[c] Calculer le moment centré d'ordre 3 puis le coefficient de dérive. Que peut-on en conclure en ce qui concerne la symétrie de la distribution ?

[les solutions]

[2.2.1] [a] Tableau des fréquences

nombre n de bourgeons par tige	2	3	4	5	6	7	8	9	10	11	12
fréquences absolues	3	5	10	15	23	19	13	5	3	3	1
fréquences relatives cumulées	0,03	0,08	0,18	0,33	0,56	0,75	0,88	0,93	0,96	0,99	1

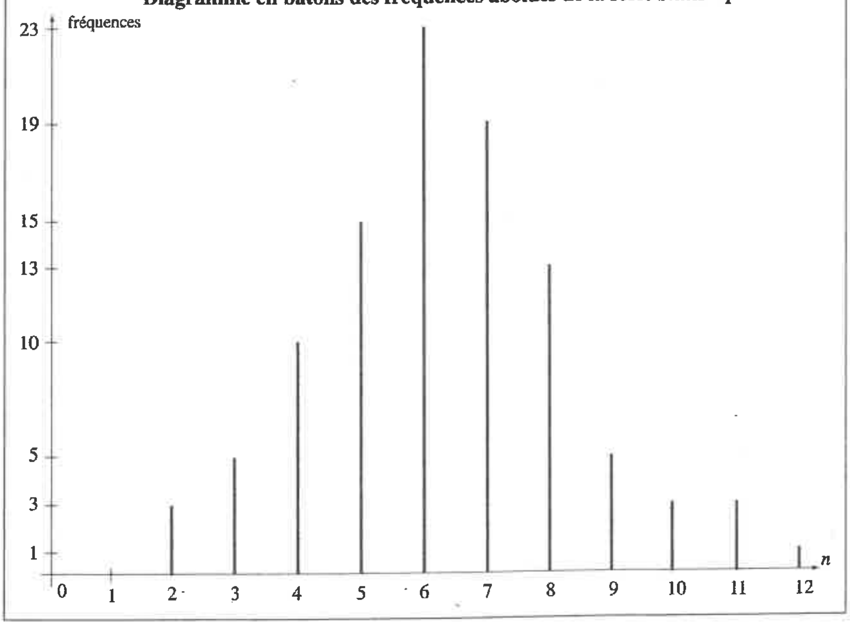


Diagramme en bâtons des fréquences abolues de la série statistique

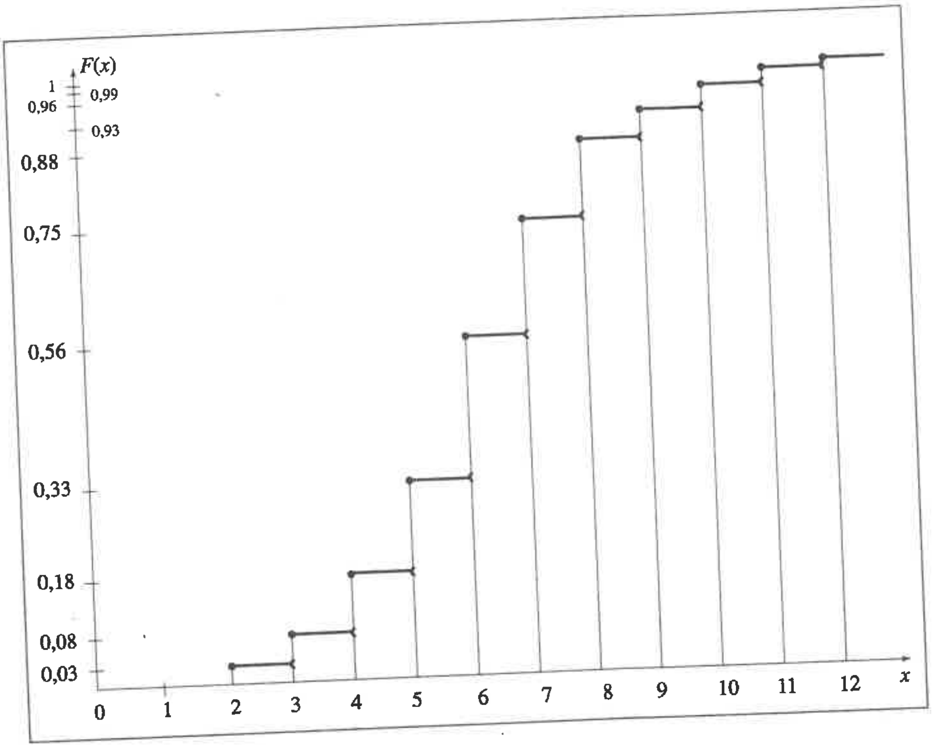
[b] Fonction de répartition

Soit $\{x'_j, \quad 1 \leq j \leq 11\}$ les valeurs de la série statistique ; ces valeurs sont :

$$2, \quad 3, \quad 4, \quad 5, \quad 6, \quad 7, \quad 8, \quad 9, \quad 10, \quad 11, \quad 12.$$

La fonction de répartition F est définie par (§ [2][b][iii])[†] :

$F(x) = 0$ si $x < 2$, $F(x) = 1$ si $x \geqslant 12$ et $F(x)$ est la fréquence relative cumulée des valeurs de la série statistique inférieure à x'_{j+1} si $x \in [x'_j, x'_{j+1}[$, $1 \leqslant j \leqslant 11$.



