تمهيد: عرفنا في الفصل السابق أن مقاييس النزعة المركزية (المتوسط الحسابي والوسيط والمنوال) تسمح لنا بالحصول على القيم المتوسطة للبيانات أو على 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، ولتوضيح ما سبق نفترض أن طالبين تحصلا على النتائج التالية في خمس مواد دراسية:

الطالب (X): 10،11،13،14،15.

الطالب (Y):8،9،13،15،18.

فمتوسط درجات الطالب (X) يساوي 12،6 وكذلك متوسط درجات الطالب (Y) يساوي 12،6 ووسيط درجات الطالب (X) يساوي 13 وكذلك وسيط درجات الطالب (Y) يساوي 13.

قد يفهم مما سبق أن الطالبين (X)و (Y) لهما نفس المستوى غير أن التمعن الجيد في الدرجات التي تحصل عليها الطالبين تبين أن الطالب (X) ناجح في كل المواد المدروسة في حين أن الطالب (Y) ناجح في ثلاث مواد فقط. إن هذه الحقيقية تبين أن مقاييس النزعة المركزية لا تعطي فكرة وافية عن اختلاف قيم الظواهر، ولا تحقق كل الأغراض التي نرغب الوصول إليها من دراستنا لذلك فإن مقاييس النزعة المركزية لا بد أن تكون مصحوبة بمقاييس أخرى لقياس مدى تباعد أو تقارب البيانات من بعضها البعض أو من متوسطها، تسمى هذه المقاييس بمقاييس التشتت.

ما معنى التشتت؟.

تشتت بيانات ظاهرة ما يقصد به درجة أو مقدار التفاوت أو الاختلاف بين مفردات هذه الظاهرة، وتعتبر بيانات الظاهرة متجانسة عندما تكون قيمتها قريبة من بعضها البعض ونقول في هذه الحالة أنها غير مشتتة. أما إذا كانت بيانات الظاهرة متباعدة وغير متجانسة فنقول أن مفردات الظاهرة مشتتة وغير مركزة. ويقاس تشتت البيانات بعدة مقاييس منها:

أولا – المدى (المطلق)

المدى لمجموعة من البيانات هو الفرق بين أكبر قيمة وأصغر قيمة لها ويرمز له بالرمز R.

المدى= أكبر قيمة – أصغر قيمة.

أما المدى للتوزيعات التكرارية فيحسب بعدة طرق منها:

المدى = مركز الفئة الأخيرة – مركز الفئة الأولى

المدى = الحد الأعلى لفئة الأخيرة – الحد الأدنى للفئة الأولى

مثال1:

أوجد المدى للبيانات التالية: 12،18،22،28،30.

الحل: المدى = أكبر قيمة – أصغر قيمة

= 30-12=22.

مثال2:

أوجد المدى للبيانات التالية 65،20،17، -04،18،19،4.

الحل:

المدى = 65 –(-4) = 69.

نلاحظ المدى في هذا المثال قد تأثر بشكل كبير جدا بالقيم المتطرفة، إذ نلاحظ أن معظم البيانات متقاربة باستثناء القيمة 65 والقيمة (-4)، فإذا استبعدنا هذه القيم المتطرفة فإن المدى يصبح

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

R = 20-14=6.

وبسبب هذا العيب فإن المدى كمقياس للتشتت لا يستخدم إلا عندما نرغب في مقياس تقريبي وسريع لتشتت البيانات دون الاهتمام بالدقة في المقياس، أو عندما يكون للبيانات المتطرفة أهمية خاصة كتوزيعات درجات الحرارة على سبيل المثال، حيث تعلن درجات الحرارة اليومية بحدها الأقصى وحدها الأدنى خلال اليوم، كما يشيع استخدام هذا المقياس في حالات مراقبة جودة الإنتاج أو متابعة المبيعات التي يحققها رجال البيع لمؤسسة ما.

أما إذا أردنا أن نقلل من أثر القيم المتطرفة فإننا نقوم باستبعادها ويمكن أن يتم ذلك باستخدام الطرق التالية:

* المدى الربيعي = الربيع الثالث – الربيع الأول.

* المدى العشري = العشر التاسع – العشر الأول.

* المدى المئيني = المئين 99 – المئين الأول.

خواص المدى:

1 – يتصف المدى بسهولة حسابه.

2 – يعتمد في حساب على قيمتين فقط هما القيمة الكبرى والقيمة الصغرى.

3- بسبب الخاصية الثانية فإن المدى شديد التأثر بالقيم المتطرفة.

