تصميم التجارب الوراثية

مقرر دراسات عليا (دكتوراه) / قسم المحاصيل الحقلية

عدد الوحدات (3)

أستاذ المادة:أ.م.د.داود سلمان مدب

يهدف تصميم التجارب الوراثية الى دراسة الطرق الإحصائية الوراثية لتقدير مكونات التبايات المظهرية المختلفة وما يترتب عليها من تقدير للمعالم الوراثية

المصدر الرئيسي

Singh,R.K. and B.D.Chaudhary92007.Biometrical methods in quantitative genetic analysis. Kalyani Publishers,New Delhi.Ludhiana.pp:102-127.

بالإضافة الى المصادر الثانوية المتضمنة البحوث والدوريات المتخصصة با

المحاضرة الأولى

مقاييس النزعة المركزية

Measures of Central Tendency

المصدر:الشيحة، عبدالله بن عبد الكريم وعدنان بن ماجد بري.(2014). مقدمة في الإحصاء والاحتمالات وتطبيقاتها باستخدام اكسل.المملكة العربية السعودية.

وتسمى أحيانا مقاييس الموقع (Measures of Position) حيث أنها تصف أين يقع التمركز في قيم التوزيع , ونلاحظ أنه في أغلب التوزيعات ذات القيمة الواحدة وجود عدد كبير من القيم الملاحظة متراكمة حول قيمة معينة في التوزيع ويقل هذا التراكم تدريجيا كلما بعدت هذه القيم عن هذه القيمة من الجانبين . هذا التراكم حول القيمة المعينة نطلق عليه النزعة المركزية للتوزيع أو القيمة المتوسطة وتسمى تلك القيمة آلتي يحدث حولها التجمع بمقياس النزعة المركزية هذا المقَياس يختلف بحسب طبيعة البيانات والهدف من تحليلها بحيث يتم اختيار أيا من هذه المقاييس المتاحة وفقا لما يتلاءم مع هذه البيانات . ومن أهم هذه المقاييس

1- الوسط الحسابي . 2- الوسيط .

3- الربيعات والعشيرات والمئينات .

4- المنوال . 5- الوسط الهندسي .

6- الوسط التوافقي .

ومن أهمها الوسط الحسابي وذلك لما له من صفات خاصة وبالذات أهميته في أجراء العديد من العمليات الرياضية في التحليل الإحصائي .

<u>1-3 الوسط الحسابي أو المتوسط Mean</u>

يعرف الوسط الحسابي بأنه نتيجة قسمة مجموع القيم على عددها . فإذا افترضنا أن سعر كيلو البصل في أحد الأسواق كان كالآتي : 500 درهم ، 600 درهم ، 750 درهم ، 950 درهم فإن متوسط سعر البصل في هذا السوق هو :

$$\frac{500+600+750+950}{4}=\frac{2800}{4}=700$$

<u>طرق حساب الوسط الحسابي :</u>

<u>أ – الحساب من واقع بيانات غير مبوبة :</u> ونقصد بذلك أن جميع المفردات آلتي نريد حساب متوسطها يمكن حسابها مباشرة دون أعداد جدول توزيع تكراري لها وبهذه الطريقة فإن الوسط الحسابي لمجموعة من القيم : $x_1, x_2, x_3, \ldots\ldots\ldots\ldots, x_n$ هو $\overline{X}$

$$\overline{X}=\frac{x_{1}+x_{2}+x_{3}+\ldots\ldots+x_n}{N}=\frac{\sum X_i}{N}$$

<u>ب – حساب الوسط الحسابي من بيانات مبوبة :</u>

<u>1–الطريقة المطولة :</u>

عندما يكون لدينا جدول توزيع تكراري (أي أن البيانات المتوفرة لدينا قد تم تقسيمها ألي فئات ثم حسبنا التكرارات أمام كل فئة بالطريقة آلتي أسلفنا شرحها عند الحديث عن الجدول التكراري) فإن المتوسط الحسابي يمكن حسابه بأكثر من طريقة وسوف نقتصر الشرح هنا على الطريقة المباشرة بينما نستعرض بقية الطرق المختصرة بعد استعراض خواص الوسط الحسابي لما لهما من علاقة .

لحساب الوسط من الجدول فإننا نحسب مراكز الفئات لتمثل الفئة بأكملها (حيث مركز الفئة هو نصف المدى بين الحد الأعلى والأدنى للفئة) وبذلك تكون لدينا قيم لـ x من 1 ألي n ويقابل كل قيمة تكرارها f أي :

$$x_1, x_2, \ldots\ldots\ldots\ldots\ldots\ldots, x_n$$
$$f_1, f_2, \ldots\ldots\ldots\ldots\ldots\ldots, f_n$$

والمتوسط الحسابي في هذه الحالة هو مجموع حاصل ضرب مراكز الفئات في تكراراتها مقسوماً على المجموع الكلى للتكرارات ونعبر عن ذلك رياضياً بالمعادلة التالية

$$\bar{X} = \frac{x_1 f_1 + x_2 f_2 + \dots\dots\dots\dots + x_n f_n}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum f_i}$$

حيث $\sum f_i = n$

ويطلق على حساب المتوسط بهذا الأسلوب أحيانا الطريقة المطولة . ومن الملاحظ أن الحساب بهذه الطريقة مزعج علاوة على انه أكثر عرضة لارتكاب العديد من الأخطاء الحسابية خصوصاً عندما يكون التعامل مع أرقام كبيرة . غير أنه بتوفر الآتي الجيب الحاسبة وآلتي تؤدى معظمها العمليات الإحصائية المختلفة فإن الطريقة المطولة أيضا أمرا اكثر قبولاً . ولعله من المفيد هنا توضيح كيفية استخدام الآلة الحاسبة في إيجاد الوسط الحسابي مستخدمين الأنواع الشائعة منها .

كي نجعل الآلة الحاسبة تعمل بخيار الإحصاء نقوم بما يلي :

1. الضغط على مفتاح Inv ثم مفتاح SD أو STAT وعندها يظهر على الشاشة في أعلى اليمين هذا النمط .
2. قم بإدخال البيانات كالتالي : أدخل قيمة x_1 ثم مفتاح الضرب × وإدخال قيمة f_1 ثم الضغط على مفتاح الذاكرة +M .
3. كرر هذه العملية مع ($f_2 \times x_2$) ألي غاية ($f_n \times x_n$) .
4. الآن بإمكانك الحصول على أية نتيجة ترغب الحصول عليها مع مراعاة انه بالنسبة ألي المفاتيح آلتي لها وظيفتين فإن وظيفة المفتاح العادية تكون بالضغط عليه مفرداً ولاستخدامه إحصائيا نضغط مفتاح Inv قبله .

وجميع الآلات آلتي بها خيار الإحصاء ستجد أن مفاتيحها تشمل التالي :

$$n = \sum f$$

$$\sum x = \sum x_i f_i$$

$\bar{X}$ = الوسط الحسابي

<u>خواص الوسط الحسابي :</u>

ويتميز الوسط الحسابي بعدة خصائص نجملها في آلاتي:.

1 . مجموع القيم في مجموعة معينة يساوى المتوسط مضروبا في عدد تلك القيم . أي أن $\sum x_i = n\bar{x}$

2 . مجموع انحرافات القيم عن وسطها الحسابي يساوى صفرا أي $\sum(\mathrm{x_i} - \bar{\mathrm{x}}) = 0$

بفك ما بين القوسين ثم باستخدام الخاصية رقم 1 نستطيع إثبات ذلك :

$$\sum(x_i - \bar{x}) = \sum x_i - \sum \bar{x} = \sum x_i - n\bar{x}$$

$$\because \sum x_i = n\bar{x} \qquad \therefore n\bar{x} - n\bar{x} = 0$$

مثال 1 :

فعلى سبيل المثال الوسط الحسابي للقيم الثلاث 2 , 6 , 13 هو : $\bar{X}$ = 7 $= \frac{21}{3}$

إذن (2 − 7) + (6 − 7) + (13 − 7)

= − 5 −1 + 6 = 0

3 . مجموع الانحرافات المربعة لمجموعة من القيم عن رقم ثابت A هو أصغر ما يمكن في حالة واحدة فقط , وهى عندما $\bar{X}$ = A

أي أن $\sum(X-A)^2 = Min$ وهو آدني عند $\bar{X}$ = A

لأثبات هذه الخاصية نفرض أن d = X- A حيث A هي أية قيمة غير المتوسط الحسابي $\bar{X}$ ، لأثبات أن $\sum d^2$ وهى $\sum(X-\bar{X})^2$ حد أدنى ، يتحقق ذلك فقط عندما $\bar{X}$ = A أي أن $\sum x^2$ هي حد أدنى .

$$\sum d^2 = \sum(X-A)^2$$

$$= \sum X^2 - 2A\sum X + nA^2$$

لكن

$$\sum X = n\bar{X}$$

$$\therefore \sum d^2 = \sum X^2 - 2nA\bar{X} + nA^2$$

بإضافة وطرح المقدار $n\bar{X}^2$ تصبح المعادلة السابقة كما يلي :

$$\sum d^2 = \sum X^2 - n\bar{X}^2 + (n\bar{X}^2 - 2nA\bar{X} + nA^2)$$

$$= \sum X^2 - n\bar{X}^2 + n(\bar{X}^2 - 2A\bar{X} + A^2)$$

$$= \sum X^2 - n\bar{X}^2 + n(\bar{X} - A)^2$$

الحد الثالث في هذه المعادلة دائماً موجب مهما كانت قيمة A (أي أكبر أم أصغر من $\bar{X}$) وعليه فإن قيمة المعادلة اصغر ما يمكن عندما يكون المقدار الثالث صفر أي بتساوى A و $\bar{X}$ وبالتالي يكون $\sum d^2 = \sum x^2$.

4− إذا ضربنا (قسمنا) كل قيمة من قيم التوزيع x_i بعدد أو (على عدد) ثابت A فإن $\bar{X}$ الوسط الحسابي سيتغير بمقدار الضرب أو القسمة على ذلك العدد.

في حالة الضرب

$$\overline{X}=\frac{\sum x_i(af_i)}{\sum f_i}=\frac{a\sum x_i f_i}{\sum f_i}=a\overline{X}$$

في حالة القسمة

$$\overline{X}=\frac{\sum x_i f_i}{\sum (af_i)}=\frac{\sum x_i f_i}{a\sum f_i}=\frac{\overline{X}}{a}$$

5- لا تتغير قيمة الوسط الحسابي إذا ضربنا وقسمنا جميع التكرارات f_i بقيمة ثابتة a أي :

$$\overline{X}=\frac{\sum x_i(af_i)}{\sum (af_i)}=\frac{a\sum x_i f_i}{a\sum f_i}=\frac{\sum x_i f_i}{\sum f_i}=\overline{X}$$

وتفيد هذه الخاصية في حساب المتوسط الحسابي عندما تتوفر لدينا التكرارات النسبية ، ذلك أن التكرارات النسبية تنتج عن قسمة f_i (كل تكرار) على العدد الثابت $\sum f_i$ = a مجموع التكرارات وهذا مالا يؤثر على قيمة الوسط الحسابي $\overline{X}$.

$$\overline{X}=\frac{\sum x_i f_i}{\sum f_i}=\frac{\sum x_i \frac{f_i}{\sum f_i}}{\sum \frac{f_i}{\sum f_i}}$$

وكما نعلم فإن مجموع النسب دائماً تساوى الواحد صحيح أي $\sum \frac{f_i}{\sum f_i}=1$

عليه فإن : $\overline{X}=\sum x_i \frac{f_i}{\sum f_i}$ نظراً لأن المقام = 1

6- إذا أضفنا ألي (أو طرحنا من) كل قيمة من قيم التوزيع التكراري قيمة ثابتة فإن $\overline{X}$ يزيد (أو ينقص) بمقدار تلك القيمة الثابتة . ويستفاد من هذه الخاصية في حساب الوسط الحسابي بطريقة الانحرافات عن وسط فرضى .

$$\overline{X}=\frac{\sum (x_i \pm A)f_i}{\sum f_i}=\overline{X}\pm A$$

إذا كانت A أي وسط حسابي افتراضي لمجموعة من القيم , حيث $d_i = X_i - A$ هو انحرافات X_i عن A فإن $X_i = d_i + A$ وبالتالي :

$$\overline{X}=\frac{\sum (d_i + A)f_I}{\sum f_i}=\frac{\sum d_i f_i + A\sum f_i}{\sum f_i}$$

$$\overline{X}=\frac{\sum d_i f_i}{\sum f_i}+A\frac{\sum f_i}{\sum f_i}=A+\frac{\sum d_i f_i}{\sum f_i}=A+\overline{d}$$

مثال2 :

من المثال السابق نجد أن $\sum(X_i - 7)^2$ هي كما يلي :

(2-7) + (6-7) + (13-7) = 62

وهذه الانحرافات المربعة ستكون أكبر من 62 عند أي قيمة أخرى غير 7 .

