



LES MÉTHODES D'ANALYSE EN RECHERCHE QUANTITATIVE : UNE INTRODUCTION AUX PRINCIPAUX OUTILS DISPONIBLES POUR LE CHERCHEUR

Thomas Rajotte¹

¹ Ergothérapeute, PhD, Professeur à l'Université du Québec à Rimouski – campus de Lévis, Québec, Canada

Adresse de contact : thomas_rajotte@uqar.ca

La **Revue Francophone de Recherche en Ergothérapie** est publiée par CARAFE, la Communauté pour l'Avancement de la Recherche Appliquée Francophone en Ergothérapie

doi:10.13096/rfre.v5n1.132

ISSN: 2297-0533. URL: <https://www.rfre.org/>



INTRODUCTION

L'analyse des données constitue l'une des étapes cruciales de tout processus de recherche (Dumas, 2000). En effet, après avoir défini son objet de recherche et son cadre de référence, le chercheur peut entamer l'opérationnalisation de son projet. À cette étape de la recherche, le choix d'une méthode permet de déterminer quels outils de collecte de données et quelles analyses seront utilisés pour répondre à la ou aux questions posées. L'arrimage de ces différentes phases implique la mise en place d'une réflexion rigoureuse de la part du chercheur. Cette réflexion repose sur une connaissance détaillée de l'éventail des outils méthodologiques disponibles.

Les principaux devis utilisés en recherche quantitative ont été présentés dans une précédente chronique (Rajotte, 2017). Ici, ce sont les principaux outils d'analyse de données quantitatives qui seront mis en avant, afin de permettre au chercheur de choisir la méthode la plus appropriée à sa situation.

L'ANALYSE DES DONNÉES QUANTITATIVES : LE COFFRE À OUTILS DU CHERCHEUR

Analyser adéquatement des données quantitatives requiert une bonne planification dans la mesure où l'on doit, avant même de procéder à la collecte de données, décider du ou des tests statistiques qui seront utilisés en fonction de la question et des données de recherche. Le coffre à outils présenté ici contient trois types d'analyse, soit : les analyses descriptives, les analyses inférentielles et les analyses corrélationnelles. Ce texte s'adresse à l'étudiant, au professionnel ou au chercheur, néophyte en méthodes quantitatives, qui souhaite avoir un éventail sommaire des outils qui lui sont accessibles. Pour une mise en œuvre approfondie de ces outils, il est fortement recommandé de se référer à des ouvrages spécialisés, comme celui de Beaucage, Simpson et Viger (2009), de Howell (2015) ou de Fortin et Gagnon (2015).

L'ANALYSE DESCRIPTIVE DES DONNÉES

L'analyse descriptive des données permet au chercheur de résumer un ensemble de données brutes à l'aide de techniques statistiques. Ce type d'analyse vise essentiellement à décrire les caractéristiques d'un échantillon et à répondre aux questions de recherche (Fortin et Gagnon, 2015). Les outils dont dispose le chercheur pour réaliser ce type d'analyse se répartissent en trois principales catégories : les mesures de tendance centrale, les mesures de dispersion et de position ainsi que les analyses de fréquences.

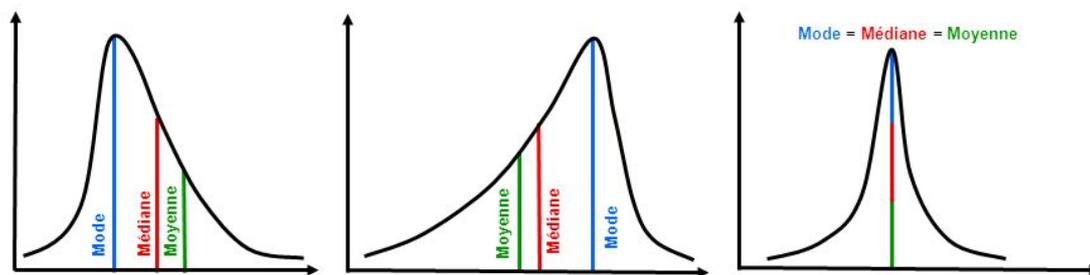
Les mesures de tendance centrale

Afin de procéder à une analyse descriptive de ses données, un chercheur peut utiliser des mesures de tendance centrale. Ces mesures constituent un indice