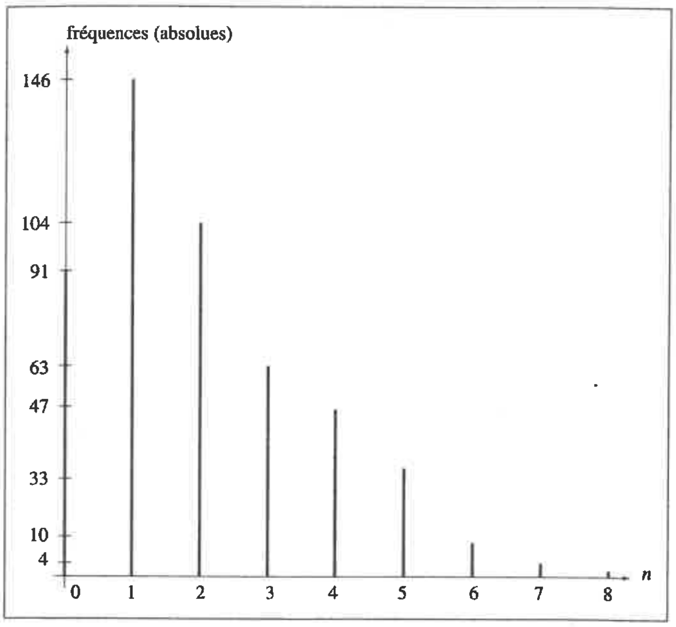
[2.2.2] [a] Tableau des fréquences

nombre n d'enfants par famille	0	1	2	3	4	5	6	7	8
fréquences absolues ou nombre de familles ayant n enfants	91	146	104	63	47	33	10	4	2
fréquences relatives cumulées	0,182	0,474	0,682	0,808	0,902	0,968	0,988	0,996	1

† Voir la note au bas de la page 15.

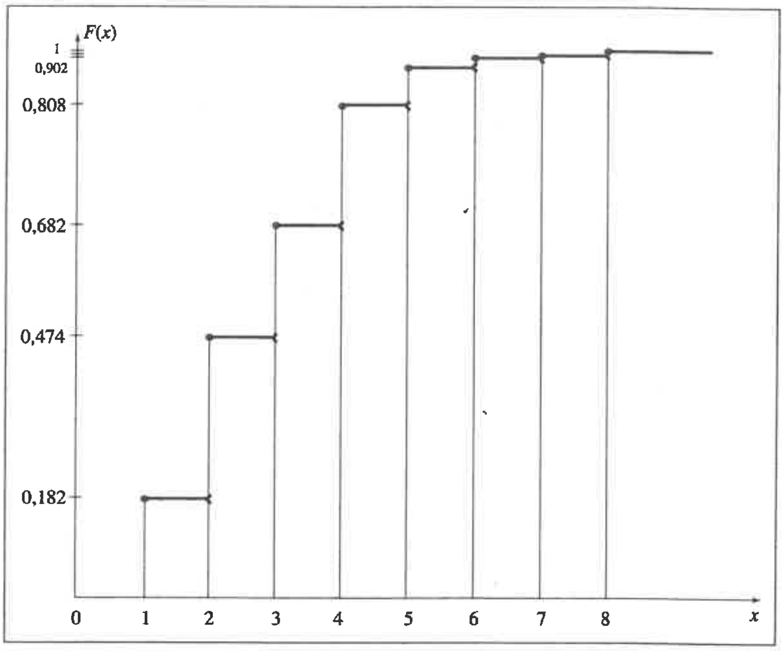
umulée des

7	8
4	2
996	1



[b] Fonction de répartition

La définition est celle données dans l'exercice **[2.2.1]** avec ici $p = 9$.



[c] Nombre moyen d'enfants par famille dans l'échantillon.

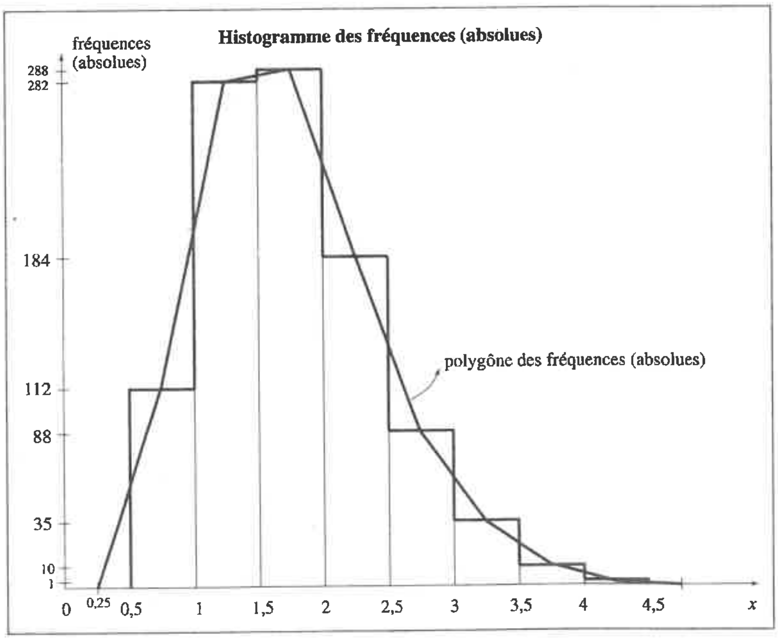
Soit m ce nombre ; alors on a :

