Programme

MQSH2023

UNIVERSITÉ TÉLUQ

8 h 45 **ACCUEIL ET MOT DE BIENVENUE**

Mot de bienvenue

Louis Laurencelle, Université du Québec à Trois-Rivières

9 h **CONFÉRENCE D'OUVERTURE**

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

Sébastien Béland, Université de Montréal

La majorité des travaux sur la fidélité des scores ont été effectués sur des items à réponse continue (Raykov, Dimitrov et Asparouhov, 2010). Cette étude consiste à évaluer l'efficacité de sept coefficients populaires sur des items unidimensionnels à réponse dichotomique. Nous allons procéder, ici, en deux étapes. Dans un premier temps, une stratégie permettant d'estimer la valeur vraie de la fidélité des scores pour réponse dichotomique sera présentée et justifiée en vue de faire des analyses de simulation. Dans un deuxième temps, nous allons procéder à des comparaisons à l'aide de la simulation informatique. Les résultats montrent que i) la stratégie de Dimitrov (2003) basée sur le modèle de la théorie de la réponse aux items à deux paramètres, ii) l'oméga total de McDonald basé sur le modèle bifactoriel et iii) un coefficient découlant de l'analyse factorielle confirmatoire présentent les erreurs et les biais les plus faibles. Cela permet de recommander l'utilisation d'indices de fidélité basés sur un modèle psychométrique plutôt que sur la simple covariance/corrélation. Fait intéressant, les coefficients pour items à réponse continue semblent bien performer avec le type de réponse qui nous intéresse dans cette communication. Cela fait écho aux travaux d'autres chercheurs tels que Robitzsch (2020) et laisse croire que les modèles de mesure pour items continus sont plus robustes que plusieurs tendent à le croire en présence d'items à réponse discrète de type ordinale (dichotomique et polychotomique).

9 h 45 **CONFÉRENCE**

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

Félix Laliberté et Éric Lacourse, Université de Montréal

Cette présentation fait un survol des capacités de la nouvelle librairie StepMix et de ses applications en sciences sociales et du comportement. Un *package* Python suivant l'API de scikit-learn pour le *clustering* basé sur des modèles et la modélisation de mélanges généralisés (analyse de classes/profils latents) de données continues et catégorielles. StepMix gère les valeurs manquantes grâce à la méthode du maximum de vraisemblance complète (FIML) et fournit plusieurs méthodes d'estimation par maximisation de l'espérance (EM) par étapes basées sur la théorie de la pseudovraisemblance. Les fonctionnalités supplémentaires incluent la prise en charge des covariables et des résultats distaux, divers utilitaires de simulation et le *bootstrap* non paramétrique, qui permet l'inférence dans des contextes semi-supervisés et non supervisés.

10 h 30 **PAUSE**

11 h **CONFÉRENCE**

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

Eric Frenette, Christiane Trottier, Vicky Drapeau et Roxane Carrière, Université Laval

Cette communication vise à présenter les 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. Ce programme porte sur l'accompagnement d'enseignants et d'entraîneurs dans l'enseignement et le transfert de cinq habiletés de vie (implanté séquentiellement) auprès d'élèves sur une période de trois ans : fixation d'objectifs, concentration, saines habitudes alimentaires, comportements sécuritaires et récupération physique et mentale.

Composée d'un groupe témoin (n = 145) et expérimental (n = 148), l'évaluation de l'implantation du programme s'est effectuée à partir de questionnaires lors de cinq temps de mesure au cours des trois années. Bien que les qualités psychométriques soient considérées acceptables à chaque temps de mesure, divers défis sont ressortis : a) qualité des données sociodémographiques, b) présence de données manquantes et c) gestion d'une collecte pendant la pandémie COVID-19.

L'évaluation de l'implantation du programme n'ayant pas permis d'atteindre tous les objectifs souhaités, une analyse des processus méthodologiques et statistiques a été effectuée : échelonnement de la mesure, pouvoir discriminant des items et de la mesure (consistance interne, courbe d'information), positionnement des items/répondants, identification des sources d'erreur par l'utilisation de la théorie de la généralisabilité, etc. Lors de la présentation, les diverses limites identifiées affectant les comparaisons longitudinales seront discutées et des pistes de solutions seront proposées pour optimiser la prochaine évaluation du programme *Gagnant pour la vie* 2.0.

