الفصل الثالث

إحصائيات وصفية لمتغيرات أحادية البعد

مثال 1 : لتكن السلسلة الإحصائية التالية التي توضح توزيع الطلاب قسم السنة الثانية رياضيات في جامعة بسكرة لسنة 1998.

العمر	التكرارات
$[18-19]$	1
$]19-20]$	10
$]20-21]$	3
$]21-22]$	2
$]22-23]$	1
المجموع	17

و هو جدول توزيع السلسلة الإحصائية.

بالنسبة لسلسلة ذات متغيرات منفصلة ، لدينا n قيمة مختلفة x_i للسلسلة ذات الطابع الحقيقي ، و k فئة مختلفة والتي نرمز لها $C_i =]a_i, a_i + 1]$.

ستكون الأقواس المربعة بهذا الشكل $C_i = [a_i, a_i + 1[$ باستثناء الفئة الأولى التي تتضمين الحد الأدنى لها أو غير ذلك من النموذج وفي هذه الحالة سيتم ادراج القيمة الحدية العليا في الفئة الأخيرة .

مثال 2 : في المثال السابق *1*: أعلاه ، توجد 5 فئات. كل فئة C_i مرتبطة برقم n_i.

وبالمثل ، كما في حالة المتغير المنفصل، يمكننا تحديد إجمالي التكرارات والترددات الجزئية والترددات التراكمية المتزايدة والترددات التراكمية المتناقصة.

$$n = \sum_{i=1}^{k} n_i.$$

مثال 3 : في المثال السابق *1:* أعلاه لدينا:

$$n = 1 + 10 + 3 + 2 + 1 = 17.$$

إجمالي التكرارات

1.3 جدول التواترات أو التواترات الجزئية

يتم حساب التواتر f_i للفئة C_i للمتغير، بالمعادلة

$$f_i = \frac{n_i}{n}.$$

مثال 1 : في المثال *1:* أعلاه ، جدول التواتر هو:

العمر C_i	$[18-19]$	$]19-20]$	$]20-21]$	$]21-22]$	$]22-23]$	المجموع
التواترات $f_i\,(\%)$	5.9	58.8	17.7	11.7	5.9	100

1.1.3. جدول التواترات التراكمية المتزايد

لكل فئة C_i من المتغير، يتم حساب التواتر التراكمي المتزايد F_i بواسطة الصيغة التالية:

$$F_i = \sum_{j \leq i} f_j.$$

مثال 2 : في المثال *1:* أعلاه ، جدول التواتر التراكمي المتزايد هو:

العمر C_i	$[18-19]$	$]19-20]$	$]20-21]$	$]21-22]$	$]22-23]$
التواتر التراكمي المتزايد $F_i\,(\%)$	5.9	64.7	82.4	94.1	100

2.1.3. جدول التواترات التراكمية المتناقص

لكل فئة C_i من المتغير، يتم حساب التواتر التراكمي المتناقص G_i بواسطة الصيغة التالية:

$$G_i = 1 - F_i,$$

حيث

$$G_i = 100 - F_i$$

إذا تم إعطاء F_i بالنسب المئوية.

مثال 3 : في المثال *1*: أعلاه ، جدول التواتر التراكمي المتناقص هو:

العمر C_i	$[18-19]$	$]19-20]$	$]20-21]$	$]21-22]$	$]22-23]$
التواتر التراكمي المتناقص $G_i\,(\%)$	94.1	35.3	17.6	5.9	0

ملاحظة 1 : ★ القيمة الأخيرة هي دائماً 1 (أو 100 إذا تم إعطاؤها بنسب مئوية) للتواترات التراكمية المتزايدة.

★ القيمة الأخيرة هي دائماً 0 بالنسبة للتواترات التراكمية المتناقصة.

2.3 مقاييس النزعة المركزية

تشير النزعة المركزية إلى موقع التوزيع . وأهم مقاييس النزعة المركزية هي: المتوسط الحسابي، الوسيط ، المنوال. وسوف نقوم بقياس هذه المقاييس بالنسبة للمجتمعات (بمعنى مجموعات تشمل جميع العناصر موضع الدراسة) وبالنسبة لعينات مسحوبة من المجتمعات وكذلك بالنسبة للبيانات المبوبة والبيانات غير المبوبة. وهي مقاييس تساعد على ما يلي:

★ التعبير عن ترتيب حجم البيانات الإحصائية،

★ معرفة اتجاه المتغير الإحصائي.

هذا يقودنا إلى وضع مكان البيانات عند مقياس محدد ومعرفة القيم التي توجد بها البيانات.

1.2.3. القيم القصوى Maximum & minimum

أنها توفر لنا الحد العلوي (أكبر قيمة la plus grande valeur) والحد الأدنى (أصغر قيمة la plus petite valeur) في البيانات التي يمكن أن تكون أيضا قيم متطرفة valeurs aberrantes. هذا يسمح لنا بتحديد مقياس للرسومات على سبيل المثال. نمثلهم بواسطة: الحد الأعلى Max (maximum) أو الحد الأدنى Min (minimum).

$$Min = \min_i (x_i), \quad Max = \max_i (x_i).$$

2.2.3. المتوسط الحسابي La moyenne

تعريف 1.2.3 : المتوسط الحسابي *La moyenne arithmétique*، أو الوسط الحسابي، وأحياناً المعدّل في الرياضيات والإحصاء هو قيمة تتجمع حولها قيم مجموعة ويمكن من خلالها الحكم على بقية قيم المجموعة، فتكون هذه القيمة هي المتوسط الحسابي. لمجتمع ما يرمز μ (الحروف اليوناني ميو) ، أما المتوسط الحسابي لعينة ما فيرمز له بالرمز $\overline{X}$.

تعريف 2.2.3 : من أجل سلسلة إحصائية بسيطة ذات متغير منفصل، المتوسط الحسابي μ هو مجموع كل القيم مقسوما على العدد الإجمالي لها. بمعنى آخر:

$$\mu = \frac{1}{n}\sum_{i=1}^{k} n_i x_i.$$

و بالنسبة لسلسلة إحصائية بسيطة ذات متغير حقيقي (أو مستمر) ، يكون المتوسط الحسابي هو مجموع مراكز الفئات مقسوماً على عدد الفئات (عدد الأشكال *modalités*). وبعبارة أخرى

$$\mu = \frac{1}{n}\sum_{i=1}^{k} n_i \left(\frac{a_i + a_{i+1}}{2}\right),$$

حيث k هو عدد القيم (الفئات على التوالي $C_i =]a_i, a_{i+1}]$ المختلفة للمتغير، و n_i هو العدد المرتبط بهذه القيم (الفئات على التوالي) و n هو إجمالي عدد الأفراد أو التكرارات.

