


---

# ANALYSER DES DONNÉES

Nadine Mandran – Juin 2022  
Laboratoire Informatique de Grenoble

---



N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Ox

1

---

---

---

---

---

---

---

---


---

# LIVE ANALYSIS

---

- Si l'analyse de données était un animal, elle serait ....

---



N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Ox

2

---

---

---

---

---

---

---

---


---

# L'ANALYSE ET VOUS

---

- Quelles difficultés rencontrées vous ?
  - En quali ?
  - En quanti ?

---



N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Ox

3

---

---

---

---

---

---

---

---

# ANALYSER

- L'analyse des données est la fin d'un cycle
- Problématique, Question de recherche, Contribution, Protocole expérimental, Production des données, ANALYSE, Conclusion
- Une bonne analyse c'est avant tout de bien poser le problème et de produire correctement des données ...

CC BY-NC-SA N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax 4

4

---

---

---

---

---

---

---

---

# PROTOCOLE EXPERIMENTAL

- Recenser l'ensemble des étapes qui vont être conduites pour construire et évaluer la proposition de recherche avec des utilisateurs
- Garantir la traçabilité des différentes étapes de l'étude menée

[http://theдре.imag.fr/wp-content/uploads/2022/02/V5\\_Rediger\\_protocole\\_experimental.docx](http://theдре.imag.fr/wp-content/uploads/2022/02/V5_Rediger_protocole_experimental.docx)

CC BY-NC-SA N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

5

---

---

---

---

---

---

---

---

# ANALYSE DES DONNEES QUALITATIVES

Problématique

Questions de recherche

Objectifs de l'expérimentation

Données qualitatives ou quantitatives

CC BY-NC-SA N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

6

---

---

---

---

---

---

---

---

## DONNÉES ?

- une **donnée Existante** hors du contexte de la recherche que le chercheur peut mobiliser pour répondre à ses questions. (e.g. un support de cours)
- Une **donnée Mesurée** données construites par et pour le chercheur sur la base de travaux théoriques dans une démarche scientifique (e.g. une note de d'utilisabilité).



N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

7

7

---

---

---

---

---

---

---

---

## DONNÉES ?

- **Factuelles** résultent de l'observation de faits ou d'activités. Elles sont mesurées par des traces d'activité sur la plateforme ou des captures vidéo de personnes en situation.
- **Déclaratives** reposent sur l'expression des personnes sur leurs expériences et leurs perceptions. Elles sont recueillies par des questions dans des questionnaires ou lors d'entretiens.



N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

8

8

---

---

---

---

---

---

---

---

## DONNÉES ?

- **Qualitatives** : Peu de connaissances, Comprendre, identifier des phénomènes, Etudier la diversité - Nombre de personnes faibles mais divers profils.
- **Quantitatives** : Connaissances sur le phénomène, Dénombrer, quantifier, tester - Méthode d'échantillonnage : quota, aléatoire, plan d'expérience



N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

9

9

---

---

---

---

---

---

---

---


**QUALITATIVES**

**Objectifs**

- Accéder au sens (Duyck, 2003) et/ou aux représentations.
- Comprendre
- Identifier des phénomènes, des notions inconnues
- Appréhender des notions inconnues
- Connaître en profondeur et avec précision un phénomène
- Observer la diversité

**« Echantillon »**

- Des profils divers pour augmenter la diversité
- Par de représentativité de la diversité

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

10

---

---

---

---

---

---

---

---


**QUALITATIVES**

**Les méthodes de PRODUCTION de données QUALITATIVES**

- Entretien semi- directif
- Focus-group ou groupe de discussion
- Auto-confrontation
- Journal de bord et « Social probes »
- Observation ....

**Les données produites**

- Audio
- Vidéo
- Carnets de notes
- Schéma...

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

11

---

---

---

---

---

---

---

---

**ANALYSE DES DONNEES  
QUANTITATIVES**

**Comment faites vous ?**

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

12

---

---

---

---

---

---


---

---

# SYNTHÈSE A CHAUD

**Une synthèse même rapidement écrite après un entretien ou une séance de focus group est essentielle, elle permet de gagner du temps lors de l'analyse des données.**

**RÉSUMÉ DES ENTRETIENS**  
 ETUDE : NOM DE L'ÉTUDE  
 Équipe/chercheur(s) : Nom, Prénom  
 Champ(s) : Nom, Prénom  
 Date et heure de rdv :  
 Nom Maternelle :  
 N° :  
 Titre de l'entretien :  
 L'entretien servira :  
 Résumé :  
 Reproduire les thèmes de la grille d'entretien  
 Noter ce que le sujet évoque pour chacun des thèmes  
 Éléments en lien avec l'objectif de l'expérimentation et de la thèse  
 Éléments nouveaux ou marquants :  
 Note de 0 à 5 :  
 0 : pas intéressant, 5 très intéressant

