Statistiques

« Objectiver un fonctionnement en quantifiant ses observations »

→ *économie, prévisions, aspect synthèse*

hubert.perrot@upmc.fr

Plan du cours :

I) Séries statistiques – Généralités
II) Paramètre de position et de dispersion
III) Ajustement de séries statistiques
IV) Indices statistiques
V) Séries statistiques à deux caractères
VI) Séries chronologiques

Chapitre 1. Séries statistique – Généralité

I) Définition

- Population : ensemble d'élément faisant l'objet d'une étude statistique.

Exemples :

- *Ensemble des élèves d'un amphi (étude sur l'âge, étude sur le lieu d'habitation, les notes …)*
- *Ensemble de pièces produites par une machine (étude sur la taille, qualité, couleur …)*
- *Ensemble des assurés d'une compagnie d'assurance (étude sur le nombre de voitures par assuré, le nombre d'accidents …)*

- Echantillon : partie de la population.

Exemples :

- *Les élèves du groupe B*
- *Pièces fabriquées un jour donné ou sélection de 100 pièces prises au hasard*
- *Assurés d'un département donné*

- Unité statistique : Un élément de la population étudié dont on veut examiner un ou plusieurs caractères. On parle aussi d'individu.

- Caractère :

  - qualitatif (nominale, ordinale) : non mesurable *(bon, mauvais; lieu d'habitation)*
  - quantitatif (discrète, continue) : mesurable à l'aide d'une variable statistique *(nombre d'individus, longueur d'une pièce fabriquée …)*

- Variable statistique :

  - discrètes : correspond à des valeurs isolées *(nombre d'individus)*
  - continues : correspond à des valeurs sur un intervalle *(longueur, …)*

- Séries statistique : ensemble des valeurs prises par une variable statistique sur l'ensemble de la population ou sur un échantillon.

II) Séries statistique d'un caractère quantitatif discret

Soit un échantillon de taille n (x_1, x_2, x_3, …, x_P) sont les valeurs prises par le caractère x, avec p le nombre de valeurs possible des x.

- Séries statistique : les valeurs prise par les n membres de l'échantillon.
- Effectif total : n
- Effectif partiel de x_i (fréquence absolue) : c'est le nombre de fois n_i qu'apparait la valeur x_i ($n_1 + n_2 + \ldots + n_p = n$).
- Fréquence relative de x_i : $f_i = \frac{n_i}{n}$ $(f_1 + f_2 + \ldots + f_p = 1)$
- Etude de la série : écart entre la plus grande et la plus petite des valeurs de x_i

Exemple :

On considère 100 familles *(l'échantillon)* de 4 enfants.
Le caractère retenu : nombre de garçons.

- Valeurs prises par x_i = {0, 1, 2, 3, 4}
- Effectif total : 100
- Série statistique : 3, 1, 3, 2, 4, 0, 1, 2, 3, 0, …

On compte les valeurs :

x_i	0	1	2	3	4	Total
n_i (effectifs partiels)	7	20	43	25	5	100

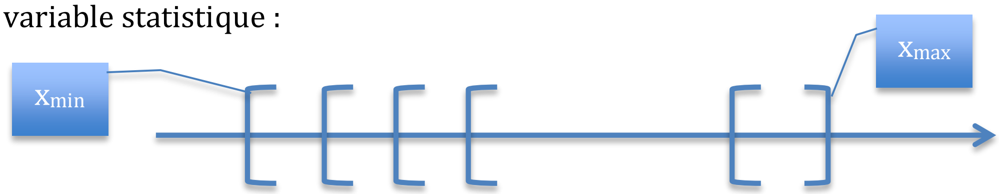
Fréquence relative de famille avec 2 garçons : $\frac{43}{100}$: 43 %

III) Séries statistique d'un caractère quantitatif continu

Le nombre p de valeurs prises par la variable statistique x_i est infini (en réalité c'est fini à cause de la précision des mesures → les f_i sont presque nulles).

- On va constituer les classes en divisant l'étendu de la série en un certain nombre d'intervalles

Donc si x_i variable statistique :



N° des classes :	**1**	**2**	**3**	**4**	**p**
Effectifs partiels :	n_1	n_2	n_3	n_4	n_p
Centre des classes :	x_1	x_2	x_3	x_4	x_p

Avec pour la classe i : un centre de classe xi et un effectif u_i.

Exemple : des nourrissons

Effectif total : 161

Pesées effectuée à 10 g près. Donné le poids entre 2,24 et 4, 45.

