Corrigé contrôle continu de statistique

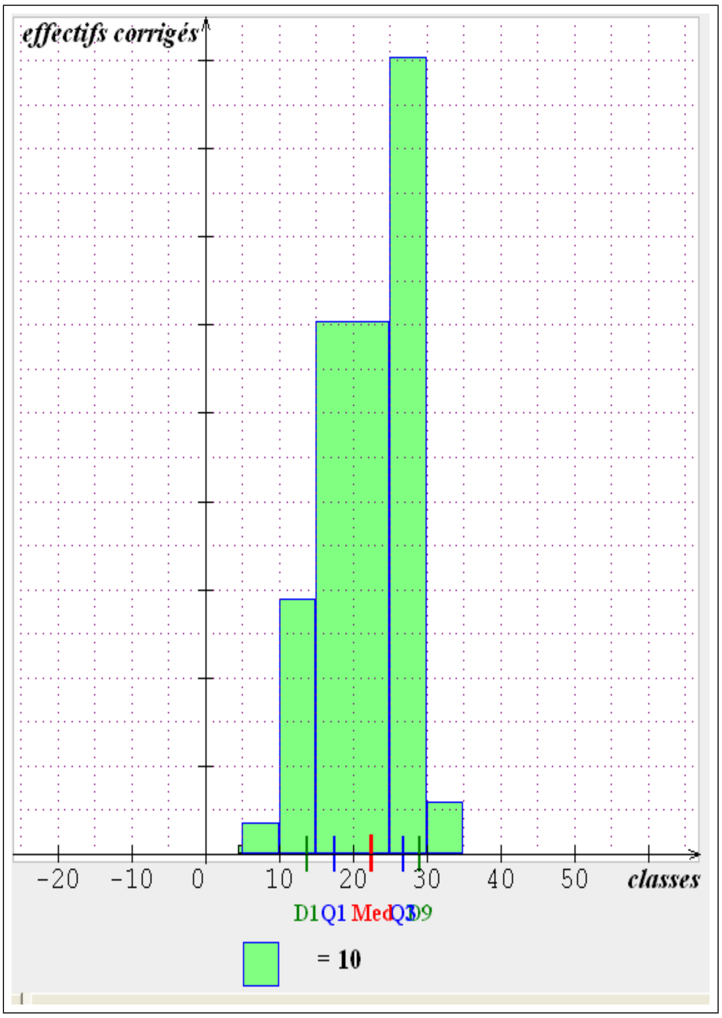
I EXERCICE-1 $(3pts)$

1. En France, le premier décile nous indique que 10% des français ont un revenu inférieur ou égal à 7075 dollars ; au Portugal, le revenu médian est de 7138 dollars, 50% des portugais ont un revenu inférieur ou égal à 7138 dollars. Il est à noter la grande différence entre les deux pays, le premier décile en France correspondant sensiblement au revenu médian au Portugal.

2. $V(x) = \frac{1}{n}\sum n_i (x_i - \overline{x})^2$ et la formule développée : $V(x) = \frac{1}{n}\sum n_i x_i^2 - \overline{x}^2$ (moyenne des carrés moins carré de la moyenne).

II EXERCICE-2

1. La représentation de cette série représentant les données relatives à un caractère quantitatif continu est un histogramme ; les classes étant d'amplitude inégale, on a utilisé la densité, $d_i = \frac{n_i}{A_i}$ et les effectifs corrigés $n_{icor} = 5d_i$, 5 étant l'amplitude minimale de classe.