concernant le regroupement des valeurs attribuées aux données autour d'une valeur centrale. Afin de résumer la distribution de fréquences, une mesure de tendance centrale emploie un seul nombre. Trois mesures de tendance centrale sont utilisées dans les recherches : le mode, la médiane et la moyenne (Fortin et Gagnon, 2015). Une synthèse de ces trois mesures est présentée et illustrée dans le tableau 1.

Tableau 1. Présentation synthétique des trois mesures de tendance centrale

Le mode	Le mode correspond à la valeur de la variable statistique qui apparaît le plus souvent au sein d'un échantillon de données.
La médiane	La médiane réfère à la valeur de la variable telle qu'il y ait autant d'observations en dessous d'elle qu'au-dessus.
La moyenne	Mesure statistique caractérisant les éléments d'un ensemble de quantités : elle exprime la grandeur qu'aurait chacun des éléments d'un ensemble s'ils étaient tous identiques sans changer la dimension de l'ensemble.



Les mesures de dispersion et de position

Afin de réaliser une analyse descriptive de ses données, un chercheur peut recourir aux statistiques mesurant la dispersion et la position des données. Ces mesures constituent un indice d'étalement des données qui rend compte de la variation de celles-ci. Fréquemment, ces mesures de dispersion établissent une relation entre l'ensemble des données et la moyenne de l'échantillon (Zheng, Plaisent, Zuccaro, Bernard, Dagfous et Favreau, 2018). Le tableau 2 présente sommairement les principales mesures de dispersion et de position.

Tableau 2. Présentation sommaire des principales mesures de dispersion et de position

L'écart-type	Mesure de dispersion autour de la moyenne. Cette mesure correspond à l'écart moyen par rapport à la moyenne.
La variance	Mesure de dispersion qui consiste en la somme des carrés des écarts par rapport à la moyenne, divisée par le nombre d'observations.
L'étendue	L'étendue représente la différence entre les valeurs extrêmes d'une distribution/d'un ensemble.
Le minimum	Le minimum correspond à la plus petite valeur d'une distribution ou d'un ensemble.
Le maximum	Le maximum réfère à la plus grande valeur d'une distribution ou d'un ensemble.
Le percentile	Le rang percentile indique le pourcentage des données ayant une valeur inférieure ou égale à la donnée considérée.

Les analyses de fréquences

Finalement, afin de réaliser une analyse descriptive des données recueillies, il est également possible de procéder à une analyse de fréquences. Ce type d'analyse consiste à calculer le nombre d'observations, pour chacune des modalités d'une variable, ou par groupe de modalités (par classe). La fréquence absolue est basée sur le nombre effectif d'observations et la fréquence relative sur le ratio entre ce nombre effectif et l'ensemble des observations (Zheng *et al.*, 2018). Par exemple, la fréquence absolue des âges dans un échantillon peut s'exprimer de la manière suivante : 16 ans : $n = 5$; 17 ans : $n = 30$; 18 ans : $n = 10$; 19 ans : $n = 5$. En fréquence relative, les données seront résumées de la manière suivante : 16 ans : $n = 10\%$; 17 ans : $n = 60\%$; 18 ans : $n = 20\%$; 19 ans : $n = 10\%$. Enfin, les données peuvent être regroupées, par exemple en deux classes de plus ou de moins de 18 ans : moins de 18 ans : $n = 35$; 18 ans et plus : $n = 15$. L'analyse de la fréquence est fondamentale pour décrire les données, car elle permet, par exemple, d'établir si les données suivent des distributions connues, comme la loi normale, ce qui influencera le choix du test statistique dans le cas d'une analyse inférentielle. Elle permet également de voir rapidement quelles sont les valeurs surprenantes. En épidémiologie, elle est utilisée pour décrire des maladies au travers de concepts comme la prévalence ou l'incidence.

