

→ METHODOLOGIE DE L'ENQUETE

Enseignant : Marc Jacquemain

Assistant

Sébastien Fontaine

- ***Les enquêtes sociologiques. Théories et pratique.***
- B. MATALON et R. GHIGLIONE
- Paris, A Colin, 1998

→ **LES GRANDES PARTIES DU COURS**

- **1. Les questions générales de validité**
- **2. Le questionnaire**
- **3. L'échantillonnage**
- **4. Indicateurs, variables et relations entre variables**
- **5. Initiation à l'opinion publique**

INTRODUCTION GENERALE A L'ENQUETE SOCIOLOGIQUE

→ DEFINITION DE L'ENQUETE

- Réaliser une enquête c'est **interroger** un certain nombre **d'individus** en vue d'une **généralisation**

- Chacun des termes entraîne une spécification de la méthode :
 - Interroger
 - Individus
 - généralisation

→ **INTERROGER**

→ **VS OBSERVER**

→ **VS EXPERIMENTER**

→ **VS ETUDE DES TRACES**

→ **Perturbation + grande. Rôle du langage**

→ **Moins de contrôle**

→ **Situation plus artificielle**

- **POURQUOI ALORS FAIRE DES ENQUETES ?**
- **Comportements inaccessibles**
- **Phénomènes mentaux (opinions, valeurs, croyances...)**
- **« la meilleure façon de connaître l'opinion d'un individu, c'est de la lui demander »**

→ **INDIVIDU**

→ **Réponses individuelles**

→ **Mais l'individu singulier n'est pas l'objet**

→ **VS interactions sociales :**

- Dynamique d'un mouvement
- Dynamique d'une institution

→ **VS Autres types d'entretien**

- Entretien d'embauche
- Entretien thérapeutique

→ **LA VALIDITE INTERNE**

- **Validité interne : les questions posées permettent-elles de mieux connaître la situation ou de mettre en évidence les mécanismes explicatifs ?**
 - **Influence de l'enquêteur sur la situation**
 - **Accès au seul comportement verbal**
 - **Compréhension du langage**
 - **Sélection des bonnes informations**

→ LA VALIDITE EXTERNE

- **Validité externe** : les réponses fournies nous permettent-elles de généraliser au-delà des personnes interrogées ?
- **Agrégation** : Les informations sont-elles suffisamment **uniformes** pour être additionnées ?
- **Généralisation** : L'**échantillon** de personnes interrogées est-il représentatif de la **population** qui nous intéresse ?

→ **VALIDITE DE L'ENTRETIEN
QUALITATIF**

→ **PRIVILEGIE LA VALIDITE INTERNE**

→ Réponses plus nuancées et moins
artificielles (spontanées)

→ Possibilités de rétro-action

→ **VALIDITE EXTERNE PLUS FRAGILE**

→ Réponses peu uniformes

→ Peu de personnes interrogées

→ **VALIDITE DE L'ENQUETE PAR QUESTIONNAIRE**

→ **VALIDITE INTERNE PLUS FRAGILE**

- Questions simples
- Réponses précodées
- Pas de rétroaction

→ **PRIVILEGIE LA VALIDITE EXTERNE**

- Questions standardisées (agrégation)
- Echantillons importants.
- Echantillons représentatifs

→ QUALITATIF

- L'objectif premier est de dégager des **configurations**
- Le **comptage** n'est pas pertinent
- L'essentiel est la **saturation** : ne pas avoir laissé passer une configuration spécifique
- Insistance sur la **compréhension**
- Exemple : les manières de vivre un divorce.

→ QUANTITATIF

- L'objectif premier est de dégager une **prévalence**
- Le **comptage** est fondamental
- L'essentiel est la **représentativité statistique** : inférer à une population à partir d'un échantillon
- Insistance sur la **causalité**
- Exemple : l'évolution des taux de divorce, la liaison avec les caractéristiques socio-démographiques?

LA VALIDITE INTERNE

- **L'ENQUETE S'EFFORCE DE PRODUIRE LES INFORMATIONS LES PLUS CONFORMES POSSIBLES A LA REALITE D'UNE SITUATION**
 - **Soit des informations factuelles**
 - Etes vous propriétaire de votre maison ?
 - Etes-vous parti en vacances à l'étranger cet été ?
 - Pour quel parti avez-vous voté lors des dernières élections législatives ?
 - **Soit des opinions (au sens très large)**
 - Etes-vous favorable à une intervention militaire de la Belgique en Irak ?
 - Etes-vous satisfait de votre travail actuel ?
 - Pensez-vous que les couples homosexuels doivent avoir les mêmes droits que les couples hétérosexuels en matière d'adoption

- **C'EST CE RAPPORT A LA REALITE QUI DISTINGUE UNE ENQUETE D'UNE TECHNIQUE DE MARKETING :**
 - L'enquête n'a pas pour objectif de « faire plaisir » au commanditaire
 - Elle n'a pas pour objectif non plus de « convaincre » les personnes interrogées
 - Elle doit s'efforcer de réduire tout ce qui risque de *biaiser* l'expression la plus sincère des faits ou des opinions.

- **EN CLAIR L'ENQUETE SOCIOLOGIQUE (QUALITATIVE OU QUANTITATIVE) A POUR OBJECTIF L'INFORMATION ET PAS LA COMMUNICATION**
 - L'information se soucie avant tout de validité et de vérité
 - La communication doit avant tout vanter les mérites d'un produit, d'une politique, d'une situation.

→ **LA VALIDITE INTERNE DANS L'ENQUETE PAR QUESTIONNAIRE
DEPEND DE LA MANIÈRE DE FORMULER LES QUESTIONS .**

- **Comment l'interviewé est-il susceptible de comprendre les questions posées ?**
- **Comment va-t-il traduire son avis en réponses ?**
- **Comment l'enquêteur va-t-il comprendre les réponses de l'interviewé ?**

→ **LES GRANDS PROBLEMES DE VALIDITE INTERNE**

- **L'usage du langage**
- **Les aspects cognitifs de la compréhension**
- **Les aspects motivationnels de la réponse**

→ **LA QUESTION DES INDICATEURS**

- **Les questions mesurent-elles bien ce qu'elles sont supposées mesurer ?**

LA VALIDITE EXTERNE

→ **PROBLEME DE L'ECHANTILLONNAGE**

- **Le problème** : Impossibilité en général d'étudier l'ensemble de la **population** qui nous intéresse (tous les Belges ou tous les Européens, ou tous les jeunes qui font du sport ou tous les acheteurs d'un produit....)

- **Quel échantillon** : comment extraire un sous-groupe dans la population pour obtenir des informations suffisamment fiables ?
- **combien de personnes ?**
- **comment les choisir ?**

→ ECHANTILLONS REPRESENTATIFS ET ECHANTILLONS BIAISES

- L'échantillon est représentatif si chaque membre de la population a la même probabilité d'en faire partie.
- L'échantillon est biaisé si chaque membre de la population n'a pas la même probabilité d'être sélectionné

- **ERREUR D'ÉCHANTILLONNAGE (PRECISION)**
- **La taille de l'échantillon** dépend de la **précision** souhaitée sur les résultats (ex : veut-on connaître les intentions de vote à 3 % près, à 5 % près, etc. ?) et du **degré de certitude** (risque de se tromper)

→ **LA VALIDITE ECOLOGIQUE**

- Peut-on transposer les résultats de l'enquête aux conditions de la **vie réelle ?**
- C'est le problème de la définition de l'enquête (ou de l'entretien) comme **situation sociale spécifique.**

- PLAN GENERAL
- 1. Formulation : les mots et les phrases
- 2. Les biais cognitifs
- 3. Les biais motivationnels
- 4. Construction générale
- 5. Eléments récapitulatifs

LE QUESTIONNAIRE ET SES TECHNIQUES

- PLAN GENERAL
- 1. Formulation : les mots et les phrases
- 2. Les biais cognitifs
- 3. Les biais motivationnels
- 4. Construction générale
- 5. Eléments récapitulatifs

– 1. FORMULATION : LES MOTS ET LES PHRASES

→ A) L'USAGE DES MOTS

– 1) Mots savant et mots de l'interviewé

- Le chercheur doit **adapter son vocabulaire** à celui des interviewés. Il doit donc éviter d'utiliser des mots qui risqueraient de ne pas être compris
- **1^{ère} difficulté : le chercheur n'a pas forcément une bonne idée du langage de ses interviewés**
 - « Connaissez-vous le volume de la population carcérale en Belgique ? » peut lui paraître simple
 - « Pouvez-vous me dire le nombre de personnes dans les prisons en Belgique ? »
- **2^{ème} difficulté : le chercheur peut vouloir un niveau de mesure qui ne correspond pas à celui que manipule ordinairement l'interviewé**
 - Il souhaite que l'on distingue les détenus en maison d'arrêt (non encore jugés) et les condamnés
 - Pour le public, ce sont tous des « détenus »

- **3^{ème} difficulté : tous les interviewés ne pratiquent pas le même niveau de langage**
 - Pour certains, la distinction entre « détention préventive » et « peine » est évidente
 - Pour d'autres, elle est difficile et pas forcément bien comprise

- **4^{ème} difficulté : parfois il n'est pas adapté de troquer le « langage savant » contre le « langage de l'interviewé » parce que, pour ce dernier, son propre langage n'est pas adapté à la situation de l'interview**
 - Exemple : le scientifique qui pose une question sur les pratiques sexuelles demandera : « avez-vous eu un rapport sexuel aux cours des dernières 48 heures ? », par exemple
 - L'interviewé n'utilisera jamais cette expression dans son propre langage courant
 - Mais il ne s'attend pas à ce que le chercheur utilise des mots comme « faire l'amour » ou « baiser » qui paraissent tout à fait incongrus dans le contexte
 - Donc, le langage « savant » du chercheur est parfois le plus approprié dans le contexte d'une enquête, alors même qu'il ne le serait pas dans le cadre d'une relation « familière ».

→ 2) Les mots et le contexte

- Les mots les plus simples peuvent voir leur sens varier en fonction du contexte
- **Le contexte culturel propre à une variante de la langue :**
 - Le mot « thé » en Grande-Bretagne désigne un repas mais il signifie une boisson aux Etats-Unis
 - Les Français et les Belges organisent différemment les mots « torchon », « essuie », « drap », « serpillière »...
- **Le contexte culturel propre à une catégorie sociale :**
 - Un « livre », dans un milieu d'étudiants désignera un objet d'au moins une centaine de pages, sur un sujet unique, et doté d'une reliure
 - Un « livre » dans un milieu populaire peut aussi bien être un magazine
 - Problème : le chercheur interviewe peut-être dans les deux types de milieu.
 - Il n'a donc pas forcément une référence unique.

- **Le contexte propre à une catégorie d'âge**
 - Exemple 1 : « vieux »
 - Les vieux commencent à 40 ans pour les moins de 20 ans
 - Pour les plus de 60 ans, « vieux » commence à 80 ans
 - Exemple 2 : « Sortir »
 - Pour les étudiants « sortir » voudra en général dire « guindailer » (mot argotique !)
 - Pour les adolescents « sortir avec » désignera en général une relation amoureuse avec un partenaire
 - Pour les plus âgés « sortir » voudra dire aller au spectacle ou au restaurant

- Les mêmes mots font donc référence à des réalités objectivables différentes : nos manières de découper le réel varient avec notre sexe, notre âge, notre milieu social, notre niveau d'études, notre expérience personnelle, etc.
- Pour « désambiguïser » un mot, le chercheur peut proposer lui-même une explication :
 - Exemple 1 : « quand je dis « sortir », je veux dire, passer une soirée à l'extérieur de chez soi ».
 - Exemple 2 : « quand je dis « livre », je ne compte pas les magazines »
 - Exemple 3 « quand je dis livre, cela inclut (ou n'inclut pas) les kindle.
 - Mais il doit faire attention au fait que la même explication devra être donnée à tous les interviewés **et donc donner des consignes claires aux enquêteurs**

→ 3) La multidimensionnalité des concepts

- On l'a vu, la plupart des concepts sont multidimensionnels et donc les mots qui représentent ces concepts peuvent être ambigus quand aux **représentations** qu'ils mobilisent
- **Exemple 1 : « Pensez-vous que l'Etat remplit correctement son rôle dans la politique de sécurité » ?**
 - Etat Vs **privé ?**
 - Etat Fédéral **VS communautés et régions ?**
 - Etat VS **partenaires sociaux ?**
 - Etat VS **administration ?**
 - Etat VS **citoyens ?**
- **Exemple 2 : « Vous considérez-vous comme un homme ? »**
 - Un membre du genre humain ?
 - Un membre du sexe masculin ?
 - Un adulte plutôt qu'un enfant ou un adolescent ?
- C'est généralement le contexte qui amènera la définition de la dimension pertinente, mais il est souvent difficile d'être univoque.

→ **B) LES PIEGES DES PHRASES**

→ 1) Les question multiples ou argumentatives.

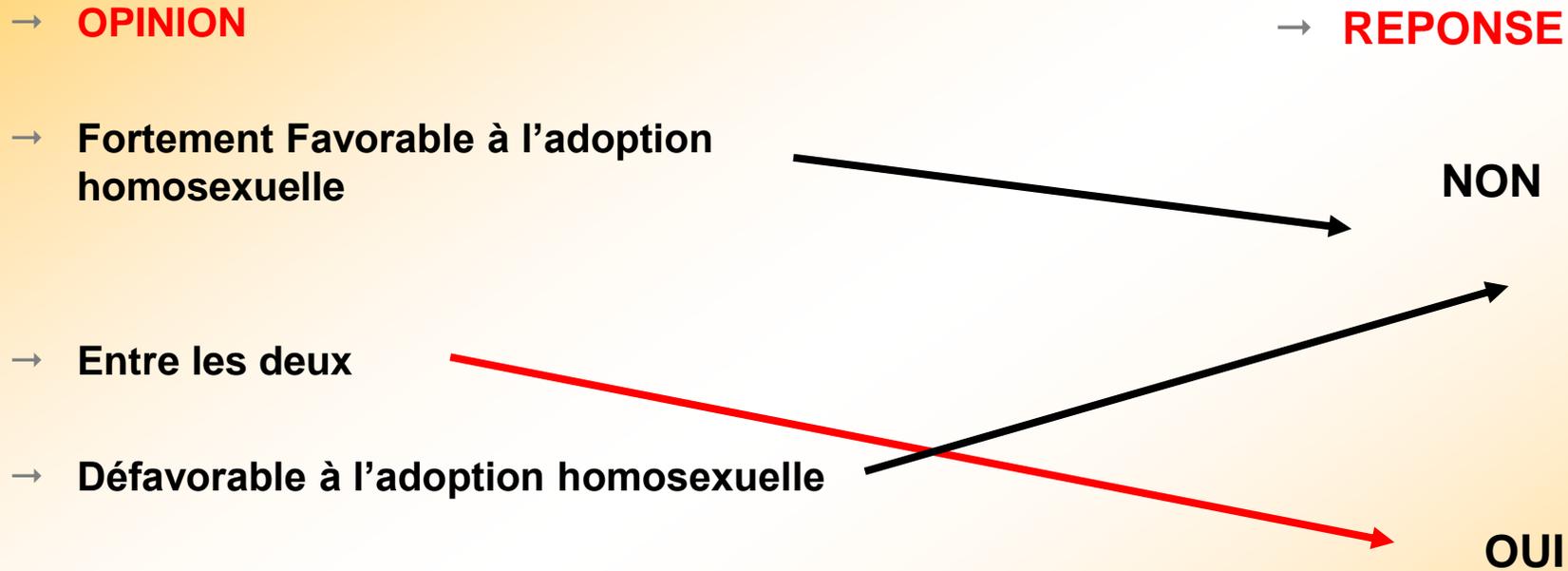
- **Exemple 1** : « **Si les juges n'étaient pas si laxistes, il y aurait beaucoup moins d'agression dans les villes** ». Décomposons :
 - Il y a beaucoup d'agression dans les villes **(1)**
 - Il pourrait y en avoir beaucoup moins **(2)**
 - Les juges sont laxistes **(3)**
 - Si (3) n'était pas vrai (2) ne le serait pas non plus. **(4)**

- Comment va réagir celui qui trouve qu'il y a peu d'agression mais que les juges sont quand même laxistes ? Ou qu'il y a beaucoup d'agressions mais que les juges font bien leur travail ?

- **Exemple 2** : « si beaucoup d'enfants disparus ne sont pas retrouvés, c'est parce que la police ne fait pas bien son travail ».
- Il y a au moins **quatre idées distinctes dans cette phrase** :
 - Il y a beaucoup d'enfants disparus **(1)**
 - Peu sont retrouvés **(2)**
 - La police ne fait pas bien son travail **(3)**
 - (3) est la cause de (2) ... et peut-être, par contagion, de (1). **(4)**
- Que fait celui qui est tout a fait d'accord avec (1) ou (2) mais pas d'accord avec (3) ou (4) ou l'inverse ?

- L'argumentation peut se trouver dans camouflée dans les **raisons** de répondre oui ou non
- Exemple : Etes-vous d'accord avec la proposition « **Les homosexuels devraient être autorisés à adopter s'ils ont déjà élevé des enfants** »
 - On peut répondre **Non** tout en étant très favorable à l'adoption par un couple homosexuel
 - On peut répondre **Non** tout en étant peu favorable à l'adoption par un couple homosexuel
 - Le « **oui** » traduira soit une attitude modérément favorable, soit qu'on choisit à défaut d'une meilleure option.
- Pourquoi ? Parce que le « oui » associe **deux idées contradictoires** :
 - Une attitude plutôt favorable à l'adoption homosexuelle
 - L'acceptation d'une restriction qui peut apparaître comme profondément injuste ou non fondée
- On est dans le cas typique ou **l'indicateur ne varie pas comme le concept que l'on cherche à estimer.**

1. FORMULATION : LES MOTS ET LES PHRASES



Les homosexuels devraient être autorisés à adopter s'ils ont déjà élevé des enfants

→ 2) Les questions à présupposés

- Les questions « **à présupposés** » sont d'une certaine façon une forme de question argumentative dans la mesure où
 - Répondre à la question « officielle »
 - Suppose que l'on a déjà admis une affirmation implicite
- Exemple classique : « **Avez-vous cessé de battre votre femme** »
 - Que l'on réponde « oui » ou « non » on est supposé implicitement avoir admis que l'on battait sa femme
 - L'interviewé n'a pas d'autre choix que de refuser la question.
- Exemple moins évident : « **pour qui allez-vous voter à la prochaine présidentielle** » ?
 - Suppose que l'on a admis que l'on irait voter.
 - C'est en fait une question « double » :
 - Irez-vous voter à la prochaine présidentielle ?
 - Si vous allez voter, pour qui voterez-vous ?

- Les présupposés peuvent ne pas forcément être perçus par l'enquêteur alors qu'ils induisent évidemment une réponse :
- Pensez-vous que les **terroristes** en Afghanistan sont sur le point d'être battus ?
 - (présupposé : qui sont les « terroristes » ?)
- Vous sentez-vous **d'abord** belge, d'abord wallon ou d'abord européen ?
 - (présupposé idéologique : exclusivisme)
- Parmi les **hommes politiques** suivants, quels sont ceux que vous souhaiteriez voir jouer un rôle plus important à l'avenir?
 - (présupposé idéologique : individualisme)

→ 3) Les questions négatives

- La **double négation** embrouille généralement l'interlocuteur puisqu'elle aboutit en fait à produire une affirmation.
- Exemple 1
 - Ne pensez-vous pas que l'Etat ne doit pas intervenir ?
 - Signifie en réalité : Pensez-vous que l'Etat **doit intervenir** ?
- Exemple 2 :
 - Il ne faudrait pas que la justice s'abstienne dans cette affaire.
 - Signifie en réalité : il faudrait que la justice **intervienne** dans cette affaire

- Le problème **d'une phrase ou question négative**, c'est qu'elle aboutit généralement à une double négation dans les faits si l'on veut répondre négativement.
- Exemple :
 - Etes-vous d'accord avec la phrase « La police ne devrait pas s'en mêler »
 - Dire « **non** » signifie que vous êtes **d'accord** avec l'intervention de la police
 - Dire « **oui** » signifie que vous n'êtes **pas d'accord** avec l'intervention
 - Le répondant est donc, de fait confronté à une double négation et risque fort de s'embrouiller.
 - Comment savoir si sa réponse ne vient pas d'une confusion ?

→ 4) Attention aux modalités de réponses !

- Elles doivent être **exhaustives** :
 - Ne pas oublier un cas possible
 - Prévoir une possibilité « Ne sait pas »
 - Eventuellement prévoir explicitement le refus de réponse
- Elles doivent être **exclusives**
 - Exemple
 - Moins de 30 ans
 - De 30 à moins de 40 ans
 - De 40 à moins de 50 ans
 - Etc...
- Echelles : **paires ou impaires ?**
 - L'échelle impaire favorise systématiquement la modalité centrale ⇒ risque de réponse « refuge »
 - L'échelle paire peut apparaître comme « choix forcé » donc susciter un rejet

→ 2. LES BIAIS COGNITIFS

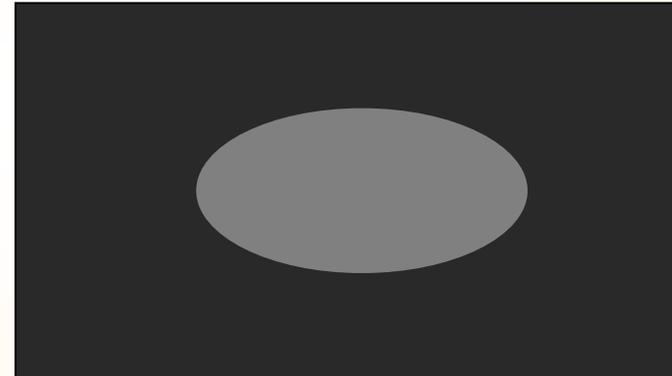
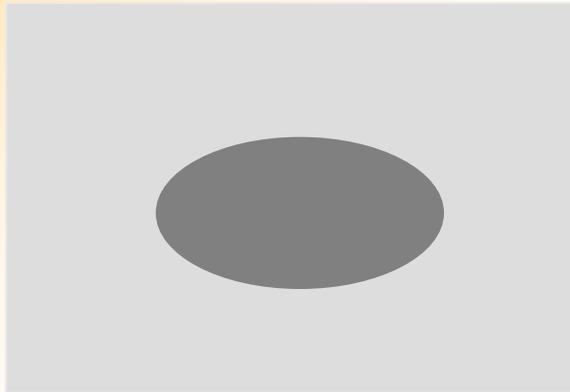
→ 1) Position du problème

- L'enquêteur recherche l'**opinion la plus « authentique » possible** des répondants.
- Pour faire simple, il veut s'approcher de ce que ses répondants « pensent vraiment » (avec toutes les difficultés liées à ce que cette expression implique et sur laquelle on reviendra)
- Il s'agit donc de contrôler au maximum tout ce qui, **dans la formulation des questions**, risque « d'induire » un certain type de réponses.
- Or, toutes les études montrent que en dépit de tous les efforts qui sont faits par un chercheur pour trouver la formulation la plus objective, la réponse à une question dépend largement à la fois du contexte dans lequel celle-ci est posée et de la façon dont elle est posée.
- Toute réponse dépend en effet du **cadre de référence** utilisé par le répondant (en anglais « **frame** ») pour la décoder.
- C'est pourquoi on parle généralement de « **framing effects** » (ce qui pourrait se traduire par « effet de cadrage » ou « effet de formatage » pour désigner les « biais cognitifs »)

2. LES BIAIS COGNITIFS

- Les biais **cognitifs** se distinguent donc des **biais motivationnels**
 - Dans le premier cas, c'est le **cadre de référence** que nous utilisons, le point de vue spécifique tel que le comprend l'interviewé qui induit une réponse. Le biais cognitif peut être inclus dans notre appareil conceptuel ou bien encore construit socialement. Une façon de faire percevoir le biais cognitif est de le construire comme une forme d'illusion d'optique. Un bon exemple d'illusion d'optique est l'effet de contraste des couleurs (voir slide suivant)
 - Le biais motivationnel, lui, est lié à notre **désir** de voir les choses sous un certain angle ou de nous présenter d'une certaine façon. L'exemple typique du biais motivationnel est les « wishfull thinking » : même si je vois le matin un ciel gris, je me persuade qu'il ne va pas pleuvoir parce que je n'ai pas envie de renoncer à ma promenade du jour. Je crois donc ce que j'ai envie de voir arriver.
- Les biais cognitifs et les biais motivationnels sont des mécanismes qui nous affectent constamment dans la vie courante. Mais il sont évidemment d'une importance capitale lorsqu'il s'agit d'interpréter la signification réelle des réponses à une enquête.

ANALOGIE VISUELLE DES BIAIS COGNITIFS



L'arrière-plan est un « cadre » qui structure notre perception de la tonalité

→ 2) Approche par un exemple

- **L'attitude à l'égard du risque** se prête particulièrement bien à la mise en évidence d'effets de « framing ». La question posée est précédée de l'explication suivante.
- « *Supposons une épidémie soudaine d'une maladie mortelle que l'on croyait éradiquée. Six cents personnes sont atteintes dans une région spécifique de l'Europe. Il ne subsiste plus qu'une quantité limitée de doses de vaccins, en principe suffisamment pour guérir à coup sûr deux cents personnes. Le temps que l'on fabrique la quantité supplémentaire, toutes les personnes atteintes seront mortes. Il faut donc distribuer la quantité disponible* »
- **Première version** : L'institut produisant le vaccin doit donc choisir deux stratégies :
 - **Soit il distribue des doses entières et deux cents personnes seront sauvées à coup sûr par une dose de vaccin. (A)**
 - **Soit il distribue le vaccin à tout le monde et chaque personne a ainsi une sur trois d'être sauvée. (B)**
- Laquelle vous paraît moralement préférable ?

- **Deuxième version** L'institut a donc le choix entre deux stratégies
 - **Soit il distribue des doses entières et il y a ainsi quatre cents personnes qui sont quoi qu'il arrive condamnées (A2)**
 - **Soit il distribue la dose à tout le monde et chaque personne a ainsi deux chances sur trois de mourir. (B2)**
- **Constat général :**
 - Les deux dilemmes sont exactement identiques mais l'un est formulé en termes de proportion de survivants et l'autre en termes de proportion de décès.
 - Toutes les enquêtes montrent que, statistiquement, les répondants préfèrent nettement A à B mais préfèrent nettement B2 à A2
 - Conclusion
 - Selon que la **même question** est formulée en termes de survie ou en termes de décès, les préférences moyennes des personnes interrogées sont inversées : la certitude de **sauver** deux cents personnes paraît préférable, mais la certitude de **laisser mourir** quatre cents personnes semble inacceptable. Mais c'est bien sûr le même choix !
 - On est très clairement dans le cadre d'un **effet de « framing »** : les répondants différencient un « verre à moitié plein » d'un « verre à moitié vide ».

- Conclusion de l'exemple : les effets de « Framing » sont souvent inévitables
 - Pour les psychologues sociaux qui les ont étudiés (Kahneman, Tverski, Nisbet, Ross...) ils s'apparentent effectivement à des **illusions d'optique**
 - En effet, un certain nombre d'entre eux, même lorsqu'ils sont expliqués, persistent.
 - Il s'agit parfois de « biais » par rapport une perception « canonique »
 - Mais parfois, il n'existe pas de « perception canonique » : la réponse est inextricablement dépendante du point de vue adopté
 - **Position de principe :**
 - **il est pratiquement impossible de poser une question sans fournir en même temps un élément de cadrage pour la réponse.**
 - **L'essentiel est donc que l'enquêteur soit conscient des éléments de cadrage fournis et qu'il s'efforce de les rendre homogènes à travers tout le questionnement**

- Un exemple célèbre :
- « *That which we call welfare by any other name would smell sweeter : an analysis of the impact of question wording on response pattern* »
 - Tom Smith (1987)
 - Public Opinion Quarterly
- Pourquoi le mot « welfare » suscite-t-il de l'aversion ?
 - Il évoque la **dépense** que représente le système pour chaque contribuable et non ce que **reçoivent** les plus pauvres
 - Le discours conservateur a donc systématiquement associé « welfare » à **gaspillage**
 - En conséquence, les enquêtes qui utilisent le mot apparaissent en moyenne beaucoup plus défavorables à l'aide aux plus pauvres que les enquêtes qui utilisent d'autres formulations.

→ 3) Cadrage par la question elle-même

→ **Exemple 1** : « *Pourquoi avez-vous acheté ce livre ?* » : le sens de la question va changer selon l'élément accentué (ou supposé essentiel)

- **Pourquoi** Indique l'idée générale de surprise
- **Acheté** Plutôt qu'emprunté ou volé
- **Ce** Pourquoi ce livre-là plutôt qu'un autre ?
- **Livre** Pourquoi un livre et pas autre chose ?

→ **Exemple 2** :

- Les Etats-Unis doivent –ils **autoriser** les discours contre la démocratie ? (A)
 - 62 % NON (soit 38 % pour autoriser)
- Les Etats-Unis doivent-ils **interdire** les discours contre la démocratie ? (B)
 - 46 % OUI (soit 54 % pour autoriser)
- Pourquoi une telle différence ?

parce que « autoriser » est un mot plus neutre et « interdire » un mot plus fort. Donc « **interdire** » n'est pas perçu comme le contraire de « **autoriser** »

- **Exemple 3** : L'importance de l'illustration explicative
 - « Pensez-vous que des légumes verts, comme par exemple, les épinards, devraient faire partie de la diète » **devient dans l'esprit du répondant**
 - « Pensez-vous que les épinards devraient faire partie de la diète ? »

- **Exemple 4** : même principe
 - « Pensez-vous que les indépendants, par exemple les plombiers ou les avocats, arrivent à cacher une partie de leurs revenus au fisc ? » **devient dans l'esprit du répondant**
 - « Pensez-vous que les plombiers et les avocats arrivent à cacher une partie de leurs revenus au fisc ? »

- Ce type de cadrage est d'autant plus important à repérer que l'utilisation d'exemples explicatifs est souvent utile. Mais le risque est toujours que le répondant assimile la substance de la question à l'exemple utilisé.

