

# Cours d'Éléments de Statistique

Jean Christophe Meunier

## MODULE 1

### De la récolte de données à l'élaboration d'une base de données

1<sup>ère</sup> Bac, Commerce Extérieur  
Année Académique 2015-2016



## A. Notions de base

# I. Unité statistique et variable

## 1. Unité statistique (ou observation)

- Egalement appelée : sujet, individu, observation...
- Unité 'observée' (humain, animal, ...) sur laquelle des données sont récoltées (ex : sondage...)
- La statistique s'opère sur un ensemble d'unités statistiques :
  - Population : ensemble d'unités statistiques appartenant à une même catégorie (ex : belge,...)
  - Echantillon : sous-groupe de la population
- N'a de sens que comme faisant partie d'un groupe (population ou échantillon)
  - Pas d'analyse statistique possible sur une unité statistique 'isolée'

3

# I. Unité statistique et variable

## 2. Variable (ou critère)

- Propriétés de l'unité statistique qui sont étudiées
- Données récoltées sur le sujet (ex: taille, poids, revenus...)
- Peut être qualitative ou quantitative

4

# I. Unité statistique et variable

- Organisation des données dans un tableau
  - En vue d'analyser les données : statistique descriptive ou inférentielle
  - Tableau à double entrée
    - Lignes : Unités statistiques (individus, sujets,...)
    - Colonnes : variables

Variables	V1	V2	V3	...
<b>Unités statistiques</b>				
<b>1</b>				
<b>2</b>				
<b>3</b>				
<b>...</b>				5

# I. Unité statistique et variable

- Ex : différentes caractéristiques recueillies pour chacun des élèves

Prénom	Sexe	Taille (cm)	Nombre de frères et sœurs	Résultat test d'anglais
Lara	F	168	1	D
Chloé	F	163	2	D
Flora	F	161	4	B
Sara	F	165	1	B
Kim	G	165	0	A
Lella	F	168	2	B
Sandra	F	161	2	D
Pauline	F	163	1	C
Anne	F	165	3	A
Laurene	F	165	1	C
Lucas	G	170	3	C
Quentin	G	175	1	D
Valentine	F	168	1	A
Clément	G	165	0	B
Hugo	G	170	1	C
Marine	F	170	0	C
Valentin	G	170	1	B
Aline	F	170	1	B
Allison	F	165	1	C
Nadia	F	163	0	D
Rudy	G	163	1	A
Jeannot	G	170	4	C
Pedro	G	175	2	B
Ahmed	G	175	4	A
Pierre	G	180	2	C
Michael	G	170	2	C
Alexandre	G	175	5	B
Loick	G	175	1	A
Antoine	G	180	0	C
Thomas	G	180	1	B
Ludovic	G	180	1	A
Karim	G	182	1	C
Paul	G	175	3	B
Maxime	G	182	2	C
Vim	G	180	0	D

Connaissance de l'anglais  
 A : Très bonne connaissance  
 B : bonne connaissance  
 C : Connaissance moyenne  
 D : Connaissance médiocre  
 E : Aucune connaissance

# I. Unité statistique et variable

- Ex : idem sur tableur Excel

	Prénom	Sexe	Taille (cm)	Nombre de frères et soeurs	Résultat test d'anglais		
1	Lara	F	168	1	D		
2	Chloé	F	163	2	D		
3	Flora	F	161	4	B		
4	Sara	F	165	1	B		
5	Kim	G	165	0	A		
6	Leila	F	168	2	B		
7	Sandra	F	161	2	D		
8	Pauline	F	163	1	C		
9	Anne	F	165	3	A		
10	Laurène	F	165	1	C		
11	Lucas	G	170	3	C		
12	Quentin	G	175	1	D		
13	Valentine	F	168	1	A		
14	Clément	G	165	0	B		
15	Hugo	G	170	1	C		
16	Marine	F	170	0	C		
17	Valentin	G	170	1	B		
18	Aline	F	170	1	B		
19	Alison	F	165	1	C		

