



**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE**



**Ministère de l'Enseignement Supérieur et de la Recherche Scientifique**

**Université Abou bekr Belkaïd – Tlemcen –**

**Faculté des Sciences Economiques, Commerciales et Sciences de Gestion**

**Département des Sciences Commerciales**

**Techniques d'Enquête  
Et  
Traitement des Données**

**Master 2 - Finance et Commerce International**

**Dr. Ouieme SOUR**

**2022 - 2023**

***TECHNIQUES D'ENQUETE ET  
TRAITEMENT DES DONNEES***

***Master 2 - Finance et Commerce International***

***Dr. Ouieme SOUR***

## **TABLE DES MATIERES**

### **INTRODUCTION**

#### **CHAPITRE 1 : Préparation du Cadre de l'Enquête**

##### **Cours 1 : Préparer les questions pertinentes**

1. Les principales méthodes de recueil d'information
2. Concepts statistiques pour la réalisation d'une enquête
3. Qu'est ce qu'une enquête ? Qu'est ce qu'une technique d'enquête ?

##### **Série « 1 » d'exercices**

##### **Cours 2 : Choix d'échantillonnage**

1. L'échantillonnage
2. Les types d'échantillonnage
  - 2.1 Méthodes d'échantillonnage aléatoires
    - 2.1.1. Echantillonnage aléatoire simple
    - 2.1.2. Echantillonnage aléatoire systématique
    - 2.1.3. Echantillonnage aléatoire par grappe
    - 2.1.4. Echantillonnage aléatoire stratifié
  - 2.2 Méthodes d'échantillonnage Non-aléatoire
    - 2.2.1 Echantillonnage non-aléatoire volontaires
    - 2.2.2 Echantillonnage non-aléatoire par quota
    - 2.2.3 Echantillonnage non-aléatoire boule de neige

##### **Série « 2 » d'exercices**

##### **Cours 3 : Types d'enquête**

##### ***Partie « 1 » : Etude qualitative – Entretien***

1. Les types d'entretien
  - 1.1 Entretien non-structuré
  - 1.2 Entretien semi-structuré
  - 1.3 Entretien structure

2. Comment préparer un entretien de recherche ?

### ***Partie « 2 » : Etude quantitative – Questionnaire***

1. Les types de questionnaire
2. Comment rédiger un questionnaire d'une enquête quantitative ?
3. Comment formuler les questions ?
  - 3.1 Questions fermées
  - 3.2 Questions ouvertes
4. Les erreurs à éviter dans un questionnaire
5. Comment diffuser le questionnaire ?

### **Série « 3 » d'exercices**

#### **Cours 4 : Le recueil des données**

1. Des données : Quantitative & Qualitative
2. Echelle de mesure

### **Série « 4 » d'exercices**

## **CHAPITRE 2 : Traitement des Données**

### **Cours 1 : Types de tableaux analysables**

1. Tableaux des données
2. Tableaux de distribution de fréquences
3. Tableaux de contingence

### **Cours 2 : l'analyse en composantes principales**

1. Qu'est ce qu'un analyse en composantes principales
2. Explication étape par étape de l'ACP
3. Résultats de l'ACP

### **Cours 3 : L'analyse factorielle des correspondantes**

1. Qu'est ce que l'analyse factorielle des correspondantes ?
2. Résultats de l'AFC

### **Cours 4 : Traitement des données**

## *Partie « 1 » : Traitement des données avec le SPSS*

1. Les logiciels de traitement des données
2. Qu'est ce que SPSS ?
3. Interface du logiciel SPSS
  - 3.1 La fenêtre de démarrage
  - 3.2 Types de fenêtres dans SPSS
  - 3.3 Comment organiser les données dans un SPSS ?
  - 3.4 Manipulation des données d'un fichier
    - 3.4.1 Insertion des variables
    - 3.4.2 Insertion des données
    - 3.4.3 Sauvegarder la base de données
4. L'analyse des données

## *Partie « 2 » : Exemple Pratique*

### **CONCLUSION**

### **REFERENCES**

## **LISTE DES FIGURES**

| <b>N°</b> | <b>Intitulé</b>                                 |
|-----------|---|
| <b>01</b> | Les Concepts Statistiques d'une Enquête         |
| <b>02</b> | Les Différentes Méthodes d'Echantillonnage      |
| <b>03</b> | Echantillonnage Aléatoire Simple                |
| <b>04</b> | Echantillonnage Aléatoire En Grappe             |
| <b>05</b> | Echantillonnage Non-Aléatoire Par Quotas        |
| <b>06</b> | Echantillonnage Non-Aléatoire en Boule de Neige |
| <b>07</b> | Les Quatre Types d'Echelles de Mesure           |
| <b>08</b> | La Fenêtre de Démarrage                         |
| <b>09</b> | Les 3 Fenêtres SPSS                             |

# ***INTRODUCTION***

Ce polycopié est un support pédagogique du module Technique d'Enquête et Traitement des Données préparées selon le Canva, et qui s'adresse principalement aux étudiants universitaires de deuxième année Master spécialité Finance et Commerce International, dispensé au Département de Sciences Commerciales, Faculté des Science Economiques, Commerciales et Sciences de Gestion, Université de Tlemcen.

Le module de techniques d'enquête couvre tous les aspects de l'enquête, en termes des cours théoriques à connaître et des cours empiriques pour savoir les appliquer sur le terrain. Ce cours, structuré et pédagogique, vise à fournir une description complète de toutes les spécificités des enquêtes en termes d'études qualitatives et d'études quantitatives sur lesquelles on s'appuie.

Ce module de Technique d'Enquête et Traitement des Données comprends deux Chapitres (Chapitre 1 : *Préparation du cadre de l'enquête* – Chapitre 2 : *Traitement des données*) et chaque chapitre comprend un ensemble de cours visant à atteindre les Objectifs de modules. OÙ à la fin de semestre, les étudiants devront être capable de :

- Collecter, traiter, diffuser et utiliser les informations statistiques.
- Concevoir et mettre en exécution des programmes d'enquête.
- Appréhender les étapes de réalisation d'une enquête.
- Connaître les méthodes d'exploitation des données recueillies à l'aide des outils statistiques adéquats.
- Renforcer les capacités des étudiants en matière d'outils et techniques d'enquêtes, depuis la conception jusqu'à l'exploitation des résultats.
- Dépouiller et analyser les résultats des dites enquêtes.

# **CHAPITRE 1 :**

## **Préparation du Cadre de l'Enquête**

*Cours « 1 »*

*Préparer les  
Questions Pertinentes*

## **1. Les principales méthodes de recueil d'information**

**Etude Documentaire** : consultation de monographies, revues, livres, banques des données...

Exemples :

- Etudier l'évolution du PIB par région à travers les statistiques annuelles.
- Etudier l'évolution des exportations algériennes sur les 20 dernières années à travers les bilans du Ministère du commerce.

**Expérimentation** : réalisation d'expérience.

Exemples :

- Etudier l'effet d'un médicament sur des personnes diabétiques
- Etudier l'effet d'un type d'entraînement sportif sur la performance de l'athlète.

**Enquête** : consultation de personnes

- **Enquête Qualitative** :

- Cerner les perceptions de la population à l'égard des inégalités sociales.
- Comprendre les représentations des jeunes diplômés relatives à la migration des cerveaux.

- **Enquête Quantitative** :

- Mesurer le niveau de vie des ménages
- Mesurer les niveaux de connaissances, les attitudes et les pratiques des jeunes..

## **2. Concepts statistiques pour la réalisation d'une enquête**

Avant de commencer la définition de l'enquête et comment la mener, il est nécessaire de connaître certains termes importants et de base dans le processus d'enquête, notamment les suivants :

➤ **Population** :

Dans les statistiques, la population est l'ensemble des éléments à partir desquels vous tirez des données pour une étude statistique. Il peut s'agir d'un groupe d'individus, d'un ensemble

d'éléments, etc. Il constitue le pool de données d'une étude. Généralement, la population fait référence aux personnes qui vivent dans une zone particulière à un moment précis. Mais dans les statistiques, la population fait référence aux données sur votre étude d'intérêt (Ardilly P., 2006). Il peut s'agir d'un groupe d'individus, d'objets, d'événements, d'organisations, etc. Vous utilisez des populations pour tirer des conclusions.

➤ **Unité statistique :**

La population objet de l'étude est composée d'individus (unité de base à laquelle on s'intéresse) appelés aussi, entités, observations ou unités statistiques. Chaque individu est supposé « se démarquer » sans ambiguïté des autres par une information claire et précise (identifiant i). L'unité statistique peut être simple (personne) ou complexe (entreprise).

➤ **Echantillon :**

Un échantillon est défini comme une représentation plus petite et plus gérable d'un groupe plus grand. Un sous-ensemble d'une population plus large qui contient les caractéristiques de cette population. Un échantillon est utilisé dans les tests statistiques lorsque la taille de la population est trop grande pour que tous les membres ou observations soient inclus dans le test. L'échantillon est un sous-ensemble non biaisé de la population qui représente le mieux l'ensemble des données.

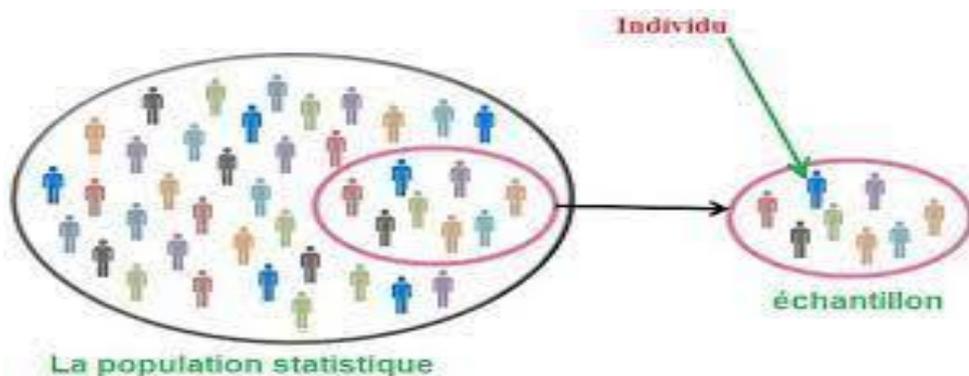
➤ **Caractéristiques d'intérêt :**

Les individus sont soumis à l'observation du chercheur à propos d'un certain nombre de caractéristiques d'intérêt. ———→ Des variables « Y » ———→ opération exhaustive / opération par sondage.

Typologies des caractéristiques d'intérêt peuvent être :

- Quantitatives : note, nombre d'années d'études, chiffre d'affaire, revenu,...
- Ou Qualitatives : opinions, comportements, filières, ...

*Figure 01 : Les Concepts Statistiques d'une Enquête*



### **3. Qu'est ce qu'une Enquête? Qu'est ce qu'une Technique d'Enquête?**

Une **Enquête** est plus qu'une méthode, elle est une stratégie de recueil de données à partir d'un outil ou support tel que le questionnaire, le guide d'entretien, etc.

Une **Technique d'Enquête** est donc l'ensemble des stratégies et habiletés qui permettent de recueillir des informations. Elle exige donc inévitablement une posture physique et mental attentionnée, intéressée et rigoureuse qui permet de respecter les suivantes :

#### **Première Etape :**

1. Choix d'un thème d'étude, d'un sujet (critères : intérêt ; caractère social, objectifs professionnels et personnels, contribution socio-économique, porteur, durable, préparation relève), souligner les concepts clefs du sujets.
2. Mener des recherches d'informations (digitales et documents physiques) les plus récentes sur le sujet tout en clarifiant les concepts pour pouvoir chercher les dimensions et indicateurs dans le but de le clarifier.
3. Repréciser votre sujet (en tentant en compte les concepts, dimensions, indicateurs et en intégrant les critères précédents) suite à la recherche digitale et documentaire.
4. Poser votre question de départ à partir du sujet reprécisé.
5. Donner une réponse provisoire à votre question (hypothèse).
6. Préciser votre objectif.
7. Préparer vous à aller vérifier si votre réponse est confirmée ou informée.

### **Deuxième Etape :**

En fonction de votre étude, vous devez savoir quel type de méthode, voire de technique d'enquête vous devez utiliser :

1. La technique d'enquête Quantitative est un instrument de mesure qui vous permet d'apprécier le niveau de satisfaction, de motivation, les comportements d'achats d'un produit, etc., elle s'adapte aux études marketing.
2. La technique d'enquête Qualitative quant à elle cherche à recueillir le maximum d'informations sur le sujet dans le but de présenter une analyse approfondie.

### **Troisième Etape :**

Le traitement et l'analyse de données recueillies se font grâce à un système de codage avec l'aide de logiciel.

## Série « 1 » d'exercices

### **Exercice : Identifier la population et l'échantillon**

1/ Le propriétaire des stands de boisson d'un stade de football voudrait vendre une nouvelle boisson. Pour savoir quelle future boisson aura le plus de succès, il choisit au hasard 80 numéros de siège et interroge leurs occupants sur leur boisson préférée.

**Identifier la population et l'échantillon de cette étude.** Choisir une réponse :

- La population est l'ensemble des visiteurs du stade qui achètent des boissons ; l'échantillon est constitué des 80 sièges sélectionnés.
- La population est l'ensemble des occupants des 80 sièges sélectionnés ; l'échantillon est constitué des différentes futures boissons.
- La population est l'ensemble des occupants des sièges du stade ; l'échantillon est constitué des occupants des 80 sièges sélectionnés.

2/ Un contrôleur de qualité dans une usine automobile veut vérifier l'épaisseur de la peinture sur une voiture. Il choisit au hasard 30 zones sur la voiture et mesure l'épaisseur de la peinture dans chacune de ces zones.

**Identifier la population et l'échantillon de cette étude.** Choisir une réponse :

- La population est l'ensemble des zones de la voiture ; l'échantillon est constitué des 30 zones sélectionnées.
- La population est l'ensemble des voitures produites ; l'échantillon est l'une de ces voitures.
- La population est l'ensemble des voitures produites ; l'échantillon est constitué des 30 zones sélectionnées.

3/ Une municipalité veut connaître l'opinion de ses administrés sur le programme de réfection de la voirie et de l'éclairage public de la ville. La responsable de la communication a sélectionné au hasard 75 noms dans l'annuaire téléphonique de la ville pour sonder ces personnes par téléphone.

**Identifier la population et l'échantillon de cette étude.** Choisir une réponse :

- La population est l'ensemble des administrés figurant dans l'annuaire ; l'échantillon est constitué des 75 personnes sélectionnées dans cet annuaire.
- La population est l'ensemble des administrés de la ville ; l'échantillon est constitué des administrés inscrits sur la liste électorale.
- La population est l'ensemble des administrés inscrits sur la liste électorale ; l'échantillon est constitué des personnes inscrites sur l'annuaire téléphonique de la ville.

4/ Lucie veut savoir si les plats proposés dans son restaurant sont servis à une température réglementaire. Elle choisit aléatoirement 70 plats dont elle mesure la température avant le service.

**Identifier la population et l'échantillon de cette étude.** Choisir une réponse :

- La population est l'ensemble des plats chauds servis dans le restaurant ; l'échantillon est composé des plats servis à une température réglementaire.
- La population est l'ensemble des 70 plats choisis au hasard ; l'échantillon est constitué des plats servis à une température réglementaire.
- La population est l'ensemble des plats servis dans le restaurant ; l'échantillon est composé des 70 plats choisis aléatoirement.

5/ Pour en savoir plus sur le nombre de livres et de magazines que ses clients adultes achètent chaque mois, une chaîne de librairie interroge 150 clients adultes choisis aléatoirement.

**Identifier la population et l'échantillon de cette étude.** Choisir une réponse :

- La population est l'ensemble des clients adultes de la chaîne de librairie ; l'échantillon est constitué de 150 clients adultes de cette chaîne sélectionnés aléatoirement.
- La population est l'ensemble des clients de la chaîne de librairie ; l'échantillon est constitué des clients adultes de cette chaîne.
- La population est l'ensemble des clients de la chaîne de librairie achetant au moins 1 livre par mois ; l'échantillon est constitué de 150 clients adultes de cette chaîne sélectionnés aléatoirement.

6/ Dans un atelier de mécanique, le contrôleur qualité sélectionne au hasard 40 tiges filetées dans la production hebdomadaire afin de vérifier leur conformité.

**Identifier la population et l'échantillon de cette étude.** Choisir une réponse :

- La population est l'ensemble des tiges filetées produites dans cet atelier ; l'échantillon est constitué des tiges filetées produites cette semaine.
- La population est l'ensemble des tiges filetées produites cette semaine ; l'échantillon est constitué de 40 de ces tiges choisies au hasard.
- La population est l'ensemble des tiges filetées produites dans le monde ; l'échantillon est constitué des tiges filetées produites dans cet atelier.

7/ Un pédiatre a choisi au hasard 10 parents parmi l'ensemble des parents de ses patients. Il leur a ensuite demandé de donner leurs opinions sur différents types de lingettes pour bébé.

**Identifier la population et l'échantillon de cette étude.** Choisir une réponse :

- La population est l'ensemble des parents dont les enfants portent des couches ; l'échantillon est l'ensemble des parents des patients du pédiatre.
- La population est l'ensemble des patients du pédiatre ; l'échantillon est constitué des 10 patients sélectionnés.
- La population est l'ensemble des parents des patients du pédiatre ; l'échantillon est constitué des 10 parents sélectionnés.

8/ Un inspecteur de la sécurité effectue des tests de qualité de l'air dans 7 salles de classe choisies au hasard dans une école primaire.

**Identifier la population et l'échantillon de cette étude.** Choisir une réponse :

- La population est l'ensemble des salles de classe de la région ; l'échantillon est constitué des 7 salles de classe sélectionnées.
- La population est l'ensemble des salles de classe de l'école primaire ; l'échantillon est constitué des 7 salles de classe sélectionnées de cette école.
- La population est l'ensemble des élèves de l'école ; l'échantillon est constitué des élèves des 7 salles de classe sélectionnées.

*Cours « 2 »*

*Choix*

*d'Echantillonnage*

## **1. L'échantillonnage**

Lorsqu'on veut travailler sur une très grande population, il est difficile d'en déterminer les propriétés et il est rarement possible d'avoir accès aux données relatives à l'ensemble de la population. On utilise alors un *échantillon* de cette population ainsi, cela réduit le volume de données à étudier. En statistique, la méthode d'échantillonnage ou la technique d'échantillonnage est le processus d'étude de la population en recueillant des informations et en analysant ces données (Ardilly P., 2006). C'est la base des données où l'espace d'échantillonnage est énorme.

Les avantages de l'échantillonnage : cout moindre - gain de temps, c'est la seule méthode qui donne des résultats dans le cas d'un test destructif.

Bien que le recensement soit la meilleure manière d'obtenir les informations les plus fidèles d'une population, on procèdera très souvent à un sondage. Voici quelques raisons d'effectuer un sondage au lieu d'un recensement :

- Lorsque la population est trop grande puisque cela engendre moins de dépenses monétaires (transport, temps, employés, etc.)
- Lorsqu'on ne bénéficie pas de beaucoup de temps
- Lorsque la population ciblée est difficilement accessible

De plus, il existe plusieurs méthodes pour créer un échantillon dans une population. Chaque méthode présente des avantages et des inconvénients en fonction du contexte et des besoins de l'étude.

Le terme échantillonnage fait référence au processus utilisé pour obtenir des données à partir d'un petit groupe (ou de sous-groupes). Une fois ces données collectées, elles peuvent être généralisées à un groupe plus large, tel que le marché cible d'une entreprise.

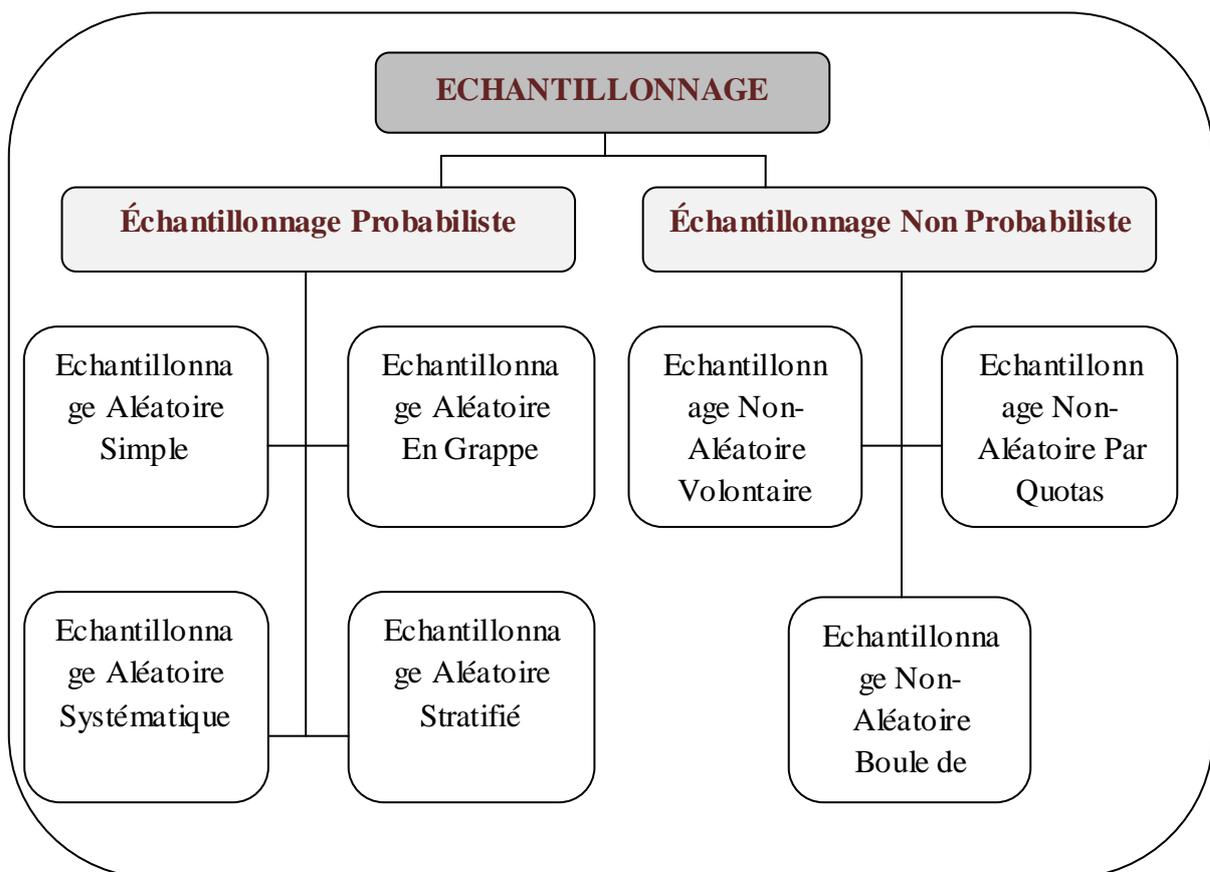
## **2. Les types d'échantillonnages**

Il existe plusieurs techniques d'échantillonnage différentes disponibles, et elles peuvent être subdivisées en deux groupes. Toutes ces méthodes d'échantillonnage peuvent impliquer de cibler spécifiquement des groupes difficiles à atteindre ou à approcher.

