Séries statistiques à deux variables

3

Olivier FINOT, 11 janvier 2018

Objectifs

Être capable :

1. de calculer une moyenne, un écart type ;
2. de calculer une médiane, une étendue, un interquartile ;
3. de calculer une fréquence conditionnelle ;
4. de réaliser un ajustement affine par méthode graphique ;
5. d'utiliser l'équation d'une droite d'ajustement fournie par un tableur.

I. Statistiques à une variable (révisions)

1) Médiane et moyenne

Définition

La **médiane** Me d'une série statistique est le nombre qui **partage la série en deux** séries ayant **le même effectif**.
La moitié (ou 50 %) des valeurs de la série sont inférieures ou égales à la médiane et l'autre moitié (50 %) lui sont supérieures ou égales.

Définition

On note $x_1, x_2, ..., x_p$ les valeurs du caractère étudié et $n_1, n_2, ..., n_p$ les effectifs correspondants.
La **moyenne** $\bar{x}$ de la série statistique est $\bar{x} = \dfrac{n_1x_1 + n_2x_2 + ... + n_px_p}{N} = \dfrac{\Sigma n_ix_i}{N}$

2) Étendue

Définition

L'**étendue** e d'une série statistique est la différence entre la plus grande et la plus petite valeur de la série.

3) Quartiles

Définition

- Le **premier quartile Q_1**, est la plus petite valeur à laquelle un quart (ou 25 %) des valeurs sont inférieures ou égales.
- Le **troisième quartile Q_3**, est la plus petite valeur à laquelle trois quarts (ou 75 %) des valeurs sont inférieures ou égales.
- L'**écart interquartile $Q_3 - Q_1$** est la différence entre les 3^e et 1^{er} quartiles : $Q_3 - Q_1$. Il regroupe au moins 50 % des effectifs de la série avec un nombre égal de valeurs réparties de part et d'autre de la médiane Me.

4) Écart type

Définition

L'**écart type σ** (sigma), fourni par la calculatrice ou le tableur, mesure la dispersion de la série autour de la moyenne $\bar{x}$.
Plus l'écart type σ est grand, plus les valeurs sont «**dispersées**» autour de la moyenne.
Inversement, plus l'écart type σ est grand, plus les valeurs sont «**resserrées**» autour de la moyenne.

II. Tableaux croisés d'effectifs

1) Rappel sur les fréquences

Définition

La fréquance f d'une population A dans une population E est le rapport des effectifs :

$$f = \frac{n_A \, (EffectifdeA)}{n_E \, (EffectifdeE)}$$

Exemple

On considère les montant des achats, en euros de $N = 200$ personnes dans une pharmacie un jour donné :

Montant des achats	[0, 20[	[20, 40[	[40, 60[	[60, 80[
Effectif n_i	52	110	30	8
Fréquence f_i	0,26	0,55	0,15	0,04

2) Fréquence conditionnelle

Activite Adhérents d'un club de sport

Parmi les 360 adhérents d'un club de sport, une enquête à donné les résultats suivants :

- 5 % des adhérents sont fumeurs et pratiquent la compétition ;
- 54 sont des fumeurs ;
- Les non-fumeurs ne pratiquant pas la compétition sont cinq fois plus nombreux que les fumeurs qui pratiquent la compétition.

1 Compléter le tableau suivant :

	Compétition (C)	Pas compétition ($\bar{C}$)	Total
Fumeurs (F)			
Non fumeurs ($\bar{F}$)			
Total			

2
a) Quelle est la proportion, notée $f(C)$ de personnes pratiquant la compétition ?

b) Déterminer la proportion $f(F)$ de fumeurs.

c) Quelle est la proportion, notée $f(F \cap C)$ de personnes qui fument **et** pratiquent la compétition ? (On l'appelle fréquence conjointe de F et C)

d) Déterminer la proportion, notée $f_c(F)$ de fumeurs parmi les personnes pratiquant la compétition ? (On l'appelle fréquence conditionnelle de F sachant C).

e) Quelle est la proportion, notée $f(F \cup C)$, des personnes qui fument **ou** qui pratiquent la compétition ? (On l'appelle fréquence de la réunion de F et C).

1)

	Compétition (C)	Pas compétition ($\bar{C}$)	Total
Fumeurs (F)	18	36	54
Non fumeurs ($\bar{F}$)	216	90	306
Total	234	126	360

2) **a)**

$$\begin{aligned} f(C) &= \frac{Effectif\ de\ C}{Effectif\ total} \\ f(C) &= \frac{234}{360} \\ f(C) &= 0{,}65 \end{aligned}$$

La proportion de personnes pratiquant la compétition est 0,65.

b)

$$\begin{aligned} f(F) &= \frac{Effectif\ de\ F}{Effectif\ total} \\ f(F) &= \frac{54}{360} \\ f(F) &= 0{,}15 \end{aligned}$$

La proportion de fumeurs est 0,15.

c)

$$\begin{aligned} f(F \cap C) &= \frac{Effectif\ des\ fumeurs\ pratiquant\ la\ compétition}{Effectif\ total} \\ f(F \cap C) &= \frac{18}{360} \\ f(F \cap C) &= 0{,}05 \end{aligned}$$

La proportion de personnes qui fument et pratiquent la compétition est 0,05.

d)

$$\begin{aligned} f_C(F) &= \frac{Effectif\ des\ fumeurs\ pratiquant\ la\ compétition}{Effectif\ des\ personnes\ pratiquant\ la\ compétition} \\ f_C(F) &= \frac{18}{234} \\ f_C(F) &\approx 0{,}08 \end{aligned}$$

La proportion de fumeurs parmi les personnes pratiquent la compétition est 0,08.

e)

$$\begin{aligned} f(F \cup C) &= \frac{fumeurs + pratiquants - fumeurs\ pratiquants}{Effectif\ total} \\ f(F \cup C) &= \frac{54 + 234 - 18}{360} \\ f(F \cup C) &= \frac{270}{360} \\ f(F \cup C) &= 0{,}75 \end{aligned}$$

La proportion de personnes qui fument ou pratiquent la compétition est 0,75.

