# Statistical Analysis of the ToothGrowth Data

*Arun K Viswanathan*

*January 17, 2015*

## Overview

This report analyzes the ToothGrowth dataset in R.

## Summary of the Tooth Growth dataset

The *Tooth Growth* dataset in R shows the effect of Vitamin C on tooth growth in guinea pigs. The data provides is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

The data contains three variables:

| Variable | Data Type | Description |
|---|---|---|
| `len` | `numeric` | Tooth length |
| `supp` | `factor` | Supplement type (VC or OJ) |
| `dose` | `numeric` | Dose in milligrams |

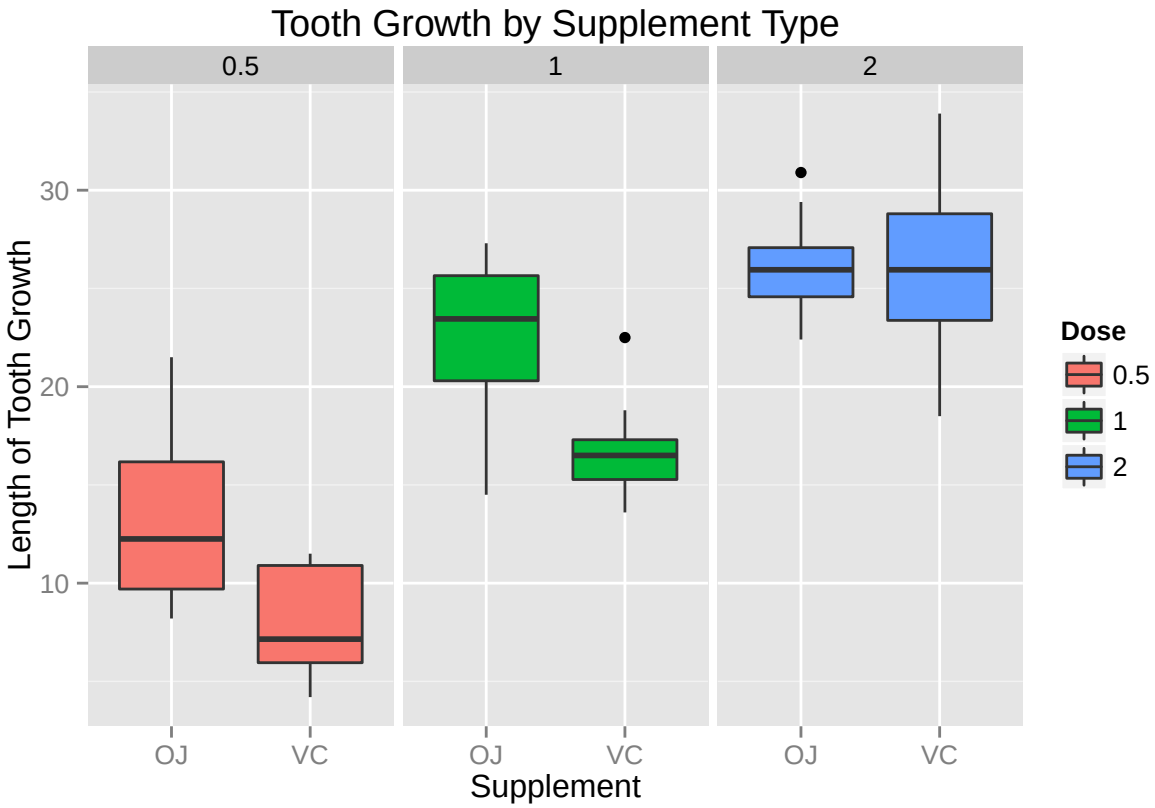
The dataset contains a total of 60 observations. Here's a summary of the raw data.

```
data <- transform(ToothGrowth, dose = as.factor(dose))
summary(data)
```

```
##       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
```

The plot below shows the tooth growth ranges for each supplement at each dosage level.

