The Effect of Vitamin C on Tooth Growth in Guinea Pigs

Vasil Yordanov aka b1ck0

16 June, 2017

Description

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).

Exploratory Data Analysis

We will start by loading the dataset and the libraries which we plan to use:

```
library(datasets)
library(ggplot2)

df = ToothGrowth
```

Next we want to see the database structure:

```
str(df)
```

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

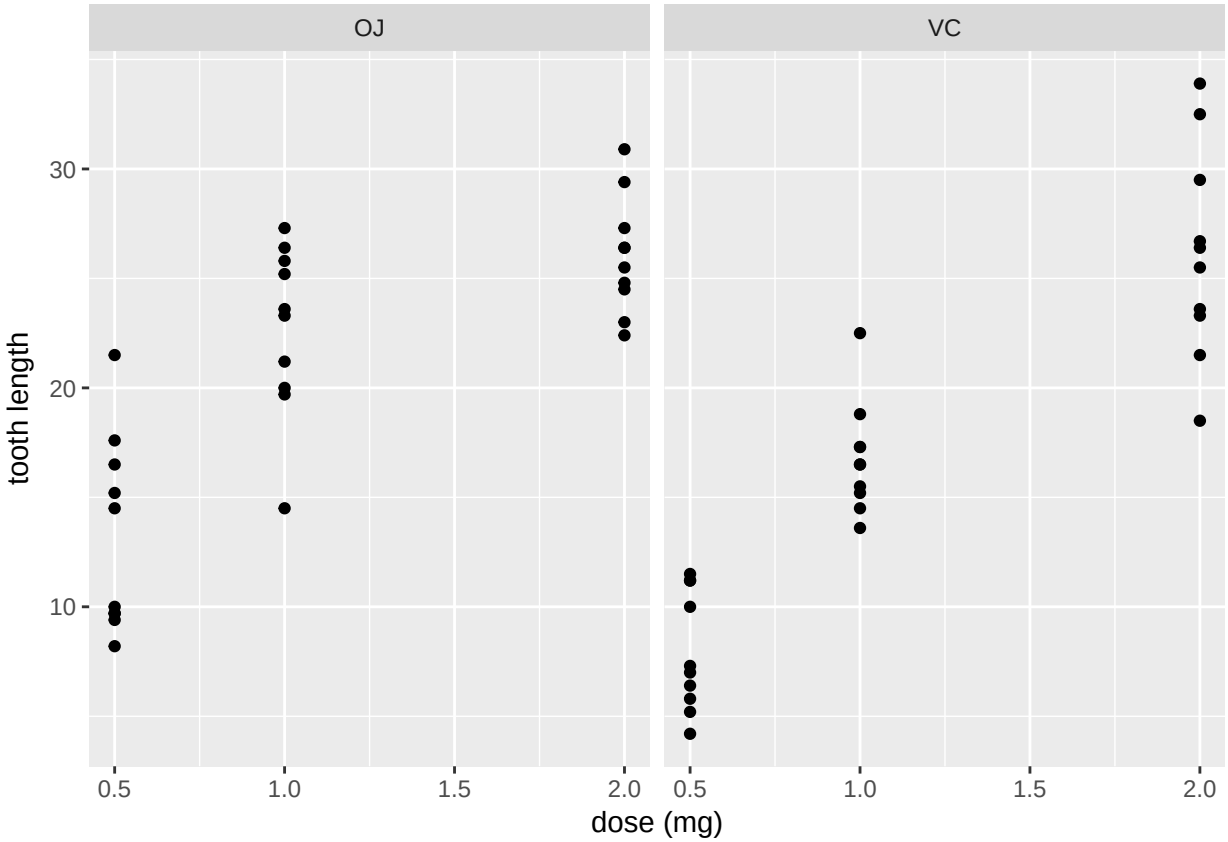
There are only three variables in this dataset with total of 60 observations. Now let's get a feeling how the data is distributed:

```
summary(df)
