

Analyse de données et rédaction de rapports

ÉLABORATION ET MISE EN OEUVRE DE PROGRAMMES

DE RECHERCHE SUR LES SYSTÈMES DE SANTÉ



*Corlien M. Varkeuisser
Indra Pathmanathan
Ann Brownlee*

*Série sur la formation à la recherche
sur les systèmes de santé*

Volume 2, Deuxième partie



Le Centre de recherches pour le développement international, société d'État créée en 1970 par une loi du Parlement du Canada, a pour mission d'appuyer l'exécution de recherches qui, dans le domaine technique et dans celui des politiques, ont pour but d'adapter les sciences et la technologie aux besoins des pays en développement. Ses cinq secteurs d'activités sont : l'environnement et les richesses naturelles, les sciences sociales, les sciences de la santé, les sciences et systèmes d'information, et les initiatives et affaires institutionnelles. Le CRDI est financé par le Parlement du Canada, mais c'est un conseil des gouverneurs international qui en détermine l'orientation et les politiques. Établi à Ottawa, Canada, il a des bureaux régionaux en Afrique, en Asie, en Amérique latine et au Moyen-Orient.

L'Organisation mondiale de la santé est un organisme spécialisé des Nations Unies responsable des questions internationales touchant la santé et la santé publique. Par l'entremise de cet organisme, créé en 1948, les professions de la santé de quelque 165 pays échangent leurs connaissances et leur expérience dans le but de faire en sorte que d'ici l'an 2000, tous les citoyens du monde jouissent d'une santé suffisante pour leur permettre de mener une vie productive sur le plan social et économique.

Par une coopération technique directe avec les États membres et en suscitant la collaboration entre ceux-ci, l'OMS favorise le développement de services de santé complets, la prévention, la lutte contre les maladies, l'amélioration des conditions environnementales, le perfectionnement du personnel de santé, la coordination et l'élaboration de recherches biomédicales et de recherches sur les services de santé ainsi que la planification et la mise en oeuvre de programmes de santé.

La Série sur la formation à la recherche sur les systèmes de santé

- Volume 1 : La recherche sur les systèmes de santé : un outil de gestion (IDRC-286f)
Ann Brownlee
- Volume 2 : Élaboration et mise en oeuvre de programmes de recherche sur les systèmes de santé
Première partie — Formulation et mise à l'essai d'une proposition (IDRC-287f.1)
Deuxième partie — Analyse de données et rédaction de rapports (IDRC-287f.2)
Corlien Varkevisser, Indra Pathmanathan et Ann Brownlee
- Volume 3 : Stratégies permettant d'intéresser les universités et les instituts de recherche à la recherche sur les systèmes de santé
(IDRC-288f)
Ann Brownlee, Lilia Duran Gonzales et Indra Pathmanathan
- Volume 4 : Gestion de la recherche sur les systèmes de santé (IDRC-289f)
Indra Pathmanathan
- Volume 5 : La recherche sur les systèmes de santé : la formation des formateurs (IDRC-290f)
Indra Pathmanathan et N.I. Nik-Safiah

Pour de plus amples renseignements au sujet de ces publications, prière d'écrire à l'adresse suivante : CRDI, Marketing et distribution, Division des initiatives et affaires institutionnelles, BP 8500, Ottawa (Ontario) Canada K1G 3H9.

This publication is also available in English.

La edición española de esta publicación también se encuentra disponible.

Série sur la formation à la recherche sur les systèmes de santé

**Volume 2 : Élaboration et mise en oeuvre de programmes de recherche
sur les systèmes de santé**

**Deuxième partie :
Analyse de données et rédaction de rapports**

Le groupe de travail technique

Ann Brownlee (États-Unis)
Lilia Duran Gonzales (Mexique)
German Gonzales (Colombie)
Yvo Nuyens (Belgique)
Indra Pathmanathan (Malaysia)
Annette Stark (Canada)
Patrick Twumasi (Ghana)
Corlien M. Varkevisser (Pays-Bas)

Deuxième partie : Analyse de données et rédaction de rapports

**Élaboration et mise en oeuvre de programmes de recherche
sur les systèmes de santé**

Série sur la formation à la recherche sur les systèmes de santé
Volume 2

Corlien M. Varkevisser
Indra Pathmanathan
Ann Brownlee

Titre original de l'ouvrage : *Designing and conducting health systems research projects. Pt. 2, Data analysis and report writing*

© International Development Research Centre 1991

© Centre de recherches pour le développement international 1993
BP 8500, Ottawa (Ontario) Canada K1G 3H9

Varkevisser, C.M.
Pathmanathan, I.
Brownlee, A.

CRDI, Ottawa, Ont. CA
OMS, Genève CH

IDRC-287f.2

Élaboration et mise en oeuvre de programmes de recherche sur les systèmes de santé. 2^e partie, Analyse de données et rédaction de rapports. Ottawa, Ont., CRDI, 1993. xxi + 168 p. : ill. (Série sur la formation à la recherche sur les systèmes de santé; v. 2)

/Enquêtes sanitaires/, /projets de recherche/, /élaboration de projets/, /services de santé/ — /cours de formation/, /traitement des données/, /analyse des données/, /risques/, /études de cas/, /tests/, /recherche-participation/, /résultats de recherche/, /rapports de recherche/, /rédaction/, /références.

CDU : 613.001.5

ISBN : 0-88936-616-0

Édition microfiche offerte sur demande.

Les opinions émises dans la présente publication sont celles des auteures et ne reflètent pas nécessairement celles du Centre de recherches pour le développement international et de l'Organisation mondiale de la santé. La mention d'une marque déposée ne constitue pas une sanction du produit; elle ne sert qu'à informer le lecteur.

Résumé

Ce volume est le deuxième d'une collection de cinq volumes de formation à la recherche sur les systèmes de santé (RSS) qui ont été rassemblés par un groupe de travail technique financé par le Centre de recherches pour le développement international et l'Organisation mondiale de la santé. Chaque volume est destiné à un groupe particulier et chacun porte sur certains aspects de la recherche sur les systèmes de santé. Le volume 2, sous forme modulaire, est le volume central qui expose, étape par étape, la manière de formuler une proposition de RSS et de la mettre à l'essai (partie I) et d'analyser les données et de rédiger un rapport (partie II). Les participants des cours choisiront, de préférence avant les cours, les problèmes de santé qui sont prioritaires pour eux dans leur travail et qui ne pourront être résolus sans un supplément d'information. Dans la majorité des cas, les participants sont groupés en équipes pour faire la recherche planifiée en plus de leurs tâches ordinaires. Un second cours est ensuite organisé sur l'analyse de données, la rédaction de rapports et l'utilisation des résultats. Ce volume intéressera tous les groupes cibles et surtout les gestionnaires de soins de santé et les chercheurs de ce domaine qui veulent exécuter des travaux de recherche sur les systèmes de santé.

Les autres volumes de la collection sont les suivants : le volume 1 traite de la nécessité de promouvoir la RSS comme outil de gestion. Y sont décrites les stratégies propres à cette promotion auprès des décideurs et des cadres supérieurs. Le volume 3 vise à aider les chercheurs de formation universitaire qui travaillent dans des universités ou des instituts de recherche et qui veulent promouvoir des programmes multidisciplinaires de RSS et y participer. Le volume 4 est un guide de gestion d'un programme de RSS. Le volume 5 aidera les personnes chargées d'organiser et de donner des cours de formation aux divers groupes cibles.

Ces cinq volumes ont pour but d'appuyer la création d'un programme national de recherche essentielle en santé. Les personnes qui s'en serviront sont incitées à les examiner d'un oeil critique et à en tirer ce qui répond à leurs besoins ou y répondrait après adaptation.

Abstract

This is the second volume of a five-volume Health Systems Research (HSR) Training Series which has been compiled by a Technical Working Group supported by IDRC and WHO. Each volume is directed toward a particular target group and each addresses specific aspects of the HSR process. Volume 2, in modular format, is the pivotal one which deals step-by-step with the development of an HSR proposal and field testing (Part I), and with data analysis and report writing (Part II). Course participants will select, preferably in advance of the course, priority health problems particular to their own situations that cannot be solved unless more information is collected. In most cases, a team of course participants will then carry out the planned research alongside their regular duties. A second workshop is then scheduled to provide information on data analysis, report writing, and utilization of results. This volume will be of interest to all target groups and especially to those health-care managers and researchers who wish to conduct HSR projects.

The other volumes in the series are: volume 1, which focuses on the need to promote the use of HSR as management tool and reviews strategies for promoting HSR among policymakers and senior managers; volume 3, a review of strategies that can assist universities or research institutes to initiate and implement multidisciplinary HSR programs; volume 4, a course outline in modular format designed to provide research managers with the skills for managing a program of HSR; volume 5, a course outline in modular format, designed to assist those whose primary responsibility is organizing and conducting training courses for the relevant target groups.

The series is designed to support a program of essential national health research. Users are encouraged to critically examine the materials and to choose or adapt them to their particular needs.

Resumen

Este es el segundo de cinco volúmenes de una serie de capacitación sobre Investigación de Sistemas de Salud (ISS), compilada por un Grupo de Trabajo Técnico que recibió el apoyo del Centro Internacional de Investigaciones para el Desarrollo (CIID) y la Organización Mundial de la Salud (OMS). Cada volumen está dirigido hacia un grupo particular y trata de aspectos específicos del proceso de ISS. El Volumen 2, en formato modular, es un elemento fundamental que trata progresivamente del desarrollo de una propuesta de ISS y su prueba sobre el terreno (Parte I). Asimismo, se trata en este volumen el análisis de datos y la redacción de informes (Parte II). Los participantes del curso seleccionarán, preferentemente con antelación al curso, problemas de salud prioritarios, específicos de sus propias situaciones e imposibles de resolver hasta que no se recopile más información. En la mayoría de los casos, un equipo de participantes del curso llevará a cabo la investigación planificada conjuntamente con sus deberes regulares. A continuación se programa otro seminario para proporcionar información sobre análisis de datos, redacción de informes y utilización de resultados. Este volumen será de interés para todos los grupos específicos y especialmente para los administradores de establecimientos de salud e investigadores que deseen realizar proyectos de investigación sobre sistemas de salud.

Los otros volúmenes en la serie son: volumen 1, centra su atención en la necesidad de promover el uso de ISS como instrumento de gestión. Asimismo, describe las estrategias para promover la ISS entre ejecutivos y gerentes principales; volumen 3, concebido para ayudar a los investigadores con educación universitaria que trabajan en universidades o institutos investigativos que deseen promover y participar en programas multidisciplinarios de ISS; volumen 4, guía para la gestión de un programa de ISS; volumen 5 servirá de ayuda a aquellos cuya responsabilidad primaria sea organizar y dictar cursos de capacitación para los grupos meta pertinentes.

La serie está diseñada para apoyar un programa esencial de investigación sobre salud a nivel nacional. Se exhorta a los usuarios a examinar críticamente los materiales y/o adaptarlos a sus necesidades particulares.

REMERCIEMENTS

Le présent volume, *Élaboration et mise en oeuvre de programmes de recherche sur les systèmes de santé*, a été rédigé à partir des documents de formation élaborés au début des années 80 dans le cadre du Projet pour le renforcement des systèmes de prestation de soins de santé, à la demande du bureau régional d'Afrique de l'OMS.

Ces documents étaient très populaires et sont couramment utilisés en atelier dans l'Ouest et le Centre de l'Afrique, ainsi qu'ailleurs dans le monde, pour former les travailleurs de la santé à l'élaboration et à la mise en oeuvre de propositions de recherche sur des problèmes particuliers. Cependant, il a fallu apporter des modifications à ces documents de formation, car il est apparu que leur contenu pouvait être adapté aux besoins d'information du processus décisionnel aux différents paliers du système de santé, et qu'un éventail plus large de méthodes de recherche devait être présenté. Les modules 1 à 17 compris dans le présent ouvrage sont des versions fortement adaptées ou renouvelées des modules du Projet. En outre, il a été jugé nécessaire de soutenir les participants non seulement pour l'élaboration d'une proposition de recherche, mais également pour le travail sur le terrain, l'analyse des données et la rédaction de rapports. Par conséquent, un ensemble de modules supplémentaires a été élaboré (modules 18 à 32).

L'adaptation et l'étoffement des modules initiaux ont été effectués au bureau régional III de l'OMS à Harare et au Public Health Institute de Kuala Lumpur, en Malaysia, d'abord de façon autonome, puis en collaboration étroite avec le groupe de recherche sur les systèmes de santé de Genève.

À Harare, le projet conjoint sur la recherche sur les systèmes de santé (entreprise conjointe de l'OMS et du Royal Tropical Institute d'Amsterdam, soutenue par le ministère de la coopération pour le développement des Pays-Bas), qui encourage les recherches sur les systèmes de santé dans tous les pays du Sud de l'Afrique, a élaboré un ensemble de modules avec l'aide de dix chercheurs de la région. Le principal groupe de chercheurs qui a élaboré le cours en Malaysia comptait deux scientifiques malaysiens et un statisticien de Sri Lanka. Ces modules ont été employés dans de nombreux ateliers dans des pays du Sud de l'Afrique, en Malaysia et dans d'autres pays et régions de 1988 à 1991, et ils ont été révisés plusieurs fois.

Depuis le début de 1989, lorsque le CRDI et l'administration centrale de l'OMS ont entrepris de soutenir l'élaboration du présent volume, les différents ensembles de modules ont été graduellement regroupés et approfondis. Corlien M. Varkevisser (Zimbabwe), Indra Pathmanathan (Kuala Lumpur) et Ann Brownlee (autrefois du Projet pour le renforcement des systèmes de prestation de soins de santé), qui s'étaient également chargées de la révision définitive, ont rédigé la version finale du texte. Cependant, de nombreuses autres personnes ont apporté leur contribution. Mentionnons d'abord les membres du groupe de travail technique, qui ont fourni des conseils précieux et partagé leur expérience en formation à la recherche sur les systèmes de santé dans différentes régions du monde, et tous ceux qui ont participé à l'élaboration des modules : L. Omondi (Botswana), M.E. Sebatane et T.K. Makatjane (Lesotho), P. Chimimba et L. Msukwa (Malawi), Maimunah Abdul Hamid et K. Mariappan (Malaysia), A. Kitua et E. Savy (Seychelles), C. Sivagnanasundram (Sri Lanka), G. Tembo (Zambie), R. Munochiveyi, P. Taylor et G. Woeik (Zimbabwe), R. Peeters (Belgique) et, enfin, M.W. Borgdorff et L. Bijmakers (bureau régional III de l'OMS, Harare), qui ont beaucoup contribué à la rédaction ou à la révision des différentes versions. Richard Hayes, Betty Kirkwood et Tom Marshall ont eu la gentillesse d'autoriser la publication de documents employés dans le cours sur la santé communautaire dans les pays en voie de développement donné dans le cadre du programme de maîtrise ès sciences de la London School of Hygiene and Tropical Medicine (modules 28 et 29).

Les animateurs des différents cours ainsi que les personnes qui y ont participé ont fait un apport important à l'élaboration des modules qui en sont issus. Enfin, une équipe dévouée a consacré de nombreuses heures au traitement de texte, à la révision, à la disposition, à la correction d'épreuves et à une foule de tâches qui ont permis la publication du présent volume.

À tous, nos remerciements.

TABLE DES MATIÈRES

Avant-propos	xi
Introduction générale	xv
Introduction au présent volume	xvii
Module 21: Préparation à l'atelier sur l'analyse de données et la rédaction de rapports	1
Module 22: Description des variables, première partie	15
Module 23: Recoupement des données quantitatives	33
Module 24: Analyse des données qualitatives	47
Module 25: Description des variables, deuxième partie	63
Module 26: Tests d'hypothèses	77
Module 27: Différences entre des groupes Partie I — Analyse d'observations non appariées	91
Module 28: Différences entre des groupes Partie II — Analyse d'observations appariées	109
Module 29: Associations entre variables Régression et corrélation	119
Module 30: Mesure du risque dans les études cas-témoins	131
Module 31: Rédaction de rapports	143
Module 32: Utilisation des constatations	159

AVANT-PROPOS

L'objectif ultime de tout processus national d'élaboration d'un système de santé consiste à donner à la population un niveau de santé qui lui permet au moins de participer activement à la vie sociale et économique de la collectivité. Pour atteindre cet objectif, les systèmes de santé actuels doivent être réorientés en vue de permettre l'affectation équitable des ressources destinées à la santé : protection totale, meilleure accessibilité aux soins de santé primaires et renvoi efficace aux soins secondaires et tertiaires. Il est également indiqué d'élaborer les mécanismes nécessaires pour favoriser la participation de la collectivité à la promotion et au maintien de la santé.

Cette réorientation des systèmes de santé pourrait nécessiter des changements à la planification des soins de santé et aux politiques gouvernementales, à l'organisation et à l'administration des services de santé et des services connexes, au financement des systèmes et des procédés ainsi qu'à la sélection et à l'application des techniques appropriées.

Pour mettre en oeuvre les changements nécessaires, les pays doivent établir les meilleures approches à adopter. Il leur faut pour cela des renseignements détaillés et exacts sur les besoins, les possibilités et les conséquences des différentes mesures recommandées. Ces renseignements se révèlent souvent insuffisants, inadéquats ou peu fiables. Par conséquent, les décisions prises sont fondées sur des hypothèses et des conclusions injustifiées, et résultent souvent en des choix mal avisés, dont les conséquences ne sont visibles qu'après la mise en oeuvre.

La recherche constitue une quête systématique d'information et de connaissances nouvelles. Elle a deux fonctions essentielles permettant d'accélérer les progrès dans le domaine de la santé. En premier lieu, la recherche fondamentale ou traditionnelle est nécessaire pour acquérir de nouvelles connaissances et technologies permettant de faire face à d'importants problèmes de santé. En second lieu, la recherche appliquée est nécessaire à l'identification des problèmes prioritaires ainsi qu'à la conception et à l'évaluation des politiques et des programmes qui rapporteront les meilleurs avantages sur le plan de la santé, au moyen des connaissances et des ressources financières et humaines dont on dispose.

Ces deux fonctions, qui constituent ce qu'on appelle la recherche essentielle sur la santé nationale, doivent servir de catalyseurs pour l'acquisition de nouvelles connaissances et l'application des connaissances actuelles, éléments essentiels du développement équitable.

Au cours des dix dernières années, les concepts et les approches de recherche s'appliquant au développement de la santé ont évolué rapidement. Bon nombre d'entre eux sont désignés par des termes spécifiques, tels que recherche opérationnelle, recherche sur les services de santé, recherche sur le personnel de santé, analyse stratégique et économique, recherche appliquée et recherche liée aux décisions. Chacun de ces domaines a fait un apport crucial à l'élaboration de la recherche sur les systèmes de santé, mais leur approche limitée et très ciblée à l'égard de la résolution des problèmes a entraîné leur intégration au sein de la recherche sur les systèmes de santé (RSS) tout en soulignant leur contribution unique à la santé dans les pays en développement.

La RSS consiste au bout du compte à améliorer la «santé» d'une collectivité, quel que soit le concept qu'on désigne par ce terme, en améliorant l'efficacité et l'efficience du système de santé en tant que partie intégrante du processus général de développement socio-économique.

L'objectif de la RSS consiste à fournir aux gestionnaires de la santé de tous les paliers les renseignements pertinents dont ils ont besoin pour résoudre les problèmes auxquels ils font face. La nature participative de cette recherche constitue l'une de ses principales caractéristiques. On prétend que la participation de toutes les parties (la collectivité, les gestionnaires et décideurs des soins de santé et les chercheurs) à

toutes les parties (la collectivité, les gestionnaires et décideurs des soins de santé et les chercheurs) à la définition du problème permet de préciser l'enquête et d'enrichir les données recueillies. De même, la participation à toutes les étapes de la recherche est essentielle pour mettre en oeuvre des solutions pratiques et acceptables à l'échelon communautaire, local, régional ou national.

Comme la RSS aborde les problèmes de santé dans le contexte général du développement social, économique et communautaire, l'apport de différentes disciplines est nécessaire. Parmi ces disciplines, on relève la démographie, l'épidémiologie, l'économie sanitaire, les sciences appliquées au domaine politique, les sciences de la gestion, les sciences sociales et du comportement, la statistique et certains aspects des sciences cliniques. Ces disciplines ont élaboré des approches de recherche spécialisées dans le but de recueillir des renseignements permettant de soutenir le développement de la santé, mais il devient de plus en plus évident que les problèmes que soulève la RSS nécessitent l'apport combiné de nombreuses disciplines et, surtout, que les chercheurs travaillant dans ces domaines spécialisés devront acquérir les compétences nécessaires pour faire partie d'équipes multidisciplinaires.

Voici les principales caractéristiques de la RSS :

- Elle est orientée vers les problèmes prioritaires en matière de santé;
- Elle est de nature participative;
- Elle est orientée vers des mesures concrètes;
- Elle comporte une approche intégrée et multidisciplinaire;
- Elle est de nature multisectorielle;
- Elle met l'accent sur l'efficacité;
- Elle vise à trouver des solutions pratiques en temps voulu;
- Sa nature itérative permet l'évaluation de l'incidence des changements prévus et la révision des plans d'action et des politiques de santé en conséquence.

Bien que ces méthodes puissent être appliquées à des problèmes semblables dans des pays différents, les constatations et les solutions à ces problèmes de santé varieront sans doute en raison de différences culturelles, sociales, économiques et politiques. C'est pour cette raison qu'il est préférable de constituer des groupes nationaux de personnes formées à la RSS dont l'orientation et le plan de travail découlent du programme national de recherche essentielle en matière de santé.

À mesure que la RSS se développe, ses applications se révèlent de plus en plus reconnues. Par conséquent, on l'intègre et on l'applique dans des secteurs particuliers de la gestion tels que l'assurance de qualité, l'évaluation technologique et la gestion des ressources.

Comme les capacités de recherche sur les systèmes de santé sont faibles, surtout dans les pays en développement, il n'est pas étonnant qu'une série de programmes de formation aient été mis sur pied ou financés au cours des dernières années par bon nombre d'organismes, notamment le Centre de recherches pour le développement international, l'Organisation panaméricaine de la santé, l'Organisation mondiale de la santé (OMS) et l'US Agency for International Development (USAID).

En outre, plusieurs programmes internationaux de santé accordent une forte priorité à l'augmentation de la capacité de recherche sur les systèmes de santé.

- Le *Programme spécial pour le développement des capacités nationales en matière de survie et de développement de l'enfant* de l'UNICEF vise à renforcer les connaissances et les compétences en matière de recherche opérationnelle au moyen de l'approche des systèmes de santé dans le but de favoriser la curiosité et des méthodes autonomes d'identification des problèmes communautaires et de recherche de solutions pratiques.

- L'objectif général du *Network of Community Oriented Educational Institutions for Health Sciences* [Réseau d'institutions éducatives communautaires en sciences de la santé] consiste à améliorer la pertinence de la formation des professionnels de la santé en permettant aux diplômés de mieux déceler et résoudre les problèmes de leur collectivité dans le cadre d'un nouveau régime de collaboration entre les universités, les gouvernements et les collectivités orienté autour d'un programme national de recherche essentielle sur la santé.
- L'*International Health Policy Program* (IHPP) [Programme international sur les politiques de santé] compte élaborer des recherches sur les politiques en matière de santé et créer des centres de formation dont le rôle sera de favoriser et de coordonner la synthèse des recherches portant sur les politiques, la diffusion de ces recherches, l'accroissement des capacités d'analyse des politiques de santé et de l'aide technique pour l'analyse des politiques et les recherches menées à ce sujet.
- L'*International Clinical Epidemiology Network* (INCLIN) [Réseau international d'épidémiologie clinique] soutient la création d'unités d'épidémiologie clinique dans les écoles de médecine des pays en développement. Le rôle de ces unités consiste à orienter l'application de principes de mesure quantitative (tirés de l'épidémiologie clinique, de la biostatistique, de l'économie sanitaire et des sciences sociales de la santé) dans les écoles de médecine sur le plan des recherches, de la formation et des services.
- L'Organisme danois de développement international (DANIDA) soutient une série d'ateliers de formation interrégionaux destinés aux directeurs de RSS et, depuis 1987, le projet conjoint de l'Organisation mondiale de la santé, du ministère de la coopération pour le développement des Pays-Bas et le Royal Tropical Institute s'emploie à développer la capacité de RSS de 14 pays du Sud de l'Afrique.

Ces initiatives d'amélioration des capacités de recherche appliquée ont reçu en 1990 un appui politique, moral et intellectuel de la Commission sur la recherche en santé au service des pays en développement, qui recommande, dans son *Plan d'action*,

Que le développement et le maintien de la capacité de recherche constitue un objectif essentiel de tous les investissements en matière de santé et de développement. Il revient aux gouvernements des pays en développement d'accorder la priorité à cette question et de fournir du soutien financier. Des incitatifs internationaux sont également nécessaires. Des échanges internationaux peuvent contribuer à renforcer la capacité de recherche des chercheurs et des établissements des pays en développement.

Dans le contexte plus large des recommandations de la Commission, le développement de la RSS repose sur trois tâches cruciales :

- Stimuler la demande de RSS;
- Renforcer la capacité des pays de mener ces recherches;
- S'assurer que cette capacité puisse être maintenue.

La *Série sur la formation à la recherche sur les systèmes de santé* est fondée sur ces objectifs.

Annette Stark, directrice
Systèmes de santé (recherche)
Division des sciences de la santé
Centre de recherches
pour le développement international

Yvo Nuyens, chef de programme
Programme de recherche et de
développement sur les systèmes de santé
Organisation mondiale de la santé

INTRODUCTION GÉNÉRALE

Un examen récent des ateliers de recherche sur les systèmes de santé (RSS) parrainés par le CRDI a permis de conclure que les objectifs du CRDI ont été atteints, mais que les documents de formation devaient être révisés et étoffés pour répondre aux besoins de groupes précis et pour orienter l'organisation de séances de suivi. En outre, le groupe consultatif mondial de l'OMS sur la RSS a décidé que la création et le maintien de capacités nationales de recherche sur les systèmes de santé constituent une question importante qui doit être soulevée dans les programmes. Il a notamment recommandé que les programmes prévoient l'évaluation et la révision périodiques des documents de formation et le soutien de programmes de formation à différents paliers des systèmes de santé.

À la suite de ces recommandations, des représentants du CRDI, de l'Organisation panaméricaine de la santé et de l'OMS se sont réunis à Ottawa en octobre 1988 pour évaluer les initiatives passées et actuelles et proposer des activités futures. Ces représentants ont reconnu que pour que la formation à la RSS puisse servir à améliorer la santé et les soins de santé, il est nécessaire d'éclaircir le contexte et les étapes du développement d'un processus de recherche efficace dans un pays donné. Ils ont également convenu que des groupes cibles devaient être choisis pour l'orientation et la formation à la RSS et que des stratégies de formation devaient être élaborées pour renforcer la capacité de recherche des pays selon leurs besoins et leur capacité actuelle.

Pour atteindre cet objectif, un groupe de travail technique a été mis sur pied et chargé de définir et de coordonner l'élaboration d'un ensemble de documents de formation de base à l'intention de cinq groupes cibles. Son travail a été orienté par :

- Une définition du groupe cible;
- Une description de la compétence et des caractéristiques de base du groupe cible;
- Le comportement attendu, notamment en ce qui concerne les compétences et l'attitude;
- Les stratégies et le contexte de formation appropriés;
- Les documents de formation disponibles.

Les travaux du groupe de travail technique ont mené à la rédaction de cinq volumes. Les utilisateurs sont invités à se familiariser avec toute la série et à mettre en oeuvre un programme de formation, de recherche et de planification et des politiques en matière de soins de santé fondés sur les besoins de leur pays.

Volume 1 : La recherche sur les systèmes de santé : un outil de gestion

À l'intention des décideurs

Ce document porte sur la nécessité de promouvoir l'utilisation de la RSS comme outil de gestion pour les décideurs. En se fondant sur une analyse des activités des pays en développement au cours des dix dernières années, il démontre comment les recherches sur les systèmes de santé peuvent améliorer le processus décisionnel et explique les étapes de la création d'un programme national de recherche efficace. En outre, il décrit des stratégies précises de promotion de la RSS auprès des décideurs et des gestionnaires supérieurs qui ont fait leurs preuves dans différentes circonstances.

Volume 2 : *Élaboration et mise en oeuvre de programmes de recherche sur les systèmes de santé*

Première partie — Formulation et mise à l'essai d'une proposition

Deuxième partie — Analyse de données et rédaction de rapports

Les participants à ce cours, qui peuvent être de simples citoyens, des travailleurs de la santé, des chercheurs et des décideurs du domaine de la santé du palier provincial ou même national, sont appelés à choisir des problèmes de santé prioritaires qui les concernent et qui ne peuvent être réglés à moins de recueillir plus de renseignements à leur sujet. De préférence, ces problèmes auront été choisis avant le début de la formation (voir le volume 1), mais ils pourront être détaillés par la suite. Dans la plupart des cas, une équipe de participants effectue la recherche pendant l'exercice normal de leurs fonctions (première partie). Un deuxième atelier est alors tenu pour fournir aux participants des renseignements sur l'analyse de données, la rédaction de rapports et l'utilisation des résultats (deuxième partie).

Ce volume traite particulièrement de l'élaboration de propositions de recherche de nature participative (collectivité, gestionnaires des soins de santé, chercheurs), de la tenue de la recherche sur le terrain, puis de l'analyse et de la diffusion des résultats. Dans ce contexte, il peut également intéresser les chercheurs débutants et les personnes qui veulent opérationnaliser la RSS dans les universités et d'autres établissements de formation.

Volume 3 : *Stratégies permettant d'intéresser les universités et les instituts de recherche à la recherche sur les systèmes de santé*

À l'intention des chercheurs d'expérience et des universitaires

Ce volume est conçu pour aider les chercheurs universitaires travaillant dans des universités ou instituts de recherche qui veulent promouvoir les programmes multidisciplinaires de recherche sur les systèmes de santé et y participer. Ce volume intéressera particulièrement les personnes qui veulent intégrer les notions de RSS dans les programmes actuels de santé et de sciences sociales et favoriser la rédaction de thèses dans ce domaine.

Volume 4 : *Gestion de la recherche sur les systèmes de santé*

À l'intention des directeurs de recherche

Ce volume est destiné aux directeurs d'instituts de recherche, de départements universitaires et d'organismes qui sont chargés de traiter les demandes de recherche ainsi que de financer et de coordonner des programmes de recherche. Il devrait permettre à ces gestionnaires d'améliorer l'apport de leur organisme au développement de la RSS dans leur pays et l'utilisation de ces recherches en vue d'améliorer la santé de la population.

Volume 5 : *La recherche sur les systèmes de santé : la formation des formateurs*

À l'intention des formateurs et des animateurs

Les chercheurs d'expérience ne sont pas tous des enseignants chevronnés. Peu d'entre eux ont de l'expérience dans l'organisation et la formation de participants pour qui la recherche constitue une activité secondaire et qui ont peu de temps pour lire ou mener des recherches.

Pour que la formation à la recherche sur les systèmes de santé soit efficace, les chercheurs d'expérience doivent se familiariser avec des méthodes élaborées et utilisées avec succès au cours des dernières décennies pour former le personnel de la santé à différents sujets importants reliés à la santé.

Ce volume est destiné aux formateurs et aux animateurs qui sont chargés d'organiser et de donner des cours de formation à l'intention des différents groupes cibles ainsi qu'à leurs assistants.

INTRODUCTION AU PRÉSENT VOLUME

La présente publication est conçue pour être utilisée avec la première partie, intitulée Formulation et mise à l'essai d'une proposition. Ce document se compose de 20 modules de formation qui orientent les participants au cours de l'élaboration d'une proposition de recherche et leur donnent des directives utiles pour la tenue de la recherche¹.

La deuxième partie, Analyse de données et rédaction de rapports, se compose de 12 modules. Ces modules de formation, comme ceux qui portent sur l'élaboration d'une proposition de recherche, peuvent être adaptés :

- au niveau de scolarité et d'expérience en recherche des participants;
- au type d'étude et aux techniques de collecte de données utilisées;
- à l'état des données au début de l'atelier.

Si les participants ont déjà reçu de la formation sur les méthodes de recherche et la statistique ou s'ils ont un peu d'expérience en recherche, les modules pourront être exposés rapidement. Dans ce cas, les exposés visent surtout à rafraîchir la mémoire des participants et à leur montrer comment utiliser correctement les analyses et tests appropriés. Certains modules peuvent être combinés ou abrégés.

Si les participants n'ont ni formation ni expérience en recherche, les exposés pourraient devoir être limités à l'essentiel nécessaire pour traiter les données recueillies. Dans ce cas, les exposés seront plus longs et devraient prévoir des périodes de questions et des exercices en classe.

Les éléments essentiels du présent volume sont les suivants :

- Le module 21 (préparation à l'atelier).
- Les modules 22, 23 et 25 (description des variables et recoupements).
- Le module 24 (analyse des données qualitatives), particulièrement les deux premières parties.
- Le module 26 (tests d'hypothèses). Insister sur les sections I, II et III et expliquer la raison d'être des tests d'hypothèses et leur fonctionnement, mais n'aborder que brièvement la section IV, qui porte sur la sélection d'un test d'hypothèses, s'il est peu probable que les participants aient besoin d'un autre test que le test t ou le test du khi carré.
- Le module 27 (différences entre des groupes : analyse d'observations non appariées). Traiter du test t ou du test du khi carré, ou des deux.
- Module 30 (mesure du risque : incidence, risque, risque relatif et risque relatif approché). Insister sur les observations non appariées.

¹ Le lecteur est invité à lire les pages d'introduction de la première partie du volume 2 pour des renseignements sur les quatre autres volumes de la Série sur la formation à la recherche sur les systèmes de santé, le but de cette série, les remerciements et une introduction sur l'utilisation de la première et de la deuxième partie du volume II, Élaboration et mise en oeuvre de programmes de recherche.

- Les modules 31 et 32 (rédaction de rapports et utilisation des constatations).

Selon le type d'études que les participants ont mené et les analyses auxquelles leurs données doivent être soumises, l'envergure des exposés peut être élargie (en détaillant les tests statistiques, par exemple, ou l'analyse des données qualitatives) ou leur ordre modifié (le module 24 peut être présenté avant le module 23 si les participants disposent surtout de données qualitatives).

- La première moitié de l'atelier (une semaine) est normalement consacrée à l'achèvement du traitement et de l'analyse des données. Tous les modules reliés à l'analyse (21 à 30) sont présentés au cours de cette semaine.

Les exposés doivent avoir lieu au moment opportun. Les modules 21 à 24 peuvent être présentés avant la fin du traitement des données. Le module 25, sur la description des variables (mesures de dispersion, écart-type, erreur-type), ne devrait être vu qu'une fois le traitement terminé, car ce module constitue une introduction aux tests statistiques.

Les modules qui portent sur les différents tests statistiques ne devraient être présentés qu'après le traitement des données, une fois que les participants auront entrepris la préparation des tableaux croisés.

- La seconde moitié de l'atelier porte sur la rédaction de rapports, la formulation des recommandations ainsi que la présentation et la discussion en plénière des grandes constatations et recommandations de l'étude. Au cours de la semaine, il n'y a normalement que deux exposés : un sur la rédaction de rapports (module 31) et l'autre sur l'utilisation des constatations (module 32). Il est préférable de présenter le dernier module juste avant que les participants rédigent le sommaire des constatations et recommandations.

Un exemple de calendrier d'un cours de deux semaines sur l'analyse de données et la rédaction de rapports figure aux pages suivantes.

Si les participants sont de niveau élevé et si les données ont été convenablement traitées avant l'atelier d'analyse de données et de rédaction de rapports, il est possible de ramener la durée de l'atelier à une semaine ou 10 jours.

**EXEMPLE DE CALENDRIER DE COURS
(utilisé dans le Sud de l'Afrique)**

**Élaboration et mise en oeuvre de programmes de recherche
sur les systèmes de santé :
Analyse de données et rédaction de rapports**

Jour/heure	Séance	Responsable
<i>Lundi</i>		
0800 - 0830	Ouverture de la séance	Coordonnateur
0830 - 0915	Présentation des résultats préliminaires et discussion	Groupe 1
0915 - 1000	Présentation des résultats préliminaires et discussion	Groupe 2
1000 - 1030	Thé	
1030 - 1115	Présentation des résultats préliminaires et discussion	Groupe 3
1115 - 1200	Présentation des résultats préliminaires et discussion	Groupe 4
1200 - 1230	Module 21 : Préparation à l'atelier sur l'analyse de données et la rédaction de rapports	Animateur
1230 - 1400	Déjeuner	
1400 - 1530	Travail en groupe	
1530 - 1600	Thé	
1600 - 1700	Module 22 : Description des variables - première partie	Animateur
1700 - 1800	Travail en groupe	
<i>Mardi</i>		
0800 - 0900	Module 23 : Recoupement des données quantitatives	Animateur
0900 - 1300	Travail en groupe (et thé)	
1300 - 1400	Déjeuner	
1400 - 1500	Module 24 : Analyse des données qualitatives	Animateur
1500 - 1800	Travail en groupe (et thé)	
<i>Mercredi</i>		
0800 - 1300	Travail en groupe (et thé)	
1300 - 1400	Déjeuner	
1400 - 1500	Facultatif : Exposé des principaux résultats du travail en groupe : objectifs révisés, principaux tableaux croisés, résultats de l'analyse qualitative	Tous les groupes

1500 - 1600	Module 25 : Description des variables - deuxième partie	Animateur
1600 - 1800	Travail en groupe (et thé)	

Jeudi

0800 - 0900	Module 26 : Tests d'hypothèses (20 min.), suivi du	Animateur
0900 - 1300	Module 27, sections I et II (test t) Travail en groupe (et thé)	
1300 - 1400	Déjeuner	
1400 - 1500	Module 27, section III (test du khi carré)	Animateur
1500 - 1800	Travail en groupe (et thé)	

Vendredi

0800 - 0900	Module 28, 29 ou 30 (selon les besoins des groupes)	Animateur
0900 - 1300	Travail en groupe (et thé)	
1300 - 1400	Déjeuner	
1400 - 1500	Module 28, 29 ou 30 (selon les besoins des groupes)	Animateur
1500 - 1800	Travail en groupe (et thé)	

Samedi

0800 - 1300	Travail en groupe (et thé)	
-------------	----------------------------	--

Lundi

0800 - 0900	Module 31 : Rédaction de rapports	Animateur
Reste de la journée	Travail en groupe	

Mardi

Toute la journée	Travail en groupe	
------------------	-------------------	--

Mercredi

0800 - 0830	Module 32 : Utilisation des constatations	Animateur
Reste de la journée	Travail en groupe	

Jeudi

0800 - 1300	Travail en groupe	
1300 - 1400	Déjeuner	
1400 - 1445	Présentation des résultats, recommandations et plan d'action	Groupe 1
1445 - 1530	Présentation des résultats, recommandations et plan d'action	Groupe 2

1530 - 1600	Thé	
1600 - 1645	Présentation des résultats, recommandations et plan d'action	Groupe 3

1645 - 1730	Présentation des résultats, recommandations et plan d'action	Groupe 4
-------------	---	----------

1730 - 1800	Cérémonie de clôture	
-------------	----------------------	--

Vendredi

0800 - 1300	Travail en groupe (achèvement des rapports)	
-------------	---	--

1300 - 1400	Déjeuner	
-------------	----------	--

1400 - 1445	Évaluation du cours de formation	
-------------	----------------------------------	--

1445 - 1730	Travail en groupe (achèvement des rapports)	
-------------	---	--

Samedi (si nécessaire)

0800 - 1300	Achèvement des rapports	
-------------	-------------------------	--

**Formation à la recherche sur les systèmes de santé
Volume 2, 2^o partie : Analyse de données et rédaction de rapports**

Module 21

**PRÉPARATION À L'ATELIER SUR L'ANALYSE
DE DONNÉES ET LA RÉDACTION DE RAPPORTS**

Module 21 : PRÉPARATION À L'ATELIER SUR L'ANALYSE DE DONNÉES ET LA RÉDACTION DE RAPPORTS

OBJECTIFS DE L'ATELIER

À la fin du présent atelier, vous devriez pouvoir :

1. **Identifier** et définir les notions et procédés fondamentaux de l'analyse et de l'interprétation des données;
2. **Analyser** et interpréter les données recueillies pour le projet de recherche que vous avez élaboré au cours du premier atelier et tirer des conclusions fondées sur les objectifs de votre étude;
3. **Rédiger** un rapport de recherche clair et concis ainsi qu'un sommaire des grandes constatations et recommandations à l'intention de tous les intervenants qui s'y intéressent;
4. **Présenter** les grandes constatations et recommandations de votre étude aux décideurs, aux gestionnaires de la santé et aux sujets de votre recherche; collaborer avec ces personnes pour mettre au point les recommandations;
5. **Préparer** un plan de diffusion des constatations et des recommandations (s'il y a lieu) à des fins de recherches supplémentaires.

I. Revue des activités sur le terrain

II. Introduction au présent atelier

III. Tâches à remplir pendant l'atelier

Examen et fin du traitement des données

Analyse des données

Rédaction du rapport

Présentation du sommaire des constatations et recommandations

Élaboration d'un plan de mise en oeuvre des résultats de la recherche

I. REVUE DES ACTIVITÉS SUR LE TERRAIN

La mise en oeuvre de votre proposition de projet de recherche a dû représenter pour vous une expérience stimulante. Vous vous êtes sans doute heurté à des obstacles imprévus au cours de vos travaux sur le terrain, mais vous avez également quelques réussites à votre actif. Si tout s'est passé comme prévu, vous avez recueilli vos données, vous en avez traité une grande partie, sinon la totalité, vous avez entrepris l'analyse et vous avez rédigé un rapport préliminaire sur vos travaux sur le terrain et leurs résultats. L'expérience pratique que vous acquerrez dans l'exécution du projet de recherche est très importante. Il vous sera utile de la partager avec les autres participants à l'atelier, car vous pouvez tirer profit de vos erreurs aussi bien que de vos réussites.

Avant de présenter un aperçu du programme de cet atelier, nous aimerions que chacun des groupes de recherche nous fasse part de ses activités.

EXERCICE : Exposé des activités sur le terrain

Présentez le rapport préliminaire que votre groupe a préparé à la fin de vos travaux sur le terrain, en suivant les directives figurant au module 20. Répondez aux questions des autres participants ou des animateurs après votre exposé.

Chaque groupe disposera d'environ 20 minutes pour son exposé, suivi d'une période de questions de 20 à 30 minutes.

II. INTRODUCTION AU PRÉSENT ATELIER

Le présent atelier fait suite à celui où vous avez élaboré votre proposition. Vous avez maintenant à vous attaquer à la difficile tâche d'analyser complètement les données que vous avez recueillies sur le terrain et de rédiger votre rapport de recherche. Ce rapport devrait contenir des recommandations pratiques et utiles fondées sur vos constatations concernant la manière de résoudre le problème étudié.

