Statistiques

Chapitre 1 : Séries statistiques – Généralités

1) Définitions

Population : Ensemble d'éléments faisant l'objet d'une étude statistique
exemples : ensemble des élèves d'un amphi (âge,lieu d'habitation…)
ensemble des pièces produites par une machine (taille/qualité…)
ensemble d'assurés (nombre de voiture, nombre d'accidents…)

Echantillon : Partie de la population
exemples : groupe A des élèves de l'amphi
pièces fabriquées un jour donné
assurés d'un département donné

Unité statistique : Un élément de la population étudiée dont on veut examiner un ou plusieurs caractères.

Caractère qualitatif : Non mesurable
exemples : qualité : bon ou mauvais, lieu d'habitation…

Caractère quantitatif : Mesurable à l'aide d'une variable statistique
exemples : nombre d'individus, longueur d'une pièce fabriquée…

Variable statistique :

Discrètes : Prend des valeurs isolées
exemple : nombre d'individus

Continues : Prend toutes les valeurs sur un intervalle
exemple : longueur

Série statistique : Ensemble des valeurs prises par une variable statistique sur l'ensemble de la population ou sur un échantillon

2) Séries statistiques d'un caractère quantitatif discret

Soit un échantillon de taille n (x1, x2, … ,xp) les valeurs possibles du caractère x.
Série statistique : les valeurs prises par les n membres de l'échantillon.

Effectif total : n

Effectif partiel de x_i : c'est le nombre n_i de fois qu'apparaît la valeur x_i .

$n1 + n2 + n3 + \cdots + np = n$

Fréquence relative de x_i :

$$fi = \frac{ni}{n}$$

Remarque : $f1 + f2 + f3 + \cdots + fp = 1$

Etendue de la série : Ecart entre la plus grande et la plus faible des valeurs de x_i .

exemple : On considère 100 familles de 4 enfants.
On va étudier le caractère : nombre de garçon(s)
Série statistique : 3,1,3,2,4,0,1,2,…

Effectif total : 100

Etendue : 4-0=4

On va compter les valeurs :

x_i	0	1	2	3	4
n_i	7	20	43	25	5

n_i effectif partiel

$$\sum_{i=1}^{5} ni = n = 100$$

Fréquence relative des familles avec deux garçons : $\frac{43}{100}$.

3) Séries statistiques d'un caractère quantitatif continue

Le nombre p de valeurs possibles de x_i est infini (en réalité c'est fini à cause de la précision des mesures mais très grand => les f_i sont presque nulles)

On va constituer des classes en divisant l'étendue de la série en un certain nombre d'intervalles.

---[----------[------------[------[----------[------[----[-------------[----

classe 1| classe 2| 3 | 4 | .... | classe p |

Centre :	x_1	x_2	x_3	....	x_p
Effectif :	n_1	n_2	n_3	....	n_p

Pesées effectuées à 10g près. Donne le poids entre 2,24kg et 4,45kg.

2,2 2,5 2,8 3,1 3,4 3,7 4,0 4,3 4,6

-----|---------|---------|-----------|-------------|-------------|-----------|------------|-------------------|--

Classes:	1	2	3	4	5	6	7	8
Centre:	2,35	2,65	2,95	3,25	3,55	3,85	4,15	4,45
Effectif:	5	11	24	40	42	20	13	6
Fréquence:	3,1	6,8	14,9	24,8	26,1	12,4	8,1	3,7

=>total = 99,9%

4) Séries statistiques d'un caractère quanlitatif

On groupe les résultats en autant de classes qu'il existe de modalités de caractère.

Exemple : Etude sur la couleur de fleurs avec 3 couleurs possibles : rouge, bleu et vert. On va définir 3 classes que correspondent à chaque couleur et on y associe un effectif.