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

13

---

---

---

---

---

---

---

---


---

---

# ANALYSE QUALITATIVE

**Deux familles d'analyse des données**

- **Analyse thématique : manuelle ou semi-automatisée**
  - Construire une liste de thème en lien avec l'objet de l'étude
  - Relever les extraits de phrases qui évoquent le thème (manuelle ou semi-automatisée : Taguette, Elan, Atlas TI, ..)
- **Analyse du contenu, du discours, automatisée**
  - Analyses statistiques du corpus sur la base d'étude du champ lexical (Sphinx), sur la base de la syntaxe ou de la grammaire (Alceste, Rtemis, Rainette)

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

14

---

---

---

---

---

---

---


---

---

---

# ANALYSE THÉMATIQUE

P.Paille et A.Muchielli, L'analyse qualitative en sciences humaines et sociales, 2012.

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

15

---

---

---

---

---

---

---

---

---


---

## ANALYSE THÉMATIQUE

Ce n'est

**Ni un codage des données => trop réducteur difficile de faire une synthèse**

**Ni une quantification des thèmes => pas d'échantillonnage statistique, pas produit pour cela**

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

16

---

---

---

---

---

---

---

---

## ANALYSE QUALITATIVE

Retranscrire les entretiens (1h => 6h) , des logiciels de retranscriptions : NOOTA <https://noota.io/>

A partir de la problématique et de la grille d'entretien, créer une grille thématique.

A partir de la **retranscription des entretiens**, identifier les thèmes qui correspondent à la problématique et aussi les thèmes qui émergent

Les thèmes sont larges, il convient de créer des sous-thèmes

Chacun des sous-thèmes est illustré par des verbatims (extraits du discours associé au code du sujet ayant prononcé le discours)

Synthèse pour chaque thème

Synthèse globale

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

17

---

---

---

---

---

---

---

---

## ANALYSE THÉMATIQUE

Comment faire ?

**1- A partir des questions de recherche, des questions du guide d'entretien, élaborer une grille d'analyse**

Pour vous aider utilisez ces trois niveaux d'analyse :

- Descriptif : « *il a pris son manuel scolaire* »
- Interprétatif : « *là je crois qu'il a compris* »
- Explicatif : « *Cet élève a vraiment progresser car il a beaucoup travaillé la géométrie ce trimestre* »

Pour les hypothétiques-déductifs, pour vous aider utilisez des modèles d'analyse existants

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

18

---

---

---

---

---

---

---

---


## ANALYSE THÉMATIQUE

---

**Comment faire ?**

- 2- A la lecture des retranscriptions ou à l'écoute des enregistrements, Identifier des thèmes pour répondre à vos questions, à affirmer ou à rejeter vos hypothèses .
- 3- Identifier aussi des nouveaux thèmes auxquels vous n'auriez pas pensés (raisonnement abductif)
- 4- Garder les *verbatim*s (extraits de phrases) pour « prouver » ce que vous avez identifié et le code de la personne associée
- 5- Elaborer ensuite une synthèse pour chaque thème
- 6- Mettre en lien avec vos questions sou hypothèses

---

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

19

---

---

---

---

---

---

---

---

---

---

## ANALYSE QUALITATIVE

---


Exemple : Une étude sur l'impact des smartphones et des tablettes

Thème « Utilisation des smartphones »  
 Sous -thème : « utilisation générale »  
 « je fais tout avec mon téléphone » (S4)

Sous thème « type d'application »  
 1- utilisation pour télécharger des films  
 2- utilisation pour voir des photos avec des amis »  
 « c'est pour voir les photos que l'on en fait entre potes (S6) »

Thème nouveau de « remplacement des tablettes par les téléphones »  
 « je ne l'utilise plus elle est tombée en panne , j'ai pas changé mon téléphone st suffisant » (S4)

---

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

20

---

---

---

---

---

---

---

---

---

---

## ANALYSE QUALITATIVE

### Points forts et Points faibles

---

Outil : Grille qui liste les éléments relatifs à une proposition et qui à deux colonnes une pour noter les points forts et l'autre les points faibles


Remarque : Outil de production pour une situation expérimentale où le nombre de personnes est réduit.

