

Méthodes quantitatives pour les sciences sociales

Mattéo Lanoë*

Semestre de printemps, année universitaire 2025-2026

Sciences Po, Programme de Bachelor

Page web du cours, régulièrement actualisée

Objectifs du cours

Bienvenue dans le cours de méthodes quantitatives pour les sciences sociales !

Ce cours de première année a plusieurs objectifs. Le principal est d'introduire les étudiants aux usages et à la lecture des sciences sociales quantitatives – comprises de manière large, avec la sociologie, la science politique, l'économie, l'histoire et le droit. Le second objectif est d'introduire les étudiants aux bases de l'analyse de données en utilisant des concepts statistiques simples, afin de mieux saisir comment décrire une variable, associer plusieurs variables et tester la force de l'association entre plusieurs variables. Enfin, par une approche pratique et critique des statistiques, il s'agira de transmettre aux étudiants les manières dont elles peuvent être utiles à l'action publique et à la recherche scientifique en sciences sociales.

Ce cours n'a pas vocation à former de futurs statisticiens. En conséquence, il n'est pas nécessaire d'avoir un bagage mathématique très poussé pour en retirer les principaux enseignements.

Organisation du cours

Le cours a lieu sur le campus de Poitiers de Sciences Po Paris, et se déroulera les mardis après-midi. Il aura lieu en salle 10.

Il est organisé en trois grands modules, composés de plusieurs séances. Chaque séance sera divisée en deux : la première partie du cours sera destinée à l'introduction des concepts statistiques de base que nous utiliserons ; la seconde partie du cours sera destinée à l'implémentation de ces concepts en R à travers l'interface de

*Doctorant en sociologie, rattaché au Centre de recherche sur les inégalités sociales (Sciences Po, CRIS/CNRS), à l'Institut national d'études démographiques (Ined) et affilié au projet de recherche européen « Genre et inégalités sociales de santé » (Gendhi). Mél : matteo.lanoe@sciencespo.fr.

RStudio, de manière guidée et pédagogique. Après chaque séance, il sera indiqué quelles manipulations et fonctions sont à retenir pour pouvoir réaliser les analyses statistiques appropriées.

L'évaluation de ce cours sera basée sur des devoirs sur table réalisés au cours du semestre, un devoir sur table final commun à tous les étudiants suivant ce cours, puis la participation.

Les évaluations réalisées au cours du semestre se passeront après la fin de chaque module pédagogique, et dureront entre 20 et 30 minutes. Elles sont au nombre de 3, et compteront pour 2/3 de la note finale. L'examen final aura lieu lors de la dernière séance du cours, la séance 12. Il comptera pour 1/3 de la note. La participation orale en cours comptera pour 10% de la note finale.

Plan des séances

Séance 1 : À quoi peuvent bien servir les statistiques ?

Les méthodes quantitatives ne sont pas que des mathématiques. Elles se basent sur un ensemble de choix politiques et de constructions scientifiques qui permettent de chiffrer ou de « mettre en chiffres » la réalité sociale. L'objectif de cette séance est d'introduire les étudiants aux principes scientifiques de ces méthodes ainsi qu'à la manière dont elles se sont construites historiquement. La fin de la séance sera dédiée à l'installation de RStudio, ainsi qu'à la découverte du logiciel. **Il n'y a pas de lecture obligatoire pour cette séance.**

Notions abordées : *quantification, épistémologie et histoire des statistiques, méthodes quantitatives*

Éléments de R à retenir : installation de R et RStudio, découverte de la console, différence entre base R et les packages dédiés, importer un jeu de données

Séance 2 : D'où viennent les données statistiques ?

Un des éléments clés pour une analyse quantitative réussie, c'est de disposer de bonnes données. L'objectif de cette séance sera d'introduire les étudiants au panorama des données quantitatives qui peuvent exister, et aux conséquences que cela peut avoir sur l'analyse. On identifiera trois grandes sources d'enquêtes différentes : les données issues d'enquêtes régulières ou *ad hoc*, les données issues d'enquêtes administratives, et les données construites à partir de « traces ». Il s'agira ainsi d'avoir un certain nombre de clés pour repérer les sources de données utilisées par l'action publique ou la littérature scientifique.