فمثلا عند A = 0

الانحرافات المربعة= (2− 0) + (6−0) + (13−0) = 209

عند A = 5 = (2−5) + ... + (13−5) = 69

عند A = 9 = (2− 9) + ... + (13−9) = 74

7− عندما نُقسم بيانات مجموعة معينة ألي اثنان أو أكثر من المجموعات , فأن المتوسط العام لمتوسطات جميع المجموعات هو الوسط المرجح لها . أي أن :

$$\bar{X}_m = \frac{\bar{x}_1 f_1 + \bar{x}_2 f_2 + \ldots\ldots + \bar{x}_n f_n}{\sum f_i} = \frac{\sum_{i=1}^{n} \bar{x}_i f_i}{\sum f_i}$$

حيث m = عدد المجموعات الفرعية .

$\bar{X}_i$ = الوسط الحسابي للمجموعة الفرعية i .

f_i = عدد القيم في المجموعة الفرعية i .

n = العدد الكلي للقيم .

8 − لا يمكن حسابه في حالة التوزيعات التكرارية المفتوحة وذلك لعدم تمكننا من تحديد مركز الفئة المفتوحة .

9− لا يمكن حسابه في حالة البيانات النوعية .

10− يتأثر بالقيم المتطرفة .

مثال 3 : الجداول التالي يعطى توزيع تكراري لظاهرة معينة ، أحسب الوسط الحسابي بالطريقة المطولة :

الفئات	5−9	10−	15−	20−	25−	30−	المجموع
التكرار	1	3	2	3	1	2	12

الحل :

Xf	مراكز الفئات x	التكرار f	الفئات
7	7	1	5 − 9
36	12	3	10 − 14
34	17	2	15 − 19
66	22	3	20 − 24
27	27	1	25 − 29
64	32	2	30 − 34
234		12	المجموع

$$\bar{X} = \frac{\sum x_i f_i}{\sum f_i}$$

$$\bar{X} = \frac{234}{12} = 19.5$$

<u>الوسط المرجح</u> : عندما يكون لدينا مجموعة من البيانات x_1 ، x_2 ، .... x_n مرجحة بأوزان مختلفة نشير أليها بالرموز w_1 ، w_2 ،........w_n . فإن المتوسط الحسابي لهذه البيانات يعطى بالمعادلة التالية :

$$\bar{X}_w = \frac{x_1 w_1 + x_2 w_2 + \ldots\ldots\ldots + x_n w_n}{w_1 + w_2 + \ldots\ldots\ldots + w_n} = \frac{\sum x_i w_i}{\sum w_i}$$

حيث w هي الأوزان المختلفة المقابلة لكل قيمة .

<u>مثال 4</u> :

تدفع شركة أجر $\frac{5}{12}$ من قوة العمل بها بمعدل 8 دينارات في اليوم وأجر $\frac{1}{3}$ قوة العمل بمعدل 9 دينارات في اليوم , وأجر $\frac{1}{4}$ قوة العمل بمعدل 12 دينار في اليوم . ما هو المتوسط المرجح للأجور المدفوعة في هذه الشركة .

<u>الحل :</u>

الوسط المرجح : $\bar{X}_w = 8 \times \frac{5}{12} + 9 \times \frac{1}{3} + 12 \times \frac{1}{4}$

$= 9.33$ دينار

2-الطريقة المختصرة :

أ- طريقة الأنحرافات : تعتمد هذه الطريقة لحساب المتوسط الحسابي على استخدام وسط فرضى ويتم ذلك كما يلي :

نختار قيمة معينة A لتكون وسطاً فرضياً للبيانات ثم نقوم بحساب انحرافات بقية مراكز الفئات عن تلك القيمة d_i فإذا كان مجموع هذه الانحرافات $\sum d_i$ مساوياً للصفر فإن ذلك يعنى تساوى الوسط الحسابي مع الفرضي . ويكون الفرضي أصغر إذا كان المجموع موجباً ، ويكون أكبر إذا كان المجموع سالباً . ويتم حساب الوسط الحسابي عن طريق قسمة مجموع الانحرافات على عددها ثم إضافة ذلك الناتج بإشارته ألي الوسط الفرضي أي أن

$$\overline{X} = A + \frac{\sum d_i f_i}{\sum f_i}$$

ويتم في العادة اختيار الوسط الفرضي في منتصف الجدول أو تلك القيمة التي تقابل اكبر تكرار أو كلاهما .

ب- طريقة الانحرافات المختصرة : ولمزيد من الاختصار في إجراء العمليات الحسابية واستعمالاً لخصائص الوسط الحسابي السالف ذكرها فإننا في حالة تساوى أطوال الفئات نستطيع قسمة جميع الانحرافات d_i على طول الفئة c فنحصل على ما يسمى بالانحرافات المختصرة ونضيف عمود ألي جدول حساباتنا نقوم فيه بحساب $d_i' = \frac{d_i}{c}$ ثم حساب المتوسط الحسابي بالمعادلة التالية :

$$\overline{X} = A + \left(\frac{\sum d_i' f_i}{\sum f_i}\right) \times c$$

مثال 5 :

أستخدم نفس المثال السابق لإيجاد الوسط الحسابي بالطريقة المختصرة (الانحرافات) .

الحل : يمكن إتباع الخطوات الآتية :

1- يتم اختيار وسط فرضى من أحد مراكز الفئات وليكن مركز الفئة المقابل لأكثر تكرار (A =22)

2- نجد انحراف مركز فئة عن الوسط الفرضي $d = X - A$

3- نطبق القانون الأتي:.

$$\overline{X} = A + \frac{\sum d_i f_i}{\sum f_i}$$

وذلك كما هو مبين بالجدول الآتي :

الفئات	التكرار F	مركزالفئه X	d	dF
9−5	1	7	−15	−15
14−10	3	12	−10	−30
19−15	3	22	0	−10
24−20	3	22	0	0
29−25	1	27	5	5
34−30	2	32	10	20
المجموع	12			−30

$$\overline{X} = A + \frac{\sum d_i f_i}{\sum f_i}$$

$$\overline{X} = 22 + \frac{-30}{12} = 22 - 2.5 = 19.5$$

ملاحظة :

يمكن إيجاد الوسط الحسابي بالطريقة المختصرة وذلك بأحد عامل مشترك من الانحرافات يرمز له k وقسمة هذه الانحرافات على العامل المشترك k وكما ذكرنا في العادة يكون هذا العامل المشترك هو طول الفئة ، وبالتالي نقوم بقسمة الانحرافات على طول الفئة C (إذا كانت متساوية) ، فيصبح الوسط الحسابي كما يلي :

$$\overline{X} = A + \left(\frac{\sum d_i' f_i}{\sum f_i} \right) \times C$$

الفئات	التكرار	مراكز الفئات	d	$d_å' = \frac{d}{k}$	$d_i' f_i$
9-5	1	7	−15	−3	−3
14 . 10	3	12	−10	−2	−6
19. 15	2	17	−5	−1	−2
24. 20	3	22	0	0	0
29 . 25	1	27	5	1	1

34 . 30	2	32	10	2	4
المجموع	12				−6

$$\overline{X} = 22 + \left(\frac{-6}{12}\right) \times 5$$

$$\overline{X} = 22 + \frac{-30}{12} = 22 - 2.5 = 19.5$$

3−2 الوسيط Median :

الوسيط يعتبر من مقاييس النزعة المركزية الأكثر توسطا في التوزيع , حيث أنه يقسم القيم بعد ترتيبها حسب حجمها إلى نصفين متساويين . ويرمز للوسيط Md بالرمز $\tilde{x}$.

أ− حساب الوسيط للبيانات غير المبوبة :

إذا كان عدد القيم فردياً (n فردى) فإن الوسيط هو القيمة التي ترتيبها $\frac{n+1}{2}$ بعد ترتيب القيم تصاعدياً أو تنازلياً .

مثال 6 :

مجموعة الأرقام : 13− , 12 , 14 , 17 , 41 وسيطها هو 14 .

حيث أن ترتيبه بعد تنسيقها بشكل تصاعدي أو تنازلي هو :

(1+5) / 2 = 6 / 2 = 3 والقيمة ذات الترتيب 3 هي 14 .

أما إذا كان عدد القيم زوجياً (n زوجي) فإن الوسيط هو معدل القيمتين اللذين ترتيبهما كما يلي : $\frac{n}{2}$, $\frac{n}{2}+1,$ بعد ترتيب القيم تصاعدياً أو تنازلياً .

مثال 7 :

مجموعة الأرقام : 41 , 17 , 14 , 66 , 13− , 12 .

لإيجاد الوسيط لها نرتبها أولا (تصاعديا أو تنازليا) : 13− ، 12 ، 14 ، 17 ، 41 ، 66

الترتيب الأول : $\frac{6}{2} = 3$

الترتيب الثاني : $\frac{6}{2} + 1 = 4$

∴ الوسيط Md $\tilde{X} = \frac{14+17}{2} = 15.5$

حساب الوسيط للقيم المبوبة :

لحساب قيمة الوسيط من بيانات توزيع تكراري فأننا نقوم أولا بحساب نصف التكرارات مبتدئين من أحد النهايات (تصاعدي أو تنازلي) . بعد تحديد ترتيب الوسيط نحدد الفئة التي يقع بداخلها وذلك بالاستعانة بجدول التكرار المتجمع (الصاعد أو النازل).

وتكون بالتالي قيمة الوسيط مساوية لبداية الفئة التي يقع بداخلها , مضافا إليها قدرا من طول الفئة يتناسب مع بعد الوسيط عن البداية . ولتسهيل حساب الوسيط فإننا نستخدم المعادلة التالية :

$$\tilde{X} = L_1 + (\frac{\frac{\sum f}{2} - F_1}{f_{med}}) \times c$$

حيث L_1 = الحد الأدنى للفئة الوسيطية (أي الفئة التي يقع فيها الوسيط)

$\sum f_i$ = عدد العناصر في البيانات (مجموع التكرارات) .

F_1 = التكرار الصاعد السابق للفئة الوسيطية .

f_{med} = تكرار الفئة الوسيطية .

c = طول الفئة الوسيطـية .

مثال 8 :

الآتي جدول توزيع تكراري ، المطلوب / حساب الوسيط .

الفئات	5–9	10–	15–	20–	25–	30–	35–	المجموع
التكرار	3	2	6	10	13	7	4	45

الحل : باستخدام التكرار الصاعد

الفئات	التكرار f	تكرار متجمع صاعد f	
5– 9	3	3	
10 – 14	2	5	
15 – 19	6	11	
20 – 24	10	21	
25 – 29	13	34	الفئة الوسطية
30 – 34	7	41	
36 – 39	4	45	

لمجموع	45	

$$22.5 = \frac{45}{2} = \frac{\sum f_i}{2} = \quad \text{ترتيب الوسيط}$$

$$\tilde{X} = L_1 + \left(\frac{\frac{\sum f_i}{2} - F_1}{f_{mad}} \right) \times C$$

$$\tilde{X} = 25 + \left(\frac{22.5 - 21}{13} \right) \times 4$$

$$\tilde{X} = 25 + \left(\frac{1.5}{13} \right) \times 4$$

$$\tilde{X} = 25 + \frac{6}{13}$$

$$\tilde{X} = 25 + 0.46 = 25.46$$

باستخدام التكرار التجمع النازل :

الفئات	التكرار	تكرار متجمع نازل	
9 − 5	3	45	
14 − 10	2	42	
19 − 15	6	40	
24 − 20	10	34	
29 − 21	13	24	الفئة الوسطية
34 − 30	7	11	
39 . 35	4	4	
المجموع	45		

$$22.5 = \frac{45}{2} = \frac{\sum f_i}{2} = \quad \text{ترتيب الوسط}$$

$$\tilde{X} = L_2 - \left(\frac{\frac{\sum f}{2} - F_1}{f_{med}} \right) C$$

حيث L_2 الحد الأعلى للقيمة الوسطية

F_1 = التكرار المتجمع النازل اللاحق للفئة الوسيطية .

