

Les données chiffrées en sciences sociales

CORINE EYRAUD

Les données chiffrées en sciences sociales

Du matériau brut
à la connaissance
des phénomènes sociaux

2^e édition

ARMAND COLIN

Illustration de couverture : ©Shutterstock

Mise en page : Kepler

NOUS NOUS ENGAGEONS EN FAVEUR DE L'ENVIRONNEMENT :



Nos livres sont imprimés sur des papiers certifiés pour réduire notre impact sur l'environnement.



Le format de nos ouvrages est pensé afin d'optimiser l'utilisation du papier.



Depuis plus de 30 ans, nous imprimons 70 % de nos livres en France et 25 % en Europe et nous mettons tout en œuvre pour augmenter cet engagement auprès des imprimeurs français.



Nous limitons l'utilisation du plastique sur nos ouvrages (film sur les couvertures et les livres).

© Armand Colin, 2008, 2024
Armand Colin est une marque de
Dunod Éditeur, 11 rue Paul Bert, 92240 Malakoff
ISBN : 978-2-200-63847-4

Sommaire

Introduction	9
1. Évolution des populations bachelière et étudiante	15
L'évolution des bacheliers depuis le début du xx ^e siècle	15
Un peu de vocabulaire	16
Deux questions fondamentales	20
Comment comparer, ou des effectifs aux pourcentages	23
Petite conclusion quant aux proportions et pourcentages	28
Bacheliers et bachelières	36
Un baccalauréat = un baccalauréat ?	37
L'évolution de la population étudiante depuis 1960	40
Variations absolues	40
Variations relatives	41
Variations relatives et coefficients multiplicateurs	46
Variations relatives moyennes	53
Les projections	53
2. Composition de la population bachelière	61
Les tableaux simples	61
Les tableaux croisés	72
Les pourcentages en lignes	75
Les pourcentages en colonnes	79
Que nous dit ce tableau croisé ?	82
Explications sociologiques des inégalités sociales devant l'école	86
3. Réussite et démocratisation scolaires	95
Réussir à l'école, réussir à l'université	96
Les indicateurs transversaux	97

Les indicateurs longitudinaux	102
Classements et palmarès	108
La démocratisation de l'enseignement	112
Constats et critiques	113
Une autre manière d'analyser la démocratisation	123
4. Idées générales sur les statistiques et applications	135
Idées générales	135
Quels calculs effectuer et quelles données utiliser ?	135
Quels points de comparaison ?	137
Quelle définition des catégories et du phénomène étudié ?	139
Quel processus de construction ?	143
La pluridimensionnalité d'un phénomène social	145
Les statistiques indissociables des enjeux sociaux	147
Mesures de la croissance, mesures de la richesse	149
Mais qu'est-ce que le PIB ?	150
D'autres indicateurs ?	156
Quelle valeur octroyer aux statistiques ?	159
Mesures du chômage, mesures de la qualité de l'emploi.	
Controverses des années 2000	160
Mais qu'est-ce qu'un chômeur ?	161
Les écarts se creusent	168
Du chômage à la qualité de l'emploi	171
Ce que ces statistiques ne nous disent pas	174
Mesures de la délinquance	176
Quels chiffres pour quelle délinquance ?	176
Quel processus de construction ?	178
Comment les chiffres travaillent les organisations et les métiers ?	182
Quoi de neuf ?	184
Une sociologie de la quantification	186
5. Connaissance des phénomènes sociaux	191
Travailler à définir un phénomène	192
Différencier cartes et territoire	192
Définir	195
Construire ses données. L'exemple de la religiosité	210
La religiosité vue par Glock : les États-Unis des années 1960	211
Être catholique en France dans les années 1980	212

Être musulman au Maroc dans les années 2000	214
Une grille d'analyse	216
Évolution de la situation religieuse aux États-Unis	216
Les inégalités sociales. Analyse de controverses des années 2000	219
Conclusion	235
Bibliographie	237
Table des encadrés	245
Index des noms d'auteurs	247
Index notionnel	249

