

Méthodes quantitatives en sciences sociales

cours de Reto Schumacher

Statistique pour sciences sociales
printemps 2012

Plan

- 1 Méthodes quantitatives en sciences sociales
- 2 Analyse statistique
- 3 Concepts de base

Plan

- 1 Méthodes quantitatives en sciences sociales
- 2 Analyse statistique
- 3 Concepts de base

Approches empiriques en sciences sociales

- **en sciences naturelles**, les chercheurs ont recours à toute une panoplie de méthodes de collecte et d'analyse de données
 - ▷ les astronomes utilisent des télescopes
 - ▷ les biologistes ont recours à la microscopie
 - ▷ les physiciens et les chimistes utilisent des appareils de mesure sophistiqués et font des expériences reproductibles
 - ▷ les météorologues analysent et interprètent les données fournies par les satellites

Approches empiriques en sciences sociales

- **en sciences naturelles**, les chercheurs ont recours à toute une panoplie de méthodes de collecte et d'analyse de données
 - ▷ les astronomes utilisent des télescopes
 - ▷ les biologistes ont recours à la microscopie
 - ▷ les physiciens et les chimistes utilisent des appareils de mesure sophistiqués et font des expériences reproductibles
 - ▷ les météorologues analysent et interprètent les données fournies par les satellites
- **et en sciences sociales ? ?**
 - ▷ les sociologues effectuent des sondages, interrogent les gens, analysent le contenu des médias etc.
 - ▷ les politologues analysent les comportements électoraux en utilisant des données de sondages, analysent les discours politiques, évaluent les politiques publiques etc.
 - ▷ les économistes analysent des données administratives, de même que des données d'enquêtes, ont également recours à la recherche expérimentale etc.

Traditions de recherche en sciences sociales

- la recherche en sciences sociales est majoritairement **empirique** ou **inductive** (\neq déductive) : une hypothèse (év. formulée de façon déductive) est confrontée à l'**observation**, dont les résultats sont généralisés par induction

Traditions de recherche en sciences sociales

- la recherche en sciences sociales est majoritairement **empirique** ou **inductive** (\neq déductive) : une hypothèse (év. formulée de façon déductive) est confrontée à l'**observation**, dont les résultats sont généralisés par induction
- depuis le 17e siècle un antagonisme entre deux courants de recherche empirique :

Traditions de recherche en sciences sociales

- la recherche en sciences sociales est majoritairement **empirique** ou **inductive** (\neq déductive) : une hypothèse (év. formulée de façon déductive) est confrontée à l'**observation**, dont les résultats sont généralisés par induction
- depuis le 17e siècle un antagonisme entre deux courants de recherche empirique :

statistique universelle

(Universalstatistik) :

mémoires (littéraires)

synthétisant la situation de l'Etat



Traditions de recherche en sciences sociales

- la recherche en sciences sociales est majoritairement **empirique** ou **inductive** (\neq déductive) : une hypothèse (év. formulée de façon déductive) est confrontée à l'**observation**, dont les résultats sont généralisés par induction
- depuis le 17e siècle un antagonisme entre deux courants de recherche empirique :

statistique universelle

(Universalstatistik) :

mémoires (littéraires)

synthétisant la situation de l'Etat



statistiques départementales

(Napoléon 1er)



Traditions de recherche en sciences sociales

- la recherche en sciences sociales est majoritairement **empirique** ou **inductive** (\neq déductive) : une hypothèse (év. formulée de façon déductive) est confrontée à l'**observation**, dont les résultats sont généralisés par induction
- depuis le 17e siècle un antagonisme entre deux courants de recherche empirique :

statistique universelle

(Universalstatistik) :

mémoires (littéraires)

synthétisant la situation de l'Etat



statistiques départementales

(Napoléon Ier)



approche qualitative

Traditions de recherche en sciences sociales

- la recherche en sciences sociales est majoritairement **empirique** ou **inductive** (\neq déductive) : une hypothèse (év. formulée de façon déductive) est confrontée à l'**observation**, dont les résultats sont généralisés par induction
- depuis le 17e siècle un antagonisme entre deux courants de recherche empirique :

statistique universelle

(Universalstatistik) :
mémoires (littéraires)

synthétisant la situation de l'Etat



statistiques départementales
(Napoléon Ier)



approche qualitative

arithmétique politique Graunt

(1662), Petty (1690) :
étude quantitative des
phénomènes sociaux



Traditions de recherche en sciences sociales

- la recherche en sciences sociales est majoritairement **empirique** ou **inductive** (\neq déductive) : une hypothèse (év. formulée de façon déductive) est confrontée à l'**observation**, dont les résultats sont généralisés par induction
- depuis le 17e siècle un antagonisme entre deux courants de recherche empirique :

statistique universelle

(Universalstatistik) :
mémoires (littéraires)

synthétisant la situation de l'Etat



statistiques départementales
(Napoléon Ier)



approche qualitative

arithmétique politique Graunt

(1662), Petty (1690) :
étude quantitative des
phénomènes sociaux



statistique morale, théorie
probabiliste, Quetelet (1848)



