مقاييس التشتت

مقدمة:

إذا كانت المتوسطات تعطينا فكرة عن التوزيع التكراري، إلا أن هذه الفكرة ليست دقيقة وكاملة، لأن المتوسط وحده لا يكفي لإعطاء فكرة دقيقة عن المجموعة من حيث طبيعتها وكيفية توزيع مفرداتها فضلا عن ذلك، فإن المتوسط لا يكفي عند عقد المقارنة بين عدة مجموعات، حيث أنه لا يظهر حقيقة المقارنة. فقد يتساوى متوسط مجموعتين في حين تختلف المجموعتان عن بعضهما كل الاختلاف فقد تكون مفردات إحدى المجموعتين متقاربة بعضها من بعض أو مبعثرة.

مثال: لدينا أجور ثلاثة أفراد بالساعات كما يلي:

$$(1)\quad 95\,,97,100,103,105\ \Rightarrow \bar{X}_1=\frac{\sum x_i}{N}=\frac{500}{5}=100$$

$$(2)\quad 50\,,75,100,125,150\ \Rightarrow \bar{X}_2=\frac{\sum x_i}{N}=\frac{500}{5}=100$$

$$(3)\quad 15\,,85,100,115,185\ \Rightarrow \bar{X}_3=\frac{\sum x_i}{N}=\frac{500}{5}=100$$

إن الوسط الحسابي لكل مجموعة هو 100 دج، ولكن تختلف هذه التوزيعات فيما بينها من حيث مدى تجانس أو تشتت القيم حول مقياس النزعة المركزية (الوسط الحسابي).

1- المدى، المدي الربيعي، نصف المدى الربيعي:

1-1 المدى: أبسط مقاييس التشتت المطلق ويتم الحصول عليه من خلال الفرق بين أكبر قيمة وأصغر قيمة في مجموعة القيم.

مثال: لدينا القيم التالية:

المجموعة الأولى: 50، 52، 56، 58، 62.

المجموعة الثانية: 50، 55، 62، 70، 98.

المدى في المجموعة الأولى: $E_1 = 62 - 50 = 12$

المدى في المجموعة الثانية: $E_2 = 98 - 50 = 48$

إذن المجموعة الثانية هي الأكثر تشتتا، إن هذه الطريقة وإن كانت سهلة عمليا، لكنها ليست دقيقة وأحيانا تكون مضللة، حيث يسبب وجود قيمة متطرفة في المجموعة زيادة كبيرة في طول المدى ويستدل منها على وجود تشتت، مع أن جميعها في الواقع مجتمعة بالقرب من بعضها ماعدا هذه القيمة الشاذة.

ملاحظة: في حالة وجود قيم شاذة يمكن استخدام ما يسمى شبيهات المدى.

- **المدى الأول:** نحصل عليه باستبعاد قيمة واحدة من كل طرفي القيم.
- **المدى الثاني:** نحصل عليه باستبعاد قيمتين من كل طرفي القيم.

هذا عن البيانات غبر المبوبة، أما في حالة البيانات المبوبة فالمدى هو الفرق بين الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة الأولى.

استنتاج:

المدى قياس مضلل للدرجة تشتت الظاهرة، لأنه يتوقف على أكبر قيمة وأصغر قيمة، فإذا كانت إحداهما أو كليهما قيم متطرفة، فإن المدى لا يعكس التشتت الفعلي للبيانات غبر المبوبة، أما في حالة البيانات المبوبة، فإنه يهمل التكرارات الواقعة عند كلا من الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة الأولى.

2-1 المدى الربيعي: وهو الفرق بين الربيعي الثالث والربيعي الأول $(Q_3 - Q_1)$، ويرمز له بالرمز: $I\,Q$ ويعتبر أحسن من المدى العام، إذ يضم 50% من مفردات المجتمع الإحصائي، ويستعمل في المقارنة بين توزيعين إحصائيين أو أكثر.