Variables ordinales : si les modalités peuvent être ordonnées
exemple : taille vestimentaire plus ou moins grandes…

Variables nominales : si elles ne peuvent être ordonnées
exemple : couleur, profession…

Variables dichotomique : avec deux modalités

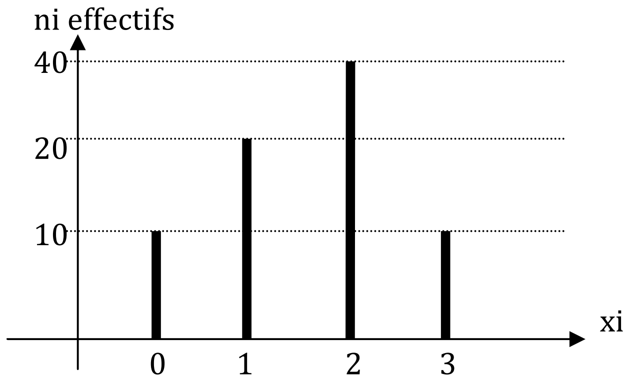
5) Représentation graphique des séries statistiques

a) Cas des variables discrètes

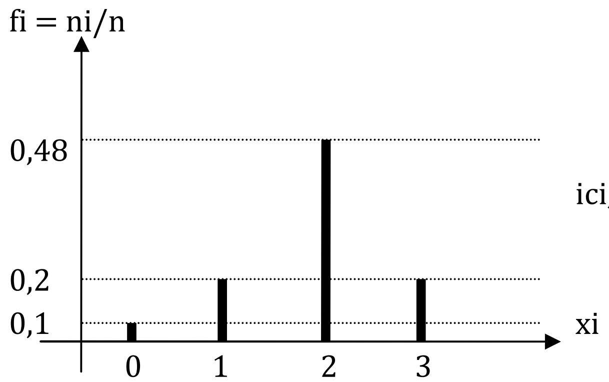
Diagrammes en bâtons :

On trace ni en fonction de xi
ou fi en fonction de xi

Exemple avec les familles de 4 enfants avec xi = nombre de garçon(s)



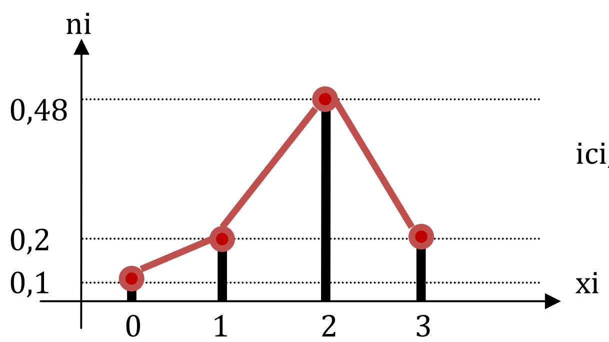
Ou avec les fréquences relatives :



ici, la somme de la hauteur des bâtons vaut 1

- <u>Polygône des effectifs et des fréquences absolues</u>

On va joindre l'extremité des bâtons par des segments de droite. Si l'on reprend l'ensemble prescedent :



ici, la somme de la hauteur des bâtons vaut 1

- <u>Polygône des fréquences relatives</u>

Même chose que le graph prescedent sauf que ni est remplacé par fi

b) Cas des variables continues

Diagramme en bâtons impossible car il y a trop de valeurs de xi.
=><u>histogrammes</u>

Deux cas possibles:
*classes égales : on va partager le domaine des valeurs de xi en classes = intervalles d'où la representation :
Exemple avec les nourrissons

```
     2,2       2,5      2,8        3,1          3,4           3,7        4,0          4,3                 4,6
-----|---------|---------|-----------|-------------|--------------|-----------|------------|-------------------|--
Classes:1         2          3            4              5             6            7                  8
Effectif:  5      11         24           40             42            20           13                 6
```

Voir graph (1) photocop

Remarques:

- Aire d'un rectangle = l . ni
- Aire totale sous l'histogramme :

$$\sum_{i=1}^{p} l.ni = l \sum_{i=1}^{p} ni = l.n$$

(avec n = effectif total)

✓ classes inégales :

classes	5-6	6-7	7-8	8-10
ni effectifs	12	13	16	6

Il faut corriger l'effectif de la classe 8-10 : 6/2=3

d'où l'histogramme :

Voir graph A

Il est à remarquer que l'on conserve l'aire totale : 1.47=47

❖ <u>Polygone des effectifs et des fréquences :</u>

Ligne brisée poignant les milieux de sommets des rectangles.