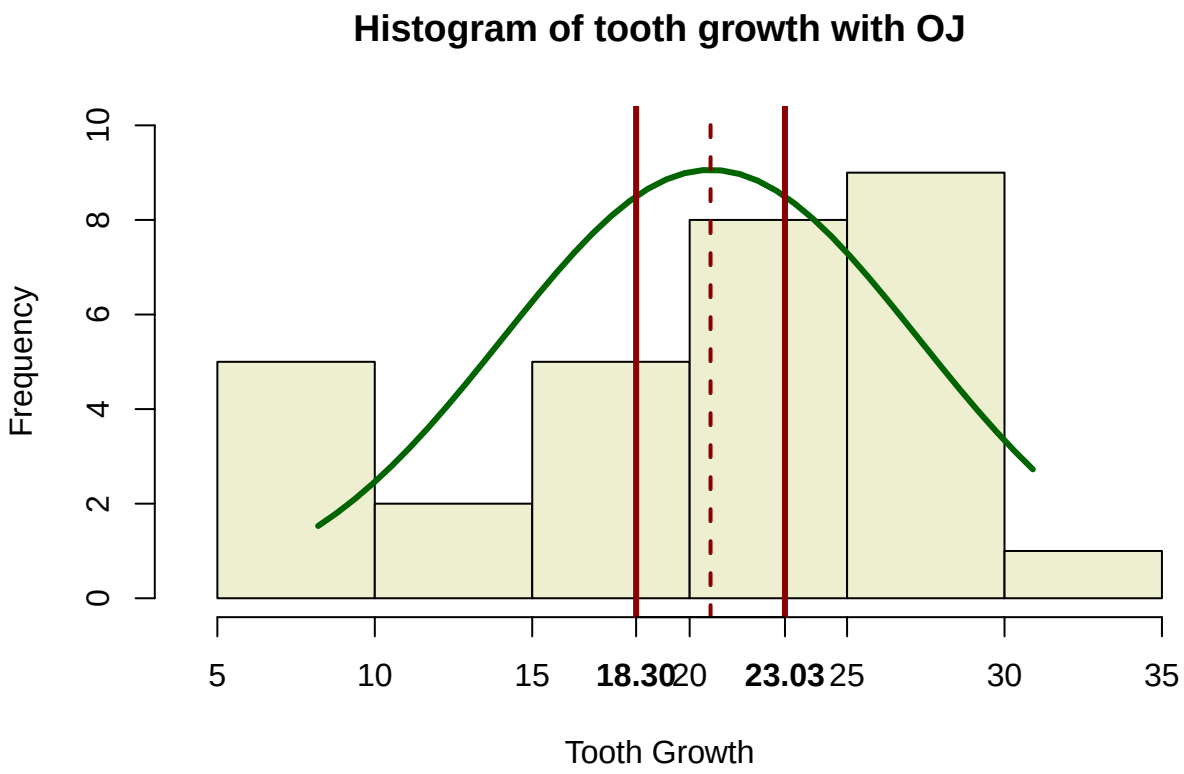
```
library(ggplot2)
g <- ggplot(data, aes(x = supp, y = len, fill = dose)) + geom_boxplot() +
    facet_grid(. ~ dose) +
    ggtitle("Tooth Growth by Supplement Type") +
    xlab("Supplement") + ylab("Length of Tooth Growth") +
    guides(fill = guide_legend(title = "Dose"))
print(g)
```

