تمهيد

عرفنا في الفصل السابق، أن مقاييس النزعة المركزية تسمح لنا بالحصول، على القيم المتوسطة للبيانات أو على مكان تجمعها، غير أن هذه المقاييس لا تكفي لوحدها لمعرفة الصفات الإحصائية اللازمة لوصف الظواهر، لأن الفروق بين قيم الظواهر، قد تزداد أو تنقص رغم تساوي المتوسطات لهذه الظواهر. لذلك ترفق مقاييس النزعة المركزية دائما بمقياس يدل على تشتت القيم وانتشارها في التوزيع، تسمى بمقاييس التشتت.

مثال:

لتكن لدينا مجموعتين من الطلاب، علامات الطلبة في المجموعتين كالتالي:

المجموعة الأولى: 6، 9، 18، 15، 12.

المجموعة الثانية: 10، 11، 12، 13، 14.

لو قمنا بحساب المتوسط الحسابي لكل مجموعة منهما، نجد أن متوسط العلامات في كل مجموعة يساوي 12، ومع ذلك هناك فرق كبير بين علامات المجموعتين، هذا الفرق يتمثل في تجانس علامات المجموعة الثانية وتشتت علامات المجموعة الأولى.

من أجل ذلك لجأ الإحصائيون إلى استخدام مقاييس أخرى، لقياس مدى تجانس البيانات أو مدى انتشارها حول مقاييس النزعة المركزية.

مفهوم التشتت:

يمكننا أن نعرف التشتت على أنه: مدى تباعد أو تناثر قيم الظاهرة (البيانات) حول أحد المتوسطات. فإذا كانت البيانات متجانسة ومتشابهة وغير متباعدة عن بعضها، يقال أنها غير مشتتة، أي مركزة حول بعضها وبالتالي حول وسطها الحسابي. أما إذا كانت البيانات متباعدة ومتباينة عن بعضها وغير متجانسة، فيقال أنها بيانات متشتتة، وغير مركزة.

ويمكننا أن نصنف مقاييس التشتت إلى مجموعتين:

- مقاييس التشتت المطلقة: وهي التي تقيس التشتت مقدرا بوحدات قيم الظاهرة نفسها.
- مقاييس التشتت النسبي: وهي التي تقيس التشتت على شكل نسب مئوية وفيما يلي مقاييس المجموعتين:

أولا– مقاييس التشتت المطلقة: Les simples mesures de dispersion

1– المدى: L'Etendue ou Range (R)

من أبسط مقاييس التشتت المطلقة، يحسب كما يلي:

في البيانات غير المبوبة: يحسب بتطبيق المعادلة التالية:

المدى لمجموعة من البيانات هو الفرق بين أكبر وأصغر قيمة في البيانات.

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

مثال (4-1):

أوجد المدى للبيانات التالية التي تمثل درجات الحرارة المسجلة في يوم ما:

24، 20، 17، 18، 19، 4-

الحل:

$$R = X_{max} - X_{min}$$
$$R = 24 - (-4) = 28$$

في البيانات المبوبة: له أكثر من صيغة منها:

المدى هو الفرق بين مركز الفئة الأخيرة ومركز الفئة الأولى.   أو:

المدى هو الفرق بين الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة الأولى.

مثال (4-2):

الجدول التالي يبين توزيع 60 مزرعة حسب المساحة المزروعة.

المساحة	20-15	25-20	30-25	35-30	40-35	45-40
عدد المزارع	3	9	15	18	12	3

المطلوب: حساب المدى للمساحة المزروعة؟

الحل:

المدى = مركز الفئة الأخيرة – مركز الفئة الأولى

= 42,5 – 17,5 = 25.

مدى المساحة المزروعة هو 25.

المدى = الحد الأعلى للفئة الأخيرة – الحد الأدنى للفئة الأولى

= 45 – 15 = 30.

مدى المساحة المزروعة هو 30.

