

Homework 1

P8130 Fall 2025

Due: September 16 at 11:59pm

Guidelines for Submitting Homework

- Your homework must be submitted through Courseworks. No email submissions!
- Only one PDF file should be submitted, including all derivations, graphs, output, and interpretations. When handwriting is allowed (this will be specified), scan the derivations and merge ALL PDF files (<http://www.pdfmerge.com/>).
- You are encouraged to use R for calculations, but you must show all mathematical formulas and derivations. Please include the important parts of your R code in the PDF file but also submit your full, commented code as a separate R/RMD file.
- To best follow these guidelines, we suggest using Word (built in equation editor), R Markdown, Latex, or embedding a screenshot or scanned picture to compile your work.

REMINDER: You are encouraged to collaborate on homework, explain things to each other, and test each other's knowledge. But **everyone must complete their own assignment and write their own solutions.**

Problem 1 (5 points)

Classify each of the following variables as binary, nominal, ordinal, discrete, or continuous:

- a. Patient survival status recorded as “alive” or “deceased”
- b. Stage of cancer at diagnosis (I, II, III, IV)
- c. Type of vaccine received (e.g., Pfizer, Moderna, Johnson & Johnson)
- d. Body temperature measured in degrees Celsius
- e. Number of emergency room visits in the past year
- f. Self-reported pain level (“none”, “mild”, “moderate”, “severe”)
- g. Systolic blood pressure (in mmHg)
- h. Diabetes status (“no diabetes”, “pre-diabetes”, “diabetes”)

Problem 2 (10 points)

In a study on recovery times following two types of minor surgeries, recovery duration (in days) was recorded for 15 patients who underwent laparoscopic surgery and for 14 patients who underwent open surgery. The recovery times are as follows:

Laparoscopic surgery:

12, 15, 10, 14, 13, 16, 20, 18, 9, 11, 17, 19, 14, 13, 15

Open surgery:

20, 25, 22, 30, 18, 27, 24, 21, 29, 26, 22, 23, 28, 24

- a. Calculate the mean, median, range, and standard deviation of recovery times for the laparoscopic surgery group.
- b. Construct and describe a box plot for the laparoscopic surgery group. Use terms such as skewness and modality to describe the distribution.
- c. Calculate the mean, median, range, and standard deviation of recovery times for the open surgery group.
- d. Construct and describe a box plot for the open surgery group using similar descriptive terms.
- e. Create side-by-side box plots comparing recovery times between the laparoscopic and open surgery groups. Label the plots clearly.
- f. Compare the two groups' recovery times based on the box plots and summary statistics. Which group tends to recover faster?
- g. Describe any potential outliers or unusual values visible in the box plots and suggest how they might affect interpretation.
- h. Discuss limitations of judging recovery differences based on these samples and suggest additional analyses that could be performed.

Problem 3 (5 points)

We say that events A and B are independent if $P(A|B) = P(A)$. Show that if $P(A|B) = P(A)$ then also $P(B|A) = P(B)$.

This answer may be hand-written.

Problem 4 (10 points)

In a community health study of 200 adults, data were collected on two behaviors: regular exercise (at least 3 times a week) and daily consumption of sugary drinks.

Define the following events:

- Event A: An adult exercises regularly.
- Event B: An adult consumes sugary drinks daily.

Survey results showed that:

- 120 adults exercise regularly.
- 70 adults consume sugary drinks daily.
- 40 adults both exercise regularly and consume sugary drinks daily.

- a. What is $P(A)$ and $P(B)$?
- b. Calculate $P(A \cup B)$, the probability that a randomly selected adult either exercises regularly or consumes sugary drinks daily (or both).
- c. Calculate $P(B|A)$.
- d. Calculate $P(A|B^c)$.
- e. Are events A and B independent? Justify your answer.
- f. If two adults are selected at random independently, what is the probability that both exercise regularly but neither consumes sugary drinks daily?

Problem 4 (10 points)

Among women aged 75 and older in a certain population, 20% have dementia. Among women in this group who have dementia, 70% show positive findings on a certain brain CT scan. Among women without dementia, about 15% also show positive findings (false positives).

- a. If a randomly selected woman in this population has a positive CT scan finding, what is the probability that she actually has dementia?
- b. If a randomly selected woman in this population has a NEGATIVE CT scan finding, what is the probability that she DOES NOT have dementia?
- c. Discuss the implications of these probabilities for interpreting test results in this population.

Show all calculations. This answer may be hand-written.