Exercises 4

1. Two-variable tests with toy data

- Consider the following three variables:
  - `A`: 34, 23, 51, 47, 34
  - `B`: 48, 27, 33, 45, 41, 35
  - `C`: 34, 53, 54, 28, 52, 29
- Choose suitable statistical tests to compare pairs `A`&`B`, `A`&`C` and `B`&`C`. Justify your choices. What hypotheses do the tests concern?
- Calculate the P-values. What can you conclude based on the observed p-values?
- The description of the data is deliberately vague. Can you come up with other plausible tests for each pair?

2. More two-variable tests

The same property was measured before medication (`test`) and after medication (`re-test`) for two groups of individuals (D and E).

- Group D
  - `test`: 5.6, 3.1, 8.7, 4.5, 6.7, 4.5
  - `re-test`: 6.1, 5.8, 8.5, 5.3, 7.2, 5.1
- Group E
  - `test`: 4.5, 3.9, 7.1, 4.3, 6.9, 8.2, 7.6
  - `re-test`: 4.9, 4.7, 7.8, 4.8, 7.5, 7.8, 8.1

Select the correct statistical tests to compare the following pairs and calculate the P-values. What hypotheses do the tests concern? What can you conclude based on the observed p-values?

- `test` and `re-test` within group D
- `test` and `re-test` within group E
- `test` between groups D and E
- `re-test` between groups D and E

3. More two-variable tests (continues)

Consider the group E `test` and `re-test` from the earlier exercise.

- Calculate the Pearson correlation coefficient and its P-value.
- Calculate the Spearman correlation coefficient and its P-value.
- What hypotheses do the tests concern? What can you conclude based on the observed p-values?

4. Advertisements

An advertisement company followed online customers to discover how effectively advertisements lure them to spend money. The data files `ads-image.csv` and `ads-video.csv` contain information on how much customers spent in total after clicking on advertisements.

- Is there statistical evidence to claim that the total amount spent by customers is different if they click on image advertisements than on video advertisements?
- Explain the assumptions you made about how the data was collected and how it affected your choice of the test.

5. Electric bikes (continues)

Continue to analyse the electric bike data from the earlier exercises. Justify your design choices, interpret the results and use your discoveries to make conclusions about the customers.

- Is there statistical evidence to claim that the travel times tend to be shorter or longer for the single than for the season ticket type?
- Is there statistical evidence to claim that the savonia ticket type differs from the others with respect to how often the electric assistance is used?