Multilevel Models

Philippe Rast

PSC 204B:
General Linear Model
February 13, 2024

UC Davis, Winter Quarter

Topics

Overview

Hierarchical or Nested Data Structures

- Participants' data often come in natural groups or clusters that are nested within each other:
  - Participants belong to the same family
  - Participants live in the same residential home
  - Participants live in the same neighborhood
  - Participants work in the same organization
  - Participants visit the same classes or courses
  - etc.
  - Participants have been assessed several times (repeated measures)

Hierarchical or Nested Data Structures

- This grouping or clustering has the effect that participants within groups or clusters are **more similar to each other**, because
  - Participants share the same upbringing
  - Participants share the same nurses, nutrition, etc.
  - Participants share the same environment
  - Participants share the same organizational climate
  - Participants share the same teachers
  - etc.
  - Or: Participants *are* the same (repeated measures)

Hierarchical or Nested Data Structures

- Grouped or clustered data are called
  - **hierarchical**, because individual observations are nested into higher-order units
  - **multilevel**, because one can distinguish different data levels, e.g., individual versus group level

- There may be more than two levels of hierarchy
  - **Cross-Sectional:** Participants (level 1) are nested into the classes (level 2), classes are nested into schools (level 3), schools are nested into districts (level 4), districts...
  - **Longitudinal:** Repeated observations (level 1) are nested into participants (level 2)
  - **Both:** Repeated academic achievement scores (level 1) for individual students (level 2), nested within a school (level 3)

Hierarchical or Nested Data Structures

- Hierarchical or multilevel data lead to **dependent observations**
- ▷ **Violation of assumption** of independent and identically distributed (iid) random variables

- Consequences for analytical approaches ignoring this dependency in OLS regression:
  - A single observation does not contribute as much information as is assumed!
  - ▷ Standard errors of parameters are biased (too small)
  - ▷ Statistical significance tests will lead to the wrong conclusions (too liberal)

Hierarchical or Nested Data Structures

Not accounting for nestedness can lead to wrong conclusions:

Ecological Fallacy

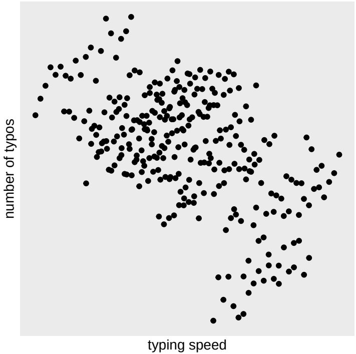
- *Ecological fallacy*: Drawing conclusions about nature of individuals from inferences about the group (cf. McCrae and John)

- Popularized by sociologist Hanan C. Selvin's (1958) paper on Durkheim's Suicide and Problems of Empirical Research
- Attributed to sociologist William S. Robinson's (1950)
- ▷ "There need be no correspondence between the individual correlation and the ecological correlation [or correlation among group means]" (p. 339)

- More generally (beyond correlation framework):
  - Results from between-person analyses will not necessarily match up with results from within-person analyses.

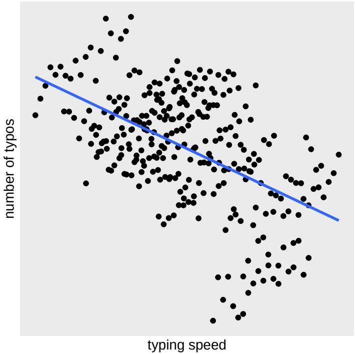
Between vs. Within

Ellen Hamaker's (2012) example on typing speed and typing errors.



Between vs. Within

Ellen Hamaker's (2012) example on typing speed and typing errors.

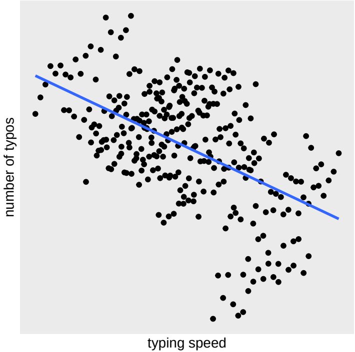



```
lm(y ~ speed, data = dat)

Coefficients:
Estimate     S.E t value  Pr(>|t|)
(Int.) 541.22  10.48   51.63   <2e-16 ***
speed   -3.10   0.29  -10.67   <2e-16 ***
---

Residual standard error: 64.65 on 298 df
Multiple R-squared:  0.2763
```

- correlation: $r = -.52$

Between vs. Within

Ellen Hamaker's (2012) example on typing speed and typing errors.



"If we were to generalize this result to the within-person level, we would conclude that if a particular person types faster, he or she will make fewer mistakes. **Clearly, this is not what we expect**: In fact, we are fairly certain that for any particular individual, the number of typos will increase if he or she tries to type faster. (p. 44)

- correlation: $r = -.52$

Between vs. Within

Ellen Hamaker's (2012) example on typing speed and typing errors.

