

VARIABLES QUANTITATIVES CONTINUES

§ 1. — Exercices de synthèse	1
§ 2. — Boîtes à moustaches et intervalles de variation	8
§ 3. — Changement d'unités	11

§ 1. — Exercices de synthèse

Exercice 1.1. La répartition des salariés français en 2004 selon la durée de trajet (en minute) pour se rendre à leur travail est donné par le tableau suivant :

DURÉE]0 ; 5]]5 ; 10]]10 ; 15]]15 ; 20]]20 ; 25]]25 ; 30]]30 ; 40]]40 ; 50]]50 ; 70]]70 ; 120]
POURCENTAGE	2,85	9,96	13,58	12,86	11,41	9,47	13,61	8,15	7,96	10,15

SOURCE : DADS, 2004 (INSEE).

(La durée de trajet maximale a été arbitrairement choisie à 120 min pour les besoins de l'exercice.)

- (a) Déterminer la population et la variables étudiée (en précisant son type).
- (b) Représenter graphiquement la distribution des proportions.
- (c) Déterminer la classe modale de la variable.
- (d) Représenter graphiquement la proportion de salariés ayant un trajet compris entre 7 et 22 minutes. Calculer cette proportion.
- (e) Calculer, à l'aide de la fonction de répartition, la proportion de salariés mettant moins de 35 minutes pour aller à leur lieu de travail. Même question pour un temps de trajet compris entre 17 et 35 minutes.
- (f) Déterminer la moyenne et l'écart-type de la variable.
- (g) Déterminer la médiane, les quartiles ainsi que les premier et neuvièmes déciles.
- (h) Représenter graphiquement la fonction de répartition.

Corrigé de l'exercice 1.1.

(a) **Population** : salariés Français en 2004.

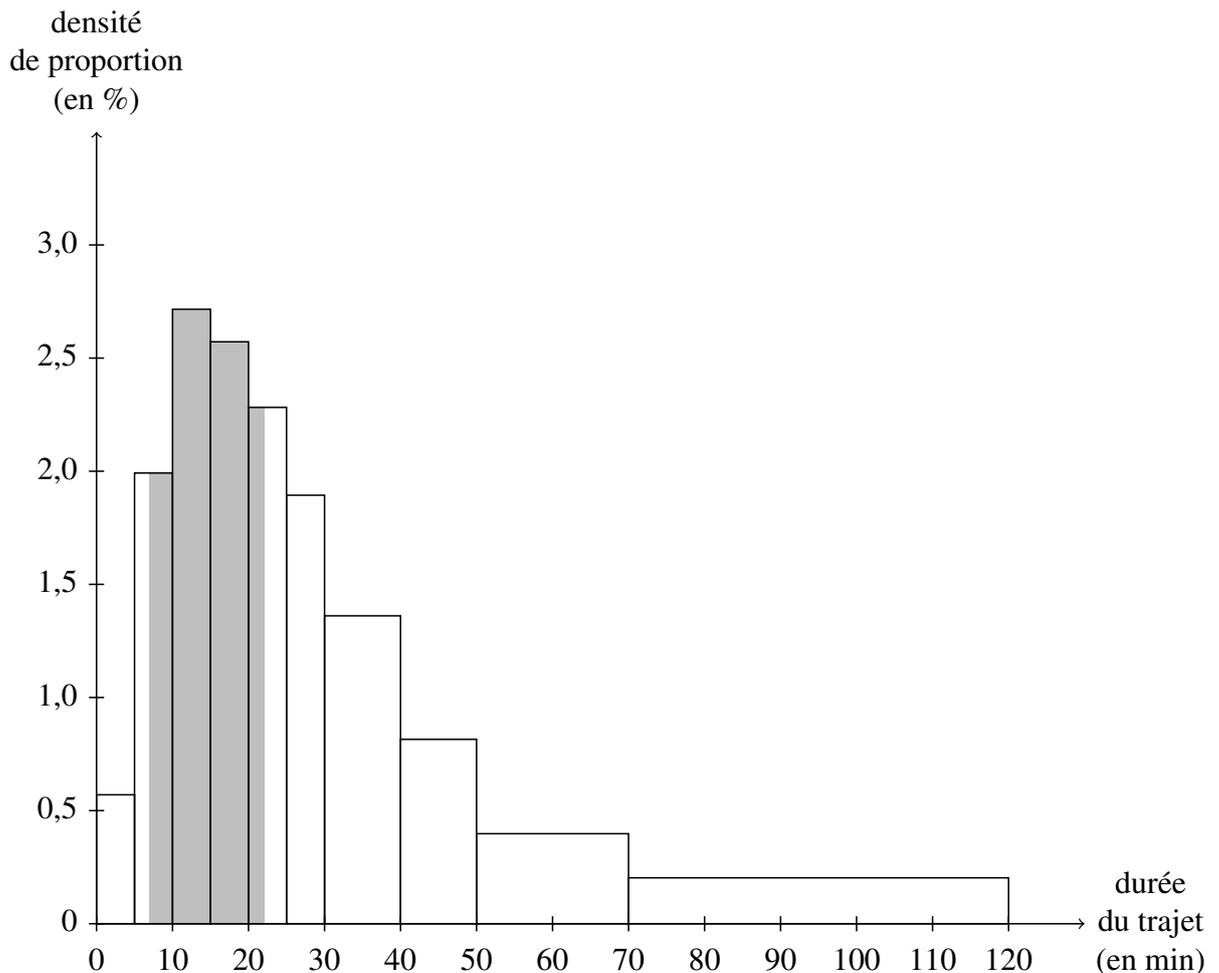
Individu : un salarié Français en 2004.