$$\tilde{X} 29 - \left(\frac{22.5 - 11}{13} \right) 4$$

$$\tilde{X} = 29 - \left(\frac{11.5}{13} \right) 4$$

$$\tilde{X} = 29 + \frac{46}{13}$$

$$\tilde{X} = 29 - 3.54 = 25.46$$

<u>الربيعــات , والعشــيرات والمئيئــات : Quartile, Deciles, and Percentiles</u>

رأينا فيما سبق أن الوسيط هو مقياس للقيمة التي تتوسط المجموعة إذا رتبت أعداد المجموعة تربيا تصاعديا أو تنازلياً . والواقع أن هناك مقاييس عدة أخرى للتوزيع التكراري والتي ليست مقاييس للنزعة المركزية ولكنها تساعد بشكل كبير في قياس التشتت والالتواء . هذه المقاييس لها صلة وثيقة بالوسيط حيث يشتركون في خاصية أن جميعها تعتمد على موقعها في التوزيع . وبتعميم فكرة الوسيط يمكننا القول إننا لو قسمنا المجموعة ألي أربعة أجزاء متساوية . هذه القيم يرمز لها بالرموز Q_1 , Q_2 , Q_3 ونسميها بالترتيب الربيع الأدنى , الوسيط , الربيع . وبالمثل فان القيم التي تقسم المجموعة ألي عشرة أجزاء متساوية تسمى بالعشيرات ويرمز لها بالرمز :

D_1 , D_2 , ... ,D_9 أما القيمة التي تقسم البيانات ألي مائة قسم متساوٍ فتسمىالمئيئات ويرمز لها بالرمز P_{10} , P_{20} , ... ,P_{90} والوسيط في هذه الحالات يساوى العشير الخامس D_5 وأيضا المئين الخمسن D_{50} . وتكون طريقة حساب هذه المقاييس مشابهة لحساب الوسيط وتستخدم نفس المعادلة السابقة , والاختلاف الوحيد هنا هو فقط تحديد الترتيب . وعليه فان الخطوات الواجب اتباعها هي الأتي:

1. تحديد الترتيب للمقياس المطلوب .

2.تحديد الفئة التي يقع بداخلها هذا المقياس .

3.تكون القيمة ممثلة لمبدأ الفئة التي يقع بداخلها مضافا إليها قدرا من طول الفئة يتناسب مع البعد عن مبدئها .

فباستخدام هذه الخطوات تجد مايلى :.

الربيع الأول (الأدنى) : لتحديد ترتيبه نقسم $\frac{\sum f}{4}$

$$Q1 = L_1 + \left(\frac{\frac{\sum f}{4} - F_1}{f_{Q1}} \right) \times c$$

الربيع الثالث (الأعلى) Q_3: لتحديد ترتيبه نقسم $\frac{3\sum f}{4}$

$$Q3 = L_1 + \left(\frac{\frac{3\sum f}{4} - F_1}{f_{Q3}} \right) \times c$$

العشير السابع =

$$D7 = L_1 + \left(\frac{\frac{7\sum f}{10} - F_1}{f_{D7}} \right) \times c$$

العشير الخامس =

$$D5 = L_1 + \left(\frac{\frac{5\sum f}{10} - F_1}{f_{D5}} \right) \times c$$

المئين الخامس والعشرين =

$$P25 = L_1 + \left(\frac{\frac{25\sum f}{100} - F_1}{f_{P25}} \right) \times c$$

المئين الخامس والسبعين =

$$P75 = L_1 + \left(\frac{\frac{75\sum f}{100} - F_1}{f_{P75}} \right) \times c$$

إيجاد الوسيط بيانياً

بمعنى إيجاد الوسيط بالرسم البياني ، فإذا رسمنا المنحنى التكراري المتجمع الصاعد أو النازل ، ومن ثم نمد خط أفقي موازى للمحور الأفقي أمام ترتيب الوسيط على المحور العمودي ، حيث يمثل المحور الأفقي الفئات أما المحور العمودي يمثل التكرارات المتجمعة(الصاعدة أو النازلة) .

نقوم بإسقاط عمود في تلك النقطة على المحور الأفقي يقابله قيمة الوسيط .

كما يمكن إيجاد الوسيط برسم المنحنيين الصاعد والنازل في آن واحد ومن نقطة التقاطع نسقط عمود على المحور الأفقي عندها تكون قيمة الوسيط .

خواص الوسيط

1- يمكن استخدامه في الجداول ذات الفئات المفتوحة .

2- لا يتأثر بوجود قيم متطرفة .

3- يمكن حسابه للبيانات النوعية .

4 -مجموع الانحرافات المطلقة للقيم عن وسيطها أقل ما يمكن.

المنوال Mode :

المنوال هو الفئة التي تبدو أكثر شيوعا أو تركيزا . ولذا فإن وجود قيمة أو بعض القيم المتطرفة أو الشاذة في التوزيع . لن يؤثر على المنوال .

مثال 9 :

إذا كان لدينا الأجر اليومي لسبعة عمال هي :. 5 ، 6 ، 7 ، 7 ، 7 ، 8 ، 10

فإن قيمة المنوال = 7(لأنها تكررت اكثر من غيرها)

أما إذا كان لدينا مجموعة القيم : 3 , 5 , 6 , 7 , 9 , 10 , 11

فإن هذه القيم ليست لها منوال .

أما إذا وجدت قيمة مكررة نفس العدد فإنه يمكن أن يكون هناك منوالان لهذه القيم .

استخراج المنوال للقيم المبوبة :

المنوال يساوى مركز الفئة المقابل للفئة المنوالية (الفئة المنوالية هي التي تقابل أكبر تكرار) . ويلاحظ على هذه الطريقة في إيجاد قيمة المنوال بأنها طريقة تقريبية وذلك لأن المنوال يتوزع على الفئة كلها ، وبذلك يجب أن يتم تقدير المنوال بصورة أدق بأحد الطريقتين الآتيتين

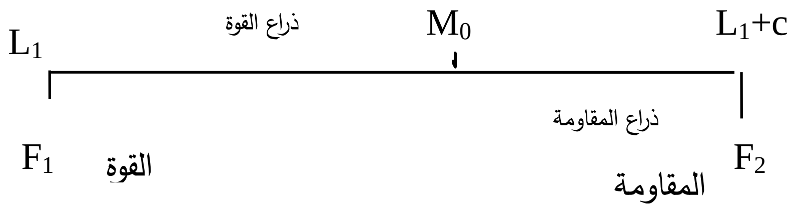
<u>الطريقة الأولى (طريق الرافعة)</u>

ويطلق عليها هذا الاسم وذلك لتشابه حساباتها مع قانون الروافع حيث يقول القانون الفيزيائي بأن توازن الرافعة عند نقطة الارتكاز يشترط أن تتحقق المتساوية :

(القوة × ذراعها = المقاومة × ذراعها)

حيث نأخذ الطريقة في الاعتبار الفئة المنوالية وتكراري الفئتين المتجاورتين بها على أساس أن تكراري هاتين الفئتين يتجاذبان المنوال ، فتميل فئة المنوال للاقتراب من الفئة ذات التكرار الأكبر .

والشكل التالي يوضح الأرتكازعند نقطة المنوال M_0 :



حيث L_1 = بداية الفئة ، L_1+c نهاية الفئة ، c طول الفئة ، و F_1 = التكرار السابق لفئة المنوال ، F_2 هو تكرار الفئة اللاحقة لفئة المنوال

ولكي يتحقق التوازن حول نقطة الارتكاز M_0 يجب تحقيق العلاقة التالية :

$$(M_0 - L_1) \times F_1 = (L_1 + c - M_0) \times F_2$$

$$M_0 F_1 - L_1 F_1 = L_1 F_2 + cF_2 - M_0 F_2$$

$$\therefore M_0 (F_1 + F_2) = L_1 (F_1 + F_2) + cF_2$$

$$M_0 = \frac{L_1(F_1 + F_2) + cF_2}{(F_1 + F_2)} = L_1 + \frac{F_2}{(F_1 + F_2)} \times c$$

<u>مثال 10 :</u>

لحساب المنوال من التوزيع التكراري لأوزان البيض في المثال السابق نجد أن التكرار السابق لفئة المنوال = 12 ، و التكرار اللاحق لفئة المنوال = 7 ، الحد الأدنى لفئة المنوال = 62 ، عليه فامنوال يمكن حسابه كما يلي :

$$63.47 = 4 \times \left(\frac{7}{12+7}\right) + 62 = \text{المنوال}$$

مثال 11 :

الجدول الأتي توزيع تكراري والمطلوب إيجاد فئة المنوال .

الفئات	5 −	10−	15−	20−	25−	30−	35−
التكرار	2	5	10	17	15	8	3

$$MO = L_1 + \left(\frac{F_2}{F_1 + F_2}\right)C$$

$$MO = 20 + \left(\frac{15}{10+15}\right)5$$

$$MO = 20 + \frac{75}{25}$$

$$MO = 20 + 3 = 23$$

أو نفرض x بعد المنوال عن نقطة الارتكاز

القوة = التكرار السابق للفئة المنوالية .

المقاومة = التكرار اللاحق للفئة المنوالية

القوة × ذراعها = المقاومة × ذراعها

$$10x = 15(5-x)$$

حيث $5 = C$ *(طول الفئة المنوالية)*

$$10x = 75 - 15x$$

$$25x = 75$$

$$x = \frac{75}{25} = 3$$

$$MO = L + x$$

$$MO = 20 + 3 = 23$$

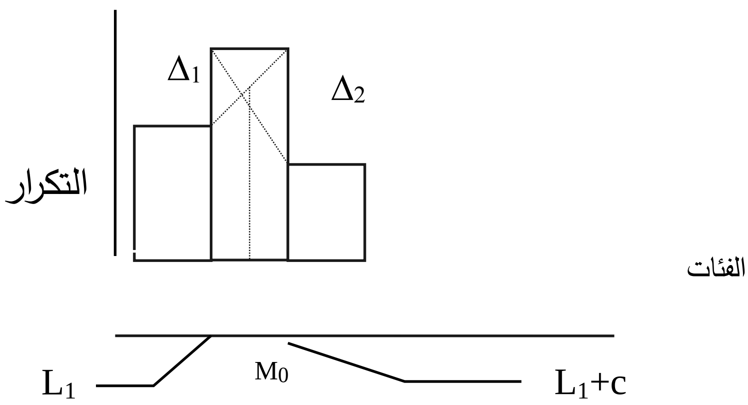
الطريقة الثانية (طريقة بيرسون)

ونطلق عليها بعض الأحيان الفروق .

تعتمد هذه الطريقة على الفرق بين الفئة المنوالية وتكرار كل من الفئتين المجاورتين لها . وهذه تسمى طريقة الفروق لبيرسون نسبة ألي كارل بيرسون الذي توصل لها. وهو يرى أن الذي يحدد موقع المنوال داخل الفئة المنوالية هو الفرق بين تكرار الفئة المنوالية وتكراري الفئتين المجاورتين أي السابقة واللاحقة لها . ويسير العمل الحسابي على نسق يتشابه ألي حد كبير من طريقة الرافعة مع اختلافاً بسيطاً في مكونات المتساوية حيث أن هذه الطريقة تُعطى كما يلي :

القوة × ذراع المقاومة = المقاومة × ذراع القوة

والشكل التالي يوضح كيفية استنتاج هذه العلاقة :



ويمكن اشتقاق العلاقة جبرياً بالخطوات التالية :

نحن نرغب في تعيين المنوال بحيث :

$$\frac{M_0 - L_1}{L_1 + c - M_0} = \frac{\Delta_1}{\Delta_2}$$

$$\therefore \Delta_2 M_0 - \Delta_2 L_1 = \Delta_1 L_1 + \Delta_1 c - \Delta_1 M_0$$

$$M_0(\Delta_1 + \Delta_2) = L_1(\Delta_1 + \Delta_2) + \Delta_1 c$$

$$\therefore M_0 \quad = \frac{L_1(\Delta_1 + \Delta_2) + \Delta_1 c}{(\Delta_1 + \Delta_2)}$$

$$M_0 \quad = L_1 + \frac{\Delta_1}{(\Delta_1 + \Delta_2)} \times c$$

حيث Δ_1 = الفرق المطلق بين تكرار فئة المنوال وتكرار الفئة السابقة .

Δ_2 = الفرق المطلق بين تكرار فئة المنوال وتكرار الفئة اللاحقة .

L_1 = الحد الأدنى لفئة المنوال ، c = طول الفئة المنوالية .

