CORRIGE CONTRÔLE CONTINU BLANC

I EXERCICE-1

1. 10% des français ont un revenu mensuel inférieur ou égal à 776 euros et 50%, un revenu mensuel inférieur ou égal à 1362 euros.
2. L'écart inter-décile est : $D9 - D1 = 1672$, donc $D9 = 1672 + 776 = 2448$ €. On en déduit que 90% des français ont un revenu inférieur ou égal à 2448 € ; par ailleurs, 80% des français ont un revenu situé entre 776 et 2448 euros.

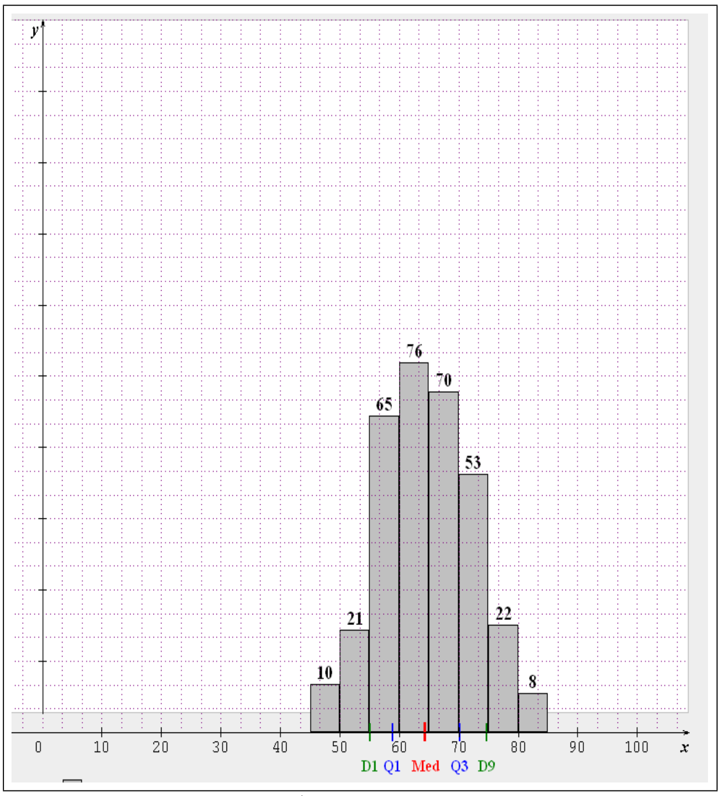
II EXERCICE-2

Attention : il y avait une erreur dans le texte : $Q_3 = 70.17$ et non 68.17.
On donne ici le tableau statistique qui permettra de répondre aux différentes questions de l'exercice.

1.

a_i	b_i	x_i	n_i	$n_i cc$	$n_i x_i$	$n_i x_i^2$	$n_i(x_i\text{-moy})^3$
45	50	47.5	10	10	475	22562.5	-50069.62
50	55	52.5	21	31	1102.5	57881.25	-37273.78
55	60	57.5	65	96	3737.5	214906.25	-23339.91
60	65	62.5	76	172	4750	296875	-711.60
65	70	67.5	70	242	4725	318937.5	1693.68
70	75	72.5	53	295	3842.5	278581.25	26054.81
75	80	77.5	22	317	1705	132137.5	47142.72
80	85	82.5	8	325	660	54450	45823.58
			325		20997.5	1376331.25	9319.89

Il s'agit d'un caractère quantitatif continu dont les classes sont d'égales amplitudes, il n'y a donc pas à corriger les effectifs.

