المحاضرة التاسعة

الموضوع : **مقاييس التشتت (الاختلاف)**

Measures of Dispersion (Variation)

عنوان المحاضرة : **التباين و الإنحراف المعياري**

عناصر المحاضرة :

– مقدمة

نشاط

– التباين والإنحراف المعياري: تعاريف وملاحظات .

نشاط

– حساب التباين والإنحراف المعياري للبيانات غير المبوبة .

– نشاط .

تطبيق

أهداف المحاضرة :

خاص : في نهاية المحاضرة

– يتعرف الطالب على مفهومي الإنحراف المعياري والتباين لبيانات سلسلة احصائية غير مبوبة ويحسبهما .

إجرائي

1– يحسب كلا من التباين والإنحراف المعياري باعتماد علاقتهما الرياضية لبيانات غير مبوبة من دون خطأ .

2– يصف اختلاف وتشتت قيم السلسلة الإحصائية معتمدا على قيم الإنحراف المعياري والتباين من دون خطأ .

ملخص المحاضرة

تتضمن المحاضرة إستعراضا لمفاهيم تشتت البيانات الإحصائية غير المبوبة إنطلاقا من تباينها وانحرافها المعياري من خلال تقديم العلاقات الرياضية الخاصة بكل من الإنحراف المعياري والتباين وتطبيقها ضمن أمثلة .

مقدمة :

ذكرنا في الدروس السابقة بعضًا من مقاييس النزعة المركزية والتي تمثل مقاييس عددية لموضع

تمركز البيانات لظاهرة ما. وقد ذكرنا بأن هذه المقاييس تستخدم لمقارنة مجموعات البيانات المختلفة. وفي الحقيقة فإن مقاييس النزعة المركزية غير كافية لإيجاد مقارنة شاملة بين مجموعات البيانات المختلفة. فقد تكون هناك مجموعات من البيانات لها نفس مقاييس النزعة المركزية (لها نفس الموضع) ولكنها تختلف في بعض الصفات الأخرى.

المثال التالي يبين لنا مجموعتين من البيانات لهما نفس المتوسط ولكنهما مختلفتان في طبيعة تشتت البيانات.

سؤال :

1- ماذا نعني بمقاييس التشتت .

2- أذكر مقاييس التشتت التي درست

الجواب :

مقاييس التشتت هي مقاييس عددية تستخدم لقياس اختلاف أو تشتت البيانات. والاختلاف أو التشتت لمجموعة من البيانات هو مقدار تفرق أو تباعد أو انتشار البيانات فيما بينها. فتشتت البيانات يكون صغيرًا إذا كانت البيانات متقاربة فيما بينها والعكس بالعكس. وأما البيانات المتساوية فلا اختلاف ولا تشتت فيها. ومقاييس التشتت تستخدم لوصف مجموعة البيانات وكذلك لمقارنة مجموعات البيانات المختلفة إذ أن مقاييس النزعة المركزية لا تكفي وحدها لوصف مجموعة البيانات أو مقارنة مجموعات البيانات المختلفة.

يعتبر التباين والانحراف المعياري من أهم وأفضل مقاييس التشتت ومن أكثرها شيوعًا واستخدامًا في التحليل الإحصائي وذلك لما يتمتعان به من خصائص وصفات إحصائية جيدة.

1- المدى: Range

2- نصف المدى الربيعي Semi-Inter-quartile Range

المقطع الأول : التباين (Variance) والانحراف المعياري (Standard Deviation):

نشاط 1 :

أوجد متوسط ووسيط كلا من السلسلتين : 59, 61, 62, 60, 58, 60 و 50, 60, 66, 54, 60, 70

الحل :

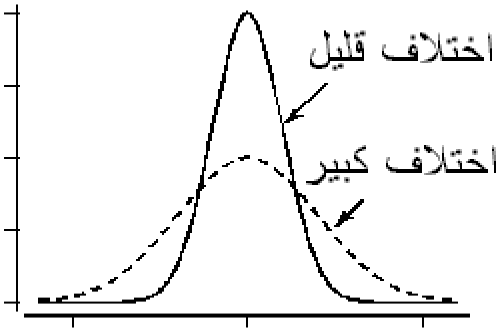
الجدول التالي يوضح مجموعتين لهما نفس المتوسط $\bar{x} = 60$ ونفس الوسيط $Med = 60$ ونفس المنوال $Mod = 60$ ولكنهما مختلفتان في طبيعة التشتت.