$\therefore$ المنوال في المثال السابق =

$$63.6 = 4 \times [(15 + 10) \div 10] + 62$$

<u>مثال 12 :</u>

أستخدم المثال السابق لإيجاد فئة المنوال

<u>الحل:</u>

$$\Delta_1 = 17 - 10 = 7$$

$$\Delta_2 = 17 - 15 = 2$$

$$Mo = L_1 + \left(\frac{\Delta_1}{\Delta_{1+}\Delta_2} \right) C$$

$$Mo = 20 + \left(\frac{7}{2+7} \right) 5$$

$Mo = 20 + \frac{35}{9}$ ، عليه $Mo = 20 + 3.88 = 23.88$

أو $\frac{\Delta_1}{\Delta_2} = \frac{X}{C - X}$ ، $\frac{7}{2} = \frac{X}{5 - X}$

$$2X = 35 - 7X$$

$9X = 35$ ، $X = \frac{35}{9} = 3.88$

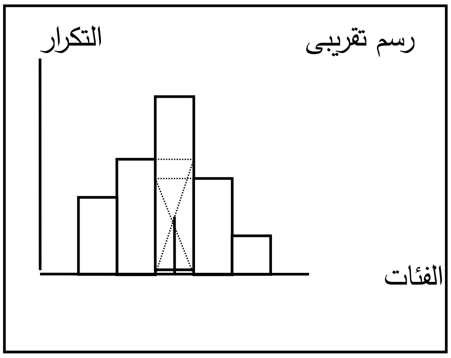
$$Mo = L_1 + X$$

$$Mo = 20 + 3.88 = 23.88$$

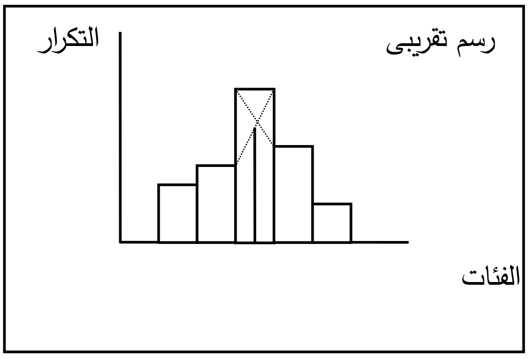
<u>الطريقة البيانية:</u>

يمكن الحصول على المنوال بيانيا بأن نرسم المدرج التكراري للثلاث مستطيلات التي تمثل تكرارات الفئة المنوالية والفئات السابقة واللاحقة لها . ثم نستخدم أحد الطريقتين الآتيتين في أجراء التوصيلات بين النقاط المختلفه في الرسم :

1- نمد قيمتي المستطلين المجاورين داخل المستطيل الذي يرمز لأكبر تكرار كما في الرسم . ونقوم بتوصيل النقط الأربعة بخطين يتقاطعان في نقطة . وعليه يبين أحداثي النقطة على المحور الأفقي قيمة المنوال .



2–من المدرج التكراري لكل من الركنين العلويين للمستطيل الممثل للفئة المنوالية مع ركني المستطيلين للفئتين المتجاورتين بخطين يتقابلان في نقطة نصلها بالإحداثي الأفقي ومن ثم نحصل على قيمة المنوال .



كذلك يمكن حساب المنوال من رسم المنحنى التكراري للتوزيع وذلك بأن نسقط عموداً من أعلى قمة بهذا المنحنى على المحور الأفقي فتكون نقطة الالتقاء هي قيمة المنوال .

خواص المنوال

1. أسهل مقاييس النزعة المركزية.
2. لايتأر بوجود القيم المتطرفة.
3. يمكن حسابه في حالة التوزيعات التكرارية المفتوحة بحيث لاتكون الفئة المنوالية أحدها .
4. يكون لبعض البيانات منوالان أو أكثر .

3–5 الوسط الهندسي Geometric Mean

يعرف الوسط الهندسي على أنه الجذر النوني لحاصل ضرب مجموعة الملاحظات أو البيانات . فعلى سبيل المثال إذا كانت $x_1 , x_2 , x_3 , ..., x_n$ هي مجموعة أرقام , فإن متوسطها الهندسي هو:

$$G = \sqrt[n]{X_1.X_2..........X_n}$$

وبأخذ لوغاريتم الطرفين نجد: $Log\ G = \frac{\sum LogX}{n}$

أما بالنسبة للبيانات المبوبة فالوسط الهندسي يُعطى بالمعادلة الآتية :

$$G = \sqrt[n]{X_1^{f_1}.X_2^{f_2}............X_n^{f_n}}$$

بأخذ لوغاريتمات الطرفين فإن :

$$Log\ G = \frac{\sum f \times \log X}{\sum f}$$

<u>مثال13 :</u>

أحسب الوسط الهندسي للبيانات الآتية

1. 90   1. 75   1. 25   2. 10   1. 50

<u>الحل:.</u>

$$Log\ G = \frac{\sum Log}{n}$$

$$Log\ G = \frac{1}{5}\left[\sum Log1.50 + Log2.10 + Log1.25 + Log1.75 = Log1.90\right]$$

$$Log\ G = \frac{1}{5}\left[0.1761 + 0.3222 + 0.5965 + 0.2430 + 0.2788\right]$$

$$Log\ G = \frac{1.1170}{5} = 0.2234$$

ومن جداول الأعداد المقابلة للوغاريتمات نجد أن $G = 1.67$

<u>استخدامات الوسط الهندسي :</u>

- عندما نرغب في إعطاء وزن متساوى لنسب التغير. ولإيضاح ذلك نفرض أن سعر سلعتين كان عند سنة معينة الأولى 2 دينار للوحدة ، والثانية 10 دينار للوحدة وبعد فترة تضاعف سعر الأولى (إلى 4 دينار) وأنخفض سعر الثانية إلى النصف (5 دينار) . ثم في فترة ثالثة أنخفض سعر الأولى عن أول سنة إلى النصف (دينار واحد)

تضاعف سعر الثانية أي أصبح (20 دينار) . عليه تكون المتوسطات الحسابية لهذه السلع خلال الثلاث فترات يساوى 6 ، 4.5 ، 10.5 بينما نجد أن متوسطها الهندسي للفترات الثلاث = 4.47 .

- ربما يكون من أهم استخدامات مبادئ الوسط الهندسي هو تطبيقاته على تحديد معدلات النمو حيث نستطيع حساب المتوسط السنوي للتغير في ظاهرة معينة بالمعادلة التالية .

$$P_n = P_0(1+r)^n$$

حيث p_0 = المجتمع بداية الفترة

P_n = المجتمع في نهاية الفترة

r = معدل التغير (معدل النمو)

n = عدد الفترات الزمنية

<u>مثال 14:</u>

الوسط الهندسي للأرقام 5 , 8 , 10 , 12 هو :

$$G = \sqrt[4]{5.8.10.12} = \sqrt[4]{4800} = 8.4$$

بينما نجد أن المتوسط الحسابي لهذه الأرقام هو8.75 .

لأي مجموعة من القيم الموجبة فان الوسط الهندسي أصغر من المتوسط الحسابي .فإذا كانت أي قيمة من هذه القيم = صفر فإن المتوسط الهندسي لن يكون ملائما . أما إذا كانت بعض القيم سالبة فإن المتوسط الهندسي يمكن حسابه ولكنه سوف لن يعنى شيئا . ومن الناحية العملية فإننا نستعين باللوغريتمات عندما نقوم بحساب الوسط الهندسي للبيانات المجمعة ولعل من أهم استعمالات الوسط الهندسي هو إيجاد متوسط نسبة التغير .

<u>مثال 15 :</u> أحسب الوسط الهندسي للتوزيع التكراري الآتـي :

الفئات	115 .	125 .	135 .	145 .	155−165	المجموع
التكرار f	10	20	40	20	10	100

الحل :

f Logx	Log x	مراكز الفئات x	التكرار f	الفئات
20. 792	2. 1792	120	10	− 115
42. 278	2. 1193	130	20	− 125
75. 844	2. 1461	140	40	− 135
42. 522	2. 1761	160	20	− 145
22. 041	2. 2. 41	160	10	165 − 155
214.477			100	المجموع

$$Log\ G = \frac{\sum f\ Logx}{\sum f}$$

$$= \frac{214.477}{100} = 2.14477$$

من جداول الأعداد المقابلة للوغاريتمات نجد أن :

$G = 139.\ 6$

مثال 16:

سكان أحد المدن كان 300.000 في عام 1980م, وبعد عشر سنوات وصل ألي 380.000 . أوجد متوسط نسبة التغير (أو النمو السنوي) ؟

الحل :

يمكن حساب معدل التغير بالمعادلة : $P_n = P_0(1+r)^n$

حيث P_0 = السكان في بداية المدة = 300000

P_n = ,, ،، نهاية المدة = 380000

r = نسبة التغير أو النمو

n = عدد السنوات

بالتعويض في المعادلة : $380{,}000 = 300{,}000(1+r)^{10}$

بأخذ لوغاريتم الطرفين نجد أن : r = 2.39 أو تقريبا 2.4 %

3-6 الوسط التوافقي:

الوسط التوافقي H هو مقلوب الوسط الحسابي لمقلوبات القيم المراد إيجاد متوسطها .

عليه فإن الوسط التوافقي لمجموعة n : $x_1 , x_2 , x_3 , ..., x_n$

$$H = \frac{n}{(\frac{1}{x_1} + \frac{1}{x_2} + ... + \frac{1}{x_n})}$$

$$H = \frac{n}{\sum \frac{1}{x_i}}$$

ولتسهيل الحساب نكتب المعادلة على الصورة

$$\frac{1}{H} = \frac{\sum \frac{1}{x}}{n} = \frac{1}{n} \sum \frac{1}{x_i}$$

مثال 17 :

الوسط التوافقي للعددين 3 ، 12 هي :

$$\frac{1}{2}(\frac{1}{3} + \frac{1}{12}) = \frac{1}{2}(\frac{5}{12}) = 4.8$$

إذن الوسط التوافقي لهذه الأعداد = 4.8

بينما الوسط الحسابي لهذه الأعداد = 7.5

بينما الوسط الهندسي لهذه الأعداد = 6.0

أما في حالة البيانات المبوبة يكون الوسط التوافقي كالتالي : $H = \frac{\sum f_i}{\sum \frac{f_i}{x_i}}$

مثال 18 :

أوجد الوسط التوافقي للأرقام 5 , 8 , 10 , 12 . ثم لاحظ الفرق بين المتوسطات الثلاث لهذه الأرقام.

الحل:

الوسط التوافقي = $\frac{1}{4}(\frac{1}{5} + \frac{1}{8} + \frac{1}{10} + \frac{1}{12}) = 7.87$

الوسط الحسابي = $\overline{X} = \frac{35}{4} = 8.75$

وعليه نستنتج أن أكبرها هو الوسط التوافقي ثم الحسابي ثم الهندسي بالترتيب .

مثال 19 :

نفرض أن المسافة بين مدينة أ ، ومدينة ب ، 200 كم ، ولنفرض بأن سيارة قطعت أل 50 كم الأولى بسرعة 25 كم/ ساعة وال 50 كم الثانية بسرعة 50 كم / ساعة ، ثم قطعت أل 50 كم الثالثة بسرعة 75 كم / ساعة ، ثم قطعت أل 50 كم الأخيرة بسرعة 100 كم / ساعة .

أحسب معدل السرعة ، بالمقارنة بين الوسط الحسابي والوسط التوافقي .

