Quantitative Methods

Computer Lab 1 - Descriptive Statistics

Work in small groups of 2 to 3 students to complete the exercises below. Feel free to ask each other for help and discuss as much as you want during the computer lab. Remember to reflect on what you are doing and what the output tells you. It's better to take your time than to rush through the exercises. If you get stuck, you are free to ask the teacher for help. You can use `JAMOVI` or `R` to complete all the exercises.

About the data

The mouse-deer, or the chevrotain, is a family of small, hoofed mammals. There are various types of mouse-deer, including the lesser mouse-deer, the greater mouse-deer and the Philippine mouse-deer. As such, one might be interested in investigating if these different types of mouse-deer are about the same size.

The dataset for this computer lab is a fictional and asub-dataset of mouse-deer data. The data file, ***mousedeer.csv***, can be found in the folder "Computer Labs" in Studium.

In the dataset you find two variables,

- **Species:** The type of mouse-deer the animal was.
- **Length:** The head-and-body length of the animal, measured in centimetres.

Exercises

Using the given dataset to answer the following questions:

1. Open the file and confirm that the data has been correctly reads.
2. Is the data you have access to a sample or a population? Discuss!
3. What is the measurement levels for the two variables in the data set and why does it matter what the measurement levels are?
4. Create a frequency table for the `Species` variable.

5. Look at the frequency table you created. What is the **mode** of the Species variable?

6. Make a bar plot of the Species variable. Compare the plot to the frequency table you previously created.

7. Make a boxplot for the length variable for each of the three types of mouse-deer.

8. Look at the boxplot you created. Does there appear to be any difference in the distribution of length of the mouse-deer between the three kinds?

9. Are there any outliers in the data? If so, what was the outlier?

10. Looking at the boxplot, approximately what was the median length of a lesser mouse-deer?

11. The 25% longest greater mouse-deer in the data, above which length were they?

12. Calculate the mean, median and standard deviation of the Length variable. Interpret those numbers.

13. Create a histogram of the Length variable and change the title of the histogram to something reasonable of your choice.

14. Also change the label of the X-axis to "Length (cm)".

15. Describe the shape of the distribution illustrated by the histogram. Is it roughly symmetrical or skewed? Is it unimodal, bimodal or multimodal?

16. Add a variable called `LengthInches` to the dataset. The new variable should be calculated as $0.393701 \times Length$.

17. Create a new data set called mousedata.long containing only the mouse-deer that are longer than the median length of mouse-deer in the data. Save the new dataset.