Different individuals!

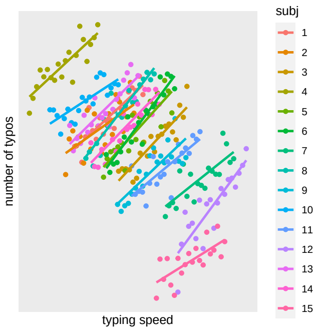
- Slowest individual (beginner) produces most typos
- Fastest individual (expert) produces fewest typos
- Generally:
  - ▷ Faster typing speed incurs more errors
- Two sources of change:

Between-Person Experts commit fewer errors (pc = average typing across individuals)

Within-Person Typing faster increases errors (pmc = changes in typing speed within-person)

Between vs. Within

Ellen Hamaker's (2012) example on typing speed and typing errors.



- Mdn: $r = .85$

Multilevel Model:

```
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ pc + pmc + (pmc | subj)
Data: dat

REML criterion at convergence: 2691.7

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 subj     (Intercept) 1095.7839 33.1026
          pmc            0.9289  0.9638  0.84
 Residual              379.9269 19.4917
Number of obs: 300, groups:  subj, 15

Fixed effects:
            Estimate Std. Error t value
(Intercept) 620.4805    20.9836  29.570
pc           -5.4448     0.5669  -9.605
pmc           5.0723     0.3163  16.039
```

Hierarchical or Nested Data Structures

- Different degrees of data dependency
- Consequences of ignoring nested structure depends on its degree:
  - If data dependency is small, approaches ignoring data dependency will hardly have en effect on statistics
  - If data dependency is large, approaches ignoring data dependency will strongly affect results of significance tests and interpretation
- Quantify data dependency
- Decide whether dependency is ignorable (small) or not (large)

Intraclass Correlation

- Amount of data dependency depends on the **proportion of variance between groups** or clusters **in relation to the total variance**:

$$\rho_{ic} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\epsilon^2} \tag{1}$$

  $\sigma_b^2 =$ between-person variance
  $\sigma_\epsilon^2 =$ residual variance or within person-variance
- ρ_{ic} is called the *intraclass correlation coefficient* (ICC, or cluster effect)

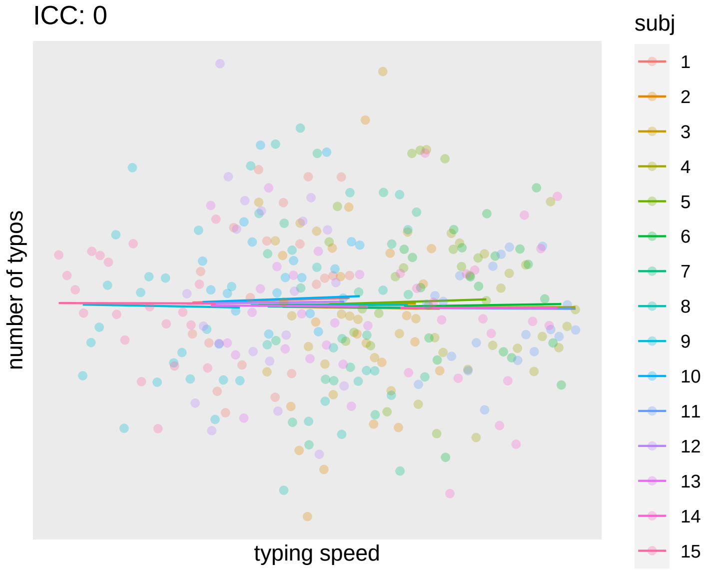
- Range of intraclass correlation

$$0 < \rho_{ic} < 1$$

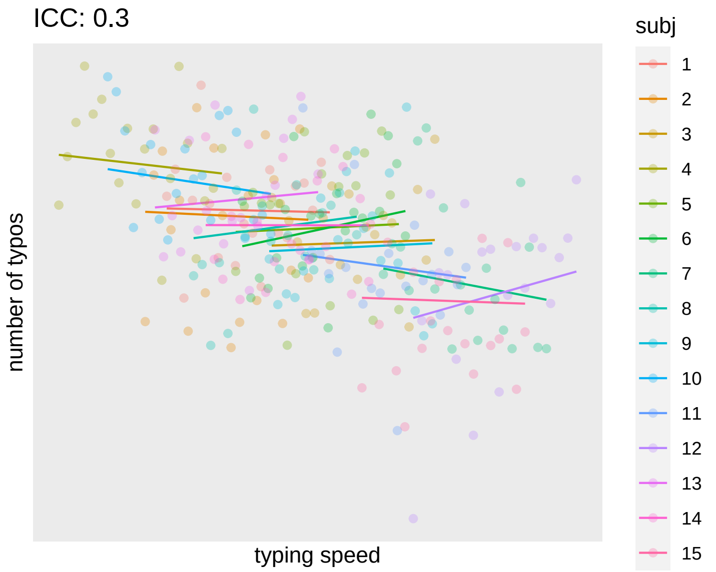
  No data dependency   Complete data dependency

