La modélisation des scores issus des enquêtes en santé à l'aide de la modélisation Rasch : promesses et défis*

Eric Dionne
Université d'Ottawa

Julie Grondin
UQAR, Campus de Lévis

*Recherche subventionnée par le Programme de soutien à la recherche du Consortium National de Formation en Santé (CNFS)

PLRS, Gatineau (Québec), Juin 2011

Plan de la présentation

- Contexte de la recherche
- Objectifs
- Modèle de mesure
- Données
- Bilan de l’analyse
- Discussion
- Conclusion

Contexte de la recherche

- Découle d'une réflexion en trois volets :
  - Constats sur la santé
  - Types de cueillettes de données
  - Choix de la méthode d'analyse

Recherche subventionnée par le Programme de soutien à la recherche
du Consortium National de Formation en Santé (CNFS)

Constats sur la santé

- La promotion de la santé et des saines habitudes de vie est au coeur des préoccupations actuelles.
- Selon certaines études, les francophones semblaient moins préoccupés par les questions de santé que les anglophones.

Recherche subventionnée par le Programme de soutien à la recherche
du Consortium National de Formation en Santé (CNFS)

Types de cueillettes de données

- Cueillette de données primaires
  - Choix instrument, échantillon, matériel, validation, etc.
  - Permet de contrôler l'origine et la qualité des données.
- Analyse de données secondaires
  - Permet d'accéder à des données, sans l'investissement en temps et en argent.

Choix de la méthode d'analyse

- Méthodes unidimensionnelles
  - Analyse du tout, sans tenir compte des parties.
  - Analyse des parties, sans tenir compte du tout.
- Méthodes multidimensionnelles
  - Permet d'analyser le tout et les parties, de même que les interrelations entre ces dernières simultanément.

Objectifs de la recherche

1. Maximiser la portée de données existantes sur la santé afin de comparer les populations francophones et anglophones :

   - Est-ce que le constat est toujours d'actualité?
   - Est-ce que le constat s'applique hors Québec?
   - Quelles variables pourraient expliquer cette différence?

Objectifs de la recherche

2. Appliquer, dans une perspective exploratoire, un modèle de mesure multidimensionnel issu de la famille des modèles de Rasch :

- Effectuer une analyse FDI
- Rendre compte de la qualité de la mesure
- Répondre au premier objectif

Modèle de mesure

- Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM)
- Permet de modéliser des données pouvant contenir plusieurs attributs/dimensions
- Conserve la structure des données

Modèle de mesure

- Calibre les données simultanément
- Utilise les corrélations entre les attributs/dimensions pour augmenter la précision de la mesure
- Permet de rendre compte de la qualité de la mesure obtenue

MRCMLM

$$P(X_{ijk} = 1; \xi | \theta_i) = \frac{\exp\left(b'_{jk}\theta_i + a'_{jk}\xi\right)}{\sum_{u=1}^{K_j} \exp\left(b'_{jk}\theta_i + a'_{jk}\xi\right)}$$

