

SIMON LAFLAMME, Ph.D.

Université Laurentienne, département de sociologie

RUN-MIN ZHOU, Ph.D.

Université Laurentienne, département de psychologie

Méthodes statistiques en sciences humaines

Avec des illustrations tirées du logiciel SPSS

Cognitio

Érisc
départole

Ancrées dans le Nouvel-Ontario, les Éditions Prise de parole appuient les auteurs et les créateurs d'expression et de culture françaises au Canada, en privilégiant des œuvres de facture contemporaine.

La collection «Cognitio» publie des manuels et des ouvrages de synthèse en sciences humaines, destinés en priorité aux programmes postsecondaires du Canada français.

Prise
de parole

Éditions Prise de parole
C.P. 550, Sudbury (Ontario)
Canada P3E 4R2
www.prisedeparole.ca

Nous reconnaissons l'aide financière du gouvernement du Canada par l'entremise du Fonds du livre du Canada (FLC) et du programme Développement des communautés de langue officielle de Patrimoine canadien. La maison d'édition remercie également la Ville du Grand Sudbury de son appui financier.

La maison d'édition remercie également l'Université Laurentienne pour sa généreuse contribution à cet ouvrage.



Patrimoine
canadien

Canadian
Heritage



Université Laurentienne
Laurentian University

Méthodes statistiques en sciences humaines

Avec des illustrations tirées du logiciel *SPSS*

DES MÊMES AUTEURS

LIVRES DE SIMON LAFLAMME

- Initiation thématique à la sociologie*, Sudbury, Éditions Prise de parole, coll. « Cognitio », 2008. Avec Jean Lafontant.
- Suites sociologiques*, Sudbury, Éditions Prise de parole, coll. « Épistémè », 2006.
- Homogénéité et distinction*, Sudbury, Éditions Prise de parole, coll. « Ancrages », 2003. Avec Ali Reguigui.
- Des biens, des idées et des personnes au Canada, 1981-1995. Analyse macrologique relationnelle*, Sudbury, Éditions Prise de parole / Paris, L'Harmattan, 2000.
- Deux groupes linguistiques, une communication de masse*, Montréal / Paris, L'Harmattan, coll. « Logiques sociales », 1997. Avec Ali Reguigui.
- Humain objet, humain sujet. Initiation à quelques notions de philosophie de l'histoire et d'épistémologie des sciences humaines*, « Série monographique en sciences humaines », Sudbury, Institut franco-ontarien, 1996.
- Communication et émotion. Essai de microsociologie relationnelle*, Paris, L'Harmattan, coll. « Logiques sociales », 1995.
- Être un être social*, Montréal, Guérin, 1994. Avec Christiane Bernier.
- La société intégrée. De la circulation des biens, des idées et des personnes*, New York / Berne, Peter Lang, Worcester Polytechnic Institute, Studies in Science, Technology and Culture, 1992.
- L'ambition démesurée. Enquête sur les aspirations et les représentations des étudiants et des étudiantes francophones du nord-est de l'Ontario*, Sudbury, Éditions Prise de parole / Institut franco-ontarien, coll. « universitaire », 1990. Avec Donald Dennie.
- Contribution à la critique de la persuasion politique*, Sillery (Québec), Presses de l'Université du Québec, 1987.

LIVRES DE RUN-MIN ZHOU

- La pensée de l'enfant et l'éducation préscolaire* (en chinois), Beijing, Presses Tuanjie, 1990.
- Le monde psychologique de l'enfant* (en chinois), Beijing, Presses de l'Université de Beijing, 1989. Avec Fuxi Fang, Ge Fang, Run-Min Zhou et Rui Huan Wu.

Méthodes statistiques en sciences humaines

Avec des illustrations tirées du logiciel *SPSS*

SIMON LAFLAMME
ET RUN-MIN ZHOU

COLLECTION COGNITIO
Éditions Prise de parole
Sudbury 2014

Nous tenons à remercier International Business Machines Corporation (IBM) de nous avoir permis de reproduire des captures d'écran de résultats générés grâce au logiciel *SPSS*. Nous reconnaissons le © *SPSS*, une compagnie IBM, sur l'ensemble des résultats ainsi reproduits.

Conception de la première de couverture : Olivier Lasser

Tous droits de traduction, de reproduction
et d'adaptation réservés pour tous pays.
Imprimé au Canada.
Copyright © Ottawa, 2014

Diffusion au Canada : Dimedia

Catalogage avant publication de Bibliothèque et Archives Canada

Laflamme, Simon, auteur

Méthodes statistiques en sciences humaines / Simon Laflamme et Run-Min Zhou.

