**Compte rendu du colloque**« **Méthodes Quantitatives en Sciences Humaines (MQSH) 2023** »

Pier-Olivier Caron

Université TÉLUQ

0000-0001-6346-5583

Agathe Bellemare-Lepage a

Université du Québec à Montréal

0000-0003-4123-8771

Juliette Casgrain a

Université du Québec à Montréal

0009-0000-4294-0293

André Achim b

Université du Québec à Montréal

0000-0003-2159-6010

Sébastien Béland b

Université de Montréal

0000-0002-0324-5076

Denis Cousineau b

Université d’Ottawa

0000-0001-5908-0402

Eric Frenette b

Université Laval

0000-0001-6503-535X

Éric Lacourse b

Université de Montréal

0000-0002-4779-9900

Félix Laliberté b

Université de Montréal

0009-0003-9073-5595

Louis Laurencelle b

Université du Québec à Trois-Rivières

0000-0003-3448-2872

Titre court : MQSH 2023

a Les deux autrices ont contribué également au présent papier.

b Les auteurs sont en ordre alphabétique.

Les auteurs approuvent ce manuscrit, et ne rapportent aucun conflit d’intérêt.

Correspondance :

Pier-Olivier Caron

Département des Sciences Humaines, Lettres et Communication

Université TÉLUQ

5800, rue Saint-Denis, bureau 1105

Montréal (Québec) H2S 3L5

Canada

Courriel: pier-olivier.caron@teluq.ca

**Résumé**

Le vendredi 9 juin 2023 s’est tenu le 12e colloque annuel Méthodes quantitatives en sciences humaines à l’Université TÉLUQ, Montréal. Sept conférenciers ont présenté leurs résultats de recherche. Sébastien Béland a évalué l'efficacité de différents coefficients populaires sur des items unidimensionnels à réponse dichotomique. Félix Laliberté a présenté *StepMix*, un packagePython et R pour le clustering basé sur des modèles et la modélisation de mélanges généralisés (classes latentes, profils latents) de données continues et catégorielles. Eric Frenette a fait part des défis méthodologiques et statistiques rencontrés lors de l'évaluation du programme longitudinal *Gagnant pour la vie* (version 1.0) implanté dans une école à concentration sport au Québec. André Achim a présenté une nouvelle technique de symétrisation de distributions asymétriques, même en présence de valeurs aberrantes. Pier-Olivier Caron a présenté une comparaison des propriétés statistiques (puissance, erreur de type I, biais) des modèles de médiations longitudinaux et transversaux. Denis Cousineau a présenté deux techniques entièrement analogues à l'ANOVA : l'analyse des proportions utilisant la transformation arcsine (ANOPA) et l'analyse des fréquences de données (ANOFA). Louis Laurencelle propose un test complet et vraisemblable sur le plan paramétrique de la différence entre deux proportions jumelées. Après trois ans d’absence, la résurrection du colloque MQSH a été un succès marqué de débats enthousiastes qui ont mis en lumière des problèmes originaux, l'importante approche critique pour les étudier, mais aussi l’originalité des solutions apportées pour les résoudre.

**Mots-clés :** Statistiques, Actes de colloque, MQSH 2023.

**Abstract**

The 12th annual meeting Quantitative Methods in the Humanities (*Méthodes quantitatives en sciences humaines*, MQSH) was held on Friday, June 9, 2023 at Université TÉLUQ, Montreal. Seven speakers presented their research findings. Sébastien Béland assessed the effectiveness of various popular coefficients on unidimensional items with dichotomous responses. Félix Laliberté presented StepMix, a Python and R package for model-based clustering and generalized mixture modeling (latent classes, latent profiles) of continuous and categorical data. Eric Frenette presented the methodological and statistical challenges encountered during the evaluation of the longitudinal program *Gagnant pour la vie* (version 1.0 ; Eng. *Win for life*) implemented in a sports-oriented school in Quebec. André Achim presented a new technique for symmetrizing asymmetrical distributions, even in the presence of outliers. Pier-Olivier Caron presented a comparison of the statistical properties (power, type I error, bias) of longitudinal and cross-sectional mediation models. Denis Cousineau presented two techniques entirely analogous to ANOVA: analysis of proportions using the arcsine transformation (ANOPA) and analysis of data frequencies (ANOFA). Louis Laurencelle proposed a complete and parametrically plausible test of the difference between two paired proportions. After a three-year absence, the resurrection of the MQSH colloquium was a success, marked by enthusiastic debates that highlighted original problems, the important critical approach to studying them, but also the originality of the solutions proposed to solve them.