- *Rule of thumb*: $\rho_{ic} > .05$ may be considered substantialerrors, etc.)
- For repeated measures data, ρ_{ic} is virtually always substantial

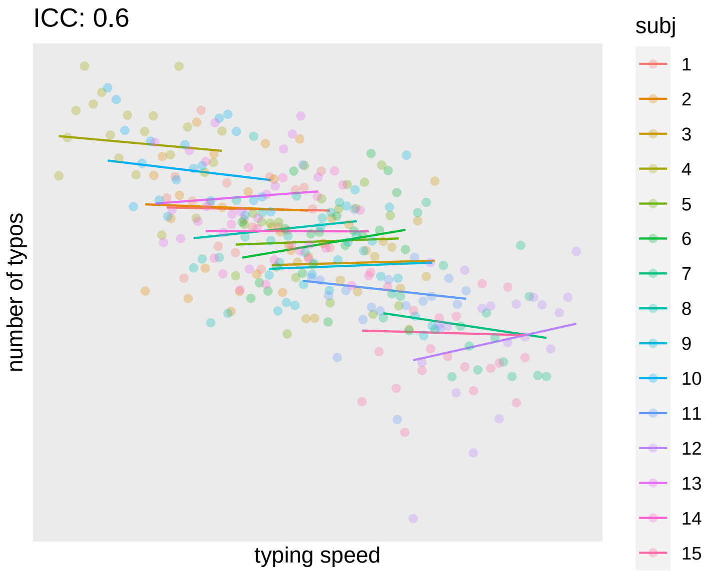
Intraclass Correlation



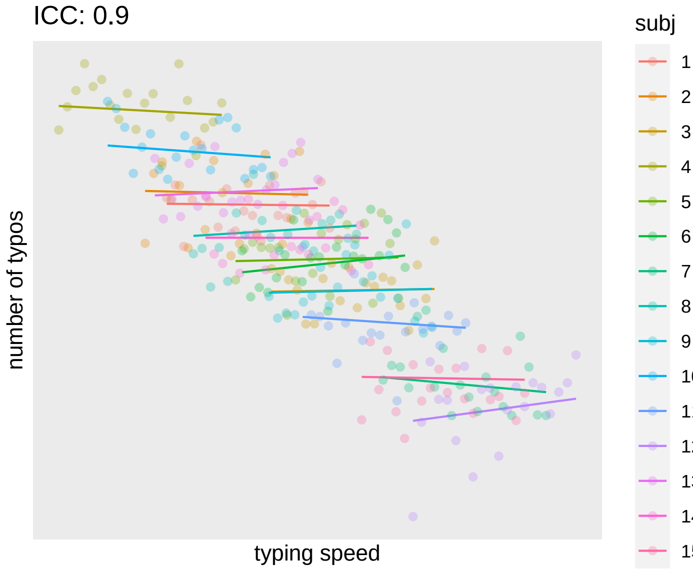
Intraclass Correlation



Intraclass Correlation



Intraclass Correlation



Cluster Effect

- Another perspective: Cluster effect

$$\mathrm{CE} = 1 + (n_i - 1)\rho_{ic}$$

▷ n_i is the average cluster size

- *Rule of thumb*: $\mathrm{CE} \geq 2$ indicates that the clustering in the data is influential
  - Example:
    $\rho_{ci} = .15$
    If $n_i = 5$ then $\mathrm{CE} = 1.6$
    If $n_i = 15$ then $\mathrm{CE} = 3.1$

    $\rho_{ci} = .55$
    If $n_i = 5$ then $\mathrm{CE} = 3.2$
    If $n_i = 15$ then $\mathrm{CE} = 8.7$

Multilevel Model Definition

- Multilevel Models capture (at least) two types of effects: Those who operate at the **cluster level** and those who operate **within the cluster**.
- ▷ The within-cluster level is typically referred to as **Level 1**
  - The Level 1 is the most granular level. E.g., individual student scores on a test
  - Those scores might be predicted by student level variables (student SES)
- ▷ In a two-level model, the cluster level is referred to as **Level 2**
  - The Level 2 captures the hierarchically next higher level
  - This could be *school* in which the students are nested.
  - We might have school-level predictors that influence how a school performs (eg. funding for the whole school)

Multilevel Model Definition

- Multilevel models capture the influence of different variables at different levels and yield:
- **Fixed Effects**:
  - Describe average effects across all clusters
  - e.g. In the typing speed example, pc captured the effect of typing proficiency across all individuals
  - e.g. pmc captured the average effect of typing speed across all individuals
- **Random Effects**:
  - These are individual departures from the mean, fixed effects
  - Expressed as variances (similar idea as in ANOVA)
  - e.g. The random intercept variance of 1095.8 in the typing speed example indicates that these individuals were spread out substantially around the mean of 620.