الحل :

$$\overline{X} = \frac{25+50+75+100}{4} = 62.5$$ كم/ساعة

وهذا غير صحيح لأن السيارة قطعت المسافة الأولى في ساعتين والمسافة الثانية في ساعة واحدة و المسافة الثالثة في $\frac{2}{3}$ ساعة أي أنها

أخذت زمنا قدره $= 2+1+\frac{2}{3}+\frac{1}{2} = 4\frac{1}{6}$ ساعة لقطع المسافة من مدينة أ ، ب

متوسط السرعة $= \frac{200}{4\frac{1}{6}} = 48$ كم/ ساعة

يمكن الحصول على هذا المتوسط كما يلي :.

$$H = \frac{N}{\frac{1}{X_1}+\frac{1}{X_2}+\frac{1}{X_3}+\frac{1}{X_4}}$$

$$H = \frac{4}{\frac{1}{25}+\frac{1}{50}+\frac{1}{75}+\frac{1}{100}=\frac{4}{0.083}} = 48$$ كم/ساعة

3-7 : العلاقة بين الوسط والوسيط والمنوال

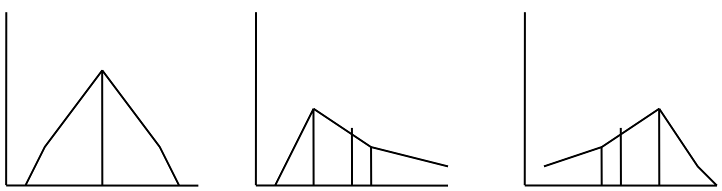
إذا كانت مجموعة البيانات متماثلة تماماً ووحيدة المنوال (لها منوال واحد) ، فإن قيم الوسط والوسيط والمنوال تكون متساوية ، والبيانات المتماثلة عند رسمها بيانياً نلاحظ أن العمود الذي يصل قمة المنحنى بقاعدته تتساوى عندة القيم الثلاث ، كما أنه يقسم الرسم ألي نصفين كل منها مرآة الآخر ، أما إذا كانت البيانات وحيدة المنوال ولكنها غير متماثلة ، فإن هذه المقاييس ستقع عند نقط مختلفة من التوزيع . فإذا كان أحد جوانب المنحني ممتداً ألي ناحية معينة فان الوسط الحسابي ينجذب ألي أبعد نقطة في نفس الاتجاه . ويقع المنوال دائماً عند أعلى نقطة في التوزيع أما الوسط والوسيط فإنهما مجذوبان في اتجاه الطرف

الممتد من التوزيع بحيث يقع الوسيط باستمرار في المساحة الفاصلة بين المنوال والوسط الحسابي . ويطلق على أي توزيع غير متماثل حول وسطه الحسابي بأنه ملتوٍ . وهذا الالتواء قد يكون ناحية اليمين أو الشمال وهذا الموضوع سوف نتناوله في الفصل القادم .

وباختصار فإن كل مقياس من هذه المقاييس الثلاث له مميزاته وصفاته التى تجعله مناسباً في توزيع معين وغير مناسب لتوزيع آخر وأكثر هذه المقاييس استخداماً هو الوسط الحسابى وبالذات في المسائل المتعلقة بالتقدير الإحصائي ، إلا أنه يُعاب عليه انه يتأثر بالقيم المتطرفة ولا يصلح في البيانات التي تحتوى على فئات مفتوحة ، أما الوسيط فانه أغلب استخداماته مع البيانات التي تحتوى على قيمة متطرفة كما يُفضل استخدام الوسيط في حالة البيانات التي تحتوى على فئات مفتوحة ، إلا أنه يعاب على الوسيط كونه لا يعتمد في حسابه على جميع القيم الواردة في التوزيع .أما المنوال فإن استخداماته مهمة في حالة اهتمامنا بدراسة الملاحظات المتكررة الحدوث ، مثلاً في أحد المصانع أو المحلات ينصب اهتمامنا على نوع محدد أو حجم معين من أحد الأصناف التي يتعاملون بها سواءاً في التخزين أو لأجل الشراء.

وتوجد علاقة تجريبية بين المقاييس الثلاث يمكن التعبير عنها كمياً كالآتي :

1. اذا كان التوزيع التكرارى قريب من التماثل أي قريب من شكل المنحنى الطبيعي (شكل الجرس) فإن : الوسط الحسابي – المنوال = 3 (الوسط الحسابي – الوسيط) .
2. أما إذا كان التوزيع التكراري متماثلاً ، فإن : الوسط الحسابي = الوسيط = المنوال .



أمثلة محلوله

مثال 1 :

يبين الجدول التالي توزيع درجات أعمال السنة لعدد 40 طالباً في مقرر الفيزياء لأحد الكليات.

الدرجات	10 −	14 −	18 −	22 −	26 −
عدد الطلاب	3	8	13	10	6

المطلوب :

1. أحسب الوسط و الوسيط و المنوال للتوزيع.
2. إذا قرر الأستاذ إعطاء تقدير ممتاز لأفضل 25% من المجموعة فما هو الحد الأدنى للدرجة التى ستمكن الطالب من الحصول على تقدير ممتاز .

الحل :

جدول التوزيع التكراري

الفئات	X	F	Xi Fi
10-14	12	3	36
14-18	16	8	128
18-22	20	13	260
22-26	24	10	240
26-30	28	6	168
		40	832

1. الوسط الحسابي

$$\overline{\mathbf{X}} = \frac{\sum \mathbf{X_i F_I}}{\mathbf{N}} = \frac{832}{40} = 20.8$$

2. الوسيط

$$\tilde{\mathbf{X}} = 18 + (\frac{20-11}{13}) \times 4 = 20.7692$$

3. المنوال

$$\mathbf{MD} = 18 + (\frac{5}{5+3}) \times 4 = 20.5$$

ترتيب الربيع الأعلى

$$\mathbf{Q}_3 = \frac{3 \times 40}{4} = 30$$

$$\mathbf{Q}_3 = 22 + (\frac{30-24}{10}) \times 4 = 24.4$$ الربيع الأعلى

$\therefore$ فإن تقدير ممتاز سيُمنح للطالب الحاصل على 24.5 درجة فما فوق أي بمعنى أن ($X \leq 24.5$ درجة) .

<u>مثال 2 :</u>

سافر أحمد من المدينة س ألي المدينة ص التي تبعد 600 كم عن الأولى ، قطع الثلث الأول من الطريق بسرعة 140 كم/ساعة ، والثلث الثاني بمعدل 120 كم / ساعة ، بينما الثلث لأخير قطعه بمعدل 100 كم / ساعة . أحسب كم كانت متوسط سرعته خلال كامل الرحلة ؟

<u>الحل :</u>

الوسط التوافقي = $$\mathbf{H} = \frac{1}{\mathbf{N}} \sum \frac{1}{\mathbf{X}}$$

$$\frac{1}{H} = \frac{1}{3}(\frac{1}{100} + \frac{1}{120} + \frac{1}{140})$$

$\therefore \quad \mathbf{H} = 117.75 \cong$ 118 كم / ساعة

<u>مثال 3 :</u>

إذا كان المتوسط العام لأحد الطلبة في أربع مواد هو 74.56 وكانت درجاته فى هذه المواد هي 70 ، 78 ، 65 ، 80 . أوجد عدد وحدات المادة الرابعة إذا علمت أن عدد وحدات المواد الأخرى هي :

3 ، 2 ، 1 على التوالي .

<u>الحل :</u>

الوسط المرجح للدرجات بالوحدات الدراسية =

$$X_w = \frac{70 \times 3 + 78 \times 2 + 65 \times 1 + 80 \times \mathbf{y}}{3 + 2 + 1 + \mathbf{y}} = 74.56$$

$$\therefore 431 + 80\mathbf{y} = 447.36 + 74.56\mathbf{y}$$

$$\therefore 5.44\mathbf{y} = 16.36$$

$$\therefore \mathbf{y} = 3.0073 \cong 3$$

عليه تكون وحدات المادة الرابعة هي 3

مثال 4 :

إذا كان المتوسط الهندسي لمجموعة من 6 قيم هو 1.24 وأن 5 من هذه القيم معلومة لدينا وهى : 1.18 ، 1.32 ، 1.27 ، 1.15 ، 1.22 . أوجد القيمة السادسة لهذه المجموعة.

الحل :

$$\mathbf{G = \sqrt[N]{Xi}}$$

$$124 = \sqrt[6]{X} = (X)^{\frac{1}{6}}$$

$$= \sqrt[6]{(1.18).(1.32)\ldots(1.22).\mathbf{X}}$$

$$= \sqrt[6]{2.7753473\mathbf{X}}$$

$$1.24 = (2.77\,X)^{\frac{1}{6}}$$

$$Log1.24 = \frac{1}{6}(Log2.77 + LogX)$$

$$6 \times Log1.24 - Log2.7753473 = LogX$$

$$\therefore \quad LogX = 0.1172127$$

$$\therefore \quad \mathbf{X} = 1.3098$$

مثال 5 :

قام مدرس مادة الكيمياء بوضع امتحان شهري لطلبته ، ثم قام بتسجيل عدد المسائل التي توصل كل طالب ألي إجابتها الصحيحة خلال عشرة دقائق فقط. وفيما يلي درجات 40 طالب :

6	8	12	5	10	9	10	5	8	7
13	9	5	10	11	10	5	6	11	10
8	8	6	7	15	10	8	8	12	8
14	5	5	7	8	14	7	9	6	5

احسب ما يلي لهذا التوزيع :

1. أحسب الوسط والوسيط والمنوال .
2. الربيعين الأدنى والأعلى ، والمئين العاشر .

الحل :

نبدأ أولاً بتكوين جدول وتوزيع التكراري لدرجات الطلبة ثم نحسب المطلوب :

الفئات	Xi	Fi	Xi Fi
7−5	6	11	66
9−7	8	12	96
11−9	10	9	90
13−11	12	4	48
15−13	14	4	56
		40	356

$$\overline{X} = \frac{\sum x_i f_i}{\sum f_i} = \frac{356}{40} = 8.9$$

ترتيب الوسيط : $\frac{\sum f}{2} = \frac{40}{2} = 20$

الوسيط = $\tilde{\mathbf{X}} = 7 + (\frac{20-11}{12}) \times 2 = 8.5$

المنوال = $\mathbf{Md} = 7 + (\frac{1}{3+1}) \times 2 = 7.5$

$$\mathbf{Q}_1 = 5 + (\frac{10-0}{11}) = \times 6.8$$

$$\mathbf{Q}_3 = 9 + (\frac{30-23}{9}) \times 2 = 10.56$$

$$\mathbf{P}_{10} = 5 + (\frac{4-0}{11}) \times 2 = 5.73$$

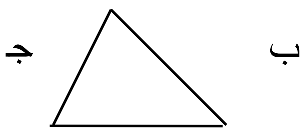
مثال 6 :

ثلاث مدن أ و ب و ج ، تقع كل منها على مسافات متساوية عن بعضها البعض .
قام أحد السائقين بالسفر من المدينة أ ألي المدينة ب بسرعة 30 كم / ساعة ، ثم من ب ألي ج بسرعة 40 كم / ساعة ، وعاد أخيرا من ج ألي أ بسرعة قدرها 50 كم /

ساعة . أحسب المتوسط التوافقي لسرعة هذا السائق خلال الرحلة كلها ، وهل يختلف هذا المتوسط لو أنه متوسطا حسابيا .

أ

الحل :



المتوسط الحسابي = $\overline{\mathbf{X}} = \frac{120}{3} = 40$

المتوسط التوافقى = $\frac{\mathbf{N}}{\sum \frac{1}{\mathbf{X}}} = \frac{3}{\frac{1}{30} + \frac{1}{40} + \frac{1}{50}} = 38.30$

$\therefore$ نلاحظ أن المتوسط التوافقي أقل من المتوسط الحسابي .

مثال 7 :

أزداد سكان أحد المجتمعات خلال الفترة من 1984 ألي 1994 من 200000 نسمة ألي 280000 نسمة . أحسب معدل النمو السنوي للسكان في هذا المجتمع خلال الفترة المذكرة .

الحل :

حيث أن هذه زيادة سنوية مركبة أي معدل للنمو فإننا نطبق صيغة المتوسط الهندسي المعطاة بالمعادلة التالية :

$$P_n = P_0(1+r)^n$$

حيثP_n هو عدد السكان في السنة النهائية .n هي عدد السنوات

P_0 هو عدد السكان في السنة الأساس . r معدل النمو السنوي

$$280000 = 200000(1+r)^{10}$$

بأخذ لوغاريتمات الطرفين ثم إيجاد اللوغاريتم المقابل نجد أن $r = 3.42\%$

وهذه هي معدل النمو السنوي لسكان هذا المجتمع خلال العشر سنوات .