Table Number : \_\_\_\_\_ First and last name : \_\_\_\_\_

Exemple :

	Strengths	Weaknesses
List the strengths and weaknesses on the following		
the work performed with the method	+	-
the collaborative work enabled with the method	+	-
The creativity enabled with the method	+	-
the serious gaming of the method	+	-
The software compared to card game	+	-

---

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

21

---

---

---

---

---

---

---

---

---

---

# ANALYSE QUALITATIVE

## Points forts et Points faibles

Méthode :  
Saisir l'ensemble des réponses  
: Ligne=participants. Colonnes  
= réponse aux questions, 2  
colonnes par éléments

Experimentation Francke en 2011 les de l'écrit d'être sur les OBL. LIENET-DAN Perren at all			
		The collaborative work enabled with the method	The method ability to give new ideas
video	not	What you like +	What you don't like +
DM	1	1	1
DM	2	1	1
DM	3	1	1
DM	4	1	1
DM	5	1	1
DM	6	1	1
DM	7	1	1
DM	8	1	1
DM	9	1	1
DM	10	1	1
DM	11	1	1
DM	12	1	1
DM	13	1	1
DM	14	1	1
DM	15	1	1
DM	16	1	1
DM	17	1	1
DM	18	1	1
DM	19	1	1
DM	20	1	1

CC BY-NC-ND N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

22

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

# ANALYSE QUALITATIVE

## Points forts et Points faibles

Méthode :  
Faire une « analyse thématique  
simplifiée » et une synthèse

The collaborative work enabled with the method		
Les points positifs		
3 cases videos (dont 2 vidéos aussi côté points négatifs)		
1. Concernant la méthode		
La méthode facilite et stimule la collaboration des formateurs	<p>great for triggering discussions</p> <p>The method promotes collaboration.</p> <p>Everyone participating.</p> <p>Good to develop team spirit</p> <p>Easy, nice, good way</p>	14
Elle permet de mieux participer, de partager des idées différentes et d'être créatifs collectivement	<p>Very good "helped to see each participant's point of view and thus come up with a collective work"</p> <p>Effective creating, compromising, selection and implementation.</p> <p>We all participate in the creation process by bringing ideas and say what we think is not good or other ideas</p> <p>emergence of the solution by speaking and interacting</p> <p>confident, exchange, share ideas,</p> <p>creativity,</p> <p>direct, open, brainstorm,</p> <p>open discussion</p> <p>friendly.</p>	17

CC BY-NC-ND N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

23

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

# ANALYSE QUALITATIVE

## Schéma

Créer une grille d'analyse pour savoir ce qui doit être repéré par rapport à la problématique

Photographier chacune des productions

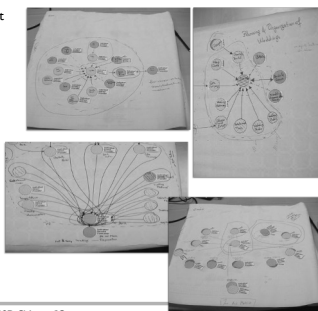
Repérer les ressemblances/les différences entre les représentations

Repérer ce qui confirme les objectifs de l'étude

Repérer les apports de la représentation des sujets par rapport à la proposition

Visionner et écouter les films

Conduire une « analyse thématique simplifiée »



CC BY-NC-ND N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

24

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---



# ANALYSE QUALITATIVE

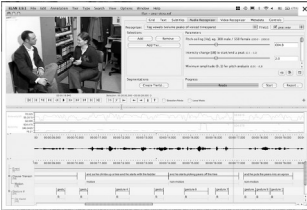
Video


Créer une grille d'analyse pour savoir ce qui doit être repérer par rapport à la problématique.

Utiliser des logiciels d'annotations de video

Analyser les annotations  
Statistiques descriptives  
Extraction de verbatims

Faire une synthèse



 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

25

---

---

---

---

---

---

---

---


# QUANTITATIVES

**Objectif**

- Quantifier
- Evaluer, valider
- Modéliser, prévoir
- Diagnostiquer

**Echantillonnage**

- Méthode d'échantillonnage aléatoire
- Méthode échantillonnage par quotas
- Plans d'expériences : groupe contrôle vs groupe expérimental (Post-Positivisme)
- Recensement

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

26

---

---

---

---

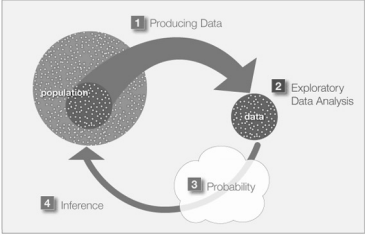
---

---


---

---

# QUANTITATIVES



Source : <https://oli.web.cmu.edu/openlearning/forstudents/freecourses/statistics>