Variable étudiée : « temps de trajet » ; c'est une variable quantitative continue qui prend des modalités entre 0 et 200 minutes.

(b) La variable étant continue, on va la représenter graphiquement à l'aide d'un histogramme. Pour cela on a besoin de déterminer les densités de proportion des classes et donc de calculer les amplitudes. Vu les questions suivantes (sur la fonction de répartition, la moyenne, la médiane), on met également dans le tableau les proportions cumulées ainsi que les centres des classes.

DURÉE]0 ; 5]]5 ; 10]]10 ; 15]]15 ; 20]]20 ; 25]]25 ; 30]]30 ; 40]]40 ; 50]]50 ; 70]]70 ; 120]
PROP. (%)	2,85	9,96	13,58	12,86	11,41	9,47	13,61	8,15	7,96	10,15
AMPLITUDE	5	5	5	5	5	5	10	10	20	50
DENS. PROP.(%)	0,57	1,992	2,716	2,572	2,282	1,894	1,361	0,815	0,398	0,203
PROP. CUM.(%)	2,85	12,81	26,39	39,25	50,66	60,13	73,74	81,89	89,85	100
CENTRE	2,5	7,5	12,5	17,5	22,5	27,5	35	45	60	95

On trace l'histogramme. Le choix de l'échelle sur l'axe des abscisses doit permettre d'aller de 0 à 200 et le choix sur l'axe des ordonnées d'aller de 0 à la plus grande densité de proportion, à savoir 2,719. Les rectangles de l'histogramme ont pour largeur l'amplitude de la classe et pour hauteur la densité de proportion :



(c) La classe modale est la classe qui a la plus grande densité de proportion, donc c'est $]10 ; 15]$.

(d) Sur l'histogramme précédent on a mis les salariés dont le temps de trajet est compris entre 7 et 22 minutes en grisé. Calculons la proportion correspondante (rappelons que l'aire d'un rectangle est égale à sa base multipliée par sa hauteur) :

$$\begin{aligned}
 P(7 \leq X \leq 22) &= \text{aire de la partie hachurée} \\
 &= \text{somme des aires des rectangles hachurés} \\
 &= 3 \times 1,992 + 5 \times 2,716 + 5 \times 2,572 + 2 \times 2,282 \\
 &= 5,976 + 13,580 + 12,860 + 4,564 \\
 &= 36,98
 \end{aligned}$$

Il y a environ 36,98 % de salariés français dont le temps de trajet est compris entre 7 et 22 minutes.

