Multivariate: Linear Mixed Models

Lecture 1: Conceptual Introduction

Tom Booth

Lecture 1

Welcome

Overview

- Week 1: Linear model recap and conceptual introduction
- Week 2: Estimating simple LMM
- Week 3: Model evaluation and building
- Week 4: Complex designs, longitudinal analyses and power
- Week 5: Generalized models and catch up

Reading

- There is a reading list on LEARN.
- You will be able to complete the assessment for LMM using the materials from lecture and labs.
- The reading is there if you want to:
  1. get a deeper understanding
  2. push for really top marks

Labs

- First couple of weeks materials are on LEARN
- We will start with data organisation and plotting
- We will then build and evaluate models
- Last couple of weeks we will look at some more complex examples, and give a more open task
  - i.e., to mimic report
- Lab answer sheets will be posted on a week lag.

Assessment

- You will need to produce a write up and code based on the methods in the course.
- More details in a couple of weeks time.
  - We need to actually do some LMM first!

R Notebooks or PDFs

- Your lecture material is provided as both a PDF slide deck, and an R Notebook.
- The R Notebook is on LEARN in HTML format.
    - Open as an HTML, you can follow the lecture, and choose whether to look at the underpinning code (e.g., the code that makes the plots) in the document.
    - You can also download the code and live edit in R, add your own notes, tweak and run the code etc.

Today

Topics

- Recap on the linear model
- Recap on LM assumptions
- Violations of assumptions due to structured data
- Overview of mixed models
- Some notes on data
- Conceptual examples (*class task, answers not in this slide deck*)
- Some more model details (*Bridge to week 2, time dependent*)

Linear Model (LM)

LM with single predictor

- Hopefully we are all familiar with:

$$y_i = \beta_0 + \beta_1 x_{1i} + \epsilon_i$$

- y_i is a vector of values for the outcome variable y for each observation
- x_{1i} is a vector of values for the predictor variable x for each observation
- β_0 is the intercept
- β_1 is the slope for the predictor x_1
- ϵ_i is a vector of individual error terms for each observation

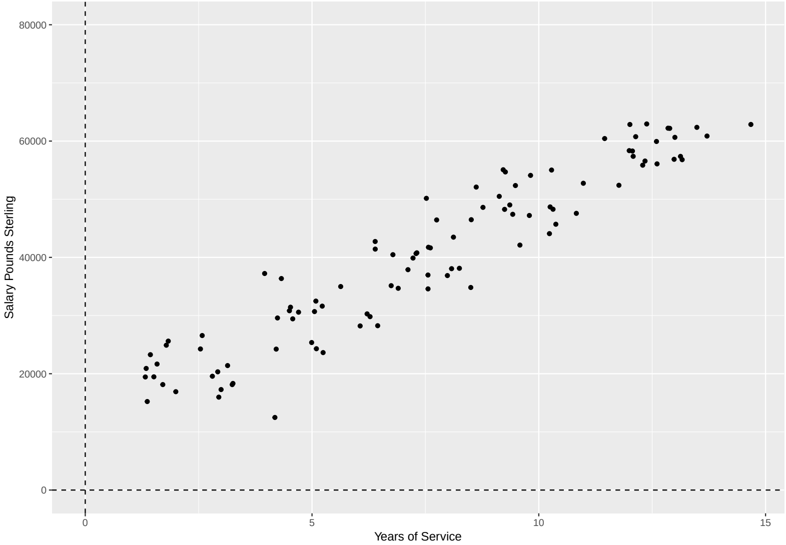
LM with a single predictor

- The intercept and slope are called coefficients or parameters.
- Together define a model line.
  - If we have multiple predictors, the coefficients define a surface.
- This provides an easy way to visualise a simple linear model.

