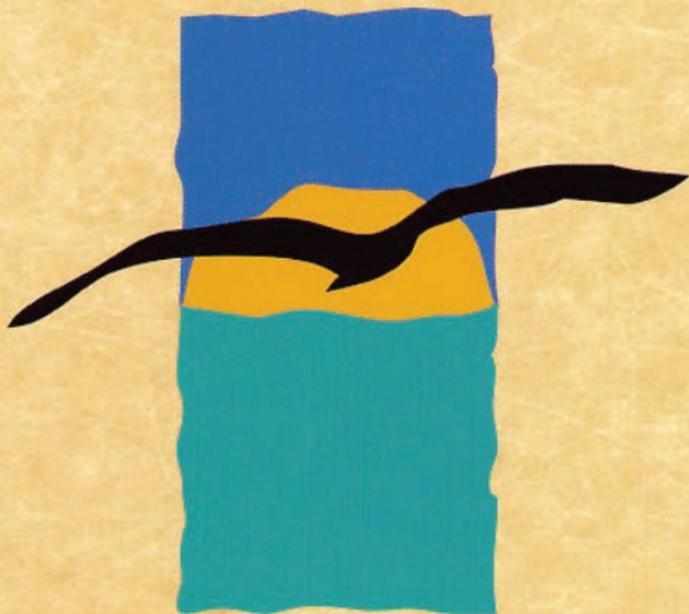


LA RECHERCHE TOURISTIQUE

INTRODUCTION À LA RECHERCHE QUANTITATIVE
PAR QUESTIONNAIRE

JEAN STAFFORD



Presses de l'Université du Québec

LA RECHERCHE TOURISTIQUE

INTRODUCTION À LA RECHERCHE QUANTITATIVE PAR QUESTIONNAIRE

© 1998 – Presses de l'Université du Québec

Édifice Le Delta I, 2875, boul. Laurier, bureau 450, Québec, Québec G1V 2M2 • Tél. : (418) 657-4399 – www.puq.ca

Tiré de : *La recherche touristique*, Jean Stafford, ISBN 2-7605-0909-5 • DA909N

Tous droits de reproduction, de traduction et d'adaptation réservés

DANS LA MÊME COLLECTION

Microéconomie du tourisme

Jean Stafford

1996, 146 pages

L'expérience touristique contemporaine

Fondements sociaux et culturels

Marc Laplante

1996, 202 pages

L'espace touristique

Sous la direction de Normand Cazalais,

Roger Nadeau et Gérard Beaudet

1999, 314 pages

PRESSES DE L'UNIVERSITÉ DU QUÉBEC

2875, boul. Laurier, Sainte-Foy (Québec) G1V 2M3

Téléphone : (418) 657-4399 • Télécopieur : (418) 657-2096

Courriel : secretariat@puq.quebec.ca

Catalogue sur Internet : <http://www.quebec.ca/puq>

Distribution :

CANADA et autres pays

DISTRIBUTION DE LIVRES UNIVERS S.E.N.C.

845, rue Marie-Victorin, Saint-Nicolas (Québec) G7A 3S8

Téléphone : (418) 831-7474 / 1-800-859-7474

Télécopieur : (418) 831-4021

FRANCE

LIBRAIRIE DU QUÉBEC A PARIS

30, rue Gay-Lussac, 75005 Paris, France

Téléphone : 33 1 43 54 49 02

Télécopieur : 33 1 43 54 39 15

SUISSE

GM DIFFUSION SA

Rue d'Etraz 2, CH-1027 Lonay, Suisse

Téléphone : 021 803 26 26

Télécopieur : 021 803 26 29



La *Loi sur le droit d'auteur* interdit la reproduction des œuvres sans autorisation des titulaires de droits. Or, la photocopie non autorisée – le «photocopillage» – s'est généralisée, provoquant une baisse des ventes de livres et compromettant la rédaction et la production de nouveaux ouvrages par des professionnels.

L'objet du logo apparaissant ci-contre est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit le développement massif du «photocopillage».

© 1998 – Presses de l'Université du Québec

Édifice Le Delta I, 2875, boul. Laurier, bureau 450, Québec, Québec G1V 2M2 • Tél. : (418) 657-4399 – www.puq.ca

Tiré de : *La recherche touristique*, Jean Stafford, ISBN 2-7605-0909-5 • DA909N

Tous droits de reproduction, de traduction et d'adaptation réservés

LA RECHERCHE TOURISTIQUE

INTRODUCTION À LA RECHERCHE QUANTITATIVE PAR QUESTIONNAIRE

JEAN STAFFORD

1999



Presses de l'Université du Québec
2875, boul. Laurier, Sainte-Foy (Québec) G1V 2M3

© 1998 – Presses de l'Université du Québec

Édifice Le Delta I, 2875, boul. Laurier, bureau 450, Québec, Québec G1V 2M2 • Tél. : (418) 657-4399 – www.puq.ca

Tiré de : *La recherche touristique*, Jean Stafford, ISBN 2-7605-0909-5 • DA909N

Tous droits de reproduction, de traduction et d'adaptation réservés

Données de catalogage avant publication (Canada)

Stafford, Jean

La recherche touristique : Introduction à la recherche quantitative par questionnaire

Comprend des réf. bibliogr.

ISBN 2-7605-0909-5

1. Tourisme – Recherche – Méthodologie. 2. Recherche – Méthodologie. 1.
Titre.

G155.A1S7221 996

338.4'791'072

C96-940894-3

Les Presses de l'Université du Québec remercient le Conseil des arts du Canada et le Programme d'aide au développement de l'industrie de l'édition du Patrimoine canadien pour l'aide accordée à leur programme de publication.

Révision linguistique : GISLAINE BARRETTE

Mise en pages : TYPO LITHO COMPOSITION INC.

Conception graphique de la couverture : CARON & GOSSELIN

1 2 3 4 5 6 7 8 9 PUQ 1999 9 8 7 6 5 4 3 2 1

Tous droits de reproduction, de traduction et d'adaptation réservés

© 1996 Presses de l'Université du Québec

Dépôt légal – 3^e trimestre 1996

Bibliothèque nationale du Québec / Bibliothèque nationale du Canada

Imprimé au Canada

© 1998 – Presses de l'Université du Québec

Édifice Le Delta I, 2875, boul. Laurier, bureau 450, Québec, Québec G1V 2M2 • Tél. : (418) 657-4399 – www.puq.ca

Tiré de : *La recherche touristique*, Jean Stafford, ISBN 2-7605-0909-5 • DA909N

Tous droits de reproduction, de traduction et d'adaptation réservés

**Ce petit livre est dédié à tous mes étudiants,
anciens, présents et futurs.**

Je sais que je ne sais pas ce que je ne sais pas ; j'envie ceux qui sauront davantage mais je sais qu'ils auront tout comme moi à mesurer, peser, déduire et se méfier des déductions produites, faire dans le faux la part du vrai et tenir compte dans le vrai de l'éternelle admixtion du faux.

Marguerite Yourcenar
L'œuvre au noir,
Paris, Gallimard, 1991, p. 159.

Remerciements

Je tiens à remercier monsieur Marcel Samson qui a participé à une première ébauche de ce livre dans les *Fascicules Téoros*.

Mille mercis à mesdames Line Lapierre et Carolle Racette pour la typographie.

Table des matières

Introduction	1
I. La recherche quantitative en tourisme	3
1. La science et la recherche en tourisme	5
2. Recherche en tourisme et industrie touristique	20
3. Les fonctions de la recherche en tourisme	24
4. Les problèmes particuliers de la recherche en tourisme	27
II. Le problème de départ de la recherche	33
1. L'analyse de la demande de recherche	35
2. La préenquête	38
3. Les objectifs et les hypothèses de départ	42
4. La planification de la recherche : les moyens nécessaires et les coûts	45
III. Les outils de la recherche et la collecte des données	51
1. Des objectifs au questionnaire	53
2. Les modes d'administration des questionnaires et la rédaction des questions	57
3. La codification des questions et le prétest	67
4. L'échantillonnage et la collecte des données	73
IV. Le traitement des données	91
1. La préparation des données	93
2. L'analyse univariée	100
3. L'analyse bivariée	105
4. L'analyse multivariée	119

V. Les résultats de la recherche	131
1. Présentation et interprétation des tableaux univariés	133
2. Présentation et interprétation des tableaux bivariés	136
3. Du «brouillon» au rapport de recherche	139
4. État actuel de la recherche en tourisme	144
Conclusion	153
Tables statistiques	155
Bibliographie	159

Liste des tableaux

Tableau 1	Niveau d'information et complexité du projet	21
Tableau 2	Types de recherche et critères d'évaluation d'une recherche	36
Tableau 3	Processus de sélection des objectifs et des hypothèses de la recherche	43
Tableau 4	Étapes et échéancier de la recherche	46
Tableau 5	Budget/coûts de la recherche	47
Tableau 6	Premières étapes d'une recherche quantitative	49
Tableau 7	Degré de satisfaction et de fidélité	53
Tableau 8	Mesure de la fréquentation d'un restaurant	56
Tableau 9	Différents niveaux de problème selon les modes d'administration de questionnaire	59
Tableau 10	Contraintes et modes d'administration de questionnaires	60
Tableau 11	Exemple de questionnaire de satisfaction	64
Tableau 12	Tableau pour déterminer la taille de l'échantillon	87
Tableau 13	Rubriques pour le calcul du taux de réponse selon le mode d'administration du questionnaire	96

Tableau 14	Redressement d'échantillon avec la technique «mécanique»	98
Tableau 15	Destination de vacances des Québécois en 1995	101
Tableau 16	Nombre de semaines de vacances au cours des six derniers mois	102
Tableau 17	Nombre de semaines de vacances au cours des six derniers mois, en pourcentage	103
Tableau 18	Le score de satisfaction	104
Tableau 19	Forme générale des tableaux croisés	105
Tableau 20	Exemple d'un tableau croisé ; la variable scolarité avec la variable vacances dans les trois dernières années	106
Tableau 21	Distribution conjointe des données (lecture/explication du tableau 20)	108
Tableau 22	Distributions marginales de la variable «années de scolarité»	109
Tableau 23	Les distributions marginales de la variable «vacances dans les trois dernières années»	109
Tableau 24	Calcul du khi carré à partir des données absolues du tableau 20	112
Tableau 25	Table abrégée de la distribution du khi carré (loi de K. Pearson)	113
Tableau 26	Analyse de variance, le nombre de semaines de vacances et le groupe d'âge du répondant	118
Tableau 27	Structure du modèle dans l'analyse factorielle	120
Tableau 28	Analyse factorielle de la satisfaction à l'Auberge du Lac Carré (matrice factorielle après rotation varimax)	122
Tableau 29	Répondants qui ont pris des vacances au cours des trois dernières années	133
Tableau 30	Importance de certains éléments dans le choix d'une destination de vacances et les moyennes pour chacune des variables	134

Tableau 31	Ordre d'importance de certains éléments dans le choix d'une destination de vacances ; classement obtenu à partir des moyennes des réponses	135
Tableau 32	Années de scolarité et la prise de vacances au cours des trois dernières années, en pourcentage	137
Tableau 33	Relations entre les destinations de vacances dans la prochaine année et certaines variables indépendantes, en pourcentage	138
Tableau 34	La rentabilité d'une recherche	148

Liste des figures

Figure 1	Évolution des écoles de pensée dans les sciences	16
Figure 2	Le système d'information touristique	24
Figure 3	Schéma de décision au regard de la nécessité de faire une enquête	37
Figure 4	Variables indépendantes et variable dépendante	61

Introduction

La recherche empirique en tourisme n'a pas encore conquis ses lettres de noblesse dans la confrérie sélecte des sciences ; la téorologie (ou science du tourisme) avance masquée, soit par des idéologies brumeuses percevant le tourisme comme «l'avenir de l'humanité» ou l'ultime «sauveteur» des régions en péril ou d'une économie moribonde, soit par un professionnalisme étroit pour lequel le touriste est seulement un «client» qu'on doit «plumer» le plus possible !

La plupart du temps, les connaissances touristiques se résument aux récits des «grands explorateurs» et aux articles des pages touristiques des grands journaux : mélange d'exotisme de pacotille et de conseils pratiques pour la survie du voyageur. Dans l'ensemble de ces discours, ce qui domine ce sont les descriptions où se mêlent les remarques générales (les clichés touristiques), les observations minutieuses (l'enflure du détail) et les constatations faites de prénotions et de préjugés.

Face à ce magma informe, à cette masse d'informations, la plupart du temps contradictoires, la téorologie doit opérer un tri, distinguer les niveaux de savoir et formuler les règles d'une connaissance scientifique du tourisme. Dans cette optique, la recherche empirique en tourisme a pour tâche de construire consciemment des «faits scientifiques».

Construire des «faits scientifiques», c'est partir d'une situation, d'un problème quelconque en tourisme et d'en faire «la théorie» à l'aide d'hypothèses, de concepts et d'indicateurs fortement *articulés*. Le caractère scientifique d'une théorie se manifeste par son caractère réfutable ; la

2 *La recherche touristique*

théorie doit pouvoir être mise à l'épreuve (selon la pensée poppérienne). Cette épreuve, ce sera l'enquête elle-même, effectuée sur le terrain. Ainsi, la boucle est bouclée : la connaissance s'établit à partir de problèmes, face auxquels des conjectures sont émises (celles-ci doivent être réfutables) et des tests permettent de confirmer ou d'infirmer totalement ou partiellement ces hypothèses.

La recherche empirique en tourisme a un rôle primordial à jouer dans l'établissement de connaissances scientifiques dans ce domaine ; les praticiens en tourisme doivent disposer d'un corps de connaissances rigoureux sur lequel fonder leurs expertises et leurs pratiques de gestion, de planification et d'intervention. La recherche empirique en tourisme est une tâche aride, humble et souvent ingrate, loin des colloques qui ne servent qu'aux relations publiques avec leurs discours bidons. Ce petit livre vise à faire connaître, étape par étape, la démarche qui servira à construire ces « faits scientifiques » et à donner des assises solides à une théorie naissante.

Chapitre 1

La recherche quantitative en tourisme

Ce chapitre traite de la recherche quantitative en tourisme, qui est la base même d'une connaissance scientifique dans ce domaine. Sans cette connaissance, le tourisme demeurera une activité économique de troisième ordre évoluant au gré des modes et des fluctuations économiques. La recherche empirique en sciences sociales et en gestion, bien qu'elle soit relativement récente, recèle d'énormes possibilités de développement. Son principal objectif est d'étudier et d'analyser les besoins des touristes de façon à éclairer les gouvernements et

les entreprises touristiques dans les décisions qu'ils devront prendre au regard des orientations, des produits et des services nécessaires dans ce domaine. Les particularités du tourisme soulèvent des problématiques nouvelles et requièrent l'adaptation des méthodes et des techniques de recherche usuelles.

1 LA SCIENCE ET LA RECHERCHE EN TOURISME

Dans cette section, nous allons aborder les fondements mêmes de la recherche quantitative en tourisme ; nous allons essayer de saisir les dimensions philosophiques et pratiques propres à ce type de recherche. Nous traiterons donc des éléments suivants :

- les raisons motivant l'étude scientifique des phénomènes touristiques ;
- la naissance de la science moderne ;
- l'émergence des sciences appliquées sur les plans économique et social ;
- les principaux thèmes d'étude.

POURQUOI LA RECHERCHE SCIENTIFIQUE EN TOURISME ?

La plupart des industries, du secteur primaire au secteur tertiaire, se sont développées grâce aux innovations scientifiques ; les technologies dérivées des sciences pures ou des sciences sociales ont servi de levier à la croissance de ces industries. Certains secteurs touristiques sont fortement dépendants de la technologie, par exemple, les transports et, à un degré moindre, la restauration.

La principale raison qui milite en faveur d'une connaissance scientifique du tourisme est la dure nécessité, dans ce domaine comme ailleurs, de faire des choix. Dans le traitement des problèmes, la démarche scientifique valorise une approche rationnelle plutôt qu'intuitive. Cette approche rationnelle insiste sur l'obligation de connaître avant d'agir, sur la primauté de la connaissance sur l'action.

Le principal intérêt de la démarche scientifique en tourisme réside dans le fait qu'elle permet d'introduire une ligne de démarcation entre un discours subjectif, souvent tissé de prénotions et de préjugés, et un discours que l'on veut objectif et dont les conclusions reposent sur une analyse froide des faits recueillis ! À la limite, si tous les discours se valent, on peut affirmer et proposer n'importe quoi ! La démarche scientifique permet d'établir une hiérarchie des discours sur le tourisme et d'en opérer un classement en privilégiant ceux qui correspondent le mieux à la réalité étudiée.

Il semble évident aujourd'hui que la connaissance préalable du marché est nécessaire au lancement d'un produit. La connaissance des environnements économique, sociologique et politique représente l'a b c du travail d'un gestionnaire du secteur privé ou du secteur public.

Ainsi, la constitution de banques de données fiables, les recherches reliées à certains problèmes importants, les prévisions et les simulations sont des outils indispensables au développement d'une entreprise, quelle qu'elle soit !

Les principaux problèmes du tourisme, à l'heure actuelle, sont l'amateurisme et l'abus d'idées simplistes. L'amateurisme, c'est la quête désarmante de recettes toutes faites et pouvant s'appliquer à toutes les situations (par exemple, les casinos). Le goût des idées simples, c'est l'échafaudage, à partir d'un fait isolé, d'une suite d'interprétations et de recommandations qui s'écroule au moindre souffle de la critique.

Ces deux fléaux sont les deux faces d'un même refus d'aller au fond des choses et de faire une analyse logique de toutes les dimensions d'un problème quelconque. Pour réduire l'attrait des idées simples et s'en libérer, il faut affirmer que le tourisme en général n'existe pas, mais que des situations concrètes existent. En posant comme principe fondamental l'autonomie des phénomènes touristiques, nous nous prémunissons contre les idées générales.

On peut affirmer sans ambages que la téorologie est née bien avant qu'un véritable «corpus» scientifique soit construit. Cette situation pose des problèmes au regard de la crédibilité des praticiens du tourisme, actuels et futurs (les étudiants en tourisme).

C'est un leurre de croire que la crédibilité et la visibilité sociale des praticiens en tourisme pourront s'établir en faisant l'économie d'un corps de connaissances scientifiques sur le tourisme. Une reconnaissance sociale durable doit passer par le biais de la reconnaissance scientifique. La reconnaissance sociale des professions du tourisme (et de la formation en tourisme) ne peut être réduite à une question d'image : elle doit aussi être une question de fond ! Il faut pouvoir disposer d'un corpus minimal de problèmes importants, d'hypothèses plausibles et de théories généralisables. En outre, nos praticiens devraient posséder des outils d'analyse et des moyens efficaces pour répondre aux besoins des industries touristiques ; ils devraient pouvoir poser des diagnostics et agir conséquemment.

Par ailleurs, sur le plan sociologique, ces connaissances scientifiques devraient assurer une plus grande cohésion des différents champs

du domaine touristique en fournissant un vocabulaire commun, des approches et des méthodologies communes. Les problèmes, les théories et les hypothèses pourraient ainsi jouer le rôle d'un «cordon sanitaire » entre les connaissances scientifiques se rapportant au tourisme et les autres connaissances.

Les solutions apparaissent dans la création de programmes de formation et de recherche ayant des liens étroits avec l'industrie touristique, programmes qui intégreraient les multiples facettes du tourisme dans une démarche multifonctionnelle et multidisciplinaire, respectueuse des différences et rigoureuse dans sa formulation.

D'ici l'an 2000, pour l'établissement de connaissances scientifiques dans le domaine du tourisme, d'énormes investissements seront nécessaires. Les techniques pratiques de l'ingénieur s'appuient sur les théories de la physique et des mathématiques. De la même façon, l'étudiant et le praticien en tourisme doivent disposer d'un corps de connaissances rigoureux pour servir d'assise à leurs expertises et à leurs pratiques de gestion et d'intervention.

NAISSANCE DE LA SCIENCE MODERNE

Depuis l'Antiquité, la connaissance oscille entre le rationalisme et l'empirisme ; disons sommairement que, pour le rationalisme, la connaissance s'acquiert par la formulation de propositions, d'idées «a priori» sur les choses, tandis que pour l'empirisme, elle a pour source l'expérience concrète de la réalité. Depuis des siècles, la science zigzague entre ces deux paradigmes fondamentaux.

Le courant rationaliste fut dominant de la Grèce antique jusqu'au Siècle des lumières (vers 1750). S'inspirant, en partie, de l'idéalisme platonicien, Aristote fut le premier penseur du rationalisme ; pour lui, la science se définissait par son universalité et n'était possible que par la formulation de concepts abstraits et généraux. Suivant cette conception de la connaissance, notre intelligence va appréhender l'essence même des choses sensibles et l'ériger en principes universels.

Par la suite (cette évolution s'est déroulée sur plusieurs siècles), le christianisme dominant a lentement assimilé, pour servir ses propres fins, l'armature rigoureuse de la pensée aristotélicienne (notamment au Moyen Âge par saint Thomas d'Aquin, 1228-1274). La pensée libre d'Aristote est devenue, par le biais de la «synthèse scolastique du vile au

XIII^e siècle¹», l'élément moteur du dogmatisme religieux qui allait peser sur les sociétés occidentales jusqu'au XIX^e siècle !

Pour la pensée scolastique, la principale réalité à expliquer, d'où découlent toutes les autres, c'est Dieu «[...] parce qu'il réalise pleinement l'être, et est source à la fois réalisatrice et explicative de tous les autres êtres²». La philosophie thomiste devient, même pour la science balbutiante, la seule référence acceptable. L'alliance entre « le trône et l'autel », le dogmatisme d'État, allait pour longtemps limiter le champ d'action de la science naissante.

L'empirisme trouva ses racines chez les sceptiques grecs qui se méfiaient des idées abstraites et des grands systèmes élaborés par les platoniciens. Pyrrhon, puis Sextus Empiricus (II^e siècle av. J.-C.) jugeaient invérifiable la connaissance de « l'essence des choses » et voyaient d'un mauvais œil la spéculation et la rationalité érigée en système. Pour les sceptiques, la science devait se bâtir sur l'observation concrète et, quand cela était possible, sur des expériences tangibles³. Ce sont là les deux éléments de la science embryonnaire.

Durant le Moyen Âge s'est développé un empirisme rampant ; la menace du bûcher était constante et l'empirisme, pour survivre, devait se teinter de religiosité et montrer « patte blanche » à la police théologique de l'Inquisition⁴. Cet empirisme discret était bien représenté par Nicolas Copernic (1473-1543), Giordano Bruno (1548-1600) et Galilée (1564-1642) ; ces trois physiciens, malgré l'opposition active du clergé de l'époque, ont fait une critique radicale de la cosmologie aristotélicienne. Il faut voir là l'acte de naissance de la science moderne.

Le Siècle des lumières, qui couvre le XVII^e siècle et la première partie du XVIII^e siècle, parachèvera la rupture entre la science et la pensée métaphysique qui tentait de la contenir. Pendant cette période, l'empirisme et le

1. Voir à ce sujet J. Pépin, « Hellénisme et christianisme » et « La philosophie patristique », dans F. Châtelet (1979), *La philosophie de Platon à St-Thomas*, tome 1, Paris, Hachette, Verviers, p. 175 à 238.
2. F.J. Thonnard (1963), « La synthèse thomiste », dans *Précis d'histoire de la philosophie*, Paris, Desclée et Cie Éditeurs pontificaux, p. 354 à 380.
3. Voir à ce sujet H. Dufrénois (1991), *Le savoir et la pratique scientifique*, Montréal, McGraw-Hill, p. 44 à 50.
4. Cette période troublée où l'absolutisme politique et le dogmatisme religieux laissaient peu de place à la pensée libre est bien décrite dans le roman de Marguerite Yourcenar (1968), *L'œuvre au noir*, Paris, Gallimard.

rationalisme sont devenus de véritables écoles de pensée, chacune d'elles raffinant sa vision du monde et ses méthodes de travail.

L'empirisme⁵ rejette tous les apriorismes et réduit la connaissance à la stricte observation par les sens. De plus, seules des expériences répétées peuvent dégager des constantes qui deviennent alors des vérités temporaires pouvant être révisées par d'autres observations et d'autres expériences.

À cette époque, le rationalisme a beaucoup de peine à se débarrasser de la pensée métaphysique. Cette école reste empreinte d'idéalisme, elle rêve d'unifier toutes les connaissances en mettant au jour les principes purs de la pensée. Le monde obéirait à une logique formelle établie sur le modèle des mathématiques et de la physique. Par exemple, pour Descartes⁶ (1596-1650), le monde s'organise autour de postulats et d'axiomes formant une longue chaîne de raisons se justifiant mutuellement. Gottfried Leibniz (1646-1716) fera une tentative (avortée) de synthèse entre les dimensions empiriques et rationnelles de la science.

C'est à Auguste Comte (1798-1857) que l'on doit d'avoir fait de l'empirisme une véritable idéologie sociale⁷. Le positivisme (l'apothéose de la pensée empirique) apparaît comme une philosophie boiteuse qui aurait, malgré tout, socialement réussi. La philosophie positive conservera du rationalisme l'idée d'un système de connaissance universel et de l'empirisme, la nécessité d'établir les lois par des expériences pratiques.

Le positivisme contient renferme certaines positions tenant à la fois de l'empirisme et du rationalisme ; Comte émet dans le «Cours de philosophie positive» l'opinion suivante : «Car si d'un côté toute théorie positive doit nécessairement être fondée sur des observations, il est également sensible, d'un autre côté que, pour se livrer à l'observation, notre esprit a besoin d'une théorie quelconque⁸.» Par la suite, avec sa «Loi des trois états », Comte va développer une espèce de métaphysique de l'empirique, de la science totalement identifiée au «Progrès» humain et social. Malgré ses lacunes et ses dérives, le positivisme aura une grande influence dans la deuxième moitié du XIX^e siècle.

5. Représenté surtout par des penseurs anglais : Francis Bacon (1561-1626), John Locke (1632-1704) et David Hume (1711-1776).
6. Voir à ce sujet J. Laporte (1945), *Le rationalisme de Descartes*, Paris, PUF.
7. Voir à ce sujet A. Comte (1963), *Discours sur l'esprit positif*, Paris, Union Générale d'Éditions.
8. A. Comte (1973), «Le cours de philosophie positive», tiré de *La science sociale*, Paris, Gallimard, p. 129.

Au positivisme initial se greffera ensuite une dimension nouvelle : le «pragmatisme». Bien que le «pragmatisme» découle de «l'utilitarisme» de Jeremy Bentham (1748-1832), c'est John Stuart Mill (1806-1873) qui lui donne la forme la plus complète. Pour les «pragmatiques», le critère de la vérité repose sur la valeur pratique, sur le succès et l'efficacité d'une théorie. Dans cette perspective, une bonne théorie est une théorie qui marche, qui fonctionne et qui est donc profitable à la société ! Le chercheur ne doit s'intéresser qu'à un problème utile et y trouver une solution pratique. C'est à partir de là que les sciences appliquées ont acquis leurs lettres de noblesse.

Le «pragmatisme», qu'il faut voir comme une forme avancée du positivisme, débouche, dans la pratique scientifique, sur «l'instrumentalisme». Pour John Dewey (1859-1952), la connaissance n'est pas un phénomène passif : c'est dans l'action de comprendre que la connaissance émerge. Cette perception a été largement répandue par le physicien Percy Bridgman. Selon lui, «[...] pour découvrir le sens d'un concept scientifique quelconque, il suffit de préciser l'opération physique qui est accomplie pour lui donner ses valeurs⁹».

«L'instrumentalisme», c'est le triomphe de la pensée opératoire : une notion, une idée, un concept prend son sens par les opérations qui vont servir à le mesurer. C'est, par exemple, la pensée opératoire qui fait qu'à la question «Qu'est-ce que l'intelligence ?», Alfred Binet (1857-1911), l'inventeur des tests d'intelligence (quotient intellectuel), répondra : «C'est ce que mesure mon test ! »

Pour «l'instrumentalisme», la notion d'intelligence, de bonheur, etc., se mesure par certaines dimensions correspondant à des aspects du concept utilisé ; chacun de ces concepts aura des indicateurs empiriques servant à mesurer l'intelligence, l'intensité du bonheur, etc. Ainsi, pour Jean Ullmo : «Aujourd'hui, comprendre, c'est être capable de refaire ; un concept est compris lorsqu'on en possède une définition opératoire¹⁰. »

Dans les années 1920, la vieille nébuleuse rationaliste est réapparue sur de nouvelles bases, véhiculée par un nouveau langage. Le «positivisme logique» se démarque de l'empirisme par le désir d'arriver à une vérification logique des théories scientifiques. De même, il rejette l'idéalisme et les relents métaphysiques qui encombraient encore le courant

9. Cité par M. Blaug (1982), *La méthodologie économique*, Paris, Économica, p. 11.

10. J. Ullmo (1969), *La pensée scientifique moderne*, Paris, Flammarion, p. 228.

rationaliste. L'objectif du Cercle de Vienne (qui regroupait certains penseurs de cette école) était de constituer un langage scientifique unifié grâce à une logique symbolique (mathématique)¹¹.

L'insistance sur la logique et la cohérence interne des théories a beaucoup influencé les scientifiques de cette époque. Par exemple, pour les sciences sociales, les notions *d'homo economicus* ou *d'homo sociologicus* laissaient entendre qu'une présentation logique de la réalité correspondait à la vérité scientifique. La leçon que ce courant nous apporte est que la réalité humaine (et même physique) n'est pas réductible à une logique pure. Les systèmes sociaux et économiques ne peuvent être expliqués uniquement par des schémas rationnels ; comme l'écrit Bruno Jarrosson : «La raison ne peut servir à établir des vérités définitives qui rendraient inutile la raison¹².»

Somme toute, le «positivisme », le «pragmatisme », «l'instrumentalisme» et le «positivisme logique » ont plusieurs choses en commun, et particulièrement le refus :

- de la métaphysique ;
- de la recherche des choses «en soi» ;
- de l'introspection ;
- des valeurs ;
- de la spéculation.

Avec le temps, ces différentes visions de la science se sont amalgamées ; l'empiriste pur qui se contentait d'observer, de constater et de mesurer n'existe plus. Est aussi porté disparu (du moins dans les milieux scientifiques) le pur rationaliste adepte d'une théorie universelle passe-partout.

Il faut aussi signaler que les différents courants de la pensée scientifique n'ont jamais été totalement homogènes : des emprunts ont été faits aux uns et aux autres. Un chercheur pouvait faire partie (souvent sans le savoir) de deux courants de pensée à la fois ou même changer sa vision des choses au cours de sa carrière scientifique.

11. Voir à ce sujet J. Bouveresse, «La théorie et l'observation dans la philosophie des sciences du positivisme logique», dans F. Châtelet (1979), *La philosophie au XX^e siècle*, tome IV, Paris, Hachette, p. 70 à 128.

12. B. Jarrosson (1992), *Invitation à la philosophie des sciences*, Paris, Seuil, p. 54.

Une certaine synthèse s'est faite : toute démarche scientifique comportera, à des degrés divers, une part de théorie rationnelle et une part de théorie empirique. Cette part théorique peut être de 90 % en mathématique et de 20 % en histoire (ce ne sont là, bien sûr, que de vagues estimations).

C'est Gaston Bachelard (1884-1962) qui a le mieux actualisé cette synthèse entre l'école rationaliste et l'école empiriste. Dans la science actuelle, il y a un subtil mélange d'un rationalisme tempéré et d'un empirisme mitigé. Le réel et le rationnel ne sont pas les membres d'un couple ennemi mais deux éléments essentiels d'une même chose. Selon Bachelard, le cheminement scientifique se fait du rationnel au réel, et les deux éléments sont transformés par cette combinaison pour produire autre chose qui est la science se faisant. C'est ce qu'il appelle : le «rationalisme appliqué¹³»

Pour Bachelard, la science n'évolue pas comme un train sur ses rails ; cette évolution apparaît comme une suite de ruptures, de dépassement d'obstacles épistémologiques¹⁴ à la connaissance. En effet, le chercheur et l'apprenti chercheur (l'étudiant) doivent faire face à des obstacles à la fois cognitifs et culturels à la connaissance scientifique ; la science avance (selon les disciplines) en surmontant ces obstacles. Retenons les plus importants : «l'expérience première », la «connaissance générale » et «l'obstacle substantialiste ».

«L'expérience première», c'est cette connaissance spontanée et immédiate acquise dans la vie quotidienne, formulée par notre entourage et amplifiée par la bouillie médiatique. Selon Bachelard : «[...] c'est l'expérience placée avant et au-dessus de la critique qui, elle, est nécessairement un élément intégrant de l'esprit scientifique¹⁵. » Cet esprit scientifique doit rejeter cet instantané et ce pittoresque propres à l'éthos journalistique, il doit maintenir une distance à l'objet et à ses symboles, car : « [...] dans la connaissance vulgaire les faits sont trop tôt impliqués dans des raisons. Du fait à l'idée, le circuit est trop court¹⁶ »

13. Voir à ce sujet G. Bachelard (1962), *Le rationalisme appliqué*, Paris, PUF.

14. On peut définir l'épistémologie comme : «[...] l'étude critique des principes, des hypothèses et des résultats des diverses sciences, destinée à déterminer leur origine logique (non psychologique), leur valeur et leur partie objective» d'après A. Virieux-Raymond (1966), *L'épistémologie*, Paris, PUF, p. 3.

15. G. Bachelard (1967), *La formation de l'esprit scientifique*, Paris, Librairie Vrin, p. 23.

16. G. Bachelard, *ibid.*, p. 44.

La «connaissance générale» est l'obstacle le plus insidieux pour la connaissance scientifique car, par son effet, tous les phénomènes sont équivalents ; elle suppose un amasement hétéroclite, une rationalité éclectique qui ne vise pas des éléments spécifiques. Pour Bachelard : «L'objectivité se détermine dans la précision et dans la cohérence des attributs, non pas dans la collection des objets plus ou moins analogues. Cela est si vrai que ce qui limite une connaissance est souvent plus important, pour les progrès de la pensée, que ce qui étend vaguement la connaissance¹⁷. » La science ne peut se limiter à l'information.

«L'obstacle substantialiste» provient de la croyance à l'intériorité des choses ; c'est l'idée que les phénomènes ou les systèmes ont une substance, une réalité intérieure qu'il faudrait découvrir. Comme l'écrit encore Bachelard : «La substantialisation d'une qualité immédiate saisie dans une intuition directe n'entrave pas moins les progrès ultérieurs de la pensée scientifique que l'affirmation d'une qualité occulte ou intime, car une telle substantialisation donne lieu à une explication aussi brève que péremptoire¹⁸. »

La science est plus une construction, une invention qu'une «découverte» au sens habituel. C'est ce qui faisait dire à Jacques Monod, en présentant les résultats de ses recherches sur la structure interne de la cellule (l'ADN ou l'acide désoxyribonucléique) : «Nous ne l'avons pas découverte, nous l'avons inventée.» Une découverte suppose quelque chose qui existait au préalable, donc la fin de la recherche, la fin de la science.

Après 1920, la science a connu un développement phénoménal ; avec l'apparition de nouvelles religions (sectes) et l'émergence de nouvelles disciplines (plus ou moins scientifiques), d'autres questions urgentes se sont posées :

- Quels sont les mécanismes propres à la connaissance scientifique ?
- Qu'est-ce que la vérité en science ?
- Comment distinguer une théorie scientifique d'une autre qui ne l'est pas ?

L'épistémologue Karl Popper a passé une grande partie de sa vie à tenter de répondre le mieux possible à ces questions.

17. G. Bachelard, *ibid.*, p. 71.

18. G. Bachelard, *op. cit.*, p. 102.

Pour Popper : « [...] la science naît dans les problèmes et finit dans les problèmes¹⁹. » Pour résoudre ces problèmes, le chercheur échafaude des propositions, des hypothèses (ce que Popper appelle des « conjectures ») qui sont des théories sur les phénomènes étudiés. Ainsi, selon lui, « les problèmes, y compris les problèmes pratiques, sont toujours théoriques. D'autre part, on ne peut considérer les théories que comme des solutions provisoires apportées à des problèmes et on ne peut les comprendre qu'en relation avec les situations génératrices de problème²⁰. »

Selon cet auteur, s'il est impossible de définir un critère de vérité, il est en revanche possible de définir un critère de démarcation (la réfutabilité) entre une théorie scientifique et une autre qui ne l'est pas. Ainsi : « Une théorie qui n'est réfutable par aucun événement qui se puisse concevoir est dépourvue de caractère scientifique. Pour les théories, l'irréfutabilité n'est pas (comme on l'imagine souvent) vertu mais défaut²¹. » Par exemple : « Tous les cygnes sont blancs » est une théorie falsifiable, car la vue d'un seul cygne noir détruit cette hypothèse. Par contre, l'hypothèse suivante : « Toutes les actions humaines sont égoïstes, motivées par l'intérêt. » Pour Popper : « [...] il est clair que cette théorie, et avec elle toutes ses variantes, n'est pas falsifiable : aucun exemple d'action altruiste ne peut réfuter l'idée selon laquelle il en existe une motivation égoïste cachée²². »

Ainsi, dans la vision poppérienne de la science, le développement des théories se fait en fonction du schéma suivant :



où

P₁ et P₂ = le problème traité avant et après ;

T = la théorie à l'essai ;

E = la discussion critique de l'erreur²³.

Le progrès des connaissances survient dans la distance entre P 1 et P2, c'est-à-dire ce qui reste du problème à étudier.