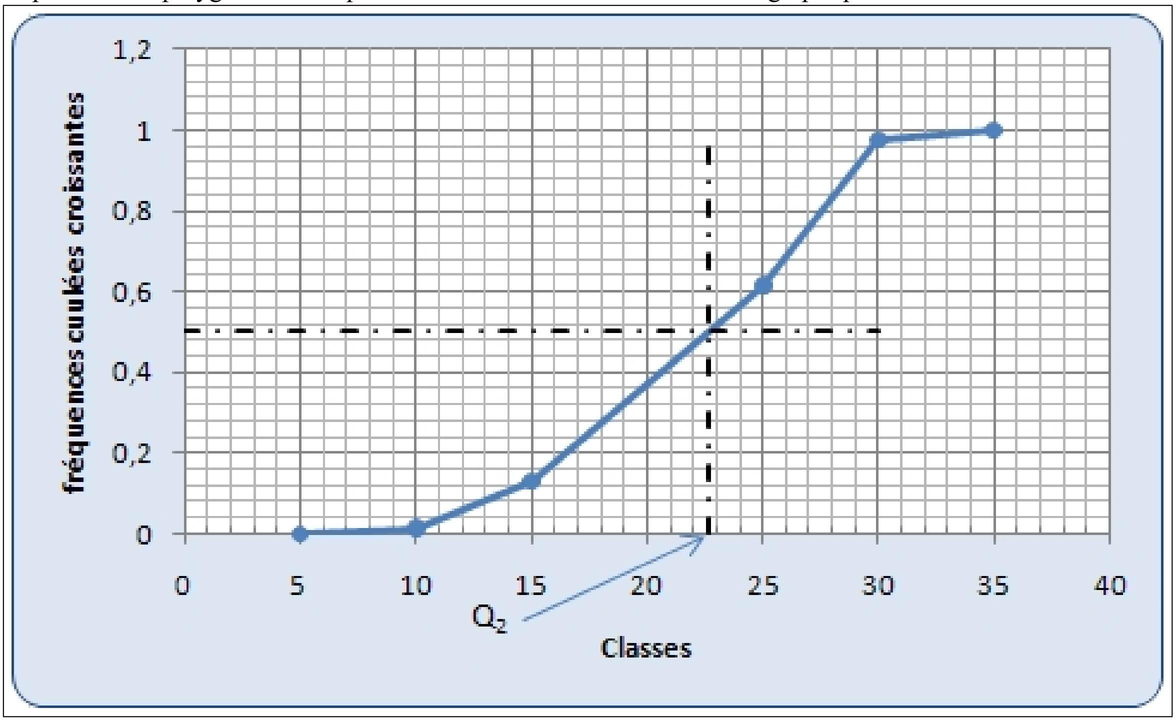


a_i	b_i	x_i	n_i	A_i	d_i	$n_i cor$	$n_i cc$	$f_i cc$	$n_i x_i$	$n_i x_i^2$
5	10	7.5	7	5	1.4	7	7	0.014	52.5	393.75
10	15	12.5	58	5	11.6	58	65	0.13	725	9062.5
15	25	20	242	10	24.2	121	307	0.614	4840	96800
25	30	27.5	181	5	36.2	181	488	0.976	4977.5	136881
30	35	32.5	12	5	2.4	12	500	1	390	12675
			500						10985	255813

La représentation de cette série représentant des données relatives à un caractère quantitatif continu est un histogramme ; les classes étant d'amplitudes inégales, on a utilisé la densité, $d_i = \dfrac{n_i}{A_i}$ et les effectifs corrigés $n_{icor} = 5d_i$, 5 étant l'amplitude minimale de classe.

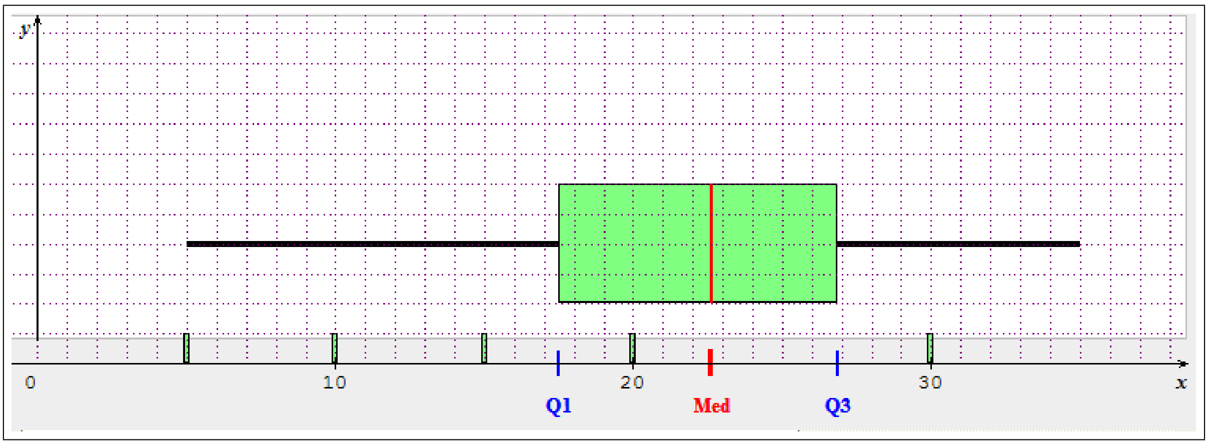
2. La classe modale est celle de plus grande densité, c'est-à dire la classe $[25; 30[$ et le mode est calculé en considérant les classes encadrant la classe modale, ce qui donne avec les notations du cours : $\begin{cases} x_1 = 25 \\ x_2 = 30 \end{cases}$, $\begin{cases} h = 181 \\ h_1 = 121 \text{ et } h_2 = 12 \end{cases}$ $\begin{cases} k_1 = h - h_1 = 60 \\ k_2 = h - h_2 = 169 \end{cases}$ et pour conclure :
$M_o = \dfrac{k_2 x_1 + k_1 x_2}{k_2 + k_1} = \dfrac{25 * 169 + 60 * 30}{229} \simeq 26.31$; comme prévu, le mode est plus petit que 27.5, le centre de la classe modale, car il est attiré par la classe de gauche, de densité plus importante.