Ligne brisée joignant les points (xi,ni) avec xi centre des classes

❖ <u>Polygone des effectifs cumulés :</u>

exemple des nourrissons :

Voir graph B

❖ <u>Polygones des frequences relatives cumulés</u>

Même chose que les graphes prescedents sauf que l'on remplace $\sum ni$ par $\sum fi$

Chapitre 2 : Paramètres de position et de dispersion

Objectif : On veut obtenir des paramètres pour condenser l'information contenue dans la serie statistique étudiée.

- Paramètres de
  - position : donnent un ordre de grandeur des mesures et l'existence de valeurs centrales autour desquelles se groupent les mesures
  - dispertion : donnent la dispersion des valeurs de la serie autour de la valeur centrale.

I. Paramètres de position :

1. Moyenne arithmétique :

Définition :

$$\bar{x} = \frac{x1 + x2 + \cdots + xn}{n} = \frac{1}{n}\sum_{i=1}^{n} xi$$

- Série à caractère discret :

On va avoir p valeurs x1,x2,...,xp et ni est l'effectif de xi
(fi sera la fréquence relative de xi)
alors :

$$\bar{x} = \frac{n1.x1 + n2.x2 + \cdots + np.xp}{n} = \frac{1}{n}\sum_{i=1}^{p} ni.xi$$

ou encore

$$\bar{x} = \sum_{i=1}^{p} \frac{ni}{n}.xi = \sum_{i=1}^{p} fi.xi$$

Exemple : famille de 4 enfants, xi=nombre de garçons

xi	0	1	2	3	4
ni effectif	7	20	43	25	5
fi (en %) f relative	7	20	43	25	5

Série statistique étudiée : 100 familles

On veut calculer la moyenne arithmétique

$$\bar{x} = \frac{x1+x2+\cdots+xn}{n}$$
$$= \frac{x1+x2+\cdots+x100}{100}$$
$$= \frac{0+2+3+2+4+2+3+\cdots\ (100\ valeurs)}{100}$$ trop long

$$= \frac{n1.x1 + n2.x2 + \cdots + np.xp}{100}$$

$$= \frac{7*0 + 20*1 + 43*2 + 25*3 + 5*4}{100} = \frac{1}{100}(20 + 86 + 75 + 20) = \frac{201}{100} = 2{,}01$$

plus rapide!

Remarque : On parle de moyenne arithmétique poudérée par les fréquences relatives fi par la relation :

$$\bar{x} = \sum_{i=1}^{p} fi.xi$$

- Série à caractère continu :

On a une série statistique qvec p classes de centre xi et d'effectif ni alors :

$$\bar{x} = \frac{n1.x1 + n2.x2 + \cdots + np.xp}{n}$$

(avec xp, centre des classes)
ou encore

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{p} ni.xi = \sum_{i=1}^{p} fi.xi$$

Même formule que dans le cas des series à caractère discret. D'une certaine façon, on rend la variable xi discrète en considérant que x ne peut prendre comme valeurs que le centre des classes.

Exemple des nourrissons :

classes	1	2	3	4	5	6	7	8
centre	2,35	2,65	2,95	3,25	3,55	3,85	4,15	4,45
effectif ni	5	11	24	40	42	20	13	6
fi (en %) $\frac{ni}{n}$	3,1	6,8	14,9	24,8	26,1	12,4	8,1	3,7

d'où

$$\bar{x} = \sum_{i=1}^{8} fi.xi = \frac{3{,}1}{100}.2{,}35 + \frac{6{,}8}{100}.2{,}65 + \cdots = 3{,}406\ kg$$

(avec xi centre des classes)

