Statistical Inference Part 2 - Vitamin C Impact on Guinea Pig Tooth Growth

Betsy Nash

January 14, 2018

Synopsis

The question posed is to analyze guinea pig tooth growth by vitamin C supplement and dose.

Specifically, “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).” - https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html (https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html)

The result of this review is that while there are too few observations to draw reliable conclusions, early indications show dose of vitamin C can play a role in tooth growth. Further research is recommended.

Loaded Libraries

```
library(dplyr)
library(tidyr)
library(reshape)
library(ggplot2)
```

Create the DataSet

The dataset is availble in R and called “ToothGrowth”.

```
data("ToothGrowth")
```

Explore the DataSet for Reasonability

The following tests were performed on the dataset. In the interest of report length, only the summarized table will be presented showing the count of observations by dose and supplement.

```
#dim(ToothGrowth)
#names(ToothGrowth)
#str(ToothGrowth)
##check no numeric values beyond acceptable boundaries for columns of interest (ie neg
ative)
#filter(ToothGrowth, dose < 0)
#UniqueSupp<-distinct(select(ToothGrowth,supp))
#UniqueDose<-distinct(select(ToothGrowth,dose))
##five number summary (min, max, etc)
#summary(ToothGrowth)
##check for na's
#sum(is.na(ToothGrowth$len))
#sum(is.na(ToothGrowth$supp))
#sum(is.na(ToothGrowth$dose))
with(ToothGrowth, table(dose, supp))
```

```
##      supp
## dose  OJ VC
##   0.5 10 10
##   1   10 10
##   2   10 10
```

These are small counts. This review will investigate if there are any statistically significant findings in the data.

The tables below show the mean and variances of tooth length, respectively, by supplement and dose.

```
#desired matrix of means
S<-ToothGrowth %>%
        group_by(supp, dose)%>%
        summarise(MeanLength = mean(as.numeric(len), na.rm=TRUE))
#now arrange
S2<-tidyr::spread(S,dose,MeanLength)
S2
```

```
## # A tibble: 2 x 4
## # Groups:   supp [2]
##     supp `0.5`   `1`   `2`
## * <fctr> <dbl> <dbl> <dbl>
## 1     OJ 13.23 22.70 26.06
## 2     VC  7.98 16.77 26.14
```

```
#desired matrix of variances
S3<-ToothGrowth %>%
        group_by(supp, dose)%>%
        summarise(VarLength = var(as.numeric(len), na.rm=TRUE))
#now arrange
S4<-tidyr::spread(S3,dose,VarLength)
S4
```

