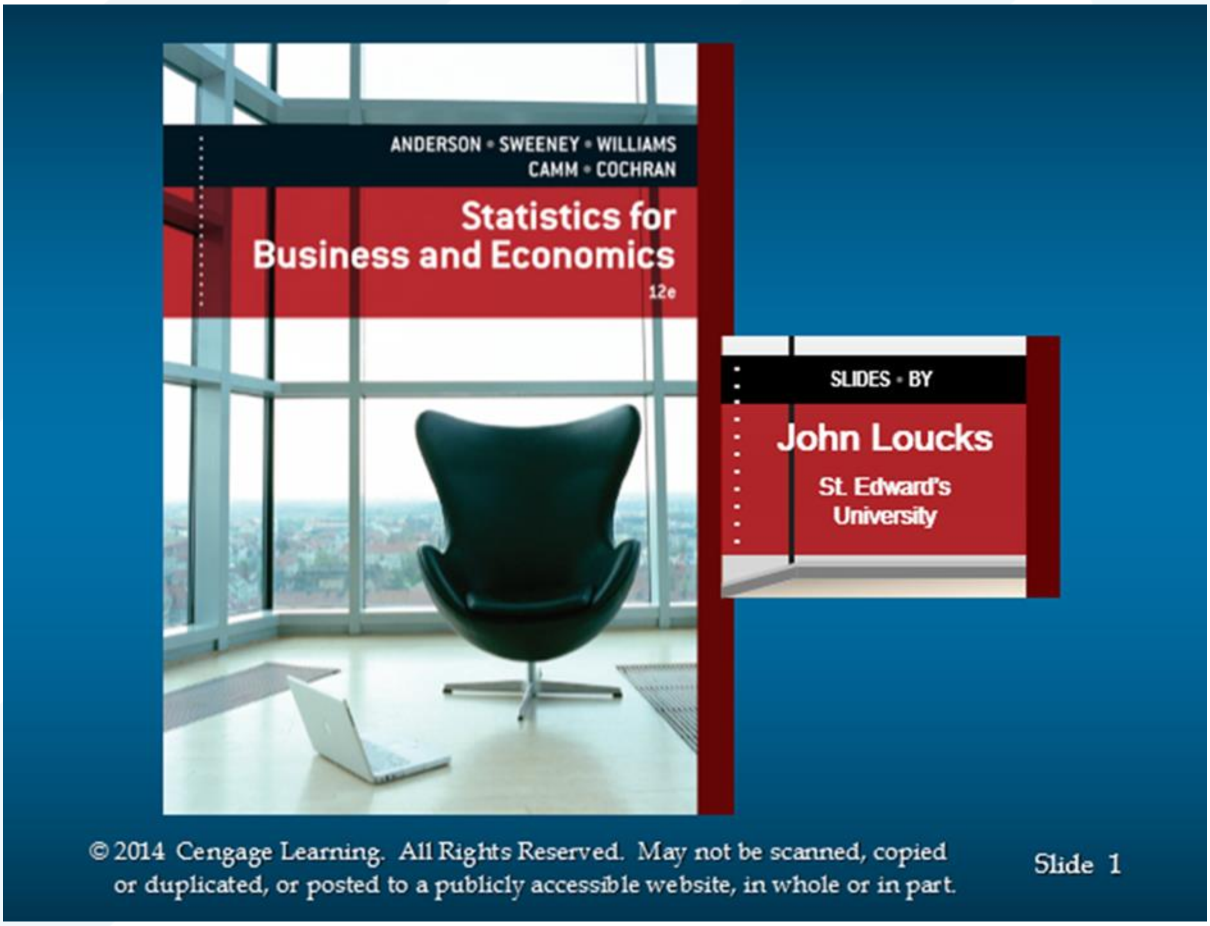
كلية إدارة الاعمال

الإحصاء 2 Statistics

الدكتور محمود محمد ديب طيوب

الفصل الثاني للعام 2024-2025

محاضرة رقم 1



التباين والانحراف المعياري:

Variance & Standard Deviation

يعدّ الانحراف المعياري من أدق مقاييس التشتّت، وهو مبني على نفس الأساس الذي بني عليه الانحراف المتوسط، أي على أساس متوسط مجموع انحرافات قيم المتغيّر عن المتوسط الحسابي وهو قيمة صالحة لقياس مدى تشتت هذه القيم، غير أن الانحراف المعياري لا يهمل إشارات الانحرافات، بل يتخلص من وجودها بطريقة رياضية مقبولة وهي إيجاد مربعات هذه الانحرافات، أي تستخدم المقدار:

$$\sigma_x^2 = \frac{\sum (x - \bar{x})^2}{n}$$

ولكن بما أن الأساس هو متوسط الانحرافات ذاتها وليس متوسط مربعاتها، لذا يكون من الضروري أخذ الجذر التربيعي لهذا المقدار، وهذا الإجراء يمكننا أيضاً من الاحتفاظ بوحدة القياس الأصلية للمتغيّر. وعلى هذا يعرف الانحراف المعياري بأنه الجذر التربيعي للتباين أي لمتوسط مربعات انحرافات القياسات عن متوسطها الحسابي ونرمزله بـ ع. أي:

$$\sigma_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

ونظراً لأن مجموع مربعات انحرافات القياسات عن متوسطها الحسابي أقل من مجموع مربعات انحرافات نفس القياسات عن أية قيمة تختلف عن المتوسط الحسابي زيادة أم نقصاناً جعل منه أدق مقياس للتشتت مقارنة بغيره من المقاييس الأخرى للتمركز إلا في حالة واحدة في حالة التوزيع الاعتدالي.

