Statistical Inference Analysis of the Effects of VitC on Tooth Growth in Guinea Pigs

Wei Wei

May 15, 2016

Synopsis

Using the ToothGrowth dataset, the effects of doses of vitamin C and delivery methods on tooth growth in guinea pigs were analyzed. It is found that higher dose (2 mg/day) of vitamin C promoted more tooth growth than lower dose (0.5 mg/day), while delivery methods did not make a difference (2 mg/day dose).

Exploratory Data Analysis

```
data(ToothGrowth)

## basic data information
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 ...
```

```
summary(ToothGrowth$len)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.20   13.08   19.25   18.81   25.28   33.90
```

The ToothGrowth dataset contains 60 observations on 3 variables (len, supp, and dose). Basic data analysis shows that the tooth growth variable (len) has a median of 19.25 and a mean of 18.81. Exploratory analysis (see Appendix Figure 1) shows that 1) there is a clear upward trend of tooth length with increasing dose of vitamin C using either delivery methods (OJ or VC), and 2) at high dose of vitamin C (2 mg/day), there is essentially no difference in the medians of tooth length using OJ or VC delivery methods. Therefore whether delivery methods affect tooth growth is not pursued further. Additionally, analyses on 1 mg/day vitamin C dose are omitted for simplicity.

Comparison of the means of tooth growth after 2 mg/day vs 0.5 mg/day vitamin C

Two sample T-tests are carried out separately on the means of tooth growth after 2 mg/day or 0.5 mg/day doses of vitamin C for the two delivery methods. The H0 hypothesis in each test is: there is no difference in the means of tooth growth after 2 mg/day or 0.5mg/day vitamin C treatment.

```
## subset individual data groups
suppressPackageStartupMessages(library(dplyr))
OJ05 <- ToothGrowth %>% filter(dose=="0.5" & supp=="OJ")
OJ2 <- ToothGrowth %>% filter(dose=="2" & supp=="OJ")
```

```
VC05 <- ToothGrowth %>% filter(dose=="0.5" & supp=="VC")
VC2 <- ToothGrowth %>% filter(dose=="2" & supp=="VC")

## independent group t-tests
## OJ as delivery method
t.test(OJ2$len, OJ05$len, paired=F, var.equal=T, alternative="two.sided")
```

```
##
##  Two Sample t-test
##
## data:  OJ2$len and OJ05$len
## t = 7.817, df = 18, p-value = 3.402e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   9.381777 16.278223
## sample estimates:
## mean of x mean of y
##     26.06     13.23
```

The Two Sample T-test shows that the mean of tooth growth is 26.06 for 2 mg/day, vs 13.23 for 0.5 mg/day vitamin C when using OJ delivery method. With a p-value of 3.402e-07, the H0 hypothesis that there is no difference in the means is rejected. The 95% confidence interval is (9.381777, 16.278223), which does not overlap with 0. It is therefore concluded with confidence that 2 mg/day vitamin C promoted more tooth growth than 0.5 mg/day when using OJ delivery method.

```
## VC as delivery method
t.test(VC2$len, VC05$len, paired=F, var.equal=T, alternative="two.sided")
```

```
##
##  Two Sample t-test
##
## data:  VC2$len and VC05$len
## t = 10.388, df = 18, p-value = 4.957e-09
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  14.48716 21.83284
## sample estimates:
## mean of x mean of y
##     26.14      7.98
```

The Two Sample T-test shows that the mean of tooth growth is 26.14 for 2 mg/day, vs 7.98 for 0.5 mg/day vitamin C when using VC delivery method. With a p-value of 4.957e-09, the H0 hypothesis that there is no difference in the means is rejected. The 95% confidence interval is (14.48716, 21.83284), which does not overlap with 0. It is therefore concluded with confidence that 2 mg/day vitamin C promoted more tooth growth than 0.5 mg/day when using VC delivery method.