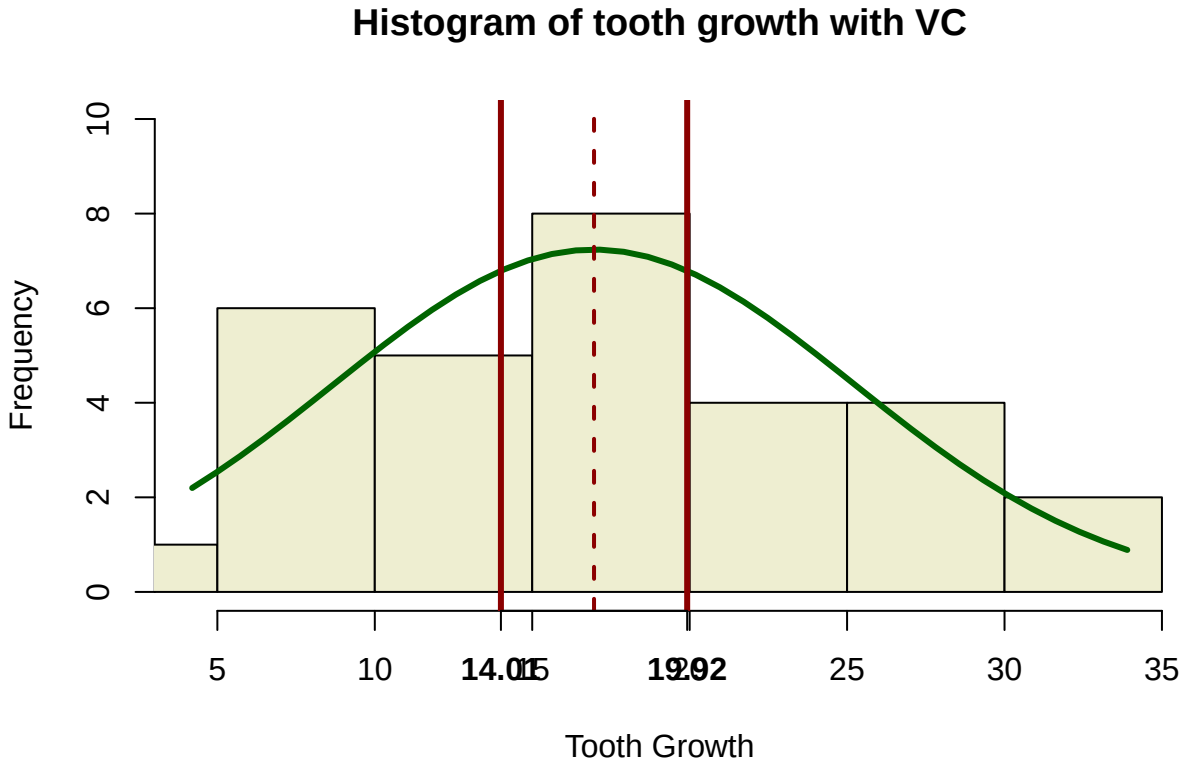



## Tooth growth analysis by supplement

Assuming that the tooth growth data is *iid*, the 95% confidence intervals for tooth growth for each supplement can be computed with the code below. The plots show the tooth growth for each supplement as a histogram.

```
conf <- lapply(split(data, data$supp), function(dataForSupp) {
    lens <- dataForSupp$len
    mean(lens) + c(-1, 1) * qnorm(0.975) * sd(lens) / sqrt(length(lens))
    })
analysis <- lapply(names(conf), function(supp) {
    lens <- data[data$supp == supp, "len"]
    h <- hist(lens, xlim = range(data$len), ylim = range(0, 10),
              main = paste0("Histogram of tooth growth with ", supp),
              xlab = "Tooth Growth", ylab = "Frequency", col = "lightyellow2")
    xfit <- seq(min(lens), max(lens), length = 40)
    yfit <- dnorm(xfit, mean = mean(lens), sd = sd(lens))
    yfit <- yfit * diff(h$mids[1:2]) * length(lens)
    lines(xfit, yfit, col = "darkgreen", lt = 1, lw = 3)
    abline(v = mean(lens), col = "darkred", lt = 2, lw = 2)
    abline(v = conf[[supp]], col = "darkred", lt = 1, lw = 3)
    axis(1, font = 2, at = conf[[supp]], labels = sprintf("%2.2f", conf[[supp]]))
    paste0("For supplement ", supp, ", the tooth growth is between ",
           sprintf("%2.2f", conf[[supp]][1]), " and ",
           sprintf("%2.2f", conf[[supp]][2]), " with 95% confidence.")
    })
```