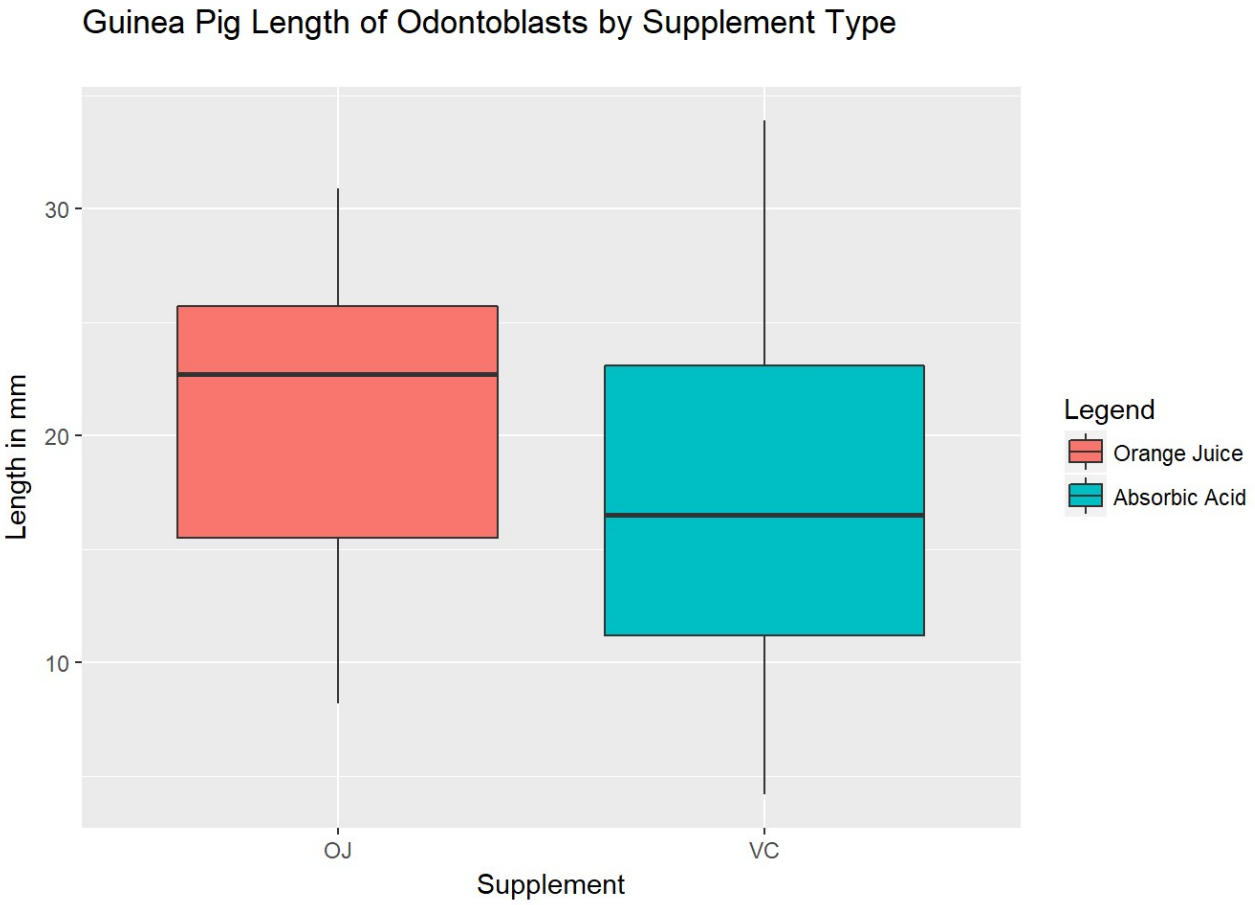
```
## # A tibble: 2 x 4
## # Groups:   supp [2]
##     supp  `0.5`       `1`       `2`
## * <fctr>  <dbl>     <dbl>     <dbl>
## 1     OJ 19.889 15.295556  7.049333
## 2     VC  7.544  6.326778 23.018222
```

There is a large spread of values in each table. The results appear to be reasonable (ie no negative or extreme values). The next steps are to chart the data and investigate for statistical significance.

