Statistical Inference: Course Project Part2

Author: Jens Berkmann

Introduction

In this report we will do investigations on the ToothGrowth dataset provided by the R package. According to the R help function the dataset provides information to study the effect of Vitamin C on tooth growth in Guinea pigs. The dataset reports 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). Let's do first some basic data analysis.

Basic Data Analysis

We first load the data and print a summary.

```
library(knitr)
data(ToothGrowth)
tg      <- ToothGrowth
names(tg)
```

```
## [1] "len"  "supp" "dose"
```

```
tg$dose <- as.factor(tg$dose)
str(tg)
```

```
## '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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

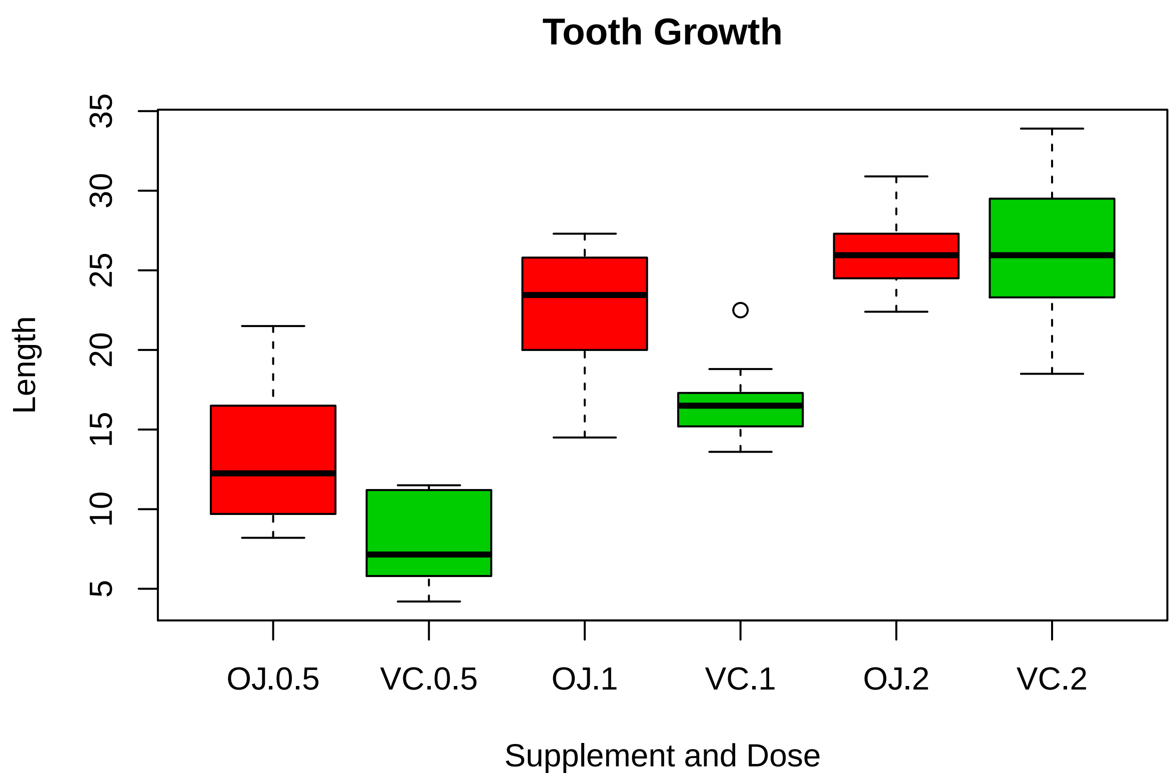
```
summary(tg)
```

