Statistical Inference Project, Part 2: Tooth Growth Analysis

Fenton Taylor

August 20, 2016

Synopsis

The ToothGrowth data frame contains data on the effects of vitamin C on tooth growth in guinea pigs. The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC).

In this analysis, we want to determine if supplement type and dose affect tooth growth in guinea pigs.

1. Load the ToothGrowth data and perform some basic exploratory data analyses.

```
# Load libraries and data
library(datasets)
library(dplyr)
library(grid)
library(ggplot2)
library(knitr)

tg <- ToothGrowth
```

```
str(tg)
```

```
## '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 ...
```

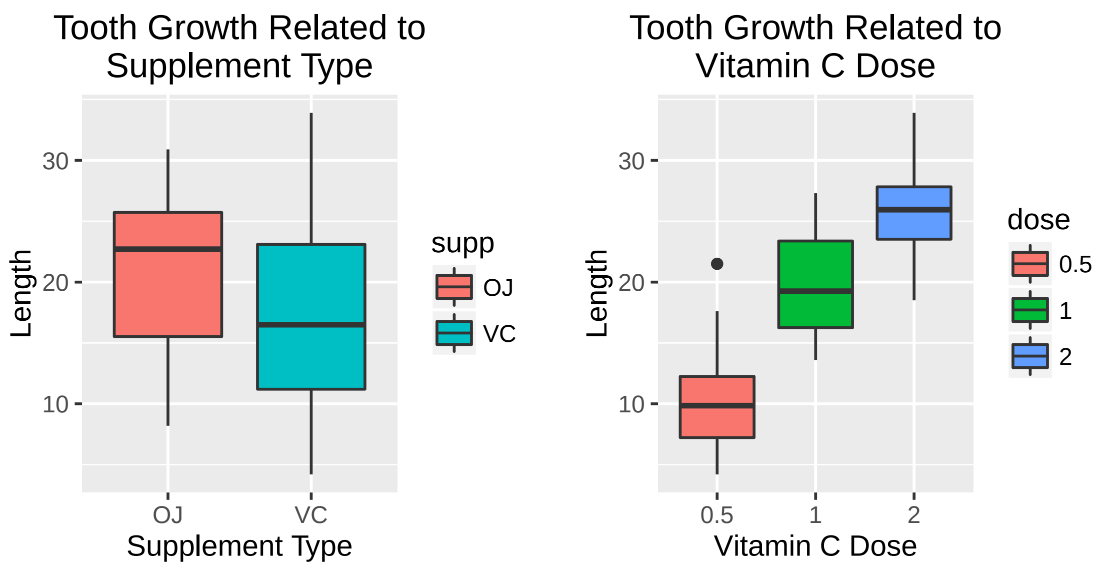
So, ToothGrowth contains 60 observations of 3 variables.

- Column 1, "len", is numeric and contains the tooth length measurement.
- Column 2, "supp", is a factor with 2 levels containing the supplement type.
- Column 3, "dose", is numeric and contains the dose size.

Note: "dose" will be converted to a factor for analysis purposes.

Let's look at some boxplots to see if there appears to be any relationship between type of supplement and tooth length, and vitamin C dose and tooth length.

```r
tg$dose <- factor(tg$dose)
p1 <- ggplot(tg, aes(x=supp, y=len)) +
      geom_boxplot(aes(fill=supp)) +
      ggtitle("Tooth Growth Related to\nSupplement Type") +
      labs(x="Supplement Type", y="Length")
p2 <- ggplot(tg, aes(x=dose, y=len)) +
      geom_boxplot(aes(fill=dose)) +
      ggtitle("Tooth Growth Related to\nVitamin C Dose") +
      labs(x="Vitamin C Dose", y="Length")

pushViewport(viewport(layout = grid.layout(1, 2)))
print(p1, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
print(p2, vp = viewport(layout.pos.row = 1, layout.pos.col = 2))
```



There certainly appears to be a positive correlation between vitamin C dose and tooth length. It appears that as dosage increases, so does tooth length. However, it is not clear if there is a significant difference in tooth growth with the different supplements.