Plot the Data

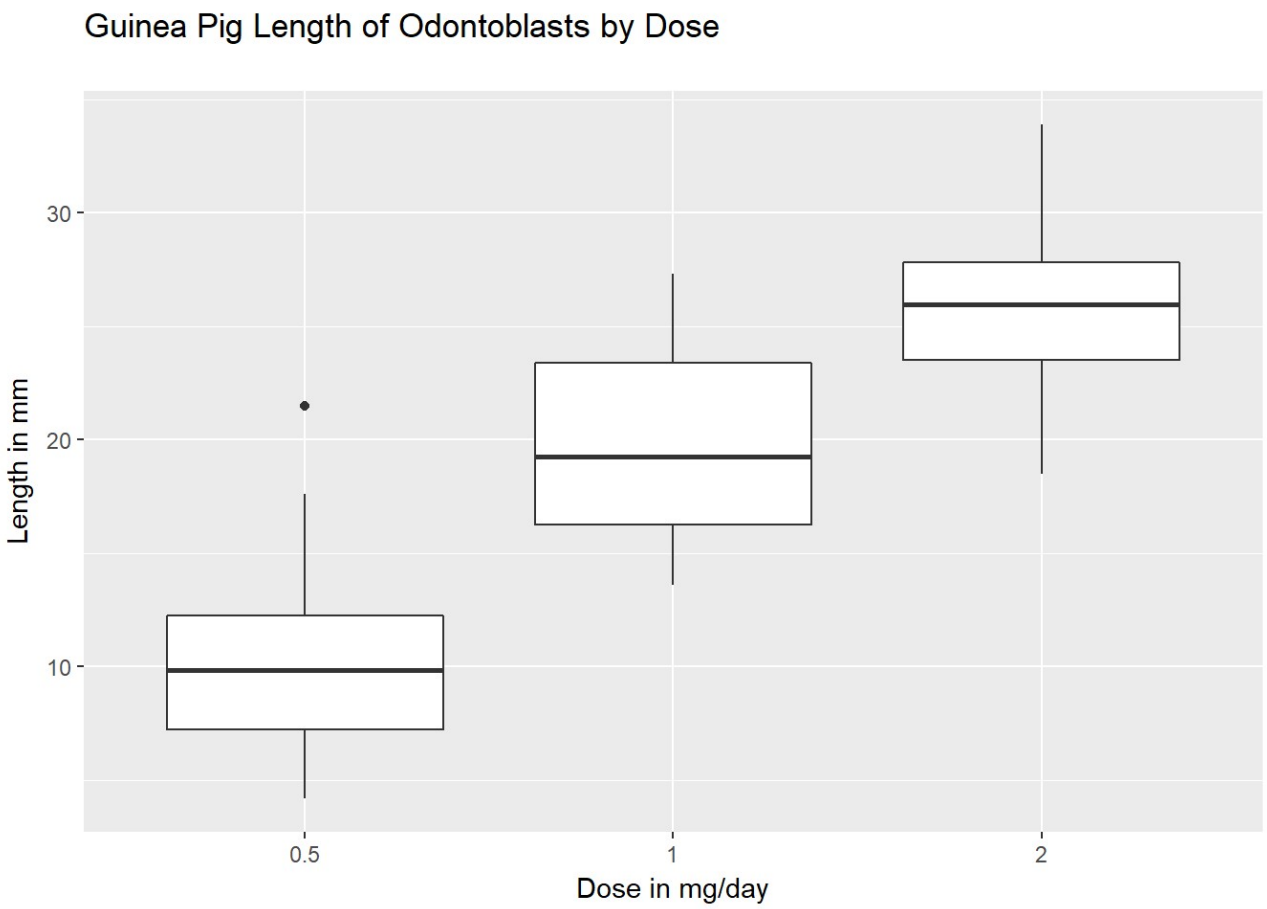
The first box plot shows tooth length by supplement. The horizonal bars represent the median.

```
MyPlot = ggplot(ToothGrowth, aes(x=supp,y=len,fill=supp))
MyPlot = MyPlot + geom_boxplot()
#label and title
MyPlot = MyPlot + ylab("Length in mm") + xlab("Supplement") + ggtitle("Guinea Pig Leng
th of Odontoblasts by Supplement Type", subtitle = "")
#legend
MyPlot= MyPlot + theme(legend.position = "right") + scale_fill_discrete(name="Legend",
labels=c("Orange Juice", "Absorbic Acid"))
MyPlot
```



The second box plot shows tooth length by dose. The horizonal bars represent the median.

