ToothGrowth Dataset Analysis

Saul Lugo

October 25, 2015

```
require(knitr)
```

```
## Loading required package: knitr
```

```
## Warning: package 'knitr' was built under R version 3.1.3
```

```
require(datasets)
opts_chunk$set(echo=TRUE)
opts_chunk$set(cache=TRUE)
```

Loading the data

```
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
```

Exploratory Data Analysis

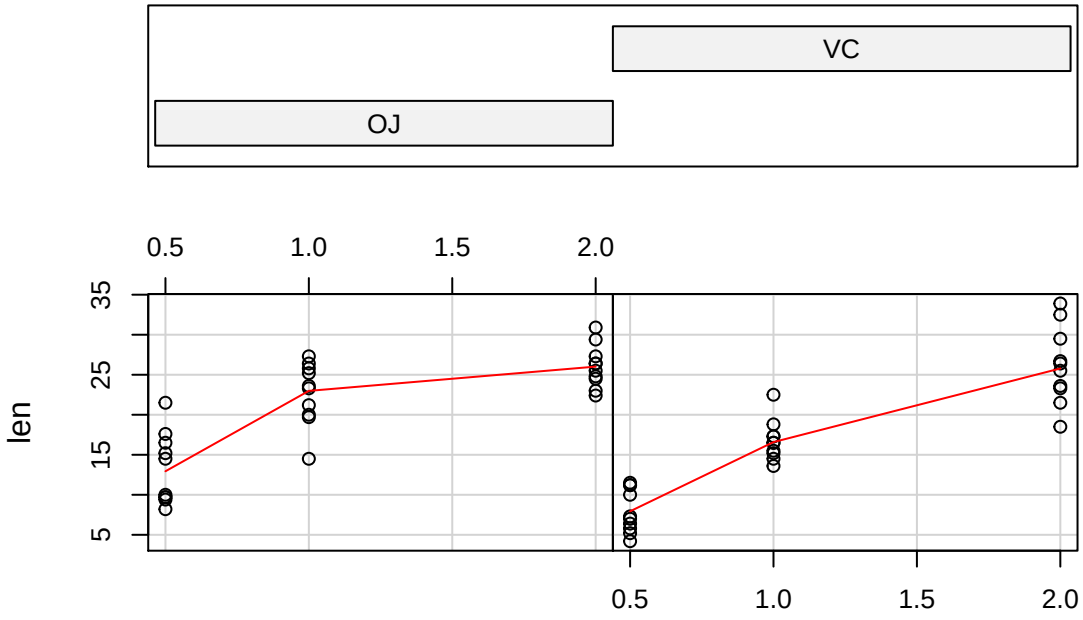
The description of the ToothGrowth dataset, as per the R documentation is the following:

- The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC).

So, given these data, we might be interested in study and compare tooth growth in the ginea pigs and its variation according to the dose levels of vitamin C and the administration method (orange juice or VC).

In order to visualize the data, let's make a plannel plot of it and the variables supp (administration method of the vitamin) and dose of the vitamin C:

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



ToothGrowth data: length vs dose, given type of supplement

The first thing we can realized from this graph is that the higher the dosse of vitamin (whichever the administration method) the higher the tooth growth is.

On the other hand, if we take the tooth growth as a random variable, it seems to be normaly distributed, since the values are equally spread around each sample mean. However it is difficult to conclude anything about how variable is the tooth growth because we have two variables, the dose and the supp that influence the result.

Summary of the Data

Let's explore the mean tooth len by vitamin dose:

```
mean(ToothGrowth$len[ToothGrowth$dose==2.0])
```

```
## [1] 26.1
```

```
mean(ToothGrowth$len[ToothGrowth$dose==1.0])
```

```
## [1] 19.735
```

```
mean(ToothGrowth$len[ToothGrowth$dose==0.5])
```

```
## [1] 10.605
```

So, without considering the supp variable, we can see that in average the tooth is larger when a dose of 2.0 mg/day has been applied, **so we can assume that the higher the dose the larger the tooth**.