2. Provide a basic summary of the data.

So, let's take a look at each combination of supplement and dose:

- orange juice: 0.5, 1, 2 mg of vitamin C per day
- ascorbic acid: 0.5, 1, 2 mg of vitamin C per day

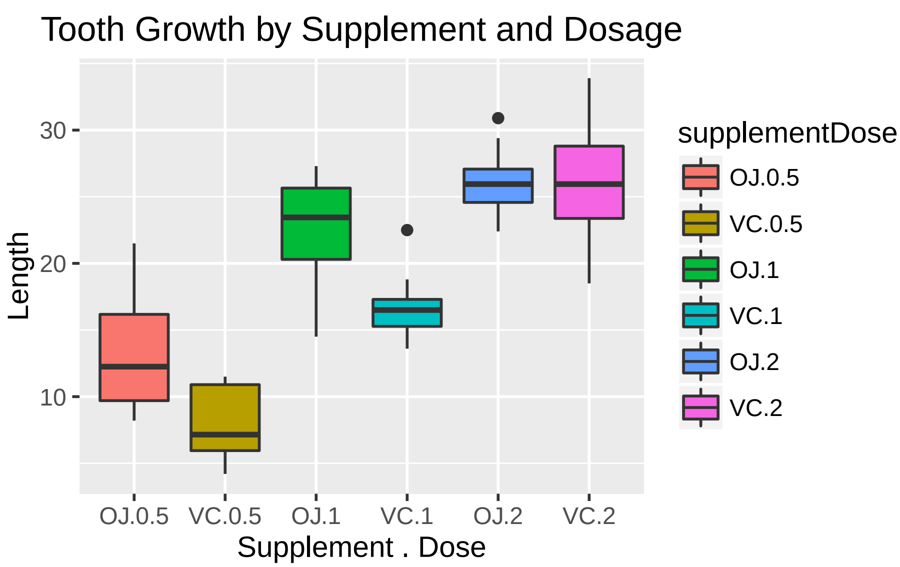
```r
tg$supplementDose <- interaction(tg$supp, tg$dose)
tgMeans <- tg %>% group_by(supplementDose) %>% summarize(meanLength=mean(len))

kable(tgMeans)
```

supplementDose	meanLength
OJ.0.5	13.23
VC.0.5	7.98
OJ.1	22.70
VC.1	16.77
OJ.2	26.06
VC.2	26.14

So, it seems like the supplement orange juice results in a greater tooth growth than ascorbic acid at dosages of 0.5 and 1.0 mg/day of vitamin C. However, the supplement type does not seem to matter at a dosage of 2.0 mg/day. Increased dosage does seem to correlate with increased tooth growth for both supplements.

The following visual representation of each supplement/dose group supports that hypothesis.

```
p3 <- ggplot(tg, aes(x=supplementDose, y=len)) +
      geom_boxplot(aes(fill=supplementDose)) +
      ggtitle("Tooth Growth by Supplement and Dosage") +
      labs(x="Supplement . Dose", y="Length")
p3
```



3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

Assumptions

1. Sample sizes are small, so it is necessary to use a t-test.
2. Variances are assumed to be unequal, so 'var.eaqual = F' will be used in each test.

T-tests and Summary

We want to determine if supplement type affects tooth growth independent of vitamin C dose, if vitamin C dose affects tooth growth independent of supplement type, and if there is any significant difference in tooth growth between supplements at each dosage. Therefore the following 7 t-tests are necessary:

- OJ vs. VC
- 0.5 vs. 1.0 mg/day
- 1.0 vs. 2.0 mg/day
- 0.5 vs. 2.0 mg/day
- OJ 0.5 vs. VC 0.5
- OJ 1.0 vs. VC 1.0
- OJ 2.0 vs. VC 2.0

```
# Create subsets for each dose level.
tg05 <- filter(tg, dose==0.5)
tg10 <- filter(tg, dose==1)
tg20 <- filter(tg, dose==2)

#######################
## T-Tests
#######################
# OJ vs. VC
t_ojvc <- t.test(len~supp, data=tg)
# 0.5 vs. 1.0 mg/day
t_0510 <- t.test(tg05$len,tg10$len, data=tg)
# 1.0 vs. 2.0 mg/day
t_1020 <- t.test(tg10$len,tg20$len, data=tg)
# 0.5 vs. 2.0 mg/day
t_0520 <- t.test(tg05$len,tg20$len, data=tg)
# 0.5 mg OJ vs. VC
t_05ojvc <- t.test(len~supp, data=tg05)
# 1.0 mg OJ vs. VC
t_10ojvc <- t.test(len~supp, data=tg10)
# 0.5 mg OJ vs. VC
t_20ojvc <- t.test(len~supp, data=tg20)

# Create list of t-test outputs to loop over and extract p-values and
# confidence intervals to display in the test summary.
t_tests <- list(t_ojvc, t_0510, t_1020, t_0520, t_05ojvc, t_10ojvc, t_20ojvc)
pvals <- numeric()
cilower <- numeric()
ciupper <- numeric()
for(i in 1:length(t_tests)){
      pvals[i] <- t_tests[[i]][[3]]
      cilower[i] <- round(t_tests[[i]][4]$conf.int[1],4)
      ciupper[i] <- round(t_tests[[i]][4]$conf.int[2],4)
}
t_summary <- data.frame("p_value"=round(pvals,6),"lower.CI"=cilower,"upper.CI"=ciupper,
            row.names = c("OJ vs. VC","0.5 vs. 1.0 mg/day","1.0 vs. 2.0 mg/day",
                          "0.5 vs. 2.0 mg/day", "0.5 mg OJ vs. VC",
                          "1.0 mg OJ vs. VC","2.0 mg OJ vs. VC"))
kable(t_summary)
```

	p_value	lower.CI	upper.CI
OJ vs. VC	0.060635	-0.1710	7.5710
0.5 vs. 1.0 mg/day	0.000000	-11.9838	-6.2762
1.0 vs. 2.0 mg/day	0.000019	-8.9965	-3.7335
0.5 vs. 2.0 mg/day	0.000000	-18.1562	-12.8338

	p_value	lower.CI	upper.CI
0.5 mg OJ vs. VC	0.006359	1.7191	8.7809
1.0 mg OJ vs. VC	0.001038	2.8021	9.0579
2.0 mg OJ vs. VC	0.963852	-3.7981	3.6381

4. Conclusion

As stated before, variances were assumed to be unequal, so the default parameter 'var.equal = FALSE' was set for each t-test. R automatically calcualted the pooled variance for each t-test.

Supplement type

In the first test, the type of supplement was not found to have a significant effect on average tooth growth, independent of dose size. However, the p-value for the orange juice vs. ascorbic acid test was 0.060635, which is fairly close to our limit of 0.05. Furthermore, the boxplot of each supplement/dose combination appeared to show significant differences in supplement types for some dose levels, so comparisons of supplement types should be made for each dose level.

At 0.5 and 1.0 mg/day of vitamin C, there is a significant difference in tooth growth between supplement types. However, at 2.0 mg/day of vitamin C, there was no significant difference in supplement type.

- At 0.5 mg/day, the mean tooth growth for guinea pigs that were given orange juice and ascorbic acid were 13.23 and 7.98, respectively. Those means are statistically different at a p-value of 0.006359.
- At 1.0 mg/day, the mean tooth growth for guinea pigs that were given orange juice and ascorbic acid were 22.7 and 16.77, respectively. Those means are statistically different at a p-value of 0.001038.
- At 2.0 mg/day, the mean tooth growth for guinea pigs that were given orange juice and ascorbic acid were 26.06 and 26.14, respectively. Those means are not statistically different at a p-value of 0.963852.

Dose Size

For each increase in dose size, independent of supplement type, there is a significant positive correlation in tooth growth.

- The mean length at 0.5, 1.0, and 2.0 mg/day are 10.605, 19.735, and 26.1, respectively.
- The difference between mean length at 0.5 and 1.0 mg/day was statistically significant at a p-value of 0.0000001.
- The difference between mean length at 1.0 and 2.0 mg/day was statistically significant at a p-value of 0.0000191.
- The difference between mean length at 0.5 and 2.0 mg/day was statistically significant at a p-value of 0.