Analysis of Tooth Growth Data

Sharon Francisco

August 10, 2016

Synopsis

This report analyzes the ToothGrowth data in R's *datasets* library. Six groups of 10 guinea pigs underwent Vitamin C treatment protocols. The length of the guinea pigs' odontoblasts (cells responsible for tooth growth) were recorded. Each subject received one of three doses of Vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods (orange juice or ascorbic acid).

The null hypotheses being tested are that there is no difference in mean tooth length for the subjects receiving vitamin C by the different delivery methods regardless of dose; receiving different doses; and by delivery method for each dose.

Install Packages and Load Libraries

The packages used in the analysis are installed below.

```
if(!require(datasets, quietly = TRUE)) install.packages("datasets")
library(datasets)
if(!require(ggplot2, quietly = TRUE)) install.packages("ggplot2")
library(ggplot2)
```

Exploratory Analysis

The R Documentation for the ToothGrowth data states it is the results of a study of the effect of Vitamin C on Tooth Growth in Guinea Pigs.

len is the length of the odontoblasts (cells responsible for tooth growth) in 60 different guinea pigs. Each animal recieved one of three dose levels (**dose**) of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods (**supp**): orange juice (OJ) and ascorbic acid, a form of vitamin C (VC).

I'll get the first 10 rows of the ToothGrowth data, get its structure, and get a summary.

```
head(ToothGrowth, 10)
```

```
##     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
## 7  11.2   VC  0.5
## 8  11.2   VC  0.5
## 9   5.2   VC  0.5
## 10  7.0   VC  0.5
```

```r
str(ToothGrowth)
```

