Statistical Inference Course project Part 2.

Aleksandr Beloushkin (aka gudun-ku)

October 25, 2015.

The gist

In this report, we're going to analyze the ToothGrowth data in the R datasets package. To provide this we need complete following tasks:

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.
4. State conclusions and the assumptions needed for your conclusions.

Necessary R packages (datasets, ggplot2,dplyr)

```r
library(datasets)
library(ggplot2)
```

Loading data and providing initial exploratory analysis

```r
data("ToothGrowth")
head(ToothGrowth,3)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
```

Summary of ToothGrowth

```r
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  
```

Getting means of Len variable between supplements

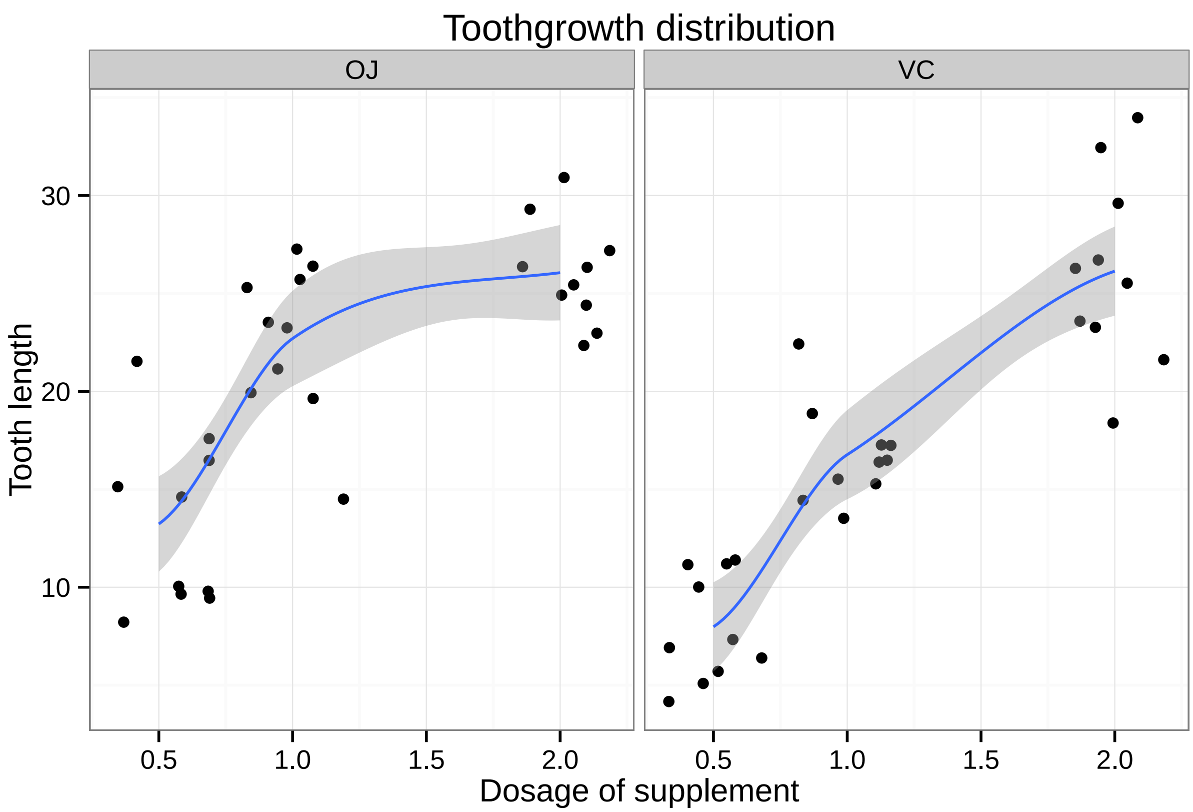
```r
means_table <- split(ToothGrowth$len, ToothGrowth$dose)
sapply(means_table,mean)
```

```
##    0.5      1      2
## 10.605 19.735 26.100
```

