جامعة الملك سعود
King Saud University

مقاييس التشتت (الاختلاف)
Measure of Dispersion (Variation)

الإحصاء والاحتمالات (١٢٠١ إحص)

الفصل الصيفي ١٤٣٨/١٤٣٧هـ

مقاييس النزعة المركزية

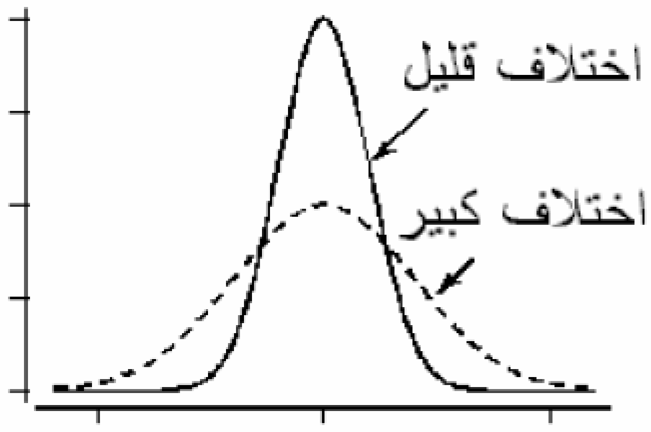
- مقاييس النزعة المركزية تمثل مقاييس عددية لموضع أو مكان تركز البيانات لظاهرة ما.
- مقاييس النزعة المركزية غير كافية لإيجاد مقارنة شاملة بين مجموعات البيانات المختلفة. فقد تكون هناك مجموعات من البيانات لها نفس مقاييس النزعة المركزية (لها نفس الموضع) ولكنها تختلف في بعض الصفات الأخرى.
- المثال التالي يبين لنا مجموعتين من البيانات لهما نفس المتوسط ولكنهما مختلفتان في طبيعة تشتتهما.

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

- بالرغم من أن المتوسط يساوي ٦٠ للمجموعتين إلا أن التشتت (أو الاختلاف) بين القيم في كل مجموعة غير متساو. فمن الواضح أن بيانات المجموعة الأولى أكثر تقارباً فيما بينها (أقل تشتتاً وتباعداً فيما بينها) من بيانات المجموعة الثانية.

مقاييس التشتت

- 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

- 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.
- تشتت البيانات يكون صغيراً إذا كانت البيانات متقاربة فيما بينها والعكس بالعكس. أما البيانات المتساوية فلا اختلاف ولا تشتت فيها.
- مقاييس 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.
- من أشهر مقاييس التشتت نذكر:
  - المدى Range
  - التباين Variance
  - الانحراف المعياري Standard Deviation
  - معامل الاختلاف (أو التغير) Coefficient of Variation

المدى Range

- يعتبر المدى من أسهل مقاييس التشتت تعريفاً وحساباً ويعطينا فكرة سريعة عن مدى تفرق البيانات.
- يعرف المدى لمجموعة من البيانات بالصيغة التالية:

$$Range = Xmax - Xmin$$

حيث أن:

Xmax= أكبر قيمة (للبيانات المفردة) = مركز الفترة العليا (للبيانات المبوبة)

Xmin= أصغر قيمة (للبيانات المفردة) = مركز الفترة الدنيا (للبيانات المبوبة)

المدى للبيانات المفردة (غير المبوبة)

مثال:

أوجد المدى للمشاهدات التالية والتي هي عبارة عن أوزان (بالكيلوجرام) مجموعة مكونة من سبعة أشخاص:

٢٥، ٣٠، ٤٠، ٤٥، ٣٥، ٥٥، ٥٠

الحل:

Xmax = 55

Xmin = 25

Range = Xmax – Xmin = 55 – 25 = 30 Kg

المدى للبيانات المبوبة

مثال:

أوجد المدى لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم.

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

الحل:

Xmax= مركز الفترة العليا = ١٨,٤٥

Xmin= مركز الفترة الدنيا = ١٣,٤٥

Range = Xmax – Xmin

= ١٨,٤٥ – ١٣,٤٥

= ٥,٠٠

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

التباين والانحراف المعياري Variance (Standard Deviation)

- يعتبر التباين والانحراف المعياري من أهم وأفضل مقاييس التشتت ومن أكثرها شيوعاً.