19. K. Popper (1981), *La quête inachevée*, Paris, Calmann-Levy, p. 185.

20. K. Popper, *ibid.*, p. 189.

21. K. Popper (1985), *Conjectures et réfutations*, Paris, Payot, p. 64.

22. K. Popper (1990), *Le réalisme et la science*, Paris, Hermann, p. 2.

23. K. Popper (1991), *La connaissance objective*, Paris, Aubier, p. 427-428.

Donc, il n'y a pas de vérité en science mais des conjectures qui succèdent à d'autres conjectures : c'est un processus sans fin. Le caractère scientifique d'une théorie se manifeste par son caractère réfutable ; la théorie doit pouvoir être mise à l'épreuve : « [...] le critère de la validité d'une théorie réside dans la possibilité de l'invalider, de la réfuter ou encore de la tester²⁴. » Ainsi, la boucle est bouclée : la connaissance s'établit à partir de problèmes, face auxquels des conjectures sont émises (celles-ci doivent être réfutables) et des tests permettent de confirmer ou d'infirmer les théories.

L'évolution de la pensée scientifique peut être résumée dans un schéma (voir la figure 1). Il faut voir dans cette présentation une manière pédagogique d'illustrer l'histoire de la pensée scientifique : il est bien évident que cette évolution n'est pas aussi linéaire que ce schéma le laisse croire. L'évolution de la pensée scientifique a connu des avancées fulgurantes de même que des régressions ; chacune des écoles a fait des emprunts aux autres (tout en les contredisant) et, dans les sciences d'aujourd'hui, il est encore possible de retrouver des relents d'empirisme naïf et des restes, encore verts, du rationalisme hautain.

Ce qu'il faut conserver à l'esprit, c'est que chacune de ces écoles avait sa propre façon de sélectionner les problèmes jugés importants et de définir ce qui était scientifique (pour l'époque) et ce qui ne l'était pas. Pour les empiristes et les positivistes est scientifique ce qui peut s'observer et se répéter dans des expériences successives. Les tenants du pragmatisme, quant à eux, valorisent les idées utiles pour la société et susceptibles de favoriser le « progrès humain » : une bonne théorie est une théorie qui marche, qui fonctionne.

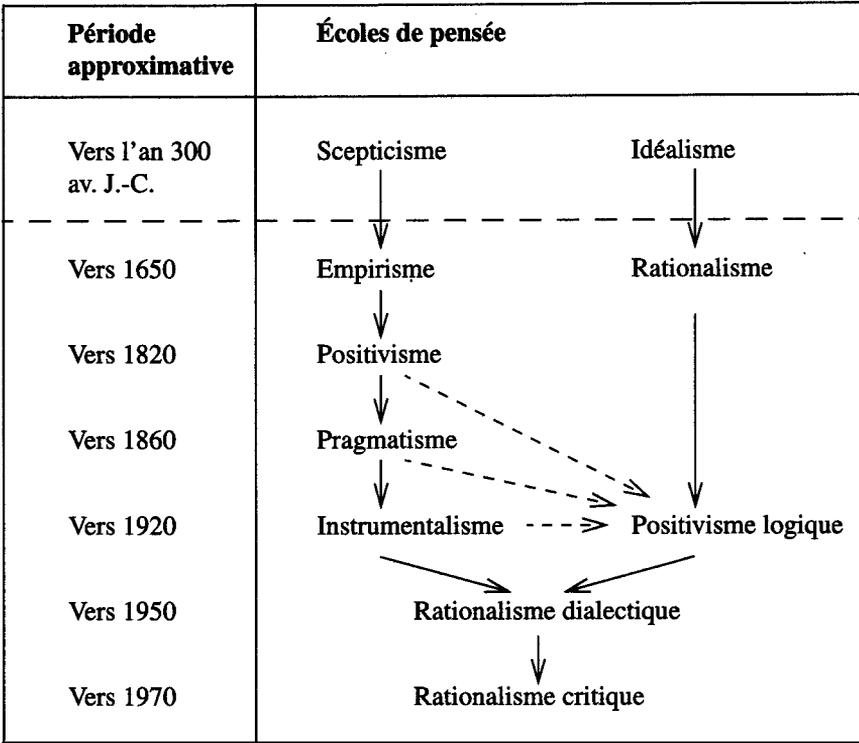
Dans le positivisme logique, la présentation de la réalité doit toujours avoir un très fort contenu logique : est vrai ce qui est rationnel. Pour l'instrumentalisme, une théorie, un concept n'a de sens que si l'on peut le mesurer à l'aide d'indicateurs bien circonscrits. Dans les formes contemporaines du rationalisme (Bachelard, Popper), l'empirique est saisi par des catégories de la pensée et tout le travail de la science consiste à construire ces catégories. Dans cette vision récente, il n'y a pas de vérité définitive mais des conjectures réfutables : si on ne peut prouver qu'une théorie est vraie, on peut prouver qu'elle est fausse.

Assistons-nous à une synthèse définitive de l'empirisme et du rationalisme ? Il semble évident que ces deux grandes familles de pensée

24. K. Popper (1985), *Conjectures et réfutations*, Paris, Payot, p. 65.

Figure 1

ÉVOLUTION DES ÉCOLES DE PENSÉE DANS LES SCIENCES



s'entendent pour dire : « [...] qu'une explication scientifique du monde doit consister en fin de compte dans le fait que les concepts, les propositions et les théories de la science entretiennent avec l'expérience un certain type de rapport (aussi indirect, éloigné et complexe qu'il puisse être) qui n'existe pas dans le cas des productions les plus typiques de la méta- physique et de la théologie²⁵. » La science continue de changer, d'évoluer toutes les heures, tous les jours ; la pensée scientifique et l'épistémologie des sciences²⁶ se raffinent sans cesse. Théorie (pensée rationnelle) et expérience (approche empirique) sont les deux dimensions désormais incontournables de la science «moderne».

25. J. Bouveresse (1973), «La théorie et l'observation dans la philosophie des sciences du positivisme logique », dans F. Châtelet, *La philosophie au XX^e siècle*, tome IV, Paris, Hachette, p. 77.

26. Voir la présentation du rationalisme interactionniste : S. Robert (1993), *Les mécanismes de la découverte scientifique*, Ottawa, Les Presses de l'Université d'Ottawa.

L'ÉMERGENCE DES SCIENCES APPLIQUÉES EN SCIENCES SOCIALES ET EN GESTION

Il est facile de dater assez précisément la naissance des sciences appliquées (en tout cas, des méthodes quantitatives) : entre 1920-1940. La Première Guerre mondiale de 1914-1918 avait provoqué de profonds changements : de nouveaux ensembles politiques et économiques ont été créés, de nouveaux marchés sont apparus. À la suite de cette guerre, des innovations technologiques ont transformé les sociétés : automobile, aviation, télégraphe et téléphone, la radio, etc.

Avant 1920, pour satisfaire les besoins de l'État et de l'industrie, la méthode statistique s'est imposée. Nous devons à Blaise Pascal (1623-1662) d'avoir formulé les fondements du calcul des probabilités qui sont la base même des statistiques ; son ambition était de bâtir une véritable «géométrie du hasard». Pour Pascal, la probabilité d'un événement est liée au nombre de chances pour que cet événement se produise. Jacques Bernouilli (1654-1705) a poussé plus loin cette théorie en énonçant la fameuse «loi des grands nombres» ; selon cette loi, si un événement est répété un grand nombre de fois, la fréquence d'apparition de cet événement sera égal à sa probabilité²⁷.

Dans la même mouvance, nous devons à Abraham de Moivre (1667-1754) la «loi binomiale» et le fameux théorème «limite centrale» qui servent de base à la théorie de l'échantillonnage. Carl Friedrich Gauss (1777-1855) va parfaire la «loi normale» (la fameuse «courbe en cloche») et définir ses principales propriétés²⁸. Enfin, une autre génération de statisticiens allait se consacrer aux relations mathématiques entre les variables et élaborer la «théorie des tests statistiques» fondée sur des hypothèses explicites. Karl Pearson (1857-1937), Ronald Fisher (1890-1940) et George Yule (1871-1951) ont élaboré une multitude de tests métriques et paramétriques pour mesurer le degré d'association entre les variables étudiées²⁹.

27. Voir à ce sujet J.-L. Boursin (1966), *Les structures du hasard*, Paris, Seuil ; J.J. Droesbeke et P. Tassi (1990), *Histoire de la statistique*, Paris, PUF, Que sais-je ?, n° 2527.

28. Pour ces propriétés, voir G. Ouellet (1985), *Statistiques*, Sainte-Foy, Éditions Le Griffon d'argile, p. 202-216.

29. Les modes de calcul et l'utilisation des tests étaient loin de faire l'unanimité ; voir au sujet de la polémique Yule/Pearson : D. Mackenzie (1991), «Comment faire une sociologie de la statistique», dans M. Callon et B. Latour, *La science telle qu'elle se fait*, Paris, Éditions La Découverte, p. 262-298.

Les statisticiens ont fourni les bases scientifiques aux études empiriques effectuées à partir d'échantillons représentatifs. C'est George Gallup qui, le premier, appliquera dans les années 1930, de façon empirique, les théories statistiques aux enquêtes sur le terrain³⁰. L'un des premiers chercheurs à utiliser les sondages pour l'étude de la consommation (les études de marché) a été Paul Lazarsfeld en 1934³¹. George Gallup et Paul Lazarsfeld ont été les véritables initiateurs de l'application des méthodes quantitatives à l'étude du comportement des consommateurs. Ces démarches allaient donner naissance à la recherche marketing actuelle.

Un quatrième facteur (après les bouleversements socio-économiques, la formulation des lois de la statistique et les applications empiriques) jouera dans le formidable développement des enquêtes par sondage : ce sont les innovations technologiques dans le domaine de l'informatique. Une fois résolu le problème de la collecte des données, il fallait trouver une solution au problème du traitement rapide de ces données.

Les premiers ordinateurs avaient la taille d'un gymnase en plus de tomber souvent en panne. L'utilisation des transistors réunis sur des pastilles de silicium (les fameuses « puces ») a entraîné la miniaturisation des appareils. Les nouveaux ordinateurs ont à la fois une grande vitesse de calcul et une mémoire quasi infinie.

Depuis 1970, sont apparus des logiciels perfectionnés de traitement et d'analyse des données ; les deux principaux sont Statistical Package for the Social Sciences (SPSS) et Statistical Analysis System (SAS). Ces logiciels très complets peuvent servir autant à l'étude descriptive des données qu'à des analyses statistiques plus complexes comme l'analyse de régression multiple ou l'analyse factorielle.

D'autres innovations plus récentes (on n'arrête pas le progrès !) vont servir à faciliter le travail de saisie et de codage des informations. Ainsi, le logiciel CATI³² (Computer Aided Telephone Interviewing) permet de faire des entrevues téléphoniques (le logiciel choisit lui-même les numéros à partir de paramètres fixés à l'avance) et de saisir directement les réponses dans un fichier de données mémorisé. Il y a donc un contrôle permanent sur la collecte des informations.

30. Voir à ce sujet A. Max (1981), *La république des sondages*, Paris, Idées-Gallimard, p. 67 à 75.

31. Voir à ce sujet P. Lazarsfeld (1970), *Philosophie des sciences sociales*, Paris, Gallimard.

32. Voir à ce sujet F. Fowler (1988), *Survey Research Methods*, Newbury Park, Sage Publications, p. 130 à 135.

LES PRINCIPAUX THÈMES D'ÉTUDE

Les principaux thèmes de la recherche appliquée en tourisme portent le plus souvent :

- sur l'étude du profil socio-économique des clientèles d'un hôtel, d'un restaurant, d'une attraction touristique, etc. ;
- sur la mesure de la satisfaction des usagers (à l'égard des divers services offerts) ;
- sur la recherche des besoins futurs et des attentes des clientèles.

Les études de type socio-économique visent à mieux comprendre les comportements des consommateurs des produits touristiques. Il s'agit de bien saisir qui sont nos clients et de comprendre leurs principales motivations d'achat. Cette approche permet de créer des « segments » de clientèle et (s'il le faut) d'adapter les produits et les services à ces clientèles particulières. Elle amène aussi le chercheur à distinguer les clients fidèles, des acheteurs irréguliers et des clients possibles (sous certaines conditions que l'enquête peut aider à mettre au jour).

La mesure de la satisfaction des usagers est, en fait, une évaluation de la performance d'un produit ou d'un service. La recherche appliquée sera donc une façon d'exercer un certain contrôle sur le déroulement de la gestion journalière. Elle isolera les éléments qui posent des problèmes particuliers et suggérera des solutions possibles. Ces enquêtes répétées sur plusieurs mois, trimestres ou années vont faciliter un suivi constant des opérations sur le terrain. Elles seront alors le véritable « tableau de bord » de la gestion avec ses « clignotants d'alerte » et ses indicateurs de performance.

Les études des besoins et des attentes sont orientées (habituellement) vers l'avenir. Il s'agit de déterminer pour divers termes (court, moyen et long) les besoins futurs des consommateurs.

Comme chacun peut l'observer, la vie sociale et économique est en perpétuelle transformation : les idées, les modes et les façons de vivre changent et ces changements affectent, souvent profondément, la mise en marché, la vente et le contenu même des produits et des services. L'entreprise, pour survivre dans un monde où la concurrence est très vive, doit s'adapter à ces changements et c'est là qu'intervient l'étude des besoins et des attentes.

Ces trois types de recherche correspondent à ce que Renaud Degon appelle les études «photographiques», les études «assurances» et les études «révélation»³³. Les premières consistent à avoir une image juste de la situation présente de façon à bien la comprendre et à prendre les bonnes décisions. Les études «assurances» ont des vertus à la fois pratiques et psychologiques : en mesurant la performance, on exerce un contrôle constant, ce qui est rassurant pour les décideurs. Enfin, les études «révélation» vont mettre au jour des façons nouvelles de percevoir (et de concevoir) un produit ou un service déjà ancien et menacé d'obsolescence. Ces trois formes de recherche appliquée garantiront une meilleure navigation dans des eaux de plus en plus houleuses.

2 RECHERCHE EN TOURISME ET INDUSTRIE TOURISTIQUE

Toute entreprise naît et grandit dans un monde où une certaine partie des éléments qui conditionnent son développement sont aléatoires. Souvent, la survie de la firme exige de prendre des décisions rapides, sans avoir une connaissance suffisante des problèmes qui se posent. La recherche vise à établir des connaissances objectives des différents environnements où se meut l'entreprise ; elle sert donc à réduire les incertitudes dans la prise de décision et l'ensemble des risques inhérents à la vie économique.

L'objectif de la recherche sera donc, pour l'industrie touristique, d'accroître les connaissances et de réduire, le plus possible, les effets des phénomènes aléatoires sur les actions projetées. L'acquisition de ces connaissances suppose l'établissement d'un système d'information construit en fonction des objectifs de la firme ; on doit aussi imaginer une façon concrète d'utiliser ces informations dans les décisions courantes.

COMPLEXITÉ ET NIVEAU D'INFORMATION

Le besoin d'information sera inégal d'une entreprise à l'autre : il variera selon les types d'entreprises, leurs volumes et la situation de concurrence. On s'entend habituellement pour dire que plus le projet touristique sera complexe, plus le besoin d'information sera élevé. La complexité peut

33. Voir à ce sujet R. Degon (1990), *Les études marketing. Pourquoi ? Comment ?*, Paris, Les Éditions d'Organisation, p. 248-254.

Tableau 1**NIVEAU D'INFORMATION ET COMPLEXITÉ DU PROJET**

Niveau d'information	Complexité		
		Élevée	Faible
Élevé		1	2
Faible		3	4

être mesurée par l'exigence d'intégrer des facteurs (économiques, sociaux et politiques) très différents.

On pourra donc rencontrer les situations suivantes (en simplifiant) :

1. Dans la situation 1, le niveau d'information est élevé (des études et recherches ont été faites) et la complexité est élevée. Dans ce cas, les risques sont réduits au minimum.
2. Dans la situation 2, la complexité est faible et le niveau d'information est élevé ; il y a donc, ici aussi, une réduction maximale des aléas.
3. Dans la situation 3, la complexité du domaine touristique est élevée et le niveau d'information est faible ; dans ce cas, les risques d'échecs sont très grands. Il est donc nécessaire de faire des recherches pour limiter les phénomènes aléatoires.
4. La situation 4 correspond à une complexité et à un niveau d'information faibles. Dans cette situation, l'intuition et l'expérience du décideur peuvent suffire pour atténuer les risques des opérations à exécuter.

On se rend compte que toute décision d'envergure exige des informations pertinentes et sûres.

LA STRUCTURATION DE L'INFORMATION

Une information pertinente et sûre est une information structurée en vue d'atteindre des objectifs précis. Une information structurée est une information construite, conceptualisée ; c'est la seule façon de lui donner un sens univoque. Par exemple, pour mesurer une surface, on peut, comme nos ancêtres le faisaient, la mesurer en pas d'homme, on peut aussi utiliser

la main, des cordes à nœuds, etc. Quand on a défini l'unité de mesure en pouces ou en centimètres, on a structuré l'information, elle est devenue la même pour tous.

En tourisme, la notion de voyage est un problème majeur. Il faut donc structurer l'information dans le sens de la plus grande rigueur conceptuelle et métrique. Par exemple, en définissant un voyage comme un déplacement effectué hors du domicile habituel, de quatre jours et plus, on choisit l'approche la plus stricte. De cette façon, on élimine les voyages de fins de semaine, les aller-retour au chalet, les visites de la fête des Mères, les «ponts» liés à de longs congés, etc. On plonge ainsi au cœur même du tourisme qui se définit par un déplacement, mais pas n'importe lequel !

Comme la valeur d'une information dépend beaucoup plus de ce qu'elle élimine que de ce qu'elle réunit, le principal travail de la recherche consistera à faire le tri dans la masse d'informations existantes pour ne conserver que l'essentiel, que ce qui est pertinent.

LE SYSTÈME D'INFORMATION TOURISTIQUE

Le tourisme est souvent perçu comme un système d'éléments interreliés³⁴ ; de la même façon, la recherche touristique peut aussi être considérée comme un système d'information où l'information structurée est habituellement pauvre et dispersée. Le rôle de la méthodologie est justement de créer un certain ordre, d'offrir une information structurée. Comme on l'a vu, cette information structurée sera le fruit d'une construction raisonnée, c'est-à-dire d'une formalisation.

Le système d'information touristique doit être conçu comme un instrument de collecte et d'analyse de données qui serviront à éclairer et à guider les actions à mener. La collecte des données peut se faire de différentes façons, et la qualité même des données recueillies peut aussi varier de façon considérable. En outre, la qualité des résultats sera assez étroitement liée à la structuration et à la définition des informations.

Un système d'information est généralement constitué de niveaux correspondants aux étapes de la saisie et de l'analyse des données pertinentes ; ces niveaux sont interreliés et dépendants les uns des autres (voir figure 2).

34. Voir R. Mill et A. Morrison (1985), *The Tourism System*, New Jersey, Prentice-Hall.

Le premier niveau peut s'appeler : « Recherche documentaire et formulation des objectifs ». Dans cette première étape, on tente d'élargir, le plus possible, l'état de nos connaissances par rapport au problème à résoudre. Il s'agit donc de faire une recherche attentive dans les journaux et les revues spécialisées ainsi que dans la littérature scientifique qui traite de ce thème. On peut aussi ajouter des entrevues avec des personnes-ressources et des informations privilégiées (non publiées) avec des personnes concernées par ce problème ou un problème similaire. Dans cette étape, on vise à relever et à comprendre toutes les dimensions du problème étudié de façon à pouvoir déterminer des objectifs précis qui pourront être atteints par les autres étapes.

Le deuxième niveau sert à la conceptualisation. C'est une étape plus formelle où seront définies des notions pouvant être utilisées dans les démarches ultérieures. C'est à ce stade que commence à se construire la méthodologie employée tout au long du processus de recherche. Les notions définies doivent avoir à la fois un contenu symbolique et un contenu métrique. Le contenu symbolique se rattache aux objectifs déterminés dans la première étape, tandis que le contenu métrique fait appel à la capacité de mesure de la notion. La définition d'un voyage, donnée en exemple à la page 22, illustre bien cette approche.

Dans le troisième niveau s'effectue la recherche empirique. À l'aide des notions opérationnelles définies dans la deuxième étape, il s'agit d'effectuer des mesures empiriques. On doit donc, sur le terrain, aller chercher des informations à partir des concepts et des notions de base. La recherche empirique peut être qualitative et quantitative. Dans le premier cas, la définition de la mesure sera sensiblement différente même si la logique d'ensemble de la démarche demeure inchangée.

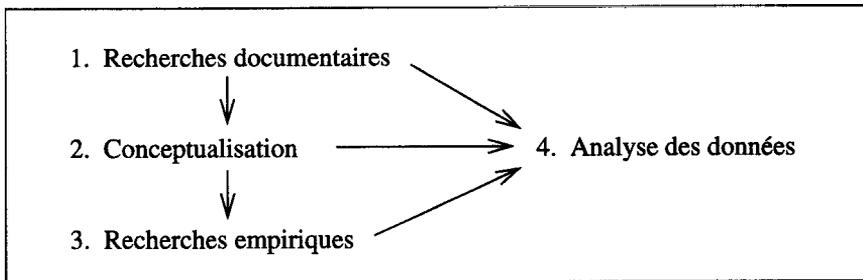
Dans le dernier niveau, il s'agit d'analyser les résultats obtenus. Cette analyse des résultats dépendra directement des objectifs désignés dans la première étape et des notions strictement définies de la deuxième étape. Les résultats n'apparaissent donc pas comme le fruit du hasard mais comme le produit d'une démarche raisonnée. Par conséquent, on peut dire que la perception des résultats est déjà contenue dans la formulation des objectifs.

Le système d'information touristique que nous proposons ici est donc une approche rationnelle de l'information. L'information se doit d'avoir une structure, un éclairage particulier et strictement défini. Ainsi, un fait touristique n'existe que par une certaine perception qui le nomme,

le mesure et l'utilise dans la communication et dans l'action. Cette distinction permet de n'accepter que les faits touristiques qui ont une certaine objectivité (celle-ci étant déterminée par les objectifs et les notions) et de rejeter les informations qui ne sont que des ragots, des préjugés ou des prénotions.

Cette formalisation est le prix à payer pour avoir des informations pertinentes en tourisme. Elle semble permettre de réduire les risques inhérents à ce domaine et de favoriser une meilleure prévision des changements à venir. Par ailleurs, une telle formalisation est une condition nécessaire pour que l'industrie touristique prenne toute sa place dans l'économie nationale.

Figure 2
LE SYSTÈME D'INFORMATION TOURISTIQUE



3 LES FONCTIONS DE LA RECHERCHE EN TOURISME

Les recherches en tourisme remplissent des fonctions multiples et quelquefois contradictoires. La plupart du temps, ces fonctions se complètent et contribuent à étendre notre compréhension des phénomènes touristiques. La première de ces fonctions³⁵ en est une de connaissance orientée vers l'élaboration de théories, d'hypothèses, de lois et de règles générales. Elle sert surtout dans le cas de la téorologie, comme discipline en émergence, à déterminer les problèmes significatifs.

35. Nous nous inspirons ici de la démarche entreprise par A. Mingat, P. Salmon et A. Wolfelsperger (1985), *Méthodologie économique*, Paris, PUE p. 35-50.

Le but de cette connaissance est de créer des faits utiles pour l'action. C'est dire que le fait sert de trait d'union entre la réflexion sur les choses et les objectifs à atteindre. Pas de fait éprouvé, pas de diagnostic valable ; pas de diagnostic valable, pas de reconnaissance sociale des diplômés et de la compétence pratique en tourisme.

L'épistémologie contemporaine nous a montré qu'un fait brut n'existe pas³⁶, et que le fait tel qu'on le perçoit habituellement est une construction de la pensée. Ainsi, pour Jean-Louis Besson : «Fait = fabriqué, construit, produit, etc. Oubliez cela, et vous tombez dans le fétichisme statistique³⁷. » Les faits sont issus d'une théorie, il est vrai que cette théorie peut être minimale ou très élaborée. Selon Carl Hempel : « [...] la maxime selon laquelle on doit rassembler les données sans être guidés par une hypothèse antérieure sur les relations entre les faits se détruit elle-même, et personne ne la suit dans une recherche scientifique³⁸. »

Cette fonction de connaissance comporte aussi une dimension didactique : elle sert à enseigner, aux actuels et futurs intervenants en tourisme, les bases du tourisme perçu comme une discipline et une pratique. La fonction de connaissance alimente ce besoin pédagogique mais il arrive, dans certains cas, que les échanges qui ont lieu entre universitaires servent à raffiner et à préciser des approches théoriques ou méthodologiques leur donnant ainsi une plus grande efficacité.

La deuxième fonction de la recherche en tourisme est une fonction technique. C'est une recherche appliquée qui a pour principal objectif de résoudre des problèmes concrets. Il s'agit pour chacun de ces problèmes de puiser dans l'arsenal méthodologique propre aux statistiques, à l'économie, à la psychosociologie, etc.

Pour l'entreprise touristique, cela veut dire :

- une observation constante des marchés, prévoir les principales tendances et réduire les risques pour les prises de décisions ;
- contribuer à diminuer les coûts et améliorer la qualité des produits touristiques ;

36. Voir à ce sujet G. Fourez (1992), *La construction des sciences*, Montréal, ERPI-Sciences, p. 32-40.

37. J.-L. Besson (1992), «Les statistiques : vraies ou fausses», dans *La Cité des chiffres, Revue Autrement*, série Sciences en société, n° 5, Paris, p. 54.

38. C. Hempel (1972), *Éléments d'épistémologie*, Paris, A. Colin, p. 19.

- développer de nouvelles sources de profits en utilisant et en adaptant les expériences faites ailleurs ;
- aider à la promotion et aux ventes ;
- favoriser la fidélité des clientèles par des études de besoins et des études de satisfaction ;
- faire des études de faisabilité et des études d'impacts.

Ces recherches utilitaires vont généralement se nourrir des théories, des concepts et des méthodologies formulées dans la fonction de connaissance.

La troisième fonction est une fonction de diagnostic et d'expertise. Normalement, le chercheur chevronné (ou celui qui l'est moins) pourra, dans un domaine donné, exprimer des conseils et des recommandations en se basant sur ses propres recherches ou sur celles des autres. Il arrive souvent que ces « donneurs de conseils spécialisés » aient également acquis une expérience pratique enviable dans une entreprise touristique quelconque. Le consultant, dans ce cas, peut s'appuyer sur des expériences déjà réalisées ailleurs ou dans des entreprises similaires.

Enfin, les chercheurs en tourisme remplissent parfois une fonction politique. Il s'agit évidemment de la politique au sens noble du terme qu'il faut bien distinguer de la « politique politicienne » (sous la férule d'un parti). Cette fonction va être soutenue par les trois premières définies plus haut. En outre, cette fonction servira à formuler des politiques générales ou particulières concernant le tourisme et cela, à l'égard d'un État, d'une région, d'une municipalité, d'une association professionnelle ou d'une entreprise. Elle contribuera à définir les objectifs stratégiques et les grandes orientations d'une politique touristique ; elle tiendra compte aussi des programmes et des projets qui serviront de véhicules à cette politique. Soulignons que cette fonction s'exerce dans un cadre multidisciplinaire (aménagement, sciences de la gestion, économie, sociologie, géographie, urbanisme, etc.).

Un chercheur en tourisme peut dans une même année, par exemple, assumer l'une ou plusieurs de ces fonctions. Dans le contexte d'une discipline en émergence, il arrive souvent que le chercheur doive assumer plusieurs rôles assez complexes. Ces fonctions se doivent d'avoir une forte interaction entre elles mais on peut s'attendre, à plus long terme, à ce qu'elles subissent une certaine compartimentation.

4 LES PROBLÈMES PARTICULIERS DE LA RECHERCHE EN TOURISME

La recherche en tourisme pose des problèmes particuliers ; selon J.R. Brent Ritchie³⁹, on peut isoler cinq types de problèmes spécifiques :

1. la valeur des données ;
2. la nature de la population étudiée ;
3. les conditions de la collecte des données ;
4. les processus de décision ;
5. la grande variété des demandeurs de recherche.

LA VALEUR DES DONNÉES

La valeur des données est souvent relative en tourisme et cela tient à trois raisons. La première de ces raisons est la grande complexité de la consommation des voyages ; ceux-ci sont faits d'une grande quantité de dépenses difficiles à retracer et à comptabiliser. On relève également un grand nombre d'intermédiaires (dans un voyage de longue durée, par exemple) et un fractionnement des dépenses elles-mêmes.

La deuxième raison est le caractère irrégulier de ce type de loisir. Cette dimension saisonnière est un obstacle à une observation continue. On peut aussi ajouter que plus on s'éloigne du moment choisi pour les vacances, plus les informations des voyageurs sont sujettes à caution. La troisième raison est liée à la difficulté de faire des prétests quand il s'agit d'événements uniques tels que l'Exposition universelle à Montréal en 1967 et les Jeux d'hiver de 2002.

LA NATURE DE LA POPULATION ÉTUDIÉE

Les touristes sont très mobiles et difficiles, si l'on vise des marchés particuliers, à retrouver dans les échantillons aléatoires (par exemple, les Américains qui sont déjà venus à Montréal).

39. J.R. Brent Ritchie (1975), «Some Critical Aspects of Measurement Theory and Practice in Travel Research», *Journal of Travel Research*, vol. 14, n° 1 ; et aussi : W.F. Van Raaij (1986), «Consumer Research on Tourism : Mental and Behaviour Constructs », *Annals of Tourism Research*, vol. 13, n° 1.

Aussi, sur le plan psychologique, le tourisme est perçu comme une activité un peu magique et féérique (exotique), difficile à couler dans le moule rationnel des questionnaires d'enquête. Par exemple, pour certaines destinations, les tests de sensibilité aux prix ne donnent pas de résultats significatifs.

Par ailleurs, il est difficile de faire des enquêtes auprès des visiteurs étrangers, car les touristes sont souvent en état d'insécurité, état créé par le changement de société, de langue, de culture et d'habitudes.

LES CONDITIONS PHYSIQUES DE LA COLLECTE DES DONNÉES

Souvent dans les recherches touristiques, dans le travail sur le terrain, on doit faire face à des obstacles particuliers : par exemple, interroger des skieurs à l'entrée du remonte-pente quand il fait -20 °C. Les conditions atmosphériques peuvent avoir des effets très négatifs.

La forme de certaines activités touristiques peut aussi représenter un obstacle dans la collecte des données ; par exemple, effectuer une enquête à la sortie d'un festival ou d'un congrès quelconque. Il faut admettre que les lieux physiques propres aux activités touristiques ou paratouristiques posent des problèmes particuliers.

LES PROCESSUS DE DÉCISION

La plupart des recherches en tourisme négligent les processus d'influence liés à la famille⁴⁰ et aux groupes d'appartenance ou de référence⁴¹. Ainsi, les études sur l'emprise des médias et de la sous-culture médiatique sur le touriste sont très rares.

De plus, dans le domaine du tourisme, il y a un écart, plus élevé que dans d'autres domaines, entre les choix effectués en situation expérimentale (dans les enquêtes) et ceux effectués dans la réalité. Les répondants ont tendance à effectuer des choix idéaux sans trop tenir compte des contraintes (famille, travail, santé, finances personnelles, etc.) qui vont limiter ces choix.

40. Voir P. Filiatrault et J.R. Brent Ritchie (1980), « Family Vacation Decision Making. A Replication and Extension », *Journal of Travel Research*, vol. XVIII, n° 4.

41. D'après T. Veblen (1970), *Théorie de la classe de loisir*, Paris, Gallimard.

Enfin, les processus décisionnels en tourisme sont multiples ; ils s'exercent dans :

- les transports ;
- les destinations ;
- l'hébergement ;
- la restauration ;
- les événements ;
- les sites à visiter ;
- les activités.

Pour chacune de ces dimensions du voyage, une foule de combinaisons sont possibles. Par exemple, on peut utiliser dans un même voyage le taxi, l'avion, une auto louée, le bateau, le train, etc. Il s'agit de multiples décisions difficiles à isoler, aspect complexifiant l'étude et l'analyse des processus de décision.

LA GRANDE VARIÉTÉ DES DEMANDEURS DE RECHERCHE EN TOURISME

Nous l'avons vu, la demande touristique est fractionnée à la fois dans l'espace, dans le temps et aussi selon la forme de l'offre touristique (transport, hébergement, restauration, etc.). Ce morcellement de l'offre multiplie d'autant les demandeurs de recherche.

Ces offreurs ayant des intérêts, des objectifs souvent différents vont créer un certain éparpillement. Cela provoque des problèmes : 1) dans la définition des objectifs et des hypothèses de recherches ; 2) dans la formulation des variables ; 3) dans l'analyse et la compréhension des résultats.

La plupart des problèmes de la recherche en tourisme, décrits plus haut, sont loin d'être insurmontables. Ils constituent à la fois un programme de travail et une illustration pratique de la spécificité de la recherche touristique. Les approches formulées pour résoudre ces problèmes vont contribuer à bâtir un ensemble de connaissances dont l'entreprise touristique va bénéficier pour croître et prospérer. Ces problèmes constituent des énigmes, au sens de Thomas Kuhn⁴², qui vont occuper les méthodologues de la recherche touristique pour les dix prochaines années.

42. Voir T. Kuhn (1972), *La structure des révolutions scientifiques*, Paris, Flammarion.

LA RECHERCHE TOURISTIQUE PEUT-ELLE SE DÉVELOPPER DE FAÇON SPÉCIFIQUE ?

Le caractère spécifique de la recherche en tourisme transparaît dans les recherches elles-mêmes... Ces recherches font appel, de plus en plus, à une documentation et à des approches scientifiques spécialisées. Ce mouvement correspond au développement habituel d'une nouvelle discipline scientifique. Cette discipline se bâtit à partir de deux mécanismes de différenciation.

À un premier niveau, la théorie tentera de se distinguer des disciplines sœurs ; par exemple, l'économie du tourisme se construira en fonction de problématiques différentes de l'économie du travail ou de l'économie de l'éducation. De la même façon, la gestion touristique créera des modèles pouvant s'appliquer directement, sur le terrain, au domaine du tourisme.

À un deuxième niveau, la théorie essayera d'établir une barrière entre la science du tourisme et le tourisme dans ses aspects extra-scientifiques. Ces aspects se résument assez souvent à des informations disparates et très inégales sur le plan de la qualité et de la fiabilité. Quand une nouvelle discipline est en émergence, une nette distanciation est souvent difficile à réaliser ; elle demeure un idéal à plus long terme !

La science du tourisme s'est institutionnalisée assez rapidement et selon un schéma assez classique du développement d'une discipline scientifique. Selon Terry Clark⁴³, l'institutionnalisation des disciplines scientifiques se réalise à travers cinq étapes successives. Dans la première étape, il y a des chercheurs isolés qui travaillent sur un objet spécifique ; dans la deuxième, ces chercheurs se rencontrent, de façon sporadique, pour échanger des informations et confronter les résultats de leurs recherches ; dans la troisième, on assiste à l'émergence d'une discipline scientifique peu structurée et de publications spécialisées ; dans la quatrième, la discipline scientifique devient plus organisée (apparaissent les études de premier cycle et les études supérieures avec une plus grande diversification des programmes) ; enfin, nous assistons à la formation d'une science devenue «adulte» et s'appuyant sur des paradigmes de plus en plus reconnus et performants.

43. T. Clark (1972), «Les étapes de l'institutionnalisation scientifique», *Revue internationale des sciences sociales*, vol. XXIV, n° 4, Paris, UNESCO.

Le tourisme, comme discipline scientifique, se situe entre la troisième et la quatrième étape. Les chercheurs sont encore trop peu nombreux et le long travail de distanciation interdisciplinaire et extrascientifique en est encore à ses premiers balbutiements. Le passage de ces différentes étapes exige beaucoup de temps et d'énergie et, de l'intérieur, on a souvent l'impression que le domaine du tourisme progresse par des pulsions suivies de longues accalmies. Dans l'ensemble, la spécificité de la téorologie semble acquise bien que beaucoup d'efforts de consolidation soient encore nécessaires.

RECHERCHE EN TOURISME ET ÉTHIQUE

La recherche dans le domaine du tourisme (comme dans tous les autres domaines de recherche) pose certains problèmes liés à l'éthique propre de la démarche de recherche. Dans la pratique courante, on doit donc respecter certaines règles de déontologie⁴⁴, ces règles s'appliquant à trois groupes :

- les demandeurs de la recherche ;
- le chercheur et ses collaborateurs ;
- la population interrogée (l'échantillon).

Au regard des commanditaires de la recherche, on doit s'efforcer d'identifier correctement les besoins réels d'information de ces clients afin d'avoir un mandat de travail le plus clair possible. Il faut aussi trouver, pour cette étude, un niveau de précision qui soit raisonnable et qui n'entraîne pas de frais trop élevés pour la firme. Les méthodes et les techniques de recherche doivent être très bien décrites et les «limites» de validité des résultats bien indiqués. Enfin, les résultats de la recherche doivent demeurer confidentiels.

Le chercheur doit informer ses collègues et collaborateurs des objectifs et de l'utilisation qui sera faite de cette recherche. Il doit aussi s'assurer d'un contrôle constant de la méthodologie utilisée. Les personnes interrogées pourront avoir connaissance des objectifs généraux de l'étude. Il faut aussi garantir aux répondants l'anonymat et la confidentialité de

44. Voir la «charte» promulguée par l'UNESCO (1975), «Problèmes moraux dans la conduite professionnelle en sciences sociales», dans *Revue internationale des sciences sociales*, vol. XXVII, n° 4, Paris, UNESCO ; et aussi : *Guide d'éthique de la recherche avec des sujets humains* (1996), Ottawa, Conseil de recherche en sciences humaines du Canada, ministère des Approvisionnement et Services Canada.

leurs réponses. De même, on demandera toujours l'autorisation pour tout enregistrement audiovisuel des entrevues ainsi que pour l'utilisation de miroirs unidirectionnels. Dans tous les cas, il faut respecter les opinions et les choix exprimés par les répondants. Ces dimensions éthiques ne doivent jamais être négligées.