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

27

---

---

---

---

---

---

---

---

## ECHANTILLONNAGE, INDIVIDUS

**Sans description de la méthode d'échantillonnage, les résultats ne sont valides que pour les données recueillies mais la généralisation n'est pas possible.**

- **Les quotas**
  - Caler les résultats sur les données du recensement de la population ou sur des informations globales sur la population étudiée.
  - Le plus pratiqué car moins couteux
- **La méthode aléatoire**
  - Tirer au hasard dans une base de sondage.
  - La probabilité d'inclusion d'un individu est connu sans biais.
  - Couteux car l'individu tiré au hasard ne doit pas être abandonné
- **L'exhaustivité (le recensement)**
  - Avoir la totalité des individus ou des actions faites par les individus
- **Le plan expérimental**
  - Le nombre de sujets est fixé a priori en fonction des facteurs expérimentaux
  - Définir la structure de l'expérience

CC BY-NC-SA N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex 28

28

---

---

---

---

---

---

---

---

---

---

## QUANTITATIVES

### Les méthodes de PRODUCTION des données quantitatives

- **Questionnaire**  
Auprès d'un échantillon correctement échantillonné (en sociologie minimum 500 personnes)  
Peut-être administré en face-à-face ou en ligne
- **Capture de traces d'activités**  
Avoir une application qui permet de tracer les interactions des utilisateurs
- **Tests utilisateurs**  
Avoir une application sur laquelle on va faire des tests précis  
Souvent utilisés en IHM, mesure de performance, de précision, erreurs, ...

**Les données quantitatives**

- **Fichier : tableau ligne\*colonne, type CSV , excel, data set de R, etc..**

CC BY-NC-SA N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

29

---

---

---

---

---

---

---

---

---

---

## QUANTITATIVES

- VOCABULAIRE DES DONNEES => CHOIX DES METHDOES D'ANALYSE DES DONNEES
  - Ligne : **Individus statistiques**
  - Colonne : **Variables**
  - A découvrir par vous mêmes ...

CC BY-NC-SA N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

30

---

---

---

---

---

---

---


---

---

---

## QUANTITATIVES

- Fichier de données
  - Utilisation d'une plateforme pour former à la construction de protocoles expérimentaux en chimie
  - Impact des niveaux de rétroaction de la plateforme sur l'apprentissage

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

31

---

---

---

---

---

---

---

---

## QUANTITATIVES

- . Quel est le thème, la sémantique des variables ?
- . Quelles caractéristiques/quels types ont les valeurs prises par ces variables ?
- . Quand sont collectées les données ?
- . Qui les a produites ?
- . Quel est leur rôle dans l'expérimentation et l'analyse des données ?
- . Quels sont les objectifs de recherche qui pourraient être à l'origine de la création de ces données ?
- . Quelle est la méthode d'échantillonnage ?

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

32

---

---

---

---

---


---

---

---

## QUANTITATIVES

- . Corrections

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

33

---

---

---

---

---

---

---

---

# QUANTITATIVES

Source : <https://oli.web.cmu.edu/openlearning/forstudents/freecourses/statistics>

CC BY-NC-SA N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

34

---

---

---

---

---

---

---

---

# ECHANTILLONNAGE, INDIVIDUS

**Sans description de la méthode d'échantillonnage, les résultats ne sont valides que pour les données recueillies mais la généralisation n'est pas possible.**

- Les quotas**
  - Caler les résultats sur les données du recensement de la population ou sur des informations globales sur la population étudiée.
  - Le plus pratiqué car moins coûteux
- La méthode aléatoire**
  - Tirer au hasard dans une base de sondage.
  - La probabilité d'inclusion d'un individu est connue sans biais.
  - Coûteux car l'individu tiré au hasard ne doit pas être abandonné
- L'exhaustivité (le recensement)**
  - Avoir la totalité des individus ou des actions faites par les individus
- Le plan expérimental**
  - Le nombre de sujets est fixé a priori en fonction des facteurs expérimentaux
  - Définir la structure de l'expérience

CC BY-NC-SA N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex 35

35

---

---

---

---

---

---

---

---

## 3 Moments dans la production des données

AVANT	PENDANT	APRES
-------	---------	-------

**l'expérience avec une plateforme**

Socio-démographiques Pré-test Facteurs expérimentaux ...	La date et heure de l'action Les actions avec la plateforme ....	Le post-test La note d'utilisabilité La durée de connexion Le nombre d'actions Le score (post-test - pre test) ...
---	--	---