مثال 1 : بإستخدام الأمثلة *1:* و *5:* من بداية الفصل ؛ لدينا:

- من أجل السلسلة المنفصلة:

$$\begin{aligned}\mu &= \frac{7\cdot 0+23\cdot 1+40\cdot 2+13\cdot 3+11\cdot 4+3\cdot 5+2\cdot 6+1\cdot 7}{100}\\ &= 2.2\\ &\simeq 2 \text{ طفل.}\end{aligned}$$

- أما السلسلة المستمرة:

$$\begin{aligned}\mu &= \frac{1\cdot 18.5+10\cdot 19.5+3\cdot 20.5+2\cdot 21.5+1\cdot 22.5}{17}\\ &= 20.0294 \text{ سنة}\\ &\simeq 20 \text{ سنة و } 10 \text{ أيام}\end{aligned}$$

ملاحظة 1 : ويطلق إسم السلسلة المتمركزة، إذا كان متوسطـها الحسابي معدوما.

3.2.3. المتوسط التوافقي Moyenne harmonique

تعريف 3.2.3 : إذاكان $x_i \neq 0$ فإننا نسمي القيمة

$$H = \frac{n}{\sum_{i=1}^{n} 1/x_i}$$

بالمتوسط التوافقي.

تمرين 1 : يقطع دراج 4 مراحل من 100 كيلومتر تبلغ سرعات هذه المراحل كما يلي: 10 كم\ساعة ، 30 كم\ساعة ، 40 كم\ساعة ، 20 كم\ساعة. ماذا كان متوسط سرعته؟

الحـــل

بحساب بسيط فإن الدراج قد أكمل المرحلة الأولى بـ 10 ساعات، والثانية بـ 3 ساعات و20 دقيقة ، والثالثة بساعتين و30 دقيقة ، والرابعة بـ 5 ساعات. لذلك غطى ما مجموعه 400 كم في

$$10 + 3h20 + 2h30 + 5h = 20h50 = 20.8333 \text{ ساعة},$$

سرعته المتوسطة V_M هي:

$$V_M = \frac{400}{20.8333} = 19.2 \text{كم\الساعة}.$$

★ إذا قمنا بحساب المتوسط الحسابي للسرعات ، نحصل عليه

$$\overline{x} = \frac{10+30+40+20}{4} = 25 \text{ كم\textbackslash الساعة}.$$

★ و منه، المتوسط التوافقي للسرعات

$$H = \frac{4}{\frac{1}{10}+\frac{1}{30}+\frac{1}{40}+\frac{1}{20}} = 19.2 \text{ كم\textbackslash الساعة}.$$

وبالتالي فإن المتوسط التوافقي هو الطريقة المناسبة لحساب متوسط السرعة.

▼

4.2.3. المتوسط الهندسي Moyenne géométrique

ليكن X السلسلة الإحصائية الكمية المنفصلة ذات الحجم (Taille) $n \in \mathbb{N}^*$، حيث $X = \{X_i\}_{1\leq i\leq n}$
من أجل كل عدد صحيح i، $1 \leq i \leq n$ لدينا $X_i > 0$.

تعريف 4.2.3 : المتوسط الهندسي للمتغير X الذي نرمز له بالرمز $\overline{X}_G$ معرف كما يلي:

$$\overline{X}_G = \left(\prod_{i=1}^{n} X_i\right)^{1/n} = (X_1X_2...X_n)^{1/n}.$$

في حالة أن كل شكل X_i يظهر بتكرار n_i فإن المتوسط الهندسي $\overline{X}$ يكتب على الشكل التالي:

$$\overline{X}_G = \left(\prod_{i=1}^{n} X_i^{n_i}\right)^{1/n} = (X_1^{n_1}X_2^{n_2}...X_n^{n_n})^{1/n},$$

حيث $n = n_1 + n_2 + ... + n_n$.

مثال 2 : لتكن السلسلة الإحصائية المعطاة في الجدول التكراري التالي

الشكل	1	2	3
التكرارات	1	2	4

المتوسط الهندسي لهذه السلسلة معطى بالشكل التالي:

$$\overline{X}_G = \left(1^12^23^4\right)^{\frac{1}{1+2+4}} = 2.2837$$

5.2.3. الوسيط Médiane

تعريف 5.2.3 : من الناحية النظرية ، فإن الوسيط هو القيمة x_i التي تبلغ فيها قيمة دالة التوزيع (أو الرسم التخطيطي المتكامل لها) 0.5 أو (50%).

لكن تطبيقيا، الوسيط فريد من نوعه. حيث يجعل من الممكن تقسيم مجموعة القيم إلى جزأين متساويين في المساحة. ولكن نادرا ما نقع على القيمة 0.5 أو (50%) عند حساب التواترات التراكمية المتزايدة.

الحساب التطبيقي للوسيط

في حالة سلسلة إحصائية منفصلة، نضع في الاعتبار جميع البيانات (n هو العدد الإجمالي) ، حتى لو كان هناك تكرار (n_i للقيمة x_i) ، نحصل على

$$x_1, x_2, ..., x_n.$$

بعد فرز هذه القيم بترتيب تصاعدي ، نحصل على سلسلة مرتبة

$$X_{(1)}, X_{(2)}, ..., X_{(n)}.$$

في هذه الحالة ينتج لدينا حالتين

★ لما n فردي في هذه الحالة قيمة الوسيط هي

$$\text{الوسيط} = X_{\left(\frac{n+1}{2}\right)},$$

★ لما n زوجي في هذه الحالة قيمة الوسيط هي

$$\text{الوسيط} = \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2}.$$

مثال 3 : لتكن السلسلة التالية

$$15,\ 8,\ 12,\ 11,\ 5,\ 7,\ 9.$$

ترتيب السلسلة تصاعديا يعطينا:

$$5,\ 7,\ 8,\ {\color{red}9},\ 11,\ 12,\ 15.$$

بما أن قيمة الحجم n فردية فإن القيمة التي تعبر عن قيمة الوسيط هي القيمة الرابعة في السلسلة المرتبة تظهر باللون الأحمر أي الوسيط يساوي لـ 9.

