Basic Inferential Data Analysis

Daniel Hertenstein

January 4, 2017

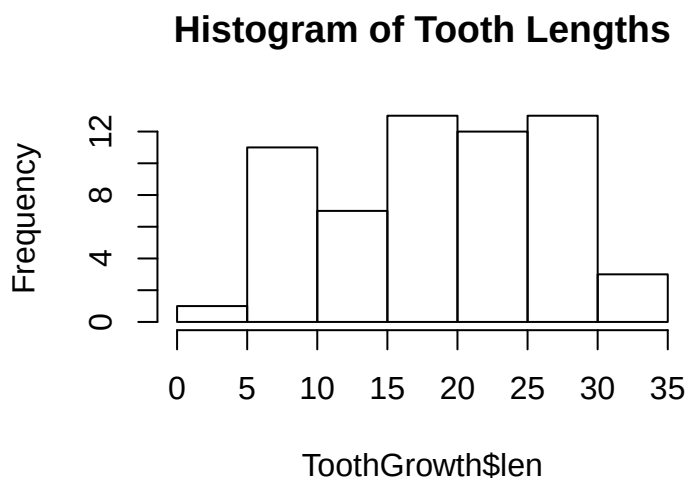
Loading the Data

The dataset that we are exploring is in the datasets library. To gain access to the dataset, we simply load the library.

```
library(datasets)
```

Exploratory Analysis

```
hist(ToothGrowth$len, main='Histogram of Tooth Lengths')
```




```
mean(ToothGrowth$supp == 'VC')
```

```
## [1] 0.5
```

```
table(ToothGrowth$dose)
```

```
##
## 0.5   1   2
##  20  20  20
```

With some simple exploratory analysis, we see: 1) the distribution of tooth lengths is fairly uniform, 2) half of the observations were for each supplement type, and 3) dosages were even divided between three amounts.

Data Summary

```
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
head(ToothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

This data frame contains sixty observations of three variables each (a tooth length, the supplement type, and a dose (mg/day))

Comparing Tooth Growth by Supp and Dose

We use two sided t-tests to compare the mean tooth length of groupings of observations grouped by supplement type and dosage. Since there are three dosage levels, we must perform three t-tests for dosage to test all combinations.

```
by_supp <- t.test(ToothGrowth$len ~ ToothGrowth$supp)
by_supp$p.value
```

```
## [1] 0.06063451
```

```
dose_0.5_1 <- t.test(ToothGrowth$len[ToothGrowth$dose == 0.5], ToothGrowth$len[ToothGrowth$dose == 1])
dose_0.5_1$p.value
```

```
## [1] 1.268301e-07
```

```
dose_0.5_1$conf.int
```

```
## [1] -11.983781  -6.276219
## attr(,"conf.level")
## [1] 0.95
```

```
dose_0.5_2 <- t.test(ToothGrowth$len[ToothGrowth$dose == 0.5], ToothGrowth$len[ToothGrowth$dose == 2])
dose_0.5_2$p.value
```

```
## [1] 4.397525e-14
```

```
dose_0.5_2$conf.int
```

```
## [1] -18.15617 -12.83383
## attr(,"conf.level")
## [1] 0.95
```

```
dose_1_2 <- t.test(ToothGrowth$len[ToothGrowth$dose == 1], ToothGrowth$len[ToothGrowth$dose == 2])
dose_1_2$p.value
```

```
## [1] 1.90643e-05
```

```
dose_1_2$conf.int
```

```
## [1] -8.996481 -3.733519
## attr(,"conf.level")
## [1] 0.95
```

Conclusions and Assumptions

We see that tooth length does not vary significantly between supplement type, but does vary significantly by dosage with higher dosage resulting in higher tooth lengths. For our t-tests, we assumed unequal variances between the subsets of the ToothGrowth data set, and that a signifiance level of 95% was acceptable acceptable.