لكن يجب التمييز بين التباين والانحراف المعياري للعينة، والتباين والانحراف المعياري للمجتمع الإحصائي الأصلي، هذا وإن استخدام العلاقة التالية:

$$\sigma_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

يعطينا تقديراً للانحراف المعياري للمجتمع الأصلي أي أن القيم الناتجة من استخدام هذه العلاقة تميل إلى أن تكون أقل من القيم الحقيقية للانحرافات المعيارية للمجتمعات الأصلية. هذا وإن تقسيم مجموع مربعات الانحرافات على n-1 بدلاً من ن تعطينا تقديراً غير متحيزاً للانحراف المعياري أي:)

$$\sigma^2{}_x = \frac{\sum (x - \bar{x})^2}{n-1}$$

ونلاحظ أن (n) يشير إلى عدد القياسات أو الدرجات أو المشاهدات، بينما يشير (n-1) إلى عدد الانحرافات عن المتوسط التي يمكن أن تتغير، ويطلق عليها اسم درجات الحرية... هذا ولتوضيح مفهوم التباين والانحراف المعياري نورد المثال التالي:

قام باحث بدراسة لأثر تعاطي عقار معين على التذكر، فاختار عينتين عشوائيتين من الطلاب عدد أفراد كل منها عشرة طلاب فأعطى الأولى العقار (مجموعة تجريبية) والأخرى لم تعط العقار (مجموعة ضابطة) وكانت نتائج الاختيار للتذكر على النحو التالي:

المجموعة التجريبية	4	9	16	38	47	48	70	83	97	104
المجموعة الضابطة	35	38	39	44	49	56	67	69	68	71

فوجد أن متوسط المجموعة التجريبية $\bar{x}_1 = 51.6$ ومتوسط المجموعة الضابطة $\bar{x}_2 = 53.6$ ولإظهار في أي المجموعتين كانت الدرجات أكثر تشتتاً فوجد أن

الانحراف المعياري للمجموعة الأولى والانحراف المعياري.

$$\sigma_x = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{38324 - \frac{516^2}{10}}{10-1}} = \sqrt{\frac{38324 - 26625.6}{9}} = 36.0530$$

وبنفس الطيقة للمجموعة الثانية يساوي $\sigma_{2} = 13.924$ نجد أن درجات المجموعة التجريبية أكثر تشتتاً وانتشاراً من المجموعة الضابطة، مما يدل على أن أداء المجموعة التجريبية في الاختبار أكثر تبايناً من أداء المجموعة الضابطة وبالتالي نستنتج أن العقار يبدو له تأثيراً كبيراً على تباين الأداء على الرغم من أن تأثيره على مستوى الأداء كان طفيفاً.

-

طرق حساب التباين والانحراف المعياري:

1- طريقة تربيع الانحرافات:

*- حالة بيانات مفردة :

$$\sigma_x^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

مثال:

لتكن لدينا درجات خمسة تلاميذ في مقرر الحساب (أصل الدرجة 10):

x: 5، 6، 7، 8، 9. لحساب التباين والانحراف المعياري:

1- حساب المتوسط الحسابي : $\bar{x} = \frac{\sum x}{n}$

$$\bar{x} = \frac{5+6+7+8+9}{5} = 7$$

2- تحسب الانحرافات عن المتوسط الحسابي $x - \bar{x}$

3- نربع الانحرافات $(x - \bar{x})^2$

4- بتقسيم مجموع مربعات الانحرافات عن المتوسط الحسابي على عدد القياسات نحصل على التباين:

$$\sigma_x^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

5- بأخذ الجذر التربيعي للتباين نحصل على الانحراف المعياري

$$\sigma_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

$(x-\bar{x})^2$	الانحرافات $x-\bar{x}$	القياسات (x)
4	2- = 7-5	5
1	1-=7-6	6
0	0=7-7	7
1	1+=7-8	8
4	2+=7-9	9

$\sum(x-\bar{x})^2$ 10

ومنه نجد أن

$$\sigma_x^2 = \frac{\sum(x-\bar{x})^2}{n} = \frac{10}{4} = 2.5$$

الانحراف المعياري =

$$\sigma_x = \sqrt{2.5} = 1.581$$

2- طريقة الدرجات الخام:

يمكن حساب التباين والانحراف المعياري باستخدام القياسات نفسها وذلك باستخدام علاقة الآتية:

$$\sigma^2 = \frac{\sum X^2}{n} - \bar{X}^2$$

مثال

لتكن لدينا درجات مجموعة من الطلاب كما يلي:

الدرجات xi	x_i^2
5	25
6	36
7	49
8	64
9	81
$\sum x = 35$	$\sum x^2 = 255$

$$\sigma^2 = \frac{\sum X^2}{n} - \bar{X}^2 = \frac{255}{5} - 7^2 = 51 - 49 = 2$$

$$\sigma = \sqrt{2} = 1.41421$$

- حساب التباين والانحراف المعياري في بيانات التوزيع التكراري (بيانات مرتّبة) :

1- حساب التباين والانحراف المعياري بطريقة الانحر افات عن المتوسط الحسابي الحقيقي.

لتكن لدينا الدرجات التالية:

قيم الترتيب Xi	12	10	8	6	4	3
التكرارات ni	1	2	4	4	3	1

لتسهيل الحسابات ننشئ الجدول المساعد التالي:

قيم X	تكرار n	x.n	$x-\bar{x}$	$(x-\bar{x})^2$	$n_i\,(x-\bar{x})^2$
3	1	3	-1.86= 6.86-3	14.90	14.90
4	3	12	2.86-	8.17	24.51
6	4	22	0.86-	0.74	2.92
8	4	32	1.14+	1.30	5.16
10	2	20	3.14+	9.85	19.70
12	1	12	5.14+	26.42	26.40
المجموع	$\sum n = 15$	**103** $\sum x.n$			**93.59** $\sum n(x-\bar{x})^2$

- حساب المتوسط الحسابي:

$$\bar{x} = \frac{\sum nx_i}{\sum n_i} = \frac{103}{15} = 6.86$$

- حساب التباين:

$$\sigma^2 = \frac{\sum n(x-\bar{x})^2}{\sum n} = \frac{93.59}{15} = 6.239333$$

- حساب الانحراف المعياري :

$$\sigma = \sqrt{6.239333} = 2.49786$$

- حساب التباين بطريقة الدرجات الخام لبيانات مرتّبة:

مثال

لنعود إلى معطيات المثال السابق:

nx^2	x^2	n	X
9	9	1	3
48	16	3	2
144	36	4	6
256	64	4	8
200	100	2	10
144	141	1	12
$\sum nx_i^2 = 801$		$\sum n_i = 15$	

ومنه يكون لدينا:

$$\sigma_x^2 = \frac{\sum nx^2}{\sum n} - \bar{x}^2 \qquad \sigma_x^2 = \frac{801}{15} - (6.86)^2 = 6.3$$

ومنه الانحراف المعياري $\sigma_x = \sqrt{6.3} = 2.497$

- خطوات حساب التباين والانحراف المعياري لبيانات مبوّبة:

- تبويب البيانات في فئات وحصر التكرارات.
- حساب مراكز الفئات x'_i
- حساب المتوسط الحسابي $\bar{x} = \frac{\sum n x'_i}{\sum n_i}$
- نحسب انحرافات مراكز الفئات عن المتوسط الحسابي $x'_i - \bar{x}$.
- نربع الانحرافات $(x'_i - \bar{x})^2$
- نضرب مربعات الانحرافات في التكرارات المقابلة لها $n(x'_i - \bar{x})^2$
- بتقسيم مجموع جداء $\sum n_i (x'_i - \bar{x})^2$ على مجموع التكرارات نحصل على التباين:

$$\sigma_x^2 = \frac{\sum n_i (x'_i - \bar{x})^2}{\sum n_i}$$

- وبأخذ الجذر التربيعي نحصل على الانحراف المعياري أي:

$$\sigma_x = \sqrt{\frac{\sum n_i (x'_i - \bar{x})^2}{\sum n_i}}$$

مثال

يبين الجدول التالي دخل 18 أسرة ألف ل.س:

الفئات	4-2	6-4	8-6	10-8	12-10	14-12	16-14	المجموع
التكرارات	1	2	4	5	3	2	1	**18**

أوجد التباين والانحراف المعياري بطريقة الانحرافات عن الوسط الحسابي:

الحل: لتسيير الحسابات نكون الجدول المساعد:

الجدول المساعد لحساب التباين والانحراف المعياري:

طريقة تربيع الانحرافات

طريقة الدرجات الخام

الفئات	n	x'_i	$n x'_i$	$x'-\bar{x}$	$(x'-\bar{x})^2$	$n(x'-\bar{x})^2$	x'^2_i	$x'^2_i n$
4-2	**1**	3	3	=8.89-3 5.89-	34.6921	34.6921	9	9
6-4	**2**	5	10	3.89-	15.1321	30.2642	25	50
8-6	**4**	7	28	1.89-	3.5321	14.2884	49	196
10-8	**5**	9	45	0.11+	0.0121	0.0605	81	405
12-10	**3**	11	33	2.11+	4.4521	13.3563	121	363
14-12	**2**	13	26	4.11+	16.8921	33.7842	169	338
16-14	**1**	15	15	6.11+	37.3321	37.3321	225	225
المجموع	**18**	-	160			163.7778		1586

$$\bar{x} = \frac{\sum x_i' n_i}{\sum n_i} = \frac{160}{18} = 8.89$$

التباين بطريقة الانحرافات:

$$\sigma_x^2 = \frac{\sum n_i (x_i' - \bar{x})^2}{\sum n_i} = \frac{163.7778}{18} = 9.0987$$

ومنه الانحراف المعياري:

$$\sigma_x = \sqrt{9.0987} = 3.017$$

التباين بطريقة الدرجات الخام:

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

$$\sigma_x^2 = \frac{1586}{18} - (8.89)^2 = 9.0987$$

- الانحراف المعياري : $\sigma_x = \sqrt{9.0987} = 3.017$

3-5 تصحيح الانحراف المعياري من خطأ التجميع: تصحيح شبرد

لقد اعتمدنا في تطبيق الحسابات في البيانات المبوّبة علاقة تستند على أن جميع الدرجات (التكرار) في فئة ما تتركز في منتصف الفئة، فإن الخطأ الناشئ عن هذا الافتراض والذي يسمى بخطأ التجميع Grouping error يكون كبيراً إذا كان طول الفئة كبيراً. وهنا يجب تصحيح هذا الخطأ باستخدام معادلة تصحيح شبرد على النحو التالي :

الانحراف المعياري بعد تصحيحه من خطأ التجميع.

$$C_S = \sqrt{\sigma_x^2 - \frac{l^2}{12}}$$

حيث أن σ^2 : التباين المحسوب من البيانات المبوّبة ℓ : طول الفئة

وقد وجد أنه عندما تكون طول الفئة L مساوية 0.49 انحراف معياري فإن معادلة شبرد تعطي فرقاً قدره 0.01 بين قيمتي الانحراف المعياري بعد وقبل تصحيحه.

أما إذا كانت سعة الفئة حوالي نصف قيمة الانحراف المعياري أي ($0.49\sigma_x$) كما ذكر وكانت العيّنة كبيرة، وإذا كان المدى حوالي ستة انحرافات معيارية فإنه يكون لدينا 12 فئة، **وإذن يجب أن يكون عدد الفئات 12 فئة هي الحد** الأدنى لحساب الانحراف المعياري بدقة في حالة العيّنات **الكبيرة فإذا كان لدينا أقل من 12 فئة يجب إجراء تصحيح شبرد لمزيد من الدقة.**

تفسير الانحراف المعياري:

لقد رأينا سابقاً أن **تشتت مجموعة من القياسات يكون صغيراً** تجمعت القياسات بدقة أكبر حول المتوسط الحسابي، **ويكون كبيراً إذا انتشرت القياسات انتشاراً واسعاً حول المتوسط**، ولذا يمكن القول: أنه إذا كان الانحراف المعياري لمجموعة من القياسات صغيراً تميل القياسات إلى التراكم حول المتوسط. وإذا كان كبيراً تنتشر القياسات انتشاراً واسعاً حول المتوسط الحسابي .

وقد يتساءل البعض عما يقصد بانحراف معياري صغير أو انحراف معياري كبير ولتوضيح ذلك يجب أن نذكر نظرية تشيبسيثيف (1821-1894) وهي أنه تقع $\left(1-\frac{1}{k^2}\right)*100$ في المائة من مجموع القياسات في مدى مقداره ك انحراف معياري عن متوسطها.

إذا كان K=2 فإنه يمكن القول بأنه تقع) $\left(1-\frac{1}{2^2}\right)*100=75\%$ على الأقل من أي مجموعة بيانات في مدى قدره انحرافان معياريان عن المتوسط.

