Multilevel Data Analysis

AQ801-3-M & Version 1.1

Hierarchical Linear Model

Contents & Structure

- Random slopes
- Explanation of random intercepts and slopes
- Specification of random slope models
- Estimation

Recap:

- Discuss the different between multiple linear regression and random intercept model.

Learning Outcomes

- **At the end of this topic, You should be able to**
- Discuss the general case of HLM.
- Interpret the output of HLM.

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:**
- *Random slope model*
- *Hierarchical Linear Model*

Leaning Outcome 1

- Discuss the general case of HLM.

Introduction

- In the previous topic, the simpler case of the hierarchical linear model (HLM) was treated, in which ***only intercepts*** are assumed to be random.
- In the more general case, slopes may also be random.
- In a study of students within schools, for example, the effect of the pupil's IQ or SES on scholastic performance could differ between schools, i.e. the slope can be different.

Random Slopes

- In the random intercept model of Topic 4, the groups differ with respect to the average value of the dependent variable: the only random group effect is the random intercept.
- But the relation between explanatory and dependent variables can differ between groups in more ways.

- For example, in the field of education (nesting structure: students within classrooms), it is possible that the effect of socio-economic status of students on their scholastic achievement is stronger than in others.
- it is possible that some subjects progress faster than others.

- In the hierarchical linear model (HLM), it is modeled by **random slopes**.
- a model with group-specific regressions of Y on one level-one variable X only, like model (4.3) but without the effect of Z
- $Y_{ij} = \beta_{0j} + \boldsymbol{\beta_{1j}} x_{ij} + R_{ij} --- (5.1)$
- The intercept β_{0j} as well as the regression coefficients, or slopes, β_{1j} are group-dependent.

- These group-dependent coefficients can be split into an average coefficient and the group-dependent deviation.
- $\beta_{0j} = \gamma_{00} + U_{0j}, \beta_{1j} = \gamma_{10} + U_{1j}$ $-----(5.2)$
- Substitution leads to the model
- $Y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + U_{0j} + U_{1j}x_{ij} + R_{ij}$ ---(5.3)

- It is assumed that the level-two residuals U_{0j} and U_{1j} as well as the level-one residuals R_{ij} have mean 0, given the values of the explanatory variable X.
- Thus, γ_{10} is the average regression coefficient just as γ_{00} is the average intercept.

- The first part of (5.3), $\gamma_{00} + \gamma_{10} x_{ij}$ is called the fixed part of the model.
- The second part, $U_{0j} + U_{1j} x_{ij} + R_{ij}$, is called the random part.
- The term $U_{1j} x_{ij}$ can be regarded as a random interaction between group and X.
- This model implies that the groups are characterized by two random effects: their intercept and their slope.

- These are called ***latent variables***, meaning that they are not directly observed but play a role “behind the scenes” in producing the observed variables.
- The variance of the level-one residuals R_{ij} is again denoted σ^2; the variances and covariance of the level-two residuals (U_{oj}, U_{1j}) are denoted as follows:

$$Var(U_{oj}) = \tau_{00} = \tau_0^2$$

$$Var(U_{1j}) = \tau_{11} = \tau_1^2$$

- $cov(U_{0j}, U_{1j}) = \tau_{01}$ ---------------- (5.4)

Heteroscedasticity

- Model (5.3) implies not only that individuals within the same group have correlated Y-values, but also that this correlation as well as the variance of Y are dependent on the value of X.
- For example, in a study of the effect of socio-economic status (SES) on scholastic performance (Y), we have schools which do not differ in their effect on high-SES children, but do differ in the effect of low SES on Y(eg. because teacher expectancy effect).

- Then for children from a high-SES background it does not matter which school they go to, but for children from a low-SES background it does.
- The school then adds a component of variance for the low-SES children, but not for the high-SES children: as a consequence, the variance of Y (for a random child at a random school) will be larger for the former than for the latter children.

- This example shows that model (5.3) implies that the variance of Y, given the value x on X, depends on x.
- This is called heteroscedasticity in statistical literature.

- An expression for the variance of (5.3) is obtained as the sum of the variances of the random variables involved plus a term depending on the covariance between U_{0j} and U_{1j} (the other random variables are uncorrelated).
- Here we also use the independence between the level-one residual R_{ij} and the level-two residuals (U_{0j}, U_{1j}).