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

We can note that there is no missing data from the dataframe. Now let's visualize these numbers:

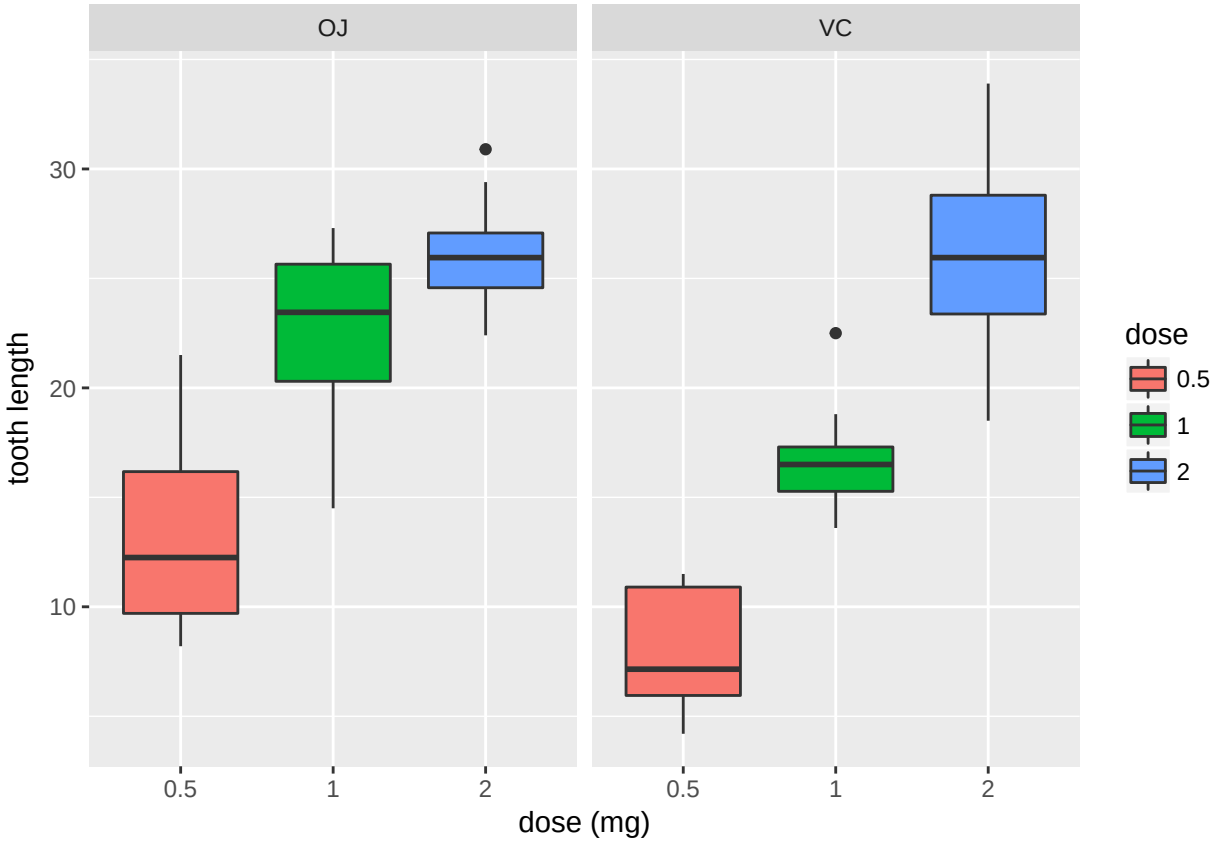
```
ggplot(df, aes(x=dose, y=len)) +
geom_point() +
facet_grid(.~supp) +
labs(x="dose (mg)", y="tooth length")
```



Here we see something interesting, it seems that the variable dose can be represented as a factor variable as in this case it take only three discrete values - 0.5, 1.0 and 1.5. Based on this we will convert it to a factor variable and plot the data again (now taking advantage of the capabilities which we get from the factor):