3-1 نصف المدى الربيعي أو الانحراف الطبيعي: يستعمل للتخلص من تأثير القيم الشاذة الدنيا منها والعليا ويرمز له بالرمز: $E\,Q$ ويحسب وفق التالي: $EQ = \frac{Q_3 - Q_1}{2}$.

مثال 1: لدينا القيم التالية: 71، 70، 69، 73، 68، 74، 77.

المطلوب: أحسب الانحراف الربيعي لهذه القيم.

الحل: - **أولا:** ترتيب المفردات تصاعديا أو تنازليا: 68، 69، 70، 71، 73، 74، 77.

- **ثانيا:** تحديد موقع كل من الربيع الثالث والأول.

$$P_{Q1}=\frac{(n+1)}{4}=\frac{(7+1)}{4}=2$$

$$P_{Q3}=\frac{3(n+1)}{4}=\frac{3(7+1)}{4}=6$$

- **ثالثا:** تحديد قيمة الربيع الثالث والأول وفق موقعهما – ترتيبهما – $Q_1=69$، $Q_3=74$.

ومنه: $EQ=\frac{Q_3-Q_1}{2}=\frac{74-69}{2}=2.5$

مثال 2: إذا كانت لدينا البيانات التالية:

الفئات	55-50	60-55	65-60	70-65	75-70	80-75	المجموع
التكرارات	30	34	40	28	12	06	150

المطلوب: حساب نصف المدى الربيعي.

الحل: - **أولا:** تكوين جدول تكرار تجميعي صاعد أو نازل.

الفئات	55-50	60-55	65-60	70-65	75-70	80-75	المجموع
التكرارات	30	34	40	28	12	06	150
ت. ت. ص	30	64	104	132	144	150	/

- **ثانيا:** تحديد موقع الربيعيات.

$$P_{Q1}=\frac{n}{4}=\frac{150}{4}=37.5$$

$$P_{Q3}=\frac{3n}{4}=\frac{3\times 150}{4}=112.5$$

– **ثالثا:** حساب الربيعيات وفق العلاقة التالية:

$$Q_1 = 1_1 + \frac{\left(\frac{\sum n_i}{4} - \sum F_1\right)}{F_{Q1}} \, .C$$

$$Q_1 = 55 + \frac{\left(\frac{150}{4} - 30\right)}{34} \, .5 = 56\,.102$$

$$Q_3 = 1_3 + \frac{\left(\frac{3\sum n_i}{4} - \sum F_1\right)}{F_{Q3}} \, .C$$

$$Q_3 = 65 + \frac{\left(\frac{3 \times 150}{4} - 104\right)}{28} \, .5 = 66\,.517$$

2- الانحراف المتوسط:

ويرمز له بالرمز: $E_{\bar{x}}$ ويقصد به مجموع متوسط انحرافات القيم عن متوسطها بغض النظر عن إشارتها وتعطى في حالة البيانات غبر المبوبة على النحو التالي: $E_{\bar{x}} = \frac{\sum|X_i - \bar{X}|}{N}$

والسبب في الاعتماد على القيمة المطلقة للانحرافات هو التخلص من الإشارات السالبة، لأن مجموع انحرافات القيم عن وسطها الحسابي تساوي الصفر.

مثال: أوجد انحرافات المتوسط للقيم التالية: 14، 16، 10، 08، 12.

الجواب: – أولا: تحديد الوسط الحسابي لهذه القيم.

$$\bar{X} = \frac{\sum X_i}{N} = \frac{60}{5} = 12$$

– ثانيا: حساب الانحرافات: $\sum|i - \bar{X}| = 2 + 4 + 2 + 4 = 12$

$$\bar{E}_X = \frac{12}{5} = 2.4$$

أما في حالة البيانات المبوبة:

لحساب متوسط الانحرافات المطلقة لبيانات مبوبة نضيف على العلاقة السابقة تكرارات الفئات كما تعوض القيم بمراكز الفئات أي: $E_{\bar{X}} = \frac{\sum n_i|X_I - \bar{X}|}{N}$

ولحساب ذلك نتبع الخطوات التالية:

1- يحسب الوسط الحسابي للتوزيع التكراري أي $|X_i - \bar{X}|$.

