ToothGrowth Data: Basic Exploratory Analyses & Hypothesis Testing

Christopher Skyi

August 23, 2015

Part II: Basic Exploratory Analyses & Hypothesis Testing Using the ToothGrowth Dataset

In part II of the final project for the Johns Hopkins' *Statistical Inference* class, I'll load the ToothGrowth data and perform some basic exploratory data analyses and do some t-tests, investigating the main and interaction effects of the variables.

Tooth Growth Dataset

The ToothGrowth dataset is a subset of the data reported in E.W. Crampton (1946) *The Growth Of The Odontoblasts Of The Incisor Tooth As A Criterion Of The Vitamin C Intake Of The Guinea Pig.* The Journal of Nutrition. The dataset was imported into R from C. I. Bliss's (1952) *The Statistics of Bioassay.* Academic Press.

Description

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).

The data is represented in data frame with 60 observations on 3 variables.

[,1] len numeric Tooth length (mm) [,2] supp factor Supplement type (VC or OJ). [,3] dose numeric Dose in milligrams.

From the original design reported by E.W. Crampton, I'm going to treat the data as **non-paired observations** and because of this, I'm going to assume, erring on the side of safety, that the **variances of the 6 groups are unequal**.

Exploratory Data Analysis

Both the plots and tables below show the following basic features of the data:

- Main Effect of Supplement on Tooth Growth
- Main Effect of Dose on Tooth Growth
- Interaction of Supplement and Dose on Tooth Growth

Exploring the Effect of Supplement on Tooth Growth: Orange Juice (OJ) or Ascorbic Acid (VC)

Tables and Plots

```
tdata %>%
  group_by(supp) %>%
  summarize(obs = n(), min = min(len),max = max(len), average = mean(len))
```

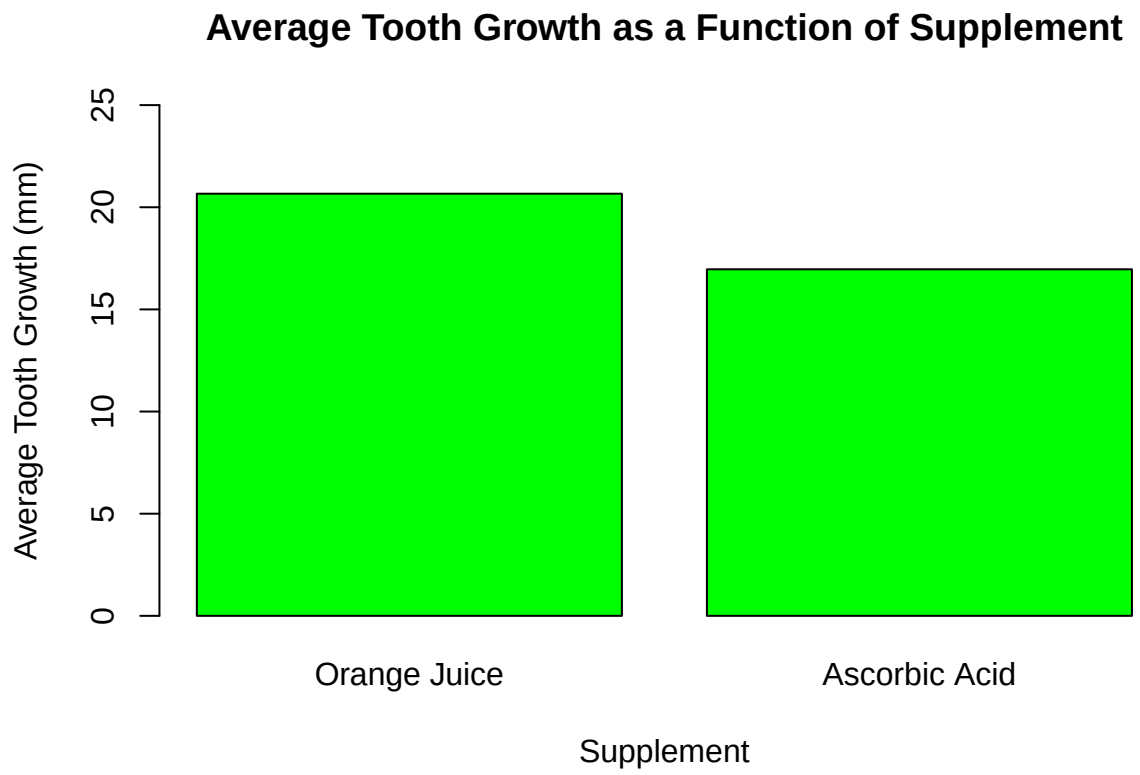
```
## Source: local data frame [2 x 5]
##
##   supp obs min  max  average
## 1   OJ  30 8.2 30.9 20.66333
## 2   VC  30 4.2 33.9 16.96333
```

On average, it appears Orange Juice promotes more growth than Ascorbic Acid, 3.7mm on average (20.66mm - 16.96mm).

