

Etudes et Diagnostic en Marketing

Master 1

G. de Lanauze

IAE - Université Montpellier

Etudes et Diagnostic en Marketing

Séances 3 et 4

Les études quantitatives

Objectifs de la séance:

1. Comprendre les objectifs des études quantitatives
2. Savoir construire un échantillon
3. Savoir rédiger un questionnaire
4. Savoir recueillir et analyser les données

Travaux dirigés Simulation d'une étude quantitative:

- **Objectif : Test d'hypothèses**
- **Réduction de l'écart entre attitudes et comportements**
- **H1:** Le consommateur sera plus sensible à un argument responsable s'il est présenté comme une obligation légale imposée à l'entreprise plutôt que comme une initiative de l'entreprise
- **H2:** L'intention d'achat d'un produit avec un avantage environnemental sera supérieure si cet argument est présenté comme une obligation légale imposée à l'entreprise plutôt que comme une initiative de l'entreprise
- **H3:** Le prix psychologique d'un produit avec un avantage responsable sera supérieur si ce dernier est présenté comme une initiative de l'entreprise plutôt que comme une obligation légale

Travail à faire

• Etapes de travail:

1. **Familiarisation: test du questionnaire**
 - Compléter le questionnaire de façon novice
 - Evaluer le questionnaire
2. **Identifier les variables et décrire le modèle d'hypothèses**
3. **Décrire le design méthodologique**
4. **Constituer la base de données**
 - Faire remplir le questionnaire à personnes par étudiant en respectant des quotas de base: 3 hommes, 3 femmes, pour chaque sexe 1 par classe d'âge: 18-34, 35_54, Plus de 55 ans
5. **Analyser les données**
 - Sur SPSS, en utilisant le fichier commun compilé transmis

Les études quantitatives

- **OBJECTIF : Mesurer et quantifier, pour décrire, expliquer, valider, prévoir**

Mesurer et quantifier des phénomènes :

- directement observables : comportements d'achat, de réachat et de consommation, chiffre d'affaires, etc.
- non observables directement : attitude du consommateur, satisfaction, risque perçu, qualité perçue, connaissance, perceptions, émotions, etc.

Deux principaux types d'objectifs

Descriptif

- On cherche à connaître le poids et les hiérarchies concernant les variables d'intérêt au sein d'une population
- Ex: Quelle part représentent les consommateurs « responsables » dans la population? Quel est le profil du consommateur responsable?

Explicatif

- On cherche à mesurer la présence et l'importance d'une relation causale entre deux variables
- Ex: les consommateurs responsables achètent-ils réellement moins? Achètent-ils différemment?

Deux types de démarches correspondantes

Objectif descriptif

- Notion d'enquête :
- On observe des phénomènes tels qu'il se présentent naturellement sans les influencer

Objectif explicatif

- Notion d'expérience :
- On provoque les phénomènes auxquels on s'intéresse en maîtrisant certains facteurs

La démarche descriptive

- **Mesure directe des variables descriptive du phénomène (tris à plat)**
- **Identification d'ensembles de variables fortement corrélées entre elles: Analyses en composantes principales**
 - Utiles pour regrouper des groupes de variables en facteurs
 - Permet de construire des outils de mesure de variables non observables
- **Catégorisation et typologies de groupes d'individus semblables pour un groupe de variables données (exemple: les typologies de consommateurs par style de vie)**
- **Nécessite l'approche d'une population par échantillonnage**

La notion d'enquête

Méthode de recueil de données primaires à partir d'un questionnaire administré à un échantillon issu d'une population cible

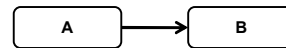
Objectif: étudier et décrire le plus précisément possible un phénomène relatif à une population cible.

Approche statistique des caractéristiques de la population sur la base de données recueillies auprès d'un échantillon de cette même population

Exemples: sondage politique, étude des comportements et des opinions, panels, ...

La démarche explicative

- **Identifier des relations causales**
 - Construire des hypothèses (lien entre deux variables)
 - Définir logiquement le sens de la causalité
 - Valider la corrélation entre les deux variables



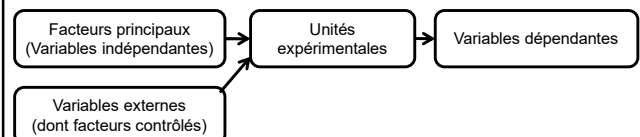
- A: variable indépendante ou explicative (antécédent)
- B: variable dépendante ou expliquée (conséquence)

La notion d'expérimentation

- **Qu'est-ce que l'expérimentation?**
- **L'expérimentation est une méthode scientifique qui consiste à tester par des expériences répétées la validité d'une hypothèse et à obtenir des données quantitatives permettant de l'affiner.**
- **Objectif: expliquer un phénomène en mettant en évidence des relations explicatives, confirmées au fur et à mesure des répétitions des tests de ces relations**

Qu'est-ce que l'expérimentation?

- **Objectif de l'expérimentation :**
 - Établir des liens de causalité en contrôlant l'environnement
 - « mesurer les effets, sur une ou plusieurs variables dépendantes, d'une ou plusieurs variables indépendantes appelées aussi facteurs principaux » (Evrard et al., 2003)



- **3 caractéristiques :**
 - Répétition des mesures
 - Affectation aléatoire (randomisation) des répondants à une situation expérimentale
 - Contrôle de certaines variables externes

Qu'est-ce qu'une variable?

Caractéristique pouvant prendre au moins 2 valeurs, quelque chose qui par définition varie.

La variation peut être qualitative:

Variable nominale

Variable nominale ordinale (ex: niveau d'études)

La variation peut être quantitative:

Variable numérique

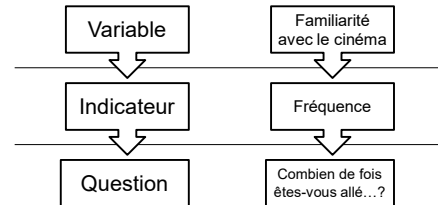
- ordinale

- d'intervalles

- de proportions (rapport)

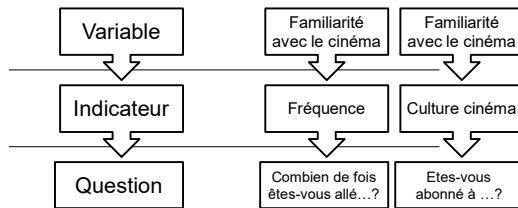
Mesurer la variable

- Affecter des nombres pour représenter les valeurs ou les propriétés (souvent cachées) de la variable
- Relier le concept (idée) à une mesure (technique) pour la rendre empiriquement observable
- Opérationnalisation



Mesurer la variable

- Lorsque la variable est abstraite et complexe, plusieurs indicateurs peuvent être pertinents pour capturer toute les facettes de la variable.



La notion d'expérience

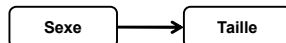
- L'expérience permet de tester une hypothèse d'un lien entre une variable A et une variable B
- L'idée est de faire varier A et de mesurer comment B réagit
- On introduit la notion de plan d'expérience pour maîtriser les facteurs concernés. En effet, dans la réalité, il existe de nombreux facteurs non maîtrisés, non contrôlés, et qui peuvent influencer les variations de B

En manipulant une ou plusieurs variable(s) indépendante(s), on va essayer de provoquer une variation des réponses des participants (donc de la VD, variable dépendante). La variable indépendante est déterminée et construite par le chercheur. Il suppose qu'elle et elle seule aura un effet sur le sens de l'hypothèse qu'il se propose de tester

La causalité

- Variables dépendantes (ou à expliquer)
- Variables indépendantes (ou prédictives, ou explicatives)

– ex:



• La causalité:

Pour pouvoir affirmer « X est la cause du phénomène », il faut :

- Une variation concomitante élevée entre X et Y. Les changements dans Y doivent toujours être associés à X.
- Une séquence temporelle. X doit précéder Y : si X est la cause de Y, il est nécessaire que X le précède chronologiquement.
- L'élimination de toutes les autres causes possibles (dites concurrentes). Pour que X cause Y, il faut que la relation existe en l'absence d'autres causes possibles. Il s'agit de maintenir les autres facteurs pouvant affecter X constants ou contrôlés.

- Une corrélation entre deux variables (mesures) ne prouve pas une relation causale • La meilleure protection contre ce type de raisonnement erroné reste toujours la réflexion théorique.

Quelles variables sont prises en compte?

- Variables manipulées :

– Les variables dont les effets des différentes modalités sur les variables expliquées sont mesurés et comparés.

– Exemple : impact du prix sur l'intention d'achat
100, 120 et 150 euros formeront autant de niveaux manipulés ou traitements.

- Variables externes :

– Toutes les variables intégrées ou non dans l'expérimentation et qui, en dehors des variables explicatives manipulées, peuvent également exercer un effet direct ou un effet indirect sur la variable à expliquer.

– Exemple : impact du prix sur l'intention d'achat
Variables qui concernent le sujet (genre, âge, expérience préalable, implication...)
Variables qui concernent l'environnement (physique, social ou temporel...)
Variables qui caractérisent la tâche à effectuer (nature, difficulté, résultat attendu...)

Les hypothèses : les liens entre les variables

• Hypothèses

– Une hypothèse est une **affirmation/prédiction testable** (vérifiable ou réfutable), à la différence des spéculations ou d'hypothèses **ad hoc**, qui sont invérifiables.

– Exemple d'hypothèse :
« Tous les corbeaux sont noirs »



– Exemple d'hypothèse ad hoc :
« La meilleure preuve qu'il existe une intelligence extraterrestre, c'est qu'elle n'a pas essayé de nous contacter ».
L'absence de preuve serait la preuve : c'est non testable.

Les étapes de l'expérimentation

• 1ère étape : la question de recherche: l'hypothèse de départ

• 2ème étape : définir le protocole expérimental

- Les facteurs manipulés
- Les niveaux retenus
- Les combinaisons de ces facteurs

• Question de la validité :

- Validité interne : mettre en évidence les effets des variables indépendantes sur les variables dépendantes
- Validité externe : généraliser les relations de causes à effets mises en évidence au delà du cadre de l'expérimentation.

• Plan d'expérience : combinaisons des niveaux sélectionnés des variables indépendantes

Unités expérimentales et plan d'expérience

• Unités testées :

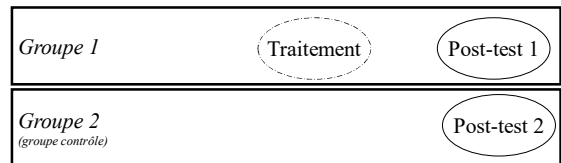
- Unités d'observations (sujets, éléments, organisations)
- Un individu, un magasin, un rayon, un emplacement publicitaire...