المجموعة	البيانات	شكل انتشار البيانات
الأولى	59, 61, 62, 60, 58, 60	– • – • – • – • – • – 60
الثانية	50, 60, 66, 54, 60, 70	– • – – • – – – – • – – – – • – – • – 60

لا حظ أنه بالرغم من تساوي مقاييس النزعة المركزية للمجموعتين إلا أن التشتت (أو الاختلاف) بين القيم في كل مجموعة غير متساو. فمن الواضح أن بيانات المجموعة الأولى أكثر تقاربًا فيما بينها (أقل تشتتًا وتباعدًا فيما بينها) من بيانات المجموعة الثانية. لذلك دعت الحاجة لإيجاد مقاييس تقيس طبيعة تشتت (أو تفرق أو اختلاف أو تباعد) البيانات فيما بينها. هذه المقاييس تسمى مقاييس التشتت أو الاختلاف.

مثال توضيحي :

الشكل التالي يوضح المضلعين التكراريين لمجموعتين من البيانات لهما نفس مقاييس النزعة المركزية ولكنهما مختلفتان في طبيعة التشتت.



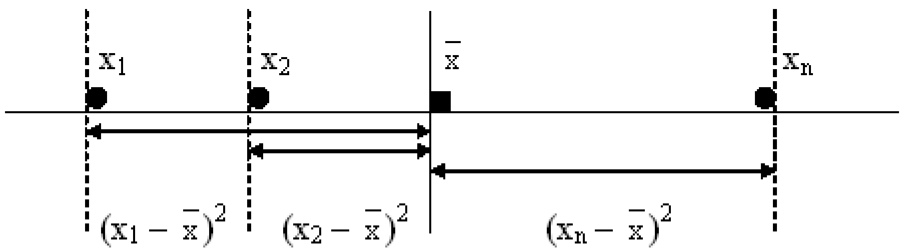
المضلعان التكراريان لتوزيعين لهما نفس مقاييس
النزعة المركزية ولكنهما مختلفين في التشتت

التباين: Variance

فكرة التباين تعتمد على تشتت أو تباعد البيانات عن متوسطها. فالتباين يكون كبيرًا إذا كانت البيانات متباعدة عن متوسطها والعكس بالعكس.

ويعرف التباين بأنه متوسط مربع انحرافات القيم عن الوسط الحسابي ويرمز له بالرمز S^2.

الشكل التالي يبين مربعات انحرافات القيم عن الوسط الحسابي.



والجدول التالي يلخص طريقة حساب انحرافات ومربعات انحرافات القيم عن المتوسط.

القيم (البيانات)	x_n	...	x_2	x_1
انحرافات القيم عن المتوسط	$x_n - \bar{x}$	...	$x_2 - \bar{x}$	$x_1 - \bar{x}$
مربع انحرافات القيم عن المتوسط	$(x_n - \bar{x})^2$	...	$(x_2 - \bar{x})^2$	$(x_1 - \bar{x})^2$

تستخدم مربعات انحرافات القيم عن الوسط الحسابي لحساب التباين. فمن الواضح أن تشتت البيانات يزداد بزيادة مربعات الانحرافات والعكس بالعكس. والسؤال الذي يتبادر للأذهان هو أي واحد من هذه المربعات ينبغي أن يستخدم لقياس التشتت؟ إن من المنطقي أن نبحث عن قيمة نموذجية تمثل هذه المربعات لاستخدامها لقياس التباين. وبناءً على دراستنا السابقة لمقاييس النزعة المركزية فإن المتوسط من أفضل المقاييس التي تمثل مجموعة البيانات. وعليه فإن متوسط مربعات انحرافات القيم عن الوسط الحسابي $\sum_{i=1}^{n}(x_i - \bar{x})^2 / n$ يعتبر أحد المقاييس التي يمكن استخدامها لقياس التباين.