(Cognitio) Comprend des références bibliographiques et un index.

Publié en formats imprimé(s) et électronique(s).

ISBN 978-2-89423-913-1.– ISBN 978-2-89423-759-5 (pdf). – ISBN 978-2-89423-888-2 (epub)

I. Sciences humaines – Méthodes statistiques – Manuels d'enseignement supérieur.

I. Zhou, R.-M., auteur II. Titre. III. Collection: Collection Cognitio

HA29.5.F7L34 2014 300.72'7 C2014-904903-X

C2014-904904-8

ISBN 978-2-89423-913-1 (Papier)

ISBN 978-2-89423-759-5 (PDF)

ISBN 978-2-89423-888-2 (ePub)

REMERCIEMENTS

Notre reconnaissance est grande envers tous les étudiants qui, au cours des années, ont suivi nos cours de méthode quantitative et nous ont permis d'affiner notre enseignement.

Elle est grande aussi envers plusieurs personnes qui ont contribué concrètement, d'une façon ou d'une autre, à la réalisation de l'ouvrage. Nous pensons à Carole G. Anderson, Danielle Barbeau-Rodrigue, Frédérique Blais, Mélanie Girard, Paul Jalbert, Leslie Laczko, Sylvie Lafrenière, Sylvie Landry, Ali Reguigui et Denise Truax de même qu'à toute l'équipe de Prise de parole. À chacun et chacune, nous exprimons clairement notre gratitude.

TABLE DES MATIÈRES

REMERCIEMENTS	5
1. INTRODUCTION GÉNÉRALE	7
1.1. Pourquoi un autre manuel	7
1.2. Que fait la statistique	8
1.3. La statistique comme outil	9
1.4. Objectif du manuel	11
1.5. Structure du manuel	12
1.6. Suggestions pour l'usage du manuel	13
PREMIÈRE PARTIE	
NOTIONS GÉNÉRALES	17
2. LES VARIABLES	19
2.1. L'analyse quantitative et les variables	19
2.2. La nature des variables	21
2.2.1. Les variables nominales	21
2.2.2. Les variables ordinales	22
2.2.3. Les variables cardinales	26
2.2.4. Comparaison entre les trois niveaux de variables	34
2.2.5. Les relations entre les variables	35
3. STATISTIQUES DESCRIPTIVES UNIVARIÉES	37
3.1. Statistiques descriptives et variables nominales	37
3.2. Statistiques descriptives et variables ordinales	44
3.3. Statistiques descriptives et variables cardinales	53

4. L'INFÉRENCE.....	79
4.1. Population et échantillon	79
4.2. Principe de probabilité	87
4.3. La distribution normale centrée réduite et la probabilité.....	88
4.4. La distribution normale centrée réduite et la standardisation	94
4.5. La distribution normale centrée réduite et l'inférence.....	98
4.5.1. L'intervalle de confiance	103
4.5.1.1. L'intervalle de confiance pour les moyennes	103
4.5.1.2. L'intervalle de confiance pour les proportions.....	109
4.5.2. Le test d'hypothèse.....	114
4.5.2.1. Le test d'hypothèse pour une seule variable :	
l'exemple de la moyenne	118
4.5.2.2. Le test d'hypothèse pour une seule variable :	
l'exemple de la proportion.....	133
4.6. Le test sur la forme de la distribution de Kolmogorov-Smirnov	138
4.7. La démarche en résumé.....	141
4.8. Inférence paramétrique et inférence non paramétrique	142
4.9. Cumul de tests inférentiels	147

DEUXIÈME PARTIE

AUTOUR DES TABLEAUX DE CONTINGENCE.....	149
INTRODUCTION SUR LES TABLEAUX DE CONTINGENCE	151
5. LE CHI-CARRÉ.....	153
5.1. La construction des tableaux de contingence	153
5.2. La question du nombre dans les tableaux de contingence	155
5.3. L'inférence des tableaux de contingence	158
5.3.1. La fréquence théorique et le hasard	159
5.3.2. La statistique chi-carré.....	160
5.3.3. Le degré de liberté.....	161
5.3.4. L'usage du chi-carré	163
5.3.5. Une fréquence théorique trop petite et	
une fréquence observée trop faible	168
5.3.6. Démarche pour l'interprétation des résultats	170
5.3.7. Quand utiliser le chi-carré ?	171
5.3.8. Les tableaux de contingence 2 x 2 :	
le correctif de Yates	173