	2,2-2,5	2,5-2,8	2,8-3,1	3,1-3,4	3,4-3,7	3,7-4,0	4,0-4,3	4,3-4,6	total
Classe	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	161
Relatives f_i	3,1	6,8	14,9	24,8	26,1	12,4	8,1	3,7	99,9 %

IV) Séries statistiques d'un caractère qualitatif

On va grouper les résultats en autant de classes qu'il existe de modalités du caractère.

Exemple : étude statistique sur des fleurs avec 3 couleurs possibles :
- 3 classes : une classe par couleur avec un effectif par classe.

Deux types de variables :
- Variable ordinale : si les modalités peuvent être ordonnées (taille vestimentaire, préférences, …)
- Variable nominale : si elles ne peuvent être ordonnées (couleurs, profession, …)

Variable dichotomiques : si on a seulement deux modalités.

V) Représentation graphique des série statistiques

1) Séries à caractère discret :

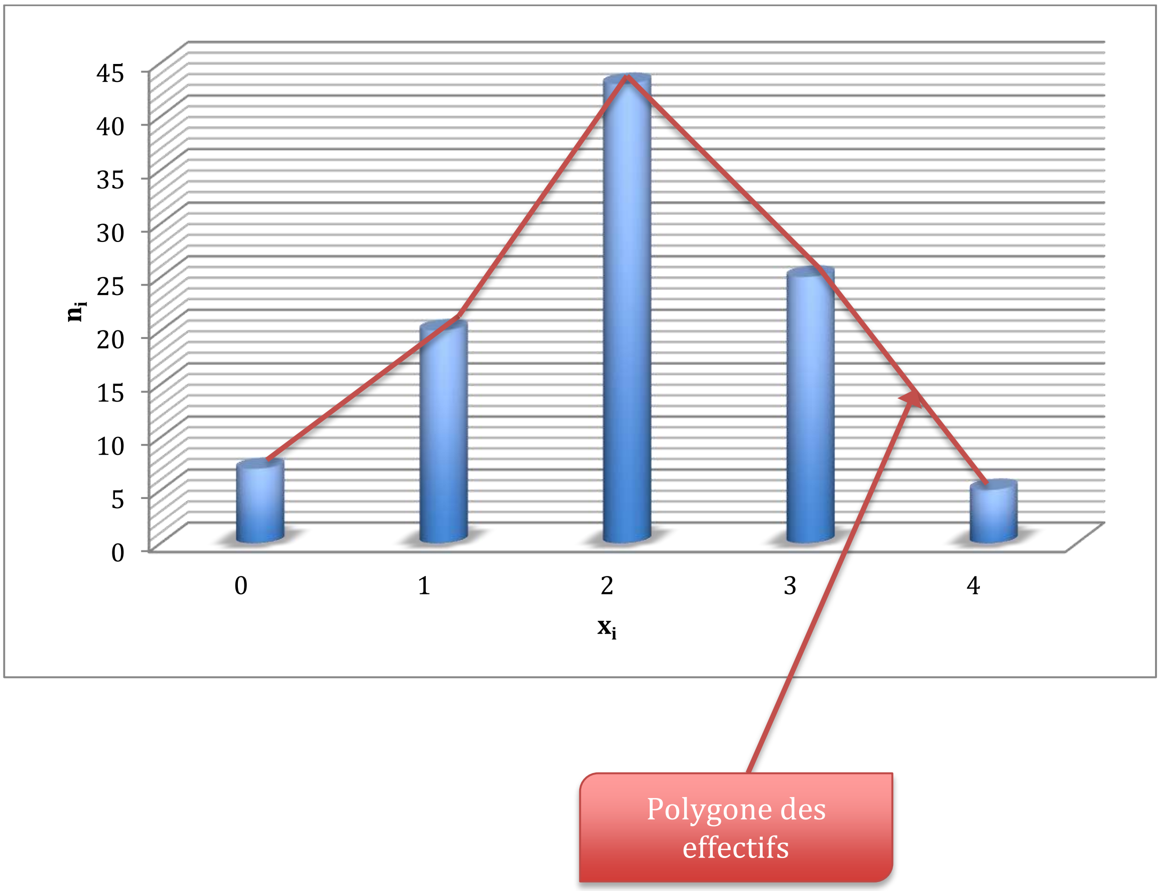
- Diagramme en bâtons : (des effectifs et des fréquences absolues)

