

# The Effect of Vitamin C on Tooth Growth in Guinea Pigs

Sreenivasulu Parimi

May 27, 2018

## Statistical Inference Course Project - Part 2: Basic Inferential Data Analysis

### A. Overview

This is for second part of the course project of the Coursera course 'Statistical Inference' which is a part of 'Data Science' specialization. In this second part, we perform basic inferential analyses using the ToothGrowth data in the R datasets package.

#### 1. Load the ToothGrowth data and perform some basic exploratory data analyses

Load the required Packages:

```
library(datasets)
library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

Load the data and run the basic exploratory analysis:

```
data("ToothGrowth")
tooth_growth <- ToothGrowth
dim(tooth_growth)

## [1] 60  3

head(tooth_growth)

##   len supp dose
## 1  4.2   VC  0.5
```

```
## 2 11.5 VC 0.5
## 3 7.3 VC 0.5
## 4 5.8 VC 0.5
## 5 6.4 VC 0.5
## 6 10.0 VC 0.5
```

```
tail(tooth_growth)
```

```
##      len supp dose
## 55 24.8 OJ 2
## 56 30.9 OJ 2
## 57 26.4 OJ 2
## 58 27.3 OJ 2
## 59 29.4 OJ 2
## 60 23.0 OJ 2
```

```
str(tooth_growth)
```

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

```
# Unique Values
```

```
unique(ToothGrowth$len)
```

```
## [1] 4.2 11.5 7.3 5.8 6.4 10.0 11.2 5.2 7.0 16.5 15.2 17.3 22.5 13.6
## [15] 14.5 18.8 15.5 23.6 18.5 33.9 25.5 26.4 32.5 26.7 21.5 23.3 29.5 17.6
## [29] 9.7 8.2 9.4 19.7 20.0 25.2 25.8 21.2 27.3 22.4 24.5 24.8 30.9 29.4
## [43] 23.0
```

