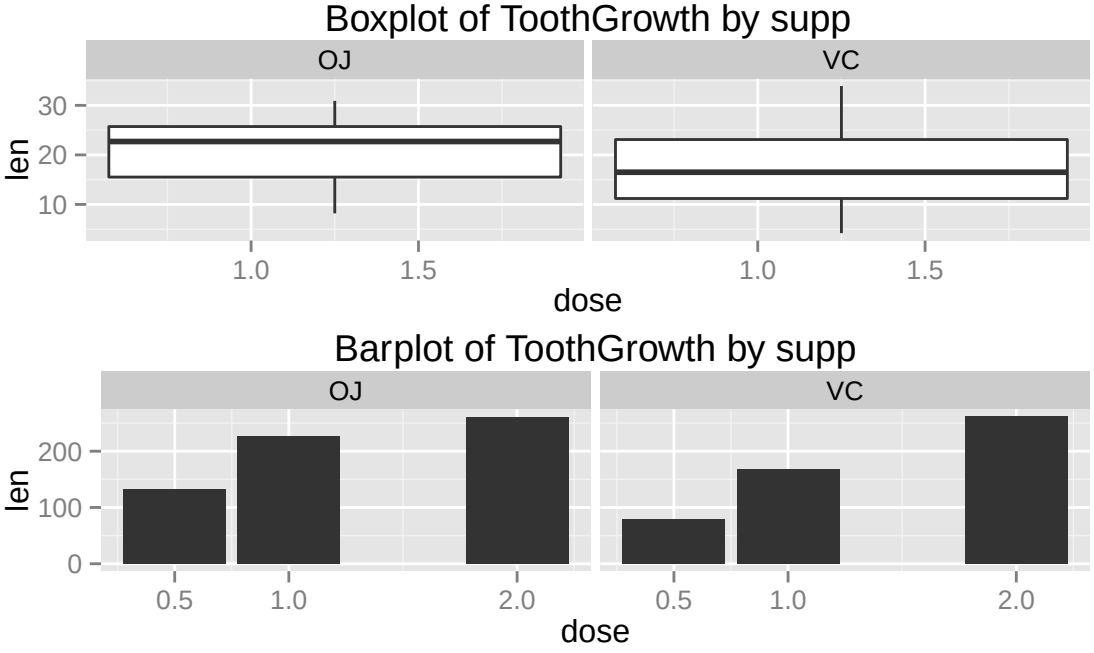
Statistical Inference - Peer Assessment 2

Cleosson Souza

27 de setembro de 2015

Analyzing the ToothGrowth data in the R datasets package

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




```
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 ...
```

There are 60 observations and 3 variables: a numeric “len”, a factor “supp”, a numeric “dose”.

The “len” variable represents Guinea Pig tooth length. The “supp” variable represents one of two delivery methods (OJ = orange juice, VC = ascorbic acid). And “dose” variable represents one of three dose levels of vitamin C in milligrams (0.5, 1, 2)

2. Provide a basic summary of the data

This summary let us know that Toothgrowth length in OJ group is longer than VC group execept maximum value.

```
oj.summary <- summary(ToothGrowth[ToothGrowth$supp=="OJ",1])
vc.summary <- (ToothGrowth[ToothGrowth$supp=="VC",1])
t.summaries <- as.data.frame(rbind(names(oj.summary), "OJ"=as.vector(oj.summary), "VC"=as.vector(vc.sum
t.summaries
```

```
##      V1      V2     V3    V4      V5   V6
##    Min. 1st Qu. Median  Mean 3rd Qu. Max.
## OJ  8.2   15.52   22.7 20.66   25.72 30.9
## VC  4.2    11.5    7.3   5.8     6.4   10
```

And This table gives us informations that means of OJ in each dose are also larger than VC. But Standard deviation of OJ group is larger too in 0.5 and 1 doses.

```
colnames(t.summaries) <- c("Min.", "1st Qu.", "Median",  "Mean", "3rd Qu.", "Max.")
t.summaries <- t.summaries[-1,]
t.mean <- aggregate(len~supp+dose, data=ToothGrowth, FUN=mean)
t.sd <- aggregate(len~supp+dose, data=ToothGrowth, FUN=sd)
t.summaries2 <- cbind(t.mean, t.sd[,3]); colnames(t.summaries2)<-c("supp", "dose", "Mean", "sd")
t.summaries2
```

```
##   supp dose  Mean       sd
## 1   OJ  0.5 13.23 4.459709
## 2   VC  0.5  7.98 2.746634
## 3   OJ  1.0 22.70 3.910953
## 4   VC  1.0 16.77 2.515309
## 5   OJ  2.0 26.06 2.655058
## 6   VC  2.0 26.14 4.797731
```

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

For this analysis, we will compare the two supplements across each dosage. This is a two sided test since we are looking at tooth growth provided that the supplement is orange juice vs ascorbic acid.

Assumption

We are going to perform these tests using the parameter paired=FALSE because the guinea pigs that received the orange juice are separate from the guinea pigs that received the ascorbic acid. We will also assume unequal variances. The hypothesized value in this case is 0 and confidence is 95%

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
g1 <- subset(ToothGrowth, dose==0.5)
g2 <- subset(ToothGrowth, dose==1)
g3 <- subset(ToothGrowth, dose==2)

t.test(len ~ supp, paired=FALSE, var.equal=FALSE, data=g1)
```

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

```
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
t.test(len ~ supp, paired=FALSE, var.equal=FALSE, data=g2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```
t.test(len ~ supp, paired=FALSE, var.equal=FALSE, data=g3)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, 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 in group OJ mean in group VC
##            26.06            26.14
```

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

From the confidence interval testing we can see the following:

> At a dosage of 0.5, the confidence interval is entirely above 0. We can reject the null hypotesys and it is reasonable assume there is difference in tooth growth between orange juice than on ascorbic acid
>
> At a dosage of 1.0, the confidence interval is also entirely above 0. We can reject the null hypotesys and it is reasonable assume there is difference in tooth growth between orange juice than on ascorbic acid
>
> At a dosage of 2.0, the confidence interval includes 0. We cannot reject the null hypotesys and it is reasonable assume there is no difference in tooth growth between orange juice than on ascorbic acid