STAT253 – Project#<2>

Ufuk GURBUZ - <150113058>

12 Mayıs 2016

<<< PROJECT EXPLAIN >>>

To simulate the sampling distribution of $\bar{x}$, we have selected 50 samples of size n = 10 with replacement from a population with mean, μ=4.4 and standard deviation, σ=2.15, and have calculated the corresponding sample means. Sample means are given as R dataset at the course webpage.

<<< DATA TABLE >>>

```
load("C:/Users/_SeriousBoy_/Desktop/Statistic Project
2/Homework_2_50samples.csv")

print(ex0717$`Sample mean`[1:10])

##  [1] 4.8 3.0 4.6 5.0 4.4 4.2 5.9 4.1 4.6 4.2

print(ex0717$`Sample mean`[10:20])

##  [2] 4.2 4.2 5.7 3.4 4.1 4.2 4.5 4.2 4.9 5.1 5.2

print(ex0717$`Sample mean`[20:30])

##  [3] 5.2 4.3 4.4 4.1 3.4 5.4 4.3 4.8 4.0 5.9 4.8

print(ex0717$`Sample mean`[30:40])

##  [4] 4.8 5.0 5.0 3.7 5.0 3.6 4.0 5.1 4.3 4.3 5.0

print(ex0717$`Sample mean`[40:50])

##  [5] 5.0 3.3 4.8 4.3 4.5 4.5 4.7 4.2 4.5 3.9 4.9
```

< - - QUESTION 1 - - >

Construct a relative frequency histogram for these 50 values of $\bar{x}$. What is the shape of this distribution?

```
h<-hist(ex0717$`Sample mean`, plot=F)
h$counts <- h$counts / sum(h$counts)*100
colors = c("red", "yellow", "green", "violet", "orange", "blue")
plot(h, col = colors,freq=TRUE, main="Histogram of Sample Means",
xlab="Sample Means (x)", ylab="Relative Frequency (%)")

x<-ex0717$`Sample mean`
xfit<-seq(min(x),max(x),length=100)
yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))
yfit <- yfit*diff(h$mids[1:2])*length(x)
lines(xfit, yfit, col="black", lwd=2)
```

Histogram of Sample Means
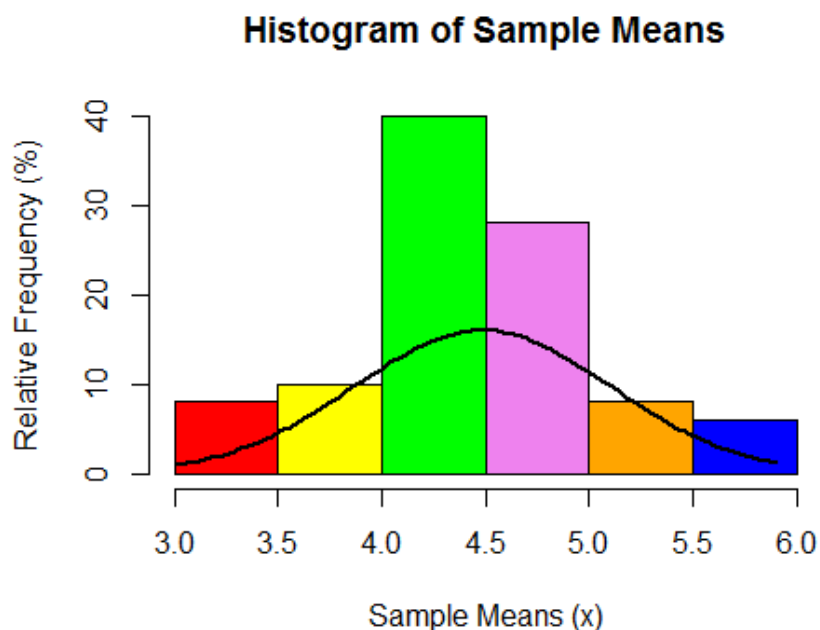
~ The Shape of This Distribution ~