Comme lors du premier atelier, nous procéderons à des exposés, à des séances de travail en groupe et à quelques plénières, mais nous nous concentrerons cette fois-ci sur le travail en groupe. Les exposés auront lieu au cours de la première semaine, qui sera consacrée à l'analyse des données. La deuxième semaine sera entièrement réservée à la rédaction du rapport, et il n'y aura que deux exposés pour vous orienter. Vers la fin de la semaine, chaque groupe présentera un sommaire de ses constatations et de ses recommandations lors d'une importante séance plénière. Un groupe restreint de décideurs et de gestionnaires qui ont commandé l'étude ou qui s'y intéressent de près seront invités à formuler des commentaires sur votre exposé au cours de cette séance.

Les modules du présent atelier comprennent plusieurs tâches importantes, qui sont illustrées dans le diagramme de la page suivante. Ce diagramme revient au début de chaque module suivant pour indiquer la tâche qui fait l'objet de l'exposé. Nous aborderons brièvement chacune de ces tâches au cours de la présente séance.

Figure 21.1. Étapes de l'analyse de données et de la rédaction de rapports.

Questions à poser	Étapes à suivre*	Éléments importants de l'étape
Quelles données ont été recueillies pour chaque objectif de recherche? Les données sont-elles complètes et exactes?	Préparer les données pour l'analyse	Revoir les activités sur le terrain Compiler les données pour chaque objectif ou population étudiée Classer les données et en vérifier la qualité Vérifier les imprimés
À quoi ressemblent les données?	Décrire les variables	Distributions Figures, moyennes
Comment peut-on résumer les données pour en faciliter l'analyse?	Faire le recoupement des données quantitatives Résumer les données qualitatives	Faire le recoupement des données par rapport aux objectifs Illustrations, texte
Pour les données qualitatives : la recherche a-t-elle pour but de décrire, de comparer ou de trouver des associations?	Déterminer le type d'analyse statistique nécessaire	Examiner les objectifs, le type d'étude et les variables Faire une description statistique des variables Choisir les tests d'hypothèses
1. Comment peut-on décrire les données?	Analyser des observations appariées et non appariées	**Test t de Student **Test t pour observations appariées **Test du khi carré **Test du khi carré de McNemar
2. Comment peut-on déterminer les différences entre les groupes?		
3. Comment peut-on déterminer les associations entre les variables?	Évaluer les associations entre les variables	**Diagramme de dispersion **Droite de régression et coefficient de corrélation **Risque relatif, risque relatif approché
Comment le rapport doit être rédigé?	Rédiger le rapport et formuler des recommandations	Préparer le canevas du rapport Rédiger les 1 ^{re} et 2 ^e versions Faire un sommaire des constatations Faire un sommaire des conclusions pour chaque objectif Formuler des recommandations Préparer un résumé analytique
Comment doit-on présenter et diffuser les constatations et les recommandations?	Présenter un sommaire et rédiger un plan de mise en oeuvre des recommandations	Discuter des sommaires avec différents groupes cibles Discuter du programme de mise en oeuvre

* Il n'est pas nécessaire de faire ces étapes dans l'ordre indiqué. L'ordre peut être adapté aux besoins de l'atelier.

** Ces éléments sont facultatifs; ils peuvent être omis s'ils ne sont pas pertinents au cours.

III. TÂCHES À REMPLIR PENDANT L'ATELIER

Examen et fin du traitement des données

Nous supposons que tous les groupes ont consacré beaucoup d'efforts au traitement de leurs données, mais il est possible que des redressements et de l'étoffement soient nécessaires. C'est là une chose normale en recherche; un pas en avant peut être suivi par un demi-pas en arrière. Même à une étape avancée de l'analyse, vous pourriez devoir regrouper et retraiter une partie des données.

Avant l'analyse, il est très important de vérifier si le traitement des données a été effectué de façon que l'information :

- est facile à manipuler;
- a été vérifiée afin d'un déceler les erreurs qui auraient pu s'y glisser pendant la collecte.

Vous devez donc vous poser les questions suivantes :

- **Les données ont-elles été classées correctement?** La numérotation des questionnaires et des listes de vérification est-elle pratique? Est-il possible de distinguer clairement les grandes catégories de répondants pour faciliter la comparaison portant sur des variables pertinentes, comme l'exigent vos objectifs de recherche?
- **L'exhaustivité et la cohérence des données ont-elles été vérifiées?** Consultez le **module 13** pour des mesures à prendre si les données sont incomplètes ou incohérentes.
- **Toutes les fiches de données ont-elles été remplies** avec vos données qualitatives (ou, si vous utilisez un ordinateur, toutes les données ont-elles été saisies)? Le nombre total de réponses correspond-il au nombre de répondants pour chaque variable?
- **Les données qualitatives ont-elles été divisées en catégories dans toute la mesure du possible?** S'il y a lieu, le codage a-t-il été effectué? (Voir le **module 13** pour diviser en catégories des questions ouvertes.) Soulignons que pour les données qualitatives, la collecte, le classement, la rédaction du sommaire et l'analyse sont, en principe, reliés entre eux. Nous le verrons en détail au **module 24**.
- Si vous avez utilisé un ordinateur pour traiter les données, **vérifiez la fréquence de chaque variable** du questionnaire. Vérifiez également les recoupements effectués par l'ordinateur. La marche à suivre figure à l'**annexe 21.1**.

Avant de revoir les procédures de traitement des données, il est fortement recommandé de faire un **INVENTAIRE** de toutes les données disponibles pour chaque **OBJECTIF**, surtout si les données nécessaires ont été recueillies au moyen de différents outils de collecte.

Exemple

Sources de données pour l'objectif 3, «Détection des faiblesses dans le fonctionnement des services de santé maternelle et infantile permettant d'expliquer la faible consommation des soins relatifs à l'accouchement.»

- Questionnaire destiné aux mères, questions 12, 15 à 19 et 23
- Discussion de groupe avec le personnel de santé, sujets 3 et 4
- Observations figurant dans les listes de vérification

Cet inventaire vous aidera à mieux organiser l'analyse des données et, plus tard, la rédaction de rapports.

Analyse des données

Avant de commencer l'analyse des données, il faut déterminer parmi nos données lesquelles sont quantitatives et lesquelles sont qualitatives.

Données quantitatives

Les données quantitatives sont exprimées en chiffres et sont habituellement présentées dans des tableaux de fréquence. À partir de vos fiches de données, vous pouvez facilement calculer le total pour chaque variable ou question, compter le nombre de réponses différentes obtenues et présenter l'information dans un tableau de fréquence (**module 22**).

Lors de l'analyse des données quantitatives, il est important de remettre en cause le but de l'étude. Consiste-t-elle à :

- **décrire des variables?**

Exemple : distribution de la grossesse chez les adolescentes dans une population déterminée.

- **établir des différences entre des groupes?**

Exemple : différences entre les personnes habitant depuis longtemps dans une certaine région et les nouveaux arrivants sur le plan du revenu ou de l'état de santé.

- **déterminer des associations entre variables?**

Exemple : l'association entre la satisfaction des infirmiers et infirmières à l'égard de leur travail et le nombre de réunions du personnel au cours de l'année.

Le tableau croisé constitue un moyen approprié de résumer et d'analyser les données (**module 23**).

Des notions statistiques plus avancées pour l'analyse des données quantitatives figurent aux **modules 25 à 30**. Après que les distributions de fréquences et différents types de recoupements ont été effectués, il faut déterminer le type d'analyse statistique nécessaire. Cette étape nécessite une description plus approfondie des variables (**module 25**) et la sélection de test d'hypothèses appropriés (**module 26**).

Les tests d'hypothèses les plus courants sont :

- Le test t de Student et le test du khi carré pour déterminer les différences entre groupes si les observations sont non appariées (**module 27**);
- Le test t pour observations appariées et le test du khi carré de McNemar pour déterminer des différences entre groupes dans le cas d'observations appariées (**module 28**).

Pour mesurer des associations entre variables, la régression et la corrélation (**module 29**) ainsi que le risque relatif approché (**module 30**) sont présentés.

Tout au long de l'analyse des données, il est important de tenir compte du fait que **nos constatations devraient permettre de répondre à nos questions et ainsi d'atteindre les objectifs de notre**

recherche. Nous voudrions tirer des conclusions et formuler des recommandations en nous fondant sur ces constatations.

Données qualitatives

Si vous vous souvenez, nous pouvons obtenir des données qualitatives par les moyens suivants :

- **par des questions ouvertes**, non divisées au préalable en catégories, dans des questionnaires visant à recueillir d'abord des données quantifiables;
- **par des entrevues peu structurées** comportant des questions ouvertes posées à des répondants importants (particuliers ou groupes);
- **par des discussions de groupe** sur certaines questions, avec une liste de points à discuter pour orienter les participants;
- **par des observations** visant à décrire le comportement du particulier ou du groupe.

Comme nous l'avons vu à l'exercice sur les motifs de l'usage du tabac au **module 13**, les réponses aux questions ouvertes peuvent être :

- **portées à une liste;**
- **divisées en catégories** (selon vos objectifs de recherche et votre bon sens, en combinant les réponses assorties en quatre ou six catégories tout au plus);
- **étiquetées ou codées;**
- **insérées**, au moyen de ces codes, **dans vos fiches de données ou dans l'ordinateur;**
- **comptées** comme toute autre donnée quantitative.

Cependant, si vous vous intéressez également au **CONTENU** de chaque réponse, par exemple parce que vous voulez entreprendre une campagne anti-tabac s'adressant à différentes catégories de fumeurs, vous analyserez en profondeur le contenu des réponses de chaque catégorie.

Remarque

L'analyse des données qualitatives se distingue surtout par le fait qu'elle est **narrative**, et non fondée sur des données numériques.

Il est souvent utile de résumer les données qualitatives dans des diagrammes, des organigrammes ou des matrices afin de faciliter l'analyse. Le **module 24** traite en détail de l'analyse des données qualitatives.

Rédaction du rapport

À la fin du présent atelier, vous devrez avoir achevé votre rapport de recherche. Ce rapport comportera les éléments suivants :

1. Une INTRODUCTION comportant un énoncé du problème, quelques données pertinentes expliquant le contexte et une étude documentaire;
2. Des OBJECTIFS;
3. Une section sur la MÉTHODOLOGIE comportant des renseignements sur la méthode de collecte des données (type d'étude, variables, outils de collecte), le moment et l'endroit où elles ont été recueillies (taille de l'échantillon, procédures d'échantillonnage); la méthode d'analyse des données et les faiblesses possibles dans la collecte et l'analyse;
4. Des CONSTATATIONS et des CONCLUSIONS;
5. Une DISCUSSION;
6. Des RECOMMANDATIONS.

Les trois derniers points seront abordés en détail au **module 31**. Vous pouvez évaluer les trois premiers dans les parties correspondantes de votre proposition de recherche.

Présentation du sommaire des constatations et recommandations

Comme l'un des objectifs de votre recherche consiste à vous assurer que des mesures seront prises en fonction de ses résultats, il est important que toutes les parties concernées aient la possibilité de discuter des constatations et des recommandations avant que le rapport soit achevé. Vous pouvez inviter des décideurs, des gestionnaires de soins de santé, du personnel de santé et des membres de la collectivité à cette discussion. Le **module 32** donne certaines indications sur la façon d'organiser une réunion à cette fin.

Élaboration d'un plan de mise en oeuvre des résultats de la recherche

Vous avez rédigé un plan d'utilisation et de diffusion des résultats au cours du dernier atelier (**module 16**). À la fin du présent atelier, nous l'évaluerons et le détaillerons.

TRAVAIL EN GROUPE (peut être effectué à différentes étapes, selon le sujet de la recherche et l'état d'avancement du traitement des données)

- Revoyez les objectifs de votre proposition de recherche et établissez une liste des différentes sources de données pour chaque objectif (questionnaires, dossiers, discussions en groupe, etc.)

Nota Si vous constatez que vous avez recueilli plus de données pour expliquer le problème étudié que ne le nécessitent vos objectifs, vous pouvez revoir vos objectifs ou en ajouter un ou deux. Cependant, si vous avez recueilli moins de données, ne laissez pas tomber d'objectifs, mais expliquez pourquoi vous n'avez pu les atteindre.

- Vérifier si l'exhaustivité, la cohérence et le codage des données ont été vérifiés. Sinon, faites-le.
- Déterminez si différentes fiches de collecte ont été préparées pour des populations différentes ou pour des catégories différentes de répondants (une pour les personnes qui se conforment et une pour celles qui ne se conforment pas, par exemple). Cela facilitera l'analyse et, s'il y a lieu, la comparaison des groupes.

Vous pouvez également attribuer une couleur à chaque sous-groupe afin de pouvoir revenir facilement aux données de base pour vérifier certaines questions.

- Vérifiez si les fiches de collecte ont été remplies et si le nombre de réponses pour chaque variable correspond au nombre de répondants.

Déterminez si toutes les données à saisir dans l'ordinateur ont bien été saisies et vérifiées (voir l'annexe 21.1).

- Vérifiez si les données qualitatives ont été divisées en catégories et résumées. Prenez note de toutes vos observations pertinentes.

Annexe 21.1. IMPRIMÉ

Ce que l'on imprime au moyen de l'ordinateur constitue le résultat de commandes exécutées par le logiciel afin d'analyser les données disponibles. La précision des informations imprimées repose donc sur :

- les données saisies,
- le logiciel employé.

En informatique, la qualité des résultats dépend de la qualité des données. Il revient donc à l'enquêteur de s'assurer que ces résultats sont exacts.

Types d'imprimés

1. Liste de données

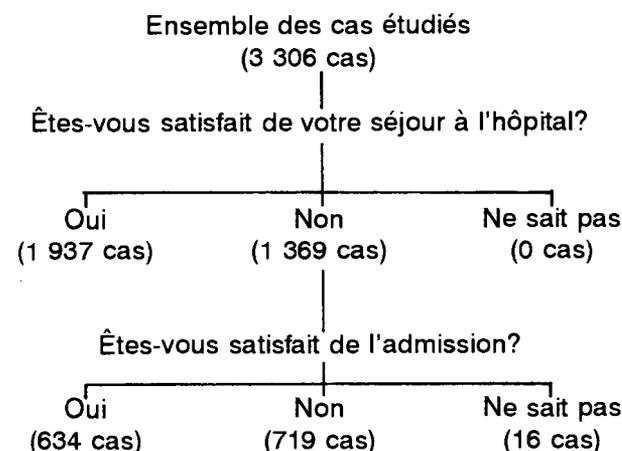
Il s'agit d'une liste des données qui ont été saisies dans l'ordinateur. Cet imprimé est utile pour apporter des corrections aux données existantes tout en les validant.

2. Fréquence des données

Cette opération permet d'obtenir le compte (et le pourcentage) de chaque variable du questionnaire. Étudiez l'exemple suivant et voyez comment il est relié aux questionnaires.

Pour s'assurer que le logiciel fonctionne correctement, l'enquêteur doit se familiariser avec la méthode ou le questionnaire utilisé et avec le processus de collecte des données. Pour vérifier la précision du logiciel, l'enquêteur peut tracer un organigramme du processus comportant des questions appropriées et un compte des variables.

Exemple



Dans cet exemple, le nombre de réponses «oui», «non» et «ne sait pas» à la question «Êtes-vous satisfait de votre séjour à l'hôpital?» doit être égal au nombre total de cas étudiés. Sinon, communiquez avec votre spécialiste en informatique ou votre statisticien.

Une fréquence devrait être obtenue pour chaque question du questionnaire. Utilisez la fréquence pour vous assurer que :

- le nombre total de réponses à chaque question est exact (il correspond à la taille de l'échantillon ou au nombre de personnes à qui la question a été posée);
- tous les codes sont pertinents à la question. **Par exemple**, il ne devrait pas y avoir de code 3-8 dans une question qui ne comporte que deux réponses possibles en plus de «inconnu» (si «inconnu» reçoit le code 9).

3. Tableau croisé

Un autre type d'imprimé courant est le **tableau croisé**. Il montre le nombre de sujets auxquels s'appliquent deux ou plusieurs des variables étudiées.

Exemple :

	Homme	Femme
Malade		
Pas malade		

Avant de l'utiliser, vérifiez les points suivants :

- Le total du tableau doit correspondre au nombre de sujets de l'échantillon;
- Le total des colonnes et des rangées doit correspondre à la fréquence de chaque variable (ainsi, le nombre d'hommes et de femmes doit correspondre à leur fréquence respective);
- De même, les chiffres correspondant à «malade» et «pas malade» doivent correspondre à cette fréquence. Sinon, il y a probablement une erreur dans le logiciel. Consultez votre spécialiste en informatique.
- Si l'imprimé porte la mention «cas manquants»,
 - il y a un mauvais code dans les données (p. ex., un code 4 lorsque seul un code 1, 2 ou 9 serait valable), ou
 - les catégories indiquées ne sont pas complètes.

Exemple

Le questionnaire comportait la réponse «inconnu», mais pas le logiciel. Par conséquent, toutes les réponses «inconnu» constituent des «cas manquants».

Dans le questionnaire, l'état civil pouvait être «marié, célibataire, divorcé ou veuf». Cependant, le logiciel ne reconnaît que «marié, célibataire ou divorcé». Toutes les personnes veuves sont donc manquantes.

Si les catégories d'âge sont 10 à 14 ans et 15 à 19 ans mais que le programmeur ait programmé par mégarde les catégories 10 à 13 ans et 15 à 19 ans, tous les sujets âgés de 14 ans sont manquants.

Notes du formateur

Module 21 : PRÉPARATION À L'ATELIER SUR L'ANALYSE DE DONNÉES ET LA RÉDACTION DE RAPPORTS

Durée et méthode d'enseignement

- | | |
|-----------|--|
| 3 heures | Exposés portant sur les travaux sur le terrain |
| 1/2 heure | Introduction et discussion |
| 1 heure* | Travail en groupe (durée dépendant du sujet de la recherche et de l'état d'avancement du traitement des données) |

Introduction et discussion

- Consacrez la première partie de la séance d'introduction aux rapports des participants sur leur travail sur le terrain. Si tous les groupes sont prêts à présenter leur rapport préliminaire, la séance peut commencer par cette activité. Cependant, si les groupes ont encore besoin d'un peu de temps pour préparer leur exposé, réservez un peu de temps avant la première séance ou au début de celle-ci.
- L'introduction à l'atelier doit clairement souligner qu'il y a différentes tâches à remplir, et parmi elles, l'analyse des données et la rédaction des rapports nécessiteront le plus de temps. Cependant, les participants doivent bien comprendre que la préparation de recommandations et leur mise en oeuvre constituent l'objectif ultime de leur projet de recherche. Vous pouvez leur demander de suggérer des décideurs et gestionnaires à inviter à l'exposé et à la discussion sur leurs constatations et recommandations à la fin de l'atelier.
- Pour la présentation du diagramme, vous pourriez employer des transparents superposés.
- Adaptez l'exposé au niveau et aux intérêts des participants. Rafraîchissez-leur la mémoire en leur donnant des exemples de traitement des questions ouvertes et en expliquant la différence entre les études descriptives, les études comparatives servant à trouver des différences entre des groupes et les études permettant à établir des associations entre des variables, **de préférence avec des exemples tirés de leur propre recherche.**
- Pour l'instant, veillez à ne pas effrayer les groupes qui ont peu d'expérience en statistique en leur donnant des détails sur les tests. Dites seulement que chaque type d'étude nécessite des tests différents.
- Soulignez l'importance de dresser une liste de toutes les données relatives à chaque objectif, y compris les données qualitatives. L'atelier accorde tellement d'importance à la préparation des tableaux que les participants auront tendance à négliger les observations et informations valables obtenues de répondants importants. Le formateur devrait demander aux participants de prendre note de ces renseignements dès maintenant (si ce n'est déjà fait) et de les inclure

dans la liste des données relatives à chaque objectif. Pendant la rédaction du rapport, vérifiez que ces données ont été analysées.

Travail en groupe

- Lisez les directives du travail en groupe avec les membres du groupe. Laissez-les revoir leurs objectifs, dresser une liste des données relatives à chaque objectif et déterminer si les objectifs sont assez précis pour s'appliquer à toutes les données pertinentes qui ont été recueillies. Parfois, les objectifs doivent être séparés, reformulés, étoffés ou intervertis afin de faciliter l'analyse. **Ne laissez jamais** le groupe omettre un objectif sans explication (dans la section sur la méthodologie) sur la raison pour laquelle cet objectif n'a pu être atteint.
- Étudiez avec les membres du groupe toutes les données disponibles afin d'en vérifier l'exhaustivité, de déceler les erreurs, etc. Assurez-vous que des fiches de données distinctes ont été préparées pour différentes populations ou pour différents sous-groupes à comparer, ou que les données sur des sous-groupes différents peuvent être facilement extraites par ordinateur.
- **En tant que formateur, prenez le temps d'assimiler toutes les données disponibles, de déterminer les faiblesses éventuelles et d'envisager différentes possibilités d'analyse.** Si vous ne le faites pas dès le début de l'atelier, il sera difficile d'orienter efficacement les groupes de façon qu'ils obtiennent des résultats idéaux à partir des données qu'ils ont recueillies.
- Les membres du groupe peuvent se diviser en sous-groupes pour achever le traitement des données, mais assurez-vous de discuter de temps à autre des problèmes éventuels et des progrès réalisés avec tout le groupe.

**Formation à la recherche sur les systèmes de santé
Volume 2, 2^e partie : Analyse de données et rédaction de rapports**

Module 22

**DESCRIPTION DES VARIABLES
PREMIÈRE PARTIE**

ÉTAPES DE L'ANALYSE DE DONNÉES ET DE LA RÉDACTION DE RAPPORTS

Questions à poser	Étapes à suivre*	Éléments importants de l'étape
Quelles données ont été recueillies pour chaque objectif de recherche? Les données sont-elles complètes et exactes?	Préparer les données pour l'analyse	Revoir les activités sur le terrain Compiler les données pour chaque objectif ou population étudiée Classer les données et en vérifier la qualité Vérifier les imprimés
À quoi ressemblent les données?	Décrire les variables	Distributions Figures, moyennes
Comment peut-on résumer les données pour en faciliter l'analyse?	Faire le recoupement des données quantitatives	Faire le recoupement des données par rapport aux objectifs Illustrations, texte
	Résumer les données qualitatives	
Pour les données qualitatives : la recherche a-t-elle pour but de décrire, de comparer ou de trouver des associations?	Déterminer le type d'analyse statistique nécessaire	Examiner les objectifs, le type d'étude et les variables Faire une description statistique des variables Choisir les tests d'hypothèses
1. Comment peut-on décrire les données?		
2. Comment peut-on déterminer les différences entre les groupes?	Analyser des observations appariées et non appariées	**Test t de Student **Test t pour observations appariées **Test du khi carré **Test du khi carré de McNemar
3. Comment peut-on déterminer les associations entre les variables?	Évaluer les associations entre les variables	**Diagramme de dispersion **Droite de régression et coefficient de corrélation **Risque relatif, risque relatif approché
Comment le rapport doit être rédigé?	Rédiger le rapport et formuler des recommandations	Préparer le canevas du rapport Rédiger les 1 ^{er} et 2 ^e versions Faire un sommaire des constatations Faire un sommaire des conclusions pour chaque objectif Formuler des recommandations Préparer un résumé analytique
Comment doit-on présenter et diffuser les constatations et les recommandations?	Présenter un sommaire et rédiger un plan de mise en oeuvre des recommandations	Discuter des sommaires avec différents groupes cibles Discuter du programme de mise en oeuvre

* Il n'est pas nécessaire de faire ces étapes dans l'ordre indiqué. L'ordre peut être adapté aux besoins de l'atelier.

** Ces éléments sont facultatifs; ils peuvent être omis s'ils ne sont pas pertinents au cours.

Module 22 : DESCRIPTION DES VARIABLES, PREMIÈRE PARTIE

OBJECTIFS

À la fin de la présente séance, vous devriez pouvoir :

1. Décrire les données sous forme de distributions, de pourcentages et de proportions;
2. Utiliser des figures pour présenter des données;
3. Expliquer la différence entre la moyenne, la médiane et le mode;
4. Calculer les fréquences, les pourcentages, les rapports, les taux, les moyennes, les médianes et les modes pour les principales variables de votre étude, s'il y a lieu;
5. Identifier d'autres variables indépendantes (autres que celles qui ont été établies lors du premier atelier) qui sont nécessaires à l'analyse de vos données, s'il y a lieu.

I. Introduction

II. Distributions

III. Pourcentages, proportions, rapports et taux

IV. Figures

V. Mesures de tendance centrale

I. INTRODUCTION

Vous avez choisi les variables pour votre étude au **module 8** en croyant qu'elles vous aideraient à décrire votre problème (variables dépendantes) ou qu'elles étaient des facteurs qui contribuaient au problème (variables indépendantes). Le but de l'analyse des données consiste à déterminer quelles variables décrivent le mieux le problème et les facteurs qui l'influencent, ainsi que la façon dont les données répondent aux questions de la recherche établies dans les objectifs.

Avant d'examiner comment les variables peuvent s'influencer mutuellement, nous devons résumer les renseignements obtenus sur chaque variable dans un tableau simple ou dans une figure.

Certaines variables peuvent avoir donné des données numériques, alors que d'autres peuvent être catégoriques et avoir produit des données divisées en catégories. Lors de l'analyse des données, il est important de déterminer avant tout le type de données à traiter. Cette étape est cruciale dans l'organisation de notre approche à l'égard des méthodes statistiques, car le type de données utilisé détermine les techniques statistiques générales à employer.

Données catégoriques

Il existe deux types de données catégoriques : les données nominales et les données ordinales.

Dans le cas des données **NOMINALES**, les variables sont divisées en un certain nombre de catégories déterminées. Ces catégories, cependant, ne peuvent pas s'englober (l'une ne peut être plus ou moins large que l'autre).

Exemple

DONNÉES NOMINALES	CATÉGORIES
Sexe	masculin, féminin
État civil	célibataire, marié, veuf, séparé, divorcé

Dans le cas des données **ORDINALES**, les variables sont également divisées en un certain nombre de catégories, mais celles-ci peuvent s'englober et on peut les classer de la plus large à la moins large, ou vice versa.

Exemple

DONNÉES ORDINALES	CATÉGORIES
Degré de connaissance	bon, moyen, médiocre
Opinion sur un énoncé	tout à fait d'accord, d'accord, en désaccord, tout à fait en désaccord

Données numériques

Les données exprimées sous forme de nombres sont des données NUMÉRIQUES.

Certaines variables numériques sont essentiellement **discrètes**, telles que le nombre d'accidents de la route ou le prix des oeufs. Les valeurs possibles ne constituent qu'une série distincte de nombres. En fait, toutes les données sont consignées de façon discrète, car on les arrondit pour les simplifier.

Exemple

- hauteur arrondie au centimètre ou au pouce le plus proche,
- température exprimée en degrés Celsius,
- âge au dernier anniversaire de naissance.

Cependant, avec de telles données, on peut prévoir l'élaboration d'instruments de mesure de plus en plus précis et une consignation plus détaillée permettant de faire en sorte que le nombre de valeurs possibles augmente à l'infini et que les données deviennent essentiellement continues. En statistique, ces données sont désignées sous le nom de données **continues**.

Les données numériques peuvent être étudiées par

- des distributions,
- des pourcentages, des proportions, des rapports et des taux,
- des figures,
- des mesures de tendance centrale.

Nous discuterons de ces opérations à tour de rôle dans le contexte des données catégoriques et numériques.

II. DISTRIBUTIONS

Une **DISTRIBUTION** est une description des données dans un tableau afin d'en faciliter la manipulation. Cette description précise la fréquence à laquelle une valeur particulière revient dans les données (c'est-à-dire le nombre de fois qu'elle s'y trouve).

Dans votre projet de recherche, vous avez déjà calculé la fréquence de toutes les variables en vous fondant sur les fiches de données et en comptant le nombre de réponses dans chaque catégorie. Résumons maintenant quelques points importants.

- **Les données CATÉGORIQUES** peuvent être divisées en catégories très simples.

Exemple 1

Pour vérifier l'exactitude d'un diagnostic de malaria, des échantillons de sang de 33 patients ont fait l'objet d'un test de dépistage des parasites de la malaria. Trois résultats étaient possibles: négatif, *P. falciparum* ou *P. vivax*.

Les résultats sont présentés dans la **distribution** suivante :

Négatif	19
<i>P. falciparum</i>	13
<i>P. vivax</i>	1
TOTAL	33

Ces données sont **NOMINALES**. On calcule la **distribution** simplement en totalisant le nombre de réponses dans chaque catégorie.

Vous devriez toujours vérifier que le nombre total de réponses correspond au nombre de sujets (répondants). Au besoin, créez une catégorie pour les réponses manquantes.

La distribution révèle qu'en réalité, plus de la moitié des patients qui ont fait l'objet d'un diagnostic clinique de malaria ne sont pas atteints de cette maladie. On peut également observer que le parasite *P. falciparum* est beaucoup plus répandu que *P. vivax*.

Exemple 2

Des praticiens de la santé de 148 établissements ruraux de santé ont été invités à répondre à la question suivante : «À combien de reprises avez-vous manqué de médicaments pour traiter la malaria au cours des deux dernières années?»

Il s'agissait là d'une question fermée comportant les réponses possibles suivantes : jamais, 1 ou 2 fois (rarement), 3 à 5 fois (parfois) et plus de 6 fois (souvent). Le nombre de réponses dans chaque catégorie a été calculé pour constituer la distribution suivante.

Jamais	47
Rarement	71
Parfois	24
Souvent	6
TOTAL	148

Dans cet exemple, les données sont **ORDINALES**. L'ordre des catégories est important du fait que le problème s'aggrave de haut en bas.

La distribution indique que la plupart des cliniques ne manquent jamais ou manquent rarement de médicaments contre la malaria, mais que ce problème survient parfois dans environ le sixième des cliniques et souvent dans quelques-unes.

- Les distributions se font essentiellement de la même façon qu'il s'agisse de données **NUMÉRIQUES** ou de données catégoriques, sauf que les données doivent être divisées en catégories. Les étapes de l'établissement de la distribution sont les suivantes :
 1. Sélection des catégories;
 2. Calcul du nombre de mesures dans chaque catégorie;
 3. Calcul et vérification des résultats.

Lors du regroupement des données, le choix des catégories peut influencer les résultats. Il faut donc faire preuve de bon sens, et il peut même être nécessaire d'apporter des changements si des renseignements risquent d'être cachés en raison d'un mauvais choix de catégories.

Exemple 3

Les centres de santé du district X font état tous les jours du nombre de cas de malaria, dont vous voulez faire le sommaire. Comparez les sommaires quotidiens et hebdomadaires suivants des mêmes données :

Tableau 22.1. Sommaires quotidiens et hebdomadaires des cas de malaria dans les centres de santé du district X.

Jour 1	9 cas		
Jour 2	12		
Jour 3	11		
Jour 4	13		
Jour 5	14		
Jour 6	13		
Jour 7	16	1 ^{ère} semaine	88 cas
Jour 8	16 cas		
Jour 9	16		
Jour 10	18		
Jour 11	19		
Jour 12	16		
Jour 13	21		
Jour 14	25	2 ^e semaine	131 cas
Jour 15	28 cas		
Jour 16	28		
Jour 17	28		
Jour 18	32		
Jour 19	21		
Jour 20	19		
Jour 21	12	3 ^e semaine	168 cas

Les sommaires quotidiens et hebdomadaires indiquent un nombre croissant de cas de malaria. Cependant, si nous présentons seulement les données hebdomadaires, l'amélioration s'étant manifestée aux jours 19, 20 et 21 ne sera pas relevée. Il serait donc préférable de présenter les données quotidiennes pour montrer le moment exact où le nombre de cas déclarés de malaria commence à baisser.

Lors du regroupement des données, il faut suivre certaines règles importantes :

- Les groupes ne doivent pas se chevaucher, autrement on ne sait plus à quel groupe correspondent les mesures.
- Il doit y avoir une continuité entre les groupes; en d'autres mots, il faut éviter les écarts. Autrement, certaines mesures ne pourront correspondre à aucun groupe.
- Les groupes doivent aller de la mesure la plus petite à la plus grande, de sorte que toutes les mesures peuvent faire partie d'un groupe.
- Les groupes devraient être de taille égale de manière à faciliter le compte dans des groupes différents.

Cependant, il est parfois acceptable de choisir des groupes de taille différente, notamment si on s'intéresse à des groupes d'âge précis (p. ex., moins d'un an, 1 à 4 ans, 5 à 14 ans).

Lorsqu'il s'agit de faire un sommaire des données, il est préférable de choisir trop de groupes plutôt que pas assez. Il est alors possible de combiner des groupes pour constituer de nouvelles catégories sans étudier à nouveau l'ensemble des données, ce que vous seriez tenu de faire pour ajouter des groupes si vous n'en aviez pas assez.

Un nombre élevé de groupes permet en général d'avoir une idée plus précise des données, mais s'il y en a trop, la vue d'ensemble est moins claire.

Exemple : 1,00–9,99, 10,00–19,99, 20,00–29,99

III. POURCENTAGES, PROPORTIONS, RAPPORTS ET TAUX

Pourcentages

Dans les tableaux de fréquence, il est souvent préférable de présenter les données sous forme de pourcentages plutôt que de nombres absolus.

Le POURCENTAGE constitue le nombre d'unités présentant une certaine caractéristique divisé par le nombre total d'unités de l'échantillon et multiplié par 100.

Les pourcentages peuvent également être appelés FRÉQUENCES RELATIVES. Les pourcentages normalisent les données, ce qui les rend plus faciles à comparer à des données semblables obtenues à partir d'un autre échantillon de taille différente.

Exemple 4

Dans un district, 82 cliniques ont été appelées à indiquer le nombre de patients traités pour la malaria au cours d'une période d'un mois. Les chercheurs ont fourni la distribution et les pourcentages (ou fréquences relatives) :

Tableau 22.2. Distribution des cliniques selon le nombre de patients traités pour la malaria au cours d'un mois.

Nombre de patients	Nombre de cliniques ^a	Fréquence relative
0 à 19	25	31 %
20 à 39	3	4 %
40 à 59	5	6 %
60 à 79	11	14 %
80 à 99	19	24 %
100 à 119	10	12 %
120 à 139	4	5 %
140 à 159	3	4 %
Total	80	100 %

^a Les données de deux cliniques sont manquantes.

Remarque

Normalement, on ne tient pas compte des données manquantes dans le calcul des pourcentages.

La fréquence des réponses de chaque groupe constitue le pourcentage des éléments à l'étude pour lesquels vous avez obtenu les données (ou, si une question est posée aux répondants, le pourcentage des répondants qui ont répondu à la question).

Cependant, le nombre de cas où les données sont manquantes (p. ex., le nombre de personnes qui n'ont pas répondu à une question) constitue un bon moyen de déterminer si la collecte des données a été effectuée correctement. Par conséquent, il faut le mentionner, par exemple dans une remarque (voir le **tableau 22.2**).

N'oubliez pas que la catégorie «ne sait pas» est une catégorie véritable qui ne doit pas entrer dans le compte des données manquantes. S'il y a lieu, elle doit figurer dans le tableau.

Lorsque le nombre total est faible, il faut être prudent lors du calcul et de l'interprétation des pourcentages, car une différence d'une unité pourrait faire varier ceux-ci considérablement.

Ainsi, dans l'**exemple 1**, le nombre total d'échantillons de sang examinés n'est que de 33. Le parasite *P. falciparum* est décelé dans 13 échantillons, ce qui donne un pourcentage de 39 %. Si l'on avait trouvé ce parasite dans 14 échantillons au lieu de 13 (ce qui aurait très bien pu arriver), le pourcentage aurait été de 42 %, ce qui représente déjà une différence de 3 %.

Si le nombre total d'échantillons examinés avait été de 330, et si 130 d'entre eux (ou 39 %) avaient contenu le parasite *P. falciparum*, une différence d'un échantillon (131 ou 129) n'aurait pas fait varier beaucoup le pourcentage.

Il est donc recommandé d'accompagner le pourcentage du nombre d'observations ou du nombre total de cas étudiés.

Proportions

Parfois, les fréquences relatives sont exprimées sous forme de proportions plutôt que de pourcentages.

Une PROPORTION constitue une expression numérique qui compare une partie des unités étudiées à l'ensemble; elle s'exprime sous forme de FRACTION ou de DÉCIMALES.

Exemple 5

Sur un total de 55 patients fréquentant une clinique lors d'une journée donnée, 22 sont de sexe masculin et 33 de sexe féminin. On peut alors dire que la proportion de patients de sexe masculin est de 22/55 ou de 2/5, ce qui équivaut à 0,40 (le numérateur est 22 et le dénominateur 55).

Lorsqu'une proportion est exprimée sous forme de décimales, cette valeur multipliée par 100 donne un pourcentage. Dans l'exemple, 0,40 est égal à 40 %.

Rapports

Un **RAPPORT** est une expression numérique qui indique un lien entre deux ou plusieurs éléments sur le plan de la quantité ou de la taille.

Dans l'**exemple 5**, le rapport des patients de sexe masculin sur ceux de sexe féminin est de 22:33, c'est-à-dire 2:3.

Taux

Un **TAUX** est la quantité ou le degré d'une chose mesuré sur une période de temps donnée.

Les taux les plus fréquemment utilisés dans le secteur de la santé sont les suivants :

- | | |
|--------------------------------|---|
| • Taux de naissance | Nombre de naissances vivantes par 1 000 personnes au cours d'une période d'un an. |
| • Taux de mortalité | Nombre de décès par 1 000 personnes au cours d'une période d'un an. |
| • Taux de mortalité infantile | Nombre de décès d'enfants par 1 000 naissances vivantes. |
| • Taux de mortalité maternelle | Nombre de décès reliés à la grossesse au cours d'un an par 100 000 naissances vivantes au cours de la même année. |

IV. FIGURES

Si votre rapport contient beaucoup de tableaux descriptifs, vous pouvez le rendre plus facile à consulter en présentant les plus importants sous forme de figures.

Les figures les plus fréquemment utilisées pour présenter des données sont les suivantes :

- | | |
|--------------------------------|---------------------------------|
| • Diagrammes à barres | } pour les données catégoriques |
| • Diagrammes circulaires | |
| • Histogrammes | } pour les données numériques |
| • Graphiques linéaires simples | |
| • Diagrammes de dispersion | |
| • Cartes | |

Nous verrons maintenant des exemples de ces figures que l'on peut utiliser pour présenter des données.

Diagrammes à barres

Les données de l'exemple 2 peuvent être présentées dans un diagramme à barres, en utilisant les fréquences absolues (voir la figure 22.1) ou les fréquences relatives (pourcentages) (voir la figure 22.2).

Figure 22.1. Fréquence des pénuries de médicaments contre la malaria dans les établissements de santé ruraux.

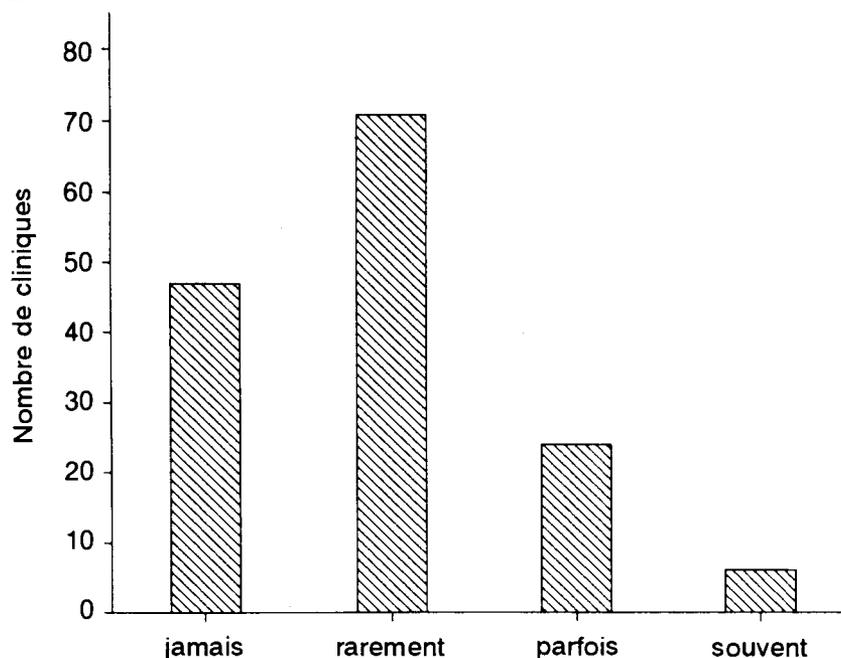
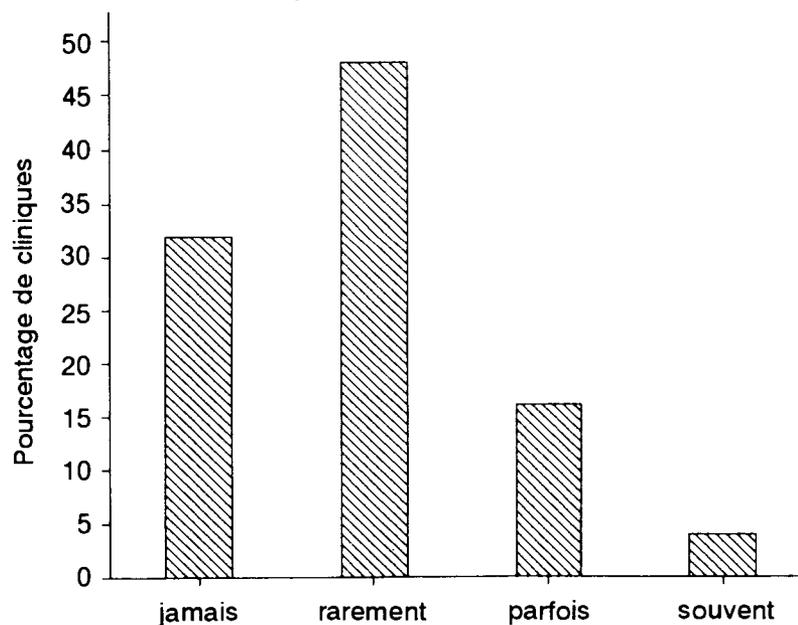


Figure 22.2. Fréquence relative des pénuries de médicaments contre la malaria dans les établissements de santé ruraux (n = 148).

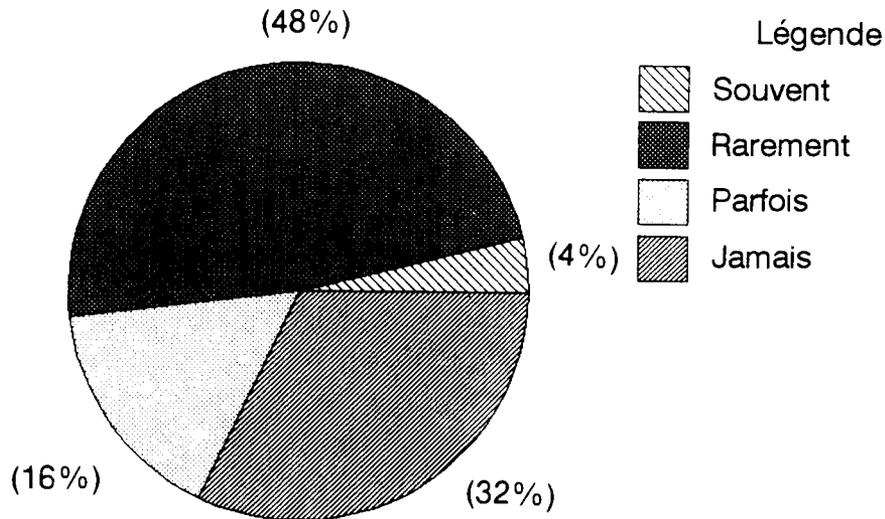


On remarque que la taille de l'échantillon doit être indiquée lorsqu'on présente les données sous forme de pourcentages.

