Statistical Inference Project: Tooth Growth Analysis

Stephanie Denis

3/25/2017

Overview

The objective of this exercise is to compare the response of tooth growth in guinea pigs to the daily dose amount and delivery method administered. The results provide strong evidence of a higher response in tooth growth when the dose amount is doubled. On the other hand, there is no discernible difference in the average tooth growth between the two delivery methods.

Data Description

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

The data set consists of 60 observations and 3 variables:

1. ***len*** - tooth length
2. ***supp*** - supplement type (orange juice OJ or vitamin C VC)
3. ***dose*** - dose amount in milligrams per day (0.5, 1, or 2)

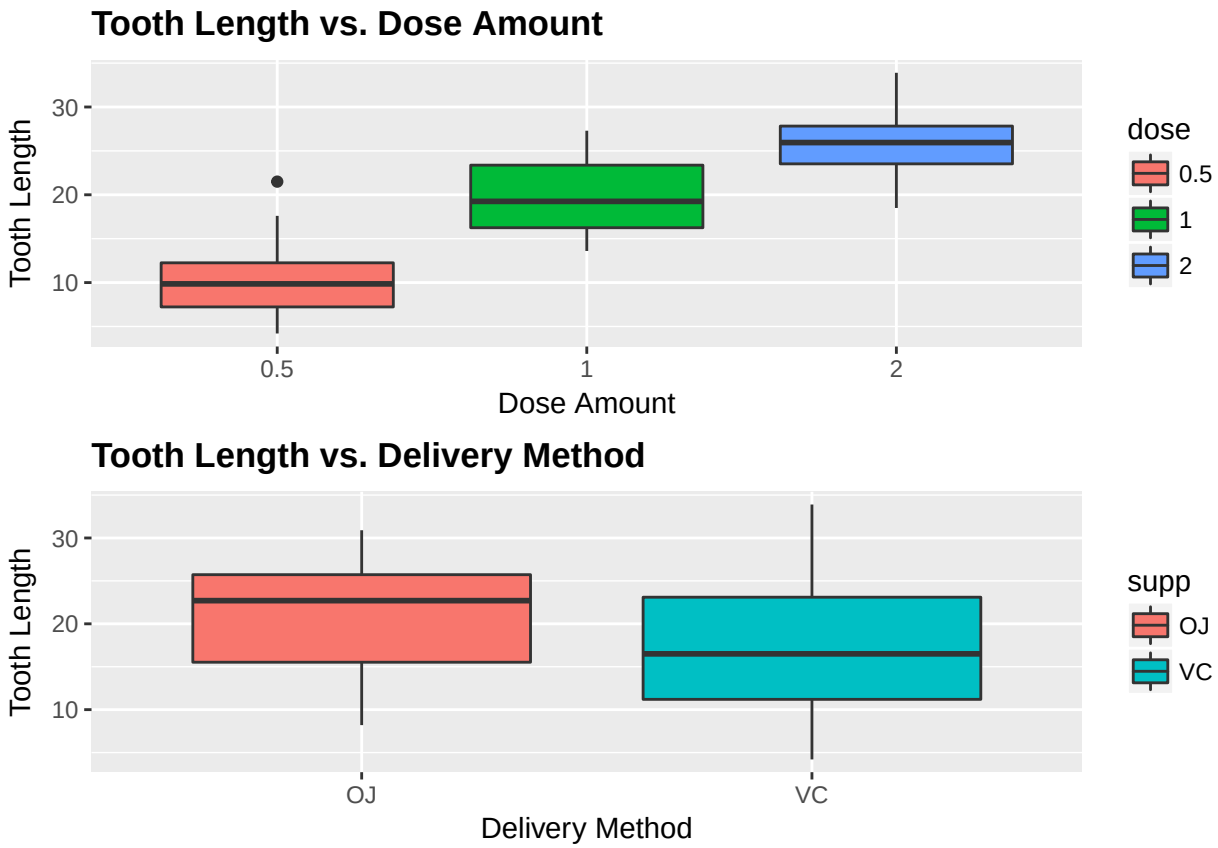
Exploratory Data Analysis

```
##   dose   mean       sd
## 1  0.5 10.605 4.499763
## 2  1.0 19.735 4.415436
## 3  2.0 26.100 3.774150
```

```
##   supp     mean       sd
## 1   OJ 20.66333 6.605561
## 2   VC 16.96333 8.266029
```

Tooth Length vs. Dose Amount: The above summary statistics show a positve relationship between the response in tooth growth and the daily dose amount. Moreover, a doze of 1 mg/day is associated with a significantly higher mean in tooth length compared to a dose of 0.5 mg/day. In the box plots illustrated below, the 2 IQRs (the 2 rectangles) do not overlap along the vertical axis, suggesting that the 2 population means differ beyond just random variation. A similar observation can be made between the group receiving 1 mg/day and the one receiving 2 mg/day.