(e) La classe contenant $x = 35$ est $]a ; b] =]30 ; 40]$ donc $a = 30$ et $b = 40$ d'où $F(a) = P(X \leq$

30) = 60,13 et $F(b) = P(X \leq 40) = 73,74$. Appliquons la formule :

$$\begin{aligned}
 F(35) &= F(x) = F(a) + \left[(F(b) - F(a)) \times \frac{x - a}{b - a} \right] \\
 &= F(30) + \left[(F(40) - F(30)) \times \frac{35 - 30}{40 - 30} \right] \\
 &= 60,13 + \left[(73,74 - 60,13) \times \frac{5}{10} \right] \\
 &= 60,13 + \left[13,61 \times \frac{1}{2} \right] \\
 &= 60,13 + 6,805 \\
 &= 66,935
 \end{aligned}$$

Il y a environ 66,935 % de salariés français dont le temps de trajet est inférieur à 35 minutes.

Pour calculer $P(17 \leq X \leq 35)$, on utilise la formule

$$P(17 \leq X \leq 35) = F(35) - F(17).$$

On a déjà calculé $F(35)$, donc il ne nous manque que $F(17)$. La classe contenant $x = 17$ est $]a; b] =]15; 20]$ donc $a = 15$ et $b = 20$ d'où $F(a) = P(X \leq 15) = 26,39$ et $F(b) = P(X \leq 20) = 39,25$. Appliquons la formule :

$$\begin{aligned}
 F(17) &= F(x) = F(a) + \left[(F(b) - F(a)) \times \frac{x - a}{b - a} \right] \\
 &= F(15) + \left[(F(20) - F(15)) \times \frac{17 - 15}{20 - 15} \right] \\
 &= 26,39 + \left[(39,25 - 26,39) \times \frac{2}{5} \right] \\
 &= 26,39 + [12,86 \times 0,4] \\
 &= 26,39 + 5,144 \\
 &= 31,534
 \end{aligned}$$

Il y a environ 31,534 % de salariés français dont le temps de trajet est inférieur à 17 minutes.

On a donc :

$$P(17 \leq X \leq 35) = F(35) - F(17) = 66,935 - 31,534 = 35,401.$$

Il y a environ 35,401 % de salariés français dont le temps de trajet est compris entre 17 et 35 minutes.

(f) Puisqu'on n'a pas les effectifs mais uniquement les proportions, on utilise la formule

$$\mu = \sum_{i=1}^k p_i x_i,$$

où p_i est la proportion de la i -ième classe (exprimée comme un nombre compris entre 0 et 1 et non comme un pourcentage) et x_i le centre de la i -ième classe. Cela donne :

$$\begin{aligned}\mu &= 0,0285 \times 2,5 + 0,0996 \times 7,5 + 0,1358 \times 12,5 + 0,1286 \times 17,5 + 0,1141 \times 22,5 \\ &\quad + 0,0947 \times 27,5 + 0,1361 \times 35 + 0,0815 \times 45 + 0,0796 \times 60 + 0,1015 \times 95 \\ &= 32,7873\end{aligned}$$

Le temps de trajet moyen des salariés français est de 32,7873 minutes.

L'écart-type est donné par la formule

$$\begin{aligned}\sigma &= \sqrt{\left(\sum_{i=1}^k p_i \times (x_i^2)\right) - \mu^2} \\ &= \sqrt{(0,0285 \times 2,5^2 + 0,0996 \times 7,5^2 + 0,1358 \times 12,5^2 + 0,1286 \times 17,5^2 \\ &\quad + 0,1141 \times 22,5^2 + 0,0947 \times 27,5^2 + 0,1361 \times 35^2 + 0,0815 \times 45^2 \\ &\quad + 0,0796 \times 60^2 + 0,1015 \times 95^2) - 32,7873^2} \\ &= \sqrt{1730,12 - 1075,0038} \\ &= \sqrt{655,1162} \\ &= 25,60.\end{aligned}$$

La dispersion du temps de trajet des salariés français autour de la moyenne est 25,60.

