Tooth Growth Analysis

Rachit Kinger

19 August 2017

Overview

Objective

This report is about the analysis of the impact on tooth growth based on the dosage and delivery method of vitamin C to the subject.

Methodology

The report does a t-test to determine whether there is a difference in tooth growth based on the following different groupings:

- dosage amount only
- delivery method only
- dosage and delivery

Results

Based on the t-test using a 99% confidence level the analysis concludes the following:

1. **Dosage**: As dosage of vitamin C increases tooth growth increases
2. **Delivery Method**: There is no difference in tooth growth based on delivery method used
3. **Dosage & Delivery**: For smaller dosage amounts (`0.5 mg` and `1.0 mg`) there is a difference in tooth growth based on delivery method used, but for higher amount (`2.0 mg`) there is no difference in tooth growth based on delivery method used

Detailed steps in the methodology

We use the `ToothGrowth` dataset found in R. The dataset has 60 observations across 3 variables:

1.`len`: the response variable, i.e. length of the tooth
2. `supp`: the delivery method used, `OJ` for orange juice, `VC` for ascorbic acid
3. `dose`: the dosage amount administered in miligrams per day. Three levels - `0.5`, `1.0` and `2.0`.

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

Exploratory data analysis

First we explore the distribution of tooth lengths on its own.

```r
stem(tooth$len)
```

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   0 | 4
##   0 | 5667789
##   1 | 00001124
##   1 | 55555677777899
##   2 | 001222333344
##   2 | 55566666667779
##   3 | 0134
```

```r
summary(tooth$len)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.20   13.07   19.25   18.81   25.27   33.90
```

We now inspect tooth length based on factors of delivery method (supp) and dosage.

```r
avg <- with(tooth, tapply(len, supp, mean))
avgSupp <- data.frame(supp = names(avg), val = avg)
rownames(avgSupp) <- NULL
avgSupp
```

```
##   supp      val
## 1   OJ 20.66333
## 2   VC 16.96333
```

We can see that the mean tooth length is higher for OJ than for VC, but we will run a t-test later to establish how confident we can be of this difference.

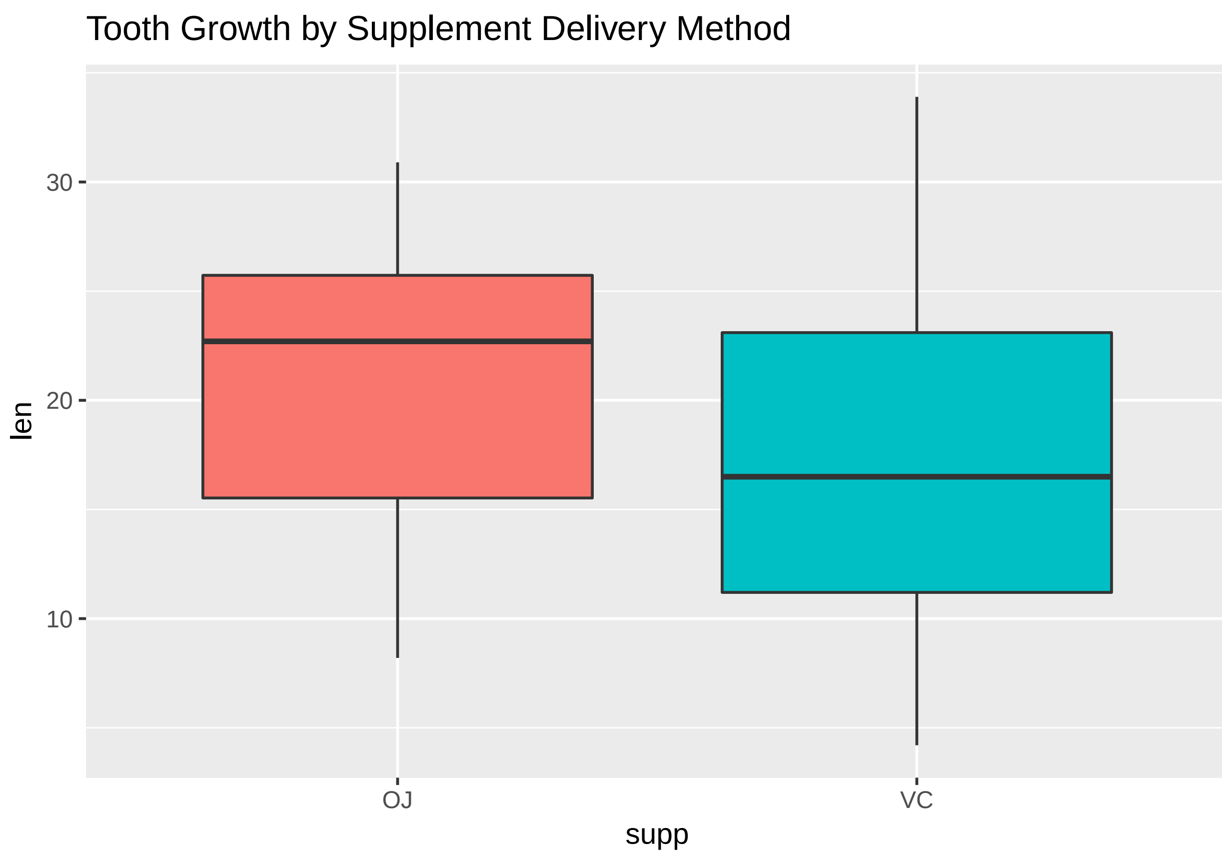
```r
avg <- with(tooth, tapply(len, dose, mean))
avgDose <- data.frame(dose = names(avg), val = avg)
rownames(avgDose) <- NULL
avgDose
```