Example

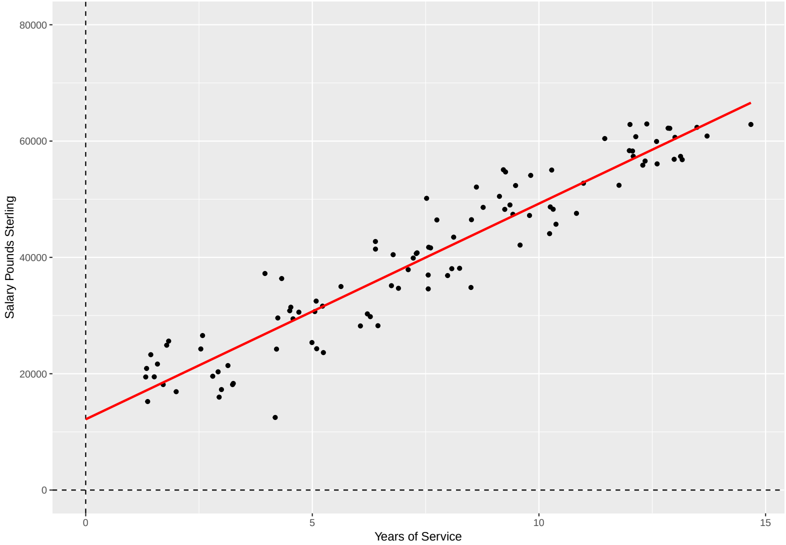
- Suppose we were looking to predict salary (y) from years of service (x) in an organisation.
- Let's look at some data:

Example



80000
60000
40000
20000
0
Salary Pounds Sterling
0
5
10
15
Years of Service

Fitting the linear model



Intercept β_0

- The intercept is the value of the outcome (y, here `Salary`), when the predictor (x, here `Service`) $= 0$
- In this example, the intercept is:

```
sal_mod <- lm(salary ~ service, data = job_df)
round(sal_mod$coefficients[1],0)
```

```
## (Intercept)
##       12161
```

- The salary for someone with 0 years experience is £12,161

Slope β_1

- The slope of the line is the change in the outcome (y, here `Salary`), per unit change in the predictor (x, here `Service`).
  - `Service` is measured in years, so...

```
sal_mod <- lm(salary ~ service, data = job_df)
round(sal_mod$coefficients[2],0)
```

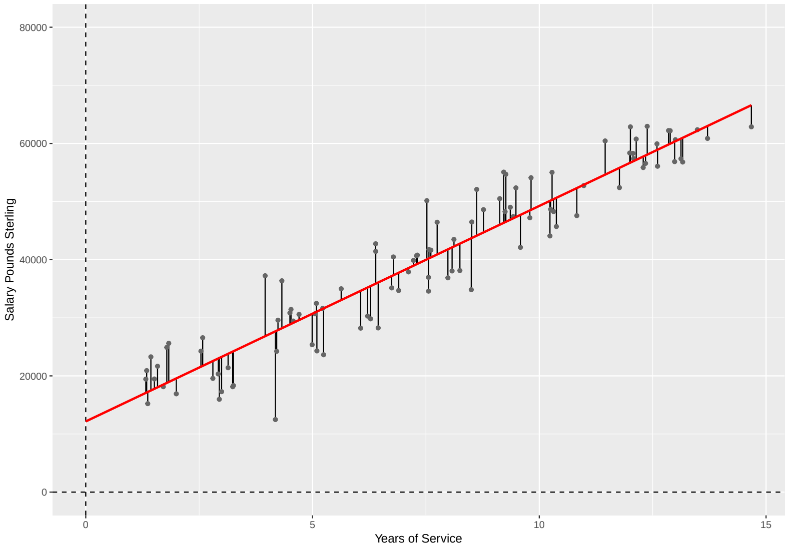
```
## service
##    3709
```

- For each year of service, salary increases by £3,709.

