الشركة القابضة
لمياه الشرب والصرف الصحى

برنامج المسار الوظيفي
املين بقطاع مياه الشرب والصرف الصحي

دليل المتدرب

برنامج

الإحصاء المتقدم

اخصائي تحليل بيانات – الدرجة الثالثة

تم إعداد المادة بواسطة الشركة القابضة لمياه الشرب والصرف الصحي
قطاع تنمية الموارد البشرية - الادارة العامة لتخطيط المسار الوظيفي
الإصدار الأول - ٢٠٢٣.

الفهرس

المحور الأول :مقاييس التشتت (الاختلاف)

Measures of Dispersion (Variation)

أولا - مفهوم مقاييس التشتت :

ان مقاييس النزعة المركزية والتي تمثل مقاييس عددية لموضع أو مكان تركز 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. فمثلاً المثال التالي يبين لنا مجموعتين من البيانات لهما نفس المتوسط ولكنهما مختلفتان في طبيعة تشتت البيانات.

مثال (١):

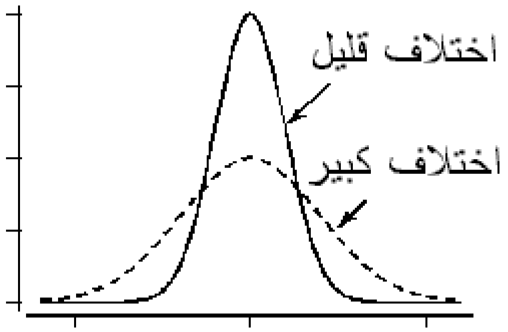
الجدول التالي يوضح مجموعتين لهما نفس المتوسط $\bar{x} = 60$ ونفس الوسيط $Med = 60$ ونفس المنوال $Mod = 60$ ولكنهما مختلفتان في طبيعة التشتت.

لمجموعة	البيانات	شكل انتشار البيانات
الأولى	59, 61, 62, 60, 58, 60	• — ● — ● — ●— ● — ● — 60
الثانية	50, 60, 66, 54, 60, 70	• — ● — — ● — — — — ●— — — — ● — — ● — 60

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 (٢):

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



شكل (١) المضلعان التكراريان لتوزيعين لهما نفس مقاييس النزعة المركزية ولكنهما مختلفين في التشتت

ثانيا - تعريف مقاييس التشتت :

مقاييس التشتت هى 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.

ثالثا – أنواع مقاييس التشتت :

ومن أشهر مقاييس التشتت نذكر:

1. **المدى: Range**
2. **نصف المدى الربيعي Semi-Inter-quartile Range**
3. **التباين: Variance**
4. **الانحراف المعياري: Standard Deviation**
5. **معامل الاختلاف (أو التغير): Coefficient of Variation**

وسنتناول شرح كلا منها بالتفصيل

١- المدى: Range

يعتبر المدى من أسهل مقاييس التشتت تعريفًا وحسابًا ويعطينا فكرة سريعة عن مدى تفرق البيانات.

تعريف :

نعرف المدى لمجموعة من البيانات على أنه الفرق بين أكبر قيمة وأصغر قيمة، أي أن المدى هو:

$$Range = X_{max} - X_{min}$$

حيث نعرف X_{max} و X_{max} كما يلي:

(أ) للبيانات المفردة:

X_{max} = أكبر قيمة

X_{min} = أصغر قيمة

(ب) للبيانات المبوبة:

X_{max} = مركز الفترة العليا

X_{min} = مركز الفترة الدنيا

ملاحظة (١):

تعرف بعض الكتب أكبر قيمة وأصغر قيمة للبيانات المبوبة كما يلي:

X_{max} = الحد الأعلى للفترة العليا

X_{min} = الحد الأدنى للفترة الدنيا

مثال (١):

أوجد المدى للمشاهدات التالية والتي هي عبارة عن أوزان (بالكيلوجرام) مجموعة مكونة من سبعة أشخاص: 50 ,55 ,35 ,45 ,40 ,30 ,25

الحل:

$$X_{max} = 55$$

$$X_{min} = 25$$

$$Range = X_{max} - X_{min} = 55 - 25 = 30 \quad \text{(كيلوجرامًا)}$$

حساب المدى من البيانات الغير مبوبة

المدى = أكبر قيمة – أصغر قيمة

<u>مثال (٢) :</u>

احسب المدى للبيانات التالية :

