Zane_Alderfer_HW3

Zane

2023-07-27

```
#question2
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
```

```
dim(ChickWeight)
```

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

The 578 is the number of rows or number of observations in the data set.

```
#question3
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)
```

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

```
mean(ChickWeight$weight)
```

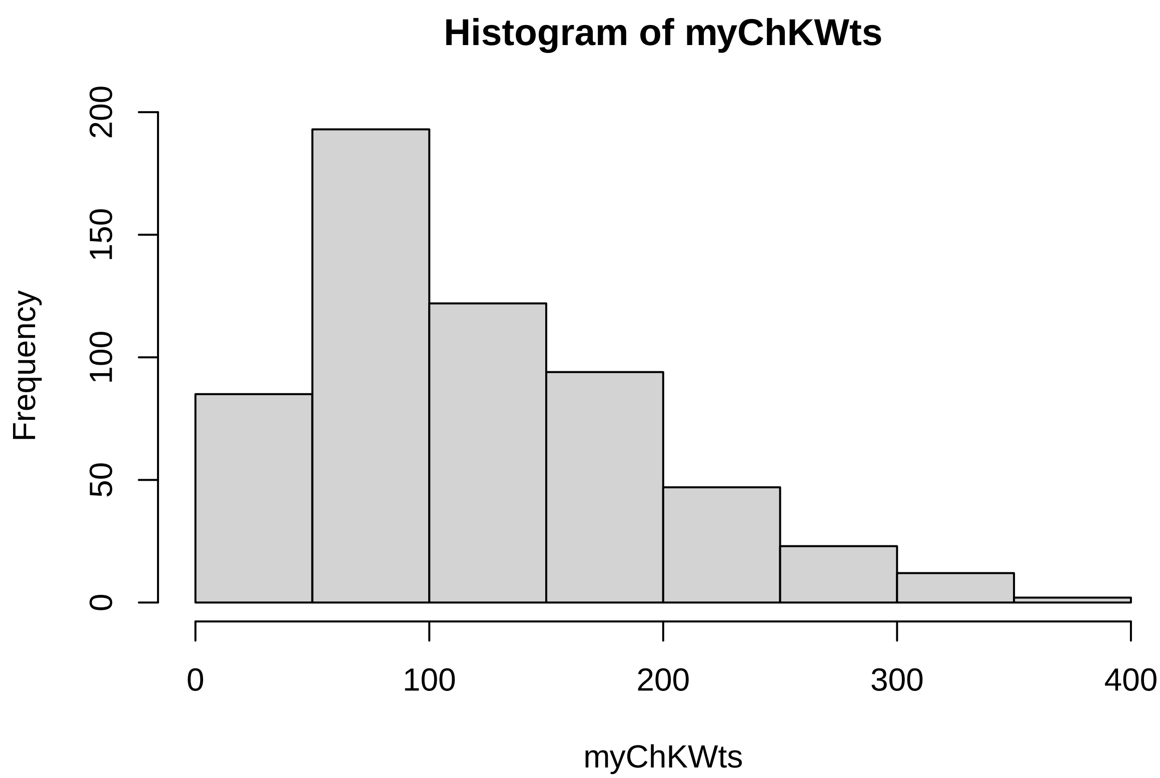
```
## [1] 121.8183
```

```
myChKWts <- ChickWeight$weight
quantile(myChKWts,0.5)
```

```
## 50%
## 103
```

Summary describes the overall dataset describing the column weight with each quantile and with the mean larger than the median, you can assume the graph has a right skew. Head functions gives the first few observations in the dataset. Mean gives the mean which matches the summary's mean and then we create a variable that stores the weight column adn then call the midpoint or 2nd quantile with the following line.

```
#question4
hist(myChKWts)
```




```
quantile(myChKWts,.025)
```

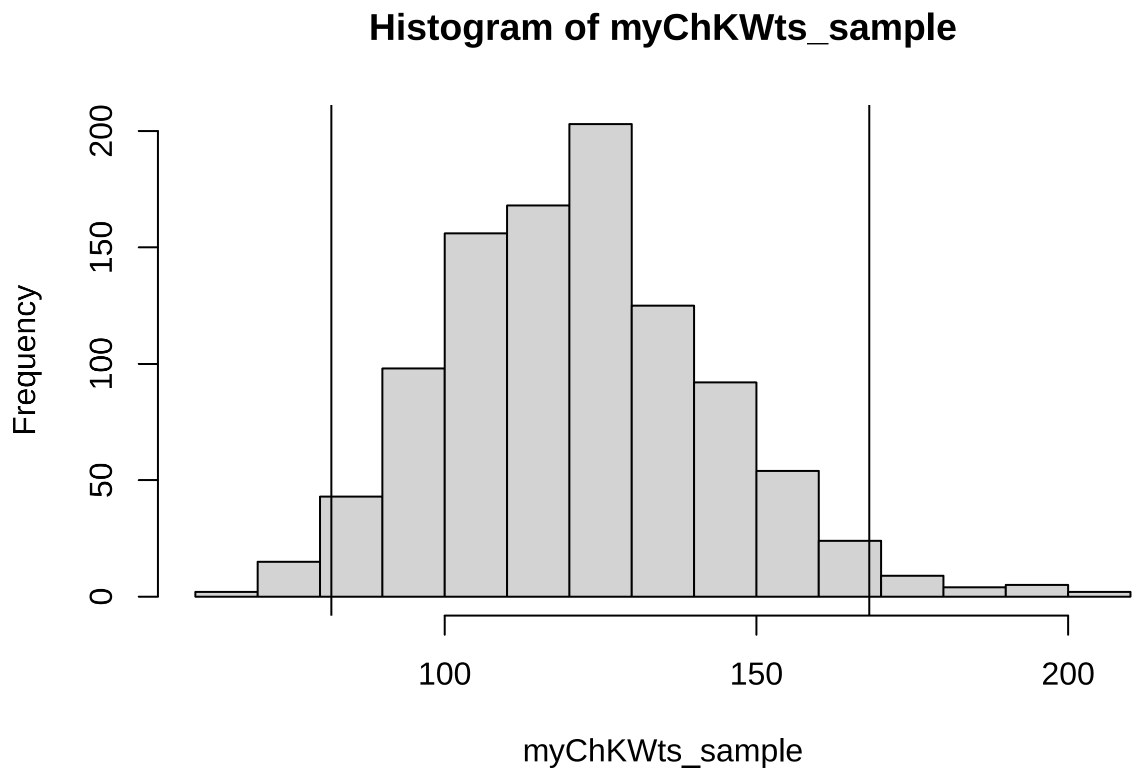
```
## 2.5%
##   41
```

```
quantile(myChKWts,.975)
```

```
##   97.5%
## 294.575
```

