Multilevel Data Analysis

AQ801-3-M & Version 1.1

The Random Intercept Model Part A

Contents & Structure

- Terminology and notation
- A regression model: fixed effects only
- Variable intercepts: fixed or random parameters? When to use random coefficient models
- Definition of the random intercept model

Recap:

- When handling multilevel data, aggregation or disaggregation method should be used?

Learning Outcomes

- **At the end of this topic, You should be able to:**
  - Discuss the different between random effect and fixed effect.
  - Design random intercept model

Key Terms You Must Be Able To Use

- If you have mastered this topic, you should be able to use the following terms correctly in your assignments and exams:
- *Regression model*
- *Intercept*
- *Fixed effects*
- *Random effects*
- *Residuals*
- *Fixed effect model*
- *Analysis of covariance model*
- *Ordinary least square (OLS)*
- *Mixed model*
- *Random Intercept model*
- *Random intercept*
- *Empty model*
- *Intraclass correlation coefficient*

- ***Within- and between-group regression coefficients.***
- ***Cross-level interaction effect***
- ***Parameter estimation***

Learning Outcome 1:

- Discuss the different between random effect and fixed effect.

Introduction

- In the preceding topics is was argued that the best approach to the analysis of multilevel data is one that represents within-group as well as between-group relations within a single analysis.
- Very often, it makes sense to conceive of the unexplained variation within groups and that between groups as random variability.

- For a study of students within schools, for example, this means that not only unexplained variation between students, but also unexplained variation between schools is regarded as random variability.
- This can be expressed by statistical models with so-called random coefficients.

Terminology and notation

- The number of groups in the data is denoted by N; the number of individuals in the groups is denoted by n_j for group j (j = 1, 2, 3,….N).
- The total number of individuals is denoted by $M = \sum_j n_j$

- Instead of “explanatory variable’ the names ‘predictor variable” and “independent variable” are also used; and “criterion” is also used for “dependent variable”.
- The dependent variable must be a variable at level one

- For group *j*,
- z_j is the explanatory variable at the group level.
- To understand the notation, it is essential to realize that the indices *i* and *j* indicate precisely what the variables depend on.
- The notation Y_{ij}, for example, indicates that the value of variable *Y* depends on group *j* and also on individual *i*.

- Since the individuals are nested within groups, the index *i* makes sense only if it is accompanied by the index *j*; to identify individual i = 1, we must know which group we are referring to.
- The notation z_j, on the other hand, indicates that the value of *Z* depends only on group *j* and not on individual *i*.

- The basic idea of multilevel modeling is that the outcome variable Y has an individual as well as a group aspect.
- This also carries through to other level-one variables.
- The mean of X in one group may be different from the mean in another group.
- In other words, X may (and often will) have a positive between-group variance.

A regression model: fixed effects only

- The simplest model is one without the random effects that are characteristic of multilevel models; it is the classical model of multiple regression.
- This model states that the dependent variable, Y_{ij} can be written as the sum of a systematic part (a linear combination of the explanatory variables) and a random residual
- $Y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 z_j + R_{ij}$ ---- (4.1)

- In this model equation, the β_s are the regression parameters: β_0 is the intercept (i.e. the value obtained if both x_{ij} and z_j are 0), β_1 is the coefficient for the individual variable X, and β_2 is the coefficient for the group variable Z.
- The variable R_{ij} is the residual (sometimes called errors)
- The model has a multilevel nature only to the extent that some explanatory variables may refer to the lower and others to the higher level.

- In designs with group sizes larger than 1, the nesting structure often cannot be represented completely in the regression model by the explanatory variables.
- Additional effects of the nesting structure can be represented by letting the regression coefficients vary from group to group.
- Thus, the coefficients β_0 and β_1 in equation (4.1) must depend on the group, denoted by j.

- This is expressed in the formula by an extra index j for these coefficients.
- This yields the model
- $Y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \beta_2 z_j + R_{ij}$ ----(4.3)
- Group j can have a higher (or lower) value of β_{0j}, indicating that, for any given value of X, they tend to have higher (or lower) values of the dependent variable Y.

- Groups can also have a higher or lower value of β_{1j}, which indicates that the effect of X on Y is higher or lower.
- Since Z is a group-level variable, it would not make much sense conceptually to let the coefficient of Z depend on the group.
- Therefore β_2 is left unaltered in this formula.

- The simplest version of model (4.3) is that where β_{0j} and β_{1j} are constant (do not depend on j), that is, the nesting structure has no effect, and we are back at model (4.1).
- If on the other hand, the coefficients β_{0j} and β_{1j} do depend on j, then these regression models may give misleading results.

- Then it is preferable to take into account how the nesting structure influences the effects of X and Z on Y.
- This can be done using the random coefficient model of this and the following topics.
- This topic examines the case where the intercept β_{0j} depends on the group

Learning Outcome 2:

- Design random intercept model

Variable intercepts: fixed or random parameters?

- Let us first consider only the regression on the level-one variable X.
- A first step toward modelling between-group variability is to let the intercept vary between groups.
- This reflects the tendency for some groups to have, on average, higher responses Y and others to have lower responses.