٩٥ – ٢٠٠ – ٢٥٠ – ٣٠٠ – ١١٠ – ٩٠ – ١٥٠ – ١٠٠ – ٣٥٠ – ٨٠

الحل :

نرتب القيم أولاً : (٨٠-٩٠-٩٥-١٠٠-١١٠-١٥٠-٢٠٠-٢٥٠-٣٠٠-٣٥٠)

المدى = ٣٥٠ – ٨٠ = ٢٧٠

حساب المدى من البيانات المبوبة

المدى = الحد الأعلى للفئة الأخيرة – الحد الأدنى للفئة الأولى

<u>مثال (٣):</u>

احسب المدى للجدول التالي :

الفئات	١٦-	٢٠-	٢٤-	٢٨-	٣٢-٣٦
عدد المبحوثين	١٠	١٥	٤٠	٢٠	١٥

الحل :

المدى = الحد الأعلى للفئة الأخيرة – الحد الأدنى للفئة الأولى

المدى = ٣٦ – ١٦ = ٢٠

<u>بعض مميزات وعيوب المدى:</u>

أ- يتميز المدى بسهولة التعريف والحساب

ب- يعيب المدى العيوب التالية:

1. **يتأثر المدى بالقيم الشاذة أو المتطرفة.**
2. **لا يأخذ المدى في الاعتبار جميع البيانات.**

<u>ملاحظة (٢):</u>

1. **وحدة المدى هي نفس وحدة البيانات الأصلية.**
2. **نظرًا لأن المدى يعتمد فقط على أكبر وأصغر قيمة ولا يأخذ في الاعتبار القيم الأخرى فهو مقياس غير جيد لقياس التشتت.**

٢- نصف المدى الربيعي Semi-Inter-quartile Range:

رأينا أن المدى يتأثر كثيرًا بالقيم الشاذة أو المتطرفة. ولذلك دعت الحاجة لإيجاد مقاييس أخرى للتشتت لا تتأثر بالقيم المتطرفة. وأحد هذه المقاييس هو نصف المدى الربيعي. وحيث أن القيم المتطرفة هي تلك القيم الصغيرة جدًا أو الكبيرة جدًا فإنه عند حساب نصف المدى الربيعي لا يؤخذ في الاعتبار ربع البيانات الصغيرة (25%) ولا ربع البيانات الكبيرة (25%). الشكل التالي يوضح موضع القيم الشاذة والمتطرفة.

¼ n = 25% ¼ n = 25%

-/-/-/-/-/-/-/-/-/-/-/-/-/-|----------------|----------------|-/-/-/-/-/-/-/-/-/-/-/-/-/-

ربع البيانات الصغيرة Q1 Q2 Q3 ربع البيانات الكبيرة

يرمز لنصف المدى الربيعي بالرمز *Q* ويعرف بالصيغة التالية:

$$Q_2 = \frac{Q_3 - Q_1}{2}$$

حيث أن Q_1 هو الربيع الأول و Q_3 هو الربيع الثالث **ويسمى نصف المدى الربيعي احيانا Q2**

بعض مميزات وعيوب نصف المدى الربيعي:

1. من المميزات أنه لا يتأثر بالقيم الشاذة أو المتطرفة.
2. من العيوب أنه لا يأخذ في الاعتبار جميع البيانات.

مثال (١) :

احسب ملخص البيانات التالية

٥,٨,٣,٢,١,٣,١٠

لاحظ ان عدد البيانات فردي

أولا يتم ترتيب البيانات تصاعديا من اليمين لليسار

١٠,٨,٥,٣,٣,٢,١

بعد ترتيب البيانات يكون :

Min = 1

Max=10

Q1 = 2القيمة الخاصة بالربيع الأول

Median =Q2 =3القيمة في المنتصف هي نفسها الربيع الثاني

Q2 =(Q3-Q1)/2

نصف المدي الربيعي Q2 = (٨-٢)÷٢= ٣

Q3=8القيمة الخاصة بالربيع الثالث

RANGE = MAX-MIN = 10-1=9

<u>ملخص النقاط الخمس :</u>

MIN	Q1	Q2	Q3	MAX
1	2	3	8	10

<u>مثال (٢) :</u>

احسب ملخص البيانات التالية

١٠٥,١٠,٣,١,٢,٣,٨,٥

لاحظ ان عدد البيانات زوجي

أولا يتم ترتيب البيانات تصاعديا من اليمين لليسار

١٠٥,١٠,٨,٥,٣,٣,٢,١

بعد ترتيب البيانات يكون :