Les deux différents types de méthodes d'échantillonnage sont :

- l'échantillonnage *probabiliste* (où la probabilité de sélection de chaque individu est connue)
- l'échantillonnage *non probabiliste* (où la probabilité de sélection de chaque individu est inconnue).

**Figure 02: Les Différentes Méthodes d'Echantillonnage**



### **2.1. Méthodes d'échantillonnage Aléatoires/ Probabiliste**

La méthode d'échantillonnage probabiliste utilise une certaine forme de sélection aléatoire. Dans cette méthode, tous les individus éligibles ont une chance de sélectionner l'échantillon à partir de l'ensemble de l'espace échantillon. Cette méthode prend plus de temps et coûte plus cher que la méthode d'échantillonnage non probabiliste. L'avantage d'utiliser l'échantillonnage probabiliste est qu'il fournit un échantillon qui représente bien la population (Bardin B., 2016). Ainsi, les données obtenues à partir d'échantillons sélectionnés aléatoirement pourront, sous certaines conditions, être généralisées pour la population étudiée.

Les types des méthodes d'échantillonnage aléatoires :

Echantillonnage aléatoire Simple

Echantillonnage aléatoire Systématique

Echantillonnage aléatoire Par Grappe

Echantillonnage aléatoire Stratifié

### **2.1.1 Echantillonnage Aléatoire Simple**

L'échantillonnage aléatoire simple est une technique d'échantillonnage dans laquelle chaque membre d'une population a une chance égale d'être choisi grâce à l'utilisation d'une méthode de sélection non biaisée. Chaque sujet de l'échantillon reçoit un numéro, puis l'échantillon est choisi au hasard.

La méthode d'échantillonnage aléatoire est l'une des formes les plus simples et les plus courantes de collecte de données, car elle fournit une représentation impartiale d'un groupe. Le sous-ensemble aléatoire d'individus sélectionnés représente un ensemble de données complet (Pierre D., 2017). Le but de l'échantillonnage aléatoire simple est de créer un sous-ensemble gérable et équilibré d'individus représentatif d'un groupe plus large qui serait autrement trop difficile à échantillonner.

Les étapes de l'échantillonnage aléatoire simple sont :

- On se fait une liste de toutes les unités statistiques de la population et on les numérote de 1 à N.
- On choisit au hasard « n » nombres différents correspondant aux « n » unités statistiques qui devront faire partie de l'échantillon.

Quelques exemples de techniques d'échantillonnage aléatoire simples comprennent les loteries, les générateurs de nombres informatiques aléatoires ou les tirages au sort.

#### **➤ Les avantages et les inconvénients**

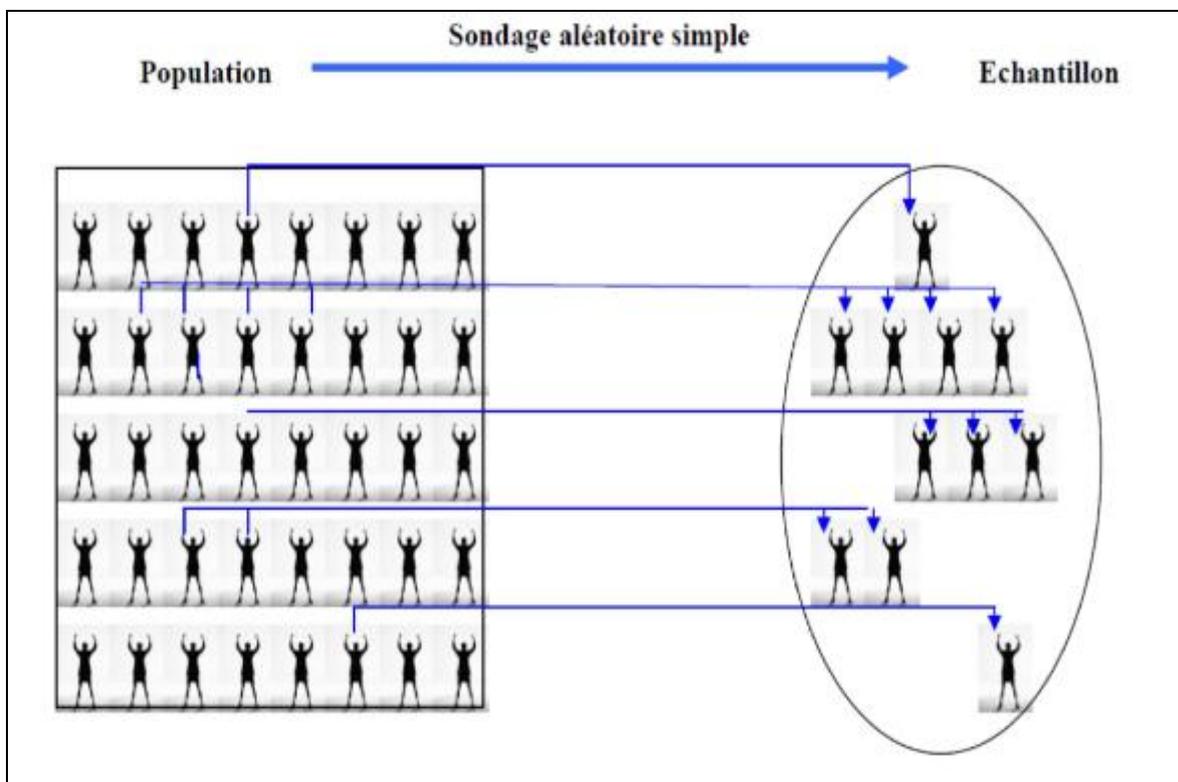
- Avantages : Simple, l'échantillon représente bien la population.

• Inconvénients : Il faut posséder une bonne base de sondage, une liste complète de toutes les unités statistiques de la population sans répétition ; aussi peut être très longue à effectuer si l'on n'utilise pas un ordinateur pour générer les grands échantillons.

➤ **Exemple**

On souhaite évaluer la satisfaction des étudiants d'une faculté qui en compte 1 000 (population) à propos de la propreté générale du lot. Pour ce faire, on décide de construire un échantillon de 400 étudiants par la méthode d'échantillonnage aléatoire. Ainsi, un ordinateur choisit au hasard le nom de 400 d'entre eux.

*Figure 03: Echantillonnage Aléatoire Simple*



**2.1.2 Echantillonnage Aléatoire Systématique**

Consiste à choisir d'abord une unité statistique d'une population au hasard puis à choisir ensuite les autres unités statistiques qui devront faire partie de l'échantillon à intervalle régulier dans la population.

Les étapes de l'échantillonnage systématique :

- On se land une liste de toutes les unités statistiques de la population et on les numérote de 1 à N.
- On détermine la taille « n » de l'échantillon voulu.
- On calcule le pas de sondage.

➤ **Les avantages et les inconvénients**

- Avantages : On peut facilement prédéterminer la taille et les éléments faisant partie de l'échantillon, L'échantillon est distribué dans des proportions égales dans la population.
- Inconvénient : De par sa caractéristique d'intervalles réguliers pour choisir les éléments, cela ne garantit pas un échantillon représentatif.

➤ **Exemple**

Supposons que les noms de 200 élèves d'une école soient triés dans l'ordre alphabétique inverse. Pour sélectionner un échantillon dans une méthode d'échantillonnage systématique, nous devons choisir une quinzaine d'étudiants en sélectionnant au hasard un numéro de départ, disons 5. À partir du numéro 5, nous sélectionnerons chaque 15e personne de la liste triée. Enfin, nous pouvons nous retrouver avec un échantillon de quelques étudiants.

### **2.1.3 Echantillonnage Aléatoire Par Grappe**

L'échantillonnage en grappes est utilisé lorsque la population cible est trop grande ou dispersée, et l'étude de chaque sujet serait coûteuse, longue et improbable. L'échantillonnage en grappes permet aux chercheurs de créer des sous-sections plus petites et plus gérables de la population présentant des caractéristiques similaires. L'échantillonnage en grappes est particulièrement utile dans les zones d'échantillonnage géographique lorsque les populations sont largement dispersées. Les chercheurs formeront des grappes basées sur une zone géographique en regroupant des individus au sein d'une communauté, d'un quartier ou d'une zone locale en une seule grappe (Vandercammen, M., 2018).

L'échantillonnage en grappes est également utilisé dans les études de marché lorsque les chercheurs ne peuvent pas collecter d'informations sur la population dans son ensemble.

Enfin, l'échantillonnage en grappes peut être utilisé pour estimer les taux de mortalité élevés, tels que les guerres, les famines ou les catastrophes naturelles.

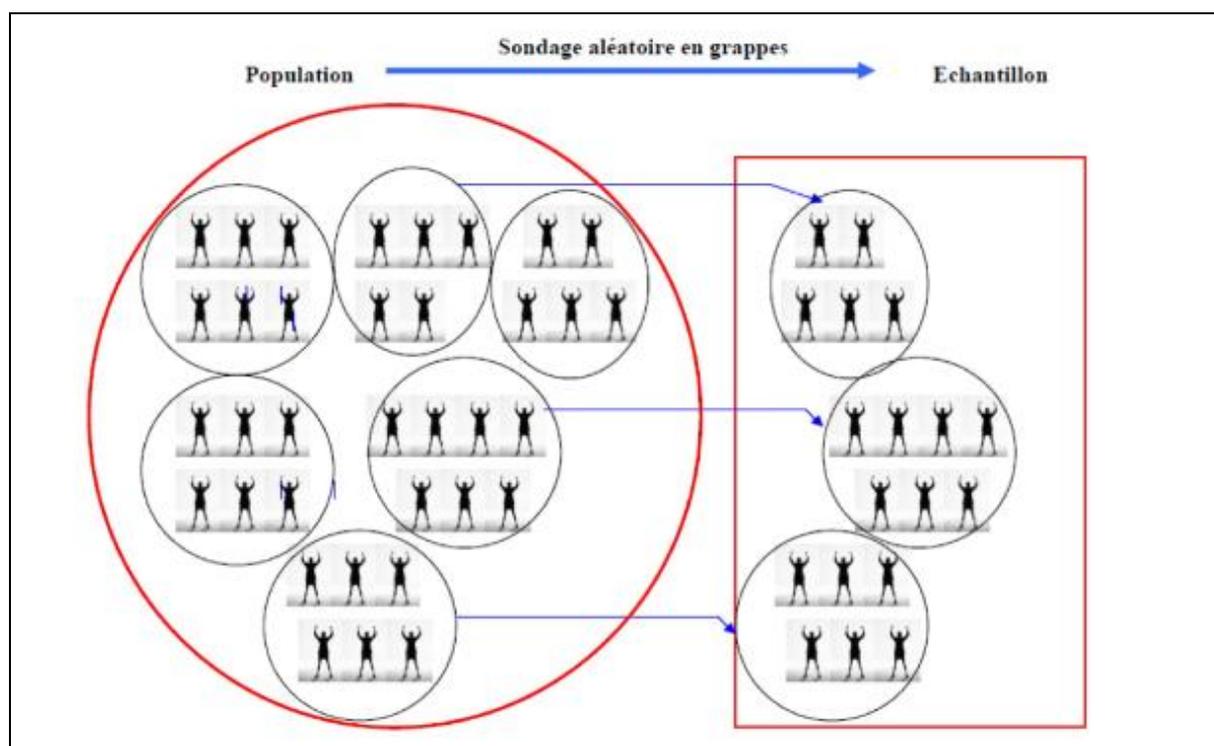
Les étapes de l'échantillonnage en Grappe :

- Choisissez la population cible que vous souhaitez étudier et déterminez la taille d'échantillon souhaitée.
- Divisez votre échantillon en grappes. Lors de la formation des clusters, assurez-vous que la population de chaque cluster est diversifiée, à une distribution de caractéristiques similaire à la distribution de la population dans son ensemble et au même nombre de membres.
- Sélectionnez les clusters par un processus de sélection aléatoire. Il est important de sélectionner au hasard parmi les clusters pour préserver la validité de vos résultats. Le nombre de grappes sélectionnées est basé sur la taille de l'échantillon.

➤ **Les avantages et les inconvénients**

- **Avantages :** Gain de temps et d'argent, l'échantillonnage en grappes est moins cher et plus rapide que les autres méthodes d'échantillonnage. Par exemple, il réduit les frais de déplacement pour de larges populations géographiques - Validité externe élevée, si votre population est correctement regroupée pour représenter toutes les caractéristiques possibles de l'ensemble de la population, vos clusters refléteront avec précision l'ensemble de la population.
- **Inconvénients :** Erreur d'échantillonnage élevée Lorsque les grappes ne reflètent pas les caractéristiques de la population ou ne servent pas de mini-représentation de la population dans son ensemble, la certitude et la précision statistiques seront moindres – Complexité, la planification des modèles d'étude pour l'échantillonnage en grappes nécessite généralement plus d'attention car les chercheurs doivent déterminer comment diviser une population plus large de manière efficace et appropriée.

**Figure 04: Echantillonnage Aléatoire En Grappe**



#### **2.1.4 Echantillonnage Aléatoire Stratifié**

L'échantillonnage stratifié est une méthode d'échantillonnage aléatoire dans laquelle les chercheurs divisent d'abord une population en sous-groupes plus petits, ou strates, en fonction des caractéristiques communes des membres, puis sélectionnent au hasard parmi ces groupes pour former l'échantillon final.

Pour maintenir cette proportion, on peut utiliser l'équation suivante :

$$\frac{\text{Nombre de membres à prendre dans cette strate}}{\text{Taille de l'échantillon}} = \frac{\text{Taille de la strate}}{\text{Taille de la population}}$$

Les étapes de l'échantillonnage stratifié :

- Définissez votre population d'intérêt et choisissez la ou les caractéristiques que vous utiliserez pour diviser vos groupes.
- Divisez votre échantillon en strates en fonction de la ou des caractéristiques pertinentes. Chaque strate doit être mutuellement exclusive, mais ensemble, elles doivent représenter l'ensemble de la population.

- Définissez la taille de l'échantillon pour chaque strate et décidez si votre échantillon sera proportionné ou disproportionné. La taille de l'échantillon dans chaque strate devrait idéalement être proportionnelle aux membres de ce groupe au sein de la population cible ou du cadre d'échantillonnage.

- Tirez un échantillon aléatoire de chaque strate et combinez-les pour former votre échantillon final.

➤ **Les avantages et les inconvénients**

- Avantages : Efficace et maniable, en organisant une population en groupes ayant des caractéristiques similaires, les chercheurs économisent du temps de collecte de données et peuvent mieux gérer un échantillon qui serait autrement trop grand pour être analysé – Précision, l'échantillonnage stratifié peut produire des estimations plus précises que l'échantillonnage aléatoire simple lorsque les membres des sous-populations sont homogènes par rapport à l'ensemble de la population. Cela donne à une étude plus de puissance statistique.

- Inconvénients : trop de différences au sein de la population, une population ne peut pas être organisée en sous-groupes s'il y a trop de différences au sein de la population ou s'il n'y a pas assez d'informations sur la population à portée de main – Planification, les chercheurs doivent s'assurer que chaque membre de la population appartient à une seule strate et que toutes les strates contiennent collectivement chaque membre de la population plus large.

## **2.2 Méthodes d'échantillonnage Non-Aléatoires/ Non-Probabiliste**

L'échantillonnage non probabiliste est une méthode d'échantillonnage qui utilise des critères non aléatoires tels que la disponibilité, la proximité géographique ou les connaissances spécialisées des personnes que vous souhaitez rechercher afin de répondre à une question de recherche. L'échantillonnage non probabiliste est utilisé lorsque les paramètres de la population sont inconnus ou impossibles à identifier individuellement. Par exemple, les visiteurs d'un site Web qui n'exige pas que les utilisateurs créent un compte pourraient faire partie d'un échantillon non probabiliste (Bardin B., 2016).

La méthode d'échantillonnage non probabiliste est une technique dans laquelle le chercheur sélectionne l'échantillon en fonction d'un jugement subjectif plutôt que d'une sélection

aléatoire. Dans cette méthode, tous les membres de la population n'ont pas la chance de participer à l'étude.

Les types des méthodes d'échantillonnage Non Aléatoire sont en outre classées en différents types, tels que l'échantillonnage de commodité, l'échantillonnage consécutif, l'échantillonnage par quotas, l'échantillonnage au jugement, l'échantillonnage en boule de neige. Ici, discutons de quelques types d'échantillonnage non probabiliste :

Echantillonnage de Volontaires

Echantillonnage par Quotas

Echantillonnage Boule de Neige

### **2.2.1 Echantillonnage Non-Aléatoire Volontaires**

L'échantillonnage de volontaire (également appelé échantillonnage auto-sélection) repose sur des participants qui acceptent volontairement de faire partie de votre recherche. Ceci est courant pour les échantillons qui ont besoin de personnes répondant à des critères spécifiques, comme c'est souvent le cas pour la recherche médicale ou psychologique. Dans l'échantillonnage d'auto-sélection, les volontaires sont généralement invités à participer par le biais d'annonces demandant à ceux qui remplissent les conditions de s'inscrire. Les volontaires sont recrutés jusqu'à ce qu'une taille d'échantillon prédéterminée soit atteinte (Pierre D., 2017).

L'auto-sélection ou l'échantillonnage volontaire comporte deux étapes :

- Faire connaître votre besoin de sujets.
- Vérifier la pertinence de chaque sujet et les inviter ou les rejeter.

**Remarque Importante** : Gardez à l'esprit que toutes les personnes qui postulent ne seront pas éligibles pour votre recherche. Il y a de fortes chances que de nombreux candidats ne lisent pas ou ne comprennent pas entièrement le sujet de votre étude, ou puissent posséder des facteurs disqualifiant. Il est important de vérifier attentivement l'éligibilité avant d'inviter des volontaires à faire partie de votre échantillon.

### **2.2.2 Echantillonnage Non-Aléatoire Par Quotas**

Dans l'échantillonnage par quota, vous sélectionnez un nombre prédéterminé ou une proportion d'unités, appelée quota. Votre quota doit comprendre des sous-groupes avec des caractéristiques spécifiques (par exemple, des individus, des cas ou des organisations) et doit être sélectionné de manière non aléatoire. Vos sous-groupes, appelés strates, doivent s'exclure mutuellement.

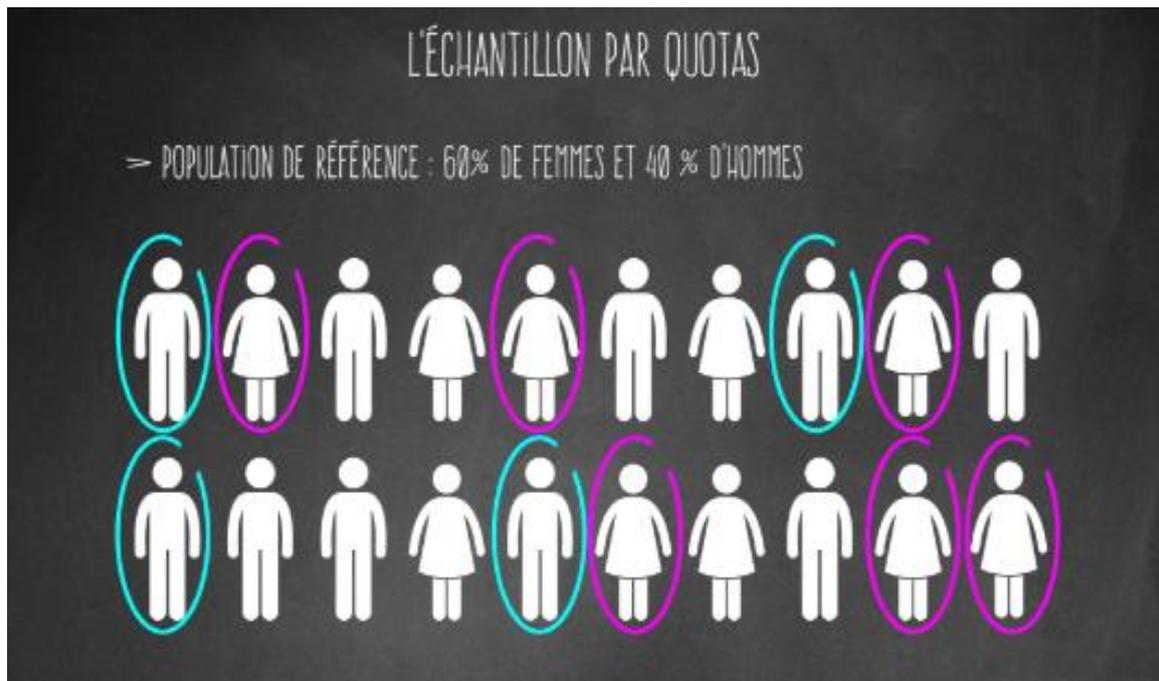
Votre estimation peut être basée sur des études antérieures ou sur d'autres données existantes, le cas échéant. Cela vous aide à déterminer le nombre d'unités à choisir dans chaque sous-groupe. Dans la phase de collecte de données, vous continuez à recruter des unités jusqu'à ce que vous atteigniez votre quota.