2- تحسب الفروق المطلقة بين مراكز كل فئة والوسط الحسابي أي:

3- يضرب كل فرق في التكرار الخاص به ثم يجمع ونحصل : $\sum n_i|X_i - \bar{X}|$.

4- وبقسمة ذلك المجموع على مجموع التكرارات نحصل على متوسط الانحرافات المطلقة للتوزيع.

مثال: التوزيع التكراري التالي بين إنتاج 60 مزرعة من الفواكه بالطن.

الفئات	$\boldsymbol{n_i}$	$\boldsymbol{X_i}$	$\boldsymbol{n_i\, X_I}$	$\boldsymbol{X_I - X}$	$\boldsymbol{n_i\, \lvert X_i - X \rvert}$
20-10	04	15	60	27	108
30-20	09	25	225	17	153
40-30	16	35	560	07	112
50-40	13	45	585	03	39
60-50	10	55	550	13	130
70-60	06	65	390	23	138
80-70	02	76	150	33	66
المجموع	60	/	2520	/	746

$$\bar{X} = \frac{\sum n_i\, X_i}{N} = \frac{2520}{60} = 42$$

$$E_{\bar{X}} = \frac{\sum n_i|X_i - \bar{X}|}{N} = \frac{746}{60} = 12.43$$

ملاحظات عامة:

يتميز عن المقاييس السابقة للتشتت بأنه يأخذ في الاعتبار جميع المفردات كما أن قيمته في الصغر كلما كبر حجم العينة.

تحسب الانحرافات المطلقة عن الوسيط كما عن المنوال.

3- التباين:

نظرا لصعوبة استخدام انحرافات القيم عن متوسطها الحسابي كأساس لقياس التشتت بسبب الإشارة السالبة الذي جعلنا نحسب الانحراف المتوسط كمتوسط مجموع الانحرافات مع إهمال الإشارة، إلا أن هناك طريقة أخرى لمربعات الفروق بين قيم المتغير الإحصائي والوسط الحسابي ويرمز له بالرمز σ^2.

في حالة البيانات غير المبوبة:

$$\sigma^2 = \frac{\sum(X - X)^2}{N} = \frac{\sum(X_i^2 - 2X_iX + X^2)}{N}$$

$$= \frac{\sum X_i^2 - 2\sum X_iX + NX^2}{N}$$

$$= \frac{\sum X_i^2}{N} - \frac{2X\sum X_1}{N} + \frac{NX^2}{N}$$

$$= \frac{\sum X_i^2}{N} - 2XX + X^2$$

$$= \frac{\sum X_i^2}{N} - 2X^2 + X^2$$

$$= \frac{\sum X_i^2}{N} - X^2$$

مثال 1: فيما يلي درجات عينة من التلاميذ: 5، 8، 11، 15، 20 العلامة من 20.

- أحسب التباين.

التلاميذ	X_i	
1	5	25
2	8	64
3	11	121
4	15	225
5	20	400

الحل: $\bar{X} = 11{,}8$

$\bar{X}^2 = 139,24$

حساب التباين:

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

$$= \frac{835}{5} - 139,24$$

$$= 167 - 139,24 = 27,76$$

- **في حالة توزيع تكراري:** فالصيغة تكون وفق التالي: $\sigma^2 = \frac{\sum n_i(X_i - X)^2}{N}$

$$= \frac{\sum n_i X_i^2}{N} - \bar{X}^2$$

مثال 2: إليك توزيع علامات 15 طالب من ذوي الاحتياجات، أحسب التباين.