5.3.9. Les tableaux de contingence 2 x 2 :	
le test exact de Fisher	176
5.3.10. Chi-carré corrigé ou test de Fisher	180
5.3.11. Test de McNemar	180
6. L'ANALYSE D'UNE SEULE VARIABLE NOMINALE	185
6.1. Le test binomial	185
6.1.1. Le test binomial sur l'égalité des fréquences	185
6.1.2. Le test binomial par rapport à une distribution théorique.....	188
6.2. Le test d'ajustement du chi-carré	191
6.2.1. Le test d'ajustement du chi-carré	
sur l'égalité des fréquences.....	191
6.2.2. Le test d'ajustement du chi-carré	
par rapport à une distribution théorique.....	195
7. LE LOG-LINÉAIRE.....	199
7.1. Au-delà du chi-carré	199
7.2. L'analyse log-linéaire hiérarchique avec <i>SPSS</i>	202
7.3. L'analyse log-linéaire généralisée avec <i>SPSS</i>	205
8. LES MESURES D'ASSOCIATION POUR	
LES TABLEAUX DE CONTINGENCE	209
8.1. Le coefficient phi.....	210
8.2. Le coefficient de contingence	216
8.3. Le coefficient V de Cramér	219
8.4. Le coefficient lambda (de Goodman et Kruskal)	223
8.5. Le coefficient kappa.....	227
8.6. Le coefficient gamma (de Goodman et Kruskal)	230
8.7. Le coefficient tau de Kendall	234
8.8. L'association linéaire	238
8.9. Le coefficient eta	239
8.10. Tests de Cochran et de Mantel-Haenszel.....	242
8.11. Tableau en forme de résumé	245
TROISIÈME PARTIE	
AUTOUR DE LA COMPARAISON	247
INTRODUCTION SUR LA COMPARAISON	249

9. LE TEST T.....	251
9.1. Deux échantillons et une mesure	251
9.2. Deux mesures et un échantillon	257
10. L'ANALYSE DE VARIANCE (ANOVA)	263
10.1. L'analyse de variance à un facteur	263
10.2. L'analyse de variance à plus d'un facteur.....	271
10.3. L'analyse de variance avec covariante.....	286
10.4. L'analyse de variance pour un échantillon et plus de deux mesures	290
10.5. L'analyse de variance mixte	295
11. L'ANALYSE DE VARIANCE MULTIVARIÉE (MANOVA).....	305
11.1. Exemple d'une analyse de variance multivariée à un facteur.....	307
11.2. Exemple d'une analyse de variance à 2 facteurs	314
12. LES TESTS W DE WILCOXON ET U DE MANN-WHITNEY	325
13. LE TEST KRUSKAL-WALLIS.....	335
14. LE TEST DE LA MÉDIANE	343
15. LE TEST DU SIGNE	349
16. LE TEST DES RANGS SIGNÉS DE WILCOXON	353
17. LE TEST DE FRIEDMAN.....	359
QUATRIÈME PARTIE	
AUTOUR DE L'ASSOCIATION	365
INTRODUCTION SUR L'ASSOCIATION	367
18. LA CORRÉLATION DE PEARSON.....	369
18.1. Une relation linéaire.....	369
18.2. La nécessité de la pente.....	373
18.3. L'intensité de la relation	375
18.4. Le calcul de la relation.....	375
18.5. Un exemple	376
18.6. Pour l'inférence.....	378
18.7. Avec le logiciel SPSS	378
18.8. Un autre exemple	380

18.9. Quand on rapporte plusieurs corrélations	381
18.10. Autres considérations	383
19. LA CORRÉLATION DE SPEARMAN	385
20. LA CORRÉLATION PARTIELLE ET LA SEMI-PARTIELLE	391
20.1. La corrélation partielle	391
20.2. La corrélation semi-partielle	395
21. L'ANALYSE DE RÉGRESSION	399
21.1. Éléments de base pour la régression	400
21.2. Illustration mathématique	404
21.3. Illustration analytique	412
21.4. La régression multiple	414
21.4.1. La contribution marginale	415
21.4.2. Les variables dichotomiques	420
21.4.3. Les méthodes de sélection des variables	424
21.4.3.1. Entrée forcée par étape	425
21.4.3.2. Élimination par bloc	429
21.4.3.3. Entrée une à une en fonction d'un critère d'insertion	429
21.4.3.4. Retrait une à une en fonction d'un critère d'exclusion	430
21.4.3.5. Entrée et exclusion combinées	432
21.4.4. Régression et interaction	435
21.4.4.1. Deux régresseurs cardinaux	436
21.4.4.2. Un régresseur cardinal et un régresseur dichotomique	439
22. L'ANALYSE DE CHEMINEMENT	447
23. LA RÉGRESSION LOGISTIQUE	453
23.1. La régression logistique binaire	453
23.2. La régression logistique multinomiale	469
23.3. La régression logistique ordinale	479
24. LA RÉGRESSION DE COX	493
25. LE COEFFICIENT DE CONCORDANCE W DE KENDALL	503
26. L'ANALYSE DISCRIMINANTE	509
27. L'ANALYSE FACTORIELLE	523