7

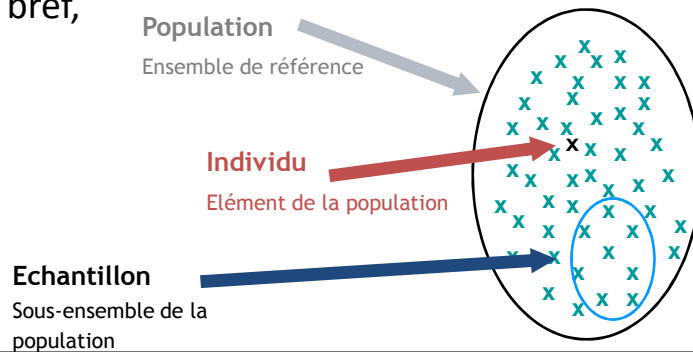
# II. Echantillon VS. Population

- Population :
  - Ensemble de référence : ensemble des individus qui partagent les mêmes caractéristiques
  - Représente **TOUS** les individus qui partagent ces caractéristiques
  - Doit être définie avec précision
    - Ex : tous les humains, tous les habitants de WSP, tous les étudiants en bac 1 'commerce extérieur' à l'EPFC,...
- Echantillon :
  - Sous-ensemble de la population (tiré par une méthode qui assure la représentativité de la population)
  - Public cible de la statistique inférentielle : plus facilement appréhendable qu'un échantillon

8

## II. Echantillon VS. Population

- Unité statistique = Individu
  - Du latin : 'indivisible'
  - Synonymes : Sujets, observations,...
  - Être humains, êtres vivants, objets concrets ou abstraits
- En bref,



## III. Types de variables

### 1. Qualitative vs. Quantitative

- Qualitative : information sur le sujet qui n'est pas mesurable ni repérable :
  - Ex : nationalité, sexe,...
- Quantitative : peut être mesurée
  - Ex : âge, revenu mensuel,...

### III. Types de variables

#### 2. Discrète vs. Continue

- Discrète : peut prendre un nombre limité de valeurs
  - variable qualitative :
    - choix entre différentes valeurs associées à différentes catégories ou à un classement
    - Ex statut marital : 1. Célibataire – 2. Marié – 3. Divorcé/séparé
  - variable quantitative :
    - choix entre différentes valeurs généralement 'entières' (pas de valeurs intermédiaires)
    - Ex variables de dénombrement : nombre d'enfants, nombre d'employés...
- Continue : peut prendre une infinité de valeurs
  - Toutes les valeurs réelles comprises entre les bornes inférieures et supérieures
    - Ex poids, taille, durée de vie, intervalle de temps,...

Rem : lorsque le nombre de modalités d'une variable discrète est important, la variable discrète peut être assimilable à une variable continue

11

### IV. Echelles de mesure

#### 1. Echelle nominale (variables qualitatives)

- Différentes catégories sans relation d'ordre/d'intensité
  - Ex : Profession : 1. Salarié, 2. Indépendant, 3. Sans emploi
  - Cas particulier : Dichotomique (choix entre deux réponses) : Oui/non, Vrai/faux, Homme/femme

12

## IV. Echelles de mesure

### 2. Echelle ordinale (variables qualitatives)

- Classement/ordre de préférence : Relation d'ordre/d'intensité mais différence entre modalités de réponse non-mesurable

- Ex : classer loisirs suivants par ordre de préférence :

Sport ..3.. Lecture ..5.. Cinéma ..2.. Sorties ..7.. Théâtre ..6.. Concert ..4..

13

## IV. Echelles de mesure

### 3. Echelle d'intervalle (variables quantitatives)

- Relation d'ordre/d'intensité et différence entre modalités de réponse mesurable

- Pas de relativité entre les modalités de réponse

- Pas de ratio entre modalités de réponse
- Zéro n'existe pas ou est une modalité comme les autres

- Ex : Température, heure,...

