

ABRÉGÉ SUR LES MÉTHODES DE RECHERCHE ET LA RECHERCHE EXPÉRIMENTALE

Louis Laurencelle



**Presses
de l'Université
du Québec**

**ABRÉGÉ SUR LES MÉTHODES DE RECHERCHE
ET LA RECHERCHE EXPÉRIMENTALE**

PRESSES DE L'UNIVERSITÉ DU QUÉBEC

Le Delta I, 2875, boulevard Laurier, bureau 450

Sainte-Foy (Québec) G1V 2M2

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

Courriel: puq@puq.ca • Internet : www.puq.ca

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

DISTRIBUTION DU NOUVEAU MONDE

30, rue Gay-Lussac, 75005 Paris, France

Téléphone: 33 1 43 54 49 02

Télécopieur: 33 1 43 54 39 15

SUISSE

SERVIDIS SA

5, rue des Chaudronniers, CH-1211 Genève 3, Suisse

Téléphone: 022 960 95 25

Télécopieur: 022 776 35 27



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 ».

ABRÉGÉ SUR LES MÉTHODES DE RECHERCHE ET LA RECHERCHE EXPÉRIMENTALE

Louis Laurencelle

2005



Presses de l'Université du Québec

Le Delta I, 2875, boul. Laurier, bur. 450
Sainte-Foy (Québec) Canada G1V 2M2

Catalogage avant publication de Bibliothèque et Archives Canada

Laurencelle, Louis, 1946- .

Abrégé sur les méthodes de recherche et la recherche expérimentale

Comprend des réf. bibliogr.

ISBN 2-7605-1362-9

1. Recherche – Méthodologie. 2. Sciences – Méthodologie.
3. Sciences – Expériences – Méthodologie. I. Titre.

Q180.55.M4L38 2005 001.4'2 C2005-940550-3

Nous reconnaissons l'aide financière du gouvernement du Canada
par l'entremise du Programme d'aide au développement
de l'industrie de l'édition (PADIÉ) pour nos activités d'édition.

Mise en pages : INFOSCAN COLLETTE QUÉBEC

Couverture : RICHARD HODGSON

1 2 3 4 5 6 7 8 9 PUQ 2005 9 8 7 6 5 4 3 2 1

Tous droits de reproduction, de traduction et d'adaptation réservés
© 2005 Presses de l'Université du Québec

Dépôt légal – 2^e trimestre 2005
Bibliothèque nationale du Québec / Bibliothèque nationale du Canada
Imprimé au Canada

PRÉFACE

La méthodologie constitue un savoir stratégique, tactique, heuristique et logistique. C'est tout ce qui reste lorsque le savoir substantif relatif à une partie d'un domaine de connaissance devient tout à coup obsolète. Par exemple, la découverte d'un neuromodulateur de l'appétit pourra rendre caduc tout un ensemble de connaissances dûment et péniblement validées par expérimentation sur les contrôles de l'appétit via les entrées-sorties énergétiques. Dans plusieurs domaines en rapide croissance, les connaissances substantives sont vite reléguées aux archives de la curiosité trompée. Heureusement, les connaissances méthodologiques restent, elles. Elles autorisent le chercheur à porter sa curiosité vers des problèmes neufs, et à les aborder d'emblée avec compétence, dans sa perpétuelle quête de compréhension d'un monde dont il est partie prenante.

Le présent ouvrage constitue un recueil des méthodes et techniques de recherche scientifique, à tout le moins les principales que l'on retrouve en sciences du comportement et de la cognition. Son auteur a voulu privilégier le caractère compréhensif et la rigueur des contenus plutôt que verser dans un jargon particulier ou dans des discussions savantes. Il s'agit d'un abrégé, en quelque sorte un *vade mecum* pour le chercheur, qui n'est pas toujours spécialiste de toutes les questions méthodologiques et dans lequel il pourra trouver les définitions conceptuelles,

formules et références bibliographiques dont il aura éventuellement besoin. Ce texte a servi et peut servir de manuel de référence pour un cours de méthodologie dans presque tous les domaines où il se fait de la recherche empirique, les contenus présentés étant universels.

L'auteur est professeur titulaire de méthodologie à l'Université du Québec à Trois-Rivières. Il agit comme consultant et méthodologue-conseil dans plusieurs équipes de recherche depuis 1969, et ce, en psychologie fondamentale et clinique, en médecine, en physiologie de l'effort, en biomécanique humaine, en psychométrie, en chiropratique, en sciences infirmières, en épidémiologie et pour les enquêtes d'opinions, etc. L'équilibre des contenus abordés dans l'*Abrégé sur les méthodes de recherche et la recherche expérimentale* reflète cette richesse et cette diversité.

Jacques Beaugrand, Ph. D.
professeur honoraire
Université du Québec à Montréal

TABLE DES MATIÈRES

Préface	vii
--------------------------	-----

Remerciements	xiii
--------------------------------	------

FONDEMENTS

La connaissance scientifique	1
Nature de la connaissance scientifique	1
But et mode de la connaissance scientifique	3
Place et buts de la recherche	4
Recherche et méthode scientifiques	4
Une taxinomie des démarches de recherche	8
Principes de classement	8
Description sommaire des schémas types de recherche	9
Distinctions entre les démarches de recherche et formes mixtes	17
Validité interne et externe d'une recherche	19
Les concepts de validité en science	19
Validité interne d'une recherche	20
Validité externe d'une recherche	21

PROCÉDURE DE LA RECHERCHE EXPÉRIMENTALE

Question de recherche et domaine de recherche	23
Élaboration de la problématique de recherche	24
Formulation de l'hypothèse de recherche ou de l'objectif de recherche	25
Nature de l'hypothèse de recherche	25
Variables en jeu	26
Formulation de l'hypothèse de recherche ou de l'objectif de recherche	29
Opérationnisme et définitions opérationnelles	30
Choix des méthodes et des instruments de mesure	31
Diversité des mesures d'un même phénomène	31
Critères pour choisir la bonne mesure	32
Choix des méthodes de contrôle et d'expérimentation	34
Contrôle des conditions environnementales	35
Contrôle de la (des) variable(s) indépendante(s) et des variables contrôlées ; postvérification	36
Effets pervers ; les techniques du simple et du double insu	37
Protocole et plan d'expérience	41
Sélection et nombre de sujets	54
Mise au point du dispositif d'expérimentation	64
Préexpérimentation (fonctions et buts)	65
Expérimentation proprement dite	66
Éthique du chercheur et ses obligations déontologiques	66
Robotisation des procédures et observation	67
Mortalité expérimentale	67
Traitement statistique et généralisation	69
Traitement statistique : synthèse de données et généralisation	69

Vérification de l'hypothèse de recherche et impact dans le domaine de recherche	72
Vérification statistique vs conclusion	72
Significativité et portée d'une expérience.	73
Impact dans le domaine de recherche	74
Annexe A	
<i>Déontologie de la recherche</i>	75
Annexe B	
<i>Éléments de la théorie des tests</i>	81
Annexe C	
<i>Quelques grandes idées sur l'échantillonnage</i>	93
Annexe D	
<i>Contenus et présentation d'un rapport de recherche</i>	99
Références	107
Index	111

REMERCIEMENTS

Il n'est jamais facile pour un auteur d'analyser la dette intellectuelle qu'il a contractée auprès des uns et des autres dans l'élaboration de son ouvrage, ni complètement satisfaisant de nommer seulement quelques personnes plus significatives. Depuis 1969, à l'Université du Québec à Montréal, où j'enseignais les méthodes de recherche au baccalauréat en psychologie, jusqu'en 2005 à l'Université du Québec à Trois-Rivières, où je me retrouve au baccalauréat et à la maîtrise dans différentes disciplines (kinésiologie, psychologie, sciences infirmières, chiropratique), chacune avec ses particularités, j'ai reçu autant que j'ai donné, et les étudiants, surtout les étudiants en recherche, ont été mes meilleurs collaborateurs.

Il reste que l'assemblage des rubriques de l'*Abrégé sur les méthodes de recherche et la recherche expérimentale* et leur rédaction ont été truffés de périls : choix des termes, définitions des concepts, développement plus ou moins extensif de certaines parties, etc. Le présent ouvrage constitue la énième version de notes de cours plusieurs fois remaniées ; cette version a bénéficié d'une aide spéciale de certains collègues, éminents méthodologues, qui ont accepté d'en faire la révision. Je remercie donc sincèrement mes collègues Yves Girouard (Université du Québec à Trois-Rivières), Jacques P. Beaugrand (Université du Québec à Montréal) et Michèle Robert (Université de Montréal) pour leurs précieuses suggestions. Quant aux points de vue,

options et définitions qui pourraient paraître discutables, controversés, voire erronés, ils ne sont le fait que de l'auteur et n'engagent en rien les personnes nommées.

L'auteur tient à remercier le Décanat des études de cycles supérieurs et de la recherche ainsi que le Département des sciences de l'activité physique de l'Université du Québec à Trois-Rivières pour leur généreuse participation à l'effort de publication.

Louis Laurencelle
mai 2005

FONDEMENTS

LA CONNAISSANCE SCIENTIFIQUE

Nature de la connaissance scientifique

La connaissance et le savoir scientifiques se distinguent des autres savoirs ou autres types de connaissance, tout comme la science, en tant que mode de pensée, de discours et d'action, se distingue des autres activités humaines. Plusieurs critères de distinction peuvent être mentionnés, dont aucun ne suffit à lui seul.

La connaissance scientifique doit d'abord être empirique, émaner de phénomènes sensibles, se rapporter à une réalité tangible, objective et observable, être confirmée par répétition. La démarche de connaissance doit être systématique, foncièrement désintéressée¹ (notamment quant à ses retombées), critique, et elle s'appuie généralement sur le principe du déterminisme

1. Nous admettons les recherches dites « commanditées », « appliquées », « orientées », même si le chercheur a un intéressement quelconque (notamment monétaire, peut-être via son employeur) dans l'objet de la recherche, comme c'est le cas de la recherche pharmaceutique en compagnie. Le critère de désintéressement mentionné est formel et s'adresse à la *démarche de recherche*, laquelle ne devrait pas être influencée, ni dans ses modalités ni dans la diffusion de ses résultats, par des intérêts autres que de connaissance.

(strict ou statistique). Enfin, la science étant du domaine public, l'expérience de connaissance et ses contenus principaux doivent pouvoir être exhaustivement décrits, communiqués et reproduits.

En fait, ce que l'on entend et que l'on admet comme « explication scientifique » de la nature et de ses phénomènes tient à une idéologie implicite et assez contraignante. Cette idéologie peut se traduire en quelques principes épistémologiques, couramment et tacitement acceptés par les chercheurs. Les voici, accompagnés d'une définition sommaire :

- Le *rationalisme* : « Rien de ce qui existe ne trouve une explication qui soit étrangère à ce que la raison humaine peut accepter » (*Le Petit Larousse*).
- Le *matérialisme* : « Rien n'existe en dehors de la matière, et que l'esprit est lui-même entièrement matériel » (*Le Petit Larousse*).

Tout est matériel : les phénomènes, les causes, les effets. Cela ne revient pas à dire que la logique ou les processus mentaux sont bannis ou qu'ils n'existent pas.

- L'*empirisme* : la connaissance doit procéder de l'expérience, à l'exclusion des idées ou systèmes a priori ou des doctrines autoritaires. Essentiellement, les données de la science viennent des sens.
- Le *déterminisme* : chaque phénomène, chaque variation observée dans la nature est attribuable à l'action d'une cause ou d'un ensemble fini de causes.

Le déterminisme strict, affirmation du règne universel du principe de causalité (mécanique), est consacré comme principe d'explication notamment dans les œuvres de Pierre Simon de Laplace (1749-1827). Ce dogme a été plus ou moins abandonné dans l'épistémologie scientifique moderne, depuis les années 1950, au profit d'une version assouplie, le déterminisme statistique. Le déterminisme statistique affirme encore la primauté du principe de causalité mais il accorde sa juste place à la complexité et au hasard. On peut le formuler comme suit : « Chaque phénomène, chaque variation observée est attribuable à l'action probable de certains

facteurs principaux en plus d'une fluctuation aléatoire plus ou moins grande, qu'on espère réduire mais qu'il est impossible d'annuler ». Les modèles d'explication proposés dans les sciences humaines (sociologie, psychologie, etc.), certains modèles en biologie, les phénomènes stochastiques (économie, démographie, météorologie, etc.), s'y retrouvent mieux dans cet accommodement du principe de causalité.

- La *cohérence* : L'explication d'un phénomène doit tenir compte des conceptions et théories admises dans le domaine.
- La *parcimonie* : Un modèle explicatif, une théorie qui met en jeu un nombre de causes réduit est préférable, plus « vraie », plus scientifique que celle ayant un nombre de causes plus élevé.
- La *falsifiabilité* (ou testabilité) : Un modèle explicatif, une théorie, un test ne sont admissibles que si on peut concevoir une expérience capable de les infirmer (les falsifier).

But et mode de la connaissance scientifique

Les buts de la science sont d'expliquer, prédire et, éventuellement, contrôler les phénomènes naturels. « L'explication » réfère ici à l'insertion des comportements, manifestations et variations observables d'un phénomène dans un modèle causal, mécaniste à la limite, selon la doctrine courante du déterminisme. L'observation et l'expérimentation sont les modes privilégiés de l'élaboration du savoir scientifique ; ceux-ci n'excluent pas la réflexion, l'induction, la déduction (voir Formulation de l'hypothèse de recherche, p. 25), au contraire, mais ils les dominent (en général).

L'étude scientifique d'un phénomène procède ordinairement par à-coups, par approximations successives, en fonction de l'initiative et du talent des chercheurs ainsi que des conditions qui peuvent ou non favoriser leurs efforts. Néanmoins, on peut identifier un schéma général, un paradigme d'étude scientifique, qui comporterait les étapes suivantes : 1) établissement d'un fait ou d'un ensemble de faits d'observation ; 2) intuition d'une hypothèse

explicative ou rattachement à un modèle explicatif (ou théorie); 3) préparation d'un devis et d'un dispositif d'expérimentation (ou de démonstration); 4) expérimentation et mesure; 5) compilation et interprétation des résultats; 6) retour inductif/déductif sur le modèle explicatif. L'élaboration de la théorie du phénomène (le but ultime de l'effort de connaissance) passe par les paliers de complexité suivants: observation et identification des faits significatifs → régularités statistiques → loi → groupe de lois en affinité → modèle (système cohérent de lois) ou théorie. Le schéma présenté en page 5 illustre globalement le processus de l'étude scientifique des phénomènes et est donné comme base de réflexion.

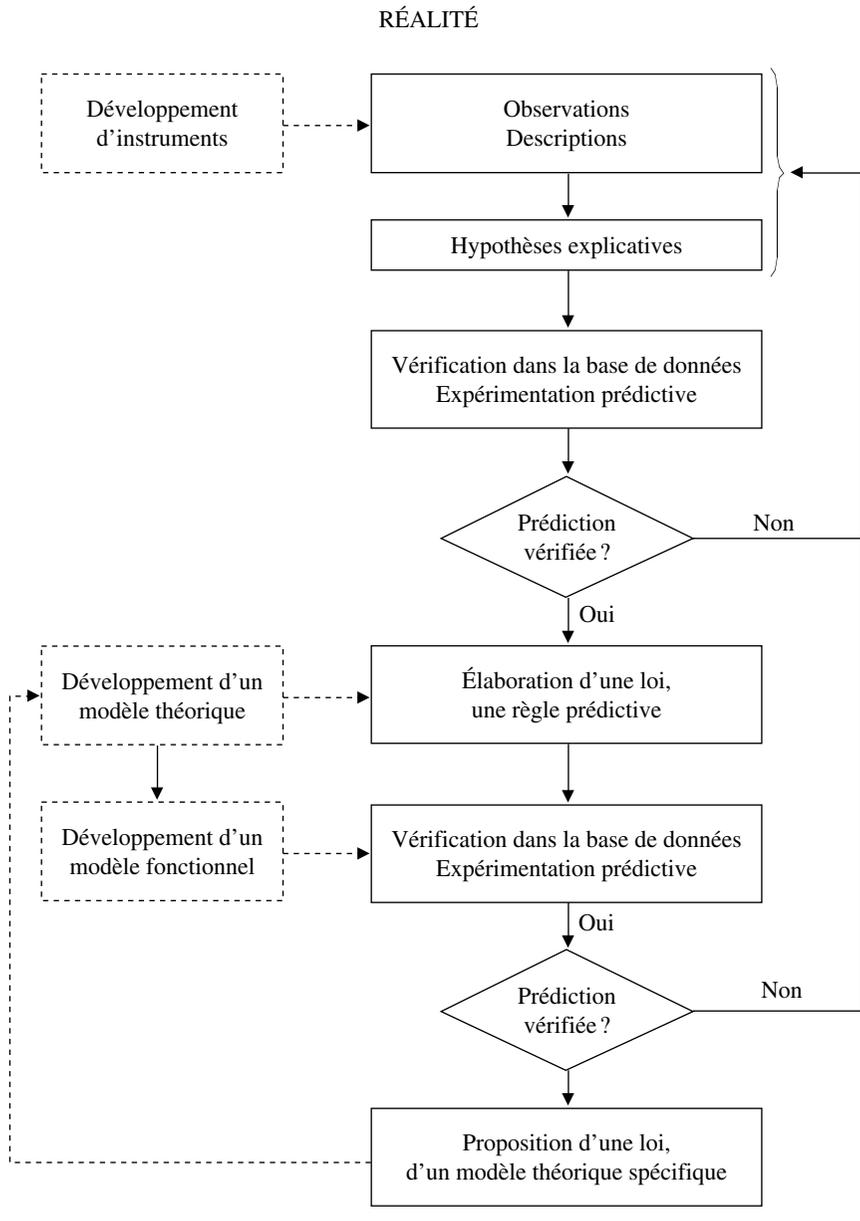
L'élément ultime de la connaissance scientifique, le modèle, en possède aussi les critères essentiels: se rapporter à une réalité tangible et observable, être critique, être déterministe. Le modèle n'est pas « la vérité » sur un phénomène, à moins que ce soit dans une pédagogie d'ailleurs dépassée sur le plan socioculturel. L'origine du modèle est souvent difficile à retracer, multiple et brumeuse: l'intuition du chercheur y a sa part, de même que les propositions même partielles de ceux qui le précèdent, l'érudition dans le domaine de recherche, la connaissance de modèles dans d'autres secteurs de la science. Le déterminisme (« Tout effet peut être assigné à l'action d'une cause ou d'un nombre déterminé de causes identifiables ») et la parcimonie (« Un modèle plus simple est préférable à un modèle équivalent plus complexe ») forment deux balises généralement reconnues dans l'élaboration de modèles en science moderne (Beaugrand, 1988a, p. 1-35).

PLACE ET BUTS DE LA RECHERCHE

Recherche et méthode scientifiques

Dans le monde de la science, évoluent plusieurs catégories de personnes qui s'adonnent à différents types d'activité: recherche, applications industrielles ou biomédicales, procédés d'usinage et

Un schéma possible des étapes du processus de recherche, de la réalité à la théorie.



de transformation de matériaux, technologies, publications, enseignement, etc. Dans cette variété foisonnante de métiers et de modes d'activité, les chercheurs et la recherche représentent le caractère distinctif de la science : ils la définissent comme une discipline et en fondent la légitimité. Ce sont les chercheurs qui, les premiers, observent la méthode de la science (en réalité, plusieurs méthodes) et la développent, et qui surveillent son application, l'intègrent dans un exercice professionnel. Le philosophe moraliste Nietzsche (1844-1900) tient les propos suivants :

À tout prendre, les méthodes scientifiques sont un aboutissement de la recherche au moins aussi important que n'importe quel autre de ses résultats ; car c'est sur l'intelligence de la méthode que repose l'esprit scientifique, et tous les résultats de la science ne pourraient empêcher, si lesdites méthodes venaient à se perdre, une recrudescence de la superstition et de l'absurdité reprenant le dessus. Des gens intelligents peuvent bien apprendre tout ce qu'ils veulent des résultats de la science, on n'en remarque pas moins à leur conversation, et notamment aux hypothèses qui y paraissent, que l'esprit scientifique leur fait toujours défaut : ils n'ont pas cette méfiance instinctive pour les aberrations de la pensée qui a pris racine dans l'âme de tout homme de science à la suite d'un long exercice. Il leur suffit de trouver une hypothèse quelconque sur une matière donnée, et les voilà tout feu tout flamme pour elle, s'imaginant qu'ainsi tout est dit. Avoir une opinion, c'est bel et bien pour eux s'en faire les fanatiques. Y a-t-il une chose inexplicquée, ils s'échauffent pour la première fantaisie qui leur passe par la tête et ressemble à une explication ; il en résulte continuellement, surtout dans le domaine de la politique, les pires conséquences ! C'est pourquoi tout le monde devrait aujourd'hui connaître à fond au moins une science ; on saurait tout de même alors ce que c'est que la méthode, et tout ce qu'il y faut d'extrême circonspection. (Nietzsche, 1968)

La recherche et la méthode, dans leurs aspects de soumission au réel, de rigueur intellectuelle et de questionnement critique, sont donc ce qui marque les savoirs scientifiques par rapport à d'autres savoirs, notamment les savoirs pratiques et les doctrines.

Buts immédiats de la recherche scientifique

Qu'elle soit entreprise seule ou menée de concert avec d'autres recherches d'un programme cadre, chaque recherche a un but global et des buts ou objectifs immédiats et spécifiques. Le but global, bien sûr, est de faire progresser l'étude d'un problème, de faire « avancer les connaissances » sur un phénomène donné. À ce titre, la découverte d'un agent causal ou d'une relation causale, dans un contexte donné, serait hautement satisfaisante. La mise au point d'une nouvelle méthode de mesure (d'un phénomène), la réfutation d'une théorie lacunaire ou erronée, la synthèse des résultats expérimentaux accumulés dans un domaine, constituent aussi des jalons significatifs dans l'avancement des connaissances.

Quant aux buts immédiats d'une recherche particulière, ils peuvent consister à :

- démontrer (prouver) expérimentalement un lien de cause à effet entre un facteur supposé d'un phénomène et une variable (un aspect) du phénomène étudié ; mettre à l'épreuve une prédiction déduite d'un modèle théorique ; faire le test d'une « vérité » déclaratoire (*i.e.* une hypothèse couramment acceptée sans fondement scientifique, et souvent considérée comme la conception naturelle d'un phénomène, sans caractère argumentatif ou hypothétique) ;
- démontrer l'existence d'une relation sérieuse (régulière, significative) entre deux ou plusieurs variables relatives au phénomène étudié ;
- identifier le phénomène étudié en dégagant ses aspects ou ses composantes les plus importants ou les plus prometteurs ;
- mettre sur pied ou améliorer les méthodes et procédés d'étude d'un phénomène.

Dans ces différents buts, on reconnaîtra des démarches de recherche correspondantes, soit, dans l'ordre : la recherche expérimentale et pseudo-expérimentale ; la recherche corrélationnelle ; la

recherche descriptive ; la recherche-développement (ou recherche méthodologique). Dans chaque cas, aussi, le chercheur prétend à une certaine généralité : les démonstrations et les effets expérimentaux, sans forcément réussir à tous les coups, doivent tenir la route et corroborer notre explication la plupart du temps. De plus, même si nous les avons énoncés positivement, ces buts sont aussi bien négatifs, puisque les conclusions d'une recherche peuvent nous conduire à rejeter une théorie ou une prédiction théorique, à déterminer dans quels cas ou dans quelles conditions un modèle théorique devient erratique ou erroné, etc.

UNE TAXINOMIE DES DÉMARCHES DE RECHERCHE

Principes de classement

Chaque recherche, d'une certaine façon, est unique : soit que la démarche utilisée en est originale, soit qu'elle diffère d'une recherche antécédente par un raffinement, un changement distinctif, soit que, tout en étant pareille à une précédente, elle en constitue une réplique et vise à en confirmer les conclusions. Toujours est-il que, dans les innombrables recherches publiées dans les revues scientifiques ou présentées dans les congrès, nous remarquons des similitudes d'une sorte ou d'une autre, et ce, quels que soient les domaines de recherche ou les disciplines académiques des chercheurs. Il est donc possible et utile de dégager ces similitudes afin d'identifier des schémas types de recherche.

Le but d'une classification de la recherche en des schémas types n'est pas d'imposer ces schémas en modèles, sous la fausse allégation que ce qui déroge à un schéma type n'est pas de la bonne recherche. C'est tout le contraire puisque, à notre avis, *toute bonne recherche consacre ses méthodes*. Le but premier de la classification est didactique et revient à fournir à l'apprenti-chercheur et au

chercheur lui-même un cadre conceptuel et un langage (un vocabulaire) qui leur permettent de réfléchir, discuter et communiquer leurs idées à propos du métier de chercheur.

Différents principes ou critères peuvent être avancés dans le dessein de classer les démarches de recherche :

- le lieu de réalisation : en laboratoire, sur le terrain (ou en milieu naturel) ;
- la finalité : recherche fondamentale (à finalité de connaissance), appliquée (à finalité de profits ou de bénéfices humains d'ordre général), industrielle (à finalité de profits ou de bénéfices privés) ;
- le « matériel » employé : recherche sur des sujets humains, animaux ou virtuels, avec des techniques d'observation quantitatives ou qualitatives, utilisant du matériel synthétique, fabriqué ou naturel, etc. ;
- le degré de contrôle exercé : modalité de recherche sans aucun contrôle jusqu'à un contrôle maximal sur la situation de recherche.

Nous retenons ce dernier principe, celui du « degré de contrôle », pour ordonner les démarches de recherche les plus courantes. À la limite inférieure, où très peu de contrôle est exercé par le chercheur, nous trouvons la recherche descriptive et ses sous-catégories. À la limite supérieure, marquée du contrôle le plus grand possible, nous trouvons la recherche expérimentale. D'autres démarches, hors taxinomie, sont aussi mentionnées, soit la recherche-développement, la recherche théorique et la recherche-action.

Description sommaire des schémas types de recherche

Les alinéas suivants, dont une partie correspond à notre taxinomie du degré de contrôle, déclinent les principales caractéristiques des schémas types de recherche que nous proposons, soit : la recherche

expérimentale, pseudo-expérimentale, corrélationnelle, évaluative, descriptive (et éthologique), la recherche-développement, la recherche théorique. Nous mentionnons aussi la recherche-action.

