Quinn Knudsen

Homework 3

Topic: **Sampling Distributions**

2)

> summary(ChickWeight)

weight Time Chick Diet

Min. : 35.0 Min. : 0.00 13 : 12 1:220

1st Qu.: 63.0 1st Qu.: 4.00 9 : 12 2:120

Median :103.0 Median :10.00 20 : 12 3:120

Mean :121.8 Mean :10.72 10 : 12 4:118

3rd Qu.:163.8 3rd Qu.:16.00 17 : 12

Max. :373.0 Max. :21.00 19 : 12

(Other):506

> colnames(ChickWeight)

[1] "weight" "Time" "Chick" "Diet"

> dim(ChickWeight)

[1] 578 4

The column names are weight, Time, Chick, and Diet. In the dimensions, 4 represents the number of columns (variables) and 578 represents the number of rows (observations).

3)

summary(ChickWeight$weight)-summary of the variable weight in the ChickWeight dataset. It shows the range of values from min to max as well as the 1st and 3rd quantile. It also displays the median and the mean as measures of central tendency for this variable.

> summary(ChickWeight$weight)

Min. 1st Qu. Median Mean 3rd Qu. Max.

35.0 63.0 103.0 121.8 163.8 373.0

head(ChickWeight$weight)-shows the first 6 values in the dataset ChickWeight for the variable weight.

> head(ChickWeight$weight)

[1] 42 51 59 64 76 93

mean(ChickWeight$weight)- shows the average value in the dataset ChickWeight for the variable weight. This is the total of all the weight rows divided by the number of rows with non-missing data.

> mean(ChickWeight$weight)

[1] 121.8183

myChkWts <- ChickWeight$weight This code stores the variable weight from the dataset ChickWeights into a new vector myChkWts.

> myChkWts <- ChickWeight$weight

quantile(myChkWts,.50)

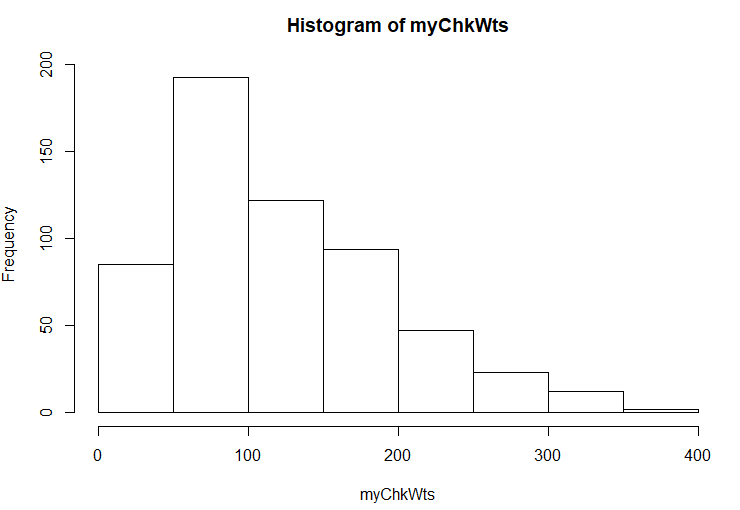
> quantile(myChkWts,.50) shows the average value in vector myChkWts which is identical to the average for the variable weight in ChickWeights as would be expected.

50%

103

4)

> hist(myChkWts)



This data is skewed right indicating more extreme values on the higher end of the spectrum (probably GMO steroid loving chickens).

> quantile(myChkWts,.025)

2.5%

41

The smallest 2.5% of the data falls below the value of 41.

> quantile(myChkWts,.975)

97.5%

294.575

The largest 2.5% of the data falls above 294.575.

> mean(myChkWts)

[1] 121.8183

The average data point is 121.8183.

> median(myChkWts)

[1] 103

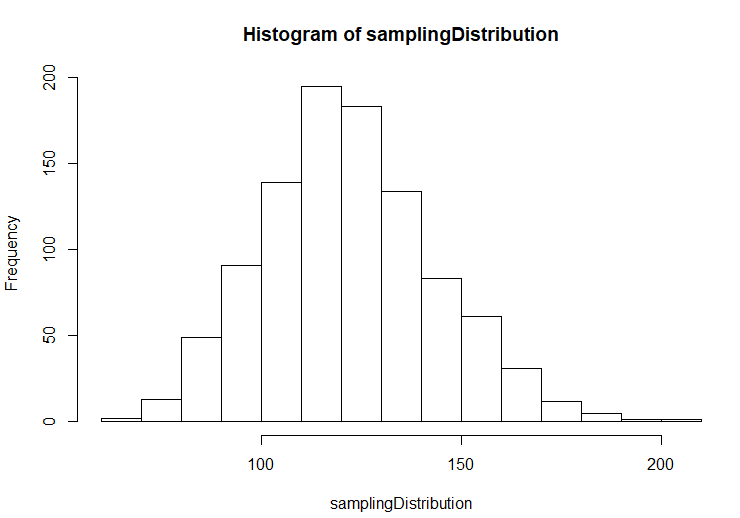
The median or middle value is 103.

95% of the observations of weight fall between 41 and 294.575.