ففي مثالنا السابق نجد أن 95% على الأقل من القياسات تقع بين $\bar{x}\mp 2\sigma_x$

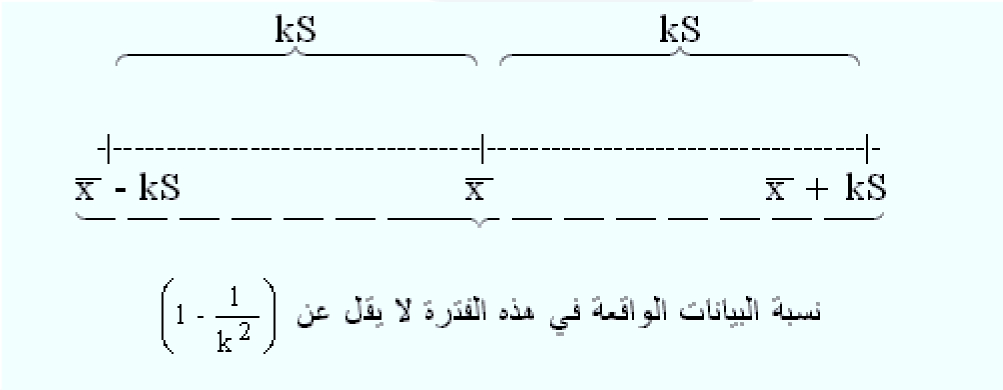
نظرية (متراجحة) تشيبيشيف Chebyshev Inequality:

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.

نظرية

إذا كان لدينا عينة من البيانات متوسطها $\bar{x}$ و انحرافها المعياري s فإن نسبة البيانات الو اقعة في الفترة $(\bar{x}-ks,\bar{x}+ks)$ لايقل عن $\left(1-\frac{1}{k^2}\right)$ حيث أن $k>1$.

الشكل التالي يوضح فكرة نظرية تشيبيشيف.



ملاحظة:

1. في بعض الأحيان نكتب الفترة $(\bar{x}-ks\,,\bar{x}+ks)$ على الصورة $\bar{x}\pm ks$.
2. **تطبق نظرية تشيبيشيف للفترات التي منتصفها (مركزها) هو المتوسط $\bar{x}$.**
3. **تستخدم نظرية تشيبيشيف بطريقتين(في كلا الحالتين لابد من معرفة قيمة k):**

أ- **تحديد النسبة (التقريبية) لعدد البيانات الو اقعة في فترة معينة.**

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

مثال

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:

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

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

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

$k = 11/5 \quad \Leftrightarrow$

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

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

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

مثال

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:

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

$$\Leftrightarrow \quad \frac{1}{k^2} = 0.25$$

$$\Leftrightarrow \quad k = \sqrt{\frac{1}{0.25}}$$

$$\Leftrightarrow \quad k = 2$$

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

$$(\bar{x} - ks, \bar{x} + ks) = (7 - 2\times5, 7 + 2\times5)$$

$$= (7 - 10, 7 + 10)$$

$$= (-3, 17)$$

خواص التباين

1- إذا كانت جميع القياسات xi متساوية وتساوي K فإن تباينها يساوي الصفر أي:

$$\sigma_k^2 = \frac{\sum_{i=1}^{n}(\chi - k)^2}{n} = 0$$

لتكن لدينا القياسات الاتية : X،5،5،5،5،5،5،5

$$\bar{X} = \frac{\sum X}{n} = \frac{35}{7} = 5$$

$$\sigma^2 = \frac{\sum x^2}{n} - \bar{X}^2 = \frac{175}{7} - 5^2 = 25 - 25 = 0$$

2- عند إضافة أو (تقديم) عدد ثابت مقداره k إلى كل قيمة من القياسات فإن قيمة تباين القياسات لا تتغير أي أن تباين القياسات الجديدة = التباين للقيم الأصلية أي أننا إذا طرحنا k من كل قياس من القياسات فإننا نحصل على قياسات جديدة yi أي أن: $yi = xi - k$ وبالتالي نحسب متوسط القياسات الجديدة أي:

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{\Sigma(xi \pm k)}{n}$$

ومنه نجد أن التباين الجديد يساوي:

$$\sigma^2 y = \frac{\sum_{i=1}^{n}(y_i \pm \bar{y})^2}{n} = \frac{\sum_{i=1}^{n}(xi - k - \bar{x} + k)^2}{n} = \frac{\sum_{i=1}^{n}(xi - \bar{x})^2}{n} = \sigma^2 x$$

وفي حالة البيانات المبوّبة:

$$\sigma^2 y = \frac{\sum_{i=1}^{n} ni(y_i - \bar{y})^2}{\Sigma ni} = \frac{\sum_{i=1}^{n}(xi - k - \bar{x} + k)^2}{\Sigma ni} = \frac{\sum_{i=1}^{n} ni(xi - \bar{x})^2}{\Sigma ni} = \sigma^2 x$$

.

3 . إذا قسمنا (أو ضربنا) كل قياس من القياسات xi بعدد أو على عدد ثابت مقداره k فإن قيمة التباين للقياسات الجديدة تتناقص أو تتضاعف بمقدار $\frac{1}{x^2}$ أو k^2 مرة أي أن:

$$y_i = \frac{x_i}{n} \Rightarrow \bar{y} = \frac{\bar{x}}{k}$$

وبذلك نجد أن:

$$\sigma^2 y = \frac{\sum_{i=1}^{n} ni\,(y_i - \bar{y})^2}{\sum_{i=1}^{n} ni} = \frac{\sum_{i=1}^{n} ni\,(\frac{xi}{k} - \frac{\bar{x}}{k})^2}{\sum_{i=1}^{n} ni}$$

$$= \frac{\frac{1}{k^2}\sum_{i=1}^{n} ni\,(x_i - \bar{x})^2}{\sum_{i=1}^{n} ni} = \frac{1}{k^2}\sigma^2 x$$

4. إن تباين القياسات عن وسطها الحسابي أصغر من تباينها عن أية قيمة أخرى شرط أن $A \neq \bar{x}$ أي أن:

<u>**حالة بيانات مفردة:**</u>

$$\sigma^2 x = \frac{\sum_{i=1}^{n}(xi - \bar{x})^2}{n} < \sigma^2 A = \frac{\sum_{i=1}^{n}(xi - A)^2}{n}$$

<u>**حالة بيانات مرتّبة:**</u>

$$\sigma^2 x = \frac{\sum_{i=1}^{n} ni(xi - \bar{x})^2}{n} < \sigma^2 A = \frac{\sum_{i=1}^{n} ni(xi - A)^2}{n}$$

<u>**حالة بيانات مبوّبة:**</u>

$$\sigma^2 x = \frac{\sum_{i=1}^{n} ni\,(x'i - x)^2}{\sum_{i=1}^{n} ni} < \sigma^2 A = \frac{\sum_{i=1}^{n} ni\,(x'i - A)^2}{\sum_{i=1}^{n} ni}$$

حيث أن: xi مركز الفئة i.   n : تكرار الفئة i. Σni : المجموع الكلي للتكرارات.