- **مزايا وعيوب المدى:**

من مزاياه أنه سهل الحساب ويكثر استخدامه عند الإعلان عن حالات الطقس، ومن عيوبه أنه يعتمد على قيمتين فقط ويتأثر بالقيم المتطرفة والشاذة.

ملاحظة: إذا أردنا أن نقلل من أثر القيم المتطرفة، فإننا نقوم باستبعادها وذلك باستخدام احدى الطرق التالية في حساب المدى:

المدى الربيعي = الربيع الثالث – الربيع الأول

المدى العشيري = العشير التاسع – العشير الأول

المدى المؤيني = المؤين 99 – المؤين الأول

2- الانحراف الربيعي (نصف المدى الربيعي): Quartile Déviation (Q)

عبارة عن نصف الفرق بين الربيعين الثالث والأول، ويحسب وفق العلاقة التالية:

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

حيث: Q_1: الربيع الأول.

Q_3: الربيع الثالث.

وقد بينا في الفصل الثالث كيفية حساب هذين الربيعان سواء في البيانات المبوبة أو غير المبوبة.

- **مزايا وعيوب الإنحراف الربيعي:**

من مزاياه أنه يبتعد عن القيم المتطرفة فهو أحسن من المدى، ولكن بالرغم من اقترابه من مركز البيانات إلا أنه يستعمل قيمتين فقط ويهمل باقي البيانات فهو لا يعكس حقيقة التشتت.

3- الانحراف المتوسط: L'Ecart moyen

لأن مقاييس التشتت هي مقاييس لقوة تجمع البيانات حول بعضها، وحيث أن التجمع يكون حول القيم المتوسطة، فإنه إذا كان مقدار الانحراف كبير دل ذلك على أن التشتت كبير والعكس صحيح. وحيث أن مجموع الانحرافات عن المتوسط يساوي صفرا، فإنه لو حسبنا القيم المطلقة لمقدار الاختلاف عن المتوسط يكون متوسط الانحرافات مقياس مناسبا للتشتت، يسمى بالانحراف المتوسط.

يحسب الانحراف المتوسط حول الوسط الحسابي كما يلي:

من بيانات غير مبوبة:

$$EM_{\bar{X}} = \frac{|x_1 - \bar{X}| + |x_2 - \bar{X}| + \cdots + |x_n - \bar{X}|}{N}$$

$$= \frac{\sum|x_i - \bar{X}|}{N}$$

مثال (4-3):

إذا كانت الطاقة التصديرية لخمس محطات لتحلية المياه بالمليون متر مكعب كما يلي:

7، 10، 2، 5، 4.

- أوجد قيمة الإنحراف المتوسط للطاقة التصديرية؟

الحل:

لحساب قيمة الإنحراف المتوسط، نقوم بحساب الوسط الحسابي، وتكوين الجدول التالي:

$$\bar{X} = \frac{\sum x_i}{N} = \frac{28}{5} = 5{,}6$$

$\lvert x_i - \bar{X}\rvert$	$(x_i - \bar{X})$	x_i
1,6	-1,6	4
0,6	-0,6	5
3,6	-3,6	2
4,4	4,4	10
1,4	1,4	7
11,6	**0**	**المجموع**

$$EM_{\bar{X}} = \frac{\sum|x_i - \bar{X}|}{N} = \frac{11{,}6}{5} = 2{,}32$$

الإنحراف المتوسط للطاقة التصديرية حول الوسط الحسابي هو: 2,32 مليون متر مكعب.

أما إذا كانت البيانات مبوبة:

$$EM_{\bar{X}} = \frac{\sum n_i\,|x_i - \bar{X}|}{N}, \qquad N = \sum n_i$$

مثال (4-4):

الجدول التالي يبين توزيع الأجر الشهري لمجموعة من العمال في شركة ما.

