المحاضرة السابعة: مقاييس التشتت

Measures of Dispersion

يشير مفهوم مقاييس النزعة المركزية إلى ميل البيانات للتمركز حول قيمة ممثلة أو نموذجية في التوزيع، وتستخدم مقاييس النزعة المركزية لغايات المقارنة بين مجموعتين من البيانات، ولوصف توزيع البيانات إن كان متماثلاً أم غير متماثل.

وهناك مقياس مكملة لمقاييس النزعة المركزية وهي مقاييس التشتت، إذ أن **مقاييس التشتت تشير إلى مدى تقارب أو تباعد أو تناثر أو اختلاف القيم عن بعضها البعض أو عن نقطة معينة كالمتوسط مثلاً.** فالتشتت يرجع إلى اختلاف قيم البيانات، لذلك فإن القيم المتشابهة لا يوجد لها تشتت، ولكن باختلاف قيم البيانات فلا بد من وجود تشتت بين هذه القيم، ويختلف مقدار التشتت من مجموعة إلى أخرى.

وتبرز الحاجة إلى مقاييس التشتت لتقديم وصف دقيق حول التوزيعات التكرارية وخصوصاً عند إجراء المقارنات بين المجموعات أو بين الأفراد، فدلالات مقاييس النزعة المركزية كالمتوسط الحسابي قد تقود إلى اتخاذ قرارات غير صحيحة.

مثال 1: لتكن لدينا البيانات الآتية لثلاث مجموعات:

الجدول (1) بيانات ثلاث مجموعات

المجموعات	الأولى	الثانية	الثالثة
القيم (البيانات)	5 , 5, 5	4, 5,7, 9,2,3	1, 9
المتوسط	5	5	5

يلاحظ أن قيمة المتوسط الحسابي لكل مجموعة يساوي (5) فإذا اكتفينا بهذا المقياس فإننا نقرر أن المجموعات الثلاث متشابهة ولكن في الحقيقة أن قيم المجموعة الثالثة أكثر تباعداً من قيم المجموعة الثانية والتي هي أكثر تباعداً من قيم المجموعة الأولى.

مثال 2: بفرض أننا اخترنا وبشكل عشوائي مجموعتين من الطلبة الذكور (10 طلاب) والإناث (10 طالبات)، قد امتحنوا في مادة اللغة العربية، وكانت علاماتهم وفق الآتي:

علامات الطلاب الذكور: 77 – 77 – 77 –74 – 73 – 71 – 68 – 67 – 66 – 65

علامات الطالبات الإناث: 100 – 94 – 84 – 77 – 77 – 67 – 63 – 57 – 55 – 41

وعند حساب مقاييس النزعة المركزية لكل مجموعة كانت وفق الآتي:

مجموعة الطلاب الذكور: المتوسط= 71.5 الوسيط = 72 المنوال = 77

مجموعة الطالبات الإناث: المتوسط= 71.5 الوسيط = 72 المنوال = 77

مناقشة: بالنظر إلى علامات المجموعتين نلاحظ للوهلة الأولى أن علاماتهم متماثلة، ولكن لو تفحصنا علامات كل طالب وطالبة لوجدنا أن هناك فرقاً كبيراً بين الطلاب والطالبات، وأن علامات الطالبات تنتشر وتتباعد عن بعضها البعض بشكل أكبر وبمدى أوسع من علامات الطلاب، وبالتالي فالاختلاف بين المفردات صفة لا يمكن أن تظهرها مقاييس النزعة المركزية بل تظهرها مقاييس التشتت.

ومنه فإن مقاييس التشتت تفيد في:

1-تزودنا بمعلومات حول تبعثر (تشتت) أو تجمع البيانات داخل التوزيع، وحول المتوسط الحسابي.

2-تعد مقاييس التشتت مكملاً لمقياس النزعة المركزية في وصف البيانات.

3-تعد مؤشراً على مدى كفاءة وفاعلية مقياس النزعة المركزية في تمثيل البيانات، فمثلاً كلما تجمعت القيم حول المتوسط الحسابي كان هذا المتوسط يمثل بشكل جيد هذه القيم والعكس صحيح، بمعنى أنه كلما كان مقياس التشتت صغيراً كان هذا مؤشراً على أن مقياس النزعة المركزية يمثل بياناته أصدق تمثيل.

