Homework 3
Syracuse University
IST 772
Summer 2021

Question 2

```
# summarize the chickweights dataset
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
```

```
# there are four variables
    # weight
    # time
    # chick
    # diet

# dimensions of the chickweights dataset
dim(ChickWeight)
```

```
## [1] 578   4
```

```
# the first number reflects the number of rows - 578
# it signifies how many observations there are
```

Question 3

```
summary(ChickWeight$weight)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    35.0    63.0   103.0   121.8   163.8   373.0
```

```
# provides the min, 1st quartile, median, mean, 3rd quartile, and max of the
weight variable

head(ChickWeight$weight)
```

```
## [1] 42 51 59 64 76 93
```

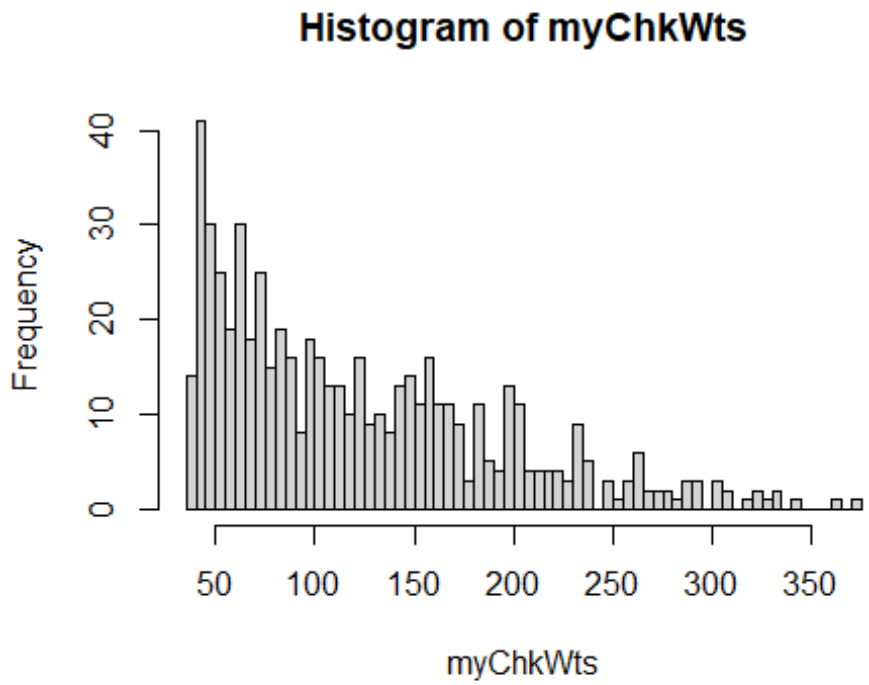
```
# shows the first 6 observations of the weight variable

mean(ChickWeight$weight)
```

```
## [1] 121.8183
```

```
# shows the mean of the weight variable

myChkWts <- ChickWeight$weight
# stores the weight variable into a new variable

quantile(myChkWts, 0.50)

## 50%
## 103

# provides the number where .50 quantile is located
```

Question 4

```
# create a histogram for the myChkWts variable
hist(myChkWts, breaks = 50)
```

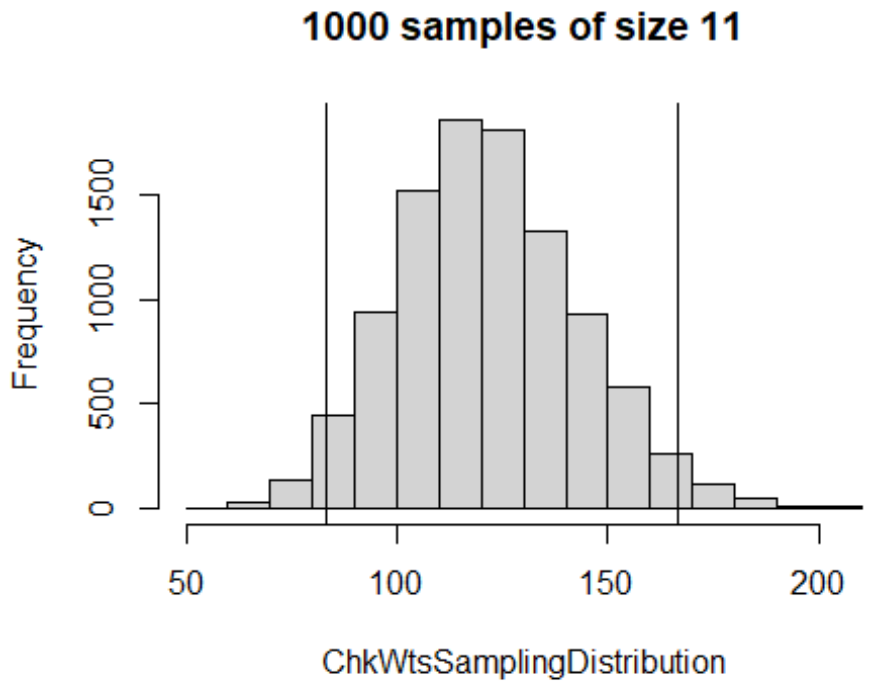



```
# display the 2.5 and 97.5 quantiles
quantile(myChkWts, c(0.025, 0.975))

##    2.5%   97.5%
##  41.000 294.575

# this is a right skewed distribution because the mean is greater than the
# median. It is not a normal distribution.

# the 2.5 quantile of 41 means that 2.5% of all of the data falls below
```

```
# the number 41. Likewise, the 97.5 quantile of 294.575 means that the top
# 2.5% of all of the data falls above the number 294.575.
```

