# FE590. Assignment #1.

## 2021-09-25

I pledge on my honor that I have not given or received any unauthorized assistance on this assignment/examination. I further pledge that I have not copied any material from a book, article, the Internet or any other source except where I have expressly cited the source.

By filling out the following fields, you are signing this pledge. No assignment will get credit without being pledged.

Name: Muhammet Furkan Isik

CWID: 10472193

Date: 09/25/2021

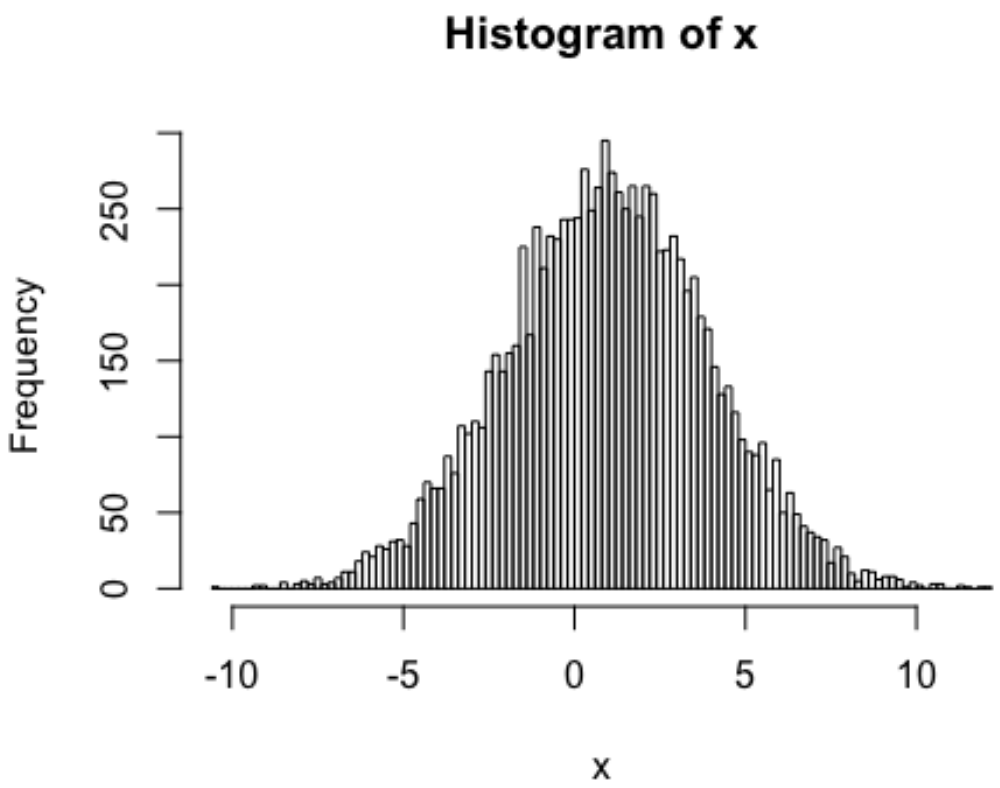
# Question 1

## Question 1.1

```
CWID = 10472193 #Place here your Campus wide ID number, this will personalize
#your results, but still maintain the reproduceable nature of using seeds.
#If you ever need to reset the seed in this assignment, use this as your seed
#Papers that use -1 as this CWID variable will earn 0's so make sure you
change
#this value before you submit your work.
personal = 10472193 %% 10000
set.seed(personal)
```

Generate a vector x containing 10,000 realizations of a random normal variable with mean 1.0 and standard deviation 3.0, and plot a histogram of x using 100 bins.

```
x= rnorm(n=10000, mean=1,sd=3)
hist(x,breaks=100)
```




(Note that the following two fields can be added wherever you desire to show a solution. You can use the first for a written response, and the second for showing R code and its output. Some questions will require just one, and some both. I will not always provide you with these, but you can add them at your discretion whereven necessary. If it makes sense to do the R code first then that's fine. If you want to include multiple of each, that's ok too. Do what you feel is necessary to answer the question fully.)

```
#response
```

## Question 1.2

Calculate the mean and standard deviation of these 10000 values. Do your answers make sense?

```
mean(x)
```

```
## [1] 0.9621471
```

```
sd(x)
```

```
## [1] 2.987446

# Mean and the standard deviation of random variable X is quite close to one
that we created above with mean=1, and sd=3
```

## Question 1.3

Using the `sample` function, take out 10 random samples of 500 observations each (with replacement). Create a vector of the means of each sample. Calculate the mean of the sample means and the standard deviation of the sample means. What do you observe about these results?

```
##  Mean of the sample means is very close to population mean, and sd of the
sample means is 0.15

sample= replicate(n=10,sample(x,500))
mean_vector= colMeans(sample)
mean(mean_vector)

## [1] 0.9457518

sd(mean_vector)

## [1] 0.1429849
```

#Question 2

##Question 2.1

Create a script that creates a vector of the values from 1 to 100 using a for loop.

```
vec= vector()

for (i in seq(from=1, to= 100)) {


  vec= append(vec,i)


}


print(vec)

##   [1]   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17
18
##  [19]  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35
36
```

```
##  [37]  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53
54
##  [55]  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71
72
##  [73]  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89
90
##  [91]  91  92  93  94  95  96  97  98  99 100
```

