JG

Nom : ....

Prénom : ....

Numéro étudiant : ....

Groupe de TD : ....

Contrôle continu – Introduction à la statistique – octobre 2024 – Année 2024-25 – Licence 1 MIASHS

Le sujet comprend 4 pages, une feuille A4 manuscrite recto-verso autorisée, calculatrices autorisées.

Composer uniquement sur le sujet. Ne pas dégrafer les feuilles. Écrire au dos d'une feuille du sujet, si besoin, en l'indiquant dans la réponse du sujet.

Exercice 1.1 *On dispose des revenus mensuels (en euros) des ménages d'un échantillon de 2020 dans une étude sur les inégalités de revenus :*

23 valeurs.

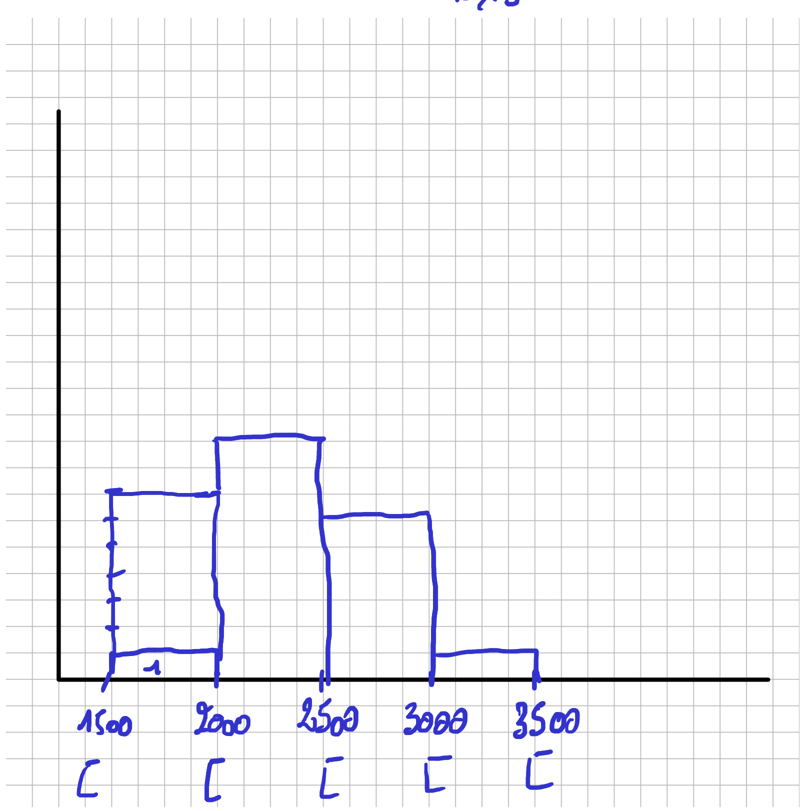
$q_{0,25}$

$$1500, 1500, 1600, 1600, 1800, 1800, 1900, 2000, 2100, 2100,$$

$$2100, 2200, 2200, 2200, 2400, 2400, 2500, 2500, 2700, 2800, 2800, 2900, 3200.$$

Me $\qquad q_{0,75}$

(i) Tracez l'histogramme des revenus mensuels



La série paraît :

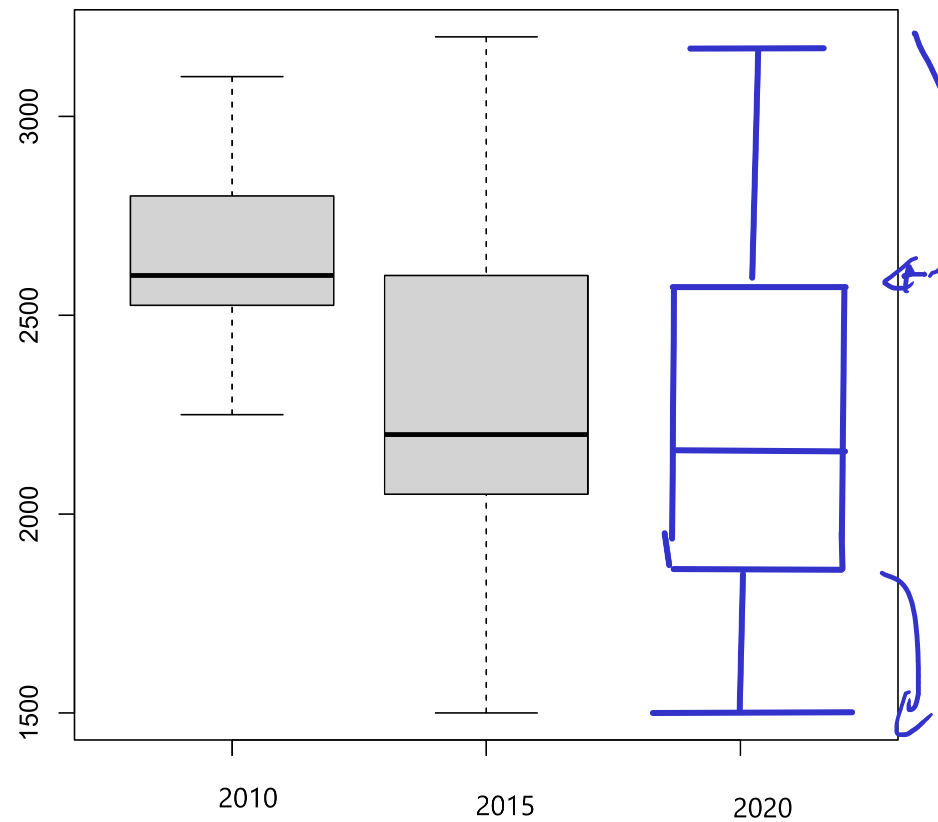
☒ centrée, ☐ décentrée vers les bas salaires, ☐ décentrée vers les hauts salaires.