3. Représenter le polygone des fréquences cumulées croissantes et donner graphiquement une estimation de la médiane.



Ce graphique permet d'estimer la médiane à environ 22.7, en prenant l'intersection du polygone des effectifs cumulés croissants avec la droite horizontale : $y = 0.5$.

4. On donne $Q_1 = 17.48$ et $Q_3 = 26.88$; pour la boîte à moustaches nous devons calculer $EIQ = Q_3 - Q_1 = 26.88 - 17.48 = 9.4$ et $1.5EIQ = 1.5 * 9.4 = 14.1$, taille maximale des moustaches ; par ailleurs calculons : $Q_1 - 5 = 17.48 - 5 = 12.48$ et $35 - Q_3 = 35 - 26.88 = 8.12$; aucune des moustaches n'exède $1.5\ EIQ$, il n'y pas de correction à effectuer.



5. Le tableau statistique permet de calculer $\overline{x} = \dfrac{1}{n}\sum n_i x_i = \dfrac{10985}{500} = 21.97$; $V(x) = \dfrac{255812.5}{500} - 21.97^2 \simeq 28.94$ et $\sigma(x) = \sqrt{V(x)} \simeq \sqrt{28.94} \simeq 5.38$

Corrigé contrôle continu de statistique

6. On a : $\overline{x} = 21.97$, $M_o \simeq 26.31$ et $M_e \simeq 22.7$, soit : $\overline{x} < M_e < M_o$, ce qui traduit un étalement à gauche, en cohérence avec la boîte à moustaches.
7. Caractère y
   a. On sait que : si $y = ax + b$, $\overline{y} = a\overline{x} + b$, $V(y) = a^2 V(x)$ et $\sigma(y) = |a|\,\sigma(x)$. Ici on a :
   $y = 0.85x$ donc on en déduit : $\overline{y} = 0.85\overline{x} = 0.85 * 21.97 = 18.68$ et $\sigma(y) = 0.85 * 5.12 = 4.35$

   b. $CV(y) = \dfrac{\sigma(y)}{\overline{y}} = \dfrac{0.85\sigma(x)}{0.85\overline{x}} = \dfrac{\sigma(x)}{\overline{x}} = \dfrac{5.38}{21.97} = 0.2449$ soit 24.49%. Les caractères x et y ont la même dispersion relative ; une variation en pourcentage ne modifie pas le coefficient de variation.

8. On estime maintenant que le temps moyen de livraison d'une pizza est de 22 mn avec un écart-type de 5 mn.
   a. L'inégalité de Bienaymé-Tchebychev nous donne en statistique une loi empirique : dans un intervalle centré autour de la moyenne, du type $[\overline{x} - k\sigma(x) ; \overline{x} + k\sigma(x)]$, il y a au moins $\left(1 - \dfrac{1}{k^2}\right) * 100\%$ d'observations. On doit calculer : $P(12 < T < 32)$; c'est un intervalle centré sur 22, la moyenne, avec une variation maximale de 10 autour de la moyenne, donc de deux écart-type ; pour $k = 2$, l'inégalité de Bienaymé-Tchebychev nous indique qu'il y a au moins $\left(1 - \dfrac{1}{2^2}\right) * 100 = 75\%$ des observations, donc que $P(12 < T < 32) \geq 0.75$.
   b. La loi normale
      i. On calcule cette probabilité en standardisant la variable aléatoire, c'est-à dire en se ramenant à la loi normale centrée réduite. On pose $Z = \frac{X-m}{\sigma} = \frac{X-22}{5}$, 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 : $P(12 < T < 32) = P\left(\frac{12-22}{5} \leq Z \leq \frac{32-22}{5}\right) = P(-2 \leq Z \leq 2) = 2F(2) - 1$, où F est la fonction de répartition de la loi normale centrée réduite, dont les valeurs sont lues dans la table ; on lit $F(2) = 0.97725$, ce qui donne : $P(12 < T < 32) \simeq 2 * 0.97725 - 1 = 0.9545$

      ii. On doit calculer : $P(T > 30) = 1 - P(T \leq 30) = 1 - P\left(Z \leq \frac{30-22}{5}\right) = 1 - F(1.6) \simeq 1 - 0.9452 = 0.054\,8$ soit 5.48%.
      Sur l'échantillon observé, la fréquence des pizzas livrées en plus de 30 min est de : $\dfrac{12}{500} = 0.024$ soit 2.4%.