Statistical Inference Course Project. Part 2: Basic Inferential Data Analysis Instructionsless

Vitalii Zhukov

Project Description

Analyze the ToothGrowth data in the R datasets package.

1. Exploratory data analyses

ToothGrowth data loading:

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

Basic exploratory data analyses:

```
head(ToothGrowth)
```

```
##    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
```

```
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
```

```
unique(ToothGrowth$len)
```

```
##  [1]  4.2 11.5  7.3  5.8  6.4 10.0 11.2  5.2  7.0 16.5 15.2 17.3 22.5 13.6
## [15] 14.5 18.8 15.5 23.6 18.5 33.9 25.5 26.4 32.5 26.7 21.5 23.3 29.5 17.6
## [29]  9.7  8.2  9.4 19.7 20.0 25.2 25.8 21.2 27.3 22.4 24.5 24.8 30.9 29.4
## [43] 23.0
```

```
unique(ToothGrowth$supp)
```

```
## [1] VC OJ
```

```
## Levels: OJ VC
```

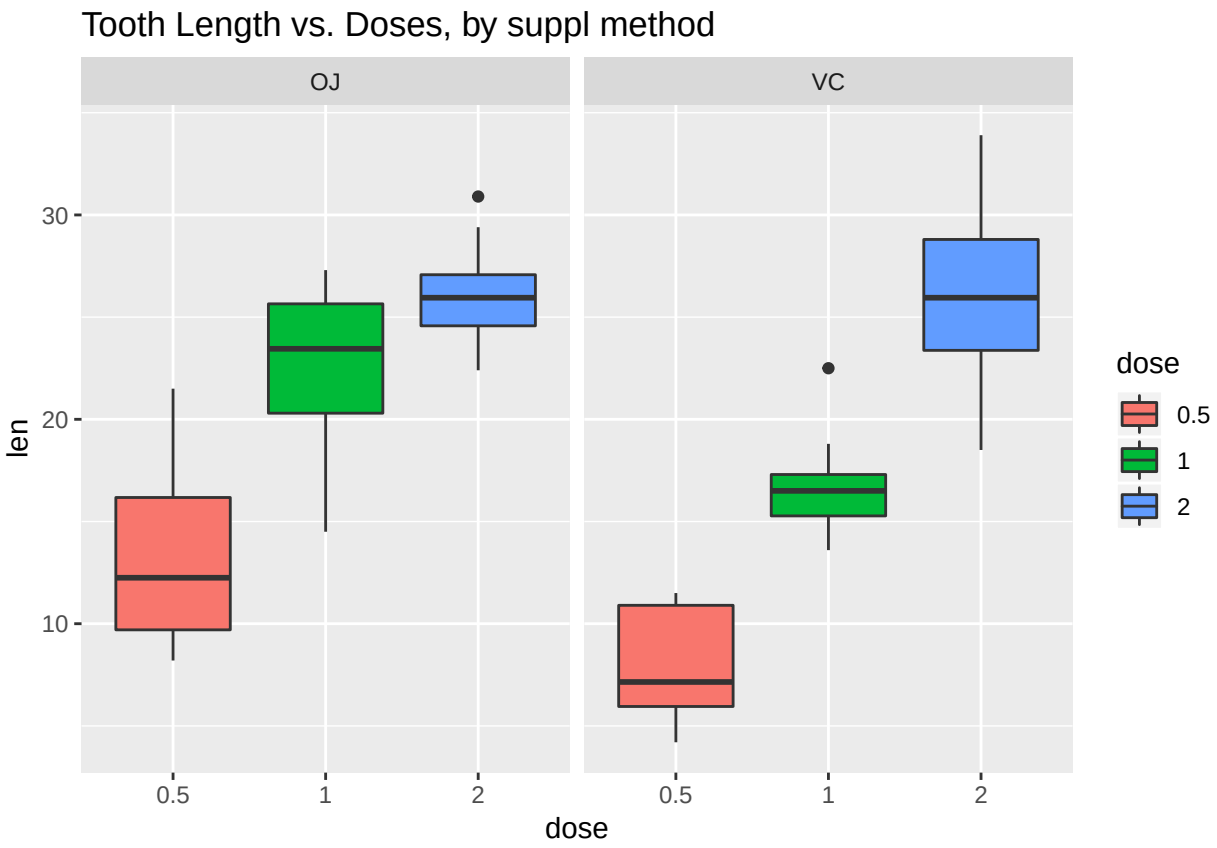
```
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

