الفصل الرابع: مقاييس التشتت

الفصل الرابع: مقاييس التشتت

يقصد بالتشتت أو الاختلاف بأنه التباعد أو التقارب الموجود بين قيم المشاهدات التابعة للمتغير. والتشتت صفة هامة من صفات أي مجموعة من البيانات الرقمية. فلا يمكن أن نتصور تساوي الإنتاج في جميع المؤسسات الصناعية أو تساوي مداخيل جميع أفراد المجتمع. فمقاييس التشتت تعطينا معلومات عن مدى انتشار (تباعد أو اقتراب) قيم السلسلة الاحصائية حول قيمة مركزية. فكلما كان مقياس التشتت كبيرا دل ذلك على عدم التجانس بين القيم ويكون مقياس التشتت صغيرا عندما يكون الاختلاف بين القيم صغيرا.

تتمثل مقاييس التشتت في المدى، الانحراف المتوسط، التباين، الانحراف المعياري ومعامل الاختلاف (التغير).

I. مقاييس التشتت البسيطة:

1- المدى:

يرمز له بالحرف W. هو أبسط مقياس لتشتت المعطيات وهو يحسب بإيجاد الفرق بين أكبر قيمة وأصغر قيمة في السلسلة الإحصائية:

$$\boxed{W = X_{max} - X_{min}}$$

وهذا المقياس غير دقيق لأنه لا يعتمد إلا على حساب قيمتين في السلسلة وبالتالي لا يعير اهتماما للقيم الأخرى في المجموعة وهو لا يعطي تصورا واضحا على مدى انتشار وتوزيع القيم داخل المجموعة.

2- المدى الربيعي (S):

إن حساب المدى يمكن أن يؤدي إلى تفسيرات خاطئة، خاصة عند وجود قيم شاردة أو حدية سواء كانت صغيرة جدا أو كبيرة جدا. ومن أجل التخلص من تأثير القيم الشاردة، نفضل استعمال المجال ما بين الربيعيات، وهو الفرق بين الربيع الثالث Q_3 والربيع Q_1.

$$\boxed{S = Q_3 - Q_1}$$

3- نصف المدى الربيعي (الانحراف الربيعي):

وهو يساوي نصف المجال ما بين الربيعيات وهو قريب جدا من الوسيط.

$$\boxed{SE = \frac{Q_3 - Q_1}{2}}$$

4- **المدى الربيعي النسبي:**

يستعمل في المقارنة بين التوزيعات الإحصائية حتى وإن كانت وحدات القياس فيها مختلفة. ويساوي المجال ما بين الربيعيات مقسوما على الوسيط (أو الربيع الثاني).

$$SR = \frac{Q_3 .\ Q_1}{Q_2} = \frac{Q_3 .\ Q_1}{Me} = \frac{S}{Me}$$

إذا كان $SR = 50\%$ ←، فإن تشتت السلسلة يكون معتدلا.

إذا كان $SR > 50\%$ ← ، فإن تشتت السلسلة يكون قويا.

إذا كان $SR < 50\%$ ← ، فإن تشتت السلسلة يكون ضعيفا.

II- **التباين:**

1. **الانحراف المتوسط(e):**

يقبس التشتت المتوسط لمختلف القيم عن وسطها الحسابي بقيم مطلقة وهو يساوي:

$$e_{\overline{X}} = \frac{\sum |\ x_i - \overline{x}\ |}{n} = \frac{\sum |\ x_i - \overline{x}\ | n_i}{\sum n_i}$$

بيانات مبوبة بيانات غير مبوبة

ويمكن كذلك حساب الانحراف المتوسط بالنسبة للوسيط:

$$e_{Me} = \frac{\sum |\ x_i - \mathrm{Me}\ |}{n} = \frac{\sum |\ x_i - \mathrm{Me}\ | n_i}{\sum n_i}$$

2. **التباين V(x):**

يقبس التشتت لمربعات الفوارق لمختلف القيم عن وسطها الحسابي وهو يساوي:

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

ويساوي كذلك بدلالة التكرارات النسبية:

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

ويمكن حسابه بطريقة أخرى كما يلي:

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

وبدلالة التكرار النسبي:

$$\boxed{V(x) = \sum x_i^2 f_i - \overline{x}^{\,2}}$$

في حالة المتغير المستمر، x_i تعبر عن مراكز الفئات.+

3- **خواص التباين:**

$$\boxed{Var(aX) = a^2\, Var(X)}$$

➢ ***مثال:*** قررت المؤسسة رفع أجور العمال بنسبة (10%).