CC BY-NC-SA N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

36

---

---

---

---

---

---

---

---



# ANALYSE QUANTITATIVE

---

- Que pensez vous de cette formule ?
- Données + Algorithme de traitement = Résultats

N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

40

---

---

---

---

---

---

---

---

# ANALYSE QUANTITATIVE

Données  
quantitatives

Valider les données

Enrichir les données

Traiter les données

Interpréter les données

Pré-traitement = 80%

- **Valider** identifier les outliers, les manquants, ...
- **Enrichir** les données créer de nouvelles variables
- **Traiter** : appliquer des méthodes statistiques pour donner du sens aux données recueillies
- **Interpréter** les résultats par rapport aux questions de recherche

N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

41

---

---

---

---

---

---

---

---

# ANALYSE QUANTITATIVE

## BONNES PRATIQUES DE VALIDATION

- Mettre en relation les données mesurées avec les valeurs attendues, *Exemple : note des élèves comprises entre 0 et 20 ou entre 0 et 10.*
- Si les valeurs mesurées sont quantitatives, rechercher les valeurs minimales et maximales
- Identifier les individus pour lesquelles les valeurs sont manquantes
- Croiser des variables entre elles, pour voir la cohérence, *Exemple : âge et date de naissance ; âge de l'élève et niveau scolaire*
- **ne pas supprimer les données : garder le fichier de départ et créer un nouveau fichier de travail**
- Sur certains logiciels, les déclarer en données manquantes ,
- Les corriger, en gardant une trace des corrections ( pgm informatique, note sur word, .... ) ,
- Faire des traitements stats en filtrant les données

N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

42

---

---

---

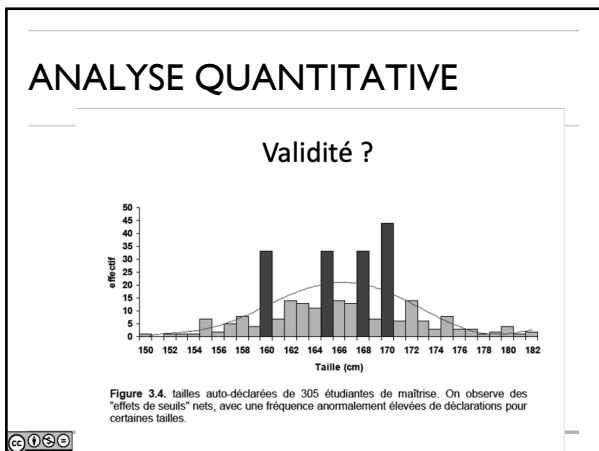
---

---

---

---

---



43

---

---

---

---

---

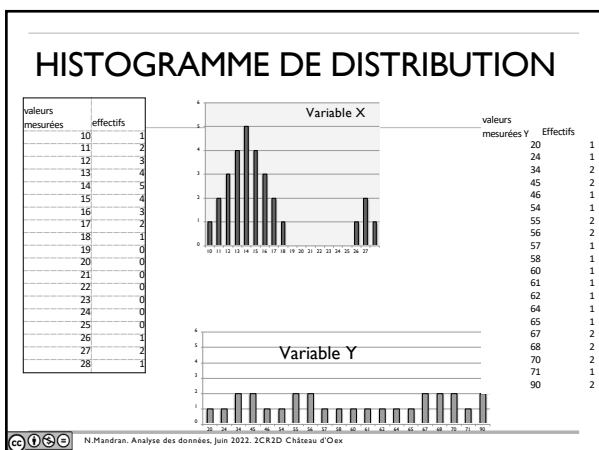
---

---

---

---

---



44

---

---

---

---

---

---

---

---

---

---

## ANALYSE QUANTITATIVE

- INDICATEURS STATISTIQUES
- Lesquels connaissez vous ?

N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

45

---

---

---

---

---

---

---

---

---


---

**STATISTIQUES DESCRIPTIVES**

---

Moyenne :  
Classe 1 : 10,4  
Classe 2 : 10,4

---

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

46

---

---

---

---

---

---

---


---

**STATISTIQUES DESCRIPTIVES**

---

Moyenne :	Médiane
Classe 1 : 10,4	Classe 1 : 11
Classe 2 : 10,4	Classe 2 : 10

---

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

47

---

---

---

---

---

---

---


---

**STATISTIQUES DESCRIPTIVES**

---

Moyenne :	Médiane	Etendue
Classe 1 : 10,4	Classe 1 : 11	Classe 1 : 11
Classe 2 : 10,4	Classe 2 : 10	Classe 2 : 1

