



UNIVERSITE GASTON BERGER

L'excellence au service du développement

UFR des Lettres et Sciences humaines
Section de sociologie

SOCIO 532 EC2 (STATISTIQUES &) INFORMATIQUE APPLIQUÉE(S) AUX SCIENCES SOCIALES

Niveau : Master 2

Semestre/Année : 1^{er} semestre 2024

Crédits : 5

Volume horaire : TD (50 heures), TPE (50 heures)

CONTACTS

Professeur :	©El Hadj Toure, PhD. Sociologie Spécialiste des statistiques sociales et méthodes de sondage
Horaire hebdomadaire :	Théorie (2 h) : Samedi 8h-10h, Local D01 Labo (2 h) : Salle multimédia (Hilaire Bouka) !
Bureau :	109B (au fond du bâtiment B, près du département de géo)
Téléphone :	(+221) 77 670 34 26 (+1) 514 377 4095
Courrier électronique :	elhadj.toure@ugb.edu.sn
Disponibilité :	Sur rendez-vous
Auxiliaire d'enseignement :	Néant
Espace virtuel (Moodle) :	http://foad.ugb.sn/course/view.php?id=895

DESCRIPTIF DU COURS

Statistiques descriptives et inférentielles, synthétiser l'information statistique, de l'analyse bivariée à l'analyse de régression linéaire approfondie. Modèles de causalité en sociologie. Analyse de tableaux croisés multivariés, corrélation et régression linéaires multiples, régression logistique binaire, régression logistique multinomiale. Application à l'aide d'un logiciel statistique (SPSS).

PLACE DU COURS DANS LE PROGRAMME

Ce cours est partie intégrante du programme de Master de sociologie. Il se situe dans les cours obligatoires. Il contribue à amener l'étudiant-e à savoir utiliser des outils statistiques et informatiques appropriés pour répondre à des questions de recherche multivariées.

Préalable : IASS 361

OBJECTIFS D'APPRENTISSAGE

De façon général, ce cours de statistiques et d'informatique appliquées aux sciences sociales présente les statistiques avancées fréquemment appliquées aux sciences sociales, l'enjeu étant d'amener les étudiants de M2 à utiliser ces outils pour répondre à des questions de recherche multivariées. De façon plus spécifique, à la fin du cours, les étudiants-e-s seront en mesure de :

- 1) Savoir synthétiser l'information statistique, et avoir un aperçu des techniques uni-bivariées et approfondir l'analyse de régression linéaire bivariée ;
- 2) Connaître, illustrer et schématiser les modèles de causalité de base utilisés dans la littérature en sociologie et dans les sciences sociales connexes ;
- 3) Appliquer ces modèles de causalité et procéder à l'élaboration en employant des méthodes, techniques et procédures statistiques appropriées,
- 4) Effectuer les analyses uni-bi-multivariées à l'aide du logiciel SPSS et être en mesure d'en interpréter statistiquement et sociologiquement les résultats ;
- 5) Savoir émettre une critique méthodique de travaux de recherche faisant appel aux statistiques multivariées.

Telles sont les compétences à acquérir dans le cadre du présent cours.

DÉMARCHE PÉDAGOGIQUE

Dans le but d'atteindre les objectifs susmentionnés, chaque séance hebdomadaire est divisée en deux temps successifs. La première période (2 h) consiste en un *exposé théorique* entrecoupé de remue-méninges et d'exercices « éclair ». Elle s'apparente à un cours magistral (CM). Sans tomber dans le piège du formalisme mathématique outrancier, le professeur présente les statistiques uni-bivariées selon une approche intuitive. Derrière chaque formule se cache, en effet, une intelligence logico-mathématique et conceptuelle. On peut y accéder à l'aide de mises en situation, de simulations simples, d'illustrations visuelles et d'analogies intuitives. Ce n'est toutefois pas un cours de maths. Les mathématiques sont plutôt utilisées pour donner sens aux phénomènes sociaux

La deuxième période (2 h), quant à elle, est consacrée à l'application en *laboratoire informatique* des connaissances théoriques préalablement acquises. Elle s'apparente à des travaux dirigés (TD). À cet effet, les étudiant-e-s utiliseront le logiciel statistique SPSS afin de procéder automatiquement à des analyses statistiques avancées. Ils pourront ainsi s'en servir dans l'étude concrète des phénomènes sociaux en manipulant des données réelles. Certaines données proviennent d'un sondage complété par les étudiants eux-mêmes ainsi que d'enquêtes de l'Agence nationale de la Statistique et de la Démographie (ANSD)¹.

