٤. مقاييس التشتت (الاختلاف)
Measures of Dispersion (Variation)

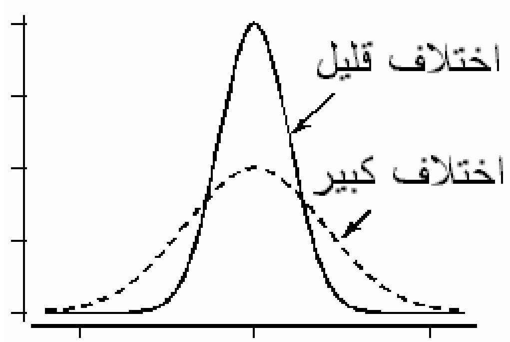
(٤-١) مقدمة:

لقد ذكرنا في الفصل السابق بعض مقاييس النزعة المركزية والتي تمثل مقاييس عددية لموضع أو مكان تركز البيانات لظاهرة ما. وقد ذكرنا بأن هذه المقاييس تستخدم لمقارنة مجموعات البيانات المختلفة. وفي الحقيقة فإن مقاييس النزعة المركزية غير كافية لإيجاد مقارنة شاملة بين مجموعات البيانات المختلفة. فقد تكون هناك مجموعات من البيانات لها نفس مقاييس النزعة المركزية (لها نفس الموضع) ولكنها تختلف في بعض الصفات الأخرى. فمثلاً المثال التالي يبين لنا مجموعتين من البيانات لهما نفس المتوسط ولكنهما مختلفتان في طبيعة تشتتهما.

مثال:

المجموعة	البيانات	المتوسط	
الأولى	59, 61, 62, 58, 60	60	-----------------●--●--●--●--●-----------
الثانية	50, 60, 66, 54, 70	60	-----------●-----●-----●-----●------●--- 60

بالرغم من أن المتوسط يساوي 60 للمجموعتين إلا أن التشتت (أو الاختلاف) بين القيم في كل مجموعة غير متساو. فمن الواضح أن بيانات المجموعة الأولى أكثر تقاربًا فيما بينها (أقل تشتتًا وتباعدًا فيما بينها) من بيانات المجموعة الثانية. لذلك دعت الحاجة لإيجاد مقاييس تقيس طبيعة تشتت (أو تفرق أو اختلاف أو تباعد) البيانات فيما بينها. هذه المقاييس تسمى مقاييس التشتت أو الاختلاف.



المضلعان التكراريان لتوزيعين لهما نفس مقاييس
النزعة المركزية ولكنهما مختلفين في التشتت

مقاييس التشتت هي مقاييس عددية تستخدم لقياس اختلاف أو تشتت البيانات. والاختلاف أو التشتت لمجموعة من البيانات هو مقدار تفرق أو تباعد أو انتشار البيانات فيما بينها. فتشتت البيانات يكون صغيرًا إذا كانت البيانات متقاربة فيما بينها والعكس بالعكس. وأما البيانات المتساوية فلا اختلاف ولا تشتت فيها. ومقاييس التشتت تستخدم لوصف مجموعة البيانات وكذلك لمقارنة مجموعات البيانات المختلفة إذ أن مقاييس النزعة المركزية لا تكفي وحدها لوصف مجموعة البيانات أو مقارنة مجموعات البيانات المختلفة. ومن أشهر مقاييس التشتت نذكر:

١. المدى: Range
٢. نصف المدى الربيعي Semi-Inter-quartile Range
٣. التباين: Variance
٤. الانحراف المعياري: Standard Deviation
٥. معامل الاختلاف (أو التغير): Coefficient of Variation

(٤-٢) المدى: Range

يعتبر المدى من أسهل مقاييس التشتت تعريفاً وحسابًا ويعطينا فكرة سريعة عن مدى تفرق البيانات. ويعرف المدى لمجموعة من البيانات بالصيغة التالية:

$$\boxed{\text{Range} = X_{max} - X_{min}}$$

حيث أن:

X_{max} = أكبر قيمة (للبيانات المفردة) = مركز الفترة العليا (للبيانات المبوبة)

X_{min} = أصغر قيمة (للبيانات المفردة) = مركز الفترة الدنيا (للبيانات المبوبة)

مثال (٤-١):

أوجد المدى للمشاهدات التالية والتي هي عبارة عن أوزان (بالكيلوجرام) مجموعة مكونة من سبعة أشخاص: 25, 30, 40, 45, 35, 55, 50

الحل:

$X_{max} = 55$
$X_{min} = 25$
$\text{Range} = X_{max} - X_{min} = 55 - 25 = 30$ (كيلوجرامًا)

مثال (٤-٢):

أوجد المدى لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (٢-٢).

