Title: Statistical Inference Project – Assignment II

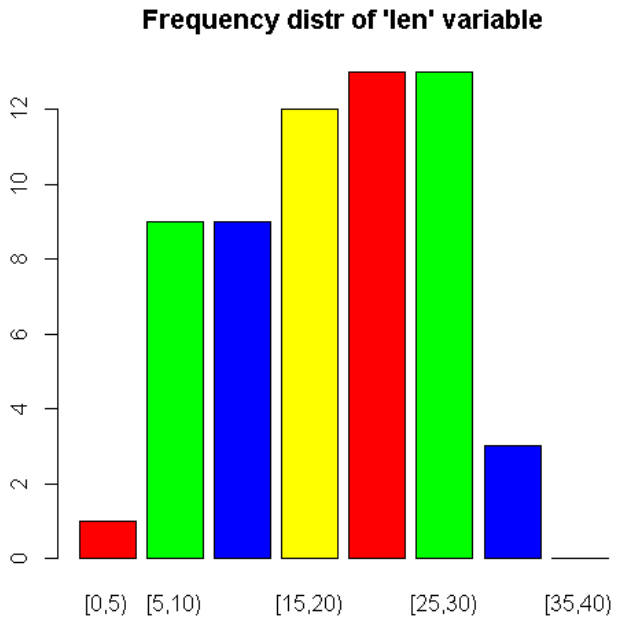
Author: Kashyap Kumar

Date: 28th Aug '16

Goal: To analyze the ToothGrowth data in the R datasets package.

Question 1: Load the ToothGrowth data and perform some basic exploratory data analyses

```
# Load dataset
library(datasets)
data(ToothGrowth)
# Find dimensions and Headers of dataset. Create a dataframe
dim(ToothGrowth)
[1] 60  3
names(ToothGrowth)
[1] "len"  "supp" "dose"
df <- data.frame (ToothGrowth)
# Find frequency distribution of 'len' values
range(df$len)                    # get range of values
[1]  4.2 33.9
breaks <- seq(0,40,by = 5)       # create buckets
freq_len <- table (cut(df$len, breaks, right = FALSE))     # create frequency table with open ended Right ends
freq_len
[0,5)  [5,10) [10,15) [15,20) [20,25) [25,30) [30,35) [35,40)
    1      9      9       12       13       13        3      0
# Draw bar chart
barplot(freq_len, col = c("red", "green", "blue", "yellow"), main = "Frequency distr of 'len' variable")
```

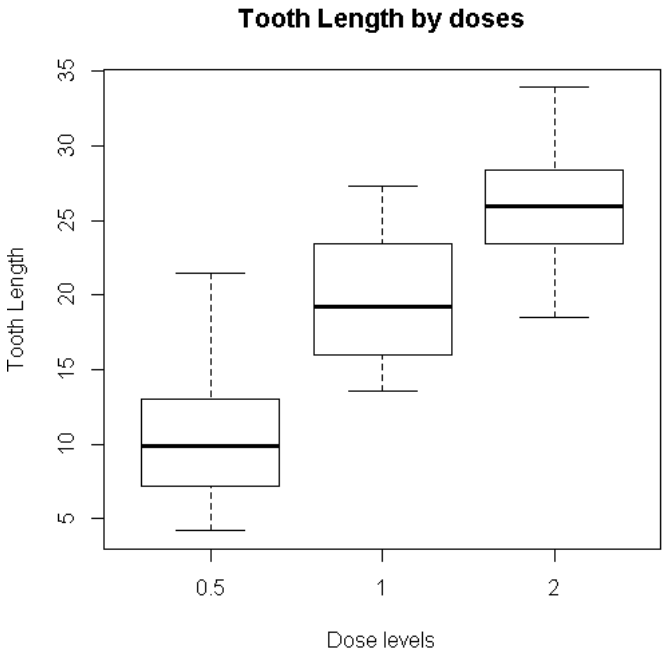


The distribution is a negatively-skewed distribution. As is shown by the result below

```
library(e1071)

skewness (df$len)

[1] -0.1425376
```