مثال 4 : لتكن السلسلة التالية

$$15,\ 8,\ 12,\ 11,\ 5,\ 7,\ 9,\ 17.$$

ترتيب السلسلة تصاعديا يعطينا:

$$5,\ 7,\ 8,\ 9,\ 11,\ 12,\ 15,\ 17.$$

بما أن قيمة الحجم $n = 8$ زوجية فإن

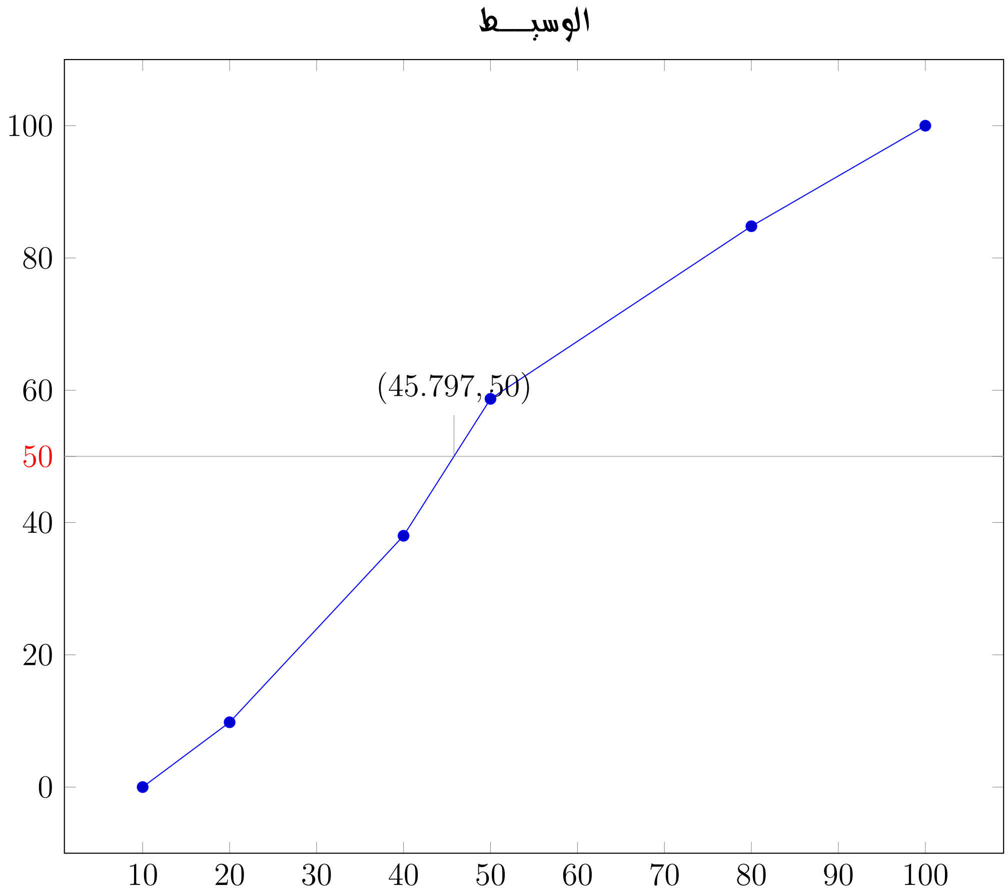
$$\text{الوسيط} = (9 + 11)/2 = 10.$$

في حالة سلسلة حقيقية أو مستمرة: فليس من الممكن دائمًا معرفة قيمة الوسيط تمامًا ولكن يمكن تحديدها تقريبيا عن طريق الاستيفاء الخطي Interpolation linéaire.

★ **بيانياً: نعلم ، بحكم تعريفه ، أن الوسيط يحتوي على تواتر تراكمي متزايد قدره** 0.5**، حتى نتمكن من إنشاء الرسم التخطيطي المتكامل أو دالة التوزيع، يكفي أن نرسم خط أفقي ترتيبته** 0.5 **يتقاطع هذا الخط مع منحنى دالة التوزيع للتواترات المتراكمة في نقطة فاصلتها تكوّن الوسيط.**

مثال 5 : لتكن السلسلة التالية

C_i	n_i	$f_i\,(\%)$	$F_i\,(\%)$
$[10;20]$	9	9.8	9.8
$]20;40]$	26	28.3	38.0
$]40;50]$	19	20.7	58.7
$]50;80]$	24	26.1	84.8
$]80;100]$	14	15.2	100
المجموع	92	100	



★ **عدديا: لتكن** $C_i =]a_i, a_{i+1}]$ **أول فئة التي يكون تواترها التراكمي** F_i **أكبر من أو يساوي القيمة** 0.5.

إذا كان $F_i = 0.5$ **فإن قيمة الوسيط هنا بديهية وتساوي طرف الفئة** a_{i+1} **وهذا نادرا ما يحدث.**

في الحالة العكسية، اذا كانت $F_i > 0.5$ **حينها نضع النقطة** $A : (a_i, F_{i-1})$ **و النقطة** $B : (a_{i+1,} F_i)$ **حيث هي قيمة التواتر المتراكم للفئة السابقة** C_i **في حالة** $i > 1$ **و إلا نعتبرها صفر في حالة** $i = 1$.

المستقيم D **المحدد بهاتين النقطتين يمر على النقطة ذات الترتيبة** 0.5 **والتي فاصلتها تشكل قيمة الوسيط، معادلة هذا المستقيم هي من الشكل:**

$$D : x - a_i = (y - F_{i-1})\frac{(a_{i+1} - a_i)}{(F_i - F_{i-1})}.$$

قيمة الوسيط (Med) **تقابل قيمة** x **لما** y **يساوي القيمة** 0.5 **يعني**

$$Med = a_i + (a_{i+1} - a_i)\frac{(0.5 - F_{i-1})}{(F_i - F_{i-1})}.$$

أو نستعمل نرية طاليس كي نجد

$$\frac{Med - a_i}{a_{i+1} - a_i} = \frac{0.5 - F_{i-1}}{F_i - F_{i-1}}.$$

مثال 6 : في المثال السابق *5*: لدينا $i = 3$, $C_i =]40; 50]$, $F_i = 0.587$, $F_{i-1} = 0.38$ ومنه قيمة الوسيط هي

$$Med = 40 + (50 - 40)\frac{(0.5 - 0.38)}{0.587 - 0.38} = 45.7971.$$

6.2.3. الربيعيات Quartiles

تعريف 6.2.3 : نسمي *quantiles* الـ $(k-1)$ قيمة التي تقسم السلسلة الإحصائية المرتبة الى k فئة من نفس الحجم.