Traditions de recherche en sciences sociales

- la recherche en sciences sociales est majoritairement **empirique** ou **inductive** (\neq déductive) : une hypothèse (év. formulée de façon déductive) est confrontée à l'**observation**, dont les résultats sont généralisés par induction
- depuis le 17e siècle un antagonisme entre deux courants de recherche empirique :

statistique universelle

(Universalstatistik) :
mémoires (littéraires)

synthétisant la situation de l'Etat



statistiques départementales
(Napoléon Ier)



approche qualitative

arithmétique politique Graunt

(1662), Petty (1690) :
étude quantitative des
phénomènes sociaux



statistique morale, théorie
probabiliste, Quetelet (1848)



approche quantitative

Rôle des méthodes quantitatives en sciences sociales

- cet antagonisme persiste jusqu'à nos jours (les "quali" contre les "quanti"), même s'il existe des tentatives de synthèse

Rôle des méthodes quantitatives en sciences sociales

- cet antagonisme persiste jusqu'à nos jours (les "quali" contre les "quanti"), même s'il existe des tentatives de synthèse
- notamment dans le monde anglosaxon, le courant quantitativiste reste dominant :

Rôle des méthodes quantitatives en sciences sociales

- cet antagonisme persiste jusqu'à nos jours (les "quali" contre les "quanti"), même s'il existe des tentatives de synthèse
- notamment dans le monde anglosaxon, le courant quantitativiste reste dominant :

Revue	période	orientation quantitative (%)	(n)
<i>International organization</i>	2008-2009	64	45
<i>International Studies Quarterly</i>	2009	60	48

Rôle des méthodes quantitatives en sciences sociales

- cet antagonisme persiste jusqu'à nos jours (les "quali" contre les "quanti"), même s'il existe des tentatives de synthèse
- notamment dans le monde anglosaxon, le courant quantitativiste reste dominant :

Revue	période	orientation quantitative (%)	(n)
<i>International organization</i>	2008-2009	64	45
<i>International Studies Quarterly</i>	2009	60	48
<i>American Political Science Review</i>	2009	84	37
<i>American Journal of Political Science</i>	2009	97	60

Rôle des méthodes quantitatives en sciences sociales

- cet antagonisme persiste jusqu'à nos jours (les "quali" contre les "quanti"), même s'il existe des tentatives de synthèse
- notamment dans le monde anglosaxon, le courant quantitativiste reste dominant :

Revue	période	orientation quantitative (%)	(n)
<i>International organization</i>	2008-2009	64	45
<i>International Studies Quarterly</i>	2009	60	48
<i>American Political Science Review</i>	2009	84	37
<i>American Journal of Political Science</i>	2009	97	60
<i>American Journal of Sociology</i>	mi08-mi09	83	36

Rôle des méthodes quantitatives en sciences sociales

- cet antagonisme persiste jusqu'à nos jours (les "quali" contre les "quanti"), même s'il existe des tentatives de synthèse
- notamment dans le monde anglosaxon, le courant quantitativiste reste dominant :

Revue	période	orientation quantitative (%)	(n)
<i>International organization</i>	2008-2009	64	45
<i>International Studies Quarterly</i>	2009	60	48
<i>American Political Science Review</i>	2009	84	37
<i>American Journal of Political Science</i>	2009	97	60
<i>American Journal of Sociology</i>	mi08-mi09	83	36

⇒ **négliger les méthodes quantitatives, c'est se priver d'un courant puissant des sciences sociales**

et les étudiants ?

Voici quelques résultats d'un questionnaire soumis aux étudiants de 1ère année en sciences sociales (BA RI, BA SO etc.)(2009) :

et les étudiants ?

Voici quelques résultats d'un questionnaire soumis aux étudiants de 1ère année en sciences sociales (BA RI, BA SO etc.)(2009) :

- Partagez-vous les affirmations suivantes : disposer de connaissances en statistique est
 - ▷ nécessaire pour mener une recherche en sciences sociales

et les étudiants ?