ثانيا: الانحراف المتوسط L'écart moyen

لأن مقاييس التشتت هي مقاييس لقوة تجمع البيانات حول بعضها، وحيث أن التجمع يكون حول القيم المتوسطة، فإنه إذا كان مقدار الاختلاف بين القيم ومتوسطها كبيرا 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 هذه الاختلافات مقياسا مناسبا لمقدار التشتت، يسمى هذا المقياس بالانحراف المتوسط.

ويعرف الانحراف المتوسط بأنه المتوسط الحسابي للقيم المطلقة لانحرافات القيم عن متوسطها الحسابي، وسوف نرمز للانحراف المتوسط في دراستنا بالرمز E_x وعليه:

إذا كانت لدينا القيم التالية: $X_1 , X_2 , X_3 , X_4 \ldots . X_n$ فإن الانحراف المتوسط لها هو:

$$E_X = \frac{\left|X_1 - \overline{X}\right| + \left|X_2 - \overline{X}\right| + \left|X_3 - \overline{X}\right| + \ldots . + \left|X_n - \overline{X}\right|}{n}$$

أو

$$E_X = \frac{\sum_{i=1}^{n} \left|X_i - \overline{X}\right|}{n}$$

أما إذا كانت البيانات مكررة أو مبوبة في جداول توزيع تكرار فإن الانحراف المتوسط لها يعطي العلاقة:

$$E_X = \frac{\sum_{i=1}^{n} ni\left|X_i - \overline{X}\right|}{\sum_{i=1}^{n} ni}$$

مثال3:

أوجد الانحراف المتوسط للبيانات التالية: 2،4،5،6،8؟

(*) أنظر خواص المتوسط الحسابي.

الحل:

$\left\| X_I - \overline{X} \right\|$	X_i
3	2
1	4
0	5
1	6
3	8
8	**—**

$$\overline{X} = \frac{\sum n_i}{n} = \frac{25}{5} = 5$$

$$E_x = \frac{\sum \left| X_i - \overline{X} \right|}{n} = \frac{8}{5} = 1.6$$

مثال4:

أوجد تشتت البيانات المبوبة في جدول التوزيع التكراري الآتي باستخدام الانحراف المتوسط؟

المجموع	8-6	6-4	4-2	2-0	الفئة
12	3	4	3	2	التكرار

الحل:

$n_i\left\| X_i - \overline{X} \right\|$	$\left\| X_I - \overline{X} \right\|$	n_iX_i	X_i	التكرار	الفئة
6.66	3.33	2	1	2	0-2
4	1.33	9	3	3	4-2
2.68	0.67	20	5	4	6-4
8.01	2.67	21	7	3	8-6
21.35		52		12	المجموع

$$\overline{X} = \frac{\sum n_i X_i}{\sum n_i} = \frac{25}{12} = 4.33$$

$$E_x = \frac{\sum n_i \left| X_i - \overline{X} \right|}{\sum n_i} = \frac{21.35}{12} = 1.78$$

ويعتبر الانحراف المتوسط أفضل من (المدى) لأنه أقل تأثر بالقيم المتطرفة غير أنه لا يستعمل بشكل واسع بسبب اعتماده على القيمة المطلقة لانحرافات القيم عن متوسطها الحسابي.

خواص الانحراف المتوسط:

1- يعتمد في حسابه على جميع القيم وليس على القيمة الكبرى والصغرى فقط.

2- لا يمكن حسابه في حالة التوزيعات التكرارية المفتوحة.

3- يتأثر بالقيم المتطرفة، لأن انحرافها عن المتوسط الحسابي يكون كبيرا.

ثالثا – التباين والانحراف المعياري La variance et l'écart type

1 – التباين La variance:

وهو عبارة عن المتوسط الحسابي لمربعات الفروق بين قيم المتغير الإحصائي ومتوسطها الحسابي، ونستخدم مربعات الفروق هنا تفاديا لاستخدام القيم المطلقة كما هو الشأن في الانحراف المتوسط.

فإذا كانت لدينا البيانات التالية: $X_1, X_2, X_3, X_4 \ldots\ldots X_n$

فإن التباين لهذه البيانات يعطي بالعلاقة

$$V_X = \frac{(X_1-\overline{X})^2+(X_2-\overline{X})^2+(X_3-\overline{X})^2+\ldots\ldots+(X_n-\overline{X})^2}{n}$$

$$V_X = \frac{\sum_{i=1}^{n} ni(X_1-\overline{X})^2}{\sum ni}$$

مثال5:

أوجد التباين للبيانات التالية: 3،7،6،1،11،5،6،9.