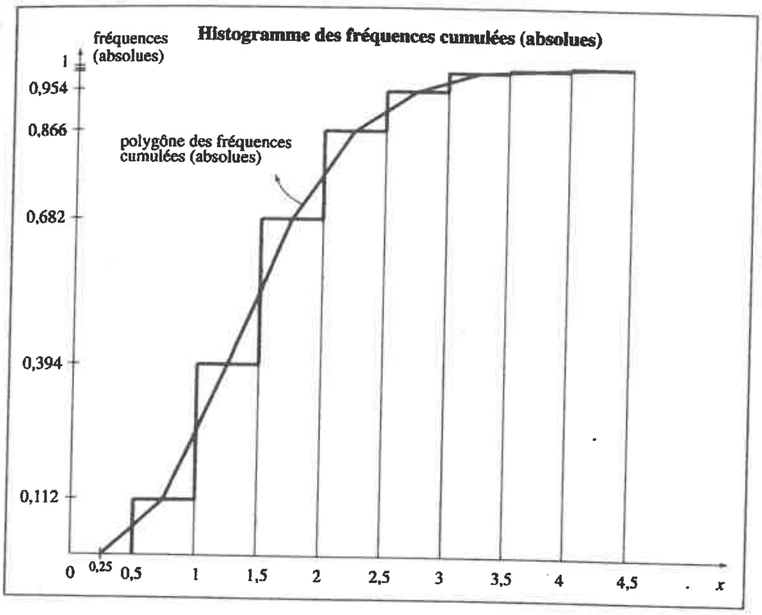
$$m = \frac{1}{500}(0 \times 91 + 1 \times 146 + 2 \times 104 + 3 \times 63 + 4 \times 47$$
$$+5 \times 33 + 6 \times 10 + 7 \times 4 + 8 \times 2) = \frac{1\,000}{500} = 2.$$

[2.2.3] [a] Centres des classes et fréquences cumulées

x	centres des classes	fréquences absolues	fréquences absolues cumulées	fréquences relatives cumulées
$0,5 \leqslant x < 1$	0,75	112	112	0,112
$1 \leqslant x < 1,5$	1,25	282	394	0,394
$1,5 \leqslant x < 2$	1,75	288	682	0,682
$2 \leqslant x < 2,5$	2,25	184	866	0,866
$2,5 \leqslant x < 3$	2,75	88	954	0,954
$3 \leqslant x < 3,5$	3,25	35	989	0,989
$3,5 \leqslant x < 4$	3,75	10	999	0,999
$4 \leqslant x < 4,5$	4,25	1	1 000	1

[b]



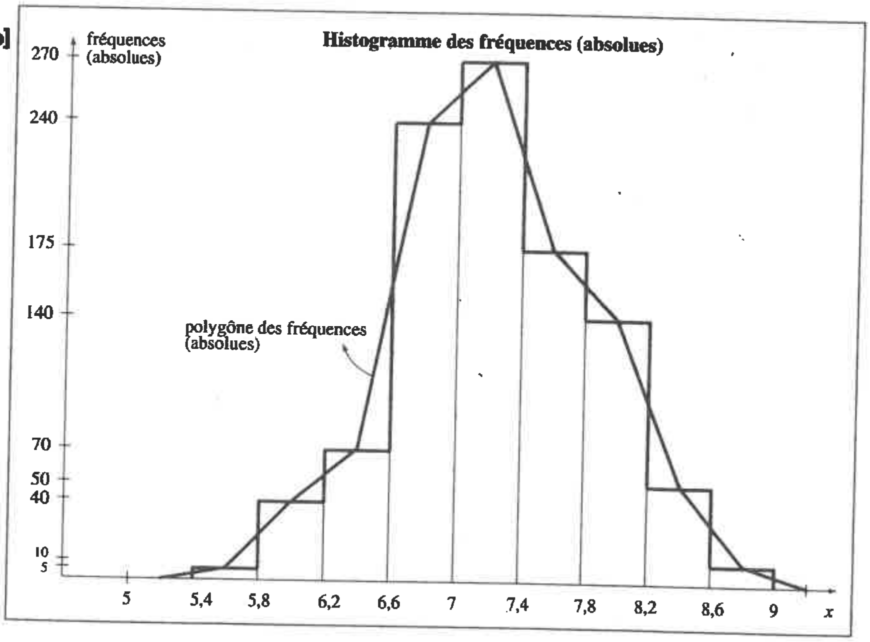


[c] L'histogramme des fréquences (absolues) nous conduit à supposer que X pourrait suivre une loi du χ^2.

[2.2.4] **[a]** L'étendue commune des classes est la distances des centres de deux classes consécutives, soit 0,4. Les extrémités des classes sont donc :

5,4 – 5,8 ; 5,8 – 6,2 ; 6,2 – 6,6 ; 6,6 – 7,0 ; 7,0 – 7,4 ; 7,4 – 7,8 ; 7,8 – 8,2 ; 8,2 – 8,6 ; 8,6 – 9,0.

[b]



[c] L'histogramme des fréquences (absolues) nous conduit à supposer que la variable aléatoire « diamètre des hématies » pourrait suivre une loi de Gauss.

[2.3.5]

$$\sum_{i=1}^{n}(x_i-\alpha)^2=\sum_{i=1}^{n}(x_i-m+m-\alpha)^2=\sum_{i=1}^{n}(x_i-m)^2+2(m-\alpha)\sum_{i=1}^{n}(x_i-m)+n(m-\alpha)^2.$$

Or :

$$\sum_{i=1}^{n}(x_i-m)=\sum_{i=1}^{n}x_i-nm=nm-nm=0.$$

Donc :

$$\sum_{i=1}^{n}(x_i-\alpha)^2=\sum_{i=1}^{n}(x_i-m)^2+n(m-\alpha)^2\geq\sum_{i=1}^{n}(x_i-m)^2,$$

et l'égalité a lieu si et seulement si $\alpha = m$ et la valeur minimum de cette somme est $\sum_{i=1}^{n}(x_i-m)^2=n\sigma^2$ où σ désigne l'écart type de la série statistique.

[2.3.6] Quitte à renuméroter les valeurs de la série statistique, on peut supposer que les valeurs x_i sont rangées par ordre de valeurs non décroissantes :

$$x_1\leq x_2\leq\cdots\leq x_{n-1}\leq x_n.$$

Rappelons que l'écart absolu de α à un nombre x est $|x-\alpha|$. Soit a et b tels que $a\leq b$; la somme des écarts absolus de α à a et b est $e=|a-\alpha|+|b-\alpha|$.

