Contrôle du 9 mars 2017 – Durée 1h30
Statistiques, L1 de psychologie

Réponses

Veuillez rendre ce sujet avec votre copie

Nom :
Prénom :
Groupe de TD :

Le formulaire et la calculatrice sont autorisés. Vous êtes invités à répondre directement sur l'énoncé, mais si vous avez besoin de plus de place, vous pouvez détailler certaines questions sur un copie que vous rendrez avec l'énoncé. Le soin de la rédaction entrera en compte dans la notation mais dans les questions où des détails ne sont pas explicitement demandés, un résultat correct, donné sans détails de calcul sera accepté.

Ce sujet, composé de trois parties ***indépendantes****, porte sur le revenu des Français, et leur vote au second tour de l'élection présidentielle de 2012.*

1 Petit échantillon d'électeurs

On interroge neuf personnes choisies au hasard sur leur vote de mai 2012 et leur revenu. On obtient les données suivantes :

Mme PARIS	Mr GOMEZ	Mme LAUNAY	Mr FOURNIER	Mme POIRIER
23449 €/an a voté Sarkozy	31285 €/an a voté Hollande	6924 €/an a voté Sarkozy	16141 €/an a voté Hollande	13738 €/an a voté Hollande

Mr PREVOST	Mme FOUCHER	Mr LEVY	Mr COURTOIS
14193 €/an a voté Sarkozy	15775 €/an a voté Hollande	20139 €/an a voté Sarkozy	0 €/an a voté Hollande

1. Au sein de cet échantillon, combien d'individus ont voté pour Nicolas Sarkozy le 6 mai 2012 ?

Ils sont 4 (Mme PARIS, Mme LAUNAY, Mr PREVOST et Mr LEVY).

2. Quelle est la proportion de femmes au sein de l'échantillon ?

Il y a 4 femmes dans l'échantillon (Mme PARIS, Mme LAUNAY, Mme POIRIER et Mme FOUCHER), ce qui correspond à une proportion $\frac{4}{9} \simeq 0{,}444$.

3. Quel est, au sein de cet échantillon de neuf personnes, le revenu annuel moyen ?

C'est $\frac{23449+31285+6924+16141+13738+14193+15775+20139}{9} \simeq 15738$.

4. Au sein de cet échantillon, quels sont la moyenne et l'écart type du revenu des individus ayant voté pour François Hollande ? Et de ceux ayant voté pour Nicolas Sarkozy ?

Calcul pour les électeurs de F, Hollande :
moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{31285+16141+\cdots+0}{5} = \frac{76939}{5} = 15387{,}8$
$m(X^2) = \frac{\sum x_i^2}{n} = \frac{31285^2+16141^2+\cdots+0^2}{5} = \frac{1676866375}{5}$
$Var(X) = m(X^2) - m(X)^2 = \frac{1676866375}{5} - \left(\frac{76939}{5}\right)^2 = 98588886{,}16$
Écart-type : $s(X) = \sqrt{Var(X)} \simeq 9929{,}19$

Vote	Hollande	Sarkozy
Moyenne	15388	16176
Écart-type	9929	6288

Calcul pour les électeurs de N, Sarkozy :
moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{23449+6924+14193+20139}{4} = \frac{64705}{4} = 16176{,}25$
$m(X^2) = \frac{\sum x_i^2}{n} = \frac{23449^2+6924^2+14193^2+20139^2}{4} = \frac{1204817947}{4}$
$Var(X) = m(X^2) - m(X)^2 = \frac{1204817947}{4} - \left(\frac{64705}{4}\right)^2 \simeq 39533422{,}69$
Écart-type : $s(X) = \sqrt{Var(X)} \simeq 6287{,}56$

5. Au sein de cet échantillon, quel est la médiane du revenu des femmes ? Et la médiane du revenu des hommes ?

Pour les femmes, la médiane de 6924, 13738, 15775 et 23449 est $\frac{13738+15775}{2} = 14756{,}5$.

	Femmes	Hommes
Revenu médian	14756,5	16141

Pour les hommes, la médiane de 0, 14193, 16141, 20139 et 31285 est 16141.

