

Assignment 2: Inference fundamentals

Course: Statistics for Engineers

Term: Fall 2025

Due date: October 17, 08:15

Objective:

The primary aim of this assignment is to apply the concepts of **confidence intervals** and **hypothesis testing** to real-world situations. By working on this task, you will gain practical experience in comparing characteristics across different populations. Each group has been assigned a specific dataset, containing samples from two populations denoted as A and B, to analyze using the R programming language. This hands-on approach will help you explore the practical aspects of statistical analysis and hypothesis testing, deepening your understanding and skills in these essential techniques. Please note that the datasets differ between groups, and we have already verified their suitability for this assignment.

Tasks

Each group must prepare a short, believable story based on the assigned dataset and then perform appropriate statistical analyses. Your story and analysis should look like a real mini-study that you carried out.

For your story and analyses, include the following:

- a) **Data visualization:** Begin by exploring and visualizing your data. Use suitable plots (histograms, boxplots, scatter plots, etc) to describe the variables in your dataset and to gain an initial understanding of possible patterns or differences between groups.
- b) **Single-group questions (pick either Group A or Group B):** Design one or two interesting and answerable research questions about one of the groups that can be addressed using confidence intervals and/or hypothesis tests.
- c) **Comparison question (compare Group A vs Group B):** Create one research question that demands comparing the means of the two groups.
- d) **For each question (both the single-group and the comparison questions):**
 - Clearly state the null hypothesis (H_0) and the alternative hypothesis (H_1), and justify your choices.
 - Choose an appropriate significance level $\alpha \in \{0.01, 0.05, 0.10\}$ and justify your choice.
 - Select a suitable statistical test (or construct a confidence interval) and verify that essential assumptions are satisfied.
 - Show all steps needed to come to a conclusion, and report key results, including, for instance, the test statistic, p -value, degrees of freedom (if applicable), etc.
- e) **Interpretation:** Provide a clear, context-based interpretation of your results in plain language. Avoid simply stating whether $p < 0.05$; explain what the findings mean in the context of your story and whether they are practically meaningful.

Submission

Each group must submit the following files on **Canvas**:

- Your **presentation** (in PDF format).
- A **log document** describing each group member's contributions (in PDF format).
- Your **R code** used within your study.

Deadline: October 17 at 08:15.

No late submissions will be accepted. Only one submission attempt is allowed.

Quick Checklist (to tick off before submitting)

- Story** is realistic, coherent, and explains why your analysis matters.
- Data visualization** is included to show initial exploration of the dataset.
- Hypotheses** are clearly stated for each question.
- Significance level α** is chosen and justified.
- Assumptions** are checked and clearly reported.
- Test statistics, p-values, and confidence intervals (CIs), etc,** are correctly reported.
- Interpretation** is written in plain language, not just statistical terms.
- Limitations** of the analysis are mentioned.
- R code** is clean, well-commented, and fully reproducible.
- Files** are correctly named: `groupX.pdf`, `groupX_log.R`, and `groupX_analysis.R`.

Oral Presentations

All group members must actively participate in the presentation. Each presentation is planned to be **10 minutes**. Your presentation should include the following:

1. Story and Data

Provide a clear and engaging presentation of your dataset and the story you designed around it. Explain briefly how your data relate to your research questions.

2. Methodology and Justification

Clearly explain your hypotheses, assumptions, and the reasoning behind the choices you made (e.g., test statistics, significance level, etc). Demonstrate that you understand the methods you applied.

3. Results and Interpretation

Present your key findings, supported by visuals (e.g., plots, tables, etc), and explain your conclusions in the context of your story. Make sure your interpretations are clear, concise, and meaningful in relation to your story.

4. Presentation Delivery

All group members must equally speak during the presentation. **You should present without reading from written notes or a prepared script. Reading directly from text will result in disqualification.**

At the end of your presentation, you will be asked a few questions about your analysis to evaluate your level of understanding of the results.

