

Mathématiques pour la gestion et statistiques

Corrigé du TP noté blanc du 22 novembre 2019

Temps disponible : 1h30

Seules les réponses aux questions à traiter sur papier apparaissent dans ce corrigé, pour les autres réponses se rapporter aux fichiers Excel

Exercice 1. On répondra sur papier aux questions 1, 3, 4, 5 (en partie), 6.

1. La population est l'ensemble des foyers français. Le caractère étudié est le budget consacré aux vacances. La variable est le montant consacré aux vacances de chaque foyer, exprimé en euros. Le caractère est quantitatif continu (discret est acceptable aussi).
3. Histogramme : hauteur sur la classe C_i égale à la fréquence relative f_i^r .

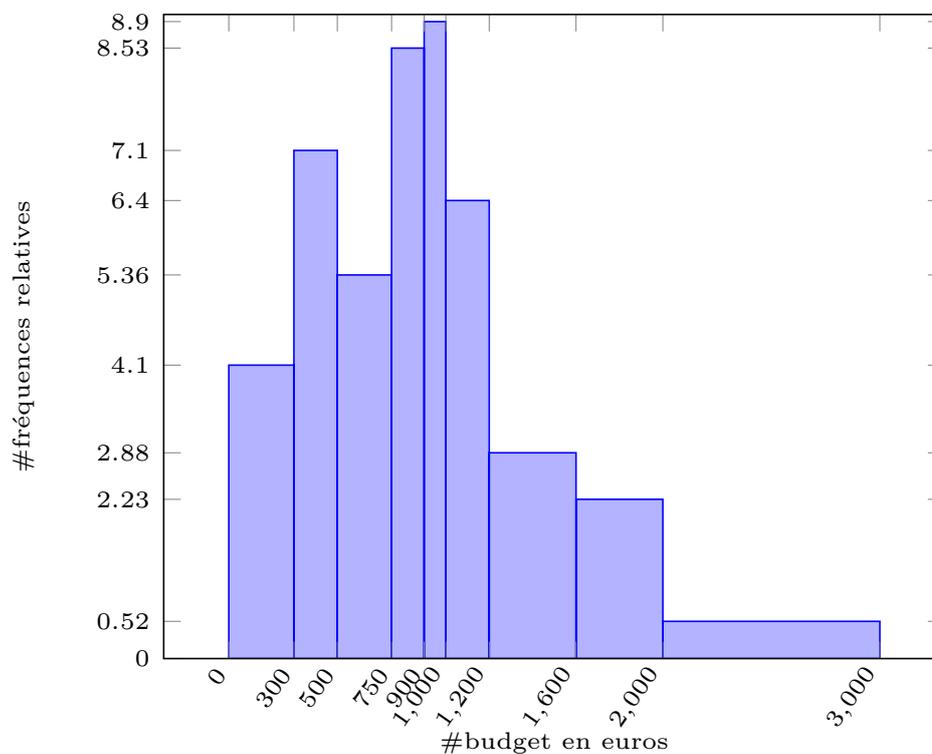


FIGURE 1. Histogramme représentatif du budget vacances des foyers français

4. La distribution est plutôt symétrique, les valeurs centrales entre 750 et 1000 euros ayant des fréquences relatives importantes (8,53 entre 750 et 900, 8,9 entre 900 et 1000 et

12 euros. L'échantillon présente tout de même une dissymétrie comme le montre la fréquence relative élevée (7,1) de la classe $[300, 500[$ (en euros).

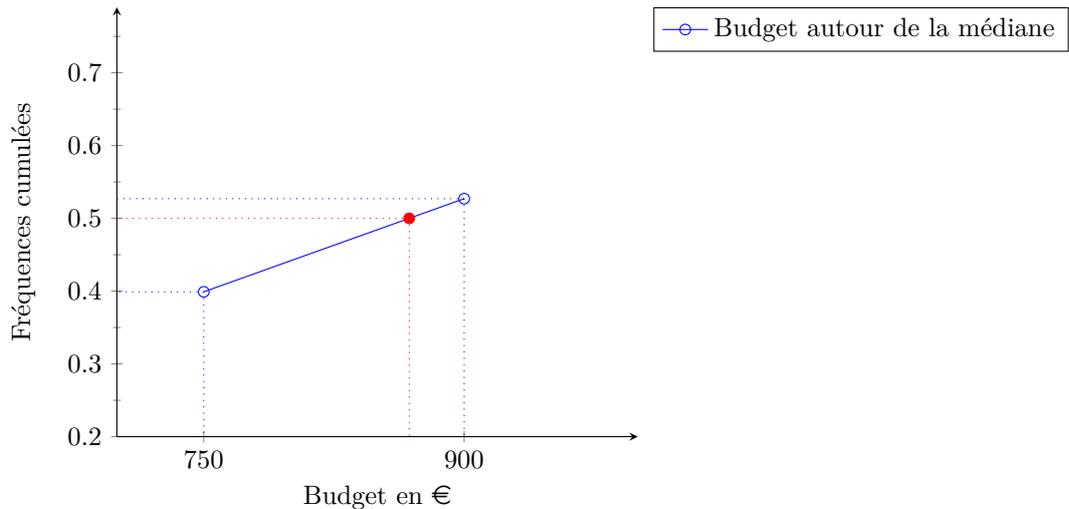
5. On reprend les valeurs trouvées dans Excel.

(a) La moyenne \bar{X} est la valeur :

$$\bar{X} = \sum_{i=1}^9 c_i f_i = (150 \times 0,123 + \dots + 2500 \times 0,052) \text{€} = 941,15 \text{€}$$

Ainsi, la tendance des foyers français est de consacrer environ 941,15 € aux vacances.

