



paramètres

Robert R. Haccoun • Denis Cousineau

Statistiques

Concepts et applications

DEUXIÈME ÉDITION REVUE ET AUGMENTÉE



Les Presses de l'Université de Montréal

Extrait de la publication

STATISTIQUES

Page laissée blanche

Robert R. Haccoun
et
Denis Cousineau

STATISTIQUES

Concepts et applications

Deuxième édition revue et augmentée

Les Presses de l'Université de Montréal

Extrait de la publication

**Catalogage avant publication de Bibliothèque et Archives nationales du Québec
et Bibliothèque et Archives Canada**

Haccoun, Robert R.

Statistiques : concepts et applications

2^e éd.

Publ. à l'origine dans la coll. : Paramètres. 2007.

ISBN 978-2-7606-2113-8

eISBN 978-2-7606-2584-6

1. Statistique mathématique. 2. Probabilités. 3. Statistique. I. Cousineau, Denis, 1967- . II. Titre.

QA276.H185 2010

519.5

C2009-942664-1

Dépôt légal : 1^{er} trimestre 2010

Bibliothèque et Archives nationales du Québec

© Les Presses de l'Université de Montréal, 2010

Les Presses de l'Université de Montréal reconnaissent l'aide financière du gouvernement du Canada par l'entremise du Programme d'aide au développement de l'industrie de l'édition (PADIÉ) pour leurs activités d'édition.

Les Presses de l'Université de Montréal remercient de leur soutien financier le Conseil des arts du Canada et la Société de développement des entreprises culturelles du Québec (SODEC).

Imprimé au Canada en janvier 2010

À la « grande fille » de son Pappy, Orli Haya Abramson.

Robert Haccoun

À Élysabeth Aguila et Richard Shifrin, pour leur patience ∞

Denis Cousineau

Page laissée blanche

TABLE DES MATIÈRES

Avant-propos	9
Comment utiliser cet ouvrage	11
Chapitre 1 : La description des données	15
Chapitre 2 : La distribution des données	33
Chapitre 3 : Les statistiques descriptives.....	61
Chapitre 4 : La position relative des observations.....	101
Chapitre 5 : La distribution normale.....	129
Chapitre 6 : La corrélation.....	149
Chapitre 7 : La régression linéaire simple.....	183
Chapitre 8 : Les concepts de l'inférence statistique.....	215
Chapitre 9 : La mécanique de l'inférence statistique	251
Chapitre 10 : Une ou deux populations? Le test t.....	293
Chapitre 11 : L'analyse de variance à un facteur	327
Chapitre 12 : L'analyse de variance factorielle.....	369
Chapitre 13 : Les statistiques non paramétriques	395
Annexe	429
Réponses aux quiz rapides	443
Bibliographie	457

Page laissée blanche

AVANT-PROPOS

Si la plupart des programmes de premier cycle exigent que les étudiants suivent un cours de base en méthodes statistiques, c'est que cette formation est essentielle pour maîtriser les aspects scientifiques d'une discipline, notamment en sciences sociales.

Ce manuel est d'abord destiné aux étudiants qui suivront peut-être un seul cours de statistiques dans leur formation, mais il pourra également servir d'entrée en matière à ceux qui suivront des cours plus avancés. De la construction d'une distribution d'effectifs jusqu'à l'analyse de variance factorielle, il explique les fondements logiques, les résultats et les interprétations que les techniques statistiques permettent et celles qu'elles ne permettent pas. Nous avons délibérément laissé de côté les méthodes plus avancées qui permettent d'analyser des données expérimentales ou corrélationnelles complexes.

L'étude des statistiques suscite souvent des appréhensions qui ne facilitent pas l'apprentissage; les formules mathématiques d'apparence complexe peuvent en rebuter plus d'un. Nous avons donc voulu proposer une approche intuitive et graduelle qui soit rassurante. Bien entendu, les statistiques s'expriment par des formules, présentées dans ce volume, mais il ne faut pas perdre de vue que ces formules servent d'abord à rendre des concepts plus concrets. C'est pourquoi ce sont les concepts et non leur expression mathématique qui sont au cœur de notre approche. *Nous préférons utiliser le concept pour expliquer la formule, plutôt que d'utiliser la formule pour expliquer le concept.*

Nous croyons que la logique statistique peut être plus facilement comprise lorsqu'elle fait appel au raisonnement de l'étudiant et qu'elle s'appuie

sur des exemples capables de susciter son intérêt. C'est dans cet esprit que l'humour est parfois mis à contribution.