(g) **Calcul de la médiane** La proportion cumulée de 50 % correspond à la classe $[a ; b] = [20 ; 25]$ donc la médiane est donnée par la formule

$$\begin{aligned}\text{médiane} &= a + \left[(b - a) \times \frac{50 - P(X \leq a)}{P(X \leq b) - P(X \leq a)} \right] \\ &= 20 + \left[(25 - 20) \times \frac{50 - 39,25}{50,66 - 39,25} \right] \\ &= 20 + \left[5 \times \frac{10,75}{11,41} \right] \\ &= 20 + [5 \times 0,9422] \\ &= 20 + 4,71 \\ &= 24,71.\end{aligned}$$

Au moins la moitié des temps de trajet est $\leq 24,71$ minutes et au moins la moitié des temps de trajet est $\geq 24,71$ minutes.

Calcul du premier quartile La proportion cumulée de 25 % correspond à la classe $[a ; b] =$

]10 ; 15] donc le premier quartile est donné par la formule

$$\begin{aligned} Q_1 &= a + \left[(b - a) \times \frac{25 - P(X \leq a)}{P(X \leq b) - P(X \leq a)} \right] \\ &= 10 + \left[(15 - 10) \times \frac{25 - 12,81}{26,39 - 12,81} \right] \\ &= 10 + \left[5 \times \frac{12,19}{13,58} \right] \\ &= 10 + [5 \times 0,8976] \\ &= 10 + 4,49 \\ &= 14,49. \end{aligned}$$

Au moins un quart des temps de trajet est $\leq 14,49$ minutes et au moins trois quart des temps de trajet est $\geq 14,49$ minutes.

Calcul du troisième quartile La proportion cumulée de 75 % correspond à la classe]a ; b] =]40 ; 50] donc le troisième quartile est donné par la formule

$$\begin{aligned} Q_3 &= a + \left[(b - a) \times \frac{75 - P(X \leq a)}{P(X \leq b) - P(X \leq a)} \right] \\ &= 40 + \left[(50 - 40) \times \frac{75 - 73,74}{82,89 - 73,74} \right] \\ &= 40 + \left[10 \times \frac{1,26}{8,15} \right] \\ &= 40 + [10 \times 0,1377] \\ &= 40 + 1,38 \\ &= 41,38. \end{aligned}$$

Au moins trois quart des temps de trajet est $\leq 41,38$ minutes et au moins un quart des temps de trajet est $\geq 41,38$ minutes.

Calcul du premier décile La proportion cumulée de 10 % correspond à la classe]a ; b] =]5 ; 10] donc le premier décile est donné par la formule

$$\begin{aligned} D_1 &= a + \left[(b - a) \times \frac{10 - P(X \leq a)}{P(X \leq b) - P(X \leq a)} \right] \\ &= 5 + \left[(10 - 5) \times \frac{10 - 2,85}{12,81 - 2,85} \right] \\ &= 5 + \left[5 \times \frac{7,15}{9,96} \right] \\ &= 5 + [5 \times 0,7179] \\ &= 5 + 3,59 \\ &= 8,59. \end{aligned}$$

