مقيـــــــــاس: إحصـــــــاء 1

المحور الخامس: مقاييـــــس التشتـت

Caractéristiques de Dispersion

تمهيد الفصل: إن مقاييس النزعة المركزية غير كافية لوحدها لتحديد خواص الظاهرة المدروسة بشكل جيد ولا سيما في مجال المقارنة بين عدة مجموعات من الظواهر المدروسة، لأنه يمكن أن يكون لتوزيعين إحصائيين نفس الوسط الحسابي ولكنهما يختلفان تماماً من حيث توزيع القيم حول هذا الوسط، وبالتالي نجد أن هناك ضرورة لقياس مقدار ابتعاد مختلف المشاهدات عن هذه القيمة المتوسطة.

مثال 1: ليكن لدينا السلسلتين الإحصائيتين التاليتين:

$$Serie\ 1: \quad 90, \quad 95, \quad 100, \quad 105, \quad 110 \Rightarrow \bar{x} = \frac{\sum x_i}{n} = \frac{500}{5} = 100$$

$$Serie\ 2: \quad 10, \quad 20, \quad 70, \quad 120, \quad 280 \Rightarrow \bar{y} = \frac{\sum y_i}{n} = \frac{500}{5} = 100$$

نلاحظ أن السلسلة الإحصائية الأولى وسطها الحسابي يساوي 100 وهو يساوي الوسط الحسابي للسلسلة الإحصائية الثانية، معنى ذلك أنه إذا اكتفينا بمقارنة الوسطين الحسابيين لكلتا المجموعتين فقد نستنتج أن السلسلتين متساويتين، غير أنه من ينظر إلى السلسلتين تظهر أنهما مختلفتين في الواقع، حيث أن بيانات السلسلة الأولى متقاربة من بعضها أو قريبة من قيمة الوسط الحسابي بينما بيانات السلسلة الثانية متباعدة عن بعضها البعض، ولهذا نلجأ إلى حساب مقاييس التشتت لمعرفة مدى تباعد القيم او تقاربها عن مركزها. ونميز عند دراستنا لمقاييس التشتت بين:

- مقاييس التشتت المطلقة: والتي تضم كل من المدى العام، الانحراف المتوسط، الانحراف الوسيط، الانحراف الربيعي والانحراف المعياري.
- مقاييس التشتت النسبية: والتي تضم معامل الاختلاف الاول (الانحراف المعياري النسبي) و معامل الاختلاف الثاني (الانحراف الربيعي النسبي).

1/ مقاييس التشتت المطلقة

1.1 المدى العام (E) *L'Etendue*: المدى العام لمجموعة من البيانات هو الفرق بين أكبر مشاهدة وأصغر مشاهدة، ويعبر عنه بالعلاقة التالية: $E = x_i(max) - x_i(min)$

ونلجأ عادة إلى حساب المدى العام عند اهتمامنا بالقيم المتطرفة (المشتتة) أو في حالة تبويب البيانات وجعلها في شكل فئات، ويتصف المدى العام ببساطة حسابه وأنه سهل الفهم ويعتمد في حسابه على قيمتين فقط، كما أنه شديد التأثر بالقيم المتطرفة.

2.1 الانحراف المتوسط ($e_{\bar{x}}$) *L'écart moyen*

يُعرف الانحراف المتوسط على أنه متوسط القيم المطلقة لانحرافات القيم عن وسطها الحسابي، ونستخدم القيمة المطلقة لتفادي الاشارة السالبة عند حساب الانحرافات، وتختلف طرق حسابه باختلاف طبيعة البيانات.

البيانات المنفصلة غير المتكررة	البيانات المنفصلة المتكررة	البيانات المتصلة
$e_{\bar{x}} = \frac{\sum \lvert x_i - \bar{x} \rvert}{n}$	$e_{\bar{x}} = \frac{\sum F_i \lvert x_i - \bar{x} \rvert}{\sum F_i}$	$e_{\bar{x}} = \frac{\sum F_i \lvert c_i - \bar{x} \rvert}{\sum F_i}$

مثال 2: أحسب الانحراف المتوسط للبيانات التالية: 5، 10، 15، 20، 25، 45.

الحل: $\bar{x} = \frac{\sum x_i}{n} = \frac{120}{6} = 20$

$$e_{\bar{x}} = \frac{\sum |x_i - \bar{x}|}{n} = \frac{|5-20| + |10-20| + \cdots + |25-20| + |45-20|}{6} = \frac{60}{6} \Rightarrow e_{\bar{x}} = \mathbf{10}$$

3.1 الانحراف الوسيط (e_{Me}) *L'écart médian*

يُعرف الانحراف الوسيط على أنه متوسط القيم المطلقة لانحرافات القيم عن وسيطها، وتختلف طرق حسابه باختلاف طبيعة البيانات.