يمكن تحديدها بيانيا باستخدام دالة التوزيع أو الرسم البياني المتكامل.

ملاحظة 2 : .

1. من أجل $k = 2$: نجد قيمة الوسيط الإحصائي *la médiane*
2. من أجل $k = 4$: نجد قيمة الربيعيات الثلاثة *les trois quartiles* الموافقة للتواترات المتراكمة 0.25 من أجل الربع الأول، 0.5 من أجل الوسيط الذي يمثل أيضا الرُبيع الثاني و 0.75 من أجل الربيع الثالث.

في حالة وجود سلسلة إحصائية منفصلة ، نرتب السلسلة تصاعديا فنميز أربع حالات:

1. $\Leftarrow n = 4p$:

أـ الرُبيع Quartile الأول يوافق

$$Q_1 = \frac{X_{(p)} + X_{(p+1)}}{2}.$$

ب- الرُبيع الثاني أو الوسيط

$$Q_2 = \frac{X_{(2p)} + X_{(2p+1)}}{2}.$$

ج- و الرُبيع الثالث

$$Q_3 = \frac{X_{(3p)} + X_{(3p+1)}}{2}.$$

2. $\Leftarrow$ $n = 4p + 1$:

أ- الرُبيع الأول يوافق

$$Q_1 = \frac{X_{(p)} + X_{(p+1)}}{2}.$$

ب- الرُبيع الثاني أو الوسيط

$$Q_2 = X_{(2p+1)}.$$

ج- و الرُبيع الثالث

$$Q_3 = \frac{X_{(3p)} + X_{(3p+1)}}{2}.$$

3. $\Leftarrow$ $n = 4p + 2$:

أ- الرُبيع الأول يوافق

$$Q_1 = X_{(p+1)}.$$

ب- الرُبيع الثاني أو الوسيط

$$Q_2 = \frac{X_{(2p)} + X_{(2p+1)}}{2}.$$

ج- و الرُبيع الثالث

$$Q_3 = X_{(3p+1)}.$$

4. $\Leftarrow$ $n = 4p + 3$:

أ- الرُبيع الأول يوافق

$$Q_1 = X_{(p+1)}.$$

ب- الرُبيع الثاني أو الوسيط

$$Q_2 = X_{(2p+2)}.$$

ج- و الرُبيع الثالث

$$Q_3 = X_{(3p+3)}.$$

في حالة سلسلة حقيقية أو مستمرة، تقسم الأرباع الثلاثة الرسم البياني للمدرج التكراري للسلسلة الإحصائية إلى أربعة أجزاء من نفس المساحة. نتبع مرة أخرى الطريقة المستخدمة لحساب الوسيط للوسيط بأخذ التواترات المتراكمة 0.25 من أجل الربيع الأول، 0.5 من أجل الوسيط الذي يمثل أيضا الربيع الثاني و 0.75 من أجل الربيع الثالث.

مثال 7 : لحساب الربيع الأول من سلسلة المثال السابق *5*: نتبع العلاقة التالية:

$$D : x - 20 = (y - 0.098)\,\frac{(40-20)}{(0.38-0.098)}.$$

قيمة الربيع الأول تقابل قيمة x لما y يساوي القيمة 0.25 يعني

$$Q_1 = 20 + (40-20)\,\frac{(0.25-0.098)}{(0.38-0.098)} = 30.78.$$

قيمة الربيع الثالث تقابل قيمة x لما y يساوي القيمة 0.75 يعني

$$Q_3 = 50 + (80-50)\,\frac{(0.75-0.587)}{(0.848-0.587)} = 68.735.$$

7.2.3. العُشير Déciles

العشريات هي $k = 9$ هي تسعة أعداد حقيقية تقسم السلسلة الإحصائية المطلوبة إلى 10 مجموعات فرعية متساوية الحجم أي تحتوي كل منها على عُشر (10%) من المعطيات أو البيانات. التسع عشور هي:

$$(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)\,.$$

مثال 8 : العشير الثاني للسلسلة الإحصائية للمثال السابق *5*: هي

$$d_2 = 20 + (40-20)\,\frac{(0.2-0.098)}{(0.38-0.098)} = 27.23.$$

8.2.3. المنوال Mode

يعتبر المنوال من أسهل مقاييس النزعة المركزية التي يمكن الحصول عليها بدون أجراء عمليات حسابية معقدة سواء كانت البيانات مبوبة او غير مبوبة او كانت بشكل توزيعات تكرارية.

تعريف 7.2.3 : المنوال *Le Mode* في الإحصاء هو القيمة الأكثر تكراراً في مجموعة من البيانات، أو في فضاء إحتمالي.

يعتمد حساب المنوال وفقاً لنوع السلسلة الإحصائية المدروسة مستمرة كانت أو منفصلة.

حساب المنوال في حالة متغير كمي منفصل

في حالة سلسلة إحصائية منفصلة أو متقطعة، يكون المنوال هو قيمة المتغير ذي أكبر عدد تكراري n_i. أو أعلى تردد أو تواتر f_i. في هذه الحالة يمكن ملاحظتها مباشرة. في الجدول التكراري أو جدول التواتر الإحصائي ، بأخذ قيمة x_i الموافقة.

مثال 9 : في المثال *5*: ، المنوال هو 2 طفل لكل عائلة

عدد الأطفال	0	1	2	3	4	5	6	7
التكرارات	7	23	40	13	11	3	2	1

ملاحظة 3 : يمكن أن يكون هناك أكثر من منوال واحد، إذا كانت هناك قيمتان مثلا لهما نفس العدد التكراري الأكبر.

حساب المنوال في حالة متغير كمي مستمر

الفئة المنوالية la classe modale هي الفئة ذات أكبر تكرار و يمكن تعريف المنوال في حالة سلسلة مستمرة أنه مركز الفئة المنوالية.

مثال 10 : في المثال السابق *1*: ، الفئة المنوالية هي $[20-19[$ سنة.