- Si $a\leq\alpha\leq b$, $e=(\alpha-a)+(b-\alpha)$ et ce résultat est indépendant de α.
- Si $\alpha\leq a\leq b$, $e=(a-\alpha)+(b-\alpha)=b-a+2(a-\alpha)\geq b-a$.
- Si $a\leq b\leq\alpha$, $e=(\alpha-a)+(\alpha-b)=b-a+2(\alpha-b)\geq b-a$.

Ainsi la somme des écarts absolus de α à a et b est minimum quand α est (quelconque) entre a et b.

Appliquons ce résultat à la série statistique.

- La somme des écarts absolus de α à x_1 et x_n est minimum lorsque α est quelconque entre x_1 et x_n.
- De même si α est aussi quelconque entre x_2 et x_{n-1}, la somme des écarts absolus de α à x_2 et x_{n-1} sera minimum.

Il en sera aussi de même pour x_3 et x_{n-2} et ainsi de suite.

On veut donc que la somme des écarts absolus de α aux valeurs x_i de la série statistique sera minimum lorsque le nombre de valeurs de la série statistique supérieures à α sera égal au nombre de valeurs de la série statistique inférieures à α ; c'est-à-dire lorsque α est la médiane.

[2.3.7] Les valeurs sont ordonnées par ordre de valeurs non décroissantes :

$$1 \quad 2 \quad 2 \quad 3 \quad 4 \quad 4 \quad 5 \quad 6 \quad 7 \quad 7 \quad 8 \quad 9 \quad 10 \quad 11 \quad 12.$$

Il y a $n = 15 = 2p + 1$ valeurs où $p = 7$.

[a] Mode

Il y a trois valeurs modales 2, 4 et 7 de fréquence absolue 2.

[b] Médiane

C'est la valeur de la suite située au rang $p + 1 = 8$, soit la valeur 6.

[c] Moyenne

$$\begin{aligned} m &= \frac{1}{15}(1+2+2+3+4+4+5+6+7+7+8+9+10+11+12) \\ &= \frac{91}{15} \simeq 6{,}07. \end{aligned}$$

[d] Écart type σ

Soit m_2 le moment d'ordre 2 :

$$\begin{aligned} m_2 &= \frac{1}{15}\big(1^2 + 2 \times 2^2 + 3^2 + 2 \times 4^2 + 5^2 + 6^2 + 2 \times 7^2 \\ &\qquad + 8^2 + 9^2 + 10^2 + 11^2 + 12^2\big) \\ &= \frac{719}{15}. \end{aligned}$$

On a alors (voir **[Théorème 2.3.3]**) :

$$\begin{aligned} \sigma^2 = \mu_2 = m_2 - m^2 &= \frac{719}{15} - \left(\frac{91}{15}\right)^2 = \frac{10\,785 - 8\,281}{225} \\ &= \frac{2\,504}{225} = 11{,}128\,8\ldots \\ \sigma &\simeq 3{,}34. \end{aligned}$$

[e] Premier quartile Q_1 et troisième quartile Q_3

Q_1 est la médiane de la suite de valeurs :

$$1 \quad 2 \quad 2 \quad 3 \quad 4 \quad 4 \quad 5,$$

donc $Q_1 = 3$.

Q_3 est la médiane de la suite de valeurs :

$$7 \quad 7 \quad 8 \quad 9 \quad 10 \quad 11 \quad 12,$$

donc $Q_3 = 9$.

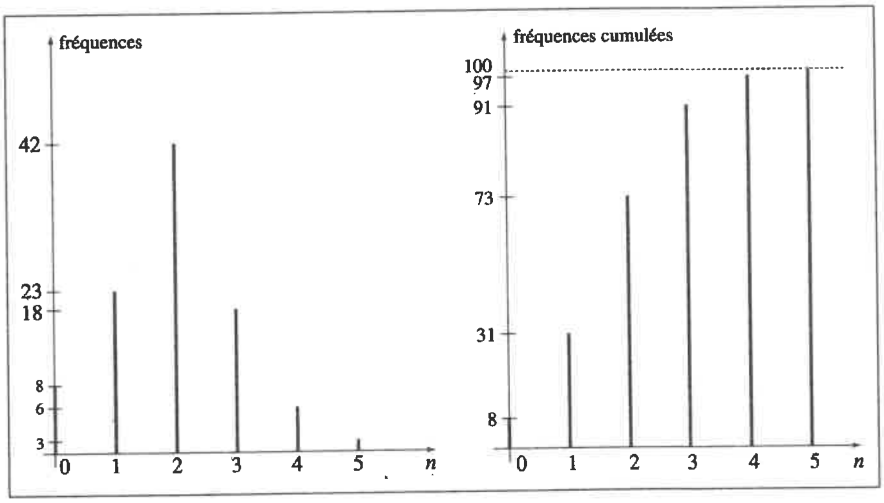
[f] Le semi-interquartile Q est (**[Définition 2.3.3]**) :

$$Q = \frac{1}{2}(Q_3 - Q_1) = \frac{1}{2}(9 - 3) = 3.$$

[2.3.8] [a] [i] Tableau de fréquences

nombre n d'enfants	0	1	2	3	4	5
fréquences absolues ou nombre de familles ayant n garçons	8	23	42	18	6	3
fréquences absolues cumulées	8	31	73	91	97	100

[ii] Diagrammes en bâtons des fréquences et des fréquences cumulées



[b] • Le mode est la valeur de plus grande fréquence ; c'est donc 2 (fréquence 42).