الانحراف المعياري Standard deviation

يعرف الانحراف المعياري بأنه الجذر التربيعي للتباين الموجب أي أنه:

* خصائص الانحراف المعياري:

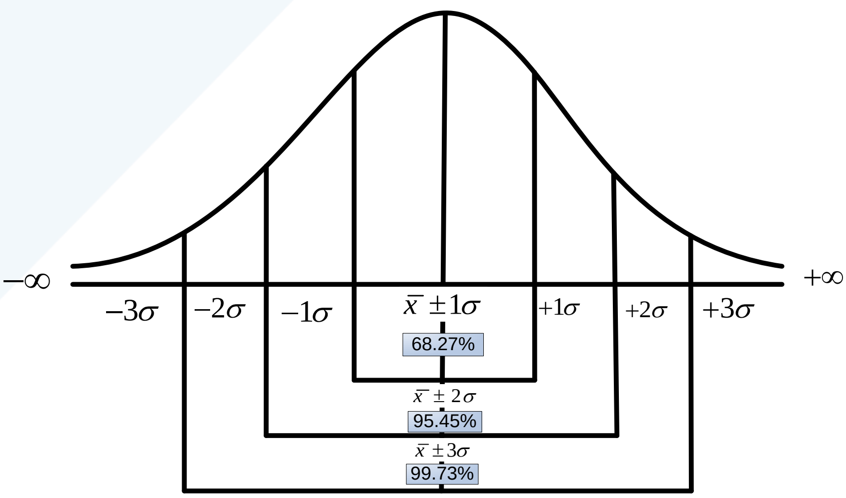
- يتمتع بجميع خصائص التباين.
- يشكل الوسط والانحراف المعياري المقياسين الأساسيين في النظرية الإحصائية فهما المفتاح الرئيسي لتوزيع هام جداً يعرف باسم التوزيع الاعتدالي وهو توزيع متماثل يتحدّد بوسطه الحسابي وانحرافه المعياري.
- يستعمل الانحراف المعياري كمقياس للتشتت المطلق بشكل واسع في الإحصاء وكمقياس للثقة ويستخدم في الارتباط والاستدلال الإحصائي.
- في حالة التوزيع المتماثل حيث الوسط الحسابي يقع في منتصف منحنى التوزيع الطبيعي أي أن المدى ما بين الوسط الحسابي ووحدات من الانحراف المعياري يحصر فيه نسبة معينة من قيم التوزيع على النحو التالي:

✓ 68.27% من الحالات تقع على بعد انحراف معياري واحد على كل جانب من الوسط الحسابي $\bar{x} \mp 1\sigma_x$

✓ 95.44% من الحالات تقع على بعد انحرافين معياريين على كل جانب من الوسط الحسابي $\bar{x} \mp 2\sigma_x$.

✓ 99.73% من الحالات تقع على بعد ثلاثة انحرافات معيارية على كل جانب من الوسط الحسابي $\bar{x} \mp 3\sigma_x$.

ويمكن توضيح ذلك بالشكل المثالي:



شكل : منحنى التوزيع الطبيعي المعياري

5-يستفاد من الانحراف المعياري في عمليات التبويب إذ يفترض أن القياسات متناظرة فإن معظمها لا يخرج عن المجال.

$$\left[\bar{x}-3\sigma_x \;\;;\;\; \bar{x}+3\sigma_x\right]$$

ويتم تقسيم هذا المجال إلى مجالات جزئية متساوية طول كل منها يساوي σ_x ، أو $\frac{2}{3}\sigma_x$ أو $\frac{1}{2}\sigma_x$ ، وذلك حسبما تكون قيمة معامل الاختلاف.

6- يستخدم الانحراف المعياري لمعرفة عدد ونسبة القيم التي توجد في مجال محدّد حول المتوسط الحسابي، أو عدد ونسبة القيم الواقعة خارج ذلك المجال.

مثال

درس باحث نتائج 15000 طالب في كلية التربية في مقرر الإحصاء نوجد أن متوسط الدرجات $\bar{x}$ = 70 درجة وبانحراف معياري $\sigma_x = 8$ درجات فإذا فرضنا أن درجات الطلاب تتوزع طبيعياً أوجد:

- **عدد الطلاب الذين تتراوح درجاتهم بين 62 و 78 درجة.**
- **عدد الطلاب الذين تتراوح درجاتهم بين 46-54 درجة.**
- **عدد الطلاب الذين درجاتهم أكثر من 86 درجة.**

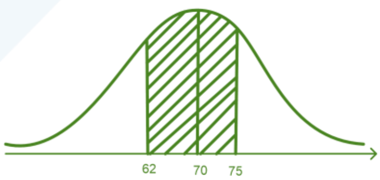
الحل:

من المعروف أن المساحة تحت منحني التوزيع الطبيعي تمثل عدد الطلاب جميعاً وبالتالي فإن الوسط الحسابي يقسم هذه المساحة إلى قسمين متساويين أي 50% من الطلاب درجاتهم أقل من 70 درجة و 50% من الطلاب درجاتهم أكبر من 70 درجة، وبالتالي نحدّد النسبة المحصورة بين 62-70 و 70-78 فنحصل على نسبة هؤلاء الطلاب وبضرب هذه النسبة بعدد الطلاب الكلي نحصل على عدد الطلاب الذين درجاتهم بين 62 و 78 وذلك كما يلي:

- نحدّد الفرق بين الوسط الحسابي ووحدات من الانحراف المعياري.

$$\bar{x}-1\sigma_x = 70-(8\times1)=62$$
$$\bar{x}+1\sigma_x = 70+(8\times1)=78$$

إذاً المجال المطلوب هو: $\bar{x}\mp1\sigma_x$



وبما أن الدرجات تتوزع طبيعياً إذاً المجال $\bar{x}\mp1\sigma_x$ يحصر 68.27% من القياسات وبالتالي عدد الطلاب ضمن هذا

المجال يساوي:

عدد الطلاب = $\frac{15000\times 68.27}{100}=10241$ طالباً

2- إن درجات 46 -54 هي أصغر من المتوسط الحسابي لذلك لحساب النسبة المحصورة بين هاتين الدرجتين، يجب حساب النسبة التي يحصرها كل منهما مع الوسط الحسابي ثم تقديم هاتين النسبتين