Recherche expérimentale et pseudo-expérimentale

Exemple. Un chercheur veut déterminer le rôle des sucres dans le métabolisme et l'endurance musculaires. Pour ce faire, il sélectionne un groupe de 16 hommes de 25 ans, en bonne santé, et il les soumet à un exercice intense de niveau sous-maximal à deux reprises, chaque fois deux heures après un petit déjeuner contrôlé. L'exercice, sur cycle-ergomètre, est poursuivi jusqu'à épuisement, la durée d'effort étant la mesure retenue. Par un cathéter installé au bras préféré, le sujet reçoit de l'eau saline isotonique. Dans la condition expérimentale, la solution saline est additionnée de glucose à 10 %. Les deux conditions, effectuées à une semaine d'intervalle, sont faites dans l'ordre « sans glucose, avec glucose » pour huit sujets, et dans l'ordre inverse pour les huit autres. Le chercheur émet l'hypothèse que l'addition de glucose va contrer sa déplétion normale en cours d'effort et prolonger l'effort, occasionnant une durée d'exercice plus grande sous la condition « avec glucose ».

Problématique. Puisée principalement dans la documentation scientifique et inspirée de la tradition d'expérimentation dans le thème de recherche, la problématique doit produire une hypothèse de recherche explicative (en vertu d'un modèle théorique) et causale, de forme schématique $VD = f(VI)$ (la forme étant enrichie au besoin).

Dans notre exemple, l'effort musculaire et le travail constituent le thème de recherche. La variable dépendante (VD), l'endurance musculaire, est concrétisée par la durée chronométrée de l'exercice (terminé à l'épuisement du sujet). Le régime d'apport veineux fournit la variable indépendante (VI), avec sa variante expérimentale (l'addition de glucose dans la saline) et sa variante neutre (la saline seulement). D'autres variables sont explicitement et implicitement contrôlées.

Méthodes. Le procédé est dit spécifiquement expérimental lorsque c'est le chercheur lui-même qui, dans le cadre de l'expérimentation, fait varier la variable indépendante (VI) et impose les conditions d'influence potentielle sur le phénomène étudié, observé par la variable dépendante (VD). Le procédé sera pseudo-expérimental si les différentes variantes de la VI, les conditions comparées, préexistent à l'expérimentation ou sont simplement invoquées (plutôt que provoquées) par le chercheur. Ce serait le cas si, par exemple, au lieu d'apprécier l'effet d'un apport externe de glucose chez des sujets sains, on comparait plutôt des sujets sains (sans apport externe) à des sujets hyperglycémiques équivalents.

La démarche sera donc dite expérimentale si la comparaison des conditions expérimentales ou des traitements se fait dans un contexte d'expérience contrôlée et en employant des groupes de sujets proprement constitués et échantillonnalement équivalents. On désignera plutôt de « quasi expérimentales » les études (qu'elles soient de *procédé expérimental* ou non) qui emploient seulement un ou quelques sujets, qui par exemple ne retiennent pas de groupe témoin, qui n'assurent pas l'équivalence échantillonnale des groupes comparés, qui se font dans un contexte « réaliste » plutôt qu'expressément contrôlé, etc.²

Produit. Le produit de la recherche expérimentale est une sanction d'un modèle explicatif du phénomène. Le modèle (ou théorie) qui sous-tend l'hypothèse de recherche est partiellement confirmé ou infirmé avec celle-ci.

2. L'ouvrage classique sur les protocoles dits quasi expérimentaux est celui de Campbell et Stanley (1966). Il reste qu'il y a un certain flottement terminologique sur ces désignations. Ainsi, P. Mercier et M. Gagnon (2005, p. 87-148) appellent « quasi » nos protocoles « pseudo » expérimentaux (ceux pour lesquels le chercheur exploite une variante expérimentale déjà existante chez ses sujets), et « pré » la plupart de nos protocoles « quasi » (ceux pour lesquels on n'a pas de groupe comparatif, ou de mesure préalable, etc.). Cependant, quels que soient les termes de vocabulaire appliqués, ce sont les distinctions opératoires entre les protocoles eux-mêmes qu'il importe de bien saisir et identifier.

Une recherche expérimentale (ou pseudo-expérimentale) bien faite peut néanmoins ne pas être concluante en raison de sa confection méthodologique, qui peut donner lieu à un manque de sensibilité expérimentale. La faute doit être imputée le plus souvent aux différentes opérationnalisations et aux procédés de mesure ou de contrôle utilisés (choix du caractère mesuré, précision ou sensibilité de l'instrument de mesure, méthode de contrôle de la VI, sélection et taille des échantillons, etc.).

Recherche descriptive (éthologique, évaluative, corrélationnelle)

Exemple. Les autorités d'une commission scolaire d'une grande ville sont préoccupées par le haut taux de « décrochage scolaire » à l'école secondaire (au-delà de 30 %) et elles commanditent une équipe de chercheurs universitaires pour les éclairer sur le problème. Y a-t-il plus d'une catégorie de décrocheurs ? De quels degrés scolaires sont-ils et quel était leur niveau de réussite scolaire ? Leur intégration à la vie de l'école est-elle différente de celle des autres élèves ? Quelle est la socioéconomie de leur école, le type d'occupation, le niveau d'instruction et de revenus de leurs parents ? Quelles raisons donnent-ils de leur décrochage ? Quel est leur mode de vie post-décrochage ? Le mandat de l'équipe est, bien sûr, de cerner et tenter de comprendre le phénomène du décrochage, et de trouver, si possible, des pistes afin d'aider les jeunes gens concernés à assurer leur avenir comme citoyens de la manière la plus efficace.

Problématique. Le but d'une recherche descriptive est de structurer l'espace des variables reliées à une question de recherche, et ce, à divers niveaux. Si la recherche est (explicitement ou implicitement) corrélationnelle, elle vise et contribue aussi à proposer des esquisses de relations explicatives, susceptibles d'investigation ultérieure. Cette démarche de recherche, basée sur une motivation et un état de la question documenté de toutes sources, n'a pas d'hypothèse

déterminante (ou restrictive) : elle a des objectifs, tels l'établissement d'une description critériée du phénomène, l'inventaire de ses conditions d'apparition, etc.

Méthodes. La problématique issue de la question de recherche peut réclamer une étude à caractère sociodémographique, auquel cas l'échantillonnage doit être représentatif et constitue une préoccupation majeure. La qualité des instruments de mesure (et leur justesse normative, le cas échéant) est aussi déterminante.

L'étude d'un cas unique n'est pas à proscrire ici (alors qu'elle est impossible en recherche expérimentale) puisqu'elle peut faire apparaître les variables souhaitées. La possibilité de généraliser les données recueillies est alors annulée.

Produit. Une recherche descriptive réussie donne aux chercheurs une image clarifiée des variables importantes dans un thème de recherche : en quoi consiste le phénomène, ce qu'il faut mesurer, les conditions ou contextes d'apparition qu'il faut surveiller. Une recherche descriptive laborieuse et abondante peut néanmoins être improductive si on l'a orientée prématurément dans un sous-espace théorique, en vertu d'une hypothèse favorisée par le chercheur : si on examine la lumière blanche à travers un filtre rouge, on ne verra que de la lumière rouge.

La démarche évaluative (dénotée « recherche *ex post facto* » par les Américains) consiste à mesurer après coup l'état d'un système pour apprécier les effets possibles d'une intervention passée : elle est structurellement apparentée à la recherche descriptive.

La recherche éthologique (Beaugrand, 1988b ; Coutu, Provost, Bowen, 2005), inspirée des études d'observation des mœurs animales initiée entre autres par Konrad Lorenz (1903-1989), est basée sur l'observation directe, non instrumentale, des sujets, animaux ou humains ; elle est, en principe, non intrusive. Cette recherche peut toutefois déboucher sur des modèles explicatifs rigoureux, voire sur des tests expérimentaux.

Recherche-développement

Exemple. Férue d'une conception théorique récente liant l'équilibre et les phénomènes de posture à des substrats psychophysiologiques de l'intégration temporelle et du rythme, via le cervelet, une chercheuse met sur pied un programme de rééducation posturale pour des patients traumatisés crâniens. Les exercices sont une séquence de mouvements rythmiques, avec appui musical, d'abord en position assise, puis debout avec support des bras, puis en pas de danse sans support. Des mesures d'équilibre dynamique et de stabilité posturale guident l'intervention. La chercheuse espère ainsi mettre à profit ses idées sur les interactions sensori-motrices en même temps qu'elle en vérifie l'efficacité thérapeutique.

Problématique. La « recherche-développement », pour n'être pas un seul « développement » et mériter le titre de démarche de recherche scientifique, doit être orientée vers l'accroissement de connaissances et l'explication des phénomènes. Aussi, la problématique, le plus souvent inspirée d'un besoin exprimé, voire commandée de l'externe, doit émerger des documentations scientifiques assorties aux différents thèmes (ou domaines) de recherche touchés et donner lieu à un système conceptuel (partiellement) enraciné dans ces documentations.

Méthodes. Les interventions à faire suivent les phases du développement. Une séquence proposée est : 1) motivation et énoncé de l'idée du « produit » à développer ; 2) identification des fonctions du produit et des contextes d'utilisation anticipés ; 3) détermination des contraintes de toute espèce (même matérielles) dans l'élaboration et l'utilisation anticipée du produit ; 4) élaboration d'un modèle conceptuel justifié du produit, assorti d'hypothèses de travail ; 5) mise au point d'un prototype du modèle ; 6) mises à l'essai du prototype ; 7) conclusions (sur le prototype, le modèle, les hypothèses de travail). Les mises à l'essai (fonctionnelle, empirique, systématique par recherche évaluative) restent, normalement, de modeste envergure.

Produit. L'étape 4 ci-dessus, ou ce qui en tient lieu, est réalisée créativement et en réunissant pour la première fois un jeu de concepts, de théories, de « pensées », un faisceau d'hypothèses de travail plus ou moins inventées, qui sont préalables et nécessaires à l'élaboration. Advenant le cas où le modèle théorique et le prototype s'avèrent intéressants, lesdites hypothèses se trouvent accréditées empiriquement et représentent alors, pour la science, le produit le plus valable de cette démarche.

Recherche théorique

Exemple. La synthèse théorique à laquelle réfère « la théorie de la relativité » restreinte d'Albert Einstein, et qui découle notamment de son postulat sur la constance universelle de la vitesse de la lumière, est l'exemple par excellence d'une recherche théorique, « sur papier ».

Problématique. Abondance et confusion marquent un thème de recherche, où pullulent les publications, les faits, notamment les faits inexplicables ou les explications et modèles explicatifs contradictoires ou disparates. Cette situation étant constatée, le but de la recherche théorique (en analyse ou en synthèse) est de « faire le ménage » en révisant les bilans empiriques, en revalidant les interprétations, en épouillant les théories et les modèles, en les confrontant et, si possible, en proposant un modèle nouveau qui synthétise tous les faits.

Méthodes. L'érudition et l'intelligence sont les instruments principaux pour cette démarche. Des étapes possibles d'une recherche en synthèse théorique sont les suivantes : 1) proposition du modèle nouveau, avec ses axiomes et ses assises empiriques principales ; 2) confrontation critique du nouveau modèle et de chacun des modèles concurrents ; 3) démonstration du pouvoir conciliateur du nouveau modèle (expliquant les différences entre les modèles antérieurs, assimilant les données laissées-pour-compte par les autres

modèles, etc.) et de sa parcimonie ; 4) démonstration du pouvoir créatif du nouveau modèle, par son étendue explicative plus grande ou sa prédiction de faits jusqu'alors négligés ou ignorés. Dans certains cas, l'étude expérimentale du nouveau modèle peut être simulée sur ordinateur, grâce à la méthode dite de Monte Carlo (Bratley, Fox et Schrage, 1987 ; Laurencelle, 2001).

Les techniques de la méta-analyse³ peuvent aider à l'intégration des données et conclusions émanant d'une série d'études, qu'il y ait ou non proposition d'un nouveau modèle.

Produit. Même si aucun modèle nouveau n'est proposé ou si le modèle proposé n'est pas retenu, le travail intellectuel considérable de sa méditation et le défi posé restent un produit important pour les artisans d'un domaine de recherche, lesquels ne prennent ordinairement pas le temps de se former un point de vue d'ensemble sur leurs activités et leurs résultats de recherche.

Recherche-action

Exemple. Une grande entreprise privée, employant plusieurs centaines de personnes, est affligée d'un taux d'absentéisme chronique très élevé. Pour contrer le phénomène, on met sur pied un régime de bonus par lequel les employés reçoivent un supplément de salaire pour chaque jour de congé de maladie non utilisé. Deux chercheurs universitaires sont engagés (un psychologue et un ingénieur en ergonomie) ; ils s'adjoignent un employé délégué par le syndicat et l'assistant du chef du personnel, afin d'implanter le régime et voir à son rodage durant un an. Le « comité d'implantation » doit faire des recommandations au bout d'un an.

La démarche dite de recherche-action est d'abord une intervention savante (ou savamment inspirée), à l'organisation de laquelle participent des chercheurs désignés et qui s'effectue en partenariat

3. L'ouvrage de F.M. Wolf (1986) fournit une bonne introduction et une bibliographie.

et en concertation avec les agents du milieu où se situe l'intervention. La priorité n'est pas à la connaissance mais à la solution, l'amélioration d'une situation problématique. C'est un bain d'expérience pour le chercheur, sans l'objectivité requise dans une recherche et sans possibilité de généralisation. Ce n'est pas, à proprement parler, une démarche de recherche scientifique⁴.

DISTINCTIONS ENTRE LES DÉMARCHES DE RECHERCHE ET FORMES MIXTES

Il n'est pas toujours facile ni nécessaire de décider si une recherche particulière est d'un type ou d'un autre. Toutefois l'exercice académique est intéressant et peut être formateur. Le classement d'une recherche n'affecte pas sa valeur. Plus bas, nous considérons aussi ce qu'on a coutume d'appeler « recherche clinique » et « recherche épidémiologique ».

Expérimental, pseudo-expérimental, corrélationnel. Comme le font voir les schémas types plus haut, les modes « expérimental » et « pseudo-expérimental » diffèrent par la nature de leur VI, celle-ci étant manipulée *actuellement* et imposée par le chercheur dans la situation de recherche expérimentale, alors que la situation pseudo-expérimentale hérite des différents états de la VI, qui *préexistent* à l'exécution de la recherche. La différence entre démarche pseudo-expérimentale et corrélationnelle tient, quant à elle, aux intentions du chercheur : celui-ci endosse tout de même un modèle causal en recherche pseudo-expérimentale (même s'il ne peut l'actualiser *in vivo*), alors qu'il n'est en quête que de régularités et dépendances mutuelles entre divers aspects d'un phénomène, en recherche corrélationnelle. Notons que « recherche corrélationnelle » n'implique pas nécessairement l'utilisation d'un coefficient de corrélation, l'adjectif étant entendu au sens large et signifiant toute espèce de

4. Plusieurs auteurs ont écrit sur ce type d'intervention, notamment Goyette et Lessard-Hébert (1987).

relation ou de dépendance entre deux ou plusieurs variables (telle qu'en évaluent l'analyse de variance, le test t , le Khi-deux sur un tableau de fréquences, le coefficient de corrélation r , etc.).

Expérimental, évaluatif. La démarche évaluative, de pratique courante, consiste à faire l'état de situation actuel, consécutif à une intervention passée et qui se situe en deçà de l'horizon de la recherche⁵. C'est donc une démarche quasi pseudo-expérimentale (quasi, parce qu'elle ne bénéficie pas des contrôles nécessaires, notamment sur l'échantillon et sur les conditions de mesure des sujets, et pseudo, parce que l'« avant » précédait la prise en charge par le chercheur), donc une recherche de format « avant-après » à laquelle il manquerait le « avant » !

Étude de cas, recherche-action. L'étude de cas est, au titre de démarche de recherche, désintéressée, critique et objective. Au contraire, l'intervention professionnelle dénommée recherche-action est intéressée, compromise avec le milieu d'étude, particulariste (plutôt que généralisante) et non critique (au sens d'une attitude de critique scientifique). Sur le plan des objectifs, la recherche-action vise d'abord le bénéfice du commanditaire, l'amélioration de la situation analysée, l'étude de cas (singulier ou non) s'adressant d'abord à la communauté scientifique et se prémunissant contre une trop forte osmose dans chaque situation examinée.

Recherche clinique. Il existe presque autant de formes de « recherche clinique » qu'il y a de démarches de recherche, le caractère distinctif tenant à la population retenue, soit des personnes symptomatiques ou atteintes d'une maladie, d'une dysfonction ou d'une infirmité. Les « essais randomisés », dans lesquels les patients sont assignés au hasard soit dans un groupe à traitement

5. La recherche évaluative est à distinguer de la *recherche longitudinale*, qui répond au schéma de l'étude descriptive étalée dans le temps et dont le but est de décrire l'évolution d'une cohorte de sujets évalués périodiquement, et de l'*étude de suivi*, par laquelle un chercheur reprend les sujets et informations constitués dans un moment de recherche antérieur, possiblement dans le cadre d'une recherche expérimentale, et en actualise les données.

expérimental, soit dans un groupe de contrôle, sont un exemple de démarche expérimentale (ou pseudo-expérimentale), alors que d'autres types de recherche, à contrôle plus flou et à participation plus créative du chercheur-intervenant, ressemblent à la recherche-action. En fait, l'expression « recherche clinique » définit d'abord un terrain (ou une population) plutôt qu'une forme spécifique de démarche de recherche.

Recherche épidémiologique. L'épidémiologie s'intéressant, à l'origine, l'étendue et l'expansion des maladies et de ses facteurs dans les populations, la « recherche épidémiologique » couvre, elle aussi, un vaste champ et ne refuse aucune démarche. L'étiologie (la recherche des causes et des facteurs des maladies) en fait partie, qu'elle soit descriptive ou (pseudo) expérimentale. Les grandes enquêtes descriptives, la mise sur pied de modèles mathématiques de diffusion, l'analyse in vitro d'échantillons prélevés sur des sites représentatifs... En somme, nonobstant certains auteurs, il n'y a pas de « recherche épidémiologique » comme telle, mais une discipline et des chercheurs qui, comme dans les autres disciplines, font de leur mieux pour atteindre leurs objectifs avec efficacité et rigueur.

VALIDITÉ INTERNE ET EXTERNE D'UNE RECHERCHE

Les concepts de validité en science

Le concept de « validité » est polysémique en science. La *validité de mesure* constitue une acception spécifique, qui elle-même se ramifie en plusieurs sous-espèces (voir « Choix des méthodes et des instruments de mesure », p. 31) ; la *validité (globale) d'une recherche* est une autre catégorie, qui englobe la « validité interne » et la « validité externe » comme sous-espèces principales. L'emploi du mot « validité » doit obligatoirement être mis en contexte. Robert (1988) et Reid (2005) s'étendent sur les types de validité et d'invalidité d'une recherche.

Validité interne d'une recherche

Le bilan de la recherche effectuée nous permet-il d'avoir confiance dans les résultats, ces derniers ont-ils une interprétation sûre et peuvent-ils être reproduits à volonté ? De telles questions fondent la validité *interne* d'une recherche.

La validité interne met en question l'organisation même de la démarche de recherche. La plupart des procédés, prescriptions et moyens qu'on retrouve en méthodologie ont trait à ce problème. Un jury de thèse, le comité éditorial d'une revue scientifique qui reçoit un projet d'article, un comité de pairs donnant un avis aux fins de subvention ont pour première tâche de juger les qualités qui fondent la validité interne de la recherche effectuée. Parmi les qualités et facteurs susceptibles de compromettre la validité interne, nommons seulement :

- la non-équivalence des groupes comparés ;
- la contagion expérimentateur/sujet ou sujet/sujet, l'effet placebo, l'effet cadre de l'expérimentation (effet Hawthorne) ;
- les effets de séquence (dans un protocole à mesures répétées) ;
- les effets de gamme (biais adaptatif des sujets à l'étendue des conditions d'expérimentation, dans un protocole à mesures répétées) ;
- la défection des sujets ;
- la pauvreté des procédés de mesure ;
- la piètre qualité des manipulations ou contrôles effectués ;
- l'incompétence exécutive (dans l'expérimentation, la mesure, le traitement de données) ;
- le manque de minutie et de rigueur de l'interprétation des résultats (confrontation des variables, relation avec les données publiées, etc.).

Validité externe d'une recherche

Les conclusions de la recherche effectuée s'appliquent-elles intégralement au thème de recherche, peut-on les étendre au « monde réel », ou n'ont-elles de valeur que dans le cadre du laboratoire ou d'un échantillonnage artificiel ? La validité *externe*, qui évolue autour de ces questions, donne lieu à deux sous-concepts, ceux de validité *échantillonnale* (les groupes de participants retenus dans l'étude représentent-ils bien la population à laquelle s'adresse le thème de recherche ?) et de validité *écologique* (les conditions et procédures de l'expérience reflètent-elles pertinemment la ou les situations du monde réel dont émane le thème de recherche ?).

En plus des deux facteurs (l'échantillon, la pertinence écologique) déjà cernés par les concepts de validité échantillonnale et écologique, on peut énumérer d'autres facteurs qui affectent la validité externe d'une recherche. L'argument clé est cependant celui-ci : comment peut-on assurer que des résultats, obtenus dans un contexte tel que celui de l'expérience réalisée, s'appliquent généralement hors de ce contexte ? L'argument est certainement épineux en sciences humaines, *i.e.* pour des problématiques cognitives/psychologiques avec des sujets humains, qui ne se contentent pas de subir les procédés d'expérimentation qu'on leur propose, mais interagissent avec eux, les interprètent et en devinent, justement ou non, l'intention.

PROCÉDURE DE LA RECHERCHE EXPÉRIMENTALE

QUESTION DE RECHERCHE ET DOMAINE DE RECHERCHE

Toute recherche se situe dans un *domaine de recherche* et est spécifiquement centrée sur une *question de recherche* : le projet de recherche est déjà bien avancé lorsque question et domaine de recherche ont été identifiés. D'où l'idée de recherche originale ?

Au départ de l'entreprise de recherche, il peut y avoir une *préoccupation de recherche*, l'intérêt de quelqu'un pour un sujet, un problème particulier, ou bien une commande (une « commandite ») de recherche, qui fixe le sujet d'étude. Ce sujet est lié ou, en général, est connexe à d'autres sujets, questions et préoccupations, qui ensemble constituent le *thème de recherche*.

C'est après avoir choisi le sujet et délimité le thème de recherche que commence le vrai travail du chercheur. La préoccupation de recherche (individuelle ou reçue de l'extérieur) se traduira d'emblée comme une question de recherche dans le thème identifié. Il s'agira alors de construire la *problématique de recherche*, par un effort d'exploration, de traduction, d'identification, d'opérationnalisation.

ÉLABORATION DE LA PROBLÉMATIQUE DE RECHERCHE

Recrutement des disciplines, secteurs et documentations touchés par le thème de recherche. Parfois, le thème de recherche sera unidisciplinaire, le domaine de recherche découlant alors directement du thème ; parfois, il sera multidisciplinaire, et le chercheur devra alors s'efforcer d'identifier clairement le domaine propre, à rattachements multiples, dans lequel sa problématique pourra s'ancrer et s'épanouir.

Compilation des études pertinentes parues, en considérant chaque fois : la question de recherche, les conditions manipulées, le protocole de recherche, les caractères observés et leurs méthodes de mesure, l'échantillonnage, l'instrumentation employée, les variables contrôlées, les méthodes statistiques, les conclusions, les difficultés déclarées ou constatées.

La problématique d'une recherche n'est pas connue d'emblée : c'est la question de recherche, le besoin, la préoccupation exprimée qui l'est. Or, puisque la recherche à faire est un travail, une opération systématique et de longue haleine, elle doit être pensée, rattachée à ses bases et planifiée afin d'en assurer la rigueur, l'efficacité, la crédibilité. Bien qu'on ne puisse guère en éclairer le processus, l'élaboration de la problématique constitue une étape nécessaire, en partie souterraine, par laquelle le chercheur prend intellectuellement possession de la question de recherche et la reformule dans le langage du domaine et du thème de recherche appropriés. Et, parmi les fruits principaux de cette élaboration, le chercheur aura veillé à :

- la détermination des aspects du phénomène retenus pour étude, le choix des facteurs ou conditions à considérer et à contrôler, le choix de la démarche de recherche (principale), la définition des variables (voir la section suivante, p. 26) ;
- l'adoption ou la délimitation d'un modèle explicatif, cadre conceptuel ou conception théorique, dans lequel ou laquelle va s'inscrire la question de recherche ;

- la reformulation précise de la question de recherche dans le langage des variables du modèle adopté, en appliquant au besoin le principe d'*opérationnisme*.

FORMULATION DE L'HYPOTHÈSE DE RECHERCHE OU DE L'OBJECTIF DE RECHERCHE

Nature de l'hypothèse de recherche

Toute démarche sérieuse se fait en vue d'un objectif ou de quelques objectifs, de même qu'on l'entreprend avec une idée préconçue de ce qui pourrait advenir, de ce qu'on voudrait obtenir : il ne faut pas confondre ces hypothèses de travail, de nature générale et partiellement subjective, et l'« hypothèse de recherche » qui, strictement parlant, est l'apanage de la recherche expérimentale et en fournit l'objectif. C'est à ce type de recherche que nous référons ici expressément.

L'hypothèse de recherche, dans sa formulation définitive, représente le modèle explicatif retenu et annonce à la fois l'expérimentation à faire, dans ses grandes lignes. C'est dire qu'elle est l'aboutissement d'un long processus constructif (l'élaboration de la problématique) et que, de préoccupation plus ou moins bien verbalisée au départ, elle doit progresser vers une expression qui joint justification théorique et conditions matérielles d'expérimentation.

L'hypothèse de recherche, qui tient en une phrase ou quelques phrases, donne en même temps une expression simplifiée, compacte, du modèle explicatif ou du cadre conceptuel sur lequel repose toute l'entreprise. Toutefois, la formulation parfois lapidaire de l'hypothèse de recherche peut masquer le caractère incomplet, mal articulé, du raisonnement du chercheur, voire quelquefois l'absence totale d'un raisonnement. Rappelons qu'à l'hypothèse de recherche s'appliquent les mêmes principes généraux qu'à toute démarche de

connaissance scientifique, notamment le déterminisme (strict ou statistique), la parcimonie (ou principe du « rasoir d'Occam »), l'opérationnisme.

Apanage de la recherche expérimentale (ou pseudo-expérimentale), l'hypothèse de recherche n'est pas requise pour les autres démarches de recherche ; elle peut même y être malvenue et encombrante. Dans certains cas, par exemple en recherche descriptive, le chercheur la donne comme son hypothèse de travail, *i.e.* comme ce qu'il anticipe observer ou inférer au terme de sa démarche. Il reste que toute démarche intelligente est ordonnée à un but, poursuit un ou plusieurs objectifs, objectifs que le chercheur devra préciser afin de répondre à la question de recherche tout en clarifiant sa problématique.