Min = 1

Max=105

Q1=(٢+٣)÷٢= ٢,٥

Median =Q2=(٣+٥)÷٢= ٤

القيمة الخاصة بالربيع الثالث Q3= (٨+١٠)÷٢=٩

RANGE = MAX-MIN =105-1=104

<u>ملخص النقاط الخمس :</u>

MIN	Q1	Q2	Q3	MAX
1	٢,٥	4	9	105

٣- التباين (Variance) والانحراف المعياري (Standard Deviation):

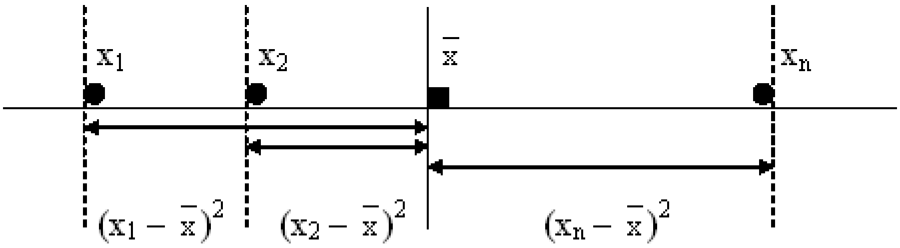
التباين (Variance) والانحراف المعيارى (Standard Deviation):

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- 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 الحسابى ويرمز لـه بـالرمز S^2. الشكل التالى يبين مربعات انحرافات القيم عن الوسط الحسابى.



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 يعتبر أحد المقاييس التي يمكن استخدامها لقياس التباين.

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

2- 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 المعيارى على أنه الجذر التربيعي الموجب للتباين ويرمز له بالرمز s.

حساب التباين والانحراف المعياري:

سنستعرض طرق حساب التباين والانحراف المعياري في حالة البيانات المفردة وفي حالة البيانات المبوبة.

أولا :التباين والانحراف المعيارى 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 المعيارى فإنه يعرف بالصيغة:

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

ملاحظة (*):

1. $0 \geq s^2$ (دائمًا) وكذلك $0 \geq s$ (دائمًا).
2. $s = 0 \Leftrightarrow s^2 = 0$ $\Leftrightarrow$ جميع قيم البيانات 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.

مثال (١):

أوجد التباين 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	$\sum_{i=1}^{n} (x_i - \bar{x})^2 = 27.83$	$\sum_{i=1}^{n} x_i^2 = 160.96$

من هذا الجدول نوجد الكميات التالية:

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

إن متوسط البيانات هو:

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

حساب تباين البيانات:

باستخدام صيغة التعريف:

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

الانحراف المعيارى هو:

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

إذا كان لدينا مجموعتان من البيانات: بحيث أن عدد بيانات المجموعة الأولى n_1 ومتوسطها $\bar{x}_1$ وتباينها s_1^2 وكان عدد بيانات المجموعة الثانية n_2 ومتوسطها $\bar{x}_2$ وتباينها s_2^2 وإذا كان $\bar{x}_1=\bar{x}_2$ (أي أن متوسطي المجموعتين متساويان) فإن تباين المجموعة الكلية المكونة من دمج هاتين المجموعتين يمكن حسابه بالصيغة التالية:

$$s^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-1}$$

مثال (٢):

أوجد تباين المجموعة الكلية المكونة من دمج المجموعتين التاليتين:

	المجموعة الأولى	المجموعة الثانية
حجم البيانات	$n_1 = 4$	$n_2 = 6$
المتوسط	$\bar{x}_1 = 5$	$\bar{x}_2 = 5$
التباين	$s_1^2 = 3$	$s_2^2 = 3.5$

الحل:

أولاً نلاحظ أن متوسطي المجموعتين متساويان.

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 1}$$

$$= \frac{(4-1)(3)+(6-1)(3.5)}{4+6-1} = 2.944$$

<u>مثال (٣)</u>:- اوجد التباين والانحراف المعياري للبيانات التالية الخاصة بالفترة الزمنية لعطل احد أجهزة القياس (1,4,5,6,3,7).