• Il y a deux valeurs en centre, ce sont les valeurs placées au rang 50 et au rang 51. Ces deux valeurs sont égales à 2 donc la médiane est 2.

$$\underbrace{\underbrace{\underbrace{0\,0\ldots0}_{8\text{ fois}}\ \underbrace{1\,1\ldots1}_{23\text{ fois}}}_{31\text{ valeurs}}\ \underbrace{2\,2\ldots2}_{42\text{ fois}}}_{73\text{ valeurs}}\ \underbrace{3\,3\ldots3}_{18\text{ fois}}\ \underbrace{4\,4\ldots4}_{6\text{ fois}}\ \underbrace{5\,5\ldots5}_{3\text{ fois}}$$

• La moyenne m est :

$$m = \frac{1}{100}(0\times 8 + 1\times 23 + 2\times 42 + 3\times 18 + 4\times 6 + 5\times 3) = \frac{200}{100} = 2.$$

[c] • Le premier quartile Q_1 est tel que 25 % des valeurs de la série statistique lui soit inférieur, donc $Q_1 = 1$.

• Le premier quartile Q_1 est tel que 75 % des valeurs de la série statistique lui soit inférieur, donc $Q_3 = 3$.

• Soit m_2 le moment d'ordre 2 :

$$m_2 = \frac{1}{100}\left(0^2 \times 8 + 1^2 \times 23 + 2^2 \times 42 + 3^2 \times 18 + 4^2 \times 6 + 5^2 \times 3\right) = \frac{524}{100} = 5{,}24.$$

La variance est donc :

$$v = m_2 - m^2 = 5{,}24 - 4 = 1{,}24.$$

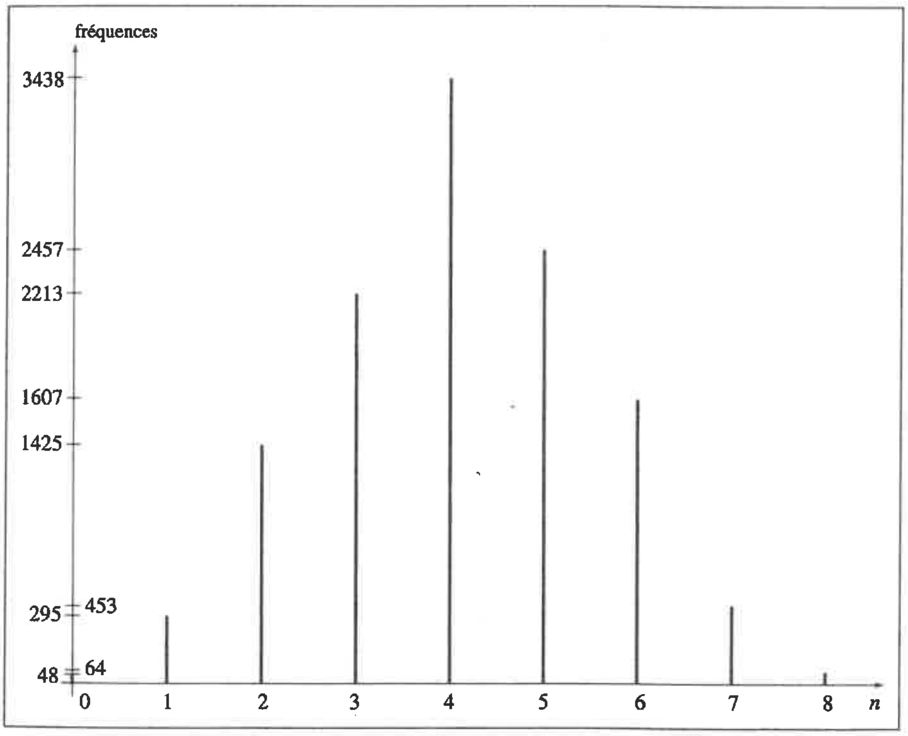
D'où l'écart type :

$$\sigma = \sqrt{1{,}24} \simeq 1{,}11.$$

[2.3.9] [a] [i] Tableau des fréquences

nombre n de garçons	0	1	2	3	4	5	6	7	8
fréquences absolues ou nombre de familles ayant n garçons	48	295	1 425	2 213	3 438	2 457	1 607	453	64
fréquences absolues cumulées	48	343	1 768	3 981	7 419	9 876	11 483	11 936	12 000

[ii] Diagramme en bâtons des fréquences (absolues)



[b] • La valeur de plus grande fréquence est 4, donc le mode est 4.

• Le nombre de familles est $n = 12\,000 = 2 \times 6\,000$; par conséquent, il y a deux valeurs situées au centre de la série statistique, ce sont les valeurs placées aux rang 6 000 et 6 001. Il en résulte que la médiane est 4.

• La moyenne m est :

$$m = \frac{1}{12\,000}[0 \times 48 + 1 \times 295 + 2 \times 1\,425 + 3 \times 2\,215 + 4 \times 3\,438 \\ +5 \times 2\,457 + 6 \times 1\,607 + 7 \times 453 + 8 \times 64]$$

$$m = \frac{49\,146}{12\,000} = 4{,}095\,5.$$

[c] Soit m_2 le moment d'ordre 2 :

$$m_2 = \frac{1}{12\,000}\Big[0^2 \times 48 + 1^2 \times 295 + 2^2 \times 1\,425 + 3^2 \times 2\,213 + 4^2 \times 3\,438 \\ + 5^2 \times 2\,457 + 6^2 \times 1\,607 + 7^2 \times 453 + 8^2 \times 64\Big]$$

$$m_2 = \frac{226\,490}{12\,000}.$$

La variance v est :

$$v = m_2 - m^2 = \frac{226\,490 \times 12\,000 - (49\,146)^2}{(12\,000)^2} = 2{,}101\,04\ldots$$

d'où l'écart type :

$$\sigma = \sqrt{v} \simeq 1{,}45.$$

[2.3.10] [a] [i] Médiane

Il y a $n = 1\,000$ observations, donc la médiane appartient à la classe de centre 7,2 et d'extrémités 7 et 7,4.