Diagrammes circulaires

Un diagramme circulaire peut être utilisé pour illustrer le même ensemble de données, ce qui permet au lecteur d'avoir un aperçu des données sous une forme différente. Le diagramme circulaire illustre la fréquence relative d'un certain nombre de possibilités, de façon que le total de ses segments totalise 100 %.

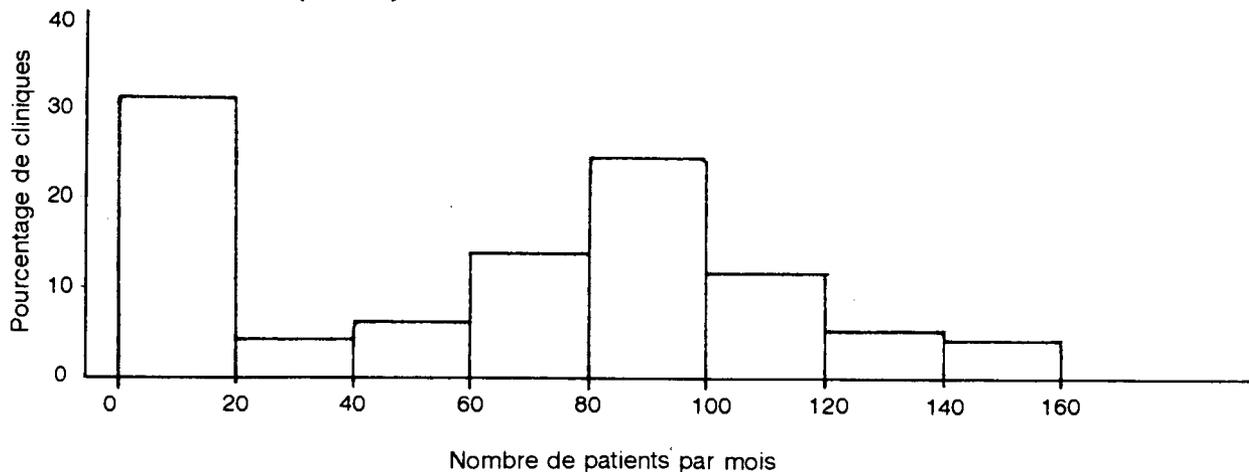
Figure 22.3. Fréquence relative des pénuries de médicaments contre la malaria dans les établissements de santé ruraux (n = 148).



Histogrammes

Les données numériques sont souvent présentées sous forme d'histogrammes, qui ressemblent beaucoup aux diagrammes à barres utilisés pour les données catégoriques. Cependant, ils s'en distinguent par le fait que les «barres» d'un histogramme sont reliées entre elles (pourvu qu'il n'y ait pas d'écart entre les données), alors que les barres d'un diagramme à barres ne le sont pas, car les différentes catégories sont des éléments distincts. Les données de l'exemple 4 sont présentées sous forme d'histogramme à la figure 22.4.

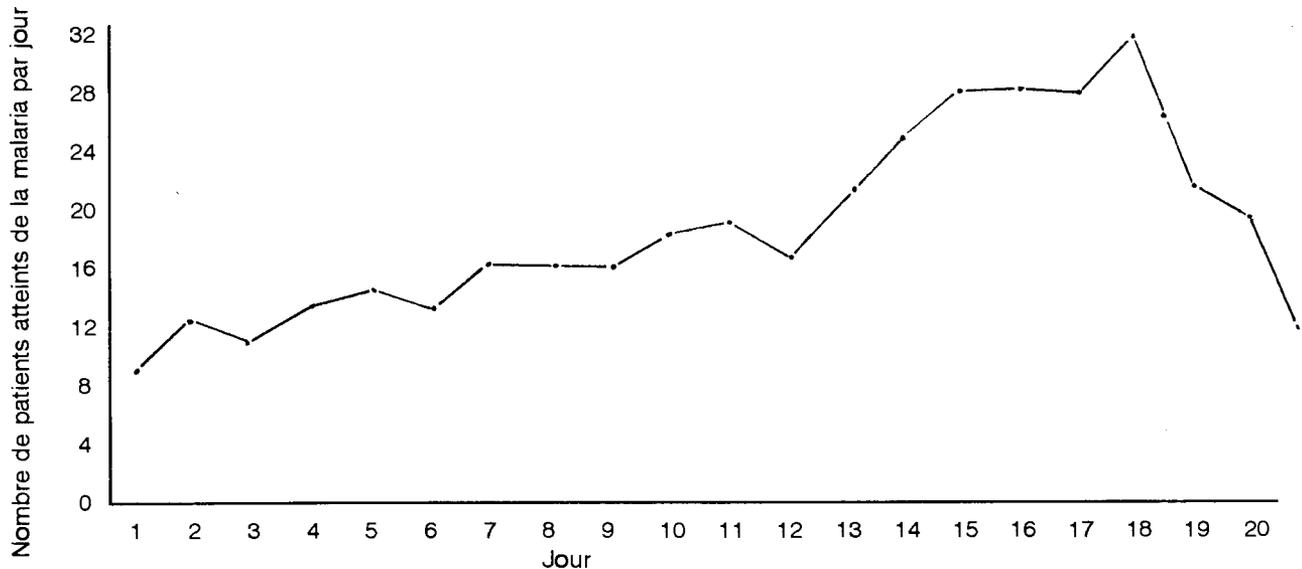
Figure 22.4. Pourcentage de cliniques fournissant des traitements à un nombre différent de patients atteints de la malaria (n = 80).



Graphiques linéaires simples

Le graphique linéaire simple est particulièrement utile pour les données numériques lorsqu'on veut illustrer une tendance sur une certaine période. Les données de l'**exemple 3** sont présentées de cette façon à la **figure 22.5**. On peut illustrer une ou plusieurs distributions dans un même graphique, pourvu que les lignes soient faciles à distinguer. Il est alors possible de les comparer.

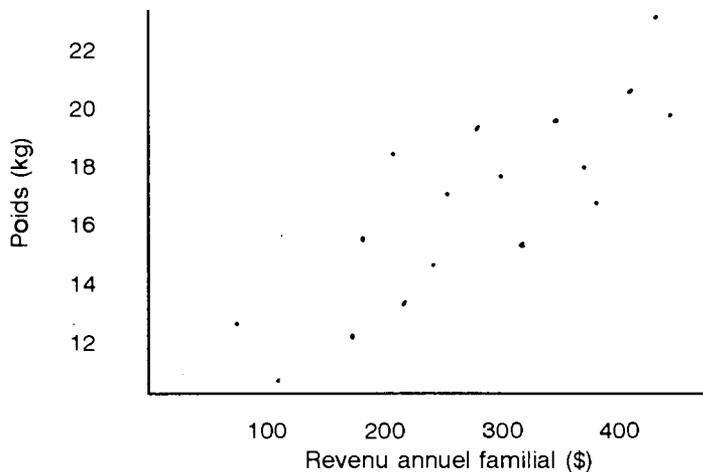
Figure 22.5. Nombre de patients atteints de la malaria par jour dans les établissements de santé du district X.



Diagrammes de dispersion

Les diagrammes de dispersion permettent de présenter des renseignements sur deux variables qui sont peut-être reliées. L'exemple suivant est utilisé au **module 29**, où on traite des notions d'association et de corrélation.

Figure 22.6. Poids des enfants de 5 ans selon le revenu annuel familial.



Remarque

Il est important que toutes les figures contenues dans votre rapport de recherche soient numérotées et titrées et que les données soient bien identifiées.

Outre les figures ci-dessus, il est possible de présenter de l'information au moyen de **CARTES**. Par exemple, la région soumise à l'étude peut être représentée sur une carte. Si l'étude porte sur l'épidémiologie du choléra, une carte pourrait indiquer la répartition géographique des cas de choléra ainsi que celle des sources d'eau protégées, illustrant ainsi la présence d'une association. Si l'étude porte sur la vaccination, une carte pourrait indiquer l'emplacement des cliniques et la proportion des personnes de moins de cinq ans qui ont été vaccinées dans chaque village, ce qui permettrait peut-être de déterminer que la distance entre les domiciles et la clinique constitue un important facteur associé à l'état de la vaccination.

V. MESURES DE TENDANCE CENTRALE

Les distributions et les histogrammes constituent des moyens utiles de présenter un ensemble d'observations sur une variable. Dans bien des cas, ils sont essentiels pour comprendre les tendances des données. Cependant, pour mieux résumer un ensemble d'observations, il est souvent utile d'employer une mesure pouvant être exprimée par un seul chiffre.

En premier lieu, il est possible de mesurer le centre de la distribution. Les trois mesures employées pour ce faire sont la **MOYENNE**, la **MÉDIANE** et le **MODE**.

Moyenne

La **MOYENNE** (ou moyenne arithmétique) constitue le total des résultats de toutes les observations divisé par le nombre d'observations. On ne peut calculer la moyenne que des données numériques.

Exemple 6

Une mesure de la taille de 7 femmes a donné les résultats suivants :

141, 141, 143, 144, 145, 146, 155 cm (total de 1 015 cm pour 7 mesures)

La moyenne est donc de $1015/7$, c'est-à-dire 145 cm.

Médiane

La **MÉDIANE** est la valeur qui divise une distribution en deux parties égales.

La médiane est utile lorsque certaines mesures sont beaucoup plus grandes ou beaucoup plus petites que les autres. La moyenne de telles mesures est biaisée en faveur de ces valeurs extrêmes; elle ne constitue donc pas une bonne mesure du centre de la distribution dans ce cas. La médiane, par contre, n'est pas influencée par les valeurs extrêmes.

La médiane se calcule de la façon suivante :

- Placer les observations en ordre ascendant;
- Compter le nombre d'observations (n);
- La médiane est la valeur correspondant à l'observation $(n + 1)/2$.

Exemple 8

Voici le poids de 7 femmes enceintes :

40, 41, 41, 43, 44, 47 et 72 kg.

La médiane est la valeur correspondant à l'observation $(7 + 1)/2$, c'est-à-dire la quatrième, qui est de 43 kg.

Soulignons que la moyenne de ces observations est de 47 kg. Il s'agit là d'une illustration du fait que la moyenne est influencée par les valeurs extrêmes (72 kg dans ce cas), contrairement à la médiane. Si le poids le plus élevé dans cette série avait été de 51 kg plutôt que de 72 kg, la médiane aurait toujours été de 43 kg, mais la moyenne aurait été de 44 kg.

Mode

Le MODE est la valeur la plus fréquente d'une série d'observations.

Le mode n'est pas très utile pour les données numériques continues. Cependant, on l'établit pour les données numériques divisées en classes.

Dans l'**exemple 4** (nombre de patients traités pour la malaria dans les cliniques), le mode est de «0 à 19», car ce résultat est le plus fréquent (25 cas sur 80).

Le mode peut également être utilisé avec des données catégoriques, qu'elles soient nominales ou ordinales.

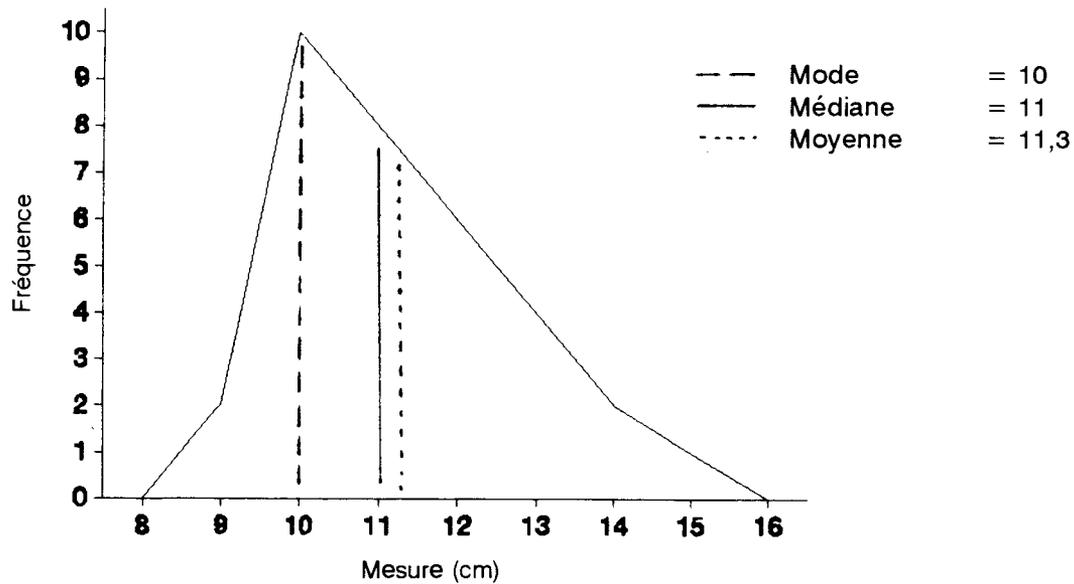
Dans l'**exemple 1** (diagnostic clinique de malaria), le mode est «négatif». Dans l'**exemple 2** (nombre de cliniques ayant subi une pénurie de médicaments), le mode est «rarement».

En résumé, la moyenne, la médiane et le mode sont des mesures de tendance centrale. La moyenne est la plus courante; elle est plus informative parce que son calcul tient compte de la valeur de toutes les observations.

Cependant, la moyenne est fortement influencée par des valeurs éloignées du centre de la distribution, alors que la médiane et le mode ne le sont pas. Le calcul de la moyenne constitue la première d'une série de procédures statistiques plus complexes visant à décrire et à analyser des données.

La **figure 22.7** montre une courbe de distribution où la moyenne, la médiane et le mode ont une valeur différente.

Figure 22.7. Moyenne, médiane et mode dans une courbe de distribution.



TRAVAIL EN GROUPE

- Décrivez vos échantillons en fonction des variables de base (sexe, âge, etc.) et des variables dépendantes (p. ex., ne se conforme pas/se conforme, usager/non-usager).
- Assurez-vous de vérifier la fréquence de toutes les variables de votre étude (à partir de votre fiche de données). Calculez les pourcentages en fonction du nombre total d'unités étudiées (ou, s'il y a lieu, calculez les proportions, les rapports ou les taux).
- Vérifiez vos objectifs pour déterminer les variables qui devraient faire l'objet de tableaux des fréquences dans votre rapport. Habituellement, on présente un tableau des fréquences pour certaines variables de base, les variables dépendantes et les variables indépendantes les plus importantes. Préparez les tableaux des fréquences.
- Si nécessaire, tracez des histogrammes, des diagrammes à barres, des diagrammes circulaires et des graphiques linéaires simples. Préparez une description et une interprétation brèves des figures.
- Calculez au besoin les moyennes, médianes et modes et interprétez les résultats.
- Familiarisez-vous avec les résultats et essayez d'en saisir toute la signification.

Notes du formateur

Module 22: DESCRIPTION DES VARIABLES, PREMIÈRE PARTIE

Durée et méthode d'enseignement

1 heure Introduction et discussion

3 heures⁺ Travail en groupe**Introduction et discussion**

- Les participants se sont probablement familiarisés avec certaines notions présentées dans le présent module, telles que les pourcentages et les proportions. En outre, les groupes ont déjà préparé des distributions (et calculé des pourcentages). Afin de maintenir l'intérêt des élèves, surtout si leur niveau de connaissances est élevé, abordez ces notions brièvement. Cependant, une attention particulière doit être accordée à la question des valeurs manquantes lors du calcul des pourcentages.
- Bien que le module contienne une définition de pourcentage, de proportion, de rapport et de taux, il est plus important de fournir des exemples ou de demander aux participants d'en imaginer.
- En présentant l'**exemple 3**, vous pouvez également demander aux participants de décrire comment ils ont regroupé les données numériques et de dire s'ils avaient trop ou trop peu de catégories.
- Il ne suffit pas de présenter les exemples; il faut également les reprendre lors d'exercices. Par exemple, demandez aux participants ce qu'est la moyenne, la médiane et le mode d'une série de mesures, au lieu de leur donner tout de suite les réponses.

Travail en groupe

- Avant que les participants ne calculent la fréquence des données au moyen de la fiche de données, demandez-leur de vérifier si les données ont été divisées correctement en catégories. Assurez-vous également que le nombre total de répondants de tous les groupes étudiés a été établi.
- Rappelez aux participants qu'il n'est nécessaire de préparer des tableaux des fréquences complets que pour les variables à décrire dans le rapport final. Habituellement, il faut préparer un tableau pour certaines variables de base, les variables dépendantes et les variables indépendantes les plus importantes. Bon nombre des autres variables de base et des variables indépendantes seront présentées sous forme de tableaux croisés (**module 23**).

**Formation à la recherche sur les systèmes de santé
Volume 2, 2^o partie : Analyse de données et rédaction de rapports**

Module 23

RECOUPEMENT DES DONNÉES QUANTITATIVES

ÉTAPES DE L'ANALYSE DE DONNÉES ET DE LA RÉDACTION DE RAPPORTS

Questions à poser	Étapes à suivre*	Éléments importants de l'étape
Quelles données ont été recueillies pour chaque objectif de recherche? Les données sont-elles complètes et exactes?	Préparer les données pour l'analyse	Revoir les activités sur le terrain Compiler les données pour chaque objectif ou population étudiée Classer les données et en vérifier la qualité Vérifier les imprimés
À quoi ressemblent les données?	Décrire les variables	Distributions Figures, moyennes
Comment peut-on résumer les données pour en faciliter l'analyse?	Faire le recoupement des données quantitatives Résumer les données qualitatives	Faire le recoupement des données par rapport aux objectifs Illustrations, texte
Pour les données qualitatives : la recherche a-t-elle pour but de décrire, de comparer ou de trouver des associations?	Déterminer le type d'analyse statistique nécessaire	Examiner les objectifs, le type d'étude et les variables Faire une description statistique des variables Choisir les tests d'hypothèses
1. Comment peut-on décrire les données?		
2. Comment peut-on déterminer les différences entre les groupes?	Analyser des observations appariées et non appariées	**Test t de Student **Test t pour observations appariées **Test du khi carré **Test du khi carré de McNemar
3. Comment peut-on déterminer les associations entre les variables?	Évaluer les associations entre les variables	**Diagramme de dispersion **Droite de régression et coefficient de corrélation **Risque relatif, risque relatif approché
Comment le rapport doit être rédigé?	Rédiger le rapport et formuler des recommandations	Préparer le canevas du rapport Rédiger les 1 ^{re} et 2 ^e versions Faire un sommaire des constatations Faire un sommaire des conclusions pour chaque objectif Formuler des recommandations Préparer un résumé analytique
Comment doit-on présenter et diffuser les constatations et les recommandations?	Présenter un sommaire et rédiger un plan de mise en oeuvre des recommandations	Discuter des sommaires avec différents groupes cibles Discuter du programme de mise en oeuvre

* Il n'est pas nécessaire de faire ces étapes dans l'ordre indiqué. L'ordre peut être adapté aux besoins de l'atelier.

** Ces éléments sont facultatifs; ils peuvent être omis s'ils ne sont pas pertinents au cours.

Module 23: RECOUPEMENT DES DONNÉES QUANTITATIVES

OBJECTIFS

À la fin de la présente séance, vous devriez pouvoir :

1. **Effectuer** tous les recoupements importants qui vous aideront à atteindre vos objectifs de recherche;
2. **Identifier** les variables confusionnelles qui doivent être envisagées dans l'étude des relations entre variables et prendre les mesures nécessaires.

- I. Introduction
- II. Types de recoupements
- III. Préparation de tableaux croisés correspondant aux objectifs de recherche
- IV. Généralités sur la préparation de tableaux
- V. Les variables confusionnelles : stratification et appariement

I. INTRODUCTION

Jusqu'à maintenant, nous avons préparé des tableaux contenant la distribution d'une variable à la fois pour décrire en partie nos données. Selon les **objectifs** de notre étude et le **type d'étude**, il est possible que nous devions **étudier les liens entre plusieurs variables à la fois** afin de décrire adéquatement le problème ou d'identifier des facteurs pouvant l'expliquer.

À cette fin, il convient de faire des recoupements au moyen de TABLEAUX CROISÉS.

Exemple 1

Nous voulons connaître l'âge auquel les grossesses se produisent chez les adolescentes et déterminer si elles sont plus fréquentes chez les filles d'âge scolaire que chez celles qui ne fréquentent pas l'école. Pour répondre à ces questions, nous préparons le tableau croisé suivant (les données sont fictives).

Tableau 23.1 Nombre de grossesses à différents âges chez les adolescentes fréquentant et ne fréquentant pas l'école (province X, 1988-1990).

Âge au début de la grossesse	Nombre de grossesses		Total
	Fréquentant l'école	Ne fréquentant pas l'école	
12 ans	2 (3 %)	1 (2 %)	3
13 ans	2 (3 %)	0	2
14 ans	5 (7 %)	2 (4 %)	7
15 ans	23 (34 %)	12 (23 %)	35
16 ans	36 (53 %)	37 (71 %)	73
Total	68 (100 %)	52 (100 %)	120

Les tableaux croisés qui décrivent un problème ou une situation, tel que celui du tableau 23.1, sont des TABLEAUX DESCRIPTIFS. Les tableaux croisés qui visent à trouver des explications possibles à un problème ou qui décrivent les résultats d'une intervention sont des TABLEAUX ANALYTIQUES.

Quelques exemples de tableaux croisés descriptifs et analytiques figurent aux pages suivantes. Vous remarquerez que dans chaque exemple, la nature et la présentation du tableau croisé sont fondées sur les objectifs de l'étude et son type.

II. TYPES DE RECOUPEMENTS

Selon les objectifs et le type de l'étude, trois types de recoupements différents peuvent se révéler nécessaires :

- **Les recoupements visant à décrire l'échantillon** permettent de décrire le problème étudié en présentant une combinaison de variables. Le **tableau 23.1** en est un exemple. Les recoupements descriptifs servent également à décrire l'échantillon étudié selon une combinaison de variables de base, comme l'âge, le sexe, la profession ou le lieu de résidence.

- Les recoupements comparant des groupes afin de déterminer les différences entre eux.
- Les recoupements permettant d'étudier des liens entre variables.

Des exemples de ces trois types de recoupements figurent ci-dessous. Soulignons que les tableaux qui ne contiennent pas encore de données sont appelées des MAQUETTES DE TABLEAUX.

Recoupements visant à décrire l'échantillon

Dans les études qui donnent des données quantitatives, qu'elles soient descriptives ou analytiques, il est fréquent de commencer par décrire les sujets compris dans les échantillons avant de présenter les résultats de l'étude. On peut le faire pour des variables distinctes dans un tableau des fréquences simple (tel qu'indiqué au **module 22**) ou pour une combinaison de variables dans un tableau croisé.

Exemple 2

Une étude a été menée sur le degré de satisfaction (variable dépendante) des médecins et infirmiers à l'égard de leur emploi dans des régions rurales et urbaines. Pour décrire l'échantillon, on a effectué un recoupement comprenant le sexe et le lieu de résidence (rural ou urbain) des médecins et des infirmiers interrogés. Cette méthode était motivée par le fait que l'analyse de l'opinion du personnel masculin et féminin devait être effectuée séparément pour les régions rurales et urbaines.

Tableau 23.2. Lieu de résidence et sexe des médecins et infirmiers.

Travailleurs de la santé		Lieu de résidence		Total
		rural	urbain	
Médecins	Hommes	8 (10 %)	35 (21 %)	43
	Femmes	2 (3 %)	16 (10 %)	18
Infirmiers	Hommes	46 (58 %)	36 (22 %)	82
	Femmes	23 (29 %)	77 (47 %)	100
Total		79 (100 %)	164 (100 %)	243

Le tableau 23.2 peut également être divisé en deux tableaux croisés séparés pour le lieu de résidence et le type de travailleur de la santé; l'un pour les hommes et l'un pour les femmes.

Dans les tableaux, les données figurent habituellement en nombres absolus et en fréquences relatives (pourcentages ou proportions). La moyenne, la médiane et le mode des données numériques (telles que l'âge) peuvent être calculés pour décrire l'échantillon.

Dans les études descriptives qui visent à quantifier une question, le tableau croisé est un moyen utile de présenter ses constatations.

Exemple 3

Une étude a été effectuée sur les facteurs causant la forte mortalité dans un hôpital. La maquette de tableau croisé suivante pourrait indiquer combien de mort-nés frais et macérés pesaient moins de 2 500 grammes et combien pesaient 2 500 grammes ou plus.

Tableau 23.3. Poids des foetus frais et macérés.

Poids du foetus	État à la naissance		Total
	Frais	Macéré	
Moins de 2 500 g			
2 500 g ou plus			
Total	(100 %)	(100 %)	

Remarque : Les données sont inscrites dans chaque cellule du tableau après la collecte.

Différenciation de groupes au moyen des tableaux croisés

Dans des études comparatives telles que des études cas-témoins, des études de cohortes ou des études quasi expérimentales, l'un des objectifs consiste à découvrir s'il y a des différences entre deux ou plusieurs groupes concernant des variables particulières. Il est alors nécessaire de préparer un tableau croisé des données.

Exemple 4

Dans le cadre d'une étude quasi expérimentale sur l'effet de la participation des mères à des discussions sur l'alimentation sur leurs connaissances à ce sujet, deux groupes de mères ont été comparés; celles qui ont participé et celles qui n'ont pas participé aux discussions sur l'alimentation. La maquette de tableau croisé suivante a été préparée :

Tableau 23.4. Nombre de participantes et de non-participantes à des discussions sur l'alimentation et niveau de connaissances sur l'alimentation.

	Niveau de connaissances sur l'alimentation			Total
	Faible	Moyen	Élevé	
Participant				(100 %)
Non-participant				(100 %)
Total				

Dans cet exemple, il faut définir clairement ce que constitue un niveau de connaissance «faible», «moyen» et «élevé» (voir le **module 8**).

Soulignons que dans le tableau croisé précédent, les groupes à comparer sont inscrits dans les rangées, tandis que les niveaux de connaissances sur l'alimentation forment les colonnes. On remplit le tableau de cette façon parce que les connaissances sur l'alimentation constituent le résultat de la participation aux discussions, et qu'elles sont donc considérées comme la variable dépendante. Les variables dépendantes forment habituellement les colonnes.

Exemple 5

Dans le cadre d'une étude cas-témoins sur la malnutrition, des enfants souffrant de malnutrition grave ont été comparés à des enfants bien nourris afin de trouver des moyens adéquats de soulager le problème de la malnutrition chez les enfants. La maquette de tableau croisé suivante a été préparée :

Tableau 23.5. Niveau des connaissances sur l'alimentation des mères ayant des enfants très mal nourris et des enfants bien nourris (12 à 35 mois).

Niveau de connaissance des mères sur l'alimentation	État des enfants		Total
	Très mal nourris	Bien nourris	
Faible			
Moyen			
Élevé			
Total	(100 %)	(100 %)	

Soulignons la différence entre les tableaux 23.4 et 23.5. Dans celui-ci, les connaissances sur l'alimentation constituent la variable indépendante, et elles forment donc les rangées, tandis que les deux groupes à comparer forment les colonnes.

D'autres tableaux croisés analytiques peuvent être préparés pour l'étude mentionnée à l'exemple 5. Chaque fois, les deux groupes (enfants très mal nourris et enfants bien nourris) peuvent former les colonnes. Des variables indépendantes différentes peuvent former les rangées, comme la source d'eau potable (protégée ou non protégée) ou l'état de la vaccination (complète ou non).

Remarques

Lors de la préparation de tableaux croisés analytiques visant à déceler des différences entre des groupes, déterminez d'abord si les groupes à comparer doivent former les rangées ou les colonnes, selon le type d'étude. (Selon la convention, les catégories de la variable dépendante forment les colonnes et celles des variables indépendantes les rangées.)

Si l'on ajoute des pourcentages au tableau croisé, il ne faut pas oublier que le total pour chaque groupe à comparer doit être de 100 %. (Voyez la différence entre les tableaux 24.4 et 24.5.)

Une fois remplis de données, les tableaux croisés sont utilisés dans les tests statistiques. Ces tests sont abordés aux **modules 27 et 28**.

Étude des liens entre variables au moyen des tableaux croisés

De nombreux projets de recherche sur les systèmes de santé visent à étudier des liens ou des associations possibles entre variables. Dans ce cas, il est important de déterminer si les variables sont indépendantes ou dépendantes.

Exemple 6

Dans le cadre d'une étude des pratiques en matière d'allaitement, vous étudiez le lien entre l'âge de la mère et la durée de l'allaitement. Le tableau croisé 23.6 se révèle alors utile.

Tableau 23.6. Durée de l'allaitement naturel chez les femmes de différentes catégories d'âge.

Âge (années)	Durée de l'allaitement naturel			Total
	0-5 mois	6-11 mois	> 12 mois	
15-19				
20-24				
25-29				
30-34				
35-39				
40 +				
Total				100 %

Le tableau 23.7 permet de déterminer s'il existe une association entre la situation d'emploi de la mère (variable indépendante) et la durée de l'allaitement naturel (variable dépendante).

Tableau 23.7. Durée de l'allaitement naturel selon la situation d'emploi de la mère.

Situation d'emploi de la mère	Durée de l'allaitement naturel			Total
	0-5 mois	6-11 mois	> 12 mois	
Employée à temps plein				
Employée à temps partiel				
Sans emploi				
Total				100 %

Les tableaux croisés analytiques préparés pour étudier des liens entre variables peuvent être soumis à une analyse statistique. Les **modules 29 et 30** abordent cette question en détail.

III. PRÉPARATION DE TABLEAUX CROISÉS CORRESPONDANT AUX OBJECTIFS DE LA RECHERCHE

Au moment de la conception de votre projet de recherche, vous avez été appelé à préparer des maquettes de tableaux pour les données que vous prévoyiez recueillir (**module 13**). Ces tableaux étaient fondés sur les objectifs et le type de l'étude.

Comme vous avez maintenant recueilli vos données et que vous avez une idée de leur qualité et de la façon de les utiliser, vous devez réexaminer de façon systématique la façon dont vous devez préparer les tableaux croisés.

Pour préparer des tableaux croisés adéquats, nous vous recommandons de suivre les étapes suivantes.

1. Étudiez chaque objectif et la méthode choisie pour recueillir les données correspondantes.
2. Formulez des phrases hypothétiques composant les conclusions que vous escomptez pour chaque objectif.

Par exemple, dans une étude descriptive sur les pratiques en matière d'allaitement naturel où l'un des objectifs spécifiques consiste à déterminer les facteurs associés au sevrage précoce, les conclusions prévues seraient les suivantes :

- «Les mères qui ont un emploi sèvent leurs enfants plus tôt que celles qui n'ont pas d'emploi.»
- Les mères qui n'ont pas participé à des discussions sur l'alimentation sèvent leurs enfants plus tôt que celles qui y ont participé.»

La formulation de conclusions possibles vous aide à :

- garder à l'esprit l'objet de chaque tableau et calcul que vous effectuez;
 - éviter de gaspiller du temps à faire des calculs et des tableaux inutiles;
 - organiser vos données de façon à faciliter la rédaction d'un rapport bien organisé.
3. Pour chaque «conclusion prévue», préparez les maquettes de tableaux croisés qui vous permettront d'arriver à la bonne conclusion.
 4. Calculez les fréquences (au moyen des fiches de données) et inscrivez les résultats dans les cellules du tableau croisé.
 5. Interprétez le tableau et rédigez une conclusion claire. Il n'est pas nécessaire de décrire le contenu du tableau en détail.

EXERCICE 1

Choisissez un objectif spécifique de chaque projet de recherche, formulez des «conclusions prévues» et préparez la maquette de tableau croisé appropriée.

IV. GÉNÉRALITÉS SUR LA PRÉPARATION DE TABLEAUX

- Assurez-vous que les catégories des variables figurant dans les tableaux ont été précisées et qu'elles s'excluent mutuellement (c.-à-d. qu'il n'y a ni chevauchement ni écart).
- Lors de la préparation d'un tableau croisé, assurez-vous que le total des colonnes et des rangées est égal à la fréquence de chaque variable.
- Vérifiez également que le total du tableau correspond bien au nombre de sujets de l'échantillon.
- Donnez à chaque tableau un titre clair. Veillez également à ce que les titres des colonnes et des rangées soient sans équivoque.

- Numérotez vos tableaux et placez-les avec les objectifs correspondants. De cette façon, vous pourrez mieux organiser votre rapport et vous assurer que vous ne répétez pas des étapes inutilement.

V. LES VARIABLES CONFUSIONNELLES : STRATIFICATION ET APPARIEMENT

Stratification

Si un ou plusieurs objectifs de votre étude consistent à évaluer des liens entre variables, il est important de déterminer si d'autres facteurs influent sur ces liens. Ces facteurs sont appelés des VARIABLES CONFUSIONNELLES (voir les **modules 8 et 9**).

Dans l'**exemple 6**, il est possible que la scolarité de la mère soit une variable confusionnelle, car on pourrait l'associer à la fois à la situation d'emploi de la mère et à la durée de l'allaitement naturel (voir le tableau 23.7).

Après avoir recueilli les données, il faut soumettre les variables confusionnelles à la STRATIFICATION, qui comprend une analyse DISTINCTE de tous les éléments de cette variable.

Dans l'**exemple 6**, cette analyse nécessite la préparation de tableaux croisés pour les mères de scolarité différente. Cela signifie que le tableau 23.7 doit être séparé en deux tableaux : l'un pour les mères de faible scolarité, c'est-à-dire moins de cinq ans, et un autre pour celles qui ont cinq années de scolarité ou plus.

Si l'on découvre une association SEMBLABLE entre la situation d'emploi et la durée de l'allaitement naturel chez les deux groupes de femmes, la scolarité de la mère n'est pas une variable confusionnelle.

L'**annexe 3** du **module 27** décrit le test statistique à employer en présence de variables confusionnelles.

Appariement

Le **module 9** (type d'étude) explique que si l'on soupçonne dès l'étape de la conception de l'étude qu'une variable sera confusionnelle, il est possible d'annuler l'effet de cette variable en APPARIANT les observations. Chaque sujet de l'étude est alors jumelé à un autre sujet dans le groupe témoin pour la variable confusionnelle en question.

Dans l'**exemple 6**, il faut choisir pour chaque mère de moins de cinq ans de scolarité une mère sans emploi de scolarité semblable. De même, pour chaque mère de cinq ans ou plus de scolarité, il faut choisir une mère sans emploi ayant cinq ans de scolarité ou plus.

Il est important de tenir compte du fait que les données sont jumelées lors de la préparation des tableaux croisés et de l'analyse.

L'analyse des observations appariées sera abordée en détail au **module 28** (où des groupes sont comparés pour en relever les différences) et au **module 30** (deuxième partie de la section III sur les études cas-témoins avec des observations appariées).

TRAVAIL EN GROUPE

- Évaluez chaque objectif spécifique et les critères de recherche qui s'y appliquent; formulez des phrases hypothétiques qui décrivent le type de conclusions attendues pour chaque objectif.
- Préparez des maquettes de tableaux croisés en déterminant s'il s'agit de :
 - décrire les sujets de l'échantillon ou le problème;
 - comparer des groupes pour en déterminer les différences;
 - trouver des associations entre variables.

Consultez les maquettes de tableaux croisés déjà préparées au premier atelier (**module 13**).

- Dans le cas des tableaux croisés analytiques, tentez d'identifier des variables confusionnelles éventuelles en faisant d'autres maquettes (stratification).
- Enfin, remplissez les maquettes de tableaux croisés avec les données et formulez une conclusion à inscrire directement sous chaque tableau.

Notes du formateur

Module 23 : RECOUPEMENT DES DONNÉES QUANTITATIVES

Durée et méthode d'enseignement

- | | |
|-----------|--|
| 1 heure | Introduction et discussion |
| 3 heures* | Travail en groupe |
| 1 heure | Exposés des groupes et discussion en plénière (facultatif) |

Introduction et discussion

- Il est recommandé d'utiliser un rétroprojecteur (ou un tableau à feuilles mobiles) pour présenter et expliquer la préparation des différents tableaux afin de retenir l'attention des participants. Ne vous contentez pas de citer les modules.
- Certains tableaux croisés présentés dans ce module contiennent des données fictives. Les exemples peuvent ainsi être plus concrets et il est possible de montrer comment les tableaux doivent être interprétés. Cependant, il faut porter une attention particulière à la préparation des tableaux (quant au contenu des colonnes et des rangées). Il est donc recommandé d'employer deux transparents superposés pour présenter le tableau : l'un de la maquette du tableau et l'autre contenant les données.
- Portez une attention particulière à la façon de lire chaque tableau : certains doivent être lus horizontalement, d'autres verticalement, selon que les groupes à comparer forment les rangées ou les colonnes. La valeur «100 %» insérée dans la plupart des tableaux indique le sens de lecture.
- L'exercice de groupe figurant à la fin de la section III donne la chance aux participants d'essayer de préparer des tableaux croisés appropriés. Essayez d'obtenir des groupes au moins un tableau des trois types (un tableau descriptif et un tableau analytique des deux types).

Assurez-vous que les groupes s'entendent sur le sens de lecture du tableau (horizontal ou vertical). S'ils ont de la difficulté à préparer des tableaux appropriés pour leur propre projet, ils devraient être invités à présenter quelques tableaux croisés en plénière après la séance de groupe pour obtenir des commentaires.

- Les généralités sur la préparation de tableaux (section IV du module) devraient être illustrées d'un exemple de tableau tiré du module ou fourni par l'un des groupes.
- La section V du module (qui traite des variables confusionnelles) doit demeurer simple. Cependant, tous les groupes devraient s'employer à déceler les variables confusionnelles afin d'éviter les conclusions non fondées.

Travail en groupe

- Passez en revue les trois premières étapes du travail avec le groupe. Les tableaux pourront être remplis par des sous-groupes.
- Précisez que les tableaux doivent être interprétés dès qu'ils sont remplis. Les conclusions provisoires doivent être inscrites sous le tableau correspondant pour faciliter la rédaction du rapport au cours de la deuxième semaine. Plus tard, ces conclusions feront l'objet de discussions avec les autres membres du groupe.
- Ensuite, les tableaux sont numérotés selon l'objectif auquel ils correspondent.

Exposés et plénière (facultatif)

- Demandez à tous les groupes de présenter au moins un tableau rempli ainsi que les conclusions établies pour chacun des objectifs qui nécessitent un tableau croisé. Les autres groupes et formateurs devraient être invités à formuler des commentaires.

**Formation à la recherche sur les systèmes de santé
Volume 2, 2^o partie : Analyse de données et rédaction de rapports**

Module 24

ANALYSE DES DONNÉES QUALITATIVES

ÉTAPES DE L'ANALYSE DE DONNÉES ET DE LA RÉDACTION DE RAPPORTS

Questions à poser	Étapes à suivre*	Éléments importants de l'étape
Quelles données ont été recueillies pour chaque objectif de recherche? Les données sont-elles complètes et exactes?	Préparer les données pour l'analyse	Revoir les activités sur le terrain Compiler les données pour chaque objectif ou population étudiée Classer les données et en vérifier la qualité Vérifier les imprimés
À quoi ressemblent les données?	Décrire les variables	Distributions Figures, moyennes
Comment peut-on résumer les données pour en faciliter l'analyse?	Faire le recouplement des données quantitatives	Faire le recouplement des données par rapport aux objectifs Illustrations, texte
	Résumer les données qualitatives	
Pour les données qualitatives : la recherche a-t-elle pour but de décrire, de comparer ou de trouver des associations?	Déterminer le type d'analyse statistique nécessaire	Examiner les objectifs, le type d'étude et les variables Faire une description statistique des variables Choisir les tests d'hypothèses
1. Comment peut-on décrire les données?		
2. Comment peut-on déterminer les différences entre les groupes?	Analyser des observations appariées et non appariées	**Test t de Student **Test t pour observations appariées **Test du khi carré **Test du khi carré de McNemar
3. Comment peut-on déterminer les associations entre les variables?	Évaluer les associations entre les variables	**Diagramme de dispersion **Droite de régression et coefficient de corrélation **Risque relatif, risque relatif approché
Comment le rapport doit être rédigé?	Rédiger le rapport et formuler des recommandations	Préparer le canevas du rapport Rédiger les 1 ^{re} et 2 ^e versions Faire un sommaire des constatations Faire un sommaire des conclusions pour chaque objectif Formuler des recommandations Préparer un résumé analytique
Comment doit-on présenter et diffuser les constatations et les recommandations?	Présenter un sommaire et rédiger un plan de mise en oeuvre des recommandations	Discuter des sommaires avec différents groupes cibles Discuter du programme de mise en oeuvre

* Il n'est pas nécessaire de faire ces étapes dans l'ordre indiqué. L'ordre peut être adapté aux besoins de l'atelier.

** Ces éléments sont facultatifs; ils peuvent être omis s'ils ne sont pas pertinents au cours.

Module 24 : ANALYSE DES DONNÉES QUALITATIVES

OBJECTIFS

À la fin de la présente leçon, vous devriez pouvoir :

1. **Décrire** des moyens efficaces d'ordonner et de résumer des données qualitatives;
2. **Indiquer** pourquoi il est essentiel de commencer à résumer et à analyser les données qualitatives au cours des travaux sur le terrain;
3. **Dresser** une liste des principales étapes de l'analyse des données qualitatives et tirer des conclusions;
4. **Rédiger un aperçu** de la méthode utilisée pour ordonner et résumer les données qualitatives et de l'analyse subséquente;
5. **Indiquer**, maintenant ou à la fin de l'analyse des données, les activités supplémentaires que vous ferez pour vérifier ou confirmer vos constatations afin de prouver leur validité.

I. Introduction

II. Procédés de traitement et de présentation des données qualitatives

III. Établissement et vérification des conclusions

IV. Stratégies de vérification des constatations visant à prouver leur validité

I. INTRODUCTION

Dans les modules précédents (9, 10, 13), nous avons souligné que les techniques de recherche qualitative servent à obtenir des indications sur des situations ou des problèmes sur lesquels nous disposons de peu de renseignements. Les techniques qualitatives telles que le recours aux questions ouvertes, les entrevues peu structurées, les discussions en groupe et les observations conviennent donc à de nombreuses études, surtout au début.

Dans les recherches sur les systèmes de santé, l'étape à laquelle sont rédigées les recommandations relatives à la mise en oeuvre des résultats de la recherche constitue une autre bonne occasion d'appliquer les techniques qualitatives.

Les techniques de recherche qualitative sont souvent utilisées pour décrire en détail des procédés, points de vue et renseignements relatifs aux questions touchant la santé parmi la population et les fournisseurs de soins. Ces techniques se prêtent également à l'étude des motifs d'un certain comportement ou de l'opinion des répondants sur des questions particulières. Quels que soient la méthode de collecte de données et ses motifs, le chercheur se retrouve habituellement avec bon nombre de pages de **texte** qu'il lui faudra analyser.

