Statistical Inference Final Project Part One

Dax Gerts

Saturday, June 20, 2015

1. *"Load the ToothGrowth data and perform some basic exploratory data analyses"*

Retrieves ToothGrowth data from "datasets" package.

```
library(datasets)
data(ToothGrowth)
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
```

```
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 ...
```

Overview of Data:

-3 Variables: Independent, supp, dose; Dependent len

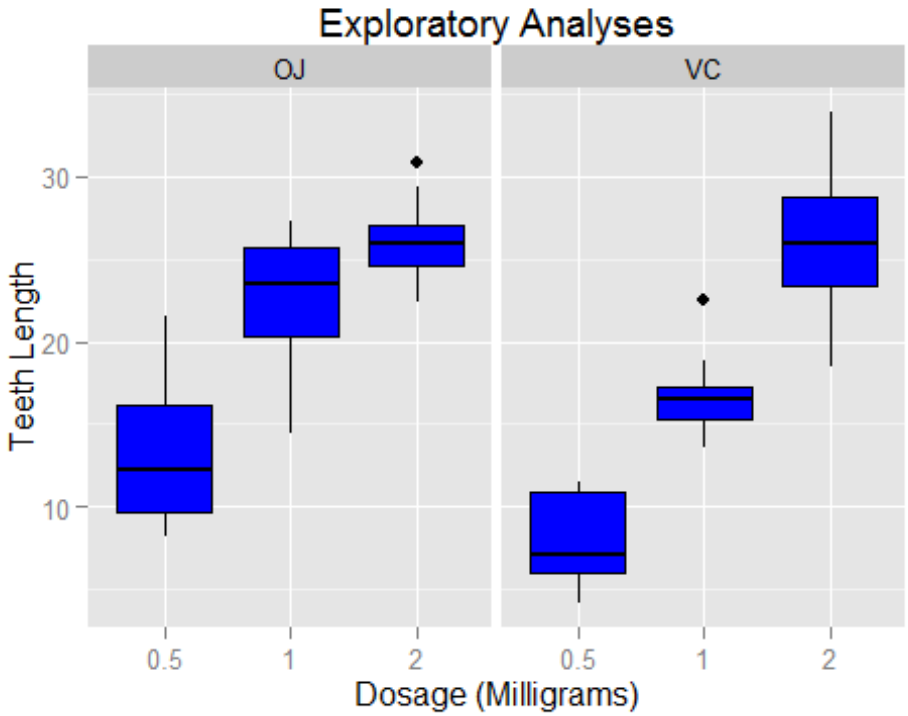
-len: 60 observerations

-supp: 2 values, "OJ" and "VC"

-dose: 3 values, 0,5, 1, and 2

Some basic exploratory data analyses.

```
library(ggplot2)
plot <- ggplot(ToothGrowth,
               aes(x=factor(dose),y=len,fill=factor(dose)))
plot + geom_boxplot(notch=F,color='black',fill='blue') + facet_grid(.~supp) +
      scale_y_continuous("Teeth Length") +
      scale_x_discrete("Dosage (Milligrams)") +
      ggtitle("Exploratory Analyses")
```



2. *"Provide a basic summary of the data"*

```
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 summary indicates that there is an even division of combinations of supp and dose obserations, totalling 10 of each (OJ-0.5,0J-1,OJ-2,VC-0.5,VC-1,VC-2)

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

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

Additionally, the mean tooth length is 18.81 and the median is 19.25. This indicates some level of skew to the right (higher values) in the data, but further analysis is necessary to confirm or deny the normality of the distribution.

3. *"Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose."*

The confidence intervals for each individual variable are shown numerically below.

Initially, three subsets are formed from each dosage group (1,2,and0.5 mg) before t-tests are conducted.

```
D0.5 <- subset(ToothGrowth,dose==0.5)
D1 <- subset(ToothGrowth,dose==1)
D2 <- subset(ToothGrowth,dose==2)

t.D0.5 <- t.test(len~supp,data=D0.5)
t.D1 <- t.test(len~supp,data=D1)
t.D2 <- t.test(len~supp,data=D2)

t.synopsis <- data.frame("Low-
Conf"=c(t.D0.5$conf[1],t.D1$conf[1],t.D2$conf[1]),"High-
Conf"=c(t.D0.5$conf[2],t.D1$conf[2],t.D2$conf[2]),"p-
value"=c(t.D0.5$p.value,t.D1$p.value,t.D2$p.value),row.names=c("dose=0.5mg","
dose=1mg","dose=2mg"))
t.synopsis
```

```
##             Low.Conf High.Conf     p.value
## dose=0.5mg  1.719057  8.780943 0.006358607
## dose=1mg    2.802148  9.057852 0.001038376
## dose=2mg   -3.798070  3.638070 0.963851589
```

This is followed by subsetting by Orange Juice or Vitamin C

4. *"State your conclusions and the assumptions needed for your conclusions."*

As indicated by both the graphical and numerical analyses above, the 2.0 milligram dose of Vitamin C showed the most statistically significant results with regards to increased tooth growth. Additionally, Vitamin C showed consistently stronger results than Orange Juice in each batch of tests, which is consistent with previouis observations. Lastly, while a 2.0 milligram dose of Vitamin C was significantly more effective than the 1.0 milligram or 0.5 milligram dose, 1.0 milligrams was also more effective than 0.5, and showed the most significant jump in effectiveness over Orange Juice.

In synopsis:

1. Vitamin C was significantly more effective than Orange Juice

2. Larger dosages (2.0mg) were more effective than smaller dosages (1.0mg & 0.5 mg)