Part 2

Phil Mai

December 4, 2016

Loading Libraries

```
library(ggplot2)
```

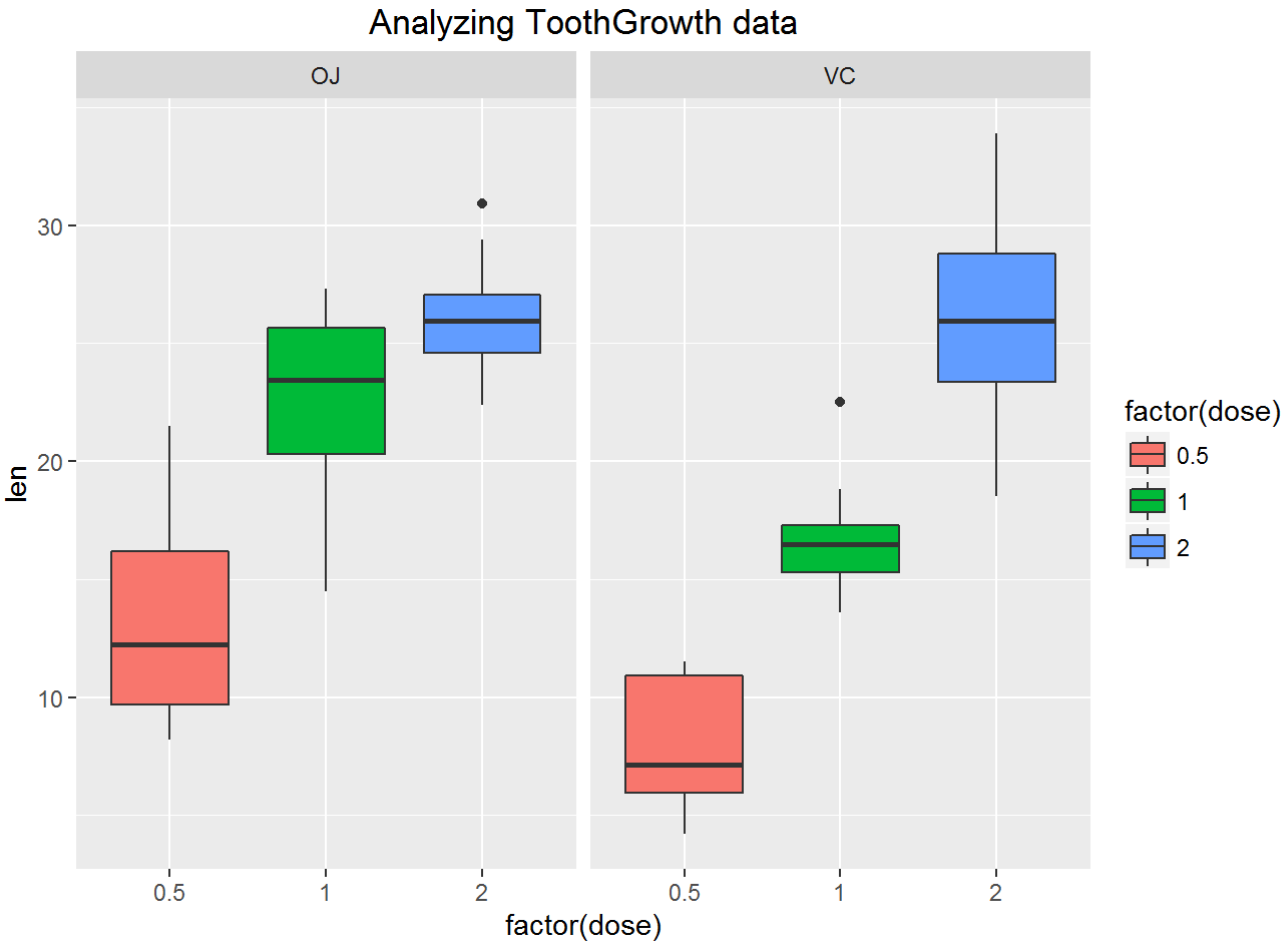
Statistical Inference Part 2

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.

Step1 - Loading the ToothGrowth Data

```
data(ToothGrowth)
ggplot(ToothGrowth, aes(x=factor(dose), y=len, fill=factor(dose)))+geom_boxplot()+facet_grid(.~s
upp)+ggtitle("Analyzing ToothGrowth data")
```



Step2 - Provide a Basic Summary

```
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 code shows a summary of the data set. It shows 2 supplements OJ and VC and 3 doses 0.5, 1, and 2 for in the dataset.

Step 3 - Confidence Internvals and/or Hypothesis Test

```
xBar<-mean(ToothGrowth$len[1:30])
yBar<-mean(ToothGrowth$len[31:60])
xVar<-(sd(ToothGrowth$len[1:30]))^2
yVar<-(sd(ToothGrowth$len[31:60]))^2
q<-(((xVar+yVar)/30)^2)/((((xVar/30)^2)+((yVar/30)^2))/29)
t<-qt(0.975, q)
yBar -xBar + c(-1,1)*t*sqrt(xVar/30 + yVar/30)
```

```
## [1] -0.1710156  7.5710156
```

Step #4 - Stating Conclusion

Through the boxplot, it can be concluded that as tooth size increases, the doses tend to be higher. The confidence interval is (-0.171, 7.571).