Les procédés et les résultats de l'analyse des données qualitatives diffèrent de ceux de l'analyse des données quantitatives, mais leurs principes sont semblables. Dans les deux cas, le chercheur doit :

- décrire l'échantillon choisi;
- classer et réduire ou coder les données (traitement des données);
- présenter des sommaires des données de façon à faciliter leur interprétation;
- tirer des conclusions;
- élaborer des stratégies de mise à l'épreuve ou de confirmation des constatations afin de prouver leur validité.

Voyons maintenant ces tâches en détail.

II. PROCÉDÉS DE TRAITEMENT ET DE PRÉSENTATION DES DONNÉES QUALITATIVES

Description de l'échantillon

Il est utile de commencer le traitement des données (ainsi que le compte rendu des constatations) par une description des répondants. Si ceux-ci sont en nombre suffisant, des données de base pertinentes peuvent être calculées, notamment l'âge, le sexe, l'occupation, la scolarité ou l'état civil, comme il est d'usage dans les études quantitatives.

Cependant, comme les données qualitatives proviennent de petits échantillons (qui se limitent parfois à quelques répondants ou à des discussions en groupe), il faut obtenir plus de renseignements pour remettre les données en contexte.

Par exemple, qui étaient les répondants, et comment ont-ils été choisis? Qui a participé aux discussions en groupe? À quel point les participants étaient-ils représentatifs des groupes qu'ils étaient censés représenter? Dans quelles circonstances les observations ont-elles été effectuées? Qui a fait l'objet d'observations (et qui n'en a pas fait l'objet)? Quelles ont été les réactions des personnes observées?

Sans ces renseignements, l'interprétation des données peut se révéler hasardeuse.

Classement et codage des données

Les modules précédents (**modules 10.B et 13**) soulignent que dès la fin de l'entrevue ou de la discussion en groupe, les notes prises (et les enregistrements, s'il y a lieu) doivent être organisées. Cette version des notes devrait refléter le mieux possible les discussions, et comprendre également vos observations et commentaires.

À l'analyse, on constate habituellement que les notes prises sur le terrain contiennent des données valables mais également des détails moins utiles et même superflus, même si la discussion a eu lieu selon des directives bien conçues. En outre, les données ne sont normalement pas présentées dans l'ordre où elles doivent l'être pour l'analyse, car les répondants peuvent sauter d'un sujet à l'autre.

Pour faciliter l'analyse, il faut **classer** et **réduire** les données. Le classement est normalement fondé sur les objectifs ou les sujets de la discussion.

Si les données sont limitées et simples, on peut se contenter d'identifier les données pertinentes qui ont trait à un même sujet en écrivant le numéro de ce sujet dans la marge. Si les données sont complexes et nombreuses, il peut être utile d'employer des codes pour le classement.

Contrairement aux données quantitatives, où les codes sont habituellement des chiffres, les codes employés pour les données qualitatives sont des **étiquettes** qu'il est facile de mémoriser.

Par exemple, dans la discussion sur le changement des habitudes de sevrage présentée au **module 10C**, nous pourrions coder les données de la façon suivante :

Typ al bébé -	Types d'aliments pour bébé
Fréq rep -	Fréquence des repas
Ag déb AM -	Âge où l'enfant commence à manger des aliments mous
Typ AM -	Types d'aliments mous
Fréq prép AM -	Fréquence de préparation d'aliments mous
Mod prép AM -	Mode de préparation des aliments mous
Rai AM tôt -	Raisons pour lesquelles l'enfant reçoit des aliments mous tôt, etc.

Les codes correspondent étroitement aux sujets présentés dans le guide de la discussion ou sur la liste de vérification des observations. Toutefois, chaque sujet peut comprendre plusieurs codes.

Remarques

Il est préférable d'effectuer le codage au crayon, car vous allez peut-être préciser ou modifier vos codes plus tard au cours de l'analyse. Inscrivez les codes dans la marge de gauche.

Dans la marge de droite de vos notes, vous pouvez inscrire des commentaires qui vous viennent à l'esprit à la lecture : conclusions, caractère incomplet des données, autres questions ou même sujets à ajouter.

Sommaires des données : tableaux et figures

Après avoir classé les données, il faut les résumer. La première étape consiste à dresser une liste des données qui vont ensemble. Si les données ont été codées, on dresse une liste de toutes les données qui ont reçu le même code.

La liste des données peut prendre une forme encore plus abrégée, de sorte qu'il est facile de répondre aux questions de la recherche.

Remarque

Au moment de dresser la liste des données, n'oubliez pas d'indiquer la source (numéro d'entrevue ou numéro de page des notes prises sur le terrain); il sera alors possible de revenir aux données dans leur contexte initial.

Il est possible de résumer encore plus les données de façon **graphique** dans un tableau (p. ex., une matrice) ou une figure (p. ex., un diagramme ou un organigramme). Cette méthode se révèle particulièrement utile pour interpréter une grande quantité de données.

Matrices

Une **MATRICE** est un tableau qui contient du texte au lieu de chiffres.

Lors de la discussion sur la modification des pratiques de sevrage, les chercheurs ont dressé une liste des réponses des jeunes mères concernant le début de l'alimentation avec des aliments mous et de celles des mères qui ont passé l'âge de procréer, en utilisant les codes décrits ci-dessus. Il ont ensuite résumé ces réponses au moyen d'une matrice.

Figure 24.1. Matrice sur le début de l'alimentation des enfants aux aliments mous chez les mères de différents groupes d'âge.*

GROUPES D'ÂGE	DÉBUT D'ALIMENTS MOUS	TYPE D'ALIMENT	FRÉQUENCE DES REPAS PAR JOUR
Jeunes mères (20-30)	intervalle : 4-7 mois moyenne : 6 mois	<ul style="list-style-type: none"> • gruau • gruau avec noix hachées • pommes de terre en purée, fruits hachés, biscuits trempés 	Une ou deux fois par jour <ul style="list-style-type: none"> • selon la disponibilité de la mère et du gardien • selon l'appétit de l'enfant
Mères ayant passé l'âge de procréer (>45)	intervalle : 5-11 mois moyenne : 8,5 mois	<ul style="list-style-type: none"> • gruau • fruits mous 	Une ou deux fois par jour <ul style="list-style-type: none"> • selon la disponibilité de la mère et du gardien • selon l'appétit de l'enfant

* Le plus jeune enfant des jeunes mères et l'enfant le plus âgé des mères ayant passé l'âge de procréer.

Grâce à cette matrice, les chercheurs peuvent conclure facilement que :

- les jeunes mères commencent à donner des aliments mous à leur enfant en moyenne 2,5 mois plus tôt que ne l'ont fait leur propre mère;
- les jeunes mères utilisent des aliments mous de sevrage plus variés que ceux qu'utilisait leur propre mère;
- les mères des deux générations donnent des aliments mous à leurs enfants à une fréquence semblable.

En d'autres mots, le seul fait de disposer les données dans un tableau permet de les analyser.

Les matrices constituent la forme la plus courante de présentation graphique des données qualitatives. On peut les utiliser pour classer les données de plusieurs façons, notamment selon :

- l'ordre (des procédés étudiés, par exemple),
- le type de répondants (comme dans l'exemple ci-dessus),
- l'endroit où les données ont été recueillies,
- le type d'activité,
- les motifs d'un certain comportement.

Les chercheurs peuvent également inscrire leurs propres commentaires dans une matrice (par exemple, sur la qualité du travail des fournisseurs de soins).

Diagrammes

Un DIAGRAMME est une figure composée de cases ou de cercles contenant des variables, ainsi que de flèches illustrant les relations entre les variables.

La discussion en groupe sur la modification des pratiques de sevrage pourrait permettre de recueillir des renseignements sur les motifs poussant les mères à sevrer leur enfant tôt ou tard. Ces renseignements peuvent se résumer de manière satisfaisante au moyen d'un diagramme (figures 24.2 et 24.4).

Figure 24.2. Motifs poussant les jeunes mères à donner des aliments mous tôt à leurs enfants.

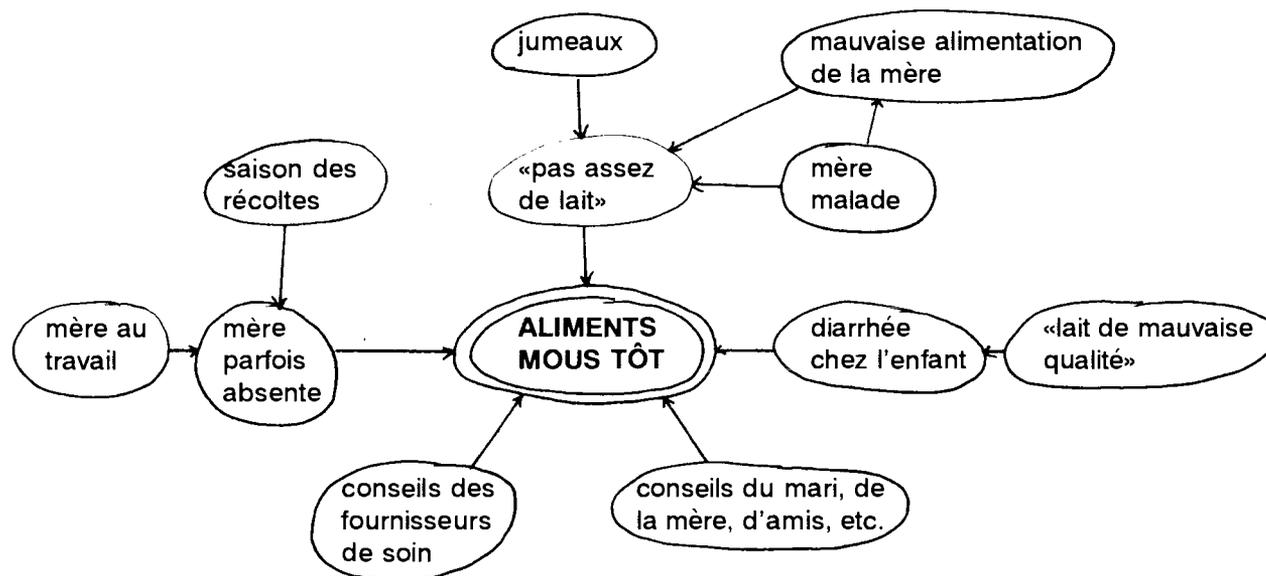
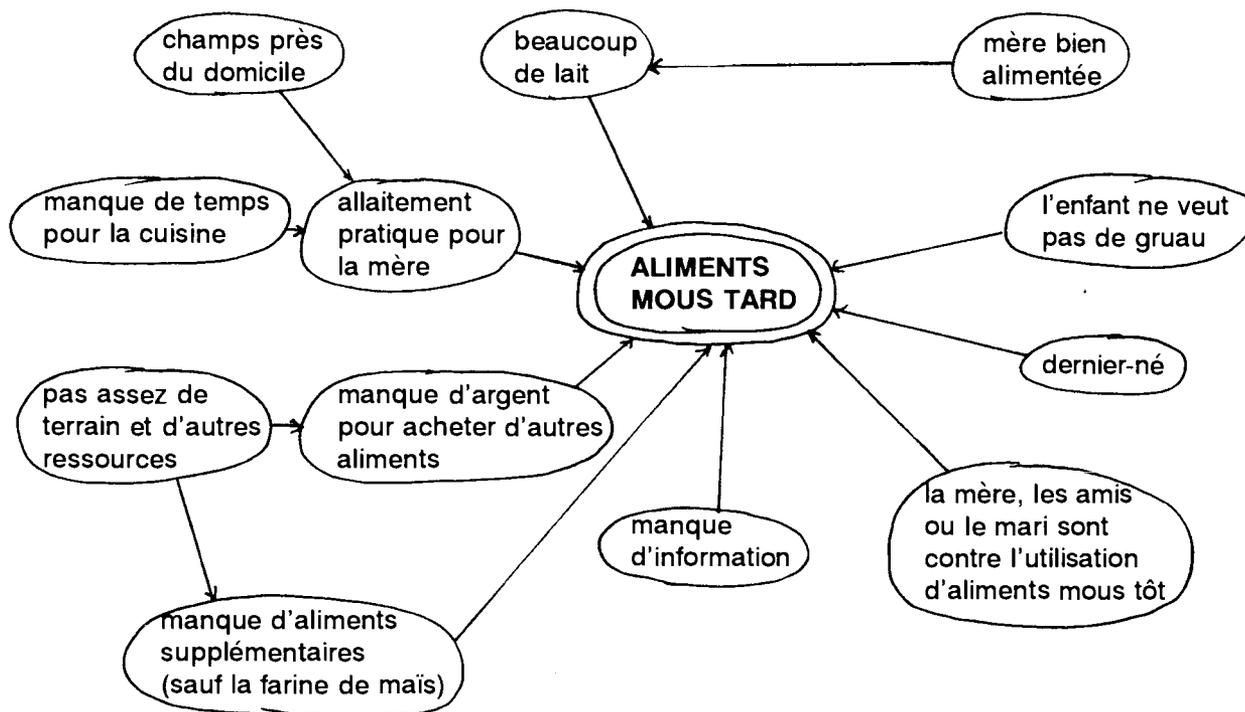


Figure 24.3. Motifs poussant les jeunes mères à donner des aliments mous tard à leurs enfants.



Ces diagrammes peuvent être préparés après une discussion en groupe, par exemple, pour les jeunes mères aussi bien que pour les mères qui ne sont plus en âge d'avoir des enfants. Il sera utile d'étudier le problème encore plus attentivement dans d'autres discussions. **Par exemple**, qu'est-ce qui fait que la mère, les amis ou le mari soit ou non en faveur de l'utilisation d'aliments mous tôt? Comment se fait-il que certaines mères savent quels aliments mous utiliser et comment les préparer et que d'autres ne le savent pas? Quelle serait la meilleure façon de leur fournir ces renseignements?

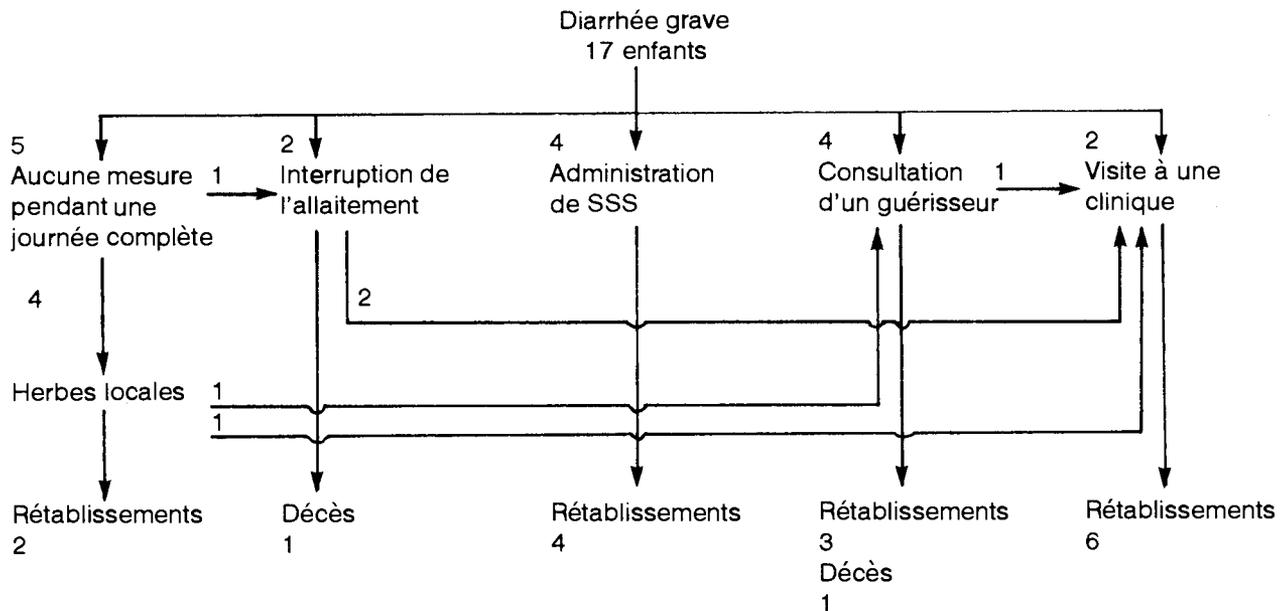
Il est possible que les diagrammes soient plus complexes que les précédents, surtout si vous y inscrivez les attitudes et les opinions des répondants. Vous pouvez modifier et élargir les diagrammes à mesure que vous obtenez de nouvelles données ou procédez à l'analyse des données recueillies. Comme les matrices, les diagrammes sont très utiles pour obtenir une vue d'ensemble des données et orienter leur analyse.

Organigrammes

Les ORGANIGRAMMES sont des diagrammes qui illustrent l'ordre logique d'actions ou de décisions.

La figure du **module 21** indiquant les différentes étapes de l'analyse de données constitue un exemple d'organigramme. En voici un autre, qui résume les mesures prises par les mères dont les enfants ont la diarrhée.

Figure 24.4. Mesures prises par 17 mères dont les enfants allaités naturellement (âgés de 6 à 18 mois) étaient atteints de diarrhée grave.



Cet organigramme révèle que les 17 mères interrogées ont pris des mesures différentes. Il est remarquable de constater que les enfants qui ont reçu une solution de sel et de sucre (quatre au total) se sont tous rétablis sans autre traitement. Trois mères ont cessé d'allaiter leur enfant, ce qui a causé un décès. Les deux autres enfants se sont rétablis après avoir été amenés à une clinique, etc.

Tableaux

Un TABLEAU est un graphique composé de rangées et de colonnes dont les cellules contiennent des données.

Parfois, les données qualitatives peuvent être divisées en catégories, comptées et disposées dans des tableaux. Les réponses à des questions ouvertes de questionnaires peuvent souvent être divisées en catégories et résumées de cette façon. Il est également possible d'analyser les réponses en tant que telles (voir le module 21).

Texte

Le texte constitue un élément important de la présentation des données qualitatives. Nous avons commencé par la présentation graphique dans des tableaux et figures parce que les chercheurs ont tendance à s'éparpiller et à se perdre dans les détails s'ils commencent par la rédaction. Les méthodes de présentation graphique telles que les matrices, les diagrammes, les organigrammes et les tableaux aident le chercheur à rester sur la bonne voie et à demeurer concis. On peut les présenter dans le texte lui-même ou en annexe.

III. ÉTABLISSEMENT ET VÉRIFICATION DE CONCLUSIONS

L'établissement et la vérification de conclusions constituent le fondement de l'analyse de données. Cependant, il ne s'agit pas d'une activité isolée. Le chercheur qui résume ses données tire continuellement des conclusions, et en modifie ou en rejette plusieurs par la suite. La rédaction permet également de trouver de nouvelles idées. Par conséquent, **la rédaction doit être entreprise le plus tôt possible**, dès le début du traitement et de l'analyse des données.

Remarque

La collecte, le traitement, l'analyse et la présentation des données qualitatives sont intimement liés, et ne constituent pas (comme dans le cas des données quantitatives) des étapes distinctes et successives. Pour cette raison, il devrait toujours être possible (et il est souvent essentiel) de revenir aux notes et de retourner sur le terrain pour vérifier ses conclusions, recueillir des données supplémentaires et obtenir les commentaires de tous les intervenants.

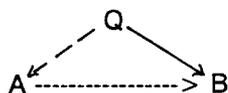
Cela dit, récapitulons brièvement les différentes étapes du traitement et de l'analyse des données qui peuvent déboucher sur des conclusions justifiées :

1. **Classement** des données selon les objectifs ou les questions de la recherche.
2. **Division en catégories** ou codage des réponses qui présentent des caractéristiques ou des tendances semblables.
3. **Présentation d'un résumé des données** dans des tableaux et figures tels que des matrices et diagrammes afin de **visualiser les relations possibles entre certaines variables**. Soulignons que les matrices sont particulièrement utiles pour découvrir ces relations.
4. **Identification des variables et des associations entre variables**. Le chercheur qui suit une méthode qualitative doit jouer le rôle d'un détective qui cherche des preuves, tient compte des éléments contradictoires et vérifie ses constatations en trouvant des preuves indépendantes jusqu'à ce qu'il ait établi des associations possibles et des liens causals entre certaines variables.

Par exemple, si l'on découvre que parmi les mères qui sèvrèrent leurs enfants tôt, bon nombre ont un emploi, on peut supposer que le fait d'occuper un emploi est un facteur qui est lié au sevrage anticipé. Des études semblables effectuées ailleurs, qui ont mené à des constatations semblables, soutiennent cette hypothèse (preuves indépendantes). Cependant, pour confirmer la véracité de cette hypothèse, il faudra qu'un très petit nombre de femmes sevrant leurs enfants tard occupent un emploi, et qu'une explication soit apportée à chacune de ces exceptions. Les mères emmènent-elles leurs enfants avec elles (garderie au travail) ou travaillent-elles près de chez elles, ce qui leur permet de donner à manger à leurs enfants pendant les pauses? Combinent-elles l'allaitement avec d'autres aliments? Dans l'affirmative, pourquoi plus de femmes ne font pas la même chose?

5. **Découverte de variables confusionnelles ou intervenantes**

Parfois, des variables semblent reliées entre elles, mais il est difficile d'expliquer leur association. Il arrive également que des variables semblent logiquement liées, mais qu'il soit impossible de déterminer la relation. Il est alors possible qu'une autre variable, la variable Q, influe sur l'association entre les deux variables concernées et doive être identifiée.



Par exemple, selon un programme de contrôle de la malaria, les villages où les maisons ont subi une vaporisation pendant l'après-midi présentent une incidence plus élevée de la malaria que les villages vaporisés en matinée (voir l'exemple 2 de l'exercice 4a, au **module 9**). Les responsables du programme n'ont pu expliquer cette association avant de faire une observation sur place, qui leur a permis de constater que les vaporisateurs utilisaient la plus grande partie de leur DDT pendant la matinée de façon à réduire leur charge pour l'après-midi.

Mentionnons également comme **deuxième exemple** la relation entre la qualité de l'eau potable et l'incidence de la diarrhée. On peut s'attendre à ce que cette incidence soit inversement proportionnelle au nombre de robinets d'un village. S'il n'y a pas de changement avec le temps, il pourrait y avoir une variable intervenante. Par exemple, il est possible que les gens détestent tellement le goût de l'eau du robinet qu'ils n'en boivent jamais, même s'ils l'utilisent pour d'autres tâches.

Remarque

De telles associations inexplicables peuvent survenir dans n'importe quelle étude. La méthode de recherche qualitative se distingue essentiellement par le fait qu'elle recherche ces associations pendant le travail sur le terrain, par la collecte et l'analyse simultanées des données. Dans les enquêtes strictement quantitatives, ce n'est généralement qu'après qu'on découvre les associations et les écarts dans les données et qu'on peut les expliquer.

6. Recherche de liens logiques

Lorsqu'un chercheur tente d'établir des associations entre toutes les variables (reliées) qui donnent un certain résultat, il veut découvrir des liens logiques. Il est particulièrement intéressant d'analyser une entreprise réussie en examinant les différentes conditions qui doivent être remplies avant d'obtenir le résultat positif. **Les recherches sur les systèmes de santé s'intéressent tellement à la résolution de problèmes qu'il est facile d'oublier que l'analyse de réussites peut nous permettre d'en apprendre beaucoup sur les causes des échecs.**

Au cours de la période limitée dont nous disposons pour la recherche, il ne sera peut-être pas possible d'établir de bons liens logiques pour étayer un résultat observé. Il est encore moins probable de pouvoir tirer de nos réseaux causaux et liens logiques des théories générales. Néanmoins, pour démontrer la validité de nos constatations, nous devons appliquer les mêmes stratégies d'essai et de confirmation que les chercheurs qui ne franchissent pas cette dernière étape.

IV. STRATÉGIES DE VÉRIFICATION DES CONSTATATIONS VISANT À EN PROUVER LA VALIDITÉ

Les chercheurs qui utilisent des modèles de recherche quantitative réduisent leurs données en chiffres et font des tests statistiques. Cela ne leur permet pas nécessairement de garantir la validité de leurs résultats; une erreur peut être survenue pendant l'échantillonnage ou la collecte des données. Par conséquent, les stratégies suivantes devraient être employées par tous les chercheurs. Cependant, elles s'appliquent particulièrement à la recherche qualitative, car la plupart servent à donner une certaine cohérence aux données recueillies.

1. **Vérifier la représentativité des données**

Bien que dans les recherches qualitatives les répondants ne fassent généralement pas l'objet d'une sélection aléatoire, ils doivent avoir été choisis systématiquement selon des règles préétablies (voir le **module 11**). Vérifiez si vous avez vraiment interrogé des répondants de toutes les catégories nécessaires pour obtenir une bonne vue d'ensemble du sujet (sans trop dépendre d'experts). Assurez-vous de ne pas généraliser de facteurs non représentatifs.

2. **Vérifier la présence d'un biais** attribuable au parti pris de l'observateur ou à l'influence du chercheur sur la recherche. Nous avons discuté de cette question en détail au **module 10A**.

3. **Vérifier les données par rapport à d'autres sources indépendantes**

Ces sources peuvent être des répondants différents, le recours à des techniques de recherche différentes pour enquêter sur le même sujet ou les résultats d'autres études semblables (voir les **modules 5, 9 et 10**). Les données devraient se confirmer ou, à tout le moins, ne pas se contredire.

Une telle vérification systématique des données en cherchant à les confirmer ou à les démentir au moyen de sources indépendantes constitue pour le chercheur l'un des principaux moyens qui lui permettent d'améliorer la validité de ses données.

Par exemple, les réponses des maris et des épouses (et des autres répondants concernés) devraient concorder sur certaines questions, notamment lequel des conjoints décide d'utiliser un moyen de contraception et choisit ce moyen, lequel décide si les filles doivent subir une clitoridectomie ou lequel gère l'argent de la famille.

4. **Comparer les données**

On peut faire de la comparaison une partie intégrante de la recherche en choisissant différentes catégories de répondants.

Par exemple, pour s'assurer que la variable A (scolarité élevée) influe sur la variable B (recours à des méthodes de contraception), il faut comparer un groupe de mères de scolarité élevée à un groupe de mères de faible scolarité. On s'attend alors à une utilisation moins fréquente des méthodes de contraception dans le second groupe.

Il est important de comparer les données lorsqu'il s'agit d'**identifier** les variables et de confirmer les associations entre elles.

5. **Recourir le plus possible à des groupes de répondants opposés**

Au sujet de la conception de l'étude et de l'échantillonnage, nous avons affirmé qu'il peut être utile de constituer des catégories de répondants qui représentent les extrêmes d'une variable particulière.

Par exemple, il serait très utile d'étudier les personnes qui sont souvent absentes et celles qui participent régulièrement aux services de lutte contre la tuberculose, et non les personnes qui sont absentes de temps à autre. Cela peut constituer le moyen le plus rapide de déterminer les principales variables qui influent sur le comportement des tuberculeux.

6. **Effectuer des recherches supplémentaires pour mettre à l'épreuve les constatations établies**

Vous n'aurez peut-être pas beaucoup de temps entre les deux ateliers pour faire des recherches supplémentaires, mais il est possible que vous trouviez les résultats de votre étude si intéressants que vous décidiez de faire une étude de suivi. Une telle étude peut être effectuée pour :

- reproduire certaines constatations;
- exclure (ou identifier) des variables intervenantes éventuelles;

- exclure des explications contradictoires en les étudiant;
- trouver des preuves contradictoires.

Des études supplémentaires entreprises pour l'une ou plusieurs de ces raisons pourraient rendre plus convaincants les résultats de votre recherche initiale.

7. **Obtenir des commentaires des répondants**

Dans les **modules 1 à 20**, nous avons souligné qu'il vous faut assurer la participation de toutes les parties concernées aux différentes étapes de la recherche. C'est important non seulement pour des raisons déontologiques et parce qu'il sera plus probable que les résultats seront mis en oeuvre, mais également parce que **la qualité de votre proposition, de vos données et de vos conclusions** s'en trouvera améliorée. Les conseils et les renseignements supplémentaires que vous obtiendrez lors des séances de discussion ne manqueront pas d'améliorer la qualité de votre rapport de recherche.

TRAVAIL EN GROUPE (Le temps nécessaire repose sur la quantité des données qualitatives recueillies)

- Vérifiez si vous avez dressé une liste de toutes les sources de données qualitatives pour chaque objectif lorsque vous avez fait un inventaire de vos données, à la séance de travail en groupe du **module 21**.
- Si votre étude porte notamment sur des groupes de répondants qui vous ont fourni uniquement des données qualitatives (p. ex., discussions en groupe ou principaux répondants), décrivez votre échantillon.
- Si vos données qualitatives s'étendent sur plusieurs pages, codez-les et attribuez des énoncés ou des observations à chaque code.

Sinon, dressez une liste des données correspondant à chaque objectif.

- Déterminez si vous devez utiliser des matrices, des diagrammes ou du texte pour résumer vos données.
- Tirez des conclusions après avoir vérifié vos constatations. Dites comment vous aimeriez mettre à l'épreuve certaines de vos conclusions, s'il y a lieu.

Ouvrages consultés

Miles, M.B., Huberman, A.M. 1984. Qualitative data analysis, a sourcebook of new methods. Sage Publications, Beverly Hills, CA, États-Unis.

Spradly, J.P. 1979. The ethnographic interview. Holt, Rinehart and Winston, New York, NY, États-Unis.

Walker, R. (éd). 1985. Applied qualitative research. Gower Publishing Company Ltd, Hants, Royaume-Uni.

Yin, R.K. 1984. Case study research: design and methods. Sage Publications, Beverly Hills, CA, États-Unis.

Nota : Le présent module est fondé principalement sur Miles et Huberman. Les sections III et IV constituent une version fortement abrégée et adaptée du chapitre VII de cet ouvrage.

Module 24 : ANALYSE DES DONNÉES QUALITATIVES

Durée et méthode d'enseignement

1 heure Introduction et discussion

1 heure* Travail en groupe

(La durée repose sur la quantité et le type de données qualitatives)

Introduction et discussion

- Si aucun des groupes ne dispose de données qualitatives à part quelques questions ouvertes faisant partie de questionnaires, vous pourriez vous concentrer sur les sections I et II afin de donner aux participants un aperçu du traitement des données qualitatives, et n'aborder que brièvement les sections III et IV.
- Toutefois, si les participants ont de l'expérience ou de la formation en recherche ou s'y intéressent, vous pouvez étudier la totalité des sections III et IV, même s'ils n'ont pas recueilli beaucoup de données qualitatives. Les procédures présentées pour tirer des conclusions et vérifier la validité s'appliquent à tous les types de recherche, et il se peut que les méthodes de vérification et de contre-vérification des données ne soient pas familières à tous les participants.
- En vous fondant sur le diagramme d'analyse que les groupes ont préparé au cours de la rédaction de leur proposition de recherche et sur l'organigramme précédant les **modules 1 à 18 et 22 à 32**, présentez d'autres exemples de tableaux ou de graphiques au moyen de transparents ou d'un tableau à feuilles mobiles.
- Si un ou plusieurs groupes ont fait beaucoup de recherche qualitative, étudiez le module en détail en utilisant des exemples tirés de leurs études. Il est probable qu'aucun participant ne se sera familiarisé avec l'analyse des données qualitatives.
- Laissez les groupes qui ont fait de la recherche qualitative décrire en séance plénière comment ils ont analysé les données recueillies lors de discussions en groupe, d'observations ou d'entrevues avec les principaux répondants. Demandez-leur de préciser les questions qu'ils ont ajoutées ou celles qu'ils ont abandonnées pendant le déroulement des entrevues, et pourquoi.

Travail en groupe

- **Tous les groupes :**

Vérifiez si leurs questions ouvertes nécessitent une analyse des réponses. Certaines questions portant sur des opinions pourraient constituer des illustrations intéressantes pour les rapports.

Prenez-en note, car les groupes pourraient oublier ces données lorsqu'il travailleront aux tableaux et aux statistiques.

Déterminez par une discussion s'il suffit de dresser une liste des énoncés pour l'analyse du contenu ou si une présentation graphique des données serait souhaitable.

- **Groupes qui disposent de données qualitatives en plus de réponses à des questions ouvertes :**

Étudiez toutes les données dont ils disposent et demandez-leur de suivre les directives du travail en groupe.

**Formation à la recherche sur les systèmes de santé
Volume 2, 2^e partie : Analyse de données et rédaction de rapports**

Module 25

**DESCRIPTION DES VARIABLES
DEUXIÈME PARTIE**

ÉTAPES DE L'ANALYSE DE DONNÉES ET DE LA RÉDACTION DE RAPPORTS

Questions à poser	Étapes à suivre*	Éléments importants de l'étape
Quelles données ont été recueillies pour chaque objectif de recherche? Les données sont-elles complètes et exactes?	Préparer les données pour l'analyse	Revoir les activités sur le terrain Compiler les données pour chaque objectif ou population étudiée Classer les données et en vérifier la qualité Vérifier les imprimés
À quoi ressemblent les données?	Décrire les variables	Distributions Figures, moyennes
Comment peut-on résumer les données pour en faciliter l'analyse?	Faire le recouplement des données quantitatives	Faire le recouplement des données par rapport aux objectifs Illustrations, texte
	Résumer les données qualitatives	
Pour les données qualitatives : la recherche a-t-elle pour but de décrire, de comparer ou de trouver des associations?	Déterminer le type d'analyse statistique nécessaire	Examiner les objectifs, le type d'étude et les variables Faire une description statistique des variables Choisir les tests d'hypothèses
1. Comment peut-on décrire les données?		
2. Comment peut-on déterminer les différences entre les groupes?	Analyser des observations appariées et non appariées	**Test t de Student **Test t pour observations appariées **Test du khi carré **Test du khi carré de McNemar
3. Comment peut-on déterminer les associations entre les variables?	Évaluer les associations entre les variables	**Diagramme de dispersion **Droite de régression et coefficient de corrélation **Risque relatif, risque relatif approché
Comment le rapport doit être rédigé?	Rédiger le rapport et formuler des recommandations	Préparer le canevas du rapport Rédiger les 1 ^{re} et 2 ^e versions Faire un sommaire des constatations Faire un sommaire des conclusions pour chaque objectif Formuler des recommandations Préparer un résumé analytique
Comment doit-on présenter et diffuser les constatations et les recommandations?	Présenter un sommaire et rédiger un plan de mise en oeuvre des recommandations	Discuter des sommaires avec différents groupes cibles Discuter du programme de mise en oeuvre

* Il n'est pas nécessaire de faire ces étapes dans l'ordre indiqué. L'ordre peut être adapté aux besoins de l'atelier.
** Ces éléments sont facultatifs; ils peuvent être omis s'ils ne sont pas pertinents au cours.

Module 25 : DESCRIPTION DES VARIABLES, DEUXIÈME PARTIE

OBJECTIFS

À la fin de la présente séance, vous devriez pouvoir :

1. Expliquer ce qu'est l'intervalle, le centile, l'écart-type, une distribution normale, l'erreur-type et un intervalle de confiance de 95 %;
2. Calculer à partir de vos propres données des intervalles, des écarts-types, des erreurs-types et des intervalles de confiance de 95 % au besoin;
3. Interpréter les résultats de ces calculs.

I. Introduction

II. Mesures de dispersion

III. Distribution normale

IV. Relation entre l'échantillon et la population

V. Représentativité de l'échantillon par rapport à la population

I. INTRODUCTION

Comme nous l'avons déjà vu, la moyenne, la médiane et le mode sont des mesures de la tendance centrale d'une variable, mais ils ne permettent pas de déterminer à quel point les mesures varient. Le présent module décrit certaines mesures de variation courantes, que l'on décrit souvent dans les manuels de statistique comme des mesures de dispersion. En outre, les notions de distribution normale, d'erreur-type et d'intervalle de confiance sont présentées. Elles nous seront utiles au moment des tests statistiques.

II. MESURES DE DISPERSION

Étendue

L'ÉTENDUE d'une série de mesures est la différence entre ses valeurs extrêmes.

Par exemple, si le poids de sept femmes enceintes était de 40, 41, 42, 43, 44, 47 et 72 kg, l'étendue serait de $72 - 40 = 32$ kg.

Bien que facile à calculer, l'étendue ne donne aucune indication sur la distribution des valeurs entre les deux extrêmes.

Si le poids de sept autres femmes enceintes était de 40, 46, 46, 46, 50, 60 et 72 kg, l'étendue serait également de $72 - 40 = 32$ kg, même si les valeurs sont très différentes de celles du premier exemple.

Centiles

Une autre façon de décrire la dispersion d'une série de mesures consiste à diviser la distribution en centiles. Le centile ne constitue en fait que le prolongement de la notion de la médiane, qui constitue également le 50^e centile.

Les CENTILES sont des points qui divisent une série de mesures en 100 parties égales.

Le 3^e centile (C_3) est la valeur sous laquelle se trouvent 3 % des mesures.

Le 50^e centile (C_{50}) est la valeur sous laquelle se trouvent 50 % des mesures.

La notion des centiles est utilisée par les nutritionnistes pour élaborer des tableaux de croissance normale pour différents pays.

Écart-type

Pour déterminer à quel point les mesures s'écartent de la valeur moyenne afin d'effectuer certains tests statistiques, il faut calculer une valeur qu'on appelle l'écart-type.

L'ÉCART-TYPE constitue la différence moyenne entre chaque mesure et la moyenne.

Pour obtenir l'écart-type d'une série de mesures, il faut suivre les étapes suivantes :

1. Calculer la moyenne de toutes les mesures;
2. Calculer la différence entre chaque mesure et la moyenne;
3. Mettre ces différences au carré;
4. Faire la somme de ces différences au carré;
5. Diviser cette somme par le nombre de mesures moins un;
6. Enfin (comme les différences ont été mises au carré), extraire la racine carrée de la valeur obtenue (afin de revenir à la même unité de mesure).

Exemple 1

Onze enfants de 3 ans sont pesés. Voici leur poids :

13, 14, 14, 15, 16, 16, 16, 17, 17, 18 et 20 kg

Le nombre de mesures (n) est de 11.

Pour calculer l'écart-type :

1. Calculer d'abord la moyenne : 16 kg.
2. Calculer ensuite la différence entre chaque mesure et la moyenne :

3, 2, 2, 1, 0, 0, 0, 1, 1, 2, 4 (c.-à-d. $16 - 13 = 3$; $16 - 14 = 2$; etc.)

3. Mettre ces valeurs au carré :

9, 4, 4, 1, 0, 0, 0, 1, 1, 4, 16 (c.-à-d. $3 \times 3 = 9$; $2 \times 2 = 4$; etc.)

4. La somme de ces différences au carré est de 40 ($9 + 4 + \dots$).
5. Diviser cette somme par le nombre de mesures moins un ($n - 1 = 10$):

$$40 / 10 = 4$$

6. Enfin, on fait la racine carrée pour obtenir l'écart-type :

$$\sqrt{4} = 2 \text{ kg}$$

Un grand écart-type signifie que les mesures sont fortement dispersées de chaque côté de la moyenne, tandis qu'un petit écart-type montre que les valeurs se situent près de la moyenne et présentent peu de variations.

Par exemple, si les poids mesurés à l'exemple 1 étaient de

10, 11, 12, 14, 16, 16, 18, 20, 21 et 22 kg

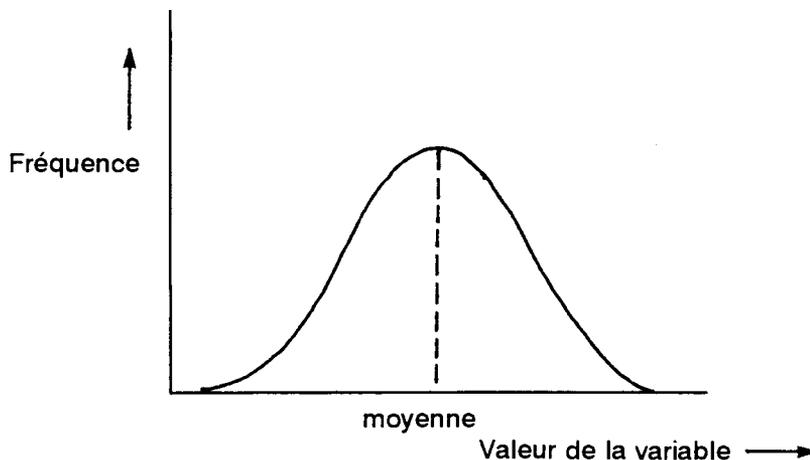
le poids moyen serait toujours de 16 kg, mais l'écart-type serait de 4 kg, ce qui représente une variation beaucoup plus grande des observations.

Une autre façon de calculer l'écart-type consiste à utiliser la formule figurant à l'annexe 25.1. Heureusement, bon nombre de calculatrices peuvent effectuer facilement ce calcul, mais il est quand même important de comprendre ce qu'il signifie.

III. DISTRIBUTION NORMALE

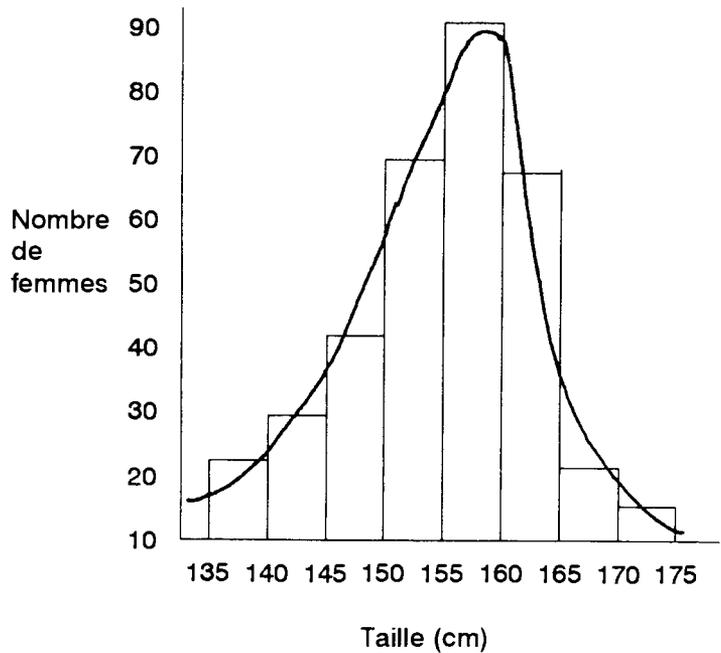
Bon nombre de variables présentent une **distribution normale**. Cette distribution s'illustre par une courbe en forme de cloche, où la plupart des valeurs sont regroupées près de la moyenne et quelques-unes près des extrémités. La distribution normale est symétrique de chaque côté de la moyenne. La moyenne, la médiane et le mode d'une distribution normale sont égaux.

Figure 25.1 Courbe de distribution normale.



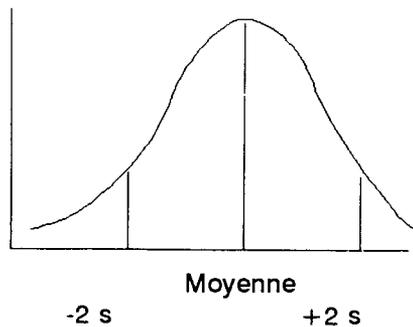
La figure suivante constitue un histogramme de la taille des femmes enceintes qui ont fréquenté une clinique prénatale, avec la courbe normale tracée par-dessus.