Il existe deux types d'échantillonnage par quota :

- L'échantillonnage par quota proportionnel est utilisé lorsque la taille de la population est connue. Cela vous permet de déterminer le quota d'individus que vous devez inclure dans votre échantillon afin d'être représentatif de votre population.
- L'échantillonnage par quota non proportionnel est utilisé lorsque la taille de la population est inconnue. Ici, c'est à vous de déterminer à l'avance le quota d'individus que vous allez inclure dans votre échantillon.

**Remarque :** Notez que l'échantillonnage par quota peut sembler similaire à l'échantillonnage stratifié, une méthode d'échantillonnage probabiliste dans laquelle vous divisez votre population en sous-groupes qui partagent une caractéristique commune. La principale différence ici est que dans l'échantillonnage stratifié, vous prélevez un échantillon aléatoire de chaque sous-groupe, tandis que dans l'échantillonnage par quotas, la sélection de l'échantillon n'est pas aléatoire, généralement via un échantillonnage de commodité. En d'autres termes, qui est inclus dans l'échantillon est laissé au jugement subjectif du chercheur.

*Figure 05: Echantillonnage Non-Aléatoire Par Quotas*

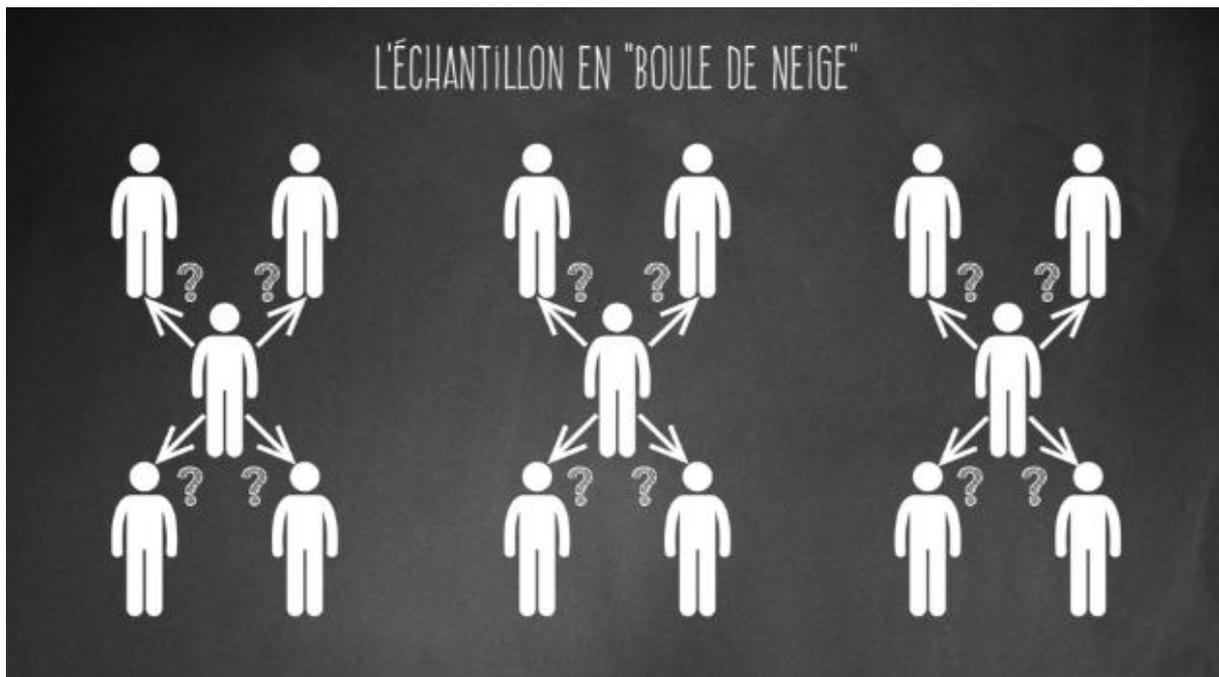


### **2.2.3 Echantillonnage Non-Aléatoire Boule de Neige**

L'échantillonnage en boule de neige est utilisé lorsque la population que vous souhaitez rechercher est difficile à atteindre, ou qu'il n'existe pas de base de données ou d'autre cadre d'échantillonnage pour vous aider à les trouver.

Pour effectuer un échantillon boule de neige, vous commencez par trouver une personne qui est prête à participer à votre recherche. Vous leur demandez ensuite de vous présenter aux autres. Alternativement, votre recherche peut impliquer de trouver des personnes qui utilisent un certain produit ou qui ont de l'expérience dans le domaine qui vous intéresse. Dans ces cas, vous pouvez également utiliser des réseaux de personnes pour accéder à votre population d'intérêt.

*Figure 06: Echantillonnage Non-Aléatoire en Boule de Neige*



## Série « 2 » d'exercices

### **Exercice 1 : Quelles méthodes d'échantillonnage utilisées ?**

1/ Aux États-Unis, tout cabinet d'avocats doit inscrire son numéro de téléphone auprès des instances du système judiciaire de l'État. Un enquêteur utilise un programme informatique pour sélectionner de façon aléatoire 500 numéros de téléphone enregistré et les cabinets d'avocats dont le numéro de téléphone a été choisi feront l'objet d'un audit.

**Quel type d'échantillon a-t-on constitué ?** Choisir une réponse :

- échantillon aléatoire simple
- Échantillon aléatoire stratifié
- Échantillon aléatoire par grappes
- Aucune de ces propositions.

2/ Chaque jour, un contrôleur de qualité dans une usine sélectionne les 10 premières pièces qu'il voit pour vérifier leur conformité.

**Quel type d'échantillon a-t-on constitué ?** Choisir une réponse :

- échantillon aléatoire simple
- Échantillon aléatoire stratifié
- Échantillon aléatoire par grappes
- Aucune de ces propositions.

3/ Une responsable qualité d'un atelier de confection veut vérifier la conformité des chemises avant leur expédition auprès d'un client. Pour cela, elle choisit au hasard 2 cartons parmi les 10 constituant la commande et vérifie la conformité des chemises dans les 2 cartons.

**Quel type d'échantillon a-t-on constitué ?** Choisir une réponse :

- échantillon aléatoire simple
- Échantillon aléatoire stratifié
- Échantillon aléatoire par grappes
- Aucune de ces propositions.

4/ Un club d'athlétisme choisit de manière aléatoires 3 athlètes dans chacune de ses équipes pour participer à une enquête sur la pratique de leur sport.

**Quel type d'échantillon a-t-on constitué ?** Choisir une réponse :

- échantillon aléatoire simple
- Échantillon aléatoire stratifié
- Échantillon aléatoire par grappes
- Aucune de ces propositions.

5/ Un enseignant souhaite sélectionner trois de ses étudiants pour participer à un défi mathématiques. Pour cela, il met les noms de tous ses étudiants dans un grand chapeau et, sans regarder, en tire 3.

**Quel type d'échantillon a-t-on constitué ?** Choisir une réponse :

- échantillon aléatoire simple
- Échantillon aléatoire stratifié
- Échantillon aléatoire par grappes
- Aucune de ces propositions.

6/ Une salle de concert propose des billets dans 3 zones : orchestre (places les plus chères), loges et balcon (places les moins chères). Les gestionnaires de cette salle souhaitent interroger 150 spectateurs au prochain concert sur leur niveau de satisfaction. Pour cette enquête, ils envisagent d'effectuer un échantillonnage stratifié, chaque zone constituant une strate.

**Exercice 2 : Quelle méthode d'échantillonnage leur permet d'obtenir un échantillon aléatoire stratifié ?** Choisir une réponse :

- Choisir comme échantillon les 150 premiers spectateurs qui arrivent.
- Utiliser un programme informatique qui sélectionne aléatoirement 150 billets. Les spectateurs possesseurs de ces tickets font alors partie de l'échantillon.
- Sélectionner aléatoirement une rangée et retenir tous les spectateurs de cette rangée dans l'échantillon. Répéter le procédé jusqu'à l'obtention de 150 spectateurs.
- Dans chaque zone, faire un échantillonnage aléatoire simple de 50 billets.

- Sélectionner chaque 10<sup>e</sup> spectateur qui arrive dans la salle jusqu'à l'obtention de 150 spectateurs.

**Exercice 3: À partir des contextes ci-dessous, déterminez le type d'échantillonnage non aléatoire utilisé**

- Une station de radio invite ses auditeurs à répondre à une question de sondage sur son site web : .....
- Un directeur de centre commercial veut connaître le montant moyen dépense par les clients lors d'une visite. Il demande à un enquêteur de conduire des entrevues auprès 200 clients, en respectant les proportions suivantes : 50% de femmes - 50% d'hommes, 1/3 de jeunes - 1/3 d'adultes - 1/3 de personnes âgées. Il doit rencontrer 100 personnes durant les jours de semaine, et 100 personnes la fin de semaine : .....
- Vous étudiez les sans-abri qui vivent dans votre ville. Vous commencez par assister à une réunion de défense du logement et vous avez une conversation avec une femme sans abri. Vous expliquez le but de votre recherche et ils acceptent de participer. Elle vous invite à un parking qui agit comme une résidence temporaire et vous donne des informations sur vous, de quelle manière elle peut ensuite vous connecter avec d'autres membres du groupe : .....

**Exercice 4 : Calculer les proportions en respectant les critères d'un échantillonnage stratifié**

Soit une population constituée de  $N = 1000$  unités statistiques répartie entre 3 classes sociales avec  $n_1 = 250$ ,  $n_2 = 400$ ,  $n_3 = 350$ . Sachant que la taille de l'échantillon à prélever  $T = 280$ , calculer les unités à prélever pour chaque classe?

*Cours « 3 »*

*Types d'Enquête*



*Partie 1 :*

*Etude Qualitative*

*Entretien*

## Les types d'entretien

Si vous menez une recherche qualitative, vous vous demandez peut-être quel est le meilleur type d'entretien pour votre étude ? Eh bien, la réponse est que cela dépend de la conception de votre recherche et de ce que vous voulez accomplir. Commençons par les différents types d'entretiens. Il existe trois types d'entretiens : non structurés, semi-structurés et structurés.

### 1.1 Entretiens Non Structurés / Non Directifs

Il s'agit d'entretiens qui se déroulent avec peu ou pas de questions d'entretien. Ils progressent souvent de la même manière qu'une conversation normale, mais cela concerne le sujet de recherche à l'étude. Il s'agit d'un style d'entretien relativement simple que les chercheurs utilisent pour établir un rapport et se sentir à l'aise avec le participant, et il est extrêmement utile lorsque les chercheurs discutent de sujets sensibles (Blanchet A. & Gotman A., 2015). Le chercheur est censé sonder les participants afin d'obtenir les informations les plus riches et les plus approfondies possibles.

#### ➤ **Les caractéristiques de l'entretien non directif**

- Non structuré.
- Entrevue de style conversationnel.
- Pas de structure particulière, l'intervieweur pose les questions au fur et à mesure qu'elles lui viennent à l'esprit.
- Pas de notation des réponses

#### ➤ **Avantages et inconvénients de l'entretien non directif**

L'entretien non structuré encourage la maturité et le type de communication ouvert. Il développe également un sens des responsabilités personnelles. Étant donné que la personne interrogée se sent détendue, elle peut divulguer des informations pertinentes qui seraient oubliées en cas de stress. L'entretien non structuré offre également la possibilité de s'éloigner du sujet, de partager trop d'informations sans rapport, ce qui peut conduire à exclure des informations spécifiques que l'intervieweur souhaite/doit connaître. Ils prennent également plus de temps qu'un entretien structuré.

## **1.2 Entretiens Semi-Structurés / Semi-Directifs**

Un entretien semi-directif est une méthode de collecte de données qui repose sur le fait de poser des questions dans un cadre thématique prédéterminé. Cependant, les questions ne sont pas définies dans l'ordre ou dans la formulation.

Les entretiens semi-structurés sont un mélange d'entretiens structurés et non structurés. Alors que quelques questions sont prédéterminées, les autres ne sont pas planifiées. Les entretiens semi-structurés sont souvent ouverts, ce qui permet une certaine flexibilité. Poser des questions fixes dans un ordre défini permet une comparaison facile entre les répondants, mais cela peut être limitant (Blanchet A. & Gotman A., 2015). Avoir moins de structure peut vous aider à voir des modèles, tout en permettant des comparaisons entre les répondants.

### **➤ Les caractéristiques de l'entretien semi-directif**

- Vous avez une première expérience en entretien. Les questions spontanées sont un défi trompeur, et il est facile de poser accidentellement une question suggestive ou de mettre un participant mal à l'aise.
- Votre question de recherche est de nature exploratoire. Les réponses des participants peuvent guider les questions de recherche futures et vous aider à développer une base de connaissances plus solide pour les recherches futures.

### **➤ Avantages et inconvénients de l'entretien semi-directif**

La combinaison d'éléments d'entretiens structurés et non structurés donne aux entretiens semi-structurés les avantages des deux : des données comparables et fiables et la possibilité de poser des questions de suivi.

La flexibilité des entretiens semi-structurés peut également diminuer leur validité. Il peut être difficile de comparer les réponses entre les participants selon la mesure dans laquelle l'intervieweur s'est éloigné de la liste de questions prédéterminée. La nature ouverte des entretiens semi-structurés peut conduire à la tentation de poser des questions orientées, ce qui entraîne un biais de l'observateur (Blanchet A. & Gotman A., 2015). À l'inverse, vos répondants peuvent également chercher à vous donner les réponses qu'ils pensent que vous voulez entendre, et aussi difficile de développer de bonnes questions d'entrevue semi-structurées.

### **1.3 Entretien Structurés / Directifs**

Les entretiens structurés sont le type d'entretien le plus systématisé. Contrairement aux entretiens semi-structurés ou non structurés, l'intervieweur utilise des questions prédéterminées dans un ordre défini.

Les entretiens structurés sont souvent fermés. Elles peuvent être dichotomiques, c'est-à-dire demander aux participants de répondre « **oui** » ou « **non** » à chaque question, ou à **choix multiples**. Bien qu'il existe des entretiens structurés ouverts, ils sont moins courants. Poser des questions fixes dans un ordre défini vous permet de comparer facilement les réponses entre les participants dans un contexte uniforme.

Cela peut vous aider à voir des modèles et à mettre en évidence des domaines pour des recherches plus approfondies, et cela peut être un outil de recherche explicatif ou exploratoire utile.

#### ➤ **Les caractéristiques de l'entretien directif**

- Vous avez déjà une compréhension très claire de votre sujet, vous possédez donc une base de référence pour concevoir des questions structurées solides.
- Vous êtes contraint en termes de temps ou de ressources et avez besoin d'analyser efficacement vos données.
- Votre question de recherche dépend d'une forte parité entre les participants, à conditions environnementales constantes.

#### ➤ **Avantages et inconvénients de l'entretien directif**

Biais réduit, la nature fixe des entretiens structurés réduit les effets de contexte et autres biais. Poser les mêmes questions dans le même ordre à tous les participants minimise le risque d'introduire un biais via l'ordre ou la nature des questions posées, ou via tout facteur environnemental. Crédibilité, fiabilité et validité accrues. Simple, économique et efficace. Bien que similaires aux questionnaires et aux enquêtes, les entretiens structurés introduisent plus de nuances et de richesse dans le sujet étudié sans représenter trop de travail pour l'enquêteur. De même, il y a moins de préparation nécessaire pour la personne interrogée, de sorte que le processus prend également moins de temps de son côté.

La rigidité des entretiens structurés signifie qu'il y a très peu de possibilités d'établir un rapport entre l'intervieweur et le participant. La formalité perçue des entretiens structurés peut rendre les participants mal à l'aise ou nerveux, ce qui peut affecter leurs réponses. Flexibilité limitée, une fois les questions sélectionnées, elles ne peuvent être modifiées ou supprimées sans nuire à la qualité de l'entretien. Même si une question est mal formulée, superflue ou inutile, elle doit quand même être présentée à tous les répondants.

## **2. Comment préparer un entretien de recherche ?**

### **➤ Définir le statut et le type d'entretien à mener**

Pour réaliser un entretien de manière professionnelle, vous devez répondre à l'avance à plusieurs questions structurantes :

- Quel est le but de mon entretien ?
- Combien de temps ai-je à ma disposition ?
- Quelles informations dois-je obtenir ?
- Quel comportement dois-je adopter ?
- Quel type d'étude utiliser ? L'étude qualitative ou quantitative ?
- Quel est le statut de mon entretien ? Entretien exploratoire, principal ou de contrôle ?
- Vers quel type d'entretien est-ce que je veux me tourner ? Entretien directif, semi-directif ou libre ?
- Qui vais-je interroger ? Pourquoi ?

### **➤ Définir un guide d'entretien**

Le guide d'entretien est utilisé dans les entretiens qualitatifs pour structurer la discussion. Il aide à déterminer un flux « logique » pour aborder les sujets. Nous recommandons de structurer les questions selon le principe de l'entonnoir :

Commencer par le sujet le plus général et le plus simple à gérer pour le répondant et aller progressivement vers les sujets plus spécifiques qui demandent réflexion.

Il sert également de liste de contrôle pour l'intervieweur afin de l'aider à couvrir tous les sujets qui doivent être couverts dans le projet.

Cependant, c'est une erreur de penser que le guide d'entretien est immuable. Il doit être amélioré au fur et à mesure des entretiens pour couvrir les sujets qui ressortent spontanément

lors des entretiens qualitatifs.

Réorganisez l'ordre des questions pour refléter la séquence naturelle des événements au cours de l'entretien.

Reformuler les questions qui posent problème.

➤ **La construction d'un guide d'entretien**

Les grandes étapes de la construction d'un guide d'entretien sont:

*1/ Définir l'objectif de l'entretien :* Tout entretien qualitatif s'inscrit nécessairement dans un projet de recherche plus large. Il peut s'agir d'études sociologiques, d'études de marché, etc. Il faut donc être très clair sur les objectifs communs de l'entretien. Il faut notamment bien définir LE sujet principal, celui autour duquel s'articulera l'entretien.

*2/ La formulation des questions :* Une fois votre séquence logique de thèmes déterminée, il ne reste plus qu'à formuler les questions. Les règles de formulation sont un peu différentes de celles d'une enquête quantitative (sondage) :

- La question doit être aussi claire et sans ambiguïté que possible.
- Évitez les formulations alambiquées et les mots trop complexes (sauf si l'entretien porte sur un sujet technique, bien sûr).
- Ne pas prendre position Utilisez des mots « neutres » pour éviter les préjugés.

*3/ Les consignes :* La dernière étape consiste à inclure des notes sous les questions. Ces notes sont destinées à vous aider à rebondir, à donner des instructions aux enquêteurs et à proposer des formulations alternatives.

*Partie 2 :*

*Etude Quantitative*

*Questionnaire*

Un questionnaire est une liste de questions ou d'éléments utilisés pour recueillir des données auprès des répondants sur leurs attitudes, leurs expériences ou leurs opinions. Les questionnaires peuvent être utilisés pour recueillir des informations quantitatives et/ou qualitatives.

Les questionnaires sont couramment utilisés dans les études de marché ainsi que dans les sciences sociales et de la santé. Par exemple, une entreprise peut demander des commentaires sur une expérience récente de service client, ou des chercheurs en psychologie peuvent enquêter sur les perceptions des risques pour la santé à l'aide de questionnaires.

## **1. Les types de questionnaires**

Les questionnaires peuvent être auto-administrés ou administrés par le chercheur. Les questionnaires auto-administrés sont plus courants car ils sont faciles à mettre en œuvre et peu coûteux, mais les questionnaires administrés par les chercheurs permettent des informations plus approfondies.

### **➤ Questionnaires auto-administrés**

Les questionnaires auto-administrés peuvent être livrés en ligne ou en format papier ou par courrier. Toutes les questions sont standardisées afin que tous les répondants reçoivent les mêmes questions avec un libellé identique.