Introduction

La connaissance d'une institution ou d'un phénomène social, quels qu'ils soient, ne se fait plus aujourd'hui sans données chiffrées, statistiques ou démographiques. Ces données permettent d'appréhender l'ampleur du phénomène, ses différentes dimensions, l'activité de l'institution, son poids, son public, son budget, son évolution, et bien d'autres choses encore. Le sociologue ne peut donc se passer ni de la statistique ni de la démographie, et ceci pour deux raisons essentielles. Tout d'abord parce que, bien utilisées, celles-ci sont, pour lui, un outil de connaissance important. Ensuite, parce que les informations chiffrées sont devenues omniprésentes dans les médias et les débats publics : chaque institution émet ses propres statistiques ; ses différents acteurs (employeurs, syndicats, associations d'usagers, etc.) produisent des chiffres souvent concurrents et utilisés pour démontrer la validité de leur propre argumentaire. La mesure de chaque phénomène social donne souvent lieu à des querelles sur le « vrai chiffre », pensons ne serait-ce qu'au chômage.

Indispensable pour comprendre le monde qui nous entoure et pour en parler, le nombre mérite d'être mieux connu et mieux approprié. Un étudiant de sociologie, et plus largement de sciences humaines et sociales, et même tout citoyen, ne peut donc se passer ni de la compréhension du mode de construction de ces statistiques, ni d'une réflexion sur la manière dont on peut construire une connaissance d'un phénomène ou d'une institution en s'appuyant sur ces données.

C'est dans ce cadre que se situent les objectifs de cet ouvrage ; nous pouvons les différencier comme suit :

- premièrement, une bonne initiation aux calculs de pourcentages et taux (y compris taux de variation) et à la lecture de tableaux statistiques ;

- deuxièmement, la maîtrise d'un minimum de termes démographiques et statistiques dont on donnera la définition (ces termes sont généralement en italique dans le texte) ;
- troisièmement, l'apport de connaissances factuelles et de points de repère quant à la structure de la société française et de certaines de ses institutions ;
- quatrièmement, un apprentissage de la démarche sociologique permettant de construire une connaissance la plus complète et la plus objective possible d'un phénomène social, et ce en utilisant ces données ;
- cinquièmement, une initiation à la lecture critique d'articles de journaux ou de travaux utilisant des données chiffrées.

Ces objectifs seront progressivement mis en œuvre en prenant appui sur de nombreux exemples choisis dans différents domaines de la vie sociale : démocratisation de l'enseignement, évolution des inégalités sociales, chômage, délinquance, intégration, religion, mais également « trou de la Sécu », mesure de la pauvreté, palmarès des lycées, etc. Cette deuxième édition réactualise, par rapport à la première édition parue en 2008, la plupart des données utilisées et intègre en références bibliographiques des travaux réalisés entre-temps sur ces différents domaines. Nous avons parfois choisi de conserver les données de la première édition, soit parce qu'elles s'intégraient à l'analyse de controverses particulières (par exemple, la controverse du milieu des années 2000 autour des chiffres du chômage dans le chapitre 4), soit parce que l'analyse réalisée dans la première édition avait sa cohérence et que les institutions concernées ne fournissent plus les données que nous avons alors utilisées (en particulier en ce qui concerne la question de la réussite ou de l'échec à l'université dans le chapitre 3).

L'utilisation d'exemples variés et réels est nécessaire à la réalisation des objectifs qui sont les nôtres, mais risque d'être insuffisante, tant, en la matière, rien ne vaut la pratique. C'est pourquoi des exercices intégralement corrigés de calculs, d'analyses de tableaux, d'analyses d'articles de journaux ou de revues accompagnent le cheminement de l'ouvrage. Ils sont conçus à la fois comme moyens d'assimiler les contenus et comme outils visant à aiguïser la réflexion. Nous vous conseillons vivement de les réaliser seuls, avant de vous reporter à leur correction en fin de chapitre.

Cet ouvrage n'est donc ni un cours de statistiques ni un manuel de techniques d'enquêtes, mais un livre de réflexion et d'analyse

sociologiques des données chiffrées et de leur utilisation. La primauté est donnée aux idées, et non aux formules. Bien sûr, cette analyse ne peut être menée sans une maîtrise minimale des outils quantitatifs, de la lecture des tableaux et du vocabulaire démographique et statistique de base. C'est ce par quoi nous commencerons¹.

1. L'auteure tient à remercier toutes celles et ceux qui l'ont aidée à concevoir cet ouvrage et, tout particulièrement, Patrick Perez, pour sa relecture critique et minutieuse, nos discussions stimulantes et ses encouragements.