Variables en jeu

La science pourrait être définie sommairement comme l'étude des variations. La nature et ses phénomènes, dans les règnes minéral, végétal et animal comme dans le monde culturel humain, nous transmettent de l'information par la *variation*¹. C'est pourquoi les modèles que nous tentons de construire pour comprendre, expliquer et contrôler la nature sont dotés d'aspects et de facteurs collectivement appelés « variables ».

-
1. À l'instar de certains prédateurs aux yeux desquels la proie immobile cesse d'exister, les scientifiques sont d'abord et surtout à l'affût de la variation, des changements, des « phénomènes » dont la propriété première est de survenir. Si l'eau d'une rivière garde sa température constante, l'étude de cette température ne nous apprend rien à propos de ce qui se passe dans la rivière. Si, malgré tout, la rivière abrite des phénomènes particuliers grâce à sa température constante, nous pouvons les découvrir en constatant une différence, une variation, entre la température particulière de la rivière et les fluctuations de température observables dans les autres cours d'eau. Même s'il recherche des régularités et des redondances dans la nature, le chercheur se nourrit d'abord de variation et s'efforce de la décrire, la comprendre, la contrôler.

On peut distinguer trois grandes classes de variables²: les variables dépendantes, les variables déterminantes (ou conditions du phénomène), qui se subdivisent elles-mêmes en variables indépendantes et variables contrôlées, et enfin les variables d'expérimentation (ou procédurales).

La variable dépendante (VD). Nous étudions un phénomène, précisément un aspect d'un phénomène (p. ex., la chaleur d'un muscle à l'effort): la fonction de mesure (ou la mesure) de cet aspect constitue la VD. Dans certaines études, il est fréquent que le chercheur s'intéresse en même temps à plusieurs VD, qui dénotent différents aspects d'un même phénomène (p. ex., chaleur et lactatémie du muscle) ou encore qui réfèrent à différents phénomènes (p. ex., la chaleur du muscle et le débit cardiaque).

La variable indépendante (VI). Dans une étude expérimentale, le chercheur veut tester ou valider une relation causale supposée qui associe un facteur supposé du phénomène à un aspect de celui-ci. Pour faire ce test, le chercheur va manipuler les conditions d'expérimentation en faisant varier lui-même une variable déterminante pour observer ensuite les effets potentiels sur le phénomène. C'est ce facteur manipulé, que le chercheur a élu parmi les conditions déterminantes, qu'on appelle spécifiquement variable indépendante (VI).

La VI proprement dite représente donc le facteur causal à l'étude, dans une recherche expérimentale. Dans certains cas, on l'a dit, particulièrement pour la recherche avec des sujets humains, il est difficile, voire éthiquement impossible au chercheur de manipuler lui-même la condition déterminante qui l'intéresse (p. ex., l'inoculation d'un agent pathogène, l'imposition du sexe du sujet, de son âge); le chercheur n'a alors d'autre recours que de procéder par variation échantillonnale, en sélectionnant des sujets qui présentent

2. Les termes « variable indépendante » et « variable dépendante », dans la nomenclature en usage, émanent de la théorie mathématique des fonctions. Dans l'expression schématique $Y = f(X)$, le terme X est dénoté variable indépendante (ou VI), et le terme Y, variable dépendante (VD). Il serait hasardeux, à notre avis, d'abuser de cette analogie purement formelle en suggérant, par exemple, que la VD *dépend* de la VI (alors que la preuve en reste à faire) tandis que la VI *dépendrait* de l'expérimentateur, une suggestion malheureuse qu'on trouve dans certains manuels.

d'emblée différents niveaux de la condition expérimentale étudiée. L'étude devient alors pseudo-expérimentale, et la VI, qu'on appelle encore de ce nom, ne répond plus strictement à sa définition.

Les variables contrôlées (VC). La problématique a mis au jour plusieurs variables déterminantes, réelles ou potentielles, du phénomène. Or, l'étude envisagée concerne une seule VI, que le modèle explicatif retenu rattache au phénomène observé. Si, en fait, d'autres conditions sont présentes, elles vont interférer dans la relation VI → VD qui nous intéresse, et il convient de les contrôler, du moins en partie. Ces variables déterminantes, situées en périphérie du modèle proposé par l'hypothèse de recherche, constituent alors des variables contrôlées. Divers modes de contrôle sont possibles : voir p. 36.

Les variables d'expérimentation (ou procédurales). Lors de l'exécution expérimentale et dans l'application du protocole d'expérience, tout ce qui ajoute ou peut ajouter à l'effet du facteur expérimental ou interférer avec lui est à prendre en compte ici : les instructions données et la préparation des sujets, la répétition de la mesure chez les mêmes sujets, le contexte physique et temporel (circadien ou saisonnier) de l'expérimentation, son contexte psychologique (effet Hawthorne, effet placebo, effet Rosenthal), le choix des matériels, l'ambiance, la rigueur de l'expérimentation, etc. Les variables dites échantillonnales sont du lot : elles recouvrent à la fois les critères, le mode et le nombre de l'échantillonnage employé, toutes choses qui influencent certainement et peuvent biaiser ou invalider les résultats obtenus. Deux éléments de contrôle méritent une mention spéciale : l'équivalence échantillonnale des groupes et l'équilibre des conditions³ comparés, et les effets de séquence engendrés par les protocoles à mesures répétées. Nous y reviendrons.

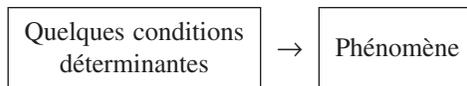
3. Les conditions expérimentales mises en comparaison diffèrent l'une de l'autre, bien sûr, mais elles ne doivent différer que par le facteur testé et doivent demeurer mutuellement équivalentes quant au reste. Par exemple, l'effet d'une drogue injectée à un animal sur sa performance pourra être comparé judicieusement en utilisant des animaux à qui on aura injecté de la saline isotonique (une substance neutre), ou l'effet d'un régime particulier d'entraînement musculaire (sur la force) devra être établi en comparant des situations à temps d'exercice égaux.

Les variables non contrôlées. Les auteurs parlent aussi de variables non contrôlées (VNC), soit des variables déterminantes du phénomène que le chercheur a négligé de contrôler ou qu'il a simplement oubliées, soit des variables procédurales ignorées lors de la mise sur pied du protocole. Ces variables se révèlent lors d'un retour critique sur l'expérience réalisée : on les invoque pour expliquer ou justifier une partie moins cohérente des résultats et, dans certains cas assez fréquents, elles peuvent même invalider les conclusions.

Formulation de l'hypothèse de recherche

L'hypothèse de recherche, mise en mots dans une phrase ou quelques phrases, est une déclaration affirmant une relation orientée entre VI et VD, dans le contexte d'un modèle explicatif, et laissant entrevoir les éléments principaux du protocole de recherche. La formule de l'hypothèse de recherche exprime le parti pris heuristique du chercheur en faveur d'une solution particulière de la question de recherche ; elle anticipe, elle interprète en quelque sorte à l'avance, les résultats de l'expérience, à l'exclusion des autres solutions.

À l'orée de la problématique de recherche, le chercheur envisage une relation, peut-être floue, entre certaines conditions déterminantes et le phénomène en général, soit :



Basée sur la problématique fournie, l'hypothèse de recherche, formulée dans le langage des variables et faisant référence à un modèle explicatif, peut être schématisée de façon quasi algébrique par l'expression :

$$f [VI | VC_1, VC_2, \dots] \rightarrow VD ;$$

cette forme souligne la relation privilégiée que le chercheur entend exploiter entre VI et VD, elle identifie les quelques variables déterminantes qu'il veut contrôler et, par la fonction f , elle indique spécifiquement le modèle théorique qui est mis à l'épreuve dans cette recherche.

Le concept d'« hypothèse de recherche » apparaît ailleurs, en méthodologie, dans le contexte quelque peu différent des tests d'hypothèses statistiques, dans lequel il dénote un changement réel dans les données, la modification d'un paramètre, plutôt qu'une fluctuation ou un effet du hasard. L'« hypothèse de recherche » est alors mise en opposition à l'« hypothèse nulle ». Le rejet éventuel de l'hypothèse nulle, grâce à une procédure ressortissant à un calcul de probabilité, confortera l'hypothèse de recherche, l'accréditera auprès du chercheur et de son auditoire. Cette acception particulière du concept d'hypothèse de recherche n'équivaut pas à celle réservée plus haut pour la recherche expérimentale.

Non seulement l'hypothèse de recherche est-elle l'affirmation d'une relation entre deux variables, d'un changement expliqué, de l'influence effective de la variable indépendante sur la variable dépendante : elle doit aussi être démontrée positivement par les résultats. Ainsi, une hypothèse formulée comme suit : « Les sujets soumis à la condition expérimentale auront des résultats semblables à ceux des témoins » (un énoncé qui revêt la forme typique de l'hypothèse nulle mentionnée ci-dessus) ne peut être démontrée expérimentalement de manière convaincante, et encore moins statistiquement. En un mot, l'hypothèse doit être conçue et formulée de façon telle que sa démonstration prenne une forme positive, se concrétise par une variation significative.

Opérationnisme et définitions opérationnelles

L'« opérationnisme » est une approche épistémologique des sciences proposée afin d'établir des descriptions et des communications universellement claires et non ambiguës. Les concepts, qu'un chercheur manipule mentalement, se réalisent finalement dans une forme concrète, au moment de l'expérimentation. L'approche opérationniste consiste donc à donner des définitions concrètes de ces concepts, notamment ceux de la VD et la VI, et de l'hypothèse de

recherche. Cette approche suggère d'exprimer ces différents concepts dans un langage dit opérationnel, qui fait explicitement référence aux moyens, méthodes et matériels mis en œuvre.

CHOIX DES MÉTHODES ET DES INSTRUMENTS DE MESURE

Diversité des mesures d'un même phénomène

Le chercheur veut déterminer quelle influence il peut exercer sur un phénomène en faisant varier les conditions et facteurs de son expérience. Pour cela, il lui faut choisir quel aspect du phénomène il va observer et par quel moyen il va le mesurer, deux choix critiques. L'hypothèse de recherche, notamment dans sa forme opérationnelle définitive, fixera ces choix.

L'aspect à observer. Chaque phénomène offre autant d'aspects à observer qu'il nous plaît d'en imaginer. Par exemple, la fatigue musculaire peut être observée à travers la force de contraction exercée, la lactatémie du muscle, la lactatémie du sang veineux ou artériel, une échelle verbale autodescriptive de fatigue, etc. On évalue le stress psychologique (ou l'anxiété) par l'activité cardiovasculaire, la performance dans une tâche d'attention, la tension musculaire passive, ou encore une échelle verbale autodescriptive, etc. La préférence du chercheur devra être guidée d'abord par l'adéquation entre l'aspect considéré et le thème de recherche, puis par la sensibilité (ou réactivité) expérimentale de l'aspect considéré.

Le procédé de mesure. Il existe, en général, plus d'une opérationnalisation possible d'un phénomène ou d'un aspect d'un phénomène. L'activité cardiovasculaire, par exemple, peut être *mesurée* par la fréquence cardiaque, la (haute) tension artérielle (au bras), le volume d'éjection systolique, etc. ; voire, la fréquence cardiaque admet elle-même différentes opérationnalisations. La performance dans une tâche, peut aussi être évaluée par le nombre d'erreurs commises, le temps pris pour compléter une unité de tâche, le nombre d'unités complétées, etc.

Critères pour choisir la bonne mesure

Au moment de faire son choix d'instruments de mesure, le chercheur doit considérer quatre points :

- la tradition expérimentale : que retrouve-t-on dans les recherches publiées, qu'est-ce qui s'est avéré payant, expérimentalement rentable ?
- la panoplie des instruments et procédés de mesure disponibles ou possibles : le chercheur, faut-il le redire, est celui dont la tâche consiste à faire avancer la connaissance, en inventant au besoin de nouveaux moyens. Il doit se soucier avant tout d'utiliser les moyens les plus efficaces d'observer et d'influencer le phénomène à l'étude, et il n'est pas incongru qu'il propose ou, au besoin, qu'il invente de nouvelles façons de faire.
- les procédés de mesure choisis doivent présenter des qualités de mesure démontrables, notamment⁴ :
 - *la fidélité* (ou précision) : toute mesure comporte une part d'imprécision, de fluctuation au hasard : la marge d'erreur (ou d'imprécision) autour d'une mesure est indiquée par l'erreur-type de mesure (σ_e) et est typique de l'instrument de mesure. La fidélité, notée par le coefficient r_{XX} (ou ρ_{XX}) tel que $0 < r_{XX} \leq 1$, indique jusqu'à quel point les différences mesurées d'un objet à l'autre (ou d'une personne à l'autre) reflètent des différences réelles entre ceux-ci.
 - *la validité* : même un instrument, une mesure précise peut refléter ou révéler autre chose que le phénomène qui nous intéresse. La validité, qui touche surtout les mesures à caractère interprétatif ou théorique (plutôt que les mesures directes), indique jusqu'à quel point les différences mesurées d'un objet à l'autre correspondent à des grandeurs du phénomène

4. La théorie et la pratique de la mesure en science sont documentées dans Laurencelle (1998).

visé. Par exemple, la valeur donnée par un (bon) pèse-personne, instrument à fidélité élevée, constitue une mesure valide (par définition) de la masse corporelle d'une personne, mais elle est d'une validité douteuse pour indiquer le degré d'obésité, voire de masse grasse, de la personne évaluée. Il existe différents sous-concepts opératoires de validité et différents procédés de validation (manifeste, échantillonnale, conceptuelle, factorielle, prédictive, concomitante, etc.). Dans le cas des instruments à mesure directe (p. ex., le pèse-personne), le concept de justesse se substitue à celui de validité et la validation se ramène alors à un calibrage.

- *la capacité discriminante*: la fonction de la mesure est d'indiquer numériquement les différences observables entre les objets ou personnes évalués. Un instrument, un procédé de mesure sera d'autant plus utile qu'il donne lieu à un nombre élevé de valeurs réellement différentes, qu'il permet un grand nombre de discriminations stables.

Dans l'annexe B, le lecteur trouvera un exposé sommaire sur la « théorie des tests » classique et sur les principaux concepts qui lui sont associés.

- les mesures obtenues doivent, si possible, être *bien distribuées*, *i.e.* convenir aux traitements statistiques envisagés. Dans la plupart des cas, le chercheur va recourir à des méthodes de traitement statistique classiques, comme le calcul de moyennes et écarts-types, et exécuter des tests de différences comme le *t* sur deux moyennes, l'analyse de variance, etc. Ces méthodes réclament des variables (ou mesures) continues et à distribution normale. Le chercheur devrait alors favoriser des mesures largement distribuées (p. ex., le temps d'exécution d'une tâche, en millisecondes) plutôt que catégorielles (réussite/échec) ou discrètes (nombre d'erreurs commises). De plus, des données documentées ou l'intuition personnelle devraient l'orienter vers des mesures à distribution symétrique avec un mode central, tel que dans la distribution normale.

Tout n'est pas perdu, toutefois, si les mesures retenues ne satisfont pas vraiment ces impératifs. Considérons en premier lieu que plusieurs tests statistiques à référence normale (comme le t de Student et l'analyse de variance) sont assez robustes pour accommoder sans souffrance de légères déviations par rapport au modèle de la loi normale; des déviations plus graves seront signalées de toute façon par les tests de précaution (p. ex., test sur l'homogénéité des variances intragroupe) qui accompagnent ces procédures. En second lieu, s'il s'avère que la mesure, ou variable, X ait une forme de distribution inadéquate, elle peut être transformée mathématiquement en une autre variable, disons X' , par un procédé quelconque, de type $f(X) \rightarrow X'$, la nouvelle variable se conformant davantage au modèle normal invoqué. En troisième lieu, même si les précautions et manœuvres suggérées sont infructueuses, *tout peut s'analyser* en trouvant éventuellement la méthode et le moyen adéquats pour le faire⁵.

CHOIX DES MÉTHODES DE CONTRÔLE ET D'EXPÉRIMENTATION

La situation de recherche constitue un lieu privilégié pour l'étude des phénomènes. D'une part, nous nous mettons en état d'attention respectueuse par rapport au phénomène observé, tout en nous équipant le mieux possible pour enregistrer scrupuleusement ses manifestations. D'autre part, nous isolons d'une façon ou d'une autre le phénomène en le plaçant dans un contexte et des conditions qui le

5. La méthode du « *scoring* normal », par exemple, consiste à remplacer chacun des résultats par une valeur à distribution normale imposée correspondant au rang centile du résultat dans l'ensemble : la littérature anglo-saxonne désigne ces scores normalisés (ou leurs espérances) par *rankits* (Sokal et Rohlf, 1981). Une approche plus générale consiste à obvier carrément à la contrainte distributionnelle en utilisant l'approche non paramétrique des tests par combinatoire (combinatoire exhaustive ou approximative) suggérée par R.A. Fisher (Edgington, 1980) et L. Laurencelle (2001).

protègent d'interférences et d'irruptions intempestives. Le contrôle de la situation de recherche constitue donc un *sine qua non* de la qualité d'une recherche, et le mot d'ordre donné au chercheur est « rigueur ».

Cet idéal de contrôle, qu'on retrouve en recherche expérimentale et qu'il faut y maintenir, n'est pas toujours réalisable ni même souhaitable dans d'autres modes d'investigation. Un chercheur clinique, par exemple, à qui sont confiés des patients présentant un syndrome nouveau à haute morbidité, va traquer les causes du problème et en poursuivre l'étiologie par tous les moyens possibles. L'impératif de la généralisation et de ses conditions requises cède alors le pas à l'urgence de la situation et à la débrouillardise du chercheur. De même, les recherches d'observation directe fuient pour ainsi dire les manœuvres de contrôle, lesquelles pourraient stériliser le phénomène étudié. Néanmoins, l'exigence de rigueur, dans le respect du contexte et des objectifs d'une recherche, reste toujours de bon aloi, et la nécessité du contrôle, ou d'un certain contrôle à tout le moins, devrait toujours habiter le chercheur.

Contrôle des conditions environnementales

La recherche expérimentale, menée ordinairement en « laboratoire » (*i.e.* dans un milieu artificiel, sous le contrôle quasi absolu du chercheur) ne pose guère de problèmes quant au contrôle des conditions d'environnement (lumière, bruit, température, etc.). Le partage d'un laboratoire par plusieurs chercheurs, la proximité peut-être bruyante de salles de cours ou d'un secrétariat, l'ingérence souvent inopinée du personnel d'entretien, l'arrivée de visiteurs, parfois de visiteurs donateurs, voilà de petits cailloux qui ne troublent que peu un ruisseau d'eau claire.

La recherche descriptive doit, en principe, respecter les conditions telles qu'on les trouve dans le milieu observé. Il reste que, pour certains protocoles de recherche – par exemple les protocoles qui touchent deux groupes ou quelques groupes de sujets observés, ou ceux qui enregistrent le même groupe de sujets à différentes époques de son développement –, une certaine uniformité, donc un certain contrôle s'imposent sur les conditions d'observation.

Contrôle de la ou des variables indépendantes et des variables contrôlées ; postvérification

Une variable peut être contrôlée de trois manières : 1) en lui faisant prendre quelques valeurs prédéterminées, appelées aussi « variantes » ; 2) en fixant la variable à une valeur donnée, constante ; 3) en faisant prendre à la variable un grand nombre de valeurs différentes, plus ou moins au hasard. On parlera de contrôle au sens strict pour les procédures 1 et 2. Quant à la procédure 3, qu'on peut qualifier d'optimiste puisqu'elle est basée sur l'effet compensatoire du hasard, elle vise à annuler les influences d'un ou de plusieurs facteurs potentiels en laissant jouer le hasard, grâce auquel les biais apportés devraient s'équilibrer. Cette procédure est appliquée généralement pour la constitution des groupes de sujets, pour le contrebalancement des suites de conditions et dans d'autres cas.

La VI est contrôlée par le chercheur selon la procédure 1. Les VCs peuvent être gérées par les procédures 1 ou 2, indifféremment ou selon les intentions du chercheur. Noter qu'une VC gérée par la procédure 1 ne constitue pas pour autant une VI même si, pragmatiquement, elle lui ressemble.

Il est fortement recommandé de valider les contrôles effectués, tant dans le cas de la VI que des VCs : pour ce faire, le chercheur ou, préférablement, une tierce personne, vient mesurer ou constater la valeur imposée réelle de la variable. Cette postvérification des contrôles peut rassurer le chercheur quant aux conditions de ses observations ; elle peut aussi, si certains contrôles tendent à déraiper ou à flancher, l'aider à imposer des contrôles plus solides ou à expliquer véridiquement les résultats obtenus.

La nécessité de vérifier la valeur effective des contrôles apparaît d'emblée dans les projets employant des sujets humains à qui, par exemple, le chercheur voudrait imposer une diète alimentaire, un régime d'exercices, un programme d'activités. Les variables physiques ne sont toutefois pas à l'abri des dérapages, tels la température ambiante qu'on voudrait fixer à tel niveau (pour un taux

d'humidité inchangé) ou l'éclairage, qui dépend non seulement de la puissance de la source lumineuse mais aussi de la réflectance présente dans le milieu ambiant. La rigueur, qui est de mise au départ de la démarche de contrôle d'une expérience, doit se doubler de vigilance afin que les contrôles déclarés dans l'énoncé de recherche soient effectifs.

Notons que l'examen critique, après coup, de l'expérience réalisée peut faire apparaître que certaines variables, certains biais non prévus sont intervenus et ont contribué à en rendre les résultats moins concluants. Ces variables intempestives faisaient sans doute partie du lot de variables que le chercheur avait globalement référées à la procédure de contrôle aléatoire, une procédure optimiste, redisons-le. Leurs effets avérés et, il va sans dire, imprévus leur confèrent le statut de « variables non contrôlées » authentiques.

Effets pervers : les techniques du simple et du double insu

Le chercheur veut attribuer les variations observées du phénomène à la chaîne causale qu'il a choisi d'étudier, *i.e.* à la VI. Or, plusieurs autres facteurs et conditions peuvent s'ajouter au facteur choisi et venir en confondre l'influence. Il incombe encore au chercheur de prévenir la difficulté par un contrôle, une manière de procéder appropriée.

- *L'effet placebo* consiste généralement en l'amélioration de l'état du patient, l'accroissement d'une performance, un changement favorable qui découle essentiellement de l'attitude du sujet parce qu'il croit en l'efficacité du traitement auquel on le soumet. L'utilisation d'un « groupe placebo », à qui on administre pour ainsi dire un traitement à vide (p. ex., une pilule inoffensive), constitue la parade d'usage pour décanter l'effet placebo. Parallèlement, la baisse de réponse du sujet qui appréhende les effets négatifs du traitement est désignée « effet nocebo ».

- *L'effet Rosenthal* (ou effet de l'expérimentateur) résume toute influence exercée par l'expérimentateur sur le sujet, tout impact direct de l'expérimentateur (en marge de la VI) sur le niveau de réponse, et ce, au gré ou à l'insu de l'expérimentateur. Cette influence, en partie inéluctable, se pervertit lorsqu'elle se double du biais de l'expérimentateur envers une catégorie de sujets, par exemple un biais favorable à l'égard du groupe expérimental : on parle aussi d'« effet Pygmalion » dans ce cas. La transposition des interventions expérimentales dans un cadre automatisé, ou robotisation de l'expérience, peut régler le problème, tout comme l'emploi d'agents d'expérimentation bien formés et le recours aux techniques d'insu (p. 40).
- *L'effet Hawthorne* dénote tout changement systématique du niveau de réponse des sujets attribuable au simple fait qu'ils participent à une recherche, qu'ils sont mesurés, observés, placés dans un contexte inhabituel. Alors que l'effet placebo découle d'une attitude (ou croyance, ou disposition émotive) des sujets eu égard à l'expérience qu'on leur propose et au comportement qu'on attend d'eux, l'effet Hawthorne n'en dépend pas et pourrait survenir même si les sujets n'étaient pas avertis de leur participation ou de leur rôle dans une expérience. On peut atténuer l'effet Hawthorne en dissimulant le plus possible le dispositif d'expérimentation aux yeux des sujets mais, lorsqu'il existe, il n'est guère possible de l'éradiquer. L'amélioration de réponse chez les sujets du groupe contrôle, parce qu'ils savent faire partie d'une expérience scientifique, est parfois désignée « effet John-Henry ».
- *L'effet de nouveauté* dénote le changement provisoire de réponse ou de performance des sujets lorsqu'ils sont soumis pour la première fois au contexte de l'expérience, au matériel ou à la tâche imposée. Cet effet, qui se manifeste le plus souvent par un niveau de réponse dégradé ou par un corrélât quelconque de stress, s'estompe d'ordinaire rapidement après un certain temps d'habituation. Quelques essais de pratique ou une simple période d'accommodation laissée au sujet suffisent souvent à obvier à cet effet.

- *L'effet chirurgical* résume toute variation dans le niveau de réponse, voire dans l'état général du sujet, qui dépend expressément des manipulations subies aux fins du contrôle expérimental. L'injection péritonéale ou intramusculaire d'une substance chez un rat cause habituellement un traumatisme, dont on doit rendre compte et qu'il faut contrôler : on utilisera par exemple un groupe « chirurgical » (ou groupe « injection ») dont les sujets se verront administrer une substance neutre. L'injection d'une drogue expérimentale chez un humain peut aussi provoquer un effet chirurgical, en plus de l'effet placebo éventuel.
- *L'effet de sélection* concerne la composition du groupe ou des groupes de participants à une étude, laquelle peut impliquer un biais de non-représentativité. Le cas le plus patent est celui du recrutement de « volontaires pour une expérience scientifique », un mode de sélection grâce auquel le chercheur recevra dans ses groupes des personnes ayant des loisirs, curieuses d'apprendre, prêtes aussi à contribuer à l'avancement des connaissances, et qui vont mettre une bonne volonté particulière à bien répondre aux demandes. Un biais positif semblable découle de la sélection d'un groupe d'élèves par l'enseignant en vue d'une étude en pédagogie, ou de patients d'un département hospitalier par le médecin responsable, etc.
- *Le biais de l'observateur* survient au moins de deux façons et peut contribuer à biaiser, *i.e.* fausser les résultats d'une expérience. Dans la première façon, l'observateur (qui est aussi parfois l'expérimentateur) doit par lui-même décider du résultat de l'expérience, en donnant une appréciation ou une catégorisation de ce que le sujet a produit comme réponse ; dans la seconde façon, même si la valeur du résultat est produite par un instrument de mesure, l'expérimentateur doit parfois intervenir dans le cours de l'expérience et décider, par exemple, quand l'essai est complet ou la tâche terminée. Dans l'un ou

l'autre cas, la préconception que l'observateur a de la situation globale et son anticipation du résultat de l'expérience peuvent et vont influencer sur son jugement.