```
##       len        supp     dose
##  Min.   : 4.2   OJ:30   0.5:20
##  1st Qu.:13.1   VC:30   1  :20
##  Median :19.2           2  :20
##  Mean   :18.8
##  3rd Qu.:25.3
##  Max.   :33.9
```

The dataframe contains of 60 rows (samples) and three columns. The column *len* contains the measured length of the teeth as a numeric value. The 2nd column *supp* is a binary factor variable distinguishing between the 2 delivery methods *orange juice (OJ)* or *ascorbic acid (VC)*. The 3rd column which we converted already into a factor variable distinguishes between the 3 possible dose levels 0.5, 1, and 2 mg.

Let us look at the variable *len* conditioned on the *(supp,dose)*-value-pairs.

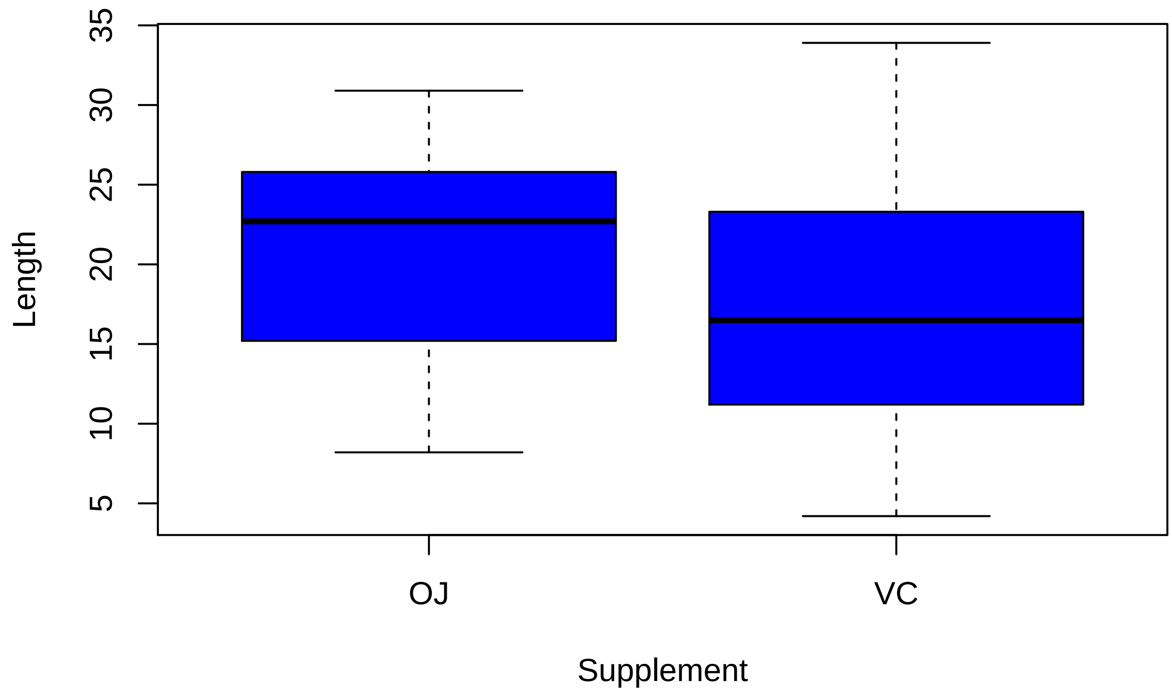
```
boxplot(len ~ interaction(supp,dose), data=tg, col=2:3,main="Tooth Growth", xlab="Supplement and Dose",
```



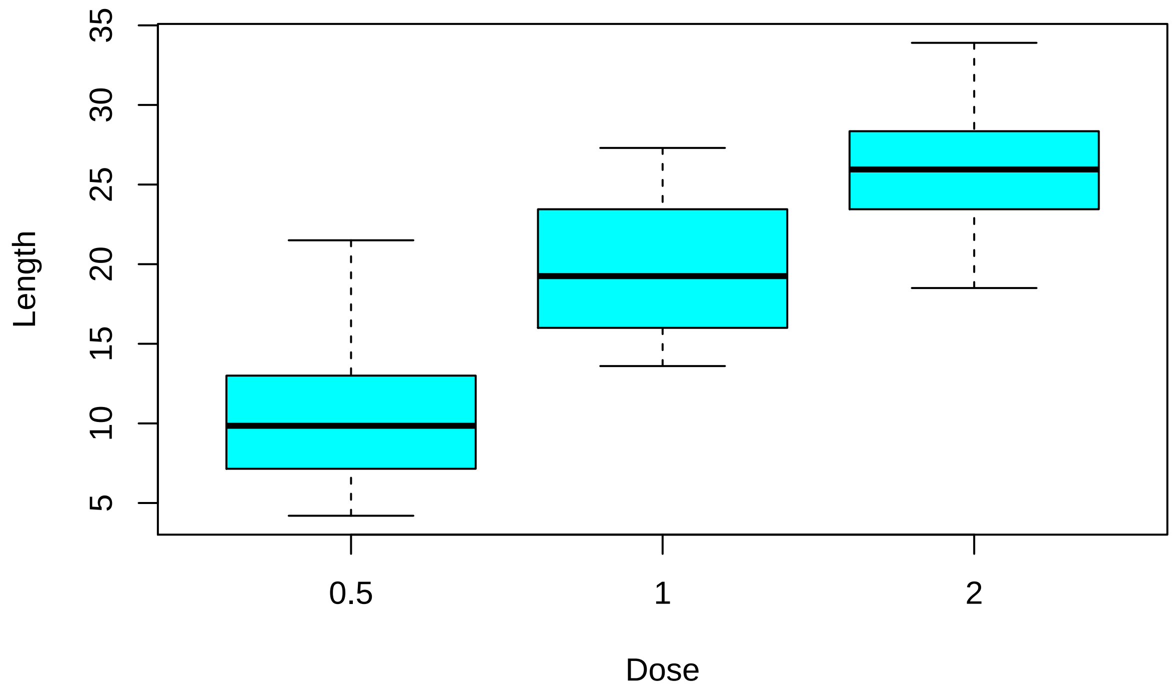
It is seen that the median of the tooth length increases if the dose is increased no matter how the dose is delivered. Apart from the highest dose (2 mg) the delivery with orange juice (OJ) also shows an increase of the tooth length.

Let us finally plot also the dependency of the tooth length on the individual variables *supp* and *dose*, respectively.

