



ENGINEERING STATISTICS

الاحصاء الهندسي

المحاضرة الرابعة



م.م. تمار ميثم عبدالوهاب

Measures of dispersion or variation

مقاييس التشتت أو الاختلاف

Measures of centering are not sufficient to describe a set of data completely. Some samples may have the same arithmetic mean despite the difference in the distribution of their data around their center. The degree of data homogeneity, such as the following two sets of data.

لا تعتبر مقاييس التمرکز كافية لوصف مجموعة من البيانات وصفاً كاملاً فقد تتساوى بعض العينات في الوسط الحسابي بالرغم من اختلاف توزيع بياناتها حول مركزها درجة تجانس البيانات، مثل المجموعتين التاليتين من البيانات.

(150، 125، 100، 50، 75)، (95، 97، 100، 103، 105)

It is noted that they have the same arithmetic mean and median, which are 100, while the two groups differ in terms of their dispersion around the center or middle. The degree to which we tend to spread digital data around a middle value is called the dispersion or distribution of data. The most important measures of dispersion are (range, variance, standard deviation, and mean deviation).

حيث يلاحظ ان لهما نفس الوسط الحسابي والوسيط وهما 100 في حين المجموعتان مختلفتان من ناحيه نشتتهما حول المركز او الوسط إن الدرجة التي نتجه بها البيانات الرقمية للانتشار حول قيمة وسطى تسمى تشتت أو توزيع البيانات .

ومن اهم مقاييس التشتت (المدى، التباين، والانحراف المعياري، الانحراف المتوسط).

Mean deviation: It is the average of the absolute deviations from the arithmetic mean and is symbolized by the symbol (M.D) and is calculated according to

الانحراف المتوسط: وهو متوسط الانحرافات المطلقة عن الوسط الحسابي ويرمز له بالرمز (M.D) ويتم حسابه وفق

for ungrouped data للبيانات غير المبوبة

$$M.D = \frac{\sum |y_i - \bar{y}|}{n}$$

Example / Find the mean deviation of the following values

جد الانحراف المتوسط للقيم التالية

$$y_i = 9, 8, 6, 5, 7$$

Sol/

y_i	9	8	6	5	7	$\sum y_i = 35; \bar{y} = 7$
$y_i - \bar{y}$	2	1	-1	-2	0	$\sum (y_i - \bar{y}) = 0$
$ y_i - \bar{y} $	2	1	1	2	0	$\sum y_i - \bar{y} = 6$

$$M.D = \frac{\sum |y_i - \bar{y}|}{n} = \frac{6}{5} = 1.2$$

for grouped data للبيانات المبوبة

$$M.D = \frac{\sum f_i |y_i - \bar{y}|}{f_i}$$

Example / Find the mean deviation of the following frequency distribution table

جد الانحراف المتوسط لجدول التوزيع التكراري الاتي

class	60-62	63-65	66-68	69-71	72-74
fi	5	18	42	27	8

Sol/

Class	Fi	yi	fi*yi
60-62	5	61	305
63-65	18	64	1152
66-68	42	67	2814
69-71	27	70	1880
72-74	8	73	584
Σ	100		6745

$$\bar{y} = \frac{\sum f_i y_i}{\sum f_i} = \frac{6745}{100} = 67.45$$

$ y_i - \bar{y} $	$f_i y_i - \bar{y} $
6.45	32.25
3.45	62.10
0.45	18.90
2.55	68.85
5.55	44.40
	226.50

$$M.D = \frac{\sum f_i |y_i - \bar{y}|}{\sum f_i} = \frac{226.50}{100} = 2.265$$

Variance and standard deviation

We notice that the sum of the deviations of the sample elements (items) from their arithmetic mean ($\sum |y_i - \bar{y}| = 0$) because some of the deviations are positive and others are negative. To overcome this problem, it was treated by taking the absolute values of the deviations in (the average deviation), and it can be treated in another way, which is squaring the values of the deviations to obtain the sum of the squares of the deviations (sum squares) and it is symbolized by (SS).

التباين والانحراف القياسي

نلاحظ ان مجموع انحرافات عناصر (مفردات) العينة عن وسطها الحسابي ($\sum |y_i - \bar{y}| = 0$) وذلك لان قسم من الانحرافات موجبة و اخرى سالبة، وللتغلب على هذه المشكلة فقد تم معالجتها باخذ القيم المطلقة للانحرافات في (الانحراف المتوسط)، ويمكن معالجتها بطريقة اخرى وهي تربيع قيم الانحرافات للحصول على مجموع مربعات الانحرافات (sum squares) و يرمز لها (SS).

$$SS = \sum (y_i - \bar{y})^2$$

In order to take into account the sample size so that we can compare samples of different sizes, SS is divided by the degrees of freedom (n-1), which leads to obtaining the variance (Variance S^2).

ولكي نأخذ في الاعتبار حجم العينة حتى يمكن المقارنة بين العينات المختلفة الاحجام يتم تقسيم SS على درجات الحرية (n-1) مما يؤدي الى الحصول على التباين (Variance S^2)

للبيانات المبوبة for grouped data

$$S^2 = \frac{\sum (yi - \bar{y})^2}{n - 1} = \frac{\sum yi^2 - \frac{(\sum yi)^2}{n}}{n - 1}$$

In the case of society, it is calculated according to

اما في حالة المجتمع فتحسب وفق

$$\sigma^2 = \frac{\sum (yi - \mu)^2}{N}$$

μ = In the middle of the computing community (وسط المجتمع الحسابي)

N = Number of community vocabulary (عدد مفردات المجتمع)

The reason for taking $(n-1)$ in the case of a sample is that when a sample is drawn, $(n-1)$ of the observations are free values, while the remaining observation (the last) must complete its deviation from the arithmetic mean to zero, so the number of free values in any sample is $(n-1)$, which is called degrees of freedom. In order to return the units of measurement for the sample items to their origin, the square root of the variance must be taken to obtain $(S = \sqrt{S^2})$, which is called the standard deviation.