الحل:

مستوى الهيموجلوبين	مركز الفترة x	التكرار f
12.95 – 13.95	**13.45**	3
13.9 5– 14.95	14.45	5
14. 95– 15.95	15.45	15
15. 95– 16.95	16.45	16
16. 95– 17.95	17.45	10
17. 95– 18.95	**18.45**	1

X_{max} = مركز الفترة العليا = 18.45

X_{min} = مركز الفترة الدنيا = 13.45

$$\text{Range} = X_{max} - X_{min}$$
$$= 18.45 - 13.45$$
$$= 5.00$$

بعض مميزات وعيوب المدى:

- **مميزات المدى:** سهل التعريف والحساب
- **عيوب المدى:**

١. يتأثر المدى بالقيم الشاذة أو المتطرفة.

٢. لا يأخذ المدى في الاعتبار جميع البيانات.

ملاحظات:

١. وحدة المدى هي نفس وحدة البيانات الأصلية.

٢. نظرًا لأن المدى يعتمد فقط على أكبر وأصغر قيمة ولا يأخذ في الاعتبار القيم الأخرى فهو مقياس غير جيد لقياس التشتت.

(٣-٤) نصف المدى الربيعي Semi-Inter-quartile Range:

رأينا أن المدى يتأثر كثيرًا بالقيم الشاذة أو المتطرفة. ولذلك دعت الحاجة لإيجاد مقاييس أخرى للتشتت لا تتأثر بالقيم المتطرفة. وأحد هذه المقاييس هو نصف المدى الربيعي. وحيث أن القيم المتطرفة هي تلك القيم الصغيرة جدًا أو الكبيرة جدًا فإنه عند حساب نصف المدى الربيعي لا يؤخذ في الاعتبار ربع البيانات الصغيرة (25%) ولا ربع البيانات الكبيرة (25%).

¼ n =25% ¼ n= 25%

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

ربع البيانات الصغيرة **Q_1** **Q_2** **Q_3** ربع البيانات الكبيرة

يرمز لنصف المدى الربيعي بالرمز Q ويعرف بالصيغة التالية:

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

حيث أن Q_1 هو الربيع الأول و Q_3 هو الربيع الثالث وقد مر معنا كيفية إيجادهما للبيانات المبوبة بالطريقة الحسابية والبيانية.

مثال (٤-٣):

أوجد نصف المدى الربيعي لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخـصا تـم تلخيص مستوى الهيموجلوبين لهم كما في مثال (٢-٢) باستخدام:

(أ) الطريقة الحسابية

(ب) الطريقة البيانية

الحل:

(أ) الطريقة الحسابية:

	مستوى الهيموجلوبين		التكرار المتجمع الصاعد	
	12.95	أقل من	0	
	13.9 5	أقل من	3	
$Q_1 \Rightarrow$	14. 95= A	أقل من	$8 = F_1$	$\leftarrow R = \frac{n}{4} = 12.5$
$Q_3 \Rightarrow$	15. 95= A^*	أقل من	$23 = F_2 = F^*_1$	$\leftarrow R^* = \frac{3n}{4} = 37.5$
	16. 95	أقل من	$39 = F^*_2$	
	17. 95	أقل من	49	
	18.95	أقل من	50	

حساب الربيع الأول Q_1:

$$R = \frac{n}{4} = \frac{50}{4} = 12.5 \ , \ A = 14.95, \ \ L = 1.00, \ \ F_1 = 8 \ , \ F_2 = 23$$

$$Q_1 = A + \left(\frac{R - F_1}{F_2 - F_1}\right) \times L = 14.95 + \left(\frac{12.5 - 8}{23 - 8}\right) \times 1.00 = 15.25$$

حساب الربيع الثالث Q_3:

