HW3_lin

IST 772 Homework 3

Due October 26, 2021 at 8:00AM EDT

Homework 3 by Nora Lin: I produced the material below with no assistance.

Exercise 2 on page 50:

```
#basic information about the 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
```

The four variables are "Weight", "Time", "Chick", and "Diet". According to the help documentation, "Weight" is a numeric vector giving the body weight of the chick(gm). "Time" is a numeric vector giving the number of days since birth wehn the measurement was made. "Chick" is an ordered factor between 18 and 48. The order is based on chicks having the same diet and grouping them by their final weight within their diets. Lastly, "Diet" is a factor with level 1-4, which shows the experimental diet the chick was on.

```
dim(ChickWeight)
```

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

The first output is 578, indicating the number of rows. I think it signifies the unique identifiers for the chicks.

Exercise 3 on page 50:

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

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

Summary command provides basic information about the variable weight within the ChickWeight dataset.The min is 35grams and the max weight was 373grams. 25th percentile of chicks was 63grams, 50th percentile was 103grams and lastly the 75th percentile weight was 163.8. The mean or average weight was 121.8 grams.

```
head(ChickWeight$weight)
```

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

The head command provides the first 6 weights in the ChickWeight dataset.

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

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

The mean of the weights of all the chiks in the ChickWeight dataset is 121.8183 grams. This is the same value as given in the mean of the summary command.

```
myChkWts <- ChickWeight$weight
```

This line stores a numeric vector of 578 values that contains all the values within the "weight" variable in our ChickWeight dataset.

