

Statistique Descriptive

Chapitre I : Vocabulaire

Anas KNEFATI

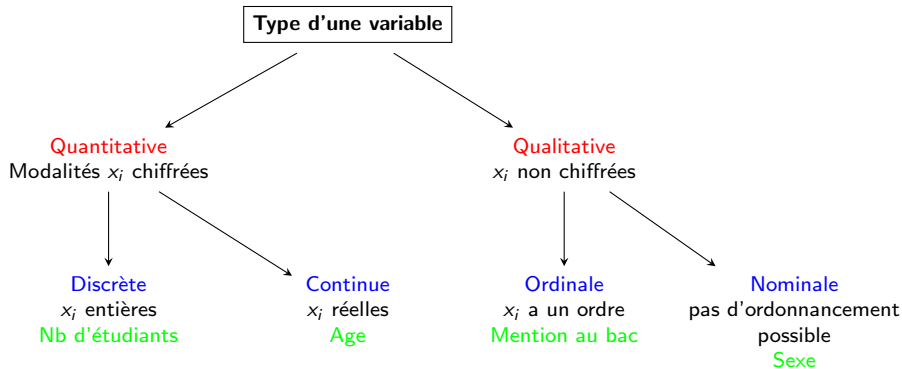
Université Rennes 2



- 1 Terminologie
- 2 Statistique descriptive
- 3 Présentation des données
- 4 Fréquence et fréquence cumulée

- 1 Terminologie
- 2 Statistique descriptive
- 3 Présentation des données
- 4 Fréquence et fréquence cumulée

- **Statistique** : Ensemble des données numériques concernant un phénomène quelconque et dont on tire certaines conclusions..
- **Population** : Ensemble que l'on observe.
 - ▶ **Individu** : Élément de la population.
 - ▶ **Taille de la population** : Nombre de ses individus.
- **Échantillon** : Sous-ensemble de la population.
- **Variable statistique (caractère)** : Critère d'observation.
 - ▶ **Modalités d'une variable** : Ses valeurs possibles.
- **Série statistique** : Ensemble d'observations d'une ou plusieurs variables faites sur une population ou un échantillon.



Plan

- 1 Terminologie
- 2 Statistique descriptive**
- 3 Présentation des données
- 4 Fréquence et fréquence cumulée

Objectif

- Résumer l'information contenue dans la série statistique.
- Suggérer des hypothèses relatives à la population

Objectif

- Résumer l'information contenue dans la série statistique.
- Suggérer des hypothèses relatives à la population

Outils utilisé

- Graphiques (Histogramme, box-plot,...), Ch2.
- Indicateurs (Moyenne, Variance, ...), Ch3.
- Tableaux (Table de fréquence, de contingence, ...), Ch2, Ch4 et Ch6.

Objectif

- Résumer l'information contenue dans la série statistique.
- Suggérer des hypothèses relatives à la population

Outils utilisé

- Graphiques (Histogramme, box-plot,...), Ch2.
- Indicateurs (Moyenne, Variance, ...), Ch3.
- Tableaux (Table de fréquence, de contingence, ...), Ch2, Ch4 et Ch6.

Type de l'outil

Dépend de :

- Nature de la série (uni ou multidimensionnelle).
- Nature de la variable (quantitative continue ou discrète, ...)

Plan

- 1 Terminologie
- 2 Statistique descriptive
- 3 Présentation des données**
- 4 Fréquence et fréquence cumulée

Présentation des données

Données brutes : On dispose de toutes les observations de la variable étudiée.

Pays	Superficie	Population	Appartenance.CEE
Allemagne	357.00	80	O
Autriche	83.80	7,6	N
Belgique	30.50	9,9	O
Danemark	43.10	5,1	O
Espagne	505.00	39,2	O
Finlande	337.00	4,9	N
France	552.00	56,5	O
Grèce	132.00	10	O
Irlande	70.30	3,5	O
Islande	103.00	0,3	N
Italie	301.00	58	O
Luxembourg	3.00	0,4	O
Norvège	324.00	4,2	N
Pays-Bas	33.90	14,9	O
Portugal	92.10	10,6	O
Royaume-Uni	244.00	57	O
Suède	450.00	8,5	N
Suisse	41.30	6,7	N

Table : Europe occidentale en 1992

Données agrégées : Dans le cas où la variable est

Qualitative ou discrète : une classe contient tous les individus ayant la même modalité

Classe	N	O
Effectifs	6	12

$$n_1 = 6 \text{ et } n_2 = 12, \text{ effectif total} = n_1 + n_2 = 18$$

continue : Une classe est un intervalle.

Classe	[0,100[[100,200[[200,350[[350,600[
Effectifs	8	2	4	4

$$n_1 = 8, n_2 = 2, n_3 = 4, n_4 = 4, n_5 = 4.$$

$$\text{Effectif total} = \sum_{i=1}^5 n_i = n_1 + n_2 + n_3 + n_4 + n_5 = 18$$

Plan

- 1 Terminologie
- 2 Statistique descriptive
- 3 Présentation des données
- 4 Fréquence et fréquence cumulée**

Fréquence et fréquence cumulée

- K classes.
- Table des fréquences et des fréquences cumulées :

N° classe	Effectifs (n_i)	Fréquences (f_i)	Fréquences cumulées (F_i)
1	n_1	f_1	F_1
2	n_2	f_2	F_2
\vdots	\vdots	\vdots	\vdots
K	n_K	f_K	F_K
Total	n	1	

- n_i : Effectif de la classe numéro (i)
- Effectif total : $n = n_1 + n_2 + \dots + n_K := \sum_{i=1}^K n_i$
- Fréquence de la $i^{\text{ème}}$ classe : $f_i = \frac{n_i}{n}$
- Fréquence cumulée de la $i^{\text{ème}}$ classe = Somme des fréquences précédentes, cumulées de 1 à i , c-à-d : $F_i = \sum_{j=1}^i f_j$

Fréquence et fréquence cumulée

Exemple : Superficie des pays de l'europe occidentale en 1992

Classe	Effectifs	f_i	F_i
[0; 100[8	0.44	0.44
[100; 200[2	0.11	0.55
[200; 350[4	0.22	0.77
[350; 600[4	0.22	0.99 \approx 1
Total	18	0.99 \approx 1	

Fréquence et fréquence cumulée

Exemple : Superficie des pays de l'europe occidentale en 1992

Classe	Effectifs	f_i	F_i
[0; 100[8	0.44	0.44
[100; 200[2	0.11	0.55
[200; 350[4	0.22	0.77
[350; 600[4	0.22	0.99 \approx 1
Total	18	0.99 \approx 1	

Remarques

- On peut remplacer f_i par $f_i * 100$ qui représente un pourcentage
- $0 \leq f_i \leq 1$
- $\sum_{i=1}^K f_i = 1$
- Fréquence cumulée n'est pas définie pour une variable qualitative nominale.