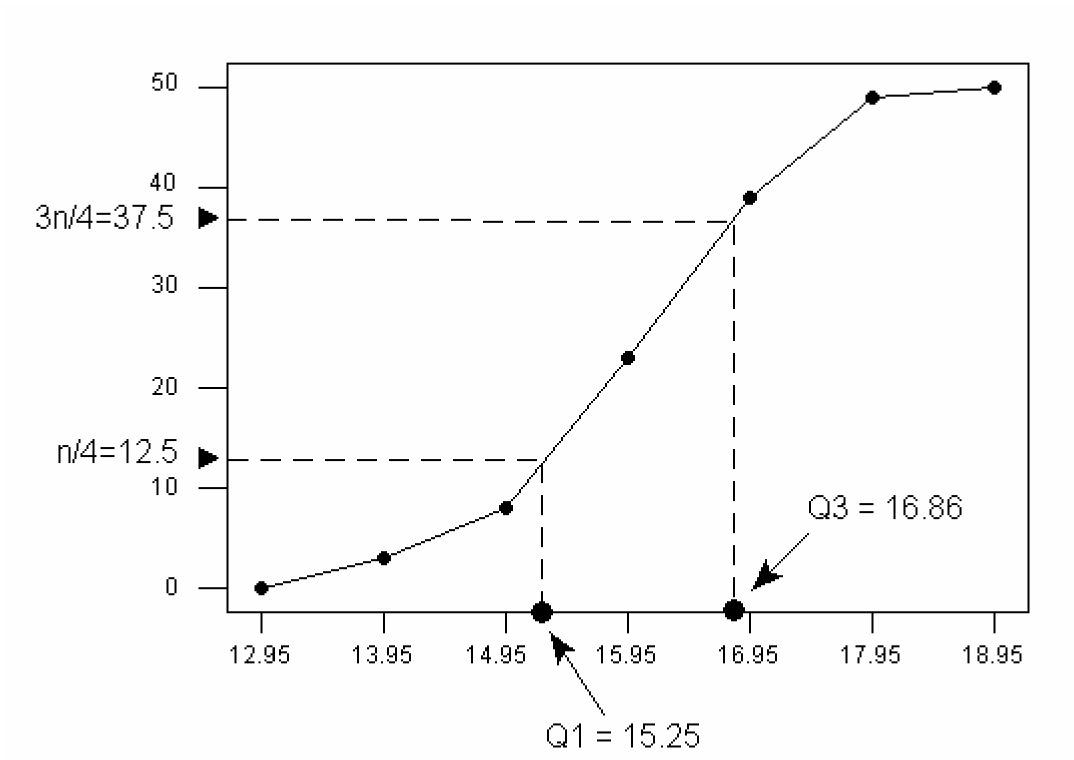
$$R^* = \frac{3n}{4} = \frac{3 \times 50}{4} = 37.5 \ , \ A^* = 15.95, \ \ L^* = 1.00, \ \ F^*_1 = 23 \ , \ F^*_2 = 39$$

$$Q_3 = A^* + \left(\frac{R^* - F_1^*}{F_2^* - F_1^*}\right) \times L^* = 15.95 + \left(\frac{37.5 - 23}{39 - 23}\right) \times 1.00 = 16.86$$

وباستخدام الصيغة فإن نصف المدى الربيعي هو:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{16.86 - 15.25}{2} = 1.61$$

(ب) الطريقة البيانية:



وباستخدام الصيغة فإن نصف المدى الربيعي هو:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{16.86 - 15.25}{2} = 1.61$$

بعض مميزات وعيوب نصف المدى الربيعي:

- **من المميزات:** لا يتأثر بالقيم الشاذة أو المتطرفة.
- **من العيوب :** لا يأخذ في الاعتبار جميع البيانات.

ملاحظة:

وحدة نصف المدى الربيعي هي نفس وحدة البيانات الأصلية.

