# Statinf Project Part 2 – ToothGrowth

Dataset

The given dataframe contains 60 observations on the effect of vitamin C on the tooth growth in guinea pigs. The first column gives the length of the tooth. The second column states the delivery method to the vitamin C: Either via orange juice (OJ) or via ascorbic acid (VC). The third and last columns contains the dose for every observation.

Exploratory Analysis

A basic summary of the data shows that there are 30 observations for each delivery method. Furthermore len seems to be a continuous variable, while dose is discrete and takes the values 0.5, 1, and 2. It later turns out that there are exactly 20 observations for each dose.

```
> summary(ToothGrowth)
      len          supp         dose
 Min.   : 4.20   OJ:30   Min.   :0.500
 1st Qu.:13.07   VC:30   1st Qu.:0.500
 Median :19.25           Median :1.000
 Mean   :18.81           Mean   :1.167
 3rd Qu.:25.27           3rd Qu.:2.000
 Max.   :33.90           Max.   :2.000
```

Two interesting questions arise from this data: Does the amount of vitamin C given to the guinea pig (the dose) affect tooth growth (len)? If that is the case, does the delivery method have an impact on the effectiveness? Plotting len against dose for the whole dataset shows that both questions are worth looking into:
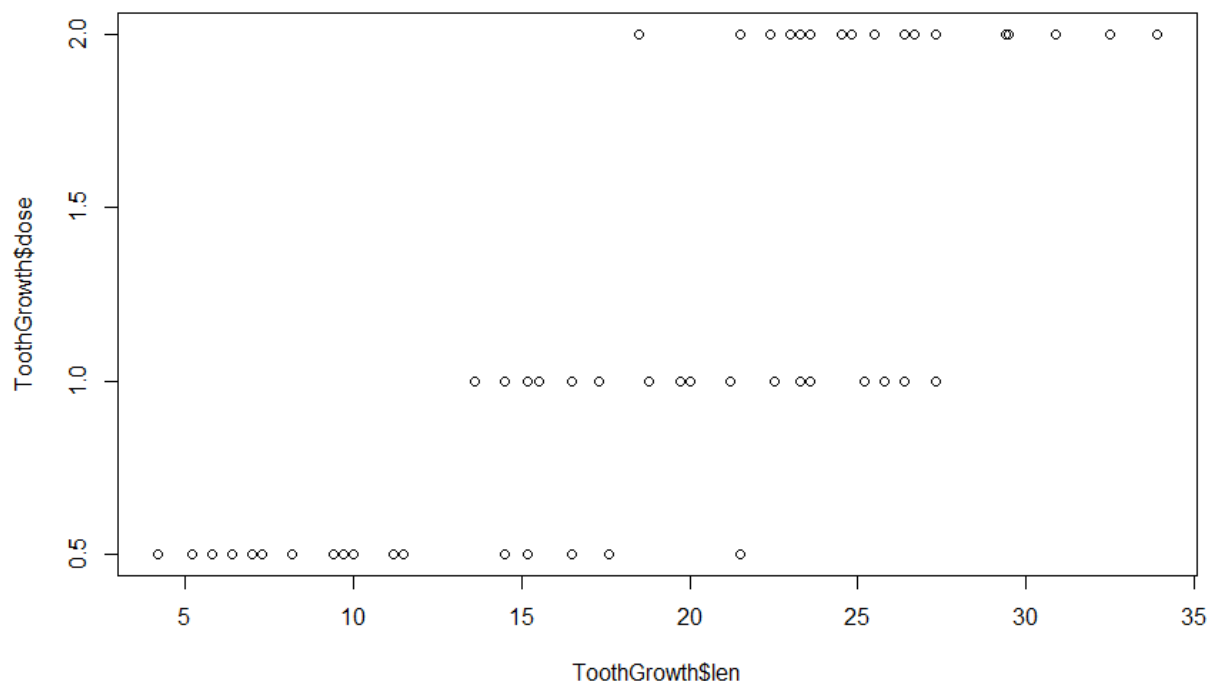
Statistical Tests

For the following tests I assume that the guinea pigs for each observation were chosen at random and that the population variance for any two groups of data in the following tests is the same.