Evaluation

Your performance will be mainly assessed based on your understanding of the content presented. After reviewing your presentation, you will be graded on a scale from 0 to 6, with increments of 0.5 points.

To be considered successful, you must achieve a **minimum score of 3 points**.

All group members are expected to be **actively involved** during the presentation. Even though this is a group assignment, the evaluation can result in **individual grades** if a member is unable to demonstrate the expected level of understanding. A member who does not participate or cannot adequately explain the analysis may fail, regardless of the group's overall performance.

Having all the elements from the **check-list** does not automatically guarantee a passing grade. However, missing some of those elements can lead to failure. To ensure a passing grade, you should include all the checklist elements and demonstrate a good level of understanding in your study.

Your presentation should thoroughly address all key elements required to complete the assigned tasks. For reference, see Section *Tasks, Submission, and Oral Presentations* above for the specific points that will be evaluated.

Datasets:

Datasets are verified to be suitable for this homework and randomly assigned to the groups. Below you can find your assigned dataset. The numbers before each case refers to the group number. For instance, **1. Urban Trees and Air Quality Improvement** refers to group one.

1. Urban Trees and Air Quality Improvement

This dataset presents air quality scores for two groups of urban areas, each consisting of 15 measurements. Group A includes areas without newly planted urban trees, while Group B includes areas where new urban trees were planted. The measurements represent air quality scores recorded after one month, where higher scores indicate better air quality.

```
##  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   63   58   68   56   70   65   65   66   61   69   68   63   62   68   71
## B   70   71   67   69   77   73   67   64   67   68   69   69   70   71   77
```

2. Green Roofs and Urban Temperature Reduction

This dataset presents the effect of green roofs on local urban temperatures for 15 city blocks measured under two different conditions. Group A shows the average daily temperature (in °F) recorded when the city blocks had conventional rooftops, while Group B shows the temperatures recorded when the same blocks featured green roofs during a heatwave.

```
##  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   85   85   85   84   86   85   85   85   83   83   85   82   83   84   84
## B   85   85   85   86   88   84   86   86   84   85   86   86   87   86   86
```

3. Shopping Discounts and Spending

This dataset consists of spending data (in euros) for 15 customers over two shopping periods. Group A represents their spending during a regular shopping week without any special offers, while Group B shows their spending during a seasonal sale with major discounts and promotions.

```
##  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   69   68   75   72   72   74   73   72   74   70   73   67   73   66   69
## B   69   72   76   73   76   72   76   71   70   74   70   76   73   76   72
```

4. Outdoor Activities and Daily Mood

This dataset contains daily mood ratings (on a 1–50 scale) for two groups of 15 university students during a typical semester. Group A represents students who spent most of their free time indoors, studying or using electronic devices, while Group B represents students who regularly spent time outdoors, such as walking, biking, or meeting friends in parks.

```
##  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   42   42   42   43   41   42   44   43   42   42   43   43   42   43   40
## B   45   42   42   42   43   45   43   44   44   44   43   43   43   42   43
```

5. Group Study vs. Individual Study on Math Test Performance

This dataset presents the exam grades of 15 students for both an initial exam and a re-exam. Group A represents the students' grades from the re-exam, where the same students studied in groups, while Group B shows the grades from the main exam, for which they studied individually.

```
##  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   47   46   46   44   50   45   51   55   48   48   52   51   53   45   47
## B   43   43   44   48   48   48   49   47   48   45   46   52   44   42   44
```

6. Online Learning vs. In-Person Learning Outcomes

This dataset contains the grades of 15 students from two groups, A and B. Group A represents students who took the course online, while Group B includes students who attended the same course in person

```
##  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   63   62   62   62   63   58   57   66   63   64   61   60   54   64   66
## B   64   60   59   61   66   64   62   62   64   64   58   62   66   63   57
```