*** The shape of this distribution is "Symmetric-Normal Distribution". Namely, the distribution of variables are regular and they create a balanced shape. The median equals mean in the shape of this distribution.

```
a=seq(0,pi,0.0001)
b=sin(a)
plot(a, b,,main="Symmetric-Normal Distribution", xlab="Sample Means (x)",
ylab="Relative Frequency (%)")
```

Symmetric-Normal Distribution
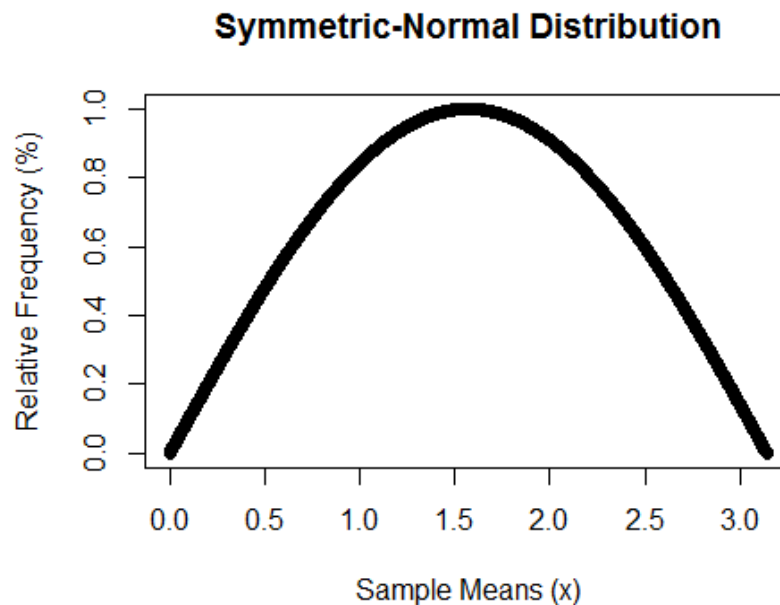
< - - QUESTION 2 - - >

Find the mean and standard deviation of the 50 values of $\bar{x}$.

~ The Five-Number Summary ~

- Let us find five-number summary of 50 values.

```
summary(ex0717$`Sample mean`)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.000   4.200   4.450   4.486   4.900   5.900
```

~ The Mean and Standard Deviation of 50 Values ~

```
mean<-mean(ex0717$`Sample mean`)
stdDev<-sd(ex0717$`Sample mean`)

print(paste("The Mean of 50 values: ", mean))

## [1] "The Mean of 50 values:  4.486"

print(paste("The Standard Deviation of 50 values: ", stdDev))

## [1] "The Standard Deviation of 50 values:  0.623030775274204"
```

< - - QUESTION 3 - - >

Compare the values calculated in part a to the theoretical mean μ and the theoretical standard deviation $\sigma\sigma\sqrt{n}$ for the sampling distribution of x̄. How close do the values calculated from the 50 measurements come to the theoretical values?

```
theoreticalMean<-4.4
sigma<-2.15
n<-10
theoreticalStdDev<-sigma/sqrt(n)

print(paste("The Theoretical Mean of 50 values: ", theoreticalMean),)

## [1] "The Theoretical Mean of 50 values:  4.4"

print(paste("The Calculated Mean of 50 values: ", mean))

## [1] "The Calculated Mean of 50 values:  4.486"

print(paste("The Theoretical Standard Deviation of 50 values: ",
theoreticalStdDev))

## [1] "The Theoretical Standard Deviation of 50 values:  0.679889696936201"

print(paste("The Calculated Standard Deviation of 50 values: ", stdDev))

## [1] "The Calculated Standard Deviation of 50 values:  0.623030775274204"
```

~ Comparison of Calculated and Theoretical Results ~

*** Firstly, I calculated mean and standard deviation. Then, I found theoretical mean and standard deviation. I compared calculated values and theoretical values. The calculated and theoretical results are almost the same for mean and standard deviation. There are some small differences between these results. Because; the specific part of population (sample) can't supplant population. The sample covers less than area according to population. So, its results can't be absolute correct. But, the choise of values is balanced. Because of this, the calculated and theoretical results are almost the same.