```
df$dose = as.factor(df$dose)

ggplot(df, aes(x=dose, y=len, fill=dose)) +
geom_boxplot() +
facet_grid(.~supp) +
labs(x="dose (mg)", y="tooth length")
```



Now it definately looks way more informative. From this plot it seems that the gunea pigs which recieved 0.5 ~ 1.0 mg of Orange Juice (OJ) have longer teeth that tose which recieved the same amount of Vitamic C (VC). However we do not observe this trend for the dosage of 2 mg. This is great as now we have our first hypothesis which we will test ## Statistical Inference

The NULL Hypothesis will be that the Orange Juice doesn't influence the teeth length of guinea pigs in dosage of 0.5 and 1.0 mg.

$$H_0 : \mu_{OJ} = \mu_{VC}$$

Now we will test our hypothesis in R:

```
subset.OJ = df$len[df$supp=="OJ" & df$dose=="0.5"] # subsetting the guinea pigs which were given Orange
subset.VC = df$len[df$supp=="VC" & df$dose=="0.5"] # subsetting the guinea pigs which were given Vitami

mean.OJ = mean(subset.OJ) # calculating the sample mean for OJ subset
mean.VC = mean(subset.VC) # calculating the sample mean for VC subset

sd.OJ = sd(subset.OJ) # calculating the sample variance for OJ subset
sd.VC = sd(subset.VC) # calculating the sample variance for VC subset

n.OJ = length(subset.OJ) # obtaining the number of observations for OJ
n.VC = length(subset.VC) # obtaining the number of observations for VC

err.OJ = (sd.OJ^2/n.OJ) # squared standard error for OJ
err.VC = (sd.VC^2/n.VC) # squared standard error for VC
```

```
dof = (err.OJ + err.VC)^2/(err.OJ^2/(n.OJ-1)+err.VC^2/(n.VC-1)) # degrees of freedom
t = (mean.OJ - mean.VC)/sqrt(err.OJ + err.VC) # t-statistic
p.value = pt(t,dof) # p-value
conf.int = (mean.OJ - mean.VC) + c(-1,1)*qt(0.975,dof)*sqrt(err.OJ+err.VC) # 95% confidence interval
```

The results from our hypothesis testing are: t-statistic = 3.1697 which results into p-value = 0.006359. This is strong evidance that our NULL hypothesis is wrong and in fact the alternative hypothesis is correct. Next we will check the same NULL hypothesis but this time we will be using the subset of guniea pigs who were given 1.0 mg of dosage. To speed up the things I will use the build in fuction in R:

```
subset.OJ = df$len[df$supp=="OJ" & df$dose=="1"] # subsetting the guinea pigs which were given Orange J
subset.VC = df$len[df$supp=="VC" & df$dose=="1"] # subsetting the guinea pigs which were given Vitamin

t.test(subset.OJ, subset.VC, var.equal = FALSE, paired = FALSE, alternative="two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  subset.OJ and subset.VC
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

Here the evidance in favor of the alternative hypothesis is even more extreme than before. We can again conclude that the alternative hypothesis is correct.

Lastly we will test what is the situation when the dosage = 2.0 mg.

```
subset.OJ = df$len[df$supp=="OJ" & df$dose=="2"] # subsetting the guinea pigs which were given Orange J
subset.VC = df$len[df$supp=="VC" & df$dose=="2"] # subsetting the guinea pigs which were given Vitamin

t.test(subset.OJ, subset.VC, var.equal = FALSE, paired = FALSE, alternative="two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  subset.OJ and subset.VC
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

This time the results are definately in favor for the NULL hypothesis.

Conclusions

Based on the statistical analysis shown above we can colcude:

1. Orange Juice in dosage = 0.5 mg increases the size of odontoblasts in guinea pigs, compared with the same amount of vitamin C
2. Orange Juice in dosage = 1.0 mg increases the size of odontoblasts in guinea pigs, compared with the same amount of vitamin C
3. There is no statistical evidance that Orange Juice in dosage = 2.0 mg increases the size of odontoblasts in guinea pigs, compared with the same amount of vitamin C