→ 4) Effet des modalités de réponse

- **Exemple 1** : à quelle classe sociale diriez-vous que vous appartenez parmi les cinq groupes suivants ? (Eurobaromètre – enquête CLEO 1997)
 - Classe ouvrière 27,3 %
 - Classe moyenne inférieure 9,3 %
 - Classe moyenne 38,3 %
 - Classe moyenne supérieure 10,4 %
 - Classe supérieure 2,9 %
 - Autre 1,6 %
 - Ne sait pas 3,1 %
 - Refuse de se classer 7,1 %

- La question comprend **trois fois la référence à la « classe moyenne »**
- La question « induit » l'interviewé à fournir une réponse et une réponse d'un certain type. Mais on peut se poser deux questions :
 - Se représente-t-il vraiment la société en termes de classes sociales ?
 - La typologie des classes proposées lui est-elle naturelle ?

- Par comparaison, voici une question posée sans référence à aucune catégorisation (CLEO-Ulg, enquête de 1997 sur 1000 Wallons adultes)
- **Exemple 2** Notre société est composée de multiples groupes ou catégories de personnes. Pouvez-vous citer un ou plusieurs groupes (jusqu'à 4) de notre société au(x)quel(s) vous appartenez ? (donc question ouverte)
- Voici les principales catégories de réponses (post-codage)
 - Groupe socio-professionnel : 28,1 %
 - Association : 20,6 %
 - Groupe défini par un attribut personnel : 12,1 %
 - Famille : 9,2 %
 - Groupe d'âge : 6,7 %
 - Classe sociale : 5,3 %
 - Divers : 18,0 %

2. LES BIAIS COGNITIFS

- Dans une question ouverte, seuls 5 % des personnes interrogées se définissent spontanément en termes de classe sociale
- Mais dans une question fermée, seuls 10 % des répondants déclarent « Ne sait pas » ou refusent de se classer.
- Quel sens donner à la réponse des 85 autres % ?
- La question « **induit** » **donc une vision de la société structurée**
 - Il y a une hiérarchie de classes sociales
 - La hiérarchie se construit notamment sur une dimension cols blancs/cols bleus
 - La classe moyenne occupe une partie déterminante de la hiérarchie
 - Chacun est capable de se positionner
- En même temps, on voit ici à quel point le passage d'une question ouverte à une question fermée peut modifier les réponses.
- **Faut-il supprimer la question « fermée » ? Non. Elle nous donne une information utilisable et comparable avec d'autres enquêtes. Et c'est souvent elle qui justifie la recherche**
- **Mais quand c'est possible, elle doit être « remise en perspective » par la question ouverte**

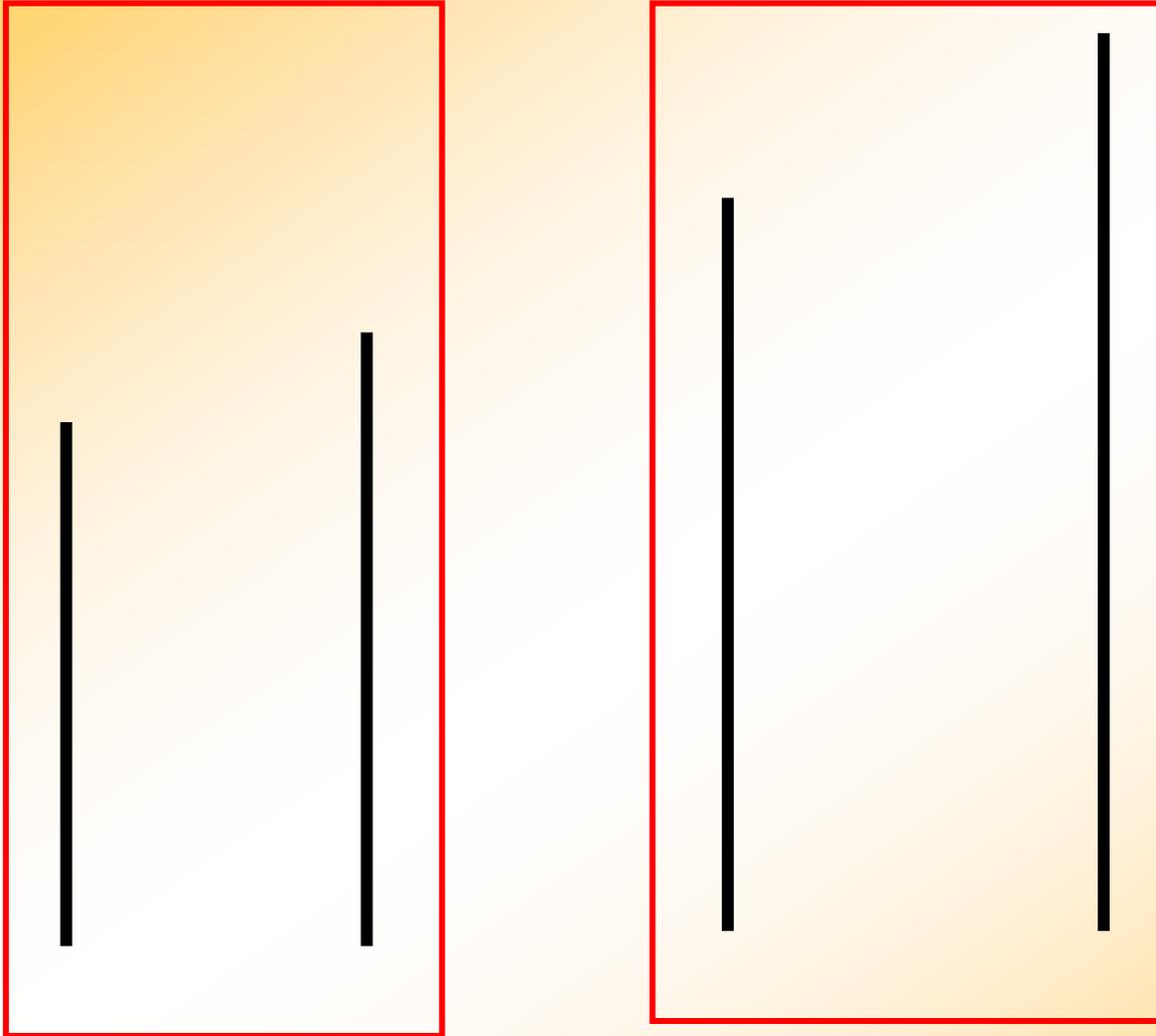
- **Exemple 3 : vérifier que les alternatives de réponse sont complètes**
- « Pensez-vous que les entreprises pourraient éviter les licenciements en période creuse ? » (A)
 - Oui : 63 %
 - Non : 22 %
 - NSP : 15 %
 - « Pensez-vous que les entreprises pourraient éviter les licenciements en période creuse ? ***Ou Pensez-vous que ces licenciements sont inévitables ?*** » (B)
 - Pourraient : 35 %
 - Inévitables : 41 %
 - NSP : 24 %
 - Explication :
 - la spécification de l'alternative « implicite » a rendu cette alternative saillante
 - Elle a aussi réduit le « **biais de positivité** » (voir plus bas)
 - Elle a accru le dilemme : le pourcentage de NSP a sérieusement augmenté

→ Exemple 4 : Même principe : expliciter les alternatives de réponse

- De manière générale :
 - Est-ce que vous préférez le thé au café ?
 - Est-ce que vous préférezle thé ou café ou vous préférez le café au thé?
 - N'induiront pas les mêmes réponses, toutes autres choses égales
- La première formulation à tendance à privilégier implicitement une des deux options : elle induit un **biais de positivité**
- La deuxième formulation insiste sur la « neutralité » de la question en mettant ouvertement les deux options sur le même pied, même si cela paraît redondant.
- L'équilibre et la complétude des échelles de réponse ne constituent donc pas des problèmes « évidents »
 - Cela demande une attention spécifique de la part du concepteur de l'enquête
 - Cela demande des pré-tests (idéalement en testant plusieurs formules, pour vérifier si elles sont équivalentes)

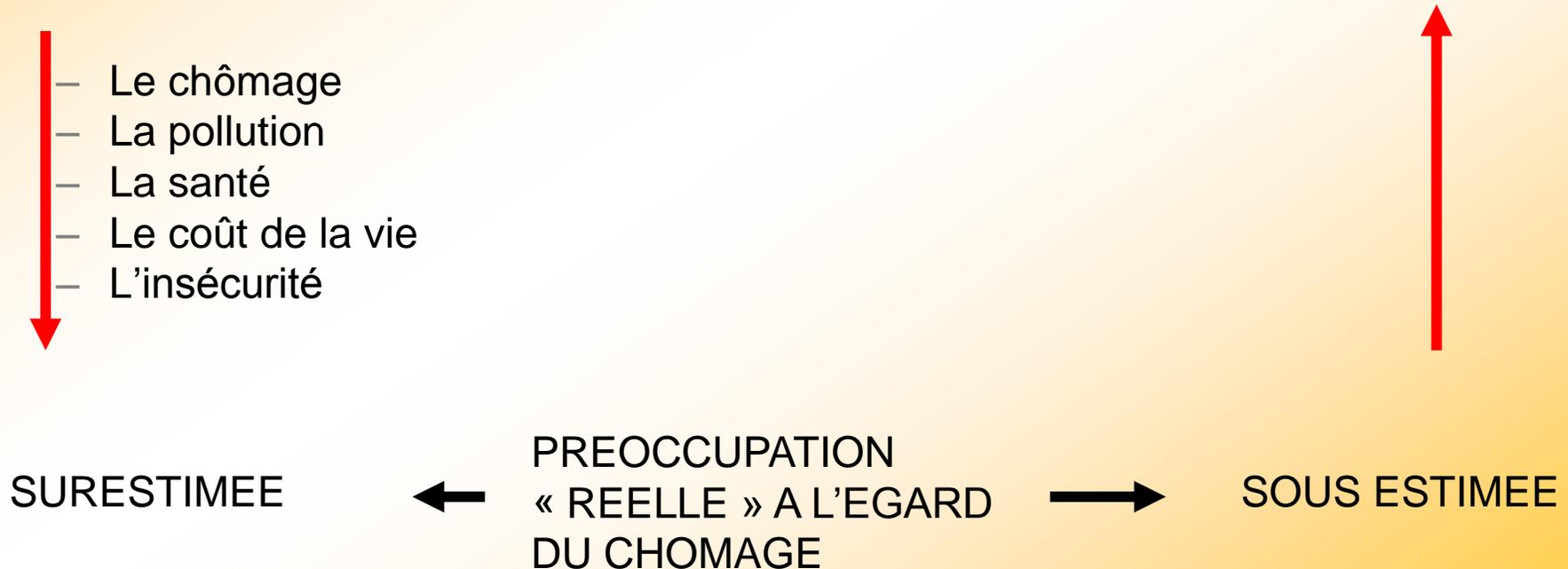
- **Exemple 5 : L'échelle de réponses** induit souvent un effet de cadrage
- Combien d'heures par jour avez-vous regardé la télévision, en moyenne, au cours de la semaine passée » ?
 - **1^{ère} échelle de réponse**
 - Moins de 1 heure
 - De 1 heure à moins de 2 heures
 - De 2 heures à moins de 3 heures
 - 3 heures ou plus
 - **2^{ème} échelle de réponse**
 - Moins de 2 heures
 - De 2 heures à moins de 3 heures
 - De trois heures à moins de 4 heures
 - 4 heures ou plus
 - Dans le cas de la deuxième échelle, le niveau moyen de réponses sera plus élevé. Pourquoi ?
 - Parce que l'échelle de réponse sert d'indice au répondant pour évaluer ce que l'enquêteur considère comme « normal »
 - Celui qui regarde la télévision plus de 4 heures se verra comme « consommateur atypique » (Couch potatoe) dans le cas de la première échelle et moins dans le cas de la deuxième échelle.

- **Exemple 6** : les modalités de réponse peuvent induire un effet de **contraste** (ou de **catégorisation**)
 - La présentation de modalités de réponses extrêmes réduira la distance apparente entre les modalités moyennes.
 - Exemple : quel degré de sympathie avez-vous pour les personnes suivantes :
 - George Bush
 - Tony Blair
 - François Mitterrand
 - Hitler
 - Oussama Ben Laden
 - Khomeyni
 - Effet de contraste :
 - Alors que les trois dernières personnes pourraient être jugées « dans l'abstrait » très différentes, la présence des trois premières va les faire paraître plutôt proches entre elles, par contraste.
 - On voit que l'effet de contraste rejoint la nécessité d'un point de référence.



EFFET DE CATEGORISATION

- **Exemple 7** : L'ordre des modalités de réponse peut produire un effet « d'ancrage » : les premières modalités vont servir à déterminer l'importance des suivantes



→ Quels mécanismes sous-tendent les effets liés à l'échelle de réponse ?

- (A) Un effet de **point de référence** :
 - De nombreuses grandeurs n'ont guère de signification pour nous dans l'absolu : nous ne pouvons les apprécier qu'à partir d'un point de référence
 - Exemple : Un pays qui fait 500.000 km² de superficie est-il un grand pays ?
 - C'est 15 fois la Belgique
 - C'est un peu moins que la France
 - C'est un 18^{ème} des Etats-Unis
 - Exemple : un salaire mensuel net de 2.300 € , est-ce beaucoup ?
 - 89 % des salariés français sont en dessous
 - Mais c'est cent fois moins que les patrons industriels les mieux payés
 - Et dix mille fois moins que le « top » de la finance mondiale
- (B) Un effet de **désirabilité sociale** (voir plus bas)
 - Nous souhaitons apparaître sous le jour le plus positif
 - Nous souhaitons en tout cas éviter d'apparaître « anormaux »
- (C) Des effets de **primauté ou de récence** (voir plus bas)
 - L'ordre des questions est aussi lié à notre capacité de remémoration

→ 5) Les effets de « Halo »

- On désignera généralement par effet de **halo** le fait qu'une question ou une modalité de réponse peut modifier l'interprétation des questions suivantes ou de l'ensemble des modalités de réponse.
- L'effet de halo provient à nouveau du fait que nos perceptions sont « structurées » et pas « analytiques » et que nous avons donc tendance à comprendre des ensembles ou des séquences en évaluant les éléments les uns par rapport aux autres.
- **Exemple 1** : légitimité de certains comportements
 - « Pensez-vous qu'il est acceptable de ne pas déclarer certains revenus au fisc » ?
 - La réponse sera différente si l'on a posé auparavant la question :
 - « Trouvez-vous acceptable que certains hommes politiques ne déclarent pas tous leurs revenus » ?

- **Exemple 2** : quels sont les adjectifs qui, dans la liste suivante, s'appliquent à Liège ?
 - Conviviale
 - Culturelle
 - **Sale**
 - Chaleureuse
 - Sociale
 - Dangereuse
 - Conviviale
 - Culturelle
 - **Propre**
 - Chaleureuse
 - Sociale
 - Dangereuse
- Le changement d'un seul adjectif dans la liste risque de modifier le choix des autres.

→ 6) Les effets d'amorçage (catégorie d'effets de halo)

→ Il y a un effet **d'amorçage** de certaines questions sur la suite : lorsque les questions antérieures vont aider l'interviewé à attribuer (erronément ou non) un sens précis aux questions suivantes.

→ **Exemple 1** : L'intérêt pour la vie politique

- Vous intéressez-vous à la vie politique ?
- Savez-vous qui est député dans votre circonscription ?
- Pouvez-vous me dire quel est le Ministre Wallon qui est en charge de l'emploi ?

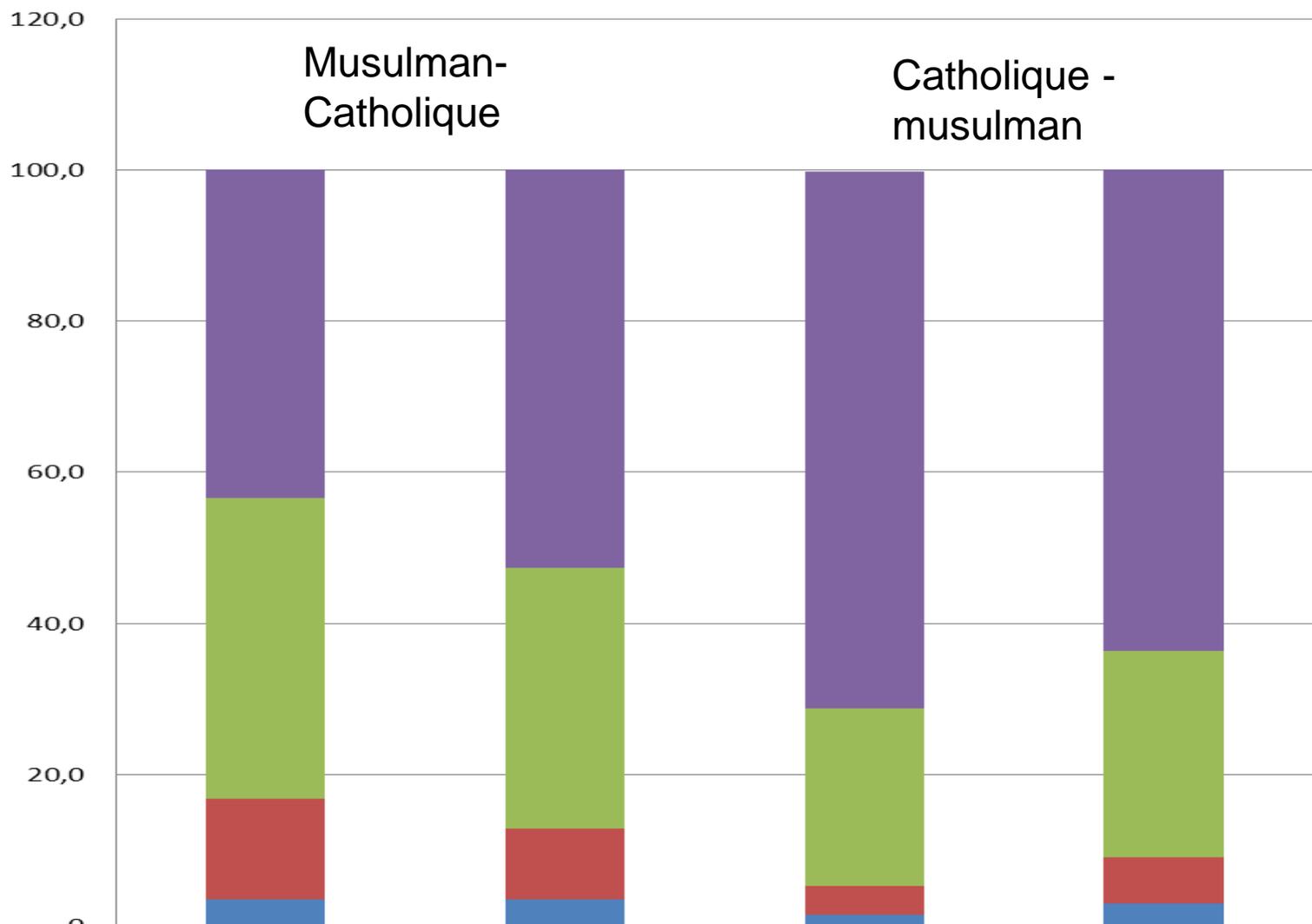


→ Poser les questions dans un sens ou dans l'autre n'est pas du tout indifférent

- Si l'on pose d'abord les autres questions, le répondant va y voir un indice de ce que l'enquêteur entend par « intérêt pour la politique ». Toutes autres choses égales, il est donc probable que l'intérêt pour la politique sera moins élevé chez ceux qui ont d'abord dû répondre aux autres questions **parce qu'ils auront dû affronter les limites de leur connaissance**.

- **Exemple 2** : Le sens de l'équité
- Voici un nouvel exemple de double question pour lesquelles l'ordre dans lequel elles sont posées modifie les réponses
 - Pensez vous que **notre pays** devrait pouvoir utiliser des restrictions commerciales pour défendre l'emploi de ses travailleurs ? (A)
 - Pensez-vous que les **autres pays** devraient pouvoir utiliser des restrictions commerciales contre les produits de notre pays pour défendre l'emploi de leurs travailleurs ? (B)

- L'expérience a montré que les répondants sont plus favorables aux restrictions commerciales dans le sens (A)-(B) que dans le sens (B)-(A)
 - Suggestion d'explication :
 - S'interdire **soi-même** des restrictions commerciales apparaît comme plus « dérangeant » que d'accepter les restrictions commerciales **des autres**.
 - Celui qui a déjà défendu les restrictions commerciales en faveur de son propre pays est poussé à les défendre, par équité, en faveur des autres
 - Celui qui s'est déjà prononcé contre les restrictions commerciales en faveur des pays concurrents devrait donc se prononcer **contre** les restrictions commerciales en faveur de son propre pays
 - Mais ici, l'équité est trop difficile et souvent hors de portée.
- Meme exemple dans la diapositive suivante (enquête réalisée par les étudiants)
 - Pensez-vous qu'un musulman vivant en Belgique devrait pouvoir exercer sa religion librement (A)
 - Pensez-vous qu'un catholique vivant en Arabie saoudite devrait pouvoir exercer sa religion librement (B)



	Pensez-vous qu'un musulman vivant en Belgique devrait pouvoir pratiquer sa religion librement ?	Pensez-vous qu'un catholique vivant en Arabie Saoudite devrait pouvoir pratiquer sa religion librement ?	Pensez-vous qu'un catholique vivant en Arabie Saoudite devrait pouvoir pratiquer sa religion librement ?	Pensez-vous qu'un musulman vivant en Belgique devrait pouvoir pratiquer sa religion librement ?
■ Tout à fait d'accord	43,4	52,7	71,0	63,6
■ Plutôt d'accord	39,8	34,5	23,5	27,3
■ Plutôt pas d'accord	13,3	9,3	3,8	6,1
■ Pas d'accord du tout	3,5	3,5	1,5	3,0

- **Exemple 3** : effet « d'exclusion » (amorçage inverse)

- Il y a un effet de « d'exclusion » lorsqu'une question spécifique est posée avant une question générale sur le même thème et que, dès lors, l'interviewé comprend que la question générale exclut la question spécifique
 - Etes-vous heureux dans votre vie familiale ? (A)
 - Etes-vous heureux en général ? (B)

- Lorsque les questions sont posées dans ce sens-là, le répondant pensera que « en général » ne porte plus sur la vie familiale. Celui qui se sent malheureux dans sa vie familiale aura donc plus tendance à se dire « heureux en général » dans le sens (A)-(B) que dans le sens (B)-(A)

- **Exemple 4** : un effet d'exclusion typique
 - Pensez-vous que les jeunes d'aujourd'hui conduisent moins bien qu'avant ? (A)
 - Pensez-vous que l'on conduit aujourd'hui moins bien qu'avant ? (B)
 - Dans le sens (A)-(B), 27 % pensent que l'on conduit moins bien qu'avant en général
 - Dans le sens (B)-(A) 34 % pensent que l'on conduit moins bien qu'avant en général
 - Pourquoi ? Dans le sens (A)-(B) la question B est implicitement comprise comme « en dehors des jeunes » (les répondants considérant généralement que ce sont les jeunes qui conduisent moins bien qu'avant)

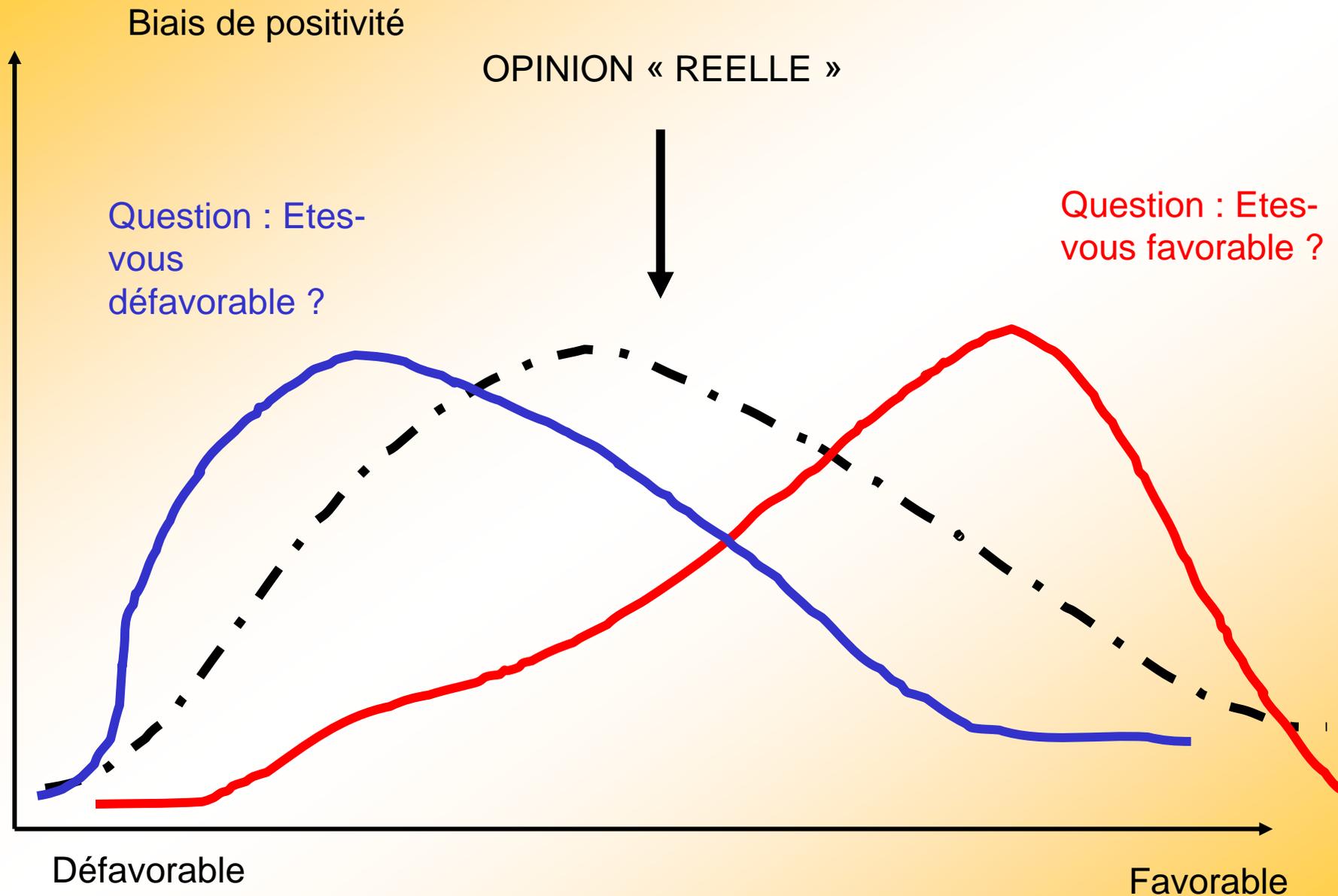
- **Qu'est-ce qui produit les effets d'amorçage et d'exclusion ?**
 - C'est la séquence **question spécifique** ⇒ **question générale**
 - L'interviewé utilise la question spécifique comme indice pour interpréter la question générale sur le même thème
 - **Sauf cas particulier, il vaut donc toujours mieux exclure une séquence question spécifique ⇒ question générale (voir la séquence quintamensionnelle de Gallup)**

→ 7) Le biais de positivité

- Le biais de positivité est sans doute le plus simple et le plus direct des effets de « framing ». Il désigne simplement le fait, que, toutes autres choses égales, il est plus facile de répondre positivement que négativement à une question, quelle que soit la question.

- Autrement dit, si l'on suppose une distribution a priori des opinions en faveur d'un objet X, le biais de positivité impliquera
 - Que l'on sur-estimera les opinions favorables à X si elles s'expriment par une réponse positive (oui, d'accord...)
 - Que l'on sous-estimera les opinions favorables à X si elles s'expriment par une réponse négative (non, pas d'accord)

- On peut illustrer le biais de positivité en comparant la distribution des réponses aux deux questions suivantes
 - Etes-vous favorable à une taxe sur les 4 x 4 ?
 - Etes-vous défavorable à une taxe sur les 4 x 4 ?



- **Interprétation du graphique :**
- si on suppose une distribution « réelle » des opinions sous la forme d'une courbe de Gauss
 - Poser la question sous forme **positive** conduit à déplacer la moyenne vers les positions favorables
 - Poser la question sous forme **négative** conduit à déplacer la moyenne vers les opinions défavorables
- **Prenons un exemple chiffré simple**
- **Ex : opinion « réelle » sur la guerre en Afghanistan : 65 % sont contre, 35 % sont favorables**
 - Supposons un biais de positivité de 10 %
- **Question 1 : êtes-vous favorable à la guerre en Afghanistan ?**
 - Oui : 45 %
 - Non : 55 %
- **Question 2 : êtes-vous opposé à la guerre en Afghanistan ?**
 - Oui : 75 %
 - Non : 25 %

- Comment tenir compte du biais de positivité ?

- **A) On peut formuler une question équilibrée en suggérant successivement deux formulations « contrebalancées » : Ex**
 - **Item 1 : « La guerre en Afghanistan est une nécessité dans la lutte contre le terrorisme mondial »**
 - **Item 2 : « La guerre en Afghanistan est une pure opération de propagande militaire américaine »**

- Avantages
 - **Chaque personne doit exprimer son avis par un acquiescement et un refus : le biais de positivité est diminué.**
 - **Au niveau de l'ensemble de l'échantillon, les biais de positivités s'équilibrent entre les opinions opposées;**

→ B) Surtout : **pratique des échelle bi-polaires**

- On ne se prononce plus en « oui » ou « non »
- Mais on choisit entre deux pôles
- Exemple
 - Comment vous situeriez-vous sur l'échelle suivante :

– 1 2 3 4 5 6 7

Guerre Contre le terrorisme

Propagande militaire

→ 8) Les biais liés à la mémoire

- Les problèmes de mémorisation dans les enquêtes peuvent être de deux types :
 - Ils peuvent concerner la **mémoire à long terme** : difficulté pour le répondant de se rappeler avec précision un certain nombre d'informations qui lui sont demandées concernant une période de temps donnée.
 - Ils peuvent concerner la **mémoire à court terme** : dans certaines conditions, difficulté de suivre jusqu'au bout la totalité de la question en restant concentré :
 - Difficulté avec les longues listes
 - En particulier lorsque la question est entendue et non lue par le répondant

→ **A) PROBLEMES DE MEMOIRE A LONG TERME**

- Les biais liés à mémorisation concernent essentiellement les **questions de fait** et non plus les questions d'opinions. Ces questions portent
 - Sur les événements vécus
 - Sur les comportements effectifs.
- Le répondant peut
 - Eprouver des difficultés se remémorer les événements eux-mêmes
 - Eprouver des difficultés à les situer précisément dans une période de temps donnée
 - Inventer des souvenirs sous la pression de la question.