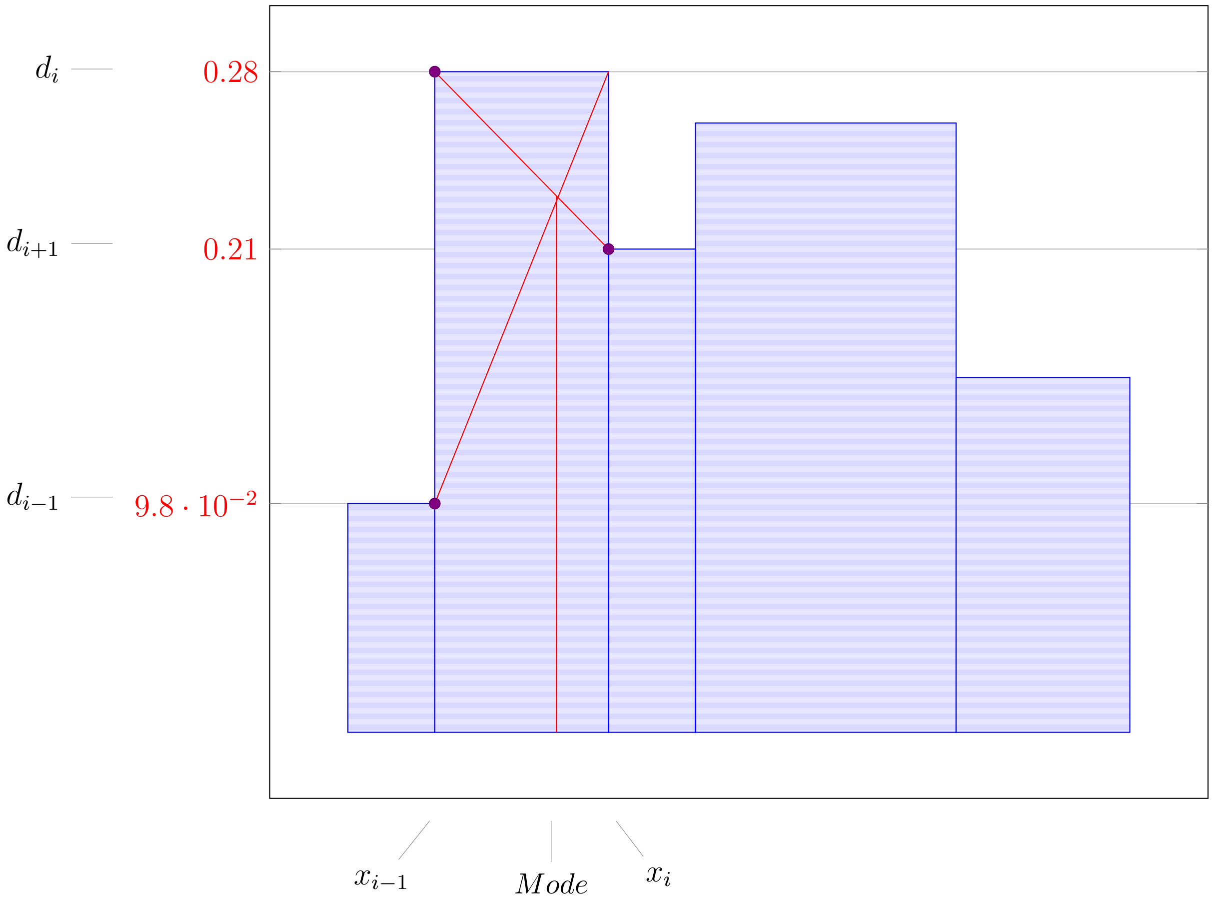
العمر العمر	التكرارات
$[18-19]$	1
$]19-20]$	10
$]20-21]$	3
$]21-22]$	2
$]22-23]$	1
المجموع	17

اذا كانت الفئات $C_i =]a_i, a_{i+1}]$ ذات التكرارات n_i في هذه الحالة نميز فرضيتين

1. اذا كانت فئات السلسلة C_i متساوية في سعتها اي طول الفئة، في هذه الحالة المنوال هو مركز الفئة المنوالية.

٢. إذا كان أطوال الفئات غير متساوية لا بد من تصحيح التكرارات المطلقة حتى تكون مساحة المستطيلات في المدرج التكراري تتناسب مع التكرار الموافق لها. و بما أن المنوال يعتمد في حسابه على التكرارات المطلقة فلابد من تصحيح التكرارات المطلقة قبل البدء في حسابه .

مثال 11 : بالنسبة لسلسلة المثال السابق 5: قيمة المنوال هي



هذا يعني أن المنوال $Mode$ محصور بين x_{i-1} و x_i مثلما أن d_i بين d_{i-1} و d_{i+1}.

جبريا، نطبق نظرية طاليس le théorème de Thalès نجد:

$$\frac{Mode - x_{i-1}}{x_i - x_{i-1}} = \frac{d_i - d_{i-1}}{(d_i - d_{i-1}) + (d_i - d_{i+1})}.$$

نتحصل على النتائج التالية

$$Mode = x_{i-1} + (x_i - x_{i-1})\frac{d_i - d_{i-1}}{(d_i - d_{i-1}) + (d_i - d_{i+1})}.$$

مثال 12 : كتطبيق عددي نأخذ سلسلة المثال السابق 5:

C_i	n_i	$f_i\,(\%)$	$F_i\,(\%)$	
$[10;20]$	9	9.8	9.8	
$]20;40]$	26	28.3	38.0	$\leftarrow i=2$
$]40;50]$	19	20.7	58.7	
$]50;80]$	24	26.1	84.8	
$]80;100]$	14	15.2	100	
المجموع	92	100		

لدينا $i=2$، $x_1=20$، $x_2=40$، $d_1=0.098$، $d_2=0.283$ و $d_3=0.207$ و منه

$$\begin{aligned} Mod &= x_1+(x_2-x_1)\frac{d_2-d_1}{(d_2-d_1)+(d_2-d_3)} \\ &= 20+(40-20)\frac{0.283-0.098}{(0.283-0.098)+(0.283-0.207)} \\ &= 34.176. \end{aligned}$$

ملاحظة 4 : من خواص المنوال أنه غير ثابت، يتأثر بطول الفئة، يفضل عندما يكون المقياس اسمي، ولا يعتمد عليه في حالة الإحصاءات اللاحقة.

