Mixed models

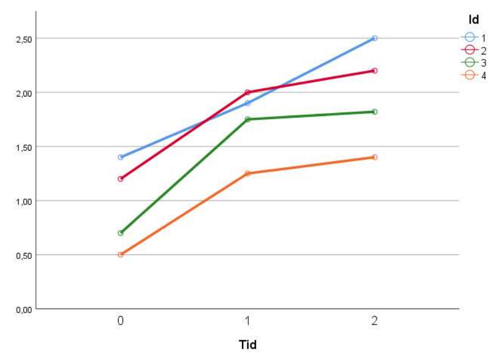
Stian Lydersen

21 November 2018

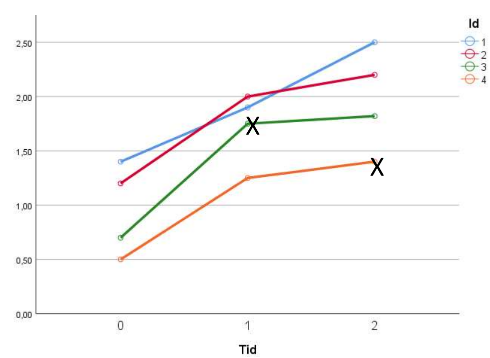
When?

- 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)

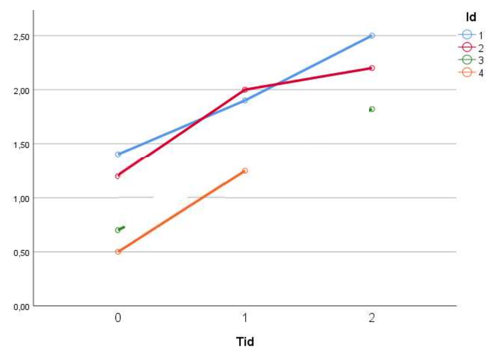
Longitudinal study – complete data



Longitudinal study – missing data



Longitudinal study – missing data



Types of missing data (Missing data mechanism)	The probability that a data value is missing (unobserved) can depend on
MCAR Missing Completely at Random (Mangler helt tilfeldig)	Neither observed or unobserved values
MAR Missing at Random (Mangler betinget tilfeldig)	Only observed values
MNAR Missing Not at Random (mangler ikke-tilfeldig)	Unobserved values (and observed values)

Norske betegnelser:
Lydersen, S: «Manglende data – ikke helt tilfeldig». Akseptert i Tidsskrift for Den norske legeforening, 2019

Types of missing data (Sterne et al. 2009)

- Missing completely at random—There are no systematic differences between the missing values and the observed values. For example, blood pressure measurements may be missing because of breakdown of an automatic sphygmomanometer
- Missing at random—Any systematic difference between the missing values and the observed values can be explained by differences in observed data. For example, missing blood pressure measurements may be lower than measured blood pressures but only because younger people may be more likely to have missing blood pressure measurements
- Missing not at random—Even after the observed data are taken into account, systematic differences remain between the missing values and the observed values. For example, people with high blood pressure may be more likely to miss clinic appointments because they have headaches

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.

Should we quit using repeated measures analysis of variance?

Repeated Measures ANOVA, R.I.P.?

Charles E. McCulloch

It is a difficult experiment to run and to analyze: What are the effects of alcohol on sleepiness and does a hormone, pregnenolone, which has been shown to enhance memory in rat experiments, help alleviate the sleepiness? Each person is tested under each of four conditions on four different visits in random order: a placebo for the drug and for the hormone, alcohol alone, hormone alone, and the combination. Each subject is also queried multiple times within a visit in the minutes after alcohol (or placebo) ingestion. Some subjects drop out of the protocol without completing all the conditions and some of the sleepiness scores are not recorded within a visit because of difficulties executing the protocol. How should the data be analyzed?

$$Y_{ijk} = \mu + \alpha_j + \beta_k + \alpha\beta_{jk} + p_i + \varepsilon_{ijk}, \qquad (1)$$

where μ is the overall mean, α_j represents the alcohol effect, β_k the pregnenolone effect, $\alpha\beta_{jk}$ the interaction effect, p_i is the person effect, and ε_{ijk} is an error term. This model hypothesizes simple person effects that raise or lower (if the effect is negative) the average sleepiness in all four conditions. Interest focuses on the interaction, because the scientific question is whether pregnenolone helps to reduce the sleep-inducing effect of alcohol.

With a mean, error term and four explanatory factors in the model, the analysis of variance would partition the variability in Y_{ijk} into four sources: person, alcohol, pregnenolone, and the

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.

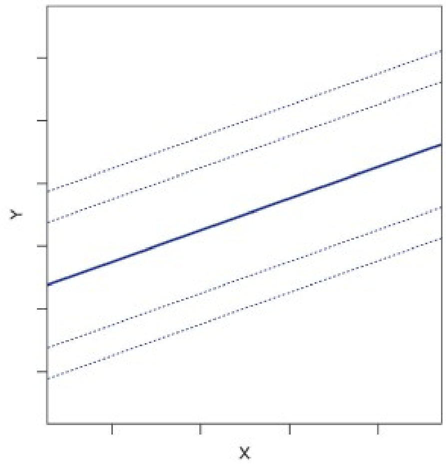


Figure 7.1 Illustration of the random intercept model, with fixed effect (solid line) and cluster variation around this line (dotted lines).

Random intercept and random slope:

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

where $b_{1i} \sim N(0, \sigma^2_{b_1})$ 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.

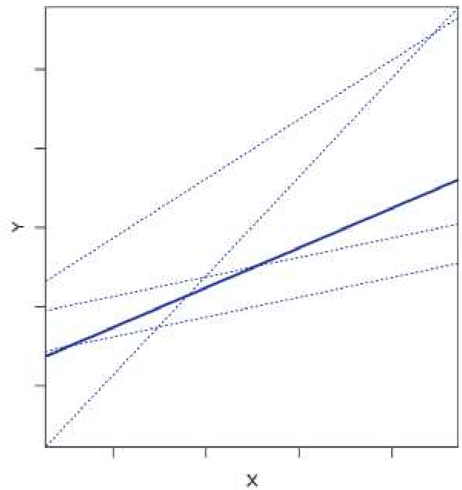


Figure 7.2 Illustration of a model with random intercept and random slope. Marginal effect (solid line) and cluster variation (dotted lines).

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

McCulloch, C.E. 2005. Repeated Measures ANOVA, R.I.P.? Chance, 18, (3) 29-33

Sterne, J.A., White, I.R., Carlin, J.B., Spratt, M., Royston, P., Kenward, M.G., Wood, A.M., & Carpenter, J.R. 2009. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*, 338, b2393

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.