Basic Statistical Inference

Lukasz Konczyk

30 kwietnia 2018

Part Two

In this part of Statistical Inference Coursera project, we will focus on basic statistical inference in data. Please look at the dataset, which we will analise, from R help:

The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice or ascorbic acid (a form of vitamin C and coded as VC).

Data loading

Now, we load data and look at summary.

```
library(datasets)
data(ToothGrowth)
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
```

```
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 ...
```

```
dim(ToothGrowth)
```

```
## [1] 60  3
```

We will compare length of tooth growth in comparision groups: supp and dose. Firstly, we must transform variable dose to factor.

```
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
summary(ToothGrowth)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

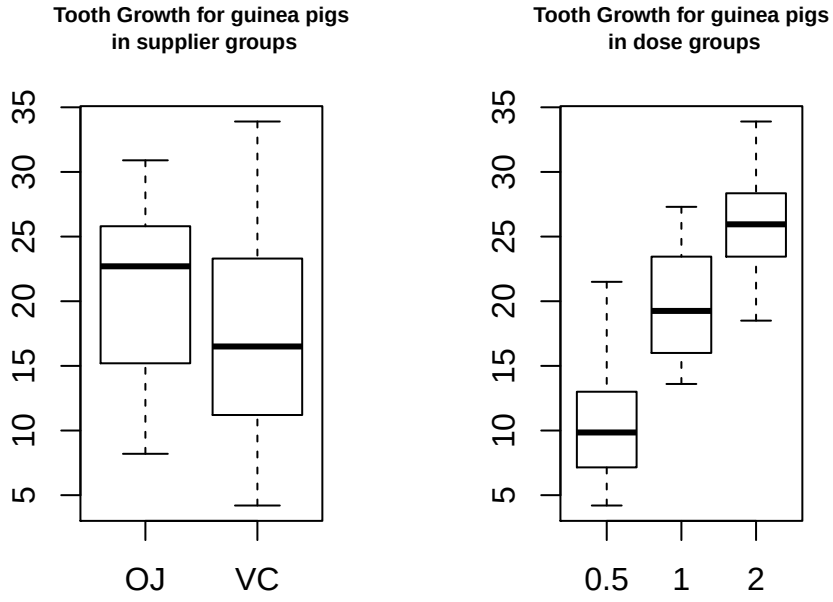
Data analysis

As in data we have tooth growth for 60 different guinea pigs, to compare impact of dose or supplier in group, we must use t test for independent groups. T test compares means between group.

Before we apply statistical tests, we prepare some basic plots.

```
par(mfrow=c(1,2))
boxplot(ToothGrowth$len~ToothGrowth$supp, main="Tooth Growth for guinea pigs\n in supplier groups",cex.m

boxplot(ToothGrowth$len~ToothGrowth$dose, main="Tooth Growth for guinea pigs\n in dose groups", cex.mai
```



Looking at first plot, we may state, that there is no influence between supply way of vitamin C. But if we look at second plot, we may conclude that there is dependancy between tooth growth and dose. So, we check our hypothesis. Firstly we check supply way.

```
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
```

T test p-value is larger than 5%. Also t test confidence interval contains 0. This confirms that there is no effect between two supply way. Now have look at dose. As we have 3 dose, we can compare only two groups

and we must do 3 tests.

```
ToothGrowth_0.5_1<-subset(ToothGrowth, dose!=2)
t.test(ToothGrowth_0.5_1$len~ToothGrowth_0.5_1$dose)
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth_0.5_1$len by ToothGrowth_0.5_1$dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

Looking at this test, we see that p-value is significant lower than 5% and confidence interval doesn't contain 0, so there is difference in tooth growth by dose 0.5 and 1.

```
ToothGrowth_0.5_2<-subset(ToothGrowth, dose!=1)
t.test(ToothGrowth_0.5_2$len~ToothGrowth_0.5_2$dose)
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth_0.5_2$len by ToothGrowth_0.5_2$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
```

Looking at this test, we see that p-value is significant lower than 5% and confidence interval doesn't contain 0, so there is difference in tooth growth by dose 0.5 and 2.

```
ToothGrowth_1_2<-subset(ToothGrowth, dose!=0.5)
t.test(ToothGrowth_1_2$len~ToothGrowth_1_2$dose)
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth_1_2$len by ToothGrowth_1_2$dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

Looking at this test, we see that p-value is significant lower than 5% and confidence interval doesn't contain 0, so there is difference in tooth growth by dose 1 and 2.

Summary

Looking overall at analysis, we can deduce following proposals:

1. Supply way doesn't influence on tooth growth guinei pigs.
2. There is significant dependancy between dose tooth growth of guinei pigs - larger dose cause bigger tooth growth.