Les questionnaires auto-administrés peuvent être rentable, facile à gérer pour petits et grands groupes, anonyme et adapté aux sujets sensibles, à son rythme mais ils peuvent aussi être ne convient pas aux personnes dont l'alphabétisation ou les compétences verbales sont limitées, susceptible d'un biais de non-réponse (la plupart des personnes invitées peuvent ne pas remplir le questionnaire), biaisé envers les personnes qui font du bénévolat parce que les demandes d'enquête impersonnelles sont souvent ignorées (Fenneteau H., 2015).

### **➤ Questionnaires administrés par les chercheurs**

Les questionnaires administrés par les chercheurs sont des entretiens qui ont lieu par téléphone, en personne ou en ligne entre les chercheurs et les répondants.

Les questionnaires administrés par les chercheurs peuvent vous aider à vous assurer que les répondants sont représentatifs de votre public cible, permettre des éclaircissements sur des questions et réponses ambiguës ou peu claires, ont des taux de réponse élevés parce qu'il est

plus difficile de refuser une entrevue lorsqu'une attention personnelle est accordée aux répondants; mais les questionnaires administrés par les chercheurs peuvent être limités en termes de ressources. Ils sont coûteuse et longue à réaliser, plus difficile à analyser si vous avez des réponses qualitatives, susceptible de contenir un biais de l'expérimentateur ou des caractéristiques de la demande.

## **2. Comment rédiger un questionnaire d'une enquête quantitative ?**

Suivez les étapes ci-dessous pour concevoir un questionnaire :

1/ Choisir le type de question d'enquête quantitative : Sélectionnez un type qui correspond à l'objectif de l'enquête. Le type de réponses données par les répondants est déterminé par le type de question posée.

2/ Identifier le groupe cible ainsi que les variables dépendantes et indépendantes : L'étape suivante consiste à choisir le public cible et les variables à mesurer.

3/ Donner une structure appropriée à l'enquête : Une fois les 2 premiers points réglés, il est maintenant temps de structurer l'enquête de manière à ce qu'elle soit simple. Des formulations simples et sans ambiguïté doivent être utilisées pour formuler les questions.

## **3. Comment formuler les questions ?**

Votre questionnaire peut comprendre des **questions ouvertes** ou **fermées** ou une combinaison des deux. L'utilisation de questions fermées limite vos réponses, tandis que les questions ouvertes permettent un large éventail de réponses (Fenneteau H., 2015). Vous devrez équilibrer ces considérations avec votre temps et vos ressources disponibles.

### **3.1 Questions fermées**

Les questions fermées ou à choix restreint offrent aux répondants un ensemble fixe de choix parmi lesquels choisir. Les questions fermées sont les meilleures pour collecter des données sur des variables catégorielles ou quantitatives.

La possibilité de réponses est limitée le plus souvent à OUI/NON ou des choix prédéfinis pour un QCM.

### ➤ Exemples des questions fermées

1/ Quel âge avez-vous?

- 15 ans ou moins
- 16–35
- 36–60
- 61 ans ou plus.

2/ Dans quelle mesure êtes-vous satisfait ou insatisfait de votre expérience d'achat en ligne ?

- Très insatisfait
- Plutôt insatisfait
- Ni satisfait ni insatisfait
- Plutôt satisfait
- Très satisfait

3/ Êtes-vous satisfait de notre service client ?

- Oui
- Non

### **3.2 Questions ouvertes**

Les questions ouvertes, permettent aux répondants de donner des réponses dans leurs propres mots. Parce qu'il n'y a aucune restriction sur leurs choix, les répondants peuvent répondre d'une manière que les chercheurs n'auraient peut-être pas envisagée autrement.

Les questions ouvertes ont quelques inconvénients. Ils exigent plus de temps et d'efforts de la part des répondants, ce qui peut les dissuader de remplir le questionnaire.

Pour les chercheurs, comprendre et résumer les réponses à ces questions peut prendre beaucoup de temps et de ressources. Vous devrez développer un schéma de codage systématique pour catégoriser les réponses, et vous devrez peut-être également impliquer d'autres chercheurs dans l'analyse des données pour une fiabilité élevée.

### ➤ Exemples des questions ouvertes

- Selon vous, quel est le plus gros obstacle à la productivité dans le travail à distance ?
- Comment proposez-vous que nous améliorions notre service pour votre expérience ?
- Selon vous, quels sont les points forts de ce produit ?

## 4. Les erreurs à éviter dans un questionnaire

Voici une liste d'erreurs à éviter lors de l'établissement d'un questionnaire d'étude de marché :

- *Les questions doubles* : Evitez de poser des questions qui appellent plusieurs réponses, gardez vos questions ciblées.
- *Utilisation de plusieurs champs* : Améliorez le confort de vos répondants. Préférez les listes plutôt que les champs.
- *Être ambigu* : Pour plus de lisibilité et de clarté, évitez le jargon ou un langage trop complexe.
- *Évitez les questions orientées* : Les questions suggestives guident les répondants vers des réponses spécifiques, même si ce n'est pas ce qu'ils ressentent vraiment, en leur fournissant explicitement ou implicitement des informations supplémentaires.
- *Évitez la négation dans une question* : Cela peut également entraîner des incompréhensions. Par exemple : "Ne souhaiteriez-vous pas avoir un téléphone dans votre maison ?".

Une fois les résultats de l'enquête obtenus et compilés sous forme de tableaux ou de texte, le porteur de projet peut analyser les éléments importants et rédiger un rapport d'étude. En fonction des résultats, il peut être nécessaire d'adapter le produit ou service aux attentes des clients qui ont été caractérisées lors de l'enquête.

## **5. Comment diffuser le questionnaire?**

*Tableau : Les Différents Modes de Diffusion du Questionnaire*

|                                      |   |
|--------------------------------------|---|
| <b>En ligne<br/>(le Web)</b>         | Les enquêtes en ligne utilisent différents canaux qui sont utilisés par le public pour envoyer des questionnaires. Cette méthode d'enquête est rentable et permet également de gagner du temps. Avec le loisir de répondre à tout moment, le public peut répondre selon son propre horaire.   |
| <b>Appel<br/>Téléphonique</b>        | Un appel téléphonique au répondant est un moyen rapide et direct de collecter des données. Cependant, cela coûte cher et le répondant peut être réticent à donner beaucoup d'informations. De plus, la taille de l'échantillon de vos données est limitée à une autre forme d'enquête.  |
| <b>En personne<br/>(Face à Face)</b> | Ce type de recherche est généralement utilisé par les chercheurs qui rencontrent l'intervenant sur leur lieu de travail ou à domicile. Dans le confort de son environnement naturel, une personne peut fournir les informations requises. Cependant, c'est un processus lent et aussi coûteux en raison de tous les déplacements que vous auriez à faire. |
| <b>E-mail</b>                        | Dans ce cas, le répondant reçoit une copie physique du questionnaire qu'il est invité à remplir et à renvoyer. Cependant, le problème avec cette méthode est qu'elle est coûteuse pour un chercheur ainsi que pour l'intervenant. L'intervenant peut également choisir de l'ignorer car cela prend du temps   |

## Prêt à commencer?

Une étape cruciale qu'ils peuvent franchir consiste à effectuer des tests préalables. Le **pré-test**, comme son nom l'indique, consiste à envoyer le sondage à un échantillon de personnes avant que la recherche proprement dite ne soit menée. Cela peut aider le chercheur à comprendre comment les répondants interagissent avec l'enquête, y compris combien de temps cela prend, quelles questions sont confuses ou peu claires et si l'une de vos questions est controversée (Fenneteau H., 2015). De cette manière, les tests préalables sont un moyen nécessaire et inestimable pour les chercheurs d'identifier les problèmes de leurs enquêtes avant que la recherche ne soit menée.

## Série « 3 » d'exercices

### Exercice 1: Reformulez les questions suivantes

Les questions suggestives sont formulées de manière à diriger les réponses des utilisateurs vers une option spécifique. Une enquête utilise une question suggestive si le chercheur veut que les répondants répondent d'une certaine manière. Le problème est que vous ne connaîtrez jamais vraiment les avis honnêtes des utilisateurs.

Dans le cas des questions suivantes, il est probable que le répondant ne sera pas en mesure de bien répondre. Alors formulez mieux chaque question :

1/ Êtes-vous satisfait (e) de notre produit et de la livraison ?

2/ Dans quelle mesure êtes-vous satisfait du salaire et des avantages professionnels de votre emploi actuel ?

3/ Regardez-vous toujours la télévision ?

- Oui
- Non

4/ Possédez-vous une tablette ?

5/ Je n'ai pas aimé visiter votre restaurant !

- Vrai
- Faux

6/ Notre restaurant a été classé 5 étoiles par nos clients. Comment noteriez-vous votre expérience?

### Exercice 2 : Soulignez la bonne réponse la bonne réponse

Q1/ L'étude quantitative vise à :

1. la représentativité des résultats
2. neutraliser l'influence de l'inconscient
3. recenser les composantes d'un problème.

Q2/ L'objectif majeur d'une étude qualitative est de :

1. pondérer les variables
2. regrouper des hypothèses ou des variables
3. formuler des hypothèses.

Q3/ La population au sens statistique :

1. est toute personne habitant un pays donné
2. l'ensemble des objets possédant l'information désirée
3. tous les individus ayant répondu au questionnaire

Q4/ La base de sondage est :

1. la liste à partir de laquelle on va sélectionner l'échantillon
2. l'urne dans laquelle on effectue le tirage
3. l'ensemble des individus qui vont figurer dans l'échantillon

Q5/ Le guide d'entretien comporte :

1. une série de thèmes précis devant être abordés par l'interviewé
2. une série de questions vagues laissées à l'appréciation de l'enquêteur
3. une liste de thèmes servant de canevas pour le déroulement de l'interview

Q6/ Le pré-test du questionnaire intervient :

1. pour préparer la rédaction du questionnaire
2. pour vérifier que le questionnaire a bien été administré
3. pour repérer les erreurs figurants dans le questionnaire

*Cours « 4 »*

*Le Recueil  
des Données*

## Après l'enquête!

Vous avez donc lancé votre enquête. Que vous disent les données, les réponses? Une enquête n'est qu'une collecte de données, elle ne vous dira rien par elle-même. La première chose que vous devrez faire est de coder votre sondage afin qu'il soit prêt pour l'analyse. L'enquête qui a été montrée au début du chapitre a été réalisée avec un stylo et du papier, elle a donc dû être saisie manuellement dans une feuille de calcul pour l'analyse. Comme nous en avons discuté, la recherche quantitative signifie généralement l'utilisation de chiffres et de techniques statistiques, de sorte que toutes les questions que vous avez posées doivent être converties.

### 1. Des données: Qualitatives et Quantitatives

Les données, il s'agit d'un ensemble de mesures ou d'observations, divisées en deux types différents : qualitatifs et quantitatifs.

**Les données qualitatives** font référence à des informations sur les qualités ou à des informations qui ne peuvent pas être mesurées. Il est généralement descriptif et textuel. Comme exemples : la couleur des yeux de quelqu'un ou le type de voiture qu'il conduit. Dans les sondages, il est souvent utilisé pour catégoriser les réponses « oui » ou « non ».

**Les données quantitatives** sont numériques. Il est utilisé pour définir les informations qui peuvent être comptées. Quelques exemples de données quantitatives incluent la distance, la vitesse, la taille, la longueur et le poids.

Il est facile de se souvenir de la différence entre les données qualitatives et quantitatives, car l'une fait référence aux qualités et l'autre aux quantités.

### 2. Échelles de mesure

Les échelles de mesure sont la façon dont les variables sont définies et catégorisées. Le psychologue Stanley Stevens a développé les quatre échelles de mesure courantes : *nominale*, *ordinale*, *intervalle* et *ratio*. Chaque échelle de mesure a des propriétés qui déterminent comment analyser correctement les données (Delacroix, E., et al, 2021). Les propriétés évaluées sont l'identité, la magnitude, les intervalles égaux et une valeur minimale de zéro.

### ➤ **Échelle Nominale**

L'échelle nominale de mesure définit la propriété d'identité des données. Cette échelle a certaines caractéristiques, mais n'a aucune forme de signification numérique. Les données peuvent être placées dans des catégories mais ne peuvent pas être multipliées, divisées, ajoutées ou soustraites les unes des autres. Il n'est pas non plus possible de mesurer la différence entre les points de données.

Des exemples de données nominales incluent la couleur des yeux et le pays de naissance.

Les données nominales peuvent à nouveau être décomposées en trois catégories :

- *Nominal avec ordre*: Certaines données nominales peuvent être sous-catégorisées dans l'ordre, comme « froid, chaud, chaud et très chaud ».
- *Nominal sans ordre*: Les données nominales peuvent également être sous-catégorisées comme nominales sans ordre, comme masculin et féminin.
- *Dichotomique*: les données dichotomiques sont définies en n'ayant que deux catégories ou niveaux, tels que "oui" et "non".

### ➤ **Échelle Ordinale**

L'échelle ordinale définit les données qui sont placées dans un ordre spécifique. Bien que chaque valeur soit classée, aucune information ne spécifie ce qui différencie les catégories les unes des autres. Ces valeurs ne peuvent pas être ajoutées ou soustraites (Delacroix, E., et al, 2021).

Un exemple de ce type de données inclurait des points de données de satisfaction dans une enquête, où «un = heureux, deux = neutre et trois = mécontent». Lorsqu'une personne a terminé dans une course, décrit également des données ordinales.

### ➤ **Échelle d'intervalle**

L'échelle d'intervalle contient les propriétés des données nominales et ordonnées, mais la différence entre les points de données peut être quantifiée. Ce type de données montre à la fois l'ordre des variables et les différences exactes entre les variables. Ils peuvent être ajoutés ou soustraits les uns des autres, mais pas multipliés ou divisés.

Par exemple, 40 degrés n'est pas 20 degrés multiplié par deux. Cette échelle est également caractérisée par le fait que le nombre zéro est une variable existante. Dans l'échelle ordinale, zéro signifie que les données n'existent pas. Dans l'échelle d'intervalle, zéro a une signification, par exemple, si vous mesurez des degrés, zéro a une température (Marc X & Tcherna J., 2018).

Cette échelle est utilisée pour quantifier la différence entre les variables, tandis que les deux autres échelles sont utilisées pour décrire uniquement des valeurs qualitatives. D'autres exemples d'échelles d'intervalle incluent l'année de fabrication d'une voiture ou les mois de l'année.

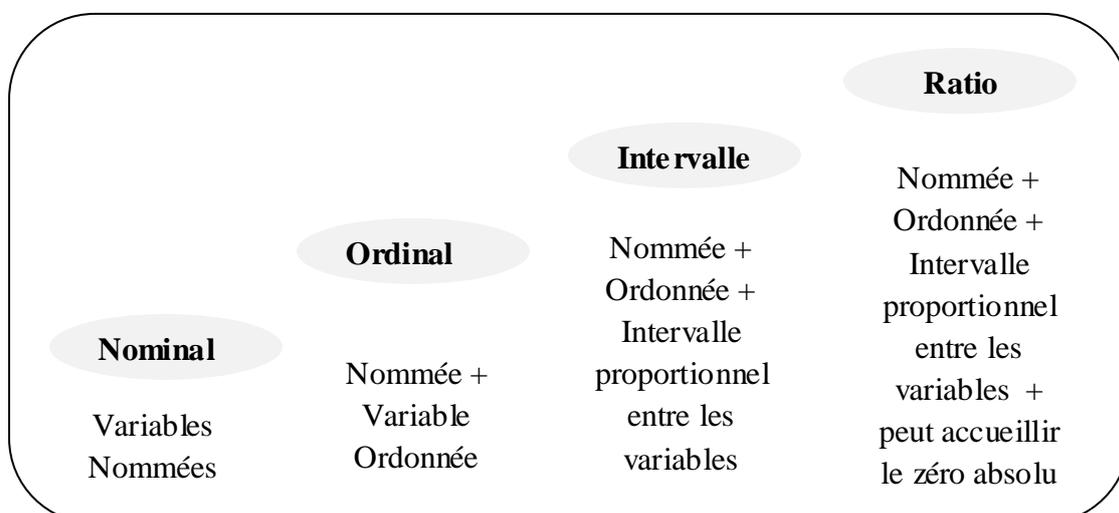
➤ **Échelle de Ratio**

Les échelles de mesure de ratio incluent les propriétés des quatre échelles de mesure. Les données sont nominales et définies par une identité, peuvent être classées dans l'ordre, contiennent des intervalles et peuvent être décomposées en valeur exacte.

Le poids, la taille et la distance sont tous des exemples de variables de ratio. Les données de l'échelle de ratio peuvent être additionnées, soustraites, divisées et multipliées. Les échelles de ratio diffèrent également des échelles d'intervalle en ce que l'échelle a un «vrai zéro». Le nombre zéro signifie que les données n'ont pas de point de valeur (Delacroix, E., et al, 2021).

Un exemple de ceci est la taille ou le poids, car quelqu'un ne peut pas mesurer zéro centimètre ou peser zéro kilo.

*Figure 07: Les Quatre Types d'Echelles de Mesure*



En résumé, les échelles nominales sont utilisées pour étiqueter ou décrire des valeurs. Les échelles ordinales sont utilisées pour fournir des informations sur l'ordre spécifique des points de données, principalement observés dans l'utilisation des enquêtes de satisfaction. L'échelle d'intervalle est utilisée pour comprendre l'ordre et les différences entre eux. Les échelles de rapport donnent plus d'informations sur l'identité, l'ordre et la différence, ainsi qu'une ventilation des détails numériques dans chaque point de données.

## Série « 4 » d'exercices

### Exercice : Pour les questions suivantes, identifier l'Echelle de Mesure ?

1/ Veuillez indiquer votre sexe :

- Masculin
- Féminin

2/ De quel groupe d'âge faites-vous partie ?

- < 18
- 18 - 28
- 29 - 39
- > 40

3/ Dans quelle mesure êtes-vous satisfait du produit que vous avez reçu ?

- Très satisfait
- Satisfait
- Insatisfait
- Très insatisfait

4/ Quelle est votre taille?

- ..... mètres

5/ En quelle année êtes-vous né?

- .....

6/ Quel sport aimes-tu pratiquer ?

- Criquet
- Football
- Volley-ball
- Autre

7/ Un enseignant enregistre l'ordre dans lequel ses étudiants terminent leurs questions. Le premier à finir, le deuxième... Quelle échelle de la mesure est-elle utilisée ?

## **CHAPITRE 2 :**

# **Traitement des Données**

*Cours « 1 »*

*Types de  
Tableaux  
Analysables*

## Utilisation de données quantitatives et qualitatives dans les statistiques !

Une fois que les chercheurs ont un ensemble de données concluant à partir de leur échantillon, ils peuvent commencer à utiliser les informations pour tirer des descriptions et des conclusions. Pour ce faire, ils peuvent utiliser à la fois des statistiques descriptives

On distingue trois types de tableaux en statistique : les tableaux de données, les tableaux de distribution des variables et enfin les tableaux de contingence.

### 1. Tableaux de Données

#### ➤ Qu'est-ce qu'un tableau de Données ?

Après avoir collecté les données, l'enquêteur doit trouver des moyens de les condenser sous forme de tableau pour étudier leurs caractéristiques saillantes. Un tel arrangement s'appelle la représentation des données.

- Ordre de série de l'ordre alphabétique
- Ordre croissant
- Ordre décroissant

Si le nombre d'observations est énorme, l'organisation des données par ordre croissant, décroissant ou sériel est une tâche fastidieuse. Ainsi, pour faciliter la compréhension et la clarté, nous pouvons tabuler les données sous la forme d'un tableau, cette table s'appelle une *table de données*.

Un tableau de données est un document composé de colonnes, de lignes et de cellules qui contiennent des valeurs spécifiques. Ils stockent des informations que les utilisateurs peuvent récupérer ultérieurement et mettre à jour si nécessaire. Le titre du tableau de données, les en-têtes de colonne et les en-têtes de ligne peuvent aider un utilisateur à comprendre plus clairement les informations du tableau (Katerine R., et al, 2019). Le document traite également les données par le biais de diverses opérations, telles que le classement, l'organisation, le filtrage et la recherche.

#### ➤ L'utilisation d'un tableau de données

Pour utiliser efficacement les tableaux de données, tenez compte de ces conseils utiles :

- Sélectionnez le bon mode de traitement des données.
- Simplifiez votre tableau de données.
- Choisissez un format accessible

## **2. Tableaux de Distribution de Fréquence**

### **➤ Qu'est-ce qu'un tableau de Distribution de Fréquence ?**

La fréquence d'une valeur est le nombre de fois qu'elle apparaît dans un ensemble de données. Une distribution de fréquence est le modèle de fréquences d'une variable. C'est le nombre de fois que chaque valeur possible d'une variable apparaît dans un ensemble de données.

Une distribution de fréquence décrit le nombre d'observations pour chaque valeur possible d'une variable. Les distributions de fréquences sont représentées à l'aide de graphiques et de tableaux de fréquences (Vandercammen, M., 2018).

### **➤ Types de distributions de fréquence**

Il existe quatre types de distributions de fréquences :

- *Distributions de fréquences non groupées* : le nombre d'observations de chaque valeur d'une variable. Vous pouvez utiliser ce type de distribution de fréquence pour les variables qualitatives.
- *Distributions de fréquence groupées* : le nombre d'observations de chaque intervalle de classe d'une variable. Les intervalles de classe sont des regroupements ordonnés des valeurs d'une variable. Vous pouvez utiliser ce type de distribution de fréquence pour les variables quantitatives.
- *Distributions de fréquences relatives* : la proportion d'observations de chaque valeur ou intervalle de classe d'une variable. Vous pouvez utiliser ce type de distribution de fréquence pour tout type de variable lorsque vous êtes plus intéressé par la comparaison des fréquences que par le nombre réel d'observations.
- *Distributions de fréquences cumulées* : somme des fréquences inférieures ou égales à chaque valeur ou intervalle de classe d'une variable. Vous pouvez utiliser ce type de distribution de fréquence pour les variables ordinales ou quantitatives lorsque vous souhaitez comprendre la fréquence à laquelle les observations tombent en dessous de certaines valeurs.