– Oubli de certains événements (*under-reporting*)

- En règle générale, l'oubli des événements vécus est une fonction exponentielle du temps écoulé depuis qu'ils ont été vécus. Par exemple, une enquête américaine commence par recenser des séjours hospitaliers, puis interroge les personnes concernées. Cela donne le résultat suivant :

– TEMPS ECOULE DEPUIS L'HOSPITALISATION	% D'OUBLI
– 1 à 10 SEM	3 %
– 11 à 20 SEM	6 %
– 21 à 30 SEM	9 %
– 31 à 40 SEM	11 %
– 41 à 50 SEM	16 %
– 51 à 53 SEM	42 %

- L'oubli des événements est également une fonction inverse de leur « saillance » définie par trois dimensions
 - Le caractère inhabituel
 - L'association avec un coût ou une rétribution élevés
 - La continuité des conséquences

– Le télescopage (over-reporting)

- Interrogées sur des événements hautement « saillants », les personnes interrogées peuvent rapporter des événements qui ne se situent pas dans la période indiquée par l'enquêteur
 - **Par exemple, si l'enquêteur demande « Au cours des deux dernières années, avez-vous subi une intervention chirurgicale importante »**
 - **L'enquêteur peut avoir tendance à inclure des interventions qui se sont passées il y a deux ans et demi**
 - **Statistiquement, cela peut produire une surestimation du nombre réel d'hospitalisations subis par les gens**

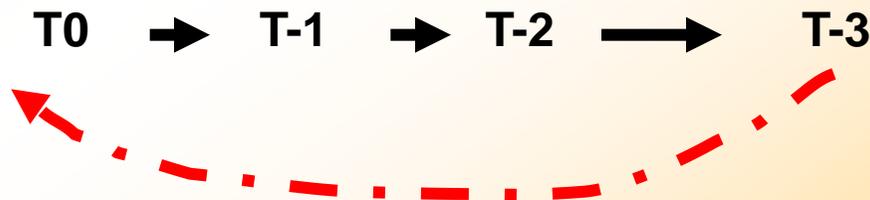
– Les faux souvenirs

- Exemple : après la vision d'un film, les sujets de l'enquête se voient poser la question : « A quelle vitesse roulait la voiture lorsqu'elle a dépassé le tracteur » ?
- Bien que la voiture n'ait dépassé aucun tracteur, 17 % des enquêtés ont évoqué l'épisode

- **Le biais de désirabilité sociale (voir plus bas)**
 - Les effets de la mémoire sont influencés aussi par le biais de désirabilité sociale :
 - On va avoir tendance (inconsciemment) à se rappeler davantage les comportements valorisants que les comportement moins valorisants
 - Ainsi, les voyages d'affaires, ou les rencontres gratifiantes, vont être surestimés, par exemple (over-reporting)
 - A l'inverse, les épisodes moins valorisants (maladies, difficultés conjugales, accidents) vont avoir tendance à être sous-évalué (under-reporting)
 - De la même façon, les comportements habituels valorisants (lecture, activités sportives) vont être être surestimés et les comportements habituels moins valorisants (quantités d'alcool ou de tabac quotidiennes, par exemple) vont être sous-évalué
 - Attention : il s'agit bien ici d'évoquer un phénomène inconscient de remémoration, indépendamment du fait que les personnes interrogées peuvent aussi délibérément mentir sur les comportements réels.

- **Comment contrôler les effets de mémoire à long terme ?**
 - **a) La solution la plus efficace consiste à multiplier les prises d'information dans le temps**
 - En procédant à des panels
 - En encourageant les enquêtés à tenir eux-mêmes un journal (voir les premières enquêtes de consommation TV, par exemple)
 - Ces méthodes permettent à l'enquêteur de poser des balises temporelles beaucoup plus précises. Par exemple « Depuis que je vous ai vu la dernière fois, avez-vous été hospitalité ? »
 - Mais bien sûr, elles ne sont disponibles que dans des cas limités
 - **b) Une deuxième stratégie consiste à tenter d'amener les effets d'oubli et de télescopage à s'annuler statistiquement**
 - Puisque l'un produit une sous-estimation des événements et l'autre une surestimation, il faut interroger sur des événements remontant à une période où les deux sont aussi fréquents l'un que l'autre
 - Pour les comportements « usuels », la période idéale est généralement estimée à entre deux semaines et un mois
 - Pour les événements plus inhabituels, il existe encore peu de recherches concluantes

- **c) Une troisième stratégie : procéder par ordre chronologique inverse**
 - La succession des questions dans l'ordre chronologique permet d'aider à la remémoration
 - Dans cette logique, chaque élément remémoré sert de « jalon » pour aider à la remémoration des autres
 - L'ordre **chronologique inverse** est en général préférable parce que l'on a moins de risque de manquer un jalon
 - Exemple : une enquête a montré que si l'on demande à des élèves de se souvenir de leurs professeurs, ils y arrivent mieux en partant des professeurs récents pour remonter que dans le sens inverse
 - Parce qu'il y a plus de chance de rencontrer un « trou » dans le passé lointain



- **d) Quatrième stratégie : multiplier les dimensions d’interrogation**
 - Chaque événement du passé peut s’appréhender sous diverses dimensions
 - Par exemple, un séjour à l’hôpital peut impliquer des souvenirs relatifs
 - À la maladie elle-même
 - Au coût que cela a impliqué
 - Aux relations avec le personnel
 - A la vie à l’hôpital, etc...
 - **Multiplier les angles d’approche peut aider l’interviewé à reconstruire un souvenir plus précis**
 - **Exemple : les enquêtes sur la violence subie ou sur les discriminations**

- e) cinquième stratégie : le cadrage de l'interviewé
 - L'expérience a montré qu'une stratégie efficace consiste simplement à **insister explicitement** sur l'importance d'une remémoration exacte
 - Attirer l'attention de l'interviewé
 - L'encourager au long de l'interview
 - Essayer d'obtenir un accord explicitement formulé de l'interviewé sur l'importance d'un souvenir précis et exact
 - Ces consignes simples peuvent se combiner aisément avec les autres stratégies. Elles amènent la personne interrogée à se « concentrer » effectivement sur la tâche

- **Problèmes d'analyse liés à l'échantillonnage temporel**
 - Indépendamment des stratégies « d'optimisation » du souvenir, il est important de noter que « l'échantillonnage temporel » des comportements modifie le sens (et parfois les possibilités) de l'analyse.
 - Par exemple, si l'on interroge les individus sur les **habitudes culturelles**, on obtient des « patterns » de corrélation inverses selon l'échantillonnage
 - Si l'on demande aux personnes interrogées, d'une part combien de fois ils sont allés au cinéma **la semaine précédent l'enquête**, et, d'autre part, combien de fois ils sont allés au théâtre, on obtiendra sur l'ensemble de la population une **corrélation négative** : plus on est allé au cinéma la semaine précédent l'enquête, et moins on a de chance d'être allé aussi au théâtre (« effet d'éviction »)
 - A l'inverse, si on leur pose la même question sur **l'année précédent l'enquête**, on a de fortes chances d'obtenir une corrélation positive : parce que les gens qui vont en moyenne le plus souvent au cinéma sont les mêmes que ceux qui vont le plus souvent au théâtre (effet de statut social).

- Il en est de même si on interroge sur certaines **habitudes alimentaires** :
- Sur le long terme, la consommation de bœuf et celle de veau sont généralement positivement corrélées (habitudes alimentaires « carnées »)
- Sur le court terme, elles sont négativement corrélées: la semaine où l'on a mangé plusieurs fois du bœuf, il y a peu de chances qu'on ait mangé plusieurs fois du veau

→ B) PROBLEMES DE MEMOIRE A COURT TERME

- La mémoire à court terme concerne la capacité que nous avons de nous souvenir de tous les éléments d'une question ou d'une liste de réponse lorsque la question ou la liste sont relativement longues
- **Effet de primauté (*primacy effect*)**
 - Lorsque l'on **voit** une longue liste de réponses possibles, on a tendance à se souvenir davantage des premières réponses et à se focaliser sur elles
 - Cet effet vient se combiner avec l'effet de halo (voir plus haut) : les premiers items servent à « ancrer » notre représentation de ce qu'est une réponse « normale » et viennent modifier la signification des items suivants
 - Une façon de réduire l'effet de primauté est de varier l'ordre des items à travers les questionnaires. Ainsi, on peut supposer que statistiquement, les différents effets de primauté seront distribués de manière différentes selon les questionnaires et tendront à s'annuler.
 - Certains dispositifs graphiques sont parfois adoptés : mettre l'ensemble des réponses possibles en cercle (principe de la « table ronde »)

– Effet de récence (*recency effect*)

- Lorsque l'on **entend** une longue liste de réponses possibles, la tendance est inverse : les derniers items entendus restent plus longtemps en mémoire et ont donc tendance à « dominer » les autres, indépendamment de « l'opinion réelle » de l'interviewé.
- Une des **solutions partielles** à cette difficulté consiste également à varier l'ordre de lecture des items au travers des interviews
- Une **autre solution** peut être de multiplier les formes de prise de connaissance de la liste:
 - En même temps que la liste est lue, par exemple, la personne interrogée peut prendre en prendre connaissance de manière visuelle
 - Cette méthode est parfois discutée : elle permet au répondant de « se balader » parmi les différentes questions, induisant parfois ainsi d'autres biais cognitifs non maîtrisés par l'enquêteur.

→ 3. LES BIAIS MOTIVATIONNELS

→ 1) Le biais « d'engagement »

- La psychologie sociale a bien mis en évidence la tendance que nous avons à persévérer, par souci de cohérence, dans une séquence d'action engagée de manière contingente : c'est la théorie de **l'engagement**.
- De manière assez similaire, en termes d'opinions, une séquence « syllogistique » peut ainsi amener les répondants, par souci de ne pas se contredire, à endosser des opinions qu'ils ne défendraient pas spontanément.
- L'exemple suivant, issu en fait d'un roman, montre comment il est possible d'induire la même personne par une suite de questions aux réponses assez évidentes, à prendre position.

– Séquence 1

- Q : La délinquance chez les jeunes vous tracasse-t-elle ?
- R : Oui
- Q : Pensez-vous qu'il y a un manque de discipline dans nos écoles secondaires ?
- R : Oui
- Q : Seriez-vous **favorable** à une réintroduction du service militaire ?
- R : Oui

– Séquence 2 (même personne)

- Q : Etes-vous tracassé par les risques de guerre aujourd'hui ?
- R : Oui
- Q : Pensez-vous qu'il est dangereux de distribuer des armes aux jeunes et de leur apprendre à tuer ?
- R : Oui
- Q : Pensez-vous qu'il est mal de forcer des gens à prendre les armes contre leur volonté ?
- R : Oui
- Q : Vous **opposeriez** vous à la réintroduction du service militaire ?
- R : Oui

- **Comment l'enquêteur doit-il se comporter vis-à-vis de ce biais ?**
 - Il faut évidemment éviter toute séquence « syllogistique » qui produirait un très lourd effet d'induction
 - Faut-il pour autant faciliter la tâche des répondants qui ne se soucient pas de la cohérence de leurs réponses ?
 - Cela dépendra sans doute partiellement de ce que l'on cherche :
 - Des « opinions » les plus libres possibles pour mettre en évidence en suite d'éventuelles « structures » d'attitudes (analyses factorielles)
 - Ou bien vérifier comment l'enquêté se comporte par rapport à des situations de « choix sous contrainte »
- L'important est que, dans un cas comme dans l'autre, l'enquêteur soit conscient du fait **que le cadre de référence qu'il aura donné induira partiellement la nature des réponses** (organisées ou non, élimination des incohérences ou non).

→ 2) Le biais de « désirabilité sociale »

- Le biais de désirabilité sociale est un des effets les plus connus des méthodologues
- Il indique que la personne interviewée cherchera, spontanément, à donner d'elle-même une image « positive » à l'enquêteur

- **Les questions de fait et de comportement**
 - On a vu plus haut que le biais de désirabilité sociale peut contribuer à produire (sans ironie) des effets de « mémoire sélective » :
 - on surévaluera les événements vécus ou les comportements qui paraissent valorisants
 - On sous-évaluera à l'inverse ce qui paraît moins valorisant
 - Certains spécialistes ont listé ce qui, dans les différentes enquêtes apparaît comme socialement valorisé et socialement dévalorisé :

 - **Valorisé :**
 - Le registre de la « **citoyenneté** » : voter aux élections, exercer un rôle dans les affaires de la cité, être informé des problèmes politiques et sociaux
 - Le registre de la « **culture** » : lecture des journaux et de livres, fréquenter les concerts, le théâtre et les musées, participer à des formations
 - Le registre de la « **morale** » : avoir un travail, manifester de l'intérêt pour les défavorisés, aider ses amis et sa famille

- **Dévalorisé :**
 - Le registre des **maladies et incapacités** : cancer, maladies vénériennes, maladies mentales, dépression
 - Le registre des **comportements « socialement mal vus »** : tout ce qui est contraire à la loi, la consommation d'alcool, l'évasion fiscale, les pratiques sexuelles
 - Le **statut financier** : le revenu, l'épargne, les propriétés.

- On remarquera en passant que certains comportements seront sous-évalués non pas simplement par biais de désirabilité sociale mais parce que la livraison de certaines informations à l'enquêteur peut-être perçue comme menaçante : si l'anonymat n'est pas respecté l'interviewé pourrait avoir à regretter certaines informations privées qu'il aurait livrées.

- **Comment réduire le biais de désirabilité sociale ?**
 - Quelques **éléments généraux** peuvent être mis en avant pour tenter de minimiser son impact, mais ils ont trait à la procédure de passation plutôt qu'à la forme des questions elles-mêmes.
 - Ne pas mettre les enquêteurs dans des situations trop difficiles à gérer avec les enquêtés (par exemple, sur des questions délicates comme les pratiques sexuelles, choisir des enquêteurs du même sexe que les répondants, autant que possible)
 - Donner le moins d'indices possibles à travers la question, les échelles de réponse, ou l'attitude de l'enquêteur lui-même de ce que l'enquêteur trouve « normal »
 - Ne pas demander plus de détails que nécessaires pour les objectifs de l'enquête, lorsqu'il s'agit de comportements « dévalorisés ».

- Un certain nombre de **techniques spécifiques** peuvent aider dans certaines circonstances
 - La technique de la **porte dans la figure** :
 - On commence par poser une question « indiscrete », par exemple : pouvez-vous me dire combien vous gagnez en net par mois ?
 - Lorsque les gens ont refusé de répondre, on pose une question moins précise, par exemple : « alors, pourriez-vous au moins vous situer parmi les tranches de revenu qui sont indiquées ? »
 - Le sujet, ayant déjà refusé une fois, sera plus enclin à répondre à la deuxième question
 - Inconvénient : la technique peut indisposer la personne interviewée
 - La réponse **indirecte** :
 - Plutôt que de demander aux gens « avez-vous déjà fraudé le fisc »
 - On leur demandera : à votre avis, y a-t-il beaucoup de gens qui fraudent le fisc ?
 - Et on extrapole la réponse de l'individu à partir de ce qu'il dit sur les autres
 - Inconvénient : il devient difficile de faire la part de ce qui est simple représentation sociale

- Dans le cas où la question est **particulièrement menaçante** , on peut inventer des dispositifs d'anonymat
 - Exemple: la méthode de la **réponse aléatoire**
 - La personne interrogée devra répondre soit à une question « menaçante », soit à une question « anodine » et sans rapport
 - La sélection de la question anodine ou menaçante ou anodine se fait par tirage au sort et l'enquêteur ne sait pas quel est le résultat du tirage (Mais il connaît la probabilité de tirer la question « cible ». Exemple : l'enquêté tire à « pile ou face »)
 - La question anodine est calibrée de telle façon que l'enquêteur puisse déterminer le taux probable de réponses positives
 - A partir de là, il peut déduire le taux probable de réponses positives à la question menaçante
 - Illustrons la méthode de la réponse aléatoire
 - Question mençante : avez-vous déjà consommé des drogues dures ? (Oui/non)
 - Question anodine : votre mère est-elle née en juin ? (Oui/non)
 - L'interviewé tire à pile ou face la question à laquelle il répondra
 - L'enquêteur ne connaît pas le résultat du tirage

- Comment cela marche-t-il ?
 - Supposons que 20 % des sujets répondent « oui ». Mais oui à quelle question ?
 - L'enquêteur sait que, statistiquement, 50 % des sujets auront tiré « votre mère est-elle née en juin ? »
 - L'enquêteur sait que, parmi elles, une sur 12 (soit environ 4 %) auront répondu « oui »
 - Il reste donc 16 % de « oui » à la question « Avez-vous déjà consommé des drogues dures »
 - Comme 50 % des sujets ont répondu à cette question, cela représente 32 % de personnes qui ont déjà consommé des drogues dures.

- Remarque : cette méthode repose évidemment sur l'idée que l'aléatoire a bien fonctionné à chaque étape
 - Mais on a remarqué que sur des sujets très sensibles, la méthode de la réponse aléatoire donne de meilleurs résultats que le questionnement direct parce que les répondants mentent trop souvent.

	Pourcentage total	Réponses estimées à la question sans rapport	Réponses estimées à la question cible	Estimation du comportement cible
	Total	Votre mère est-elle née en juin ?	Avez-vous déjà pris de l'héroïne ?	Pourcentage de ceux qui ont répondu à la question cible
Oui	20	4	16	32
Non	80	46	34	68
% total	100	50	50	100

→ 4. CONSTRUCTION DU QUESTIONNAIRE : SCHEMA GENERAL

→ 1) Plan général du questionnaire

- Sélection des différentes thématiques
- Sélection des informations « latérales » importantes (variables de cadrage sociologique)
- Mise en séquence des différents blocs de question
 - **Les questions « menaçantes » doivent si possible venir à la fin (ex : revenus)**
 - Eviter de sauter d'une thématique à l'autre
 - Attention à l'effet d'ordre des thématiques

→ 2. La construction des blocs de question

- **A) questions ouvertes ou questions fermées ?**
- **Les questions ouvertes**
 - Permettent les réponses inattendues
 - Evitent la monotonie du long questionnaire fermé
 - Renforcent le sentiment d'être écouté chez l'interviewé
 - Autorisent un retravail lors du codage
- **Les questions fermées**
 - Donnent un cadre de référence (exemple : une liste des avantages/inconvénients d'une voiture)
 - Homogénéisent le degré de détail (ex : enquêtes sur la consommation)
 - Evitent les oublis
 - Evitent le travail (très lourd) de postcodage
- **Dans tout questionnaire quantitatif, limiter l'usage des questions ouvertes est une nécessité pratique.**

- **B) Comment séquencer les questions dans un bloc ?**

- **1° Passer du général au spécifique**
 - Une question trop spécifique va influencer la manière dont la question générale est perçue
 - Exemple : si on veut des questions sur les valeurs en général et si on commence par une question sur les valeurs religieuses, les questions suivantes seront interprétées en termes de religiosité
 - Si on veut des questions sur les pratiques politiques et qu'on commence par une question sur le vote, les pratiques politiques seront perçues comme se limitant au champ électoral

- **2° Passer des questions ouvertes aux questions fermées**
 - Pour la même raison : la question fermée va « réduire » l'interprétation de la question ouverte
 - Exemple à éviter
 - Accepteriez-vous un homosexuel comme voisin ? OUI – NON
 - Si vous deviez donner trois mots pour dire ce que l'homosexualité représente pour vous, quels mots choisiriez-vous ? (question ouverte)

→ 3. La formulation des questions

- La formulation des questions d'enquête pose deux grands types de difficulté, une fois les objectifs du questionnaire et sa structure clairement définis
 - **Adopter le langage adéquat :**
 - Eviter les pièges grammaticaux
 - Utiliser les mots adéquats par rapport à la population visée
 - Eviter toute ambiguïté
 - **Eviter au maximum les biais de réponse**
 - Biais cognitifs
 - Biais motivationnels
 - **Prendre conscience du cadrage dans l'interprétation des réponses.**

→ 4) La mise en cohérence

- Les différents blocs sont-ils en ordre logique ?
- Les transitions entre thématiques sont-elles claires ?
- N’y a-t-il pas des questions « menaçantes » qui viennent trop tôt dans le questionnaire ?

→ 5) Le prétest

- Vérifier comment les formulations « passent » auprès des sujets
- Vérifier que les éléments de référence sont compris comme prévus
- Vérifier quel type de public peut avoir des difficultés
- Tester la durée de passation
- Corriger le questionnaire en conséquence

→ **NB : le prétest ne sera pas détaillé au cours théorique. Mais il faut le considérer comme une étape indispensable de la validation du questionnaire.**

→ 5. QUELQUES ELEMENTS DE CONCLUSION

→ 1) Principaux points généraux à garder à l'esprit dans la formulation

- Complexité et adaptation du vocabulaire
- Complexité de la structure de la phrase
- Possibilité de présupposés
- Questions à double détente
- Possibilités de non réponse voire de refus de répondre
- Usage ou non de filtres
- Echelles de réponses : modalités exhaustives et mutuellement exclusives

→ 2) Les principaux biais à tenter d'éviter

- Biais de positivité
- Biais de désirabilité sociale
- Effets de halo
- Effets de contraste
- Effets des échelles de réponse
- Problèmes de mémorisation

→ 3) De manière générale : méthodologie négative

- Le plus souvent, il s'agira moins de déterminer une liste d'opérations à faire qu'une liste d'erreurs à éviter
- La plupart des sources de biais possibles dans l'expression des opinions ont été recensées par la littérature
- Mais beaucoup restent discutées
- En particulier : la notion même « d'opinion publique » reste controversée (voir dernière partie du cours)
- Néanmoins, on peut s'accorder sur un schéma général

→ 4. La nature du débat dans l'espace public

- S'agissant de questions d'opinion, il faut parfois accepter une formulation « non orthodoxe » si elle correspond à la manière dont le débat public est posé.
- Exemple : êtes-vous d'accord avec la phrase suivante :
 - « Les dysfonctionnements judiciaires en Belgique ont empêché la découverte de réseaux de pédophilie »
 - La question **présuppose** qu'il y a eu des dysfonctionnements
 - On ne sait pas bien sur quoi porte la question : les dysfonctionnements ou les réseaux ?
 - Que va répondre quelqu'un qui pense qu'il y a eu des dysfonctionnements mais pas de réseau ?
 - **MAIS : il peut être utile de poser la question ainsi parce que c'est ainsi que les gens sont habitués à l'entendre.**

→ 5. Le souci de comparabilité

- Exemple : dans quelle classe sociale vous situez-vous ?
 - Classe ouvrière
 - Classe moyenne inférieure
 - Classe moyenne
 - Classe moyenne supérieure
 - Classe supérieure

- La question
 - Mélange des conceptualisations différentes (marxiste/ libérale)
 - Sur-représentation des catégories moyennes

- MAIS : c'est devenu une formulation « classique » donc il est difficile de l'éviter si on veut pouvoir comparer avec d'autres enquêtes.

TECHNIQUES D'ECHANTILLONNAGE

La validité externe

→ 1. NOTIONS GENERALES

- **1^{ère} étape : définition de la population**
- **La population** à étudier va dépendre de la question qu'on se pose.
- **Exemples**
 - **Marketing de parfums : on ciblera les femmes**
 - Sauf si l'on pense que le choix est plutôt fait par le partenaire masculin
 - **Enquête sur les victimes des discriminations racistes : on interrogera les publics « cibles »**
 - **Enquête sur les intentions de vote : on interrogera les électeurs (belge de 18 ans au moins)**
 - **Enquête sur les conditions de vie des étudiants universitaires, etc.**

- Une fois la population définie ses **Limites** peuvent être difficiles à fixer :
 - Les immigrés : quelles nationalités ?
 - Si on cible les jeunes : quel âge ?
 - Qu'est-ce que le public « cible » des discriminations ?
 - Tous les étrangers ?
 - Seulement les étrangers de certaines régions du monde ?
 - Toutes les minorités visibles même belges ?

- Attention : en théorie des sondages « **population** » n'a donc pas le même sens que dans la vie courante

- **Enquête exhaustive** : on a la possibilité d'interroger la **totalité de la population visée**
- **Avantage** : on connaît les « valeurs réelles » de la variable qui nous intéresse
- **Inconvénients** :
 - c'est souvent **impossible** (exemple : test de lampes électriques)
 - Le **coût** est presque toujours exorbitant.
- **Attention** : lorsqu'on fait une enquête exhaustive, la théorie de l'échantillonnage ne s'applique évidemment plus. Elle n'a plus de raison d'être

- **Échantillon** : sous-ensemble de la population à partir de laquelle on tente d'inférer des mesures sur la population elle-même.
- **Erreur d'échantillonnage** : toute estimation à partir d'un échantillon s'écarte dans une certaine mesure des vraies valeurs de la population.
- Même le meilleur échantillon donne donc toujours une réponse à l'intérieur d'un « **intervalle de confiance** »

- **Techniques d'échantillonnage** : ensemble des méthodes destinées à sélectionner un échantillon qui, **en fonction des contraintes de coût et des contraintes techniques**, minimise l'erreur d'échantillonnage.
- En d'autres mots, obtenir l'erreur d'échantillonnage la plus petite possible en fonction des moyens dont on dispose
- L'erreur d'échantillonnage s'apparente donc à l'erreur de mesure : si on mesure vingt fois la même distance entre deux villes, les mesures ne seront pas identiques au millimètre près.

→ Synthèse du problème de l'échantillonnage

- Je dois obtenir des informations sur une **population** donnée :
exemple, je cherche à savoir si les Liégeois sont satisfaits de la gestion communale
- Je n'ai pas la possibilité d'interroger la **totalité** de la population
 - Soit parce que cela prendrait trop de temps
 - Soit parce que ce serait trop coûteux
 - Soit parce que je ne la connais pas tout entière (n'y a-t-il pas un certain nombre de clandestins, par ex ?)
 - Etc.

- Comment extraire un **sous-ensemble de cette population**
 - Qui me donne une **approximation satisfaisante** des informations que je cherche à obtenir
 - Compte tenu des **moyens** dont je dispose
- Remarque : la question du « bon » échantillon est donc toujours un rapport entre la **qualité de l'information et le coût pour l'obtenir**. Il n'y a pas de solution « parfaite » (sauf si je dispose de statistiques exhaustives)

- **Échantillon représentatif** : tous les membres de la population ont **la même probabilité** de figurer dans l'échantillon
 - Ou en tous les cas, ils ont une probabilité connue de figurer dans l'échantillon

- **Échantillon biaisé** : lorsque les membres de la population n'ont pas tous la même probabilité de figurer dans l'échantillon

→ Exemples de biais :

- j'interroge un échantillon de personnes âgées : les plus « désengagés » auront davantage tendance à ne pas répondre
- J'interroge un échantillon d'électeurs : les électeurs extrême-droite répondront moins que les autres.
- Je mène une enquête sur les habitudes des automobilistes en les repérant aux pompes à essence : la probabilité d'être interrogé est proportionnelle au kilométrage parcouru.
- J'interroge mon public par correspondance : 15 % seulement répondent. Sont-ils représentatifs de l'ensemble ?

- **Fondamental: ne pas confondre l'erreur d'échantillonnage et le biais d'échantillonnage**
- **L'erreur d'échantillonnage** est la marge d'imprécision qui existe dans tous les échantillons **même s'ils sont représentatifs**. Elle contraint à situer toute mesure dans un **intervalle de confiance**
- **Le biais d'échantillonnage** est une erreur systématique qui conduit à **sur-représenter ou sous-représenter certaines catégories de la population dans l'échantillon** ce qui veut dire que mon estimation est faussée (biaisée)

- **Exemple** : je veux un échantillon représentatif de cent personnes issu de la population étudiante de l'université
- Si j'utilise la liste des étudiants et que je tire cent noms selon une procédure aléatoire rigoureuse, il est très peu probable que mon échantillon ait exactement l'âge moyen ou la proportion garçons/filles qui prévaut dans la population : **erreur d'échantillonnage**
 - Si j'interroge les étudiants en passant dans l'auditoire, ceux qui viennent plus souvent au cours seront sur-représentés : **biais systématique** (qui s'ajoute à l'erreur d'échantillonnage).

- **Comment va-t-on présenter alors un résultat ?**
- **Si on pose une question sur les intentions de vote et que l'on obtient 15% d'électeurs CDH**
- **On dira que pour un degré de certitude donné, (en général 95 %) le pourcentage « réel » dans la population (si mon échantillon n'est pas biaisé) est compris entre (par exemple) 13,5 % et 16,5 % : intervalle de confiance**
- **Cet intervalle de confiance est variable pour chaque enquête**
- **Il est indispensable de le mentionner pour prendre la mesure de ce que l'on sait.**

- **Taux de sondage** : rapport entre la taille de l'échantillon et la taille de la population
 - Ex : si on interroge 3.500 personnes en Wallonie, le taux de sondage est de 1 pour mille.
- **Unité de sondage** : unités disponibles pour tirer l'échantillon
- **Unité d'analyse** : unités sur lesquelles porte l'enquête
 - Les deux peuvent différer : par exemple, on dispose d'une liste d'adresses mais on s'intéresse aux individus; ou, à l'inverse, on dispose d'une liste électorale mais on souhaite interroger des jeunes de moins de 18 ans. (voir plus loin l'échantillonnage par **grappes**)

- **Échantillon aléatoire ou probabiliste** : on choisit un certain nombre de personnes au moyen d'une procédure formalisée de tirage au sort
 - Cela implique que l'on dispose d'une liste de la population : une **base de sondage**.
 - Exemple : on peut disposer d'une base de sondage pour tirer un échantillon aléatoire d'étudiants de l'Ulg.