- This model is halfway between (4.1) and (4.3) in the sense that the intercept β_{0j} does depend on the group but the regression coefficient of X, β_1 is constant:
- $Y_{ij} = \beta_{0j} + \beta_1 x_{ij} + R_{ij}$----- (4.4)
- For simplicity here the effect of Z or other variables is omitted

- This model is depicted in Figure 4.1 in the next slide.
- The group-dependent intercept can be split into an average intercept and the group-dependent deviation:
- $\beta_{0j} = \gamma_{00} + U_{0j}$
- the notations for the regression coefficients is changed here, and the average intercept is called γ_{00} while the regression coefficient for X is called γ_{10}.

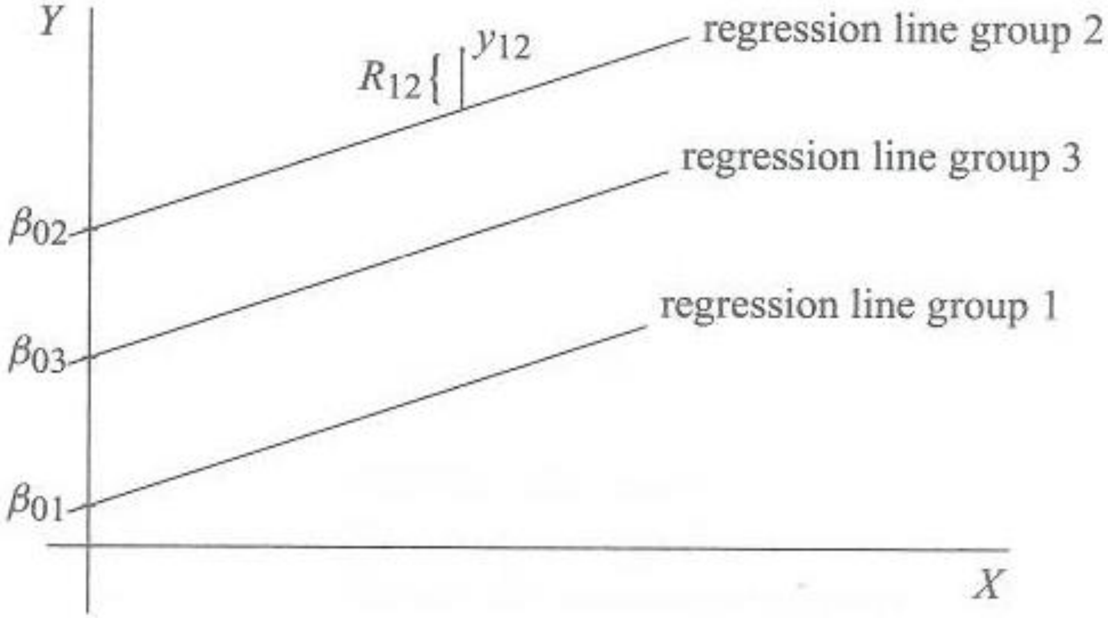

Figure 4.1: Different parallel regression lines. The point y_{12} is indicated with its residual R_{12}.

- Substitution now leads to the model
- $Y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + U_{0j} + R_{ij} ---- -(4.5)$
- The values U_{0j} are the main effects of the groups: conditional on an individual having a given X-value and being in group j, the Y-value is expected to be U_{0j} higher than in the average group.

- Model (4.5) can be understood in two ways:
- (1) as a model where the U_{0j} are fixed parameters, N groups, of the statistical model.
- (2) as a model where the U_{0j} are independent and identically distributed random variables.
- Note that the U_{0j} are the unexplained group effects, which also may be called group residuals, controlling for the effects of variable X

- This is the simplest random coefficient regression model.
- It is called the random intercept model because the group-dependent intercept, $\gamma_{00} + U_{0j}$, is a quantity that varies randomly from group to group.

- Note that models (4.1) and (4.2) are OLS models, or fixed effects models, which just take the nesting structure account at the minimum level: by the use of a group-level variable Z_j), whereas models of type (1) above are OLS models that do take the nesting structure into account.

- The latter kind of OLS model has a much larger number of regression parameters, since in such models N groups lead to N-1 regression coefficients
- It is important to distinguish between these two kind of OLS models in discussing how to handle data with a nested structure

When to use random coefficient models

- Which of these two interpretations is the most appropriate in a given situation depends on the focus of the statistical inference, the nature of the set of N groups, the magnitudes of the group sample size n_j, and the population distributions involved.

- if the groups are regarded as unique categories and the researcher wishes primarily to draw conclusions pertaining to each of these N specific categories, then it is appropriate to use the analysis of covariance (fixed effects) model.
- Examples are groups defined by gender of ethnic background

Populations and populations

- Having chosen to work with a random coefficient model, the researcher must be aware that more than one population is involved in the multilevel analysis.
- Each level corresponds to a population.
- For example, for a study of students in schools, there is a population of schools and a population of students: for voters in municipalities, there is a population of municipalities and a population of voters.

Summary of Main Teaching Points

- It is conceive of the unexplained variation within groups and that between groups as random variability.
- Additional effects of the nesting structure can be represented by letting the regression coefficients vary from group to group

Question and Answer Session

Q & A

What we will cover next

- **Random Intercept Model Part B**