الفصل الأول: مفاهيم أساسية في الإحصاء

الفصل الأول: مفاهيم أساسية في الإحصاء
Chapter 1: Basic concepts in statistics

الكلمات المفتاحية:

الإحصاء الوصفي، الإحصاء الاستدلالي، المجتمع، العينة، معطيات كمية، معطيات وصفية، معطيات متقطعة، معطيات مستمرة، وسيط احصائي، الإحصائية، أخذ العينات، العينات الاحتمالية، جدول تكراري، جدول تكراري تراكمي، مدرج تكراري، مضلع تكراري، مخطط صندوقي، رسو بيانية، أعمدة بيانية، مخطط الساق والأوراق، نزعة مركزية، المتوسط، الوسيط، المنوال، التشتت، الانحراف المعياري، المدى، المدى الربيعي، معامل الاختلاف، متراجحة تشيبتشيف، القيم المعيارية.

ملخص:

نهتم في هذا الفصل بتوضيح المفاهيم الأساسية لعلم الإحصاء وأنواع المعطيات وطرق الإحصاء الممكن تطبيقها على كل منها. كما نهتم أيضاً بمصادر الحصول على المعطيات وطرق أخذ العينات والطرق التي تستخدم من أجل تنظيم وإظهار المعطيات. وأخيراً نهتم بدراسة مركز وتشتت المعطيات وهي من الخصائص الهامة جداً للمعطيات التي تساعد على توصيفها واستثمارها.

الأهداف تعليمية:

يتعرف الطالب في هذا الفصل على:

- المفاهيم الأساسية في علم الإحصاء
- أنواع المعطيات ومصادر الحصول عليها
- تنظيم المعطيات وعرضها
- مقاييس النزعة المركزية وطرق حسابها (المتوسط، الوسيط، ...)
- مقاييس التشتت وطرق حسابها (المدى، التشتت، ...)

المخطط:

1. مدخل إلى الإحصاء.
2. تنظيم المعطيات وعرضها.
3. مقاييس النزعة المركزية.
4. مقاييس التباين.

1. مدخل إلى الإحصاء introduction to statistic:

يُعرف علم الإحصاء على أنه ذلك الفرع من العلوم المكون من مجموعة من الطرق الرياضية المستخدمة في جمع وتنظيم وعرض وتلخيص المعلومات والمعطيات ومن ثم تحليلها وفق طرق وأساليب علمية للحصول على استدلالات وقرارات سليمة على ضوء هذا التحليل.

ينقسم علم الإحصاء إلى قسمين أساسيين هما:

الإحصاء الوصفي descriptive or deductive statistics: يهتم بتنظيم المعطيات وتلخيصها وعرضها، والغرض من التنظيم والتلخيص والعرض هو المساعدة على فهم المعطيات. وهو لا يتضمن أي تعميم أو استنتاج من المعطيات.

الإحصاء الاستدلالي inductive statistics or statistical inference: عبارة عن مجموعة من الطرق العلمية لسبر معالم مجتمع بناء على معلومات يتم الحصول عليها من عينة إحصائية مأخوذة وفق طرق إحصائية علمية مناسبة.

1.1. مفاهيم أساسية في علم الإحصاء general concepts in statistics:

1.1.1. المجتمع population:

المجتمع هو المجموعة الكلية من الأشياء (تسمى عناصر المجتمع) المراد دراستها والتي لها خصائص مشتركة حيث سيتم عمل بعض الاستدلالات والنتائج حولها. ونسمي عدد أفراد المجتمع بحجم المجتمع ويرمز له عادة بالرمز N. والمجتمع إما أن يكون محدوداً (منتهياً) مثل مجتمع طلاب كلية الجامعة الافتراضية السورية، أو غير محدود (غير منته) مثل المجتمع المكون من النجوم في السماء.

2.1.1. العينة sample:

العينة هي مجموعة مكونة من عدد محدد من أفراد المجتمع يتم اختيارهم بطريقة مناسبة بحيث تمثل المجتمع تمثيلاً جيداً وذلك لدراسة صفات المجتمع إذ يُستدل بصفات العينة على صفات المجتمع. ونسمي عدد أفراد العينة بحجم العينة sample size ويرمز لها عادة بالرمز n.

قد تكون الحاجة ضرورية لأخذ عينة بدلاً من دراسة المجتمع كله، على سبيل المثال أخذ عينة من دم الإنسان لفحصها حيث أننا لا تستطيع فحص كل دمه لأن ذلك يؤدي إلى موته. كذلك قد تؤدي دراسة المجتمع كله إلى فقدان عناصره أو إتلافها ومن هنا يجب أخذ عينه صغيرة منها، على سبيل المثال عند فحص كمية من البيض نأخذ عينة منها ونقوم بكسرها لنرى فيما إذا كان البيض سليماً أم لا.

3.1.1. المعطيات data:

هي مجموعة المشاهدات أو الملاحظات المأخوذة من الدراسة الإحصائية. ويمكن تقسيم المعطيات إلى نوعين:

- **معطيات كمية** وتتألف من أرقام تمثل عدد أو قياس ما، مثل معطيات أطوال مجموعة من الأشخاص وتقاس بالسنتيمتر، رواتب الموظفين وتقاس بالليرة السورية. ويمكن تقسيم المعطيات الكمية إلى نوعين:
  - **معطيات متقطعة** وهي معطيات تنتج عندما يكون عدد القيم الممكنة للمعطيات هو إما عدد محدود أو قابل للعد، مثل علامات طالب في الجامعة الافتراضية.
  - **معطيات مستمرة** وهي معطيات تنتج عندما يكون عدد القيم الممكنة للمعطيات عدد غير محدود وغير قابل للعد (مجال مستمر)، مثل درجات الحرارة على مدار السنة لمدينة دمشق.

- **معطيات وصفية** حيث يمكن تقسيمها إلى فئات تتميز فيما بينها ببعض الخصائص غير الرقمية، مثل معطيات المستوى التعليمي (أمّي أو يقرأ ويكتب: ابتدائية، متوسطة، ثانوية، جامعية، أعلى من جامعية)، معطيات الحالة الاجتماعية (متزوج أو أعزب أو أرمل أو مطلق)، معطيات الجنس (ذكر أو أنثى) لمجموعة من الأشخاص.

4.1.1. المتحول variable:

المتحول هو مقدار كمي أو وصفي ويستخدم لقياس خاصية أو مميزة معينة لعناصر المجتمع أو العينة. يتم قياس المتحول على أفراد المجتمع أو العينة وتتغير قيمته من فرد إلى آخر. من أمثلة المتحولات: طول الشخص، عدد أولاد الشخص (متحول كمي)، فصيلة الدم للشخص، المستوى التعليمي للشخص (متحول وصفي).

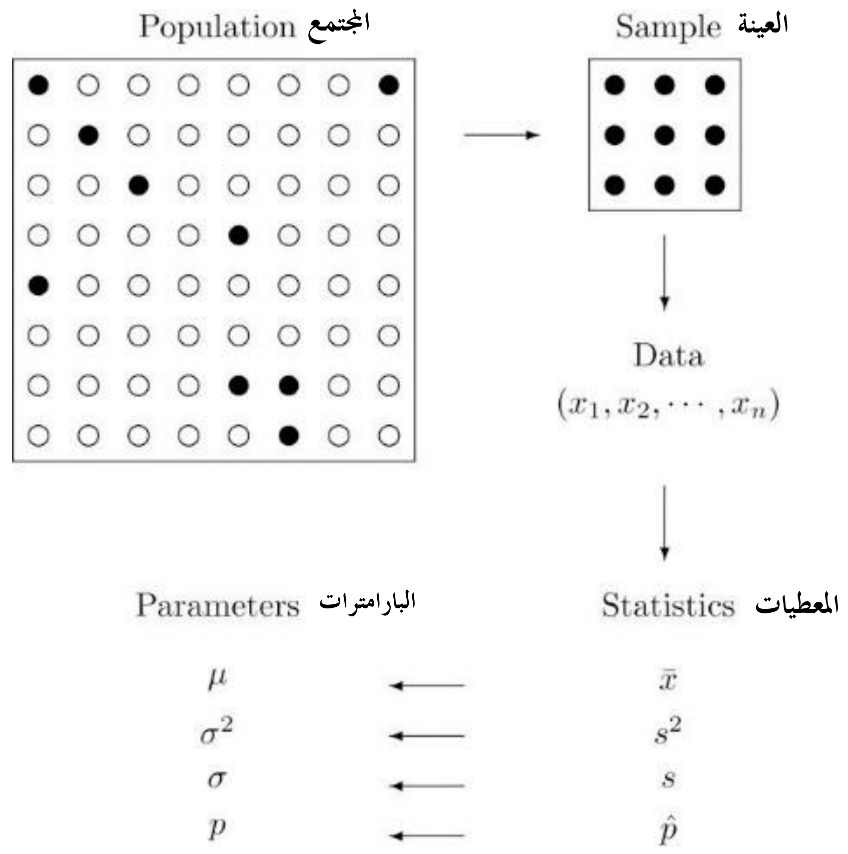
5.1.1. المعلمة (البارامتر أو الوسيط) parameter:

المعلمة هي سمة أو خاصية رقمية تصف المجتمع، وبما أننا لا نستطيع أخذ كل عينات المجتمع فإن المعلمة غير معروفة. المعلمات هي قيم محددة fixed (حتى لو كنا لا نعرفهم). مثال متوسط العمر في دولة معينة أو نسبة الذين يدخنون بصفة دائمة في مجتمع معين.

6.1.1. الإحصائية statistic:

الإحصائية هي سمة أو خاصية رقمية تصف عينة من المجتمع. بالطبع عينات مختلفة تُنتج إحصائيات مختلفة المعلمات. مثال متوسط العمر لعينة مكونة من 500 شخص في دولة معينة أو نسبة الذين يدخنون بصفة دائمة لعينة مكونة من 200 شخص في مجتمع معين.

إن العلاقة بين المجتمع قيد الدراسة والعينة المأخوذة من هذا المجتمع ربما هي أهم مفهوم في الإحصاء. هذه العلاقة يبينها الشكل التالي:



2.1. تجميع المعطيات collecting data:

تعتبر طرق تجميع (تحصيل) المعطيات من أهم المراحل التي يعتمد عليها البحث الإحصائي، كما أن تجميع المعطيات بأسلوب علمي صحيح، يترتب عليه الوصول إلى نتائج دقيقة في التحليل. هناك مصدرين للحصول منها على المعطيات هما:

- **المصادر الأولية (الميدانية):** وهي المصادر التي نحصل منها على المعطيات بشكل مباشر، حيث يقوم الباحث نفسه بجمع المعطيات من ميدان المعرفة أو الدراسة التي يهتم بها، فعندما يهتم الباحث بجمع معطيات عن الأسرة، يقوم بإجراء مقابلة مع رب الأسرة ويتم الحصول منه مباشرة على معطيات خاصة بأسرته، مثل معطيات المنطقة التابع لها، والحي الذي يسكن فيه، والجنسية، والمهنة، والدخل الشهري، وعدد أفراد الأسرة، والمستوى التعليمي، ... وهكذا.

  يتميز هذا النوع من المصادر بالدقة والثقة في المعطيات، لأن الباحث هو الذي يقوم بنفسه بجمع المعطيات، ولكن من مساوئها أنها تحتاج إلى وقت ومجهود كبير، كما أنها مكلفة من الناحية المادية.
- **المصادر الثانوية (التاريخية):** وهي المصادر التي نحصل منها على المعطيات بشكل غير مباشر، بمعنى آخر يتم الحصول عليها بواسطة أشخاص آخرين أو أجهزة أو هيئات رسمية متخصصة، مثل سجلات المواليد والوفيات والسجلات الصادرة عن هيئة الأمم والبنك الدولي وغيرها. على سبيل

المثال نستطيع معرفة الحاصلين على شهادة الثانوية العامة في سورية من خلال الاطلاع على السجلات الخاصة بإدارة الامتحانات.

من مزايا هذا النوع من المصادر، توفير الوقت والجهد والمال، إلا أن درجة ثقة الباحث فيها ليست بنفس الدرجة في حالة المصادر الأولية.

3.1. إجرائيات أخذ العينات sampling procedures:

العينة السليمة هي العينة الممثلة للمجتمع الذي اختيرت منه، وعملية أخذ العينات بطريقة غير مناسبة واحدة من أسوأ الأخطاء الممكن ارتكابها. ويمكن تقسيم العينات وفقا لأسلوب اختيارها إلى نوعين هما العينات الاحتمالية والعينات غير الاحتمالية.

1.3.1. العينات الاحتمالية probability sampling:

هي العينات التي يتم اختيار عناصرها وفقا لقواعد الاحتمالات، بمعنى آخر هي التي يتم اختيار عناصرها من مجتمع الدراسة بطريقة عشوائية، بهدف تجنب التحيّز الناتج عن اختيار العناصر، ومن أهم أنواع العينات الاحتمالية ما يلي:

1. **العينة العشوائية البسيطة simple random sample:** تؤدى هذه الطريقة إلى احتمال اختيار أي عنصر من أفراد المجتمع كعنصر من عناصر العينة، كما أن لكل عنصر فرصة متساوية لاختياره ضمن العينة.
2. **العينة العشوائية المنتظمة systematic random sample:** يتم فيها اختيار الحالة الأولى من العينة بطريقة عشوائية ثم يتم اختيار بقية الحالات على أبعاد رقمية منتظمة أو متساوية بين الحالات.
3. **العينة العشوائية الطبقية stratified random sample:** يتم في هذه الطريقة تقسيم المجتمع إلى عدة مجموعات نسمي كل منها طبقة تشترك بنفس الخواص (الجنس، العمر، ...).
4. **العينة العشوائية العنقودية cluster random sample:** يتم في هذه الطريقة تقسيم المجتمع إلى مقاطع، بعدها يتم اختيار بعض هذه المقاطع عشوائياً ومن ثم نختار جميع العناصر من تلك المقاطع.