Finding the best fit linear model

- The best fit linear model for our data will be the line that minimizes the distance of each point to the line.
  - That is, the model that has the smallest error, or residuals
  - This is the **least squares** solution
- *An aside for later, if all assumptions are met, the least squared solution is the maximum likelihood solution. This we will return to when we discuss model estimation,*

Finding the best fit linear model



Fitted values

- The residual term (ϵ_i) for any individual is the difference between their observed y and the model fitted value (or predicted value) $\hat{y}$

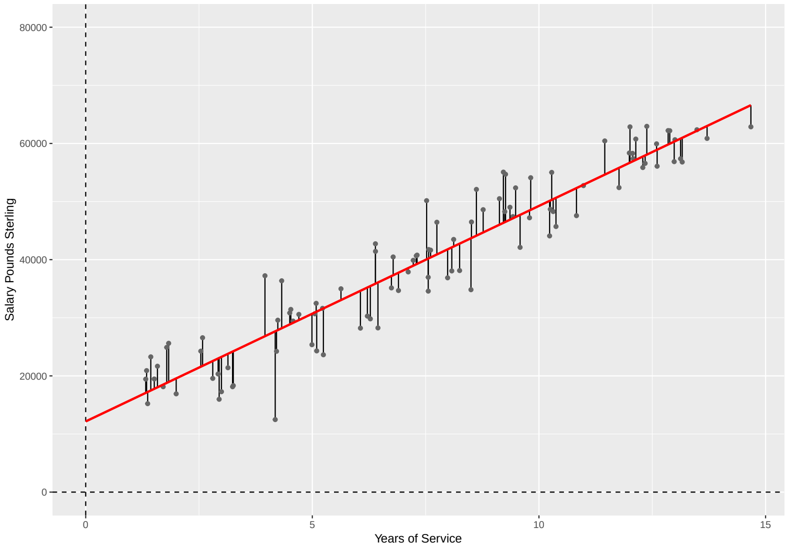
```
job_df$salary[1] - fitted(sal_mod)[1]
```

```
##        1
## 4841.989
```

```
residuals(sal_mod)[1]
```

```
##        1
## 4841.989
```

Fitted values (red line)



Things to keep in mind

1. β_0 and β_1 are single constant values.
    - They are the same for all observations in our data.
    - *Question*: When might β_1 not be constant?
2. Anything with a subscript i is at the level of the individual observation.
3. x_{1i} are individual values of the predictor, y_i the outcome.
4. There is an ϵ_i associated with every observation.

Assumptions & Data Structure

LM Assumptions

- LM have the following assumptions:
  - Linearity
  - Homoscedasticity (constant error variance)
  - Normality of residuals
  - Independence of residuals

Thinking about data structure

	Cross-Sectional	**Repeated Measures**	**Longitudinal**
Level 2	Classroom	Subject	Participant
Level 1	Pupil	Trial	Time

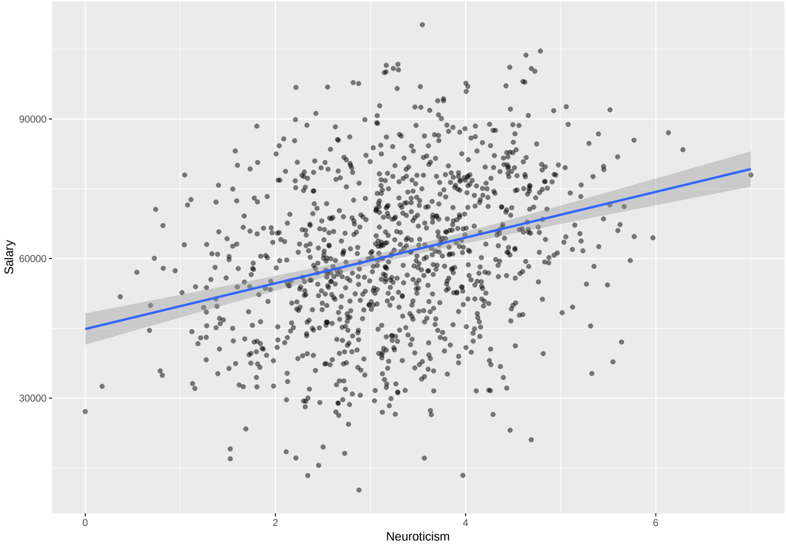
It can get more complicated

	Cross-Sectional	Repeated Measures	Longitudinal
Level 3	School	Condition	Family
Level 2	Classroom	Subject	Participant
Level 1	Pupil	Trial	Time

