

MIASS 231
MATHÉMATIQUES
(APPLIQUÉES AUX SCIENCES SOCIALES) 3

© El Hadj Touré, Ph D. Sociologie
Section de sociologie, UGB de St-Louis

Leçon 2

Statistiques et variables

Au programme

- Comprendre ce que sont les statistiques et les variables en survolant les concepts statistiques de base
 - Statistique, statistiques et statistiques sociales
 - Passage des données aux statistiques
 - Statistiques descriptives vs statistiques inférentielles
 - Échantillon vs population; statistique vs paramètre
 - Classification des variables: quantitative vs qualitative
 - Mesure des variables : nominale, ordinale, intervalles, rapport
 - Distinction des variables quantitatives selon leur caractère discret ou continu

Statistique & statistiques

■ La Statistique

- Branche des mathématiques (discipline formelle) dont l'objet est le traitement méthodique de données numériques en vue de les décrire et d'en tirer des inférences (lois de probabilités)

■ Une statistique(s)

- Nombre décrivant une caractéristique d'une personne, un groupe, une chose ou d'un phénomène: note maximale MIASS 241 (20); taux de chômage au Sénégal en 2020-t4 (16.7%) selon l'ANSD

■ Les statistiques

- Techniques, méthodes et procédures utilisées pour calculer ces résumés numériques (nombres) et en dégager des considérations générales. C'est cette acception qui sous-tend ce cours

Statistiques & sciences sociales

■ Les statistiques sociales

- Techniques, méthodes et procédures utilisées pour résumer et généraliser des informations relatives aux phénomènes sociaux

■ Lien historique entre les statistiques et la sociologie

- Adolphe Quételet, l'inventeur de la «physique sociale» (1835)
 - ❖ Physique sociale = Sociologie = étude mathématique des phénomènes sociaux: le crime comme résultante de la pauvreté
 - ❖ À la recherche de l'«homme moyen» (indice de masse corp.)
- Émile Durkheim, le père de la « sociologie quantitative »
 - ❖ Les statistiques au cœur de l'administration de la preuve (1894)
 - ❖ Exemple d'analyse statistique appliquée aux phénomènes sociaux: le suicide dépend du degré de cohésion sociale (1897)

Données & statistiques

Définition & Comparaison

- Les données sont des matériaux bruts qui nécessitent un traitement pour qu'elles soient intelligibles
- Les statistiques sont des informations obtenues au sortir d'un traitement de données numériques
- La métaphore du photographe et de l'image offre une illustration intéressante du lien données/statistiques
 - Vous pouvez parler de vous-mêmes au photographe en mille mots, une simple image rend compte de l'essentiel de vous
 - Les statistiques peuvent être considérées comme l'image captée des données préalablement bien choisies

Données & statistiques

Microdonnées & données agrégées

- Les microdonnées sont des données non résumées
 - Ex: réponses à une enquête par questionnaire (liste de tous les participants avec leurs choix de réponses)
 - Les microdonnées n'ont pas encore fait l'objet d'un traitement
- Les données agrégées sont des données groupées ou résumées selon des critères précis d'ordre :
 - géographique | spatial (États, provinces, villes), temporel (intervalles d'années), sociologique (sexe, groupe d'âges...)
 - Si le critère est spatial, on a des données (agrégées) écologiques
 - Les données agrégées fournissent de l'information statistique

Statistiques descriptives & inférentielles

- Statistiques descriptives: l'objet est de décrire un ensemble de données, surtout d'échantillon, afin de les rendre plus intelligibles, communicables, utiles (réduction)
 - Ex: Pourcentages, moyenne, écart-type, score z...
 - Statistiques inférentielles: l'objet est d'inférer les informations sur un échantillon à une population en cherchant à savoir si elles sont significatives (généralisation)
 - Ex: Intervalle de confiance, marge d'erreur, test chi-carré, test t...
- «Les statistiques descriptives sont plus fondamentales que les statistiques inférentielles [certes]» (Fox, 1999: 7), mais les statistiques inférentielles constituent la finalité...