تمارين

1_ فيما يلي توزيع تكراري :

الفئات	−10	−15	−20	−30	−40	60−50	المجموع
التكرار	100	120	320	280	160	20	1000

المطلوب حساب :

أ – الوسط الحسابي ب – الوسط التوافقي ج – الوسيط د – المنوال

2 _ أحسب الوسط الحسابي للتوزيع الآتي :

الفئات	−8	−10	−12	−14	−16	−18	المجموع
التكرار	6	8	15	15	10	6	60

3 – الآتي توزيع تكراري :

الفئات	20-16	25-21	30-26	35-31	40-36	45-41	50-46
التكرار	80	44	100	200	40	20	16

المطلوب / حساب الوسط الحسابي ، المنوال

4– أحسب الوسيط للتوزيع الآتي :

الفئات	أقل من 5	−5	−7	−9	اكثر من 11
التكرار	5	7	3	3	7

5 . الآتي توزيع تكراري :

الفئات	−10	−20	−30	−40	−50	−60	المجموع
التكرار	5	15	30	25	15	10	100

المطلوب حساب :

أ – الوسط الحسابي. ب – الوسيط . جـ –المنوال

6 . أحسب الوسط الهندسي للقيم الآتية :

x	70	80	90	100	110	120
f	20	30	40	50	60	70

7. إذا كان أحد المتسابقين يجب أن يقطع 300 كم على الشكل الآتـي :

المسافة الأولى 100 كم بسرعة 100 كم / ساعة

المسافة الثانية 100 كم بسرعة 160 كم / ساعة

المسافة الثالثة 100 كم بسرعة 40 كم / ساعة

أحسب الوسط التوافقي بسرعة هذا التسابق

8- أحسب الوسط الهندسي للتوزيع الآتي:

الفئات	-10	-20	-30	-40	-50	-60	المجموع
التكرار	20	30	50	60	25	15	200

الفصل الرابع

مقاييس التشتت (الاختلاف)
Measures of Dispersion (Variation)

(4−1) مقدمة:

لقد ذكرنا في الفصل السابق بعضًا من 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 (1):

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

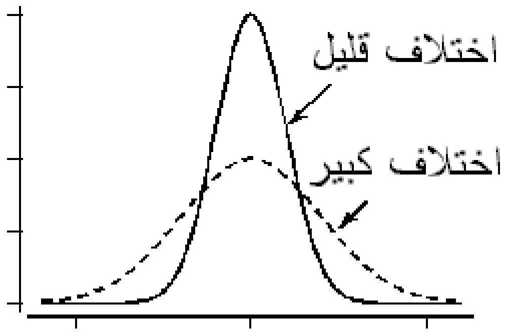
المجموعة	البيانات	شكل انتشار البيانات
الأولى	59, 61, 62, 60, 58, 60	• — ● — ● — ●— ● — ● — 60
الثانية	50, 60, 66, 54, 60, 70	• — ● — — ● — — — — ●— — — — ● — — ● — 60

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 (2):

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



شكل (1) المضلعان التكراريان لتوزيعين لهما نفس مقاييس
النزعة المركزية ولكنهما مختلفين في التشتت

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. 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

(2−4) المدى: Range

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

تعريف (1):

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

$$Range = X_{max} - X_{min} \tag{1}$$

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 (1):

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

X_{max} = الحد الأعلى للفترة العليا

X_{min} = الحد الأدنى للفترة الدنيا

مثال (3):

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

<u>الحل:</u>

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

مثال (4):

أوجد المدى لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (2) من الفصل الثاني.

الحل:

مستوى الهيموجلوبين	مركز الفترة x	التكرار f
12.95 – 13.95	**13.45**	3
13.9 5– 14.95	14.45	5
14. 95– 15.95	15.45	15
15. 95– 16.95	16.45	16
16. 95– 17.95	17.45	10
17. 95– 18.95	**18.45**	1

$$X_{max} = \text{مركز الفترة العليا} = 18.45$$

$$X_{min} = \text{مركز الفترة الدنيا} = 13.45$$

$$Range = X_{max} - X_{min} = 18.45 - 13.45 = 5.00$$

بعض مميزات وعيوب المدى:

أ- يتميز المدى بسهولة التعريف والحساب

ب- يعيب المدى العيوب التالية:

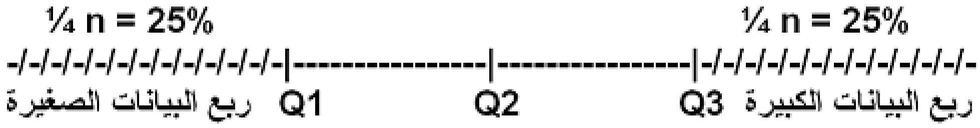
1. يتأثر المدى بالقيم الشاذة أو المتطرفة.
2. لا يأخذ المدى في الاعتبار جميع البيانات.

ملاحظة (2):

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.

(4−3) 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.



يرمز لنصف المدى الربيعي بالرمز Q ويعرف بالصيغة التالية:

$$Q = \frac{Q_3 - Q_1}{2} \qquad (2)$$

حيث أن Q_1 هو الربيع الأول و Q_3 هو الربيع الثالث وقد مر معنا كيفية إيجادهما للبيانات المبوبة بالطريقة الحسابية والبيانية.

مثال (5):

أوجد نصف المدى الربيعي لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (2) من الفصل الثاني باستخدام:

(أ) الطريقة الحسابية.

(ب) الطريقة البيانية.

الحل:

(أ) الطريقة الحسابية:

حساب الربيع الأول Q_1:

$$R = \frac{n}{4} = \frac{50}{4} = 12.5$$

$$A = 14.95,\quad L = 1.00,\quad F_1 = 8\ ,\ F_2 = 23$$

$$Q_1 = A + \left(\frac{R - F_1}{F_2 - F_1}\right) \times L = 14.95 + \left(\frac{12.5 - 8}{23 - 8}\right) \times 1.00 = 15.25$$

حساب الربيع الثالث Q_3:

$$R^* = \frac{3n}{4} = \frac{3 \times 50}{4} = 37.5$$

$$A^* = 15.95,\quad L^* = 1.00,\quad F_1^* = 23\ ,\ F_2^* = 39$$

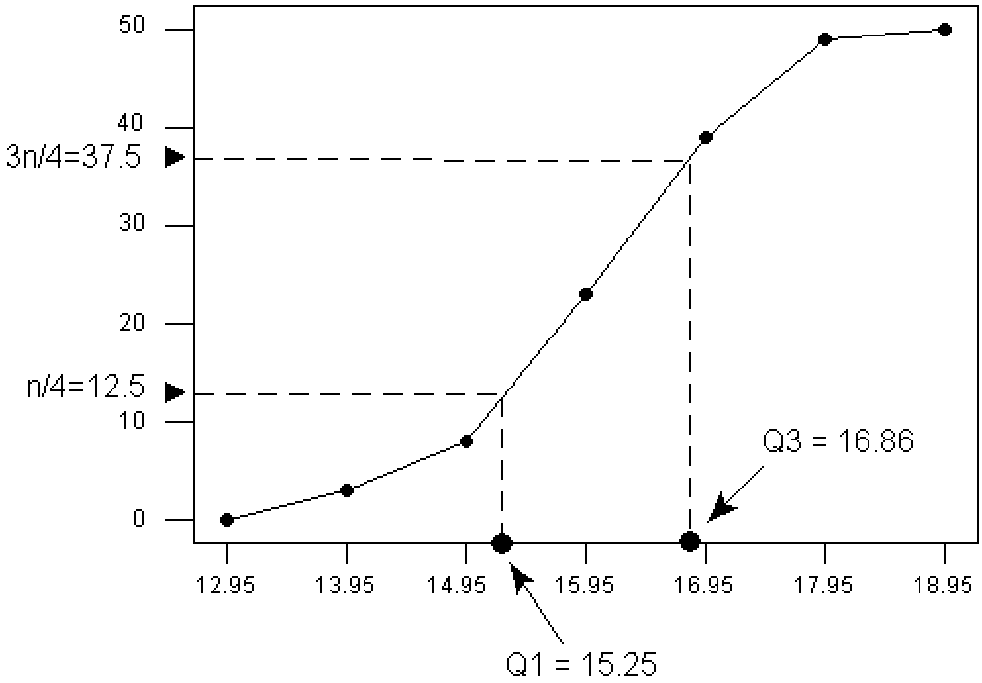
$$Q_3 = A^* + \left(\frac{R^* - F_1^*}{F_2^* - F_1^*}\right) \times L^* = 15.95 + \left(\frac{37.5 - 23}{39 - 23}\right) \times 1.00 = 16.86$$

وباستخدام الصيغة فإن نصف المدى الربيعي هو:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{16.86 - 15.25}{2} = \frac{1.61}{2} = 0.805$$

	مستوى الهيموجلوبين		التكرار المتجمع الصاعد	
	12.95	أقل من	0	
	13.9 5	أقل من	3	
	14. 95= A	أقل من	8 = F_1	← $R = \frac{n}{4} = 12.5$
$Q_1 \Rightarrow$	15. 95= A^*	أقل من	23 = F_2 = F^*_1	← $R^* = \frac{3n}{4} = 37.5$
$Q_3 \Rightarrow$	16. 95	أقل من	39 = F^*_2	
	17. 95	أقل من	49	
	18.95	أقل من	50	

(ب) الطريقة البيانية:



وباستخدام الصيغة فإن نصف المدى الربيعي هو:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{16.86 - 15.25}{2} = \frac{1.61}{2} = 0.805$$

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.

ملاحظة (3):

وحدة نصف المدى الربيعي هي نفس وحدة البيانات الأصلية.

(4-4) التباين (Variance) والانحراف المعياري (Standard Deviation):

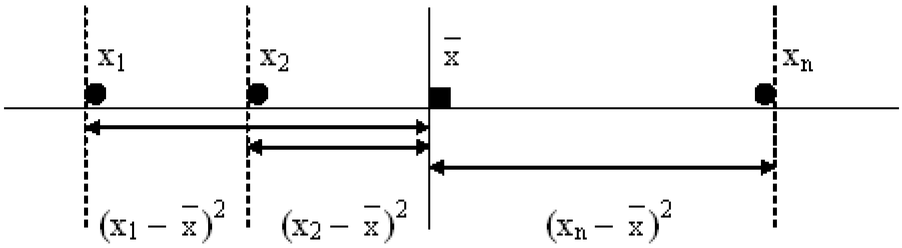
يعتبر التباين والانحراف المعياري من أهم وأفضل مقاييس التشتت ومن أكثرها شيوعًا واستخدامًا في التحليل الإحصائي وذلك لما يتمتعان به من خصائص وصفات إحصائية جيدة.

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

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.

حساب التباين والانحراف المعياري:

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

أولاً: 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:

$$s^2 = \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \cdots + (x_n-\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} \tag{3}$$

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:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} \tag{4}$$

ملاحظة (4):

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:

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

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^2 - n\left(\bar{x}\right)^2\right\} \qquad (6)$$

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 (6):

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

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:

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

حساب تباين العينة:

(أ) باستخدام التعريف:

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

(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:

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

بعض خصائص التباين والانحراف المعياري:

1. يخضع التباين والانحراف المعياري لبعض العمليات الجبرية. فإذا كان s^2 و s هما

علـى الترتيـب التبـاين والانحـراف المعيـاري للبيانـات $x_1,x_2,\ldots,x_n$ وكـان a و b مقدارين ثابتين، فإن:

أ- تبـاين البيانـات $x_1 \pm b, x_2 \pm b, \cdots, x_n \pm b$ هـو s^2 وأمـا انحرافهـا المعيـاري فهو S. لذلك فإن التباين والانحراف المعياري لا يتأثران بإضافة أو طرح مقدار ثابت من جميع المشاهدات.

ب- تبـاين البيانـات $ax_1, ax_2, \cdots, ax_n$ هـو a^2s^2 وأمـا انحرافهـا المعيـاري فهـو $|a|s$. حيث $|a|$ هي القيمة المطلقة للقيمة a.

ج- تبـاين البيانـات $ax_1 \pm b, ax_2 \pm b, \cdots, ax_n \pm b$ هـو a^2s^2 وأمـا انحرافهـا المعياري فهو $|a|s$.

د- تباين المقدار الثابت يساوي الصفر.

ويمكن تلخيص هذه الخاصية في الجدول التالي:

المشاهدات	التباين	الانحراف المعياري
$x_1,x_2,\ldots,x_n$	s^2	s
$x_1 \pm b, x_2 \pm b, \cdots, x_n \pm b$	s^2	s
$ax_1, ax_2, \cdots, ax_n$	a^2s^2	$\|a\|s$
$ax_1 \pm b, ax_2 \pm b, \cdots, ax_n \pm b$	a^2s^2	$\|a\|s$

مثال (7):

	المشاهدات	الانحراف المعياري	التباين
: x	2, 6, 4, 3, 5	s = 1.581	s^2= 2.5
: x+5	7, 11, 9, 8, 10	1.581	2.5
: 3x	6, 18, 12, 9, 15	$\|3\|\times1.581=4.743$	$9\times2.5=22.5$
: 3x +5	11, 23, 17, 14, 20	$\|3\|\times1.581=4.743$	$9\times2.5=22.5$

مثال (8):

إذا كـــان التبـــاين للمشـــاهدات $x_1,x_2,\ldots,x_n$ هـــو 36 فـــإن التبـــاين للمشـــاهدات

$\frac{x_1-10}{2},\frac{x_2-10}{2},\ldots,\frac{x_n-10}{2}$ هـو $\left(\frac{1}{2}\right)^2 36=\frac{36}{4}=9$ وأمـا الانحـراف المعيـاري فهـو $\sqrt{9}=3$.

