Statistical Inference Course Project – Data Analysis

Jerry Dumblauskas

April 10, 2016

Load the ToothGrowth data (part of the datasets in R) and perform some basic exploratory data analyses and come up with some results.

```
# source the data
data(ToothGrowth)
```

Provide a basic summary of the data.

```
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 ...
```

```
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
```

```
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
```

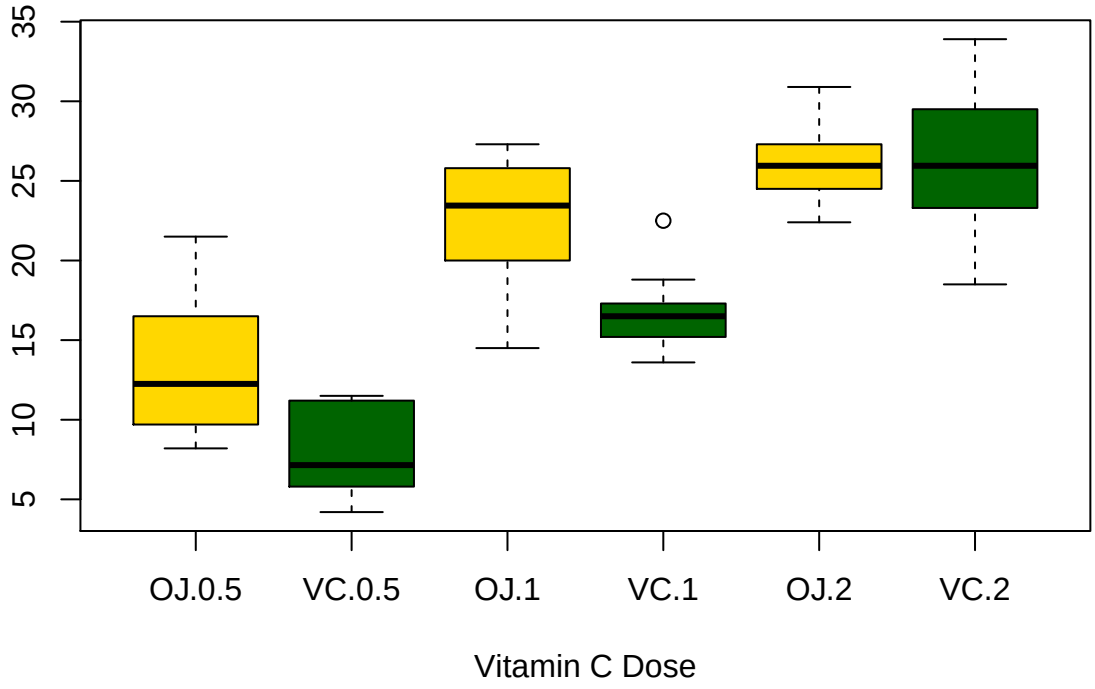
```
unique(ToothGrowth$supp)
```

```
## [1] VC OJ
## Levels: OJ VC
```

```
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

```
boxplot(len~supp*dose, data=ToothGrowth, xlab="Vitamin C Dose", col=(c("gold","darkgreen")))
```



Perform some analysis

looking at the data (per dose) it is close, so let's say "different supp types have no effect on length" (H0)

```
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
```

Here, the p is 6 percent (kind of low >23 percent chance of messing up), but the confidence interval contains 0, so we cannot make a clear acception or rejection of H0

Dose has an impact on tooth growth (.5 v 2)

```
d_05 <- ToothGrowth[which(ToothGrowth$dose==.5),1]
d_20 <- ToothGrowth[which(ToothGrowth$dose==2),1]
t.test(d_05, d_20, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  d_05 and d_20
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean of x mean of y
##    10.605    26.100
```

low p, test contains 0 – so this H0 is accepted!

Conclusions

the higher the dose, the more growth. The basic test showed no statistically different data between the supp types. In future analysis, if the dose type is restricted (like getting rid of low dose), it may be possible to infer differences in the supplement types.