Homework #3\_FA2024\_BMEN7340

First\_Last\_7340\_HW3\_FA24

2024-09-30

# clean R environment

rm(list = ls(all=TRUE))  
graphics.off()  
shell("cls")

library(readxl)  
library(ggplot2)  
library(epiR)  
library(MASS)  
library(nnet)  
library(car)  
library(PairedData)  
library(pwr)  
library(dplyr)  
library(FSA) #posthoc

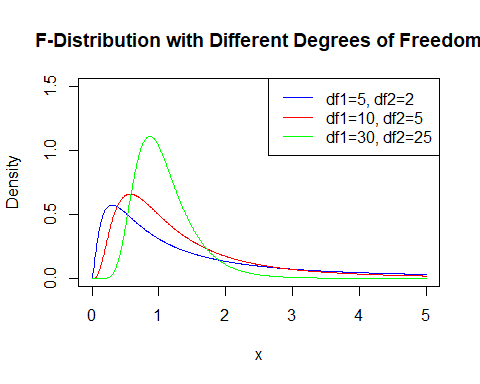
# Question 1:

## What is an F-distribution?

## In what situations is the F-distribution applied?

## The R code below visually compares the shapes of the F-distribution for various pairs of degrees of freedom. How do the degrees of freedom influence the shape of the F-distribution?

# Generating 100 random numbers from an F-distribution with 5 and 2 degrees of freedom  
set.seed(111)  
# Set up the x values  
x <- seq(0, 5, by = 0.01)  
# Calculate the density for different degrees of freedom  
y1 <- df(x, df1 = 5, df2 = 2) # F(5, 2)  
y2 <- df(x, df1 = 10, df2 = 5) # F(10, 5)  
y3 <- df(x, df1 = 30, df2 = 25) # F(30, 25)  
  
# Plotting the F-distributions  
plot(x, y1, type = "l", col = "blue", ylim = c(0, 1.5),   
 ylab = "Density", xlab = "x",   
 main = "F-Distribution with Different Degrees of Freedom")  
lines(x, y2, col = "red")  
lines(x, y3, col = "green")  
# Adding a legend  
legend("topright", legend = c("df1=5, df2=2", "df1=10, df2=5", "df1=30, df2=25"),   
 col = c("blue", "red", "green"), lty = 1)



# Design a study:

## Research Question and Introduction:

Formulate a research question relevant to your field of study. Provide a brief introductory paragraph that includes the context of the case and the units of measurement used.

## Identify Variables:

Determine the independent variable (IV) and the dependent variable (DV) in your study.

## Data Levels:

Specify the levels of measurement for both the IV and DV.

## Outcome Measurement:

Based on your research question, clearly define the outcome you intend to measure.  
# Statistical Analysis:  
Select an appropriate statistical analysis method based on the outcome. Consider options such as a two-sample mean test (parametric or non-parametric), one-way ANOVA (parametric or non-parametric), or correlation (parametric or non-parametric).

## Hypotheses Formulation:

Based on the chosen statistical method, formulate the null hypothesis (Ho) and the alternative hypothesis (Ha).

## Data Generation:

Assuming your samples are drawn using simple random sampling (SRS), use R code to generate a simulated dataset with a sample size greater than 30.

## Data Visualization:

Employ appropriate descriptive statistical tools to visualize the data using histograms, scatterplots, or boxplots.

## Hypothesis Testing:

Use the relevant inferential statistical tool to test your hypotheses.

## Assumptions Check:

List the assumptions related to your statistical method and verify that they are met.

## Report R Output:

Present the R output, including the test statistic, degrees of freedom (df), confidence interval (CI), and p-value, where applicable.

## Power Analysis:

Calculate the statistical power of your study.

## Conclusion:

Summarize the findings of your study, ensuring to include the content of your research and the units of measurement.

### To generage a normally distibuted random sample with sample size more than 30

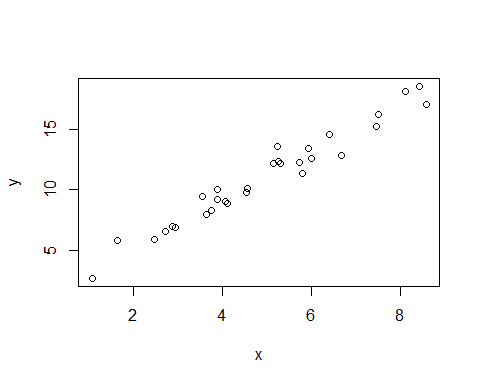
# Set the sample size  
sample\_size <- 30  
# Generate a normally distributed random sample  
random\_sample <- rnorm(sample\_size, mean = 20, sd = 2.5)  
# You can adjust the mean and sd parameters to fit your needs.

### To generate pairs of 𝑥and𝑦for a correlation study

# Set the sample size  
sample\_size <- 30  
# Generate x values from a normal distribution  
set.seed(123) # Set seed for reproducibility  
x <- rnorm(sample\_size, mean = 5, sd = 2)  
# Generate y values with some noise  
y <- 1 + 2 \* x + rnorm(sample\_size, mean = 0, sd = 1)  
# Combine x and y into a data frame and give a name "data"  
data <- data.frame(x = x, y = y)  
# Display the first few rows of the data  
head(data)

## x y  
## 1 3.879049 9.184562  
## 2 4.539645 9.784219  
## 3 8.117417 18.129959  
## 4 5.141017 12.160167  
## 5 5.258575 12.338732  
## 6 8.430130 18.548900

plot(x, y)



### To generate a non-normally distributed random sample.

# Set the sample size  
sample\_size <- 30   
# Generate an exponentially distributed random sample  
set.seed(123) # Set seed for reproducibility  
exp\_sample <- rexp(sample\_size, rate = 1) # rate parameter  
# Display the first few values  
head(exp\_sample)

## [1] 0.84345726 0.57661027 1.32905487 0.03157736 0.05621098 0.31650122

shapiro.test(exp\_sample)

##   
## Shapiro-Wilk normality test  
##   
## data: exp\_sample  
## W = 0.82484, p-value = 0.0001926