2.3.1. العينات غير الاحتمالية non probability sampling:

هي العينات التي يتم اختيار عناصرها بطريقة غير عشوائية، حيث يقوم الباحث باختيار عناصر العينة بالصورة التي تحقق الهدف من المعاينة، مثل اختيار عينة من المزارع التي تنتج التمور من النوع السكري، وأهم أنواع العينات غير الاحتمالية:

1. **العينة العمدية purposive sample:** يتم اختيار الحالات بناء على هدف خاص لدى الباحث. مثال تحليل محتوى مجلة محددة، دراسة متعمقة لبعض حالات التخلف العقلي.

2. **العينة الحصصية quota sample:** تتطلب معرفة مسبقة لمجتمع الدراسة من حيث تكوين المجموعات داخله، وعملية الاختيار في كل مجموعة لا ترتبط بقواعد معينة ولكن لقناعة الباحث بشرط أن تمثل كل مجموعة في العينة حسب تمثيلها في مجتمع الدراسة. مثال تحديد الباحث فئات المجتمع (ذكور وإناث) ثم يختار عدد ثابت من كل فئة إذ يقرر اختيار عشرة ذكور وخمس إناث.

2. تنظيم المعطيات وعرضها organization and display of data:

بعد تجميع المعطيات سواء من المصادر التاريخية أو المصادر الميدانية فإنها تكون معطيات أولية غير منظمة تصعب دراستها أو استنتاج أي معلومة مفيدة منها. ولذلك تدعو الحاجة إلى تنظيم المعطيات وتلخيصها بصورة يسهل فهمها، ولتوضيح ذلك نعتبر المثال التالي:

مثال 1: ليكن لدينا علامات طلاب الجامعة الافتراضية في مادة الإحصاء والاحتمالات التالية:

23, 50, 38, 42, 63, 75, 10, 33, 26, 39, 35, 47,
43, 52, 56, 59, 64, 77, 15, 21, 51, 54, 72, 68,
36, 65, 52, 60, 27, 34, 47, 48, 55, 58, 59, 62,
51, 48, 50, 41, 57, 65, 54, 43, 56, 44, 30, 46,
67, 53

المعطيات السابقة لا يمكن دراستها والاستفادة منها. فعلى سبيل المثال ما هو عدد الطلاب الناجحين (علامتهم فوق ال 60)؟ أو ما هو عدد الطلاب الذين تتراوح علامتهم بين 65 و75 درجة؟ بالتالي فإن أول مرحلة من مرحلة من التحليل الاحصائي هي تنظيم هذه المعطيات في جداول تكرارية.

1.2.الجداول التكرارية frequency distributions:

الجدول التكراري عبارة عن جدول يلخص المعطيات فهو يبين تكرار كل مفردة من مفردات العينة في حال كان عدد المفردات ذات القيم المختلفة غير كبير.

مثال 2: أوجد الجدول التكراري لعينة من القياسات تمثل طول الساق ل 10 أنواع من النباتات بالسنتيمتر:
. 62, 59, 68, 62, 58, 59, 65, 62, 65,65

الحل:

طول الساق	التكرار
58	1
59	2
62	3
65	3
68	1
المجموع	**10**

أما في الحالة التي يكون عدد المفردات المختلفة كبير يتم توزيع قيم المفردات على فئات classes ويحدد عدد المعطيات التي تنتمي إلى كل فئة.

- نسمي عدد المعطيات التي تنتمي إلى كل فئة بتكرار frequency تلك الفئة، ويرمز للتكرار بالرمز f
- نسمي الحد الأدنى للفئة أقل قيمة يمكن أن تنتمي إلى تلك الفئة.
- نسمي الحد الأعلى للفئة أكبر قيمة يمكن أن تنتمي إلى تلك الفئة.

طريق تكوين الفئات المنتظمة للمعطيات الكمية:

الهدف من تكوين الفئات هو تجميع القيم المتقاربة في مجموعات وحيث لا توجد قواعد ثابتة لتحديد طول الفئات ولا عددها، وعادة يتراوح عدد الفئات من 5 إلى 20 فئة. ولتحديد عدد الفئات وطول كل فئة يعتمد على الخبرة، هذا ويمكن استخدام العلاقة التالية في إيجاد العدد الأعظمي لعدد الصفوف (n عدد المعطيات):

$$C = 3.3\log(n) \approx \sqrt{n}$$

لتوضيح كيفية تكوين الفئات المنتظمة نأخذ معطيات المثال الأول (علامات طلاب الجامعة الافتراضية في مادة الإحصاء والاحتمالات):

- نختار عدد الفئات $C = 3.3\log(50) \approx 6.66$ بالتالي نأخذ $C = 7$.
- نحسب مدى المعطيات (أكبر قيمة – أصغر قيمة) Range=77−10=67.
- نحسب طول الفئة بتقسيم المدى على عدد الفئات، أي L = 67/7 = 9.6 نقربها فتصبح 10.
- نختار أصغر قيمة لتكون بداية الفئة الأولى (10 في مثالنا)، نضيف إليها طول الفئة فنحصل على بداية الفئة الثانية (في مثالنا 20 = 10+10) وفي كل مرة نضيف إلى بداية فئة ما بد طول الفئة نحصل على بداية الفئة التي تليها.
- لإيجاد نهاية كل فئة نضيف إلى بدايتها طول الفئة مطروحاً منه واحد (في مثالنا نهاية الفئة الأولى هي 19).

مثال 3: ليكن لدينا علامات طلاب الجامعة الافتراضية في مادة الإحصاء والاحتمالات السابقة. الجدول التالي يبين الجدول التكراري لعلامات الطلاب ($n = 50$):

التكرار	العلامة (M)
2	10 ≤ M < 20
4	20 ≤ M < 30
7	30 ≤ M < 40
10	40 ≤ M < 50
16	50 ≤ M < 60
8	60 ≤ M < 70
3	70 ≤ M < 80
50	**المجموع**

ملاحظة 1: مجموع تكرار الفئات لعينة من المعطيات يساوي حجم العينة.

ويمكن بالاعتماد على جدول التوزيع التكراري الحصول على الجداول التالية:

1. **جدول التوزيع التكراري النسبي relative frequency distribution**، حيث يتم تعريف التكرار النسبي للفئة بالعلاقة التالية:

Relative frequency = frequency/sample size = f/n

مثال 4: ليكن لدينا علامات طلاب الجامعة الافتراضية في مادة الإحصاء والاحتمالات السابقة. الجدول التالي يبين الجدول التكراري النسبي لعلامات الطلاب:

العلامة (M)	التكرار النسبي
$10 \le M < 20$	0.04
$20 \le M < 30$	0.08
$30 \le M < 40$	0.14
$40 \le M < 50$	0.2
$50 \le M < 60$	0.32
$60 \le M < 70$	0.16
$70 \le M < 80$	0.06
المجموع	1

2. **جدول التوزيع التكراري المئوي percentage frequency distribution**، حيث يتم الحصول على التكرار المئوي للفئة بضرب التكرار النسبي بالعدد 100، أي:

$$\text{Percentage frequency} = \frac{f}{n} * 100$$

مثال 5: ليكن لدينا علامات طلاب الجامعة الافتراضية في مادة الإحصاء والاحتمالات السابقة. الجدول التالي يبين الجدول التكراري المئوي لعلامات الطلاب:

العلامة (M)	التكرار المئوي
$10 \le M < 20$	4%
$20 \le M < 30$	8%
$30 \le M < 40$	14%
$40 \le M < 50$	20%

$50 \le M < 60$	32%
$60 \le M < 70$	16%
$70 \le M < 80$	6%
المجموع	**100%**

تعريف 1 (مركز الفئة): يتم تعريف مركز الفئة midpoint بالعلاقة التالية:

مركز الفئة = (الحد الأعلى للفئة + الحد الأدنى للفئة)/2

يمكن تجميع الجداول السابقة للمعطيات في جدول واحد يشمل الفئات ومراكز الفئات والتكرار النسبي والتكرار المئوي كما يبنه الجدول التالي لتوزيع علامات الطلاب:

العلامة (M)	مركز الفئة	التكرار	التكرار النسبي	التكرار المئوي
$10 \le M < 20$	14.5	2	0.04	4%
$20 \le M < 30$	24.5	4	0.08	8%
$30 \le M < 40$	34.5	7	0.14	14%
$40 \le M < 50$	44.5	10	0.2	2%
$50 \le M < 60$	54.5	16	0.32	32%
$60 \le M < 70$	64.5	8	0.16	16%
$70 \le M < 80$	74.5	3	0.06	6%
المجموع		**50**	**1**	**100%**

3. جدول التوزيع التكراري التراكمي: cumulative frequency distribution

في كثير من الأحيان نحتاج للإجابة على التساؤلات التالية: ما هو عدد الطلاب الذين تكون علامتهم أصغر من قيمة معينة؟ أو ما هو عدد الطلاب الذين تكون علامتهم أكبر من قيمة معينة؟ أو ما هو عدد الطلاب الذين تكون علامتهم محصورة بين قيمتين محددتين؟

للإجابة على تلك التساؤلات يتم استخدام ما يسمى جداول التوزيع التراكمية وهي نوعان الصاعدة والهابطة.

جدول التوزيع التكراري التراكمي الصاعد:

نعرف التكرار التراكمي الصاعد للفئة على أنه عدد المعطيات التي تقل عن الحد الأعلى للفئة، وهي تساوي إلى تكرار الفئة مضافاً إليها مجموع تكرارات الفئات السابقة لها.

يتم استخدام جدول التكرار التراكمي الصاعد لإيجاد عدد المعطيات التي تكون قيمها أصغر من قيمة معينة، كما يستخدم أيضاً لإيجاد عدد المعطيات التي قيمها محصورة بين قيمتين محددتين.

مثال 6: ليكن لدينا علامات طلاب الجامعة الافتراضية في مادة الإحصاء والاحتمالات السابقة. أوجد جدول التكرار التراكمي الصاعد لعلامات الطلاب.

الحل:

عدد الطلاب الذين علامتهم أقل من 10 يساوي 0

عدد الطلاب الذين علامتهم أقل من 19 يساوي 2

عدد الطلاب الذين علامتهم أقل من 29 يساوي $6 = 4 + 2$

عدد الطلاب الذين علامتهم أقل من 39 يساوي $13 = 7 + 4 + 2$

...

بوضع النتائج السابقة في جدول نحصل على جدول التكرار التراكمي الصاعد لعلامات الطلاب.

التكرار التراكمي الصاعد	العلامة
0	أقل من 10
2	أقل من 20
6	أقل من 30
13	أقل من 40
23	أقل من 50
39	أقل من 60
47	أقل من 70
50	أقل من 80
50	**المجموع**

يمكن من الجدول السابق استنتاج أن عدد الطلاب الذين علامتهم أقل من 60 هو 39، وأن عدد الطلاب الذين تتراوح علامتهم بين 40 و70 هو $70 = 13 - 47$

جدول التوزيع التكراري التراكمي الهابط:

نعرف التكرار التراكمي الهابط للفئة على أنه عدد المعطيات التي تزيد عن أو تساوي الحد الأدنى للفئة.

مثال 7: ليكن لدينا علامات طلاب الجامعة الافتراضية في مادة الإحصاء والاحتمالات السابقة. أوجد جدول التكرار التراكمي الهابط لعلامات الطلاب.

الحل:

عدد الطلاب الذين علامتهم أكبر أو تساوي 10 يساوي 50

عدد الطلاب الذين علامتهم أكبر أو تساوي 20 يساوي $48 = 2 - 50$

عدد الطلاب الذين علامتهم أكبر أو تساوي 30 يساوي $44 = 4 - 2 - 50$

عدد الطلاب الذين علامتهم أكبر أو تساوي 30 يساوي 37 = 7 4 – 2 – 50

...

بوضع النتائج السابقة في جدول نحصل على جدول التكرار التراكمي الهابط لعلامات الطلاب.

العلامة	التكرار التراكمي الهابط
أكبر أو يساوي 10	50
أكبر أو يساوي 20	48
أكبر أو يساوي 30	44
أكبر أو يساوي 40	37
أكبر أو يساوي 50	27
أكبر أو يساوي 60	11
أكبر أو يساوي 70	3
أكبر أو يساوي 80	0
المجموع	**50**

2.2. العرض البياني graphical representation:

إن تنظيم وتلخيص وعرض المعطيات على شكل جداول تكرارية يعطي صورة شاملة وواضحة عن المعطيات الأولية وتوزيعاتها التكرارية، ولكن عرض هذه الجداول التكرارية بشكل بياني يعطي فكرة أوضح وأسرع عن أشكال التوزيعات التكرارية.

يتم عرض التوزيعات التكرارية بيانياً باستخدام العديد من المعطيات graphs أهمها:

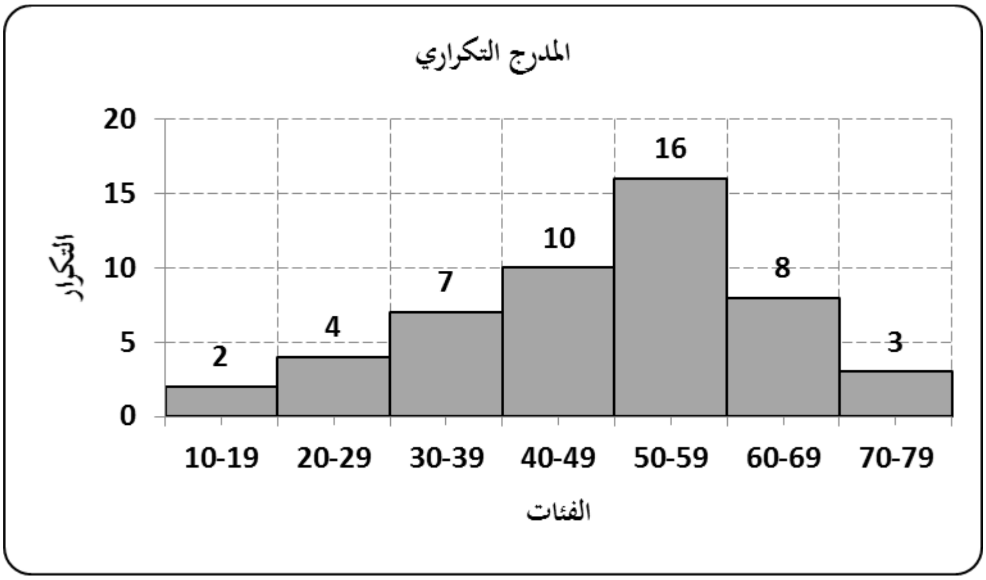
- المدرج التكراري
- المضلع التكراري
- المنحني التكراري
- المنحني التكراري التراكمي الصاعد والهابط.