- On peut dire qu'il y a 8 heures de différence entre 8h et 16h mais on ne peut pas dire que 16h est deux fois plus tard que 8h

- Ex : Echelle de Likert

- Sur échelle de 1 à 7, évaluer dans quelle mesure vous êtes favorable à :

1	2	3	4	5	6	7
Très défavorable					Très favorable	

14

## IV. Echelles de mesure

### 4. Echelle de rapport (variables quantitatives)

- Relation d'ordre/d'intensité et différence entre modalités de réponse mesurable
- Relativité entre les modalités de réponse
  - Zéro représente l'absence de la caractéristique mesurée
  - Calcul de ratio possible entre les scores de deux individus
    - Calcul du ratio en divisant le score d'un sujet par celui d'un autre (permet d'estimer les valeurs relatives)
- Ex : quantité (argent, poids,...)
  - Avec 12000 euros, Jack est deux fois plus riche que Jimmy qui n'a que 6000 euros

15

## V. Variables vs. échelles de mesure

Échelle de mesure	Variable	
<b>Nominale</b> - Différentes catégories	Qualitative	Discrète
<b>Ordinale</b> - Classement/ordre de préférence	Qualitative	Discrète
<b>Intervalle</b> - Différence mais pas de relativité - Zéro inexistant ou non significatif	Quantitative	Discrète ou Continue
<b>Rapport</b> - Différence et relativité - Zéro = absence de la variable considérée	Quantitative	Discrète ou Continue

16



## Exercice

- Pour chacune des questions précisez le type de variable et l'échelle de mesure

Questions
1. Selon vous, quelle est la température idéale pour faire du sport ?
2. Quel est votre statut socio-professionnel (1. Employé ; 2. Indépendant ; 3. Sans emploi) ?
3. Vers quelle heure (en heure 'pile') vous levez vous le matin ?
4. Quel âge avez-vous ?
5. Sur une échelle de 1 à 10, évaluez la qualité du produit X.
6. Etes-vous un homme/une femme ?
7. Combien d'employés compte votre entreprise ?
8. Classer les véhicules suivants par ordre de taille : moto, voiture, camionnette, camion.
9. Dans quelle branche exercez-vous votre profession (1. Sciences humaines ; 2. Sciences économiques ; 3. Sciences exactes) ?
10. En combien de temps courez-vous le 100 mètres ?

17

## Exercice

Questions	Type variable		Échelle de mesure
1. Selon vous, quelle est la température idéale pour faire du sport ?	Quantitative	Continue	Intervalle
2. Quel est votre statut socio-professionnel (1. Employé ; 2. Indépendant ; 3. Sans emploi) ?	Qualitative	Discrète	Nominale
3. Vers quelle heure (en heure 'pile') vous levez vous le matin ?	Quantitative	Discrète	Intervalle
4. Quel âge avez-vous ?	Quantitative	Continue	Rapport
5. Sur une échelle de 1 à 10, évaluez la qualité du produit X.	Quantitative	Discrète	Intervalle
6. Etes-vous un homme/une femme ?	Qualitative	Discrète	Nominale dichotomique
7. Combien d'employés compte votre entreprise ?	Quantitative	Discrète	Rapport
8. Classer les véhicules suivants par ordre de taille : moto, voiture, camionnette, camion.	Qualitative	Discrète	Ordinale
9. Dans quelle branche exercez-vous votre profession (1. Sciences humaines ; 2. Sciences économiques ; 3. Sciences exactes) ?	Qualitative	Discrète	Nominale
10. En combien de temps courez-vous le 100 mètres ?	Quantitative	Continue	Rapport

18

## VI. Séries

### 1. Séries statistiques

- Ensemble d'observations dans l'espace à un moment donné

- une même variable (ex: poids) est récoltée auprès d'un ensemble d'individus

Variables	V1
Unités statistiques	
1	
2	
3	
...	

### 2. Séries chronologiques

- Ensemble d'observations dans le temps

- une même variable récoltée plusieurs fois (à intervalle régulier ou non) auprès d'un même individu/ensemble d'individus

Variables	V1 tps1	V1 tps2	V1 tps 3	...
Unités statistiques				
1				
...				

19

## B. Récolte de données

20

## I. Objectif de la récolte

- Recueil d'informations destiné à étudier un phénomène particulier :
  - Préférences du consommateur
  - Satisfaction par rapport à un service ou produit
  - Profil du client
  - Gestion des ressources humaines
  - ...