- *Techniques du simple insu et du double insu.* L'expérimentateur qui, en exerçant un jugement personnel ou par observation, note le comportement de ses cobayes ou de ses sujets d'expérience tendra à favoriser d'une manière ou d'une autre la confirmation de son hypothèse de recherche. Ce biais peut aussi se manifester dans le dépouillement d'un questionnaire, dans l'application d'un procédé de mesure complexe, etc., toutes situations dans lesquelles interviennent le jugement et la décision de l'expérimentateur. Une tendance biaisante semblable peut survenir aussi dans les traitements mêmes appliqués aux participants, l'expérimentateur encourageant davantage ou encadrant d'un soin accru les sujets soumis à la condition expérimentale.

La technique du simple insu consiste à taire à la personne chargée d'effectuer la mesure du sujet la nature du groupe ou de la condition expérimentale dont ce sujet relève : l'expérimentateur lui-même se trouve ordinairement exclu de cette opération. La technique du double insu ajoute à la précaution du simple insu l'autre précaution par laquelle l'expérimentateur (ou l'agent d'expérimentation) ignore lui-même à quel groupe appartient le sujet, de quelle condition expérimentale il relève. On parle aussi de simple insu lorsque le participant est tenu dans l'ignorance de la nature du traitement qu'il reçoit, et de double insu si cette ignorance est partagée avec l'expérimentateur.

Ces techniques, le simple insu et le double insu, supposent une organisation globale de l'expérience dans laquelle le chercheur accepte de tenir le rôle d'un gestionnaire, s'efforce de recruter des auxiliaires compétents et se garde d'intervenir lui-même dans la routine expérimentale.

Protocole et plan d'expérience

Les contrôles actifs que le chercheur exerce dans sa situation de recherche concernent la VI et aussi, facultativement, une ou plusieurs VCs. La manière particulière dont ces contrôles sont concrétisés tout au long de la recherche constitue le *protocole de recherche*. Le protocole de recherche est donc l'organisation globale du contrôle que le chercheur met en œuvre, en faisant varier et en agencant les différents contrôles (VI et VCs) prévus. Le *plan d'expérience*, ou *plan de recherche*, est un schéma représentant ce protocole et cette organisation.

Protocole de recherche et plan d'expérience sont souvent employés de manière interchangeable. Si différence il y a, elle tient au fait que le protocole renferme l'expérimentation faite ou à faire de façon plus détaillée, notamment dans sa chronologie. Protocole est donc presque synonyme d'agenda d'expérimentation. Le plan de recherche, d'autre part, fige l'expérimentation dans un schéma, un tableau définitif, qui présente les parties majeures de l'organisation expérimentale et leur agencement.

La recherche expérimentale exploite généralement des plans (ou protocoles) comparatifs, grâce auxquels le chercheur peut obtenir, comparer et juger les effets des différentes manipulations qu'il opère sur la situation de recherche. Les protocoles utilisant un seul groupe en posttest, voire un seul groupe en prétest et en posttest, relèvent plutôt de démarches quasi expérimentales.

Plans à groupes indépendants (simple ou factoriels)

Le plan de base, pour ainsi dire, consiste en un dispositif expérimental tel que p (≥ 2) groupes de sujets sont constitués, chaque groupe recevant un traitement distinct. L'ensemble des traitements constitue la variable indépendante (ou variable de classification). Si on dénote par la lettre A cette variable, les différents groupes peuvent être identifiés et désignés par les symboles A_1, A_2, \dots, A_p , qui dénotent aussi les p variantes expérimentales appliquées.

Une première variable de classification (p. ex., la VI) peut être combinée à une seconde (p. ex., une VC ou une deuxième VI), chaque combinaison donnant lieu à une situation expérimentale distincte et étant réalisée avec un groupe de sujets spécifique : il s'agit alors d'un plan factoriel à deux variables (ou facteurs, ou dimensions). Identifiant une variable de classification par A (avec p niveaux ou variantes) et l'autre par B (avec q niveaux), le plan factoriel peut être symbolisé $A \times B$ et comporte $p \times q$ groupes de sujets. Le schéma ci-contre illustre ce plan.

Schéma du plan factoriel à deux dimensions ($A \times B$)

	B ₁	B ₂	...	B _q
A ₁	A ₁ B ₁	A ₁ B ₂		A ₁ B _q
A ₂	A ₂ B ₁			
...			A _i B _j	
A _p				A _p B _q

Par exemple, chaque sujet affecté au groupe A₂B₁ réalise l'expérience en étant soumis au niveau 2 de la variable A et au niveau 1 de la variable B.

Nous pouvons construire ainsi des plans à trois dimensions ($A \times B \times C$), avec $p \times q \times r$ groupes ; à quatre dimensions ($A \times B \times C \times D$), avec $p \times q \times r \times s$ groupes, etc.

Les plans d'expérience factoriels combinent les niveaux de deux ou plusieurs variables pour en étudier les effets sur la variable dépendante. De tels plans permettent non seulement au chercheur de déterminer les *effets principaux* (ou « marginaux ») de chaque variable, ou facteur, sur la VD, mais ils lui donnent aussi la possibilité de déceler des *effets d'interaction* entre les facteurs. Dans un plan $A \times B$, par exemple, le profil des différences de moyennes entre les niveaux de A (*i.e.* entre A₁, A₂ et A₃) peut rester semblable d'un niveau de B (*i.e.* B₁) à l'autre (B₂), ou bien les effets de A peuvent différer en fonction de B : c'est cette

« différence de différences » qu'on désigne par le concept d'interaction, et qui représente souvent la clé même d'une recherche expérimentale et son résultat le plus intéressant.

La méthode de prédilection appliquée pour constituer les différents groupes de sujets, dans la ligne de la recherche expérimentale proprement dite, consiste en deux phases. Le chercheur recrute d'abord un grand échantillon de sujets homogènes, qui répondent aux critères d'admissibilité convenus, puis il procède à leur affectation au hasard parmi les conditions et combinaisons de conditions. Nous revenons sur la question de l'échantillonnage plus loin (voir Sélection et nombre de sujets, p. 54).

Plans avec mesures répétées (simples ou factoriels)

Pour certaines expérimentations, le chercheur peut trouver avantageux, voire nécessaire, d'appliquer à chaque sujet d'un groupe deux ou plusieurs mesures successives, de mesurer le même sujet de manière répétée : ces mesures peuvent correspondre chacune à une condition expérimentale différente, ou encore être répétées sous la même condition imposée afin d'étudier l'effet de répétition ou du temps. Le plan avec mesures répétées de base est réalisé avec un groupe de sujets soumis à p conditions successives. On peut désigner les différentes conditions par les variantes A_1, A_2, \dots, A_p , et le plan lui-même par A_R , le suffixe R indiquant que la variable (ou VI) A est évaluée par la mesure répétée des mêmes sujets.

Toujours en utilisant un seul groupe de sujets, les conditions de contrôle peuvent être au nombre de deux, pour un plan factoriel $A_R \times B_R$, de trois, pour un plan $A_R \times B_R \times C_R$, etc.

Plans factoriels mixtes

L'expérimentateur a tout loisir de bâtir un plan de recherche et une structure de contrôle qui, à ses yeux, augurent bien de son expérimentation. Aussi l'organigramme des contrôles peut-il être compliqué à volonté, en mêlant les types de variables et parfois en entravant

certaines de leurs combinaisons. Les plans mixtes associent justement d'une à plusieurs variables de groupes et d'une à plusieurs variables répétées. Le plan mixte de base, dénoté $A \times B_R$ dans notre algèbre des plans, est d'ailleurs, et de loin, le plus répandu des plans de recherche : en voici le schéma.

Schéma du plan factoriel mixte à deux dimensions ($A \times B_R$)

	(1 ^{re} mesure) B_1	(2 ^e mesure) B_2	...	(q ^e mesure) B_q
(Groupe 1) A_1	A_1B_1	A_1B_2		A_1B_q
(Groupe 2) A_2	A_2B_1			
...			A_iB_j	
(Groupe p) A_p				A_pB_q

Le protocole paradigmatique qui consiste à comparer un groupe Expérimental à un groupe Témoin, et ce, en prétest (avant l'application du traitement au groupe Expérimental) et en posttest, correspond précisément au plan $A \times B_R$, tout comme d'autres protocoles connexes.

Les plans factoriels mixtes supérieurs, comprenant trois variables ou plus, se notent par exemple $A \times B \times C_R$, $A \times B_R \times C_R$, $A \times B \times C \times D_R$, etc. Dans cette notation, il est d'usage de réserver à la dernière ou aux dernières lettres le suffixe R pour mesures répétées.

Plans complexes (non factoriels)

En marge des plans factoriels, qui représentent des protocoles dans lesquels les p , q , ... niveaux des variables de classification ont été croisés, il existe d'autres plans, représentant d'autres stratégies

d'expérimentation et d'autres modes de contrôle. La documentation de ces plans se trouve principalement dans les traités d'analyse statistique (l'« analyse de variance ») plus avancés : nous y référons le lecteur (Kirk, 1994 ; Winer, Brown et Michels, 1991).

Mentionnons tout de même trois espèces, plus courantes, parmi ces plans complexes. Les plans dits « à facteurs emboîtés », dans lesquels les niveaux d'un facteur sont emboîtés dans l'autre (plutôt que croisés) : par exemple, les variantes A_1, A_2 et A_3 sont combinées à B_1 alors que A_4, A_5 et A_6 le sont à B_2 , etc. Les plans factoriels incomplets sont une autre espèce et ils correspondent à un plan factoriel dans lequel on supprime stratégiquement certaines cellules (afin d'en alléger la réalisation) ; on connaît aussi ces plans sous la dénomination de plans à « interactions confondues ». Enfin, les plans à contrebalancement, examinés plus loin, constituent une dernière espèce, qui concerne cette fois les protocoles à mesures répétées, et dans lesquels l'ordre même des séquences de mesure des sujets représente un contrôle ainsi qu'un facteur possible de classification.

Plans quasi expérimentaux

Le schéma à séquence temporelle :

$$M_{-t} \dots M_{-2} M_{-1} \Downarrow M_1 M_2 \dots M_t$$

dans lequel les lettres indicées M_i dénotent différentes mesures périodiques et la barre fléchée indique une intervention, le moment d'application d'un traitement, constitue un plan quasi expérimental, qu'on retrouve en recherche clinique par exemple. Un tel plan peut être appliqué à un groupe de personnes présentant le même syndrome, voire, à la limite, à un sujet unique.

Il existe toute une flore de protocoles qu'on peut désigner par le vocable de quasi ou préexpérimentaux : certains peuvent exploiter un groupe témoin, constitué cependant de sujets non équivalents aux

sujets expérimentaux, d'autres n'ayant pas de prétest ou soumettre cycliquement les sujets à l'application puis au retrait du traitement. Que dire encore d'un traitement (clinique) appliqué à des sujets qui présentent le syndrome étudié mais à des degrés divers ?

Les trois références fournies à l'orée de ce texte donnent plus de détail ; l'ouvrage déjà cité de Campbell et Stanley (1966) constitue la bible sur cette question.

Mérites et inconvénients des plans (ou protocoles) avec mesures répétées

Le chercheur obtient les mesures de plusieurs sujets dans différentes conditions, puis il compare l'effet des conditions en comparant les *moyennes* observées : c'est la différence entre ces moyennes qui l'informe de l'influence possible des conditions comparées.

Comparons la moyenne des données de n_1 sujets dans une condition A à celle de n_2 autres sujets dans une condition B. Quel est l'impact sur ces moyennes des différences observées entre les sujets *dans le même groupe* (la même condition) ? Quel est l'impact des différences observées entre les sujets *d'une condition à l'autre* ? Quel est l'impact de la non-équivalence possible des groupes comparés ?

Comparons à présent la moyenne des données de n sujets mesurés dans la condition A à celle des mêmes n sujets mesurés aussi dans la condition B. Quel est l'impact sur ces moyennes des différences réelles d'un sujet à l'autre ? Quel est l'impact de la variation des mesures du même sujet d'une condition à l'autre ? Comment se manifeste ici l'« erreur de mesure » ?

Les protocoles à mesures répétées, quand la situation de recherche l'autorise, sont généralement avantageux puisque, en soustrayant les particularités individuelles des sujets dans la comparaison de leurs moyennes de groupe, la part de hasard et de variabilité échantillonnale

apportée par le recrutement des sujets se trouve annulée⁶. Restent les effets de séquence, potentiellement nocifs, qu'il importe de neutraliser.

La prise de deux ou de plusieurs mesures sur la même personne (ou le même animal), d'une condition à l'autre, présente des avantages statistiques importants mais elle a l'inconvénient d'encourir des *effets de séquence* qui modifient le niveau de réponse des sujets : apprentissage ou habituation à la tâche, fatigue, anticipation, usure, contagion de mesure, contagion de traitement, etc. Ces effets, quand ils existent, ne peuvent pas être annulés : l'insertion d'intervalles temporels entre les moments de mesure peut les atténuer. Il est toutefois pensable de neutraliser les effets de séquence en les équilibrant à travers les sujets évalués⁷. Le schéma indiquant l'ordonnement des conditions de mesure qui vise à un tel but s'appelle plan contrebalancé ou équilibré (ou plan à contrebalancement).

Le principe du contrebalancement est justifiable par une algèbre grossière, qui attache un effet additionné de $+s$ à chaque occasion de répétition. Ainsi, la mesure d'un sujet i quelconque sous la condition C , qu'on pourrait dénoter X_{iC} en général, serait, si elle apparaissait comme 3^e mesure, *i.e.* après deux répétitions,

-
6. La variation attachée à chaque mesure est ordinairement rapportée à un seul concept global de variance d'erreur, symbolisé par σ_e^2 . Cependant, cette variation provient de plusieurs sources, les deux plus importantes étant la variabilité interindividuelle (reflétant la capacité, le niveau de réponse particulier de chaque sujet) et la fluctuation instrumentale, l'incertitude de la mesure elle-même. Les protocoles à mesures répétées épongent au moins partiellement la variabilité interindividuelle, du fait qu'il s'y trouve une corrélation positive, disons ρ , entre les mesures des sujets d'une condition à l'autre. Grâce à cette corrélation (laquelle serait nulle si l'on recourait à des groupes de sujets indépendants), la variance de la *différence entre les moyennes*, d'une condition à l'autre, se trouve réduite, comme l'exprime sa formule de calcul : $2\sigma_e^2/n \times (1 - \rho)$: une différence de moyennes donnée a donc plus de chances d'être trouvée significative.
 7. Notons toutefois que certains biais imputables à la répétition des mesures ne sont pas corrigés par le contrebalancement, notamment les effets d'anticipation (voir la méthode des permutations minimales, p. 52), les effets de gamme et les effets d'interaction particuliers entre chaque sujet et la séquence de conditions à laquelle on le soumet.

représentée par $X_{i,C} + s_1 + s_2$ ou encore $X_{i,C} + 2s$ (s étant la valeur de l'effet de séquence, supposée constante). En modifiant la séquence des traitements appliqués d'un sujet à l'autre en vue d'un juste équilibrage des effets de séquence, la *moyenne* des mesures pour chaque condition se trouvera affectée d'un effet de séquence moyen, disons (p. ex., $\bar{X}_c + \bar{s}$), qui sera le même d'une condition à l'autre et permettra ainsi de les comparer équitablement.

Méthodes de contrebalancement

Le contrebalancement d'un protocole de conditions en mesures répétées, *i.e.* d'une expérience soumettant le même groupe de sujets aux différentes conditions d'expérimentation, consiste à former des sous-groupes de sujets et à soumettre ceux-ci à des séquences de conditions complémentaires, de telle sorte que, au total des sujets et dans la comparaison des moyennes de conditions, le moment d'apparition de chaque condition ait une influence équilibrée.

Supposons une expérience impliquant k conditions différentes, avec mesures répétées : chaque sujet du groupe (expérimental) est donc soumis, successivement, à chaque condition. Il s'agit alors de répartir les n sujets du groupe en m sous-groupes, de sorte que chaque sous-groupe comporte $n' = n/m$ ($n' \geq 1$) sujet(s) : à la limite, un seul sujet suffit par sous-groupe. L'ordre des conditions successives auxquelles les sujets sont soumis est alors modifié pour chaque sous-groupe, en veillant à ce que les effets de séquence potentiels soient équilibrés de façon optimale.

Pour ordonner k conditions distinctes, l'expérimentateur dispose de $k!$ (k factorielle) permutations ou séquences différentes, où $k! = k \times (k - 1) \times \dots \times 1$. On peut envisager trois stratégies distinctes de contrebalancement :

- 1) le *contrebalancement complet* utilise $k!$ séquences correspondant aux $k!$ permutations des k conditions. Un protocole comportant $k = 4$ conditions générerait $4! = 24$ séquences distinctes, et requerrait donc soit 24, soit 48, soit 72 sujets, etc., à raison de 1, 2 ou 3 sujets par séquence.

- 2) le *contrebalancement partiel*, ou aléatoire, consiste à sélectionner, parmi les $k!$ permutations possibles de k conditions, un sous-ensemble équilibré ou à peu près équilibré ; la vérification du caractère équilibré incombe alors à l'expérimentateur. Ce type de protocole est parfois utilisé lorsque le nombre total (N) de sujets disponibles n'est pas un multiple du nombre (k) de conditions à évaluer.
- 3) le *contrebalancement en carré latin* – la technique la plus utilisée – exploite k séquences de k conditions chacune, formant un arrangement carré tel que chaque condition n'apparaisse qu'une seule fois dans chaque position ordinale. Il existe plusieurs carrés latins possibles pour chaque valeur de k^8 . Le protocole maître présenté ci-dessous illustre un carré latin d'ordre 4 :

Schéma d'un plan en carré latin pour les conditions A_1 , A_2 , A_3 et A_4

	Moment 1	Moment 2	Moment 3	Moment 4
Sous-groupe 1	A_1	A_2	A_3	A_4
Sous-groupe 2	A_2	A_4	A_1	A_3
Sous-groupe 3	A_3	A_1	A_4	A_2
Sous-groupe 4	A_4	A_3	A_2	A_1

Le lecteur peut vérifier, en exploitant le procédé algébrique suggéré plus haut, que le plan ci-dessus est vraiment équilibré.

8. L'expérimentateur peut être rassuré quant aux différents carrés latins disponibles, leur nombre étant fourni par la formule $C_k k!(k-1)!$, où $C_3 = 1$, $C_4 = 4$, $C_5 = 56$, $C_6 = 9408$. Par exemple, avec $k = 4$ conditions ou symboles, on peut former $C_4 \cdot 4!3! = 576$ carrés latins différents.

Pour une plus grande simplicité d'écriture, les séquences de conditions sont ordinairement réduites à la série équivalente de numéro de conditions. Voici par exemple deux nouveaux carrés latins d'ordre 4, présentés dans cette forme réduite :

Carré latin d'ordre 4

1	2	3	4
2	3	4	1
3	4	1	2
4	1	2	3

Carré latin d'ordre 4

3	2	1	4
4	1	2	3
1	3	4	2
2	4	3	1

L'expérimentateur pourra souhaiter, dans un plan en carré latin, éviter que, par exemple, la même paire de numéros $i : j$ ne se répète, comme c'est le cas systématique dans le plan ci-dessus, à gauche ; le plan présenté à droite n'a pas cet inconvénient.

Pour l'expérimentateur en herbe, il est de bonne guerre de se pratiquer à construire des plans, particulièrement des plans contrebalancés, et de réfléchir à « ce qui se passe » chez les sujets en cours d'exécution de ces plans⁹.

9. L'arrangement en carré latin a, à l'origine, été conçu pour la recherche en agronomie et il correspondait alors à un quadrillage de la surface arable pour l'expérimentation. Le chercheur, dans ce contexte, disposait de k « rangées » et k « colonnes » pour ses k types de semences, par exemple : l'ordre des rangées, permutable de $k!$ façons, était lui-même un enjeu et une source de variation potentiellement importante. Tel n'est pas le cas dans nos protocoles à mesures répétées, pour lesquels les rangées représentent des sous-groupes de sujets (assignés chacun à une séquence de k conditions), et dont l'ordre ne correspond ordinairement à rien d'utile dans l'intelligence de l'expérience en cours ou de ses résultats. Les traités de statistique qui abordent l'analyse de variance des plans en carré latin ou gréco-latin font souvent fi de ces nuances, et l'usager ne s'y reconnaît pas.

Il convient aussi de mentionner deux autres méthodes de contrebalancement moins courantes, qui trouvent leur application dans des situations particulières : le contrebalancement en carré gréco-latin et le contrebalancement par permutations minimales.

- 4) Le *carré gréco-latin* est un carré latin double formé par l'union de deux carrés latins « orthogonaux », soit des carrés dont les éléments sont couplés une seule fois l'un à l'autre. L'épithète de « gréco-latin » vient de ce que, naguère, les éléments d'un carré étaient notés par des lettres grecques, ceux de l'autre par des lettres romaines : on emploie dorénavant des lettres (romaines) et des chiffres (arabes). Ce type d'arrangement convient aux expérimentations mettant en jeu deux variables de contrôle, disons V_1 et V_2 , qui comportent chacune k niveaux, avec mesures répétées. Au lieu de former les k^2 combinaisons possibles des deux variables et de les contrebalancer, l'idée consiste ici à contrebalancer chaque variable, puis à réunir les deux schémas de contrebalancement pour en former un seul (dans lequel certaines combinaisons sont évidemment sacrifiées). Ce procédé très économique assure que les effets de séquence de chaque variable sont équilibrés de même que les effets principaux conjointement évalués, et ce, de manière compacte.

Les deux illustrations suivantes indiquent comment former un carré gréco-latin, d'ordre 3 et d'ordre 4 respectivement¹⁰. Comme pour les plans en carré latin, il s'agit pour l'expérimentateur d'assigner un sous-groupe de sujets à chaque ligne du carré, *i.e.* à chaque séquence de combinaisons de conditions telles que symbolisées dans les cellules du carré.

Remarquons que, dans chaque carré gréco-latin (à droite), toute combinaison lettre-chiffre n'apparaît qu'une seule fois.

10. Il est possible de former un carré gréco-latin pour tout k , sauf $k = 2$ et $k = 6$. Le lecteur peut trouver des listes de tels carrés dans les publications spécialisées (Fisher et Yates, 1963).

Exemple de formation d'un carré gréco-latin d'ordre 3

1	2	3	+	A	B	C	=	A1	B2	C3
3	1	2		B	C	A		B3	C1	A2
2	3	1		C	A	B		C2	A3	B1

Exemple de formation d'un carré gréco-latin d'ordre 4

1	2	3	4	+	A	B	C	D	=	A1	B2	C3	D4
3	4	1	2		B	A	D	C		B3	A4	D1	C2
4	3	2	1		C	D	A	B		C4	D3	A2	B1
2	1	4	3		D	C	B	A		D2	C1	B4	A3

Les *permutations minimales* sont, par définition, des permutations des nombres 1 à k telles que la corrélation que chacune entretient avec l'ordre naturel est nulle ou minimale. Dans un protocole comportant $k = 4$ conditions, dénotées 1, 2, 3 et 4, supposons (à toutes fins utiles) que l'ordre naturel soit (1, 2, 3, 4) : il s'agira de trouver une ou plusieurs séquences (j_1, j_2, j_3, j_4) telles que la corrélation entre ces séquences et la séquence (1, 2, 3, 4) soit la plus proche possible de zéro. Il est évident que, si une permutation (j_1, j_2, \dots, j_k) est minimale, son inverse $((j_k, j_{k-1}, \dots, j_1))$ l'est aussi. Par exemple, pour $k = 4$, les permutations (2, 4, 1, 3) et (3, 1, 4, 2) sont minimales ; pour $k = 5$, les permutations (1, 5, 4, 3, 2), (2, 5, 3, 1, 4), (4, 3, 2, 1, 5) ainsi que leurs inverses sont minimales, etc.¹¹

Il arrive que certaines variables indépendantes, ou variables de contrôle, constituent en elles-mêmes un ordre intrinsèque, une séquence naturelle : prenons l'exemple de 5 niveaux de

11. L'article de Clément et Laurencelle (1979) documente la question et fournit des listes étendues de permutations minimales pour $4 \leq k \leq 15$.

dosage d'un stimulant endocrinien, 4 niveaux de charge dans un travail de manutention, 6 intensités lumineuses croissantes dans une tâche perceptivo-motrice. Dans de tels cas, l'administration des k conditions dans leur ordre naturel peut être contre-indiquée, puisqu'elle pourrait induire chez le sujet une accommodation, une anticipation, une adaptation quelconque, reflétant le fait que la séquence des conditions imposées obéit à une progression. Les méthodes de contrebalancement normal, pour lesquelles l'ordre réel des conditions reste sans importance, ne sont pas appropriées ici, et l'expérimentateur doit veiller à contrecarrer cette tendance à l'accommodation, en rendant l'ordre des conditions successives imprévisible. C'est le rôle du présent type de contrebalancement.

Le schéma ci-dessous illustre un contrebalancement à permutations minimales pour $k = 5$ conditions (ordonnées). Le plan, qui exploite deux permutations distinctes en plus de leurs inverses, suppose donc l'utilisation de quatre sous-groupes. Le lecteur pourra vérifier aussi que plan est parfaitement équilibré, et ce, malgré qu'il ne soit ni carré ni latin ! L'équilibrage d'un tel plan, un mérite secondaire ici, incombe à l'expérimentateur.

Exemple de plan contrebalancé par permutations minimales ($k = 5$)

(Les chiffres dans le tableau dénotent les numéros d'ordre des 5 conditions imposées)

	Moment 1	Moment 2	Moment 3	Moment 4	Moment 5
Sous-groupe 1	1	5	4	3	2
Sous-groupe 2	2	3	4	5	1
Sous-groupe 3	4	3	2	1	5
Sous-groupe 4	5	1	2	3	4

Sélection et nombre de sujets

Type d'étude et de généralisation envisagée versus type d'échantillonnage. Ce ne sont pas toutes les recherches qui posent le problème de l'échantillonnage des participants. Les enquêtes globales par exemple, les recensements, rejoignent tous les membres admissibles de la population, toutes les unités échantillonnales : opération qui présente une logistique difficile et onéreuse, mais nul problème de choix ou de nombre. De même, afin de réaliser une étude exploratoire, de mode quasi (ou pré) expérimental, le chercheur pourra recruter quelques participants admissibles, sans plus.