Cette deuxième édition maintient l'approche et l'esprit de l'édition originale, mais elle contient des changements importants : non seulement la présentation graphique et les textes ont été entièrement révisés, mais plusieurs sections ont été refondues pour les rendre plus claires, notamment celles décrivant les aspects plus complexes portant sur l'inférence statistique. On y trouve aussi de nouveaux contenus, dont un chapitre additionnel sur l'*analyse non paramétrique*.

Comme dans la première édition, chaque chapitre est ponctué de « quiz rapides » qui permettent aux étudiants de vérifier leur niveau de maîtrise des concepts et se termine par des questions à choix multiples. On y trouve évidemment les réponses aux uns et aux autres.

Le site Internet (www.pum.umontreal.ca/ca/fiches/978-2-7606-2113-8.html) reste actif. On y trouve, pour chaque chapitre du livre, des banques de données, des exercices et les commandes d'analyse pour le logiciel SPSS ; on y trouve également des discussions sur l'interprétation des résultats produits par le logiciel.

Nous tenons à exprimer nos remerciements aux personnes qui nous ont aidés à préparer cette nouvelle édition. Merci d'abord à Corinne Zacharyas pour sa lecture aussi attentive que généreuse, et à M. Élie Haccoun pour ses précieux conseils. Merci tout spécialement à nos familles et amis qui ont eu à supporter les écarts d'humeur qu'implique une entreprise de ce genre. Enfin, nous exprimons notre reconnaissance aux Presses de l'Université de Montréal, à son directeur, M. Antoine Del Busso, et à notre éditrice, Mme Natacha Monnier, pour leur soutien indéfectible.

Robert R. Haccoun
Denis Cousineau

COMMENT UTILISER CET OUVRAGE

Cet ouvrage explique les concepts et la logique statistiques; il est principalement destiné aux étudiants du premier cycle universitaire qui étudient la statistique. La présentation des concepts et techniques se fait de façon progressive: les premiers chapitres s'attachent à des aspects élémentaires (la construction d'une distribution, la nature des mesures de tendances centrales, les indices de dispersion, etc.), les chapitres suivants passent à des aspects plus élaborés (la logique de l'inférence, le test t, l'analyse de variance, etc.).

Afin de favoriser le développement graduel d'une compréhension intégrée de la statistique, chaque chapitre débute par un exemple qui fait le lien avec les connaissances acquises dans le chapitre antérieur. Chaque nouvel élément ou concept abordé dans le chapitre est illustré par des exemples concrets et simples, la plupart étant extraits de la vie quotidienne, certains étant même plutôt humoristiques. Il s'agit donc d'une approche volontairement «conviviale» de la statistique qui vise à susciter l'intérêt des étudiants pour lesquels cette matière semble trop souvent rébarbative.

Cependant, cet ouvrage n'est pas un livre de mathématiques! Les présentations et les explications n'exigent généralement pas une formation poussée en mathématiques. Si on y présente, comme il se doit, des formules statistiques de façon formelle, les explications qui les accompagnent décrivent leurs logiques plutôt que leurs dérivations algébriques. Et lorsque certaines preuves mathématiques sont requises, elles sont isolées du texte et placées dans des encadrés. Dans certains encadrés, le lecteur trouvera de brèves biographies, des anecdotes et d'autres exposés qui illustrent ou montrent l'origine des concepts qui sont abordés dans les chapitres.

Après avoir lu chaque chapitre, le lecteur devra normalement être en mesure de comprendre l'utilisation de la technique décrite, sa logique et les

interprétations qui peuvent ou non être faites à partir des résultats qu'elle génère.

Dans tous les chapitres, nous avons présenté des « quiz rapides ». Ces courts exercices permettent au lecteur de tester ses connaissances au fur et à mesure de la lecture de l'ouvrage. Les réponses à ces exercices se trouvent à la fin du volume.