البيانات المتصلة	البيانات المنفصلة المتكررة	البيانات المنفصلة غير المتكررة
$e_{Me} = \frac{\sum F_i \lvert c_i - Me \rvert}{\sum F_i}$	$e_{Me} = \frac{\sum F_i \lvert x_i - Me \rvert}{\sum F_i}$	$e_{Me} = \frac{\sum \lvert x_i - Me \rvert}{n}$

مثال 3: أحسب الانحراف الوسيط لبيانات الجدول التكراري التالي:

x_i	F_i	Fcc	$\lvert x_i - Me \rvert$	$F_i \lvert x_i - Me \rvert$
10	5	0 / 5	30	150
20	10	15	20	200
30	25	40	10	250
40	35	$Rme = 45$ / 75	0	0
50	15	90	10	150
$\sum$	**90**	-	-	**750**

$$Rme = \frac{\sum Fi}{2} = 45 \Rightarrow Me = 40 \qquad e_{Me} = \frac{\sum F_i |x_i - Me|}{\sum F_i} = \frac{750}{90} = \mathbf{8,33}$$

4.1 الانحراف الرُبيعي (e_Q) *L'écart quartile*

يُعرف الانحراف الرُبيعي على أنه نصف المدى ما بين الرُبيع الأول Q_1 والرُبيع الثالث Q_3، في حالة البينات المتصلة لحساب قيم الرُبيع الأول والثالث نتبع نفس طريقة حساب الوسيط (الربيع الثاني) حيث نعتمد على قيم Fcc ورُتبة الرُبيع لتحديد الفئة الرُبيعية، ومنه:

الربيع الأول: $$RQ_1 = \frac{\sum F_i}{4} \Rightarrow Q_1 = L_0 + \frac{\frac{\sum F_i}{4} - F_1}{F_2 - F_1} \times k$$

الربيع الثالث: $$RQ_3 = \frac{3\sum F_i}{4} \Rightarrow Q_3 = L_0 + \frac{\frac{3\sum F_i}{4} - F_1}{F_2 - F_1} \times k$$

ويُحسب الانحراف الرُبيعي وفقاً للعلاقة التالية: $\boldsymbol{e_Q = \frac{Q_3 - Q_1}{2}}$

مثال 4: أحسب الانحراف الرُبيعي للبيانات التالية:

$Classes$	F_i	Fcc
10 - 20	6	0 / 6
20 - 30	10	16
30 - 40	26	$RQ_1 = 25$ / 42
40 - 50	38	$RQ_3 = 75$ / 80
50 - 60	20	100
$\sum$	**100**	-

$$RQ_1 = \frac{\sum F_i}{4} \Rightarrow Q_1 = L_0 + \frac{\frac{\sum F_i}{4} - F_1}{F_2 - F_1} \times k = 30 + \frac{25 - 16}{42 - 16} \times 10 = \mathbf{33,46}$$

$$RQ_3 = \frac{3\sum F_i}{4} \Rightarrow Q_3 = L_0 + \frac{\frac{3\sum F_i}{4} - F_1}{F_2 - F_1} \times k = 40 + \frac{75 - 42}{80 - 42} \times 10 = \mathbf{48,68}$$

$$e_Q = \frac{Q_3 - Q_1}{2} = \frac{48,68 - 33,46}{2} = \mathbf{7,61}$$

5.1 الانحراف المعياري (σ_x) *L'écart type*

الانحراف المعياري هو الجذر التربيعي لمتوسط مربعات انحرافات القيم عن وسطها الحسابي، ويُحسب بعدة طرق وفقاً لطبيعة البيانات الإحصائية.

البيانات المتصلة	البيانات المنفصلة المتكررة	البيانات المنفصلة غير المتكررة
$\sigma_x = \sqrt{\dfrac{\sum F_i(c_i-\bar{x})^2}{\sum F_i}}$	$\sigma_x = \sqrt{\dfrac{\sum F_i(x_i-\bar{x})^2}{\sum F_i}}$	$\sigma_x = \sqrt{\dfrac{\sum(x_i-\bar{x})^2}{n}}$

ملاحظات:

- مربع الانحراف المعياري يسمى التباين $V_x = \sigma_x^2$؛
- يرمز للانحراف المعياري بالنسبة للعينة بالرمز S_x ويُطرح العدد 1 من عدد المفردات $(n-1)$ أو من عدد التكرارات $(\sum F_i - 1)$ في حساب الانحراف المعياري أو التباين للعينة.