On calcule la médiane M par interpolation linéaire (§ [3][A][b]) :

$$M = 7 + \frac{500 - 355}{270} \times 0{,}4 = 7{,}214\,8\ldots \simeq 7{,}215.$$

[ii] Moyenne m

$$m = \frac{7\,250}{1\,000} = 7{,}25.$$

[b] [i] Quartile et semi-interquartile

Le premier quartile Q_1 appartient à la classe de centre 6,8 et d'extrémités 6,6 et 7. On calcule Q_1 par interpolation linéaire (§ [3][B][b]) :

$$Q_1 = 6{,}6 + \frac{250 - 115}{240} \times 0{,}4 = 6{,}825.$$

Le troisième quartile Q_3 appartient à la classe de centre 7,6 et d'extrémités 7,4 et 7,8. Par interpolation linéaire, on a :

$$Q_3 = 7{,}4 + \frac{750 - 625}{175} \times 0{,}4 = 7{,}685\,7\ldots \simeq 7{,}686.$$

Le semi-interquartile est :

$$Q = \frac{1}{2}(Q_3 - Q_1) \simeq 0{,}430\,5.$$

Écart moyen

$$e_m = \frac{1}{1\,000}[5 \times 1{,}65 + 40 \times 1{,}25 + 70 \times 0{,}85 + 240 \times 0{,}45 + 270 \times 0{,}05$$
$$+ 175 \times 0{,}35 + 140 \times 0{,}75 + 50 \times 1{,}15 + 10 \times 1{,}55]$$

$$e_m = \frac{478{,}5}{1\,000} = 0{,}478\,5.$$

Écart type

Le moment d'ordre 2 est :

$$m_2 = \frac{1}{1\,000}\big[5 \times (5{,}6)^2 + 40 \times (6{,}0)^2 + 70 \times (6{,}4)^2 + 240 \times (6{,}8)^2 + 270 \times (7{,}2)^2$$
$$+ 175 \times (7{,}6)^2 + 140 \times (8{,}0)^2 + 50 \times (8{,}4)^2 + 10 \times (8{,}8)^2\big]$$

$$m_2 = \frac{52\,928{,}8}{1000} = 52{,}928\,8.$$

La variance est donc :

$$v = m_2 - m^2 = 52{,}928\,8 - 52{,}526\,5 = 0{,}366\,3.$$

D'où l'écart type :

$$\sigma = \sqrt{v} = 0{,}605\,2\ldots \simeq 0{,}605.$$

[2.3.11] Ordonnons la suite des mesures par ordre de valeurs non décroissantes :

15,86 15,88 15,88 15,88 15,88 15,91 15,91
15,91 15,93 15,95 15,96 15,96 16,01 19,81.

[a] Étendue de la série statistique (§ **[2][c]**) :

$$E = 19{,}81 - 15{,}86 = 3{,}95.$$

La valeur suspecte (valeur aberrante) : 19,81. En supprimant cette valeur, la nouvelle série est donc :

15,86 15,88 15,88 15,88 15,88 15,91 15,91
15,91 15,93 15,95 15,96 15,96 16,01,

et l'étendue de celle-ci est donc :

$$E' = 16{,}01 - 15{,}86 = 0{,}15.$$

[b] Moyenne m et écart type σ de la nouvelle série :

$$m = 15{,}9169\ldots \simeq 15{,}92$$
$$\sigma = 0{,}0415\ldots \simeq 0{,}042.$$

[c]

$$m - 2\sigma = 15{,}8337\ldots \simeq 15{,}83$$
$$m + 2\sigma = 16{,}000\ldots \simeq 16{,}00.$$

Toutes les valeurs de la nouvelle série statistique sauf la dernière valeur, c'est-à-dire 16,01, sont dans l'intervalle $[m - 2\sigma, m + 2\sigma]$.

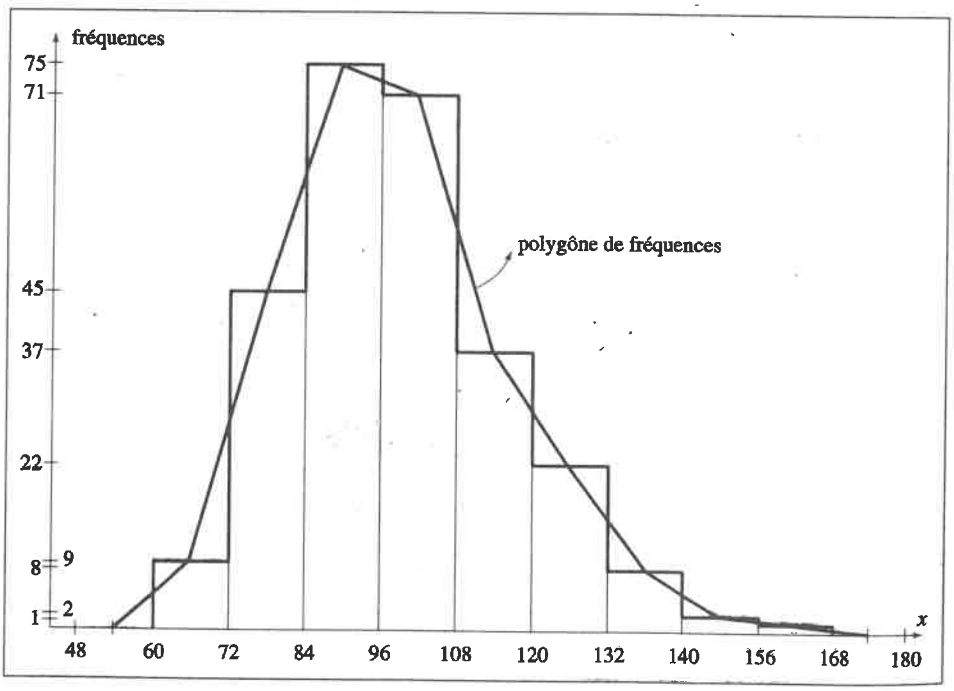
On élimine la valeur 16,01 et on considère alors la série de mesures :

15,86	15,88	15,88	15,88	15,88	15,91
15,91	15,91	15,93	15,95	15,96	15,96.