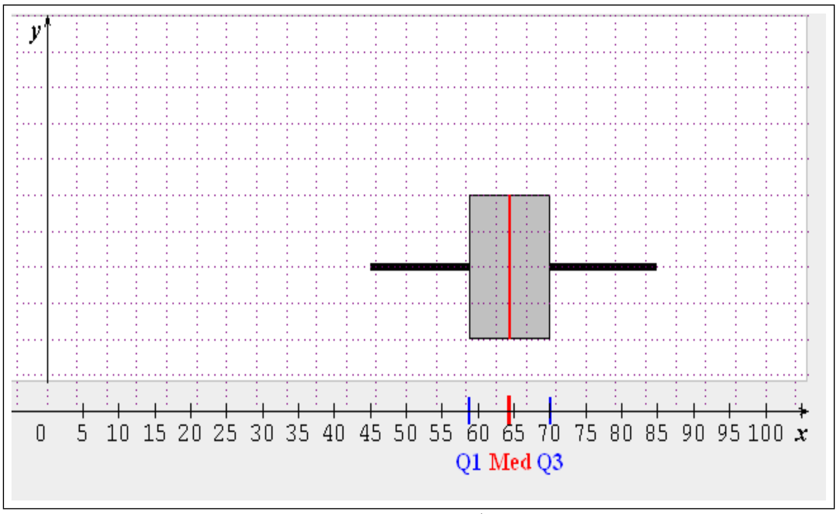


Histogramme

2. La classe modale est la classe de plus grande densité, donc dans le cas d'amplitudes égales, de plus grand effectif, c'est donc la classe $[60; 65[$. Il reste à calculer le mode à l'intérieur de la classe modale en considérant les classes encadrant la classe modale, ce qui donne avec les notations du cours : $\begin{cases} x_1 = 60 \\ x_2 = 65 \end{cases}$, $\begin{cases} h = 76 \\ h_1 = 65 \text{ et } h_2 = 70 \end{cases}$ $\begin{cases} k_1 = h - h_1 = 11 \\ k_2 = h - h_2 = 6 \end{cases}$ et pour conclure :
$M_o = \dfrac{k_2 x_1 + k_1 x_2}{k_2 + k_1} = \dfrac{6 * 60 + 11 * 65}{17} \simeq 63.24$; comme prévu, le mode est plus grand que 62.5, car il est attiré par la classe de droite.

3. Le premier quartile de cette série est localisé grâce aux effectifs cumulés croissants, il y a au moins 25% de la série avant Q_1, donc $\frac{325}{4} = 81.25$ observations, Q_1 est donc dans la classe $[55; 60[$; il reste à effectuer une interpolation linéaire, par exemple en écrivant l'alignement des trois points : $A(55, 31)$, $B(60, 96)$ et $M\,(Q_1; 81.25)$, ce qui donne : $\frac{96-31}{60-55} = \frac{81.25-31}{Q_1-55}$ soit : $Q_1 - 55 = \frac{81.25-31}{65} * 5$ et $Q_1 = 55 + \frac{50.25}{13} = 58.87$; il y a 25% des observations qui sont inférieures ou égales à 58.87.

4. On calcule l'écart inter-quartile : $EIQ = Q_3 - Q_1 = 70.17 - 58.87 = 11.3$; par convention on limite la taille des moustaches à $1.5EIQ$, pour éliminer les observations abérrantes. $1.5EIQ = 1.5 * 11.3 = 16.95$; la valeur minimale du caractère est de 45 et la valeur maximale de 85, ce qui donne à priori pour les tailles des moustaches : $\begin{cases} \text{à gauche } Q_1 - Min\,(x) = 58.87 - 45 = 13.87 \leq 16.95 \\ \text{à droite } Max\,(x) - Q_3 = 85 - 70.17 = 14.83 \leq 16.95 \end{cases}$; les moustaches ne posent pas de problème, on.garde comme extrémité le minimum 45 et le maximum 85.



Boîte à moustaches

5. $C_Y = \dfrac{Q_1 + Q_3 - 2M_e}{Q_3 - Q_1} = \dfrac{58.87 + 70.17 - 2 * 64.38}{9.3} \simeq 3.01 \times 10^{-2}$. Ce coefficient sans unité, est positif ; il est du signe du numérateur et nous indique que $Q_1 + Q_3 > 2M_e$ soit que $\frac{Q_1+Q_3}{2} > M_e$, donc que la médiane est à gauche du milieu de la boite à moustache : la série est légèrement asymétrique, étalée à droite.

6. Le coefficient β_1 de Pearson, défini par : $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = \dfrac{\mu_3^2}{V\,(x)^3} \dfrac{(9319.89/325)^2}{60.71^3} \simeq 3.68 \times 10^{-3}$, μ_2 et μ_3 désignant respectivement les moments centrés d'ordre 2 et 3 $(\mu_2 = V\,(x))$. β_1 est proche de zéro, la série est très peu asymétrique ; comme μ_3 est positif, elle est légèrement asymétrique étalée à droite.

7. Le tabeau statistique permet de calculer la moyenne de cette série : $\overline{x} = \frac{1}{n} \sum n_i x_i = \dfrac{20997.5}{325} \simeq 64.61$.

   Remarque : on a $M_o \leq M_e \leq \overline{x}$, ce qui est cohérent avec un étalement à droite.