الحل:- ان اوليات جدول الحل يكون وفق الاتي

المجموع	جدول البيانات						
26	7	3	6	5	4	1	x_i
136	49	9	36	25	16	1	${x_i}^2$

ثم نعوض بقانون التباين

$$s_x^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} {x_i}^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right\} = \frac{1}{5}\left\{136 - \frac{1}{6}(26)^2\right\} = 4.666$$

وبانحراف معياري

$$s_x = \sqrt{s_x^2} = \sqrt{4.666} = 2.160$$

<u>ثانيا: التباين والانحراف المعياري في حالة البيانات المبوبة</u>:

فيحسب وفق الصيغة التالية <u>ولحالة البيانات المستمرة</u> <u>grouped frequency distribution)</u> .

$$s_x^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{k} f_i {m_i}^2 - \frac{1}{n}\left(\sum_{i=1}^{k} f_i m_i\right)^2\right\}$$

وان $n = \sum_{i=1}^{k} f_i$ مجموع التكرارات وk تمثل عدد الفئات في التوزيع ،

<u>وفي حالة كون التوزيع من نوع ((ungrouped frequency distribution)) والذي يتعامل مع المتغير المتقطع</u> يبقى القانون نفسه مع تغيير m_i بالمتغير x_i والذي يبين قيمة المتغير المتقطع للفئة i .

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

مثال(٤):- اوجد التباين والانحراف المعياري للبيانات المبوبة التالية الخاصة بأرباح عدد من الاسر الريفية المنتجة في احدي قري الصعيد ولعينة حجمها (25).

classes	f_i
12-	2
22-	5
32-	9
42-	6
52-62	3

الحل :- نكون الجدول التالي لان المتغير مستمر وبالاعتماد على مراكز الفئات

classes	f_i	m_i	$f_i m_i$	$f_i {m_i}^2$
12-	2	17	34	578
22-	5	27	135	3645
32-	9	37	333	12321
42-	6	47	282	13254
52-62	3	57	171	9747
المجموع	25		955	39545

$$s_x^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{k} f_i {m_i}^2 - \frac{1}{n}\left(\sum_{i=1}^{k} f_i m_i\right)^2\right\} = \frac{1}{24}\left\{39545 - \frac{1}{25}(955)^2\right\}$$

$$s_x^2 = 127.666$$

اما الانحراف المعياري فيمثل جذر التباين وفق الصيغة

$$s_x = \sqrt{s_x^2} = \sqrt{127.666} = 11.298$$

حيث ان mi هي عبارة عن مركز الفئة ومركز الفئة الاولي =(الفئة الاولي + الفئة الثانية) ÷ ٢ = (١٢+٢٢)÷٢ = ١٧ وهكذا لباقي الأرقام

مثال(٥):- اوجد التباين والانحراف المعياري للبيانات المبوبة التالية الخاصة بعدد الزبائن المنتظرين امام الصراف الالي في مطار القاهرة الدولي ولعينة حجمها (13) زبون .

classes	2	3	4	5
f_i	1	4	7	1

الحل :- نكون الجدول التالي لان المتغير متقطع وبالاعتماد على قيم x_i .

Classes(x_i)	f_i	$f_i x_i$	$f_i {x_i}^2$
2	1	2	4
3	4	12	36
4	7	28	112
5	1	5	25
المجموع	13	47	177

$$s_x^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{k} f_i {x_i}^2 - \frac{1}{n}\left(\sum_{i=1}^{k} f_i x_i\right)^2\right\} = \frac{1}{12}\left\{177 - \frac{1}{13}(47)^2\right\}$$

$$s_x^2 = 0.59$$

وبانحراف معياري

$$s_x = \sqrt{s_x^2} = \sqrt{0.59} = 0.77$$

٤ – معامل الاختلاف (التغير): Coefficient of Variation:

ذكرنا سابقًا أن التباين والانحراف المعياري من المقاييس المفيدة لقياس التشتت لتوزيع متغير ما. ولكن في كثير من الأحيان نكون مهتمين بمقارنة التشتت والاختلاف لتوزيعي متغيرين مختلفين. وبما أن التباين والانحراف المعياري مقياسان يعتمدان على وحدة البيانات فإنه يصعب استخدامهما لمقارنة تجانس المجموعات المختلفة من البيانات وذلك لاختلاف الوحدة المستخدمة. وبشكل عام فإن مقاييس التشتت التي ذكرناها آنفًا تكون غير مناسبة لمقارنة تجانس مجموعات البيانات المختلفة في الحالتين التاليتين:

1. إذا كانت وحدتا المتغيرين مختلفتين حيث لا نستطيع مقارنة الوحدات المختلفة ببعض.
2. إذا كان متوسطا المتغيرين مختلفين وذلك لأن تباين توزيع المتغير ذي المتوسط الصغير ينـــزع لأن يكون صغيراً والعكس بالعكس.

