Practice Problem Set - 0 (R, Probability and Sampling Distributions)

ECO 204 (Spring 2025)
Instructor: Shaikh Tanvir Hossain

§

We will go through these problems in the class.

1 Basic R Exercises

1. Open R Studio, create a R script, then change the directory with the command `setwd()`, and save the script as `PS_0_1.R`. Also close the R studio and open it again and run the following commands in the script.

```
# create an object x = 10 and object y = 20 and and add them and store value in a new object called z
x <- 10
y <- 20
z <- x + y
# create a vector of numbers with the values 1, 2, 3
data <- c(1, 2, 3)

# create a vector of numbers with the values 1, 2, 3, 4, 5
data <- c(1:5) # you can also use seq() function, use this to create an even sequence

# create a string
name <- "John Doe"

# create a matrix
mat <- matrix(1:6, nrow = 2, ncol = 3)

# Create a list that can store different types of objects
my_list <- list(
  number = 42,
  name = "Alice",
  values = c(3, 4, 5)
)
print(my_list)

# Create a simple data frame
df <- data.frame(
  id = 1:3,
  name = c("Alice", "Bob", "Charlie"),
  score = c(85, 90, 88)
)
print(df)

# Create a factor (categorical variable) called temperature
temp <- factor(c("low", "medium", "high", "medium", "low"))


# Define a simple function to calculate the square of a number
square <- function(x) {
  return(x^2)
}

# Test the function
result <- square(5)
print(result)

# If Else condition
num <- 10
if (num %% 2 == 0) {
  print("Even number")
} else {
  print("Odd number")
}
```

2. Now open a Markdown, create a document with Title, Author, Date, save the RMD file (which is short for RMarkdown file). This question explains how to create a R Markdown file. You will submit R Markdown files whenever you will submit the problem set, where it has some description of the problem and the solution code. It's an excellent way to combine writings and code.

   Otherwise, usually in class or for practice we will use R Scripts.

2 Problems Related to Normal Distributions

For the following problems, you can find the codes in the file `PS_0_3.R` file

3. Using the function `rnorm()` in R, generate a random sample of size 100 from $\mathcal{N}(10, 2)$, and plot the histogram of the sample. Increase the sample size to 1000 and plot the histogram. What do you observe?

4. Using the function `pnorm()` in R, for a random variable $X \sim \mathcal{N}(20, 3)$, calculate the probability $\mathbb{P}(X \leq 21) = ?$

5. Using the function `qnorm()` in R, for a random variable $X \sim \mathcal{N}(20, 3)$, calculate the value x such that $\mathbb{P}(X \leq x) = 0.95$ (Here I am looking for a value x, which is also called quantile)

3 Problems Related to Sampling Distribution of Sample Means

6. Suppose we have data that shows that the average expenditures of the EWU students were $\$10,348$. Use $\$10,348$ as the population mean and suppose a survey research firm will take a sample of 100 students to investigate the nature of the expenditures. Also for the following questions you can assume all the students' observations are independent.

   (a) Assume the population standard deviation is $\$2500$, and also assume independence then what is the mean and variance of the sample mean $\overline{X}$.

   (b) Again assume the population standard deviation is $\$2500$, and moreover the expenditure data is normally distributed, then what is the sampling distribution of $\overline{X}$, what is the mean and variance of the sampling distribution.

   (c) Now assume we don't know the population standard deviation, however we know the sample standard deviation of the 100 sample, and that is $\$2800$, what is the distribution of the T statistic, what is the parameter of the distribution.

   (d) Answer following questions assuming the scenario in (b)

      i. What is the probability the sample mean will be within $\pm\$200$ of the population mean?

      ii. What is the probability the sample mean will be greater than $\$12,000$?

   (e) Answer following questions assuming the scenario in (c)

      i. What is the probability the sample mean will be within $\pm\$200$ of the population mean?

      ii. What is the probability the sample mean will be greater than $\$12,000$?

   (f) Finally if we don't assume the normality of the data, then what can we say about the distribution of the sample mean $\overline{X}$? What assumption we need to use this large sample approximation.

   Note that: In all of the above problems we don't have any estimation, or we didn't get any value of the estimate. But in question 3) you can get some estimates... how?

Remarks:

References: