

Durée de l'épreuve : 2h00 - Calculatrice autorisée, aucun document n'est permis

*** Partie statistique ***
14 points

Vous menez une étude sur les français, la littérature et l'intelligence. Vous interrogez 450 individus représentatifs de la population française de plus de 18 ans. Vous leur demandez combien de temps ils passent chaque semaine, en moyenne, à lire des romans. Vous leur faites ensuite passer un test de QI simplifié (coefficient intellectuel), qui est censé fournir un niveau d'intelligence. Un score de 80 indique un niveau d'intelligence maximal. Le tableau suivant présente les résultats croisés de ces deux variables.

$X_i \backslash Y_j$	[0 ; 20[10	[20 ; 40[30	[40 ; 60[50	[60 ; 80] 70	Total
[0h ; 2h[1	73	112	32	9	226
[2h ; 4h[3	32	44	34	21	131
[4h ; 6h[5	3	7	47	36	93
Total	108	163	113	66	450

X_i = temps moyen par semaine à lire des romans (en heures)
 Y_j = score de coefficient intellectuel (de 0 à 80)

- Déterminez la valeur de n_{32} , f_{31} , $n_{.2}$ et $f_{.1}$. (1 point)
- Calculez la moyenne, la variance et l'écart type des X_i et des Y_j . (1,5 point)
- Calculez la médiane (par interpolation linéaire) et le mode des Y_j . (1 point)

$$\text{Rappel : } Mo = L + a \times \left[\frac{e_s - e_s}{(e_s - e_s) + (e_s - e_s)} \right]$$
- Comparez la moyenne, le mode et la médiane des Y_j . Quelle information peut-on en tirer sur la symétrie de la distribution des QI ? Vérifiez cette interprétation en calculant les deux coefficients d'asymétrie de Pearson s et β_1 . Enfin, calculez et interprétez le coefficient d'aplatissement β_2 de Pearson des Y_j . (3 points)
- Déterminez le coefficient de corrélation et le coefficient de détermination entre X_i et Y_j . Le coefficient de corrélation est-il statistiquement significatif ? Expliquez ces résultats et tirez une conclusion de cette étude. (3 points)
- Déterminez l'équation de régression $Y = aX + b$. D'après cette équation, si quelqu'un vous dit qu'il passe en moyenne 5h30 par semaine à lire des romans, quel score de QI devrait-il atteindre ? Ce résultat est-il fiable ? (1,5 point)

- 7) Vous choisissez un individu au hasard parmi ces 450 interrogés. Quelle est la probabilité :
- qu'il lise des romans plus de 2h par semaine ?
 - qu'il ait obtenu un score de QI inférieur à 40 sur 80, sachant qu'il lit des romans moins de 4h par semaine ?
 - qu'il lise des romans plus de 2h par semaine, sachant qu'il a un score de QI supérieur à 40 sur 80 ?

(1,5 point)

8) Vous décidez de changer d'approche statistique en agrégeant les résultats. Vous croisez les deux variables binaires suivantes : le fait d'être ou non un gros lecteur de romans (plus de 4 heures par semaine *versus* moins de 4 heures par semaine) et le fait d'obtenir ou non un score de QI élevé (plus de 40 sur 80 *versus* moins de 40 sur 80). Vérifier l'existence d'un lien statistiquement significatif entre ces deux nouvelles variables.

(1,5 point)

*** Partie probabilités ***
6 points

Dans une salle d'attente, toute en longueur, 13 fauteuils forment une ligne. 13 candidats à un poste sont assis et attendent d'être reçus pour un entretien d'embauche. 6 sur 13 des candidats sont des femmes et 7 sont des hommes. Toutes ces personnes ont choisi une place au hasard parmi celles restantes, en fonction de leur heure d'arrivée. Le directeur ouvre la porte et annonce qu'il va recevoir les candidats un par un, du premier au dernier de la ligne formée.

- 9) Quelle est la probabilité qu'il y ait une alternance masculine et féminine des candidats reçus ?

(2 points)

- 10) Quelle est la probabilité que toutes les femmes soient assises les unes à côté des autres ?

(2 points)

- 11) Imaginons que cette configuration se reproduise à l'identique 5 jours de suite, chaque matin d'une semaine donnée (même salle d'attente, avec toujours 6 femmes et 7 hommes qui s'assoient aléatoirement). Quelle serait la probabilité que toutes les femmes soient assises les unes à côté des autres 2 jours sur les 5 ?

(2 points)