرياض تكون قيمه أعداداً صحيحة غير سالبة ينتج عن التساؤل بـ "كم"، فلو كان لدى أسرة X من مدينة الرياض خمسة أطفال، فعندئذٍ هذه القيمة تنتج عن السؤال: كم عدد الأطفال لدى الأسرة X؟

ج- المتغيّر الذي يرصد الطول (مقدَّراً بالسنتيمتر) لدى أفراد بلد ما، فإنَّ قيم هذا المتغيّر هي أعداد حقيقية، ومن الممكن أن يكون لمجموعة قيمه الفترة [45 , 246]، وينتج عن التساؤل بـ "كم"، فلو أخذ شخص ما X من ذلك البلد وكان طوله 176 سنتيمتر، فعندئذٍ هذه القيمة تنتج عن السؤال: كم طول الشخص X؟

2- **المتغيّر النوعيّ** Qualitative Variable، وهو متغيّر تكون قيمه عبارة عن رموز أو أسماء أو أرقام دالَّةً على نوع أو اسم أو صفة أو تمييز، وهذه القيم تنتج عن السؤال بـ "**ما**" أو "**كيف**".

إنَّ البيانات التي تنتج عن هذا المتغيّر تُدعى **بيانات نوعيَّة** Qualitative Data، فعلى سبيل المثال:

أ- المتغيّر الذي يرصد ألوان الزهور في حديقة معيّنة هو متغيّر نوعي، والقيم التي تنتج عنه (أحمر، أصفر، أبيض و...)، ونحصل عليها بالسؤال: **ما** لون الزهرة؟

ب- المتغيّر الذي يرصد فصيلة الدم لدى البشر تكون مجموعة قيمه رموزاً O، A، B وAB، ونحصل على هذه القيم بالسؤال: **ما** فصيلة دم الشخص X؟

ج- المتغيّر الذي يرصد الرقم الجامعي لطالب في جامعة الملك سعود هو متغيّر نوعي، والقيم التي تنتج عنه هي أرقام من قبيل 436......، 437..... و.... وهذه الأرقام تميّز الطّالب ولا تعني مقداراً كميّاً له، ونحصل على هذه القيم بالسؤال: **ما** رقم الطالب X؟

1-1-7-3- نتيجة

من الأمثلة السابقة نستنتج أنَّه يمكننا تصنيف المتغيّرات النوعية والكميَّة إلى نوعين أيضاً وذلك بحسب عدد البيانات الناتجة عن الدراسة، حيث لاحظنا أنَّ مجموعات القيم للمتغيّرات (النوعية أو الكميَّة) قد تكون منتهية أو غير منتهية قابلة للعدَّ، أو أن تكون غير منتهية وغير قابلةٍ للعدَّ، وبناء على ذلك يمكن تصنَّف المتغيّرات (وبغض النظر عن نوعها) في أحد النوعين الآتيين:

1- **متغيّر متقطِّع** Discrete Variable، وهو ذلك المتغيّر (نوعي أو كميّ) الذي تكون مجموعة قيمه منتهيةً أو غير منتهيةٍ ولكن قابلة للعدّ، ومن الأمثلة على ذلك:

أ- لنفترض أنَّ عدد السّيارات المباعة يومياً من معرض ما وعلى مدى عشرة أيام كانت كالتالي:

3	2	5	0	0	2	4	2	5	1

فعندئذ بفرض أنَّ X هو متغيّر يرصد عدد السّيارات المباعة يومياً من هذا المعرض وعلى مدى هذه الأيام العشرة، فإنَّه سيكون لهذا المتغيّر ست قيم فقط هي: 0، 1، 2، 3، 4 و5، ومن ثمَّ X هو متغيّر متقطِّع.

ب- لنفترض أنَّنا نقوم بقذف قطعة نقود معدنيّة حتى الحصول على شعار لأول مرَّة، وأنَّ X هو متغيّر يرصد عدد التجارب التي نفِّذت حتى الحصول على شعار لأول مرَّة، فإنَّه سيكون لهذا المتغيِّر عدداً غير منتهٍ ولكن قابل للعدّ من القيم هي 1، 2، 3 و...، وهذه القيم تكوِّن مجموعة الأعداد الطبيعية كاملة $\{1,2,3,...\}$، ومن ثمَّ X هو متغيّر متقطِّع.

ج- لنفترض أنَّ فريق طبي يجري بحثاً حول أنواع فصائل الدم الأكثر انتشاراً في فصل من فصول السنة الأولى المشتركة مكوَّنٍ من عشرين طالباً، وأنَّ النتائج كانت كالآتي:

A	B	O	B	AB	O	A	AB	O	O
B	O	A	A	AB	O	A	A	O	A

فعندئذ بفرض أنَّ X هو متغيّر يرصد نوع فصيلة الدم لطالب من هذا الفصل، فإنَّه سيكون لهذا المتغيّر أربع قيم فقط هي: O، A، B وAB، ومن ثمَّ X هو متغيّر متقطِّع.

د- لنفترض أنَّنا نقوم بقذف قطعة نقود معدنيّة حتى الحصول على شعار لأول مرَّة، وأنَّ X هو متغيِّر يرصد النتائج التي سنحصل عليها حتى الحصول على شعار لأول مرَّة، فإنَّه سيكون لهذا المتغيّر عدداً غير منتهٍ قابلٍ للعدّ من القيم، وهي:

T, HT, HHT, HHHT, HHHHT, ...

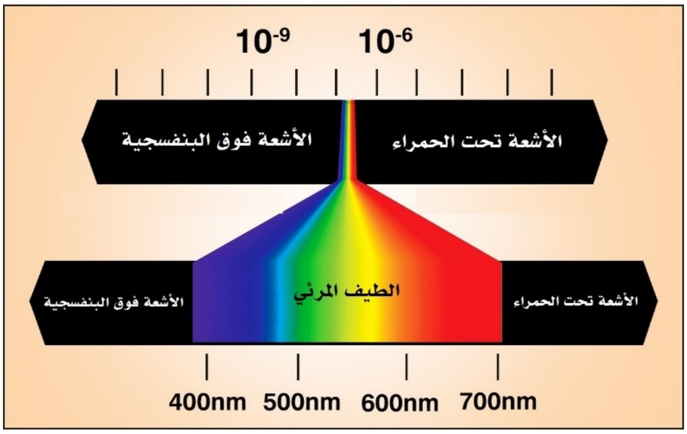
ومن ثمَّ X هو متغيِّر متقطِّع.

2- متغيّر مستمرّ (أو متّصل)، Continuous Variable وهو ذلك المتغيّر (نوعي أو كميّ) الذي تكون مجموعة قيمه غير قابلة للعدّ (**وبالتالي غير منتهية أيضاً**)، ومن الأمثلة على ذلك:

أ- بفرض أنَّ X هو متغيّر يرصد عمر الإنسان في القرن الأخير 1916-2016، فعندئذ نجد أنَّ قيم هذا المتغيِّر تقع في الفترة $[0 , 179]$، وهي مجموعة غير قابلة للعدّ، ومن ثمَّ X هو متغيّر مستمر.

ب- بفرض أنَّ X هو متغيّر يرصد الوقت المستغْرق من قبل طالب لإنهاء اختباره (**بزمن 120 دقيقة**) في مقرَّر معيّن، فعندئذ نجد أنَّ قيم هذا المتغيِّر تقع في الفترة $(0 , 120]$، وهي مجموعة غير قابلة للعدّ، ومن ثمَّ X هو متغيّر مستمر.

ج- بفرض أنَّ X هو متغيّر يرصد اللون الذي له طول موجي يقع في الفترة $[400nm , 700nm]$ (nm **تعني: نانو متر** $1nm = 10^{-9}m$)، فعندئذ نجد أنَّ هذا المتغيّر يرصد القسم الأكبر من ألوان الطيف الضوئي المرئي (**انظر الشكل الآتي**)، ومن ثمَّ عدد الألوان التي يرصدها هذا المتغيّر غير قابلة للعدّ، وبالتالي فإنَّ هذا المتغيّر النوعي هو متغيّر مستمر.



الشكل [1-2]

الشّكلان الآتيان يوضّحان لنا تصنيفات المتغيّرات والبيانات.

الشكل [1-3-a]

الشكل [1-3-b]

مراحل تنفيذ الدراسة الإحصائية

علم الإحصاء من العلوم المهمة وتتكون عملية الإحصاء من مجموعة من المراحل التي تتم متتالية وهي ما يلي:

ان اي عملية احصائية تمر بعدة مراحل:

واولها هو تحديد الظاهرة المطلوب دراستها .

وثانيا جمع المعلومات والبيانات المطلوبة حول هذه الظاهرة .

وثالثا التأكد من البيانات وتصنيفها وتقسيمها بطريقة يسهل دراستها .

ورابعا تحليل هذه البيانات من خلال ادوات الاحصاء المعروفة.

وخامسا عمل الاستنتاجات اللازمة التي تساعد في حل المشكلة.

سادسا وأخيرا حيث يتم تطبيق العملية الإحصائية في الواقع العملي على المجتمع الاحصائي.

1-2- تنظيم البيانات الخام وتمثيلها:
Organizing and Representing Raw Data

الآن، وبعد إتمام عملية جمع البيانات تأتي المرحلة التّالية، وهي البحث في كيفية التعامل مع هذه البيانات. حيث يتمّ التحقُّق أولاً من اكتمال البيانات، فإذا كانت هناك بعض البيانات المفقودة فعندئذٍ يجب النّظر في كيفية تحصيلها ثانية أو تقديرها إن أمكن ذلك، وإلاَّ سوف تُستثنى من الدراسة (**علماً أنَّ عملية التقدير للبيانات المفقودة تقع خارج إطار هذا الكتاب، ولذلك سوف لن نتطرق إلى عملية تقدير البيانات المفقودة**).

إنَّ البيانات التي نحصل عليها قد تكون على أشكال مختلفة، فمنها على شكل قيم عددية مُفْردة، وبعضها الآخر قد يَعْرِض تَغيُّرَ ظاهرةٍ ما مع مرور الزمن أو مع مسمَّيات كالبلدان، أو المدن، أو مع كليهما معاً، وبعضها الآخر قد يكون مجمَّعاً في جداول. لذلك سنبحث في تنظيم البيانات وفق اتجاهين:

الأول: يهتمّ بتنظيم البيانات المُفْرَدة التي تنتج مباشرة عن الدراسة الإحصائية (**كميَّةً كانت أم نوعيّة**) في جداول تُدعى الجداول التكرارية، ومن ثمَّ تمثيل هذه البيانات في عروض بيانية مناسبة.

الثاني: يهتمّ بتجميع البيانات المُفْرَدة الكميَّةً فقط في جداول من نوع خاص تُدعى جداول التوزيع التكرارية، حيث يُقال عن البيانات المقدَّمة بهذه الجداول إنَّها بيانات مجمَّعة (**أو مبوَّبة، أو مجدولة**)، ومن ثمَّ تمثيل بيانات هذه الجداول في عروض بيانية مناسبة.

1-2-1- البيانات الخام Raw Data

لدى تنفيذ دراسة إحصائية معيّنة حول ظاهرة ما وجمع البيانات حول هذه الظاهرة تنتج لدينا بيانات مُفْرَدة تُدعى **بيانات خام**.

إذا كان عدد البيانات صغيراً فإنَّه يمكن التعامل مع هذه البيانات بشكل مباشر (**مع كل قيمة على انفراد**) لدراستها، وأمّا إذا كان عدد البيانات كبيراً نسبياً، فإنَّه قد يكون من الصعب التعامل معها بشكل مباشر، ولذلك لا بدَّ من تقديم طرائق تسهل التعامل مع هذه البيانات كي يتمّ الاستفادة منها بأقصى قدر ممكن.

فيما يلي نقدِّم بعض الأمثلة على بيانات خام قبل الشروع في تقديم طرائق عرضها جدولياً.

1-2-1-1- أمثلة

1- لدى الاطلاع على تقديرات 60 طالباً في مقرَّر الإحصاء وجدنا المُعطيات الآتية:

C	A	D	B	D	A	F	D	C	A
D	D	D	C	C	B	C	F	F	C
A	B	C	D	D	A	D	A	A	B

D	C	F	D	C	B	C	C	B	C
B	D	A	B	B	C	B	A	D	C
C	C	F	C	B	D	C	D	B	F

فنلاحظ أنَّ هذه البيانات هي بيانات خام نوعيّة.

2- لقد سُئل 30 طالباً من كلية X عن عدد الحوادث المرورية التي حصلت معه خلال الفصل الدراسي الأول لهذا العام فكانت الإجابات كما يلي:

0	0	1	3	1	0	1	2	2	3
2	0	1	2	1	1	1	1	2	1
1	0	0	0	3	2	2	1	0	0

فنلاحظ أنَّ هذه البيانات هي بيانات خام كميّة.

3- البيانات الآتية تمثِّل الطول لخمسين طالباً جامعياً (مقدَّرة بالسنتيمتر):

140	155	168	171	158	168	159	149	172	145
155	154	166	169	168	158	149	172	168	166
156	166	149	157	156	159	167	166	169	171
170	159	168	168	167	157	154	166	169	168
158	157	172	155	154	166	168	167	171	168

وهذه البيانات هي بيانات خام كميَّة، ولكنَّها تتبع مجموعة بيانات مستمرّة (أو متّصلة). حيث نلاحظ أنَّ كمية البيانات الخام المقدَّمة أعلاه لا تعدّ كبيرة نسبياً إلاَّ أنَّه يصعب أخذ انطباع سريع عن سلوك هذه القيم بشكلٍ مباشر، فعلى سبيل المثال: هل كل قيمة في هذه المجموعة تتكرَّر بالقدر نفسه؟

بالطّبع قد يقوم المرء بإجراء تعداد لكل بيان من هذه البيانات ومن ثمَّ إجراء مقارنة بين هذه التّعدادات، ولكنَّ ذلك سيستغرق وقتاً لا بأس به. لذلك اقترح تنظيم هذه البيانات بطريقة معيّنة حتى يسهل على المرء الاستفادة منها واستنباط سلوك هذه البيانات بأقل وقت وجهد ممكن. من هنا جاءت فكرة صبّ البيانات في جداول يذكر في أحد أعمدتها البيان المميَّز (الرمز أو العدد الممثِّل)، وفي العمود المقابل يذكر التعداد وفي العمود الذي يليه يدوَّن عدد يعبِّر عن تعداد هذا البيان.

1-2-2- التّمثيل الجدولي للبيانات الخام Table Representation of Raw Data

إنَّ تمثيل البيانات الخام جدولياً يعني صبّ هذه البيانات في جدول بتصميم معيَّن، وهذا الجدول يُدعى **الجدول التكراري** للبيانات. فإذا أردنا صبّ مجموعة بيانات خام في جدول تكراري نقوم بإدراج جدول يحتوي على خمسة أعمدة، وهذه الأعمدة تُخصَّص على النحو الآتي:

1- يدوَّن في العمود الأول **ممثِّل** لكل نوع من الأسماء أو الرموز أو الأعداد (ولتكن x_1 و x_2 و... و x_m على افتراض أنَّ عدد الممثِّلين m) حسب طبيعة البيانات التي قيد الدراسة، فعلى سبيل المثال لدينا:

أ- الرمز A هو ممثِّل لكل الرموز A في مجموعة بيانات المثال (1) من الفقرة (1-2-1-1)،

ب- العدد 0 هو ممثِّل لكل الأعداد 0 في مجموعة بيانات المثال (2) من الفقرة (1-2-1-1).

2- يدوَّن في العمود الثّاني **التّعداد** Tally لكل ممثِّل (من الأسماء أو الرّموز أو الأعداد) في مجمل البيانات التي قيد الدراسة، ويتمّ ذلك برسم خط عمودي عن كل بيان موافق للاسم أو الرمز أو العدد، وإذا أصبح لدينا أربعة خطوط عمودية فإنَّ الخط الخامس يحزمها على النحو ~~||||~~.

3- يدوَّن في العمود الثّالث عدد يُعبّر عن تعداد كل ممثِّل (**من الأسماء أو الرموز أو الأعداد**) في مجمل البيانات التي قيد الدراسة، وهذا العدد يُدعى **التكرار** Frequency (**وسنرمز لها بـ** f_1 **و** f_2 **و ... و** f_m **على افتراض أنَّ عدد الممثِّلين** m).

4- يدوَّن في العمود الرّابع عدد يعبّر عن نسبة تكرار كل ممثِّل (**من الأسماء أو الرموز أو الأعداد**) إلى العدد الكلي للبيانات التي قيد الدراسة، وهذا العدد يُدعى **التّكرار النّسبي** Relative Frequency. أي إنَّ التكرار النسبي يساوي تكرار النوع مقسوماً على المجموع الكلي للتكرارات.

5- يدوَّن في العمود الخامس عدد يعبّر عن حاصل ضرب العدد 100 في التكرار النسبي لكل ممثِّل (**من الأسماء أو الرموز أو الأعداد**) ويقرأ كنسبة مئوية، وهذا العدد يُدعى **التكرار المئوي** Percent Frequency. أي إنَّ التكرار المئوي يساوي التكرار النسبي مضروباً في 100.

وهكذا يمكننا عرض نتائج صب البيانات في جدول تكراري نموذجي وفقاً للعرض الآتي:

الجدول [1-1]

الممثِّل	التعداد	التكرار
x_1	| ⋯	f_1
x_2	|| ⋯	f_2
⋮	⋮	⋮
x_m	~~||||~~ ⋯	f_m
sum	---------------	$\sum f_i$

نشير هنا إلى أنّه في حال البيانات الكميّة ليس بالضرورة أن يكون للممثّلين x_1 ، x_2 ،... و x_m ترتيب تصاعدي في الجدول، ولكن يفضَّل ذلك.

1-2-2-1- أمثلة

1- لنقم بصبّ البيانات الموجودة في المثال (1) من الفقرة (1-1-2-1) في جدول تكراري، فنجد أنَّ لهذا الجدول العرض الآتي:

الجدول [1-2]

التقدير	التعداد	التكرار
A	~~||||~~ ||||	9
B	~~||||~~ ~~||||~~ ||	12

C	||| 𝍸 𝍸 𝍸	18
D	𝍸 𝍸 𝍸	15
F	| 𝍸	6
Total	----------------	**60**

نلاحظ هنا سهولة الحصول على المعلومة الخاصّة بالتكرارات، فعلى سبيل المثال إذا أردنا الحصول على عدد الطلاب الحاصلين على تقدير C نجده بكل سهولة يساوي 18، وأنَّ نسبة الطلاب الذين حصلوا على تقدير A يشكلون 15% من إجمالي عدد الطلاب، وأنَّ 25% من الطلاب حصلوا على التقدير D.

2- لتكن لدينا البيانات الآتية و الناتجة عن فحص فصيلة الدم لستين شخصاً.

B	A	B	A	B	O	A	O	AB	B
A	O	A	AB	O	A	AB	O	A	AB
A	B	B	B	A	AB	O	A	AB	A
B	AB	A	A	AB	A	A	O	B	B
AB	A	B	O	A	B	A	AB	A	AB
A	B	A	A	AB	A	O	A	B	B

فلو قمنا بصبِّ هذه البيانات في جدول تكراري على النحو السابق، فإنّنا سنجد له العرض الآتي:

الجدول [1-3]

رمز فصيلة الدم	التعداد	التكرار
A	|||| 𝍸 𝍸 𝍸 𝍸	24
B	𝍸 𝍸 𝍸	15
AB	|| 𝍸 𝍸	12
O	|||| 𝍸	9
Total		**60**

3- في إحدى المدارس أخذت عيّنة مكوَّنةً من 40 طالباً من أجل دراسة عدد المرّات التي أصيب فيها الطالب بنزلة برد (**انفلونزا**) خلال موسم الشّتاء في عام 1438 هـ، فكانت النّتائج كما يلي:

1	0	1	3	2	1	2	2	1	0
2	2	0	1	0	1	2	0	2	1
1	2	1	3	1	2	1	2	1	1
1	0	1	1	2	0	0	1	1	2

فلو قمنا بصبِّ هذه البيانات في جدول تكراري، فإنّنا سنجد له العرض الآتي.

الجدول [1-4]

عدد مرات الإصابة	التعداد	التكرار

0	||| 卌	8
1	卌 卌 卌 |||	18
2	|| 卌 卌	12
3	||	2
Total		**40**

1-2-2-2- ملاحظات

1- إنَّ مجموع التكرارات النسبيّة يجب أن يساوي الواحد تماماً، ولكن عند تنفيذ بعض الحسابات نضطر إلى إجراء عمليّة تدوير للأرقام، وفي هذه الحالة قد لا نحصل على مجموعٍ يساوي الواحد تماماً، فيكون المجموع أكبر أو أصغر من الواحد بقليل.

2- إنَّ مجموع التّكرارات المئويّة يجب أن يساوي المئة تماماً، ولكن إذا ما حصلت عمليّة تدوير للأرقام فإنَّ مجموع التكرارات المئوية قد لا يساوي المئة تماماً، فيكون لدينا مجموع أكبر أو أصغر من المئة بقليل.

3- بعد الانتهاء من صبِّ البيانات يمكن الاستغناء عن عمود التعداد لأنَّ عمود التكرار يؤدي الغاية نفسها، وأمَّا في حال تقديم البيانات مجمَّعة في جدول تكراري فإنَّه (**وفي معظم الحالات**) لا يدرج عمود التعداد معه بسبب عدم وجود مبرر لظهوره، وبذلك يتبقى لدينا جدول بأربعة أعمدة فقط.

نوع آخر من التمثيلات الجدولية للبيانات تقدِّمها لنا الفقرة الآتية.

1-2-2-3- جدول الساق والأوراق Stem and Leafs Table

توجد طريقة أخرى لصب البيانات الخام في جدول مشابه للجداول التكرارية وتُدعى **عرض السّاق والأوراق**، والغاية منــه تصنيف البيانات في مجموعات جزئيــة، كأن نبيّن القيم ما بين 0 و9 في مجموعة جزئية واحدة، والقيم ما بين 10 و19 في مجموعة جزئية ثانية، والقيم ما بين 20 و29 في مجموعة جزئية ثالثة، و هكذا دواليك.

إنَّ صب البيانات الخام في جدول السّاق والأوراق يتمّ من خلال بناء جدول بعمودين أحدهما يخصَّص للساق والآخر للأوراق، وذلك على النحو الآتي:

بفرض أنَّه لدينا بيانات كميّة مكوَّنةً من أعداد صحيحة من خانتين على الأكثر، فعندئذٍ نضع في عمود الساق القيمة 0، وفي العمود المقابل لها (**عمود الأوراق**) نضع جميع القيم المكوَّنة من خانة واحدة فقط. بعد ذلك ننتقل إلى الأرقام المكوّنة من خانتين فنبدأ بالقيـم التي خانـة العشرات لها هي الأصغر،
فنضع العدد المكوّن لخانة العشرات في عمود الساق وأمّا الأجزاء الأخرى المكوّنة لخانة الآحاد من العدد نفسه فنضعها في عمود الأوراق، فإذا ما انتهينا من هذه ننتقل إلى الأرقام الأكبر والمكوَّنة من خانتين فنقوم بتطبيق الإجراء السابق عليها أيضاً، وهكذا دواليك حتى الانتهاء من عملية الصب. هذا ويفضَّل ترتيب القيم في عمود الأوراق تصاعدياً في العرض الأخير للبيانات، والمثال الآتي يوضِّح لنا ذلك.

1-2-2-4- مثال

لتكن لدينا البيانات الآتية التي تمثِّل درجات اختبار المنتصف لـ 24 طالباً:

13	3	27	22	8	11	14	17	7	21	6	12
16	6	5	25	25	23	15	6	24	5	19	13

عندئذٍ باستخدام الطّريقة التي تمَّ شرحها أعلاه نجد أن للبيانات المعطاة جدول الساق والأوراق الآتي:

الجدول [1-5]

الساقStem	الأوراق Leafs حسب موضعها في البيانات المقدَّمة	الأوراق بعد ترتيبها تصاعدياً
0	3 , 8 , 7 , 6 , 6 , 5 , 6 , 5	3 , 5 , 5 , 6 , 6 , 6 , 7 , 8
1	3 , 1 , 4 , 7 , 2 , 6 , 5 , 9, 3	1 , 2 , 3 , 3, 4, 5 , 6 , 7 , 9
2	7 , 2 , 1 , 5 , 5 , 3 , 4	1 , 2 , 3 , 4 , 5 , 5 , 7

في الواقع هذه ليست كلَّ التمثيلات الجدولية للبيانات حيث يوجد نوع آخر من التمثيلات الجدولية للبيانات يسمَّى **جدول التوزيع التكراري** وسنأتي على شرحه لاحقاً في هذا الفصل.

1-2-3- التّمثيل الشّرائطي للبيانات الخام Bar Chart Representation of Raw Data

يعدّ تمثيل البيانات الخام باستخدام الشّرائط العمودية (**ويُعرف باسم التمثيل بالأعمدة أيضاً**) أو الشرائط الأفقية من أحد التمثيلات الجيّدة للبيانات الخام، وذلك لأنَّها تعطي انطباعاً سريعاً حول طبيعة البيانات الخام وسلوكها، وسبب ذلك أنَّه من طبيعة الإنسان سرعة الملاحظة عند النظر إلى المشاهد والرسومات.

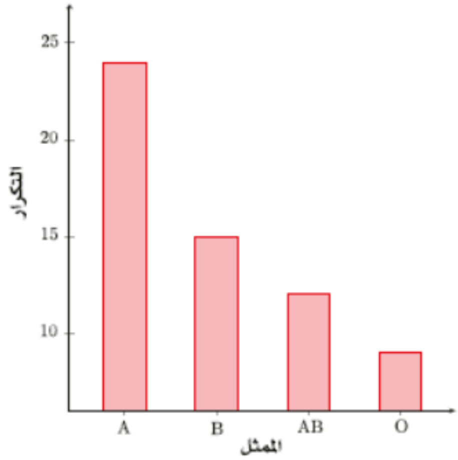
من أجل التمثيل الشّرائطي لمجموعة بيانات خام (**وبغض النظر عن نوعها اسميّةً، رموزاً أو عددية**) نقوم برسم محورين متعامدين XoY، ومن ثمَّ يُدوَّن أسفل المحور الافقي oX المُمثَّل لكل صنف في البيانات الخام (**فإن كانت أسماءً كُتِبت الأسماء، وإن كانت رموزاً وضِعَت الرموز وإن كانت أعداداً سُجِّلت الأعداد**)، وأمّا المحور العمودي oY فيدوَّن عليه قيم تكرارات البيانات الخام. بعد ذلك يُرسم عمود فوق كل مُمثِّل بيانات بارتفاع قدره

يساوي قيمة تكرار هذا الممثِّل مع الأخذ بالحسبان أن تكون هذه الأعمدة منفصلة بعضها عن البعض الآخر بتباعد ثابت (**وغالباً ما تؤخذ بمقدار وحدة قياس**). في هذه الحالة نحصل على التمثيل بالشرائط العمودية أو التمثيل بالأعمدة.

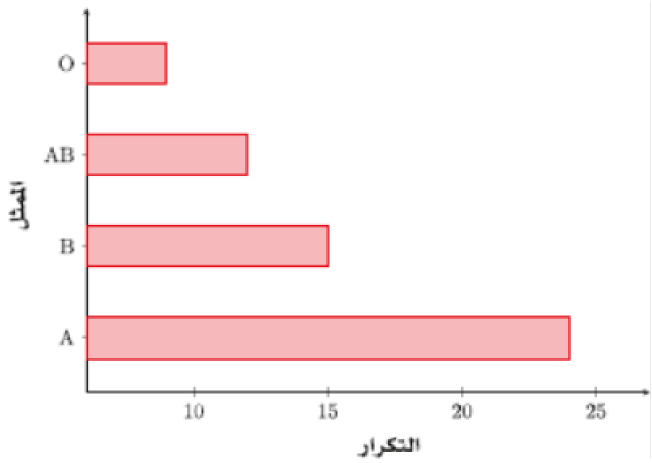
أمّا إذا أردنا تمثيل البيانات بالشرائط الأفقية فإنَّنا نعكس العمليات التي تمت على المحورين المتعامدين، فيصبح المحور oY من أجل تدوين الممثِّل لكل صنف في البيانات الخام في حين يستخدم المحور oX لتدوين قيم تكرارات البيانات الخام، وفي هذه الحالة لا يُقال عن التمثيل الناتج إنَّه تمثيل بالأعمدة للبيانات الخام. بعد ذلك يُرسم شريط أفقي إلى جانب كل مُمثِّل للبيانات بطول قدره يساوي قيمة تكرار هذا الممثِّل، ومع الأخذ بالحسبان أن تكون هذه الشرائط منفصلة بعضها عن البعض الآخر بتباعد ثابت أيضاً.

1-2-3-1- مثال

بالعودة إلى المثال (2) من (**1-2-2-1**) فإنّنا نجد تمثيل البيانات الخام المُعطاة باستخدام الشّرائط العموديّة (**أو التمثيل بالأعمدة**) له الشـكل [1-4-a]، وأمّا التمثيل باستـخـدام الشّـرائط الأفقيّة فله الشكل [1-4-b].



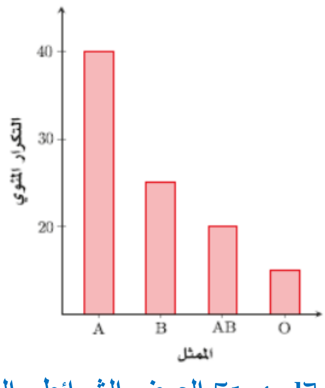
الشكل [1-4-a]



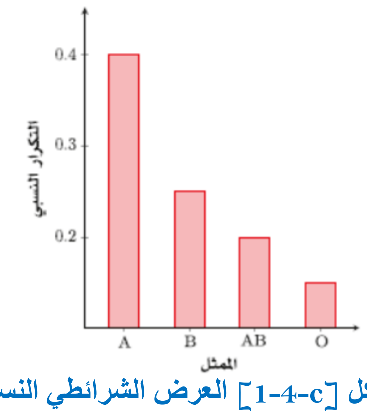
الشكل [1-4-b]

1-2-3-2- ملاحظة

يمكن استخدام التكرارات النّسبيّة والتكرارات المئوية بدلاً من التكرارات في التمثيل الشرائطي أيضاً، حيث تستبدل قيم التكرارات بقيم التكرارات النسبيّة أو التكرارات المئوية، فعلى سبيل المثال نجد من أجل المثال السابق أنَّ العرض الشرائطي النسبي والمئوي لهما الشكلين الآتيين:



الشكل [1-4-d] العرض الشرائطي المئوي



الشكل [1-4-c] العرض الشرائطي النسبي

1-2-3-3- التّمثيل بالشّرائط البيانيّة المزدوجة Pair Bar Charts Representation

يُعدّ التمثيل البياني بالشرائط المزدوجة من التمثيلات البيانية المهمّة عند مقارنة مجموعتين من البيانات أو أكثر (**ظاهرتين أو أكثر**) مع بعضها البعض الآخر، وهذه الطريقة في التمثيل لها الخطوات نفسها التي استخدمت من أجل التمثيل الشرائطي، ولكن لكل مجموعة بيانات على حدى مع وضع الأشرطة الممثِّلة للنوع الواحد ملاصقة أو قريبة بعضها من البعض الآخر، ويُميَّز أحدهما عن الآخر بالتّلوين المختلف أو التظليل المختلف.\

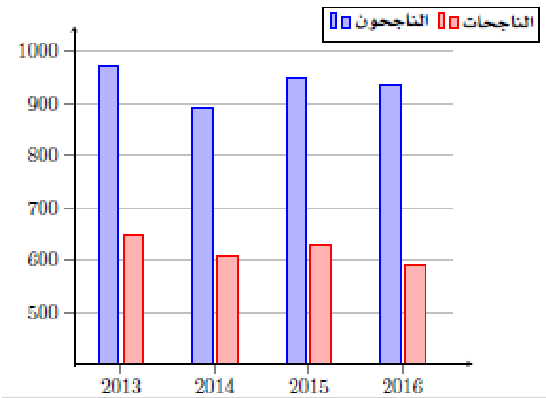
1-2-3-4- مثال

على سبيل المثال يمكننا أن نقارن أعداد الطلاب والطالبات الذين اجتازوا اختبار مقرَّر الإحصاء في السنة الأولى المشتركة في جامعة الملك سعود وذلك خلال الأعوام الدراسية 2013 وحتى 2016. حيث لدينا البيانات الخاصّة بذلك كما في الجدول المجاور.

الجدول [1-6]

العام الدراسي	عدد الطلاب الناجحون	عدد الطالبات الناجحات
2013	970	650
2014	890	610
2015	950	630
2016	935	590
Total	3745	2480

فعندئذٍ نجد أنَّ التمثيل بالشرائط المزدوجة للبيانات المقدَّمة أعلاه له الشكل الآتي.