Figure 25.2. Taille des femmes enceintes ayant fréquenté une clinique prénatale.



Les variables à distribution normale présentent une importante caractéristique : 95 % des mesures ont une valeur qui s'écarte de la moyenne d'environ deux fois l'écart-type(s).¹ Cette particularité est illustrée à la figure ci-dessous.

Figure 25.3. Variable à distribution normale.



Exemple 2

Si la taille moyenne d'un groupe de 120 femmes est de 158 cm et si l'écart-type est de 3 cm, 95 % des femmes mesurent de 152 à 164 cm (si l'on suppose que la taille suit le modèle normal). En d'autres mots, 2,5 % des femmes (c'est-à-dire 3 femmes) ont une taille inférieure à 152 cm et 2,5 % ont une taille supérieure à 164 cm.

¹ Pour être plus précis, disons que 95 % des mesures ont une valeur égale à la moyenne $\pm 1,96$ écart-type.

Bon nombre de tests statistiques exigent que les variables suivent le modèle normal. Par conséquent, il est important de vérifier la distribution de vos variables pour déterminer lesquelles ne sont pas conformes au modèle normal. Parmi les exemples fournis dans les modules jusqu'à maintenant, plusieurs ne suivent pas la distribution normale (p. ex., voir l'exemple 4 du module 22).

IV. RELATION ENTRE L'ÉCHANTILLON ET LA POPULATION

Au début d'une étude, il est habituellement nécessaire de choisir un échantillon parmi la population étudiée. La population sera donc décrite à partir des renseignements recueillis auprès de l'échantillon. En d'autres mots, vous tenterez de généraliser les conclusions tirées de l'échantillon. De toute évidence, il n'est possible de le faire que si l'échantillon est choisi de manière à être représentatif de la population.

Toute valeur d'une variable obtenue de l'échantillon (p. ex., la moyenne) peut donc être considérée comme une estimation de la valeur correspondante pour la population.

Par exemple, si la taille moyenne des 120 femmes faisant partie de l'échantillon est de 158 cm, on peut supposer que cette valeur constitue une bonne estimation de la taille moyenne de l'ensemble des femmes faisant partie de la population étudiée.

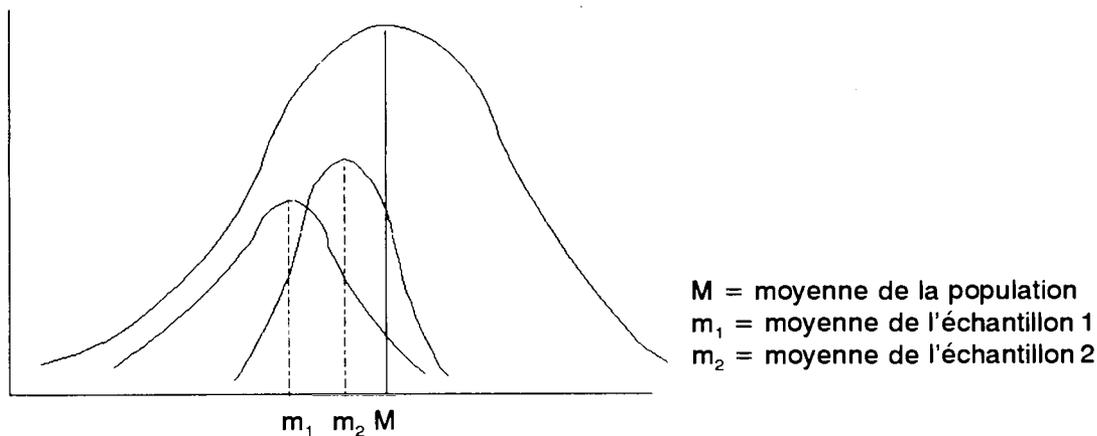
Cependant, il est peu probable que la moyenne de l'échantillon soit identique à la moyenne de la population.

Si vous choisissez un autre échantillon de 120 mères, il est possible que la moyenne soit de 157 cm, donc différente de la moyenne du premier échantillon. Cette valeur diffère probablement de la véritable taille moyenne de la population totale dans laquelle a été choisi l'échantillon.

Ce phénomène représente la **variation de l'échantillonnage**.

La figure suivante montre une courbe de la distribution de la population, avec les courbes de deux échantillons différents.

Figure 25.4. Courbe de distribution de la population et deux courbes de distribution d'échantillons pour une variable conforme au modèle normal.



Soulignons qu'un échantillon représentatif présentera une distribution de forme semblable à celle de la population, qui pourra cependant se trouver n'importe où sous elle.

V. REPRÉSENTATIVITÉ DE L'ÉCHANTILLON PAR RAPPORT À LA POPULATION

Pour découvrir la mesure dans laquelle une valeur de l'échantillon s'écarte de la valeur s'appliquant à la population, il est possible de placer la valeur de l'échantillon dans un intervalle qui, selon toute probabilité, contiendra également la valeur de la population.

Cet intervalle s'appelle **L'INTERVALLE DE CONFIANCE**.

L'INTERVALLE DE CONFIANCE constitue l'ensemble de données qui est le plus susceptible de comprendre la valeur de la population. Les deux extrémités de cet intervalle sont les **limites de confiance**.

Exemple

Un intervalle de confiance de 95 % de 152 à 164 cm pour la taille moyenne d'une population de femmes signifie qu'il est probable à 95 % que la moyenne de la population réelle, qu'il est impossible de déterminer exactement à moins de mesurer la taille de toutes les femmes, se situe entre 152 et 164 cm (152 et 164 cm constituent les limites de confiance).

Le calcul de l'intervalle de confiance tient compte de l'**erreur-type**. L'erreur-type donne une estimation de la mesure dans laquelle la moyenne de l'échantillon varie par rapport à la moyenne de la population. On la calcule à partir de l'écart-type.

Nous verrons maintenant comment calculer :

- l'erreur-type et l'intervalle de confiance d'une moyenne pour une sécurité de 95 % (données numériques),
- l'erreur-type et l'intervalle de confiance d'un pourcentage pour une sécurité de 95 % (données catégoriques).

Erreur-type et intervalle de confiance d'une moyenne pour une sécurité de 95 %

En présence de données numériques, il est possible d'évaluer à quel point la moyenne de l'échantillon diffère de la moyenne de la population.

L'erreur-type pour la moyenne se calcule en divisant l'écart-type par la racine carrée de la taille de l'échantillon :

$$\text{écart-type} / \sqrt{\text{taille de l'échantillon}} \text{ ou } s / \sqrt{n}$$

S'il s'agit d'une variable conforme au modèle normal, on peut supposer qu'environ 95 % de la moyenne de tous les échantillons possibles se trouvent dans un intervalle de deux fois l'écart-type de chaque côté de la moyenne. En d'autres mots, il est probable à 95 % que la moyenne de la population, que l'on veut estimer le plus précisément possible, se trouve dans un intervalle de deux fois l'écart-type autour de la moyenne de l'échantillon.

Pour faire une description statistique des variables, on présente habituellement la moyenne de l'échantillon plus ou moins deux fois l'écart-type. C'est ce qu'on appelle l'**intervalle de confiance pour une sécurité de 95 %**. Cela signifie que le chercheur est certain à environ 95 % que la moyenne de la population se trouve dans cet intervalle.

Dans l'**exemple 1**, le poids d'un échantillon aléatoire de 11 enfants de trois ans a été calculé. La moyenne de l'échantillon était de 16 kg et l'écart-type de 2 kg.

L'erreur-type est égale à

$$2 / \sqrt{11} = 0,6 \text{ kg}$$

L'intervalle de confiance pour une sécurité de 95 % est égal à

$$16 \pm (2 \times 0,6) = 14,8 \text{ à } 17,2 \text{ kg}$$

Cela signifie qu'il est probable à 95 % que le poids moyen de tous les enfants de trois ans de la population se situe entre 14,8 et 17,2 kg.

Soulignons que plus l'échantillon est grand, plus l'erreur-type est faible et plus l'intervalle de confiance est étroit. L'avantage d'étudier un échantillon important réside dans le fait que la moyenne de l'échantillon constituera une meilleure estimation de la moyenne de la population.

Dans l'exemple précédent, si la taille de l'échantillon était de 20 (au lieu de 11), l'erreur-type serait égale à

$$2 / \sqrt{20} = 0,45 \text{ kg}$$

et l'intervalle de confiance du poids moyen pour une sécurité de 95 % serait de 15,1 à 16,9 kg.

Erreur-type et intervalle de confiance d'un pourcentage pour une sécurité de 95 %

Dans la section précédente, nous avons calculé l'erreur-type et l'intervalle de confiance de la moyenne d'un échantillon pour une sécurité de 95 % à partir de **données numériques**. Nous ferons maintenant la même chose pour un pourcentage calculé à partir de **données catégoriques**.

Exemple 3

Parmi un échantillon de 120 patients atteints de tuberculose choisi parmi la population totale de tuberculeux du pays, on a constaté que 28 (ou 23,3 %) ne suivaient pas leur traitement ambulatoire. Les 92 autres (ou 76,7 %) s'y conformaient de façon satisfaisante. Nous voulons maintenant calculer l'erreur-type du pourcentage de personnes ne suivant pas leur traitement (23,3 %). On le fait de la façon suivante.

Si p représente l'un des pourcentages (23,3 %) et $100 - p$ l'autre pourcentage (76,7 %), l'erreur-type du pourcentage est obtenue en les multipliant, en divisant le résultat par la taille de l'échantillon et en extrayant la racine carrée de ce nombre.

La formule de calcul de l'erreur-type d'un pourcentage est la suivante :

$$\sqrt{p(100 - p) / n}$$

Dans notre **exemple**, nous avons donc :

$$\sqrt{23,3 \times 76,7 / 120} = \sqrt{14,89} = 3,9$$

Nous voulons également calculer l'intervalle de confiance du pourcentage de personnes qui ne suivent pas leur traitement dans l'ensemble du pays.

L'intervalle de confiance pour une sécurité de 95 % est de

$$23,3 \% \pm (2 \times 3,9), \text{ c'est-à-dire de } 15,5 \% \text{ à } 31,1 \%$$

Il est donc probable à 95 % que parmi la population de tous les tuberculeux du pays où l'échantillon de 120 personnes a été choisi, de 15,5 à 31,1 % des malades ne suivent pas leur traitement ambulatoire.

Soulignons qu'au lieu de pourcentages, il est possible d'utiliser des proportions, qui peuvent avoir une valeur de 0 à 1. La formule de calcul de l'erreur-type serait alors la suivante :

$$\sqrt{p(1 - p) / n}$$

Dans notre **exemple**, le résultat serait le suivant :

$$\sqrt{0,233 \times 0,767 / 120} = 0,039$$

TRAVAIL EN GROUPE

- Calculez l'étendue, l'écart-type et l'intervalle de confiance pour une sécurité de 95 % de votre série de données numériques la plus importante.

Interprétez les résultats.

- Calculez l'intervalle de confiance pour une sécurité de 95 % de vos séries de données catégoriques les plus importantes. Ce calcul peut s'appliquer aux données numériques si vous les avez regroupées en catégories.

Interprétez les résultats.

Prenez note des résultats de votre travail. Vous en aurez besoin lors des tests statistiques.

Annexe 25.1. Formule de calcul de l'écart-type

$$\text{Écart-type (s)} = \sqrt{(\sum x^2 - (\sum x)^2 / n) / (n - 1)}$$

où x = valeur
 x^2 = carré de la valeur
 Σ = somme
 n = nombre d'observations

En appliquant cette formule à l'**exemple 1** (poids de 11 enfants de trois ans, page 5), nous constatons qu'elle n'est pas aussi difficile qu'elle en a l'air. Les valeurs de x sont :

13, 14, 14, 15, 16, 16, 16, 17, 17, 18, 20 kg et n = 11

Le carré de chaque valeur donne :

169, 196, 196, 225, 256, 256, 256, 289, 289, 324, 400

La somme de tous les carrés ($\sum x^2$) est de 2 856

La somme de toutes les valeurs ($\sum x$) est de 176

Donc, $(\sum x)^2 = 176^2 = 30\,976$

et $(\sum x)^2 / n = 30\,976 / 11 = 2\,816$

$$\text{Écart-type} = \sqrt{(2\,856 - 2\,816) / 10} = \sqrt{4} = 2 \text{ kg}$$

On peut calculer l'écart-type plus facilement en utilisant la fonction automatique dont sont munies certaines calculatrices.

Notes du formateur**Module 25 : DESCRIPTION DES VARIABLES, DEUXIÈME PARTIE****Durée et méthode d'enseignement**

1 heure Introduction et discussion

3 heures* Travail en groupe

Introduction et discussion

- C'est dans ce module que sont présentées pour la première fois des notions statistiques nouvelles pour plusieurs participants, sinon pour tous, en fonction de leur scolarité. Il est donc recommandé de procéder lentement, de bien expliquer les formules et de présenter des exemples simples.
- Assurez-vous que tous les participants comprennent la notion de variation de l'échantillonnage (section IV), car elle est essentielle pour saisir le fonctionnement des tests d'hypothèses (module suivant).
- Si les participants ont peu d'expérience en statistique, vous pouvez exclure de votre exposé les première et deuxième partie de la section V. Cependant, les groupes qui mènent une étude descriptive sans comparaisons pourraient devoir calculer l'erreur-type d'une différence (première partie) ou l'erreur-type d'un pourcentage (deuxième partie) au lieu de faire des tests statistiques. Dans ce cas, vous devriez discuter de cette section dans votre exposé.

Travail en groupe

- Même si certains calculs ne sont pas appropriés pour certains groupes, ces derniers devraient les faire au moins une fois avec des données de l'échantillon.
- Il n'est pas nécessaire de calculer l'erreur-type de différences ou de pourcentages si des tests statistiques seront effectués pour établir les différences entre des groupes ou pour évaluer des associations entre variables (voir les **modules 27 à 30**).

**Formation à la recherche sur les systèmes de santé
Volume 2, 2^e partie : Analyse de données et rédaction de rapports**

Module 26

TESTS D'HYPOTHÈSES

ÉTAPES DE L'ANALYSE DE DONNÉES ET DE LA RÉDACTION DE RAPPORTS

Questions à poser	Étapes à suivre*	Éléments importants de l'étape
Quelles données ont été recueillies pour chaque objectif de recherche? Les données sont-elles complètes et exactes?	Préparer les données pour l'analyse	Revoir les activités sur le terrain Compiler les données pour chaque objectif ou population étudiée Classer les données et en vérifier la qualité Vérifier les imprimés
À quoi ressemblent les données?	Décrire les variables	Distributions Figures, moyennes
Comment peut-on résumer les données pour en faciliter l'analyse?	Faire le recoupement des données quantitatives Résumer les données qualitatives	Faire le recoupement des données par rapport aux objectifs Illustrations, texte
Pour les données qualitatives : la recherche a-t-elle pour but de décrire, de comparer ou de trouver des associations?	Déterminer le type d'analyse statistique nécessaire	Examiner les objectifs, le type d'étude et les variables Faire une description statistique des variables Choisir les tests d'hypothèses
1. Comment peut-on décrire les données?	Analyser des observations appariées et non appariées	**Test t de Student **Test t pour observations appariées **Test du khi carré **Test du khi carré de McNemar
2. Comment peut-on déterminer les différences entre les groupes?	Évaluer les associations entre les variables	**Diagramme de dispersion **Droite de régression et coefficient de corrélation **Risque relatif, risque relatif approché
3. Comment peut-on déterminer les associations entre les variables?	Rédiger le rapport et formuler des recommandations	Préparer le canevas du rapport Rédiger les 1 ^{re} et 2 ^e versions Faire un sommaire des constatations Faire un sommaire des conclusions pour chaque objectif Formuler des recommandations Préparer un résumé analytique
Comment le rapport doit être rédigé?	Présenter un sommaire et rédiger un plan de mise en oeuvre des recommandations	Discuter des sommaires avec différents groupes cibles Discuter du programme de mise en oeuvre
Comment doit-on présenter et diffuser les constatations et les recommandations?		

* Il n'est pas nécessaire de faire ces étapes dans l'ordre indiqué. L'ordre peut être adapté aux besoins de l'atelier.

** Ces éléments sont facultatifs; ils peuvent être omis s'ils ne sont pas pertinents au cours.

Module 26 : TESTS D'HYPOTHÈSES

OBJECTIFS

À la fin de la présente séance, vous devriez pouvoir :

1. Expliquer ce que sont les tests d'hypothèses et à quoi ils servent;
2. Au moyen de tableaux, choisir les tests d'hypothèses appropriés pour différentes séries de données;
3. Choisir des tests d'hypothèses appropriés pour vos données.

I. Introduction

II. Tests d'hypothèses

III. Fonctionnement des tests d'hypothèses

IV. Sélection d'un test d'hypothèses

I. INTRODUCTION

Comme nous l'avons déjà mentionné, l'analyse et l'interprétation des résultats de notre étude doivent correspondre aux objectifs de celle-ci. Le **module 24** explique comment faire des tableaux croisés appropriés aux objectifs de la recherche.

Supposons que selon une étude sur la consommation de tabac, 30 % des hommes de l'échantillon sont fumeurs, alors que seulement 20 % des femmes sont fumeuses. Comment interpréter ce résultat?

- Cette différence observée de 10 % peut constituer une **différence véritable**, qui existe également dans la population de laquelle l'échantillon a été tiré.
- Cette différence peut également être **due au hasard**; en effet, il n'y a pas de différence entre les hommes et les femmes, mais l'échantillon d'hommes était différent de l'échantillon de femmes. On peut également affirmer que la différence observée est attribuable à la variation de l'échantillonnage.
- Enfin, cette différence de 10 % peut être attribuable à des lacunes sur le plan de la conception de l'étude (que l'on désigne sous le nom de **biais**); elle ne se serait pas produite si l'étude avait été bien conçue.

Lors de la conception de notre étude, nous avons tenté d'éviter le biais en choisissant un type d'étude approprié, les bonnes techniques de collecte de données, une bonne procédure d'échantillonnage, etc. Cependant, il est toujours souhaitable de déterminer les biais qui auraient pu se produire pendant la mise en oeuvre de l'étude et influencer les résultats.

Si l'on arrive à la conclusion qu'une différence observée entre deux groupes ne découle pas d'un biais, il faut déterminer si cette différence est véritable. On ne peut le faire qu'en excluant les risques de hasard (variation de l'échantillonnage; voir le **module 25**). On y arrive en faisant un test d'hypothèses.

II. TESTS D'HYPOTHÈSES

Les **TESTS D'HYPOTHÈSES** permettent d'évaluer la probabilité qu'une observation (p. ex., une différence entre deux groupes) soit due au hasard.

En d'autres mots, les tests d'hypothèses servent à déterminer si une observation peut s'appliquer à la population de laquelle l'échantillon étudié a été tiré.

Les tests d'hypothèses varient selon le type de données. Dans le présent module, deux organigrammes vous aideront à choisir un test d'hypothèses qui convient à vos données. Le premier doit être utilisé si vous comparez des groupes pour en déterminer les différences. Utilisez le deuxième si vous voulez mesurer des associations entre variables.

Dans les modules suivants, certains des tests d'hypothèses les plus courants seront étudiés en détail. Pour l'instant, examinons le fonctionnement des tests d'hypothèses.

III. FONCTIONNEMENT DES TESTS D'HYPOTHÈSES

L'utilisation des tests d'hypothèses est fondée sur le raisonnement suivant :

- Supposons que vous avez observé une différence entre deux groupes (ou une association entre deux variables) dans votre échantillon.
- Vous voulez savoir si cette différence entre les deux groupes (ou l'association entre les variables) représente une différence véritable (ou une association véritable) dans la population de laquelle l'échantillon a été tiré, ou si elle découle du hasard (en raison de la variation de l'échantillonnage).
- Pour répondre à cette question, vous déterminez à quel point il est probable que votre observation soit attribuable au hasard, s'il n'existe aucune différence entre les deux groupes dans la population totale.

Si vous mesurez une association entre deux variables, vous déterminez à quel point il est probable que votre observation soit attribuable au hasard (c.-à-d. qu'elle se soit produite en raison de la variation de l'échantillonnage).

S'il est **peu probable** (probabilité de moins de 5 % ou inférieure à un pourcentage préétabli¹) que votre observation soit due au hasard, vous rejetez cette éventualité et concluez qu'il existe une différence (ou une association) véritable. Vous pouvez alors affirmer que cette différence (ou association) est **significative**.

S'il est **probable** (probabilité de 5 % ou plus) que votre observation soit due au hasard, vous ne pouvez pas conclure qu'il existe une différence (ou une association) véritable. Cette différence (ou association) **n'est pas significative**.

Remarque

En statistique, l'hypothèse selon laquelle des groupes ne présentent aucune différence véritable (ou des variables ne sont pas associées) constitue l'**hypothèse nulle**.

Voici des exemples d'hypothèses nulles :

- Il n'y a aucune différence dans l'incidence de la rougeole chez les enfants vaccinés et non vaccinés.
- Les hommes ne fument pas plus que les femmes.
- Il n'y a aucune association entre le revenu des familles et la malnutrition chez les enfants.

Il est important de souligner que le fait qu'une différence ou une association soit «significative» ne veut pas dire qu'elle est importante. La différence la plus petite et la moins pertinente se révélera significative si l'échantillon est assez grand. Par contre, si l'on utilise un petit échantillon, des différences importantes pourraient ne pas être significatives.

¹ Vous pouvez choisir vous-même la valeur de p : par exemple, 0,10, 0,05, 0,01 ou même 0,001, selon le degré de certitude désiré. Le pourcentage de 5 % est arbitraire. Ainsi, le chercheur peut choisir un pourcentage de 1 % s'il désire réduire le risque que son observation soit due au hasard.

La **PROBABILITÉ** qu'une observation soit due au hasard est exprimée habituellement par une **valeur de p**.

La valeur de p est exprimée sous forme de proportion. Une probabilité de 5 % correspond à une valeur de p de 0,05.

Par conséquent, une différence ou une association est considérée significative si $p < 0,05$. En d'autres mots, si l'hypothèse nulle selon laquelle il n'y a pas de différence entre les deux groupes est vraie, il y aura une différence dans vos données 5 fois ou moins tous les 100 échantillons étudiés.

IV. SÉLECTION D'UN TEST D'HYPOTHÈSES

Le test d'hypothèses choisi doit correspondre au but de l'étude et au type de données recueilli. Le processus de sélection est expliqué au moyen des tableaux 26.1 et 26.2.

Remarque

Avant d'utiliser un test statistique, énoncez l'hypothèse nulle appliquée aux données faisant l'objet du test. Vous pourrez ainsi interpréter les résultats du test.

Pour établir des différences entre des groupes, utilisez le tableau 26.1. Vous devez déterminer si vous avez des observations appariées ou non appariées. Dans bien des études, les échantillons de différents groupes sont choisis de façon indépendante, sans que l'on tente de faire correspondre les unités étudiées dans les différents échantillons. Dans ce cas, on obtient des observations et des échantillons non appariés.

Cependant, il arrive que deux séries d'observations soient recueillies de telle sorte que les observations d'une série sont couplées à celles de l'autre série. Il s'agit alors d'observations et d'échantillons appariés (voir les **modules 9 et 23**).

Les sections suivantes, qui expliquent comment utiliser les deux tableaux, contiennent des exemples d'échantillons appariés et non appariés.

Différences entre des groupes (tableau 26.1)

Avant de choisir le test à utiliser pour déterminer si les différences entre des groupes sont significatives, il faut envisager plusieurs questions. Il faut déterminer d'abord si les observations sont appariées ou non appariées (consulter les **modules 9 et 23** au besoin); ensuite, il faut déterminer si les données sont nominales, ordinales ou numériques (les définitions de ces termes figurent au **module 22**).

Pour les **données nominales** (appariées ou non), le test d'hypothèses à employer repose sur la taille de l'échantillon. Il n'existe pas de définition précise de ce qui constitue un échantillon «petit» ou «grand». Cependant, dans le cas des observations non appariées, il est préférable de recourir à la **méthode exacte de Fisher** plutôt qu'au **test du khi carré** si l'échantillon comprend moins de 40 sujets ou si l'une des cellules du tableau qui doit être dressé contiendra une valeur inférieure à 5. Dans les présents modules de formation, nous n'envisageons que les tests appliqués à de grands échantillons. Le **test du khi carré** sera abordé au **module 27**, et le **test du khi carré de McNemar** sera étudié au **module 28**.

Exemple d'échantillon non apparié

Dans une étude sur l'efficacité du vaccin contre la rougeole, l'équipe de recherche a décidé d'étudier 100 personnes atteintes âgées de 1 à 5 ans se présentant à une clinique et 100 patients du même âge qui n'avaient pas la rougeole. En comparant les deux groupes quant aux vaccinations reçues, l'équipe a constaté que le taux de vaccination parmi les personnes atteintes de la rougeole était inférieur à celui des patients non atteints. Le **test du khi carré** a été employé pour déterminer si cette différence était significative.

Tableau 26.1. Sélection d'un test d'hypothèses pour établir les différences entre des groupes.

	Observations non appariées	Observations appariées
Données nominales Petit échantillon Grand échantillon	Méthode exacte de Fisher Test du khi carré* (module 27)	Test des signes Test du khi carré de McNemar* (module 28)
Données ordinales Deux groupes Plus de deux groupes	Test de Wilcoxon à deux échantillons ou test U de Mann-Whitney Test de Kruskal-Wallis (analyse de la variance à un critère de classification)	Test de Wilcoxon pour observations appariées Test de Friedman (analyse de la variance à deux critères de classification)
Données numériques Deux groupes Plus de deux groupes	Test t* (module 27) Test F	Test t pour observations appariées* (module 28)

* Les tests marqués d'un astérisque seront abordés dans les modules indiqués.

Source : Adapté de Riegelman, R.K. 1981. *Studying a study and testing a test*, Little, Brown and Company, Boston, MA, États-Unis, p. 246.

Exemple d'échantillon apparié

Dans une autre étude dont l'objectif était le même, les chercheurs pensaient que l'âge et le sexe pouvaient influencer sur la susceptibilité à la rougeole et sur la probabilité d'avoir été vacciné. Par conséquent, pour chaque personne atteinte de la rougeole se présentant à une clinique, un patient du même âge et du même sexe non atteint de la rougeole a été sélectionné à partir de la liste des patients ambulatoires. Ainsi, le taux de vaccination de 100 paires de sujets a été évalué. Le **test du khi carré de McNemar** a été employé pour analyser les données.

Pour les données ordinales, le test d'hypothèses à utiliser repose sur la comparaison de deux groupes ou de plusieurs groupes.

Les tests à utiliser pour comparer deux groupes sont fondés sur le classement des données : le test de Wilcoxon à deux échantillons, qui donne des résultats équivalant à ceux du test U de Mann-Whitney pour les observations non appariées, et le **test de Wilcoxon pour observations appariées**. Nous ne traiterons pas de ces tests dans les modules de formation; si vous voulez les utiliser, consultez un manuel de statistique (voir l'**annexe 26.1** : le chapitre 10 du manuel *Statistics at square one* de Swinscow est très clair et facile à comprendre). Vous pouvez également consulter votre formateur ou un statisticien.

Exemple d'étude où les données sont classées

Dans une étude quasi expérimentale visant à étudier l'effet d'une campagne d'éducation sur la santé portant sur le traitement de la diarrhée, deux groupes de villages ont été sélectionnés. Le premier était composé de villages où la campagne a été tenue, et le deuxième de villages où n'a été donnée aucune séance. Au cours de l'analyse, les villages ont été classés par ordre décroissant du degré de connaissances sur le traitement adéquat de la diarrhée. Le **test de Wilcoxon à deux échantillons** a été effectué pour déterminer si les deux groupes de villages présentaient une différence significative.

Pour les **données numériques**, comme pour les données ordinales, le choix du test d'hypothèses dépend du nombre de groupes comparés (deux ou plusieurs).

Aux **modules 27 et 28**, nous discuterons de l'exécution et de l'interprétation d'un **test t** et d'un **test t pour observations appariées** pour la comparaison de deux groupes. Si vous comparez plus de deux groupes, vous devriez consulter un manuel de statistique (voir l'**annexe 26.1**).

Exemple d'échantillon non apparié, deux groupes

Dans le cadre d'une étude sur l'alimentation, on mesure le poids de 142 enfants de cinq ans vivant dans des régions rurales et de 171 enfants de cinq ans vivant dans des régions urbaines. Le poids moyen pour chacun des deux échantillons est calculé et comparé au moyen du **test t**.

Exemple d'échantillon non apparié, plus de deux groupes

Le poids moyen des quatre groupes suivants d'enfants de cinq ans est comparé : garçons des régions rurales, garçons des régions urbaines, filles des régions rurales et filles des régions urbaines. Dans ce cas, le test approprié est le **test F**.

Exemple d'échantillon apparié, deux groupes

Le poids moyen d'hommes et de femmes adultes est comparé en tenant compte de la taille. Ainsi, pour chaque homme d'une certaine taille, une femme de la même taille est choisie de façon à comparer leur poids. Dans ce cas, le **test t pour observations appariées** est employé.

Associations entre variables (tableau 26.2)

Tableau 26.2. Sélection d'un test d'hypothèses pour mesurer des associations entre variables.

Données nominales	Test du khi carré (si l'échantillon est assez grand)* (modules 27 et 28)	Risque relatif approché ou estimation du risque relatif* (module 30)
Données ordinales ou numériques lorsque aucune relation linéaire n'est prévue	Coefficient de corrélation des rangs de Spearman ou de Kendall	Caractère significatif du coefficient de corrélation des rangs de Spearman ou de Kendall
Données numériques lorsqu'une relation linéaire est prévue	Coefficient de corrélation de Pearson (r)* (module 29)	Caractère significatif du coefficient de corrélation de Pearson (r)* (module 29)

* Les tests marqués d'un astérisque seront abordés dans les modules indiqués.

Source : Adapté de Riegelman, R.K. 1981. *Studying a study and testing a test*, Little, Brown and Company, Boston, MA, États-Unis, 1981, p. 246.

Déterminez si vos données sont nominales, ordinales ou numériques. Si elles sont numériques, déterminez s'il pourrait y avoir une relation linéaire. Dans les cas des données numériques, le terme «relation linéaire» signifie que l'association est telle que la variable dépendante varie de manière constante par rapport à la variable indépendante de sorte que les points d'un diagramme de dispersion, lorsqu'ils sont joints, forment une droite.

Pour **les données nominales**, le **risque relatif** constitue une bonne mesure de l'association, qu'on utilise souvent dans les études cas-témoins et dans les études de cohortes. Le **module 30** traite du calcul du **risque relatif approché** comme estimation du risque relatif dans les études cas-témoins (observations appariées et non appariées).

Exemple

Dans une étude cas-témoins sur la tuberculose où vous étudiez le type d'emploi, vous constatez que les mineurs sont plus susceptibles de contracter la tuberculose que les agriculteurs. Un test du khi carré confirme que la différence sur le plan de l'incidence de la tuberculose entre les mineurs et les agriculteurs est significative. Le calcul du **risque relatif approché** vous aiderait à exprimer à quel point les mineurs sont plus susceptibles de contracter la tuberculose que les agriculteurs.

Pour **les données ordinales**, le **coefficient de corrélation de Spearman ou de Kendall** peut être calculé et testé pour vérifier s'il est significatif. Pour calculer ces coefficients, consultez un manuel de statistique (voir l'annexe 26.1; p. ex., Swinscow, 1983, chapitre 12).

Pour **les données numériques**, lorsqu'on soupçonne la présence d'une relation linéaire, on peut calculer et vérifier le caractère significatif du coefficient de corrélation de Pearson. Le **module 29** illustre la marche à suivre.

Exemple

Vous voulez déterminer si le poids des enfants de cinq ans est associé au revenu de la famille. Vous soupçonnez l'existence d'une relation linéaire entre les variables «revenu de la famille» et «poids» selon laquelle le poids augmente lorsque le revenu de la famille augmente.

Remarque

Une association significative ne prouve pas nécessairement l'existence d'une relation causale. Cependant, elle peut amener le chercheur à vérifier s'il existe vraiment une telle relation.

EXERCICES de sélection d'un test

Au moyen des tableaux, déterminez les tests appropriés pour les recherches suivantes.

Exercice 1

Une étude sera menée pour comparer l'effet d'un nouveau médicament hypotenseur sur la pression sanguine diastolique d'un échantillon par rapport à l'effet d'un placebo sur un échantillon de contrôle non apparié (Riegelman 1981, p. 243).

Exercice 2

Une étude sera effectuée pour découvrir si les femmes enceintes vivant dans un logement dépourvu d'eau courante courent un risque plus élevé d'avoir un enfant mort-né et, dans l'affirmative, pour mesurer à quel point cette association est forte.

Exercice 3

Une étude a été entreprise dans le but de déterminer si une perte de poids significative s'est produite chez les sujets qui ont suivi un cours d'un an sur le traitement du diabète, et si le poids perdu était relié au poids initial. Le tableau suivant donne le poids initial (x) et le poids après un an de traitement (y) de 16 adultes chez qui le diabète a été décelé il y a peu de temps.

Poids initial (x) en livres	Poids après 1 an (y) en livres	Poids initial (x) en livres	Poids après 1 an (y) en livres
140	115	120	123
160	130	145	143
180	135	150	125
120	125	160	140
132	112	160	135
146	130	149	120
190	160	129	119
200	160	150	132

Exercice 4

À partir d'études antérieures, il est établi que 30 % des couples admissibles d'une circonscription sanitaire utilisent une méthode de contraception. Après un programme d'éducation de masse, sur 90 couples admissibles choisis au hasard, 40 utilisaient une méthode de contraception. L'agent d'éducation sur la santé veut savoir si son programme a eu une incidence sur le groupe cible.

TRAVAIL EN GROUPE

En vous fondant sur les objectifs spécifiques de votre étude et sur votre liste de variables et en utilisant les tableaux croisés déjà faits, choisissez les tests d'hypothèses auxquels vous devrez soumettre vos données.

Annexe 26.1. Ouvrages consultés

Barker, D.J.P., 1982. Practical epidemiology (3^e éd.). Churchill Livingstone, Édimbourg, Royaume-Uni.

Bradford Hill, A. 1984. A short textbook of medical statistics (11^e éd.). Hodder and Stoughton, Londres, Royaume-Uni.

Fisher, A., Laing, J., Stoeckel, J. 1983. Handbook for family planning operations research design. Population Council, New York, NY, États-Unis.

Kelsey, J.L., Thompson, W.D., Evans, A.S., 1986. Methods in observational epidemiology. Oxford University Press, Oxford, Royaume-Uni.

Kidder, L.H., Judd, C.M., 1986. Research methods in social relations. CBS College Publishing, New York, NY, États-Unis.

Kleinbaum, D.G., Kupper, L.L., Morgenstern, H. 1982. Epidemiologic research - principles and quantitative methods. Van Nostrand Reinhold, New York, NY, États-Unis.

McMahon, B., Pugh, T.F. 1970. Epidemiology - principles and methods. Little, Brown and Co., Boston, MA, États-Unis.

Riegelman, R.F. 1981. Studying a study and testing a test. Little, Brown and Co., Boston, MA, États-Unis.

Schlesselman, J.J., 1982. Case control studies - design, conduct, analysis. Oxford University Press, Oxford, Royaume-Uni.

Swinscow, T.D.V., 1980, Statistics at square one (7^e éd.). British Medical Association, Londres, Royaume-Uni.

Notes du formateur

Module 26 : TESTS D'HYPOTHÈSES

Durée et méthode d'enseignement

1/2 heure* Introduction et discussion

1 heure* Travail en groupe

* Une période d'une demi-heure à une heure devrait être ajoutée à la séance plénière ou au travail en groupe si le formateur demande aux participants de faire les exercices de sélection d'un test d'hypothèses.

Introduction et discussion

- Le présent module n'a pas à être présenté en entier. Comme l'indiquent les objectifs, le but principal de cette séance est de faire comprendre aux participants **ce que sont les tests** d'hypothèses et **pourquoi** ils doivent y recourir. En outre, les participants devraient pouvoir utiliser les tableaux pour choisir les tests d'hypothèses appropriés pour différentes séries de données.

Les exemples d'utilisation des deux tableaux **ne devraient pas** être présentés.

- Vous pouvez expliquer l'**utilité** des tests d'hypothèses en présentant un tableau croisé (par exemple, le nombre de fumeurs et de non-fumeurs parmi les hommes et les femmes) et en demandant comment interpréter la différence entre les hommes et les femmes (30 % par rapport à 20 %). Il serait même préférable de prendre un tableau croisé préparé par l'une des équipes de recherche et de poser à celle-ci la même question.
- Une fois que les participants ont bien compris à quoi servent les tests d'hypothèses, demandez-leur de donner des exemples de résultats (tableaux croisés) tirés de leurs études et devant être soumis à des tests d'hypothèses.
- Soulignez qu'il existe deux tests d'usage courant : le test t et le test du khi carré. Tous les autres tests mentionnés au tableau 26.1 sont utilisés plus rarement dans les recherches du type que les participants effectuent le plus souvent.
- Si les participants sont assez avancés pour apprendre eux-mêmes à utiliser les deux tableaux, vous pouvez leur demander de faire quelques exercices, ou même les cinq. Demandez-leur de prendre quelques minutes pendant la séance plénière pour utiliser les tableaux afin de choisir les tests appropriés, et demandez les réponses à des volontaires. Une autre méthode consisterait à donner plusieurs ou tous les exercices à faire au début du travail en groupe, en aidant les membres du groupe à se familiariser avec les tableaux.
- Après avoir présenté le module 26, nous vous recommandons d'aborder une partie du module suivant, c'est-à-dire le test t ou le test du khi carré. Les notions abordées dans le présent module apparaîtront alors plus concrètes aux participants.

Travail en groupe

- Laissez les participants choisir les tableaux croisés qu'ils doivent soumettre à des tests d'hypothèses. Les participants devraient également choisir le test approprié pour chaque tableau croisé.

Tests suggérés pour les études mentionnées dans les exercices

Exercice 1

L'étude porte sur des **échantillons**. On veut déterminer si les différences entre les résultats ou les valeurs de la variable dépendante, c'est-à-dire le changement de pression sanguine diastolique, sont significatives. Il s'agit là de **données numériques** (tableau 26.1).

Il y a une comparaison (deux groupes) et les échantillons sont **non appariés**. Par conséquent, le **test t** doit être effectué.

Exercice 2

On suppose qu'il s'agit d'**échantillons**. On étudie le caractère significatif des **différences**, et le résultat ou la variable dépendante étudiée est le nombre de mortalités périnatales, c'est-à-dire des données **nominales** (tableau 26.1). Ces études sont appliquées à de grands échantillons qui sont **non appariés**. Il faut donc choisir le **test du khi carré**.

Pour déterminer le caractère significatif de l'**association**, il faut utiliser le tableau 26.2. Comme il s'agit de **données nominales**, il faut calculer le **risque relatif approché** ou le **risque relatif**.

Le même tableau de deux cellules sur deux employé pour le test du khi carré peut également être utilisé dans ce cas.

Exercice 3

Cet exemple comprend des échantillons, des différences, le poids comme variable dépendante et des données numériques (tableau 26.1). L'échantillon n'était que de 16 patients, mais les mesures ont été prises deux fois pour chaque patient. Il s'agit donc d'un **échantillon apparié**. Dans ce cas, il est jumelé à lui-même.

Le test à choisir est donc le **test t pour observations appariées**. Pour savoir si la perte de poids était reliée au poids initial, il faut déterminer le degré **d'association**.

Comme les données en question sont des **données numériques**, il faut calculer le **coefficient de corrélation de Pearson (r)**. S'il est nécessaire de calculer le caractère significatif de l'association, on utilise le tableau 26.2, qui suggère de déterminer le caractère significatif du **r de Pearson**.

Exercice 4

Dans cette étude, un seul **échantillon** a été choisi. On le compare à des données connues sur la population. Comme ces données sont **nominales**, le test porte sur des différences entre proportions ou entre pourcentages.

**Formation à la recherche sur les systèmes de santé
Volume 2, 2^e partie : Analyse de données et rédaction de rapports**

Module 27

**DIFFÉRENCES ENTRE DES GROUPES
PREMIÈRE PARTIE — ANALYSE D'OBSERVATIONS NON
APPARIÉES**

ÉTAPES DE L'ANALYSE DE DONNÉES ET DE LA RÉDACTION DE RAPPORTS

Questions à poser	Étapes à suivre*	Éléments importants de l'étape
Quelles données ont été recueillies pour chaque objectif de recherche? Les données sont-elles complètes et exactes?	Préparer les données pour l'analyse	Revoir les activités sur le terrain Compiler les données pour chaque objectif ou population étudiée Classer les données et en vérifier la qualité Vérifier les imprimés
À quoi ressemblent les données?	Décrire les variables	Distributions Figures, moyennes
Comment peut-on résumer les données pour en faciliter l'analyse?	Faire le recoupement des données quantitatives Résumer les données qualitatives	Faire le recoupement des données par rapport aux objectifs Illustrations, texte
Pour les données qualitatives : la recherche a-t-elle pour but de décrire, de comparer ou de trouver des associations?	Déterminer le type d'analyse statistique nécessaire	Examiner les objectifs, le type d'étude et les variables Faire une description statistique des variables Choisir les tests d'hypothèses
1. Comment peut-on décrire les données?	Analyser des observations appariées et non appariées	**Test t de Student **Test t pour observations appariées **Test du khi carré **Test du khi carré de McNemar
2. Comment peut-on déterminer les différences entre les groupes?	Évaluer les associations entre les variables	**Diagramme de dispersion **Droite de régression et coefficient de corrélation **Risque relatif, risque relatif approché
3. Comment peut-on déterminer les associations entre les variables?	Rédiger le rapport et formuler des recommandations	Préparer le canevas du rapport Rédiger les 1 ^{re} et 2 ^e versions Faire un sommaire des constatations Faire un sommaire des conclusions pour chaque objectif Formuler des recommandations Préparer un résumé analytique
Comment le rapport doit être rédigé?	Présenter un sommaire et rédiger un plan de mise en oeuvre des recommandations	Discuter des sommaires avec différents groupes cibles Discuter du programme de mise en oeuvre
Comment doit-on présenter et diffuser les constatations et les recommandations?		

* Il n'est pas nécessaire de faire ces étapes dans l'ordre indiqué. L'ordre peut être adapté aux besoins de l'atelier.

** Ces éléments sont facultatifs; ils peuvent être omis s'ils ne sont pas pertinents au cours.

**Module 27 : DIFFÉRENCES ENTRE DES GROUPES
PREMIÈRE PARTIE - ANALYSE D'OBSERVATIONS NON APPARIÉES**

OBJECTIFS

À la fin de la présente séance, vous devriez pouvoir :

1. Déterminer quand utiliser le test t (observations non appariées) et le test du khi carré;
2. Calculer la valeur de t et de khi carré;
3. Utiliser les tables des valeurs de t et de khi carré pour déterminer si les valeurs de t et de khi carré sont significatives;
4. Déterminer si vos données peuvent être soumises à ces tests et, dans l'affirmative, choisir celui qui convient;
5. Soumettre vos données à ces tests.