Somme toute, la pédagogie est axée sur la compréhension conceptuelle et l'interprétation contextualisée des statistiques sociales bi-multivariées. Comme le suggèrent Boudon (2002) et Lazarsfeld (1955), une association statistique bivariée ne peut être interprétée comme une relation causale sans contrôler ou considérer le rôle d'une variable tierce. Afin de faciliter leur apprentissage, les étudiant-e-s devront se préparer par les *lectures hebdomadaires suggérées*. De plus, après chaque séance hebdomadaire, ils pourront s'autoévaluer en ligne à travers un *test quiz*.

¹ <https://www.ansd.sn/>, consulté le 15 mars 2021. L'agence sénégalaise offre un portail qui permet d'accéder à des données statistiques portant, entre autres, sur la situation sociale et économique du pays.

SUPPORT DE COURS

Avant chaque séance hebdomadaire, les étudiant-e-s ont la possibilité de télécharger les textes suggérés, les présentations PowerPoint (sous divers formats) et les applications à effectuer au labo (SPSS). Les applications sont illustrées visuellement dans un cahier de laboratoire, de telle sorte que les étudiant-e-s peuvent suivre facilement les démonstrations du professeur et réaliser les exercices pratiques proposés. Plus tard après les labos, les informations contenues dans le cahier de laboratoire devraient permettre aux étudiants de répéter les mêmes démonstrations. Tous ces supports de cours sont disponibles sur le site web (Moodle).

Moodle est une plateforme de gestion de cours et d'apprentissage en ligne. On peut y accéder via ce lien : <http://foad.ugb.sn/course/view.php?id=895>. *Si vous avez déjà utilisé Moodle, mettez directement vos coordonnées pour accéder au cours.* Si vous avez oublié votre mot de passe, vous rendre à l'IFOAD auprès de Mme Kassé. Pour un premier accès, vous pouvez aussi saisir foad.ugb.sn dans l'adresse URL d'un navigateur Internet.

Au demeurant, les notes de cours résument certes l'essentiel de la partie théorique du cours. Mais, elles n'en fournissent pas le contenu intégral, le style télégraphique étant utilisé dans les présentations PowerPoint. Par conséquent, la présence en classe et au labo est obligatoire pour maîtriser la matière et augmenter les chances de réussir le cours. **Le succès est entre vos mains !**

Autre matériel obligatoire : l'achat d'une calculatrice de base. Elle est indispensable à la passation des examens, tout comme à la réalisation des exercices « éclair » proposés dans le cours théorique.

MODALITÉS D'ÉVALUATION

Réalisés en équipe de deux à trois tout au plus, deux *travaux pratiques* (TPs) seront proposés aux étudiant-e-s tout au long de ce cours. Les TPs consistent en des analyses statistiques bi-multivariées à effectuer sur la base de données sociales réelles à l'aide du logiciel SPSS. L'enjeu est de répondre à des questions de recherche relationnelles et causales en appliquant des modèles fréquemment employés en sociologie. Ces travaux offrent aux étudiants l'occasion de mettre en pratique les techniques abordées et d'en présenter, interpréter statistiquement et sociologiquement les résultats. Les étudiant-e-s auront au moins deux semaines pour réaliser les TPs et les remettre en format imprimé auprès du professeur en classe, conformément à la programmation.

Organisé sur table, l'*examen* porte sur l'ensemble du cours. Il est individuel, dure 2 heures et se fera à « livres et cahiers ouverts »². L'examen consiste en des questions objectives de compréhension, de calculs et problèmes pratiques à résoudre ainsi qu'en l'interprétation de résultats statistiques obtenus à l'aide du logiciel SPSS. Il n'y aura aucun logiciel à manipuler.