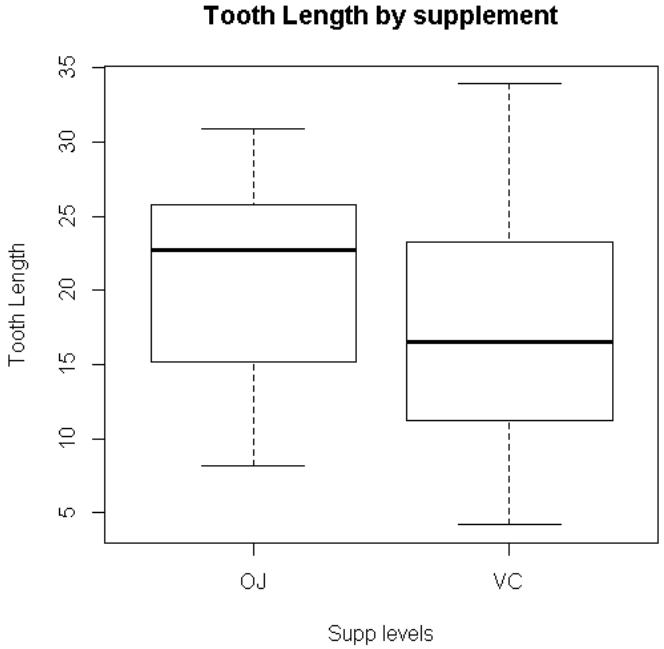
Question 2: Provide a basic summary of the data.

```
# Get basic summary figures

summary (df)

          len       supp       dose
 Min.   : 4.20   OJ:30   Min.   :0.500
 1st Qu.:13.07   VC:30   1st Qu.:0.500
 Median :19.25           Median :1.000
 Mean   :18.81           Mean   :1.167
 3rd Qu.:25.27           3rd Qu.:2.000
 Max.   :33.90           Max.   :2.000


# Compare Tooth length as per dosage levels

boxplot(len~dose, data=df, main="Tooth Length by doses", ylab="Tooth Length", xlab="Dose levels")
```



The box-plot graph indicates a strong correlation between Tooth-length and dosage levels. The length increases with the dosage.

```
# Compare Tooth length as per supplement levels

boxplot(len~supp, data=df, main="Tooth Length by supplement", ylab="Tooth Length", xlab="Supp levels")
```



The box-plot graph indicates that most of Tooth-length data is evenly distributed among OJ and VC, length being in similar range, hence signifying independence.

Question 3: Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)

```
# Compare tooth growth by supp, use the Welch two sample t-test on OJ and VC

# H0: There is no difference in Tooth Growth whether it's OJ or VC.

t.test (len~supp, data = df)plot

Welch Two Sample t-test

data:  len by supp

t = 1.9153, df = 55.309, p-value = 0.06063

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

 -0.1710156  7.5710156

sample estimates:

mean in group OJ   mean in group VC

    20.66333          16.96333
```

The P-Value for this test is 0.06, which is greater than 0.05 and signifies: We have insufficient evidence to reject H0. 95% confidence interval is (-0.171,7.571) which contains 0 – which weakens the possibility of Supplements OJ and VC having different impacts on Tooth-Length.

```
### Compare tooth growth by two levels of dose, use the Welch two sample t-test

# Test tooth length against dose levels of 0.5 and 1.0, use the Welch two sample t-test

# H0: No difference in impact on Tooth length, of dose levels

t.test (len~dose, data = df[df$dose==0.5 | df$dose==1,])

Welch Two Sample t-test

data:  len by dose

t = -6.4766, df = 37.986, p-value = 1.268e-07

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

 -11.983781  -6.276219

sample estimates:

mean in group 0.5   mean in group 1

       10.605          19.735
```

95% confidence interval: (-11.984,-6.276) - does not have 0 within; It signifies bias in impact of dose levels on Tooth length. Dose of 0.5 has less impact compared to dose of 1.0, as shown by negative CI.

p-value: 1.268e-07 < 0.05, Hence, we can reject the null hypothesis of “No difference in impact”.

```
# Test tooth length against dose levels of 1.0 and 2.0, use the Welch two sample t-test

# H0: No difference in impact on Tooth length, of dose levels

t.test (len~dose, data = df[df$dose==1 | df$dose==2,])

Welch Two Sample t-test

data:  len by dose

t = -4.9005, df = 37.101, p-value = 1.906e-05

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

 -8.996481  -3.733519

sample estimates:

mean in group 1  mean in group 2

       19.735         26.100
```

95% confidence interval: (-8.996481, -3.733519)

p-value: -8.996481 -3.733519 < 0.05.

The results are similar to those of the previous test.

Hence, we can reject the null hypothesis of “No difference in impact”.

Conclusion: 1. Tooth lengths are not impacted by supplement type; OJ / VC

2. Increase in dosage amounts, leads to increase in Tooth length.