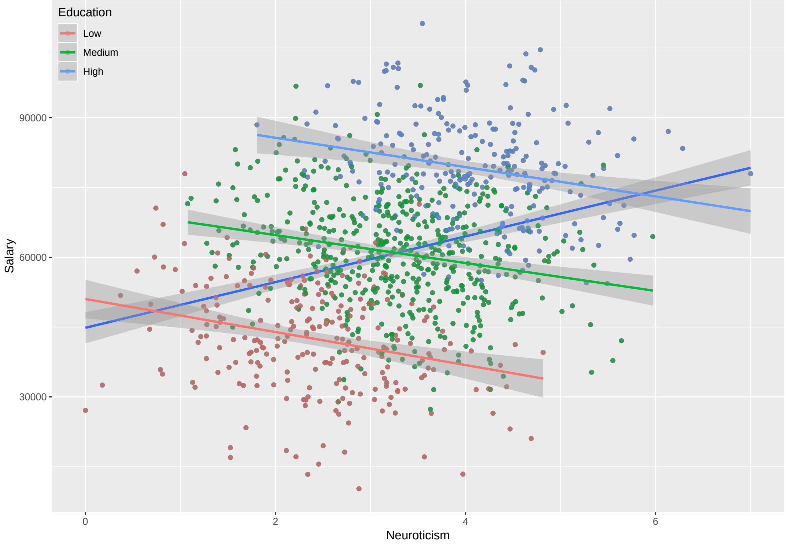
Why is clustering a problem?

1. Clustering of any of these types leads to data where the residuals will not be independent.
2. Non-independence leads to an underestimation of standard errors for model estimates.
3. In turn, this leads to an increased risk of Type I errors.
4. We miss the chance to ask interesting questions.
5. We may make errors in our conclusions.
   - **Ecological Fallacy**: Inferences about individuals deduced from inferences about groups.
   - **Atomistic Fallacy**: Inferences about groups deduced from inferences about individuals.

Simpson's Paradox (1)



Simpson's Paradox (2)



Imperfect solutions (we will come back to "pooling language")

- Ignore it!
    - Run a normal linear model.
    - Sometimes called *complete pooling* in that all responses are bundled together.
- Aggregation
    - Explicitly collapse levels to remove structure.
- Stratification
    - Run a separate model within each level cluster/grouping variable.
    - Sometimes called *no pooling*
- Better solution...

Linear mixed models (LMM)

Terminology

- Linear mixed models are referred to in a wide variety of ways:
  - General linear mixed-models
  - Random coefficient models
  - Hierarchical linear models (HLM)
  - Multi-level models (MLM)
  - Multi-level regression
  - Variance components models
- These are all pretty much the same thing.

Broad concept

- When we have clustered/structured data, we need to fit a model that takes this into account.
  - That models dependencies
- To do this, our model includes **fixed** and **random** effects.

Fixed vs. random effects

FIXED EFFECTS

Average estimates across the whole sample for model parameters.
Fixed = the same for all.

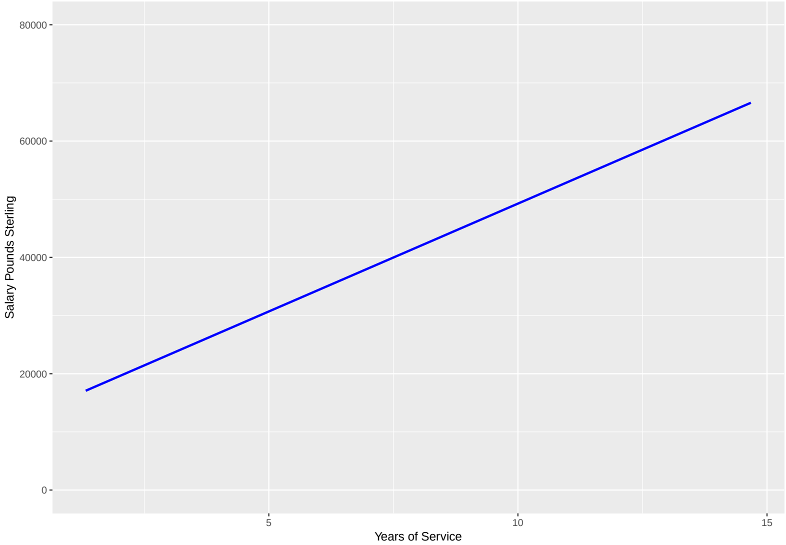
RANDOM EFFECTS

Variability around the average parameter (fixed effects) due to clustering.
Random = varying between clusters.

