

CORRIGÉ

Exercice 5 : Croisement de deux variables quantitatives

Exercice 1 Rappel de statistique univariée

TABLE 1 – POIDS ET QUANTITÉ DE CHOCOLAT CONSOMMÉE POUR 200 ÉTUDIANTS D’AES AU SEIN DE L’UNIVERSITÉ DE BORDEAUX (MESURANT TOUS 1,75M).

		Poids (en kg)				Total
		[55 ; 65[[65 ; 75[[75 ; 95[[95 ; 115[
Quantité	[0 ; 50[27	24	6	3	60
de chocolat consommée	[50 ; 150[30	30	38	2	100
(en grammes par semaine)	[150 ; 450[1	2	14	23	40
	Total	58	56	58	28	200

Source : Inspiré par le Département de Sciences Politiques, Université d’Amsterdam.

1. Caractériser la série statistique X (quantité de chocolat consommée), et représenter graphiquement sa distribution.

Pour caractériser une distribution statistique (série statistique), il est nécessaire de présenter :

- La population : 200 étudiants d’AES au sein de l’Université de Bordeaux.
- Par conséquent, l’unité statistique est un étudiant d’AES de l’Université de Bordeaux.
- Le caractère étudié : nous étudions ces étudiants selon la quantité hebdomadaire de chocolat qu’ils consomment.

Sachant que nous étudions une quantité (mesurée en grammes), il s’agit bien d’un caractère quantitatif (quantifiable, dénombrable) et continu dans la mesure où celui-ci est susceptible de prendre une infinité de valeurs possibles dans l’intervalle de définition $[0; 150[$. Toutefois, et afin de simplifier cette distribution, toutes les valeurs ont été regroupées en classes.

- Nous observons enfin, que l’amplitude des ces dernières varie, il faudra en tenir compte par la suite. La représentation graphique d’un caractère quantitatif continu est l’histogramme.
- Attention : le calcul des densités est nécessaire en présence de classes d’amplitudes différentes.

- $d_i = \frac{n_i}{a_i}$ en sachant que $a_i = b - a$

2. Quelle est la proportion des étudiants consommant moins de 80 grammes de chocolat par semaine ? Quelle est la quantité de chocolat hebdomadaire consommée par les 10% d’étudiants d’AES qui en consomment le moins ?

On va utiliser l’interpolation linéaire afin de déterminer ces valeurs :

- Nous avons : $\frac{80-50}{\alpha-0,30} = \frac{150-50}{0,80-0,30}$

Ainsi : $30 = 200(\alpha - 0,30) = 200\alpha - 60$

Enfin : $90/200 = \alpha = \mathbf{0,45}$

Donc : 45% des étudiants d’AES observés consomment strictement moins de 80 grammes de chocolat par semaine.

- Nous avons : $\frac{\beta-0}{0,10-0} = \frac{50-0}{0,30-0}$

Enfin : $\beta = 0,10\left(\frac{50}{0,30}\right) = \mathbf{1,5}$

Donc : les 10% d'étudiants qui consomment le moins de chocolat ingèrent strictement moins de 1,5 grammes de chocolat par semaine.

3. Quelle est la consommation moyenne et médiane de chocolat des étudiants d'AES ?

— La moyenne : $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i n_i$

$$\text{Ainsi : } \bar{x} = \frac{23500}{200}$$

$$\bar{x} = \mathbf{117,5}$$

En moyenne les étudiants d'AES observés consomment 117,5 grammes de chocolat par semaine.

— La médiane (par interpolation linéaire) : $\frac{\gamma-50}{0,50-0,30} = \frac{150-50}{0,80-0,30}$

$$\text{Ainsi : } \gamma = 200 \cdot 0,20 + 50 = \mathbf{90}$$

La moitié des étudiants d'AES observés consomme strictement moins de 90 grammes de chocolat par semaine, lorsque l'autre moitié consomme au moins 90 grammes de chocolat par semaine.

4. La consommation de chocolat est-elle plus stable chez les étudiants d'AES ou en Sociologie ? Sachant que dans cet autre cursus le nombre d'étudiants considérés est de 150, la consommation hebdomadaire moyenne de chocolat est de 124g par étudiant et la variance est de 5000 grammes au carré.

Analyser la stabilité de la consommation de chocolat entre les deux cursus revient à comparer la dispersion au sein des deux distributions.

