Effectiveness of Vitamiin C Delivery Methods on Tooth Growth

Cha Li

06/20/2015

Introduction

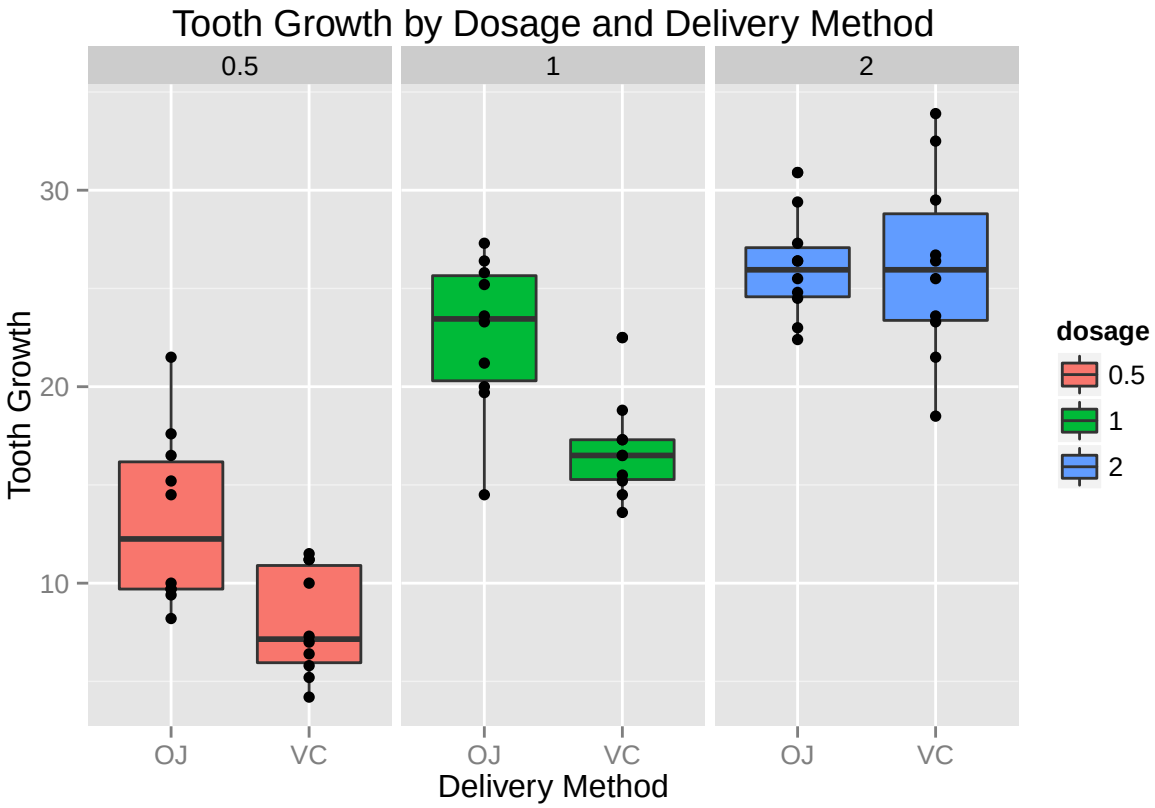
The purpose of this analysis is to examine the effect of vitamin C on tooth growth in guinea pigs. There are two control variables in the ToothGrowth dataset: delivery method and dosage. This report will focus on the effectiveness of the two delivery methods at varying dosage levels.

```
# load useful libraries and the dataset
library(datasets)
library(dplyr)
library(ggplot2)

data(ToothGrowth)
```