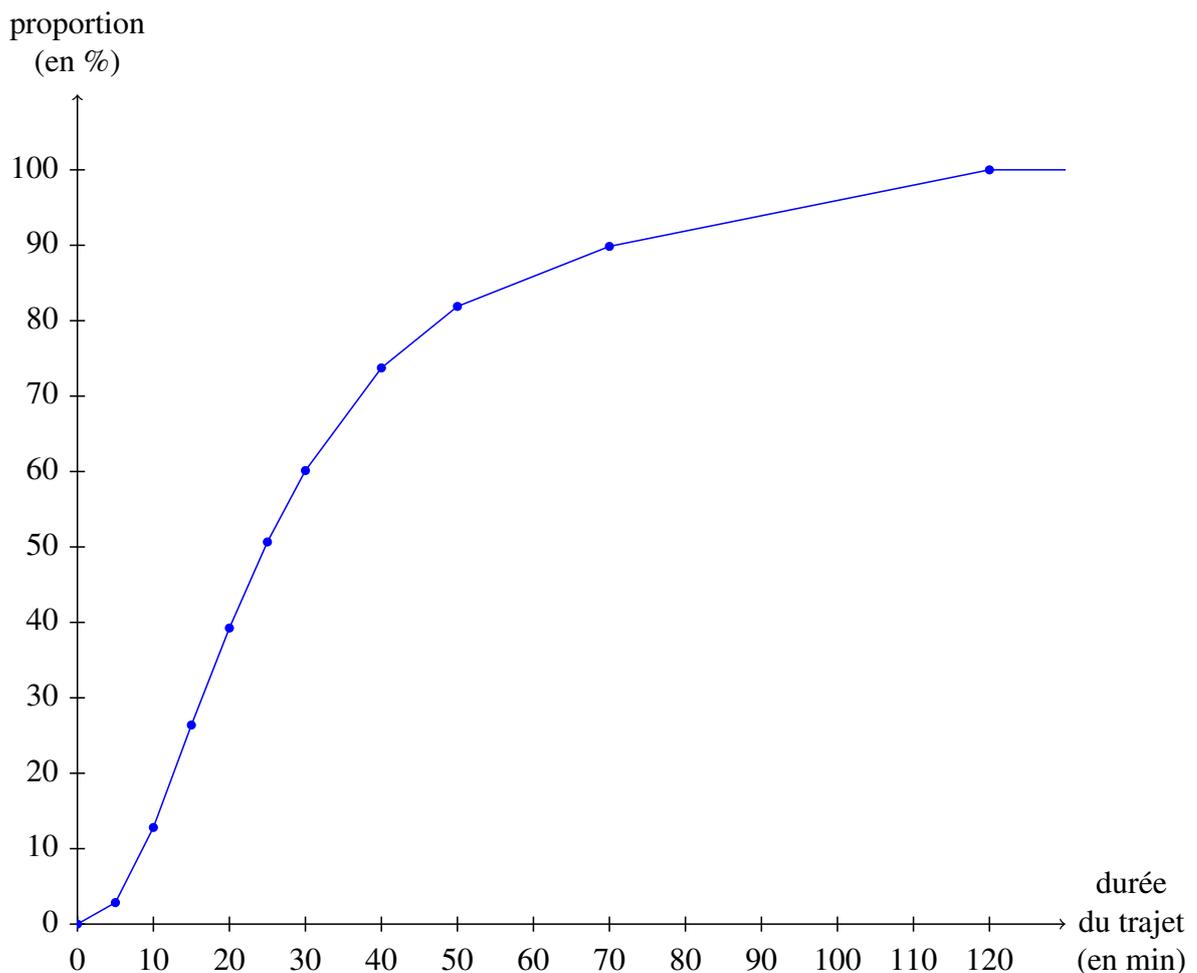
Au moins 10 % des temps de trajet est $\leq 8,59$ minutes et au moins 90 % des temps de trajet est $\geq 8,59$ minutes.

Calcul du neuvième décile La proportion cumulée de 90 % correspond à la classe $]a ; b] =]70 ; 120]$ donc le premier décile est donné par la formule

$$\begin{aligned}
 D_9 &= a + \left[(b - a) \times \frac{90 - P(X \leq a)}{P(X \leq b) - P(X \leq a)} \right] \\
 &= 70 + \left[(120 - 70) \times \frac{90 - 89,85}{100 - 89,85} \right] \\
 &= 70 + \left[50 \times \frac{0,15}{10,15} \right] \\
 &= 70 + [50 \times 0,0148] \\
 &= 70 + 0,74 \\
 &= 70,74.
 \end{aligned}$$

Au moins 90 % des temps de trajet est $\leq 70,74$ minutes et au moins 10 % des temps de trajet est $\geq 70,74$ minutes.

