

ST344: Professional Practice of Data Analysis

Prof Jane Hutton, Prof Bärbel Finkenstädt Rand, Claudia Viaro

Lab Report 1: data and questions

Academic year 2023-2024

Exploratory or initial data analysis is very important in professional practice. However, it is often regarded as either trivial or routine. This practical aims to check that you can use appropriate summary statistics and sensible graphs or tables, and can think about data collection issues. It also allows you to think about the format of a very brief report, and make your own choice. You do not have to use R-markdown.

Discussion of different approaches with other people can be very useful. However, please remember to write your own report. I expect much of your discussions will not be relevant to your final report.

Initial data analysis

There are four columns in ST344y2023Week1.csv. Each column is a separate set of data on independent subjects. For this practical only, I have not used real data. I have created data which reflects issues with which I have had to deal. The column names and content are:

PhysicsExam Marks for students in a physics examination

NurseDays Number of days work missed due to illness by nurses in 2022

Heights The heights of cricket players

Magazine The number of issues of the magazine 'Football Monthly' read in 2022

For each of these four separate data sets, provide a *suitable* set of summary statistics, a figure, with a sentence or two to describe the data.

For one data set, provide a poor set of summary statistics and figure. Explain why the summary statistics and figure are poor quality.

Speed Data: basic table

1. Create a table which summarises the age, sex, activities and shared interests with partner for all 551 people in the Speed Data study. Use sex for columns, as well as other columns you regard as relevant. Think carefully about what is important for a readable and informative table.
2. Write a brief paragraph about the table.

Word, table, figure limits.

Initial data analysis: The upper limits are 150 words, five tables and five figures.

Speed Data: basic table: The word limit is 100 words. Only one table is expected

The report

The only requirement is that the report is saved as a pdf with the file name ST344u1234567Week1.pdf, where '1234567' is replaced by your student number. Also include your student number in the report. You should not include any R-scripts or RMarkdown files for this assignment.

The report is uploaded to the course moodle.

The deadline for submission is Thursday 12 October, 13:00