• Plan expérimental ou plan d'expérience :

- Affectation aléatoire des individus aux cellules de test
- Affectation aléatoire des traitements aux différentes cellules
- Contrôle des facteurs externes à l'expérimentation
- Mesurer l'effet principal des facteurs et l'effet de leurs interactions.

Les différents plans factoriels

Plan expérimental : plan « après seulement » avec un groupe de contrôle



• Expérimentation simple

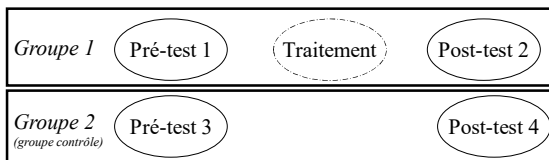
• Postulat : les événements extérieurs, non mesurés non contrôlés par le chercheur, vont influencer de la même manière les répondants des deux groupes

• Comparaison des scores de post-test

• Attribution de la différence au seul effet de traitement

Les différents plans factoriels

• Plan expérimental : plan « avant-après » avec un groupe de contrôle



• Effet de X : [Pt2-Pt1]-[Pt4-Pt3]

• Pt4-Pt3 : capture les effets d'histoire, de maturation, de régression, d'instrumentation et de sélection

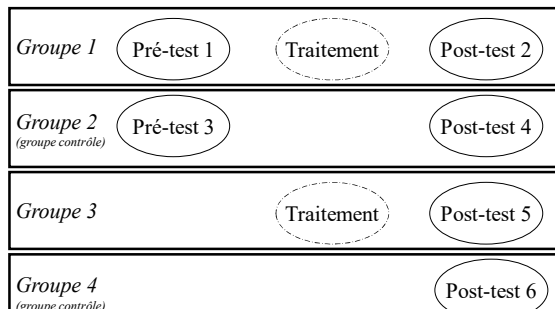
• Hypothèse : les biais mesurés sur le groupe de contrôle sont directement transposables au groupe expérimental.

• Inconvénient: effet de la 1^{ère} mesure sur l'expérimentation

Les différents plans factoriels

• Plan expérimental : plan Solomon

• Intérêt: capture des effets éventuels de la première mesure



Les différents plans factoriels

- **Exemple :**

- Un opérateur de téléphonie mobile souhaite tester deux offres d'abonnement comprenant deux types de forfaits (4 heures mensuelles soir et week-end illimités et 4 heures mensuelles 10 SMS par mois offert) et deux modèles d'appareils (Samsung et Sagem).
- Variable à expliquer : l'intention d'achat

	Forfait 1	Forfait 2
Samsung	Groupe 1	Groupe 2
Sagem	Groupe 3	Groupe 4

- **Attention:** le nombre de cellules peut rapidement devenir élevé
- **Exemple:** pour 4 facteurs et 3 niveaux = 81 cellules expérimentales !!!

Les différents biais liés à l'expérimentation

- **Effet d'histoire :** biais provoqué par des événements extérieurs à l'expérimentation et qui peuvent perturber les mesures.
- **Effet de maturation :** biais induit par tout changement intervenu dans les unités expérimentales entre deux mesures (vieillesse, fatigue, faim...).
- **Effet de test :** biais provoqué par le processus d'expérimentation lui-même.
 - Souhait du répondant de rationaliser ses réponses par rapport aux réponses fournies avant
 - Processus de mesure : le fait de mesurer un phénomène une première fois peut entraîner chez le répondant une plus grande sensibilité au sujet étudié
- **Effet d'instrumentation :** biais provoqué par l'instrument de mesure utilisé (questionnaire mal conçu, personnel mal formé...)
- **Effet de régression statistique :** biais provoqué par la tendance des scores extrêmes à se rapprocher de la moyenne pendant l'expérimentation.
- **Effet de sélection :** biais provenant de la sélection des sujets et de la façon dont ces derniers sont affectés aux conditions de l'expérience (aléatoire, appariement).
- **Effet de mortalité expérimentale :** biais provoqué par l'autoélimination ou la disparition naturelle des sujets pendant l'expérimentation
- **Effet de contamination :** biais qui se produit lorsqu'un individu qui fait partie de l'expérience a connaissance de l'objet de l'étude ou d'un aspect de l'étude qu'il n'aurait pas dû connaître ou bien encore lorsque plusieurs individus échangent entre eux au cours de l'expérimentation : toute communication ou interaction entre les individus est de nature à fausser leurs réponses.

La prise en compte des autres variables susceptibles d'influencer la VD

1. **Maintien de la variable parasite à un niveau constant (ex: n'interroger que des hommes)**
 2. **Variation systématique de la variable (contrôle par équilibrage des groupes expérimentaux, ex: respecter les quotas par sexe dans chaque groupe expérimental)**
 3. **Aléatorisation, randomisation ou contrôle par variation au hasard**
- **Contrôler également:**
 - **Les effets de rang ou de dépendance séquentielle (contre-balancement)**
 - **Le facteur « sujets », c'est-à-dire le choix des personnes interrogées**

Les fondamentaux des études quantitatives: échantillon et questionnaire

Les études quantitatives

- **Les enquêtes par sondage: l'approche d'une population par un échantillon**
 - Un recueil des données par questionnaire
 - Des questions simples pour les personnes interrogées
 - Des questionnaires directifs
 - Support : courrier, mail, téléphone
 - Un traitement statistique des données (mesure de l'erreur et tests de significativité)
 - *Généralement, nécessité de mener une phase qualitative au préalable*

Démarche de l'étude quantitative

- **Etudier une population au travers d'un échantillon représentatif de cette population:**
 - Définir les objectifs et le protocole de la recherche
 - Définir la population
 - Sélectionner l'échantillon
 - Collecter les données, généralement au travers d'un questionnaire
 - Analyser les données (statistiques)

L'échantillonnage

- L'échantillonnage consiste à choisir un groupe restreint et représentatif de répondants auprès de qui conduire la recherche au lieu d'interroger toute la population
- L'échantillonnage réduit le coût et le temps pour mener la recherche
- Un échantillonnage correctement mené fournit des informations fiables au sujet de la population

Quelques définitions

Population cible:

Population à étudier / à laquelle le chercheur souhaite pouvoir généraliser ses résultats

Unité d'étude:

plus petite unité dont est constitué l'échantillon

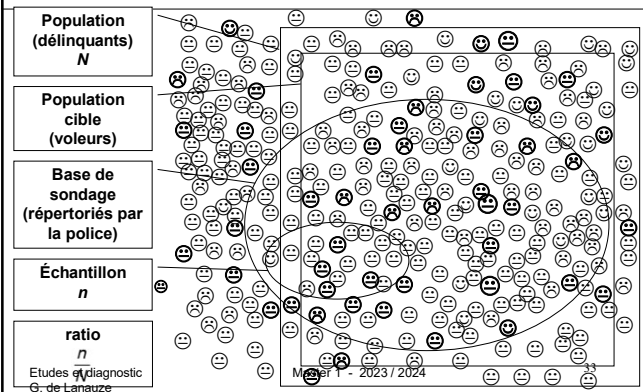
Base de sondage

Liste des unités d'échantillonnage à partir de laquelle l'échantillon est construit

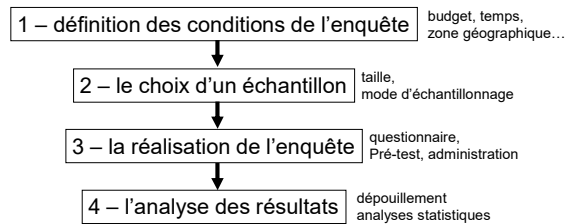
Méthode d'échantillonnage

Méthode de sélection des unités statistiques dans la base de sondage

Échantillonnage



L'étude par questionnaire: les différentes étapes



Le principe de l'échantillonnage

- Approcher une population en n'interrogeant qu'une partie de cette population: l'échantillon

- Deux questions:

Comment sélectionner les répondants ?

Question de la composition de l'échantillon



représentativité

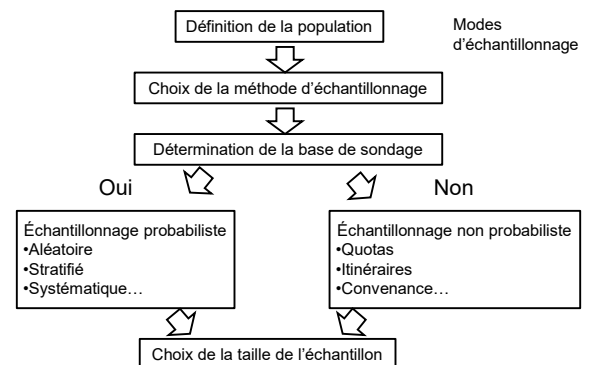
Combien de répondants ?

Question du volume ou de la taille de l'échantillon



précision

De la population à l'échantillon



La population cible

- **Définir la population cible**
 - C'est la population totale pour laquelle on a besoin de l'information
 - Il faut définir les unités qui composent la population sous forme de caractéristiques les identifiant :
 - Nature des données dont on a besoin : sur des personnes, des points de vente, hôpitaux...
 - Emplacement géographique : périmètre géographique (région, canton...)
 - Période de référence : Date
 - Autres caractéristiques dont on veut pouvoir tenir compte : caractéristiques sociodémographiques par exemple
- **L'enquête prend en compte une population différente : la population observée :**
 - Certains membres de la population cible ne sont pas observés par exemple du fait du coût de la collecte des données
- **Les conclusions ne s'appliqueront qu'à la population réellement observée**

La base de sondage

Liste des unités statistiques qui permet d'avoir accès à la population

Principal type de base de sondage: Les nomenclatures

- **Liste de noms et d'adresses qui donnent directement accès à des unités**
- **Exemple :**
 - Liste des clients de l'entreprise
 - Liste des étudiants inscrits à l'UM
 - Registre des entrées
- **Doit être complète et à jour**
- **Aucun membre de la population observée ne devrait en être exclu ni y être représenté plusieurs fois**
- **Aucune unité ne faisant pas partie de la population ne doit y figurer (décédé...)**

Les méthodes d'échantillonnage

- **Comment faire pour que l'échantillon représente au mieux la population mère ?**

- Méthodes d'échantillonnage aléatoires / probabilistes
- Méthodes d'échantillonnage empiriques

Peut-on obtenir la liste de toutes les unités d'analyse ?

OUI

Il est possible d'avoir recours à des méthodes probabilistes

NON

Il s'agit d'avoir recours à des méthodes empiriques.

Considérer les contraintes matérielles (coûts, temps)

Qu'est-ce qu'un échantillon probabiliste?