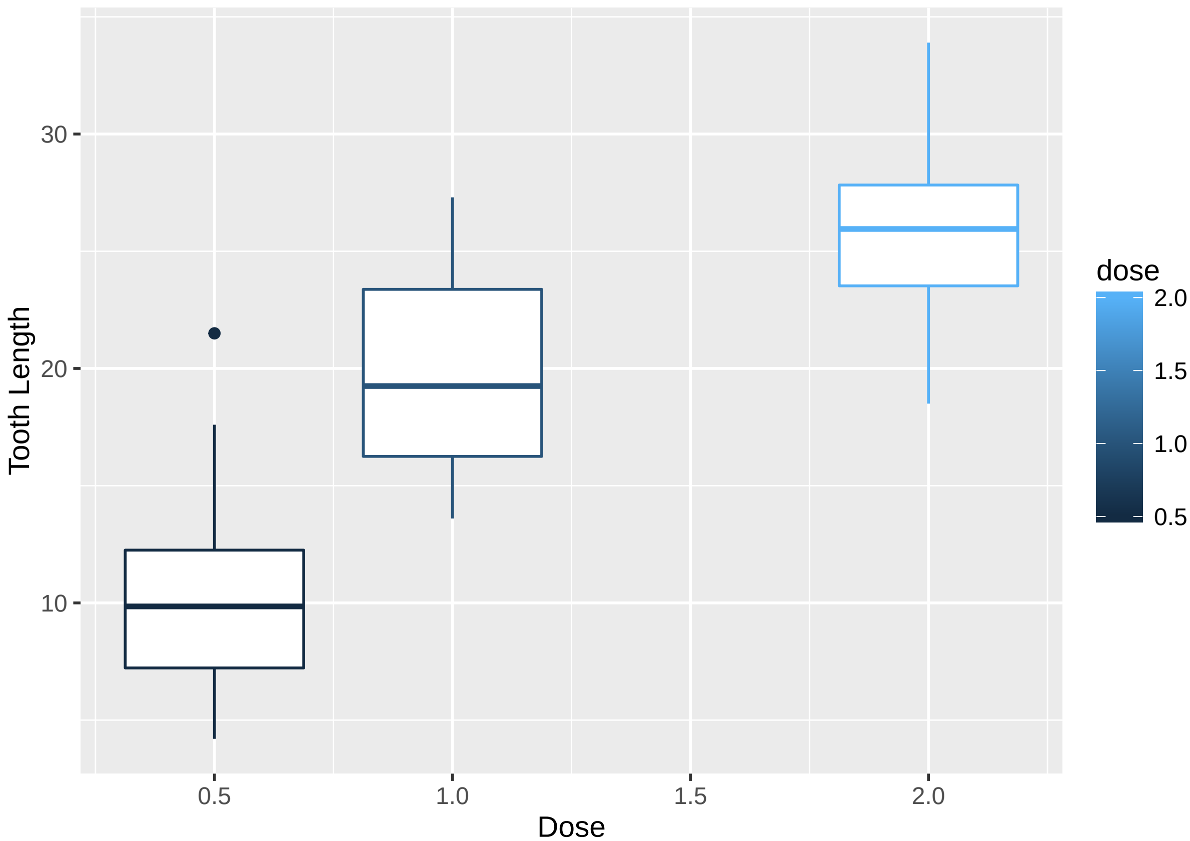
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```r
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
```

Now I'll make some graphs. I'll look at **dose** and **len**.

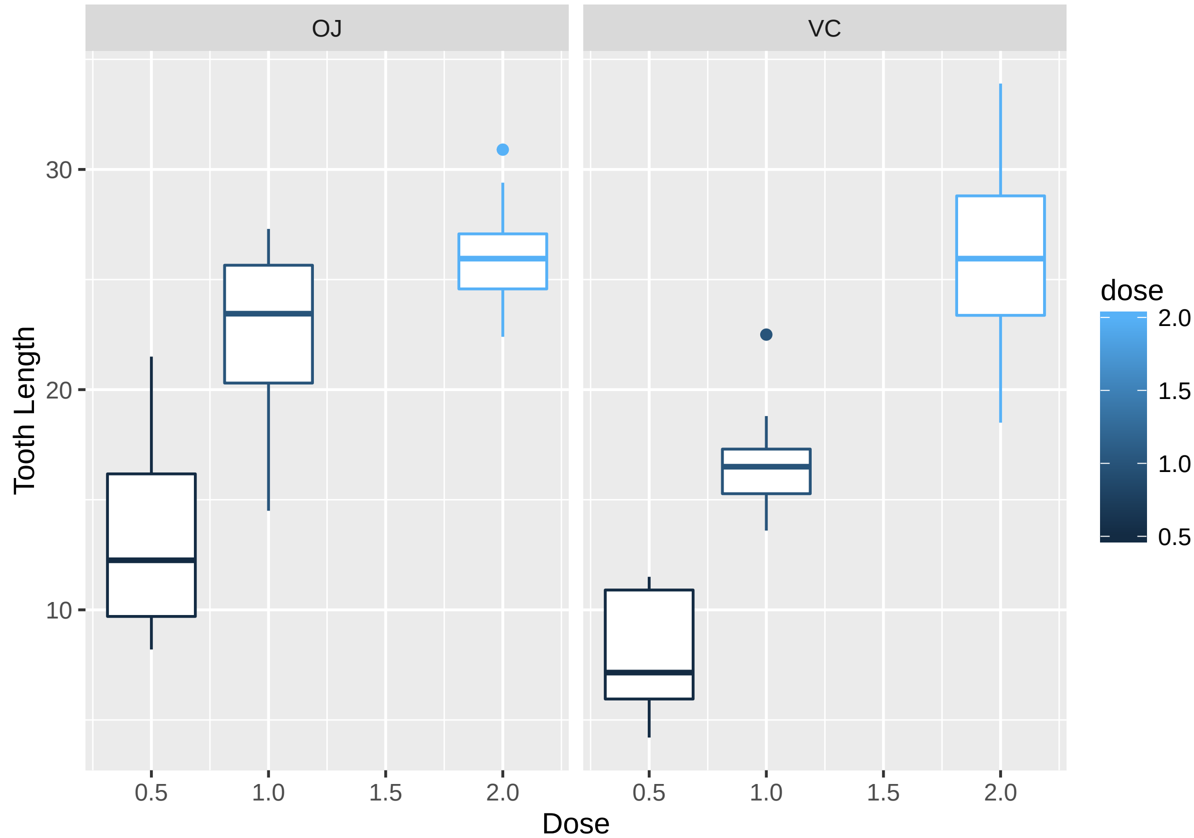
```r
grph1 <- ggplot(ToothGrowth, aes(dose, len))
grph1 <- grph1 + geom_boxplot(aes(group = dose, color = dose)) + labs(x = "Dose", y = "Tooth Length")
grph1
```



It looks like the higher the dose of vitamin C, the longer the length.

I'll see what effect **supp** has.

