

Méthodes Quantitatives pour les Sciences Humaines et Sociales

Introduction &

Séance 1 : Les Statistiques, un outil pour l'argumentation

E. Madinier

2025

Plan du cours

- 1 Introduction au cours
- 2 Les statistiques : au-delà des mathématiques
- 3 Les statistiques comme outil de description
- 4 Les statistiques comme outil de comparaison
- 5 La quantification en pratique
- 6 Conclusion

Introduction au cours

Objectifs du cours

- Développer une familiarité avec les méthodes quantitatives
- Applications pratiques et professionnelles
- Approche non-traumatique des statistiques :
 - Ne fonctionne que si vous participez activement.

Modalités d'évaluation

- Contrôle continu (30% de la note finale)
 - 3 évaluations de 30 minutes
 - Dates : séances 3, 6, 9
- Examen final (50% de la note finale)
 - Durée : 1h30
 - Séance 12
- Exposé en groupe (20% de la note finale)
 - Durée : 20 minutes
 - Présentation de deux jeux de données et d'un projet de "recherche".

Les statistiques : au-delà des mathématiques

Une histoire sociale et politique

- Les origines des statistiques :
 - Parmi les premiers textes écrits retrouvés ont trouvé des recensements du bétail, des informations sur son cours
 - Usage généralisé du terme en français vers 1785, dérivé de l'italien *statista* ("homme d'État, statiste"), emprunté à l'allemand *Statistik* : science de l'État
 - Développement mathématiques aux XVIIIe et XIXe siècles
- Points clés sur l'histoire de la statistique :
 - Tension initiale entre trois traditions : allemande, anglaise et française
 - Période révolutionnaire française (1789-1815) comme moment clé
 - Transition vers une approche quantitative et une intégration croissante des mathématiques
 - Développement des bureaux officiels

"La politique des grands nombres" par Alain Desrosières.

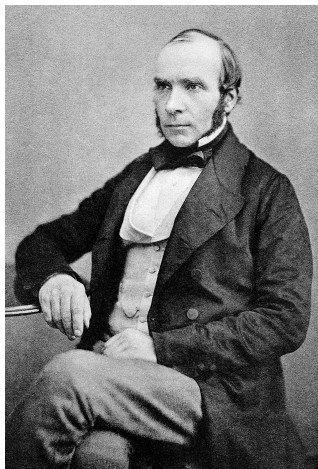
- Rôle social et politique :
 - Outil de gouvernement
 - Construction d'une représentation commune
 - Instrument d'unification nationale

→ **Les statistiques servent un discours argumentatif, souvent au service d'objectifs politiques et/ou économiques**

John Snow, fondateur de l'épidémiologie moderne

Epidémie de choléra de 1854 à Londres

- John Snow rejette de la théorie dominante des miasmes ("mauvais air")
- Méthodes d'investigation :
 - Entretiens avec les résidents locaux
 - Création d'une carte des cas de choléra
 - Analyse statistique comparative des compagnies d'eau et des pompes utilisées



John Snow, fondateur de l'épidémiologie moderne

Découvertes principales :

- Identification de la pompe de Broad Street comme source
- Taux de mortalité plus élevé près de la pompe
- Taux de choléra 14 fois plus élevé chez les utilisateurs de Southwark et Vauxhall
- Majorité des victimes buvaient régulièrement à la pompe

John Snow, fondateur de l'épidémiologie moderne

Rien n'apparaît lors de l'analyse chimique de l'eau mais la démonstration est suffisamment convaincante pour que la pompe soit désactivée par les autorités locales.

La présentation des résultats est un enjeu majeur.

→ Puits creusé à 0,9 mètres d'une fosse septique qui fuyait



Les acteurs des statistiques

- Administrations
- Chercheurs et chercheuses
- Personnel politique
- Entreprises
- Journalistes
- **Un exemple d'usage pour chaque catégorie ?**

Les statistiques comme outil de description

Brève histoire de la statistique en France

- 1801 : Premier recensement moderne en France.
 - Comptage systématique de la population
 - Collecte d'informations démographiques de base
 - Un outil d'administration publique
- 1840 : création de la Statistique Générale de France et l'indice de la "vie chère"
- 1941 : le gouvernement Vichy crée le Service national des statistiques et met en place le numéro d'inscription au répertoire des personnes physiques (numéro de sécurité sociale).
- 1946 : l'INSEE est créé après la Libération.