- **Principe d'équiprobabilité : chaque unité statistique a la même probabilité d'être tiré au sort**
- **Nécessité d'une Base de sondage : liste EXHAUSTIVE des individus de la population de référence**
 - **La méthode aléatoire ou élémentaire**
 - Tirer « n » participants parmi N (tirage au sort, table de nombre au hasard)
 - **La méthode aléatoire systématique**
 - Calculer le taux de sondage ; Choisir les individus en fonction de ce taux (N/n avec N = taille population et n = taille échantillon)
 - **La méthode stratifiée**
 - Tirer « n » participants parmi des sous-ensembles définis selon un critère pertinent « managérialement »

Méthodes non aléatoires

- **On distingue :**
 - **l'échantillonnage à l'aveuglette ou de commodité :** Ex.: .. Déguster un échantillon de vin.
 - **L'échantillonnage de volontaires :** Ex : Expériences médicales ou psychologiques.
 - **L'échantillonnage au jugé :** cette méthode implique la sélection d'individus en fonction de l'idée qu'on se fait de la composition de la population. On le fait pour des essais auprès des groupes cibles.
 - **L'échantillonnage par quotas :** il est largement utilisé dans les enquêtes d'opinion et les études de marché notamment parce qu'il ne suppose pas de liste des individus de la population. On parle aussi d'**échantillonnage dirigé ou par choix raisonné**. On demande aux enquêteurs de faire un nombre d'entrevues dans divers groupes établis en fonction du secteur géographique, de l'âge, du sexe ou d'autres caractéristiques... L'enquêteur doit respecter son quota.

Echantillons non probabilistes ou empiriques selon la méthode des quotas

- **Construire l'échantillon selon des caractéristiques prédéfinies**
- **Respecter les proportions sur chaque caractéristique**
- **Base de sondage : la liste exhaustive des individus de la population de référence n'existe pas**

AVANTAGES :

- Moins coûteux
- Mise en œuvre plus rapide
- Modèle réduit de la population
- Exhaustivité de la population non nécessaire
- Méthode la plus utilisée

INCONVENIENTS :

- Précision des résultats non connue
- Représentatif pour les critères retenus seulement
- Nécessité de statistiques à jour
- Liberté de choix de l'enquêteur

La méthode des quotas

- **Hypothèse sous-jacente** : si l'échantillon est représentatif de la population étudiée du point de ces critères, il le sera aussi du point de vue des caractéristiques sur lesquelles porte l'enquête.
 - Procédure :
 - Connaître la composition de la population mère
 - Constituer l'échantillon en fixant des « quotas » proportionnels aux poids relatifs de chaque catégorie d'individus
 - Quotas simples
 - Restituer les fréquences sur chaque variable de contrôle, sans se soucier de la répartition à l'intérieur de chaque catégorie
 - Quotas croisés
 - Restituer les fréquences à l'intérieur de chacun des croisements entre catégories définissant les variables de contrôle

Exemple : La méthode des quotas croisés

Répartition de la population mère sur les critères de sexe et d'utilisation			
	Hommes	Femmes	%
Utilisateurs	29%	22%	51%
Non utilisateurs	11%	38%	49%
%	40%	60%	100%

- si on désire interroger 500 personnes, on obtient la répartition suivante :

Constitution de l'échantillon			
	Hommes	Femmes	N
Utilisateurs	?	?	?
Non utilisateurs	?	?	?
N	?	?	?

Exemple : La méthode des quotas croisés

Répartition de la population mère sur les critères de sexe et d'utilisation			
	Hommes	Femmes	%
Utilisateurs	29%	22%	51%
Non utilisateurs	11%	38%	49%
%	40%	60%	100%

- si on désire interroger 500 personnes, on obtient la répartition suivante :

Constitution de l'échantillon			
	Hommes	Femmes	N
Utilisateurs	145	110	255
Non utilisateurs	55	190	245
N	200	300	500

Le redressement d'échantillon

- **Redressement par suppression**

- Principe : supprimer aléatoirement des répondants parmi les catégories surreprésentées
- Conséquences :
 - Réduction de la taille de l'échantillon
 - « Gâchis » des informations collectées
- Exemple :

Constitution de l'échantillon			
	Échantillon souhaité	Échantillon obtenu	Échantillon redressé
Utilisateurs	145 (63%)	120 (52%)	?
Non utilisateurs	85 (37%)	110 (48%)	?
%	230	230	?

Le redressement d'échantillon

- **Redressement par suppression**

- Principe : supprimer aléatoirement des répondants parmi les catégories surreprésentées
- Conséquences :
 - Réduction de la taille de l'échantillon
 - « Gâchis » des informations collectées
- Exemple :

Constitution de l'échantillon			
	Échantillon souhaité	Échantillon obtenu	Échantillon redressé
Utilisateurs	145 (63%)	120 (52%)	120 (63%)
Non utilisateurs	85 (37%)	110 (48%)	70 (37%)
%	230	230	190

Le redressement d'échantillon

- **Redressement par pondération**

- Principe : affecter un poids différent à chacune des données en fonction de la catégorie
 - Poids supérieur à 1 : catégorie sous-représentée
 - Poids inférieur à 1 : catégorie sur-représentée
- Conséquences :
 - Conserve toutes les données
 - Maîtriser un logiciel statistique
- Exemple :

Constitution de l'échantillon				Poids utilisé pour le redressement
	Effectif obtenu	% obtenu	% souhaité	
Utilisateurs	120	52,174%	63%	?
Non utilisateurs	110	47,826%	37%	?
%	230	100%	100%	?

Le redressement d'échantillon

• Redressement par pondération

- Principe : affecter un poids différent à chacune des données en fonction de la catégorie
 - Poids supérieur à 1 : catégorie sous-représentée
 - Poids inférieur à 1 : catégorie sur-représentée
- Conséquences :
 - Conserver toutes les données
 - Maîtriser un logiciel statistique
- Exemple :

Constitution de l'échantillon				Poids utilisé pour le redressement
	Effectif obtenu	% obtenu	% souhaité	
Utilisateurs	120	52,174%	63%	1,2075
Non utilisateurs	110	47,826%	37%	0,7736
Etudes et diagnostic G. de Lanauze	230	Master 1 - 2023 / 2024	100	49

Exemple d'échantillonnage par quotas

Exemple: Clientèle potentielle d'une voiture familiale de 10 chevaux

	Echantillon sur 1000 personnes	
Hommes	75%	750
Femmes	25%	250
20-30 ans	20%	200
31-40 ans	50%	500
41-50 ans	30%	300
Patrons de l'industrie et du commerce	10%	100
Cadres sup et professions libérales	50%	500
Cadres moyens	40%	400

Plan de sondage correspondant

CSP	Patrons		Cadres sup et professions libérales		Cadres moyens		Totaux
Sexe	H	F	H	F	H	F	
20-30	15	5	75	25	60	20	200
31-40	38	12	188	62	150	50	500
41-50	22	8	113	37	90	30	300
Totaux	100		500		400		1000

La taille de l'échantillon

• Elle dépend de deux facteurs:

- La précision souhaitée et le coût de l'enquête
 - Les coûts terrains sont des coûts variables, directement dépendants de la taille de l'échantillon
 - La précision statistique est liée à la taille absolue de l'échantillon, non au rapport entre la taille de l'échantillon et la population totale
 - La précision est proportionnelle au carré de la taille de l'échantillon

Population	Marge d'erreur			Intervalle de confiance		
	10%	5%	1%	90%	95%	99%
100	50	80	99	74	80	88
500	81	218	476	176	218	186
1000	88	278	906	215	278	400
10000	96	370	4900	264	370	623
100000	96	383	8763	270	383	660
1M et +	97	384	9513	271	384	664

L'objectif d'estimation: le choix du degré de précision souhaité

- La taille de l'échantillon dépend des objectifs en termes d'estimation
- Le chercheur doit préalablement définir le cadre de l'estimation souhaitée:

- L'information disponible sur la probabilité connue de réalisation de l'événement étudié
- La précision attendue
- Le niveau de confiance

Taille de l'échantillon

- Pour estimer la taille de l'échantillon il faut préalablement définir:

- La probabilité de réalisation de l'événement
Connaître l'occurrence de l'événement dans la population par domaine d'estimation (*urbain-rural par ex.*)
- La marge d'erreur
Spécifier le degré de précision souhaité (par un intervalle de confiance, (marge d'erreur)) (*ex. à 95%*)
- Le niveau de confiance
Spécifier le degré de confiance (*ex. 5%*)

Combien de personnes de la population jugent votre produit sympathique?

- **la marge d'erreur :**

C'est la déviation (positive ou négative) que vous vous autorisez dans les résultats de votre échantillon comparé avec le groupe-cible entier.

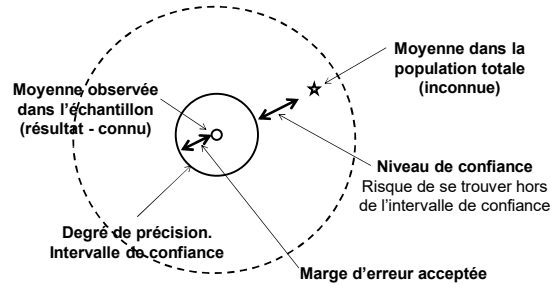
Si vous choisissez une ME de 5% et que vous obtenez 90% sur votre échantillon, vous acceptez que la réponse de la population soit comprise entre 85% et 95% (intervalle de confiance)

- **le niveau de confiance :**

C'est le risque que la réponse dans la population ne soit pas incluse entre 85% et 95% (autrement dit dans quelle mesure vous pouvez être sûr de la marge d'erreur).

Il reste 5% de chance que votre résultat soit faux

Illustration graphique



La probabilité de l'occurrence

- Si vous savez déjà qu'en général, les gens trouvent vos produits sympathiques (par exemple par le résultat d'études antérieures), alors votre probabilité de vous tromper est moindre et vous pouvez interroger moins de répondants pour un même niveau de précision.

