Executive Summary

We're going to analyze the ToothGrowth dataset. There are 3 variables (len, supp, dose) and 60 observations.

The description of the dataset from R hep file: The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

The description of the 3 variables are:

[,1] len numeric Tooth length

[,2] supp factor Supplement type (VC or OJ).

[,3] dose numeric Dose in milligrams.

Analysis

A basic summary of the data is below. Let's take a more in-depth look into this data.

```r
data(ToothGrowth)
suppressWarnings(suppressMessages(library(dplyr)))
agg<-group_by(ToothGrowth, supp, dose)
agg_sum<-summarise(agg, len_mean=mean(len), sd=sd(len), qty=n())
agg_sum
```

```
## Source: local data frame [6 x 5]
## Groups: supp
##
##   supp dose len_mean       sd qty
## 1   OJ  0.5    13.23 4.459709  10
## 2   OJ  1.0    22.70 3.910953  10
## 3   OJ  2.0    26.06 2.655058  10
## 4   VC  0.5     7.98 2.746634  10
## 5   VC  1.0    16.77 2.515309  10
## 6   VC  2.0    26.14 4.797731  10
```

We've just grouped the data by supplement type and dose and took the mean of the len of the tooth measured by these groups. We've also calculate the standard deviation for each group.

Doing a simple linear regression calculation, we get the following:

```r
summary(with(ToothGrowth, lm(len~ supp + dose)))
```

```
##
## Call:
## lm(formula = len ~ supp + dose)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.600 -3.700  0.373  2.116  8.800
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.2725     1.2824   7.231 1.31e-09 ***
## suppVC       -3.7000     1.0936  -3.383   0.0013 **
```

```
## dose          9.7636     0.8768  11.135 6.31e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.236 on 57 degrees of freedom
## Multiple R-squared:  0.7038, Adjusted R-squared:  0.6934
## F-statistic: 67.72 on 2 and 57 DF,  p-value: 8.716e-16
```

We can see the length of tooth measured are positively influenced by supplement type OJ (orange juice, the intercept) and dose, and negatively influenced by supplement type VC (ascorbic acid) when compared to the intercept (which is orange juice).

Let's take the following null hypothesis: - Dose doesn't influence the growth of teeth.

Since we have a low number of observations for each case (about 10), we're going to adopt a T-statistic approach instead of Z-statistic.

```r
t.test(len ~ dose, data=filter(ToothGrowth, dose<=1.0))$conf.int
```

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

```r
t.test(len ~ dose, data=filter(ToothGrowth, dose>=1.0))$conf.int
```

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

As we can see, the 95% confidence interval for the mean between doses 0.5mg and 1.0mg are between -11.983 and -6.276. Similarly, the 95% confidence interval for the mean between doses 1.0mg and 2.0mg are between -8.996 and -3.733. We can conclude, since we don't have zero included in the 95% confidence interval we can reject the null hypothesis, and state that dose influences the growth of the teeth.

With regards to the supplement variable, let's assume the following null hypothesis: - Supplement type does not influence the growth of the teeth.

```r
t.test(len ~ supp, data=ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by 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 in group OJ mean in group VC
##         20.66333         16.96333
```

In fact, since the 95% confidence interval includes 0 (-0.171, 7.571), and we have a p-value about 0.06, we cannot reject the null hypothesis. Let's investigate deeper and see why is this.

Trying to hold the variable dose constant:

```r
t.test(len ~ supp, dose==0.5, data=ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```r
t.test(len ~ supp, dose==1.0, data=ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```r
t.test(len ~ supp, dose==2.0, data=ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.0461, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

We can realize that for 0.5mg and 1.0mg, the 95% confidence intervals do not include zero and p-values are very low (below 0.007, 0.7%). So for these cases, we reject the null hypothesis. This is not the case when dose is equal to 2.0mg. In this case, the supplement type doesnt make any difference. The means are similar, p-value is about 10% (a high value). So, for dose equal 2.0mg we cannot reject the null hypothesis.

Conclusion

We can conclude that for lower doses (0.5mg and 1.0mg), the supplement type makes difference and influences teeth growth. Similarly, the dose itself is highly relevant and highly correlated with teeth growth as well. But for higher doses, the supplement type doesnt make much difference and doesnt influence the growth of the teeth of the pigs, so we only account for the dose itself (which remains relevant). Due to the low number of observations for each case, we assumed a T distribution instead of a normal one (Z).