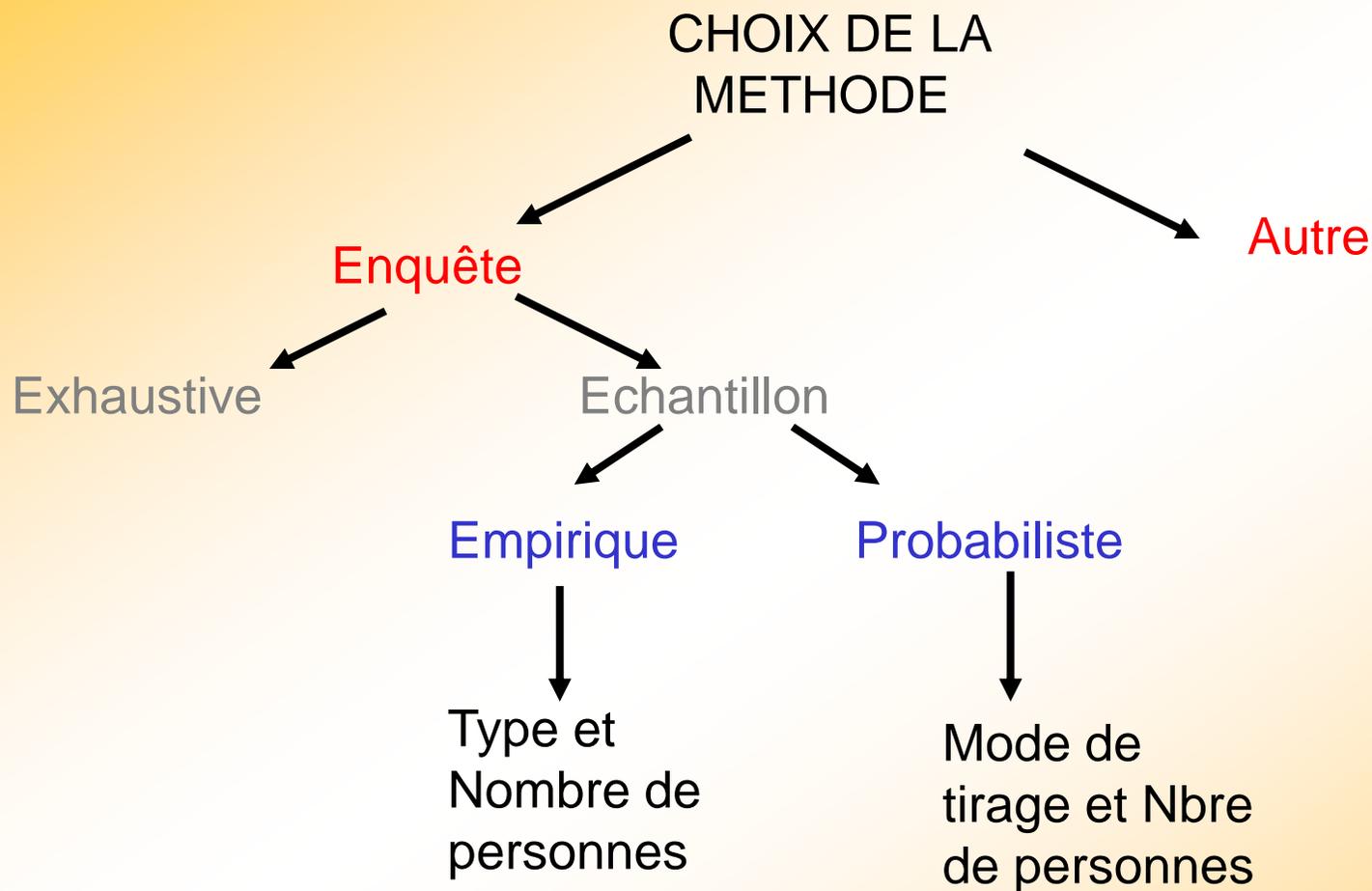
- **Échantillon empirique ou raisonné** : on sélectionne un échantillon de telle façon qu'il constitue une « bonne image » de la population étudiée.
 - Exemple : on ne dispose pas d'une liste de personnes à interroger, mais on veillera à ce que l'échantillon comporte la même proportion d'hommes et de femmes que la population considérée, la même proportion d'actifs et d'inactif, la même structure par âge et la même structure en fonction du niveau d'études (méthode des quotas).

→ **Échantillons probabilistes**

- on peut estimer l'erreur possible d'échantillonnage de manière rigoureuse grâce au calcul des probabilités.
- Mais les conditions sont difficilement remplies et le coût est souvent élevé.
- Les biais sont en général plus facilement contrôlables
- La qualité dépend bien sûr du taux de réponse.

→ **Échantillons empiriques**

- Il n'y a pas de calcul possible sur l'erreur d'échantillonnage, sauf par analogie
- La méthode est souvent moins coûteuse et parfois la seule possible (absence de toute base de sondage)
- Il est plus difficile de contrôler les biais.
- Certains échantillons empiriques « imitent » le hasard (méthode des itinéraires)



Paramètres
du choix :

- Information disponible
- Qualité souhaitée
- Budget

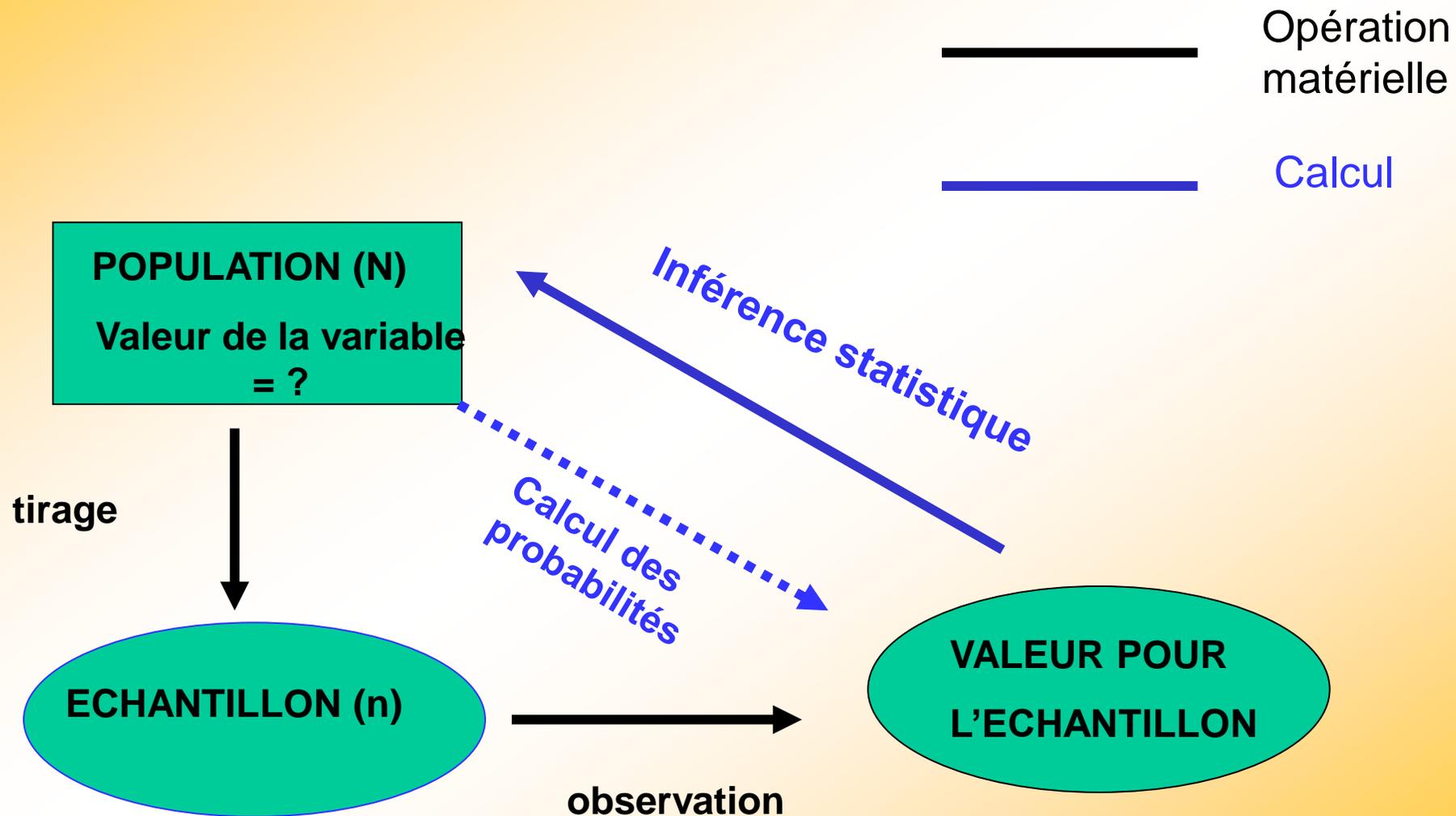
→ 2. L'ECHANTILLONNAGE PROBABILISTE

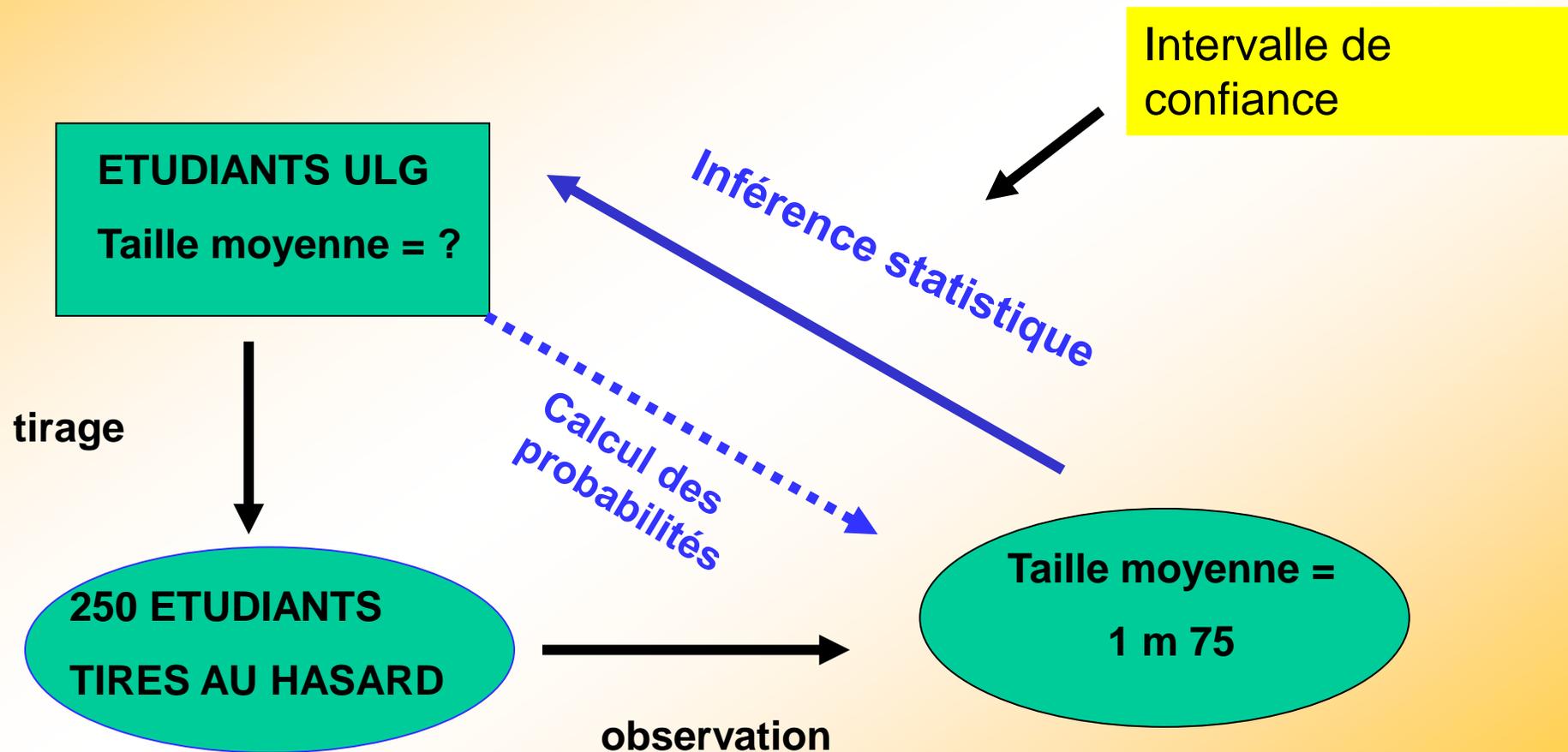
→ Le principe :

- si l'on applique une logique rigoureuse de tirage au sort, alors,
- on peut utiliser le **calcul des probabilités** pour faire des **inférences** précises sur la population.

→ Probabilité et inférence

- Le **calcul des probabilités** nous permet de dire quelle chance on a de tirer un échantillon donné connaissant la population
- **L'inférence statistique** nous permet de raisonner pour savoir quelle est la population la plus « vraisemblable » sachant que l'on a obtenu aléatoirement tel ou tel échantillon





- **Illustrons cette démarche :**
- **Que peut me dire le calcul des probabilités ?**
 - Si ma population est une population de boules blanches et noires dans une urne
 - Pour fixer les idées, on va supposer que j'ai quatre boules, **deux blanches et deux noires**
 - Que peut me dire le calcul des probabilités sur la chance que j'ai de tirer deux boules noires, par tirage successif **sans remise** ?
 - 1^{ère} boule : j'ai une chance sur deux de tirer une noire
 - 2^{ème} boule : j'ai une chance sur trois de tirer encore une noire
 - J'ai donc après deux tirages $\frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$

- **Qu'est-ce que l'inférence ? Le raisonnement symétrique :**
- Je sais que j'ai quatre boules dans l'urne
 - Je ne sais pas combien de blanches et de noires
 - Mais en **tirant un échantillon à l'aveugle, je tombe sur une blanche et une noire**
 - Que puis-je dire sur la population ?

1^{er} cas : population = 2 noires et 2 blanches

Quels sont les tirages possibles ?

1 ^{ère} boule	2 ^{ème} boule	Probabilité
Noire	Noire	$\frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$
Noire	Blanche	$\frac{1}{2} \times \frac{2}{3} = \frac{1}{3}$
Blanche	Noire	$\frac{1}{2} \times \frac{2}{3} = \frac{1}{3}$
Blanche	Blanche	$\frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$

2^{ème} cas : population = 1 noire et 3 blanches

Quels sont les tirages possibles ?

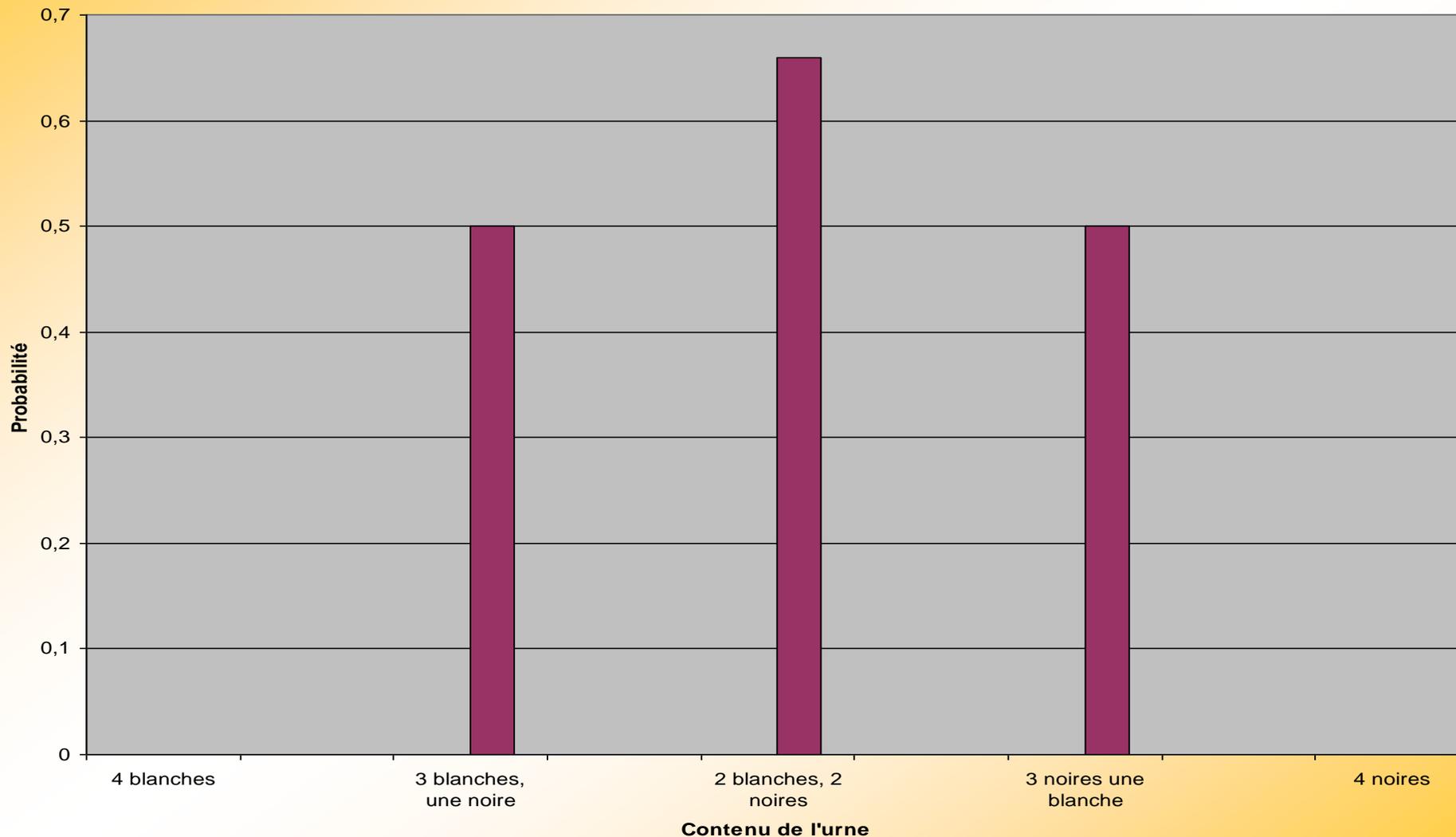
1 ^{ère} boule	2 ^{ème} boule	Probabilité
Noire	Noire	$\frac{1}{4} \times 0 = 0$
Noire	Blanche	$\frac{1}{4} \times 1 = \frac{1}{4}$
Blanche	Noire	$\frac{3}{4} \times \frac{1}{3} = \frac{1}{4}$
Blanche	Blanche	$\frac{3}{4} \times \frac{2}{3} = \frac{1}{2}$

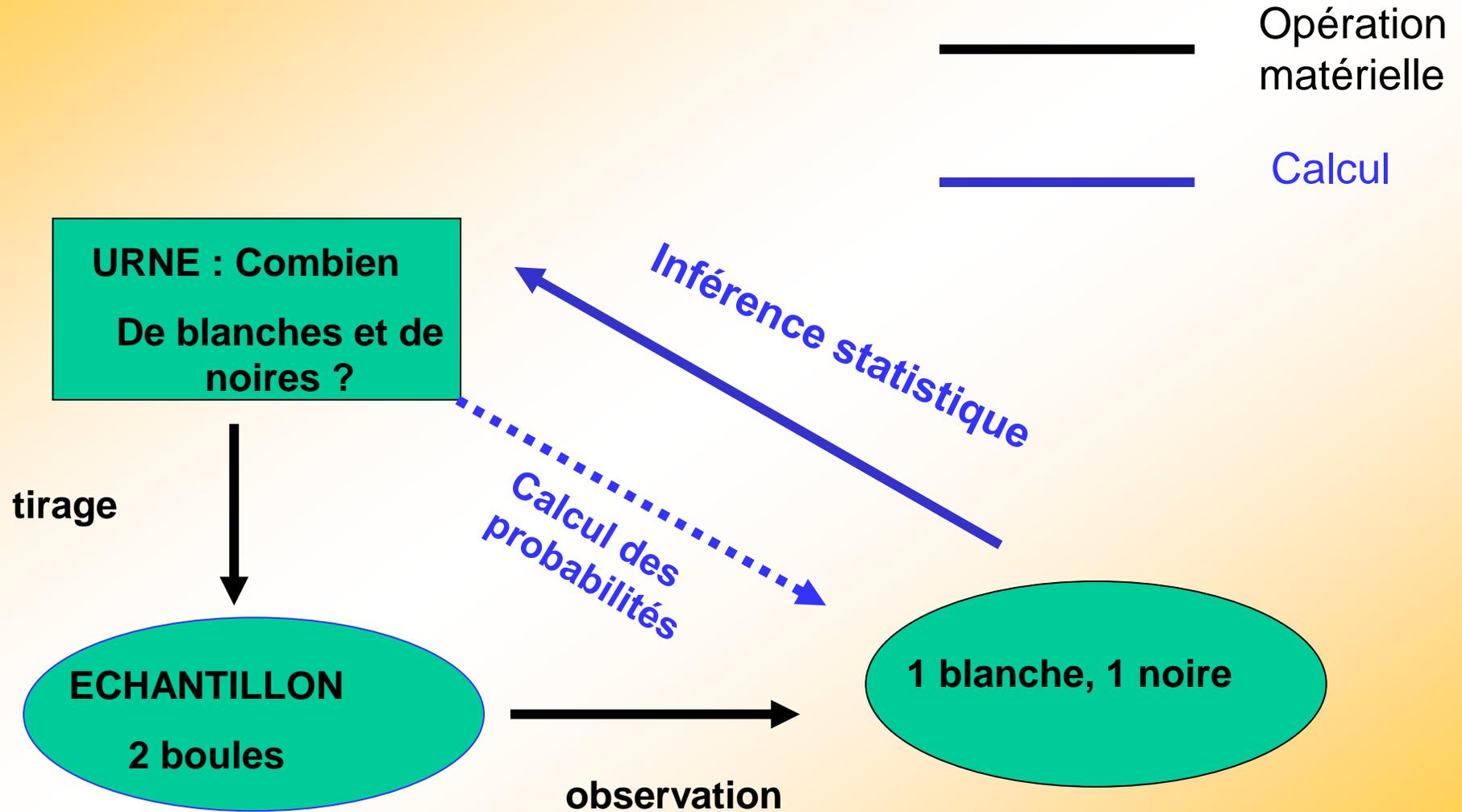
3^{ème} cas : population = 3 noires et 1 blanche

Quels sont les tirages possibles ?

1 ^{ère} boule	2 ^{ème} boule	Probabilité
Noire	Noire	$\frac{3}{4} \times \frac{2}{3} = \frac{1}{2}$
Noire	Blanche	$\frac{3}{4} \times \frac{1}{3} = \frac{1}{4}$
Blanche	Noire	$\frac{1}{4} \times \frac{2}{3} = \frac{1}{4}$
Blanche	Blanche	$\frac{1}{4} \times 0 = 0$

Probabilité d'avoir une blanche et une noire selon la population





- **Base théorique de l'inférence : la loi des grands nombres.**
- **Expression intuitive : plus on répète une épreuve, plus la fréquence relative de chaque résultat s'approche de sa probabilité**
- **Exemple 1 : si une pièce de monnaie est parfaitement équilibrée, il existe un nombre N de lancers de la pièce pour laquelle la différence entre la proportion de piles et de faces est inférieure à ε quel que soit ε .**

- Exemple 2 : si un dé est parfaitement équilibré, il existe un nombre N de lancers du dé pour lequel la différence entre la proportion de « 5 » et $1/6$ est inférieure à ε quel que soit ε
- **Condition : on connaît la probabilité a priori pour chaque face de la pièce de monnaie ($1/2$) et pour chaque face du dé ($1/6$)**

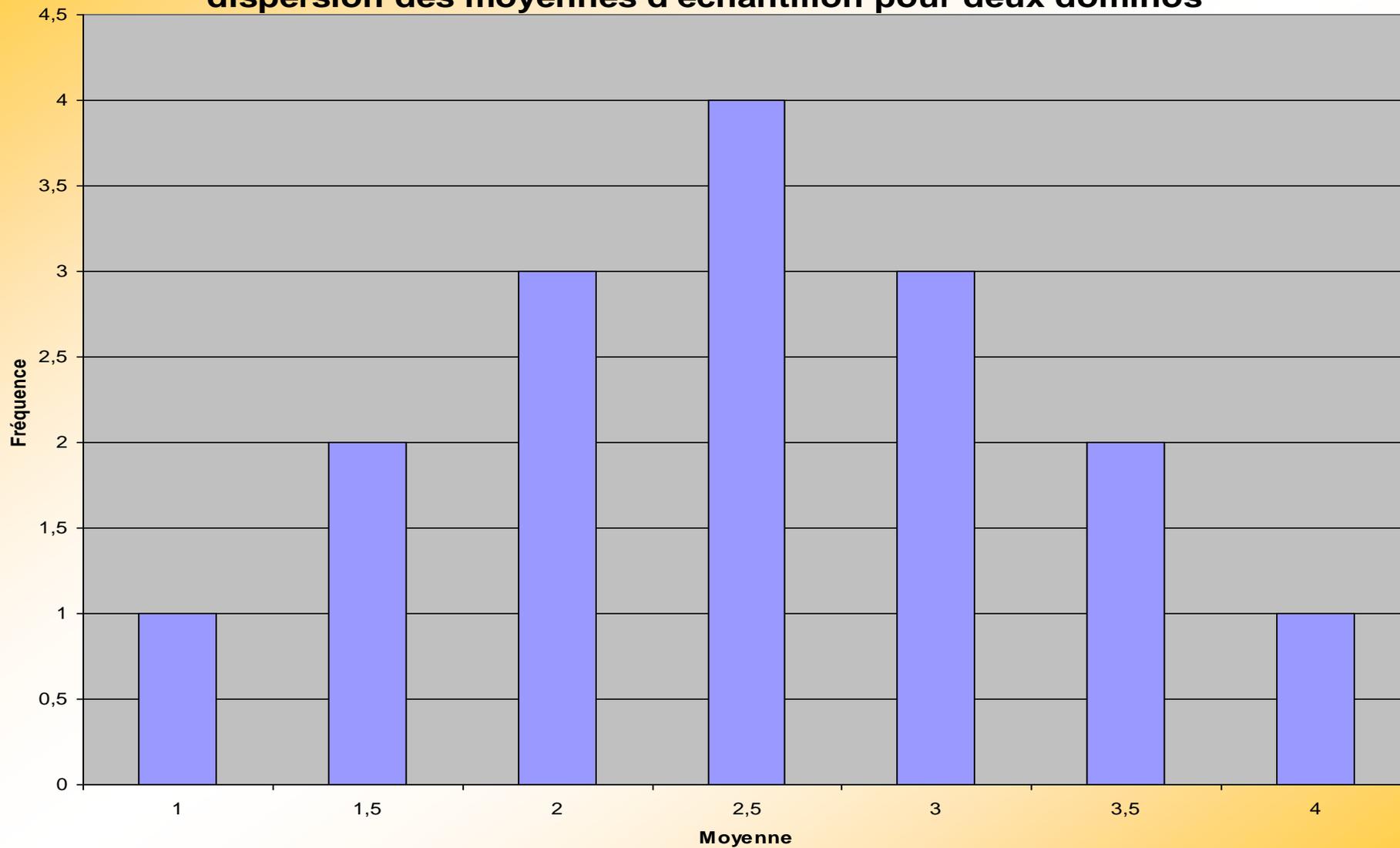
- **Approche statistique intuitive :**
- **Si j'ai quatre dominos marqués de 1 à 4. Somme des quatre dominos est donc $(1 + 2 + 3 + 4) = 10$ et la moyenne = $4/10$ soit 2,5**
- **Une épreuve : je tire un domino, je le replace et j'en tire un deuxième.**
- **Quelles sont les valeurs possibles de mes dominos ?**

E	1 1	1 2	1 3	1 4	2 1	2 2	2 3	2 4	3 1	3 2	3 3	3 4	4 1	4 2	4 3	4 4
Σ	2	3	4	5	3	4	5	6	4	5	6	7	5	6	7	8
M	1	1,5	2	2,5	1,5	2	2,5	3	2	2,5	3	3,5	2,5	3	3,5	4

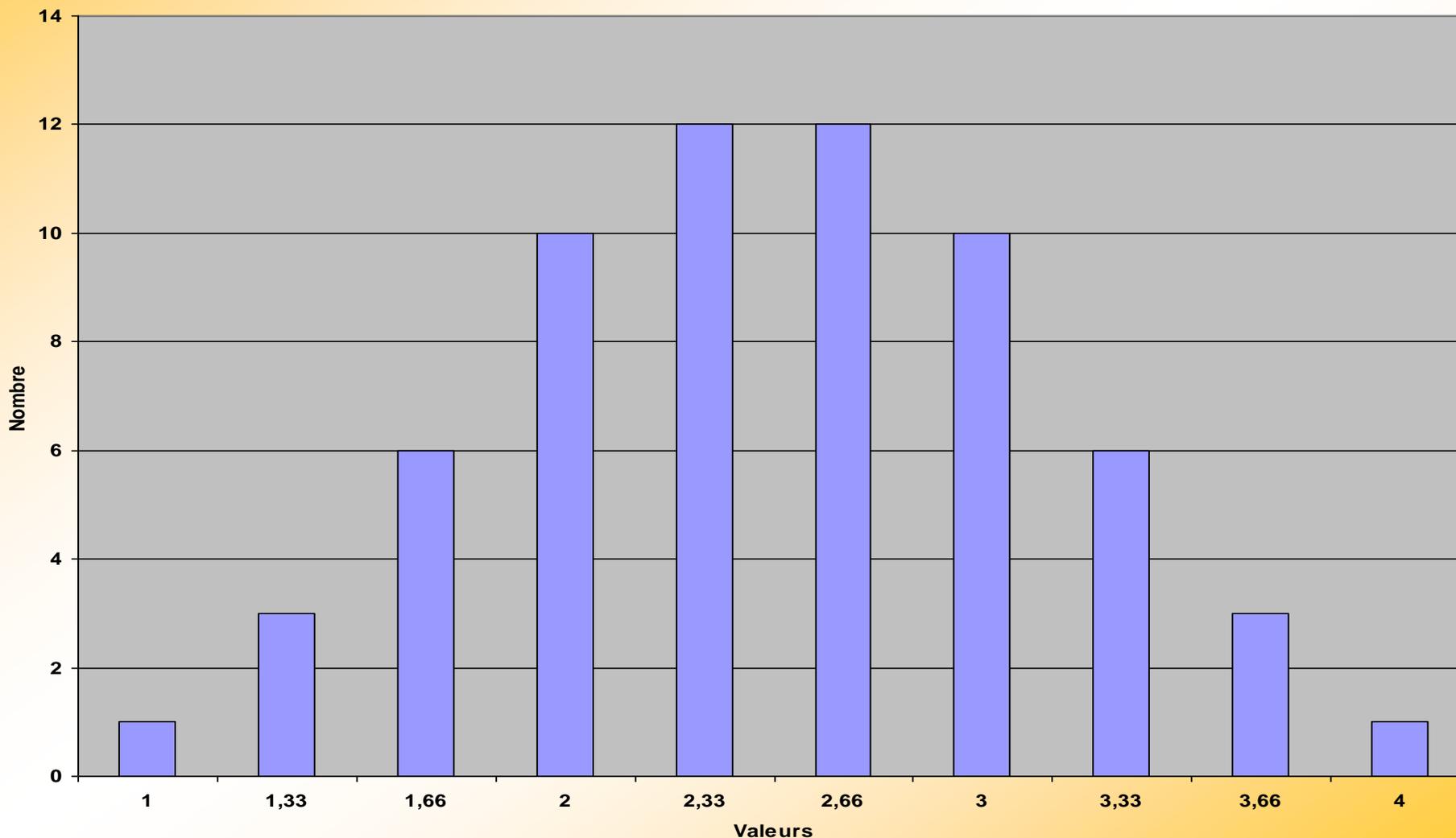
- **Que voit-on ?**
- Que la valeur moyenne **la plus probable** pour l'échantillon est la moyenne de la population : $P(2,5) = 4/16$
- Si on augmente à l'infini le nombre des valeurs, la moyenne des échantillons possibles se distribue comme une courbe de Gauss autour de la moyenne de la population.

