Statistical Inference Course Project Part 2

by Szymon Tomczyk

```
library(tidyverse)
library(datasets)
```

Part 2 Basic Inferential Data Analysis

In this part we will use the ToothGrowth dataset studying the impact of dose and delivery method of Vitamin C on tooth growth in guiney pig. I will perform an exploratory analysis, look at the summary of the data and do some basic hyptotheis testing.

Loading the data

```
data("ToothGrowth")
```

Exploratory analysis

Look at the structure of the data and the summary statistics

```
str(ToothGrowth)
```

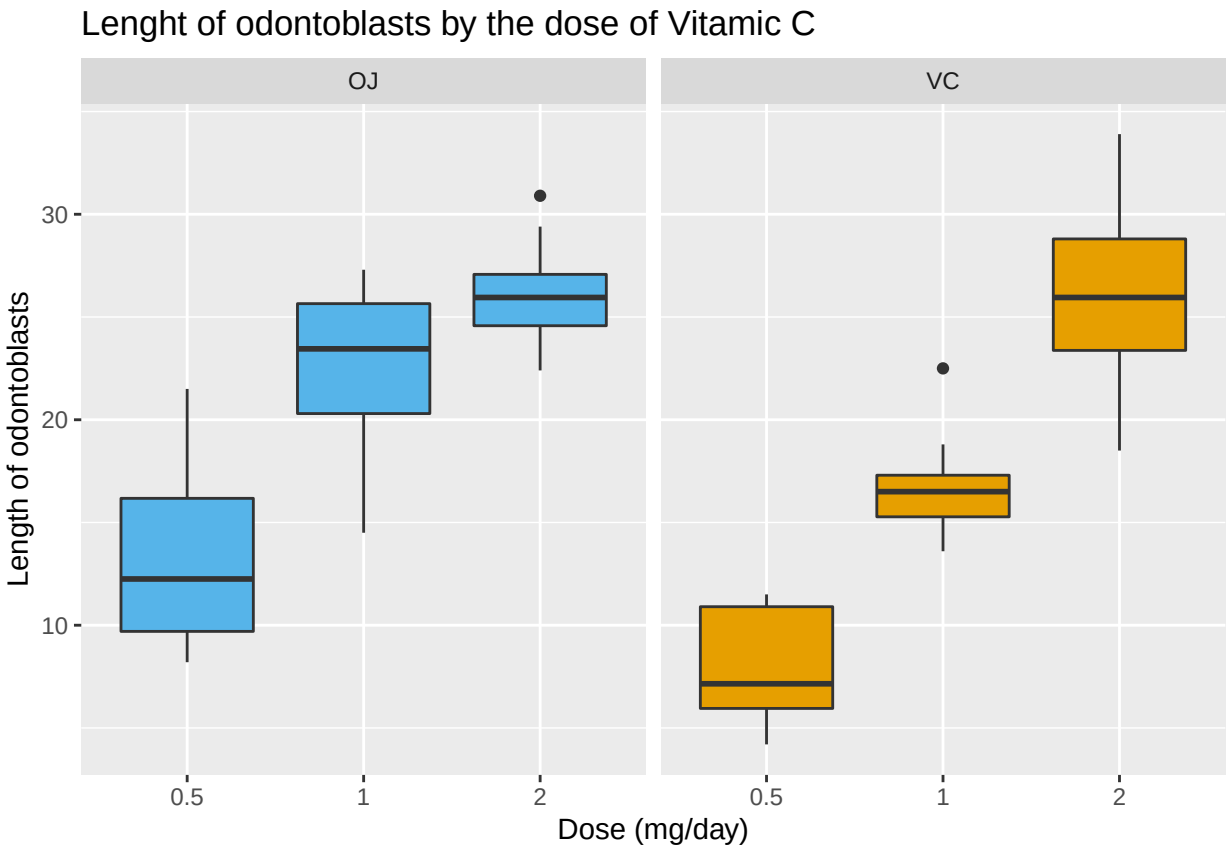
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
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
```

Plot of the Lenght of the odontoblasts by the method of delivery: VC - ascorbic acid, OJ - orange juice.

```
ggplot(ToothGrowth, aes(x = as.factor(dose), y = len)) +
        geom_boxplot(aes(fill=supp)) +
        scale_fill_manual(values=c("#56B4E9", "#E69F00"))+
        facet_wrap(.~supp)+
        labs(title = "Lenght of odontoblasts by the dose of Vitamic C") +
        xlab("Dose (mg/day)") +
        ylab("Length of odontoblasts") +
        theme(legend.position = "none")
```



Hypothesis Testing

Test if there is a difference in tooth growth between the group that was given ascorbic acid and orange juice.

```
t.test(ToothGrowth$len ~ ToothGrowth$supp)
```

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

The 95% confidence interval contains the difference between the means of two groups so we cannot reject the null hipothesis

Test if there is difference in tooth growth between the group that was given low (0.5) and high (2) dose of Vitamin C.

```r
ToothGrowth.sub <- filter(ToothGrowth, dose  %in% c(.5, 2))
t.test(ToothGrowth.sub$len ~ ToothGrowth.sub$dose)
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth.sub$len by ToothGrowth.sub$dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

The observed p-value is very low, much lower than alpha = 0.05. Based on this fact we cannot assume that the means of the populations are equal. We reject the null hypthesis

Conclusions

Given the data we analyzed we can conclude that:

1. The method of delivery does not impact the tooth growth in guinea pig
2. The dose impacts the tooth growth in guinea pig

Assumptions

1. The samples come from independent identicaly distributed populations
2. The samples are representative of the population