### ➤ **Comment faire un tableau de fréquence**

Les distributions de fréquences sont souvent affichées à l'aide de tableaux de fréquences. Un tableau de fréquence est un moyen efficace de résumer ou d'organiser un ensemble de données. Il est généralement composé de deux colonnes : Les valeurs ou intervalles de classe  
Leurs fréquences La méthode de création d'un tableau de fréquences diffère entre les quatre types de distributions de fréquences.

### **3. Tableau de Contingence**

#### ➤ **Qu'est-ce qu'un tableau de contingence ?**

Un tableau de contingence affiche les fréquences des combinaisons de deux variables catégorielles. Les analystes désignent également les tableaux de contingence comme des tableaux croisés et des tableaux à double entrée. Les tableaux de contingence classent les résultats pour une variable en lignes et l'autre en colonnes. Les valeurs aux intersections des lignes et des colonnes sont des fréquences pour chaque combinaison unique des deux variables (Katerine R., et al, 2019).

#### ➤ **Distributions marginales et conditionnelles dans les tableaux de contingence**

Les tableaux de contingence sont un moyen excellent de trouver des distributions marginales et conditionnelles. Ces deux distributions sont des types de distributions de fréquence.

##### *1/ Répartition marginale*

Ces distributions représentent la distribution de fréquence d'une variable catégorielle sans tenir compte des autres variables. Sans surprise, vous pouvez trouver ces distributions dans les marges d'un tableau de contingence.

##### *2/ Répartition conditionnelle*

Pour ces distributions, vous spécifiez la valeur de l'une des variables dans le tableau de contingence, puis évaluez la distribution des fréquences pour l'autre variable. En d'autres termes, vous conditionnez la distribution de fréquence pour une variable en définissant une valeur pour l'autre variable. Cela peut sembler compliqué, mais c'est facile en utilisant un tableau de contingence.

➤ **Exemple de tableau de contingence**

L'exemple de tableau de contingence ci-dessous affiche les ventes d'ordinateurs dans notre magasin fictif. Plus précisément, il décrit les fréquences de vente en fonction du sexe du client et du type d'ordinateur acheté. C'est un tableau à double entrée.

**Tableau: Les Fréquences des ventes d'ordinateurs par Sexe et Type d'ordinateur**

|         | PC | Mac | Totales |
|---------|----|-----|---------|
| Homme   | 66 | 40  | 106     |
| Femme   | 30 | 87  | 117     |
| Totales | 96 | 127 | 223     |

Dans ce tableau de contingence, les colonnes représentent les types d'ordinateurs et les lignes représentent les sexes. Les valeurs des cellules sont des fréquences pour chaque combinaison de sexe et de type d'ordinateur. Les totaux sont dans les marges.

Il est facile de voir comment les tableaux à double entrée organisent vos données et brosent un tableau des résultats. Vous pouvez facilement voir les fréquences pour toutes les combinaisons de sous-ensembles possibles ainsi que les totaux pour les hommes, les femmes, les PC et les Mac.

Par exemple, 66 hommes ont acheté des PC tandis que les femmes ont acheté 87 Mac. De plus, il y a 117 femmes, 106 hommes, 96 ventes de PC, 127 ventes de Mac et un grand total de 223 observations dans l'étude.

La *distribution marginale* du sexe sans tenir compte du type d'ordinateur est la suivante :  
Homme : 106 - Femme : 117

La *distribution marginale* des types d'ordinateurs est la suivante :  
PC : 96 - Mac : 127

La distribution conditionnelle du type d'ordinateur pour les femmes est la suivante :  
PC : 30 - Mac : 87

Alternativement, la distribution conditionnelle du genre pour les Mac est la suivante :  
Homme : 40 - Femmes : 87

*Cours « 2 »*

*L'Analyse en  
Composantes  
Principales*

L'analyse en composantes principales (ACP) est un outil indispensable pour la visualisation et la réduction de la dimensionnalité pour la science des données, mais elle est souvent enfouie dans des mathématiques compliquées.

## **1. Qu'est-ce que l'Analyse en Composantes Principales ?**

L'analyse en composantes principales, ou ACP, est une méthode de réduction de la dimensionnalité qui est souvent utilisée pour réduire la dimensionnalité de grands ensembles de données, en transformant un grand ensemble de variables en un plus petit qui contient encore la plupart des informations dans le grand ensemble.

Réduire le nombre de variables d'un ensemble de données se fait naturellement au détriment de la précision, mais l'astuce dans la réduction de la dimensionnalité consiste à échanger un peu de précision contre la simplicité. Parce que les ensembles de données plus petits sont plus faciles à explorer et à visualiser et rendent l'analyse des points de données beaucoup plus facile et plus rapide pour les algorithmes d'apprentissage automatique sans variables étrangères à traiter (Kurita, T., 2014).

L'ACP peut donc être considéré comme une méthode d'exploration de données car elle permet d'extraire facilement des informations de grands ensembles de données. Elle peut être utilisée à plusieurs fins, notamment :

- L'étude et la visualisation des corrélations entre les variables, afin d'éventuellement limiter le nombre de variables à mesurer par la suite ;
- L'obtention de facteurs non corrélés qui sont des combinaisons linéaires des variables de départ, afin d'utiliser ces facteurs dans des méthodes de modélisation telles que la régression linéaire, la régression logistique ou l'analyse discriminante ;
- La visualisation des observations dans un espace à deux ou trois dimensions, afin d'identifier des groupes homogènes d'observations, ou au contraire des observations atypiques.

Pour résumer, l'idée de l'ACP est simple : Réduire le nombre de variables d'un ensemble de données, tout en préservant autant d'informations que possible.

## **2. Explication étape par étape de l'ACP**

### **➤ Étape 1 : Normalisation**

Le but de cette étape est de normaliser l'étendue des variables initiales continues afin que chacune d'entre elles contribue de manière égale à l'analyse. Plus précisément, la raison pour laquelle il est essentiel d'effectuer une normalisation avant l'ACP, est que cette dernière est assez sensible aux variances des variables initiales.

Autrement dit, s'il y a de grandes différences entre les plages de variables initiales, les variables avec des plages plus grandes domineront sur celles avec de petites plages (par exemple, une variable qui se situe entre 0 et 100 dominera sur une variable qui se situe entre 0 et 1), ce qui conduira à des résultats biaisés. Ainsi, la transformation des données à des échelles comparables peut éviter ce problème.

### **➤ Étape 2 : Calcul de la matrice de covariance**

Le but de cette étape est de comprendre comment les variables de l'ensemble de données d'entrée s'écartent de la moyenne les unes par rapport aux autres, ou en d'autres termes, de voir s'il existe une relation entre elles. Parce que parfois, les variables sont fortement corrélées de telle sorte qu'elles contiennent des informations redondantes. Ainsi, afin d'identifier ces corrélations, nous calculons la matrice de covariance.

La matrice de covariance est une matrice symétrique  $p \times p$  (où  $p$  est le nombre de dimensions) qui a comme entrées les covariances associées à toutes les paires possibles des variables initiales (Maćkiewicz, A., & Ratajczak, W., 1993).

Que nous disent les covariances que nous avons comme entrées de la matrice sur les corrélations entre les variables ? C'est en fait le signe de la covariance qui compte :

- Si positif alors : les deux variables augmentent ou diminuent ensemble (corrélées)
- Si négatif alors : l'un augmente quand l'autre diminue (Inversement corrélé)

Maintenant que nous savons que la matrice de covariance n'est rien de plus qu'un tableau qui résume les corrélations entre toutes les paires de variables possibles, passons à l'étape suivante.

➤ **Étape 3 : Calculer les vecteurs propres et les valeurs propres de la matrice de covariance pour identifier les principales composantes**

Les vecteurs propres et les valeurs propres sont les concepts d'algèbre linéaire que nous devons calculer à partir de la matrice de covariance afin de déterminer les principales composantes des données. Avant d'aborder l'explication de ces concepts, commençons par comprendre ce que nous entendons par composants principaux.

Les composantes principales sont de nouvelles variables qui sont construites comme des combinaisons linéaires ou des mélanges des variables initiales. Ces combinaisons sont faites de manière à ce que les nouvelles variables (c'est-à-dire les composantes principales) ne soient pas corrélées et que la plupart des informations contenues dans les variables initiales soient comprimées ou compressées dans les premières composantes (Holland, S. M., 2019).

Ainsi, l'idée est que les données à 10 dimensions vous donnent 10 composants principaux, mais ACP essaie de mettre le maximum d'informations possibles dans le premier composant, puis le maximum d'informations restantes dans le second.

Une chose importante à réaliser ici est que les composantes principales sont moins interprétables et n'ont pas de sens réel puisqu'elles sont construites comme des combinaisons linéaires des variables initiales.

➤ **Étape 4 : Vecteur de fonctionnalités**

Comme nous l'avons vu à l'étape précédente, le calcul des vecteurs propres et leur classement par leurs valeurs propres par ordre décroissant, permettent de retrouver les composantes principales par ordre de signification.

Dans cette étape, ce que nous faisons est de choisir de conserver toutes ces composantes ou de rejeter celles de moindre importance (de faibles valeurs propres), et de former avec celles qui restent une matrice de vecteurs que nous appelons vecteur de caractéristiques. Ainsi, le vecteur caractéristique est simplement une matrice qui a pour colonnes les vecteurs propres des composantes que nous décidons de conserver (Shlens, J., 2014). Cela en fait la première étape vers la réduction de la dimensionnalité, car si nous choisissons de ne conserver que  $p$  vecteurs propres (composants) sur  $n$ , l'ensemble de données final n'aura que  $p$  dimensions.

### ➤ **Étape 5 : Recadrer les données selon les axes des composantes principales**

Dans les étapes précédentes, à part la normalisation, vous n'apportez aucune modification aux données, vous sélectionnez simplement les composants principaux et formez le vecteur de caractéristiques, mais l'ensemble de données d'entrée reste toujours en termes d'axes d'origine (c'est-à-dire en termes de les variables initiales).

Dans cette étape, qui est la dernière, le but est d'utiliser le vecteur caractéristique formé à l'aide des vecteurs propres de la matrice de covariance, pour réorienter les données des axes d'origine vers ceux représentés par les composantes principales (d'où le nom d'analyse en composantes principales).

## **3. Résultats de l'Analyse en Composantes Principales**

L'ACP fournit des résultats relatifs aux variables et aux observations.

**1/ Statistiques descriptives** : le tableau de statistiques descriptives présente pour toutes les variables sélectionnées des statistiques simples. Sont affichés le nombre d'observations, le nombre de données manquantes, le nombre de données non manquantes, la moyenne, et l'écart-type (non biaisé).

**2/ Matrice de corrélation/de covariance** : ce tableau correspond aux données qui sont ensuite utilisées pour les calculs. Le type de corrélation dépend de l'option qui a été choisie dans l'onglet « Général » de la boîte de dialogue. Dans le cas de corrélations, les corrélations significatives sont affichées en gras.

**3/ Test de sphéricité de Bartlett** : les résultats du test de sphéricité de Bartlett sont affichés. Ils permettent de valider ou d'infirmer l'hypothèse selon laquelle les variables ne sont pas significativement corrélées.

**4/ Mesure de précision de l'échantillonnage de Kaiser-Meyer-Olkin** : ce tableau donne pour chaque variable la valeur de la mesure KMO ainsi que le KMO global. L'indice KMO varie entre 0 et 1. Une valeur faible correspond au cas où il n'est pas possible d'extraire de facteurs synthétiques (ou variables latentes). Autrement dit, les individus ne permettent pas de faire ressortir le modèle que l'on pouvait imaginé préalablement (l'échantillon est « inadéquat »).

5/ **Valeurs propres** : les valeurs propres et le graphique correspondant sont affichés. Le nombre de valeurs propres est égal au nombre de valeurs propres non nulles.

6/ **Contributions** : les contributions sont une aide à l'interprétation. Les variables ayant le plus influencé la construction des axes sont celles dont les contributions sont les plus élevées.

7/ **Cosinus carrés** : comme pour les autres méthodes factorielles, l'analyse des cosinus carrés permet d'éviter des erreurs d'interprétation dues à des effets de projection. Si les cosinus carrés associés aux axes utilisés sur un graphique sont faibles, on évitera d'interpréter la position de l'observation ou de la variable en question.

8/ **Les coordonnées des observations** dans le nouvel espace sont ensuite affichées. Si des données supplémentaires ont été sélectionnées, elles sont affichées en fin de tableau.

9/ **Contributions** : ce tableau fournit les contributions des observations à la construction des composantes principales.

10/ **Cosinus carrés** : dans ce tableau sont affichés les cosinus carrés entre les vecteurs observations et les axes factoriels.

Dans le cas où une rotation a été demandée, les résultats de la rotation sont affichés, avec en premier la **matrice de rotation** appliquée aux coordonnées des variables. Suivent ensuite les pourcentages modifiés de variabilité associés à chacun des axes concernés par la rotation.

L'Analyse en Composantes Principales (ACP) est une technique utilisée pour identifier un plus petit nombre de variables non corrélées appelées composantes principales à partir d'un plus grand ensemble de données. La technique est largement utilisée pour mettre l'accent sur la variation et capturer des modèles forts dans un ensemble de données. Inventée par Karl Pearson en 1901, l'analyse en composantes principales est un outil utilisé dans les modèles prédictifs et l'analyse exploratoire des données. ACP est une méthode de la famille de l'analyse des données et plus généralement de la statistique multi variée, qui consiste à transformer des variables liées entre elles (dites « corrélées » en statistique) en nouvelles variables dé-corrélées les unes des autres. Ces nouvelles variables sont nommées « composantes principales » ou axes principaux. Elle permet au statisticien de résumer l'information en réduisant le nombre de variables.

*Cours « 3 »*

*L'Analyse*

*Factorielle des  
Correspondances*

## **1. Qu'est-ce que l'Analyse Factorielle des Correspondances ?**

L'Analyse Factorielle des Correspondances (ACF) est une méthode qui permet d'étudier l'association entre deux variables qualitatives. Cette méthode est basée sur l'inertie.

Le but de l'Analyse Factorielle des Correspondances consiste à représenter un maximum de l'inertie totale sur le premier axe factoriel, un maximum de l'inertie résiduelle sur le second axe, et ainsi de suite jusqu'à la dernière dimension.

Quatre approches de l'Analyse Factorielle des Correspondances sont proposées :

- Analyse Factorielle des Correspondances classique (AFC)
- Analyse Non Symétrique des Correspondances (ANSC)
- Analyse Factorielle des Correspondances utilisant la distance de Hellinger (HD)
- Analyse des Correspondances Détendancée (ACD)

## **2. Résultats de l'Analyse Factorielle des Correspondances**

**1/ Tableau de contingence** : le tableau de contingence est affiché. Le diagramme en bâtons en 3 dimensions en est la représentation graphique.

**2/ Inertie par case** : le tableau des inerties par case est affiché. La somme des inerties est égale à la statistique du  $\chi^2$  divisée par la fréquence totale (somme des cellules du tableau de contingence).

**3/ Test d'indépendance** entre les lignes et les colonnes : ce test permet de déterminer, sur la base de la statistique du  $\chi^2$ , si l'on doit rejeter l'hypothèse nulle selon laquelle les lignes et les colonnes du tableau sont indépendantes. Une interprétation détaillée est fournie automatiquement.

**4/ Valeurs propres et pourcentages d'inertie** : les valeurs propres et le graphique (scree plot) correspondant sont affichés. Seules les valeurs propres non triviales sont affichées. Si un filtrage a été demandé, il est appliqué aux résultats qui suivent. Remarque : la somme des valeurs propres affichées est égale à l'inertie totale. A chaque valeur propre correspond un axe principal représentant un pourcentage donné de l'inertie totale. On peut ainsi mesurer le pourcentage cumulé d'inertie totale correspondant à un nombre croissant de dimensions.

Une série de résultats est ensuite affichée, d'abord pour les points lignes, puis pour les points colonnes :

**1/ Poids, distances et distances quadratiques à l'origine, inerties et inerties relatives :** ce tableau contient des statistiques de base pour les points-lignes (puis les points-colonnes).

**2/ Profils :** dans ce tableau sont affichés les profils, ainsi que la moyenne des profils.

**3/ Distances du  $\chi^2$  :** dans ce tableau sont affichées les distances du  $\chi^2$  entre les profils.

**4/ Coordonnées principales :** dans ce tableau sont affichées les coordonnées principales. Ces coordonnées sont utilisées pour la création des graphiques symétriques (ou barycentriques) et asymétriques (ou pseudo-barycentriques).

**5/ Coordonnées standard :** ces coordonnées correspondent aux précédentes à un facteur près. Le facteur est la racine carrée de l'inverse de la valeur propre correspondante. Ces coordonnées sont utilisées pour la création des graphiques asymétriques (ou barycentriques).

**6/ Contributions :** les contributions sont une aide à l'interprétation. Les modalités ayant influencé le plus la construction des axes sont celles dont les contributions sont les plus élevées. On pourra se contenter d'interpréter les résultats des modalités pour lesquelles les contributions sont supérieures aux poids relatifs affichés dans la première colonne.

**7/ Cosinus carrés :** comme pour les autres méthodes factorielles, l'analyse des cosinus carrés permet d'éviter des erreurs d'interprétation dues à des effets de projection. Si les cosinus carrés associés aux axes utilisés sur un graphique sont faibles, on évitera d'interpréter la position du point-ligne ou du point-colonne correspondant.

L'analyse factorielle des correspondances (AFC) est une méthode statistique d'analyse des données qui permet d'analyser et de hiérarchiser les informations contenues dans un tableau rectangulaire de données et qui est aujourd'hui particulièrement utilisée pour étudier le lien entre deux variables qualitatives. Elle a été mise au point à partir des années 1960 par Jean-Paul Benzécri et son équipe, d'abord à la faculté des sciences de Rennes2, puis à celle de Jussieu à Paris au sein du laboratoire de statistique multidimensionnelle. Elle se rattache à la famille des analyses factorielles qui regroupe différentes méthodes d'analyses de grands tableaux rectangulaires de données, visant toutes à identifier et à hiérarchiser des facteurs corrélés aux données placées en colonnes.

*Cours « 4 »*

*Traitement  
des Données*

*Partie 1 :*

*Traitement des Données*

*« SPSS »*

## 1. Les logiciels de traitement des données

Les logiciels de traitement des données sont nombreux et, pour la plupart, très bien construits et très faciles à utiliser. On va citer les plus importants:

- **Le logiciel Excel** (Microsoft) : est le plus connu et le plus utilisé ; la version la plus récente contient une partie des procédures statistiques utilisées dans les analyses des données.
- **StatBox et Question** : mis au point par la firme Grimmer Logiciels, sont des logiciels conçus spécialement pour l'analyse des données d'enquête ; ces logiciels fonctionnent à partir du logiciel Excel de Microsoft.
- **Le Sphinx** : est un logiciel utilisé surtout pour la recherche marketing.
- **Minitab** : est un logiciel statistique puissant qui propose un grand nombre de procédures statistiques.
- **Le logiciel SAS** (*système d'analyse statistique*) a été conçu au départ pour le calcul économique et les modèles de régression ; par la suite, on l'a adapté de façon à y inclure les méthodes les plus connues de l'analyse des données.
- **Le logiciel SPSS** (*Statistical Package for the Social Sciences*) a été créé, au tout début, pour les besoins des psychologues. Avec le temps (cette entreprise existe depuis 1965), on a intégré un grand nombre de procédures statistiques tout en facilitant le travail de manipulation des données.