À partir de 1946, toutes les communes ont été recensées exhaustivement à intervalles irréguliers (6-9 ans). Depuis 2004, le recensement concerne successivement toutes les communes de moins de 10 000 habitants et 40 % des logements des communes de plus de 10 000 habitants au cours d'une période de cinq ans.

L'exemple du chômage en France

L'INSEE utilise la définition suivante pour mesurer le chômage

Bureau international du travail (BIT)

Une personne est au chômage si elle âgée de 15 ans ou plus et répond simultanément à trois conditions :

- être sans emploi durant une semaine donnée
- être disponible pour prendre un emploi dans les deux semaines
- avoir cherché activement un emploi au cours des quatre dernières semaines ou en avoir trouvé un qui commence dans moins de trois mois.

Le chômage au sens du BIT est donc indépendant de la situation administrative à France Travail.

L'exemple du chômage en France

L'INSEE définit alors le "halo" autour du chômage : composé de personnes sans emploi qui, soit recherchent un emploi mais ne sont pas disponibles dans les deux semaines pour travailler, soit n'ont pas effectué de démarche active de recherche d'emploi dans le mois précédent mais souhaitent travailler, qu'elles soient disponibles ou non.

Par exemple :

- Une personne travaille à temps très partiel et souhaite travailler plus (l'INSEE parle "sous-emploi").
- Une personne souhaite un emploi, mais n'effectue plus de démarches actives de recherche d'emploi

Conclusion

Les statistiques, même dites descriptives, ne sont pas "neutres", particulièrement en sciences sociales.

Non seulement les définitions peuvent être manipulées, mais le choix des quantités étudiées n'est pas neutre.

Par exemple, parler du nombre de chômeurs (2 333 000) en mentionnant ou non le nombre d'emplois vacants (520 000) ne produira pas le même discours.

Les statistiques comme outil de comparaison

Le développement des essais cliniques

En 1721 en Angleterre, Lady Mary Wortley Montagu, propose à sept prisonniers condamnés à mort de subir l'inoculation (ancêtre du vaccin) de la variole. Ils ont survécu et n'ont pas contracté la variole.

⇒ Elle en déduit que l'inoculation est efficace et propose de l'inoculer à la population générale.



Le développement des essais cliniques

En 1721 en Angleterre, Lady Mary Wortley Montagu, propose à sept prisonniers condamnés à mort de subir l'inoculation (ancêtre du vaccin) de la variole. Ils ont survécu et n'ont pas contracté la variole.

⇒ Elle en déduit que l'inoculation est efficace et propose de l'inoculer à la population générale.

Quel est le problème avec cette approche ?



Le développement des essais cliniques

En 1721 en Angleterre, Lady Mary Wortley Montagu, propose à sept prisonniers condamnés à mort de subir l'inoculation (ancêtre du vaccin) de la variole. Ils ont survécu et n'ont pas contracté la variole.

⇒ Elle en déduit que l'inoculation est efficace et propose de l'inoculer à la population générale.

Quel est le problème avec cette approche ?

Pas de groupe témoin (*placebo*) pour évaluer si ce résultat était dû à l'inoculation ou à un autre facteur.

Par exemple :



Le développement des essais cliniques

En 1721 en Angleterre, Lady Mary Wortley Montagu, propose à sept prisonniers condamnés à mort de subir l'inoculation (ancêtre du vaccin) de la variole. Ils ont survécu et n'ont pas contracté la variole.

⇒ Elle en déduit que l'inoculation est efficace et propose de l'inoculer à la population générale.

Quel est le problème avec cette approche ?

Pas de groupe témoin (*placebo*) pour évaluer si ce résultat était dû à l'inoculation ou à un autre facteur.