##Question 2.2

Create a script that creates a vector of the values from 1 to 100 however you like (just make it different from the above approach)

```
a=1
new_vec= c()

while (a<=100) {

  new_vec= append(new_vec,a)
  a=a+1

}

new_vec
```

```
##   [1]   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17
18
##  [19]  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35
36
##  [37]  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53
54
##  [55]  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71
72
##  [73]  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89
90
##  [91]  91  92  93  94  95  96  97  98  99 100
```

## Question 3

Download and use the “obese.csv” file found under Modules section of Canvas. The data frame has 102 rows and 3 columns. It contains data from a random sample of Mexican-American adults in a small California town.

```
data= read.csv("/Users/metuhead/Desktop/FA590/obese.csv")
```

## Question 3.0

Find the mean and standard deviation of the variables ‘obese’ and ‘bp’ in this data set.

```
mean(data$obese)
```

```
## [1] 1.313039
```

```
mean(data$bp)
```

```
## [1] 127.0196
```

```
sd(data$obese)
```

```
## [1] 0.2578387
```

```
sd(data$bp)
```

```
## [1] 18.18441
```

```
means= list(mean(data$obese), mean(data$bp))
names(means) = c("obese_mean", "bp_mean")


sds= list(sd(data$obese),   sd(data$bp))
names(sds)= c("obese_sd", "bp_sd")


print(c(means, sds))
```

```
## $obese_mean
## [1] 1.313039
##
## $bp_mean
## [1] 127.0196
##
## $obese_sd
## [1] 0.2578387
##
## $bp_sd
## [1] 18.18441
```

## Question 3.1

The variable sex is an integer code with 0 representing male and 1 representing female. Use the `table` function operation on the variable `sex’ to display how many men and women are represented in the sample.

```r
table(data$sex)
```

```
##
##  0  1
## 44 58
```

## Question 3.2

The `cut` function can convert a continuous variable into a categorical one. Convert the blood pressure variable `bp` into a categorical variable called `bpc` with break points at 80, 120, and 240. Rename the levels of `bpc` using the command `levels(bpc) <- c("low", "high")`.

```r
bpc=cut(data$bp, breaks = c(80,120,240))
levels(bpc)= c("low_bp", "high_bp")

data$bpc= bpc

bpc
```

```
##   [1] high_bp high_bp high_bp high_bp high_bp high_bp high_bp low_bp
high_bp
##  [10] high_bp low_bp  low_bp  high_bp low_bp  high_bp low_bp  high_bp
low_bp
##  [19] high_bp high_bp high_bp high_bp high_bp low_bp  low_bp  low_bp
low_bp
##  [28] low_bp  high_bp high_bp high_bp high_bp high_bp low_bp  high_bp
high_bp
##  [37] high_bp low_bp  high_bp high_bp low_bp  low_bp  low_bp  high_bp
high_bp
##  [46] high_bp high_bp high_bp high_bp low_bp  high_bp low_bp  high_bp
low_bp
##  [55] low_bp  low_bp  high_bp high_bp high_bp high_bp low_bp  low_bp
low_bp
##  [64] high_bp high_bp high_bp low_bp  low_bp  low_bp  low_bp  low_bp
low_bp
##  [73] low_bp  high_bp high_bp high_bp low_bp  low_bp  high_bp high_bp
high_bp
##  [82] low_bp  low_bp  low_bp  high_bp low_bp  low_bp  low_bp  low_bp
low_bp
##  [91] high_bp high_bp high_bp low_bp  low_bp  high_bp low_bp  high_bp
high_bp
## [100] high_bp high_bp high_bp
## Levels: low_bp high_bp
```

## Question 3.3

Use the `table` function to display a relationship between `sex` and `bpc`.

```
table(data$sex, data$bpc)
```

```
##
##     low_bp high_bp
##   0     16      28
##   1     28      30
```

## Question 3.4

Now cut the `obese` variable into a categorical variable `obesec` with break points 0, 1.25, and 2.5. Rename the levels of `obesec` using the command `levels(obesec) <- c("low", "high")`.

Use the `ftable` function to display a 3-way relationship between `sex`, `bpc`, and `obesec`.

```
obesec= cut(data$obese, breaks= c(0,1.25,2.5) )
levels(obesec)= c("low_obs","high_obs")
#obesec

data$obesec= obesec


ftable(data$sex, data$bpc,data$obesec)
```

```
##            low_obs high_obs
##
## 0 low_bp        12        4
##   high_bp       15       13
## 1 low_bp        14       14
##   high_bp        4       26
```

Which group do you think is most at risk of suffering from obesity? Why? (Note that the why is just as important as getting the group right. This holds true for the course as a whole, not just this problem.)

```
# Comparing the genders, we can say that 1( female),  have  more obesity than
males (0)

# The group, who is most at risk is:  female(1) who have  high blood pressure
because they have the highest obesity numbers among all the categories.
```