Population & échantillon

Population

- Ensemble d'individus, ayant une ou plusieurs caractéristiques communes, sur lesquels porte une étude
 - Les individus peuvent être des personnes, universités, pays, régions, cultures, dossiers criminels, élections, guerres, etc.
- La taille d'une population s'exprime par la lettre N
- Recensement = enquête menée auprès d'une population entière (recensement décennal sénégalais)
- « L'unité d'analyse est la personne, l'objet ou l'évènement que le chercheur étudie » (Fox: 5)

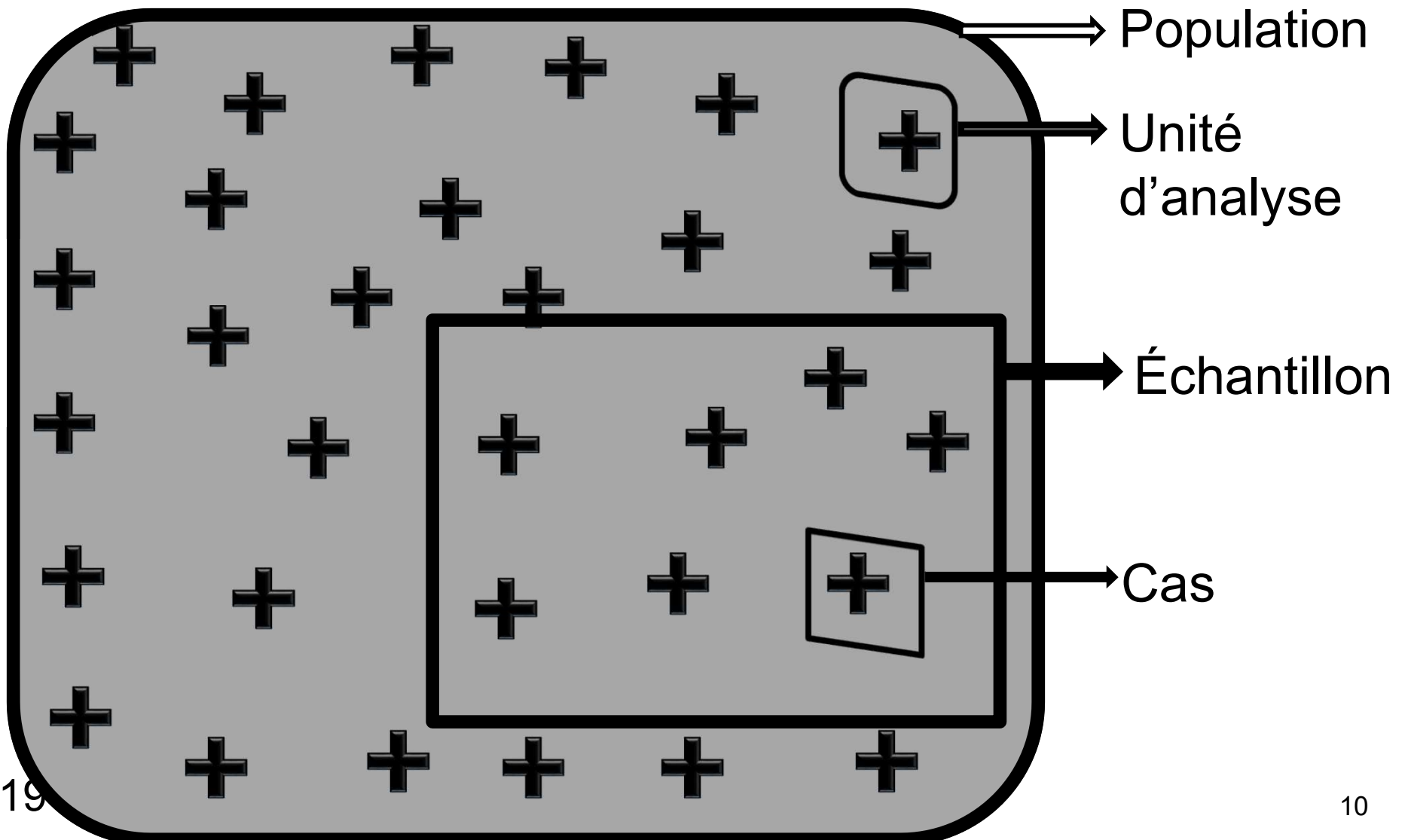
Population & échantillon

Échantillon

- Sous-ensemble tiré d'une population, à partir duquel l'on tente de généraliser les attributs à la population parente
 - Les chercheurs travaillent habituellement sur des échantillons
 - L'échantillon doit être le plus représentatif possible
- La taille d'un échantillon est désignée par la lettre n
- Sondage = Enquête menée auprès d'un échantillon (sondage électoral, sondage auprès des ménages)
- « Le cas est l'unité spécifique à propos de laquelle on collecte de l'information [statistique] » (Fox: 5)