أوجد تباين الأجر الجديد علما أن $Var(X) = 4,8$.

➢ ***الحل:*** نرمز للأجر الجديد ب (y_i)

$$y_i = x_i + 0,1x_i$$

$$\rightarrow y_i = 1,1\, x_i$$

$$Var(y_i) = Var(1,1\, x_i)$$

$$= (1,1\,)^2\, Var(x_i)$$

$$= (1,1\,)^2\, 4,8$$

$$\boxed{Var(y_i) = 5,808}$$

➢ ***البرهان أن:*** $\sum[n_i(x_i - \overline{x})] = 0$

$$\sum[n_i(x_i - \overline{x})] = \sum n_i x_i - \sum n_i \overline{x}$$

$$= \sum n_i x_i - \overline{x} \sum n_i$$

$$= \sum n_i x_i - \boxed{\frac{\sum n_i x_i}{\sum n_i}} . \sum n_i$$

$$= \sum n_i x_i - \sum n_i\, x_i$$

$$= 0$$

$$\boxed{\rightarrow \sum [n_i(x_i - \overline{x})] = 0}$$

4- الانحراف المعياري (σ):

يساوي الجذر التربيعي للتباين:

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

ويعتبر أحسن مقياس للتشتت لأنه يتعامل مع جميع قيم السلسلة الإحصائية.

III- معامل الإختلاف أو التغير (CV):

يصعب علينا مقارنة سلسلتين أو أكثر، والتي تكون فيها وحدات القياس مختلفة ولذلك وجب إيجاد خاصية للمقارنة لا تتأثر بوحدة القياس. وهذه الخاصية تدعى معامل الاختلاف، وهي تعبر على نسبة كل من الانحراف المتوسط، الانحراف المعياري أو التباين، إلى الوسط الحسابي، وهو يساوي:

$$\boxed{CV = \frac{e_{\overline{X}}}{\overline{X}} \times 100 = \frac{e_{Me}}{\overline{X}} \times 10 = \frac{V(x)}{\overline{X}} \times 100 = \frac{\sigma}{\overline{X}} \times 100}$$

كلما كبر معامل الاختلاف، كلما دل ذلك على قوة التشتت بين مفردات التوزيع. وكلما صغر كلما دل ذلك على ضعف التشتت.

- إذا كان $CV = 35\%$ ←يكون التشتت معتدلا.
- إذا كان $CV > 35\%$ ← يكون التشتت قويا (بيانات غير متجانسة).
- إذا كان $CV < 35\%$ ← يكون التشتت ضعيفا (بيانات متجانسة).

ويمكن معرفة مدى تشتت السلسلة باستعمال الربيعيات من خلال المقاييس التالية:

➢ ***ملاحظة:***

إن معامل الاختلاف (CV) والمدى الربيعي النسبي (SR) عبارة عن مقاييس نسبية للتشتت، أما الباقي فهي مقاييس مطلقة.

➢ ***مثال:***

الجدول (4-1): حساب التباين

الأجور	عدد العمال n_i	x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$n_i(x_i - \overline{x})^2$
3-5	10	4	-4	16	160
5-7	20	6	-2	4	80
7-9	40	8	0	0	0
9-11	20	10	2	4	80
11-13	10	12	4	16	160
Σ	**100**	-	-	-	**480**

احسب التباين علما أن: $\overline{X} = 8$

$$Var(X) = \frac{480}{100} = \boxed{4,8}$$

البرهان:

$$Var(X) = \frac{\sum[n_i(x_i - \overline{x})^2]}{\sum n_i} = \frac{\sum (n_i x_i^2)}{\sum n_i} - \overline{x}^{\,2}$$

لدينا:

$$Var(X) = \frac{\sum[n_i(x_i - \overline{x})^2]}{\sum n_i}$$

$$= \frac{\sum\left[n_i\left(x_i^2 + \overline{x}^2 - 2x_i\overline{x}\right)\right]}{\sum n_i}$$

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

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

$\overline{X}$

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

$$= \frac{\sum(n_i x_i^2)}{\sum n_i} - \overline{X}^2$$