2. ECHANTILLONNAGE PROBABILISTE (16)

dispersion des moyennes d'échantillon pour deux dominos



Distribution des valeurs pour trois dominos



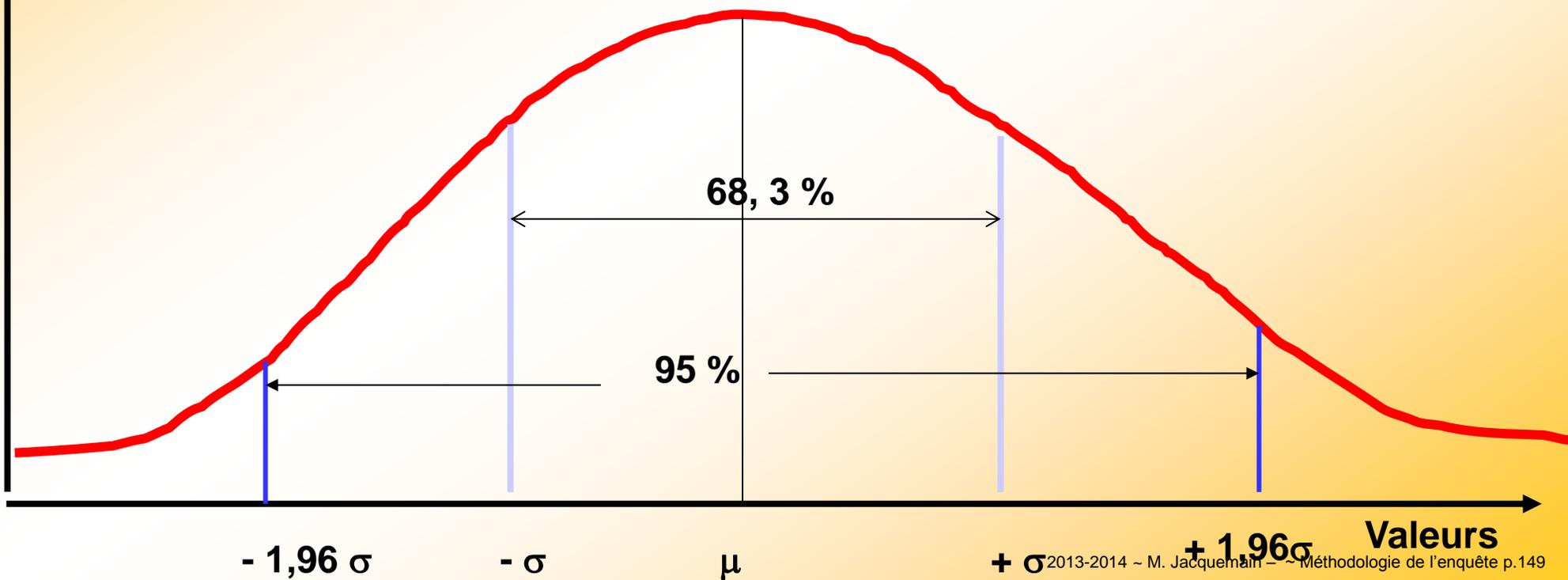
FAITES L'EXERCICE

2. ECHANTILLONNAGE PROBABILISTE (18)

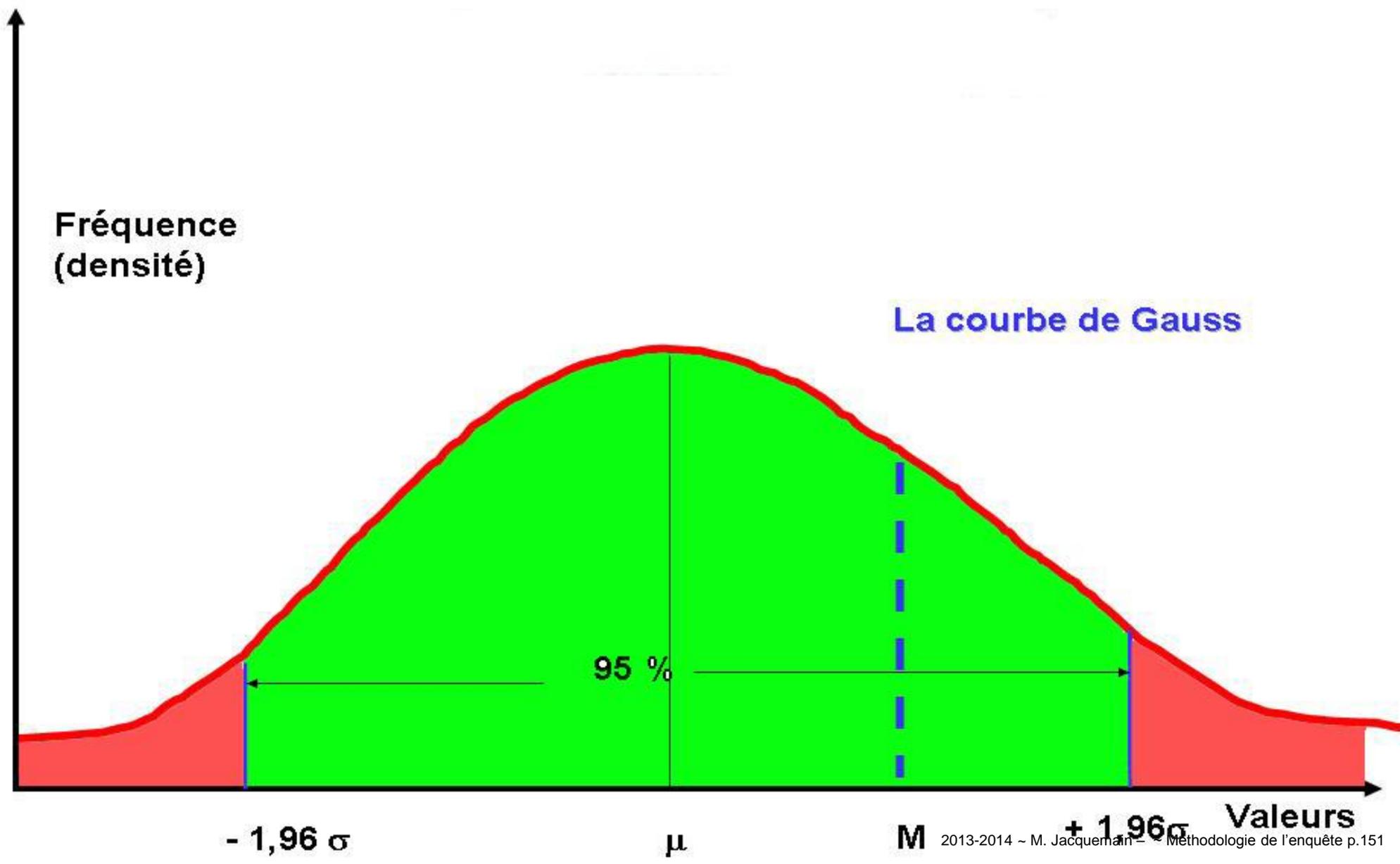
La probabilité de se situer entre deux valeurs est la surface sous la courbe.

Fréquence
(densité)

La courbe de Gauss



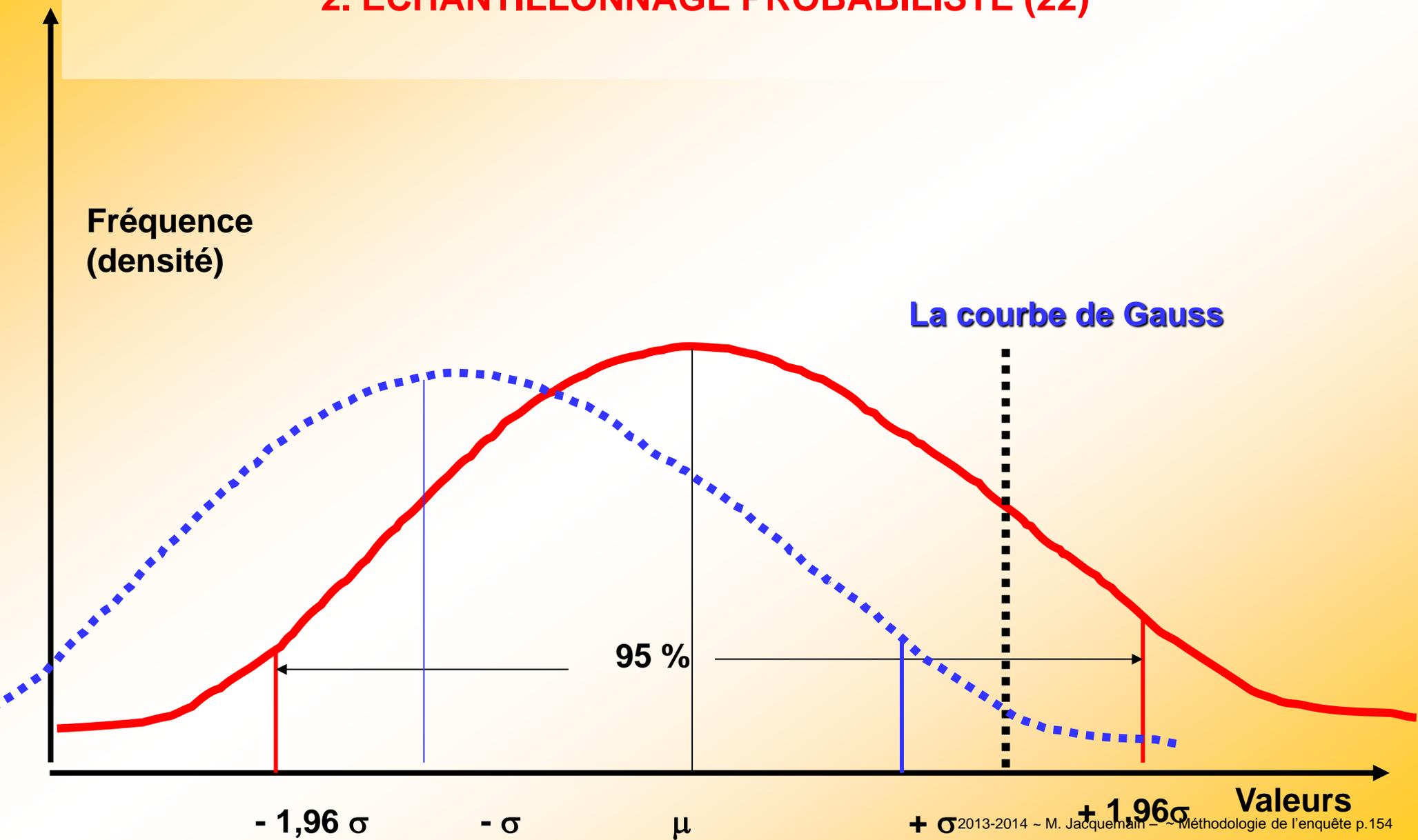
- **Qu'est-ce que cela veut dire ?**
- **Dans le cas d'un tirage aléatoire, 95 % des échantillons présenteront une moyenne (M) comprise entre**
 $\mu - 1,96 \sigma$ et $\mu + 1,96 \sigma$
 - **μ = moyenne de la population**
 - **σ = écart-type de la population (estimateur : celui de l'échantillon)**
 - **M = moyenne de l'échantillon**
- **Toujours dans le cas d'un tirage aléatoire, environ 68 % des échantillons présenteront une moyenne comprise entre**
- **$\mu - \sigma$ et $\mu + \sigma$**



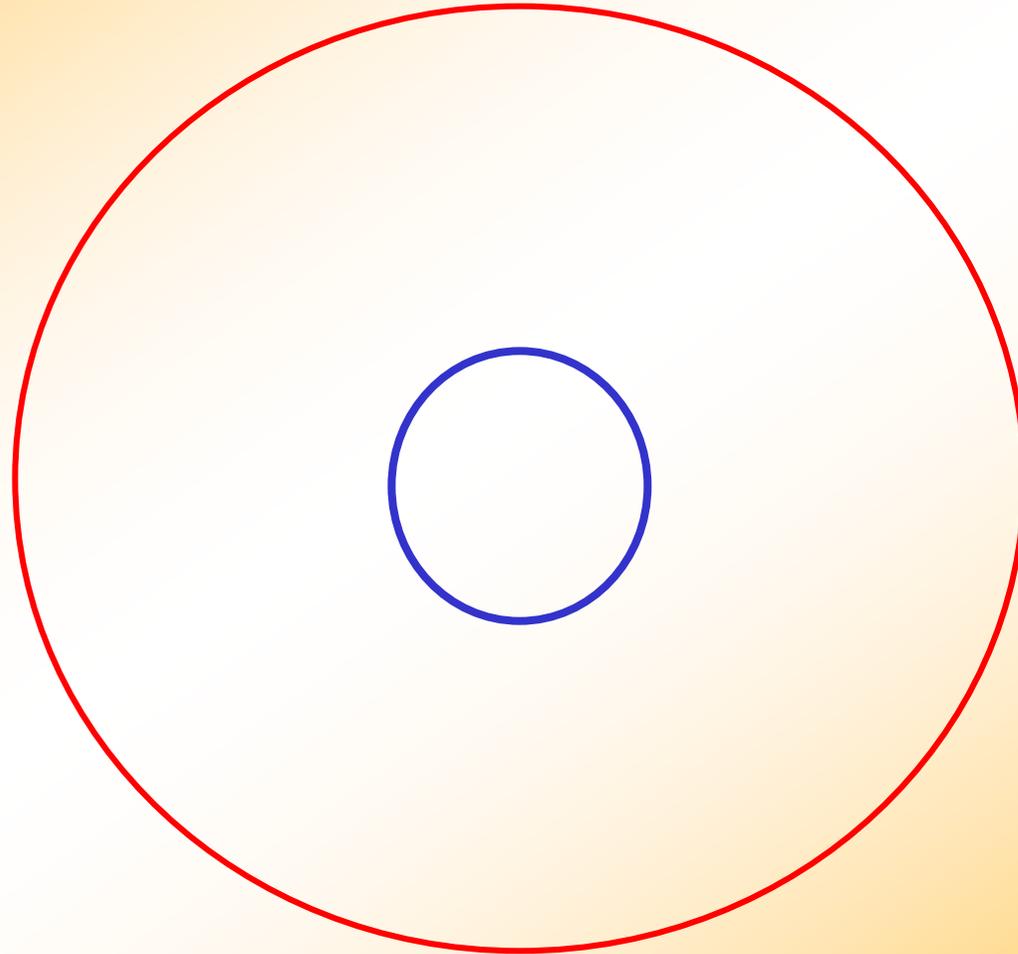
- **Interprétation :**
- **95 %** représente le **degré de certitude** : c'est le risque de nous tromper que nous acceptons de prendre
- L'intervalle [$M - 1,96 \sigma$, $M + 1,96 \sigma$] est **l'intervalle de confiance** : ce sont les limites entre lesquelles se situera la vraie valeur **pour un degré de certitude donné**. L'intervalle de confiance est ce qui vient de l'erreur d'échantillonnage.

- **En conséquence : plus nous exigeons un degré de certitude élevé, plus l'intervalle de confiance (la marge d'erreur) sera grande. Inversement, si nous acceptons un degré de certitude moins élevé, plus l'intervalle de confiance sera réduit.**
- **Exemple : pour un degré de certitude de 68,3 %, l'intervalle de confiance est compris entre $[M - \sigma, M + \sigma]$**

2. ECHANTILLONNAGE PROBABILISTE (22)



Si on passe
du **grand**
cercle au
petit cercle



Le degré de
certitude ?

L'intervalle de
confiance ?

→ **Exemple :**

→ **intervalle de confiance d'une proportion.**

→ **On connaît la formule pour l'écart-type de la distribution d'une proportion pour un échantillon probabiliste :**

$$\rightarrow \sigma = \sqrt{p(1-p)/n} \times \sqrt{(N-n)/(N-1)}$$

- **N** = population; **n** = échantillon **p** = proportion dans l'échantillon
- Cela permet de vérifier que lorsque **n** est petit par rapport à la population, la deuxième moitié de la formule tend vers 1.

- **Exercice :**
- **j'interroge trois cents habitants de Wallonie sélectionnés par une procédure aléatoire**
- - **J'obtiens 20 % de diplômés du supérieur**
 - **Que puis-je dire du pourcentage réel au sein de la population, avec un **degré de certitude** de 95 % ?**
- **Il est compris dans l'intervalle**
 - **$0,2 \pm 1,96 \times \sqrt{(0,2 \times 0,8)/300}$ soit $0,2 \pm 0,045$**
 - **Qu'il est compris entre 15,5 % et 24,5 %**

- **Comment puis-je réduire cet intervalle de confiance?**

- **En augmentant la taille de l'échantillon**
 - Exemple : si je prends 2.500 personnes, je peux dire avec un degré de certitude de 95 % que cette proportion est comprise entre 18,4 % et 21,6 %

- **En acceptant un degré de certitude plus faible**
 - Exemple : avec trois cents personnes, si j'accepte un degré de certitude de 68,3 %, je peux dire que la proportion est comprise entre 17,7 % et 22,3 %

- **Récapitulons l'évaluation de l'erreur d'échantillonnage pour un sondage aléatoire**
- **1. La précision de la mesure (étroitesse de l'intervalle de confiance) dépend du degré de certitude que nous exigeons**
- **2. Cette précision dépend également de la taille de l'échantillon**
- **3. Elle ne dépend pas, sauf cas particulier, du taux de sondage : 1000 personnes sur un million ou sur trois cents millions, cela revient au même.**
- **4. Remarquons que la précision, dans le cas d'une proportion, n'est pas la même pour toutes les proportions : elle diminue à mesure que l'on se rapproche du 50/50**

- **Exemple d'intervalle de confiance sur deux proportions différentes.**

- **Supposons que j'interroge mille personnes tirées au hasard pour connaître la proportion d'électeurs d'un parti X**

- **Si j'obtiens 10 %,**
 - **$p(1-p) = 0,9 \times 0,1 = 0,09$ et $\sigma = 0,0094$**
 - **L'intervalle de confiance (au seuil de 95 %) : $\pm 1,8\%$**

- **Si j'obtiens 50 %**
 - **$p(1-p) = 0,5 \times 0,5 = 0,25$ et $\sigma = 0,0158$**
 - **L'intervalle de confiance (au seuil de 95 %) : $\pm 3,1\%$**

- **Les biais dans les échantillons probabilistes.**
- **Rappel : un biais n'est pas l'erreur d'échantillonnage.** Il signifie que tous les membres de la population n'ont pas la même probabilité de faire partie de l'échantillon.
- **Le biais suppose donc que l'échantillon n'est pas représentatif.** L'erreur d'échantillonnage (et donc l'intervalle de confiance) est présente aussi dans l'échantillon représentatif.
- **Quand le biais est-il important ?** Quand on ne peut pas être raisonnablement sûr que les personnes non choisies auraient répondu la **même chose** que les personnes choisies.

- **Les sources de biais dans l'échantillon probabiliste**

- **1. La liste ne correspond pas à la population**
 - **Ex : on veut interroger l'ensemble des Wallons mais on part de l'annuaire téléphonique. Sont écartés tous ceux qui n'ont pas de téléphone fixe. Risquent-ils de répondre autrement que ceux qui en ont un ?**
 - **Oui, ce ne sont pas en général les mêmes catégories (revenu, âge, mode de vie...)**
 - **Exemple : les sondages pour les élections américaines ne trouvent pas les jeunes qui n'utilisent que le GSM**

→ 2. Les absences

- Les personnes que l'on trouve difficilement sont généralement plus actives. Donc, on risque de sur-représenter les inactifs.

→ 3. Les refus de répondre (taux de réponse)

- Le refus de répondre est généralement plus important dans les catégories disposant de moindres ressources culturelles : on risque de surestimer le poids des plus éduqués.

→ Si un échantillon tiré est aléatoire, l'échantillon des **répondants** ne l'est pas forcément. L'enjeu est donc que les deux soient le plus proches possible.

- **Comment corriger les biais ?**

- **Absence de la personne à interroger ou non réponse :**
 - **Nécessité de plusieurs essais** avant d'abandonner
 - Tirer un **échantillon de remplacement**
 - Possibilité d'estimation par **extrapolation** : vérifier si les réponses obtenues après un seul essai, deux essais, « n » essais sont identiques.
 - si oui, les non réponses n'impliquent pas de biais
 - Si non, on peut tenter d'extrapoler les réponses à l'essai « n+1 ».

- **Importance des biais : estimation de la valeur de la variable dépendante.**
- **Tout biais est important pour l'estimation d'une variable (résultats brut ou tris à plat)**
- **Exemple : supposons qu'on veuille estimer l'électorat des différents partis**
 - **Si certains électeurs (FN par ex) répondent moins que les autres, l'estimation est biaisée**
 - **Si certains électeurs (les jeunes, par ex) sont plus difficiles à toucher, l'estimation est biaisée (les jeunes votent plus Écolo, par ex).**

- **Importance des biais : relations entre variables**
- **Même exemple : je m'intéresse à la relation entre l'âge et le vote.**
 - Si les FN répondent moins que les autres mais qu'il n'y a pas de biais d'âge, on peut supposer que les FN répondront moins **dans toutes les catégories d'âge** et donc la relation âge/vote est préservée.
 - Si les plus âgés répondent moins que les autres, mais que il n'y a pas de biais liés au vote FN (ou pour un autre parti) on peut supposer que les plus âgés répondront moins, **indépendamment du parti pour lequel ils votent.**
 - Pour que la **relation** soit affectée, il faut donc que le taux de réponse soit biaisé **simultanément** selon la variable indépendante et la variable dépendante.
- **Conclusion : les biais d'échantillonnage sont plus dommageables pour estimer les valeurs des variables que pour valider des relations entre les variables**

- **Tirage d'un échantillon aléatoire.**
- **Principe** : lorsqu'on dispose d'une liste, il s'agit d'extraire de cette liste, au hasard, un certain nombre de noms jusqu'à ce qu'on obtienne l'échantillon souhaité.
- **Différentes méthodes de sélection au hasard sont disponibles :**
 - On utilise une table de nombres au hasard
 - On tire en fonction du dernier chiffre
 - Tirage par lettre : on prend tous les gens dont le nom commence par une certaine lettre
 - Tirage par date de naissance : on prend tous les gens qui sont nés un même jour de l'année
 - Etc.
- **Toujours vérifier qu'il ne peut pas y avoir un biais dans la méthode de tirage. Ex : les noms des étrangers ne commencent pas statistiquement par les mêmes lettres que les noms des nationaux.**

→ 3. ECHANTILLONNAGE PROBABILISTE A PLUSIEURS DEGRES

- **1. L'échantillon stratifié**
- **Principe** : on tire non pas un échantillon global mais un échantillon dans chaque strate de la population.
- **Exemple** : si je veux étudier les étudiants de l'Ulg, au lieu de tirer un échantillon sur l'ensemble, je peux tirer un premier échantillon parmi les étudiants de Bac, un échantillon parmi les étudiants de Master, un échantillon parmi les étudiants de 3^{ème} cycle
- **Utilité** : permettre de faire des inférences sur des groupes peu nombreux dans la population. A condition d'utiliser des taux de sondage différents dans les différentes strates.

- **Exemple 1 : je veux interroger 1000 citoyens belges adultes.**
- **Echantillonnage aléatoire simple : la répartition la plus probable sera 550 Flamands, 350 Wallons et 100 Bruxellois.**
- **Je pourrai donc dire peu de choses des Bruxellois (échantillon faible)**
- **Je peux donc décider de tirer **séparément** 333 Flamands, 333 Bruxellois et 333 Wallons. L'intervalle de confiance sera ainsi meilleur pour les Bruxellois**

- Évidemment, l'ensemble n'est plus représentatif de la population **belge**
- Si je veux un échantillon représentatif de la population belge, je dois donc **pondérer** mes résultats en fonction de la région d'origine

- **Exemple 2 :**
- Je veux un échantillon **d'étudiants ingénieurs** dans l'ensemble du pays.
- Si je tire un échantillon global, la proportion de **filles** sera faible et ce que je pourrai dire des filles sera plus imprécis que ce que pourrai dire des garçons

- **Je peux choisir de tirer par exemple 500 étudiantes parmi l'ensemble des filles et 500 étudiants parmi l'ensemble des garçons**
- **Bien sûr, il s'agit alors de deux échantillons différents : je n'ai pas un échantillon de 1000 étudiants mais deux échantillons de 500.**
- **Si je veux recomposer un échantillon représentatif, je dois repondérer en fonction inverse du taux de sondage.**

→ 2. Echantillonnage par grappes.

→ **Grappe** : ensemble d'unités que l'on peut tirer simultanément parce qu'elles ont un lien entre elles.

→ **Exemples** :

- Un ménage = une grappe d'individus
- Une entreprise = une grappe de salariés
- Une classe = une grappe d'étudiants
- Un immeuble = une grappe de logements.

- **Caractéristiques de l'échantillonnage par grappe**
- **L'unité d'analyse et l'unité de sondage diffèrent.**
 - **Ex 1 : je m'intéresse à des individus, mais j'ai une liste d'adresses :**
 - **Unité d'échantillonnage : logement (ou ménage)**
 - **Unité d'analyse = individus.**
 - **Ex 2 : je m'intéresse à des élèves mais j'ai une liste de classes dans les écoles**
 - **Unité d'analyse = élèves**
 - **Unité de sondage = classe ou école**
 - **Ex 3 : (inverse) : je m'intéresse au comportement d'achat des ménages mais j'ai une liste de personnes**
 - **Unité de sondage = individus**
 - **Unité d'analyse = ménage (grappe).**

- **Avantages du sondage par grappes**
- **Un moyen sensiblement plus facile de toucher certaines personnes**
 - Les lycéens, par classe
 - Les jeunes, si on ne dispose que des listes électorales, etc.
- **C'est parfois la seule base de données disponibles :**
 - Si on veut toucher telle catégorie professionnelle, on ira voir dans les entreprises
- **Inconvénients du sondage par grappe**
- **Les personnes d'une même grappe ont souvent une « proximité » particulière: effet de grappe**
 - les élèves d'une même classe, les membres d'un même club sportif...
- **Si on tire trop peu de grappes, la loi des grands nombres cesse de s'appliquer.**

- **Qui interroger dans la grappe ?**
- **En principe** : tous les membres d'une grappe doivent être interrogés sinon le sondage n'est plus représentatif : les membres ont une probabilité différente d'être interrogés selon la taille de la grappe
 - **Illustration** : supposons qu'on tire une classe sur mille aléatoirement et un élève par classe. Dans une classe de 15, chaque élève aura une chance sur 15.000 d'être sélectionné. Dans une classe de 30, il aura une chance sur 30.000.
- **En pratique** : l'effet de grappe est souvent considéré comme plus problématique que la sous-représentation de certaines personnes. On tentera alors de n'interroger qu'une personne par grappe.

