# ToothGrowth

Danny Fraser

October 2015

## Summary

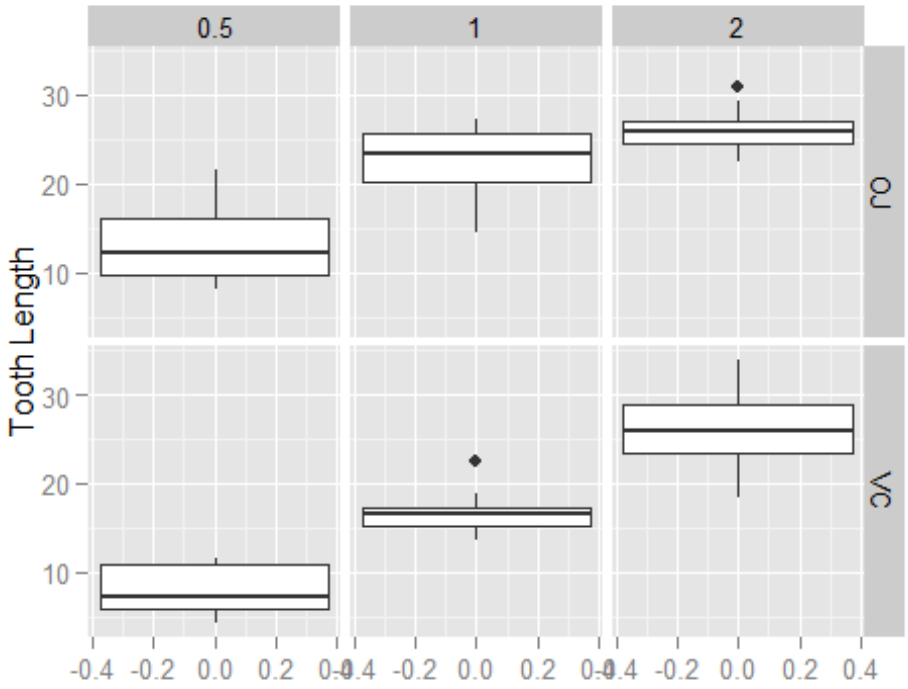
The `ToothGrowth` dataset in R is a small dataset showing the response of tooth length in guinea pigs at three dose levels (0.5, 1, and 2mg) of Vitamin C via 2 delivery methods (orange juice (OJ) or ascorbic acid (VC)).

This paper is a short statistical exploration of this dataset.

## Exploration

A quick exploratory plot of the data shows that the dosage is a more significant contributor to the change in length than the delivery method:

```
    qplot(data = ToothGrowth, facets = supp ~ dose,
        geom = "boxplot", x = 0, y = len,
        xlab = "", ylab = "Tooth Length")
```




And in a tabular form:

```
## Source: local data frame [6 x 4]
##
##     supp  dose mean(len) observations
##   (fctr) (dbl)     (dbl)        (int)
## 1     OJ   0.5     13.23           10
## 2     OJ   1.0     22.70           10
## 3     OJ   2.0     26.06           10
## 4     VC   0.5      7.98           10
## 5     VC   1.0     16.77           10
## 6     VC   2.0     26.14           10
```

In order to understand the impact that each factor has on the length of the teeth, we will need to apply some statistical tests to the data.

## Statistics

We are going to find the standard deviation of each group of results and establish with a 95% confidence interval which combination of dose and delivery method is the most effective.

The following table shows the 95% confidence interval upper and lower bounds for each pair of dose and delivery method. The R code is included for clarity (note that the `dplyr` library is used here).

```
    tgInterval <- ToothGrowth %>%
        group_by(supp, dose) %>%
        summarise(
            mean(len),
            lower=t.test(len, conf.level = .95)$conf.int[1],
            higher=t.test(len, conf.level=.95)$conf.int[2]
        )
    print(tbl_df(tgInterval))
```

```
## Source: local data frame [6 x 5]
##
##     supp  dose mean(len)     lower    higher
##   (fctr) (dbl)     (dbl)     (dbl)     (dbl)
## 1     OJ   0.5     13.23 10.039717 16.420283
## 2     OJ   1.0     22.70 19.902273 25.497727
## 3     OJ   2.0     26.06 24.160686 27.959314
## 4     VC   0.5      7.98  6.015176  9.944824
## 5     VC   1.0     16.77 14.970657 18.569343
## 6     VC   2.0     26.14 22.707910 29.572090
```

## Finding the Most Effective Treatment

It's clear from this table (and the exploratory boxplot) that the 2mg dose has the gretest effect. It's more difficult, though, to discern which of the delivery methods has the greatest effect at this dose.

First let's take the details of the t-test for both high dose scenarios:

```
    OJ_high <- filter(ToothGrowth, supp == "OJ", dose == 2)
    VC_high <- filter(ToothGrowth, supp == "VC", dose == 2)

    t_data_OJ_high <- t.test(OJ_high$len)
    t_data_VC_high <- t.test(VC_high$len)
```

Now let's take the p-values of each of these. The lower p-value will be the most effective treatment - i.e. it has the lowest probability of having occurred in the null hypothesis.

For the Orange Juice:

```
    pt(t_data_OJ_high$statistic, t_data_OJ_high$parameter, lower.tail =
FALSE)

##            t
## 9.166757e-11
```

And for the Ascorbic Acid:

```
    pt(t_data_VC_high$statistic, t_data_VC_high$parameter, lower.tail =
FALSE)

##            t
## 1.683983e-08
```

Both are extremely small vlaues, but this test shows that the Orange Juice delivery method is marginally more effective than the Ascorbic Acid.