وسبب اخذ ($n-1$) في حالة العينة هو انه عند سحب عينة فان ($n-1$) من المشاهدات هي قيم حرة اما المشاهدة المتبقية (الأخيرة لابد ان يكمل انحرافها مجموع الانحرافات عن الوسط الحسابي الى الصفر، لذلك فان عدد القيم الحرة في أي عينة هو ($n-1$) وهو ما سمي بدرجات الحرية

ولاجل إعادة وحدات القياس لمفردات العينة الى اصلها يجب ان يؤخذ الجذر التربيعي للتباين للحصول على $(S = \sqrt{S^2})$ وهو ما يسمى بالانحراف القياسي.

Derivation الاشتقاق

$$SS = \Sigma(yi - \bar{y})^2 = \Sigma(yi^2 - 2yi\bar{y} + \bar{y}^2)$$

$$= \Sigma yi^2 - 2\Sigma yi * \left(\frac{\Sigma yi}{n}\right) + n\left(\frac{\Sigma yi}{n}\right)^2$$

$$= \Sigma yi^2 - 2\frac{(\Sigma yi)^2}{n} + n\left(\frac{\Sigma yi}{n}\right)^2$$

$$SS = \Sigma yi^2 - \frac{(\Sigma yi)^2}{n}$$

$$S^2 = \frac{SS}{n-1} = \frac{\Sigma(yi - \bar{y})^2}{n-1} = \frac{\Sigma yi^2 - \frac{(\Sigma yi)^2}{n}}{n-1}$$

$$\rightarrow S = \sqrt{S^2} = \sqrt{\frac{\Sigma(yi - \bar{y})^2}{n-1}} = \sqrt{\frac{\Sigma yi^2 - \frac{(\Sigma yi)^2}{n}}{n-1}}$$

for grouped data للبيانات المبوبة

$$S = \sqrt{S^2} = \sqrt{\frac{\Sigma fi(yi - \bar{y})^2}{\Sigma fi - 1}} = \sqrt{\frac{\Sigma fiyi^2 - \frac{(\Sigma fiyi)^2}{\Sigma fi}}{\Sigma fi - 1}}$$

Example/ Find the standard deviation of the following data using the shorthand method.

جد الانحراف القياسي للبيانات التالية باستخدام الطريقة المختصرة

$y_i = 9, 8, 6, 5, 7$

Sol/

y_i	9	8	6	5	7	$\Sigma y_i = 35$
y_i^2	81	64	36	25	49	$\Sigma y_i^2 = 255$

$$S = \sqrt{\frac{\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}}{n - 1}} = \sqrt{\frac{255 - \frac{(35)^2}{5}}{5 - 1}} = \sqrt{\frac{255 - 245}{4}} = \sqrt{\frac{10}{4}} = 1.58$$

Example/ Find the standard deviation of the following data:

جد الانحراف القياسي للبيانات التالية

Class	Fi
60-62	5
63-65	18
66-68	42
69-71	27
72-74	8
Σ	100

Sol/

Class	fi	Yi	$yi - \bar{y}$	$(yi - \bar{y})^2$	$fi * (yi - \bar{y})^2$
60-62	5	61	-6.45	41.6025	208.0125
63-65	18	64	-3.45	11.9025	214.2450
66-68	42	67	-0.45	0.2025	8.5050
69-71	27	70	2.55	6.5025	175.5675
72-74	8	73	5.55	30.8025	246.4200
Σ	100	$\bar{y} = 67.45$			$\Sigma=852.7500$

The most important properties of variance and standard deviation

اهم خواص التباين والانحراف القياسي

1. When adding or subtracting a fixed number K from the values of observations, it does not affect the value of variance and standard deviation of those observations.

عند اضافة او طرح عدد ثابت K من قيم المشاهدات، فانه لا يؤثر على قيمة التباين و الانحراف القياسي لتلك المشاهدات.

If $X_i = y_i + k$ or $X_i = y_i - k$ then $S_x^2 = S_y^2$ or $S_x = S_y$

Example/

$$X_i = y_i + 3 = 11, 6, 5, 15, 13$$

$$y_i = 8, 3, 2, 12, 10 \quad \Sigma y_i = 35$$

$$y_i^2 = 64, 9, 4, 144, 100 \quad \Sigma y_i^2 = 321$$

$$S = \sqrt{\frac{\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}}{n-1}} = \sqrt{\frac{321 - \frac{(35)^2}{5}}{5-1}} = \sqrt{\frac{321 - 245}{4}} = \sqrt{19} = 4.36$$

$$X_i = 11, 6, 5, 15, 13 \quad \Sigma X_i = 50$$

$$X_i^2 = 121, 36, 25, 225, 169 \quad \Sigma X_i^2 = 576$$

$$S_X = \sqrt{\frac{\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}}{n-1}} = \sqrt{\frac{576 - \frac{(50)^2}{5}}{5-1}} = \sqrt{\frac{576 - 500}{4}} = \sqrt{19} = 4.36$$

2. If each of the observation values is multiplied by a constant number (k), then
- The variance of the resulting values = the variance of the original values * the square of k .
 - The standard deviation of the resulting values = the standard deviation of the original values * k .

إذا ضربت كل قيم من قيم المشاهدات بعدد ثابت (k) فإن

تباين القيم الناتجة = تباين القيم الأصلية * مربع k

الانحراف القياسي للقيم الناتجة = الانحراف القياسي للقيم الأصلية * k