Chapitre 2

Le problème de départ de la recherche

Une étape importante de la démarche de recherche consiste à définir ce que l'on veut, ce que l'on cherche exactement. Il s'agit d'esquisser le plan de travail.

Dans un premier temps, il faut faire une analyse de la demande de recherche : qui fait la demande, pourquoi, quels sont les besoins d'information ? Une fois que l'on a cerné le problème de départ, on doit effectuer une «préenquête» qui aidera à faire le tour de la question et à élaborer une stratégie de recherche. La «pré-enquête» fera ressortir les objectifs les plus importants et les hypothèses de

*départ. Enfin, on doit planifier la
recherche et établir les moyens nécessaires
et estimer les coûts de la démarche.*

1 L'ANALYSE DE LA DEMANDE DE RECHERCHE

L'analyse de la demande de recherche comprend des aspects théoriques et des aspects pratiques. Sur le plan théorique, on doit s'efforcer de bien saisir ce que veut le commanditaire ; sur le plan pratique, il faut considérer les coûts (équipement et personnel) et le temps disponible. La plupart du temps la demande de recherche se résume à des questions simples : connaissance des caractéristiques socio-économiques de la clientèle, mesure de la satisfaction à l'égard d'un produit ou d'un service, étude de la notoriété, etc.

Il y a aussi des situations plus complexes où une véritable psychanalyse du client serait nécessaire. Ainsi, il peut arriver que :

- le client ne sache pas ce qu'il veut ;
- le client ne sache pas ce qu'il veut tout en ne sachant pas qu'il ne sait pas ;
- le client sache ce qu'il veut mais ne veuille pas le dire. Le besoin de recherche peut donc obéir à plusieurs mobiles :
 - besoin de connaissance ;
 - pouvoir sur les autres (associés ou concurrents) ;
 - besoin de sécurité contre l'angoisse liée à l'incertitude de la conjoncture économique et sociale ;
 - besoin de camoufler un problème plus important ;
 - besoin de surseoir à une décision difficile (remettre à plus tard).

Il est évident que le chercheur ne doit pas faire une fouille approfondie des mobiles de son client éventuel ; il doit s'en tenir à ce qui lui est officiellement présenté. L'analyse de la demande ne doit porter que sur des éléments objectifs. Ces éléments devraient tenir sur deux ou trois pages et pouvoir être analysés avec une grille comprenant certains critères. Ces critères apparaissent dans le tableau 2.

Suivant cette grille, on se rend compte que plus le niveau conceptuel est élevé, plus la démarche méthodologique devient complexe et cette situation a des effets directs sur les coûts estimés et sur la période de temps nécessaire pour réaliser l'étude. Un niveau conceptuel élevé exigera une lecture approfondie de la littérature scientifique sur le sujet et la mise en place d'instruments de mesure particuliers.

Tableau 2**TYPES DE RECHERCHE ET CRITÈRES D'ÉVALUATION D'UNE RECHERCHE**

Types de recherche	Niveaux des difficultés sur le plan conceptuel	Démarches méthodologiques	Coûts estimés	Temps disponible
<ul style="list-style-type: none"> ▪ Caractéristiques Socio-économiques ▪ Étude de satisfaction ▪ Étude de notoriété 	Faible	Descriptive	Faibles	Faible
Recherche des besoins présents et futurs	Moyen	Descriptive et explicative	Moyens	Moyen
Étude des motivations	Élevé	Explicative	Moyens à élevés	Élevé
Étude des décisions d'achat	Très élevé	Explicative	Élevés	Élevé

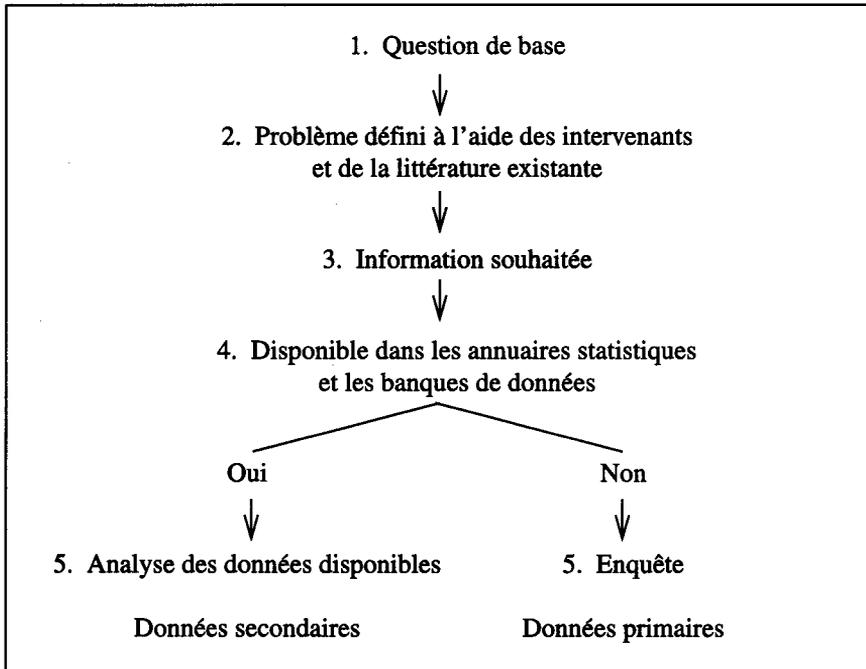
L'analyse de la demande correspond à une étude de pré faisabilité ; à partir de là, trois solutions sont possibles :

1. faire la recherche selon la demande formulée ;
2. ne pas faire la recherche ;
3. faire la recherche en reconsidérant et en réduisant les buts définis préalablement.

L'analyse de la demande permet donc de voir si la recherche est faisable sur le plan théorique et si elle est réalisable en tenant compte des ressources proposées.

Si l'on répond par l'affirmative aux points 1 ou 3, on peut passer à la préenquête.

Nous nous plaçons ici dans le contexte d'une recherche quantitative par sondage. Normalement, l'étude du problème de départ nous amènera à déterminer les renseignements et les informations qui seront utiles pour résoudre ce problème. Ce cheminement se fera de la façon suivante :

Figure 3**SCHÉMA DE DÉCISION AU REGARD DE LA NÉCESSITÉ DE FAIRE UNE ENQUÊTE**

À partir de la question de base, le problème est défini plus précisément. C'est à ce moment qu'apparaît plus clairement le type d'information souhaitée pour comprendre et résoudre ce problème. Ces informations peuvent exister dans les annuaires statistiques, dans les banques de données gouvernementales ou privées ainsi que dans les divers centres de documentation. Si ces informations existent, une grande partie de la recherche est déjà terminée (la plus lourde et la plus longue : la collecte des données). Si ces informations n'existent pas, il faudra les recueillir à l'aide d'une enquête sur le terrain.

Les données déjà existantes dans les bottins statistiques et les banques de données sont appelées «données secondaires », car elles ne sont pas construites par le chercheur mais par des services administratifs qui répondent à des besoins macro-économiques d'information. Nous allons aborder brièvement le traitement des données secondaires en tourisme pour ensuite nous attarder aux données primaires fournies par les enquêtes.

Les données secondaires sont très abondantes en tourisme, depuis une dizaine d'années, tant sur le plan national qu'international. Les données secondaires sont des statistiques administratives dont la seule utilité, semble-t-il, est d'alimenter les modèles explicatifs et les modèles prévisionnels. Ce genre de données pose souvent des problèmes sur les plans théorique et méthodologique.

Les modèles explicatifs servent à représenter la réalité touristique en la schématisant. Il s'agit d'étudier les relations entre certaines variables sélectionnées afin de dégager des règles ou des lois de comportement qui seront utiles dans l'établissement de programmes et de politiques touristiques.

Les données primaires, quant à elles, sont construites par le chercheur lui-même en fonction d'objectifs explicites. Les notions de base et les définitions sont repensées et adaptées à partir de problèmes précis. Ce type de problèmes ne peut être résolu par l'utilisation de données secondaires, car elles sont générales et routinières et ne conviennent pas pour l'analyse des situations complexes comme la motivation ou l'intérêt manifesté par les touristes. Le reste de cet ouvrage sera consacré à la manière dont se construisent ces «données primaires».

2 LA PRÉENQUÊTE

Une fois le but général de la recherche défini avec le client-commanditaire, il faut établir la problématique de cette recherche à l'aide d'une préenquête. La préenquête servira à définir de façon détaillée les besoins nécessaires au regard de l'information. Voici les objectifs de la préenquête :

- étudier tous les aspects d'un thème, d'un problème ou d'une question ;
- établir une hiérarchie entre ces aspects ;
- préciser clairement ce qui doit être étudié ;
- définir le degré de précision souhaitée.

La préenquête vise à «faire le tour» de la question étudiée. La préenquête est une démarche qualitative qui se veut objective ; cette objectivité devrait apparaître dans la volonté d'explorer toutes les dimensions d'un problème ou d'une question. Ce désir d'objectivité est une façon de lutter

contre les préjugés, les prénotions et les «habitudes mentales» qui influencent souvent, de façon implicite, notre jugement.

Sur le plan méthodologique, la préenquête va ouvrir nos horizons face au problème étudié. Si notre préenquête est bien faite, tous les éléments du problème auront été répertoriés. Il s'agira par la suite d'analyser ces éléments ; de tracer les limites, de retenir certains éléments et d'en rejeter d'autres à partir de certains critères de sélection. Ces critères, que nous verrons plus loin, permettront d'arriver à définir la piste de recherche la plus intéressante, celle qui est réalisable sur le plan empirique.

La collecte des informations dans la préenquête doit se faire à quatre niveaux :

1. interroger ceux qui «vivent» le problème de près ou de loin ;
2. recherche théorique sur la littérature existante ;
3. recherche documentaire ;
4. visite des lieux (si possible).

Pour ces quatre niveaux, il s'agit de recueillir :

- des faits,
- des idées,
- des images,
- des «points de vue »,
- des impressions,
- des perceptions,
- des questions,
- des explications,
- des analyses.

Dans le premier niveau, nous devons effectuer des entrevues avec des personnes concernées par le problème. Par exemple, dans le cas de la recherche de la satisfaction de la clientèle d'un hôtel, il faudrait interroger les membres de la direction, certains employés, certains clients. Il s'agit d'entrevues exploratoires ; dans ce contexte, l'intervieweur doit poser des questions assez larges du genre «Quelles sont les causes de l'insatisfaction ?» L'entrevue doit ressembler à une conversation, et les mêmes thèmes doivent être abordés avec toutes les catégories de répondants. L'intervieweur doit prendre des notes et enregistrer les entretiens.

Dans la préenquête (niveau deux), il faut faire une recherche théorique préalable. Celle-ci peut porter sur :

- la théorie générale du consommateur ;
- une théorie particulière, la consommation touristique propre à l'hôtellerie, à la restauration, aux attractions ;
- la recherche de cas semblables, de situations similaires dans d'autres entreprises touristiques ou dans des entreprises d'autres secteurs ;
- une recension des recherches empiriques déjà réalisées sur le même problème dans les cinq dernières années. Cette lecture de la littérature scientifique sur le sujet peut nous apporter des renseignements précieux sur les objectifs, les hypothèses, les concepts, les variables utilisées et les résultats obtenus. Ces renseignements nous permettront d'éviter des erreurs et d'être plus efficaces.

Le troisième niveau de la préenquête consiste à effectuer une recherche documentaire minimale. Celle-ci doit porter, en premier lieu, sur les sources internes à l'entreprise :

- statut juridique ;
- nombre et classification des employés ;
- chiffre d'affaires ;
- résultats des ventes par secteur ;
- procès-verbaux des réunions ;
- documents internes, etc.

La lecture de ces documents devrait nous éclairer sur les buts de l'entreprise, ses stratégies et sur les moyens dont elle dispose.

La recherche documentaire peut aussi porter sur des sources externes de type macro-économique :

- situation économique des entreprises de même taille ;
- données des associations patronales et syndicales ;
- données officielles du ministère du Tourisme, du Bureau de la statistique du Québec, de Statistique Canada ;
- données officielles de l'Office mondial du tourisme (OMT), de l'Organisation de coopération et de développement économique (OCDE).

Le quatrième niveau de la préenquête consiste (si cela est possible) à visiter les lieux comme tels : hôtel, restaurant, attraction, musée ou même le quartier. Cet aspect est souvent négligé, pourtant il arrive souvent que la visite d'un lieu de travail nous apporte une meilleure compréhension du problème à l'étude que la lecture d'une dizaine de documents. La disposition des lieux, l'utilisation de l'espace, les couleurs et les objets, les entrées et les sorties nous aident à mieux saisir ce qui fait problème pour les usagers.

Une fois les données qualitatives de la préenquête recueillies, il faut faire une analyse de ces informations. Cette analyse se fait habituellement en trois temps. Premièrement, il faut présenter de façon synthétique les informations ; la meilleure méthode est de faire une liste de deux ou trois pages des éléments qui ont été abordés. Par exemple, selon la préenquête, certaines causes seront avancées concernant l'insatisfaction présumée de la clientèle :

- mauvaise qualité des services ;
- absence de formation des employés ;
- incompréhension des dirigeants ;
- problème de rentabilité ;
- les prix ;
- la situation économique générale ;
- la négligence des employés ;
- une mauvaise compréhension des besoins des consommateurs ;
- la situation générale du tourisme ;
- l'inutilité de certains services ;
- l'absence de services jugés essentiels ;
- etc.

Deuxièmement, il faut regrouper chacun de ces éléments dans un thème donné ; ainsi, on pourrait classer les explications à saveur économique (prix, rentabilité, situation économique générale) dans le thème « économie ». Il faudrait faire la même chose pour les autres éléments. À la fin, on devrait se retrouver avec cinq ou six thèmes.

Troisièmement, il faut établir la hiérarchie entre les thèmes définis. Cette hiérarchie va permettre de dégager la piste de recherche la plus féconde. Cette approche permet de sélectionner ce qui est vraiment important. Le classement thématique sert à regrouper les choses semblables et à

écarter les éléments contradictoires. La classification hiérarchique permet d'illustrer que les thèmes qui arrivent en dernier sont les moins intéressants. Les critères de rejet peuvent être les suivants :

- contradiction avec les autres thèmes ;
- ampleur du thème ;
- thème difficile à conceptualiser ;
- thème imprécis, trop vague ;
- thème dont l'étude exige beaucoup de temps et des coûts élevés.

En résumé, la préenquête servira à recueillir le plus d'information possible sur le problème à étudier. Par la suite, il faut effectuer un classement, un choix parmi toutes les idées recensées. La sélection des thèmes (et sous-thèmes) importants se fera en fonction de raisons théoriques et de raisons techniques. Les raisons théoriques pourront être les suivantes :

- thème trop complexe ;
- thème peu mesurable par les moyens actuels ;
- mauvaise compréhension du thème par le chercheur ;
- thème contradictoire par rapport aux autres.
- Les raisons techniques concerneront :
 - la disponibilité en termes de temps ;
 - les coûts à assumer ;
 - l'absence ou la rareté des techniques de recherche.

L'une (ou plusieurs) de ces raisons servira de critère de sélection.

3 LES OBJECTIFS ET LES HYPOTHÈSES DE DÉPART

On l'a vu, la préenquête sert à accumuler de l'information, l'étape suivante consiste à réduire cette information à l'essentiel. Ce travail de réduction va surtout porter sur les thèmes prioritaires. Ce processus de raffinement doit se faire de façon systématique¹ : un exemple apparaît dans le tableau 3.

1. Voir à ce sujet J. Chevrier, «La spécification de la problématique », dans Benoît Gauthier (et autres) (1987), *Recherche sociale : de la problématique à la collecte des données*, Sainte-Foy, Presses de l'Université du Québec, p. 49 à 77.

Tableau 3
PROCESSUS DE SÉLECTION DES OBJECTIFS ET DES
HYPOTHÈSES DE LA RECHERCHE

Problème de départ	Thèmes choisis	Questions spécifiques	Objectifs	Hypothèses
Baisse des clientèles et baisse de la «fidélisation» des clients.	Faiblesse de certains services jugés essentiels. Insatisfaction générale de la clientèle.	Quel est le niveau de satisfaction des clients ?	Mesurer le niveau de satisfaction des clients.	On suppose que la satisfaction est reliée à la qualité des services.
		Qu'est-ce qui influence la satisfaction ?	Rechercher la ou les variables qui influencent le plus le niveau de satisfaction.	Le taux de satisfaction pourrait se modifier en fonction de certaines variables : — âge ; — scolarité ; — profession ; — etc. Le degré de fidélisation dépendra du taux de satisfaction.

Quelle que soit l'ampleur de la préenquête, on doit réduire tous les aspects recensés (du problème initial) à quelques éléments faciles à comprendre ; ces éléments vont servir de base à la planification de l'ensemble des opérations de recherche. Dans le tableau 3, nous avons un problème de départ ; à partir de celui-ci, des thèmes particuliers ont été sélectionnés parmi tous ceux recueillis dans la préenquête. La conclusion de la préenquête, c'est de parvenir à définir des questions spécifiques, à déterminer un ou des objectifs précis et à formuler une ou des hypothèses.

L'hypothèse peut prendre la forme d'une relation entre une variable factuelle et un concept ; par exemple, nous pouvons supposer que le revenu brut familial influence le niveau de satisfaction des usagers d'un service quelconque. L'hypothèse peut aussi contenir une relation entre deux concepts ; on peut émettre l'hypothèse que le niveau de satisfaction déterminera le niveau de fidélité des clients. Enfin, il est possible de formuler des hypothèses plus complexes qui réunissent plusieurs concepts ;

par exemple, le niveau social et le niveau culturel des touristes vont influencer le choix d'une destination de voyage.

Il faut souligner que dans la plupart des recherches appliquées (près de 80%), il n'y a pas d'hypothèse, elles se contentent d'avoir un ou plusieurs objectifs qui serviront de ligne directrice à l'ensemble de la recherche. Cela dit, l'hypothèse peut être définie comme étant : « [...] une prédiction consistant à mettre en relation une variable et un comportement. Elle s'exprimera sous la forme : « telle variable a tel effet sur tel comportement ». Une bonne hypothèse est une prédiction précise qui peut être opérationnalisée de façon simple². »

Une hypothèse scientifique doit répondre simultanément à deux critères :

- être réfutable (c'est la contrainte nécessaire définie par Karl Popper, voir page 14) ;
- être opératoire : l'hypothèse doit pouvoir s'actualiser dans des démarches concrètes de recherche, dans des processus empiriques de mesure.

Certains auteurs ont élaboré des critères assez complets d'évaluation des hypothèses ; Gérald Zaltman et Philip Burger définissent ainsi près de seize critères rassemblés en quatre niveaux d'analyse :

- les critères formels ;
- les critères sémantiques ;
- les critères méthodologiques ;
- les critères épistémologiques³.

Cette approche semble bien complexe ; la plupart du temps, compte tenu de la préenquête (si celle-ci est bien faite), l'hypothèse va s'imposer d'elle-même. Si le problème de départ est bien défini, si la sélection des thèmes (à partir des résultats de la préenquête) est effectuée de façon rigoureuse, on devrait parvenir à des questions spécifiques et, de là, à des objectifs assez précis et à des hypothèses cohérentes avec l'ensemble de la démarche. De toute façon, il n'y a pas de recette sûre pour parvenir à des

2. J.-P. Rossi (1989), *La méthode expérimentale en psychologie*, Paris, Dunod, p. 16.

3. G. Zaltman et P. Burger (1975), *Marketing Research*, Illinois, The Dryden Press, p. 585 à 594.

hypothèses intéressantes ; la formulation des hypothèses est une démarche qualitative⁴ qui dépend, en grande partie, des capacités intellectuelles du chercheur.

4 LA PLANIFICATION DE LA RECHERCHE : LES MOYENS NÉCESSAIRES ET LES COÛTS

Une fois l'étape théorique de formulation de la problématique (préenquête) terminée, il faut planifier la recherche sur le plan pratique en tenant compte du temps disponible et des coûts à assumer. Dans cette section, nous allons voir comment planifier l'administration d'une recherche appliquée ; cette gestion « [...] ne peut se faire qu'en fonction de trois critères nécessaires et interdépendants : qualité de réalisation, respect des délais, maîtrise des coûts⁵. »

L'étude minutieuse du problème de base de la recherche permettra d'aboutir à un diagnostic assez complet. Ce diagnostic devrait comprendre une définition passablement précise de l'information souhaitée. Par la suite, il sera possible d'élaborer un échéancier des différentes étapes de la recherche. En règle générale, un échéancier type comprendra les étapes qui apparaissent dans le tableau 4.

Il s'agit, bien sûr, d'une recherche type ; des recherches complexes exigeront des périodes de temps beaucoup plus longues : de deux mois à une année (ou plus). La préparation de l'enquête et des instruments de recherche exige, dans ce cas type, 10 jours de travail, soit 25 % du temps global requis.

Le travail «sur le terrain » est l'étape qui demande le plus de temps (38 % de l'ensemble). C'est aussi l'étape la plus difficile en raison des aléas, des difficultés non prévues, qui apparaissent toujours dans cette partie de l'enquête qui est très empirique.

Le traitement informatique et statistique des données réclame environ 12 % du temps global de la recherche. C'est une phase très technique mais relativement facile à exécuter et à contrôler ! C'est aussi l'étape la

4. Cette démarche qualitative est bien expliquée dans J. Fourastié (1966), *Les conditions de l'esprit scientifique*, Paris, Gallimard, p. 133 à 139.

5. Y. Harvatopoulos, Y. Livian et P. Sarnin (1989), *L'art de l'enquête*, Paris, Éditions Eyrolles, p. 31.

plus standardisée, en ce sens que les opérations à effectuer laissent peu de place à l'imagination et à la fantaisie.

Enfin, en dernier lieu, on doit analyser, interpréter et commenter les résultats en fonction des objectifs de départ et des besoins exprimés par les demandeurs. La formulation des conclusions et des recommandations réclame beaucoup de jugement, d'expérience et une grande capacité de synthèse. Habituellement, pour ce moment crucial, le temps manque toujours... Il faut bien préciser que le temps alloué à chacune des parties correspond à une espèce de moyenne et qu'il peut varier considérablement d'une enquête à l'autre.

Tableau 4
ÉTAPES ET ÉCHÉANCIER DE LA RECHERCHE

Étapes	Démarche méthodologique	Temps requis	%
I. Préparation de la recherche	1. Rencontre avec le client. 2. Préenquête et devis. 3. Définition de la population étudiée. 4. Détermination du mode d'administration du questionnaire. 5. Préparation du questionnaire. 6. Prétest et version finale du questionnaire. 7. Formation des intervieweurs. 8. Échantillonnage.	10 jours	25
II. Réalisation de la recherche	9. Réalisation des entrevues. 10. Contrôle des entrevues. 11. Vérification des questions des questionnaires. 12. Codage des questions fermées et ouvertes.	15 jours	38
III. Préparation des données	13. Entrée des données dans l'ordinateur. 14. Préparation et épuration de la banque des données. 15. Programmation. 16. Production de tableaux informatisés.	5 jours	12
IV. Analyse des données	17. Analyse des données. 18. Interprétation et recommandations. 19. Rédaction du rapport de recherche. 20. Présentation des résultats.	10 jours	25
Durée totale		40 jours	100

Parallèlement à la distribution des étapes de la recherche dans le temps, il faut aussi faire une estimation des coûts pour chacune des étapes (voir tableau 5). Il s'agit de coûts types qui peuvent varier selon les exigences des clients⁶. Par exemple, pour l'étape 2, l'impression du questionnaire, si elle est réalisée sur du papier luxueux en trois couleurs, les coûts vont doubler, etc.

Tableau 5
BUDGET/COÛTS DE LA RECHERCHE

Étapes	Contenu	Coûts	%
1	Rencontre avec le client ; préparation du questionnaire ; pré-test ; version finale du questionnaire (français et anglais) ; préparation des instructions aux intervieweurs et du matériel.	4 000,00\$	9
2	Impression du questionnaire.	1 000,00 \$	2
3	Préparation de l'échantillon (5 000 numéros) n = 1 500 questionnaires complétés.	1 500,00 \$	3
4	Collecte des données ; entrevues ; coordination ; contrôle de la qualité ; location de lignes téléphoniques.	15 000,00 \$	33
5	Frais d'interurbain (calculés en utilisant les rabais de week-end et de soirée).	9 000,00 \$	20
6	Codification des données ; entrée de données ; plan d'analyse ; traitement informatique ; production des tableaux.	4 000,00\$	9
7	Analyse des données ; rédaction du rapport ; préparation d'une présentation ; la présentation des résultats.	7 500,00 \$	17
8	Reprographie, coordination, frais d'administration, frais divers.	3 000,00\$	7
Budget total		45 000,00 \$	100
Coût moyen par entrevue		30,00 \$	

6. Pour une comparaison des coûts d'une enquête selon les pays, voir : pour la France : O. Badot (1990), *L'étude marketing dans une économie de tohu-bohu*, Paris, ESF, p. 138-139 ; pour la Belgique : C. Javeau et C. Vigneron (1989), *Les secrets des sondages enfin révélés*, Bruxelles, Éditions Labor, p. 133 à 137 ; pour les États-Unis, C. Emory et D. Cooper (1991), *Business Research Methods*, Boston, Irwin, p. 125 et 126.

On remarque, dans le tableau 5, que les coûts les plus élevés concernent la collecte des données et les frais téléphoniques (à eux deux, ils s'approprient 53 % du coût total de l'enquête).

La préparation de l'enquête (étape 1) et l'analyse des résultats (étape 7) drainent 26 % de l'ensemble des sommes prévues pour l'enquête. Les étapes plus « techniques » sont, en général, moins coûteuses que les autres (soit les étapes 2, 3, 6 et 8).

Comme nous pouvons le constater, à la lecture des étapes et des coûts, la recherche empirique est un instrument lourd et coûteux ! C'est aussi un outil, nous le verrons plus loin, qui amène aux décideurs une grande quantité d'informations privilégiées sur l'environnement externe de l'entreprise et sur l'état des marchés qu'elle vise.

Les premières étapes de la recherche sont résumées dans le tableau 6. Dans la première étape, il faut bien s'entendre avec le client-commanditaire : on fait une recherche quantitative avec questionnaire et un échantillon représentatif ou l'on fait autre chose (recherches qualitatives, étude des données secondaires, etc.). Si le choix d'une recherche quantitative est accepté, le chercheur doit procéder à une préenquête ; celle-ci peut être plus ou moins approfondie selon la capacité de payer du client.

La préenquête aura des objectifs théoriques et pratiques ; les objectifs théoriques consisteront à bien formuler les thèmes prioritaires, les questions vraiment importantes pour résoudre en tout ou en partie le problème de départ de la recherche. Les objectifs pratiques viseront à déterminer les étapes concrètes de l'enquête et les coûts de ces opérations.

Après l'étude de préfaisabilité, le client-commanditaire peut refuser d'aller plus loin parce que cela ne correspond pas à ses besoins ou parce que les coûts de la recherche sont trop élevés. De même, après l'étude de faisabilité, un refus du client est possible. La plupart du temps, on doit négocier (à la baisse) le devis présenté ; à ce moment, on réduira l'ampleur de la recherche, la taille de l'échantillon et la profondeur de l'analyse des données.

Il arrive aussi que le client, très pressé d'obtenir des résultats, veuille réduire les délais, ce qui amènera une discussion (à la hausse cette fois) des coûts, car il faudra avoir recours à plus de personnel de recherche dans un laps de temps plus court.

Tableau 6**PREMIÈRES ÉTAPES D'UNE RECHERCHE QUANTITATIVE**

Étapes	Étude de préféabilité		Étude de faisabilité : préenquête	
	1	2	3	4
Objectifs	Définition de l'objectif général de la recherche.	Faire le « tour de la question » : recherche des thèmes importants.	Hiérarchie et sélection des thèmes prioritaires.	Déterminer les étapes empiriques et les coûts.
Démarches de recherche	1-3 rencontre(s) avec le client-commanditaire en vue de bien cerner le problème de départ.	Préenquête (terrain) : — recherche scientifique ; — recherche documentaire ; — entrevues ; — aspects physiques.	Formulation des questions spécifiques, des objectifs et hypothèses.	— Estimation de la taille de l'échantillon. — Les coûts en personnel. — Le temps nécessaire.
Documents à produire	Étude de préféabilité de la recherche (de 3 à 5 pages) ; — synthèse des discussions ; — objectif général ; — ordre de grandeur du coût total. Lettre d'intention concernant la préenquête à faire.	Étude de faisabilité (devis final) comprenant : — les résultats de la préenquête ; — les objectifs spécifiques et les hypothèses (s'il y a lieu) — un devis détaillé des coûts et du temps nécessaire ; — le contrat final.		

3 Chapitre

Les outils de la recherche et la collecte des données

Une fois que les étapes sont définies et que les coûts sont établis et acceptés par les demandeurs, il faut effectuer une planification serrée des différentes étapes de la recherche empirique¹. Cette organisation de la recherche va surtout servir à créer des liens étroits entre les objectifs de départ et les outils mêmes de l'enquête.

1. Voir à ce sujet A. Pizam, «Planning a Tourism Research Investigation», dans B. Ritchie et C. Goeldner (1987), *Travel, Tourism and Hospitality Research*, N.Y., John Wiley and Sons.

Le devis de recherche (temps et coûts) constitue lui-même un premier élément de planification de l'enquête. Par la suite, on doit construire les principaux outils de la recherche. De ces outils, le questionnaire est bien l'instrument principal : il faut choisir un mode d'administration qui convienne, rédiger les questions, définir une procédure de codification et effectuer un prétest.

Dans toute recherche quantitative par questionnaire, l'échantillonnage est une étape cruciale : la validité des résultats dépendra de la façon dont l'échantillon a été bâti. L'échantillonnage est une application directe de lois statistiques, aussi certaines règles de base doivent être appliquées avec rigueur. Les meilleures hypothèses, le questionnaire le mieux rédigé, ne pourront jamais suppléer des lacunes importantes dans l'échantillonnage.

1 DES OBJECTIFS AU QUESTIONNAIRE

Dans cette section, on doit établir le lien entre les objectifs (et les hypo-thèses) de la recherche et l'outil de recherche qu'est le questionnaire. On doit maintenant élaborer une certaine conceptualisation de la recherche en fonction des objectifs généraux. Cette conceptualisation consistera à déterminer les variables importantes et à trouver une façon de les mesurer.

Donnons un exemple pratique de la conceptualisation d'une recherche. Supposons que je m'intéresse à la fidélité des clients envers ma station touristique ; je veux connaître les éléments qui déterminent cette fidélité de façon à pouvoir agir sur eux, de manière à intensifier, si possible, cette fidélité de ma clientèle. Guy Serraf définit cette fidélité par l'énoncé suivant : «La qualité des personnes qui ont des goûts stables, une conduite régulière, des attachements affectifs durables². »

Ici, on cherche à connaître comment et pourquoi les clients sont fidèles ; pour trouver une réponse, il faut donc rechercher une autre notion qui soit déterminante. À ce stade, je peux émettre l'hypothèse que les clients qui sont les plus satisfaits sont aussi les plus fidèles. On arrive ainsi à une grille de lecture qui peut prendre la forme du tableau 7.

Tableau 7

DEGRÉ DE SATISFACTION ET DE FIDÉLITÉ

Degré de fidélité	Degré de satisfaction	Élevé	Moyen	Faible
Élevé		1	2	3
Moyen		4	5	6
Faible		7	8	9

2. G. Serrai (1985), *Dictionnaire méthodologique du marketing*, Paris, Les Éditions d'Organisation, p. 113.

Cette typologie permet d'effectuer une classification des diverses clientèles en trois groupes. Dans le premier groupe, on a les cas 1, 5 et 9 ; ce sont les cas les plus logiques :

1. ce sont les clients qui sont à la fois très satisfaits et très fidèles ;
5. ceux qui sont plus ou moins satisfaits et plus ou moins fidèles ;
9. ceux qui sont insatisfaits et infidèles.

Dans le deuxième groupe, on a les cas extrêmes 3 et 7 ; ce sont des cas un peu bizarres et difficiles à expliquer a priori :

3. ici, la satisfaction est faible mais la fidélité est élevée ;
7. là, on est très satisfait mais la fidélité est faible.

Enfin, dans le troisième groupe, nous avons les cas 2, 4, 6 et 8 ; ce sont les clients entre deux eaux : pas tout à fait fidèles, ni tout à fait satisfaits. Ces clients regroupent des fidèles potentiels qu'il serait facile d'atteindre si l'on trouvait le défaut de la cuirasse !

L'étape suivante consiste à faire des hypothèses spécifiques (si cela est possible) pour chacun des types de la grille. Par exemple, on peut faire l'hypothèse que les gens qui sont très satisfaits tout en étant peu fidèles (cas 7) sont empêchés de l'être par la distance à parcourir ou des revenus plus faibles, etc. La plupart des variables socio-économiques usuelles suffiront à vérifier ces sous-hypothèses.

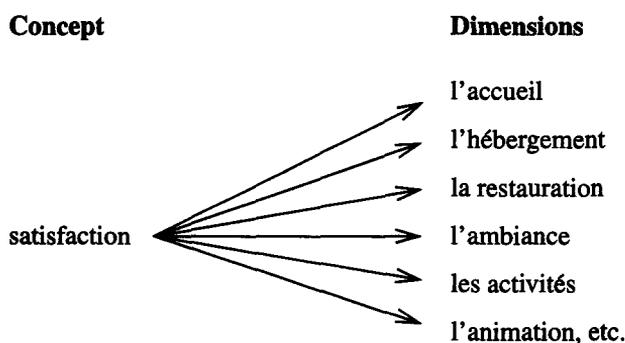
Les objectifs et les hypothèses étant établis, on doit revenir aux deux notions de base, soit la fidélité et la satisfaction. Comme ces notions sont abstraites, il est nécessaire de les préciser, de les rendre opérationnelles. Et la seule façon de rendre une notion concrète, c'est de lui donner des composantes qui la définiront de façon empirique. Ces composantes, ce sont les aspects particuliers et tangibles de la notion théorique.

Le concept de base de la recherche doit être une construction théorique tout en ayant des qualités opératoires³. C'est Paul Lazarsfeld qui a le mieux cerné la démarche de conceptualisation. Voici ce qu'il en dit :

3. L'approche «constructiviste» est très intéressante à cet égard ; voir à ce sujet J.-L. Le Moigne, «Épistémologies constructivistes et sciences de l'organisation», dans A.C. Martinet (et autres) (1990), *Épistémologies et sciences de gestion*, Paris, Economica, p. 81 à 140.

«L'une des principales tâches de la recherche sociale empirique est de traduire les concepts en indices⁴.» Pour Lazarsfeld, cette «traduction» se fait en « [...] quatre phases majeures : la représentation imagée du concept, la spécification des dimensions, le choix des indicateurs observables et la synthèse des indicateurs constituant les indices »⁵.

Cette démarche est devenue un paradigme fondamental de la méthodologie des études empiriques⁶. Pour revenir à l'exemple de tout à l'heure, le concept de satisfaction se définira, au sens littéraire, par «un sentiment de bien-être, un plaisir qui résulte de l'accomplissement de ce que l'on attend⁷». Pour rendre cette définition opératoire, il faut lui donner des aspects, des dimensions qui peuvent être les suivantes :



De cette façon, le concept devient concret et s'inscrit dans une réalité : c'est le passage difficile de l'abstrait au concret⁸.

4. P. Lazarsfeld (1970), «L'apport conceptuel des analyses d'enquêtes à la sociologie générale », dans *Qu'est-ce que la sociologie ?*, Paris, Gallimard, p. 20.
5. P. Lazarsfeld (1965), «Des concepts aux indices empiriques », dans *Le vocabulaire des sciences sociales*, Paris, Mouton, p. 28.
6. Plusieurs ouvrages de méthodologie en font mention ; voir R. Quivy et L. Van Campenhoudt (1988), *Manuel de recherches en sciences sociales*, Paris, Dunod, p. 114-117 ; F. De Sigly (1992), *L'enquête et ses méthodes : le questionnaire*, Paris, Nathan, p. 30-32 ; A. Gilles (1994), *Éléments de méthodologie et d'analyse statistique pour les sciences sociales*, Montréal, McGraw-Hill, p. 23-27.
7. D'après le *Petit Robert*.
8. Pour d'autres exemples pratiques, voir U. Sekaran (1992), *Research Methods for Business*, N.Y., John Wiley and Sons, p. 383-388.

Il faut ensuite trouver des indicateurs pour chacune des dimensions du concept ; par exemple, la satisfaction face à l'accueil (réception) peut être mesurée par des indicateurs tels que les suivants :

- la disponibilité ;
- la rapidité ;
- la qualité des informations fournies ;
- l'amabilité ou la gentillesse, etc.

À ce stade, un nouveau découpage est toujours possible ; par exemple, l'indicateur « les activités » pourrait se mesurer pour l'activité 1, l'activité 2, l'activité 3, etc.

La démarche suivante consiste à trouver un niveau de mesure acceptable pour les indicateurs. Pour ne donner qu'un exemple, un indicateur aussi simple que la fréquentation d'un restaurant comporte plusieurs niveaux de mesure (comme on peut le voir au tableau 8).

Tableau 8

MESURE DE LA FRÉQUENTATION D'UN RESTAURANT

Niveaux de mesure	Questions	Réponses
Échelle nominale	Mangez-vous au restaurant ?	Oui Non
Échelle ordinale	Mangez-vous au restaurant ?	Jamais Rarement De temps en temps Souvent
Échelle par intervalles	Combien de fois avez-vous mangé au restaurant au cours des derniers six mois ?	Jamais 1-10 11-20 21-30 31-40 41-50 51 et plus
	Combien de fois environ avez-vous mangé au restaurant au cours des derniers six mois ?	Réponse littérale

Dans le premier niveau, l'information obtenue est vague et peu intéressante. Dans le deuxième, il y a une ambiguïté possible entre «De temps en temps» et «Rarement» ; les réponses seront difficiles à interpréter. Le premier exemple du troisième niveau donne une information plus précise en indiquant la période de consommation (six derniers mois) ; l'échelle de mesure est graduée par intervalles égaux (à l'exception du dernier).

