

R2.08 - Statistique descriptive Cours 0 - Introduction

A. Ridard

A propos de ce document

- Pour naviguer dans le document, vous pouvez utiliser :
 - le menu (en haut à gauche)
 - l'icône en dessous du logo IUT
 - les différents liens
- Pour signaler une erreur, vous pouvez envoyer un message à l'adresse suivante :
anthony.ridard@univ-ubs.fr

Plan du cours

1 Individus et variables

2 Tableau de données

La **statistique** est la science qui s'occupe du recueil¹ et de l'**analyse des données** issues de l'observation de phénomènes **aléatoires**².

Elle permet de **décrire, expliquer, faire des prévisions** et **prendre des décisions**. C'est un outil essentiel pour la compréhension et la gestion des phénomènes³ complexes.



En résumé

Les **méthodes statistiques** sont utilisées pour **analyser les variations** observées dans le but de **maitriser au mieux l'incertitude** au niveau des données !

-
1. Cette notion sort du cadre de ce cours
 2. En opposition aux phénomènes déterministes
 3. Issus de l'économie à la biologie en passant par la psychologie, et bien sûr les sciences de l'ingénieur

Plus précisément, ces méthodes statistiques relèvent de :

- la **statistique descriptive** qui a pour but de *décrire* et d'*expliquer* les données, à l'aide de tableaux, de graphiques et de résumés numériques.
- la **statistique inférentielle** dont l'objectif est de *faire des prévisions* et de *prendre des décisions*, en contrôlant le risque d'erreur.



Au programme et à la mode !

- La statistique inférentielle (BUT 3C) repose sur les Probabilités (BUT 2)
- Des **méthodes algorithmiques**^a complètent la panoplie du statisticien^b

a. Plus connues aujourd'hui, grâce à l'essor de l'informatique, sous le nom de *machine learning*

b. Ou encore *data scientist*

1 Individus et variables

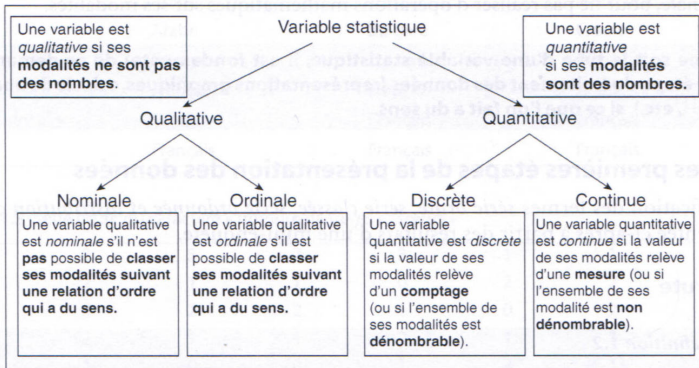
2 Tableau de données

En statistique, une **population**⁴ désigne un ensemble d'objets équivalents.

Ces objets sont appelés des **individus** ou **unités statistiques**, et ils sont décrits par un ensemble de caractéristiques appelées **variables**.

4. Ce terme est hérité des premières applications de la statistique à la démographie

Les variables étudiées peuvent être classées selon leur nature :



- 1 Individus et variables
- 2 **Tableau de données**

En général⁵, la population est trop vaste pour pouvoir être observée de manière exhaustive, on ne dispose que d'une partie appelée **échantillon**.

Un échantillon de n individus caractérisés par p variables est présenté sous la forme d'un tableau à n lignes et p colonnes :

id_etud	sexe	was_boursier	type_lycee	dpt_lycee	type_bac	mention_bac	classe_moy_bac	moy_bac	lv_1	lv_2
1	M	non	public	56	BG	TB	[16, 20]	18	19,18	12,94
2	M	non	prive	56	BG	TB	[16, 20]	18	19	14,6
3	M	non	prive	99	BG	TB	[16, 20]	18	16,99	14,73
4	M	non	public	56	BG	TB	[16, 20]	18	14,98	16,1
5	M	oui	public	28	BT	TB	[16, 20]	18	17,08	14,41
6	M	non	public	76	BG	B	[14, 16[15	12,92	10,35
7	M	non	public	56	BG	B	[14, 16[15	16,8	11,89
8	M	non	prive	35	BG	B	[14, 16[15	10,38	12,45
9	M	oui	public	75	BG	B	[14, 16[15	14,59	12,55
10	M	oui	public	56	BG	B	[14, 16[15	15,01	13,1

5. Sauf s'il s'agit d'un recensement