- Cette probabilité d'occurrence est p (la probabilité de non occurrence est de 1-p = q)

- Si vous savez que 80% des gens jugent vos produits sympathiques: p = 0,8 et q = 0,2

- Si vous n'avez aucune information, alors c'est « pile ou face », ou encore 50/50, et alors: p = q = 0,5 (cas le plus fréquent)

La taille de l'échantillon

La taille d'échantillon se calcule donc avec la formule suivante:

$$n = t^2 * p * (1-p) / e^2$$

n: Taille d'échantillon minimale pour l'obtention de résultats significatifs pour un événement et un niveau de risque fixé
t: valeur type du Niveau de confiance (la valeur type du niveau de confiance de 95 % sera 1,96)
p: Probabilité de réalisation de l'événement
e: Marge d'erreur (généralement fixée à 5 %)

t est une constante tirée d'une table de distribution normale centrée réduite qui dépend du niveau de confiance (1-α)

α	1%	5%	10%	20%
t	2.575	1.96	1.64	1.282

Ainsi, pour un événement ayant une probabilité de réalisation de 40 %, en prenant un niveau de confiance de 95 % et une marge d'erreur de 5 %, la taille d'échantillon devra être de $n = 1,96^2 * 0,4 * 0,6 / 0,05^2 = 368,79$ soit 369

Taille de l'échantillon

- Pour estimer la taille de l'échantillon dans le cas de proportions, on peut utiliser la formule suivante :

$$P\left(\bar{Y} \pm t_{1-\alpha} \sqrt{\left(1 - \frac{n}{N}\right) \frac{s^2(y)}{n}}\right) = (1 - \alpha)$$

\bar{Y} valeur estimée

n la taille de l'échantillon

$t_{1-\alpha}$ la valeur critique associée à α

N la taille de la population totale

$s^2(y)$ la variance estimée de la proportion à estimer

On déduit:

$$m = t_{1-\alpha} \sqrt{\left(1 - \frac{n}{N}\right) \frac{s^2(y)}{n}}$$

est la précision en points de %

$$s^2(y) = pq, \text{ cas d'une loi binomiale}$$

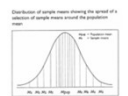
$$1 - \frac{n}{N} \approx 1$$

d'où:

$$n = t_{1-\alpha}^2 \frac{pq}{m^2}$$

Distribution théorique des échantillons dans une population donnée

Population	individus	A	B	C	D	E	F	G	H	I	J	Moy.
	Score	0	1	2	3	4	5	6	7	8	9	4,5
Échantillon	Sample Means	Samples of 2 Cases		Samples of 4 Cases		Samples of 6 Cases						
	.5	1										
	1.0	1										
	1.5 - 1.75	2		2								
	2.0 - 2.67	5		10		2						
	2.75 - 3.25	3		25		10						
	3.33 - 4.00	8		43		52						
	4.17 - 4.83	5		50		82						
	5.00 - 5.67	8		43		52						
	5.75 - 6.25	3		25		10						
	6.33 - 7.0	5		10		2						
	7.25 - 7.5	2		2								
	8.0	1										
	8.5	1										
	No. of samples	45		210		210						
Mean of s. means	4.5		4.5		4.5							
% sample means > 4.00 and < 5.00	11		24		39							



La question de la taille de la population

La formule simplifiée de calcul de la taille de l'échantillon

$$(n = t^2 * p * (1-p) / e^2) \frac{p*(1-p)}{(\frac{e}{1,96})^2}$$

ne s'applique qu'à partir d'un certain seuil. Au-delà, la taille de la population ne joue plus.

Illustration:

<https://fr.checkmarket.com/calculateur-taille-echantillon/>

Quelle est la taille de ma population, et quel doit être mon échantillon?

Population de : 10? 40? 100? 10.000 100.000 1M 11M 67M

La taille d'échantillon : le cas des faibles populations

- Cas intermédiaire entre le recensement et le sondage
- $n/N > 15\%$ (1/7 de la population, en fait autour de 10.000)
- $n' = (n \times N) / (n + N)$

$$n = \frac{p*(1-p)}{(\frac{e}{1,96})^2}$$

- Exemple : 577 députés, $e = 5\%$, oui ou non à un projet

$$n = \frac{0,5*(1-0,5)}{(\frac{e}{1,96})^2}$$

- $n = 384$ (ce qui représente 66 % de la population)
- $n' = (384 \times 577) / (384 + 577) = 230$
- Un échantillon de 230, au lieu de 384, est suffisant ici pour une erreur maximum de 5 points.

La taille d'échantillon

- La précision a tendance à augmenter avec la taille de l'échantillon.
- Pour un niveau de précision souhaité, il est possible de déterminer la taille minimale de l'échantillon.
- La précision statistique d'un sondage est plus liée à la taille absolue de l'échantillon qu'au rapport entre la taille de l'échantillon et celle de la population totale (1000 français ou 1000 belges). La taille de l'échantillon est indépendante de celle de la population quand l'échantillon représente moins de 10% de la population
- La précision des estimations ne varie pas proportionnellement à la taille de l'échantillon, mais à la racine carrée de celle-ci.
- La taille de l'échantillon n'est pas, à elle seule, une garantie de validité des estimations tirées du sondage.

- <https://fr.checkmarket.com/blog/taille-d-echantillon-enquete/>

Autre exemple

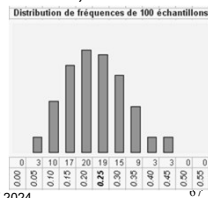
- Pour une enquête sur les stations service n'atteignant pas 40 stations comment déterminer la taille de l'échantillon pour avoir des bons résultats?
- Si la population est si petite, il faut en fait inclure presque chaque station service dans la recherche. Pour un niveau de confiance de 95% et une marge d'erreur de 5% on a besoin de 37 stations sur 40.

La question des groupes

- Je suis en train d'évaluer un programme. J'ai une clientèle admissible de 31000 personnes réparties dans 17 régions. Toutes les régions n'ont pas le même poids. L'outil de calcul m'indique un échantillon de 380 avec un niveau de confiance de 95% et une marge d'erreur de 5%. Est-ce correct de répartir mon échantillon en pourcentage de personnes présents dans chaque région? Ou dois-je faire mon échantillonnage par région?
- Ça dépend du niveau auquel vous voulez tirer vos conclusions. Si vous êtes satisfaits avec des conclusions globales, votre échantillon sera 380 au total. Si vous voulez des conclusions par région avec marge d'erreur de 5% et niveau de confiance 95%, vous devez calculer les échantillons par région.

Les erreurs d'échantillonnage

- Erreurs d'échantillonnage
 - Provenant des méthodes
 - Provenant de la taille de l'échantillon
- Erreurs de recueil de l'information, ex: consignes respectées, homogénéité du recueil,...(NON ESTIMABLES)
 - Erreurs dues au questionnaire
 - Erreurs dues aux enquêteurs
 - Erreurs dues aux enquêtés



• Erreur totale = Erreur d'échantillonnage + Erreur d'observation + Erreur due au défaut de couverture et aux non réponses

Le taux de réponse

- **Pourcentage de personnes qui répondent effectivement à un sondage qu'elles reçoivent**
- **Aide à déterminer le nombre total de sondages que vous devez envoyer pour obtenir le nombre requis de sondages terminés.**
 - Les taux de réponse varient beaucoup en fonction de différents facteurs : relation avec le public cible, longueur et complexité du sondage, incitations, thème du sondage, etc.
 - Pour les sondages en ligne pour lesquels il n'existait aucune relation préalable avec les destinataires, on estime qu'un taux de réponse atteignant 20 à 30 % est remarquable. Il est plus probable et plus prudent de compter sur un taux de réponse de l'ordre de 10 à 15 % si vous n'avez encore jamais interrogé votre population.
 - Par exemple, si vous avez besoin que 100 utilisatrices de shampooing répondent à votre sondage et si vous estimez que 10 % des femmes auxquelles vous enverrez ce sondage y répondront, il vous faut alors l'envoyer à $100 \times 10 = 1000$ utilisatrices de shampooing.

La collecte des données: rédaction d'un questionnaire

- **Doit remplir deux fonctions:**
 - Traduire les objectifs de la recherche en questions spécifiques
 - Inciter les personnes à coopérer et fournir les informations correctement
- **Administration du questionnaire:**
 - Face-à-face, Auto-administrée par voie postale, Par téléphone, Auto-administrée par internet
- **La rédaction du questionnaire**
 - Longueur et structure du questionnaire
 - Le principe de l'entonnoir
 - Les types de questions
 - Questions ouvertes
 - Questions fermées
 - Questions préformées
 - Échelles d'attitudes



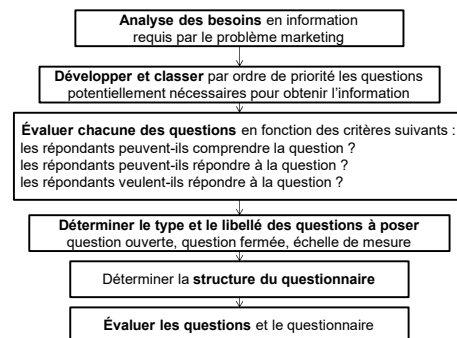
Le questionnaire, outil de collecte et outil de mesure

- **Le but du questionnaire est de recueillir les informations souhaitées auprès de l'échantillon défini et de les présenter sous forme de réponses numériques correspondant à la réponse donnée par chaque répondant**
- **Les questions (items) du questionnaire sont des instruments de mesure**
- **Toutes les informations obtenues ont vocation à rejoindre un tableau présentant en colonnes les réponses à chaque question pour chaque répondant (en ligne)**
- **Ce tableau sera la base des analyses statistiques**

Les étapes de la construction d'un questionnaire

R. A. Peterson, Marketing Research (1982)

pris dans Le marketing, Fondements et Pratique, Dubois & Jolibert, Economica



Les questions du questionnaire

Trois problèmes principaux:

- **Obtenir des chiffres**
 - Immédiat pour les questions appelant des chiffres
 - Problèmes pour les questions dites ouvertes
- **Définir l'instrument de mesure**
 - Assez simple pour les variables concrètes
 - Plus délicat pour les variables abstraites: échelles de mesure
- **Anticiper le processus d'analyse**
 - Sous quelle forme et avec quel degré de précision l'information sera-t-elle obtenue?

Les types de questions

- 1 - **Question fermée dichotomique** : l'interviewé dispose de 2 réponses possibles.
- 2- **Question à choix multiple** : l'interviewé choisit la ou les réponses qui lui conviennent le mieux dans la liste qui lui est proposée.
- 3- **Question ouverte** : ici la réponse est libre.
- 4- **Question à échelle d'attitude** : pour répondre à une question de ce type, la personne doit préciser sa réponse selon une échelle de valeurs.
- 5- **Question à classement ou à notation** : ce type de question propose à la personne sondée de réaliser un classement en attribuant une note aux réponses.

Les types de questions

- **1 Les questions ouvertes** :
 - on distingue deux types de questions ouvertes : celles qui amènent une réponse sous la forme de chiffres et celles qui amènent une réponse sous la forme d'un discours.

AVANTAGES	INCONVENIENTS
- Elles permettent d'obtenir des réponses spontanées et approfondies sur n'importe quel sujet. - Elles se prêtent à l'étude des attitudes. - Elles sont utiles dans une phase préparatoire pour établir la liste des réponses possibles d'une question à choix multiples.	- Grosse proportion de non réponse. - Superficialité de certaines réponses. - Le regroupement des réponses jugées voisines est long, coûteux et souvent subjectif. - Subjectivité de l'enquêteur lors de sa synthèse.

Les questions de type qualitatif

- **Les questions ouvertes de type qualitatif** : Les questions fermées présentées précédemment sont faciles à traiter mais elles supposent d'avoir anticipé toutes les réponses possibles.