1.2.2. المدرج التكراري histogram:

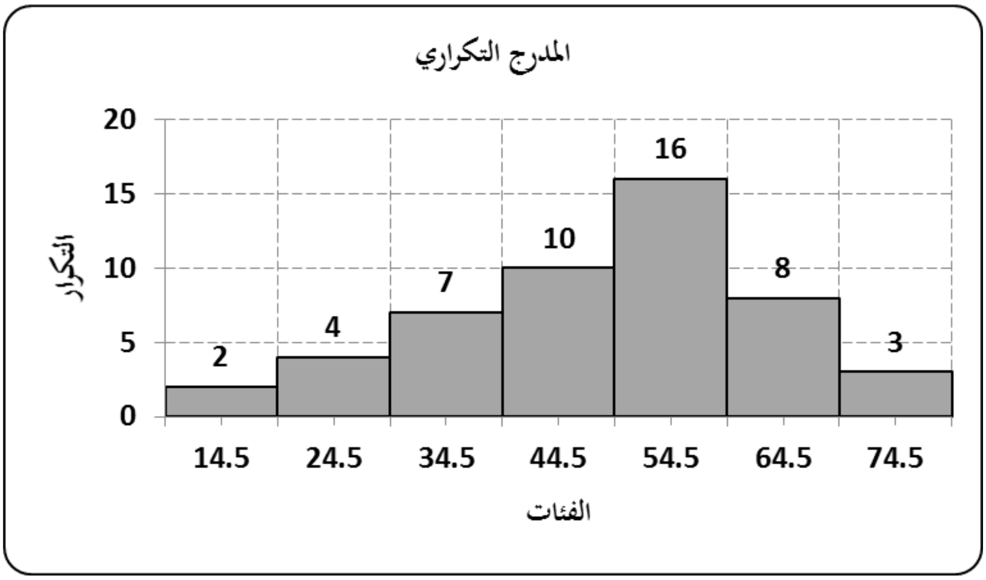
يمثل المحور الأفقي الفئات ويمثل المحور الشاقولي تكرارها. نرسم مستطيلات متلاصقة على الفئات قاعدتها طول الفئة وارتفاعها عبارة عن تكرار هذه الفئات.

مثال 8: ليكن لدينا علامات طلاب الجامعة الافتراضية في مادة الإحصاء والاحتمالات السابقة. أوجد المدرج التكراري لعلامات الطلاب.

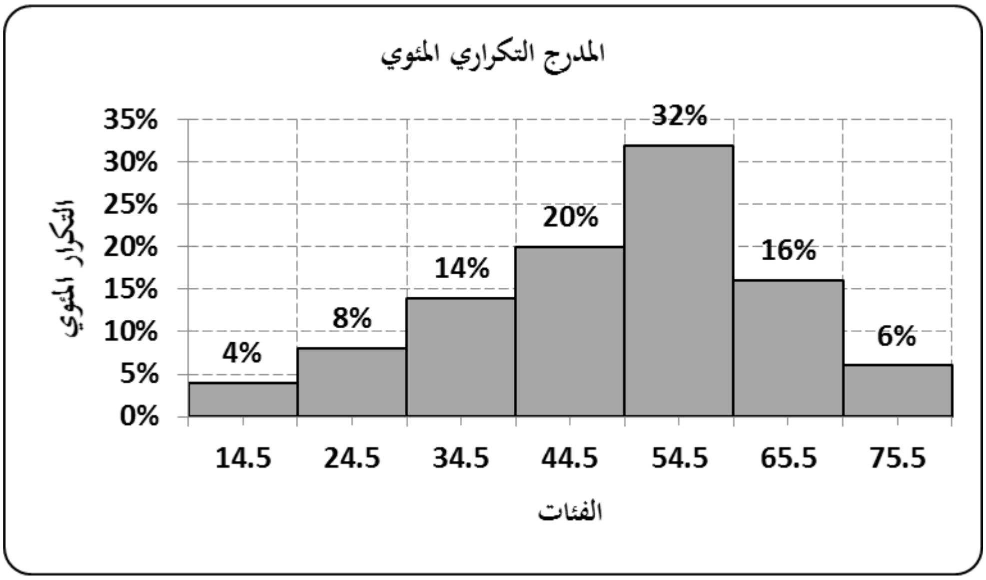
الحل:



من الملاحظ أن كل مستطيل من المدرج التكراري محدد بالحد الأدنى والحد الأعلى للفئة الموافقة. ولكن في كثير من الأحيان يتم استخدام نقطة المنتصف للفئة بدلاً من استخدام الحدين الأدنى والأعلى للفئة كما هو مبين في الشكل التالي:



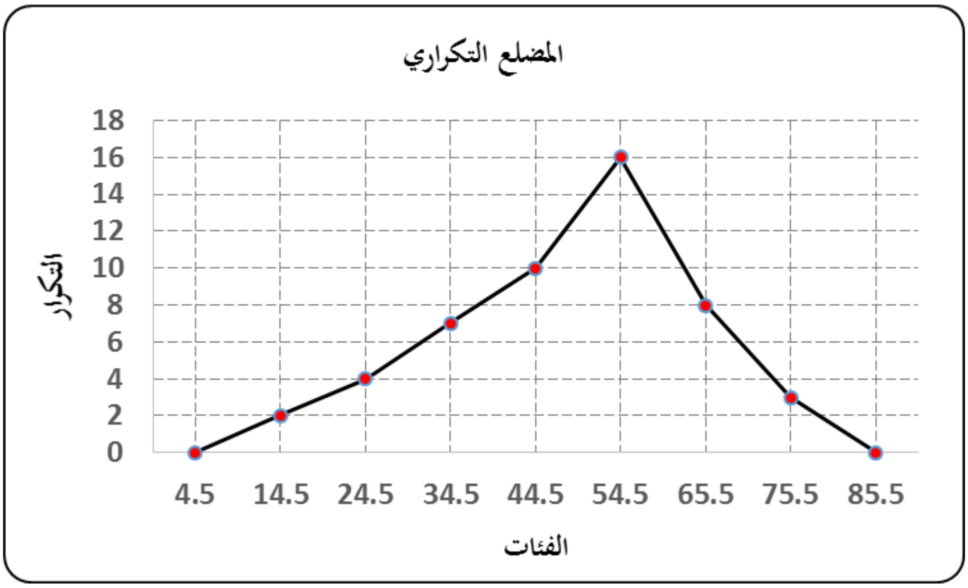
من الممكن أيضاً استخدام التكرار المئوي للفئات لنحصل عنها على المدرج التكراري المئوي كما يبينه الشكل التالي:



2.2.2. المضلع التكراري polygon:

يمثل المحور الأفقي الفئات ويمثل المحور الشاقولي تكرارها كما هو الحال بالنسبة للجدول التكراري ولكن بدلاً من رسم مستطيل يمثل ارتفاع التكرار نضع نقطة واحدة عند منتصف الفئة ارتفاعها يمثل التكرار. وفي الأخير نقوم بوصل النقاط التي حصلنا عليها بخطوط مستقيمة فنحصل على المضلع التكراري.

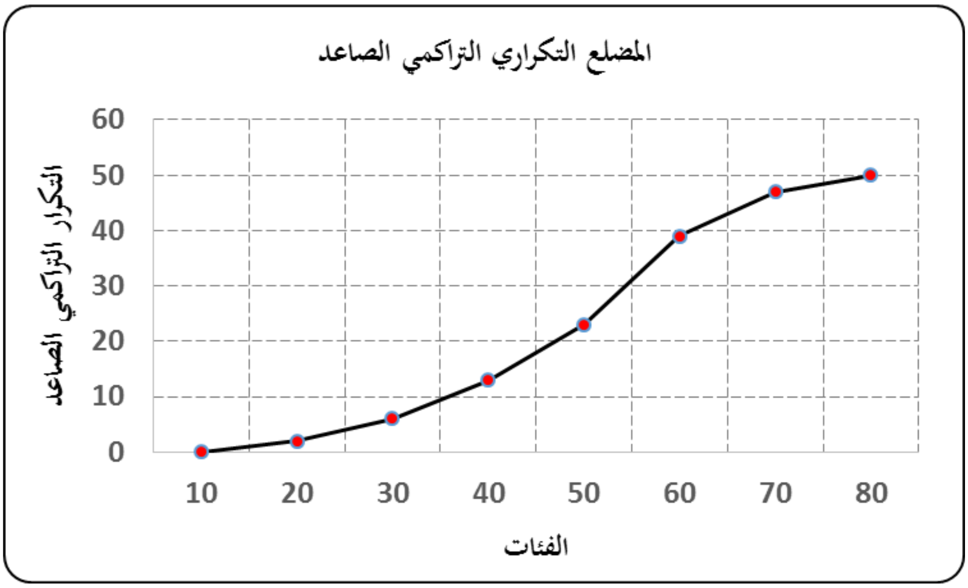
مثال 9: ليكن لدينا علامات طلاب الجامعة الافتراضية في مادة الإحصاء والاحتمالات السابقة. أوجد المضلع التكراري لعلامات الطلاب.



3.2.2. المنحني التكراري التراكمي cumulative frequency curve:

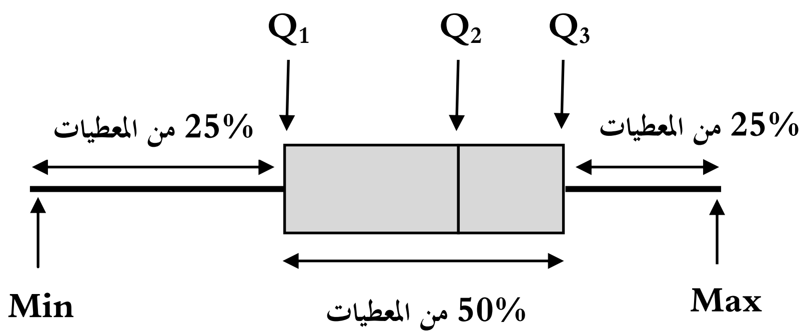
وهي نوعان تكرارية تراكمية صاعدة وتكرارية تراكمية هابطة. في كلا الحالتين، يمثل المحور الأفقي الفئات بحدودها الدنيا ويمثل المحور الشاقولي التكرار التراكمي الصاعد للفئات (الصاعدة) أو التكرار التراكمي الصاعد للفئات (الهابطة).

مثال 10: ليكن لدينا علامات طلاب الجامعة الافتراضية في مادة الإحصاء والاحتمالات السابقة. أوجد المنحني التكراري التراكمي لعلامات الطلاب.



4.2.2. المخطط الصندوقي box plot:

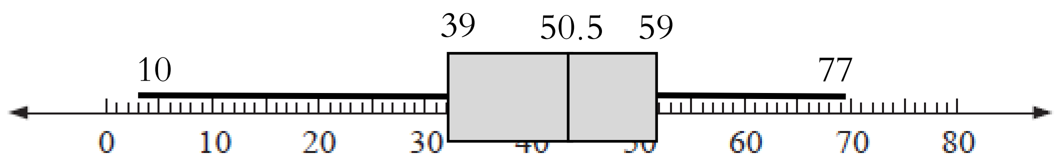
المخطط الصندوقي هو طريقة للتمثيل البياني للمعطيات العينة من خلال تمثيل القيم الإحصائية الخمس المحددة للعينة وهي: القيمة الصغرى min، الربيع الأول (الأدنى) Q_1 ويساوي القيمة التي يسبقها ربع المعطيات (أو 25% من المعطيات)، الوسيط median Q_2 ويساوي القيمة التي يسبقها نصف المعطيات (أو 50% من المعطيات)، الربيع الثالث (الأعلى) Q_3 ويساوي القيمة التي يسبقها ثلاثة أرباع المعطيات (أو 75% من المعطيات)، والقيمة العظمى max.



مثال 11: ليكن لدينا علامات طلاب الجامعة الافتراضية في مادة الإحصاء والاحتمالات السابقة. أوجد المخطط الصندوقي لعلامات الطلاب.

الحل:

$$Min = 10,\ Max = 77,\ Q_1 = 39,\ Q_2 = 50.5,\ Q_3 = 59$$



3.2. الرسوم البيانية chart:

تُستخدم الرسوم البيانية المختلفة أكثر الأحيان لتوضيح الجداول الإحصائية وجعل الرؤية للعلاقة بين المتغيرات أكثر سهولة. أهم الرسوم المستخدمة هي الخط البياني والأعمدة البيانية والرسوم الدائرية.

1.3.2. الخط البياني line chart:

الخط البياني عبارة عن خط منكسر يمثل عادة مسار المعطيات تم تحصيلها على فترات زمنية مختلفة. يمثل المحور الأفقي في هذه الحالة الزمن ويمثل المحور الشاقولي قيم المعطيات.

