STAT 1910 - Winter 2022$_{\bar{x}}$

Project

Your Name: Jonathon Meney **Your Student ID: 348074**

You are given a small data set in a file pDat.csv. This file consists of 3 columns and 40 rows. Following are the first few rows.

```
      Gender Age  BMI
4872  Female  67 26.0
12722 Female  55 24.0
1729  Female  72 31.6
2611  Female  77 26.0
6116    Male  70 32.6
8974  Female  86 21.5
```

In this file, the first row is for the column names. All these forty people are the population of a small town. You will work only on BMI (Body Mass Index). I will suggest using R or a calculator to do all the calculations. However, using R will make your life easy.

1. Compute the μ and σ for BMI only for these data.
   $\mu = 27.1525$
   $\sigma = 4.902499$

I draw ten random samples, each of size five, from this population of 40. These ten random samples are stored in ten separate files, rs1.csv, rs2.csv, and so on.

2. Compute the mean and standard deviation of these ten random samples.
   Sample 1: $\bar{x} = 27.42$ $s = 2.392070$
   Sample 2: $\bar{x} = 27.48$ $s = 5.622455$
   Sample 3: $\bar{x} = 29.10$ $s = 3.941446$
   Sample 4: $\bar{x} = 25.14$ $s = 2.740073$
   Sample 5: $\bar{x} = 31.02$ $s = 6.989063$
   Sample 6: $\bar{x} = 24.76$ $s = 4.531335$
   Sample 7: $\bar{x} = 23.94$ $s = 1.728583$
   Sample 8: $\bar{x} = 29.08$ $s = 6.616041$
   Sample 9: $\bar{x} = 26.92$ $s = 5.372802$
   Sample 10: $\bar{x} = 25.80$ $s = 3.321897$

3. Now, compute the mean and standard deviations of these ten sample means and sample standard deviations. Don't forget to use an appropriate formula for $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$ for n=5.
   $\mu_{\bar{x}} = 27.066$ $\sigma_{\bar{x}} = 2.076989$

4. What percentage of the 10 sample means found in (3) lie in the interval $\mu_{\bar{x}} - \sigma_{\bar{x}}$ to $\mu_{\bar{x}} + \sigma_{\bar{x}}$? In the interval $\mu_{\bar{x}} - 2\sigma_{\bar{x}}$ to $\mu_{\bar{x}} + 2\sigma_{\bar{x}}$? In the interval $\mu_{\bar{x}} - 3\sigma_{\bar{x}}$ to $\mu_{\bar{x}} + 3\sigma_{\bar{x}}$?
   $\mu_{\bar{x}} - \sigma_{\bar{x}}$ to $\mu_{\bar{x}} + \sigma_{\bar{x}}$: 70%
   $\mu_{\bar{x}} - 2\sigma_{\bar{x}}$ to $\mu_{\bar{x}} + 2\sigma_{\bar{x}}$: 100%
   $\mu_{\bar{x}} - 3\sigma_{\bar{x}}$ to $\mu_{\bar{x}} + 3\sigma_{\bar{x}}$: 100%

5. I took another random sample of size 15 from this population of 40. This random sample is stored in the file rs15.csv. Find the 95% confidence interval for the mean BMI μ. Assume that the BMI of the 15 subjects is normally distributed.
   The 95% confidence interval is 22.62610 to 26.56056.

R Code:

```
data <- read.csv ("pDat.csv")

BMI <- data$BMI

# Q1
Mean <- mean(BMI)
stanDiv <- sqrt((sum(BMI**2) - ((sum(BMI))**2)/length(BMI))/length(BMI))

# Q2
s1 <- read.csv ("rs1.csv")
s1M <- mean(s1$BMI)
s1SD <- sd(s1$BMI)

s2 <- read.csv ("rs2.csv")
s2M <- mean(s2$BMI)
s2SD <- sd(s2$BMI)

s3 <- read.csv ("rs3.csv")
s3M <- mean(s3$BMI)
s3SD <- sd(s3$BMI)

s4 <- read.csv ("rs4.csv")
s4M <- mean(s4$BMI)
s4SD <- sd(s4$BMI)
```

```
s5 <- read.csv ("rs5.csv")
s5M <- mean(s5$BMI)
s5SD <- sd(s5$BMI)

s6 <- read.csv ("rs6.csv")
s6M <- mean(s6$BMI)
s6SD <- sd(s6$BMI)

s7 <- read.csv ("rs7.csv")
s7M <- mean(s7$BMI)
s7SD <- sd(s7$BMI)

s8 <- read.csv ("rs8.csv")
s8M <- mean(s8$BMI)
s8SD <- sd(s8$BMI)

s9 <- read.csv ("rs9.csv")
s9M <- mean(s9$BMI)
s9SD <- sd(s9$BMI)

s10 <- read.csv ("rs10.csv")
s10M <- mean(s10$BMI)
s10SD <- sd(s10$BMI)

# Q3
sMeans <- c(s1M, s2M, s3M, s4M, s5M, s6M, s7M, s8M, s9M, s10M)

meanOfSample <- mean(sMeans)

stanDivOfSample <- (stanDiv/sqrt(5)) * sqrt(35/39)

# Q5
samOf15 <- read.csv("rs15.csv")
samBMI <- samOf15$BMI
mean(samBMI)
CI(samBMI, ci=0.95)
```