On trace n_i en fonction de x_i ou f_i en fonction de x_i

Exemple des familles de 4 enfants : (x_i = nombre de garçons)

x_i	0	1	2	3	4	Total
n_i (effectifs partiels)	7	20	43	25	5	100

Cas discret



- <u>Polygone</u> (des effectifs et de fréquences absolues) : On joint l'extrémité des bâtons par des segments

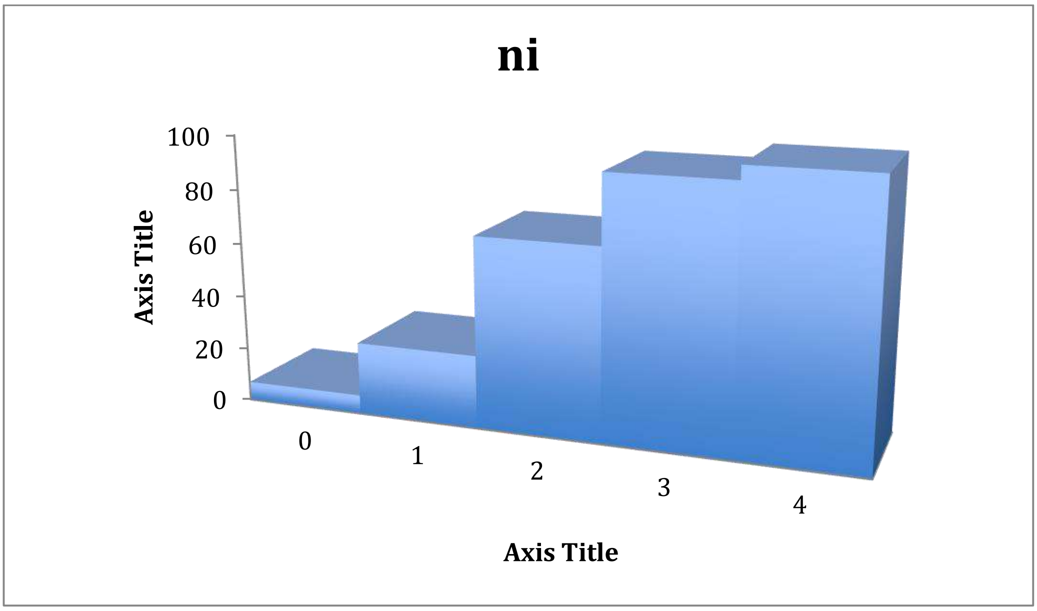
- <u>Polygone</u> (des fréquences relatives) : idem sauf que $n_i \rightarrow f_i$

- <u>Diagramme cumulé</u> :
    - <u>Effectif cumulé :</u> On va additionner les effectifs n_i, jusqu'à la $i^{ème}$ valeur x_i du caractère : $n_1+n_2+...+n_i$
    - <u>Fréquence relative cumulée :</u> idem sauf qu'on additionne le f_i

<u>Exemple :</u>

x_i	0	1	2	3	4	Total
Σn_i (effectifs cumulés)	7	27	70	95	100	100

Diagramme cumulé des effectifs



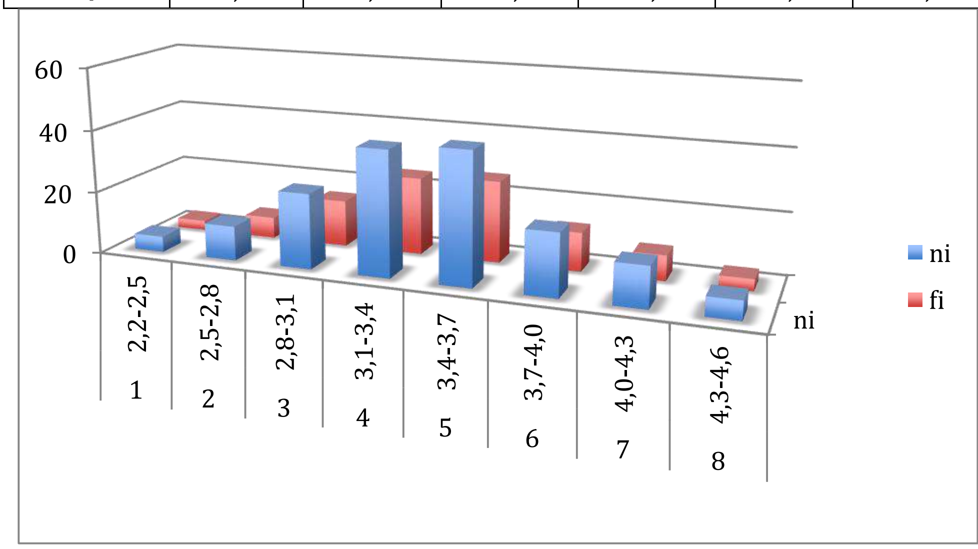
2) Caractère continu

Le diagramme en bâton est impossible car on a trop de valeurs de x_i :

