Serge Alalouf

METHODES STATISTIQUES

TABLE DES MATIÈRES

PRÉFACE

À l'étudiant

Ce livre a été conçu — et utilisé, sous forme de notes de cours — depuis plusieurs années à l'École des sciences de la gestion pour un cours d'introduction à la statistique. Il est donc résolument appliqué et, dans un sens plutôt large, appliqué *à la gestion*. Je reconnais que ce mandat peut signifier plusieurs choses, et qu'il se réalise différemment selon le sens qu'on lui donne. Pour commencer, y a-t-il vraiment des applications *propres à la gestion* ? La gestion est une activité multidisciplinaire qui touche à plusieurs domaines ; toute technique statistique est susceptible de se révéler un jour utile en gestion. Son domaine n'étant pas facilement cerné, la « statistique pour gestionnaires » ne peut pas être définie par un ensemble de techniques. Elle se caractérise plutôt par un point de vue général qui met l'accent sur les concepts qui unifient les différentes méthodes et ne s'attarde pas sur les techniques pointues qui ne concernent que quelques disciplines spécialisées. Le livre comprend donc des exemples et des problèmes qui, à l'image de la gestion elle-même, relèvent d'une variété de disciplines.

Je suis parfaitement conscient du fait que je m'adresse à des lecteurs pour qui la statistique n'est qu'une nécessité, peut-être même un mal nécessaire. Je me concentre donc sur les applications plutôt que sur la théorie. Ce que cela veut dire, c'est que je n'aborde aucun concept qui n'aboutit pas — directement ou indirectement — à une application utile, une formule. Ce qui ne veut pas dire, cependant, que je me borne à montrer comment appliquer des formules. Il y en a, des formules, et bien sûr il faut savoir les appliquer. Mais une formule ne s'applique que dans un contexte donné, un contexte qui doit être clairement défini et qui ne peut l'être que dans un langage théorique. C'est pour cela que la théorie joue malgré tout un rôle important dans une approche néanmoins appliquée.

Heureusement, c'est une théorie abordable : elle n'exige pas beaucoup de mathématiques — une maîtrise de l'algèbre élémentaire, la capacité de manipuler des inégalités et des valeurs absolues, une connaissance rudimentaire des fonctions exponentielles et (rarement) des logarithmes. Elle comprend, cependant, quelques concepts nouveaux, des idées fondamentales qui ne vous seront pas familières et qui exigeront un certain effort d'abstraction.

Mais attention : la nécessité d'un tel effort n'est pas immédiatement évidente. Cela donne une fausse impression de facilité. Voici ce qui peut se passer. Vous avez un problème à résoudre; vous trouvez un exemple semblable traité dans le chapitre ; vous suivez la même démarche ; vous obtenez la bonne réponse. Une réussite ? Cela dépend. Au-delà de la démarche, il y a la théorie qui la justifie. Sans un effort conscient de votre part, la justification risque de vous échapper. Votre but devrait être de développer la capacité de passer d'un concept théorique à la solution d'un problème et non de passer d'un problème résolu à la solution d'un problème semblable. C'est cette habileté qui vous facilitera les choses plus tard, et c'est celle que ce livre s'emploie à développer.

À l'enseignant

L'origine de ce livre remonte à plusieurs années, alors qu'un comité constitué de membres du Département de mathématiques de l'UQÀM et de l'École des sciences de la gestion recevait pour mandat de définir un cours apte à convenir à la majorité des étudiants de gestion. Il s'agissait d'un cours de 45 heures, pour certains le seul cours de statistique au baccalauréat. L'approche naturelle aurait été de se limiter à un nombre restreint (mais quand même important) de techniques et de s'en tenir essentiellement aux applications de formules. Si cette approche est réalisable dans certains domaines — là où l'ensemble des techniques statistiques est relativement circonscrit — , elle se révèle quasiment impossible en gestion, un domaine dont le champ d'application est vaste et les techniques statistiques diverses et nombreuses. Tout choix limité se révèlerait inadéquat.

J'y ai donc renoncé, et privilégié plutôt une approche qui consiste à limiter le nombre de techniques pour faire place aux concepts fondamentaux — variable aléatoire, loi de probabilité, espérance mathématique, variance, distribution d'échantillonnage — sur lesquelles reposent les techniques.

