Q1

Huiyan Li

3/29/2020

A. Calculate the population mean

```
population = c(21, 22, 23, 24, 25)
# mean = Sum of number divide by the number of item
populationMean = mean(population) # Output: 23
populationMean
```

```
## [1] 23
```

B. Calculate the population variance

```
populationVariance = sum((population - populationMean)**2)/5  # Output: 2
populationVariance
```

```
## [1] 2
```

C. Find all combination from this population with replacement and sample size of n = 3

Using for loop to generate all possible combination of the population

```
combinationList = vector()
for (valA in c(21:25)) {
  for (valB in c(21:25)) {
    for (valC in c(21:25)) {
      combinationList <- c(combinationList, c(valA, valB, valC))
    }
  }
}
combinationList <- matrix(combinationList, ncol = 3, byrow=TRUE)
colnames(combinationList) <-  c("Col1", "Col2", "Col3")
```

D. Calculate the sample mean of each row of sample 3 from part C

For looping through each row and compute its mean, and save it to the X_bar vector object

```
X_bar = vector()
for (row in 1:nrow(combinationList)) {
  sampleMean = mean(c(combinationList[row,0:3]))
  X_bar <- c(X_bar, sampleMean)
}
```

E. Construct a frequency table based on the column X bar

Using built in function data.frame to auto-compute the frequency

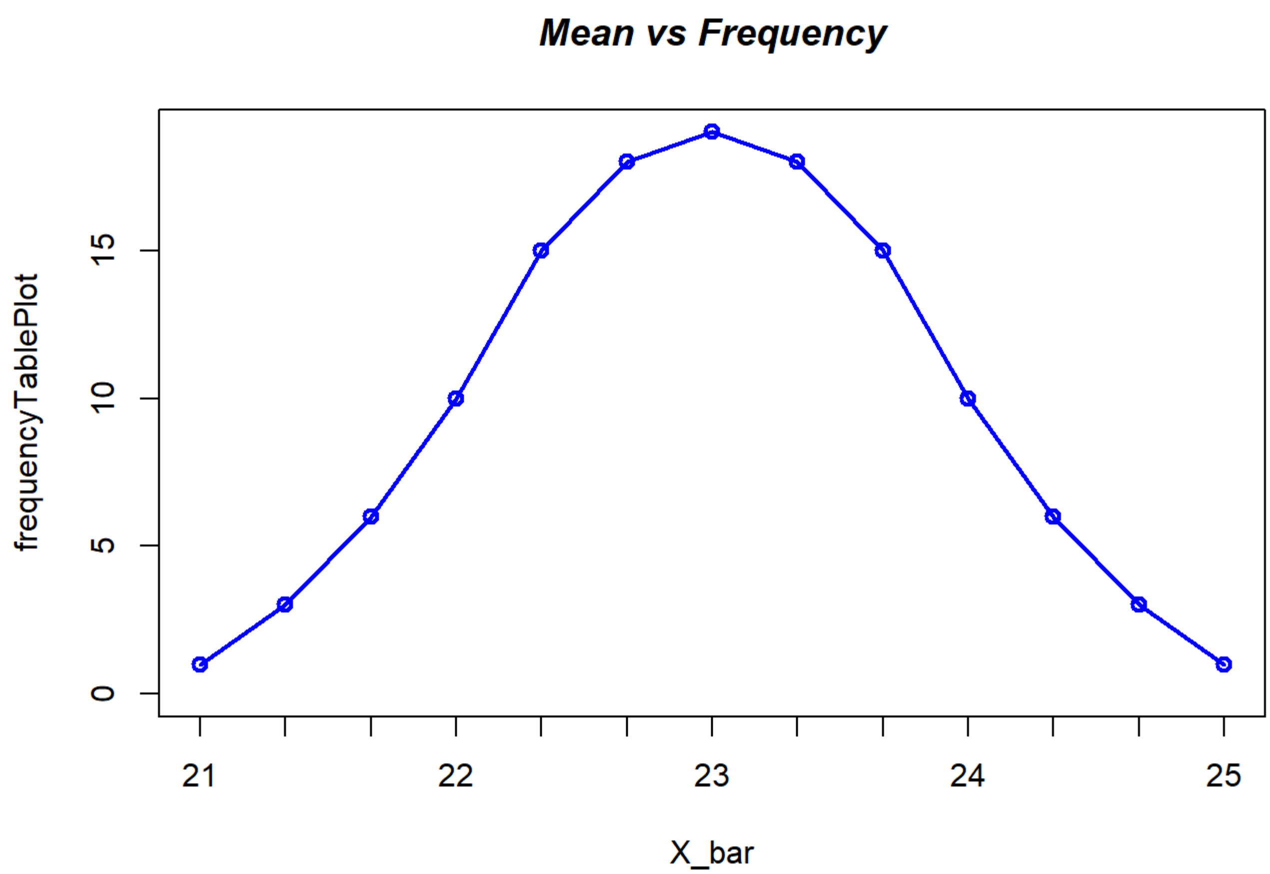
```
frequencyTable = as.data.frame(table(X_bar))
frequencyTable
```

```
##               X_bar Freq
## 1                21    1
## 2  21.3333333333333    3
## 3  21.6666666666667    6
## 4                22   10
## 5  22.3333333333333   15
## 6  22.6666666666667   18
## 7                23   19
## 8  23.3333333333333   18
## 9  23.6666666666667   15
## 10               24   10
## 11 24.3333333333333    6
## 12 24.6666666666667    3
## 13               25    1
```

F. Plot these proportions against the values and connect the points using a non-linear line.

Convert the X_bar to table format which also list out the frequency per value This table format is easier to plot

```
frequencyTablePlot = table(X_bar)
plot(frequencyTablePlot, type="o", col="blue")
title(main="Mean vs Frequency", col.main="Black", font.main=4)
```



The output graph looks very much like normal distribution

G. Using the table of proportions or otherwise, calculate the mean of these 125 numbers (values under X bar) and compare it to your answer of 1(a)

```
X_barMean = mean(X_bar) # Output : 23
X_barMean
```

```
## [1] 23
```

The output from the mean of X_bar is identical to the mean that we computed from part A

H. Using the table of proportions or otherwise, calculate the variance of these 125 numbers. Use the population variance formula (i.e. divide by 125 not 124). What is the relationship of this answer to your answer of 1(b)?

```
X_barVariance = sum((X_bar - X_barMean)**2)/125  # Output: 0.666667
X_barVariance
```

```
## [1] 0.6666667
```

The original variance computed from 1.B was 2, which is very different than the 0.6666 computed from this. This is because involves more data set, and the variance computed from $\overline{X}$ is much more reliable due to larger sample size, which result in a smaller variance.

I. Which theorem did you demonstrate empirically in part f, g and h?

F, G, H, demonstrate the Central Limit Theorem, which states
Sampling distribution of the sample means approaches a normal distribution as the sample size gets larger — no matter what the shape of the population distribution.