**Keywords :** Statistics, Proceedings, MQSH 2023.

**Compte rendu du colloque**« **Méthodes Quantitatives en Sciences Humaines (MQSH) 2023** »

L’objectif de cet article est de faire le compte rendu du 12e colloque Méthodes Quantitatives en Sciences Humaines (MQSH) ayant eu lieu le 9 juin 2023 à l’Université TÉLUQ à Montréal. Le compte rendu sera organisé de sorte à présenter les conférences dans l’ordre où elles ont eu lieu.

**Mot de bienvenue**

Le colloque commence par le mot de bienvenu de Louis Laurencelle (Université du Québec à Trois-Rivières). Après les salutations d’usage, Laurencelle souligne sa fierté de voir ressusciter, après trois ans d’arrêt dû à la pandémie mondiale de COVID-19, cette rencontre annuelle, qu’il juge importante pour la santé et le progrès des méthodes quantitatives en sciences humaines au Québec et en milieux francophones. Par la suite, Laurencelle fait état du développement du colloque MQSH depuis sa création en 2009 où il imaginait un colloque voué, non pas aux applications concrètes des méthodes quantitatives, mais aux méthodes elles-mêmes et à la compétence professionnelle de ceux qui les pratiquent.

Selon lui, ce travail des spécialistes de la mesure et des approches quantitatives en sciences humaines repose à la fois sur la compétence, le sens critique, et la créativité. Compétence, critique et créativité : ces trois « C » pourraient être la bannière des quantitativistes que pourrait servir l’« association MQSH » si elle existait. Elle pourrait promouvoir tous ceux et toutes celles qui exercent la fonction de conseiller et conseillères en méthodes quantitatives. Il importe que ces personnes soient à la hauteur et soient reconnues comme telles dans leur milieu. C’était l’objectif du colloque en 2009, ce l’est toujours en 2023.

**Comparaison de coefficients de fidélité des scores populaires pour items à réponse dichotomique : une étude de simulation**

La conférence de Sébastien Béland (Université de Montréal) a pour but de comparer l’efficacité de huit coefficients de fidélité sur des items unidimensionnels à réponse dichotomique. Il souligne d’entrée de jeu que la majorité des écrits sur la fidélité des scores ont été réalisés sur des items à réponse continue. Ainsi, il est difficile d’avancer des conclusions concernant l’efficacité de coefficients de fidélité sur des items à réponse dichotomique. Béland commence par proposer une stratégie permettant d’estimer la valeur vraie de la fidélité des scores pour trouver le meilleur estimateur en présence de données dichotomiques. La méthode quadrature rectangulaire est retenue comme étant la plus précise.

Des comparaisons sont effectuées par la suite à l’aide de la simulation informatique. Le devis de simulation de cette deuxième étude croise trois longueurs de test (15, 40, 65) et cinq tailles d’échantillon (100, 300, 500, 1000, 3000). Pour chacune des combinaisons, 1000 matrices sont générées pour un total de (3×5)×1000=15000 matrices. Les résultats témoignent que, en comparaison aux autres coefficients, la stratégie basée sur le modèle à deux paramètres, l’oméga total basé sur le modèle bifactoriel et un coefficient découlant de l’analyse factorielle confirmatoire ont les taux d’erreurs les plus faibles. Au contraire, le coefficient GLB présente des erreurs importantes, ce qui est soutenu par des études antérieures. Ce coefficient n’est par conséquent pas recommandé pour des données dichotomiques. De plus, les résultats et de nombreux écrits scientifiques montrent que le de MacDonald et le de Cronbach ne sont pas des coefficients interchangeables.

Béland conclut en avançant que les coefficients basés sur des modèles pour réponses continues sont pertinents pour analyser des réponses dichotomiques et recommande leur utilisation. Ces résultats suggèrent que les modèles de mesure pour items continus sont plus robustes que plusieurs tendent à le croire.

