Statistical Inference Course Project: Basic Inferential Data Analysis

This is the project for the statistical inference class. The project consists of two parts:

1. Simulation exercises.
2. Basic inferential data analysis.

Basic Inferential Data Analysis

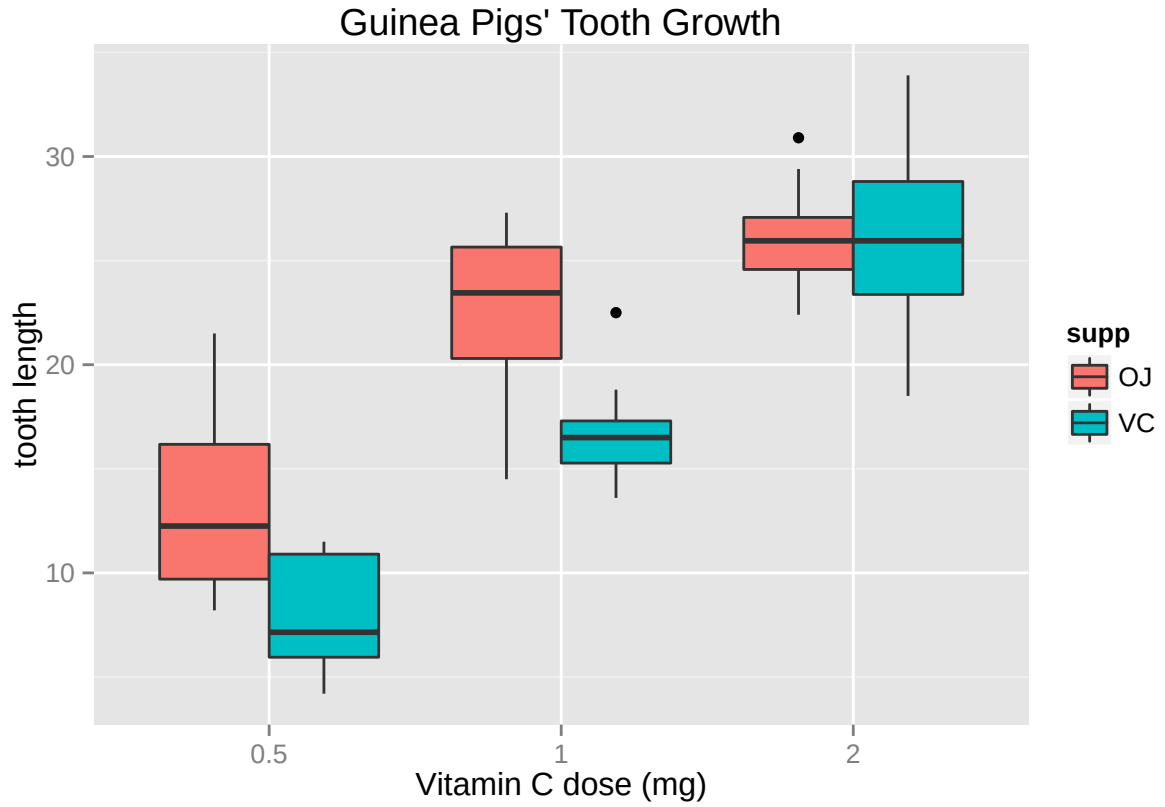
Now in the second portion of the class, we're going to analyze the ToothGrowth data in the R datasets package.

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 hypothesis tests to compare tooth growth by supp and dose. (Use the techniques from class even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

1. Load the ToothGrowth data and perform some basic exploratory data analyses and 2. Provide a basic summary of the data. The growth dataset looks at the effect of vitamin C on tooth growth in guinea pigs. The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

```
if (!require(datasets)) install.packages('datasets'); require(datasets)
if (!require(ggplot2)) install.packages('ggplot2'); require(ggplot2)
if (!require(plyr)) install.packages('plyr'); require(plyr)
if (!require(reshape2)) install.packages('reshape2'); require(reshape2)

ggplot(ToothGrowth,(aes(x = factor(dose), y = len, fill = supp))) + geom_boxplot()+xlab("Vitamin C dose
```




```
cdata <- ddply(ToothGrowth, c("dose", "supp"), summarise,
               N    = length(len),
               min  = min(len),
               q.25 = quantile(len, c(0.25)),
               mean = mean(len),
               median = median(len),
               q.75 = quantile(len, c(0.75)),
               max = max(len),
               sd   = sd(len),
               se   = sd / sqrt(N) )
```

As seen in the box plot above, the general relationship between tooth length and Vitamin C dose is positive. As the dosage of either supplement administered increases, the tooth length increases. At lower dosages, orange juice seems to outperform the ascorbic acid supplememt. A summary of the boxplot values can be seen in the table below.

Basic summary data

```
##   dose supp  N  min  q.25  mean median  q.75  max    sd     se
## 1  0.5   OJ 10  8.2  9.70 13.23  12.25 16.18 21.5 4.460 1.4103
## 2  0.5   VC 10  4.2  5.95  7.98   7.15 10.90 11.5 2.747 0.8686
## 3  1.0   OJ 10 14.5 20.30 22.70  23.45 25.65 27.3 3.911 1.2368
## 4  1.0   VC 10 13.6 15.27 16.77  16.50 17.30 22.5 2.515 0.7954
## 5  2.0   OJ 10 22.4 24.57 26.06  25.95 27.08 30.9 2.655 0.8396
## 6  2.0   VC 10 18.5 23.38 26.14  25.95 28.80 33.9 4.798 1.5172
```

3. Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose. (Use the techniques from class even if there's other approaches worth considering) Since the groups are independent, we will be using t confidence intervals. This interval assumes unequal variance across the two groups. We will compare tooth growth by supp and dose.

```
dose.5<-ToothGrowth[ToothGrowth$dose==0.5,]
dose1<-ToothGrowth[ToothGrowth$dose==1.0,]
dose2<-ToothGrowth[ToothGrowth$dose==2.0,]
d1<-t.test(len~supp , paired = FALSE,var.equal = FALSE, data = dose.5)$conf
d2<-t.test(len~supp , paired = FALSE,var.equal = FALSE, data = dose1)$conf
d3<-t.test(len~supp , paired = FALSE,var.equal = FALSE, data = dose2)$conf
```

We will test for the null hypothesis: mean OJ is equal to mean VC and alternate hypothesis: mean OJ is not equal to mean VC for each of the doses.

Performing the t-test for dose 0.5 mg, the intervals are 1.7191 and 8.7809, suggesting that the means are not equal and that the tooth length tends to be longer for the OJ supplements. Therefore, we reject the null hypothesis and accept the alternate hypothesis.

Performing the t-test for dose 1.0 mg, the intervals are 2.8021 and 9.0579, suggesting that the means are not equal and that the tooth length tends to be longer for the OJ supplements. Therefore, we reject the null hypothesis and accept the alternate hypothesis.

Performing the t-test for dose 2.0 mg, the intervals are -3.7981 and 3.6381, suggesting that the the differences between the two datasets can possibly have a zero difference. Thus, there is a possibility that the means could be equal. Therefore, we accept the null hypothesis and reject the alternate hypothesis.

4.State your conclusions and the assumptions needed for your conclusions. In this analysis we performed t-tests using the confidence interval of 95%. This assumes unequal variance across the two groups.

This analysis statistically corroborates the initial assessments in Question 1 and 2. There is a difference for tooth length with supplement type and dosages under 2mg. Guinea pigs that received dosages of <2mg of the orange juice supplement tended to have longer teeth when compared to the ascorbic acid supplement of the same dosage. However, at the max dose of 2mg both supplements performed the same.