Statistical Inference Project Part 2

Ryan Wissman

Sunday, January 25, 2015

Coursera Statistical Inference Course Project

This is part two of the project for the class on statistical inference. For this project we will be analyzing the ToothGrowth dataset found in R. We will address the following questions using this data:

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.

Loading data and initial analysis (Q1)

Load the data.

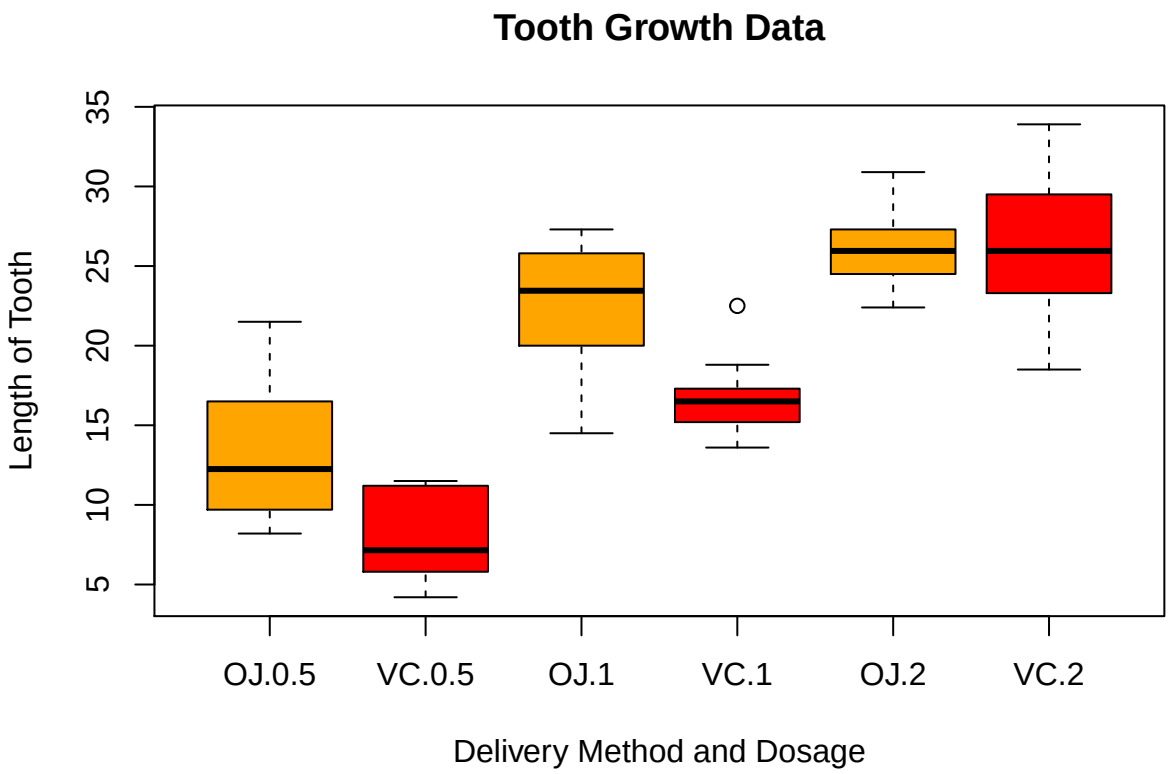
```
library(datasets)
data(ToothGrowth)

#Get the first few rows of the data
head(ToothGrowth,2)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
```

The data shows the effect of vitamin c on tooth growth in guinea pigs. We know that the length of the tooth is given by (len), as well as the delivery method (OJ) or (VC), and the three dose levels (0.5mg, 1mg, and 2mg). A great way to quickly visualize this data would be to create a box and whisker plot illustrating the tooth growth by dosage level and delivery method.

```
boxplot(len ~ supp * dose, data=ToothGrowth, col=(c("orange", "red")),
        ylab="Length of Tooth", xlab="Delivery Method and Dosage",
        main="Tooth Growth Data")
```



After examining the above chart we could hypothesize that guinea pigs experienced more tooth growth when given orange juice when compared to ascorbic acid. Furthermore, tooth growth was further increased when exposed to higher dose levels.

Data Summary (Q2)

We will now perform a very basic summary of the data.

```
#set dose to factor to get count
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
summary(ToothGrowth)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

We will also look at the mean for each dosage level and delivery method

```
with(ToothGrowth, tapply(len, list(supp,dose), mean))
```

```
##      0.5     1     2
## OJ 13.23 22.70 26.06
## VC  7.98 16.77 26.14
```

Look at the data this way furthers our hypothesis that orange juice is the superior delivery method to promote tooth growth at 0.5mg and 1mg. There appears to be little difference at 2mg.

Analysis and Hypothesis Test (Q3)

We will now test our hypothesis that orange juice is the superior vitamin C delivery method and results in better tooth growth among guinea pigs than ascorbic acid. This is our Null hypothesis. The alternative hypothesis will be that neither delivery method promotes toothgrowth in a superior way. We will perform a t-test on these data as the size of the data is quite small (60 total observations, 10 guinea pigs per supplement).

We will reject accept the null hypothesis if the t-test shows that the true difference for each supplement is larger or less than zero.

```
#t-test for 0.5mg
OJ_min <- ToothGrowth[ToothGrowth$supp=="OJ" & ToothGrowth$dose=="0.5", 1]
VC_min <- ToothGrowth[ToothGrowth$supp=="VC" & ToothGrowth$dose=="0.5", 1]
t.test(OJ_min, VC_min)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ_min and VC_min
## 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 of x mean of y
##     13.23      7.98
```

After this test we accept the null hypothesis as the difference is not equal to zero, it appears that OJ promotes larger tooth growth at 0.5mg. p-value is $< .01$. We could infer this with the initial chart.

```
#t-test for 1mg
OJ_min <- ToothGrowth[ToothGrowth$supp=="OJ" & ToothGrowth$dose=="1", 1]
VC_min <- ToothGrowth[ToothGrowth$supp=="VC" & ToothGrowth$dose=="1", 1]
t.test(OJ_min, VC_min)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ_min and VC_min
## 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 of x mean of y
##     22.70     16.77
```

After this test we accept the null hypothesis as the difference is not equal to zero, it appears that OJ promotes larger tooth growth at 1mg. p-value is $< .01$. We could infer this with the initial chart.

```
#t-test for 2mg
OJ_min <- ToothGrowth[ToothGrowth$supp=="OJ" & ToothGrowth$dose=="2", 1]
VC_min <- ToothGrowth[ToothGrowth$supp=="VC" & ToothGrowth$dose=="2", 1]
t.test(OJ_min, VC_min)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ_min and VC_min
## 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
```

After this test we reject the null hypothesis as the difference could be zero at 2mg dosage level. p-value is $> .05$. We could infer this with the initial chart.

Conclusion (Q4)

For dosage levels less than 2mg (0.5mg and 1mg) we accepted the null hypothesis and determined that there is a true statistical difference between vitamin c delivery methods in promoting tooth growth among guinea pigs. At the dosage level of 2mg there was no statistical difference between orange juice and ascorbic acid as a delivery method for vitamin c with regards to promoting tooth growth. Thus, at 2mg we can reject the null hypothesis.