مقاييس التشتت

- المدى.
- المدى الربيعي
- الانحراف المتوسط
- التباين والانحراف المعياري.

عند مقارنة مجموعتين من البيانات يمكن استخدام أحد مقاييس النزعة المركزية كـ $\overline{X}$ وقد نجد تساوي بين متوسطي المجموعتين مع أنه وربما يوجد اختلاف كبير بين المجموعتين من حيث مدى تقارب وتباعد البيانات فيما بينها

مثال: اذا كان هناك فوجين حيث علاماتهم كالتالي:

الفوج الأول: $3 - 17 - 10 - 2 - 18 \leftarrow \overline{X} = 10$

الفوج 02: $9 - 9.5 - 10 - 10.5 - 11 \leftarrow \overline{X} = 10$

- من الوهلة الأولى نجد تساوي بين المتوسطين ولكن بعد ملاحظة البيانات نجد أن نقاط الفوج 1 متباعدة وغير متجانسة مقارنة بنقاط الفوج 2. وبهذا نحتاج الى مقاييس تقيس لنا مدى تشتت البيانات عن بعضها او عن متوسطها الحسابي: نحتاج الى مقاييس التشتت.

المدى: étendu

يرمز له بـ E وهو الفرق بين أعلى قيمة وأدنى قيمة للمتغير الإحصائي. $E = X_{max} - X_{min}$

أما في حالة فئات فالمدى هو الفرق بين الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة الأولى، من

مزايا المدى:

- بسيط وسهل الحساب.
- فعال في حالة أحوال الطقس
- يستخدم في مقاربة الجودة

عيوب المدى:

- يعتمد فقط على قيمتين ويهمل بقية القيم.
- يتأثر بالقيم المتطرفة

حسب المثال السابق

$$E_{ف_1} = X_{max} - X_{min}$$
$$= 18 - 2 = 16$$

$$E_{ف_2} = 11 - 9 = 02$$

المدى الربيعي:

يعتمد المدى على قيمتين متطرفتين لذلك يلجأ الإحصائيون إلى استخدام المدى الربيعي.

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

من مزاياه أنه لا يتأثر بالقيم الشاذة.

- يستعمل في حالة وجود قيم شاذة.
- بسيط وسهل الحساب

Scanné avec CamScanner

عيوبه :
- لا يأخذ كل القيم بعين الاعتبار
مثال: 1 - 2 - 3 - 4 - 7 - 10 - 12 - 15 - 20

$$RQ_1 = \frac{n+1}{4} = 2.5 \rightarrow Q_1 = 2.5$$

$$RQ_3 = \frac{(n+1)\times 3}{4} = 7.5 \rightarrow Q_3 = 13.5 = \frac{12+15}{2}$$

$$EQ = \frac{Q_3 - Q_1}{2} = \frac{13.5 - 2.5}{2} = 5.5$$

الانحراف المتوسط :
يعرف بأنه المتوسط الحسابي للانحرافات المطلقة عن المتوسط الحسابي

$$EM = \frac{\sum |x_i - \overline{X}|}{n}$$

مثال أحسب الانحراف المتوسط للبيانات التالية
4 , 5 , 2 , 10 , 7

$$\overline{X} = \frac{\sum x_i}{n} = \frac{4+5+2+9+10}{5} = \frac{30}{5} = 6$$

$\|x_i - \overline{X}\|$	$x_i - \overline{X}$	x_i
4	-4	2
2	-2	4
-1	-1	5
3	3	9
4	4	10
14	0	3

$$EM = \frac{14}{5} = 2,8$$

في حالة بيانات مبوبة ،

$$EM = \frac{\sum n_i |C_i - \overline{x}|}{n}$$

مثال السابق :

$\overline{x} = 39.5$

$n_i \cdot \lvert C_i - \overline{x} \rvert$	$C_i - \overline{x}$	C_i	n_i	الحدود
239,2	-29.9	10	8	
199	-19.9	20	10	
118,8	-9.9	30	12	
2,5	0,1	40	25	
353,5	10,1	50	35	
201	20,1	60	10	
1114	0	/	100	Σ

$$EM = \frac{1114}{100} = 11,14$$

مزايا الانحراف المتوسط :

. يأخذ جميع القيم بعين الاعتبار ولكن يصعب التعامل معه رياضيا.

التباين . La Variance $V(x)$

يعتبر أهم مقاييس التشتت و أكثرها استخداما حيث يقيس مجموع مربعات انحرافات القيم عن متوسطها الحسابي له هذه الأهمية لأنه يأخذ جميع القيم بعين الإعتبار و تسهل معالجته رياضيا في حالة بيانات غير مبوبة

$$V(x) = \frac{\sum (x - \overline{x})^2}{n}$$

مثال :