2. 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 (9):

أوجد تباين المجموعة الكلية المكونة من دمج المجموعتين التاليتين:

	المجموعة الأولى	المجموعة الثانية
حجم العينة	$n_1=4$	$n_2=6$
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.

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

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

ثانياً: التباين والانحراف المعياري للبيانات المبوبة:

إذا كان لدينا بيانات عددها n وكانت هذه البيانات ملخصة في جدول تكراري بحيث أن:

أ- عدد الفترات هو k.

ب- مراكز الفترات هي $x_1, x_2, \ldots, x_k$.

ج- تكرارات الفترات هي $f_1, f_2, \ldots, f_k$.

بطريقة مشابهة لحسب المتوسط للتوزيع التكراري فإن التباين للتوزيع التكراري المبوب يمكن حسابه بشكل تقريبي بالصيغة التالية:

$$s^2 = \frac{\sum_{i=1}^{k} f_i \left(x_i - \bar{x}\right)^2}{n-1} \qquad (7)$$

حيث أن:

$$n = \sum_{i=1}^{k} f_i \ , \quad \bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum xf}{n} \ ,$$

كما يمكن استخدام إحدى الصيغتين الحسابيتين التاليتين:

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{k} x_i^2 f_i - \left(\sum_{i=1}^{k} x_i f_i\right)^2 / \, n\right\} \qquad (8)$$

$$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{k} x_i^2 f_i - n\left(\bar{x}\right)^2\right\} \qquad (9)$$

ويكمن تلخيص عمليتي إيجاد المتوسط والتباين باستخدام الجدول التالي:

الفترة	مركز الفترة x	التكرار f	$x\,f$	$x^2\,f$	$f\,(x-\bar{x})^2$
الفترة رقم 1	x_1	f_1	$x_1\,f_1$	$x_1^2\,f_1$	$f_1\,(x_1-\bar{x})^2$
الفترة رقم 2	x_2	f_2	$x_2\,f_2$	$x_2^2\,f_2$	$f_2\,(x_2-\bar{x})^2$
:	:	:	:	:	:
:	:	:	:	:	:
الفترة رقم k	x_k	f_k	$x_k\,f_k$	$x_k^2\,f_k$	$f_k\,(x_k -$ ر$)^2$
المجموع		$\sum f = n$	$\sum xf$	$\sum x^2 f$	$\sum_{i=1}^{k} f_i\,(x_i-\bar{x})^2$

مثال (10):

أوجد التباين والانحراف المعياري لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (2) من الفصل الثاني.

الحل:

نلخص إيجاد التباين باستخدام الجدول التالي:

مستوى الهيموجلوبين	مركز الفترة x	التكرار f	$x f$	$x^2 f$	$f(x-\bar{x})^2$ $f(x-16.01)^2$
12.95 – 13.95	13.45	3	40.35	542.708	19.6608
13.9 5– 14.95	14.45	5	72.25	1044.013	12.1680
14. 95– 15.95	15.45	15	231.75	3580.538	4.7040
15. 95– 16.95	16.45	16	263.20	4329.640	3.0976
16. 95– 17.95	17.45	10	174.50	3045.025	20.7360
17. 95– 18.95	18.45	1	18.45	340.403	5.9536
المجموع		$\sum f$ $= 50$	$\sum xf$ $= 800.5$	$\sum x^2 f$ $=12882.33$	$\sum_{i=1}^{k} f_i (x_i - \bar{x})^2$ $= 66.320$

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum xf}{n} = \frac{800.5}{50} = 16.01$$

حساب التباين باستخدام صيغة التعريف:

$$s^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{n-1} = \frac{66.320}{50-1} = 1.3535$$

حساب التباين باستخدام الصيغة الحسابية الأولى:

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

$$= \frac{1}{50-1}\left(12882.33 - \frac{(800.5)^2}{50}\right)$$

$$= \frac{1}{49}(12882.33 - 12816.005)$$

$$= \frac{66.325}{49} = 1.3536$$

حساب التباين باستخدام الصيغة الحسابية الثانية:

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

$$= \frac{1}{50-1}\{12882.33 - (50)(16.01)^2\}$$

$$= \frac{1}{49}\{12882.33 - 12816.005\}$$

$$= \frac{66.325}{49} = 1.3536$$

حساب الانحراف المعياري:

$$s = \sqrt{s^2} = \sqrt{1.3536} = 1.163$$

(4−5) 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 ينزع لأن يكون صغيرًا 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 تجانسًا والعكس بالعكس. ويعرف معامل الاختلاف للعينة التي متوسطها $\overline{x}$ وانحرافها المعياري s بالصيغة التالية:

$$C.V. = \frac{s}{\overline{x}} \qquad (10)$$

مثال (11):

الجدول أدناه يتضمن بيانات إحدى الدراسات التي طبقت على خمسة أشخاص لقياس الوزن (بالكيلوجرام) والطول (بالسنتيمتر). أي البيانات أكثر تشتتًا نسبيًا (أقل تجانسًا)؛ 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 تجانسًا من بيانات الأطوال.

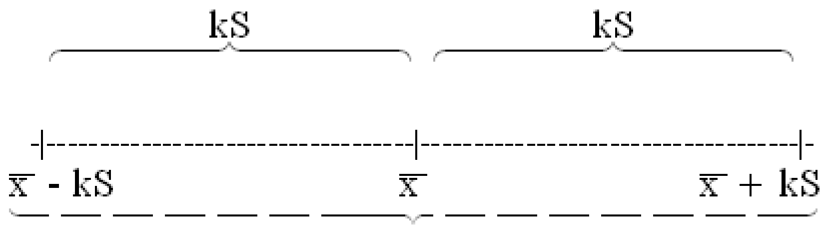
(6–4) نظرية (متراجحة) تشيبيشيف Chebychev Inequality:

إن نظرية تشيبيشيف من النظريات المفيدة إذا أنها تعطينا حدًا أدنى لنسبة البيانات الواقعة في فترة معينة عند معرفة متوسط البيانات وانحرافها المعياري دون الحاجة لمعرفة البيانات الأصلية أو التوزيع الذي أخذت منه العينة.

نظرية (1):

إذا كان لدينا عينة من البيانات متوسطها $\bar{x}$ وانحرافها المعياري s فإن نسبة البيانات الواقعة في الفترة $\left(\bar{x} - ks, \bar{x} + ks\right)$ لايقل عن $\left(1 - \frac{1}{k^2}\right)$ حيث أن $k > 1$.

الشكل التالي يوضح فكرة نظرية تشيبيشيف.



نسبة البيانات الواقعة في هذه الفترة لا يقل عن $\left(1 - \frac{1}{k^2}\right)$

ملاحظة (5):

1. في بعض الأحيان نكتب الفترة $\left(\bar{x} - ks , \bar{x} + ks\right)$ على الصورة $\bar{x} \pm ks$.
2. تطبق نظرية تشيبيشيف للفترات التي منتصفها (مركزها) هو المتوسط $\bar{x}$.
3. تستخدم نظرية تشيبيشيف بطريقتين(في كلا الحالتين لابد من معرفة قيمة k):
   - أ- تحديد النسبة (التقريبية) لعدد البيانات الواقعة في فترة معينة.
   - ب- تحديد فترة يقع فيها ما لا يقل عن نسبة معينة.

مثال (12):

إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x} = 7$ وانحرافها المعياري s=5 فما هي نسبة البيانات الواقعة في الفترة (18 , 4–)؟

الحل:

أولاً نلاحظ أن منتصف الفترة المعطاة (4,18–) هو المتوسط $\bar{x} = 7$ لذلك نستطيع تطبيق نظرية تشيبيشيف. والآن:

$$\left(\bar{x} - ks , \bar{x} + ks\right) = (-4 , 18) \quad \Rightarrow \quad \bar{x} + ks = 18$$

$$\Leftrightarrow \quad 7 + k\,(5) = 18$$

$$\Leftrightarrow \quad 5k = 11$$

$$\Leftrightarrow \quad k = 11/5$$

$$\Leftrightarrow \quad \left(1 - \frac{1}{k^2}\right) = \left(1 - \frac{1}{(11/5)^2}\right)$$

$$\Leftrightarrow \quad \left(1 - \frac{1}{k^2}\right) = 0.7934$$

لذلك فإن نسبة البيانات الواقعة في الفترة (4,18–) لا تقل عن 79.34%.

مثال (13):

إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x}=7$ وانحرافها المعياري $s=5$ فأوجد فترة يقع فيها ما لا يقل عن 75% من البيانات.

الحل:

$$\left(1-\frac{1}{k^2}\right)=0.75 \quad \Leftrightarrow \quad \frac{1}{k^2}=1-0.75$$

$$\Leftrightarrow \quad \frac{1}{k^2}=0.25$$

$$\Leftrightarrow \quad k=\sqrt{\frac{1}{0.25}}$$

$$\Leftrightarrow \quad k=2$$

وبالتالي فإن الفترة التي يقع فيها ما لا يقل عن 75% من البيانات هي:

$$\left(\bar{x}-ks,\bar{x}+ks\right) = (7-2\times5\ ,\ 7+2\times5)$$

$$= (7-10\ ,\ 7+10)$$

$$= (-3\ ,\ 17)$$

(7−4) الدرجات (القيم) المعيارية : Standard Scores (Values)

نستطيع بكل يسر وسهولة استخدام قيمتي مشاهدتين في نفس المجموعة لمقارنتهما ببعض. فمثلاً، نستطيع أن نقول بأن أداء الطالب الحائز على الدرجة 85 في اختبار مقرر ما أفضل من أداء الطالب الحائز على الدرجة 80 في نفس الاختبار إذا كان الطالبان في نفس الشعبة. وفي المقابل، لا نستطيع القول بأن أداء الطالب الحائز على الدرجة 85 في اختبار مقرر ما في الشعبة التي يدرسها المدرس (أ) أفضل من أداء طالب آخر حائز على الدرجة 80 في نفس الاختبار ولكنه في شعبة أخرى يدرسها المدرس (ب). من هنا، نرى أنه من الضروري إيجاد قيم لا تعتمد على الوحدات ويمكن استخدامها لمقارنة البيانات في المجموعات المختلفة. هذه القيم التي لا تعتمد على الوحدات نسميها بالقيم (أو الدرجات) المعيارية.

تعريف (2):

لتكن $x_1, x_2, \ldots, x_n$ عينه من البيانات حجمها n ومتوسطها $\bar{x}$ وانحرافها المعياري s.

نعرف الدرجة (القيمة) المعيارية للمشاهدة x_i بالصيغة التالية:

$$z_i = \frac{x_i - \bar{x}}{s}\,; i = 1, 2, \dots, n \qquad (11)$$

أي أن الدرجات المعيارية للبيانات $x_1, x_2, \dots, x_n$ هي:

$$z_1 = \frac{x_1 - \bar{x}}{s}\,,\; z_2 = \frac{x_2 - \bar{x}}{s}, \dots,\; z_n = \frac{x_n - \bar{x}}{s}$$

ملاحظة (6):

1. القيمة $z_i = \frac{x_i - \bar{x}}{s}$ هي الدرجة المعيارية للمشاهدة الأصلية x_i.
2. المشاهدة الأصلية للدرجة المعيارية z_i هي $x_i = \bar{x} + s\, z_i$.
3. الدرجات المعيارية هي قيم عديمة الوحدة ولذلك فإنها تستخدم للمقارنة بين المشاهدات المختلفة في المجموعات المختلفة للبيانات.
4. متوسط الدرجات المعيارية =0.
5. الانحراف المعياري للدرجات المعيارية يساوي =1.

مثال (14):

إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x} = 7$ وانحرافها المعياري $s = 5$ فأوجد:

1. الدرجة المعيارية للقيمة $x = 9$.
2. القيمة الأصلية للدرجة المعيارية $z = 0.1$.

