Mixed models

Stian Lydersen

11 March 2015

Revised 11 March 2015

When?

- Relevant when analysing **repeated measures** within clusters, such as:
  - Students within classes
  - Repeated measures in the same patients
- Alternative analysis methods:
  - Repeated measures ANOVA
  - Mixed models
  - Generalized estimating equations (GEE)

Repeated measures ANOVA

- Only complete cases are included in the analysis
- Unbiased only if data are missing completely at random (MCAR)
- The underlying mathematical model is not transparent
- Was an attractive method before computers became powerful
- Ought to be in the museum.

Mixed models

- Includes all subjects, also those with missing data at some time point(s)
- Unbiased under the less restrictive missing at random (MAR) assumption
- Transparent mathematical model

Generalized estimating equations

- A useful alternative to Mixed models if the outcome is for example binary (logistic regression) or count (Poisson regression).
- Unbiased only if data are MCAR

Standard linear regression:

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij}$$

where subscript ij denotes observation j within cluster i,

and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$.

The parameters β_0 and β_1 represent fixed effects.

Random intercept model:

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + b_{0i} + \varepsilon_{ij}$$

where $b_{0i} \sim N(0, \sigma_{b_0}^2)$ is the random effect of cluster i.

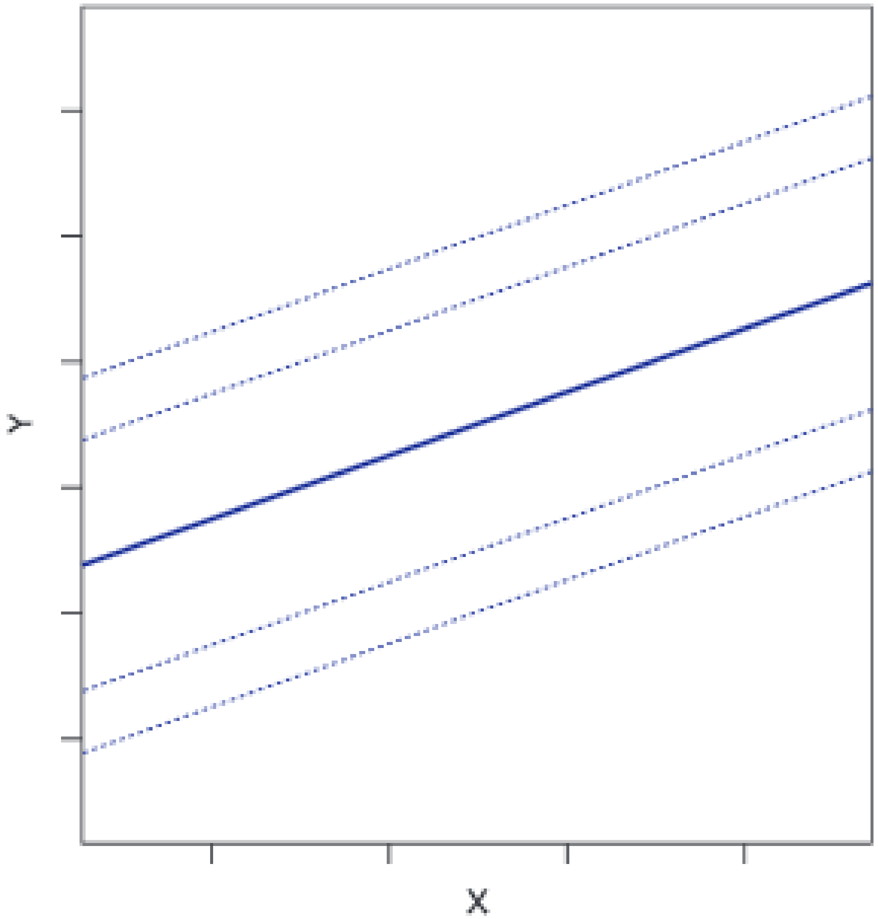

Figure 7.1 Illustration of the random intercept model, with fixed effect (solid line) and cluster variation around this line (dotted lines).

Random intercept and random slope:

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + b_{0i} + b_{1i} x_{ij} + \varepsilon_{ij}$$

where $b_{1i} \sim N(0, \sigma_{b_1}^2)$ represents the random slope for cluster i.

Note 1:

In (almost) every statistics package, the default is to assume the random effects $b_{0i}, b_{1i} (\ldots)$ to be independent. This is completely unrealistic: Generally, their covariances are nonzero. So their variance-covariance matrix must be specified as unstructured.

Note 2:

Adding one or more random slopes causes a large increase in the number of parameters, and make estimation computationally very demanding or impossible.