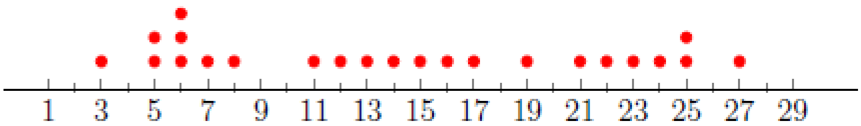


الشكل [1-5]

1-2-3-5- العرض النقطي للبيانات Dot Plots of Data

يعدّ تمثيل البيانات باستخدام النقاط إحدى الطرائق المستخدمة في تمثيل البيانات، ويتمّ ذلك من خلال وضع نقطة فوق القيمة الممثِّلة (أو الرمز الممثِّل) عن كل بيان يتوافق مع هذه القيمة (أو الرمز). إنَّ هذه الطريقة تستخدم عادة من أجل البيانات الكميّة، وذلك بجعل المحور الحقيقي أساساً للتمثيل فوقه.

على سبيل المثال لو أردنا تمثيل بيانات المثال 1-2-2-4 باستخدام العرض النقطي فإنَّنا سنجد له العرض الآتي:



الشكل [1-6]

1-2-4- التّمثيل بالقطاعات الدّائريّة (أو القرص الدائري) Pie Chart

يُنظر إلى التمثيل بالقطاعات الدّائريّة (أو القرص الدّائري) على أنَّه أحد الأشكال البيانية الواسعة الانتشار في دراسات الإحصاء الوصفي عندما يكون عدد ممثّلي البيانات قليلاً، وأمّا إذا كان عدد ممثّلي البيانات كبيراً فعندئذٍ تصبح الفائدة منه شبه معدومة.

يُلجأ عادة إلى استخدام هذه الطريقة عندما نكون بحاجة لتقسيم الكل إلى k من الأجزاء، وأمَّـا لرسمها فإنَّنا نقوم أولاً برسم دائرة بنصف قطر مُثبَّت (غالباً يكون عموديّاً) يُعدّ مبدأً لقياس الزّاوية عنـه، ثمَّ تُحسـب زوايا القطاعات الدائرية α_i مقدَّرة بالدّرجات Degrees، وتأخذ إلى يمين العمود السابق باتجاه دوران عقارب السّاعة، وبحيث يكون للممثِّل (أو الجزء) i قطّاع دائري زاويته تُحسب باستخدام العلاقة الآتية:

$$\alpha_i := \frac{n_i}{n} \times 360 \qquad [1-1]$$

علماً أنَّ n هو عدد البيانات الخام المُعطاة و n_i هو عدد العناصر (أو البيانات) التابعة للممثِّل (أو الجزء) i، بمعنى آخر، فإنَّنا نحصل على قيمة الزّاوية للقطاع التابع للممثِّل (أو الجزء) i من خلال ضرب قيمة التكرار النسبي لهذا الممثِّل في 360، والمثالان الآتيان يوضّحان لنا ذلك.

1-2-4-1 مثال

1- بالرّجوع إلى المثال (2) من (1-2-2-1) وباستخدام العلاقة [1-1] نجد أنَّ:

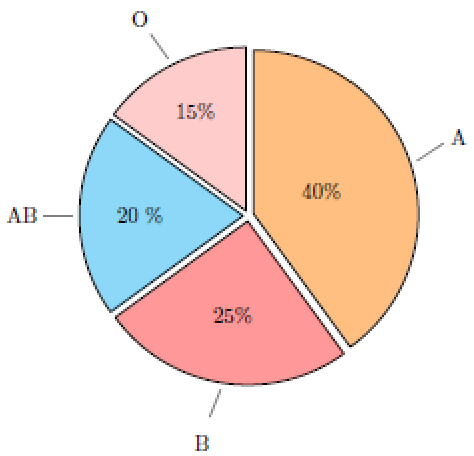
زاوية القطاع الممثِّل لبيانات A هي $\alpha_{\mathbf{A}} = \frac{24}{60} \times 360 = 144^o$

زاوية القطاع الممثِّل لبيانات B هي $\alpha_{\mathbf{B}} = \frac{15}{60} \times 360 = 90^o$

زاوية القطاع الممثِّل لبيانات AB هي $\alpha_{\mathbf{AB}} = \frac{12}{60} \times 360 = 72^o$

زاوية القطاع الممثِّل لبيانات O هي $\alpha_{\mathbf{O}} = \frac{9}{60} \times 360 = 54^o$

ومن ثمَّ يكون لدينا العرض الجانبي للقطاعات الدائرية.



الشكل [1-7]

2- بالرجوع إلى المثال (3) من (1-2-2-1) وباستخدام العلاقة [1-1] نجد أنَّ:

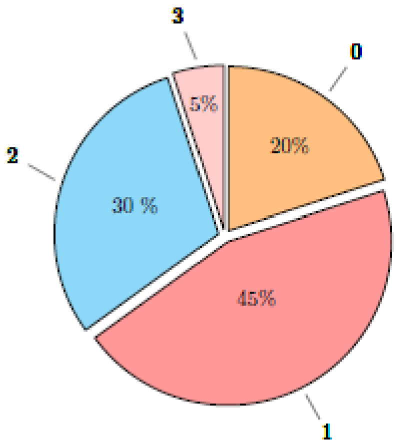
زاوية القطاع الممثِّل لبيانات 0 هي $\alpha_0 = \frac{8}{40} \times 360 = 72^o$

زاوية القطاع الممثِّل لبيانات 1 هي $\alpha_1 = \frac{18}{40} \times 360 = 162^o$

زاوية القطاع الممثِّل لبيانات 2 هي $\alpha_2 = \frac{12}{40} \times 360 = 108^o$

زاوية القطاع لممثِّل البيانات 3 هي $\alpha_3 = \frac{2}{40} \times 360 = 18^o$

ومن ثمَّ يكون لدينا العرض الجانبي للقطاعات الدائرية.



الشكل [1-8]

1-3- التّوزيعات التّكرارية
Frequency Distributions

لقد قدّمنا فيما سبق شرحاً مفصّلاً للجداول التكرارية حيث لاحظنا أنَّ تلك الجداول تعطينا تصوّراً سريعاً حول سلوك البيانات، وأكثر من ذلك فقد كانت تهتمّ بسلوك ممثِّل البيانات نفسه، فتظهر لنا تكراره وتكراره النسبي والمئوي إذا رغبنا في ذلك. إلاَّ أنَّه من أجل البيانات الكميّة المستمرّة (أو المتّصلة) خصوصاً فقد لا نكون قادرين على استخدام هذه الطريقة في العرض بسبب أنَّ قيمها تنتمي إلى مجموعات قد تكون غير قابلة للعدّ، وربما لا تكرَّر قيمة البيان لأكثر من مرَّةٍ واحدةٍ أيضاً. لذلك في مثل هذه الحالات لا يعود اهتمامنا مركّزاً على قيمة البيان نفسه وإنّما على الفترة التي ينتمي إليها هذا البيان، ومن ثمَّ يصبح اهتمامنا منصباً على بيان ممثِّل لكلِّ فترةٍ من الفترات التي ستصبّ فيها البيانات. بمعنى آخر، يكون لدينا تجزئة لمجموعة البيانات في فترات جزئية واهتمامنا يكون منصباً على قيمة ممثِّلة وحيدة (**تُدعى مركز الفئة – سيرد ذكرها لاحقاً**) لكل فترة من هذه الفترات الجزئية. عادة يُطلق اسم "**فئة** Class" على كل فترة جزئية من هذه الفترات.

إنَّ تمديد الجداول التكرارية إلى جداول تكرارية ذوات فئات يجعلها أكثر فاعلية في تنفيذ بعض العمليات الحسابية على البيانات وخاصّة عندما يصبح عدد البيانات كبيراً، فعلى سبيل المثال تصوّر لو أنَّك تقوم بدراسة على مجموعة بيانات مكوَّنة من مليون قيمة عددية أو أكثر فكم من الوقت ستحتاج لاستنباط سلوك هذه البيانات أو الحصول على بعض المميّزات العددية لها؟

قبل البدء في بناء جداول التوزيع التكرارية لا بدَّ لنا من تقديم بعض المفاهيم التي لا بدَّ منها لبناء مثل هذه الجداول، وسنبدأها بالتعريف الآتي.

1-3-1- تعريف (المدى Range)

لتكن لدينا x_1، x_2، … و x_n بيانات مُعطاة، ولنرمز لأصغر وأكبر قيمة في هذه البيانات بـ x_s و x_ℓ على الترتيب. عندئذٍ يُعرَّف **المدى** لهذه البيانات (**وسنرمز له بـ R**) على أنَّه الفرق بين أكبر وأصغر قيمة في هذه البيانات. أي إنَّه لدينا:

$$\mathbf{R} = x_\ell - x_s \qquad [1\text{-}2]$$

1-3-2- تعريف (الفئة Class)

الفئة من أجل بيانات مُعطاة هي فترة من مجموعة الأعداد الحقيقية لها طول موجب تماماً وتحتوي على بعضٍ من البيانات المُعطاة، ويقال عن طرفها الأيسر إنَّه الحد الأدنى للفئة في حين يُقال عن طرفها الأيمن إنَّه الحد الأعلى للفئة.

من التّسميات الأخرى للفئة Interval أو Category أو Group.

1-3-3- بناء جدول التوزيع التكراري

بالرجوع إلى موضوع تمديد الجداول التكرارية فإنَّ عملية تمديد هذه الجداول وفقاً للآلية التي سنقدِّمها بعد قليل تُعطينا ما يُعرف باسم "**جداول التوزيع التكرارية**"، ولهذه الجداول نماذج مختلفة ولكن معظم هذه الجداول تحتوي على الأعمدة المقدّمة في الجدول الآتي:

الجدول [1-7]

التكرار المتجمِّع الصاعد للفئة	التكرار المئوي للفئة	التكرار النسبي للفئة	تكرار الفئة	تعداد الفئة	مركز الفئة	الحدود الفعليّة للفئة	الحدود العمليّة للفئة	رقم الفئة
…	…	…	…	…	…	…	…	1
…	…	…	…	…	…	…	…	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
المجموع				------- -	------- -	-------- -	-------- -	Total

نشير هنا إلى أنَّ الفئات لجداول التوزيع التكرارية يمكن أن تُعرض (أو تُقدَّم) وفقاً لأحد نوعين من الفئات:

أ- جداول توزيع تكرارية تحتوي على فئات ذات أطوال مختلفة، وهذا النوع لن نقوم بدراسته في هذا الكتاب.

ب- جداول توزيع تكرارية تحتوي على فئات ذات أطوال متساوية، وهذا النوع من الجداول سيكون محور دراستنا في هذا الكتاب.

قبل البدء في كيفية بناء جدول التوزيع التكراري نودّ التنويه إلى أنَّنا سنشرح بناء جدول التوزيع التكراري من أجل الحالات البسيطة التي تكون فيها قيمة المدى للبيانات كبيرة نسبياً، والأمثلة التي سنقدّمها ستكون على قيم صحيحة للبيانات، وكذلك لن نتطرق إلى الحالات التي تستوجب معالجة خاصة في بناء جدول التوزيع التكراري.

الآن، ومن أجل بناء جدول توزيع تكراري نتبع الخطوات الآتية:

1 - تعيين عدد فئات جدول التوزيع التكراري:

إذا قدِّم لنا عدد الفئات (وليكن k فئة) الواجب استخدامها في **جداول التوزيع التكراري** من قبل الجهة الطالبة لدراسة المسألة الإحصائية، فإنّنا نلتزم بهذا العدد للفئات ولا نجري أيّ تعديل عليه إلاَّ لضـرورة تتطلّبها شـروط بناء جداول التوزيع التكراريـة. لكن إذا لم يقدَّم عـدد الفئـات الواجب دراسـتها فإنَّنـا نختار عدد الفئات k بحيث لا يقلُّ عن خمس فئات ولا يزيد عن عشرين فئة، وذلك لأنَّ جداول التوزيع التكرارية التي تحتـوي على أقلِّ من خمس فئـات تعدّ قليلة الفائـدة، وأمّا التي تحتوي على أكثـر من عشـرين فئة فإنَّهـا تكون متعبـة في الدراســة. إنَّ هذه التّوصيـة تتحقَّـق عمليّاً باسـتخدام

العلاقة [1–3] الآتية (مستخدمة في بعض البرامج الإحصائية) لحساب عدد الفئات k ما دام عدد المشاهدات (البيانات) n أكبر أو يساوي 32 مشاهدة (بطبيعة الحال من غير المرغوب تجميع البيانات في جداول توزيع تكرارية إذا كان عددها أقل من 32):

$$k = \lfloor 3.322 \log n \rfloor \qquad [1-3]$$

علماً أنَّ $\lfloor x \rfloor$ هو أكبر عدد صحيح أصغر أو يساوي x ، فعلى سبيل المثال $\lfloor 5 \rfloor = 5$ وكذلك لدينا:

$$\lfloor 3.322 \log 31 \rfloor = \lfloor 4.9543 \rfloor = 4 \qquad \& \qquad \lfloor 3.322 \log 32 \rfloor = \lfloor 5.00011 \rfloor = 5$$

$$\lfloor 3.322 \log 63 \rfloor = \lfloor 5.97741 \rfloor = 5 \qquad \& \qquad \lfloor 3.322 \log 64 \rfloor = \lfloor 6.00013 \rfloor = 6$$

2- تعيين سعة وحدود الفئات في جدول التوزيع التكراري:

بفرض أنَّه لدينا بيانات عددها n ولها مدى **R** وبعدد فئات k، فعندئذٍ نحصل على **سعة الفئة الفعليّة (وسنرمز لها بـ C)** باستخدام العلاقة الآتية:

$$\mathbf{C} = \frac{\mathbf{R} + \textit{one measuring unit}}{k} \qquad [1\text{-}4]$$

وننوّه هنا إلى أنَّ سعة الفئة النّاتجة عن الحساب السّابق يمكن تقريبها بالزّيادة قليلاً إلى قيمة أكبر بحيث تسمح لنا القيمة الجديدة للسعة بصبّ أسهل للبيانات أو تنسيق أفضل للفئات، ولكن يجب عدم المبالغة في زيادة السعة للفئات، وذلك لأنَّ الزيادة المبالغ فيها قد تؤدي إلى توليد فئات تصبح معها الفئات الأخيرة خارج نطاق البيانات.

3- تعيين حدود الفئات الفعليّة (أو الحقيقية) Class Boundaries:

الآن، وبعد تحديد السعة لكل الفئات الفعليّة **C**، فإنَّنا نقوم بتعيين حدود هذه الفئات كما يلي:

أ ـ نجعل الحد الأدنى لأول فئة فعليّة مساوياً للقيمة الناتجة عن طرح 0.5 من أصغر قيمة في البيانات.

ب- نضيف قيمة **C** إلى الحد الأدنى لهذه الفئة فنحصل على الحد الأعلى للفئة الفعليّة الأولى.

ج- نجعل الحد الأدنى للفئة الفعليّة التالية (**الثانية**) مساوياً للحد الأعلى للفئة الفعليّة السابقة (**الأولى**)، ومن ثمَّ نضيف قيمة **C** إلى الحد الأدنى لهذه الفئة فنحصل على الحد الأعلى للفئة الفعليّة الثانية.

د- نقوم بتطبيق الفقرة السابقة (ج) من أجل جميع الفئات المتبقية فنحصل على الحدود الدنيا والعليا للفئات الفعليّة لجدول التوزيع التكراري.

4- تعيين حدود الفئات العمليّة (أو التّجريبيّة) Class Limits:

يُعيَّن الحد الأدنى للفئة العمليّة ذات الرقم i من خلال إضافة نصف وحدة قياس إلى الحد الأدنى للفئة الفعليّة ذات الرقم i من أجل كل $i = 1, 2, \ldots, k$ (**لاحظ هنا أنَّ الحد الأدنى لأول فئة عمليّة سيوافق أصغر قيمة في البيانات المُعطاة**)، وأمَّا الحد الأعلى للفئة العمليّة ذات الرقم i فإنَّه يُعيَّن من خلال طرح نصف وحدة قياس من الحد الأعلى للفئة الفعليّة ذات الرقم i.

تجدر الإشارة هنا إلى الملاحظات الآتية:

أ- إنَّ عملية صبّ البيانات في جدول توزيع تكراري تتمّ في الفئات الفعليّة حصراً لأنَّه لدى عملية تحصيل البيانات سيكون لدينا خطأ مرتكب من وسائل القياس المعتمدة (**مهما بلغت من دقّة**) قد لا تصل إلى القيمة الحقيقية لقياس المشاهدة، ولذلك تمّ الاتفاق على أنَّ نصف وحدة الدقّة المعتمدة في القياس ستغطي هذا الخطأ زيادةً أو نقصاناً، بمعنى أنَّه بهذه العملية سيُغطي أكبر خطأ محتمل لدى أخذ البيانات، ولهذا السبب سمّيت هذه الفئات بالفئات الفعليّة.

ب- إذا وافقت قيمة x من قيم البيانات الحد الأدنى لفئة فعليّة فإنَّها تفرغ في هذه الفئة، وأمّا إذا وافقت هذه القيمة الحد الأعلى للفئة الفعليّة فإنّها تفرغ في الفئة الفعليّة التالية، ولهذا السبب سنكتب (**وعلى سبيل التوضيح**) الفئة الفعليّة التي حدّها الأدنى a وحدّها الأعلى b بالشكل $a \to b$ للدلالة على أنَّ القيمة b لا تتبع هذه الفئة وإنَّما تتبع الفئة الفعليّة التالية.

ج- وفقاً لاستخدام العلاقة [1-4] في تعيين سعة الفئة قد يحصل أنَّ قيمة x أو أكثر من قيم البيانات تساوي الحد الأعلى للفئة الفعليّة الأخيرة، فإذا وافقت القيم المتبقية الحد الأعلى للفئة الفعليّة الأخيرة فإنَّنا نقوم بتحميلها في الفئة الفعليّة الأخيرة.

د- من أجل فئة فعليّة حدّها الأدنى a وحدّها الأعلى b ستكون للفئة العمليّة الموافقة حدّ أدنى α قيمته تزيد على a بمقدار نصف وحدة الدّقة المعتمدة، وأمّا حدّها الأعلى β فإنَّه سينقص عن b بمقدار نصف وحدة الدّقة المعتمدة، ولذلك تُكتب الفئة العمليّة بالشكل $\alpha - \beta$ (**بدون سهم** $\rightarrow$) للدلالة على أنَّه من أجل هذا النوع من الفئات يكون كل من الحدّ الأدنى والأعلى منتمياً لها.

هـ- عندما يقدَّم جدول توزيع تكراري لمجموعة بيانات (**أي أنّ البيانات الخام التي صُبَّت فيه ليست ظاهرة – خفية**) فإنَّه يمكن الاستغناء عن العمود الخاص بالفئات العمليّة لعدم الحاجة له في أية دراسة لاحقة تتعلَّق بهذا الجدول.

و- إذا حصل لدينا فئات فعليّة خالية من البيانات عند عمليّة صبّ البيانات فإنَّنا نقوم بتغيير عدد الفئات (**وغالباً بزيادة عددها**) حتى تختفي جميع الفئات الخالية. بالطبع سيرافق هذا التغيير لعدد الفئات تغيُّر في سعتها أيضاً، ولذلك يجب الانتباه في حال التعديل على السعة الجديدة ألاّ يؤدي ذلك إلى فئات خالية من البيانات أيضاً.

4- تعيين مراكز الفئات Class Centers:

إنَّ مركز الفئة يساوي نصف مجموع حدّيها الأعلى والأدنى (**أية فئة كانت العمليّة أو الفعليّة**)، ويُنظر إلى هذه القيمة (**أي إلى مركز الفئة**) على أنَّها الممثِّل لكل البيانات التي ستنتمي إلى هذه الفئة، ولذلك سوف نلاحظ أنَّ لهذه القيمة دوراً مهمّاً جداً لدى حساب القيم العددية المميِّزة للبيانات المجمّعة في جداول توزيع تكرارية.

5- تدوين التّعدادات للفئات Class Tallies:

لقد قمنا سابقاً بشرح تدوين التعدادات من أجل الجداول التكرارية، وهنا يستخدم ذلك الشرح بآليّة مماثلة، ولكن هنا يرسم خط من أجل كل قيمة بيان تنتمي إلى الفئة الفعليّة (**وليس إلى الفئة العمليّة**)، وأخيراً نشير إلى أنَّ هذا العمود يستخدم عندما يكون لدينا عمليّة صبّ لبيانات خام (**بيانات كميّة**) في جدول توزيع تكراري، وفيما عدا ذلك لا ضرورة لهذا العمود ضمن جدول التوزيع التكراري حيث يمكن حذفه (**أو الاستغناء عنه**).

6- تعيين قيم التكرارات للفئات Frequencies of Classes:

لقد قمنا سابقاً بشرح تعيين قيم التكرارات للفئات من أجل الجداول التكرارية، ويستخدم ذلك الشرح من أجل جداول التوزيع التكرارية وفقاً للآلية نفسها.

7- تعيين قيم التكرارات النسبيّة للفئات Relative Frequencies of Classes:

لقد قمنا سابقاً بشرح طريقة تعيين قيم التكرارات النسبيّة للفئات من أجل الجداول التكرارية، وتستخدم تلك الطريقة نفسها من أجل جداول التوزيع التكرارية أيضاً.

8- تعيين قيم التكرارات المئوية للفئات Percentile Frequencies of Classes:

لقد قمنا سابقاً بشرح طريقة تعيين قيم التكرارات المئوية للفئات من أجل الجداول التكرارية، وتستخدم تلك الطريقة نفسها من أجل جداول التوزيع التكرارية أيضاً.

9- تعيين قيم التكرارات المتجمِّعة الصاعدة للفئات:

Ascending Cumulative Frequencies of Classes

العمود التالي والأخير من جدول التوزيع التكراري هو عمود التكرارات المتجمِّعة الصاعدة، وتدوَّن القيم في هذا العمود على النّحو الآتي:

مقابل الفئة الأولى يدوَّن تكرار الفئة الأولى نفسه (**ويُدعى التكرار المتجمِّع الصاعد للفئة الأولى**) لأنَّه يمثِّل كل التكرارات التي أقل من حدِّها الأعلى. أمَّا مقابل الفئة الثانية فيتمّ تدوين حاصل مجموع التكرارين للفئتين الأولى والثانية (**ويُدعى التكرار المتجمِّع الصاعد للفئة الثانية**) وهو يمثِّل كلّ التكرارات التي أقلّ من حدِّها الأعلى. يدوَّن مقابل الفئة الثّالثة حاصل مجموع التكرارات للفئات الأولى والثانية والثالثة (**ويُدعى التكرار المتجمِّع الصاعد للفئة الثالثة**) وهو يمثِّل كلّ التكرارات التي أقلّ من حدِّها الأعلى، وهكذا دواليك حتى الانتهاء من جميع الفئات. إنَّ القيم المدوَّنة في هذا العمود تُدعى التكرارات المتجمِّعة الصّاعدة.

1-3-4- مثال

البيانات الآتية تمثِّل عدد الأميال المقطوعة لكلّ لتر من البنزين لأربعين سيارة حديثة.

12	16	15	12	19	17	20	16	14	13
12	20	12	15	20	14	16	15	12	18
20	18	16	14	12	20	18	16	14	12
16	17	16	15	16	18	15	16	15	16

والمطلوب صبّ هذه البيانات في جدول توزيع تكراري.

الإجابة: بسبب كِبر نموذج جدول التوزيع التكراري (**كجدول نموذجي للتدريب**) سنقوم (**وعلى سبيل التوضيح**) بتجزئة الجدول إلى جدولين.

لدينا أكبر وأصغر قيمة في البيانات هما $x_\ell = 20$ و $x_s = 12$ على الترتيب، ومن ثمّ نجد باستخدام العلاقة [1-2] أنَّ المدى للبيانات المُعطاة يساوي:

$$\mathbf{R} = x_\ell - x_s = 20 - 12 = 8$$

والآن لتعيين عدد الفئات نلاحظ أنَّه لدينا 30 مشاهدة، ومن ثمَّ باستخدام العلاقة [1-3] يكون عدد الفئات لجدول التوزيع التكراري المطلوب هو: 5

$$k = \lfloor 3.\ 322 \,\mathbf{log}\, 40 \rfloor = \lfloor 5.322 \rfloor = 5$$

ومن ثمَّ بحسب العلاقة [1-4] تكون سعة الفئة لجدول التوزيع التكراري المطلوب هي:

$$\mathbf{C} = \frac{\mathbf{R}+1}{k} = \frac{9}{5} = 1.8 \approx 2$$

والآن باتّباع الخطوات التي قمنا بشرحها سابقاً فإنَّنا سنميِّز البيانات التي سيتمّ صبّها باستخدام لون خاص لكل فئة (**وذلك على سبيل التّوضيح فقط**) فيكون لدينا العرض الآتي للبيانات المُعدّة للصبّ:

12	**16**	**15**	**12**	**19**	**17**	**20**	**16**	**14**	**13**
12	**20**	**12**	**15**	**20**	**14**	**16**	**15**	**12**	**18**
20	**18**	**16**	**14**	**12**	**20**	**18**	**16**	**14**	**12**
16	**17**	**16**	**15**	**16**	**18**	**15**	**16**	**15**	**16**

وبصبّ هذه البيانات سنحصل على جدول التوزيع التكراري الآتي الذي يحتوي على الفئات الفعليّة، الفئات العمليّة، مراكز الفئات، التعدادات والتكرارات:

الجدول [1-9-a]

رقم الفئة	الحدود العمليّة للفئة	مركز الفئة	تعداد الفئة	تكرار الفئة
1	12 – 13	12.5	𝍷𝍷𝍷 𝍸	8
2	14 – 15	14.5	𝍸 𝍸	10
3	16 – 17	16.5	𝍷𝍷 𝍸 𝍸	12
4	18 – 19	18.5	𝍸	5
5	20 – 21	20.5	𝍸	5
Total	--------------	-----	------	40

وأمّا من أجل جدول التوزيع التكراري الذي يحوي التكرار النسبي، المئوي والمتجمّع الصاعد فإنَّنا سنكون بحاجة إلى عمود الفئات الفعليّة والتكرارات، فيكون لدينا الجدول الآتي.

الجدول [1-9-b]

رقم الفئة	الحدود العمليّة للفئة	تكرار الفئة	التكرار المتجمّع الصاعد الفئة
1	12 – 13	8	8
2	14 – 15	10	8+10=18
3	16 – 17	12	8+10+12=30
4	18 – 19	5	8+10+12+5=35
5	20 – 21	5	8+10+12+5+5=40
Total		40	المجموع

1-3-5- ملاحظة

بما أنَّ التّعدادات تستخدم عند صبّ البيانات فقط، فإنَّه إذا كانت قيم التكرارات معلومة فعندئذٍ يمكن استنتاج قيم التكرارات النسبيّة والتكرارات المئوية عند اللزوم، ولذلك تحذف الأعمدة الخاصّة بالتعدادات والتكرارات النّسبيّة والتكرارات المئويّة من جداول التوزيع التكرارية، وكذلك سـنلاحظ لاحقاً أنَّ العروض البيانية وحساب القيم العددية المميّزة لبيانات جداول التوزيع التكرارية تعتمد على الفئات الفعليّة فقط، ولذلك يحذف العمود الخاص بالفئات العمليّة من جداول التوزيع التكرارية مالم تكن هناك عملية صبّ لبيانات خام في الجدول. لذلك سنستخدم لاحقاً جداول توزيع تكرارية لها العرض المختزل الآتي:

الجدول [1-10]

رقم الفئة	الحدود الفعليّة للفئة	مركز الفئة	تكرار الفئة	التكرار المتجمّع الصاعد للفئة

1	$a_0 \to a_1$	x_1	f_1	$F_1 = f_1$
2	$a_1 \to a_2$	x_2	f_2	$F_2 = f_1 + f_2$
$\vdots$	$\vdots \;\vdots\vdots$	$\vdots$	$\vdots$	$\vdots \;\; \vdots \;\; \vdots \;\; \vdots$
k	$a_{k-1} \to a_k$	x_k	f_k	$F_k = f_1 + f_2 + \ldots + f_k$
Total	----------------	--------	$\sum f_i$	المجموع

فعلى سبيل التوضيح يُعرض جدول التوزيع التكراري لبيانات المثال السابق **بشكله المختزل** على النحو الآتي:

رقم الفئة	الحدود الفعليّة للفئة	مركز الفئة	تكرار الفئة	التكرار المتجمّع الصاعد للفئة
1	$11.5 \to 13.5$	12.5	8	8
2	$13.5 \to 15.5$	14.5	10	18
3	$15.5 \to 17.5$	16.5	12	30
4	$17.5 \to 19.5$	18.5	5	35
5	$19.5 \to 21.5$	20.5	5	40
Total	----------------	--------	40	المجموع

1-4- التّمثيلات البيانيّة لجداول التّوزيعات التّكراريّة:
Graphical Representations of Frequency Distributions Tables

في كثير من الحالات يكون التعامل مع جداول التوزيع التكرارية أمراً شاقاً، وخاصة إذا كان عدد الفئات كبيراً، لذلك يحاول المرء عرض نتائجه من خلال أشكال بيانية مناسبة يَسْهل معها فهم طبيعة وسلوك بيانات هذه الجداول. من النماذج المستخدمة في هذه العروض البيانية النموذج الآتي.

1-4-1- المدرَّج التكراري Histogram

ينظر إلى المدرجات التكرارية على أنّها من الأشكال البيانية المفيدة في تمثيل بيانات الجداول التكرارية، ويتميَّز بسهولته وبساطته. يتكوَّن المدرَّج التكراري من مستطيلات متلاصقة ترسم فوق الفئات الفعليّة للبيانات وبحيث يكون ارتفاع كل مستطيل متناسباً مع قيمة التكرار للفئة التي رُسم عليها، حيث تُمثِّل الفئات الفعليّة على المحور الأفقي (**المحور oX**)، وأمّا على المحور العمودي عليه (**أي المحور الرأسي oY**) فيتمّ تدوين قيم التكرارات للفئات الفعليّة.

في الواقع يُمكِّننا الرسم البياني للمدرَّج التكراري من معرفة شكل توزيع البيانات وانتشارها وأين تتمركز البيانات بشكل سريع وسهل، والمثال الآتي يوضِّح لنا ذلك.

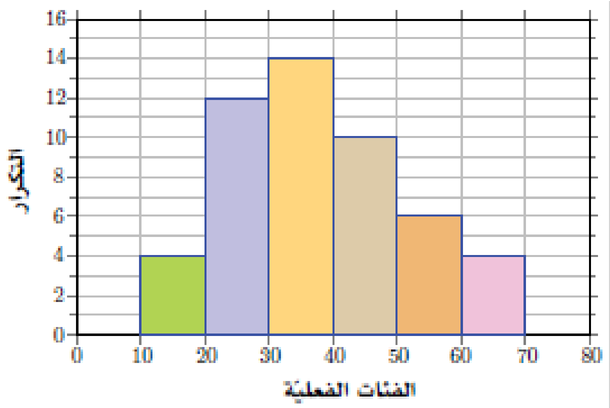
1-4-1-1- مثال

ليكن لدينا جدول التوزيع التكراري الآتي:

الجدول [1-11]

رقم الفئة	الحدود الفعليّة للفئة	مركز الفئة	تكرار الفئة	التكرار المتجمِّع الصاعد للفئة
1	$10 \rightarrow 20$	15	4	4
2	$20 \rightarrow 30$	25	12	16
3	$30 \rightarrow 40$	35	14	30
4	$40 \rightarrow 50$	45	10	40
5	$50 \rightarrow 60$	55	6	46
6	$60 \rightarrow 70$	65	4	50
Total	------------	------	50	المجموع

والآن لرسم المدرَّج التكراري نقوم برسم أعمدة مستطيلة فوق الفئات الفعليّة بحيث يكون ارتفاع كل مستطيل يساوي تكرار الفئة التي رُسم عليها، فنحصل على الشكل [1-9-a] للمدرَّج التكراري.

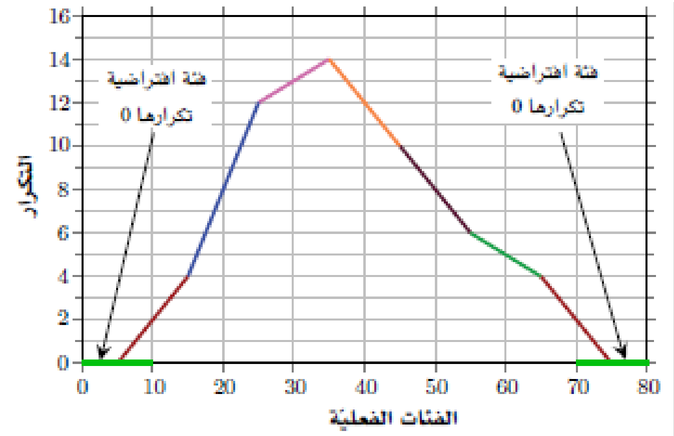


الشكل [1-9-a]

1-4-2- المضلَّع التّكراري Frequency Polygon

تعدّ المضلّعات التكرارية من الأشكال البيانية المهمّة في تمثيل البيانات الكميَّة أيضاً، وهو رسمٌ بياني مكوَّن من محورين متعامدين حيث تُمثَّل الفئات الفعليّة على المحور الأفقي (المحور oX) في حين تدوَّن قيم التكرارات للفئات الفعليّة على المحور الرّأسي (المحور oY)، ويتمّ تشكيل هذا المضلَّع من خــلال الوصل بقطع مستقيمة بين تلك النقاط التي إحداثياتها على محور الفئات هي مراكز الفئات، وأمّا إحداثياتها على محور التكرارات فهي قيم التكرارات المقابلة لتلك الفئات، وبعد ذلك إغلاق هذا المضلَّع الناتج إلى محور الفئات من خلال وصل بداية المضلَّع الناتج إلى مركز فئة افتراضية (وهمية) سابقة لأول فئة تكرارها معدوم، وبعد ذلك وصل نهاية هذا المضلَّع إلى مركز فئة افتراضية (وهمية) لاحقة بآخر فئة تكرارها معدوم أيضاً.

