Statistiques Descriptives

Devoir Surveillé 1

Durée : une heure.
Les calculatrices sont autorisées ; tout autre matériel électronique (comme le téléphone portable) est interdit.

Exercice 1. (10 pts)

Table 1 donne la répartition d'un échantillon de quarante personnes par classe d'âge et de masse.

		masse (y)		
		$[50, 70[$	$[70, 90[$	$[90, 100]$
âge (x)	$[20, 30[$	2	8	2
	$[30, 40[$	5	10	5
	$[40, 60]$	2	5	1

Table 1: *Tableau de contingence de l'âge mesuré en années et de la masse mesurée en kg d'un échantillon de quarante personnes*

Partie A. Une seule des trois réponses proposées à chacun des cinq items suivants est exacte ; entourez sans justification la réponse qui vous semble correcte et rayez les autres. Une réponse juste apporte un point ; une réponse fausse enlève 0,5 points ; une absence de réponse n'apporte ni n'enlève de points.

1. L'âge modal[a] arrondi à l'unité est : (a) 32 (b) 33 (c) 34.

   L'âge modal déterminé par la méthode des diagonales est : $30 + 10 \times 8/(8 + 16) \approx 33,3$; la bonne réponse est (b).

2. L'âge médian[b] est : (a) 34 (b) 35 (c) 36.

   L'âge médian déterminé par interpolation linéaire est : $30 + (10/0,5) \times 0,2 = 34$; la bonne réponse est (a).

3. Le neuvième décile de l'âge est : (a) 45 (b) 50 (c) 55.

   Le neuvième décile est 50, le milieu entre 40 (huitième décile) et 60 (dixième décile) ; la bonne réponse est (b).

4. Le quotient $n_{3,2}/n_{\bullet,2}$ est égal à : (a) 5/20 (b) 5/23 (c) 5/40.

   $n_{3,2}/n_{\bullet,2} = 5/(5 + 10 + 8)$; la bonne réponse est (b).

5. Ce quotient représente la proportion de personnes

   (a) entre 40 et 60 ans parmi les 70-90 kg
   (b) entre 90 et 100 kg parmi les 30-40 ans
   (c) ayant antre 40 et 60 ans et pesant de 70 à 90 kg

   $n_{3,2}/n_{\bullet,2}$ représente la proportion de personnes âgées de 40 à 60 ans parmi les 70-90 kg ; la bonne réponse est (a).

Partie B.

âge	$[20, 30[$	$[30, 40[$	$[40, 60]$
masse moyenne	$\bar{y}_1 \approx 79,2$	$\bar{y}_2 \approx 78,7$	$\bar{y}_3 \approx 76,9$

Table 2: *Masses moyennes conditionnelles à l'âge*

1. Déterminez la masse moyenne dans chaque classe d'âge.

   La masse moyenne des 20-30 ans est $\bar{y}_1 = \left(\sum_{j=1}^{3} n_{1,j} y_j\right)/n_{1,\bullet} = (2\times 60+8\times 80+2\times 95)/12 \approx 79,2$. On détermine ainsi les masses moyennes conditionnelles données par Table 2.

2. Déterminez la distribution marginale de la masse.

   La distribution marginale des effectifs, relative à la masse, est donnée par Table 3.

masse	$[50, 70[$	$[70, 90[$	$[90, 100]$
effectif	$n_{\bullet,1} = 9$	$n_{\bullet,2} = 23$	$n_{\bullet,3} = 8$

Table 3: *Distribution marginale des effectifs, relative à la masse*

3. Calculez de deux manières différentes la masse moyenne.

   La masse moyenne est la moyenne des masses de Table 3 : $\bar{\bar{y}} = \left(\sum_{j=1}^{3} n_{\bullet,j} \times y_j\right)/n = (9\times 60+23\times 80+8\times 95)/40 = 78,5$. De façon équivalente, on peut calculer la moyenne pondérée des moyennes conditionnelles de Table 2 : $\bar{\bar{y}} = \left(\sum_{i=1}^{3} n_{i,\bullet} \times \bar{y}_i\right)/n \approx (12\times 79,2+20\times 78,7+8\times 76,9)/40 \approx 78,5$.