Voici quelques résultats d'un questionnaire soumis aux étudiants de 1ère année en sciences sociales (BA RI, BA SO etc.)(2009) :

- Partagez-vous les affirmations suivantes : disposer de connaissances en statistique est
 - ▷ nécessaire pour mener une recherche en sciences sociales
 - non : 3 %
 - en partie : 25 %
 - oui : 72 %

et les étudiants ?

Voici quelques résultats d'un questionnaire soumis aux étudiants de 1ère année en sciences sociales (BA RI, BA SO etc.)(2009) :

- Partagez-vous les affirmations suivantes : disposer de connaissances en statistique est
 - ▷ nécessaire pour mener une recherche en sciences sociales
 - non : 3 %
 - en partie : 25 %
 - oui : 72 %
 - ▷ nécessaire pour bien comprendre les rapports de recherche et articles de revue qui s'appuient sur des études empiriques

et les étudiants ?

Voici quelques résultats d'un questionnaire soumis aux étudiants de 1ère année en sciences sociales (BA RI, BA SO etc.)(2009) :

- Partagez-vous les affirmations suivantes : disposer de connaissances en statistique est
 - ▷ nécessaire pour mener une recherche en sciences sociales
 - non : 3 %
 - en partie : 25 %
 - oui : 72 %
 - ▷ nécessaire pour bien comprendre les rapports de recherche et articles de revue qui s'appuient sur des études empiriques
 - non : 3 %
 - en partie : 36 %
 - oui : 61 %

et les étudiants ?

- Êtes-vous intéressé par l'analyse statistique ?

et les étudiants ?

- Êtes-vous intéressé par l'analyse statistique ?
 - non : 63 %
 - oui : 37 %

et les étudiants ?

- Êtes-vous intéressé par l'analyse statistique ?
 - non : 63 %
 - oui : 37 %
- Suivriez-vous un cours de statistique si celui-ci n'était pas obligatoire ?

et les étudiants ?

- Êtes-vous intéressé par l'analyse statistique ?
 - non : 63 %
 - oui : 37 %
- Suivriez-vous un cours de statistique si celui-ci n'était pas obligatoire ?
 - non : 80 %
 - oui : 20 %

et les étudiants ?

- Êtes-vous intéressé par l'analyse statistique ?

non : 63 %

oui : 37 %

- Suivriez-vous un cours de statistique si celui-ci n'était pas obligatoire ?

non : 80 %

oui : 20 %

→ **l'approche statistique est chargée de préjugés négatifs**

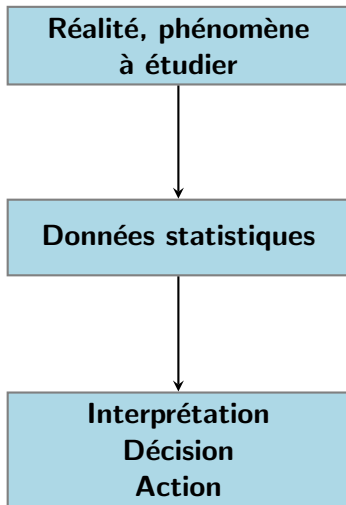
Pourquoi quantifier ?

- Objets des sciences sociales
 - ▷ organisation de la société, dynamique de groupe
 - ▷ comportement, tradition, norme sociale
 - ▷ ...
- Concepts qualitatifs \implies Pourquoi quantifier ?

Pourquoi quantifier ?

- Objets des sciences sociales
 - ▷ organisation de la société, dynamique de groupe
 - ▷ comportement, tradition, norme sociale
 - ▷ ...
- Concepts qualitatifs \implies Pourquoi quantifier ?
- **Trois raisons principales :**
 - ▷ Description : **pouvoir synthétique des chiffres**
 - ▷ Observation : **objectivité**
 - Décrire comment mesurer implique de spécifier le concept étudié
 - Une fois défini comment mesurer, tout le monde devrait trouver le même résultat
 - ▷ Traitement de l'information : **efficacité**

Qu'est-ce que la statistique ?