7. Stress Levels in Urban vs. Rural Populations

This dataset presents the stress levels of 15 individuals in two different time periods. Group A represents the initial stress levels of individuals living in rural areas, while Group B shows their stress levels after residing in urban areas for one year.

```
##  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   65   69   67   64   67   66   68   65   68   67   66   66   67   68   67
## B   68   66   69   68   68   69   67   70   70   68   69   69   69   67   70
```

9. Screen Time and Academic Performance

This dataset contains information on the academic performance of two groups of students, A and B, each consisting of 15 individuals. Group A includes students who spend a high amount of time on screens, while Group B consists of students who spend significantly less time on screens.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   91   95   94   94   98   94   93   91   96   94   92   93   94   93   95
## B   98   99   99   95   97   99   95   98   100  93   98   96   93   96   96
```

10. Music and Workplace Productivity

This dataset presents the productivity levels of two groups of employees, A and B, each consisting of 15 individuals. Group A consists of employees who do not listen to music during work hours, while Group B includes employees who listen to music while working.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   73   70   74   69   74   72   73   72   70   71   69   73   72   69   69
## B   69   70   67   67   78   74   81   80   80   73   83   76   74   77   71
```

11. Exercise Frequency and Sleep Quality

This dataset contains measurements used to study sleep quality for 15 individuals. Group A includes measurements when they do not engage in regular exercise, while Group B shows their measurements after participating in a regular exercise program for a specified period.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   56   56   51   60   53   65   51   56   61   58   55   58   61   49   57
## B   60   66   56   63   58   61   63   56   64   61   58   58   57   64   54
```

12. Social Media Usage and Self-Esteem

This dataset consists of measurements of self-esteem levels for two groups, A and B, each containing 15 individuals. Group A includes individuals who spend a high amount of time on social media, while Group B consists of individuals who spend minimal time on social media.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   56   60   59   58   48   61   65   54   58   63   56   54   59   55   57
## B   58   54   60   52   60   55   59   63   62   51   58   58   61   61   51
```

13. Workplace Flexibility and Job Satisfaction

This dataset contains measurements of job satisfaction for 15 employees. Group A includes measurements taken before a change in their supervisor, while Group B consists of their measurements after the change in leadership.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   8    12   15   13   11   14   8    11   8    12   18   9    8    4    9
## B   9    5    16   5    16   6    21  12   11   14   6    3    15   16   13
```

14. Study Time and Exam Performance

This dataset presents the performance of students on the TOEFL exam for two groups, A and B, each consisting of 15 individuals. Group A comprises students who studied for 12 hours per week over a period of three months, while Group B includes students who studied for 10 hours per week during the same duration.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   69   85   71   77   78   74   74   73   76   75   75   76   71   75   74
## B   74   80   76   73   75   81   76   78   71   67   72   78   79   82   70
```

15. Dietary Supplements and Energy Levels

This dataset contains information on the energy levels of individuals in two groups, A and B, each comprising 15 participants. Group A consists of individuals who take dietary supplements, while Group B includes individuals who do not take any supplements.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   26   21   11   18   18   23   22   20   15   20   29   22   15   18   21
## B   22   20   24   24   18   16   17   23   16   30   29   27   27   27   25
```

16. Impact of Meditation on Stress Reduction

This dataset presents measurements of stress reduction for two groups, A and B, each consisting of 15 individuals. Group A includes individuals who did not practice meditation, while Group B consists of individuals who practised meditation daily for one month.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   24   28   28   23   28   31   26   25   32   26   26   32   29   30   26
## B   38   30   33   32   23   35   28   20   37   25   36   27   23   38   32
```

17. Reading Speed: E-Books vs. Paper Books

This dataset presents measurements of reading speed for 15 individuals. Group A includes the reading speeds of individuals while reading e-books, while Group B captures their reading speeds while reading traditional paper books.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   55   53   50   53   54   52   53   56   54   56   57   53   54   55   52
## B   55   53   54   54   54   54   53   52   54   55   56   52   52   52   54
```

18. Sleep Quality: Smartphone Users vs. Non-users Before Bed

This dataset contains measurements of sleep quality for two groups, A and B, each comprising 15 individuals. Group A consists of individuals who use smartphones before going to sleep, while Group B includes individuals who do not use smartphones before bed.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   23   27   27   26   18   28   28   19   19   24   31   20   28   26   24
## B   40   25   32   32   23   37   27   34   29   19   25   24   28   34   34
```

