JHU Statistical Inference Course Course Project Part 2

Part 2: Basic Inferential Data Analysis Instructions

The Instructions for part 2 of the project are as follows:

> Now in the second portion of the project, 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.

Question 1 & 2

First off lets load the data and libraries:

```
data("ToothGrowth")
library(ggplot2)
```

Lets start by taking a look at the head of the data and a summary to get a feel for what it contains:

```
head(ToothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
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
```

The data appears to be contain three columns. The first is the length of the tooth, the second is the name of a supplement and third is the supplement dosage.

Lets use table to see what the breakdown for the supplement and dosage:

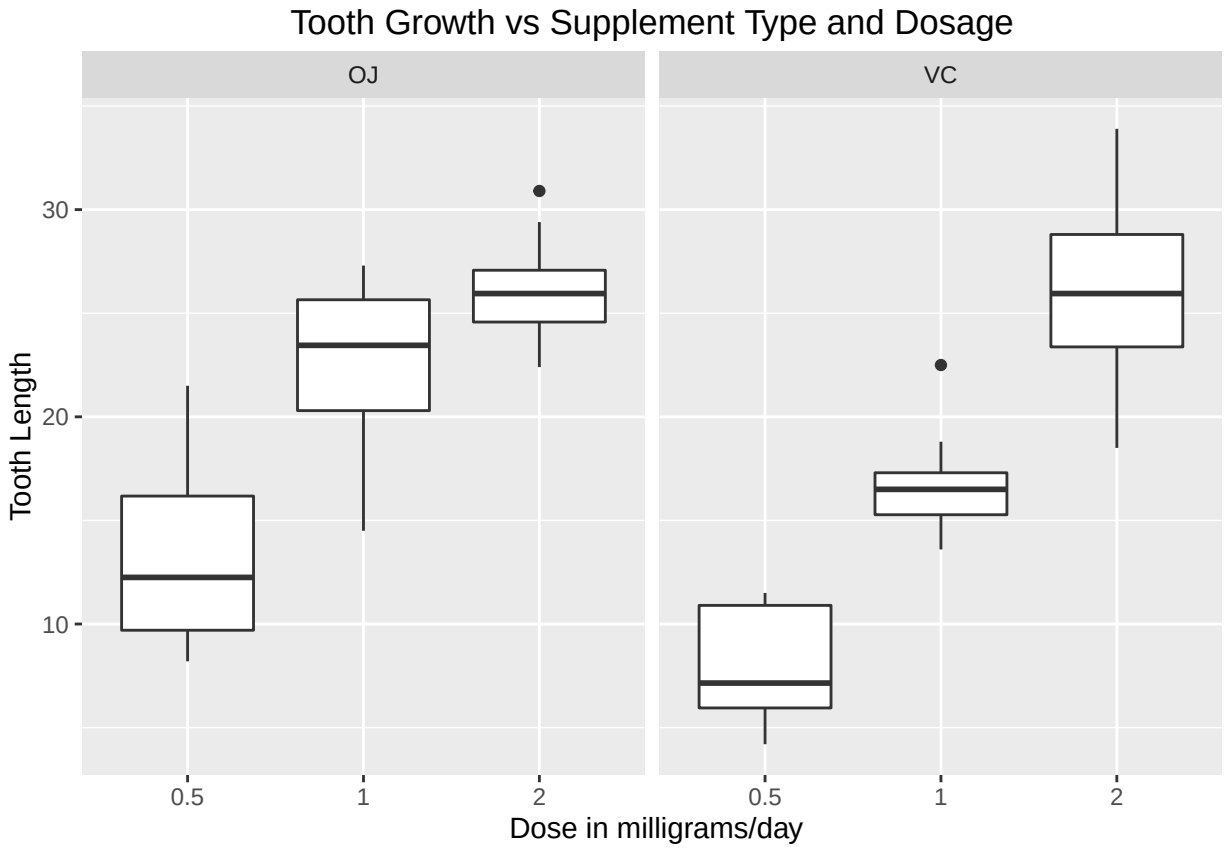
```
table(ToothGrowth$supp,ToothGrowth$dose)
```

```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

So we appear to have 3 dosages(0.5,1.0,2.0) for each supplement with 10 samples each.

Next we can use a box plot to get a feel for the relationship between supplement type and dosage to tooth length.