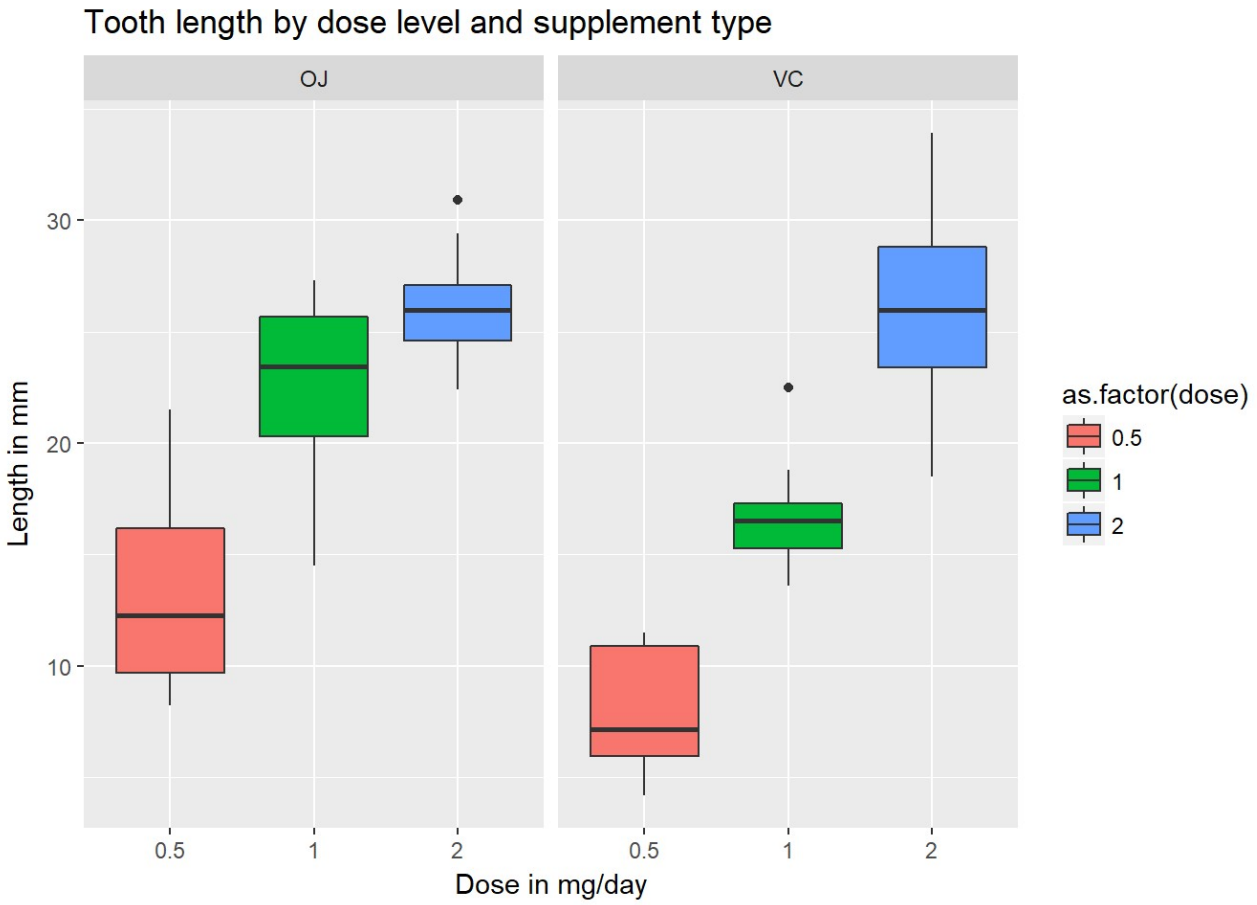
```
MyPlot2 = ggplot(ToothGrowth, aes(x=as.factor(dose),y=len))
MyPlot2 = MyPlot2 + geom_boxplot()
#Labels
MyPlot2= MyPlot2 + xlab("Dose in mg/day") +ylab("Length in mm") + ggtitle("Guinea Pig
Length of Odontoblasts by Dose", subtitle = "")
MyPlot2
```



There is quite a bit of overlap in the box plot by supplement (first plot). This is one indication the data by supplement is more similar than different. Note the more distinct separation of the box plots by dose (second plot). This indicates there may be some evidence in the data that dose has an impact on tooth length.

The next plot digs deeper and provides a side-by-side of supplement and dose.

```
ggplot(ToothGrowth, aes(as.factor(dose), len)) +
        geom_boxplot(aes(fill = as.factor(dose))) +
        facet_grid(. ~ supp) +
        xlab("Dose in mg/day") +
        ylab("Length in mm") +
        ggtitle('Tooth length by dose level and supplement type')
```



The dose of 2 mg/day shows consistent results in tooth length by supplement type. The next step is see if the difference in tooth length is significantly different than 0.

Hypothesis Testing

This section addresses the question 'Does the data provide strong evidence that the average difference in tooth length across two groups is different than 0?'. The groups used in the comparisons are as follows.

Note the conservative assumption is made that the variances are not equal.

