

the ethical problems raised by attempts to probe more deeply into people's emotional lives.

Initially, my research with Brian Jackson was an attempt to understand and locate our own experience in the broader context of the lives of others in similar circumstances. However, my recent (holiday) reading of working-class autobiographies (Caroline Steedman and Lorna Sage) now encourages me to reverse this process, so that instead of autobiography being the starting point of research, the end-point of my career in research may be a further - and hopefully more fully informed - excursion into autobiography.

Ray Pahl and Liz Spencer, Friends and Personal Communities: Methodological Issues

Our paper is based on research we have recently carried out on friendship and friend-like relationships. We wanted to explore the actual set of social relationships in which people are embedded and to liberate respondents from received categories. The paper describes an innovative method for generating and mapping personal communities, and discusses the development of some new typologies.

End Note

Martyn Hammersley, Pasts, Presents and Futures for Qualitative Research

Views of the future of qualitative research are even more obviously social constructions than are accounts of its past. At the same time, any history of qualitative research hints at what its future might be. So, in a situation where exponents of qualitative research seem to have very different images of its past, there will also be divergent ideas about its future; both about what is likely to happen, and about what should happen.

In this paper I look at some recent versions of the history of qualitative research, including that underpinning this conference, and consider what implications they carry for its future. The divergences amongst these histories raise questions about the nature of qualitative research, and about its relationship to the rest of sociology - in particular, to quantitative work and social theory.

ON-GOING RESEARCH RECHERCHE EN COURS

ANALYSE MULTINIVEAUX

par

Tom Snijders

(ICS, Université de Groningen,

t.a.b.snijders@ppsw.rug.nl)

Abstract: Multi-Level Analysis. This article is based on a one-day course that the author has presented in Lille and in Paris to French-speaking sociologists who are just now becoming familiar with the method. **Multi-Level Analysis, Statistical Modelling, Statistical Education.**

Résumé. Ce article est basé sur un cours d'un journée que l'auteur a donné à Lille et à Paris pour des sociologues francophones qui commencent tout juste à se familiariser avec cette méthode. **Analyse multiniveau, Modélisation statistique, Éducation statistique.**

Bien des recherches en sciences sociales portent sur des entités de niveaux différents; plus spécifiquement, sur les influences exercées par le contexte (niveau supérieur) sur l'individu (niveau inférieur).

Les contextes peuvent être de nature, par exemple:

- spatiale: quartiers, communautés, états, ...
- organisationnelle: entreprises, branches d'entreprises, écoles, classes scolaires, ...
- sociale: groupes de référence, réseaux sociaux,
- temporelle: âges, moments d'enquête, ...

Dans l'analyse des données portant sur l'influence des contextes sur les individus, il faut tenir compte de la diversité des genres d'entités analysées.

Dans une structure à deux niveaux (disons, individus et contextes), on distingue l'ensemble des individus et l'ensemble des contextes; les individus étant insérés (*nested*) dans les contextes.

Par conséquent, il n'y a pas une simple matrice rectangulaire de données, mais les données ont une structure emboîtée.

Parfois on parlera de groupes comme synonyme de contextes. Ne pas tenir compte de cette diversité des entités analysées peut mener à des erreurs d'agrégation:

erreur écologique -- interprétation des différences entre les groupes comme des différences entre les individus; par exemple, si les régions à majorité protestante montrent des taux élevés de suicide comparés aux régions à majorité catholique, on ne peut pas conclure que les individus protestants se suicident plus que les individus catholiques (voir van Poppel et Day dans *American Sociological Review*, 1996, 61).

erreur atomistique -- interprétation des différences entre les individus comme des différences entre les groupes; par exemple, si, dans chaque pays, les individus qui ont une scolarisation prolongée ont tendance à gagner des salaires plus importants, on ne peut pas conclure que les pays à scolarisation moyenne prolongée sont plus riches. (voir, par exemple, Alker: A typology of ecological fallacies, dans M. Dogan et S. Rokkan, *Quantitative ecological analysis in the social sciences*, 1969, Cambridge, MA: MIT Press).