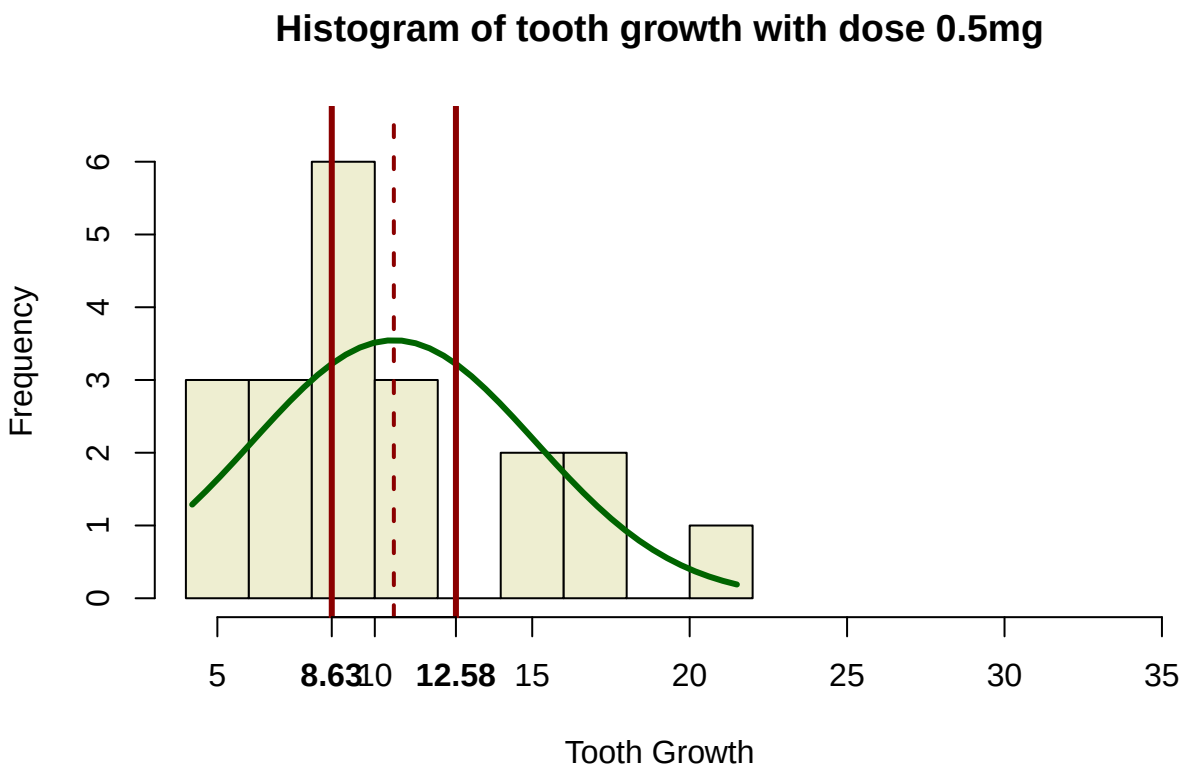

```
For supplement OJ, the tooth growth is between 18.30 and 23.03 with 95% confidence.
For supplement VC, the tooth growth is between 14.01 and 19.92 with 95% confidence.
```

