Statistical Inference Project - Part 2

Andrej Schoeke

2015-Oct- 24

Abstract

This analysis tries to evaluate the Tooth Growth Data set ##link. The main question is whether the dose as well as the delivery method of Vitamine C affects differnce in teeth growth in guniea pigs.

Exploration of the data set

First, let's gain a basic understanding of what our data set looks like:

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
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
```

```
table(ToothGrowth$supp, ToothGrowth$dose)
```

```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

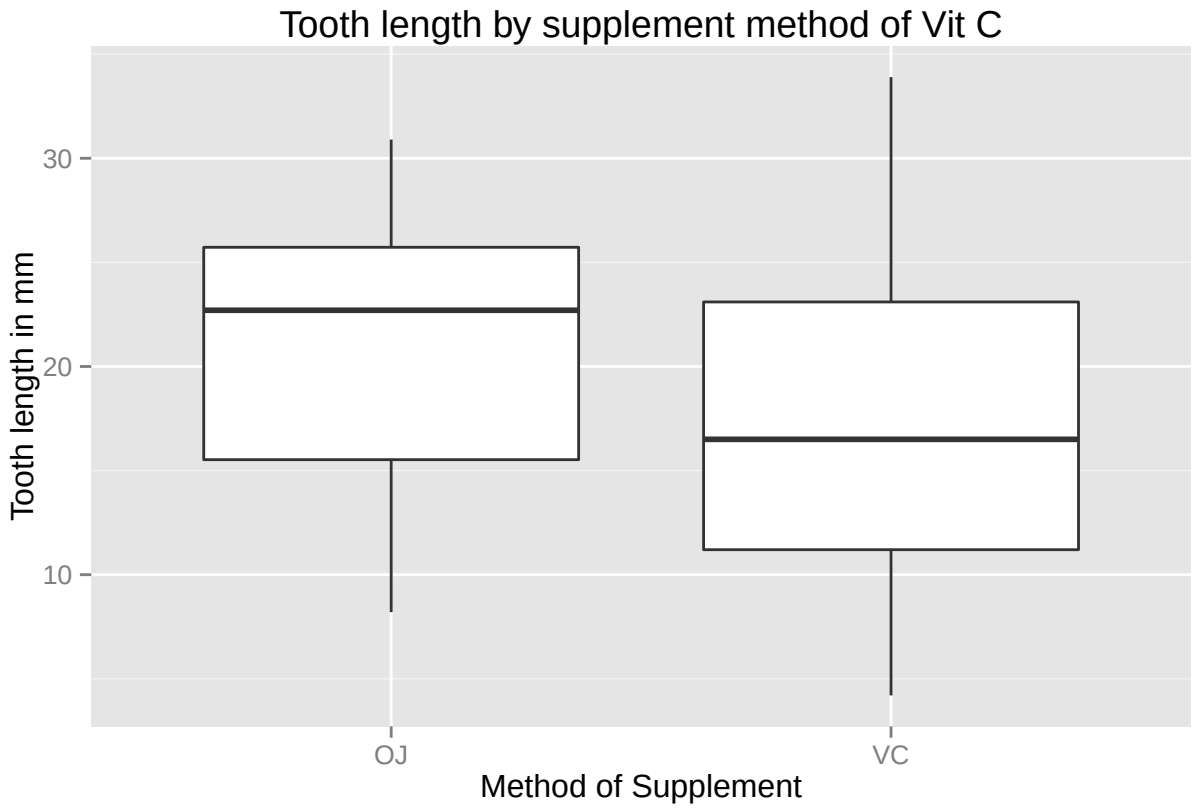
```
ToothGrowth %>% group_by(supp) %>% summarise(avg = mean(len), sd = sd(len))
```

```
## Source: local data frame [2 x 3]
##
##   supp      avg       sd
## 1   OJ 20.66333 6.605561
## 2   VC 16.96333 8.266029
```