19. Hydration and Athletic Performance

This dataset presents hydration measurements for 15 athletes. Group A includes measurements taken when the athletes consume more than 2 liters of water per day, while Group B consists of their measurements for when they drink less than 2 liters of water per day.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   55   52   45   50   56   55   49   55   50   46   54   53   53   50   52
## B   50   49   56   55   57   51   49   56   58   53   55   55   48   54   50
```

20. Gender and Emotional Intelligence

This dataset contains measurements of emotional intelligence for two groups, A and B, each consisting of 15 individuals. Group A includes male participants, while Group B consists of female participants.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   49   48   48   49   49   50   47   52   51   48   49   48   48   51   50
## B   48   45   51   50   48   49   47   51   42   49   41   53   50   42   50
```

21. Noise Level and Task Performance

This dataset contains measurements of employee performance for two groups, A and B, each comprising 15 individuals. Group A consists of employees working in noisy environments on the second floor of the building, while Group B includes employees working in quiet environments on the top floor.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   43   43   43   41   45   43   43   45   44   40   41   44   43   46   43
## B   45   45   44   47   43   45   44   45   45   46   44   44   45   43   44
```

22. Coffee Consumption and Productivity

This dataset presents measurements of productivity levels for two groups, A and B, each consisting of 15 individuals. Group A comprises individuals who consume coffee regularly, while Group B includes individuals who do not drink coffee.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   31   26   27   32   26   30   27   28   27   29   29   30   28   28   32
## B   25   26   37   33   24   26   39   21   31   27   39   32   28   31   30
```

23. High-Protein Diet and Muscle Gain

This dataset presents measurements of muscle gain for 15 individuals. Group A includes data for individuals following a moderate-protein diet, while Group B consists of their measurements after they have followed a high-protein diet for a specified period.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   7    6    6    3    5    7    4    8    7    7    5    7    5    5    4
## B   6    6    5    7    8    6    8    7    9    9    8    6    10   6    6
```

24. Team Sports vs. Individual Sports on Motivation

This dataset presents measurements of work-related motivation levels for two groups, A and B, each consisting of 15 individuals. Group A includes individuals who participate in team sports, while Group B comprises individuals who engage in individual sports.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   88   85   84   82   91   89   89   83   89   93   87   85   87   81   84
## B   92   91   86   86   91   85   86   87   90   90   85   88   93   90   90
```

25. Vitamin D Supplementation and Immunity

This dataset presents the reported frequency of illness for two groups, A and B, each consisting of 15 individuals. Group A includes individuals who take vitamin D supplements, while Group B comprises individuals who do not take any supplements.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   14    7   14   11   15   10   24   12   13   12    6   16    9   11   15
## B   14    9   19   20   24   12   21    7   15   16   17   17   11   18   13
```

26. Bilingualism and Cognitive Flexibility

This dataset contains measurements of cognitive flexibility for two groups, A and B, each consisting of 15 individuals. Group A includes bilingual individuals, while Group B consists of monolingual individuals.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   101  102   94  104  107   94   96   93   94   107  103   95   97  102   96
## B   93   101   97   91  101   94  108   97   96  101  100   88  102   91   98
```

27. Plant-Based Diet and Blood Pressure

This dataset presents relative measurements of re-scaled blood pressure for two groups, A and B, each consisting of 15 individuals. Group A includes individuals following a plant-based diet, while Group B comprises individuals following a mixed diet.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   33   35   37   30   36   33   41   41   36   29   39   42   33   41   38
## B   42   36   37   34   39   42   37   41   34   39   38   46   51   40   32
```

28. Standing vs. Sitting Desks on Back Pain

This dataset presents reported frequencies of back pain for two groups, A and B, each consisting of 15 individuals. Group A includes individuals who use standing desks, while Group B comprises individuals who use traditional sitting desks.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   96   93   93   93   93   91   91   96   94   92   94   96   96   94   94
## B  100  105   86   93  110   89   85   92   92   92   89   97  104   95  100
```