Les outils standards de l'analyse statistique (analyse de régression, analyse de variance, etc.) ne sont pas adaptés à ce type de données: ils sont faits pour un seul genre d'entité d'analyse, pour des inférences concernant une seule population; ils sont basés sur la supposition que les résidus (parties inexplicables) sont indépendants.

L'analyse multinationale est un ensemble de techniques statistiques conçues pour l'analyse de données:

- de structure emboîtée, donc de plusieurs niveaux ordonnés de nature hiérarchique, chaque niveau correspondant à un type d'entité d'analyse;
- utilisant des modèles statistiques à plusieurs sources de variance (de variation inexplicables).

Les inférences statistiques dans l'analyse multinationale portent sur plusieurs ensembles d'entités d'analyse, correspondant aux différentes sources de variation.

Exemples de niveaux emboîtés

Individus dans groupes:

- élèves dans des écoles
- employés dans des entreprises
- habitants dans des communautés
- patients de médecins
- enfants dans familles
- etc.

Groupes dans groupes:

- classes dans des écoles
- branches dans des entreprises
- thérapeutes dans des cliniques
- etc.

Données longitudinales ou multivariées:

- séries d'observations sur chacun des individus.

Meta-analyse:

- répondants dans des enquêtes
- enquêtes dans des groupes de recherche.

Trois ou plusieurs niveaux:

- élèves dans des classes dans des écoles
- employés dans des branches dans des entreprises
- patients de médecins dans des cliniques
- etc.

Extensions:

- relations entre individus dans des groupes
- explication des variances au lieu de l'explication des moyennes: modélisation de l'hétérogénéité des variances (*hétéroscédasticité*).

Par convention, on appelle niveau inférieur, ou niveau premier, le niveau le plus détaillé.

L'idée de base de l'analyse multinationale:

- les relations entre les variables individuelles peuvent différer selon les contextes;
- ces différences entre les contextes peuvent elles-mêmes être expliquées par des variables contextuelles.

Par conséquent, on pourra utiliser dans l'analyse des variables définies à chaque niveau (variables individuelles ainsi que contextuelles).

L'analyse multinationale comporte surtout des variantes de l'analyse de régression ou de l'analyse de variance, conçues pour les données à structure emboîtée, où la variable dépendante est une variable du niveau inférieur: le modèle linéaire hiérarchique (*hierarchical linear model*).

Exemple: extrait de la thèse de Peter Mühlau, *The governance of the employment relation* (2000, ICS, Université de Groningen).

Données: caractéristiques des emplois des 3,735 employés dans 46 usines au Japon (Lincoln et Kalleberg, 1990).

Cet exemple porte sur les effets de différentes méthodes de gestion du personnel sur l'investissement des employés dans l'organisation.

Variables explicatives:

- niveau individuel -- variables de contrôle: âge, formation, ancienneté dans l'entreprise; variable de recherche: orientation coopérative du supérieur hiérarchique immédiat, encouragement du travail en groupe et objectivité dans l'évaluation du travail;
- niveau usine -- existence d'un marché du travail interne, réglementation des contrats et promotions.

Variable dépendante:

- investissement des employés dans l'organisation.

Les populations impliquées: population d'usines, population d'employés. L'inférence statistique vise ces deux populations.

L'analyse multinationale est basée sur des modèles aux coefficients aléatoires, ce qui veut dire que l'inférence ne vise pas ces 46 usines particulières, mais plutôt la population des usines dont les 46 seraient considérées comme un échantillon.

Comment analyser un tel ensemble de données groupées (donc, ici, comment analyser les effets des méthodes de gestion du personnel sur l'investissement des employés à l'égard de l'organisation)?