Le deuxième exemple demande une réponse littérale. Une fois ces réponses rassemblées, on pourra fabriquer des intervalles qui tiendront compte de la moyenne des réponses obtenues. Cette façon de faire permet d'éviter certains problèmes ; par exemple, dans le niveau trois, si 90 % des répondants se retrouvent dans les trois premières catégories, les autres sont inutiles et l'information se retrouvera «tassée» dans ces trois premières catégories. Dans ce dernier cas, il y aura une perte d'information.

Dans les enquêtes, il faut choisir le niveau de mesure qui est le plus performant, qui permet de drainer le plus d'informations possible. Il arrive souvent que les aspects psychologiques du questionnaire priment sur les niveaux de mesure ; par exemple, dans le cas des revenus des personnes interrogées, il est plus facile de proposer une échelle que de demander le revenu (ou le salaire) littéral ; ici, l'échelle par intervalles permet d'atténuer quelque peu le caractère embarrassant de cette demande.

Dans la construction des indicateurs et la rédaction des questions, il faut avoir à l'esprit la précision de la mesure souhaitée. Dans tous les cas, il faut viser la plus grande précision possible. La qualité de l'analyse statistique des données dépendra du niveau de mesure atteint ; par exemple, l'échelle nominale ne permet pas d'utiliser, dans l'analyse des résultats, certaines statistiques descriptives (moyenne, écart type, etc.) ou des tests plus complexes (corrélation, etc.)⁹.

2 LES MODES D'ADMINISTRATION DES QUESTIONNAIRES ET LA RÉDACTION DES QUESTIONS

Avant de rédiger le questionnaire, il faut déterminer de quelle façon celui-ci sera administré. Il y a trois grands modes d'administration des questionnaires¹⁰. Le premier mode comprend les questionnaires envoyés par la

9. Voir à ce sujet J.-P. Crauser, Y. Harvatopoulos et P. Sarnin (1989), *Guide pratique d'analyse des données*, Paris, Les Éditions d'Organisation, p. 19-22.

10. Voir à ce sujet Y. Négro (1987), *L'étude du marché*, Paris, Vuibert, p. 71-72.

poste et ceux qui sont employés avec des échantillons captifs. Les groupes captifs sont formés de personnes réunies pour une raison quelconque et à qui l'on demande de répondre à un questionnaire. Il peut s'agir, par exemple, de touristes dans un autobus, d'élèves dans une classe, d'ouvriers lors d'une réunion syndicale, etc.

Le deuxième mode réunit les recherches où un intervieweur interroge une personne dans une relation «face à face». Le troisième mode concerne les enquêtes où les entrevues se font par téléphone. La rédaction du questionnaire sera différente pour chacun de ces modes. Le tableau 9 résume ces différences.

Chacun des modes d'administration de questionnaire imposera ses règles et comportera des difficultés particulières. Par exemple, un questionnaire administré par la poste exigera près de deux mois pour la collecte des données. Si l'on veut avoir les résultats rapidement (en moins de 20 jours, par exemple), on doit nécessairement faire des entrevues par téléphone, poser des questions plus simples et en moins grand nombre.

Les entrevues «face à face» donnent de meilleurs résultats, habituellement, que les deux autres techniques de collecte des données ; elles permettent d'aborder en profondeur des problèmes complexes tout en apportant des données très riches en termes qualitatifs et quantitatifs. C'est aussi l'approche la plus coûteuse en temps et en argent (il s'agit en fait de la même chose).

Le problème de départ, le temps et l'argent disponibles, vont orienter les choix vers l'un ou l'autre de ces modes d'administration de questionnaires. On peut donc établir une nouvelle typologie en fonction de ces trois contraintes (voir tableau 10).

Par conséquent, si l'on a un problème complexe à résoudre, il faut s'attendre à ce que cela demande beaucoup de temps et d'argent, l'inverse étant également vrai. Le client et le chercheur se doivent de composer avec ces contraintes. En fonction de ces éléments, l'échec d'une recherche est assez facile à prévoir : si l'on a un problème complexe et que l'on tente de l'étudier en investissant peu de temps et d'argent, on est certain de passer à côté et d'arriver à des résultats non significatifs.

À partir de l'objectif poursuivi, et en tenant compte de ces contraintes, on adoptera l'un de ces modes d'administration des questionnaires. Une fois cette étape franchie, on doit rédiger des questions pour chacune des composantes de la notion choisie.

Tableau 9
DIFFÉRENTS NIVEAUX DE PROBLÈME SELON LES MODES
D'ADMINISTRATION DE QUESTIONNAIRE

Critères types	Rédaction sémantique	Indications techniques	Biais possibles à éviter	Nombre de questions et durée moyenne
Correspondance et groupes captifs	<ul style="list-style-type: none"> • Questions précises. • pas trop de questions ouvertes. • Complexité assez grande. • Ouverture sémantique assez grande 	<ul style="list-style-type: none"> • Doivent être très nombreuses et homogènes. • Indications faciles à comprendre. • Indications très espacées. 	<ul style="list-style-type: none"> • Non-réponse à certaines questions importantes. • Le questionnaire peut être rempli par une autre personne que celle choisie. 	<ul style="list-style-type: none"> • 50 à 100 questions. • 10 à 40 minutes.
Entretien personnel « face à face »	<ul style="list-style-type: none"> • Questions plus générales. • Complexité très grande. • Ouverture sémantique très grande. 	<ul style="list-style-type: none"> • Indications peuvent être réduites au minimum. • La formation de l'intervieweur va suppléer. 	<ul style="list-style-type: none"> • La personnalité, le sexe et l'âge de l'intervieweur vont jouer un rôle difficile à évaluer et sont la cause de certains biais. 	<ul style="list-style-type: none"> • 50 à 300 questions. • 30 minutes à 2 heures.
Entrevue par téléphone	<ul style="list-style-type: none"> • Questions très précises. • Questions très courtes. • Complexité faible. • Ouverture sémantique faible. 	<ul style="list-style-type: none"> • Indications très formelles. • Indications très simplifiées. • Peu de choix de réponse. • Phrases courtes. 	<ul style="list-style-type: none"> • Refus de répondre. • Mauvaise compréhension des questions. • Ambiguïté de l'interprétation des réponses. 	<ul style="list-style-type: none"> • Moins de 50 questions. • 10 à 20 minutes au maximum.

Tableau 10**CONTRAINTES ET MODES D'ADMINISTRATION DE QUESTIONNAIRES**

Modes d'administration	Contraintes	Problème de départ, niveau de complexité	Temps disponible	Coûts
Correspondance et groupes captifs		Plus ou moins complexe	1 mois à 3 mois (ou plus)	10 000 \$ à 30 000 \$
Entretien personnel «face à face»		Très complexe	1 mois à 6 mois (ou plus)	30 000 \$ à 60 000 \$
Entrevues par téléphone		Complexité très faible	7 jours à 21 jours (ou plus)	20 000 \$ à 40 000 \$

Tout en rédigeant les questions, il faut réfléchir à la façon d'analyser les réponses des personnes interrogées. Tout questionnaire doit comprendre au moins deux parties. Dans une première partie, on traitera des thèmes, des objectifs et des hypothèses déjà formulés dans la préenquête (la «problématique» de la recherche). Dans une deuxième partie, on posera des questions permettant de bien cerner les dimensions socio-économiques du répondant. Ces questions factuelles apparaissent dans tous les questionnaires ; elles concernent les aspects suivants ;

- l'âge ;
- le sexe ;
- la scolarité ;
- la profession ;
- les revenus ;
- la localisation géographique ;
- la situation familiale, etc.

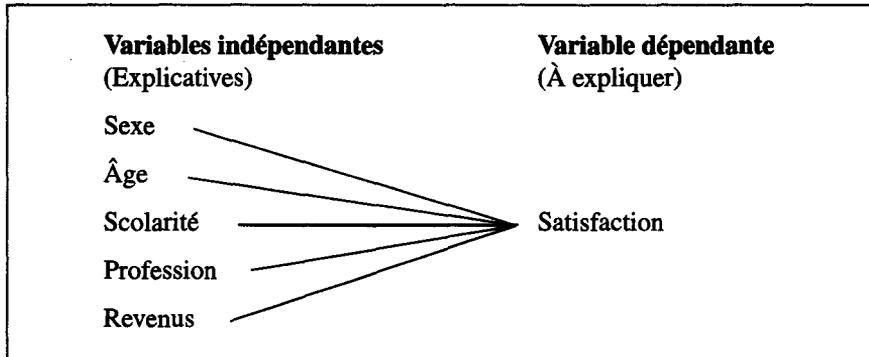
Ces questions factuelles serviront à segmenter les répondants en sous-groupes et à déterminer s'il y a des ressemblances ou des différences entre les sous-groupes ainsi créés. Ces questions sont appelées les variables «indépendantes» par rapport à la ou aux variables dépendantes.¹¹ Par

11. Voir à ce sujet J : H. Guay (1991), *Sciences humaines et méthodes quantitatives*, Montréal, Éditions Beauchemin, p. 42 à 44.

exemple, dans une recherche, la variable dépendante (à expliquer) sera la satisfaction ; celle-ci variera en fonction de l'âge, du sexe, de la scolarité (les variables indépendantes), etc.

Figure 4

VARIABLES INDÉPENDANTES ET VARIABLE DÉPENDANTE



Dans la figure 4, nous voyons une illustration de cet exemple ; pour la variable indépendante « sexe », on pourra vérifier si les répondants de sexe féminin ont un pourcentage de satisfaction différent des répondants de sexe masculin. Nous pourrions faire la même chose pour les autres variables indépendantes qui apparaissent dans la figure 4.

Le statut de variable dépendante n'est pas « coulé dans le béton ». Il est possible, dans certains cas, qu'une variable dépendante devienne, pour une autre variable dépendante, une variable indépendante. Par exemple, la fidélité de la clientèle d'un hôtel peut être définie par le nombre de séjours effectués par le client au cours des trois dernières années ; cette variable dépendante peut être transformée en variable indépendante par rapport au degré de satisfaction. Ainsi, on peut mesurer l'effet de la fidélité sur la satisfaction des usagers.

LES TYPES DE QUESTIONS

Le questionnaire peut varier considérablement selon la formulation des questions. Voyons les formes les plus connues :

1. La question ouverte

Ouverture totale	Que pensez-vous de la politique touristique du gouvernement actuel ?
Ouverture numérique	– Revenus bruts annuels ? – Nombre de séjours dans un hôtel au cours des six derniers mois ?

Dans le premier cas, l'ouverture est totale : le répondant peut disposer de cinq à dix lignes pour répondre ; l'avantage ici, c'est que toutes les opinions peuvent être recensées, et l'inconvénient, c'est la difficulté de codifier (de façon numérique) les réponses. Dans le deuxième cas, la personne indique, de façon littérale, ses revenus et on procède à la codification des réponses en strates après la sortie des tableaux de fréquences simples. Ici, le principal désavantage est psychologique : une stratification a priori garantit mieux l'anonymat du répondant.

2. Question fermée à une seule variable

Fermée à une seule variable	Combien possédez-vous de téléviseurs dans votre logement
	0. Aucun
	1. Un
	2. Deux
	3. Trois
	4. Quatre et plus

Avec ce genre de question, le répondant ne peut choisir qu'une des cinq possibilités. Ici, la codification numérique des réponses est simple et facile. Le principal danger est de choisir des strates trop larges qui rendent inutiles les réponses données. Par exemple, si je demande «le nombre de séjours à l'hôtel au cours des derniers six mois ? », et que le choix de réponse est le suivant :

- 0. Aucun ;
- 1. 1 à 10 séjours ;
- 2. 11 à 20 séjours ;
- 3. 21 à 30 séjours ;
- 4. 31 séjours et plus.

si 80 % des personnes questionnées répondent « 1 à 10 séjours », cette question perdra beaucoup de son intérêt pour l'analyse des données.

Dans ce dernier cas, on laisse une porte ouverte ; il faudrait normalement que plus de 80 % des réponses soient situées dans les catégories 0 à 4. Dans le cas contraire, il s'agit d'une question ouverte déguisée. Il existe un très grand nombre de variantes 12 des quatre types présentés ici. Dans le tableau 11, nous avons un exemple complet d'un questionnaire sur la satisfaction.

Tableau 11

EXEMPLE DE QUESTIONNAIRE DE SATISFACTION

Les Motels Internationaux dans le but de mieux connaître et de mieux servir sa clientèle effectuent actuellement un sondage. Votre nom a été choisi au hasard parmi la liste des clients des trois dernières années. Vos réponses sont assurées de l'anonymat et resteront confidentielles. Aidez-nous à vous aider.

Pour indiquer votre degré de satisfaction, entourez le chiffre qui correspond à votre réponse.

	1	2	3	4	5	6
	Très insatisfait	Insatisfait	Plus ou moins satisfait	Satisfait	Très satisfait	Ne s'ap- plique pas
La réception						
1. L'accueil	1	2	3	4	5	6
2. Disponibilité	1	2	3	4	5	6
3. Rapidité	1	2	3	4	5	6
4. Informations	1	2	3	4	5	6
La chambre						
5. Confort	1	2	3	4	5	6
6. Propreté	1	2	3	4	5	6
7. Température	1	2	3	4	5	6
8. Silence	1	2	3	4	5	6
9. Décoration	1	2	3	4	5	6
10. Dimensions	1	2	3	4	5	6
11. Salle de bain	1	2	3	4	5	6
12. Rapport qualité/prix	1	2	3	4	5	6

12. Voir à ce sujet J. Moscarola (1990), *Enquêtes et analyse des données*, Paris, Vuibert-Gestion, p. 112 à 134.

Tableau 11 (Suite)

EXEMPLE DE QUESTIONNAIRE DE SATISFACTION

	1	2	3	4	5	6
	Très insatisfait	Insatisfait	Plus ou moins satisfait	Satisfait	Très satisfait	Ne s'ap- plique pas
Petit déjeuner à la chambre						
13. Ponctualité	1	2	3	4	5	6
14. Qualité	1	2	3	4	5	6
15. Quantité	1	2	3	4	5	6
16. Service	1	2	3	4	5	6
17. Rapport qualité/prix	1	2	3	4	5	6
Petit déjeuner au restaurant						
18. Qualité	1	2	3	4	5	6
19. Quantité	1	2	3	4	5	6
20. Service	1	2	3	4	5	6
21. Rapport qualité/prix	1	2	3	4	5	6
Restauration						
22. Accueil	1	2	3	4	5	6
23. Choix du menu	1	2	3	4	5	6
24. Qualité	1	2	3	4	5	6
25. Quantité	1	2	3	4	5	6
26. Service	1	2	3	4	5	6
27. Ambiance	1	2	3	4	5	6
28. Rapport qualité/prix	1	2	3	4	5	6
29. Face à l'ensem- ble de votre séjour êtes-vous :	1	2	3	4	5	
30. Recommanderiez-vous ce motel à des parents ou amis ?	Oui	1				Non 2
31. Combien de séjours dans un hôtel avez-vous réalisés dans la dernière année ?	Nombre de séjours					

Tableau 11 (Suite)

EXEMPLE DE QUESTIONNAIRE DE SATISFACTION

32. Indiquez le nombre de voyages (de trois jours et plus) effectués au cours de la dernière année.

Nombre de voyages _____

33. Remarques : _____

Voulez-vous indiquer

- 34. Votre âge : _____
- 35. Votre sexe : Féminin 1 Masculin 2
- 36. Votre profession : _____
- 37. Votre entreprise : _____
- 38. Vos revenus bruts : _____
- 39. Votre adresse : _____
- 40. Vous êtes marié(e) : Oui 1 Non 2
- 41. N° de chambre : _____

Merci de votre collaboration.

Dans l'ordre de présentation des questions, certaines règles de base sont à respecter :

- Poser les questions les plus faciles à répondre au tout début du questionnaire (les questions générales).
- Placer à la fin du questionnaire les questions jugées difficiles, celles qui exigent une forte implication psychologique.
- Regrouper les questions selon les grandes dimensions du questionnaire (définies dans la conceptualisation préalable), par exemple, les questions sur l'hébergement, celles sur la restauration, etc.

Normalement, le questionnaire devrait avoir quatre parties :

1. Une introduction indiquant :

- l'objectif général de la recherche ;
- qui administre cette recherche (entreprise, firme spécialisée, gouvernement, chercheurs universitaires, etc.)
- une phrase ou deux incitant les personnes à répondre ;
- l'assurance que les réponses seront anonymes et confidentielles.

2. Les grandes dimensions du questionnaire.

3. D'autres questions ponctuelles, plus disparates (s'il y a lieu).

4. Des variables indépendantes (surtout les principales caractéristiques socio-économiques du répondant). Tout questionnaire doit contenir au moins cinq ou six de ces variables. Il peut aussi s'agir de variables indépendantes plus complexes telles qu'un indice de qualité de vie (calculé à partir de plusieurs variables socio-économiques).

3LA CODIFICATION DES QUESTIONS ET LE PRÉTEST

Pour Jean-Jacques Lambin : «La codification a pour objectif de mettre les informations recueillies dans une forme compatible avec une procédure de traitement et d'analyse informatisée¹³. » La procédure de codification consiste, pour une question donnée, à attribuer une valeur numérique à chacune des réponses possibles ; le code utilisé doit être exhaustif (couvrir toutes les possibilités présentées) et chacune des valeurs du code doit être exclusive.

Par exemple, à la question suivante : «Combien avez-vous dépensé, au total, pour vos dernières vacances ? », on aura le code de réponses suivant :

Codes	Catégories
0	Aucune dépense
1	500 \$ et moins
2	501\$ à 1000\$
3	1001\$ à 1500\$
4	1 501\$ à 2000\$
5	2 001\$ et plus

13. J.-J. Lambin (1990), *La recherche marketing*, Paris, McGraw-Hill, p. 224.

Ici, toutes les catégories possibles sont représentées et sont mutuellement exclusives.

Toutes les questions apparaissant dans le questionnaire doivent avoir un tel code. Par la suite, la réponse sera transcrite sur une feuille de codification ou directement sur le fichier informatique des données.

La feuille de codification est formée de lignes. Il y aura une feuille de codification par répondant et autant de lignes qu'il y a de variables dans le questionnaire. Cette feuille prendra la forme suivante :

Feuille de codification

Projet : _____ Numéro du questionnaire : ____

V1		V16		V31		V46	
V2		V17		V32		V47	
V3		V18		V33		V48	
V4		V19		V34		V49	
V5		V20		V35		V50	
V6		V21		V36		V51	
V7		V22		V37		V52	
V8		V23		V38		V53	
V9		V24		V39		V54	
V10		V25		V40		V55	
V11		V26		V41		V56	
V12		V27		V42		V57	
V13		V28		V43		V58	
V14		V29		V44		V59	
V15		V30		V45		V60	

Prenons, à titre d'exemple, les douze premières questions qui apparaissent dans le questionnaire du tableau 10, et supposons que deux clients, choisis au hasard, ont rempli le questionnaire.

Pour indiquer leur degré de satisfaction, ils devaient entourer le chiffre qui correspondait à leur réponse.

Répondant 1

	1 Très insatisfait	2 Insatisfait	3 Plus ou moins satisfait	4 Satisfait	5 Très satisfait	6 Ne s'ap- plique pas
La réception						
1. Accueil	1	2	3	④	5	6
2. Disponibilité	1	2	③	4	5	6
3. Rapidité	1	2	3	④	5	6
4. Informations	1	②	3	4	5	6
La chambre						
5. Confort	1	2	3	④	5	6
6. Propreté	1	2	3	4	⑤	6
7. Température	1	2	③	4	5	6
8. Silence	1	②	3	4	5	6
9. Décoration	1	2	3	④	5	6
10. Dimensions	1	2	3	4	⑤	6
11. Salle de bain	1	2	3	④	5	6
12. Rapport qualité/prix	1	2	3	④	5	6

Répondant 2

	1 Très insatisfait	2 Insatisfait	3 Plus ou moins satisfait	4 Satisfait	5 Très satisfait	6 Ne s'ap- plique pas
La réception						
1. Accueil	1	2	3	4	⑤	6
2. Disponibilité	1	2	3	④	5	6
3. Rapidité	1	2	③	4	5	6
4. Informations	1	2	③	4	5	6
La chambre						
5. Confort	1	2	3	④	5	6
6. Propreté	1	2	3	④	5	6
7. Température	1	②	3	4	5	6
8. Silence	1	②	3	4	⑤	6
9. Décoration	1	2	3	4	⑤	6
10. Dimensions	1	2	3	4	5	6
11. Salle de bain	1	2	3	④	5	6
12. Rapport qualité/prix	1	2	3	④	5	6

Dans le cas du 1^{er} répondant on écrira, dans la feuille de codification :

Projet : _____ Hôtel _____ Numéro du questionnaire : 1

V1	4	V16		V31		V46	
V2	3	V17		V32		V47	
V3	4	V18		V33		V48	
V4	2	V19		V34		V49	
V5	4	V20		V35		V50	
V6	5	V21		V36		V51	
V7	3	V22		V37		V52	
V8	2	V23		V38		V53	
V9	4	V24		V39		V54	
V10	5	V25		V40		V55	
V11	4	V26		V41		V56	
V12	4	V27		V42		V57	
V13		V28		V43		V58	
V14		V29		V44		V59	
V15		V30		V45		V60	

Dans la feuille du 2^e répondant on inscrira :

Feuille de codification du répondant n° 2

Projet : _____ Hôtel _____ Numéro du questionnaire : 2

V1	5	V16		V31		V46	
V2	4	V17		V32		V47	
V3	3	V18		V33		V48	
V4	3	V19		V34		V49	
V5	4	V20		V35		V50	
V6	4	V21		V36		V51	
V7	2	V22		V37		V52	
V8	2	V23		V38		V53	
V9	5	V24		V39		V54	
V10	5	V25		V40		V55	
V11	4	V26		V41		V56	
V12	4	V27		V42		V57	
V13		V28		V43		V58	
V14		V29		V44		V59	
V15		V30		V45		V60	

Dans cette codification, la distribution des codes se fera, pour chacune des lignes (de chacune des feuilles de codification), de la façon suivante :

Variables	N° de la question
V 1	Q 1 : Accueil
V2	Q2 : Disponibilité
V3	Q3 : Rapidité
V4	Q4 : Information
V5	Q5 : Confort
V6	Q6 : Propreté
V7	Q7 : Température
V8	Q8 : Silence
V9	Q9 : Décoration
V 10	Q 10 : Dimension
V11	Q 11 : Salle de bain
V12	Q 12 : Rapport qualité/prix

On transpose, dans les lignes, les codes qui correspondent aux réponses fournies. Chacune des variables du questionnaire aura une place déterminée dans les lignes de la feuille de codification. On doit suivre normalement, dans la suite des codes, l'ordre de présentation des questions dans le questionnaire.

En règle générale, il y a une ligne par variable. Dans chacune des colonnes, il y a dix possibilités (0 à 9). Il est possible en utilisant un code impair, c'est-à-dire un code numérique utilisant des chiffres impairs, d'entrer plusieurs informations dans la même colonne. Par exemple, à la question : « Quel(s) moyen(s) de transport avez-vous utilisé lors de votre dernier voyage (3 jours et plus) ? »

Code apparaissant dans le questionnaire	Code utilisé pour la codification
0. Aucun	0. Aucun
1. Avion	1. Avion
2. Auto	3. Auto
3. Autobus	5. Autobus

Avec le code impair, nous avons les possibilités suivantes :

- 0. Aucun
- 1. Avion
- 3. Auto

4. Avion et auto
5. Autobus
6. Avion et autobus
8. Auto et autobus
9. Avion, auto et autobus

Cette approche est très précise mais gourmande pour ce qui est du nombre de colonnes utilisées. Par la suite, à l'aide d'un logiciel informatique, il est possible d'établir des strates qui collent mieux à la réalité.

La codification des réponses doit suivre des règles communes :

- la codification doit être uniforme pour l'ensemble des questionnaires ;
- la formulation du code doit tenir compte des objectifs poursuivis par la recherche ;
- il faut viser l'économie dans l'utilisation des colonnes (quand c'est possible) ;
- il est important de standardiser certains libellés des réponses ; par exemple :
 - 0 = oubli/refus de répondre ;
 - 9 = ne s'applique pas ;
- la distribution des codes de chacune des questions doit correspondre, autant que possible, à l'ordre de présentation des questions dans le questionnaire ;
- la codification des questions ouvertes doit se faire, soit par la même personne, soit par un jury de codeurs, pour éviter les distorsions.

LE PRÉTEST DU QUESTIONNAIRE

Le prétest servira à tester de façon tangible le questionnaire. Ce test a surtout pour objectif de faire ressortir les obstacles à la compréhension des questions et d'évaluer la qualité du questionnaire dans son ensemble. Le prétest devrait porter sur un mini-échantillon de dix à vingt personnes. Ces personnes doivent correspondre sur le plan socio-économique, à la population ou à la sous-population qui sera interrogée ultérieurement. Le mode d'administration (auto-administré, par intervieweur face à face ou par téléphone) du questionnaire, dans le prétest, doit s'approcher (le plus possible) de celui qui sera utilisé dans la réalité.

Une fois le questionnaire complété, la personne responsable du prétest doit demander au répondant quels sont ses commentaires, les questions qui posent des problèmes, etc. Par la suite, il faut analyser l'ensemble des résultats du prétest : le nombre d'oubli et de non-réponses est un bon indicateur de la valeur des questions. Pour chacune des questions, il faut :

- évaluer la qualité sémantique de la question et des réponses possibles : (est-ce que les mots utilisés entraînent l'incompréhension ou la confusion ?) ;
- voir si la codification des réponses est appropriée.

Il faut aussi considérer le degré d'attention que suscite le questionnaire ; voir s'il est trop long, si certaines questions sont trop agressives ou exigent une adhésion immédiate. Si besoin est, il faut reprendre la formulation des questions à problème de façon à les rendre acceptables pour les répondants. Il faut parvenir à établir un juste équilibre entre l'aspect communicationnel du questionnaire et la nécessaire recherche d'information. Dans certains cas (études complexes), il faudra passer plusieurs prétests avant d'obtenir une version définitive.

4 L'ÉCHANTILLONNAGE ET LA COLLECTE DES DONNÉES

L'échantillonnage constitue, dans la recherche touristique, une étape extrêmement importante. Il s'agit de choisir dans une vaste population, des personnes qui seront, par leurs réponses, représentatives de cette population. La collecte des données doit aussi être soigneusement contrôlée.

LES TROIS RÈGLES D'OR DE L'ÉCHANTILLONNAGE

Pour être représentatif d'une population, l'établissement de l'échantillon doit obéir à des règles assez strictes. Ce sont les trois règles d'or de l'échantillonnage. Les voici :

1. Il faut que chacun des membres de la population mère (la population totale visée) ait une chance raisonnable d'être choisi pour faire partie de l'échantillon.
2. Il faut que le choix des personnes, dans la formation de l'échantillon, soit vraiment fait au hasard.
3. Il faut que la taille de l'échantillon soit suffisamment grande.

À ces trois règles s'ajoutent deux contraintes :

1. On doit assurer un contrôle très serré dans la collecte des données (au regard des listes utilisées, du mode de tirage et des personnes choisies).
2. Le taux de réponse au questionnaire doit être supérieur à 50 %.
Nous allons voir en détail chacune de ces règles.

LA BASE DE SONDAGE

Pour se conformer à la première règle, il faut établir ce qu'on appelle «la base de sondage» ; cette base est formée d'une ou de plusieurs listes de personnes faisant partie de la population globale. Par exemple, pour une ville (ou une région), la base pourrait être l'annuaire téléphonique ou la liste des électeurs (si elle est récente).

Pour une entreprise, la base de sondage pourrait être la liste des clients des derniers six mois ou de la dernière année. Dans le cas de l'hôtellerie, par exemple, pour une étude de satisfaction, la base de sondage sera formée par les personnes ayant loué des chambres dans la dernière année (à condition, bien sûr, d'avoir enregistré leurs noms et leurs adresses).

L'établissement de ces listes peut être une tâche très simple ou bien très compliquée. Dans le cas des employés d'une entreprise, il faudrait utiliser la liste à jour des employés, dans le cas d'une université, nous pourrions utiliser la liste des étudiants inscrits au trimestre en cours : établir la base de sondage est relativement simple, dans ces deux cas, car il s'agit d'obtenir les autorisations nécessaires auprès des responsables administratifs.

Quand des listes à jour n'existent pas, le travail d'élaboration de l'échantillon devient beaucoup plus difficile. Si je veux interroger les Québécois qui ont effectué un séjour en France dans la dernière année, je dispose bien sûr des annuaires téléphoniques de l'ensemble du Québec ; pour rejoindre un répondant possible (ayant voyagé en France dans la dernière année), il faudra probablement interroger, au préalable, une cinquantaine de personnes. Pour avoir 1 000 répondants qui vont satisfaire à la condition (avoir voyagé en France), nous devons atteindre 50 000 personnes, ce qui entraînera des coûts astronomiques.

Dans cette première étape, le travail le plus important consistera à trouver et à rassembler ces listes. Par la suite, il faudra «nettoyer» ces listes, les épurer en enlevant les personnes décédées ou qui n'habitent plus la ville ou la région. Beaucoup de listes de clients sont périmées ou déficientes par rapport à l'adresse ou au numéro de téléphone. C'est un long travail de patience de reconstituer cette population de clients, d'acheteurs ou de citoyens. La plupart du temps, la liste réelle n'existe pas : pour un restaurant, il s'agira de la liste des clients qui consommeront un repas le midi et le soir dans les deux prochaines semaines (dans le prochain mois ou la prochaine année).

LE CHOIX DES RÉPONDANTS

Le choix des répondants doit être fait absolument au hasard. Il ne s'agit pas ici du «hasard culturel» où les personnes sont choisies suivant leur apparence ou de certains signes culturels distinctifs mais d'un hasard qui soit systématique. Le hasard systématique est un hasard «construit» de telle façon que le tirage des personnes est totalement aléatoire au sens mathématique du terme.

Les problèmes liés au tirage des répondants sont très variés et il existe de multiples techniques permettant d'en venir à bout ; il faut dans chaque cas trouver une technique qui s'adapte bien à notre problème. On distingue les méthodes probabilistes et les méthodes non probabilistes. Les méthodes probabilistes fournissent une certaine garantie que tous les membres de la population mère auront une chance raisonnable de faire partie de l'échantillon. Les méthodes non probabilistes ne permettent pas de calculer la probabilité réelle qu'une personne aurait d'être choisie ; ce sont des méthodes qui s'appliquent à des cas particuliers.

Les méthodes probabilistes peuvent être regroupées en quatre catégories :

- l'échantillon aléatoire pur ;
- l'échantillon aléatoire systématique ;
- l'échantillon stratifié ;
- l'échantillon par grappes et aréolaire.

L'échantillon aléatoire pur est un mode de tirage «avec remise», ce qui veut dire qu'en théorie, à chacun des tirages, la probabilité est toujours la même pour chacun des individus (dans une population donnée)

d'être choisi. Sur le plan pratique, cette méthode est difficilement applicable (et rarement appliquée). Il faut allouer à chacun des membres de la population un nombre arbitraire, ensuite on doit, à partir d'une table de hasard, choisir les individus qui feront partie de l'échantillon. Pour une entreprise (grande ou petite), il est possible, à l'aide d'un logiciel informatique, d'utiliser cette méthode si la liste des employés ou des clients est complète et informatisée.

Cette méthode est souvent utilisée pour les sondages téléphoniques ; dans ce cas, il faut tenir compte des trois premiers numéros qui indiquent le quartier, les quatre autres sont générés par un logiciel spécial. Si la population étudiée s'étend sur un vaste territoire, il faut ajouter l'indicatif régional. La plupart du temps, on doit produire un très grand nombre de numéros, car près de la moitié des numéros tirés devront être rejetés (commerces, services privés et publics, entreprises, boîtes vocales, cellulaires, télécopieurs, etc.).

L'échantillon aléatoire systématique est plus facile à utiliser. Par exemple, si l'on doit faire une recherche pour une compagnie aérienne sur la qualité du service en vol, nous pouvons choisir les personnes à interroger à l'intérieur même de l'avion. Dans l'avion, on peut numéroter les sièges et en choisir 50 au hasard en tenant compte d'un intervalle. Par exemple, s'il y a 250 places dans l'avion, l'intervalle sera :

$$\frac{250}{50} = 5$$

On choisira un nombre de départ entre les numéros 1 et 5 ; supposons que ce numéro soit le 3, on aura :

première personne interrogée = siège n° 3
 deuxième personne interrogée = siège n° 3 + 5 = siège n° 8
 troisième personne interrogée = siège n° 8 + 5 = siège n° 13
 quatrième personne interrogée = siège n° 13 + 5 = siège n° 18

On continuera jusqu'à ce qu'on atteigne 50 personnes, et ce seront les personnes qui ont les sièges correspondant à ces numéros qui seront choisies et interrogées. S'il y a dix vols dans la semaine, on aura donc 500 questionnaires remplis.

Pour une liste de noms (clients, usagers, etc.), on procédera de la même façon. La plupart du temps, pour les recherches dans le secteur du tourisme, ces listes n'existent pas. Par exemple, dans le cas d'un événement touristique, on doit faire une estimation du nombre total des par-

ticipants en se basant sur les années passées ou sur des événements similaires. Supposons que l'événement a attiré 50 000 personnes l'année passée et que, cette année, nous voulons rejoindre 300 personnes (l'échantillon) ; dans ce cas, l'intervalle sera de :

$$\frac{50\ 000}{300} = 166,6 \text{ ou } 167.$$

On doit donc choisir un nombre de départ entre 1 et 167. Supposons encore que ce nombre soit 42 ; on interrogera donc le 42e participant qui franchira la porte de sortie et, pour tous les autres, on ajoutera l'intervalle de 167. Il faudra noter les refus de répondre et les personnes inadmissibles (âgées de moins de 18 ans, ne parlant ni français ni anglais, etc.).

Le problème du tirage peut devenir très compliqué si les participants quittent l'événement en même temps (par exemple, lors d'un concert) ; il faut alors prévoir plusieurs équipes d'enquêteurs qui travailleront très rapidement. La plupart du temps, il faut faire appel à d'autres méthodes d'échantillonnage. Quand la population est hétérogène, la méthode de stratification sera utilisée. Cette méthode tente de tenir compte des différents sous-groupes que renferme la population mère. Prenons, par exemple, une chaîne hôtelière comptant 8 000 employés. On veut interroger un certain nombre d'entre eux relativement à leurs besoins de formation en tourisme et l'on dispose des données suivantes :

N (population mère) = 8 000 employés

n (échantillon souhaité) = 300 employés

Types d'emplois :

	%
Entretien général	10
Entretien des chambres	25
Serveurs/serveuses	18
Accueil	10
Administration	7
Cuisines	30
TOTAL :	100

Échantillon stratifié

Entretien général :	0,10 x 300 =	30
Entretien de chambre :	0,25 x 300 =	75
Serveurs/serveuses :	0,18 x 300 =	54
Accueil :	0,10 x 300 =	30
Administration :	0,07 x 300 =	21
Cuisines :	0,30 x 300 =	90
TOTAL :		300

Pour chacune des strates, par exemple l'entretien général, il faudra utiliser la méthode de l'échantillonnage aléatoire systématique pour le choix des répondants. Ainsi, la strate entretien général comprend 800 employés (0,10 x 8 000) et l'intervalle sera de :

$$\frac{800}{30} =$$

On agira de la même façon pour les autres strates.

La stratification utilisée peut aussi reposer sur des hypothèses et des estimations. Par exemple, un Salon du tourisme présentant les principales destinations québécoises a lieu les jours suivants :

Taille de la population (N)

Jours	Nombre estimé	%
Jeudi soir	12000	12,6
Vendredi soir	18000	19,0
Samedi	30000	31,6
Dimanche	35000	36,8
TOTAL :	95000	100

Échantillon (n) = 500

Jours	Nombre par strates	Intervalle
Jeudi soir	500 x 0,126 = 63	190
Vendredi soir	500 x 0,19 = 95	190
Samedi	500 x 0,316 = 158	190
Dimanche	500 x 0,368 = 184	190
TOTAL :	500	

L'intervalle correspond donc à :

$$\frac{95\ 000}{500} = 190.$$

Ici, on fait l'hypothèse que les visiteurs seront différents (au regard de certaines caractéristiques socio-économiques) selon les journées d'ouverture du Salon. Là aussi, le tirage se fera strate par strate. Par exemple, pour le jeudi soir, il faudra trouver un nombre de départ dans l'intervalle 1 à 190 ; supposons que ce nombre est 100, ce sera la 100^e personne qui sera interrogée et, ensuite, ce sera la 290^e, etc. Cette méthode tiendra donc compte des différences entre les sous-groupes de la population mère ; elle suppose, toutefois, une recherche préalable pour bien connaître (ou estimer) le nombre et la taille de ces sous-groupes.

Une autre méthode de tirage souvent utilisée est la méthode par grappes et aréolaire. Expliquons par un exemple : je veux interroger 400 employés de restaurants qui comptent moins de 20 employés dans la ville de Montréal et le nombre total de restaurants est de 850. Supposons que la moyenne des employés par restaurant soit de 15. On aura donc :

$$\frac{400}{15} = 26,6 \text{ ou } 27$$

(27 restaurants x 15 employés = 405 employés) restaurants qui devront être tirés des 850 qui existent ;

$$\frac{850}{27} = 31$$

sera l'intervalle du tirage. On choisit un nombre entre 1 et 31 ; disons le numéro 10 et tous les employés de ce restaurant qui porte le numéro 10 (sur une liste numérotée de 1 à 850) seront interrogés. Il y a donc une grande économie de temps et de déplacements sur le territoire.