على سبيل المثال لو رجعنا إلى المثال السابق (1-4-1-1) نجد أنَّ المضلَّع التكراري لبيانات جدول التوزيع التكراري المُعطى له الشكل الآتي.



الشكل [1-9-b]

أخيراً نشير إلى أنَّه توجد حالات يغلق فيها المضلَّع النّاتج إلى بداية الفئة الأولى ونهاية الفئة الأخيرة إلاَّ أنَّنا لن نتطرق إلى النّقاش الخاص بهذه الحالة.

1-4-3- مضلَّع التّكرار المتجمِّع الصَّاعد

Ascending Cumulative Frequency Polygon (ACFP)

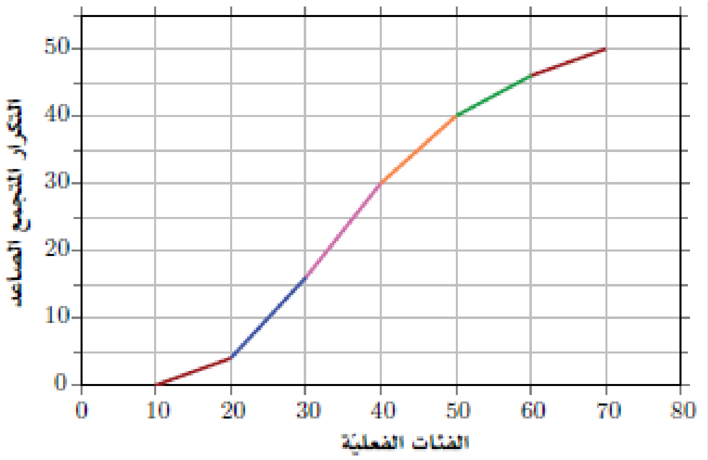
لرسم مضلَّع التّكرار المتجمِّع الصّاعد نقوم برسم محورين متعامدين، ومن ثمَّ يُمثَّل على المحور الأفقي الفئات الفعليّة في حين يُمثَّل على المحور الرّأسي التكرارات المتجمِّعة الصّاعدة، وبعد ذلك تُعيَّن مجموعة نقاط على النحو الآتي:

النقطة الأولى مسقطها على المحور الأفقي هو الحد الأعلى للفئة الفعليّة الأولى، ومسقطها على محور التكرارات يساوي قيمة التكرار المتجمِّع الصاعد للفئة الفعليّة الأولى.

النقطة الثانية مسقطها على المحور الأفقي هو الحد الأعلى للفئة الفعليّة الثانية، ومسقطها على محور التكرارات يساوي قيمة التكرار المتجمِّع الصاعد للفئة الفعليّة الثانية.

و هكذا دواليك حتى يتمّ تعيين النقطة الأخيرة التي مسقطها على المحور الأفقي هو الحد الأعلى للفئة الفعليّة الأخيرة، ومسقطها على محور التكرارات يساوي قيمة التكرار المتجمِّع الصاعد للفئة الفعليّة الأخيرة. بعد ذلك نصِل بقطع مستقيمة بين النقاط التي تمَّ تعيينها آنفاً، وأخيراً نقوم بإغلاق بداية الضلَّع الناتج إلى بداية الفئة الأولى فقط دون إغلاقه من اليمين (**أي إنَّه يغلق من طرف واحد فقط**).

على سبيل المثال لو رجعنا إلى المثال (1-4-1-1) فإنَّنا نجد لمضلَّع التكرار المتجمِّع الصاعد لتلك البيانات الشكل الآتي.



الشكل [1-9-c]

1-4-4- المنحنى التّكراري Frequency Curve

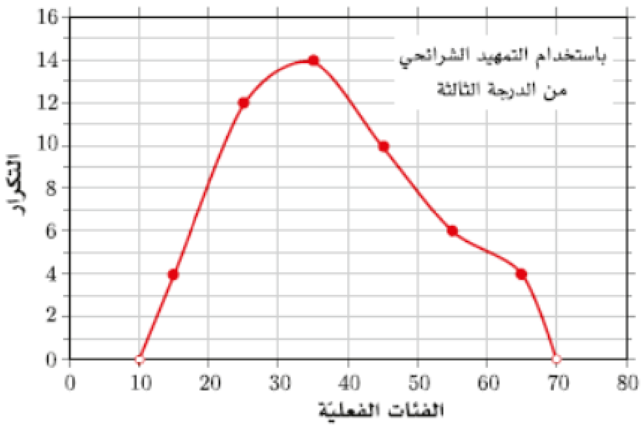
إنَّ هذه الطريقة تماثل طريقة العرض باستخدام المضلّعات التكرارية من حيث تعيين النّقاط التي سيمرّ منها الرّسم البياني، ولكنَّ الإغلاق من اليسار يكون إلى بداية الفئة الفعليّة الأولى، في حين يكون الإغلاق من اليمين إلى نهاية الفئة الفعليّة الأخيرة، ونحصل على هذا المنحنى من خلال تمهيد الخط المنكسر ليصبح منحنياً وفقاً لإحدى الطرائق الآتية:

1- جعل الفترات التي تبنى عليها هذا الخطوط المنكسرة قصيرة جداً مع زيادة عددها بشكل كبير جداً، ولذلك فإنَّ هذه الطريقة نادرة الاستخدام في المجالات التّطبيقية.

2- التمهيد اليدوي للخطوط المستقيمة لتصبح منحنية وملائمةً عند مرورها بالنقاط الممثِّلة للبيانات، وهذه الطريقة قليلة الاستخدام أيضاً لأنَّها تحتاج إلى مهارة في الرسم.

3- استخدام التمهيد الشرائحي Spline من الدرجة الثانية أو الثالثة، ومن ثمَّ النظر أيهما أنسب للعرض وأقل تناقضاً مع واقع البيانات، وهذه الطريقة تعطينا نتائج ممتازة لعمليّة التمهيد وهي متوفّرة في بعض البرامج الإحصائية، ومنها برنامج Curve Expert.

على سبيل المثال لو رجعنا إلى المثال السابق (1-4-1-1) نجد أنَّ المنحنى التكراري لبيانات جدول التوزيع التكراري المُعطى له الشكل الآتي.



الشكل [1-9-d]

1-3-4- مثال مثال توضيحي اخر:

يكن لدينا العينة العشوائية من مجتمع إحصائي أعمار 45 شخصا لأقرب سنة كما يلي:

45	54	48	35	63	19	65	41	27	51	48	32	58	30	61
39	73	36	41	36	60	52	46	40	64	40	53	24	57	41
33	42	38	51	53	35	44	41	49	55	68	57	48	45	29

المدى المطلق = أكبر قيمة - أصغر قيمة = 73 - 19 = 54.

البيانات غير مرتبة، فلا بد من تصنيفها وترتيبها وفق شكل يعطي معنى أدق ويمكن أن نبدأ في تجميع الأعمار في فئات نختارها بشكل ملائم، فمن أجل ذلك نجد:

أ- المدى المطلق للبيانات:

لدينا أكبر قيمة في هذا الجدول هي 73 وأصغر قيمة هي 19، والفرق بينهما نسميه بالمدى المطلق:

73 - 19 = 54

ب- نحدد عدد الفئات وطول كل فئة:

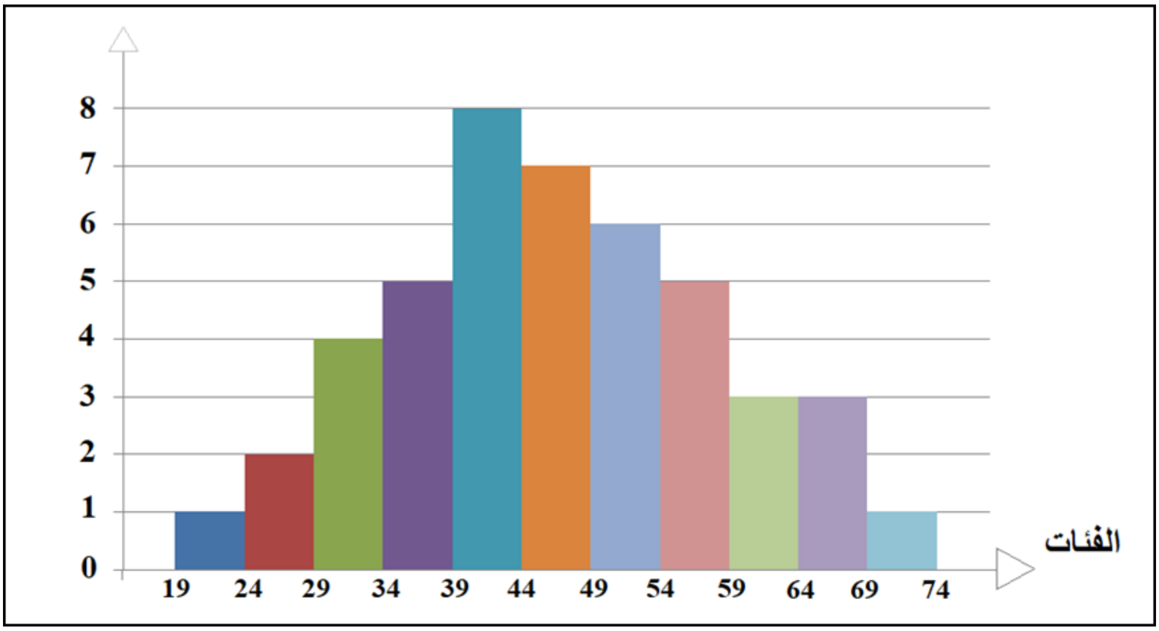
طول الفئة (عدد صحيح دوما) = "المدى المطلق" /"عدد الفئات" = 54/11 = 4.9 ≈ 5.

عدد الفئات = 11 ، اختياري حسب المطلوب.

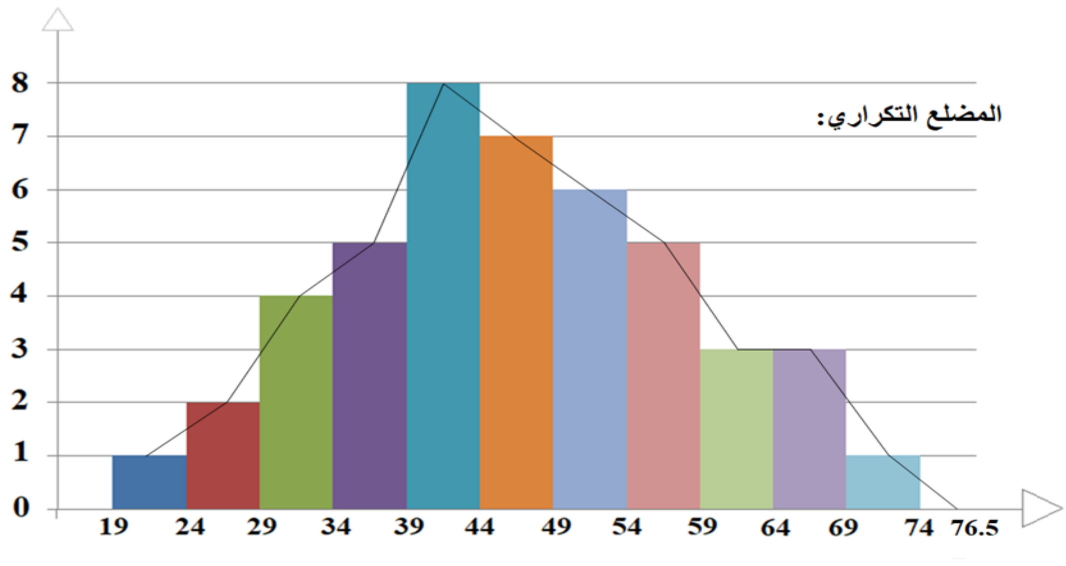
ويكون جدول التوزيع التكراري:

التكرار	مجال الفئة	الفئة
1	[19،24[	1
2	[24،29[	2
4	[29،34[	3
5	[34،39[	4
8	[39،44[	5
7	[44،49[	6
6	[49،54[	7
5	[54،59[	8
3	[59،64[	9
3	[64،69[	10
1	[69،74[	11

ويكون المنحني التكراري (طريقة المستطيلات



و المضلع التكراري (الخط المنكسر):

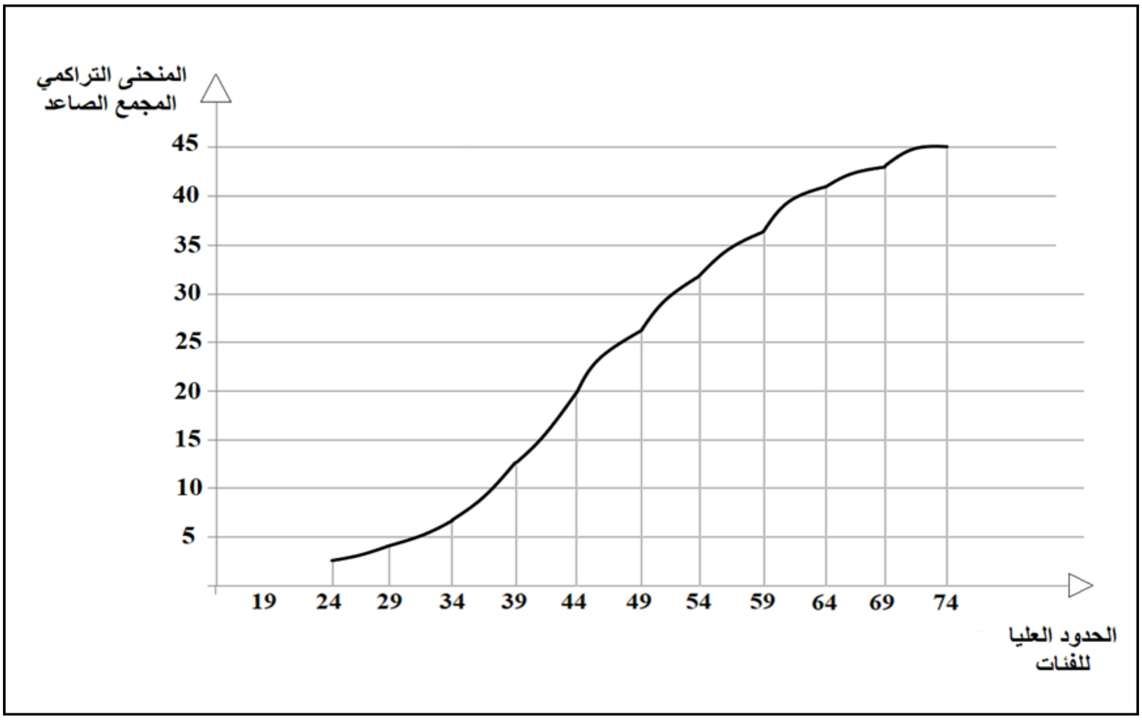


والتوزيع التكراري التراكمي :

الحدود العليا للفئات	التكرار fi	التكرار التراكمي fi

التكرار التراكمي fi	التكرار fi	الحدود العليا للفئات
1	1	24
3	2	29
7	4	34
12	5	39
20	8	44
27	7	49
33	6	54
38	5	59
41	3	64
44	3	69
45	1	74

ورسمه البياني:



تمرين1:

لدينا درجات 30 طالب في امتحان مادة مبادئ الإحصاء لطلبة السنة الأولى كما في الجدول التالي:

39	54	48	59	45	34	58	40	62	37
48	54	75	46	49	48	58	54	50	56
43	57	41	45	47	61	49	44	68	54

المطلوب:

1- اكتب جدول المدرج التكراري

2- ارسم المدرج التكراري

3- ارسم المضلع التكراري

4- ارسم المنحنى التكراري

5- ارسم المنحنى التكراري التراكمي الصاعد

الحل

1) **طول الفئة** $= \frac{\text{المدى المطلق}}{\text{عدد الفئات}} = \frac{\mathbf{41}}{\mathbf{7}} \cong 6$

المدى المطلق $\mathbf{75 - 34 = 41}$

الفئة	مجال الفئة	التكرار
1	[34.40[	3
2	[40.46[	6
3	[46.52[	8
4	[52.58[	5
5	[58.64[	6
6	[64.70[	1
7	[70.76[	1

تمرين2:

للبيانات التالية تمثل أجوًا يومية لــ (40) عامل في مصنع.

110-115-117-113-145-137-200-117-113-250

225-115-145-113-230-225-117-250-113-200

113-175-180-185-190-195-240-245-237-248

148-137-188-194-165-167-173-209-213-219

والمطلوب:

1- اكتب جدول التوزيع التكراري
2- ارسم المدرج التكراري
3- ارسم المضلع التكراري
4- اكتب جدول التوزيع التكراري التراكمي ومثله بيانًا.

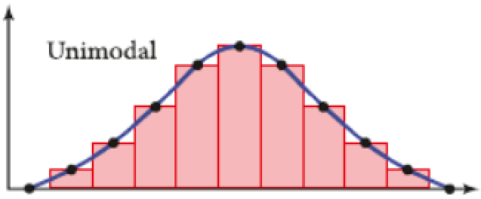
1-5- أشكال التّوزيعات التّكراريّة
Shapes of Frequency Distributions

في الواقع توجد أشكال عديدة للتوزيعات التكرارية، وشكلها يتبع طبيعة انتشار البيانات ونزوعها نحو قيمة أو موضع ما، وهذه الأشكال تعطي دلالات هامّة عن المؤثِّرات التي خضعت لها البيانات. في هذه الجزئية سوف لن نخوض بعيداً في هذه المسائل ولكن سنعرض بشكل موجز وبسيط أهم التوصيفات التي تذكر بها الأشكال البيانية للتوزيعات التكرارية، وسنبدأها بالفقرة الآتية.

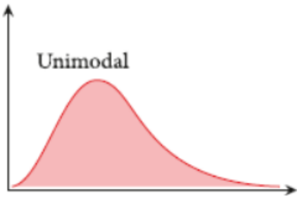
1-5-1- التوزيعات ذات المناويل Distributions that have Modes

يمكن للمرء التمييز بين نوعين من التوزيعات هما:

1- توزيعات تملك قمّة (أو ذروة) واحدة، وهذه التوزيعات تُدعى توزيعات أحادية المنوال، والعرضين الآتيين يوضِّحان ذلك.

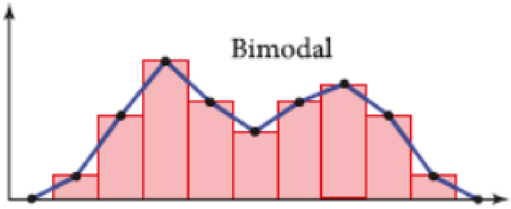


الشكل [1-10-b]

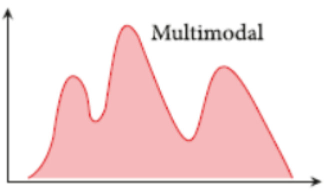


الشكل [1-10-a]

2- توزيعات تملك أكثر من قمّة (أو ذروة)، وهذه التوزيعات تُدعى توزيعات متعددة المناويل.



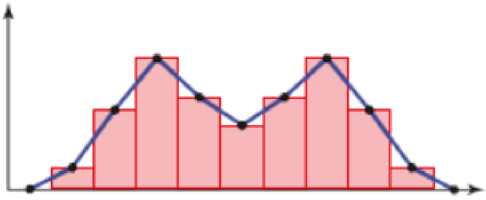
الشكل [1-11-b]



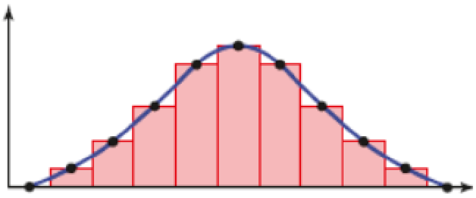
الشكل [1-11-a]

1-5-2- الالتواء والتناظر لتوزيع تكراري
Skewness and Symmetry of Frequency Distributions

يُقال عن توزيع تكراري إنَّه متناظر إذا انطبق على نفسه تمام الانطباق لدى طيّه على محورٍ مارٍ من منتصف قاعدته، وفي حال عدم تحقّق الانطباق التام فإنَّه يُقال عن التوزيع التكراري إنَّه ملتوٍ، والعروض الآتية توضِّح لنا ذلك.

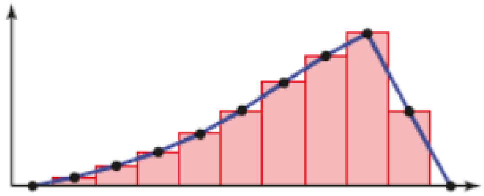


الشكل [1-12-b]

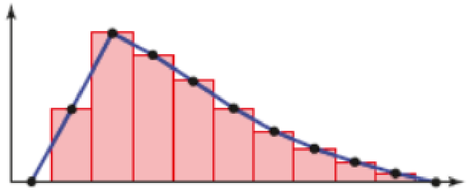


الشكل [1-12-a]

توزيعان تكراريان متناظران (أو متماثلان) Symmetric Frequency Distributions



توزيع تكراري ملتو نحو اليسار
الشكل [1-13-b]



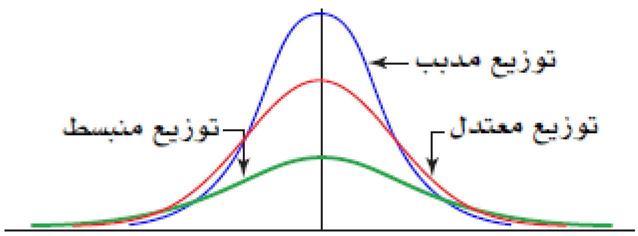
توزيع تكراري ملتو نحو اليمين
الشكل [1-13-a]

1-5-2- التدبب والتفلطح لتوزيع تكراري
Sputtering and Flattening of Frequency Distributions

تصنَّف التوزيعات التكرارية في ثلاثة أنواع (**انظر الشكل الآتي [1-14]**) هي:

أ- التوزيعات **المدببة**   ب- التوزيعات **المعتدلة**   ج- التوزيعات **المنبسطة**

علماً أنَّ مفهوم الاعتدال للتوزيعات بُنِيَّ على مدى تقارب شكله من شكل دالَّة الكثافة للتوزيع الطبيعي المعيَّاري (**سنأتي على ذكر دالَّة الكثافة للتوزيع الطبيعي المعيَّاري في الفصل الأخير**)، وأمّا من أجل الحكم على تدبب أو تفلطح توزيع تكراري ما فإنَّنا نحتاج لمعيَّار مُحدّد لن نتطرق إليه هنا.



الشكل [1-14]

1-5-3- تفسير شكل التوزيع

في الواقع يمكن تقديم بعض التفسيرات لأشكال التوزيعات، ومنها:

1- إذا كان للتوزيع التكراري شكل متناظر معتدل فإنَّ ذلك يعني أنَّ البيانات تتوزع طبيعياً على وجه التقريب، وأنَّ الأخطاء المرتكبة لدى عمليّة القياس هي على الغالب أخطاء عشوائية (**غير متعمَّدة**).

2- إذا كان للتوزيع التكراري عدَّة مناويل فإنَّ ذلك يدلِّل على وجود عدَّة أسباب فاعلة ومؤثِّرة في التجربة (**أو المسألة**) المولّدة للبيانات، وعادة يكون عدد هذه الأسباب الفاعلة والمؤثِّرة مساوياً لعدد المناويل في التوزيع.

3- إذا كان للتوزيع التكراري شكل ملتوٍ نحو اليمين فإنَّ ذلك يعني على الغالب تنفيذ عمليات فرز بشأن البيانات بحيث يستثنى منها البيانات التي تقل عن مقدار محدَّد تفرضه طبيعة الدراسة الإحصائية، وأمَّا إذا كان شكل التوزيع ملتوٍ نحو اليسار فإنَّ ذلك يعني على الغالب تنفيذ عمليات فرز بشأن البيانات بحيث يستثنى منها البيانات التي تزيد عن مقدار محدَّد تفرضه طبيعة الدراسة الإحصائية.

4- إذا كان شكل التوزيع التكراري منبسطاً فإنَّ ذلك يعني على الغالب تنفيذ عمليات فرزٍ بشأن البيانات بحيث يُستثنى منها البيانات التي تقل وتزيد عن قيمة محدَّدة تفرضها طبيعة الدراسة الإحصائية، وكذلك يدلِّل على أنَّ البيانات تتبعثر بعيداً عن قيمة المتوسط لهذه البيانات وذلك تبعاً لمقدار تفرطح (**أو تفلطح**) هذا التوزيع.

تمارين الفصل الأول

1- وضح الفرق بين العيّنة الإحصائية والمجتمع.

2- لماذا نحتاج إلى تبويب (أو صبّ) البيانات في جداول تكرارية، وضح ذلك بالتفصيل.

3- صنف المتغيّرات الآتية من حيث كونها متغيّرات كميَّة أو متغيّرات نوعيّة:
أ- عدد الطلاب في جامعة الملك فهد للبترول والمعادن.
ب- فصيلة الدم لعيّنة مكونة من 50 طالباً.
ج- أنواع التمور في مزرعة معيّنة.
د- جنسية مجموعة من الأشخاص يعيشون بالمملكة العربية السعودية.
هـ- تقديرات مجموعة من الطلاب في مقرَّر الفيزياء.
و- عدد الكيلومترات التي يقطعها الطلاب للذهاب إلى جامعة الملك سعود.
ز- أعمار الطلاب في جامعة الملك عبد العزيز.
ح- عدد الوجبات الغذائية المقدَّمة في مطعم معيَّن في يوم ما.
ك- مستوى الخدمة المقدَّم من فندق معيَّن في مكة المكرمة.

4- حدد فيما إذا كانت العبارات الآتية صحيحة أم خاطئة:
أ- العمر يعدّ مثالاً على متغيّر نوعي.
ب- عدد الكليات في جامعة الملك سعود يعدّ متغيّراً متقطِّعاً.
ج-عدد الدقائق التي يقطعها الطالب للوصول إلى جامعة الإمام محمد بن سعود الإسلامية تعدّ متغيّراً متّستمراً.
د- لا يوجد علاقة بين المجتمع الإحصائي والعيّنة الإحصائية.
هـ- المتغيّر الذي يمكن تمثيل قيمه على شكل فترات يسمى متغيّر نوعي.

5- عدد أنواع البيانات الإحصائية، ومن ثمَّ اذكر مثالاً على كل نوع منها.

6- ليكن لدينا جدول التوزيع التكراري الآتي:

رقم الفئة	الفئات الفعليّة	مراكز الفئات	التكرار	التكرار المتجمع الصاعد
1	$10 \rightarrow 14$	12	2	2
2	$14 \rightarrow 18$	16	4	6
3	$18 \rightarrow 22$	20	8	14
4	$22 \rightarrow 26$	24	16	30
5	$26 \rightarrow 30$	28	10	40
Total			40	

والمطلوب ما يلي:
أ- رسم المدرَّج التكراري لبيانات هذا الجدول.
ب- رسم المضلَّع التكراري لبيانات هذا الجدول.
ج- رسم مضلَّع التكرار المتجمِّع الصاعد لبيانات هذا الجدول.
د- رسم المنحنى التكراري لبيانات هذا الجدول مستخدماً التمهيد باليد.

7- ليكن لدينا جدول التوزيع التكراري الآتي:

الفئات الفعليّة	$6 \to 9$	$9 \to 12$	$12 \to 15$	$15 \to 18$	$18 \to 21$	$21 \to 24$	المجموع
التكرار	4		16			4	**50**
التكرار النسبي		0.12					
التكرار المئوي				24 %			
التكرار المتجمع الصاعد					46		

والمطلوب إكمال مُعطيات هذا الجدول.

-8 أخذت عيّنة مكونة من 100 شخص وتمّ تصنيفهم حسب عدد مرات الذّهاب لنادي رياضي معيّن خلال شهرٍ، فكانت النتائج كما في الجدول التكراري الآتي:

عدد الأشخاص	عدد مرات الذهاب إلى النادي
23	أقل من أربع مرّات
40	ما بين أربع إلى سبع مرّات
28	ما بين ثمان إلى إحدى عشرة مرّة
6	ما بين اثنتي عشرة إلى ثلاث عشرة مرّة
3	أكثر من أربع عشرة مرّة
100	**Total**

والمطلوب ما يلي:
أ- قم بتمديد هذا الجدول بحيث يحتوي على التكرارات النسبيّة والمئويّة.
ب- ما هي نسبة الأشخاص الذين يذهبون إلى النّادي سبع مرّات على الأكثر خلال الشّهر؟

-9 أخذت عيّنة عشوائية من إحدى المدارس مكوّنة من 50 طالباً، وقُرأت قيمة الوزن لكل واحدٍ منهم (**مقدَّرةً بالكيلوجرام**)، فكانت لدينا النتائج الآتية:

28	24	25	26	33	33	23	27	34	30
21	28	23	27	33	27	22	31	26	21
27	31	35	23	25	26	22	34	35	31
27	31	35	23	33	27	22	27	34	30
27	22	23	27	26	33	33	26	22	34

والمطلوب ما يلي:
أ- صبُّ هذه البيانات في جدول توزيع تكراري.

ب- رسم المدرَّج التكراري، المضلَّع التكراري ومضلَّع التكرار المتجمّع الصاعد لبيانات هذا الجدول.

ج- باستخدام مضلَّع التكرار المتجمّع الصاعد للبيانات قدِّر التكرار المتجمّع الصاعد للقيمة 25.5.

د- حدِّد شكل التوزيع من حيث كونه متماثلاً أم لا.

-10 البيانات الآتية تُمثِّل عيِّنة نتائج استبيانات مُصنَّفة في ثلاثة أنواع A، B وC كما يلي:

B	A	C	C	A	C	B	A	C	B
C	C	C	B	C	B	C	B	C	C
B	C	C	C	A	B	C	C	B	C

والمطلوب ما يلي:

أ- صبُّ هذه البيانات في جدول تكراري تظهر فيه التكرارات، التكرارات النسبيّة والتكرارات المئوية.

ب- تقديم العرض الشرائطي (**بالأعمدة**) لهذه البيانات.

11- أُخِذت عيّنة مكوَّنة من 80 شخصاً، وسُئل كلُّ واحدٍ منهم عن آخر مؤهلٍ علميٍّ حصل عليه فكانت النّتائج موضحّة كما في الجدول الآتي:

مستوى التعليم	ابتدائي	متوسطة	ثانوي	جامعي	دراسات عليا	المجموع
عدد الأشخاص	10	16	18	30	6	**80**

والمطلوب تمثيل بيانات هذا الجدول باستخدام القطاعات الدّائريّة.

12- من أحد الحقول الزّراعيّة تم أخذ عيّنة مكونة من 40 زهرة وتمّ تصنيف هذه الزّهور بحسب اللون، فكانت لدينا النتائج الآتية:

أصفر	أحمر	أبيض	أصفر	أصفر	أحمر	أبيض	بنفسجي
أصفر	أبيض	أحمر	بنفسجي	أصفر	أصفر	أحمر	بنفسجي
أحمر	أصفر	أبيض	أصفر	بنفسجي	أحمر	أبيض	أصفر
أصفر	أحمر	بنفسجي	أحمر	أصفر	أحمر	أبيض	أبيض
أصفر	أبيض	أصفر	بنفسجي	أصفر	أحمر	أصفر	بنفسجي

والمطلوب ما يلي:

أ- صبُّ هذه البيانات في جدولٍ تكراريٍّ يظهر فيه التكراري النسبي والمئوي لهذه البيانات.

ب- تمثيل هذه البيانات بالعرض الشّرائطي (**بالأشرطة الأفقية**).

ج- تمثيل هذه البيانات بالقرص الدّائري وموضِحاً قيم زوايا القطاعات الدائرية الناتجة.

د- ما هي نسبة الزّهور غير الحمراء؟

13- الجدول الآتي يوضّح وقت الانتظار (**بالدقائق**) في طوارئ أحد المستشفيات لعيّنةٍ من المرضى.

وقت الانتظار	عدد المرضى
$0 \rightarrow 7$	5
$7 \rightarrow 14$	15
$14 \rightarrow 21$	20
$21 \rightarrow 28$	15
$28 \rightarrow 35$	5
Total	

بناءً على معطيات الجدول السّابق اختر الإجابة الصحيحة لكلٍ من العبارات الآتية:

أ- سعة (**أو طول**) الفئة لبيانات الجدول هي: 8 ، 7 ، 6 أو 5.