Par exemple :

- La prison a une excellente cantine et les prisonniers sont en meilleure santé que la population générale ?



Certaines hypothèses semblent plus crédibles que d'autres.



Le développement des essais cliniques

En 1721 en Angleterre, Lady Mary Wortley Montagu, propose à sept prisonniers condamnés à mort de subir l'inoculation (ancêtre du vaccin) de la variole. Ils ont survécu et n'ont pas contracté la variole.

⇒ Elle en déduit que l'inoculation est efficace et propose de l'inoculer à la population générale.

Quel est le problème avec cette approche ?

Pas de groupe témoin (*placebo*) pour évaluer si ce résultat était dû à l'inoculation ou à un autre facteur.

Par exemple :

- La prison a une excellente cantine et les prisonniers sont en meilleure santé que la population générale ?
- La maladie circule moins en prison du fait de l'isolement ?

Certaines hypothèses semblent plus crédibles que d'autres.



Le développement des essais cliniques

En 1747 James Lind, un médecin écossais, réalise le premier essai clinique.

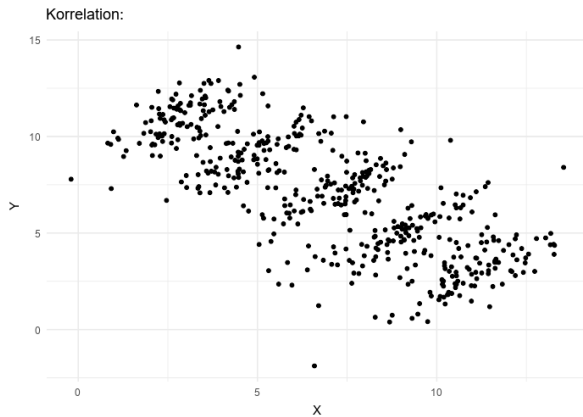
- Sur un navire douze marins scorbutiques sont divisés en six groupes. Tous reçoivent tous le même régime alimentaire à un aliment près.

- ① Un quart de cidre,
- ② Vingt-cinq gouttes d'élixir de vitriol
- ③ Six cuillères de vinaigre
- ④ Un demi-litre d'eau de mer
- ⑤ deux oranges et un citron
- ⑥ Une pâte épicée accompagnée d'eau d'orge.

→ après six jours les marins du groupe 5 sont rétablis, et quelques effets positifs sont observés dans le groupe 1.

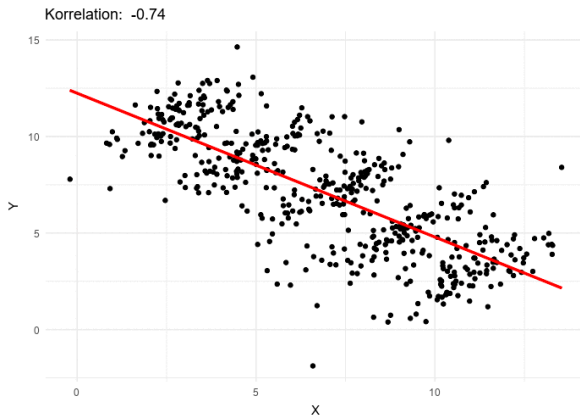
Paradoxe de Simpson

*Données artificielles pour l'illustration



Paradoxe de Simpson

*Données artificielles pour l'illustration



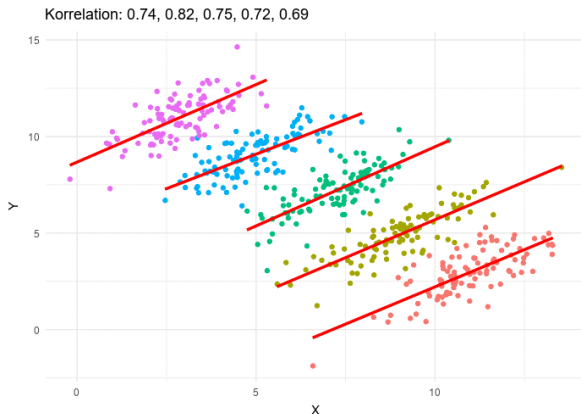
Paradoxe de Simpson

*Données artificielles pour l'illustration



Paradoxe de Simpson

*Données artificielles pour l'illustration



Exemple : le taux de vaccination COVID et les taux de rémissions à l'hôpital aux États-Unis. Les sous-groupes étant des groupes avec des niveaux de risques de comorbidités croissants.