- **Qui interroger dans la grappe ? 1^{er} exemple**
- **Je souhaite interroger les personnes en maison de repos**
 - Je possède une liste de 1000 maisons de repos
 - Je tire un échantillon aléatoire de 100 maisons
 - Qui interroger
 - Si j'interroge une personne par institution, alors difficulté d'échantillonnage
 - Dans une maison de 30 personnes, un résident a une chance sur 300 d'être interrogé
 - Dans une maison de 10 personnes, un résident a une chance sur 100 d'être interrogé
 - Mon échantillon n'est plus représentatif
 - Si j'interroge tous les membres de chaque maison de repos, chacun a une chance sur 10 d'être interrogé
 - Mon échantillon est représentatif
 - Mais : effet de grappe.

- **Qui interroger dans la grappe ? (suite)**
- **Solution possible** : si l'on connaît la taille des grappes, chacune est tirée avec une probabilité proportionnelle à sa taille.
- **Exemple** : on veut interroger 1000 personnes en Wallonie, et on ne dispose pas de liste générale mais on a accès aux registres communaux
 - On fait un tirage aléatoire de 50 communes et on interroge 20 personnes par commune
 - On organise le tirage de façon que chaque commune ait une probabilité d'être tirée proportionnelle à sa taille
 - Chaque personne aura une probabilité identique d'être sélectionnée.

- Exemple : on doit tirer dix communes pour obtenir notre échantillon.
- Comparons
 - Une commune A de 10.000 habitants
 - Une commune B de 20.000 habitants
 - On tire 100 personnes dans chaque commune.
 - **Si leur commune est sélectionnée** les habitants de A ont deux fois plus de chances d'être tirés que les habitants de B
 - Mais la commune A a deux fois **moins** de chances d'être sélectionnée que la commune B
 - Donc au total , un habitant de A a la même chance d'être sélectionné qu'un habitant de B

→ 4. ECHANTILLONNAGE EMPIRIQUE

- **Principe :**
- on ne connaît pas la probabilité a priori pour une unité d'être tirée (absence de liste possible)
- Il n'y a donc pas de base pour une estimation de « l'erreur d'échantillonnage ».
- On s'efforce alors d'assurer la représentativité par « **raisonnement** »
- On envisage les différents biais possibles et on s'efforce de les éliminer ou du moins de les réduire.

→ 1. Echantillonnage par quotas

- Le principe : on construit un échantillon qui **ressemble** le plus possible à la population initiale.
- On va donc s'arranger pour que, dans l'échantillon, la distribution des valeurs de certaines variables soient identiques à leur distribution au sein de la population
- Les valeurs que l'on souhaite sauvegarder dans l'échantillon seront le plus souvent
 - La distribution des sexes (même proportion de femme et d'hommes que dans la population)
 - La distribution des niveaux d'études
 - La distribution des statuts socio-professionnels
 - La distribution des âges

- **Echantillonnage par quotas (2)**
- **Exemple (fictif) : un échantillon de 1000 personnes de 18 ans et + en Wallonie**
- **Je devrai interroger**
 - **490 hommes**
 - **510 femmes**
- **Je devrai interroger**
 - **400 personnes ayant au plus un diplôme d'études primaires,**
 - **200 personnes ayant un diplôme d'études secondaires inférieures,**
 - **200 personnes ayant un diplôme d'études secondaires supérieures**
 - **200 personnes ayant un diplôme d'études supérieures**

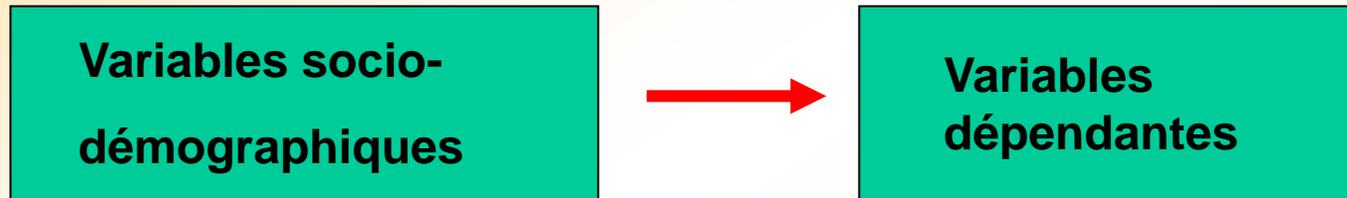
- **Echantillonnage par quotas (3)**

- **Je devrai interroger**
 - 100 étudiants
 - 250 retraités ou préretraités
 - 150 Ménagères
 - 500 « actifs » (parmi lesquels les chômeurs)

- **Etc. (voir les différents âges)**

- **Remarque : tous ces quotas sont des exemples fictifs (même s'ils se rapprochent de la réalité wallonne)**

- **Echantillonnage par quotas (4)**
- **Quelle base théorique pour justifier ce mode d'échantillonnage ?**
 - **Aucune base probabiliste**
 - **Mais on fait l'hypothèse d'une relation**



- **Qu'est-ce que cela veut dire ?**
- **On fait l'hypothèse (**forte**) que si l'échantillon est constitué comme la population sur les variables sociodémographiques, il réagira comme elle sur les opinions que l'on investigue**

- **Echantillonnage par quotas (5)**
- **1^{ère} précaution** : un plan d'échantillonnage. Il faut éviter de se retrouver en « queue d'enquête » à devoir trouver deux pensionnés de moins de 65 ans ou un étudiant de plus de (50 ans par exemple)
- **2^{ème} précaution** : possibilité de construire des **quotas croisés**. **Exemple** :
 - Au lieu de prendre un quota hommes/femmes et un quota actifs/inactifs
 - On impose de respecter la proportion d'actifs et d'inactifs parmi les femmes et les hommes
 - On aura donc quatre quotas (hommes actifs, homme inactifs, femmes actives, femmes inactives) à la place de deux.

- **Echantillonnage par quotas (6)**

- **2^{ème} précaution : être très attentif aux biais possibles :**
 - Avec les mêmes quotas, on trouvera des personnes différentes à des heures différentes ⇒ nécessité **d'échantillonnage temporel**
 - Avec les mêmes quotas, on ne trouvera pas les mêmes personnes à des endroits différents ⇒ nécessité **d'échantillonnage spatial**
 - Éviter la constitution de « **réseaux de répondants** » : les enquêteurs peuvent être tentés de réinterroger à chaque enquête les mêmes personnes parce qu'ils savent qu'elles acceptent de répondre.
 - Le **taux de refus** est plus élevé puisqu'on ne peut pas recontacter les personnes ⇒ attention aux biais de non réponses
 - Attention aux « **biais d'apparence** » : les enquêteurs peuvent sélectionner les personnes sur leur apparence physique.

→ 2. L'échantillonnage sur place (1)

- **Principe** : on sélectionnera les personnes en fonction d'un lieu déterminé qu'elles fréquentent
 - Un complexe de cinéma
 - Un site touristique
 - Une ville
 - Une salle de sport

- **Utilité** : c'est la seule manière de construire un échantillon lorsque la population est définie comme une « clientèle » pour laquelle aucune liste n'est disponible.

- **Difficulté spécifique** : on ne peut utiliser aucun quota pour contrôler la qualité de l'échantillon puisque, précisément, on ne connaît pas les caractéristiques sociodémographiques de la population.

→ 2. Echantillonnage sur place (2)

→ Biais possibles :

→

→ la clientèle change selon les moments (ex : les grands magasins) ⇒ nécessité **d'échantillonnage temporel**

→ S'il y a plusieurs lieux possibles (une chaîne de magasins, par ex), nécessité **d'échantillonnage spatial**

→ S'il y a plusieurs types d'activité ⇒ nécessité d'échantillonnage en fonction des **types d'activités**.

→ Les gros consommateurs seront plus souvent interrogés (voir l'exemple de la pompe à essence) ⇒ éventuellement pondérer en **fonction inverse de la consommation**

- **3. La méthode des itinéraires (1) ou *Random walking*.**
- **Utilité :** tenter d'imiter autant que possible l'aléatoire lorsqu'on ne dispose pas de liste et qu'on veut interroger dans un certain espace.
- **Exemple :** on veut interroger deux cents personnes dans une ville
 - On choisit un point de départ aléatoire
 - On prend la 2^{ème} rue à droite
 - On choisit la 13^{ème} maison
 - On interroge la personne de rang 1 dans le ménage
 - On prend la 2^{ème} rue à gauche
 - On choisit la 7^{ème} maison
 - On interroge la personne de rang 2 dans le ménage
 - On prend la 2^{ème} rue à droite et on recommence la procédure avec la personne de rang 3, etc.

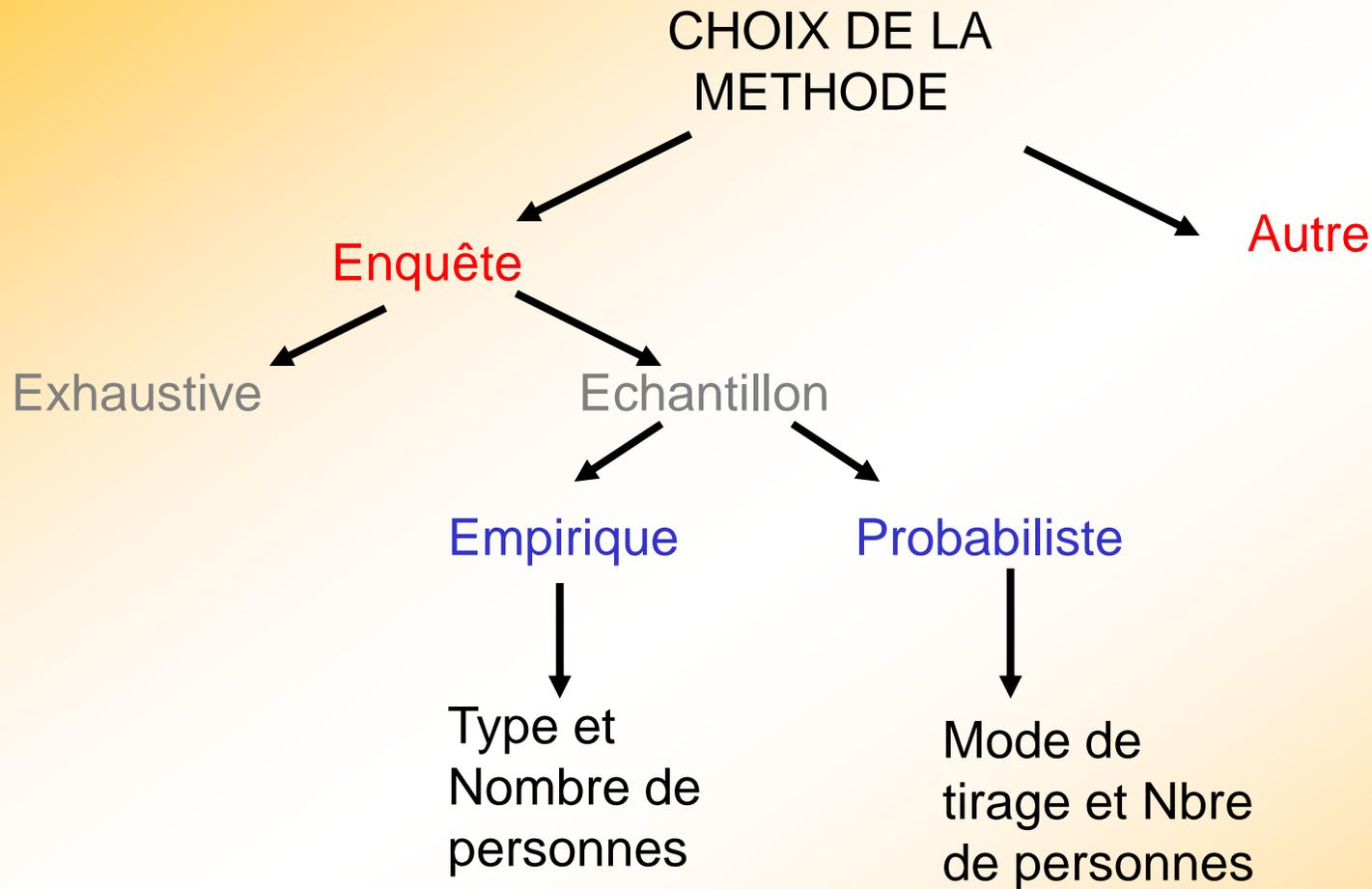
→ **3. La méthode des itinéraires (2)**

→ **Biais possibles et précautions :**

- **Prévoir les différents cas de figure, pour que l'enquêteur n'ait pas de choix à faire (algorithme)**
- **Si la personne que l'on veut interroger n'est pas disponible, il faut également une procédure déterminée (on revient, ou on passe à l'adresse suivante)**
- **Le point de départ et la procédure doivent être choisis de façon à ce que toute la superficie visée puisse être couverte.**

→ **Remarque :**

- **Il s'agit bien d'un hasard, mais on ne connaît pas les probabilités a priori des différentes personnes d'être sélectionnées. On ne peut donc pas parler d'échantillon aléatoire.**



Paramètres du choix :

- Information disponible
- Qualité souhaitée
- Budget

ANALYSE DES RESULTATS

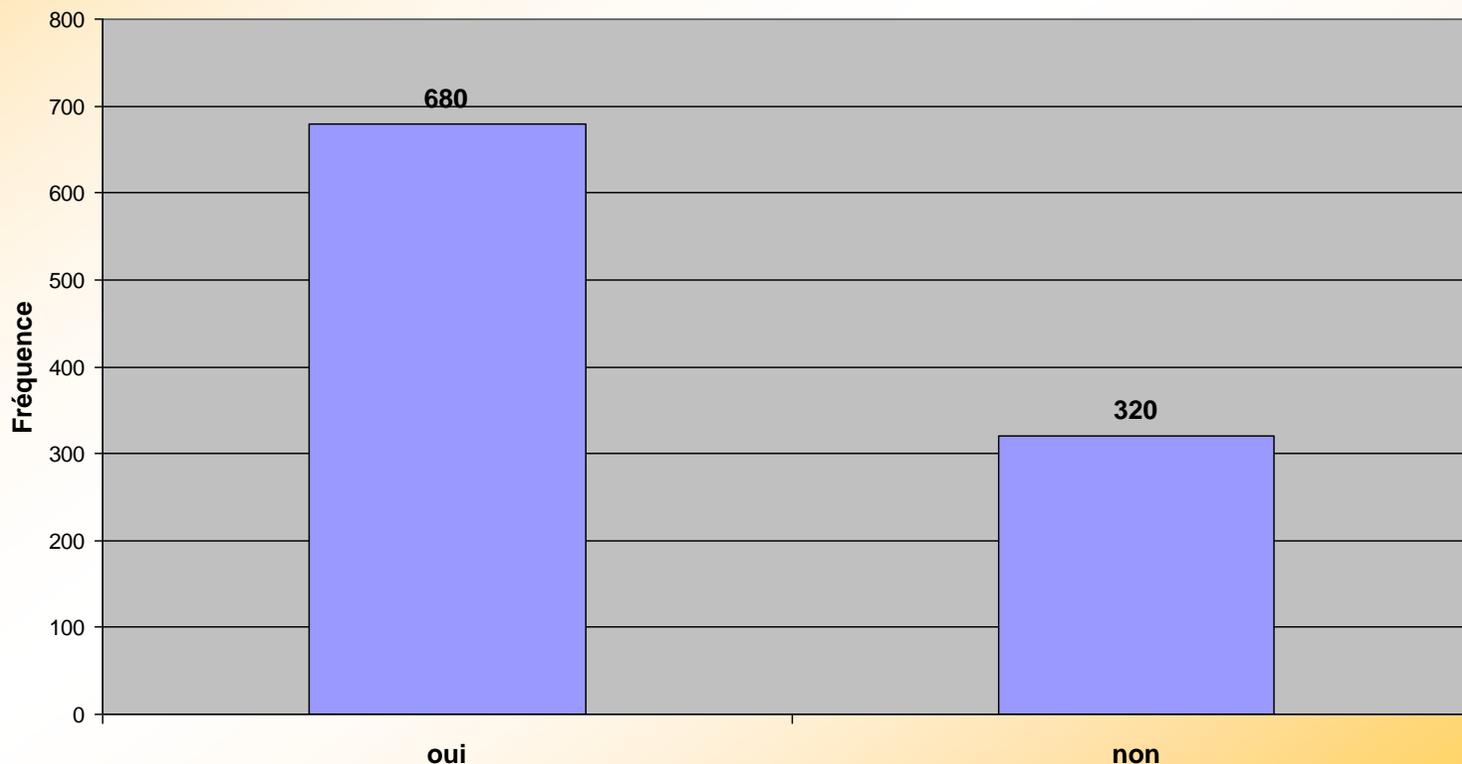
- 1. Tris à plat**
- 2. Tris croisés : relations entre variables**
- 3. Analyse multivariée : relations entre plusieurs variables**
- 4. Recompositions de variables**
- 5. L'interprétation de la causalité : l'exemple de l'âge**

→ **1. TRIS A PLAT**

1. Accepteriez-vous un homosexuel comme voisin ?

Résultat général

**ATTENTION : L'EXEMPLE QUI SUIT EST ILLUSTRATIF
TOUS LES CHIFFRES SONT FICTIFS**



- **Quelle première opération sur ces résultats ?**
- **Calcul d'un intervalle de confiance**
 - **Etape 1** : calcul de l'écart-type
 - $[(0,68) \times (0,32) / 1000]^{1/2} = 0,0147$
 - **Etape 2** : calcul d'un intervalle de confiance au seuil de probabilité (degré de certitude) de 0,05 :
 - $0,0147 \times 1,96 = 0,029$
- **Qu'est-ce que cela veut dire ?**

→ **2. TRIS CROISES : RELATIONS ENTRE DEUX
VARIABLES**

- **Que va vouloir savoir le sociologue une fois cette vérification faite sur le pourcentage lui-même?**
- **Il va vouloir vérifier si la distribution des résultats est la même dans les différentes catégories des **variables explicatives** qui lui paraissent éventuellement pertinentes**
- **La vérification d'une **relation entre deux variables** est donc la première étape de construction d'un modèle explicatif**
- **1^{ère} hypothèse : les femmes sont-elles plus tolérantes que les hommes ?**

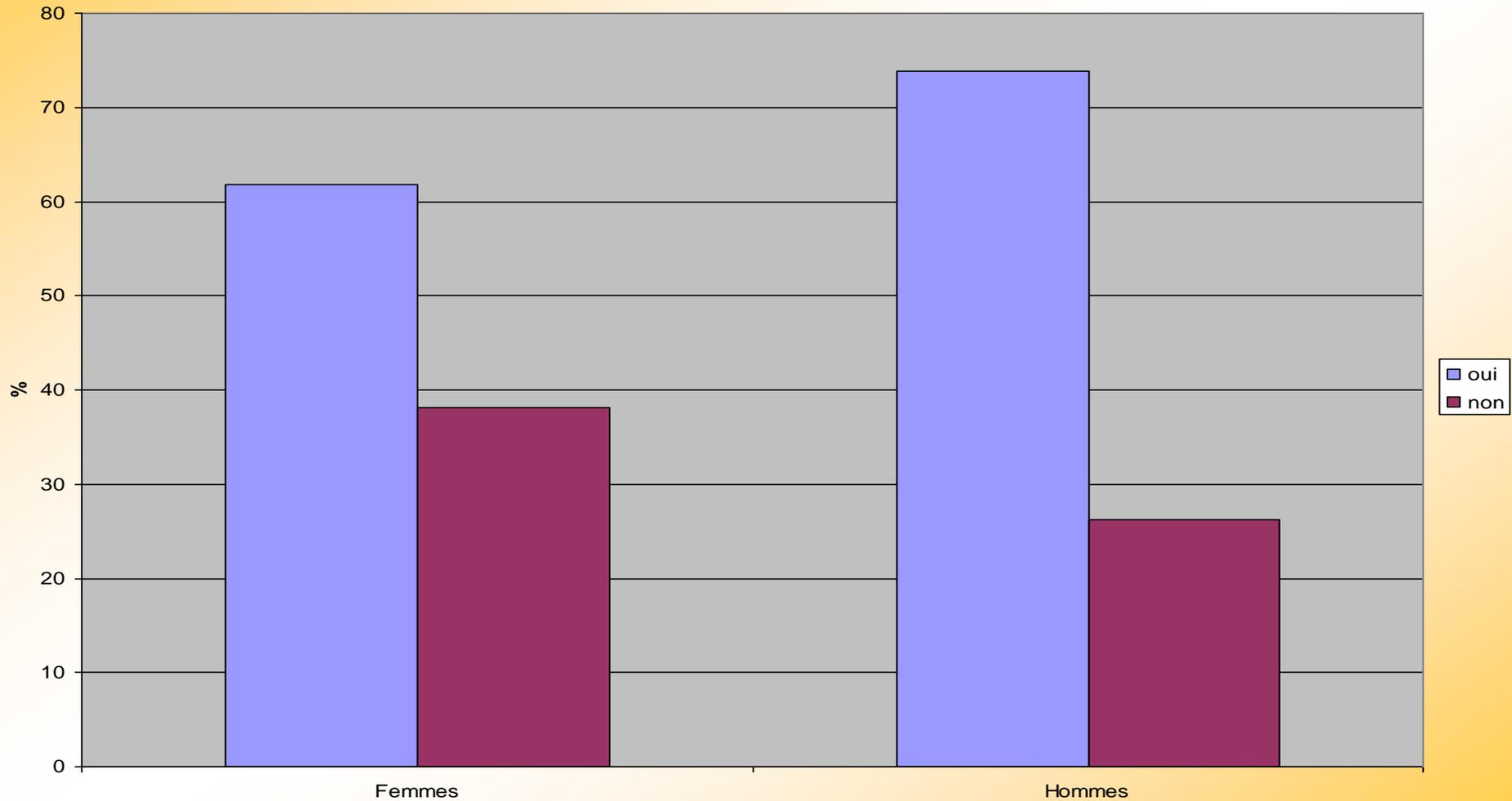
Accepteriez-vous un homosexuel comme voisin ?			
	Tableau des fréquences		
	Femmes	Hommes	TOTAL
Oui	300	380	680
Non	185	135	320
TOTAL	485	515	1000

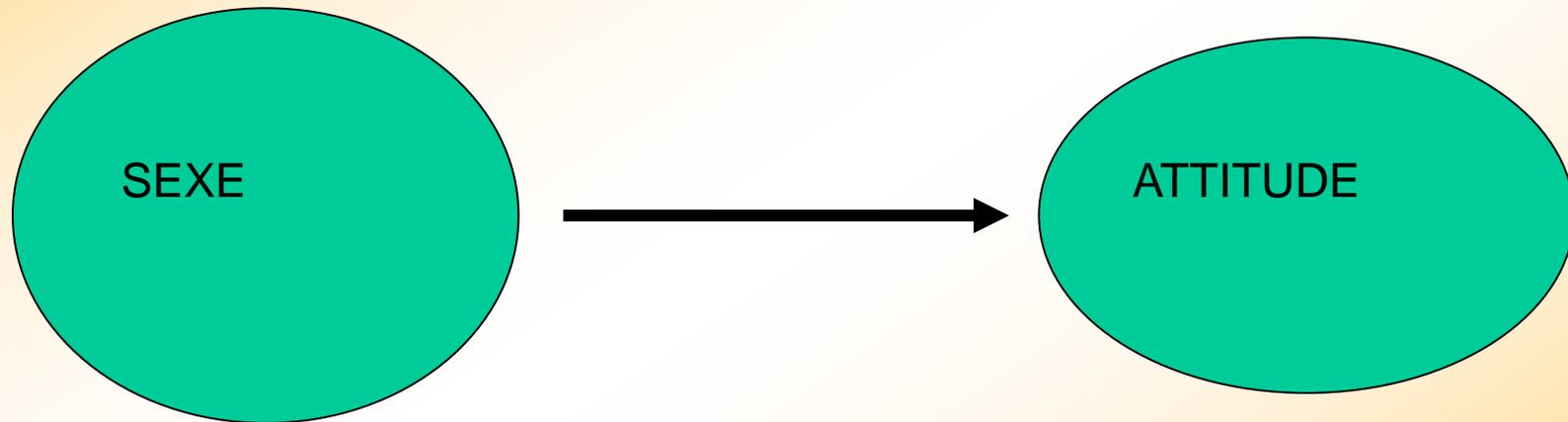
Accepteriez-vous un homosexuel comme voisin

Pourcentage parmi les hommes et parmi les femmes

	Femmes	Hommes	
oui	61,8	73,8	
non	38,1	26,2	
	100	100	

2. Accepteriez-vous un homosexuel comme voisin ? Ventilation par sexe





- Attention au flou du vocabulaire :
 - on parle aussi de variables indépendantes pour désigner les variables explicatives
 - Et de variables dépendantes pour désigner les variables expliquées
 - Evitons de confondre avec l'idée « d'indépendance » entre deux variables
- Pourquoi ai-je dit qu'il y a apparemment une relation statistique ?
 - Parce que comme, pour les pourcentages : il se pose une problème **d'inférence**
 - Quelle probabilité avais-je de tirer par hasard un échantillon dans lequel les deux variables sont liées à partir d'une population où elles ne le sont pas ?
 - Autrement dit, quelle chance avais-je de tirer un

- Deux variables sont **indépendantes** si le fait de connaître la valeur de l'une ne modifie pas la distribution des valeurs de l'autre

- Exemple :
 - La variable « yeux bleus ou bruns »
 - et la variable « cheveux blonds ou noirs »
 - sont **indépendantes**, si la probabilité d'avoir des cheveux blonds est **la même** quelle que soit la couleur des yeux (et symétriquement

- Donc, si j'ai dans une population donnée
 - 30 % de cheveux blonds (et 70 % de cheveux bruns)
 - 50 % d'yeux bruns (et 50 % d'yeux bleus)
 - Et que les deux variables sont **indépendantes**

- Je dois avoir 30 % de cheveux blonds **parmi** les yeux bruns (et autant parmi les yeux bleus)

- **Quelle est la probabilité d'avoir à la fois des yeux bruns et des cheveux blonds (si les variables sont indépendantes) ?**
- **Réponse : il y a 30 % de cheveux blonds parmi les 50 % de yeux bruns**
- **Il y a donc au total : $30 \% \times 50 \% = 15 \%$ de gens qui ont à la fois les cheveux blonds et les yeux bruns**
- **Dans le cas présent, les deux variables sont bien indépendantes puisque la proportion des différentes couleurs de cheveux est la même en fonction des différentes couleurs des yeux**
- **Connaître la valeur d'une variable ne m'indique rien sur la distribution de l'autre**

	Yeux bleus	Yeux bruns	TOTAL
Cheveux blonds	15	15	30
Cheveux bruns	35	35	70
TOTAL	50	50	100

- **Comment calculer l'écart à l'indépendance dans le cas de variables ?**
- On définit une variable aléatoire Chi2 qui vaut
 - $\Sigma [(F_{obs} - F_{th})^2 / F_{th}]$
- Le calcul des probabilités nous permet de calculer la probabilité que **CHI2 > x** pour un nombre donné de **degrés de liberté**.
- Comment calcule-t-on le nombre de degrés de liberté dans une table à deux entrées ?
- C'est le nombre de cellules dont il faut connaître la fréquence, **les marges étant données** pour que **toutes les valeurs des cellules soient fixées**.
- Soit (Nombre de lignes - 1) x (Nombre de colonnes - 1)

ma population il y a le même pourcentage d yeux bleus ou brun quelle que soit la couleur des cheveux ?