(b) La médiane est la valeur telle que la moitié des foyers français consacre aux vacances un budget supérieur et l'autre moitié inférieur à cette valeur. Ici, elle se trouve dans la classe numéro 3, à savoir $[750, 900[$ €. Pour la calculer, on regarde la fréquence cumulée dans la classe médiane $[750, 900[$ €, donc la figure suivante :



La médiane est l'abscisse du point rouge, dont l'ordonnée est 0,5. On note b_i la borne supérieure de la i -ième classe et F_i la fréquence cumulée de la i -ième classe ; on a que b_{i-1} est aussi la borne inférieure de la i -ième classe. Pour la classe médiane on a $i = 4$. On utilise le théorème de Thalès, qui affirme :

$$\frac{0,5 - F_3}{F_4 - F_3} = \frac{me - b_3}{b_3 - b_4},$$

donc :

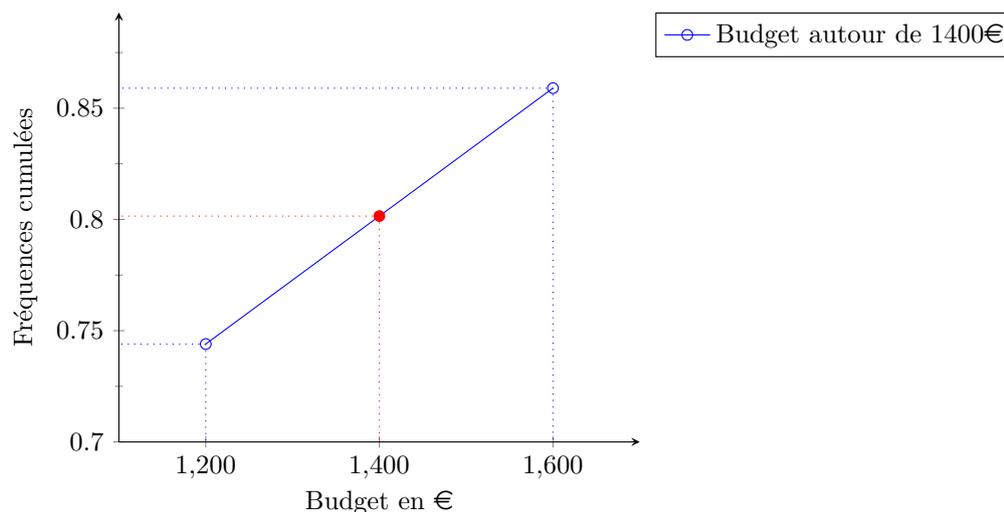
$$\begin{aligned} me &= \frac{0,5 - F_3}{F_4 - F_3} \times (b_3 - b_4) + b_3 = \\ &= \frac{0,5 - 0,399}{0,527 - 0,399} \times (900 - 750) + 750 = 868,35 \text{€}. \end{aligned}$$

(c) La classe modale est la numéro 2 à savoir $[300, 500[$ €. Cela veut dire que, parmi les tranches de budget considérées, $[300, 500[$ € est la plus fréquente.

6. On note $P(a \leq X \leq b)$ le pourcentage de foyers dont le budget vacances est compris entre les deux valeurs a et b . Pareil on pourra noter $P(X \leq b)$ et $P(X \geq a)$ le pourcentage de foyers dont le budget vacances est au plus b ou au moins a .

(a) On a $P(X \geq 1000) = 1 - P(X \leq 1000) = 1 - F_5 = 1 - 0,616 = 0,384$.

(b) La valeur 1400 est dans la classe numéro 7. On regarde la fréquence cumulée sur cette classe :



Le pourcentage $P(X \leq 1400)$ est l'ordonnée y du point rouge. On cherche donc y , pour le trouver on utilise de nouveau Thalès, qui affirme :

$$\frac{y - F_6}{F_7 - F_6} = \frac{1400 - b_6}{b_7 - b_6},$$

donc :

$$\begin{aligned} y &= \frac{1400 - b_6}{b_7 - b_6} \times (F_7 - F_6) + F_6 = \\ &= \frac{1400 - 1200}{1600 - 1200} \times (0,859 - 0,744) + 0,744 = 0,8015. \end{aligned}$$

Donc $P(X \leq 1400) = 0,8015$.

On peut aussi argumenter que $P(X \leq 1400)$ est la somme de $P(X \leq 1200)$ et du pourcentage correspondant à la moitié de la classe $[1200, 1600[€$, à savoir 0,115. La moitié de 0,115 est 0,0575 donc :

$$P(X \leq 1400) = P(X \leq 1200) + 0,0575 = 0,744 + 0,0575 = 0,8015.$$

Cette réponse est juste ; elle est plus rapide certes mais elle relève d'une observation ponctuelle plutôt que d'une méthode générale : c'est pourquoi elle n'a pas été proposée en priorité.

- (c) On a $P(100 \leq X \leq 1400) = P(X \leq 1400) - P(X \leq 1000)$. On vient de calculer $P(X \leq 1400) = 0,8015$ et on sait que $P(X \leq 1000) = 0,616$. Donc :

$$P(100 \leq X \leq 1400) = -0,616 = 0,1855.$$

Exercice 2. Voir le fichier Excel. Le coefficient de Cramer montre un lien très fort entre les deux variables.

Exercice 3. On se rapportera au fichier Excel. Le nuage de points montre une répartition autour d'une droite. Le coefficient de corrélation linéaire montre une corrélation forte entre les deux variables, ce qui justifie la possibilité d'utiliser un modèle de régression linéaire dans ce contexte.