The histogram below (with accompaning R code) shows the average growth in the Orange Juice and Ascorbic Acid groups:

```
supp.data <- tdata %>%
  group_by(supp) %>%
  summarize(avg = mean(len))

# barplot
barplot(supp.data$avg, xlab="Supplement", ylim = c(0, 25), col = "green", main = "Average Tooth Growth a
        ylab="Average Tooth Growth (mm)", names.arg=c("Orange Juice", "Ascorbic Acid"))
```



Testing for Significance: T-Tests

A t-test(df = 29, 0.95, t = 3.3026) (**see code below**) comparing the difference between the Orange Juice (OJ) and Ascorbic Acid (VC) groups was significant at $p < 0.05$ (p-value = 0.00255), with a 95 percent confidence interval of 1.408659mm to 5.991341mm.

Conclusion: This result supports the conclusion that Orange Juice (OJ) is significantly better than Ascorbic Acid (VC) in promoting tooth growth in guinea pigs (averaged over the dose levels).

T-Test R Code:

```
# t-test
group.oj <- filter(tdata, supp == "OJ")
group.vc <- filter(tdata, supp == "VC")
t.test(group.oj$len - group.vc$len, paired = FALSE, var.equal = FALSE)
```

Exploring the Effect of Dose on Tooth Growth: 0.5mg, 1mg, and 2 mg

Tables and Plots

```
tdata %>%
  group_by(dose) %>%
  summarize(obs = n(), min = min(len),max = max(len), average = mean(len))
```

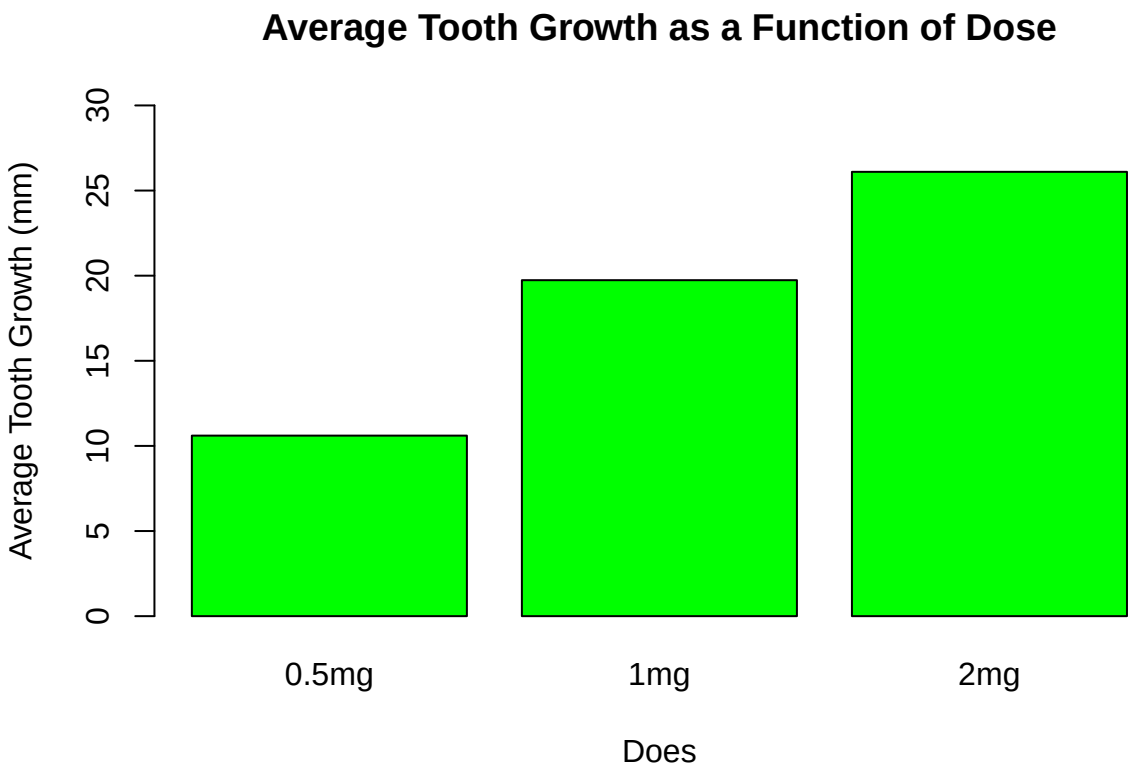
```
## Source: local data frame [3 x 5]
##
##   dose obs  min  max average
## 1  0.5  20  4.2 21.5  10.605
## 2    1  20 13.6 27.3  19.735
## 3    2  20 18.5 33.9  26.100
```