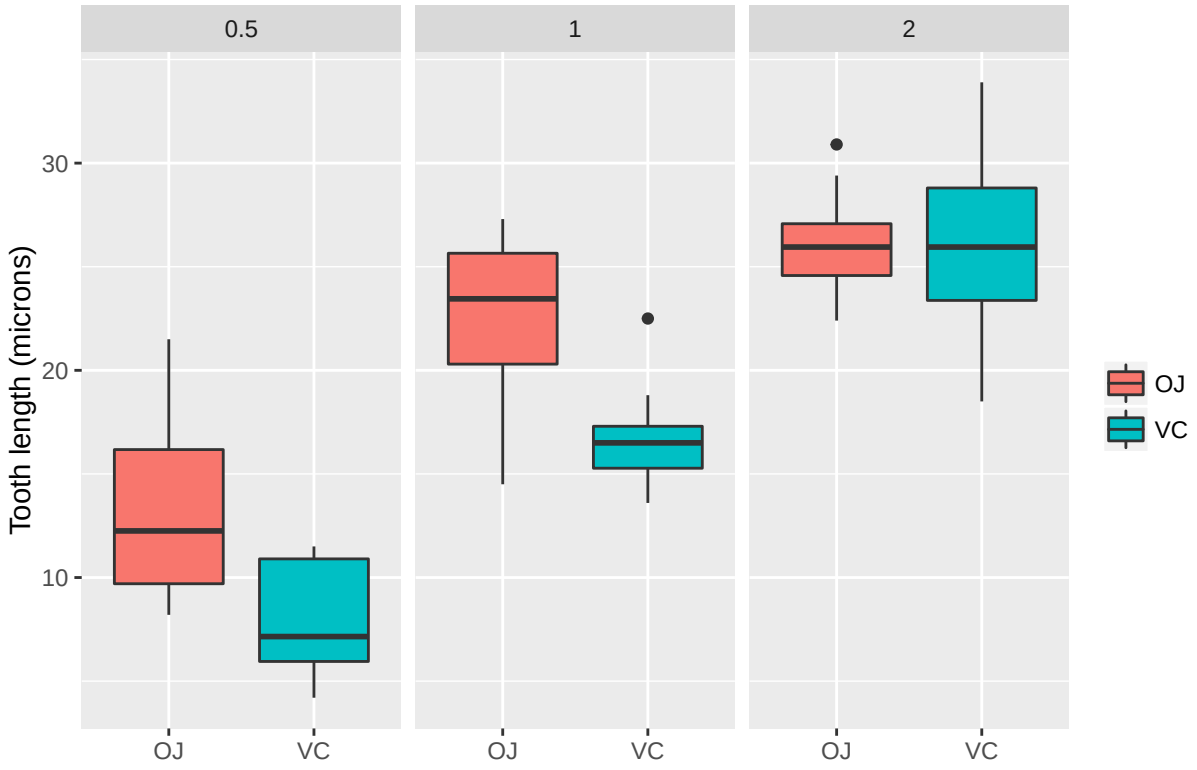
Assumptions made using Two Sample T-test

The t-interval assumes that the data is independently and identically distributed. To attest this, the distribution of the entire data in the tooth length variable is plotted (see Appendix Figure 2). The distribution of the data appears to be roughly mount-shaped and symmetric. Therefore the use of the t-interval test is justified. Additional assumptions are that the variances are equal in the compared groups.

Appendix

```
## Appendix figure 1. plot len in respect to dose and supp
library(ggplot2)
ToothGrowth$dose <- factor(ToothGrowth$dose)
g <- ggplot(ToothGrowth, aes(x=supp, y=len))
g <- g + geom_boxplot(aes(fill=supp))
g <- g + facet_wrap(~ dose)
g <- g + labs(x='', fill='', y='Tooth length (microns)',
title='Effects of vitamin C dose and delivery method on tooth growth')
g <- g + ggtitle("Figure 1. Effects of vitamin C dose and delivery method on tooth growth")
g
```

Figure 1. Effects of vitamin C dose and delivery method on tooth growth




```
## Appendix figure 2. distribution of the len variable
g <- ggplot(ToothGrowth, aes(x=len))
g <- g + geom_histogram(binwidth=3, alpha=0.75)
g <- g + ggtitle("Figure 2. Histogram of tooth growth data")
g
```

Figure 2. Histogram of tooth growth data