ب- مركز الفئة الثالثة هو: 15.5 ، 16.5 ، 17.5 أو 19.5.

ج- الحد الأدنى للفئة العمليّة الثانية هو: 6 ، 6.5 ، 7 أو 7.5.

د- الحد الأعلى للفئة العمليّة الثالثة هو: 19.5 ، 20 ، 20.5 أو 21.5.
هـ- التكرار النسبي للفئة الرابعة هو: 0.20 ، 0.22 ، 0.25 أو 0.27.
و- حجم العينة هو: 48 ، 58 ، 60 أو 65.
ز- التكرار المئوي للفئة الأخيرة هو: 7.8 ، 8.3 ، 8.5 أو 9.2.
ح- التكرار المتجمّع الصاعد للفئة الأولى هو: 0 ، 5 ، 15 أو 20.
ط- توزيع هذه البيانات: ملتوٍ نحو اليمين، ملتوٍ نحو اليسار أم متناظر.
ك- توزيع هذه البيانات: أحادي المنوال، ثنائي المنوال أم لا منوال له.

14- الجدول الآتي يعطينا مبيعات شركتين A وB (مقدَّرة بالملايين) خلال خمسة أشهر.

الشهر	مبيعات الشركة A	مبيعات الشركة B
يناير	5.6	3.8
فبراير	4.7	3.3
مارس	3.9	2.8
إبريل	4.5	3.5
مايو	6.1	4.2

والمطلوب ما يلي:

أ- مثّل بيانات مبيعات الشركة A باستخدام الشرائط الأفقية.
ب- مثّل البيانات السابقة باستخدام الأعمدة البيانية المزدوجة.

15- الجدول الآتي يمثّل مبيعات أحد المقاهي بالرّيال خلال فترتين من اليوم ولمدّة أسبوع كامل.

اليوم	مبيعات الفترة الأولى	مبيعات الفترة الثانية	المبيعات الكلية
السبت	530	150	**680**
الاحد	670	270	**940**
الإثنين	550	160	**710**
الثلاثاء	630	250	**880**
الأربعاء	620	230	**850**
الخميس	740	490	**1230**
الجمعة	850	530	**1380**

والمطلوب ما يلي:

أ- مثّل بيانات المبيعات الكلية باستخدام الشرائط العمودية.

ب- مثّل البيانات السابقة باستخدام الأعمدة البيانية المزدوجة.

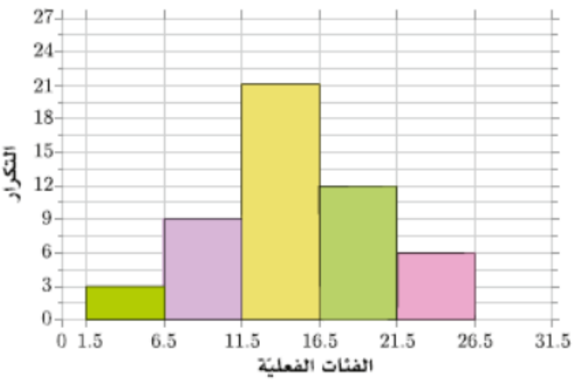
16- البيانات الآتية تمثّل عدد الكيلومترات التي يقطعها أربعون مهندساً للوصول إلى مقر عملهم.

31	12	7	13	11	25	20	10	3	5
2	16	17	32	11	18	17	12	10	19
3	18	15	12	5	3	8	7	9	7
12	6	7	15	15	6	9	2	14	12

والمطلوب ما يلي:

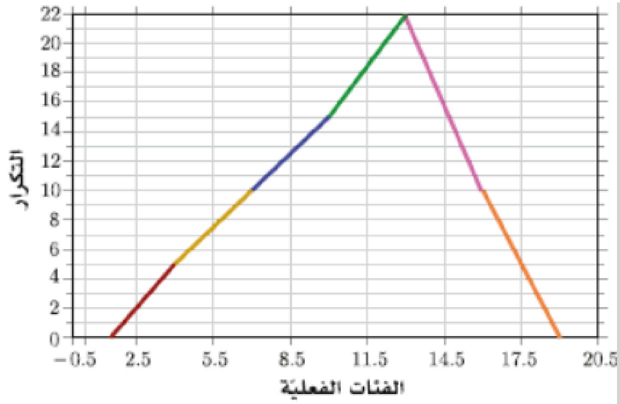
أ- مثّل البيانات المعطاة باستخدام العرض النقطي.
ب- صبُّ هذه البيانات في جدولِ توزيع تكراري تكون فيه الفئة الفعليّة الأولى هي $0 \rightarrow 7$.
ج- كم عدد المهندسين الذين يقطعون 21 كيلو متر على الأكثر للوصول إلى مقرِّ عملهم؟
د- كم عدد المهندسين الذين يقطعون 14 كيلو متر على الأقل للوصول إلى مقرِّ عملهم؟

17- لتكن لدينا بيانات مقدَّمة من خلال العرض البياني الآتي:



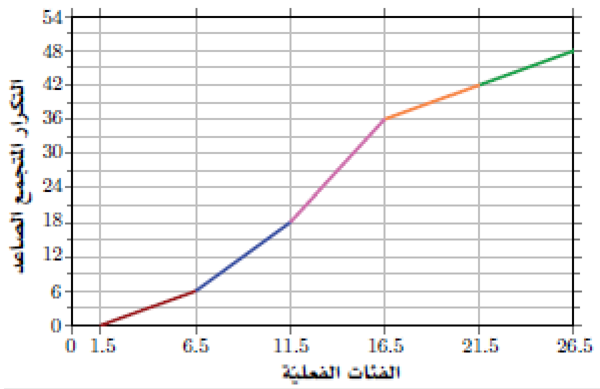
والمطلوب ما يلي:
أ- قم ببناء جدول التوزيع التكراري لهذه البيانات.
ب- ارسم المضلَّع التكراري ومضلَّع التكرار المتجمِّع الصاعد لهذه البيانات.
ج- هل يوحي هذا الشكل إلى أنَّ توزيع البيانات ملتوٍ، ولماذا؟

18- لتكن لدينا بيانات مقدَّمة من خلال العرض البياني الآتي:



والمطلوب ما يلي:
أ- قم ببناء جدول التوزيع التكراري لهذه البيانات.
ب- ارسم المدرَّج التكراري ومضلَّع التكرار المتجمِّع الصاعد لهذه البيانات.
ج- إلى أية جهةٍ يلتوي فيها شكل توزيع البيانات؟

19- لتكن لدينا بيانات مقدَّمة من خلال العرض البياني الآتي:



والمطلوب ما يلي:

أ- قم ببناء جدول التوزيع التكراري لهذه البيانات.

ب- ارسم المدرَّج التكراري والمضلَّع التكراري لهذه البيانات.

ج- هل يوحي لك هذا الشكل إلى وجود التواء في توزيع البيانات، ولماذا

الفصل الثاني

مقاييس الموضع للبيانات

Location Measures of Data

مقدِّمة:

لقد لاحظنا أنَّ طرائق العرض البيانية لها أهمية خاصة عند تقديم المعلومات الإحصائية فهي تُعطي وصفاً عاماً وسريعاً للبيانات الإحصائية، إلاَّ أنَّ فوائدها الاستقرائية تبقى قليلة، فإذا نظرنا إلى المضلَّع التكراري لمجموعة بيانات عينة فإنَّه يعطينا تصوراً عن شكل المضلَّع التكراري لهذه البيانات، واستقراؤنا له يقف عند الفرض أنَّه يوجد تشابه ما بين المضلَّع الممثِّل لهذه البيانات والمضلَّع الممثِّل لمجتمع البيانات التي أُخذت منه هذه العينة، ولذلك كان لا بدَّ من تقديم معايير عدديّة تحدَّد لنا وبدّقة سلوك البيانات. من المعايير التي تقوم بهذه المهمّة ما يُعرَّف باسم "**مقاييس الموضع**" وعلى وجه الخصوص مقاييس النزعة المركزيّة (**والتي تعدّ جزأً من مقاييس الموضع**).

2-1- مقاييس النزعة المركزية

2-2- الربيعيَّات

2-3- المئينات

2-4- الأعداد الخمسة والتمثيل الصندوقي للبيانات

2-1- مقاييس النزعة المركزيّة

Measures of Central Tendency

في الواقع لكلِّ مجتمع إحصائي مَعَالم Parameters تميّزه، وعادة تكون هذه المَعَالم مجهولةً كلياً أو جزئياً، ولكي نتمكن من دراسة المجتمع الإحصائي يجب علينا تقدير مَعَالمه المجهولة. إنَّ الوسائط التي تقوم بهذه المهمّة تُدعى الإحصاءات (**وسنستخدم كمفردة لها كلمة "إحصاءة" Statistic**)، وهذه الإحصاءات تتعامل مع بيانات العينات مباشرة. لذلك لا بدّ لنا أن نتعرَّف أولاً على مفهومي المَعْلَمة والإحصاءة.

2-1-1- تعريف (المَعْلَمة Parameter)

المَعْلَمة هي قيمةٍ عدديةٍ تُميِّز المجتمع (**ومنها على سبيل المثال: المتوسط والانحراف المعياري للمجتمع- سنأتي على ذكرهما لاحقاً-**)، وهذه القيمة تكون على الغالب مجهولة ويجب تقديرها من بيانات عينة تسحب من هذا المجتمع.

2-1-2- تعريف (الإحصاءة Statistic)

الإحصاءة هي قيمةٍ عدديةٍ تُميّز العينة (**ومنها على سبيل المثال: المتوسط والانحراف المعياري للعينة**)، وتحسب هذه القيمة من بيانات العينة، ومن ثمَّ فهي قيمة يمكن الحصول عليها بالحساب، وبالتالي يُنظر إليها على أنَّها قيمة معلومة وتستخدم كقيمةٍ تقريبيةٍ للمَعْلَمة.

الآن بالعودة إلى الحوار السّابق فإنَّ المشكلة التي سنقف عندها هي كيفية قياس درجة الاختلاف بين قيمة مَعْلَمة المجتمع وما يقابلها من قيمة الإحصاءة التي ستستخدم كتقدير لقيمة هذه المَعْلَمة، ولهذا السبب كان لا بدَّ من وجود مقاييس كميَّة تُمكّننا من تقدير جيد لمَعَالم المجتمع الإحصائي.

فيما يلي سنوجّه اهتمامنا على استخراج قيمة أو أكثر من مجموعة البيانات للاستدلال من خلالها على حقائق الظاهرة التي تمثِّلها مجموعة البيانات ككلّ، وهذا يتطلب منّا الآتي:

1- البحث عن قيمة عددية (**أو قيم عددية**) تُمثِّل مركز جذب لهذه البيانات، بمعنى آخر، تبدو البيانات كلّها أو بعضها ينزع نحو هذه القيمة (**أو نحو هذه القيم**)، وهذا المبحث سيكون محور هذا الفصل من الكتاب.

2- البحث عن قيمة عددية توضِّح مدى تبعثر قيم البيانات عن تلك القيمة التي تنزع إليها البيانات، وهذا المبحث سيكون محور الفصل الثالث من هذا الكتاب.

3- البحث عن قيم عددية تدل على شكل التوزيع من حيث الالتواء والتفلطح، بمعنى أنَّ هذه القيم تحدِّد لنا وبدقّة إن كان التوزيع ملتوياً أم لا، وكذلك إن كان التوزيع مفلطحاً، مدبباً أم معتدلاً، وهذا المبحث لن نتناوله في كتابنا هذا.

كما يوجد أنوَّاع أخرى من المقاييس التي تساعدنا في عملية الاستقراء لنتائج البيانات، إلا أنَّها أقل أهمية من المقاييس السابقة.

لقد تحدثنا قبل قليل حول ضرورة البحث عن القيمة العددية (**أو القيم العددية**) التي تنزع نحوها البيانات التي قيد الدراسة. إنَّ هذه القيمة (**أو القيم**) تندرج تحت مفهوم مقاييس النزعة المركزيّة والتي تقدّمها لنا الفقرة الآتية.

2-1-3- تعريف (مقياس النزعة المركزيّة)

لتكن لدينا مجموعة بيانات مُعطاة. عندئذٍ كل قيمة عددية تنزع إليها البيانات كلياً أو جزئياً (**تُبدو كمركز جذب للبيانات**) تُدعى **مقياساً للنزعة المركزيّة**.

من التعريف السابق نلاحظ أن القيمة العددية التي تمثِّل مقياساً للنزعة المركزيّة تُعبِّر في الواقع عن موضع تمركُز توزيع البيانات، ومن المقاييس التي تهتمّ بهذه الدراسة **المتوسّط** (**أو الوسط الحسابي**)، **المتوسّط الهندسي**، **المتوسّط التوافقي**، **الوسيط والمنوال**، ولكن في كتابنا هذا سنقوم بالتركيز على استخدام **المتوسّط**، **الوسيط والمنوال** فقط، والتي سنقوم بتقديمها تباعاً.

2-1-4- تعريف (المتوسّط Mean)

إنَّ المتوسط لمجموعة بيانات يمثِّل في الواقع مركز ثقل هذه البيانات، ولهذا فإنَّه يُعدّ من أهم مقاييس النزعة المركزية، ولتعريفه سنناقش الحالات الأربع الآتية:

1– إذا كانت البيانات المُعطاة **خام** من قبيل x_1 ، x_2 ، ... و x_n (**قد لا تكون مختلفة بعضها عن البعض الآخر**)، فعندئذٍ يُعرَّف المتوسط لهذه البيانات على أنَّه قيمة عددية (**يُرمز له بـ** $\bar{x}$) تحسب باستخدام العلاقة الآتية:

$$\bar{X} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad [2-1-a]$$

2- إذا كانت البيانات المُعطاة **مفردة (خام)** من قبيل x_1، x_2، ... و x_k ولها أوزان w_1، w_2، ... و w_k، فعندئذٍ يُعرَّف المتوسط لهذه البيانات على أنَّه قيمة عددية (**ويُرمز له بـ $\bar{X}$ أيضاً**) تحسب باستخدام العلاقة الآتية:

$$\bar{X} = \frac{w_1 \cdot x_1 + w_2 \cdot x_2 + \cdots + w_k \cdot x_k}{w_1 + w_2 + \cdots + w_k} = \frac{\sum_{i=1}^{n} w_i \cdot x_i}{\sum_{i=1}^{n} w_i} \qquad [2-1-b]$$

إنَّ القيمة العددية $\bar{X}$ التي حُسِبت بالعلاقة السابقة تُدعى المتوسط الموزون للبيانات x_1، x_2، ... و x_k.

3- إذا كانت البيانات مُعطاة من خلال جدول تكراري كما في الجدول [1-1] (**في الفصل الأول**)، فعندئذٍ يُعرَّف المتوسط لتلك البيانات على أنَّه قيمة عددية (**ويُرمز لها بـ $\bar{X}$ أيضاً**) تحسب باستخدام العلاقة الآتية:

$$\bar{X} = \frac{1}{\sum_{i=1}^{m} f_i} \sum_{i=1}^{m} f_i \cdot x_i \qquad [2-1-c]$$

علماً أنَّ x_1، x_2، ... و x_m هي قيم الممثّلين للبيانات.

لاحظ هنا أنَّ عمل المتوسط لبيانات جدول تكراري يماثل تماماً عمل المتوسط الموزون حيث تقوم التكرارات بعمل الأوزان فقط، ولكن الفارق بينهما هو أنَّ قيم التكرارات أعداد طبيعية دوماً في حين أنَّ قيم الأوزان هي أعدد حقيقية (**أي ليس بالضرورة أن تكون أعداداً طبيعية**).

4- إذا كانت البيانات المُعطاة **مجمَّعةً** في جدول توزيع تكراري بـ k فئة كما في الجدول [1-10] (**انظر الفصل الأول**)، فعندئذٍ يحسب المتوسط (**ويُرمز له بـ $\bar{X}$ أيضاً**) لبيانات ذلك الجدول من خلال العلاقة الآتية:

$$\bar{X} = \frac{1}{\sum_{i=1}^{k} f_i} \sum_{i=1}^{k} f_i \cdot x_i \qquad [2-1-d]$$

علماً أنَّ x_1، x_2، ... و x_k هي مراكز الفئات.

إنَّ العلاقة الأخيرة [2-1-d] توضِّح لنا أنَّه يمكن تمثيل كل فئة من فئات الجدول التكراري بمركزها كما ذكرنا ذلك في الفصل السابق، وأنَّ لهذا المركز وزناً يساوي تكرار الفئة.

2-1-4-1- أمثلة

1- قام تاجر بشراء أربعة من الإبل كلّ على انفراد بثمنٍ قدره 5950، 6950، 7950 و 6250 ريال. وأراد بيعها دفعةً واحدة بسعر الحبّة الواحدة، فما هو الحد الأدنى للمبلغ الذي يجب أن يطلبه من الشاري في الحبّة الواحدة حتّى لا يقع في أية خسارة؟

الإجابة: حتى لا يقع التاجر في أية خسارة يجب ألاّ يقل المبلغ الذي سيطلبه في الحبّة الواحدة عن قيمة متوسط الثمن الذي دفعه في الإبل الأربعة، ومن ثمَّ بتطبيق العلاقة [2-1-a] يكون لدينا:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{5950+6950+7950+6250}{4} = \frac{27100}{4} = 6775$$

إذن عليه أن يطلب 6775 ريالاً على الأقل ثمناً لكلّ رأس من الإبل حتى لا يقع في أية خسارة.

2- قام طالب بحساب الزمن (**مقدّراً بالدقيقة**) الذي يستغرقه الطريق من منزله إلى عمادة السّنة الأولى المشتركة، وذلك على مدى 12 يوماً، فحصل على القيم الآتية:

32	28	36	28	37	39	25	39	27	42	33	37

فما هو متوسط الزمن اللازم لتنقل الطالب من منزله إلى العمادة؟

الإجابة: بتطبيق العلاقة [2-1-a] نجد أنَّ متوسط الزمن اللازم لتنقل الطالب من منزله إلى العمادة يساوي:

$$\bar{X} = \frac{37+33+42+27+39+25+39+37+28+36+28+32}{12} = \frac{403}{12} = 33.58$$

إذن يحتاج الطالب بالمتوسط إلى 33.58 دقيقة للتنقل من منزله إلى عمادة السّنة الأولى المشتركة.

3– إذا كان لدى طالب خمسة مقرَّرات دراسيّة، وسنفترض أنَّ لكل مقرَّر وزن خاص به مقدَّم في الجدول الآتي، وأنَّ الطالب قد نجح في هذه المقرَّرات الخمسة وكانت درجاته في كل مقرَّر كما هو مدوَّن في الجدول الآتي، فما هو متوسط درجات هذا الطالب في المقرَّرات الخمسة هذه ؟

الجدول [2-1]

المقرَّر	لغة إنكليزية	مهارات الاتصال	تقن	إحصاء	تطوير الذات
الوزن w	3	2.25	2	1.3	1
الدرجة النهائية للطالب x	87	92	80	93	98

الإجابة: بناءً على المعطيات المقدَّمة نجد أنَّ متوسط درجات هذا الطالب في المقرَّرات الخمسة هو المتوسط الموزون لهذه الدرجات، ومنه بتطبيق العلاقة [2-1-b] يكون لدينا:

$$\bar{X} = \frac{1}{\sum_{i=1}^{k} w_i}\sum_{i=1}^{k} w_i \cdot x_i = \frac{(3\times 87)+(2.25\times 92)+(2\times 80)+(1.3\times 93)+(1\times 98)}{9.55} = 88.681$$

4- لدى زيارة مريض مستوصف صحي معيَّن يترتَّب على معاينته وطبابته مبلغاً محدَّداً (**مقدَّراً بوحدة نقدية ما**) حسب نوع المعاينة التي يطلبها. لقد تمَّ رصد عدد زيارات المرضى لذلك المستوصف خلال يوم فكانت المعطيات كما في الجدول التكراري الآتي:

الجدول [2-2]

عدد الزائرين	تكلفة المعاينة والطبابة	العيادة
16	55	العينية
35	45	الأذن والأنف والحنجرة
20	85	العظمية
14	65	الباطنية
15	75	العصبية
100	------------------	**Total**

فما هو متوسط التكلفة للزائر الواحد في ذلك اليوم؟

الإجابة: من المعلوم أنَّ قيمة المتوسط لبيانات جدول تكراري يُعطى بالعلاقة [2-1-c]، ومن ثمَّ يكون متوسط التكلفة للزائر الواحد يساوي:

$$\bar{X} = \frac{1}{\sum_{i=1}^{m} f_i} \sum_{i=1}^{m} f_i \cdot x_i$$
$$= \frac{(16\times 55)+(35\times 45)+(20\times 85)+(14\times 65)+(15\times 75)}{100} = \frac{6190}{100} = 61.90$$

إذن معدَّل التكلفة للزائر الواحد هو 61.90 ريالاً.

5– لتكن لدينا بيانات إحصائية مُعطاة من خلال جدول التوزيع التكراري الآتي الذي يعرض عدد الزبائن الذين اشتروا بضائع (مقدّرة بالريال) من سوق تجاري في يوم معيَّن، فما هو متوسط مبلغ المبيعات لكل زبون؟

الجدول [2-3-a]

(عدد الزبائن) تكرار الفئة	(نطاق المبلغ الذي اشترى به الزبون) الحدود الفعليّة للفئة	رقم الفئة
125	$0 \to 100$	1
48	$100 \to 200$	2
32	$200 \to 300$	3
25	$300 \to 400$	4
27	$400 \to 500$	5
257	--	**Total**

الإجابة: من المعلوم أنَّ قيمة المتوسط لبيانات جدول ذو فئات يُعطى بالعلاقة [2-1-d]، ومن ثمَّ يتوجب علينا تعيين مراكز الفئات لإتمام عملية الحساب حيث لدينا:

الجدول [2-3-b]

1	2	3	4	5	رقم الفئة
$0 \to 100$	$100 \to 200$	$200 \to 300$	$300 \to 400$	$400 \to 500$	الحدود الفعليّة للفئة
50	150	250	350	450	مركز الفئة
125	48	32	25	27	تكرار الفئة

ومنه يكون متوسط مبلغ المبيعات لكل زبون يساوي:

$$\bar{X} = \frac{1}{\sum_{i=1}^{k} f_i} \sum_{i=1}^{k} f_i \cdot x_i$$
$$= \frac{(125 \times 50) + (48 \times 150) + (32 \times 250) + (25 \times 350) + (27 \times 450)}{257} = \frac{42350}{257} = 164.79$$

إذن متوسط مبلغ المبيعات لكل زبون هو 164.79 ريالاً.

6– بالرجوع إلى المثال (1-3-4) (عدد الأميال المقطوعة لكل لتر من البنزين). ما هو متوسط الأميال المقطوعة لكل لتر من البنزين وفقاً لجدول توزيعها التكراري ؟

الإجابة: بتطبيق العلاقة [2-1-d] نجد أنَّ متوسط بيانات جدول التوزيع التكراري الخاص بذلك المثال يساوي:

$$\bar{X} = \frac{1}{\sum_{i=1}^{k} f_i} \sum_{i=1}^{k} f_i \cdot x_i$$
$$= \frac{(8 \times 12.5) + (10 \times 14.5) + (12 \times 16.5) + (5 \times 18.5) + (5 \times 20.5)}{40} = \frac{638}{40} = 15.95$$

7- لتكن لدينا مجموعة من البيانات الخام الآتية التي تمثِّل الطول (مقدَّراً بالسنتيمتر) لخمسين طالباً.

175	190	175	172	160	182	155	143	179	160
177	184	179	171	183	165	184	168	160	182
190	168	182	168	157	153	182	175	159	169
149	168	173	162	183	162	144	191	178	162
199	140	185	176	169	166	151	167	160	197

ولنقم بصبِّ هذه البيانات في جدول توزيع تكراري كما في الجدول الآتي:

الجدول [2-4]

رقم الفئة	الحدود الفعليّة للفئة	مركز الفئة	تكرار الفئة	التكرار المتجمّع الصاعد للفئة
1	$140 \to 150$	145	4	4
2	$150 \to 160$	155	5	9
3	$160 \to 170$	165	16	25
4	$170 \to 180$	175	11	36
5	$180 \to 190$	185	10	46
6	$190 \to 200$	195	4	50
Total	------------	--------	50	-----------------

ما هو متوسط الطول لهؤلاء الطلاب وفقاً للبيانات الخام والمجمَّعة، ثمَّ ناقش النتيجة؟

الإجابة: بتطبيق العلاقة [2-1-a] نجد أنَّ متوسط الطول لهؤلاء الطلاب وفقاً للبيانات الخام يساوي:

$$\bar{X} = \frac{175 + 190 + 175 + \ldots + 167 + 160 + 197}{50} = \frac{8529}{50} = 170.58$$

وأمّا بعد صبِّها في جدول التوزيع التكراري فنجد بتطبيق العلاقة [2-1-d] أنَّ متوسط الطول لهؤلاء الطلاب يساوي:

$$\bar{X} = \frac{1}{\sum_{i=1}^{k} f_i} \sum_{i=1}^{k} f_i \cdot x_i = \frac{(4\times 145) + (5\times 155) + \cdots \qquad 195)}{50} = \frac{8550}{50} = 171$$

نلاحظ هنا وجود اختلاف في قيمة المتوسط بين البيانات الخام والمجمَّعة، والسبب في ذلك يعود إلى أنَّ العلاقة [2-1-d] لا تأخذ جميع القيم في الحسبان وإنَّما تكتفي بالقيم الممثِّلة للفئات (التي هي مراكز الفئات) فقط. لهذا السبب، فإنَّه إذا ما وجد اختلاف في قيمتي المتوسط (للخام والمجمَّعة) فإنَّ قيمة المتوسط الناتجة عن البيانات الخام تكون أكثر دقَّة من تلك القيمة الناتجة عن ذات البيانات بعد تجميعها في جدول توزيع تكراري.

▶

2-1-4-2- ملاحظة

إذا كانت x_1، x_2، ... و x_n بيانات خام مُعطاة، فعندئذ المجموع الجبري للفروق $x_1 - \bar{X}$، $x_2 - \bar{X}$، ... و $x_n - \bar{X}$ يساوي الصفر دوماً، وهذا يعني أنَّ المجموع الجبري لانحرافات قيم البيانات عن متوسطها معدوم دوماً، وذلك لأنَّ:

$$\sum_{i=1}^{n}\left(x_i - \bar{x}\right) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{X} = \sum_{i=1}^{n} x_i - n\cdot\bar{X} = n\cdot\bar{X} - n\cdot\bar{X} = 0$$

فعلى سبيل المثال، لو أخذنا البيانات: 4, 5, 6, 1 و 4 فإننا نجد متوسطها هو 4، ومن ثمّ يكون لدينا:

$$(4 - 4) + (5 - 4) + (6 - 4) + (1 - 4) + (4 - 4) = 0 + 1 + 2 - 3 + 0 = 0$$

2-1-4-3- مزايا المتوسط

إنَّ المتوسط أكثر مقاييس النزعة المركزيّة استخداماً وذلك يعود لسهولة حسابه وتعريفه بعلاقة رياضيه بسيطة، ومن أهمّ المزايا (أو المحاسن) التي يتمتَّع بها هي:

1- يأخذ بالحسبان جميع القياسات التي تخضع للدراسة والبحث، ولذلك يُفضَّل استخدامه عندما يكون الاهتمام منصباً على القيمة العددية التي تأخذ جميع القياسات بالحسبان وليس الحصول على قيمة نموذجية مُمثِّلة لها فقط.

2- يخضع للعمليات الجبرية المعروفة، فلو كان a وb عددين حقيقيين مع $a \neq 0$، فعندئذٍ:

أ- المتوسط للقيم $a\cdot x_1$، $a\cdot x_2$، ... و $a\cdot x_n$ يساوي $a\cdot\bar{X}$،

ب- المتوسط للقيم $x_1 + b$، $x_2 + b$، ... و $x_n + b$ يساوي $\bar{X} + b$،

ومن ثمَّ ينتج لدينا أنَّ المتوسط للقيم $a\cdot x_1 + b$، $a\cdot x_2 + b$، ... و $a\cdot x_n + b$ يساوي $a\cdot\bar{X} + b$.

3- إذا عُلِمت قيمته فإنَّه يمكن حساب مجموع البيانات إذا كان عددها معلوماً، أو حساب عدد البيانات إذا كان المجموع لها معلوماً.

4- إذا اضطررنا إلى تمديد حجم البيانات، فإنَّه بإضافة قيم تساوي قيمة متوسط البيانات الأصل لا تتغيَّر قيمة هذا المتوسط.

5- يفضّل استخدام المتوسط عندما يكون التوزيع متماثلاً (أو متناظراً) على وجه التقريب، وكذلك عندما تكون جميع البيانات معلومة ولا يوجد فيها بيانات مفقودة.

2-1-4-4- عيوب المتوسط

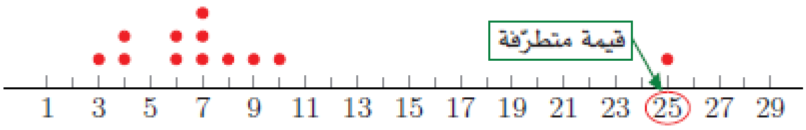
في الواقع للمتوسط عيوب (أو مساوئ) عديدة من أهمِّها ما يلي:

1- إذا فقدت إحدى أو بعض قيم البيانات فإنَّ المتوسط يصبح عديم التطبيق (حتى إذا عُلم ترتيبها بين البيانات).

2- تأثُّره بالقيم المتطرفة (أو القيم المنعزلة Outlier Values وسنأتي على تعيين هذه القيم لاحقاً في هذا الفصل)، فعلى سبيل المثال لو قامت مجموعة من التلاميذ بجمع تبرعات من أجل عمل خيري، وبفرض أنَّ هذه التبرعات كانت على النحو الآتي (مقدّرةً بالريال):

4	3	7	4	7	25	6	7	6	10	9	8

وبتمثيل هذه القيم على امحور الحقيقي يصبح لها العرض الآتي حيث نلاحظ أنَّ القيمة 25 تقع بعيداً على الطرف الأيمن من المحور.



الشكل [2-1]

وهنا نلاحظ بوضوح تام أنَّ كل تبرعات التلاميذ كانت أقل من 11 ريال باستثناء أحدهم قد تبرع بـ 25 ريالاً، فلو استثنينا هذه القيمة 25 لوجدنا أنَّ متوسط باقي القيم هو 6.45 ريالاً فقط، ولكن عندما تبرع أحد الطلاب بـ 25 ريالاً أصبحت قيمة المتوسط 8 ريالات، وكأنَّما قامت القيمة المتطرفة (25) بسحب قيمة المتوسط نحوها ممّا أظهر لنا نتيجة لا تُعبِّر عن واقع التبرعات للتلاميذ.

إذن، فإذا وجدت قيم متطرفة فإنَّنا على الأقلّ غير قادرين على إعطاء قيمة مقنعة ومقبولة لقيمة المتوسط، ولذلك فمن غير المرغوب فيه استخدام المتوسط كمقياس للنزعة المركزيّة في هذه الحالة.

لقد لاحظنا فيما سبق أنَّه إذا فقدت إحدى أو بعض البيانات فإنَّنا سنقف عاجزين عن حساب قيمة المتوسط للبيانات المُعطاة. لذلك كان من الضرورة البحث عن طرائق أخرى تقوم بهذه المهمّة. إنَّ المفهوم الآتي يقدِّم لنا حلاً جزئياً للمشكلة السابق ذكرها، ويعدّ من المفاهيم البديلة للمتوسط في حال عدم إمكانية استخدامه أو عدم القبول بنتيجته.

