

**L'analyse multi-niveaux
avec MlwiN**

Alain Marchand
Groupe de recherche sur les aspects
sociaux de la santé et de la prévention
(GRASP) et
École de Relations industrielles
Université de Montréal

© Alain Marchand 2000

1

Introduction

La recherche en sociologie implique souvent
la présence de données hiérarchiques:

- Travailleurs nichés dans des usines
- Étudiants nichés dans des écoles
- Patients nichés dans des hôpitaux
- Longitudinale: même mesure répétée dans le temps
- Plus complexe: travailleurs nichés dans les usines,
usines nichées dans industries

2

- Dans ce type de recherche il y a des variables qui sont définies au niveau des individus (travailleurs) et au niveau des groupes (usines)

- La question de recherche se pose souvent ainsi: Comment les variables individuelles et de groupes influencent-elles une variable dépendante mesurée au niveau des individus.

3

Les approches antérieures

Modèles linéaires traditionnels basés sur les moindres carrés ordinaires avec deux approches:

- Imputation des variables de groupes aux individus
- Agrégation des variables individuelles au niveau des groupes

4

Ces deux approches sont insatisfaisantes

Approche imputation

- Violation de l'assomption d'indépendance.
- Covariation entre les observations partageant un même contexte
- MCO produit estimé instable; erreurs types biaisées (ie. trop grandes)

5

Approche agrégation

- Perte de l'information concernant la variation intra groupe (80-92%)
- Produit généralement des corrélations plus fortes : Effet Robinson
- On ne peut plus conclure sur les individus sans commettre l'erreur écologique. Les variables individuelles ne sont plus de niveau individuel

Solution: les modèles multi-niveaux

6

Bref historique

- La formalisation de ces modèles est connue depuis au moins les années 60. (Elston et Grizzle, 1962: Biométrie).

- Ils apparaissent dans plusieurs disciplines:

Sociologie: modèles linéaires multi-niveaux ou hiérarchiques
Biométrie: modèles à effets mixtes; modèles à effets aléatoires
Économétrie: modèles de régression à coefficients aléatoires

- C'est seulement depuis une vingtaine d'années qu'il est techniquement possible d'estimer ces modèles

7

Les logiciels disponibles

- HLM
- MLwiN (mon préféré)
- Lisrel
- SPSS
- R
- Stata xtreg
- BMDP (BMDP3V)
- SAS (MIXED)
- VARCL
- MIXREG, MIXOR, MIXNO, MIXPREG (gratuit)
- EGRET

8

Modèle pour une variable dépendante normalement distribuée

9

Modèle de composition de la variance

Modèle à deux niveaux avec $j=1 \dots K$ niveau 2 et $i=1 \dots n_j$ niveau 1

Niveau 1 : $Y_{ij} = \beta_{0j} + \epsilon_{ij}$

β_{0j} = moyenne dans le groupe j
 ϵ_{ij} = résidu pour individu i du groupe j

Niveau 2 : $\beta_{0j} = \gamma_{00} + \mu_{0j}$

γ_{00} = moyenne pour l'ensemble des groupes
 μ_{0j} = résidu pour le groupe j

Alors: $Y_{ij} = \gamma_{00} + \mu_{0j} + \epsilon_{ij}$

10

Postulat:

μ_{0j} et ϵ_{ij} sont non-corrélées et distribuées normalement avec comme moyenne 0 et des variances σ^2_{μ} , σ^2_{ϵ} estimées par les données.

Corrélation intra-classe

$$\rho_i = \sigma^2_{\mu} / (\sigma^2_{\mu} + \sigma^2_{\epsilon})$$

ρ_i = proportion (%) de la variabilité de la variable dépendante Y_{ij} qui est entre les groupes.

11

Avec variables indépendantes

Niveau 1 : X_{pji} ($p=1 \dots, P$) Niveau 2 : Z_{qj} ($q=1 \dots, Q$)

Niveau 1 : $Y_{ij} = \beta_{0j} + \beta_{pj} X_{pji} + \epsilon_{ij}$

Niveau 2 : $\beta_{0j} = \gamma_{00} + \gamma_{0q} Z_{qj} + \mu_{0j}$

Si on pose que β_{0j} varie entre les groupes et que les pentes de niveau 1 β_{pj} sont constantes, soit $\beta_{pj} = \gamma_{p0}$

Alors : $Y_{ij} = \gamma_{00} + \gamma_{p0} X_{pji} + \gamma_{0q} Z_{qj} + (\mu_{0j} + \epsilon_{ij})$

X_{pji} expliquent variation intra et inter groupes
 Z_{qj} expliquent variation inter groupes

12

Modèle à coefficients aléatoires

Pentes aléatoires au niveau 2

$$Y_{ij} = \gamma_{00} + \gamma_{p0} X_{prij} + \gamma_{0q} Z_{qj} + (\mu_{pj} X_{prij} + \mu_{0j} + \epsilon_{ij})$$

Pentes aléatoires niveau 2 avec interaction

$$Y_{ij} = \gamma_{00} + \gamma_{p0} X_{prij} + \gamma_{0q} Z_{qj} + \gamma_{0q} Z_{qj} X_{prij} + (\mu_{pj} X_{prij} + \mu_{0j} + \epsilon_{ij})$$