Salary example

- Suppose in our salary example we have employees measured within 5 different companies.
  - For simplicity in the following plots we plot a single employee from each company.
- What possible patterns do we think we could see?

LM: Fixed intercept and slope

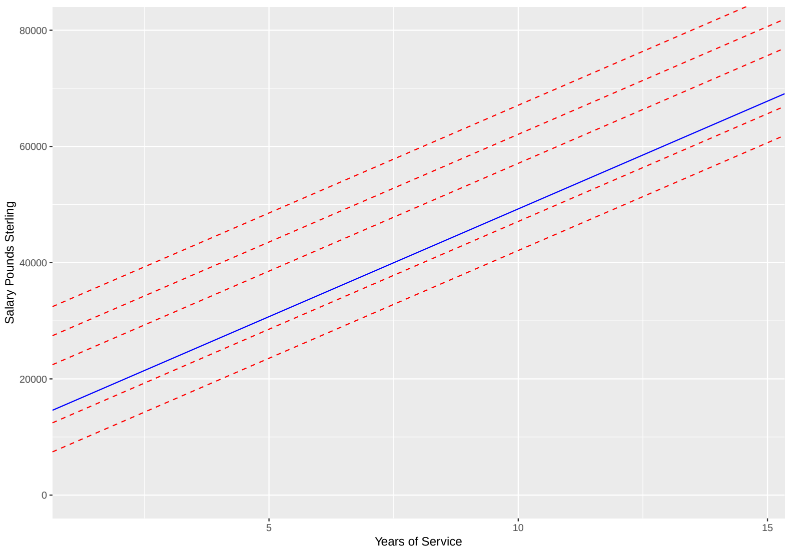


LM: Fixed intercept and slope

$$Salary_i = \beta_0 + \beta_1 Service_i + \epsilon_i$$

- Fixed intercept (β_0)
- Fixed slope (β_1)

Random (varying) intercept, fixed (same) slope



Random (varying) intercept, fixed (same) slope

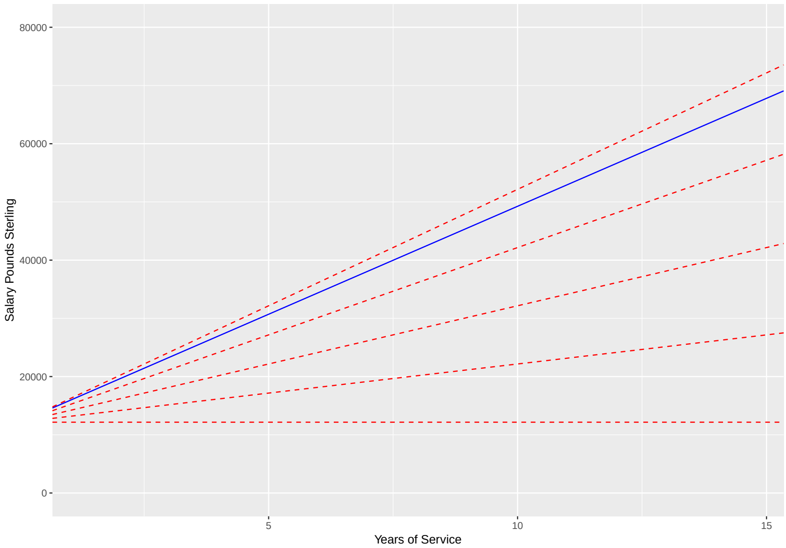
$$Salary_{ij} = \beta_{0j} + \beta_1 Service_i + \epsilon_{ij}$$

$$\beta_{0j} = \gamma_{00} + \upsilon_{0j}$$

- Random intercept (β_{0j}): Overall average (γ_{00}) and a residual term (υ_{0j})
    - Based on the residuals, we can estimate the variance (σ^2_{0j}) in the intercepts.
- Fixed slope (β_1)
- Presented as a single equation:

$$Salary_{ij} = \gamma_{00} + \beta_1 Service_i + \upsilon_{0j} + \epsilon_{ij}$$