- Le chi-carré mesure précisément **l'écart à l'indépendance**
- La question devient donc : quelle **probabilité** avais-je d'obtenir une telle valeur du chi-carré dans mon échantillon si, dans la population, il n'y a pas de relation entre les deux variables
- Plus cette probabilité est **faible**, plus il est raisonnable de supposer **qu'il y a une relation** dans l'ensemble de la population
- En sciences sociales, on fixe souvent (mais c'est une **convention**) un seuil de **0,05**
- **Synthèse : si la probabilité d'obtenir une valeur donnée de Chi carré est inférieure à 5 % on**

	Yeux bleus	Yeux bruns	TOTAL
Cheveux blonds	16	14	30
Cheveux bruns	34	36	70
TOTAL	50	50	100

	Yeux bleus	Yeux bruns	TOTAL
Cheveux blonds	5	25	30
Cheveux bruns	45	25	70
TOTAL	50	50	100

→ 1^{er} Cas

- Valeur du Chi2 = $1/15 + 1/15 + 1/35 + 1/35 = 0,196$
- **P = 0,69**
- **On ne peut pas rejeter l'indépendance**

→ 2^{ème} cas

- Valeur du CHI2 = $100/15 + 100/15 + 100/35 + 100/35 = 19,6$
- **P = 0,00009**
- **On peut rejeter l'indépendance**

→ **PLUS LA VALEUR DU CHI2 EST ELEVÉE, PLUS SA PROBABILITÉ EST FAIBLE, ET MIEUX LA RELATION ENTRE LES VARIABLES EST ATTESTÉE**

Exemple : 1 degré de liberté

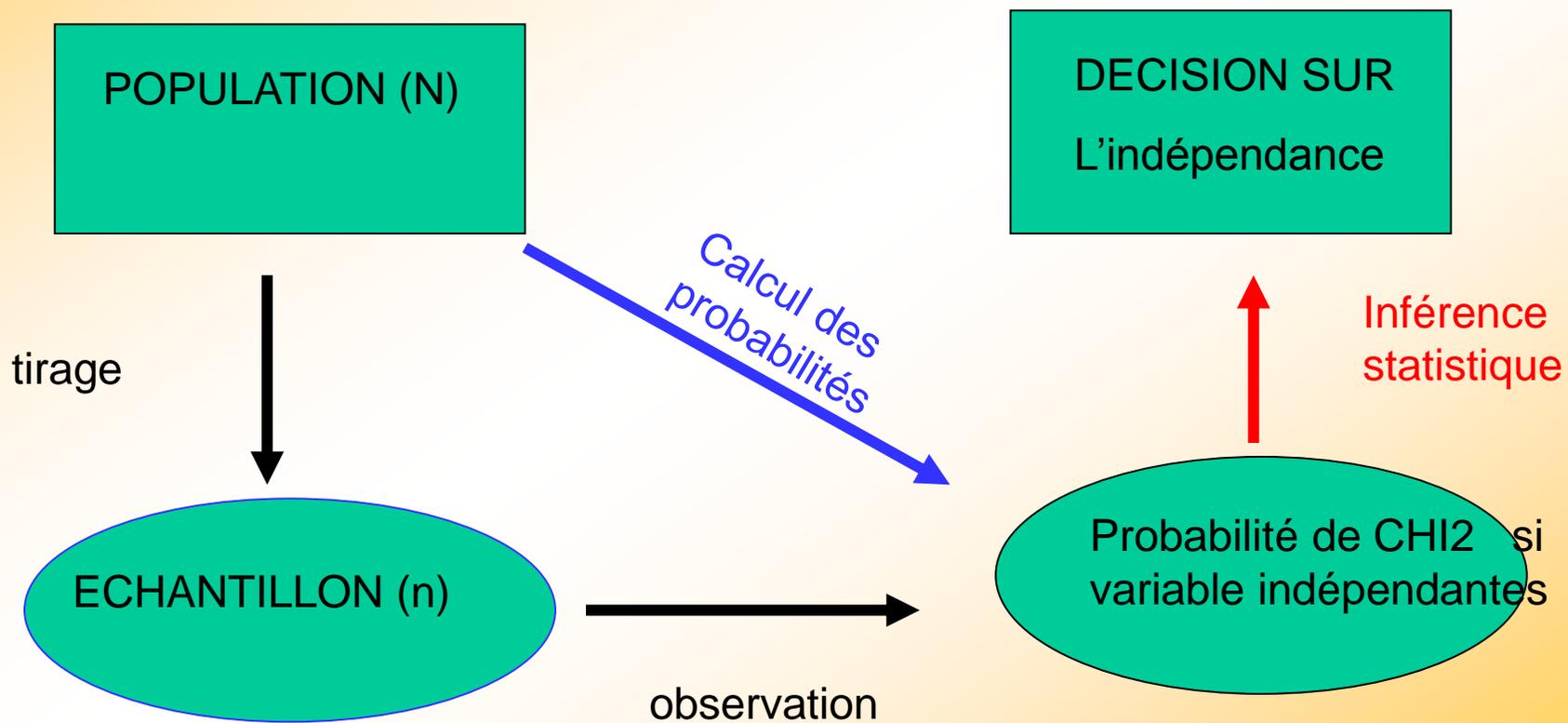
Accepteriez-vous un homosexuel comme voisin ?			
	Tableau des fréquences		
	Hommes	Femmes	TOTAL
Oui	380	300	680
Non	135	185	320
TOTAL	515	485	1000

Accepteriez-vous un homosexuel comme voisin ?			
	Fréquences observées		
	Femmes	Hommes	TOTAL
Oui	300	380	680
Non	185	135	320
TOTAL	485	515	1000
	Fréquences à l'indépendance		
	Femmes	Hommes	total
oui	329,8	350,2	680
Non	155,2	164,8	320
total	485	515	1000
P			0,00259647

- On va donc calculer la variable Chi2 dans l'exemple qui précède de la manière suivante :
- $[(300 - 329,8)^2 / 329,8] + [(185 - 155,2)^2 / 155,2] + [(380 - 350,2)^2 / 350,2] + [(135 - 164,8)^2 / 164,8] = X$
- Dans le cas d'un échantillon probabiliste, on peut calculer la probabilité suivante :
- $P(\text{Chi2} > x)$ si les deux variables sont indépendantes dans la population.
- Si cette $P < 0,05$, que conclut-on ?

Exemple 2 : 4 degrés de liberté

Accepteriez-vous un homosexuel comme voisin				
Fréquences observées				
	Primaire	Secondaire	Supérieur	
Oui	80	200	240	520
Non	60	100	60	220
Cela dépend	60	100	100	260
	200	400	400	1000
Fréquences à l'indépendance				
	Primaire	Secondaire	Supérieur	
Oui	104	208	208	520
Non	44	88	88	220
Cela dépend	52	104	104	260
	200	400	400	1000
P				0,00073613

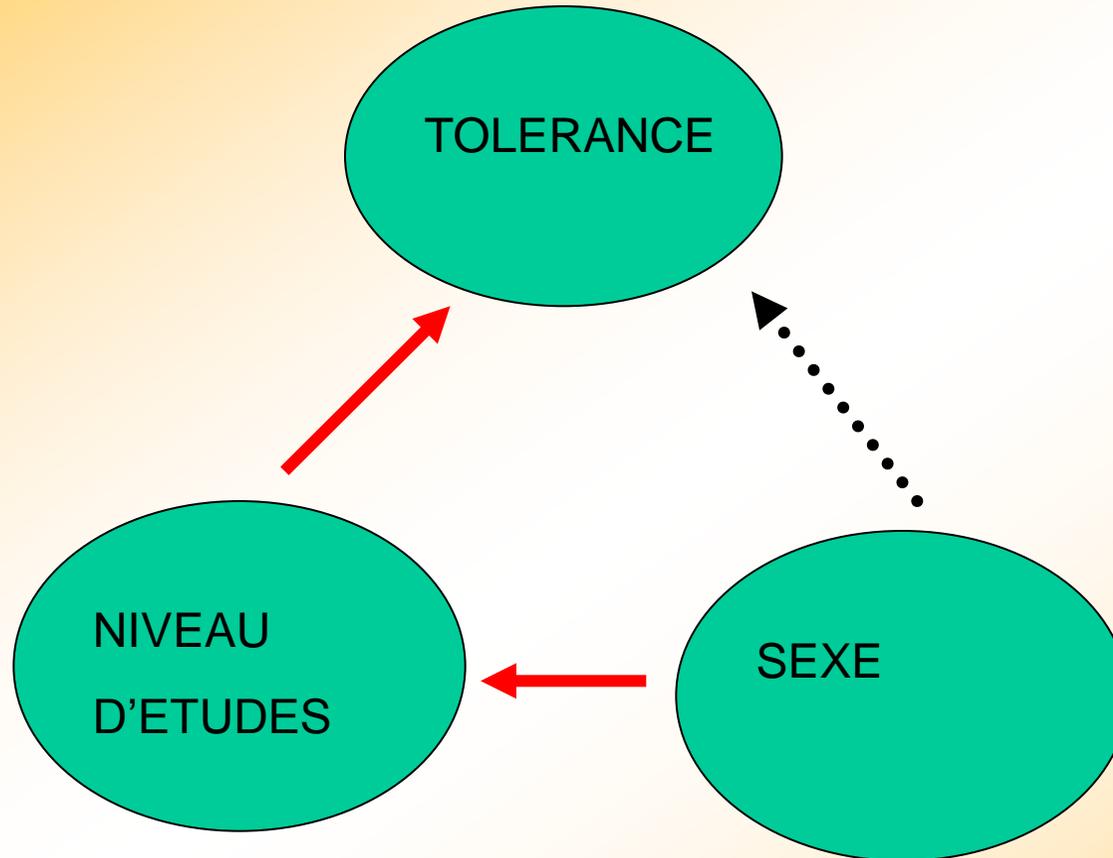


- Remarque sur la nature des variables en sciences sociales
- On s'est limité ici au cas particulier où les variables sont **qualitatives** ou **ordinales**
- C'est le plus souvent le cas en sociologie, mais dans d'autres disciplines, on aura souvent des variables **quantitatives** (voir la première partie du cours)
- Exemple-type : en économie, on raisonne sur des **revenus, des prix, des taux** qui sont des variables quantitatives
- Ce sont d'autres tests que le test de Chi2 qui s'appliquent alors mais le principe intuitif est le même.

→ **3. ANALYSE MULTIVARIEE : RELATIONS ENTRE PLUSIEURS VARIABLES**

- **Comment interpréter la relation ?**
- **1. Est-ce une vraie relation ?**
- **Nécessité de contrôler certaines variables interférentes possibles**
- **Que signifie contrôler les variables interférentes ?**
 - ⇒ **Cela veut dire vérifier la relation pour chacune des catégories de la variable potentiellement interférente (variable de contrôle)**

- **Exemple :**
- **Supposons que nous découvrons dans nos chiffres une relation entre le **niveau d'études** et l'attitude à l'égard de l'homosexualité**
- **Supposons par ailleurs que le niveau d'études ne soit **pas le même** parmi les hommes et les femmes**
- **La relation entre sexe et tolérance à l'homosexualité est-elle alors due au niveau d'étude ou au sexe ?**
- **Pour le savoir, on analyse la relation entre sexe et tolérance **à l'intérieur** de chaque niveau d'études**



		Accepteriez vous un homosexuel comme voisin ?					
		RAPPEL : chiffres fictifs					
		Femmes			Hommes		
	Educ +	Educ -	Total Femmes	Educ +	Educ -	Total Hommes	
oui	160	140	300	329	51	380	
non	40	145	185	83	52	135	
	200	285	485	412	103	515	

Accepteriez vous un homosexuel comme voisin. Pourcentages					
		Femmes		Hommes	
	Educ +	Educ -	Educ +	Educ -	
oui	0,8	0,49	0,8	0,49	
non	0,2	0,51	0,2	0,51	

- L'analyse « **multivariée** » consiste précisément à construire des modèles où l'on prend en compte simultanément **l'effet de plusieurs variables**.
- L'analyse multivariée permet par exemple de vérifier si la relation que l'on a décelée entre deux variables ne vient pas d'une **variable interférente**
- Elle permet aussi de vérifier si la relation entre deux variables, tout en étant « réelle », ne serait pas **influencée** par une troisième.

- Prenons l'exemple d'une relation entre deux variables qui **s'inverse** lorsqu'on fait intervenir une troisième
- Un certain nombre d'études ont montré qu'il y avait une **relation positive** entre la pratique sportive et le tabac : plus on fait du sport et plus on a de chance d'être fumeur !
- Cette relation paradoxale a été obtenue à une époque où il y avait **à la fois** beaucoup plus de fumeurs et beaucoup plus de pratiquants sportifs parmi les hommes que parmi les femmes
- Lorsqu'on décompose les résultats en fonction du **sexe**, la relation entre pratique sportive et tabac s'inverse.

Analyse des résultats

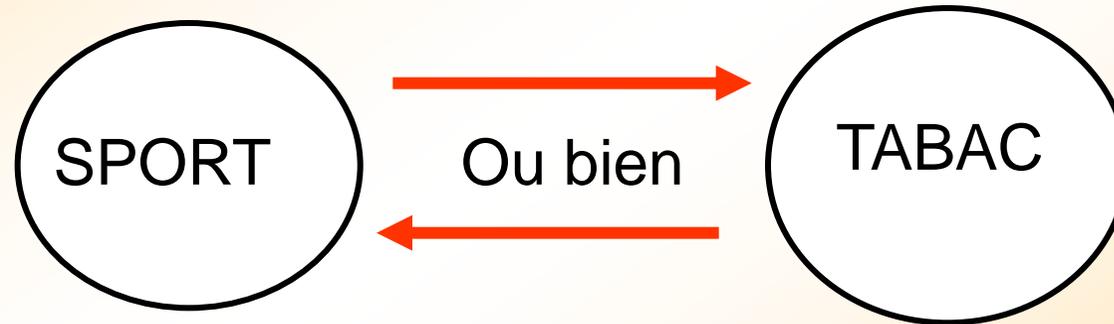
	Fumeurs	Non Fumeurs	Total
Sportifs	60 %	40 %	100 %
Non sportifs	40 %	60 %	100 %

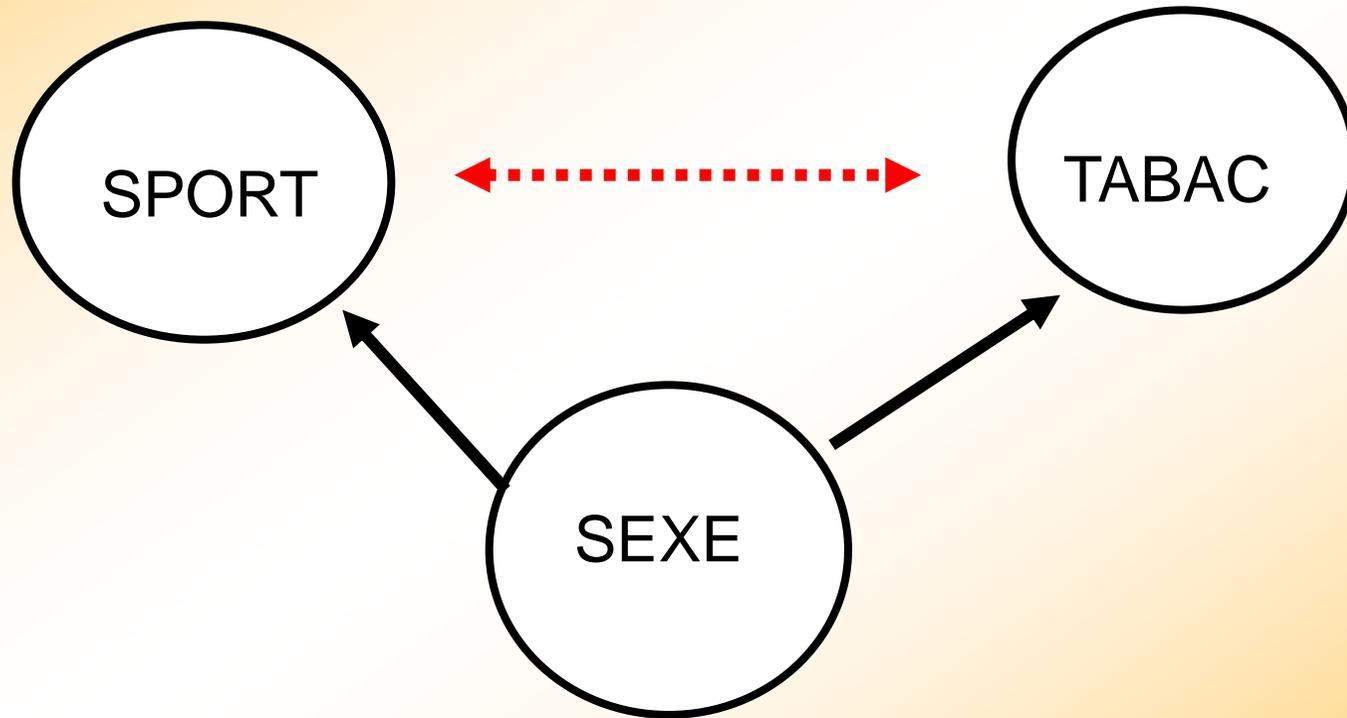
Analyse des résultats

	Fumeurs	Non Fumeurs	Total
Sportifs	30	20	50
Non sportifs	20	30	50
TOTAL	50	50	100

	HOMMES			FEMMES		
	Fumeurs	Non fumeurs	TOTAL	Fumeuses	Non Fumeuses	TOTAL
Sportifs	30	10	40	0	10	10
Non Sportifs	10	0	10	10	30	40
TOTAL	40	10	50	10	40	50

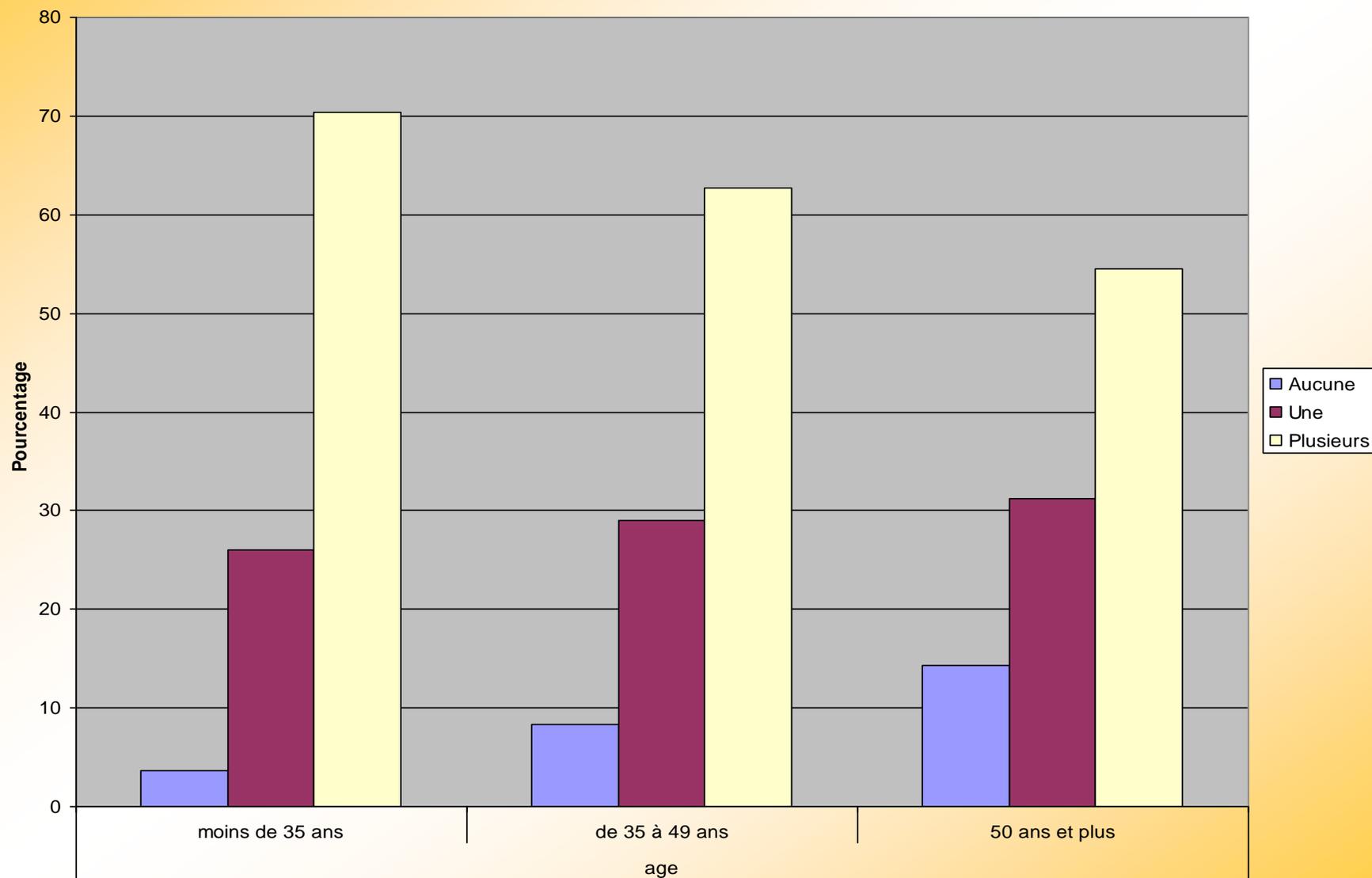
	HOMMES			FEMMES		
	Fumeurs	Non fumeurs	TOTAL	Fumeuses	Non Fumeuses	TOTAL
Sportifs	75 %	25 %	100 %	0 %	100 %	100 %
Non sportifs	100 %	0 %	100 %	25 %	75 %	100 %



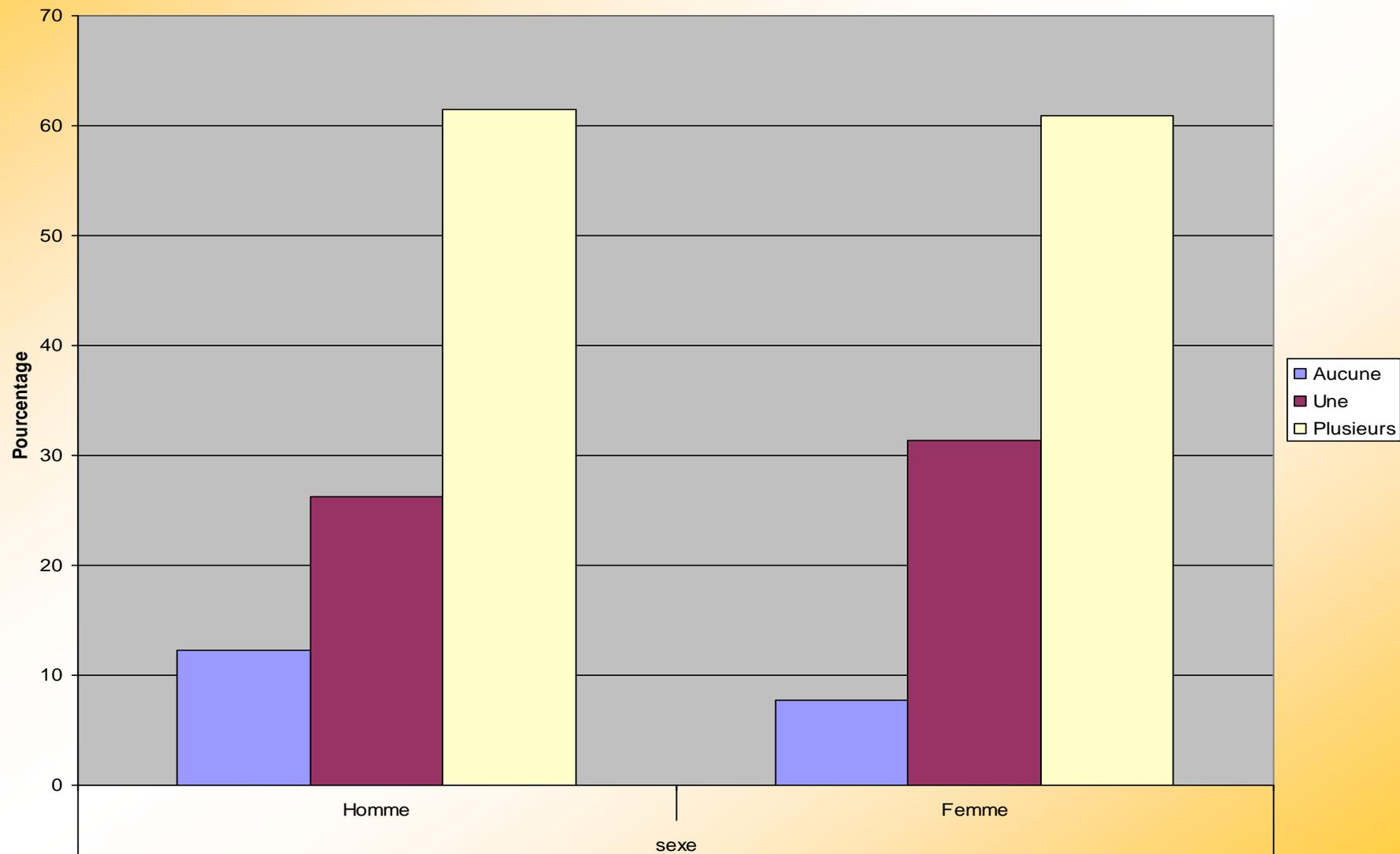


→ http://fr.wikipedia.org/wiki/Paradoxe_de_Simpson

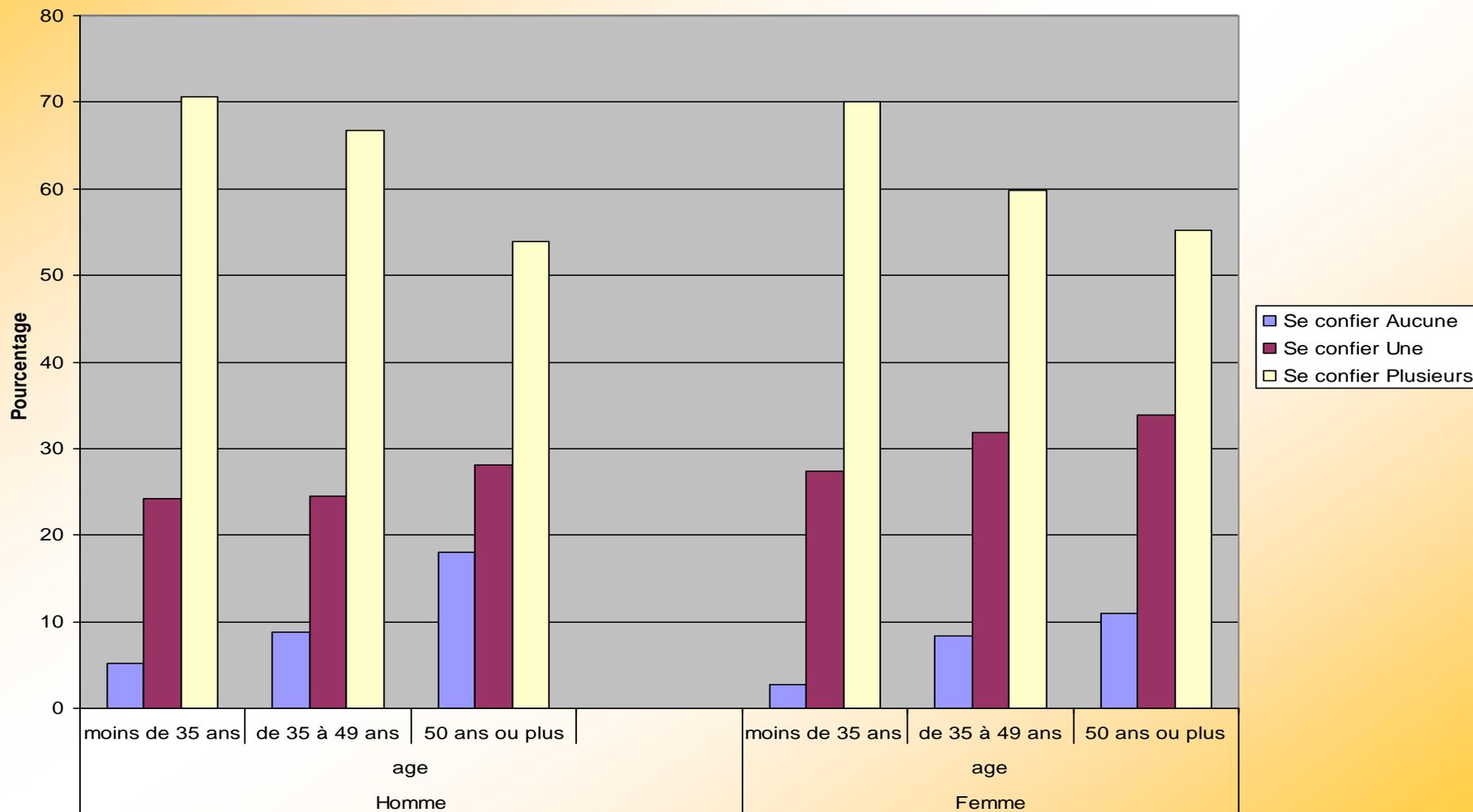
A combien de personnes peut-on se confier ? Selon l'âge



A combien de personnes peut-on se confier ? Selon le sexe



A combien de personnes peut-on se confier ? Selon l'âge et le sexe

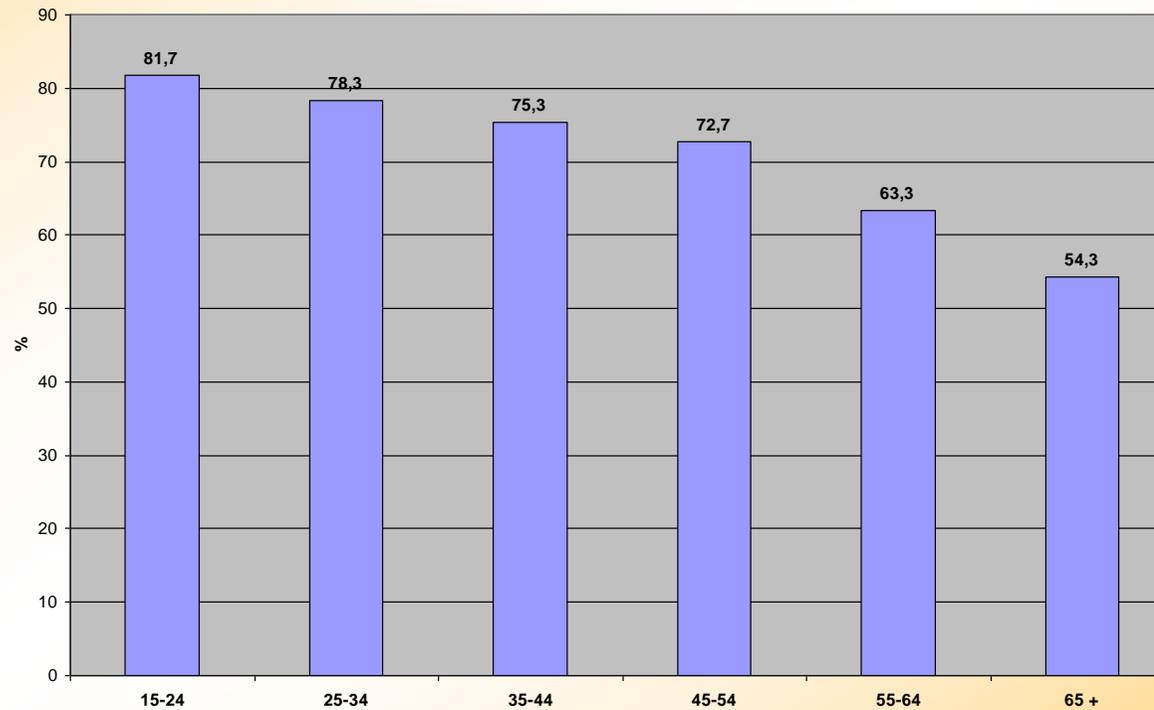


- **2. Si on pense qu'il s'agit d'une « vraie » relation, comment l'interpréter ?**
- Dans l'exemple de la tolérance à l'homosexualité, on peut supposer que le niveau d'études accroît la tolérance, parce que, de manière générale, davantage d'éducation « élargit » les principes moraux.
- Peut-on imaginer une relation dans **l'autre sens** ?
- Seule la **nature des variables** nous renseigne sur le **sens** de la relation causale (du moins dans le cas d'une enquête unique).
- Dans le cas d'une enquête répétée, on peut déduire le sens de la relation de la succession temporelle : la causalité va toujours du **passé vers le futur**.

