PARTIE I : STATISTIQUES

Mardi 23 octobre 2018 - Durée : 2 heures

L'utilisation des calculatrices est autorisée. On arrondira les résultats numériques à 10^{-2} *près.*

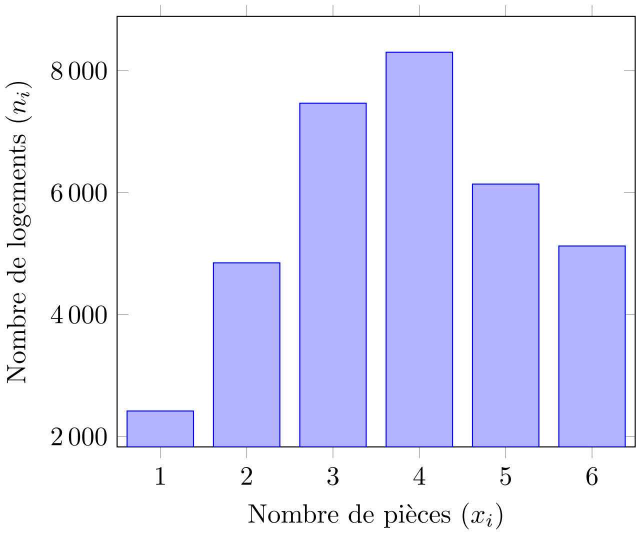
Exercice n° 1 - La distribution des logements en France en 2015 (6 points)

On considère la série suivante concernant le nombre de logements disponibles (en milliers) par nombre de pièces en France en 2015 (Source : Insee, RP2015 exploitation principale.).

Nombre de pièces (x_i)	1	2	3	4	5	6	Ensemble
Nombre de logements (n_i)	2 420	4 850	7 467	8 303	6 141	5 126	34 307

x_i	n_i	f_i	f_{icc}	f_{icd}	$x_i * n_i$
1	2 420	0.07	0.07	1.00	2 420.21
2	4 850	0.14	0.21	0.93	9 699.91
3	7 467	0.22	0.43	0.79	22 399.87
4	8 303	0.24	0.67	0.57	33 211.40
5	6 141	0.18	0.85	0.33	30 704.81
6	5 126	0.15	1.00	0.15	30 756.47
Ensemble	34 307	1.00			129 192.67

1. Caractériser le type de caractère étudié.
   Il s'agit d'une variable quantitative discrète car les modalités de la variable sont des entiers.
2. Représenter graphiquement cette série. En déduire le mode de la série. Donnez-en sa signification.



   Nous pouvons lire sur le graphique que la modalité 4 a la hauteur la plus élevée : il s'agit de la modalité ayant l'effectif le plus élevé. On trouve donc que $Mo = 4$ avec un effectif de 8 303.
3. Donner le tableau des fréquences (f_i), des fréquences cumulées croissantes (f_{icc}) et des fréquences cumulées décroissantes (f_{icd}).
   On calcule les fréquences qui rapportent chaque observation à l'effectif total ($f_i = n_i/n$). Ainsi, $f_1 = n_1/n = 2\ 420/34\ 307 = 0,07$. On calcule ensuite les fréquences cumulées croissantes : $f_{1cc} = f_1$ et $f_{2cc} = f_{1cc} + f_2 = 0,07 + 0,14 = 0,21$. On calcule également les fréquences cumulées décroissantes : $f_{1cd} = 1$ et $f_{2cd} = 1 - f_1 = 1 - 0,07 = 0,93$.

4. En déduire le pourcentage de logements d'au moins de trois pièces puis le pourcentage de logements de plus de cinq pièces.

   On en déduit que le parc immobilier comprend 43 % de logements d'au moins de trois pièces qui correspond à la fréquence cumulée croissante associée à la modalité 3. De même, le parc immobilier comprend 33 % de logements de plus de cinq pièces qui correspond à la fréquence cumulée décroissante associée à la modalité 5.

5. A l'aide du tableau des fréquences, en déduire la médiane de la série. Donnez-en sa signification.

   On cherche la médiane. Nous sommes en présence d'une variable discrète. La valeur médiane de la série est obtenue lorsque sa fréquence cumulée croissante est supérieure ou égale à 0.5 car $F(Me) = 0.5$. On trouve que $Me = 4$. Ainsi, 50 % des logements a un nombre de pièces inférieur ou égal à 4, 50 % des logements a un nombre de pièces supérieur ou égal à 4.

6. Calculer la moyenne arithmétique $\bar{x}$.

$$\bar{x} = \frac{x_i * n_i}{n} = \frac{129\ 192{,}67}{34\ 307} = 3,77$$

7. En déduire la forme de la distribution de la série.

   Nous comparons la médiane avec la mode pour caractériser la distribution de la série. Ici, on trouve que la médiane est égale au mode : la série est symétrique.