⇒ Certains logiciels d'enquêtes permettent de définir des questions qualitatives "enrichissables"

⇒ La personne interrogée peut choisir sa réponses parmi celles proposées ou donner une réponse nouvelle : ces réponses nouvelles seront ajoutées à la liste initiale

En dehors des aspects ludiques, quels outils informatiques utilisez-vous le plus souvent ?

Traitement de texte : Word WordPerfect Autre
 Logiciel tableur : Excel Kaleidagraph Autre
 Navigateur : Netscape Internet Explorer Autre

Les questions de type numérique

3. **Les questions ouvertes de type numérique** : attendent une réponse chiffrée. Il peut s'agir d'une quantité, d'un âge, d'un prix, d'une note, etc. Il est indispensable, pour obtenir une réponse valable, de bien indiquer l'unité attendue et, lorsqu'il s'agit d'une note, le sens retenu (1 pouvant être la meilleure note ou la moins bonne).

*Quelle note /10 donneriez-vous à l'accueil téléphonique ? ... /10
 Combien d'heures utilisez-vous Internet ?*

... Heures Par jour Par semaine Par mois.

Avantages :

- numérique plus riche et plus précis qu'une échelle,
- numérique ⇒ classe sur une échelle possible, pas l'inverse,
- Le choix a un effet sur le taux de réponse.

Inconvénients :

- pour certaines questions les personnes refusent de donner une valeur exacte (salaire) ⇒ proposer des tranches en plus.

Les questions ouvertes

- **Questions post codées**

Qu'est-ce qui caractérise selon vous une bonne émission à la télévision ?

- **Questions pré codées**

Quels sont les types d'émissions que vous préférez à la télévision ? (cochez la case qui se rapproche le plus de la réponse donnée)

débat politique /_/

film /_/

jeux /_/

sport... /_/

autres, préciser.....

Les questions fermées

a- Questions dichotomiques

Avez-vous entendu parler d'entreprises qui ont changé de nom ?

Oui Non

Regardez-vous la télévision ?

- oui /_/_
- non /_/_

Les questions fermées

b- Questions à choix multiple et à réponse unique

- A votre avis, à qui s'adresse une entreprise lorsqu'elle change de nom ? (citez le principal public)

à son personnel à ses banquiers
 à ses clients au marché financier
 à ses actionnaires au grand public
 à ses fournisseurs autre (préciser) :

- A quel moment de la journée regardez-vous le plus la télévision ?

- au petit déjeuner /_/_
- au déjeuner /_/_
- dans l'après-midi /_/_
- le soir /_/_

Les questions fermées

Questions multichotomiques à réponses multiples (plusieurs réponses possibles)

- Quel type d'émissions préférez-vous à la télévision ? (3 réponses maxi)

débat politique /_/_ jeux /_/_
 film /_/_ sport /_/_

Action - regularly taken (%)	Reasons for action (% of total respondents)					
	To protect the environment	Convenience	To save money	For my health	Other	Another reason
Turn off lights I'm not using	95.7	41	4.8	72.3	0.3	32.6
Recycle items other than glass	95.1	72.4	8.9	2.1	1.4	12.7
Recycle glass	95.3	65.4	6.5	1.4	0.7	12.2
Buy energy efficient light bulbs	96.2	36.4	3.1	48.3	0.3	1.5
Buy organic food	43.7	12.5	0.5	3.2	36.3	1.2
Walk/bike to work	43.6	14.2	16.6	12.7	56.2	5.3
Use public transport	36.9	6.9	28	4.8	1.7	2.7
Take part in a campaign about an environmental issue	17.6	10.1	0.3	0.3	0.7	0

Avantages et inconvénients respectifs

Les questions multichotomiques à réponse unique :

- Avantages** : la simplicité du questionnement, du dépouillement, du traitement statistique.
- Inconvénients** : la frustration en cas d'hésitation, la mémorisation difficile pour le répondant.

Pouvez-vous indiquer approximativement le revenu mensuel de votre foyer ?

- moins de 500,000 Entre 500,000 et 1 500,000
 entre 1 500,000 et 2 500,000 plus de 2 500,000

Les questions multichotomiques à réponses multiples :

- Avantages** : facilité et richesse de la collecte, possibilité de mesurer la notoriété assistée.
- Inconvénients** : risque d'induire des réponses inconnues du répondant.

Parmi les marques d'eau minérale ci-dessous, cochez celles que vous connaissez, ne serait-ce que de nom :

- Sabrine Malne Vichy Hayet
 Safia Vittel Marwa Najet
 Evian Oktor Fourat Garsi

Les échelles de mesure d'attitudes

• Une échelle de mesure est un procédé permettant d'assigner à chaque unité statistique un code selon des règles bien déterminées.

Présentation des échelles de mesure

- Échelle au format Likert : modalités comprises entre « pas du tout ... » et « tout à fait ... »

Exemple : indiquez votre niveau de satisfaction vis-à-vis du produit X ?

- Pas du tout satisfait Plutôt pas satisfait Ni pas satisfait, ni satisfait
- Plutôt satisfait Tout à fait satisfait

- Échelle au format Osgood (ou échelle sémantique différentielle)

Exemple : le prix du produit X vous semble

Bon marché |-----| Cher

- Échelle sous forme de pictogramme

- Échelle de notation

- Exemple : Indiquez votre niveau de satisfaction vis-à-vis de votre banque par une note comprise entre 0 (très mauvaise) et 10 (excellente).

Echelles unidimensionnelles d'attitudes

Échelle de Likert

Pour moi, regarder la télévision, c'est avant tout :

Pas du tout d'accord Plutôt pas d'accord Ni d'accord ni pas d'accord Plutôt d'accord Tout à fait d'accord

Pour se divertir après une journée de travail.....X.....
 Pour être informé(e) des événements de la journée.....X.....

Le différentiel sémantique

Pour moi, l'émission Y est :

Extrêmement Extrêmement
 Agréable 1 2 3 4 5 6 7 désagréable
 Intéressante 1 2 3 4 5 6 7 inintéressante

Échelle de Likert

	Tout à fait d'accord	Plutôt d'accord	Plutôt pas d'accord	Pas du tout d'accord
•Je m'intéresse à la vie des entreprises				
•Quand j'apprends qu'une entreprise change de nom, ça attire mon attention				
•C'est un plaisir pour moi que de suivre les évolutions des entreprises				
•Pour moi, le nom d'une entreprise exprime ce qu'elle est				
•Dans le monde actuel, il est important de bien connaître les entreprises				
•Pour une entreprise, le changement de nom est une décision grave				

Etudes et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

85

Différenciateur sémantique

L'entreprise BIS est devenue VEDIORBIS à la suite de son rachat par VEDIOR

	Extrême-ment	Très	Assez	Ni/ni	Assez	Très	Extrême-ment
Rajeuni	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Réussi	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Vieilli
Raté

Etudes et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

86

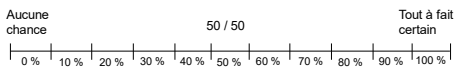
Les échelles à catégories spécifiques

Échelle d'intention

Si j'avais la possibilité d'avoir accès aux chaînes câblées :

Je regarderais plus souvent la télévision 1 2 3 4 5 Je ne regarderais pas plus souvent la télévision

Si j'avais la possibilité d'avoir accès aux chaînes câblées, il y a (cochez ci-dessous) % de chances pour que je regarde plus souvent la télévision :



Etudes et diagnostic
G. de Lanauze

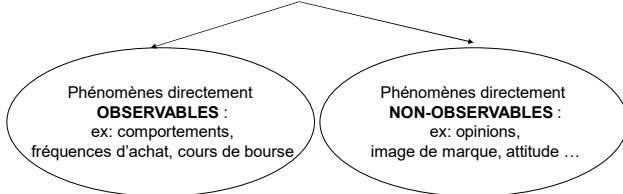
Master 1 - 2023 / 2024

87

La mesure des variables non directement observables

Problématique et principe de la mesure

QUE MESURE-T-ON ?



Etudes et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

89

La mesure des phénomènes non directement observables

- **ex:** Le chercheur qui enquête sur la relation qui existe entre la confiance vis-à-vis de la marque et la fidélité du consommateur vis-à-vis de cette même marque cherchera dans un premier temps à mesurer soigneusement ces deux concepts (confiance et fidélité) à partir de plusieurs items ; puis en fonction de leurs réponses aux différents items de mesure, les répondants se verront attribuer une valeur numérique, reflet de leur position par rapport au concept ainsi mesuré.
- **Pourquoi pas une question directe ?**
 - Différentes dimensions, difficilement réductibles à une seule question
 - Concept latent : pas directement observé, pas directement mesuré.

Etudes et diagnostic
G. de Lanauze

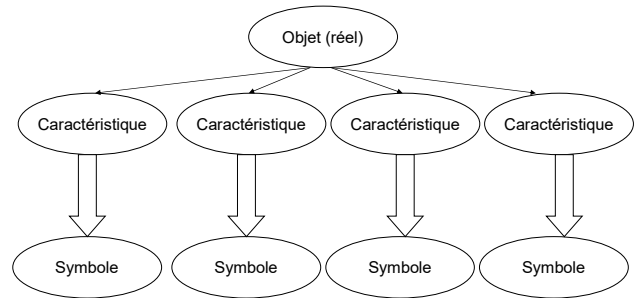
Master 1 - 2023 / 2024

90

Autre exemple

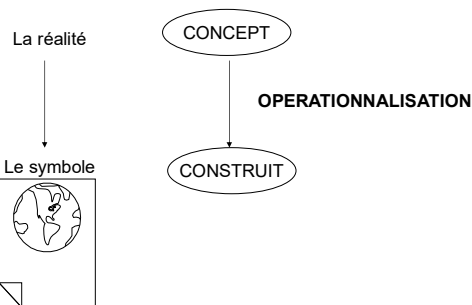
- Identifier les consommateurs innovateurs
- 1ère technique : « êtes-vous innovateur ? »
- 2ème technique : identifier et mesurer les différentes manifestations du comportement innovateur
 - J'aime acheter des produits nouveaux
 - J'aime tester et essayer les dernières nouveautés
 - J'en sais plus que les autres sur les derniers produits nouveaux
 - Etc...
- Source de ces items :
 - Littérature ou du corpus théorique
 - Etudes qualitatives
 - Test d'études quantitatives