---

 N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

48

---

---

---

---

---

---

---

---



## STATISTIQUES DESCRIPTIVES

Moyenne :	Médiane	Etendue	Ecart type
Classe 1 : 10,4	Classe 1 : 11	Classe 1 : 11	Classe 1 : 3,61
Classe 2 : 10,4	Classe 2 : 10	Classe 2 : 1	Classe 2 : 0,53

N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

49

---

---

---

---

---

---

---

---

---

---

## STATISTIQUES DESCRIPTIVES

Moyenne :	Médiane	Etendue	Ecart type
Classe 1 : 10,4	Classe 1 : 11	Classe 1 : 11	Classe 1 : 3,61
Classe 2 : 10,4	Classe 2 : 10	Classe 2 : 1	Classe 2 : 0,53

**Classe 1**

Note	5	6	7	8	9	10	11	12	13	14	15	16
Bar	1	1	1	1	1	1	1	1	1	1	1	1

Classe 1	Classe 2
8	10
12	11
16	10
9	10
6	11
12	11
14	11
11	10
10	10

**Classe 2**

Note	5	6	7	8	9	10	11	12	13	14	15	16
Bar	1	1	1	1	1	1	1	1	1	1	1	1

N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

50

---

---

---

---

---

---

---

---

---

---

## STATISTIQUES DESCRIPTIVES

**Classe 1**

Note	5	6	7	8	9	10	11	12	13	14	15	16
Bar	1	1	1	1	1	1	1	1	1	1	1	1

**Positionnement**

Moyenne :	Classe 1 : 10,4
	Classe 2 : 10,4
Médiane	Classe 1 : 11
	Classe 2 : 10

**Classe 2**

Note	5	6	7	8	9	10	11	12	13	14	15	16
Bar	1	1	1	1	1	1	1	1	1	1	1	1

**Dispersion**

Etendue	Classe 1 : 11
	Classe 2 : 1
Ecart type	Classe 1 : 3,61
	Classe 2 : 0,53

N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oex

51

---

---

---

---

---

---

---

---

---

---

## FREQUENCES, POURCENTAGES ET QUARTILES

- Fréquences : nombre d'occurrences d'une valeur
- Pourcentages : nombres d'occurrences d'une valeur/effectif total \*100
  - Valide si effectif total (n) >=100
- Quartiles : découpage de la distribution en 4 parties égales
  - 1er quartile est la donnée de la série qui sépare les 25 % inférieurs des données (Q1) ;
  - 2e quartile est la donnée de la série qui sépare les 50 % inférieurs des données (Q2 ⇔ Médiane) ;
  - 3e quartile est la donnée de la série qui sépare les 75 % inférieurs des données (Q3) ;
  - le 4e quartile est le maximum (Q4)
- La distance interquartile = Q3-Q1
- Déciles : découpage de la distribution en 10 parties égales
- Percentiles : découpage de la distribution en 100 parties égales

N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

52

---

---

---

---

---

---

---

---

---

---

## BOX PLOT

- Barre du bas : Q1-1,5 IQ
- Deuxième barre : Q1
- Troisième barre : médiane
- Quatrième barre : Q3
- O : à plus de 1,5 distance inter quartile (IQ)
- \* : à plus de 3 distances inter quartile

N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

53

---

---

---

---

---

---

---

---

---

---

## BOX PLOT

délai par type de traitement

Délai de guérison et type de traitements

- Délai de guérison en jours : Quantitative
- Traitement : Qualitative à 5 modalités (placebo, T2, T3, T4, T5)

	Min	Q1	Median	Mean	Q3	max	Etendue	Ecart type
Placebo	5	7,0	7,5	7,5	8,0	10,0	5,0	2,0
T2	3	4,2	5,5	5,0	6,0	6,0	3,0	2,5
T3	3	4,0	4,0	4,3	4,8	6,0	3,0	2,6
T4	3	4,3	5,5	5,2	6,0	7,0	4,0	3,0
T5	3	5,3	6,5	6,2	7,0	9,0	6,0	3,5

N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

54

---

---

---

---

---

---

---

---

---

---

## DÉCILES ET PERCENTILES

Nombre de ressources ouvertes lors de l'exercice de math

	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Groupe 1	2	3	4	15	10	11	12	15	20	35
Groupe 2	2	7	10	15	20	21	22	24	28	29
Groupe 3	0	0	0	0	0	0	0	0	4	20

Qu'observez vous ?  
 Quelles différences entre les groupes ?  
 Que faire du groupe 3 ?  
 Comment recoder cette variable ?