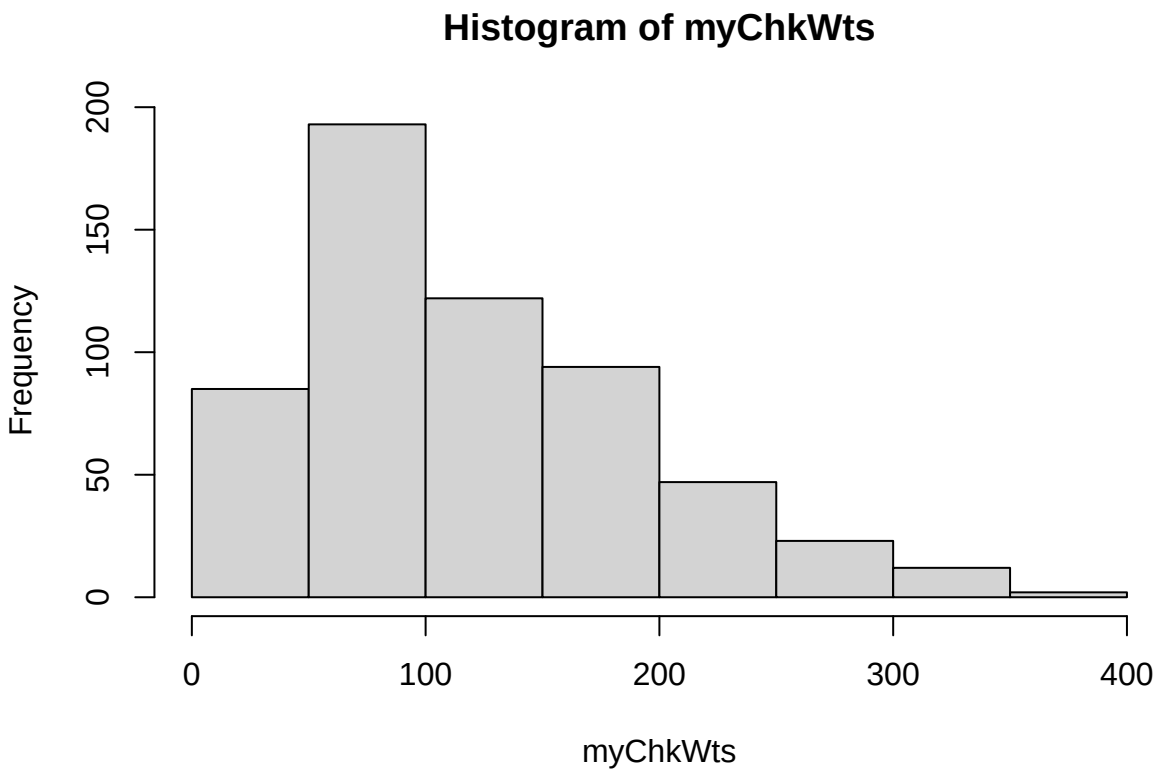
```
quantile(myChkWts,0.50)
```

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

The 50th percentile of chick weights is 103 grams. This is the same value as our summary median value.

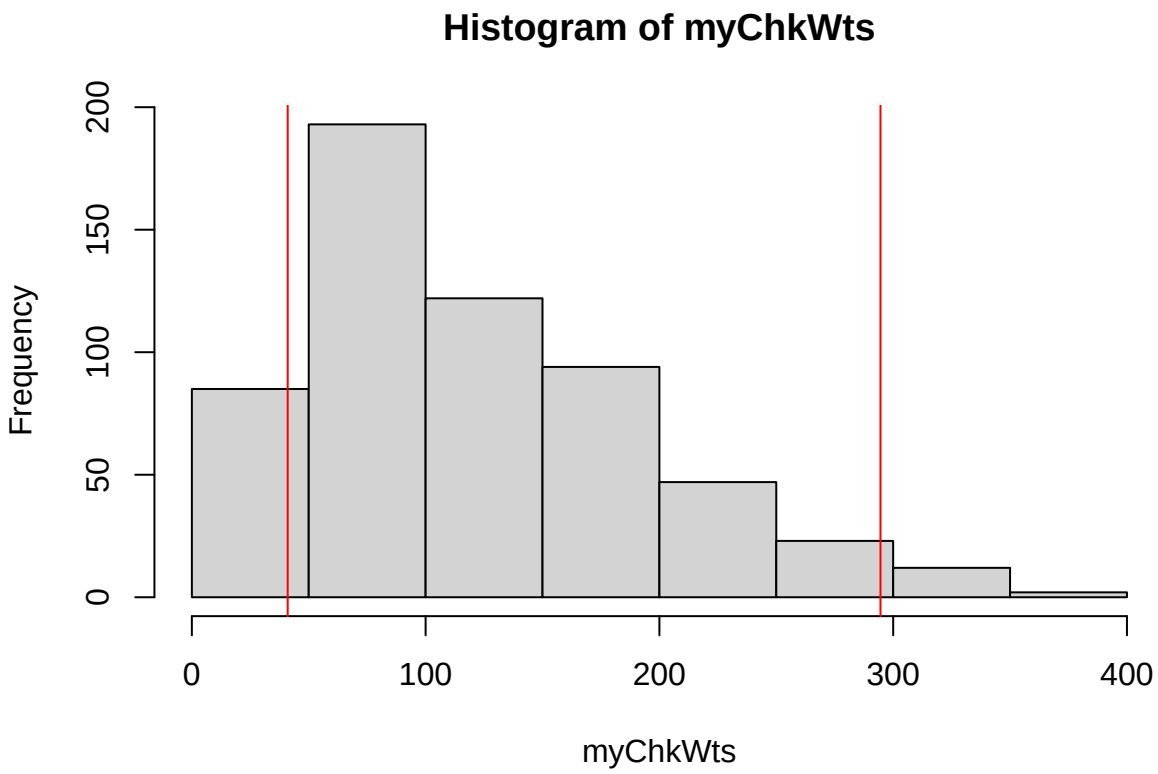
Exercise 4 on page 50:

```
hist(myChkWts)
```



Histogram of the distribution of weights of the chicks in ChickWeight dataset

```
hist(myChkWts)
abline(v=quantile(myChkWts,0.025),col='red')
abline(v=quantile(myChkWts,0.975),col='red')
```