11 h 45 **DÎNER ET RÉSEAUTAGE**

13 h 15 **CONFÉRENCE**

Stratégie pour identifier des valeurs aberrantes dans des distributions asymétriques

André Achim, Université du Québec à Montréal

Des valeurs aberrantes sur une variable modifient souvent le coefficient d'asymétrie de ses mesures observées. Pour une distribution déjà asymétrique, il est difficile de déterminer si les dernières valeurs du côté étiré appartiennent ou non à la distribution. Une solution est proposée pour assurer la symétrie de la variable transformée avec un minimum d'influence de possibles valeurs aberrantes. Il s'agit de produire une transformation de type y=f(x+k) qui ramène deux valeurs témoins de la distribution observée (par exemple celles aux 5^{e} et 95^{e} centiles) à égale distance de la médiane. La fonction f de transformation peut être fixe (ex. : log10 pour une asymétrie positive) de sorte que la transformation ne dépende que de la constante appropriée k à ajouter aux données avant d'appliquer f. La fonction f peut aussi avoir un paramètre à optimiser, comme l'exposant optimal pour symétriser les deux valeurs témoins par rapport à la médiane étant donnée la constante k (exposant optimal > 1 pour une distribution étirée à gauche). Il n'est alors pas besoin de d'abord inverser une distribution étirée à gauche. En fixant le zéro de la variable transformée à sa médiane et son échelle telle que les valeurs témoins transformées soient à la distance de la médiane correspondant à la cote z des rangs utilisés, on peut traiter les valeurs aux extrémités de la distribution transformée comme des pseudocotes z pour décider si elles appartiennent ou non à la distribution.

14 h **CONFÉRENCE**

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

Pier-Olivier Caron, Université TÉLUQ

Le temps joue un rôle fondamental dans les processus de médiation tels que les modèles de médiation par panel à décalage croisé (CLPM) et d'autres modèles longitudinaux. Lorsque cela est impossible, les chercheurs praticiens peuvent opter pour un modèle de médiation transversale (CSM), même s'il est statistiquement douteux de découvrir des phénomènes longitudinaux à l'aide de modèles transversaux. Cependant, peu d'études ont été menées pour déterminer leurs propriétés statistiques et leur comparabilité. Par conséquent, à l'aide de simulations de Monte-Carlo, cet article évalue l'erreur de type I, la puissance, le biais et l'accord entre le modèle CLPM et le modèle CSM dans le cadre d'un véritable modèle CLPM. Les effets indirects (αβ), les effets directs (γ), les trajectoires autorégressives (τ) et les tailles d'échantillon (n) ont été modifiés, ce qui a donné lieu à 97 scénarios différents (485 au total) répliqués 5 000 fois. Les résultats montrent que tous les modèles ont le taux d'erreur de type I attendu, à l'exception d'une triple interaction de τ×γ×α (et dans une moindre mesure avec β). S'ils sont tous élevés, il y a une forte probabilité d'effets indirects parasites dans la CSM. Lorsque les trajectoires autorégressives sont très faibles, la CSM manque de puissance, comme on peut s'y attendre. Dans d'autres cas, la puissance de la CSM est faible (50 %) à excellente (98,3 %), en raison d'un τ élevé et d'un αβ élevé. Des recommandations sont proposées pour l'utilisation de la CSM afin de mettre en évidence les effets indirects longitudinaux.

14 h 45 **PAUSE**

15 h 15 **CONFÉRENCE**

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

Denis Cousineau, Université d'Ottawa, et Louis Laurencelle, Université du Québec à Trois-Rivières

Deux nouvelles méthodes d'analyse seront présentées : l'analyse des proportions utilisant la transformation arcsine (ANOPA) et l'analyse des fréquences de données (ANOFA). Ces deux techniques sont entièrement analogues aux ANOVA pour les moyennes de données continues. 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. Les deux techniques disposent d'une mesure de la grandeur (ou taille) d'effet, identique à l'êta carré et qui s'interprète de la même façon. Ces techniques permettent enfin d'illustrer les résultats par des graphiques d'effets avec barres d'erreur et d'effectuer une analyse de puissance statistique a priori. L'ANOPA est applicable aussi aux devis à mesures binaires répétées. La présentation couvre les grandes lignes de ces nouvelles techniques, lesquelles permettent à n'importe qui familier de l'ANOVA de transférer ses habiletés et les appliquer à d'autres types de mesures. Les lacunes des analyses logistiques parfois utilisées sur ces types de données seront survolées.

16 h **FIN DE LA JOURNÉE**

Mot de clôture

Pier-Olivier Caron, Université TÉLUQ