Remarque : calcul plus rapide sur l'exemple des nourissons
On va utiliser la moyenne sur le numero de la classe
On remarque que :

$$xi = M + a.zi$$

(avec zi numero de la classe)
Dans cet exemple : M=2,05 kg ; a= 0,3kg
ou $x1 = 2{,}05 + 0{,}3 * 1 = 2{,}35kg$
$x2 = 2{,}05 + 0{,}3 * z2 = 2{,}65kg$
...
alors

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{8} ni.xi = \frac{1}{n}\sum_{i} ni(M + a.zi)$$

$$\bar{x} = \frac{1}{n}\sum_{i} ni.M = \frac{1}{n}\sum_{i} ni.a.zi$$

$$\bar{x} = \frac{M}{n}\sum_{i} ni.\frac{a}{n}\sum_{i} ni.zi$$

d'où

$$\bar{x} = M + a\bar{z}$$

moyenne arithmétique du numero des classes

Donc ici,

$$\bar{z} = \frac{5 * 1 + 11 * 2 + 24 * 3 + \cdots}{161}$$

2. Médiane :

On va ordonner les valeurs de la série statistique et on va prendre la valeur du milieu (médiane). On va ainsi trouver la médiane notée Me.

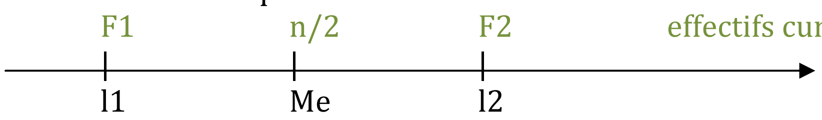
- Série à caractère discret :

Exemples : Série 3,3,4,5,7,7,9 (7 valeurs, valeur médiane : 5)

Série 3,3,4,5,6,7,7,9 (8 valeurs, valeur médiane : $\frac{5+6}{2} = 5{,}5$)

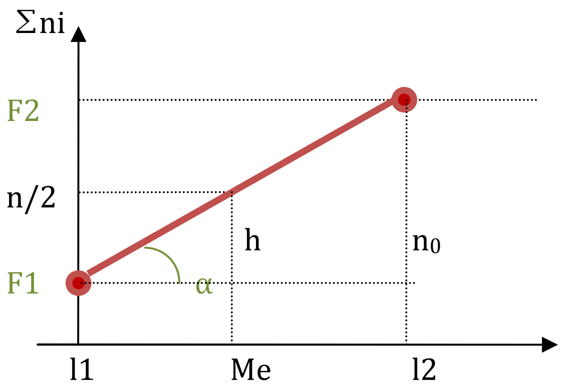
- Série à caractère continu :

On va considérer que la série est déjà ordonnée par classes et la médiane Me va tomber dans une certaine classe [l1 , l2]
Cette classe correspond à la moitier de l'effectif total.



où F1 effectif cumulé en l1
F2 effectif cumulé en l2
n effectif total
n_0 = F2-F1 = effectif de la classe

On suppose les n_0 valeurs de la classe [l1 , l2 [uniformément répartis dans cette classe.
Si l'on reprend le diagramme cumulé :



On veut calculer Me = ?

-détermination graphique par rapport au diagramme cumulatif

-on a $tg\ \alpha = \frac{h}{Me-l1} = \frac{n0}{l2-l1}$

Or : $\frac{n}{2} = \mathrm{F1} + \mathrm{h} = \mathrm{F1} = \frac{F2-F1}{l2-l1}(\mathrm{Me} - \mathrm{l1})$

d'où $\frac{n}{2} - \mathrm{F1} = \frac{F2-F1}{l2-l1}(\mathrm{Me} - \mathrm{l1})$

⇔ $\mathrm{Me} - \mathrm{l1} = (\frac{n}{2} - \mathrm{F1})\frac{l2-l1}{F2-F1}$

⇔ $\mathrm{Me} = \mathrm{l1} + (\frac{n}{2} - \mathrm{F1})\frac{l2-l1}{F2-F1}$

Exemple des nourissons :

classes	2,2 à 2,5	2,5 à 2,8	2,8 à 3,1	3,1 à 3,4	3,4 à 3,7	3,7 à 4,0	4,0 à 4,3	4,3 à 4,6	TOTAL
ni	5	11	24	40	42	20	13	6	161
Σni	5	16	40	80	122	142	155	161	

Dans quelle classe se trouve la médiane?