Les quiz 1 à 8 sont *obligatoires* et les étudiant-e-s les réaliseront directement sur le site web (Moodle), tout au plus une semaine après leur disponibilité. Chacun des quiz porte sur des questions

² Mais attention, les examens ne sont pas conçus pour que vous ayez toute la latitude de consulter vos notes de cours. Si vous n'êtes pas prêts avant, vous ne le serez pas pendant les examens. À vous de vous organiser! Une bonne pratique, par exemple, consiste à prendre une feuille et y noter les formules et concepts statistiques essentiels. L'objectif est simplement de ne pas vous encourager à mémoriser aveuglement la matière, mais plutôt à la comprendre et à faire preuve de *conscience réflexive*, pour reprendre l'expression du sociologue Giddens (1987). Ci-après un document qui comprend des conseils pour réussir un examen à livres ouverts: <https://infomaitres-mberube.profweb.ca/wp-content/uploads/2020/11/Fiche-pour-examen-a-%CC%80-livre-ouvert.pdf>.

objectives de compréhension en lien avec une leçon. Les questions se présentent sous diverses formes : vrai/faux, choix multiples, numérique, appariement.




Les composantes de l'évaluation sont l'objet d'une pondération.






Tableau 1. Résumé des évaluations

Items	Pondération	Disponibilité	Remise	Statut
8 quiz (1 à 8)	20%	Après chaque séance	Une semaine après	Individuel
TP#1	20%	Semaine 2	Semaine 4	En équipe
TP#2	20%	Semaine 5	Semaine 7	En équipe
Examen	40%		Semaine 10	Individuel
Total	100%			





DÉROULEMENT DU COURS





Sous réserve de modification, ci-dessous le contenu et le calendrier détaillé des activités du cours :

Semaine 1	
Sam.	<ul style="list-style-type: none"> Présentation du cours, de son contenu et des modalités d'évaluation  <p>Leçon 1. Rappel : Statistiques descriptives et inférentielles univariées</p> <ul style="list-style-type: none"> Description statistique : distribution de %, tendance centrale et variation Inférence statistique : estimation par intervalle de confiance d'un paramètre <p><i>Lectures suggérées : Cousineau & Harding (2017)</i></p> <p>Quiz 1 à faire en ligne (date limite le 17 février).</p>
10 fév.	 <p>Labo 1. À l'aide de SPSS, exécuter des analyses descriptives se rapportant aux distributions de pourcentages, mesures de tendance centrale, de variation et de forme d'une distribution. Exécuter des analyses inférentielles en calculant l'erreur-type, la marge d'erreur et l'intervalle de confiance d'une moyenne, les représenter graphiquement par la barre d'erreur afin de comparer des groupes eu égard à une variable quantitative d'intérêt. Lire et interpréter les sorties SPSS en termes de description des données d'échantillon et de généralisation des résultats à la population. Présenter les résultats sous forme textuelle et dans des tableaux publiables à l'aide du logiciel Word.</p>
Semaine 2	
Sam.	 <p>Leçon 2. Comment synthétiser l'information statistique ?</p> <ul style="list-style-type: none"> De la nécessité de recoder les variables et de synthétiser l'information statistique Création de tableaux synthétiques Création de graphes synthétiques Création de variables synthétiques <p><i>Lectures suggérées : Martin, 2009, pp. 42-62</i></p> <p>Quiz 2 à faire en ligne (date limite le 24 février).</p>

17 fév.	 Labo 2. À l'aide du logiciel Word, construire des tableaux synthétiques présentables. À l'aide du logiciel Excel, construire des graphes synthétiques présentables à partir de résultats SPSS. À l'aide du logiciel SPSS, créer des variables synthétiques, « recoder » des variables de façon à réduire leurs valeurs. Lire et interpréter les sorties SPSS. TP#1 à faire : régression linéaire approfondie, ANOVA, synthèse de l'information statistique (séances 2, 3)
Semaine 3	
Sam. 24 fév.	 Leçon 3. De l'analyse bivariée. Régression linéaire approfondie <ul style="list-style-type: none"> • Quelques éléments de rappel : Questions relationnelles et techniques statistiques bivariées, incluant la régression linéaire simple • Régression linéaire simple avec une VI qualitative dichotomique codée 0/1 et parallèle avec l'ANOVA • Régression linéaire simple avec une VI qualitative non dichotomique (création de variables factices) et parallèle avec l'ANOVA • Interprétation et utilité des coefficients de régression standardisés (bêta) <p><i>Lectures suggérées : Fox (1999 : 253-287)</i></p> <p>Quiz 3 à faire en ligne (date limite le 2 mars)</p>
	 Labo 3. À l'aide du logiciel SPSS, estimer un modèle de régression linéaire simple incluant une VI qualitative dichotomique codée 0/1 ; créer des variables factices dichotomiques pour une VI qualitative non dichotomique et les inclure dans un modèle bivariée. Interpréter les coefficients bêta et justifier leur intérêt en termes de comparaison d'effets. Examiner, lire et interpréter les sorties SPSS en termes de prédiction. Présenter les résultats sous forme textuelle et dans des tableaux publiables à l'aide du logiciel Word.
Semaine 4	
Sam. 2 mars	 Leçon 4. Introduction à l'analyse causale <ul style="list-style-type: none"> • Élaboration : de l'analyse bivariée à l'analyse multivariée • Les trois conditions de la causalité • Aperçu, illustration et schématisation des modèles de causalité usuels en sociologie : convergence, explication (relation fallacieuse), modération, discrimination • Vue d'ensemble des techniques statistiques multivariées associées à ces modèles <p><i>Lectures suggérées : Tacq (1997: 8-19)</i></p> <p>Quiz 4 à faire en ligne (date limite le 9 mars)</p> <p>Date limite pour la remise du TP#1 (23h59)</p>
	 Labo 4. À l'aide du logiciel SPSS, créer et gérer des fichiers de syntaxes; effectuer des procédures statistiques de façon à préparer des analyses bi-multivariées subséquentes : « créer une variable avec la condition IF, « calculer une variable » complexe à partir d'autres variables, créer des variables composites, filtrer ou « sélectionner » des observations, et « scinder » un fichier afin de comparer des groupes. Lire et interpréter les sorties SPSS, présenter les résultats sous forme textuelle et dans des tableaux et