```
##   dose    val
## 1  0.5 10.605
## 2    1 19.735
## 3    2 26.100
```

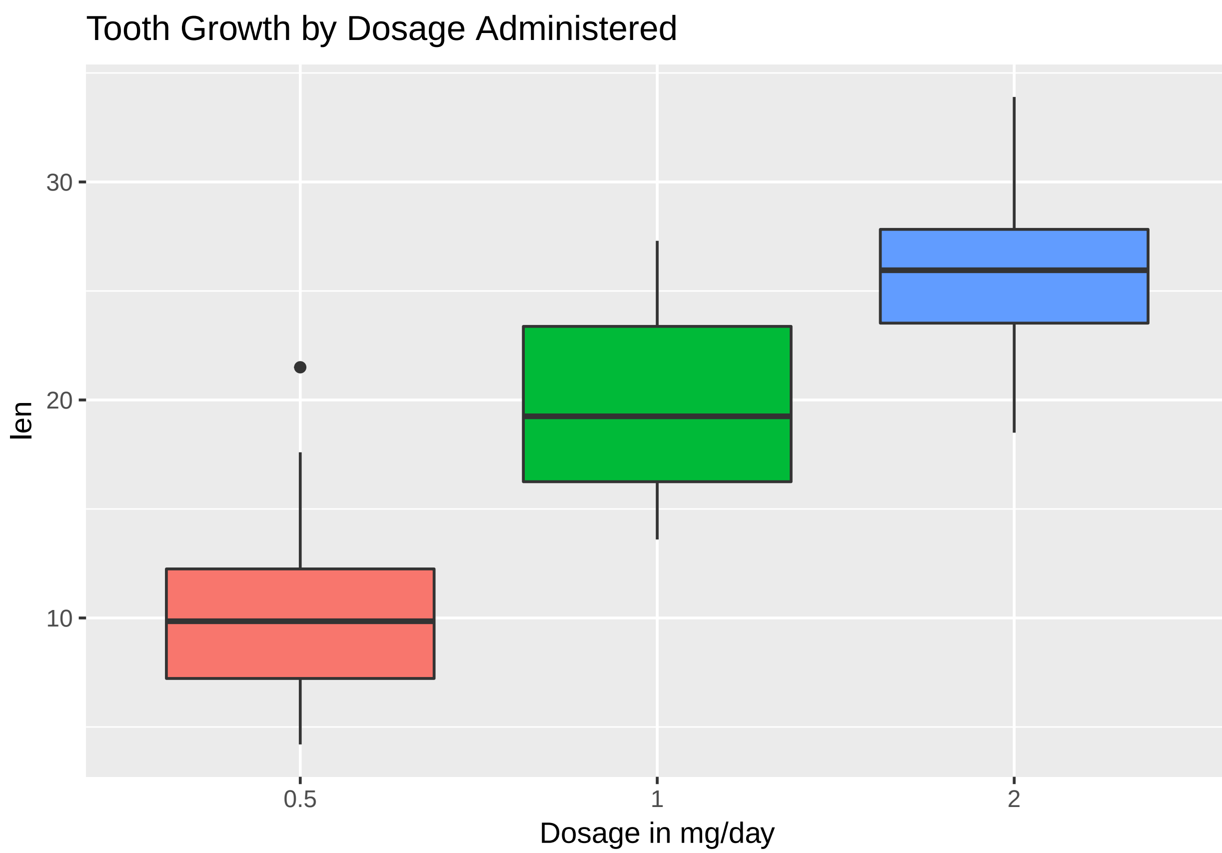
Here again, we can see that the mean tooth length increases as the dosage increases from `0.5 mg` to `1 mg` to `2 mg`. This is in line with our expecations but we will run a t-test later to establish whether this difference is real.

The following plots help further probe our theory.

