Exploratory Analysis of the Tooth Growth dataset

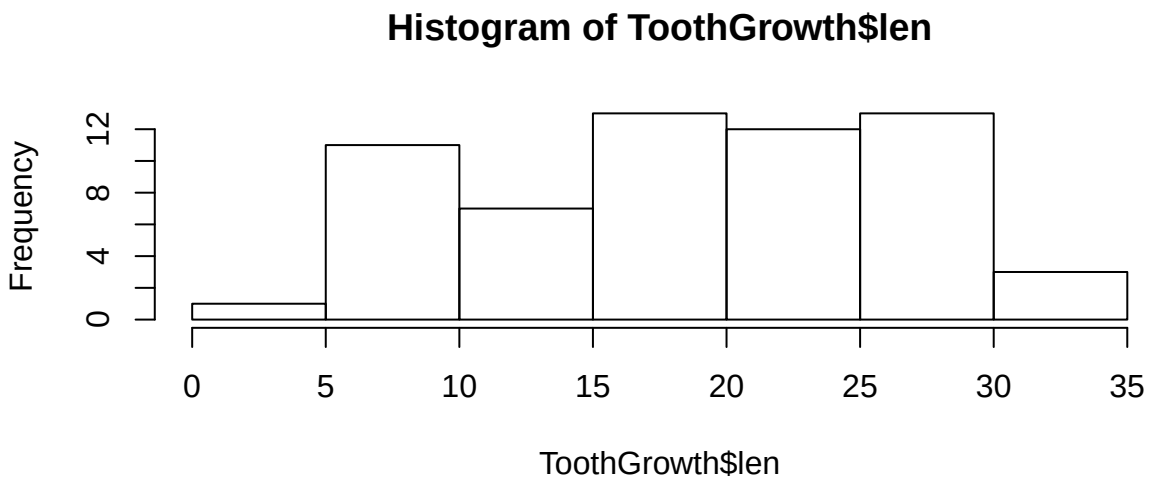
Andrew J. Dyck

Data Summary

The Tooth Growth dataset includes 60 observations of tooth growth with two additional categorical variables:

- Supplement Type (supp): Either VC or OJ
- Dose (dose): Dose level of Vitamin C

```
data(ToothGrowth)
hist(ToothGrowth$len)
```




```
aggregate(ToothGrowth$len, by=list(ToothGrowth$dose), FUN=mean)
```

```
##   Group.1      x
## 1     0.5 10.605
## 2     1.0 19.735
## 3     2.0 26.100
```

```
aggregate(ToothGrowth$len, by=list(ToothGrowth$supp), FUN=mean)
```

```
##   Group.1        x
## 1      OJ 20.66333
## 2      VC 16.96333
```

```
aggregate(ToothGrowth$len, by=list(ToothGrowth$supp, ToothGrowth$dose), FUN=mean)
```

```
##   Group.1 Group.2     x
## 1      OJ     0.5 13.23
## 2      VC     0.5  7.98
## 3      OJ     1.0 22.70
## 4      VC     1.0 16.77
## 5      OJ     2.0 26.06
## 6      VC     2.0 26.14
```

The overall distribution for tooth growth appears to be normal, with a fairly wide distribution around a mean of 18.81 and a variance of 58.51. The mean tooth growth appears to differ betwen supplement type and dose. Tests of these mean differences are explored in the following section.

Statistical Tests

Starting with a confidence interval of the overall mean below, we see that 95% of observations fall in the following range:

```
t.test(ToothGrowth$len)$conf.int[1:2]
```

```
## [1] 16.83731 20.78936
```

This is interesting as we can see from the previous section that the group mean for some combinations of Supplement method and dose are outside of this range.

Dose

For example, let's investigate 95% confidence intervals for the vitamin C dose (This is just exploratory at this point and I draw no conclusions):

```
t.test(ToothGrowth[which(ToothGrowth$dose==1.0), 'len'])$conf.int[1:2]
```

```
## [1] 17.66851 21.80149
```

```
t.test(ToothGrowth[which(ToothGrowth$dose==2.0), 'len'])$conf.int[1:2]
```

```
## [1] 24.33364 27.86636
```

The 95% confidence interval for doses of 0.5 is entirely below the confidence interval for the entire sample. Similarly, the 95% confidence interval for doses of 2.0 is entirely above the interval on the full sample. It appears that there may be something to the Vitamin C dose that affects tooth growth, but further analysis is required.

Again, this is exploratory at this point and no direct conclusions are drawn. So, let's try a two-sided T-test of dose to see if a doubling of the dose from 1.0 to 2.0 results in a significant change in tooth growth. The test here is that the difference in means is equal to zero (the null hypothesis). We specify that the alternate hypothesis is that the tooth growth is greater when dose is 2.0.

```
t.test(ToothGrowth[which(ToothGrowth$dose==1.0), 'len'],
       ToothGrowth[which(ToothGrowth$dose==2.0), 'len'],
       alternative = 'l')
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth[which(ToothGrowth$dose == 1), "len"] and ToothGrowth[which(ToothGrowth$dose == 2)
## t = -4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
```

```
##      -Inf -4.17387
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

It seems that we can concluded that average tooth growth is indeed higher with a dose of 2.0 compared to 1.0.

Supplement method

Next, let's consider a T-test to test the hypothesis that the mean tooth growth is the same between supplement methods:

```
t.test(ToothGrowth[which(ToothGrowth$supp=='OJ'), 'len'],
       ToothGrowth[which(ToothGrowth$supp=='VC'), 'len'])
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth[which(ToothGrowth$supp == "OJ"), "len"] and ToothGrowth[which(ToothGrowth$supp ==
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

With this sample size, we get a p-value of 0.06 for this T-test, so we cannot say that there is a statistical difference between the OJ and VC delivery methods at the $\alpha = 0.05$ level of confidence. It appears that the supplement method doesn't make a difference in tooth growth

Conclusion

In conclusion, it appears from the data collected in the ToothGrowth dataset, that tooth growth in Guinea Pigs is affected by the dose of Vitamin C and that the supplement method doesn't make much difference at all.

The assumptions that are required for this analysis are that the population data does indeed follow a normal distribution. If this assumption is not true, then the application of the T-test would be inappropriate for the analysis here. Additionally, I've assumed here that the sample size of 60 guinea pigs is sufficient to make the claims in the paragraph above. In fact, with a p-value on the supplement method of 0.06, it's likely that if the sample size were to be increased, I may be able to comment on the difference between supplement methods from a statistical perspective with more conviction.