الحل:

1. الدرجة المعيارية للقيمة $x = 9$ هي:

$$z = \frac{x - \bar{x}}{s} = \frac{9-7}{5} = \frac{2}{5} = 0.4$$

2. القيمة الأصلية للدرجة المعيارية $z = 0.1$ هي:

$$x = \bar{x} + s\,z = 7 + 5\times 0.1 = 7 + 0.5 = 7.5$$

مثال (15):

إذا كانت درجة أحد الطلاب في مقرر الإحصاء تساوي 82 ودرجته في مقرر الرياضيات تساوي 89، وإذا كان متوسط درجات الطلاب في مقرر الإحصاء يساوي 75 بانحراف

معياري يساوي 10 ومتوسط درجات الطلاب في مقرر الرياضيات يساوي 81 بانحراف معياري يساوي 16، ففي أي المقررين كان أداء الطالب أفضل؟

الحل:

نلخص إيجاد الدرجات المعيارية في الجدول التالي:

المقرر	المتوسط $\bar{x}$	الانحراف المعياري s	الدرجة x	الدرجة المعيارية $z = \frac{x - \bar{x}}{s}$
الإحصاء	75	10	82	$z = \frac{82-75}{10} = 0.7$
الرياضيات	81	16	89	$z = \frac{89-81}{16} = 0.5$

بما أن الدرجة المعيارية لمقرر الإحصاء 0.7 أكبر من الدرجة المعيارية لمقرر الرياضيات 0.5 فإن أداء الطالب في مقرر الإحصاء أفضل من أدائه في مقرر الرياضيات بالرغم من أن درجته في مقرر الإحصاء أقل من درجته في مقرر الرياضيات.

(8-4) حل أمثلة الفصل الرابع بإستخدام إكسل:

مثال (1):

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

<u>**الحل:**</u>

ادخل البيانات في صفحة من إكسل كالتالي:

=MAX(A2:A8) - MIN(A2:A8)

L	K	J	I	H	G	F	E	D	C	B	A	
											الوزن	1
									30	= المدى	25	2
											30	3
											40	4
											45	5
											35	6
											55	7
											50	8

لحساب المدى استخدمنا التعريف $R = MAX(x_1,x_2,...,x_n) - MIN(x_1,x_2,...,x_n)$

لاحظ اننا أدخلنا المشاهدات على شكل مجال A2:A8.

مثال (2):

أوجد المدى لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (2) من الفصل الثاني.

الحل:

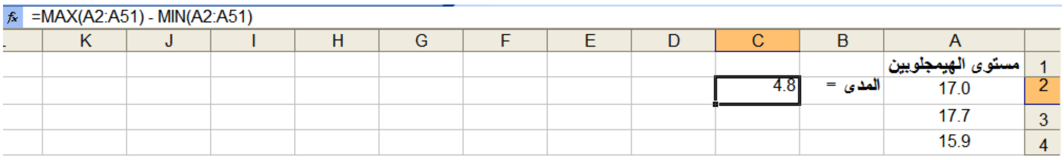
بالطرح المباشر

F	E	D	C	B	A	
			التكرار	مركز الفترة	مستوى الهيموجلوبين	1
من مركز الفئات	5	= المدى	3	13.45	12.95 – 13.95	2
من حدود الفئات	6	أو المدي =	5	14.45	13.9 5– 14.95	3
			15	15.45	14. 95– 15.95	4
			16	16.45	15. 95– 16.95	5
			10	17.45	16. 95– 17.95	6
			1	18.45	17. 95– 18.95	7

الحل بطريقة اخرى:

يتميز إكسل بمقدرته على التعامل مع عدد كبير جدا من المشاهدات والتي لانحتاج إلى تلخيصها في جدول تكراري إلا إذا كنا نحتاج ذلك لغرض إيجاد توزيع المشاهدات. لذلك من الأفضل التعامل مع المشاهدات كما هي في إكسل.

سوف نوجد المدى لمستوى الهيمجلوبين من المشاهدات 50 مباشرة كالتالي:



=MAX(A2:A51) - MIN(A2:A51)

	A	B	C
1	مستوى الهيمجلوبين		
2	17.0	المدى =	4.8
3	17.7		
4	15.9		

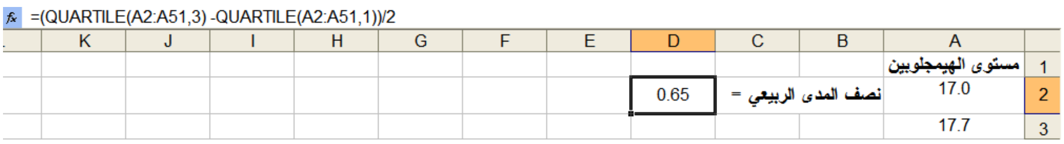
وهذه النتيجة أدق من النتيجة السابقة.

مثال (3):

أوجد نصف المدى الربيعي لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا.

الحل:

نستخدم الصيغة: =(QUARTILE(A2:A51,3) - QUARTILE(A2:A51,1))/2



=(QUARTILE(A2:A51,3) -QUARTILE(A2:A51,1))/2

	A	B	C	D
1	مستوى الهيمجلوبين			
2	17.0	نصف المدى الربيعي =		0.65
3	17.7			

مثال (4):

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

7.1, 2.5, 2.5, 5.4, 8.3

الحل:

التباين يحسب بالصيغة =VAR($x_1,x_2,...,x_n$) حيث تدخل المشاهدات على شكل مجال A2:A51 ويحسب الإنحراف المعياري من الصيغة =STDEV($x_1,x_2,...,x_n$) حيث تدخل المشاهدات على شكل مجال A2:A51 ايضا.

ملاحظة: بعد حساب التباين يمكن حساب الإنحراف المعياري بأخذ الجزر التربيعي بإستخدام الصيغة =SQRT(Number) حيث Number هو قيمة التباين.

	A	B	C
1	الوزن		
2	7.1	التباين =	=VAR(A2:A51)
3	2.5	الإنحراف المعياري =	=STDEV(A2:A51)
4	2.5	الإنحراف المعياري	
5	5.4	من التباين =	=SQRT(C2)
6	8.3		

مثال (5):

أوجد التباين والانحراف المعياري لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا.

الحل:

من المشاهدات 50 نوجد التباين والإنحراف المعياري كما في المثال السابق

	A	B	C
1	مستوى الهيمجلوبين		
2	17	التباين =	=VAR(A2:A51)
3	17.7	الإنحراف المعياري =	=STDEV(A2:A51)
4	15.9		
5	16.2		

	A	B	C
1	مستوى الهيمجلوبين		
2	17.0	التباين =	1.216526531
3	17.7	الإنحراف المعياري =	1.102962615
4	15.9		

مثال (6):

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

رقم الشخص	1	2	3	4	5
الوزن	69	59	65	67	65

158 165 155 162 164	الطول

الحل:

لحساب معامل الإختلاف نستخدم الصيغة

$$=STDEV(x_1,x_2,...,x_n)/AVERAGE(x_1,x_2,...,x_n)$$

	A	B	C	D	E
1	رقم الشخص	الوزن	الطول		
2	1	69	164	معامل الإختلاف للوزن =	=STDEV(B2:B6)/AVERAGE(B2:B6)
3	2	59	162	معامل الإختلاف للطول =	=STDEV(C2:C6)/AVERAGE(C2:C6)
4	3	65	155		
5	4	67	165		
6	5	65	158		

	A	B	C	D	E
1	رقم الشخص	الوزن	الطول		
2	1	69	164	معامل الإختلاف للوزن =	0.05756396
3	2	59	162	معامل الإختلاف للطول =	0.026163786
4	3	65	155		
5	4	67	165		
6	5	65	158		

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 تجانسًا من بيانات الأطوال.

(4–9) تمارين:

1. البيانات التالية عبارة عن ألوان عينة لنوع من الزهور:

حمراء، بيضاء، صفراء، زرقاء، زرقاء، بيضاء، بيضاء، خضراء، حمراء، صفراء، بيضاء، زرقاء، حمراء، صفراء، خضراء، زرقاء، خضراء، خضراء، بيضاء، زرقاء، صفراء، زرقاء، بيضاء، حمراء، خضراء، خضراء، صفراء، حمراء، خضراء، حمراء، حمراء، حمراء، زرقاء، زرقاء، صفراء، زرقاء، صفراء، صفراء، خضراء، صفراء، حمراء، حمراء، حمراء، خضراء، بيضاء، خضراء، بيضاء، بيضاء صفراء، خضراء

أ. أي من مقاييس التشتت يستخدم مع هذه البيانات؟

ب. أوجد مقياس أو مقاييس تشتت للبيانات إذا وجد.

ج. هل يمكن حساب مدى لهذه البيانات.

2. البيانات التالية تمثل أطوال طلاب أحد شعب 101 إحص (لأقرب سم):

166	170	167	169	166	166	168	168	170	166	167
168	170	168	166	169	169	168	169	170	167	168
168	167	167	167	169	167	170	169	169	167	166
169	167	167	167	170	169	170	170	168	170	168
169	170	166	166	169	166					

أ. أي من مقاييس التشتت يستخدم مع هذه البيانات؟

ب. أوجد مقياس أو مقاييس تشتت للبيانات.

ج. أوجد معامل الإختلاف والدرجات المعيارية للبيانات وأوجد فترة يقع فيها ما لا يقل عن 75% من البيانات.

د. أستخدم إكسل لحل الفقرات (ب و ج).

3. جمع عالم تصنيف نباتات عينة من خمسة انواع من النباتات في رحلة برية وحين فرزها وجد التالي:

الصنف	الأول	الثاني	الثالث	الرابع	الخامس
العدد (نبتة)	20	35	15	25	10

أ. أي من مقاييس التشتت يستخدم مع هذه البيانات؟

ب. أوجد مقياس أو مقاييس تشتت للبيانات.

ج. أحسب معامل الاختلاف إذا وجد.

4. يجري طبيب صيدلي تجارب لأحد العقاقير على عينة من الفئران وسجل النتائج التالية:

مجال الوزن (جرام)	100 − 51	150 −101	200 − 151	250 − 201
التكرار (عدد الفئران)	20	25	15	10

أ. أي من مقاييس التشتت يستخدم مع هذه البيانات؟

ب. أوجد مقياس أو مقاييس تشتت للبيانات.

ج. أحسب معامل الاختلاف والدرجات المعيارية إذا وجد أي منها.

د. أستخدم إكسل لحل الفقرات (ب و ج).

5. البيانات التالية تمثل تقديرات 100 طالبًا في شعبتين من مادة 101 إحص.

```
B B D B B E E C A F D F C B F C A B B E
B B D D B D B D A D D E D D D C A B F C
F B B A B B E E D E B B D B B E E C A F
D F C B F C A B B E B B D D B D B D A D
D E D D D C A B F C F B B A B B E E D E
```

أ. أي من مقاييس التشتت يستخدم مع هذه البيانات؟

ب. أوجد مقياس أو مقاييس تشتت للبيانات.

ج. هل يمكن حساب درجات معيارية لهذه البيانات.

6. البيانات التالية عبارة عن أوزان 50 سمكة (بالكيلوجرام):

```
2.5 4.0 1.5 1.5 2.5 3.0 4.0 2.5 2.5 2.0
2.5 1.5 1.5 2.5 4.0 1.5 3.0 2.0 3.5 1.5
2.5 1.0 3.0 2.0 1.5 2.5 3.5 4.0 2.5 3.0
3.5 1.5 2.5 3.5 2.0 1.5 3.5 2.5 2.5 2.5
2.0 2.5 3.5 3.5 2.0 2.5 2.0 2.0 1.5 1.5
```

بإستخدام إكسل أوجد:

أ. المدى.

ب. نصف المدى الربيعي.
ج. التباين.
د. الإنحراف المعياري.
هـ. معامل الإختلاف.
و. الدرجات المعيارية.

7. سجل أحد الطلاب في 6 مواد لفصل دراسي والجدول التالي يبين عدد الساعات لكل مادة وتقدير الطالب في كل مادة

المادة	1	2	3	4	5	6
عدد الساعات	2	3	3	5	1	2
الدرجات	95	83	75	80	70	90

أ. أي من مقاييس التشتت يمكن حسابه من هذه البيانات؟
ب. أحسب على الأقل ثلاثة من مقاييس التشتت إذا وجد.
ج. احسب معمل الاختلاف إذا وجد.
د. هل يمكن حساب درجات معيارية للبيانات؟ وأحسبها إذا أمكن ذلك.
هـ. حل الفقرات (ب) و (ج) و (د) بإكسل إذا أمكن ذلك.

8. البيانات التالية عبارة عن الرواتب الشهرية لمنسوبي أحد الشركات بآلاف الريال:
1 6 12 7 3 8 11 6 12 7 2 13 8 2 15 9 3 7
14 11

أي مقاييس تشتت تمثل هذه البيانات بشكل عادل.