9.2.3. المركز الحسابي Le milieu

تعريف 8.2.3 : المركز الحسابي *Le milieu* هو مركز المجال الحقيقي المحدد بالقيم القصوى للسلسلة ويحسب بالمعادلة التالية:

$$Milieu=\frac{\min\limits_{1\le j\le n}(x_j)+\max\limits_{1\le j\le n}(x_j)}{2}.$$

مثال 13 : في حالة السلسلة المستمرة للمثال السابق *1*: فإن المركز الحسابي لهذه السلسلة يعطينا:

$$(23+17)/2=20$$

3.3 مقاييس التشتت

تعرف مقاييس التشتت على أنها مجموعة من الدوال الإحصائية التي تستخدم في تحديد مقدار انحراف البيانات الإحصائية عن بعضها البعض، أو عن قيمتها الوسطية والتي تسمى

بالوسط الحسابي للقيم، وتعد هذه المقاييس هامة في عملية صنع القرار، لأنها تعطي معلومات دقيقة عن مدى تجانس العينات الإحصائية، وتربط بين ما هو موجود، وبين ما كان متوقع الحدوث، كما تسهم هذه المقاييس الإحصائية في المقارنة بين عدة مجموعات من البيانات الإحصائية وفق النتائج التي تصدر عنها.

ليكن المثال التالي الذي نوضح به أهمية مقاييس التشتت، لدينا تقييم أستاذين (أ) و (ب) لـ 6 طلبة فهل نستطيع الإستنتاج من الوسيط والمتوسط الحسابي أو المنوال؟

رقم الطالب	نقطة الاستاذ a	نقطة الاستاذ b
1	7	0
2	11	20
3	9	9
4	13	10
5	10	10
6	10	11
المنوال	10	10
الوسيط	10	10
المتسط الحسابي	10	10

نلاحظ جيدا تساوي الوسيط والمتوسط الحسابي و المنوال في القيمة 10 يعني أنهما ينقطان بنفس الطريقة لكن ضمنيا ليس هو الحال فهم يختلفون تماما في طريقة التنقيط

لذلك من المفيد مقارنة القيم بمقاييس أخرى تعطي ترتيبا لحجم اختلاف القيم بينها، هذه المقاييس تجعل من الممكن التعبير عن تشتت البيانات حول الوسط الحسابي.

1.3.3. مقاييس التشتت المطلقة

إن مقاييس النزعة المركزية لا تكفي لوحدها لوصف البيانات وإجراء المقارنات بين التوزيعات التكرارية، لأنها لا تعطينا فكرة عن مدى تجانس أو عدم تجانس البيانات، فعند إجراء مقارنة بين ظاهرتين يمكن أن يتساوى متوسطهما الحسابي، ورغم ذلك نجد أن انتشار البيانات في الظاهرتين مختلف كثيرا لأن البيانات غير متجانسة، لهذا وجدت مقاييس أخرى تعطينا فكرة عن مدى تباعد البيانات عن بعضها البعض، تسمى هذه المقاييس بمقاييس التشتت.

المدى L'étendue

يعتبر المدى من أسهل مقاييس التشتت تعريفا وحسابا ويعطينا فكرة سريعة عن مدى تفرق البيانات.

تعريف 1.3.3 : في الإحصاء، يطلق اسم المدى *L'étendue* على **طول أصغر مجال** يضم جميع عناصر البيانات. ويتم حسابه بطرح العينة الصغرى من العينة الكبرى.

$$e = \max_{1 \leq i \leq n} (x_i) - \min_{1 \leq i \leq n} (x_i)$$

بما أن المدى يعتمد فقط على قيمتين من كامل العينة الإحصائية فإنه لايقدم معلومات كافية عن مقدار تشتت العينة إلا إذا كان حجم العينة صغيراً.

مثال 1 : نأخذ المثال السابق *1*: نجد

$$e = 23 - 17 = 6.$$

ملاحظة 1 : (بعض مميزات وعيوب المدى) .
- سهل التعريف والحساب.
- يتأثر بالقيم الشاذة أو المتطرفة.
- لا يأخذ في اعتباره كل البيانات.

المدى الربيعي L'écart inter-quartile

تعريف 2.3.3 : المدى الربيعي *L'écart inter-quartile* **هو الفرق بين الربيع الثالث والربيع الأول.**

$$IQ = Q_3 - Q_1.$$

مثال 2 : نأخذ المثال السابق *5*: نجد قيمة المدى الربيعي هي

$$IQ = 3 - 1 = 2.$$

أما بالنسبة للمثال *1*: نجد

$$IQ = 68.735 - 30.78 = 37.955.$$

الانحراف المتوسط بالنسبة للمتوسط الحسابي

تعريف 3.3.3 : الانحراف المتوسط بالنسبة للمتوسط الحسابي أو *L'écart absolu moyen* هو البعد المتوسط لقيم المتغير الإحصائي عن المتوسط الحسابي وهو الأكثر استعمالا.

من عيوبه أنه لا يفرق بين القيم التي تكون أكبر أو أقل من المتوسط، و يحسب بالطريقة التالية:

★ من أجل سلسلة منفصلة

$$E = \frac{1}{n}\sum_{i=1}^{k} n_i \left|x_i - m\right|$$

حيث m هو المتوسط الحسابي للسلسلة.

★ من أجل سلسلة متصلة أو حقيقية

$$E = \frac{1}{n}\sum_{i=1}^{k} n_i \left|\left(\frac{a_i + a_{i+1}}{2}\right) - m\right|.$$

مثال 3 : حساب الانحراف المتوسط بالنسبة للمتوسط الحسابي للسلسلة التالية

x_i	3	5	6	8	10
n_i	1	3	4	1	1

حيث $\overline{X}$ هو المتوسط الحسابي

$$\overline{X} = \frac{13 + 35 + 46 + 18 + 110}{10} = 6,$$

الإنحراف المتوسط هو

$$E = \frac{13 + 31 + 40 + 12 + 14}{10} = 1.2.$$

في المتوسط ، تختلف القيم المرصودة ، أكثر أو أقل ، بمقدار 1.2 عن متوسط قيمة السلسلة البالغ 6.

الإنحراف المتوسط عن الوسيط

تعريف 4.3.3 : الانحراف المتوسط بالنسبة للوسيط أو *L'écart absolu médian* هو البعد المتوسط لقيم المتغير الإحصائي عن الوسيط.

