# Statistiques

Marc SAGE

mardi 5 novembre – jeudi 28 novembre

## Table des matières



Introduction : activités *partir un bon pied* p. 162, activité *découvrir* 1 p. 163
Exercices : vrai-faux 29, 32, 33 p. 180
D. M. : exo 38 p. 181
D. S. : exos 34, 35, 37 p. 180-181

Suggestion de lecture : *Le grand trucage* de Lorraine Data.

La statistique est l'étude méthodique des faits sociaux par des procédés numériques (définition du XVIIIe). Vient de l'italien *statista*, "homme d'état" : la statistique est à l'origine une connaissance utile à l'homme d'état, lequel a affaire à des populations de grand effectif.

En seconde, la part mathématique est quasi-nulle (même si l'on calculera des moyennes) : on ne fait que de la *lecture* et de la *représentation*.

# 1 Représentation d'une série statistique

On étudie une ***population*** formée d'***individus*** (par exemple : les enfants nés en 2000, les voitures produites par Peugeot en octobre 2007). L'étude porte sur un certain ***caractère*** des individus (couleur des yeux, poids, n° INSÉÉ, date de production, salaire, pays d'origine...) dont la valeur peut être ***numérique*** (on parle alors de caractère ***quantitatif*** : taille, salaire...) ou non (on parle alors de caractère ***qualititatif*** : couleur, date de naissance...)

★ L'étude porte sur un *caractère*, elle oublie complètement la *singularité* des individus. ★

L'***effectif*** d'une valeur donnée est le nombre d'individus dont le caractère possède cette valeur, sa ***fréquence*** est son effectif rapporté à l'effectif total. Lorsque l'on peut ordonner les valeurs, on définit l'***effectif cumulée croissant*** (resp. ***fréquence cumulée croissante***) d'une valeur donnée $v$ par la somme des effectifs (resp. fréquence) des valeurs $\leq v$.

**Exemple 1.** Étudions par exemple la couleur des yeux de six personnes :

| individu | Alice | Bob | Charles | Denise | Émilie | Francis |
|---|---|---|---|---|---|---|
| valeur du caractère | bleu | bleu | marron | vert | vert | vert |

.

Oublions la singularité de la population pour ne retenir que les valeurs et les effectifs :

| valeur | bleu | vert | marron | TOTAL |
|---|---|---|---|---|
| effectif | 2 | 3 | 1 | 6 |
| fréquence | $\frac{1}{3} \simeq 0,33$ | $\frac{1}{2} = 0,50$ | $\frac{1}{6} \simeq 0,17$ | 1 |
| angle (en °) | 120 | 180 | 60 | 360 |

.

On peut représenter cette série qualitative par un ***diagramme en bâtons***, avec les valeurs en abscisse (régulièrement espacées) et les effectifs en ordonnée. On peut également la représenter par un ***digramme circulaire*** où l'aire de chaque secteur est proportionnelle à l'effectif, ce qui revient à dire que l'angle au sommet est proportionnel à l'effectif.

**Exemple 2.** Lorsqu'on étudie un caractère quantitatif, on peut regrouper les valeurs par ***classes*** de valeurs. Intéressons-nous par exemple au temps de trajet moyen des élèves d'une classe de leur domicile à leur lycée :

| classe de valeurs (en min.) | $[0,5[$ | $[5,10[$ | $[10,15[$ | $[15,20[$ | $[20,25[$ | $[25,30[$ | TOTAL |
|---|---|---|---|---|---|---|---|
| effectif | 2 | 5 | 4 | 10 | 6 | 3 | 30 |
| effectif cumulé croissant | 2 | 7 | 11 | 21 | 27 | 30 | |

.

Dans ce cas, on représente la série par un ***histogramme*** (= diagramme de rectangles) avec en abscisse les valeurs (à l'échelle) de sorte que l'aire de chaque rectangle soit proportionnel à l'effectif correspondant. Cela suppose de se donner une aire de référence arbitraire (choisie pour simplifier les calculs). Lorsque les classes sont de même largeur, les hauteurs sont directement proportionnelles à l'effectif. Par exemple, on peut regrouper comme suit :

| classe de valeurs | $[0,15[$ | $[15,20[$ | $[20,30[$ |
|---|---|---|---|
| effectif | 11 | 10 | 9 |
| aire | 55 | 50 (référence) | 45 |
| largeur | 15 | 5 | 10 |
| hauteur | $\frac{11}{3} \simeq 3,7$ | 10 | 4,5 |

.

(Sur les histogrammes, visualiser que les rectangles contiennent un liquide : regrouper les classes revient à lever certaines cloisons, le liquide se stabilise alors tout seul à une certaine hauteurv – celle du nouveau rectangle.)

## 2 Indicateurs d'une série quantitative

Une série statistique contient beaucoup d'information (même si on a déjà oublié les individus). Pour faciliter sa lecture et sa comparaison avec d'autres, on va lui associer des nombres qui en "donnent une idée"

La ***valeur moyenne*** est la somme des valeurs (coefficientée par les effectifs) rapportée à la somme des effectif.