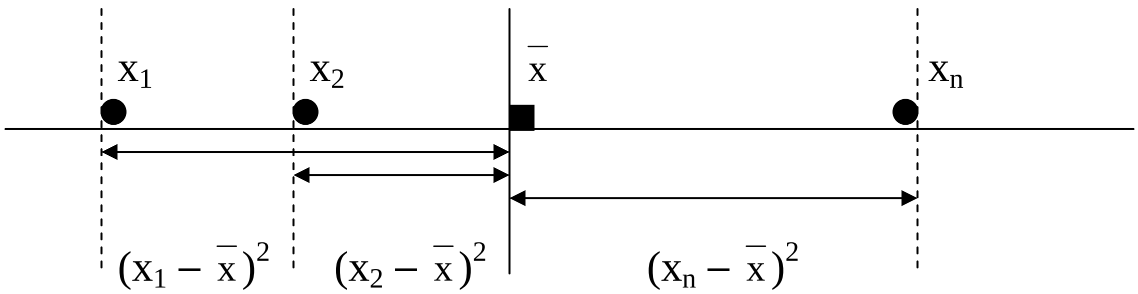
(٤-٤) التباين (Variance) والانحراف المعياري (Standard Deviation):

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

التباين: Variance

فكرة التباين تعتمد على تشتت أو تباعد البيانات عن متوسطها. فالتباين يكون كبيرًا إذا كانت البيانات متباعدة عن متوسطها والعكس بالعكس.

ويعرف التباين بأنه متوسط مربع انحرافات القيم عن الوسط الحسابي ويرمز له بالرمز S^2.



x_1	x_2	$\cdots$	x_n	القيم (البيانات)
$x_1 - \bar{x}$	$x_2 - \bar{x}$	$\cdots$	$x_n - \bar{x}$	انحرافات القيم عن المتوسط
$(x_1 - \bar{x})^2$	$(x_2 - \bar{x})^2$	$\cdots$	$(x_n - \bar{x})^2$	مربع انحرافات القيم عن المتوسط

الانحراف المعياري:

إن التباين من أهم وأفضل مقاييس التشتت ولكنه يقاس بوحدة البيانات الأصلية المربعة. وفي كثير من الأحيان نرغب في استخدام مقياس للتشتت يقاس بوحدة البيانات الأصلية ويتمتع بخصائص إحصائية جيدة مثل التباين. وأحد هذه المقاييس هو الانحراف المعياري. ويعرف الانحراف المعياري على أنه الجذر التربيعي للتباين ويرمز له بالرمز S.

حساب التباين والانحراف المعياري:

أولاً: التباين والانحراف المعياري للبيانات المفردة (غير المبوبة):

إذا كانت $x_1, x_2, \ldots, x_n$ عينه حجمها n وكان متوسطها هو $\bar{x}$ فإن تباين العينة يعرف كما يلي:

$$S^2 = \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$$

وأما الانحراف المعياري فإنه يعرف بالصيغة:

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

ملاحظات:

١. $S^2 \geq 0$ (دائمًا) وكذلك $S \geq 0$ (دائمًا).

٢. $S^2 = 0 \Leftrightarrow S = 0 \Leftrightarrow$ جميع قيم العينة متساوية (لا يوجد اختلاف بين القيم).

٣. وحدة S^2 هي وحدة البيانات الأصلية المربعة.

٤. وحدة S هي نفس وحدة البيانات الأصلية.

٥. يمكن حساب التباين بالصيغة الحسابية التالية:

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

ولحساب تباين العينة باستخدام الصيغة الحسابية السابقة فإننا نحتاج إلى معرفة الكميات التالية فقط دون الحاجة لمعرفة البيانات الأصلية:

- حجم العينة = n .
- مجموع البيانات = $\sum_{i=1}^{n}x_i$.
- مجموع مربعات البيانات = $\sum_{i=1}^{n}x_i^2$.

والصيغة الحسابية السابقة تستخدم لحساب تباين العينة وذلك لسببين هما:

١. لأنها أكثر سهولة.

٢. لأنها أكثر دقة في الحساب عندما يكون هناك تقريب في حساب متوسط العينة.

مثال (٤-٤):

أوجد تباين العينة والانحراف المعياري لمجموعة الأوزان (بالكيلوجرام) التالية:

7.1, 2.5, 2.5, 5.4, 8.3

الحل:

نلخص الحل في الجدول التالي:

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	x^2
7.1	1.94	3.7636	50.41
2.5	−2.66	7.0756	6.25
2.5	−2.66	7.0756	6.25
5.4	0.24	0.0576	29.16
8.3	3.14	9.8596	68.89
$\sum_{i=1}^{n} x_i = 25.8$	0.00	$\sum_{i=1}^{n}(x_i - \bar{x})^2 = 27.832$	$\sum_{i=1}^{n} x_i^2 = 160.96$