Première partie

Des données brutes aux pourcentages et taux

Nous nous situons, tout au long de cette partie, dans une logique de découverte : il s'agit de faire apparaître les différents problèmes que pose l'utilisation des statistiques, en même temps que l'on apprend (ou revoit) les modes de calcul (essentiellement pourcentages et taux de variation) et que l'on s'initie au vocabulaire statistique et démographique de base et à la lecture et interprétation de tableaux. Ces différentes dimensions seront abordées à partir de statistiques scolaires et universitaires.

Chapitre 1

Évolution des populations bachelière et étudiante

Ce premier chapitre est consacré à la présentation des opérations de base de la statistique et du vocabulaire afférent, à partir de l'étude de populations particulières que sont les populations bachelière et étudiante. En nous familiarisant avec le calcul de pourcentages, nous nous trouverons confrontés aux premières difficultés et aux premiers pièges. Apprendre à les reconnaître, à les maîtriser ou les éviter nous permettra de poser les premières pierres pour une démarche de travail et d'analyse rigoureuse des statistiques.

L'évolution des bacheliers depuis le début du xx^e siècle

Nous allons partir d'un tableau simple (tableau 1.1).

Tableau 1.1 Bacheliers par sexe (effectifs) selon l'année d'obtention du diplôme (de 1920 à 2022) et population de référence (effectifs)

Années	Bacheliers			17 ans
	Masculins	Féminins	Ensemble	
1920	9 190	1 326	10 516	691 600
1930	11 858	3 708	15 566	668 000
1935	8 560	3 379	11 939	351 100
1940	18 485	9 292	27 777	706 600
1945	15 596	11 719	27 315	660 000
1950	19 039	14 106	33 145	642 500
1960	31 635	29 864	61 499	543 500
1970	79 074	88 233	167 307	838 300
1980	92 760	129 669	222 429	852 500
1985	108 481	144 569	253 050	844 500
1990	168 576	215 424	384 000	882 700
1995	213 571	266 429	480 000	765 500
2000	227 544	271 456	499 000	804 800
2005	233 849	272 759	506 608	829 100
2010	252 430	279 338	531 768	818 100
2015	296 007	322 837	618 844	796 500
2020	351 463	371 508	722 971	834 800
2022	323 220	343 620	666 840	837 700

Champ : France métropolitaine jusqu'en 2000, France métropolitaine + Départements et régions d'outre-mer (DROM) hors Mayotte entre 2005 et 2010, France métropolitaine + DROM y compris Mayotte à partir de 2011.

Source : Ministère de l'Éducation nationale.

Un peu de vocabulaire

Ce tableau est un tableau en *effectifs* ou nombre *absolus* ou, mieux encore, en *valeurs absolues*. Attention les informations ne sont pas toujours fournies pour les mêmes intervalles de temps (parfois 5 ans, parfois 10 ans).

La population de référence (dernière colonne) est ici la population qui atteint l'âge de 17 ans telle et telle année. Ainsi 691 000 individus

ont, en France, atteint cet âge en 1920. En quelle année sont-ils nés ? $1920 - 17 = 1903$. En démographie, l'ensemble des individus nés la même année est appelé une *génération*, on parlera donc de la génération 1903.

Attardons-nous quelques instants sur cette colonne, et regardons son évolution entre 1920 (génération 1903) et 1980 (génération 1963). On retrouve différents événements historiques qu'a connus la France :

- baisse importante des effectifs de 17 ans en 1935, soit la génération 1918 : on parle en démographie de *génération creuse* ou, plus souvent, de *classes creuses* puisque cela concerne les naissances pendant toute la guerre de 1914-1918 ;
- les guerres sont souvent suivies d'un « rattrapage » de naissances : les *génération*s ou *classes pleines*, c'est le cas pour la génération 1923 (qui a 17 ans en 1940) ;
- on peut faire les mêmes remarques pour la génération, moins nombreuse, qui a 17 ans en 1960 (génération creuse de 1943), ou pour la génération, beaucoup plus nombreuse, qui a 17 ans en 1970 (génération pleine de 1953, après-guerre et « baby-boom »).

Cela dit, chaque génération — c'est-à-dire chaque population qui a comme caractéristiques communes de résider en France et d'avoir 17 ans la même année — se différencie du point de vue d'autres caractéristiques. Certains sont de sexe masculin, d'autres de sexe féminin, ou encore certains ont le baccalauréat, d'autres ne l'ont pas. On appelle ces caractéristiques des *caractères*, des *critères*, ou, terme très usité en sociologie et en statistique, des *variables*.