لذلك دعت الحاجة إلى مقياس لا يعتمد على وحدة المتغير ويقيس ما يسمى بالتشتت النسبي. وأحد هذه المقاييس هو ما يسمى بمعامل الاختلاف أو معامل التغير. فمعامل الاختلاف هو أحد مقاييس التشتت النسبي وهو مقياس عديم الوحدة ويستخدم لمقارنة التشتت النسبي أو التجانس لمجموعات البيانات المختلفة. فمجموعة البيانات ذات معامل الاختلاف الأكبر يكون تشتتها النسبي أكبر أي أنها تكون أقل تجانساً والعكس بالعكس. ويعرف معامل الاختلاف للعينة التي متوسطها $\bar{x}$ وانحرافها المعياري s بالصيغة التالية:

$$C.V. = \frac{s}{\bar{x}}$$

مثال (١):

الجدول أدناه يتضمن بيانات إحدى الدراسات التي طبقت على خمسة أشخاص لقياس الوزن (بالكيلوجرام) والطول (بالسنتيمتر). أي البيانات أكثر تشتتاً نسبياً (أقل تجانساً)؛ بيانات الأوزان أم بيانات الأطوال؟

رقم الشخص	1	2	3	4	5
الوزن	69	59	65	67	65
الطول	164	162	155	165	158

الحل:

أولاً نوجد المتوسط $\bar{x}$ والانحراف المعياري s لكل من بيانات الأوزان وبيانات الأطوال كما مر معنا سابقاً. نلخص الحسابات في الجدول التالي:

البيانات	المتوسط $\bar{x}$	الانحراف المعياري s	معامل الاختلاف $C.V. = \frac{s}{\bar{x}}$
الأوزان	65.0 kg	3.7417 kg	0.0576
الأطوال	160.8 cm	4.2071 cm	0.026

بما أن معامل الاختلاف لبيانات الأوزان أكبر من معامل الاختلاف لبيانات الأطوال فإن التشتت النسبي لبيانات الأوزان أكبر من التشتت النسبي لبيانات الأطوال. أي أن بيانات الأوزان أقل تجانساً من بيانات الأطوال.

المحور الثانى الارتباط والانحدار

١- معامل الارتباط Correlation Coefficient :

يقاس الارتباط بين متغيرين بمقياس إحصائي يسمى " معامل الارتباط " ويعكس هذا المقياس درجة أو قوة العلاقة بين المتغيرين واتجاه هذه العلاقة. وتنحصر قيمة معامل الارتباط بين + ١، - ١. فإذا كانت قيمة معامل الارتباط تساوي + ١ فمعنى ذلك أن الارتباط بين المتغيرين طردي تام، وهو أقوى أنواع الارتباط الطردي بين متغيرين. وإذا كانت قيمة معامل الارتباط تساوي – ١ فمعنى ذلك أن الارتباط بين المتغيرين عكسي تام، وهو أقوى أنواع الارتباط العكسي بين متغيرين. وإذا كانت قيمة معامل الارتباط تساوي صفر، فمعنى ذلك أنه لا يوجد ارتباط بين المتغيرين. وكلما اقتربت قيمة معامل الارتباط من + ١ أو – ١ كلما كان الارتباط قوياً، وكلما اقترب من الصفر كلما كان الارتباط ضعيفاً.

