

# Statistical Inference - Tooth Growth Data Analysis

*Keith S Messina*

*August 21, 2015*

## Introduction

This analysis uses the ToothGrowth dataset in the R datasets package to determine the statistical significance of two different delivery methods of Vitamin C and their effect on tooth odontoblast growth in guinea pigs.

## Summary of Data

The data used in this analysis shows the response in length of the tooth odontoblasts in 10 guinea pigs to three dose levels of Vitamin C delivered in two different ways: orange juice and ascorbic acid.

There are 60 observations of three variables in the dataset: length of the tooth odontoblasts (number), supplement type (factor with two levels: VC or OJ), dose of Vitamin C (number in milligrams).

A further summary of the data can be seen below:

```
library(datasets)
summary(ToothGrowth)
```

```
##      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
```

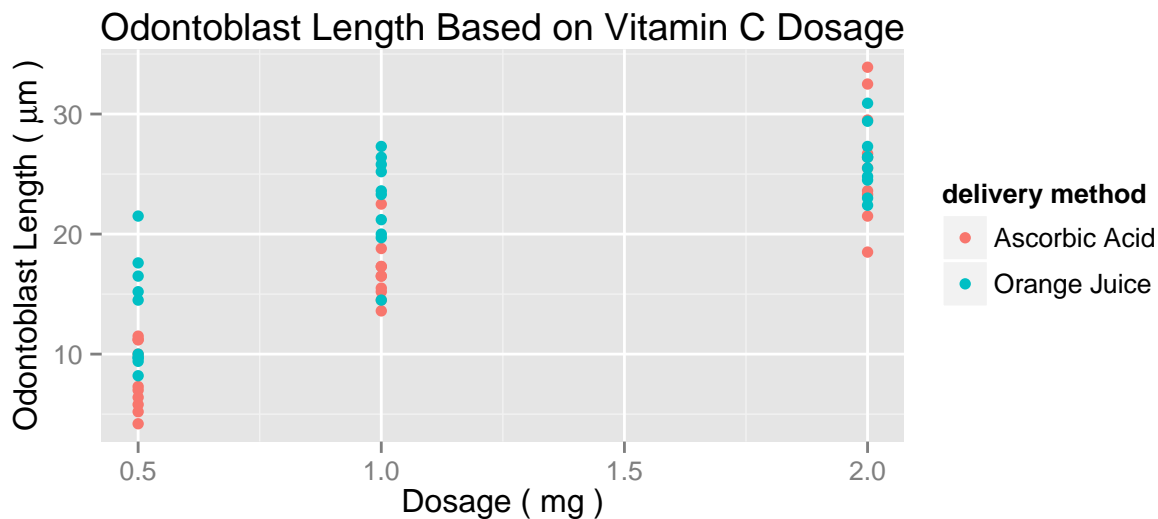
We can tidy up the data a bit and give the variables and factors more appropriate names.

```
tidyTooth <- ToothGrowth
colnames(tidyTooth) <- c("length", "delivery method", "dosage size")
tidyTooth$`delivery method` <- as.character(tidyTooth$`delivery method`)
tidyTooth$`delivery method`[tidyTooth$`delivery method` == "OJ"] <-
  "Orange Juice"
tidyTooth$`delivery method`[tidyTooth$`delivery method` == "VC"] <-
  "Ascorbic Acid"
tidyTooth$`delivery method` <- as.factor(tidyTooth$`delivery method`)
```

## Exploratory Analysis

Exploring the data via a plot of the data we can see if we can spot any trends worth investigating in the data.

```
library(ggplot2)
ggplot(data = tidyTooth) +
  geom_point(aes(x = `dosage size`, y = length, color = `delivery method`)) +
  labs(list(title = "Odontoblast Length Based on Vitamin C Dosage", x = "Dosage ( mg )",
    y = expression(paste("Odontoblast Length ( ", mu, "m )"))))
```



From the data, we can see what appears to be a trend of increasing length of tooth odontoblasts when we increase the dosage of Vitamin C. From just a cursory review, it appears that the Ascorbic Acid delivery method has a greater effect on odontoblast length, but we'll take a look at the data a bit more closely to make that call.

## Data Analysis

First, we'll calculate some standard values for the data. Throughout the analysis of this data, we are assuming that the observations are independent and identically distributed (iid), as these computations only hold if the observations are iid. The mean of the data for the **Orange Juice** delivery method is 20.66 and mean of the data for the **Ascorbic Acid** delivery method is 16.96. The standard deviation and variance of the data for the **Orange Juice** delivery method is 6.61 and 43.63 respectively and the standard deviation and variance of the data for the **Ascorbic Acid** delivery method is 8.27 and 68.33 respectively. There are 30 observations using the **Orange Juice** delivery method and 30 observations using the **Ascorbic Acid** delivery method.

The 95% confidence level of the mean for each delivery method are given below. As we can see, there is an overlap of the mean in the delivery methods, which says that there may not be enough evidence to confidently proclaim that one delivery method is better than the other.

Orange Juice: 18.1967763, 23.1298904

Ascorbic Acid: 13.8767475, 20.0499192

A two-sided t-test of the data gives us the following information:

```
t.test(tidyTooth$length[tidyTooth$"delivery method" == "Ascorbic Acid"],
       tidyTooth$length[tidyTooth$"delivery method" == "Orange Juice"],
       alternative = "two.sided", paired= FALSE)

##
##  Welch Two Sample t-test
##
## data:  tidyTooth$length[tidyTooth$"delivery method" == "Ascorbic Acid"] and tidyTooth$length[tidyTooth$"delivery method" == "Orange Juice"]
## t = 0, df = 58, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.272224  4.272224
```

```
## sample estimates:  
## mean of x mean of y  
## 16.96333 16.96333
```