5)

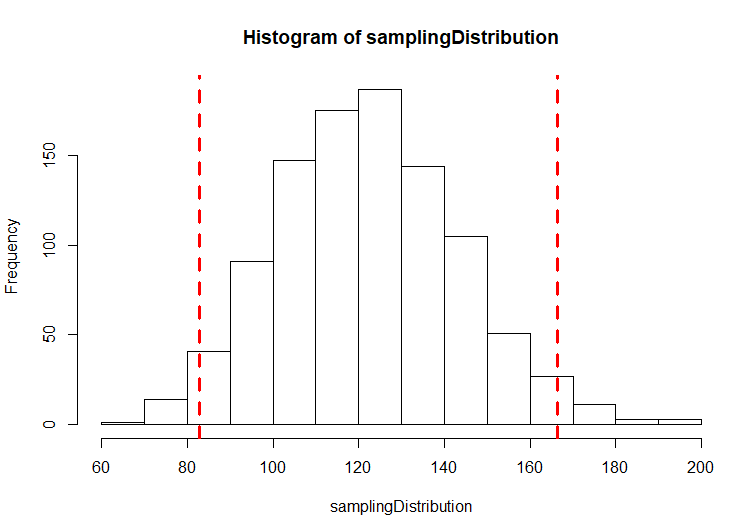
samplingDistribution<-replicate(1000,mean(sample(myChkWts,size=11,replace = TRUE)),simplify = TRUE)

hist(samplingDistribution)



abline(v=quantile(samplingDistribution,.025),col="red", lwd=3, lty=2)

abline(v=quantile(samplingDistribution,.975),col="red", lwd=3, lty=2)



6)

The means of means provide a normal distribution, whereas the exercise 4 quantiles are based upon one sample of raw data. The means of means will transform the shape of the histogram into a normal distributed figured and reduces the skewed histogram found in the raw data. The range of observed values is narrower by using the means of means shrinking the width of the quartiles and minimizing the effect of outlier or extreme samples.

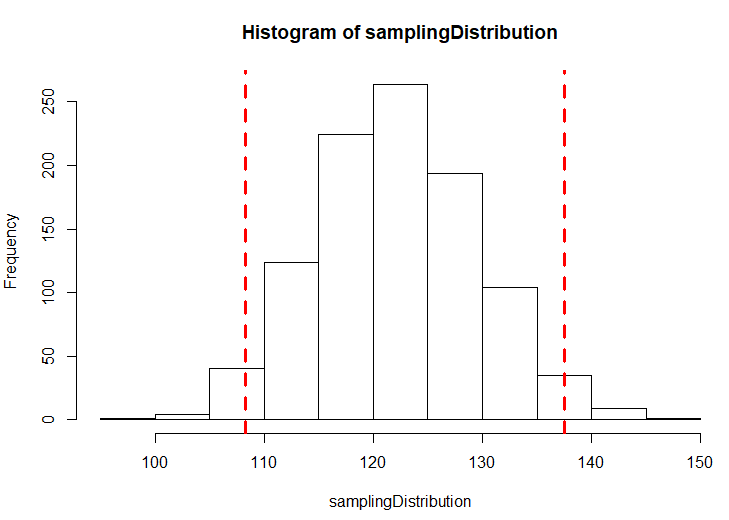
7)

samplingDistribution<-replicate(1000,mean(sample(myChkWts,size=100,replace = TRUE)),simplify = TRUE)

hist(samplingDistribution)

abline(v=quantile(samplingDistribution,.025),col="red", lwd=3, lty=2)

abline(v=quantile(samplingDistribution,.975),col="red", lwd=3, lty=2)



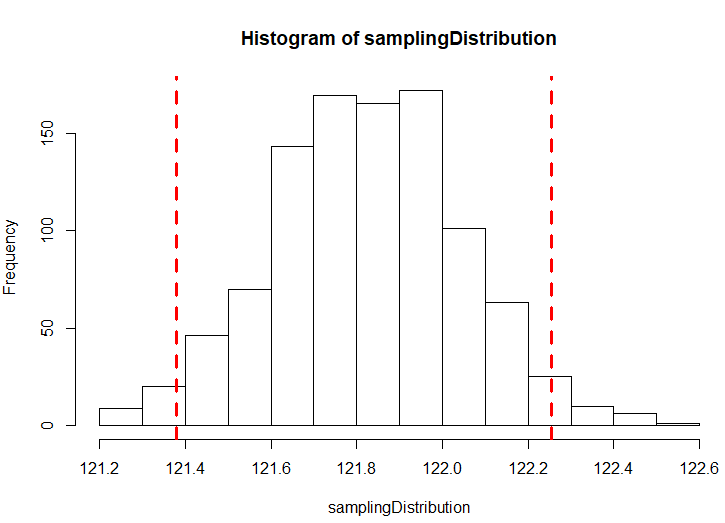
Again, the width of the quantiles narrows. With increased sample size comes a diminished impact by outlier or extreme observed values. With each tenfold increase in sample size, the quantile range will decrease. The distribution will continuously converge around the true mean of the population.

samplingDistribution<-replicate(1000,mean(sample(myChkWts,size=100000,replace = TRUE)),simplify = TRUE)

hist(samplingDistribution)

abline(v=quantile(samplingDistribution,.025),col="red", lwd=3, lty=2)

abline(v=quantile(samplingDistribution,.975),col="red", lwd=3, lty=2)



samplingDistribution<-replicate(1000,mean(sample(myChkWts,size=10000000,replace = TRUE)),simplify = TRUE)

hist(samplingDistribution)

abline(v=quantile(samplingDistribution,.025),col="red", lwd=3, lty=2)

abline(v=quantile(samplingDistribution,.975),col="red", lwd=3, lty=2)

