Report 2

Report 2

Intro

This document is a report to submit analysis asked for in the Coursera Statistical Inference Project, Question 1.

Question

Now in the second portion of the class, we're going to analyze the *ToothGrowth* data in the *R datasets* package.
1. Load the *ToothGrowth* data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
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**)
4. State your conclusions and the assumptions needed for your conclusions.

Solution

Let's load and summarize the data

```
# load and summarize data
library(datasets)
data(ToothGrowth)
summary(ToothGrowth)
```

```
##       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
```

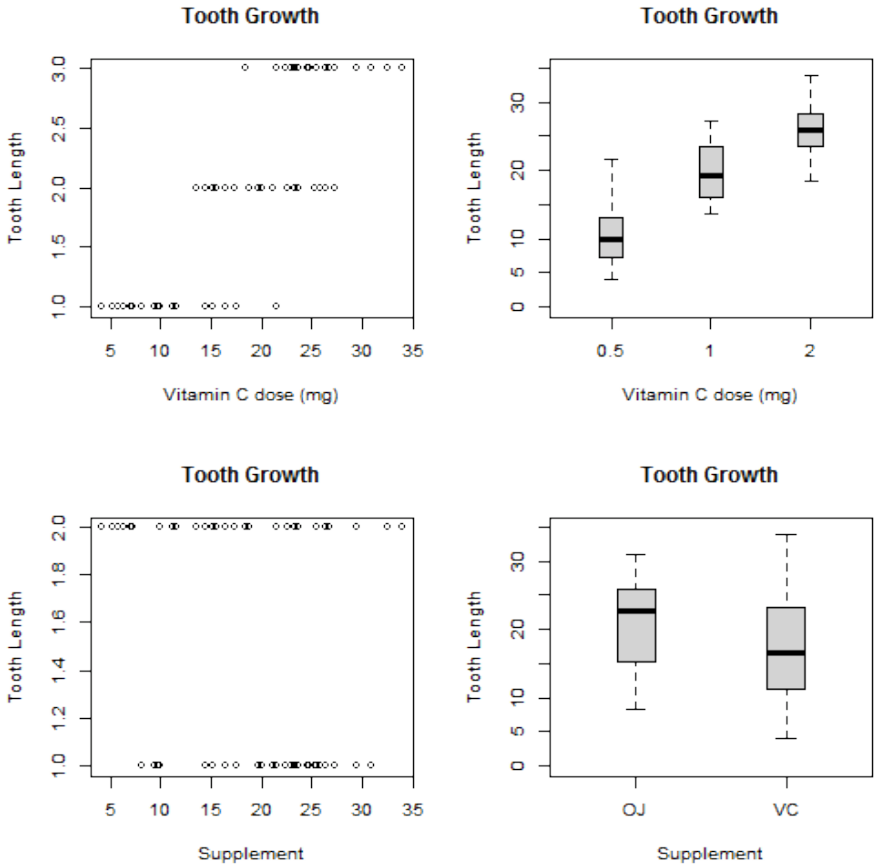
For analysis to proceed, we have to understand the variables. They are:

- **len**: Tooth length (*numeric*)
- **supp**: Supplement type (*factor*);
- **dose**: Dose of vitamin C (*numeric*). Next, lets convert dose to a factor variable

```
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
```

Let's get a good idea of the data by drawing some plots

```
par(mfrow=c(2,2))
plot(ToothGrowth$len,ToothGrowth$dose,ylab="Tooth Length",xlab="Vitamin C dose (mg)",main="T
with(ToothGrowth,boxplot(len~dose,boxwex=0.25,col="lightgray",ylim=c(0,35),ylab="Tooth Lengt
plot(ToothGrowth$len,ToothGrowth$sup,ylab="Tooth Length",xlab="Supplement",main="Tooth Growt
with(ToothGrowth,
     boxplot(len~supp,boxwex=0.25,col="lightgray",ylim=c(0,35),
             ylab="Tooth Length",xlab="Supplement",main="Tooth Growth"))
```



From the plots it seems like dose has a direct correlation with tooth length whereas no such relation seems to exist between supplement and length

The results of the hypothesis are shown below. All tests are performed at the 0.05 significance level.

```
# Hypothesis test regarding differences between the two levels of supp:
with(ToothGrowth,t.test(len~supp))
```

```
##
##  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
```

```
# Hypothesis tests regarding differences between the levels of dose, pairwise:
with(ToothGrowth,pairwise.t.test(len,dose,p.adjust="none"))
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  len and dose
##
##   0.5     1
## 1 6.7e-09 -
## 2 < 2e-16 1.4e-05
##
## P value adjustment method: none
```

From the results shown above we can derive the following conclusions:

- Cannot reject null hypothesis at 0.05 significance;
- There is sufficient evidence to reject the null hypotheses in all comparisons of levels of the factor **dose**, at the 0.05 significance level.

The above results suggest that the mean tooth growth is directly correlated with the doese amount and increases with increase in dosage