Ce sondage peut être fait aussi à deux degrés : dans un premier temps, on choisit le restaurant et dans un deuxième, on opère un tirage parmi les employés. La méthode à deux ou plusieurs degrés revient à faire plusieurs tirages, ce qui augmente la précision mais diminue l'efficacité du tirage par grappes (et ses qualités).

Le tirage aréolaire est une simple variante du tirage par grappes. Par exemple, dans un quartier densément peuplé, on peut effectuer un tirage des maisons à appartements et interroger tous les habitants de 18 ans et plus qui vivent à ces endroits.

Ici, la population mère est formée du nombre des maisons à appartements du secteur. Supposons qu'il y ait 500 maisons à appartements de 6 logements chacune et que chacun de ces logements soit habité, en moyenne, par deux personnes et que la taille de l'échantillon soit fixée à 120 personnes. On aura dans ce cas, $n = 120$; on doit donc choisir 10 maisons à appartements, chacune d'elles abritant 12 personnes ; le calcul est le suivant :

$$\frac{120}{12} = 10$$

maisons à appartements = échantillon (10 maisons à appartements \times 6 appartements \times 2 personnes = 120 personnes) ; on aura :

$$\frac{500}{10} = 50,$$

ce qui correspond à l'intervalle du tirage des maisons à appartements. Ces maisons devront être numérotées sur un plan, et l'on suivra la méthode habituelle de tirage de l'échantillon aléatoire systématique.

Pour les sondages à plusieurs degrés, on doit opérer un tirage au niveau du logement ou du foyer. Dans le tirage d'un numéro de téléphone ou d'une adresse de logement, il faut savoir qui interroger dans ce logement ou à ce numéro de téléphone. La technique la plus simple est de choisir la personne dont l'anniversaire de naissance est le plus récent. Une technique plus complexe, l'approche de Kish¹⁴, suppose que l'on fasse la liste de toutes les personnes du foyer à partir de l'âge (ou d'une autre caractéristique) et que l'on effectue, à partir de cette liste, un tirage aléatoire.

LES MÉTHODES NON PROBABILISTES

On recourra aux méthodes non probabilistes lorsque les méthodes probabilistes se révéleront très difficiles, sinon impossibles à utiliser. Même dans ce cas, il faut se rapprocher, dans la confection de l'échantillon, le plus possible des règles des méthodes probabilistes. La méthode par quotas est une méthode non probabiliste souvent utilisée quand il y a absence de listes ou pour des sujets d'études très spécifiques.

14. Voir à ce sujet A. Tremblay (1991), *Sondages : histoire, pratique et analyse*, Boucherville, Gaëtan Morin, Éditeur, p. 182-184.

Donnons un exemple d'échantillon par quotas. Supposons que, dans une petite ville, nous voulons connaître l'opinion des résidents sur un nouveau projet touristique ; dans cette ville, nous disposons des données suivantes sur les résidents :

		%
Sexe	Féminin	50
	Masculin	50
Âge	18-34	40
	35-54	40
	55 et plus	20
Scolarité	Primaire et secondaire	40
	Collégial	40
	Université	20

Supposons, en outre, que nous voulons interroger 300 résidents. Nous allons construire l'échantillon en fonction des données existantes et nous bâtirons une maquette de cette population qui prendra la forme suivante :

Nom de l'intervieweur :											
Zone géographique :											
Date :											
Note du directeur d'enquête :											
Numéro de la personne interrogée											
Variables		1	2	3	4	5	6	7	8	9	10
Sexe	Féminin	x	x	x	x	x					
	Masculin						x	x	x	x	x
Âge	18-34	x	x				x	x			
	35-54			x	x				x	x	
	55 et plus					x					x
Scolarité	Secondaire et moins				x	x				x	x
	Collégial		x	x				x	x		
	Universitaire	x					x				

Chacune de ces maquettes peut contenir 10 répondants ; il faudra donc construire 30 de ces maquettes pour voir un échantillon de 300 personnes. Le personnel d'enquête (sur le terrain) devra respecter les critères qui apparaissent dans cette maquette. Ainsi, la personne numéro 1 devra être une femme de 18 à 34 ans et ayant une scolarité universitaire ; à l'inverse, la personne numéro 10 devra être de sexe masculin, avoir 55 ans ou plus et avoir une scolarité secondaire (ou moins). L'intervieweur n'est pas obligé de commencer par le numéro 1, mais il doit choisir des personnes qui correspondent aux critères définis sous chacun des numéros.

D'autres contraintes peuvent aussi être imposées aux personnes qui mènent les entrevues : on peut obliger les intervieweurs à suivre un itinéraire précis à l'intérieur de la ville et à sonder des groupes particuliers. Quand cette méthode est bien utilisée, elle donne des résultats similaires à ceux obtenus par les méthodes probabilistes : l'Institut de démographie d'Allensbach¹⁵, en Allemagne, a effectué des comparaisons expérimentales systématiques pour la même enquête (et avec le même questionnaire) sur les résultats obtenus par une méthode probabiliste et ceux obtenus par la méthode des quotas. Il semble que les résultats soient presque identiques. L. Kish propose¹⁶, pour les sondages par quotas, un facteur de correction de 1,4 sur l'erreur possible sur les réponses ; ainsi, un sondage par quotas ayant une erreur de 3 % verrait cette marge d'erreur se transformer de la façon suivante : $3\% \times 1,4 = 4,2\%$.

Dans les autres méthodes non probabilistes, on cite habituellement les échantillons de convenance, de commodité ou discrétionnaires ; ils sont souvent utilisés dans les recherches exploratoires ou pour tester des échelles d'opinions ou d'attitudes. Dans tous les cas, les résultats des recherches effectuées à partir de ces méthodes ne peuvent être représentatifs de l'ensemble de la population étudiée ; ces résultats resteront toujours contestables. Beaucoup de chercheurs novices (ou malins) finissent par oublier l'origine de leurs données et formulent des théories ou des «tendances» qui présentent la perfection tenace des illusions.

15. Expériences citées par E. Noelle (1966), *Les sondages d'opinion*, Paris, Les Éditions de Minuit, p. 170-177.

16. Mesure citée par R. Delbès et E. Teyssonnière (1983), *Études de marché*, Paris, J. Delmas et Cie, p. C6.

LA TAILLE DE L'ÉCHANTILLON

Le choix de la taille de l'échantillon est lié, d'une part, à la marge d'erreur que l'on est prêt à accepter sur les réponses et, d'autre part, aux coûts des entrevues à faire. Donald Lehmann ¹⁷ propose une formule simple pour estimer les coûts qui dépendent de la taille ; la formule est la suivante :

$$n = \frac{B - CF - CV}{CME} ; \text{ où :}$$

- n = la taille de l'échantillon ;
- B = budget total alloué pour la recherche ;
- CF = les coûts fixes (locaux, frais de bureau, employés, etc.) ;
- CV = les coûts variables (interurbains, déplacements, poste, etc.) ;
- CME = le coût moyen d'une entrevue estimée à partir des expériences passées.

Donnons un exemple :

- Budget = 40 000 \$;
- CF = 12 000 \$
- CV = 5 000 \$;
- CME = 30 \$.

On aura donc :

$$n = \frac{40\,000\ \$ - 12\,000\ \$ - 5\,000\ \$}{30\ \$} = 767.$$

La taille de l'échantillon serait, dans ce cas, de 767 répondants. Nous le verrons plus loin : «[...] le choix d'un échantillon (effectif et structure) est toujours un compromis entre coût et précision¹⁸.»

Dans le tirage d'un échantillon, on doit toujours s'attendre à un écart, plus ou moins élevé, entre les résultats obtenus dans l'échantillon et les résultats réels dans la population mère. Ce décalage indique qu'il y a toujours une certaine probabilité (un degré de malchance) de ne pas retrouver toutes les caractéristiques de la population mère dans l'échantillon tiré. De la même façon, les projections des résultats de l'échantillon à l'ensemble de la population mère seront toujours soumises à une certaine imprécision.

17. D. Lehmann (1979), *Market Research and Analysis*, Illinois, Richard D. Irwin Inc., p. 259.
18. J. Antoine (1981), *Le sondage, outil du marketing*, Paris, Dunod, p. 179.

Cette imprécision (ou cette précision) dépendra des fluctuations d'échantillonnage. Ces fluctuations ont été étudiées par les statisticiens qui ont élaboré la «théorie des distributions d'échantillonnage». À partir de la loi des grands nombres : «le théorème d'or de Bemouilli montre en effet que, sous certaines conditions, la fréquence observée dans un échantillon tend vers sa probabilité théorique, et la moyenne arithmétique observée converge «en probabilité » vers l'espérance mathématique théorique¹⁹. » En d'autres mots, si le nombre des observations est assez grand, la fréquence observée s'éloignera très peu de la probabilité réelle de l'événement et l'écart entre la population mère et l'échantillon pourra être estimé assez facilement.

Plus concrètement, pour définir la taille de l'échantillon, nous avons besoin de trois choses :

1. Un niveau de confiance (Z) de l'échantillon : c'est la possibilité d'obtenir un échantillon dont les caractéristiques (socio-économiques, par exemple) seraient très différentes de la population mère.
2. Une marge d'erreur théorique (E) sur les réponses : c'est l'intervalle de confiance d'une réponse quelconque.
3. Une estimation de la réponse (Q) de la population à une question binaire.

Le niveau de confiance de l'échantillon oscille habituellement entre 95 et 99 % ; ce qui veut dire que la probabilité de se tromper sera de 5 % dans le premier cas et de 1 % dans le deuxième cas. Un niveau de confiance à 95 % signifie que dans 95 % des cas, les résultats obtenus seront contenus dans les intervalles de la marge d'erreur déjà établie. Ce niveau de confiance suppose que les résultats obtenus se situeront à $\pm 1,96$ écart type de la moyenne de l'échantillon (99 % correspondra à 2,58 écarts types de la moyenne).

La marge d'erreur théorique sur les réponses varie habituellement entre 1 et 10 % (10 % est une valeur limite par convention). Plus la marge d'erreur sera petite, plus la taille de l'échantillon sera grande, et inversement. Le troisième élément consiste à faire une hypothèse théorique à une question de type binaire (oui ou non) ; par exemple, à la question «Recommanderiez-vous cet hôtel à des parents ou amis ? » on suppose la pire situation possible : c'est-à-dire que les clients choisis répondront au

19. J. Rose (1993), *Le hasard au quotidien*, Paris, Éditions du Seuil, p. 168.

hasard (à «pile ou face », ce qui correspond au hasard absolu), à savoir que 50 % répondront oui et que l'autre 50 % répondront non.

La formule pour trouver la taille de l'échantillon sera la suivante²⁰ :

$$n = \frac{Z^2 P (1 - P)}{E^2} \quad \text{où :}$$

n = taille de l'échantillon ;

Z = niveau de confiance de l'échantillon (écart type) ;

E = marge d'erreur théorique sur les réponses ;

P = proportion estimée à une réponse binaire.

Si nous acceptons un niveau de confiance de 95 % (un écart type de 1,96) et une marge d'erreur sur les réponses de 5 %, nous appliquerons la formule suivante :

$$n = \frac{1,96^2 \times 0,50 (1 - 0,50)}{0,05^2} = 384 \text{ personnes.}$$

Nous devons donc choisir de façon aléatoire 384 personnes. Supposons aussi que les résultats du sondage indiquent que 65 % des répondants recommanderaient à des parents et amis notre hôtel. À ce moment, pour calculer la marge d'erreur, nous inversons la formule ; nous obtenons donc :

$$E = Z \sqrt{\frac{P (1 - P)}{n}}$$

Alors, l'erreur sera :

$$E = 1,96 \sqrt{\frac{0,65 \times (1 - 0,65)}{384}} = 0,047$$

ou 4,7 %.

Dans notre rapport final, nous pourrions écrire : si nous prenions 100 échantillons de 384 personnes, il n'y en aurait que 5 (ou 5 %) dont les résultats à une question binaire seraient en dehors de la fourchette $\pm 4,7$ %. En d'autres mots, nous avons 95 % des chances de ne pas nous tromper en avançant que les clients recommanderaient (le cas échéant) notre hôtel à

20. Voir I. Bradley et J. South (1994), *Introductory Statistics for Business and Economics*, Illinois, The Dryden Press, p. 217.

des parents et amis dans une proportion de 65 % \pm 4,7 % (le résultat réel se situerait entre 60,3 % et 69,7 %).

Nous allons maintenant à l'aide de cette formule étudier les cas limites.

Premier cas (la taille la plus faible) :

Z = 95 % ou 1,96 écart type ;

P = 0,50 ;

E = 10% (la probabilité la plus forte qu'il soit possible d'accepter).

La taille sera donc de :

$$n = \frac{1,96^2 \times 0,50 (1 - 0,50)}{0,10^2} = 96 \text{ personnes.}$$

Deuxième cas (la taille la plus élevée) :

Z = 95 % ou 1,96 écart type ; P = 0,50 ;

E = 1 % (la probabilité la plus forte acceptée).

La taille sera donc de :

$$n = \frac{1,96^2 \times 0,50 (1 - 0,50)}{0,01^2} = 9\ 604 \text{ personnes.}$$

Supposons que le coût moyen par entrevue soit de 30 \$; nous aurons dans le premier cas : $96 \times 30 \$ = 2\ 880 \$$ et dans le deuxième cas : $9\ 604 \times 30 \$ = 288\ 160 \$$. Nous pouvons constater ici qu'une précision très élevée coûte énormément cher. La situation moyenne sera de choisir une marge d'erreur entre 5 et 1 %. La majorité des sondages administrés sur l'ensemble du Québec adopteront une marge d'erreur de 3 % ; dans ce dernier cas, nous aurons :

$$n = \frac{1,96^2 \times 0,50 (1 - 0,50)}{0,03^2} = 1\ 067 \text{ personnes.}$$

Le coût sera donc de $1\ 067 \times 30 \$ = 32\ 010 \$$ (si nous conservons le même coût moyen par questionnaire).

Il est sage de prévoir, dans le tirage dans les listes, un nombre plus grand de répondants que le nombre obtenu par la formule. Par exemple, si nous prévoyons interroger 1 067 personnes, il faut faire le tirage en tenant compte des personnes malades, en voyage ou déménagées ; il faut donc

ajouter des personnes à la taille préalablement fixée. Ce facteur de correction peut être plus ou moins élevé dépendant de la qualité des listes mises à notre disposition. Dans tous les cas, il est prudent d'utiliser un facteur de correction d'au moins 25 %. Nous obtiendrons donc : $1\ 067 \times 1,25 = 1\ 334$; il faudra, par conséquent, «tirer» 1 334 personnes pour obtenir (plus ou moins) 1 067 questionnaires complétés.

Certaines firmes de recherche utilisent des tableaux résumant les calculs pour déterminer rapidement la taille de l'échantillon. Les résultats de ces calculs sont présentés au tableau 12.

Tableau 12

TABLEAU POUR DÉTERMINER LA TAILLE DE L'ÉCHANTILLON

Les tailles des échantillons avec divers niveaux de confiance et avec des résultats estimés de 0,50 (oui) et de 0,50 (non) à une question binaire et pour diverses marges d'erreurs sur les réponses.

Erreur acceptée	Taille des échantillons	
	Niveau de confiance de 95 %	Niveau de confiance de 99 %
1	9 604	16 641
2	2 401	4 260
3	1067	1 849
4	600	1 040
5	384	666
6	267	462
7	196	340
8	150	260
9	118	205
10	96	166

On peut constater, à l'examen des colonnes du tableau 12, que la précision est onéreuse. Il faut toujours se demander si l'on a besoin d'un niveau de précision élevé et, dans le choix de la taille de l'échantillon, revenir aux objectifs initiaux de la recherche et tenir compte du budget alloué pour cette étude.

Ces résultats sont valables uniquement si la population est relativement homogène et si les autres règles de l'échantillonnage sont strictement appliquées. Il arrive souvent, dans le cas des échantillons stratifiés, qu'il faille «gonfler» certaines strates pour pouvoir les comparer aux autres ; dans ce cas, la taille de l'échantillon peut facilement quadrupler.

En pratique, la taille de l'échantillon sera établie : «en fonction de quatre paramètres :

- le degré de précision escompté ;
- le degré de spécificité de l'étude ;
- le budget disponible ;
- la taille psychologique d'un échantillon²¹».

La dimension psychologique est importante pour rassurer les commanditaires de la recherche sur la valeur globale du sondage. En résumé, les procédures d'échantillonnage se résument à cinq étapes :

1. Définir le plus clairement possible la population visée et bien cerner ses principales caractéristiques, et cela, en fonction des hypothèses et des objectifs de l'enquête.
2. Établir une liste réelle ou virtuelle de la population mère ; cette liste servira de base de sondage.
3. Opérer une vérification minutieuse de la base de sondage : il s'agit de «nettoyer» en enlevant les personnes qui, pour une raison ou une autre, sont dans l'incapacité de répondre (décès, maladie, divorce, voyage, etc.).
4. Déterminer la taille de l'échantillon en tenant compte du budget disponible et du «confort psychologique » des commanditaires ; ajouter à cette taille 25 % de noms supplémentaires qui serviront à atténuer les insuffisances de la liste utilisée.
5. Établir une méthode de tirage aléatoire (systématique, stratifiée, etc.) et, à l'aide de cette méthode, procéder au «tirage» des personnes qui feront partie de l'échantillon.

LA COLLECTE DES DONNÉES

Dans la collecte des données, il y a un certain nombre d'étapes importantes à suivre :

1. le choix de la période de collecte des informations ;
2. la préparation immédiate de la collecte ;
3. les niveaux de contrôle de la collecte des données.

21. O. Badot (1990), *L'étude marketing dans une économie de tohu-bohu*, Paris, ESF Éditeur, p. 98.

Pour ce qui est de la première étape, il faut (en général) bien choisir la période dans l'année où l'on administrera l'enquête : il faut éliminer les mois de l'été (entre le 20 juin et le 10 septembre), la période des fêtes de Noël et du jour de l'An ainsi que les mini-vacances, (Pâques, etc.). Dans la plupart des cas, il faut aussi tenir compte des jours de la semaine : par exemple, du lundi au jeudi. De même, il est judicieux de choisir des plages horaires dans la journée : les entrevues téléphoniques se font habituellement entre 17 h 30 et 20 h 30.

Toute enquête requiert une certaine préparation psychologique des répondants ; des avis d'une enquête prochaine peuvent être publiés dans les journaux locaux, les revues professionnelles, les journaux d'entreprise ou annoncés par d'autres médias. Ces avis vont limiter l'effet de surprise et augmenter la tolérance des répondants. Pour les entrevues face à face, il est bon d'envoyer une lettre au préalable et, par la suite, de convenir d'un rendez-vous par téléphone. Ces démarches exigent beaucoup de temps et de doigté, elles doivent être adaptées à l'objet de l'étude et aux situations particulières. Par exemple, dans un hôtel, une affichette et une lettre dans la chambre peuvent prévenir les clients qu'une enquête est en cours.

Le dernier élément à considérer, c'est le contrôle exercé sur la collecte elle-même. Ce contrôle doit s'effectuer à plusieurs niveaux :

1. Au niveau du questionnaire, selon le mode d'administration, par des prétests réels et approfondis.
2. Par le contrôle des entrevues « sur le terrain » ; le directeur d'enquête doit vérifier de façon systématique, pour chacun des intervieweurs, si les entrevues ont bien été réalisées et de quelle façon. Une recherche sans contrôle de terrain peut entraîner une contestation sérieuse des résultats obtenus.
3. Pour chacune des personnes choisies, selon le mode d'administration du questionnaire, on doit prévoir un contrôle du taux de réponse. Ce contrôle va s'exercer par une fiche-personne comprenant les rubriques suivantes (elles ne sont pas exhaustives) :
 1. questionnaire complété ;
 2. refus de répondre ;
 3. la personne choisie ne parle ni le français, ni l'anglais ;
 4. moins de 18 ans ;
 5. la personne choisie est dans l'incapacité de répondre (malade, sénile, etc.).

D'autres catégories pourraient être ajoutées pour un sondage téléphonique. Nous verrons le calcul du taux de réponse dans le chapitre 4 intitulé «Le traitement des données ».

4. Enfin, par le contrôle de la codification des données ; il faut s'assurer que la codification respecte les règles usuelles et répond aux besoins de la recherche.

4

Chapitre

Le traitement des données

Le traitement des données est l'étape la plus intéressante d'une recherche empirique ; c'est à ce moment que l'on peut enfin considérer les résultats obtenus et voir si l'on a bien atteint ses objectifs de départ. Nous allons voir, dans ce chapitre, quatre niveaux dans le traitement des données. Dans le premier niveau, il s'agit de préparer les données brutes pour l'analyse statistique. Le deuxième niveau consiste en une étude univariée des questions du questionnaire. Dans le troisième niveau, on observe les relations entre les principales variables de la recherche : c'est l'analyse bivariée. Le quatrième niveau consiste à approfondir

l'étude des relations entre variables en utilisant des techniques statistiques plus complexes : c'est l'analyse multivariée.

1 LA PRÉPARATION DES DONNÉES

Après la collecte des données sur le terrain, plusieurs étapes doivent être franchies avant de pouvoir faire une analyse statistique complète des résultats recueillis. Ces étapes sont les suivantes :

- acceptation ou rejet des questionnaires ;
- choix d'un logiciel et entrée des données ;
- calcul de la marge d'erreur a posteriori ;
- calcul du taux de réponse ;
- étude des principales variables socio-économiques et, s'il y a lieu, redressement de l'échantillon.

La première étape de la préparation des données consiste à vérifier si chacun des questionnaires est bien complété, c'est-à-dire si le répondant a répondu à toutes les questions. L'approche la plus rigoureuse est de rejeter tous les questionnaires incomplets. Le responsable de l'enquête peut aussi établir une règle d'acceptation en fonction du nombre de questions non répondues dans le questionnaire ; ce pourcentage peut aller de 10 à 25 % du nombre total des questions. Le critère d'acceptation peut aussi être établi en fonction des questions jugées vitales par les chercheurs (le «cœur» du questionnaire). En outre, le nombre de rejets peut dépendre de la taille de la population échantillonnée ; si la taille de l'échantillon est petite, la décision de rejeter plusieurs questionnaires pourrait mettre en péril le caractère représentatif de l'échantillon.

Pour la deuxième étape, il s'agit de choisir un logiciel et de créer un fichier pour l'entrée des données. Une fois les données recueillies, il faut les rendre «digestibles» par l'ordinateur ; elles doivent être traduites en langage informatique : c'est l'étape de la codification. On doit établir un code de réponse question par question ; par exemple, à la question «Quel est votre sexe ? », le sexe féminin aura le code 1 et le sexe masculin, le code 2, et l'on inscrira le code 0 (ou 9) si la personne interrogée refuse ou oublie de répondre à cette question. De cette façon, chacune des possibilités de répondre, pour chacune des questions, aura été prévue et aura son code numérique propre. C'est sous cette forme codifiée et chiffrée que les résultats des entrevues seront introduits dans la mémoire de l'ordinateur.

Le fichier des données sera rédigé en suivant le protocole de codification. Normalement, l'entrée des réponses du questionnaire dans le fichier des données devrait suivre l'ordre d'apparition des questions dans

le questionnaire ; de plus, les codes des réponses devraient être les mêmes que ceux établis dans la rédaction du questionnaire¹.

Une fois les données mémorisées par l'ordinateur, on doit envisager un plan de traitement informatique permettant de créer des tableaux statistiques standardisés (uniformes). Ce plan de traitement suppose que l'on pourra créer des tableaux de fréquence ainsi que des tableaux «croisés» (mise en relation de deux variables ou plus) et utiliser la plupart des techniques statistiques usuelles. À l'heure actuelle, ce travail est grandement facilité grâce à des logiciels simples qui permettent un traitement informatique et statistique des données avec le minimum de programmation.

Deux de ces logiciels dominent le marché : le Statistical Package for the Social Sciences (SPSS) et le Statistical Analysis System (SAS) ; ils fonctionnent sur de gros ordinateurs centraux et sur des micro-ordinateurs. Ces logiciels sont formés de programmes servant à traiter, à transformer, à lire, à éditer et à analyser des données d'enquêtes statistiques. Ce sont des outils inestimables en raison de leur facilité d'apprentissage ; leur rapidité ; leur précision.

Grâce à eux, nous pouvons créer des dizaines de tableaux et analyser en profondeur les relations entre les diverses variables de l'enquête.

Après l'entrée des données dans un fichier informatique, il devient possible de calculer la marge d'erreur théorique, a posteriori, sur les réponses.

LA MARGE D'ERREUR

La marge d'erreur, sur une proportion pour une question donnée de l'échantillon, sera établie à partir de la formule suivante :

$$E = P \pm Z \sqrt{\frac{PQ}{n}} ;$$

où :

E = marge d'erreur ;

P = proportion observée dans l'échantillon ;

Z = nombre d'écarts types (pour un seuil de confiance de 95 %) ;

Q = 1 - P ;

n = taille de l'échantillon.

1. À ce sujet voir le chapitre III, partie 3 : la codification des questions.

Donnons un exemple :

- 15 % des répondants à un sondage achèterait le forfait XYZ (P) ;
- le seuil de confiance est de 95 % ou de 1,96 écart type (Z) ;
- la taille de l'échantillon est de 1 500 personnes (n).

L'application de la formule se fera de la façon suivante :

$$E = 0,15 \pm 1,96 \sqrt{\frac{(0,15)(1 - 0,15)}{1\,500}}$$

$$E = 0,15 \pm 1,96 (0,0092)$$

$$E = 0,15 \pm 0,018$$

L'estimation de la proportion réelle des acheteurs dans la population mère (la population globale) oscillera donc entre 0,13 (0,15 - 0,018 ou 13 %) et 0,17 (0,15 + 0,018 ou 17 %).

Pour lire ces résultats, on procédera de la façon suivante : si l'on prenait 100 échantillons de 1 500 répondants, il y en aurait au plus cinq dont les résultats (0,15 d'acheteurs potentiels du forfait XYZ) seraient à l'extérieur de la fourchette 0,15 ± 0,02.

LE TAUX DE RÉPONSE

À ce moment de la recherche, il est utile de calculer le taux de réponse de l'ensemble de l'échantillon ; si ce taux est inférieur à 50%, on peut mettre en doute le caractère représentatif de l'échantillon. Un taux de réponse faible oblige les chercheurs à effectuer un nouveau tirage et à rejoindre un plus grand nombre de répondants ! Le taux de réponse se calcule à partir des rubriques utilisées lors du contrôle de la collecte des données.

Le calcul variera selon le nombre de rubriques et celui-ci sera différent selon le mode d'administration du questionnaire. Ces différentes possibilités apparaissent dans le tableau 13.

Nous remarquons que le sondage téléphonique comprend plus de rubriques que les autres modes d'administration. Chacune des personnes sélectionnées pour faire partie de l'échantillon doit avoir une fiche-réponse : cette fiche comprendra les rubriques décrites au tableau 13, selon le mode choisi d'administration de questionnaire.

Tableau 13**RUBRIQUES POUR LE CALCUL DU TAUX DE RÉPONSE
SELON LE MODE D'ADMINISTRATION DU QUESTIONNAIRE**

Mode d'administration	Rubriques utilisées
Entrevue par téléphone	1. Questionnaire complété. 2. Pas de service à ce numéro. 3. Numéro commercial ou institutionnel. 4. Problème de langue. 5. Absence prolongée hors du Québec. 6. Incapacité de répondre (âge, maladie). 7. Numéro de télécopieur, boîte vocale, cellulaire, pagette. 8. Pas de réponse après 5 appels. 9. Refus de répondre.
Envoi postal	1. Questionnaire complété. 2. Retour du questionnaire non complété pour cause : <ul style="list-style-type: none"> • absence prolongée ou • problème de langue ou • maladie ou décès. 3. Retour du questionnaire pour «mauvaise adresse» ou déménagement. 4. Questionnaire non retourné après une lettre et trois appels (refus de répondre).
Autres modes	1. Questionnaire complété. 2. Problème de langue. 3. Incapacité de répondre (sénilité ou maladie). 4. Refus de répondre.

Le calcul du taux de réponse se fera à partir de la formule suivante :

$$\text{Taux de réponse} : \frac{\text{NQC}}{\text{NUO} - \text{UI}} \times 100$$

où :

NQC = nombre de questionnaires complétés ;

NUO = nombre d'unités (personnes) d'observation initialement tirées pour faire partie de l'échantillon ;

UI = les unités d'observations inadmissibles.

Voici un exemple : supposons un sondage téléphonique portant sur 500 répondants (unités initialement choisies dans la population globale). Par le biais des fiches-réponses, nous arrivons au décompte suivant :

Rubriques	Nombre	%
1. Nombre de questionnaires complétés.	800	53,3
2. Pas de service à ce numéro.	40	2,7
3. Numéro commercial ou institutionnel.	220	14,7
4. Problème de langue.	60	4,0
5. Absence prolongée hors du Québec.	10	0,6
6. Incapacité de répondre.	12	0,8
7. Numéro de télécopieur, etc.	64	4,3
8. Pas de réponse après 5 appels.	144	9,6
9. Refus de répondre.	150	10,0
	1 500	100

Appliquons la formule de tout à l'heure et nous aurons :

$$\frac{800}{1\ 500 - 406} \times 100 = 73,1 \%$$

Ici, le chiffre 800 correspond au nombre de questionnaires complétés et 1 500 nous ramène à la taille initiale de l'échantillon ; le nombre 406 est obtenu par l'addition des rubriques 2 à 7. Cette formule suppose que les unités inadmissibles (les rubriques 2 à 7) sont des erreurs qui ne dépendent pas de l'équipe de recherche mais représentent plutôt des problèmes intrinsèques à toutes les recherches empiriques. Le taux de réponse doit être égal ou supérieur à 50 % ; idéalement, il doit être le plus élevé possible. Normalement, un taux inférieur à 50 % remet en cause le caractère représentatif de l'échantillon ; à ce moment-là l'équipe de recherche doit effectuer un nouveau tirage et retourner sur le terrain pour augmenter sensiblement le nombre des questionnaires complétés.

En partant des résultats obtenus, il est possible aussi de calculer un facteur de corrections ; celui-ci est obtenu en divisant le nombre d'unités d'observation initialement tirées par le nombre total de questionnaires complétés. Donc, ici on aura :

$$\frac{1\ 500}{800} = 1,875$$

Ce résultat indique que si l'on veut obtenir (la prochaine fois) 1 500 questionnaires complétés, l'échantillon initial devra comprendre 2 812 numéros de téléphone (ou $1\ 500 \times 1,875 = 2\ 812$).

LE REDRESSEMENT DE L'ÉCHANTILLON

Si le taux de réponse est satisfaisant, nous pouvons faire une première analyse sommaire des données. Cette analyse devrait porter sur les variables qui permettent de cerner le profil socio-économique des répondants. L'objectif de cette préanalyse des données est d'examiner si, pour ces variables, la structure de l'échantillon correspond à celle observée dans la population totale (quand cela est possible). Si la structure de l'échantillon, pour l'une ou plusieurs variables, diffère de celle observée dans la population globale, il faudra opérer un redressement de l'échantillon. D'après Max Deroo : «Redresser un échantillon, c'est au moment de l'élaboration des résultats chiffrés (dépouillement), utiliser une information supplémentaire pour améliorer la qualité des estimations, c'est-à-dire pour réduire l'erreur d'échantillonnage².»

Pour redresser un échantillon, il faut comparer certaines variables de la population globale avec ces mêmes variables dans l'échantillon. Il existe deux grandes techniques de pondération : la première est dite «mécanique », car il s'agit de supprimer ou de reproduire des questionnaires ; la deuxième consiste à calculer un coefficient de pondération applicable aux résultats bruts de l'enquête. Voyons la première technique avec l'exemple qui apparaît dans le tableau 14.

Tableau 14
REDRESSEMENT D'ÉCHANTILLON AVEC LA TECHNIQUE
«MÉCANIQUE»

Variable Âge	Pourcentage de la population globale	l'aille de l'échantillon		Questionnaires	
		Théorique	Réelle	Manquants	En trop
30 ans et moins	25	125	110	15	
31-40 ans	30	150	180		30
41-50 ans	27	135	120	15	
51-60 ans	10	50	40	10	
61 ans et plus	8	40	30	10	
Total	100	500	480	50	30

2. M. Deroo et A.-M. Dussaix (1980), *Pratique et analyse des enquêtes par sondage*, Paris, PUF, p. 127.

Dans la première colonne apparaît la variable de contrôle : l'âge et ses diverses strates définies à l'avance. La deuxième colonne montre le pourcentage, pour chacune des strates, dans la population étudiée. La troisième colonne présente la taille théorique que l'on devrait obtenir et la quatrième, la taille réelle de l'échantillon. Les deux dernières colonnes indiquent les questionnaires qui manquent et ceux qui sont en trop dans certaines strates.

Pour les questionnaires manquants, il s'agit, par exemple, dans la strate d'âge 30 ans et moins, de choisir au hasard systématique 15 questionnaires parmi les 110 questionnaires complétés et de les reproduire intégralement. Pour la strate 31 à 40 ans, nous opérons de la même manière (tirage au hasard), mais ces 30 questionnaires seront détruits ! Dans l'exemple du tableau 13, il serait indiqué de regrouper les deux dernières strates, car dans le cas des 61 ans et plus, on devrait reproduire un questionnaire sur trois, ce qui est une proportion assez élevée. Normalement, la limite est de reproduire au maximum 15 à 20 % de l'échantillon initial ; une proportion plus élevée pourrait entraîner des biais importants.

Si l'on veut conserver les 30 questionnaires en trop (dans la strate 31-40), il faut gonfler la taille de l'échantillon pondéré à 600 personnes. De cette façon, pour la strate 30 ans et moins, le total pondéré devra être de 150 ($600 \times 25\%$) ; il faudra donc doubler 40 questionnaires au lieu de 15 (voir le tableau 14). Le risque ici est de reproduire un trop grand nombre de questionnaires et de dépasser les limites permises. Cette approche serait valable avec un échantillon initial plus élevé (2 000 personnes et plus).

La deuxième technique, celle des coefficients de pondération, peut s'exercer sur une seule variable ou sur plusieurs variables de l'enquête.

La formule du coefficient de pondération sera la suivante :

$$\text{Coefficient de pondération} = \frac{\text{Taille théorique de l'échantillon}}{\text{Taille réelle de l'échantillon}}$$

La taille théorique est trouvée grâce à des informations connues sur la population totale à étudier. Par exemple, dans les données du tableau 14, les 30 ans et moins forment 25 % de la population globale (par exemple, les clients d'une agence de voyages ; ces informations doivent être connues). Le coefficient de pondération sera donc, pour cette strate d'âge, le suivant :

$$\text{Coefficient de pondération} = \frac{125}{110} = 1,14.$$

Ce coefficient sera multiplié aux résultats obtenus à une autre question ; par exemple, si à la question «Êtes-vous propriétaire d'une résidence secondaire ? », on obtient :

avant la pondération

pour les 31 ans et moins = 15 répondants disent «oui », soit 13,6 % ;

après la pondération

pour les 31 ans et moins = $15 \times 1,14 = 17$ disent « oui », soit 15,4 %.

avant la pondération

pour les 31-40 ans = 45 répondants disent « oui », soit 25 % ;

après la pondération

pour les 31-40 ans = $45 \times \frac{150}{180} = 37$ disent « oui », soit 20,5 %.

Comme on le voit ici, le redressement de l'échantillon permet de corriger certaines lacunes dans l'échantillonnage et d'accroître le caractère représentatif de l'échantillon. Il existe des logiciels qui vont effectuer automatiquement la pondération à partir des variables choisies et des coefficients calculés ; dans les tableaux croisés où ces variables seront impliquées ces pondérations vont aussi s'exercer de façon mécanique. Avec le logiciel de traitement de données le plus connu (SPSS), les commandes « compute » et « poids » permettent d'appliquer facilement les coefficients de pondération.

2 L'ANALYSE UNIVARIÉE

L'analyse univariée consiste à produire et à interpréter un tableau pour chacune des variables de la recherche. Selon Jean-Pierre Védérine : «Les analyses univariées concernent le traitement d'une seule variable. Elles ont pour objet de synthétiser et d'analyser les informations disponibles sur cette variable³. » Un tableau univarié comprendra certains éléments tels que :

- les modalités de la variable (les codes prévus pour les réponses possibles à cette question) ;

3. J.-P. Védérine (1991), *Le traitement des données en marketing*, Paris, Les Éditions d'Organisation, p. 25.

- les effectifs pour chacune des modalités de réponse ;
 - les fréquences en pourcentage ;
 - les fréquences cumulées.
- Donnons un exemple.

Tableau 15***DESTINATION DE VACANCES DES QUÉBÉCOIS EN 1995***

Destinations	Effectifs	Pourcentage (%)	Pourcentage cumulé
Québec	333	51,2	51,2
Canada	107	16,5	67,7
États-Unis	97	14,9	82,6
Europe	49	7,5	90,1
Autres	64	9,8	100
TOTAL	650	100	

Dans le tableau 15, nous voyons que la question sur les destinations de vacances comprend cinq choix et pour chacun d'eux, nous avons les réponses des personnes interrogées : par exemple, 333 personnes choisissent le Québec comme destination de vacances, ce qui correspond à 51,2% de l'ensemble des répondants. La fréquence cumulée permet d'associer certains choix de réponse : ici, 67,7 % des répondants vont séjourner soit au Québec, soit au Canada.