21

## II. Qualité d'une récolte

- Se définit par :
  - Objectivité :
    - Reflète objet d'étude et pas désir de l'expérimentateur : Biais de confirmation d'hypothèse (ex : test médicament)
  - Pertinence
    - En rapport avec l'objectif
  - Précision
    - Ex : I-Pad vs. Tablette
  - Fiabilité
    - Représentativité de l'échantillon
  - Actuelle

22

### III. Origine des données

- Données primaires
  - Données spécifiquement récoltées pour l'étude
  - L'expérimentateur récolte lui-même données
- Données secondaires
  - Déjà collectées et ré-exploitées pour une seconde utilisation
  - Ex : données de recensement, institut d'études de marché...

23

### IV. Méthode de récolte

#### 1. Le questionnaire (sondage, survey)

- Papier-crayon, en ligne, par téléphone, poste...
- Questions ouvertes (liberté de réponse) ou fermées (choix entre plusieurs propositions)
- Questionnaire préexistant ou nouvellement créé
- Facilité de passation : échantillon important
- Objectifs divers : comparaisons entre groupes, relations entre variables...

24

## IV. Méthodes de récolte

### 2. L'observation

- 'Enregistrer' l'activité de l'individu dans son cours normal et sans volonté de la modifier
  - Ex : attrait pour quelle vitrine dans rue commerçante
- Sans biais : le sujet ne se sait pas observé
  - Comportement tout à fait naturel
  - Caméra, vitre sans teint,...
  - Mais question éthique (violation vie privée) !!!
- L'observation doit être rigoureuse et systématique
  - Définition du comportement identique pour tout le monde (gestes, temps passé devant rayon,...) → Grille d'observation avec unité de codage

25

## IV. Méthodes de récolte

### 2. L'observation

- Ex grille d'observation : potentiel d'influence

	Très souvent	Moyennement	Presque jamais
<b>Influence directe (Production)</b>			
- Prend des initiatives en rapport direct avec la tâche (par ex., propose des idées nouvelles)			
- Parle souvent			
- Argumente longuement sa position			
- Maintient une position cohérente			
- Manifeste beaucoup d'assurance (au moins apparente)			
- Se présente plus ou moins comme un expert de la question traitée (fait état de ses connaissances ou de son expérience)			
- Se range à l'opinion des autres			
- Parle très peu			
<b>Influence indirecte (Facilitation)</b>			
- Se soucie de faire s'exprimer chacun			
- Fait le point, résume ce qui a été fait jusque là			
- Propose une manière de procéder, une méthode de travail (ne pas confondre ces deux comportements !)			

26

## IV. Méthodes de récolte

### 3. L'expérimentation

- A partir d'une situation naturelle, faire varier l'un ou l'autre paramètre
- Plan expérimental : séquence d'une expérimentation
  - Ex : attitude du consommateur // nouveau produit
    - 1. Rien
    - 2. Changer le goût
    - 3. Changer la texture
    - 4. Changer l'emballage
    - ...
  - Ex : effet d'une promotion sur un produit
    - Nombre d'achats avant, pendant et après la promotion
- Doit être appréhendé à partir d'une mesure quantifiable (changement d'attitude sur une échelle d'évaluation, nombre d'achats...)

27

## C. Cas pratique : élaboration d'un questionnaire

28

## I. Le questionnaire permet d'étudier

- Ce que les gens sont
  - Caractéristiques/profil du client (variables socioculturelles)
- Ce que les gens font
  - comportements de consommation, d'utilisation et d'achat
- Ce que les gens pensent
  - jugement global, préférence, satisfaction, intentions
- Comment les gens se décident
  - motivations et freins :
    - psychologiques (caractère hédoniste, rationnel,...), financiers (pouvoir d'achat ,...), techniques (compétence à utiliser le produit,...)
  - critères de choix entre les marques, sources d'information et de conseils
  - degré de préméditation de l'achat : réflexion avant l'achat

29

## I. Le questionnaire permet d'étudier

- Comportement:
  - acheter ou non
- Opinions :
  - pour ou contre banque X
- Attitude :
  - rapport à l'argent
- Intentions/Motivation :
  - épargne pension
- Attente :
  - par rapport à un produit/service
- Caractéristiques/profil du sujet
  - Age, sexe, profession ? (données sociodémographiques)

30

## II. Structure d'un questionnaire

### 1. Introduction

- Paragraphe décrivant objet de l'étude
- Données signalétiques/variables sociodémographiques

### 2. Corps du questionnaire

- Comportements, attitudes, opinions, intentions, motivations, attentes...
  - Ce que les gens 'sont', 'font', 'pensent', 'comment ils se décident',...

