

Cours de statistiques descriptives

Georges Vinaver

georges.vinaver@univ-evry.fr

- **Bibliographie: Massoni, statistiques descriptives avec Excel, Vuibert**

programme

- 1. Vocabulaires des statistiques
- 2. Graphiques
- 3. Moyennes
- 4. Dispersion
- 5. Concentration, indice de Gini.
- 6. Statistiques doubles
- 7. Régression
- 8. Séries chronologiques

Deux types de statistiques:

- La statistique mathématique: tests modèles économétriques, etc.
- La statistique descriptive.

Objet de la statistique descriptive

a. organisation, présentation de données, représentations en tableaux et par graphiques

b. extraction d'informations pertinentes, synthèse numérique (moyennes, écart-type)

c. recherche de liens entre données (régression)

d. critique des modèles, interprétation des données et des calculs

e. décision

Vocabulaire des statistiques

- Ensemble étudié: population
- Éléments de cet ensemble: individus
- Objet de l'étude: caractère
- Valeurs prises par le caractère: modalités
- Ensemble des individus ayant même modalité ou groupe de modalités: classe

Types de caractères

- Qualitatif: non mesuré par un nom
- Quantitatif: mesuré par un nombre

- Présentation fonctionnelle:
- Soit A une population, X un caractère et i un individu. On note $X(i)$ ou x_i la valeur du caractère X pour l'individu i . Le caractère X est une application de l'ensemble des modalités.

- Exemples de caractères qualitatifs
 - a. couleur d'une voiture dans un parking
 - population: les voitures du parking
 - caractère: la couleur
 - modalités: bleu, vert, etc.
 - b. Décision finale à un examen.
 - population: un amphithéâtre
 - individus: étudiants
 - caractère: décision
 - modalités: ajourné, passable, AB,

Exemples de caractère quantitatifs

a. température journalière d'un m
population: les jours de maladie
modalités: [36,41]

b. température des malades d'un
hôpital un jour j donné.
population: les malades hospitalisé
jour j
modalités: [36,41]

c. note à l'examen de statistiques
étudiants de votre amphitheâtre.
population: étudiants de l'amphi
modalités: [[0 ; 20]]

Tableaux de données individuelles

- (caractère qualitatif ou quantitatif)
- On met en colonne les individus i et caractères x_i .

individu	valeur
1	x_1
2	x_2
3	x_3
.....

Tableau par valeurs

- En colonne 1, on liste les modalités; en colonne 2, les effectifs, en colonne 3, les fréquences. On peut aussi mettre en colonne 4 les fréquences cumulées croissantes et en colonne 5 les effectifs cumulés croissants.

valeur	effectif	fréquence	cumu
y_1	n_1	f_1	F_1
y_2	n_2	f_2	F_2
y_3	n_3	f_3	F_3
.....
y_p	n_p	f_p	100%
total	n	100%	

Formules

$$\sum n_i = n, \sum f_i = 100\% = 1$$

- Fréquences: $f_i = n_i/n$

- Cumul croissant: $F_i = f_1 + f_2 + \dots + f_i =$

Statistiques regroupées par intervalles

- Pour les caractères quantitatifs, on peut avoir :
 - Une présentation individuelle
 - Une présentation par valeurs (perte d'information, mais on a un gain de lisibilité)
 - Une présentation par intervalles: on regroupe plusieurs valeurs dans un intervalle. En général, on prend des intervalles ouverts à gauche et fermés à droite : $[a, b[$. Toutefois, le dernier intervalle peut être fermé à droite.
- Ainsi, une extrémité d'intervalle n'appartient qu'à un seul intervalle.

- Un intervalle dont on relève l'effectif s'appelle une classe.
- Soit la classe

$$[a, b[$$

Le nombre $b-a$ est son amplitude.

Le nombre $(a+b)/2$ est son centre

Soit n son effectif. On appelle densité effectif corrigé de la classe le nombre $n/(b-a)$

Exemple

Statistique par valeurs

note	eff
0	2
1	4
2	5
3	7
4	12
5	11
6	3
7	5
8	2
9	1
10	1
	53

La même, avec regroupement
par intervalles

$[0,3[$	11
$[3,6[$	30
$[6,10]$	12
	53