مثال 5: أحسب الانحراف المعياري للبيانات التالية: 2، 8، 14، 20، 26.

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
2	-12	144
8	-6	36
14	0	0
20	6	36
26	12	144
Σ	-	360

أ/ حساب الوسط الحسابي:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{70}{5} = 14$$

ب/ حساب الانحراف المعياري بثلاثة طرق :

$$\sigma_x = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n}} = \sqrt{\frac{360}{5}} = \mathbf{8,49}$$

مثال 6: أحسب الانحراف المعياري والتباين للمجتمع وللعينة لبيانات الجدول التكراري التالي:

Classes	0 – 4	4 – 8	8 – 12	12 - 16	16 – 20
Fi	30	15	20	15	30

الحل:

Classes	F_i	c_i	$F_i \times c_i$	$(c_i-\bar{x})$	$(c_i-\bar{x})^2$	$F_i(c_i-\bar{x})^2$
0 – 4	30	2	60	-8	64	1920
4 – 8	15	6	90	-4	16	240
8 – 12	20	10	200	0	0	0
12 – 16	15	14	210	4	16	240
16 – 20	30	18	540	8	64	1920
$\sum$	**110**	-	**1100**	-	-	**4320**

$$\bar{x} = \frac{\sum F_i c_i}{\sum F_i} = \frac{1100}{110} = \mathbf{10}$$

$$\sigma_x = \sqrt{\frac{\sum F_i(c_i-\bar{x})^2}{\sum F_i}} = \sqrt{\frac{4320}{110}} = \mathbf{6,27} \Rightarrow V_x = \sigma_x^2 = (6{,}27)^2 = \mathbf{39,27}$$

$$S_x = \sqrt{\frac{\sum F_i(c_i-\bar{x})^2}{\sum F_i - 1}} = \sqrt{\frac{4320}{109}} = \mathbf{6,30} \Rightarrow V_x = S_x^2 = (6{,}30)^2 = \mathbf{39,63}$$

2. مقاييس التشتت النسبية:

إذا كانت مقاييس التشتت المطلقة تقيس نفس الظاهرة محل الدراسة فإننا نلجأ إلى مقاييس التشتت النسبية في حالة المقارنة بين تشتت توزيعات لظواهر مختلفة أو نفس التوزيع لكن على مستويات مختلفة من أجل التخلص من تأثير اختلاف وحدات القياس وكذلك التخلص من تأثير اختلاف قيم المتوسطات والربيعيات.

ولإجراء ذلك نستخدم مقياس يسمى بمعامل الاختلاف وله صورتان: معامل الاختلاف الأول (الانحراف المعياري النسبي) ومعامل الاختلاف الثاني (الانحراف الربيعي النسبي).

1.2 معامل الاختلاف الأول (CV_1) *Coefficient de Variation 1*

يُسمى أيضاً بالانحراف المعياري النسبي، نحتاج في حسابه إلى كل من المتوسط الحسابي والانحراف المعياري للعينة وفقاً للعلاقة التالية:

$$CV_1 = \frac{S_x}{\bar{x}} \times 100$$

مثال 7: بالعودة لمعطيات المثال رقم 6 أحسب معامل الاختلاف الأول

$$CV_1 = \frac{S_x}{\bar{x}} \times 100 = \frac{6,30}{10} \times 100 = \mathbf{63\%}$$

مثال 8: بعد تصحيح أوراق الامتحانات لمجموعتين من الطلبة وُجد أن متوسط علامات المجموعة الأولى هو 13بانحراف معياري مقداره 3، بينما متوسط علامات المجموعة الثانية هو 8بانحراف معياري قدره 2. أي المجموعتين أكثر تشتتاً؟

الحل: للإجابة على السؤال يجب حساب معامل الاختلاف الأول في كلتا المجموعتين ومقارنة النتائج.

$$CV_1(G_1) = \frac{S_x}{\bar{x}} \times 100 = \frac{3}{13} \times 100 = \mathbf{23,08\%}$$

$$CV_1(G_2) = \frac{S_x}{\bar{x}} \times 100 = \frac{2}{8} \times 100 = \mathbf{25\%}$$

من خلال النتائج المتحصل عليها نستنتج أن علامات المجموعة الثانية أكثر تشتتاً (تباعداً فيما بينها) مقارنة بعلامات المجموعة الأولى.

2.2 معامل الاختلاف الثاني (CV_2) *Coefficient de Variation 2*

يُسمى أيضاً بالانحراف الربيعي النسبي، نحتاج في حسابه إلى حساب أولاً قيم الربيعات الثلاثة، ويُحسب وفقاً للعلاقة التالية:

$$CV_2 = \frac{Q_3 - Q_1}{Me} \times 100$$

مثال 9: قارن بين تشتت التوزيعين التاليين:

المجموعة 1 (G1): 8، 12، 16، 20، 24، 28، 32.