```r
library(ggplot2)
g <- ggplot(tooth, aes(y= len))
g + geom_boxplot(aes(x = supp, fill = supp), show.legend = FALSE) +
    ggtitle("Tooth Growth by Supplement Delivery Method")
```



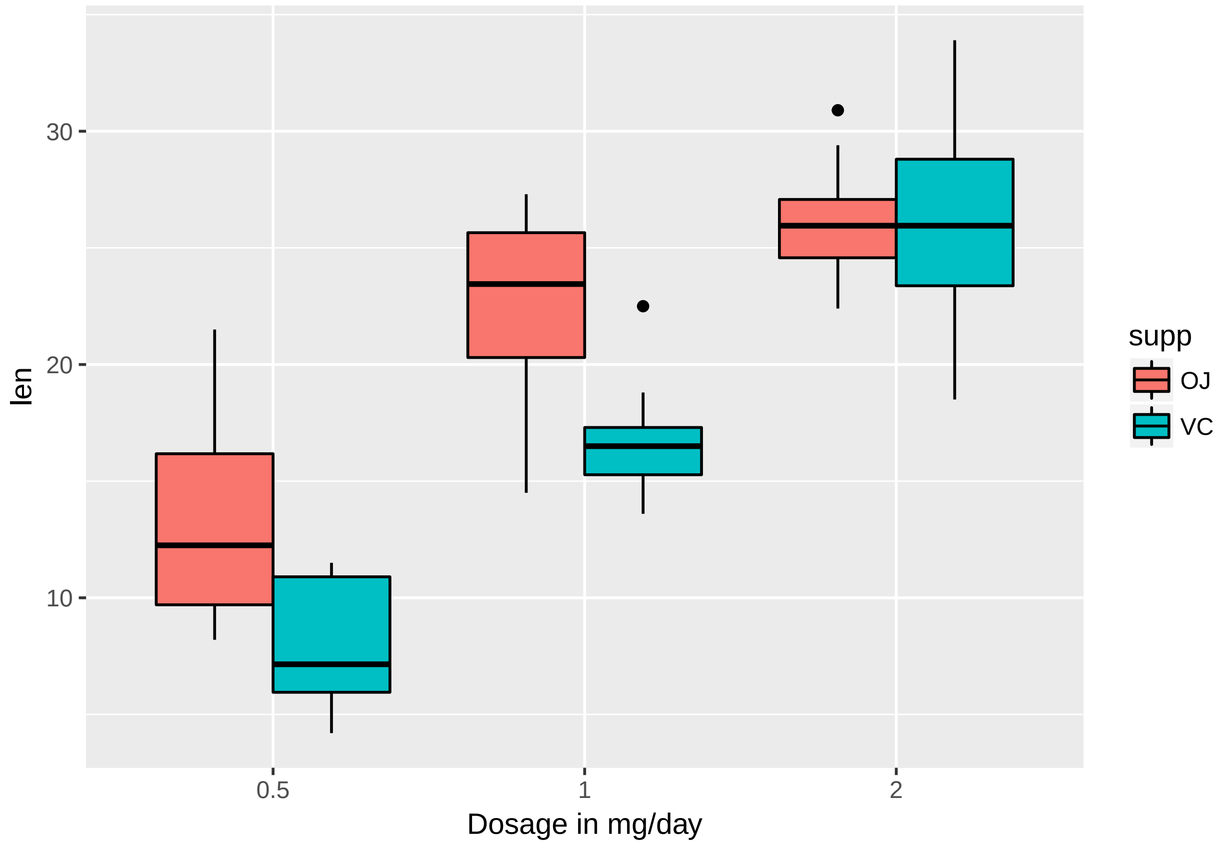

```
g + geom_boxplot(aes(x = as.factor(dose), fill = as.factor(dose)), show.legend = FALSE) +
    ggtitle("Tooth Growth by Dosage Administered") + xlab("Dosage in mg/day")
```



This graph strengthens our theory further that there is an mean tooth growth increases with increase in dosage amount.

In our next graph we combine both factors together to inspect whether for a specific dosage amount the delivery method makes a difference or not.