Population & échantillon

Illustration des liens



Paramètre & statistique

Un paramètre

- Valeur calculée sur des données d'une population N
- Une lettre grecque désigne habituellement un paramètre:
 - ❖ μ = moyenne (mu)
 - ❖ σ = écart type (sigma)
 - ❖ π = proportion (pi)
- Par ex., l'espérance de vie moyenne au Sénégal μ était de 67.9 ans en 2019. Cette valeur est un paramètre, car elle est calculée sur des données de recensement

Paramètre & statistique

Une statistique

- Valeur calculée sur des données d'un échantillon n
- Une lettre romaine désigne habituellement une statistique:
 - ❖ \bar{X} = moyenne (X barre)
 - ❖ s = écart type
 - ❖ p = proportion
- Par ex., un sondage auprès de $n = 100$ cas peut estimer le revenu mensuel moyen des Sénégalais \bar{X} à 90 000 Fcfa. Cette valeur est une statistique, car elle est calculée sur des données de sondage

Variable & constante

Une variable

- Caractéristique ou propriété qui varie en fonction...
 - des personnes : sexe, situation matrimoniale, âge, etc.
 - des universités: nombre d'étudiants, de professeurs, de départements, taux moyen de réussite, etc.
 - des pays ou provinces: revenu par tête d'habitant, nombre d'habitants, taux de criminalité, etc.
 - des dossiers criminels: nature des crimes, récidive, etc.
- «Une caractéristique constitue une variable si elle possède au moins deux valeurs ou attributs» (Fox: 5)
 - Sexe: féminin, masculin | âge: 30 ans, 31ans, etc.

Variable & constante

Une constante

- Caractéristique qui a une valeur fixe ou une seule valeur
 - Pour les étudiants du cours MIASS 231, les données ci-dessous sont des constantes:
 - ❖ Section= Sociologie
 - ❖ Université d'appartenance = Université Gaston Berger
 - ❖ Région d'appartenance?
- Un même concept peut être une variable dans une étude et une constante dans une autre étude
- Les statistiques (sociales) opèrent habituellement sur des variables et non sur des constantes

Variables quantitative & qualitative

- Une variable quantitative est une variable dont les valeurs possibles sont des nombres (quantité)
 - Ex: revenu en dollars, scolarité en années
- Une variable est dite qualitative lorsque ses valeurs sont plutôt des catégories ou modalités (qualité)
 - Ex: sexe, statut matrimonial
- Une variable quantitative peut être transformée en variable qualitative
 - Ex: revenu en classes de dollars (faible; moyenne; élevée)
- Certains outils statistiques conviennent à des variables quantitatives, d'autres à des variables qualitatives

Population, échantillon, variable & constante

Remue-méninges

→ Un sociologue du crime veut étudier la nature de tous les crimes enregistrés en 2020 par le service de police de Dakar. Il s'intéresse à tous les dossiers criminels

1. Quelle est la population étudiée?
2. Quelle est l'unité statistique (unité d'analyse)?
3. L'étude est-elle effectuée par recensement ou sondage?
4. Quelle est la variable centrale ici? Est-elle quantitative ou qualitative? Quelles sont ses valeurs possibles?
5. Y a-t-il une ou des constantes?

Échelles de mesure

Définition & classification

- « La mesure est la procédure qui nous permet de trouver les valeurs d'une variable pour des cas différents» (Fox:9)
 - Une échelle est la série des valeurs possibles d'une variable X
 - Les valeurs mesurées dans les faits sont nommées scores X_i

- La façon dont les variables sont mesurées détermine le choix approprié des statistiques: 4 échelles de mesure
 - Variable nominale
 - Variable ordinale
 - Variable d'intervalles
 - Variable de rapport

}	Qualitative ou catégorielle
}	Quantitative ou métrique

Échelles de mesure

Une variable nominale

Caractéristiques

- ❖ Les valeurs sont des catégories nommées, non ordonnées
- ❖ L'assignation des nombres aux catégories est arbitraire et n'obéit à aucun ordre (classification)

Origine (catégorie)

Aucune (pas de premier rang)

Échelons (gradation)

Aucun

Nombres (nature)

Repères (assignation, identification)

Relation entre les valeurs

Aucune (pas d'opérations arithmétiques)

Exemples

Sexe (1. Femme; 2. Homme), statut civil, appartenance politique, religion

Échelles de mesure

Une variable ordinale

Caractéristiques

- ❖ Les valeurs sont nommées et surtout résultent d'une mise en ordre
- ❖ Elle a un niveau de mesure supérieur à la variable nominale

Origine (catégorie)