الانحراف المعياري: Standard Deviation

إن التباين من أهم وأفضل مقاييس التشتت ولكنه يقاس بوحدة البيانات الأصلية المربعة. وفي كثير من الأحيان نرغب في استخدام مقياس للتشتت يقاس بوحدة البيانات الأصلية ويتمتع بخصائص إحصائية جيدة مثل التباين. وأحد هذه المقاييس هو الانحراف المعياري. ويعرف الانحراف المعياري على أنه الجذر التربيعي الموجب للتباين ويرمز له بالرمز s.

المقطع الثاني : حساب التباين والانحراف المعياري:

نستعرض في هذا الدرس طرق حساب التباين والانحراف المعياري في حالة البيانات المفردة غير المبوبة

التباين والانحراف المعياري للبيانات المفردة (غير المبوبة):

إذا كانت $x_1, x_2, \ldots, x_n$ عينه حجمها n وكان متوسطها هو $\bar{x}$ فإن تباين العينة يعرف كما يلي:

$$s^2 = \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \cdots + (x_n-\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$$

لاحظ أننا قسمنا على المقدار $(n-1)$، وهو ما يسمى بدرجات الحرية، بدلاً من القسمة على عدد البيانات n في الصيغة السابقة وذلك لكي نحصل على مقياس يتمتع بصفات إحصائية جيدة.

وأما الانحراف المعياري فإنه يعرف بالصيغة:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$$

ملاحظة

1. $s^2 \geq 0$ (دائمًا) وكذلك $s \geq 0$ (دائمًا).
2. $s^2 = 0 \Leftrightarrow s = 0 \Leftrightarrow$ جميع قيم العينة متساوية (لا يوجد اختلاف بين القيم).
3. وحدة التباين، s^2، هي وحدة البيانات الأصلية المربعة.
4. وحدة الانحراف المعياري، s، هي نفس وحدة البيانات الأصلية.
5. يمكن حساب التباين بإحدى الصيغتين الحسابيتين التاليتين:

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right\}$$

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2\right\}$$

ولحساب تباين العينة باستخدام الصيغتين الحسابيتين السابقتين فإننا نحتاج إلى معرفة الكميات التالية فقط دون الحاجة لمعرفة البيانات الأصلية:

1. حجم العينة = n.
2. مجموع البيانات = $\sum_{i=1}^{n} x_i$.
3. مجموع مربعات البيانات = $\sum_{i=1}^{n} x_i^2$.

والصيغتين الحسابيتين السابقتين تستخدمان لحساب تباين العينة وذلك لسببين هما:

1. لأنهما أكثر سهولة من صيغة التعريف.
2. لأنها أكثر دقة في الحساب من صيغة التعريف عندما يكون هناك تقريب في حساب متوسط

العينة.

مثال مباشر :

أوجد تباين العينة والانحراف المعياري لمجموعة الأوزان (بالكيلوجرام) التالية :

7.1, 2.5, 2.5, 5.4, 8.3

الحل:

نلخص الحل في الجدول التالي:

x_i	الانحراف $(x_i-\bar{x})=(x_i-5.16)$	مربع الانحراف $(x_i-\bar{x})^2$	x^2
7.1	1.94	3.7636	50.41
2.5	−2.66	7.0756	6.25
2.5	−2.66	7.0756	6.25
5.4	0.24	0.0576	29.16
8.3	3.14	9.8596	68.89
$\sum_{i=1}^{n} x_i = 25.8$	0.00	$\sum_{i=1}^{n}(x_i-\bar{x})^2 = 27.8$	$\sum_{i=1}^{n} x_i^2 = 160.9$