Fixed (same) intercept, random (varying) slope



Fixed (same) intercept, random (varying) slope

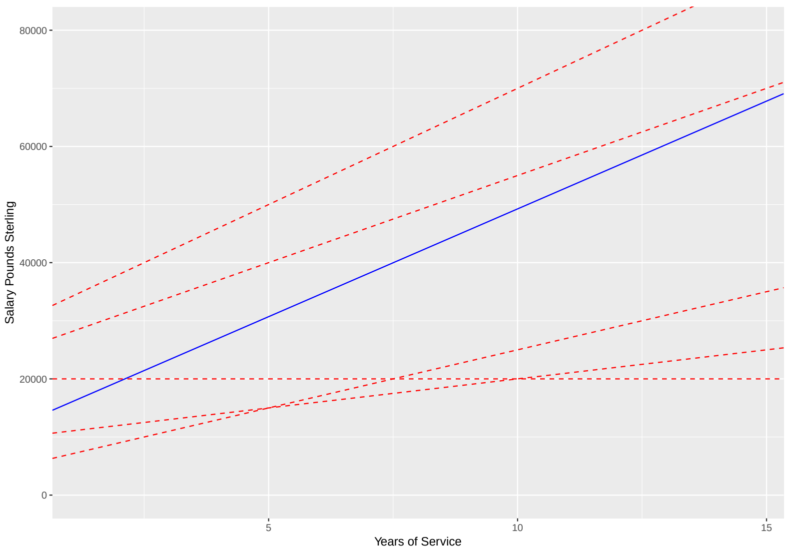
$$Salary_{ij} = \beta_0 + \beta_{1j} Service_{ij} + \epsilon_{ij}$$

$$\beta_{1j} = \gamma_{10} + \upsilon_{1j}$$

- Fixed intercept (β_0)
- Random slope (β_{1j}): Overall average slope (γ_{10}) and a residual term (υ_{1j})
- Presented as a single equation:

$$Salary_{ij} = \beta_0 + \gamma_{10} Service_{ij} + \upsilon_{1j} + \epsilon_{ij}$$

Random (varying) intercept, random (varying) slope



Random (varying) intercept, random (varying) slope

$$Salary_{ij} = \beta_{0j} + \beta_{1j} Service_{ij} + \epsilon_{ij}$$

$$\beta_{0j} = \gamma_{00} + \upsilon_{0j}$$

$$\beta_{1j} = \gamma_{10} + \upsilon_{1j}$$

- Random intercept (β_0): Overall average (γ_{00}) and a residual term (υ_{0j})
- Random slope (β_{1j}): Overall average slope (γ_{10}) and a residual term (υ_{1j})
- Presented as a single equation:

$$Salary_{ij} = \gamma_{00} + \gamma_{10} Service_{ij} + \upsilon_{0j} + \upsilon_{1j} + \epsilon_{ij}$$

Steps in LMM

- In lecture 3, we will consider in detail the approach to model building in LMM.
- Here I wanted to make brief reference to the overall process of analysis.
- Assuming we have already established our theory, developed our hypotheses, and constructed a suitable study (easy right!!!!!!), then we. . . .

1. Tidy our data.
2. Describe and visualize the data.
3. Build and test our model (lecture 3)
4. Interpret the model (all lectures!)

- Lab today is (1) and (2), so I wanted to take a few minutes to flag a couple of things.

A practical comment on data

Wide vs long format

- Who is aware of the distinction between wide and long data?

Example wide data

Table 3 : Wide Data 2 Participants

ID	Gender	Trial_1	Trial_2	Trial_3
ID1001	Male	10.0	12.5	18
ID1002	Female	7.5	10.0	5

- How many variables do we have?