27.1. La logique de l'analyse factorielle.....	524
27.2. L'extraction de facteurs.....	524
27.2.1. L'extraction du premier facteur	525
27.2.2. L'extraction du deuxième facteur et des facteurs subséquents	525
27.3. Le nombre de facteurs.....	526
27.4. La rotation des facteurs	526
27.5. Deux types d'analyse factorielle	527
27.5.1. Un exemple fictif de l'analyse factorielle confirmatoire	527
27.5.2. Un exemple fictif de l'analyse factorielle exploratoire	539
28. L'ANALYSE DE REGROUPEMENT	545
28.1. La méthode « nuées dynamiques ».....	547
28.2. La méthode « Classification TwoStep ».....	552
BIBLIOGRAPHIE.....	559
ANNEXE RÉCAPITULATIVE SUR LES STATISTIQUES INFÉRENTIELLES EN FORME DE TABLEAU	563
ANNEXE RÉCAPITULATIVE SUR LES SYMBOLES	565
INDEX	569

1. INTRODUCTION GÉNÉRALE

1.1. Pourquoi un autre manuel

Il existe de nombreux manuels de statistique. Et la plupart d'entre eux ont de belles qualités. Malgré cela, nous avons choisi d'en produire un autre. Nous avons choisi de le faire pour plusieurs raisons.

Il y a, d'abord, que nous ne rencontrons pas d'ouvrage devant lequel nos étudiants trouvent aisément ce dont ils ont besoin pour comprendre théoriquement ce qu'ils peuvent ou doivent faire et pour interpréter pratiquement les résultats des analyses dans le cadre de leurs préoccupations de chercheur virtuel ou actif.

Il y a, ensuite, que bon nombre de nos collègues viennent vers nous, manuel en main, pour savoir quoi faire avec telle ou telle variable, devant telle ou telle problématique, ou pour comprendre les pages que génère un logiciel, le manuel qu'ils portent avec eux ne leur étant pas utile.

Il y a, troisièmement, que nous enseignons les statistiques depuis plusieurs années, souvent ensemble tous les deux devant un groupe d'étudiants, et que nous avons mis au point quelques stratégies pédagogiques dont nous pensons qu'elles pourraient être utiles en dehors de nos salles de classe.

Il y a, quatrièmement, que, depuis longtemps, nous produisons des travaux de recherche et dirigeons des thèses, et que cela nous a indiqué à quel point il importait d'aborder la statistique dans la perspective pratique du chercheur ou de son lecteur qui ne sont pas forcément férus de statistiques, mais qui ont envie ou besoin d'en comprendre les rudiments dans le contexte d'analyses particulières.

Il y a, cinquièmement, qu'on dénombre peu d'ouvrages renfermant l'ensemble des techniques statistiques les plus usuelles, du non-paramétrique au paramétrique, de la régression simple au log-linéaire, du test t à l'analyse de variance qui comporte plus d'un facteur, des différences selon les groupes aux différences pour des échantillons appariés.

Il y a, sixièmement, que la plupart des manuels rendent compte des techniques statistiques, mais ne les raccordent pas aux résultats que génèrent des logiciels alors que l'immense majorité des chercheurs, pour presque toutes leurs analyses, sont exposés à ce type de présentation des résultats.

Il y a, septièmement, que les ouvrages dans lesquels la statistique est enseignée d'après les résultats que génèrent les logiciels sont généralement trop peu explicites pour ce qui est des motifs du recours à telle ou telle technique et en ce qui a trait aux finalités statistiques elles-mêmes.

Il y a, huitièmement, que, depuis longtemps, collègues et anciens étudiants nous réclament cet ouvrage et qu'il nous semble le leur devoir.

1.2. Que fait la statistique

Quand on l'observe en surplomb, on constate que la statistique sait faire bien peu de choses. Sur un plan descriptif :

- i. elle offre des façons de porter des jugements quantitatifs sur des informations qui ont été obtenues dans un ensemble d'individus ;
- ii. elle permet de juger quantitativement du lien entre deux ensembles d'informations ou plus.