On average, it appears that increasing Dose has a positive effect on tooth growth, with an average increase of 9.13mm between the 0.5mg and 1mg dose (19.735mm - 10.605mm), an increase of 15.49mm between the 0.5mg and 2mg dose (26.100mm - 10.605mm) and 6.365mm between the 1mg and 2mg dose (26.100mm - 19.735mm).

The histogram below (with accompaning R code) shows the average grow in the 0.5mg, 1mg, and 2 mg groups:

```
dose.data <- tdata %>%
  group_by(dose) %>%
  summarize(avg = mean(len))

# barplot
barplot(dose.data$avg, xlab="Does", ylim = c(0, 30), col = "green", main = "Average Tooth Growth as a Fu
        ylab="Average Tooth Growth (mm)", names.arg=c("0.5mg", "1mg", "2mg"))
```



Testing for Significance: T-Tests

A t-test(df = 19, 0.95, t = 6.9669) (**see code below**) comparing the difference between the 0.5mg and 1mg groups (9.13mm) was significant at $p < 0.05$ (p-value = 1.225e-06), with a 95 percent confidence interval of 6.387121mm to 11.872879mm.

Another t-test(df = 19, 0.95, t = 11.291) (**see code below**) comparing the difference between the 0.5mg and 2mg groups (15.49mm) was significant at $p < 0.05$ (p-value = 7.19e-10), with a 95 percent confidence interval of 12.6228mm to 18.3672mm.

A final t-test(df = 19, 0.95, t = 4.6046) (**see code below**) comparing the difference between the 1mg and 2mg groups (6.36mm) was significant at $p < 0.05$ (p-value = 0.0001934), with a 95 percent confidence interval of 3.471814mm to 9.258186mm.

Conclusion: These results support the conclusion that averaged over type of supplement, a 2mg dose significantly promotes better tooth growth than either a 1mg or 0.5mg dose and that a 1mg dose significantly promotes better tooth growth than a 0.5mg dose in guinea pigs.

T-Test R Code:

```
# t-test
group.05 <- filter(tdata, dose == 0.5)
group.10 <- filter(tdata, dose == 1)
group.20 <- filter(tdata, dose == 1)
t.test(group.10$len - group.05$len, paired = FALSE, var.equal = FALSE)
t.test(group.20$len - group.05$len, paired = FALSE, var.equal = FALSE)
t.test(group.20$len - group.10$len, paired = FALSE, var.equal = FALSE)
```