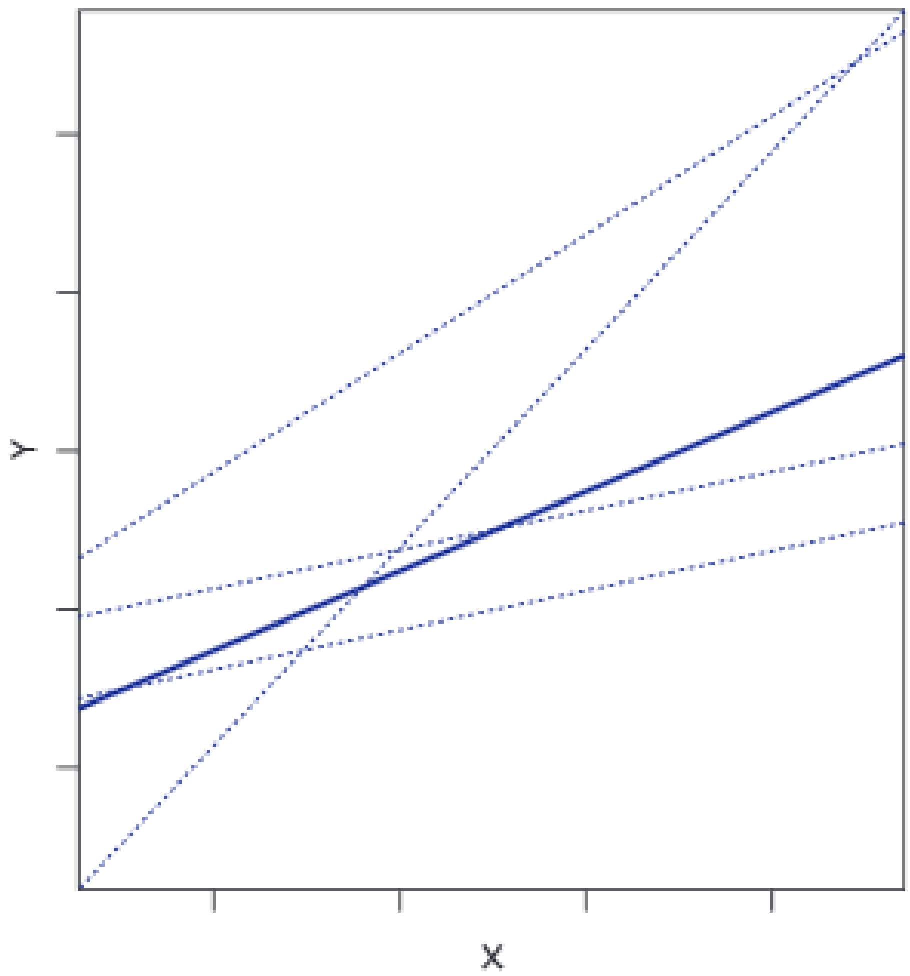

Figure 7.2 Illustration of a model with random intercept and random slope. Marginal effect (solid line) and cluster variation (dotted lines).

Multilevel analysis in SPSS (with two levels)

The data file must be in “long” format, that is, one line per observation within the cluster. You can convert the file from “wide” to “long” format using ***Data***, ***Restructure***, and *restructure selected variables into cases.*

Choose ***Analyze***, ***Mixed Models*** and ***Linear***.

Move the group or cluster variable to *Subjects*. ***Continue***.

Add the outcome variable in *Dependent Variable*.
Continuous covariates go into *Covariate(s)*, and categorical covariates go into *Factor(s)*. Dichotomous covariates may alternatively go into *Covariate(s)*.

Click ***Fixed***. Put the middle button on *main Effects*. Move all variables from *Factors and Covariates* to *Model*. Then move the interactions, if any, into the *model*, after setting the middle button on *Interaction*.

After ***Continue*** click ***Random***. In the lower part move the group or cluster variable to the right. In the upper part move to the right the variables (if any) for which you want a random slope.
Important:
Mark ***Include Intercept***, because otherwise, there will be no random intercept.
Covariance Type must be put on *Unstructured* (important if you included (at least) one random slope)

In order to get the regression coefficients, click on ***Statistics*** and ***Parameter estimates*.** After ***Continue*** and ***OK*** the analysis will be performed.

References

Thoresen, M. 2012, "Longitudinal Analysis," *In Medical statistics in clinical and epidemiological research*, M. Veierød, S. Lydersen, & P. Laake, eds., Oslo: Gyldendal Akademisk, pp. 259-287.

Thoresen, M. & Gjessing, H. K. 2012, "Mixed Models," *In Medical statistics in clinical and epidemiological research*, M. Veierød, S. Lydersen, & P. Laake, eds., Oslo: Gyldendal Akademisk, pp. 231-258.

Twisk, J.W.R. 2013. *Applied longitudinal data analysis for epidemiology a practical guide*, 2nd ed.