Sur un plan inférentiel, elle permet d'estimer l'erreur de se tromper quand les données proviennent d'un échantillon et qu'on espère extrapoler les observations, qu'il s'agisse :

- i. d'informations sur un seul objet,
- ii. de comparaisons entre des groupes,
- iii. de la quantification de liens.

La statistique est donc simple par son approche globale. Ce qui la rend compliquée, c'est qu'il y a de nombreuses façons de quantifier les liens et de procéder à des comparaisons. Mais si on a bien à l'esprit sa façon globale d'approcher les données, on ne se laisse pas distraire par les ramifications de ces approches.

1.3. La statistique comme outil

Il y a des chercheurs en statistique. Ce sont des mathématiciens qui se consacrent à l'évolution de leur discipline, peaufinant les modèles existants et en créant de nouveaux. Mais, pour la plupart des chercheurs qui doivent recourir aux méthodes quantitatives, la statistique n'est qu'un outil, elle est un ensemble de techniques mis à la disposition des chercheurs qui souhaitent décrire quelque chose qui n'est pas statistique. Il y a peu de statisticiens en dehors des mathématiques, mais il y a des utilisateurs de la statistique dans presque toutes les disciplines. Toute façon de décrire des données n'est pas statistique ; mais il y a peu de disciplines universitaires au sein desquelles l'ensemble des chercheurs puisse se dispenser de faire usage de la statistique.

Ce manuel s'adresse aux utilisateurs de la statistique, non pas à ses fabricants. Il s'adresse aux chercheurs pour lesquels la statistique est un outil, un instrument de description. Mais l'outil statistique – comme tous les autres instruments – permet de produire des résultats d'autant plus élégants et solides que l'ouvrier sait le manipuler. En ce sens, s'il est vrai que la statistique représente un mode de description dans la plupart des disciplines, elle peut aussi façonner. L'utilisateur de la statistique décrit en façonnant. Pour obtenir une moyenne sur les manières de penser des individus, il faut quantifier ces manières de penser. Cette quantification des pensées est déjà du façonnement. Elle rendra possible des dénombrements, puis des calculs, par exemple, de moyennes ; donc elle permettra de décrire. Le chercheur qui sait bien employer l'outil pourra aussi établir des liens entre des pensées, entre ces pensées et d'autres objets, pourra les comparer selon les caractéristiques des personnes qui les expriment, pourra prendre des ensembles de pensées et les associer à des ensembles de déterminants. Tout cela lui permettra de décrire ce qui se donne à l'observation, mais en le modulant au gré de son habileté à manier son outil.

La statistique est un outil de description qui commande une quantification de ce qui se donne à l'observation. Elle est un outil dont l'usage dépend des habiletés du chercheur, et donc un outil qui décrit en fabriquant, selon la compétence de l'ouvrier. Cela n'est pas un mal en soi. Il n'y a pas de description du monde, quelles que soient les méthodes qui la sous-tendent, qui ne suppose simultanément quelque façonnage. La statistique n'est qu'une de ces méthodes ; elle n'est qu'une méthode parmi d'autres. Comme toutes les autres méthodes, tous les autres outils d'analyse empirique, elle a ses forces et ses faiblesses. C'est pourquoi, d'ailleurs,

il est souvent sage d'accompagner l'analyse quantitative d'autres types d'analyse, et donc d'enrichir une étude de ce qu'elle fait de mieux et de combler ses lacunes en puisant ailleurs. La statistique est un outil qui est au service des chercheurs dans d'innombrables disciplines, au service de leurs préoccupations théoriques, de leurs suspicions, de leurs craintes, de leurs espoirs. En cela, elle n'est pas différente des autres outils. Et tout comme les autres outils, elle ne vaut que ce qu'on sait en faire et que ce qu'elle permet de faire. La statistique est un outil scientifique important, mais elle ne peut faire que ce qu'elle sait faire : une analyse quantifiée du monde est souvent fort utile, mais elle est souvent critiquable. Comme l'est toute autre analyse. Une analyse statistique, par ailleurs, ne vaut jamais plus que les données sur lesquelles elle repose. On aura beau recourir aux techniques les plus sophistiquées, si les données que ces techniques manipulent, au point de départ, ne sont pas valables, les résultats qu'elles produisent ne vaudront rien. À l'inverse, même les données les moins critiquables ne peuvent être analysées, dans le cadre étroit de la statistique, que si les techniques qui les exploitent prennent en considération les fins de la recherche. En dehors d'un cadre théorique, les statistiques sont peu utiles.

