Analysis of Tooth Growth

Barret Miller

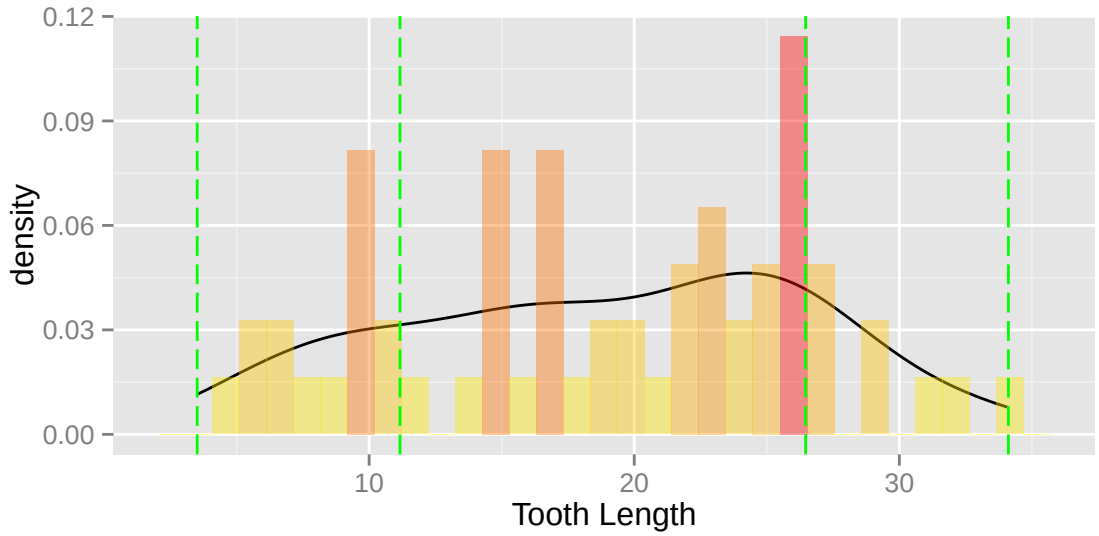
Importing and Exploring the Dataset

We'll take a look at the ToothGrowth dataset in R's datasets library. This dataset contains observations of the tooth length of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2mg) with each of two delivery methods (orange juice and ascorbic acid). Let's take a look at the overall skew of the data. If we want to use t-tests, then we need to make sure the data is roughly mound shaped and symmetrical about a central point. Based on this graph, we should be safe to use the t confidence intervals without worrying about taking a log or using other methods of analysis.

```
require(datasets); require(graphics); data(ToothGrowth); head(ToothGrowth)
```

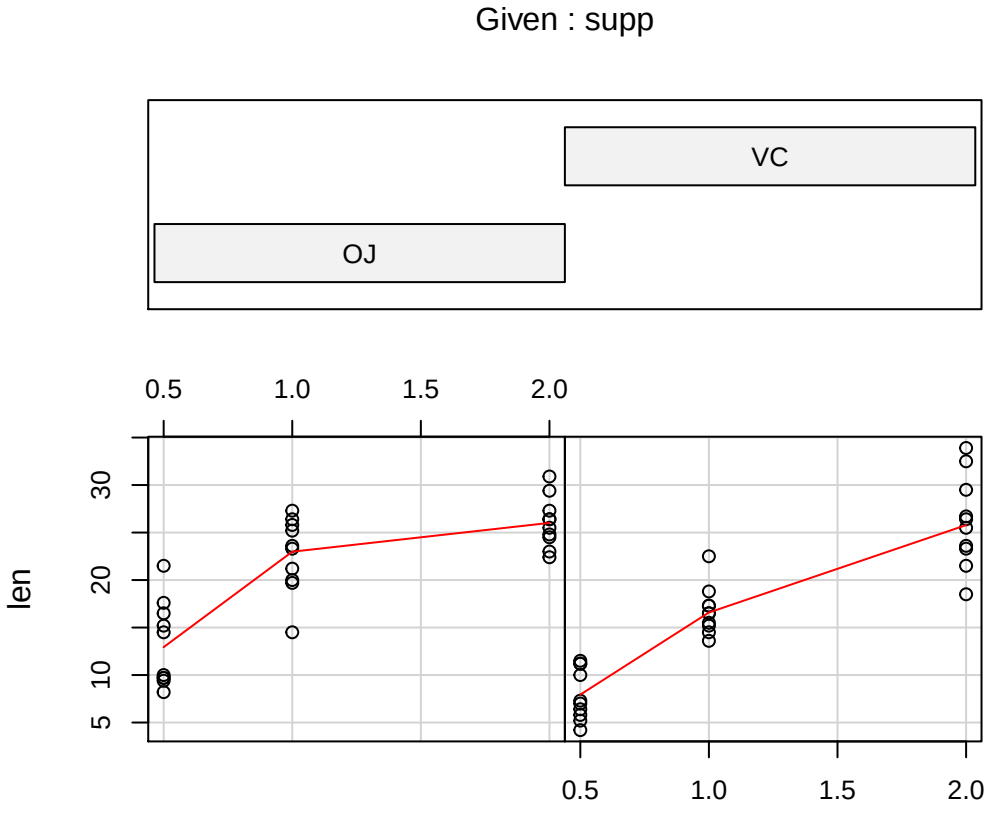
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
suppressMessages(require(ggplot2))
m0 <- mean(ToothGrowth$len)
s0 <- sd(ToothGrowth$len)
p0 <- qplot(ToothGrowth$len, geom = "blank", xlab="Tooth Length") +
   geom_line(aes(y = ..density..), stat = 'density') +
   geom_histogram(aes(y=..density.., fill = ..density..), alpha = 0.4) +
   scale_fill_gradient("Count", low = "yellow", high = "red", guide = F) +
   geom_vline(xintercept = m0+c(-1*s0, s0, -2*s0, 2*s0),
              colour="green", linetype = "longdash")
suppressMessages(print(p0))
```



We can also run the suggested plot in the documentation of the ToothGrowth dataset to get an idea of trends without doing any kind of statistical test.