Exercice n$^{\text{o}}$ 2 - La population active française en 2016 (14 points)

On considère la série suivante concernant la population active (en milliers) par grandes tranches d'âge des individus pour l'ensemble des personnes actives de plus de 15 ans en France en 2016 (source : Insee, Enquête Emploi 2016).

Classes d'âge (x_i)	Population active (n_i)
15 à 29 ans	6 051,20
30 à 39 ans	7 074,50
40 à 49 ans	7 850,20
50 à 59 ans	7 016,40
60 ans à 69 ans	1 563,90
Ensemble	29 556,20

a_i	b_i	n_i	A_i	d_i	$n_i corr$	f_i	f_{icc}	c_i	$n_i * c_i$	c_i^2	$n_i * c_i^2$
15	29	6 051.2	14	432.23	3 890.1	0.20	0.20	22	133 126.4	484.00	2 928 780.8
30	39	7 074.5	9	786.06	7 074.5	0.24	0.44	34.5	244 070.3	1 190.25	8 420 423.6
40	49	7 850.2	9	872.24	7 850.2	0.27	0.71	44.5	349 333.9	1 980.25	15 545 358.6
50	59	7 016.4	9	779.60	7 016.4	0.24	0.95	54.5	382 393.8	2 970.25	20 840 462.1
60	69	1 563.9	9	173.77	1 563.9	0.05	1.00	64.5	100 871.6	4 160.25	6 506 215.0
Ensemble		29 556.2				1			1 209 795.9		54 241 240.1

1. Caractériser le type de caractère étudié.

   Il s'agit d'une variable quantitative continue.

2. Calculer la valeur modale (Mo) de la série. Donnez-en la signification.

   La classe modale est la classe présentant la densité la plus élevée : il s'agit de la classe [40, 49]. On pose donc $x_1 = 40$ et $x_2 = 49$. On définit les trois valeurs correspondant respectivement à la hauteur de la classe modale (h), la hauteur de la classe précédant la classe modale (h_1) et la hauteur de la classe suivant la classe modale (h_2). Soit $h = 7\ 850.2$; $h_1 = 7\ 074.5$ et $h_3 = 7\ 016.4$. On calcule la distance relative de la hauteur de la classe modale par rapport à la hauteur de deux autres classes : $k_1 = h - h_1 = 7\ 850{,}2 - 7\ 074{,}5 = 775.50$ et $k_2 = h - h_2 = 7\ 016{,}4 - 7\ 074{,}5 = 833.80$. On obtient donc $Mo = [(x_1 \times k_2) + (x_2 \times k_1)] / (k_1 + k_2) = 44,34$. Ainsi, on trouve que le nombre d'actifs sur le marché du travail est statistiquement le plus élevé à 44 ans.

3. Donner le tableau des fréquences (f_i) et des fréquences cumulées croissantes (f_{icc}).

   On calcule les fréquences qui rapportent chaque observation à l'effectif total ($f_i = n_i/n$). Ainsi, $f_1 = n_1/n = 6\ 051{,}2/29\ 556{,}2 = 0,20$. On calcule ensuite les fréquences cumulées croissantes : $f_{1cc} = f_1$ et $f_{2cc} = f_{1cc} + f_2 = 0,20 + 0,24 = 0,44$.

4. En déduire la valeur médiane (Me), le premier quartile (Q_1) et le troisième quartile (Q_3). Donnez la signification de chaque paramètre.

   On commence par calculer la médiane. La classe médiane est la classe dont la fréquence cumulée croissante est au moins égale à 0,50 : il s'agit de la classe [40, 49]. On définit les trois points $A(40, 0, 44); B(49; 0, 71)$ et $Me(Me, 0, 5)$. On trouve que

$$Me = x_A + \frac{(x_B - x_A) \times (0,5 - f(x_A))}{(f(x_B) - f(x_A))} = 40 + \frac{(49 - 40) \times (0,5 - 0,44)}{(0,71 - 0,44)} = 41,89$$

   On trouve que 50 % des actifs français ont moins de 42 ans, l'autre moitié ayant plus de 42 ans. Calculons ensuite Q_1. La classe de Q_1 est la classe dont la fréquence cumulée croissante est au moins égale à 0,25 : il s'agit de la classe [30, 39]. On définit les trois points $A(30, 0, 20); B(39; 0, 44)$ et $Q_1(Q_1, 0, 25)$. On trouve que

$$Q_1 = x_A + \frac{(x_B - x_A) \times (0,25 - f(x_A))}{(f(x_B) - f(x_A))} = 30 + \frac{(39 - 30) \times (0,25 - 0,20)}{(0,44 - 0,20)} = 31,70$$