- Histogramme :
    - Classer les égales : amplitude des classes identique

Exemple des nourrissons :

Intervalles	2,2-2,5	2,5-2,8	2,8-3,1	3,1-3,4	3,4-3,7	3,7-4,0	4,0-4,3	4,3-4,6
Classe	1	2	3	4	5	6	7	8
n_i	5	11	24	40	42	20	13	6
f_i	3,1	6,8	14,9	24,8	26,1	12,4	8,1	3,7



- Classes inégales

Exemple

Classes	5-6	6-7	7-8	8-10
n_i	12	13	16	6

N = 47

Problème d'amplitude pour la dernière classe : l'idéal est d'avoir quedes classes d'amplitude donc la classe [8-10] est divisée en 2 avec 6/2 = 3 comme effectif

D'où le nouveau tableau :

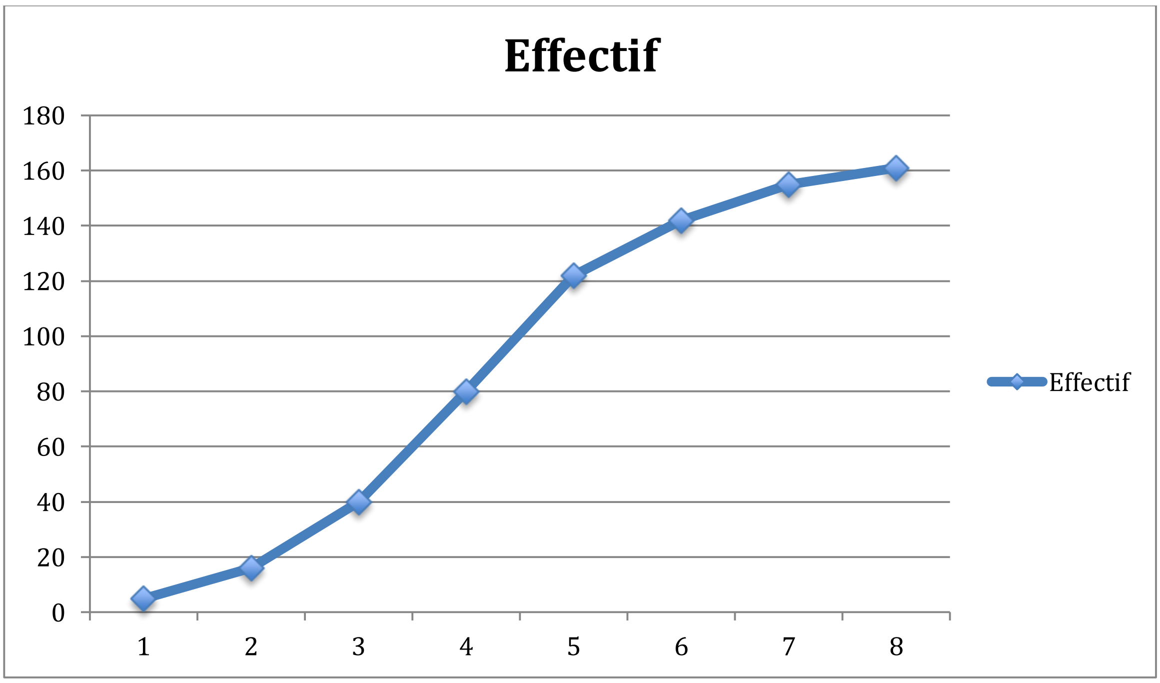
Classes	5-6	6-7	7-8	8-9	8-10
n_i	12	13	16	3	3

D'où l'histogramme

…

- <u>Polygones des effectifs et des fréquences relatives</u> :

Ligne brisée qui va rejoindre les milieux des sommets des rectangles



Chapitre 2 . Paramètre de position et de dispersion

Objectif : on veut obtenir des paramètres pour condenser l'information contenue dans une série statistique

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

I) Paramètre de position

Soit une série statistique prenant n valeurs $x_1, x_2, \ldots, x_n$

1) Moyenne arithmétique

On aura $\bar{x} = \frac{x_1+x_2+\cdots+x_n}{n} = \frac{1}{n}\sum_{(i=1)}^{n} n_i$

- Séries à caractère discret :