Tooth Length vs. Delivery Method: The average tooth growth is higher when the delivery method is OJ. However, in the box plots illustrated below, the 2 IQRs overlap along the vertical axis, suggesting that the 2 population means may not differ significantly.



Hypothesis Testing

Assumptions

1. ***Independence:*** i.i.d random sample of guinea pigs, less than 10 percent of the population
2. ***Sample size:*** $n = 60$, large enough sample
3. ***Error rate:*** the probability of a Type 1 error rate is $\alpha = 0.05$

Tooth Length vs. Dose Amount

$H_0 : \mu_{0.5mg} = \mu_{1mg}$; $H_A : \mu_{0.5mg} < \mu_{1mg}$

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

The mean difference in tooth growth between the two dose amounts is -9.13. The 95 percent Student's T-confidence interval is below zero, suggesting less tooth growth in the group receiving 0.5 mg/day. Given

the very low p-value, we can reject H_0 and accept H_A that the mean tooth growth is *lower* for the group receiving 0.5 mg/day than the group receiving 1 mg/day.

$H_0 : \mu_{1mg} = \mu_{2mg}$; $H_A : \mu_{1mg} < \mu_{2mg}$

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

The mean difference in tooth growth between the two dose amounts is -6.36. Similary to the previous comparison, the 95 percent Student's T-confidence interval is also below zero, suggesting less tooth growth in the group receiving 1 mg/day compared to the group receiving 2 mg/day. Given the very low p-value, we can reject H_0 and accept H_A that the mean tooth growth is *lower* for the group receiving 1 mg/day.

Tooth Length vs. Delivery Method

$H_0 : \mu_{CV} = \mu_{OJ}$; $H_A : \mu_{CV} < \mu_{OJ}$

```
##
##  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 mean difference in tooth growth for the group recieving the VC method is -3.7 compared to the group receiving the OJ method. The 95 percent Student's T-confidence interval contains zero, suggesting that we cannot rule out zero as the possibility for population difference between the two groups. Given the p-value is 0.06, we fail to reject H_0 at the 5 percent confidence level and conclude that the data do not provide convincing evidence of a difference between the average tooth growth when the delivery method is OJ and the average tooth growth when the delivery method is VC.

Conclusion

The results provide strong evidence of a higher response in tooth growth when the dose amount is doubled, for example from 0.5 mg/day to 1 mg/day and similarly from 1 mg/day to 2 mg/day. On the other hand, there is no discernible difference in the average tooth growth between the two delivery methods.

Appendix: R Code

```
#upload & check data structure
data(ToothGrowth)
str(ToothGrowth)
```

```
#load packages
library(dplyr); library(ggplot2); library(grid); library(gridExtra); library(plyr)
```

```
#summary statistics
ddply(ToothGrowth, ~ dose, summarise, mean = mean(len), sd = sd(len))
```

```
ddply(ToothGrowth, ~ supp, summarise, mean = mean(len), sd = sd(len))
```

```
#convert dose from numeric to factor
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
#boxplot by dose amount
p1 <- ggplot(aes(x = dose, y = len), data = ToothGrowth) +
        geom_boxplot(aes(fill = dose)) +
        labs(x = "Dose Amount", y = "Tooth Length") +
        ggtitle("Tooth Length vs. Dose Amount") +
        theme(plot.title = element_text(face="bold"))
#boxplot by delivery method
p2 <- ggplot(aes(x = supp, y = len), data = ToothGrowth) +
        geom_boxplot(aes(fill = supp)) +
        labs(x = "Delivery Method", y = "Tooth Length") +
        ggtitle("Tooth Length vs. Delivery Method") +
        theme(plot.title = element_text(face="bold"))
#panel
grid.arrange(p1, p2, nrow = 2, ncol = 1)
```

```
#t-test by dose amount
subdata <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5, 1))
g1 <- subdata$len[subdata$dose==0.5]
g2 <- subdata$len[subdata$dose==1]
difference <- g1 - g2
dmn <- round(mean(difference), 2)
t.test(len ~ dose, data = subdata)
```

```
#remove
rm(subdata, g1, g2, difference, dmn)
#t-test by dose amount
subdata <- subset(ToothGrowth, ToothGrowth$dose %in% c(1, 2))
g1 <- subdata$len[subdata$dose==1]
g2 <- subdata$len[subdata$dose==2]
difference <- g1 - g2
dmn <- round(mean(difference), 2)
t.test(len ~ dose, data = subdata)
```

```
#t-test by delivery method
g1 <- ToothGrowth$len[ToothGrowth$supp=="VC"]
g2 <- ToothGrowth$len[ToothGrowth$supp=="OJ"]
difference <- g1 - g2
dmn <- round(mean(difference), 2)
t.test(len ~ supp, data = ToothGrowth)
```