29. Creativity: Structured vs. Unstructured Work

This dataset presents measurements of creativity levels for two groups, A and B, each consisting of 15 individuals. Group A includes measurements for individuals before the implementation of new rules aimed at increasing structure in the work environment, while Group B includes measurements taken after individuals have been working in structured environments.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   36   34   31   29   24   36   34   28   31   32   34   39   31   27   32
## B   36   33   41   40   34   39   29   33   29   46   45   44   40   41   45
```

30. Community Cycling Program and Daily Bicycle Activity

This dataset presents the daily distance traveled by bicycle (in kilometers) for two groups of participants, each consisting of 15 individuals. Group A includes participants who actively joined a new community cycling program and are encouraged to cycle daily, while Group B includes participants who did not join the program and mostly follow their usual commuting habits using cars or public transport.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   18   24   22   23   17   22   25   25   23   19   21   20   19   19   18
## B   24   25   22   27   22   22   26   23   28   20   22   20   25   26   25
```

31. Work break and productivity score

This dataset consists of productivity scores for 15 employees. Group A represents productivity measured during a standard 8-hour workday with no breaks beyond lunch. Group B represents productivity measured during a modified schedule where employees took short breaks every 90 minutes throughout the day.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   91   96   92   100  94   95   96   93   99   99   100  95   94   100  104
## B   103  95   104  103  104  103  98   97   98   98   98   101  101  104  93
```

32. Video Games and Problem-Solving Skills

This dataset presents measurements of problem-solving skills for two groups, A and B, each consisting of 15 individuals. Group A includes individuals who are frequent video game players, while Group B comprises individuals who do not play video games.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   92   79   81   85   89   82   80   83   87   86   88   90   73   83   85
## B   83   81   87   75   83   91   92   92   84   90   89   87   86   87   82
```

33. Water Conservation and Daily Household Water Usage

This dataset presents the daily water usage (in liters) for two groups of households, each consisting of 15 households. Group A includes households that actively implement water-saving measures (e.g., low-flow faucets, rainwater collection, shorter showers), while Group B includes households that do not consistently follow water-saving practices.

```
## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   17   22   13   17   19   23   22   24   19   16   9    14   25   23   16
## B   24   22   21   15   23   27   25   26   28   21   19   20   15   16   16
```

34. Household Recycling and Weekly Recycled Waste

This dataset presents the weekly amount of recycled waste (in kilograms) for two groups of households, each consisting of 15 households. Group A includes households that actively participate in recycling programs, while Group B includes households that do not consistently recycle.

```
##  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   26   26   26   26   27   30   26   29   28   25   29   27   24   27   23
## B   28   25   28   25   30   28   33   25   27   30   31   28   31   25   28
```

35. Staffing Levels and Quality of Elderly Care

This dataset presents the average daily well-being scores (on a standardized scale) of residents from two groups of elderly care homes, each with 15 observations. Group A represents care homes currently experiencing staff shortages, while Group B includes homes with adequate staffing levels.

```
##  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   28   33   29   31   30   30   30   29   26   31   29   29   28   27   26
## B   27   34   29   33   30   31   28   35   27   30   36   34   32   30   29
```

36. Housing Queue Time and Tenant Satisfaction

This dataset presents tenant satisfaction scores (on a scale from 0 to 100) for two groups of renters, each consisting of 15 individuals. Group A represents tenants who obtained their apartments through the regular public housing queue, while Group B includes those who accessed housing through alternative or fast-track systems

```
##  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15]
## A   57   58   54   57   57   55   56   56   54   54   59   57   59   58   54
## B   54   60   57   59   62   60   57   55   60   59   57   59   55   56   59
```