Statistical Inference Final Project Part 2: ToothGrowth

By James Lamb

I. Introduction

In this brief report, I present a basic summary of the ToothGrowth dataset available in the `datasets` package in R. I use basic plotting, R summary functions, confidence intervals and hypothesis tests to describe some characteristics of the 60-observation dataset.

II. Exploratory Analysis

The code to load the data is:

```
library(datasets)
data(ToothGrowth)
```

Figure 1 in the Appendix shows some preliminary evidence that guinea pigs who received higher doses of Vitamin C had higher rates of tooth growth. There is no overlap in the distributions of tooth length by dose, providing strong evidence that the treatment does have a impact on tooth growth.

III. Basic Summary of the Data

Figure 2 in the Appendix gives a table of basic descriptive statistics for the ToothGrowth dataset. The mean guinea pig had tooth length of 18.81 mm (I assume this is mm. The help documentation does not say.). There were an equal numbers of 0.5, 1.0 and 2.0 mg doses administered. The standard deviation of tooth length is 7.65 mm.

The tooth length data do not appear to be normally distributed when looking at the histogram in Figugre 3. However, a Jarque-Bera test of tooth length suggests that the distribution of tooth length does roughly conform to normality.

IV. Using Hypothesis Tests and Confidence Intervals

In an attempt to compare the means of tooth length along two dimensions (supplement type and dosage), I created the following table of means and 95% confidence intervals.

```
##                 Mean Lower Limit (95%) Upper Limit (95%)
## Supp = OJ      20.66            18.300             23.03
## Supp = VC      16.96            14.005             19.92
## Dose = 0.5 mg  10.61             8.633             12.58
## Dose = 1.0 mg  19.73            17.800             21.67
## Dose = 2.0 mg  26.10            24.446             27.75
```

There is very little overlap in the 95% confidence interval for the means of the OJ vs. VC supplement groups, suggesting that their means are statistically different. Similarly, as can also be observed in Figure 1, the means of tooth length for the three different dosages are also statistically different from each other.

V. Necessary Assumptions

The confidence intervals above were constructed using the 95% two-sided (97.5% one-sided) quantile of the normal distribution. Therefore, the underlying assumption of this analysis is that the data are drawn from a normal distribution centered at the respective sample means and with variance approximated by the sample variances.

In order to interpret measures of differences in tooth *length* as representative or changes in tooth *growth*, it was also necessary to assume that pre-treatment tooth length for all subjects was roughly equal or that, at the very least, subjects were randomly placed into various treatment groups.

VI. Conclusion

This brief and informal analysis gives some preliminary evidence that the Vitamin C treatment did have a statistically significant impact on tooth growth. Subjects who received higher doses experienced greater tooth lengths.

I apologize to my readers for the brevity of this report. I spent more time than I expected to learning LaTeX for this and the Regression models report. Please grade kindly and offer constructive criticism in the space provided in the evaluation sheets.

The R markdown document with the underlying code used to generate this report can be found at this GitHub repo.

Appendix: Plots and Result Tables

Figure 1. Boxplot of Tooth Length by Dose

```
library(datasets)
data(ToothGrowth)
```


**Boxplot of Tooth Length by Dose**

Figure 2. ToothGrowth Descriptive Statistics

```
## Loading required package: boot
```

```
##                    len supp      dose
## nbr.val         60.0000   NA 60.00000
## nbr.null         0.0000   NA  0.00000
## nbr.na           0.0000   NA  0.00000
## min              4.2000   NA  0.50000
## max             33.9000   NA  2.00000
## range           29.7000   NA  1.50000
## sum           1128.8000   NA 70.00000
## median          19.2500   NA  1.00000
## mean            18.8133   NA  1.16667
## SE.mean          0.9875   NA  0.08119
## CI.mean.0.95     1.9760   NA  0.16245
## var             58.5120   NA  0.39548
## std.dev          7.6493   NA  0.62887
## coef.var         0.4066   NA  0.53903
```

Figure 3. Histogram of Tooth Length

Tooth Length
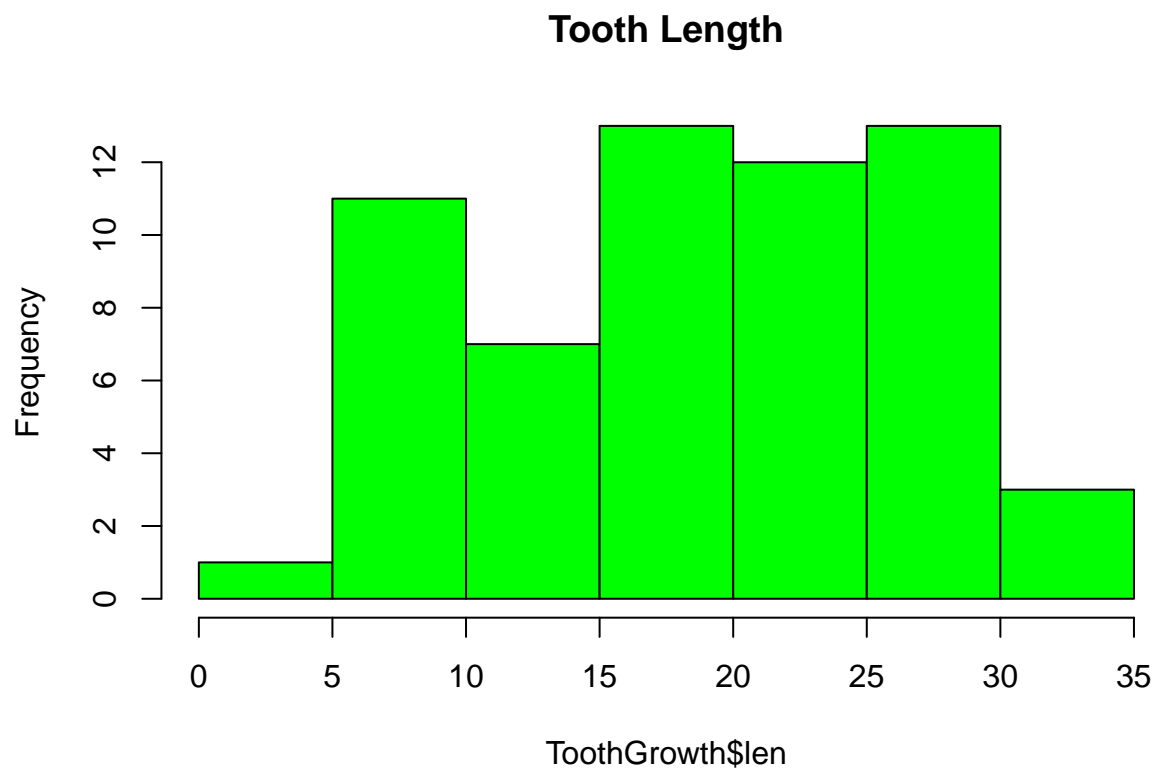
```
## [1] Jarque-Bera Test | JB Stat 2.59 p-value 0.27347

## [1] Excess Kurtosis = -1.04251 | Skewness = -0.14254
```

References

[1] Phillippe Crosjean et. al, "Pastecs: Package for Analysis of Space-Time Ecological Series", *CRAN*, 2014. Retrieved from: http://bit.ly/1pnSI2r