2-1-5- تعريف (الوسيط Median)

من أجل تعريف الوسيط لمجموعة بيانات مُعطاة سنناقش الحالات الثلاث الآتية:

1– إذا كانت البيانات المُعطاة خام فعندئذٍ يُعرَّف الوسيط لهذه البيانات على أنَّه القيمة التي تقع وسط البيانات بعد ترتيبها تصاعدياً (وسوف نرمز له بـ $\tilde{X}$)، فلو كانت x_1 ، x_2 ، ... و x_n بيانات خام مرتبة تصاعدياً، فإنَّ قيمة الوسيط لهذه البيانات تحسب باستخدام إحدى العلاقتين الآتيتين:

أ- إذا كان عدد البيانات n فردياً فيكون لدينا:

$$\tilde{X} : X_{\frac{n\ 1}{2}} \qquad [2\text{–}2\text{-a}]$$

ب- إذا كان عدد البيانات n زوجياً فيكون لدينا:

$$\tilde{X} : \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} \qquad [2\text{–}2\text{-b}]$$

2– إذا كانت البيانات مُعطاة من خلال جدول تكراري كما في الجدول [1-1] (في الفصل الأول)، وكانت القيم الممثِّلة للبيانات قابلة للترتيب فيما بينها (بمعنى أنّنا سنركز هنا على البيانات الكميّة)، فعندئذٍ نقوم بإعادة تنسيق الجدول بحيث تصبح فيه قيم ممثّلي البيانات x_1، x_2، ... و x_n مرتَّبة تصاعدياً، أي يكون لدينا في هذه الحالة $x_1 < x_2 < ... < x_m$، ومن ثمَّ ندرج عموداً للتكرارات التراكميّة كما في الجدول الآتي:

الجدول [2-5]

التكرار المتجمِّع الصاعد	التكرار	الممثِّل
f_1	f_1	x_1
$f_1 + f_2$	f_2	x_2
$\vdots \;\; \vdots \;\; \vdots \;\; \vdots\vdots$	$\vdots$	$\vdots$
$\sum f_i = f_1 + f_2 + ... + f_m$	f_m	x_m
-------------------------	$\sum f_i$	sum

عندئذ لتعيين الوسيط لبيانات هذا الجدول (وسنرمز له بـ $\tilde{x}$ أيضاً) سنناقش الحالتين الآتيتين:

أ– إذا كان $\sum f_i = n$ عدداً فردياً فإنَّنا نأخذ كقيمة للوسيط ذلك الممثِّل x_i المقابل لأول قيمة أكبر أو يساوي $\frac{n+1}{2}$ في عمود التكرار المتجمّع الصاعد.

ب- إذا كان $\sum f_i = n$ عدداً زوجياً فعندئذ:

i. إذا كان التكرار المتجمّع الصاعد لممثِّل x_i يساوي $n/2$ تماماً، فإنَّنا نأخذ المتوسط للممثِّلين x_i و x_{i+1} كقيمة للوسيط.

ii. إذا كان التكرار المتجمّع الصاعد لممثِّل x_i هو أول قيمة أكبر تماماً من $n/2$، والتكرار المتجمّع الصاعد للممثِّل x_{i-1} أصغر تماماً من $n/2$، فإنَّنا نأخذ الممثِّل x_i نفسه كقيمة للوسيط.

3– إذا كانت البيانات المُعطاة مجمَّعةً في جدول توزيع تكراري بـ k فئة كما في الجدول [1-10] (في الفصل الأول)، فعندئذٍ يُحسب الوسيط لبيانات ذلك الجدول (وسنرمز له بـ $\tilde{X}$ أيضاً) من خلال العلاقة الآتية:

$$\tilde{X} : L \quad \frac{\frac{\sum f_i}{2} - (\tilde{F} \quad \tilde{f}_j)}{\tilde{f}} \times C \qquad [2-3]$$

علماً أنَّ: $\tilde{L}$ هو الحد الأدنى للفئة (الفعليَّة) الوسطيّة، وأمّا الفئة الوسطيّة فهي أول فئة تكرارها المتجمِّع الصاعد أكبر من نصف مجموع التكرارات.

$\tilde{f}$ هو تكرار الفئة الوسطيّة.

$\tilde{F}$ هو التكرار المتجمّع الصاعد للفئة الوسطيّة.

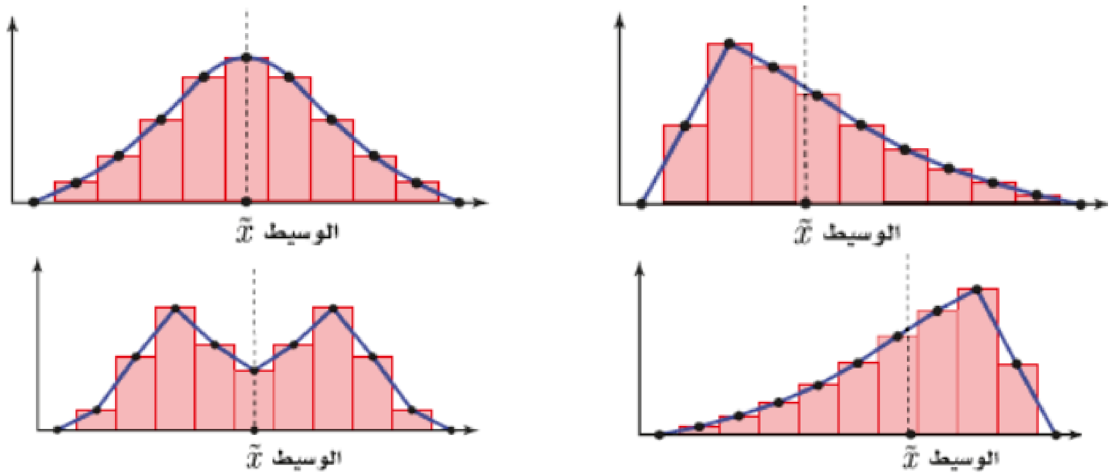
C هي سعة الفئة الوسطيّة (هنا لم نرفق دليل لـ C أو رمز خاص بها لأنَّنا نتعامل مع فئات متساوية السعة)

$\sum_{i=1}^{k} f_i$ هو مجموع التكرارات.

2-1-5-1- ملاحظات

1- قد يلجأ البعض إلى أخذ **مركز** الفئة الوسيطية كقيمة تقريبية للوسيط $\tilde{x}$ (**على سبيل التبسيط**)، ولكن هذا قد يؤدي إلى استنتاجات خاطئة لدى استخدامها من أجل إجراء مقارنة بينها وبين مقاييس النزعة المركزيّة الأخرى أو لدى الاستدلال بها عن شكل توزيع البيانات.

2- كما هو واضح من تعريف الوسيط، فإنَّ الوسيط يأخذ موضعه في وسط البيانات المرتَّبة، ومن ثمَّ فإنَّ 50% من البيانات ستقع على يساره ومثل ذلك على يمينه، ومن أجل مجموعة بيانات مجمَّعة في جدول توزيع تكراري يُعبَّر عن الوسيط هندسياً على أنَّه القيمة على محور الفئات التي إذا رُسم عندها عموداً فإنَّه سيقسم التوزيع (**أو المدرَّج التكراري**) إلى قسمين متساويين (**وليس بالضرورة منطبقين**)، ويكون الانطباق في حالة التوزيعات المتناظرة فقط (للتوضيح انظر الأشكال الآتية).



الشكل [2-2]

2-1-5-2- أمثلة

1- تقدَّم تسعة طلاب للاختبار النهائي في مقرَّر دراسي وكانت نتائجهم على النحو الآتي:

45	27	37	43	42	47	42	45	39

فما هي الدرجة التي تجاوزها 50% من الطلاب؟

الإجابة: نلاحظ أنَّ الإجابة على هذا السؤال تتمّ بتعيين الوسيط لهذه الدرجات، ولذلك لنقم أولاً بترتيبها تصاعديا فيكون لها الترتيب الآتي:

27	37	39	42	42	43	45	45	47	**القيم بعد ترتيبها**
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	**رموز القيم بعد ترتيبها**

وبما أنَّ عدد البيانات فردي فإنَّه بتطبيق العلاقة [2-2-a] نجد أنَّ قيمة الوسيط يساوي:

$$\tilde{x} = x_{\frac{n+1}{2}} = x_5 = 42$$

إذن، فالدرجة التي تجاوزها 50% من الطلاب هي 42.

2- رصدت أعمار عشرة بطاريات حاسب محمول فكانت لدينا البيانات الآتية مقدَّرة بالشهر:

13	8	9	17	14	14	3	12	5	11

فما هي قيمة الوسيط لأعمار البطاريات؟

الإجابة: لتعيين قيمة الوسيط لأعمار بطاريات الحواسيب سنقوم بترتيب البيانات التي لدينا تصاعديا فيكون لها الترتيب الآتي:

قيم الأعمار بعد ترتيبها	17	14	14	13	12	11	9	8	5	3
رموز القيم بعد ترتيبها	x_{10}	x_9	x_8	x_7	x_6	x_5	x_4	x_3	x_2	x_1

وبما أنَّ عدد البيانات زوجي فإنَّه بتطبيق العلاقة [2-2-b] نجد أنَّ قيمة الوسيط يساوي:

$$\tilde{X} = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} = \frac{X_5 + X_6}{2} = \frac{11+12}{2} = 11.5$$

إذن، فإنَّ 50% من بطاريات الحواسيب التي خضعت للدراسة لم تعمّر أكثر من 11.5 شهراً.

3- بالرجوع إلى المثال (4) من (2-1-4-1) **زيارة مريض لمستوصف صحي**)، فإننا نجد بعد ترتيب القيم الممثِّلة لتكلفة المعاينة تصاعدياً أنَّ لبيانات الجدول التكراري العرض الآتي:

الجدول [2-6]

التكرار المتجمِّع الصاعد	**عدد الزائرين**	**تكلفة المعاينة والطباعة**	**العيادة**
35	35	45	الأذن والأنف والحنجرة
51	16	55	العينية
65	14	65	الباطنية
80	15	75	العصبية
100	20	85	العظمية
-------------	**100**	-----------------	**Total**

فما هي قيمة الوسيط لبيانات هذا الجدول؟

الإجابة: بتطبيق القاعدة الخاصّة بتعيين الوسيط لبيانات الجداول التكرارية نجد أنَّ القيمة الممثِّلة التي تكرارها المتجمِّع الصاعد أكبر أو يساوي 50 هي القيمة التي تمثِّل تكلفة وطبابة المعاينة العينية، ومن ثمَّ قيمة الوسيط لبيانات هذا الجدول هي $\tilde{X}$ 55.

4- بالرجوع إلى المثال (7) من (2-1-4-1- **الطول مقدَّراً بالسنتيمتر لخمسين طالباً**-)، فنجد بعد ترتيب البيانات تصاعدياً أنَّ لها العرض الآتي:

140	153	160	162	168	171	175	179	183	189
143	155	160	165	168	172	176	182	183	190
144	157	160	166	168	173	177	182	184	191

149	159	162	167	169	175	178	182	184	197
151	160	162	168	169	175	179	182	185	199

ما هي قيمة الوسيط لطول هؤلاء الطلاب وفقاً للبيانات الخام وللبيانات المجمَّعة في الجدول [2-4]، ثمَّ ناقش النتيجة؟

الإجابة: بما أنَّ عدد البيانات زوجي فإنَّه بتطبيق العلاقة [2–2-b] نجد قيمة الوسيط للبيانات الخام تساوي:

$$\tilde{x} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{x_{25} + x_{26}}{2} = \frac{169 + 171}{2} = 170$$

وبعد صبّ هذه البيانات في الجدول [2-4] نجد أنَّ الفئة الرابعة هي الفئة الوسيطيّة، ومن ثمَّ بتطبيق العلاقة [2–3] تكون قيمة الوسيط لهذه البيانات بعد تجميعها في الجدول [2-4] تساوي:

$$\tilde{x} = \ldots \frac{\frac{1}{2}\sum_{i=1}^{n} f_i - (\ldots)}{\ldots} \times \mathbf{C} = 170 + \frac{25 - (36 - 11)}{11} \times 10 = 170$$

نشير هنا إلى أنَّه من الممكن أن تتوافق قيمة الوسيط للبيانات الخام مع قيمة الوسيط لبيانات جدول التوزيع التكراري الممثِّل لها، ولكن في معظم الحالات تكون قيمة الوسيط للبيانات الخام مختلفةً عن قيمة الوسيط للبيانات ذاتها بعد تجميعها في جداول توزيع تكرارية.

تعقيباً على ملاحظة سابقة نلاحظ أنَّه لو أخذنا مركز الفئة الوسيطية كقيمة تقريبية لقيمة وسيط بيانات الجدول فإنَّنا سنجده يساوي 165، وهكذا نجد أنَّنا قد ارتكبنا خطأً ليس بالقليل بسبب كِبَر الفارق بينها وبين القيمة التي حُسِبت لوسيط بيانات جدول التوزيع.

2-1-5-3- مزايا الوسيط

1- الوسيط سهل التعريف والحساب.

2- لا يتأثر بالقيم المتطرفة لأنَّه يعتمد على القيم الواقعة في الوسط فقط.

3- لا يعتمد على جميع القيم في حسابه، ومن ثمَّ تغيّر قيمة أو أكثر من القيم غير الواقعة في الوسط لا تؤثر في قيمة الوسيط.

4- إذا فُقدت بعض قيم البيانات المرتّبة غير الواقعة في الوسط فإنَّه يمكن استخدام الوسيط.

2-1-5-4- عيوب الوسيط

1- الوسيط أقل دقَّة من المتوسط لأنَّه لا يأخذ كل القيم بالحسبان.

2- إذا فقدت إحدى القيم المرتَّبة الواقعة في الوسط فإنَّه من غير الممكن حساب الوسيط.

أخيراً نشير إلى أنَّه يُفضّل استخدام الوسيط إذا كان الاهتمام منصباً على إيجاد قيمة ممثِّلة للقيمة التي تنزع إليها البيانات بدلاً من الاهتمام بالمجموع الكلي، وكذلك عندما يكون التوزيع ملتوياً.

الآن، إذا كانت البيانات المفقودة واقعة في الوسط بعد ترتيبها تصاعدياً (ولكن معلوم موضعها) فعندئذٍ من غير الممكن استخدام المتوسط والوسيط، ولذلك يمكن اللجوء إلى المقياس الآتي لاستخدامه كمقياس للنزعة المركزيّة.

2-1-6- تعريف (المنـوال) (Mode)

من أجل تعريف المنوال لمجموعة بيانات مُعطاة سنناقش الحالات الثلاث الآتية:

1– إذا كانت البيانات المُعطاة **خام** فعندئذٍ يُعرَّف المنوال (وسنرمز له بـ $\hat{X}$) لهذه البيانات على أنَّه البيان (قد يكون كميّاً أو نوعيّاً) الأكثر تكراراً.

2– إذا كانت البيانات مُعطاة من خلال جدول تكراري كما في الجدول [1-1] (في الفصل الأول)، فعندئذٍ يُعرَّف المنوال لبيانات ذلك الجدول (وسنُرمز له بـ $\hat{X}$ أيضاً) على أنَّه الممثِّل الذي يقابله أكبر قيمة تكرار (الممثِّل هو قيمة عددية إذا كانت البيانات كميّة، ورمزٌ أو رقمٌ أو صفةٌ إذا كانت البيانات نوعيّة.

3– إذا كانت البيانات المُعطاة **مجمَّعةً** في جدول توزيع تكراري بـ k فئة كما في الجدول [1-10]، فعندئذٍ تُحسب قيمة المنوال لبيانات ذلك الجدول (وسنُرمز له بـ $\hat{X}$ أيضاً) من خلال العلاقة الآتية:

$$\hat{X} = \hat{L} + \frac{d_1}{d_1 + d_2} C \qquad [2-4]$$

علماً أنَّ: $\hat{L}$ هو الحد الأدنى للفئة المنوالية،

d_1 هو الفرق بين تكرار الفئة المنواليّة وتكرار الفئة السابقة لها مباشرةً،

d_2 هو الفرق بين تكرار الفئة المنواليّة وتكرار الفئة اللاحقة بها مباشرةً،

C هو سعة الفئة المنواليّة.

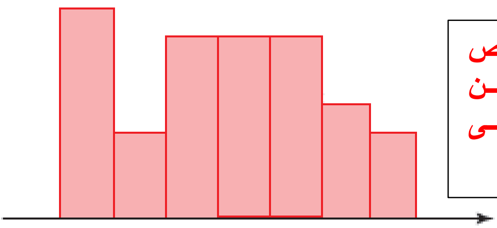
علماً أنَّ الفئة المنواليّة هي تلك الفئة التي تكرارها أكبر من تكرار الفئة السابقة لها واللاحقة بها مباشرةً، وعلى ألا تكون هذه الفئة طرفيّة. أي إنَّ الفئة الأولى والأخيرة في التوزيع التكراري لا ينظر إليهما كفئتين منواليّتين. كذلك إذا كان لفئة ملاصقة لفئة طرفية تكرار يساوي تكرار الفئة الطرفية فإنَّنا لن ننظر إليها كفئة منوالية أيضاً.

2-1-6-1- ملاحظات الفقرة رقم (1) حُذفِت وحلّ مكانها الملاحظة رقم (2)

1- من تعريف المنوال يُلاحظ أنَّ المنوال قد لا يكون موجوداً في حال أنَّ جميع البيانات مختلفة عن بعضها البعض الآخر أو أنَّ لجميع القيم الممثِّلة التكرار نفسه. كذلك يُلاحظ أنَّه حتى في حال وجوده قد لا يكون وحيداً في حال تساوي التكرار لقياسين على الأقل من هذه القياسات حيث يكون لدينا منوالين على الأقل في هذه الحالة.

2- إذا حصل أنَّه لدينا فئات داخلية متجاورة ذات تكرار متساوٍ وأعظمي بالنسبة لتكراري الفئتين السابقة واللاحقة بهذه الفئات، فإنَّنا سنتعامل مع هذه الفئات المتجاورة على أنّها فئة منواليّة بسعةٍ تساوي مجموع السعات لهذه الفئات (انظر الشكل الآتي [2-3-b]).

تـمّ نقـل الشـكل الخـاص بهـذه الملاحظـة مـن الصـفحة التاليـة إلـى مكانها الصحيح.



الشكل [2-3-b] فئات داخلية متجاورة ذات تكرار متساوٍ

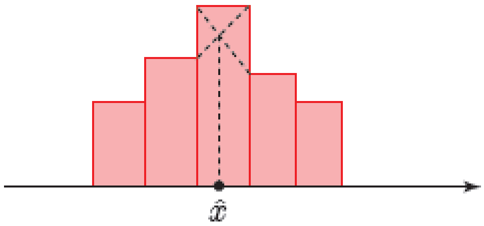
هنا نلاحظ أنَّه إذا أصبح لجميع الفئات في جدول توزيع تكراري التكرار نفسه فعندئذٍ لا يمكن حساب المنوال، ومن ثمَّ لا وجود للمنوال في هذه الحالة.

3- إذا كان المدرَّج التكراري مُعطى فإنَّه يمكن تعيين المنوال هندسياً على النحو الآتي:

أ- نصل بقطعة مستقيمة بين النهاية اليمنى لقمة الفئة الفعليَّة المنوالية مع النهاية اليمنى لقمَّة الفئة السابقة لها.

ب- نصل بقطعة مستقيمة بين النهاية اليسرى لقمة الفئة الفعليَّة المنوالية مع النهاية اليسرى لقمَّة الفئة اللاحقة بها.

ج- نسقط عمود من نقطة تقاطع المستقيمين الناشئين على محور الفئات فتكون النقطة الموافقة لهذا المسقط هي قيمة للمنوال.



الشكل [2-3-a] القيمة المقدَّرة للمنوال

إنَّ هذه الطريقة لتعيين المنوال تحتاج إلى دقة في الرسم واستخدام مقاييس دقيقة، ولذلك لا تستخدم إلاَّ عند الضرورة.

2-1-6-2- أمثلة

1- لدى وزن عشرة صناديق متماثلة تحتوي على برتقال وجدنا أوزانها كما يلي:

8	8.5	8.25	7.85	8	7.75	7.95	8	8.45	8.25

فما هو منوال هذه البيانات؟

الإجابة: نلاحظ أنَّ القيمة الأكثر تكراراً بين البيانات المُعطاة هي 8 ، وهي قيمة وحيدة ومن ثمَّ لدينا منوال وحيد في هذه البيانات هو $\hat{X} = 8$.

2- في سباق للسيَّارات عالية السرعة في حلبة مُحدَّدة، دوِّنت الأزمنة الآتية لثمانية متسابقين (**الزمن مقدَّر بالدقيقة**):

13.62	12.79	12.88	13.09	13.00	12.95	13.82	13.57

فما هو منوال هذه البيانات؟

الإجابة: نلاحظ هنا أنَّ لجميع القيم التكرار نفسه، ومن ثمَّ لا يوجد منوال لهذه البيانات.

3- بالرجوع إلى المثال (7) من (2-1-4-1- **الطول مقدَّراً بالسنتيمتر لخمسين طالباً-**)، ما هي القيمة المنوال (**أو القيم المنوالية**) لطول هؤلاء الطلاب وفقاً للبيانات الخام وللبيانات المجمَّعة في الجدول [2-4]، ثمَّ ناقش النتيجة؟

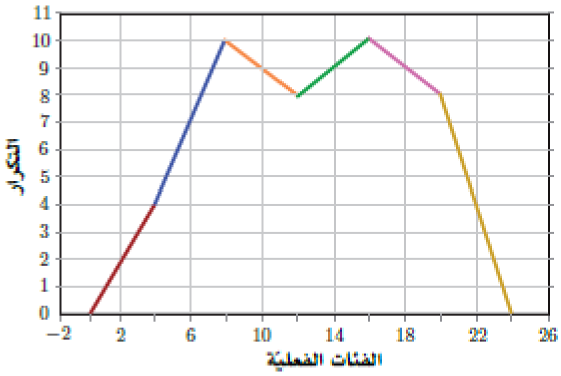
الإجابة: بخصوص البيانات الخام نجد فيها ثلاث قيم لها أكبر قيمة تكرار (**لكل منها أربعة تكرَّارات**)، ولذلك سيكون لتلك البيانات ثلاثة مناويل هي $\hat{X}_1 = 160$ و $\hat{X}_2 = 168$ و $\hat{X}_3 = 182$.

وبعد صبّ تلك البيانات في الجدول [2-4] نجد أنَّ الفئة الثالثة هي الفئة المنوالية، ومن ثمَّ تكون قيمة المنوال لتلك البيانات تساوي:

$$\hat{X} = \hat{L} + \frac{d_1}{d_1 + d_2} C = 160 + \frac{11}{11+5} \times 10 = 166.88$$

نلاحظ هنا أنَّه من غير الممكن مناقشة قيم المناويل للبيانات الخام مع قيمة المنوال للبيانات ذاتها بعد تجميعها في الجدول [2-4] وذلك بسبب عدم وحدانية المنوال للبيانات الخام.

4- لتكن لدينا بيانات مقدَّمة من خلال المضلَّع التكراري الآتي:



الشكل [2-4]

فما هي قيمة المنوال لبيانات هذا المضلَّع؟

الإجابة: من أجل ذلك لنقم ببناء جدول التوزيع التكراري الموافق لهذا المضلَّع، فنجد له العرض الآتي:

الجدول [2-7]

رقم الفئة	الحدود الفعليّة للفئة	مركز الفئة	تكرار الفئة	التكرار المتجمّع الصاعد للفئة
1	$2 \rightarrow 6$	4	4	4
2	$6 \rightarrow 10$	8	10	14
3	$10 \rightarrow 14$	12	8	22
4	$14 \rightarrow 18$	16	10	32
5	$18 \rightarrow 22$	20	8	40
Total	-------------	---------	**40**	------------------

فنلاحظ وجود فئتين منواليتين في هذا الجدول، ومن ثمَّ تملك بيانات هذا الجدول منوالين هما:

$$\hat{X}_1 = \hat{L} + \frac{d_1}{d_1 + d_2} C = 6 + \frac{6}{6+2} \times 4 = 9$$

$$\hat{X}_2 = \hat{L} + \frac{d_1}{d_1 + d_2} C = 14 + \frac{2}{2+2} \times 4 = 16$$

2-1-6-3- مزايا المنوال

1- تأثرهِ بتغيّر بعض قيم البيانات قد لا يكون كبيراً بالقدر الذي يحصل للمتوسط.

2- يمكن استخدامه في حالة البيانات النوعيّة، فعلى سبيل المثال إذا كان لون العيون لمعظم النَّاس في بلد ما هو اللون البني، فعندئذٍ يمكننا القول إنَّ اللون المنوالي لعيون هؤلاء النَّاس هو اللون البني.

3- في حال فقدان بعض البيانات الواقعة في الوسط (**بعد ترتيبها تصاعدياً أو تنازلياً**) فإنَّ للمنوال دور جيد في تعيين القيمة الممثِّلة لتمركز البيانات.

4- كما أنَّه يشير في كثير من الحالات إلى عدد العوامل المؤثِّرة في توليد البيانات من خلال عدد المناويل التي ستظهر للبيانات عند تعيينها.

2-1-6-4- عيوب المنوال

1- إنَّ المنوال هو أقل مقاييس النزعة المركزيّة دقَّة واستخداماً.

2- من مساوئ المنوال أنَّ قد لا يكون موجوداً، وحتى في حال وجوده فقد لا يكون وحيداً، وهذا يعني أنَّه لا يمكن تعريفه بشكل وحيد كمقياس للنزعة المركزيّة.

3- يعدُّ المنوال غير ذي جدوى عمليّاً إذا كان عدد البيانات قليلاً (**هذا إنَّ وجِدَ أصلاّ**)، وأمَّـا في حالة البيانات كبيرة العدد ومن أجل جداول التوزيع التكرارية فيمكن الاستفادة منه لكون تأثرهِ بتغيّر بعض قيم البيانات قد لا يكون كبيراً بالقدر الذي يُلغي استخدامه.

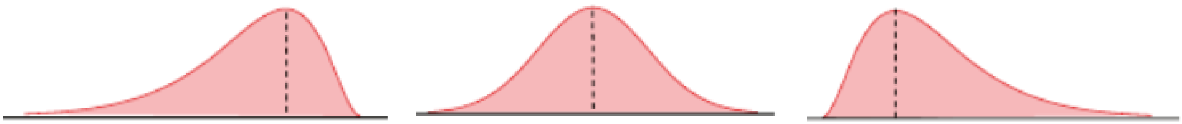
2-1-6-5- ملاحظات

1- نشير هنا إلى ما يقوم به البعض بأخذ مركز الفئة المنوالية كقيمة تقريبيّة لقيمة منوال بيانات جداول التوزيع التكرارية، فلو أخذنا مركز الفئة المنوالية في الجدول [2-4] كقيمة تقريبيّة لقيمة منوال بيانات ذلك الجدول فإنَّنا سنجده يساوي 165، وهذه القيمة تقلّ بمقدار 1.88 عن القيمة المحسوبة في المثال (3) من (2-1-6-2)، وكذلك لا تتوافق مع أية قيمة من قيم المناويل للبيانات الخام نفسها في المثال المذكور آنفاً. إذن فمن غير المقبول استخدام مركز الفئة المنوالية كقيمة تقريبية لقيمة منوال بيانات جدول التوزيع التكراري.

2- من أجل التوزيعات التكرارية المتماثلة يكون لدينا $\tilde{X} = \hat{X} = \bar{X}$، وأمَّا إذا كان التوزيع قريب من التماثل فإنَّه يكون $\tilde{X} \approx \hat{X} \approx \bar{X}$، وفي هذه الحالة وجد تجريبياً أنَّ:

$$\bar{X} - \hat{X} = 3(\bar{X} - \tilde{X})$$

وأمّا التوضُّع النسبي لقيم المتوسط والوسيط والمنوال في حالة التوزيعات التكرارية (**التوزيعات لمجموعات بيانات مستمرَّة**) المختلفة تُظهرها لنا الأشكال الآتية:



$\bar{X} < \tilde{X}$ $\quad$ $\bar{X} = \hat{X} = \tilde{X}$ $\quad$ $\hat{X} < \tilde{X}$

الشكل [2-5]

2-2- الربيعيَّات
Quartiles

لقد قدَّمنا فيما سبق بعض مقاييس النزعة المركزيّة، حيث لاحظنا أنَّ كل مقياس من هذه المقاييس يشغل موضعاً على محور القيم (**أو الفئات**)، ومن ثمَّ هذه المقاييس التي قدِّمت هي مقاييس موضع أيضاً. لكن يوجد في الواقع مقاييس موضع أخرى لا تنتمي لمجموعة مقاييس النَّزعة المركزيّة، ومنها على سبيل المثال الربيعيَّات والمئينات التي سنقدِّمها تباعاً، ولكن من أجل البيانات الكميّة الخام فقط وبشيء من الإيجاز أيضاً.

2-2-1- تعريف (الربيعيَّات)

لتكن x_1، x_2، ... و x_n بيانات خام **مرتبة تصاعدياً**. عندئذٍ تُعرَّف الربيعيَّات على أنَّها تلك القيم التي تقسم البيانات المُعطاة إلى أربع شرائح (**أو أجزاء**) متساوية الأعداد من البيانات (**أي في كل شريحة العدد نفسه من قيم البيانات**).

2-2-1-1 ملاحظات

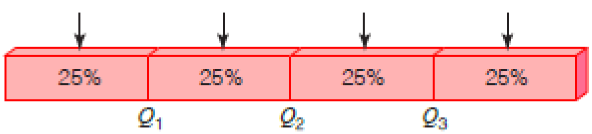
من هذا التعريف يتضح لنا وجود ثلاثة ربيعيَّات وهي:

1- القيمة التي يقع قبلها 25% وبعدها 75% من البيانات المرتبة وتُدعى **الرُبيْعي الأول** First Quartile (**أو الرُبيْعي الأدنى** Lower Quartile)، ويُرمز له بالرمز $\mathbf{Q}_1$.

2- القيمة التي يقع قبلها 50% وبعدها 50% من البيانات المرتبة وتُدعى **الرُبيْعي الثاني** Second Quartile، ويُرمز له بـ $\mathbf{Q}_2$، ويُلاحظ هنا أنَّ الرُبيْعي الثاني هو الوسيط نفسه، أي أنَّ $\mathbf{Q}_2 = \tilde{x}$.

3- القيمة التي يقع قبلها 75% وبعدها 25% من البيانات المرتبة وتُدعى **الرُبيْعي الثالث** Third Quartile (**أو الرُبيْعي الأعلى** Upper Quartile)، ويُرمز له بـ $\mathbf{Q}_3$.

ومن أجل مجموعة البيانات $A = \{1, 2, \cdots$ يمكننا تقديم العرض الهيكلي الآتي للرُبيْعيات الثلاث.



الشكل [2-6]

2-2-2- تعيين الربيعيَّات

في الواقع توجد طرائق عديدة لتعيين الربيعيَّات، وهذه الطرائق تتفاوت في دقّة نتائجها وفقاً للقاعدة الرياضياتية المستخدمة في الحساب، وهذا بدوره يجعل نتائج حساب الربيعيَّات باستخدام البرامج الإحصائية تختلف من برنامج إلى آخر. في كتابنا هذا سنقدّم إحدى هذه الطرائق التي تعدّ من أكثر الطرائق دقةً، وهي مستخدمة في برامج إحصائية كثيرة أيضاً. إنَّ هذه الطريقة تقوم على الخطوات الآتية:

الخطوة الأولى: لتكن x_1، x_2، ... و x_n (مع $3 \le n$) بيانات خام **مرتبة تصاعدياً**، فعندئذ نضع بالتعريف:

$$q_r := \frac{r\,(n+1)}{4} \quad ; r = 1, 2, 3 \qquad [2-5]$$

والذي يُدعى **رتبة** الرُبيْعي r (**وهذه القيمة تحدِّد موضع الرُبيْعي بين البيانات المرتَّبة**).

الخطوة الثانية: تُعيَّن قيمة الرُبيْعي $\mathbf{Q}_r$ مع $r = 1, 2, 3$ على النحو الآتي:

بفرض أنَّ k هو الجزء الصحيح من العدد q_r، وأنَّ الباقي من هذا العدد يساوي s، فعندئذٍ يُحسب الرُبيْعي $\mathbf{Q}_r$ من خلال العلاقة الآتية:

$$\mathbf{Q}_r = x_k + s(x_{k+1} - x_k) \quad ; r = 1, 2, 3 \qquad [2-6]$$

نلاحظ هنا أنَّه إذا كان الباقي $s = 0$ أو كان $x_{k+1} = x_k$ فعندئذٍ سيكون لدينا $\mathbf{Q}_r = x_k$.

2-2-2-1- أمثلة

1- لدينا درجات اختبار نهائي (الدرجة القصوى 50) لخمسة وعشرين طالباً كما هو آتٍ.

48	37	25	45	44
44	49	29	42	48
45	50	39	36	47
41	48	38	15	47
43	47	39	25	48

فما هي قيم الربيعيَّات الثلاثة $\mathbf{Q}_1$، $\mathbf{Q}_2$ و $\mathbf{Q}_3$ الخاصّة بهذه البيانات؟

الإجابة: من أجل تعيين قيم الربيعيَّات الثلاثة $\mathbf{Q}_1$، $\mathbf{Q}_2$ و $\mathbf{Q}_3$ نقوم أولاً بترتب البيانات التي لدينا تصاعدياً، فنجد لها التسلسل الآتي:

15	37	42	45	48
25	38	43	47	48
25	39	44	47	48
29	39	44	47	49
36	41	45	48	50

وبما أنَّ $n = 25$ فإنَّ رتبة الرُبيْعي $\mathbf{Q}_1$، $\mathbf{Q}_2$ و $\mathbf{Q}_3$ ستكون على الترتيب هي:

$$q_1 = \frac{n+1}{4} = \frac{25+1}{4} = 6.5$$

$$q_2 = \frac{2(n+1)}{4} = \frac{2(25+1)}{4} = 13$$

$$q_3 = \frac{3(n+1)}{4} = \frac{3(25+1)}{4} = 19.5$$

فنجد من أجل:

أ- الرُبيْعي الأول $\mathbf{Q}_1$ (**أي أنَّ** $r = 1$) أنَّ الجزء الصحيح من العدد q_1 هو $k = 6$، وأمّا الباقي من هذا العدد يساوي $s = 0.5$، ومن ثمَّ يكون لدينا:

$$\mathbf{Q}_1 = x_k + s(x_{k+1} - x_k) = x_6 + 0.5(x_7 - x_6)$$
$$= 37 + 0.5\,(38 - 37) = 37.5$$

أي أنَّ 25% من الطلاب قد حصلوا على درجة دون 37.5، وأمَّا الباقي (**وهم 75%**) فقد حصلوا على درجة أعلى من 37.5 في هذا الاختبار.