On ne peut pas éluder la théorie sous prétexte que les étudiants en gestion n'en ont pas besoin. Je crois qu'au contraire, la théorie, même dans un cours appliqué, a sa place pour plusieurs raisons :

- Une bonne compréhension de la théorie est nécessaire pour assurer un minimum de confiance et d'autonomie, pour reconnaître ce que des contextes disparates ont en commun, pour savoir quelle technique appliquer dans quelle situation. En ce sens, un cours plus théorique est également plus « pratique ».
- Un contenu théorique encourage une réflexion sur ce qu'est, fondamentalement, la statistique, soit la science du savoir empirique. L'information, qu'elle soit le fruit d'une expérience scientifique, le résultat d'un sondage, ou une simple accumulation de données opérationnelles, ne peut être interprétée qu'à la lumière des méthodes qui l'ont générée. Une approche étroitement utilitaire ne permet pas de développer cette essentielle sensibilité à l'incertitude qui accompagne l'acquisition des connaissances.
- L'intérêt culturel de ces réflexions justifierait l'existence du cours, même aux yeux de ceux — probablement assez nombreux — qui n'auront jamais l'occasion de s'en servir de façon pratique. Il ajoute une composante académique qui pourrait être à sa place dans tout programme universitaire.

Cette théorie, est-il possible de la comprendre sans un bagage mathématique important ? Oui, selon mon expérience et celle de mes collègues. On peut dépouiller la statistique de tout son contenu mathématique (ce qu'on fait presque, dans ce livre) sans pour autant la réduire à des recettes. Il reste d'importants concepts qui peuvent fort bien s'exprimer en langage ordinaire avec l'aide de quelques notions élémentaires de mathématiques. Et s'ils sont parfois difficiles à saisir, c'est généralement leur sens concret — et non mathématique — qui pose problème. Par exemple, la définition *mathématique* d'« indépendance » ne présente aucune difficulté. Mais son sens intuitif et concret est loin d'être banal. C'est pourtant celui-là qu'il est nécessaire de comprendre. De même pour les notions d'espérance, de variance, de loi de probabilité : c'est sur le plan des applications qu'elles se révèlent subtiles et problématiques.

Ces notions sont nouvelles et parfois abstraites. D'où l'importance de les illustrer par des applications concrètes. Les exemples et problèmes ont donc été choisis avec un double objectif :

- Montrer à quoi sert la statistique. Car malgré ce biais favorable à une composante théorique importante, il fallait bien sûr garder l'œil sur l'utilité immédiate potentielle du contenu. Il serait d'ailleurs inadmissible de présenter la théorie statistique sans montrer clairement ce qu'on peut en faire.
- Concrétiser les notions théoriques.

(Il arrive, parfois, que ces deux objectifs ne puissent pas être atteints en même temps. Si un problème présenté comme une « application » pratique semble tiré par les cheveux, c'est qu'il est particulièrement apte à concrétiser une idée.)

Ces deux objectifs sont poursuivis inlassablement tout au long de cet ouvrage. Ils expliquent le choix des problèmes. Ils expliquent aussi le choix de sujets, entre autres l'ampleur de l'espace réservé aux techniques d'échantillonnage.

Échantillonnage

L'échantillonnage répond aux *deux* objectifs. Tout d'abord, c'est l'une des applications les plus répandues, et si l'espace qui lui est réservé dans ce livre semble démesuré, il n'est pas disproportionné par rapport à ce qui se fait dans l'entreprise. Je ne prétends pas pouvoir chiffrer le « pourcentage » de l'activité statistique consacré à l'échantillonnage, mais si l'on se fie aux demandes d'aide adressées aux statisticiens, il est clair qu'il s'en fait beaucoup. (À l'instar des vérificateurs, pour qui l'échantillonnage est désormais un outil incontournable, les gestionnaires semblent de plus en plus disposés à se fier à des données d'échantillonnage pour informer leurs décisions.)

Ensuite, l'échantillonnage est une problématique qui permet d'illustrer les concepts théoriques de la façon la plus concrète qui soit. Le paramètre à estimer est une moyenne ordinaire avant d'être une espérance mathématique ; ou une proportion, avant d'être une probabilité de succès. La notion de distribution d'échantillonnage — difficile de prime abord — est plus claire et concrète dans un contexte d'échantillonnage.

