R Stats Bootcamp

2.11 - T-tests

Megan Lewis

2025-04-10

R stats bootcamp - Module 2

Schedule:

- ~~Session 7: Explore data~~
- ~~Session 8: Distributions~~
- ~~Session 9: Correlation~~
- ~~Session 10: Regression~~
- **Session 11: T-test**
- Session 12: ANOVA

R Stats Bootcamp

The t-test and t-distribution are widely considered to be at the very foundation of modern statistic science and they form an important foundation for the practice of statisitics.

Student's T-test

Session 11 objectives:

- The question of the t-test
- Data and assumptions
- Graphing
- Tests and alternatives
- Practice exercises

The question of the T-test

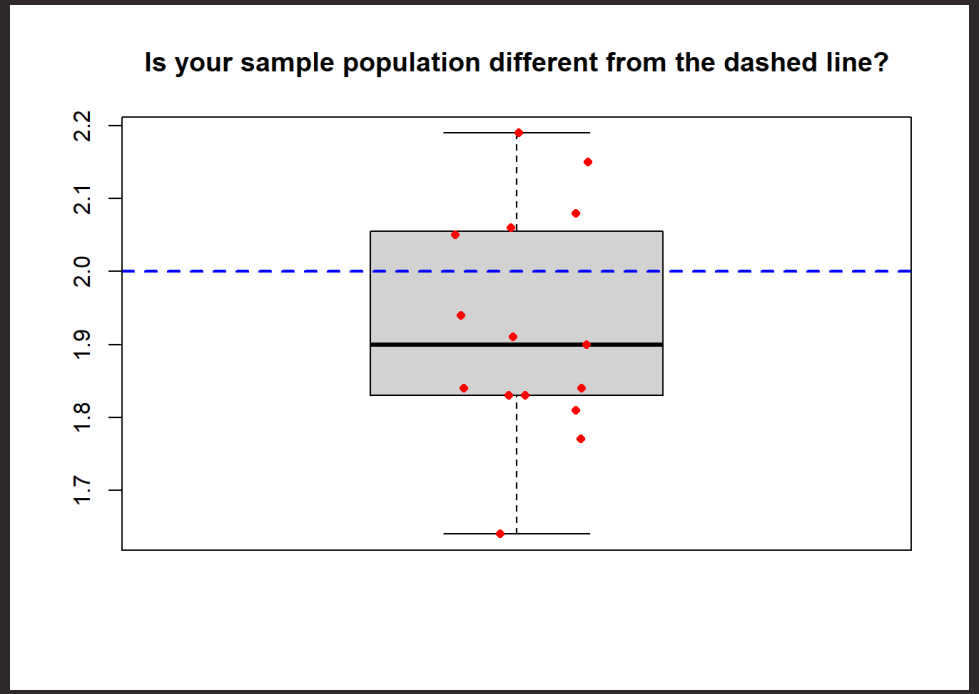
- Three common versions:
  - Are the means of the two samples different?
  - Did the sample come from a population exhibiting the known mean?
  - Is the mean difference between paired observations significantly differ from zero?

Independent samples

Off to R!

Comaparing sample to known mean

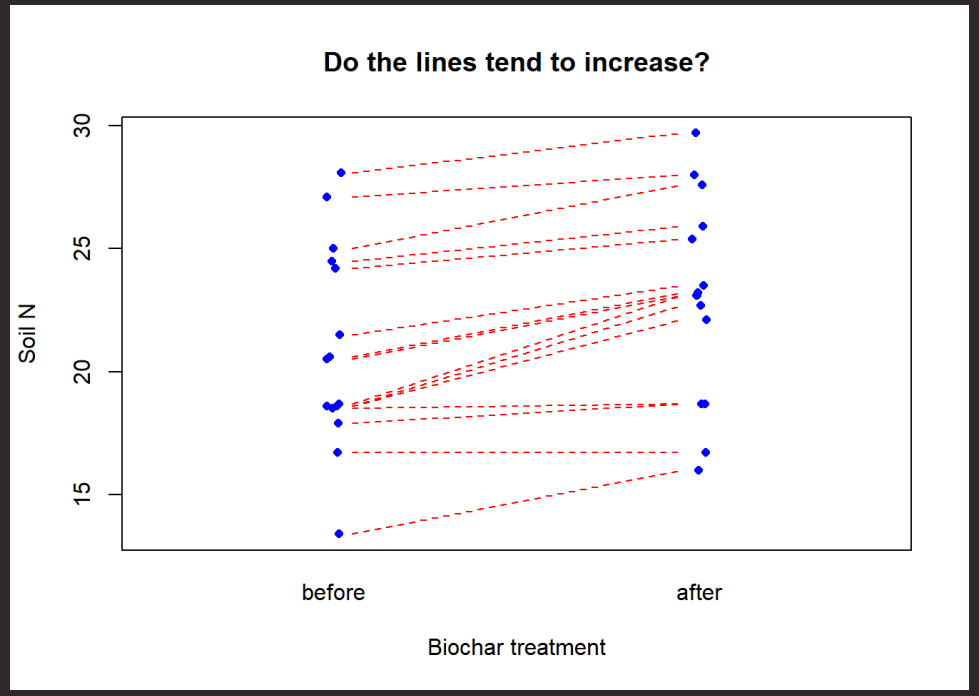
- Measured a numeric variable
- Comparing to a mean value