Exemple de recherche ayant abordé les trois catégories d'analyses descriptives

En 2017, Dumont a élaboré une recherche présentant les trois catégories d'analyse descriptive. En fait, le projet de cette auteure était d'explorer l'intérêt de l'utilisation des technologies mobiles (TM) pour améliorer le rendement occupationnel des personnes ayant un trouble du spectre de l'autisme (TSA). Cette ergothérapeute a ainsi utilisé des analyses de fréquence pour décrire les caractéristiques des répondants ayant participé à son projet d'étude (sexe, âge, scolarité, rôle principal, milieu de vie et capacité à lire/écrire). Le tableau suivant présente les fréquences absolues et relatives pour différentes variables, de différentes natures (dichotomique, discrète ou continue).

Tableau 3. Caractéristiques des répondants – tiré de Dumont (2017)

	Initial (n = 32) n (%)	Final (n = 26) n (%)
Sexe		
Masculin	23 (72)	17 (65)
Féminin	9 (28)	9 (35)
Âge		
6 à 10,9 ans	4 (12)	2 (8)
11 à 14,9 ans	6 (19)	5 (19)
15 à 17,9 ans	11 (34)	9 (34)
18 ans et plus	11 (34)	10 (38)
Scolarité		
Maternelle	1 (3)	1 (4)
Primaire	4 (13)	2 (8)
Secondaire	6 (19)	3 (12)
Collégiale	2 (6)	2 (8)
Spéciale	19 (59)	18 (69)
Rôle principal		
Étudiant	28 (87)	22 (85)
Travail adapté	1 (3)	1 (4)
À domicile	3 (10)	3 (12)
Milieu de vie		
Domicile	25 (78)	19 (73)
Résidence spécialisée	7 (22)	7 (27)
Sait lire et écrire		
Oui	23 (72)	17 (65)
Non	9 (28)	9 (35)

Dumont a également utilisé des mesures de dispersion et de position (écart-type, minimum et maximum) ainsi que deux mesures de tendances centrales (moyenne et médiane), pour mettre en lumière les caractéristiques de l'intervention réalisée en lien avec l'utilisation des technologies mobiles. Celles-ci sont présentées dans le tableau ci-dessous.

Tableau 4. Caractéristiques de l'intervention – tiré de Dumont (2017)

	Moy	ÉT	Méd	Min	Max
Nombre d'interventions directes (n = 26)	8,3	7,2	7	1	38
Nombre de jours entre les évaluations initiale et finale (n = 26)	120,7	48,5	99	47	207
Temps total d'interventions directes (minutes) (n = 26)	508,9	648,6	400	60	3370
Temps total d'interventions indirectes (minutes) (n = 26)	529,2	905,5	147,5	60	3965

Moy : moyenne ; ÉT : écart-type ; Méd : médiane ; Min : minimum ; Max : maximum

L'ANALYSE INFÉRENTIELLE DES DONNÉES

L'analyse inférentielle des données correspond à la branche de la statistique qui a pour objet de généraliser les résultats obtenus à partir d'un échantillon à l'ensemble de la population d'où il a été tiré (Amyotte et Côté, 2017). Le recours à ce type d'analyse est déterminé par des considérations de divers ordres (économique, pratique, etc.).

Ainsi, il peut être trop onéreux, voire impossible d'avoir accès à tous les constituants d'une population donnée. Pour ces raisons, il peut être judicieux d'inférer la valeur d'une caractéristique d'une population à partir de la valeur de cette même caractéristique mesurée auprès d'un échantillon (Dumas, 2000).

Dans le cadre de cette chronique méthodologique, seulement deux types d'analyses inférentielles seront décrits : les tests t ainsi que les analyses de variance (ANOVA). Par ailleurs, il importe de mentionner qu'en fonction de ses besoins spécifiques, le chercheur peut mettre en œuvre un éventail accru d'analyses inférentielles.