- **2^{ème} hypothèse** : on va maintenant tester s'il y a une relation entre l'âge des personnes interrogées et leur tolérance à l'égard de l'homosexualité.
- **L'âge des personnes interrogées est très souvent une des variables les plus discriminantes par rapport aux attitudes.**
- **Comment interpréter ce résultat ?**

Accepteriez-vous un homosexuel comme voisin ?						
Enquete 2000 réponse par âge						
AGE	15-24	25-34	35-44	45-54	55-64	65 +
% de OUI	81,7	78,3	75,3	72,7	63,3	54,3

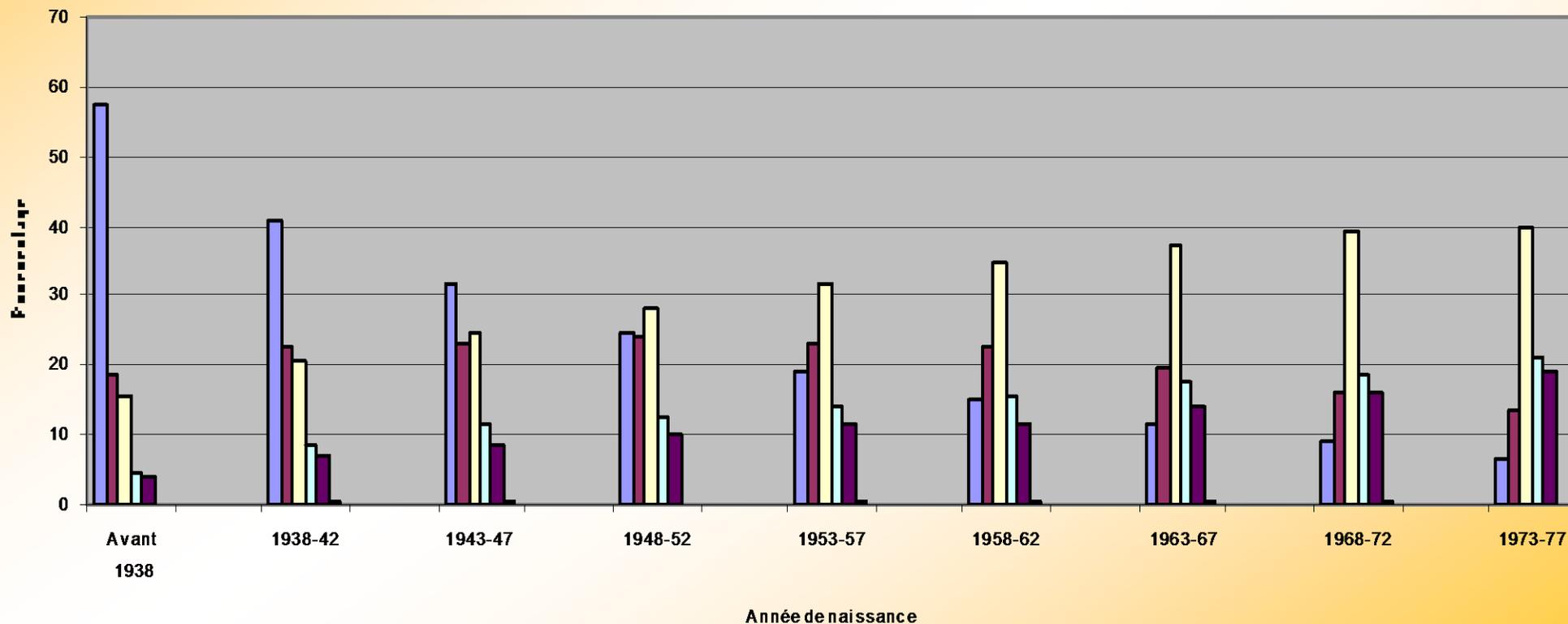
3. Accepteriez-vous un homosexuel comme voisin ? Ventilation par âge



Accepteriez-vous un homosexuel comme voisin							
Chiffres réels							
	15-24	25-34	35-44	45-54	55-64	65 +	
oui	98	94	113	109	114	152	680
non	22	26	37	41	66	128	320
	120	120	150	150	180	280	1000
Chiffres théoriques en cas d'indépendance							
	15-24	25-34	35-44	45-54	55-64	65 +	
oui	81,6	81,6	102	102	122,4	190,4	680
non	38,4	38,4	48	48	57,6	89,6	320
	120	120	150	150	180	280	1000
							3,9734617E-06
							0,000004

- **Première étape : y a-t-il un risque de variable interférente ?**
- Oui : **le niveau d'études** est fortement lié à l'âge.
- Pourquoi ? Parce que le niveau d'études moyen au sein de la population n'a **pas cessé d'augmenter** au cours des cinquante dernières années.
- Donc les jeunes d'aujourd'hui sont en moyenne **beaucoup plus éduqués** que les plus âgés.
- Il faudrait donc d'abord vérifier que la relation entre l'âge et la tolérance à l'égard de l'homosexualité est conservée **à l'intérieur** de chaque niveau d'études.

Niveaux scolaires par âge en communauté française en 2002 (chiffres réels !)



On sait donc déjà que, s'il y a une relation entre le niveau d'études et la tolérance à l'égard de l'homosexualité, il faut contrôler le niveau d'études puisque celui-ci est distribué de façon très différente parmi les différentes tranches d'âge.

- Ici encore on peut donc faire une analyse multivariée : à l'intérieur de chaque niveau d'étude, vérifier si la liaison entre âge et tolérance reste significative (et la même !)
- On peut imaginer que le niveau d'éducation
 - Vienne annuler la relation entre l'âge et la tolérance : on est plus tolérant parce qu'on est plus éduqué quel que soit l'âge
 - Entre en interaction avec l'âge pour déterminer la tolérance : par exemple,
 - Parmi les moins éduqués la tolérance varierait fortement avec l'âge
 - Parmi les plus éduqués, la tolérance serait indépendante de l'âge.

- De la même façon, il faudrait contrôler si le **sexe** intervient dans la relation entre âge et tolérance à l'égard de l'homosexualité.
- Pourquoi ? Parce que **l'âge moyen** des femmes dans notre population est supérieur à celui des hommes.
- Si notre échantillon est correctement constitué, il est donc fort probable que les femmes seront plus nombreuses dans les **tranches d'âge élevées** que dans les tranches d'âge les plus basses.
- Donc le fait que la tolérance n'est pas la même entre hommes et femmes **influence** probablement la relation entre l'âge et la tolérance.

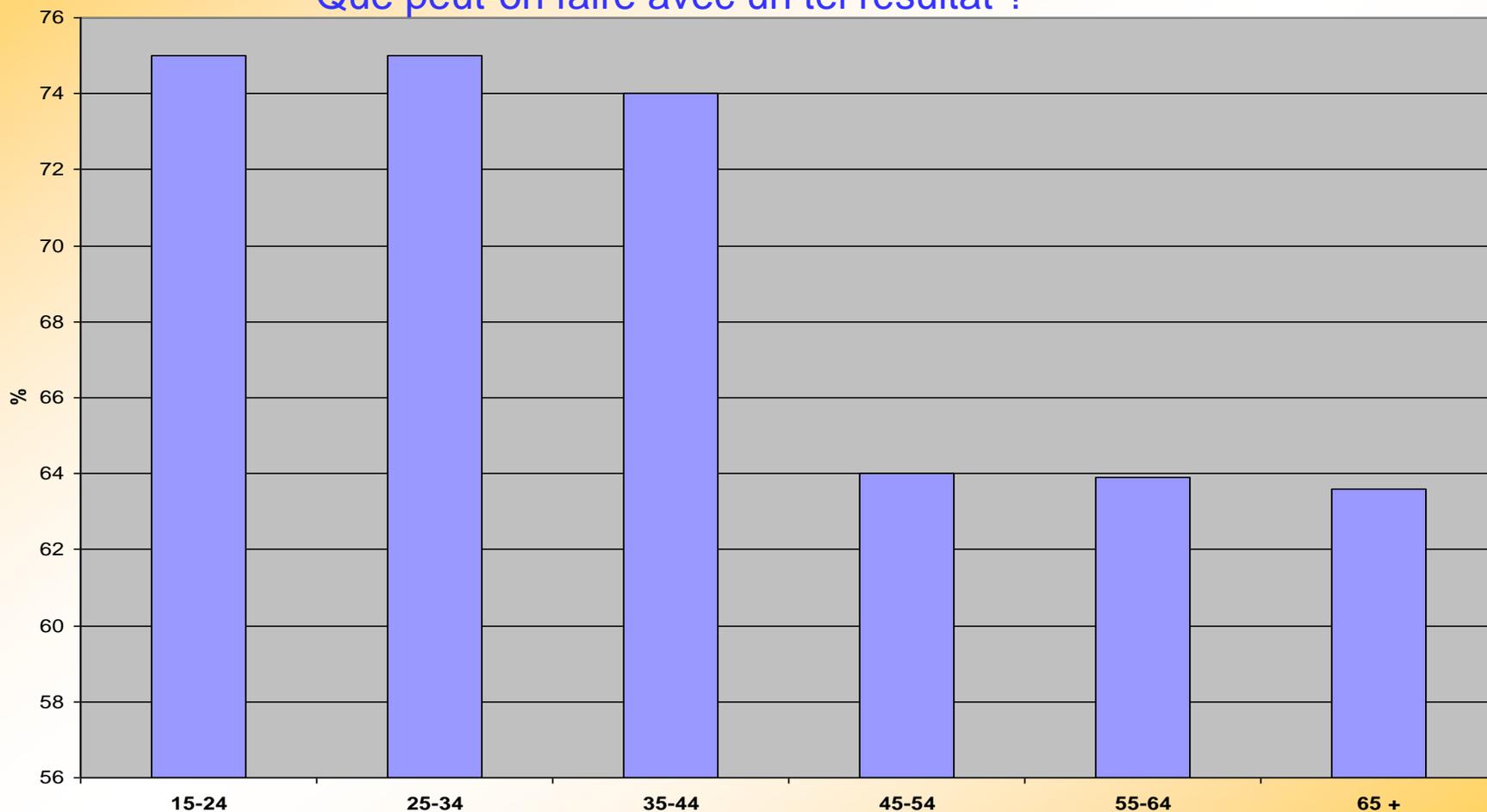
→ 4. RECOMPOSITION DE VARIABLES

Accepteriez-vous un homosexuel comme voisin ?						
Enquête 2000 réponse par âge différentes						
	15-24	25-34	35-44	45-54	55-64	65 +
	75	75	74	64	63,9	63,6

Test de Chi2 : P = 0,31

Accepteriez-vous un homosexuel comme voisin ? Distribution bi-polarisée

Que peut-on faire avec un tel résultat ?



Accepteriez-vous un homosexuel comme voisin ?				
Enquête 2000 bloc d'âges.				
	Moins de 45	45 et +		
OUI	291	389	680	
NON	99	221	320	
TOTAL	390	610	1000	
	265,2	414,8	680	
	124,8	195,2	320	
	390	610	1000	
			0,01198894	

Quel est l'intérêt de dichotomiser la variable ?

→ 5. INTERPRETATION DE LA CAUSALITE : L'EXEMPLE DES EFFETS LIES A L'AGE

- **Lorsqu'on contrôle le niveau d'études des répondants**
 - **Lorsqu'on contrôle le sexe » des répondants**
 - **Les plus âgés continuent à être moins tolérants que les plus jeunes.**
-
- **Nous en arrivons donc à l'hypothèse qu'il y a une « vraie » relation entre l'âge et la tolérance (même si elle est partiellement influencée par d'autres variables)**
 - **Pouvons nous induire le sens de cette relation ?**
 - **Oui : s'il y a une causalité, elle va dans le sens âge \Rightarrow tolérance.**
 - **Peut-on en effet imaginer une relation dans le sens tolérance \Rightarrow âge ?**

- Que veut dire cette relation ? Nous sommes devant une question **d'interprétation**
- Nous ne pouvons supposer une relation de causalité entre deux variables si nous sommes incapables de lui trouver une signification.
- L'analyse en termes de **causalité** (*Explication = Erklären*) doit donc se doubler d'une analyse en termes de **signification** pour les personnes interrogées (*Compréhension : = Verstehen*)
- Or la signification de l'influence de l'âge est précisément très questionnée en **sociologie**.
- **Que signifie une relation entre l'âge des répondants et une attitude ou opinion ?**

- **1^{ère} Hypothèse : effet de cycle de vie : 3 aspects principaux**
- **L'âge physiologique (physique ou mental)**
 - La consommation médicale
 - La capacité à apprendre ou à changer
 - L'activité sexuelle, etc.
- **Le statut professionnel, avec ce qu'il implique (revenu, budget temps...)**
 - Etudiant
 - Actif
 - Retraité
- **Le statut familial**
 - Vie sous la tutelle des parents ou autonomes
 - Vie en couple ou isolé
 - Présence ou absence d'enfants
 - Soutien des enfants, des parents

- 2^{ème} hypothèse : effet de génération
- Ici l'âge n'est plus un indicateur de situation physique, mentale ou sociale
- L'âge intervient ici comme indicateur de **l'année de naissance**
- Les opinions et attitudes d'une personne dépendent en effet crucialement des **valeurs dominantes durant sa période de socialisation.**

- Par exemple : La théorie de Ronald Inglehart (*The silent revolution*)
- Une personne socialisée durant une période de **guerre ou de crise économique dure** aura tendance à s'accrocher davantage à des valeurs de sécurité matérielle ou psychologique (matérialistes)
- Une personne socialisée durant une période de **profonde imprégnation religieuse** aura tendance à attacher plus d'importance au conformisme moral et aux règles communautaires
- De manière générale, les **attitudes fondamentales** (tolérance/intolérance, individualisme/solidarisme, sécurité/insécurité...), **les valeurs et les pratiques culturelles** ont tendance à s'enraciner tôt dans la vie et à changer peu par la suite.

- L'effet de cycle de vie et de génération peuvent se contrarier
- Exemple : la consommation de **Science-Fiction**
- En termes de **cycle de vie** c'est une littérature que l'on tend à abandonner en vieillissant. On devrait donc trouver plus de lecteurs chez les plus jeunes
- En termes de **générations** c'est une littérature qui était davantage présente dans les années 60-80. Donc elle devrait être plus représentée chez ceux qui avaient entre 12 et 35 ans à l'époque et qui ont grosso modo entre 40 et 65 ans aujourd'hui.
- Dans cette hypothèse on pourrait voir les deux effets jouer **en sens contraire** et la lecture de science-fiction s'étaler de manière plus ou moins égale sur les différents âges.

- Revenons à la relation entre âge et tolérance à l'homosexualité :
 - Effet de cycle de vie
 - Effet de génération ?

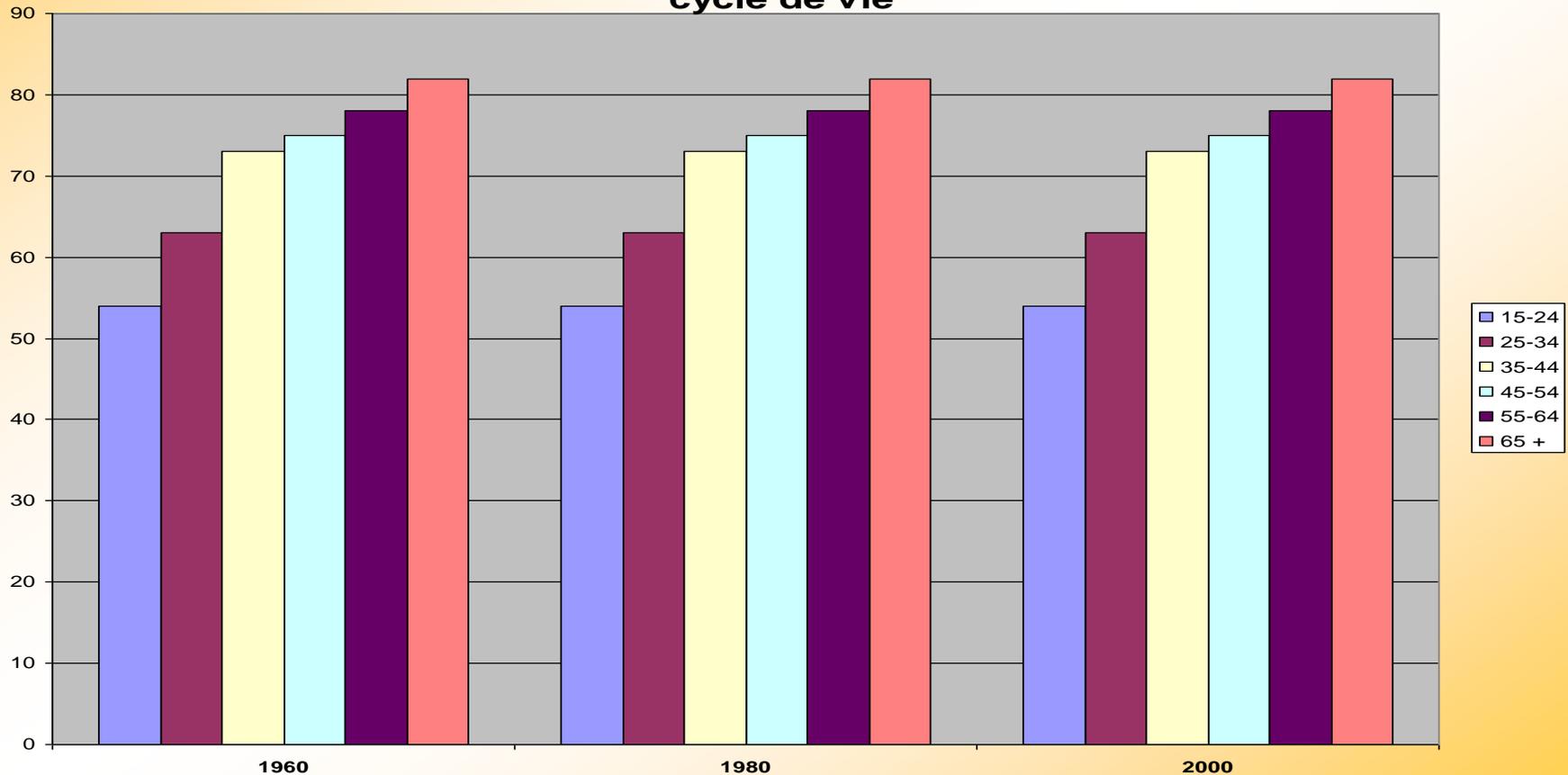
- Les deux sont possibles

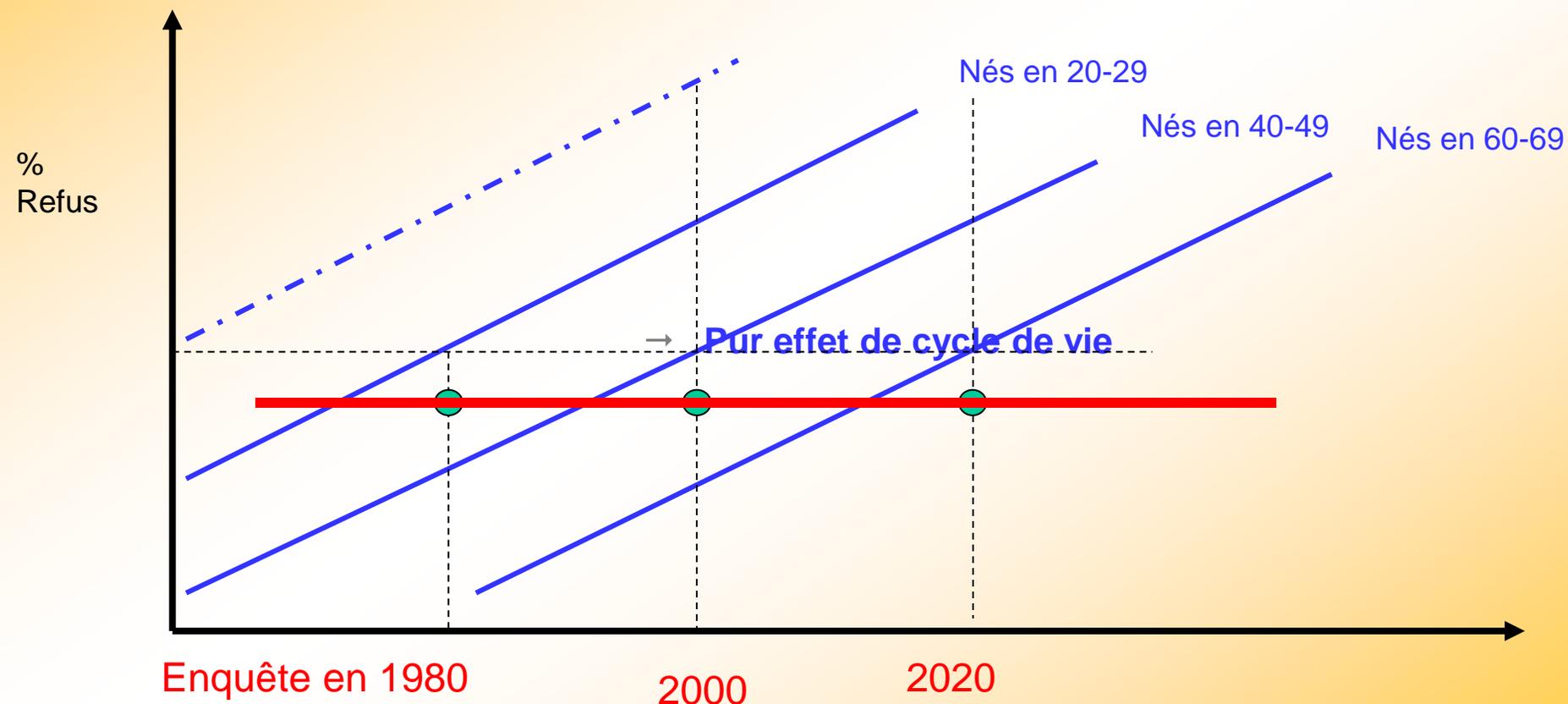
- **Cycle de vie** : en vieillissant, on devient moins tolérant à la différence ou aux attitudes minoritaires

- **Génération** : les plus âgés ont été socialisés dans une période où l'homosexualité était globalement beaucoup moins bien tolérée.

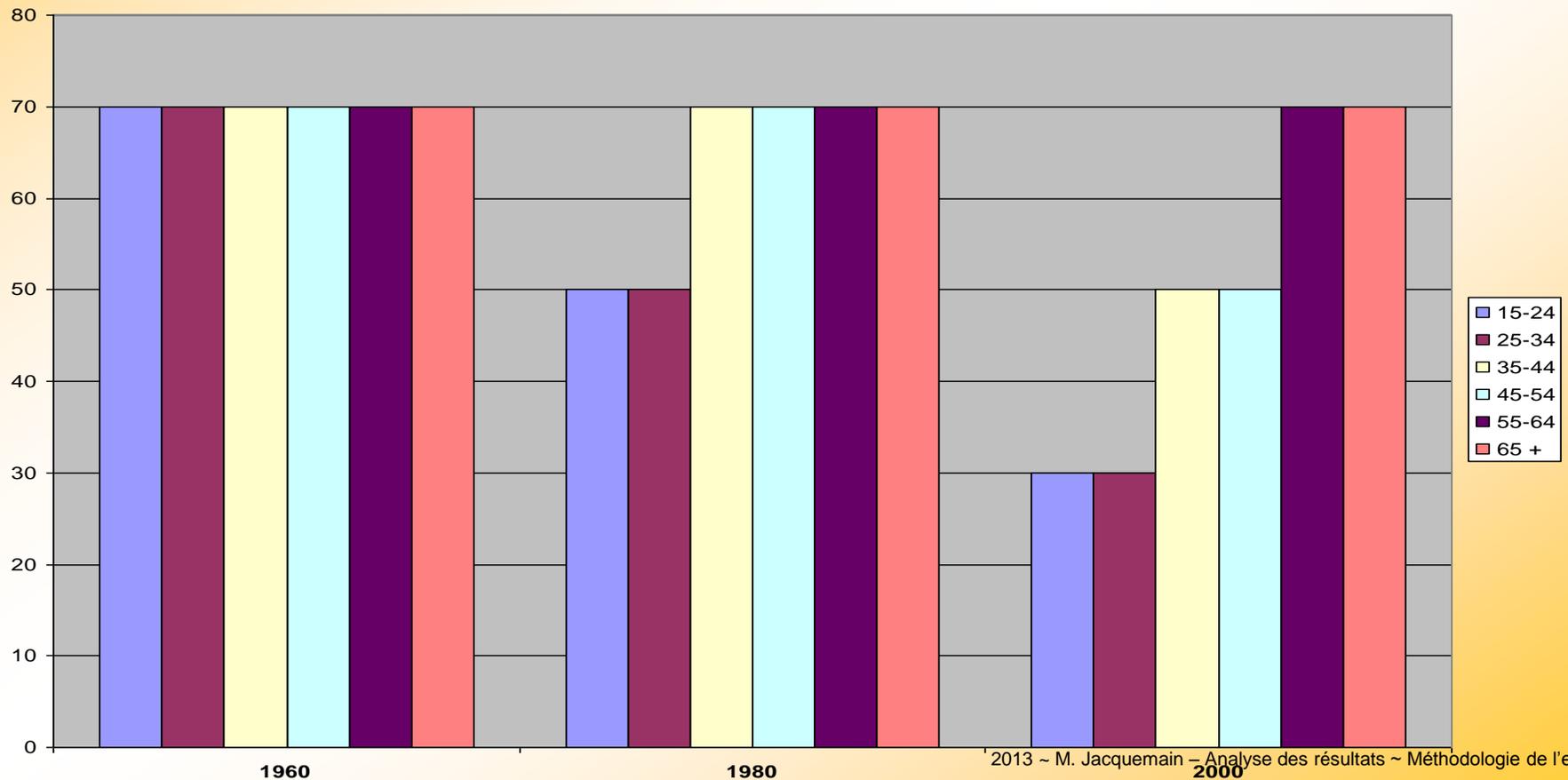
- Les deux effets peuvent donc s'additionner.

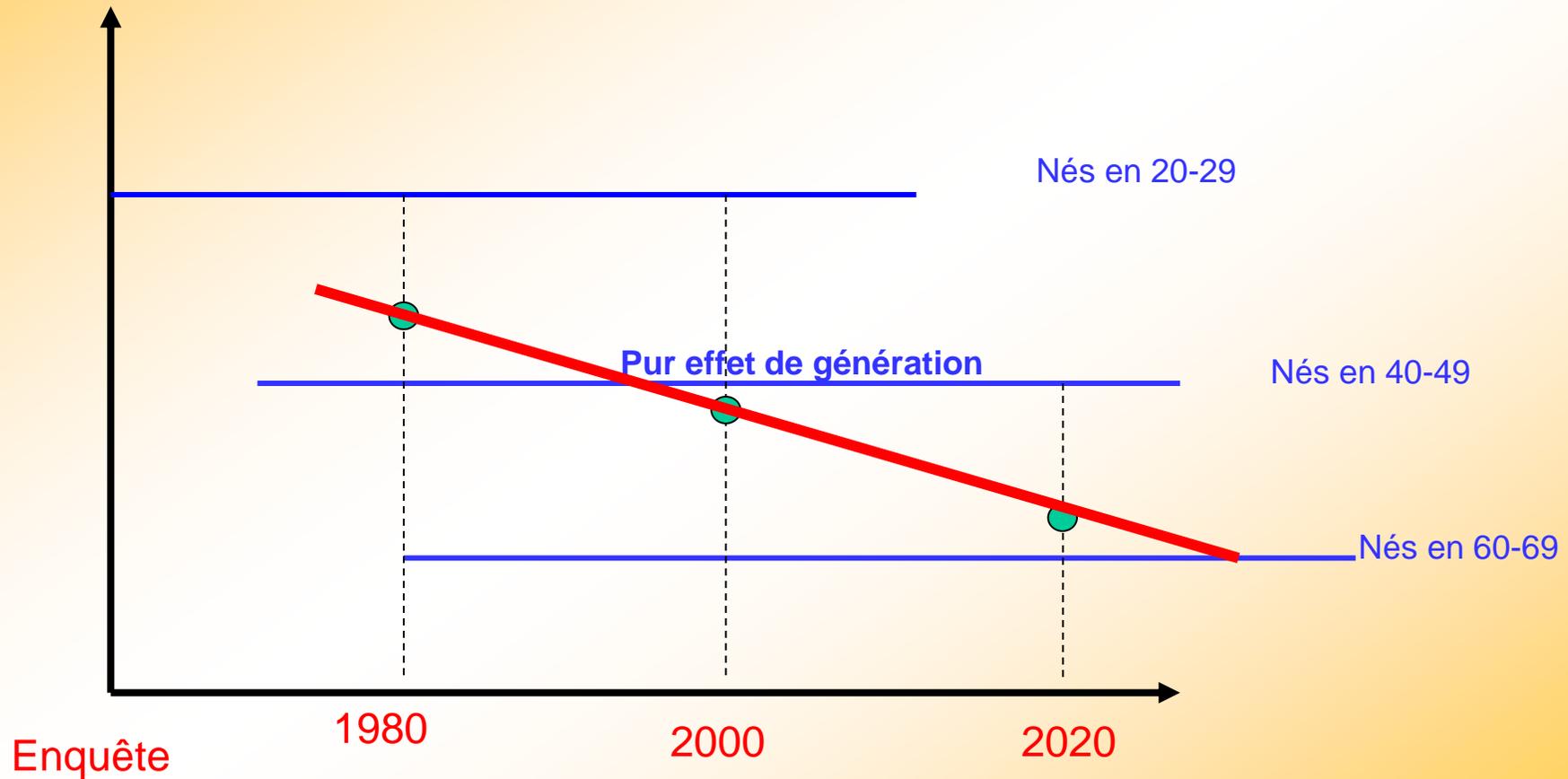
Refuseriez-vous un homosexuel comme voisin ? Pur effet de cycle de vie





Refus d'un homosexuel comme voisin Pur effet de génération





Refuseriez-vous un homosexuel comme voisin ? Effet de cycle de vie et de génération