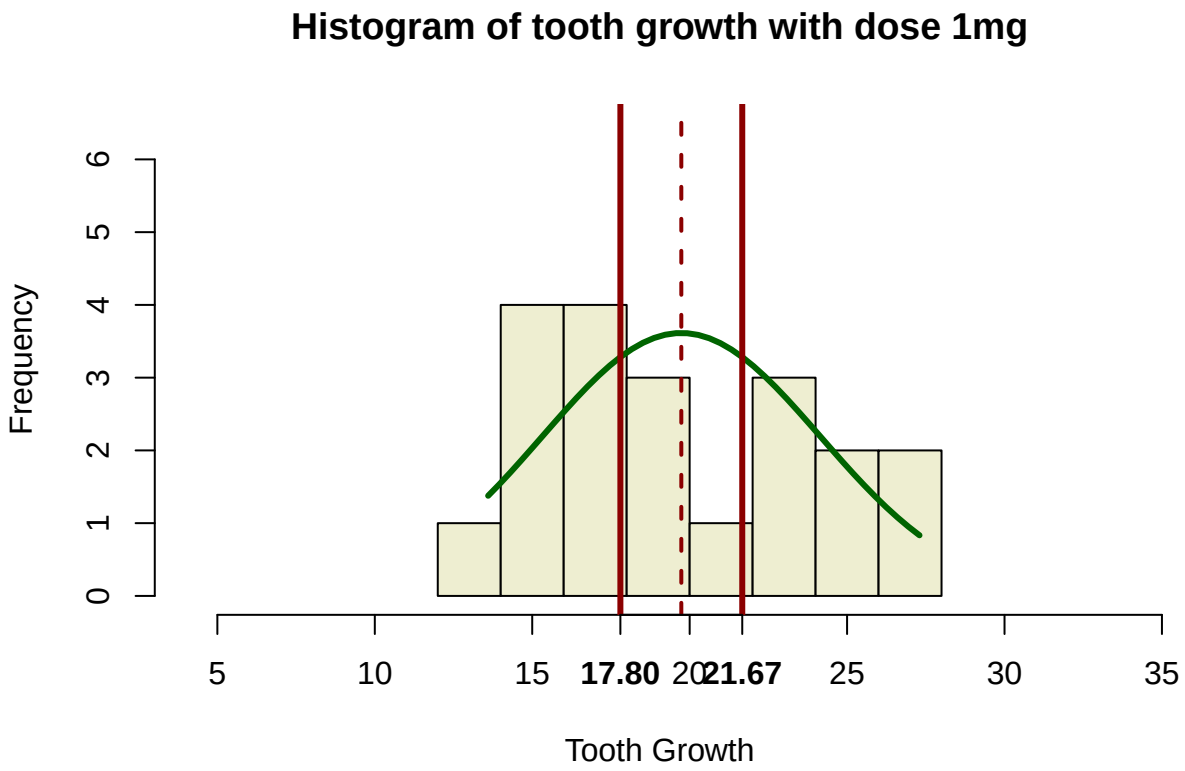
It can be clearly seen that orange juice (OJ) produces greater tooth growth compared to ascorbic acid (VC).