Si l'on veut compter un ensemble, une population d'êtres humains, on peut décider de compter par exemple tous ceux qui résident sur tel territoire à l'instant t (on prend alors comme *caractère* la résidence sur un territoire), on peut aussi dénombrer ceux qui ont plus de 60 ans (on prend comme *critère* l'âge), ou compter les individus de sexe féminin (on prend comme *variable* le sexe). Dans le tableau 1.1, la population de référence a été définie à partir de deux variables : résidant en France et ayant 17 ans telle année. Au sein de cette population, on peut constituer des sous-ensembles : être bachelier ou ne pas l'être, être bachelier fille ou bachelier garçon, etc. Cette opération de définition d'ensembles, de sous-ensembles, et de classement d'individus (humains ou non : entreprises...) dans ces ensembles et sous-ensembles est l'opération de base de la statistique.

Ces caractéristiques qui font classer les individus dans tel ou tel sous-ensemble, ne sont pas toutes du même ordre. Certaines sont

repérées par des *mots* : sexe, état matrimonial, profession ou religion ; on parle alors de *variables qualitatives* ou *nominales*, ou encore, dans un langage moins statistique, de *catégories*. Certaines de ces caractéristiques sont repérées par des *chiffres* : âge, revenu, nombre de salariés d'une entreprise ou surface d'un logement ; il s'agit des *variables quantitatives* ou *numériques*. Finalement, certaines de ces caractéristiques sont repérées par des *mots ordonnés sur une échelle* ; on parle alors de *variables ordinales* très utilisées dans les enquêtes d'opinion ou d'attitude : « Qu'avez-vous pensé de cette exposition ? Vous l'avez trouvée : très intéressante/intéressante/assez peu intéressante/pas intéressante du tout ? » La distinction entre les différents types de variables est importante car leur traitement statistique, autrement dit le type de calcul que l'on peut effectuer, est différent. Dans ce chapitre, nous nous intéresserons exclusivement à des variables qualitatives et à la manière la plus courante de les traiter : le calcul de pourcentages.

Le nombre de sous-ensembles que l'on construit à partir d'un critère s'appelle le nombre de *modalités* de la variable. Ce nombre est parfois défini par le droit, et donc dépendant du pays et du moment historique : il y a ainsi légalement en France aujourd'hui quatre états matrimoniaux (célibataire, marié, divorcé, veuf), mais on peut également intégrer le PACS ; la variable « état matrimonial » a donc quatre ou cinq modalités. Mais le plus souvent, *le nombre de modalités dépend du degré de détail souhaité* : on peut ranger les âges en trois catégories ou trois classes (par exemple : 0-19 ans, 20-64 ans, plus de 65 ans), par tranches de 5 ans ou de 10 ans, etc. La modalité qui regroupe le plus d'individus, soit la modalité la plus fréquente, s'appelle le *mode*. La *catégorie* d'âge *modale* en France aujourd'hui est la catégorie des 20-64 ans. Finalement, l'ensemble des modalités d'une variable qualitative s'appelle une *nomenclature*.

● Sexe ou genre ?

Jusqu'aux années 1970, les recherches sur les différences et inégalités entre les hommes et les femmes utilisaient uniquement le terme « sexe ». Le terme de « genre » (« *gender* ») est apparu aux États-Unis dans le discours des mouvements féministes des années 1970 et avec le développement des « *Women's Studies* » en sciences sociales. La volonté initiale était de souligner que les rapports de domination et les différences de comportements entre hommes et

femmes n'étaient pas naturels. Il n'est pas naturel qu'une fille soit douce, sensible et attirée par les garçons, et qu'un garçon soit énergique, ambitieux et attiré par les filles. La socialisation fabrique des individus masculins/féminins et crée des rapports de domination entre les deux. Or le terme « sexe » a, en langue anglaise, une forte dimension biologique ; utiliser le terme « *gender* » permettait d'insister sur la dimension sociale et l'on différençiait ainsi le sexe biologique et le genre social.