Off to R!

Paired samples

- Individuals comprising two samples are not independent
- Typical experiment design may include:
  - measurement of some variable before and after a treatment
  - measuring plots that are paired spatially



Off to R!

Principle assumptions

- Gaussian distribution of observations WITHIN each sample
- Heteroscedasticity (equal variance in each sample)
- Independence of observations

Evaluating and testing assumptions

- T-test thought to be somewhat robust to violations of assumptions
  - i.e., if assumption of Gaussian distribution or heterscedasticity are violated a little, it is not likely to biad results
  - Assumption of independence is always of high importance!

Gaussian distribution of observations WITHIN each sample

- Typically first test assumption of Gaussian distribution
  - Graphical evaluation - histogram, Q-Q plots, shapiro-test
- HOWEVER:
  - Gaussian assumption doesn’t apply to ALL OBSERVATIONS TOGETHER, but to EACH SAMPLE SEPARATELY
  - Two samples come from DIFFERENT POPULATIONS

Off to R!

Heteroscedasticity assumption

- Typically examine graphically or through calculation and comparision of descrtiptive statistics
- In case of two sample t-test, a method exisits to pool the standard deviation if variances aren’t equal
- If pooled SD is used and t-test conducted assuming unequal variances (not necessary to test this assumption)

Independence assumption

- EXTREMELY IMPORTANT
- Related to making an INFERENCE on a POPULATION of interest via SAMPLING one or more populations of interest
- If not met, consider paired t-test

Graphing

- 2 independent samples
  - Boxplot or similar showing central tendency
- 1 sample
  - Boxplot indicating mean as reference
- 2 paired samples
  - Boxplot (but hides pairwaise relationships)

Examples of the t-test and alternatives

- t-test for two independent samples
- 1 sample t-test
- paired samples t-test
- Mann-Whitney U-test

Off to R!

Mann Whitney U-test

- Alternative to t-test when data cannot meet assumptions of t-test
- T-test is robust, especially when sample size is large or deviation from assumptions is similar for both samples
- But when sample size less than ~30, and skew or samples are dissimilar - Mann Whiteny U-test is a good choice
- Two sample and one sample exist

Off to R!

Practice Exercises

Practice exercise 1

- Pick the appropriate form of t-test to ask whether male reported and actual `height` are the same.
- Perform the test, make a great graph to illustrate, and report your results in the technical style. Show all required code.

Practice exercise 2

- Devise a similar test to the one in the previous question using the `weight` variables in females.
- Formally state your hypothesis, perform the test, make a great graph to illustrate, and report your results in the technical style.
- Show all required code.

Practice exercise 3

- Calculate the difference between reported `height` and reported `weight` for all study subjects and place the result in a new numeric vector.
- Use a t-test to discover whether the degree of discrepancy between reported height differs between males and females.
- Report your results in the technical style. Show all required code.

Practice exercise 4

- Devise a way to examine the question whether taller people tend to self report height similarly to whorter people.
- Discuss your approach and present any evidence, graphical or otherwise, to resolve your question.

Practice exercise 5

- The subjects in this dataset were students at the University of California Davis in the Psychology Department. The average height of adult men in California has been estimated as 176.5 cm.
- Test whether males in our dataset is different. State your conclusion and results and briefly discuss an explanation for the pattern (i.e., the difference or lack of difference) that you observe. Comment on sampling assumptions when you do so.

Practice exercise 6

- Write a plausible practice question involving any aspect of data handling, graphing or analysis for the t-test framework to ask a novel question for the Davis student height data.