Multilevel Model Definition

Formal Definition

- *Notation*: Fixed effects are denoted using Greek letters, random effects are denoted using Latin letters
- A linear multilevel model can be defined as :

Level 1: $y_{ij} = b_{0i} + b_{1i}x_{ij} + e_{ij}$

- y_{ij} is the outcome variable of cluster i and individual j
- b_{0i} is the cluster-specific or random intercept of cluster i
- b_{1i} is the cluster-specific or random slope of cluster i
- x_{ij} level 1 predictor (e.g. student SES)

Multilevel Model Definition

Formal Definition

- At level 2, we have two models, one for the intercept and one for the slope

  $$\text{Level 2: } \begin{aligned} b_{0i} &= \beta_0 + u_{0i} \\ b_{1i} &= \beta_1 + u_{1i} \end{aligned}$$

  - β_0 is the fixed intercept
  - u_{i0} is the individual departure from the fixed *intercept* of cluster i
  - β_1 is the fixed slope
  - u_{i1} is the individual departure from the fixed *slope* of cluster i

- Both levels:

$$\begin{aligned} \text{Level 1: } y_{ij} =& b_{i0} + b_{i1}x_{ij} + e_{ij} \\ \text{Level 2: } b_{0i} =& \beta_0 + u_{0i} \\ b_{i1} =& \beta_1 + u_{i1} \end{aligned}$$

Multilevel Model Definition

Formal Definition

- At level 2, we have two models, one for the intercept and one for the slope

  Level 2: $$\begin{aligned} b_{0i} &= \beta_{00} + \beta_{01} w_i + u_{0i} \\ b_{1i} &= \beta_{10} + \beta_{11} z_i + u_{1i} \end{aligned}$$

  - β_0 is the fixed intercept
  - u_{i0} is the individual departure from the fixed *intercept* of cluster i
  - β_1 is the fixed slope
  - u_{i1} is the individual departure from the fixed *slope* of cluster i
  - β_{01} effect of predictor w_i on intercept
  - β_{11} effect of predictor z_i on intercept
  - w_i and z_i could be the same

- Both levels:

$$\begin{aligned} \text{Level 1: } y_{ij} =& b_{i0} + b_{i1} x_{ij} + e_{ij} \\ \text{Level 2: } b_{0i} =& \beta_{00} + \beta_{01} w_i + u_{0i} \\ b_{i1} =& \beta_{10} + \beta_{11} z_i + u_{1i} \end{aligned}$$

Multilevel Model Definition: Random Effects

- The level 1 and level 2 models may be combined to a complete model, namely
  $y_{ij} = \beta_{00} + \beta_{01} w_i + u_{0i} + (\beta_{10} + \beta_{11} z_i + u_{1i}) x_{ij} + e_{ij}$

- There are two sources of variation:
  - Residual variance (within-person): $e_{ij} \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right)$
  - Random effects variance (between-person): $\mathbf{u} = \begin{bmatrix} u_0 \\ u_1 \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \Psi\right)$
- Residuals are "drawn" from a normal distribution with a fixed residual variance σ_ϵ^2
- Random effects are also from a (multivariate) normal distribution: Ψ encodes random effects variances and covariances:
  $\mathbf{\Psi} = \begin{bmatrix} \sigma_{int}^2 & \sigma_{int,slp} \\ \sigma_{int,slp} & \sigma_{slp}^2 \end{bmatrix}$

Multilevel Model Definition

Formal Definition

- Note: Because random effects are deviations from fixed effects, individual trajectories are of the same functional form, e.g., linear, just as the mean trajectory
- If the mean trajectory is modeled using two parameters (initial level, linear slope), random effects may exist in initial level and linear slope
  - ▷ Even though random effects can provide individual Level-1 estimates, they assume that every cluster follows the same functional form

Multilevel Model / Linear Mixed Models

- What are linear *mixed* models?