المجموع	]100-90]	]90-80]	]80-70]	]70-60]	]60-50]	الأجر 10^3 دج
100	5	10	20	40	25	n_i

- أوجد الانحراف المتوسط لأجور العمال؟

الحل:

لحساب الإنحراف المتوسط نقوم أولا بحساب الوسط الحسابي، وتكوين الجدول التالي:

الأجر 10^3 دج	n_i	x_i	$n_i x_i$	$\lvert x_i - \bar{X} \rvert$	$n_i \lvert x_i - \bar{X} \rvert$
]60 – 50]	25	55	1375	13	325
]70 – 60]	40	65	2600	3	120
]80 – 70]	20	75	1500	7	140
]90 – 80]	10	85	850	17	170
]100 – 90]	5	95	475	27	135
المجموع	**100**	/	**6800**	/	**890**

$$\bar{X} = \frac{\sum x_i n_i}{N} = \frac{6800}{100} = 68\,.$$

متوسط الأجر الشهري لهؤلاء العمال هو: 68 ألف دج.

$$EM_{\bar{X}} = \frac{\sum n_i \lvert x_i - \bar{X} \rvert}{N} = \frac{890}{100} = 8{,}9$$

الانحراف المتوسط لأجور العمال حول الوسط الحسابي هو 8,9 ألف دج.

وقد يحسب الانحراف المتوسط حول الوسيط، في هذه الحالة تكون صيغته كالتالي:

في حالة البيانات غير المبوبة:

$$EM_{Me} = \frac{\sum \lvert x_i - Me \rvert}{N}$$

في حالة البيانات المبوبة:

$$EM_{Me} = \frac{\sum n_i \lvert x_i - Me \rvert}{N}, \quad N = \sum n_i$$

- **خواص الانحراف المتوسط:**

- يعتمد في حسابه على جميع القيم.
- لا يمكن حسابه في حالة الجداول المفتوحة.

- أقل تأثر بالقيم المتطرفة من المدى.
- لا يستعمل بشكل واسع بسبب اعتماده على القيمة المطلقة (أي إهمال الإشارات الجبرية للانحرافات).

4- الانحراف المعياري والتباين: L'Ecart-Type et Variance

إن الفكرة الأساسية لهذين المقياسين هي أنه بدلا من إهمال الإشارات الجبرية مثل عند حساب الانحراف المتوسط، نحاول التخلص من هذه الإشارات بطريقة أخرى أكثر صلاحية، وذلك بتربيع الانحرافات.

- **الانحراف المعياري:**

يعتبر الانحراف المعياري من أهم مقاييس التشتت لأنه يحتوي على مفهوم جبري للانحرافات، وهو من الأساليب الإحصائية الرياضية الحديثة لقياس التشتت، ومن أكثرها استعمالا.

يعرف الانحراف المعياري على أنه مقياس للتشتت، يقيس لنا مدى تباعد أو اختلاف أو انحراف قيم الظاهرة المدروسة عن وسطها الحسابي، كما يمكن تعريف الانحراف المعياري رياضيا، بأنه الجذر التربيعي لمتوسط مربعات انحرافات القيم عن الوسط الحسابي.

يتم حساب الانحراف المعياري لبيانات غير مبوبة كما يلي:

$$\sigma_x = \sqrt{\frac{\sum(\mathrm{x_i} - \bar{\mathrm{X}})^2}{\mathrm{N}}}$$

كما يمكن حسابه بالصيغة المختصرة التالية:

$$\sigma_x = \sqrt{\frac{\sum \mathrm{x_i}^2}{\mathrm{N}} - (\bar{\mathrm{X}})^2}$$

مثال (4-5):

بالرجوع إلى المثال السابق (4-3) الطاقة التصديرية لخمس محطات لتحلية المياه بالمليون متر مكعب هي: 7، 10، 2، 5، 4. أوجد قيمة الانحراف المعياري للطاقة التصديرية؟

الحل:

لحساب قيمة الانحراف المعياري، نقوم بحساب الوسط الحسابي، وتكوين الجدول التالي:

$$\bar{X} = \frac{\sum \mathrm{x}_i}{\mathrm{N}} = \frac{28}{5} = 5{,}6$$

x_i^2	$(x_i - \bar{X})^2$	$(x_i - \bar{X})$	x_i
16	2,56	-1,6	4
25	0,36	-0,6	5
4	12,96	-3,6	2
100	19,36	4,4	10
49	1,96	1,4	7
194	**37,2**	**0**	المجموع