As described in number 3, because the mean is larger than the median, there is a right skew. The 2.5% quantile is 41 and the 97.5% quantile is 294.575. These two quantiles describe the two standard deviations to the left and right of the midpoint or 2nd quantile.

```r
#question5
myChKWts_sample <- replicate(1000,mean(sample(myChKWts,size = 11, replace = TRUE)),simplify = TRUE)
hist(myChKWts_sample)
abline(v=quantile(myChKWts_sample,.025))
abline(v=quantile(myChKWts_sample,.975))
```




```r
#question6
quantile(myChKWts,.025)
```

```
## 2.5%
##   41
```

```r
quantile(myChKWts,.975)
```

```
##   97.5%
## 294.575
```

```r
quantile(myChKWts_sample, .025)
```

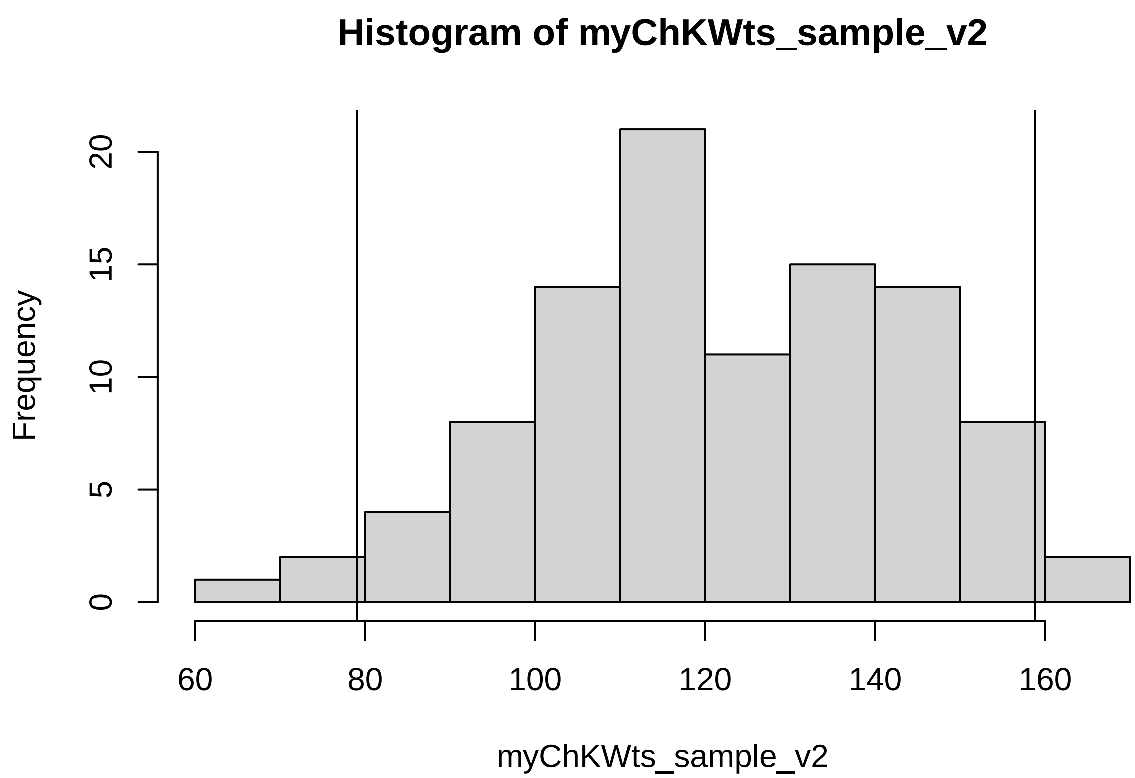
```
##     2.5%
## 81.81591
```

```r
quantile(myChKWts_sample, .975)
```

```
##    97.5%
## 168.1045
```

With the raw data, the means and quantiles are adjusted to account for outliers versus a mean sampling dataset approaches a more normal distribution due to the law of large numbers which essentially eliminates factors such as outliers that skew means and quantiles.

```r
#question7
myChKWts_sample_v2 <- replicate(100,mean(sample(myChKWts,size = 11, replace = TRUE)),simplify = TRUE)
hist(myChKWts_sample_v2)
abline(v=quantile(myChKWts_sample_v2,.025))
abline(v=quantile(myChKWts_sample_v2,.975))
```



As mentioned in exercise 6, the law of large numbers will ultimately bring a sample towards normalcy if there's enough means of a sample dataset being used. For example, with flipping a coin, if you flip it 10 times in 1 trial, you very well could flip 9 heads and 1 tail but if you flipped a coin 10 times with 100 trials, you'll find that mean results of these trials will get closer to 5 heads and 5 tails as the most common result.