La recherche expérimentale, qui est typiquement réalisée en laboratoire, utilise typiquement aussi des protocoles comparatifs. Par exemple, un groupe de personnes soumis à l'influence d'un régime expérimental est comparé à un groupe équivalent d'autres personnes, choisies selon les mêmes critères, qui ne reçoivent aucun traitement spécial, soit le groupe témoin. L'intention du chercheur est, en ce cas, de déterminer si les niveaux de réponse ou les mesures des sujets sont modifiés par l'imposition du régime expérimental, et si cette modification est généralisable d'une fois à l'autre, dans les mêmes conditions : on vise ici la *généralisation fonctionnelle* des résultats. Par contraste, dans une recherche descriptive, le chercheur a (souvent) pour but de déterminer la valeur d'un caractère dans une population, de faire le point sur certaines variables dans la population : pour ce faire, à défaut de mesurer la population entière, il va y prélever un échantillon, un groupe. La perspective de recherche, ici, n'est pas comparative ; pour que les données obtenues dans le groupe aient quelque valeur et nous renseignent de façon juste sur la population, il importe que le groupe choisi soit représentatif : on peut parler ici de *généralisation démographique*. Pour le cas expérimental, dans lequel l'information est obtenue essentiellement par une comparaison des groupes en présence, la représentativité n'est pas un enjeu réel ; les groupes constitués doivent avant tout être mutuellement comparables, ou échantillonnalement équivalents.

Échantillonnage au hasard vs répartition (affectation) au hasard. En recherche expérimentale, le mot clé pour le recrutement des sujets et la formation des groupes, est l'*équivalence échantillonnale*: en dehors des conditions expérimentales imposées aux différents groupes, chaque sujet devrait pouvoir se retrouver indifféremment dans l'un ou l'autre groupe, et les groupes constitués devraient présenter, en principe ou en fait, les mêmes caractéristiques échantillonnales et les mêmes dispositions à répondre¹². Dans certains cas, le chercheur peut s'assurer de cette équivalence en imposant à tous un prétest, par les résultats duquel il peut vérifier l'équivalence ou encore constituer des groupes équilibrés, soit par jumelage individuel, soit par égalisation des niveaux de réponse moyens.

La méthode privilégiée de formation des groupes, en recherche expérimentale, se déroule en deux phases : en phase 1, sont recrutés des participants qui sont admissibles et présentent des caractéristiques le plus homogènes possible, puis, en phase 2, les participants sont répartis au hasard à travers les groupes. Notons que la *répartition au hasard* est une affaire sérieuse, qui s'appuie préférentiellement sur un mécanisme de hasard (jeton ou dé) ou sur un procédé couplé à une table de nombres au hasard¹³.

Autre mode de formation des groupes, l'échantillonnage au hasard proprement dit signifie que les sujets sont tirés au hasard à partir d'un bassin, réel ou virtuel, une liste, contenant *tous* les éléments admissibles de la population. Ce mode d'échantillonnage, généralement très onéreux, vise d'abord à assurer la représentativité des sujets et des groupes (par rapport à la population mère); il concerne donc les recherches à problématique descriptive, les enquêtes, les sondages, plutôt que la recherche expérimentale comme telle.

12. On ne peut pas appliquer *stricto sensu* une telle exigence aux groupes pseudo-expérimentaux, formés de participants qui possèdent déjà la variante expérimentale sous enquête, p. ex., dans la comparaison hommes vs femmes, symptomatiques vs asymptomatiques, etc.. Pour ces cas, le chercheur doit néanmoins veiller à équilibrer les autres caractéristiques pertinentes des sujets recrutés.

13. Comme on en trouve dans Laurencelle et Dupuis (2000).

Types de recrutement et précautions à prendre. Les traités sur l'échantillonnage établissent plusieurs catégories différentes, à savoir l'échantillonnage probabiliste ou non probabiliste, puis l'échantillonnage au hasard simple, stratifié, stratifié proportionnel, par quotas, etc. Presque toutes ces catégories concernent la recherche descriptive et influencent le degré de généralisation (par argument de représentativité) qu'on peut en tirer. L'annexe C pose le vocabulaire et les grands concepts reliés à l'échantillonnage dans les études descriptives.

Il reste que, pour le recrutement des sujets en recherche expérimentale, l'impératif de l'équivalence échantillonnale impose de vraies contraintes. Par exemple, il serait dangereux de recruter les 10 premiers participants pour le traitement expérimental, et les 10 suivants dans le groupe de contrôle, à moins d'assurer une discrétion parfaite (verbale et générale) des participants entre eux. De même, il serait risqué que les premiers sujets soient tous assignés à une même condition, quelle qu'elle soit, alors que l'expérimentateur n'a pas encore bien assimilé la routine expérimentale, etc.

Nombre de sujets. En recherche expérimentale, l'intention du chercheur porte sur la généralisation des effets des conditions imposées ; peut-il affirmer que les différences qu'il observe entre ses groupes se reproduiront pour d'autres groupes comparables ? Dans cette perspective, il importe que les groupes expérimentaux soient constitués à partir d'un bassin de sujets homogènes, à caractéristiques échantillonnales semblables¹⁴ : ce précepte a pour but de réduire les sources de variation hétérogènes dans les données, d'assurer en fin de compte une plus grande sensibilité du devis de recherche aux effets expérimentaux. Dans ce contexte, pourquoi est-il important de mesurer un « bon nombre » de sujets, et combien en faut-il ? Nous proposons trois ordres de considérations à cet effet.

14. Dans les nombreuses recherches réalisées dans nos universités, ces sujets homogènes sont ordinairement nos bons vieux « étudiants de premier cycle ». Quant aux recherches à caractère biologique ou médical pratiquées sur les animaux, on utilise ordinairement des souris ou des rats achetés dans des fermes d'élevage, et dont les caractéristiques même génétiques sont soigneusement contrôlées.

La précision des moyennes obtenues. Les sujets, même ceux issus d'une population homogène, diffèrent quelque peu les uns des autres, et leurs mesures (X_1, X_2 , etc.) varient. Soit μ_X , le niveau typique ou moyen des mesures dans la population, et σ_X^2 leur variance. La mesure de chaque sujet, X_i , nous informe sur μ_X , avec un taux de variabilité, ou une marge d'imprécision, proportionnelle à σ_X . Au lieu d'un seul sujet, le chercheur va mesurer un *groupe*, *i.e.* un échantillon de sujets, obtenant les mesures $\{X_1, X_2, \dots, X_n\}$. La mesure globale qu'il va retenir, pour caractériser le groupe et vérifier l'effet des conditions expérimentales, est la moyenne (arithmétique), \bar{X} . Or, l'incertitude de cette moyenne peut être quantifiée par son erreur-type $\sigma_{\bar{X}}$ ou, au carré, par sa variance d'erreur $\sigma_{\bar{X}}^2$, et elle dépend essentiellement du nombre de données utilisées, n . Les lignes suivantes donnent, en esquisse, la démonstration de cette variance d'erreur : étant donné sa grande importance, nous invitons le lecteur à en étudier les implications.

$$\begin{aligned} \text{var}(\bar{X}) &= \text{var}\left(\sum_{i=1}^n X_i/n\right) \\ &= \text{var}(X_1 + X_2 + \dots + X_n)/n^2 \\ &= \left[\sum_{i=1}^n \sigma^2(X_i) + \sum_{i=1}^n \sum_{j \neq 1}^n \sigma(X_i)\sigma(X_j)\rho(X_i, X_j) \right] / n^2 \end{aligned}$$

Or, les éléments échantillonnés X_i proviennent de la même population, dans laquelle μ (ou μ_X) comme σ^2 (ou σ_X^2) sont constants ; les quantités $\rho(X_i, X_j)$, aussi notées $\rho_{i,j}$, dénotent les corrélations entre les variables de l'échantillon. Admettant provisoirement que $\rho_{i,j} = \bar{\rho}$ (une corrélation « moyenne »), l'équation précédente devient :

$$\begin{aligned} \text{var}(\bar{X}) &= [n\sigma^2 + n(n-1)\sigma^2\bar{\rho}]/n^2 && \text{(éq. 1)} \\ &= \sigma^2[1 + (n-1)\bar{\rho}]/n \end{aligned}$$

Si les éléments sont pigés au hasard dans la population, la corrélation entre leurs valeurs (p. ex., $\rho_{i,j}$, entre X_i et X_j) tourne près de zéro et elle tend en général vers zéro, d'où $\bar{\rho} = 0$. Par conséquent,

Lorsque des éléments sont tirés au hasard d'une population ayant pour moyenne μ_X et pour variance σ_X^2 , la moyenne de leurs valeurs (X_1, X_2, \dots, X_n) tend vers μ_X et elle a pour variance d'erreur :

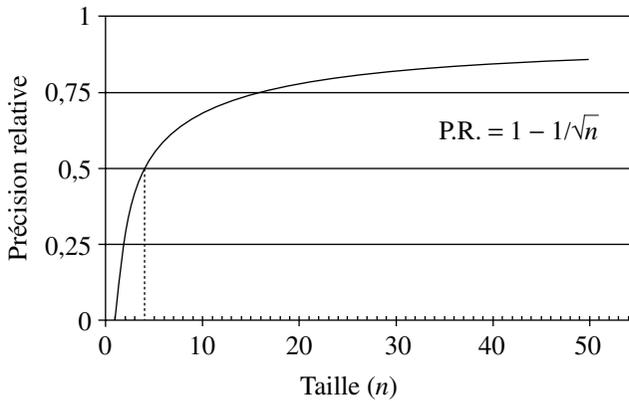
$$\text{var}(\bar{X}) = \sigma_X^2/n. \quad (\text{éq. 2})$$

L'*erreur-type* d'une moyenne, notée aussi $\sigma(\bar{X})$, est la racine carrée de sa variance d'erreur et égale donc σ_X/\sqrt{n} . Cette erreur-type indique la marge d'imprécision qu'a la moyenne observée \bar{X} pour estimer μ_X , selon $\bar{X} \pm \sigma(\bar{X})$. On voit que, par exemple, pour *doubler la précision de la moyenne*, i.e. réduire de moitié son erreur-type et sa marge d'imprécision, il faut *quadrupler le nombre d'éléments* échantillonnés au hasard qu'elle contient.

La précision de la moyenne varie selon la « loi de l'inverse de la racine carrée (du nombre de données) ». Un graphique de la fonction « $1 - 1/\sqrt{n}$ », selon n , illustre cette loi (voir page suivante). Par exemple, on réduit l'imprécision de la moyenne par un facteur de 50 % en utilisant $n = 4$ sujets, tel que l'illustre le graphique présenté à la page suivante ; pour gagner les prochains 50 % de réduction, il nous faudrait ajouter une infinité de nouveaux sujets. En sciences expérimentales, la coutume occulte qui fixe entre 8 et 15, et souvent à 10, le nombre de « cobayes » par condition d'expérimentation se trouve confortée par ces considérations. En effet, si l'expérimentateur est assez riche ou assez scrupuleux pour songer à recruter un plus grand nombre de sujets, mieux vaut pour lui consacrer plutôt son effort à recruter des groupes de sujets *plus homogènes*¹⁵.

15. Ce faisant, il contribue à réduire la quantité σ_X^2 , qui reflète surtout la variabilité inter-individuelle, et il réduit donc en même temps (par son numérateur) l'erreur-type de la moyenne.

En général, pour n croissant, la moyenne \bar{X} fluctue dans un intervalle de plus en plus étroit (selon $1/\sqrt{n}$) et s'approche de ce fait de sa valeur limite, μ_X . L'examen de l'équation 1 montre toutefois que la loi mentionnée n'est valide que si les corrélations $\rho(X_i, X_j)$ sont nulles, ce qui est garanti par l'échantillonnage au hasard (ou, d'une certaine façon, par la répartition au hasard), qui assure que chaque nouvelle observation est statistiquement indépendante (donc non corrélée) par rapport aux précédentes.



L'échantillonnage par grappes, très répandu en sciences humaines, contrevient à cet impératif et engendre au contraire des corrélations positives dans les données. À la limite, $\rho_{i,j} > 0$ et $\rho \rightarrow 1$; dans ce cas hypothétique, l'examen de l'équation 1 montre que $\text{var}(\bar{X}) \rightarrow \sigma_X^2$, soit que l'accumulation de données ne réduit pas (ou pas assez) l'imprécision de la moyenne¹⁶.

16. La perte de précision de la moyenne (ou de toute autre statistique comparable, comme la proportion, la variance, la corrélation) n'est pas la seule conséquence d'une telle situation. Toutes les procédures de test d'hypothèses, tel le test t de différence entre les moyennes de deux groupes indépendants, se trouvent faussées, et d'une puissance faussement gonflée, par la présence de corrélation dans les données. Cela est vrai aussi de l'analyse des tableaux de fréquences par le Khi-deux.

Le principe de l'échantillonnage par grappes consiste à recruter, non pas des éléments individuels de la population, mais des groupes plus ou moins importants, tels qu'on les rencontre dans leur milieu naturel : homme et femme d'un couple, élèves d'une même classe, patients d'un même département en centre hospitalier. Même si les grappes sont elles-mêmes tirées au hasard (dans une population de grappes), il reste qu'il y a vraisemblablement des similitudes d'un ordre ou d'un autre entre les éléments d'une même grappe, donc de la corrélation positive, ce qui entraîne des difficultés parfois insurmontables dans l'interprétation statistique. Par exemple, dirons-nous que notre moyenne est basée sur 50 individus, ou bien sur les membres de 25 couples ?

La puissance statistique, un concept que nous n'examinons pas ici, reflète jusqu'à quel point, dans des conditions expérimentales données, notre procédure statistique est sensible aux effets expérimentaux et apte à produire une déclaration de résultat significatif. La taille d'échantillon (n) influence positivement la puissance. Noter que l'impact de la taille est d'autant plus marqué que les effets à détecter sont subtils, que les différences apportées (Δ) sont légères (par rapport à la variabilité intrinsèque des données, *i.e.* $\Delta/\sigma_X \rightarrow 0$).

La puissance d'un test d'hypothèses dans une situation donnée, dénotée P , est la probabilité que l'hypothèse de non-différence (l'hypothèse nulle) soit rejetée, et le test déclaré significatif, alors qu'il existe une vraie différence paramétrique dans les résultats (p. ex., entre les moyennes de deux groupes). Les facteurs qui influencent positivement la puissance sont : la différence réelle entre les conditions comparées (Δ), la taille échantillonnale (n) et le seuil de signification (α), tandis que la variabilité intrinsèque des données, reflétée par l'écart-type σ_X , l'amoindrit. Sauf pour certains cas spéciaux, le calcul de la puissance s'avère une tâche sérieuse, qui comporte aussi une certaine part d'arbitraire. Pour se familiariser avec le concept et la technique de calcul, le lecteur peut se référer à W.C. Guenther (1973), J. Cohen (1988), M.M. Desu et D. Raghavarao (1990) ; Laurencelle (2005b) aborde l'*estimation empirique* de la puissance comme un outil stratégique permettant au chercheur de replanifier son expérimentation.

La représentativité échantillonnale. Dans une population dite « homogène », tous les éléments ont des caractéristiques semblables, à des différences aléatoires près. Dans la « grande population », par contre, telle que celle visée par les études descriptives, les enquêtes et sondages, les recensements, etc., on admet (ou on constate) qu'il existe des catégories et sous-catégories d'éléments, définies par des caractéristiques distinctives¹⁷ : cette population est dite « hétérogène ». C'est pour ce type de population qu'on doit se préoccuper de représentativité (voir l'Annexe C, p. 93).

La représentativité (des échantillons) est primordiale, car elle garantit que les mesures et informations tirées de l'échantillon renseignent adéquatement sur la population, *i.e.* que les moyennes et autres indices statistiques obtenus sont *non biaisés*¹⁸ (quelle que soit leur précision).

N'importe quel groupe d'éléments tirés d'une population inhomogène n'est pas forcément représentatif de celle-ci. Pour que l'échantillon soit représentatif, il faut que sa composition reflète effectivement ou virtuellement la composition de la population, qu'il en constitue une copie miniature. Par composition, on entend sa structure en catégories et sous-catégories, une structure pertinente en fonction de la problématique et des objectifs de la recherche entreprise. La planification d'une étude à caractère sociodémographique ou épidémiologique déborde le présent exposé, de même

17. Par exemple, le sexe, l'âge, le niveau de revenu familial, le degré de scolarité atteint, etc. La pertinence des subdivisions de la « population » en catégories dépend essentiellement de l'objet même de l'enquête et de l'influence appréhendée des catégories sur les résultats.

18. Nous parlons ici du biais *échantillonnal*, qui découlerait d'une composition de l'échantillon déséquilibrée par rapport à la population, composition dans laquelle certaines sous-populations seraient surreprésentées et d'autres sous-représentées. La question du biais statistique (ou mathématique) est tout autre. Ainsi, pour des échantillons qui seraient dûment tirés au hasard et représentatifs de la population, les indices (ou estimateurs) usuels de moyenne, proportion et variance sont non-biaisés, tandis que l'écart-type et la corrélation sont biaisés, en ce sens que $E(s_X) < \sigma_X$ et $E(r_{X,Y}) < \rho_{X,Y}$ (si $\rho_{X,Y} > 0$), la valeur de ces biais étant connue pour des populations « normales ».

que la discussion des concepts associés : unité échantillonnale, critères de représentativité, modes et sous-modes d'échantillonnage, biais et précision des moyennes obtenues selon le mode, etc.

Le tableau présenté à la page suivante illustre différentes configurations d'échantillons possibles, tirés d'une population inhomogène définie, en même temps que les risques découlant d'un échantillonnage trop peu nombreux.

Schéma d'une population inhomogène constituée de cinq sous-populations et de la composition probable d'échantillons de tailles croissantes tirés au hasard de cette population

	Population (N = 10 000)	Répartitions possibles selon différentes tailles n (Intervalle de confiance à 50 %)		
		$n = 10$	$n = 50$	$n = 100$
Strate A	5000	4 – 6 (5)	23 – 27 (25)	47 – 53 (50)
Strate B	2500	2 – 3 (2)	10 – 14 (12)	22 – 27 (25)
Strate C	1250	0 – 1 (1)	4 – 7 (6)	10 – 14 (12)
Strate D	1000	0 – 1 (1)	4 – 6 (5)	8 – 12 (10)
Strate E	250	0 – 0 (0)	0 – 1 (1)	1 – 3 (2)

L'intervalle $c_1 - c_2$ indique les effectifs possibles de chaque strate, tels qu'on les retrouverait à peu près dans 50 % des échantillons tirés. La valeur juxtaposée (c) indique l'effectif le plus probable.

La méthode employée pour fixer les intervalles de confiance (à 50 %) est une approximation, par la loi binomiale, de la méthode exacte qui exploiterait une loi hypergéométrique multiple. L'approximation suppose que, au lieu de N = 10 000 éléments répartis tel qu'indiqué, la population est inépuisable et composée des strates respectant les fractions spécifiées.

Il reste qu'un échantillon total de $n = 1$ personne ne peut pas être dit représentatif d'une population quelconque, homogène ou hétérogène. De plus, un échantillon de $n (> 1)$ personnes, qui comporterait $n_j = 0$ élément dans la catégorie retenue j , ne peut pas être dit représentatif de la population mère.

La probabilité qu'une strate, une portion spécifique, de la population soit absente de l'échantillon de n éléments (c'est-à-dire, que $n_{\text{strate}} = 0$) peut être facilement calculée. Supposons que la proportion d'éléments de cette strate dans la population est p . Alors, si on pige au hasard n éléments dans la population (infinie), l'échantillon pourra contenir zéro élément de la strate mentionnée selon une probabilité :

$$\Pr(n_{\text{strate}} = 0) = (1 - p)^n. \quad (\text{éq. 3})$$

Prenons l'exemple des échantillons mentionnés dans le tableau ci-dessus. Pour la strate E, à proportion $p = 0,025$, la probabilité de tirer au hasard un échantillon de $n = 10$ éléments sans aucun élément de la strate E est de $(1 - 0,025)^{10} \approx 0,78$; pour $n = 50$, de 0,28; pour $n = 100$, de 0,08.

On conçoit, et on peut montrer statistiquement, que la taille n influence la représentativité de l'échantillon et, par conséquent, le biais. Une règle pratique consiste à repérer la catégorie ou sous-catégorie la moins nombreuse de la population et qu'il nous importe d'inclure dans l'échantillon, puis d'assurer que notre recrutement échantillonnal en retiendra au moins n_j (≥ 1).¹⁹

Outre l'annexe C, le lecteur est invité à consulter les ouvrages de base sur l'échantillonnage, tant pour les méthodes de planification et de recrutement que pour les corrélats statistiques. Signalons notamment les classiques de L. Kish (1967) et W.G. Cochran (1977). Vallerand et Hess (2000) présentent aussi un chapitre intéressant et une bibliographie sur l'échantillonnage et ses modes. Quant à la représentativité, Laurencelle (2005c) discute plus sérieusement le concept et propose un test de non-représentativité pour des échantillons tirés de populations infinies ou finies.

19. Pour $n_j \geq 1$, nous suggérons le procédé suivant, basé sur une inversion de l'équation 3. Soit p_j , la portion d'éléments de cette catégorie (ou sous-catégorie) dans la population, et un coefficient d'assurance α ($0 < \alpha < 1$). Alors, si la pige est faite au hasard pur, on doit échantillonner $n \geq \log(1-\alpha) / \log(1-p_j)$, moyennant quoi on est assuré à $100\alpha\%$ que l'échantillon comporte au moins 1 élément de ladite catégorie. Par exemple, avec $p_j = 0,1$ et $\alpha = 0,99$, on calcule $\log(1-0,99) / \log(1-0,1) \approx 43,71$, soit une taille minimale de $n \geq 44$. D'autres méthodes sont aussi utilisées.

Mise au point du dispositif d'expérimentation

Les appareils, méthodes de mesure, méthodes de contrôle étant choisis et séparément mis au point, il faut ensuite trouver, réserver et préparer le lieu d'expérimentation, ce qui peut présenter un défi peu banal dans le cas d'une recherche sur le terrain (p. ex., dans une école, dans un lieu public, dans un milieu sauvage). Dans le cas d'une expérimentation avec des participants humains, outre la nécessité du consentement éclairé et les précautions déontologiques d'usage (voir l'Annexe A, p. 75), il importe d'apporter un soin particulier aux « instructions » adressées à chaque participant, par lesquelles le contexte global de la recherche lui est communiqué en même temps que l'attitude et les comportements qu'on attend de lui. Ces instructions sont préparées d'avance, rédigées et, si possible, récitées mot à mot à chaque sujet.

C'est ordinairement la première fois que les appareils, contrôles, mesures choisis se trouveront réunis dans le contexte défini d'une recherche, comme c'est aussi la première fois que le chercheur (notamment l'étudiant-chercheur) ou l'agent d'expérimentation opérera dans le contexte qu'on vient de mettre sur pied. Or, chaque participant doit être traité, chaque donnée doit être obtenue dans des conditions qui sont à la fois identiques et optimales. Cela signifie que, entre autres, l'encadrement du *premier participant* doit être aussi assuré et compétent que celui des derniers participants, et que les problèmes et ratés du dispositif d'expérimentation (contrôles et mesures) doivent avoir été repérés, maîtrisés et éliminés d'entrée de jeu. Il y a, dans cette préparation matérielle du dispositif d'expérimentation, un travail de patience, de rigueur et d'inventivité qui reste incontournable.

PRÉEXPÉRIMENTATION (FONCTIONS ET BUTS)

Comme l'indique son nom, la préexpérimentation précède l'expérimentation et constitue à la fois une phase de rodage et une précaution nécessaires à toute recherche. La préexpérimentation se fait « comme pour vrai », en employant d'un à quelques sujets pour chaque condition, ces sujets étant évidemment « brûlés » pour l'expérience elle-même.

La préexpérimentation vise deux ordres d'objectifs spécifiques. Les objectifs de premier ordre touchent le bon fonctionnement de l'expérience, son script et sa chorégraphie, pour emprunter le langage des arts de la scène. Non seulement tous les contrôles, toutes les mesures doivent-elles fonctionner parfaitement, mais aussi l'expérimentateur (et ses assistants, s'il y a lieu) doit y évoluer avec aplomb et compétence. Script et chorégraphie font référence aux échanges verbaux (instructions aux sujets, réponses aux questions, feedbacks si autorisés, etc.) et aux manipulations, déplacements, interventions à effectuer dans le contexte de l'expérience. Si besoin est, la préexpérimentation devrait être prorogée jusqu'à satisfaction.

Les objectifs de second ordre se rapportent au contenu, au programme même de la recherche. La réalisation « comme pour vrai » va faire apparaître pour la première fois le contexte réel et concret de l'expérience, par exemple la durée d'expérimentation pour un participant, le délai requis pour passer d'une condition ou d'une phase expérimentale à l'autre, la compréhension effective des instructions et consignes par les sujets, les aspects rebutants ou pénibles des traitements imposés ou des mesures. Finalement, malgré leur petit nombre et le fait qu'on doive (généralement) les écarter des résultats proprement dits, les données obtenues en préexpérimentation restent indicatives et peuvent servir à signaler des problèmes. Par exemple, la performance des sujets dans différentes conditions peut « plafonner », *i.e.* produire des scores toujours maximaux ou presque maximaux, en raison d'une trop grande

facilité de la tâche, ou bien les données peuvent contredire totalement l'hypothèse de recherche, auquel cas la problématique mérite sans doute d'être repensée, etc.

Une expérience bien planifiée mais menée sans préexpérimentation risque fort d'être considérée, après coup, comme la préexpérimentation qu'on aurait dû faire. Un tel ratage, pourtant facile à éviter, peut se révéler fatal lorsque la recherche doit être conduite dans un délai assez court ou qu'elle implique des patients « gaspillés » ou un budget irrécouvrable.

Par ailleurs, même quand il y a eu préexpérimentation, chaque expérience, même concluante, n'est jamais irréprochable (puisqu'on a dû y faire des choix) ni définitive (puisqu'elle comporte des aspects statistiques et fluctuants). Mieux vaut, en général, un train de petites expérimentations, conçues et exécutées si possible par des chercheurs et dans des laboratoires différents, plutôt que la grande expérimentation dans laquelle on a tenté de tout prévoir. La recette de R.A. Fisher (1890-1962) pour le progrès soutenu en recherche est : « significativité et réplication ».