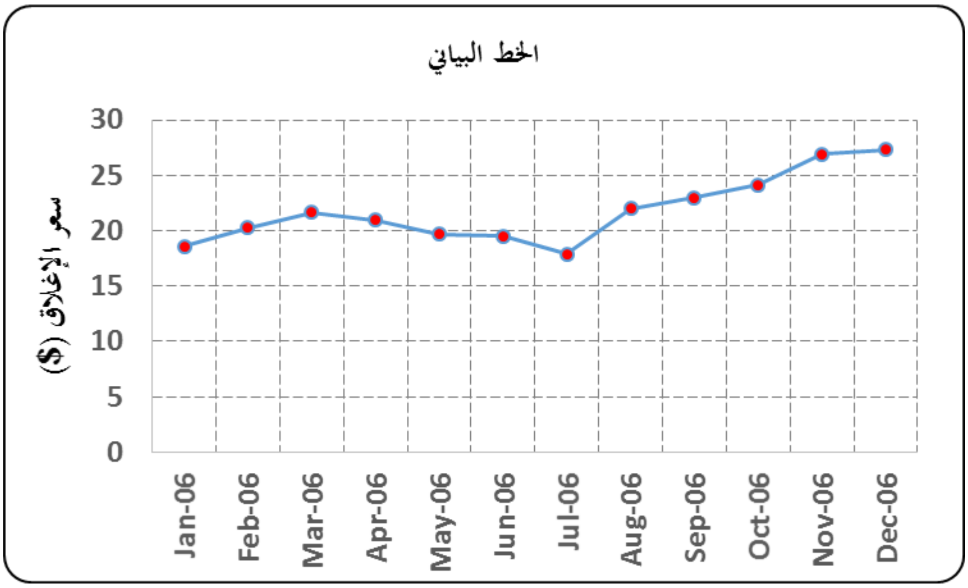
إذا كان لدينا أكثر من ظاهرة وقراءات قيم المعطيات لكل منها مأخوذة في الفترة الزمنية نفسها ويراد المقارنة بينها فإننا نرسم أكثر من خط للظواهر على نفس الرسمة.

مثال 12: تمثل معطيات الجدول التالي سعر إغلاق أسهم إحدى الشركات الأميركية بنهاية كل شهر من شباط 2006 ولغاية كانون الأول من 2006:

الشهر	سعر الإغلاق	الشهر	سعر الإغلاق
1/06	18.57	7/06	17.88
2/06	20.24	8/06	21.99
3/06	21.67	9/06	22.98
4/06	20.95	10/06	24.13
5/06	19.68	11/06	26.91
6/06	19.53	12/06	27.33

أوجد الخط البياني لمعطيات الجدول المذكور.

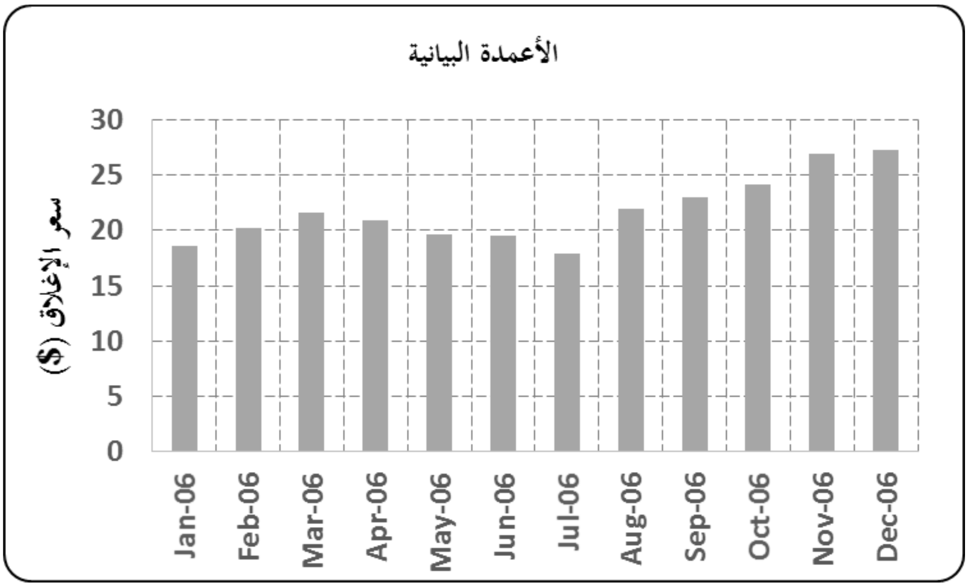
الحل:



2.3.2. الأعمدة البيانية bar chart:

الأعمدة البيانية عبارة عن مستطيلات عرضها متساوي وارتفاعها تمثل قيم المعطيات. وتستخدم عادة لتمثيل ظاهرة واحدة (الأعمدة البيانية البسيطة)، أو للمقارنة بين ظاهرتين أو مجموعة من الظواهر ومقارنة التطور فيما بينها (الأعمدة البيانية المزدوجة). كما أنه ليس من الضروري أن تكون قراءات قيم المعطيات مقاسة مع مرور الزمن.

مثال 13: أوجد الأعمدة البيانية لمعطيات الجدول السابق.

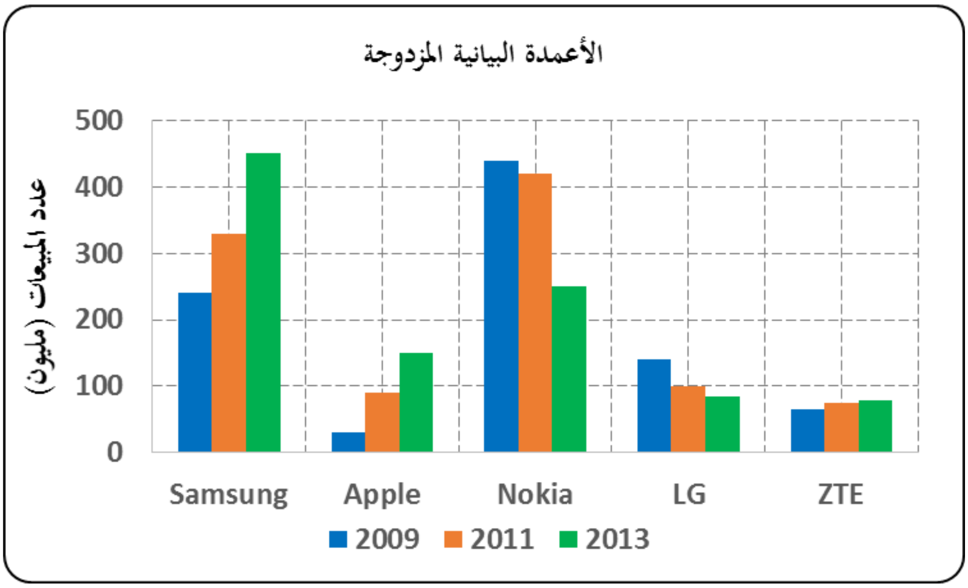


مثال 14: تمثل معطيات الجدول التالي مبيعات أشهر 5 شركات للأجهزة الخلوية بين الأعوام 2009 و2013.

الشركة/العام	2009	2011	2013
Samsung	240	330	450
Apple	30	90	150
Nokia	440	420	250
LG	140	100	85
ZTE	65	75	78

أوجد الأعمدة البيانية المزدوجة لمعطيات الجدول السابق.

الحل:



3.3.2.الرسوم الدائرية Pie chart:

الرسوم الدائرية عبارة عن دائرة يتم تقسيمها إلى قطاعات زاوية sectors، وكل قطاع زاوي يمثل فئة معينة من المعطيات ومساحة القطاع الزاوي متناسبة مع تكرار تلك الفئة f . ويتم حساب القطاع الزاوي لفئة معينة بالعلاقة التالية:

$$Angle = 360\text{x}\frac{f}{n}$$

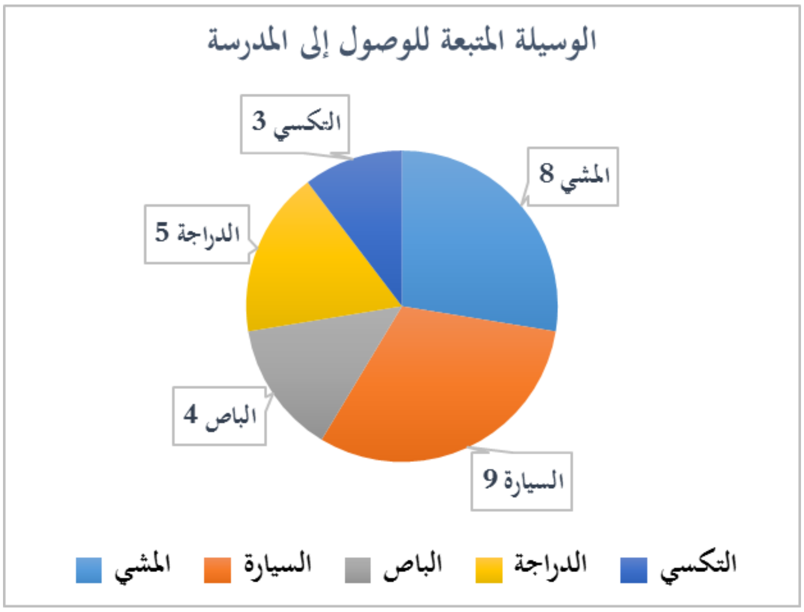
مثال 15: سئل طلاب صف في المدارس عن كيفية الوصول إلى المدرسة (الوسيلة المستخدمة). تم تجميع الأجوبة وحصلنا على الجدول التالي:

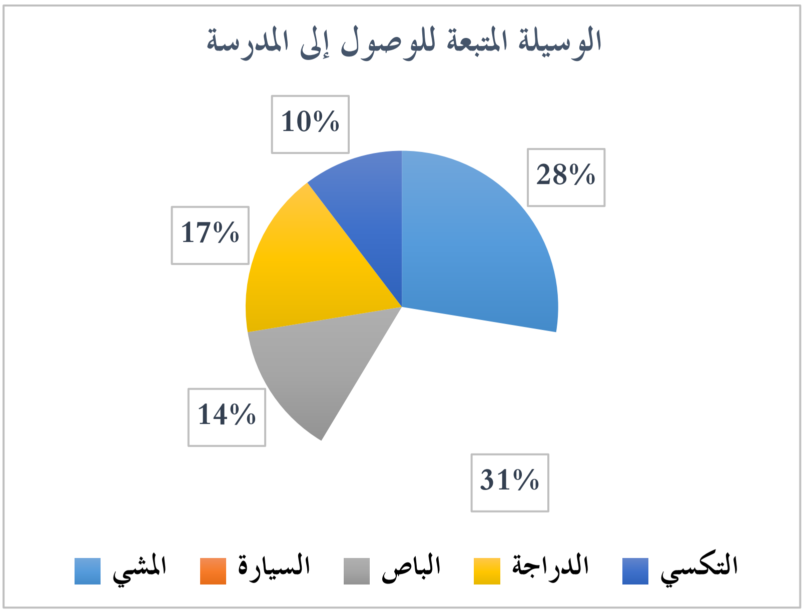
الوسيلة المستخدمة	عدد الطلاب

المشي	8
السيارة	9
الباص	4
الدراجة	5
التكسي	3

أوجد الرسوم الدائرية الموافقة.

الحل:





4.3.2. مخطط الساق والأوراق stem-and-leaf plots:

مخطط الساق والأوراق هو طريقة أخرى لتمثيل المعطيات الكمية بيانياً، فهي تساعد على تكوين فكرة جيدة عن المدى الذي تغطيه المعطيات وكيفية تمركزها. يتم تمثيل كل مفردة على شكل ورقة وساق، حيث يتم تشكيل الورقة من آخر رقم على اليمين من قيمة المفردة والساق يتكون من بقية الأرقام. على سبيل المثال القيمة 147 الورقة هي الرقم 7 والساق هو الرقم 14، أما بالنسبة للقيمة 5.8 الورقة هي الرقم 8 والساق هو الرقم 5.

مثال 16: ليكن لدينا علامات طلاب الجامعة الافتراضية في مادة الإحصاء والاحتمالات السابقة. أوجد مخطط الساق والأوراق لعلامات الطلاب.

الحل:

من الملاحظ أن علامات الطلاب تتغير بين 10 و77، بالتالي نشكل الساق من الأرقام 1 إلى 7 كما هو مبين في الشكل أقصى اليسار.

نستعرض القيم واحدة تلك الأخرى مثلاً نبدأ بالقيمة 23 فنضع الرقم 3 أمام الرقم 2 الموجود في الساق. وهكذا نستمر حتى تنتهي القيم ويكون الناتج كما هو مبين في الشكل في الوسط.

ضمن كل ساق نعيد ترتيب الأوراق تصاعدياً كما هو مبين في الشكل أقصى اليمين.

1	
2	
3	
4	
5	
6	
7	

1	0 5
2	3 6 1 7
3	8 3 9 5 6 4 0
4	2 7 3 7 8 8 1 3 4 6
5	0 2 6 9 1 4 2 5 8 9 1 0 7 4 6 3
6	3 4 8 5 0 2 5 7
7	5 7 2

1	0 5
2	1 3 6 7
3	0 3 4 5 6 8 9
4	1 2 3 3 4 6 7 7 8 8
5	0 0 1 1 2 2 3 4 4 5 6 6 7 8 9 9
6	0 2 3 4 5 5 7 8
7	2 5 7

مثال 17: أوجد مخطط الساق والأوراق للمعطيات التالية:

13.5	13.2	10.5	14.7	13.2	12.2	12.6	19.1	14.5	12.3	13.9	16.2
12.1	11.0	15.1	10.1	14.1	12.8	19.7	12.8	13.3	11.3	15.4	17.4
13.5	15.4	15.1	17.1	12.4	16.3	11.3	10.6	15.7	13.3	23.8	22.7
14.8	23.5	19.7	14.1								

الحل:

10	1 5 6
11	0 3 3
12	1 2 3 4 6 8 8
13	2 2 3 3 5 5 9
14	1 1 5 7 8
15	1 1 4 4 7
16	2 3
17	1 4
18	
19	1 7 7

20	
21	
22	**7**
23	**5 8**

3. مقاييس النزعة المركزية measures of central tendency:

مقاييس النزعة المركزية هي مقاييس عددية تستخدم لقياس موضع تركز أو تجمع المعطيات إذ أن معطيات أي ظاهرة تنزع (تميل) في الغالب إلى التركز والتجمع حول قيم معينة. وتُستخدم مقاييس النزعة المركزية لتلخيص المعطيات عددياً إذ أنها تعتبر قيم نموذجية أو مثالية للمعطيات. كما أن هذه المقاييس تُستخدم لوصف مجموعة المعطيات وكذلك لمقارنة مجموعات المعطيات المختلفة. ومن أهم هذه المقاييس نذكر: الوسط الحسابي أو المتوسط، الوسيط، والمنوال.