وهناك مقاييس عدة للتشتت: المدى والمدى الربيعي والانحراف الربيعي والانحراف المتوسط والانحراف المعياري والتباين. وسيتم التركيز على كل من المدى والانحراف المعياري والتباين.

أولاً-المدى Range:

يعد المدى أسهل وأبسط مقياس من مقاييس تشتت البيانات الكمية، ويعرف **المدى بأنه الفرق بين أكبر قيمة وأصغر قيمة في البيانات**، ويفيد المدى في إعطاء صورة مبسطة للتشتت في الدرجات والتفاوت بينها، ويرمز له اختصاراً بالرمز (R).

<u>-حساب المدى:</u>

1-حساب المدى لبيانات غير مبوبة:

المدى (R) = أكبر قيمة – أصغر قيمة

مثال 3: احسب المدى للبيانات الآتية: 5، 10، 9، 15، 14، 3، 7

المدى (R) = أكبر قيمة – أصغر قيمة = 15 – 3 = 12

بالعودة إلى المثال 2 إذ كانت:

علامات الطلاب الذكور: 77 – 77 – 77 – 74 – 73 – 71 – 68 – 67 – 66 – 65

علامات الطالبات الإناث: 100 – 94 – 84 – 77 – 77 – 67 – 63 – 57 – 55 – 41

وعند حساب المدى لعلامات الطلاب الذكور = 77 – 65 = 12

وعند حساب المدى لعلامات الطالبات الإناث = 100 – 41 = 59

مناقشة: نجد أن المدى لعلامات الطلاب أصغر منه لعلامات الطالبات، وهذا يعني أن تشتت علامات الطلاب أصغر من تشتت علامات الطالبات، بمعنى أوضح أن الطلاب كانت علاماتهم متقاربة أكثر من علامات الطالبات التي كانت أكثر تباعداً أو تناثراً.

مثال 4: أوجد المدى لكل مجموعة من المجموعتين الآتيتين:

المجموعة الأولى: 7 – 4 – 6 – 7 – 7 – 6 – 6 – 20

المجموعة الثانية: 20 – 14 – 16 – 5 – 9 – 4 – 18 – 12

المدى للمجموعة الأولى = 20 – 4 = 16 المدى للمجموعة الثانية = 20 – 4 = 16

مناقشة: إن المدى لكلا المجموعتين متساوٍ ولكن عند النظر إلى بيانات المجموعة الثانية نجد أن الاختلاف بين بياناتها أكبر من الاختلاف بين بيانات المجموعة الأولى، فأغلب بيانات مجموعة الأولى متقاربة تتألف من الرقمين 6 و7 في معظمها، لذلك يمكن أن يكون المدى مقياساً مضللاً في بعض الأحيان لاعتماده على القيمتين المتطرفتين اللتين غالباً ما تكونان شاذتين أو متطرفتين.

2–حساب المدى لفئات القيم:

يحسب المدى لفئات القيم من خلال: **المدى (R) = الحد الأعلى للفئة الأخيرة – الحد الأدنى للفئة الأولى**

مثال 5: احسب المدى لبيانات 20 طالباً في مادة العلوم:

الجدول (2) بيانات 20 طالباً في مادة العلوم

التكرار	الفئات
2	10 – 14
3	14 – 18
5	18 – 22
6	22 – 26
4	26 – 30
$\Sigma f = 20$	

الحد الأدنى للفئة الأولى

الحد الأعلى للفئة الأخيرة

المدى (R) = الحد الأعلى للفئة الأخيرة – الحد الأدنى للفئة الأولى = 30 – 10 = 20

-خواص المدى:

- يعد المدى أسهل وأبسط مقاييس التشتت، ولذلك فهو يستخدم في الأوساط العامة.
- يعطي فكرة سريعة عن طبيعة البيانات ويستخدم كثيراً في مراقبة جودة الإنتاج، وكذلك في وصف الأحوال الجوية.
- يتأثر بالقيم المتطرفة أو الشاذة، لذلك لا يفضل استخدامه في حال وجود قيم متطرفة.
- لا يمكن حسابه من الجداول التكرارية المفتوحة.
- لا يصلح استخدامه إذا كان التوزيع البياني شديد الالتواء.