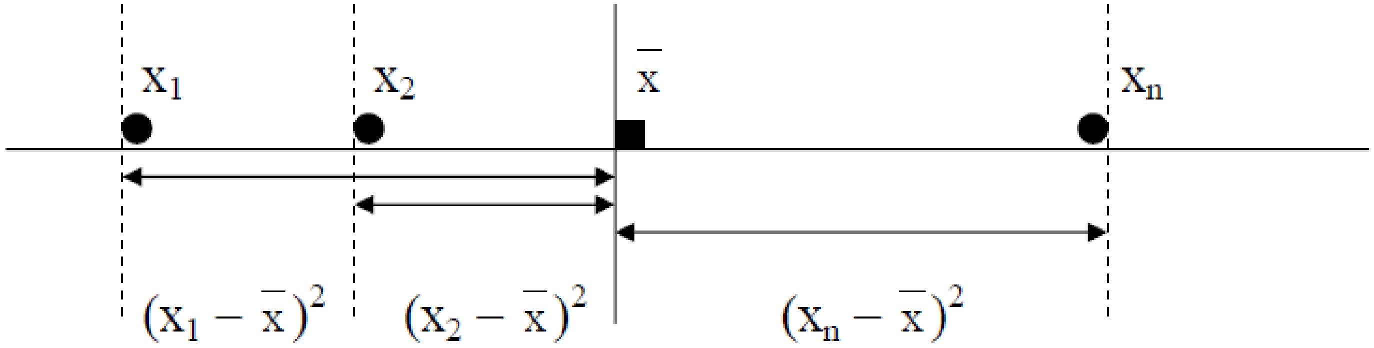
التباين Variance

- فكرة التباين تعتمد على تشتت أو تباعد البيانات عن متوسطها.
- التباين يكون كبيراً إذا كانت البيانات متباعدة عن متوسطها والعكس بالعكس.
- يعرف التباين بأنه متوسط مربع انحرافات القيم عن الوسط الحسابي ويرمز له بالرمز S^2.
- يقاس التباين بوحدة البيانات الأصلية المربعة.

الانحراف المعياري Standard Deviation

- في 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.

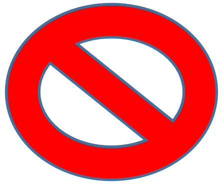
Variance (Standard Deviation) التباين والانحراف المعياري



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

التباين والانحراف المعياري للبيانات المفردة (غير المبوبة)

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



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

- يمكن حساب **التباين** بصيغة **أكثر سهولة**:

$$s^2 = \frac{1}{n-1}\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)$$

- **الانحراف المعياري** يعرف بالصيغة:

$$\boxed{S = \sqrt{S^2}}$$

التباين والانحراف المعياري للبيانات المفردة (غير المبوبة)

مثال:

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

$$s^2 = \frac{1}{n-1}\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)$$

٨,٣ ، ٥,٤ ، ٢,٥ ، ٢,٥ ، ٧,١

الحل:

x_i	x^2
7.1	50.41
2.5	6.25
2.5	6.25
5.4	29.16
8.3	68.89
$\sum_{i=1}^{n} x_i = 25.8$	$\sum_{i=1}^{n} x_i^2 = 160.96$

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

$$S^2 = \frac{1}{4}\left(160.96 - \frac{(25.8)^2}{5}\right) = \frac{1}{4}\left(160.96 - \frac{665.64}{5}\right)$$

$$= \frac{1}{4}(160.96 - 133.13) = \frac{27.83}{4} = 6.958$$

$$S = \sqrt{S^2} = \sqrt{6.958} = 2.6378 \text{ (كيلوجراماً)}$$

التباين والانحراف المعياري للبيانات المفردة (غير المبوبة)

إذا كان لدينا مجموعتان من البيانات بحيث أن:

- عدد بيانات المجموعة الأولى هو n_1 ومتوسطها هو $\bar{x}_1$ وتباينها هو S_1^2
- عدد بيانات المجموعة الثانية هو n_2 ومتوسطها هو $\bar{x}_2$ وتباينها هو S_2^2
- متوسطا المجموعتين متساويان

فإن تباين المجموعة الكلية المكونة من دمج هاتين المجموعتين يمكن حسابه بالصيغة التالية:

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

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	$S_1^2 = 3$	$S_2^2 = 3.5$

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

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

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 $f_1, f_2, \ldots, f_k$

الفترة	مركز الفترة x	التكرار f	$x\,f$	$x^2\,f$
الفترة رقم 1	x_1	f_1	$x_1\,f_1$	$x_1^2\,f_1$
الفترة رقم 2	x_2	f_2	$x_2\,f_2$	$x_2^2\,f_2$
:	:	:	:	:
:	:	:	:	:
الفترة رقم k	x_k	f_k	$x_k\,f_k$	$x_k^2\,f_k$
المجموع		$\sum f = n$	$\sum x\,f$	$\sum x^2 f$

- فإن التباين للتوزيع التكراري المبوب يمكن حسابه:

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

التباين والانحراف المعياري للبيانات المبوبة

مثال:

أوجد التباين والانحراف المعياري لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصاً تم تلخيص مستوى الهيموجلوبين لهم.

مستوى الهيموجلوبين	مركز الفترة x	التكرار f	$x\,f$	$x^2 f$
12.95 – 13.95	13.45	3	40.35	542.708
13.9 5– 14.95	14.45	5	72.25	1044.013
14. 95– 15.95	15.45	15	231.75	3580.538
15. 95– 16.95	16.45	16	263.20	4329.640
16. 95– 17.95	17.45	10	174.50	3045.025
17. 95– 18.95	18.45	1	18.45	340.403
المجموع		$\sum f = 50$	$\sum x f = 800.5$	$\sum x^2 f = 12882.33$

$$\overline{x} = \frac{\sum xf}{\sum f} = \frac{\sum xf}{n} = \frac{800\ .5}{50} = 16\ .01$$

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

$$= \frac{1}{50-1}\left(12882.33 - \frac{(800.5)^2}{50}\right)$$

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

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

Coefficient of Variation (التغير) معامل الاختلاف

- التباين والانحراف المعياري مقياسان يعتمدان على وحدة البيانات فإنه يصعب استخدامهما لمقارنة تجانس المجموعات المختلفة من البيانات وذلك لاختلاف الوحدة المستخدمة .
- لذلك دعت الحاجة إلى مقياس لا يعتمد على وحدة المتغير ويقيس ما يسمى بالتشتت النسبي .
- أحد هذه المقاييس هو ما يسمى بمعامل الاختلاف أو معامل التغير .
- معامل 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 تجانساً والعكس بالعكس .
- ويعرف معامل الاختلاف للعينة التي متوسطها $\bar{x}$ وانحرافها المعياري S بالصيغة التالية:

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

Coefficient of Variation (التغير) معامل الاختلاف

مثال:

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

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

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

الحل:

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 تجانساً من بيانات الأطوال.

نظرية (متراجحة) تشيبيشيف Chebychev Inequality

- نظرية تشيبيشيف تعطي حداً 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 البيانـــات الواقعة في الفترة $(\bar{x} - kS \,,\, \bar{x} + kS)$ لايقل عن $\left(1 - \frac{1}{k^2}\right)$ حيث أن $k>1$.

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

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

نظرية (متراجحة) تشيبيشيف Chebychev Inequality

مثال ١ :

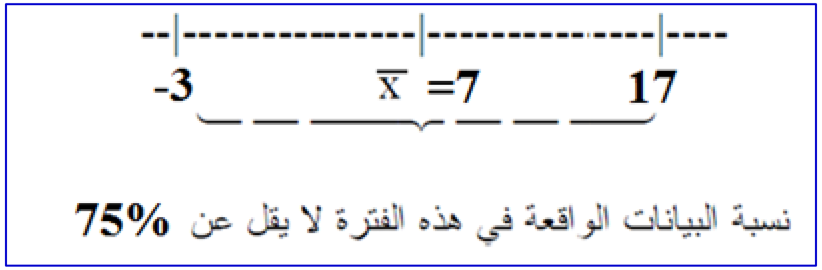
إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x}=7$ وانحرافها المعياري S=5 ، فأوجد فترة يقع فيها ما لا يقل عن 75% من البيانات.

الحل:

$$\left(1-\frac{1}{k^2}\right)=0.75 \Leftrightarrow \frac{1}{k^2}=1-0.75 \Leftrightarrow \frac{1}{k^2}=0.25 \Leftrightarrow k^2=\frac{1}{0.25} \Leftrightarrow k^2=4 \Leftrightarrow k=2$$

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

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



نظرية (متراجحة) تشيبيشيف Chebychev Inequality

مثال ٢ :

إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x}=7$ وانحرافها المعياري S=5 ، فأوجد فترة يقع فيها ما لا يقل عن 88.89% من البيانات.