31

## II. Structure d'un questionnaire

- Exemples de données signalétiques (variables sociodémographiques) :
  - Âge, sexe
  - Statut marital : marié(e), concubinage, séparé(e)/divorcé(e), veuf(ve),...
  - Niveau d'études : primaire, secondaire inférieur/supérieur, supérieur non-universitaire, supérieur universitaire
  - Statut professionnel : employé, indépendant, ss emploi
  - Composition de famille : nombre d'enfants (âge et sexe)
  - Habitation : zone rurale/urbaine, maison/appartement,...

32



### III. Élaboration en 4 étapes

#### 1. Cibler l'objet d'étude

- Répertorier les éléments d'informations souhaités
  - comportement, attitude, opinion, caractéristique...
- Avant d'élaborer les questions
- Définir la visée de l'étude : exploratoire ou confirmatoire
  - Ex exploratoire : quel est le profil du consommateur ?
  - Ex confirmatoire (pour valider/confirmer une hypothèse) : le 'pro-écologie' est-il plus regardant à sa consommation d'énergie ?

33

### III. Élaboration en 4 étapes

#### 2. Création des questions (1)

- Soigner la formulation pour une bonne compréhension
  - Formulation correcte et précise
    - Intérêt pour l'opéra chez les jeunes : 'Aimez-vous la musique ?'
    - Avez-vous peur des animaux ? Lapin OK, Lion KO
  - Éviter équivoque/ambiguïté
  - Éviter termes complexes ou techniques
- Choix du type de question
  - Questions ouvertes
    - Qualitativement riches mais dépouillement difficile (analyse de contenu, création de catégories)
  - Questions fermées
    - Facilité de passation et de traitement de données mais moins riches en information

34

## III. Élaboration en 4 étapes

### 2. Création des questions (2)

- Eviter biais de réponses
  - Biais de désirabilité sociale/conformisme :
    - Tendance à donner réponses socialement acceptables
      - » Surtout sur sujets 'sensibles' : FN, Emprunte écologique,...
    - Mesures à prendre pour contrer biais :
      - » Questions neutres et non-jugeantes : ménager la 'sensibilité'
  - Effet de halo :
    - Contamination des réponses les unes par les autres
      - » Par organisation logique de la pensée : tendance à répondre de la même manière à des questions similaires
    - Mesure à prendre pour contrer biais
      - » Randomiser l'ordre des variables
      - » inverser le sens des échelles

35

## III. Élaboration en 4 étapes

### 3. Pré-test

- Soumettre questionnaire à un petit nombre
- Intérêt du pré-test = évaluer la qualité du questionnaire
  - Longueur du questionnaire et durée de passation
  - Compréhension et formulation : vocabulaire, pertinence des questions, consignes données
  - Questions filtres, modalités des réponses
  - Structure du questionnaire
- Récolter feedbacks et suggestions
- Corriger

36

### III. Élaboration en 4 étapes

#### 4. Administration du questionnaire

- Préserver anonymat
- Expliquer contexte/but de l'étude
- Choix du mode de passation
  - Papier-crayon
  - En ligne
  - Par téléphone
  - Envoi postal

37

### Exercice

- Elaboration d'un questionnaire
  - Définissez l'objet d'étude
    - Répertorier les éléments d'informations souhaités
    - Définir la visée de l'étude : exploratoire/confirmatoire
  - Demandez trois 'données sociodémographiques' pertinentes par rapport à l'objet d'étude
    - Age, sexe, ...
  - Posez cinq questions pertinentes par rapport à l'objet d'étude
    - Ce que les gens 'sont', 'font', 'pensent', 'comment ils se décident',...

38