أنواع الارتباط :

بالطبع عرفنا أن قيمة معامل الارتباط محصورة في الفترة المغلقة [-١ ، ١] وتتحدد نوعية الارتباط من الجدول التالي :

قيمة معامل الارتباط	نوع الارتباط
١+	ارتباط طردي تام
من ٠,٧ إلى أقل من ١+	ارتباط طردي قوى
من ٠,٤ إلى أقل من ٠,٧	ارتباط طردي متوسط
من صفر إلى أقل من ٠,٤	ارتباط طردي ضعيف
صفر	الارتباط منعدم
١-	ارتباط عكسي تام
من -٠,٧ إلى أقل من -١	ارتباط عكسي قوى
من -٠,٠٤ إلى أقل من -٠,٧	ارتباط عكسي متوسط
من صفر إلى أقل من -٠٤.	ارتباط عكسي ضعيف

- **من اهم أنواع طرق حساب معامل الارتباط**

أ- الارتباط الخطي البسيط لبيرسون

ب- معامل ارتباط الرتب لسبيرمان

أ- الارتباط الخطي البسيط :

معامل بيرسون للارتباط الخطي البسيط

يفترض بيرسون Pearson أن المتغيرين كميان، وأن العلاقة بينهما خطية (أي تأخذ شكل خط مستقيم).

مثال (١) :

البيانات التالية تمثل أعمار ثمانية من الناخبين ودخولهم اليومية بالدولار، والمطلوب حساب معامل بيرسون للارتباط الخطي بين الأعمار والدخول.

الأعمار x : 25 32 29 43 38 51 47 35

الدخول y : 10 18 15 35 40 62 100 50

الحل :

لحساب معامل بيرسون للارتباط الخطي يلزم حساب المجاميع:

$\sum x , \sum y , \sum xy , \sum x^2 , \sum y^2$ **لذلك يتم تنظيم حساب هذه المجاميع كما في الجدول التالي:**

الأعمار x	الدخول y	xy	x^2	y^2
25	10	250	625	100
32	18	576	1024	324
29	15	435	841	225
43	35	1505	1849	1225
38	40	1520	1444	1600
51	62	3162	2601	3844
47	100	4700	2209	10000
35	50	1750	1225	2500
300	330	13898	11818	19818

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\ \sqrt{n\sum y^2 - (\sum y)^2}}$$

$$= \frac{8(13898) - (300)(330)}{\sqrt{8(11818) - (300)^2}\ \sqrt{8(19818) - (330)^2}}$$

$$= \frac{111184 - 99000}{\sqrt{94544 - 90000}\ \sqrt{158544 - 108900}}$$

$$= \frac{12184}{\sqrt{4544}\ \sqrt{49644}}$$

$$= \frac{12184}{15019.6}$$

$$r = 0.81$$

من جدول الارتباط السابق حدد نوع الارتباط

ب - معامل سبيرمان لارتباط الرتب :

معامل سبيرمان لارتباط الرتب :

Spearman rank Correlation Coefficient

لحساب معامل سبيرمان لارتباط الرتب يقوم بترتيب كل من المتغيرين ترتيباً تصاعدياً أو تنازلياً (أما تصاعدياً لكلا المتغيرين أو تنازلياً لكليهما). وفي حالة الترتيب التصاعدي تأخذ أقل قيمة من قيم المتغير الرتبة رقم ١ ، والقيمة الأعلى منها مباشرة الرتبة رقم ٢ وهكذا (بالنسبة لكل من المتغيرين). أما في حالة الترتيب التنازلي تأخذ أكبر قيمة من قيم المتغير الرتبة رقم ١ ، والقيمة الأقل منها مباشرة الرتبة رقم ٢ وهكذا (بالنسبة لكل من المتغيرين). وعند تساوي قيمتين (أو أكثر) من قيم المتغير نعطي كل قيمة رتبة مختلفة (كما لو كانت القيم غير متساوية) ثم نحسب متوسط هذه الرتب، ويعطى هذا المتوسط لكل من هذه القيم المتساوية.

وبعد ترتيب المتغيرين نحسب الفروق بين رتب كل من المتغيرين (ونرمز للفروق بالرمز d) ثم نقوم بتربيع هذه الفروق ونحصل على مجموعها أي نحصل على $\sum d^2$ ثم نعوض في معامل سبيرمان لارتباط الرتب والذي يأخذ الشكل التالي :

$$r_s = 1 - \frac{6(\sum d^2)}{n(n^2-1)}$$

حيث : $\sum d^2$ هو مجموع مربعات الفروق بين رتب المتغيرين، n هي عدد أزواج القيم.