N'Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax 55

55

---

---

---

---

---

---

---

---

---

---

---

---

## TABLEAUX CROISÉS

■ Mettre en relation deux variables qualitatives

Fréquences	Men	Women	Total
Humanities	40	100	<b>140</b>
Natural Science	110	100	<b>210</b>
Social sciences	80	140	<b>220</b>
<b>Total</b>	<b>230</b>	<b>340</b>	<b>570</b>

Pourcentages	Men	Women	Total
Humanities	7,0	17,5	<b>24,6</b>
Natural Science	19,3	17,5	<b>36,8</b>
Social sciences	14,0	24,6	<b>38,6</b>
<b>Total</b>	<b>40,4</b>	<b>59,6</b>	<b>100,0</b>

N'Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax 56

56

---

---

---

---

---

---

---

---

---

---

---

---

## TABLEAUX CROISÉS

■ Mettre en relation deux variables qualitatives

% lignes	Men	Women	Total
Humanities	28,6	71,4	<b>100</b>
Natural Sciences	52,4	47,6	<b>100</b>
Social sciences	36,4	63,6	<b>100</b>
<b>Total</b>	<b>40,4</b>	<b>59,6</b>	<b>100</b>

% colonnes	Men	Women	Total
Humanities	17,4	29,4	<b>24,6</b>
Natural Science	47,8	29,4	<b>36,8</b>
Social Sciences	34,8	41,2	<b>38,6</b>
<b>Total</b>	<b>100,0</b>	<b>100,0</b>	<b>100</b>

N'Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax 57

57

---

---

---

---

---

---

---

---

---

---

---

---

## TESTS STATISTIQUES

- **Tester quoi ?**
  - Comparer des indicateurs statistiques
    - Moyennes, Proportions, Variance, ...
  - Tester des liaisons entre variables
    - Corrélation, indépendance, ...
  - Tester des modèles
    - Régression (le poids est une fonction du poids)

N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

58

---

---

---

---

---

---

---

---

## CHOIX DES TESTS

- Méthodes de tests sont liées
  - À la combinaison rôles et types des variables
  - Aux distributions des variables (normales et non normales)

N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

59

---

---

---

---

---

---

---

---

## QUELQUES TESTS

		Condition d'application	Variable(s) explicative(s)
<i>Type de la variable à expliquer</i>	Qualitative		Qualitative Test d'indépendance du Chi2 (couleur des yeux et couleur des cheveux)
	Quantitative	La variable à expliquer SUIT une loi normale (test de shapiro)	Test de Student (2 groupes) Anova (plusieurs groupes)
	Quantitative	La variable à expliquer NE SUIT PAS une loi normale (test de shapiro)	Test de Wilcoxon ou test de Kruskal et Wallis

N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

60

---

---

---

---


---

---

---

---

## LES TESTS STATISTIQUES



**HYPOTHESES STATISTIQUES**

H0 (hypothèse nulle) : moyenne des notes des garçons = moyenne des notes des filles.  
 H1 (hypothèse alternative) : moyenne des notes des garçons  $\neq$  **moyenne** des notes des filles.

=> Posent deux questions ...  
 Est ce que la moyenne des notes que j'ai observée dans la classe est différente ou non entre les garçons et les filles ?  
 Est ce que résultat est généralisable d'autres classes ? Voir à l'ensemble des individus ?

**STATISTIQUE**: l'étude de l'égalité de deux moyennes suit une distribution d'un t de Student.

- À partir des données observées => calcul le t de student observé (To)
- Compare ce To à un t théorique qui serait obtenu dans le cas de l'hypothèse H0.

Permet d'obtenir la probabilité d'obtenir un t de student supérieur à celui que j'ai observé dans ma classe, si on refaisait des relevés de notes dans une autre classe.

**CONCLUSIONS**

- Si le t observé est supérieur au t théorique au seuil de 1%: le test est significatif au seuil de 1% Autrement dit, c'est la probabilité de trouver un t Observé supérieur au t théorique dans le cas d'une autre classe
- Le langage habituel: rejeter H0: autrement dit les moyennes ne sont pas égales
- Le plus simple à comprendre: on accepte H1 avec un risque de se tromper égal à 1%.