**StepMix : une nouvelle librairie Python et R pour les modèles de mélange fini avec des variables externes**

La conférence de Félix Laliberté (Université de Montréal), étudiant d’Éric Lacourse (Université de Montréal), consiste à présenter l’utilisation de StepMix à l’aide d’une analyse de classes latentes. StepMix est une nouvelle librairie Python et R pouvant servir à modéliser des modèles de mélange fini, dont des modèles structurels qui incluent des variables externes. La libraire permet d’identifier l’hétérogénéité dans une population par la création de sous-groupes latents, c’est-à-dire des sous-groupes d’individus qui ne peuvent être directement observés dans les données. Laliberté souligne les nombreux avantages de StepMix en comparant cette librairie aux autres librairies déjà existantes. La spécificité de StepMix réside principalement dans le fait que les sous-groupes latents cernés peuvent ensuite être liés à des variables externes (covariables et/ou variables dépendantes distales) en utilisant les approches modernes et robustes d’estimation en étapes de la famille des modèles de mélange finis généralisés. StepMix est aussi facile d’utilisation, permet la gestion des données manquantes à l’aide de la méthode du maximum de vraisemblance complète (FIML) et rend possible l'inférence dans des contextes d’apprentissage automatique non supervisé par l’utilisation du bootstrap non paramétrique.

StepMix possède une interface modulable polyvalente permettant plusieurs méthodes d'estimation par maximisation de l'espérance (EM) par étapes basées sur la théorie de la pseudovraisemblance. Laliberté présente d’abord les approches classiques, telles que l’approche en une étape et l’approche en trois étapes dite « naïve ». Il souligne divers problèmes, notamment que l’approche en une étape tend à rendre le modèle de mesure ininterprétable. Ensuite, Laliberté présente les approches modernes, comme l’approche en trois étapes avec correction BCH ou ML et l’approche en deux étapes. L’approche en trois étapes avec correction BCH ou ML permet d’estimer les paramètres du modèle structurel corrigés de l’incertitude d’assignation aux classes. Or, comme le mentionne Laliberté, cette approche est surtout disponible sur des logiciels payants, ce qui contribue au fait que les approches classiques sont encore utilisées dans les écrits aujourd’hui. Puisque StepMix est disponible dans les logiciels libres d’accès R et Python, les approches en étapes modernes sont désormais plus accessibles.

Laliberté conclut en mentionnant que diverses fonctionnalités sont encore à développer comme les erreurs types et diverses analyses utilisant des données longitudinales (p. ex., analyse de transitions latentes), mais que le logiciel est déjà fort prometteur.

**Défis rencontrés en évaluation : le cas du programme *Gagnant pour la vie* 1.0**

Cette présentation d’Eric Frenette (Université Laval) fait état des défis méthodologiques et statistiques rencontrés lors de l’évaluation du programme *Gagnant pour la vie 1.0* et fait un retour sur la qualité de la mesure. Ce programme implanté dans une école secondaire avec concentration sport porte sur l'accompagnement d'enseignants et d'entraîneurs dans l'enseignement et le transfert de cinq habiletés de vie auprès d’élèves sur une période de trois ans : fixation d'objectifs, concentration, saines habitudes alimentaires, comportements sécuritaires et récupération physique et mentale.