ط1):

$$\sigma_x = \sqrt{\frac{\sum(x_i - \bar{X})^2}{N}} = \sqrt{\frac{37,2}{5}} = \sqrt{7,44} = 2,73$$

أو:

ط2):

$$\sigma_x = \sqrt{\frac{\sum x_i^2}{N} - (\bar{X})^2} = \sqrt{\frac{194}{5} - (5,6)^2} = \sqrt{38,8 - (31,36)^2}$$

$$\sigma_x = \sqrt{7,44} = 2,73$$

الانحراف المعياري للطاقة التصديرية حول الوسط الحسابي هو: 2,73 مليون متر مكعب.

أما **الانحراف المعياري لبيانات مبوبة** فيحسب كما يلي:

$$\sigma_x = \sqrt{\frac{\sum n_i (x_i - \bar{X})^2}{N}}$$

كما يمكن حسابه بالصيغة المختصرة التالية:

$$\sigma_x = \sqrt{\frac{\sum n_i x_i^2}{\sum n_i} - (\bar{X})^2}$$

مثال (6-4):

بالرجوع إلى المثال (4-4) الذي يبين توزيع الأجر الشهري لمجموعة من العمال في شركة ما. أوجد الانحراف المعياري لأجور العمال؟

الحل:

لحساب الانحراف المعياري نقوم أولا بحساب الوسط الحسابي، وتكوين الجدول التالي:

الأجر 10^3 دج	n_i	x_i	$n_i x_i$	$n_i (x_i - \bar{X})^2$	$n_i x_i^2$
]60 – 50]	25	55	1375	4225	75625
]70 – 60]	40	65	2600	360	169000
]80 – 70]	20	75	1500	980	112500
]90 – 80]	10	85	850	2890	72250
]100 – 90]	5	95	475	3645	45125
المجموع	**100**	/	**6800**	**12100**	**474500**

$$\bar{X} = \frac{\sum x_i n_i}{N} = \frac{6800}{100} = 68\,.$$

متوسط الأجر الشهري لهؤلاء العمال هو: 68 ألف دج.

ط1):

$$\sigma_x = \sqrt{\frac{\sum n_i (x_i - \bar{X})^2}{N}} = \sqrt{\frac{12100}{100}} = \sqrt{121} = 11$$

أو:

ط2):

$$\sigma_x = \sqrt{\frac{\sum n_i x_i^2}{\sum n_i} - (\bar{X})^2} = \sqrt{\frac{474500}{100} - (68)^2} = \sqrt{121} = 11$$

الانحراف المعياري لأجور العمال حول الوسط الحسابي هو 11 ألف دج.

- **خصائص الانحراف المعياري:**

1. الانحراف المعياري للمقدار الثابت يساوي صفرا، أي أنه إذا كانت لدينا: $x: a, a, \dots, a$ حيث a مقدار ثابت فإن:

$$\sigma(a) = 0$$

2. إذا كان الانحراف المعياري للقيم $x_1, x_2, \dots, x_n$ هو $\sigma(x)$، فإنه إذا أضيفت أو طرحت قيمة ثابتة (a) إلى أو من جميع القيم فإن الانحراف المعياري لا يتغير.

$$\sigma(x + a) = \sigma(x)$$
$$\sigma(x - a) = \sigma(x)$$

3. إذا كان الانحراف المعياري للقيم $x_1, x_2, \dots, x_n$ هو $\sigma(x)$ فإنه إذا ضربت كل قيمة بالمقدار a فإن الانحراف المعياري يتأثر بالمقدار نفسه:

$$\sigma(ax) = a\,\sigma(x)$$
$$\sigma(ax + c) = a\,\sigma(x)$$

4. لا يمكن ايجاده في حالة التوزيعات التكرارية المفتوحة.