Si n = 161 ⇒ n/2 = 80,5

Donc on a :

l1 = 3,4 l2 = 3,7 F1 = 80 F2 = 122 n/2 = 80,5

d'où $\mathrm{Me} = 3{,}4 + (80{,}5 - 80) \cdot \frac{3{,}7-3{,}4}{122-80}$

$\mathrm{Me} = 3{,}4 + 0{,}5\,(\frac{0{,}3}{42}) = 3{,}403$ kg

Remarque : Graphiquement sur le polygône des fréquences relatives cumulées, la médiane correspond à l'abscisse qui a ½ en ordonnée.

3. Quartiles:

La médiane sépare la série en 2 groupes de même effectif
Les quartiles séparent la série en 4 groupes de même effectif

- Cas des séries à caractère discret :

Exemples : 4,5,6,11,13,14,16
Q1 Q2 Q3

A gauche de Q1 : 7/4=1,75 termes
On prend 1
A droite de Q1 : on aura aussi un seul terme

Exemples : 4,5,6,7,11,13,14,15,16
Q1 Q2 Q3

A gauche de Q1 : 9/4 termes = 2,25 Donc on va prendre 2 termes
Pour Q3 on prend deux termes au dessus.

- Cas des séries à caractère continu :

On peut se servir du polygône des fréquences relatives cumulées.

Pour la médiane on regarde la valeur à 0,5
Pour les quartiles, on va regarder à ¼ et ¾.

Ou on peut utiliser les formules :
$Q1 = l1 + \left(\frac{n}{4} - F1\right)\frac{l2-l1}{F2-F1}$

Ici la classe [l1 , l2 [est celle où se situe le premier quartile.
Pour Q3, on va prendre $\frac{3n}{4}$

Remarque : on peut définir aussi des déciles (10) et des quantiles/centiles (100)

4. Mode ou valeur dominante :

Le mode est la valeur la plus fréquente (effectif le plus important).

- Cas des séries à caractère discret :

Exemple : 2,2,5,7,9,9,9,10,10,11
valeur dominante ou mode : 9
notation : Mo=9

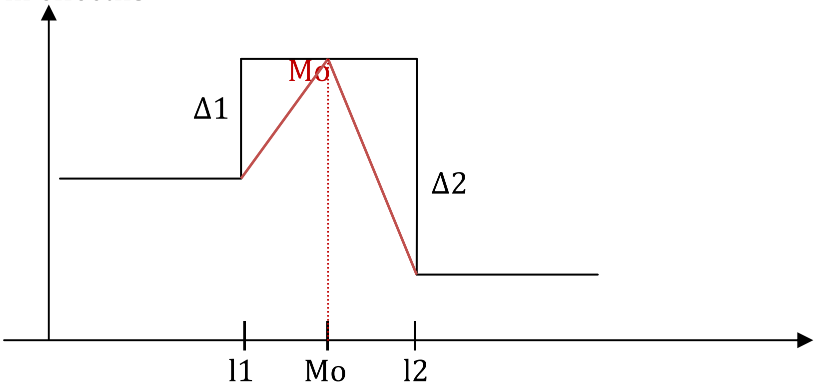
Remarque : on peut avoir plusieurs modes possibles
Exemple : 2,2,5,7,9,9,9,10,10,10,11
Mo = 9 et 10

- Cas des séries à caractère continu :

On va parler de classe modale ou dominante. Il s'agit de la classe avec l'effectif le plus important.

On veut calculer la valeur du mode.