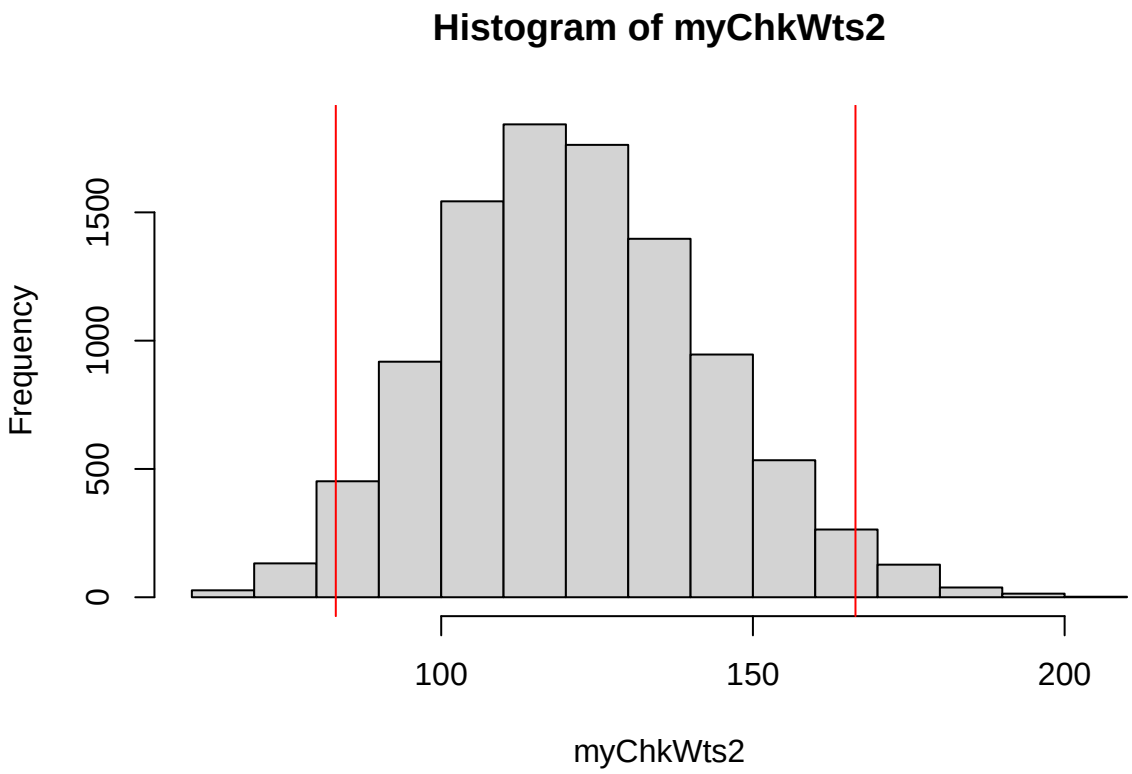



```
quantile(myChkWts,c(0.025, 0.975))
```

```
##    2.5%   97.5%
##  41.000 294.575
```

The 2.5% and 97.5% quantiles are 41grams and 294.575grams respectively. These values represent the tails of the Weights distribution. The shape of the distribution is skewed to the right. The mean is 121 and the median is 103. Since the mean is greater than the median, the distribution is skewed to the right. ## Exercise 5 on page 50:

```
myChkWts2 <- replicate(10000,mean(sample(myChkWts,size=11,replace=TRUE)))
hist(myChkWts2)
abline(v=quantile(myChkWts2,0.025),col='red')
abline(v=quantile(myChkWts2,0.975),col='red')
```




```
quantile(myChkWts2,c(0.025, 0.975))
```