Désagrégation:

Oubliez les groupes, faites une analyse sur le niveau des individus, désagréguez les variables des groupes jusqu'au niveau des individus: régression selon le principe des moindres carrés ordinaires (MCO). Désavantages: fausse supposition d'indépendance des résidus dans les groupes; erreurs-types et tests incorrects; généralement, ceci implique qu'on fait comme s'il y avait plus d'information qu'il n'y en a en vérité.

Agrégation:

Agrégez toutes les variables à leur moyennes par groupe: MCO au niveau des groupes.

Désavantages: perte d'information; les relations à l'intérieur des groupes sont perdues de vue; on ne tient pas compte des tailles différentes des groupes.

Analyse de covariance:

Comme régression sur le niveau individuel, mais utilisant les groupes comme facteur.

Désavantages: l'inférence vise ces groupes particuliers et pas la population des groupes (modèle aux effets fixes au lieu d'aléatoires); facteur à 46 niveaux, ce qui est bien élevé.

Procédure à deux étapes:

D'abord faites une régression MCO dans chaque groupe, ensuite utilisez les coefficients estimés pour chaque groupe comme variable dépendante dans une analyse inter-groupes. Ceci est une forme plus subtile d'agrégation.

Désavantages: les coefficients par groupe sont très peu fiables (régression à peu de cas); on ne tient pas compte des tailles différentes des groupes.

Mieux:

Utilisez le *Modèle Linéaire Hiérarchique* (MLH) avec différences aléatoires entre les employés et différences aléatoires entre les usines. Dans cette approche, les rôles des employés ainsi que des usines en tant que sources de variations inexpliquées sont exprimés dans le modèle statistique.

Points fondamentaux du paradigme multinationale:

- L'effet de X (orientations coopératives du supérieur) sur Y peut être différent selon les groupes (coefficients aléatoires); ces différences pourraient être expliquées par des caractéristiques des groupes (*interactions entre-niveaux*, voir ci-dessous).

- L'effet entre-groupes de X sur Y pourrait être différent de l'effet intra-groupe (*utilisez la valeur moyenne de X par groupe comme variable explicative*).

- L'effet intra-groupe est la différence attendue de la valeur de Y entre deux individus dans le même groupe qui ont une différence de taille 1 pour X.

- L'effet entre-groupes est la différence attendue de la moyenne par groupe pour Y entre deux groupes qui ont une différence de taille 1 dans leur moyenne pour X.

PHASE 1: DECOMPOSITION DE LA VARIANCE

Le premier modèle pour l'investissement des employés dans l'organisation incorpore seulement les effets des variables de contrôle (âge, formation, ancienneté dans l'entreprise), et les effets aléatoires des employés et des usines.

Variance niveau usines: $\tau^2 = 0,63$

Variance niveau employés: $\sigma^2 = 4,65$.

Donc, la proportion de la variance totale au niveau des usines est de:

$$0,63 / (0,63 + 4,65) = 12\%$$

(coefficient résiduel intra-classe).

Ceci montre que les usines sont responsables pour une importante partie de la variation dans l'engagement des employés à l'organisation.

ELABORATION DES DEUX POINTS DU PARADIGME MULTI-NIVEAUX

1. Différences aléatoires entre les groupes des valeurs moyennes de Y et des effets de X sur Y:

$$Y = \beta_{0j} + \beta_{1j} X_{ij} + R_{ij}$$

où

i = individu

j = groupe

β_{0j} = intercept dépendant du groupe

β_{1j} = coefficient de régression dépendant du groupe (pente).

Les effets spécifiques aux groupes sont caractérisés par:

-- leurs espérances (effets fixes),

-- leurs variances (variance de l'intercept, variance de la pente),

-- et leur covariance (covariance intercept - pente).

Le fait que le but des inférences est la population des groupes et non les groupes individuels s'exprime dans le modèle mathématique de la manière suivante: ce ne sont pas les coefficients individuels β_{0j} et β_{1j} qui sont les paramètres du modèle, mais plutôt leur distribution: espérances, variances, covariance. (On fait la supposition d'une loi normale multivariée.)