```
unique(ToothGrowth$supp)
```

```
## [1] VC OJ
## Levels: OJ VC
```

```
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

The variable 'dose' can be converted into a factor variable as it seems that it is rather a level than a numeric.

```
# convert variable dose from numeric to factor
```

```
tooth_growth$dose <- as.factor(tooth_growth$dose)
```

```
str(tooth_growth)
```

```
## '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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

## B. Visualization

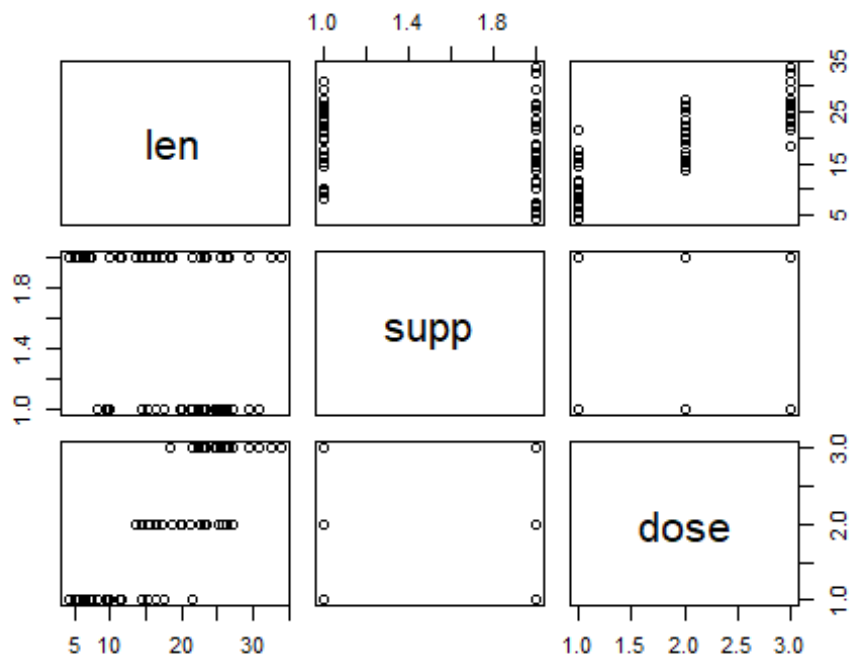
### 2. Provide a basic summary of the data

Summary statistics for the data:

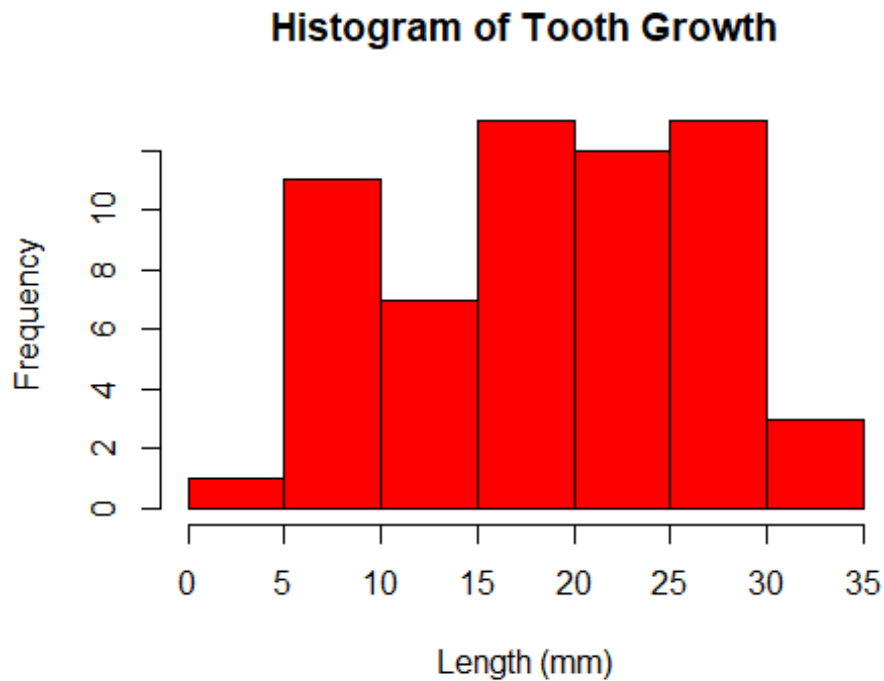
```
summary(tooth_growth)

##      len      supp      dose
##  Min.   : 4.20    OJ:30    0.5:20
##  1st Qu.:13.07    VC:30    1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90

