R for Psychology Research

Week 7 - Exercises

Marcus Lindskog

1. Descriptive statistics

For all the descriptive statistics exercises, please use the `bfi` data set in the `psych` package

1. Use functions from Base R to calculate the *variance*, *standard deviation*, *maximum*, *minimum*, *range*, *mean*, *quantiles* of the variables `E1` and `O1`.
2. Use functions from the `tidyverse` to create a data frame called `summary_statistics` that contain the same descriptive statistics as in 1.
3. Use functions from the `psych` package to calculate the same descriptive statistics as in 1 for all variables in the `bfi` data set.
4. Modify your script in 2 to calculate the same descriptive statistics but for each gender separately.
5. Modify your script in 3 to calculate the same descriptive statistics but for each education level separately.

2. Basic inferential statistics

6. Assign the `anorexia` data set from the `MASS` package to a data frame called `anorexia_data`. Inspect the structure of the data set.
7. Calculate descriptive statistics for the Pre- and Post-test variables and plot the data using `ggplot2`.
8. Perform a paired t-test on the Pre- and Post-test variables and interpret the results.
9. Assign the `mtcars` data set from Base R to a data frame called `car_data`. Inspect the structure of the data set.
10. Calculate descriptive statistics for the `mpg` variable for each of the two categories in the `am` variable separately.
11. Perform an independent t-test on `mpg` with `am` as group and interpret the results.
12. For your `car_data` calculate the correlation between car price and fuel economy (mpg).
13. Check whether the obtained coefficient is statistically significant at 1% level.
14. Calculate Spearman's correlation coefficient for the same variables, and determine if it is statistical significance.
15. Create a new data frame, `auto_num`, that contains only columns with numeric values from `car_data`. Use functions from `tidyverse` to accomplish this.
16. Use functions from the `psych` package to calculate all pairwise correlations in `auto_num`.
17. Use the `corrgram` function from the `corrgram` package to create a default correlogram to visualize correlations between variables in the `car_data` data frame. Spend some time customizing your correlogram to learn about the functionality of the `corrgram` function.
18. A research is interested in examining if there is an association between ownership of a mac computer and geographical region in the USA. The research collects the following data. Perform a χ^2 test on the data.

Region	Mac	No Mac
North East	12	14
South West	21	18
Mid West	17	18

21. BONUS EXERCISE: Visit http://guessthecorrelation.com and play. Are your intuitions about correlations correct?

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.

1. Write a script that simulates 1000 experiments in which two groups of 100 participants each have done an IQ-test under slightly different conditions. Compare the mean IQ-level between the two groups in each experiment and use `ggplot2` to plot the distribution of p-values from the 1000 experiments.
2. Write a script that simulates 1000 studies in which a sample of 150 participants are measured on two variables X and Y. You are free to choose the parameters of X and Y as you see fit, but they should be the same for all experiments. For each experiment, calculate Pearson's r, Kendall's τ, and Spearman's ρ and determine which of the three coefficients are statistically significant (use $p < .05$). Finally, use `ggplot2` to make a bar graph that that plots the proportion of experiments with a statistically significant outcome for each of the three coefficients.
3. A casino owner has reached out to you. She thinks that one of her poker machines might have been manipulated. She has collected data from the machine for 365 consecutive days. For each day she has calculated the number of Spades, Hearts, Diamonds, and Clubs the machine has drawn. If the machine has not been manipulated on a particular day, she expects the proportion of the four cards to be equal. Her data is available to you in the file `poker_data.csv`. Import the file and perform the appropriate statistical test for each day. Summarize your findings in a graph that plots the proportion of days that the machine might have been manipulated. Should the casino owner be worried?