

# Analysing and Interpreting Psychological Data II

## - PSYC402

Margriet Groen and Rob Davies

2022-01-14



# Contents

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>Intro</b>   | <b>5</b> |
| 1.1      | Analysis labs and ‘pre-lab work’ . . . . .                                   | 5        |
| <b>2</b> | <b>Week 11: Recap of the linear model and practising data-wrangling in R</b> | <b>7</b> |
| 2.1      | Lectures . . . . .   | 7        |
| 2.2      | Reading . . . . .  | 8        |
| 2.3      | Pre-lab activities . . . . .   | 9        |



# Chapter 1

## Intro

This is a collection of tuition material written for Psychology postgraduates at Lancaster University. At the moment the content represents the “lab materials” for the PSYC402 module. They feature tuition of programming with R, building on the skills you developed last term.

### 1.1 Analysis labs and ‘pre-lab work’

Some parts should be completed before you attend the lab session (watching lectures, reading chapters, pre-lab activities). All the links to the different materials and activities are also in the ‘to-do list’ for the relevant week on Moodle.



## Chapter 2

# Week 11: Recap of the linear model and practising data-wrangling in R

Written by Margriet Groen (partly adapted from materials developed by the PsyTeachR team at the University of Glasgow)

Before we start covering new material, we want to spend some time on recapping the basic concepts of the linear model (correlation, simple regression, multiple regression). You all come from different educational backgrounds and therefore have vastly different knowledge of, and experience with statistics. Therefore, please follow your own judgement as to whether you feel you want to/need to revisit material outlining the theoretical background to and the practical implementation in R for these topics. Below we provide some guidance as to materials that are relevant. **Just to be clear: We don't expect you to watch and/or read and/or do everything, please have a look at what you feel you need and spend some time with those materials.**

### 2.1 Lectures

**The linear model was discussed in weeks 6 to 9 of PSYC401, so that is a good place to start.**

Alternatively, if you don't feel confident about the material, these recorded lectures might help.

- **The linear model: theory (~30 min)** An introduction to the linear model and linear regression. I follow material as discussed in Chapter 4

of Bodo Winter's book *Statistics for Linguists: An Introduction using R* (see below under 'Reading').

- **How to build a linear model in R (~30 min)** In this video I demonstrate how to build a linear model in R by talking you through a simple linear regression script (you can download it here [stats\\_linearModel\\_howTo.R](#)). If you are unclear on what different parts of the `lm()` function do, or how to read the output, this video might help clarify that.
- **Multiple regression: theory (~35 min)** An introduction to multiple regression. I follow material as discussed in Chapter 5 of Bodo Winter's book *Statistics for Linguists: An Introduction using R* (see below under 'Reading').
- **Centering and standardising (~5 min)** Brief explanation of what centering and standardising are.

## 2.2 Reading

### 2.2.1 Miller & Haden (2013)

#### Link

**Chapter 10** gives you a brief overview of what correlation and regression are. **Chapter 11** introduces correlation in more detail. **Chapters 12 and 14** provide accessible overviews of simple and multiple regression, respectively. All these chapters are really short but provide a good basis to understanding. We consider this the minimum level of understanding you should acquire.

### 2.2.2 Winter (2020)

#### Link

**Chapter 4** provides an excellent conceptual introduction to the linear model and also explains how this is implemented in R (highly recommended).

**Chapter 5** takes a slightly different approach to the one taken in Miller & Haden (2013) to introducing correlation. If you already understand the basic theory behind correlation, this will be an interesting read. Chapter 5 also clearly explains what centering and standardizing are and why you need to bother with these linear transformations.

**Chapter 6** provides an excellent overview of multiple regression and also explains how this is implemented in R.



## 2.3 Pre-lab activities

After having watched the lectures and read the textbook chapters you'll be in a good position to try these activities. Completing them before you attend your lab session will help you to consolidate your learning and help move through the lab activities more smoothly.

### 2.3.1 Pre-lab activity 1: Visualising the regression line

Have a look at **this visualisation of the regression line** by Ryan Safner.

In this shiny app, you see a randomly-generated set of data points (within specific parameters, to keep the graph scaled properly). You can choose a slope and intercept for the regression line by using the sliders. The graph also displays the residuals as dashed red lines. Moving the slope or the intercept too much causes the generated line to create much larger residuals. The shiny app also calculates the sum of squared errors (SSE) and the standard error of the regression (SER), which calculates the average size of the error (the red numbers). These numbers reflect how well the regression line fits the data, but you don't need to worry about those for now.

In the app he uses the equation  $Y = aX + b$  in which  $b$  is the intercept and  $a$  is the slope.

This is slightly different from the equation you saw during the lecture. There we talked about  $Y = b_0 + b_1X + e$ . Same equation, just different letters. So  $b_0$  in the lecture is equivalent to  $b$  in the app and  $b_1$  in the lecture is equivalent to  $a$  in the app.

Pre-lab activity questions:

1. Change the slider for the intercept. How does it change the regression line?
2. Change the slider for the slope. How does it change the regression line?
3. What happens to the residuals (the red dashed lines) when you change the slope and the intercept of the regression line?

### 2.3.2 Pre-lab activity 2: Data-wrangling in R

In PSYC401, you've already learned how to read in data, how to select variables and how to compute summary statistics, so re-visiting the PSYC401 materials is a good place to start.

RStudio also provides some useful interactive tutorials that take you through the basics:

- **The Basics** Start here to learn how to inspect, visualize, subset and transform your data, as well as how to run code.
- **Work with Data** Learn how to extract values from a table, subset tables, calculate summary statistics, and derive new variables.
- **Visualize Data** Learn how to use ggplot2 to make any type of plot with your data. The tutorials on **Exploratory Data Analysis** and **Scatter-plots** are particularly relevant.

Please note that there are often different ways to do the same or similar things in R. This means you might encounter slightly different functions or styles of coding in different materials. This is not something to worry about. Just make sure you're clear on what a bit of code achieves and choose the function/style that you feel most comfortable with.

### 2.3.3 Pre-lab activity 3: Getting ready for the lab class

#### 2.3.3.1 Remind yourself of how to access and work with the RStudio Server.

- **Video on how to access the RStudio Server by Padraic**
- I highly recommend using R Projects to structure your workflow. You could create an R project for each week of the module. Have a look at section **8 Workflow: projects of R for Data Science** by Hadley Wickam and Gareth Golemund for an introduction.

#### 2.3.3.2 Get your files ready

Files to download for the lab activities as well as the lab activities themselves will be added later this week.