$n = 5$

$\sum_{i=1}^{n} x_i = 25.8$

$\sum_{i=1}^{n}(x_i - \bar{x})^2 = 27.832$

$\sum_{i=1}^{n} x_i^2 = 160.96$

متوسط العينة هو:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{25.8}{5} = 5.16 \text{ (كيلوجرامًا)}$$

حساب تباين العينة:

(أ) باستخدام التعريف:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{27.832}{5-1} = 6.958 \text{ (كيلوجرامًا مربعًا)}$$

(ب) باستخدام الصيغة الحسابية:

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{160.96 - \frac{(25.8)^2}{5}}{5-1} = \frac{160.96 - 133.128}{4} = 6.958$$

الانحراف المعياري هو:

$$S=\sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = S=\sqrt{6.958} = 2.6378 \text{ (كيلوجرامًا)}$$

بعض خصائص التباين والانحراف المعياري:

١. يخضع التباين والانحراف المعياري لبعض العمليات الجبرية كما يلي:

المشاهدات	الانحراف المعياري	التباين
$x_1, x_2, \ldots, x_n$	S	S^2
$x_1 \pm b$, $x_2 \pm b$, …, $x_n \pm b$	S	S^2
ax_1 , ax_2 , …, ax_n	$\|a\|$ S	$a^2 S^2$
$ax_1 \pm b$, $ax_2 \pm b$, …, $ax_n \pm b$	$\|a\|$ S	$a^2 S^2$

- مثال:

	المشاهدات	الانحراف المعياري	التباين
: x	2, 6, 4, 3, 5	S = 1.581	S^2= 2.5
: x+5	7, 11, 9, 8, 10	1.581	2.5
: 3x	6, 18, 12, 9, 15	$\|3\| \times 1.581 = 4.743$	$9 \times 2.5 = 22.5$
: 3x +5	11, 23, 17, 14, 20	$\|3\| \times 1.581 = 4.743$	$9 \times 2.5 = 22.5$

- مثال:

إذا كان التباين للمشاهدات $x_1, x_2, \ldots, x_n$ هو 36 فإن التباين للمشاهدات $\frac{x_1-10}{2}, \frac{x_2-10}{2}, \ldots, \frac{x_n-10}{2}$ هو $\left(\frac{1}{2}\right)^2 36 = \frac{36}{4} = 9$ وأما الانحراف المعياري فهو $\sqrt{9}=3$.

٢. إذا كان لدينا مجموعتان من البيانات بحيث أن عدد بيانات المجموعة الأولى n_1 ومتوسطها $\bar{x}_1$ وتباينها S_1^2 وكان عدد بيانات المجموعة الثانية n_2 ومتوسطها $\bar{x}_2$ وتباينها S_2^2 وإذا كان $\bar{x}_1 = \bar{x}_2$ (أي أن متوسطي المجموعتين متساويان) فإن تباين المجموعة الكلية المكونة من دمج هاتين المجموعتين يمكن حسابه بالصيغة التالية:

$$\boxed{S^2 = \frac{(n_1-1)\,S_1^2 + (n_2-1)\,S_2^2}{n_1+n_2-1}}$$

- مثال:

أوجد تباين المجموعة الكلية المكونة من دمج المجموعتين التاليتين:

	المجموعة الأولى	المجموعة الثانية
حجم العينة	$n_1 = 4$	$n_2 = 6$
المتوسط	$\bar{x}_1 = 5$	$\bar{x}_2 = 5$
التباين	$S_1^2 = 3$	$S_2^2 = 3.5$

الحل:

$$S^2 = \frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 1}$$

$$= \frac{(4-1)(3) + (6-1)(3.5)}{4 + 6 - 1} = 2.944$$

ثانياً: التباين والانحراف المعياري للبيانات المبوبة:

إذا كان لدينا بيانات عددها n وكانت هذه البيانات ملخصة في جدول تكراري بحيث أن:

- عدد الفترات هو k
- مراكز الفترات هي $x_1, x_2, \ldots, x_k$
- تكرارات الفترات هي $f_1, f_2, \ldots, f_k$

بطريقة مشابهة لحسب المتوسط للتوزيع التكراري فإن التباين للتوزيع التكراري المبـوب يمكـن حسابه بشكل تقريبي بالصيغة التالية:

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

حيث أن:

$$n = \sum_{i=1}^{k} f_i\,, \quad \bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum xf}{n}\,,$$

كما يمكن استخدام الصيغة الحسابية التالية:

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

ويكمن تلخيص عمليتي إيجاد المتوسط والتباين باستخدام الجدول التالي:

الفترة	مركز الفترة x	التكرار f	$x f$	$x^2 f$	$f(x-\overline{x})^2$
الفترة رقم 1	x_1	f_1	$x_1 f_1$	$x_1^2 f_1$	$f_1 (x_1-\overline{x})^2$
الفترة رقم 2	x_2	f_2	$x_2 f_2$	$x_2^2 f_2$	$f_2 (x_2-\overline{x})^2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
الفترة رقم k	x_k	f_k	$x_k f_k$	$x_k^2 f_k$	$f_k (x_k-\overline{x})^2$
المجموع		$\sum f = n$	$\sum x f$	$\sum x^2 f$	$\sum_{i=1}^{k} f_i (x_i-\overline{x})^2$

مثال (٤-٥):

أوجد التباين والانحراف المعياري لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (٢-٢).

الحل:

مستوى الهيموجلوبين	مركز الفترة x	التكرار f	$x f$	$x^2 f$	$f(x-\overline{x})^2$ $f(x-16.01)^2$
12.95 – 13.95	13.45	3	40.35	542.708	19.6608
13.9 5– 14.95	14.45	5	72.25	1044.013	12.1680
14. 95– 15.95	15.45	15	231.75	3580.538	4.7040
15. 95– 16.95	16.45	16	263.20	4329.640	3.0976
16. 95– 17.95	17.45	10	174.50	3045.025	20.7360
17. 95– 18.95	18.45	1	18.45	340.403	5.9536
المجموع		$\sum f = 50$	$\sum x f = 800.5$	$\sum x^2 f = 12882.33$	$\sum_{i=1}^{k} f_i (x_i-\overline{x})^2 = 66.320$

$$\overline{x} = \frac{\sum xf}{\sum f} = \frac{\sum xf}{n} = \frac{800.5}{50} = 16.01$$

حساب التباين باستخدام صيغة التعريف:

$$S^2 = \frac{\sum_{i=1}^{k} f_i (x_i-\overline{x})^2}{n-1} = \frac{66.320}{50-1} = 1.3535$$

حساب التباين باستخدام الصيغة الحسابية:

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k} x_i^2 f_i - \frac{\left(\sum_{i=1}^{k} x_i f_i\right)^2}{n}\right) = \frac{1}{50-1}\left(12882.33 - \frac{(800.5)^2}{50}\right)$$

$$= \frac{1}{49}(12882.33 - 12816.005) = \frac{66.325}{49} = 1.3536$$

حساب الانحراف المعياري:

$$S = \sqrt{S^2} = \sqrt{1.3536} = 1.163$$

(٤-٥) معامل الاختلاف (التغير): Coefficient of Variation:

ذكرنا سابقًا أن التباين والانحراف المعياري من المقاييس المفيدة لقياس التشتت لتوزيع متغير ما. ولكن في كثير من الأحيان نكون مهتمين بمقارنة التشتت والاختلاف لتوزيعي متغيرين مختلفين. وبما أن التباين والانحراف المعياري مقياسان يعتمدان على وحدة البيانات فإنه يصعب استخدامهما لمقارنة تجانس المجموعات المختلفة من البيانات وذلك لاختلاف الوحدة المستخدمة. وبشكل عام فإن مقاييس التشتت التي ذكرناها آنفًا تكون غير مناسبة لمقارنة تجانس مجموعات البيانات المختلفة في الحالتين التاليتين:

١. إذا كانت وحدتا المتغيرين مختلفتين حيث لا نستطيع مقارنة الوحدات المختلفة.