Tests for different mean len by dosage:

For these test I subset the data in three ways:

```
> tg0.5_1 <- subset(ToothGrowth, dose %in% c(0.5, 1))
> tg0.5_2 <- subset(ToothGrowth, dose %in% c(0.5, 2))
> tg1_2 <- subset(ToothGrowth, dose %in% c(1, 2))
```

```
> t.test(len ~ dose, paired=F, var.equal=T, data=tg0.5_1)

        Two Sample t-test
data:  len by dose
t = -6.4766, df = 38, p-value = 1.266e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.983748  -6.276252
sample estimates:
mean in group 0.5    mean in group 1
         10.605             19.735
```

This first test shows that the mean for a dosage of 0.5 is clearly lower than the mean for a dosage of 1. The 95% confidence interval is entirely negative and the p-value is very low.

```
> t.test(len ~ dose, paired=F, var.equal=T, data=tg0.5_2)

        Two Sample t-test
data:  len by dose
t = -11.799, df = 38, p-value = 2.838e-14
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -18.15352 -12.83648
sample estimates:
mean in group 0.5    mean in group 2
         10.605             26.100
```

The result of the second test also shows a clearly higher mean for a dosage of 2 than for 0.5. This is again supported by an entirely negative 95% confidence interval and an extremely low p-value.

```
> t.test(len ~ dose, paired=F, var.equal=T, data=tg1_2)

        Two Sample t-test
data:  len by dose
t = -4.9005, df = 38, p-value = 1.811e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -8.994387 -3.735613
sample estimates:
mean in group 1 mean in group 2
         19.735             26.100
```

The third test similarly shows that the mean teeth length for a dosage of 2 is higher than for a dosage of 1. However the difference is not as big as in the first test: The two means are closer together and the confidence interval is not as negative as for 0.5 vs. 1. The p-value is still very small.

Test for different mean len by delivery method:

The data is again split into subsets:

```
> tg0.5 <- subset(ToothGrowth, dose == 0.5)
> tg1 <- subset(ToothGrowth, dose == 1.0)
> tg2 <- subset(ToothGrowth, dose == 2.0)
```

```
> t.test(len ~ supp, paired=F, var.equal=T, data=tg0.5)

        Two Sample t-test
data:  len by supp
t = 3.1697, df = 18, p-value = 0.005304
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.770262 8.729738
sample estimates:
mean in group OJ mean in group VC
           13.23            7.98
```

The first test finds a difference in mean len for a dosage of 0.5. Orange juice seems to result in significantly longer teeth than ascorbic acid. This difference is supported by an entirely positive 95% confidence interval and a low p-value.

```
> t.test(len ~ supp, paired=F, var.equal=T, data=tg1)

        Two Sample t-test
data:  len by supp
t = 4.0328, df = 18, p-value = 0.0007807
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.840692 9.019308
sample estimates:
mean in group OJ mean in group VC
           22.70           16.77
```

The second test paints a similar picture for a dosage of 1. Again orange juice has a higher mean len. This is supported by an entirely positive 95% confidence interval and a very low p-value.

```
> t.test(len ~ supp, paired=F, var.equal=T, data=tg2)

        Two Sample t-test
data:  len by supp
t = -0.0461, df = 18, p-value = 0.9637
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.722999  3.562999
sample estimates:
mean in group OJ mean in group VC
           26.06           26.14
```

The difference in means for the two delivery methods is not evident in the third test anymore, which was performed on the data for a dosage of 2. The two means are now nearly identical, the 95% confidence interval neatly centers around 0, and the p-value is nearly 1.

Conclusions

The first set of tests shows that in general a higher dosage of vitamin C, regardless of delivery method, leads to longer teeth in the guinea pigs. However this effect seems to be "wearing off": The difference between a dosage of 1 and 2 is not as big as the difference between 0.5 and 1.

In the second set of tests it was shown that orange juice seems to be the more effective delivery method for dosages of 0.5 and 1. The mean teeth length for dosages that high delivered via orange juice were significantly higher than the mean teeth length for those pigs that received their vitamin C via ascorbic acid. No difference was detected for a dosage of 2 however.