Data & Probability Assignment

Instructions

- You should submit a single PDF file through Brightspace. If you decide to work on the markdown template provided, knit it to PDF and submit it. Make sure the R code snippets are compiled.

- A penalty will be applied to late submissions. Submissions more than 5 days late will not be accepted.

- Whenever you use R, you should include the code and its output.

- You are encouraged to use R to help answer the questions, although it is not required. However, you must also provide clear explanations of your reasoning. Submitting only R code, even if it produces the correct answer, will not be sufficient to receive full credit.

- This is an individual assignment. Do not copy it from someone or somewhere, and do not share your work with anyone else! If fraud is suspected (e.g., if your answers are too similar to the answers of someone else) the Board of Examiners will be notified.

- The deadline is 07 December 2025, 23:59 CET.

Questions

1. (1.5 pts) Write R code to answer all the questions related to the `tips`. You can get started working on this question using the RMarkdown template posted on Brightspace. You should refer to the help file of the dataset to know about the variables. This will help you answer the questions.

   (a) Find the number of people who visited the restaurant.

   (b) Create a column called tip_rate that would show the proportion of the tip to the bill and add it to the dataframe.

   (c) If you were to work at this restaurant, which shift on which day would you prefer? Support your response with a plot. *Hint: Your motivation for doing this job until you are hired by a big tech company is making money.*

   (d) What is the probability that there will be a group of size 6 on sunday for dinner?

2. (3 pts) You are playing a dice game using two six-sided dice. The dice are **not fair**, but for simplicity, assume that both dice have the same bias. The probability of each side appearing on a single die is:

Side	1	2	3	4	5	6
Probability	0.05	0.10	0.15	0.20	0.25	0.25

The rules of the game are as follows:

- You and the dealer toss a fair coin to decide who starts the game. If the coin lands on Heads, then you start the game. Otherwise, the dealer starts the game.
- You and the dealer each roll two dice with the starter rolling first .
- Let the total sum of your dice be your score.
- The winner is determined as follows:
    - If both players have the same total, the game ends in a **tie**.
    - If only one player rolls a total of 7 or 11, that player wins immediately.
    - If only one player's total exceeds 11, the other player wins.
    - If none of the above occurs, the game **continues to the next round**.

(a) Calculate the probability mass function of the discrete random variable T representing the total of the dice.

(b) At the end of the first round, what is the probability that:

   i. The game ends in a tie?

   ii. You win immediately?

   iii. The game continues to the next round?

Make sure that you clearly explain the assumptions you are making to calculate these probabilities.

(c) Eventually, what is the probability that:

   i. The game ends in a tie?

   ii. You win immediately?

Make sure that you clearly explain the assumptions you are making to calculate these probabilities.

(d) Is it possible that this game continues forever? Why? Explain.

Hint: You may use simulation to calculate the probabilities. Particulary, we highly recommend using R for the probability distribution of T.

3. (1 pts) A company has developed a "Marten Car Alarm Detector", designed to alert car owners when malicious stone martens are chewing on the cables of their vehicles. In northern Europe, there are 500,000 stone martens, but only 500 are estimated to tamper with cars. To test the product before launching it to the market, a car owner conducts trials:

- The alarm correctly detects a car-tampering marten 99% of the time.
- The alarm incorrectly triggers for a harmless marten 1% of the time.

(a) If the alarm goes off, what is the probability that a stone marten is actually tampering with the car?

(b) Based on your calculations, would you, as a car owner, be confident in buying this alarm system?

4. (2 pts) A bubble tea cup is 15 cm tall. Because the tapioca pearls settle toward the bottom, the linear density of pearls (in pearls per cm of height) is given by

$$f(h) = 50 - 2h, \qquad 0 \le h \le 15,$$

where h is the height measured from the bottom of the cup.

(a) How many pearls are in the cup?

(b) Let H be the continuous random variable of height of a randomly selected pearl. Find the probability density function of H.

(c) Find the cumulative distribution function of H.

(d) If a straw reaches down to a depth of 10 cm, what is the probability that the drinker swallows a pearl?

5. (1 pts) X is the result of rolling a fair 4-sided die and Y is defined as follows:

$$Y = \begin{cases} \text{the result of rolling a fair 8-sided die,} & \text{if } X \text{ is even,} \\ X, & \text{if } X \text{ is odd.} \end{cases}$$

(a) Verify that X and Y are not independent.

(b) Find $\mathbb{E}[XY]$.

6. (1.5 pts) Let W be the weight (in kg) of a person in a population, with mean 70 and standard deviation 15. Consider a random sample of n people. Let $\bar{W}_n$ denote the sample mean weight.

(a) What is the distribution that approximates $\bar{W}_n$? Give its parameters. Explain why?

(b) Using R, approximate the probability that the sample mean weight is between 75 kg and 80 kg for the following sample sizes:

    i. $n = 100$

    ii. $n = 1000$

    iii. $n = 10000$

(c) Discuss how the probability changes as the sample size increases.