EXPÉRIMENTATION PROPREMENT DITE

Éthique du chercheur et ses obligations déontologiques

Activité professionnelle responsable, la recherche doit être éthique et axée principalement sur le bien-être immédiat et futur des personnes et de la société. De la conception du projet jusqu'à la diffusion de ses résultats et au-delà, le chercheur doit se préoccuper explicitement de la dimension éthique de ses interventions, de leur impact, du risque qu'elles font courir aux sujets et de sa propre capacité à en assumer les conséquences. À cette obligation éthique du chercheur sont associés presque partout un code de déontologie, composé d'une série de règles coercitives, et un comité de surveillance, qui a le pouvoir d'approuver, sanctionner ou bannir un projet. La règle déontologique ayant la plus grande évidence est

celle qui impose au chercheur, désireux de recruter un participant pour son expérience, d'obtenir au préalable son consentement libre et éclairé. L'Annexe A développe davantage cette question.

Robotisation des procédures et observation

Quand elle est préparée avec minutie et adéquatement rodée en préexpérimentation, l'expérience se déroule de façon quasi automatique : aussi, il y a peu à en dire ici. Mentionnons seulement deux facettes complémentaires de sa réalisation. D'une part, les données accumulées seront d'autant plus crédibles, et l'expérience réalisée un succès, que sa réalisation se sera approchée d'une exécution robotique (sans devenir rébarbative pour les participants !), *i.e.* toujours la même, dans des conditions le plus possible identiques, avec une attitude et des interventions calquées sur un modèle inchangé. D'autre part, la recherche constitue en elle-même un lieu et une occasion privilégiés pour l'observation du phénomène, quel que soit le contexte – laboratoire ou milieu naturel – où elle se déroule. C'est pourquoi, même s'il reste impassible devant les événements et tient son rôle dans le déroulement de l'expérience, le chercheur doit garder les yeux et l'esprit ouverts et se tenir à l'affût de ce qui s'y passe : réactions verbales ou physiologiques du sujet, raté des appareillages, bruits ou incidents dans l'environnement, etc. Les observations faites, même si elles sont qualitatives et sans portée intrinsèque, pourront guider le chercheur dans son interprétation et sa discussion des résultats.

Mortalité expérimentale

L'expression « mortalité expérimentale » dénote, au sens large, la perte ou défection de sujets après qu'ils aient été recrutés et inscrits dans le ou les groupes de l'expérience. Cette perte de sujets peut être ramenée aux trois catégories suivantes :

- *mortalité aléatoire* : les sujets font défaut au hasard, pour des raisons ou des motifs n'ayant rien à voir avec les conditions d'expérimentation ou leur type échantillonnal (déménagement, accident fortuit, etc.) ;

- *mortalité idiosyncrasique* : les sujets font défaut vraisemblablement à cause d'une disposition particulière de leur type échantillonnal, d'une maladie, etc. ;
- *mortalité expérimentale* proprement dite : les sujets font défaut (vraisemblablement) à cause de l'impact de la condition expérimentale (*i.e.* du traitement) à laquelle on les a assignés.

Les cas de perte aléatoire de sujets ne sont évidemment pas inquiétants : s'il est possible de leur trouver des remplaçants possédant les mêmes caractéristiques échantillonales, tant mieux. Les pertes de sujets ressortissant aux deux dernières catégories sont plus sensibles. En effet, la défection (voire la mort) d'un sujet constitue un « résultat » qu'il est impossible d'ignorer. Le remplacement pur et simple des absents ne résout en rien le problème engendré et servirait seulement à l'occulter. La difficulté reste entière.

Dans le cas de certaines analyses statistiques, il existe des procédés mathématiques qui permettent de pallier à peu près correctement aux données manquantes (par interpolation, extrapolation ou winsorisation²⁰) ; l'emploi légitime de l'un ou l'autre de ces procédés doit être chaque fois justifié en regard de la raison alléguée de la perte de données, de l'information transmise par le cas concerné et de l'utilisation de la ou des données dans les traitements statistiques prévus²¹.

20. Sommairement, l'interpolation consiste à suppléer une valeur manquante dans un contexte qui fournit ses valeurs adjacentes inférieure et supérieure ; l'extrapolation, en projetant la valeur manquante par le recours à un modèle hypothétique (p. ex., « S'il avait complété les 30 minutes d'expérimentation, le sujet aurait produit des réponses au même taux que dans les 20 premières minutes ») ; la winsorisation, en supposant que la valeur manquante est au moins aussi élevée (ou aussi basse) que la plus élevée (ou basse) obtenue, et en inscrivant cette valeur.

21. L'insertion d'une donnée artificielle, en effet, repose sur le principe que cette donnée n'ajoute pas d'information nouvelle dans la synthèse statistique des vraies données et qu'elle n'ajoute pas un degré de liberté de plus dans les procédures d'inférence statistique dans lesquelles ce paramètre est utilisé.

TRAITEMENT STATISTIQUE ET GÉNÉRALISATION

Traitement statistique : synthèse de données et généralisation

Le traitement de données, consécutif à l'expérimentation, se fait en deux phases : production d'une synthèse des résultats obtenus, ou bilan, et vérification statistique de l'hypothèse de recherche. Les calculs de statistique descriptive suffisent au bilan : moyennes et écarts-types, corrélations, tableaux de fréquences, etc. La vérification de l'hypothèse de recherche est en fait une généralisation, ou un essai de généralisation, des résultats obtenus dans l'expérience : le but est de montrer que les conditions étudiées dans l'expérience produisent systématiquement des différences telles qu'observées, plutôt qu'elles ne les produisent par accident ou au hasard. Cette généralisation n'est pas un geste trivial, étant donné la variabilité intrinsèque des données et les différences interindividuelles qui caractérisent presque tous les domaines d'étude.

Louis Laurencelle (2005a) présente les concepts de statistique inférentielle les plus importants en recherche ; on y trouve aussi, à titre d'illustrations, quelques méthodes de test parmi les plus fréquemment employées.

Pour faire saisir l'enjeu et les principaux concepts liés à la généralisation statistique, nous présentons trois cas-types (volontairement exagérés), relatifs à une expérience imaginaire. Cette expérience comporterait deux groupes équivalents de 10 sujets chacun : un groupe soumis au traitement expérimental, le groupe expérimental, et un groupe témoin. Une fois l'expérience réalisée dans des conditions optimales, nous obtenons 20 mesures, soit 10 mesures dans chaque groupe. Voici les trois situations hypothétiques, sous forme de tableaux de résultats individuels.

En **situation 1**, la différence entre les groupes est nettement marquée ; il est clair que les sujets en condition expérimentale (E) performant mieux que les témoins (T). Aussi, on peut anticiper que

	Groupe expérimental	Groupe témoin	
Situation 1	13 13 13 13 13 13 13 13 13 13 $\bar{X} = 13,00$ $s_X = 0,00$	9 9 9 9 9 9 9 9 9 9 $\bar{X} = 9,00$ $s_X = 0,00$	$t = + \infty$
Situation 2	12 14 9 2 15 3 7 10 11 16 $\bar{X} = 9,90$ $s_X = 4,77$	12 3 5 2 9 16 14 6 8 7 $\bar{X} = 8,20$ $s_X = 4,61$	$t = + 0,810$
Situation 3	12 11 16 12 21 8 18 21 19 7 $\bar{X} = 14,50$ $s_X = 5,19$	10 9 9 6 1 12 16 1 15 11 $\bar{X} = 10,00$ $s_X = 4,29$	$t = + 2,113$

la $n + 1^e$ donnée du groupe E sera $x_{n+1} \approx 13$, et de même avec $x_{n+1} \approx 9$ pour le groupe T. D'où on peut sûrement conclure à une différence systématique et réelle (*i.e.* non aléatoire) entre les conditions.

En **situation 2**, les moyennes avantagent le groupe E. Cependant, il y a chevauchement réciproque des données d'un groupe à l'autre. De plus, l'incertitude des moyennes, indiquée par les écarts-types assez élevés, est forte en comparaison de leur différence mutuelle²². Par un test d'hypothèses (le test t de différence entre deux moyennes indépendantes), la différence se révèle « non significative » au seuil de 5 %²³. En fait, ces ensembles de valeurs ont été constitués par tirage aléatoire, dans un tableau de nombres de 1 à 16.

En **situation 3**, les moyennes avantagent le groupe E, comme en situation 2. Il y a chevauchement réciproque des groupes ; cependant, le rapport (Différence mutuelle entre les moyennes) : (Imprecision de chaque moyenne) est plus élevé. Au test t , la

22. Les valeurs estimatives des erreurs-types sont respectivement égales à $4,77/\sqrt{10} \approx 1,51$ et $4,61/\sqrt{10} \approx 1,46$, versus la différence observée de 1,70 entre les moyennes.

23. La « valeur critique » du t avec dl (degrés de liberté) = $n_1 + n_2 - 2 = 18$ est de 2,101 (pour un test bilatéral) au seuil $\alpha = 0,05$ et de 2,878 au seuil $\alpha = 0,01$ (voir Laurencelle et Dupuis, 2000).

différence apparaît tout juste significative au seuil de 5 % (en mode bilatéral). En fait, l'ensemble des valeurs T (du groupe témoin) a été constitué comme en situation 2 ; pour l'ensemble E, on a procédé de même, en additionnant 5 à chaque donnée. Il y a donc là une différence réelle qui est présente, qui aurait pu ne pas apparaître significative ou ne pas être détectée si on avait utilisé +3 au lieu de +5 ou si on avait employé des groupes de sujets moins nombreux (*i.e.* $n < 10$).

Le traitement statistique des données afin de vérifier des différences expérimentales est un art qui conjugue pratique et théorie et qu'il faut acquérir patiemment. En tout cas, méfions-nous des interprétations hâtives !

Les opérations concrètes associées au traitement statistique des données (relevé d'un appareil de mesure, transcription de données dans un tableau, inscription sur tableur informatique, fixation des paramètres de commande d'un progiciel de calcul et analyse statistique) sont, chacune à sa façon, susceptibles d'erreur. Il est fortement recommandé au chercheur de valider lui-même, de A à Z, le cheminement de ses données jusqu'à leur conclusion produite²⁴. En plus de rassurer le chercheur quant aux erreurs de manipulation et de transformation possibles, cette validation rafraîchira la compréhension qu'a le chercheur de tout ce processus et lui en redonnera la responsabilité, laquelle est trop souvent laissée aux « spécialistes », voire aux ordinateurs.

Les données manquantes ou aberrantes

Il arrive que les tableaux de données accumulées dans une recherche aient une ou plusieurs données ou manquantes, ou aberrantes, leur inclusion dans les traitements statistiques posant problème.

24. Un procédé de validation souvent utilisé consiste à sélectionner, plus ou moins au hasard, un sujet ou un groupe, et à en suivre la trace à chaque étape de manipulation des données, en vérifiant au besoin les calculs effectués.

Par « donnée manquante », on désigne le résultat d'une mesure qui n'a pas pu être obtenu ou enregistré, soit en raison d'une panne du dispositif de mesure, soit par la faute du participant (absence, mortalité expérimentale, comportement contraire). Par « donnée aberrante » (on dit aussi « donnée folle »), on désigne une valeur de résultat remarquable et peu plausible, qui peut provenir d'un ratage ou d'un dérèglement provisoire du dispositif de mesure ou d'affichage, d'une erreur de transcription, parfois d'un faux mouvement du sujet.

Deux mises en garde s'imposent ici : une donnée manquante ne doit pas (généralement) être remplacée par la valeur 0 (attention aux logiciels de traitement statistique), et on ne doit pas corriger d'autorité et arbitrairement une valeur aberrante (p. ex., en lui substituant la moyenne). La gestion des cas de données provenant de la défection des sujets dépend, on l'a vu, du type de perte impliqué. Quant à la gestion des données aberrantes, deux solutions existent : l'une consiste à déclarer cette donnée comme perdue et à la traiter comme donnée manquante, l'autre à enquêter sur la situation ou analyser le problème, puis à « corriger » en conséquence, si possible, la valeur suspecte. Dans tous les cas, mieux vaut prendre conseil !

VÉRIFICATION DE L'HYPOTHÈSE DE RECHERCHE ET IMPACT DANS LE DOMAINE DE RECHERCHE

Vérification statistique vs conclusion

Les données enregistrées en expérimentation servent, au moment du traitement statistique, à nous indiquer s'il est apparu des différences tel que prévu par l'hypothèse de recherche, et si ces différences sont significatives, sérieuses, généralisables. Pour confirmer (ou valider) l'hypothèse de recherche, il est donc nécessaire que les tests statistiques appropriés soient eux-mêmes valides et significatifs. Toutefois, la seule obtention de tests significatifs ne suffit pas à démontrer la vérité ou la pertinence de l'hypothèse de recherche :

encore faut-il que les conditions d'expérimentation en garantissent l'interprétation. Prenons l'exemple de deux groupes, un groupe expérimental (E) et un groupe témoin (T), chacun constitué de 10 sujets *sans répartition aléatoire*. Si, *avant* l'imposition du traitement au groupe E, on mesurait les sujets et on obtenait une différence significative (en faveur du groupe E), l'obtention d'un test significatif *après* expérimentation ne nous permettrait pas de conclure honnêtement en faveur de l'influence du traitement expérimental. Si les effets de séquence n'ont pas été contrôlés dans un protocole à mesures répétées, si les sujets ont été mesurés à des moments différents de la journée ou dans un environnement qui a changé, en un mot, si toutes les précautions d'usage n'ont pas été prises en vue d'en assurer la validité interne, les résultats d'expérience ne seront pas généralisables et n'auront pas la crédibilité nécessaire pour valider l'hypothèse, quels que soient les traitements statistiques qu'on leur fera subir.

Significativité et portée d'une expérience

Une expérience peut ne pas produire de résultats significatifs pour différentes raisons :

- l'hypothèse et la conception théorique sous-jacente sont erronées ou incomplètes ;
- trop de variables interférentes n'ont pas été contrôlées ;
- la puissance statistique n'était pas suffisante ;
- la VD telle que définie n'est pas suffisamment sensible ;
- l'organisation expérimentale est en cause (protocole, type d'échantillonnage, lieu d'expérimentation, etc.).

Un résultat peut, par contre, être significatif par accident. Ainsi, on affirme par exemple qu'une différence obtenue est significative au seuil de 5 % : la clause « au seuil de 5 % » signifie, littéralement, qu'on peut et qu'on va retrouver une différence aussi grande dans (au plus) 5 % des cas *même lorsque aucun traitement, aucune variation systématique ne sera intervenue* et que seul le

hasard sera à l'œuvre ! La prudence, donc, est de mise, et un résultat significatif ne constituera une loi que lorsque d'autres expérimentations dans la même veine l'auront reproduit et confirmé. Ainsi, le principe de Fisher, qui propose le double critère de la significativité et de la réplication des résultats d'expérience, demeure essentiel à tout progrès réel de la recherche expérimentale.

Impact dans le domaine de recherche

L'impact d'une recherche particulière, de ses résultats, de ses conclusions sur le thème et le domaine de recherche est difficile à évaluer, sauf dans le cas d'une réussite exceptionnelle. Le plus souvent, même l'obtention de résultats significatifs – qui confirment une conception théorique et font certainement plaisir au chercheur – n'apporte pas de progrès décisif dans l'étude du thème de recherche. Paradoxalement, ces progrès vont souvent provenir de résultats inattendus, survenus en cours d'expérimentation, ou bien d'une remise en question des méthodes de contrôle ou de mesure consécutive à des résultats négatifs ou non significatifs. Il faut peut-être se le remettre en mémoire : le but de la recherche n'est pas d'obtenir des résultats significatifs, voire de confirmer nos hypothèses, mais c'est vraiment de mieux connaître et comprendre le phénomène étudié.

Beaugrand (1988a) discute aussi de l'insertion d'une recherche et d'un programme de recherche dans l'élaboration de théories, en science.

ANNEXE A

Déontologie de la recherche

La déontologie est l'ensemble des règles et des devoirs qui régissent une profession, la conduite de ceux qui l'exercent, les rapports entre ceux-ci et leurs clients ou le public.

(Le Petit Larousse)

Le chercheur, au-delà des infractions d'ordre commun ou des crimes qu'il peut commettre, est imputable aussi de ses manquements aux règles déontologiques qui gouvernent son action.

La déontologie du chercheur, au Canada, est établie par divers organismes (CRSHC, CRSNG, IRSC, IRSST, etc.), leurs conventions se recoupant largement les unes les autres. La Société canadienne de Psychologie et l'American Psychological Association, par exemple, ont aussi leur code de déontologie du chercheur. Les universités ont elles-mêmes un code général et un comité institutionnel de déontologie : le chercheur (professeur, professionnel, étudiant) doit remplir un formulaire d'aval déontologique qui, dans le cadre d'une demande de subvention, doit être soumis et accepté conjointement à la demande et qui est une pièce obligée pour le dépôt d'un mémoire de recherche ou d'une thèse.

QUAND SE POSE LA QUESTION DÉONTOLOGIQUE ?

La question déontologique se pose toujours, dès que le chercheur (étudiant de 1^{er} cycle ou en recherche, chercheur universitaire, chercheur industriel) se met en action et réalise des opérations qui utilisent des humains ou des animaux ou qui se répercutent ou peuvent se répercuter sur eux.

- La recherche met-elle en jeu des risques non négligeables d'atteinte à l'intégrité physique, morale, sociale ou psychologique des personnes ?
- Les bénéfices probables, attendus de la recherche, compensent-ils les inconvénients ou dommages vécus par chacun pour y participer ?

Les risques peuvent découler de l'expérimentation ou tenir à l'état de santé des participants. Quant à la compensation, elle peut n'être qu'anticipée ou future, ou ne toucher qu'une catégorie de participants, par exemple les sujets du groupe expérimental ; dans ces cas, elle ne peut pas servir à justifier les dommages ou inconvénients encourus par les sujets.

RÈGLES DÉONTOLOGIQUES

De l'élaboration à la publication des résultats d'une recherche, la « bonne conduite morale » du chercheur obéit à certaines règles. La liste suivante, tirée de Sabourin et Bélanger (1988), en donne une énumération :

Règle 1 – La bonne intention

Aucune recherche sur la personne humaine ne doit être entreprise si elle n'a pas pour but ultime l'acquisition de connaissances susceptibles de contribuer à l'amélioration de l'état et des conditions de vie de l'individu et de la société.

Règle 2 – L'importance du risque anticipé

Aucune recherche n'est justifiable si elle fait courir au sujet des risques démesurés.

Règle 3 – La conduite compétente

Le chercheur a l'obligation d'élaborer le meilleur projet de recherche dont il est capable.

Règle 4 – Le consentement du participant

Aucune recherche ne doit se faire sur la personne humaine si celle-ci n'a pas donné un consentement libre.

Règle 5 – Le consentement éclairé

Aucune recherche ne doit se faire sur la personne humaine si celle-ci n'a pas donné un consentement éclairé.

Règle 6 – La responsabilité continue

Une fois la recherche terminée, le chercheur doit veiller à l'élimination des effets consécutifs négatifs.

Règle 7 – La responsabilité globale

Le chercheur est pleinement responsable de la conduite de la recherche.

Règle 8 – L'objectivité

Le chercheur doit analyser et interpréter objectivement les résultats.

Règle 9 – La prudence dans les conclusions

Le chercheur doit reconnaître la portée limitée de ses interprétations et veiller à ce qu'elles ne soient pas utilisées à des fins autres que celles prévues.

Règle 10 – L'anonymat et la confidentialité

Le chercheur a le devoir de protéger l'anonymat des sujets et de maintenir la confidentialité des données.

L'obligation faite au chercheur d'une observance déontologique est relativement récente. Elle est tributaire en partie de l'importance grandissante des enjeux et des risques de la recherche, surtout dans le domaine médical. Elle dépend fortement aussi de la judiciarisation croissante de notre société (en vertu de laquelle les citoyens recourent volontiers et plus fréquemment aux tribunaux). Cette obligation est incontournable.

En pratique, il importe au chercheur de savoir quand la question déontologique pose problème et de s'acquitter de son obligation en respectant les règles applicables à sa situation. On peut distinguer trois catégories de recherches du point de vue déontologique :

- les recherches qui ne mettent en œuvre aucun moyen, dispositif ou intervention comportant un risque d'atteinte à l'intégrité (physique, morale, sociale, psychologique) du participant et qui recrutent parmi une population « normale » ;
- les recherches qui mettent en œuvre des moyens, dispositifs ou interventions qui portent atteinte ou risquent de porter atteinte à l'intégrité du participant ;
- les recherches qui recrutent des participants dans une population dite vulnérable.

Par exemple, une recherche classique portant sur la mémorisation de listes de mots et employant comme participants des étudiants universitaires de premier cycle tombe dans la première catégorie et ne pose pas problème du point de vue déontologique.

Dans la seconde catégorie, une autre recherche, utilisant encore des étudiants universitaires, leur propose d'ingérer une boisson alcoolisée afin de vérifier comment un taux d'alcoolémie donné influence la perception d'une scène ambiguë (ou la mémorisation, ou le temps de réaction à choix); l'administration d'un questionnaire sur les « tendances suicidaires » ou les « relations familiales », ou d'un test de personnalité (projectif ou non), tombe aussi dans cette catégorie.

La troisième catégorie regroupe toutes les recherches et études qui impliquent des populations vulnérables ou des personnes dites à risque, quelle qu'en soit la définition. Les études sur la mémoire des personnes diabétiques, le temps de réaction des personnes âgées, toute expérimentation sur des enfants, ainsi de suite et, enfin, la plupart des problématiques reliées à la santé physique ou mentale s'y retrouvent.

Les recherches associées aux catégories deux et trois exigent du chercheur qu'il motive son entreprise et en justifie les moyens en regard des risques potentiels et des bénéfices escomptés pour l'ensemble des participants. Cette exigence est d'ordre éthique et elle est imposée par les comités déontologiques dont se sont pourvus tous les organismes qui encadrent les chercheurs (gouvernements, universités, industries).

L'étudiant de cycles supérieurs ou le professeur qui entreprend une recherche dans une université québécoise doit remplir et soumettre un formulaire d'aval déontologique pour tout projet d'expérimentation impliquant l'utilisation de sujets humains ou animaux. Le formulaire (il en existe ordinairement un pour les expériences avec des humains, un autre pour les animaux) est disponible sur le site Internet de l'institution ou il peut être obtenu auprès du service de la recherche, en même temps que la politique institutionnelle correspondante. Chaque université a mis en place des instances compétentes pour recevoir ces formulaires, analyser les projets et leur attribuer la sanction déontologique appropriée.

Pour une expérimentation que l'étudiant doit réaliser dans un cours (p. ex., un cours de méthodes de recherche), d'autres procédures sont en place, habituellement chapeautées par le professeur et avalisées au niveau du département ou de la faculté.

LECTURES COMPLÉMENTAIRES

ALLARD, M., S. DION et S. BOUCHARD (2005). « La recherche et l'éthique », dans S. Bouchard et C. Cyr (dir.), *Recherche psychosociale. Pour harmoniser recherche et pratique*, Sainte-Foy, Presses de l'Université du Québec, p. 483-508.

BOUCHARD, G., S. SABOURIN, Y. LUSSIER et J. WRIGHT (2000). « Les perspectives déontologiques en recherche », dans R.J. Vallerand et U. Hess (dir.), *Méthodes de recherche en psychologie*, Boucherville, Gaëtan Morin, p. 483-506.

DOUCET, H. (2002). *L'éthique de la recherche*, Montréal, Presses de l'Université de Montréal.

SABOURIN, M. et D. BÉLANGER (1988). « Règles de déontologie en recherche », dans M. Robert (dir.), *Fondements et étapes de la recherche scientifique en psychologie* (3^e édition), Saint-Hyacinthe, Edisem, p. 367-397.

SALES, B.D. et S. FOLKMAN (dir.). (2000). *Ethics in research with human participants*, Washington D.C., American Psychological Association.

Voir aussi les sites : <www.crsh.ca> <www.crsng.ca> <www.irsc.ca>

ANNEXE B

Éléments de la théorie des tests

QU'EST-CE QUE LA THÉORIE DES TESTS ?

Que vaut la mesure produite par un test psychologique ? Le score obtenu à un test a-t-il des propriétés comparables à d'autres mesures, comme le poids, le taux d'alcool dans le sang (alcoolémie), la force de flexion du bras droit ? Le QI obtenu à partir d'un test (p. ex., le EIHM) est-il plus précis, préférable, au QI dérivé d'un autre test (p. ex., le WISC-III) ? Le score d'une échelle d'anxiété générale représente-t-il vraiment l'anxiété ?

Pour les besoins de la pratique clinique, de la sélection de personnel autant que de la recherche, les psychologues ont développé des tests et des procédés de mesure de toutes sortes. Ils ont été confrontés aussi à la contestation, à l'incrédulité, aux critiques, et ils ont dû développer une approche réfléchie et scientifiquement cohérente de la mesure. Les psychologues scolaires les ont rejoints, avec les psychologues industriels, les docimologues, les orienteurs ; vers les années 1950, cet effort collectif a produit ce qu'on appelle désormais « la théorie classique des tests ».

Réduite à sa plus simple expression, la théorie des tests est constituée par une doctrine algébrique de la mesure, basée sur la notion d'une erreur de mesure aléatoire, une collection un peu floue de concepts relatifs aux qualités de la mesure, et

un arsenal de techniques et de procédés méthodologiques pour élaborer, analyser et valider les instruments de mesure, particulièrement pour les mesures en sciences humaines et en sciences de la vie.

Nous présentons ici, en esquisse, le modèle de base de la mesure (selon la théorie des tests), les concepts liés de fidélité et de précision (avec celui d'erreur-type de la mesure), les méthodes d'estimation les plus utiles pour le coefficient de fidélité et l'erreur-type de mesure, et un aperçu du concept polysémique de validité.

LE MODÈLE DE BASE DE LA MESURE

Mesurer consiste à appliquer un procédé de mesure (un test) à un objet (une personne) pour obtenir un indice numérique d'une de ses qualités. Une définition de mesure : « une fonction d'information qui exprime symboliquement la grandeur d'une dimension d'un objet » suggère (1) que la mesure n'exprime pas tout l'objet mais rien qu'une dimension définie de l'objet ; (2) que la « grandeur » de ce trait chez l'individu évalué peut et va être convertie en une quantité symbolique, un nombre le plus souvent ; (3) qu'il s'agit d'une fonction d'information sujette à toutes sortes d'erreurs et de biais.