```
g + geom_boxplot(aes(x = as.factor(dose), fill = supp), color = "black") +
    xlab("Dosage in mg/day")
```



This plot indicates that at lower dosage levels there is a difference in tooth growth based on the delivery method (OJ > VC) but as the dosage level increases this difference diminishes. Our t-test can indicate whether these differences are significant or not.

Hypothesis testing

Based on the above analysis we have following hypothesis:

- **Dosage**
  - H_{01}: There is no difference in tooth length as the dosage increases
  - H_{a1}: The difference in tooth length as the dosage increases `> 0`
- **Delivery method**
  - H_{02}: There is no difference in tooth length based on delivery method used
  - H_{a2}: The difference in tooth length based on delivery method `> 0`
- **Dosage & Delivery method**
  - H_{03}: As the dosage level increases, there is no difference in tooth length based on delivery method used
  - H_{a3}: As the dosage level increases, the difference in tooth length based on delivery method used is `> 0`

We want to 99% confidence levels in our results, therefore our α is `.01`.

Running significance tests

Testing Hypothesis 1

Our first t-test is on dosage amounts and we will test the difference in mean tooth growth based on two groups at a time in the following order:
1. Test 1 is for dosage level `0.5 mg` and `1 mg`
2. Test 2 is for dosage level `1 mg` and `2 mg`

```
test1 <- t.test(len ~ dose, alternative = "less", paired = FALSE,
                var.equal = FALSE, data = subset(tooth, dose != 2), conf.level = .99)
test1 #mu(dose =0.5) < mu(dose=1.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is less than 0
## 99 percent confidence interval:
##       -Inf -5.706443
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```
test2 <- t.test(len ~ dose, alternative = "less", paired = FALSE,
                var.equal = FALSE, data = subset(tooth, dose != 0.5), conf.level = .99)
test2 #mu(dose =1.0) < mu(dose=2.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is less than 0
## 99 percent confidence interval:
##       -Inf -3.207299
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

Result 1

Since `0` is not included in either of the confidence intervals of our tests we reject H_{01} and accept H_{a1} which says that there is difference in tooth growth based on the dosage levels.

Testing Hypothesis 2

We will test whether there is a difference in mean tooth growth based on the delivery method used.

```
test3 <- t.test(len ~ supp, alternative = "greater", paired = F, var.equal = F, data = tooth, conf.level
test3
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
```

```
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 99 percent confidence interval:
##  -0.9280805        Inf
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

Result 2

At 99% confidence levels our confidence interval **includes `0`** so we fail to reject H_{02}.

Testing Hypothesis 3

We will test whether there is a difference in tooth growth based on delivery method, but only at lower levels of dosage. We will run this test in the following steps:

1. step 1: test difference in delivery method for dosage 0.5mg
2. step 2: test difference in delivery method for dosage 1mg
3. step 3: test difference in delivery method for dosage 2mg

```r
test4 <- t.test(len ~ supp, alternative = "greater", paired = F, var.equal = F,
                data = subset(tooth, dose == 0.5), conf.level = .99)
test5 <- t.test(len ~ supp, alternative = "greater", paired = F, var.equal = F,
                data = subset(tooth, dose == 1), conf.level = .99)
test6 <- t.test(len ~ supp, alternative = "greater", paired = F, var.equal = F,
                data = subset(tooth, dose == 2), conf.level = .99)

test4 #OJ > VC ("greater" implies OJ > VC because in levels(tooth$supp) the first level is OJ)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.003179
## alternative hypothesis: true difference in means is greater than 0
## 99 percent confidence interval:
##  0.9384774       Inf
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```r
test5 #OJ > VC ("greater" implies OJ > VC because in levels(tooth$supp) the first level is OJ)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.0005192
## alternative hypothesis: true difference in means is greater than 0
## 99 percent confidence interval:
##  2.113624      Inf
## sample estimates:
## mean in group OJ mean in group VC
```

```
##            22.70            16.77
```

```
test6 #OJ = VC
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.5181
## alternative hypothesis: true difference in means is greater than 0
## 99 percent confidence interval:
##  -4.629237       Inf
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

Result 3

Since in the first two steps our confidence interval did not include `0` we can reject H_{03} for dosage amount `0.5 mg` and `1 mg` and accept our alternate hypothesis that there is a difference in tooth growth based on delivery method.

However, for the thir step our confidence interval **includes** `0` so we fail to reject H_{03} in this instance and hence conclude that, at higher dosage levels (i.e. `2 mg`) there is no difference in tooth growth based on delivery method.

END OF REPORT