- From (5.3) and (5.4), we obtain the result

$$\operatorname{var}\left(Y_{ij} \middle| x_{ij}\right) = \tau_0^2 + 2\tau_{01} x_{ij} + \tau_1^2 x_{ij}^2 + \sigma^2 \quad --(5.5)$$

- Similarly, for two different individuals (i and i', with $i \neq i'$) in the same group,

$$\operatorname{cov}(Y_{ij}, Y_{i'j} \mid x_{ij}, x_{i'j}) = \tau_0^2 + \tau_{01}(x_{ij} + x_{i'j}) + \tau_1^2 x_{ij} x_{i'j} \quad --(5.6)$$

- Formula (5.5) implies that the residual variance of Y is minimal for $x_{ij} = \frac{-\tau_{01}}{\tau_{11}}$

Do not force τ_{01} to be 0!

- All of the preceding discussion implies that the group effects depend on x: according to (5.3), this effect is given by $U_{0j} + U_{1j}x$
- This is illustrated by Figure 5.1, a hypothetical graph of the regression of school achievement (Y) on intelligence (X).

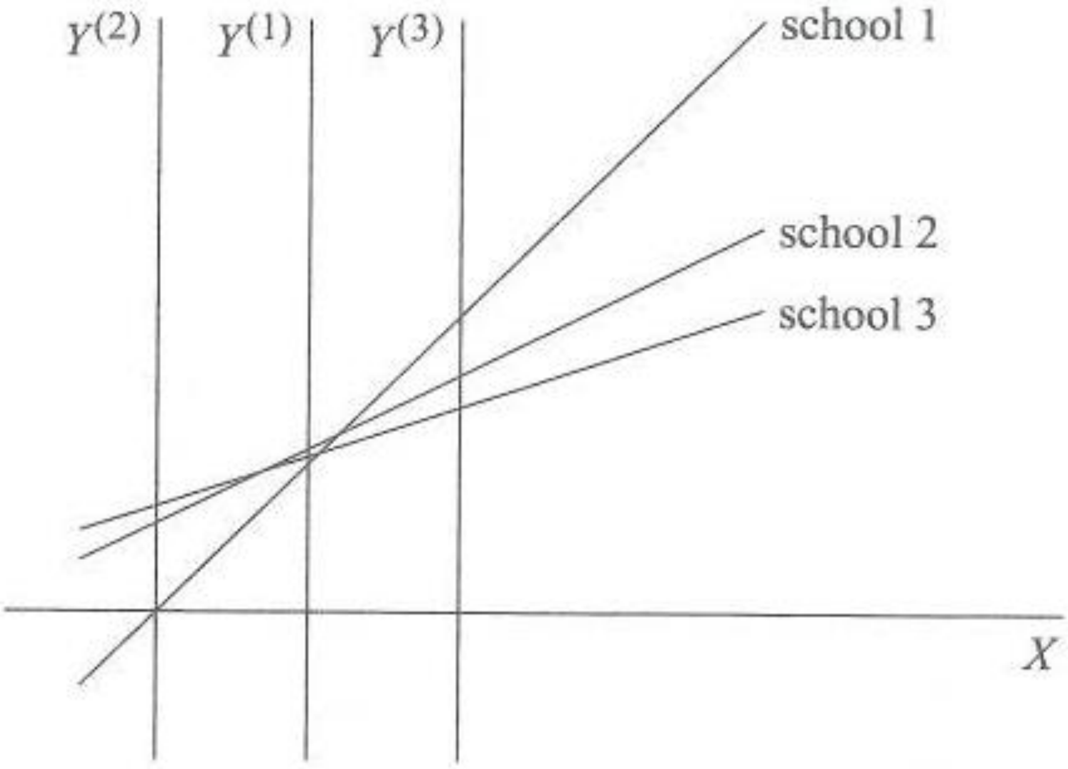

Figure 5.1: Different vertical axes.

- It is clear that there are slope differences between the three schools.
- Looking at the $Y^{(1)} - axis$, there are almost no intercept differences between the schools.
- But if we add a value 10 to each intelligence score x, then the Y-axis is shifted to the left by 10 units: the $Y^{(2)} - \text{axis}$

- Now, school 3 is the best, school 1 the worst: there are strong intercept differences.
- If we had subtracted 10 from the x-scores, we would have obtained the $Y^{(3)} - axis$, again with intercept differences but now in reverse order.
- This implies that the intercept variance τ_{00}, as well as the intercept-by-slope covariance τ_{01}, depend on the origin (0-value) for the X-variable.