ثانياً-الانحراف المعياري Standard deviation:

يعد الانحراف المعياري أهم مقاييس التشتت وهو يقوم في جوهره على حساب انحرافات الدرجات عن متوسطها كما تدل تسميته عليه. **فالانحراف المعياري هو الجذر التربيعي لمجموع مربعات انحرافات القيم عن متوسطها مقسوماً على عددها.** في حين أن **مجموع مربعات انحرافات القيم عن متوسطها مقسوماً على عددها يدعى التباين.** ومنه فإن الانحراف المعياري هو الجذر التربيعي للتباين، أو بمعنى آخر إن التباين هو مربع الانحراف المعياري، ويرمز للانحراف المعياري اختصاراً بالرمز (S)، ويرمز للتباين بالرمز (S^2).

-حساب الانحراف المعياري:

هناك طرائق عدة لحساب الانحراف المعياري منها الطريقة أو الأسلوب المباشر، أسلوب المربعات، أسلوب الانحرافات، وأسلوب الانحرافات المختصرة، وسنقتصر على حساب الانحراف المعياري للبيانات المبوبة وغير المبوبة وفق الأسلوب المباشر.

1-حساب الانحراف المعياري للبيانات غير المبوبة (وفق الأسلوب المباشر):

نحسب الانحراف المعياري وفق الأسلوب المباشر للبيانات غير المبوبة وفق الخطوات الآتية:

- نحسب المتوسط الحسابي للقيم $\bar{X}$.
- نحسب انحراف القيم عن المتوسط وذلك بطرح كل قيمة من متوسط القيم.
- نربع كل انحراف من انحرافات القيم.

- نحسب مجموع مربع انحرافات القيم.
- نحسب الانحراف المعياري وفق المعادلة الآتية:

$$S = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n}}$$

التباين يساوي

$$S = \frac{\Sigma(X-\bar{X})^2}{n}$$

مثال 6: لتكن لدينا مجموعة من القيم والمطلوب حساب الانحراف المعياري.

الجدول (3) حساب الانحراف المعياري لمجموعة من القيم

الدرجات X	$X-\bar{X}$	$(X-\bar{X})^2$
5	−2	4
7	0	0
9	+2	4
4	−3	9
10	+3	9
6	−1	1
8	+1	1
$\Sigma x = 49$ $\bar{X} = 7$	0	28

1-نحسب المتوسط= مجموع القيم/عددها

2-نحسب انحراف القيم عن المتوسط

3-نربع انحرافات القيم

4-نجمع مربع انحرافات القيم

5-نطبق معادلة الانحراف المعياري

لاحظ أن انحراف القيم عن المتوسط يساوي الصفر

الحل:

- نحسب المتوسط الحسابي للقيم وهو مجموع القيم على عددها ويساوي 7.
- نحسب انحراف القيم عن المتوسط $(X-\bar{X})$ وذلك بطرح كل قيمة من متوسط القيم.
- نربع كل انحراف من انحرافات القيم.
- نحسب مجموع مربع انحرافات القيم.
- نحسب الانحراف المعياري وفق المعادلة الآتية:

$$S = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n}} = \sqrt{\frac{28}{7}} = \sqrt{4} = 2$$

بما أن الانحراف المعياري يساوي 2، فإن التباين يساوي 4.