Bien que les qualités psychométriques soient considérées comme acceptables à chaque temps de mesure, Frenette mentionne que les chercheurs ont rencontré cinq défis principaux. Le premier concerne la qualité des données sociodémographiques. Selon le présentateur, pour des jeunes de 12 à 17 ans, les questionnaires étaient trop nombreux ou trop longs à remplir. Cela a occasionné un nombre important de données incomplètes qui pouvaient difficilement être déduites. Il explique que d’autres données étaient remplies de façon inconstante selon les temps de mesure, notamment la date de naissance, la taille (système métrique ou impérial) et le sexe. Deux solutions ont été apportées pour de futures collectes, soit de s’assurer de la codification des variables dans le sondage en ligne et d’utiliser un menu déroulant plutôt que de demander l’écriture dans les questionnaires. Le deuxième défi concerne la présence de données manquantes, puisque sur les 212 participants, seulement 53 ont répondu à tous les temps de mesure. La pandémie de COVID-19 a contribué à cet enjeu, car des collectes de données ont été retardées. Notamment, lors de la dernière, les élèves athlètes en secondaire 5 avaient quitté l’école et n’ont, pour la plupart, pas répondu aux questionnaires. Le test Little’s MCAR a permis d’indiquer que les données semblent manquantes de façon aléatoire et les méthodes d’imputation multiple (*Markov Chain Monte Carlo method and linear regression*) ont été utilisées pour estimer les données manquantes. Le troisième défi concerne la mesure des saines habitudes alimentaires. Frenette explique que le questionnaire utilisé était protégé (payant), comportait un nombre important de variables dont le calcul de scores composites peut parfois être déduit, mais pas pour la majorité des variables. Le quatrième défi exposé concernait le groupe témoin que le présentateur considérait comme inutile. Effectivement, il a expliqué que le groupe témoin provenait d’une seconde école où le programme n’avait pas été implanté. Toutefois, un programme similaire a été mis en place, ce qui a contaminé le groupe témoin. Le dernier défi présenté est que peu de résultats significatifs ont été observés, ce qui, ajouté aux autres défis mentionnés en début de présentation, limite considérablement la dissémination des résultats.

L'évaluation de l'implantation du programme n'ayant pas permis d'atteindre tous les objectifs souhaités, une analyse des processus méthodologiques et statistiques est effectuée. Ces analyses révèlent qu’il était difficile d’évaluer un effet indirect chez les élèves/athlètes puisque le programme est destiné aux enseignants et entraîneurs. Également, Frenette explique qu’un devis longitudinal amène inévitablement diverses sources potentielles d’erreur de mesure, par exemple une implantation différente du programme et la mouvance des élèves (attrition). Il a mentionné qu’il est important que la mesure puisse discriminer le niveau des élèves/athlètes.

Le présentateur conclut avec plusieurs pistes de réflexion concernant la qualité de la mesure de ce projet de recherche. Il mentionne notamment que les questionnaires utilisés n'étaient pas adaptés au contexte longitudinal et qu'il était nécessaire d'améliorer le pouvoir discriminant de la mesure afin de détecter des changements.

**Stratégie pour identifier des valeurs aberrantes dans des distributions asymétriques.**

André Achim (Université du Québec à Montréal) commence d’entrée de jeu par réorienter sa présentation qui s’est précisée sur la « Symétrisation de distributions asymétriques même en présence de valeurs aberrantes », nouveau titre de la présentation. Celle-ci porte sur la transformation de données par une nouvelle technique, nommée reSym, pour symétriser des distributions avec données aberrantes. En partant de ces propres travaux sur l’estimation de la dimensionnalité en analyse factorielle exploratoire, Achim souligne que les distributions asymétriques posent problème, car elles peuvent créer des facteurs fantômes. La technique reSym permet de symétriser automatiquement les distributions (pour celles qui le peuvent) hors de l’influence des valeurs aux extrêmes. Il s’agit, après avoir polarisé l’asymétrie de la variable pour qu’elle soit positive (éparse à droite) puis ramené le minimum à 0, de produire une transformation telle que la constante fasse que la transformation ramène deux valeurs témoins de la distribution observée (p. ex., celles aux 5e et 95e centiles) à égale distance de la médiane aussi transformée.

Achim note qu’en fixant la médiane de variable transformée à zéro et son échelle telle que la distance de la médiane à chacun des témoins ait une cote correspondant aux centiles utilisés, on peut traiter les valeurs aux extrémités de la distribution transformée comme des pseudo-cotes pour décider si elles appartiennent ou non à la distribution.

Achim termine en exposant plusieurs considérations à prendre en compte. D’abord, il rappelle l’importance de pouvoir rapporter des résultats, comme les moyennes, dans la métrique originale pour faciliter l’interprétation ; il précise que reSym donne la transformation inverse pour chaque variable transformée. Il recommande ensuite une transformation commune en présence de plusieurs groupes. Pour de petits échantillons, il est possible de spécifier d’omettre un petit nombre de valeurs à chaque extrémité au lieu de spécifier des rangs centiles. Il a également expliqué que reSym s’applique aussi à des données ordinales avec des observations multiples à chaque rang et avise si la symétrie n’est pas atteinte. L’approche reSym est disponible en MATLAB en contactant l’auteur et le sera bientôt en R.