من هذا الجدول نوجد الكميات التالية:

$$n = 5,\ \sum_{i=1}^{n} x_i = 25.8\ ,\ \sum_{i=1}^{n}(x_i-\bar{x})^2 = 27.832\ ,\ \sum_{i=1}^{n} x_i^2 = 160.96$$

إن متوسط العينة هو:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{25.8}{5} = 5.16$$ (كيلوجرامًا)

حساب تباين العينة:

(أ) باستخدام التعريف:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{27.832}{5-1} = 6.958$$ (كيلوجرامًا مربعًا)

(ب) باستخدام الصيغة الحسابية الأولى:

$$s^2=\frac{1}{n-1}\left\{\sum_{i=1}^{n}x_i^{\ 2}-\left(\sum_{i=1}^{n}x_i\right)^2/\ n\right\}=\frac{1}{5-1}\left\{160.96-(25.8)^2/\ 5\right\}$$

$$=\frac{160.96-133.128}{4}=\frac{27.832}{4}=6.958$$

(ج) باستخدام الصيغة الحسابية الثانية:

$$s^2=\frac{1}{n-1}\left\{\sum_{i=1}^{n}x_i^{\ 2}-n\ (\bar{x})^2\right\}=\frac{1}{5-1}\left\{160.96-(5)(5.16)^2\right\}$$

$$=\frac{160.96-133.128}{4}=\frac{27.832}{4}=6.958$$

الانحراف المعياري هو:

$$s=\sqrt{s^2}\ =\ \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n\ -1}}\ =\ s=\sqrt{6.958}=2.6378$$ (كيلوجرامًا)

نشاط موضعي :

أجب بصحيح أو خطأ

1- تباين سلسلة إحصائية يكون دائما موجب .

2- الانحراف المعياري لا يكون دائما أصغر من التباين.

3- إذا كان الانحراف لسلسلة معدوما فإن كل قيم هذه السلسلة متساوية.

اختر الجواب الصحيح

نعتبر سلسلتين إحصائيتين S_1 و S_2 حيث وسط S_1 هو5 ، تباين S_1 هو 1 وسط S_2 هو9 ، تباين S_2 هو 1

V تباين سلسلة S تتضمن قيم S_1 وقيم S_2.

V=.........

1) 1

2) 1,5

3) لا يستطيع حساب V (لا يوجد معطيات كافية)

نشاط اجمالي :

استخرجت كشفا بأوقات المكالمات الهاتفية (بالثواني) التي أجريتها في مدة أسبوع ، فكانت النتائج كالآتي :

$-20-78-28-69-96-77-181-210-6-57-3-93-134-67-224-37-68$

$-8-111-12-118-88-87-12-89-45-32-9-1-123-258-28-7-39-101$

$58-27-61-94-86-42-98-7-32-47-76-107-27-75-47$

1) احسب المتوسط الحسابي، و التباين والانحراف المعياري لهذه السلسلة.

2) أملئ الجدول التالي :

الوقت	التكرار
$[0,20[$	
$[20,40[$	
$[40,60[$	
$[60,80[$	
$[80,100[$	
$[100,120[$	
$[120,160[$	
$[160,200[$	
$[200,260[$	

3) احسب المتوسط الحسابي، و التبيان والانحراف المعياري لهذه السلسلة بعد جمع النتائج في الجدول السابق .

4) أملئ الجدول التالي باعتماد البيانات السابقة :

الوقت	التكرار
$[0,30[$	
$[30,60[$	
$[60,90[$	
$[90,120[$	
$[120,240[$	

5) أحسب المتوسط الحسابي ، و التبيان والانحراف المعياري للسلسلة باستعمال النتائج الجدول الثاني.

6) لماذا نحصل على نتائج مختلفة برغم أننا استعملنا نفس السلسلة الإحصائية؟