Supplement Type

The t-test is used for evaluation. The selected alpha is 5%. The null hypothesis is that the difference in the means between the two supplement types is 0. If the p-value is greater than the selected alpha, then we fail to reject the null hypothesis. In other words, the null hypothesis stands. The second test is if the confidence interval includes 0, fail to reject the null hypothesis.

```
n<-30
degfree<-n-1
alpha<-0.05

#mean
SampleMeanSupp<-ToothGrowth %>%
        group_by(supp)%>%
        summarise(MeanLength = mean(as.numeric(len), na.rm=TRUE))
SampleMeanOJ<-SampleMeanSupp[1,2]
SampleMeanVC<-SampleMeanSupp[2,2]
DeltaSupp = SampleMeanOJ - SampleMeanVC

#variance
SampleVarSupp<-ToothGrowth %>%
        group_by(supp)%>%
        summarise(VarLength = var(as.numeric(len), na.rm=TRUE))
SampleVarOJ<-SampleVarSupp[1,2]
SampleVarVC<-SampleVarSupp[2,2]
PooledSD<-sqrt(((n-1)*SampleVarOJ +(n-1)*SampleVarVC)/(n + n -2))

#test the two supplement groups for significance, OJ vs VC
TSuppDiff<-t.test(len ~ supp, data=ToothGrowth, var.equal = FALSE, paired=FALSE)
TSuppDiff
```

```
##
##  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 p-value is > 5%. The confidence interval also includes 0. Therefore, the null hypothesis is not rejected (stands). This means the data does not provide strong evidence that supplement type impacts tooth length.

Dose

The same test is performed across 3 pairs: doses (0.5 & 1), doses (1 & 2), and doses (0.5 & 2).

The test for doses (0.5 & 1):

```
DataDose051 <- subset(ToothGrowth, dose %in% c(0.5, 1))
TDose051Diff<-t.test(len ~ dose, data=DataDose051, var.equal = FALSE, paired=FALSE)
TDose051Diff
```

```
##
##  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 p-value is < 5% and the confidence interval does not contain 0. Therefore, reject the null hypothesis. The data has strong evidence that the average tooth length of guinea pigs is different for each dose level (0.5 & 1 mg/day).

The test for doses (1 & 2):

```
DataDose12 <- subset(ToothGrowth, dose %in% c(1, 2))
TDose12Diff<-t.test(len ~ dose, data=DataDose12, var.equal = FALSE, paired=FALSE)
TDose12Diff
```

```
##
##  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 p-value is < 5% and the confidence interval does not contain 0. Therefore, reject the null hypothesis. The data has strong evidence that the tooth length of guinea pigs is different for each dose level (1 & 2 mg/day).

The test for doses (0.5 & 2):

```
DataDose052 <- subset(ToothGrowth, dose %in% c(0.5, 2))
TDose052Diff<-t.test(len ~ dose, data=DataDose052, var.equal = FALSE, paired=FALSE)
TDose052Diff
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

The p-value is < 5% and the confidence interval does not contain 0. Therefore, reject the null hypothesis. The data has strong evidence that the tooth length of guinea pigs is different for each dose level (0.5 & 2 mg/day).

In each dose test case, the tooth length does vary by dose and it is statistically significant.

Challenge the Results

There are two challenges to the data: 1) number of observations, and 2) the study of tooth length over time for each guinea pig (pair-wise data).

The number of observations is noted as low. A power test can be performed to determine the probability of rejecting the null hypothesis when it is false. The higher the power the better, and the selected target is at least 80%.

```
power.t.test(n=30,delta=as.numeric(DeltaSupp), sd=as.numeric(PooledSD),sig.level = 0.0
5, type="one.sample",strict=TRUE)
```

```
##
##      One-sample t test power calculation
##
##               n = 30
##           delta = 3.7
##              sd = 7.482001
##       sig.level = 0.05
##           power = 0.7447435
##     alternative = two.sided
```

The power is < 80%. It is recommended to increase the observation count to have more reliable conclusions.

Conclusion

This review investigated if vitamin C, in varying supplement types and dose, has an impact on length of odontoblasts for guinea pigs. The early indication is that dose does have statistially significant evidence in the data to impact tooth length. This indication is based on a limited number of observations. Further research is recommended with a larger number of subjects to draw more reliable conclusions.