Tests d'hypothèses

Les tests d'hypothèses occupent une place importante dans ce livre, bien plus que ne laisserait croire le fait qu'un seul chapitre porte ce nom. C'est que le sujet est traité en filigrane tout au long du livre. Pourquoi cette façon de faire ? D'abord, parce qu'il me semble évident que les étudiants — la plupart d'entre eux, du moins — auxquels le livre s'adresse n'effectueront pas de tests dans le sens formel. Ce qui importe, c'est qu'ils en saisissent les idées fondamentales : Pourquoi H_o est telle hypothèse et non son contraire ? Qu'est-ce qu'on conclut quand on rejette H_o et quand on ne rejette pas H_o ? Quel est le sens de la valeur p ? Quelles sont les probabilités d'erreur ? Ces idées qui se bousculent doivent être présentées graduellement. C'est pour cela que je commence à en parler tôt, sans formalisme, et surtout en misant sur un acquis important et souvent négligé, à savoir, le fait que tout le monde comprend l'idée fondamentale au cœur d'un test d'hypothèse : On rejette une hypothèse lorsque les données observées s'écartent trop de celles attendues sous l'hypothèse. Une idée très

familière, un raisonnement utilisé quotidiennement, qui comprend également l'idée d'attacher une probabilité à l'écart observé. Il est bon d'exploiter cet acquis avant qu'il ne soit noyé par des notions nouvelles de région critique, de types d'erreur, de niveau de test, etc. Je veux permettre à l'étudiant de consolider et verbaliser ces idées progressivement, les approcher à petits pas, dans le cadre d'exemples et d'exercices de probabilités, sans jargon. C'est pour cela qu'on en parle dès les premiers chapitres, comme de simples applications des probabilités, immédiatement accessibles. L'argument principal est dominé par le gros bon sens ; seul le calcul des probabilités fait appel à des notions mathématiques. Certes, cette approche a un prix : pas de loi de Student, pas de test d'égalité de moyennes, pas de test sur une variance ou sur l'égalité de deux variances, etc. Ces questions ne manquent pas d'intérêt, mais le manque de temps impose des sacrifices. Je me suis laissé guider par un seul principe : chaque idée n'est traitée que dans ses grandes lignes. La *notion* de test d'hypothèse, amplement illustrée par les tests sur une moyenne et une proportion, importe plus que ses diverses applications. Même chose pour les intervalles de confiance : on se limite à de grands échantillons et un niveau de confiance à 95 %. Mes collègues et moi avons constaté que plusieurs raffinements (petits échantillons, niveaux de confiance variables, etc.) détournent l'attention des étudiants au détriment des idées fondamentales.

Remerciements

Cet ouvrage a bénéficié de contributions des nombreux collègues qui m'ont fait profiter de l'expérience qu'ils ont vécue au fil des ans en enseignant le cours MAT2080 de l'UQÀM. Ils m'ont bien sûr signalé les erreurs et inévitables coquilles qui se produisaient à chaque nouvelle édition. Leurs commentaires m'ont permis d'ajuster le contenu et la forme des notes de cours qui ont abouti à ce livre. Ils m'ont indiqué ce qui passe et ce qui ne passe pas, ce qui est trop théorique, ce qui n'est pas suffisamment expliqué.

Je leur en suis à tous reconnaissant. Je prendrai le risque de les nommer (et je m'excuse d'ores et déjà auprès de ceux que j'aurais malencontreusement oubliés) : Marie-Claude Audet, Michel Bitton, Roger Brousseau, Mourad Dahhou, Alain Desgagné, Jean-Pierre Dion, René Ferland, Simon Guillotte, Ricardo Herrera, Lotfi Khribi, Richard Labonté, Denis Laferrière, Geneviève Lefebvre, Yves Léger, Mohamed Nassim, Alexandre Pilotte, Pascale Rousseau, Glenn Shorrock, François Watier. Je signale en particulier Roger Brousseau pour avoir persisté jusqu'aux toutes dernières semaines. Ils ont tous contribué à aplanir les maladresses.

Finalement, à Michel Adès et à Hassan Younes, mes proches collaborateurs de longue date, je tiens à exprimer ma profonde gratitude pour l'indéfectible appui, moral et matériel, qu'ils m'ont donné au fil des années. Michel, toujours prêt à collaborer, rédigeait des solutions d'examens et réussissait, avec son œil de lynx, à dénicher des coquilles même après moult vérifications. Hassan a pendant des années collaboré avec moi à la coordination du cours, sa part grandissant et la mienne diminuant avec le temps, comme des vases communicants. Michel et Hassan ont tous deux pris part à l'évolution du cours depuis sa première mouture jusqu'à celle-ci. Ils ont défendu sa philosophie et réussi à convaincre les sceptiques. J'apprécie leur aide et encore plus leur amitié.