5. بالإضافة إلى الخصائص السابقة فهو يتمتع بخاصية هامة وهي أنه في حالة التوزيع الطبيعي فإن المدى بين الوسط الحسابي والانحراف المعياري يحصر نسب معينة من قيم التوزيع على الشكل التالي:

%15 من البيانات في المجال $\bar{x} \pm \frac{2}{3}\sigma_x$

%68,27 من البيانات في المجال $\bar{x} \pm \sigma_x$

%95,45 من البيانات في المجال $\bar{x} \pm 2\sigma_x$

%99,73 من البيانات في المجال $\bar{x} \pm 3\sigma_x$

- **التباين:**

هو العزم المركزي الثاني، أو بعبارة أخرى هو الوسط الحسابي لمجموع مربع انحرافات القيم عن وسطها الحسابي (أي مربع الانحراف المعياري). ويدل على مدى تشتت قيم المتغير الإحصائي حول أحد مقاييس النزعة المركزية (الوسط الحسابي).

التباين لبيانات غير مبوبة:

يمكن حساب التباين بالطريقة المباشرة:

$$V(x) = \sigma^2 = \frac{\sum(x_i - \bar{X})^2}{N}$$

أو بالطريقة المختصرة، حيث انطلاقا من العلاقة السابقة نجد:

$$V(x) = \frac{\sum(x_i - \bar{X})^2}{N}$$

$$V(x) = \frac{\sum(x_i^2 - 2\,x_i\bar{X} + (\bar{X})^2)}{N}$$

$$= \frac{\sum x_i^2}{N} - 2\,\bar{X}\,\bar{X} + \frac{N(\bar{X})^2}{N}$$

$$V(x) = \frac{\sum x_i^2}{N} - (\bar{X})^2$$

$$V(x) = \overline{X^2} - (\bar{X})^2$$

مثال (4–7): بالرجوع إلى المثال السابق (4–3) الذي يبين الطاقة التصديرية لخمس محطات تحلية المياه نجد:

$$V(x) = \sigma^2 = 7{,}44$$

التباين للطاقة التصديرية حول الوسط الحسابي هو: 7,44 مليون متر مكعب.

التباين لبيانات مبوبة:

يمكن حساب التباين بالطريقة المباشرة:

$$V(x) = \frac{\sum n_i\,(x_i - \bar{X})^2}{N}$$

أما الصيغة المختصرة للتباين فهي:

$$V(x) = \frac{\sum n_i\,x_i^2}{N} - (\bar{X})^2$$

مثال (4–8): بالرجوع إلى المثال السابق (4–4) الذي يبين توزيع الأجر الشهري ل 100 عامل نجد:

$$V(x) = \sigma^2 = 121$$

التباين لأجور العمال حول الوسط الحسابي هو 121 ألف دج.

- **خصائص التباين:**

انطلاقا من الخصائص الموضحة سابقا للانحراف المعياري، يمكننا أن نستنتج خصائص التباين التالية:

$$V(a) = 0$$

$$V(x + a) = V(x)$$

$$V(ax) = a^2\,V(x)$$

$$V(ax + c) = a^2\,V(x)$$

ثانيا– مقاييس التشتت النسبية:

إن المقاييس السابقة لا تسمح لنا بمقارنة درجة التشتت بين صفتين تنتميان إلى نفس المجموعة، ويعبر عنهما بنفس الوحدات القياسية، طالما أن المتوسط يختلف في كل منهما. كما لا تسمح لنا هذه المقاييس بمقارنة درجة التشتت، بين ظاهرتين مختلفتين لا تشتركان بنفس الوحدات القياسية، فلو أردنا المقارنة بين تشتت ظاهرتين أو أكثر من هذا النوع، فإننا نلجأ إلى مقاييس جديدة تسمى: مقاييس التشتت النسبي والتي نحصل عليها بقسمة التشتت المطلق على المتوسط الذي حسب حوله مضروبا في 100.