Pentes aléatoires niveau 1 et 2

$$Y_{ij} = \gamma_{00} + \gamma_{p0} X_{prij} + \gamma_{0q} Z_{qj} + \gamma_{0q} Z_{qj} X_{prij} + (\mu_{pj} X_{prij} + \epsilon_{ij} X_{prij} + \mu_{0j} + \epsilon_{ij})$$

Pentes aléatoires niveau 1 seulement

$$Y_{ij} = \gamma_{00} + \gamma_{p0} X_{prij} + \gamma_{0q} Z_{qj} + (\epsilon_{ij} X_{prij} + \mu_{0j} + \epsilon_{ij})$$

13

- Pour un modèle à coefficients aléatoires la variation totale du niveau 1 ou 2 est une fonction quadratique. Ex: 1 pente
- niveau 2: $\sigma_{\mu 0}^2 x_0^2 + 2 \sigma_{\mu 01} x_0 x_{1ij} + \sigma_{\mu 1}^2 x_{1ij}^2$
- niveau 1: $\sigma_{\epsilon 0}^2 x_0^2 + 2 \sigma_{\epsilon 01} x_0 x_{1ij} + \sigma_{\epsilon 1}^2 x_{1ij}^2$

14

Méthode d'estimation

Moindres carrés itératifs généralisés (IGLS)

- La fonction de vraisemblance est vue comme étant composée des quantités aléatoires et des coefficients de régression.
- Ces quantités sont posées comme étant connues et les valeurs de départ de la procédure itérative sont ceux estimés par les moindres carrés ordinaires (MCO).
- Les estimés obtenus par IGLS sont biaisés lorsque la taille de l'échantillon est petite.
- À la convergence, les estimés sont ceux du maximum de vraisemblance.

15

Moindres carrés itératifs généralisés restreints (RIGLS)

- RIGLS est analogue à IGLS mais tient compte des variations échantillonnales des quantités aléatoires et des coefficients de régression.
- Dans un premier temps, la partie fixe du modèle estimé est soustraite et dans un deuxième temps, la vraisemblance de la partie aléatoire est maximisée.
- Lorsque la taille de l'échantillon est grande, RIGLS et IGLS donnent des résultats identiques.
- À la convergence, les estimés sont ceux du maximum de vraisemblance restreint.

16

Autres méthodes
(simulation)

- Bayésienne (Méthode Monte Carlo de la chaîne de Markov)
- Bootstrap

17

Les tests d'hypothèses

- Plusieurs tests d'hypothèses peuvent être calculés mais certains dépendent de la méthode d'estimation.

18

IGLS

- Test du rapport de vraisemblance pour l'évaluation de l'ensemble du modèle (coefficients aléatoires et de régression) ou pour comparer deux modèles.
- Test de T ou Z pour un seul coefficient de régression.
- Pour la partie aléatoire, test du rapport de vraisemblance. Pour un test pourtant seulement sur σ^2_ϵ , test de T, Z ou χ^2
- Dans un modèle à deux niveaux, on recommande pour les variables indépendantes de niveau 2 d'utiliser le test de T avec comme degrés de liberté $J-S_q-1$, ou J =nombre de groupes et S_q =Nombre de variables de niveau 2. Si J est très grand, $T \rightarrow Z$

19

RIGLS

- Les valeurs de déviance (-2 log de vraisemblance) ne sont pas directement utilisable pour calculer un test d'hypothèse sur l'ensemble du modèle (partie fixe et aléatoire) car RIGLS maximise seulement la partie aléatoire.
- Test de Wald pour l'évaluation de l'ensemble coefficients de régression.
- Test de T ou Z pour un seul coefficient de régression.
- Pour la partie aléatoire, test du rapport de vraisemblance ou test de Wald. Pour un test pourtant seulement sur σ^2_ϵ , test de T, Z ou χ^2

20

Exemple

Simard et Marchand (1997). Ergonomics, 40, 2: 172-188.

Effets des facteurs micro et macro organisationnels sur le niveau de prudence des équipes de travail.

Modèle de composition de la variance

$$\text{prudence}_{ij} = \gamma_{00} + \gamma_{p0} \text{micro}_{pj} + \gamma_{0q} \text{macro}_{qj} + (\mu_{0j} + \epsilon_{ij})$$

n_2 = 97 établissements manufacturiers
 n_1 =1061 équipes de travail

21

Autres applications possibles de l'analyse multi-niveaux

- Dépendante nominale, multinominale et ordonnée
- Séries chronologiques
- Composante principale
- Factorielle exploratoire et confirmatoire
- Cheminement de la causalité
- Régression de survie
- Meta analyse
- Longitudinale pour VD dichotomique
- Multivarié VD dichotomiques
- Multivarié mixte: multinominale et ordonnée avec une ou plusieurs variables continues
- Modèles Poisson, log-log, probit

28

Ressources sur l'internet

Pour l'Amérique du nord:

<http://www.medent.umontreal.ca/multilevel/>

Pour l'Europe:

<http://www.ioe.ac.uk/multilevel/>

Liste électronique de discussions:

mailbase@mailbase.ac.uk

Corps du message: join multilevel prénom nom.

29