$\bar{x}-2\sigma_x$

$70-(2\times 8)=54$

الدرجة 54 تقع ضمن المجال $\bar{x}-2\sigma_x$ وضمن 50% من المنحني وهذا المجال يحصر 95.45% أي ضمن $\frac{95.45}{2}$ = 47.725 %

أما الدرجة 46 تقع ضمن المجال $\bar{x}-3\sigma_x$ وهذا المجال يحصر 99.73% أي ضمن $\frac{99.73}{2}$ = 49.865%

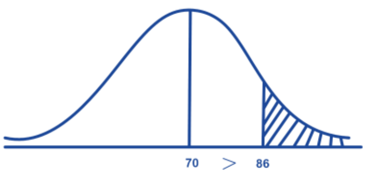
وبالتالي تكون النسبة المحصورة بين 46-54 تساوي

% 2.14 = 47.725 - 49.865

ومنه نجد أن عدد الطلاب = $\frac{15000\times 2.14}{100}$ = 362 طالباً

3- إن النسبة التي تحصرها الدرجة 86 فما فوق تساوي نصف المساحة مطروحاً منها المسافة بين المتوسط الحسابي والدرجة 86 أي :

$\Leftarrow \bar{x}+2\sigma_x$ 70 + 8×2 = 86 درجة وذلك كما في الشكل التالي:



إن النسبة المحصورة من $\bar{x}+2\sigma_x$ هي 47.725% فإن النسبة المتممة إلى 50% تساوي 50 – 47.725 = 2.275 %

ومنه نجد عدد الطلاب الذين درجاتهم أكثر من 86 درجة يساوي : عدد الطلاب = $\frac{15000\times 2.275}{100}=342$ طالب

• -المقاييس النسبية للتشتت: معاملات الاختلاف

إن جميع مقاييس التشتّت السابقة الذكر تكون قيمتها معطاة بدلالة وحدات قياس المتغيّر، فهي صالحة إذن لمقارنة المجموعات التي لها نفس الوحدات وبشرط أن تكون متوسطاتها متقاربة، لأن التشتّت مقياس يعتمد على الانحراف عن المتوسط الحسابي. ومن هنا نشأت الحاجة إلى مقاييس نسبية للتشتت لمقارنة توزيعات ليس لها نفس الوحدات أو متوسطاتها مختلفة اختلافاً كبيراً لأن مقاييس التشتّت النسبي تكون خالية من وحدات القياس ومن أهمها:

1- معامل الاختلاف ويرمزله بـC.v ويعطى بالعلاقة التالية:

$$100\times\frac{\text{حراف المعياري}}{\text{توسط الحسابي}} = C.V\% = \frac{\sigma_x}{\bar{x}}\times 100$$

مثال

لتكن نتائج الامتحانات النهائية لمقرر الإحصاء والرياضيات لطلاب السنة الأولى في كلية الادارة على النحو الآتي:

	الإحصاء	الرياضيات
الوسط الحسابي	78	73
الانحراف المعياري	8	7.6

في أي المقررين كان تشتت الدرجات أكثر؟

الحل:

تحسب معامل الاختلاف لكل من المقررين فنجد أن:

معامل الاختلاف للإحصاء = $\frac{8}{70}\times 100 = 10.25\%$

معامل الاختلاف للرياضيات = $\frac{7.6}{73}\times 100 = 10.41\%$ أي أن التشتّت لدرجات الفلسفة كان أكثر منه في درجات الإحصاء.

2- معامل الاختلاف الربيعي:

يتعذر في بعض التوزيعات التكرارية حساب المتوسط الحسابي والانحراف المعياري كما في التوزيعات المفتوحة، وقد لا يكون هذان المقياسان أنسب المقاييس في بعض التوزيعات لذا نلجأ إلى مقاييس تشتت نسبية أخرى.

ومن هذه المقاييس معامل الاختلاف الربيعي ويحسب بالعلاقة التالية:

معامل الاختلاف الربيعي =

$$Cq_v = \frac{q_3 - q_1}{q_3 + q_1}.100$$

كما يصلح هذا المقياس لجميع التوزيعات .

مثال

إذا كان الربيع الأول $q_1 = 9$ والربيع الثالث $q_3 = 13.555$ أوجد معامل التشتّت الربيعي لهذا السلسلة؟ معامل الاختلاف الربيعي = $\frac{9 - 13.555}{9 + 13.555} \times 100 = 20.19\%$

$$Cq_v = \frac{q_3 - q_1}{2.Me}.100$$

3- معامل الاختلاف المئنيني = $\frac{\text{ئ ذ ين 99} - \text{الم ئ ذ ين 1}}{2}$

$$Cq_v = \frac{q_{99} - p_1}{2}$$

هام جدا جدا

8-5-العلاقات التجريبية بين مقاييس التشتّت:

مقاييس التشتّت كما أوضحنا أن هذه المقاييس تصف التغير والتباعد بين البيانات وأنها تعتمد على الانحرافات فيما بينها، أو على الانحرافات عن الوسط الحسابي، بذلك فإننا نتوقع أن لا تتأثر هذه المقاييس بإضافة أية قيمة موجبة كانت، أم سلبية الى البيانات الأولية. هذا، وتوجد علاقتان تجريبيتان تربطان بين مقاييس التشتّت هما:

في التوزيعات متوسطة الالتواء هناك علاقات اعتبارية (تقريبية) بين مقاييس التشتت السابقة كالآتي :

الانحراف المتوسط = $\frac{٤}{٥}$ × الانحراف المعياري $M.D=\frac{4}{5}s$ أو الانحراف المعياري = $\frac{٥}{٤}$ × الانحراف المتوسط $s=\frac{5}{4}M.D$

الانحراف الربيعي = $\frac{٢}{٣}$ × الانحراف المعياري $Q=\frac{2}{3}s$ أو الانحراف المعياري = $\frac{٣}{٢}$ × الانحراف الربيعي $s=\frac{3}{2}Q$

الانحراف المتوسط = $\frac{٦}{٥}$ × الانحراف الربيعي $M.D=\frac{6}{5}Q$ أو الانحراف الربيعي = $\frac{٥}{٦}$ × الانحراف المتوسط $Q=\frac{5}{6}M.D$

هذه العلاقات الاعتبارية تمكننا من حساب قيم تقريبية لبقية مقاييس التشتت متى عُلم أحدها [وذلك في حالة صلاحيتها .. أي في حالة التوزيعات التكرارية متوسطة الالتواء]