الحل:

$(X_i-\overline{X})^2$	$(X_i-\overline{X})$	Xi
9	-3	3
1	1	7
0	0	6
25	-5	1
25	5	11
1	-1	5
0	0	6
9	3	9
70	**——**	**48**

حساب المتوسط الحسابي:

$$\overline{X} = \frac{\sum Xi}{n} = \frac{48}{8} = 6$$

حساب التباين:

$$V_x = \frac{\sum (X_i - \overline{X})^2}{n} = \frac{70}{8} = 8.75$$

- في بعض الأحيان عندما يكون المتوسط الحسابي للبيانات عبارة عن كسر، فإن عملية حساب التباين تكون عرضة للأخطاء الحسابية لذلك فإنه تم تطوير طريقة مختصرة لحساب التباين.

طريقة مختصرة لحساب التباين:

انطلاقا من العلاقة المتوصل إليها سابقا:

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

يمكن كتابة

$$V_x = \frac{\sum (X_i{}^2 - 2\overline{X}_i + \overline{X}^2)}{n}$$

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

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

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

ومنه

$$\boxed{V_x = \frac{\sum X_i^2}{\sum ni} - \overline{X}^2}$$

أما في حالة البيانات المبوبة فإن العلاقة تصبح

$$\boxed{V_x = \frac{\sum n_i X_i^2}{\sum hi} - \overline{X}^2}$$

2 – الانحراف المعياري:

يعتبر الانحراف المعياري من أهم المقاييس الإحصائية للتشتت، وهو أكثر استخداما في النظريات والقوانين الإحصائية، لأنه يعطي فكرة سليمة ومنطقية عن ظاهرة التشتت، ويعرف الانحراف المعياري بأنه الجذر التربيعي لمتوسط مجموع مربع انحراف القيم عن متوسطها، أي أنه الجذر التربيعي للتباين.

سوف نرمز للانحراف المعياري في دراستنا بالرمز (S_x).

الانحراف المعياري لبيانات مفردة

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

الانحراف المعياري لبيانات متكررة أو مبوبة

$$Sx = \sqrt{\frac{\sum hi(X_i - \overline{X})^2}{\sum hi}}$$

أما الصيغة المختصرة للانحراف المعياري فتعطى بالعلاقات التالية:

الانحراف المعياري لبيانات مفردة

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

الانحراف المعياري لبيانات متكررة أو مبوبة

$$Sx = \sqrt{\frac{\sum hiX_i^{\,2}}{\sum hi} - \overline{X}^2}$$

مثال6:

أوجد التباين والانحراف المعياري بالصيغة الأصلية ثم بالصيغة المختصرة للبيانات المبوبة في الجدول الإحصائي التالي:

الفئة	8-4	13-9	18-14	23-19	28-24	المجموع
التكرار	3	4	6	2	4	19

الحل:

الفئة	التكرار	X_i	n_{iXi}
8-4	3	6	18
13-9	4	11	44
18-14	6	16	96
23-19	2	21	42
28-24	4	26	104
المجموع	**19**	—	**304**

$(X_i - \overline{X})$	$N_i(X_i - \overline{X})$	$N_i(X_i - \overline{X})_2$
-10	-30	300
-5	-20	100
0	0	0
5	10	50
10	40	400
—	—	**850**

X_i^2	n_{iXi^2}
36	108
121	484
256	1536
441	882
676	2704
—	**5714**

$$\overline{X} = \frac{\sum h_i X_i}{\sum h_i} = \frac{304}{19} = 16$$

التباين بالصيغة الأصلية

$$\overline{X} = \frac{\sum n_i (X_i - \overline{X})^2}{\sum n_i} = \frac{850}{19} = 44.74$$

الانحراف المعياري بالصيغة الأصلية

$$S_x = \sqrt{\frac{\sum n_i (X_i - \overline{X})^2}{\sum X_i}} = \sqrt{44.73} = 6.69$$

التباين بالصيغة المختصرة

$$V_x = \frac{\sum n_i X_i^2}{\sum n_i} - \overline{X}^2$$

$$= \frac{5714}{19} - (16)^2 = 300.74 - 256 = 44.74$$

$$S_x = \sqrt{\frac{\sum h_i X_i^2}{\sum h_i} - \overline{X}^2}$$

الانحراف المعياري بالصيغة المختصرة

$$S_x = \sqrt{\frac{5714}{19} - 256} = \sqrt{44.73} = 6.69$$

خصائص الانحراف المعياري:

1 – إذا كان الانحراف المعيار للقيم $X_1, X_2, X_3, X_4 \ldots\ldots X_n$ هو S_x

فإنه إذا أضيفت أو طرحت قيمة ثابتة (a) إلى أو من جميع القيم فإن الانحراف المعياري لا يتغير

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

بتربيع الطرفين

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

بضرب الطرفين في (n)

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

نطرح قيمة ثابتة (a) من جميع القيم فنحصل على قيم جديدة $y_i = X_i - a$

$$\begin{aligned}
nSy_{ii}^2 &= \sum (X_i - a)^2 - n(\overline{X} -_i a)^2 \\
&= \sum \; (X_i^2 - 2aX_i + a^2) - n(\overline{X} - a)^2 \\
&= \sum X_i^2 - 2a\sum X_i + na^2 - n(\overline{X}^2 - 2a\overline{X} + a^2) \\
&= \sum X_i^2 - 2a\sum X_i + na^2 - n\overline{X}^2 + 2an\overline{X} - na^2 \\
&= \sum X_i^2 - 2an\frac{\sum X_i}{n} - n\overline{X}^2 + 2anX_i + na^2 - na^2 \\
&= \sum X_i^2 - 2a\sum X_i + na^2 - n\overline{X}^2 - 2an\overline{X} - na^2 \\
&= \sum X_i^2 - n\overline{X}^2 = nS_x^2
\end{aligned}$$

2 – يأخذ الانحراف المعياري نفس وحدة القياس للمتغير الأصلي (كلغ، متر، لتر ....) لذلك لا يمكن استخدامها كأساس للمقارنة بين تشتت توزيعين لهما وحدات قياس مختلفة.

3 – بما أن الانحراف المعياري يتأثر بالمتوسط الحسابي لبيانات الظاهرة فإنه لا يمكن استخدامه للمقارنة بين تشتت بيانات توزيعين لهما متوسط حسابي مختلف ولو كان هذين التوزيعين من نفس النوعية

4 – لا يمكن إيجاده بالنسبة للتوزيعات التكرارية المفتوحة من البداية أو النهاية.

رابعا – معامل الاختلاف Coefficient de variation

إذا كانت الظواهر من صفات مختلفة أو إذا كانت متوسطاتها متباعدة، فإن المقارنة اعتمادا على الانحراف المعياري ستكون غير منطقية وغير واقعية، ولهذا السبب وجدت مقاييس أخرى سميت مقاييس التشتت النسبي تعتمد على تمييز البيانات وتقيس التشتت كنسبة مئوية للمتوسط، أهم هذه المقاييس هو معامل الاختلاف.

ويحسب كما يلي:

$$\text{معامل الاختلاف} = \frac{\text{الانحراف المعياري}}{\text{المتوسط الحسابي}} \times 100 \qquad CV = \frac{Sx}{\overline{X}} \times 100$$

مثال7:

إذا كان متوسط درجات مجموعة من الطلبة في مادة ما هو 15 بانحراف معياري 3 ومتوسط درجاتهم في مادة أخرى هو 8 بانحراف معياري 2، فأي الدرجات في نظرك أكثر تشتتا؟

الحل:

إذا اعتمدنا على الانحراف المعياري فإننا نحكم على أن درجات المادة الأولى أكثر تشتتا (Sx =3) من درجات المادة الثانية (Sx =2)، وهذا غير صحيح لأننا إذا أدخلنا المتوسط الحسابي لدرجات الطلبة في المادتين في الحسبان سنحصل على النتائج التالية:

$$CV_1 = \frac{3}{15} \times 100 = 20\%$$

$$CV_2 = \frac{2}{8} \times 100 = 25\%$$

أي أن درجات المادة الثانية أكثر تشتتا

مثال8:

ينتج مصنع نوعين من المصابيح الكهربائية، فإذا علمت أن المتوسط الحسابي والانحراف المعياري لعمر المصباح في كل نوع هما:

Sx_1 = 300 ساعة $\qquad \overline{X}_1$ = 1500 ساعة

Sx_2 = 325 ساعة $\qquad \overline{X}_2$ = 1800 ساعة

أي المصابيح لها مدة حياة أكثر تشتت؟

الحل:

معامل الاختلاف للنوع الأول: $CV_1 = \frac{300}{1500} \times 100 = 20\%$

معامل الاختلاف للنوع الثاني: $CV_2 = \frac{325}{1800} \times 100 = 18\%$

أي أن النوع الأول من المصابيح الكهربائية لها مدة حياة أكثر تشتتا.