CC BY-NC-SA N.Mandran, Analyse des données, Juin 2022, 2CR2D Château d'Oax

61

---

---

---

---

---

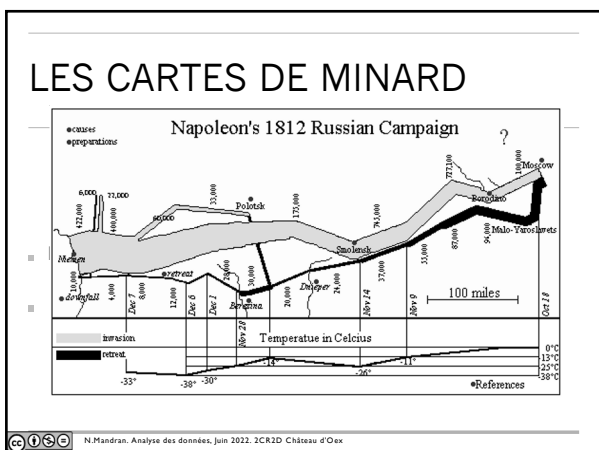
---

---

---

---

---



62

---

---

---

---

---

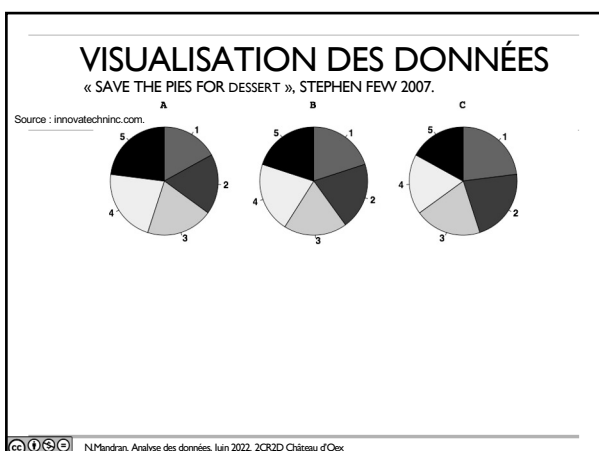
---

---

---

---

---



63

---

---

---

---

---

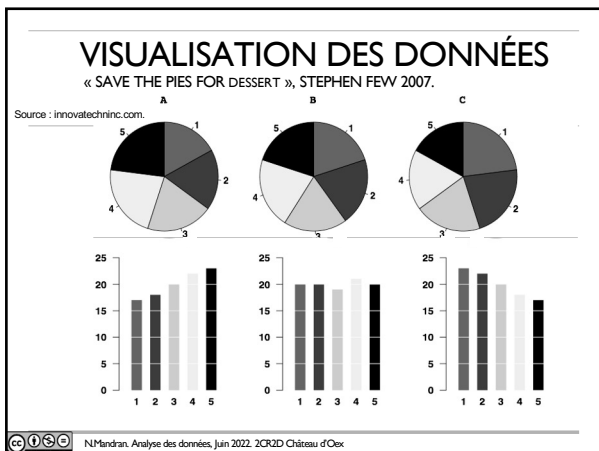
---

---

---

---

---



64

---

---

---

---

---

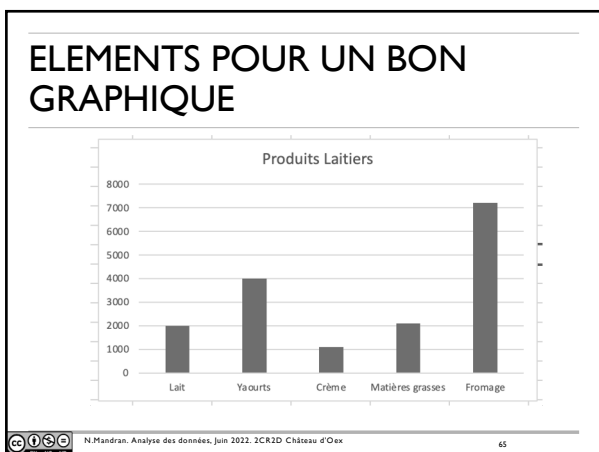
---

---

---

---

---



65

---

---

---

---

---

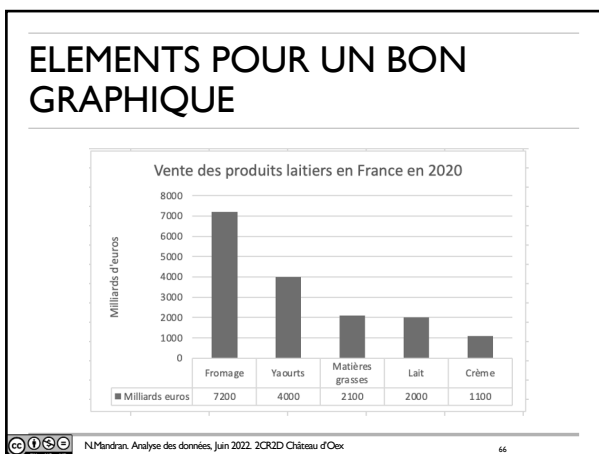
---

---

---

---

---



66

---

---

---

---

---

---

---

---

---

---