وبشكل عام عندما يكون لدينا ظاهرتين أو أكثر، ونريد أن نقارن بين تشتيتهما، تواجهنا ثلاث حالات هي:

1. لدينا نفس وحدة القياس ونفس الوسط الحسابي: في هذه الحالة نقارن بين مقياسي التشتت للظاهرتين.
2. لدينا المجموعتان مقدرتان بدلالة وحدات قياسية واحدة، ولكن المتوسط يختلف في كل منهما: في هذه الحالة لابد من حساب تشتت جديد لكل مجموعة وهو التشتت النسبي، مما يسمح لنا بالحصول على نسب مئوية قابلة للمقارنة فيما بينها حيث النسبة الأكبر تدل على تشتت أكبر.
3. أن تكون المجموعتان مقدرتان بوحدات قياسية مختلفة: وفي هذه الحالة أيضا لابد من استخدام مقاييس التشتت النسبي، وهي:

1- المدى النسبي:

يمكن حسابه بالصيغة التالية:

$$R_p = \frac{R}{\overline{X}} * 100$$

2- الانحراف الربيعي النسبي:

ويعرف أيضا بمعامل الإختلاف الربيعي، فإذا رمزنا له بالرمز: C.V.Q

$$C.V.Q = \frac{Q}{Me} \times 100$$
$$= \frac{Q_3 - Q_1}{2\,Me} \times 100$$

3- الانحراف المتوسط النسبي:

وهو يحسب حول كل من الوسط الحسابي والوسيط.

* حول الوسط الحسابي:

$$EM_{\bar{X}p} = \frac{EM_{\bar{X}}}{\overline{X}} * 100$$

* حول الوسيط:

$$EM_{Me.p} = \frac{EM_{Me}}{Me} * 100$$

4- الانحراف المعياري النسبي (معامل الاختلاف):

ويعرف بمعامل الاختلاف ويرمز له بالرمز: CV وهو يستعمل بشكل واسع في التحليل الإحصائي ويعرف معامل الاختلاف بأنه حاصل قسمة الانحراف المعياري على الوسط الحسابي مضروبا في 100.

وصيغته كالتالي:

$$CV = \frac{\sigma}{\bar{X}} * 100$$

مثال (4-9):

أعطت العينتان (أ) و(ب) البيانات التالية:

- **العينة الأولى:** $\delta = 8$ ، $\bar{X} = 160$
- **العينة الثانية:** $\sum_{i=1}^{20} X_i = 1200$ ، $\sum_{i=1}^{20} X_i^2 = 72680$

المطلوب: أوجد أي العينتين أكثر تشتتا؟

الحل:

العينة الأولى:

$$\delta = 8 \qquad \bar{X} = 160$$

العينة الثانية:

$$\sum_{i=1}^{20} X_i = 1200 \text{ kg} \quad \Rightarrow \quad \bar{X} = \frac{\sum X_i}{N} = \frac{1200}{20} = 60 \text{ kg}$$

$$\sum_{i=1}^{20} {X_i}^2 = 72\,680 \text{ kg} \quad \Rightarrow \quad \delta = \sqrt{\overline{X^2} - (\bar{X})^2} = \sqrt{\frac{72680}{20} - (60)^2}$$

$$\delta = \sqrt{34} = 5.83$$

من أجل توضيح أي العينتين أكثر تشتتا، نستخدم معامل الاختلاف:

بالنسبة للعينة الأولى:

$$CV = \frac{\delta}{\bar{X}}.100 = \frac{8}{160}.100 = 5\%$$

بالنسبة للعينة الثانية:

$$CV = \frac{\delta}{\bar{X}}.100 = \frac{5.83}{60}.100 = 9.72\%$$

$$CV_{(\text{أ})} < CV_{(\text{ب})}$$

بما أن معامل اختلاف العينة الأولى أقل من معامل اختلاف العينة الثانية فإن العينة الأولى هي الأكثر تجانسا (أي أقل تشتتا).