On a p valeurs des effectifs $n_1, n_2, n_3, \ldots, n_p$

Moyenne des $\bar{x}$ $= \frac{n1x1+n2x2+\cdots+npxp}{n} = \frac{1}{n}\sum_{(i=1)}^{p} n_i x_i = \frac{1}{n}\sum_{(i=1)}^{p} f_i x_i$
Remarque : on parle de moyenne pondérée par les f_i

- Séries à caractère continu :

On a p classes, d'effectifs n_i et de centre x_i

Moyenne des x $= \frac{n1x1+n2x2+\cdots+npxp}{n} = \frac{1}{n}\sum_{(i=1)}^{p} n_i x_i$ (x_i est le centre de classe n°i)
Même formule que dans le cas discret, on rend la variable x discrète en considérant que celle-ci n peut prendre comme valeur que le centre des classes.

Exemple :

Avec n = 161 (cas des nourrissons)
$\bar{x} = \frac{1}{n}\sum_{(i=1)}^{p} n_i x_i = \frac{1}{n}\sum_{(i=1)}^{p} f_i x_i$

Apres calcul on a $\bar{x} = 3{,}406\ kg$

Remarque : astuce de calcul !

On remarque que le centre x_i des classes est corrélé avec le n° des classes z_i
En effet, on peut écrire que $x_i = M + a.z_i$ (M = Valeur initiale)

On a $x_i = 2{,}35$ a = 0,3 et $z_i = 1$ → d'ou M = 2,05

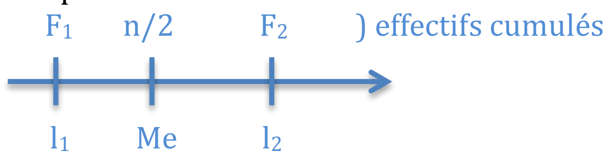
On a $\bar{x} = \frac{1}{n}\sum_{(i=1)}^{p} n_i x_i = \frac{1}{n}\sum_{(i=1)}^{p} n_i\,(M + a.z_i)$
$= M \times \frac{n}{n} + a \times \frac{1}{n}\sum_{(i=1)}^{p} n_i\, z_i) = M + a\bar{z}$

2) Médiane

On commence par ordonner les valeurs de la série et on prend la valeur qui sépare l'effectif en deux.

Notation = Me (Médiane)

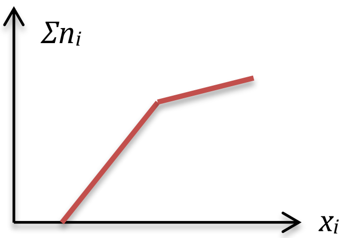
- Série à caractère discret :
  Deux exemples :
  3, 3, 4, 5, 7, 7, 9 → Me = 5 (3 valeurs au dessus, 3 valeurs en dessous)
  3, 3, 4, 5, 6, 7, 7, 9 → 5 < Me < 6 : Me = 5,5 par exemple
- Série à caractère continu :
  La série est déjà ordonnée par classes alors la médiane va tomber dans une certaine classe $[l_1, l_2[$:



où F_1 : effectif cumulé en l_1
F_2 : effectif cumulé en l_2
N : effectif total

$h_0 = F_2 - F_1$ = effectif de la classe $[l_1, l_2[$

On peut tracer le diagramme des effectifs cumulés :



On a tg $(\alpha) = \frac{h}{Me - l_1} = \frac{h_0}{l_2 - l_1} = \frac{F_2 - F_1}{l_2 - l_1}$

On a $\frac{n}{2} = F_1 + h$ et $h = \frac{F_2 - F_1}{l_2 - l_1} \times (Me - l_1)$

Donc $\frac{n}{2} = F_1 + \frac{Me - l_1}{l_2 - l_1} \times (F_1 - F_2)$

Exemple

Or $F_1 = 80$ et $F_2 = 122$

Donc $Me = 3{,}4 + (80{,}5 - 80) \times \frac{3{,}7-3{,}4}{122-80}$
$= 3{,}4 + 0{,}5 \times \frac{0{,}3}{42} = 3{,}503$ kg

remarque …

3) Quartiles

La médiane sépare la série en deux groupes de même effectif
Les quartiles séparent la série en quatre groupes de même effectif

- Cas des séries à caractère discret :

Exemples : 4, 5, 6, 11, 13, 14, 16.
$Q_1 = 5$, $Q_2 = 11$, $Q_3 = 14$
4, 5, 6, 7, 11, 13, 14, 15, 16.
$Q_1 = 6$, $Q_2 = 11$, $Q_3 = 14$