٢. إذا كان متوسطا المتغيرين مختلفين وذلك لأن تباين توزيع المتغير ذي المتوسط الصغير ينزع لأن يكون صغيرًا 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 أو معامل التغير. فمعامل الاختلاف هو أحد مقاييس التشتت النسبي وهو مقياس عديم الوحدة ويستخدم لمقارنة التشتت النسبي أو التجانس لمجموعات البيانات المختلفة. فمجموعة البيانات ذات معامل الاختلاف الأكبر يكون تشتتها النسبي أكبر أي أنها تكون أقل تجانسًا والعكس بالعكس. ويعرف معامل الاختلاف للعينة التي متوسطها $\bar{x}$ وانحرافها المعياري S بالصيغة التالية:

$$\boxed{C.V. = \frac{S}{\bar{x}}}$$

مثال (٤-٦):

الجدول أدناه يتضمن بيانات إحدى الدراسات التي طبقت على خمسة أشخاص لقياس الوزن (بالكيلوجرام) والطول (بالسنتيمتر). أي البيانات أكثر تشتتًا نسبًا (أقل تجانسًا) بيانات الأوزان أم بيانات الأطوال؟

رقم الشخص	1	2	3	4	5
الوزن	69	59	65	67	65
الطول	164	162	155	165	158

الحل:

أولاً نوجد المتوسط $\bar{x}$ والانحراف المعياري S لكل من بيانات الأوزان ولبيانات الأطوال كما مر معنا سابقاً. نلخص الحسابات في الجدول التالي:

البيانات	المتوسط $\bar{x}$	الانحراف المعياري S	معامل الاختلاف $C.V. = \frac{S}{\bar{x}}$
الأوزان	65.0 kg	3.7417 kg	0.0576
الأطوال	160.8 cm	4.2071 cm	0.026

بما أن معامل الاختلاف لبيانات الأوزان أكبر من معامل الاختلاف لبيانات الأطوال فإن التشتت النسبي لبيانات الأوزان أكبر من التشتت النسبي لبيانات الأطوال. أي أن بيانات الأوزان أقل تجانسًا من بيانات الأطوال.

(٤-٦) نظرية (متراجحة) تشيبيشيف **Chebychev Inequality**:

إن نظرية تشيبيشيف من النظريات المفيدة إذا أنها تعطينا حدًا أدنى لنسبة البيانات الواقعة في فترة معينة عند معرفة متوسط البيانات وانحرافها المعياري دون الحاجة لمعرفة البيانات الأصلية أو التوزيع الذي أخذت منه العينة. ونص النظرية هو:

> إذا كان لدينا عينة من البيانات متوسطها $\bar{x}$ وانحرافها المعياري S فإن نسبة البيانات الواقعة في الفترة $(\bar{x} - kS , \bar{x} + kS)$ لايقل عن $\left(1 - \frac{1}{k^2}\right)$ حيث أن $k>1$.

$$\underbrace{\bar{x} - kS \qquad\qquad \bar{x} \qquad\qquad \bar{x} + kS}$$

نسبة البيانات الواقعة في هذه الفترة لا يقل عن $\left(1 - \dfrac{1}{k^2}\right)$

ملاحظات:

١. تطبق نظرية تشيبيشيف للفترات التي منتصفها (مركزها) هو المتوسط.

٢. تستخدم نظرية تشيبيشيف بطريقتين(في كلا الحالتين لابد من معرفة قيمة k):

أ- تحديد النسبة (التقريبية) لعدد البيانات الواقعة في فترة معينة.

ب- تحديد الفترة التي يقع فيها ما لا يقل عن نسبة معينة.

مثال (٤-٧):

إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x} = 7$ وانحرافها المعياري $S=5$ فما هي نسبة البيانات الواقعة في الفترة $(-4 , 18)$؟

الحل:

أولاً نلاحظ أن منتصف الفترة المعطاة $(-4 \ , \ 18)$ هو المتوسط $\bar{x} = 7$ لذلك نستطيع تطبيق نظرية تشيبيشيف. والآن:

$$(\bar{x} - kS \ , \bar{x} + kS) = (-4 , 18) \quad \Rightarrow \quad \bar{x} + kS = 18$$

$$\Leftrightarrow \quad 7 + k(5) = 18$$

$$\Leftrightarrow \quad 5k = 11$$

$$\Leftrightarrow \quad k = 11/5$$

$$\Leftrightarrow \quad \left(1 - \frac{1}{k^2}\right) = \left(1 - \frac{1}{(11/5)^2}\right)$$

$$\Leftrightarrow \quad \left(1 - \frac{1}{k^2}\right) = 0.7934$$

لذلك فإن نسبة البيانات الواقعة في الفترة $(-4 , 18)$ لا تقل عن 79.34%.