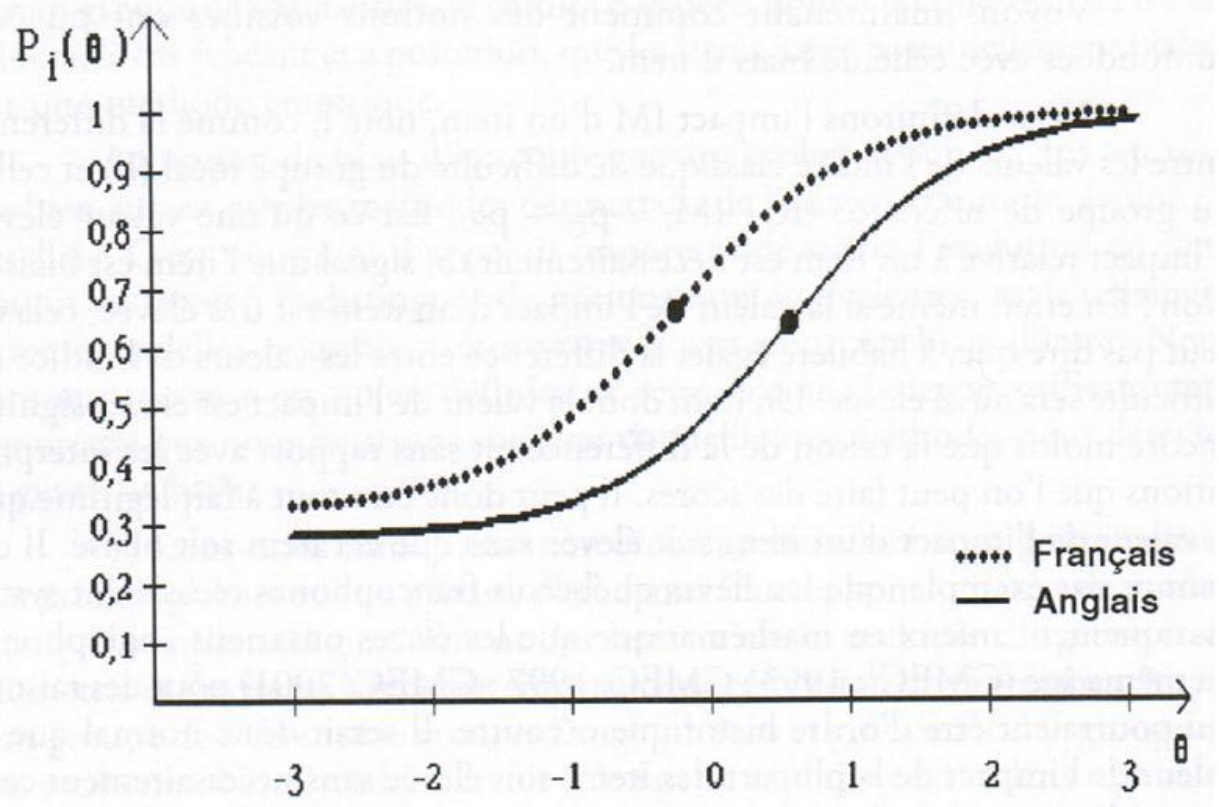
- $P(X_{ijk})$: probabilité qu'un sujet i, répondent à l'item j, dans la catégorie de réponse k
- θ_i (i = 1 à D): position du sujet par rapport aux différents traits latents modélisés
- ξ: vecteur de la position des paramètres d'items (difficulté de l'item et points d'ancrages)
- a_{jk}: vecteur associé à la catégorie k de l'item j qui modélise les relations linéaires entre les éléments du vecteur ξ
- b_{jk}: vecteur associé à la catégorie k de l'item j qui modélise le score obtenus pour les différents traits latents
- K_j: nombre de catégories de réponses associé à l'item j

Données

- Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ) de Statistique Canada
- Année retenue: an 3 (sur 6 ans)
- 331 variables, 31 963 sujets
- Variables : les problèmes de santé, l’activité physique, l’accès aux soins, etc.

Bilan de l'analyse
(Idéalement...)

- Graphique FDI sur l'adoption de pratiques dangeureuses au regard de la santé chez les francophones et les anglophones...



(Bertrand et Blais, 2004, p.284)

Bilan de l'analyse
(En réalité...)

- Les analyses sont toujours en cours.
- Plusieurs problèmes (défis…) rencontrés:
  - **Différences entre les variables disponibles et celles dont nous avons besoin** *(p. ex.: les données linguistiques)*
  - **Analyse difficile à opérationnaliser** *(nombreuses variables, différentes sources / instruments / scores, en plus des recodages...)*
  - ➢ **Tailles d'échantillon difficiles à modéliser** *(une analyse multidimensionnelle est plus riche, mais plus longue à converger; les tailles d'échantillon sont trop grandes pour ce type d'analyse; les sous-échantillons sont trop petits pour les objectifs de la recherche...)*
  - ➢ **Croisements difficiles** *(p. ex.: soit les données linguistiques sont disponibles, soit les données sur la santé...)*

Discussion

- Il est parfois difficile d'opérationnaliser une analyse de données secondaires:
  - Les données ne répondent pas aux présentes questions de recherche (type de variable, format, précision).
  - Tailles d'échantillon (peuvent être un avantage, ou être très lourdes à gérer...)
  - Un sous-échantillon peut être trop petit...
  - Il est difficile de s'approprier le construit à l'origine des données.
  - Problèmes de mesure : qualité des données initiales ?

Conclusion

- (Idéalement) On aurait aimé dire que:
  - Les modèles de Rasch permettent une anayse adéquate de ce type de données en rendant compte de la qualité de la mesure obtenue.

ET / OU

  - Le modèle multidimensionnel est un outil à explorer davantage pour l'analyse de données secondaires.

Coordonnées des auteurs

eric.dionne@uottawa.ca

julie_grondin@uqar.qc.ca