1.3. المتوسط mean:

يعتبر المتوسط من أهم وأفضل مقاييس النزعة المركزية ومن أكثرها شيوعاً واستخداماً في التحليل الإحصائي وذلك لما يتمتع به من خصائص وصفات إحصائية جيدة.

تعريف 2: إذا كان عدد المعطيات (حجم العينة) هو n وكانت قيم العينة هي $x_1, x_2, \ldots, x_n$، فإن المتوسط والذي يُرمز له بالرمز $\bar{x}$ ويعرف بالصيغة التالية:

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

مثال 18: أوجد المتوسط لعلامات طالب في الجامعة الافتراضية في المواد التي سجل فيها حيث حصل على العلامات التالية (8 مواد): 65, 75, 80, 95, 82, 74, 72, 67.

الحل:

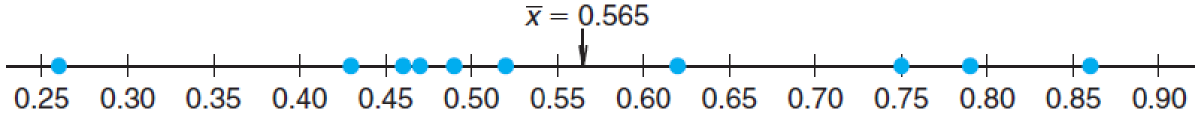
$$\bar{x} = \sum_{i=1}^{8} \frac{x_i}{8} = \frac{65+75+80+95+82+74+72+67}{8} = 76.25$$

مثال 19: أوجد المتوسط لعينة من 10 القيم: 0.26, 0.43, 0.47, 0.49, 0.52, 0.75, 0.79, 0.86, 0.62, 0.46.

الحل:

$$\bar{x} = \frac{0.26 + 0.43 + \ldots + 0.62 + 0.86}{10} = 0.565$$

ملاحظة 2: يمثل متوسط عينة من القيم مركز ثقل centroid تلك العينة، بمعنى أنه النقطة التي عندها يتوازن نظام من الأوزان قيمها هي قيم المعطيات الفردية. يبين الشكل التالي مركز ثقل قيم المثال 2.



ملاحظة 3: عندما يتم ظهور المفردة أكثر من مرة في العينة (تكرارها أكبر من الواحد)، وبفرض أن قيم المفردات المختلفة في العينة هي $x_1, x_2, \ldots, x_k$ وأن تكرار كل منها هو $f_1, f_2, \ldots, f_k$ على الترتيب، فإن المتوسط يحسب كما يلي:

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} f_i x_i}{n} = \frac{f_1 x_1 + f_2 x_2 + \ldots + f_k x_k}{n}$$

مثال 20: أوجد المتوسط لعينة من القياسات تمثل طول الساق لـــ 10 أنواع من النباتات بالسنتيمتر والمعطى بالجدول التالي (مثال 2):

طول الساق	التكرار
58	1
59	2
62	3
65	3
68	1
المجموع	**10**

الحل:

$$\bar{x} = \frac{\sum_{i=1}^{5} f_i x_i}{10} = \frac{1x58 + 2x59 + 3x62 + 3x65 + 1x68}{10} = 62.5$$

ملاحظة 4: عندما تكون معطيات العينة منظمة ضمن جدول تكراري باستخدام الفئات، فإنه يتم حساب المتوسط باستخدام العلاقة السابقة وحيث تكون القيم x_i عبارة عن مراكز الفئات.

مثال 21: أوجد المتوسط لعلامات طلاب الجامعة الافتراضية في مادة الإحصاء والاحتمالات المبينة في جدول التكرار التالي:

العلامة (M)	مركز الفئة	التكرار
$10 \le M < 20$	14.5	2
$20 \le M < 30$	24.5	4
$30 \le M < 40$	34.5	7
$40 \le M < 50$	44.5	10

$50 \le M < 60$	54.5	16
$60 \le M < 70$	64.5	8
$70 \le M < 80$	74.5	3
المجموع		**50**

الحل:

$$\bar{x} = \frac{2x14.5 + 4x24.5 + ... + 8x64.5 + 3x74.5}{50} = 48.5$$

خواص المتوسط:

ليكن لدينا العينة $x_1, x_2, \ldots, x_n$ والتي متوسطها $\bar{x}$:

1. المجموع الجبري لانحرافات القيم عن المتوسط $\bar{x}$ يساوي الصفر، أي أن:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = (x_1 - \bar{x}) + (x_2 - \bar{x}) + ... + (x_n - \bar{x}) = 0$$

2. إضافة أو طرح نفس القيمة b إلى قيم المعطيات التي متوسطها $\bar{x}$ يُنتج متوسطاً مقداره $\bar{x} \pm b$، أي أن: $\overline{(x \pm b)} = \bar{x} \pm b$.

3. ضرب قيم المعطيات التي متوسطها $\bar{x}$ بنفس القيمة a يُنتج متوسطاً مقداره $a\bar{x}$، أي أن: $\overline{ax} = a\bar{x}$.

4. إذا كان لدينا مجموعتان من المعطيات بحيث أن عدد معطيات المجموعة الأولى هو n_1 ومتوسطها هو $\overline{x_1}$وكان عدد معطيات المجموعة الثانية هو n_2 ومتوسطها هو $\overline{x_2}$ فإن متوسط المجموعة الكلية $\bar{x}$ المكونة من دمج هاتين المجموعتين يمكن حسابه كما يلي:

$$\bar{x} = \frac{n_1 \overline{x_1} + n_2 \overline{x_2}}{n_1 + n_2}$$

ملاحظة 5: يُمكن تجميع الخاصتين 2 و 3 كما يلي: $\overline{(ax \pm b)} = a\bar{x} \pm b$.

مثال 22: ليكن لدينا مجموعتان من المعطيات بحيث أن عدد معطيات المجموعة الأولى هو 8 ومتوسطها هو 4 وعدد معطيات المجموعة الثانية هو 12 ومتوسطها هو 8 فما هو متوسط المجموعة الكلية المكونة من دمج هاتين المجموعتين؟

الحل:

$$\bar{x} = \frac{n_1 \overline{x_1} + n_2 \overline{x_2}}{n_1 + n_2} = \frac{8x4 + 12x8}{8 + 12} = \frac{128}{20} = 6.4$$

مثال 23: إذا كان متوسط العينة $x_1, x_2, \ldots, x_n$ هو 25 فما هو المتوسط للعينة $\frac{x_1 - 5}{2}, \frac{x_2 - 5}{2}, \ldots, \frac{x_n - 5}{2}$؟

الحل:

المتوسط هو : $10 = \frac{25-5}{2} = \frac{\bar{x}-5}{2}$.

من ميزات المتوسط أنه: سهل التعريف والحساب، وحيد لمجموعة المعطيات الواحدة، وأخيرا يأخذ بعين الاعتبار كافة معطيات العينة.

ومن مساوئ المتوسط: يتأثر بالقيم الشاذة singular values، وهو غير معرف بالنسبة للمعطيات الوصفية إذ لا يمكن حسابه إلا من المعطيات الكمية فقط.

2.3. الوسيط median:

يعتبر الوسيط أحد مقاييس النزعة المركزية المشهورة، ويعرف الوسيط لمجموعة من المعطيات على أنه تلك القيمة التي تتوسط المعطيات عند ترتيبها تصاعدياً أو تنازلياً، أي أنه تلك القيمة التي تقسم المعطيات بعد ترتيبها إلى جزأين متساويين فتكون المعطيات في الجزء الأول تقل عن أو تساوى الوسيط والمعطيات في الجزء الثاني تزيد عن أو تساوى الوسيط. أي أن 50% من المعطيات تساوي أو تقل عن الوسيط وأن 50% من المعطيات تساوي أو تزيد عن الوسيط.

تعريف 3: إذا كان عدد المعطيات (حجم العينة) هو n وكانت قيم العينة هي $x_1, x_2, \ldots, x_n$ والمرتبة ترتيباً تصاعدياً، فإن الوسيط والذي يُرمز له بالرمز x ويعرف بالصيغة التالية (نميز حالتين حسب القيمة n هل هي عدد فردي أم زوجي):

$$x = \begin{cases} x_{(n+1)/2}, & if \;\; n \; is \; odd \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}), & if \;\; n \; is \; even \end{cases}$$

مثال 24: أوجد الوسيط لعلامات طالب التالية: 65, 75, 80, 95, 82, 74, 72, 67, 79.

الحل:

بعد ترتيب العلامات تصبح على النحو التالي: ~~65, 67, 72, 74~~, 75, ~~79, 80, 82, 95~~. بما أن عدد العينات $n = 9$ هو عدد فردي فإن الوسيط هو :

$$x = x_{(n+1)/2} = x_{(9+1)/2} = x_5 = 75$$

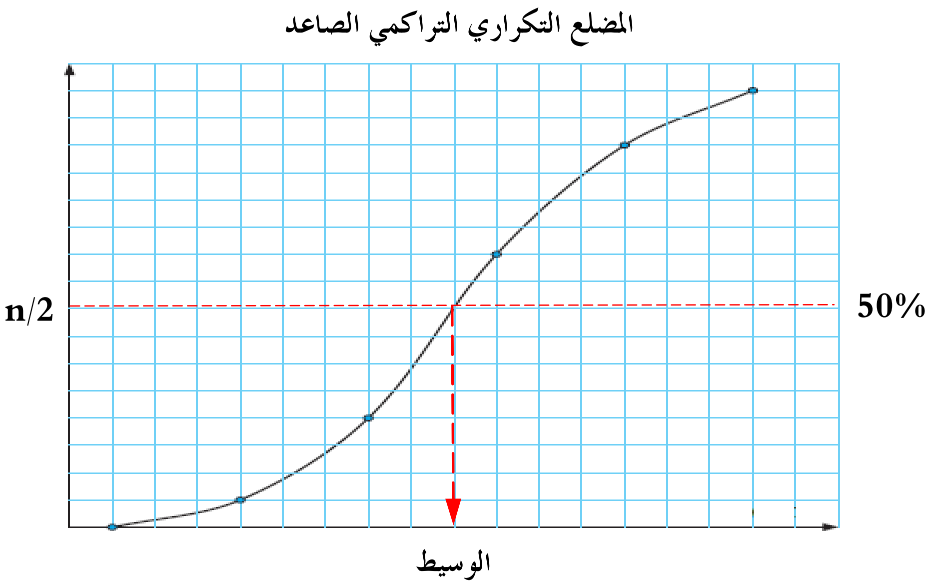
مثال 25: أوجد الوسيط لعلامات طالب في 10 مواد والتي هي التالية: 65, 75, 80, 95, 82, 74, 72, 67, 79, 67.

الحل:

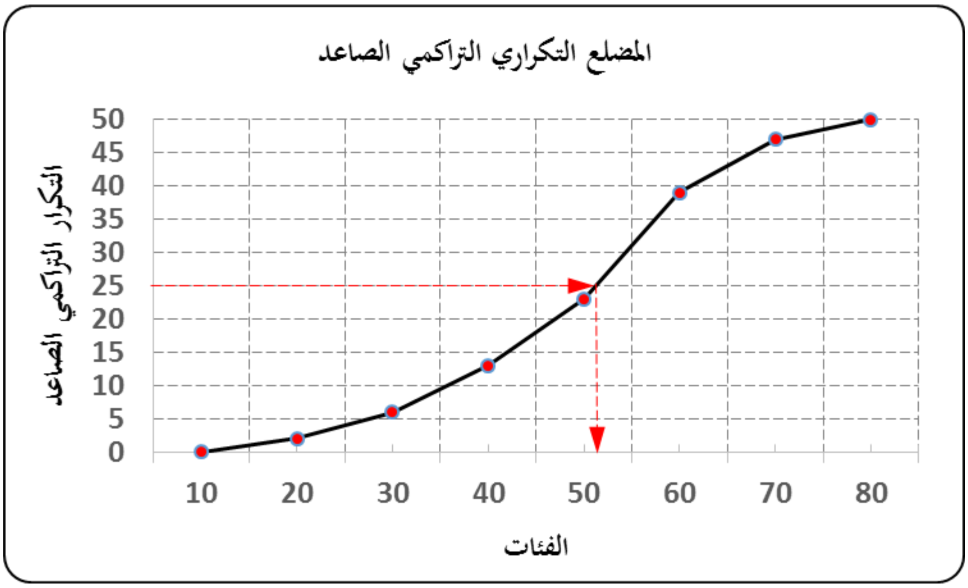
بعد ترتيب العلامات تصبح على النحو التالي: ~~65, 67, 67, 72~~, 74, 75, ~~79, 80, 82, 95~~. بما أن عدد العينات $n = 10$ هو عدد فردي فإن الوسيط هو :

$$x = \frac{1}{2}(x_{n/2} + x_{n/2+1}) = \frac{1}{2}(x_5 + x_6) = \frac{1}{2}(74 + 75) = 74.5$$

ملاحظة 6: عندما تكون معطيات العينة منظمة ضمن مضلع تكراري تراكمي صاعد، فإنه يمكننا حساب القيمة التقريبية للوسيط بيانياً، ويتم ذلك عن طريق تحديد موقع ترتيب الوسيط المساوي إلى $n/2$ (سواء كان عدد معطيات العينة n فردياً أو زوجياً) على المحور الشاقولي، ومن هذا الموقع نرسم خطاً افقياً يتقاطع مع المضلع في نقطة. ومن هذه النقطة نرسم عموداً يتقاطع مع المحور الأفقي في نقطة تمثل القيمة التقريبية للوسيط، كما يبينه الشكل التالي:



مثال 26: أوجد القيمة التقريبية لوسيط علامات طلاب الجامعة الافتراضية في مادة الإحصاء والاحتمالات المبينة في المضلع التكراري التراكمي التصاعدي التالي:



الحل:

من المضلع التكراري التراكمي التصاعدي نستطيع قراءة الوسيط والذي يساوي تقريباً 51.

من ميزات الوسيط أنه: سهل التعريف والحساب، وحيد لمجموعة المعطيات الواحدة، وأخيرا الوسيط أقل تأثراً من المتوسط بالقيم الشاذة.

