

(S3) statistique descriptive !

Construction d'un histogramme

→ à partir d'une série statistique (réalisation d'un échantillon)

Cas d'une variable quantitative qui prend des valeurs entières ou réelles (cas discret ou continu).

① Découpage en classe de l'ensemble des valeurs, en présence

② Calcul des effectifs d'occurrences de ces classes

Souvent, ces 2 premières étapes sont "déjà" effectuées lorsque la série statistique est donnée par un tableau d'effectifs.

Exemple: les notes au premier partiel de la promo sont les suivantes

Notes	$[5, 7[$	$[7, 9[$	$[9, 12[$	$[12, 14[$
Effectif	10	31	33	26

nb d'occurrences = fréquence

→ on en déduit que la promo compte $(10 + 31 + 33 + 26)$ étudiants
100


→ ce tableau est plus compact
que la donnée de toutes
les notes. (mais moins précis)

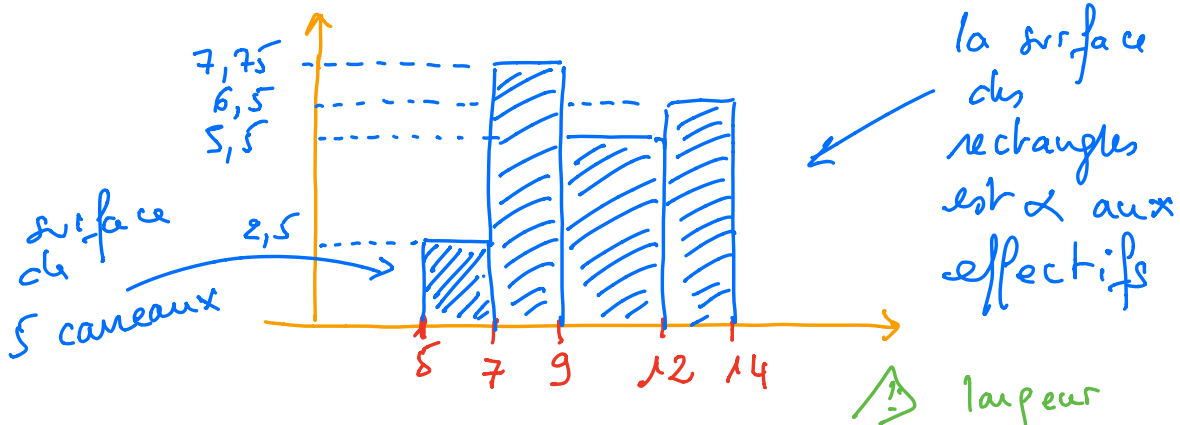
ex 6 5,5 6 5 5,5 5 6 5,5 6,5 6

10 notes $\in [5, 7[$
moyenne $[5,7]$

→ sans ces détails, une note représentant
la première classe $[5,7[$ serait une
valeur moyenne entre ses bornes $[5,7]$

L'histogramme approche le graphe de
la densité de la loi associée à la
population

⚠ 1 caneau représente (1 surface
de 1 caneau) représente un effectif
de 2 notes 



$$(7-5) \times \text{hauteur} = 10 \text{ notes } \text{variable}$$

⚠ parfois la série statistique (la réalisation d'un échantillon) est fournie avec des fréquences relatives

$$\text{freq relative} \quad \left(f_i = \frac{n_i}{n} \right) \begin{array}{l} \leftarrow \text{effectif de la classe} \\ \leftarrow \text{effectif total} \end{array}$$

$$\left(\sum_i n_i = n \quad \sum_i f_i = 1 \right)$$

Un histogramme n'est donc (pas) un diagramme en colonnes / barres / bâtons où les effectifs seraient représentés en ordonnées !

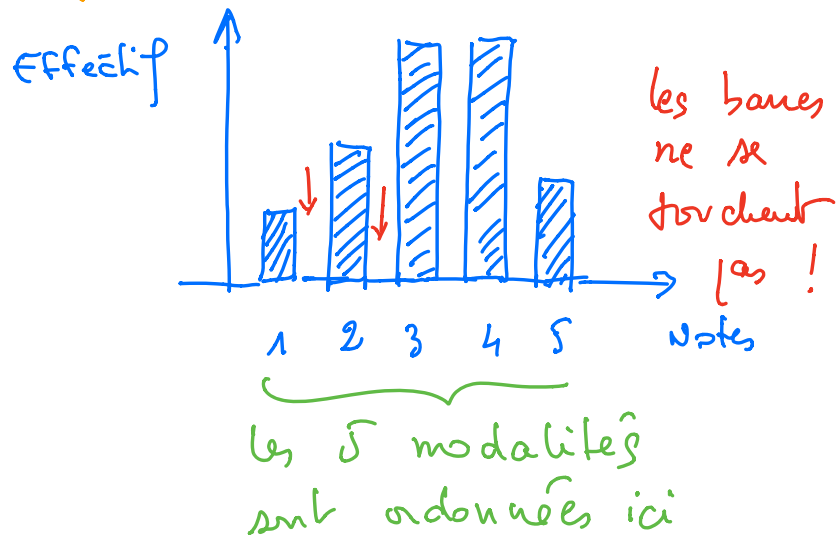
Cas d'une variable qualitative. Dans ce cas on parle de modalités (et non de valeurs numériques) pour les valeurs