مثال(١) :-اوجد معامل ارتباط سبيرمان لقيم المتغيرين

س :٦,٨,١٠,٧,٩

ص:٩,١٠,٧,٦,٨

الحل باستخدام الترتيب التنازلي

عدد مفردات البيانات=	س	ص	ترتيب س	ترتيب ص	d س- ص	d2
	٦	٩	٥	٢	٣	٩
	٨	١٠	٣	١	٢	٤
	١٠	٧	١	٤	٣-	٩
	٧	٦	٤	٥	١-	١
	٩	٨	٢	٣	١-	١
٥						
						٢٤

$$r_s = 1 - \frac{6(\sum d^2)}{n(n^2-1)}$$

$$r_s = 1 - \frac{6(24)}{5(5^2-1)} \qquad r_s = 1 - \frac{6\times(24)}{5\times 24}$$

$$r_s = 1 - \frac{144}{120} \qquad r_s = 1 - 1.20$$

r= -0.20

من جدول الارتباط السابق حدد نوع الارتباط

2- تحليل الانحدار (كوسيلة للتنبؤ) :

- يتناول تحليل الانحدار دراسة العلاقة بين متغيرين أحدهما المتغير التابع والاخر المتغير المستقل.
- المتغير التابع هو المتغير المراد تقديره والتنبؤ به خلال دراسة علاقته بالمتغير المستقل . فعلى سبيل المثال يمكن لمحلل البيانات تقدير او توقع قيمة المصروفات البيعية (المتغير التابع) في ضوء قيمة المبيعات (المتغير المستقل). حيث تزيد المصروفات البيعية كلما زادت المبيعات , ويعرف تحليل الانحدار في هذه الحالة بالانحدار البسيط . حيث يتم دراسة العلاقة بين متغيرين فقط أحدهما المتغير التابع والاخر هو المتغير المستقل.

نموذج الانحدار الخطى البسيط :

تأخذ معادلة الانحدار الخطى البسيط الشكل التالي :

ص = أ + ب س

حيث ان

ص = المتغير التابع المراد التنبؤ بقيمته.

أ = مقدار ثابت ويمثل قيمة "ص" عندما تكون قيمة "س" تساوى صفر.

ب = ميل خط الانحدار والذي يشير الى التغير في قيمة "ص" لكل وحدة تغير في "س".

س = المتغير المستقل

مثـــــــــــــــــــال (١)

بفرض ان محلل البيانات يرغب في تحليل المصروفات البيعية المسجلة بالدفاتر ، ومن واقع بيانات المنشاة محل التحليل استطاع المحلل تحديد ما يلي :

١- قيمة المصروفات البيعية الثابتة تبلغ ١٠٠٠٠ ج (ويمثل هذا المبلغ في معادلة الانحدار قيمة "أ")

٢- ان المصروفات البيعية "ص" تزيد كلما زادت المبيعات "س" وبفرض ان المحلل استطاع من بيانات المنشاة في السنوات السابقة تحديد ان كل جنيه زيادة في المبيعات يؤدى الى زيادة في المصروفات البيعية بمقدار ٠,٠٥ ج (ويمثل هذا المبلغ في معادلة الانحراف قيمة "ب ")

وفى ضوء المعلومات السابقة يمكن التعبير عن معادلة الانحدار كما يلى :

ص = أ + ب س

المصروفات البيعية (ص) = ١٠٠٠٠ج (أ) + [٠,٠٥ (ب) × قيمة المبيعات (س)]

وبفرض ان قيمة المبيعات كانت ١٠ مليون جنيه ، فان المحلل يستطيع توقع ان قيمة المصروفات البيعية تبلغ ٥١٠٠٠٠ج ، وذلك كما يلي :

ص = ١٠٠٠٠ + (٠,٠٥ × ١٠٠٠٠٠٠٠)

= ١٠٠٠٠ + ٥٠٠٠٠٠ = ٥١٠٠٠٠ جنيه.

مثـــــــــــال (٢)

بفرض ان المحلل كان بصدد تحليل مصروف عمولة البيع بإحدى المنشآت التي يقوم بتحليل حساباتها ولقد توافرت له البيانات التالية عن مصروف عمولة البيع وقيمة المبيعات خلال خمس سنوات قبل السنة محل الفحص.