# Structure
plot(tooth_growth)
```



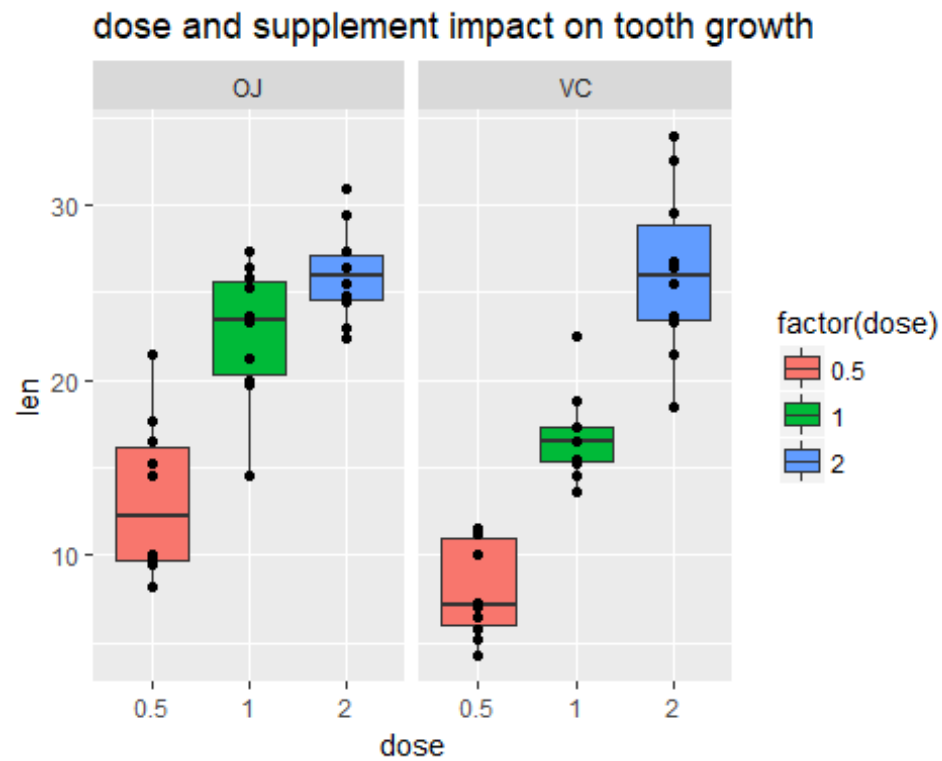
```
# Tooth Growth Histogram
hist(tooth_growth$len, col = "red", main = "Histogram of Tooth Growth", xlab =
"Length (mm)", ylab = "Frequency")
```



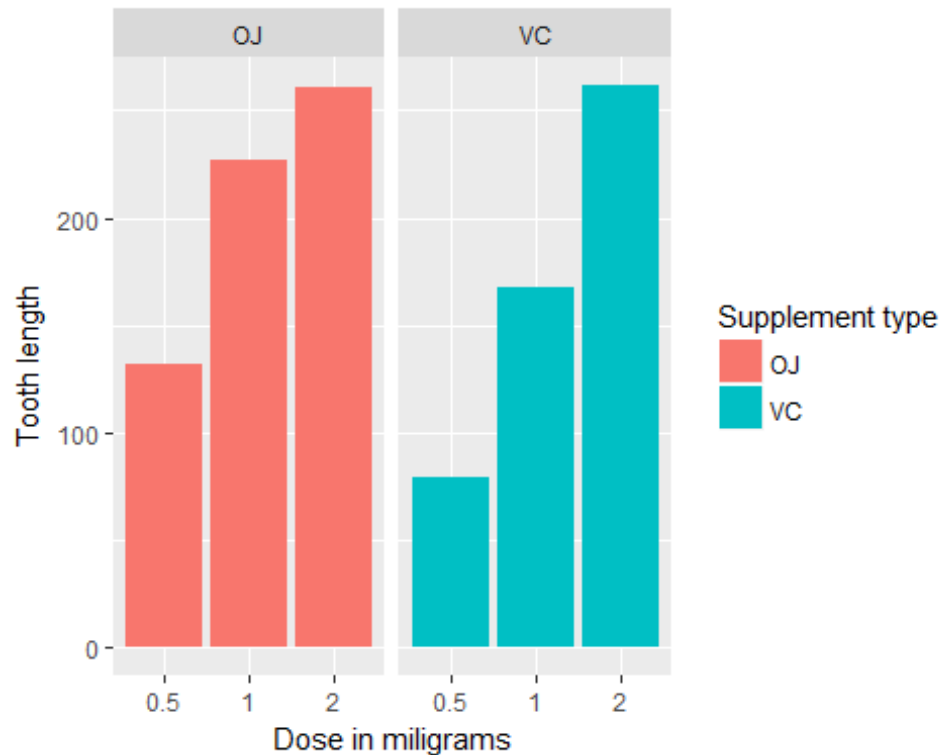
So far our analysis says that there are 60 observations, 2 types of supplements (OJ - Orange Juice & VC - Ascorbic Acid), 3 dosage sizes (0.5, 1.0, & 2mg), with more than half of the tooth length observations falling within the range of 15 - 30 mm.

#### Impact of dosage and supplement on the tooth growth

```
# Box plot
ggplot(tooth_growth, aes(x=dose, y=len)) +
  geom_boxplot(aes(fill=factor(dose))) + geom_point() + facet_grid(.~supp) +
  ggtitle("dose and supplement impact on tooth growth")
```