- from this, we can learn two things:
- (1) since the origin of most variables in the social sciences is arbitrary, in random slope models the intercept-by-slope covariance should be a free parameter estimated from the data, and not *a priori* constrained to the value 0 (i.e. left out of the model)

- (2) In random slope models we should be careful with the interpretation of the intercept variance and the intercept-by-slope covariance, since the intercept refers to an individual with x=0.
- For the interpretation of these parameters it is helpful to define the scale for X so that x=0 has an interpretable meaning, preferably as a reference situation

- In nesting structures of individuals within groups, it is often convenient to let x=0 correspond to the overall mean of the population or the sample – for example, if X is IQ at the conventional scale with mean 100, it is advisable to subtract 100 to obtain a population mean of 0.

Interpretation of random slope variances

- For the interpretation of the variance of random slopes, τ_1^2, it is also illuminating to take the average slope, $\gamma_{10},$ into consideration.
- Model (5.3) implies that the regression coefficient, or slope, for group j is $\gamma_{10} + U_{1j}$.
- This is a normally distributed random variable with mean γ_{10} and standard deviation $\tau_1 = \sqrt{\tau_1^2}$

- Since about 95% of the probability of a normal distribution is within two standard deviations from the mean, it follows that approximately 95% of the groups have slopes between $\gamma_{10} - 2\tau_1$ and $\gamma_{10} + 2\tau_1$
- Conversely, about 2.5% of the groups have a slope less than $\gamma_{10} - 2\tau_1$ and 2.5% have a slope steeper than $\gamma_{10} + 2\tau_1$.

Leaning Outcome 2

- Interpret the output of HLM.

Example 5.1 A random slope for IQ

- We continue our study of the effect of IQ on a language test score.
- Recall that IQ is here on a scale with mean 0 and its standard deviation in this data set is 2.04.
- A random slope of IQ is added to the model, that is, the effect of IQ is allowed to differ between classes.
- The model is an extension of model (5.3): a fixed effect for the class average on IQ is added.

- The model reads
- $Y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + \gamma_{01} \bar{x}_{.j} + U_{0j} + U_{1j} x_{ij} + R_{ij}$
- The result can be read from Table 5.1.
- Note that the “Level-two random part” heading refers to the random intercept and random slope which are random effects associated with the level-two units (the class), but that the variable that has the random slope, IQ, is itself a level-one variable.

Coding to run random slope model

- options validvarname=any;
- libname mdadat XLSX '/folders/myfolders/MDA/MDA2.XLSX';
- proc mixed data=mdadat.sheet1 covtest;
- class schoolnr;
- model lang_post = IQ_verb Iqmean /s;
- random Int IQ_verb / type=un sub=schoolnr s;
- repeated / type =cs subject=schoolnr r;
- run;

Table 5.1: Estimates for random slope model.

Fixed effect	Coefficient	S.E.
γ_{00} = Intercept	41.127	0.234
γ_{10} = Coefficient of IQ	2.480	0.064
γ_{01} = Coefficient of $\overline{\text{IQ}}$ (group mean)	1.029	0.262
Random part	**Parameters**	**S.E.**
Level-two random part:		
$\tau_0^2 = \text{var}(U_{0j})$	8.877	1.117
$\tau_1^2 = \text{var}(U_{1j})$	0.195	0.076
$\tau_{01} = \text{cov}(U_{0j}, U_{1j})$	–0.835	0.217
Level-one variance:		
$\sigma^2 = \text{var}(R_{ij})$	39.685	0.964
Deviance	24,864.9	

- Figure 5.2 presents a sample of 15 regression lines, randomly chosen according to the model of Table 5.1.
- Should the value of 0.195 for the random slope variance be considered to be high?
- The slope standard deviation is $\sqrt{0.195} = 0.44$, and the average slope is $\gamma_{10} = 2.48$