Tidy data

- In labs we will be using the `tidyverse` set of packages for data manipulation and plotting.
- Tidy data has a number of basic principles (*described in lab*)
- The short version:
  - Each variable is a column.
  - Each observation forms a row.
- Think again about our structured data, and the wide format. . . .

Example wide data

Table 4 : Wide Data 2 Participants

ID	Gender	Trial_1	Trial_2	Trial_3
ID1001	Male	10.0	12.5	18
ID1002	Female	7.5	10.0	5

- **Variables**: Gender, Trial and Score
- **Observations**: Participants * n.trials = (6)

Example Long Data

Table 5 : Long Data 2 Participants

ID	Gender	Trial	Score
ID1001	Male	1	10.0
ID1002	Female	1	7.5
ID1001	Male	2	12.5
ID1002	Female	2	10.0
ID1001	Male	3	18.0
ID1002	Female	3	5.0

tidyverse for conversion

```
long <- wide %>%
  gather(key = Trial,
         value = Score,
         Trial_1:Trial_3) %>%
  separate(Trial, c("label", "Trial"), "_", convert = T) %>%
  select(-label)
```

Descriptive statistics by clustering variable

- When we have higher level units we can cluster by, it can be of interest to look at descriptive statistics at this level.
- So for example, if a participant completes multiple reaction time trials, we might want to know the mean and standard deviation per participant.
- We can do this simply in `tidyverse` using `group_by()` and `summarise()` functions.

Descriptive statistics by clustering variable

```
long %>%
  group_by(ID) %>%
  summarise(
    N.trials = n_distinct(Trial),
    mean_score = mean(Score),
    sd_score = sd(Score)
  )
```

Descriptive statistics by clustering variable

```
summary <-
  long %>%
  group_by(ID) %>%
  summarise(
    N.trials = n_distinct(Trial),
    mean_score = mean(Score),
    sd_score = sd(Score)
  )
summary
```

```
## # A tibble: 2 x 4
##   ID     N.trials mean_score sd_score
##   <chr>     <int>      <dbl>    <dbl>
## 1 ID1001        3       13.5     4.09
## 2 ID1002        3        7.5     2.5
```

Descriptive statistics by clustering variable

```r
kable(summary, "latex",
      caption = "Descriptive Statistics by ID")
```

Table 6 : Descriptive Statistics by ID

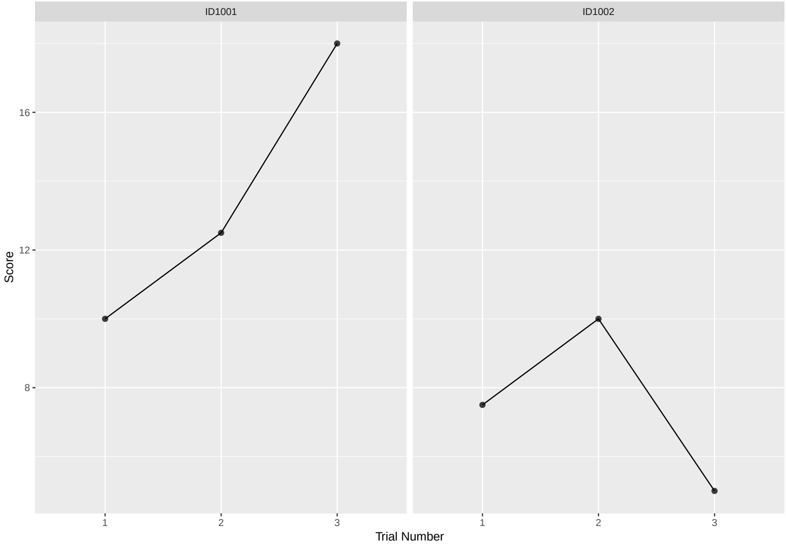
ID	N.trials	mean_score	sd_score
ID1001	3	13.5	4.092676
ID1002	3	7.5	2.500000

Plots by clustering variable

- As well as describing data, we may want to visualise data by clustering variable.
- Again this can achieved within the `tidyverse` using `ggplot2` and in particular, the `facet_wrap()` argument.

```
long %>%
  ggplot(aes(x=factor(Trial), y=Score, group = ID)) +
  geom_point(size = 2, alpha= .75) +
  geom_line() +
  labs(x="Trial Number", y="Score") +
  facet_wrap(~ID)
```

Plots by clustering variable



Study Examples: How well do we understand levels?