من عيوبه أنه لا يفرق بين القيم التي تكون أكبر أو أقل من الوسيط، و يحسب بالطريقة التالية:

★ من أجل سلسلة منفصلة

$$E = \frac{1}{n} \sum_{i=1}^{k} n_i \left| x_i - Mediane \right|$$

★ من أجل سلسلة متصلة أو حقيقية

$$E = \frac{1}{n} \sum_{i=1}^{k} n_i \left| \left(\frac{a_i + a_{i+1}}{2} \right) - Mediane \right|$$

مثال 4 : من المثال السابق *5*: نجد قيمة الإنحراف المتوسط عن الوسيط

$$E = 1.52 \ \text{طفل لكل عائلة}$$

أما بالنسبة للمثال *1*: فنجد

$$E = 0.827 \ \text{سنة}.$$

مقياس التباين Variance

يتميز التباين Variance بأخذ عينات من مجتمع الدراسة من أجل إطلاق الحكم وإعطاء معلومات إحصائية معينة، ويعتمد هذا النوع على الوسط الحسابي في قوانينه الرياضية، ويمكن أن يكون لبيانات إحصائية مبوبة، أو لبيانات إحصائية غير مبوبة.

تعريف 5.3.3 : هو مقياس إختلاف البيانات وتشتتها، وهو متوسط مربعات انحرافات القيم عن وسطها الحسابي، ويرمز له بالرمز S^2 بالنسبة للعينات و σ^2 بالنسبة للمجتمع الإحصائي ويحسب من الصيغة الرياضية الآتية:

★ في حالة سلسلة بسيطة منفصلة

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{k} n_i \left(x_i - \mu \right)^2$$

حيث μ هو المتوسط الحسابي للمجتمع، أما تباين العينة S^2 فيحسب

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i \left(x_i - \overline{X}\right)^2$$

هنا $\overline{X}$ يمثل المتوسط الحسابي للعينة

★ في حالة سلسلة مستمرة ممثلة بفئات من الشكل $C_i =]a_i, a_{i+1}[$ ، فإننا نعوض في المعادلة السابقة قيمة x_i بمركز الفئة $c_i = (a_i + a_{i+1})/2$ نتحصل على:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{k} n_i c_i^2 - \frac{2\mu}{n}\sum_{i=1}^{k} n_i c_i + \frac{\mu^2}{n}\sum_{i=1}^{k} n_i.$$

يمكن كتابة المعادلة السابقة كما يلي:

$$\begin{aligned}\sigma^2 &= \frac{1}{n}\sum_{i=1}^{k} n_i c_i^2 - 2\mu^2 + \mu^2 \\ &= \frac{1}{n}\sum_{i=1}^{k} n_i c_i^2 - \mu^2.\end{aligned}$$

وهي العبارة الأكثر استعمالا لحساب التباين. كما يمكن حساب تباين العينة S^2 في هذه الحالة بالصيغة:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i \left(c_i - \overline{X}\right)^2$$

حيث

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} n_i \left(\frac{a_i + a_{i+1}}{2}\right) = \frac{1}{n}\sum_{i=1}^{k} n_i c_i.$$

مثال 5 : من أجل سلسلة الأمثلة السابقة على الترتيب *5*: و *1*: نجد

$$\begin{aligned}\sigma^2 &= \frac{1}{100}(70^2 + 231^2 + 402^2 + 133^2 + 114^2 + 35^2 + 26^2 + 17^2) - 2.2^2 \\ &= 1.88\end{aligned}$$

في حين

$$\begin{aligned}S^2 &= \frac{1}{100-1}(70^2 + 231^2 + 402^2 + 133^2 + 114^2 + 35^2 + 26^2 + 17^2) - 2.2^2 \\ &= 1.898\end{aligned}$$

و

$$\begin{aligned}\sigma^2 &= \frac{1}{17}(118.5^2+1019.5^2+320.5^2+221.5^2+122.5^2)-20.0294^2\\ &= 0.955.\end{aligned}$$

$$\begin{aligned}S^2 &= \frac{1}{17-1}(118.5^2+1019.5^2+320.5^2+221.5^2+122.5^2)-20.0294^2\\ &= 1.0146\end{aligned}$$

الإنحراف المعياري

وهو من أدق هذه المقاييس، وأكثرها استخداما، كما أنه سهل الاحتساب، ويمثل الجذر التربيعي الموجب للتباين، ويعرف على أنه الجذر التربيعي لمتوسط مجموع مربعات انحرافات قيم المتغير العشوائي عن وسطها الحسابي، ويتم حسابه رياضياً عن طريق قانون خاص، ويتميز بأنه موجب القيمة دائماً.

تعريف 6.3.3 : الانحراف المعياري *L'écart-type* هو القيمة الأكثر استخداما من بين مقاييس التشتت الإحصائي لقياس مدى التبعثر الإحصائي، أي أنه يدل على مدى امتداد مجالات القيم ضمن مجموعة البيانات الإحصائية. عادة ما يرمز له بالرمز σ ورياضيا هو الجذر التربيعي للتباين:

$$\sigma=\sqrt{\sigma^2}$$

بالنسبة للإنحراف المعياري للمجتمع. و بالنسبة للإنحراف المعياري للعينات الذي يرمز له بالرمز sd حيث

$$sd=\sqrt{S^2}.$$

2.3.3. معلمات التشتت النسبية

فأن التباين والانحراف المعياري من المقاييس المفيدة لقياس التشتت لتوزيع متغير ما. ولكن في كثير من الأحيان نكون مهتمين بمقارنة التشتت والاختلاف التوزيعي لمتغيرين 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 آنفا تكون غير مناسبة لمقارنة تجانس مجموعات البيانات المختلفة في الحالتين التاليتين:

1. إذا كانت وحدتا المتغيرين مختلفتين حيث لا نستطيع مقارنة الوحدات المختلفة.

2. إذا كان متوسطا المتغيرين مختلفين وذلك لأن تباين توزيع المتغير ذي المتوسط الصغير ينزع لأن يكون صغير؟ والعكس بالعكس. لذلك دعت الحاجة إلى مقاييس أخرى لا تعتمد على وحدة المتغير وتقيس بما يسمى بالتشتت النسبي. وأحد هذه المقاييس هو ما يسمى بمعامل الاختلاف أو معامل التغير.