L'analyse statistique devrait aussi bien susciter l'enthousiasme du chercheur qu'inviter ce chercheur à entretenir un esprit critique.

On aime rabâcher qu'on peut faire dire n'importe quoi aux statistiques. Peut-être ! Mais les statistiques, elles, ne peuvent pas dire n'importe quoi. Elles ne peuvent pas dire qu'il y a eu 3 000 homicides au Canada, en 2010, quand il y en a eu 554. Elles ne peuvent pas dire que le taux d'homicide au Canada augmente depuis 40 ans quand il diminue. Elles ne peuvent pas dire que, au Canada, en 2008, le salaire moyen des personnes qui travaillent à plein temps était, chez les femmes, à 50 % du salaire des hommes quand ce quotient était de 71 %. Les spécialistes des disciplines à l'intérieure desquelles ces statistiques sont produites peuvent en questionner la pertinence, peuvent en chercher les déterminants, peuvent même proposer d'autres modes de compilation – en faisant valoir normalement des raisons théoriques. Mais ces statistiques elles-mêmes sont bien ce qu'elles sont. Le chercheur qui les livre doit pouvoir indiquer comment les données ont été obtenues, théoriquement et empiriquement, et comment ont été faites les compilations. On peut trouver que 554 meurtres, c'est peu ou c'est beaucoup ; mais on ne peut pas dire que le chiffre n'est pas 554, à moins de mettre en lumière des erreurs méthodologiques ou des biais théoriques. On peut trouver qu'un salaire des femmes à 71 % de celui des hommes, c'est là un trop grand écart ou que cela se comprend aisément ; mais on ne

peut pas dire que ce ne sont pas 71 %, à moins de faire valoir des problèmes théoriques ou méthodologiques. L'usage des statistiques en dehors d'un cadre théorique est douteux. On peut peut-être faire dire n'importe quoi aux statistiques, mais il reste que ce sont elles qui montrent s'il y a un écart de salaire entre les hommes et les femmes, s'il existe une corrélation entre la consommation de cigarettes et la probabilité de contracter le cancer du poumon, si le taux de chômage croît avec les années, si les performances scolaires varient en fonction du milieu. Le caractère polémique de la recherche n'est pas le propre de la statistique ; il est inhérent à la démarche scientifique elle-même.

1.4. Objectif du manuel

L'objectif de notre ouvrage est triple :

- i. fournir les principes de base pour l'usage de diverses statistiques ;
- ii. illustrer comment interpréter les résultats qui, à la lumière de ces principes, sont générés par le logiciel le plus utilisés en sciences sociales : *Statistical Package for the Social Sciences (SPSS)* ;
- iii. indiquer comment rapporter les résultats.

Notre ouvrage se veut donc une initiation à la statistique, mais surtout un guide pour l'usage des techniques statistiques. Il ne représente pas une contribution à la statistique en elle-même. Il est destiné à ceux qui veulent apprendre les méthodes quantitatives ou à ceux qui ont besoin d'un outil qui leur indiquera quelles techniques employer dans quelles circonstances, comment interpréter les résultats et comment livrer ces résultats dans un texte.

Puisqu'il en est ainsi, nous n'inventons rien de proprement statistique dans cet ouvrage. Nous sommes infiniment tributaires de nombreux auteurs. Toutes les formules qui jalonnent le texte ont été inventées par d'autres auteurs. Nous avons fait comme dans la plupart des manuels : nous avons traité ces formules comme si elles appartenaient au domaine public, et n'avons donc pas constamment renvoyé aux créateurs, ni aux innombrables ouvrages dans lesquels on retrouve ces formules (généralement sans référence)¹. Nous avons toutefois, autant que faire se peut, associé à leur auteur les techniques quand la dénomination le rendait possible : test de McNemar, test de Fisher, tau de Kendall, corrélation de Pearson...

¹ Nous signalons au passage que, dans le logiciel *SPSS*, pour presque toutes les techniques statistiques, il est possible, dans les registres d'aide, de trouver les formules qui sont mises en application.