Premier rang

Échelons (gradation)

Inégaux

Nombres (nature)

Ordre, hiérarchisation (pas de quantité)

Relation entre les valeurs

< , = , >

Mais pas d'opérations arithmétiques

Niveau d'études (1. primaire; 2. secondaire; 3. postsecondaire), assiduité (toujours, souvent, parfois, jamais), revenu en classes

Exemples

Échelles de mesure

Une variable d'intervalles

Caractéristiques

- ❖ Les valeurs peuvent être nommées et ordonnées
- ❖ Requiert une unité de mesure standard et accepte les différences d'intervalles

Origine (point zéro)

Zéro arbitraire, relatif (convention)

Échelons (gradation)

Égaux (égalité relative)

Nombres (nature)

Valeurs métriques (quantité)

Relation entre les valeurs

$<$, $=$, $>$ mais aussi $-$, $+$

Toutes les opérations sauf les rapports

Exemples

Calendrier grégorien, musulman (en années), température (en °C)

Échelles de mesure

Une variable de ratio (rapport ou proportion)

Caractéristiques

- ❖ Se mesure grâce à une unité standard et surtout accepte les rapports
- ❖ Les variables ratios ont le niveau de mesure le plus élevé

Origine (point zéro)

Zéro absolu, non arbitraire

Échelons (gradation)

Égaux (égalité absolue)

Nombres (nature)

Valeurs métriques (quantité)

Relation entre les valeurs

$<$, $=$, $>$; $-$, $+$ mais aussi $/$, $*$

Toutes les opérations arithmétiques

Exemples

Âge en années, revenu en dollars, scolarité en années

Échelles de mesure

Comparaison & remue-méninges

Type de variables	Échelles de mesure	Les valeurs servent-elles à nommer, identifier?	Y a-t-il un ordre intrinsèque entre les valeurs?	Y a-t-il une unité de mesure (diff. d'intervalles)?	Peut-on calculer des rapports?
Qualitative	Nominale	Oui	Non	Non	Non
	Ordinale	Oui	Oui	Non	Non
Quantitative	Intervalles	Oui	Oui	Oui	Non
	Ratio	Oui	Oui	Oui	Oui

→ Interprétez les résultats obtenus par trois étudiants, en termes nominal, ordinal, intervallaire, proportionnel

– Bineta = 100%; Mamadou = 85%; Fatou = 50%

Variables discrète & continue

Une variable discrète

- Quantité qui ne peut prendre qu'un nombre restreint de valeurs (séparées) sur la droite des nombres réels
- Elle résulte ordinairement d'un dénombrement (0;1;2...)
 - Nombre d'enfants par famille
 - Nombre d'amis d'une origine ethnique différente
 - Nombre de buts marqués dans un match de football
 - Nombre de départements dans une faculté
 - Nombre de cours choisis annuellement par les étudiants
- Habituellement, une variable discrète est quantitative