```
ggplot(ToothGrowth, aes(x=factor(dose), y=len))+
  facet_grid(.~supp)+xlab(" Dose in milligrams/day")+
  ylab("Tooth Length")+
  guides(fill=FALSE)+geom_boxplot()+
  ggtitle("Tooth Growth vs Supplement Type and Dosage")
```



Question 3

Next we are going to perform t-tests to determine if any significant differences exist between the two supplements and within the dosages for each supplement.

We will compare supplements at all dosages and perform a t-test to determine if any significant difference exists between them.

```
t.test(len~supp, data=ToothGrowth)
```

```
##
##  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 t-test returned a p-value of ***0.061***. For a 95% two-sided significance level we would accept the null hypothesis that there is not a difference in tooth length between the two supplements.

We will next compare the dosages within each supplement to see if any significant differences exist. First subset the data by supplement and pairs of doses.

```
vcData12<-subset(ToothGrowth, supp == "VC" & dose != 2)
vcData13<-subset(ToothGrowth, supp == "VC" & dose != 1)
vcData23<-subset(ToothGrowth, supp == "VC" & dose != 0.5)



ojData12<-subset(ToothGrowth, supp == "OJ" & dose != 2)
ojData13<-subset(ToothGrowth, supp == "OJ" & dose != 1)
ojData23<-subset(ToothGrowth, supp == "OJ" & dose != 0.5)
```

Perform t-tests on VC data.

```
t.test(len~dose, data=vcData12)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -7.4634, df = 17.862, p-value = 6.811e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.265712  -6.314288
## sample estimates:
## mean in group 0.5   mean in group 1
##              7.98             16.77
```

```
t.test(len~dose, data=vcData13)
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  len by dose
## t = -10.388, df = 14.327, p-value = 4.682e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -21.90151 -14.41849
## sample estimates:
## mean in group 0.5   mean in group 2
##              7.98             26.14
```

```
t.test(len~dose, data=vcData23)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -5.4698, df = 13.6, p-value = 9.156e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.054267  -5.685733
## sample estimates:
## mean in group 1 mean in group 2
##           16.77           26.14
```

All the t-tests have p-values < 0.025. For 95% two-sided significance level we reject the null hypotheses and find a significant difference in tooth growth for different doses of supplement VC.

Perform t-tests on OJ data.

```
t.test(len~dose, data=ojData12)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -5.0486, df = 17.698, p-value = 8.785e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.415634  -5.524366
## sample estimates:
## mean in group 0.5   mean in group 1
##             13.23             22.70
```

```
t.test(len~dose, data=ojData13)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -7.817, df = 14.668, p-value = 1.324e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -16.335241  -9.324759
```

```
## sample estimates:
## mean in group 0.5   mean in group 2
##             13.23             26.06
```

```r
t.test(len~dose, data=ojData23)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -2.2478, df = 15.842, p-value = 0.0392
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.5314425 -0.1885575
## sample estimates:
## mean in group 1 mean in group 2
##           22.70           26.06
```

The t-tests comparing 0.5 mg/day to 1.0 mg/day and 0.5 mg/day to 2.0 mg/day have p-values < 0.025. For 95% two-sided significance level we reject the null hypotheses and find a significant difference in tooth growth for these pairs of doses of supplement OJ.

However, the comparison of 1.0 mg/day to 2.0 mg/day has a p-value > 0.025. For 95% two-sided significance level we accept the null hypotheses and do not find a significant difference in tooth growth for these doses of supplement OJ.

Question 4

For 95% two-sided significance level we failed to reject the null hypothesis that tooth growth between the two supplements in the same. As a result, no significant difference was found between the two supplements and tooth growth. This indicates the two supplements did not have statistically significant difference in tooth growth.

Regarding the VC supplement, for 95% two-sided significance level we rejected the null hypotheses for each pairing of supplement dose. This indicates the dosing levels of VC had statistically significant differences in tooth growth.

Regarding the OJ supplement, for 95% two-sided significance level we rejected the null hypotheses for each pairing of supplement dose except for 1.0 mg/day and 2.0 mg/day. This indicates the dosing levels of OJ had statistically significant differences in tooth growth between 0.5 mg/day and 1.0 mg/day and 0.5 mg/day and 2.0 mg/day but not 1.0 mg/day and 2 mg/day.