(ii) Calculez la médiane des revenus et les quartiles des revenus.

- $Me =$ 2200
- $q_{0.25} =$ 1800
- $q_{0.75} =$ $\frac{2500+2700}{2} = 2600$

$IQ = 2600 - 1800$
$= 800$
$1,5\, IQ = 1200$

(iii) Tracer le boxplot des données sur le graphique suivant.



(iv) Les deux autres boxplots représentent les revenus des mêmes ménages à des dates différentes participant à la même étude. Comparez la répartition des revenus aux différentes dates.

- Les plus bas salaires baissent (Me et $q_{0,25}$)
- L'étendue des salaires augmente -

Exercice 1.2 *Le tableau ci-dessous présente la distribution de la consommation d'alcool Y en fonction des tranches d'âge X des assurés d'une assurance :*

$X \setminus Y$	*Aucun*	*Modéré*	*Élevé*
18-25	*15*	*30*	*10*
26-35	*20*	*25*	*15*
36-45	*25*	*20*	*5*
46-55	*30*	*15*	*8*
56+	*35*	*10*	*2*
	65	25	10

$n = 100$

(i) Définir la population étudiée, l'unité statistique, les caractères étudiés et leur nature.

- Population : des assurés
- Individu : un assuré
- Variable X :
  - caractère étudié : âge
  - nature : continu quantitatif (regroupement en classes)
- Variable Y :
  - caractère étudié : consommation d'alcool
  - nature : qualitatif (discret).

(ii) Comment s'appelle le tableau ci-dessus ?

tableau de contingences

(iii) (a) Déterminer la distribution en fréquences de consommation d'alcool.

Y	**Aucun**	**Modéré**	**Élevé**
fréquences	0,47	0,38	0,15

(b) Comment s'appelle cette distribution ?

distribution marginale

(iv) (a) Déterminer la distribution en fréquences du nombre de consommation d'alcool pour les assurés de plus 46 ans.

Y	**Aucun**	**Modéré**	**Élevé**
fréquences	0,65	0,25	0,1

(b) Comment s'appelle cette distribution ?

distribution conditionnelle (de la consommation sachant l'âge supérieur à 46)

(v) On regroupe à présent les informations de la façon suivante :

Y	**Aucun**	**Modéré**	**Élevé**
$[18;45]$	60	75	30
$[46+[$	65	25	10

$n = 265$

(a) Remplir le tableau des fréquences correspondant aux données ainsi que les marges en arrondissant au centième :

$X \setminus Y$	**Aucun**	**Modéré**	**Élevé**	Marge
$[18;45]$	0,23	0,28	0,11	0,62
$[46+[$	0,25	0,09	0,04	0,38
Marge	0,48	0,37	0,15	

(b) Quelle est la formule qui permet de calculer la statistique du Chi-deux :

☐ $\cancel{D_{\chi^2} = N \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{f_{jk} - f_{j\bullet} f_{\bullet k}}{f_{j\bullet} f_{\bullet k}}}$ au carré

☐ $D_{\chi^2} = N \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{(f_{jk} - f_{j\bullet} f_{\bullet k})^2}{f_{jk}}$

☐ $\cancel{D_{\chi^2} = \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{(f_{jk} - f_{j\bullet} f_{\bullet k})^2}{f_{jk}}}$

☐ $\cancel{D_{\chi^2} = \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{f_{jk} - f_{j\bullet} f_{\bullet k}}{f_{j\bullet} f_{\bullet k}}}$

☒ $D_{\chi^2} = N \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{(f_{jk} - f_{j\bullet} f_{\bullet k})^2}{f_{j\bullet} f_{\bullet k}}$ marginales

☐ $\cancel{D_{\chi^2} = \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{(f_{jk} - f_{j\bullet} f_{\bullet k})^2}{f_{j\bullet} f_{\bullet k}}}$

(c) Calculer la statistique D_{χ^2} au centième. On donnera le détail des calculs.

$$D_{\chi^2} = 265 \left(\frac{(0,23 - 0,62 \times 0,48)^2}{0,62 \times 0,48} + \frac{(0,28 - 0,62 \times 0,37)^2}{0,62 \times 0,37} + \dots + \dots + \dots + \frac{(0,04 - 0,38 \times 0,15)^2}{0,38 \times 0,15} \right)$$

$$= 20,65$$

(d) On peut supposer que $D_{\chi^2} \approx 20$. Calculer le *V de Cramer* et conclure.

$$V = \sqrt{\frac{20}{265 \times \min(c-1, l-1)}} = \sqrt{\frac{20}{265}} = 0,27$$

3   2

$\min(3-1, 2-1) = \min(2,1) = 1$

$\Rightarrow$ liaison moyenne entre âge et consommation d'alcool.