ب- الرُبيْعي الثاني $\mathbf{Q}_2$ نجد أنَّ الجزء الصحيح من العدد q_2 هو $k = 13$، وأمّا الباقي من هذا العدد يساوي $S = 0$، ومن ثمَّ يكون لدينا:

$$\mathbf{Q}_2 = x_k + s(x_{k+1} - x_k) = x_{13} = 44$$

أي أنَّ 50% من الطلاب قد حصلوا على درجة دون 44، وأمَّا الباقي (**وهم 50% أيضاً**) فقد حصلوا على درجة أعلى من 44 في هذا الاختبار.

ج- الرُبيْعي الثالث $\mathbf{Q}_3$ نجد أنَّ الجزء الصحيح من العدد q_3 هو $k = 19$، وأمّا الباقي من هذا العدد يساوي $S = 0.5$، ومن ثمَّ يكون لدينا:

$$\mathbf{Q}_3 = x_k + s(x_{k+1} - x_k) = x_{19} + 0.5(x_{20} - x_{19})$$
$$= 47 + 0.5\ (48 - 47) = 47.5$$

أي إنَّ 75% من الطلاب قد حصلوا على درجة دون 47.5، وأمَّا الباقي (**وهم 25%**) فقد حصلوا على درجة أعلى من 47.5 في هذا الاختبار.

2- لدى معاينة مجموعة مكوَّنة من 20 رجلاً بالغاً من أجل معرفة أوزانهم وجدنا البيانات الآتية (**مقدَّرةً بالكيلو غرام**):

86	95	66	76
87	98	77	92
85	98	78	65
73	106	77	88
70	115	77	102

فما هي قيم الربيعيَّات الثلاثة $\mathbf{Q}_1$، $\mathbf{Q}_2$ و $\mathbf{Q}_3$ الخاصّة بهذه البيانات؟

الإجابة: من أجل تعيين قيم الربيعيَّات الثلاثة $\mathbf{Q}_1$، $\mathbf{Q}_2$ و $\mathbf{Q}_3$ نقوم أولاً بترتيب البيانات التي لدينا تصاعدياً، فنجد لها التسلسل الآتي:

65	77	86	98
66	77	87	98
70	77	88	102
73	78	92	106
76	85	95	115

وبما أنَّ $n = 20$ فإنَّ رتبة الرُبيْعي $\mathbf{Q}_1$ ، $\mathbf{Q}_2$ و $\mathbf{Q}_3$ ستكون على الترتيب هي:

$$q_1 = \frac{n+1}{4} = \frac{20+1}{4} = 5.25$$
$$q_2 = \frac{2(n+1)}{4} = \frac{2(20+1)}{4} = 10.5$$
$$q_3 = \frac{3(n+1)}{4} = \frac{3(20+1)}{4} = 15.75$$

فنجد من أجل:

أ- الرُبيْعي الأول Q_1 (أي إنَّ $r = 1$) أنَّ الجزء الصحيح من العدد q_1 هو $k = 5$، وأمّا الباقي من هذا العدد يساوي $S = 0.25$، ومن ثمَّ يكون لدينا:

$$Q_1 = x_k + s(x_{k+1} - x_k) = x_5 + 0.25(x_6 - x_5)$$
$$= 76 + 0.25\ (77 - 76) = 76.25$$

أي أنَّ 25% من هؤلاء الرجال لهم أوزان دون 76.25، وأمَّا الباقي (**وهم 75%**) فإنَّ أوزانهم أكثر من 76.25.

ب- الرُبيْعي الثاني Q_2 نجد أنَّ الجزء الصحيح من العدد q_2 هو $k = 10$، وأمّا الباقي من هذا العدد يساوي $S = 0.5$، ومن ثمَّ يكون لدينا:

$$Q_2 = x_k + s(x_{k+1} - x_k) = x_{10} + 0.5(x_{11} - x_{10})$$
$$= 85 + 0.5\ (86 - 85) = 85.5$$

أي أنَّ 50% من هؤلاء الرجال لهم أوزان دون 85.5، وأمَّا النصف الباقي منهم فلهم أوزان أكثر من 85.5.

ج- الرُبيْعي الثالث Q_3 نجد أنَّ الجزء الصحيح من العدد q_3 هو $k = 15$، وأمّا الباقي من هذا العدد يساوي $S = 0.75$، ومن ثمَّ يكون لدينا:

$$Q_3 = x_k + s(x_{k+1} - x_k) = x_{15} + 0.5(x_{16} - x_{15})$$
$$= 95 + 0.75\ (98 - 95) = 97.25$$

أي أنَّ 75% من هؤلاء الرجال لهم أوزان دون 97.25، وأمَّا الباقي (**وهم** 25%) فإنَّ أوزانهم أكثر من 97.25.

▶

2-2-2-2- ملاحظات

1- تجدر الإشارة هنا إلى أنَّ قيمة الرُبيْعي قد لا تكون موجودةً بين قيم البيانات المُعطاة.

2- إنَّ القيمتين Q_1 و Q_3 ليستا من مقاييس النَّزعة المركزيّة.

3- من الاستخدامات المهمّة للرُبيْعيات تحديد ما إذا كانت قيمة x من مجموعة بيانات مُعطاة هي قيمة متطرفة Extreme Value (**أو قيمة منعزلة Outlier Value**) أم لا. حيث يُقال عن قيمة x من مجموعة بيانات مُعطاة إنَّها متطرفةٌ إذا حقَّقت إحدى العلاقتين الآتيتين:

$$[2\text{–}7\text{-a}]\quad x < Q_1 - 1.5(Q_3 - Q_1)$$

$$[2\text{–}7\text{-b}]\quad x > Q_3 + 1.5(Q_3 - Q_1)$$

فإذا كانت العلاقة [2–7-a] مُحقَّقةً، فعندئذٍ يُقال إنَّ القيمة x متطرفة بصغرها، وأمَّا المقدار $Q_1 - 1.5(Q_3 - Q_1)$ فإنَّه يُدعى أدنى حاجز Lowest Fence (**لأنَّه حاجز يفصل بين القيم المتطرفة بصغرها وباقي القيم غير المتطرفة**)، وسنرمز له بـ **LF**، أي أنَّه لدينا:

$$LF = Q_1 - 1.5(Q_3 - Q_1)$$

وهو يمثِّل في الواقع **الحد الأعلى للقيم المتطرفة بصغرها** من أجل مجموعة بيانات مُعطاة.

أمّا إذا كانت العلاقة [2–7-b] محققةً، فعندئذٍ يُقال إنَّ القيمة x متطرفة بكبرها، وحينئذ يُقال عن المقدار $Q_3 + 1.5(Q_3 - Q_1)$ إنَّه أعلى حاجز Highest Fence (**لأنَّه حاجز يفصل بين القيم المتطرفة بكبرها وباقي القيم غير المتطرفة**)، وسنرمز له بـ **HF**، أي أنَّه لدينا:

$$\mathbf{HF} = \mathbf{Q}_3 + 1.5(\mathbf{Q}_3 - \mathbf{Q}_1)$$

و هو يمثِّل في الواقع **الحد الأدنى للقيم المتطرفة بكبرها** من مجموعة بيانات مُعطاة.

مثال 2-2-2-3-

لتكن لدينا البيانات الآتية:

6	7	1	4	7	5	5	4	5	3
4	5	7	3	4	15	4	5	4	4
7	5	5	6	5	7	1	6	4	7

فهل يوجد في هذه البيانات قيم متطرفة؟

الإجابة: من أجل تعيين القيم المتطرفة علينا حساب $\mathbf{Q}_1$ و $\mathbf{Q}_3$، وهذا يتطلب منا ترتيب البيانات التي لدينا أولاً، فنجد لها العرض الآتي:

1	1	3	3	4	4	4	4	4	4
4	4	5	5	5	5	5	5	5	5
6	6	6	7	7	7	7	7	7	15

ومنه بحساب $\mathbf{Q}_1$ و $\mathbf{Q}_3$ نجد أنَّ $\mathbf{Q}_1 = 4$ و $\mathbf{Q}_3 = 6.25$، وكذلك:

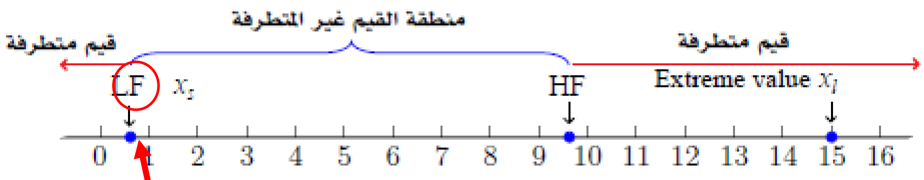
$$\mathbf{Q}_3 - \mathbf{Q}_1 = 6.25 - 4 = 2.25$$

ومن ثمَّ يكون لدينا:

$$\mathbf{LF} = \mathbf{Q}_1 - 1.5(\mathbf{Q}_3 - \mathbf{Q}_1) = 4 - (1.5 \times 2.25) = 0.625 < 1$$

$$\mathbf{HF} = \mathbf{Q}_3 + 1.5(\mathbf{Q}_3 - \mathbf{Q}_1) = 6.25 + (1.5 \times 2.25) = 9.625 < 15$$

وبذلك تكون القيمة 15 هي قيمة **متطرفة** بكبرها ولا وجود لقيم متطرفة بصغرها (والعرض البياني الآتي يوضِّح لنا ذلك أيضاً.



الشكل [2-7]

▶

2-3- المئينات
Percentiles

يعدّ المئين من مقاييس الموضع المهمّة وذلك لأنَّه يوفر معلومات حول كيفية امتداد البيانات من أصغر قيمة إلى أكبر قيمة، فعلى سبيل المثال يُذكر في كثير من الأحيان في جدول نتائج تقديم الطلبات لمؤسسة ما أنَّ 85% من المتقدمين استوفوا الشروط المطلوبة للتعيين فقط. أو أن يُقال إنَّ 72% فقط من الطلاب حقَّقوا درجة النجاح في المقرّر X.

2-3-1 تعريف (المئينات)

لتكن x_1 ، x_2 ، ... و x_n بيانات **مرتبة تصاعدياً**. عندئذٍ تُعرَّف المئينات على أنَّها تلك القيم التي تقسم البيانات إلى مئة شريحة (**أو جزء**) متساوية الأعداد من البيانات (**أي في كل شريحة العدد نفسه من قيم البيانات**).

2-3-2- ملاحظات

1- من هذا التعريف يتضّح لنا وجود تسعة وتسعين **مئيناً**، علماً أنَّه من أجل $r = 1, 2, \dots, 99$ فإنَّ القيمة التي يقـع قبلها **%** r من البيانات (**المرتبة تصاعدياً**) وبعدها **%** $(100 - r)$ تُدعـى **المئين** ذي الرقم r (rth Percentile **أو المئيني الرائي** r-Percentile)، ويُرمز لها بـ $\mathbf{P}_r$.

2- من تعريف المئين نلاحظ أنَّ $\mathbf{P}_{25} = \mathbf{Q}_1$ ، $\mathbf{P}_{50} = \mathbf{Q}_2 = \tilde{x}$ و $\mathbf{P}_{75} = \mathbf{Q}_3$.

3- بما أنَّ عدد المئينات كبير (**لدينا 99 مئيناً**) فإنَّ هذا النوع من المقاييس يكون أكثر وضوحاً كلما ازداد عدد البيانات، وتكون جميع المئينات قابلةً للحساب وأكثر دقَّة عندما يكون عدد البيانات أكبر أو يساوي 99، وما سنقدِّمه من أمثلة على بيانات قليلة العدد سيكون من باب التوضيح والتبسيط.

2-3-3- تعيين المئينات

إنَّ طريقة تعيين المئينات تُماثل طريقة تعيين الربيعيَّات تماماً، وذلك باتباع الخطوات الآتية:

الخطوة الأولى: بفرض أنَّ x_1 ، x_2 ، ... و x_n بيانات **مرتبة تصاعدياً**، فإنّنا نضع بالتعريف:

$$p_r = \frac{r\ (n+1)}{100} \quad ; r = 1, 2, ..., 99 \qquad [2\text{-}8\text{-}a]$$

و هذه القيمة p_r تُدعى **رتبة** المئين r في البيانات المرتبة تصاعدياً، و عملها تحديد موضع هذا المئين بين البيانات المرتّبة تصاعدياً. عندئذٍ لتعيين قيمة المئين $\mathbf{P}_r$ نقوم بما يلي:

الخطوة الثانية: بفرض أنَّ k هو الجزء الصحيح من المقدار p_r و أنَّ الباقي منه يساوي s، فعندئذٍ تُحسب قيمة المئين $\mathbf{P}_r$ من خلال العلاقة الآتية:

$$\mathbf{P}_r = x_k + s(x_{k+1} - x_k) \quad ; r = 1, 2, ..., 99 \qquad [2\text{-}8\text{-}b]$$

نلاحظ هنا أنَّه إذا كان $s = 0$ أو $x_{k+1} = x_k$ فإنَّه سيكون لدينا $\mathbf{P}_r = x_k$ فقط.

2-3-3-1- أمثلة

1- تقدَّم لنا البيانات الآتية الطول لخمسين طالباً في عمادة السّنة الأولى المشتركة.

185	169	173	179	173	178	147	171	173	172
169	201	184	160	163	189	190	158	149	173
172	149	177	158	169	172	171	168	181	178
195	181	171	195	190	167	185	173	180	179
169	181	172	196	152	175	179	169	175	175

فما هي قيم المئينات $\mathbf{P}_{15}$ ،$\mathbf{P}_{43}$ و $\mathbf{P}_{87}$ الخاصّة بهذه البيانات؟

الإجابة: من أجل تعيين قيم المئينات علينا أولاً ترتيب البيانات المُعطاة تصاعدياً، فنجد لها التسلسل الآتي:

147	158	169	171	172	173	177	179	184	190
149	160	169	171	172	173	178	180	185	195
149	163	169	171	173	175	178	181	185	195
152	167	169	172	173	175	179	181	189	196
158	168	169	172	173	175	179	181	190	201

البيانات بعد ترتيبها تصاعدياً

وبما أنَّ $n = 50$ فإنَّ رتبة المئين $\mathbf{P}_{15}$ ،$\mathbf{P}_{43}$ و $\mathbf{P}_{87}$ ستكون على الترتيب هي:

$$p_{15} = \frac{15(n+1)}{100} = \frac{15(50+1)}{100} = 7.65$$

$$p_{43} = \frac{43(n+1)}{100} = \frac{43(50+1)}{100} = 21.93$$

$$p_{87} = \frac{87(n+1)}{100} = \frac{87(50+1)}{100} = 44.37$$

ومن ثمَّ يكون لدينا:

$$\mathbf{P}_{15} = x_k + s(x_{k+1} - x_k) = x_7 + 0.65(x_8 - x_7)$$
$$= 160 + 0.65(163 - 160) = 161.95$$

$$\mathbf{P}_{43} = x_{(k)} + s(x_{(k+1)} - x_{(k)}) = x_{21} + 0.93(x_{22} - x_{21})$$
$$= 172 + 0.93(172 - 172) = 172$$

$$\mathbf{P}_{87} = x_k + s(x_{k+1} - x_k) = x_{44} + 0.37(x_{45} - x_{44})$$
$$= 189 + 0.37(190 - 189) = 189.37$$

2- عيّن قيم المَئينات $\mathbf{P}_{25}$ ، $\mathbf{P}_{50}$ و $\mathbf{P}_{75}$ والخاصّة ببيانات المثال (2) من (2-2-2-1).

الإجابة: من أجل تعيين قيم المئينات علينا أولاً ترتيب البيانات المُعطاة تصاعدياً، فنجد لها التسلسل الآتي:

65	77	86	98
66	77	87	98
70	77	88	102
73	78	92	106
76	85	95	115

الآن، وبما أنَّ $n = 20$ فإنَّ رتبة المَئين $\mathbf{P}_{25}$ ، $\mathbf{P}_{50}$ و $\mathbf{P}_{75}$ ستكون على الترتيب هي:

$$p_{25} = \frac{25(n+1)}{100} = \frac{25(20+1)}{100} = 5.25$$

$$p_{50} = \frac{50(n+1)}{100} = \frac{50(20+1)}{100} = 10.5$$

$$p_{75} = \frac{75(n+1)}{100} = \frac{75(20+1)}{100} = 15.75$$

ومن ثمَّ يكون لدينا:

$$\mathbf{P}_{25} = x_5 + 0.25\,(x_6 - x_5) = 76 + 0.25(77 - 76) = 76.25$$

$$\mathbf{P}_{50} = x_{10} + 0.5\,(x_{11} - x_{10}) = 85 + 0.5(86 - 85) = 85.5$$

$$\mathbf{P}_{75} = x_{15} + 0.75(x_{16} - x_{15}) = 95 + 0.75(98 - 95) = 97.25$$

وبالرجوع إلى قيم الربيعيَّات الثلاثة التي قمنا بحسابها في ذلك المثال $\mathbf{Q}_1 = 76.25$ ، $\mathbf{Q}_2 = 85.5$ و $\mathbf{Q}_3 = 97.25$ نجد تطابق نتائجها مع قيم المئينات $\mathbf{P}_{25}$ ، $\mathbf{P}_{50}$ و $\mathbf{P}_{75}$ بشكل تام.

3- لتكن لدينا البيانات الآتي والتي تمثِّل درجات 23 طالباً في مقرَّر مبادئ في الإحصاء والاحتمالات (الدرجة التامّة 100):

98	88	75	69	53	91	81	73	68	36	95	86
75	68	49	88	77	71	65	89	79	72	66	

والمطلوب ما يلي:

أ- ما هي الدرجة التي لن يتجاوزها 40% من الطلاب؟

ب- ما هي الدرجة التي سيتجوزها 15% من الطلاب فقط؟

الإجابة: من أجل الإجابة على الأسئلة السابقة سنقوم بترتيب البيانات المُعطاة تصاعدياً، فنجد لها التسلسل الآتي:

36	68	73	81	91
49	68	75	86	95
53	69	75	88	98
65	71	77	88	
66	72	79	89	

الآن من أجل الفقرة:

أ- إنَّ الدرجة التي لن يتجاوزها 40% من الطلاب هي قيمة المئين $\mathbf{P}_{40}$ ، ورتبته تساوي:

$$p_{40} = \frac{40(23+1)}{100} = 9.6$$

ومن ثمَّ يكون لدينا:

$$\mathbf{P}_{40} = x_9 + 0.6\,(x_{10} - x_9) = 71 + 0.6(72 - 71) = 71.6$$

وبما أنَّ القيم العشرية للدرجات تجبر بالزيادة إلى العدد الصحيح التالي فإنَّ 60% من الطلاب سيكون لهم درجات أعلى 72.

ب- إنَّ الدرجة التي سيتجاوزها 15% من الطلاب فقط هي قيمة المئين $\mathbf{P}_{85}$ ، ورتبته تساوي:

$$p_{85} = \frac{85(23+1)}{100} = 20.4$$

ومن ثمَّ يكون لدينا:

$$\mathbf{P}_{85} = x_{20} + 0.4\,(x_{21} - x_{20}) = 89 + 0.4(91 - 89) = 89.8$$

وبالتالي فإنَّ 15% من الطلاب فقط سيكون لهم درجات أعلى 90.

2-4- الأعداد الخمسة والتمثيل الصندوقي للبيانات
Five Numbers and the Box Plot Representation of Data

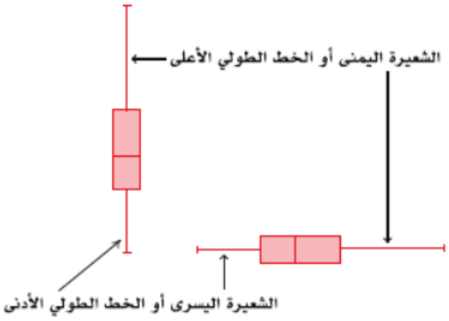
يقصد بالأعداد الخمسة في الإحصاء الوصفي القيم الآتية:

1- أصغر قيمة في البيانات x_s،

2- أكبر قيمة في البيانات x_ℓ،

3- قيم الربيعيَّات الثلاثة Q_1، Q_2 وQ_3.

إنَّ هذه الأعداد الخمسة تساعدنا في وصف تمركّز البيانات وانتشارها وشكل توزيعها، وأمّا العرض (أو التمثيل) الصندوقي Box Plot للبيانات فهو عرض رسومي للبيانات يقوم على أساس مُلخَّص الأعداد الخمسة السابقة، ويتكوَّن من صندوق وخطين طوليين يتوسطانه على طرفيه يُمنةً ويُسرى (أو من الأعلى والأدنى) يُدعيان "الشُعَيْرتان Whiskers". ويُقدَّم هذا العرض وفقاً لأحد الشكلين الآتيين:



الشكل [2-8]

أمَّا نهاية الشُعَيْرة اليمنى (وسنرمز لها بـ x_e) واليسرى (وسنرمز لها بـ x_a) فإنَّهما يعيّنان كما يلي:

1- إذا كانت مجموعة البيانات لا تحتوي قيم متطرفة، فعندئذٍ يكون $x_a = x_s$ و $x_e = x_\ell$.

2- إذا كانت مجموعة البيانات تحتوي على قيم متطرفة بصغرها فقط، فعندئذٍ يكون لدينا $x_a = \mathbf{LF}$. أي أنَّ نهاية الشُعيرة اليسرى x_a توافق الحد الأعلى للقيم المتطرفة بصغرها، ومن ثمَّ تبلغ الشُعيرة اليسرى الطول الأعظمي لها. أمَّا x_e فيكون مساوياً لـــ x_ℓ في هذه الحالة.

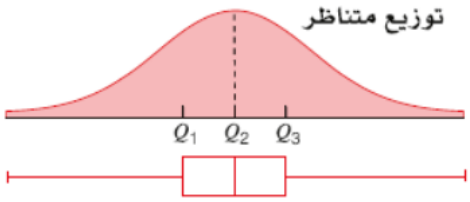
3- إذا كانت مجموعة البيانات تحتوي على قيم متطرفة بكبرها فقط، فعندئذٍ يكون لدينا $x_e = \mathbf{HF}$. أي أنَّ نهاية الشُعيرة اليمنى x_e توافق الحد الأدنى للقيم المتطرفة بكبرها، ومن ثمَّ تبلغ الشُعيرة اليمنى الطول الأعظمي لها. أمَّا x_a فيكون مساوياً لـــ x_s في هذه الحالة.

4- إذا كانت مجموعة البيانات تحتوي على قيم متطرفة بصغرها وأخرى متطرفة بكبرها، فعندئذٍ يكون لدينا $x_a = \mathbf{LF}$ و $x_e = \mathbf{HF}$، وفي هذه الحالة تبلغ الشُعيرتان اليمنى واليسرى الطول الأعظمي لهما.

2-4-1- ملاحظات

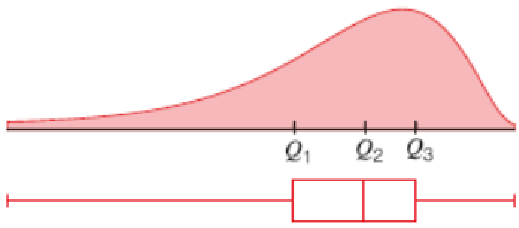
1- عند تقديم الرسم الصندوقي يُشار إلى كل قيمة متطرفة بنجمة (*) أو نقطة (•).

2- إذا كان توزيع البيانات متناظراً، فعندئذٍ سيتمركز الصندوق وخط الوسيط بين نقطتي النهاية للشُعيرّتين (انظر الشكل التوضيحي الآتي).

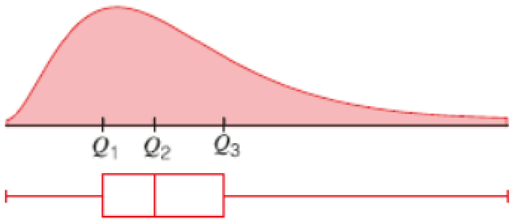


الشكل [2-9-a]

3- إذا كان توزيع بيانات مستمرة (**في جداول توزيع تكرارية، أو من خلال العروض البيانية الموافقة لها**) ملتوياً نحـو اليســار فعنـدئذٍ سـينزاح الصنـدوق لليمين باتجـاه نهايـة الشُـعيرة اليمنى (**انظر الشكل [2-9-b]**)، وأمّا إذا كان توزيع البيانات ملتوياً نحو اليمين فإنَّ الصندوق سينزاح إلى اليسار باتجاه نهاية الشُعيرة اليسرى (**انظر الشكل [2-9-c]**). أمَّا خط الوسيط فيمكن له أخذ أية وضعية داخل الصندوق في هاتين الحالتين.



الشكل [2-9-b]



الشكل [2-9-c]

بهذا يمكننا استخدام العرض الصندوقي للبيانات لمعرفة إن كان توزيع البيانات ملتوياً أم لا.

2-4-2- أمثلة

1- لتكن لدينا مجموعة البيانات الآتية:

8, 4, 7, 6, 9, 1, 5, 9, 3, 17, 5, -7

والمطلوب تعيين الأعداد الخمسة لهذه المجموعة من البيانات، ومن ثمَّ رسم المخطط الصندوقي لها.

الإجابة: من أجل تعيين الأعداد الخمسة سنقوم أولاً بترتيب البيانات تصاعدياً فنجد لها العرض الآتي:

-7, 1, 3, 4, 5, 5, 6, 7, 8, 9, 9, 17

ومن ثمَّ تكون الأعداد الخمسة لهذه البيانات هي:

$$x_s = -7 \quad \& \quad x_\ell = 17 \quad \& \quad \mathbf{Q}_1 = 3.25 \quad \& \quad \tilde{x} \ [illegible] \ 5 \quad \& \quad \mathbf{Q}_3 = 8.75$$

وأمّا من أجل تحديد القيم المتطرفة بصغرها والتي تكون أصغر من نهاية الشُعيرة اليسرى، فلدينا:

$$\mathbf{LF} = 3.25 - 1.5(8.75 - 3.25) = -5$$

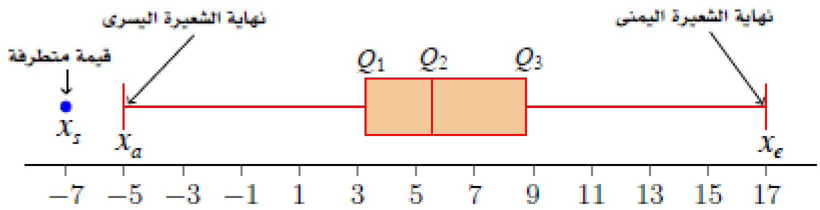
أي إنَّ القيم المتطرفة بصغرها هي تلك القيم التي تكون أصغر من 5- ، ومن ثمَّ لدينا قيمة متطرفة بصغرها في البيانات المقدَّمة هي $x_s = -7$.

كذلك من أجل تحديد القيم المتطرفة بكبرها التي تكون أكبر من نهاية الشُعيرة اليمنى لدينا:

$$\mathbf{HF} = 8.75 + 1.5(8.75 - 3.25) = 17$$

أي إنَّ القيم المتطرفة بكبرها هي تلك القيم التي تكون أكبر من 17، ومن ثمَّ لا وجود لقيمة متطرّفة بكبرها في البيانات المقدَّمة.

هكـذا نجـد أنَّ الشُـعيرتان اليمنـى واليسـرى تبلغـان الطـول الأعظـم لهمـا حيـث لـدينا $x_a = \mathbf{LF} = -5$ و $x_\ell =$ [illegible] $= \mathbf{HF} = 17$، ومن ثمَّ يكون التمثيل الصندوقي للبيانات المُعطاة العرض الآتي:



الشكل [2-10]

2- لتكن لدينا مجموعة البيانات الآتية:

$$2,\ 5,\ 7,\ 4,\ 8,\ 9,\ 4,\ 8,\ 5,\ 6,\ 15,\ 4$$

والمطلوب تعيين الأعداد الخمسة لهذه المجموعة من البيانات، ومن ثمَّ رسم المخطط الصندوقي لها.

الإجابة: من أجل تعيين الأعداد الخمسة سنقوم أولاً بترتيب البيانات تصاعدياً فنجد لها العرض الآتي:

$$2,\ 4,\ 4,\ 4,\ 5,\ 5,\ 6,\ 7,\ 8,\ 8,\ 9,\ 15$$

ومن ثمَّ تكون الأعداد الخمسة لهذه البيانات هي:

$x_s = 2$ & $x_\ell = 15$ & $\mathbf{Q}_1 = 4$ & [illegible] 5 & $\mathbf{Q}_3 = 8$

وأمَّا من أجل تحديد القيم المتطرفة بصغرها فلدينا:

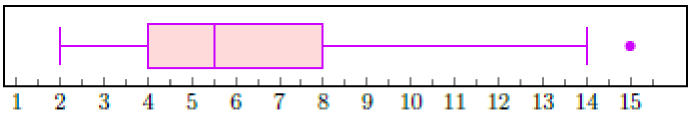
$$\mathbf{LF} = 4 - 1.5(8-4) = -2$$

ومن ثمَّ لا وجود لقيم متطرفة بصغرها في البيانات المُعطاة، ولذلك سيتوقف الطرف الأيسر من الشُعيرة اليسرى عند أصغر قيمة للبيانات.

كذلك من أجل تحديد القيم المتطرفة بكبرها لدينا:

$$\mathbf{HF} = 8 + 1.5(8-4) = 14$$

ومن ثمَّ لدينا قيمة متطرفة بكبرها في البيانات المقدَّمة وهي $x_\ell = 15$، وهكذا نجد أنَّ الشُعيرة اليمنى تبلغ الطول الأعظم لها، وبالتالي يصبح للتمثيل الصندوقي للبيانات المُعطاة الشكل الآتي.



الشكل [2-11]

تمارين الفصل الثاني

1- قام شخص بشراء أربعةِ أنواعٍ مختلفةٍ من التفاح بأسعار متباينة للكيلو غرام الواحد هي: 5.95، 6.95، 7.95 و6.25. فما هو متوسط سعر الكِيلو غرام الواحد من التفاح الذي اشتراه هذا الشخص؟

2 - بفرض أنَّ 1، 3، 5، 7، 2، 3، 6، 3، 2 و 8 تمثِّل أوزان عشرة صناديق من الفاكهة (مقدّرةَ بالكيلوغرام)، فعندئذٍ احسب المتوسط والوسيط، ومن ثمَّ عيّن المنوال (إن وجِد) لهذه البيانات؟

3- الجدول الآتي يقدِّم لنا درجات خمسين طالباً في أحد المقرّرات الدراسية:

أعداد الطلاب الحاصلين على هذه الدرجة التكرار	نطاق الدرجات الفئات الفعليّة
3	$49.5 \to 59.5$
12	$59.5 \to 69.5$
10	$69.5 \to 79.5$
17	$79.5 \to 89.5$
8	$89.5 \to 99.5$
50	**المجموع**

والمطلوب ما يلي:

أ- تعيين الفئات المنوالية إن وجدت وحساب قيم هذه المناويل.

ب- تعيين الفئة الوسيطية وحساب قيمة الوسيط.

ج- حساب المتوسط لبيانات هذا الجدول.

4- إذا كان لدينا خمس مشاهدات بمتوسط $\bar{x} = 10$ ، فما هو مجموع هذه المشاهدات؟

5- إذا كان لدينا عينة متوسطها $\bar{x} = 10$ ومجموع بياناتها 100، فما هو عدد البيانات في هذه العينة؟

6- إذا كان لدينا مشاهدات x_1 ، x_2 ، x_3 و x_4 متوسطها $\bar{x} = 6$ ، فما هو متوسط المشاهدات الآتية:

$$\frac{x_1}{4} + 0.5 \quad \frac{x_4}{4} + 0.5 \quad \text{و} \quad \frac{x_3}{4} + 0.5 \;،\; \frac{x_2}{4} + 0.5 \;،$$

7- لتكن لدينا مجموعات البيانات الآتية (علماً أنَّ ? تشير إلى القيم المفقودة في البيانات):

a) 2, 4, 4, 7, 7, 7, ?, 8, 10, 13, 15, 19, 21
b) 1, 4, 5, 7, ?, 8, 13, 13, 15, ?, 22, 25, ?
c) 20, 24, 31, 27, 28, 30, 23, 25
d) 20, 23, 21, 121, 28, 30, − 20, 28
e) 20, 21, 21, 37, 18, − 25, 23, 35
f) 4, 5, 7, 9, ?, ?, 11, 15, 18, 23

ما هو مقياس النزعة المركزيّة المناسب والأفضل لكل من البيانات السابقة، ولماذا؟ ومن ثمَّ عيّن قيمته.