2–حساب الانحراف المعياري من فئات الدرجات (بيانات مبوبة):

نحسب الانحراف المعياري وفق الأسلوب المباشر للبيانات المبوبة في فئات وفق الخطوات الآتية:

- نحسب مركز كل فئة.
- نحسب المتوسط الحسابي للفئات السابقة وفق المعادلة الآتية:

$$\bar{X} = \frac{\Sigma(x_i f)}{\Sigma f}$$

- نطرح مركز كل فئة من المتوسط $(X - \bar{X})$.
- نربع انحراف مركز كل فئة عن المتوسط $(X - \bar{X})^2$.
- نضرب مربع الانحراف بتكرار كل فئة $f(X - \bar{X})^2$ ثم نجمع العمود .
- نحسب الانحراف المعياري وفق المعادلة الآتية:

$$\mathbf{S} = \sqrt{\frac{\boldsymbol{\Sigma f(X_i - \bar{X})^2}}{\boldsymbol{\Sigma f}}}$$

مثال 7: لتكن لدينا مجموعة من فئات القيم والتي تمثل درجات 16 طالباً في مادة دراسية والمطلوب حساب الانحراف المعياري

1-نحسب مركز الفئة

4-نربع الانحرافات

الجدول (4) حساب الانحراف المعياري لمجموعة من الفئات

5-نضرب مربع الانحراف بالتكرار

الفئات	f	X_i	fX_i	$X_i - \bar{X}$	$(X - \bar{X})^2$	$f(X - \bar{X})^2$
2 – 4	4	3	12	−2.87	8.23	32.92
4 – 6	5	5	25	−0.87	0.75	3.75
6 – 8	3	7	21	1.13	1.27	3.81
8 – 10	4	9	36	3.13	9.79	39.16
$\bar{X}$=5.87	$\sum$f=16		94		20.04	79.64

2-نوجد المتوسط

6-نطبق معادلة الانحراف المعياري

3-نطرح مركز الفئة من المتوسط

الحل:

- نحسب مركز كل فئة.
- نحسب المتوسط الحسابي للفئات السابقة وفق المعادلة الآتية:

$$\bar{X} = \frac{\Sigma(x_i f)}{\Sigma f} = \frac{94}{16} = 5.87$$

نطرح مركز كل فئة من المتوسط $(X - \bar{X})$.

- نربع انحراف مركز كل فئة عن المتوسط $(X - \bar{X})^2$.

- نضرب مربع الانحراف بتكرار كل فئة $f(X-\bar{X})^2$ ثم نجمع العمود .
- نحسب الانحراف المعياري وفق المعادلة الآتية:

$$S=\sqrt{\frac{\Sigma f(X_i-\bar{X})^2}{\Sigma f}}=\sqrt{\frac{79.64}{16}}=\sqrt{4.97}=2.22$$

-الانحراف المعياري النسبي (معامل الاختلاف) Coefficient of Variation:

يعد معامل الاختلاف (C.V) من أهم مقاييس التشتت النسبية، ويستخدم بشكل واسع في التحليل الإحصائي **لمقارنة مقدار تشتت مجموعتين (أو أكثر) من القيم** إذا كانت تختلف في متوسطها أو في وحدات قياسها، وحتى عندما تكون وحدات القياس غير مختلفة. **وهو عبارة عن النسبة المئوية للانحراف المعياري منسوبة إلى المتوسط الحسابي** وبالطبع كلما كانت قيمة هذا المعامل كبيرة دل ذلك على وجود تشتت كبير بين مفردات التوزيع والعكس صحيح، ويحسب وفق المعادلة الآتية:

$$\textbf{معامل الاختلاف النسبي} = \frac{\text{الانحراف المعياري}}{\text{المتوسط الحسابي}}\times 100$$

$$C.V=\frac{S}{\bar{X}}\times 100$$

مثال 8: قارن بين تشتت درجات مجموعتين من الطلاب في مادة دراسية، وحدد أيهما أكثر تشتتاً.

العينة الأولى: المتوسط 50، الانحراف المعياري 10

العينة الثانية: المتوسط 45، الانحراف المعياري 8

لا نستطيع الاعتماد على قيمة الانحراف المعياري فقط للقول إن تشتت درجات المجموعة الأولى أكبر من تشتت درجات المجموعة الثانية، وذلك بسبب اختلاف المتوسطات الحسابية في المجموعتين، لذلك سيصبح من الضروري استخدام معامل الاختلاف النسبي لغرض المقارنة، كما يأتي:

معامل الاختلاف النسبي (للمجموعة الأولى) = $\frac{10}{50}\times 100 = 20$%

معامل الاختلاف النسبي (للمجموعة الثانية) = $\frac{8}{45}\times 100 = 17.77$%

وبالاعتماد على معامل الاختلاف النسبي نستطيع القول بأن تشتت درجات المجموعة الأولى أكبر من تشتت درجات المجموعة الثانية لأن معامل اختلافها أكبر، في حين أنه كان من الممكن أن نصل إلى استنتاج خاطئ لو اعتمدنا على قيم الانحراف المعياري فقط.