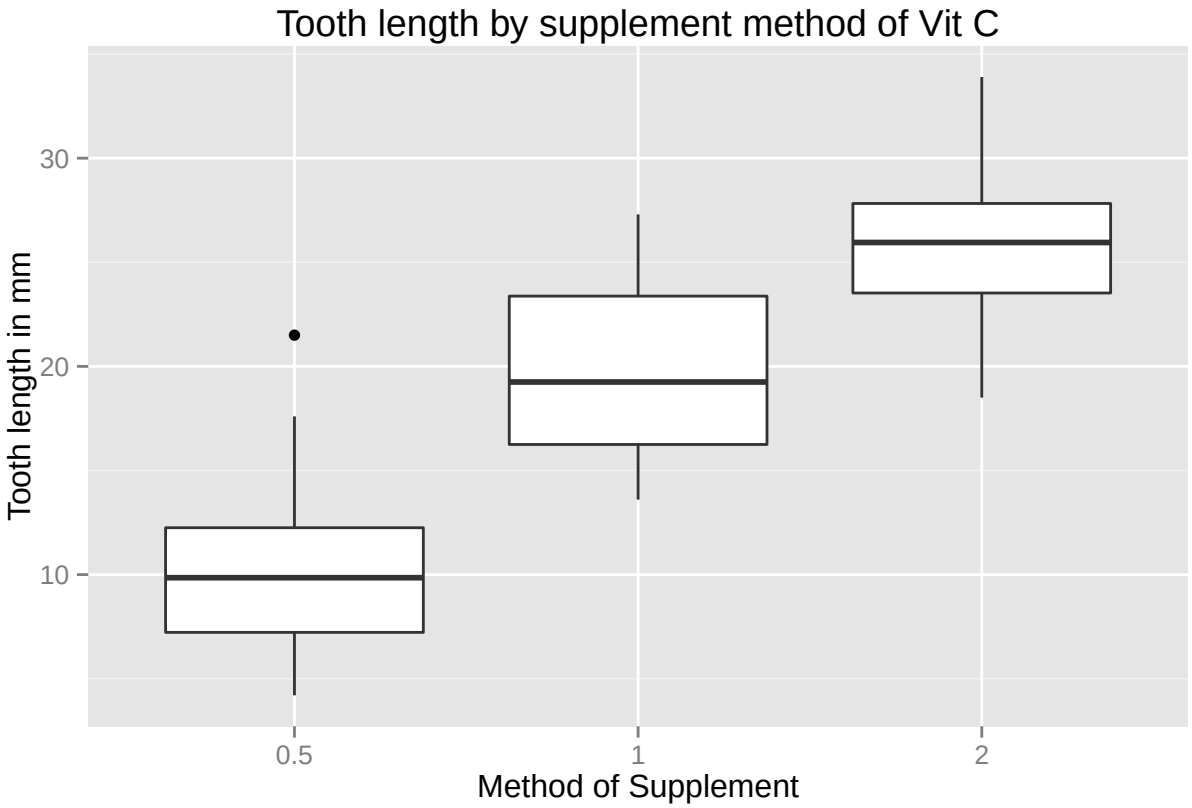
```
ToothGrowth %>% group_by(dose) %>% summarise(avg = mean(len), sd = sd(len))
```

```
## Source: local data frame [3 x 3]
##
##   dose    avg       sd
## 1  0.5 10.605 4.499763
## 2  1.0 19.735 4.415436
## 3  2.0 26.100 3.774150
```