Plot the data

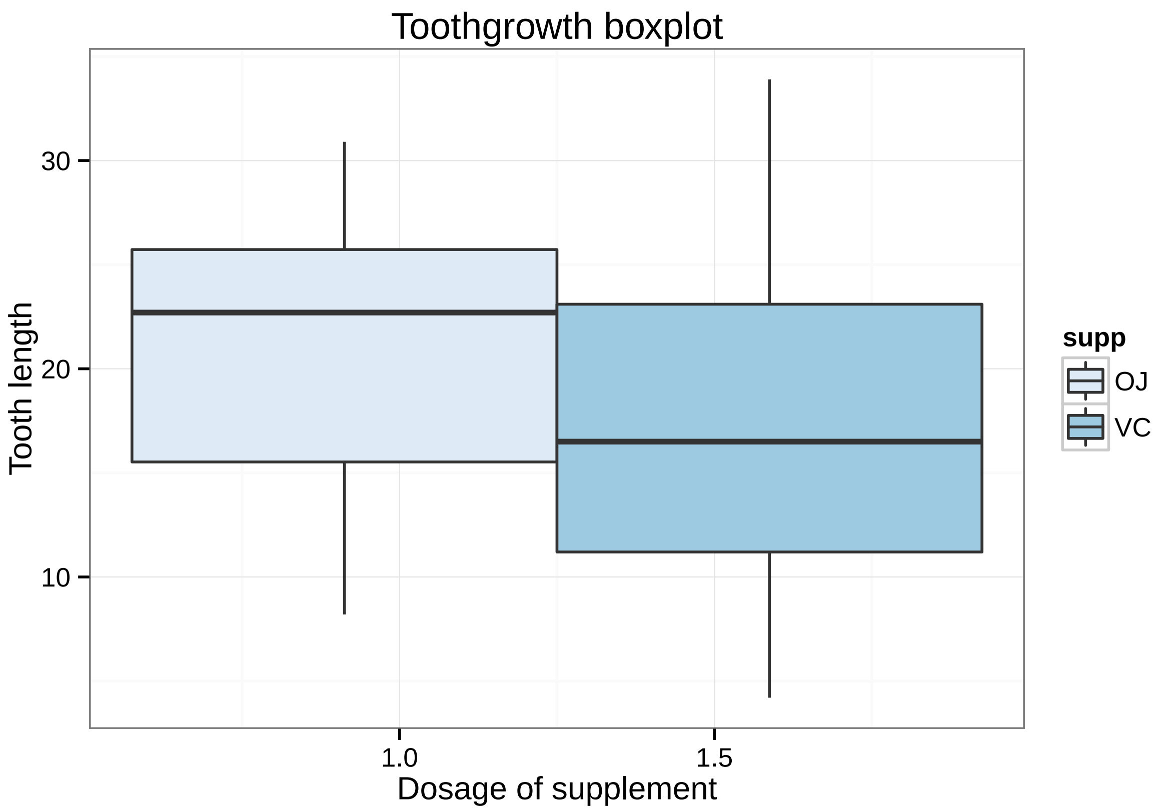
```r
g <- qplot(dose,len, data=ToothGrowth,
geom = "jitter",
main="Toothgrowth distribution",
xlab="Dosage of supplement",
ylab="Tooth length")

g <- g + facet_wrap(~supp)
g <- g + stat_smooth()
g <- g + theme_bw()
g
```



From first sight we can see the difference between two distributions in their form on plots, showing possible effect from changing supplies, and it would be even clearer shown if we form a boxplot.

```r
g <- ggplot(aes(y = len, x = dose, fill = supp), data = ToothGrowth)
g <- g + geom_boxplot()
g <- g + scale_fill_brewer()
g <- g + theme_bw()
g <- g + ggtitle("Toothgrowth boxplot")
g <- g + xlab("Dosage of supplement")
g <- g + ylab("Tooth length")
g
```



So, our hypothesis to check first if there is an approved difference between how each kind of supplement really affects the tooth growth. Our first null hypothesis than states that null hypothesis states that really there is no difference between using two supplements.

Using confidence intervals and hypothesis tests to compare tooth growth by supp and dose.

```r
len<-ToothGrowth$len
supp<-ToothGrowth$supp
dose<-ToothGrowth$dose
sapply(means_table, var)
```

```
##      0.5        1        2
## 20.24787 19.49608 14.24421
```

Difference between two groups

The variances of the two groups appears to be significant, so in the T-test we will not assume equality of variances,so our Null Hypothesis: OJ and VC show no significant changes in Tooth growth, thus we'll first test where both pairs are assumed to be FALSE

```
t.test(len[supp=="OJ"], len[supp=="VC"], paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len[supp == "OJ"] and len[supp == "VC"]
## 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 of x mean of y
##  20.66333  16.96333
```

The OJ and VC p-value is 0.06, which no significantly greater than 5%. This result shows that the means of tooth length between OJ and VC group are not statistically significantly different at confidence level 0.05 (P = 0.06), and the 95% confidance interval contains 0. Thus we cannot reject the null hypothesis.

Affect of supplement dosage

Let's test variables assumed to be equal, the null hypotheses of equal means between the two groups, versus the alternative hypothesis that the two means are different. For that we will use two-sided test, and that fact that the confidence interval is absolutely equivalent to the t-test.

```
t.test(len[dose==2], len[dose==1], paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  len[dose == 2] and len[dose == 1]
## 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
```

The p-value of this test is 0; which states that the null hypothesis is FALSE. The 95% confidence interval is (-0.1710156 7.5710156) proves the value of zero is not included. i.e we rejected null hypothesis.

Conclusions

This study of tooth growth effected by supplement type and dose reflects that there is a relationship between tooth length and amount of supplement dosage, approved that with dosage increasing there is an approwed positive effect on tooth length. However, we couln't approve a significant difference between effects of different supplement types. (Which were shown in our graphs and T test).