```
boxplot(len ~ supp, data=tg, col=4, xlab="Supplement", ylab="Length")
```




```
boxplot(len ~ dose, data=tg, col=5, xlab="Dose", ylab="Length")
```



From the last figure there is evidence that increasing the dose leads to stronger tooth growth since the 50% intervals are not overlapping. The effect of the delivery method (supplement) on tooth growth is not that strong since the 50% intervals are significantly overlapping.

As a final explanatory data analysis task we will compute the mean and standard deviation within each of the 6 groups.

```
s<-split(tg$len,list(tg$supp,tg$dose))
sapply(s,mean)
```

```
## OJ.0.5 VC.0.5   OJ.1   VC.1   OJ.2   VC.2
##  13.23   7.98  22.70  16.77  26.06  26.14
```

```
sapply(s,sd)
```

```
## OJ.0.5 VC.0.5   OJ.1   VC.1   OJ.2   VC.2
##  4.460  2.747  3.911  2.515  2.655  4.798
```

Confidence Intervals and Hypothesis Testing

In order to quantitatively study the effect of the delivery method (supplement) and dose on tooth growth we setup various hypothesis tests. Since the tested pigs are all different within each of the 3*2 groups we conduct independent t-tests. Moreover, we believe that no assumption on variance-equality can be made within each tested group of 10 pigs, we therefore apply t-tests for independent groups and unequal variances. Let's first turn to the question whether supplement has any effect on tooth growth. The NULL-Hypothesis is that the difference of the two means is equal.

```r
t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = tg)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.915, df = 55.31, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.171  7.571
## sample estimates:
## mean in group OJ mean in group VC
##            20.66            16.96
```

Looking at the results we see that the difference in means is roughly 4. The confidence interval (CI) includes zero which indicates no strong evidence that H_0 should be rejected. Moreover, the p-value (~6%) is also slightly above the confidence level (5%) which neither gives strong evidence to reject H_0. We therefore stick to H_0 and conclude that it does not have a significant effect on the tooth growth whether Vitamin C is delivered in the form of orange juice or in the form of ascorbic acid.

We now turn to the question whether the dose of vitamin C has a significant effect on tooth growth. Let us start by comparing a dose of 0.5 mg with a dose of 1 mg. From the explanatory graphs it is quite evident that increasing the dose does lead to higher tooth growth, we therefore conduct one-sided tests with $H_0 : \mu_{0.5} - \mu_1 = 0$ and $H_\alpha = \mu_{0.5} - \mu_1 < 0$.

```r
tg_1 <- tg[tg$dose=="0.5",]$len
tg_2 <- tg[tg$dose=="1",]$len
tg_3 <- tg[tg$dose=="2",]$len
t.test(tg_1,tg_2, paired = FALSE, var.equal = FALSE, alt="less",data = tg)
```

```
##
##  Welch Two Sample t-test
##
## data:  tg_1 and tg_2
## t = -6.477, df = 37.99, p-value = 6.342e-08
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##    -Inf -6.753
## sample estimates:
## mean of x mean of y
##     10.61     19.73
```

It is seen that the test statistics is negative and is contained in the half-open CI which gives rise to reject H_0. Moreover, the p-value is fairly which gives also evidence to reject H_0. We conclude that there is strong evidence that μ_1 is larger μ_0 meaning that higher doses lead to larger tooth growth.

We do an analogous test by comparing a dose of 1.0 mg with a dose of 2 mg. We conduct a one-sided tests with $H_0 : \mu_1 - \mu_2 = 0$ and $H_\alpha = \mu_1 - \mu_2 < 0$.

```r
t.test(tg_2,tg_3, paired = FALSE, var.equal = FALSE, alt="less",data = tg)
```

```
##
```

```
##  Welch Two Sample t-test
##
## data:  tg_2 and tg_3
## t = -4.901, df = 37.1, p-value = 9.532e-06
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##    -Inf -4.174
## sample estimates:
## mean of x mean of y
##     19.73     26.10
```

For similar reasons as above we reject H_0. The p-value is larger for this test compared to the previous one, however still too small in order to arise by accident. We conclude that there is strong evidence that μ_2 is larger μ_1 meaning that even further increase the dose leads to larger tooth growth.

Finally We do an analogous test by comparing a dose of 0.5 mg with a dose of 2 mg. We conduct a one-sided tests with $H_0 : \mu_1 - \mu_3 = 0$ and $H_\alpha = \mu_1 - \mu_3 < 0$.

```
t.test(tg_1,tg_3, paired = FALSE, var.equal = FALSE, alt="less",data = tg)
```

```
##
##  Welch Two Sample t-test
##
## data:  tg_1 and tg_3
## t = -11.8, df = 36.88, p-value = 2.199e-14
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##    -Inf -13.28
## sample estimates:
## mean of x mean of y
##     10.61     26.10
```

Due to the tiny p-value and CI-enclosure we reject H_0.

All in all, We conclude that there is strong evidence that increasing the dose leads to larger tooth growth. However, the for the underlying supplement there is no strong evidence that is has an effect on tooth growth.