Statistiacl Inference Project part2

Fatemeh Abyarjoo

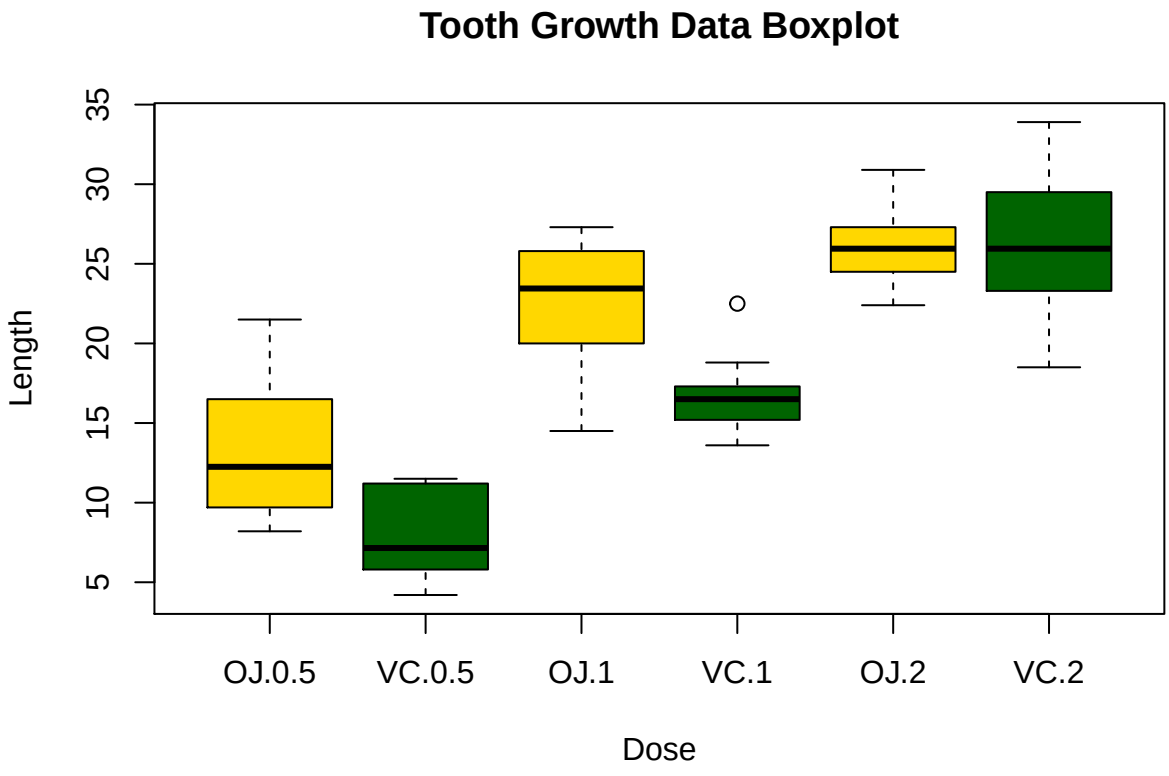
Saturday, April 25, 2015

1- Load the ToothGrowth data and perform some basic exploratory data analyses

```
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
```

```
boxplot(len ~ supp * dose, data = ToothGrowth, xlab = "Dose", ylab = "Length", col=(c("gold","darkgreen
```



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
```

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

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

```
mean(ToothGrowth$len)
```

```
## [1] 18.81333
```

```
sd(ToothGrowth$len)
```

```
## [1] 7.649315
```

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)

```
H1 <- t.test(len~supp, paired=F, var.equal=T, data=ToothGrowth)
H1
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

```
H2 <- t.test(len~supp, paired=F, var.equal=F, data=ToothGrowth)
H2
```

```
##
##  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
```

```
Ttest_results <- data.frame("Confidence Interval Low"=c(H1$conf[1],H2$conf[1]),"Confidence Interval High"=c(H1$conf[2],H2$conf[2]),
  "P Value"=c(H1$p.value,H2$p.value),row.names=c("Equal","Unequal"))
Ttest_results
```

```
##         Confidence.Interval.Low Confidence.Interval.High    P.Value
## Equal                -0.1670064                 7.567006 0.06039337
## Unequal              -0.1710156                 7.571016 0.06063451
```

4- State your conclusions and the assumptions needed for your conclusions.

Based on T test results, it can be concluded with the larger doses of either orange juce or ascorbic acid, grooth of the tooth is statistically different. The p value for lower doses (0.5 and 1) is 0.004239 which is bigger than 0.05 so orange juice has more effect on tooth growth than ascorbic acid. But at the higher dose (2), the p-value = 0.9639 which is bigger than 0.05 so the rate of tooth growth is not statistically different between orange juice and ascorbic acid.

```
OJset = subset(ToothGrowth, supp == "OJ");VCset= subset(ToothGrowth, supp == "VC")
Highdose  <- t.test(subset(OJset, dose == 2.0)$len, subset(VCset, dose == 2.0)$len)
Highdose
```

```
##
##  Welch Two Sample t-test
##
## data:  subset(OJset, dose == 2)$len and subset(VCset, dose == 2)$len
## t = -0.0461, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

```
Lowdose  <- t.test(subset(OJset, dose < 2.0)$len, subset(VCset, dose < 2.0)$len)
Lowdose
```

```
##
##  Welch Two Sample t-test
##
## data:  subset(OJset, dose < 2)$len and subset(VCset, dose < 2)$len
## t = 3.0503, df = 36.553, p-value = 0.004239
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.875234 9.304766
## sample estimates:
## mean of x mean of y
##    17.965    12.375
```