مثال (٤-٨):

إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x} = 7$ وانحرافها المعياري $S=5$ فأوجد فترة يقع فيها ما لا يقل عن 75% من البيانات.

الحل:

$$\left(1-\frac{1}{k^2}\right)=0.75 \Leftrightarrow \frac{1}{k^2}=1-0.75$$
$$\Leftrightarrow \frac{1}{k^2}=0.25$$
$$\Leftrightarrow k=\sqrt{\frac{1}{0.25}}$$
$$\Leftrightarrow k=2$$

وبالتالي فإن الفترة التي يقع فيها ما لا يقل عن 75% من البيانات هي:

$$(\overline{x}-kS\ ,\ \overline{x}+kS) = (7-2\times5\ ,\ 7+2\times5) = (7-10\ ,\ 7+10) = (-3\ ,\ 17)$$

(٤-٧) الدرجات (القيم) المعيارية:

لتكن $x_1, x_2, \ldots, x_n$ عينه من البيانات حجمها n ومتوسطها $\overline{x}$ وانحرافها المعيـــاري S.

نعرف الدرجة (القيمة) المعيارية للمشاهدة x_i بالصيغة التالية:

$$\boxed{z_i=\frac{x_i-\overline{x}}{S} \quad i=1, 2, \ldots, n}$$

أي أن الدرجات المعيارية للبيانات $x_1, x_2, \ldots, x_n$ هي:

$$z_1=\frac{x_1-\overline{x}}{S},\ z_2=\frac{x_2-\overline{x}}{S},\ \ldots,\ z_n=\frac{x_n-\overline{x}}{S}$$

ملاحظات:

١. $z_i=\frac{x_i-\overline{x}}{S}$ هي الدرجة المعيارية للمشاهدة الأصلية x_i.

٢. المشاهدة الأصلية للدرجة المعيارية z_i هي $x_i=\overline{x}+S\,z_i$.

٣. الدرجات المعيارية هي قيم عديمة الوحدة ولذلك فإنها تستخدم للمقارنة بـــين المـــشاهدات المختلفة في المجموعات المختلفة للبيانات.

٤. متوسط الدرجات المعيارية =0.

٥. الانحراف المعياري للدرجات المعيارية يساوي =1.

مثال (٤-٩):

إذا كان لدينا مجموعة من البيانات متوسطها $\overline{x}=7$ وانحرافها المعياري S=5 فأوجد:

١. الدرجة المعيارية للقيمة $x=9$.

٢. القيمة الأصلية للدرجة المعيارية $z = 0.1$.

الحل:

١. الدرجة المعيارية للقيمة $x = 9$ هي:

$$z=\frac{x-\bar{x}}{S} = \frac{9-7}{5}=\frac{2}{5}=0.4$$

٢. القيمة الأصلية للدرجة المعيارية $z = 0.1$ هي:

$$x = \bar{x} + S\,z = 7 + 5\times0.1 = 7 + 0.5 = 7.5$$

مثال (٤-١٠):

إذا كانت درجة أحد الطلاب في مقرر الإحصاء تساوي 82 ودرجته في مقرر الرياضيات تساوي 89، وإذا كان متوسط درجات الطلاب في مقرر الإحصاء يساوي 75 بانحراف معياري يساوي 10 ومتوسط درجات الطلاب في مقرر الرياضيات يساوي 81 بانحراف معياري يساوي 16، ففي أي المقررين كان أداء الطالب أفضل؟

الحل:

المقرر	المتوسط $\bar{x}$	الانحراف المعياري S	الدرجة x	الدرجة المعيارية $z=\frac{x-\bar{x}}{S}$
الإحصاء	75	10	82	$z=\frac{82-75}{10}=0.7$
الرياضيات	81	16	89	$z=\frac{89-81}{16}=0.5$

بما أن الدرجة المعيارية لمقرر الإحصاء 0.7 أكبر من الدرجة المعيارية لمقرر الرياضيات 0.5 فإن أداء الطالب في مقرر الإحصاء أفضل من أدائه في مقرر الرياضيات بالرغم من أن درجته في مقرر الإحصاء أقل من درجته في مقرر الرياضيات.