ومن مساوئ الوسيط: لا يأخذ بعين الاعتبار جميع المعطيات إذ أنه يعتمد فقط على القيم التي في المنتصف وعلى ترتيب المعطيات بغض النظر عن قيمها، وهو غير معرف بالنسبة للمعطيات الوصفية إذ لا يمكن حسابه إلا من المعطيات الكمية فقط.

3.3. المنوال mode:

المنوال هو أحد مقاييس النزعة المركزية شائعة الاستخدام ولاسيما في حالة المعطيات الوصفية.

تعريف 4: يعرف المنوال لمجموعة من المعطيات على أنه تلك القيمة التي تتكرر أكثر من غيرها أي أنها القيمة ذات التكرار الأكبر (إن وجدت)، ويرمز للمنوال عادة بالرمز Mod.

ملاحظة 7: يُمكن لعينة من المعطيات ألا يكون لها منوال، كما يُكن أن يكون لها أكثر من منوال.

مثال 27: يبين الجدول التالي معطيات إحدى الدراسات التي طبقت على 8 أشخاص لقياس العمر (بالسنة) والوزن (بالكلغ) والطول (بالسنتيمتر) وفصيلة الدم. أوجد منوال المعطيات المختلفة.

الشخص	1	2	3	4	5	6	7	8
العمر	25	32	27	23	24	32	25	37
الوزن	66	68	70	71	77	72	65	68
الطول	166	163	164	160	159	161	168	169
فصيلة الدم	A	B	AB	O	O	AB	A	O

الحل:

المعطيات	المنوال
العمر	25 و 32
الوزن	68
الطول	لا يوجد
فصيلة الدم	O

ملاحظة 8: عندما تكون معطيات العينة منظمة بجدول تكراري نستطيع حساب المنوال بطريقة حسابية. من أجل ذلك نعرف الفئة المنوالية بأنها الفئة ذات التكرار الأكبر (ويمكن أن يكون هناك فترة منوالية واحدة أو عدة فترات منوالية أو قد لا يوجد فترة منوالية)، ويكون المنوال في حال وجوده عبارة عن مركز الفئة.

مثال 28: ليكن لدينا علامات طلاب الجامعة الافتراضية في مادة الإحصاء والاحتمالات السابقة.

العلامة (M)	**التكرار**
$10 \leq M < 20$	2
$20 \leq M < 30$	4
$30 \leq M < 40$	7
$40 \leq M < 50$	10
$50 \leq M < 60$	16
$60 \leq M < 70$	8
$70 \leq M < 80$	3
المجموع	**50**

أوجد قيمة المنوال حسابياً.

الحل:

الفترة المنوالية هي $[50, 60[$ لأنها الفترة الأكبر تكراراً (16)، بالتالي المنوال يساوي 54.5 (مركز الفئة).

من ميزات المنوال: أنه سهل التعريف والحساب والمنوال أقل تأثراً من المتوسط بالقيم الشاذة، ويمكن حساب المنوال بالنسبة للمعطيات الوصفية.

ومن مساوئ الوسيط: لا يأخذ بعين الاعتبار جميع المعطيات إذ أنه يعتمد فقط على المعطيات ذات التكرار الأكثر، وقد لا يوجد منوال لمجموعة من المعطيات أو قد يكون هناك أكثر من منوال.

4. مقاييس التباين measures of variability:

وجدنا سابقاً بعض مقاييس النزعة المركزية والتي تمثل مقاييس عددية لموضع أو مكان تركز المعطيات لظاهرة ما، وقد ذكرنا بأن هذه المقاييس تُستخدم لمقارنة مجموعات المعطيات المختلفة. في الحقيقة فإن مقاييس النزعة المركزية غير كافية لإيجاد مقارنة شاملة بين تلك المجموعات، فقد تكون هناك مجموعات من المعطيات لها نفس مقاييس النزعة المركزية ولكنها تختلف في بعض الصفات الأخرى. يبين المثال التالي مجموعتين من المعطيات لهما نفس المتوسط ولكنهما مختلفتان في طبيعة تشتتهما.

مثال 29: ليكن لدينا العينتين التاليتين:

العينة الأولى: 60 ,59 ,58 ,57 ,56 ,55 ,54 ,53 ,52 ,51 ,50

العينة الثانية: 75 ,71 ,67 ,63 ,59 ,55 ,51 ,47 ,43 ,39 ,35

بالرغم من أن المتوسط يساوي 55 للعينتين إلا أن التشتت أو الاختلاف بين القيم في كل عينة غير متساوٍ. ومن الواضح أن معطيات العينة الأولى أكثر تقارباً فيما بينها (أقل تشتتاً وتباعداً فيما بينها) من معطيات العينة الثانية. لذلك دعت الحاجة لإيجاد مقاييس تقيس طبيعة تشتت (أو تفرق أو تباعد) المعطيات فيما بينها. هذه المقاييس تسمى مقاييس التشتت.

يكون تشتت المعطيات صغيراً إذا كانت قيمها متقاربة فيما بينها والعكس بالعكس. تستخدم مقاييس التشتت لوصف مجموعة المعطيات وكذلك لمقارنة مجموعات المعطيات المختلفة، إذ أن مقاييس النزعة المركزية لا تكفي وحدها لعملية الوصف أو المقارنة. من أشهر مقاييس التشتت نذكر: المدى، المدى الربيعي، التباين، الانحراف المعياري ومعامل الاختلاف.

1.4. المدى range:

يعتبر المدى من أسهل مقاييس التشتت تعريفًا وحساباً ويعطينا فكرة سريعة عن مدى تفرق المعطيات.

تعريف 5: يعرف المدى لمجموعة من المعطيات بالصيغة التالية: $Range = x_{\max} - x_{\min}$، حيث أن $x_{\max}$ تعبر عن أكبر قيمة للمعطيات و $x_{\min}$ تعبر عن أصغر قيمة.

مثال 30: أوجد المدي للعينة التالية: 19 ,26 ,37 ,41 ,45 ,17 ,34 ,22.

الحل: $Range = 45 - 19 = 26$

من ميزات المدى أنه: سهل التعريف والحساب.

من مساوئ المدى: لا يأخذ المدى في الاعتبار كافة القيم، كما أنه يتأثر بالقيم الشاذة.

2.4. المدى الربيعي interquartile range:

رأينا أن المدى يتأثر كثيراً بالقيم الشاذة ولذلك دعت الحاجة لإيجاد مقاييس أخرى للتشتت لا تتأثر بالقيم المتطرفة. أحد هذه المقاييس هو المدى الربيعي، حيث أن القيم المتطرفة هي تلك القيم الصغيرة أو الكبيرة، بالتالي فإنه عند حساب المدى الربيعي لا يؤخذ في الاعتبار ربع المعطيات الصغيرة (25%) ولا ربع المعطيات الكبيرة (25%). أي أنه يهتم بقيمتين، الأولى هي تلك التي يقل عنها ربع عدد أفراد العينة فقط، والثانية هي تلك التي يزيد عنها ربع عدد أفراد العينة فقط، وهما يسميان على الترتيب الربيع الأول أو الأدنى والربيع الأعلى أو الثالث.

تعريف 6: نرمز للمدى الربيعي بالرمز IRQ ويعرف بالصيغة التالية: $IRQ = Q_3 - Q_1$، حيث Q_1 يعبر عن الربيع الأول ويساوي القيمة التي يسبقها ربع المعطيات (أو 25% من المعطيات) و Q_3 يعبر عن الربيع الثالث ويساوي القيمة التي يسبقها ثلاثة أرباع المعطيات (أو 75% من المعطيات).

¼ n =25% ¼ n= 25%

-/-/-/-/-/-/-/-/-/-|----------------|----------------|-/-/-/-/-/-/-/-/-/-

ربع البيانات الصغيرة Q_1 Q_2 Q_3 ربع البيانات الكبيرة

ملاحظة 9: يقسم الوسيط العينة المرتبة إلى نصفين النصف الأدنى والنصف الأعلى، الربيع الأول هو وسيط النصف الأدنى لتلك العينة والربيع الثالث هو وسيط النصف الأعلى.

مثال 31: ليكن لدينا العينة التالية: 7, 3, 1, 7, 6, 9, 3, 8, 5, 8, 6, 3, 7, 1, 9. أوجد كل من الوسيط والربيع الأول والربيع الثالث والمدى الربيعي.

الحل:

بعد ترتيب قيم العينة تصبح على النحو التالي: ~~1, 1, 3, 3, 3, 5, 6~~, 6, ~~7, 7, 7, 8, 8, 9, 9~~ وبما أن $n = 15$ عدد فردي فإن الوسيط هو العينة التي ترتيبها 8 = (n+1)/2 = (15+1)/2، أي أنه يساوي 6 (العينة الثامنة).

من أجل حساب الربيع الأول نحسب الوسيط للعينة ~~1, 1, 3~~, 3, ~~3, 5, 6~~ بالتالي $Q_1 = 3$، ومن أجل حساب الربيع الثالث نحسب الوسيط للعينة ~~7, 7, 7~~, 8, ~~8, 9, 9~~ بالتالي $Q_1 = 8$. أخيراً المدى الربيعي $IRQ = Q_3 - Q_1 = 8 - 3 = 5$.

مثال 32: ليكن لدينا العينة التالية:

6, 4, 9, 15, 5, 13, 7, 12, 8, 10, 4, 1, 13, 1, 6, 4, 5, 2, 8, 2

أوجد كل من الوسيط والربيع الأول والربيع الثالث والمدى الربيعي.

الحل:

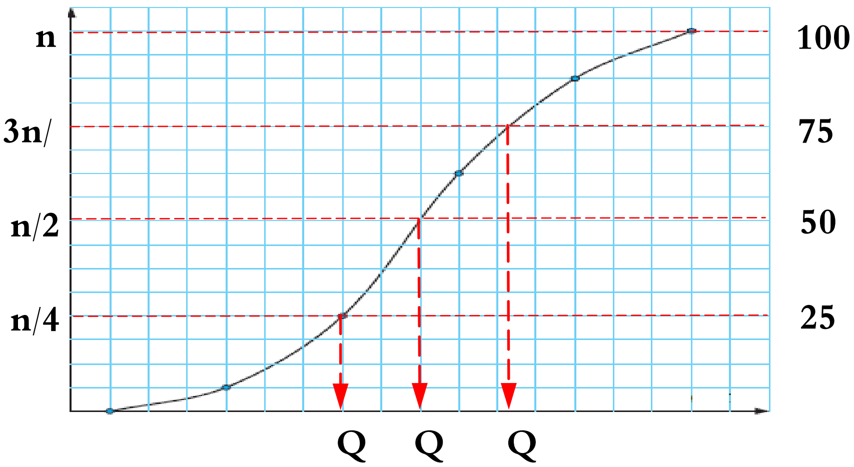
بعد ترتيب قيم العينة تصبح على النحو التالي: ~~1, 1, 2, 2, 4, 4, 4, 5, 5,~~ 6, 6, ~~7, 8, 8, 9, 10, 12, 13, 13, 15~~ وبما أن $n=20$ عدد زوجي فإن الوسيط يساوي 6 = (6+6)/2.

من أجل حساب الربيع الأول نحسب الوسيط لعينة النصف الأدنى ~~1, 1, 2, 2,~~ 4, 4, ~~4, 5, 5, 6~~ بالتالي وبملاحظة أن $n=10$ عدد زوجي، $Q_1=(4+4)/2=4$، ومن أجل حساب الربيع الثالث نحسب الوسيط لعينة النصف الأعلى ~~6, 7, 8, 8,~~ 9, 10, ~~12, 13, 13, 15~~ بالتالي $Q_3=(9+10)/2=9.5$.

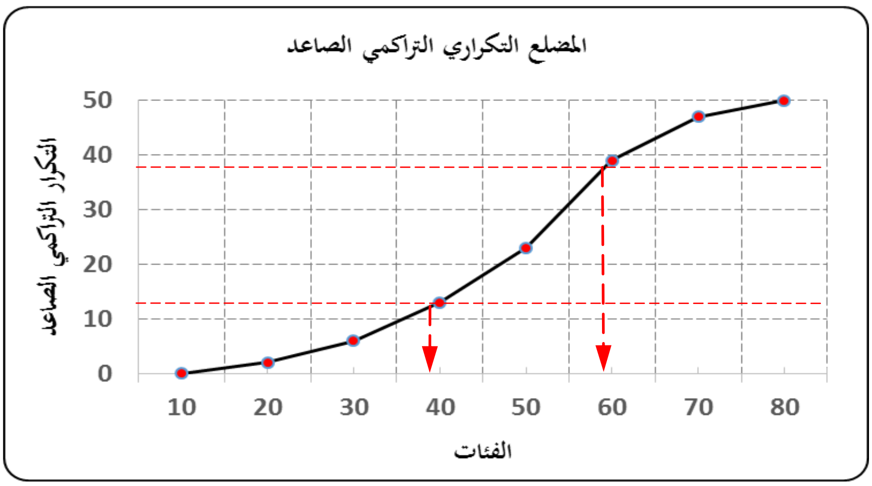
أخيراً المدى الربيعي $IRQ=Q_3-Q_1=9.5-4=5.5$.

ملاحظة 10: عندما تكون معطيات العينة منظمة ضمن مضلع تكراري تراكمي صاعد، فإنه يمكننا حساب القيمة التقريبية للمدى الربيعي بيانياً، ويتم ذلك عن طريق حساب كل من الربيع الأول Q_1 (الموافق لـ $n/4$) والربيع الثالث Q_3 (الموافق لـ $3n/4$) بطريقة مشابهه تماماً لحساب الوسيط (الموافق لـ $n/2$) التي وجدناها سابقاً، كما يبينه الشكل التالي:

المضلع التكراري التراكمي



مثال 33: أوجد القيمة التقريبية للمدى الربيعي لعلامات طلاب الجامعة الافتراضية في مادة الإحصاء والاحتمالات المبينة في المضلع التكراري التراكمي التصاعدي التالي:



الحل:

من المضلع التكراري التراكمي التصاعدي نستطيع قراءة الربيع الأول والذي يساوي تقريباً 39، والربيع الثالث والذي يساوي تقريباً 59. بالتالي المدى الربيعي يساوي 20 = 39 − 59.