Synonyms: *hierarchical* linear models
linear *multilevel* models
linear *variance components* models

- Mixed models contain both *fixed* and *random effects*
  - ▷ *mixed* models
- Hierarchical and multilevel
- Variance components refers to the fact that residual variance is split in systematic and error variance, i.e., different components

Multilevel Model / Linear Mixed Models

- Why are these models linear?
  - Because predictor variables enter linearly
  - Thus, the model is linear in its parameter estimates, i.e., the associations between predictor variables and the outcome variable are considered to be linear
  - Note: Like in OLS regression, one may still model polynomial associations (e.g., quadratic, cubic) by raising the predictor variable to the according power
- Can be expanded to intrinsically nonlinear models

Linear Mixed Models: Random Effects

Formal Definition

- More general mixed effects notation (Laird-Ware form):

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{u}_i + \boldsymbol{\epsilon}_i$$
$$\boldsymbol{u}_i \sim N(\mathbf{0}, \boldsymbol{\psi})$$
$$\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2 \boldsymbol{\Omega})$$

  $\sigma^2 \boldsymbol{\Omega}$ is also referred to as the **R**-Matrix.
  $\boldsymbol{\Omega}$ may take different forms

- Of interest are most often not the individual parameters, but rather whether
  - Random effects show statistically significant variance across clusters
  - Significant variance implies that, e.g., schools differ reliably in initial level and in the amount of linear slope, i.e., develop differentially

Multilevel Models

- Advantages of *multilevel models*
  - They solve the "unit of analysis"-problem

Example: In comparing residential homes, observations are based on individuals (level 1), but interest may lie in differences between residential homes (level 2)

  - Allow for simultaneous analyses at different levels

Example: From a behavioral genetics perspective, it seems interesting to study both individual (level 1) and family (level 2) differences in intelligence

  - Inclusion of explanatory variables at different levels

Example: Language achievement may be explained by both the composition of classes (e.g., number of students, percentage foreigners at level 2) and individual variables (e.g., foreigner?)

  - Regularization via shrinkage

▷ Property of estimation which takes into account reliability of estimates within a given group/individual

Hierarchical Data

Two types of examples:

- Classic multilevel models with educational data

- Multilevel models for longitudinal research
  + Some thoughts about longitudinal modeling

- Multilevel models are not restricted to these two areas
  - Whenever we have clustering!
  - Could also be in linguistics, with word-types or in neuroscience with homonymous brain regions etc.

Highschool and Beyond (HSB) Example

- See Rmd example on Canvas

Longitudinal Data

- Effectiveness in studying change

Example: Development of intelligence

- Cross-sectional model:

$$y_i = \beta_C x_i + \epsilon_i$$

where y_i is the intelligence score of individual i
x_i is the age of individual i
β_C is the regression of intelligence on age ϵ_i is a residual

- Then, β_C represents the difference in average y across two sub-populations which differ by 1 year of age

Longitudinal Data

- *Longitudinal model:*

$$y_{ij} = \beta_C x_{i1} + \beta_L (x_{ij} - x_{i1}) + \epsilon_{ij}$$

where y_{ij} is the intelligence score of individual *i* at age j
x_{i1} is the age of individual *i* at first measurement occasion
β_C is the cross-sectional regression of intelligence on age
x_{ij} is the age of individual *i* at *j*th measurement occasion
β_L is the regression of intelligence on age changes
ϵ_{ij} is a residual

- Then, β_L represents the average longitudinal change in *y* across individuals who changed by 1 year of age

Longitudinal Data

- To estimate how individuals change with time from cross-sectional data, we must make the strong assumption that $\beta_C = \beta_L$
- With longitudinal data, this strong assumption is unnecessary, because both β_C and β_L can be estimated
- Even if $\beta_C = \beta_L$, longitudinal data tend to be statistically more powerful
  - The basis of inference about β_C is a comparison of individuals with a particular value of age to others with a different age.
  - By contrast, β_L is estimated by comparing a person's intelligence at two (or more) times.
  - Longitudinally, each person can be thought of as serving as her own control, thus canceling out the influence of unmeasured characteristics in estimating β_L, whereas they tend to obscure the estimation of β_C

Longitudinal Data

- Distinguish the degree of variation in y across time for one person from the variation in y among people
- With *cross-sectional* data, the estimate of one person must draw on data from other individuals: individual differences are ignored and subsumed with error
- With *repeated measures*, strength can be borrowed from observations across time for one person and for other persons
- If there are little individual differences, individual estimates may also rely on data from others. Else, we might prefer to use only data for a specific person.

Longitudinal Data

- In practice, longitudinal data are oftentimes *highly unbalanced*:

▷ An equal number of measurements is not available for all subjects (drop-out, attrition)

▷ Measurements are not taken at the same, fixed time points

▷ Traditional analysis approaches (ANOVA) rely on balanced data: Attrition/disbalance leads to a great loss of information

▷ Mixed effects models don't have this problem: Use of all available information (assuming attrition is random)

Longitudinal Hierarchical Models

Example: The BOLSA Data

Bonn Longitudinal Study on Aging (BOLSA)

- Seven measurement occasions with decreasing sample size

Table: BOLSA Data

	1965	1967	1969	1972	1976	1980	1984
N=	221	188	158	127	80	46	30

- Two cohorts:
  - 113 individuals aged 63.3 years at T1
  - 108 individuals aged 72.4 years at T1
- Focus here: WAIS subtests Digit Symbol (DS; only at T1 – T6) and Block Design (BD)

