Assignment 2

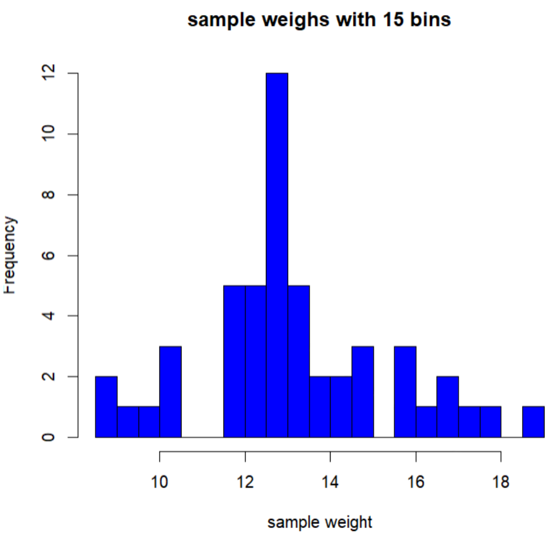
1. Generate a histogram of sample weights with 15 bins. Does the data look normal? Generate a histogram with 7 bins. Does the data look normal?

Solution:

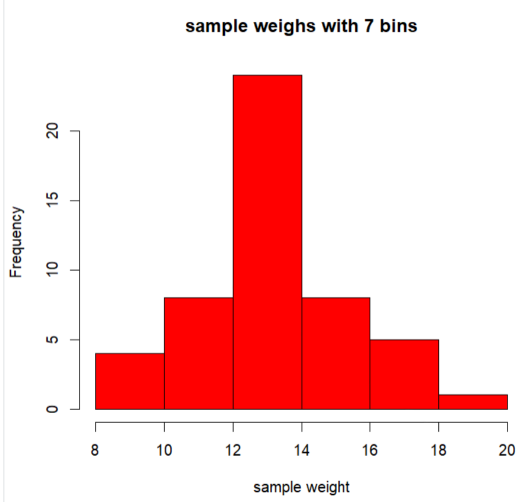
Coding: (using breaks to set bins)

```
#Assignment 2

#I want to know what is working directory
getwd()
#change to working directory I want to
setwd('coding_r/413hw/hw2')
#Using R reading csv file
PgWeight = read.csv('Assignment 2 PenguinWeights.csv')

#Problem 1a generate a histogram of sample weighs with 15 bins
str(PgWeight)
x = PgWeight$PenguinWeights
hist(x,breaks = 15, xlab = 'sample weight', main = 'sample weighs with 15 bins', col = 'blue')
##Problem 1b generate a histogram of sample weighs with 7 bins
hist(x,breaks = 7, xlab = 'sample weight', main = 'sample weighs with 7 bins', col = 'Red')
```

Histgram with15 bins:



Histogram with 7 bins



Because the graph shows that it's very close to normal distribution for both of the two histograms. So, I think the data is follow normal distribution.

2. Determine the 95% confidence interval for mean of the penguin weights.

Solution:

Confidence Interval = $\bar{x} \pm Z_{\alpha/2} * \frac{s}{\sqrt{n}}$, $Z_{0.025} = 1.96$, mean = 13.134, standard deviation = 2.217, n = 50

Confidence Interval = $13.134 \pm 1.96 * \frac{2.217}{\sqrt{50}}$ = (12.5195,13.7485).

Coding:

```
> x_bar = mean(x) #sample mean
> std = sd(x) #sample deviation
> z_value = qnorm(0.975) # get z value
> up_bnd = x_bar + z_value*std/sqrt(50) # get upper bound of 95% confidence interval
> up_bnd
[1] 13.74883
> low_bnd = x_bar - z_value*std/sqrt(50) # get lower bound of 95% confidence interval
> low_bnd
[1] 12.5198
> |
```

From R we can see that our calculation is extremely close to r result, so the confidence interval in R is (12.5198,13.74883).

3. Test the hypothesis that the sample is from a population of penguins with a mean weight of 13.5. Specify the p-value of the test and your interpretation on the test results.

Solution:

(1) By hand:

Null hypothesis: $\mu = 13.5$

Alternate hypothesis: $\mu \neq 13.5$

Let α value be 0.05 for this two-tail test. (This means p-value cannot smaller than 0.05, otherwise we have enough confidence to reject the null hypothesis,)

$Z_{\frac{0.05}{2}} = \pm 1.96$

By using the mean of this sample, we need to know the Z value of this sample mean.

$Z_{test} = \frac{x - \mu}{\sigma/\sqrt{n}} = \frac{13.1343 - 13.5}{2.217/\sqrt{50}} = -1.1664$

$P_{value} = 2 * \Phi^{-1}(x = -1.1664) = 0.2434$

Because test score is -1.1664, it's not less than -1.96 and it's not larger than 1.96. Also p-value is larger than 0.05. So, We do not have enough evidence to reject the null hypothesis.

(2) Using R coding:

```
> t.test(x, mu = 13.5, conf.level = 0.95)

	One Sample t-test

data:  x
t = -1.1663, df = 49, p-value = 0.2491
alternative hypothesis: true mean is not equal to 13.5
95 percent confidence interval:
 12.50424 13.76438
sample estimates:
mean of x 
 13.13431 
```

In this outcome, we can see that the p-value is 0.2491 and it is not less than 0.05. So, we do not reject null hypothesis.