Le test t ou test de Student

Il existe de nombreux tests visant à étudier la relation entre des données de deux ou plusieurs populations. Le test t est un test paramétrique¹ visant à déterminer la différence entre les moyennes de deux populations, et ce, relativement à une variable prédéterminée (Dumas, 2000). Ce test statistique est l'un des plus couramment utilisés dans le domaine de la santé et de l'éducation (Fortin et Gagnon, 2015).

En fonction du type d'échantillon choisi dans le protocole de recherche, il existe deux types de tests t , soit le test t pour échantillons indépendants ou le test t pour échantillons dépendants/appariés. Le test t pour échantillons indépendants sert à déterminer l'évolution d'une variable à partir d'une comparaison entre deux groupes différents, tandis que le test t pour échantillons dépendants/appariés vise plutôt à étudier le comportement d'une variable continue qui a été évaluée à deux occasions auprès d'un même groupe d'individus (Fortin et Gagnon, 2015).

L'origine du test t : un souci provenant de la qualité de la bière irlandaise

Le test t a été développé par William Gosset et fut diffusé pour la première fois dans la revue *Biometrika* (1908). À ce moment, Gosset travaillait en tant que chimiste à la brasserie Guinness à Dublin en Irlande. Ce test statistique visait essentiellement à permettre un contrôle, ainsi que des comparaisons, de la qualité gustative des bières noires (stout) produites au sein de la brasserie. Afin d'éviter d'éventuels conflits de travail avec ses collègues, Gosset a publié ses travaux sous le pseudonyme *Student* (sans aucun prénom associé). Cette publication est à l'origine de l'autre appellation du test t , le test de Student.

L'analyse de variance (ANOVA)

L'analyse de la variance (ANOVA) permet d'examiner une variable quantitative à expliquer en fonction des effets d'une ou de plusieurs variables nominales, c'est-à-dire en fonction de catégories (comme le genre, une couleur, une maladie). L'analyse de la variance multiple (MANOVA) permet d'analyser plusieurs variables à expliquer simultanément (Howell, 2015). Selon les cas, on souhaitera peut-être approfondir l'analyse, en vérifiant si certains facteurs ont un effet sur la variable à

¹ Tel que mentionné par Dumas (2000), le terme « paramètre » ou « paramétrique » signifie que la mesure ou le test effectué est réalisé auprès d'une population.

expliquer, après avoir enlevé la variance due à d'autres facteurs quantitatifs continus, dits covariables ; il s'agira alors d'une analyse de la covariance (ANCOVA). Si l'étude cherche à établir une relation linéaire entre une variable à expliquer et une ou plusieurs variables explicatives continues, le modèle de régression linéaire simple ou multiple peut être appliqué. Ces différents types d'analyse sont présentés à l'intérieur du tableau 5.

Tableau 5. Description synthétisée de certains termes se rapportant aux analyses inférentielles

L'analyse de variance (ANOVA)	Test statistique paramétrique visant à déterminer les différences entre deux groupes ou plus (habituellement trois) par rapport à une variable dépendante.
L'analyse de covariance (ANCOVA)	L'analyse statistique visant l'effet d'une variable dépendante (ou de plusieurs), et ce, en contrôlant l'effet potentiel d'une covariable (variable contrôlée par le chercheur).
L'analyse de variance multiple (MANOVA)	L'analyse de variance multiple (MANOVA) est une extension de l'ANOVA qui permet de prendre en compte une combinaison de variables dépendantes plutôt qu'une seule.