**Détecter les médiations longitudinales à l’aide d’analyses transversales : une étude sur l’impact de la spécification de modèle sur la puissance statistique et le taux d’erreur de type I**

La présentation de Pier-Olivier Caron (Université TÉLUQ) a comme objectif de vérifier les propriétés statistiques (puissance, erreur de type I, biais) d’un modèle de médiation transversale (CSM) comparé aux modèles de médiation par panel à décalage croisé (CLPM). Le présentateur commence par préciser les limites d’utiliser des modèles erronés, et ce, connu d’emblée, pour répondre à des questions d’ordre quantitatives.

Contrairement aux recommandations, plusieurs études utilisent un modèle transversal plutôt qu’un modèle longitudinal. La comparabilité des propriétés statistiques de CSM et de CLPM n’est pas claire. Pour les examiner et les comparer, 97 scénarios (485 au total) répliqués 5000 fois sont réalisés à l’aide d’une simulation de Monte-Carlo menée avec R. Les effets indirects (αβ), les effets directs (γ), les trajectoires autorégressives (τ) et les tailles d'échantillon (n) sont manipulés. Les résultats montrent que le modèle CSM peut détecter un effet indirect longitudinal lorsque les trajectoires autorégressives sont élevées. Dans ces mêmes conditions, la puissance de CSM est acceptable ; le CSM et le CLPM s’accordent sur l’effet indirect. Lorsque les trajectoires autorégressives sont très faibles, le CSM manque de puissance. Il y a un risque de biais avec le CSM par rapport au CLPM puisqu’une surestimation des faux positifs et une sous-estimation des vrais positifs sont attendues. De plus, il y a un taux d'erreur de type I alarmant pour l’effet indirect du CSM si et sont non nuls.

Caron conclut en mentionnant qu’il ne recommande tout de même pas l’utilisation de médiations transversales pour détecter des médiations longitudinales, à moins que les ressources des chercheurs soient limitées, qu’il soit impossible d’appliquer un modèle longitudinal ou qu’il s’agisse d’une première étape exploratoire.

**L'analyse de variance appliquée au traitement des proportions et des fréquences**

Cette présentation de Denis Cousineau (Université d’Ottawa) et de Louis Laurencelle (Université du Québec à Trois-Rivières) fait état de deux nouvelles méthodes d’analyses analogues aux ANOVA, soit l'analyse des proportions (ANOPA) et l'analyse des fréquences de données (ANOFA). Elles consistent à décomposer une statistique de variation globale en effets et en erreurs, donnant lieu à un tableau sommaire de type ANOVA. Ces techniques permettent d'examiner des effets principaux et d'interaction, en plus de décomposer un effet global en effets simples, de fractionner une interaction et de produire des ensembles de contrastes orthogonaux.

Cousineau débute sa présentation par un prologue concernant la vraie contribution de Fisher, soit l’idée d’effet qui permet de simplifier l’investigation des données (l’effet colonne, l’effet ligne et l’effet d’interaction). Il présente ensuite l’ANOFA dans le cas de devis à mesures répétées. Le seul postulat préalable à cette analyse est qu’il est nécessaire de connaître la taille de l’échantillon et que les données peuvent être classifiées. Pour procéder, on débute en classifiant des participants en classe d’appartenance, puis on regarde combien tombent dans chacune des classes. On calcule le ratio de vraisemblance et on applique deux fois le négatif du log ratio pour obtenir la statistique , soit . L’innovation vient du fait qu’il est possible de décomposer de façon additive sans perdre aucune information, et ce, pour plusieurs facteurs. On obtient également des intervalles de confiance et une taille d’effet qui peut être utilisée pour des analyses de puissance a priori. Cousineau illustre finalement la méthode avec une démonstration incluant des graphiques d’effets avec barres d’erreur.