Si on trace l'histogramme et on represente la classe modale (la classe où se trouve le mode) ni effectifs



Le mode est défini de la façon suivante :
On considére que les deux segments de droite en rouge ont la même pente en valeur absolue.
Si la pente est identique, alors : $\left|\frac{\Delta 1}{Mo - l1}\right| = \left|\frac{\Delta 2}{l2 - Mo}\right|$
On va calculer Mo en fonction de l1 , l2 , Δ1 , Δ2 .
A partir de la relation precedente, on a :

$$\Delta 1\ (\,l2 - Mo\,) = \Delta 2\ (\,Mo - l1\,)$$

$$\Leftrightarrow \quad \Delta 1.l2 - \Delta 1.Mo = \Delta 2.Mo - \Delta 2.l1$$

$$\Leftrightarrow \quad \Delta 1.l2 + \Delta 2.l1 = Mo\ (\Delta 1 + \Delta 2)$$

$$\Leftrightarrow \quad Mo = \frac{\Delta 1.l2 + \Delta 2.l1}{\Delta 1 + \Delta 2}$$

d'où $\quad Mo = l1 + \frac{\Delta 1}{\Delta 1 + \Delta 2}\,(l2 - l1)$

avec [l1 , l2 [classe modale et dominante
Δ1 = excèdent d'effectif de la classe modale par rapport à la classe inferieur
Δ2 = excèdent d'effectif de la classe modale par rapport à la calsse supérieur

Exemple des nourrissons :
classe modale : [3,4 – 3,7 [, n5=42
Δ1 = 42 – 40 = 2
Δ2 = 42 – 20 = 22

d'où $\quad Mo = 3{,}4 + \frac{2}{2+22}\,(3{,}7 - 3{,}4)$

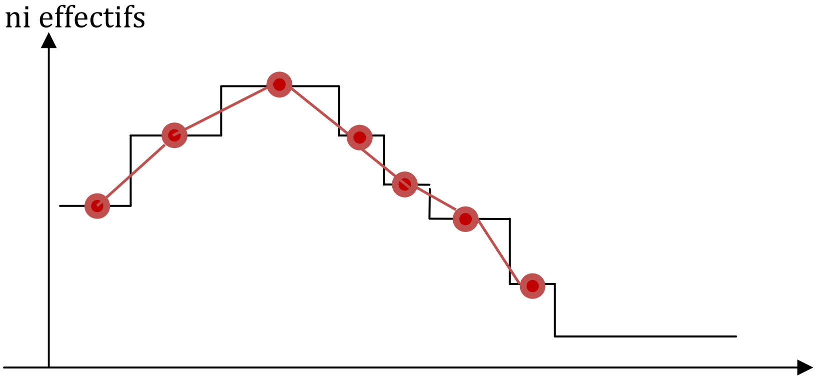
$$Mo = 3{,}4 + \frac{2}{24}\,(0{,}3)$$

$$Mo = 3{,}425$$

Remarques :

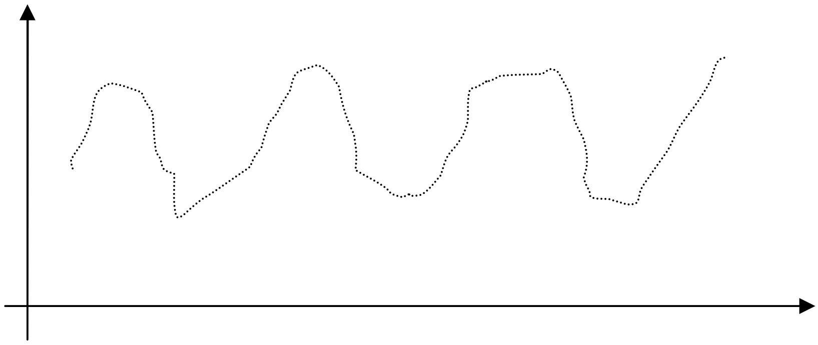
- Séries unimodales :

Histogramme :
ni effectifs



Une seule classe dominante
donc serie unimodale!

- Séries plurimodales :



- Si série symétrique :

On peut montrer que

$$\bar{x} = Me = Mo$$