À la fin de chaque chapitre, une série de questions à choix multiple est présentée et les réponses figurent à leur suite. Ces questions permettent aux lecteurs d'évaluer leur connaissance des concepts dans une forme qui s'apparente à celle des examens universitaires. Les réponses à ces questions exigent peu de calculs arithmétiques, voire n'en exigent aucun. L'accent est mis sur la compréhension des concepts et des techniques plutôt que sur la computation mécanique des formules pertinentes.

Comme accompagnement à ce manuel, des fichiers contenant plusieurs banques de données, des explications du logiciel d'analyse statistique SPSS et des exercices pouvant être analysés avec ce logiciel sont disponibles sur le site Internet des Presses de l'Université de Montréal à l'adresse suivante : www.pum.umontreal.ca/ca/fiches/978-2-7606-2113-8.html. Les principales règles de syntaxe et les fonctionnalités du logiciel SPSS qui se rattachent à chacun des 13 chapitres du volume sont expliquées et exemplifiées dans cette partie du site Internet. Les professeurs qui ne prévoient pas de sessions de pratique pour les étudiants pourront simplement ignorer cet ajout au site Internet.

Ainsi, en lisant chacun des chapitres de cet ouvrage, en faisant les quiz rapides, en répondant aux questions de fin de chapitre et en exécutant les exercices du site Internet, le lecteur pourra mettre immédiatement en pratique ses connaissances.

CHAPITRE 1

LA DESCRIPTION DES DONNÉES

La description et l'inférence en statistique	15
L'organisation d'une banque de données pour l'analyse statistique	17
Les variables.....	19
Les variables indépendantes et les variables dépendantes.....	19
Les échelles de mesure	20
Les variables (échelles) nominales.....	20
Les variables (échelles) ordinales.....	22
Les variables (échelles) à intervalles.....	23
Les variables (échelles) de rapport.....	25
Les relations entre les diverses échelles de mesure.....	26
Sommaire du chapitre.....	27
Exercices de compréhension	27

Page laissée blanche

CHAPITRE 1

LA DESCRIPTION DES DONNÉES

Les statistiques sont un inventaire de techniques et de procédures qui permettent d'organiser et de faire le sommaire d'une masse d'informations afin d'en dégager des conclusions utiles à la compréhension d'un phénomène.

LA DESCRIPTION ET L'INFÉRENCE EN STATISTIQUE

Les statistiques se divisent en deux branches, complémentaires et inter-reliées: celles qui permettent une description des informations; et celles qui permettent, à partir de ces descriptions, de faire des inférences. Les *statistiques descriptives* font le sommaire et simplifient l'information dans le but de la clarifier et de révéler ses tendances lourdes. L'*inférence statistique* est une série de procédures qui se servent de ces descriptions pour tirer des conclusions plus générales sur le phénomène à l'étude.

Tous les phénomènes mesurés peuvent être analysés statistiquement, à condition que l'information soit exprimée numériquement. C'est donc dire que les statistiques ne doivent utiliser que des informations quantitatives.

Données quantitatives

Nous avons tous l'habitude de mesurer les choses en nous servant de chiffres. Notre âge, notre poids, le montant de nos dettes, le nombre d'enfants dans notre famille ou notre température corporelle peuvent tous être mesurés quantitativement. Les attitudes, les opinions, les croyances, la personnalité et les comportements peuvent aussi être mesurés quantitativement. En principe, on se sert d'un questionnaire comme instrument de mesure. Chaque réponse possible à une question est décrite par un chiffre. Par exemple, dans un questionnaire qui mesure la satisfaction au travail, on pourrait demander aux employés d'une compagnie d'indiquer leur degré d'accord ou de désaccord avec des phrases telles que : « Aller au travail m'est très désagréable », « Si je le pouvais, je donnerais ma démission aujourd'hui », etc. Les réponses possibles sont : « Totalelement en accord » (indexé par le chiffre 1), « Plutôt en accord » (chiffre 2), « Ni en accord ni en désaccord » (chiffre 3), « Plutôt en désaccord » (chiffre 4), et « Totalelement en désaccord » (chiffre 5). Ainsi, les personnes ayant plus de satisfaction au travail auraient tendance à être en désaccord avec ces énoncés et, par conséquent, fourniraient des réponses plus près de 5, alors que les réponses des personnes ayant peu de satisfaction au travail seraient concentrées autour de 1. Ces procédures permettent de « quantifier » les attitudes, de les exprimer numériquement. Ce faisant, il devient possible d'en faire une analyse statistique.