```
require(graphics)
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,
       xlab = "ToothGrowth data: length vs dose, given type of supplement")
```



Confidence Intervals

In this section, we'll take a look at t confidence intervals. I'm going to acknowledge that the guinea pigs are the same 10 in each test and compare confidence intervals for differences between each of the different doses and suppliment delivery methods. First, I'll divide up the dataset into groups based on dose and delivery. Then, I'll use the t.test() function in R to compare each group, and we'll examine the 95% confidence intervals of expected differences beween the groups.

```
vcHalf<-ToothGrowth[1:10,1]; vcOne<-ToothGrowth[11:20,1]; vcTwo<-ToothGrowth[21:30,1]
ojHalf<-ToothGrowth[31:40,1]; ojOne<-ToothGrowth[41:50,1]; ojTwo<-ToothGrowth[51:60,1]
vc<-c(vcHalf, vcOne, vcTwo)
oj<-c(ojHalf, ojOne, ojTwo)
half<-c(vcHalf,ojHalf)
one<-c(vcOne,ojOne)
two<-c(vcTwo,ojTwo)

vcVSoj <- t.test(oj, vc, paired=T)
halfVsOne <- t.test(one, half, paired=T)
halfVsTwo <- t.test(two, half, paired=T)
```

```
oneVsTwo <- t.test(two, one, paired=T)

sprintf("OJ vs VC between %2.2f and %2.2f with a mean of %2.2f",
        vcVSoj$conf.int[1], vcVSoj$conf.int[2], vcVSoj$estimate)
```

```
## [1] "OJ vs VC between 1.41 and 5.99 with a mean of 3.70"
```

```
sprintf("Half vs One mg between %2.2f and %2.2f with a mean of %2.2f",
        halfVsOne$conf.int[1], halfVsOne$conf.int[2], halfVsOne$estimate)
```

```
## [1] "Half vs One mg between 6.39 and 11.87 with a mean of 9.13"
```

```
sprintf("One vs Two mg between %2.2f and %2.2f with a mean of %2.2f",
        oneVsTwo$conf.int[1], oneVsTwo$conf.int[2], oneVsTwo$estimate)
```

```
## [1] "One vs Two mg between 3.47 and 9.26 with a mean of 6.36"
```

Conclusions

I assume normality of the underlying population and that all samples in our dataset were independent and identically distributed. I also assume that the paired tests are comparing the same subject at two points in time, varying only the parameter we are testing at a time between the datapoints. With those assumptions, OJ is a better delivery mechanism for increasing tooth growth than ascorbic acid, although this makes the smallest difference with an average of 3.7 with a 95% confidence interval of between 1.41 and 5.99. The biggest difference can be made by increasing the dose. Going from half a milligram to 1 milligram increased the tooth growth by the largest margin, with an average of 9.13 and a 95% confidence interval of between 6.39 and 11.87. Increasing from 1mg to 2mg made a significant difference as well, although you start to see diminishing returns with an average increase of 6.36 and a 95% confidence interval of between 3.47 and 9.26.