Analyses multiniveaux, une introduction

Cours SOL 6210

© Claire Durand, 2024

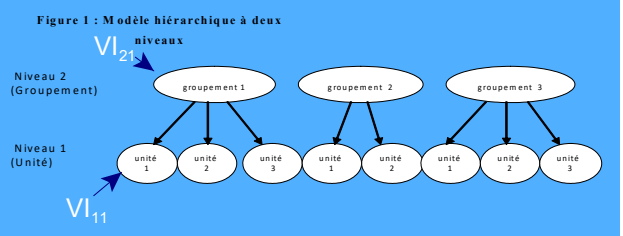
1

Quand utilise-t-on l'analyse multiniveau?

- Il n'y a pas indépendance entre les données parce que...
  - Certains individus appartiennent au même groupe:
    - À la même équipe de travail
    - À la même classe
    - À la même famille, au même quartier
  - Certaines informations ont été recueillies dans le temps (mesures répétées)
    - Mêmes informations recueillies auprès d'un individu à plusieurs moments
  - Plusieurs informations ont été recueillies au même moment pour un même individu.
    - Ces informations sont interdépendantes.
  - À chaque moment, plusieurs informations ont été recueillies:
    - Le temps au niveau 2, les informations au niveau 1.

2

Illustration...

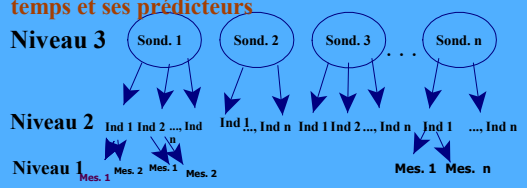


- La VD est TOUJOURS au niveau 1.
- Les VI peuvent être aux deux niveaux.

3

La combinaison de fichiers de données

- Au niveau 3: caractéristiques du sondage (incl. la formulation des questions) et l'évolution dans le temps et ses prédicteurs



- Au niveau 2: Les individus et leurs caractéristiques (âge, sexe, éducation, région)
- Au niveau 1: La confiance et ses objets (police, religion, syndicats, écoles,)

4

A noter...

- Il peut y avoir plus de deux niveaux (pour le moment quatre au maximum - disponible dans HLM et MLWin).
- La variable dépendante est toujours au niveau 1.
- La variable dépendante peut être de plusieurs types (linéaire, logistique, multinomiale, etc).
- Le niveau supérieur doit être un échantillon (sinon, on devrait utiliser une analyse de variance).
  - Il faut donc qu'il y ait un nombre d'unités minimum au niveau supérieur (minimum 40, de préférence) puisque l'on fait une régression à chaque niveau.

5

Le type d'effet d'une variable sur une autre peut être tenu en compte

- Cet effet
  - Peut être le même quelque soit le groupe: x influence y de la même manière quel que soit le groupe auquel la personne appartient; les moyennes des groupes sont similaires: équivaut à régression ordinaire.
  - Peut se produire sur les intercepts: x influence y de la même manière mais la moyenne des groupes est différente (modèle à intercepts aléatoires): par défaut.
  - Peut se produire également sur les pentes: x n'influence pas y de la même manière dans les divers groupes (modèle à intercepts et à pentes aléatoires) → ce qui veut dire qu'il y a hétéroscédasticité.

6

Avantages: information produite

- On peut savoir
  - Quelle proportion de la variance de la v.d (y) peut être attribuée aux regroupements (niveau 2) et quelle proportion au niveau 1 (individu, mesure, etc.).
    - Exemple: palmarès des écoles: quelle part de la variation dans les résultats des élèves peut être attribuée à l'école et quelle part peut être attribuée aux différences individuelles entre les élèves?
  - Quelle proportion de la variation entre les individus et entre les groupes peut être attribuée à des facteurs agissant au niveau des individus ou au niveau des groupes.
    - Exemple: quelle proportion de la variation dans les résultats scolaires des élèves peut être attribuée à des facteurs individuels (préparation des élèves, niveau de scolarité de leurs parents) et à des facteurs au niveau de l'école (expérience de leurs enseignants, approche pédagogique, climat de l'école, niveau de scolarité moyen des parents de l'école).

7

Avantages : données manquantes

- Données manquantes non problématiques avec l'analyse multiniveaux: on utilise toute l'information disponible:
  - Avantage évident si on compare aux analyses qui demandent d'avoir des informations à tous les temps de mesure: mesures répétées, séries chronologiques, analyses d'items.

8