I. Introduction

II. Test t

III. Test du khi carré (χ^2)

I. INTRODUCTION

En examinant les tableaux croisés de vos principales variables (voir le **module 23**), vous avez probablement observé des différences entre les groupes. Vous pouvez déterminer si ces différences sont dues au hasard ou si elles sont véritables (significatives).

Pour le faire, vous pouvez effectuer deux tests :

- le test t,
- le test du khi carré.

Le **test t** s'utilise avec des données NUMÉRIQUES pour comparer les moyennes de deux groupes.

Le **test du khi carré** s'utilise avec des données CATÉGORIQUES pour comparer des proportions de données revenant dans deux ou plusieurs groupes.

Ces deux tests sont valables pour des observations NON APPARIÉES. Pour les observations appariées, on utilise deux tests différents, selon que les données sont catégoriques ou numériques. Des renseignements sur ces tests figurent au **module 28**.

II. TEST T

Le **test t**, également appelé **test t de Student**, s'utilise avec des données numériques pour déterminer si une différence observée entre les moyennes de deux groupes peut être considérée significative.

Exemple 1

Dans une province, la proportion des femmes accouchant par césarienne est très élevée. Une étude est donc entreprise pour en découvrir les raisons. Comme une petite taille est reconnue comme l'un des facteurs de risque reliés aux accouchements difficiles, le chercheur peut tenter de découvrir s'il y a une différence entre la taille moyenne des femmes de cette province qui ont eu un accouchement normal et celle des femmes qui ont accouché par césarienne. L'hypothèse nulle serait qu'il n'y a aucune différence entre la taille moyenne des deux groupes de femmes. Supposons que le chercheur a obtenu les résultats suivants :

Tableau 27.1. Taille moyenne de femmes ayant eu un accouchement normal et de femmes ayant accouché par césarienne.

	Nombre de femmes étudiées	Taille moyenne (cm)	Écart-type
Accouchement normal	60	156	3,1
Césarienne	52	154	2,8

Le test t serait le bon moyen de déterminer si la différence observée de 2 cm peut être considérée significative. Pour effectuer le test t, il faut franchir trois étapes :

1. Calculer la valeur de t;
2. Consulter une table des valeurs de t;
3. Interpréter les résultats.

1. Calcul de la valeur de t

Pour calculer la valeur de t, il faut :

- Calculer la différence entre les moyennes.**
Dans l'exemple précédent, la moyenne est de $156 - 154 = 2$ cm.
- Calculer l'écart-type** pour chaque groupe (la notion d'écart-type et la marche à suivre pour le calculer ont été abordés au **module 25**). Supposons que les écarts-types figurant dans la dernière colonne du tableau 27.1 ont été calculés.
- Calculer l'erreur-type de la différence entre les deux moyennes.**

L'erreur-type de la différence se calcule au moyen de la formule suivante :

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Où : s_1 est l'écart-type du premier échantillon,
 s_2 est l'écart-type du deuxième échantillon,
 n_1 est la taille du premier échantillon,
 n_2 est la taille du deuxième échantillon.

Si nous supposons que les femmes ayant accouché normalement constituent l'échantillon 1 et que les femmes ayant subi une césarienne forment l'échantillon 2, l'écart-type de la différence est de :

$$\sqrt{\frac{3,1^2}{60} + \frac{2,8^2}{52}} = 0,56$$

- Diviser la différence entre les moyennes par l'erreur-type** de la différence. Le résultat constitue la **valeur de t**.

Dans l'exemple précédent, $t = \frac{2}{0,56} = 3,6$

On peut exprimer le tout par une seule formule :

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

où \bar{X}_1 constitue la moyenne du premier échantillon et \bar{X}_2 la moyenne du deuxième échantillon.

2. Consultation d'une table des valeurs de t

Une fois la valeur de t calculée, il faut consulter une table des valeurs de t qui permet de déterminer si l'hypothèse nulle est rejetée ou non. On en trouve une à l'**annexe 27.1**.

- a) En premier lieu, choisir le **seuil de signification** (voir le **module 26**). Il ne faut pas oublier que ce seuil constitue la probabilité de déceler une différence due au hasard lorsqu'il n'y a pas de différence réelle. On choisit habituellement un seuil de 0,05.
- b) En deuxième lieu, déterminer le nombre de **degrés de liberté** pour le test en question. Les degrés de liberté sont dérivés de la taille de l'échantillon, dont on doit tenir compte pour le test t. Plus l'échantillon est grand (et plus le nombre de degrés de liberté est élevé), plus la différence nécessaire pour rejeter l'hypothèse nulle sera petite.

Le mode de calcul des degrés de liberté diffère selon le test statistique. Pour le **test t de Student**, le nombre de degrés de liberté représente la somme de la taille des deux échantillons moins 2.

Ainsi, pour l'**exemple 1**, le nombre de degrés de liberté est de :

$$v = 60 + 52 - 2 = 110.$$

Remarque : Cette méthode de calcul est approximative. Pour obtenir la méthode exacte, consulter un manuel de statistique.

- c) En troisième lieu, localiser dans la table la valeur de t correspondant au seuil de signification et au nombre de degrés de liberté.

Dans notre exemple, nous relevons la valeur de t correspondant à $p = 0,05$ et à $v = 110$, qui est de 1,98.

3. Interprétation du résultat

Comparons maintenant la valeur absolue de t calculée à l'étape 1 (c.-à-d. la valeur de t sans le signe) avec la valeur de t tirée de la table à l'étape 2. Si la valeur calculée de t est **supérieure** à la valeur de la table, p est **inférieur** à la valeur indiquée en haut de la colonne. Nous **rejetons** donc l'**hypothèse nulle** et nous concluons qu'il existe une différence significative entre les deux moyennes.

Si la valeur calculée de t est **inférieure** à la valeur tirée de la table, p est **supérieur** à la valeur indiquée en haut de la colonne. Nous **acceptons** donc l'**hypothèse nulle** et nous concluons qu'il n'existe pas de différence significative entre les deux moyennes.

Dans notre exemple, la valeur de t calculée à l'étape 1 est de 3,6, ce qui est supérieur à la valeur de t tirée de la table à l'étape 2 (1,98). Ainsi, p est inférieur à 0,05, et nous rejetons l'hypothèse nulle pour conclure que la différence observée de 2 cm entre la taille moyenne des femmes ayant eu un accouchement normal et celle des femmes ayant subi une césarienne est significative.

Il est possible d'exprimer cette conclusion de différentes façons :

- Nous pouvons dire que la probabilité que la différence observée de 2 cm entre les deux groupes soit due au hasard est inférieure à 5 %.
- Nous pouvons également affirmer que la différence entre les deux groupes est égale à 3,6 fois l'erreur-type.

Pour comparer la moyenne de plus de deux groupes (p.ex., taille des femmes habitant dans des régions urbaines, semi-urbaines et rurales), on ne peut utiliser le **test t de Student**. Dans ce cas, il faut employer le **test F**, que nous n'aborderons pas dans ces pages.

III. TEST DU KHI CARRÉ (χ^2)

Lorsqu'on dispose de données catégoriques, le test du khi carré sert à déterminer si les différences observées entre des proportions de deux groupes peuvent être considérées significatives.

Exemple 2

Dans une étude des facteurs influant sur la fréquentation des cliniques prénatales, vous constatez que 64 % des femmes qui habitent à moins de 10 kilomètres de la clinique s'y rendent, par rapport à seulement 47 % de celles qui habitent à plus de 10 kilomètres. Cela suggère que les femmes qui vivent à proximité des cliniques reçoivent plus de soins prénataux. Les résultats complets figurent ci-dessous :

Tableau 27.2. Fréquentation des cliniques prénatales par des femmes habitant près et loin des cliniques.

Distance de la clinique	Ont fréquenté	N'ont pas fréquenté	Total
Moins de 10 km	51 (64 %)	29 (36 %)	80 (100 %)
10 km ou plus	35 (47 %)	40 (53 %)	75 (100 %)
Total	86	69	155

D'après ce tableau, il semble y avoir une différence quant à la fréquentation des cliniques prénatales entre les femmes qui vivent à proximité et celles qui vivent loin de la clinique (64 % par rapport à 47 %). Nous voulons maintenant savoir si cette différence est significative.

Le test du khi carré peut nous donner ce renseignement. Ce test mesure la différence entre les fréquences observées et les fréquences théoriques (ou espérées) **si l'hypothèse nulle (c.-à-d. l'hypothèse selon laquelle il n'y a aucune différence) était vraie.**

Pour effectuer un test du khi carré, il faut suivre les trois étapes suivantes :

1. Calculer la valeur de χ^2 ;
2. Consulter une table des valeurs de χ^2 ;
3. Interpréter les résultats.

1. Calcul de la valeur de χ^2

Suivre les étapes suivantes :

- a) Calculer la fréquence théorique (f_t) pour chaque cellule.

Pour trouver la fréquence théorique f_t d'une cellule, on multiplie le total de la rangée par le total de la colonne et on divise par le total général :

$$E = \frac{\text{total de la rangée} \times \text{total de la colonne}}{\text{total général}}$$

- b) Pour chaque cellule, soustraire la fréquence théorique de la fréquence observée (f_o).

$$f_o - f_t$$

- c) Pour chaque cellule, mettre au carré le résultat de ($f_o - f_t$) et diviser par la fréquence théorique f_t .
- d) Additionner les quotients de l'étape c) de toutes les cellules.

La formule de calcul de la valeur du khi carré (étapes b) à d)) est la suivante :

$$\chi^2 = \sum \frac{(f_o - f_t)^2}{f_t}$$

où f_o est la fréquence observée (indiquée dans le tableau),
 f_t est la fréquence théorique (à calculer),
 Σ (la somme de) représente l'addition de ($f_o - f_t$) pour toutes les cellules du tableau.

Dans le cas d'un tableau de deux cellules sur deux (qui contient donc quatre cellules), la formule est la suivante :

$$\chi^2 = \frac{(f_o(1) - f_t(1))^2}{f_t(1)} + \frac{(f_o(2) - f_t(2))^2}{f_t(2)} + \frac{(f_o(3) - f_t(3))^2}{f_t(3)} + \frac{(f_o(4) - f_t(4))^2}{f_t(4)}$$

2. Consultation d'une table des valeurs du χ^2

Comme pour le test t, la valeur calculée de χ^2 doit être comparée à une valeur théorique afin de déterminer s'il faut rejeter ou non l'hypothèse nulle. L'annexe 27.2 contient une table des valeurs théoriques de χ^2 .

- a) On détermine d'abord la valeur de p, habituellement 0,05.
- b) Ensuite, il faut calculer le nombre de degrés de liberté. Dans le cas du test du χ^2 , le nombre de degrés de liberté est relié au nombre de cellules, c'est-à-dire au nombre de groupes ou de variables que l'on compare. On le calcule en multipliant le nombre de rangées (r) moins 1 par le nombre de colonnes (c) moins 1 :

$$v = (r - 1) \times (c - 1)$$

Pour un tableau simple de deux cellules sur deux, le nombre de degrés de liberté est de 1 (c'est-à-dire $v = (2 - 1) \times (2 - 1) = 1$).

- c) Ensuite, il faut localiser la valeur de χ^2 correspondant à la valeur de p et au nombre de degrés de liberté dans la table afin de déterminer si la valeur de χ^2 est significative ou non.

3. Interprétation des résultats

Comme pour le test t, l'hypothèse nulle est rejetée si $p < 0,05$, ce qui est le cas lorsque la valeur calculée de χ^2 est supérieure à la valeur théorique de χ^2 indiquée dans la table.

Appliquons maintenant le test du khi carré aux données de l'exemple 2 (fréquentation des cliniques prénatales). On obtient les résultats suivants :

Étape 1 a)

Les fréquences théoriques pour chacune des cellules sont calculées comme suit :

$$\begin{aligned} f_t(1) &= 86 \times 80/155 = 44,4 & f_t(2) &= 69 \times 80/155 = 35,6 \\ f_t(3) &= 86 \times 75/155 = 41,6 & f_t(4) &= 69 \times 75/155 = 33,4 \end{aligned}$$

Pour simplifier, les fréquences observées et théoriques sont indiquées dans le tableau suivant :

Tableau 27.3. Fréquentation des cliniques prénatales, et fréquences observées et théoriques.

Distance de la clinique	Ont fréquenté	N'ont pas fréquenté	Total
Moins de 10 km	$f_o(1) = 51$ $f_t(1) = 44,4$	$f_o(2) = 29$ $f_t(2) = 35,6$	80
10 km ou plus	$f_o(3) = 35$ $f_t(3) = 41,6$	$f_o(4) = 40$ $f_t(4) = 33,4$	75
Total	86	69	155

Soulignons que les fréquences théoriques constituent les valeurs espérées, compte tenu du nombre total de 80 et de 75 femmes des deux groupes, **si l'hypothèse nulle**, selon laquelle il n'y a pas de différence entre les deux groupes, **était vraie**.

Étapes 1 b) à d)

$$\begin{aligned} \chi^2 &= \frac{(51-44,4)^2}{44,4} + \frac{(29-35,6)^2}{35,6} + \frac{(35-41,6)^2}{41,6} + \frac{(40-33,4)^2}{33,4} \\ &= 0,98 + 1,22 + 1,05 + 1,30 = 4,55 \end{aligned}$$

Étape 2

Comme nous avons un tableau simple de deux cellules sur deux, le nombre de degrés de liberté (ν) est de 1.

Utilisons maintenant la table des valeurs de khi carré de l'annexe 27.2. Nous avons choisi un seuil de signification de 5 % (valeur de $p = 0,05$).

Comme $\nu = 1$, nous localisons le chiffre correspondant dans la colonne où $p = 0,05$. La valeur indiquée est de 3,84. Notre valeur de 4,55 est **supérieure** à 3,84, ce qui signifie que la valeur de p est **inférieure** à 0,05 (elle est même inférieure à 0,01).

Étape 3

On peut maintenant conclure que les femmes vivant à moins de 10 km de la clinique fréquentent celle-ci plus souvent que les femmes vivant à plus de 10 km, et que cette différence est significative.

Dans votre rapport final, il est important de présenter vos données clairement et de formuler avec soin vos conclusions fondées sur des tests statistiques.

Pour l'exemple précédent, vous pouvez présenter le tableau 2 dans votre rapport et formuler vos conclusions de la manière suivante :

«Le tableau 2 indique que 64 % des femmes vivant à moins de 10 km de la clinique ont obtenu des soins prénataux pendant leur grossesse, par rapport à 47 % des femmes vivant à 10 km ou plus de la clinique la plus proche. Cette différence est significative ($\chi^2 = 4,55$; $p < 0,05$).»

Remarque

- Le test du khi carré ne peut être utilisé que si l'échantillon est assez grand. En règle générale, la taille de l'échantillon doit être d'au moins 40 et la fréquence théorique de chaque cellule devrait être d'au moins 5. Sinon, il faut utiliser la méthode exacte de Fisher (voir le chapitre 9 du manuel *Statistics at Square One* de Swinscow, mentionné à l'annexe 26.1). Si le tableau est plus grand, la fréquence théorique d'une cellule sur cinq peut être inférieure à 5.
- Contrairement au test t, le test du khi carré peut également servir à comparer plus de deux groupes. Dans ce cas, il faut un tableau comptant trois rangées et colonnes ou plus, et non seulement un tableau de deux rangées et deux colonnes.

Dans l'exemple précédent, on pourrait déterminer la différence entre trois distances : moins de 5 km, de 5 à 10 km et plus de 10 km. Ces données seraient alors présentées dans un tableau de deux cellules sur trois. Le nombre de degrés de liberté serait de $(3 - 1) \times (2 - 1) = 2$.

Formule rapide

Pour les tableaux de deux cellules sur deux, on peut employer une méthode rapide pour calculer le khi carré au lieu de suivre l'étape 1 indiquée plus haut.

Si les différents chiffres du tableau sont représentés par les lettres suivantes :

	Affection		Total
	+	-	
Exposition			
Oui	a	b	e
Non	c	d	f
Total	g	h	n

La formule rapide pour le calcul du khi carré est la suivante :

$$\chi^2 = \frac{n(ad-bc)^2}{efgh}$$

TRAVAIL EN GROUPE

Si vos données proviennent d'observations non appariées, déterminez le test d'hypothèses approprié et effectuez l'analyse nécessaire.

Annexe 27.1. Distribution du t de Student

La première colonne contient le nombre de degrés de liberté. Les autres colonnes donnent la probabilité (P) que t dépasse la valeur indiquée. Le même chiffre s'applique aux valeurs négatives.

Degrés de liberté	Valeur de t si p = 0,05	Valeur de t si p = 0,01
1	12,71	63,66
2	4,30	9,92
3	3,18	5,84
4	2,78	4,60
5	2,57	4,03
6	2,45	3,71
7	2,36	3,50
8	2,31	3,36
9	2,26	3,25
10	2,23	3,17
11	2,20	3,11
12	2,18	3,05
13	2,16	3,01
14	2,14	2,98
15	2,13	2,95
16	2,12	2,92
17	2,11	2,90
18	2,10	2,88
19	2,09	2,86
20	2,09	2,85
21	2,08	2,83
22	2,07	2,82
23	2,07	2,81
24	2,06	2,80
25	2,06	2,79
30	2,04	2,76
40	2,02	2,70
60	2,00	2,66
120	1,98	2,62
Infini	1,96	2,58

Si la valeur calculée de t (sans tenir compte du signe) est **supérieure** à la valeur indiquée dans la table, la valeur de p est **inférieure** à celle qui figure dans le haut de la colonne.

Dans ce cas, l'hypothèse nulle, selon laquelle il n'y a pas de différence, est **rejetée**, et on peut conclure qu'il existe une différence significative.

Annexe 27.2. Table des valeurs de χ^2

Degrés de liberté	Valeur de χ^2 si p = 0,05	Valeur de χ^2 si p = 0,01
1	3,84	6,63
2	5,99	9,21
3	7,81	11,34
4	9,49	13,28
5	11,07	15,09
6	12,59	16,81
7	14,07	18,48
8	15,51	20,09
9	16,92	21,67
10	18,31	23,21
11	19,68	24,72
12	21,03	26,22

Si la valeur calculée de χ^2 est **supérieure** à la valeur indiquée dans la table, la valeur de p est **inférieure** à celle qui figure dans le haut de la colonne.

Dans ce cas, l'hypothèse nulle, selon laquelle il n'y a pas de différence, est **rejetée**, et on peut conclure qu'il existe une différence significative.

Annexe 27.3. Les variables confusionnelles : Test du khi carré de Mantel-Haenszel

Dans le tableau 27.4 sont présentés les résultats d'une étude sur l'incidence de la schistosomiase parmi les habitants de deux villages.

Tableau 27.4. Incidence de la schistosomiase dans les villages A et B.

	Village A	Village B	Total
Positif	80 (32 %)	80 (32 %)	160
Négatif	170 (68 %)	170 (68 %)	340
Total	250 (100 %)	250 (100 %)	500

Il semble que l'incidence de la schistosomiase soit la même dans les deux villages (32 %).

Cependant, les chercheurs soupçonnent la présence d'une variable confusionnelle. Ils séparent donc le tableau 27.4 en deux tableaux (27.5 et 27.6). En additionnant les chiffres des tableaux 27.5 et 27.6, on obtient le tableau 27.4.

Tableau 27.5. Incidence de la schistosomiase chez les enfants de 5 à 19 ans dans les villages A et B.

	Village A	Village B	Total
Positif	37 (62 %)	73 (38 %)	110
Négatif	23 (38 %)	117 (62 %)	140
Total	60 (100 %)	190 (100 %)	250

$\chi^2 = 9,08$; 1 degré de liberté; $p < 0,01$.

Tableau 27.6. Incidence de la schistosomiase chez les personnes de 20 ans et plus dans les villages A et B.

	Village A	Village B	Total
Positif	43 (23 %)	7 (12 %)	50
Négatif	147 (77 %)	53 (88 %)	200
Total	190 (100 %)	60 (100 %)	250

$\chi^2 = 2,78$; 1 degré de liberté; $p > 0,05$.

D'après les tableaux 27.5 et 27.6, on peut conclure que :

- dans chaque catégorie d'âge, l'incidence de la schistosomiase est plus élevée dans le village A que dans le village B;
- l'incidence de la schistosomiase est plus élevée chez les enfants que chez les adultes;
- dans le village A, il y a relativement peu d'enfants et beaucoup d'adultes par rapport au village B.

L'âge est considéré comme une variable confusionnelle parce qu'il est relié à la variable qui nous intéresse (l'incidence de la schistosomiase) et aux groupes comparés (habitants du village A ou B).

Cet exemple illustre un élément très important de l'analyse des données : il peut être trompeur de réunir des données différentes. Dans cet exemple, le regroupement des catégories d'âge a caché une différence importante. Dans d'autres situations, le regroupement des données peut laisser croire qu'une différence ou une association existe alors que ce n'est pas le cas, ou que cette différence ou association est en fait le contraire de celle qu'on relève.

Il est donc important d'analyser les données séparément pour chaque groupe d'âge. Les valeurs appropriées de χ^2 (avec correction de continuité) pour comparer l'incidence de la maladie dans les villages A et B figurent aux tableaux 27.5 et 27.6. La différence d'incidence est significative chez les enfants, mais pas chez les adultes.

Test du χ^2 de Mantel-Haenszel

Il est souvent utile d'effectuer un test qui permet de réunir les résultats tirés de chacun des tableaux tout en tenant compte du facteur confusionnel (l'âge dans notre exemple). Le test du χ^2 de Mantel-Haenszel permet de le faire.

Pour chacun des tableaux de deux cellules sur deux, nous utilisons la notation suivante :

a	b	e
c	d	f
g	h	n

On calcule les trois valeurs suivantes pour chaque tableau, puis on en fait la somme :

1. La valeur observée de a, $f_o(a)$;
2. La valeur théorique de a, $f_t(a)$, qui est égale à eg / n ;
3. La variance de a, $s^2(a)$, qui est égale à $efgh / (n^2 (n - 1))$.
La valeur de χ^2 avec correction de continuité est de

$$\chi^2 = \frac{(f_o(a) - f_t(a) - 0,5)^2}{s^2(a)} \text{ avec 1 degré de liberté}$$

Dans l'exemple, les calculs sont les suivants :

	$f_o(a)$	$f_t(a) = eg / n$	$s^2(a) = efgh / (n^2 (n - 1))$
Enfants	37	26,4	$110 \times 140 \times 60 \times 190 / 250^2 \times 249 = 11,3$
Adultes	43	38	$50 \times 200 \times 190 \times 60 / 250^2 \times 249 = 7,3$
Total	80	64,4	18,6

$$\chi^2 = \frac{(80 - 64,4 - 0,5)^2}{18,6} = \frac{15,1^2}{18,6} = 12,25 \quad (p < 0,001)$$

On peut donc conclure que l'incidence de la schistosomiase est différente dans les villages A et B et que cette différence est significative (on ne pouvait arriver à cette conclusion à partir du tableau 27.4, où les données sur les adultes et les enfants étaient réunies).

Validité du test du χ^2 de Mantel-Haenszel

Le test du χ^2 de Mantel-Haenszel est un test approximatif. Il est plus difficile de déterminer sa validité que celle du test du khi carré ordinaire. Il faut calculer deux valeurs supplémentaires pour chaque tableau et en faire la somme :

1. $\min(e, g)$, c'est-à-dire la plus petite des valeurs e et g;
2. $\max(0, g - f)$, c'est-à-dire 0 si g est plus petit ou égal à f et g - f si g est plus grand.

Ces deux sommes doivent s'écarter du total des valeurs théoriques, $f_i(a)$, d'au moins 5. Voici le calcul pour l'exemple précédent :

	min (e, g)	max (0, g - f)
Enfants	60	0
Adultes	50	0
Total	110	0

Les deux sommes sont de 110 et 0, qui s'écartent de 64,4 ($f_i(a)$) de plus de 5. On peut donc utiliser le test de Mantel-Haenszel.

Notes du formateur

Module 27 : DIFFÉRENCES ENTRE DES GROUPES PREMIÈRE PARTIE - ANALYSE D'OBSERVATIONS NON APPARIÉES

Durée et méthode d'enseignement

1 heure	Introduction et discussion
3 heures*	Travail en groupe

Introduction et discussion

- Nous vous conseillons de présenter le test t ou le test du χ^2 immédiatement après avoir terminé le **module 26**; les participants auront alors une meilleure idée de ce qu'est un test d'hypothèses et de la façon de s'en servir. Il est probable que toutes les équipes de recherche utiliseront le test du χ^2 , mais que certaines n'aient pas besoin du test t. Par conséquent, vous pourriez présenter le test du χ^2 avec le **module 26** et le test t lors d'une autre séance.
- Expliquez les tests d'hypothèses lentement, étape par étape, de façon à ne pas effrayer les participants qui ont peu d'expérience. Soulignez qu'il n'est pas important de comprendre pourquoi les valeurs de t et de χ^2 sont calculées de cette façon (bien que le mode de calcul ne soit pas toujours fondé sur des notions précises, comme pour ce qui est des degrés de liberté). Il est suffisant pour eux de savoir comment faire le calcul. Attention aux formules : ne les présentez qu'après avoir expliqué le calcul étape par étape.
- Si vous voulez, vous pouvez prendre des exemples tirés des travaux des équipes elles-mêmes au lieu des exemples 1 et 2 du module. Cependant, utilisez des exemples simples, c.-à-d. des tableaux de deux cellules sur deux et de petits chiffres.
- Veillez à bien expliquer comment utiliser la table des valeurs de t et la table des valeurs de χ^2 et comment interpréter les résultats. Donnez aux participants le temps de résoudre eux-mêmes les problèmes avant de leur donner la réponse.
- Portez une attention particulière à la formulation des conclusions fondées sur des tests d'hypothèses, que les résultats soient significatifs ou non.
- L'**annexe 27.3** ne doit être présentée que si les participants sont très avancés. Les équipes peuvent la lire et utiliser au besoin le test dont on y traite.

Travail en groupe

Assurez-vous que chaque participant a la chance de faire un test statistique lui-même.

**Formation à la recherche sur les systèmes de santé
Volume 2, 2^e partie : Analyse de données et rédaction de rapports**

Module 28

**DIFFÉRENCES ENTRE DES GROUPES
DEUXIÈME PARTIE — ANALYSE D'OBSERVATIONS
APPARIÉES**

ÉTAPES DE L'ANALYSE DE DONNÉES ET DE LA RÉDACTION DE RAPPORTS

Questions à poser	Étapes à suivre*	Éléments importants de l'étape
Quelles données ont été recueillies pour chaque objectif de recherche? Les données sont-elles complètes et exactes?	Préparer les données pour l'analyse	Revoir les activités sur le terrain Compiler les données pour chaque objectif ou population étudiée Classer les données et en vérifier la qualité Vérifier les imprimés
À quoi ressemblent les données?	Décrire les variables	Distributions Figures, moyennes
Comment peut-on résumer les données pour en faciliter l'analyse?	Faire le recoupement des données quantitatives Résumer les données qualitatives	Faire le recoupement des données par rapport aux objectifs Illustrations, texte
Pour les données qualitatives : la recherche a-t-elle pour but de décrire, de comparer ou de trouver des associations?	Déterminer le type d'analyse statistique nécessaire	Examiner les objectifs, le type d'étude et les variables Faire une description statistique des variables Choisir les tests d'hypothèses
1. Comment peut-on décrire les données?	Analyser des observations appariées et non appariées	**Test t de Student **Test t pour observations appariées **Test du khi carré **Test du khi carré de McNemar
2. Comment peut-on déterminer les différences entre les groupes?	Évaluer les associations entre les variables	**Diagramme de dispersion **Droite de régression et coefficient de corrélation **Risque relatif, risque relatif approché
3. Comment peut-on déterminer les associations entre les variables?	Rédiger le rapport et formuler des recommandations	Préparer le canevas du rapport Rédiger les 1 ^{re} et 2 ^e versions Faire un sommaire des constatations Faire un sommaire des conclusions pour chaque objectif Formuler des recommandations Préparer un résumé analytique
Comment le rapport doit être rédigé?	Présenter un sommaire et rédiger un plan de mise en oeuvre des recommandations	Discuter des sommaires avec différents groupes cibles Discuter du programme de mise en oeuvre
Comment doit-on présenter et diffuser les constatations et les recommandations?		

* Il n'est pas nécessaire de faire ces étapes dans l'ordre indiqué. L'ordre peut être adapté aux besoins de l'atelier.
** Ces éléments sont facultatifs; ils peuvent être omis s'ils ne sont pas pertinents au cours.

**Module 28 : DIFFÉRENCES ENTRE DES GROUPES
DEUXIÈME PARTIE - ANALYSE D'OBSERVATIONS APPARIÉES**

OBJECTIFS

À la fin de la présente séance, vous devriez pouvoir :

1. Déterminer les recherches qui nécessitent l'appariement des sujets;
2. Choisir et utiliser les tests d'hypothèses appropriés dans le cadre d'études utilisant des données appariées.

I. Introduction

II. Test t pour observations appariées

III. Test du khi carré de McNemar

I. INTRODUCTION

La notion des observations appariées a été abordée au **module 23**. L'exemple présenté dans ce module portait sur des données nominales. Cependant, il est également possible de faire des observations appariées lorsque les données sont numériques. Le présent module décrit les tests les plus courants auxquels on soumet les observations appariées :

- le **test t pour observations appariées** pour les données numériques;
- le **test du khi carré de McNemar** pour les données nominales.

Qu'entend-on par appariement?

La notion de l'appariement des sujets est illustrée dans les exemples suivants :

Exemple 1

Un chercheur veut découvrir si des étudiants auxquels on enseigne au moyen de matériel audio-visuel obtiennent, en moyenne, des notes plus élevées que ceux auxquels on enseigne sans ce matériel. Pour réduire l'effet des variables confusionnelles telles que la situation sociale, la connaissance préalable des sujets et le quotient intellectuel, chaque étudiant du groupe utilisant du matériel audio-visuel et un étudiant de l'autre groupe sont appariés pour ces variables.

Exemple 2

Au cours d'un sondage sur l'alimentation, un exercice de contrôle de qualité est effectué pour vérifier que les deux observateurs mesurent le poids des enfants de la même façon. Dans ce cas, chaque enfant est couplé avec lui-même.

Exemple 3

Une équipe de recherche compare des comptes d'oeufs de schistosomiase dans deux villages. Elle constate que ce compte varie selon l'âge et le sexe. Elle veut donc s'assurer que les échantillons sont comparables quant à l'âge et au sexe en sélectionnant les sujets par paires, un sujet provenant de chaque village.

II. TEST T POUR OBSERVATIONS APPARIÉES

Au **module 27**, une comparaison de la moyenne d'échantillons a été effectuée pour des observations numériques non appariées au moyen du **test t**. Lorsqu'on a des observations appariées (ou couplées), on compare la moyenne des échantillons au moyen d'un test t modifié connu sous le nom de **test t pour observations appariées**.

Dans ce test, on utilise une seule série de différences entre les observations appariées plutôt que les deux séries initiales. La valeur de t se calcule alors comme suit :

$$t = \frac{\text{différence moyenne}}{\text{erreur-type}}$$

Le nombre de degrés de liberté est égal à la taille de l'échantillon moins 1 (ou au nombre d'observations appariées moins 1).

Pour interpréter le résultat, on utilise la même table des valeurs de t que pour le test t employé pour les observations non appariées (voir l'annexe 27.1).

Appliquons le test t pour observations appariées au sondage sur l'alimentation mentionné dans l'exemple 2. En voici les résultats :

Tableau 28.1. Résultats d'un exercice de contrôle de qualité effectué dans le cadre d'un sondage sur l'alimentation.

Enfant	Poids (kg)		Différence entre A et B (kg)
	Observateur A	Observateur B	
1	18,6	17,7	0,9
2	17,1	14,5	2,6
3	14,3	12,4	1,9
4	23,2	20,7	2,5
5	18,4	16,8	1,6
6	14,9	14,4	0,5
7	16,6	14,1	2,5
8	14,8	17,1	-2,3
9	21,5	21,2	0,3
10	24,6	21,9	2,7
11	17,4	16,6	0,8
12	15,7	13,6	2,1
13	16,1	14,5	1,6
14	12,9	11,2	1,7
15	12,3	16,0	-3,7
16	19,4	20,4	-1,0
17	19,3	17,5	1,8
18	24,8	22,2	2,6
19	14,3	15,1	-0,8
20	13,4	10,9	2,5

Cette étude aurait l'hypothèse nulle suivante : si les observateurs A et B mesuraient le poids de tous les enfants de la population parmi laquelle les 20 enfants de l'échantillon ont été choisis, il n'y aurait aucune différence entre les mesures. En d'autres mots, la différence moyenne entre A et B serait de zéro.

Nous pouvons considérer cette série de 20 différences (colonne «Différence entre A et B») comme un échantillon de la population de différences qui aurait été obtenue si les observateurs avaient mesuré l'ensemble de la population.

Pour effectuer le test d'hypothèses, la valeur de t doit être calculée et comparée à la valeur théorique de la table des valeurs de t afin de déterminer la probabilité que ce résultat soit attribuable au hasard.

Voici la marche à suivre :

1. Calculer la différence moyenne pour l'échantillon. Il s'agit de la somme des différences divisée par le nombre de mesures :

$$\text{Différence moyenne} = 1,04$$

2. Calculer l'écart-type des différences (**module 25**).

$$\text{Écart-type} = 1,77$$

3. Calculer l'erreur-type (**module 25**) :

$$\text{Erreur-type} = \frac{\text{Écart-type}}{\sqrt{\text{taille de l'échantillon}}} = \frac{1,77}{\sqrt{20}} = 0,40$$

4. La valeur de t est égale à la différence moyenne divisée par l'erreur-type :

$$t = \frac{1,04}{0,40} = 2,60$$

5. Consulter la table des valeurs de t à l'**annexe 27.1**.

Le nombre de degrés de liberté est égal à la taille de l'échantillon (c.-à-d. le nombre de paires d'observations) moins 1, soit $20 - 1 = 19$.

Selon la table, la probabilité est $< 0,05$, ce qui nous permet de conclure qu'il existe une différence significative entre les observateurs.

III. TEST DU KHI CARRÉ DE MCNEMAR

Le **test du khi carré de McNemar** s'utilise avec des données NOMINALES pour comparer des PROPORTIONS d'observations appariées. Soulignons que la disposition du tableau est différente de celle qu'on utilise avec des échantillons non appariés.

Le **tableau 28.2** illustre les résultats d'une étude cas-témoins menée dans le but de déterminer les causes d'une épidémie de choléra en Italie. Pour chaque cas de choléra constaté à l'hôpital, un sujet du même sexe, de la même catégorie d'âge (à la dizaine près) et du même quartier a été recherché.

Cependant, la disposition du **tableau 28.2** n'est pas adéquate, car elle ne tient pas compte du fait que les malades et les personnes en bonne santé ont été appariés.

La bonne disposition figure au **tableau 28.3**.

Tableau 28.2. Consommation de poisson ou de fruits de mer par des personnes atteintes du choléra et des personnes en bonne santé dans un délai de 5 jours avant le début de la maladie (disposition inadéquate).

	Malades	Personnes en bonne santé	Total
Ont consommé	42 (55 %)	15 (20 %)	57
N'ont pas consommé	34 (45 %)	61 (80 %)	95
Total	76 (100 %)	79 (100 %)	152

Table 28.3. Consommation de poisson ou de fruits de mer par paires de sujets atteints du choléra et de personnes en bonne santé dans un délai de 5 jours avant le début de la maladie (bonne disposition).

Personnes en bonne santé	Malades		Total
	Ont consommé	N'ont pas consommé	
Ont consommé	12 (16 %)	3 (4 %)	15
N'ont pas consommé	30 (39 %)	31 (41 %)	61
Total	42	34	76 (100 %)

Source : Baine, W.B., Mazzotti, M., Greco, D., Izzo, E., Zampieri, A., Angioni, G., Di Gioia, M., Gangarosa, E.J., Pocchiari, S. 1974. Epidemiology of cholera in Italy in 1973, Lancet, ii (déc.), 1370-1374.

Comment doit-on interpréter le **tableau 28.3**?

Chez 12 paires de sujets, tant le malade que la personne en bonne santé ont consommé des fruits de mer, et chez 31 paires, ni le malade ni la personne en bonne santé n'en a consommé. Par conséquent, ces 43 paires ne nous permettent pas de déterminer si la consommation de fruits de mer constitue un facteur de risque du choléra. Cependant, chez 30 paires (39 %), le malade a mangé des fruits de mer, alors que la personne en bonne santé n'en a pas mangé, mais chez seulement 3 paires (4 %), la personne en bonne santé a mangé des fruits de mer mais pas le malade. Il semble donc que la consommation de fruits de mer constitue un facteur de risque du choléra.

Avant d'accepter cette conclusion, nous devons effectuer un test d'hypothèses pour évaluer la probabilité que ces résultats sont attribuables uniquement au hasard ou à la variation de l'échantillonnage. Dans ce cas, le test d'hypothèses approprié est le **test du khi carré de McNemar** :

$$\chi^2 = \frac{(|r - s| - 1)^2}{r + s} \text{ avec 1 degré de liberté}$$

où r est le nombre de réponses +-,
 s est le nombre de réponses -+,
 $|r - s|$ est la différence entre r et s exprimée sous forme de chiffre positif, sans égard au fait que r ou s soit la variable la plus grande.

Remarque

Le test du χ^2 n'est valable que si $(r + s)$ est supérieur à 10.

Il est possible d'appliquer ce test aux données de notre exemple car $r + s$ ($30 + 3$) est supérieur à 10.

Le khi carré est calculé comme suit :

$$\chi^2 = \frac{(30 - 3 - 1)^2}{30 + 3} = \frac{26^2}{33} = 20,5 \text{ avec 1 degré de liberté.}$$

D'après la table des valeurs de χ^2 (**annexe 27.2**), $p < 0,01$. Cela signifie que s'il n'existait aucune association réelle entre la consommation de fruits de mer et le fait de contracter le choléra (hypothèse nulle), la probabilité d'obtenir un résultat aussi contradictoire que celui du **tableau 28.3** serait inférieure à 1 pour 100. Par conséquent, nous rejetons l'hypothèse nulle et concluons que la consommation de fruits de mer constitue un facteur de risque du choléra¹.

TRAVAIL EN GROUPE

Si vos données ont été recueillies par des observations appariées, choisissez le test statistique approprié et effectuez les calculs et l'analyse nécessaires.

¹Une partie du présent module est tirée d'un cours donné dans le cadre du programme de maîtrise en sciences de la London School of Hygiene and Tropical Medicine.

Notes du formateur

**Module 28 : DIFFÉRENCES ENTRE DES GROUPES
DEUXIÈME PARTIE - ANALYSE D'OBSERVATIONS APPARIÉES****Durée et méthode d'enseignement**

1/2 heure	Introduction et discussion
2 heures	Travail en groupe

Introduction et discussion

- Si aucune équipe n'a d'observations appariées et si les participants ont peu d'expérience en statistique, il n'est pas nécessaire de présenter ce module.
- Expliquez très lentement et étape par étape comment calculer les valeurs de t et de χ^2 . Il est plus important que les participants comprennent comment faire les calculs que de leur expliquer pourquoi il faut les faire de cette façon.
- Laissez aux participants le temps de lire le **tableau 28.3**. Il s'agit du premier tableau où les chiffres représentent des **paires** d'observations.
- Assurez-vous que les participants savent comment utiliser les tables des valeurs de t et de χ^2 et interpréter les résultats.

**Formation à la recherche sur les systèmes de santé
Volume 2, 2^e partie : Analyse de données et rédaction de rapports**

Module 29

**ASSOCIATIONS ENTRE VARIABLES :
RÉGRESSION ET CORRÉLATION**

ÉTAPES DE L'ANALYSE DE DONNÉES ET DE LA RÉDACTION DE RAPPORTS

Questions à poser	Étapes à suivre*	Éléments importants de l'étape
Quelles données ont été recueillies pour chaque objectif de recherche? Les données sont-elles complètes et exactes?	Préparer les données pour l'analyse	Revoir les activités sur le terrain Compiler les données pour chaque objectif ou population étudiée Classer les données et en vérifier la qualité Vérifier les imprimés
À quoi ressemblent les données?	Décrire les variables	Distributions Figures, moyennes
Comment peut-on résumer les données pour en faciliter l'analyse?	Faire le recoupement des données quantitatives	Faire le recoupement des données par rapport aux objectifs Illustrations, texte
	Résumer les données qualitatives	
Pour les données qualitatives : la recherche a-t-elle pour but de décrire, de comparer ou de trouver des associations?	Déterminer le type d'analyse statistique nécessaire	Examiner les objectifs, le type d'étude et les variables Faire une description statistique des variables Choisir les tests d'hypothèses
1. Comment peut-on décrire les données?		
2. Comment peut-on déterminer les différences entre les groupes?	Analyser des observations appariées et non appariées	**Test t de Student **Test t pour observations appariées **Test du khi carré **Test du khi carré de McNemar
3. Comment peut-on déterminer les associations entre les variables?	Évaluer les associations entre les variables	**Diagramme de dispersion **Droite de régression et coefficient de corrélation **Risque relatif, risque relatif approché
Comment le rapport doit être rédigé?	Rédiger le rapport et formuler des recommandations	Préparer le canevas du rapport Rédiger les 1 ^{re} et 2 ^e versions Faire un sommaire des constatations Faire un sommaire des conclusions pour chaque objectif Formuler des recommandations Préparer un résumé analytique
Comment doit-on présenter et diffuser les constatations et les recommandations?	Présenter un sommaire et rédiger un plan de mise en oeuvre des recommandations	Discuter des sommaires avec différents groupes cibles Discuter du programme de mise en oeuvre

* Il n'est pas nécessaire de faire ces étapes dans l'ordre indiqué. L'ordre peut être adapté aux besoins de l'atelier.

** Ces éléments sont facultatifs; ils peuvent être omis s'ils ne sont pas pertinents au cours.

Module 29 : ASSOCIATIONS ENTRE VARIABLES : RÉGRESSION ET CORRÉLATION

OBJECTIFS

À la fin de la présente séance, vous devriez pouvoir :

1. Illustrer la relation entre deux variables numériques dans un diagramme de dispersion;
2. Interpréter une droite de régression;
3. Calculer et interpréter le coefficient de corrélation;
4. Soumettre le coefficient de corrélation à un test d'hypothèses.

- I. Introduction
- II. Diagrammes de dispersion
- III. Relations linéaires : droite de régression
- IV. Coefficient de corrélation
- V. Caractère significatif du coefficient de corrélation

I. INTRODUCTION

Lorsqu'il s'agit d'étudier les associations entre variables, il faut distinguer les données nominales, ordinales et numériques (voir le tableau 26.1 du **module 26**).

Le **module 30** traite des associations entre données nominales, notamment dans le cas d'études cas-témoins.

Pour ce qui est des associations entre données ordinales, pour lesquelles on peut calculer et déterminer le caractère significatif du **coefficient de corrélation des rangs de Spearman** ou de **Kendall**, consultez un manuel de statistique (voir l'**annexe 26.1**).