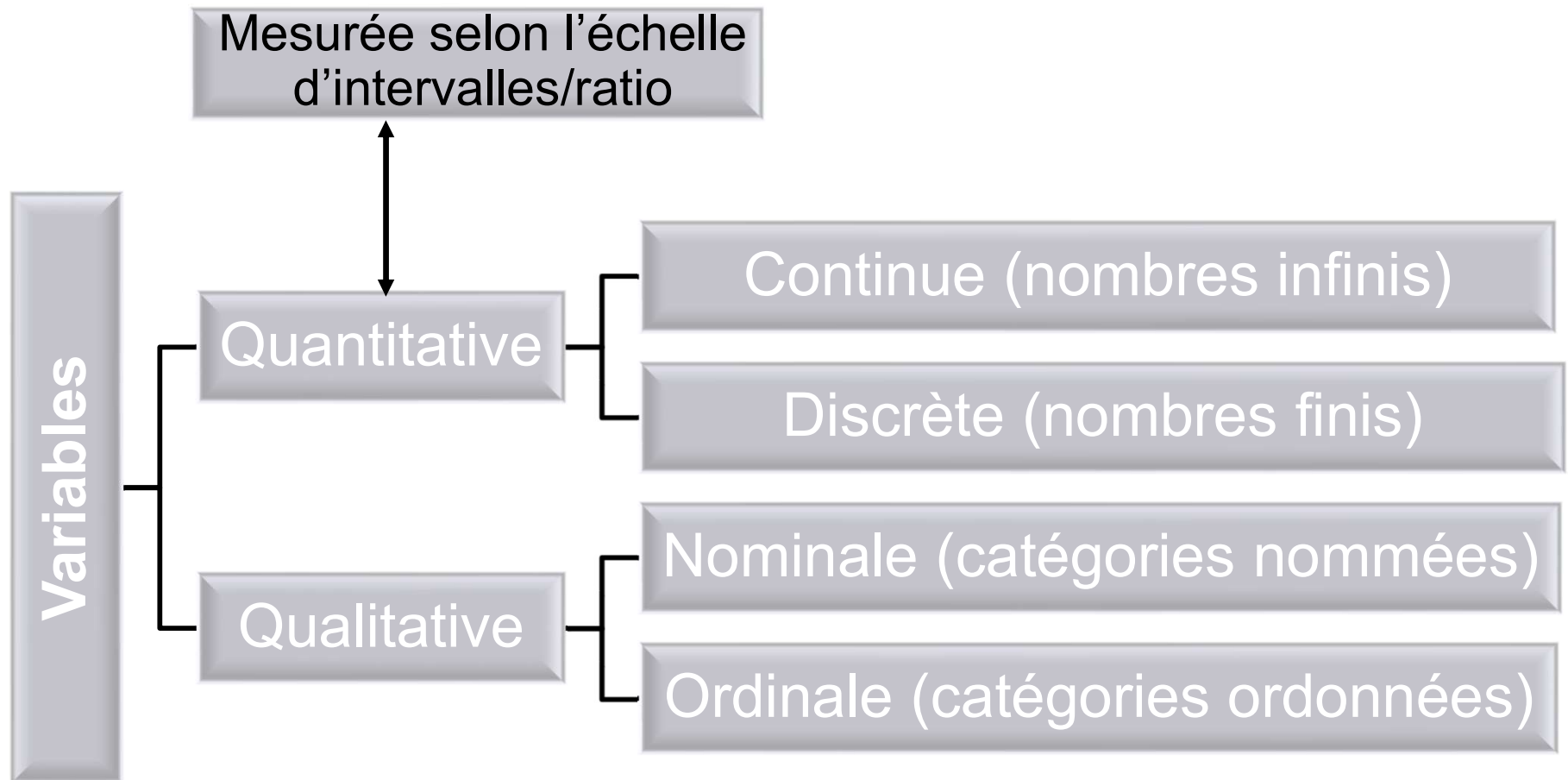
Variables discrète & continue

Une variable continue

- Quantité qui peut prendre n'importe quelle valeur (avec décimales) dans un intervalle donné de nombres réels
- Valeur exacte
 - Valeur précise à l'infini que l'on ne connaîtra jamais
 - Ex: la taille = 1,782394873627847362... m
- Valeur rapportée
 - Valeur produite par un procédé de mesure
 - Ex: la taille = 1,78 m
- Habituellement, une variable continue est quantitative

Classification des variables

Synthèse



Classification des variables

Remue-méninges

- Nommez les variables suivantes (étiquette de l'indicateur) et dites laquelle est discrète, continue, nominale, ordinale. Justifiez!
- Depuis combien de temps êtes-vous marié(e)?
 - Combien de membres compte-t-elle votre famille?
 - La religion est-elle importante? (très, assez, peu, pas du tout)
 - À quelle religion appartenez-vous?

Classification des variables

Remue-méninges (2)

- Un chercheur veut étudier la consommation de drogue chez les jeunes. Quelles sont les questions qu'il peut poser pour obtenir des informations sur l'importance de la consommation?
- Variable qualitative nominale (précisez les valeurs)
 - Variable qualitative ordinale (précisez les valeurs)
 - Variable quantitative continue (précisez les valeurs)
 - Variable quantitative discrète (précisez les valeurs)

Classification des variables

Caractéristiques d'une variable qualitative

→ Les catégories d'une variable doivent être mutuellement exclusives et collectivement exhaustives (Fox:16-17)

■ Mutuellement exclusives

- Chaque cas observé tombe dans une et une seule catégorie
 - ❖ Un individu ne peut pas être à la fois homme et femme

■ Collectivement exhaustives

- L'ensemble des catégories inclut tous les cas observés, chaque cas tombant dans une catégorie
- La valeur résiduelle « autre » assure ce principe d'exhaustivité

Tout prochainement

- À faire cette semaine
 - Complétez le quiz 2 (obligatoire)
 - Réalisez les 10 exercices récapitulatifs et vérifiez les solutions en vous référant au corrigé (optionnel)
- Prochaine leçon
 - Distribution de fréquences et de pourcentages