Notions abordées : *types de données, population, échantillonnage (aléatoire ou non), marge d'erreur*

Éléments de R à retenir : aucun pour cette séance

Séance 3 : Décrire une variable

Pour comprendre une variable et réaliser des tests statistiques ensuite, il faut d'abord comprendre comment elle est construite et distribuée. Cette séance a pour ambition d'introduire les étudiants aux indicateurs de tendance centrale et aux indicateurs de dispersion. Il s'agira ainsi de bien comprendre quels sont les enjeux derrière l'utilisation de ces concepts statistiques d'analyse univariée, de leurs avantages et de leurs défauts à avoir en tête quand on mène une analyse. **Il n'y a pas de lecture obligatoire pour cette séance.**

Notions abordées : *moyenne, médiane, écart-type, variance, quantile*

Éléments de R à retenir : `mean()`, `median()`, `sqrt()`, `quantile()`, `summarise()`

Séance 4 & 5 : Relier des variables entre elles

Comprendre la distribution d'une variable n'est que la première étape dans la construction d'une analyse quantitative d'un jeu de données. Pour opérer des comparaisons entre variables, il faut avant tout comprendre leur nature : on ne peut pas comparer de la même manière des variables qualitatives (genre ou catégorie d'âge, par exemple), et des variables quantitatives (poids ou taille). Ces séances seront dédiées aux manières de comprendre quelle est la nature d'une variable, et aux manières de visualiser une association entre différentes variables, en tableau par exemple.

Notions abordées : *tableau croisé, comparaison de moyennes, covariance, corrélation*

Éléments de R à retenir : `str()`, `sapply(df$x, class)`, `table(df$x)`, `prop.table()`, `cor()`

Séance 6 & 7 : Construire et catégoriser des données

Dans le cadre de ses missions, un chercheur ou un fonctionnaire d'une administration publique peut avoir besoin de mener ses propres enquêtes. Or, construire un questionnaire et construire des données n'est pas sans obstacles ou difficultés. Ces deux séances auront pour objectif d'introduire les étudiants aux problèmes récurrents dans la construction de questionnaires d'enquête : questions ouvertes ou questions fermées, ordre et formulation des questions, impact du Règlement général sur la protection des données (RGPD). La deuxième séance sera dédiée aux problèmes de recodage de données existantes, et des choix éthiques, scientifiques et politiques à réaliser lors de ce même recodage.

Notions abordées : *questionnaire, catégorisation, classe statistique, recodage*

Lecture obligatoire : Desrosieres, Alain. « La portée sociologique des diverses phases du travail statistique ». *Des mesures*, édité par Jean-Louis Besson et Maurice Comte, Presses universitaires de Lyon, 1986.

Éléments de R à retenir : `cut()`, `factor()`, `mutate()`, `as.character()`, `as.numeric()`, `is.na`

Séance 8 & 9 : Tester la significativité de l'association entre deux variables

Faire des associations entre des variables est à la base du travail statistique. Toutefois, ces associations peuvent parfois naître du hasard, et pour vérifier de manière rigoureuse que deux variables sont bien corrélées, il faut opérer plusieurs tests statistiques de routine qu'il faut savoir interpréter. Cette séance visera donc à introduire les étudiants aux tests canoniques en sciences sociales quantitatives, et à faire en sorte qu'ils soient capables de les mettre en oeuvre puis d'interpréter leurs résultats.

Notions abordées : *tests statistiques, erreur-type, intervalle de confiance*

Éléments de R à retenir : `t.test()`, `chisq.test()`

Séance 10 : Introduction à la régression linéaire

Afin de familiariser les étudiants avec une méthode très populaire en sciences humaines et sociales, cette séance visera à introduire aux bases conceptuelles et géométriques de la régression linéaire. Cette introduction utilisera des modèles univariés, et consistera principalement à apprendre à lire les résultats d'un modèle de régression.

Notions abordées : *nuage de points, droite de régression, R^2 , coefficients p (p -value)*

Séance 11 : Sémiologie graphique et visualisation de données

La dernière séance aborde les enjeux de visualisation des données. En particulier, elle distingue le traitement des données spatiales (qui relève de la cartographie) du cas des données non-spatiales : chronologie, catégories, diagrammes, matrices, résultats de traitements statistiques, etc.

Notions abordées : *cartographie, graphiques, visualisation des données (dataviz), échelles*

Éléments de R à retenir : aucun pour cette séance

Séance 12 : Examen final

Bibliographie indicative du cours

Beraud Martin, Olivier. et al. *L'empire des chiffres : Sociologie de la quantification*. Paris, Armand Colin. « Hors collection », (2020)

Beraud Martin, Olivier. « Chiffre ». *Chiffre*, Anamosa, 2023. p.1-96.

Bugeja-Bloch, Fanny, et Marie-Paule Couto. *Les Méthodes quantitatives*. Presses universitaires de France, 2021.

Desrosières, Alain. *La politique des grands nombres : Histoire de la raison statistique*. Paris, La Découverte. « Poche / Sciences humaines et sociales », (2010)

Llaudet, Elena, et Kosuke Imai. *Data Analysis for Social Science : A Friendly and Practical Introduction*. Princeton University Press, 2023.

Weber, Florence., et al. *Écrire, compter, mesurer : Du Moyen Âge aux bouleversements actuels*. Paris, Éditions Rue d'Ulm. « Sciences sociales », (2023)