فمثلاً : • إذا كان $s = 30$ فإن : $Q=\frac{2}{3}s=\frac{2}{3}\times 30=\underline{\underline{20}}$ ، $M.D=\frac{4}{5}s=\frac{4}{5}\times 30=\underline{24}$

• وإذا كان $Q = 20$ فإن : $s=\frac{3}{2}Q=\frac{3}{2}\times 20=\underline{\underline{30}}$ ، $M.D=\frac{6}{5}Q=\frac{6}{5}\times 20=\underline{24}$

• وإذا كان $M.D = 24$ فإن : $Q=\frac{5}{6}M.D=\frac{5}{6}\times 24=\underline{20}$ ، $s=\frac{5}{4}M.D=\frac{5}{4}\times 24=\underline{30}$

الدرجات المعيارية:

لقد لاحظنا فيما سبق أن قيمة الدرجة التي يحصل عليها طالب أو فرد في اختبار ما أو مقياس ما هي قيمة اختيارية، أي لا يكون لها معنى إلا في إطار مجموعة الدرجات التي حصل عليها أقران هذا الفرد. ولذلك فإنه من المرغوب فيه في معظم الأحيان أن نحول هذه الدرجات الخام إلى نوع آخر من الدرجات مثل الرتب المئينية حتى يمكن مقارنتها بغيرها من الدرجات. كما يمكن الاستفادة من المتوسط والانحراف المعياري في مقارنة درجة معينة في اختيار ما بدرجات مجموعة مرجعية في الاختبار أو المقياس نفسه، ويفصل في أغلب الأحيان إجراء عملية تمويل الدرجة الخام بحيث نأخذ في العدّ متوسط درجات المجموعة المرجعية وانحرافها المعياري أي نحول الدرجة الخام إلى انحرافات معيارية أعلى أو أدنى من المتوسط كوحدة قياس وحينئذ تسمى الدرجات المحولة الدرجات المعيارية، فالدرجات الخام تفيدنا بمعلومات عن عدد النقاط التي حصل عليها الطالب في اختبار ما، ولكنها لا تقدم أية أدلة عن مدى تفوق أو ضعف الطالب في الأداء في الاختبار، كما لا تسمح الدرجات الخام بمقارنة أدائه بأداء غيره من الطلاب.

وتحسب الدرجة المعيارية بالعلاقة الآتية: $Z = \frac{x_i - \bar{x}}{\sigma_x}$

الدرجة المعيارية = (جات الخام – المتوسط الحسابي) / (الانحراف المعياري)

مثال(5-22): نفترض أننا حصلنا المعطيات الآتية المتعلّقة باختيار بعض المقررات.

الاختيار	الدرجة الخام	المتوسط	الانحراف المعياري
احصاء	80	85	10
محاسبة	65	55	5
إدارة عامة	75	60	15

وبحساب الدرجة المعيارية لكل مقرر نجد أنها تساوي:

إحصاء = $\frac{85-80}{10} = \frac{-5}{10} = -0.5$

محاسبة = $\frac{55-65}{5} = \frac{10}{5} = +2$

إدارة عامة = $\frac{60-75}{15} = \frac{15}{15} = +1$

-خصائص الدرجة المعيارية:

- مجموع الدرجات المعيارية = $\sum_{i=1}^{n} Z_i = 0$
- متوسط توزيع الدرجات المعيارية = صفر أي $D = \frac{\sum Z}{n} = 0$
- الدرجات الخام التي تقل عن المتوسط تقابلها درجات معيارية سالبة والدرجات الخام التي تزيد عن المتوسط تقابلها درجات معيارية موجبة.
- مجموع مربعات الدرجات المعيارية = العدد الكلي للدرجات .
- أي $\sum Z^2 = n$

وهذه الخاصية تكون صحيحة فقط إذا حسبنا الانحراف المعياري باستخدام (n) في المقام بدلاً من n-1 أي:

$$\sum Z^2 = \frac{\sum (x-\bar{x})^2}{\sigma_x^2}$$
$$= \frac{1}{\sigma_x^2} \sum (x-\bar{x})^2$$
$$= \frac{n}{\sum (x-\bar{x})^2} * \sum (x-\bar{x})^2$$

$$\sum Z^2 = n$$

5- الانحراف المعياري وتباين توزيع الدرجات المعيارية يساوي الواحد الصحيح أي أن $\sigma_{Z_i}^2 = \sigma_Z = 1$

وذلك كما يلي:

$$\sigma_z^2 = \frac{\sum (Z-\bar{Z})^2}{n}$$

ولكن
$$\bar{Z}=0$$

إذاً
$$\sigma_z^2=\frac{\sum z^2}{n}$$

وبما أن
$$\sum z^2=n$$

إذن
$$\sigma_z^2=\frac{n}{n}=1$$

1- إذا حسبنا الدرجات المعيارية من عينات عشوائية فإن مدى هذه الدرجات يكون دالة لحجم العيّنة مفاده تتراوح الدرجات المعيارية للعينات الكبيرة بين

-2

(-3 و +3)

الدرجة التائية

من بين العيوب الرئيسية المعيارية (z) أنه يصعب على الشخص غير المتخصص تفسيرها، لاسيما مثلاً حصول بعض الطلاب على الدرجة المعيارية صفر، أو درجة معيارية سالبة وكذلك الحصول على درجات معيارية كسرية.

وللتخلص من هذه العيوب في تفسير الدرجات المعيارية فقد اقترح العالم ثورنديك الدرجة التائية ويمكن تعريفها بأنها مجموعة من الدرجات التي يكون متوسطها (50) وانحرافها المعياري (10) .