Cette étape permet aussi de transformer certaines variables, en réduisant les catégories afin de faciliter les analyses subséquentes (surtout pour les tableaux croisés). Voyons un exemple.

Tableau 16**NOMBRE DE SEMAINES DE VACANCES AU COURS
DES SIX DERNIERS MOIS**

Nombre de semaines	Effectifs	Pourcentage	Pourcentage ajusté	Pourcentage cumulé
0	2	0,2	0,4	0,4
1	32	2,7	6,0	6,4
2	127	10,8	23,7	30,1
3	94	8,0	17,6	47,7
4	145	12,4	27,1	74,8
5	59	5,0	11,0	85,8
6	27	2,3	5,0	90,8
7	4	0,3	0,7	91,6
8	19	1,6	3,6	95,1
9	3	0,3	0,6	95,7
10	6	0,5	1,1	96,8
11	1	0,1	0,2	97,0
12	8	0,7	1,5	98,5
16	1	0,1	0,2	98,7
Ne sais pas	7	0,6	1,3	100
Ne s'applique pas	636	54,3	—	
Total	1 171	100	100	

Moyenne : 5,024

Médiane : 4,0

Mode : 4,0

Écart type : 11,036

Écart type de la population : 0,477

Dans le tableau 16, la quatrième colonne nous donne le pourcentage ajusté ; ce pourcentage ne tient pas compte de la mention «Ne s'applique pas ». Les nouveaux pourcentages seront donc calculés à partir des 535 personnes qui ont eu droit à des vacances au cours des six derniers mois. Sous le tableau apparaissent quelques statistiques usuelles fournies par le logiciel statistique SPSS. Nous voyons que la moyenne, le mode et la médiane ont des valeurs très proches : ce qui indique que notre distribution des fréquences est quasi normale⁴.

4. Pour mesurer de façon détaillée la «normalité» de la distribution, il faudrait faire appel au coefficient d'asymétrie de Yule et au coefficient d'aplatissement (Kurtosis) ; voir à ce sujet J. Blard-Laborde (1994), *Initiation à la statistique descriptive*, Paris, Les Éditions d'Organisation, p. 93 à 95.

L'écart type de la population n'est qu'une estimation de celui-ci ; il est calculé par la formule $\frac{\sigma}{\sqrt{n}}$. Ici, nous aurons donc :

$$\frac{11,036}{\sqrt{535}} = 0,447.$$

Cette statistique nous permet d'estimer les intervalles de confiance de la moyenne observée par rapport à la population totale. Avec un niveau de confiance de 95 % (1,96 d'écart type), la moyenne estimée de la population mère sera de 5,024 semaines (moyenne de l'échantillon) \pm 1,96 (0,477), soit 5,024 \pm 0,93492. La moyenne réelle du nombre de semaines de vacances pour l'ensemble du Québec se situera donc dans la fourchette des 5,95892 semaines et 4,08908 semaines.

Par la suite, il faudra (grâce à la commande «recode» de SPSS) refaire les catégories du tableau 16 en tenant compte du pourcentage cumulé. Une des possibilités de «recodification» pourrait être la suivante :

Tableau 17

NOMBRE DE SEMAINES DE VACANCES AU COURS DES SIX DERNIERS MOIS, EN POURCENTAGE

Nombre de semaines	Pourcentage
Deux semaines et moins	23,7
Trois semaines	17,6
Quatre semaines	27,1
Cinq semaines et plus	31,6
TOTAL	100

Le tableau 17 permettra de croiser la variable « semaines de vacances » avec des variables socio-économiques (par exemple, les revenus) ou avec d'autres types de variables.

L'étude univariée, en plus de permettre de reformuler les variables en réduisant les catégories (en passant du tableau 16 au tableau 17), sert à éliminer les catégories qui deviendront inutiles pour les analyses bivariées et multivariées. Ces catégories sont les mentions «sans réponse », «ne sais pas », «ne s'applique pas », etc. Nous pouvons aussi être amené à réduire les catégories des variables ordinales telles que les revenus ou

l'âge lorsqu'il y a trop peu de répondants dans certaines catégories de ces variables.

L'analyse univariée peut aussi nous permettre de créer des variables qui peuvent nous aider à en synthétiser d'autres ; ces variables sont appelées des « scores ». Le score correspond, pour chacun des répondants, à l'addition d'un certain nombre de réponses ; celles-ci étant codées de façon numérique (0, 1, 2, 3, 4, 5, n), il est possible, en les agrégeant, d'obtenir un chiffre global qui devient le score du répondant.

Donnons l'exemple d'un score de satisfaction construit à l'aide des variables du tableau 18 ; ici, le score sera établi par l'addition de ces cinq variables de l'enquête.

Tableau 18

LE SCORE DE SATISFACTION

Satisfaction	Très insatisfait	Insatisfait	Plus ou moins satisfait	Satisfait	Très satisfait
1. Accueil	1	2	3	4	5
2. Hébergement	1	2	3	4	5
3. Restauration	1	2	3	4	5
4. Animation	1	2	3	4	5
5. Ambiance générale	1	2	3	4	5

Un répondant très satisfait pour l'ensemble des aspects mesurés aura le score 25 ; en revanche, un répondant très insatisfait obtiendra la note 5. L'ensemble des personnes interrogées se distribuera dans une fourchette allant de 5 (minimum) à 25 (maximum). Une fois ces calculs exécutés (à l'aide de la commande «compute» de SPSS), on peut construire une typologie des diverses clientèles en fonction des scores obtenus ; ainsi, on peut diviser ce score en cinq groupes quasi égaux :

1. groupe 1 : très insatisfaits, note 5 à 8 ;
2. groupe 2 : insatisfaits, note 9 à 12 ;
3. groupe 3 : plus ou moins satisfaits, note 13 à 16 ;
4. groupe 4 : satisfaits, note 17 à 20 ;
5. groupe 5 : très satisfaits, note 21 à 25.

On peut, par la suite, donner un qualificatif quelconque à chacun de ces groupes ; ainsi :

1. le groupe 1 devient : « les hostiles » ;
2. le groupe 2 devient : « les mécontents » ;
3. le groupe 3 devient : « les tièdes » ;
4. le groupe 4 devient : « les rassasiés » ;
5. le groupe 5 devient : « les comblés ».

L'analyse univariée consiste donc à scruter chacune des variables, à la fois pour en faire l'analyse (nous reviendrons sur cet aspect au chapitre 5), mais aussi pour transformer certaines variables (en réduisant les catégories des réponses) et en créer d'autres : les scores. Ces étapes serviront à préparer les analyses bivariées et multivariées que nous ferons plus tard.

3 L'ANALYSE BIVARIÉE

L'analyse bivariée consiste à croiser deux variables de la recherche. Les tableaux croisés prennent, en général, la forme suivante (tableau 19)⁵ :

Tableau 19

FORME GÉNÉRALE DES TABLEAUX CROISÉS

Modalités de x	Modalités de y	y1	y2	yn	total en %
x1		%	%	%	100
x2		%	%	%	100
xn		%	%	%	100

Il s'agit, dans un même espace, de rapporter deux variables, une variable indépendante x et une variable dépendante y ; on fait habituellement l'hypothèse que la variable x détermine, jusqu'à un certain point, la variable y . Dans les sciences sociales, il ne peut s'agir que d'une « implication faible⁶ » d'une variable sur l'autre. Les variables indépendantes

5. Nous nous inspirons ici d'un exemple de B. Py (1987), *Statistique descriptive*, Paris, Économica, p. 158.

6. R. Boudon (1971), *Les mathématiques en sociologie*, Paris, PUF, p. 27-36.

sont la plupart du temps des données factuelles qui définissent le profil du consommateur : sexe, âge, scolarité, profession, revenu, habitat, etc.

Selon François de Singly : «La variable indépendante doit être disposée en lignes et la variable dépendante en colonne. Les pourcentages doivent être calculés en fonction de la variable indépendante ou de celle qui joue ce rôle⁷. » Il faut en tenir compte dans la création des tableaux bivariés avec un logiciel statistique.

Voyons un exemple concret de tableau bivarié créé avec le logiciel SPSS⁸.

Tableau 20
EXEMPLE D'UN TABLEAU CROISÉ ;
LA VARIABLE SCOLARITÉ AVEC LA VARIABLE
VACANCES DANS LES TROIS DERNIÈRES ANNÉES

			Vacances dans les trois dernières années		
			Oui	Non	Total
Les années de scolarité	Primaire	Count	25	55	80
		% within Les années de scolarité	31,3%	68,8%	100,0%
		% within Vacances dans les trois dernières années.	3,1 %	15,3 %	6,8 %
		% of Total	2,1 %	4,7 %	6,8 %
	Secondaire	Count	182	143	325
		% within Les années de scolarité	56,0 %	44,0 %	100,0 %
		% within Vacances dans les trois dernières années.	22,5 %	39,7 %	27,8 %
		% of Total	15,6 %	12,2 %	27,8 %
12-13 années		Count	243	77	320
		% within Les années de scolarité	75,9 %	24,1 %	100,0 %
		% within Vacances dans les trois dernières années.	30,0 %	21,4 %	27,4 %
		% of Total	20,8 %	6,6 %	27,4 %
14-17 années		Count	257	65	322
		% within Les années de scolarité	79,8 %	20,2 %	100,0 %
		% within Vacances dans les trois dernières années.	31,7 %	18,1 %	27,5 %
		% of Total	22,0 %	5,6 %	27,5 %
18 années et plus		Count	103	20	123
		% within Les années de scolarité	83,7 %	16,3 %	100,0 %
		% within Vacances dans les trois dernières années.	12,7 %	5,6 %	10,5 %
		% of Total	8,8 %	1,7 %	10,5 %
Total		Count	810	360	1170
		% within Les années de scolarité	69,2 %	30,8 %	100,0 %
		% within Vacances dans les trois dernières années.	100,0 %	100,0 %	100,0 %
		% of Total	69,2 %	30,8 %	100,0 %

7. F. de Singly (1992), *L'enquête et ses méthodes : le questionnaire*, Paris, Nathan-Université, p. 97.

8. Statistical Package for the Social Sciences.

Tableau 20 (Suite)**EXEMPLE D'UN TABLEAU CROISÉ ;
LA VARIABLE SCOLARITÉ AVEC LA VARIABLE
VACANCES DANS LES TROIS DERNIÈRES ANNÉES**

Chi-Square Tests			
	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	116,7251	4	,000
Likelihood Ratio	112,836	4	,000
Linear-by-linear Association	97,209	1	,000
N of Valid Cases	1170		
Symmetric Measures			
		Value	Approx. Sig
Nominal by Nominal	Phi	,316	,000
	Cramer's V	,316	,000
	Contingency Coefficient	,301	,000
N of Valid Cases		1170	

1. 0 cells (.0 %) have expected count less than 5. The minimum expected count is 24,62.

Dans le tableau 20, la variable «Années de scolarité» est la variable indépendante et la variable «Vacances au cours des trois dernières années » est la variable dépendante. Ici, les observations portent sur deux variables ; à ce moment, on distingue habituellement la distribution conjointe, les distributions marginales et les distributions conditionnelles⁹.

9. Voir à ce sujet S. Alalouf, D. Labelle et J. Ménard (1990), *Introduction à la statistique appliquée*, Montréal, Éditions Addison-Wesley, p. 70-73 ; B. Goldfarb et C. Pardoux (1993), *Introduction à la méthode statistique*, Paris, Dunod, chapitre 3 ; A. Gilles (1994), *Éléments de méthodologie et d'analyse statistique pour les sciences sociales*, Saint-Laurent, McGraw-Hill Éditeurs, p. 227-236.

La distribution conjointe se lit de la façon suivante : dans chacune des cellules, nous trouvons les observations qui sont conjointes à la variable dépendante et à la variable indépendante. Si nous considérons les données du tableau 20, nous avons dans la première cellule du tableau les éléments suivants :

Tableau 21

***DISTRIBUTION CONJOINTE DES DONNÉES
(LECTURE/EXPLICATION DU TABLEAU 20)***

	Oui, départ en vacances au cours des trois dernières années	Les éléments de la première cellule de gauche du tableau 20
Niveau de scolarité primaire	25	Nombre absolu : ce sont les 25 personnes, avec une scolarité de niveau primaire qui ont pris des vacances au cours des trois dernières années.
	31,3	Ce pourcentage correspond à une lecture en ligne (horizontale) ; le calcul : $25/80 \times 100$. Le pourcentage de gens de niveau primaire qui ont pris des vacances.
	3,1	Ce pourcentage correspond à une lecture en colonne (verticale) ; le calcul : $25/810 \times 100$. Parmi tous ceux qui ont pris des vacances au cours des trois dernières années, 3,1 avait une scolarité de niveau primaire.
	2,1	Ce pourcentage correspond aux personnes qui ont pris des vacances et qui ont une scolarité primaire par rapport à l'ensemble des répondants de la recherche ; le calcul : $25/1170 \times 100$.

Les distributions marginales (dans le tableau 20) sont les marges (les totaux) qui apparaissent pour la variable x et pour la variable y . Ainsi, les valeurs marginales pour la variable indépendante (le niveau de scolarité) sera :

Tableau 22**DISTRIBUTIONS MARGINALES DE LA VARIABLE
«ANNÉES DE SCOLARITÉ»**

Niveaux	Primaire	Secondaire	12-13 années	14-17 années	18 années et plus	Total
Valeur absolue	80	325	320	322	123	1 170
Pourcentage	6,8	27,8	27,4	27,5	10,5	100

De la même façon, les valeurs marginales pour la variable dépendante (vacances prises au cours des trois dernières années) seront :

Tableau 23**LES DISTRIBUTIONS MARGINALES DE LA
VARIABLE «VACANCES DANS LES TROIS
DERNIÈRES ANNÉES»**

Réponse	Oui	Non	Total
Valeur absolue	810	360	1 170
Pourcentage	69,2	30,8	100

Les distributions conditionnelles correspondent à un premier niveau (minimal) d'interprétation statistique des données. Par exemple, pour le tableau 20, les distributions conditionnelles de la variable «vacances dans les trois dernières années » correspondent à la distribution conditionnelle de cette variable pour chacun des niveaux de la variable indépendante «années de scolarité » (tableau 22).

Inversement, les distributions conditionnelles de la variable «années de scolarité » correspondront, pour chacun des niveaux de scolarité, aux deux possibilités de la variable «vacances dans les trois dernières années », soit oui ou non (tableau 23). C'est une première étape pour avancer des hypothèses de relation entre les deux variables, c'est-à-dire évaluer la force des relations entre ces variables. C'est une démarche intuitive qui servira à préparer l'étude des relations à partir de tests statistiques.

L'analyse bivariée consiste aussi à évaluer l'influence d'une variable (la variable indépendante) sur une autre variable (la variable dépendante). L'objectif est de donner une certaine explication aux variations observées dans les tableaux croisés ; en d'autres mots, selon

P.E. Green et D.S. Tull¹⁰, nous cherchons par une formule (ou une fonction) à «déterminer le degré global d'association» entre les deux variables choisies ainsi qu'à définir le «degré de confiance» que l'on peut atteindre sur le plan statistique. Des tests statistiques vont permettre d'évaluer cette dépendance.

LE TEST DU KHI CARRE

Dans un premier temps, le principal test à utiliser est le test du khi carré (ou khi deux). Ce test a pour principal mérite de pouvoir être appliqué à plusieurs types de variables (nominale, ordinale, par intervalle ou un mélange de celles-ci) ; c'est donc un test universel. Le test du khi carré est un test «d'indépendance» ; ainsi, on dira : «que les deux variables x et y sont indépendantes en probabilité si : $\text{Prob}(x = i, y = j) = \text{Prob}(x = i) \times \text{Prob}(y = j)$ pour tout couple (i, j)»¹¹. Ce qui signifie que l'on compare les résultats observés dans l'enquête à une fréquence théorique ; si la fréquence observée et la fréquence théorique coïncident, on doit rejeter l'hypothèse d'un lien statistique entre les deux variables.

La formule, pour le calcul du khi carré est la suivante¹² :

$$\chi^2 = \sum \frac{(\text{effectifs observés} - \text{effectifs théoriques})^2}{\text{effectifs théoriques}}$$

Les effectifs théoriques sont calculés de la façon suivante :

$$\text{Effectif théorique pour une cellule donnée} = \frac{\text{total colonne} \times \text{total rangée}}{\text{grand total}}$$

Il faut aussi considérer le degré d'ouverture du tableau étudié : les «degrés de liberté» ; ceux-ci se calculent par la formule :

10. P.E. Green et D.S. Tull (1974), *Recherche et décisions en marketing*, Grenoble, Presses universitaires de Grenoble, p. 347.

11. M. Tenenhaus (1994), *Méthodes statistiques en gestion*, Paris, Dunod-Entreprise, p. 32.

12. Une formule simplifiée de calcul est :

$$\chi^2 = \left(\sum \frac{\text{effectifs observés}^2}{\text{effectifs théoriques}} \right) - n$$

où n est la taille de l'échantillon global. Voir H. Fenneteau et C. Biales (1993), *Analyse statistique de données : applications et cas pour le marketing*, Paris, Éditions Ellipses, p. 44.

$dl = (\text{nombre de lignes dans le tableau} - 1) \times (\text{nombre de colonnes dans le tableau} - 1)^{13}$.

Le test du khi carré, pour être valable, doit obéir à certaines contraintes ¹⁴ :

- s'appliquer à des échantillons indépendants ;
- porter sur les effectifs observés (et non sur les pourcentages) ;
- chacun des effectifs, pour chacune des cellules du tableau, doit théoriquement être égal ou supérieur à cinq individus.

Il doit aussi suivre certaines étapes ¹⁵ :

1. Formulation des hypothèses propres à ce tableau :
 H_0 : indépendance entre les deux variables ;
 H_1 : dépendance entre les deux variables.
2. Détermination du seuil de décision pour vérifier les hypothèses (0,10, 0,05, 0,01, 0,001, etc.).
3. Faire une conclusion au sujet de l'hypothèse (H_0 ou H_1).
4. Interprétation « littéraire » du contenu du tableau (en fonction du test) : expliquer le «sens» des différences observées (ce dernier point sera examiné en détail dans la deuxième section du chapitre 5).

Donnons un exemple à partir des données du tableau 20. Nous nous rappelons que nous avons croisé, dans ce tableau, la variable indépendante «niveau de scolarité » avec la variable «vacances prises au cours des trois dernières années ». À partir des données brutes, nous allons appliquer la formule du khi carré.

13. Il faut noter que les «totaux» des colonnes et des rangées ne font pas partie du calcul des degrés de liberté.
14. Voir à ce sujet J : P. Crauser, Y. Harvatopoulos et P. Samin (1989), *Guide pratique d'analyse des données*, Paris, Les Éditions d'Organisation, p. 93.
15. Voir à ce sujet A. Siegel (1994), *Practical Business Statistics*, Boston, Irwin, p. 704.

Tableau 24**CALCUL DU KHI CARRÉ À PARTIR DES DONNÉES
ABSOLUES DU TABLEAU 20**

Effectifs observés	Effectifs théoriques	(Effectifs observés – effectifs théoriques) ²	$\frac{(\text{Eff. obs.} - \text{eff. théo.})^2}{\text{Effectifs théoriques}}$
25	55	900	$\frac{900}{55} = 16,36$
182	225	1 849	$\frac{1 849}{225} = 8,22$
243	221	484	$\frac{484}{221} = 2,19$
257	223	1 156	$\frac{1 156}{223} = 5,18$
103	85	324	$\frac{324}{85} = 3,81$
55	25	900	$\frac{900}{25} = 36,00$
143	100	1 849	$\frac{1 849}{100} = 18,49$
77	98	441	$\frac{441}{98} = 4,5$
65	99	1 156	$\frac{1 156}{99} = 11,68$
20	38	324	$\frac{324}{38} = 8,53$
TOTAL $\chi^2 = 114,96$			

Nous nous rappelons que l'effectif théorique est obtenu par la formule :

$$\frac{\text{« total colonne } \times \text{ total rangée »}}{\text{grand total}}$$

Ainsi, pour la cellule «niveau primaire/oui » du tableau 20, le calcul sera :

$$\frac{810 \times 80}{1 170} = 55.$$

Nous voyons ici que, dans cette cellule, l'effectif théorique 55 est différent de l'effectif observé 25. L'interprétation du test se fait par la

comparaison des effectifs observés et des effectifs théoriques pour chacune des cellules du tableau ; plus la différence est accentuée, plus est forte l'influence de la variable indépendante sur la variable dépendante.

Le résultat global du test (tableau 24) donne 114,96 ; le nombre de degrés de liberté sera de $(5 - 1) \times (2 - 1) = 4$ dl. Il faut par la suite consulter la table de distribution du khi carré de Karl Pearson ; cette table apparaît dans le tableau 25.

Tableau 25

**TABLE ABRÉGÉE DE LA DISTRIBUTION DU KHI CARRÉ
(LOI DE K. PEARSON)**

Degrés de liberté	Marge d'erreur : seuils limites		
	0,10	0,05	0,01
1	2,71	3,84	6,64
2	4,60	5,99	9,21
3	6,25	7,82	11,34
4	7,78	9,49	13,28
5	9,24	11,07	15,09
6	10,64	12,59	16,81
7	12,02	14,07	18,48
8	13,36	15,51	20,09
9	14,68	16,92	21,67
10	15,99	18,31	23,21
etc.	etc.	etc.	etc.

Suivons les étapes du test du khi carré (en fonction des données du tableau 20 et des résultats calculés du tableau 24) :

Hypothèse H_0 : la scolarité n'influe pas sur la prise de vacances au cours des trois dernières années.

Hypothèse H_1 : la scolarité influe sur la prise de vacances au cours des trois dernières années.

Détermination du seuil de décision, d'acceptation ou de refus de l'hypothèse H_0 ; par convention, ce seuil se situe habituellement entre 0,10 et 0,01 (entre 10 et 1 %).

Nous devons ensuite lire sur la table abrégée du tableau 25 à n degrés de liberté les valeurs associées à différents seuils limites. Le résultat global du khi carré (voir tableau 24) est de 114,96 pour 4 dl ; à ce niveau au seuil limite de 0,01, nous avons 13,28 : ce chiffre est

très inférieur à 114,96. Nous devons donc rejeter l'hypothèse H_0 (de l'indépendance des deux variables) et accepter l'hypothèse H_1 ; dans le cas du tableau 20, la probabilité de se tromper (en acceptant H_1 et en rejetant H_0) est très faible et bien inférieure à 0,01.

4. Nous verrons plus loin comment interpréter en profondeur les résultats du tableau 20 ; en attendant, nous pouvons dire que les différences observées (dans le tableau 20), selon les niveaux de scolarité et par rapport à la prise de vacances au cours des trois dernières années, ne dépendent pas des fluctuations de l'échantillonnage mais d'une influence réelle du niveau de scolarité sur la prise de vacances au cours des trois dernières années.

LA LECTURE DES TESTS AVEC LE LOGICIEL SPSS

Dans le tableau 20, nous avons (sous le tableau) les principaux tests statistiques calculés à l'aide du logiciel SPSS. Sous le vocable «Chi-Square», quatre informations apparaissent.

« Pearson Chi-Square » :	test du khi carré mis au point par le mathématicien Karl Pearson.
« Value » :	résultat du calcul du khi carré par le logiciel SPSS (voir le calcul «manuel» effectué dans le tableau 24 ¹⁶).
«df» :	degrés de liberté : $(1 - 1)(c - 1)$ ou $(5 - 1)(2 - 1) = 4$.
«Asymp. Sig. (2-sided) » :	signification ou interprétation du résultat du test selon la table du khi cané ; il faut noter que la table intégrée au logiciel SPSS va jusqu'à trois décimales. Il y aurait donc ici moins d'une chance sur 1000 !

«Likelihood Ratio » :	cette adaptation du test de K. Pearson va tenir compte à la fois du nombre des catégories de chacune des variables et de la taille de l'échantillon ; on doit le considérer uniquement quand la population échantillonnée est petite ; quand la taille de l'échantillon est grande, le «Pearson» et le «Likelihood Ratio» vont être très semblables.
« Value » :	résultat du calcul du khi carré selon la méthode du «Likelihood Ratio».
«df» :	degrés de liberté.
«Asymp. Sig. (2-sided) » :	signification ou interprétation du test selon la table de la distribution du khi carré de Karl Pearson.

16. Ce résultat diffère (un peu) du calcul effectué dans le tableau 24 ; cette faible différence réside dans le fait que ce logiciel calcule les résultats avec cinq décimales alors que dans le tableau 24, nous nous sommes contentés de deux décimales.

«Linear-by-linear Association» :	cette statistique permet de vérifier la relation linéaire entre les deux variables qualitatives. Ce test doit être utilisé quand la population est relativement grande et quand les variables sont ordinales ¹⁷ . Ce test ne pourrait s'appliquer ici, car la variable «prise de vacances au cours des trois derniers mois » est nominale.
« Value» :	résultat du calcul établi à partir du coefficient de corrélation de Karl Pearson.
«df» :	degrés de liberté.
«Asymp. Sig. (2-sided) » :	signification du test de corrélation.

« Minimum expected frequency » :	<p>cette donnée correspond à l'effectif théorique le plus faible de l'ensemble des cellules ; nous avons déjà effectué ce calcul dans le tableau 24 (nous avons arrondi à 25). Le calcul exact dans la cellule qui correspond, au tableau 20, à «niveau de scolarité primaire » et « non » pour la variable « prise de vacances dans les trois dernières années », sera (pour l'effectif théorique de cette cellule) :</p> $\frac{360 \times 80}{1\ 170} = 24,615.$ <p>Lorsque la fréquence théorique est inférieure à 1, cela indique que dans une ou plusieurs cellules du tableau, le nombre des répondants est inférieur à 5 ; à ce moment-là, une des conditions d'application du test n'est pas remplie.</p>
----------------------------------	--

AUTRES TESTS STATISTIQUES DE RELATION ENTRE VARIABLES

Sous le tableau 20 (page 107), nous avons d'autres tests statistiques ; ces tests ne sont utilisables que si l'hypothèse H_1 (l'interdépendance des variables étudiées) est confirmée. Le test «Phi de Cramer » est un coefficient mesurant l'intensité de la relation entre deux variables dichotomiques ; une variable dichotomique ne doit avoir que deux modalités, ce qui donne donc un tableau à deux lignes et deux colonnes ¹⁸. Le calcul du Phi se fera à partir du résultat du khi carré ; nous obtiendrons alors :

$$\text{Phi} = \sqrt{\frac{\chi^2}{n}} \quad \text{ou} \quad \chi^2 = n \text{Phi}^2.$$

17. Voir à ce sujet L. Martin (1994), *Analyse et traitement des données avec SPSS : analyse descriptive et tests statistiques*, Trois-Rivières, Les Éditions SMG.

18. Voir à ce sujet L. Dayhaw (1963), *Manuel de statistique*, Ottawa, Éditions de l'Université d'Ottawa, chapitre XII.

Plus le coefficient Phi se rapproche de +1 ou de -1, plus les liens sont forts entre les deux variables ; la valeur zéro indique l'absence ou la faiblesse de la relation. Ici (pour le tableau 20), le Phi donne 0,31586, ce qui montre une intensité «assez moyenne» de la relation entre les deux variables du tableau.

Le coefficient «V de Cramer » est une autre mesure d'association de variables qualitatives (nominales). Il se calcule par la formule suivante :

$$V \text{ de Cramer} = \sqrt{\frac{\chi^2}{n(k-1)}};$$

où

χ^2 est le résultat du khi carré ;

n est la taille de l'échantillon ;

k est le plus petit chiffre selon le compte des lignes et des colonnes

(par exemple, dans le tableau 20, ce sera : colonnes - 1 ou 2 - 1 = 1 = k).

Le V de Cramer se lit de la même façon que le Phi, c'est-à-dire entre 0 et 1 ; la valeur 1 indiquant une relation parfaite entre les deux variables. Le V de Cramer permet, jusqu'à un certain point, de relativiser la taille de l'échantillon. Dans le cas du tableau 20, le coefficient V de Cramer obtenu nous montre une relation «moyenne» entre les deux variables.

Un autre test souvent utilisé est le coefficient de contingence familièrement appelé le «C» ! Il mesure lui aussi le degré d'association entre deux variables. Il s'applique quand il y a plusieurs lignes et plusieurs colonnes. La lecture du coefficient se fera entre zéro et une valeur maximale inférieure à un. Le calcul se fait à partir de la formule :

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

où :

χ^2 est le résultat du khi carré ;

n est la taille de l'échantillon.

La valeur maximale du coefficient de contingence est donnée par la formule suivante (cette valeur dépendra de la taille du tableau) :

$$\sqrt{\frac{(r-1)}{r}} \quad (\text{voir à ce sujet la table}^{19}).$$

19. Voir la table (table B page 157) de correction pour le coefficient de contingence en fonction de la taille du tableau étudié c'est-à-dire le nombre de ligne et de colonne.

Ici, on aura le coefficient corrigé en fonction de la formule :

$$C \text{ corrigé} = \frac{C}{\text{valeur maximale selon la taille du tableau ;}}$$

ce qui donnera :

$$\frac{0,30119}{0,752} = 0,4005.$$

Ce coefficient indique une «bonne» relation entre les deux variables du tableau 20.

Le coefficient de contingence doit être perçu comme un «indice de corrélation» entre les deux variables ; selon Jean-Jacques Pinty et Claude Faultier : «On ne saurait lui attribuer une précision comparable à celle du coefficient de corrélation de Bravais-Pearson²⁰.» Ce test n'aura donc, dans l'ensemble, qu'une valeur indicative. D'ailleurs, la plupart du temps, les tests de relation Phi et V de Cramer ainsi que le coefficient de contingence sont très redondants et, dans certains cas, difficiles à interpréter correctement. Cette difficulté résulte d'un désir d'obtenir un degré de précision inaccessible avec les données (le plus souvent numériques/qualitatives) obtenues par les enquêtes par questionnaire²¹.

L'ANALYSE DE LA VARIANCE

Dans une recherche empirique, il est parfois utile de pouvoir comparer des sous-groupes (d'âge, de régions ou autres) d'un même échantillon par rapport aux réponses apportées à une question importante pour l'enquête. Cette méthode repose sur «l'hypothèse nulle» selon laquelle les moyennes des différents sous-groupes de cet échantillon sont égales pour une question particulière.

20. J.-J. Pinty et C. Gaultier (1971), *Dictionnaire pratique de mathématiques et statistique en sciences humaines*, Paris, Éditions universitaires, p. 68.
21. Cette véritable hantise de la *précision vise à mimer* (de façon surtout symbolique) la rigueur des *sciences exactes*. Voir à ce sujet les titres suivants : F. Hayek (1986), *Scientisme et sciences sociales*, Paris, Plon-Agora ; A. Moles (1990), *Les sciences de l'imprécis*, Paris, Seuil.

Dans le tableau 26, nous avons un exemple d'analyse de variance (obtenue à l'aide du logiciel SPSS).

Tableau 26

ANALYSE DE VARIANCE, LE NOMBRE DE SEMAINES DE VACANCES ET LE GROUPE D'ÂGE DU RÉPONDANT

		Descriptives							
		N	Mean	Std. Deviation	Std. Error	95 % Confidence Interval for Mean			
						Lower Bound	Upper Bound	Minimum	Maximum
Nombre de semaines de vacances	L'âge du répondant 65 ans et plus	53	2,6226	1,3042	,1791	2,2632	2,9821	1,00	4,00
	55-64 ans	84	3,0476	1,0966	,1196	2,8096	3,2856	1,00	4,00
	45-54 ans	119	3,2689	,9361	8,581E-02	3,0990	3,4288	1,00	4,00
	35-44 ans	172	2,8372	1,1172	8,519E-02	2,6691	3,0054	1,00	4,00
	25-34 ans	143	2,5175	1,0737	8,979E-02	2,3400	2,6950	1,00	4,00
	24 ans et moins	55	2,1818	1,1400	,1537	1,9736	2,4900	1,00	4,00
	Total	626	2,7987	1,1346	4,535E-02	2,7097	2,8878	1,00	4,00
		ANOVA							
			Sum of Squares	Df	Mean Square	F	Sig.		
Nombre de semaines de vacances	Between Groups		65,652	5	13,130	11,016	,000		
	Within Groups		738,987	620	1,192				
	Total		804,639	625					

Le test F (de «Fisher») indique que les différences observées entre les groupes étudiés (par rapport au nombre de semaines de vacances) sont statistiquement significatives. D'autres tests, tels que le «Cochrans» et le «Bartlett-Box », permettent de confirmer (ou non) cette interprétation des résultats. B s'agit donc d'une méthode simple pour comparer les résultats selon les groupes des répondants.

4 L'ANALYSE MULTIVARIÉE

L'analyse multivariée a pour principal objectif d'étudier les relations entre plus de deux variables. Dans la plupart des cas, plusieurs indicateurs vont permettre de mieux cerner les besoins des consommateurs de produits et de services touristiques ; de la même façon, il semble plus juste d'évaluer la satisfaction des usagers, de ces produits et de ces services, à partir de plusieurs critères. Il peut être intéressant, par la suite, d'établir une certaine hiérarchie entre ces indicateurs ou ces critères : il s'agira alors d'une analyse statistique descriptive s'appliquant à plusieurs variables. Nous pouvons aussi vouloir mesurer, par exemple, l'effet de chacun des indicateurs choisis sur les besoins exprimés par les consommateurs : nous avons alors affaire à une analyse statistique explicative.

Il arrive souvent que l'on doive pousser plus loin les analyses statistiques des données afin d'extraire tout le «jus» de ces résultats. Nous allons étudier les deux approches les plus utilisées dans le domaine du tourisme : l'analyse factorielle (surtout l'analyse en composantes principales) et l'analyse de régression multiple. Le principal intérêt de ces méthodes réside dans le fait qu'elles permettent de tenir compte en même temps d'un assez grand nombre de variables. On peut ainsi obtenir un classement des variables selon un certain ordre logique (l'analyse factorielle) ou chercher l'effet de plusieurs variables indépendantes (explicatives) sur une variable dépendante (à expliquer).

L'ANALYSE FACTORIELLE

L'analyse factorielle est une technique multivariée d'analyse des données visant les objectifs suivants :

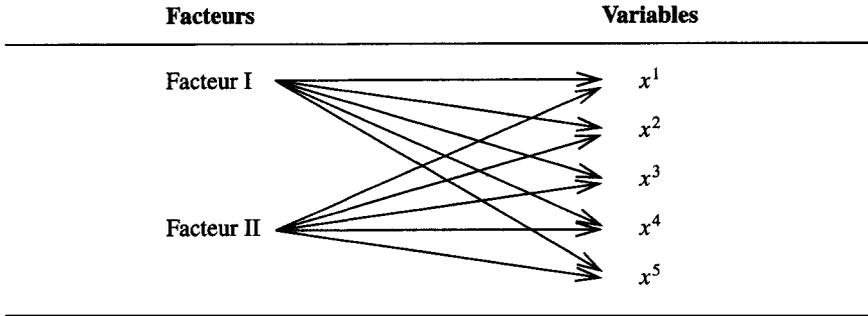
1. Étudier de façon systémique les interrelations avec un assez grand nombre de variables.
2. Réunir ces variables dans des groupes restreints appelés «facteurs».
3. Établir une hiérarchie de ces facteurs basée essentiellement sur la valeur explicative de chacun d'eux.

L'analyse factorielle est surtout une méthode de classement ; elle repose principalement sur l'hypothèse que certains facteurs (en nombre très limité) vont «expliquer» presque totalement le problème étudié.

La structure du modèle dans l'analyse factorielle prendra la forme suivante :

Tableau 27

STRUCTURE DU MODÈLE DANS L'ANALYSE FACTORIELLE



On se rend compte dans ce schéma que les cinq variables sont toutes reliées aux deux facteurs ; l'objet de l'analyse factorielle est d'indiquer quel facteur particulier a le plus d'influence sur telle ou telle variable²². L'analyse se fera à partir d'une matrice de corrélation des variables, du calcul des points initiaux de chacune de ces variables au regard des facteurs et, enfin, à la saturation des facteurs par la rotation sur les axes²³.

Pour mieux comprendre, donnons un exemple. Nous voulons obtenir une classification des variables mesurant la satisfaction après un séjour dans une station touristique. Les clients doivent répondre en fonction de l'échelle suivante :

- 22. Voir à ce sujet J. Kim et C. Mueller (1987), *Introduction to Factor Analysis*, Beverly Hills, Sage Publications.
- 23. Pour une présentation très complète de la méthodologie de l'analyse factorielle, voir G. Thomson (1950), *L'analyse factorielle*, Paris, PUF ; J. Van de Geer (1971), *Introduction to Multivariate Analysis for the Social Sciences*, San Francisco, Freeman and Co.

1 Très insatisfait	2 Insatisfait	3 Plus ou moins satisfait	4 Satisfait	5 Très satisfait
Variables mesurées : — attrait du site ; — confort ; — accueil ; — prix du séjour ; — animation des activités de plein air ; — animation des soirées ; — ambiance des soirées ; — ambiance générale ; — nombre d'animateurs ; — compétence des animateurs ; — qualité de l'hébergement ; — propreté de l'hébergement ; — qualité de l'alimentation ; — quantité des portions alimentaires ; — équilibre des menus ; — sécurité ; — ambiance des activités de plein air.				

Ces 17 variables mesurent la satisfaction, et l'analyse factorielle de la satisfaction donne les résultats suivants (tableau 28).