8. Le tableau permet de calculer $V\,(x) = \frac{1}{n} \sum n_i x_i^2 - \overline{x}^2$, si on utilise la formule dévéloppée : $V\,(x) = \frac{1376331.25}{325} - \left(\dfrac{20997.5}{325}\right)^2 = 60.71$, soit un écart-type $\sigma\,(x) = \sqrt{60.71} = 7.79$

9. Déterminer la proportion d'observations situées dans l'intervalle : Pour déterminer l'intervalle $[\overline{x} - 1.5\sigma\,(x)\,; \overline{x} + 1.5\sigma\,(x)]$, on commence par calculer : $64.61 - 1.5 * 7.79 \simeq 52.93$ et $64.61 + 1.5 * 7.79 = 76.30$; cet intervalle contient l'intervalle $[55; 75]$, soit un nombre d'observations de : $295 - 31 = 264$ observations (on a utilisé les effectifs cumulés croissants) ; il reste à ajouter les effectifs des intervalles $[52.93; 55]$ et $[75; 76.3]$, les "morceaux " de classe dont on va estimer l'effectif en utilisant la densité : la densité de la classe $[50; 55]$ est : $d = \frac{21}{5} = 4.2$, soit 4.2 observations par unité d'amplitude donc $4.2 * (55 - 52.93) \simeq 8.7$ observations dans $[52.93; 55]$; de même dans $[75; 76.3]$, on trouve : $1.3 * \frac{22}{5} = 5.72$ observations ; finalement un total de : $264 + 8.7 + 5.7 = 278.4$, soit une proportion de : $\frac{278.4}{325} = 0.856\,6$, soit 85.66% des observations situées à moins d'un écart-type et demi de la moyenne.

10. On calcule cette probabilité en standardisant la variable aléatoire, c'est-à dire en se ramenant à la loi normale centrée réduite. On pose $Z = \frac{X-m}{\sigma}$, où m et σ désignent respectivement la moyenne et l'écart-type de X, alors Z suit la loi $\mathcal{N}(0; 1)$. On obtient : $P(64.6 - 1.5 * 7.8 \leq X \leq 64.6 + 1.5 * 7.8) = P\left(\frac{64.6-1.5*7.8-64.6}{7.8} \leq Z \leq \frac{64.6+1.5*7.8-64.6}{7.8}\right) = P\,(-1.5 \leq Z \leq 1.5) = 2F\,(1.5) - 1$, où F est la fonction de répartition de la loi normale centrée réduite, dont les valeurs sont lues dans la table ; on lit $F\,(1.5) = 0.9332$

t	0,00	0,01	0,02
0,0	0,5000	0,5040	0,5080
0,1	0,5398	0,5438	0,5478
0,2	0,5793	0,5832	0,5871
0,3	0,6179	0,6217	0,6255
0,4	0,6654	0,6591	0,6628
0,5	0,6915	0,6950	0,6985
0,6	0,7257	0,7290	0,7324
0,7	0,7580	0,7611	0,7642
0,8	0,7881	0,7910	0,7939
0,9	0,8159	08186	0,8212
1,0	0,8413	0,8438	0,8461
1,1	0,8643	0,8665	0,8686
1,2	0,8849	0,8869	0,8888
1,3	0,9032	0,9049	0,9066
1,4	[illegible]	0,9207	0,9222
1,5	0,9332	0,9345	0,9357
1,6	[illegible]	0,9463	0,9474

soit $P\left(-1.5 \leq Z \leq 1.5\right) = 2F\left(1.5\right) - 1 \simeq 2 * 0.9332 - 1 = 0.8664$, soit 86.64%.

11. L'inégalité de Bienaymé-Tchebychev nous donne en statistique une loi empirique : dans un intervalle centré autour de la moyenne, du type $\left[\overline{x} - k\sigma\left(x\right); \overline{x} + k\sigma\left(x\right)\right]$, il y a au moins $\left(1 - \frac{1}{k^2}\right) * 100\%$ observations, soit ici pour $k = 1.5$, au moins $\left(1 - \frac{1}{1.5^2}\right) * 100 \simeq 55.56\%$.

III EXERCICE-3

x_i	0	1	2	3	4	5	6	7
n_i	25	43	30	37	27	35	2	1
n_icc	25	68	98	135	162	197	199	200

1. Il s'agit d'un caractère quantitatif discret.
2. On utilise les effectifs cumulés croissants pour déterminer les deux termes centraux, l'observation de rang 100 et celle de rang 101, car l'effectif de l'échantillon est pair ; par convention la médiane est la moyenne arithmétique de ces deux termes. Les deux termes cherchés sont des 3 donc la médiane est 3.
   50% des familles de l'échantillon ont eu un nombre d'accidents inférieur ou égal à 3.