Longitudinal Hierarchical Models

Example: The BOLSA Data

- Data structure: "Wide" vs. "Long"
- Hierarchical models often require a data-transformation
- From wide:

```
> head(bolsa)
VPNR KOHORTE SCHULE NOBS DS1 DS2 DS3 DS4 DS5 DS6 BD1 BD2 BD3 BD4 BD5 BD6 BD7 V1
1601       0      1    4  41  42  39  39  NA  NA  35  33  31  32  NA  NA  NA  0
1602       0      1    5  37  37  33  32  30  NA  28  29  24  23  26  NA  NA  0
1603       0      1    5  40  39  37  33  31  NA  23  24  23  23  18  NA  NA  0
1604       0      1    4  29  33  29  30  NA  NA  18  17  20  19  NA  NA  NA  0
1605       0      2    4  32  32  32  33  NA  NA  26  23  26  24  NA  NA  NA  0
1606       0      2    4  28  28  27  31  NA  NA  24  22  20  21  NA  NA  NA  0
```

Longitudinal Hierarchical Models

Example: The BOLSA Data

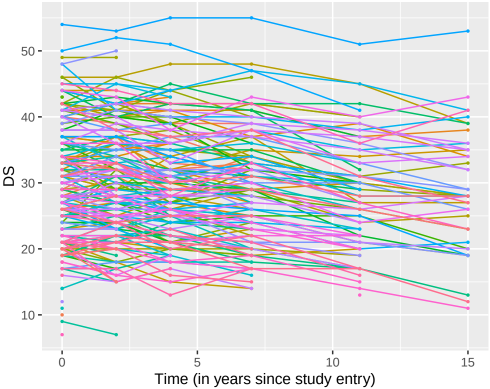
- Data structure: "Wide" vs. "Long"
- Hierarchical models often require a data-transformation
- From wide to long:

```
> head(bolsa.univ, n=20)
    VPNR KOHORTE SCHULE NOBS trial ds bd time
1   1601       0      1    4     0 41 35    0
1.1 1601       0      1    4     1 42 33    2
1.2 1601       0      1    4     2 39 31    4
1.3 1601       0      1    4     3 39 32    7
1.4 1601       0      1    4     4 NA NA   11
1.5 1601       0      1    4     5 NA NA   15
2   1602       0      1    5     0 37 28    0
2.1 1602       0      1    5     1 37 29    2
2.2 1602       0      1    5     2 33 24    4
2.3 1602       0      1    5     3 32 23    7
2.4 1602       0      1    5     4 30 26   11
2.5 1602       0      1    5     5 NA NA   15
3   1603       0      1    5     0 40 23    0
3.1 1603       0      1    5     1 39 24    2
3.2 1603       0      1    5     2 37 23    4
3.3 1603       0      1    5     3 33 23    7
3.4 1603       0      1    5     4 31 18   11
3.5 1603       0      1    5     5 NA NA   15
4   1604       0      1    4     0 29 18    0
4.1 1604       0      1    4     1 33 17    2
```

Longitudinal Hierarchical Models

Example: The BOLSA Data

- Individual trajectories in **DS** ("Spaghetti Plot")

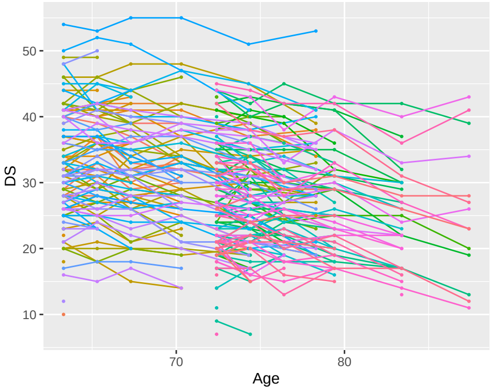


- Considerable heterogeneity (both in initial level and slope)
- Trajectories appear to be roughly linear
- On average, DS performance seems to decline slightly

Longitudinal Hierarchical Models

Example: The BOLSA Data

- Individual trajectories in **DS** ("Spaghetti Plot")

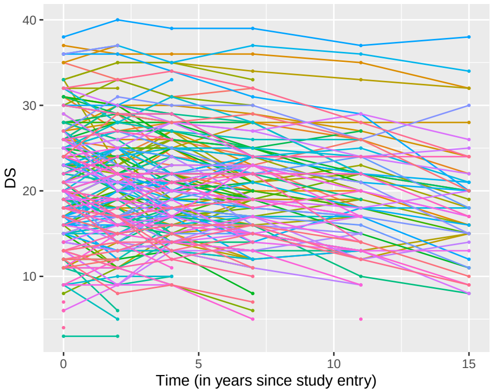


- Considerable heterogeneity (both in initial level and slope)
- Trajectories appear to be roughly linear
- On average, DS performance seems to decline slightly