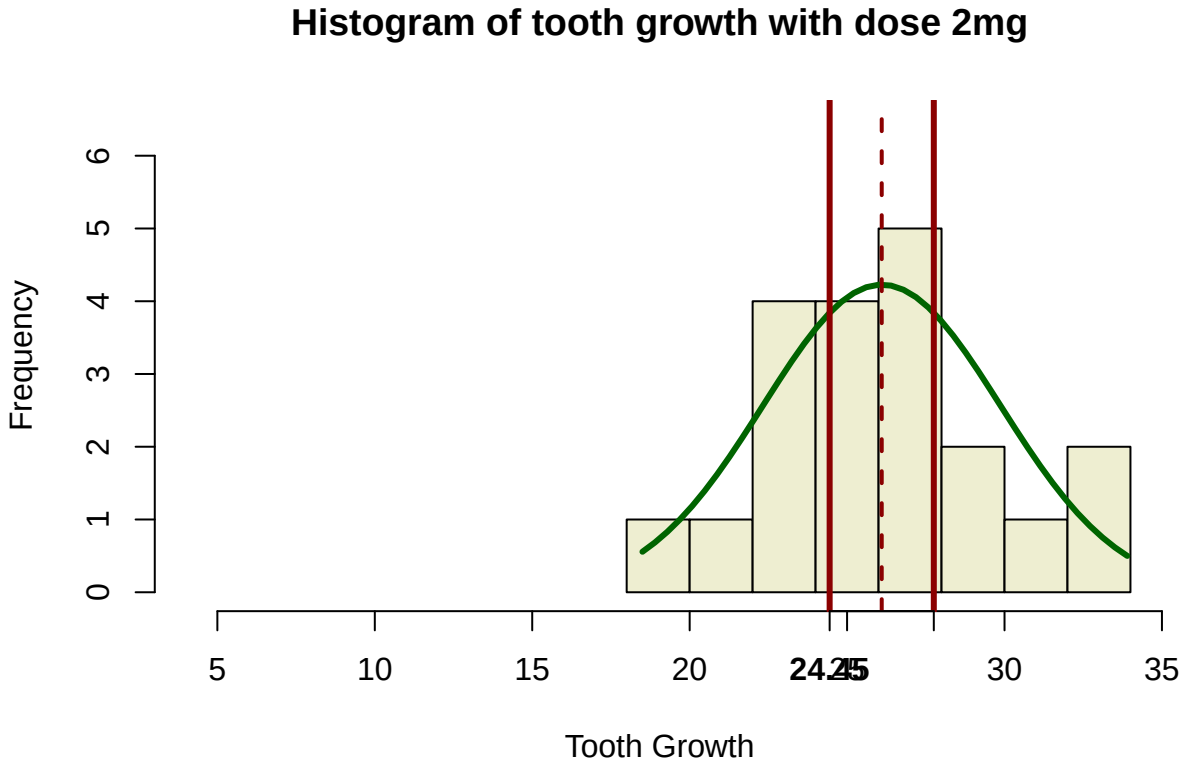
## Tooth growth analysis by dose

Assuming that the tooth growth data is *iid*, the 95% confidence intervals for tooth growth for each dose can be computed with the code below. The plots show the tooth growth for each dose as a histogram.

```
conf <- lapply(split(data, data$dose), function(dataForDose) {
    lens <- dataForDose$len
    mean(lens) + c(-1, 1) * qnorm(0.975) * sd(lens) / sqrt(length(lens))
    })
analysis <- lapply(names(conf), function(dose) {
    lens <- data[data$dose == dose, "len"]
    h <- hist(lens, breaks = 8, xlim = range(data$len), ylim = range(0, 6.5),
              main = paste0("Histogram of tooth growth with dose ", dose, "mg"),
              xlab = "Tooth Growth", ylab = "Frequency", col = "lightyellow2")
    xfit <- seq(min(lens), max(lens), length = 40)
    yfit <- dnorm(xfit, mean = mean(lens), sd = sd(lens))
    yfit <- yfit * diff(h$mids[1:2]) * length(lens)
    lines(xfit, yfit, col = "darkgreen", lt = 1, lw = 3)
    abline(v = mean(lens), col = "darkred", lt = 2, lw = 2)
    abline(v = conf[[dose]], col = "darkred", lt = 1, lw = 3)
    axis(1, font = 2, at = conf[[dose]], labels = sprintf("%2.2f", conf[[dose]]))
    paste0("For dose ", dose, "mg, the tooth growth is between ",
           sprintf("%2.2f", conf[[dose]][1]), " and ",
           sprintf("%2.2f", conf[[dose]][2]), " with 95% confidence.")
    })
```


Histogram of tooth growth with dose 0.5mg
Frequency
0
1
2
3
4
5
6
5
8.63
10
12.58
15
20
25
30
35
Tooth Growth


Histogram of tooth growth with dose 1mg
Frequency
0
1
2
3
4
5
6
5
10
15
17.80
20
21.67
25
30
35
Tooth Growth




For dose 0.5mg, the tooth growth is between 8.63 and 12.58 with 95% confidence.

For dose 1mg, the tooth growth is between 17.80 and 21.67 with 95% confidence.

For dose 2mg, the tooth growth is between 24.45 and 27.75 with 95% confidence.

It can be clearly seen that increasing the dose increases the tooth growth.