$N_i\ X_i$	NX_i^2		X_i	N_i	الفئات
12	144	144	12	01	14-10
48	768	256	16	03	18-14
100	1000	400	20	05	22-18
96	2304	576	24	04	26-22
56	1568	784	28	02	30-26
312	6784	/	/	15	المجموع

$$\bar{X} = \frac{\sum N_i X_i}{N} = \frac{312}{15} = 20.8$$

$$\sigma^2 = \frac{\sum n_i X_i}{N} - \bar{X}^2$$

$$= \frac{5784}{15} - 432,64$$

$$= 452,26 - 432,64 = 19,62$$

4- الانحراف المعياري:

يعتبر الانحراف المعياري من أهم مقاييس التشتت لأنه يحتوي على مفهوم جبري للانحرافات وهم من الأساليب الإحصائية الرياضية الحديثة لقياس التشتت ومن أكثرها استعمالا.

ويعرف على أنه مقياس للتشتت يقيس لنا مدى تباعد أو اختلاف أو انحراف قيم المتغير الإحصائي عن وسطه الحسابي، كما يمكن تعريفه رياضيا بانه الجذر التربيعي للتباين ويرمز له بالرمز: σ

- **حالة البيانات غير المبوبة:** $\sigma = \sqrt{\frac{\sum X_i^2}{N} - \bar{X}^2}$

مثال: أوجد الانحراف المعياري للقيم التالية: 6، 14، 11، 19، 10.

الحل: أولا: - إيجاد المتوسط الحسابي لهذه القيم $\bar{X} = \frac{\sum X_i}{N} = \frac{60}{5} = 12$

ثانيا: - تربيع القيم يكون كالتالي: 36، 196، 121، 361، 100.

ثالثا: - حساب الانحراف المعياري: $\sigma = \sqrt{\frac{\sum X_i^2}{N} - \bar{X}^2}$

$$\sigma = \sqrt{\frac{814}{5} - 144}$$

$$\sigma = 4,33$$

- **حالة البيانات المبوبة:** $\sigma = \sqrt{\frac{\sum n_i X_i^2}{N} - \bar{X}^2}$

مثال: يمثل الجدول التالي توزيع تكراري، لأوزان 50 فردا.

الفئات	54-50	58-54	62-58	66-62	70-66	74-70	78-74
التكرارات	5	8	11	12	7	4	3

المطلوب: أحسب الانحراف المعياري لهذا التوزيع.

الحل:

الفئات	n_i	x_i	$n_i\ x_i$	X_i^2	$n_i\ x_i^2$
54-50	5	52	260	2704	13520
58-54	8	56	448	3136	25088
62-58	11	60	660	3600	39600
66-62	12	64	768	4096	49152
70-66	7	68	476	4624	32368
74-70	4	72	288	5184	20736
78-74	3	76	228	5776	17328
المجموع	50	/	3128	29120	197792

- **حساب المتوسط الحسابي:** $\bar{X} = \frac{\sum N_i X_i}{N} = \frac{3128}{50} = 62.56$

- **حساب الانحراف المعياري:** $\sigma = \sqrt{\frac{\sum n_i X_i^2}{N} - \bar{X}^2}$

$$\sigma = \sqrt{3955,84 - 3919,38}$$

$$\sigma = 6,03$$

1-4 خصائص الانحراف المعياري: يعتبر من أهم مقاييس التشتت وأكثرها استخداما في الإحصاء وذلك لأسباب عديدة تتعلق بعملية الاستنتاج الإحصائي سواء في التقديرات أو في اختيار الفروض.

يعطي وزنا أكثر للقيم المتطرفة، وذلك بالمقارنة بمتوسط الانحرافات المطلقة، وهذا يعني أنه يتأثر بها أكثر منه.

يدخل في تركيب العديد من المقاييس مثل: الالتواء، الدرجة المعيارية ومعامل الارتباط.

وبفضل استخدام الانحراف المعياري، حين لا يكون قياس تشتت الظاهرة هو نهاية التحليل الإحصائي، بل بداية عمليات إحصائية أخرى أكثر أهمية، ونعني بها الاستنتاج الإحصائي بشقيه التقديرات والاختبارات الإحصائية.