2 Grand échantillon d'électeurs de François Hollande

On s'intéresse désormais à un plus grand échantillon, qui ne comprend que des électeurs de François Hollande, et dont on étudie le revenu annuel (exprimé en milliers d'euros) :

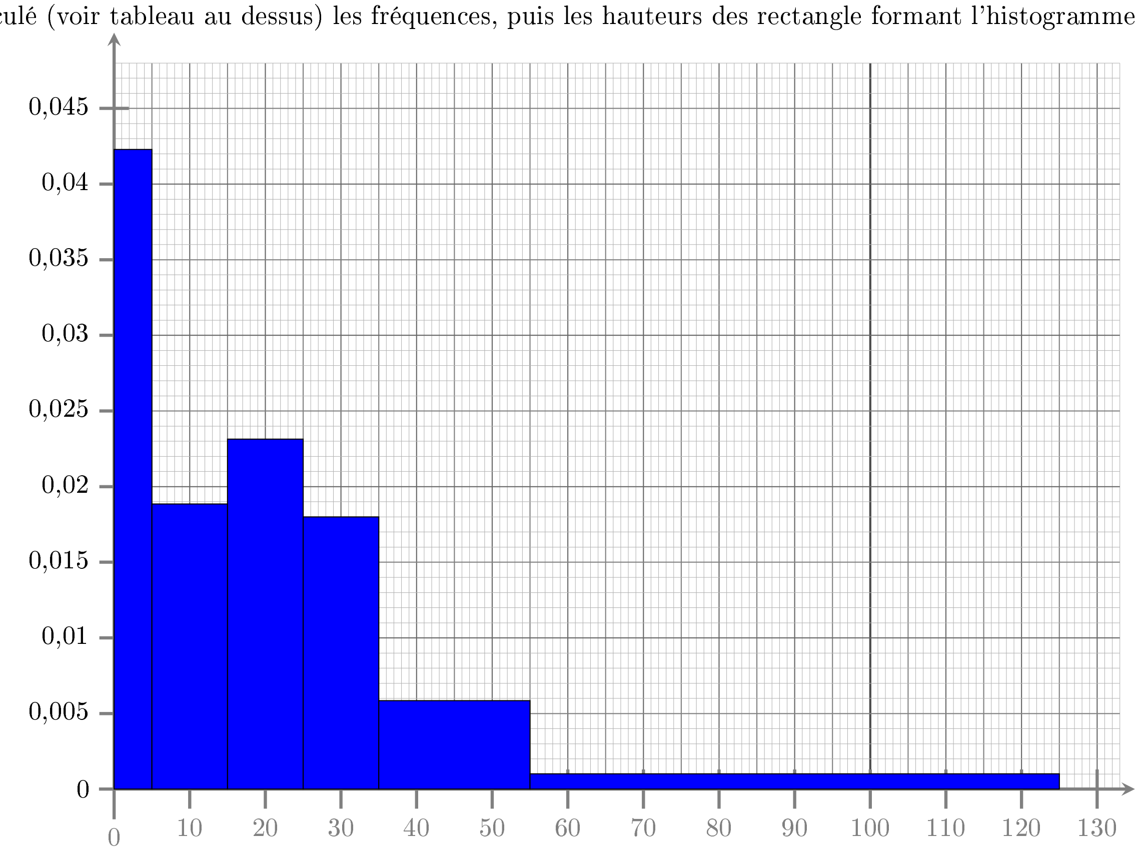
Revenu (en milliers d'€/an)	[0 ; 5[	[5 ; 15[	[15 ; 25[	[25 ; 35[	[35 ; 55[	[55 ; 125[
Effectif	74	66	81	63	41	25
Fréquence	0,2114	0,1886	0,2314	0,18	0,1171	0,0714
Hauteur	0,04229	0,01886	0,02314	0,018	0,00586	0,00102
Fréquence Cumulée	0,2114	0,4	0,6314	0,8114	0,9286	1

1. Combien y a t'il d'individus dans cet échantillon ?

Il y en a $74 + 66 + 81 + 63 + 41 + 25 = 350$.