```
grph1 <- grph1 + facet_grid(. ~ supp)
grph1
```



For the two lowest doses, 0.5 and 1.0 mg/day, delivery by *OJ* seems to result in longer lengths. For doses of 2.0 mg/day, the differences are less pronounced.

Hypothesis Testing

Two sets of hypotheses will be tested to determine the statistical significance in the difference of mean tooth lengths for the two different delivery methods (orange juice and ascorbic acid) and each dosage level (0.5, 1, and 2 mg/day).

Confidence Intervals

All confidence intervals will be constructed at a 95% confidence level, and all hypotheses will be tested using alpha = 0.05.

Assumptions

T-tests will be used under the assumption that the guinea pigs were randomly assigned to the doses and delivery methods and are independent and identically distributed (iid). The confidence intervals will be constructed using unpaired intervals. The variances in tooth length are assumed to be unequal for the different groups.

Delivery Method

The first test will evaluate the difference in mean tooth length for the two delivery methods, regardless of dose.

The Null Hypothesis (H-Null) is that the mean tooth lengths are equal for the delivery methods, orange juice (mu-OJ) and ascorbic acid (mu-VC):

H-Null: mu-OJ = mu-VC

The Alternative Hypothesis (H-Alt) is that the mean tooth lengths are not equal for the delivery methods:

H-Alt: mu-OJ != mu-VC

Here is the t-test:

```
tstSupp <- t.test(len ~ supp, data = ToothGrowth,
                  paired = FALSE,var.equal = FALSE)
tstSupp
```

```
##
##  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
```

The p-value, 0.0606345, is greater than alpha, 0.05, and the confidence interval, -0.1710156, 7.5710156, contains 0, so we fail to reject the **Null Hypothesis: mu-OJ = mu-VC**.

Doses

The next series of tests will evaluate the differences in mean tooth length for each dose given, 0.5, 1, and 2 mg/day.

The Null Hypotheses will be that the mean tooth lengths are equal for all doses:

1. H-Null-1: mu-half = mu-one
2. H-Null-2: mu-one = mu-two

The Alternative Hypotheses will be that the mean tooth lengths are unequal for all doses:

1. H-Alt-1: mu-half != mu-one
2. H-Alt-2: mu-one != mu-two

First, I'll subset the data based on dose.

```
half <- subset(ToothGrowth, dose == 0.5)
one <- subset(ToothGrowth, dose == 1)
two <- subset(ToothGrowth, dose == 2)
```

Here is the first t-test:

```
tstDose1 <- t.test(one$len, half$len, paired = FALSE,
                   var.equal = FALSE)
tstDose1
```

```
##
##  Welch Two Sample t-test
##
## data:  one$len and half$len
## t = 6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.276219 11.983781
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

The p-value is less than alpha, 0.05, and the confidence interval, 6.2762187, 11.9837813, does not contain 0, so we can reject the Null Hypothesis, and accept the **Alternative Hypothesis (H-Alt-1): mu-half != mu-one**.