Example 1: Intervention study

- A research team conducts an intervention study on exercise. They want to know if total hours exercise increases whether someone uses an in gym personal trainer, or has one-to-one sessions from home.
- They randomly assign 100 people to each condition.
- They measure number of hours exercised one week pre, and two weeks post, a 3 week training intervention period.
- They also measure a set of demographics.

Questions

1. What is the outcome?
2. What are the predictors?
3. What aspect of the study creates dependent structure?
4. Due to (3) what are the potential random effects?

Example 2: Longitudinal study

- A research team is interested in change in aggressive behaviour across adolescents.
- They measure aggression using a questionnaire measure every year from age 7 to age 17.
- The children in the study come from different areas of the same city.

Questions

1. What is the outcome?
2. What are the predictors?
3. What aspect of the study creates dependent structure?
4. Due to (3) what are the potential random effects?

Example 3: Experimental study

- **Disclaimer** I am not a linguist, better example from *actual* linguistic researchers in future lectures.
- A research team is interested in the non-word reaction time in a lexical decision task.
- They are interested in whether real word neighbourhood density is predictive of RT.
- Conduct an experiment with 30 trials per participant.

Questions

1. What is the outcome?
2. What are the predictors?
3. What aspect of the study creates dependent structure?
4. Due to (3) what are the potential random effects?

A few more details on models

“Complex”Multi-level“” theories

- LMM allow us to construct multi-level theories.
- So can think about whether phenomena at one level predict outcomes at another...
  - e.g. does sector predict variation in starting salary?
  - **What might this look like?**

Random intercept with predictor

$$Salary_{ij} = \beta_{0j} + \beta_1 Service_i + \epsilon_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Sector_j + \upsilon_{0j}$$

- Our random intercept (β_{0j}) is now predicted by Sector, and the effect captured in the γ_{01} coefficient.
  - This is identical to a β coefficient.
- Presented as a single equation:

$$Salary_{ij} = \gamma_{00} + \beta_1 Service_i + +\gamma_{01} Sector + \upsilon_{0j} + \epsilon_{ij}$$

“Complex”Multi-level“” theories

- And these can get elaborate
  - e.g. does the effect of service on salary change dependent on the organisations market performance. -**What might this look like?**

Random slope, cross-level interaction

$$Salary_{ij} = \beta_0 + \beta_{1j} Service_{ij} + \epsilon_{ij}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Performance_j + \upsilon_{1j}$$

- The (γ_{11}) coefficient now expresses the degree to which the β_{1j} slope, or the effect of `Service` on `Salary` is dependent on organisational performance.
- Presented as a single equation:

$$Salary_{ij} = \beta_0 + \gamma_{10} Service_{ij} + \gamma_{11} Service_{ij} Performance_j + \upsilon_{1j} + \epsilon_{ij}$$

“Complex”Multi-level“” theories

- We can also consider the covariance between random effects.
  - Put another way, is there a positive covariance between intercept and slope variation.
  - Do people who have higher intercept values (higher starting salary), also have a faster increase in pay (more positive slope)
  - **What might this look like?**

Covarying random variances

- Consider the random intercepts, random slopes model:

$$Salary_{ij} = \gamma_{00} + \gamma_{10} Service_{ij} + \upsilon_{0j} + \upsilon_{1j} + \epsilon_{ij}$$

- υ_{0j} and υ_{1j} are assumed independent of ϵ_{ij}, but they are allowed to covary.
- Remember, from these we can estimate variances $\sigma^2_{\upsilon_0}$ and $\sigma^2_{\upsilon_1}$ respectively.
- The associated covariance matrix is generally **not** assumed to be zero.

$$\begin{array}{ll} \sigma^2_{\upsilon_{01}} & \\ \sigma^2_{\upsilon_{12}} & \sigma^2_{\upsilon_{02}} \end{array}$$

That's all for today