Tooth Growth - Statistical Inference Part 2

Jackie Milhans

July 24, 2015

We're going to analyze the ToothGrowth data in the R datasets package. Load the ToothGrowth data and perform some basic exploratory data analyses.

Data Summary

The ToothGrowth dataset is the response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

Let's provide a basic summary of the data. First we load the dataset.

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

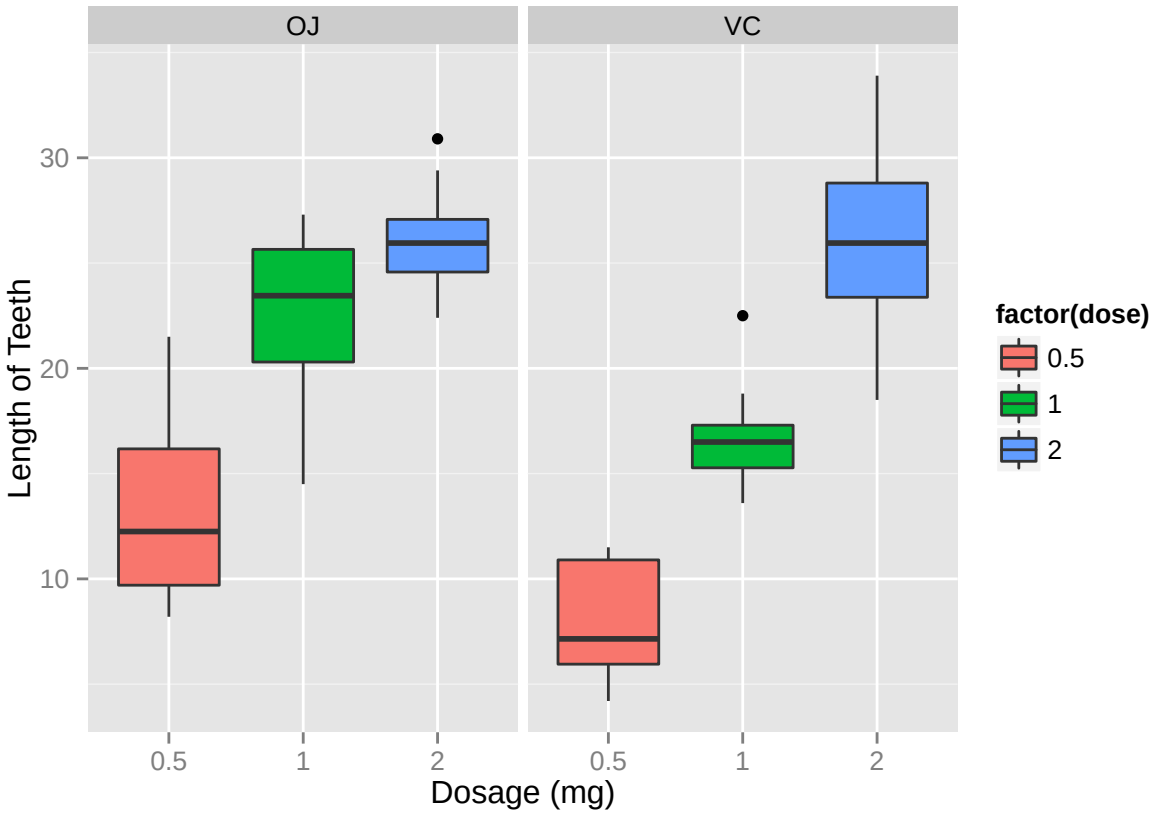
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
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
```

The summary of the dataset shows that there are two supplements, OJ and VC. There are 3 doses: 0.5, 1, and 2.

```
library(ggplot2)
ggplot(ToothGrowth, aes(x=factor(dose), y=len, fill=factor(dose))) +
        geom_boxplot() +
        facet_grid(.~supp) +
        labs(main="Analyzing ToothGrowth Dataset", y="Length of Teeth", x="Dosage (mg)")
```



The plots show that a high dosage leads to longer teeth.

Confidence Intervals

We use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

Here, we use the hypothesis test, t-test.

We can also look at the data by splitting into the different dosages.

```
dose1 <- subset(ToothGrowth, dose==0.5)
dose2 <- subset(ToothGrowth, dose==1.0)
dose3 <- subset(ToothGrowth, dose==2.0)
```

First, we look at the smallest dose.

```
t.test(len~supp, data=dose1, paired=FALSE)
```

```
##
##  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
```

Finally, the largest dose is examined.

```
t.test(len~supp, data=dose3, paired=FALSE)
```

```
##
##  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
```

Conclusions

For the lower dosage data, we can see that the OJ produces higher tooth growth. At the highest dosage, there is little difference between the two supplements.

For a higher dose, the p-value is greater than 0.05, and the confidence interval lies between (-3.79807, 3.63807). The CI includes 0, so it shows that there is no significant difference between the two supplements.

For the smallest dosage, the p-value is lower that 0.05, and the confidence interval does not include 0. Therefore, we can see that there is a significant difference in growth with the smaller dose.

APPENDIX

Here, we use the hypothesis test by supplement for all data, t-test.

```
t.test(len~supp, data=ToothGrowth, paired=FALSE)
```

```
##
##  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
```

Dosage of 1.0 mg.

```
t.test(len~supp, data=dose2, paired=FALSE)
```

```
##
##  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
```