Question 5

```
# create a sampling distribution from chick weights
ChkWtsSamplingDistribution <- replicate(10000, mean(sample(myChkWts, size =
11, replace = TRUE)))

# show a histogram of the sampling distribution
hist(ChkWtsSamplingDistribution, main = '1000 samples of size 11')
abline(v = quantile(ChkWtsSamplingDistribution, 0.025))
abline(v = quantile(ChkWtsSamplingDistribution, 0.975))
```



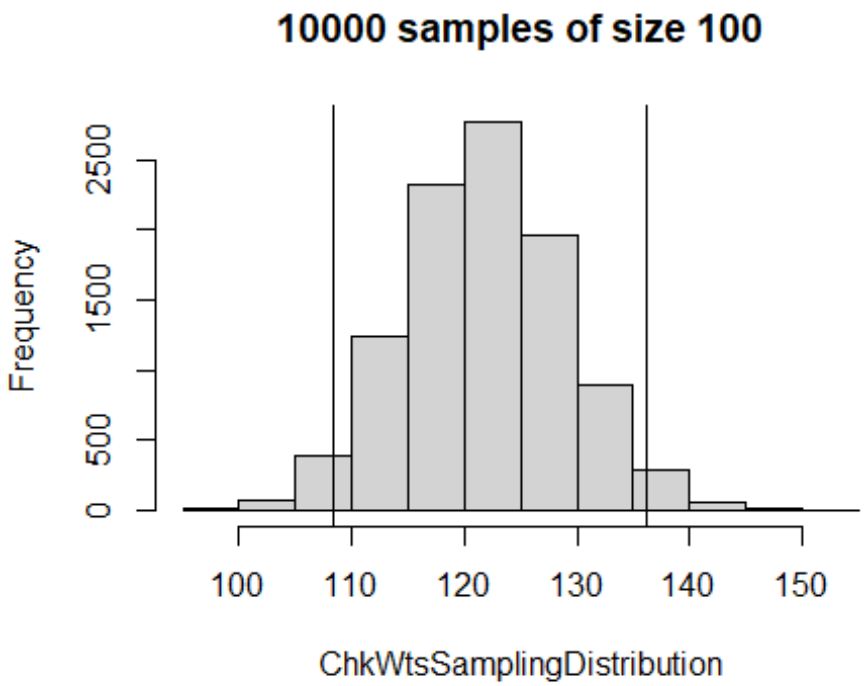
Question 6

```
# the sampling distribution is different because by taking the mean of many
# samples, it becomes a normal distribution which converges on the mean
# of the raw data. The Quantiles are different because the sampling
distribution
# is still taking the mean of a sample.
```

Question 7

```
# create a sampling distribution from chick weights
ChkWtsSamplingDistribution <- replicate(10000, mean(sample(myChkWts, size =
100, replace = TRUE)))

# show a histogram of the sampling distribution
hist(ChkWtsSamplingDistribution, main = '10000 samples of size 100')
```

```r
abline(v = quantile(ChkWtsSamplingDistribution, 0.025))
abline(v = quantile(ChkWtsSamplingDistribution, 0.975))
```




```r
# taking a sample of 100 versus a sample of 11 means that the sample will be
# a better representation of the population. With a smaller sample size there
# is a greater chance for sampling error.
```