4. Calculez de deux manières différentes la variance des masses.

   La variance de la masse peut être obtenue directement grâce à Table 3 : $\mathrm{var}(y) = \left(\sum_{j=1}^{3} n_{\bullet,j} \times y_j^2\right)/n - \bar{\bar{y}}^2 = (9\times 60^2+23\times 80^2+8\times 95^2)/40 - 78,5^2 = 132,75$.

   On peut aussi conditionner par l'âge. Pour les 20-30 ans la variance de la masse est : $v_1 = \left(\sum_{j=1}^{3} n_{1,j} \times y_j^2\right)/n_{1,\bullet} - \bar{y}_1^2 \approx (2\times 60^2+8\times 80^2+2\times 95^2)/12 - 79,2^2 \approx 103,5$; les autres variances conditionelles sont données par Table 4.

âge	$[20, 30[$	$[30, 40[$	$[40, 60]$
variance de la masse	$v_1 \approx 103,5$	$v_2 \approx 154,7$	$v_3 \approx 118,4$

Table 4: *Variance de la masse conditionnellement à l'âge*

   La variance de la masse, $\mathrm{var}(y)$, est la somme de deux termes :

   - la moyenne des variances conditionnelles de Table 4 : $\left(\sum_{i=1}^{3} n_{i,\bullet} \times v_i\right)/n \approx (12\times 103,5+20\times 154,7+8\times 118,4)/40 \approx 132,06$
   - la variance des moyennes conditionnelles de Table 2 : $\left(\sum_{i=1}^{3} n_{i,\bullet}\left(\bar{y}_i^2 - \bar{\bar{y}}\right)^2\right)/n \approx (12\times(79,2-78,5)^2+20\times(78,7-78,5)^2+8\times(76,9-78,5)^2)/40 \approx 0,69$.

Exercice 2. (10 pts)

Table 5 indique la taille (x) et la masse (y) d'un échantillon de quatre personnes.

taille (cm)	160	150	160	180
masse (kg)	60	70	80	80

Table 5: *Taille et masse d'un échantillon de quatre personnes*

1. Déterminez les coordonnées du point moyen.

   Le point moyen G a pour coordonnées $\bar{x} = (\sum_{i=1}^{4} x_i)/4 = 162,5$ et $\bar{y} = (\sum_{i=1}^{4} y_i)/4 = 72,5$

A. Lourme, Faculté d'économie, gestion & AES, Université Montesquieu - Bordeaux IV. http://alexandrelourme.free.fr

[a]déterminé par la méthode des diagonales

[b]obtenu par interpolation linéaire

2. Déterminez la variance de la taille et celle de la masse.

   La variance de la taille est : $\operatorname{var}(x) = \left(\sum_{i=1}^{4} x_i^2\right)/4 - \bar{x}^2 = 118,75$; celle de la masse est : $\operatorname{var}(y) = \left(\sum_{i=1}^{4} y_i^2\right)/4 - \bar{y}^2 = 68,75$.

3. Quelle serait la variance de la taille si les tailles étaient données en mètres ?

   La variance est multipliée par 10^{-4} lorsqu'on convertit les tailles de centimètres en mètres.

4. Quelle est la covariance des deux variables ?

   La covariance des deux variables est : $\operatorname{cov}(x,y) = \left(\sum_{i=1}^{4} x_i y_i\right)/4 - \bar{x}\bar{y} = 43,75$.

5. Quel est leur coefficient de corrélation linéaire ?

   Leur coefficient de corrélation linéaire est : $\rho = \operatorname{cov}(x,y)/\sqrt{\operatorname{var}(x) \times \operatorname{var}(y)} \approx 0,4842$.

6. Déterminez l'équation de la droite d'ajustement de la masse en la taille.

   La droite d'ajustement de la masse (y) en la taille (x) a pour équation : $\mathscr{D} : y = 0,37 \times x + 12,63$. En effet $\alpha = \operatorname{cov}(x,y)/\operatorname{var}(x) \approx 0,37$ et $\bar{y} - \alpha\bar{x} \approx 12,63$.

7. Calculez et interprétez le coefficient de détermination.

   Puisque la régression est linéaire, le coefficient de détermination est le carré du coefficient de corrélation linéaire ; il vaut donc environ $0,23$: l'ajustement linéaire de la masse par la taille explique 23% de la variabilité de la masse.