La conceptualisation des variables non observables



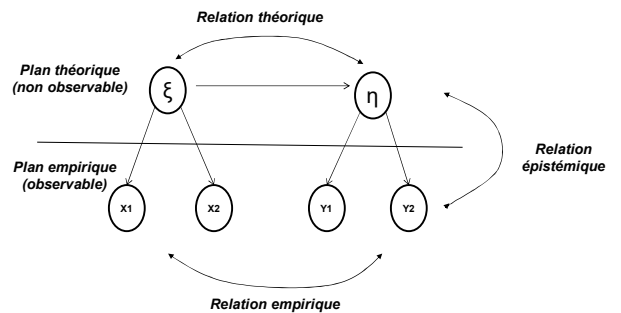
Mesure = règles d'attribution de nombres à des caractéristiques des objets

L'opérationnalisation



La question de la mesure

- Schéma général de la mesure: les chiffres et la réalité



Générer les échelles

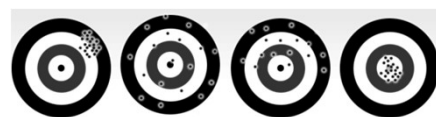
- Les items sont des questions
- S'assurer que suffisamment de questions sont posées pour refléter toutes les facettes d'un concept: multi-items
- S'assurer que les items sont bien corrélés au construit : validité convergente et au construit seul: validité discriminante

Caractéristiques d'une échelle

- Une « bonne » échelle de mesure est :

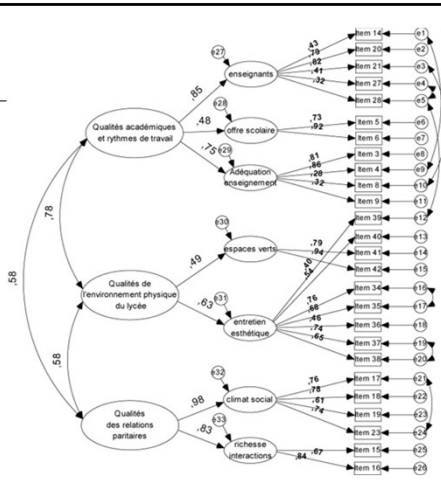
- Fiable
- Valide

- Si la fiabilité porte sur la capacité d'une échelle à reproduire le même résultat lorsqu'elle est appliquée plusieurs fois à la mesure du même concept, la validité peut être définie comme le degré avec lequel elle mesure parfaitement le construit étudié.



Exemple: La satisfaction envers le lycée

Les rectangles renvoient aux variables observées (relatives aux items de l'échelle). Les ellipses correspondent aux variables latentes (facteurs ou sous-dimensions). Les cercles représentent les erreurs de mesures.



Etudes et diagnostic
G. de Lanauze

Caractéristiques d'une échelle de mesure

Type d'échelle	propriétés	Statistiques descriptives (exemples)
Nominale	Relations d'équivalence	Mode, fréquence
Ordinale	Relations d'ordre	Médiane, quartiles
Intervalle	Pas de zéro naturel	Moyenne, écart-type
Ratio	Toutes opérations arithmétiques ok	Moyenne, écart-type

Exemples:
De proportion (quantités, poids)
D'intervalle (température)
Ordinales (préférences)
Nominales (sexe, nationalité)

Etudes et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

98

Les types d'échelles

Échelles nominales (ex : région, sexe)
Fréquences, mode, Khi Deux...

Échelles ordinales (relation d'ordre)
Médiane, quartiles, corrélations...

Échelles d'intervalle (échelle likert)
Écart type, moyenne, corrélation, test t, régression ...

Échelles de ratio (temps, distance, revenu)
Moyenne géométrique...

échelles
qualitatives

échelles
métriques

Etudes et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

99

Organisation du questionnaire

- ▶ **Commencer par des questions simples et non-offensive puis continuer en introduisant progressivement de la complexité**
 - technique de l'entonnoir
- ▶ **Vérifier que les questions s'appliquent à tous**
 - Utiliser des questions directionnelles ou filtre
- ▶ **Donner des instructions claires pour remplir le questionnaire**
- ▶ **Soigner la forme (police, taille, espaces, structure)**

Etudes et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

100

Le libellé des questions

- **Il doit être:**
 - Précis
 - Neutre
 - Simple
- **Il peut être:**
 - Objectif ou subjectif
 - Positif ou négatif

Etudes et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

101

Règles liées au fond

- **Quelle échelle retenir ?**
 - a- Méthodologie : fiabilité / validité
 - b- Pratique : simplicité / rapidité

Échelles	Fiabilité	Validité	Rapidité	Simplicité
Dichotomique	Très bonne	Faible	Excellente	Excellente
Rangs	Moyenne	Moyenne	Bonne à moyenne	Bonne à moyenne
Différentiel sémantique	Très bonne	Bonne	Bonne	Moyenne à faible
Icônes	Bonne	Bonne	Bonne	Bonne
Likert	Bonne	Bonne	Assez bonne	Assez bonne
Supports sémantiques	Très bonne	Très bonne	Bonne	Bonne

Etudes et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

Prévoir la codification et les niveaux de mesure

– Les niveaux de mesure

- a- Les échelles nominales ou catégorielles

« Quelle est votre situation familiale ? »

Célibataire	<input type="checkbox"/>	1	SITFAM	3
Vie maritale (Union libre)	<input type="checkbox"/>	2		
Marié	<input checked="" type="checkbox"/>	3		
Divorcé, séparé, veuf	<input type="checkbox"/>	4		
Autre (précisez) : _____	<input type="checkbox"/>	9		

Études et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

Prévoir la codification et les niveaux de mesure

– b- Les échelles ordinales

« Classer les 4 types de produits suivants sur chaque caractéristique. Indiquez votre classement dans le tableau ci-dessous, en donnant 1 au produit que vous jugez le "meilleur" sur la caractéristique, puis 2 au suivant, puis 3, et 4, enfin, au produit que vous jugez le moins bon. »

Caractéristiques	Huile d'olive	Beurre	Huile de tournesol	Margarine
Facilité de digestion	2	1	4	3
Goût	1	2	4	3
Facilité d'utilisation	3	4	1	2
Qualités nutritionnelles	1	3	2	4

Études et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

La rédaction du questionnaire: le vocabulaire

• Le Vocabulaire utilisé

– Le vocabulaire technique est à éviter : Comment évaluez-vous les performances d'un moteur à double arbre à came en tête ?

- Exemple : Pensez-vous que la distribution des produits d'agriculture biologique soit suffisante ? «
- Le terme distribution a un sens précis en marketing. Le dictionnaire nous donne une définition différente : c'est l'acte de répartition à des personnes...
- Les Produits d'Agriculture Biologique » (PAB) désignent l'ensemble de la production alimentaire obtenue selon un cahier des charges très précis, qui exclut notamment l'utilisation de tous les éléments résultant d'une quelconque synthèse chimique. Ce sont des produits mal connus en France et très souvent confondus avec les produits allégés ou les produits diététiques. L'utilisation de cette expression est donc source d'erreur.

Études et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

La rédaction du questionnaire: le vocabulaire

Le vocabulaire ambigu est à bannir

- **Tout mot ou expression dont la compréhension prête à confusion, soit parce qu'il/elle possède intrinsèquement plusieurs sens, soit parce qu'il/elle est subjectif(ve).**

- Exemple : La perception du temps est également subjective. Il s'agit donc d'éviter les adverbes tels que : souvent, rarement, de temps en temps, occasionnellement...
- Soit la question suivante : Au cours d'un mois normal, vous venez faire vos courses dans cet hypermarché ?
 - Très rarement
 - Occasionnellement
 - Régulièrement
 - Souvent
 - Ne sait pas

Études et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

Le style de rédaction

Style et compréhension des questions

» Interdire les doubles négations

- Exemple : Ne pensez-vous pas que les grandes surfaces ne donnent pas assez de conseils à leurs clients ?

» Éviter le biais d'acquiescement

- Exemple : Pensez-vous que la guerre des prix dans le transport aérien soit une bonne chose pour le consommateur ?

Études et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

Les impératifs de forme

Être bref

- Exemple : Si, selon vous, les étudiants mangent mal en général, pensez-vous que des actions telles que « la semaine du goût », des séances d'information sur la nutrition ou des réductions dans des magasins d'alimentation pourraient être utiles ?

- **Style et pertinence des réponses** : lorsque l'individu attribue de l'importance au sujet sur lequel il est interrogé, ses mécanismes psychologiques de défense interviennent si on utilise une approche directe

- Exemple : Prenez-vous habituellement un petit-déjeuner ? Si oui, dans la liste ci-dessous que prenez-vous en général ?

Études et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

La structure du questionnaire

- Ordre d'apparition des différents éléments du questionnaire
 - Phase 1 : présentation de l'enquêteur et de l'enquête
 - Phase 2 : questions d'introduction
 - Phase 3 : questions qualifiantes
 - Phase 4 : questions de mise en route
 - Phase 5 : questions spécifiques
 - Phase 6 : questions d'identification
- Présence de questions spéciales
 - Les questions filtre
 - Les questions piège

Le pré-test

- **Étape essentielle**
- **Interroger un échantillon de 15 à 30 personnes en conditions réelles**
 - Interroger les enquêtés
 - Interroger aussi les enquêteurs
- **Objectifs du pré-test :**
 - Vérifier la durée d'administration
 - S'assurer de la clarté des questions
 - Prendre conscience des difficultés éventuelles
 - ... dans une phase réversible du processus !!!

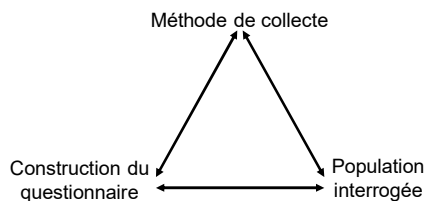
Les biais dans les enquêtes

- **Les biais liés à la compréhension du questionnaire**
- **Les biais psychologiques**
 - L'effet de désirabilité sociale
 - L'effet de conformité ou de non-conformité
 - La réaction de défense
 - L'effet d'ordre
 - L'effet d'acquiescement

L'administration du questionnaire

Les différentes méthodes d'administration

- **Deux catégories de méthodes**
 - Les méthodes auto-administrées
 - Les méthodes assistées
- **Trois questions interdépendantes**



Exemples de modes d'administration

- **L'enquête à domicile (long et coûteux)**
- **L'enquête dans la rue (doit être court et rapide)**
- **L'enquête téléphonique (lassitude des répondants)**
- **L'enquête postale (taux de réponse faible)**
- **L'enquête par internet (court et simple)**
- **L'enquête en salle (manque de diversité des répondants)**