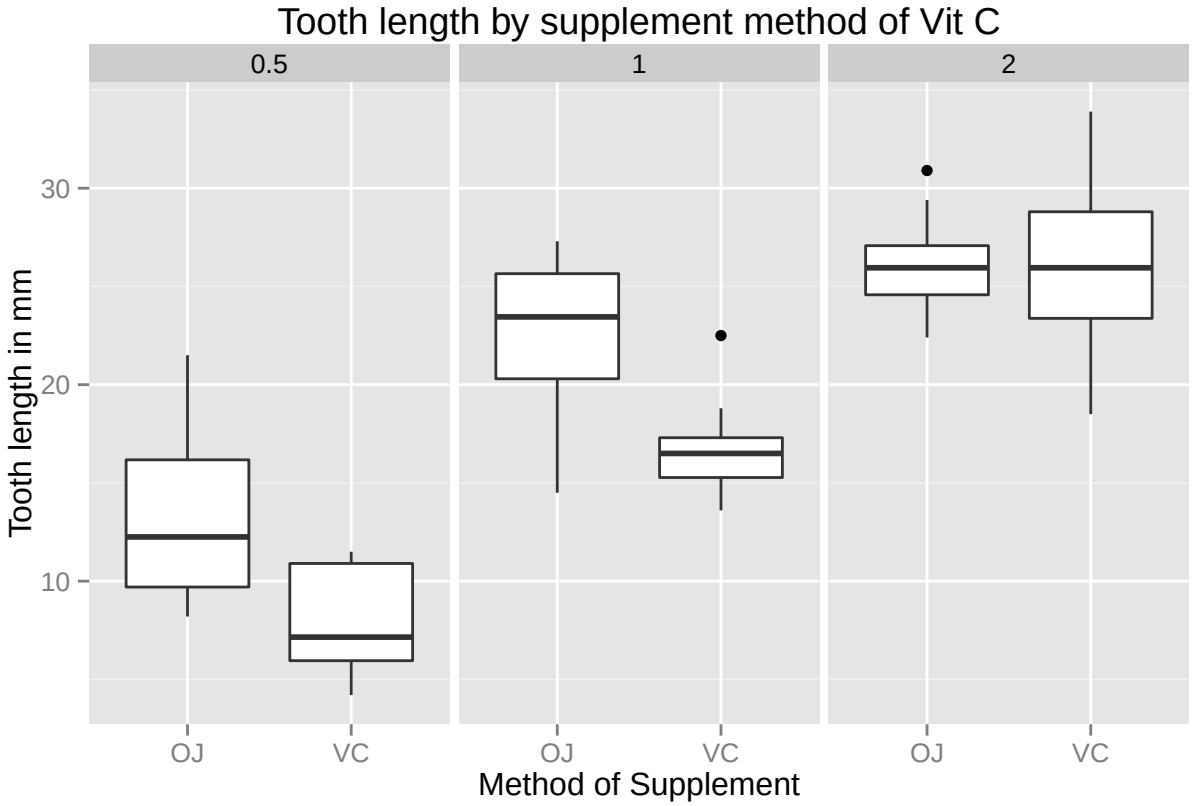
```
ggplot(ToothGrowth,aes(supp, len)) + geom_boxplot() +
    labs(title ="Tooth length by supplement method of Vit C",
         x = "Method of Supplement", y = "Tooth length in mm")
```




```
ggplot(ToothGrowth,aes(factor(dose), len)) + geom_boxplot() +
    labs(title ="Tooth length by supplement method of Vit C",
         x = "Method of Supplement", y = "Tooth length in mm")
```




```
ggplot(ToothGrowth,aes(supp, len)) + geom_boxplot() + facet_grid(. ~ dose) +
    labs(title ="Tooth length by supplement method of Vit C",
         x = "Method of Supplement", y = "Tooth length in mm")
```



We have clear an straight data set with 3 variables: len for the length of the teeth, supp for the method of supplementation and the dose being either 0.5, 1.0, or 2.0. For each dose-by-supp group we have 10 observations. On first sight, there seems not to be a a huge difference between the two supplemental methods. However, the variance is rather high in the Vitamin C group.

If we compare the dosage groups, it seems that more is generally better in this case. However, if we plot the length by dosage and supplemental method, we see that the effect could be driven by the lower dosages in the Vitamin C group.

Analysis

We will be using t-tests to assess our data set.

Assumptions

The t-test has a couple of assumptions which we will state but not test here. If all or part of them are violated, the tests might gives us wrong results.

Assumptions of the t-test: * The underlying population sidtribution follows the normal distribution * The data must be continous * The two samples are randoms sample * The variance of the two populations is equal * The two samples are independent

Assumptions of the one way Analysis of Variance: * The observations have to be independent * The residuals are normally distributed * Variances are euqually distributed, homogenic.

Tests

Effect of supplement method

Our null hypothesis is that there is no difference between the supplement methods of pure Vitamine C and Orange juice.

```
# T-Test with DF adjustment of Welch for unequal variances.
 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
```

According to our test, there is a small but not quite significant difference between the two methods od supplying Vitamine C. With p = 0.06, we can't reliable make any statement about it.

Effect of dosage

The t-test does not lend itself to investigate the effect of mutlilevel factors. For this reason, we use the oneway ANOVA.

```
oneway.test(len ~ dose, data = ToothGrowth)
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  len and dose
## F = 68.401, num df = 2.000, denom df = 37.743, p-value = 2.812e-13
```

```
pairwise.t.test(ToothGrowth$len, ToothGrowth$dose, p.adj= "none")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  ToothGrowth$len and ToothGrowth$dose
##
##   0.5     1
## 1 6.7e-09 -
## 2 < 2e-16 1.4e-05
##
## P value adjustment method: none
```

The results indicates that dosage by itself makes a significant difference as $p > 0.05$. Furthermore, we can say, that more is better by using pair-wise t-tests.

Conclusion

Our analyses showed, that the dosage made a significant difference in the development of the guinea pigs' teeth length. On the other hand, we could not find a significant difference in the way, the Vitamine C was provided. Albeit, the p-value is low and could indicate sample size that is just a bit to small and therefore prone to outliers.