	schémas publiables à l'aide du logiciel Word.
Semaine 5	
Sam. 9 mars	 Leçon 5. Analyse de tableaux croisés multivariés <ul style="list-style-type: none"> • Élaboration : procédure d'introduction d'une variable-contrôle qualitative dans une relation entre deux variables qualitatives • Paul Lazarsfeld et les modèles d'élaboration de tableau : Explication (relation fallacieuse), interprétation (médiation), spécification (modération), dissimulation • Construction de tableaux partiels et calcul du gamma partiel • Schématisation et interprétation des chi-carrés, gammas aux fins de diagnostic causal <p><i>Lectures suggérées : Fox (1999 : 291-318)</i></p> <p>Quiz 5 à faire en ligne (date limite le 16 mars)</p>
	 Labo 5. Configurer les sorties de SPSS aux normes de présentation APA et à l'aide de ce logiciel, visualiser une relation multivariée à l'aide de diagrammes en barres complexes, construire un tableau d'ordre zéro et des tableaux partiels, calculer les chi-carrés, les gammas des tableaux partiels et le gamma partiel. Lire et interpréter les sorties SPSS de façon à établir un diagnostic causal correct en termes de spécification, mais aussi d'explication et de dissimulation. Présenter les résultats sous forme textuelle et dans des tableaux et schémas publiables à l'aide du logiciel Word. <p>Retour sur le TP#1</p> <p>TP#2 à faire : tableaux multivariés, régression linéaire multiple (séances 4, 5, 6)</p>
Semaine 6	
Sam. 16 mars	 Leçon 6. Régression et corrélation linéaire multiples <ul style="list-style-type: none"> • Diagramme de dispersion et modélisation d'une relation entre deux variables quantitatives, en contrôlant l'effet d'une autre variable quantitative (convergence) • Prédiction à l'aide du plan de régression • Schématisation du modèle et interprétation des bêtas, r et r² multiple aux fins de diagnostic causal • Vérification d'un postulat de plus : l'absence de multicollinéarité <p><i>Lectures suggérées : Fox (1999 : 319-337)</i></p> <p>Quiz 6 à faire en ligne (date limite le 23 mars)</p>
	 Labo 6. À l'aide de SPSS, estimer un modèle de régression linéaire multiple incluant deux prédicteurs ou plus, sortir le diagramme de dispersion des résidus pour procéder au diagnostic du modèle multivarié, vérifier la multicollinéarité au moyen de la matrice de corrélations et des statistiques de la tolérance et du facteur d'inflation de la variance (VIF=Variance inflation factor). Lire et interpréter les sorties SPSS de façon à établir un diagnostic causal en termes de convergence (effet additif) ou non. Présenter les résultats sous forme textuelle et dans des tableaux et schémas publiables à l'aide du logiciel Word.