### 1. Qu'est-ce que SPSS ?

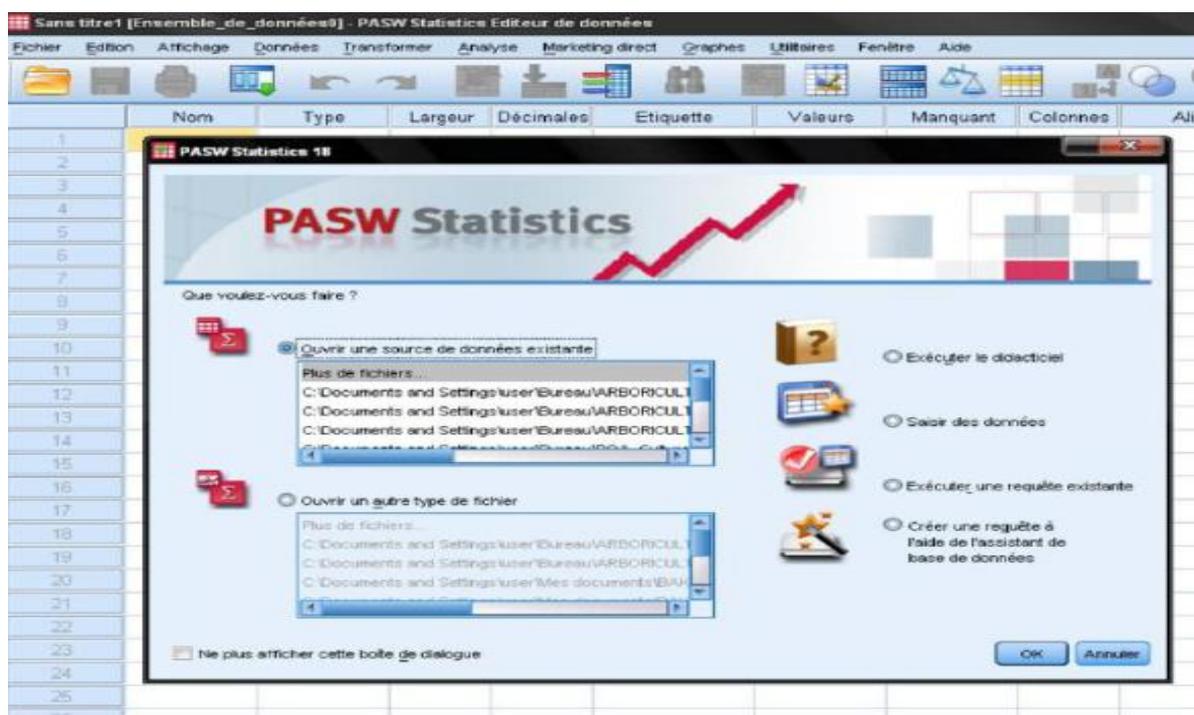
Le logiciel SPSS est un instrument particulièrement adapté à la mise en œuvre des techniques d'analyse de données statistiques. Il favorise la gestion des données dans un environnement graphique convivial combinant des menus descriptifs et des boîtes de dialogue. De plus, cet environnement offre un langage de commande qui permet d'écrire des programmes pour optimiser les tâches de production. Il permet de traiter efficacement les données et d'effectuer des analyses sur de grandes bases de données. Il offre plusieurs possibilités d'organisation et de synthèse des informations statistiques (Gilbert R., 2002).

## **2. Interface du logiciel SPSS**

### **3.1. La Fenêtre de Démarrage**

La fenêtre de démarrage se présente comme l'indique l'image ci-dessous. Cette fenêtre donne la liste des fichiers SPSS récemment ouverts. Pour ouvrir un fichier, il suffit de faire un double clic sur son nom (Gilbert R., 2002).

*Figure 08 : La Fenêtre de Démarrage*



### **3.2. Types de fenêtre dans SPSS**

Le logiciel SPSS est composé de trois fenêtres: la fenêtre des données (ou fenêtre de l'éditeur), celle des résultats (souvent appelée Viewer) et celle des syntaxes.

#### **➤ La fenêtre Éditeur de données**

Présente le contenu d'une base de données que vous avez préalablement sélectionné. La fenêtre Éditeur de données comprend deux onglets (Gilbert R., 2002) :

- L'onglet affichage des données :

Permet de voir la base de données, où les observations sont présentées en lignes et les variables sont en colonnes. Chaque cellule présente la modalité que prend une variable.

- L'onglet affichage des variables :

Permet de voir toutes les variables présentes dans la base de données, leurs noms, types, valeurs, mesures...

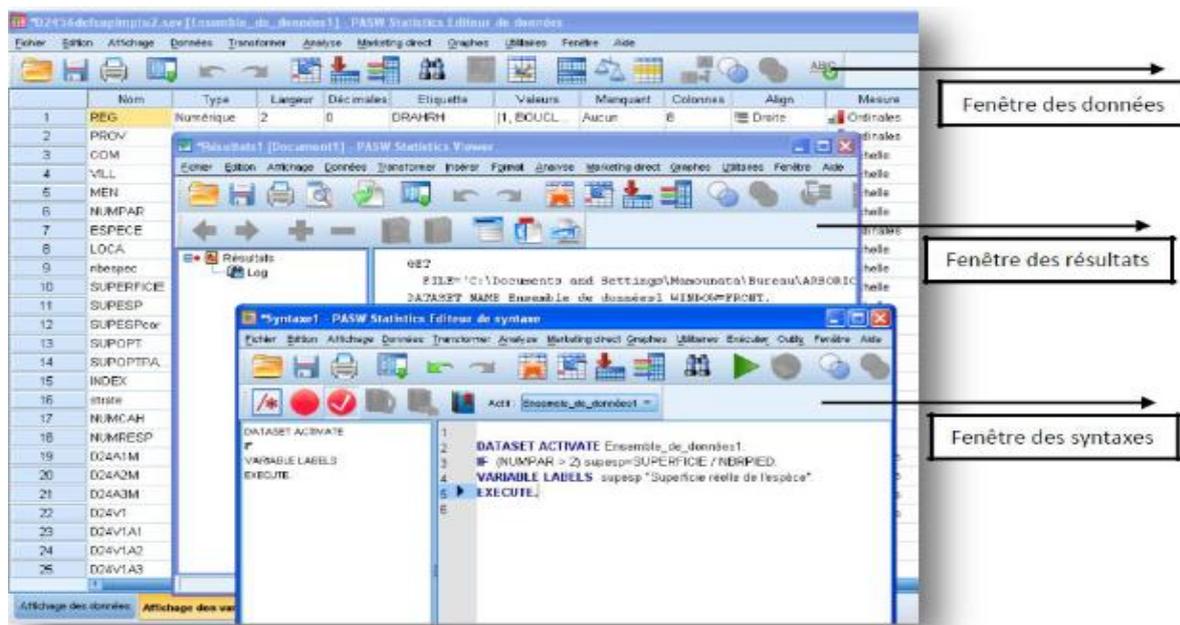
➤ **La fenêtre des résultats :**

Elle affiche les résultats des commandes exécutées par SPSS. C'est dans cette fenêtre que nous observerons les sortis de tableaux, de graphiques, etc. Cette fenêtre s'ouvre automatiquement dès que vous lancez une procédure qui doit afficher des résultats.

➤ **La fenêtre des syntaxes :**

Elle permet de saisir les commandes SPSS. En effet, SPSS présente une configuration Windows avec menu déroulant (on peut donc travailler avec la souris) et une possibilité de saisir les programmes en langage SPSS. SPSS permet en effet d'obtenir la syntaxe de toute opération avec le bouton coller.

*Figure 09 : Les 3 Fenêtres SPSS*



### 3.3. Comment organiser les données dans SPSS ?

Les données doivent être organisées de la façon suivante :

- *Chaque ligne* représente un participant; si le participant a été observé plusieurs fois, on doit retrouver plusieurs colonnes, une par observation.
- *Chaque colonne* représente une variable, soit:
  - La ou les variables indépendantes identifiant la condition expérimentale, le numéro de sujets, et possiblement des informations sur le sujet (tel le sexe, l'âge, etc.) selon les besoins;
  - La (ou les) variable dépendante, c'est à dire le résultat de la mesure.
- L'ordre des colonnes n'est pas important.
- Vous pouvez donner les noms que vous voulez à vos colonnes. Utiliser des noms significatifs (tel sexe, âge, etc.). Par contre, SPSS ne permet pas l'utilisation de symboles spéciaux dans le nom d'une colonne (le point, la virgule, le dièse, etc.), ne permet pas d'utiliser plus que huit lettres, et finalement, un nom de colonne doit commencer par une lettre.

### 3.4. Manipulations des données d'un fichier

#### 3.4.1 Insertion des Variables

➤ **Caractéristiques d'une variable :**

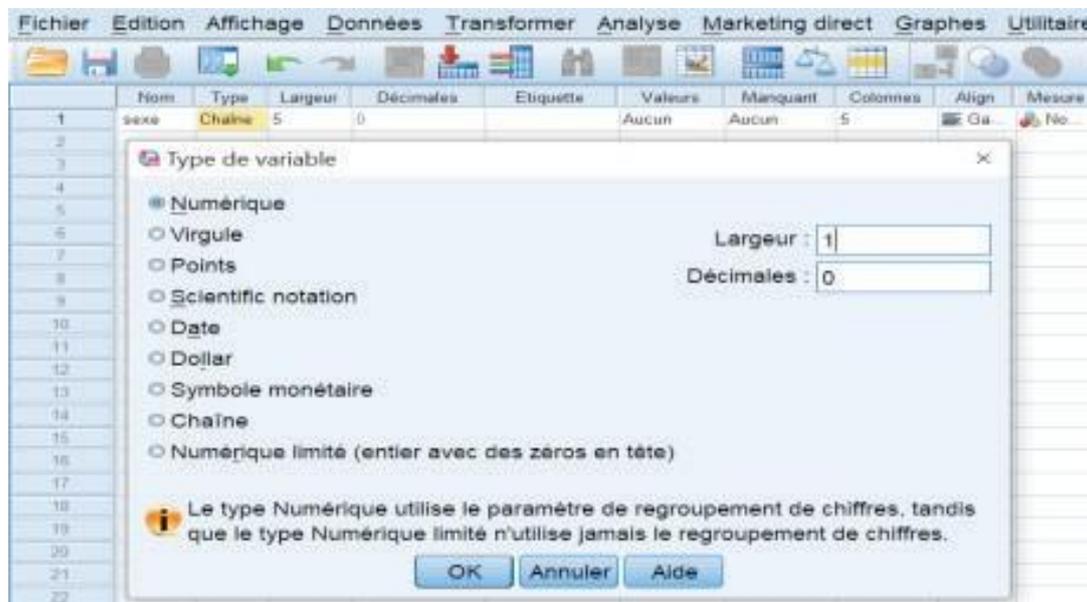
|   | Nom  | Type      | Largeur | Décimales | Etiquette | Valeurs      | Manquant | Colonnes | Align  | Mesure    | Rôle   |
|---|------|-----------|---------|-----------|-----------|--------------|----------|----------|--------|-----------|--------|
| 1 | REG  | Numérique | 2       | 0         | DRAHRH    | {1, BOUCL... | Aucun    | 8        | Droite | Ordinales | Entrée |
| 2 | PROV | Numérique | 2       | 0         | Province  | {1, BAM}...  | Aucun    | 8        | Droite | Ordinales | Entrée |

**Nom :** Taper directement dans la cellule le nom de votre variable. Le nom doit être unique, peut contenir jusqu'à 64 caractères et il ne doit pas contenir d'espaces.

**Type, largeur, décimale :** Par défaut, les variables sont codées par des chiffres: **numériques**

- Cette boîte de dialogue vous permet de préciser le type de variable, sa largeur et le nombre de décimales.

- Pour définir la variable, cliquez sur l'un des *types*, ensuite exécutez les autres selon les choix désirés.



- *Largeur de la variable* détermine le nombre de chiffres que comportent les données.

- *Les décimales* correspondent aux nombres de chiffres après la virgule.

**Etiquette :** L'étiquette va servir à donner plus d'explication au nom de la variable.

- L'étiquette peut contenir jusqu'à 256 caractères.

- L'étiquette peut contenir des espaces et des caractères réservés qui ne sont pas autorisés dans les noms de variables.

**Valeur:** Pour chaque valeur (le codage) de la variable, vous pouvez y associer une description (la modalité)

- La description peut contenir jusqu'à 60 caractères.

- Le premier rectangle contient la valeur numérique de l'une des modalités de la variable.
- Le deuxième rectangle présente l'étiquette de cette modalité.

**Valeurs manquantes :** L'absence d'une donnée (case vide) est considérée systématiquement comme une valeur manquante.

**Colonnes alignement :**

- *Colonnes* : Correspondent à la largeur de la colonne.
- *Alignement* : Permet d'aligner les données selon trois positions à l'intérieur de la cellule : droite, gauche, et centrée. Une donnée numérique est alignée à droite par défaut.

**Mesure :** Il y a trois types de mesure :

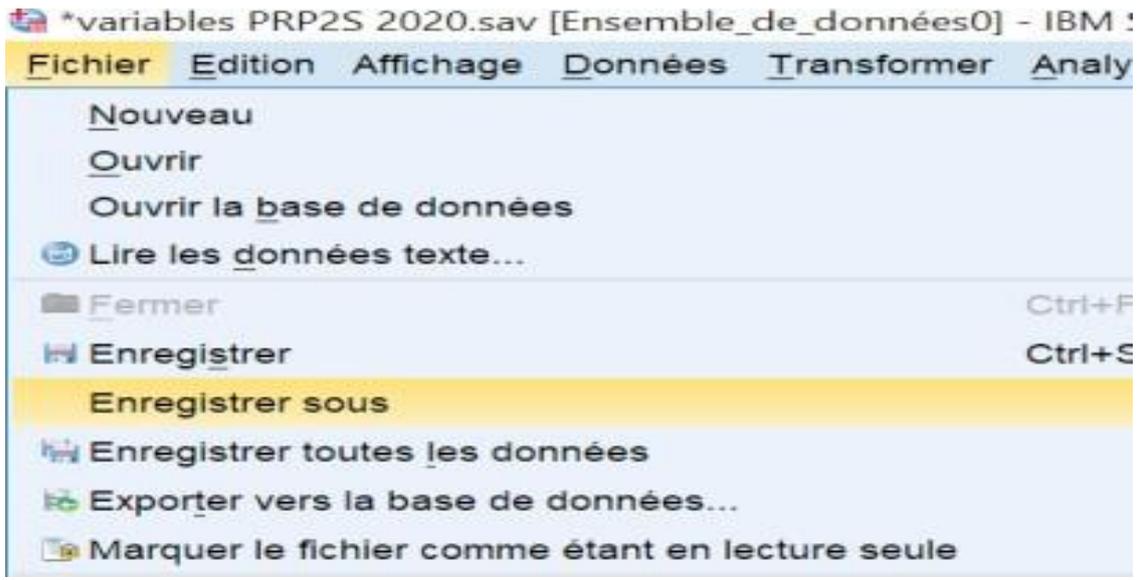
- *Echelle* : Pour les variables numériques discrètes ou continues.
- *Ordinales* : Pour les variables qualitatives ordinales.
- *Nominales* : Pour les variables qualitatives nominales.

### **3.4.2 Insertion des Données**

L'insertion des données se fait à partir de la fenêtre servant à cette opération (affichage de données) et en fonction de la définition des variables et, à partir des réponses aux questionnaires.

### **3.4.3 Sauvegarder la base de données**

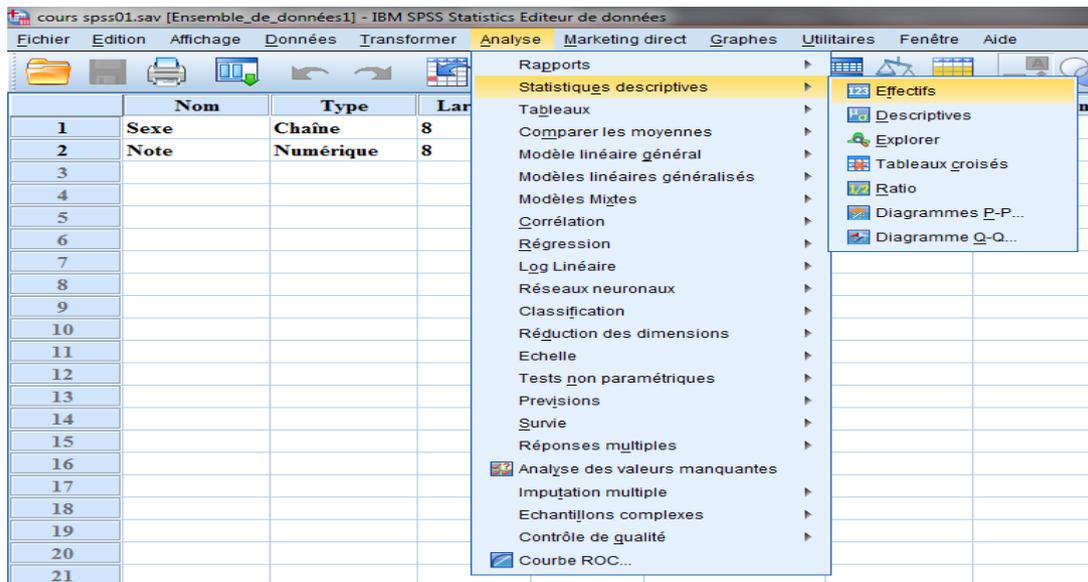
Une des opérations à ne pas l'oublier, une fois que votre base de données est construite, c'est sans aucun doute la sauvegarde. Si vous ne sauvegardez pas, vous risquez de perdre votre travail.



#### 4. L'analyse des données

L'analyse la plus courante est la procédure *effective*. Elle sert à analyser et afficher les différents paramètres statistiques. A partir des menus dans l'éditeur de données sélectionnez:

- Analyse
- Statistiques descriptives
- Effectifs.



### ➤ **Fréquences, Moyennes et Médianes**

Vous pouvez également obtenir des informations sur les mesures de tendance centrale et de dispersion présentées dans le rappel théorique de la procédure Descriptive.

#### **Tendance centrale :**

- *Moyenne* : moyenne des observations
- *Médiane* : valeur sous laquelle on trouve 50 % de la distribution et au-dessus de laquelle on trouve l'autre 50 %. Cette mesure fait du sens uniquement lorsque la variable est ordinale ou continue.
- *Mode* : valeur qui revient le plus souvent dans la distribution
- *Somme* : somme de toutes les observations

#### **Dispersion :**

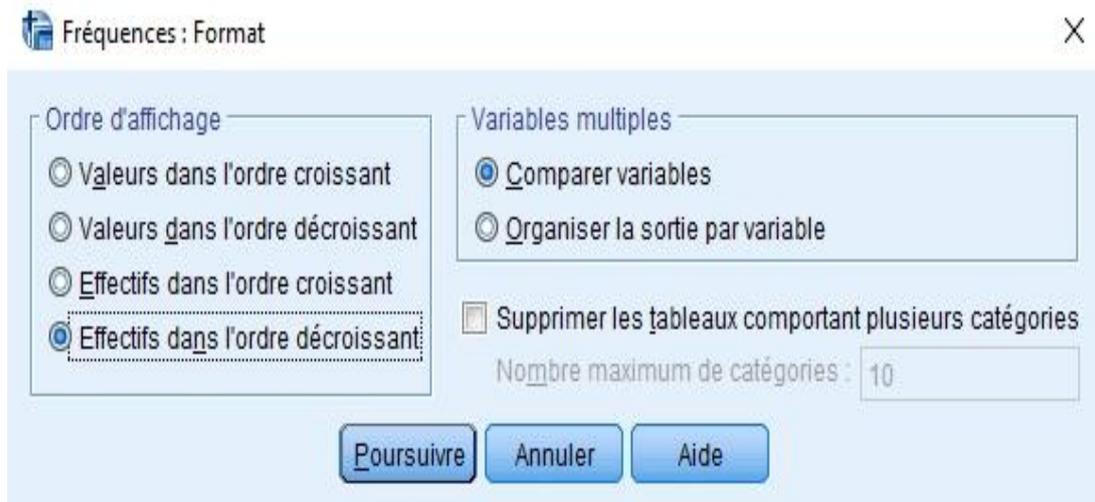
- *Écart-type* : écart-type de la distribution
- *Variance* : variance de la distribution
- *Étendue* : indique l'écart entre la plus petite et la plus grande valeur
- *Minimum* : plus petite valeur de la variable
- *Maximum* : plus grande valeur de la variable
- *E. S. moyenne* : erreur-type (standard) de la moyenne

### ➤ **Bouton Graphiques: Représentez graphiquement la distribution des observations**

Vous permet de réaliser des graphiques sans passer par le menu *Graphes*. Vous pouvez même choisir le graphique qui représentera le mieux vos données: *diagramme en bâtons*, *diagramme en secteurs* ou *histogramme*.

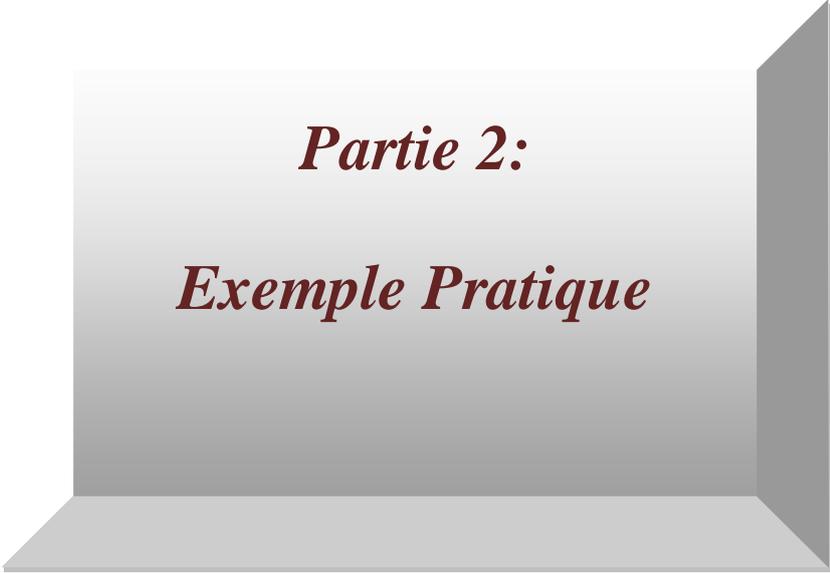
### ➤ **Bouton Format : Classification des tables de fréquences**

Le principal intérêt de cette option est de trier la table de fréquences de manière ascendante ou descendante en fonction des valeurs possibles de la variable, ou encore l'effectif. On trouve les choix d'ordre d'affichage dans la partie gauche de la boîte de dialogue ouverte en cliquant sur le bouton *Poursuivre*.



- *Tri par les valeurs possibles de la variable* : Cette option classera la table par valeur croissante ou décroissante des valeurs possibles de la variable choisie.