Les méthodes auto-administrées

Méthodes d'auto-administration	Avantages	Inconvénients	Commentaire
En dépôt	-Simple -Coût faible	-Aucun contrôle sur les retours -Faible incitation à répondre	-Vu comme une simple opération publicitaire
Par courrier (questionnaire, lettre d'accompagnement et enveloppe-réponse)	-Représentativité géographique -Coût limité -Possibilité long questionnaire -Neutralité	-Gestion du taux de retour -Risque de sur-représentation de certaines catégories -Mauvais contrôle -Problème de non-réponses	-Adapté pour les clientèles captives
Sur borne interactive	-Agréable et vivant pour les répondants -Coûts de saisie économisés	-Peu de contrôle -Maintenance du matériel	-Adapté pour les enquêtes sur un point de vente ou les tests d'évaluation
En réseau	-Processus optimisé -Très rapide	-Problème d'accès -Identification des répondants	-Très adapté en intranet et extranet, plus contestable en internet

Etudes et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

115

Les méthodes assistées

Méthodes assistées	Avantages	Inconvénients	Commentaire
Par téléphone (un enquêteur + un répondant au domicile, évolue avec les portables)	-Couverture géographique -Rapidité -Coût unitaire assez raisonnable et constant -Caractère standardisable -Limite les biais liés à l'interaction sociale	-Conditions peu agréables pour le répondant -Le questionnaire ne doit pas être trop long, ni trop personnel -Échelles difficilement utilisables -Difficultés à joindre certains types de répondants	-Adapté pour les sondages nationaux "éclairés"
En face à face (un enquêteur + un répondant au domicile, dans la rue ou variantes)	-Grande interaction avec le répondant (aide, reformulation) -Pas de limitation dans les thèmes -Possibilité de montrer ou de faire tester -Rapidité, souplesse	-Biais liés à l'interaction sociale -Logistique lourde, fiabilité des enquêtes -Coûteux	-Adapté pour des sujets impliquants

Etudes et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

116

Les différentes méthodes d'administration

- **Deux remarques**
 - Les modes hybrides
 - Consiste à utiliser un enquêteur pour expliquer mais l'enquêté répond seul aux questions
 - Favorise l'implication des enquêtés
 - Ne pas employer deux modes de collecte de l'information
 - Ne pas avoir recours à deux modes de collecte (par exemple, postal et téléphone)
 - Est-ce que les écarts obtenus sont dus aux variations du phénomène étudié ou constituent un effet méthode ?

Etudes et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

117

Internet, nouvel outil d'étude

- **Domaines d'application**
 - Évaluation de sites web et de services en ligne
 - Études de satisfaction clients
 - Pré-tests et post-test de communication
 - Tests de concepts ou de produits

Avantages	Limites
<ul style="list-style-type: none"> • Coût inférieur • Rapidité • Possibilité de tester des éléments multimédias • Autonomie et confort du répondant 	<ul style="list-style-type: none"> • Échantillonnage difficile sur certaines cibles (séniors, pays émergents) • Attitude technophile des internautes

Etudes et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

118

Le traitement des données

L'analyse des résultats

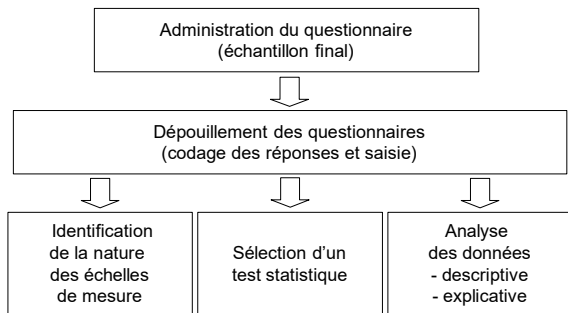
- **L'analyse des données est l'étape finale de l'étude marketing**
- **Utilisation de nombreuses techniques : analyses de contenu pour les études qualitatives ou analyses statistiques pour les études quantitatives**
- **Logiciels spécifiques comme Alceste (études qualitatives) ou SPSS (études quantitatives)**

Etudes et diagnostic
G. de Lanauze

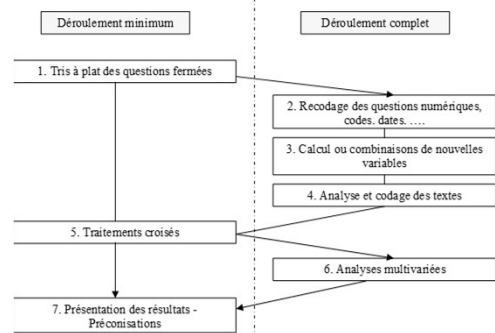
Master 1 - 2023 / 2024

120

Les étapes de l'analyse des données dans les études quantitatives



Le dépeuillement de l'enquête : démarche générale



Le dépeuillement de l'enquête : indicateurs clés

- **Le taux de réponse**
 - Ratio, exprimé en pourcentage, entre les individus ayant répondu au questionnaire et l'ensemble des individus contactés.
 - Sensible au mode de collecte et à l'implication des répondants.
 - Optimisation du taux de réponse :
 - une **attention particulière** à l'accroche ou à la lettre d'accompagnement (élément décisif dans la décision de participer à l'enquête).
 - prévoir des **relances** auprès des non-répondants.
 - proposer une **incitation concrète** :
 - les résultats de l'enquête en retour
 - la participation à un tirage au sort
 - un bon d'achat
 - une rétribution financière directe.

Le dépeuillement de l'enquête : indicateurs clés

- **Le taux d'achèvement**
 - Ratio, exprimé en pourcentage, entre les individus ayant répondu complètement au questionnaire et ceux l'ayant commencé.
 - Indicateur de progression sur Internet
- **Le taux de remplissage**
 - Ratio, exprimé en pourcentage, entre le nombre de questions auxquelles le répondant a réellement répondu et le nombre total des questions proposées dans l'enquête.
 - Dans un questionnaire auto-administré, le taux de remplissage est acceptable vers 75 %. (en face-à-face au moins, 85 - 90 %).

Le dépeuillement de l'enquête : indicateurs clés

- **La variété des réponses**
 - Identifier les individus qui donnent des réponses un peu trop systématiques afin de les isoler de l'analyse ou de les supprimer.
- **L'abondance des textes**
 - Evaluer la longueur totale (en nombre de mots) des réponses fournies aux questions ouvertes textuelles.

Le dépeuillement de l'enquête : pratiques et codification

- **Cas 1 : la question fermée (1 réponse possible)**
 - Quelle est votre marque de lessive préférée ?
 - Ariel (codification : 1)
 - SuperCroix (codification : 2)
 - Dash (codification : 3)
 - Omo (codification : 4)
 -
- **Cas 2 : la question fermée (plusieurs réponses)**
 - Disposez-vous des biens suivants ?
 - Réfrigérateur (Codification oui : 2 et non : 1)
 - Machine à laver (Codification oui : 2 et non : 1)
 - Sèche-linge (Codification oui : 2 et non : 1)
 - ...
 - Autant de variables que de modalités !!!

Le dépouillement de l'enquête : pratiques et codification

- **Cas 3 : la question fermée (1 réponse possible) - échelle**
 - Les marques de lessive sont indispensables.
 - Pas du tout d'accord (codification : 1)
 - Plutôt pas d'accord (codification : 2)
 - Ni d'accord, ni pas d'accord (codification : 3)
 - Plutôt d'accord (codification : 4)
 - Tout à fait d'accord (codification : 5)
 - Echelle de quasi-intervalle
 - Considérée comme une échelle d'intervalle
 - Et donc : possibilité de calculer une moyenne, une médiane, un écart-type....

Les méthodes de traitement des données

- **Méthodes descriptives**
 - Tris à plat, tris croisés, moyennes et dispersion, calculs de significativité des écarts, corrélations
- **Méthodes de réduction de données**
 - Résumé de plusieurs variables en facteurs (ACP)
- **Méthodes de classification des individus ou des objets (typologies)**
- **Les méthodes explicatives**
 - Régression, analyse de variance

L'analyse univariée

- **Tris à plat**
 - Tableau contenant l'effectif des réponses obtenues et la fréquence (exprimée généralement en pourcentage) pour chaque modalité de réponses proposée.
 - La fréquence relative de la modalité x_i : rapport, en %, entre l'effectif d'une valeur individuelle du caractère et l'effectif total. D'où la formule :
$$F_i = n_i / N$$
- **Traitement des non-réponses :**
- **Difficile de supprimer tous les questionnaires**
 - Somme des citations différentes du nombre de répondants
 - Remplacement des valeurs manquantes

L'analyse univariée

Elle concerne l'étude d'une variable X

- La description des données : tendance centrale et dispersion
- L'inférence : comparer les valeurs observées à une ou des valeurs prédéterminées (correspondant à un objectif à atteindre ou un résultat d'étude précédente)

Identification
de la nature
des échelles



Test
statistique
(signification
du test)

L'analyse univariée

- **Tris à plat**
 - Variables qualitatives : diagramme en tuyaux d'orgue
 - Variables quantitatives continues : histogrammes
 - Variables quantitatives discrètes : diagramme en bâtons
- **Variables quantitatives :**
 - Moyenne
 - Médiane
 - Ecart-type

» Un exemple :

Mode d'hébergement		
Hôtel	104	19,4%
Camping	113	21,1%
Location / gîte	140	26,1%
Famille / amis	179	33,4%
Total	536	100,0%

L'analyse univariée

Variable Objectifs description	Métrique	Ordinale	Nominale
Tendance centrale	Moyenne	Médiane	Mode
dispersion	Écart type	Fractile	Fréquence
Tests statistiques	Test de moyenne	Test de - Kolmogorov Smirnov	Test du Chi deux Test binomial

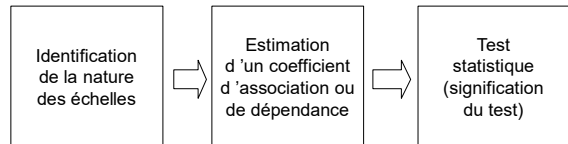
L'analyse bi-variée = les tris croisés

- **Principe:**
- **Il s'agit de comparer chaque modalité de la question A en fonction de ses réponses à la question B.**
 - Exemple : voir comment chaque tranche d'âge apprécie la qualité d'un produit.
- **Il s'agit de déterminer si les réponses à la question B varient en fonction des réponses à la question A. On étudie donc les liens statistiques entre A et B.**
- **N.B : on parle bien de lien statistique, de relation, et non pas d'influence de A sur B.**

L'analyse bivariée

Deux objectifs :

- Décrire : Tester l'association entre 2 variables ex: « Quand A est grand, B est grand » (ex: l'analyse de corrélation)
- Expliquer : Tester la causalité entre 2 variables: « A influence B » (ex: l'analyse de régression)



L'analyse bivariée

- **Mettre en évidence des relations entre deux variables analysées simultanément**
- **Construction de tableaux croisés**
 - Tableau à double entrée, en croisant les réponses à deux questions
 - Total colonne et total ligne : résultat des tris à plat
 - Trois cas :