Semaine 7	
Sam. 23 mars	 Leçon 7. Régression logistique binaire <ul style="list-style-type: none"> • Parallèle entre régression linéaire et régression logistique binaire • Quels liens entre probabilité, cote et logit ? • Régression binaire pour une variable dépendante (VD) dichotomique • Extension du modèle simple au modèle de régression binaire multiple (discrimination) <p><i>Lectures suggérées : Morgan & Teachman (1988)</i></p> <p>Quiz 7 à faire en ligne (date limite le 30 mars)</p>
	 Labo 7. À l'aide de SPSS, comparer l'analyse tabulaire bivariée et la régression logistique binaire simple au moyen du rapport de cotes, calculer les probabilités prédites et le taux de classement. Estimer un modèle de régression logistique binaire multiple en incluant deux prédicteurs ou plus. Lire et interpréter les sorties SPSS de façon à établir un diagnostic causal en termes de discrimination ou non. Présenter les résultats sous forme textuelle et dans des tableaux et schémas publiables à l'aide du logiciel Word. <p>Date limite de remise du TP#2</p>
Semaine 8	
Sam. 30 mars	 Leçon 8. Régression logistique multinomiale <ul style="list-style-type: none"> • Régression binaire vs. régression multinomiale • Régression multinomiale pour une VD qualitative à trois catégories ou plus • Calcul des probabilités prédites • Extension du modèle simple au modèle de régression multinomiale multiple (discrimination) <p><i>Lectures suggérées : Hegre (2011)</i></p> <p>Quiz 8 à faire en ligne (date limite le 6 avril)</p>
	 Labo 8. À l'aide de SPSS, comparer l'analyse tabulaire bivariée et la régression logistique multinomiale simple au moyen du rapport de cotes, calculer les probabilités prédites et le taux de classement. Estimer un modèle de régression logistique multinomiale multiple en incluant deux prédicteurs ou plus. Lire et interpréter les sorties SPSS de façon à établir un diagnostic causal en termes de discrimination ou non. Présenter les résultats sous forme textuelle et dans des tableaux et schémas publiables à l'aide du logiciel Word. <p>Retour sur le TP#2</p>
Semaine 9	
Sam. 6 avr.	<p>Synthèse du cours et révision en vue de la préparation de l'examen</p> <p>Rencontres individuelles</p>



Semaine 10

Sam. 13 avr.	Examen sur table : séances 7 & 8 Salle D01, 8h-10h!
-----------------	---

RÉFÉRENCES BIBLIOGRAPHIQUES

- Boudon, R. 2002. *Les méthodes en sociologie*, Paris : PUF, p.5-40.
- Cousineau, D., B. Harding (2017). "Pourquoi les statistiques sont-elles difficiles à enseigner et à comprendre? Quelques réflexions". *Revue de psychoéducation*, 46(2), p.397-419.
- Fox, W. (1999). *Statistiques sociales*, Québec, Les Presses de l'Université Laval, Traduit de l'anglais et adapté par L.M. Imbeau.
- Hegre, H. (2011). "Multinomial Logistic Regression". Translation of chapters 3 in Hegre, H. 2011. *Logistisk regresjon: binomisk, multinomisk og rangert*. 21 p.
- Martin, Olivier (2009). *L'analyse de données quantitatives*. Armand Colin, coll. « 128 », 124 p.
- Morgan, S. P. et J. D. Teachman (1988), "Logistic Regression: Description, Examples, and Comparisons". *Journal of Marriage and Family*, 50 (4): 929-936.
- Tacq, J. 1997. "A Number of research Examples and their Basic Format". In *Multivariate Analysis Techniques in Social Science: From Problem To Analysis*, London: Sage Publications, p.8-19.
- Touré, El. H. (2022). *Statistiques sociales avancées avec SPSS*. Cahier de laboratoire. St-Louis : Université Gaston Berger.

ASPECTS RÉGLEMENTAIRES

Remise des travaux

Aucun retard dans la remise des travaux ne sera toléré, à moins d'une raison sérieuse acceptée par le professeur. Pour la remise des TP, une pénalité de 10 % de la note du travail en question sera appliquée par jour de retard. Pour le quiz, il sera fermé une semaine après sa mise en ligne.

Qualité de la langue

La notation prend en compte autant la justesse des calculs que la qualité de la rédaction dans les réponses et interprétations. Lorsque la qualité de la langue française fait défaut, une pénalité s'applique jusqu'à concurrence de 10% de la note obtenue.

Méfiez-vous du plagiat et de la tricherie

Le plagiat, sous toutes ses formes, ainsi que la tricherie sont pris très au sérieux par le professeur.