   On trouve que 25 % des actifs français ont moins de 32 ans, les 75 % des actifs restants ayant plus de 32 ans. Calculons enfin Q_3. La classe de Q_3 est la classe dont la fréquence cumulée croissante est au moins égale à 0,75 : il s'agit de la classe [50, 59]. On définit les trois points $A(50, 0, 71); B(59; 0, 95)$ et $Q_3(Q_3, 0, 75)$. On trouve que

$$Q_3 = x_A + \frac{(x_B - x_A) \times (0,75 - f(x_A))}{(f(x_B) - f(x_A))} = 50 + \frac{(59 - 50) \times (0,75 - 0,71)}{(0,95 - 0,71)} = 51,53$$

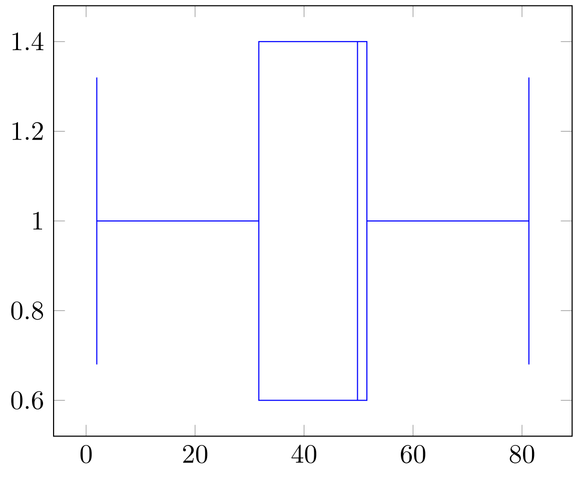
   On trouve que 75 % des actifs français ont moins de 51 ans, les 25 % des actifs restants ayant plus de 51 ans.

5. A partir des résultats obtenus, représenter la dispersion de la série à l'aide d'une boîte à moustaches. En déduire graphiquement la forme de la distribution. Ce résultat est-il confirmé après avoir calculé le coefficient de Yule C_Y ?

   Calculons les bornes des moustaches :

$$a_1 = Q_1 - [1,5 \times (Q_1 - Q_3)] = 31,70 - [1,5 \times (31,70 - 51,53)] = 1,96 \text{ et}$$
$$a_2 = Q_3 + [1,5 \times (Q_1 - Q_3)] = 51,53 + [1,5 \times (31,70 - 51,53)] = 81,27$$

   Après avoir ordonné dans l'ordre croissant a_1, Q_1, Me, Q_3 et a_3, on trouve que :



   La médiane étant plus proche de Q_3 que de Q_1, la distribution est étalée à gauche. Ce résultat est bien confirmé par le calcul du coefficient de Yule qui est négatif :

$$C_Y = \frac{(Q_1 - Me) + (Q_3 - Me)}{(Q_1 - Q_3)} = \frac{(31.70 - 49.81) + (51.53 - 49.81)}{(51.53 - 31.70)} = -0.03 < 0$$

6. Calculer la moyenne arithmétique $\bar{x}$, la variance $V(x)$ et l'écart-type σ de la série. En déduire le coefficient de variation (CV). On calcule la moyenne arithmétique $\bar{x}$:

$$\bar{x} = \frac{x_i * n_i}{n} = \frac{1\ 209\ 795{,}9}{29\ 556{,}2} = 40,93$$

On calcule ensuite la variance $V(x)$:

$$V(x) = \frac{n_i * c_i^2}{n} - (\bar{x})^2 = \frac{54\ 241\ 240{,}1}{29\ 556{,}2} - (40,93)^2 = 159,8$$

On en déduit l'écart-type σ :

$$\sigma = \sqrt{V(x)} = \sqrt{158,9} = 12,64$$

On en déduit le coefficient de variation :

$$CV(x) = \frac{\sigma}{\bar{x}} = \frac{12,64}{40,93} = 0,31$$

7. Supposons que la population active augmente de 5 %. Quelle est la nouvelle moyenne et le nouvel écart-type ?

   On suppose que toutes les catégories d'âge sont concernées par une hausse de 5 % des actifs ce qui correspond à une multiplication de 1,05. La moyenne ainsi que l'écart-type seront affectés par cette hausse :

$$\bar{x}' = \bar{x} \times 1,05 = 40,93 \times 1,05 = 42,98 \text{ et } \sigma(x)' = \sigma(x) \times 1,05 = 12,64 \times 1,05 = 13,27$$