Tableau 28
ANALYSE FACTORIELLE DE LA SATISFACTION
À L'AUBERGE DU LAC CARRÉ
(MATRICE FACTORIELLE APRÈS ROTATION VARIMAX)

Facteurs et variables	Coefficients des variables	Variance %	
		Réelle	Interne
Facteur I : Animation			
Animation des soirées	0,77		
Animation des activités de plein air	0,75	36,1	53,9
Compétence des animateurs	0,70		
Nombre d'animateurs	0,63		
Facteur II : Alimentation			
Qualité de l'alimentation	0,84		
Équilibre des menus	0,68	12,9	19,3
Portions alimentaires	0,38		
Facteur III : Ambiance			
Ambiance dans les activités de plein air	0,80		
Ambiance des soirées	0,73	8,8	13,2
Ambiance générale	0,71		
Accueil	0,43		
Facteur IV : Hébergement			
Propreté	0,73		
Qualité	0,62	6,7	10,0
Confort	0,60		
Facteur V : Facteur résiduel			
Prix du séjour	0,48		
Attrait du site	0,40	2,5	3,6
Sécurité	0,34		
TOTAL :		67	100

Les résultats de l'analyse factorielle révèlent que l'animation est le facteur le plus important puisqu'il «explique» 53,9 % de l'ensemble de la satisfaction à l'égard de cette station touristique ; il s'agit donc d'un facteur clé pour la fidélisation de la clientèle. Les coefficients, à la suite des variables, indiquent l'importance de cette variable (dans ce facteur) par rapport aux autres. Dans le facteur I, les trois premières variables ont sen-

siblement le même poids ; le nombre d'animateurs semble ici une variable moins importante que les autres.

L'alimentation est le deuxième facteur. Elle compte pour 19,3 % de l'ensemble de la satisfaction. Ce facteur joue donc aussi un rôle assez important dans la satisfaction des usagers de cette station touristique. On remarque ici une nette hiérarchie entre les variables incluses dans ce facteur : la qualité de l'alimentation est vraiment la variable essentielle et les portions alimentaires jouent un rôle plus négligeable.

L'ambiance, sous toutes ses formes, et l'accueil ont une fonction plus faible, dans l'ensemble de la satisfaction, avec 13,2%. L'hébergement forme le quatrième facteur et «explique» 10% de la satisfaction totale. Enfin, il reste un facteur résiduel regroupant les variables jouant un rôle mineur dans l'explication de la satisfaction.

À partir de cette analyse factorielle, on remarque surtout que l'animation et l'alimentation monopolisent 73,2 % de la satisfaction totale. Ils représentent, par conséquent, les deux grands éléments de l'image de marque de cette station touristique. Ce sont des variables qui vont être motrices pour fidéliser les clients et surtout pour en attirer de nouveaux.

Jusqu'ici, nous avons surtout commenté la variance interne des différents facteurs de l'analyse factorielle. La variance «réelle» se rapporte à la qualité du modèle présenté ; dans cet exemple, le modèle dans son ensemble «explique» 67 % de la satisfaction globale des usagers, il reste donc 33 % de cet ensemble qui reste «inexpliqué ». Cette partie «inexpliquée» regroupe soit des variables reliées directement (ou indirectement) à la satisfaction mais difficiles à mesurer, soit des éléments de cette satisfaction impossibles à traduire par des questions tangibles (des aspects psychanalytiques, par exemple).

Des tests²⁴ (par exemple, le test de Kaiser-Meyer-Olkin, le test de Bartlett et le coefficient alpha de Cronbach) permettent d'évaluer la cohérence interne et la validité du modèle créé (l'explication de ces tests dépasse les objectifs de cet ouvrage). L'analyse factorielle des correspondances²⁵ permet de relier, dans un même espace, des variables de nature différente ; par exemple, ce type d'analyse juxtaposera les variables de

24. Voir à ce sujet H. Laforge (1981), *Analyse multivariée*, Montréal, Éditions Études Vivantes, chapitre VII.

25. Voir à ce sujet J.-P. Benzecri (et autres) (1979), *L'analyse des correspondances*, Paris, Dunod ; X. Bry (1995), *Analyses factorielles simples*, Paris, Économica, chapitre IV.

satisfaction du tableau 28 avec la région d'origine des répondants au questionnaire.

L'analyse factorielle, comme on l'a vu, sert donc comme technique de classement en regroupant les variables dans des facteurs ; elle établit un ordre entre les facteurs (par le pourcentage de la variance expliquée) de même qu'une hiérarchie entre les variables d'un même facteur. C'est une technique extrêmement utile et pratique lorsqu'on veut étudier un grand nombre de variables et comprendre la logique de leurs relations.

L'ANALYSE DE RÉGRESSION SIMPLE ET MULTIPLE

Pour bien saisir et comprendre la réalité, nous avons besoin de modèles simples contenant un petit nombre de variables, car il est difficile de réfléchir sur un grand nombre de ces variables. Un modèle peut être verbal ou plus abstrait, établi à partir de relations mathématiques entre les variables. Le modèle aura donc la définition suivante : « Schéma conceptuel permettant de représenter (plus ou moins formellement) un système de relations entre différents éléments ou variables qui interviennent dans un phénomène et peuvent être en interdépendance²⁶. »

Relativement aux modèles créés par l'analyse de régression, deux critiques peuvent être formulées. La première concerne le caractère « formaliste » de ce type d'approche ; ces modèles ont en effet un caractère réducteur et visent souvent à évacuer toutes les dimensions d'un problème qui ne sont pas strictement quantitatives²⁷. La deuxième critique est liée au fait que les modèles de régression reposent sur des hypothèses statistiques qui sont, la plupart du temps, impossibles à respecter dans les recherches empiriques²⁸.

26. G. Serraf, *Dictionnaire méthodologique du marketing*, Les Éditions d'Organisation, Paris, 1985, p. 160.

27. Voir à ce sujet B. Walliser et C. Prou, *La science économique*, Seuil, Paris, 1988, chapitre 1, partie 4.

28. Ces hypothèses supposent : 1) que la variable dépendante et les variables indépendantes sont mesurées sans erreur (ou avec une erreur très minime) ; 2) que le terme d'erreur est une variable totalement aléatoire ; 3) que la distribution du terme d'erreur suit une loi normale ; 4) que la moyenne du terme d'erreur est égale à zéro ; 5) que la variance des termes d'erreur est constante ; 6) que, pour la régression multiple, les variables indépendantes ne doivent pas être fortement corrélées. Voir à ce sujet D. Gujarati (1992), *Essentials of Econometrics*, New York, McGraw-Hill, chapitres 4 et 5 ; G. Ansion (1988), *Économétrie pour l'entreprise*, Paris, Éditions Eyrolles, p. 32-33.

Par conséquent, c'est avec beaucoup de réserves qu'il faut aborder les modèles de comportement utilisables pour la consommation touristique. Ainsi, il faut considérer les résultats de ces modèles comme une information supplémentaire et non comme une vérité définitive ; il faut donc faire une analyse «compréhensive» de ces résultats chiffrés²⁹.

Le principal intérêt de l'analyse de régression est de nous permettre de comprendre et de simuler la réalité observée. Si les variables indépendantes étudiées sont fortement reliées à la variable dépendante recherchée, nous avons une explication du problème ; ces relations forment alors un «modèle d'action », et il est facile, en changeant certaines valeurs dans l'équation, de faire varier de façon expérimentale les résultats.

L'équation de régression prend la forme suivante :

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n + E$$

où

Y représente la variable à expliquer ou à prévoir ;

b_0 est un terme constant ;

b_1, b_2, b_3, b_n sont des coefficients de régression générés par le modèle mathématique ;

x_1, x_2, x_3, x_n sont des variables explicatives indépendantes ;

E est un terme d'erreur.

Donnons tout de suite un exemple. Supposons un modèle, construit à partir d'une enquête, qui serait formé de six variables :

Y : nombre de nuitées dans la station touristique
(variable dépendante) ;

RBF : le revenu brut familial (variable indépendante) ;

SDS : le score de satisfaction (variable indépendante) ;

SCOL : niveau de scolarité (variable indépendante) ;

NE : nombre d'enfants (variable indépendante) ;

DST : distance parcourue pour se rendre à la station touristique
(variable indépendante).

29. Voir à ce sujet F. Nicosia et Y. Wind (1977), *Behavioral Models for Market Analysis : Foundations for Marketing Action*, Illinois, The Dryden Press.

Ici, le revenu brut familial (RBF) est codifié de la façon suivante :

- 1 : moins de 19 999 \$;
- 2 : 20 000 \$ à 29 999 \$;
- 3 : 30 000 \$ à 39 999 \$;
- 4 : 40 000 \$ à 49 999 \$;
- 5 : 50 000 \$ à 59 999 \$;
- 6 : 60 000 \$ et plus.

Le score de satisfaction (SDF) est formé de l'échelle suivante (il s'agit de la satisfaction lors d'un séjour antérieur) :

- 1: très insatisfait ;
- 2: insatisfait ;
- 3 : plus ou moins satisfait ;
- 4 : satisfait ;
- 5 : très satisfait.

Le niveau de scolarité (SCOL) :

- 1: primaire ;
- 2: secondaire ;
- 3: collégial ;
- 4: universitaire.

Le nombre d'enfants (NE) :

- 1 : aucun ;
- 2 : 1 enfant ;
- 3 : 2 enfants ;
- 4 : 3 enfants ;
- 5 : 4 enfants et plus.

La distance parcourue (DST) :

- 1 : 80 kilomètres et moins ;
- 2 : 81 à 110 ;
- 3 : 111 à 140 ;
- 4 : 141 à 170 ;
- 5 : 171 à 200 ;
- 6 : 201 et plus.

D'après les calculs, effectués par ordinateur, le carré du coefficient de corrélation multiple entre les variables (mesuré par le R^2 de Bravais-Pearson) est de 0,57³⁰ ce qui indique une relation assez forte entre la variable dépendante et les variables indépendantes du modèle. Il est en outre possible de mesurer, par un test, le caractère significatif de chacune des variables dans l'équation³¹. Ce test (le test «t» de Student) donne les résultats suivants :

Variabiles	Tests significatifs
RBF	Oui
SDF	Oui
SCOL	Non
NE	Non
DST	Non

Donc, il semble ici que la scolarité, le nombre d'enfants et la proximité de la station touristique du domicile familial n'ont pas d'influence sur les nuitées. L'équation de régression, avec les variables expliquées, donne les résultats suivants :

$$\text{Nuitées} = 0,75 + 1,15 (\text{RBF}) + 0,80 (\text{SDS})$$

Cette équation signifie que chacun des niveaux supérieurs du revenu brut familial fournit 1,15 nuitée supplémentaire alors que chacun des niveaux supérieurs de la satisfaction nous accorde 0,80 nuitée. Ces deux variables sont donc importantes pour expliquer le nombre de nuitées dans cette station touristique³².

30. Selon Frank et Green, le carré du coefficient de corrélation multiple «représente la proportion de variation dans la variable dépendante expliquée par l'association linéaire nette de toutes les variables indépendantes comprises dans l'analyse », dans R. Frank et P. Green (1973), *Marketing et méthodes quantitatives*, Paris, Dunod, p. 84. Le R^2 (appelé aussi coefficient de détermination) veut dire, dans ce cas, que les variables «revenu brut familial » et «score de satisfaction » expliquent à 57 % le nombre des nuitées. Le coefficient de corrélation multiple R , qui correspond à la racine carrée du R^2 , est de 0,75. Le coefficient de corrélation multiple se lit de la façon suivante : -1 indique une corrélation parfaite négative, 0 une absence de relation entre les variables et +1 une corrélation parfaite positive entre ces mêmes variables.
31. Ce test est bien expliqué dans A.F. Siegel (1994), *Practical Business Statistics*, Illinois, Irwin, p. 479 et suivantes.
32. Pour valider cette analyse, une analyse serrée des résidus (des erreurs du modèle) serait nécessaire ; voir à ce sujet S. Chatterjee et B. Price (1977), *Regression Analysis by Example*, N.Y., John Wiley and Sons ; A. St-Pierre (1986), *Méthodes analytiques appliquées aux problèmes de gestion*, Saint-Jean-sur-Richelieu, Éditions Bo-Pré ; R. Darlington (1990), *Regression and Linear Models*, N.Y., McGraw-Hill.

À partir de cette équation, nous pouvons faire certaines simulations ; par exemple, on peut imaginer deux extrêmes :

1. Une personne ayant un RBF de 19 999 \$ ou moins et ayant le score le plus faible de satisfaction ; on aura à ce moment :
Nuitées = $0,75 + 1,15 (1) + 0,80 (1)$
soit ici 2,7 (ou 3) nuitées.
2. Une personne ayant un RBF de 60 000 \$ ou plus et un niveau de satisfaction très élevé ; on obtiendra alors :
Nuitées = $0,75 + 1,15 (6) + 0,80 (5)$
soit 11,65 (ou 12) nuitées.

Nous pouvons ainsi effectuer toutes sortes de simulations se rapprochant des moyennes du revenu brut familial et de la moyenne de la satisfaction des répondants. À la limite, en connaissant à l'avance le revenu brut familial et le niveau de satisfaction au regard d'un séjour antérieur, on peut prévoir la durée du séjour de tous les clients les plus fidèles.

L'analyse de régression facilite donc le travail de recherche à deux niveaux : elle permet d'opérer une sélection parmi un ensemble de variables et de choisir (à l'aide des tests) celles qui sont les plus significatives. Deuxièmement, elle peut simuler un grand nombre de situations probables et peut ainsi nous aider à comprendre certains aspects de la réalité.

L'ANALYSE DISCRIMINANTE

L'analyse discriminante servira à classer les répondants, en fonction de certaines variables indépendantes (revenu, scolarité, etc.), dans deux ou plusieurs groupes. C'est une méthode de classement multivariée. A contrario, elle peut aussi être utilisée pour vérifier, par exemple, s'il y a des différences significatives entre les visiteurs et les non-visiteurs d'une région quelconque.

L'analyse discriminante est une façon de mesurer si des groupes existent bien par rapport à certaines variables. Les résultats de l'analyse discriminante, s'ils sont positifs, permettront de faire des prédictions à partir des variables indépendantes choisies. L'analyse discriminante, bien qu'assez semblable à l'analyse de régression multiple, s'en distingue par le fait que la variable dépendante est, la plupart du temps, une variable nominale (partants/non-partants, consommateurs/non-consommateurs, etc.).

Comme nous avons pu le voir, les méthodes d'analyses multivariées sont nombreuses et très variées. Elles sont très utiles pour compléter et approfondir les résultats recueillis par l'enquête. Ces méthodes peuvent en outre nous aider à vérifier des hypothèses et à en formuler de nouvelles. Elles servent donc à tirer le maximum des données recueillies.

Le principal intérêt de ces méthodes réside dans leur capacité de traiter plusieurs variables à la fois et de donner une image claire des relations entre ces variables ; de cette façon, elles renforcent la valeur explicative de ces données. Seule une approche explicative donnera un caractère vraiment scientifique aux recherches touristiques.

Il faut cependant souligner que ces méthodes sont très exigeantes sur le plan mathématique ; elles demandent un haut degré de formalisation des variables utilisées et à ce niveau, il faut le reconnaître, la recherche touristique est à ses premiers balbutiements.

5 Chapitre

Les résultats de la recherche

Nous voici au terme d'un long périple. Malgré les méandres voulus ou subis, malgré les essais et erreurs, toute recherche empirique doit aboutir à un document final qui résume les efforts et les résultats obtenus. Au bout de la route, à la toute fin, le rapport final doit faire état des résultats concrets qui répondront aux objectifs et aux hypothèses qui ont servi de point de départ à cette recherche. Il doit être rédigé pour être lu et compris ne serait-ce que par un petit nombre de personnes ! Le meilleur essai littéraire qui reste dans un tiroir ne sera jamais un véritable roman.

Dans ce chapitre, nous verrons comment présenter et interpréter les tableaux univariés et bivariés. Le rapport de recherche est une affaire à la fois de rigueur et d'imagination ; il faut passer des intuitions vagues aux idées maîtresses, du brouillon au rapport définitif. Comme dans tout essai, science (empirique) et art d'écrire sont étroitement mêlés. Enfin, nous allons esquisser un portrait de la recherche actuelle dans le domaine du tourisme.

1 PRÉSENTATION ET INTERPRÉTATION DES TABLEAUX UNIVARIÉS

Dans un premier temps, on doit créer un tableau univarié pour chacune des variables du questionnaire. Chacun de ces tableaux devrait comprendre obligatoirement :

- un numéro d'ordre ;
- un titre décrivant la variable ;
- les différentes catégories de la variable (les «codes» de cette variable) ;
- le nombre de répondants observés pour chacune des catégories (si la taille de l'échantillon dépasse 500 répondants, ce nombre peut être facultatif) ;
- le pourcentage des répondants pour chacune des catégories.

Donnons un exemple i de tableau univarié.

Tableau 29

RÉPONDANTS QUI ONT PRIS DES VACANCES AU COURS DES TROIS DERNIÈRES ANNÉES

Vacances	Pourcentage
Oui	69,2
Non	30,8
Total	100

Devant ce tableau, le seul commentaire possible est le suivant : «Qu'une majorité (près de 70%) des répondants ont pris des vacances au cours des trois dernières années. »

Pour rendre le travail de l'analyse univariée plus intéressant, il est souhaitable de présenter des tableaux synthétiques regroupant plusieurs variables dont les réponses (les catégories) sont établies de la même façon. Donnons, à nouveau, un exemple (tableau 30).

1. Tous les tableaux qui suivent, dans cette partie, sont tirés de la recherche suivante : M. Samson et J. Stafford (avec la collaboration de Daniel Blier) (1995), *Vacances et tourisme 1995. Enquête auprès d'un échantillon de Québécois et de Montréalais sur les comportements de vacances*, ministère des Affaires municipales du Québec.

Tableau 30

**IMPORTANCE DE CERTAINS ÉLÉMENTS DANS
LE CHOIX D'UNE DESTINATION DE VACANCES
ET LES MOYENNES POUR CHACUNE DES VARIABLES**

Éléments	Importance				Total (%)	Moyennes*
	Très	Assez	Peu	Pas du tout		
Centre de ski alpin	6,9	11,5	17,2	64,4	100	3,4
Pistes de ski de fond	7,5	15,4	19,5	57,6	100	3,3
Sentiers pédestres	17,1	31,3	16,0	35,6	100	2,7
Activités nautiques	23,7	29,2	15,5	31,6	100	2,6
Terrains de golf	9,8	11,4	18,4	60,4	100	3,3
Parcs gouvernementaux	25,6	34,7	14,6	25,1	100	2,4
Sites et monuments historiques	26,2	38,0	16,6	19,2	100	2,3
Musées d'art	17,6	28,0	23,9	30,5	100	2,7
Autres attraits touristiques	25,0	39,9	14,2	20,9	100	2,4
Fêtes et festivals	19,9	36,1	20,2	23,8	100	2,5
Événements sportifs	12,6	20,3	25,0	42,1	100	3,0
Événements musicaux	19,4	32,1	19,2	29,3	100	2,6
Pourvoirie	17,4	17,2	20,1	45,3	100	2,9
Pistes cyclables	22,2	29,3	13,7	34,8	100	2,6

* Ici, la mention «très important» = 1, «assez important» = 2, «peu important» = 3 et «pas du tout important» = 4 ; donc, une moyenne faible indique une grande importance accordée à la variable, par exemple, les «Sites et monuments historiques» est la variable la plus importante pour les répondants.

Dans ce tableau, en plus du commentaire d'usage (qui consiste à présenter les éléments essentiels du tableau), il est possible d'effectuer des comparaisons en recherchant les similitudes et les divergences perceptibles. Nous pouvons aussi, en considérant les moyennes, montrer l'ordre d'importance des diverses variables du tableau 30 ; cet ordre d'importance des variables apparaît dans le tableau 31.

Tableau 31

**ORDRE D'IMPORTANCE DE CERTAINS ÉLÉMENTS
DANS LE CHOIX D'UNE DESTINATION DE VACANCES ;
CLASSEMENT OBTENU A PARTIR DES MOYENNES DES RÉPONSES**

Ordre	Indicateurs	Moyennes	Indices*
1	Sites et monuments historiques	2,3	82*
2	Parcs gouvernementaux	2,4	86
2	Autres attraits touristiques	2,4	86
3	Fêtes et festivals	2,5	89
4	Activités nautiques	2,6	93
4	Événements musicaux	2,6	93
4	Pistes cyclables	2,6	93
5	Sentiers pédestres	2,7	96
5	Musées d'art	2,7	96
6	Pourvoirie	2,9	103*
7	Événements sportifs	3,0	107
8	Terrains de golf	3,3	118
8	Pistes de ski de fond	3,3	118
9	Centre de ski alpin	3,4	121

* Les indices sont calculés de la façon suivante : $\frac{\text{moyenne}}{\text{moyenne des moyennes}} \times 100$.

Par exemple pour les « Sites et monuments historiques » on aura : $\frac{2,3}{2,8} \times 100 = 82$.

À partir des résultats de ce tableau, en utilisant la moyenne des moyennes du tableau 31 (qui est de 2,8), il est possible d'interpréter ces résultats de la façon suivante : les moyennes supérieures à 2,8 indiqueraient une moins grande importance accordée à ces variables². Ces formes de loisir seraient des activités résiduelles.

2. Cette technique est tirée de D. Lehman (1979), *Market Research and Analysis*, Illinois, Irwin, p. 334.

L'intérêt des tableaux 30 et 31 est qu'ils permettent de présenter un grand nombre de variables dans le même tableau, facilitant ainsi les commentaires et l'analyse des informations. L'analyse univariée suppose que l'on fasse un tableau pour chacune des variables ou des tableaux synthétiques regroupant plusieurs variables. Une méthode simple consiste à utiliser une forme de tableau préparée à l'avance et photocopiée en plusieurs exemplaires (voir un exemple de tableau univarié «vierge» ci-dessous).

TABLEAU

	Total
	100 %

Par la suite, nous aurons, selon le nombre de variables dans le questionnaire, un nombre x de tableaux univariés ; nous verrons, dans la troisième section, comment effectuer un choix parmi cet ensemble de tableaux.

2 PRÉSENTATION ET INTERPRÉTATION DES TABLEAUX BIVARIÉS

Pour l'analyse bivariée, nous suivons la même méthode et appliquons toujours la même règle : présenter les tableaux de la façon la plus «lisible» possible. Donnons un exemple (tableau 32).

Tableau 32**ANNÉES DE SCOLARITÉ ET LA PRISE DE VACANCES
AU COURS DES TROIS DERNIÈRES ANNÉES, EN POURCENTAGE**

Scolarité	Vacances	Oui	Non -	Total
Primaire		31,3	68,7	100
Secondaire		56,0	40,0	100
12-13 années		75,9	24,1	100
14-17 années		79,8	20,2	100
18 années et plus		83,7	16,3	100

Khi carré : 116,72 ; dl : 5 ; Signification : 0,0000

Ce tableau présente les mêmes résultats que le tableau 20 (page 106) mais il est plus facile à lire et à comprendre : la variable indépendante est en ligne (la scolarité) et la variable dépendante en colonne (la prise de vacances), le total 100 % indique le sens de la lecture.

Pour le tableau 32, le commentaire pourrait être le suivant : « Nous voyons que la prise de vacances augmente suivant la progression des années de scolarité. » L'interprétation pourrait se faire en comparant chacun des niveaux de scolarité et s'énoncer de la façon suivante : « Le test du khi carré indique que les différences observées entre les niveaux de scolarité sont significatives au plan statistique³. On voit une nette différence entre le niveau primaire et le niveau secondaire ; la différence demeure notable entre la strate du "secondaire" et celle des "12-13 années" de scolarité ; pour les autres niveaux (supérieurs) de scolarité, les différences sont moins prononcées⁴. »

Il est possible aussi de présenter plusieurs tableaux croisés dans un même tableau synthétique ; voyons un exemple de cela dans le tableau 33.

3. E faut se rappeler qu'une des « contraintes » dans l'utilisation du khi carré est d'avoir un minimum théorique de cinq répondants par case du tableau ; donc ici cinquante répondants. La taille de l'échantillon pour cette recherche était de 1 070 répondants.
4. Pour les modes de lecture des tableaux simples et croisés, les commentaires et les interprétations, voir H. Jacquart (1986), *Qui ? Quoi ? Comment ? ou la pratique des sondages*, Paris, Éditions Eyrolles, p. 281 à 296 ; F. De Singly (1992), *L'enquête et ses méthodes : le questionnaire*, Paris, Nathan-Université, p. 95 à 110 ; J. Blard-Laborderie (1994), *Initiation à la statistique descriptive*, Paris, Les Éditions d'Organisation, p. 55-58.

Tableau 33

**RELATIONS* ENTRE LES DESTINATIONS DE VACANCES
DANS LA PROCHAINE ANNÉE ET CERTAINES VARIABLES
INDÉPENDANTES, EN POURCENTAGE**

Variables		Québec	Canada	États- Unis	Europe	Autres	T%
Régions	Montréal	41,0	15,2	19,7	11,8	12,3	100
	Couronne de Montréal	63,2	7,0	15,8	5,2	8,8	100
	Québec	46,4	16,7	14,3	13,1	9,5	100
	Ouest	56,8	21,0	7,4	3,7	11,1	100
	Est	61,0	23,3	10,2	0,0	8,5	100
	Centre	58,4	18,2	15,6	3,9	3,9	100
Âge	24 ans et moins	36,1	19,7	19,7	4,9	11,4	100
	25-34 ans	54,5	13,2	13,2	7,4	12,0	100
	35-44 ans	61,9	12,9	12,9	6,5	5,8	100
	45-54 ans	44,1	18,3	15,0	8,6	14,0	100
	55-64 ans	50,7	20,3	13,0	5,8	10,2	100
	65 ans et plus	47,2	20,8	13,2	15,1	3,7	100
Langue maternelle	Français	58,6	13,2	12,9	6,6	8,7	100
	Anglais	7,7	38,5	34,6	7,7	11,5	100
	Autres	14,3	28,6	10,7	25,0	21,4	100
Scolarité	Primaire	61,5	23,1	15,4	0,0	0,0	100
	Secondaire	62,0	13,0	10,2	3,7	11,1	100
	12-13 ans	55,8	17,2	12,9	4,9	9,2	100
	14-17 ans	47,8	14,8	18,1	10,4	8,8	100
	18 années et plus	31,4	22,9	18,6	14,3	12,8	100
Revenus	Moins de 20 000 \$	65,1	14,5	10,8	4,8	4,8	100
	20 000 \$-40 000 \$	61,5	10,3	12,8	4,5	10,9	100
	40 001 \$-60 000 \$	55,4	16,5	18,2	5,8	4,1	100
	60 001 \$-80 000 \$	38,6	26,3	14,0	8,8	12,3	100
	80 001 \$ et plus	38,6	11,4	25,0	11,4	13,6	100
Nombre d'enfants	0	46,6	17,6	12,8	11,0	12,0	100
	1	51,4	19,4	20,8	2,8	5,6	100
	2 et plus	59,2	13,2	16,1	4,0	7,5	100

* Les relations exprimées ici sont toutes significatives (test du khi carré) au seuil de 0,05 ou moins.

Dans ce tableau, nous avons six variables indépendantes (régions, âge, langue maternelle, scolarité, revenus, nombre d'enfants) reliées aux

destinations de vacances projetées pour la prochaine année. Ce type de tableau⁵ permet donc une grande économie dans la présentation des principaux résultats de l'enquête. Là aussi, il est possible d'utiliser une forme standardisée dans la formulation des tableaux (voir l'exemple ci-dessous) :

TABLEAU

		Total en %

Khi carré : , **dl :** , **Signification :**

Dans l'analyse des tableaux bivariés, il faut conserver les tableaux dont le seuil de signification au test du khi carré est égal ou inférieur à 0,05. Dans certains cas, si plusieurs tableaux croisés sont significatifs (par rapport aux tests), nous pouvons élever cette limite à 0,01. A l'inverse, si très peu de tableaux sont significatifs (sur le plan statistique), nous pouvons baisser le seuil d'acceptation à 0,10 (en se disant que c'est la valeur limite par convention).

3 DU «BROUILLON» AU RAPPORT DE RECHERCHE

Ce court chapitre tente de répondre à une question très simple : que faire avec la pile de tableaux des résultats de l'enquête ? Cette partie est, la plupart du temps, esquivée dans les livres de méthodologie de la recherche ; cette omission n'est pas due au hasard, car cette opération dépend plus

5. Voir à ce sujet N. Lemieux, G. Roy et J.G. Savard (1991), *Méthodes quantitatives*, Montréal, Éditions Études Vivantes, p. 248-250.

des capacités de l'auteur de faire des synthèses et de rédiger (l'art d'écrire) que des connaissances méthodologiques. Il arrive quelquefois que cette étape particulière constitue un grand achoppement pour la mise en forme et la publication des résultats.

Devant les deux piles de tableaux (les tableaux univariés et bivariés), une seule règle peut s'appliquer, à ce moment, pour l'auteur (ou les auteurs) du rapport de recherche : il doit se donner une période de temps pour réfléchir ! Cette réflexion doit porter sur les objectifs et les hypothèses de la recherche : cela représente un retour aux sources mêmes de la problématique de l'étude, aux concepts et aux dimensions utilisés. Ces dimensions (et les indicateurs de ces dimensions) seront les principaux guides dans l'établissement du rapport de recherche ; elles constitueront les idées maîtresses et les grandes parties de ce rapport (les chapitres).

Donnons l'exemple d'une étude de la satisfaction des clients d'un hôtel.

1. Dans un premier temps, il faut montrer les niveaux (de très satisfaits à très insatisfaits) de satisfaction face à l'accueil, l'hébergement, la restauration et les services offerts (ces indicateurs correspondront aux sections de ce premier chapitre) ; il faut choisir et regrouper les tableaux qui illustrent la satisfaction (ou l'insatisfaction) pour ces indicateurs. Pour chacun des tableaux (qu'il renferme une ou plusieurs variables), on doit écrire un commentaire (lié aux données chiffrées) et une certaine «interprétation» qui donne un certain «sens» à l'information présentée.
2. Dans un deuxième temps, le rapport devrait contenir un ou plusieurs chapitres présentant les tableaux bivariés qui résultent du croisement des principales variables indépendantes (âge, sexe, scolarité, revenus, etc.) avec les variables dépendantes liées à la satisfaction (selon les indicateurs mentionnés plus haut). Les variables indépendantes servent alors à segmenter la population étudiée (les clients de l'hôtel) et à comparer le niveau de satisfaction selon les catégories de la variable indépendante (par exemple, le sexe, le niveau de satisfaction des femmes comparé au niveau de satisfaction des hommes face à l'hébergement, etc.).
3. Il est possible d'ajouter au rapport une troisième partie : nous pouvons décider que certaines variables dépendantes serviront de variables indépendantes dans l'analyse de la satisfaction. Par exemple, nous pouvons segmenter les répondants en fonction des habitudes de

voyage ou des destinations choisies et considérer ces segments par rapport à la satisfaction à l'égard de l'hébergement (ou à l'égard d'autres dimensions). Pour ces tableaux bivariés ce sont les tests statistiques qui indiqueront la force des relations entre les segments définis et les niveaux de satisfaction.

Avec cette approche, tranquillement, le brouillon, par étapes successives, par le choix des tableaux à placer dans telle ou telle section deviendra un véritable rapport de recherche. Nous revenons à cette règle incontournable : l'analyse et l'interprétation des résultats consistent à lire les différents tableaux pour dégager des réponses aux diverses questions posées au début de la recherche. Il est possible de s'en servir pour vérifier des hypothèses rigides ou, plus simplement, pour atteindre les objectifs fixés au tout début de l'enquête.

Après un ou deux brouillons (ou plus), nous arrivons enfin au rapport final de recherche ; ce rapport devrait normalement contenir les éléments suivants :

LE PLAN DU RAPPORT DE RECHERCHE

1. Page titre
2. Table des matières
 - liste des tableaux
 - liste des figures (s'il y a lieu)
3. Les «faits saillants»
4. L'introduction
5. La problématique
6. La méthodologie
7. Les résultats (chapitre)
8. Les recommandations
9. La conclusion
10. Les annexes

La page titre devrait indiquer le titre succinct de l'enquête ; en voici un exemple : «Étude de la satisfaction des clients de l'Hôtel Beauséjour ». Le nom de l'auteur (ou des auteurs, ou de la firme de recherche) devrait aussi apparaître ainsi que le mois et l'année.

Dans la table des matières figureront toutes les parties du rapport de recherche ainsi que les pages correspondantes (il en va de même pour la liste des tableaux et des figures présentés dans le texte).

Les «faits saillants» résument les aspects les plus importants des résultats de l'enquête. Ils ont le rôle d'éveiller l'intérêt des lecteurs et de les inciter à lire la suite du rapport. En règle générale, cette partie ne devrait pas contenir de tableaux. La rédaction des «faits saillants » réclame un style «télégraphique» où la concision est la qualité dominante.

L'introduction a pour rôle de présenter le point de départ de la recherche (par exemple, la méconnaissance ou la baisse de la satisfaction d'un hôtel donné) et les grands traits de la démarche méthodologique (recherche quantitative par questionnaire, échantillon représentatif, etc.).

La problématique doit décrire les objectifs et les hypothèses retenus en fonction de la préenquête ; elle doit aussi aborder brièvement le cadre conceptuel : les concepts choisis, les dimensions qui définissent ces concepts et les indicateurs empiriques (du terrain) permettant de cerner ces dimensions.

La partie 6 portant sur la méthodologie est essentielle pour assurer la crédibilité des résultats ; elle doit contenir :

- une description des grandes parties du questionnaire (selon les dimensions) ;
- le prétest (description et taille) ;
- la planification de l'échantillonnage, c'est-à-dire :
 - la base des données (listes, annuaires, etc.) ;
 - le mode de tirage aléatoire ;
 - les lieux, dates, heures pour la collecte des données ;
 - les types de contrôle effectués lors de cette collecte ;
 - le seuil de confiance de l'échantillon ;
 - l'erreur acceptée sur les réponses ;
 - le taux de réponse pour l'ensemble de l'échantillon ;
 - la pondération des résultats (s'il y a lieu) ;
 - le mode de traitement et d'analyse des données.

Il y a plusieurs façons de présenter les résultats ; le plus simple est de suivre l'ordre des parties (dimensions) du questionnaire ; par exemple, pour une étude de la satisfaction de la clientèle d'un hôtel :

- chapitre 1 : «La satisfaction à l'égard de l'accueil» ;
- chapitre 2 : «La satisfaction à l'égard de l'hébergement» ;
- chapitre 3 : «La satisfaction à l'égard de la restauration» ;
- chapitre 4 : «La satisfaction à l'égard des autres services offerts».

Dans chacun des chapitres, on pourrait présenter entre trois ou quatre tableaux univariés, entre trois ou quatre tableaux bivariés selon l'intérêt que pourraient susciter ces tableaux. La tendance actuelle, dans la rédaction des rapports de recherche, est d'être de plus en plus descriptif ; le temps d'analyse coûte cher, et il résulte de cette situation une perte d'informations, car un grand nombre de résultats ne sont pas exploités comme ils le devraient.

Un rapport de recherche devrait donc contenir, en moyenne, entre 30 et 40 tableaux ; un rapport plus extensif et volumineux pourrait présenter plus de 60 tableaux. Le nombre de tableaux dépendra de l'ampleur de la recherche, du nombre des variables étudiées et, finalement, de la qualité intrinsèque des résultats. La taille du rapport (en fait le nombre de pages) variera également selon les personnes auxquelles il s'adresse. Ainsi, le «monde des affaires» s'intéressera surtout aux résultats applicables dans l'immédiat ; ce qui suppose un rapport court avec peu de tableaux⁶ ; le secteur gouvernemental (public, parapublic, communautaire) a soif d'informations chiffrées, donc beaucoup de tableaux ; le «monde universitaire» se passionne pour les débats théoriques et méthodologiques (de là l'importance des parties 4, 5, 6, 9 et 10 pour ce milieu).

De même, le contenu d'un rapport de recherche doit être adapté au public auquel il s'adresse. Ainsi, les recommandations peuvent porter sur des aspects théoriques et pratiques ; il s'agit de dégager les conséquences qu'entraînent les résultats pour l'entreprise, l'organisme, le produit ou l'ensemble de la société.

Quant à la conclusion, elle doit porter à la fois sur le cheminement méthodologique et sur les principaux résultats obtenus : on résume la démarche et les principaux acquis de l'étude réalisée. Il faut penser que selon une mauvaise habitude, fortement répandue, beaucoup de lecteurs commenceront... par la fin ! C'est la raison pour laquelle cette importante partie doit ressembler, d'une certaine façon, à un communiqué de presse.

Dans les annexes, on devra placer les éléments les plus arides d'une recherche : le «protocole» de codification, le questionnaire, les calculs,

6. Il est recommandé aussi d'écrire deux types de rapport : un très court (une dizaine de pages) avec les grands traits de l'enquête et un autre, très élaboré, comprenant tous les éléments cités plus haut. Le premier est destiné aux décideurs de l'entreprise, aux responsables qui doivent prendre les décisions ; le deuxième est préparé pour les exécutants, ceux qui doivent mettre en oeuvre les opérations de fixation de prix, de mise en marché, de publicité et de vente.

les tests de validité, etc. En fait, une seule annexe est vraiment obligatoire : c'est une copie du questionnaire ; les autres (dont les cartes géographiques) pourront apparaître au gré de l'auteur (ou des auteurs) ou de l'éditeur du texte.

La dernière étape de ce long processus consiste à faire connaître et à diffuser les résultats de la recherche. Il s'agit d'une véritable «socialisation» de la recherche ; d'après Jacques Antoine : «Socialiser, c'est prendre toutes les dispositions nécessaires pour que le «message» contenu dans les résultats de l'enquête soit effectivement reçu, perçu et assimilé⁷. » Cette tâche est encore plus importante en tourisme où souvent les chercheurs sont perçus comme des empêcheurs de tourner en rond. On doit donc utiliser tous les moyens possibles (plusieurs résumés du rapport traitant de résultats spécifiques, réunion avec les décideurs et les exécutants, vulgarisation de la démarche utilisée, etc.), pour sensibiliser les personnes concernées de façon à optimiser ces résultats !

