

ToothGrowth Project Report

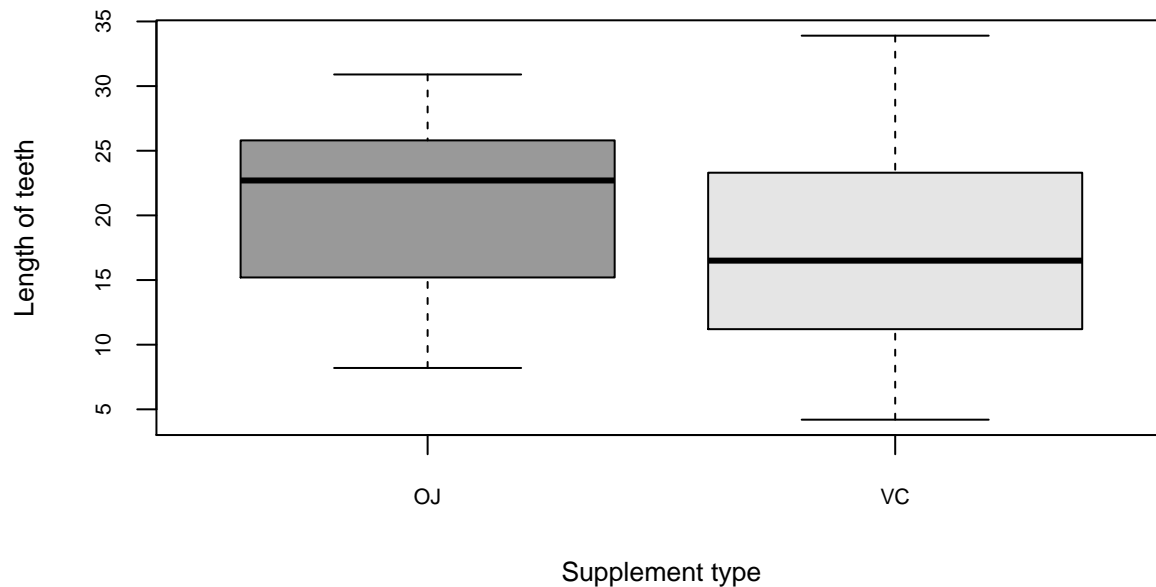
Jose V. Die

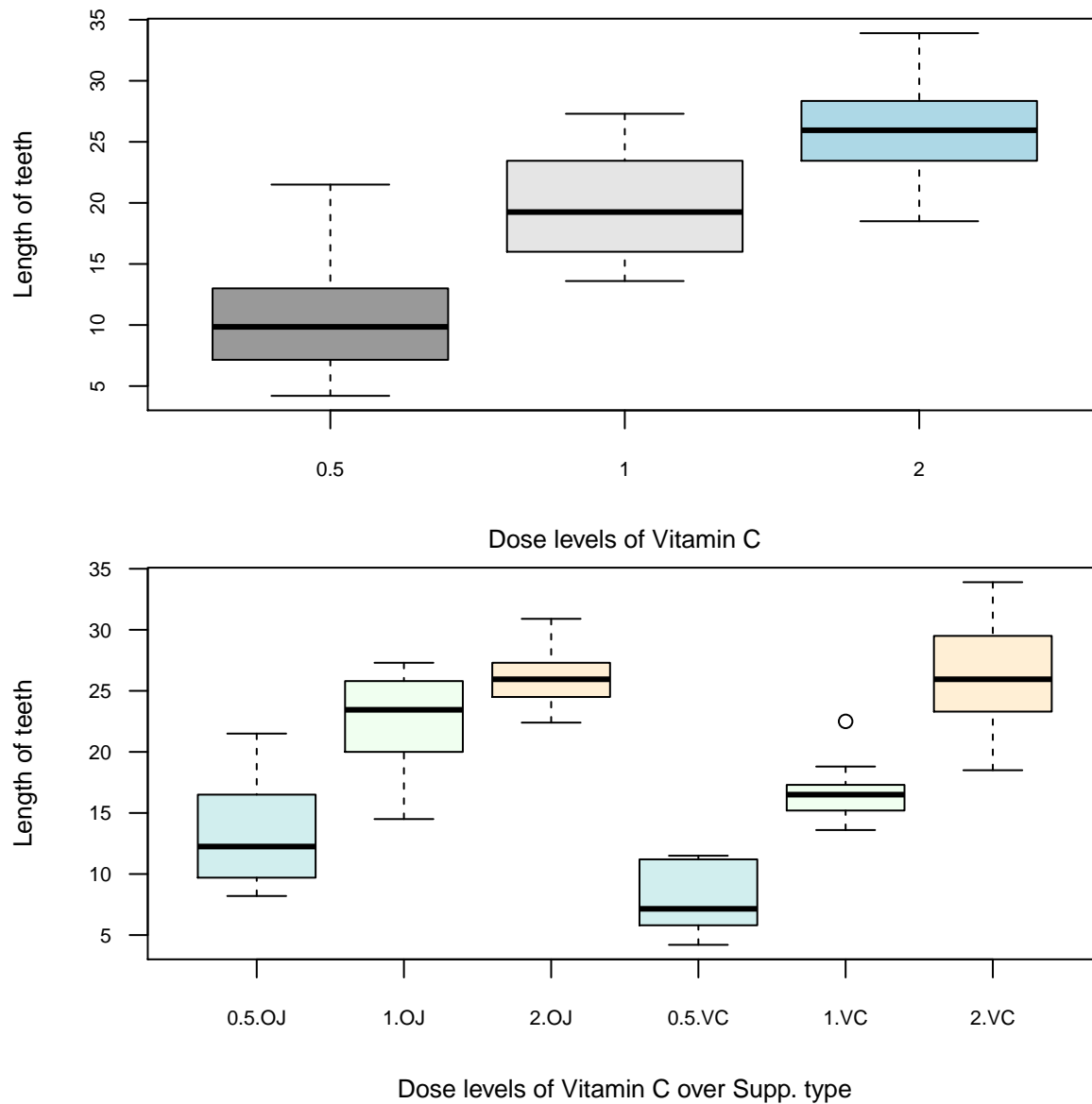
February 27, 2016

Overview

In this project we will analyze the ToothGrowth data in the R datasets package. We want to perform some basic exploratory data analysis in order to extract some basic features from the data. This will allow us to set some hypothesis that will be further tested.

1. We will take a quick look at the data : head, tail, summary. (See supporting Appendix material for code)
2. EDA. We will plot several plots: (1) Length of teeth vs. Supplement type (orange juice or ascorbic acid); (2) Length of teeth vs. Dose levels of Vitamin C; (3) Length of teeth vs. Dose levels of Vitamin C over Supp. type.





From the plots, we will test 3 hypothesis related to:

- * (1) Vitamin C supplement type,
- * (2) dose levels and,
- * (3) interaction between dose levels, and type of Vitamin C

Hypothesis 1: We will test the hypothesis that *there is not difference between the type of supplement of Vitamin C*. This is, mean of lengths of teeth from guinea pigs treated with orange juice is not different than the mean from guinea pigs treated with ascorbic acid. **For the whole Assignment, we will assume equal variance between groups.**

```
OJ <- dat[which(dat$supp=="OJ"),1]
VC <- dat[which(dat$supp=="VC"),1]
statistic <- t.test(OJ,VC, paired = FALSE, var.equal = TRUE )
```

The 95% Confidence Intervals, -0.1670064, 7.5670064 contain 0. Also, the P -value is 0.0603934. So, we cannot reject the null hypothesis (difference in means is equal to 0).

Hypothesis 2: There is not difference in the dose levels of Vitamin C. We will set the 3 subsets based on the dose levels: low, medium and high.

```
low <- dat[which(dat$dose==0.5),1]
medium <- dat[which(dat$dose==1),1]
high <- dat[which(dat$dose==2),1]
```

2.1. H_0 : There is no difference in teeth length between low levels vs. medium levels.

```
sx2 <- t.test(medium,low, paired = FALSE, var.equal = TRUE )
```

P value is ~ 0 . So, we reject the null hypothesis and accept the alternative hypothesis.

2.2. H_0 : There is no difference in teeth length between medium levels vs. high levels.

```
sx3 <- t.test(high, medium, paired = FALSE, var.equal = TRUE )
```

P value is ~ 0 . We reject the null hypothesis and accept the alternative hypothesis.