Now, let's explore the mean of the tooth len by supplement type:

```
mean(ToothGrowth$len[ToothGrowth$supp=="OJ"])
```

```
## [1] 20.66333
```

```
mean(ToothGrowth$len[ToothGrowth$supp=="VC"])
```

```
## [1] 16.96333
```

We can see that the tooth are larger in average when the vitamin is administrated through Orange Juice (OJ).

Using Confidence Interval to Analyze Tooth Growth by Supp and Dose

From the Exploratory Data Analysis and the Summary of the Data we did before, it is pretty clear that the dose of vitamin C has a clear impact in the Ginea Pig Tooth Growth. With this data, we can be confidence that the higher the dose of vitamin C we administrate per day to the pigs the larger their teeth will grow.

However, we still have the issue of discover if administrating the vitamins through orange juice or through ascorbic acid (VC) have also an impact that could make a difference in the tooth growth of the pigs.

So, assuming that the higher dose will cause the teeth to grow larger, then it is just a matter of analyze if there is a clear difference in tooth growth when we use one supp or the other (OJ or VC). For this, let's calculate the **T Confidence Inverval** of the differences in tooth growth between the supps OJ and VC. We will use the following assumptions during the calculation:

- We **are not going** to assume that the test is paired, because in the documentation of the dataset says that it contains **"The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC)."**

So, apparently the 60 pigs are different and we could assume that were taken randomly from a general population.

- Also we are going to assume that the variance in tooth growth is not the same between the group treated with OJ and the group treated with VC. We are assuming this because, after plotting the variables as we did in the Exploratory Data Analysis, it is not clear how variable are the results.

Following these assumptions, we can make the following calculation of the **T Confidence Interval**:

```
# let's first build two vectors of tooth length subsetting the data by the supp variable
vc <- ToothGrowth$len[ToothGrowth$supp=="VC"]
oj <- ToothGrowth$len[ToothGrowth$supp=="OJ"]

# now, let's calculate the T CI using the assumptions explained before
t.test(vc,oj,paired=FALSE,var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  vc and oj
## 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:
##  -7.5710156  0.1710156
## sample estimates:
## mean of x mean of y
##  16.96333  20.66333
```

We got again (as we got in the data summary) that in average the tooth growth is larger when the OJ supp is used (mean of OJ tooth length vector is 20.66 vs 16.93 of the VC tooth length vector).

However, the **T Confidence Interval** is -7.57 to 0.17 with 95% confidence interval. As we can see 0 is within the interval, so **we cannot rule out that the mean of the differences in tooth length between the two supp could be 0**. It is true that most of the interval is negative, which means that it is highly probable that when using the OJ supp the tooth will growth larger than when using the VC supp, **but we cannot be 100% sure**.

Conclusions

From the Exploratory Data Analysis, looking a the plot of the tooth growth in the ginea pigs after administrate doses of vitamin C, we can conclude that the higher the per day dose the larger the teeth grow. Also this conclusion is supported by comparing the average tooth growth by doses of 0.5, 1.0 and 2.0 mg/day. In average the tooth growth is larger when the 2.0 mg/day dose was used.

The second issue that concern this study is if the vitamin supplement used, orange juice (OJ) or acid ascorbic (VC) has an impact in the tooth growth.

As we observed from the T Confidence Interval calculation, in average the difference in the tooth growth is mostly negative when we subtract the length vector of OJ from the VC vector. This means that in average the tooth growth is higher when using the OJ supp.

The t confidence interval is -7.57 to 0.17, so it is **mostly negative** which supports the hypotesis that the tooth growth is higher with the OJ supp, but **as 0 is inside the interval, we cannot rule out the possibility that the mean of the difference of tooth growth between both supps could be 0**.

So, we are highly incline to take that the OJ as a supp have a larger impact in tooth growth than the VC supp, but we **cannot** be 100% sure.