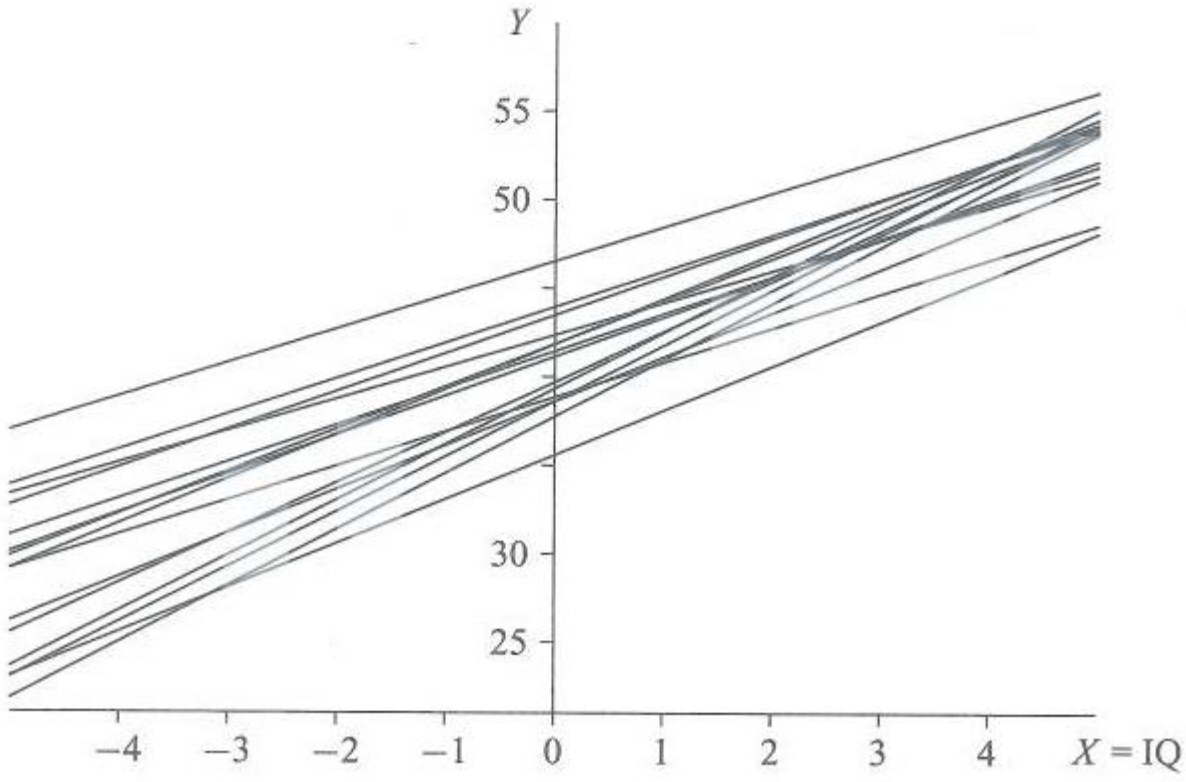

Figure 5.2: Fifteen random regression lines according to the model of Table 5.1 (with randomly chosen intercepts and slopes).

- The values of average slope $\pm$ two standard deviations range from 1.60 to 3.36.
- This implies that the effect of IQ is clearly positive in all classes, but high effects of IQ are more than twice as large as low effects.
- This may indeed be considered an important difference.
- (as indicated above, “high” and “low” are respectively understood here as those value occurring in classes with the top 2.5% and the bottom 2.5% of the class-dependent effects)

Interpretation of Intercept-slope covariance

- The correlation between random slope and random intercept is $\frac{-0.83}{\sqrt{8.88(0.195)}} = -0.63$
- Recall that all variables are centered (have zero mean), so that the intercept corresponds to the language test score for a pupil with average intelligence in a class with average mean intelligence (grand mean).

- The negative correlation between slope and intercept means that classes with a higher performance for a pupil of average intelligence have a lower within-class effect of intelligence (U_{1j} is deviational variable)
- Thus, the higher average performance tends to be achieved more by higher language scores of the less intelligent, than by higher scores of the more intelligence students.

- To investigate how the contribution of classrooms to students' performance depends on IQ, consider the equation implied by the parameter estimates:
- $Y_{ij} = 41.13 + 2.4801 IQ_{ij} + 1.029\overline{IQ_{.j}} + U_{oj} + U_{1j} IQ_{ij} + R_{ij}$
- Recall from Example 4.2 that the standard deviation of the IQ score is about 2, and the mean is 0.