D'où le poids atomique :

$$15{,}9091 \simeq 15{,}91.$$

[2.3.12] [a] Histogramme des fréquences (absolues)



[b] La classe de plus grande fréquence ou classe modale est la classe de centre 90 et d'extrémités 84 et 96.

Moyenne

$$m = \frac{1}{270}[9 \times 66 + 45 \times 78 + 75 \times 90 + 71 \times 102 + 37 \times 114$$
$$+22 \times 126 + 8 \times 138 + 2 \times 150 + 1 \times 162]$$

$$m = \frac{26\,652}{270} \simeq 98{,}71.$$

Écart type

$$m_2 = \frac{1}{270}\left[9 \times (66)^2 + 45 \times (78)^2 + 75 \times (90)^2 + 71 \times (102)^2 + 37 \times (114)^2\right.$$
$$\left.+22 \times (126)^2 + 8 \times (138)^2 + 2 \times (150)^2 + 1 \times (162)^2\right]$$

$$m_2 = \frac{2\,712\,888}{270} \simeq 10\,047{,}73.$$

La variance est donc[†] :

$$v = m_2 - m^2 = 303{,}849\,8\ldots$$

D'où l'écart type :

$$\sigma = \sqrt{v} \simeq 17{,}43.$$

[c] [i] Médiane

La médiane M est dans la classe de centre 102, d'extrémités 96 et 108 puisque la fréquence cumulée de cette classe est 200 $\left(\frac{n}{2} = 135\right)$; M est déterminée par interpolation linéaire (§ [3][A][b]) :

$$M = 96 + \frac{135 - 129}{71} \times 12 \simeq 97{,}014.$$

[ii] Les quartiles Q_1 et Q_3

Q_1 est dans la classe de centre 90 et d'extrémités 84 et 96, d'où :

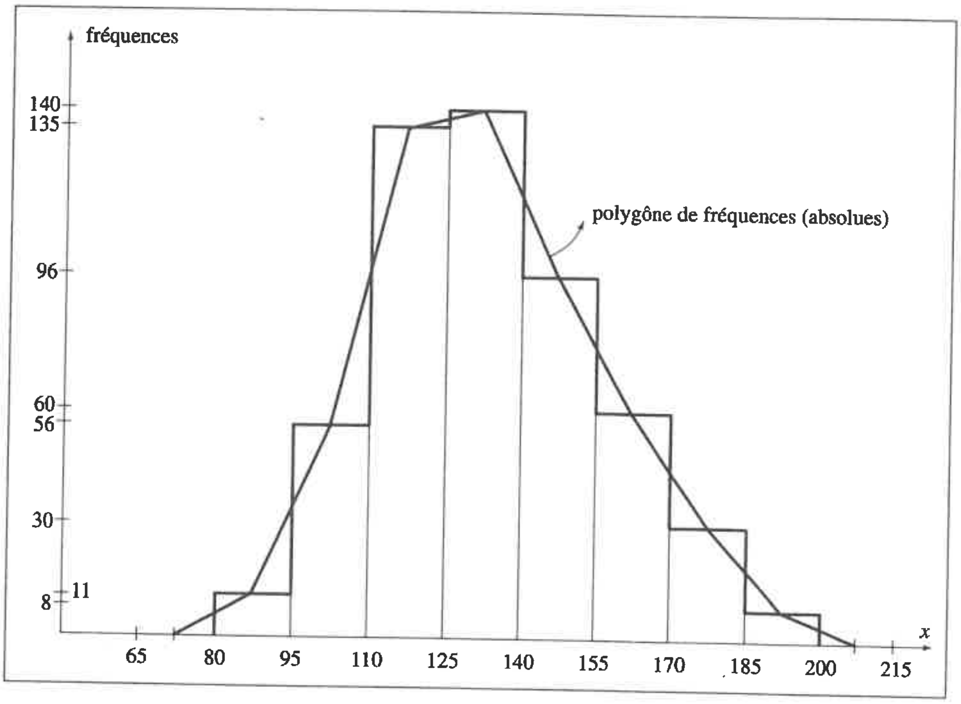
$$Q_1 = 84 + \frac{67{,}5 - 54}{75} \times 12 = 86{,}16.$$

Q_3 est dans la classe de centre 114 et d'extrémités 108 et 120, d'où :

$$Q_1 = 108 + \frac{202{,}5 - 200}{37} \times 12 \simeq 108{,}81.$$

† Pour calculer v, on écrit $m_2 - m^2$ sous forme fractionnaire et ensuite on fait l'approximation de σ.

[2.3.13] [a] Histogramme et polynôme des fréquences (absolues)



[b] Tableau indiquant les centres, les fréquences et les fréquences cumulées des classes.

centres	87,5	102,5	117,5	132,5	147,5	162,5	177,5	192,5
fréquences	11	56	135	140	96	60	30	8
fréquences cumulées	11	67	202	342	438	498	528	536

[c] La classe de plus grande fréquence ou classe modale est la classe de centre 132,5. Les limites de cette classe modale sont 125 et 140.