Les statistiques que nous allons étudier dans ce livre sont mises à profit pour faciliter la compréhension de phénomènes aussi diversifiés que la croissance économique d'une société, les comportements sociaux, l'efficacité d'une technique chirurgicale, ou même les réactions chimiques. Les cognitivistes utilisent les statistiques pour déduire l'organisation du cerveau et ses liens avec la pensée. Les psychologues font appel aux statistiques afin de mieux comprendre les caractéristiques individuelles comme la personnalité, l'intelligence ou le comportement déviant à l'école. Les sociologues s'en servent pour mieux comprendre la violence sociale ou la relation entre les idéologies et l'éducation. Les experts en marketing y recourent afin d'analyser et d'améliorer les stratégies de mise en marché. Dans le monde des affaires, on s'en sert pour planifier les inventaires ou pour établir les marges de profit. Et ce sont les statistiques qui déterminent, en grande partie, les décisions des gouvernements.

Nous lisons tous les jours dans les journaux des résultats de sondages. Ces sondages guident, dans une certaine mesure, les décisions concernant les activités des institutions, publiques ou privées, l'impact de ces activités se répercutant sur presque chacun de nous : étudiants, consommateurs, travailleurs. Tous, nous sommes personnellement affectés par les statistiques

et un grand nombre de décisions qui nous touchent trouvent leur origine dans le résultat d'une analyse statistique.

En analyse statistique, les informations (quantitatives) sont recueillies, organisées et soumises à des procédures arithmétiques. Le résultat final de ces procédés est une simplification de l'information qui permet de dégager des tendances afin de mieux comprendre le phénomène étudié et d'en tirer des conclusions utiles. Les statistiques nous permettent de voir la forêt malgré les arbres!

Les analyses statistiques offrent la possibilité de mieux comprendre les caractéristiques des individus (l'intelligence, la sociabilité), des groupes (la performance des équipes ou la compétitivité des entreprises), ou des communautés plus larges (le degré de pauvreté dans différents pays, le coût des logements dans différentes villes). La source des données (les individus, les équipes, les entreprises, les villes, etc.) se nomme le *sujet d'analyse* ou l'*unité d'analyse, ou encore l'observation*. Le sujet d'analyse définit donc l'origine de l'information. Les conclusions, par conséquent, s'appliqueront exclusivement à cette source. Ainsi, lorsque nous mesurons la densité des populations dans les villes, le sujet d'analyse est la ville et les conclusions s'appliquent aux villes. Lorsque les informations sont recueillies auprès des individus, le sujet d'analyse est l'individu. Si nous mesurons le comportement des chiens, le sujet d'analyse est le chien. Chaque sujet d'analyse fournit une ou des observations. Ainsi, lorsque nous analysons l'intention de vote de 1 000 citoyens, nous avons 1 000 observations.

L'ORGANISATION D'UNE BANQUE DE DONNÉES POUR L'ANALYSE STATISTIQUE¹

Les informations fournies par les sujets sont généralement organisées sous la forme d'un tableau comprenant des colonnes et des lignes (rangées). Chaque sujet d'analyse (chaque répondant à un sondage par exemple) occupe une ligne du tableau. Les variables (chaque question du sondage) occupent les colonnes. À l'intersection de chaque colonne et de chaque

1. Le site Internet du livre (www.pum.umontreal.ca/ca/fiches/978-2-7606-2113-8.html) contient plusieurs banques de données, y compris celle portant sur les salaires des joueurs de hockey. Un extrait de ce dossier est reproduit au Tableau 1.1.

rangée, nous trouvons une cellule. C'est dans cette cellule que sera placée la valeur numérique qui sera analysée.