Certes il en aurait été autrement si notre ouvrage avait constitué une contribution à la statistique; il aurait alors fallu citer les statisticiens avec lesquels nous serions entrés en discussion dans une entreprise de création. Mais tel n'est pas la fin d'un manuel de statistique. Constamment, dans le même esprit, au cours de la rédaction de l'ouvrage, nous nous sommes interrogés sur ce que nous devons à divers auteurs. Dans bien des cas, il était impossible de fournir une réponse tant les emprunts sont importants. Comment identifier ce que nous devons à un manuel qui nous a initiés à la statistique et dont nous nous sommes nous-mêmes servi pour enseigner, comment signaler les pages qui marquent cette influence dont nous avons joui? Comment reconnaître ce que nous devons à des ouvrages classiques. Nous ne le savons pas. Nous sommes redevables à plusieurs auteurs de notre façon même de nous exprimer. Nous ne savons plus ce que nous devons aux livres de Gilbert et Savard², Siegel et Castellan³, Wonnacott et Wonnacott⁴, Diekhoff⁵, Baillargeon⁶, Darlington⁷, Howell⁸, et aux livres de bien d'autres auteurs dont on trouvera les titres en bibliographie. Nous savons que c'est beaucoup. Nous leur devons ce que l'élève doit au maître et cela est peu définissable autant qu'infini. Sans leur enseignement, nous n'aurions jamais pu produire notre ouvrage.

1.5. Structure du manuel

Notre manuel comporte 4 grandes parties. La première fait état des principes généraux. Elle distingue les variables, initie aux statistiques descriptives univariées et présente les notions d'inférence. La deuxième partie se consacre aux tableaux de contingence; elle va du chi-carré au log-linéaire et

² Norma Gilbert et Jean-Guy Savard, *Statistiques*, 2^e édition revue et corrigée, Laval (Québec), Études vivantes, [1976] 1992.

³ Sidney Siegel et N. John Castellan Jr., *Nonparametric Statistics for Behavioral Sciences*, 2^e édition, New York, McGraw-Hill, [1956] 1988.

⁴ Thomas H. Wonnacott et Ronald J. Wonnacott, *Statistique. Économie, gestion, sciences, médecine (avec exercices d'application)*, 4^e édition, Paris, Économica, [1972] 1991.

⁵ George Diekhoff, *Statistics for the Social and Behavioral Sciences: Univariate, Bivariate, Multivariate*, Dubuque (IA), Brown, 1992.

⁶ Gérald Baillargeon, *Introduction à la statistique descriptive. Avec applications dans différents secteurs de la technologie, de la gestion et du contrôle de la qualité*, volume 1, Trois-Rivières (Québec), SMG, Science, mathématique, gestion, 1981 et *Introduction au calcul des probabilités. Notions sur les probabilités, les lois binomiale, de Poisson, normale et une introduction aux notions sur l'échantillonnage*, volume 2, Trois-Rivières (Québec), SMG, Science, mathématique, gestion, 1981.

⁷ Richard B. Darlington, *Regression and Linear Models*, New York, McGraw-Hill, coll. «Psychology», 1990.

⁸ David C. Howell, *Méthodes statistiques en sciences humaines*, 2^e édition, Bruxelles, De Boeck, 2008, p. 73-79.

comprend les mesures d'association qui peuvent se rapporter à ces tableaux de contingence. La troisième partie porte sur les statistiques qui ont principalement pour fin d'établir l'inféribilité de comparaisons de mesures de tendance centrale; elle s'étend des tests *t* aux analyses de variance et du test de Wilcoxon au test de Friedman. La quatrième partie présente les statistiques qui ont trait à l'association et à son inféribilité en dehors des tableaux de contingence; c'est là que sont décrites les mesures de corrélation, les analyses de régression, l'analyse de cheminement, le coefficient de concordance, l'analyse factorielle et l'analyse de regroupement.

À l'intérieur de chacune de ces parties, pour chacune des techniques, le lecteur trouvera une séquence récurrente arrimée aux objectifs du manuel: une présentation des principes de la technique et des motifs de son utilisation; une illustration de résultats générés par le logiciel *SPSS* et de la manière de les interpréter; un exemple de la manière de rapporter les résultats.

Chaque fois que, au fil des pages, nous présentons une technique statistique, nous indiquons comment obtenir des résultats avec le logiciel *SPSS* et comment les interpréter. Mais ce travail suppose que l'utilisateur dispose déjà d'une matrice de données. Nous tenons pour acquis que le lecteur sait comment saisir des données dans une matrice dont les lignes représentent des individus et les colonnes, des variables.