possibles

On se contente d'un diagramme en bâton
où la val d'effectifs de chaque modalité

(on parle de variable ordonale si les
modalités sont naturellement ordonnées)
est donnée en ordonnée

Notes (sur 5)	1	2	3	4	5
Effectif	2	4	7	7	3



RQ une variable qualitative est nominale si ses
modalités ne peuvent être classées

(S4) Statistique descriptive

Calculs de paramètres de position
Quartiles, Quantiles, Médiane

Idée on ordonne les valeurs présentes dans une série statistique

Ex (après tri) médiane

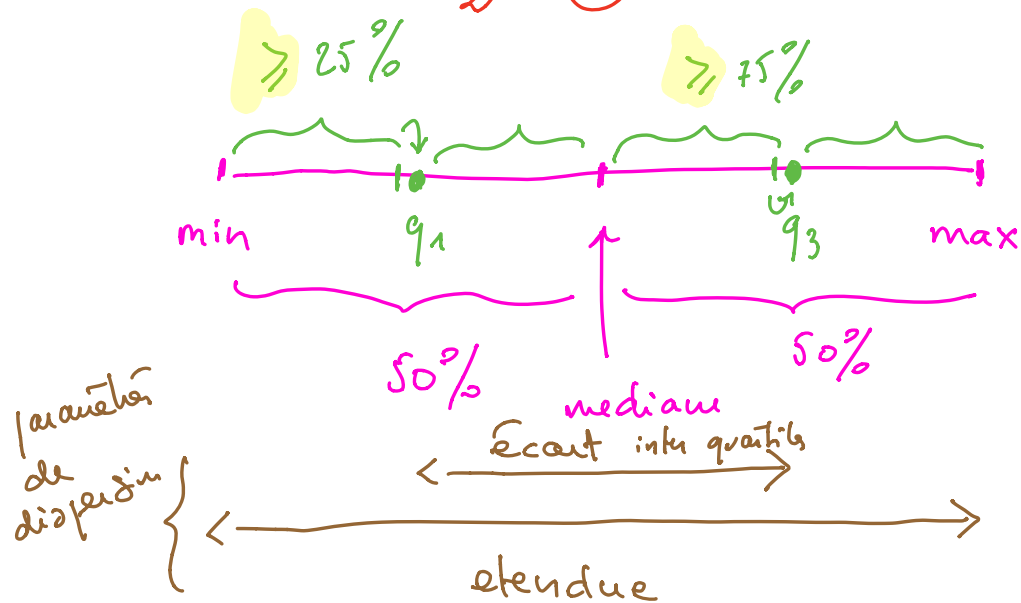
3 5 6 11 **11** 13 14 15 17

2 pour série de n effectif

Pour une série avec n nb pair

3 5 6 **11** **13** 14 15 17

$$\frac{11+13}{2} = 12 \rightarrow \text{médiane}$$



Représentation : boîte à moustaches.

La médiane est un paramètre de tendance
centrale qui a des propriétés différentes
 de la moyenne

Si la série statistique est donnée par
 les $y_1 \dots y_n$ $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$

Si des effectifs sont fournis

$(x_1, n_1) \dots (x_i, n_i) \dots (x_p, n_p)$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^p n_i x_i \quad \text{avec} \quad n = \sum_{i=1}^p n_i$$

$$\bar{x} = \sum_{i=1}^p \left(\frac{n_i}{n} \right) x_i \quad \text{freq relatives } \frac{n_i}{n} = f_i$$

De même les paramètres de dispersion
 sont souvent donnés par la
 variance et l'écart-type

$$\sigma_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

$$\rightarrow \sigma_y^2 = \left(\frac{1}{n} \sum_{i=1}^n y_i^2 \right) - \bar{y}^2$$

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

Notion de modes maximum

(local) de la distribution

\leadsto distributions $\begin{cases} \text{monomodale} \\ \text{bimodales} \end{cases}$