UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

STC 122 - Assignment 4 submission instructions

Instructions

- Answer the questions that follow and save all your code in a single R script. Name the script **Assignment4.R** where **Assignment4** is the filename and **.R** is the file extension.

- Two submissions are required for this Assignment, namely a code submission and an interpretation submission.

Submission 1: Code

- Submit your R script on Gradescope before **Tuesday, 10 October 2023, 23:59**.

- Multiple code submissions are allowed and your autograded results will be available shortly after each submission.

- **Ensure all variables are named correctly, as incorrectly named variables will not be awarded any marks.** (Remember variable names are case sensitive.)

- **Ensure your code does not contain any syntax errors.** If your code produces errors when run, the autograder will not be able to mark it.

- Any code commented out will be considered rough work and will not be marked.

- Once you have completed your submission, ensure the file is submitted on Gradescope with the correct file name, **Assignment4.R** where **Assignment4** is the filename and **.R** is the file extension. **The autograder will only be able to grade your submission if you use the correct filename.**

- Students must add **detailed** explanations of their working as comments in their code. This must be done for every question in the assignment. Students should also add citations for any resources they make use of in their assignment. Refer to the document `commenting_and_citing_in_your_code.pdf` on ClickUP.

Submission 2: Interpretation

- Submission 2 will be an invigilated assessment.

- Full details regarding the date, time and format of this assessment will be communicated as soon as possible on ClickUP.

Guidelines

- This assignment covers Inference for Numerical Data.

- The assignment is based on Labs 3, 4, 5 and 7 from the Openintro website. The labs are available on ClickUP. You should work through these labs before attempting this assignment.

- You should work through Chapters 11 to 13 and Chapters 19 to 23 in the prescribed textbook in order to understand the underlying theory to do this assignment.

- This assignment is based on the online CoffeeTruck game from the Statistics Department at Grinnell College, Iowa, USA. It is not necessary to play the game in order to do this assignment but see https://www.stat2games.sites.grinnell.edu/ to access the game and for more details regarding the game.

- Where applicable, answer the questions below by typing the appropriate **code** in the R script template provided on ClickUP. Some questions are theoretical and no coding is needed to answer those questions.

- **You should not use the `inference` function from the `statsr` package to answer any questions in the assignment**.

- For this assignment you should use the dataset in the file CoffeeTruck.csv that is available on ClickUP. The dataset contains the coffee sales in the above-mentioned online coffee truck game. The variables in the dataset are defined as follows:

    - `Location`: the location of the coffee truck for the day.
    - `Music`: the genre of music playing in the coffee truck for the day.
    - `Price`: Price of the coffee.
    - `Sales`: Number of coffee sales.
    - `Income`: Income for the coffee sales.
    - `Cost`: Cost for the coffee sales of the day.
    - `Profit`: the profit made for the day.

Questions

(Adapted from lab created by Shonda Kuiper from the Statistics Department at Grinnell College, Iowa, USA)

Let's assume that you and your best friend, Jo, live in the bustling city of Beanville. Jo's love for coffee is contagious, and you two share a dream of starting your own coffee business.

With a passion for all things caffeinated, Jo has started brewing up a strategy to purchase a coffee truck and serve her delightful concoction to customers throughout the city. While you don't want to waste any time getting started with this entrepreneurial adventure, you are hesitant to make any major purchases before developing a careful business strategy.

There are multiple factors that can affect coffee sales and profit:

- ***Location:*** *The accessibility and proximity of a coffee shop to potential customers can greatly influence sales, as convenient locations attract more foot traffic.*

- ***Music:*** *The ambiance created by the type of music played in a coffee shop can influence customers' moods and perceptions, impacting their likelihood to stay longer and make additional purchases.*

Data has been collected and you are tasked to investigate the influence of the different locations and music on the coffee sales.

Question 1

Test whether the `Profit` variable follows a normal distribution by using the Anderson-Darling test. Save the p-value of the test into a variable called `Q1`

Question 2

Simulate the sampling distribution of the average profit made with coffee sales using the following steps:

- Set the seed value to 15.

- In a for loop, take a random sample from the `Profit` variable. Let the size of the bootstrap sample be the same as the size of the `Profit` variable.

- For each random sample, calculate the mean of the random sample.

- Repeat the previous two steps 1000 times.

- Save your response into a variable called `Q2`

Question 3

Consider the sampling distribution obtained in `Q2`

  a) Calculate the theoretical probability of observing a sample mean between 326 and 347. Save your answer into a variable called `Q3a`

  b) Calculate the empirical probability of observing a sample mean between 326 and 347. Save your answer into a variable called `Q3b`

Question 4

Using the `t.test` function, calculate a 90% confidence interval for the average profit made in a day through sales with the coffee truck. Save the upper limit of the confidence interval into a variable called `Q4`

Question 5

Will it be profitable to have the coffee truck at the City Hall?

Using the `t.test` function, answer the above question by making use of an appropriate hypothesis test. Save the p-value of the test into a variable called `Q5`

Question 6

Is there a difference in the average profit made between the different locations for the coffee truck?

Answer this question by making use of an appropriate hypothesis test.

  a) Calculate the mean square error. Save your answer into a variable called `Q6a`

  b) Calculate the test statistic value. Save your answer into a variable called `Q6b`

Question 7

Will you make more profit by playing Alternative music rather than HipHop music in the Coffee truck?

Answer this question by making use of an appropriate hypothesis test. Let μ_1 be the average profit made when playing Alternative music and let μ_2 be the average profit made when playing HipHop music. Assume that the population variances are equal.

  a) Calculate the standard error required to calculate the test statistic value for the hypothesis test. Save your answer into a variable called `Q7a`

  b) Calculate the test statistic value. Save your answer into a variable called `Q7b`

  c) Calculate the p-value. Save your answer into a variable called `Q7c`