. 2 . 10 . 7 . 5 . 4 . 2

حساب $\overline{x}$:

$$\overline{x} = \frac{\sum x_i}{n} = \frac{30}{6} = 5$$

$$V(x) = \frac{\sum (x_i - \overline{x})^2}{n}$$

$$= \frac{(2-5)^2 + (10-5)^2 + (7-5)^2 + (4-5)^2 + (2-5)^2}{6}$$

$$= \frac{9 + 25 + 4 + 1 + 9}{6} = \frac{48}{6} = 8$$

في حالة بيانات مبوبة :

$$V(x) = \frac{\sum n_i (x_i - \overline{x})}{N}$$

يمكن تبسيط علاقة التباين و كتابتها بشكل آخر

$$V(x) = \sigma^2 = \frac{\sum n_i (x_i - \overline{x})^2}{N}$$

$$= \frac{\sum n_i (x_i^2 + \overline{x}^2 - 2x\overline{x})}{N}$$

$$= \frac{\sum n_i x^2 + \sum n\bar{x}^2 - \sum 2n_i x \bar{x}}{N}$$

$$= \frac{\sum n_i x^2 + \bar{x}^2 \sum n_i - 2\bar{x} \sum n_i x}{N}$$

$$= \frac{\sum n_i x_i^2}{N} + \frac{N\bar{x}^2}{N} - 2\bar{x}\frac{\sum n_i x_i}{N}$$

$$= \frac{\sum n_i x_i^2}{N} + \bar{x}^2 - 2\bar{x}^2$$

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

مثال

أحسب التباين والإنحراف المعياري :

$n(x-\bar{x})^2$	$(x-\bar{x})^2$	$n_i \times x_i$	n_i	x_i
11,52	5,76	0	2	0
5,88	1,96	3	3	1
6,8	0,16	10	5	2
2,16	0,36	18	6	3
7,62	2,56	12	3	4
6,76	6,76	5	1	5
40,74	/	48	20	$\sum$

$$\bar{X} = \frac{\sum n_i x_i}{N} = \frac{48}{20} = 2,4$$

حساب التباين :

ط1 :
$$V(x) = \frac{\sum ni(xi-\bar{x})^2}{N}$$
$$= \frac{34.8,74}{20} = 2,034\ 1.74$$

ط2 :
$$V(x) = \frac{\sum ni \times xi^2}{N} - \bar{x}^2$$
$$= \frac{150}{20} - (2,4)^2$$
$$= 7,5 - 5,76 = 1,74.$$

الانحراف المعياري :
$$\sigma = \sqrt{1,74} = 1,31$$

خصائص التباين :

- التباين دائما موجبا أو مساوي له
$$V(x) \geq 0$$
- a ثابت حقيقي $V(ax) = a^2 V(x)$
- لكل ثابت حقيقي a : $V(a) = 0$
- a, b ثابتان $V(ax+b) = a^2 V(x)$
- اذا كان X مستقل عن Y فإن
$$V(x+y) = V(x) + V(y)$$

Scanné avec CamScanner

معامل الاختلاف Coefficient de Variation

مقاييس التشتت التي تم التطرق إليها كلها مطلقة ولكن في بعض الأحيان نحتاج إلى مقارنة تشتت مجموعتين أو أكثر و قد تكون هذه المجموعات بوحدات قياس مستقلة

و قد يكون الاختلاف في متوسطاتها هنا نلجأ إلى مقاييس تشتت نسبي

- من أبرز مقاييس التشتت النسبي نجد معامل الاختلاف

$$CV = \frac{\sigma}{\bar{X}} \times 100$$

مثال قام أحد مراقبي الإنتاج بأخذ عينة عشوائية من انتاج يوم واحد لآلة معينة وكان حجم الإنتاج 200 وحدة فوجد $\sigma_x = 100g$ و $\bar{X} = 1000g$ ومن عينة أخرى من نفس الحجم لنفس الآلة لكن في يوم آخر $\bar{Y} = 960$, $\sigma_y = 150$ نحدد في أي الأيام اعطت هذه الآلة انتاج أقل تشتتا.

الحل: من أجل المقارنة نعتمد على مقياس تشتت نسبي.

مثلاً CV :

$$CV_1 = \frac{\sigma_x}{\bar{X}} \times 100 = \frac{100}{1000} \times 100 \; 10\%$$

$$CV_2 = \frac{\sigma_x}{\bar{X}} \times 100 = \frac{150}{960} = 15,61$$

أي أن انتاج الآلة في اليوم الأول أقل تشتتا وأكثر تجانسا عنه في اليوم الثاني.

مثال:

أحسب التباين في الجدول التالي:

$n_i c_i^2$	$n_i c_i$	c_i	n_i	الفئات
122500	350	350	1	]400, 300]
405000	900	450	2	]500, 400]
302500	550	550	1	]600, 500]
422500	650	650	1	]700, 600]
1125000	1500	750	2	]800, 700]
722500	850	850	1	]900, 800]
3100000	4800	✓	8	Σ

$$\bar{x} = \frac{\Sigma n_i c_i}{N} = \frac{4800}{8} = 600$$

$$V(x) = \frac{3100000}{8} - 360000 = 27500.$$

$$\sigma = \sqrt{V(x)}$$

$$\sigma = 165,83$$

$$CV = \frac{\sigma}{\bar{x}} = \frac{165,83}{600} \times 100 = 27,63\ \%$$

Scanné avec CamScanner