DEUST GNM - Corrigé Examen 2008

Tous les résultats sont arrondis à 2 décimales.

1. La population étudiée est celle des $N = 31$ cerisiers. Les variables `Diamètre` (X) et `Hauteur` (Y) sont quantitatives continues.

2. Pour le diamètre :

$$\begin{aligned}\bar{X} =& \frac{8.3 + 8.6 + \cdots + 20.6}{31} = \frac{410.70}{31} \simeq 13.25 \\ \mathrm{Var}(X) =& \frac{8.3^2 + 8.6^2 + \cdots + 20.6^2}{31} - 13.25^2 \simeq 9.49 \\ s_X =& \sqrt{\mathrm{Var}(X)} \simeq 3.08.\end{aligned}$$

De même, on trouve : $\bar{Y} = 2356/31 = 76$, $\mathrm{Var}(Y) \simeq 39.29$ et $s_Y \simeq 6.27$.

3. On commence par calculer la covariance de X et Y :

$$\mathrm{Cov}(X, Y) = \frac{8.3 \times 70 + 8.6 \times 65 + \cdots = 20.6 \times 87}{31} - 13.25 \times 76 \simeq 9.93.$$

On en déduit :

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{s_x s_Y} \simeq 0.51.$$

Donc X et Y varient dans le même sens et leur liaison est relativement forte.

4. On obtient :

	[8, 11]	]11, 12]	]12, 14]	]14, 17]	]17, 21]
[60, 70]	4	1	1	0	0
]70, 80]	2	6	3	4	3
]80, 90]	2	0	2	0	3

5,11, 12. On en déduit pour X : (on remarque que les proportions cumulées somment à 99.99% à cause des erreurs d'arrondis)

X	[8, 11]	]11, 12]	]12, 14]	]14, 17]	]17, 21]
Effectifs	8	7	6	4	6
Centres	9.5	11.5	13	15.5	19
Proportions (%)	25.81	22.58	19.35	12.90	19.35
Prop. cum.	25.81	48.39	67.74	80.64	99.99

Pour Y :

Y	[60, 70]	]70, 80]	]80, 90]
Effectifs	6	18	7

6. A partir du tableau de la question **5** :

$$\bar{X} = \frac{9.5 \times 8 + 11.5 \times 7 + \cdots + 19 \times 6}{31} \simeq 13.24$$

On obtient un résultat voisin (au centième près !) de celui de la question **2**, mais pas égal : on a perdu de l'information en résumant les données individuelles en classes.

$$\mathrm{Var}(X) = \frac{9.5^2 \times 8 + 11.5^2 \times 7 + \cdots + 19^2 \times 6}{31} - 13.24^2 \simeq 11.44.$$

On en déduit $s_X \simeq 3.38$. Là encore, on constate une erreur qui vient du fait qu'on a résumé nos données en tableau.

7. Conditionnellement aux classes de `Hauteur` [60, 70] et]80, 90] : (Dans la distribution des proportions de $X|Y \in [60, 70]$, on trouve un total de 100.01 à cause des erreurs d'arrondis).

$X\|Y \in [60, 70]$	[8, 11]	]11, 12]	]12, 14]	]14, 17]	]17, 21]	Total
Effectifs	4	1	1	0	0	6
Proportions (%)	66.67	16.67	16.67	0	0	100.01
Amplitude	3	1	2	3	4	
Densité (Prop/Ampli)	22.22	16.67	8.34	0	0	
$X\|Y \in [80, 90]$	[8, 11]	]11, 12]	]12, 14]	]14, 17]	]17, 21]	Total
Effectifs	2	0	2	0	3	7
Proportions (%)	28.57	0	28.57	0	42.86	100
Amplitude	3	1	2	3	4	
Densité (Prop/Ampli)	0.10	0	0.14	0	0.11	

8. On fait un histogramme. C'est l'aire qui est proportionnelle à la proportion, pas la hauteur (puisque les classes sont d'amplitudes différentes).

9. Les deux distributions conditionnelles sont très différentes : celle de $X|Y \in [60, 70]$ charge plutôt les petites valeurs, tandis que celle de $X|Y \in [80, 90]$ charge plutôt les grandes valeurs. Ceci montre que les variables X et Y sont liées et non indépendantes. Ce résultat est en accord avec la corrélation de 51% qu'on a trouvée à la question **3**.

10. Soit p la proportion d'arbres de diamètre plus petit ou égal à 11.5 inches.

$$p = 25.81 + \frac{1}{2} \times 22.58 \simeq 37.10\%.$$

11. C'est un histogramme qu'il faut faire.

12. Voir question **5**.

13. La médiane Q_2 appartient à la classe]12, 14]. On a :

$$50 = F(a) + \frac{F(b) - F(a)}{b - a} \times (Q_2 - a)$$

avec : $a = 12$, $b = 14$, $F(a) = 48.39$, $F(b) = 67.74$. D'où $Q2 \simeq 12.17$: 50% des arbres ont un diamètre inférieur à 12.17 inches.

14. En mètres :

$$\begin{aligned}\bar{X} =& 13.25 \times 0.02540 = 0.34 \text{ mètres}\\ \mathrm{Var}(X) =& 9.49 \times 0.02540^2 = 0.0061 \text{ mètres}^2\\ s_X =& 3.08 \times 0.02540 = 0.08 \text{ mètres.}\end{aligned}$$