Le terme de genre s'est ensuite largement diffusé à partir des années 2000 avec les « *Gender Studies* ». Il y a toutefois des débats autour de la définition du genre, et donc du sexe. Plusieurs auteurs ont souligné que cette définition posait problème, principalement parce qu'elle suppose que le sexe est purement biologique, c'est-à-dire que chaque individu peut être biologiquement catégorisé « mâle » ou « femelle ». Plusieurs travaux ont souligné que ces catégories étaient moins claires et étanches que cela, certains ont même retourné la définition considérant que « le genre précède le sexe » (en particulier Delphy, 2001, mais aussi Claude-Mathieu, 1989), voulant ainsi signifier que les différences biologiques n'ont aucune signification en elles-mêmes et que le sexe lui-même est construit socialement (l'assignation d'un sexe à la naissance par l'état civil – institution sociale - est déjà un acte social). C'est ce que disait Simone de Beauvoir en 1949 en introduction du tome II du *Deuxième sexe* : « On ne naît pas femme : on le devient ». Finalement, des auteurs, en particulier français, préfèrent aujourd'hui encore parler de sociologie des rapports sociaux de sexe plutôt que de sociologie du genre, considérant que le second choix ne met pas au premier plan, voire peut occulter, le fait qu'il s'agit de rapports sociaux et non pas d'identité (de genre).

Ceci étant, on peut trouver aujourd'hui indifféremment les termes « inégalités femmes/hommes », « inégalités de sexe » ou « inégalités de genre » (quand on parle d'« inégalités », on parle forcément de rapports sociaux).

Si dans cet ouvrage nous avons choisi de parler de « sexe » et non de « genre », c'est uniquement parce que les enquêtes de la statistique publique (Insee, Ined...) ont gardé la variable « sexe » (le sexe inscrit dans l'état civil des individus) avec deux modalités possibles (homme ou femme). D'autres chercheuses et chercheurs font le choix dans leurs enquêtes par questionnaires d'utiliser la variable du « genre » en se basant sur l'autodéfinition des individus et en lui donnant de plus nombreuses modalités (« femme », « homme », « personne non binaire », « autre », « ne se prononce pas », etc.) ; ce sont justement et généralement celles et ceux qui travaillent sur « l'identité de genre ».

Deux questions fondamentales

Revenons maintenant à notre tableau, et regardons le sous-ensemble des bacheliers. En 1920, 10 516 personnes ont obtenu le baccalauréat. Le fait d'avoir obtenu ou non le bac est une *variable, qualitative* ou *nominale*, qui possède deux modalités : avoir obtenu ou ne pas avoir obtenu. Prenons les chiffres de 1920 et ceux de 1970 : 10 516 personnes ont obtenu leur baccalauréat en 1920, 167 307 l'ont obtenu en 1970, soit beaucoup plus. Mais entre 1920 et 1970, la population de référence dans son ensemble (dernière colonne) a augmenté (de 691 600 à 838 300) ; la croissance du nombre de bacheliers pourrait être entièrement ou majoritairement imputable à la croissance de la population âgée de 17 ans. Deux questions très simples se posent donc :

- 10 516 bacheliers en France en 1920, *est-ce beaucoup ?* ;
- 167 307 bacheliers en 1970, *est-ce plus ou moins* (et beaucoup plus ou beaucoup moins) *que* 10 516 bacheliers en France en 1920 ?

On doit se poser la première question à chaque fois que l'on voit ou entend un chiffre, on se la pose d'ailleurs plus ou moins automatiquement. On doit se poser la seconde à chaque fois que l'on a plusieurs chiffres. Il s'agit en fait des deux faces d'une même question, car la « notion » de « beaucoup » a toujours un sens par rapport à quelque chose : c'est « beaucoup » par rapport à « peu ». Prenons, dans l'encadré ci-dessous, un exemple simple.

● Le nombre de véhicules que détient mon ménage

La notion de *ménage* est au fondement de très nombreuses statistiques, il faut donc en connaître la définition. Elle part de l'unité de vie : un ménage est constitué d'une ou de plusieurs personnes vivant dans un même logement autonome, quelles que soient leurs relations. Ainsi trois étudiants vivant en cohabitation sont considérés par l'INSEE comme un ménage, tout comme une personne vivant seule.

Mais, savez-vous ce qu'est *l'Insee* ? L'Insee (ou INSEE) est l'Institut national de la statistique et des études économiques. Il a été créé en 1946 et dépend du ministère de l'Économie et des Finances. Il est chargé de « collecter, produire et diffuser des informations sur l'économie et la société française afin que tous les acteurs intéressés puissent les utiliser pour effectuer des études, faire des prévisions et prendre des décisions » (site internet de l'Insee).