2. Représenter ces données par un histogramme. *Vous préciserez quels calculs vous avez du effectuer (par exemple en remplissant des lignes du tableau en haut de page).*

On a calculé (voir tableau au dessus) les fréquences, puis les hauteurs des rectangle formant l'histogramme :



3. En détaillant vos calculs, déterminer la moyenne et l'écart-type du revenu des ces électeurs de François Hollande.

moyenne : $m(X) = \frac{\sum_i c_i\, n_i}{n} = \frac{2{,}5\times74+10\times66+20\times81+30\times63+45\times41+90\times25}{350} = \frac{8450}{350} \simeq 24{,}14$
$m(X^2) = \frac{\sum_i c_i^2\, n_i}{n} = \frac{2{,}5^2\times74+10^2\times66+20^2\times81+30^2\times63+45^2\times41+90^2\times25}{350} = \frac{381687{,}5}{350}$
$Var(X) = m(X^2) - m(X)^2 = \frac{381687{,}5}{350} - \left(\frac{8450}{350}\right)^2 \simeq 507{,}66$
Écart-type : $s(X) = \sqrt{Var(X)} \simeq 22{,}53$

4. Calculer les fréquences cumulées. *Les reporter dans le tableau en haut de la page*
5. En détaillant vos calculs, déterminer la médiane et les quartiles du revenu des ces électeurs de François Hollande.

médiane : $15 + \frac{25-15}{0{,}631-0{,}4}(0{,}5 - 0{,}4) \simeq 19{,}33$ (car dans la classe $[15\,;\,25[$).
premier quartile : $Q_1 \simeq 5 + \frac{15-5}{0{,}4-0{,}211}(0{,}25 - 0{,}211) \simeq 7{,}06$ (car dans la classe $[5\,;\,15[$).
troisième quartile : $Q_3 \simeq 25 + \frac{35-25}{0{,}811-0{,}631}(0{,}75 - 0{,}631) \simeq 31{,}61$ (car dans la classe $[25\,;\,35[$).

6. Quelle est environ la proportion, au sein de cet échantillon, d'électeurs gagnant plus que 41000€ par an ?

$F_X(41) = 0{,}8114 + \frac{0{,}9286-0{,}8114}{55-35}(41 - 35) \simeq 0{,}847$ car $41 \in [35\,;\,55[$
Donc la proportion demandée vaut $\mathbb{P}_r[X > 41] = 1 - \mathbb{P}_r[X \leqslant 41] = 1 - F_X(41) \simeq 1 - 0{,}847 = 0{,}153$.

7. Au sein de cet échantillon, combien y a t'il environ d'électeurs qui gagnent entre 41000€ et 55000€ par an ?

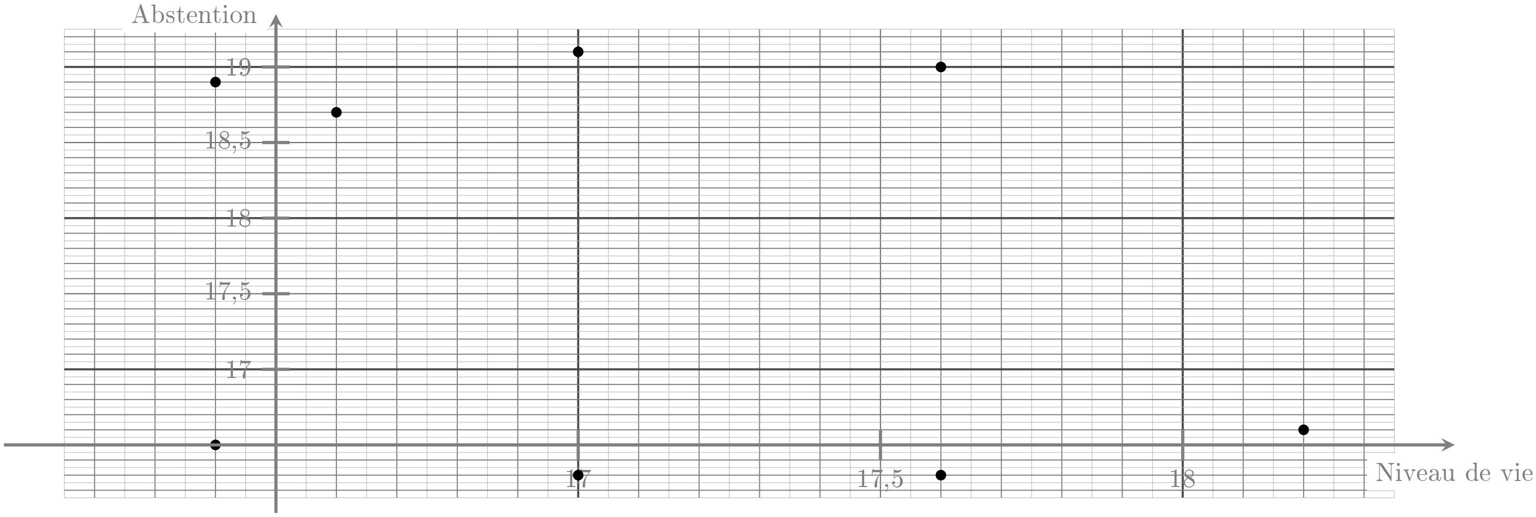
$\mathbb{P}_r[41 < X < 55] = \mathbb{P}_r[X < 55] - \mathbb{P}_r[X \leqslant 41] \simeq 0{,}9286 - 0{,}847 \simeq 0{,}082$.
Cela correspond donc à environ $0{,}082 \times 350 \simeq 29$ personnes.