Longitudinal Hierarchical Models

Example: The BOLSA Data

- Individual trajectories in **BD** ("Spaghetti Plot")

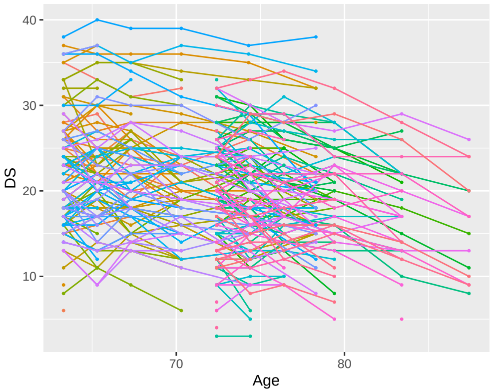


- Considerable heterogeneity (both in initial level and slope)
- Trajectories appear to be roughly linear
- On average, BD performance seems to decline slightly

Longitudinal Hierarchical Models

Example: The BOLSA Data

- Individual trajectories in **BD** (“Spaghetti Plot”)



- Considerable heterogeneity (both in initial level and slope)
- Trajectories appear to be roughly linear
- On average, BD performance seems to decline slightly

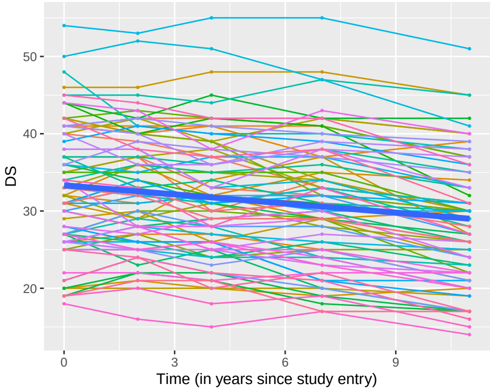
Longitudinal Hierarchical Models

Example: The BOLSA Data

- Focus is on change in DS for those subjects with complete data for the first five measurement occasions ($N = 80$, longitudinal time period: 11 years)
  - ▷ Note that the outcome variable is continuous
- We want to model change in one variable (univariate case), not in several variables simultaneously (multivariate case)
  - ▷ In the following examples we will only use complete cases

Linear Mixed Models: Fixed Effects

- Individual trajectories in DS across 11 years ($N = 81$)



Linear mixed model mean trajectory:
$DS = 33.33 - 0.39 \times Time$

Linear Mixed Models: Fixed Effects

- Statistical analyses:
  - OLS Regression (ignoring data dependency):
    $DS = 33.33 - 0.39 \times Time$
  - Mixed model (taking data dependency into account):
    $DS = 33.33 - 0.39 \times Time$
  - In this example, (fixed) parameter estimates are the same, why?
  - In general, if and only if the data are balanced, estimation of (fixed) parameters using OLS regression or mixed models will give identical results

Linear Mixed Models: Fixed Effects

- Fixed parameter estimates for the BOLSA data are the same, but what about the standard errors and statistical significance?

	OLS	Mixed
Intercept	33.33	33.33
$S.E.$	0.64	0.89
t	51.97	37.20
Slope	-0.39	-0.39
$S.E.$	0.10	0.03
t	-3.78	-12.47

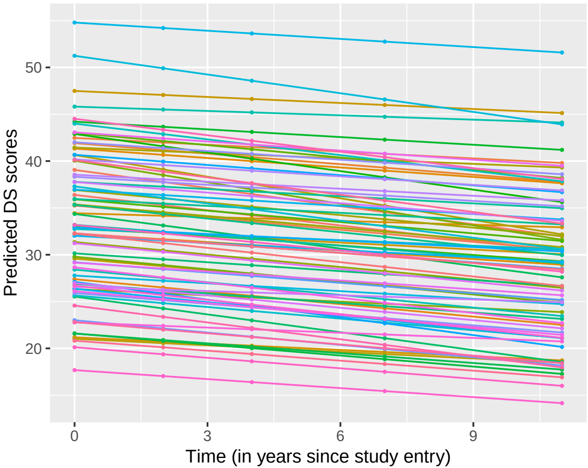
1. Considerable differences in standard errors
2. Considerable differences in t values

Linear Mixed Models: Fixed Effects

- Why are the standard errors different and with that the results of statistical significance tests?
  - Like with repeated measures ANOVA, from longitudinal data strength may be borrowed because individuals function, in a sense, as their own controls
  - Hence, like in repeated measures ANOVA, in mixed models the mean trajectory reflects the mean of individual trajectories, and not of unrelated observations (data points)
  - However, whereas in repeated measures ANOVA a saturated model is estimated, i.e., a parameter for each mean, in linear mixed models mean changes are related linearly to time (more parsimonious)