- (h) Pour représenter la fonction de répartition, on place, pour chacune des bornes des classes, la proportion cumulée (par exemple, le point 25 a pour ordonnée 50,66 %) puis on les relie par des segments de droites :



§ 2. — Boîtes à moustaches et intervalles de variation

Exercice 2.1. Le temps de trajet domicile-travail (en minutes) déclaré par les travailleurs américains a été récolté en 2000 par l'U.S. Census Bureau :

DURÉE]0 ; 5]]5 ; 10]]10 ; 15]]15 ; 20]]20 ; 25]]25 ; 30]]30 ; 35]]35 ; 45]]45 ; 60]]60 ; 90]]90 ; 150]
PROP. CUM.(%)	3,37	14,4	29,4	45,22	59,71	65,5	78,69	84,6	92,01	97,22	100

SOURCE : <http://www.census.gov/prod/2004pubs/c2kbr-33.pdf>

- (a) Représenter la boîte à moustache.
 (b) Déterminer l'intervalle de variation à 75 %.
 (c) Déterminer l'intervalle de variation à 95 %.

Corrigé de l'exercice 2.1.

(a) Pour dessiner la boîte à moustache, on a besoin des trois quartiles : Q_1 , la médiane et Q_3 .

Calcul de la médiane. La proportion cumulée de 50 % correspond à la classe $]a ; b] =]20 ; 25]$ donc la médiane est donnée par la formule

$$\begin{aligned}
 \text{médiane} &= a + \left[(b - a) \times \frac{50 - P(X \leq a)}{P(X \leq b) - P(X \leq a)} \right] \\
 &= 20 + \left[(25 - 20) \times \frac{50 - 45,22}{59,71 - 45,22} \right] \\
 &= 20 + \left[5 \times \frac{4,78}{14,49} \right] \\
 &= 20 + [5 \times 0,3299] \\
 &= 20 + 1,65 \\
 &= 21,65.
 \end{aligned}$$

Calcul du premier quartile. La proportion cumulée de 25 % correspond à la classe $]a ; b] =]10 ; 15]$ donc le premier quartile est donnée par la formule

$$\begin{aligned}
 Q_1 &= a + \left[(b - a) \times \frac{25 - P(X \leq a)}{P(X \leq b) - P(X \leq a)} \right] \\
 &= 10 + \left[(15 - 10) \times \frac{25 - 14,40}{29,40 - 14,40} \right] \\
 &= 10 + \left[5 \times \frac{10,60}{15,00} \right] \\
 &= 10 + [5 \times 0,7067] \\
 &= 10 + 3,53 \\
 &= 13,53.
 \end{aligned}$$

Calcul du troisième quartile. La proportion cumulée de 75 % correspond à la classe

$]a; b] =]30; 35]$ donc le troisième quartile est donnée par la formule

$$\begin{aligned}
 Q_3 &= a + \left[(b - a) \times \frac{75 - P(X \leq a)}{P(X \leq b) - P(X \leq a)} \right] \\
 &= 30 + \left[(35 - 30) \times \frac{75 - 65,50}{78,69 - 65,50} \right] \\
 &= 30 + \left[5 \times \frac{9,50}{13,19} \right] \\
 &= 30 + [5 \times 0,7202] \\
 &= 30 + 3,60 \\
 &= 33,60.
 \end{aligned}$$

Boîte à moustache Voici la boîte à moustache représentant l'intervalle interquartile pour la durée du trajet :



- (b) Rappelons que $75\% = 0,75$. Pour déterminer l'intervalle de variation à $1 - \alpha = 0,75$, on calcule d'abord $\alpha = 1 - 0,75 = 0,25$ puis $\alpha/2 = 0,125$ et $1 - \alpha/2 = 0,875$. L'intervalle de variation à 75% est alors l'intervalle $[q_{0,125}; q_{0,875}]$. Il nous reste donc à calculer les quantiles d'ordre $0,125$ et $0,875$.

Calcul du quantile d'ordre 0,125. La proportion cumulée de $12,5\%$ correspond à la classe $]a; b] =]5; 10]$ donc le quantile d'ordre $0,125$ est donnée par la formule

$$\begin{aligned}
 q_{0,125} &= a + \left[(b - a) \times \frac{12,5 - P(X \leq a)}{P(X \leq b) - P(X \leq a)} \right] \\
 &= 5 + \left[(10 - 5) \times \frac{12,5 - 3,37}{14,40 - 3,37} \right] \\
 &= 5 + \left[5 \times \frac{9,13}{11,03} \right] \\
 &= 5 + [5 \times 0,8277] \\
 &= 5 + 4,14 \\
 &= 9,14.
 \end{aligned}$$