Hypothesis 3: There is not difference in the Dose levels of Vitamin C over Supp. type.

```
attach(dat)
subset1 <- dat[supp=="VC" & dose==0.5,1]
subset2 <- dat[supp=="OJ" & dose==0.5,1]
sx4 <- t.test(subset2, subset1, paired = FALSE, var.equal = TRUE )
subset1 <- dat[supp=="VC" & dose==1,1]
subset2 <- dat[supp=="OJ" & dose==1,1]
sx5 <- t.test(subset2, subset1, paired = FALSE, var.equal = TRUE )
subset1 <- dat[supp=="VC" & dose==2,1]
subset2 <- dat[supp=="OJ" & dose==2,1]
sx6 <- t.test(subset1, subset2, paired = FALSE, var.equal = TRUE )
```

3.1 Low levels Vitamin C

We will reject the H_0 . With low dose levels of Vitamin C, orange juice results in higher values of teeth length. P value = 0.0053.

3.2 Medium levels Vitamin C

We will reject the H_0 . With medium dose levels of Vitamin C, orange juice results in higher values of teeth length. P value ~ 0 .

3.3 High levels Vitamin C

We cannot reject the null hypothesis for high levels. P value = 0.9637098.

Conclusions:

1. Higher dose levels of Vitamin C result in longer teeth.
2. Type of supplement of Vitamin C, matters. Orange juice results in higher values of teeth length. However, results are statistically significant only for low or medium dose levels of Vitamin C.
3. There are no differences between means from Orange juice or ascorbic acid for high Vitamin C levels. (This explains why we could not find statistically differences in our hypothesis1).

Supporting Appendix Material

Load data

```
data(ToothGrowth)
dat <- ToothGrowth
#help("ToothGrowth")
str(ToothGrowth)
```

```
## '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 ...
```

Quick look at the data

```
head(dat)
```

```
##      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
```

```
tail(dat)
```

```
##      len supp dose
## 55 24.8   OJ  2
## 56 30.9   OJ  2
## 57 26.4   OJ  2
## 58 27.3   OJ  2
## 59 29.4   OJ  2
## 60 23.0   OJ  2
```

```
summary(dat)
```

```
##      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
```

Results statistics for Hypothesis Testing

```
statistic # Hypothesis 1
```

```
##
## Two Sample t-test
##
## data: OJ and VC
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.1670064 7.5670064
## sample estimates:
## mean of x mean of y
## 20.66333 16.96333
```

```
sx2 # Hypothesis 2: low levels vs. medium levels
```

```
##
## Two Sample t-test
##
## data: medium and low
## 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:
## 6.276252 11.983748
## sample estimates:
## mean of x mean of y
## 19.735 10.605
```

```
sx3 # Hypothesis 2: medium levels vs. high levels
```

```
##
## Two Sample t-test
##
## data: high and medium
## 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:
## 3.735613 8.994387
## sample estimates:
## mean of x mean of y
## 26.100 19.735
```

```
sx4 # Hypothesis 3: levels: low level over Supp Types
```

```
##
## Two Sample t-test
##
## data: subset2 and subset1
## 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 of x mean of y
## 13.23 7.98
```

```
sx5 # Hypothesis 3: levels: medium level over Supp Types
```

```
##  
## Two Sample t-test  
##  
## data: subset2 and subset1  
## 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 of x mean of y  
## 22.70 16.77
```

```
sx6 # Hypothesis 3: levels: high level over Supp Types
```

```
##  
## Two Sample t-test  
##  
## data: subset1 and subset2  
## t = 0.046136, df = 18, p-value = 0.9637  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -3.562999 3.722999  
## sample estimates:  
## mean of x mean of y  
## 26.14 26.06
```

EDA. Code for Boxplots

```
boxplot(dat$len~dat$supp, col=c("grey60", "grey90"), ylab="Length of teeth",  
        xlab="Supplement type", cex.lab = 0.8, cex.axis = 0.7)  
  
boxplot(dat$len~dat$dose, col=c("grey60", "grey90", "lightblue"), ylab="Length of teeth", xlab="Dose level",  
        cex.lab = 0.8, cex.axis = 0.7)  
  
boxplot(dat$len~dat$dose*dat$supp, ylab="Length of teeth",  
        xlab="Dose levels of Vitamin C over Supp. type", ylab="Length of teeth",  
        col=rep(c("lightcyan2", "honeydew1", "papayawhip"),2), las=1, cex.lab=0.8, cex.axis=0.7)
```