L'***étendue*** est la différence entre les valeurs maximale et minimale.

Une ***valeur médiane***[1] (resp. ***première valeur quartile***[2], resp. ***troisième valeur quartile***[3]) est une valeur $v$ telle que :

1. au moins 50% (resp. 25%, resp. 75%) de la population a un caractère de valeur $\leq v$ ;
2. au moins 50% (resp. 75%, resp. 25%) de la population a un caractère de valeur $\geq v$.

L'***intervalle interquartile*** [4] est la différence entre les valeurs quartiles supérieure et inférieure. (Il contient donc une moitié de la population environ.)

En pratique, si l'on note $N$ l'effectif total, *la* valeur médiane[5] est définie par :

1. (si $N$ est impair) la valeur du $\frac{N+1}{2}$-ième individu ;
2. (si $N$ est pair) la moyenne des valeurs des $\frac{N}{2}$-ième et $\left(\frac{N}{2}+1\right)$-ième individus.

De même, si pour tout réel $x$ on note[6] $\lceil x \rceil$ le plus petit entier $\geq x$, alors *la* valeur quartile inférieure[7] (resp. supérieure[8]) est définie par la valeur du $\left\lceil \frac{N}{4} \right\rceil$-ième individu (resp. $\left\lceil \frac{3N}{4} \right\rceil$-ième individu)

**Remarque.** Valeurs moyenne et quartiles sont des indicateurs de *position* (ils indiquent *où se situe* la série), étendue et intervalle interquartile sont des indicateurs de *dispersion* (ils indiquent *comment se répartit* la série).

★ Ces indicateurs sont des *constructions intellectuelles* qui ne remplaceront jamais la réalité et la singularité des populations étudiées.

★ Pire : en sélectionnant la population, on peut modifier les indicateurs pour affirmer des choses sur les caractère étudié (chômage, mortalité, immigration...) –> *cf.* exercice 1 p. 164.

**Exemple.** On étudie les notes d'une classe de vingt-six élèves :

| valeur | 3 | 5 | 7 | 8 | 10 | 11 | 13 | 14 | 17 | TOTAL |
|---|---|---|---|---|---|---|---|---|---|---|
| effectif | 1 | 2 | 1 | 5 | 4 | 1 | 7 | 3 | 2 | 26 |
| effectif cumulé croissant | 1 | 3 | 4 | 9 | 13 | 14 | 21 | 24 | 26 | |

.

La note moyenne vaut $\frac{1+2\cdot5+7+5\cdot8+4\cdot10+11+7\cdot13+3\cdot14+2\cdot17}{1+2+1++4+1+7+3+2} = \frac{276}{26} = \frac{138}{13} \simeq 10,6$.

L'effectif total $26 = 2\cdot13$ étant pair, la note médiane est la moyenne entre les treizième et quatorzième notes, à savoir $\frac{10+11}{2} = 10,5$.

Puisque $\left\lceil \frac{26}{4} \right\rceil = \lceil 7,5 \rceil = 8$, la note quartile inférieure est la huitième, à savoir 8.

Puisque $\left\lceil 3\frac{26}{4} \right\rceil = \lceil 22,5 \rceil = 23$, la note quartile inférieure est la vingt-troisième, à savoir 14.

L'intervalle interquartile vaut donc $14-8=6$.

L'étendue vaut $17-3=14$.

On peut regrouper les valeur par classes et dessiner un histogramme :

| classe | $[0,5[$ | $[5,10[$ | $[10,15[$ | $[25,30[$ | TOTAL |
|---|---|---|---|---|---|
| effectif | 1 | 8 | 15 | 2 | 26 |

.

[1] ou ***deuxième valeur quartile***

[2] ou ***valeur quartile inférieure***

[3] ou ***valeur quartile supérieure***

[4] Les valeurs quartiles séparent la population en quatre parties chacune appelé un ***quartile***. Attention à l'abus de langage qui idenfie une portion d'individu (un quartile) avec une valeur extrémale de cette portion (une valeur quartile).

[5] on vérifie bien sûr que *la* médiane est *une* médiane

[6] On a par exemple les égalités $\lceil 4 \rceil = 4$, $\lceil 17 \rceil = 17$, $\lceil 15,7 \rceil = 16$ et $\lceil 72,001 \rceil = 73$.

[7] on vérifie bien sûr que *la* quartile inférieure est *une* quartile inférieure

[8] on vérifie bien sûr que *la* quartile supérieure est *une* quartile supérieure

On peut aussi regrouper selon des classes de largeurs différentes :

| classe | $[0, 8[$ | $[8, 14[$ | $[14, 20[$ |
|---|---|---|---|
| effectif | 4 | 17 | 5 |
| aire | 12 | 51 | 15 (référence) |
| largeur | 8 | 6 | 6 |
| hauteur | $1,5$ | $6,5$ | $2,5$ |