Observation statistique

Collecter données existantes,
Enquête, Expérimentation

Analyse statistique

Statistique descriptive
Statistique inférentielle

Observation statistique

- Expérience : répéter une « épreuve » dans un environnement contrôlé et relever les résultats
- Enquête : observations factuelles auprès d'un ensemble de sujets

Observation statistique

- Expérience : répéter une « épreuve » dans un environnement contrôlé et relever les résultats
- Enquête : observations factuelles auprès d'un ensemble de sujets

En sciences sociales

- Pas d'expériences (environnement non contrôlable)
- Observation de faits de la vie courante dans le temps et/ou l'espace et/ou parmi un ensemble d'individus. On procède en général par enquêtes

Saisie et sources de données : Enquêtes exhaustives et partielles

- **Enquête exhaustive**, recensement :
 - ▷ précis, exhaustif, mais coûteux et lourd

Saisie et sources de données : Enquêtes exhaustives et partielles

- **Enquête exhaustive**, recensement :
 - ▷ précis, exhaustif, mais coûteux et lourd
- **Enquête partielle**, sondage :
 - ▷ souple, permet plus facilement d'obtenir des données spécifiques
 - ▷ problème du choix de l'échantillon, de sa représentativité

Saisie et sources de données : Données primaires et secondaires

- **Statistiques primaires**
 - ▷ obtenues par des activités dont le but premier est la collecte de données.
 - ▷ exemples : enquêtes de l'OFS, panel de ménages, sondages, ...

Saisie et sources de données : Données primaires et secondaires

- **Statistiques primaires**
 - ▷ obtenues par des activités dont le but premier est la collecte de données.
 - ▷ exemples : enquêtes de l'OFS, panel de ménages, sondages, ...
- **Statistiques secondaires ou administratives**
 - ▷ résultant d'autres activités administratives
 - ▷ exemples : assurances, rapports de police, douane, registres d'état-civil, fisc, ...

Plan

- 1 Méthodes quantitatives en sciences sociales
- 2 Analyse statistique**
- 3 Concepts de base

Statistique descriptive et statistique inférentielle

- **Statistique descriptive**
 - ▷ Décrire les données

Statistique descriptive et statistique inférentielle

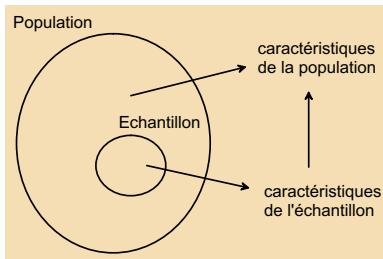
- **Statistique descriptive**
 - ▷ Décrire les données

- **Statistique inférentielle**
 - ▷ Tirer des enseignements sur la « réalité » (population, phénomène) d'où proviennent les données
 - ▷ Évaluer la fiabilité des résultats

Statistique descriptive

- Organiser et analyser les données pour en dégager la structure et en donner des présentations synthétiques :
 - ▷ tableaux (de fréquences)
 - ▷ graphiques
 - ▷ résumés numériques
- Permet de cerner les **caractéristiques des données**
- Parfois suffisant pour éclairer la réalité étudiée :
- Exemple :
 - ▷ pyramide des âges ;
 - ▷ répartition des accidents de la route selon leur cause

Statistique inférentielle



- En général, les données disponibles ne portent que sur
 - ▷ une partie (échantillon) de la population
 - ▷ quelques manifestations d'un phénomène
- On observe une partie, mais on s'intéresse au tout (population, phénomène)

Exemple d'inférence statistique

- Sondage à la veille de votations
 - ▷ On observe le taux f d'acceptation des personnes interrogées
 - ▷ On veut connaître le taux p pour l'ensemble de la population

Exemple d'inférence statistique

- Sondage à la veille de votations
 - ▷ On observe le taux f d'acceptation des personnes interrogées
 - ▷ On veut connaître le taux p pour l'ensemble de la population
- On **infère** ou **induit** les caractéristiques du tout à partir de l'observation d'une partie

Exemple d'inférence statistique

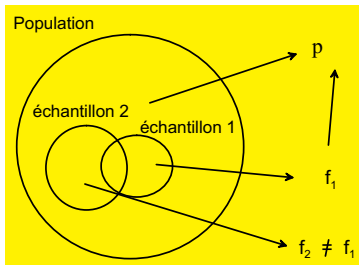
- Sondage à la veille de votations
 - ▷ On observe le taux f d'acceptation des personnes interrogées
 - ▷ On veut connaître le taux p pour l'ensemble de la population
- On **infère** ou **induit** les caractéristiques du tout à partir de l'observation d'une partie
- Le **résultat de l'induction dépend de l'échantillon observé**
⇒ **Incertitude**

