Statistical Inference Project, Part 2: Tooth Growth

Overview

The goal of Part 2 of the project is to investigate tooth growth data for guinea pigs given different doses of vitamin c and orange juice.

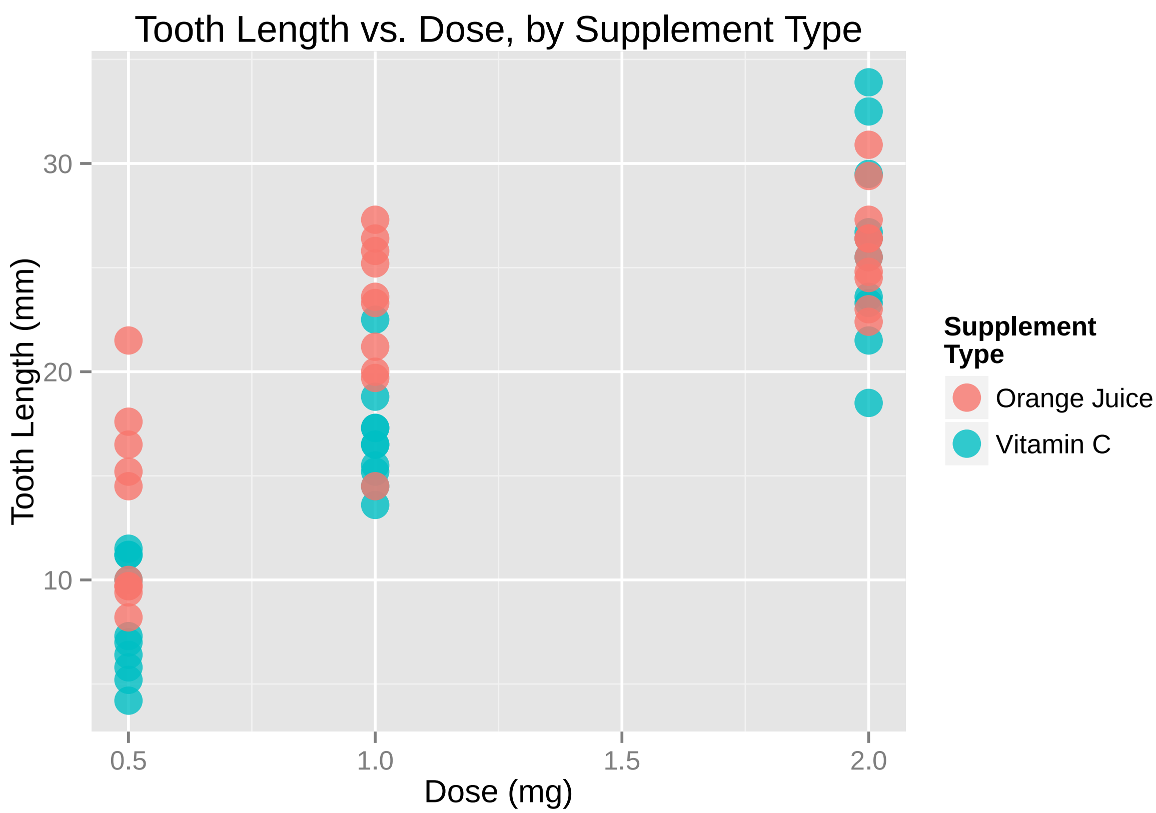
Assumptions

In order for this analysis to be meaningful, we must assume that the data is “roughly” normal. When the data is grouped by dose and supplement type, the distribution of the data is not clear. We will assume that the data is distributed sufficiently in order to use t-tests for hypothesis testing.

Exploratory Analysis

The data is easily visualized using a barplot:

The data are grouped by supplement type (orange juice and vitamin C), and the average tooth length of each group is plotted vs. dose given. We can also view all of the data rather than just the averages, which shows the same characteristics:



Based on these plots, it appears that orange juice is more effective than vitamin c for doses of 0.5 mg and 1.0 mg. For doses of 2.0 mg, orange juice appears to be equally effective as vitamin c. It is apparent that increasing dosage for either supplement leads to increased tooth length. We can also look at the mean and variance for each group, which will be useful for later testing:

```
## Source: local data frame [6 x 4]
## Groups: dose
##
##   dose supp avg.len       var
## 1  0.5   OJ   13.23 19.889000
## 2  0.5   VC    7.98  7.544000
## 3  1.0   OJ   22.70 15.295556
## 4  1.0   VC   16.77  6.326778
## 5  2.0   OJ   26.06  7.049333
## 6  2.0   VC   26.14 23.018222
```

Confidence Intervals and Hypothesis Testing

The tooth length data will be compared by supplement type at each dosage level. Since each group has 10 samples, we will use a t-test for testing the hypotheses. This requires data that appear to be "roughly" normal. Although it's difficult to tell how the data are distributed having so few samples, I will assume that the data meet the requirements to perform the t-tests.

The data are split into six groups according to the table in the previous section. Note that the data are not paired (each group is independent) and the variances between each group are not constant. The null hypothesis for these t-tests is $H_0 : \bar{x}_{OJ} = \bar{x}_{VC}$, with alternative hypothesis $H_\alpha : \bar{x}_{OJ} \neq \bar{x}_{VC}$. The t-tests are performed with the following code:

```
t.test(oj0.5,vc0.5)$conf
```

```
## [1] 1.719057 8.780943
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(oj1.0,vc1.0)$conf
```

```
## [1] 2.802148 9.057852
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(oj2.0,vc2.0)$conf
```

```
## [1] -3.79807  3.63807
## attr(,"conf.level")
## [1] 0.95
```

The confidence intervals for the first two tests (dosages of 0.5 mg and 1.0 mg) do not contain 0, and so we reject the null hypothesis and accept the alternative hypothesis: The groups given orange juice in doses of 0.5 mg and 1.0 mg have different tooth length than those given equal amounts of vitamin C. The mean tooth length of the groups given orange juice is greater than the groups given vitamin C for 0.5 mg and 1.0 mg doses.

The confidence interval of the final t-test does contain 0 and so we fail to reject the null hypothesis: there is no difference in tooth length between the groups given orange juice and vitamin C in doses of 2.0 mg. From these tests, we conclude that orange juice appears to be more effective than vitamin C for tooth growth in guinea pigs for doses of 0.5 mg and 1.0 mg. However, when given doses of 2.0 mg, orange juice and vitamin C appear to be equally effective for tooth growth. These conclusions assume that the distribution of tooth length of each group is “roughly” normal, allowing us to perform the t-tests. Also, we have assumed that the groups were not paired (as appears to be true from the data documentation).