- *Tri par l'effectif* : Lorsque vous choisissez le tri croissant, la table de fréquences débutera par la valeur ayant le plus petit effectif et se terminera par la valeur ayant le plus grand effectif d'apparition. Pour obtenir l'inverse, vous sélectionnez le tri décroissant.



*Partie 2:*

*Exemple Pratique*

➤ **Etape 1 :**

D'abord, il faut aller à « *Affichage des Données* ».

1. On saisie les modalités de la variable « Sexe » et les valeurs de la variable « Note » de 01 à 12.
2. Introduire les caractéristiques de cette variable qui ressemblent à celles de « Note ».

|    | Sexe | Note  | var |
|----|------|-------|-----|-----|-----|-----|-----|-----|-----|
| 1  | 1.00 | 10.00 |     |     |     |     |     |     |     |
| 2  | 2.00 | 8.00  |     |     |     |     |     |     |     |
| 3  | 2.00 | 12.00 |     |     |     |     |     |     |     |
| 4  | 1.00 | 13.00 |     |     |     |     |     |     |     |
| 5  | 2.00 | 15.00 |     |     |     |     |     |     |     |
| 6  | 1.00 | 18.00 |     |     |     |     |     |     |     |
| 7  | 2.00 | 9.00  |     |     |     |     |     |     |     |
| 8  | 2.00 | 15.00 |     |     |     |     |     |     |     |
| 9  | 2.00 | 7.00  |     |     |     |     |     |     |     |
| 10 | 1.00 | 13.00 |     |     |     |     |     |     |     |
| 11 | 1.00 | 9.00  |     |     |     |     |     |     |     |
| 12 | 1.00 | 16.00 |     |     |     |     |     |     |     |

3. Dans « *Afficher les variables* », on aura :

|   | Nom  | Type      | Largeur | Décimales | Etiquette          | Valeurs        | Manquant | Colonnes | Align    | Mesure      |
|---|------|-----------|---------|-----------|--------------------|----------------|----------|----------|----------|-------------|
| 1 | Sexe | Chaîne    | 8       | 0         | Sexe               | {1.00, Masc... | Aucun    | 8        | ☰ Gauche | 🔴 Nominales |
| 2 | Note | Numérique | 4       | 2         | Note d'examen /... | Aucun          | Aucun    | 8        | ☰ Droite | 🟡 Echelle   |
| 3 |      |           |         |           |                    |                |          |          |          |             |
| 4 |      |           |         |           |                    |                |          |          |          |             |

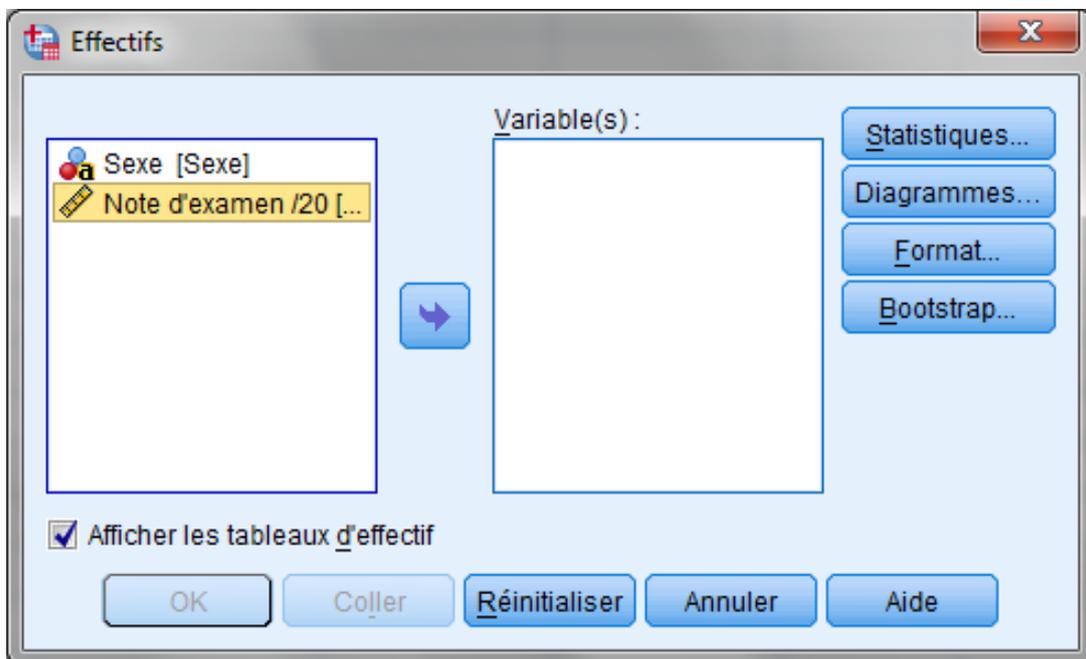
➤ **Etape 2 :**

L'analyse la plus courante est la procédure effective, A partir des menus dans **l'éditeur de données** sélectionnez :

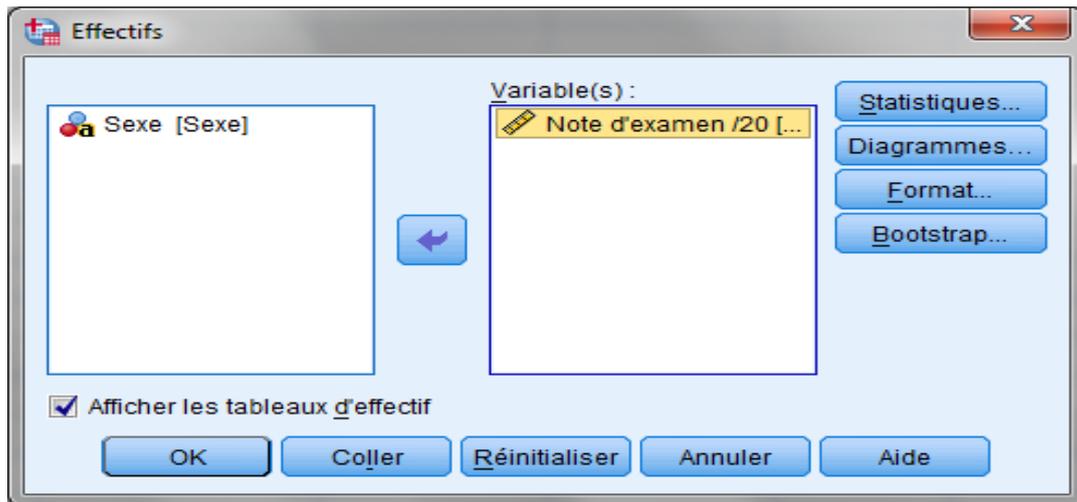
1. Analyse
2. Statistiques descriptives
3. Effectifs.

Dans la boîte de dialogue **Effectifs** :

1. Sélectionnez la variable « **Note d'examen** ».
2. Transférez-la dans le champ variable (s).

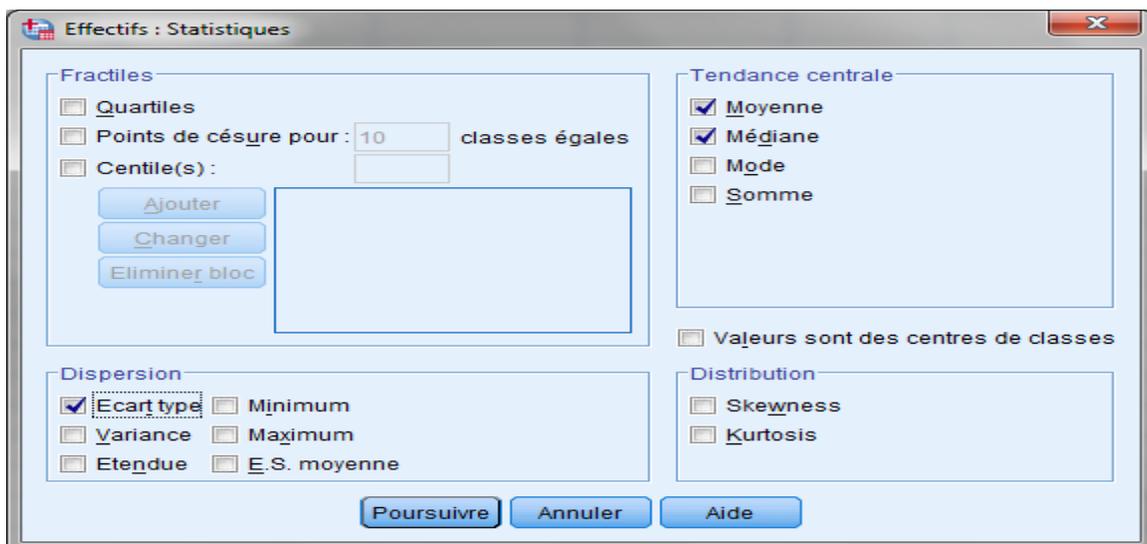


3. Puis appuyez sur Statistiques.



➤ **Etape 3 :**

1. Sélectionnez dans l'encadré « *Tendance centrale* » la **Moyenne et la Médiane**.



2. Sélectionnez dans l'encadré « *Dispersion* » l'écart-type.

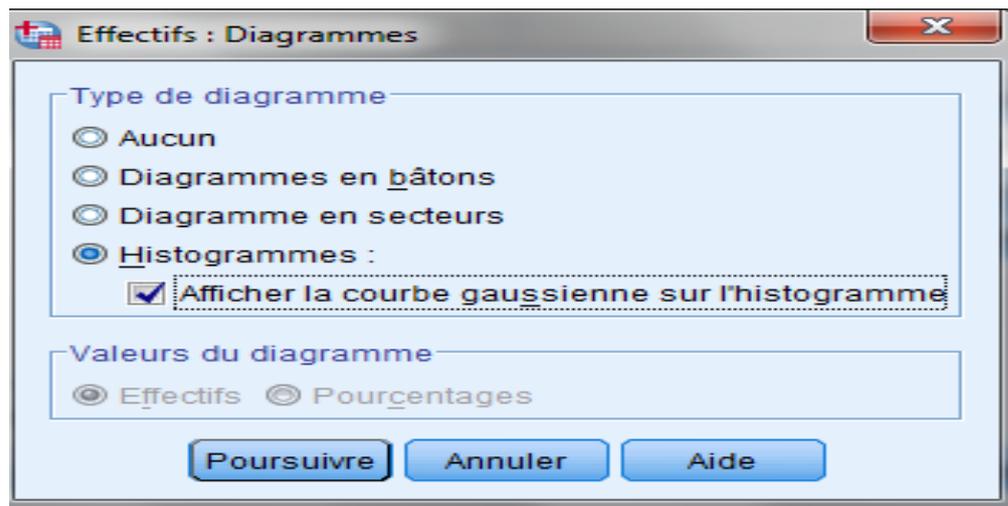
3. Cliquez sur « Poursuivre ».

➤ **Etape 4 :**

1. Sélectionnez « *Diagramme* »

2. Dans l'encadré « Type de diagramme »

3. Sélectionnez « Histogrammes »



4. Sélectionnez « Afficher la courbe gaussienne »
5. Cliquez « Poursuivre »
6. En fin cliquez « Ok »

➤ **Etape 5 :**

**Fenêtre des Résultats :** C'est une fenêtre qui indique les résultats.

Cette fenêtre, est une analyse descriptive, représente la moyenne, la médiane et l'écart-type.

| <b>Effectifs</b>   |            |         |
|--|------------|---------|
| [Ensemble_de_données1] C:\Users\DCT\Documents\cours spss01.sav |            |         |
| <b>Statistiques</b>  |            |         |
| Note d'examen /20  |            |         |
| N  | Valide     | 12      |
|  | Manquante  | 0       |
|  | Moyenne    | 12.0833 |
|  | Médiane    | 12.5000 |
|  | Ecart-type | 3.50216 |

**Note d'examen /20**

|        |              | Effectifs | Pourcentage | Pourcentage valide | Pourcentage cumulé |
|--------|--------------|-----------|-------------|--------------------|--------------------|
| Valide | 7.00         | 1         | 8.3         | 8.3                | 8.3                |
|        | 8.00         | 1         | 8.3         | 8.3                | 16.7               |
|        | 9.00         | 2         | 16.7        | 16.7               | 33.3               |
|        | <u>10.00</u> | 1         | 8.3         | 8.3                | 41.7               |
|        | 12.00        | 1         | 8.3         | 8.3                | 50.0               |
|        | <u>13.00</u> | 2         | 16.7        | 16.7               | 66.7               |
|        | <u>15.00</u> | 2         | 16.7        | 16.7               | 83.3               |
|        | 16.00        | 1         | 8.3         | 8.3                | 91.7               |
|        | 18.00        | 1         | 8.3         | 8.3                | 100.0              |
|        | Total        | 12        | 100.0       | 100.0              |                    |

Les colonnes de ce tableau représentent :

**Colonne 1:** les observations non-répétitives.

**Colonne 2 :** l'effective, les notes 9, 13,15 sont répétées 2 fois, par contre le reste une seule fois.

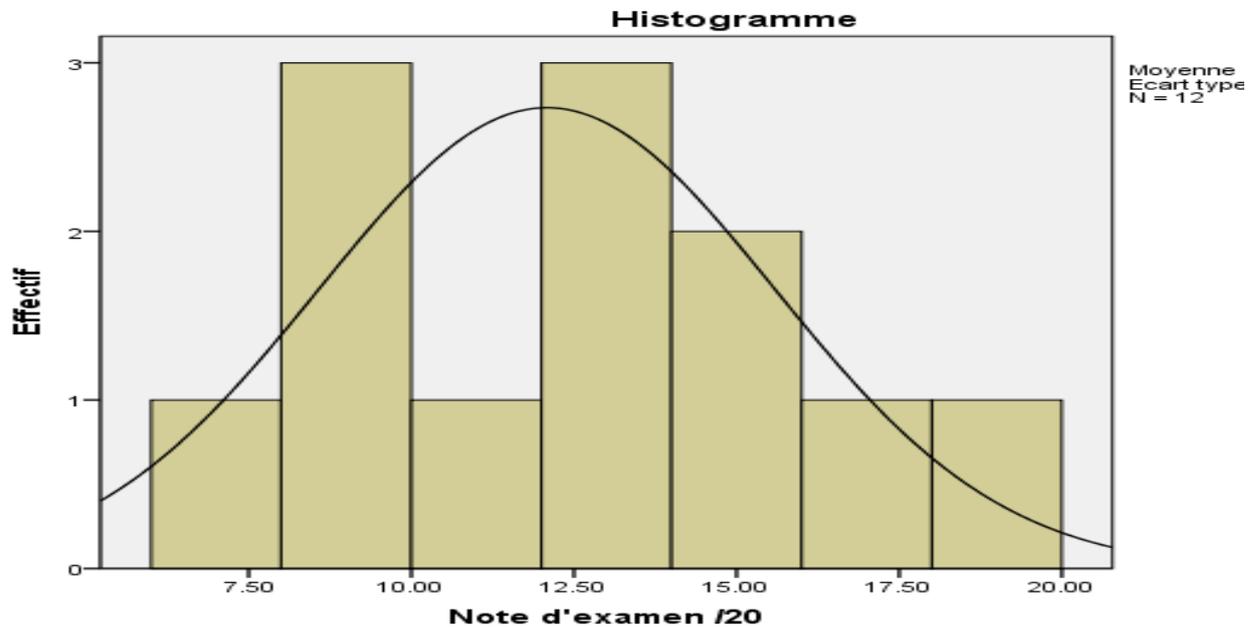
**Colonne 3:** le pourcentage de chaque observation par rapport à (100 %).

La note 7 est de 7%, la note 13 est 15,7%.

**Colonne 5:** pourcentage cumulé :

- $8,3+8,3$  (pourcentage valide)=16,7
- $16,7+16,7$  (pourcentage valide) =33,33
- $33,33+ 8,3$  (pourcentage valide) = 41,7
- etc.....

➤ Etape 6 :



L'**histogramme** montre le pourcentage des diverses observations. La note en abscisse et le pourcentage correspondant en ordonnée.

Le graphique montre aussi la courbe de Gauss. Les observations sont presque normalement distribuées.

➤ Etape 7 :

Après, faisons la procédure **Effective** de la variable « **Sexe** »

1. Transfert de la variable « Sexe » vers le champ « Variables ».
2. Appuyer sur « OK ».



➤ **Etape 8 :**

| Statistiques |           |    |
|--------------|-----------|----|
| Sexe         |           |    |
| N            | Valide    | 12 |
|              | Manquante | 0  |

| Sexe   |      |           |             |                    |                    |
|--------|------|-----------|-------------|--------------------|--------------------|
|        |      | Effectifs | Pourcentage | Pourcentage valide | Pourcentage cumulé |
| Valide | 1.00 | 6         | 50.0        | 50.0               | 50.0               |
|        | 2.00 | 6         | 50.0        | 50.0               | 100.0              |
| Total  |      | 12        | 100.0       | 100.0              |                    |

Comme on peut le voir :

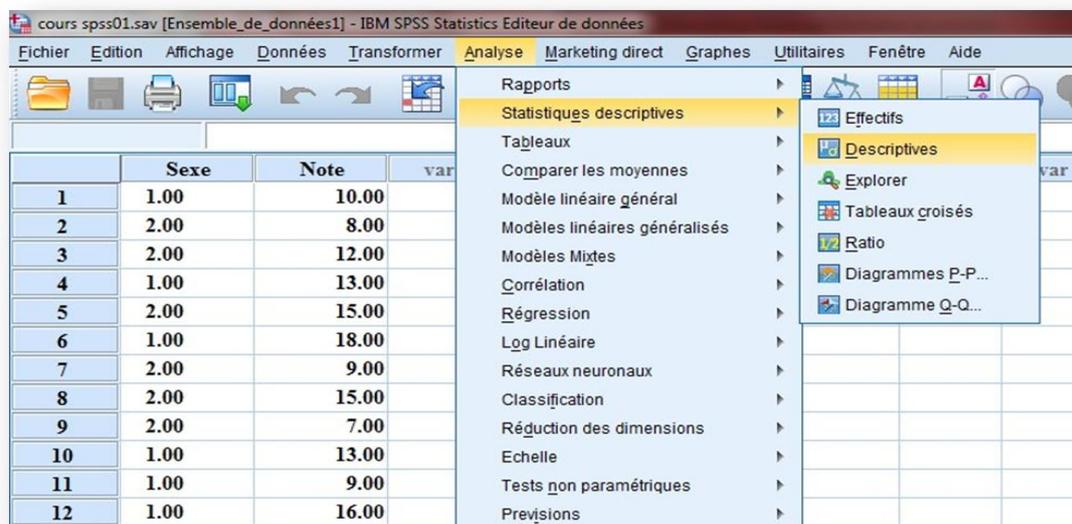
1. L'**effectif** des élèves est de 12.
2. Le **pourcentage** 50% féminin (2) et 50% masculin (1).

➤ **Etape 9 :**

Maintenant procédant à la procédure descriptive.