Dans le présent module, nous étudierons les associations entre données numériques lorsqu'on soupçonne la présence d'une relation linéaire.

II. DIAGRAMMES DE DISPERSION

La première étape de l'examen d'une relation entre deux variables numériques portant sur les mêmes sujets consiste à tracer un **DIAGRAMME DE DISPERSION**.

Exemple 1

Dans le cadre d'une étude sur l'alimentation tenue dans un important district rural, un échantillon de 20 enfants de 5 ans ont été pesés et le revenu de leur famille estimé. Voici les données recueillies.

Tableau 29.1. Poids de 20 enfants de 5 ans et revenu de leur famille.

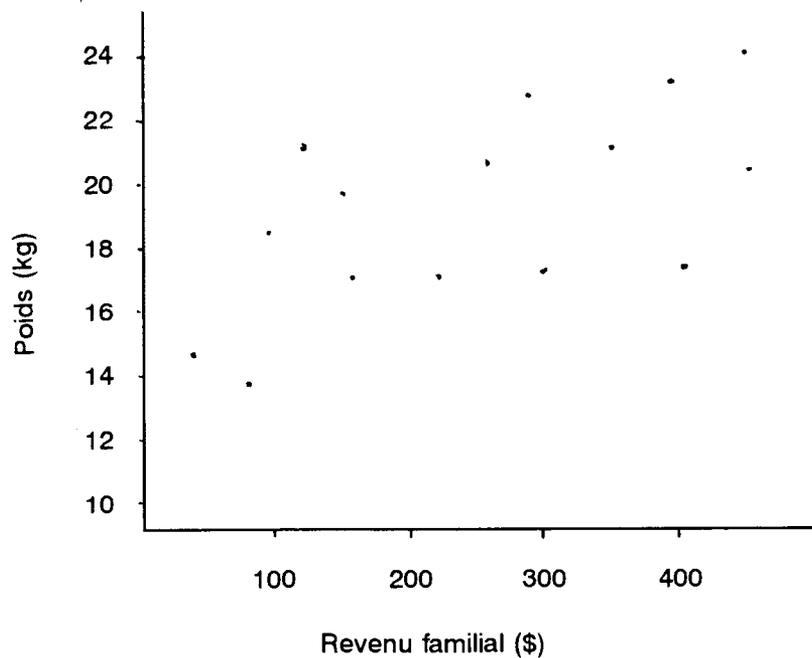
Revenu familial (\$/an)	Poids (kg)	Revenu familial (\$/an)	Poids (kg)
130	15,5	225	18,1
200	19,8	95	17,4
345	21,5	130	17,9
245	16,8	330	17,0
155	12,6	295	18,7
300	16,6	170	16,0
360	18,1	250	18,2
105	18,7	355	16,4
80	13,1	220	15,4
275	20,1	175	17,6

L'objectif de la recherche consistait à déterminer si le poids et le revenu familial étaient reliés dans cet échantillon d'enfants. Il serait possible de diviser les enfants en deux catégories de revenu, soit une catégorie à revenu élevé (p. ex., 200 \$ ou plus) et une catégorie à revenu faible (p. ex., moins de 200 \$), et de calculer puis de comparer le poids moyen dans chaque catégorie. Pour déterminer si la différence entre les deux catégories est significative, il faut faire un test t (voir le **module 27**).

Après avoir fait cette analyse, vous pourriez conclure que les enfants provenant de familles à faible revenu présentaient un poids inférieur, en moyenne, à celui des enfants provenant de familles à revenu élevé. Cependant, il serait plus intéressant de tenir compte de toutes les mesures et de déterminer si les variables «revenu familial» et «poids des enfants de cinq ans» sont associées (c.-à-d. si nous pouvons prédire la valeur d'une variable en sachant la valeur de l'autre variable).

On peut tracer le diagramme de dispersion suivant :

Figure 29.1. Poids et revenu familial de 20 enfants de 5 ans.



Remarques sur la préparation des diagrammes de dispersion

1. Pour déterminer l'association d'une variable dépendante avec une variable indépendante, on place normalement la variable dépendante sur l'axe vertical (l'axe des y) et la variable indépendante sur l'axe horizontal (l'axe des x). Parfois, il est difficile de déterminer quelle variable est dépendante, auquel cas le choix de l'axe est arbitraire.
2. Choisir les échelles de sorte que les points occupent une partie raisonnable du diagramme.
3. Si l'un des axes ne commence pas par zéro, l'indiquer clairement en «séparant» l'axe (comme il est indiqué pour l'axe des poids dans l'exemple précédent).
4. Titrer clairement chaque axe.
5. Les points doivent être assez gros pour être facilement visibles.

III. RELATIONS LINÉAIRES : DROITE DE RÉGRESSION

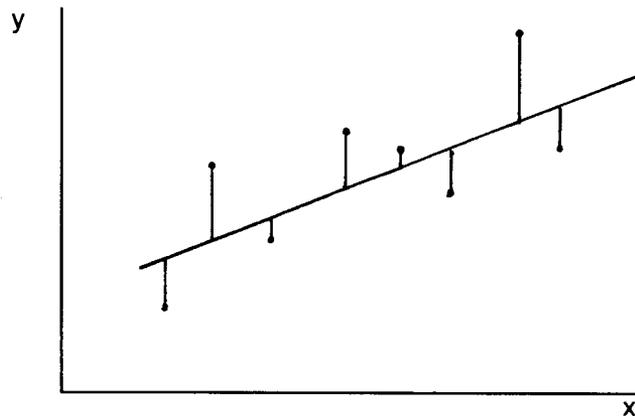
Dans le diagramme de dispersion ci-dessus, le poids semble augmenter avec la hausse du revenu familial. On peut tracer une droite à travers le nuage de points pour résumer simplement la relation entre ces deux variables. On peut le faire sans calcul, au moyen d'une règle transparente; cependant, cette technique est plutôt subjective. La **MÉTHODE DES MOINDRES CARRÉS** permet d'obtenir une droite le plus objectivement possible, au moyen de la technique décrite ci-dessous.

Toute droite tracée dans un graphique peut être représentée par l'équation suivante :

$$y = a + bx$$

Chaque point de la droite a une valeur x et une valeur y , et l'équation nous renseigne sur la relation entre ces valeurs. Des droites différentes correspondent à des valeurs différentes de a et de b . La valeur de a nous donne L'ORDONNÉE À L'ORIGINE de la droite (c.-à-d. la distance entre le point où la droite touche l'axe des y et zéro), et b constitue la PENTE de la droite.

Pour tracer la droite, il faut déterminer la valeur de a et de b . En gros, on les choisit de façon à minimiser l'écart entre les points et la droite. (Plus précisément, la valeur de a et de b doit permettre de minimiser la somme des carrés de ces écarts, d'où le nom «méthode des moindres carrés».)



L'annexe 29.1 explique comment calculer les valeurs de a et de b à partir des données. Certaines calculatrices et certains logiciels le font automatiquement.

Dans notre exemple, nous calculons, au moyen d'une calculatrice appropriée :

$$a = 15,09 \quad b = 0,00984$$

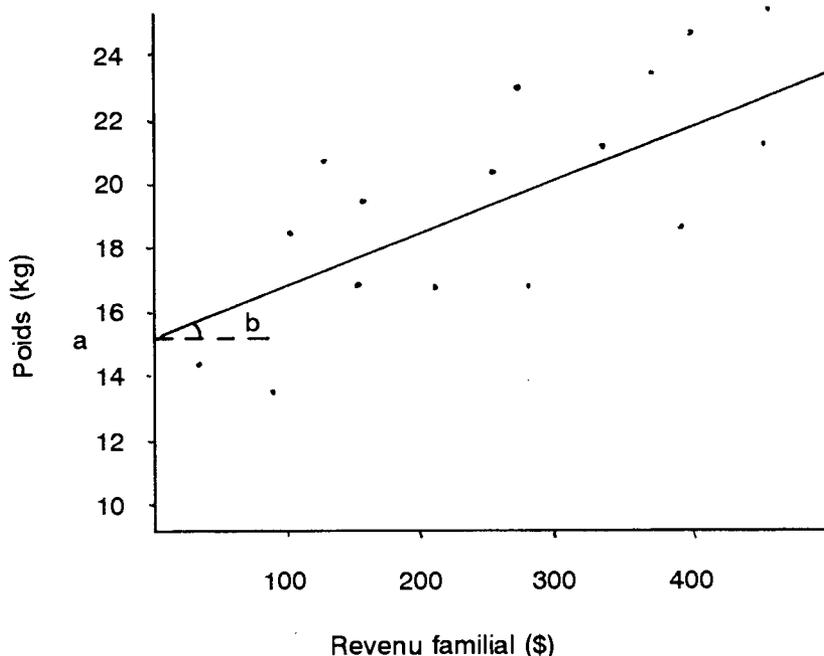
L'équation de notre droite est donc

$$y = 15,09 + 0,00984x$$

Pour tracer cette droite dans le diagramme de dispersion, on choisit deux valeurs de x et on trouve les valeurs correspondantes de y en utilisant l'équation. On trace ensuite les deux points dans le diagramme et on les relie par une ligne droite.

Exemple

$$\begin{aligned} x = 0 & \text{ donne } y = 15,09 \\ x = 400 & \text{ donne } y = 15,09 + (0,00984 \times 400) = 19,03 \end{aligned}$$



Cette droite désigne la **RÉGRESSION LINÉAIRE** du poids par rapport au revenu familial.

Interprétation d'une droite de régression

La droite de régression permet d'évaluer la valeur **moyenne** de y pour une valeur donnée de x . Par exemple, elle révèle que les enfants dont la famille a un revenu de 200 \$ par année pèsent **en moyenne** environ 17 kg. Certains pèsent moins, d'autres plus.

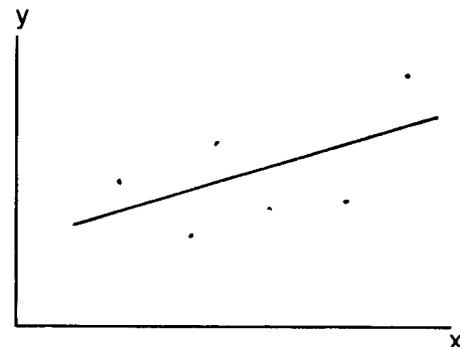
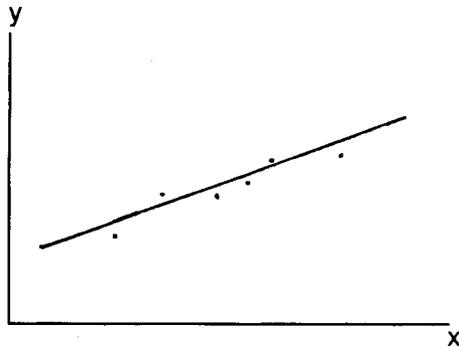
La pente b , appelée **COEFFICIENT DE RÉGRESSION**, nous donne l'augmentation de la valeur moyenne de y correspondant à une augmentation donnée de x . Dans notre exemple, le poids moyen augmente de 0,00984 kg (environ 10 g) pour chaque hausse de 1 \$ du revenu familial (soit environ 1 kg de poids par 100 \$ d'augmentation).

Mises en garde :

- On ne doit tenter de tracer une droite dans le diagramme de dispersion que si ce dernier suggère que la relation entre les deux variables est à peu près linéaire. On peut recourir à des méthodes plus complexes pour faire correspondre des courbes aux données.
- Il est périlleux d'**extrapoler** la droite de régression au-delà de l'intervalle de données. Dans notre exemple, l'extrapolation de la droite jusqu'à un revenu de 2 000 \$ par année donnerait un poids moyen de 34,8 kg, ce qui est évidemment absurde.

IV. COEFFICIENT DE CORRÉLATION

Prenons les deux diagrammes de dispersion suivants :

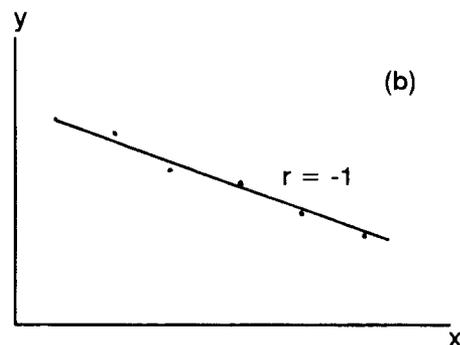
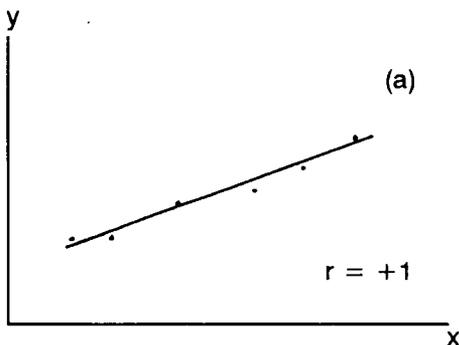


Les coefficients de régression b (c.-à-d. la pente des droites) sont identiques dans ces deux exemples, mais les points sont plus éloignés de la droite dans le deuxième. Il est évident que le lien entre les variables y et x est beaucoup plus étroit dans le premier diagramme.

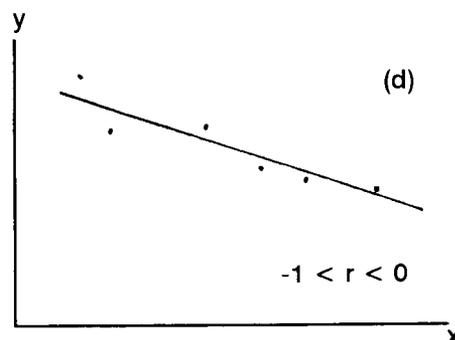
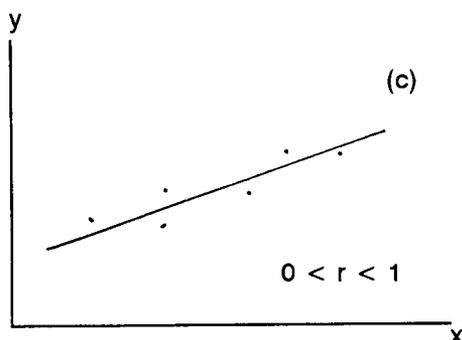
Le but du COEFFICIENT DE CORRÉLATION DE PEARSON (r) consiste à mesurer la précision de la relation linéaire entre deux variables.

Des calculatrices et logiciels informatiques calculent r automatiquement, en même temps que a et b . L'annexe 30.1 montre comment calculer r . Le coefficient de corrélation revêt les propriétés suivantes :

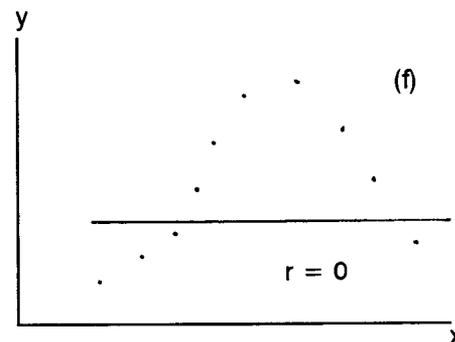
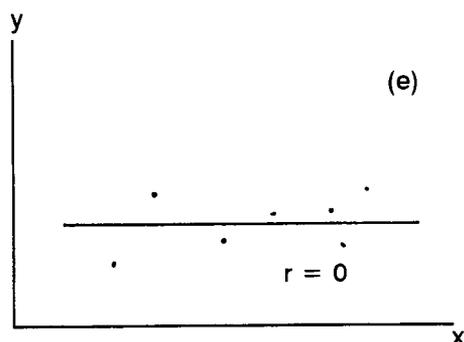
1. Quelle que soit la série de données, r se situe entre -1 et 1 .
2. Si $r = 1$ ou -1 , la relation est parfaite, c'est-à-dire que tous les points se trouvent exactement sur la droite de régression. Si $r = 1$, la variable y augmente proportionnellement à la variable x (la pente monte). Si $r = -1$, la variable y baisse proportionnellement à x (la pente baisse).



3. Si r se trouve entre 0 et 1, la droite de régression monte, mais les points sont dispersés autour d'elle. Plus r se rapproche de 1, plus les points se rapprochent de la droite. C'est la même chose pour les valeurs négatives de r , entre 0 et -1, mais dans ce cas, la pente baisse.



4. Si $r = 0$, il n'y a pas de relation **linéaire** entre y et x . Peut-être qu'il n'y a alors aucune relation entre ces deux variables (c.-à-d. que la valeur de x ne nous dit rien sur la valeur de y). Cependant, nous obtiendrions également $r = 0$ s'il existait une relation **non linéaire** entre y et x (voir le diagramme f).



5. On peut interpréter r en affirmant que son carré (r^2) mesure la proportion de la variabilité de la variable y qui découle de sa relation linéaire avec la variable x .

Si l'on revient à notre exemple du poids et du revenu familial, une calculatrice nous donne :

$$r = 0,414$$

Ce résultat est positif (ce qui donne une pente positive), mais il est très éloigné de 1 (ce qui laisse croire que les données seront très dispersées autour de la droite).

V. CARACTÈRE SIGNIFICATIF DU COEFFICIENT DE CORRÉLATION

La droite de régression et la valeur r ont été calculées à partir d'un échantillon de seulement 20 enfants. Les résultats peuvent donc faire l'objet d'une erreur d'échantillonnage et s'écartent probablement de la vraie droite de régression et de la vraie valeur de r , que nous aurions obtenue si nous avions étudié tous les enfants de 5 ans du district.

Il s'agit donc de savoir s'il existe vraiment un lien entre le poids et le revenu. Peut-être que dans la population entière d'enfants de 5 ans, le diagramme de dispersion ressemblerait au diagramme e) ci-dessus (pas de corrélation entre y et x), et que la corrélation positive constatée d'après notre échantillon serait due au hasard.

Pour déterminer si c'est le cas, il faut soumettre r à un test d'hypothèses. L'hypothèse nulle est la suivante : il n'existe aucune relation linéaire entre y et x . Le calcul se fait par la formule suivante :

$$t = r \text{ fois } \sqrt{\frac{n-2}{1-r^2}}$$

On compare ensuite cette valeur de t à celle d'une table de distribution de t avec $(n - 2)$ degrés de liberté, où n représente le nombre d'observations.

Selon notre exemple, $n = 20$ $r = 0,414$

$$\text{Donc, } t = 0,414 \times \sqrt{\frac{18}{1 - 0,414^2}} = 1,93$$

Comme $t_{18,0,05} = 2,10$, cette relation n'est pas significative ($p > 0,05$). Il s'agit d'un «cas limite», car la valeur de p est légèrement supérieure à la limite conventionnelle de 0,05.

Association et cause

Soulignons que l'existence d'une association statistique ne prouve pas qu'une augmentation de x **cause** une augmentation de y , ou qu'une augmentation de y cause une augmentation de x , même si cette association est forte. Une faiblesse fondamentale des observations réside dans le fait qu'elles peuvent démontrer une association, mais pas une cause. Pour démontrer une relation causale, il faut effectuer une étude expérimentale¹.

¹La plus grande partie du présent module est tirée d'un cours donné dans le cadre du programme de maîtrise en sciences de la London School of Hygiene and Tropical Medicine.

Annexe 29.1. Calcul du coefficient de régression et du coefficient de corrélation

Soit l'équation de régression

$$y = a + bx$$

Les variables a et b (le coefficient de régression) sont calculées à partir des données, comme suit :

$$b = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\Sigma(x - \bar{x})^2} = \frac{\Sigma xy - (\Sigma x)(\Sigma y) / n}{\Sigma x^2 - (\Sigma x)^2 / n}$$

$$a = \bar{y} - b\bar{x}$$

où n est le nombre d'observations,
 \bar{x} est la moyenne de toutes les valeurs de x,
 \bar{y} est la moyenne de toutes les valeurs de y.

On calcule le coefficient de corrélation au moyen de la formule suivante :

$$r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\sqrt{\Sigma(x - \bar{x})^2 \Sigma(y - \bar{y})^2}} = \frac{\Sigma xy - (\Sigma x)(\Sigma y) / n}{\sqrt{(\Sigma x^2 - (\Sigma x)^2 / n) (\Sigma y^2 - (\Sigma y)^2 / n)}}$$

**Module 29 : ASSOCIATIONS ENTRE VARIABLES : RÉGRESSION
ET CORRÉLATION**

Durée et méthode d'enseignement	
1 heure	Introduction et discussion
3 heures	Travail en groupe

Introduction et discussion

- Le présent module ne doit être présenté que si au moins une équipe de recherche en a besoin pour analyser ses données ou si les participants ont des connaissances suffisantes en statistique.

**Formation à la recherche sur les systèmes de santé
Volume 2, 2^e partie : Analyse de données et rédaction de rapports**

Module 30

**MESURE DU RISQUE
DANS LES ÉTUDES CAS-TÉMOINS**

ÉTAPES DE L'ANALYSE DE DONNÉES ET DE LA RÉDACTION DE RAPPORTS

Questions à poser	Étapes à suivre*	Éléments importants de l'étape
Quelles données ont été recueillies pour chaque objectif de recherche? Les données sont-elles complètes et exactes?	Préparer les données pour l'analyse	Revoir les activités sur le terrain Compiler les données pour chaque objectif ou population étudiée Classer les données et en vérifier la qualité Vérifier les imprimés
À quoi ressemblent les données?	Décrire les variables	Distributions Figures, moyennes
Comment peut-on résumer les données pour en faciliter l'analyse?	Faire le recoupement des données quantitatives Résumer les données qualitatives	Faire le recoupement des données par rapport aux objectifs Illustrations, texte
Pour les données qualitatives : la recherche a-t-elle pour but de décrire, de comparer ou de trouver des associations?	Déterminer le type d'analyse statistique nécessaire	Examiner les objectifs, le type d'étude et les variables Faire une description statistique des variables Choisir les tests d'hypothèses
1. Comment peut-on décrire les données?	Analyser des observations appariées et non appariées	**Test t de Student **Test t pour observations appariées **Test du khi carré **Test du khi carré de McNemar
2. Comment peut-on déterminer les différences entre les groupes?	Évaluer les associations entre les variables	**Diagramme de dispersion **Droite de régression et coefficient de corrélation **Risque relatif, risque relatif approché
3. Comment peut-on déterminer les associations entre les variables?	Rédiger le rapport et formuler des recommandations	Préparer le canevas du rapport Rédiger les 1 ^{re} et 2 ^e versions Faire un sommaire des constatations Faire un sommaire des conclusions pour chaque objectif Formuler des recommandations Préparer un résumé analytique
Comment le rapport doit être rédigé?	Présenter un sommaire et rédiger un plan de mise en oeuvre des recommandations	Discuter des sommaires avec différents groupes cibles Discuter du programme de mise en oeuvre
Comment doit-on présenter et diffuser les constatations et les recommandations?		

* Il n'est pas nécessaire de faire ces étapes dans l'ordre indiqué. L'ordre peut être adapté aux besoins de l'atelier.

** Ces éléments sont facultatifs; ils peuvent être omis s'ils ne sont pas pertinents au cours.

Module 30 : MESURE DU RISQUE DANS LES ÉTUDES CAS-TÉMOINS

OBJECTIFS

À la fin de la présente séance, vous devriez pouvoir :

1. **Définir** l'incidence, le risque et le risque relatif;
2. **Évaluer** le risque relatif dans le cadre d'une étude cas-témoins au moyen de la mesure appropriée :
 - Risque relatif approché pour les observations non appariées;
 - Test de McNemar pour les observations appariées.

I. Introduction

II. Incidence, risque et risque relatif

III. Estimation du risque relatif dans une étude cas-témoins

I. INTRODUCTION

Dans une étude cas-témoins, le chercheur compare un groupe de «cas» présentant le problème qu'il veut étudier (p. ex., une maladie) avec un groupe de «témoins», qui ne présentent pas ce problème. Il cherche ainsi à déterminer les facteurs qui auraient pu causer ce problème.

Avant d'étudier l'analyse appropriée, voici quelques notions importantes.

II. INCIDENCE, RISQUE ET RISQUE RELATIF

Les définitions d'incidence et de risque ne font pas l'unanimité, mais dans le présent module, nous utiliserons les définitions suivantes :

L'INCIDENCE est le nombre total de *nouveaux* cas d'un problème défini (p. ex., une maladie) qui se produisent pendant une période donnée et dans une population donnée.

Exemple

Le nombre total de nouveaux cas de tuberculose dans le district A en 1987 était de 273. On peut dire que l'**incidence** de la tuberculose dans le district A était de 273 en 1987.

Le TAUX D'INCIDENCE est le nombre total de nouveaux cas d'un problème défini qui se produisent pendant une certaine période, divisé par la «population à risque».

Exemple

Le district A compte une population de 200 000 personnes. Le **taux d'incidence** de la tuberculose en 1987 y était donc de 273/200 000, ou de 137/100 000 par année.

Le RISQUE et le **taux d'incidence** désignent la même chose.

Exemple

Le **risque** de contracter la tuberculose dans le district A en 1987 était de 137/100 000 par année.

Il est possible que le risque ne soit pas le même pour les différents sous-groupes de la population. Il pourrait être de 100/100 000 par année pour les agriculteurs et de 200/100 000 pour les mineurs. Dans ce cas, les mineurs courent un risque deux fois plus grand de contracter la tuberculose.

On peut conclure que le fait d'être mineur constitue un **facteur de risque** pour contracter la tuberculose, et que ce facteur comporte un **risque relatif** de 2.

Un **FACTEUR DE RISQUE** est un facteur dont la présence est associée à l'augmentation du risque d'une maladie ou d'un état.

Il est important de souligner que la présence d'un facteur de risque ne signifie pas qu'il existe un **lien causal** entre ce facteur et l'état.

Pour déterminer le risque relatif, il faut tenir compte de deux sous-groupes de la population étudiée : un sous-groupe présentant le facteur de risque, et un autre chez qui ce facteur est absent.

Le **RISQUE RELATIF** est le risque de contracter la maladie en question dans le groupe présentant le facteur de risque divisé par le risque de contracter cette maladie dans le groupe chez qui le facteur de risque est absent.

Plus le risque relatif est élevé, plus il est probable que le facteur de risque soit une cause et ne soit pas attribuable à des variables confusionnelles.

III. ESTIMATION DU RISQUE RELATIF DANS UNE ÉTUDE CAS-TÉMOINS

Comme nous l'avons déjà mentionné, des études cas-témoins sont menées dans le but de déceler des facteurs de risque pour certaines maladies ou affections. Pour analyser les résultats d'une étude cas-témoins, il faut constituer des tableaux croisés où les malades et les personnes en bonne santé forment les colonnes et les différentes variables considérées comme facteurs de risque possibles forment les rangées (voir le **module 23**).

Si une différence est observée entre les malades et les personnes en bonne santé quant à une variable précise (facteur de risque), un test du khi carré peut être effectué pour déterminer si cette différence est significative (voir le **module 27**).

Cependant, ce test ne permet pas de mesurer à quel point est étroite cette relation entre les deux variables (présence ou absence du facteur de risque et présence ou absence de la maladie). Par conséquent, il faut mesurer le degré d'association, en calculant le **RISQUE RELATIF**.

Cette analyse nous permet de résoudre des problèmes pratiques. Si l'on connaît un facteur de risque et chez qui il est présent et si l'on connaît également le risque relatif, on peut déterminer dans quelle mesure l'incidence de la maladie peut être réduite par des mesures préventives (si l'on suppose que le facteur de risque est une cause de la maladie).

Exemple

Si l'on sait que les fumeurs courent un risque 10 fois plus grand de contracter le cancer du poumon que les non-fumeurs, on peut supposer qu'une campagne d'information sur la santé qui ferait passer le pourcentage de fumeurs dans la population adulte de 40 % à 35 % entraînerait une baisse spectaculaire de l'incidence du cancer du poumon.

Dans les études cas-témoins, il arrive souvent qu'on ne calcule pas le risque relatif parce que l'incidence de la maladie n'a pas été mesurée dans la population où l'échantillon a été choisi. Cependant, il est possible de faire une **estimation** du risque relatif.

Voyons maintenant la marche à suivre pour estimer le risque relatif à partir d'observations non appariées et d'observations appariées.

Estimation du risque relatif - observations non appariées

Dans les cas d'observations non appariées, on peut calculer le RISQUE RELATIF APPROCHÉ, pourvu que les deux conditions suivantes soient respectées :

- La maladie comporte une faible incidence dans la population, tant dans le groupe à risque que dans le groupe qui n'est pas à risque. Une incidence inférieure à 0,05 serait conforme.
- Le groupe témoin est représentatif de la population.

Exemple 1

Dans une étude cas-témoins sur la consommation de tabac en tant que facteur de risque pour le cancer du poumon, les données suivantes ont été obtenues :

Tableau 30.1. Fumeurs et non-fumeurs parmi les personnes atteintes du cancer du poumon et les personnes en bonne santé.

	Personnes atteintes du cancer du poumon	Personnes en bonne santé
Fumeurs (+)	350	184
Non-fumeurs (-)	45	216
Total	395	400

D'après ce tableau, plus de personnes atteintes du cancer du poumon que de personnes en santé fument. Cette constatation ne peut être attribuée à la variation de l'échantillonnage ($\chi^2 = 164$; $v = 1$; $p < 0,001$).

Le risque relatif de contracter le cancer du poumon qui est associé à la consommation de cigarettes peut être estimé en calculant le RISQUE RELATIF APPROCHÉ.

$$\text{Risque relatif approché} = \frac{350 \times 216}{184 \times 45} = 9,1$$

Cela signifie que les fumeurs étudiés couraient un risque 9,1 plus élevé de contracter le cancer du poumon que les non-fumeurs.

Les étapes à suivre pour effectuer ce type d'analyse figurent ci-dessous :

Étape 1 Préparer un tableau :

Facteur de risque	Cas	Témoins	Total
Présent (+)	a	b	e
Absent (-)	c	d	f
Total	g	h	n

Étape 2 Déterminer s'il existe une différence entre les malades et les personnes en bonne santé quant à la variable soupçonnée d'être un facteur de risque. Effectuer un test du χ^2 pour déterminer si cette différence est significative.

Étape 3 Estimer le risque relatif en calculant le risque relatif approché. Ce dernier constitue un rapport de deux rapports, c'est-à-dire le rapport des malades qui présentent le facteur de risque (a) sur les malades qui ne présentent pas ce facteur (c), divisé par le rapport des personnes en bonne santé qui présentent le facteur de risque (b) sur celles qui ne présentent pas ce facteur (d). Cela donne la formule suivante :

$$\text{Risque relatif approché} = \frac{\text{risque pour les malades présentant le facteur de risque}}{\text{risque pour les personnes en santé présentant le facteur de risque}} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

Remarques

- Avant d'utiliser cette formule, assurez-vous que le tableau d'où proviennent les données est disposé de la même façon que l'exemple ci-dessus.
- Ne calculez pas le risque relatif approché si le test du χ^2 montre que la différence entre les malades et les personnes en bonne santé n'est pas significative.

Estimation du risque relatif - observations appariées

Dans certaines études cas-témoins, les personnes en santé et les malades sont appariés. Il faut en tenir compte lors de l'analyse. Reprenons l'exemple du **module 28**.

Exemple 2

Dans le cadre d'une étude cas-témoins visant à trouver les causes d'une épidémie de choléra en Italie, le tableau suivant a été préparé :

Tableau 30.2. Consommation de poisson ou de fruits de mer par des personnes atteintes du choléra et des personnes en bonne santé dans un délai de 5 jours avant le début de la maladie.

	Personnes atteintes du choléra		Total
	Ont consommé	N'ont pas consommé	
Personnes en santé			
Ont consommé	12	3	15
N'ont pas consommé	30	31	61
Total	42	34	76

Source : Baine, W.B., Mazzotti, M., Greco, D., Izzo, E., Zampieri, A., Angioni, G., Di Gioia, M., Gangarosa, E.J., Pocchiari, S. 1974. Epidemiology of cholera in Italy in 1973. Lancet, ii (déc.), 1370-1374.

Pour chaque malade (confirmé par analyse bactériologique), une personne en bonne santé du même sexe, du même âge (à la dizaine près) et du même quartier a été choisie.

Nous avons découvert que chez 30 paires, le malade a mangé des fruits de mer, mais pas la personne en bonne santé, alors que chez seulement 3 paires, la personne en bonne santé a mangé des fruits de mer mais le malade n'en a pas mangé. Comme nous l'avons vu au **module 28**, cette association n'est pas attribuable à la variation de l'échantillonnage (χ^2 de McNemar = 20,5; $p < 0,001$).

Dans ce cas, le risque relatif est estimé à :

$$\frac{30}{3} = 10$$

En d'autres mots, les personnes qui ont mangé des fruits de mer couraient un risque 10 fois plus grand de contracter le choléra que celles qui n'en ont pas mangé.

Les étapes de cette analyse sont décrites ci-dessous :

Étape 1 Préparer un tableau :

	Malades	
Personnes en bonne santé	Facteur de risque +	Facteur de risque -
Facteur de risque +	q	r
Facteur de risque -	s	t

Étape 2 Déterminer s'il existe une association entre le facteur de risque et la maladie ou l'état. Il faut alors comparer r et s dans le tableau. Pour déterminer si cette association est significative (et non due à une variation de l'échantillonnage), on effectue un test du χ^2 de McNemar.

Étape 3 Estimer le risque relatif en utilisant la formule suivante :

$$\text{Risque relatif} = \frac{s}{r}$$

TRAVAIL EN GROUPE

Si vous avez fait une étude cas-témoins, estimez le risque relatif en calculant le risque relatif approché (si vos observations ne sont pas appariées) ou le rapport du nombre de paires contradictoires (si vos observations sont appariées) pour différents facteurs de risque.

N'oubliez pas que vous ne pouvez faire cette estimation que si les facteurs de risque des malades et des personnes en bonne santé présentent une différence significative.

Annexe 30.1. Estimation du risque relatif en tenant compte de l'effet d'une variable confusionnelle

Une étude cas-témoins est entreprise dans une clinique dans le but de déterminer si l'allaitement naturel protège les enfants de 12 à 23 mois contre la malnutrition.

Les résultats suivants sont obtenus :

Tableau 30.3. Résultats d'une étude cas-témoins visant à déterminer si l'allaitement naturel protège les enfants de 12 à 23 mois contre la malnutrition.

	Malades (mal nourris)	Enfants en santé (bien nourris)	Total
Pas d'allaitement naturel (-)	100	76	176
Allaitement naturel (+)	100	124	224
Total	200	200	400

$$\chi^2 = 5,36; v = 1; p < 0,05; \text{risque relatif} = \frac{100 \times 124}{100 \times 76} = 1,6$$

On soupçonne que l'âge est une variable confusionnelle pour les raisons suivantes :

- Les enfants de 12 à 17 mois sont plus souvent mal nourris que ceux de 18 à 23 mois;
- Plus d'enfants de 12 à 17 mois sont allaités naturellement que d'enfants de 18 à 23 mois.

Idéalement, il aurait fallu tenir compte de l'âge lors de la conception de l'étude, mais malheureusement, on ne l'a pas fait.

Comment analyser ces données?

Suivons ces trois étapes :

Étape 1 Séparons le tableau en deux parties.

Tableau 30.4. Résultats pour les enfants de 12 à 17 mois.

	Malades	En santé	Total
Pas d'allaitement naturel (-)	50 (42 %)	22 (28 %)	72
Allaitement naturel (+)	70	58	128
Total	120 (100 %)	80 (100 %)	200

$$\chi^2 = 3,58; v = 1; p > 0,05; \text{risque relatif} = 1,9$$

Tableau 30.5. Résultats pour les enfants de 18 à 23 mois.

	Malades	En santé	Total
Pas d'allaitement naturel (-)	50 (63 %)	54 (45 %)	104
Allaitement naturel (+)	30	66	96
Total	80 (100 %)	120 (100 %)	200

$$\chi^2 = 5,2; \nu = 1; p < 0,05; \text{risque relatif} = 2,0$$

Soulignons que pour chacun des sous-groupes, le risque relatif de l'absence d'allaitement naturel est supérieur à celui de l'ensemble du groupe (environ 2). Supposons que nous voulons faire un test d'hypothèses d'ensemble et estimer le risque relatif général :

Étape 2 Test du χ^2 de Mantel-Haenszel (voir l'annexe 27.3).

	$f_o(a)$	$f_1(a)$ (eg/n)	$s^2(a)$ (egfh/n ² (n - 1))
12 à 17 mois	50	43,2	11,1
18 à 23 mois	50	41,6	12,0
Total	100	84,8	23,1

$$\chi^2 = \frac{(f_o(a) - f_1(a) - 0,5)^2}{s^2(a)} = \frac{(100 - 84,8 - 0,5)^2}{23,1} = 9,35; \nu = 1; p < 0,01$$

Étape 3 Le risque relatif est estimé comme suit :

	ad/n	bc/n
12 à 17 mois	14,5	7,7
18 à 23 mois	16,5	8,1
Total	31,0	15,8

$$\text{Risque relatif} = \frac{ad/n}{bc/n} = \frac{31,0}{15,8} = 2,0$$

Notes du formateur

Module 30 : MESURE DU RISQUE DANS LES ÉTUDES CAS-TÉMOINS

Durée et méthode d'enseignement	
1 heure	Introduction et discussion
2 heures	Travail en groupe

Introduction et discussion

- Pour que votre exposé soit plus facile à suivre, n'employez pas les termes cas et témoins, mais parlez plutôt de personnes malades et de personnes en santé, ou de personnes atteintes du choléra et de personnes en bonne santé.
- Si l'un des groupes a fait une étude cas-témoins, tirez-en un exemple pour illustrer comment calculer et interpréter le risque relatif approché.
- **L'annexe 30.1** ne devrait pas être présentée lors de l'exposé. Les groupes peuvent l'utiliser au besoin.
- Si aucun des groupes n'a d'observations appariées et si les participants ont peu d'expérience en statistique, vous pouvez laisser tomber la dernière partie de la section sur l'estimation du risque relatif (observations appariées).

Formation à la recherche sur les systèmes de santé
Volume 2, 2^e partie : Analyse de données et rédaction de rapports

Module 31

RÉDACTION DE RAPPORTS

ÉTAPES DE L'ANALYSE DE DONNÉES ET DE LA RÉDACTION DE RAPPORTS

Questions à poser	Étapes à suivre*	Éléments importants de l'étape
Quelles données ont été recueillies pour chaque objectif de recherche? Les données sont-elles complètes et exactes?	Préparer les données pour l'analyse	Revoir les activités sur le terrain Compiler les données pour chaque objectif ou population étudiée Classer les données et en vérifier la qualité Vérifier les imprimés
À quoi ressemblent les données?	Décrire les variables	Distributions Figures, moyennes
Comment peut-on résumer les données pour en faciliter l'analyse?	Faire le recoupement des données quantitatives Résumer les données qualitatives	Faire le recoupement des données par rapport aux objectifs Illustrations, texte
Pour les données qualitatives : la recherche a-t-elle pour but de décrire, de comparer ou de trouver des associations?	Déterminer le type d'analyse statistique nécessaire	Examiner les objectifs, le type d'étude et les variables Faire une description statistique des variables Choisir les tests d'hypothèses
1. Comment peut-on décrire les données?	Analyser des observations appariées et non appariées	**Test t de Student **Test t pour observations appariées **Test du khi carré **Test du khi carré de McNemar
2. Comment peut-on déterminer les différences entre les groupes?	Évaluer les associations entre les variables	**Diagramme de dispersion **Droite de régression et coefficient de corrélation **Risque relatif, risque relatif approché
3. Comment peut-on déterminer les associations entre les variables?	Rédiger le rapport et formuler des recommandations	Préparer le canevas du rapport Rédiger les 1 ^{re} et 2 ^e versions Faire un sommaire des constatations Faire un sommaire des conclusions pour chaque objectif Formuler des recommandations Préparer un résumé analytique
Comment le rapport doit être rédigé?	Présenter un sommaire et rédiger un plan de mise en oeuvre des recommandations	Discuter des sommaires avec différents groupes cibles Discuter du programme de mise en oeuvre
Comment doit-on présenter et diffuser les constatations et les recommandations?		

* Il n'est pas nécessaire de faire ces étapes dans l'ordre indiqué. L'ordre peut être adapté aux besoins de l'atelier.

** Ces éléments sont facultatifs; ils peuvent être omis s'ils ne sont pas pertinents au cours.

Module 31 : RÉDACTION DE RAPPORTS

OBJECTIFS

À la fin de la présente séance, vous devriez pouvoir :

1. Dresser une liste des principaux éléments d'un rapport de recherche;
2. Préparez le canevas de votre projet de recherche;
3. Rédiger par étapes une version préliminaire de votre rapport;
4. Examiner le rapport final pour s'assurer qu'il est complet, qu'il ne contient pas de chevauchements et que le style est clair et facile à lire;
5. Formuler des recommandations fondées sur votre recherche.

I. Étapes de la préparation du rapport : considérations préliminaires

II. Rédaction du rapport

I. ÉTAPES DE LA PRÉPARATION DU RAPPORT : CONSIDÉRATIONS PRÉLIMINAIRES

Le public

L'objectif d'un rapport de recherche consiste à fournir des renseignements au lecteur. Il est donc important de commencer par répondre aux questions suivantes :

QUI lira le rapport de recherche?
POURQUOI voudra-t-on le lire?

Dans les recherches sur les systèmes de santé, il est particulièrement important de garder en tête les besoins des lecteurs, car il ne s'agit pas seulement de chercheurs, mais également de gestionnaires de la santé et de dirigeants communautaires. Bon nombre de documents de recherche conçus pour des scientifiques ne conviennent pas aux gestionnaires et aux profanes. Par conséquent, il faut s'efforcer de rédiger des rapports en termes simples qui décrivent avec précision les constatations.

En outre, il est important de présenter non seulement les conclusions scientifiques, mais également des recommandations précises qui tiennent compte des caractéristiques du système local de santé ainsi que des contraintes, de la faisabilité et de l'utilité des solutions proposées. La collectivité et les gestionnaires préfèrent savoir comment régler un problème plutôt que d'apprendre simplement que ce problème existe.

Les rapports doivent répondre aux **BESOINS DES LECTEURS** : dirigeants communautaires, gestionnaires de la santé et chercheurs.

Comment le lecteur lit un rapport de recherche

Il est plus facile de rédiger un bon rapport lorsqu'on tient compte de la méthode de lecture des personnes qui le liront. La recherche a été effectuée dans le but de recueillir des renseignements : ceux-ci doivent donc constituer le pivot du rapport. Ces renseignements doivent être résumés dans les **conclusions**. La plupart des lecteurs commenceront par lire les conclusions; si elles sont intéressantes, utiles et agréablement présentées, le lecteur s'attardera aux autres sections du rapport. Celles-ci visent à étayer les conclusions en aidant le lecteur à répondre à deux questions fondamentales :

- **Comment ces renseignements permettront d'améliorer la santé de la collectivité?** (En d'autres mots, quelles sont les caractéristiques du problème en cause et du système de santé où ce problème se manifeste, et comment ces renseignements permettront de résoudre ou de minimiser ce problème?)
- **Ces constatations sont-elles réalistes** (c.-à-d. valables et fiables)? La conception de la recherche, l'échantillonnage, les méthodes de collecte des données et l'analyse des données confirmeront la validité et la fiabilité des constatations.