La théorie des tests pose que chaque mesure X est composée à la fois d'une partie qui reflète la valeur réelle, ou valeur vraie (V) de la personne, et d'une autre partie (e) qui représente une *fluctuation*, une quantité aléatoire, changeante d'une fois à l'autre, et qui contamine peu ou prou la mesure :

$$X_{i,o} = V_i + e_o ; \quad (1)$$

la mesure de l'individu i à l'occasion o est la somme de la valeur vraie de l'individu et d'une « erreur » aléatoire liée à l'occasion de mesure.

C'est ainsi que, si Robert, par exemple, est mesuré 5 fois de suite (dans des conditions comparables), il va obtenir 5 mesures différentes, les 5 ayant la même composante V_{Robert} caractéristique de Robert, et chacune ayant une fluctuation e_o nouvelle, qui reflète l'imprécision du test ou l'instabilité foncière de ce qui est mesuré.

Les variances observée, vraie et d'erreur ; le coefficient de fidélité ; l'erreur-type de mesure ; la capacité discriminante

Pourquoi mesure-t-on les personnes, les objets ? C'est généralement afin de les classer, de les choisir, d'en déterminer la valeur relative. Supposons maintenant que, en utilisant un test particulier (p. ex. le test d'intelligence EIHM, ou bien une mesure du VO₂ max), nous mesurons N personnes. Nous obtenons donc les mesures $\{X_1, X_2, \dots, X_N\}$, à raison d'une mesure par personne. La variance des mesures s_X^2 (qu'on dénomme ici « variance observée ») est un indice de dispersion, elle reflète jusqu'à quel point notre test parvient à séparer, discriminer, classer les personnes l'une par rapport à l'autre. Or, on l'a vu, chaque mesure (d'une personne) est composée à la fois d'une valeur « vraie » et d'une quantité aléatoire, qui contribue elle aussi à la variance. En fait, on peut démontrer algébriquement l'égalité suivante :

$$s_X^2 = s_V^2 + s_e^2, \quad (2)$$

qui présente la « variance observée » comme étant la somme de la « variance vraie » (ou variance des valeurs vraies) et de la « variance d'erreur » (ou variance des erreurs).

Si, pour un instrument de mesure ou un test, les fluctuations, imprécisions, « erreurs » de mesure étaient inexistantes ou très petites, nous aurions

$$s_X^2 = s_V^2$$

et nous serions alors assurés que chaque personne évaluée reçoit sa juste mesure et que le classement produit reflète les différences réelles entre les sujets. Par contre, si nous avons

$$s_X^2 = s_e^2,$$

cela voudrait dire que chaque personne reçoit pour mesure un nombre aléatoire et que les personnes seront classées au pur hasard l'une par rapport à l'autre. Dans un tel cas, le classement obtenu après une première série de mesures (*i.e.* en plaçant les personnes de la première à la dernière, dans l'ordre des scores obtenus) sera complètement chamboulé dans une deuxième ou dans une troisième série de mesures des mêmes personnes.

L'examen de l'équation des variances (2) suggère ainsi qu'un test, un procédé de mesure, sera d'autant plus sûr, utile, précis, que la portion de variance vraie incluse dans la variance observée sera élevée, ou que la portion de variance d'erreur y sera faible. Le **coefficient de fidélité**, dénoté r_{XX} , est défini équivalamment par :

$$r_{XX} = s_V^2 / s_X^2 \quad (3a)$$

$$= 1 - s_e^2 / s_X^2 \quad (3b)$$

$$= s_V^2 / (s_V^2 + s_e^2) \quad (3c)$$

La **fidélité** peut être globalement définie comme cette propriété d'un test ou d'un instrument de mesure grâce à laquelle les mesures produites reflètent réellement les différences entre les personnes mesurées. Quant au coefficient de fidélité r_{XX} , c'est la proportion de variance vraie présente dans les mesures. Les formules symboliques ci-dessus sont faciles à interpréter. On constate aussi que la fidélité peut varier depuis $r_{XX} = 0$, dénotant un test ou des mesures ne contenant que du bruit ou des valeurs aléatoires, jusqu'à $r_{XX} = 1$, indiquant que chaque mesure reflète exactement la position ou la valeur de la personne évaluée.

Que les mesures soient plus ou moins « précises », selon leur degré de fidélité, implique que la valeur X_i attribuée à une personne mesurée par un test sera tout au plus une approximation de sa valeur vraie V_i .¹ En fait, l'*imprécision* associée à chaque mesure $X_{i,o}$ est tributaire de sa composante e_o , qui est tantôt positive tantôt négative

1. Pour les curieux, la corrélation entre la valeur observée $X_{i,o}$ et la valeur vraie V_i est estimée par $\sqrt{r_{XX}}$.

et varie de manière à peu près normale. La marge de variation de e_o est indiquée par son écart-type, s_e , et elle est appelée l'**erreur-type de mesure**, caractéristique de chaque test ou instrument de mesure. Par exemple, pour un test de QI doté d'une moyenne normative (imposée) de 100, d'un écart-type normatif de 15 et d'une fidélité $r_{XX} = 0,84$, l'erreur-type de mesure (voir plus bas) sera de $s_e = 6,0$. Supposons maintenant que, à ce test, Robert obtienne un QI global de 105. L'erreur-type de mesure nous indique une marge d'imprécision de 105 ± 6 . De plus, si on suppose que l'erreur de mesure est distribuée normalement, on peut affirmer par exemple que le « vrai QI » de Robert se situe dans l'intervalle $(X_i \pm 2s_e)$ avec une probabilité d'environ 95 %, soit dans le cas présent : $QI = (105 \pm 2 \times 6) = (93 ; 117)$. Au contraire du concept de fidélité, généralement confiné aux sciences humaines, le concept d'erreur-type de mesure est utilisé dans toutes les sciences.

Si chaque mesure était parfaitement précise, selon une fidélité (r_{XX}) de 1 et une erreur-type de mesure (s_e) de 0, chaque personne mesurée (ou chaque objet) se verrait décerner une valeur numérique fixe, sa valeur vraie, et tous les objets pourraient être classés, ou catégorisés numériquement, de façon sûre. Le nombre de catégories, ou classes de mesure, qu'on peut former à partir d'un instrument de mesure est appelé sa **capacité discriminante**. Théoriquement, cette capacité (virtuelle) est définie comme le quotient de l'étendue de mesure (E) divisée par l'unité de mesure (u), soit E/u . En réalité, cependant, grâce à son imprécision caractérisée par l'erreur-type de mesure, la valeur attribuée à une personne (un objet) est incertaine, flottante, et elle marque plutôt un intervalle numérique qui inclut la valeur vraie²; cette imprécision réduit aussi le nombre de catégories différentes parmi lesquelles les objets peuvent être classés avec sûreté. La capacité discriminante réelle (Dr) d'un instrument ou procédé de mesure est alors le nombre d'intervalles numériques, ou catégories, parmi lesquels tous les objets peuvent

2. L'intervalle de mesure inclut la « valeur vraie » si et seulement si la mesure n'est pas biaisée, *i.e.* en l'absence de toute erreur systématique.

être classés, en assurant une probabilité d'au moins 0,50 que l'objet occupe l'intervalle comprenant sa valeur vraie. Laurencelle (1998) fournit la formule d'estimation suivante :

$$D_r \approx \frac{2,654}{\sqrt{1 - r_{XX}}} \quad (4)$$

Bien entendu, plus grande est la capacité discriminante d'une mesure ou d'un procédé de mesure, plus grande aussi est son utilité pratique.

Méthodes d'estimation de la fidélité

Dans ce document d'introduction, nous nous contentons de présenter, sans les justifier mathématiquement, les méthodes d'estimation les plus courantes, avec leur recette. Les références bibliographiques fourniront au lecteur exigeant toute l'information qu'il souhaite.

La méthode du test-retest. La méthode la plus courante et la plus sûre pour obtenir une estimation du coefficient de fidélité r_{XX} d'un test ou d'un instrument de mesure est celle dite du test-retest. Elle consiste à recruter un échantillon raisonnable de personnes, à les mesurer une première fois puis, après un délai à déterminer, à les mesurer à nouveau, obtenant ainsi pour chaque personne les mesures X_1 (au test) et X_2 (au retest). La corrélation $r(X_1, X_2)$, entre les mesures du test et du retest, est un estimateur du coefficient r_{XX} , soit

$$\hat{r}_{XX} = r(X_1, X_2). \quad (5)$$

Par « échantillon raisonnable », on entend, pour cette méthode comme pour les deux autres plus loin, un groupe de personnes suffisamment nombreux (*i.e.* 30 et plus, idéalement quelques centaines), choisies surtout pour représenter l'étendue des valeurs (vraies) possibles, depuis celles situées au bas de l'échelle jusqu'au plus haut. Quant au délai entre test et retest, il doit assurer que les conditions du retest soient, matériellement comme pour chaque

participant, les mêmes que celles qui prévalaient au test : les possibles effets d'apprentissage, de lassitude, d'évolution de la personne, etc., doivent être considérés. Des auteurs donnent aussi à cet estimateur le nom de coefficient de stabilité temporelle.

La méthode de la corrélation entre les parties. La plupart des tests psychologiques sont de structure composite, les composantes étant des items, des questions, des sous-tests. Pour attribuer un score X_i à la personne testée, les points mérités à chaque item sont additionnés. En désignant par y_1 la valeur de l'item 1, par y_2 pour l'item 2, jusqu'à y_t pour le dernier item « t », nous avons

$$X_i = y_{1i} + y_{2i} + \dots + y_{ti}. \quad (6)$$

Cette structure, qu'on retrouve aussi ailleurs (p. ex. dans les épreuves d'habileté ou de performance en plusieurs essais), permet l'estimation de la fidélité *sans avoir à remesurer les personnes en retest*.

Pour appliquer la méthode de la corrélation entre les parties, il faut d'abord allouer les items en deux parties constituées d'un nombre égal d'éléments différents. Par exemple, à partir d'une échelle psychologique de 20 items, on peut constituer la partie 1 (p_1) en utilisant tous les items impairs (*i.e.* 1, 3, 5, ..., 19), la partie 2 (p_2) avec tous les items pairs (*i.e.* 2, 4, 6, ..., 20). Pour chaque personne testée, on calcule alors ses deux scores partiels, p_1 et p_2 , par la somme des items correspondants, on mesure la corrélation $r(p_1, p_2)$ entre ces deux séries de scores et, enfin, on applique la formule d'estimation³ :

$$\hat{r}_{XX} = kr(p_1, p_2) / [1 + (k - 1)r(p_1, p_2)]; \quad (7)$$

dans cette formule, « k » représente le quotient du nombre total d'items dans le test sur le nombre d'items dans chaque partie. Dans l'exemple développé ci-dessus, qui correspond à la bien nommée « méthode pairs/impairs », $k = 20/10 = 2$.

3. Dite « formule d'allongement » ou d'extrapolation de Spearman-Brown.

Les deux parties reconstituées, p_1 et p_2 , doivent pouvoir être jugées équivalentes, ou de compositions équivalentes. Leur taille relative importe peu : par exemple, pour un test d'arithmétique élémentaire constitué de 17 additions à deux chiffres, la partie p_1 pourrait englober les items 1 à 5, la partie p_2 les items 13 à 17 ; dans un tel cas, la valeur de k dans la formule ci-dessus serait égale à $17/5 = 3,4$.

Notons que la formule (7), ci-dessus, est d'application générale et qu'elle permet d'estimer la fidélité d'une mesure obtenue par le total de plusieurs items équivalents, en faisant varier le nombre d'items. Par exemple, si un examen d'arithmétique basé sur 10 problèmes (ou un test de précision basé sur 10 tirs au but) a une fidélité r_{XX} , le même examen augmenté à 20 ingrédients verra sa fidélité accrue : dans ce cas, $k = 20/10 = 2$, et $= 2r_{XX}/(1 + r_{XX})$. Le phénomène inverse, de réduction de la fidélité, advient lorsqu'on diminue le nombre d'ingrédients (entraînant $k < 1$).

La méthode du Alpha de Cronbach. La méthode du Alpha de Cronbach généralise la méthode de corrélation entre les parties. En effet, pourquoi utiliser deux parties plutôt que trois ou cinq, avec leurs intercorrélations ? Pourquoi placer tel item dans une partie plutôt que dans l'autre ? Le Alpha (en grec, « α ») considère séparément chaque item, chaque question, et base son estimation sur les $t \times (t - 1)/2$ intercorrélations de tous les items :

$$\hat{r}_{XX} = \alpha = \frac{t}{t-1} \left[1 - \frac{\sum_j s_j^2}{s_X^2} \right]; \quad (8)$$

la quantité t indique le nombre total d'items (ou questions, ou éléments) dans le test ; la somme (Σ) au numérateur concerne chacune des variances des t items (s_j^2), et s_X^2 est la variance observée, celle du score total tel qu'obtenu par l'addition des items (voir équation 6). Le coefficient α de Cronbach est aussi appelé coefficient de consistance interne (d'un test).

Méthodes d'estimation de l'erreur-type de mesure

L'erreur-type de mesure, égale à la racine carrée de la variance d'erreur dans le modèle de la théorie des tests, indique la marge d'imprécision entourant chaque mesure obtenue. Nous rapportons ici trois méthodes d'estimation ; les deux premières supposent, comme il arrive souvent, que nous disposions de deux séries de mesures équivalentes des mêmes personnes, telles que celles exploitées plus haut par la « méthode du test-retest ».

Par les différences entre test et retest. Si, pour chaque personne, nous disposons d'une mesure en test (X_1) et d'une en retest (X_2), nous pouvons en calculer la variation, la différence,

$$d_i = X_{i,1} - X_{i,2}$$

puis obtenir ensuite l'écart-type des valeurs d_i pour l'ensemble des personnes, soit s_d . L'estimateur qui en résulte,

$$s_e = s_d / \sqrt{2}, \quad (9)$$

est celui recommandé, à la fois pour sa précision et sa justesse.

Par le produit des écarts-types du test et du retest. Une autre méthode, réputée excellente, pour estimer l'erreur-type de mesure consiste à appliquer l'équation :

$$s_e = \sqrt{s_1 s_2 (1 - r_{XX})}, \quad (10)$$

basée elle aussi sur les mesures du test (l'écart-type s_1) et du retest (l'écart-type s_2) ; cet estimateur utilise le coefficient de fidélité r_{XX} .

Par la formule de définition de la fidélité. La formule (3b), qui définit le coefficient de fidélité, permet, après inversion, d'obtenir une estimation de la variance d'erreur, donc de l'erreur-type de mesure, soit :

$$s_e = s_X \sqrt{1 - r_{XX}}; \quad (11)$$

cet estimateur utilise seulement l'écart-type des scores, s_X (la racine carrée de la variance observée), et le coefficient de fidélité. Quoique moins performante, cette méthode est parfois la seule disponible ; on n'a pas toujours en main deux séries de mesures équivalentes des mêmes personnes. Par exemple, si on estime la fidélité d'un test composite par son coefficient α , il suffira de reprendre l'écart-type s_X , puis de calculer $s_e = s_X \sqrt{1 - \alpha}$.

Les concepts de validité

La validité est un concept plus complexe que celui de fidélité, et qui se rapporte non pas à l'instrument de mesure lui-même ou aux propriétés des mesures qu'il produit, mais à l'interprétation qu'on veut leur conférer. Jusqu'à quel point les différences mesurées par le test ou le procédé de mesure reflètent-elles les degrés d'intensité du caractère visé ? Pour les différentes personnes évaluées, que signifie au juste le classement résultant des scores obtenus ?

La question de validité ne se pose pas de la même façon pour tous les instruments de mesure. Par exemple, une épreuve d'arithmétique (20 problèmes d'addition à deux chiffres) constitue assurément un témoin de l'habileté en arithmétique, de même que le degré lu au thermomètre fournit assurément un indice de la température : ce sont des instruments à mesure directe. D'un autre côté, dans un test d'anxiété générale, la réponse autodéscriptive à 12 questions évoquant diverses situations de vécu personnel dépend du sérieux, de l'honnêteté, de l'aptitude à transposer, de l'interprétation qu'en fait le répondant, et dans ce cas la validité du score doit être démontrée. C'est aussi le cas pour tous les tests d'aptitudes, pour les échelles psychologiques (motivation, préjugés, introversion/extraversion, ...), pour les mesures prédictives, etc., soit l'ensemble des mesures indirectes ou interprétatives.

Il n'existe pas de « formule » de validité, en général, ni de méthode unique qui permette d'établir une fois pour toutes la validité d'un test. Il n'existe que des démonstrations de la validité, par lesquelles les métrologistes et les chercheurs tentent de prouver

que le test mesure réellement et utilement ce qu'il est supposé mesurer. Il est d'usage de regrouper ces démonstrations en catégories. En voici une taxinomie.

Validité de contenu. La démonstration de la validité de contenu vise à établir que, *grosso modo*, le contenu du test (ou procédé de mesure) reflète adéquatement la dimension visée, est une image ou une reconstitution de la réalité (biologique, psychologique, ...) concernée.

On distingue communément trois sous-catégories de la validité de contenu. La **validité manifeste** est indiquée simplement par l'adéquation directe, visible, entre contenus du test et réalité visée. La démonstration que la structure et la composition du test sont représentatives de la dimension visée, en sont une image analytiquement fidèle, établit la **validité échantillonnale** du test. Enfin, la preuve que les mesures produites par le test sont principalement influencées par la dimension (biologique, psychologique, ...) voulue, la démonstration que la « variance vraie » du test correspond au phénomène que le test est censé refléter, établissent la **validité conceptuelle** (en anglais, le *construct validity*) du test. L'analyse factorielle, une méthode sophistiquée de traitement d'un tableau d'intercorrélations, est parfois appliquée à cette fin, et l'on peut alors parler de validité factorielle.

Validité critérielle. La seconde grande catégorie de démonstration de la validité est la validité par critère : en plus de considérer le test à valider et ses mesures, on y a recours à un critère, *i.e.* une autre caractéristique, une autre mesure, un autre test, qui sert de critère de validation.

La **validité concomitante** (certains auteurs disent concurrente) forme la première sous-catégorie. Typiquement, le test à valider (disons X) est comparé à un autre test censé mesurer la même caractéristique (disons Y) : la corrélation entre les séries de mesures issues des deux tests indique la validité des deux. Quant à la seconde sous-catégorie, la **validité prédictive**, elle utilise un critère, une mesure, qui serait une conséquence, un effet, un corrélat, de la qualité évaluée par le test à valider. Par exemple, si ce

dernier est un test d'aptitude X, la mesure de la performance correspondante Y peut constituer un critère ; ou encore, si le test à valider (X) en est un d'intelligence, la moyenne scolaire (Y) pourrait en être une « conséquence » plausible. Encore ici, on utilise volontiers, mais non nécessairement, la corrélation $r(X,Y)$, pour indiquer la validité prédictive du test. (Notons que le critère n'a pas besoin d'être situé dans le futur du test, sauf lorsqu'on veut démontrer la capacité du test à prédire une performance à venir.)

RÉFÉRENCES

- ALLAIRE, D. et L. LAURENCELLE (1998). « Comparaison Monte Carlo de la précision de six estimateurs de la variance d'erreur d'un instrument de mesure », *Lettres statistiques*, 10, 27-50.
- LAURENCELLE, L. (1998). *Théorie et techniques de la mesure instrumentale*, Sainte-Foy, Presses de l'Université du Québec.
- LORD, F.M. et M.R. NOVICK (1968). *Statistical Theories of Mental Test Scores*, Reading (MA), Addison-Wesley.
- THORNDIKE, R.L. (dir.) (1971). *Educational Measurement* (2^e édition), Washington (DC), American Council on Education.

ANNEXE C

Quelques grandes idées sur l'échantillonnage

QUAND DOIT-ON SE PRÉOCCUPER D'ÉCHANTILLONNAGE ?

Toutes les études n'ont pas pour programme et objectif de faire le portrait d'une population, d'obtenir des informations et des descriptions qui s'appliquent à tous : cette obligation est le lot particulier des enquêtes globales, comme en entreprennent périodiquement les gouvernements auprès de leurs citoyens, et certaines grandes entreprises avec leurs employés ou clients. C'est aussi le cas des chercheurs qui, faute de temps ou de moyens financiers, doivent se contenter de mesurer seulement quelques personnes mais en gardant l'ambition de généraliser les résultats ainsi obtenus à toute la population. Le mode de sélection de ces personnes et leur nombre sont les deux paramètres qui peuvent assurer la *représentativité* de l'échantillon observé et en garantir la *généralisation* à la population.

LES TYPES DE POPULATION

Une « population » est un ensemble d'éléments, les unités échantillonnables, qui doivent répondre à certains critères d'admissibilité : p. ex. tous les citoyens payeurs d'impôts au

Québec, tous les étudiants de premier cycle universitaire en biologie, tous les diabétiques de type II inscrits à Diabète Québec, etc. Du point de vue de l'échantillonnage, la population peut être dite :

- *homogène* : les membres de la population présentent tous des caractéristiques individuelles comparables du point de vue de l'objet d'étude. Hormis une variabilité possible d'une personne à l'autre, on ne peut les regrouper conceptuellement en sous-groupes distincts ;
- *hétérogène non recensée* : la population comprend des classes de personnes, des sous-populations qui peuvent diverger quant à l'objet d'étude, des sous-groupes d'individus dont les réponses ou les mesures sont susceptibles de différer systématiquement. Toutefois, soit on n'en connaît pas la répartition dans la population, soit ces sous-groupes sont inaccessibles au recrutement sélectif ;
- *hétérogène recensée* : les sous-populations, ou sous-groupes, qui composent la population à l'étude sont connues, recensées et séparément accessibles.

Le recrutement dans une population qui serait homogène – un cas plutôt inhabituel dans les études descriptives – peut être mené comme on le fait en recherche expérimentale, *i.e.* en repérant et observant un petit nombre de participants admissibles ($n = 12$ à 30), sans plus. Pour une population hétérogène toutefois, cet échantillonnage naïf ferait courir le risque d'obtenir un groupe non représentatif, dont la composition démographique ne correspond pas à celle de la population.

ÉCHANTILLONS PROBABILISTES VS ÉCHANTILLONS NON PROBABILISTES

Par échantillon probabiliste, on entend tout échantillon obtenu au moyen d'un procédé de tirage au hasard dans la population ; l'échantillon non probabiliste n'implique pas l'utilisation expresse du hasard.

Les bonnes méthodes d'échantillonnage, celles qui assurent la représentativité de l'échantillon et la généralisation des données, utilisent le tirage au hasard : par exemple, dans l'échantillonnage au hasard simple, le chercheur dispose d'un registre, une liste, dans laquelle *tous* les membres de la population sont inscrits, et il y pige un par un, au hasard, ses n participants. L'échantillonnage par grappes, l'échantillonnage aléatoire stratifié, l'échantillonnage systématique (sur une liste censée aléatoire) sont aussi probabilistes.

Parmi les méthodes non probabilistes, on peut mentionner le volontariat (autosélection des personnes admissibles, intéressées à participer), l'échantillonnage par quotas (recrutement d'un nombre approprié de personnes dans chaque catégorie échantillonnale, sans tirage au hasard), l'échantillonnage accidentel (ou occasionnel, où l'on retient, p. ex. les n premiers individus rencontrés), etc.

Échantillonnage et représentativité

On ne peut pas assurer la représentativité d'un échantillon, on peut seulement en fournir les conditions favorables. Ces conditions incluent le fait que chaque élément de la population doit avoir une probabilité égale d'apparaître dans l'échantillon, ou chaque catégorie, une probabilité proportionnelle à son importance dans la population. Les échantillons non probabilistes ne respectent pas cet impératif.

Une autre condition nécessaire pour asseoir la représentativité de l'échantillon réside dans son nombre d'éléments : il est difficile pour un échantillon d'être représentatif d'une population, homogène ou non, s'il contient seulement deux ou quelques participants ! Le biais échantillonnal, qu'on peut définir spécifiquement¹ comme :

$$B = E(\bar{X}) - \mu_X,$$

1. Une définition générale du biais devrait tenir compte de toutes les caractéristiques présentes dans la population, caractéristiques qu'il s'agirait de retrouver dans l'échantillon. La définition spécifique que nous offrons se limite au concept de moyenne, ou « tendance centrale », comme c'est l'usage en pareil cas.

peut survenir lorsqu'on échantillonne une population hétérogène par une méthode qui ne garantit pas l'égalité de chances de toutes ses sous-populations ; par échantillonnage au hasard simple, ce biais est égal à 0.

La précision, quant à elle, est mesurée par une erreur-type, *i.e.* une marge d'imprécision, laquelle diminue à mesure que la taille globale n augmente. Cependant, l'effet du nombre n sur la précision (l'erreur-type) des statistiques de l'échantillon dépend lourdement du mode de sélection employé et de la composition de la population. La fameuse « loi de l'inverse de la racine carrée » en est un repère, soit :

$$\sigma(\bar{X}) = \sigma_X / \sqrt{n};$$

selon cette loi, qui s'applique aussi bien à une proportion, une variance, une corrélation, une médiane qu'à une moyenne, il faut quadrupler la taille d'échantillon pour doubler la précision de la moyenne. Or, cette loi (et cette formule) concerne seulement l'échantillonnage au hasard simple, dans lequel les éléments sont tirés au hasard et indépendamment les uns des autres. Dans un échantillon par grappes, par exemple, les éléments d'une grappe tendent à corrélérer positivement l'un avec l'autre, ce qui compense l'effet de la taille n , comme le montre la formule générique² :

$$\sigma(\bar{X}) = \sigma_X / \sqrt{n} \times \sqrt{1 + (n-1)\bar{\rho}};$$

à la limite, si la corrélation moyenne était parfaite ($\rightarrow 1$), l'erreur-type de la moyenne serait aussi grande que l'écart-type et nul avantage ne découlerait d'un échantillonnage nombreux. C'est le contraire qui se produit dans l'échantillonnage stratifié : si la variable de stratification choisie (p. ex. l'âge, le niveau socioéconomique) est

2. L'effet perturbateur de l'échantillonnage par grappes peut être plus facilement estimé en utilisant la *corrélacion intraclasse* ρ_1 , définie ici par $\rho_1 = \sigma_G^2 / (\sigma_G^2 + \sigma_e^2)$, où σ_G^2 est la variance observée d'une grappe à l'autre, et σ_e^2 la variance entre les éléments. Dans ce cas et en supposant des échantillons constitués de n éléments répartis également en k grappes, l'erreur-type de la moyenne équivaut à $\sigma(\bar{X}) = \sigma_X / \sqrt{n} \times \sqrt{1 + (n-k)\rho_1/k}$.

efficace, *i.e.* si elle corrèle suffisamment avec la variable à l'étude (p. ex. l'opinion électorale, la consommation d'électricité), la précision des moyennes augmentera plus vite que par la loi de l'inverse de la racine carrée. La formule symbolique suivante,

$$\sigma(\bar{X}) = \sigma_X / \sqrt{n} \times \sqrt{1 - \eta^2 + \frac{\eta^2}{k^2}},$$

dans laquelle η (êta) dénote une corrélation entre variable de stratification et variable dépendante, et k indique le nombre de strates pratiquées, permet d'espérer une précision presque parfaite moyennant une bonne stratégie de stratification et un peu de chance.