وتحسب بالعلاقة التالية:

$$T=50+10*z$$

مثال

بالعودة إلى معطيات المثال السابق نجد أن: الإحصاء $T=50+10\times(-0.5)=50-5=45$ وهذا يعني أن الدرجة 45 تقل عن المتوسط بمقدار نصف انحراف معياري أي تناظر درجة معيارية = -0.5 لأن

الدرجة الخام = المتوسط العام لدرجات المقرر+ الانحراف المعياري × الدرجة المعيارية

$$x_i=\bar{x}+\sigma_x*z$$
)

مثال

نعود إلى معطيات المثال السابق:

الدرجة المعيارية للغة العربية -0.5 والمتوسط العام $\bar{x}=85$ والانحراف المعياري $\sigma_x=10$ ومنه نجد:

$$\begin{aligned} x_i &= \bar{x}+\sigma_x * z \\ &= 85+10*(-0.5) \\ &= 85-5=80 \end{aligned}$$

1- معيار اختبار الاستعداد الدراسي (SAT) Scholastic Aptitudetest:

درجة SAT: $$SAT = 100\times z + 500$$

2- معيار اختبار القبول في الكليات:

College entrance examination Broad (CEEB)

درجة (CEEB) = $$CEEB = 500 + 100 * z$$

3- معيار اختبار بيان الدراسات العليا: Graduate Record examination:

- درجة GRE = $$GRE = 500 + z * 100$$

- العزوم المركزية:

تحسب العزوم المركزية منسوبة إلى الوسط الحسابي ويعرف العزم المركزي من الرتبة k وفق العلاقات الآتية:

-بيانات مفردة:

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

-بيانات مرتّبة:

$$M_k = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^k}{\sum n_i}$$

-بيانات مبوّبة:

$$M_k = \frac{\sum_{i=1}^{n} \left(x_i' - \bar{x}\right)^k}{\sum n_i}$$

حيث أن k تأخذ القيم : $K=0,1,2,3,\ldots\ldots\ldots$

خواص العزوم المركزية:

-إذا كان k=0 نجد العزم المركزي يساوي:

$$M_o = \frac{\sum \left(x_i - \bar{x}\right)^o}{n} = \frac{n}{n} = 1$$

$$= \frac{\sum n_i \left(x_i - \bar{x}\right)^o}{\sum n_i} = \frac{n}{n} = 1$$

$$= \frac{\sum n_i \left(x_i' - \bar{x}\right)^o}{\sum n_i} = 1$$

-إذا كان k=1 نجد أن:

$$M_1 = \frac{\sum \left(x_i - \bar{x}\right)^1}{n} = \frac{0}{n} = 0$$

$$= \frac{\sum n_i \left(x_i' - \bar{x}\right)^1}{\sum n_i} = \frac{0}{n} = 0$$

-إذا كان k=2 نحصل على التباين:

$$M_2 = \frac{\sum \left(x_i - \bar{x}\right)^2}{n} = \sigma^2$$

.

.

$$= \frac{\sum n_i \left(x_i' - \bar{x}\right)^2}{\sum n_i} = \sigma^2$$

-إذا كان k=3 نحصل على العزم الثالث:

$$M_3 = \frac{\sum(x_i - \bar{x})^3}{n}$$

$$= \frac{\sum n_i (x_i' - \bar{x})^3}{\sum n_i}$$

-إذا كان k=4 نحصل على العزم الرابع:

$$M_4 = \frac{\sum(x_i - \bar{x})^4}{n}$$

$$= \frac{\sum n_i (x_i' - \bar{x})^4}{\sum n_i}$$

***ملاحظة*: تميز العزم الثالث بأن قيمته يمكن أن تكون موجبة أو سالبة لان أسه مفرد.**

فإذا كانت قيمته موجبة يعني ذلك أن القياسات تميل نحو اليمين عن الوسط الحسابي أي أن التوزيع ملتو نحو اليمين. وإذا كانت قيمته سالبة يعني ذلك أن القياسات تميل نحو اليسار عن الوسط الحسابي أي أن التوزيع ملتو نحو اليسار. وإذا كانت قيمته تساوي الصفر يعني أن القياسات متناظرة حول الوسط الحسابي وأن توزيعها يكون متناظراً ولهذا يستخدم العزم الثالث كمقياس كتحديد توزع البيانات حول متوسطها الحسابي. أما العزم الرابع قيمته دائماً موجبة لان أسه زوجي ويستخدم لدراسة درجة تطاول أو تفرطح التوزيع التكراري للبيانات.

مثال

لنعود إلى بيانات درجة الطلاب في الإحصاء في كلية الاقتصاد لحساب العزوم كافة وفق الآتي:

الجدول المساعد لحساب العزوم المركزية لبيانات مبوّبة

الفئات	التكرار	x'_i	الانحرافات	العزم صفر $k=0$	العزم الأول $k=1$	العزم الثاني $k=2$	العزم الثالث $k=3$	العزم الرابع $k=4$
الفئات	ni	x'_i	$(x'_i-\bar{x})$	$n_i(x'_i-\bar{x})^0$	$n_i(x'_i-\bar{x})^1$	$n_i(x'_i-\bar{x})^2$	$n_i(x'_i-\bar{x})^3$	$n_i(x'_i-\bar{x})^4$
10-20	6	15	= -42.12	6	-252.72	10644.5664	-448349.1368	18884465.64
20-30	14	25	-32.12	14	-449.68	14443.7216	-463932.3378	14901506.69
30-40	26	53	-22.12	26	-575.12	12721.6544	281402.99	6224634.257
40-50	43	45	-12.12	43	-521.16	6316.4592	-76555.4855	927852.484
50- 60	51	55	-2.12	51	-108.12	229.2144	-485.93452	1030.1812
60- 70	47	65	+7.88	47	+370.36	2918.4368	+22997.282	18121.582
70- 80	29	75	+17.88	29	+518.52	9271.1376	+156767.9403	2963930.772
80- 90	27	85	+27.88	27	+752.76	20986.9488	+585116.1325	16313037.78
90-100	7	95	+37.88	7	+265.16	10044.2608	+380476.5991	14412453.57
Σ	250	-		250	-1906.8 +1906.8 0	87576.4	-242508.3	74810130.06

-حساب العزم صفر k=0

$$k_0 = \frac{\sum n_i \left(x_i' - \bar{x}\right)^o}{\sum n_i} = \frac{250}{250} = 1$$

-حساب العزم الأول k=1:

$$k_1 = \frac{\sum n_i \left(x_i' - \bar{x}\right)^1}{\sum n_i} = \frac{0}{250} = 0$$

-حساب العزم k=2:

$$k_2 = \frac{\sum n_i \left(x_i' - \bar{x}\right)^2}{\sum n_i} = \frac{87576.4}{250} = 350.3$$

وهي قيمة التباين.

-حساب العزم الثالث k=3:

$$k_3 = \frac{\sum n_i \left(x_i' - \bar{x}\right)^3}{\sum n_i} = \frac{-242508.3}{250} = -9709.3$$

-حساب العزم الرابع k=4:

$$k_4 = \frac{\sum n_i \left(x_i' - \bar{x}\right)^4}{\sum n_i} = \frac{73975063.7}{250} = 295900.26$$

نهاية المحاضرة رقم 1