Statistical Inference Project Part 2

Jacob Townson

July 20, 2015

Overview

In this report I will be analyzing the ToothGrowth data set in R by doing different statistical methods learned in the John Hopkins Statistical Inference course on Coursera. This analysis will begin with a simple look at the data, then go more in depth by finding confidence intervals and doing hypothesis tests to find more detailed information. Finally, I will state the conclusions of my findings. Note I do not have an appendix section because I merged the code with my explanations and still did not exceed the 3 page limit.

Introduction to the Data

First I must load the data using this R code:

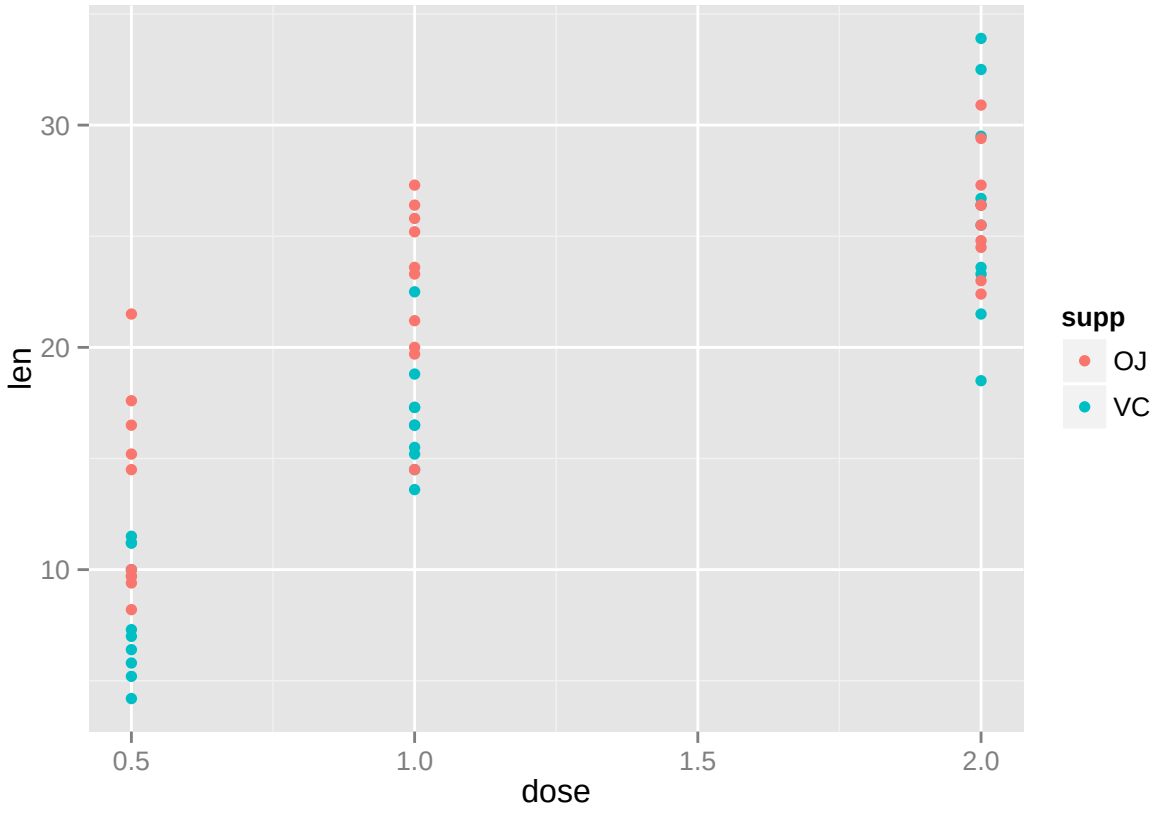
```
data("ToothGrowth")
```

The description of the data is given as such:

"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)"

So we are looking at dose levels, delivery methods, and length of teeth for the guinea pigs. To begin my analysis, let's look at a simple plot of the dosage vs. length for each of the two methods.

```
qplot(x = dose, y = len, colour = supp, data = ToothGrowth)
```



Based simply off of this graph, one could assume that the supplement method of using orange juice is better for smaller doses, whereas both methods of delivery seem to be almost equal for a dosage of 2. One could assume this because our points for orange juice are higher for the smaller doses.

Analysis

So let's see if generally the orange juice method is higher than the supplement method by finding a confidence interval. To do this, I will be using t tests to compare the two methods at each dosage. Luckily, the data is divided up in it's data frame by method of delivery and dosage. Note, I am not using a paired t test because the two methods of delivery are independent of each other (different guinea pigs for each dosage and each supplement method).

```
g1 <- ToothGrowth$len[1:10]; g2 <- ToothGrowth$len[31:40]
t.test(g1, g2)
```

```
##
##  Welch Two Sample t-test
##
## data:  g1 and g2
## 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
```

This test covers the dosage of 0.5 mg. The test states that the confidence interval is between -8.780943 and -1.719057. This confidence interval does not contain 0, so we can rule out 0 as the difference between the populations. The test also tells us that the mean of the 2nd estimate is higher, thus at a dosage level of 0.5, the orange juice method is greater than the supplement method.

```
g1 <- ToothGrowth$len[11:20]; g2 <- ToothGrowth$len[41:50]
t.test(g1, g2)
```

```
##
##  Welch Two Sample t-test
##
## data:  g1 and g2
## 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
```

This test covers the dosage of 1 mg. The test states that the confidence interval is between -9.057852 and -2.802148. This confidence interval does not contain 0, so we can rule out 0 as the difference between the populations. The test also tells us that the mean of the 2nd estimate is higher, thus at a dosage level of 1, the orange juice method is greater than the supplement method.

```
g1 <- ToothGrowth$len[21:30]; g2 <- ToothGrowth$len[51:60]
t.test(g1, g2)
```

```
##
##  Welch Two Sample t-test
##
## data:  g1 and g2
## 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:
##  -3.63807  3.79807
## sample estimates:
## mean of x mean of y
##     26.14     26.06
```

This test covers the dosage of 2 mg. The test states something different in this case. For this set of data with a dosage of 2 mg, the confidence interval does in fact contain 0. Thus, we cannot conclude that either method is better than the other because it is possible for their population difference to be 0. In fact, according to the test, even both means are extremely close.

In conclusion, my initial hypothesis was right based off of the original plot. The orange juice method is better for tooth growth than vitamin supplements for doses of 0.5 and 1, but the methods are equivalent for a dose of 2.