

Analyse statistique - Séance 3 : Construire un tableau statistique

Aubin Poissonnier-Beraud

Villes et environnements urbains - Université Lumière Lyon 2

2025

Pierru, E. et Spire, A. (2008) . Le crépuscule des catégories socioprofessionnelles.
Revue française de science politique, Vol. 58(3), 457-481.

Les normes d'édition d'un tableau

La construction d'un tableau statistique répond à des normes d'édition que vous devrez appliquer.

Créer un tableau explicite

Les éléments contenus dans un tableau statistique doivent le rendre le plus explicite possible, c'est-à-dire qu'il ne manque pas (ou très peu) des informations nécessaires pour le lire et l'interpréter correctement.

Le tableau doit comporter :

- ▶ Un titre clair introduisant la statistique présentée.
- ▶ Un contenu avec le noms des variables, de leurs modalités, et les unités précisées.
- ▶ Une ligne ou une colonne d'ensemble, une ligne une colonne de total.
- ▶ Une légende précisant la source des données et l'auteur.ice du tableau, le champ de la population étudiée et une note de lecture.
- ▶ Des annotations si nécessaires sur certaines complexités, définitions, ou choix dans les méthodes de calculs.

Le titre

Le titre doit au mieux résumer la statistique présenté, en indiquant si c'est un graphique ou un tableau avec un index, le type de statistique, le phénomène social étudié et potentiellement la population et la date.

- ▶ Tableau 1 : Origine sociale des étudiant.es en 2021.
- ▶ Tableau 2 : Répartition du statut d'emploi de la population active en 1970 et 2020.
- ▶ Tableau 3 : Statut d'emploi selon le niveau de diplôme en 1970 et 2020.

Le contenu

Le contenu correspond au coeur du tableau et il doit donner la valeur des statistiques présentées, leur unité, les variables concernées et leurs modalités.

- ▶ PCS des parents/Parts en pourcentages.
- ▶ Statut d'emploi/Effectifs en millions et parts en pourcentages.
- ▶ Niveau de diplôme - statut d'emploi/Parts en pourcentages.

La ligne ou la colonne d'ensemble

La ligne ou la colonne d'ensemble doit résumer l'information au niveau de l'ensemble de la population.

- ▶ Pour les parts, part dans l'ensemble de la population plutôt que dans le sous-groupe : part des très favorisé.es parmi l'ensemble des étudiant.es plutôt que chez les étudiant.es des Grandes Écoles.
- ▶ Pour les indicateurs, leur valeur pour l'ensemble de la population : la moyenne d'âge pour l'ensemble de la population plutôt que pour les propriétaires.

La ligne ou la colonne de total

- ▶ Pour les parts, la somme des parts qui doit faire 100%.
- ▶ Pour les effectifs, la somme des effectifs : 30 m d'actif.ves au total décomposés en employé.es et non-employé.es.

La légende

La légende doit permettre de savoir d'où proviennent les données, qui a réalisé les traitements, le champ de la population étudiée avec ses potentielles restrictions et une note de lecture qui garantit que les lecteur.ices pourront correctement lire le tableau.

- ▶ Source : Enquête Logement 2020, INSEE | A. POISSONNIER 2024
- ▶ Champ : ménages habitant dans un logement ordinaire en France métropolitaine
- ▶ Lecture : En 2020, 58,6% des ménages étaient propriétaires de leur logement et 41,4% locataires. Pour les ménages constitués d'un couple sans enfant, 70,5% étaient propriétaires et 29,5% locataires.

L'annotation

L'annotation permet de préciser ou re-préciser des définitions, des choix de calcul, de codage ou de population.

- ▶ Primo-accédant : ménage étant devenu propriétaire pour la première fois et ayant acheté à crédit.
- ▶ Le taux d'endettement moyen a été calculé en divisant la somme des salaires, des pensions et des rentes par le coût total du crédit plus la taxe foncière.
- ▶ Les étudiant.es dont les deux parents sont employés à l'étranger ont été retiré.es de la population d'étude pour ne garder que les emplois pouvant être classés dans la nomenclature des PCS.

Tri croisé : variable expliquée et explicative

Définir le statut attribué à chaque variable

La construction du tri croisé suppose d'attribuer (dans votre tête) un statut à chaque variable : laquelle est censée expliquer l'autre.

- ▶ Je vous conseille de systématiquement mettre la variable explicative en ligne et la variable expliquée en colonne, puis de faire des pourcentages en ligne.
- ▶ Il arrive parfois qu'on doive inverser, par exemple pour des soucis de présentation : si les modalités de la variable expliquée sont trop nombreuses ou contiennent beaucoup de texte, le tableau peut vite devenir illisible. On doit donc inverser cet ordre, puis faire des pourcentages en colonne.
- ▶ Peu importe votre choix, il faut néanmoins être cohérent sur l'ordre, le type de pourcentage et la lecture (cf Séance 2).

Une procédure hybride

En théorie, il est possible de construire et de mettre en forme un tableau statistique uniquement avec R. Extraire les données et utiliser un logiciel interactif tiers comme Excel ou Libre Office Calc évite d'avoir à apprendre comment le faire.

- ▶ Obtenez d'abord les données brutes sur R
- ▶ Copiez ces données dans votre logiciel tiers
- ▶ Faites la mise en forme

Un exemple de tableau fait avec Excel

Origine sociale des élèves des grandes écoles en France 2016-2017 (en %)

Distribution des élèves des grandes écoles par profession et catégorie socio-professionnelle (PCS) de leurs parents
en France pour l'année scolaire 2016-2017

Parts en %	Type de Grande école					
Catégorie sociale	Écoles d'ingénieur	Écoles de commerce	École normale supérieure	Instituts d'études politiques	Autres grandes écoles	Ensemble
PCS défavorisées	9	7	6	8	9	8
PCS moyennes	17	18	14	14	18	16
PCS favorisées	11	10	8	10	10	10
PCS très favorisées	63	65	72	68	63	66
Total	100	100	100	100	100	100

Source : Ministère de l'Enseignement supérieur et de la Recherche, 2021, Statistica | A. POISSONNIER 2024

Champ : élèves inscrits dans une formation "Grandes écoles" en 2016-2017

Lecture : en 2016, 9% des étudiant-es des écoles d'ingénieur étaient issu-es des PCS défavorisées.