Calcul du quantile d'ordre 0,875. La proportion cumulée de $87,5\%$ correspond à la

classe $]a ; b] =]45 ; 60]$ donc le quantile d'ordre 0,875 est donnée par la formule

$$\begin{aligned} q_{0,875} &= a + \left[(b - a) \times \frac{87,5 - P(X \leq a)}{P(X \leq b) - P(X \leq a)} \right] \\ &= 45 + \left[(60 - 45) \times \frac{87,5 - 84,60}{92,01 - 84,60} \right] \\ &= 45 + \left[15 \times \frac{2,90}{7,41} \right] \\ &= 45 + [15 \times 0,3914] \\ &= 45 + 5,87 \\ &= 50,87. \end{aligned}$$

Conclusion L'intervalle de variation à 75 % est $[9,14 ; 50,87]$; cet intervalle contient 75 % des données.

- (c) Rappelons que 95 % = 0,95. Pour déterminer l'intervalle de variation à $1 - \alpha = 0,95$, on calcule d'abord $\alpha = 1 - 0,95 = 0,05$ puis $\alpha/2 = 0,025$ et $1 - \alpha/2 = 0,975$. L'intervalle de variation à 95 % est alors l'intervalle $[q_{0,025} ; q_{0,975}]$. Il nous reste donc à calculer les quantiles d'ordre 0,025 et 0,975.

Calcul du quantile d'ordre 0,025. La proportion cumulée de 2,5 % correspond à la classe $]a ; b] =]0 ; 5]$ donc le quantile d'ordre 0,025 est donnée par la formule

$$\begin{aligned} q_{0,025} &= a + \left[(b - a) \times \frac{2,5 - P(X \leq a)}{P(X \leq b) - P(X \leq a)} \right] \\ &= 0 + \left[(5 - 0) \times \frac{2,5 - 0}{3,37 - 0} \right] \\ &= 0 + \left[5 \times \frac{2,50}{3,37} \right] \\ &= 0 + [5 \times 0,7418] \\ &= 0 + 3,71 \\ &= 3,71. \end{aligned}$$

Calcul du quantile d'ordre 0,975. La proportion cumulée de 97,5 % correspond à la classe $]a ; b] =]90 ; 150]$ donc le quantile d'ordre 0,975 est donnée par la formule

$$\begin{aligned} q_{0,975} &= a + \left[(b - a) \times \frac{97,5 - P(X \leq a)}{P(X \leq b) - P(X \leq a)} \right] \\ &= 90 + \left[(150 - 90) \times \frac{97,5 - 97,22}{100 - 97,22} \right] \\ &= 90 + \left[60 \times \frac{0,28}{2,78} \right] \\ &= 90 + [60 \times 0,1007] \\ &= 90 + 6,04 \\ &= 96,04. \end{aligned}$$

Conclusion L'intervalle de variation à 95 % est $[3,71 ; 96,04]$; cet intervalle contient 95 % des données.

§ 3. — Changement d'unités

Exercice 3.1. La température moyenne à Paris entre 1961 et 1990 est de $11,71\text{ }^{\circ}\text{C}$ avec un écart-type de $5,47$ tandis que dans la ville de Trappes, elle est de $10,30\text{ }^{\circ}\text{C}$ avec un écart-type de $5,35$. Sachant que $[\text{F}] = [\text{C}] \times \frac{9}{5} + 32$, convertir les données précédentes en degrés Fahrenheit.

Corrigé de l'exercice 3.1. La température moyenne à Paris est de $11,71 \times \frac{9}{5} + 32 = 53,08\text{ }^{\circ}\text{F}$ et celle à Trappes est de $10,30 \times \frac{9}{5} + 32 = 50,54\text{ }^{\circ}\text{F}$.

Pour déterminer les écart-types, on ne prend pas en compte le $+ 32$ lors de la conversion. L'écart-type à Paris est donc de $5,47 \times \frac{9}{5} = 9,85\text{ }^{\circ}\text{F}$ et celui à Trappes de $5,35 \times \frac{9}{5} = 9,63\text{ }^{\circ}\text{F}$.