Exemple d'inférence statistique

- Sondage à la veille de votations
 - ▷ On observe le taux f d'acceptation des personnes interrogées
 - ▷ On veut connaître le taux p pour l'ensemble de la population
- On **infère** ou **induit** les caractéristiques du tout à partir de l'observation d'une partie
- Le **résultat de l'induction dépend de l'échantillon observé**
⇒ **Incertitude**
- On doit évaluer cette incertitude

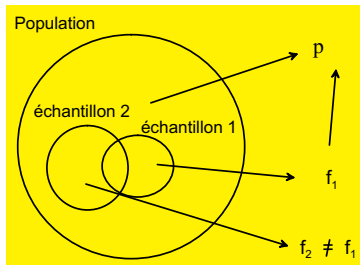
Exemple d'inférence statistique, suite

- On peut estimer p avec la valeur observée f



Exemple d'inférence statistique, suite

- On peut estimer p avec la valeur observée f



- Estimations différentes selon l'échantillon
- Exemple** : Population de taille $m = 10'000$ dont 4'000 acceptants Sur 100 personnes interrogées, on peut avoir
 - ▷ 100 acceptants ($f = 1$) ou
 - ▷ 0 acceptants ($f = 0$) ou
 - ▷ tout nombre $\in [0, 100]$

Evaluer la confiance

- Il est indispensable d'**évaluer la confiance** dans le résultat
- Une estimation ou un test d'hypothèse est sans intérêt si l'on ne sait rien sur sa validité

Evaluer la confiance

- Il est indispensable d'**évaluer la confiance** dans le résultat
- Une estimation ou un test d'hypothèse est sans intérêt si l'on ne sait rien sur sa validité
- La confiance est déterminante pour :
 - ▷ le choix entre plusieurs résultats (on préfère un taux estimé à partir d'un échantillon de taille 500 plutôt que 50)
 - ▷ la prise de décision

Évaluer la confiance

- Il est indispensable d'**évaluer la confiance** dans le résultat
 - Une estimation ou un test d'hypothèse est sans intérêt si l'on ne sait rien sur sa validité
 - La confiance est déterminante pour :
 - ▷ le choix entre plusieurs résultats (on préfère un taux estimé à partir d'un échantillon de taille 500 plutôt que 50)
 - ▷ la prise de décision
 - L'évaluation de la confiance se fonde sur le **calcul de probabilités**
- ⇒ **Résultats statistiquement significatifs** :
- Résultat statistique **significatif** si la probabilité qu'il soit du au seul hasard (choix d'un échantillon atypique) est faible

Deux aspects de la statistique inférentielle

- **Statistique inférentielle**

- ▷ Branche de la statistique qui s'occupe des problèmes de l'inférence
- ▷ en particulier de l'évaluation de l'incertitude

Deux aspects de la statistique inférentielle

- **Statistique inférentielle**

- ▷ Branche de la statistique qui s'occupe des problèmes de l'inférence
- ▷ en particulier de l'évaluation de l'incertitude

- La statistique inférentielle comprend deux volets :

- ▷ **Estimation** : quantification de paramètres numériques
- ▷ **Test d'hypothèses** : validation d'énoncés sur la réalité étudiée

Plan

- 1 Méthodes quantitatives en sciences sociales
- 2 Analyse statistique
- 3 Concepts de base**

Objets statistiques et attributs

- **Objet**, unité statistique, individu, observation, cas (qui ?)
 - ▷ Unité primaire d'intérêt
 - ▷ Exemples : l'individu, la famille, le pays, la personne-période, . . .

Objets statistiques et attributs

- **Objet**, unité statistique, individu, observation, cas (qui ?)
 - ▷ Unité primaire d'intérêt
 - ▷ Exemples : l'individu, la famille, le pays, la personne-période, ...
- **Attributs**, variables : les caractéristiques observées (quoi ?).
 - ▷ Exemples : Sexe, nationalité, âge de l'individu.
salaire mensuel, dépenses de consommation, profit, ...

Population et échantillon

- **Population**, processus
 - ▷ Ensemble de toutes les unités primaires, (réalisations possibles du processus).
 - ▷ Peut être infini (infinité d'instants au cours du temps)

Population et échantillon

- **Population**, processus
 - ▷ Ensemble de toutes les unités primaires, (réalisations possibles du processus).
 - ▷ Peut être infini (infinité d'instants au cours du temps)

- **Échantillon**
 - ▷ Sous-ensemble observé de la population
 - ▷ Échantillon aléatoire (si tous les cas ont même probabilité d'être sélectionné)

Modalités d'une variable

- **Modalités** : différents états possibles de la variable

Exemples :

variable	modalités
canton	ZH, BE, ... , GE, JU
mil. soc.	populaire, cadre moyen, cadre supérieur, prof. libérale
#personnes	1, 2, 3, 4, 5, ...
revenu	tout $x \in \mathbb{R}_+$

Niveaux de mesure

- **Variable catégorielle** (qualitative)

Niveaux de mesure

- **Variable catégorielle** (qualitative)
 - ▷ **Nominale** : valeurs (modalités) catégorielles non ordonnées. Exemples : Nationalité, Etat civil, ...

