Statistical Inference CP1 Part1

Instructions

Analyze the Tooth Growth dataset provided in the R datasets package.

Steps to be competed:

```
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.
```

Exploratory Data Analysis

The fist step is to load the dataset from the existing R libraries. This dataset contains the results from a tooth growth study that measured the tooth growth in quinea pigs who were administered vitamin C.

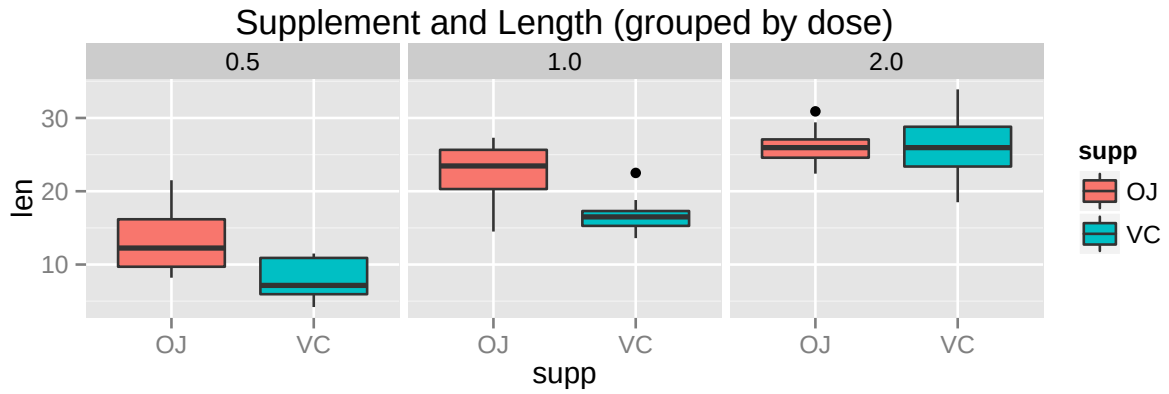
```
library(datasets)
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

By using the STR function, the data set reveals 3 Fields Len, Supp, and Dose over a time frame of 60 observations.

Out of the 60 Observations, what's the split of observations my dose and Suppliment?

As show in the Appendix, the sample size is the same for each dose and supplement combination.



Looking at a box plot comparing tooth growth by supplement dose, it can be seen that the amount of each supplement can have am impact on tooth growth.

Confidence Intervals

The question? Is there a significant statistical difference of tooth growth when compared dose amount between the two supplements. We will utilize the t confidence interval to help determine this. We will make a comparison for each dose level (0.5, 1.0, 2.0). It is assumed that the 2 test were performed on the same 10 guinea pigs. This allow us to use the difference for our calculations. See Appendix for Calculation ofMean and sd.

Confidence Interval: See Appendix for Calc

Confidence Interval for 0.5 mg Dose: 1.2635 9.2365

Confidence Interval for 1.0 mg Dose: 1.2635 9.2365

Confidence Interval for 2.0 mg Dose: -4.329 4.169

Based on the 95% confidence intervals, teh 2.0 dose contains 0 while the 0.5 and 1.0 does not. This means that there is not a statistical significant difference between either methods of treatment OJ or VC in the 2.0 mg dose. However, that does not apply for the 0.5 and 1.0 does. This supports the data as represented in the box plat above.

Hypothesis Testing

To test the hypothesis that dose does have an effect on teeth growth, we can again use the t-test to get this information.

```
Stat_05<-t.test(Diff_05)$statistic
Stat_10<-t.test(Diff_10)$statistic
Stat_20<-t.test(Diff_20)$statistic
```

Hypothesis Test for 0.5 mg Dose: 2.9791

Hypothesis Test for 1.0 mg Dose: 2.9791

Hypothesis Test for 2.0 mg Dose: -0.0426

Based on the t-test the 2.0 mg dose is the only one we can reject as saying: A 2.0 mg dose of either OJ or VC does not indicate a statistical significant difference in tooth grown in guinea pigs.

Conclusions

As shown in both the Confidence Interval and Hypothesis testing, 0.5 mg and 1.0 mg dose does show signs of statistical significant difference in growth of puinea pig teeth. However the 2.0 mg dose does not have a statistical significant difference in growth.

Apendix

Method used to split data

```
OJ_05 <- ToothGrowth[ToothGrowth$dose == 0.5 & ToothGrowth$supp == 'OJ', 1]
VC_05 <- ToothGrowth[ToothGrowth$dose == 0.5 & ToothGrowth$supp == 'VC', 1]
OJ_10 <- ToothGrowth[ToothGrowth$dose == 1.0 & ToothGrowth$supp == 'OJ', 1]
VC_10 <- ToothGrowth[ToothGrowth$dose == 1.0 & ToothGrowth$supp == 'VC', 1]
OJ_20 <- ToothGrowth[ToothGrowth$dose == 2.0 & ToothGrowth$supp == 'OJ', 1]
VC_20 <- ToothGrowth[ToothGrowth$dose == 2.0 & ToothGrowth$supp == 'VC', 1]
```

Size of each sample:

OJ 0.5 milligrams = 10 Samples

VC 0.5 milligrams = 10 Samples

OJ 1.0 milligrams = 10 Samples

VC 1.0 milligrams = 10 Samples

OJ 2.0 milligrams = 10 Samples

VC 2.0 milligrams = 10 Samples

t-Test Calculations

```
n <- 10

#Using the difference of each supp dose, Calculate mean and STd
Diff_05 <- OJ_05 - VC_05
Mean_05 <- mean(Diff_05)
std_05 <- sd(Diff_05)

Diff_10 <- OJ_05 - VC_05
Mean_10 <- mean(Diff_10)
std_10 <- sd(Diff_10)

Diff_20 <- OJ_20 - VC_20
Mean_20 <- mean(Diff_20)
std_20 <- sd(Diff_20)
```

Confidence Calc

```
Conf_05<-t.test(Diff_05)$conf
Conf_10<-t.test(Diff_10)$conf
Conf_20<-t.test(Diff_20)$conf
```