- Hence, students with an IQ among the bottom few percent or the top few percent have IQ scores of about ± 4
- Substituting these values in the contribution of the random effects gives $U_{oj} \pm 4U_{1j}$.
- It follows from equations (5.5) and (5.6) that for students with $IQ \pm 4$, we have
- $var\left(Y_{ij} \middle| IQ_{ij} = -4\right) = 8.88 + 2x(-0.835)x(-4) + (-4)^2 x0.195 + 39.69 = 58.37$

- $cov\left(Y_{ij}, Y_{i'j} \middle| IQ_{ij} = -4, IQ_{i'j} = 4\right) = 8.88 - 16x0.195 = 5.76$
- $var\left(Y_{ij} \middle| IQ_{ij} = 4\right) = 8.88 - 8x0.835 + 16x0.195 + 39.69 = 45.01$
- And therefore,
- $P\left(Y_{ij}, Y_{i'j} \middle| IQ_{ij} = -4, IQ_{i'j} = 4\right) = \frac{5.76}{\sqrt{58.37(45.01)}} = 0.11$

- Hence, the language test scores of the most intelligent and the least intelligent students in the same class are positively correlated over the population of classes: classes that have relatively good results for the less able tend to have relatively good results for the more able students.
- This positive correlation corresponds to the result that the value of IQ for which the variance given by (5.5) is minimal, is outside the range from -4 to +4.

- For the estimates in Table 5.1, this variance is (with some rounding)
- $Var\left(Y_{ij} \middle| IQ_{ij} = x\right) = 8.88 - 1.67 + 0.195x^2 + \sigma^2$
- Taking the derivative of this function of x equating it to 0 yields that the variance is minimal for $x = \frac{1.67}{0.39} = 4.3$, just outside the IQ range from -4 to +4.

- This again implies that the classes tend mostly to perform either higher, or lower, over the entire range of IQ.
- This is illustrated also by Figure 5.2 (which, however, also contains some regression lines that cross each other within the range of IQ, illustrating that the random nature of these regression lines will lead to exceptions to this pattern)

Explanation of random intercepts and slopes

- The aim of regression analysis is to explain variability in the outcome (i.e. dependent) variable.
- Explanation is understood here in a quite limited way – as being able to predict the value of the dependent variable from knowledge of the values of the explanatory variables.

- The unexplained variability in single-level multiple regression analysis is just the variance of the residual term.
- Variability in multilevel data, however, has more complicated structure.
- This is related to the fact, that several populations are involved in multilevel modeling: one population for each level.

- Explaining variability in a multilevel structure can be achieved by explaining variability between individuals but also by explaining variability between groups; if there are random slopes as well as random intercepts, at the group level one could try to explain the variability of slopes as well as intercepts.

- In the model defined by (5.1) – (5.3), some variability in Y is explained by the regression on X, that is, by the term $\gamma_{10}x_{ij}$; the random coefficients U_{0j}, U_{1j} and R_{ij} each express different parts of the unexplained variability.

- $\sigma^2 = var(R_{ij})$ is the residual variance, which can be diminished by including other level-one variables.
- Since group compositions with respect to level-one variables can differ from group to group, inclusion of such variables may also diminish residual variance at the group level.

- U_{0j} and U_{1j} are not directly observed, i.e. latent variables.
- To reduce the unexplained variability associated with the above 2 terms, just to expand equations (5.2) by predicting the group dependent regression coefficients β_{0j} and β_{1j} from level-two variables Z.

- This leads to regression formulas for β_{0j} and β_{1j} on the variable Z given by
- $\beta_{0j} = \gamma_{00} + \gamma_{01} z_j + U_{0j}$ -------(5.7)
- $\beta_{1j} = \gamma_{10} + \gamma_{11} z_j + U_{1j}$ -------(5.8)
- The βs are treated as dependent variables in the regression models.
- These are "latent regressions" because βs cannot be observed without error.

- Equation (5.7) is called an **intercepts as outcomes** model.
- Equation (5.8) is called a **slope as outcomes** model.

Summary of Main Teaching Points

- Under the hierarchical linear model (HLM), slopes may also be random.
- The relation between explanatory and dependent variables can differ between groups in more ways.

Question and Answer Session

Q & A

What we will cover next

- **Testing and Model Specification**