-ملاحظات على الانحراف المعياري:

- الانحراف المعياري حساس لبعد أو قرب العلامات من المتوسط ولذلك كلما صغرت قيمته دل ذلك على طبيعة البيانات مقاربة ومتراكمة حول المتوسط، وبالتالي التشتت قليل والعكس هو الصحيح.
- يأخذ في الحسبان جميع القيم في حسابه وبالتالي يستخدم عند المقارنة بين المجموعات والعينات الإحصائية والاستنتاجات الإحصائية المتعلقة بالفرضيات.
- يتصف بالكثير من الخواص الجبرية، ولذلك فاستخداماته كثيرة أكثر من أي مقياس آخر للتشتت.
- يتأثر الانحراف المعياري بالقيم الشاذة أو المتطرفة.
- قيمته دوماً أكبر من أي مقياس تشتت آخر ما عدا المدى.
- قيمته دوماً أصغر في العينة عنه في المجتمع الذي سحبت منه هذه العينة.
- يصعب إيجاد الانحراف المعياري في حالة الفئات المفتوحة.

-الدرجة المعيارية Z Scores:

إن حساب الدرجات المعيارية مهم جداً عندما نريد مقارنة تحصيل متعلّم واحد في أكثر من مادة، ومقارنة القيم التي تنتمي لأكثر من مجموعة إحصائية والتي لا تصلح للمقارنة أساساً، كأن نقارن تحصيل طالبين من شعبتين مختلفتين في مادة ما، أو المقارنة بين دخل موظفين اثنين في بلدين.

هنا المقارنة المباشرة غير عادلة ولا يمكن إجراؤها من خلال القيم الظاهرة، ولابد من إجراء عملية تحويلية لهذه القيم، وتحويلها إلى درجات معيارية يمكن أن نقيم المقارنة من خلالها. إن **الدرجة المعيارية للعلامة الأصلية: هي عدد الانحرافات التي تنحرفها العلامة الأصلية عن الوسط الحسابي لعلامات المجموعة، ويرمز لها الرمز (ز) أو (Z).**

ويمكن تحويل الدرجة الخام إلى درجة معيارية من خلال المعادلة الآتية:

$$Z = \frac{X-\overline{X}}{S}$$

إذ Z: الدرجة المعيارية X: الدرجة الخام $\overline{X}$: المتوسط S: الانحراف المعياري.

مثال 9: حصل طالب في مادة الإحصاء على درجة قدرها 85، علماً أن متوسط درجات جميع الطلاب في مادة الإحصاء 71 درجة وانحراف معياري 10، وحصل هذا الطالب في مادة مفاهيم هندسية على درجة قدرها

82 علماً أن متوسط درجات جميع الطلاب في مادة مفاهيم هندسية 77 درجة وانحراف معياري 12، في أي من المقررين كان تحصيله أفضل؟

الحل: عند إجراء المقارنة للدرجات الخام نجد أن تحصيل الطالب في مادة الإحصاء أفضل من تحصيله في مادة مفاهيم هندسية، ولكن حتى تكون المقارنة عادلة لا بد من تحويل الدرجات الخام إلى درجات معيارية وفق الآتي:

مادة الإحصاء $Z1 = \frac{X1-X1}{S1} = \frac{85-71}{10} = 1.4$

مادة مفاهيم هندسية $Z2 = \frac{X2-X2}{S2} = \frac{82-77}{12} = 0.41$

يتضح من خلال مقارنة الدرجات المعيارية أن تحصيل الطالب أعلى في مادة الإحصاء مقارنة مع مادة مفاهيم هندسية، لأن الدرجة المعيارية لمادة الإحصاء Z1 أكبر من الدرجة المعيارية لمادة مفاهيم هندسيةZ2 .