S'il y a plusieurs variables au niveau de l'individu, quelques unes pourraient n'avoir que l'effet fixe (variance de la pente = 0; effet égal dans tous les groupes), d'autres pourraient avoir une effet fixe et un effet aléatoire.

Les variables au niveau du groupe auront, normalement, seulement un effet fixe.

2. Décomposer la variable X (aux valeurs x_{ij}) en les deux variables suivantes:

\bar{X} moyenne pour X par groupe,

et

$X - \bar{X}$ déviation intra-groupe.

et inclure tous les deux dans le modèle (la moyenne de groupe à effet fixe seulement). Le coefficient de régression intra-groupe est l'effet fixe de la variable déviation intra-groupe; le coefficient de régression entre-groupes est l'effet fixe de la moyenne de groupe.

Ceci implique, pour chaque variable explicative du niveau 1, la nécessité de décider les choix suivants:

1. Seulement un effet fixe, ou bien fixe et aléatoire?
2. Coefficients de régression intra-groupes égaux ou inégaux?

PHASE 2: MODELISATION AU NIVEAU INFÉRIEUR

On introduit l'effet de la variable individuelle:

X = orientation coopérative du supérieur.

Régression MLH de l'engagement à l'organisation sur

X = orientation du supérieur.

Moyenne X par usine
effet fixe 0.11 (e.t. 0.06)

Déviation X intra-groupe
effet fixe 0.14 (e.t. 0.03)
variance pente 0.018

variance intercept 0.58
covariance intercept-pente 0.062

variance résiduelle (niveau 1) 4.25

Effets intra-groupe et entre-groupes:
 ceux-ci ne sont pas significativement différents -- l'effet de l'orientation du supérieur n'apparaît pas, en réalité, comme un effet au niveau de l'usine.

Pour la reste de l'analyse, retenir la variable originale X (sans décomposition en parties entre-groupes et intra-groupe).

effet fixe X X 0.13 (écart type 0.03)
 variance pente X 0.018

variance intercept 0.59
 covariance intercept-pente 0.065

variance résiduelle (niveau 1) 4.25

Interprétation variance intercept:
 l'intercept est la valeur de la droite de régression pour X = 0; ces valeurs varient entre les groupes avec un écart-type de $(0.59)^{1/2} = 0.77$.

Interprétation effet X:
 l'effet moyenne est 0.13 et l'écart-type est $(0.018)^{1/2} = 0.13$; dans 95% des groupes l'effet par groupe β_{ij} a des valeurs s'écartant de moins de 2 écarts types de la moyenne, donc entre:
 0.13 - 0.26 = - 0.13 et 0.13 + 0.26 = 0.39.

PHASE 3: MODELISATION AU NIVEAU SUPERIEUR

Explication des effets de groupe β_{0j} et β_{1j} avec des variables définies au niveau du groupe (*régression latente*): ici, existence de marchés du travail internes et réglementation des contrats et promotions.

Partie fixe
 Individu 0.11 (0,03)
 Orientation du supérieur

Usine 0.29 (0,12)
 Réglementation 0.39 (0,11)
 Marchés internes du travail

Interaction individu x usine

Or. sup. x réglementation Marchés internes x réglm. 0,09 (0,02)
 0,26 (0,11)

Partie aléatoire
 niveau 2 0,14
 variance intercept 0,000
 variance pente 0,000
 covariance intercept - pente
 niveau 1 4,28
 variance résiduelle

Pour qui est intéressé dans les formules:

individu = entité au niveau 1 = i,
 groupe = entité au niveau 2 = j.