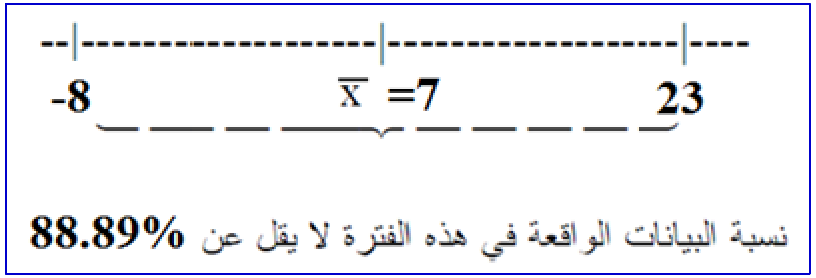
الحل:

$$\left(1-\frac{1}{k^2}\right)=0.8889 \Leftrightarrow \frac{1}{k^2}=1-0.8889 \Leftrightarrow \frac{1}{k^2}=0.1111 \Leftrightarrow k^2=\frac{1}{0.1111}$$

$$\Leftrightarrow k^2=9 \Leftrightarrow k=3$$

فإن الفترة التي يقع فيها ما لا يقل عن 88.89% من البيانات هي:

$$(\bar{x}-kS\,,\,\bar{x}-kS)=(7-3\times 5\,,7+3\times 5)=(7-15\,,7+15)=(-8\,,\ 23)$$



نظرية (متراجحة) تشيبيشيف Chebychev Inequality

مثال ٣:

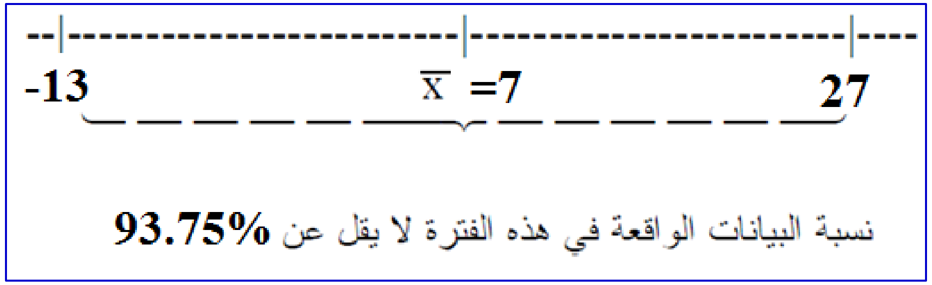
إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x}=7$ وانحرافها المعياري S=5 ، فأوجد فترة يقع فيها ما لا يقل عن 93.75% من البيانات.

الحل:

$$\left(1-\frac{1}{k^2}\right)=0.9375 \Leftrightarrow \frac{1}{k^2}=1-0.9375 \Leftrightarrow \frac{1}{k^2}=0.0625 \Leftrightarrow k^2=\frac{1}{0.0625} \Leftrightarrow k^2=16 \Leftrightarrow k=4$$

فإن الفترة التي يقع فيها ما لا يقل عن 93.75% من البيانات هي:

$$(\bar{x}-kS\,,\,\bar{x}-kS)=(7-4\times 5\,,7+4\times 5)=(7-20\,,7+20)=(-13\,,\ 27)$$



نظرية (متراجحة) تشيبيشيف Chebychev Inequality

مثال ٤ :

إذا كان لدينا مجموعة من البيانات متوسطها $\overline{x}=7$ وانحرافها المعياري S=5 ، فما هي نسبة البيانات الواقعة في الفترة (–٤ ، ١٨)؟

الحل:

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

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

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

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

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

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

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

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

الدرجات (القيم) المعيارية

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

نعرف الدرجة (القيمة) المعيارية للمشاهدة x_i بالصيغة التالية:

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

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

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

الدرجات (القيم) المعيارية

مثال:

إذا كانت درجة أحد الطلاب في مقرر الإحصاء تساوي ٨٢ ودرجته في مقرر الرياضيات تساوي ٨٩، وإذا كان متوسط درجات الطلاب في مقرر الإحصاء يساوي ٧٥ بانحراف معياري يساوي ١٠ ومتوسط درجات الطلاب في مقرر الرياضيات يساوي ٨١ بانحراف معياري يساوي ١٠ ، 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 ٠,٧ أكبر من الدرجة المعيارية لمقرر الرياضيات ٠,٥ ، 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.