Á retenir

A et B sont deux sous-populations d'une population E.

- $\mathbf{f(B)}$ est la **fréquence marginale** de B :

$$f(B) = \frac{Effectif\ de\ B}{Effectif\ de\ E}$$

- $\mathbf{f(A \cap B)}$ est la **fréquence conjointe** de A et B :

$$f(A \cap B) = \frac{Effectif\ du\ croisement\ de\ A\ et\ de\ B}{Effectif\ de\ E}$$

- $\mathbf{f(A \cup B)}$ est la **fréquence de la réunion** de A et B :

$$f(A \cup B) = \frac{Eff.\ de\ A + Eff.\ de\ B - Eff.\ du\ croisement\ de\ A\ et\ de\ B}{Effectif\ de\ E}$$

ou

$$f(A \cup B) = f(A) + f(B) - f(A \cap B)$$

- $\mathbf{f_B(A)}$ est la **fréquence conditionnelle** de A **sachant** B :

$$f_B(A) = \frac{Effectif\ du\ croisement\ de\ A\ et\ de\ B}{Effectif\ de\ B}$$

ou

$$f_B(A) = \frac{f(A \cap B)}{f(B)}$$

III. Statistiques à deux variables

1) Définition et représentation graphique

Á retenir

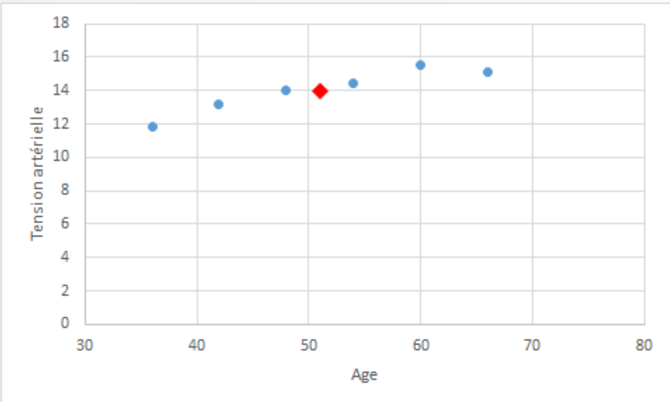
- Lorsqu'on étudie **deux caractères statistiques** sur une même population, on obtient une **série statistique double**.
- La représentation d'une série statistique double forme un **nuage de points**.
- Le **point moyen** G d'un nuage de points a pour coordonnées $(\bar{x};\bar{y})$.

Exemple

Le tableau suivant donne, dans une population féminine, la moyenne de la tension artérielle en fonction de l'âge :

Âge en années : x_i	36	42	48	54	60	66
Tension maximale : y_i	11,8	13,2	14	14,4	15,5	15,1

- La moyenne des abscisses est : $\bar{x} = 51$;
- La moyenne des ordonnées est : $\bar{y} = 14$
- Les coordonnées du point moyen G sont donc : $(51; 14)$.



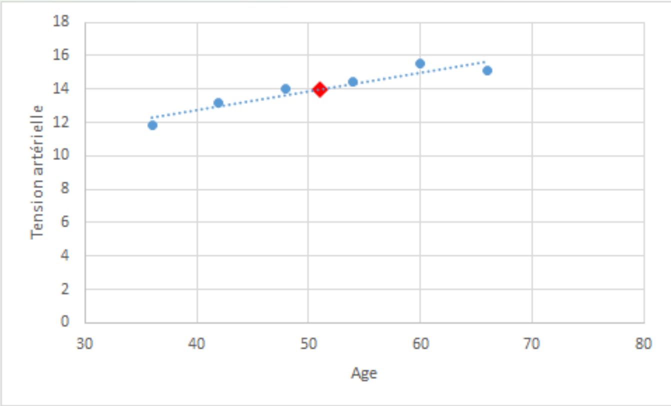
2) Ajustement affine

Á retenir

- Si le nuage de points a une forme «allongée», on peut calculer un **ajustement affine** du nuage.
- On obtient ainsi une **droite d'ajustement** (ou droite de régression) qui passe par le point moyen G et au plus près des autres points du nuage.

Exemple

La droite d'ajustement obtenue grâce au tableur passe par le point moyen G dont nous avons calculé les coordonnées.



3) Prévisions

Méthode

- La droite d'ajustement donne la «tendance» de l'évolution de la grandeur y en fonction de celle de x.
- En supposant que la tendance se poursuive, il est possible d'estimer une valeur future par lecture graphique ou à partir de l'équation de la droite.

Exemple

En prolongeant la droite d'ajustement obtenue on peut tenter d'estimer la tension artérielle à un âge plus avancé.