Régression au niveau 1 (dans groupe j)

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + R_{ij}$$

Régressions latentes au niveau 2 entre groupes

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + U_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Z_j + U_{1j}$$

La substitution donne:

$$Y_{ij} = \gamma_{00} + \gamma_{01} Z_j + U_{0j} + (\gamma_{10} + \gamma_{11} Z_j + U_{1j}) X_{ij} + R_{ij}$$

$$= \gamma_{00} + \gamma_{01} Z_j + \gamma_{10} X_{ij} + \gamma_{11} Z_j X_{ij} \quad (\text{partie fixe})$$

$$+ U_{0j} + U_{1j} X_{ij} + R_{ij} \quad (\text{partie aléatoire})$$

Paramètres de la partie fixe: coefficients de régression γ

Paramètres de la partie aléatoire:

variances et covariances U: $\tau_0^2, \tau_1^2, \tau_0$
 variance résiduelle = var(R): σ^2

Conclusion:

-- les effets d'interaction entre-niveaux sont significatifs;
-- ces effets font disparaître la variance pente, et diminuent la variance intercept.

Donc, les différences entre les usines dans les effets de l'orientation du supérieur peuvent être expliquées par le degré de réglementation et l'existence des marchés internes du travail.

En sommaire

Phase préalable: groupez les cas analysés dans des groupes qui représentent le contexte social pertinent.

Phase 1: déterminez la proportion de variance qui varie entre les groupes.

Phase 2: pour les variables importantes au niveau 1, distinguez les effets intra-groupe et entre-groupes, et déterminez la variabilité de leurs effets entre les groupes [coefficients aléatoires].

Phase 3: expliquez les différences entre les groupes (différences dans les effets de X sur Y; coefficients de régression qui sont spécifiques aux groupes) par des variables définies au niveau du contexte [diminution de la variance des coefficients aléatoires].

Pour les données longitudinales, l'analyse multinationale également fournit une approche bien flexible.

1. Moments (occasions) fixes d'enquête:

-- covariations propres à chaque occasion utilisables sans problèmes;
-- données manquantes dans Y sur certaines occasions ne font pas de problème.

2. Moments (occasions) variés d'enquête:

-- modèle pour une population de courbes;
-- nombres et temps variés d'enquête ne font pas de problème.

Cette approche fournit souvent des résultats plus facilement interprétables que l'approche traditionnelle de l'analyse de variance multivariée.

Pour les données de structure emboîtée, quand pourrait-on utiliser l'analyse multinationale; c'est-à-dire, le modèle linéaire hiérarchique? Plus généralement: quand utiliser les modèles aux coefficients aléatoires?

Pourrait-on considérer les entités (cas) sur chaque niveau comme des représentants d'une certaine population?

Y a-t-il des effets inobservés qui pourraient être considérés comme diverses sources de variance inexpliquée?

Sont les résidus (termes inexpliqués) distribués approximativement selon la loi normale?

Règle pratique:

Utilisation des modèles aux coefficients aléatoires questionnable pour moins de 5 groupes; utilisation des modèles aux coefficients fixes (ANCOVA) questionnable pour plus que 10 groupes.

Message méthodologique:

-- L'analyse multinationale fournit une combinaison automatique de la variabilité intra-groupe et celle inter-groupe (écarts-type corrects pour données emboîtées, en contraste avec régression MCO).
-- Faites une distinction entre la régression intra-groupe et celle entre-groupes.

-- Pentes aléatoires sont intéressantes. Ce sont des effets inexpliqués du contexte (si niveau 2 = contexte); explication peut-être possible sur la base des interactions entre-niveaux.

LOGICIELS

Mlwin

HLM

SAS

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

LITTÉRATURE

A. S. Bryk et S. W. Raudenbush (1992), *Hierarchical Linear Models, Applications and Data Analysis Methods*. London: Sage.

T. A. DiPrete et J. D. Forristal (1994), "Multilevel models: Methods and substance", *Annual Review of Sociology*, 20, 331 - 357.

T. A. B. Snijders et R. J. Bosker (1999), *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.