Exploratory Analysis

The graphy below groups the treatments by dosage and then within each dosage level the data is split again by delivery method. In the .5mg and 1mg scenarios we can see that there appears to be a difference in tooth growth between the delivery methods. This does not seem to be the case, however, in the 2mg scenario.

```
# create a new column which is just, the other two factors combined. this is
# for easy graphing
working_data <- ToothGrowth %>%
    mutate(treatment=interaction(ToothGrowth$supp, ToothGrowth$dose)
           , dose=as.factor(dose)
           , supp=as.factor(supp)) %>%
    rename(method=supp, dosage=dose, growth=len)

ggplot(working_data) +
    geom_boxplot(aes(y=growth, x=method, fill=dosage)) +
    geom_point(aes(y=growth, x=method)) +
    facet_grid(.~dosage, ) +
    ggtitle("Tooth Growth by Dosage and Delivery Method") +
    labs(x="Delivery Method", y="Tooth Growth")
```



We can also see here that the variances between delivery method within each dosage is very different. This is an important observation that will be used later on for building confidence intervals.

```
summary_data <- working_data %>%
    group_by(dosage, method) %>%
    summarize(n=n(), mu_growth=mean(growth), var_growth=var(growth))
print(summary_data)
```

```
## Source: local data frame [6 x 5]
## Groups: dosage
##
##   dosage method  n mu_growth var_growth
## 1    0.5     OJ 10     13.23  19.889000
## 2    0.5     VC 10      7.98   7.544000
## 3      1     OJ 10     22.70  15.295556
## 4      1     VC 10     16.77   6.326778
## 5      2     OJ 10     26.06   7.049333
## 6      2     VC 10     26.14  23.018222
```

Parameters and Assumptions

Before we build confidence intervals and perform hypothesis tests, we make the following assumptions:

1. Although the same 10 guinea pigs were used, we will assume that the trials between delivery methods for each dosage are independent. The dataset description implied some amount of time went by between treatments.

2. The sample sizes are small for each group so we'll assume a T-distribution.
3. The variance between delivery methods given a dosage value is very different.

Confidence Intervals and Hypothesis Testing

Using the three points above, the 95% confidence intervals will be defined assuming a two-sample unpooled t-distribution. The confidence interval will be built using the following equation:

$$(\hat{X}_1 - \hat{X}_2) \pm t_{\alpha/2, n+m-2} \times \sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}$$

$$n = m = 10, \alpha = .05, df = 18$$

The critical result from defining the interval is when 0 is outside the interval. If this occurs we can reject the null hypothesis, $H_0 : \mu_1 = \mu_2$, in favor of the alternate hypothesis, $H_A : \mu_1 \neq \mu_2$.

```
t_score <- qt(.975, 18)

# get the means and variances we need for the first dosage level (.5,g)
dosage <- summary_data %>% filter(dosage==0.5)
delta <- dosage$mu_growth[1] -dosage$mu_growth[2]

# the confidence interval
ci_dosage1 <- delta + c(-1, 1) * t_score * sqrt(
    ((dosage$var_growth[1]/dosage$n[1])) + (dosage$var_growth[2]/dosage$n[2]))

# repeat the process for the 1mg and 2mg
dosage <- summary_data %>% filter(dosage==1)
delta <- dosage$mu_growth[1] -dosage$mu_growth[2]
ci_dosage2 <- delta + c(-1, 1) * t_score * sqrt(
    ((dosage$var_growth[1]/dosage$n[1])) + (dosage$var_growth[2]/dosage$n[2]))

dosage <- summary_data %>% filter(dosage==2)
delta <- dosage$mu_growth[1] -dosage$mu_growth[2]
ci_dosage3 <- delta + c(-1, 1) * t_score * sqrt(
    ((dosage$var_growth[1]/dosage$n[1])) + (dosage$var_growth[2]/dosage$n[2]))

results <-rbind(ci_dosage1, ci_dosage2, ci_dosage3)
```

Results

```
##                 [,1]     [,2]
## ci_dosage1  1.770262 8.729738
## ci_dosage2  2.840692 9.019308
## ci_dosage3 -3.722999 3.562999
```

The calculated confidence intervals show more precisely that there are significant differences in tooth growth between the delivery methods with .5mg (dosage1) and 1mg (dosage2) dosages and that there is no significant difference with the 2mg dosage. Essentially, we reject the null hypothesis in the first two cases but fail to reject it in the third case.