Here is the second t-test:

```
tstDose2 <- t.test(two$len, one$len, paired = FALSE,
                   var.equal = FALSE)
tstDose2
```

```
##
##  Welch Two Sample t-test
##
## data:  two$len and one$len
## t = 4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.733519 8.996481
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

The p-value is less than alpha, 0.05, and the confidence interval, 3.7335195, 8.9964805, does not contain 0, so we can reject the Null Hypothesis, and accept the **Alternative Hypothesis (H-Alt-1): mu-one != mu-two**.

Delivery Method by Dose

The last set of tests will compare the differences in mean tooth lengths for delivery method by dose.

The Null Hypotheses will be that the mean tooth lengths are equal for each dose for both delivery methods:

1. H-Null-1: mu-half-OJ = mu-half-VC
2. H-Null-2: mu-one-OJ = mu-one-VC
3. H-Null-3: mu-two-OJ = mu-two-VC

The Alternative Hypotheses will be that the mean tooth lengths are not equal for each dose for the delivery methods:

1. H-Alt-1: mu-half-OJ != mu-half-VC
2. H-Alt-2: mu-one-OJ != mu-one-VC
3. H-Alt-2: mu-two-OJ != mu-two-VC

I'll subset the data by dose and delivery method.

```
halfOJ <- subset(ToothGrowth, dose == 0.5 & supp == "OJ")
halfVC <- subset(ToothGrowth, dose == 0.5 & supp == "VC")
oneOJ <- subset(ToothGrowth, dose == 1 & supp == "OJ")
oneVC <- subset(ToothGrowth, dose == 1 & supp == "VC")
twoOJ <- subset(ToothGrowth, dose == 2 & supp == "OJ")
twoVC <- subset(ToothGrowth, dose == 2 & supp == "VC")
```

Here is the first t-test:

```
tstDS1 <- t.test(halfOJ$len, halfVC$len, paired = FALSE,
                 var.equal = FALSE)
tstDS1
```

```
##
##  Welch Two Sample t-test
##
## data:  halfOJ$len and halfVC$len
## 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
```

The p-value, 0.0063586, is less than alpha, 0.05, and the confidence interval, 1.7190573, 8.7809427, does not contain 0, so we can reject the Null Hypothesis, and accept the **Alternative Hypothesis (H-Alt-1): mu-half-OJ != mu-half-VC**.

Here is the second t-test:

```
tstDS2 <- t.test(oneOJ$len, oneVC$len, paired = FALSE,
                 var.equal = FALSE)
tstDS2
```

```
##
##  Welch Two Sample t-test
##
## data:  oneOJ$len and oneVC$len
## 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
```

The p-value, 0.0010384, is less than alpha, 0.05, and the confidence interval, 2.8021482, 9.0578518, does not contain 0, so we can reject the Null Hypothesis, and accept the **Alternative Hypothesis (H-Alt-2): mu-one-OJ != mu-one-VC**.

Here is the third t-test:

```
tstDS3 <- t.test(twoOJ$len, twoVC$len, paired = FALSE,
                 var.equal = FALSE)
tstDS3
```

```
##
##  Welch Two Sample t-test
##
## data:  twoOJ$len and twoVC$len
## 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 of x mean of y
##     26.06     26.14
```

The p-value, 0.9638516, is greater than alpha, 0.05, and the confidence interval, -3.7980705, 3.6380705, contains 0, so we cannot reject the **Null Hypothesis (H-Null-3): mu-two-OJ = mu-two-VC**.

Summary

Based on the assumptions listed above and the t-tests performed, the delivery method, orange juice or ascorbic acid, does not result in a differences in the mean tooth lengths, regardless of dose (i.e., mu-OJ = mu-VC for all doses). There is also no significant difference in the means of tooth lengths for the guinea pigs receiving 1 or 2 mg/day of Vitamin C by the different delivery methods (i.e., mu-two-OJ = mu-two-VC). For the animals receiving 0.5 or 1 mg/day, delivery by orange juice resulted in a difference in the mean tooth lengths (i.e., mu-one-OJ != mu-one-VC).

For dose-to-dose comparisons regardless of delivery method (e.g., mu-half = mu-one), the guinea pigs given 1 mg/day had different mean tooth lengths than the ones given 0.5 mg/day (i.e., mu-half != mu-one). The guinea pigs given 1 mg/day and 2 mg/day did not have different mean tooth lengths (i.e., mu-one = mu-two).