8- ما هو المنوال (في حال وجوده) لكل من مجموعات المشاهدات الآتية:

أ- مجموعة المشاهدات 3، 1، 3، 5، 7، 11، 6، 9 و 8

ب- مجموعة المشاهدات 7، 5، 1، 6، 11، 17، 12 و 21

ج- مجموعة المشاهدات 3، 3، 3، 3، 3 و 3

د- مجموعة المشاهدات 1، 3، 5، 1، 3، 5 ، 1، 3 و 5

9- أي مقياسٍ من مقاييس النزعة المركزيّة يمكن استخدامه مع فصائل الدم O، A، B وAB لمجموعة من الأشخاص.

10- سجَّل أحد الطلاب في خمسة مقرّرات لفصل دراسي، وكانت الدرجات التي حصل عليها في كل مقرَّر والأوزان الموافقة لها مقدَّمة في الجدول الآتي.

المقرَّر	O	N	M	L	K
الدرجة	95	73	84	94	85
الوزن	1	2	3	3	2

والمطلوب حساب المعدَّل التراكمي (المتوسط الموزون) للطالب في ذلك الفصل؟

11- لتكن لدينا البيانات الآتية لعينة عشوائية:

4, 7, 5, 4, 8, 5, 17, 8, 7, 13, 2, 4

والمطلوب ما يلي:

أ- تعيين المنوال (أو المناويل في حال وجودها)، ومن ثمَّ حساب المتوسط وبعد ذلك الوسيط أيضاً.

ب- احسب الربيعي الأول والثالث لهذه البيانات.

ج- احسب المئين الــ 65 لهذه البيانات.

د- بيّن فيما إذا كانت توجد قيم متطرفة في البيانات المُعطاة أم لا.

ه- تقديم الرسم الصندوقي لهذه البيانات موضِّحاً عليه جميع المسمّيات والرموز.

و- بناءً على شكل المخطط الصندوقي للبيانات. هل تلاحظ وجود التواء في توزيع البيانات وإلى أية جهة؟

12- لدى معاينة أوزان 20 طفلاً أعمارهم دون العام الواحد وجدنا البيانات الآتية (مقدَّرةً بالكيلو غرام):

10.2	7.7	11.5	7.0
8.8	7.7	10.6	7.3
6.5	7.8	9.8	8.5
9.2	7.7	9.8	8.7
7.6	6.6	9.5	8.6

والمطلوب ما يلي:

أ- تعيين الربيعيَّات الثلاثة Q_1 و Q_2 و Q_3.

ب- تعيين القيم المتطرفة إن وجدت.

ج- تقديم الرسم الصندوقي لهذه البيانات موضِّحاً عليه جميع المسمّيات والرموز.

د- بناءً على شكل المخطط الصندوقي للبيانات. هل تلاحظ وجود التواء في توزيع البيانات وإلى أية جهة؟

13- تقدَّم لنا البيانات الآتية الطول لثلاثين شخصاً بالغاً (مقدَّرةً بالسنتميتر).

178	181	120	171	172	169	158	177	147	172
173	149	158	195	189	163	160	184	201	169
172	173	171	147	178	173	179	218	169	185

والمطلوب ما يلي:

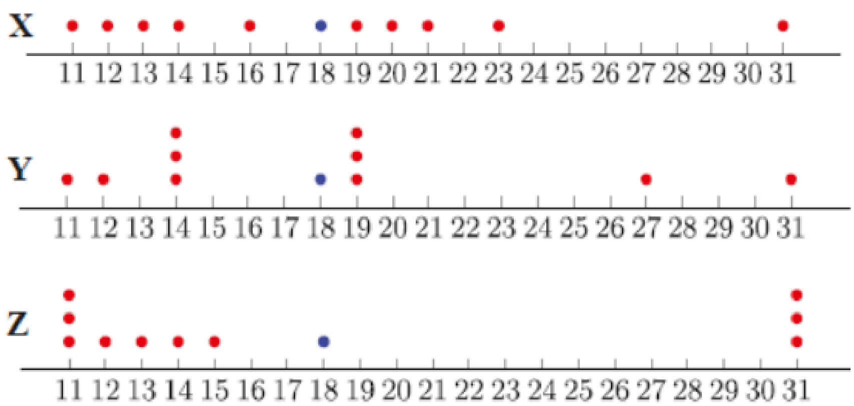
أ- تعيين الربيعيَّات الثلاثة Q_1 و Q_2 و Q_3.

ب- تعيين القيم المتطرفة إن وجدت.

ج- تقديم الرسم الصندوقي لهذه البيانات موضِّحاً عليه جميع المسمّيات والرموز.

د- بناءً على شكل المخطط الصندوقي للبيانات. هل تلاحظ وجود التواء في توزيع البيانات وإلى أية جهة؟

14- لتكن لدينا مجموعات البيانات الموضَّحة في العروض الآتية:



والمطلوب ما يلي:

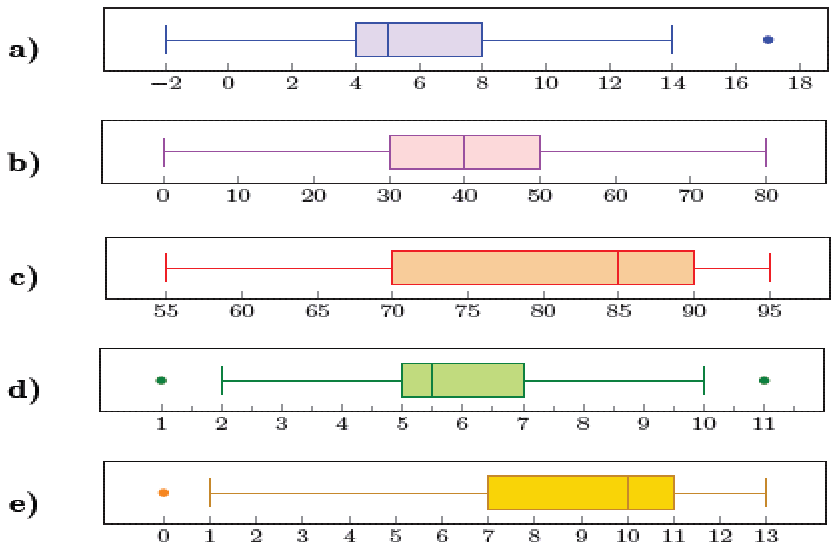
أ- ماذا تمثِّل النقطة الزرقاء على الرسم المقدَّم؟

ب- أي من هذه البيانات تملك منوالاً (أو مناويل)، ثمَّ عيّنها في حال وجودها؟

ج- أيها أكثر تبعثراً حول المتوسط، ولماذا؟

د- عيّن موضع الوسيط لكل منها على الرسم.

15- لدينا مجموعات لبيانات مقدَّمة من خلال المخططات الصندوقية الآتية:



والمطلوب ما يلي:

أ- عيِّن الأعداد الخمسة الخاصة بكل مخطط من هذه المخططات.

ب- أي من هذه الأشكال يدلّ على توزيع متناظر؟

ج- أي من هذه الأشكال يدلّ على توزيع ملتوٍ نحو اليمين؟

د- أي من هذه الأشكال يدلّ على توزيع ملتوٍ نحو اليسار؟

هـ- أي من هذه الأشكال يشير إلى وجود قيم متطرفة بكبرها؟

و- أي من هذه الأشكال يشير إلى وجود قيم متطرفة بصغرها؟

الفصل الثالث

مقاييس الاختلاف للبيانات

Variability Measures of Data

مقدِّمة:

لقد ذكرنا في الفصل السّابق أنَّه من المقاييس الضّروريّة للتعرّف على سلوك البيانات الإحصائيّة هي تلك المقاييس التي تهتمّ بتبعثر البيانات حول مقياس نزعتها المركزية (**وعلى وجه الخصوص حول متوسطها**)، وذلك لأنَّه قد يكون لدينا مجموعتي بيانات أو أكثر من ذات الطبيعة (**لها وحدة القياس نفسها**) ولها قيمة مقياس النزعة المركزيّة نفسه، ولكن تبعثُر البيانات حول مقاييس نزعتها المركزيّة مختلف من مجموعة بيانات إلى أخرى، وبالتالي استخدام مقياس النزعة المركزيّة من أجل المقارنة بين سلوك هذه المجموعات من البيانات يصبح غير ذي جدوى. هذا من جانب، ومن جانبٍ آخر فقد يكون لقيم البيانات في تلك المجموعات وحدات قياس مختلفة، ومن ثمَّ عمليّة المقارنة بين سلوك هذه المجموعات من البيانات يصبح غير ذي جدوى أيضاً حتى لو علمنا كيفية تبعثر البيانات حول مقياس نزعتها المركزيّة لتلك المجموعات من البيانات. لذلك كان لا بدَّ من تقديم مقاييس تساعدنا في مقارنة سلوك البيانات حتى في حال اختلاف وحدات القياس لبيانات تلك المجموعات من البيانات. إنَّ المقاييس التي تساعدنا في مقارنة سلوك البيانات حتى في حال اختلاف وحدات القياس تندرج تحت اسم " **مقاييس الاختلاف للبيانات**".

3-1- مقاييس التشتت

3-2- مُعاملات من أجل مقارنة التشتت لمجموعتي بيانات أو أكثر

3-3- الدرجة المعيارية Z

3-1- مقاييس التّشتت

Dispersion Measures

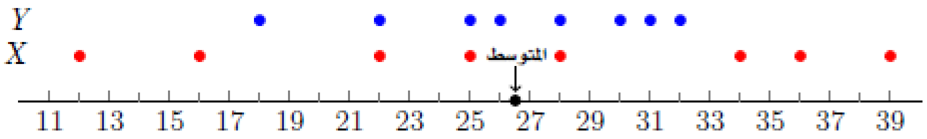
كما ذكرنا في مقدِّمة هذا الفصل أنَّه من الممكن أن يكون لدينا مجموعتي بيانات أو أكثر من ذات الطبيعة (لها وحدة القياس نفسها) ولها قيمة مقياس النزعة المركزيّة نفسه، ولكن تبعثّر البيانات حول مقاييس نزعتها المركزيّة مختلف من مجموعة بيانات إلى أخرى، ولتوضيح ذلك سنأخذ المثال الآتي.

ليكن X وY متغيرين راصدين لدرجات الحرارة في مدينتين A وB على الترتيب، علماً أنَّ قياس درجة الحرارة (مقدَّرةً بالدرجات المئوية Celsius) قد بدأ الساعة السادسة صباحاً، وأُخِذ قياساً كل ثلاث ساعاتٍ ليومٍ كامل.

الجدول [3−1]

الوقت	3	24	21	18	15	12	9	6
قيم المتغيّر X	16	22	28	34	36	39	25	12
قيم المتغيّر Y	18	26	28	30	31	32	25	22

فنجد أنَّ متوسط درجات الحرارة في كلا المدينتين خلال هذا اليوم يساوي 26.5 إلا أنَّ تبعثر قيم درجات الحرارة حول متوسطها يختلف من مدينة إلى أخرى حيث نلاحظ أنَّ بيانات المتغيّر Y قريبة من المتوسط على خلاف بيانات المتغيّر X التي تتناثر مبتعدةً عن المتوسط (انظر الشكل الآتي).



الشكل [3-1]

ممّا سبق يتبيّن لنا ضرورة وجود مقاييس تُحدّد لنا بدقة كيفية تبعثر البيانات حول مقياس نزعتها المركزيَّة ليتكوَّن لدينا انطباعاً أكثر وضوحاً حول سلوك البيانات.

3-1-1- تعريف (مقياس التشتت) (Measure of Dispersion)

إنَّ المقياس الذي يهتمّ بتحديد مقدار كمّي لقياس مدى تبعثر البيانات حول مقياس نزعتها المركزية يُدعى **مقياساً للتّشتت**.

نقدم فيما يلي بعض مقاييس التشتت.

3-1-2- الانحراف المعيّاري Standard Deviation

لقد لوحظ أنَّه يمكن النظر إلى قيم الفروقات بين البيانات ومتوسطها كمقياس للتّشتت، ولكنَّ المجموع الجبري لهذه الفروق يساوي الصّفر دوماً، ولذلك طُرِحَت فكرة استخدام مربّعات قيم تلك الفروقات كمقياس للتّشتت، فكانت فكرة تقديم الانحراف المعيّاري الذي سنمهِّد له من خلال التعريف الآتي.

3-1-2-1- تعريف (التّباين) (Variance)

لتكن لدينا بيانات عيِّنة، فعندئذٍ لتعريف **التّباين** لهذه البيانات (ويُرمز له بـ S^2)، ويدعى بالتباين العملي Empirical Variance أيضاً) سنميِّز بين الحالتين الآتيتين:

1- إذا كانت البيانات خام من قبيل x_1 ، x_2 ، ... و x_n بمتوسط $\overline{X}$، فإنَّ **التّباين** يُعطى من خلال العلاقة الآتية:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2 \qquad [3-1-a]$$

وبالتعويض في العلاقة السابقة عن قيمة المتوسط $\overline{X}$ بما يساويها بدلالة المجموع وعدد البيانات، ومن ثمَّ إجراء بعض التعديلات على العلاقة يصبح للعلاقة السابقة العرض الآتي:

$$S^2 = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right] \qquad [3\text{–}1\text{-}b]$$

وهذه العلاقة تستخدم قيم البيانات مباشرةً دون اللجوء إلى حساب المتوسط أو معرفة قيمته.

2- إذا كانت البيانات **كميّة** ومجمَّعة في جدول تكراري لـ m ممثِّل كما في الجدول [1-1]، وكان متوسط بيانات الجدول يساوي $\overline{X}$، فإنَّ قيمة **التّباين** لبيانات ذلك الجدول تُعطى من خلال العلاقة الآتية:

$$S^2 = \frac{1}{\left(\sum_{i=1}^{m} f_i\right) - 1}\sum_{i=1}^{m} f_i (x_i - \overline{X})^2 \qquad [3\text{–}2]$$

علماً أنَّ x_i و f_i هما قيمة الممثِّل i وتكراره على الترتيب.

3- إذا كانت البيانات مجمَّعة في جدول توزيع تكراري بـ k فئة وكان متوسطها $\overline{X}$، فإنَّ قيمة **التّباين** تُعطى من خلال العلاقة الآتية:

$$S^2 = \frac{1}{\left(\sum_{i=1}^{k} f_i\right) - 1}\sum_{i=1}^{k} f_i (x_i - \overline{X})^2 \qquad [3\text{–}3]$$

علماً أنَّ x_i و f_i هما مركز وتكرار الفئة i على الترتيب.

3-1-2-2- تعريف (الانحراف المعيّاري Standard Deviation)

يُعرف **الانحراف المعيّاري** على أنَّه الجذر التّربيعي الموجب للتباين، ويُرمز له بـ S، أي أنَّه لدينا:

$$S := +\sqrt{S^2} \qquad [3\text{–}4]$$

وتجدر الإشارة هنا إلى أنَّ **الانحراف المعيّاري** يَستخدم وحدة قياس البيانات نفسها.

3-1-2-3- ملاحظات

1- إنَّ استخدام رمز التّربيع فوق الرّمز S للدّلالة على أنَّ التّباين هو مقدار غير سالب، وأنَّ القيمة النّاتجة عنه تقرأ بالوحدة المربّعة لوحدة القياس المستخدمة في البيانات، فعلى سبيل المثال إذا كانت وحدة القياس للبيانات هي الكيلو غرام فإنَّ قيمة التّباين لهذه البيانات تقرأ بالكيلو غرام المربّع، ولهذا السبب لا يستخدم التّباين بحد ذاته كمقياس للتّشتت.

2- إنَّ مجموع مربعات انحرافات قيم البيانات x_1 ، x_2 ، ... و x_n عن أي عدد حقيقي a يكون أصغرياً عندما يكون $a = \overline{X}$، ولهذا السبب يُنظر إلى **الانحراف المعيّاري** (**النّاتج عن مفهوم التّباين**) على أنَّه أفضل مقياس للتّشتت.

3-1-2-4- أمثلة

1- ليكن X متغيّراً راصداً لدرجات 10 طلاب في اختبارٍ قصيرٍ (**من 10 درجات**):

6 7, 4, 9, 7, 8, 10, 9, 3, 7,

والمطلوب حساب الانحراف المعيّاري لدرجات الطلاب.

الإجابة: من أجل تبسيط عملية الحساب وكسب الوقت في الإجابة على مثل هذه المسائل يُفضَّل بناء جدول حسابات على النحو الآتي إذا كنّا سنستخدم المتوسط في حساب الانحراف المعيّاري:

الجدول [3–2]

i	x_i	$x_i - \overline{X}$	$(x_i - \overline{X})^2$
1	7	0	0
2	4	-3	9
3	9	2	4
4	7	0	0
5	8	1	1
6	10	3	9
7	9	2	4
8	3	-4	16
9	7	0	0
10	6	-1	1
Total	**70**	**0**	**44**

فنجد أنَّ متوسط درجات الطّلاب $\overline{X}$ يساوي:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{7+4+9+\cdots+7+6}{10} = \frac{70}{10} = 7$$

ومن أجل **الانحراف المعيّاري** لدرجات الطلاب علينا القيام أولاً بحساب التّباين لبيانات درجات الطّلاب فنجد باستخدام العلاقة [3–1-a] أنَّ:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2$$

$$= \frac{(7-7)^2 + (4-7)^2 + \ldots + (7-7)^2 + (6-7)^2}{9}$$

$$= \frac{44}{9} = 4.\overline{8}$$

ومنه ينتج لدينا أنَّ الانحراف المعيّاري لدرجات الطّلاب يساوي:

$$S = +\sqrt{S^2} = +\sqrt{4.9} = 2.21$$

2- عندما نتسوّق في أحد المتاجر للمواد التّموينيّة لاحظنا أنَّه كُتِبَ على إحدى السلع عبارة "الوزن الصافي ... كغ ± … كغ"، وللتّحقّق من ذلك قمنا بوزن 15 قطعة من هذه السّلعة فوجدنا القيم الآتية (**مقدَّرةً بـ كغ**):

4.87, 5.2, 5.1, 4.95, 4.85, 4.85, 4.75, 4.97, 5.15, 5.05, 5.2, 5.05, 5.12, 4.98, 4.91

والمطلوب تعيين الوزن الصافي الذي يُفترض كتابته على علب هذه السلعة مع تحديد قيمة ± … التي كتبت إلى جانبها.

الإجابة: إنَّ قيمة الوزن الصّافي الذي يُفترض كتابته على علب هذه السّلعة يقصد به متوسط الأوزان لجميع قطع هذه السّلعة، وأمّا العبارة ± … كغ فيقصد بها أنَّ معظم هذه السّلع يمكن لها أن تقل أو تزيد عن المتوسط بالمقدار المشار إليه وهو في الواقع قيمة الانحراف المعيّاري لأوزان هذه السّلعة، ولكن لحساب الانحراف المعيّاري يجب علينا حساب التّباين أولاً حيث سنستخدم العلاقة [3–1-b]، ولأجل ذلك سنستخدم الجدول الآتي الذي يسهّل علينا إنجاز العمليات الحسابية.

الجدول [3-3]

i	X_i	X_i^2	i	X_i	X_i^2

1	4.87	23.7169
2	5.20	27.04
3	5.10	26.01
4	4.95	24.5025
5	4.85	23.5225
6	4.85	23.5225
7	4.75	22.5625
8	4.97	24.7009
9	5.15	26.5225
10	5.05	25.5025
11	5.20	27.04
12	5.05	25.5025
13	5.12	26.2144
14	4.98	24.8004
15	4.91	24.1081
Total	**75**	**375.268**

ومنه يكون لدينا:

$$S^2 = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]$$

$$= \frac{15(375.268) - (75)^2}{15 \times 14} = \frac{0.2682}{210} = 0.01914$$

ومن ثمَّ ينتج أنَّ قيمة الانحراف المعيّاري لأوزان هذه السّلعة يساوي:

$$S = +\sqrt{S^2} = +\sqrt{0.01914} = 0.138 \quad Kg$$

وبهذا يصبح للعبارة "الوزن الصّافي ... كغ $\pm$ L كغ" العرض الآتي:

الوزن الصافي 5 كغ $\pm$ 0.138 كغ

3- بالعودة إلى بيانات المثال (4) من (2-1-4-1) سنقوم بحساب الانحراف المعيّاري لتلك البيانات.
لقد حسبنا سابقاً قيمة المتوسط لتلك البيانات، وكان لدينا $\bar{X} = 61.90$، ومن أجل إتمام عملية الحساب للتباين يُفضَّل بناء جدول حسابات على النّحو الآتي:

الجدول [3-4]

العيادة	تكلفة المعاينة والطبابة x_i	عدد الزائرين f_i	$(x_i - \bar{X})^2$	$f_i(x_i - \bar{X})^2$
العينية	55	16	47.61	761.76
الأذن والأنف والحنجرة	45	35	285.61	9996.35
العظمية	85	20	533.61	10672.2
الباطنية	65	14	9.61	134.54
العصبية	75	15	171.61	2574.15
Total	**325**	**100**	-----------	**24139**

ومن ثمَّ يكون لدينا:

$$S^2 = \frac{1}{\left(\sum_{i=1}^{m} f_i\right) - 1}\sum_{i=1}^{m} f_i(x_i - \bar{X})^2 = \frac{24139}{100 - 1} = 243.828$$

ومنه نجد أنَّ قيمة الانحراف المعيّاري للبيانات تساوي:

$$S = +\sqrt{S^2} = +\sqrt{243.828} = 15.615$$

4- الجدول الآتي يقدَّم عدد الحوادث المروريّة خلال شهرٍ معيّنٍ لكلّ فئةٍ عمريّةٍ في مدينة ما X.

الجدول [3-5]

الفئة العمرية	عدد الحوادث
$18 \longrightarrow 24$	37
$24 \longrightarrow 30$	28
$30 \longrightarrow 36$	17
$36 \longrightarrow 42$	8
$42 \longrightarrow 48$	5
Total	**95**

والمطلوب حساب متوسط عدد الحوادث في هذه المدينة، ومن ثمَّ حساب الانحراف المعيّاري المرافق له.

الإجابة: من أجل متوسط والانحراف المعيّاري سنقدّم جدول الحسابات الآتي:

الجدول [3-6]

i	مراكز الفئات x_i	التكرارات f_i	$f_i \times x_i$	$(x_i - \overline{X})^2$	$f_i(x_i - \overline{X})^2$
1	21	37	777	44.823	1658.452
2	27	28	756	0.483	13.5247
3	33	17	561	28.143	478.4314
4	39	8	312	127.803	1022.424
5	45	5	225	299.463	1497.315
Total	------------------	**95**	**2631**	------------	**4670.147**

والآن لنحسب المتوسط، فنجد:

$$\overline{X} = \frac{1}{\sum_{i=1}^{k} f_i} \sum_{i=1}^{k} f_i \times x_i = \frac{(37 \times 21) + \cdots + (5 \times 45)}{95} = \frac{2631}{95} = 27.695$$

وأمّا لحساب التّباين فإنَّنا سنستخدم العلاقة [3-2]، فيكون لدينا:

$$S^2 = \frac{1}{\left(\sum_{i=1}^{k} f_i\right) - 1} \sum_{i=1}^{k} f_i (x_i - \overline{X})^2$$

$$= \frac{37(21 - 27.695)^2 + \ldots + 5(45 - 27.695)^2}{94}$$

$$= \frac{37(44.823)^2 + \cdots + 5(299.463)^2}{94} = \frac{4670.147}{94} = 49.68$$

وبالتالي فإنَّ قيمة الانحراف المعيّاري تساوي:

$$S = +\sqrt{S^2} = +\sqrt{49.68} = 7.05$$

3-1-2-5- مزايا الانحراف المعيّاري

1- سهل الحساب.

2- إنَّ الانحراف المعيّاري هو أفضل مقاييس التّشتت بلا منازع كما ذكرنا ذلك سابقاً، ويأخذ بالحسبان جميع قيم البيانات.

3-1-2-6- عيوب الانحراف المعيّاري

1- إنَّ الانحراف المعيّاري يتأثر بالقيم المتطرفة، وذلك لأنَّه إذا وجِدَتْ قيم متطرّفة فإنَّ المتوسط سيتأثر بها، ومن ثَمَّ سينتقل هذا التأثير على قيمة الانحراف المعيّاري أيضاً.

2- إذا فُقِدتْ إحدى أو بعض البيانات فعندئذٍ يُصبح الانحراف المعيّاري عديم الفائدة.

3-1-3- المدى كمقياس للتشتت The range as measure of dispersion

لقد لاحظنا أنَّه في حال فقدان بعض قيم البيانات فإنّه علينا البحث عن مقياس آخر للتّشتت، في الواقع وُجِدَ أنَّ المدى (والذي قُدّم تعريفه سابقاً) يقدِّم لنا بعض الحلول الجزئيّة للمشكلات التي ذكرناها سابقاً بخصوص الانحراف المعياري.

لقد قمنا في الفصل الأول بتقديم تعريف المدى لمجموعة من البيانات الخام (التعريف 1-3-1) من خلال العلاقة [2-1]، وأمَّا إذا كانت البيانات مجمَّعة في جدول تكراري أو جدول توزيع تكراري فعندئذ يُقدّم المدى كما في التعريف الآتي.

3-1-3-1- تعريف المدى لبيانات مجمعة

إذا كانت لدينا بيانات كميّة مجمَّعة في:

1- جدول تكراري كما في الجدول [1-1]، فعندئذ تُعطى قيمة **المدى** لبيانات ذلك الجدول من خلال العلاقة الآتية:

$$\mathbf{R} = x_l - x_s \qquad [3\text{-}5\text{-}a]$$

علماً أنَّ x_s و x_l هما أصغر وأكبر قيمة ممثّل للبيانات على الترتيب.

2- جدول توزيع تكراري كما في الجدول [1-10]، فعندئذ تُعطى قيمة **المدى** لبيانات ذلك الجدول من خلال العلاقة الآتية:

$$\mathbf{R} = x_k - x_1 \qquad [3\text{-}5\text{-}b]$$

علماً أنَّ x_1 و x_k هما مركز الفئة الأولى والأخيرة على الترتيب.

في الواقع يُعدّ المدى مقياساً للتّشتت في حال فقدان بعض البيانات غير الواقعة على أطراف البيانات بعد ترتيبها، أي أنَّه عندما تكون أكبر وأصغر قيمة في البيانات ليست في عداد البيانات المفقودة. لكن لا يُنظر إلى هذا المقياس على أنَّه مقياسٌ جيدٌ للتّشتت رغم أنَّه يُعطي صورةً عن مدى تشتت مجموعة من البيانات، إذ إنَّه وفي كثير من الحالات (وخاصة لدى العينات كبيرة الحجم) لا يُظهر لنا بوضوحٍ توضَّع البيانات حول متوسطها لأنَّه يعتمد على الفرق بين أكبر وأصغر قيمةٍ فقط، والمثال الآتي يوضِّح لنا ذلك.

3-1-3-2- مثال
لنأخذ مجموعتي البيانات الآتيتين:

الجدول [3-7]

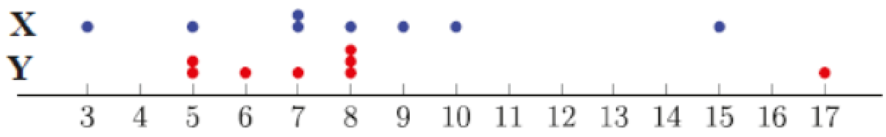
3	8	5	10	7	15	7	9	مجموعة البيانات **X**
7	8	5	8	17	6	8	5	مجموعة البيانات **Y**

فنجد أنَّ لكل من مجموعتي البيانات X و Y المتوسط نفسه $\bar{X} = \bar{Y} = 8$ والمدى نفسه أيضاً:

$$R_X = 15 - 3 = 12$$

$$R_Y = 17 - 5 = 12$$

والشكل الآتي يوضِّح لنا كيفية انتشار البيانات حول المتوسط لكل من مجموعتي البيانات X و Y



الشكل [3-2]

وأمَّا لحساب الانحراف المعيّاري فلدينا:

الجدول [3-8]

i	x_i	$(x_i - \bar{X})^2$	y_i	$(y_i - \bar{Y})^2$
1	3	25	7	1
2	8	0	8	0
3	5	9	5	9
4	10	4	8	0
5	7	1	17	81
6	15	49	6	4
7	7	1	8	0
8	9	1	5	9
Total	**64**	**90**	**64**	**104**

ومن ثمَّ نجد التّباين لمجموعة البيانات X و Y على الترتيب يساوي:

$$S_X = \sqrt{\frac{\sum_{i=1}^{8}(x_i - \bar{X})^2}{7}} = \sqrt{\frac{90}{7}} = 3.5857 \quad S_Y = \sqrt{\frac{\sum_{i=1}^{8}(y_i - \bar{Y})^2}{7}} = \sqrt{\frac{104}{7}} = 3.8545$$

وهكذا نلاحظ وجودَ تفاوتٍ واضحٍ في تبعثر البيانات حول متوسطها لكلٍّ من مجموعتي البيانات.

3-1-3-3- مزايا المدى كمقياس للتّشتت

1- إنَّه بسيط جداً وسهل الحساب.

2- يكثر استخدامه لسهولة فهمه من عامّة النّاس، حيث يستخدم لدى الإعلان عن حالات المُناخ: كدرجات الحرارة، الرطوبة والضغط الجوي.

3- له استخدامات في بعض الدراسات الإحصائية المهمّة مثل مراقبة الجودة.

3-1-3-4- عيوب المدى كمقياس للتّشتت

1- يعتمد على قيمتين فقط، ولا يأخذ جميع القيم في الحسبان، ومن ثمَّ تكون قيمته أقل دقةً من الانحراف المعيّاري.

2- يتأثر بالقيم المتطرفة بشكل كبير جدّاً لأنَّه في الأصل يعتمد على القيم التي تقع على الأطراف.

3- يصبح عديم الاستخدام إذا فقدت أصغر أو أكبر قيمة في البيانات (هنا "أو" لا تفيد الحصر).

3-1-4- المدى الربيعي Interquartile Range

بناءً على ما سبق نلاحظ أنّه لا بدَّ من البحث عن مقياس آخر يتجاوز السلبيّات للمقياس السّابق، وفي هذا الإطار نجد أنَّ ما يعرف باسم "المدى الرُبَيْعي" يقدّم لنا حلاً جزئياً آخر للمشكلات السابقة كحالة وجود قيم متطرّفة في البيانات أو فقدان أصغر أو أكبر قيمة في البيانات.

3-1-4-1- تعريف (المدى الرُبَيْعي)

لتكن x_1 ، x_2 ، ... و x_n بيانات خام لعيّنة مُعطاة، فعندئذٍ يُعرَّف **المدى الربيعي** (ويُرمز له IQR) من خلال العلاقة الآتية:

$$\mathbf{IQR} = \mathbf{Q}_3 - \mathbf{Q}_1 \qquad [3\text{–}6]$$

علماً أنَّ $\mathbf{Q}_1$ و $\mathbf{Q}_3$ هما الربيعي الأول والثّالث للبيانات على الترتيب.

لاحظ أنَّ هذا المقياس يَحسب المدى لنصف عدد البيانات الواقعة في الوسط بعد ترتيبها تصاعدياً، ومن ثمَّ فإنَّ القيم المتطرفة ستصبح خارج نطاق هذا المقياس. من جهة أخرى، فعلى الرغم من أنَّ هذا المقياس أفضل من المدى إلاَّ أنَّه يعتمد في قراره على قيمتين من البيانات فقط ولا يأخذ في الحسبان مواضع جميع البيانات، ولهذا السبب يُنظر إلى هذا المقياس على أنَّه مقياس ضعيف للتّشتت أيضاً، ولكنَّه يُعدّ مقبولاً في حال فقدان بعض البيانات المعلوم ترتيبها وغير الموافقة لقيمتي الرُبيْعيين الأول والثالث.

أخيراً نشير إلى أنَّ نصف قيمة المدى الربيعي $\mathbf{Q} = \mathbf{IQR}/2$ تدعى **الانحراف الربيعي** Quartile Deviation، وتستخدم كمقياس للتّشتت أيضاً.

3-1-4-2- أمثلة

1- لتكن لدينا البيانات الإحصائية الآتية:

8, 5, 3, 11, 5, 4, 7, 5, 3, 9, 6

والمطلوب حساب المدى الربيعي لهذه البيانات.