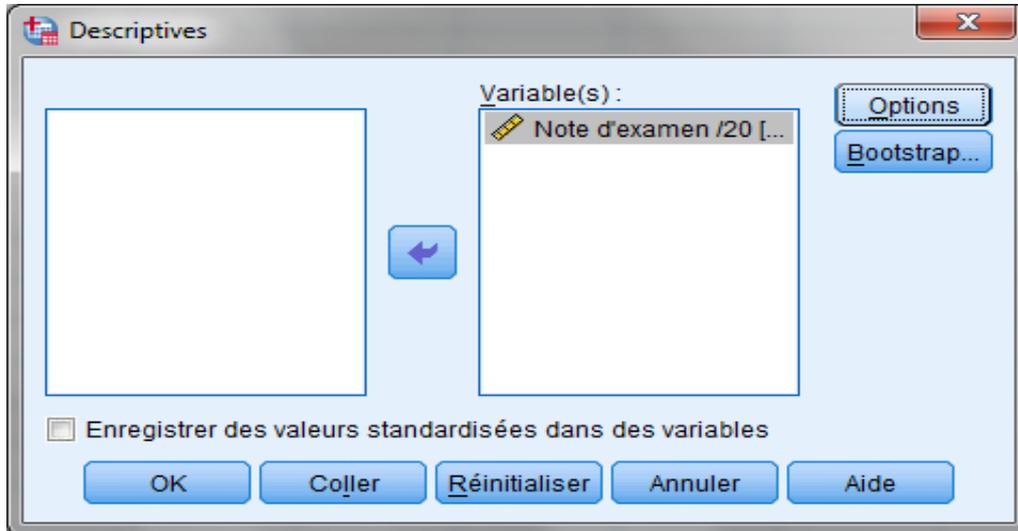
A partir de la barre de menu Cliquez sur:

2. Analyse
2. Statistiques descriptives
2. Descriptives

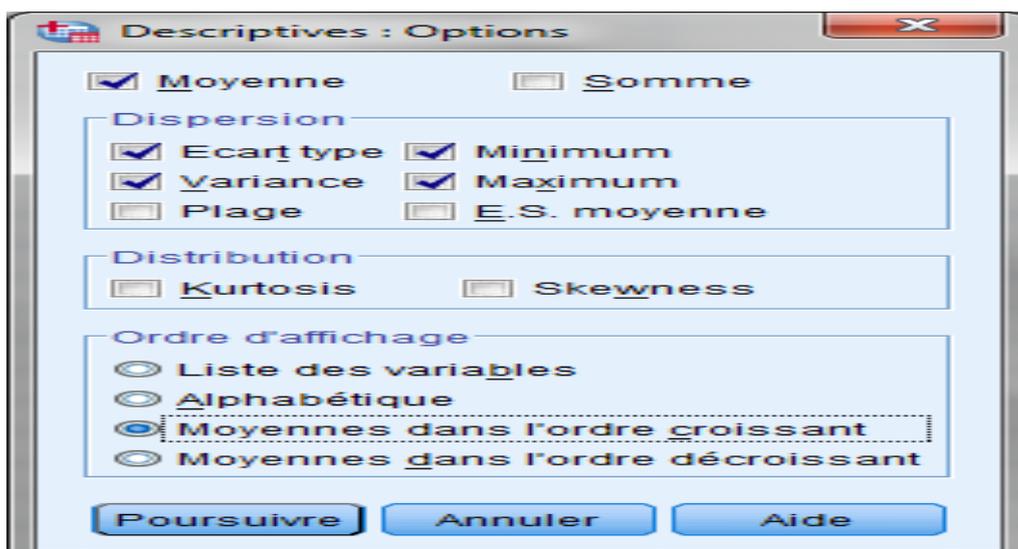


➤ **Etape 10 :**

1. Transférer la variable « Note d'examen » vers le champ « Variables » puis cliquez sur « Option ».



2. Cochez « Moyenne ».
3. **Dispersion** : Cochez : *Ecart-type, Variance, Minimum, Maximum*
4. **Ordre d'affichage** : Cochez : *Moyenne dans l'ordre croissant*
5. Sectionnez « Poursuivre ».
6. Enfin OK

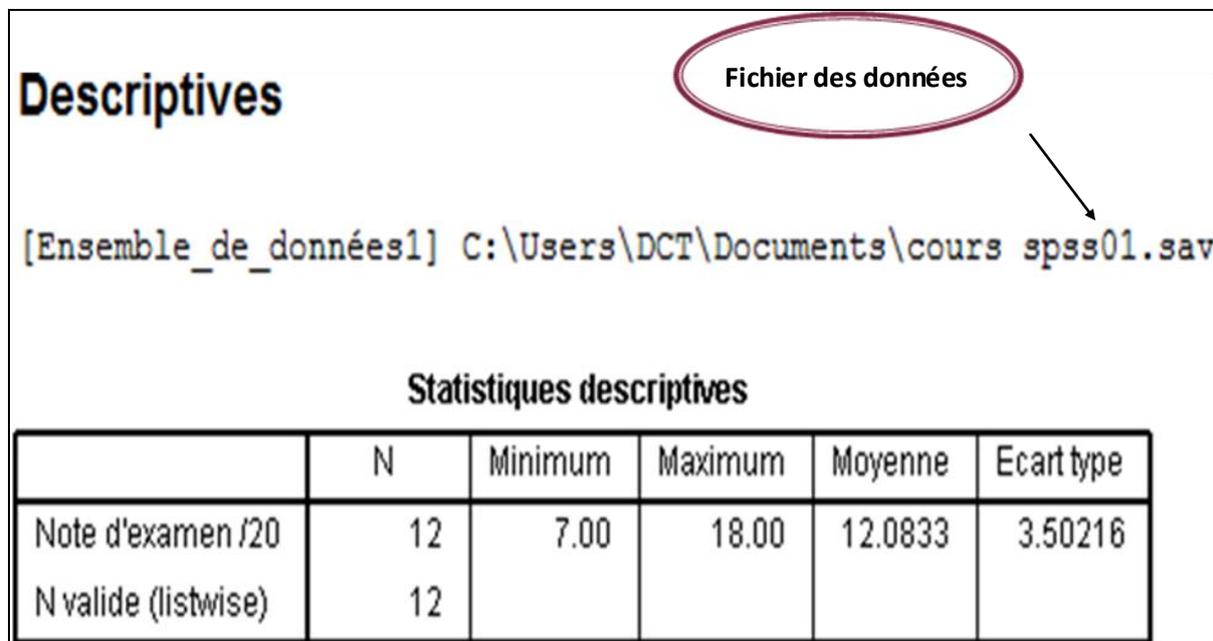


➤ **Etape 11 :**

Lire les Résultats : Ici, c'est la *Fenêtre des Résultats*, l'effective n'est pas indiqué sur le tableau.

Sur ce dernier, on peut voir :

- N le nombre de Notes (N=12).
- Le *Minimum* (7,00).
- Le *Maximum* (18,00).
- La *Moyenne* des Notes 12,08/20.
- L'*Ecart-type* (3,50).



The screenshot shows the SPSS 'Descriptives' window. At the top, the title 'Descriptives' is on the left, and 'Fichier des données' is circled in red on the right. Below this, the data source is listed as '[Ensemble\_de\_données1] C:\Users\DCT\Documents\cours spss01.sav'. The main content is a table titled 'Statistiques descriptives' with the following data:

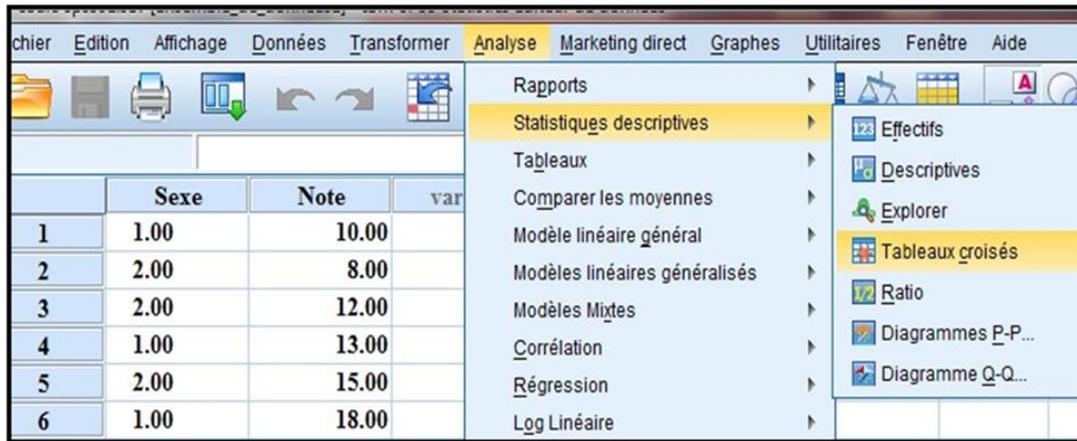
|                     | N  | Minimum | Maximum | Moyenne | Ecart type |
|---------------------|----|---------|---------|---------|------------|
| Note d'examen /20   | 12 | 7.00    | 18.00   | 12.0833 | 3.50216    |
| N valide (listwise) | 12 |         |         |         |            |

➤ **Etape 12 :**

• **Croisement des Variables**

Le croisement des variables est une autre opération d'analyse des données. Dans notre exemple, il s'agit de croisement entre la variable « *Sexe* » et la variable « *Note d'examen* » :

1. Analyse
2. Statistiques descriptives
3. Tableaux croisés



- **La Fenêtre des Résultats / Viewer**

**Tableau croisé Sexe \* Note d'examen /20**

Effectif

|       |      | Note d'examen /20 |      |      |       |       |       |       |       | Total |       |
|-------|------|-------------------|------|------|-------|-------|-------|-------|-------|-------|-------|
|       |      | 7.00              | 8.00 | 9.00 | 10.00 | 12.00 | 13.00 | 15.00 | 16.00 |       | 18.00 |
| Sexe  | 1.00 | 0                 | 0    | 1    | 1     | 0     | 2     | 0     | 1     | 1     | 6     |
|       | 2.00 | 1                 | 1    | 1    | 0     | 1     | 0     | 2     | 0     | 0     | 6     |
| Total |      | 1                 | 1    | 2    | 1     | 1     | 2     | 2     | 1     | 1     | 12    |

1. La lecture de la table ci-dessus, *Tableau croisé Sexe \* Note d'examen*, on peut voir le nombre de chaque note pour chaque sexe.

**Exemple:**

La note 9 est saisie une fois pour les 2 sexes, par contre la note 13 est saisie 2 fois pour le sexe Masculin et 0 fois pour le sexe Féminin, la note 10 est saisie 0 fois pour Féminin et 1 fois pour Masculin etc.

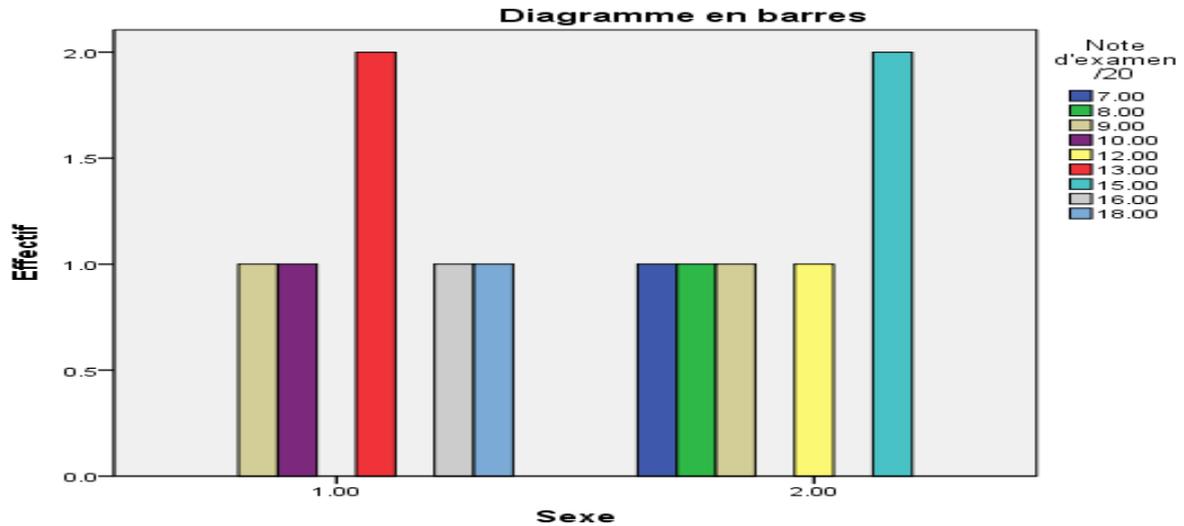
2. La dernière ligne (*Total*), on donne le nombre de chaque note pour les 2 sexes.

**Exemple:**

Pour la note 9, le nombre total est de 2.

- **Diagramme en Barre**

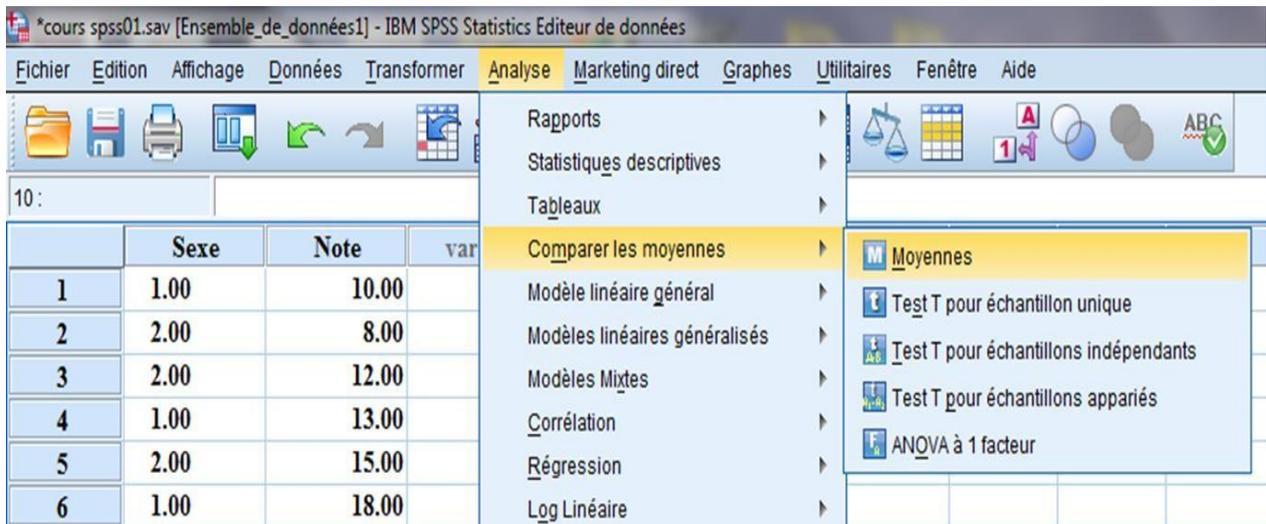
Ce diagramme traduit le contenu de la table précédente. On constate la présence de 2 barres très hautes qui représentent les notes 13,00 et 15,00. Rouge 13,00, effectif 2 pour le sexe masculin et turquoise 15,00, effectif 2 pour le sexe féminin.



➤ **Etape 13 :**

On veut maintenant comparer, La **moyenne – L'écart type** de chaque variable.

1. Analyse.
2. Comparer les moyennes.
3. Moyennes.



- Fenêtre des Résultats / Viewer

Dans cette fenêtre on aura les tableaux suivants :

→ **Moyennes**

[Ensemble\_de\_données1] C:\Users\DCT\Documents\cours spss01.sav

**Observation Calculer Récapituler**

|                     | Observations |             |          |             |       |             |
|---------------------|--------------|-------------|----------|-------------|-------|-------------|
|                     | Inclus       |             | Exclu(s) |             | Total |             |
|                     | N            | Pourcentage | N        | Pourcentage | N     | Pourcentage |
| Note d'examen /20 * | 12           | 100.0%      | 0        | .0%         | 12    | 100.0%      |
| Sexe                |              |             |          |             |       |             |

**Tableau de bord**

Note d'examen /20

| Sexe  | Moyenne | N  | Ecart-type |
|-------|---------|----|------------|
| 1.00  | 13.1667 | 6  | 3.43026    |
| 2.00  | 11.0000 | 6  | 3.52136    |
| Total | 12.0833 | 12 | 3.50216    |

**La moyenne** du sexe « male (1) » est de **13.16** est supérieure à celle du sexe « femelle (2) » de valeur **11,00**



***CONCLUSION***

L'enquête correspond à une méthode de collecte de données consistant à interroger des individus qui appartiennent à une population choisie ou à un échantillon représentatif de cette population-mère. Le travail d'enquête est très exigeant ; il comporte plusieurs étapes essentielles. Lors de chacune d'elles, le chercheur doit être vigilant afin d'assurer la qualité des données recueillies.

L'avantage principal de l'enquête est qu'elle permet de recueillir des données de nature très variée :

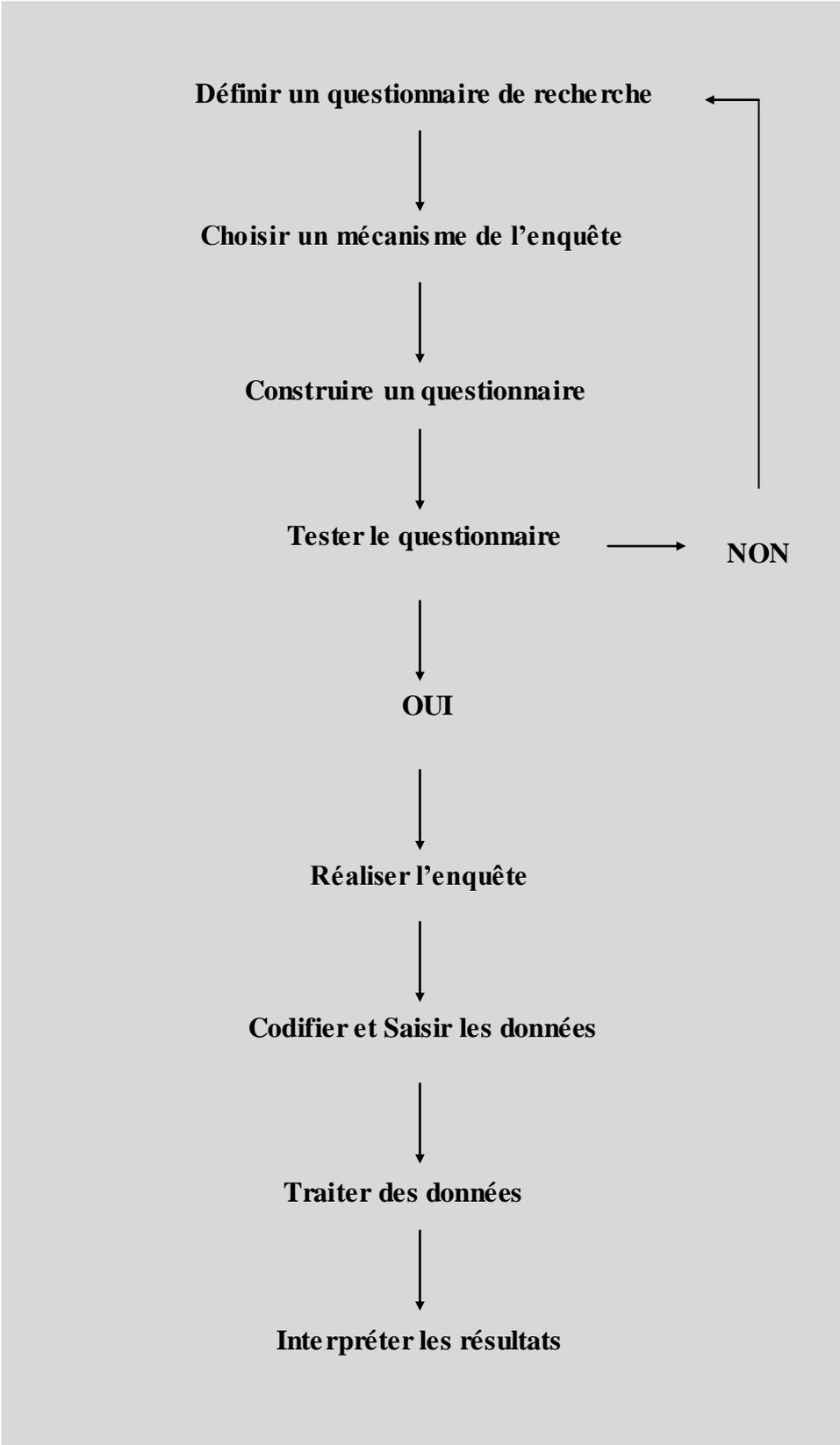
- Des données se rapportant à la personne enquêtée i.e. des données sur le profil socio-économique : l'âge, le revenu, le niveau d'instruction, l'affiliation religieuse, l'appartenance ethnique, la profession, etc.
- Des données concernant le ménage.
- Des données sur le milieu de vie ou le voisinage.
- Des données concernant les opinions.

L'enquête prend la forme d'un entretien ou d'un questionnaire entre le répondant et l'enquêteur. Ainsi de manière conventionnelle, on distinguera plusieurs types d'enquête :

- Les enquêtes par interview.
- Les enquêtes par questionnaire postal.
- Les enquêtes par téléphone.
- Les enquêtes par interception
- Etc.

Chacun de ces types comporte des caractéristiques bien précises, des avantages et des inconvénients. La maîtrise de l'ensemble de cet arsenal est indispensable pour qui prétend avoir recours à l'enquête. Le choix définitif d'une méthode d'enquête plutôt que d'une autre dépendra des objectifs de l'enquête elle-même.

D'une manière générale, pour bien faire cela, nous pouvons résumer toutes les étapes que nous avons abordées précédemment dans les deux chapitres avec ce schéma :



# *REFERENCAS*

- Ardilly P. (2006), Les techniques de sondage. Edit Technip, Paris.
- Bardin B. (2016). Cours de pratique des enquêtes. Master. Congo-Brazzaville.
- Blanchet A. & Gotman A. (2015). L'entretien. 2ème édition Armand, Colin, Paris.
- Delacroix, E., Jolibert, A., Monnot, É. & Jourdan, P. (2021). Chapitre 5. La construction d'une échelle de mesure. Dans : E. Delacroix, A. Jolibert, É. Monnot & P. Jourdan (Dir), Marketing Research: Méthodes de recherche et d'études en marketing (pp. 121-148). Paris.
- Fenneteau H. (2015). Enquête : Entretien et Questionnaire. 3ème édition Dunod, Paris.
- Gilbert R. (2002). Traitement statistique des données d'enquête avec introduction à SPSS. Département d'économétrie, Université de Genève.
- Holland, S. M. (2019). Principal Components Analysis (PCA). Department of Geology, University of Georgia, Athens, GA 30602-2501.
- Katerine Romeo, E Pissaloux, F Serin. (2019) Tableaux accessibles dans les documents numériques The Matique Informatique Robotique Imagerie. CNRIUT'2019 Congrès National de la Recherche des IUT, IUT de Toulon, France.
- Kurita, T. (2014). Principal Component Analysis (PCA). Computer Vision, A Reference Guide.
- Maćkiewicz, A., & Ratajczak, W. (1993). Principal Components Analysis (PCA). Computers & Geosciences, 19, 303-342.
- Marc X & Tcherna J. (2018). Étudier l'opinion Sondages : La révolution du numérique et des big data. 2ème édition PUG, Paris.
- Pierre Duchesne. (2017). Chapitre 2. Échantillonnage. Université de Montréal.
- Shlens, J. (2014). A Tutorial on Principal Component Analysis.
- Vandercammen, M. (2018). Chapitre 12. Analyse des données d'enquête. Dans : M. Vandercammen, Études de marchés: Méthodes & outils (pp. 359-444). Louvain-la-Neuve: De Boeck Supérieur.