Cas	Exemple	Traitement à mettre en œuvre	Test statistique validant la relation
Fermée x fermée	Sexe x sport pratiqué	Test du Chi-Deux	Chi-deux
Fermée x numérique	Sexe x objectif de salaire	Analyse de la variance (généralisation du test de différence de moyenne)	Test de Fisher
Numérique x numérique	Age x objectif de salaire	Régression	Coefficient de corrélation

La méthodologie

- **Le type de traitement diffère selon que l'on croise des questions qualitatives (nominales ou ordinales) ou des questions numériques.**
- **Trois méthodes de traitements peuvent être employées selon le type des questions :**

Cas	Exemple	Traitement à mettre en œuvre	Test statistique validant la relation
Nominale X Nominale	Genre x Sport pratiqué	Tri croisé	Chi deux
Nominale x Numérique	Genre x Objectif de salaire	Analyse de variance	Test de Fisher
Numérique X Numérique	Âge x Objectif de salaire	Régression	Coefficient de corrélation

L'analyse bivariée

- **Test du Chi-Deux : 2 variables non métriques**
 - Test d'indépendance des variables : si le test échoue, il n'y a pas d'indépendance entre les variables donc il y a dépendance
 - Etapes :
 - Etape 1 : construire le tableau croisé des effectifs observés
 - Etape 2 : construire le tableau des effectifs attendus/ théoriques
 - Etape 3 : calcul du Chi-deux calculé
 - Etape 4 : lecture du chi-deux dans la table statistique (degré de liberté, seuil de risque désiré)
 - Si le Chi-deux calculé est supérieur au Chi-deux lu dans la table statistique, alors le test d'indépendance des variables est négatif et les variables sont dépendantes.
 - Exemple :

Le test du Chi-deux

Le test du Chi-deux indique si la relation entre les deux variables est significative.

	Une fille	Un garçon	Total
Jogging	43,9%	56,4%	100,0%
Vélo	86,7%	23,3%	100,0%
VTT	51,9%	48,9%	100,0%
Sci alpin	60,6%	39,2%	100,0%
Snowboard	25,7%	74,3%	100,0%
Tennis	31,4%	68,6%	100,0%
Football	0,0%	100,0%	100,0%
Rando montagne	58,1%	41,9%	100,0%
Danse	89,3%	10,7%	100,0%
Natation	79,2%	20,8%	100,0%
Autres	41,1%	58,9%	100,0%
Total	48,6%	51,4%	

$\chi^2 = 45,1$; $df = 87,66$; $p < 10^{-10}$ (TS)

NS = non significatif PS = peu significatif
S = significatif TS = très significatif

Aides à l'interprétation, les cases colorées nous montrent les informations essentielles (les Chi-deux partiels les plus forts).

- en bleu, les sur-représentations
- en rose, les sous-représentations

Dans cet exemple, les pratiques sportives sont liées au genre : les garçons pratiquent particulièrement le football, les filles pratiquent plutôt la danse et la natation...

L'analyse bivariée

- **Test de différence de moyennes : 1 métrique x 1 non métrique**
 - Si, pour un risque donné, t est en valeur absolue plus grand que la valeur lue dans la table, la différence entre les deux sous-échantillons n'est pas due au hasard.
 - Exemple :
 - Consommation moyenne : 220 gr / mois
 - Consommation moyenne pour les hommes : 200 gr / mois
 - Consommation moyenne pour les femmes : 250 gr / mois

La télévision évaluée...

Chaîne de télévision regardée le plus souvent	Note globale
Arte	8,21
France 5	8,20
Canal +	7,38
France 3	6,75
M6	6,60
France 2	6,55
TF1	5,91

p = 0,1%; F = 25,11 (TS) La relation est très significative.

Entre fermées et numériques

- Le tableau de moyennes croisées va nous présenter pour chaque catégorie :
- la moyenne
 - l'écart-type
 - l'effectif
 - éventuellement la somme ou le pourcentage

Tableau de moyennes croisées

	SALAIRE OBJECTIF			
	Moyenne	Ecart-type	Part	Effectif
Une fille	21 474,03	6 462,39	45,0%	77
Un garçon	28 776,60	14 972,49	55,0%	94
Total	25 488,30	12 433,23	100,0%	171

L'écart-type nous indique la dispersion du phénomène dans la catégorie étudiée. Ici, avec un écart-type de près de 15000 F, les objectifs de salaire des garçons sont très hétérogènes.

Le test du F de Fisher

A partir d'une analyse de variance, le test F de Fisher nous indique si la relation entre les deux variables est significative.

Aides à l'interprétation, les cases colorées nous montrent les catégories dont les moyennes sont statistiquement différentes de la moyenne générale (par le test de Student).

Objectif de salaire en fonction du genre

Genre	Moyenne	Ecart-type
Une fille	21 474,03	6 462,39
Un garçon	28 776,60	14 972,49
Total	25 488,30	12 433,23

$t = 0,1\%$; $F = 15,88$ (TS)

Le test de Fisher est significatif si la probabilité de rejet (p) est < 5%.

Dans cet exemple, l'objectif de salaire varie en fonction du genre, d'une manière très significative :

- les filles ont un objectif de salaire nettement inférieur à la moyenne

L'analyse bivariée

- **Corrélation : 2 variables métriques**

- « racine du produit des coefficients directeurs »

- Interprétation :

- Si $|r| = 1$; dépendance totale entre les variables
- Si $r = 0$; indépendance entre les variables
- Si $|r| < 1$; dépendance partielle

- Signe de r

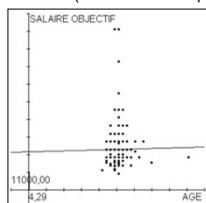
- r positif = les variables varient dans le même sens
- r négatif = les variables varient en sens contraire

- Calcul de l'indicateur t

- On considère, avec un seuil de risque de α %, que le coefficient de corrélation r est significatif quand le t calculé est supérieur au t α lu sur la table, et qu'il existe donc une relation entre les variables.

Entre deux questions numériques

Pour représenter le lien statistique entre 2 questions numériques, on utilise un nuage de points, avec une variable en abscisse (ici l'âge) et l'autre en ordonnée (le salaire escompté).



La droite de régression linéaire, de type $y = ax + b$, permet de décrire la tendance, c'est à dire l'évolution générale du salaire objectif en fonction de l'âge.

Equation de la droite de régression dans notre exemple :

$$- \text{OBJECTIF} = 76,06 \times \text{AGE} + 23827,07$$

Le coefficient de corrélation indique la dépendance entre les deux variables. Le signe du coefficient montre le sens de la relation (positive ou négative). Compris entre 0 et 1, le coefficient nous informe de la force de la dépendance. En Sciences Humaines, à partir de 0,5, on peut parler d'une relation très forte.

Dans cet exemple, avec un coefficient égal à 0,01, l'objectif de salaire n'est pas lié à l'âge, d'une manière significative.

Les analyses multivariées

- **Traitement simultané de plusieurs variables**
- **Visent à étudier de façon synthétique les relations entre variables soit au sein d'un groupe de variables (méthodes descriptives) soit entre groupes de variables (méthodes explicatives)**
- **Résultats souvent difficiles à interpréter par les non spécialistes, lecture des résultats plus ardue**

Les analyses multivariées

- **Méthodes descriptives**
 - Toutes les variables sont considérées au même niveau
 - Décrire le plus simplement possible un ensemble complexe d'information avec une perte minimale d'informations
 - Deux principales méthodes :
 - Analyse factorielle
 - Analyse en composantes principales (ACP)
 - Analyse factorielle des correspondances (AFC)
 - Analyse typologique (dendrogramme)

Les analyses multivariées

- **Méthodes explicatives**
 - Régression
 - Analyse de variance
 - Analyse discriminante
 - Analyse conjointe
- **Méthodes avancées**
 - Modèles de causalité
 - Analyses structurelles

EXEMPLE DE TABLEAU DE DONNÉES
TABLEAU FORMAT SPSS

Individus	Caractères
1	1101 1 11 1 66 2
2	1102 1 11 1 35 4
3	1103 1 11 1 33 3
4	1104 1 11 1 49 3
5	1105 1 11 1 41 3
6	1106 1 11 1 72 2
7	1107 1 11 1 50 3
8	1108 1 11 1 49 4
9	1109 1 11 1 74 1
10	1110 1 11 1 43 3
11	1111 1 11 1 65 3
12	1112 1 11 1 55 1
13	1113 1 11 1 72 1
14	1114 1 11 1 47 3
15	1115 1 11 1 49 1
16	1116 1 11 1 71 1
17	1117 1 11 1 48 3
18	1118 1 11 1 68 3
19	1119 1 11 1 46 3
20	1120 1 11 1 52 3
21	1121 1 11 1 81 1
22	1122 1 11 1 80 1
23	1123 1 11 1 42 3
24	1124 1 11 1 72 2
25	1125 1 11 1 58 1
26	1126 1 11 1 37 3
27	1127 1 11 1 41 4
28	1128 1 11 1 52 3
29	1129 1 11 1 31 4
30	1130 1 11 1 65 2
31	1131 1 11 1 60 1
32	1132 1 11 1 64 1
33	1133 1 11 1 39 3
34	1134 1 11 1 46 3
35	1135 1 11 1 65 1

TRAITEMENT ET ANALYSE DES DONNÉES

Exemple : Analyse descriptive de 3 variables quantitatives :

a. Tableau de données et commande de analyse sur SPSS

Diagram illustrating the SPSS interface for descriptive analysis:

- Tableau de données**: The data editor window showing the dataset.
- Commande de analyse**: The 'Descriptives' dialog box with 'Display' checked and 'Order by' set to 'Ascending means'.
- Variables analysées**: The list of variables selected for analysis in the dialog box.

Illustration / Application :

Voir le questionnaire de départ...

Les différentes étapes de l'analyse sous spss

- Nettoyer la base de données
- Charger les données dans SPSS
- Vérifier la qualité des échelles (fiabilité, validité convergente et divergente)
- Calculer les variables abstraites
- Calculer les scores descriptifs pour chaque variable pertinente
- Tester les relations entre les variables sur la base des hypothèses posées (tris croisés, corrélations, régressions)

- **Données collectées et recodées – échantillon final de x? répondants :**
- **Tableau Spss : variables x données (répondants)**
- **Définition de la nature des variables**
- **Analyses:**
 - Tris à plat
 - Tris croisés:
 - Nominale x nominale : ex?
 - Nominale x métrique: ex?
 - Métrique x métrique: ex?

Etudes et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

151

- Nominale x nominale : ex?
- Contrôle : scénario x sexe \Rightarrow Chi-deux
- Nominale x métrique: ex?
- Hypothèse: scénario x attribut d'image \Rightarrow Comparaison de moyennes
- Métrique x métrique: ex?
- Responsable x attributs d'image \Rightarrow Corrélations

Etudes et diagnostic
G. de Lanauze

Master 1 - 2023 / 2024

152