من ميزات المدى الربيعي أنه: لا يتأثر بالقيم الشاذة.

من مساوئ المدى: لا يأخذ المدى الربيعي في الاعتبار كافة قيم العينة.

3.4. التشتت variance والانحراف المعياري standard deviation:

يعتبر التباين والانحراف المعياري من أهم وأفضل مقاييس التشتت ومن أكثرها شيوعاً واستخداماً في التحليل الإحصائي وذلك لما يتمتعان به من خصائص وصفات إحصائية جيدة.

1.3.4. التشتت:

تعتمد فكرة التباين على تشتت أو تباعد قيم المعطيات عن متوسطها، فالتباين يكون كبيراً إذا كانت قيم المعطيات متباعدة عن متوسطها والعكس بالعكس. ويعرف التباين بأنه متوسط مربع انحرافات القيم عن الوسط الحسابي.

تعريف 7: إذا كان عدد المعطيات (حجم العينة) هو n وكانت قيم العينة هي $x_1, x_2, \ldots, x_n$ ومتوسطها $\bar{x}$ فإن التباين والذي يُرمز له بالرمز s^2 ويعرف بالصيغة التالية:

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

مثال 34: ضمن تجربة اختبار انحراف مقياس الـ PH. تم تجميع معطيات عن المقياس بقياس درجة ال PH لمادة محايدة $(PH = 7.0)$. عينه من حجم $n = 10$ تم أخذها في هذا الخصوص، أعطت النتائج التالية:

7.07, 7.00, 7.10, 6.97, 7.00, 7.03, 7.01, 7.01, 6.98, 7.08

متوسط العينة $\bar{x}$ هو:

$$\bar{x} = \frac{7.07 + 7.00 + \ldots + 7.08}{10} = 7.025$$

تشتت العينة s^2 هو:

$$s^2 = \frac{(7.07 - 7.025)^2 + (7.00 - 7.025)^2 + \ldots + (7.08 - 7.025)^2}{9} = 0.00194$$

يمكن حساب التشتت باستخدام صيغة حسابية مختلفة (أكثر سهولة وأكثر دقة)، بملاحظة أن:

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1} = \sum_{i=1}^{n} \frac{(x_i^2 - 2x_i\bar{x} + \bar{x}^2)}{n-1} = \sum_{i=1}^{n} \frac{x_i^2}{n-1} - \sum_{i=1}^{n} \frac{-2x_i\bar{x}}{n-1} + \sum_{i=1}^{n} \frac{\bar{x}^2}{n-1}$$

وأن المقدار:

$$\sum_{i=1}^{n} \frac{-2x_i\bar{x}}{n-1} = -2\bar{x}\sum_{i=1}^{n} \frac{x_i}{n-1} = -2\bar{x}\sum_{i=1}^{n} \frac{nx_i}{n(n-1)} = -\frac{2n\bar{x}}{n-1}\sum_{i=1}^{n} \frac{x_i}{n} = -\frac{2n\bar{x}^2}{n-1}$$

وأن:

$$\sum_{i=1}^{n} \frac{\bar{x}^2}{n-1} = \frac{n\bar{x}^2}{n-1}$$

بالنتيجة:

$$s^2 = \sum_{i=1}^{n} \frac{x_i^2}{n-1} - \frac{n\bar{x}^2}{n-1}$$

في المثال السابق:

$$s^2 = \sum_{i=1}^{10} \frac{x_i^2}{9} - \frac{10\bar{x}^2}{9} = \frac{7.07^2 + 7.00^2 + \ldots + 7.08^2}{9} - \frac{10(7.025)^2}{9} = 0.00194$$

ملاحظة 11: عندما يتم ظهور المفردة أكثر من مرة في العينة (تكرارها أكبر من الواحد)، وبفرض أن قيم المفردات المختلفة في العينة هي $x_1, x_2, \ldots, x_k$ وأن تكرار كل منها هو $f_1, f_2, \ldots, f_k$ على الترتيب، فإن التشتت يحسب كما يلي:

$$s^2 = \sum_{i=1}^{k} \frac{f_i(x_i - \bar{x})^2}{n-1}$$

أو باستخدام العلاقة:

$$s^2 = \sum_{i=1}^{k} \frac{f_i x_i^2}{n-1} - \frac{n\bar{x}^2}{n-1}$$

مثال 35: أوجد التشتت لعينة من القياسات تمثل طول الساق لـــ 10 أنواع من النباتات بالسنتم والمعطى بالجدول التالي (مثال 2):

طول الساق	التكرار
58	1
59	2
62	3
65	3
68	1
المجموع	**10**

الحل:

$$\bar{x} = \frac{\sum_{i=1}^{5} f_i x_i}{10} = \frac{1x58 + 2x59 + 3x62 + 3x65 + 1x68}{10} = 62.5$$

$$s^2 = \sum_{i=1}^{5} \frac{f_i x_i^2}{9} - \frac{10\bar{x}^2}{9} = \frac{1x58^2 + 2x59^2 + 3x62^2 + 3x65^2 + 1x68^2}{9} - \frac{10x62.5^2}{9}$$

$$s^2 = 10.5$$

ملاحظة 12: عندما تكون معطيات العينة منظمة ضمن جدول تكراري باستخدام الفئات، فإنه يتم حساب التشتت باستخدام العلاقة السابقة وحيث تكون القيم x_i عبارة عن مراكز الفئات.

مثال 36: أوجد التشتت لعلامات طلاب الجامعة الافتراضية في مادة الإحصاء والاحتمالات المبينة في جدول التكرار التالي:

العلامة (M)	مركز الفئة	التكرار
$10 \le M < 20$	14.5	2
$20 \le M < 30$	24.5	4
$30 \le M < 40$	34.5	7
$40 \le M < 50$	44.5	10
$50 \le M < 60$	54.5	16
$60 \le M < 70$	64.5	8
$70 \le M < 80$	74.5	3
المجموع		**50**

الحل:

$$\bar{x} = \frac{2x14.5 + 4x24.5 + 7x34.5 + 10x44.5 + 16x54.5 + 8x64.5 + 3x74.5}{50}$$

$$\bar{x} = 48.5$$

$$s^2 = \sum_{i=1}^{7} \frac{f_i x_i^2}{49} - \frac{50\bar{x}^2}{49} = \frac{2x14.5^2 + 4x24.5^2 + \ldots + 3x74.5^2}{49} - \frac{50x48.7^2}{49}$$

$$s^2 = 229.8$$

2.3.4.الانحراف المعياري:

إن التباين من أهم وأفضل مقاييس التشتت ولكنه يقاس بوحدة المعطيات الأصلية المربعة. وفي كثير من الأحيان نرغب في استخدام مقياس للتشتت يقاس بوحدة المعطيات الأصلية، أحد هذه المقاييس هو الانحراف المعياري. ويعرف الانحراف المعياري على أنه الجذر التربيعي للتباين.

تعريف 8: إذا كان عدد المعطيات (حجم العينة) هو n وكانت قيم العينة هي $x_1, x_2, \ldots, x_n$ وتشتتها s^2 فإن الانحراف المعياري والذي يُرمز له بالرمز s ويعرف بالصيغة التالية:

$$s=\sqrt{s^2}=\sqrt{\sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{n-1}}$$

أو عندما يتم ظهور المفردة أكثر من مرة في العينة، وبفرض أن تكرار كل منها هو $f_1, f_2, \ldots, f_k$ على الترتيب:

$$s=\sqrt{s^2}=\sqrt{\sum_{i=1}^{k}\frac{f_i(x_i-\bar{x})^2}{n-1}}$$

مثال 37: أوجد الانحراف المعياري لعينة المثال 11.

الحل:

$$s=\sqrt{0.00194}=0.044$$

خواص التشتت والانحراف المعياري:

يخضع التشتت والانحراف المعياري إلى العمليات الجبرية التالية:

العينة	الانحراف المعياري	التشتت
$x_1, x_2, \ldots, x_n$	s	s^2
$x_1 \pm b, x_2 \pm b, \ldots, x_n \pm b$	s	s^2
$ax_1, ax_2, \ldots, ax_n$	$\lvert a \rvert s$	a^2s^2
$ax_1 \pm b, ax_2 \pm b, \ldots, ax_n \pm b$	$\lvert a \rvert s$	a^2s^2

مثال 38:

العملية	العينة	الانحراف المعياري	التشتت

x	4, 6, 5, 8, 9, 7	1.871	3.5
$x+2$	6, 8, 7, 10, 11, 9	1.871	3.5
$3x$	12, 18, 15, 24, 27, 21	3x1.871 = 5.612	9x3.5 = 31.5
$2x+3$	11, 15, 13, 19, 21, 17	2x1.871 = 3.742	4x3.5 = 14

مثال 39: إذا كان تشتت العينة $x_1, x_2, \ldots, x_n$ هو 25 فما هو تشتت العينة $\frac{x_1-5}{2}, \frac{x_2-5}{2}, \ldots, \frac{x_n-5}{2}$؟

الحل:

التشتت هو:

$$s^2 = \left(\frac{1}{2}\right)^2 \text{x} 25 = \frac{25}{4} = 6.25$$

والانحراف المعياري هو:

$$s = \sqrt{6.25} = 2.5$$

1. إذا كان لدينا مجموعتان من المعطيات بحيث أن عدد معطيات المجموعة الأولى هو n_1 وتشتتها هو s_1^2 وكان عدد معطيات المجموعة الثانية هو n_2 وتشتتها هو s_2^2 فإن تشتت المجموعة الكلية $\bar{x}$ المكونة من دمج هاتين المجموعتين يمكن حسابه كما يلي:

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-1}$$

مثال 40: ليكن لدينا مجموعتان من المعطيات بحيث أن عدد معطيات المجموعة الأولى هو 8 وتشتتها هو 4 وعدد معطيات المجموعة الثانية هو 13 وتشتتها هو 8 فما هو تشتت المجموعة الكلية المكونة من دمج هاتين المجموعتين؟

الحل:

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-1} = \frac{7\text{x}4 + 12\text{x}8}{8+13-1} = 6.2$$

4.4. معامل الاختلاف coefficient of variation:

إن التباين والانحراف المعياري من المقاييس المفيدة لقياس التشتت لتوزيع متغير ما، ولكن في كثير من الأحيان نهتم بمقارنة التشتت والاختلاف لتوزيعي متغيرين مختلفين. وبما أن التباين والانحراف المعياري مقياسان يعتمدان على وحدة المعطيات فإنه يصعب استخدامهما لمقارنة تجانس المجموعات المختلفة من المعطيات في الحالتين التاليتين:

- إذا كانت وحدتا المتغيرين مختلفتين حيث لا نستطيع مقارنة الوحدات المختلفة.
- إذا كان متوسطا المتغيرين مختلفين وذلك لأن تباين توزيع المتغير ذي المتوسط الصغير ينزع لأن يكون صغيراً والعكس بالعكس.

لذلك دعت الحاجة إلى مقياس لا يعتمد على وحدة المتغير ويقيس ما يسمى بالتشتت النسبي. وأحد هذه المقاييس هو ما يسمى بمعامل الاختلاف ويستخدم لمقارنة التشتت النسبي أو التجانس لمجموعات المعطيات المختلفة. فمجموعة المعطيات ذات معامل الاختلاف الأكبر يكون تشتتها النسبي أكبر أي أنها تكون أقل تجانساً والعكس بالعكس.

تعريف 9: ليكن لدينا عينة من المعطيات متوسطها $\overline{x}$ وانحرافها المعياري s، نرمز بمعامل الاختلاف للعينة المذكورة بالرمز CV ويعرف بالصيغة التالية: $CV = \frac{s}{\overline{x}}$.

مثال 41: يبين الجدول التالي معطيات إحدى الدراسات التي طبقت على 8 أشخاص لقياس العمر (بالسنة) والوزن (بالكلغ) والطول (بالسنتيمتر). أوجد المعطيات الأكثر تجانساً معطيات الأعمار أم معطيات الأوزان أم معطيات الأطوال؟

الشخص	1	2	3	4	5	6	7	8
العمر	25	32	27	23	24	32	25	37
الوزن	66	68	70	71	77	72	65	68
الطول	166	163	164	160	159	161	168	169

الحل:

المعطيات	المتوسط	الانحراف المعياري	معامل الاختلاف
الأعمار	28.125	4.970	0.177
الأوزان	69.625	3.815	0.055
الأطوال	163.750	3.694	0.023

من الجدول السابق نستنتج أن معطيات الأطوال هي الأكثر تجانساً لأن معامل الاختلاف للأطوال أصغر من معامل الاختلاف للمعطيات الأخرى.

5.4. متراجحة تشيبيشيف Chebechev inequality:

تعتبر متراجحة تشيبيشيف من النظريات المفيدة التي تعطينا حداً أدنى لنسبة المعطيات التي تقع ضمن مجال معين عند معرفة متوسط المعطيات وانحرافها المعياري دون الحاجة لمعرفة المعطيات الأصلية أو التوزيع الذي أخذت منه العينة.

مبرهنة 1: ليكن لدينا عينة من المعطيات متوسطها $\bar{x}$ وانحرافها المعياري s فإن نسبة المعطيات الواقعة ضمن المجال $[\bar{x}-ks, \bar{x}+ks]$ لا تقل عن $\left(1-\dfrac{1}{k^2}\right)$، حيث $k>1$.

ملاحظة 13: نطبق متراجحة تشيبيشيف للمجالات التي منتصفها هو المتوسط.

ملاحظة 14: من الواضح رؤية أن نسبة المعطيات الواقعة خارج المجال $[\bar{x}-ks, \bar{x}+ks]$ هي على الأكثر $\dfrac{1}{k^2}$.

يبين الجدول التالي بعض القيم ل k والقيمة $1-\dfrac{1}{k^2}$ الموافقة لها:

k	$\sqrt{2}$	$\sqrt{3}$	**2**	**3**
$1-\dfrac{1}{k^2}$	$\dfrac{1}{2}$	$\dfrac{2}{3}$	$\dfrac{3}{4}$	$\dfrac{8}{9}$

مثال 42: لدينا عينة من المعطيات متوسطها 25 وانحرافها المعياري 4.5 ماهي نسبة المعطيات الواقعة ضمن المجال [16, 34]؟

الحل:

من الملاحظ أن مركز المجال [16, 34] هو 25 = 2/(16+34)، أي متوسط العينة المعطاة، بالتالي يمكننا تطبيق متراجحة تشيبيشيف.