LES ANALYSES CORRÉLATIONNELLES

Dans le cadre de certains projets de recherche, les variables sont étudiées et analysées sans qu'il y ait de manipulation expérimentale. Ce type de projet implique que le chercheur observe des variables et mesure leurs valeurs sans mettre en place une intervention ou une expérimentation. Par la suite, le chercheur tentera de mettre au jour des relations entre les variables, en utilisant un coefficient de corrélation. Le coefficient de corrélation de Pearson permet de mesurer la force de la relation linéaire entre deux variables quantitatives, normalement distribuées (cf. plus haut, les analyses de fréquence) (Fortin et Gagnon, 2015). Si le coefficient de corrélation est positif, cela signifie que lorsque la valeur de l'une des variables évaluées augmente, la valeur de la seconde variable évaluée augmentera aussi. Par contre, si le coefficient de corrélation est négatif, cela veut dire qu'une augmentation de la valeur de l'une des variables correspond à une diminution de celle de l'autre variable considérée (Bourque et El Adlouni, 2016).

Un outil complémentaire du chercheur désirant réaliser une étude corrélative est l'utilisation de la régression multiple. La régression multiple consiste à évaluer la relation entre une variable (souvent appelée la variable prédite ou la variable dépendante) et deux autres variables ou plus (souvent appelés « prédicteur » ou variables indépendantes).

CONCLUSION : QUE RETENIR DE CETTE CHRONIQUE MÉTHODOLOGIQUE ?

Au terme de cette chronique, il importe de souligner que les outils disponibles au chercheur souhaitant réaliser une étude quantitative sont nombreux. Cette chronique s'adressait au débutant et visait essentiellement à dresser un portrait des principales analyses permettant de mettre en œuvre un projet de recherche. Le choix de l'analyse statistique à appliquer découle de l'arrimage entre les questions, les objectifs de recherche et le devis choisi. S'il y a une chose à retenir de cette chronique, c'est que tout chercheur bénéficie d'un éventail diversifié d'outils d'analyse adaptés à toutes sortes de situations. Dans le domaine de la recherche, notre coffre à outils est bien plus rempli que ce que l'on pourrait croire. Il serait désolant de voir un chercheur passer à côté d'une piste d'étude innovante, par manque de connaissance concernant les analyses statistiques disponibles. Après ce premier portrait, nous invitons donc les lecteurs à faire d'autres pas et à explorer plus avant ce riche monde des méthodes quantitatives.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Amyotte, L. et Côté, C. (2017). *Complément de méthodes quantitatives. Applications à la recherche en sciences humaines* (2^e éd.). Montréal, QC : ERPI.
- Beaucage, C., Simpson, A. et Viger, Y. B. (2009). *Épidémiologie appliquée*. Montréal : G. Morin.
- Bourque, J. et El Adlouni, S.E. (2016). *Manuel d'introduction à la statistique appliquée aux sciences sociales*. Québec, QC : Presses de l'Université Laval.
- Dumas, C. (2000). L'analyse des données de base. Dans R. J. Vallerand et U. Hess. (dir.), *Méthodes de recherche en psychologie*. Boucherville, QC : Gaëtan Morin.
- Dumont, C. (2017). Les technologies mobiles pour les personnes ayant un trouble du spectre de l'autisme : une étude exploratoire. *Revue francophone de recherche en ergothérapie*, 3(2), 39-62.
- Fortin, M.-F. et Gagnon, J. (2015). *Fondements et étapes du processus de recherche. Méthodes quantitatives et qualitatives* (3^e éd.). Montréal, QC : Chenelière Éducation.
- Howell, D. C. (2015). *Méthodes statistiques en sciences humaines*. Louvain-la-Neuve, BE : De Boeck.
- Rajotte, T. (2017). La recherche expérimentale en éducation : quelles sont les limites à considérer ? *Revue francophone de recherche en ergothérapie*, 3(2), 63-73.
- Student (1908). The probable error of a mean. *Biometrika*, VI(1), 1-25.
- Zheng, B., Plaisent, M., Zuccaro, C., Bernard, P., Daghfous, N. et Favreau, S. (2018). *L'analyse des données de sondage avec SPSS. Un guide d'introduction*. Québec, QC : Presses de l'Université du Québec.