Échantillonnage dans une population restreinte (ou finie)

Les « populations » auxquelles réfèrent les statisticiens sont habituellement très nombreuses. Dénotons par N la taille de la population et par n celle de l'échantillon qu'on y prend : si la fraction échantillonnale est infime, soit $n/N \approx 0$, on peut considérer la population comme pratiquement infinie ; sinon elle est dite finie, ou restreinte. Certains échantillonnages, comme dans les enquêtes auprès des médecins omnipraticiens du Québec ou chez les concessionnaires d'automobiles d'une région, font appel à des populations finies, et l'échantillon prélevé peut atteindre une fraction de 5 à 20 %. La précision des statistiques dérivées de telles populations finies est comparativement plus grande, comme le montre l'erreur-type de la moyenne, ci-dessous :

$$\sigma(\bar{X}) = \frac{\sigma_X}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}.$$

En somme, sauf pour certains cas restreints, comme en recherche expérimentale avec des groupes artificiels tirés au hasard d'une pseudo-population homogène, le chercheur doit progresser prudemment dans l'élaboration de son plan d'échantillonnage : étudier sa population cible, veiller à contrer les biais de sélection potentiels, réfléchir et se documenter sur l'impact statistique de son mode d'échantillonnage. Chaque étude descriptive pose un problème nouveau.

RÉFÉRENCES

- ALLAIRE, D. (1995). *Comprendre la statistique* (Vol. 1), Sainte-Foy, Presses de l'Université du Québec, chap. 1 et 2.
- COCHRAN, W.G. (1977). *Sampling Techniques* (3^e édition), New York, Wiley.
- KISH, L. (1967). *Survey Sampling*, New York, Wiley.
- LAURENCELLE, L. (2005). « La représentativité d'un échantillon et son test par le Khi-deux », *Lettres statistiques*, 12, 16 p.
- VOYER, J.-P., P. VALOIS et B. RÉMILLARD (2000). « La sélection des participants », dans R.J. Vallerand et U. Hess (dir.), *Méthodes de recherche en psychologie*, Boucherville, Gaëtan Morin, p. 91-129.

ANNEXE D

Contenus et présentation d'un rapport de recherche

PRÉSENTATION MATÉRIELLE GLOBALE

Le texte est produit (par traitement de texte) sur des feuilles de format lettre (8,5" × 11")¹, au recto, avec une police de caractères sobre, typiquement la police Times à 12-13 points ou Arial à 11-12 points.

Le texte doit généralement respecter des marges de 1,5" à gauche et en haut, 1" à droite et en bas ; il est écrit à interligne double. La pagination apparaît normalement à droite, à 1" du haut et du bord droit de la page. Les pages liminaires (depuis la page titre jusqu'à l'Introduction exclusivement) sont numérotées en chiffres romains ; l'Introduction démarre, en chiffres arabes, au numéro de page 1. Sur la page titre de chaque section ainsi que la première page de texte de la section (p. ex., Sommaire, Table des matières, Liste des tableaux, Introduction, etc.), le numéro de page est escamoté. Enfin, le texte débutant une section commence à 2,5" du haut de la page. Dans les rapports de recherche de petit volume, plutôt que de séparer le

1. Les dimensions, fournies en pouces (") puisque le papier utilisé répond au système anglais, peuvent être converties en centimètres par le facteur 2,54. Ainsi, 1,5" vaut 3,81 cm.

texte en chapitres distincts, on peut enchaîner les sections l'une à l'autre, en ménageant un espacement significatif pour chaque début de section.

Les titres et sous-titres des différentes sections doivent respecter une hiérarchie esthétique et uniforme.

La structure et le contenu de la page titre, de même que certains autres aspects du mémoire ou de la thèse, sont ordinairement régis par les instances auprès desquelles le document doit être déposé : le rédacteur doit s'y conformer obligatoirement.

SÉQUENCE TYPIQUE DES SECTIONS DU RAPPORT

Le rapport de recherche, le mémoire et la thèse comportent différentes sections, qui respectent la séquence suivante, à de menues variations près :

- la page titre définit la page i ;
- le Sommaire (ou Résumé) facultatif ;
- la Table des matières ;
- la Liste des tableaux facultatif ;
- la Liste des figures facultatif ;
- les Remerciements facultatif ;
- l'Introduction
(ou Introduction et problématique) définit la page 1 ;
- le Contexte théorique (ou titre à l'avenant) ;
- la Méthode (ou Méthodologie) ;
- les Résultats (ou Résultats et interprétation) ;
- la Discussion (ou Discussion et conclusion) selon ;

- la Conclusion selon ;
- les Références ;
- les Appendices (ou Annexes) facultatif.

INDICATIONS SUR LE SOMMAIRE

En un seul paragraphe, on présente un résumé de la question de recherche, du projet, des résultats et des grandes conclusions. Le Sommaire ne déborde pas deux pages.

INDICATIONS SUR L'INTRODUCTION

En une ou quelques pages, l'auteur présente son sujet et son intérêt de recherche, en indiquant les grands thèmes et les auteurs principaux qu'il abordera. Il peut fournir aussi un aperçu des sections qui suivent.

Une fonction importante de cette section est de poser le vocabulaire qui sera employé dans tout le rapport.

INDICATIONS SUR LE CONTEXTE THÉORIQUE

L'auteur doit développer ici sa question de recherche de A à Z afin d'en étayer la problématique. Ce développement, dont la logique incombe à l'auteur, s'appuie essentiellement sur la documentation scientifique : référence à des articles, reprise des conclusions, comparaison des méthodes, etc. Les variables les plus importantes du thème de recherche sont cernées, les méthodologies inventoriées.

Dans le cas d'un traitement plus fouillé du sujet, les expériences publiées sur un thème proche du sujet de recherche sont revues en détail et, au besoin, critiquées.

Cette section se termine normalement par l'énoncé d'une hypothèse de recherche, pour des recherches de type expérimental, ou des objectifs immédiats de recherche, pour les autres types.

INDICATIONS SUR LA MÉTHODE

On décrit ici en détail les interventions méthodologiques qu'on a mises en place pour réaliser concrètement la recherche. Cette section inclut les sous-sections (facultatives) : Participants, Instruments de mesure, Matériels et tâche, Disposition du lieu d'expérimentation, Protocole d'expérimentation, Calendrier de la recherche, Déroulement d'une séance de mesures, Postvérification des contrôles, Méthodes d'analyse statistique employées. (Les titres sont purement indicatifs.)

Les descriptions fournies doivent être claires et exhaustives : ce sont elles qui permettent d'asseoir la crédibilité des conclusions et garantissent la reproductibilité de l'expérience. Notons qu'on emploie ordinairement les verbes à l'indicatif présent dans cette section.

La section Participants doit faire état du consentement libre et éclairé que chaque personne a accordé en s'enrôlant dans le protocole de recherche ; l'observance des normes déontologiques applicables doit aussi y être déclarée.

INDICATIONS SUR LES RÉSULTATS (OU RÉSULTATS ET DISCUSSION) OU LA DISCUSSION

Le texte présente les résultats (aussi, tableaux de moyennes ou pourcentages, figures, etc.), en les « lisant » avec le lecteur et en les interprétant à mesure : des données ou tableaux n'ont d'existence que si le texte y réfère. Les données brutes elles-mêmes sont

ordinairement reléguées dans un Appendice s'il y a lieu, de même que les calculs détaillés s'il s'agit de calculs ou de traitements mathématiques hors de l'ordinaire.

Si l'abondance des résultats le justifie, on peut séparer les sections Résultats et Discussion, les résultats étant d'abord sommairement interprétés, puis synthétisés et discutés dans une section subséquente ; sinon, la discussion se fait au fur et à mesure de la présentation des résultats. Si les résultats se présentent en deux ou plusieurs catégories, la section peut se clôturer sur une Discussion générale.

Notons que les discussions doivent faire intervenir la documentation pertinente (vue en une section précédente, ou nouvelle) et s'articuler sur les hypothèses ou objectifs de la recherche.

Les tableaux sont numérotés 1, 2, etc., selon l'ordre d'apparition dans le rapport ; de même, indépendamment, pour les figures.

INDICATIONS SUR LA CONCLUSION

Bref résumé, encore, avec une conclusion à propos de l'atteinte des objectifs de recherche ou de la vérification de l'hypothèse de recherche. Il est d'usage, quoique non requis, d'ajouter des recommandations (ou des généralisations, si l'expérience s'est avérée positive), soit d'ordre pratique, soit pour la recherche future. Il est d'usage aussi de faire état des limites de la recherche, qui imposent en même temps des limites à la généralité de la conclusion et au poids des recommandations.

Dans un très court rapport, on peut adjoindre la conclusion à la discussion, dans une section Discussion et conclusion.

INDICATIONS SUR LES RÉFÉRENCES

On énumère ici, par ordre alphabétique d'auteurs, les références documentaires utilisées dans le rapport. Il existe quelques modèles de rédaction d'une référence. En voici trois importants exemples :

Référence à un article

BECKER, L.J. et C. SELIGMAN (1981). « Welcome to the energy crisis », *Journal of Social Issues*, 37, 1-7.

Référence à un livre

BERNIER, J.J. (1985). *Théorie des tests : Principes et techniques de base* (2^e édition), Boucherville, Gaëtan Morin.

Référence à un chapitre dans un volume « collectif »

RIVARD, V. et S. BOUCHARD (2005). « Les protocoles à cas unique », dans S. Bouchard et C. Cyr (dir.), *Recherche psychosociale* (2^e édition), Sainte-Foy, Presses de l'Université du Québec, p. 207-243.

INDICATIONS POUR LES APPENDICES

On peut ajouter un Appendice (ou Annexe) A, un Appendice B, etc., contenant du matériel complémentaire. Par exemple, un appendice peut contenir tous les tableaux de données brutes (dont la numérotation est solidaire de celle des tableaux du rapport), un autre qui reproduit *in extenso* un questionnaire ou un test psychologique, un qui donne le détail des calculs d'une analyse (statistique, psychométrie, etc.) complexe, etc.

Le style et le format de présentation des appendices sont ordinairement « plus libres » : on y retrouve parfois des sorties d'ordinateur.

RÉFÉRENCES

Les normes de rédaction et de présentation esquissées ici s'appuient essentiellement sur la première référence ci-après. L'apprenti rédacteur peut aussi consulter les autres références. Les rapports, mémoires et thèses déjà publiés, respectant ces mêmes normes et qui sont disponibles en bibliothèque constituent autant de modèles intéressants à examiner.

AMERICAN PSYCHOLOGICAL ASSOCIATION (2001). *Publication Manual of the American Psychological Association* (5^e édition), Washington, DC.

PROVOST, M.A., M. ALAIN, Y. LEROUX, et Y. LUSSIER (2002). *Normes de présentation d'un travail de recherche*, Trois-Rivières, SMG.

Guide de présentation des travaux, sur <www.uqtr.ca/dsap/Maitrise/index.html >.

RÉFÉRENCES

RÉFÉRENCES PRINCIPALES

- BOUCHARD, S. et C. CYR (dir.) (2005). *Recherche psychosociale*, (2^e édition), Sainte-Foy, Presses de l'Université du Québec.
- COUTURE, M. et R.-P. FOURNIER (dir.) (1997). *La recherche en sciences et en génie. Guide pratique et méthodologique*, Sainte-Foy, Presses de l'Université Laval.
- KERLINGER, F.N. (1986). *Foundations of Behavioral Research* (3^e édition), New York, Holt, Rinehart and Winston.
- ROBERT, M. (dir.) (1988). *Fondements et étapes de la recherche scientifique en psychologie* (3^e édition), Saint-Hyacinthe (Qc), Edisem et Paris, Maloine.
- VALLERAND, R.J. et U. HESS (dir.) (2000). *Méthodes de recherche en psychologie*, Boucherville, Gaëtan Morin.

AUTRES RÉFÉRENCES

- BEAUGRAND, J. (1988a). « Démarche scientifique et cycle de la recherche », dans M. Robert (dir.), *Fondements et étapes de la recherche scientifique en psychologie* (3^e édition), Saint-Hyacinthe (Qc), Edisem et Paris, Maloine, p. 1-35.

- BEAUGRAND, J. (1988b). « Observation directe du comportement », dans M. Robert (dir.), *Fondements et étapes de la recherche scientifique en psychologie* (3^e édition), Saint-Hyacinthe (Qc), Edisem et Paris, Maloine, p. 277-310.
- BRATLEY, P., B.L. FOX et L.E. SCHRAGE (1987). « A guide to simulation » (2^e édition), New York, Springer Verlag.
- CAMPBELL, D.T. et J.C. STANLEY (1966). *Experimental and Quasi-experimental Designs for Research*, Chicago, Rand McNally.
- CLÉMENT, G. et L. LAURENCELLE (1979). « Contrebalancement par permutations minimales », *Lettres statistiques*, 5, chap. 3, 9 p.
- COCHRAN, W.G. (1977). *Sampling Techniques* (3^e édition), New York, Wiley.
- COHEN, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2^e édition), Mawsah, Lawrence Erlbaum Associates.
- COUTU, S. M.A. PROVOST et F. BOWEN (2005). « L'observation systématique des comportements », dans S. Bouchard et C. Cyr (dir.), *Recherche psychosociale*, (2^e édition), Sainte-Foy, Presses de l'Université du Québec, p. 321-359.
- DESU, M.M. et D. RAGHAVARAO (1990). *Sample Size Methodology*, New York, Academic Press.
- EDGINGTON, E.S. (1980). *Randomization Tests*, New York, Marcel Dekker.
- FISHER, R.A. et F. YATES (1963). *Statistical Tables* (6^e édition), Edinburgh, Oliver and Boyd.
- GOYETTE, G. et M. LESSARD-HÉBERT (1987). *La recherche-action : ses fonctions, ses fondements et son instrumentation*, Sainte-Foy, Presses de l'Université du Québec.
- GUENTHER, W.C. (1973). *Concepts of Statistical Inference*, New York, McGraw-Hill.
- KIRK, R.R. (1994). *Experimental Design : Procedures for the Behavioral Sciences* (3^e édition), Belmont (CA), Brooks/Cole.

- KISH, L. (1967). *Survey Sampling*, New York, Wiley.
- LAURENCELLE, L. (2005a). « Les analyses statistiques », dans S. Bouchard et C. Cyr (dir.), *Recherche psychosociale*, (2^e édition), Sainte-Foy, Presses de l'Université du Québec, p. 361-406.
- LAURENCELLE, L. (2005b). « L'estimation de la puissance statistique à partir des données d'expérience », *Lettres statistiques*, 12, 22 p.
- LAURENCELLE, L. (2005c). « La représentativité d'un échantillon et son test par le Khi-deux », *Lettres statistiques*, 12, 16 p.
- LAURENCELLE, L. (2001). *Hasard, nombres aléatoires et méthode Monte Carlo*, Sainte-Foy, Presses de l'Université du Québec.
- LAURENCELLE, L. (1998). *Théorie et techniques de la mesure instrumentale*, Sainte-Foy, Presses de l'Université du Québec.
- LAURENCELLE, L. et F.A. DUPUIS (2000). *Tables statistiques expliquées et appliquées* (2^e édition), Sainte-Foy, Le Griffon d'argile.
- MERCIER, P. et M. GAGNON (2005). « Les protocoles de recherche pré, quasi et expérimentaux », dans S. Bouchard et C. Cyr (dir.), *Recherche psychosociale* (2^e édition), Sainte-Foy, Presses de l'Université du Québec, p. 87-148.
- NIETZSCHE, FRIEDRICH W. (1968). *Humain, trop humain*, Paris, Gallimard.
- REID, L. (2005). « Les sources d'invalidité et de biais », dans S. Bouchard et C. Cyr (dir.), *Recherche psychosociale*, (2^e édition), Sainte-Foy, Presses de l'Université du Québec, p. 27-86.
- ROBERT, M. (1988). « Validité, variables et contrôle », dans M. Robert (dir.), *Fondements et étapes de la recherche scientifique en psychologie* (3^e édition), Saint-Hyacinthe, Edisem et Paris, Maloine, p. 79-118

- SABOURIN, M. et D. BÉLANGER (1988). « Règles de déontologie en recherche », dans M. Robert (dir.), *Fondements et étapes de la recherche scientifique en psychologie* (3^e édition), Saint-Hyacinthe, Edisem, p. 367-397.
- SOKAL, R.R. et F.J. ROHLF (1981). *Biometry*, San Francisco, Freeman.
- WINER, B.J., D.R. BROWN et K.M. MICHELS (1991). *Statistical Principles in Experimental Design* (3^e édition), New York, McGraw-Hill.
- WOLF, F.M. (1986). *Meta-analysis : Quantitative Methods for Research Synthesis*, Beverly Hills (CA), Sage.

INDEX

A

alpha de Cronbach, 88

B

biais (d'estimation) statistique, 61

biais de l'observateur
voir Effets pervers

biais échantillonnal et volontariat, 95

biais échantillonnal
définition, 95
et représentativité
échantillonnale, 61

C

capacité discriminante *voir* Mesure

carré gréco-latin, 51

carré latin, 49

chirurgical (effet)
voir Effets pervers

contrebalancement
avec permutations minimales, 52
complet, 48

définition, 48
en carré gréco-latin, 51
en carré latin, 49
et calcul symbolique
d'équilibrage, 47
et répartition des sujets, 48
partiel, 49

corrélation intraclasse et
échantillonnage par grappes, 96
Cronbach (alpha de), 88

D

déontologie
définition, 75
formulaire d'aval
déontologique, 79
obligation faite au chercheur, 66
règles de conduite, 76
vs contrôle expérimental, 35

déterminisme, 1, 2
statistique, 2

données manquantes, 68, 71
techniques de remplacement, 68

double insu *voir* Insu

E

échantillonnage

- accidentel (occasionnel), 95
- au hasard (simple), 55, 95
- au hasard et erreur-type, 59, 96
- dans une population finie, 97
- et précision des moyennes, 58, 96
- et représentativité, 95
- et type de population, 93
- par grappes, 59, 96
- par quotas, 95
- probabiliste vs non
 - probabiliste, 94
- stratifié et erreur-type, 96

effets de séquence *voir* Effets pervers, Contrebalancement

effets pervers

- biais de l'observateur, 39
- chirurgical, 39
- effets de séquence, 47
- Hawthorne, 38
- nocebo, 37
- nouveauté, 38
- placebo, 37
- Rosenthal, 38
- sélection (des sujets), 39

effets principaux, 42

équivalence échantillonnale

- contraintes en recherche expérimentale, 56
- définition, 55
- en recherche pseudo-expérimentale, 55

erreur-type (d'une moyenne)

- exemple numérique, 70

erreur-type (d'une moyenne)

- et échantillonnage au hasard (simple), 96
- et échantillonnage en grappes, 59, 96
- et échantillonnage stratifié, 96

et nombre de sujets, 57

- et plan avec mesures répétées, 47
- tirée d'une population finie, 97
- vs type d'échantillonnage, 96

erreur-type de mesure *voir* Mesures techniques d'estimation, 89

éthique (du chercheur)

voir Déontologie

explication scientifique (principes), 2

F

fidélité *voir* Mesure

- et formule d'allongement d'un test, 88
- formules d'estimation, 84
- par corrélation entre deux parties, 87
- par le alpha de Cronbach, 88
- par test-retest, 86

formule d'allongement, 88

G

généralisation

- démographique, 54
- et test statistique, 69
- et vérification de l'hypothèse de recherche, 69
- fonctionnelle, 54, 56

H

Hawthorne (effet) *voir* Effets pervers

hypothèse de recherche, 25

- causes de non-vérification, 73
- définition, 29
- et conditions de vérification, 72
- nécessairement positive, 30
- vs H_1 (en statistique), 30
- vs objectifs de recherche, 26

I

insu (technique du simple ou double), 40
interaction (effets d'), 42

J

justesse (de mesure) *voir* Validité de mesure

M

mesure
capacité discriminante, 33, 85
erreur-type de mesure, 32, 85
exemple, 85
et aspect du phénomène, 31
et bonne distribution, 33
et distribution normale, 34
et procédé de mesure, 31
et théorie des tests, 81
et transformation normalisante, 34
fidélité, 32, 84
validité, 32 *voir* Validité (de mesure)
validité critérielle, 91
validité de contenu, 91
méta-analyse, 16
mortalité expérimentale, 67

N

nombre de sujets
et précision des moyennes, 57
et puissance statistique, 60
et représentativité échantillonnale, 62
minimal par échantillonnage au hasard, 63
nouveau(té) (effet de) *voir* Effets pervers

O

observation (directe)
et recherche éthologique, 13
et recherche expérimentale, 67
opérationnisme
définition, 30
et hypothèse de recherche, 26

P

parcimonie, 3, 4, 16
permutations minimales, 52
placebo (effet) *voir* Effets pervers
plan d'expérience
à facteurs emboîtés, 45
à groupes indépendants (simple ou factoriel), 41
à séquence temporelle, 45
avec contrebalancement, 45
avec mesures répétées (simple ou factoriel), 43
contrebalancé *voir* Contrebalancement
définition, 41
équilibré *voir* Contrebalancement mixte (p. ex., A × Br), 43
plan factoriel incomplet (à interactions confondues), 45
plan quasi (ou pré) expérimental, 45
plans factoriels et interaction, 42
sélection et répartition des sujets, 43
précision (de l'estimation) échantillonnale *voir* Erreur-type (d'une moyenne)
protocole d'expérience *voir* Plan d'expérience
protocole *vs* plan d'expérience, 41
puissance statistique, 60
estimation, 60
facteurs, 60

R

rasoir d'Occam *voir* Parcimonie

recherche (étude) de suivi, 18

recherche à cas unique, 13

vs recherche-action, 18

recherche-action

définition, 16

exemple, 16

vs recherche à cas unique, 18

recherche clinique, 18

recherche corrélacionnelle

définition, 12

n'égale pas « corrélation », 17

vs expérimentale, 17

recherche descriptive

définition, 12

et échantillonnage, 55

et généralisation démographique,
54

et modes d'échantillonnage, 56

exemple, 12

recherche-développement

définition, 14

exemple, 14

recherche épidémiologique, 19

recherche éthologique

définition, 13

recherche évaluative

définition, 13

vs expérimentale, 18

recherche *ex post facto*

voir Recherche évaluative

recherche expérimentale

définition, 11

et « robotisation », 67

et généralisation fonctionnelle, 54

et observation, 67

et plans d'expérience comparatifs,
41

exemple, 10

exemple statistique, 69

recherche longitudinale, 18

recherche préexpérimentale *voir*

Recherche quasi expérimentale

recherche pseudo-expérimentale

définition, 11

exemple, 11

vs expérimentale, 17

recherche quasi expérimentale

définition, 11

recherche théorique

définition, 15

exemple, 15

répartition aléatoire *voir* Équivalence

échantillonnale

représentativité échantillonnale

biais échantillonnal, 61

et effet de sélection, 39

et nombre de sujets, 62

et taille minimale pour

l'échantillonnage au hasard, 63

et test de représentativité, 63

et type d'échantillonnage, 95

et validité externe, 21

vs nature de la population, 61

Rosenthal (effet) *voir* Effets pervers

S

sélection (des sujets) *voir* Effets

pervers

par volontariat, 39

sélection des sujets *voir*

Échantillonnage

dans un plan d'expérience

à groupes indépendants, 43, 55

significativité statistique

et hypothèse de recherche, 72

et puissance statistique, 60

vs seuil de signification, 73

simple insu *voir* Insu

Spearman-Brown, 88

T

- test-retest
 - et la fidélité, 86
 - et l'erreur-type de mesure, 89
- thème de recherche, 23
- théorie des tests (classique), 81

V

- validité (de mesure), 90 *voir* Mesure
 - conceptuelle, 91
 - concomitante, 91
 - critérielle, 91
 - de contenu, 91
 - échantillonnale, 91
 - factorielle, 91
 - justesse et calibrage, 33
 - manifeste, 91
 - prédictive, 91
- validité externe d'une recherche, 21
 - et règles 6 et 7, 77
 - représentativité échantillonnale, 21
 - validité écologique, 21
 - validité interne d'une recherche
 - facteurs d'invalidité, 20
- validité interne d'une recherche, 20
 - et règle 3, 77
 - facteurs d'invalidité, 73
- variable contrôlée
 - définition, 28
 - et postvérification, 36
 - trois modes de contrôle, 36
- variable dépendante
 - définition, 27
- variable d'expérimentation
 - voir* Variable procédurale
- variable indépendante
 - définition, 27
 - en recherche pseudo-expérimentale, 27
 - et causalité en recherche expérimentale, 27
- variable non contrôlée, 28
 - apparition après coup, 37
- variable procédurale
 - voir* Effets pervers
 - définition, 28
- variables échantillonales, 28
- variance d'erreur *voir* Erreur-type
- vérification de l'hypothèse de recherche *voir* Hypothèse de recherche

M

anuel de référence dans un cours d'initiation à la recherche de premier ou de deuxième cycle, l'*Abrégé* est écrit dans un style concis et rigoureux. Il offre à l'utilisateur une vue d'ensemble et les principaux outils de la méthodologie scientifique. De la question de recherche jusqu'à la vérification de l'hypothèse : nature des variables en jeu, techniques de contrôle, effets pervers et méthodes d'insu, choix des instruments de mesure, modes d'échantillonnage et nombre de sujets, plans d'expérience, techniques de contrebalancement, l'auteur décrit, une à une, les étapes de la réalisation d'une recherche.

Indispensable pour l'étudiant en recherche, qui en complétera les contenus selon la spécificité de sa discipline et les besoins de sa formation, l'*Abrégé* sera encore utile au chercheur professionnel, qui n'est pas forcément un spécialiste de tous les aspects de la méthodologie et qui y trouvera une réponse, ou un début de réponse, à ses questions.

*LOUIS LAURENCELLE, Ph. D.,
a pratiqué les méthodes
de recherche comme
professeur, chercheur
et méthodologue-conseil
en différentes disciplines
depuis 1969. Il est
présentement professeur au
Département des sciences
de l'activité physique
à l'Université du Québec
à Trois-Rivières.*



ISBN 2-7605-1362-9