الإجابة: من أجل ذلك يجب حساب الربيعي الأول والثالث، وهذا يتطلب ترتيب البيانات أولاً حيث لدينا:

البيانات بعد ترتيبها	11	9	8	7	6	5	5	5	4	3	3
رموز البيانات بعد ترتيبها	x_{11}	x_{10}	x_9	x_8	x_7	x_6	x_5	x_4	x_3	x_2	x_1

وبما أنَّ $n = 11$ فإنَّ رتبة الرُبيْعي $\mathbf{Q}_1$ و $\mathbf{Q}_3$ ستكون على الترتيب هي:

$$q_1 = \frac{n+1}{4} = \frac{11+1}{4} = 3 \quad \& \quad q_3 = \frac{3(n+1)}{4} = \frac{3(11+1)}{4} = 9$$

ومن ثمَّ يكون لدينا:

$$\mathbf{Q}_1 = x_k + s(x_{k+1} - x_k) = x_3 + 0(x_4 - x_3) = 4$$

$$\mathbf{Q}_3 = x_k + s(x_{k+1} - x_k) = x_9 + 0(x_{10} - x_9) = 8$$

وهكذا يكون المدى الربيعي لمجموعة البيانات هو:

$$\mathbf{IQR} = \mathbf{Q}_3 - \mathbf{Q}_1 = 8 - 4 = 4$$

2- بالرجوع إلى بيانات الجدول [7-3] حيث لدينا $n = 8$ فإنَّنا نجد ما يلي:

أ- من أجل البيانات X: نجد للبيانات بعد ترتيبها تصاعدياً العرض الآتي:

البيانات X بعد ترتيبها	15	10	9	8	7	7	5	3
رموز البيانات بعد ترتيبها	x_8	x_7	x_6	x_5	x_4	x_3	x_2	x_1

ومن ثمَّ يكون لرتبة الرُبيْعي $\mathbf{Q}_1$ و $\mathbf{Q}_3$ القيمتين الآتيتين على الترتيب:

$$q_1 = \frac{n+1}{4} = \frac{8+1}{4} = 2.25 \quad \& \quad q_3 = \frac{3(n+1)}{4} = \frac{3(8+1)}{4} = 6.75$$

ومنه يكون لدينا:

$$\mathbf{Q}_1 = x_k + s(x_{k+1} - x_k) = x_2 + 0.25(x_3 - x_2) = 5 + 0.5 = 5.5$$

$$\mathbf{Q}_3 = x_k + s(x_{k+1} - x_k) = x_6 + 0.75(x_7 - x_6) = 9 + 0.75 = 9.75$$

وهكذا يكون المدى الربيعي لمجموعة البيانات هو:

$$\mathbf{IQR} = \mathbf{Q}_3 - \mathbf{Q}_1 = 9.75 - 5.5 = 4.25$$

ب- من أجل البيانات Y: فنجد للبيانات الترتيب التصاعدي الآتي:

البيانات Y بعد ترتيبها	17	8	8	8	7	6	5	5
رموز البيانات بعد ترتيبها	x_8	x_7	x_6	x_5	x_4	x_3	x_2	x_1

ومن ثمَّ يكون لرتبة الرُبيْعي $\mathbf{Q}_1$ و $\mathbf{Q}_3$ القيمتين الآتيتين على الترتيب:

$$q_1 = \frac{n+1}{4} = \frac{8+1}{4} = 2.25 \quad \& \quad q_3 = \frac{3(n+1)}{4} = \frac{3(8+1)}{4} = 6.75$$

ومنه يكون لدينا:

$$\mathbf{Q}_1 = x_k + s(x_{k+1} - x_k) = x_2 + 0.25(x_3 - x_2) = 5 + 0.25 = 5.25$$

$$\mathbf{Q}_3 = x_k + s(x_{k+1} - x_k) = x_6 + 0.75(x_7 - x_6) = 8 + 0 = 8$$

وهكذا يكون المدى الربيعي لمجموعة البيانات هو:

$$\mathbf{IQR} = \mathbf{Q}_3 - \mathbf{Q}_1 = 8 - 5.25 = 2.75$$

▶

بذلك نجد أنَّ مجموعة البيانات X تتبعثر حول متوسطها بشكلٍ أكبر من تبعثر مجموعة البيانات Y حول متوسطها، وهكذا نلاحظ أنَّ ما عَجِزَ عنه المدى (**في إظهار فروق التّشتت لمجموعتي البيانات**) أنجزه المدى الربيعي.

3-1-4-3- مزايا المدى الرُبيْعي

1- سهل الحساب.

2- لا يتأثّر بالقيم المتطرفة.

3- يمكن حسابه حتى في حال فقدان بعض البيانات المعلوم ترتيبها والتي لا تؤثّر على حساب الربيعي الأول والثالث.

3-1-4-4- عيوب المدى الرُبيْعي

1- لا يعدّ مقياساً دقيقاً للتشتت (**شأنه شأن المدى**) لأنّه لا يأخذُ بالحسبان جميع قيم البيانات.

2- إذا فقدت بعض البيانات المعلوم ترتيبها والتي تؤثّر على حساب الربيعي الأول أو الثالث يُصبح عديم الاستخدام.

3-2- مُعاملات من أجل مقارنة التّشتت لمجموعتي بيانات أو أكثر
Coefficient of Comparison of Dispersion of two or More Data Sets

لقد لاحظنا أنَّ مقاييس التّشتت تعتمد على وحدة القياس المستخدمة من أجل البيانات، ولذلك يصعب علينا إجراء المقارنة بين تشتت مجموعتي بيانات أو أكثر عندما تكون وحدات القياس المستخدمة من أجل كل منها مختلفة عن الأخرى، ولذلك كان لابد من وضع معيًّار يُمكِّننا من الحكم على مثل هذه المقارنات حتى في حال كانت وحدات القياس مختلفة بعضها عن البعض الآخر. يقدِّم لنا المعيًّار الآتي حلاً لهذه الإشكاليّة.

3-2-1- تعريف مُعامل التَغيُّر (أو الاختلاف) Coefficient of Variation

لتكن لدينا مجموعة بيانات عيّنة بمتوسط $\overline{X} \neq 0$ وانحراف معيًّاري S، فعندئذٍ يُعرَّف **مُعامل التَغيُّر** (**والذي يُرمز له بـ** CV) لهذه البيانات من خلال العلاقة الآتية:

$$\mathbf{CV} = \frac{S}{\overline{X}} \times 100\ \% \qquad [3-7]$$

ويقرأ النّاتج كنسبةٍ مئويّةٍ كما هو واضح.

لاحظ أنّ مُعامل التغيُّر يقدِّم لنا قيمة تجعلنا نشعر بمدى التغيّر الحاصل لمتغيّرٍ ما، ويفيدنا في مقارنة إحصائية لدرجة التّباين من سلسلة بيانات إلى أخرى حتى ولو كانت وحدات القياس تختلف بشكل كبير بعضها عن البعض الآخر، وذلك لأنَّ نسبة الانحراف المعيّاري إلى المتوسط الحسابي تلغي خاصيّة وحدة القياس المستخدمة (**فمثلاً: متر على متر، كيلوغرام على كيلوغرام، فولط على فولط أو .... حيث يختفي أثر وحدة القياس**) وإظهار مفعول التّباين بآن واحد، ومن ثمَّ تقديم هذه القيمة كنسبة مئوية خاصة بالبيانات.

أخيراً نشير إلى أنَّ |CV| (**القيمة المطلقة لمُعامل الاختلاف**) تُعرف باسم "**الانحراف المعيّاري النسبي**" Relative Standard Deviation (**RSD**)، ويُعبَّر عنها كنسبةٍ مئويّةٍ أيضاً.

3-2-1-1- مثال

لتكن لدينا مجموعتي بيانات تمثِّل الطول والوزن لستة أطفال في سن العاشرة، ومقدَّمتين من خلال الجدول الآتي:

الجدول [9-3]

الشخص	F	E	D	C	B	A
الطول X بـ سم	129	119	137	128	114	123
الوزن Y بـ كغ	33	31	34	29	31	25

فهل تشير هذه البيانات إلى أنَّ تبعثر قياسات الطول حول متوسطها أصغر من تبعثر قياسات الوزن حول متوسطها؟

الإجابة: نلاحظ أنَّ وحدة القياس بين مجموعتي البيانات مختلفة، وبالتالي لا يمكننا استخدام الانحراف المعيّاري من أجل إجراء المقارنة المطلوبة، ولذلك سنستخدم معامل الاختلاف للحصول على القرار.
الآن، بحساب قيم المتوسط والانحراف المعيّاري لكل من مجموعتي البيانات نجد أنَّ:

$$\overline{X} = 125 \ cm \qquad \& \qquad S_X = 8.12 \ cm$$

$$\overline{y} = 30.5 \ Kg \qquad \& \qquad S_Y = 3.21 \ Kg$$

وبالتالي تكون قيمة مُعامل التغيُّر لمجموعة البيانات X هي:

$$\mathbf{CV}_X = \frac{S_X}{\overline{X}} \times 100 = \frac{8.12}{125} \times 100 = 6.496 \ \%$$

في حين نجد قيمة مُعامل التغيُّر لمجموعة البيانات Y هي:

$$\mathbf{CV}_Y = \frac{S_Y}{\overline{y}} \times 100 = \frac{3.21}{30.5} \times 100 = 10.525 \ \%$$

وهكذا نجد أنَّ قيمة مُعامل تغيُّر الوزن أكبر من قيمة مُعامل تغيُّر الطول لهذه العيِّنة من الأطفال، ومن ثمَّ تبعثر قياسات الطّول حول متوسطها أقلّ من تبعثر قياسات الوزن حول متوسطها.

▶

الآن، وفي حال تعذر حساب المتوسط أو التّباين لسبب ما (**مثل فقدان بعض البيانات**) فإنَّه يجب البحث عن مُعامل آخر ينجز لنا عملية المقارنة. إنَّ المعيّار الآتي الذي يُحسب بدلالة الرّبيعيين الأول والثالث يمكن استخدامه بدلاً من مُعامل التغيُّر.

3-2-2- مُعامل التّشتت Coefficient of Dispersion

لتكن لدينا مجموعة بيانات عينة مُعطاة، فعندئذٍ يُعرَّف **مُعامل التّشتت** (**والذي يُرمز له بـ** CD) لبيانات هذه العينة من خلال العلاقة الآتية:

$$CD := \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100\ \% \qquad [3-8]$$

علماً أنَّ Q_1 و Q_3 هما الرُبيْعي الأول والثالث للبيانات على الترتيب.

مثال 3-2-2-1-

تقدَّمت مجموعة من الطلّاب لاختبارين (**مقابلة وتحريري**) في مُقرَّر دراسي، فكانت لهم النتائج الآتية التي تظهر فقدان بعض الدرجات (**المعلوم ترتيبها**) في اختبار المقابلة:

الجدول [3-10]

الطالب	J	I	H	G	F	E	D	C	B	A
X نتائج الاختبار التحريري	25	36	37	45	43	49	35	48	42	35
Y نتائج اختبار المقابلة	19	18	17	?	15	13	?	10	8	05

فهل تشير هذه المُعطيات إلى أنَّ درجات الاختبار التحريري تتبعثر حول متوسطها أكثر ممّا هو لدى اختبار المقابلة؟

الإجابة: بالطبع من كون إحدى مجموعتي البيانات تحتوي على بيانات مفقودة فليس من المنطقي أن نستخدم مُعامل التغيّر من أجل البيانات الكاملة ومُعامل التّشتت من أجل البيانات المنقوصة لكي نعطي قرارنا في عملية المقارنة، وإنَّما علينا استخدام المعيَّار نفسه للوصول إلى القرار الصحيح. لذلك سنقوم بحساب مُعامل التّشتت لكل من مجموعتي البيانات، ولأجل ذلك سنقوم أولاً بترتيب مجموعة بيانات الاختبار التحريري فقط لأنَّ بيانات اختبار المقابلة قُدِّمَت مرتبةً، فيكون لدينا العرض الآتي:

نتائج الاختبار التحريري بعد ترتيبها	49	48	45	43	42	37	36	35	35	25
رموز القيم بعد ترتيبها	x_{10}	x_9	x_8	x_7	x_6	x_5	x_4	x_3	x_2	x_1

فنجد من أجل هذه البيانات أنَّ لرتبة الرُبيْعيين Q_1 و Q_3 القيمتين الآتيتين على الترتيب:

$$q_1 = \frac{n+1}{4} = \frac{10+1}{4} = 2.75 \qquad \& \qquad q_3 = \frac{3(n+1)}{4} = \frac{3(10+1)}{4} = 8.25$$

ومن ثمَّ يكون لدينا:

$$Q_1 = x_2 + 0.75\,(x_3 - x_2) = 35 + 0.75\,(35 - 35) = 35$$

$$Q_3 = x_8 + 0.25\,(x_9 - x_8) = 45 + 0.25\,(48 - 45) = 45.75$$

وبالتالي تكون قيمة مُعامل التّشتت لبيانات الاختبار التحريري هي:

$$CD_X = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100 = \frac{45.75 - 35}{45.75 + 35} \times 100 = 13.31 \ \%$$

وأمّا من أجل رتبة الرُبيْعيين Q_1 و Q_3 الخاصّة باختبار المقابلة فنجد لهما القيمتين الآتيتين على الترتيب:

$$q_1 = \frac{n+1}{4} = \frac{10+1}{4} = 2.75 \qquad \& \qquad q_3 = \frac{3(n+1)}{4} = \frac{3(10+1)}{4} = 8.25$$

ومن ثمَّ يكون لدينا:

نتائج اختبار المقابلة (مرتَّبة)	19	18	17	?	15	13	?	10	8	5
رموز القيم المرتَّبة	y_{10}	y_9	y_8	y_7	y_6	y_5	y_4	y_3	y_2	y_1

$$Q_1 = y_2 + 0.75\,(y_3 - y_2) = 8 + 0.75\,(10 - 8) = 9.5$$

$$Q_3 = y_8 + 0.25\,(y_9 - y_8) = 17 + 0.25\,(18 - 17) = 17.25$$

ومنه نجد أنَّ قيمة مُعامل التّشتت لبيانات اختبار المقابلة تساوي:

$$CD_Y = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100 = \frac{17.25 - 9.5}{17.25 + 9.5} \times 100 = 28.97 \ \%$$

وهكذا ينتج لدينا أنَّ بيانات اختبار المقابلة تتبعثر حول متوسطها بشكل أكثر بكثير ممّا هو عليه الحال من أجل بيانات الاختبار التحريري.

▶

3-3- الدرجة المعيّارية Z
Z-Score

الآن، وقبل ختام هذا الفصل سنقدّم مفهوماً متعلّقاً بقيمة المتوسط والانحراف المعيّاري لمجموعة بيانات، ويُدعى **الدرجة المعيّارية Z** (وسنستخدم عبارة الدرجة المعيّارية على سبيل الاختصار).

من فوائد الدرجة المعيّارية أنَّها تُعطينا صورةً عن موضع البيان بالنسبة إلى متوسط البيانات، ولذلك يمكننا أن نقارن بين قيمتين لكل منهما موضع نسبيٍّ مختلف في مجموعتي بيانات مختلفتين (وقد يكون لهما قيم مختلفة للمتوسط)، فعلى سبيل المثال يمكننا مقارنة مستوى أداء طالب في جامعة ما مع مستوى أداء طالب آخر من جامعة أخرى، أو مقارنة مستوى أداء طالبٍ في مدرسةٍ حكوميّةٍ مع مستوى أداء طالبٍ آخر من مدرسةٍ أهليّة.

3-3-1- تعريف (الدرجة المعيّارية)

لتكن x_1 و x_2 و $\ldots$ x_n بيانات عيّنة بمتوسط $\overline{X}$ وانحراف معيّاري S، فعندئذٍ **الدرجة المعيّارية** Z لقيمة x_i من هذه البيانات (والتي سنرمز لها بـ Z_{x_i}) تُعرَّف من خلال العلاقة الآتية:

$$Z_{x_i} = \frac{x_i - \overline{X}}{S} \quad ; i = 1, 2, \ldots, n \qquad [3\text{–}9]$$

إنَّ العملية التي نُفِّذت على القيمة x_i بوساطة العلاقة [3-9] تُدعى عمليّة استعيار للقيمة x_i ولذلك سُميَّت Z_{x_i} بالدرجة المعيّاريّة لـ x_i. في الواقع إنَّ عملية الاستعيار تعني أنَّه لو حُسِبت الدرجات المعيّارية لجميع عناصر العيّنة فإنَّه سيُلاحظ أنَّ البيانات الناتجة عن هذه الدرجات المعيّارية لها متوسط يساوي الصفر وانحراف معيَّاري يساوي الواحد.

3-3-2- أمثلة

1- في إطار التقدّم لوظيفة مدرّس لدى وزارة التعليم تقدَّم خريج جامعي A من كلية X في جامعةٍ ما F حيث كان متوسط المعدَّل التراكمي لدفعته في الكلية يساوي 4.6 بانحراف معيَّاري يساوي 0.3، وكذلك تقدَّم خريج جامعي آخر B من كلية Y في جامعةٍ ما G حيث كان متوسط المعدَّل التراكمي لدفعته في الكلية يساوي 4.2 بانحراف معيَّاري يساوي 0.4.

فإذا علمت أنَّ المعدَّل التراكمي للشخص A يساوي 4.7 في حين كان المعدَّل التراكمي للشخص B يساوي 4.4، فهل يمكننا الادّعاء أنَّ أداء الشخص A أفضل من أداء الشخص B ؟

الإجابة: هنا نلاحظ أنَّ الردّ على هذا التساؤل يتطلب معرفة مستوى كل متقدِّم في كليّته لأنَّهما لا يخضعان لتعليم موحَّد سواءً من حيث البيئة المحيطة بالمتقدم أو من حيث طبيعة العلم الذي تلقاه، ولذلك الإجابة على السؤال المطروح ليست بهذه البساطة، ويجب أن يُؤخذ في الحسبان المعدَّل التراكمي لزملاء كل واحد منهم، ومن ثمَّ يُبْحث في الإجابة بناءً على هذه المُعطيات. من هذا الحوار نلاحظ أنَّ الإجابة على السؤال المطروح تكمن في حساب الدرجة المعيّارية لكل من هذين المتقدمين حيث نجد الآتي:

$$Z_{\mathbf{A}} = \frac{4.7 - 4.6}{0.3} = 0.\overline{3} \qquad \& \qquad Z_{\mathbf{B}} = \frac{4.4 - 4.2}{0.4} = 0.5$$

وهكذا نجد أنَّ الدرجة المعيّارية للشخص B أعلى من الدرجة المعيّارية للشخص A بفارقٍ كبيرٍ، وهذا يعني أنَّ أداء الشخص B أفضل من أداء الشخص A رغم أنَّ المعدَّل التراكمي للشخص B أقل من المعدَّل التراكمي للشخص A.

2- لتكن لدينا مجموعتي البيانات الآتيتين:

A) 8, 2, 9, 3, 3

B) 8, 6, 7, 8, 11, 9, 6, 9

ولنقم بحساب الدرجة المعيّارية للقيمة $x = 8$ في كلٍّ من مجموعتي البيانات المُعطاة.

الإجابة: من أجل مجموعة البيانات (A) و (B) نجد أنَّ متوسطها وانحرافها المعيّاري على الترتيب هي:

$$\overline{X}_{\mathbf{A}} = 5 \text{ and } S_{\mathbf{A}} = 2.8981$$

$$\overline{X}_{\mathbf{B}} = 8 \text{ and } S_{\mathbf{B}} = 1.5811$$

ومن ثمَّ تكون الدرجة المعيّارية للقيمة $X = 8$ في مجموعة البيانات (A) تساوي:

$$Z_{8,\mathbf{A}} = \frac{x - \overline{X}}{S} = \frac{8 - 5}{2.8981} = 1.0352$$

والدرجة المعيّارية للقيمة $X = 8$ في مجموعة هذه البيانات (B) تساوي:

$$Z_{8,B} = \frac{x - \overline{X}}{S} = \frac{8 - 8}{1.5811} = 0$$

وهكذا نجد أنَّ الدرجة المعيّارية للقيمة $X = 8$ في مجموعة البيانات (A) أكبر من الدرجة المعيّارية للقيمة نفسها في مجموعة البيانات (B).

لاحظ هنا أنَّه لو قمنا بحساب الدرجات المعيارية لبيانات كلٍّ من المجموعتين السابقتين فإنَّنا سنجدها كما في الجدول الآتي:

i	المشاهدة في A	درجتها المعيارية	المشاهدة في B	درجتها المعيارية
1	8	1.035161	8	0
2	2	-1.03516	6	-1.26491
3	9	1.380215	7	-0.63246
4	3	-0.69011	8	0
5	3	-0.69011	11	1.897365
6			9	0.632455
7			6	-1.26491
8			9	0.632455
Mean	**5**	**0**	**8**	**0**
Standard Deviation	**2.89828**	**1.000061**	**1.58114**	**0.999999**

فإنّنا نلاحظ بوضوح أنَّ متوسط الدرجات المعيارية يساوي 0 في حين أنَّ قيمة الانحراف المعياري للدرجات المعيارية يساوي تقريباً 1 (الفرق عن الواحد ناتج عن عمليات التقريب أثناء الحساب).

▶

3-3-3- ملاحظات

1- إذا كنّا بصدد تعيين الدرجة المعيارية لـ x_i كأحد أفراد مجتمع إحصائي حجمه N، فإنّنا نعوّض في العلاقة [3-9] عن $\overline{X}$ بمتوسط المجتمع m، وعن S بالانحراف المعياري للمجتمع s.

2- نلاحظ أنَّ قيمة هذا المعيَّار هي مؤشِّر على انحراف القيمة x_i عن المتوسط، ومن ثمَّ فإنَّها تُحدِّد موقع x_i من المتوسط اتجاهاً وبعداً، فالاتجاه تُحدِّده إشارة (– أو+)، فإذا كانت قيمة Z_{x_i} موجبة فإنَّ ذلك يعني أنَّ x_i أكبر من المتوسط، والعكس إذا كانت قيمة Z_{x_i} سالبة. أمَّا البُعد فتعني كِبَر القيمة المطلقة لـ Z_{x_i}، فكلما كَبُرت القيمة المطلقة لـ Z_{x_i} دلَّ ذلك على ازدياد ابتعاد القيمة x_i عن المتوسط.

3- إنَّ الدرجة المعيارية لـ x_i تمثّل الموضع النسبي لـ x_i داخل توزيع البيانات، ومن ثمَّ فمن أجل عينتين متمايزتين (أو مجتمعين متمايزين) سيكون للدرجة المعيارية دور فعّال للمقارنة بين عناصر العينتين (أو المجتمعين).

4- يُؤخذ على الدرجة المعيّارية أنَّها عند قيمة الصفر يكون للبيان الذي استعيّرت قيمته درجة قيمة المتوسط $\overline{X}$، ولذلك قد لا يكون معناها واضحاً لدى الكثير، وعلاوةً على ذلك فقد تكون هذه الدرجة المعيّارية سالبةً أيضاً، وهذا بدوره يجعل تفسيرها من أجل بعض الحالات غير واضح أو غير ذي معنى، ولذلك اقترح تقديم مقياس آخر يتجاوز هذه السلبيات يدعى الدرجة المعياريّة -t (أو الدرجة المعيارية التائيّة)، ولكننا لن نتطرق إلى هذا المقياس في كتابنا هذا.

تمارين الفصل الثالث

1- عندما نتسوّق في أحد المتاجر للمواد التّموينيّة لاحظنا أنَّه كُتِبَ على بعض السلع عبارة "الوزن الصافي 2 كغ ± 100 غ"، فما هو تفسير هذه الكتابة؟

2- إذا كانت x_1 ، x_2 ، ... و x_n مع $1 \leq n$ مجموعة بيانات مُعطاة، فعندئذٍ:

أ- هل يمكن لقيمةٍ عدديةٍ x أن تكون مقياساً للنَّزعة المركزيَّة لهذه المجموعة من البيانات علماً أنَّها توافق x_1 (**أكبر قيمة في البيانات**) أو x_s (**أصغر قيمة في البيانات**) ؟

ب- بفرض أنَّ $\overline{X}$ هي قيمة المتوسط لهذه المجموعة من البيانات، فمن أجل أي قيمة لـ n تصبح $\overline{X} = x_1$ أو $\overline{X} = x_s$ ؟

ج- من أجل أيّة قيمة لـ n تصبح قيمة الانحراف المعيّاري لهذه المجموعة من البيانات تساوي الصفر؟

3- لتكن لدينا مجموعة بيانات مرتَّبة لعينة مُعطاة كما يلي:

? 19 14 12 12 9 ? 6 3.75 3.55 3.50

علماً أنَّ (?) ترمز إلى قيمة مفقودة ضمن الترتيب المقدَّم، فعندئذٍ:

أ- استخدم المقياس الملائم لتعيين القيمة التي تنزع إليها هذه البيانات.

ب- استخدم المقياس الملائم لتعيين القيمة التي تعبِّر عن تبعثر هذه البيانات.

ج- بفرض أنَّ القيمة الأخيرة من البيانات معلومة وتساوي 39، فهل تفضِّل استخدام المدى كمقياس لتشتتها عوضاً عن المدى الربيعي؟ ولماذا؟

4- لتكن لدينا مجموعة بيانات مرتَّبة لعينة مُعطاة كما يلي:

21 21 ? 15 ? 12 ? 6 6 6 ? 3

علماً أنَّ (?) ترمز إلى قيمة مفقودة ضمن الترتيب المقدَّم، فعندئذٍ:

أ- استخدم المقياس الملائم لتعيين القيمة التي تنزع إليها هذه البيانات.

ب- استخدم المقياس الملائم لتعيين القيمة التي تعبِّر عن تبعثر هذه البيانات.

5- لتكن لدينا مجموعة البيانات الآتية التي تمثِّل الوزن لعينة مكوَّنة من 48 شخصاً بالغاً (**مقدَّرةً بالكيلو غرام**):

76	66	75	92	74	73	62	80	76	66	79	60
75	84	77	84	77	71	81	75	84	65	61	82
77	71	81	75	92	74	73	62	80	76	66	79
84	77	71	81	75	84	81	75	84	65	61	82

والمطلوب ما يلي:

أ- حساب المدى، المدى الرُبيْعي ومعامل التشتت للبيانات المُعطاة.

ب- قم بسحب عينات عشوائية من البيانات بحجم $n = 12$، $n = 24$ و $n = 36$، ومن ثمَّ احسب المتوسط والانحراف المعيّاري لكلّ منها، وبعد ذلك قارن بين النتائج المتقابلة. ماذا تلاحظ؟

6- لتكن لدينا البيانات 8, 9, 6, 9, 5, 4, 8, 6, 7, 8 لعينة حجمها 10، والمطلوب حساب المتوسط والانحراف المعيّاري لهذه البيانات.

7- ليكن لدينا مجموعة بيانات عينة مقدَّمة من خلال جدول توزيع تكراري له العرض الآتي:

التكرار المتجمِّع الصاعد للفئة	تكرار الفئة	مركز الفئة	حدود الفئة	رقم الفئة
	5		$50 \longrightarrow 55$	1
13			$55 \longrightarrow 60$	2
	12			3
40				4
	17			5
70				6
المجموع				**Total**

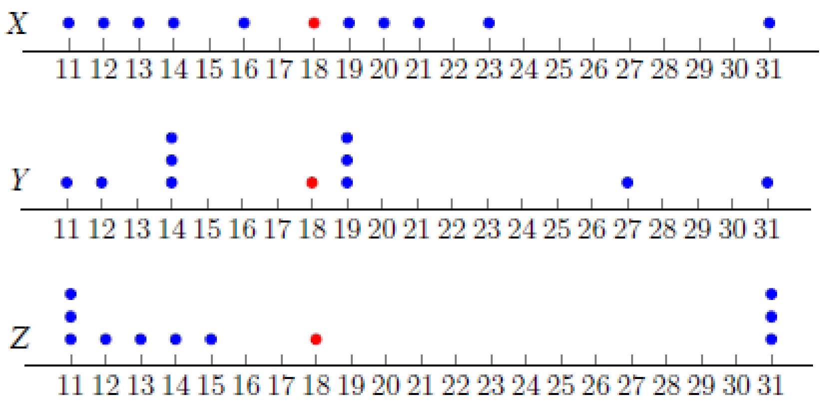
والمطلوب ما يلي:

أ- إكمال جدول التوزيع التكراري السابق.

ب- حساب المتوسط.

ج- حساب الانحراف المعيّاري.

8- لتكن لدينا ثلاث مجموعات بيانات (**ممثَّلةً نقطياً باللون الأزرق فقط**) لعينات مقدَّمة كما في العروض الآتية:



والمطلوب ما يلي:

أ- حساب المتوسط، المدى، الانحراف المعيّاري والمدى الربيعي لكل من مجموعات البيانات المُعطاة.

ب- لو أضيفت القيمة 18 إلى كل مجموعة من مجموعات البيانات السابقة، فهل تتغيَّر قيم المتوسطات لمجموعات البيانات المُعطاة، ولماذا؟

ج- حساب معامل التغيُّر لكل من مجموعات البيانات المُعطاة، ومن ثمَّ بيّن أيها أقل تبعثراً من الأخرى.

د- حساب الدرجة المعيّارية للقيمة $x = 14$ في كل من مجموعات البيانات المُعطاة، ماذا تلاحظ؟

9- ليكن لدينا المضلَّع التكراري الآتي لمجموعة بيانات عينة:

والمطلوب ما يلي:

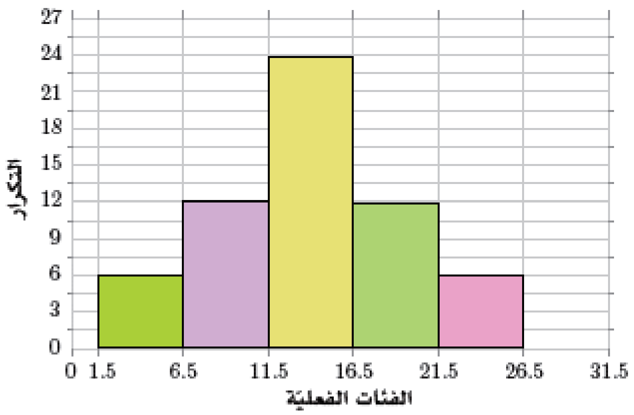
أ- بناء جدول التوزيع التكراري لبيانات هذا المضلَّع التكراري.

ب- حساب المتوسط لبيانات هذا المضلَّع التكراري.

ج- حساب المدى للبيانات الممثَّلة بهذا المضلَّع التكراري.

د- حساب الانحراف المعيّاري للبيانات الممثَّلة بهذا المضلَّع التكراري.

10- ليكن لدينا المدرَّج التكراري الآتي لمجموعة بيانات عيّنة:



والمطلوب ما يلي:

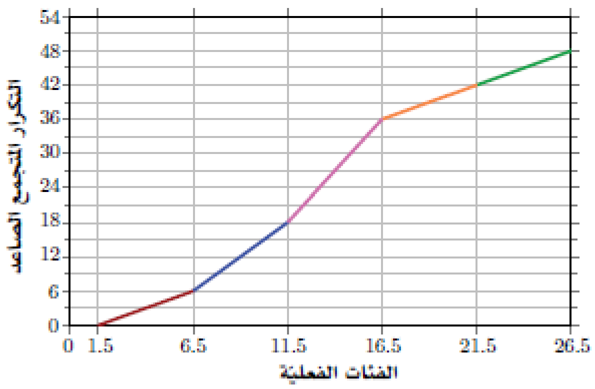
أ- بناء جدول التوزيع التكراري للبيانات الممثَّلة بهذا المدرّج التكراري.

ب- حساب المتوسط، الوسيط والمنوال للبيانات الممثَّلة بهذا المدرّج التكراري.

ج- حساب المدى للبيانات الممثَّلة بهذا المدرّج التكراري.

د- حساب الانحراف المعيّاري للبيانات الممثَّلة بهذا المدرّج التكراري.

11- لتكن لدينا بيانات مقدّمة من خلال العرض البياني الآتي:



والمطلوب ما يلي:

أ- بناء جدول التوزيع التكراري للبيانات الخاصة بمضلَّع التكرار المتجمّع الصّاعد المعطى.

ب- حساب المتوسط، للبيانات الممثَّلة بمضلَّع التكرار المتجمّع الصّاعد المعطى.

ج- حساب المدى للبيانات الممثّلة بمضلَّع التكرار المتجمّع الصّاعد المعطى.

د- حساب الانحراف المعيّاري للبيانات الممثّلة بمضلَّع التكرار المتجمّع الصّاعد المعطى.

12- ينتج مصنع أربعة أنواع من العقاقير الدوائية A، B، C وD. إنَّ هذه العقاقير تخضع لعدد من الاختبار المعمليّة كلّ حسب نوعه حتى يمكن التثبّت من صلاحيته للتسويق، فإذا علمت أنَّ البينات المتعلِّقة بهذه الاختبارات كانت كما في الجدول التكراري الآتي:

العقار الدوائي	تكلفة الاختبار الواحد (بالريال السعودي)	عدد الاختبارات التي أجريت على العقار
A	555	6
B	495	6
C	850	3
D	750	5
Total		

والمطلوب ما يلي:

أ- تقديم التمثيل الشرائطي والدائري لبيانات هذا الجدول على أساس الكلفة الإجماليّة للاختبار.

ب- حساب الوسيط وتعيين المنوال (أو المناويل) لبيانات هذا الجدول.

ج- حساب متوسط كلفة الاختبارات والانحراف المعياري المرافق لهذه الكلفة.