معامل الاختلاف

تعريف 7.3.3 : معامل الاختلاف *Le coefficient de variation* في نظرية الاحتمالات والإحصاء، هو مقياس لتشتت أو تبعثر توزيع الاحتمال أو توزيع التكرار. يتم تعريف معامل الاختلاف كنسبة الانحراف المعياري إلى الوسط الحسابي للتوزيع و يرمز له بالرمز v .

$$v = \frac{\sigma}{m}.$$

مثال 6 : من أجل السلسلة السابقة في المثال *5*: نجد

$\sigma = 1.371$ طفل لكل عائلة $\quad v = 0.623$

بالنسبة للمثال *1*: نجد :

$\sigma = 0.977$ سنة $\quad v = 0.0487.$

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 تجانس؟ والعكس بالعكس

معامل عدم التماثل

هو المفهوم الأكثر أهمية في كل نظرية الاحتمالات. وهو معامل عدم التماثل والتلف coefficient d'asymétrie الذي يساهم في حساب وكشف المتغيرات العشوائية. يتم حساب هذه

القيم بواسطة الصيغ التالية.

$$\sigma_3 = \frac{1}{n}\sum_{i=1}^{k} n_i\left(x_i - m\right)^3$$

أو

$$\sigma_3 = \frac{1}{n}\sum_{i=1}^{k} n_i\left(c_i - m\right)^3,$$

أو حسب طبيعة السلسلة التي متوسطها الحسابي m و معيارها الإنحرافي σ

$$\gamma_3 = \frac{s_3}{\sigma^3}$$

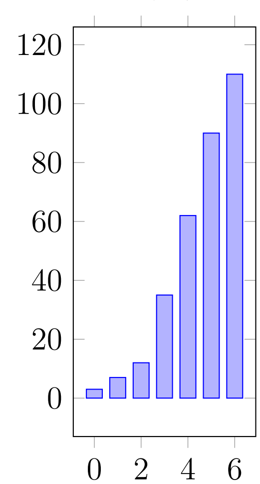
من أجل معرفة ماهية معامل عدم التماثل ، يجب أن يكون لديك المعلومات التالية: حجم اللحظة المركزية وحجم الانحراف المعياري. بالإضافة إلى ذلك ، من الضروري أن تكون σ_3، التي يُشار إليها بالمعامل نفسه ، أقل من اللانهاية. خلاف ذلك ، فإن جميع الحسابات لا معنى لها. بواسطة σ_3 ليس المقصود عدد محدود. من أجل العثور على الحل الأمثل ومعرفة ما هو معامل عدم التماثل بطريقة أو بأخرى ، سوف تحتاج إلى استخدام عدد قليل من الصيغ. من المرغوب فيه أن تكون النتيجة التي تحصل عليها قريبة من الصفر.

★ اذا كان $\gamma_3 < -0.5$ فإن السلسلة dissymétrique غير متماثلة نحو اليمين أنظر الرسم 1.

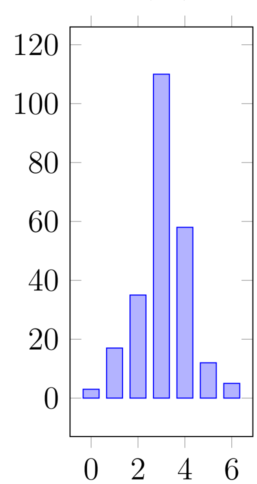
★ اذا كان $\gamma_3 \in [-0.5, 0.5]$ فإن السلسلة متناظرة، أنظر الرسم 2.

★ اذا كان $\gamma_3 > -0.5$ فإن السلسلة غير متماثلة نحو اليسار، أنظر الرسم 3.

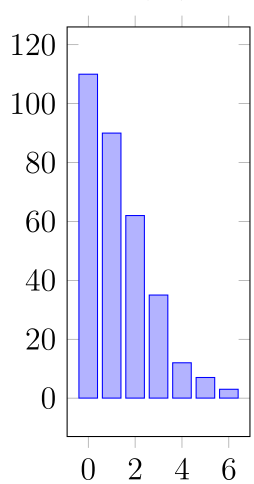
الرسم 1



الرسم 2



الرسم 3



تمارين مفتوحة

4.3 سلسلة التمارين رقم 3

تمرين 1 : سحبت عينة من 30 مزرعة للتعرف على مردوديتها من القمح (بالطن) خلال موسم ما، فكانت النتائج كاالتي

30	14	20	20	17	25	20	14	12	16	17	16	12	15	20
12	20	15	14	25	20	17	15	20	14	15	12	16	14	20

- عين المجتمع الإحصائي، الوحدة الإحصائية وطبيعة المتغيرة
- إذا أخذنا عدد الفئات هو 6، أحسب قيمة كل من التكرارات، التواترات والتواترات بالنسبة المئوية(٪) و التواترات التراكمية المتزايدة والمتناقصة.

الحــل

المجتمع الإحصائي	الوحدة الإحصائية	الصفة	طبيعتها	المتغير
المزارع	المزرعة	مردودية القمح	كمية	متصل

- تحديد طول السلسلة:

المدى = أكبر مشاهدة ـ أصغر مشاهدة

$$30 - 12 = 18$$

- تحديد أطوال الفئات:

طول الفئة = المدى \ عدد الفئات

$$18/6 = 3$$

- وضع الجدول التكراري

الفئات	n_i	f_i	$f_i\%$	التكرارات التراكمية المتزايدة	التكرارات التراكمية المتناقصة
12 − 15	9	0.3	30	9	30
15 − 18	10	0.333	33.33	19	21
18 − 21	8	0.2667	26.67	27	11
21 − 24	0	0	0	27	3
24 − 27	2	0.0667	6.67	29	3
27 − 30	1	0.0333	3.33	30	1
المجموع	30	1	100	−	−−

تمرين 2 : البيانات التالية تمثل فئات الأجور (بالألف دينار) لـ 50 عامل مبينة على النحو التالي

فئات الأجور	40 − 60	60 − 80	80 − 100	100 − 120	120 − 140	المجموع
التكرارات	8	12	20	6	4	50

- ما هو عدد العمال الذين تقل أجورهم عن 80 ألف دينار.
- ما هو عدد العمال الذين تقل أجورهم عن 55 ألف دينار.
- ما هي نسبة العمال الذين يتقاضون أجرا يزيد عن 90 ألف دينار.
- ما هي نسبة العمال الذين يتقاضون أجرا بين 55 و 90 ألف دينار.
- ما هي نسبة العمال الذين يتقاضون أجرا يقل عن 90 ألف دينار.

الحـــل

1− عدد العمال الذين تقل أجورهم عن 80 ألف دينار: هم 20 عامل (مباشرة من الجدول)

2− عدد العمال الذين تقل أجورهم عن 55 ألف دينار:

- حساب طول الفئة المعنية :$[40-60[$

$$60-40=20$$

- حساب الفرق في الرائب:

$$40-55=15$$

- تطبيق القاعدة الثلاثية نجد:

$$\frac{15 \cdot 8}{20} = 6$$

وهو عدد العمال الذين تقل أجورهم عن 55 ألف دينار.