3 Niveau de vie et taux d'abstention

Dans cette partie, on cherche à mettre en évidence un lien entre le revenu des électeurs et le fait de voter ou s'abstenir. Pour cela, on compare les taux d'abstention dans les différents départements de Bourgogne-Franche-Comté (chiffres du ministère de l'intérieur) au niveau de vie moyen par habitant (chiffres de l'INSEE) :

Département	21 (Côte d'Or)	25 (Doubs)	39 (Jura)	58 (Nièvre)	70 (Haute-Saône)	71 (Saône-et-Loire)	89 (Yonne)	90 (Territoire de Belfort)
Niveau de vie (X)	18,2	17,6	17	16,4	16,4	16,6	17	17,6
Taux d'abstention (Y)	16,6%	16,3%	16,3%	18,9%	16,5%	18,7%	19,1%	19%
rang X'	8	6,5	4,5	1,5	1,5	3	4,5	6,5
rang Y'	4	1,5	1,5	6	3	5	8	7

1. Représenter ces données sous la forme d'un nuage de points.



2. Déterminer le coefficient de corrélation linéaire entre le niveau de vie et le taux d'abstention.

On calcule d'abord les moyennes, écart-types et covariance :
moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{18{,}2+17{,}6+\cdots+17{,}6}{8} = \frac{136{,}8}{8} = 17{,}1$ $\quad m(X^2) = \frac{\sum x_i^2}{n} = \frac{18{,}2^2+17{,}6^2+\cdots+17{,}6^2}{8} = \frac{2342{,}24}{8}$
$Var(X) = m(X^2) - m(X)^2 = \frac{2342{,}24}{8} - \left(\frac{136{,}8}{8}\right)^2 \simeq 0{,}37$
$m(Y) = \frac{\sum y_i}{N} = \frac{141{,}4}{8} = 17{,}675$ $\quad m(X\,Y) = \frac{\sum x_i\,y_i}{N} = \frac{2416{,}18}{8} \simeq 302{,}022$
$Cov(X;Y) = m(X\,Y) - m(X)m(Y) = \frac{2416{,}18}{8} - \frac{136{,}8}{8}\frac{141{,}4}{8} \simeq -0{,}22$
$Var(Y) = m(Y^2) - m(Y)^2 = \frac{2511{,}9}{8} - \left(\frac{141{,}4}{8}\right)^2 \simeq 1{,}58$
d'où le coefficient de corrélation linéaire $r = \frac{\mathrm{Cov}(X;Y)}{\mathrm{s}(X)\mathrm{s}(Y)} \simeq \frac{-0{,}22}{\sqrt{0{,}37}\sqrt{1{,}58}} \simeq -0{,}288$.

3. Déterminer aussi le coefficient de corrélation des rangs de Spearman.

On a calculé les rangs (reporté dans le tableau en haut de page),
le coefficient de corrélation des rangs de Spearman est donc $1 - \left(6\,\frac{(8-4)^2+(6{,}5-1{,}5)^2+\cdots+(6{,}5-7)^2}{8\,(8^2-1)}\right) \simeq -0{,}06$

4. Si l'on voulait estimer le taux d'abstention dans un département où le niveau de vie est 16,7, quelle droite pourrait-on utiliser ? Calculer l'équation de cette droite, et commenter la pertinence de son utilisation.

On utiliserait la droite $D_{Y|X}$, mais ça n'aurait pas une grande pertinence car les variables sont très peu corrélées.
Pour son équation on pose $a = \frac{Cov(X,Y)}{Var(X)} \simeq \frac{-0{,}22}{0{,}37} \simeq -0{,}595$ et $b = m(Y) - a\,m(X) \simeq 17{,}675 + 0{,}595 \times 17{,}1 \simeq 27{,}849$
D'où l'équation de la droite $D_{Y|X}$: $Y = -0{,}595\,X + 27{,}849$