Soulignons qu'un rapport qui met l'accent sur la méthode plutôt que sur les conclusions pourrait intéresser un public de chercheurs, mais pas les gestionnaires.

Achèvement de l'analyse des données

Avant d'entreprendre l'aperçu et la rédaction de la version préliminaire de votre rapport, vous devez revoir votre analyse des données et répondre à plusieurs des questions suivantes :

- **Les conclusions correspondent-elles aux objectifs spécifiques? Sont-elles complètes?**

Les étapes de l'analyse de données auraient dû aboutir à

- une ou plusieurs conclusions exprimées sous forme de phrases claires;
- un ou plusieurs tableaux analytiques accompagnés de statistiques descriptives ou de tests statistiques pertinents pour étayer les conclusions.

Examinez vos conclusions afin de vérifier

- que tous les objectifs spécifiques y sont mentionnés;
- qu'elles tiennent compte de tous les aspects des objectifs;
- qu'elles sont pertinentes et appropriées compte tenu des objectifs.

- **Est-il nécessaire de préparer d'autres tableaux analytiques?**

Si les conclusions sont incomplètes, préparez d'autres maquettes de tableaux analytiques et analysez les données tel qu'indiqué aux **modules 22 à 30**.

- **Toutes les données qualitatives ont-elles été utilisées pour étayer et préciser les conclusions tirées des tableaux?**

Après avoir terminé cet examen, il vous reste deux autres tâches à remplir :

- **Énoncez vos conclusions définitives pour chaque objectif.**

Au cours des premières étapes de l'analyse, chaque tableau analytique a été accompagné d'une conclusion. Ces conclusions doivent maintenant être revues, combinées si possible et énoncées de façon que les principales constatations de l'étude soient facilement identifiables pour un lecteur qui lira le rapport en diagonale. Souvent, il est possible d'inclure les données numériques les plus importantes (pourcentages, moyennes, etc.).

- **Choisissez les tableaux qui apparaîtront dans le corps du texte.**

Seuls quelques tableaux devraient être inclus dans le corps du texte. Le tableau doit servir à illustrer une conclusion importante ou à l'étayer. Si possible, combinez des données provenant de plusieurs tableaux analytiques dans un seul tableau ou dans quelques-uns et présentez un tableau sommaire dans le rapport. (Au besoin, des tableaux plus détaillés peuvent être fournis en annexe.) Le titre de chaque tableau doit expliquer le plus brièvement possible le contenu du tableau. Les titres de colonnes et de rangées doivent être brefs mais sans équivoque.

Réunissez les conclusions et les tableaux relatifs à chaque objectif spécifique. Vous êtes maintenant prêt à rédiger le rapport.

II. RÉDACTION DU RAPPORT

Le but du rapport consiste à expliquer les faits au lecteur d'une façon simple, logique et ordonnée. Il faut éviter de distraire le lecteur et de le plonger dans la confusion.

Lors de la rédaction, il faut tenir compte des aspects suivants :

- le CONTENU,
- le STYLE,
- la DISPOSITION du rapport;
- la PREMIÈRE VERSION,
- la DEUXIÈME VERSION,
- le rapport final.

Voyons chacun de ces éléments de la préparation du rapport.

Le contenu : les principaux éléments d'un rapport de recherche

Le rapport de recherche doit contenir les éléments suivants :

- Page titre ou page couverture,
- Sommaire des constatations et recommandations,
- Remerciements (facultatifs),
- Table des matières,
- Liste des tableaux et figures (facultative),
- Liste des abréviations (facultative),
- 1. Introduction,
- 2. Objectifs,
- 3. Méthodologie,
- 4. Constatations et conclusions,
- 5. Discussion,
- 6. Recommandations,
- Ouvrages consultés,
- Annexes (outils de collecte de données, tableaux).

Les constatations et conclusions, la discussion des constatations et les recommandations composeront la partie la plus importante de votre rapport, qui doit être rédigé en partant de zéro. Pour l'introduction, vous pouvez vous fonder dans une grande mesure sur votre proposition de recherche, en résumant, révisant et parfois étoffant certaines sections.

Nous vous recommandons donc vivement de **commencer par les constatations et les conclusions**. Nous verrons cependant chaque élément du rapport dans l'ordre où il apparaîtra.

- **Page couverture**
La page couverture doit contenir le titre, le nom des auteurs avec leur titre et leur poste, l'établissement qui publie le rapport ainsi que le mois et l'année de publication. L'établissement sera probablement celui qui a mené l'étude, par exemple l'unité de recherche du ministère de la Santé ou un institut de recherche.
- **Sommaire**
Le sommaire ne peut être rédigé qu'après avoir achevé la première ou même la deuxième version du rapport. Il doit contenir :

- une brève description du problème,
- les principaux objectifs,
- l'endroit où l'étude a été menée,
- le type d'étude et les méthodes employées,
- les principales constatations et conclusions,
- les principales recommandations (ou la totalité).

C'est le sommaire qui sera lu en premier (et il est probable que les décideurs du domaine de la santé, dont le temps est précieux, ne liront que le sommaire). Par conséquent, il faut consacrer beaucoup de temps et d'attention à sa rédaction. Vous devrez peut-être en rédiger plusieurs versions qui feront l'objet de discussions en équipe.

Comme vous aurez collaboré avec plusieurs groupes au cours de la rédaction et de la mise en oeuvre de votre proposition de recherche, vous pouvez envisager la possibilité de rédiger des sommaires différents pour chacun de ces groupes. Par exemple, vous pourriez préparer des sommaires différents pour les décideurs et les gestionnaires de la santé, pour le personnel de santé subalterne, pour le grand public (journaux, télévision) et pour les spécialistes (articles dans des publications scientifiques). (Voir le **module 32**.)

- **Remerciements**

Vous pouvez remercier les personnes qui vous ont apporté une aide technique ou financière lors de la conception et de l'exécution de votre étude. Vous pouvez également remercier votre employeur, qui vous a permis de consacrer du temps à l'étude, ainsi que les répondants. D'habitude, les remerciements figurent tout de suite après la page couverture, ou à la fin du rapport, avant la liste des ouvrages consultés.

- **Table des matières**

La table des matières est essentielle, car elle donne au lecteur une vue d'ensemble des principales sections de votre rapport avec leur numéro de page, ce qui lui permet de parcourir le rapport dans l'ordre qu'il désire ou d'escamoter certaines sections.

- **Liste des tableaux et figures (facultative)**

Si vous avez beaucoup de tableaux ou de figures, il est utile d'en dresser une liste sous la forme d'une table des matières, en indiquant les pages.

- **Liste des abréviations (facultative)**

Vous pouvez également dresser une liste des abréviations et acronymes utilisés dans votre rapport, s'il y en a beaucoup.

Ces trois dernières parties doivent être préparées en dernier, car vous devez indiquer le numéro de page des chapitres et sections dans la table des matières et vous assurer qu'il n'y a pas d'erreur dans la numérotation des figures et des tableaux.

1. INTRODUCTION

L'introduction constitue une partie relativement facile du rapport, que vous pouvez rédiger après avoir préparé la première version des constatations. Elle doit contenir des renseignements de base sur le pays étudié, l'état de santé de sa population et des données sur les services de santé relatives au problème étudié. Vous pouvez réviser légèrement ou faire des ajouts à la section correspondante de votre proposition de recherche et vous en servir.

L'énoncé du problème tiré de votre proposition de recherche, révisé ou accompagné de commentaires ou de données supplémentaires fondées sur votre recherche, vient ensuite. Il doit contenir un paragraphe sur ce que vous espérez accomplir avec les résultats de l'étude.

Un bref examen de la documentation portant sur votre sujet d'étude doit également figurer dans le rapport (consultez le **module 5** et votre proposition de recherche). Cette section doit comprendre des éléments pertinents pour aider le lecteur à :

- comprendre le problème en fournissant un aperçu des renseignements disponibles sur lui,
- comprendre les méthodes d'étude ou de résolution du problème.

REMARQUE : Cette section ne doit pas constituer un résumé de tous les documents et ouvrages sur le sujet. Soyez sélectif et n'oubliez pas que cette section doit étayer votre étude et non prouver que vous pouvez consulter des documents.

2. OBJECTIFS

Les objectifs généraux et spécifiques doivent être mentionnés. Au besoin, adaptez légèrement leur style et leur ordre. Cependant, ne les changez pas complètement. Si vous n'avez pu atteindre certains objectifs, dites-le dans la section sur la méthodologie et dans la discussion sur les constatations.

3. MÉTHODOLOGIE

Vous devez décrire en détail la méthodologie employée pour la collecte des données, en mentionnant :

- le type d'étude,
- les variables sur lesquelles portent les données,
- la population d'où vient l'échantillon,
- la taille de l'échantillon et la méthode d'échantillonnage,
- les techniques de collecte des données :
 - les sources de données (cartes, ménages, registres des cliniques, etc.),
 - les personnes ayant recueilli les données et leur méthode,
 - les procédés d'analyse des données, y compris les tests statistiques (s'il y a lieu).

Si vous vous êtes écarté du modèle initial présenté dans votre proposition de recherche, vous devez expliquer dans quelle mesure et pour quelle raison. Les conséquences de cet écart sur les objectifs de votre étude doivent être indiquées. Décrivez également les données de faible qualité, qui peuvent causer des biais.

4. CONSTATATIONS ET CONCLUSIONS

La présentation systématique de vos constatations et conclusions par rapport à vos objectifs de recherche constitue le pivot de votre rapport.

Vous pouvez joindre à une description des constatations quelques tableaux ou graphiques qui les résument. Le rapport sera plus agréable à lire si vous illustrez certaines constatations par des exemples en citant les répondants ou en fournissant des observations et des études de cas effectuées pendant les travaux sur le terrain.

5. DISCUSSION

Les constatations peuvent être discutées par objectif ou par groupe de variables connexes. Des conclusions d'autres études qui viennent soutenir ou contredire les vôtres devraient également être mentionnées. En outre, il est important de décrire les limites de l'étude. Vous pouvez également formuler quelques conclusions générales.

Remarque

Le texte et les annexes doivent contenir des renseignements suffisants pour que les professionnels puissent déterminer comment vous démontrez vos constatations et vos conclusions. Le rapport doit être suffisamment clair pour qu'il soit possible de répéter l'étude au besoin.

6. RECOMMANDATIONS

Les recommandations doivent constituer une suite logique à la discussion sur les constatations. Elles peuvent être résumées en fonction des groupes qu'elles concernent, notamment :

- les décideurs,
- les gestionnaires de la santé et de domaines connexes au palier des districts ou à un palier inférieur,
- le personnel de la santé et de domaines connexes qui pourrait mettre en oeuvre les recommandations,
- les clients possibles,
- le grand public.

N'oubliez pas que les groupes susceptibles de prendre des mesures s'intéresseront particulièrement à cette partie.

Pour rédiger vos recommandations, ne vous fondez pas uniquement sur les constatations de votre étude, mais également sur des renseignements provenant d'autres sources et sur de l'information portant sur d'autres facteurs connexes. Il convient de discuter des recommandations avec tous les intéressés avant d'en faire la formulation finale.

Si vos recommandations sont courtes, vous pourriez les inclure toutes dans votre sommaire des constatations et recommandations au lieu de les faire figurer dans une section distincte.

• **Ouvrages consultés**

Dans le texte, les renvois peuvent être numérotés dans l'ordre où ils se trouvent, puis figurer dans cet ordre dans la partie sur les ouvrages consultés. On peut également écrire le nom de l'auteur dans le texte suivi de la date de publication entre parenthèses : (Shan 1990). Dans la liste des ouvrages consultés, les publications sont alors placées en ordre alphabétique du nom de famille de l'auteur principal (voir le module 5).

Vous pouvez choisir l'une ou l'autre méthode, mais si vous voulez publier un article, vous devez suivre la méthode prescrite par la publication.

- **Annexes**

Les annexes doivent contenir tous les renseignements supplémentaires permettant aux spécialistes de suivre vos procédures de recherche et d'analyse de données.

Des renseignements utiles à certains lecteurs mais qui n'intéressent pas le lecteur moyen peuvent également figurer en annexe.

Parmi des exemples de renseignements à inclure dans les annexes, on relève :

- des tableaux mentionnés dans le texte, mais qui n'y ont pas été ajoutés par souci de brièveté,
- des listes de critères et de définitions et des organigrammes,
- des listes d'hôpitaux, de districts, de villages, etc. qui ont participé à l'étude,
- les outils de collecte des données.

Style

N'oubliez pas que le lecteur :

- n'a pas de temps à perdre;
- doit accorder son attention à une foule d'autres questions urgentes;
- ne connaît probablement pas très bien le «jargon» de la recherche.

Suivez donc les règles suivantes :

- **Simplifiez.** Tenez-vous-en à l'essentiel.
- **Justifiez.** Ne faites aucune affirmation qui ne soit fondée sur des faits.
- **Quantifiez.** Évitez les termes «grand», «petit», etc., et dites plutôt «près de 75 %», «un sur trois», etc.
- **Soyez précis.**
- **Cherchez à informer, pas à impressionner.** Évitez les exagérations.
- **Faites des phrases courtes.**
- **N'utilisez pas trop d'adverbes et d'adjectifs.** Soyez uniforme quant aux temps (passé, présent). Évitez la formulation passive si possible.
- **Soyez clair, logique et systématique.**

Présentation du rapport

S'il est bien présenté, votre rapport :

- fera bonne impression;
- sera agréable à lire;
- donnera une idée de l'organisation du contenu de sorte que le lecteur pourra déterminer rapidement ce qu'il veut lire en premier.

Une attention particulière doit être portée aux points suivants :

- Page titre attrayante, table des matières claire,

- Uniformité des marges et de l'interligne,
- Uniformité des titres et sous-titres (p. ex., majuscules soulignées pour les titres de chapitres, majuscules pour les titres de grandes sections, minuscules soulignées pour les titres de sous-sections, etc.),
- Dactylographie et photocopies de qualité, sans coquilles (pour de plus amples renseignements, voir Keithly et Schreiner 1971).
- Figures et tableaux numérotés et titrés, avec des titres pour les colonnes et les rangées,
- Précision et cohérence des citations et des renvois.

Préparation de la première version du rapport

Préparez un canevas écrit. Ce canevas vous aidera à organiser vos idées et constitue une étape essentielle dans la production d'un rapport logique et ordonné.

Le canevas doit contenir :

- Le titre des principales sections du rapport,
- Le titre des sous-sections,
- L'objet de chaque section,
- Une liste des tableaux et figures (s'il y a lieu) illustrant chaque section.

Le canevas du chapitre sur les constatations et conclusions sera le plus difficile. Pour vous aider à structurer vos constatations d'une manière logique et cohérente, discutez avec les membres de votre équipe des liens entre vos constatations et conclusions et vos objectifs et variables.

- La première partie de ce chapitre constitue habituellement une description de l'échantillon, précisant notamment l'emplacement, l'âge, le sexe et d'autres variables de base pertinentes.
- Ensuite, selon le type d'étude, vous pouvez présenter d'autres renseignements sur le problème ou sur les variables dépendantes.
- Enfin, vous pouvez inclure une analyse des différentes variables indépendantes par rapport au problème.

Vous pouvez commencer par faire une liste des titres et des sous-titres, en laissant beaucoup d'espace entre eux pour griffonner quelques mots clés sur ce que vous comptez écrire sous chacun. Il est préférable de numéroté les sections et sous-sections.

Par exemple, dans une étude sur la malnutrition, le chapitre «Constatations et conclusions» pourrait prendre la forme suivante :

CHAPITRE 4 : CONSTATATIONS ET CONCLUSIONS

4.1 DESCRIPTION DE L'ÉCHANTILLON

4.2 ÉTENDUE ET VARIATION SAISONNIÈRE DE LA MALNUTRITION DANS LE DISTRICT X

4.3 CAUSES POSSIBLES DE LA MALNUTRITION

- 4.3.1 Disponibilité réduite d'aliments
- 4.3.2 Utilisation non optimale des aliments
- 4.3.3 Incidence élevée de maladies contagieuses
- 4.3.4 Accès limité aux services de santé maternelle et infantile et aux soins de santé
- 4.3.5 Conclusions

Ce genre de numérotation présente une grande souplesse et permet l'ajout de titres et sous-titres au besoin. Il vous permet d'avoir en tête un aperçu du processus lorsque différents membres de l'équipe travaillent en même temps à différentes parties du rapport. Si vos constatations sont très détaillées et si vous avez quatre ou cinq niveaux de sous-titres, vous pouvez diviser les constatations en plusieurs chapitres. Vous pouvez également simplifier la numérotation des sous-sections si le sujet auquel elles se rapportent est évident. Cependant, ne le faites pas avant la version finale, car la numérotation vous aidera à garder votre rapport en ordre lorsque différents membres de l'équipe travailleront à différentes sections.

Les tableaux et figures se trouvant dans le corps du texte doivent être numérotés et titrés. Il est préférable d'utiliser d'abord le numéro de la section où se trouve le tableau. Pour la dernière version, vous pouvez choisir de numérotter les tableaux et figures selon l'ordre où ils apparaîtront.

N'inclure que les tableaux et les figures qui illustrent les principales constatations et qui doivent être expliqués en détail dans le corps du texte. D'autres peuvent être présentés en annexe, ou même omis s'ils n'apportent rien d'intéressant.

Il est inutile de décrire en détail les tableaux figurant dans le rapport. Ne présentez que les principales conclusions.

La première version n'est jamais définitive. Attardez-vous donc au contenu plutôt qu'au style. Cependant, il est souhaitable de structurer le texte en paragraphes dès le départ et de formuler chaque phrase avec clarté et précision.

Remarques

Ne commencez jamais à rédiger votre rapport avant d'avoir préparé un canevas. Assurez-vous que toutes les sections portent des titres et des numéros conformes au canevas avant de les envoyer à la dactylographie. Affichez le canevas pour que tous les participants soient tenus au courant des ajouts et des changements.

Tapez la première version à double interligne avec des marges assez larges afin de pouvoir facilement insérer des commentaires et des corrections.

Faites plusieurs copies de la première version de façon à en disposer de quelques-unes pour travailler et d'une copie dans laquelle vous insérerez les révisions définitives que vous apporterez avant de renvoyer le texte à la dactylographie.

Préparation de la deuxième version

Une fois la première version des constatations et des conclusions terminée, tous les membres du groupe d'étude et animateurs devraient en faire une lecture critique et formuler des commentaires à leur sujet.

Lors de la lecture de cette version préliminaire, posez-vous les questions suivantes :

- Toutes les constatations importantes ont-elles été incluses?
- Les conclusions sont-elles reliées logiquement aux constatations? Si certaines constatations se contredisent, font-elles l'objet d'une discussion et d'une explication? Des faiblesses ont-elles été relevées sur le plan de la méthodologie?
- Le texte contient-il des chevauchements qu'il faudrait supprimer?
- Serait-il possible d'abrégier le texte? On peut généralement améliorer un texte en l'abrégeant. Certaines parties moins pertinentes peuvent figurer en annexe. Vérifiez si des paragraphes descriptifs peuvent être raccourcis et présentés comme conclusions.
- Les données figurant dans le texte correspondent-elles à celles des tableaux? Les tableaux sont-ils uniformes (même nombre de répondants par catégorie), sont-ils numérotés en ordre et portent-ils un titre clair?
- L'ordre des paragraphes et des sous-sections est-il logique et cohérent? Y a-t-il un lien logique entre les sections et paragraphes successifs? La formulation des constatations et des conclusions est-elle précise et claire?

Les auteurs de chaque section peuvent préparer une deuxième version préliminaire en tenant compte de tous les commentaires reçus. Cependant, vous pourriez envisager de nommer deux rédacteurs parmi les membres de l'équipe pour rédiger la version définitive.

Entre-temps, les autres membres du groupe peuvent s'occuper des sections du début. L'INTRODUCTION, les OBJECTIFS et la MÉTHODOLOGIE de la proposition de recherche initiale peuvent souvent être utilisés après révision et adaptation (voir la page 8).

Vous pouvez maintenant passer à la première version du sommaire (voir la page 7).

Préparation du rapport final

Il serait souhaitable de demander à l'un des autres groupes et aux animateurs de lire la deuxième version de votre rapport et de l'évaluer en fonction des critères mentionnés à la section précédente. Vous pouvez ensuite passer à la préparation du rapport final. Cette fois-ci, vous devez faire très attention à la présentation (structure, style, orthographe).

Soyez uniforme dans l'usage des temps de verbes. Les descriptions de la situation sur le terrain peuvent être rédigées au passé ou au présent (p. ex., «Cinq ménages possèdent moins d'un acre de terrain»). Les conclusions sont habituellement au présent (p. ex., «les interdits alimentaires ont très peu de répercussions sur l'alimentation des jeunes enfants. Les poissons et viandes interdits dans certains clans reviennent rarement dans le régime quotidien»).

Pour vous assurer que le texte se lit bien, passez-le en revue rapidement et lisez la première phrase de chaque paragraphe. Si vous avez une bonne idée de l'organisation et des résultats de votre étude, vous pouvez conclure que vous avez fait de votre mieux.

TRAVAIL EN GROUPE

1. **Préparez un canevas de votre rapport** sur un tableau à feuilles mobiles, après avoir examiné vos objectifs, vos sources d'information et les résultats de l'analyse des données. Assurez-vous que le chapitre sur les constatations est composé de sections et de sous-sections ordonnées logiquement. Numérotez les sections et sous-sections proposées. Affichez le canevas au mur, dans un endroit visible. Laissez assez d'espace entre les lignes pour les ajouts (de nouvelles parties, par exemple) et les changements.
2. **Entrenez la rédaction, en commençant par les constatations et les conclusions.** Déterminez avec votre animateur si vous interprétez les données en présentant les résultats selon la variable, l'objectif ou la population étudiée. Si vous avez de la difficulté à choisir une méthode de présentation, prenez quand même note de vos constatations et de vos interprétations. Lors de la deuxième version, vous pourrez déterminer comment réorganiser et abrégier le rapport. **Répartissez la rédaction** parmi des sous-groupes d'une ou deux personnes.
3. **Rédigez les chapitres de l'introduction** (contexte, énoncé du problème et revue documentaire), des objectifs et de la méthodologie, en adaptant ce que vous aviez préparé pour votre proposition.
4. **Enfin, rédigez le sommaire** en suivant les indications données au cours de la séance. Prenez au moins une demi-journée pour le faire.
5. **Informez-vous des progrès** de la rédaction et de la dactylographie en prenant des notes sur le tableau à feuilles mobiles où se trouve le canevas de votre rapport.
6. **Examinez la première version avec l'ensemble de l'équipe** et tentez d'en déceler les lacunes, les chevauchements, etc. avant la préparation de la deuxième version. Demandez à l'animateur d'un autre groupe de lire le rapport avant de l'envoyer à la dactylographie.

Ouvrages consultés

Huth, E.J. 1982. How to write and publish papers in the medical sciences (2^e éd.). Williams & Wilkins, Baltimore, MD, États-Unis.

Keithly, Schreiner. 1971. A manual of style for the preparation of papers and reports. South Western Publishing Co.

Osborn, M.L. 1971. The preparation of a scientific paper. British Journal of Medical Education, 5, 248-252.

Paton, A. 1979. Write a paper. In How to do it: 1. Articles publiés dans le British Medical Journal. British Medical Association, Londres, Royaume-Uni, pp. 207-212.

Notes du formateur

Module 31 : RÉDACTION DE RAPPORTS

Durée et méthode d'enseignement	
1 heure	Introduction et discussion
Plusieurs jours	Travail en groupe

Introduction et discussion

- Inscrivez le canevas des rapports de recherche sur un transparent et discutez-en point par point. Soulignez que les constatations et conclusions, la discussion et les recommandations sont les parties les plus importantes.
- Prenez un exemple de l'un des groupes pour présenter un canevas possible du chapitre sur les constatations et les conclusions, avec des titres et sous-titres.

Expliquez la numérotation, en vous assurant que la disposition des titres et sous-titres est cohérente afin que vous puissiez utiliser cet exemple pour illustrer plus tard la présentation à adopter.

- Demandez aux participants de suggérer des critères sur lesquels ils pourraient se fonder pour évaluer leur première version, avant de leur donner des directives.
- Donnez des exemples tirés des propositions de recherche préparées par différents groupes pour expliquer comment l'énoncé du problème, les objectifs et la méthodologie doivent être adaptés en vue du rapport final.

Soulignez que le futur employé dans la proposition doit être remplacé par le présent et le passé, si vous croyez que certains groupes pourraient négliger cet aspect.

Travail en groupe

- Assurez-vous que tous les groupes préparent un canevas de leur rapport, en se fondant sur le modèle présenté dans le module. Demandez aux groupes d'afficher leur canevas pour que tous leurs membres puissent en prendre connaissance.
- Les sections sur les constatations et conclusions, la discussion et les recommandations sont celles qui nécessiteront le plus de temps. Certains groupes pourraient croire qu'il serait préférable d'en changer l'ordre. Ainsi, pour certains projets, il serait plus logique d'avoir les sections 4. Constatations, 5. Discussion et conclusions et 6. Recommandations. Dites aux participants qu'ils peuvent préparer leur canevas et présenter leur travail de la façon la plus appropriée à leur situation. Discutez avec les groupes de la façon de structurer cette partie de leur travail le plus logiquement possible.

- Les constatations et les conclusions devraient être rédigées en premier. La rédaction des chapitres d'introduction ne devrait être entreprise qu'une fois que le groupe a en main une bonne version préliminaire.
- Assurez-vous que les membres du groupe ont des aptitudes en rédaction, notamment en les faisant rédiger en équipes de deux. Les participants qui n'ont jamais écrit pourraient avoir besoin de beaucoup de soutien. Demandez-leur d'écrire quelques paragraphes que vous réviserez ensuite.
- Les groupes qui n'ont pas d'expérience en rédaction auront besoin de directives explicites sur les aspects à vérifier lors de l'étude de la première version, notamment la présentation du rapport.
- Il est préférable de discuter des recommandations possibles au cours de la rédaction des constatations et conclusions. Prenez note immédiatement de ces idées (de préférence sur un tableau à feuilles mobiles) pour vous en servir lors de la formulation des recommandations.
- Tous les animateurs devraient formuler des commentaires sur le sommaire préparé par les groupes. Chacun devrait également lire et commenter au moins un rapport préliminaire complet d'un autre groupe avant qu'il ne soit envoyé à la dactylographie.

**Formation à la recherche sur les systèmes de santé
Volume 2, 2^e partie : Analyse de données et rédaction de rapports**

Module 32

UTILISATION DES CONSTATATIONS

ÉTAPES DE L'ANALYSE DE DONNÉES ET DE LA RÉDACTION DE RAPPORTS

Questions à poser	Étapes à suivre*	Éléments importants de l'étape
Quelles données ont été recueillies pour chaque objectif de recherche? Les données sont-elles complètes et exactes?	Préparer les données pour l'analyse	Revoir les activités sur le terrain Compiler les données pour chaque objectif ou population étudiée Classer les données et en vérifier la qualité Vérifier les imprimés
À quoi ressemblent les données?	Décrire les variables	Distributions Figures, moyennes
Comment peut-on résumer les données pour en faciliter l'analyse?	Faire le recoupement des données quantitatives Résumer les données qualitatives	Faire le recoupement des données par rapport aux objectifs Illustrations, texte
Pour les données qualitatives : la recherche a-t-elle pour but de décrire, de comparer ou de trouver des associations?	Déterminer le type d'analyse statistique nécessaire	Examiner les objectifs, le type d'étude et les variables Faire une description statistique des variables Choisir les tests d'hypothèses
1. Comment peut-on décrire les données?	Analyser des observations appariées et non appariées	**Test t de Student **Test t pour observations appariées **Test du khi carré **Test du khi carré de McNemar
2. Comment peut-on déterminer les différences entre les groupes?	Évaluer les associations entre les variables	**Diagramme de dispersion **Droite de régression et coefficient de corrélation **Risque relatif, risque relatif approché
3. Comment peut-on déterminer les associations entre les variables?	Rédiger le rapport et formuler des recommandations	Préparer le canevas du rapport Rédiger les 1 ^{re} et 2 ^e versions Faire un sommaire des constatations Faire un sommaire des conclusions pour chaque objectif Formuler des recommandations Préparer un résumé analytique
Comment le rapport doit être rédigé?	Présenter un sommaire et rédiger un plan de mise en oeuvre des recommandations	Discuter des sommaires avec différents groupes cibles Discuter du programme de mise en oeuvre
Comment doit-on présenter et diffuser les constatations et les recommandations?		

* Il n'est pas nécessaire de faire ces étapes dans l'ordre indiqué. L'ordre peut être adapté aux besoins de l'atelier.

** Ces éléments sont facultatifs; ils peuvent être omis s'ils ne sont pas pertinents au cours.

Module 32 : UTILISATION DES CONSTATATIONS

OBJECTIFS

À la fin de la présente séance, vous devriez pouvoir :

1. **Présenter** un sommaire de vos constatations et recommandations aux gestionnaires de la santé et à d'autres intéressés;
2. **Rédiger** un plan d'action en vue de la mise en oeuvre des recommandations.

I. Introduction

II. Présentation des résultats de la recherche à des gestionnaires de la santé

III. Présentation des principales constatations et recommandations à d'autres intéressés

IV. Préparation d'un plan d'action

I. INTRODUCTION

Il y a plusieurs façons de promouvoir l'utilisation des constatations de votre recherche :

- Présentez les résultats de la recherche aux gestionnaires de la santé et aux dirigeants communautaires;
- Présentez les résultats à d'autres intéressés;
- Rédigez un plan d'action pour promouvoir la mise en oeuvre des recommandations formulées dans votre étude.

Le moyen le plus efficace d'encourager l'utilisation de vos résultats pourrait consister à combiner ces trois possibilités.

II. PRÉSENTATION DES RÉSULTATS DE LA RECHERCHE À DES GESTIONNAIRES DE LA SANTÉ

Le but des recherches sur les systèmes de santé consiste à fournir aux gestionnaires des renseignements qui faciliteront la prise de décisions. Une étape très importante du processus de recherche sur les systèmes de santé consiste donc en la présentation du rapport aux gestionnaires concernés afin que vous puissiez discuter avec eux des constatations et des recommandations. Il devrait s'agir d'un exposé en personne, où vous aurez la possibilité de discuter. Il ne suffit habituellement pas d'envoyer une copie du rapport ou du document de recherche aux gestionnaires.

Pour faire en sorte que votre exposé soit utile et productif,

1. Choisissez l'auditoire avec soin.

Assurez-vous d'inviter les gestionnaires concernés et leur personnel important, notamment du personnel clinique, des superviseurs, du personnel sur le terrain, du personnel provenant d'autres secteurs et d'organismes intéressés et, s'il y a lieu, des dirigeants communautaires¹. Il peut se révéler nécessaire de faire plusieurs exposés, en en modifiant le style et le contenu selon l'auditoire.

2. Assurez-vous qu'une période suffisante a été réservée à l'exposé et à la discussion.

Une durée minimum d'une demi-heure devrait être prévue pour l'exposé comme pour la discussion. Cependant, la discussion pourrait durer plus longtemps si elle porte sur les possibilités de mise en oeuvre des recommandations.

3. Organisez votre exposé.

Votre exposé devrait comprendre les éléments suivants :

¹ Un exposé distinct peut être préparé à l'intention de l'ensemble de la collectivité ou des patients qui ont participé à votre étude (voir la **section III**). Si les dirigeants communautaires sont des décideurs importants qui pourraient utiliser vos résultats et devraient participer avec les gestionnaires de la santé à la planification de la mise en oeuvre des recommandations, ils devraient être invités à ce premier exposé.

- Une brève introduction, comprenant l'énoncé du problème, les objectifs de l'étude, l'échantillon et les outils de collecte de données employés;
- Les principales constatations, placées dans un ordre logique (par exemple, en commençant par une description du problème suivie par les principales variables qui influent sur lui);
- Les recommandations, dans un ordre semblable.

Il est préférable de rédiger des recommandations distinctes à l'intention des décideurs, des gestionnaires de la santé, du personnel de la santé et des membres de la collectivité. Il est alors plus facile de discuter avec chaque groupe.

N'oubliez pas que votre auditoire tient avant tout à savoir comment résoudre le problème en question. Par conséquent :

- Évitez le jargon technique;
- Ne bombardez pas l'auditoire de statistiques. Présentez les constatations pertinentes qui doivent faire l'objet de rectifications. Cependant, vous pourriez présenter un ou deux tableaux sur transparents pour étayer vos principales conclusions, ou illustrer celles-ci par une observation intéressante.
- Soyez très précis concernant vos recommandations sur les mesures nécessaires pour résoudre le problème. Si votre recherche révèle qu'il existe plusieurs possibilités viables, décrivez-les en mentionnant leurs avantages et leurs inconvénients possibles. (Si vous présentez votre étude à des personnes susceptibles, vous pourriez sembler «leur dire quoi faire». Vous pouvez alors présenter différentes possibilités en illustrant clairement que les renseignements que donnent vos constatations pourront servir à choisir parmi ces possibilités, puis laissez les gestionnaires et la collectivité faire le choix définitif.)

4. Préparez le matériel audio-visuel approprié.

Préparez un nombre suffisant de copies de votre sommaire dactylographié des constatations et recommandations pour les distribuer aux personnes présentes. Si l'exposé a lieu **après** la préparation du rapport final, gardez à la portée de la main quelques copies du rapport que vous pourrez remettre aux personnes intéressées.

Préparez des transparents, des diapositives ou des tableaux à feuilles mobiles pour souligner les points les plus importants de votre exposé (p. ex., le problème, les principaux objectifs, les grandes constatations et les recommandations qui doivent donner lieu à des mesures de la part des participants).

5. Soyez disposé à discuter de la logique et de la faisabilité de vos recommandations.

Soyez également prêt à recevoir des suggestions, des critiques et des réactions et à apporter les modifications appropriées en conséquence.

6. Demandez à deux membres de l'équipe de rédiger un procès-verbal.

Assurez-vous de bien prendre note des discussions, surtout en ce qui concerne les décisions et les mesures de suivi qui ont fait l'objet d'une entente. Ce procès-verbal pourrait ensuite être distribué à tous les participants ainsi qu'à des personnalités importantes qui étaient invitées mais qui n'ont pu se rendre à l'exposé.

TRAVAIL EN GROUPE, PREMIÈRE PARTIE

1. Préparez votre exposé à l'intention d'un groupe de personnes choisies. Donnez-vous assez de temps pour la préparation : au moins une demi-journée pour rédiger le canevas de l'exposé et les notes dont vous aurez besoin et pour adapter votre sommaire des grandes constatations et des recommandations que vous distribuerez aux participants au besoin. Réservez une autre demi-journée à la dactylographie, aux corrections ainsi qu'à la préparation des photocopies des sommaires, des documents à distribuer et des transparents ou des tableaux à feuilles mobiles.
2. L'enquêteur principal peut se charger de la totalité de l'exposé, ou en assigner différentes parties à d'autres membres du groupe.

III. PRÉSENTATION DES PRINCIPALES CONSTATATIONS ET RECOMMANDATIONS À D'AUTRES INTÉRESSÉS (EXEMPLES)

Grand public ou patients

Si votre recherche porte sur un problème qui touche les membres de la collectivité ou des patients et si vous avez interrogé un certain nombre de ces personnes, il est souhaitable que vous leur rendiez compte également de vos grandes constatations et recommandations.

Votre exposé peut avoir plusieurs objectifs :

- Informer les dirigeants communautaires, les patients ou le grand public de vos constatations;
- Vérifier si ces personnes sont d'accord avec vos conclusions sur la nature, l'ampleur et les causes du problème selon les renseignements dont ils disposent;
- Recueillir des renseignements supplémentaires sur les questions qu'il reste à étudier;
- Informer ces personnes au sujet des recommandations formulées à l'endroit des décideurs et gestionnaires;
- Insister sur des recommandations qui les concernent, ou leur demander si ces recommandations sont opportunes et réalisables et de prendre des mesures;
- Demander des suggestions sur des mesures à prendre.

Autres chercheurs

Bien que l'adoption de mesures concrètes soit le but premier de votre recherche, il ne faut pas pour autant négliger sa valeur scientifique. Les chercheurs de départements de santé et de services sociaux, par exemple, pourraient s'intéresser à votre méthodologie et aux résultats de votre recherche. Il est donc important de les inviter à un exposé officiel de vos constatations ou de faire part de ces constatations lors de colloques ou de réunions organisées par des établissements de recherche intéressés. Vous pouvez également envisager de faire publier votre recherche dans un bulletin local ou international.

Médias

Pour susciter de l'intérêt et du soutien à l'égard de votre recherche, il est bon de communiquer vos constatations aux médias. N'oubliez pas que ceux-ci sont attirés par des informations qui ont des aspects humains; cherchez donc à mettre l'accent sur ces aspects. Cependant, évitez de vous mettre à dos les gestionnaires de la santé, les fournisseurs de soins de santé ou le grand public en divulguant des renseignements aux médias. Vous remettriez ainsi en question l'utilisation de vos constatations et la coopération future de ces intervenants. Pour éviter cette situation, faites participer des gestionnaires de la santé et des dirigeants communautaires à la publication des résultats.

IV. RÉDACTION D'UN PLAN D'ACTION

Au moment de la rédaction de votre proposition de recherche, vous avez discuté en groupe des résultats concrets que vous attendiez de votre étude et de la façon dont vous tenteriez de favoriser l'utilisation des résultats.

Une étape importante de ce processus consistait à faire participer tant les gestionnaires de la santé que la population cible à la formulation des recommandations.

Vous devriez maintenant déterminer comment vous pouvez contribuer à la mise en oeuvre de ces recommandations. Passez en revue la partie de votre proposition de recherche où vous avez formulé des suggestions sur la façon dont vos constatations pourraient être utilisées et développez ou adaptez-les.

1. Discutez des questions suivantes :

- Quelles recommandations pouvez-vous mettre en oeuvre vous-même sans autorisation ou soutien supplémentaire?
- Comment pouvez-vous procéder à la mise en oeuvre?
- Pour quelles recommandations avez-vous besoin du soutien de responsables? (Précisez les responsables en question et le soutien nécessaire.)
- De quelle façon pouvez-vous favoriser ce soutien?

2. Rédigez un plan d'action pour l'année à venir ou, si possible, pour une période plus longue, en précisant :

- les activités qui peuvent être tenues sans autorisation, le moment où elles auront lieu, les responsables et la façon dont elles seront tenues;
- d'autres mesures suggérées et les étapes à suivre pour favoriser leur mise en oeuvre.

3. Si certaines mesures nécessitent une planification détaillée de la part de plusieurs intervenants, vous pourriez envisager de tenir un atelier de planification. Les différents groupes participants pourraient alors collaborer à l'étude détaillée des constatations et des recommandations et élaborer un plan d'action. L'exposé à l'intention des gestionnaires et des dirigeants communautaires pourrait être élargi pour comprendre cette étape de planification, ou vous pourriez prévoir un atelier distinct après que les décideurs auront eu le temps d'étudier les résultats et leurs répercussions.

L'atelier pourrait durer quelques heures ou quelques jours, selon l'ampleur de l'étude et la nature des mesures à planifier. Si vous prévoyez un atelier de longue durée, vous pourriez envisager de faire une courte visite à quelques-uns des emplacements où a eu lieu la recherche avant que le groupe ne commence la planification en tant que telle.

Une fois les plans rédigés, assurez-vous que les groupes d'étude décident des activités et des tâches à remplir, qui en sera chargé, où elles seront tenues et quelles ressources seront nécessaires.

TRAVAIL EN GROUPE, DEUXIÈME PARTIE

1. S'il y a lieu, prévoyez un exposé destiné aux membres de la collectivité ou aux patients et déterminez comment vous obtiendrez des commentaires sur les constatations et les recommandations de votre recherche.
2. Préparez un plan de mise en oeuvre des résultats de votre recherche.
3. Si le plan de mise en oeuvre prévoit la participation de plusieurs personnes ou organismes, envisagez la tenue d'un «atelier de planification». Commencez à le planifier, s'il y a lieu.

Remarque : Pendant que la version finale de votre rapport est dactylographiée, vous aurez peut-être le temps de faire ce travail en groupe. Autrement, vous devrez le faire après l'atelier. Essayez de faire le plus de planification possible pendant l'atelier.

Notes du formateur

Module 32 : UTILISATION DES CONSTATATIONS

Durée et méthode d'enseignement

20 minutes Introduction et discussion

Introduction et discussion

Il est important de présenter ce module avant que les participants ne commencent à préparer le sommaire de leurs constatations et recommandations.

Remarque

Soulignez que le sommaire doit être inséré juste après les remerciements dans le rapport.

Travail en groupe

Lors de la présentation des projets de recherche, il est recommandé que **les participants distribuent le sommaire comprenant toutes les recommandations à tous les autres participants au cours et aux invités d'honneur**, afin de permettre à ceux-ci de mieux participer à la discussion. Des détails sur les objectifs spécifiques de l'étude, la méthodologie et les constatations (p. ex., certains tableaux et graphiques essentiels) peuvent être présentés au moyen de tableaux à feuilles mobiles ou de transparents.

Siège social du CRDI

CRDI, BP 8500, Ottawa (Ontario) Canada K1G 3H9

Bureau régional pour l'Afrique centrale et occidentale

CRDI, BP 11007, CD Annexe, Dakar, Sénégal

Bureau régional pour le Moyen-Orient et l'Afrique du Nord

CRDI, BP 14 Orman, Giza, Le Caire, Égypte

Bureau régional pour l'Afrique du Sud

CRDI, Ninth Floor, Braamfontein Centre, corner Bertha and Jorissen streets, Braamfontein, 2001 Johannesburg, Afrique du Sud

Bureau régional pour l'Afrique orientale et australe

CRDI, PO Box 62084, Nairobi, Kenya

Bureau régional pour l'Asie du Sud-Est et de l'Est

CRDI, Tanglin PO Box 101, Singapore 9124, République de Singapour

Bureau pour l'Asie du Sud

CRDI, 11 Jor Bagh, New Delhi 110003, Inde

Bureau régional pour l'Amérique latine et les Antilles

CRDI, Casilla de Correos 6379, Montevideo, Uruguay

Veillez adresser vos demandes d'information au sujet du CRDI et de ses activités au bureau de votre région.

L'AUTEURE

Ann Brownlee, MA, PhD, est sociologue de la médecine spécialisée dans la recherche sur les systèmes de santé, la planification et l'évaluation et d'autres aspects interculturels des soins de santé. Pendant plusieurs années, elle a occupé le poste de coordonnatrice de la recherche et de l'évaluation pour le Projet pour le renforcement des systèmes de prestation de soins de santé en Afrique centrale et occidentale, où elle a collaboré étroitement avec le bureau régional d'Afrique de l'OMS et avec ses collègues d'Afrique et d'ailleurs en vue de l'élaboration d'un programme de formation à la RSS et de subventions et de la publication d'un cours de formation à la RSS, précurseur du présent volume. Elle est actuellement experte-conseil en santé internationale pour des organismes tels que l'OMS, le CRDI et Wellstart, et elle enseigne à l'Université de Californie à San Diego.