```
# Bar graph
ggplot(data=tooth_growth, aes(x=dose, y=len, fill=supp)) +
  geom_bar(stat="identity",) + facet_grid(. ~ supp) + xlab("Dose in milligrams")
+ ylab("Tooth length") + guides(fill=guide_legend(title="Supplement type"))
```



The above graphs shows that dose has an effect on tooth length. When the dosage is high at 2 mg, the mean value of tooth growth appears to be similar between OJ and VC, however, when the dosage is 0.5 mg or 1 mg, the chart definitely shows that OJ has a obvious positive impact on tooth growth compared to VC.

## C. Statistical Inference

### 3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)

In order to cross verify if the above insights drawn from above visual/graphical analysis are statistically valid, we perform Hypothesis/T-Tests for the tooth length as the outcome predicted by three separate vectors.

#### Effect on tooth growth/length by supplement types

- $H_0$ : Tooth length is not affected by supplement types delivery method.
- $H_a$ : Tooth length is affected by supplement types delivery method.

```
t.test(len ~ supp, data=tooth_growth)
```

```
##
## 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 0.06063 which is greater than the significance level of 0.05 and the 95% confidence interval (-0.1710156 7.5710156) which includes 0. This indicates that we can't reject the  $H_0$  null hypothesis that supplement types (OJ and/or VC) seems to have no impact on Tooth growth based on this test. So we can conclude that different supplement types have no effect on tooth length.

### Effect on tooth growth/length by various dosage

- $H_0$ : Tooth length is not affected by dose level

```
t.test(len ~ dose, data=subset(tooth_growth, dose %in% c(0.5, 1.0)))
```

```
##
## 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
```

```
t.test(len ~ dose, data=subset(tooth_growth, dose %in% c(0.5, 2.0)))
```

```
##
## 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
```

```
t.test(len ~ dose, data=subset(tooth_growth, dose %in% c(1.0, 2.0)))
```

```
##
## 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
```

For all the above three dose level pairs, the p-value is less than 0.05, and the 95% confidence interval doesn't include 0. This indicates that we can reject the  $H_0$  null hypothesis, and establish that increasing the dose level leads to an increase in tooth length. The mean tooth length increases on raising the dose level.

### Effect on tooth growth/length by supplement types and various dosage

- $H_0$ : Tooth length is not affected by supplement types at a 0.5 mg dose
- $H_1$ : Tooth length is not affected by supplement types at a 1.0 mg dose
- $H_2$ : Tooth length is not affected by supplement types at a 2.0 mg dose

```
t.test(len ~ supp, data = filter(tooth_growth, dose == 0.5), paired = F,
var.equal = F)
```

```
##
## Welch Two Sample t-test
##
## data: len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##      13.23      7.98
```

```
t.test(len ~ supp, data = filter(tooth_growth, dose == 1.0), paired = F,
var.equal = F)
```

```
##
## Welch Two Sample t-test
##
## data: len by supp
## 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 in group OJ mean in group VC
##      22.70      16.77
```

$H_0$  and  $H_1$ : Since the p-value is less than 0.05 and the 95% confidence interval doesn't cross/include 0 for above two tests, we can reject  $H_0$  and  $H_1$  with at least a 95% confidence interval.



```
t.test(len ~ supp, data = filter(tooth_growth, dose == 2.0), paired = F,
var.equal = F)

##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##           26.06           26.14
```

$H_2$ : Since the p-value is greater than 0.05 and the 95% confidence interval crosses/includes zero, we can not reject the  $H_2$  within a 95% confidence interval.

#### 4. State your conclusions and the assumptions needed for your conclusions

##### Conclusions

We can come to the following conclusions based insights drawn from above analysis.

- Supplement type has no impact of thooth growth
- There is a strong evidence includes that inreasing the dose level leads to increased tooth growth

##### Assumptions

We assumed the following in order to come to above conclusions

- supp is independent from dose
- The experient was done with random assignment of guinea pigs to different dose levels and supplement type
- Populations of guinea pigs were independent