— Nous allons dans un premier temps déterminer la variance et l'écart-type de notre distribution :

$$V(x) = \frac{1}{n} \sum_{i=1}^n n_i (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n n_i (x_i)^2 - \bar{x}^2$$

$$\text{Et : } \sigma_x = \sqrt{V(x)}$$

$$\text{Ainsi pour nos étudiants d' AES : } V(x) = \frac{4637500}{200} - \overline{117,5^2}$$

$$\text{Enfin : } V(x) = 18187,50 - 13806,25 = \mathbf{4\ 381,25}$$

$$\text{Et : } \sigma_x = \sqrt{4381,25} = \mathbf{66,19}$$

En moyenne, l'écart entre la consommation hebdomadaire de chocolat des étudiants d'AES et la consommation moyenne de ces mêmes étudiants est de 66,19 grammes.

Pour les étudiants de sociologie, nous savons qu'en moyenne ils consomment plus de chocolat par semaine est plus élevée ($\bar{x} = 124$), et nous savons également qu'en moyenne, le carré des écart par rapport à la consommation moyenne de chocolat hebdomadaire est de 5000 grammes au carré ($V(x) = 5000$).

Très rapidement, nous pouvons déterminer l'écart-type de cette seconde distribution :

$$\sigma_x = \sqrt{5000} = \mathbf{70,71}$$

Ainsi en moyenne, l'écart entre la consommation de chocolat hebdomadaire des étudiants en sociologie et la consommation moyenne de ces mêmes étudiants est de 66,64 grammes.

Toutefois, les valeurs de références dans le calcul de la dispersion pour chacune des séries étant différentes ($\bar{x}_{AES} \neq \bar{x}_{Socio}$), les écart-types ne sont pas comparables en l'état.

Nous devons neutraliser l'effet de ces différences entre les moyennes, et ainsi déterminer les coefficients de variation (CV) afin de comparer les dispersions entre ces distributions :

$$CV_x = \frac{\sigma_x}{\bar{x}}$$

$$\text{Ainsi : } CV_{AES} = \frac{66,19}{117,5} = \mathbf{0,56}$$

$$\text{Et : } CV_{Socio} = \frac{70,71}{124} = \mathbf{0,57}$$

$$\text{Enfin : } CV_{AES} < CV_{Socio}$$

Nous pouvons conclure qu'effectivement la consommation hebdomadaire de chocolat est plus stable (moins dispersée, moins hétérogène) chez les étudiants d'AES par rapport à ceux de sociologie.

5. Dans le cadre d'une étude sur l'alimentation des étudiants, l'espace santé des étudiants de Bordeaux souhaite savoir si les variations de la consommation de chocolat sont principalement dues aux variations au sein de chacun des deux cursus ou entre les différents cursus. Proposer une solution.

Il s'agit de calculer la variance globale et d'observer quel est le terme dominant dans sa composition (entre la variance intra-groupes et inter-groupes).

Dans un premier temps, nous déterminons la moyenne globale (la moyenne des moyennes pondérées des effectifs de chaque groupe) :

$$\bar{x}_{globale} = \frac{1}{K} \sum_{k=1}^K n_k \bar{x}_k \quad (n_k \text{ étant l'effectif du groupe } k, \text{ et } K \text{ étant l'effectif de l'ensemble des groupes}).$$

$$\text{Ainsi : } \bar{x}_{globale} = \frac{1}{200+150} [(200 \cdot 117,50) + (150 \cdot 124)]$$

$$\bar{x}_{globale} = \mathbf{120,29}$$

Sur l'ensemble des étudiants interrogés dans les deux cursus, la consommation hebdomadaire moyenne de chocolat était de 120,29 grammes.

$$\text{Calculons désormais la variance globale : } V(x)_{globale} = \frac{1}{K} \sum_{k=1}^K n_k V(x_k) + \frac{1}{K} \sum_{k=1}^K n_k (\bar{x}_k - \bar{x}_{globale})^2$$

Le premier terme représente la variance au sein des groupes (intra, $\overline{V(x_k)}$) et le second, la variance entre les différents groupes (inter, $V(\bar{x}_k)$).

$$\text{Ainsi : } V(x)_{globale} = \frac{1}{350} [(200 \cdot 4381,25) + (150 \cdot 5000)] + \frac{1}{350} [200(117,50 - 120,29)^2 + 150(124 - 120,29)^2]$$

$$\text{Enfin : } V(x)_{globale} = \mathbf{4646,43 + 10,35 = 4656,78}$$

$$\text{En proportion de la variance globale : } V(x)_{globale} (\%) = \mathbf{99,78\% + 0,22\%}$$

Dans la mesure où la variance intra domine très largement la variance inter dans la composition de la variance globale (99,78%), nous pouvons conclure que la dispersion totale de la consommation hebdomadaire de chocolat est essentiellement due à l'hétérogénéité de la consommation entre les étudiants des deux cursus (par rapport à l'hétérogénéité entre les étudiants des deux cursus).