Statistical Inference Course Project Part 2

Scott Lin

January 21, 2015

Description

This project is to analyze the ToothGrowth data set and run some hypothesis testing on the supplement and dosage.

Q1: Load the ToothGrowth Data Set

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

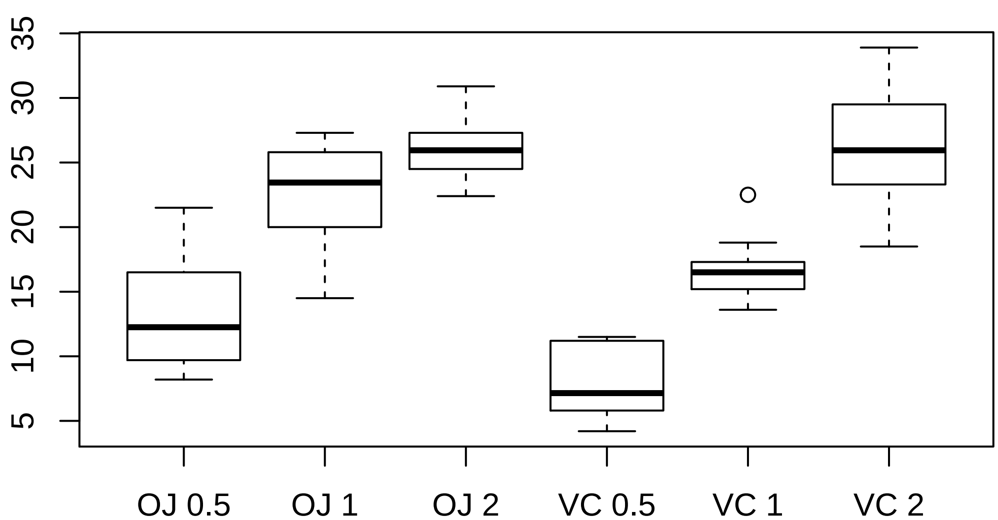
Q2: Provide a basic summary of the data

From the following R Code:

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

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

I have figured out that the data frame includes 3 columns with 3 doses for 60 observations. That's 10 observation per combined supp/dose pair. It's not a paired test as there are no columns that describe the subject. I've combined the doses with the supp and made a box/whisker plot from it.



First level observations is that both drugs get more "potent" as the doses increase. VC seems to be weaker in lower doses but has a wider range for a dose of 2.0. My geusses would be that OJ outperforms VC at lower doses, but higher doses will probably be inconclusive as the box/whisker plots overlap completely.

Q3: 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).

State your assumptions:

I'm assuming that:

- I'm going to use a T-test due to small sample size of 10
- This is an unpaired test as there are no values indicating whot he subject of the tests are
- Given the small sample size, I'm assuming it's a subset of the data and we don't know the true variance of the population, so i'm going to assume all variances are unequal.
- Longer tooth is better
- The object, is to find the drug with the dosage that grows the longest teeth.
- It's clear from the box/whisker plot, the teeth grows longer, the higher the dosage so I'll concentrate at comparing dosage for the drugs at every level

There are 3 dosage options so there will be 3 tests and I'm going to test each and one of them individually.

Did VC 0.5 grow longer teeth than OJ 0.5? (H0 is VC 0.5):

```r
VC05 <- ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 0.5, "len"]
OJ05 <- ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 0.5, "len"]
t.test(VC05,OJ05,paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  VC05 and OJ05
## 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:
##  -8.780943 -1.719057
## sample estimates:
## mean of x mean of y
##      7.98     13.23
```

The p-value is at a significance level (0.006). The t value is -3.16 and the confidence level is below 0 completely. We can reject the null hypothesis and say OJ at dose 0.5 is better.

Did VC 1.0 grow longer teeth than OJ 1.0? (H0 is VC 1.0):

```r
VC10 <- ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 1.0, "len"]
OJ10 <- ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 1.0, "len"]
t.test(VC10,OJ10,paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  VC10 and OJ10
## 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:
##  -9.057852 -2.802148
## sample estimates:
## mean of x mean of y
##     16.77     22.70
```

The p-value is at a significance level (0.001). The t value is -4.03 and the confidence level is below 0 completely. We can reject the null hypothesis and say OJ at dose 1.0 is better.

Did VC 2.0 grow longer teeth than OJ 2.0? (H0 is VC 2.0):

```r
VC20 <- ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 2.0, "len"]
OJ20 <- ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 2.0, "len"]
t.test(VC20,OJ20,paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  VC20 and OJ20
## 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:
##  -3.63807  3.79807
## sample estimates:
## mean of x mean of y
##     26.14     26.06
```

The p-value is at a insignificant level (0.96). The t value is 0.04 and the confidence level is surrounds 0 at -3.6 to 3.8. It's inconclusive at the 2.0 dosage level, and we can not reject the null hypothesis.

Q4: State your conclusions. (Assumptions stated above)

I conclude that for both VC and OJ, higher doses leads to longer teeth. For both 0.5 and 1.0, OJ is clearly better at delivering the vitamin C required for longer teeth. At the 2.0 level, it's inconclusive which one is better, but I would choose OJ due to its similar mean, and smaller standard deviation.

Appendix

Let's see if it's true that increasing dosage is better:

```
t.test(VC05, VC10, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  VC05 and VC10
## t = -7.4634, df = 17.862, p-value = 6.811e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.265712  -6.314288
## sample estimates:
## mean of x mean of y
##      7.98     16.77
```

VC 1.0 is better than VC 0.5.

```
t.test(VC10, VC20, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  VC10 and VC20
## t = -5.4698, df = 13.6, p-value = 9.156e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.054267  -5.685733
## sample estimates:
## mean of x mean of y
##     16.77     26.14
```

VC 2.0 is better than VC 1.0

```
t.test(OJ05, OJ10, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ05 and OJ10
## t = -5.0486, df = 17.698, p-value = 8.785e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.415634  -5.524366
## sample estimates:
## mean of x mean of y
##     13.23     22.70
```

OJ 1.0 is better than OJ 0.5

```r
t.test(OJ10, OJ20, paired=FALSE, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ10 and OJ20
## t = -2.2478, df = 15.842, p-value = 0.0392
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.5314425 -0.1885575
## sample estimates:
## mean of x mean of y
##     22.70     26.06
```

OJ 2.0 is better than OJ 1.0, not by much though. The p-value is 0.04.