R for Psychology Research

Week 8 - Exercises

Marcus Lindskog

1. Regression

- The file "Album_Sales.csv" contains a fictional data set (Field, 2017) with the following variables
  - **adverts -** Amount (in thousands of pounds) spent promoting the album before release.
  - **sales -** Sales (in thousands of copies) of each album in the week after release.
  - **airplay -** How many times songs from the album were played on a prominent national radio station in the week before release.
  - **attract -** How attractive people found the band's image (1 to 10).

  Field, A. P. (2017). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). London: Sage. [Fictional data set]

1. Calculate descriptiv statistics and zero order correlations between the four variables.
2. Create two different multiple linear regression models that both predict the **sales** variable, but with different predictors. Evaluate the models and model parameters.
3. Calculate standardized beta coefficients with 95% confidence intervals for the two models.
4. Compare the two models, which can better explain the data.
5. The data set `mtcars` is available in base R. Build a regression model that regresses the variables `cyl`, `disp`, and `hp` on `mpg`. What is the adjusted R-square of the model? aRe there any significant predictors?
6. Update (using `update()`) the model in 5 with a term that is the interaction between `drat` and `qsec`. What is the adjusted R-square of the model? aRe there any significant predictors? Does the new model provide a better fit to the data?
7. Calculate standardized beta coefficients with 95% confidence intervals for the model in 6.
8. Using the `mtcars` data set, build a regression model that regresses the variables `cyl`, `disp`, and `hp` on `am`. Note that `am` is a categorical variable. aRe there any significant predictors?

2. ANOVA

- The following code simulates data for a repeated measures design (you need to load `tidyverse`).

```
Long <- rnorm(40, 123, 12)
Medium <- rnorm(40, 138, 12)
Short<- rnorm(40, 140, 22)
ID <- 1:40

within_subjects_data <- tibble(ID = factor(ID), Length_Long = Long,
                               Length_Medium = Medium,
                               Length_Short = Short)
```

9. Transform the data set into a format that you can analyse using `aov()`.
10. Run a repeated measures ANOVA to determine the effect of Length on the response variable.

- The following code simulates data for a 2x2 between subjects design (you need to load `tidyverse`).

```
DV <- c(rnorm(40, 123, 12), rnorm(40, 138, 12))
IV1 <- rep(c("Long", "Short"), each = 40)
IV2 <- rep(rep(c("Wide", "Narrow"), each = 20), 2)
ID <- 1:80


two_by_two_design_data <- tibble(ID,DV, IV1, IV2)
```

11. Transform the data set into a format that you can analyse using `aov()`.
12. Run an ANOVA to determine the effect of IV1, IV2 and their interaction on DV.

3. Examination Exercises

The solution to the exercises in this section should be handed in as a part of the examination. Your solution should be contained in a single R-script that is emailed to `marcus.lindskog@psyk.uu.se`. Your code should be well commented and easy to follow. Answers to any questions below should be written as a comment in the R-script after the code that produces the answer.

- The following code simulates data for a between subjects design (you need to load `tidyverse`).

```
Resp_A <- rnorm(40, 10, 1)
Resp_B <- rnorm(40, 11, 1.5)
Resp_C <- rnorm(40, 15, 3)

between_subjects_data <- tibble(Response_A = Resp_A,
                                Response_B = Resp_B, Response_C = Resp_C)
```

1. Determine the effect of Condition on Response and run appropriate Post Hoc tests.
2. Use `ggplot2` to visualize the results.
3. Transform the F-table into a tibble that could be used for further processing.
4. Assume that assumptions for parametric analyses aRe **not** met and run an appropriate test determine the effect of Condition on Response.

- The file "Alcohol_Attitudes.csv" contains a fictional data set (Field, 2017) with the following variables in which participants indicated their attitudes towards the drink on a scale from -100 to 100.
  - **beerpos** - Attitude towards beer after watching positive imagery.
  - **beerneg** - Attitude towards beer after watching negative imagery.
  - **beerneut**- Attitude towards beer after watching neutral imagery.
  - **winepos** - Attitude towards wine after watching positive imagery.
  - **wineneg** - Attitude towards a wine after watching negative imagery.
  - **wineneut** - Attitude towards a wine after watching neutral imagery.
  - **waterpos** - Attitude towards a water after watching positive imagery.
  - **waterneg** - Attitude towards a water after watching negative imagery.
  - **waterneut** - Attitude towards a water after watching neutral imagery.
  - **participant** - Participant ID.

  Field, A. P. (2017). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). London: Sage. [Fictional data set]

5. Run a 3x3 repeated measures ANOVA on to test the effect of **Drink** and **Imagery** on attitudes. For the ANOVA you should use functions available in Base R.

6. Plot the data using `ggplot2` and run appropriate Post Hoc Tests to disentangle any significant effects. For this, other packages than those in Base R may be used.

- The file "Fat_Gain.csv" contains a data set that provides participants' non-deliberate exercise activities (NEA) and their fat gain after a period of 8 weeks of overeating. The variables aRe
  - **NEA -** Energy use (in calories) from non-deliberate exercise activities.
  - **Fat -** Fat gain (in kg) after 8 weeks of excessive calorie intake.

7. Use linear regression to determine if NEA reduces fat gain. Does NEA reduce fat gain?

8. Calculate standardized beta coefficients with 95% confidence intervals for the model in 7.