Chaque colonne contient une seule variable. Si nous demandons à 10 personnes de fournir une réponse à trois questions, la banque de données comprendra 10 lignes et 3 colonnes. En lisant horizontalement, sur une ligne, nous observons la valeur obtenue par un sujet d'analyse pour chaque variable. À l'inverse, avec une lecture verticale, le long d'une colonne, nous obtenons la réponse de tous les sujets sur une variable. Par exemple, le Tableau 1.1 présente une banque de données sur les salaires annuels des joueurs de hockey. Chaque ligne du tableau identifie un joueur de hockey spécifique. Chaque colonne définit une variable différente. À l'intersection de chaque rangée et de chaque colonne, nous trouvons une cellule qui indique la valeur de la variable X pour le joueur Y. Ainsi, en lisant les colonnes consacrées à l'athlète décrit dans la rangée 1, nous voyons son nom (variable « joueur »), son équipe (variable « équipe »), la position qu'il occupe au jeu (variable « position ») et son salaire annuel (variable « salaire »).

Tableau 1.1			
Banque de données organisées pour une analyse statistique			
<i>Joueurs</i>	<i>Équipes</i>	<i>Position</i>	<i>Salaire (\$US)</i>
Joueur 1	Flames de Calgary	G	400 000
Joueur 2	Canadiens de Montréal	C	3 300 000
Joueur 3	Predators de Nashville	G	900 000
Joueur 4	Sénateurs d'Ottawa	G	1 500 000
Joueur 5	Canucks de Vancouver	G	425 000
Joueur 6	Avalanches du Colorado	C	11 000 000
Joueur 7	Blues de Saint-Louis	G	165 000

Il est relativement facile d'organiser des banques de données en se servant de logiciels tels que Word ou Excel, ou de programmes d'analyse statistique spécialisés tels que SPSS ou SAS. L'important est de se souvenir de cette règle: *chaque rangée définit un sujet différent et chaque colonne, une variable diffé-*

rente. Une fois la banque de données construite, en se servant d'un logiciel de traitement de texte (comme le Bloc-notes) ou d'un chiffrier (tel qu'Excel), il est facile de l'importer dans un logiciel d'analyse statistique (tel que SPSS).

LES VARIABLES

Une variable est une caractéristique que l'on mesure et qui sera soumise à des analyses. On l'appelle *variable* parce que les sujets d'analyse peuvent lui attribuer des valeurs différentes. L'âge, le sexe, le quotient intellectuel (QI) et la condition sociale, entre autres, sont des variables. Le QI est une variable parce qu'il peut être différent selon les personnes. L'anxiété est une variable puisque certains peuvent être très anxieux, d'autres très calmes et d'autres encore peuvent se situer quelque part entre ces deux extrêmes. Le genre – homme versus femme – est lui aussi une variable.

Lorsque toutes ces valeurs sont identiques, la variable prend le nom de *constante*. Si la variable ne peut prendre qu'une valeur unique, cette variable devient une constante. Par exemple, lorsque nous mesurons le degré de sociabilité des femmes, le sexe, qui est habituellement une variable, devient une constante (toutes les personnes mesurées étant des femmes).

Quiz rapide 1.1

Le site Internet du livre contient une banque de données sur les joueurs de la Ligue nationale de hockey (les sujets de cette analyse). Prenez seulement la partie des données qui porte sur les Canucks de Vancouver et identifiez les variables et les constantes. Si vous n'utilisez pas le site WEB, répondez au quiz en vous référant au Tableau 1.1.

Les variables indépendantes et les variables dépendantes

Les chercheurs ou intervenants qui font de l'expérimentation distinguent les *variables indépendantes* des *variables dépendantes*. L'expérimentateur contrôle ou choisit la façon dont la variable indépendante varie afin de déterminer le degré d'impact que cette variable indépendante a sur la variable dépendante. La variable dépendante est la réaction du sujet d'analyse à la variable indépendante. Par exemple, une professeure cherche à déterminer si les étudiants réussissent mieux leurs examens lorsqu'elle y convertit

Page laissée blanche

Ce livre a été imprimé au Québec en janvier 2010 sur du papier
entièrement recyclé sur les presses de Marquis imprimeur.