

Inferential Statistics

Lecture 12: R practical 2: comparing samples

Jérémy Hervelin

Pforzheim University
j.hervelin@lb.hs-pforzheim.de

30 November 2026 15:30–17:00
Winter semester 2026/27

Adapted from the teaching materials of Prof. Dr. Thilo Klein.

Our 90-minute route

- First 15 minutes: retrieve prerequisites and introduce the question.
- Next 35 minutes: worked examples and guided calculations.
- Next 25 minutes: paired activity and interpretation.
- Final 15 minutes: knowledge check, discussion and summary.

Practical objectives

- Connect the statistical question to an appropriate R command.
- Inspect assumptions and data before interpreting output.
- Report magnitude, uncertainty and limits in plain English.
- Prerequisites: Lectures 11. Work in pairs, then write your own interpretation.

Your materials

Open the supplied practical script in the course R folder. Keep `data/synthetic_youth.csv` in that folder.

The script uses base R and runs without additional packages. Exercises marked TODO are comments: fill in your own code below each prompt.

Synthetic youth cohort: read the dictionary

- 240 fictional young people, generated with a documented seed. No confidential records or empirical programme results.
- Variables: ID, age, education, random programme assignment, employment, monthly earnings, and interview scores before and after.
- Earnings are zero for non-employed people; interview scores use arbitrary score units.
- The generator defines the data-producing assumptions.

A reproducible workflow

- Set the working directory to the supplied R folder.
- Read the data, inspect names and types, then summarise and plot.
- Save commands and interpretations together in your script.
- Never edit observations to obtain a desired p-value.

Preserve the pairing

Compare before and after interview scores for the same synthetic people in the comparison group. Calculate differences and check that a paired t test matches a one-sample t test of differences.

Compare independent groups

Compare final interview scores in the simulated programme and comparison groups. Use Welch's test. Keep the subtraction order consistent in the estimate, interval and interpretation.

Compare changes across groups

Compute one change per person, then compare those changes between independently assigned groups. Explain why this is different from a before-after test within one group.

Employment shares in percentage points

Calculate each group's employment share and programme-minus-comparison gap. Distinguish percentage points from a relative percentage difference. Describe uncertainty rather than reporting the point estimate alone.

Return to the original PISA scores

The supplied vectors of 8 and 10 scores are retained. Compare Welch and pooled t tests; report both intervals. Explain the equal-variance assumption without using a non-significant F test as proof.

Interpret output in three sentences

- Estimate: what is the quantity, direction, size and unit?
- Uncertainty: what does the confidence interval or p-value imply under the assumptions?
- Limits: what does the design or model prevent us from concluding?

Common debugging checks

- Object not found: run the data import and earlier lines first.
- File not found: check your working directory and the relative data path.
- Unexpected sign: check which vector or group is subtracted from which.
- Warnings: read them; sparse expected counts are a statistical issue, not just a coding nuisance.

Practical submission

- Submit a script that runs from a clean R session.
- Include answers to the independent exercise and a short interpretation.
- Identify one assumption you can inspect in the data and one that requires knowledge of the design.
- Final exam: 18 December 2026, 13:00–14:00. Review the statistical methods and interpretation covered in this practical.

Independent practice

- Practise the five online questions for this lecture; 80% provides access to the next practice module.
- Course materials remain accessible regardless of your practice progress.
- Explain your reasoning before comparing answers.
- Review the method, assumptions and interpretation as well as the calculation.

Sources and teaching notes

Adapted from Prof. Dr. Thilo Klein's supplied Inferential Statistics materials.

Revised teaching edition: Jérémy Hervelin, Pforzheim University, winter semester 2026/27.

New youth-employment, programme and interview-score examples are synthetic illustrations.