علينا الآن إيجاد قيم k: $[\bar{x}-ks, \bar{x}+ks]=[16, 34]$، بالتالي $\bar{x}-ks=25-4.5k=16$ بحل هذه المعادلة نحصل على $k=2$. والقيمة $\left(1-\dfrac{1}{k^2}\right)=\left(1-\dfrac{1}{2^2}\right)=\left(1-\dfrac{1}{4}\right)=0.75$.

أي أن نسبة المعطيات الواقعة ضمن المجال [16, 34] لا تقل عن 75%.

مثال 43: لدينا عينة المثال السابق. أوجد مجال يقع فيها ما لا يقل عن 50% من المعطيات.

الحل:

$$\left(1-\frac{1}{k^2}\right)=0.5 \Rightarrow \frac{1}{k^2}=0.5 \Rightarrow k^2=2 \Rightarrow k=\sqrt{2}\simeq 1.41$$

وبالتالي فإن المجال الذي يقع فيها ما لا يقل عن 50% من المعطيات هو:

$$\left[\bar{x}-ks, \bar{x}+ks\right]=\left[25-\sqrt{2}\text{x}4.5, 25+\sqrt{2}\text{x}4.5\right]\approx[18.6, 31.4]$$

6.4.القيم المعيارية standard values:

نواجه في كثير من الأحيان أثناء دراستنا لكثير من الظواهر مقارنة مفردتين من مفرداتها وبحيث تصعب هذه المقارنة خاصة إذا كانت المفردتان تنتميان إلى ظاهرتين مختلفتين من حيث المتوسط والانحراف المعياري.

على سبيل المثال لمقارنة علامة طالب في إحدى المواد بعلامته في مادة أخرى، فمثلاً إذا حصل أحد الطلاب على علامة 85 من 100 في مادة الفيزياء وكانت علامته في الرياضيات 80 من 100، يتبادر إلى الذهن أن مستوى تحصيل الطالب في الفيزياء أفضل منه في الرياضيات. لهذا اتفق الإحصائيون على ان نقيس بعد كل منهما عن وسطها الحسابي بانحرافها المعياري وهذا ما يعرف بالقيمة المعيارية.

تعريف 10: ليكن لدينا عينة من المعطيات $x_1, x_2, \ldots, x_n$ حجمها n ومتوسطها $\bar{x}$ وانحرافها المعياري $s \neq 0$، نعرف القيمة المعيارية للقيمة x_i (المفردة الأصلية) كما يلي:

$$z_i=\frac{x_i-\bar{x}}{s}, \quad i=1, 2, \ldots, n$$

- تكون القيمة المعيارية موجبة إذا كانت المفردة فوق المتوسط ويكون سالبة إذا كانت تحته.
- المفردة الأصلية للقيمة المعيارية z_i هي: $x_i=\bar{x}+sz_i$.
- متوسط القيم المعيارية يساوي الصفر.
- الانحراف المعياري للقيم المعيارية يساوي الواحد.

مثال 44: ليكن لدينا عينة من المعطيات متوسطها $\bar{x}=9$ وانحرافها المعياري $s=6$، أوجد القيمة المعيارية للقيمة 12 ومن ثم أوجد القيمة الأصلية للقيمة المعيارية $z=0.1$.

الحل:

القيمة المعيارية للقيمة $x=12$ هي: $z=\dfrac{x-\bar{x}}{s}=\dfrac{12-9}{6}=0.5$

القيمة الأصلية للقيمة المعيارية $z=0.1$ هي: $x=\bar{x}+sz=9+6\text{x}0.1=9.6$

مثال 45: في امتحان لمقرر الإحصاء كان المتوسط لعلامات طلاب صف 70 وانحراف معياري 10 وفي امتحان لمقرر الرياضيات للصف نفسه كان المتوسط لعلامات الصف 60 والانحراف المعياري 6. فإذا كانت علامة أحد الطلاب في امتحان الإحصاء 80 وعلامته في الرياضيات 72 ففي أي الامتحانين كان مستوى تحصيله أفضل؟

الحل:

القيمة المعيارية للقيمة $x=80$ في مادة الإحصاء هي: $z=\dfrac{x-\bar{x}}{s}=\dfrac{80-70}{10}=1$

القيمة المعيارية للقيمة $x=72$ في مادة الرياضيات هي: $z=\dfrac{x-\bar{x}}{s}=\dfrac{72-60}{6}=2$

بما أن القيمة المعيارية لمقرر الرياضيات 2 أكبر من القيمة المعيارية لمقرر الإحصاء 1 فإن مستوى تحصيله في مقرر الرياضيات أفضل من مستوى تحصيله في مقرر الإحصاء بالرغم من أن علامته في مقرر الإحصاء أفضل من علامته في مقرر الرياضيات.

تمارين

1. لتكن المعطيات التالية:

15, 11, 12, 13, 10, 12, 14, 16, 12

أوجد ما يلي:

a) المتوسط
b) الوسيط
c) الربيع الأول والثالث
d) المنوال
e) التشتت
f) الانحراف المعياري

2. يبين الجدول التالي الفئات وتكراراتها

الفئات	[10,20[	[20,30[	[30,40[	[40,50[	[50,60[	[60,70[	[70,80[
التكرارات	12	10	11	10	15	11	7

أوجد ما يلي:

a) المتوسط
b) الوسيط
c) الربيع الأول والثالث
d) المنوال
e) التشتت
f) الانحراف المعياري

3. تبين القيم التالية أعمار، مقدرة بالساعات، للمبات اقتصادية 40 watt, 110 volt والتي أخذت من اختبارات قسرية. أوجد المخطط الصندوقي لتلك الأعمار

919	1196	785	1126	936	918	1156	920
948	1067	1092	1162	1170	929	950	905
972	1035	1045	855	1195	1195	1340	1122
938	970	1237	956	102	1157	978	832
1009	1157	1151	1009	765	958	902	1022
1333	811	1217	1085	896	958	1311	1037
702	923						

4. لتكن معطيات التمرين السابق. أوجد ما يلي:

a) الجدول التكراري لأعمار اللمبات وذلك باستخدام عدد مناسب من الفئات.

b) جدول التكرار التراكمي الصاعد لأعمار اللمبات.

c) المدرج التكراري لأعمار اللمبات.

d) المنحني التكراري التراكمي لأعمار اللمبات.

5. عينة من القراءات تم أخذها من مجتمع أعطت الجدول التالي:

x	20	25	28	30	32	35
f	5	25	20	25	15	10

عينة ثانية من نفس المجتمع أعطت ما يلي:

$$n_2 = 50, \quad \sum x = 1620, \quad \sum x^2 = 55928$$

أوجد المتوسط والتشتت العينة الكلية المكونة من دمج هاتين العينتين.

6. عينة من طلاب الجامعة الافتراضية حجمها 300، متوسط أوزانهم 58 kg وتشتت مقداره $16\ kg^2$، المطلوب:

a) المجال الذي يحوي على الأقل 200 وزناً من أوزان العينة.

b) المجال الذي يقع خارجه على الأكثر 50 وزناً من أوزان العينة.

c) العدد الأعظمي للطلاب الذين تقع أوزانهم خارج المجال [53, 63].

d) العدد الأصغري للطلاب الذين تقع أوزانهم ضمن المجال [48, 68].

العلامة العظمى: 100 **علامة النجاح: 50** **المدة: ساعة واحدة**

السؤال الأول: اختر الإجابة الصحيحة **(80) درجة**

1. إذا كان 35 قيمة الحد الأعلى لفئة طولها 10، بالتالي فإن الحد الأدنى للفئة هو:

a) 20

b) 25

c) 30

d) لا شيء مما سبق

2. متوسط العينة التالية 10, 20, 30, 40, 50 هو:

a) 15

b) 25

c) 35

d) 30

3. أياً من المقاييس التالية ليس مقياس للنزعة المركزية؟

a) المتوسط

b) الوسيط

c) المدى

d) المنوال

4. أياً من المقاييس للنزعة المركزية التالية لا يمكن إيجاده بيانياً؟

a) المتوسط

b) الوسيط

c) المنوال

d) لا شيء مما سبق

5. أياً من المقاييس للنزعة المركزية التالية يمكن أن يأخذ أكثر من قيمة من أجل عينة واحدة؟

a) المتوسط

b) الوسيط

c) المنوال

d) لا شيء مما سبق

6. ما يحدث لمتوسط عينة من القيم (n قيمة) إذا طرحنا من كل قيمة من قيم العينة المقدار 5؟

a) لا يتغير المتوسط

b) يزداد المتوسط بمقدار 5

c) ينقص المتوسط بمقدار 5

d) ينقص المتوسط بمقدار 5/n

7. الانحراف المعياري لعينة من 50 قيمة هو 8. إذا ضربنا كل قيمة من قيم العينة بالقيمة 2، تصبح قيمة الانحراف المعياري:

a) 4

b) 8

c) 16

d) لا شيء مما سبق

8. أياً من المقاييس التالية هو مقياس نسبي للتباين؟

a) الانحراف المعياري

b) التشتت

c) معامل الاختلاف

d) كل ما سبق

9. إذا كان متوسط عينة من القيم هو 25 ومعامل الاختلاف هو 0.1، عندئذ الانحراف المعياري هو:

a) 6.25

b) 2.5

c) 0.25

d) لا شيء مما سبق

10. وسيط العينة التالية 1 ,4 ,12 ,5 ,7 ,6 ,5 ,8 ,6 ,12 ,5 هو:

a) 6

b) 5

c) 7

d) 8

11. منوال العينة التالية 1 ,4 ,12 ,5 ,7 ,6 ,5 ,8 ,6 ,12 ,5 هو:

a) 6

b) 5

c) 7

d) لا شيء مما سبق

12. مدى العينة التالية 1 ,4 ,12 ,5 ,7 ,6 ,5 ,8 ,6 ,12 ,5 هو:

a) 12

b) 1

c) 11

d) 8

13. أياً من المقاييس التالية ليس مقياس للتباين؟

a) المدى

b) الربيعيات quartiles

c) التشتت

d) الانحراف المعياري

14. أياً من مقاييس التباين التالية التي تتأثر أكبر ما يمكن بالقيم الشاذة؟

a) التشتت

b) الانحراف المعياري

c) المدى الربيعي

d) المدى

ليكن لدينا جدول التكرار التالي:

الفترة	التكرار
[0, 9[	40
[10, 19[	50
[20, 29[	70
[30, 39[	40

15. طول الفترة هو:

a) 10

b) 9

c) 11

d) تختلف من فترة لأخرى

16. التكرار النسبي للفترة]10, 19[هو:

a) 0.9

b) 0.45

c) 0.25

d) لا يمكن إيجاده بالمعطيات المتوفرة

السؤال الثاني: **(20) درجة**

عينة من طلاب الجامعة الافتراضية حجمها 200، متوسط أوزانهم 60 kg وتشتت مقداره 25 kg2،
المطلوب:

a) المجال الذي يحوي على الأقل 150 وزناً من أوزان العينة.

b) المجال الذي يقع خارجه على الأكثر 25 وزناً من أوزان العينة.

c) العدد الأعظمي للطلاب الذين تقع أوزانهم خارج المجال [52, 68].

d) العدد الأصغري للطلاب الذين تقع أوزانهم ضمن المجال [48, 72].

الحل:

a) $\left(1-\frac{1}{k^2}\right)$ من نسبة المعطيات تقع ضمن المجال $[\bar{x}-ks, \bar{x}+ks]$.

$$1-\frac{1}{k^2}=\frac{150}{200}=\frac{3}{4}$$
$$\frac{1}{k^2}=1-\frac{3}{4}=\frac{1}{4}$$
$$k=2$$

بالتالي 150 وزناً من أوزان العينة تقع ضمن المجال: $[60-2(5), 60+2(10)]=[50, 70]$.

b) نعين k بحيث يكون $\frac{1}{k^2}=\frac{25}{200}=\frac{1}{8}$، بالتالي $\bar{x}=60=\frac{54+66}{2}$. أي أن 25 وزناً من أوزان العينة على الأكثر تقع خارج المجال

$[60-5\sqrt{8}, 60+5\sqrt{8}]=[45.86, 74.14]$.

c) المجال [52, 68]: مركزه $\bar{x}=60=\frac{52+68}{2}$ ونصف قطره $ks=8=\frac{68-52}{2}$، بالتالي $5k=8$ أو $k=8/5=1.6$.

وبالتالي العدد الأعظمي للطلاب تقع أوزانهم خارج الفترة [52, 68] هو:

$$n(\frac{1}{k^2}) = 200(\frac{1}{(1.6)^2}) = 78.125$$

78 طالباً.

d) المجال $[48,\ 72]$: مركزه $\bar{x} = 60 = \frac{48+72}{2}$ ونصف قطره $ks = 12 = \frac{72-48}{2}$، بالتالي $5k = 12$ أو $k = 12/5 = 2.4$.

وبالتالي العدد الأصغري للطلاب تقع أوزانهم داخل المجال $[48,\ 72]$ هو:

$$n(1-\frac{1}{k^2}) = 200(1-\frac{1}{(2.4)^2}) = 165.28$$

165 طالباً.

توجيه في حال الخطأ: راجع الفقرة 4.5

السؤال الأول	الإجابة الصحيحة	توجيه في حال الخطأ
1	b	الفقرة 2
2	d	الفقرة 1.3
3	c	الفقرة 3
4	a	الفقرة 1.3 و2
5	d	الفقرة 3
6	c	الفقرة 1.3
7	c	الفقرة 4.3
8	c	الفقرة 4
9	b	الفقرة 4
10	a	الفقرة 2.3
11	b	الفقرة 3.3
12	c	الفقرة 1.4
13	b	الفقرة 4
14	d	الفقرة 4
15	a	الفقرة 2
16	c	الفقرة 2