1.6. Suggestions pour l'usage du manuel

Dans la plupart des établissements d'enseignement, la matière qui doit être couverte dans le cadre d'un cours en particulier est déjà prévue. Les professeurs et les étudiants trouveront aisément cette matière dans le manuel que nous proposons. Il est possible que des enseignants eussent privilégié un ordre autre que celui que nous proposons, mais il nous semble peu probable qu'ils ne reconnaissent pas quelque pertinence à ordonner le contenu d'un manuel de statistique comme nous l'avons fait. Si la matière qui doit être dispensée n'est pas *a priori* définie, et que l'intention est d'initier aux contenus élémentaires les plus usités, alors nous proposons de suivre la séquence suivante:

- i. dans la première partie:
 - a. chapitre 2 sur les variables;
 - b. chapitre 3 sur les statistiques univariées;
 - c. chapitre 4 sur l'inférence;
- ii. dans la deuxième partie:
 - d. chapitre 5 sur le chi-carré;

- e. sections 8.1. et 8.3. du chapitre 8 sur quelques mesures d'association liées aux tableaux de contingence;
- iii. dans la troisième partie :
 - f. chapitre 9 sur les tests t;
 - g. sections 10.1. et 10.4. du chapitre 10 sur l'analyse de variance;
- iv. dans la quatrième partie :
 - h. chapitre 18 sur la corrélation Pearson;
 - i. sections 21.1., 21.2. et 21.3. du chapitre 21 sur la régression simple.

Les autres sections du manuel se greffent à ces contenus :

- i. soit dans un esprit de complexification, par exemple en présentant des outils destinés aux analyses multivariées plutôt que bivariées (analyse de variance à plus d'un facteur, analyse de variance mixte, analyse de variance multivariée, régression multiple);
- ii. soit en décrivant des techniques pour exploiter des données auxquelles n'ont pas accès les instruments les plus communs, des techniques, par exemple, comme :
 - a. des tests inférentiels pour une seule variable nominale (test binomial, test d'ajustement du chi-carré);
 - b. des tests inférentiels pour des variables dépendantes non cardinales (W de Wilcoxon ou U de Mann-Whitney, Kruskal-Wallis, test sur la médiane, test du signe, test des rangs signés de Wilcoxon, de Friedman);
 - c. des mesures d'association pour les variables non cardinales (coefficients gamma, tau de Kendall, corrélation de Spearman);
 - d. un test d'association pour une variable cardinale et une variable nominale (coefficient eta);
- iii. soit les deux à la fois, par exemple en montrant comment examiner les rapports :
 - a. entre plusieurs variables nominales (log-linéaire);
 - b. entre plusieurs variables ordinales (coefficient de concordance W de Kendall);
 - c. entre une variable dépendante cardinale et plusieurs variables indépendantes cardinales en insérant des variables dichotomiques (dans le prolongement de la régression multiple);

- d. entre une variable dépendante non cardinale et plusieurs variables indépendantes (analyse discriminante, régression logistique);
- iv. soit en dépeignant des outils destinés à d'autres usages (coefficient lambda, coefficient kappa, association linéaire, tests de Cochran et de Mantel-Haenszel, analyses de variance avec covariante, corrélation partielle, diverses méthodes pour la sélection des variables dans les analyses de régression ou dans le log-linéaire, examen de l'interaction des variables indépendantes dans l'analyse de régression, analyse de cheminement, régression Cox, analyse factorielle, analyse de regroupement).

Si la matière n'est pas définie par un programme, le professeur orientera sa classe vers la complexification (i) et lui fera découvrir des techniques de base qui se rapportent à des situations autres que celles étudiées quand auront été présentés les outils les plus communs (ii). Dans les cours plus avancés ou pour répondre à des attentes spécifiques, l'enseignant abordera les matières qui à la fois complexifient et se préoccupent de données non cardinales (iii) et celles dont les usages sont plus spécialisés (iv).

L'enseignant, au passage, qui ne dispose pas de données pour faire pratiquer ses étudiants ou pour évaluer leur apprentissage, trouvera de nombreux fichiers dans le logiciel *SPSS*. Une fois que le logiciel est installé dans un ordinateur, ces fichiers se situent dans le dossier « Samples » (Programmes\IBM\SPSS\Statistics\[numéro de la version du logiciel]\Samples). Des sites comme celui de Statistique Canada, par ailleurs, tiennent plusieurs matrices à la disposition du public, notamment des fichiers de microdonnées à grande diffusion.

L'autodidacte abordera le manuel comme il le ferait s'il était guidé par un enseignant.

La personne qui dispose déjà des notions élémentaires, qui, par exemple, est déjà au fait de toute la matière couverte par la première partie, et qui ou bien souhaite se rafraîchir la mémoire sur quelque technique ou bien voudrait découvrir ce qu'elle pourrait faire de certaines données, pourra parcourir le manuel au gré de ses questionnements. Si l'ensemble de l'ouvrage correspond à une architecture, les chapitres – et même plusieurs sections de ces chapitres – des deuxième, troisième et quatrième parties ont été conçus pour être lus séparément.