Linear Mixed Models: Random Effects

- Plot of model-based predicted DS Scores (N = 80)



- Individual differences in initial level
- Individual differences in linear slope

Linear Mixed Models: Random Effects

- Results of random effects estimation

Parameter	Estimate	S.E.	Z	*p*
Intercept Variance	62.75	10.22	6.14	$<.01$
Slope Variance	0.04	0.01	3.13	$<.01$
Error Variance	2.88	0.26	10.95	$<.01$

- Both the intercept and especially the slope variance are statistically significant
- This implies that there are reliable individual differences in initial DS level and the amount of linear change in DS across time

Covariances between random effects

- Random effects are, in essence, individual differences variables
- We might be interested in the association between random effects, in this case:
    - Is there a relation between initial level and change?
    - ▷ Is the initial level related to change (very common in aging research)
- Individuals starting out at a high performance level change more/less than those who start out with a lower performance level
- For the BOLSA DS data, a correlation of $-.13$ (not significant) between initial level and linear change is estimated.

Covariances between random effects

- A "problem" in longitudinal linear mixed models is that the covariance between intercept and slope depends on how intercept is defined
- Thus, the covariance is different if level is defined as initial level, intermediate level, or end level
  - ▷ Centering!

  - ▷ Watch out when interpreting the covariance
- Also, often covariance can be influenced by ceiling or floor effects
  - ▷ Specifically for repeated measures: E.g. learning maxes out scale, or sometimes it is hard to measure decline in those who already start out with a very low performance

Explanatory Variables

- Once we have found significant fixed and/or random effects, the next logical step is to try to "explain" these effects
- For the BOLSA DS data, we might ask, e.g., whether cross-sectional age differences can explain fixed and random effects in initial level and slope
- Note that continuous explanatory variables have to be grand mean centered in order to keep the fixed effects interpretable
- But since age cohort is a dichtomy ($0 =$ younger cohort, $1 =$ older cohort), 0 is a meaningful value, since then the fixed effects represent effects for the younger cohort

Explanatory Variables

- Results with cohort as an explanatory variable

	Without cohort	With cohort
Intercept	33.33*	34.98*
Slope	-0.39*	-0.33*
Cohort	–	-3.88*
Slope×Cohort	–	-0.15*
Var(Level)	62.75*	60.22
Var(Slope)	0.04*	0.04*
Corr(LS)	-0.13	-0.23

Assessment of Model Fit

- Linear mixed modeling often involves trying several different models for the data at hand an then decide which model fits the data best
  - ▷ Comparisons can be part of a "competition" among theories
  - ▷ Comparisons could be result of exploratory analyses
- For linear mixed models, two models may be compared
  - According to their -2×log-likelihood via Likelihood-Ratio tests
  - According to Akaikes Information Criterion (AIC)
- Both indices are **relative**, not absolute, i.e., they make sense only in the comparison of two (or more) models
- For both indices, lower values indicate better model fit AIC penalizes parametrization and encourages parsimony

Model Fit and Model Comparison

- **Likelihood Ratio tests** based on $-2\times$log-likelihood is useful in two models that are **nested**, i.e., which are the same with respect to the involved effects, but different in the number of estimated parameters (e.g., by estimating one covariance between random effects more). The difference in $-2\times$log-likelihood may then be tested for statistical significance using a χ^2-test
- **AIC** (as well as other information criteria such as BIC) can also be used to compare non-nested models as long as dependent **data is exactly the same** (same individuals; same N)
  - ▷ No clear-cut thresholds in deciding which model fits better, except for smaller AIC is better.
- **Alternative**: Comparison via **cross-validation** that allows comparing non-nested models across different model types
  - ▷ In contrast to SEM: Emphasis is on *relative* fit while SEM focuses on *absolute* fit

Model Comparison

BOLSA Example

- We can compare the two BOLSA models: $\mathcal{M}_{\text{Without Cohort}}$ vs. $\mathcal{M}_{\text{With Cohort}}$
- AIC:

```
> AIC(fitWC,  fitWO )
df        AIC
fitWC   8 1996.382
fitWO   6 2004.352
```

- Likelihood Ratio Test:

```
> anova( fitWC,  fitWO )
refitting model(s) with ML (instead of REML)
Data: comp_filtered
Models:
fitWO: ds ~ time + (1 + time | VPNR)
fitWC: ds ~ KOHORTE * time + (1 + time | VPNR)
      npar    AIC    BIC   logLik deviance Chisq Df Pr(>Chisq)
fitWO    6 2000.8 2024.8 -994.43   1988.8
fitWC    8 1991.9 2023.8 -987.95   1975.9 12.95  2   0.001541 **
```