LAB 1 — PROBABILITY

For all questions below, provide all programming code and plots in the report.

Primer Material

1. Let's pretend a researcher conducts a study to examine whether socioeconomic status influences success in school. The researcher does not want to drive too far so she samples participants from around her neighborhood. Is the good or bad research practice. Why? (1 Mark)

2. Let's pretend you have some data [4,7,2,4,8,6,5,1,0,9,10,3, 8,9,5,2,7,3,2,5,7]. (2 Marks).

   a. Provide a summary of the mean, median, mode, quartiles, range, sample variance, sample standard deviation, sample standard error, and sample coefficient of variation. 1 mark.

   b. Plot a histogram and box plot. 1 mark.

3. Let's say some data follows a Normal Distribution with mean of 1.5 and standard deviation of 0.5 $\mathcal{N}(1.5, 0.5^2)$.

   a. Use the formula $f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ to plot the Normal Distribution PDF (vary through different x). 1 marks.

   b. Use the formula $\Phi(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$ to plot the Normal Distribution CDF. 1 mark

   c. calculate the probability of x being less than 1.5

   d. calculate the probability of x being greater than 1

   e. calculate the probability of x being between 1.1 and 1.6

NOTE: In R you can call the error function with, erf(). To access this function, type the following in your console:
install.packages("NORMT3")
library(NORMT3)
You now have access to the function, erf().

4. Let's sample n scores from a probability distribution and estimate the sample mean and the standard error of the mean. Let's sample from $\mathcal{N}(10.0, 2.5^2)$ using the rnorm() function.

   a. Sample 10 values and estimate the sample mean and the standard error of the mean (0.5 mark)

b. Now calculate a z-score for a value of 11.0 (0.5 mark)

c. Sample 10000 values and estimate the sample mean and the standard error of the mean (0.5 mark)

d. Now calculate a z-score for a value of 11.0 (0.5 mark)

e. What do you notice about the standard error with a changing n? Interpret. (1 mark)

f. What do you notice about the t-score with a changing n? Interpret. (1 mark)

5. Lets sample 20 values from group 1 ($\mathcal{N}(10.0, 2.5^2)$) and 20 values for group 2 ($\mathcal{N}(12.0, 2.5^2)$).

a. Estimate the sample mean difference and the standard error of the mean difference (0.5 mark)

b. Now calculate a t-score for a difference value of 0.0 (i.e., $\mu_1 - \mu_2 = 0$) (0.5 mark)

c. Interpret the t-score (1 mark)

6. Find Welch's Confidence Interval for the following data (2 marks):

Disorder	Patients (n)	Treatment time ($\bar{x}$)	Std. Deviation (s)
Schizophrenia	18	4.7	9.3
Bipolar	10	8.8	11.5

Probability

7. Find the point probabilities below, and show your work:

a. The probability of graduating from an undergraduate degree is 80%. What is the probability of not graduating (0.5 marks)

b. 4/10 individuals get a job out of high school and 4/10 go to college. What probability of people do neither? (0.5 marks)

c. A jar contains 12 red marbles and 8 black marbles. Two marbles are drawn *with replacement* from the jar (i.e., the first marble was put back in before the second draw). What is the probability that both of the marbles are black? (1 mark)

d. A jar contains 12 red marbles and 8 black marbles. Two marbles are drawn *without replacement* from the jar (i.e., the first marble was not put back in before the second draw). What is the probability that both of the marbles are black? (2 mark)

e. The probability of finishing a 5km race is 90%. The probability of finishing a race and running under a 20 minute 5km race is 15%. What is the probability that a person will run a 25km race given they finished the race? (1 mark)

f. A man goes to the mall. The probability that he buys: a) a pair of pants is 0.4, b) a shirt is 0.3, c) both pants and a shirt is .2. What is the probability that the man leaves the mall with pants, a shirt, or both? (1 mark)

g. For the question above, what is the probability that the man leaves the mall with nothing (1 mark)?

8. Sample 5,000 points from a normal distribution $\mathcal{N}(5.0, 2.0^2)$:

a. Plot the histogram (1 mark)

b. Show the data with a QQ plot (1 mark)

c. Perform a Shapiro-Wilk Test (1 mark)

d. Interpret the plots and results from a-c (1 mark)

9. Sample 5,000 points from a Beta distribution, where $\alpha = 6$ and $\beta = 4$:

a. Plot the histogram (1 mark)

b. Show the data with a QQ plot (1 mark)

c. Perform a Shapiro-Wilk Test (1 mark)

d. Interpret the plots and results from a-c (1 mark)

10. The average height and weight of the 2016 San Antonio Spurs is 78.8 (SD = 3.668) inches and 211 (SD = 25.904) lbs, respectively. The correlation between height and weight is ρ = 0.81.

a. Sample 1000 data points drawn from a bivariate (joint) normal distribution and show the data with a scatter plot. (1 mark)

b. What is the probability a player is over 85 inches tall? (1 mark)

c. What is the probability a player is under 190 lbs? (1 mark)

d. What is the probability a player is between 200 and 220 lbs? (1 mark)

e. What is the probability that a player is under 75.5 inches given they are 200 lbs? (1 mark)

f. What is the probability that a player is over 250 lbs given they are 86 inches? (1 mark)