Niveaux de mesure

- **Variable catégorielle** (qualitative)
 - ▷ **Nominale** : valeurs (modalités) catégorielles non ordonnées. Exemples : Nationalité, Etat civil, ...
 - ▷ **Ordinale** : valeurs catégorielle ordonnées Exemples : Niveau de satisfaction (faible, moyen, fort), classe socio-professionnelle ?

Niveaux de mesure

- **Variable catégorielle (qualitative)**
 - ▷ **Nominale** : valeurs (modalités) catégorielles non ordonnées. Exemples : Nationalité, Etat civil, ...
 - ▷ **Ordinale** : valeurs catégorielle ordonnées Exemples : Niveau de satisfaction (faible, moyen, fort), classe socio-professionnelle ?
 - ▷ **Binaire** : deux modalités. Exemples : Sexe (homme, femme), Favorable (oui, non),...

Niveaux de mesure

- **Variable catégorielle** (qualitative)
 - ▷ **Nominale** : valeurs (modalités) catégorielles non ordonnées. Exemples : Nationalité, Etat civil, ...
 - ▷ **Ordinale** : valeurs catégorielle ordonnées Exemples : Niveau de satisfaction (faible, moyen, fort), classe socio-professionnelle ?
 - ▷ **Binaire** : deux modalités. Exemples : Sexe (homme, femme), Favorable (oui, non),...
- **Variable métrique** (quantitative)

Niveaux de mesure

- **Variable catégorielle** (qualitative)
 - ▷ **Nominale** : valeurs (modalités) catégorielles non ordonnées. Exemples : Nationalité, Etat civil, ...
 - ▷ **Ordinale** : valeurs catégorielle ordonnées Exemples : Niveau de satisfaction (faible, moyen, fort), classe socio-professionnelle ?
 - ▷ **Binaire** : deux modalités. Exemples : Sexe (homme, femme), Favorable (oui, non),...
- **Variable métrique** (quantitative)
 - ▷ **Intervale** : la différence entre 2 valeurs fait sens, mais origine (0) arbitraire. Exemples : Température, Heure, Date, ...

Niveaux de mesure

- **Variable catégorielle** (qualitative)
 - ▷ **Nominale** : valeurs (modalités) catégorielles non ordonnées. Exemples : Nationalité, Etat civil, ...
 - ▷ **Ordinale** : valeurs catégorielle ordonnées Exemples : Niveau de satisfaction (faible, moyen, fort), classe socio-professionnelle ?
 - ▷ **Binaire** : deux modalités. Exemples : Sexe (homme, femme), Favorable (oui, non),...
- **Variable métrique** (quantitative)
 - ▷ **Intervale** : la différence entre 2 valeurs fait sens, mais origine (0) arbitraire. Exemples : Température, Heure, Date, ...
 - ▷ **Ratio** : Ratio de 2 valeurs fait sens (implique origine non arbitraire) Exemples : Age, Revenu, Taille du ménage, ...

Variables discrètes et continues

- **variables discrètes**
 - ▷ modalités dénombrables
 - ▷ exemples : nombre d'enfants, taille du ménage

Variables discrètes et continues

- **variables discrètes**
 - ▷ modalités dénombrables
 - ▷ exemples : nombre d'enfants, taille du ménage
- **variables continues**
 - ▷ il existe une modalité entre toute paire de modalités
 - ▷ exemples : revenu, temps consacré à une tâche, ...

Variables discrètes et continues

- **variables discrètes**
 - ▷ modalités dénombrables
 - ▷ exemples : nombre d'enfants, taille du ménage
- **variables continues**
 - ▷ il existe une modalité entre toute paire de modalités
 - ▷ exemples : revenu, temps consacré à une tâche, ...
- Pratiquement, les variables quantitatives discrètes qui admettent un **grand nombre de valeurs** sont, en statistique, traitées comme des **variables continues**.
 - ▷ exemples : âge, nombre de votants, ...