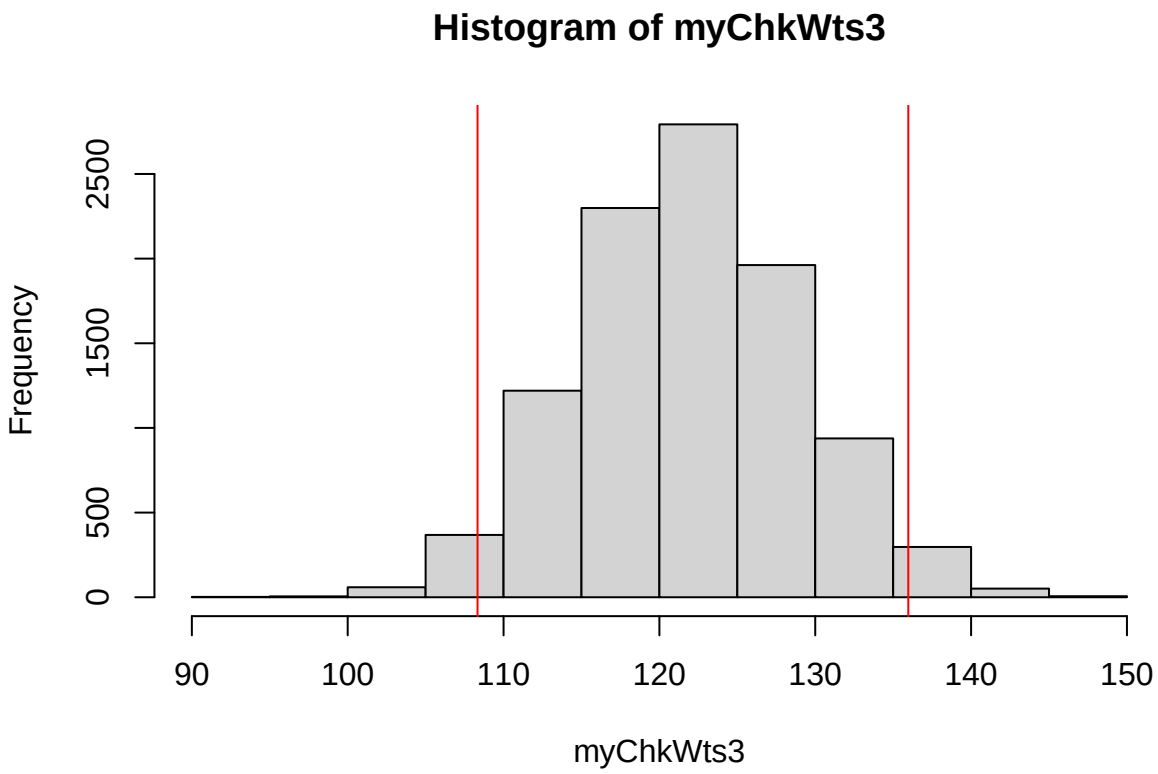
```
##      2.5%     97.5%
##  83.09091 166.45455
```

Exercise 6 on page 51:

2.5% and 97.5% quantiles for the distribution of raw weight data was 41grams and 294grams. 2.5% and 97.5% quantiles for the distribution of sampling mean weight data was 82.5grams and 166grams. The distribution of sampling means provides a picture oft he long-term trends of the sampling. It is a new dataset with solely sample means from the raw data. The quantiles are different between the raw data and the sampling distribution of means because the the raw data captures a limited amount of data, while the sampling distribution is shows the means of 10,000 repititions.

Exercise 7 on page 51:

```
myChkWts3 <- replicate(10000,mean(sample(myChkWts,size=100,replace=TRUE)))
hist(myChkWts3)
abline(v=quantile(myChkWts3,0.025),col='red')
abline(v=quantile(myChkWts3,0.975),col='red')
```




```
quantile(myChkWts3,c(0.025, 0.975))
```

```
##   2.5%  97.5%
## 108.33 135.97
```

The 2.5% and 97.5% quantiles are 108.09grams and 136.14grams respectively. In comparison, this distribution follows the bell curve. While the histrogram from exercise is slightly right skewed. This is because we are using a large sample size of 100 instead of 11. This follows the law of large numbers.

End of Homework 3