المجموعة 2 (G2): 2، 7، 12، 17، 22، 27، 32.

$$\boldsymbol{G_1}: RQ_1 = \frac{(n+1)}{4} = 2 \Rightarrow Q_1 = \mathbf{12};\quad Rme = \frac{(n+1)}{2} = 4 \Rightarrow Q_2 = \mathbf{20};\quad RQ_3 = \frac{3(n+1)}{4} = 6 \Rightarrow Q_3 = \mathbf{28}$$

$$\boldsymbol{CV_2(G_1) = \frac{Q_3 - Q_1}{Me} \times 100 = \frac{28-12}{20} \times 100 = 80\%}$$

$$\boldsymbol{G_2}: RQ_1 = \frac{(n+1)}{4} = 2 \Rightarrow Q_1 = \mathbf{7};\quad Rme = \frac{(n+1)}{2} = 4 \Rightarrow Q_2 = \mathbf{17}; RQ_3 = \frac{3(n+1)}{4} = 6 \Rightarrow Q_3 = \mathbf{27}$$

$$\boldsymbol{CV_2(G_2) = \frac{Q_3 - Q_1}{Me} \times 100 = \frac{27-7}{17} \times 100 = 117,65\%}$$

$$\boldsymbol{CV_2(G_2) > CV_2(G_1)}$$

نستنتج أن قيم المجموعة 2 أكثر تشتتاً (تباعدا) فيما بينها مقارنة بقيم المجموعة 1.

مثال 10: أحسب معامل الاختلاف الأول والثاني للبيانات التالية (حجم العينة 120):

80-70	70-60	60-50	50-40	40-30	30-20	20-10	الفئات
10	5	10	15	20	25	15	$Fr(\%)$

Classes	$Fr(\%)$	Fr	Fi	Ci	Fcc	$Fi \times$ Ci	$(c_i - \bar{x})$	$(c_i - \bar{x})^2$	$F_i(c_i - \bar{x})^2$
10 - 20	15	0,15	18	15	0	270	-23,5	552,25	9940,5
20 - 30	25	0,25	30	25	18 RQ_1	750	-13,5	182,25	5467,5
30 - 40	20	0,20	24	35	48 RQ_2 72	840	-3,5	12,25	294
40 - 50	15	0,15	18	45	RQ_3 90	810	6,5	42,25	760,5
50 - 60	10	0,10	12	55		660	16,5	272,25	3267
60 - 70	5	0,05	6	65	102	390	26,5	702,25	4213,5
70 - 80	10	0,10	12	75	108	900	36,5	1332,25	15987
$\sum$	**100**	**1,00**	**120**	-	120	**4620**	-	-	**39930**

$$\bar{x} = \frac{\sum F_i c_i}{\sum F_i} = \frac{4620}{120} = 38{,}5 \quad S_x = \sqrt{\frac{\sum F_i (c_i - \bar{x})^2}{\sum F_i - 1}} = \sqrt{\frac{39930}{119}} = 18{,}32$$

$$\boldsymbol{CV_1 = \frac{S_x}{\bar{x}} \times 100 = \frac{18,32}{38,5} \times 100 = 47,58\%}$$

$$RQ_1 = \frac{\sum F_i}{4} = 30 \Rightarrow [20 - 30[\text{ فئة ربيعية أولى };\quad Q_1 = L_0 + \frac{\frac{\sum F_i}{4} - F_1}{F_2 - F_1} \times k = 20 + \frac{30 - 18}{48 - 18} \times 10 = \mathbf{24}$$

$$RQ_2 = \frac{\sum F_i}{2} = 60 \Rightarrow [30 - 40[\text{ فئة ربيعية ثانية };\quad Q_2 = L_0 + \frac{\frac{\sum F_i}{2} - F_1}{F_2 - F_1} \times k = 30 + \frac{60 - 48}{72 - 48} \times 10 = \mathbf{35}$$

$$RQ_3 = \frac{3\sum F_i}{4} = 90 \Rightarrow [40 - 50[\text{ فئة ربيعية ثالثة};\ Q_3 = L_0 + \frac{\frac{3\sum F_i}{4} - F_1}{F_2 - F_1} \times k = 40 + \frac{90 - 72}{90 - 72} \times 10 = \mathbf{50}$$

$$\boldsymbol{CV_2 = \frac{Q_3 - Q_1}{Q_2} \times 100 = \frac{50 - 24}{35} \times 100 = 74,29\%}$$