```
table(ToothGrowth$dose, ToothGrowth$supp) # split of cases between different (dose levels and delivery
```

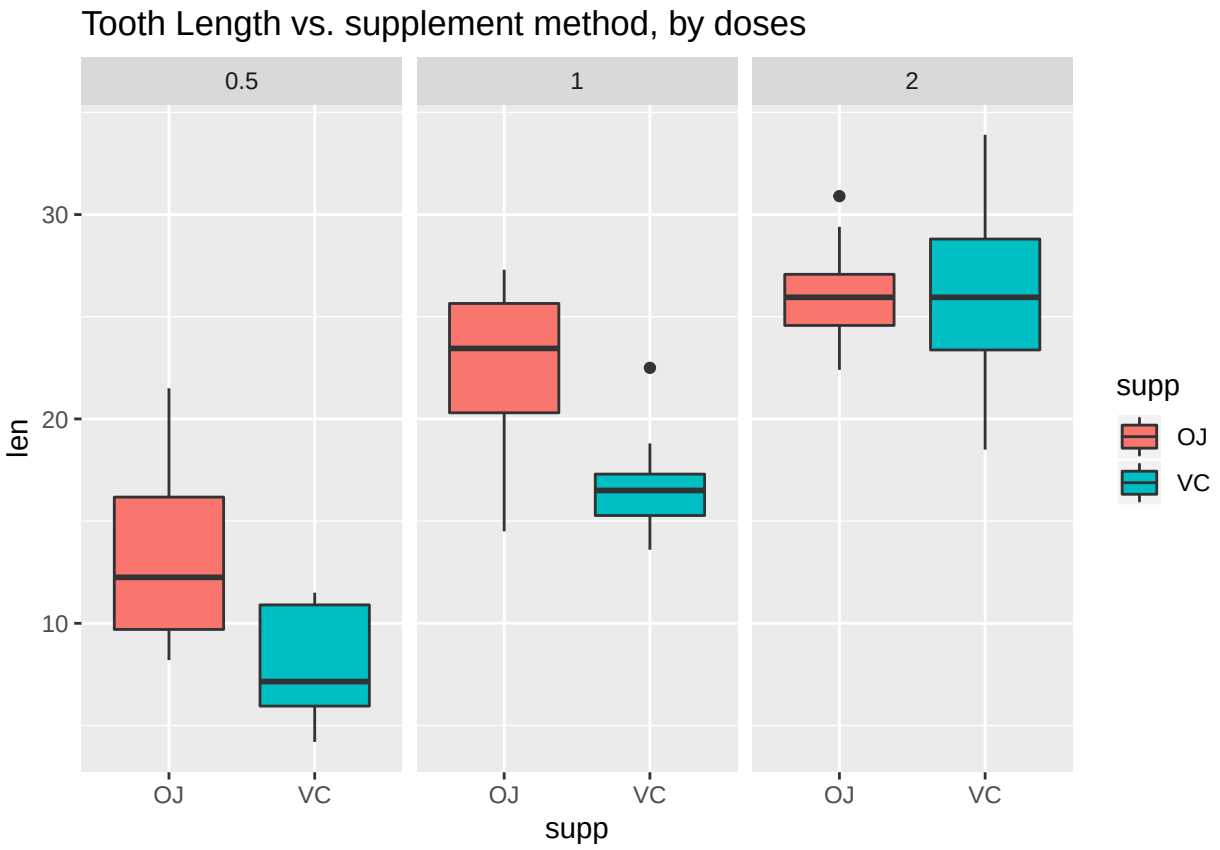
```
##
##       OJ VC
##   0.5 10 10
##   1   10 10
##   2   10 10
```

```
library(ggplot2)
##ggplot(aes(x=dose, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=dose))
ToothGrowth$dose = as.factor(ToothGrowth$dose)
g <- ggplot(ToothGrowth, aes(x=dose, y= len))
g + geom_boxplot(aes(fill=dose)) +facet_grid(.~supp) + ggtitle("Tooth Length vs. Doses, by suppl method
```



Function of supplement type

```
g <- ggplot(ToothGrowth, aes(x=supp, y= len))
g + geom_boxplot(aes(fill=supp)) + facet_grid(.~dose) +
     ggtitle("Tooth Length vs. supplement method, by doses")
```



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

t-test

Check for group differences due to different supplement type, assuming unequal variances between the two groups.

```
t.test(len ~ supp, data = ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by 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 p-value is 0.06, and the confidence interval contains zero. This indicates that we can not reject the null hypothesis that the different supplement types have no effect on tooth length.

Create three sub-groups as per dose level pairs

```
ToothGrowth.doses_0.5_1.0 <- subset (ToothGrowth, dose %in% c(0.5, 1.0))
ToothGrowth.doses_0.5_2.0 <- subset (ToothGrowth, dose %in% c(0.5, 2.0))
ToothGrowth.doses_1.0_2.0 <- subset (ToothGrowth, dose %in% c(1.0, 2.0))
```

Check for group differences due to different dose levels (0.5, 2.0), assuming unequal variances between the two groups

```
t.test(len ~ dose, data = ToothGrowth.doses_0.5_2.0)
```

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

Check for group differences due to different dose levels (0.5, 1.0), assuming unequal variances between the two groups

```
t.test(len ~ dose, data = ToothGrowth.doses_0.5_1.0)
```

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

Check for group differences due to different dose levels (1.0, 2.0),assuming unequal variances between the two groups

```
t.test(len ~ dose, data = ToothGrowth.doses_1.0_2.0)
```

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

For all three dose level pairs, the p-value is less than 0.05, and the confidence interval does not contain zero. The mean tooth length increases on raising the dose level. This indicates that we can reject the null

hypothesis, and establish that increasing the dose level leads to an increase in tooth length.

Conclusions

Supplement type has no effect on tooth growth. Inreasing the dose level leads to increased tooth growth.

Assumptions

The experiment was done with random assignment of guinea pigs to different dose level categories and supplement type to control for confounders that might affect the outcome. Members of the sample population, i.e. the 60 guinea pigs, are representative of the entire population of guinea pigs. This assumption allows us to generalize the results. For the t-tests, the variances are assumed to be different for the two groups being compared. This assumption is less stronger than the case in which the variances are assumed to be equal.