Cousineau présente par la suite l’analyse de proportions utilisant la transformation arcsin qu’il nomme ANOPA. Cette analyse peut être utilisée dans plusieurs types de devis, notamment à groupes indépendants et à mesures répétées. Considérant qu’une proportion amalgame deux mesures, le nombre de succès et le nombre total d’essais , soit . Toutefois, comme la proportion n’a pas des caractéristiques échantillonnales désirables, une transformation arcsin est appliquée qui permet de diminuer le risque d’asymétrie. La distribution sera normale et s’étendra de 0 à 1.57 () et la variance sera constante, contrairement à *p* ce qui permet de respecter les postulats de normalité et d’homogénéité des variances de l’ANOVA. De la variance théorique, il est aussi possible d’obtenir une erreur type théorique. Comme à son habitude, le présentateur montre une interprétation visuelle à l’aide d’un graphique à barres d’erreur. Cousineau conclut sa présentation en soulignant que la philosophie de l’ANOVA, traditionnellement utilisée pour analyser des moyennes, peut être appliquée pour les proportions et les fréquences.

En guise de double séance, Laurencelle présente « Tester la différence entre deux proportions jumelées versus le test de McNemar » (avec Cousineau comme collaborateur). Laurencelle débute sa présentation en mentionnant que comparer deux proportions jumelées comporte trois défis. Le premier, on connaît parfaitement le comportement statistique d’une proportion basée sur son paramètre , une constante logée entre 0 et 1. Toutefois, les proportions habituellement étudiées en sciences humaines émanent de groupes de sources (ou répondants) diverses, chaque source ayant une valeur originale et pour la plupart du temps instable. Le deuxième, quant à vérifier si une deuxième proportion estimée sur un autre groupe diffère de la première, on doit y supposer des différents, ce qui entraîne des propriétés distributionnelles des proportions différentes, notamment de leur variance. Le troisième, comment compare-t-on des proportions corrélées?

Laurencelle affirme qu’il n’existe pas de méthode directe et exacte pour comparer deux proportions jumelées. La procédure habituellement invoquée est le célèbre test de McNemar. Parmi les paires de données 0 ou 1 disponibles, McNemar retient les sources qui sont passées de 0 à 1, et les qui ont fait le chemin contraire, et il teste l’équilibre entre ces deux sous-groupes par un test de type binomial à paramètre d’autres auteurs ont tenté d’améliorer cette procédure). Ainsi, le test de McNemar statue sur la tendance du changement d’option *chez ceux qui changent*, ignorant totalement les et qui ne changent pas, quel que soit leur nombre. Le test de McNemar est par définition exact, mais ne s’applique qu’indirectement et maladroitement à la comparaison de avec , selon le présentateur.

Laurencelle propose un test complet et vraisemblable sur le plan paramétrique de la différence . Il s’agit de

où la proportion est , la variance est et la corrélation est (ou la corrélation pour des données binaires). La chaîne Fisher-Zubin-Anscombe utilise la fonction trigonométrique pour exprimer une proportion (de distribution binomiale) en une variable de distribution presque normale, avec variance stabilisée (voire constante) et asymétrie réduite d’un facteur de 2. Le test proposé, calqué sur l’étalon du test normal de la différence entre deux moyennes jumelées, met aussi en jeu la corrélation entre les deux séries binaires comparées. Les propriétés de précision (respect du seuil ) et de puissance (atteinte facile de 1) de ce test sont distinctement meilleures que celles du test McNemar.

Laurencelle conclut que la même transformation et le même principe de calcul, basé sur cette variable, permettent à l’utilisateur d’importer les proportions dans l’ensemble des devis d’examen de l’analyse de variance classique, comme le montrera l’exposé de mon collègue Denis Cousineau.

**Conclusion**

Après chaque présentation, des discussions approfondies et stimulantes ont incité une réflexion plus poussée et permis de partager un échange considérable de connaissances. En ce sens, la résurrection du colloque MQSH a été un succès marqué de quelques débats enthousiastes qui ont mis en lumière des problèmes originaux, l'importante approche critique pour les étudier (plutôt que seulement technique), mais aussi l’originalité des solutions apportées pour les résoudre. Cette dynamique a parfaitement répondu à l'objectif initial de ce colloque. P.-O. Caron, chargé de l'organisation de cette édition du colloque MQSH, a généreusement offert ses services pour prendre en charge l'organisation de la prochaine édition.