السنة	مصروف عمولة البيع بالألف جنيه (ص)	قيمة المبيعات بالألف جنيه (س)
٢٠١٥	١٠	٢٠٠
٢٠١٦	١٢	٢٥٠
٢٠١٧	١٥	٣٢٠
٢٠١٨	١٨	٤٠٠
٢٠١٩	٢٠	٤٣٠

المطلـــــــــــوب /

١- تقدير معادلة انحدار ص / س باستخدام طريقة المربعات الصغرى

٢- تقدير القيمة المتوقعة لمصروفات عمولة البيع في السنة محل الفحص (سنة ٢٠٢٠) اذا علمت ان قيمة المبيعات الفعلية قد بلغت في سنة الفحص ٤٨٠ الف جنيه.

الحـــــــــــــــــــل

١- تقدير معادلة انحدار ص / س باستخدام طريقة المربعات الصغرى :

ص = أ + ب س

ولتقدير قيمة كل من أ ، ب يتم حل المعادلتين التاليتين :

مجـ ص = ن أ + ب مجـ س

مجـ س ص = أ مجــ س + ب مجــ س٢

ويتم حل المعادلتين السابقتين لإيجاد قيمة كل من أ ، ب كالتالي :

السنة	مصروف عمولة البيع بالألف جنيه (ص)	قيمة المبيعات بالألف جنيه (س)	س ص	س٢
٢٠١٥	١٠	٢٠٠	٢٠٠٠	٤٠٠٠٠
٢٠١٦	١٢	٢٥٠	٣٠٠٠	٦٢٥٠٠
٢٠١٧	١٥	٣٢٠	٤٨٠٠	١٠٢٤٠٠
٢٠١٨	١٨	٤٠٠	٧٢٠٠	١٦٠٠٠٠
٢٠١٩	٢٠	٤٣٠	٨٦٠٠	١٨٤٩٠٠
إجمالي	٧٥	١٦٠٠	٢٥٦٠٠	٥٤٩٨٠٠

وبالتعويض في المعادلتين السابقتين :

٧٥ = ٥ أ + ١٦٠٠ ب (١)

٢٥٦٠٠ = ١٦٠٠ أ + ٥٤٩٨٠٠ ب (٢)

وبقسمة ١٦٠٠ على ٥ يعطى الناتج ٣٢٠ ، وبضرب المعادلة الاولى في ٣٢٠ وطرحها من الثانية ينتج :

٢٥٦٠٠ = ١٦٠٠ أ + ٥٤٩٨٠٠ ب

٢٤٠٠٠ = ١٦٠٠ أ + ٥١٢٠٠٠ ب

١٦٠٠ = ٣٧٨٠٠ ب

$$ب = \frac{١٦٠٠}{٣٧٨٠٠} = ٠,٠٤٢٣$$

وبالتعويض بقيمة (ب) في المعادلة الاولى نحصل على قيمة "أ"

٧٥ = ٥أ + ١٦٠٠ × ٠,٠٤٢٣

٧٥ = ٥أ + ٦٨

٧٥ - ٦٨ = ٥ أ

$$أ = \frac{٧}{٥} = ١,٤$$

ونظرا لان القيمة بالألف جنيه فان أ = ١٤٠٠ جنيه.

ويكون بذلك تقدير معادلة الانحدار (ص = أ + ب س) كما يلي :

ص = ١,٤ + ٠,٠٤٢٣ س

<u>٢- تقدير القيمة المتوقعة لمصروف عمولة البيع في السنة محل الفحص ٢٠٢٠ بفرض ان قيمة المبيعات الفعلية قد بلغت في نفس السنة ٤٨٠ ألف جنيه</u>

= ١٤٠٠ + (٠,٠٤٢٣ × ٤٨٠٠٠٠)

= ١٤٠٠ + ٢٠٣٠٤ = ٢١٧٠٤ جنيه

للاقتراحات والشكاوى قم بمسح الصورة (QR)

قام بإعداد الإصدار الأول من هذا البرنامج:

المهندس / أيمن أبو العلا خليفه	**شركة مياه الشرب والصرف الصحي بشمال وجنوب سيناء**
الأستاذ / محمد إسماعيل	**شركة مياه الشرب والصرف الصحي بقنا**
الأستاذة / مريام طلعت سعد	**شركة مياه الشرب والصرف الصحي بالأقصر**

المنسق

المهندسة / حورية سعيد حسين	**شركة الصرف الصحي بالقاهرة**