Comparer méthodiquement

- Qu'est-ce qu'une comparaison valide ?
- Danger des conclusions hâtives basées sur des données agrégées
- Comment tester la crédibilité de certaines hypothèses ?
- Nécessité d'analyser les données à différents niveaux
- Importance de considérer les variables cachées
- L'importance du contexte

Vous devez toujours adopter un regard critique sur les chiffres qu'on vous présente.

La quantification en pratique

Approches

"Il y a $x\%$ de chômeurs en France"



Décrire le monde



Statistiques descriptives

"Augmenter le SMIC augmente
le chômage en France"



exposer des liens entre plusieurs quantité



Statistiques inférentielles

Approches

Formulation et validation des hypothèses :

- Approche hypothético-déductive
 - Théorie $\rightarrow$ Hypothèses $\rightarrow$ Test empirique
- Approche inductive
 - Observation $\rightarrow$ Patterns $\rightarrow$ Théorisation

De la normativité en science sociale

Étude descriptive

- Se concentre sur ce qui *est*
- **Questions types :**
 - Que se passe-t-il ?
 - Comment cela fonctionne ?
 - Quels sont les effets ?
 - Quelles sont les causes ?

Étude normative

- Se concentre sur ce qui *devrait être*
- **Questions types :**
 - Est-ce souhaitable ?
 - Que faudrait-il faire ?
 - Comment améliorer ?
 - Quelle est la meilleure option ?
- Implique des jugements de valeur
- Propose des recommandations

De la normativité en science sociale

Exemple : Analyse d'une politique de lutte contre le décrochage scolaire

Approche descriptive

- Mesure et analyse :
 - Taux de décrochage par zone
 - Coût du dispositif
 - Nombre d'élèves concernés
 - Effets mesurables

Résultat type :

"Le dispositif touche 1500 élèves pour un coût de 2M€/an. Le taux de décrochage a diminué de 3 points dans les zones concernées."

Approche normative

- Évaluation et préconisations :
 - Pertinence des objectifs
 - Justice de la répartition
 - Efficience du dispositif
 - Améliorations souhaitables

Résultat type :

"Le dispositif devrait être étendu aux zones périurbaines. Il faudrait augmenter le budget et renforcer l'accompagnement individuel."

Les nouveaux outils statistiques

- La croissance des capacités de calculs offre de nouvelles possibilités :
- Nouvelles méthodes : machine learning
- Nouvelles bases de données : big data, NLP, images, etc...
- Nouvelles méthodes de collecte de données : web scraping, image satellites...

Étude de cas

Sur google scholar : cherchez le papier *Electoral Accountability and Corruption : Evidence from the Audits of Local Governments*, C. Ferra and f. Finan (2011)

Ne vous attendez pas à comprendre l'intégralité du papier, le but est d'apprendre à extraire des informations de ce genre de document.

Lisez l'introduction et la conclusion du papier.

- Quel est la question de recherche ?
- Quels sont les résultats ?
- Que comprenez vous de la méthodologie ?
- Quelles critiques pouvez vous y apporter ?

Lisez les sections 3 et 4.

- Qu'apprenez vous sur la méthodologie ?
- Idem pour la section 5.

Faut-il des limites au nombre de mandats ?

Conclusion

Points clés à retenir

- Les statistiques comme outil d'argumentation
- L'importance du contexte social et historique
- La construction des données n'est jamais neutre
- Prochaine séance : Description d'une variable

Ressources

- Lectures recommandées

- Desrosières, A. (1993). La politique des grands nombres
- Porter, T. M. (1995). Trust in Numbers
- Joshua D. Angrist et Jörn-Steffen Pischke (2010). The Credibility Revolution in Empirical Economics : How Better Research Design is Taking the Con out of Econometrics