Médiane

L'effectif total est $n = 536$, or $\dfrac{n}{2} = 268$, par conséquent la médiane M appartient à la classe modale. On détermine M par interpolation linéaire (§ **[3][A][b]**) :

$$M = 125 + \frac{268 - 202}{140} \times 15 \simeq 132{,}07.$$

[d] Moyenne

Soit m la moyenne de la série statistique (connue par un partage en classes) :

$$m = \frac{1}{536}[11 \times 87{,}5 + 56 \times 102{,}5 + 135 \times 117{,}5 + 140 \times 132{,}5$$
$$+96 \times 147{,}5 + 60 \times 162{,}5 + 30 \times 177{,}5 + 8 \times 192{,}5]$$

$$m = \frac{71\,890}{536} \simeq 134{,}12.$$

Écart type

$$m_2 = \frac{1}{536}\left[11 \times (87{,}5)^2 + 56 \times (102{,}5)^2 + 135 \times (117{,}5)^2 + 140 \times (132{,}5)^2\right.$$
$$\left.+96 \times (147{,}5)^2 + 60 \times (162{,}5)^2 + 30 \times (177{,}5)^2 + 8 \times 192{,}5)^2\right]$$

$$m_2 = \frac{9\,908\,900}{536} \simeq 18\,486{,}75.$$

La variance est :

$$v = m_2 - m^2 = \frac{142\,998\,300}{(536)^2} \simeq 497{,}74.$$

D'où l'écart type :

$$\sigma = \sqrt{v} \simeq 22{,}31.$$

Coefficient de dérive

Le coefficient de dérive (ou coefficient d'asymétrie de Fisher) est défini par (§ [3][C][b]) :

$$d = \frac{\mu_3}{\sigma^3},$$

où μ_3 est le moment centré d'ordre 3.

On trouve :

$$\mu_3 = \frac{2\,134\,360{,}263}{536} = 3\,982{,}015\,416.$$

D'où :

$$d = \frac{\mu_3}{\sigma^3} \simeq 0{,}36.$$

[Remarque] Le polygône des fréquences montre une assez bonne symétrie de la distribution expérimentale. Le coefficient $d \simeq 0{,}36$ indique que l'ensemble des valeurs de la série statistique est légèrement plus étalé du côté des valeurs supérieures à la moyenne.

On notera aussi que la moyenne et la médiane appartiennent à la même classe qui est la classe modale.

[2.3.14] On sait que (§ **[3][C][a]**) :

$$m_q = \frac{1}{n}\sum_{i=1}^{n} x_i^q$$

$$\mu_q = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^q, \quad q \in \mathbb{N}^*,$$

où m désigne la moyenne de la série statistique.

Pour $q = 0$, la formule est évidente. Supposons q entier ≥ 1. On a :

$$(x_i - m)^q = \sum_{k=0}^{q}(-1)^k \frac{q!}{k!(q-k)!} x_i^{q-k} m^k$$

$$\mu_q = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{k=0}^{q}(-1)^k \frac{q!}{k!(q-k)!} x_i^{q-k} m^k\right).$$

Et en échangeant les signes de sommation :

$$\mu_q = \frac{1}{n}\sum_{k=0}^{q}\left((-1)^k \frac{q!}{k!(q-k)!} m^k \sum_{i=1}^{n} x_i^{q-k}\right)$$

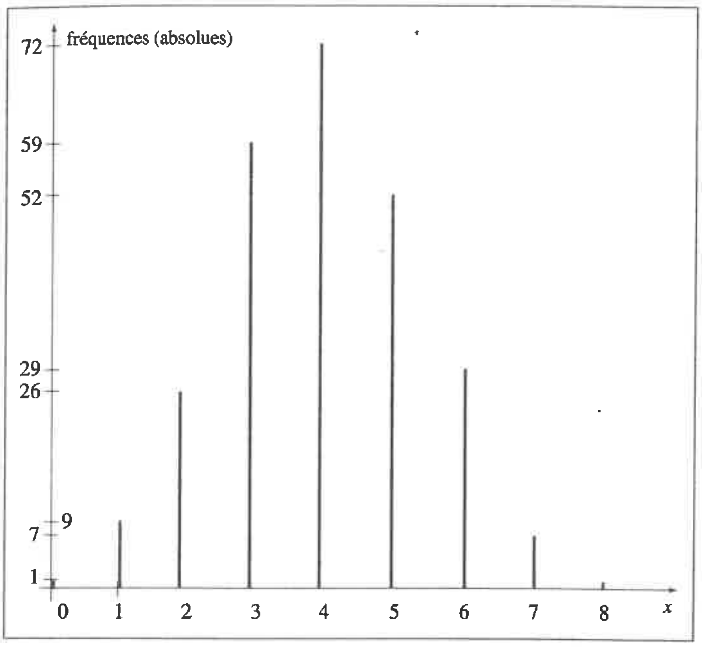
$$\mu_q = \sum_{k=0}^{q}\left((-1)^k \frac{q!}{k!(q-k)!} m_1^k \frac{1}{n}\sum_{i=1}^{n} x_i^{q-k}\right), \quad \text{car } m_1 = m..$$

D'où :

$$\mu_q = \sum_{k=0}^{q}(-1)^k \frac{q!}{k!(q-k)!} m_1^k m_{q-k}, \quad \forall q \in \mathbb{N}^*.$$

e la distri-
leurs de la
ıoyenne.
qui est la

[2.3.15] [a] Diagramme en batôns des fréquence (absolues) de la série statistique



[b] [i] Valeur modale

La valeur de plus grande fréquence ou valeur modale est 4.

[ii] Moyenne m

$$m = \frac{1}{256}[0 \times 1 + 1 \times 9 + 2 \times 26 + 3 \times 59 + 4 \times 72 + 5 \times 52 + 6 \times 29 + 7 \times 7 + 8 \times 1]$$

$$m = \frac{1\,017}{256} \simeq 3{,}97.$$