4 ÉTAT ACTUEL DE LA RECHERCHE EN TOURISME⁸

La science est une institution servant à produire des connaissances ou, utilisant un vocabulaire plus contemporain, un système secrétant de l'information. Dans cette perspective la recherche n'est qu'un mode de traitement spécialisé de l'information. La question centrale n'est pas de faire ou de ne pas faire des recherches mais de savoir si on peut vivre et prospérer dans un système complexe comme la société actuelle sans information sur ce système.

En ce sens le tourisme au Québec est un cas type de système n'ayant pas ou peu d'information systématique (scientifique) sur lui-même ; c'est une activité qui semble fonctionner toute seule sans que l'on connaisse les mécanismes de son fonctionnement !

Dans la plupart des congrès et des colloques portant sur le tourisme, on insiste lourdement sur la concertation entre les partenaires impliqués dans ce secteur ; cette concertation sera impossible tant que les intervenants n'auront pas une bonne connaissance d'eux-mêmes et des différents

7. J. Antoine (1981), *Le sondage, outil de marketing*, Paris, Dunod, p. 197.

8. Cette section s'inspire d'un article déjà publié et toujours d'actualité J. Stafford (1982), «Crise de la recherche, crise du tourisme », dans *Téoros*, vol. 1, n° 2, Montréal, Université du Québec à Montréal.

environnements où ils sont appelés à travailler. Cette connaissance, cette information, seule une recherche systématique est en mesure de la fournir.

Cette section est consacrée uniquement à l'analyse des causes qui limitent et freinent le développement des recherches sur le tourisme au Québec. Nous avons isolé trois types de causes pouvant favoriser les blocages face à la recherche scientifique dans le domaine du tourisme. Le premier type regroupe les causes conjoncturelles : ce sont celles qui dépendent directement de la crise économique et de la crise sociale que nous vivons actuellement. Le deuxième type touche surtout les causes tendanciennes : elles font référence aux principaux acteurs-décideurs du tourisme, à leurs comportements face à la recherche scientifique. Enfin, le troisième type isole les causes structurelles, c'est-à-dire les problèmes spécifiques que pose une véritable téorologie (ou science du tourisme).

Ces trois types de causes correspondent à des niveaux de complexité et d'abstraction différents ; ils font aussi appel à une périodicité différente pour chacune des causes étudiées. Les causes conjoncturelles sont des variables exogènes donc plus faciles à saisir ; elles sont aussi reliées à l'histoire socio-économique récente du Québec. Les causes tendanciennes sont endogènes et supposent une tendance lourde au niveau des attitudes des acteurs-décideurs face à la recherche scientifique ces dernières années. Les causes structurelles sont inhérentes aux difficultés de formaliser une théorie cohérente du tourisme ; c'est un problème à très long terme et il aura un impact important sur l'avenir de cette nouvelle discipline qu'est la téorologie. Ces trois types de causes interagissent ensemble et permettront d'expliquer, en tout ou en partie, les blocages actuels au niveau de la recherche en tourisme. En fonction de cette analyse nous tenterons d'indiquer les issues possibles à la crise de la recherche en tourisme au Québec.

LES CAUSES CONJONCTURELLES

Depuis les années 1980 tous les indicateurs socio-économiques démontrent que nous sommes entrés dans une lente et profonde récession. Cette récession est surtout le produit de facteurs multiples ; selon Michel Godet :

Les crises résultent soit de l'évolution de tendances contradictoires (consommation, production selon le célèbre exemple de Malthus), soit de l'émergence d'une nouvelle tendance qui vient perturber l'équilibre en place (saut technologique par exemple)⁹.

9. M. Godet (1982), *Demain les crises. De la résignation à Ponti -fatalité*, Paris, Hachette, p. 16.

Toute crise suppose aussi une augmentation de l'incertitude générale, de la confusion et du désarroi à tous les niveaux de la structure sociale (politique, économique, social). La crise actuelle peut aussi être l'indice d'un changement profond qui amènerait un approfondissement de la récession ; cette récession correspondrait ainsi à une conjoncture d'assez longue période. Certains auteurs¹⁰ expliquent cette récession en l'assimilant à la phase descendante (contraction) d'un cycle long dit de Konradieff (du nom de son inventeur) ; le cycle long de Konradieff dure une cinquantaine d'années. Selon cette théorie :

- la période 1950-1975 correspond à l'expansion du cycle, donc une période de croissance et de relative prospérité ;
- la période 1975-2000 correspond à la contraction du cycle, décroissance, récession durant environ 25 années.

Cette conjoncture d'assez longue période peut être expliquée par un ensemble de facteurs :

- les limites atteintes dans les gains de productivité dans la plupart des pays occidentaux ;
- la diminution de certaines formes de consommation (achats alimentaires, achat de biens durables ou semi-durables) ;
- la croissance rapide puis la décroissance brutale des dépenses de l'État ;
- la croissance du secteur des services.

La conjonction des ces quatre facteurs aboutit à une société restreinte où l'incertitude domine, une société précaire qui réduit l'espace social et politique.

Mais toute crise comporte aussi des éléments positifs. Un des aspects positifs de la crise est de montrer «que ce qui allait de soi, ce qui semblait fonctionnel, efficace, comporte des carences et des vices »¹¹. Elle indique qu'il y a une non-connaissance ou une méconnaissance des différents environnements. La crise peut aussi entraîner les décideurs, après le constat de non-fonctionnalité, à des attitudes régressives ; ces attitudes vont amener, très souvent, une rigidité des règles et normes anciennes ; elles provoquent un blocage à tous les changements susceptibles d'améliorer la situation.

10. Voir à ce propos : Konradieff (1992), *G. Les grands cycles de la conjoncture*, Paris, Économica ; D. Arnould (1989), *Analyse des crises économiques*, Paris, Dunod ; M. Bruno et J. Taïeb (1992), *Crises d'hier, crise d'aujourd'hui*, Paris, Nathan.

11. E. Morin (1976), «Pour une crisologie » in la Revue *Communications*, n° 24. Paris, Seuil, p. 159.

La limitation des recherches dans le domaine du tourisme au Québec est une conséquence directe de ces attitudes provoquées par la crise conjoncturelle qui sévit actuellement. Pour la plupart des organismes opérant dans le domaine du tourisme l'environnement devient plus opaque qu'auparavant. Cette opacité, due à la non-information, creuse un fossé de plus en plus large entre les décideurs et l'évolution des marchés et des besoins des populations. Cette perte de contact va provoquer, à son tour, des crises organisationnelles. C'est un cercle vicieux : les causes conjoncturelles amènent une crise de la recherche ; cette crise de la recherche va jouer un rôle important dans la crise actuelle du tourisme au Québec. Les décideurs doivent faire du «pilotage à vue» sans savoir réellement où ils vont !

LES CAUSES TENDANCIELLES

Les causes tendanciennes vont s'ajouter aux causes conjoncturelles décrites plus haut et augmenter l'opacité entre le système touristique et ses différents environnements. Les causes tendanciennes reposent, en grande partie, sur l'incapacité des différents niveaux de gouvernements et des organismes oeuvrant dans le champ du tourisme à évaluer correctement les coûts de la non-information ; les uns et les autres font une mauvaise évaluation des effets de la non-connaissance vis-à-vis des besoins de l'ensemble de la population en matière de tourisme.

Cette évaluation des coûts peut se faire au niveau macro-économique et au niveau micro-économique. Au niveau macro-économique il s'agit de faire une comparaison qualitative et/ou quantitative entre les coûts de l'information (des recherches en tourisme) et les coûts de la non-information (les coûts de l'inaction).

Au niveau micro-économique l'évaluation des coûts de l'inaction passe par une analyse avantages-coûts de chacune des recherches en tourisme. Cette analyse doit pouvoir mesurer la rentabilité de ces recherches en prenant en considération (d'une façon systématique) les dimensions internes et les dimensions externes propres à toute recherche et en dégageant les avantages et les inconvénients d'un point de vue monétaire et non-monétaire. Ce mode d'évaluation peut être schématisé de la façon suivante (tableau 34)¹².

12. Nous nous inspirons ainsi de la démarche de Jérôme Bon et Albert Louppe (1980), *Marketing des services publics : l'étude des besoins de la population*, Paris, Les Éditions d'Organisation, p. 110.

Tableau 34**LA RENTABILITÉ D'UNE RECHERCHE**

Éléments	Valeur interne à l'organisme		Valeur externe à l'organisme (valeur sociale)	
	Avantages	Inconvénients	Avantages	Inconvénients
Éléments monétaires	<ul style="list-style-type: none"> • Meilleure connaissance du marché • Meilleures décisions • Meilleure planification 	<ul style="list-style-type: none"> • Coût global de la recherche • Coûts en temps • Coûts en surcharge de personnel • Augmentation générale des coûts d'opération 	<ul style="list-style-type: none"> • Effets d'entraînement sur d'autres secteurs liés au produit ou au service • Recettes induites par une meilleure image extérieure 	<ul style="list-style-type: none"> • Plus grande compétition entre organismes • Monopole de l'information
Éléments non monétaires	<ul style="list-style-type: none"> • Rationalité • Meilleure connaissance des objectifs • Démarche à moyen et à long termes • Plus grande implication du personnel 	<ul style="list-style-type: none"> • Caractère «réductionniste» des recherches • Changement dans les habitudes de travail • «Froideur» organisationnelle 	<ul style="list-style-type: none"> • Meilleure accessibilité du service produit • Meilleure image à l'extérieur au plan professionnel 	<ul style="list-style-type: none"> • Abandon de certains objectifs jugés peu rentables • Élimination du bénévolat et de la gratuité

Ce tableau tient compte de tous les aspects entrant dans l'évaluation d'une recherche. Si les inconvénients monétaires et non monétaires, internes et externes, dépassent les avantages monétaires et non monétaires, internes et externes, l'organisme (la firme ou le service gouvernemental) n'a pas intérêt à effectuer des recherches en matière de tourisme. Dans le cas où les avantages dépassent les inconvénients, l'organisme a un intérêt réel à avoir accès à l'information souhaitée (les coûts de l'inaction dépassent les coûts de l'action).

L'impuissance tendancielle à évaluer valablement les coûts de la non-information scientifique va avoir un impact négatif important sur l'ensemble du système touristique québécois. Dans ce système où l'on

peut plus facilement parler de « désinformation » que d'information, la concertation des acteurs, si étroitement reliée au succès de toute politique touristique, dépend aussi de la capacité de produire de l'information scientifique sur ce milieu.

LES CAUSES STRUCTURELLES

Les causes structurelles du manque actuel de recherche scientifique en tourisme apparaissent dans les difficultés d'établir une véritable « science du tourisme » (une téorologie) ; ces difficultés sont considérables et elles limitent, et limiteront pour longtemps, la structuration de cette nouvelle discipline ! Ces causes structurelles sont des problèmes épistémologiques spécifiques à toute démarche scientifique.

La meilleure façon de cerner ces causes structurelles, ces problèmes épistémologiques, c'est de répondre à la question : c'est quoi une science ? Une science c'est, avant tout, des chercheurs qui formulent des hypothèses sur la réalité étudiée dans un domaine donné et tentent, avec des instruments de recherche et d'analyse, d'infirmer ou de confirmer ces hypothèses. Ces chercheurs travaillent à l'intérieur d'un cadre de pensée, d'un courant d'idées, qu'on appelle un paradigme ; les membres de ce paradigme possèdent les mêmes valeurs sociales et les mêmes orientations théoriques et méthodologiques et luttent en rangs serrés contre les autres paradigmes à l'intérieur d'une même discipline.

Le paradigme va déterminer les questions jugées importantes pour un groupe de chercheurs et déterminer aussi les moyens de les étudier¹³.

La puissance de la méthode scientifique tient essentiellement au fait que l'esprit scientifique accepte un paradigme et par conséquent limite son champ d'investigation. À partir du moment où l'on accepte un point de vue restreint, commandé par des théories, on peut déterminer les problèmes que l'on veut résoudre et y concentrer toutes ses énergies. Auparavant, on faisait face à un ensemble de questions si mal définies qu'il devenait impossible d'y voir clair¹⁴.

Le paradigme va ainsi permettre une délimitation stricte de l'objet d'étude ; celui-ci sera le résultat : « D'un processus continu de rupture

13. Voir à ce sujet : T.S. Kuhn (1972), *La structure des révolutions scientifiques*, Paris, Flammarion.

14. G. Fourez (1974), *La science partisane*, Grenoble, Duculot, p. 76.

avec les prénotions du sens commun, avec les connaissances vagues, mythiques ou idéologiques¹⁵. »

La principale cause structurelle de l'absence de recherche fondamentale en tourisme est que la téorologie (ou la science du tourisme) est préparadigmatique. Il y a, bien sûr, des recherches qui se font en tourisme mais la plupart d'entre elles sont axées sur des résultats concrets et immédiats ; ce sont des démarches ponctuelles orientées vers des solutions pratiques ; ces approches du phénomène touristique sont nécessaires mais non suffisantes pour former un ou plusieurs paradigmes.

Le tourisme n'est pas un objet scientifiquement défini à travers une démarche théorique et méthodologique ; il n'y a pas de consensus sur les questions vitales pour le développement d'une science du tourisme. Les seules approches valables dans ce domaine sont des recherches sectorielles qui se font par le biais des études de marketing, de sociologie et d'économie du tourisme. Au plan formel, certains efforts de systématisation ont été faits mais ce sont surtout des essais de francs-tireurs de la recherche en tourisme et non de groupes de recherche organisés.

Le caractère non marchand du tourisme est surtout représenté par les tenants du tourisme social ; ce courant de pensée est surtout basé sur une certaine conception de l'éthique sociale. Actuellement le tourisme social est beaucoup plus une idéologie qu'un paradigme scientifique. Il véhicule les mêmes erreurs épistémologiques et les mêmes faiblesses théoriques et méthodologiques que la sociologie du loisir défendue naguère par Joffre Dumazedier¹⁶. Une dimension critique apparaît aussi dans l'étude du développement du tourisme dans les pays en voie de développement¹⁷ mais il en faut plus pour constituer un paradigme.

Le caractère préparadigmatique de la téorologie est un obstacle important au développement de cette nouvelle discipline ; il limite la capacité de cerner les problèmes importants dans ce domaine et réduit d'autant la possibilité de définir d'une façon spécifique l'objet d'étude du tourisme. Le tourisme, sans des paradigmes structurés, demeurera un

15. P. Bruyne, J. Herman, J. Schoutheete (1974), *Dynamique de la recherche en sciences sociales*, Paris, PUF, p. 48.
16. Voir la critique pénétrante et définitive de la sociologie dumazédienne : M.F. Lanfant (1972), *Les théories du loisir*, Paris, PUE
17. Voir à ce sujet : «Regards sur le tourisme dans les pays en développements», dans la revue *Téoros*, vol. 13, n° 2, Montréal, Université du Québec à Montréal, 1994.

phénomène résiduel, à jamais insaisissable scientifiquement et toujours à la remorque des autres sciences sociales¹⁸.

POUR SORTIR DE LA CRISE

Les causes que nous avons décrites et analysées contribuent à réduire le nombre des recherches et amènent une perte considérable d'information pertinente sur le milieu touristique québécois ; celui-ci devient une zone d'ombre où prédomine le pilotage à vue alors que nos concurrents, les États voisins, disposent des informations nécessaires au développement de leur industrie touristique.

Au niveau global, ces causes s'additionnent et expliquent le degré zéro de la recherche en tourisme au Québec. Au niveau de l'analyse, chacune de ces causes du manque de recherche en tourisme est différente et exige des solutions différentes !

La lutte contre les causes conjoncturelles est déjà engagée : il existe déjà une multitude d'études sur les marchés, les attitudes et les comportements des touristes mais ces recherches sont éparées, dispersées, sans ligne directrice.

La lutte contre les causes conjoncturelles doit passer par une compréhension de la complexité du phénomène touristique ; il faudrait s'adapter à cette complexité au lieu de régresser vers des règles et des méthodes toutes faites. Face au désarroi, à l'indécision qu'entraîne la crise actuelle il faudrait proposer de nouvelles orientations, de nouvelles approches ; c'est le moment d'inventer et/ou d'adopter de nouvelles formes de gestion et d'animation dans le monde du tourisme. La philosophie de la « société restreinte » doit être rejetée au profit d'une société ouverte, axée sur la connaissance et l'analyse des différents environnements.

La réduction des causes tendanciennes dépend de l'évolution des mentalités des acteurs-décideurs du milieu du tourisme ; ceux-ci se doivent de privilégier une gestion de l'information centrée sur la connaissance scientifique des différents éléments du système touristique québécois. Il faut sensibiliser et convaincre les acteurs-décideurs de l'importance cruciale d'une information pertinente et des coûts de la non-information face aux différents environnements.

18. J. Stafford (1986), « Les paradigmes de la recherche en téorologie : étude, analyse et critique », dans *Loisir et société*, vol. 18, n° 2, Sainte-Foy, Presses de l'Université du Québec.

Cette entreprise de sensibilisation doit se faire :

1. par un accroissement des liens entre les chercheurs et les acteurs-décideurs ;
2. par la diffusion et la socialisation des informations, c'est-à-dire s'assurer que les informations soient connues, comprises, assimilées et débouchent sur des décisions et des actions ;
3. par une formation continue des acteurs-décideurs, formation orientée vers la gestion de l'information et les coûts de la non-information.

Il serait illusoire d'espérer apporter rapidement des solutions aux causes structurelles du manque de recherche en tourisme au Québec. Le caractère préparadigmatique de la téorologie va limiter pendant longtemps le développement de cette nouvelle science du tourisme.

Les seules solutions valables reposent sur la critique et l'approfondissement des embryons de paradigmes qui existent déjà ; cette critique est essentielle car, selon le grand épistémologue Gaston Bachelard : «La connaissance cohérente est un produit non pas de la raison architectonique, mais de la raison polémique¹⁹ »

L'analyse des causes conjoncturelles, tendanciennes et structurelles du manque de recherche en tourisme nous donne un tableau d'ensemble des obstacles à une connaissance structurée dans ce domaine. Pour apporter des solutions il faut percevoir ces obstacles d'une façon dynamique. Les causes conjoncturelles sont limitées dans le temps, elles supposent des réponses immédiates et concrètes.

Les causes tendanciennes dépendent de l'évolution des attitudes des acteurs-décideurs face à la recherche, ces attitudes ne sont pas immuables et pourraient changer rapidement si des efforts soutenus sont faits dans ce sens. Les solutions aux causes structurelles exigeront un très long travail d'élaboration théorique et méthodologique.

Il ne faut pas désespérer, on peut croire qu'à très long terme de nouveaux paradigmes du tourisme pourraient émerger à partir d'une théorie générale du temps et d'une théorie générale de l'espace (environnement). Temps et espace intimement liés seraient alors les deux axes d'une science du tourisme de l'avenir.

19. G. Bachelard (1962), *La philosophie du non*, Paris, PUF.

Conclusion

Le caractère spécifique de la recherche en tourisme transparaît dans les recherches elles-mêmes... Ces recherches font appel à une documentation et à des approches de plus en plus spécialisées. Ce mouvement correspond au développement habituel d'une nouvelle discipline scientifique. Il s'agit d'établir des problématiques qui collent aux différentes situations des milieux touristiques.

La mission dévolue à la recherche empirique en tourisme est, on l'a vu, de construire des «faits scientifiques » difficiles à réfuter (sinon par d'autres «faits» nouveaux) et utiles pour la prise de décision et l'action concrète. Dans cette perspective, la théorie essayera d'établir une barrière entre la science du tourisme et le tourisme dans ses aspects extra- scientifiques. Ces aspects se résument souvent à des informations disparates et très inégales sur les plans de la qualité et de la fiabilité.

La recherche quantitative comporte un grand nombre d'étapes (problématique, questionnaire, échantillonnage, collecte des données, codification, traitement informatique et statistique des données, analyse et interprétation des résultats) qui doivent être bien synchronisées. La qualité ultime de la recherche dépendra, en grande partie, de cette synchronisation. Il faut porter une attention soutenue à chacune des parties en ne perdant jamais de vue l'ensemble de la démarche de recherche (les objectifs déterminés au départ).

La construction de «faits scientifiques» est un travail de longue haleine ; elle a pour principal mérite de donner une grande crédibilité aux

résultats obtenus, donc de rendre plus acceptables les orientations (décisions, choix) prises à partir de ces résultats. Selon les «idéologues» du tourisme, l'industrie du voyage serait la «Grande Industrie » du XXI^e siècle ; si cette prophétie devait se réaliser, les recherches empiriques en tourisme permettraient de comprendre et de canaliser ce mouvement, de voir un peu plus clair dans le monde souvent nébuleux du tourisme contemporain.

Tables statistiques

© 1998 – Presses de l'Université du Québec

Édifice Le Delta I, 2875, boul. Laurier, bureau 450, Québec, Québec G1V 2M2 • Tél. : (418) 657-4399 – www.puq.ca

Tiré de : *La recherche touristique*, Jean Stafford, ISBN 2-7605-0909-5 • DA909N

Tous droits de reproduction, de traduction et d'adaptation réservés

**TABLE A : TABLEAU DE LECTURE DU KHI CARRÉ
(DISTRIBUTION DU KHI CARRÉ
SELON LA LOI DE K. PEARSON)**

DL	P = 0,20	0,10	0,05	0,02	0,01
01	1,642	2,706	3,841	5,412	6,635
02	3,219	4,605	5,991	7,824	9,210
03	4,642	6,251	7,815	9,837	11,345
04	5,989	7,779	9,488	11,668	13,277
05	7,289	9,236	11,070	13,388	15,086
06	8,558	10,645	12,592	15,033	16,812
07	9,803	12,017	14,067	16,662	18,475
08	11,030	13,362	15,507	18,168	20,090
09	12,242	14,684	16,919	19,679	21,666
10	13,442	15,987	18,307	21,161	23,209
11	14,631	17,275	19,675	22,618	24,725
12	15,812	18,549	21,026	24,054	26,217
13	16,985	19,812	22,362	25,472	27,688
14	18,151	21,064	23,685	26,873	29,141
15	19,311	22,307	24,996	28,259	30,578
16	20,465	23,542	26,292	29,633	32,000
17	21,615	24,769	27,587	30,995	33,409
18	22,760	25,989	28,869	32,346	34,805
19	23,900	27,204	30,144	33,687	36,191
20	25,038	28,412	31,410	35,020	37,566
21	26,171	29,615	32,343	36,343	38,932
22	27,301	30,813	33,924	37,659	40,289
23	28,429	32,007	35,172	38,968	41,638
24	29,553	33,196	36,415	40,270	42,980
25	30,675	34,382	37,652	41,566	44,314
26	31,795	35,563	38,885	42,856	45,642
27	32,912	36,741	40,140	44,140	46,963
28	34,027	37,916	41,337	45,419	48,278
29	35,139	39,087	42,557	46,693	49,588
30	36,250	40,256	43,773	47,962	50,892

**TABLE B : LES FACTEURS DE CORRECTION POUR LE
COEFFICIENT DE CONTINGENCE C EN FONCTION
DE LA TAILLE DES TABLEAUX ÉTUDIÉS**

Taille du tableau	Facteurs	Taille du tableau	Facteurs	Taille du tableau	Facteurs
2 × 2	0,707	3 × 9	0,843	6 × 6	0,913
2 × 3	0,685	3 × 10	0,846	6 × 7	0,930
2 × 4	0,730	4 × 4	0,866	6 × 8	0,936
2 × 5	0,752	4 × 5	0,863	6 × 9	0,941
2 × 6	0,765	4 × 6	0,877	6 × 10	0,945
2 × 7	0,774	4 × 7	0,888	7 × 7	0,926
2 × 8	0,779	4 × 8	0,893	7 × 8	0,947
2 × 9	0,783	4 × 9	0,898	7 × 9	0,952
2 × 10	0,786	4 × 10	0,901	7 × 10	0,955
3 × 3	0,816	5 × 5	0,894	8 × 8	0,935
3 × 4	0,786	5 × 6	0,904	8 × 9	0,957
3 × 5	0,810	5 × 7	0,915	8 × 10	0,961
3 × 6	0,824	5 × 8	0,920	9 × 9	0,943
3 × 7	0,833	5 × 9	0,925	9 × 10	0,966
3 × 8	0,838	5 × 10	0,929	10 × 10	0,949

Bibliographie

- ABBEY, D. *et al.* (1992), *Connaître ses clients et leurs besoins*, Québec, Les Publications du Québec.
- AGRESTI, A. *et al.* (1997), *Statistical Methods for the Social Sciences*, N.J., Prentice-Hall.
- ANTOINE, J. (1981), *Le sondage, outil de marketing*, Paris, Dunod.
- AUDET, M. *et al.* (1986), *La production des connaissances scientifiques de l'administration*, Québec, Presses de l'Université Laval.
- BADOT, O. (1990), *L'étude marketing dans une économie de tohu-bohu*, Paris, ESF.
- BALFET, M. (1997), *L'Audit Marketing touristique*, Paris, Economica.
- BERTHIER, N. (1978), *Le sondage d'opinion*, Paris, Librairies Techniques.
- BERTRAND, R. (1986), *L'analyse statistique des données*, Québec, Presses de l'Université du Québec.
- BLANCHET, A. *et al.* (1992), *L'enquête et ses méthodes : l'entretien*, Paris, Nathan-Université.
- BLARD-LABORDERIE, J. (1994), *Initiation à la statistique descriptive*, Paris, Les Éditions d'Organisation.
- BRADLEY, I. et J. SOUTH (1994), *Introductory Statistics for Business and Economics*, Illinois, The Dryden Press.
- BRENT RITCHIE, J.R. (1994), «Roles of Research in Tourism Management», dans GOELDNER, C. et J.R. BRENT RITCHIE, *Travel, Tourism and Hospitality Research*, New York, John Wiley and Sons.

- BRY, X. (1995), *Analyses factorielles simples*, Paris, Économica.
- CADOTTE, E.R. *et al.* (1988), «Key factors in guest satisfaction», dans *The Cornell Hotel and Restaurant Administration Quarterly*, vol. 28, n° 4, Cornell, Maine,
- CANNON, J. (1994), «Issues in Sampling and Sample Design. A Managerial Perspective», dans GOELDNER, C. et J.R. BRENT RITCHIE, *Travel, Tourism and Hospitality Research*, New York, John Wiley and Sons.
- CHADWICK, R. (1994), «Concepts, Definitions and Measures Used in Travel and Tourism Research », dans GOELDNER, C. et J.R. BRENT RITCHIE, *Travel, Tourism and Hospitality Research*, New York, John Wiley and Sons.
- CHON, K. *et al.* (1991), «Functional and Symbolic Congruity Approaches to Consumer Satisfaction/Dissatisfaction in Tourism », dans *Journal of the International Academy of Hospitality Research*, vol. 3, n° 1, Virginia Polytechnic Institute and State University.
- CRAUSER, J.P., HARVATOPOULOS, Y. et P. SARNIN (1989), *Guide pratique d'analyse des données*, Paris, Les Éditions d'Organisation.
- CROMPTON, J.L. et P.K. ANKOMAN (1993), « Choice Set Propositions in Destination Decisions », dans *Annals of Tourism Research*, vol. 20, New York, Pergamon Press.
- DANN, G. *et al.* (1988), « Methodology in Tourism Research», numéro spécial de *Annals of Tourism Research*, vol. 15, n° 1, New York, Pergamon Press.
- DARMON, R.Y. *et al.* (1991), *Les fondements de la recherche commerciale*, Boucherville, Gaëtan Morin, Éditeur.
- D'ASTOUS, A. (1993), *Introduction à l'analyse des données issues d'une enquête*, Montréal, Guérin-Universitaire.
- DAYHAW, L. (1963), *Manuel de statistique*, Ottawa, Éditions de l'Université d'Ottawa.
- DEGON, R. (1990), *Les études marketing. Pourquoi ? Comment ?*, Paris, Les Éditions d'Organisation.
- DELBÈS, R. et E. TEYSSONNIÈRE (1983), *Études de marché*, Paris, J. Delmas et Cie.
- DEROO, M. et A.M. DUSSAIX (1980), *Pratique et analyse des enquêtes par sondage*, Paris, PUE
- EMORY, C. et D. COOPER (1991), *Business Research Methods*, Boston, Irwin.
- FENNETEAU, H. et C. BIALÈS (1993), *Analyse statistique des données : applications et cas pour le marketing*, Paris, Éditions Ellipses.

- FITZGIBBON, J. (1994), «Market Segmentation Research in Tourism and Travel », dans GOELDNER, C. et J.R. BRENT RITCHIE, *Travel, Tourism and Hospitality Research*, New York, John Wiley and Sons.
- FODNESS, D. (1994), «Measuring Tourist Motivation », dans *Annals of Tourism Research*, vol. 21, New York, Pergamon Press.
- FOURNIS, Y. (1974), *Les études de marché*, Paris, Dunod.
- FRANK, R. et P. GREEN (1973), *Marketing et méthodes quantitatives*, Paris, Dunod.
- GIANNELLONI, J.-L. et Éric VERNETTE (1995), *Études de marché*, Paris, Vuibert.
- GILLES, A. (1994), *Éléments de méthodologie et d'analyse statistique pour les sciences sociales*, Montréal, McGraw-Hill.
- GOLDFARB, B. et C. PARDOUX (1993), *Introduction à la méthode statistique*, Paris, Dunod.
- GRAVEL, R. (1994), *La méthodologie du questionnaire*, Québec, Éditions Bo-Pré.
- GREEN, P.E. et D.S. TULL (1974), *Recherche et décisions en marketing*, Grenoble, Presses universitaires de Grenoble.
- GREEN, S. et al. (1997), *Using SPSS for Windows. Analyzing and Understanding Data*, N.J., Prentice-Hall.
- GUAY, J.H. (1991), *Sciences humaines et méthodes quantitatives*, Montréal, Beauchemin.
- GUNN, C.A. (1994), «A Perspective on the Purpose and Nature of Tourism Research Methods », dans C. GOELDNER et J.R. BRENT RITCHIE, *Travel, Tourism and Hospitality Research*, New York, John Wiley and Sons.
- HARVATOPOULOS, Y., LIVIAN, Y. et P. SARNIN (1989), *L'art de l'enquête*, Paris, Éditions Eyrolles.
- HERMET, G. et A. JOLIBERT (1995), *La part de marché*, Paris, Économica.
- HOWELL, D. (1998), *Méthodes statistiques en sciences humaines*, Bruxelles, De Boeck-Université.
- HOWITT, D. et al. (1997), *A Guide to Computing Statistics with SPSS for Windows*, N.J., Prentice-Hall.
- HURST, F. (1994), «Enroute Surveys», dans GOELDNER, C. et J.R. BRENT RITCHIE, *Travel Tourism and Hospitality Research*, New York, John Wiley and Sons.
- JACQUART, H. (1988), *Qui ? Quoi ? Comment ? Ou la pratique des sondages*, Paris, Éditions Eyrolles.

- JAVEAU, C. (1971), *L'enquête par questionnaire*, Bruxelles, Éditions de l'Université Libre de Bruxelles.
- JAVEAU, C. et C. VIGNERON (1989), *Les secrets des sondages enfin révélés*, Bruxelles, Éditions Labor.
- KIM, J. et C. MUELLER (1987), *Introduction to Factor Analysis*, Beverly Hill, Sage Publications.
- KING, B. *et al.* (1993), «Social Impacts of Tourism. Host Perceptions », dans *Annals of Tourism Research*, vol. 20, New York, Pergamon Press.
- LABELLE, D. et J. MÉNARD (1990), *Introduction à la statistique appliquée*, Montréal, Éditions Addison-Wesley.
- LAGARDE, de J. (1995), *Initiation à l'analyse des données*, Paris, Dunod.
- LAMBIN, J.J. (1990), *La recherche marketing*, Paris, McGraw-Hill.
- LAZARSELD, P. (1965), *Le vocabulaire des sciences sociales*, Paris, Mouton.
- LAZARSELD, P. (1970), *Philosophie des sciences sociales*, Paris, Gallimard.
- LEMIEUX, N., ROY, G. et J.G. SAVARD (1991), *Méthodes quantitatives*, Montréal, Éditions Études Vivantes.
- LEVIN, R. *et al.* (1998), *Statistics for Management, N.J.*, Prentice-Hall.
- MARTINET, A.C. *et al.* (1990), *Épistémologie et sciences de gestion*, Paris, Économica.
- MAX, A. (1981), *La république des sondages*, Paris, Idées-Gallimard.
- MARTIN, L. (1994), *Analyse et traitement des données avec SPSS : analyse descriptive et tests statistiques*, Trois-Rivières, Les Éditions S.M.G.
- McDOUGALL, G. *et al.* (1994), « Scaling and Attitude Measurement in Tourism and Travel Research », dans GOELDNER, C. et J.R. BRENT RITCHIE, *Travel, Tourism and Hospitality Research*, New York, John Wiley and Sons.
- MILL, R. et A. MORRISON (1985), *The Tourism System*, New Jersey, Prentice-Hall.
- MOSCAROLA, J. (1990), *Enquêtes et analyse des données*, Paris, Vuibert.
- MUCCHIELLI, R. (1967), *Le questionnaire dans l'enquête psycho-sociale*, Paris, Librairies Techniques.
- NEGRO, Y. (1987), *L'étude du marché*, Paris, Vuibert.
- NOELLE, E. (1966), *Les sondages d'opinion*, Paris, Les Éditions de Minuit.
- OLSON, R. (1994), «Research Needs of the Restaurant Industry», dans GOELDNER, C. et J.R. BRENT RITCHIE, *Travel, Tourism and Hospitality Research*, New York, John Wiley and Sons.

- OUELLET, G. (1985), *Statistiques*, Québec, Éditions Le Griffon d'Argile.
- PELLETIER, J. (1979), *Détermination de la taille d'un échantillon*, Québec, Ministère du Tourisme, de la Chasse et de la Pêche.
- PERRIEN, J. et al. (1983), *Recherche en marketing*, Chicoutimi, Gaëtan Morin, Éditeur.
- PINTY, J.J. et C. GAULTIER (1971), *Dictionnaire pratique de mathématiques et statistiques en sciences humaines*, Paris, Éditions Universitaires.
- PLAISENT, M. et al. (1998), *SPSS 7.5 pour Windows*, Québec, Presses de l'Université du Québec.
- PY, B. (1987), *Statistique descriptive*, Paris, Économica.
- QUIVY, R. et L. VAN CAMPENHOUDT (1988), *Manuel de recherche en sciences sociales*, Paris, Dunod.
- ROSE, J. (1993), *Le hasard au quotidien*, Paris, Seuil.
- ROSS, G.F. (1993), «Destination Evaluation and Vacation Preferences », dans *Annals of Tourism Research*, vol. 20, New York, Pergamon Press.
- RYAN, C. (1995), *Research Tourist Satisfaction*, London, Routledge.
- SALEH, F. et al. (1992), «Client perception of hotels : A multi-attribute approach », dans *Tourism Management*, vol. 13, n° 2, New York.
- SAMSON, M. et J. STAFFORD, avec la collaboration de D. BIBEAU (1986), *L'effet de la retraite sur les départs en vacances des nouveaux retraités*, Etudes et documents n° 47, Montréal, INRS-Urbanisation.
- SAMSON, M. et J. STAFFORD, avec la collaboration de D. BLIER (1995), *Vacances et tourisme 1995. Enquête auprès d'un échantillon de Québécois et de Montréalais sur les comportements de vacances*, Québec, Ministère des Affaires municipales du Québec.
- SEKARAN, U. (1992), *Research Methods for Business*, New York, John Wiley Inc.
- SERRAF, G. (1985), *Dictionnaire méthodologique du marketing*, Paris, Les Éditions d'Organisation.
- SIEGEL, A. (1994), *Practical Business Statistics*, Boston, Irwin. SINCICH, T. (1996), *Business Statistics by Example*, N. J., Prentice-Hall.
- SINGLY, de F. (1992), *L'enquête et ses méthodes : le questionnaire*, Paris, Nathan.
- SKRHAK, B. et A. VERCASSON (1989), *Méthodes statistiques pour la gestion*, Paris, Les Éditions d'Organisation.

- STAFFORD, J. *et al.* (1988), *Profil socio-économique et mesure de la satisfaction des usagers dans les établissements membres du Mouvement québécois des camps familiaux*, collection «Études urbaines et touristiques », n° 14, Département d'études urbaines et touristiques, Montréal, Université du Québec à Montréal.
- STAFFORD, J. *et al.* (1988), *Étude et analyse socio-économique de la clientèle des bases membres de Réseau plein-air*, collection «Études urbaines et touristiques », n° 15, Département d'études urbaines et touristiques, Montréal, Université du Québec à Montréal.
- STAFFORD, J. et S. GAGNON, avec la collaboration de M. BLANCHETTE (1995), *Besoins de formation et satisfaction au travail du personnel syndiqué de l'hôtellerie et de la restauration au Québec*, Études, matériaux et documents n° 9, Département d'études urbaines et touristiques, Montréal, Université du Québec à Montréal.
- TENENHAUS, M. (1994), *Méthodes statistiques en gestion*, Paris, Dunod-Entreprise.
- TREMBLAY, A. (1991), *Sondages : histoire, pratique et analyse*, Boucherville, Gaëtan Morin, Éditeur.
- VAN VRACEM, P. et M. GAUTHY-SINECHAL (1988), *Études de marché et sondage d'opinion*, Bruxelles, Éditions De Boeck.
- VÉDRINE, J.P. (1991), *Les traitements des données en marketing*, Paris, Les Éditions d'Organisation-Université.
- WONNACOTT, T. et R. WONNACOTT (1988), *Statistique*, Paris, Économica.
- ZALTMAN, G. et P. BURGER (1975), *Marketing Research*, Illinois, The Dryden Press.