Exploring the Interaction of Dose & Supplement on Tooth Growth

Tables and Plots

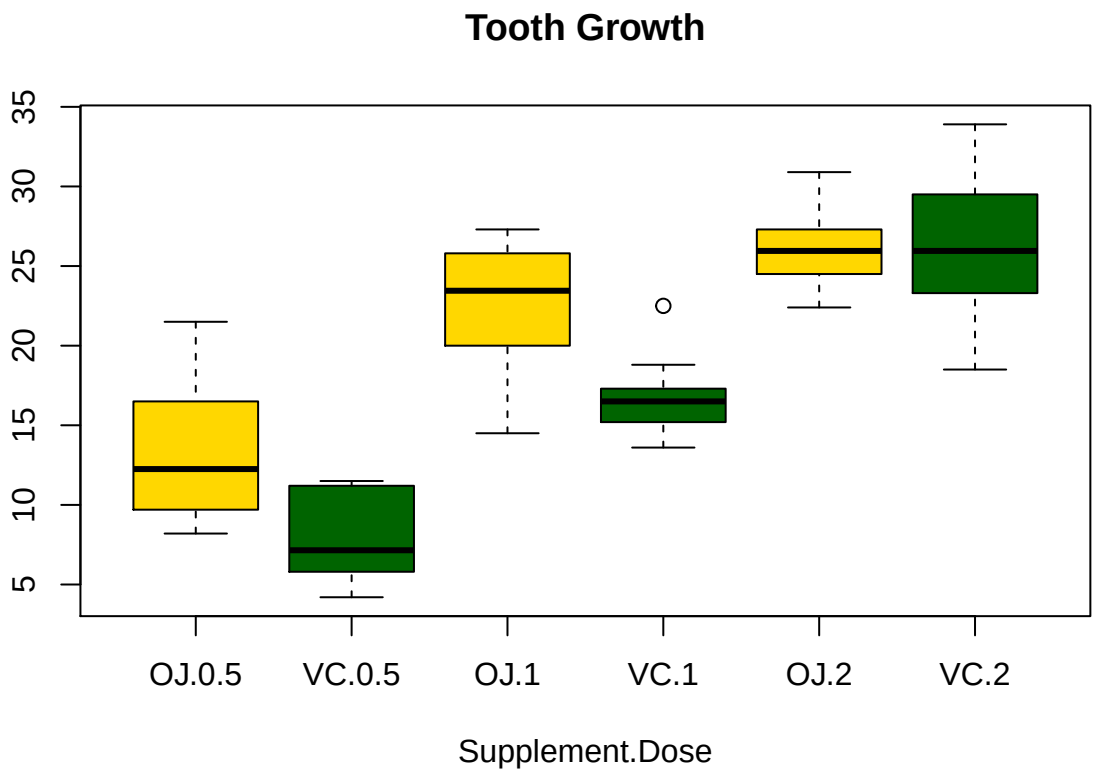
```
tdata %>%
  group_by(supp,dose) %>%
  summarize(obs = n(), min = min(len),max = max(len), average = mean(len))
```

```
## Source: local data frame [6 x 6]
## Groups: supp
##
##   supp dose obs  min  max average
## 1   OJ  0.5  10  8.2 21.5   13.23
## 2   OJ    1  10 14.5 27.3   22.70
## 3   OJ    2  10 22.4 30.9   26.06
## 4   VC  0.5  10  4.2 11.5    7.98
## 5   VC    1  10 13.6 22.5   16.77
## 6   VC    2  10 18.5 33.9   26.14
```

On average, the table data suggests that Orange Juice (OJ) is more effective than Ascorbic Acid (VC) in promoting more growth in each dose level except in the 2.0mg level.

The boxplot below (with accompaning R code) shows this interaction:

```
boxplot(len~supp*dose, data=tdata, notch=FALSE,
        col=(c("gold","darkgreen")),
        main="Tooth Growth", xlab="Supplement.Dose")
```



Testing for Significance: T-Tests

A t-test(df = 9, 0.95, t = 2.9791) (**see code below**) comparing the difference between Orange Juice (OJ) and Ascorbic Acid (VC) in the 0.5mg dose condition was significant at $p < 0.05$ (p-value = 0.01547), with a mean difference of 5.25mm and a 95 percent confidence interval of 1.263458mm to 9.236542mm.

Another t-test(df = 9, 0.95, t = 3.3721) (**see code below**) comparing the difference between Orange Juice (OJ) and Ascorbic Acid (VC) in the 1mg dose condition was significant at $p < 0.05$ (p-value = 0.008229), with a mean difference of 5.93mm and a 95 percent confidence interval of 1.951911mm to 9.908089mm.

A final t-test(df = 9, 0.95, t = -0.042592) (**see code below**) comparing the difference between Orange Juice (OJ) and Ascorbic Acid (VC) in the 2mg dose condition was **not** significant at $p < 0.05$ (p-value = 0.967).

Conclusion: These results support the conclusion that Orange Juice (OJ) is more effective than Ascorbic Acid (VC) in promoting more growth in each dose level except in the 2.0mg level.

T-Test R Code:

```r
# t-test
group.oj.05 <- filter(tdata, dose == 0.5 & supp == "OJ")
group.vc.05 <- filter(tdata, dose == 0.5 & supp == "VC")
group.oj.1 <- filter(tdata, dose == 1 & supp == "OJ")
group.vc.1 <- filter(tdata, dose == 1 & supp == "VC")
group.oj.2 <- filter(tdata, dose == 2 & supp == "OJ")
group.vc.2 <- filter(tdata, dose == 2 & supp == "VC")
t.test(group.oj.05$len - group.vc.05$len, paired = FALSE, var.equal = FALSE)
t.test(group.oj.1$len - group.vc.1$len, paired = FALSE, var.equal = FALSE)
t.test(group.oj.2$len - group.vc.2$len, paired = FALSE, var.equal = FALSE)
```