Part 2: Basic Inferential Data Analysis

Tomasz Dworowy

```
knitr::opts_chunk$set(warning=FALSE, message=FALSE)
```

```
if (!require("ggplot2"))
  install.packages("ggplot2")
library(ggplot2)
if (!require("pander"))
  install.packages("pander")
library(pander)
```

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

Overview

Goal of this analysis is to perform exploratory analyses and hypothesis testing to compare tooth growth coused by supplement type and dose on ToothGrowth R data set.

Exploratory analisis

data summary

```
summary(ToothGrowth)
```

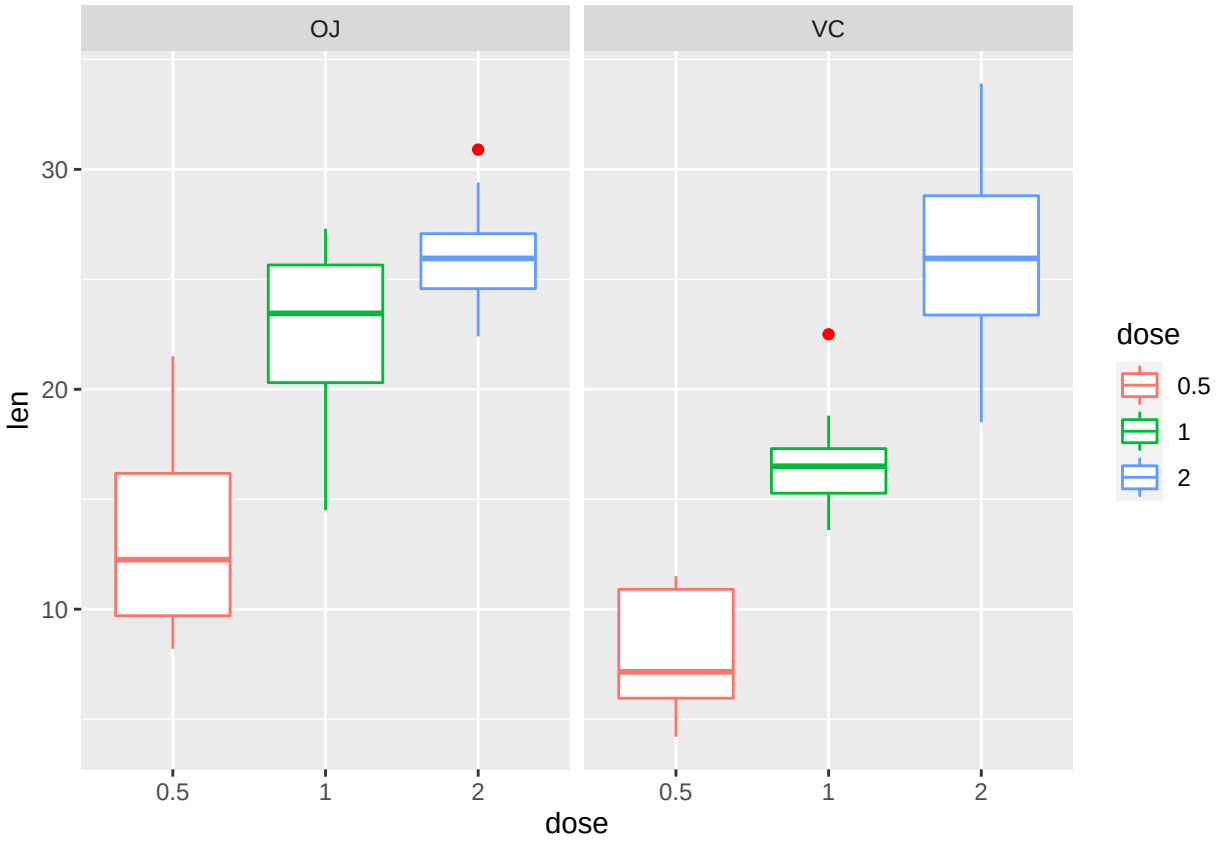
```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

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

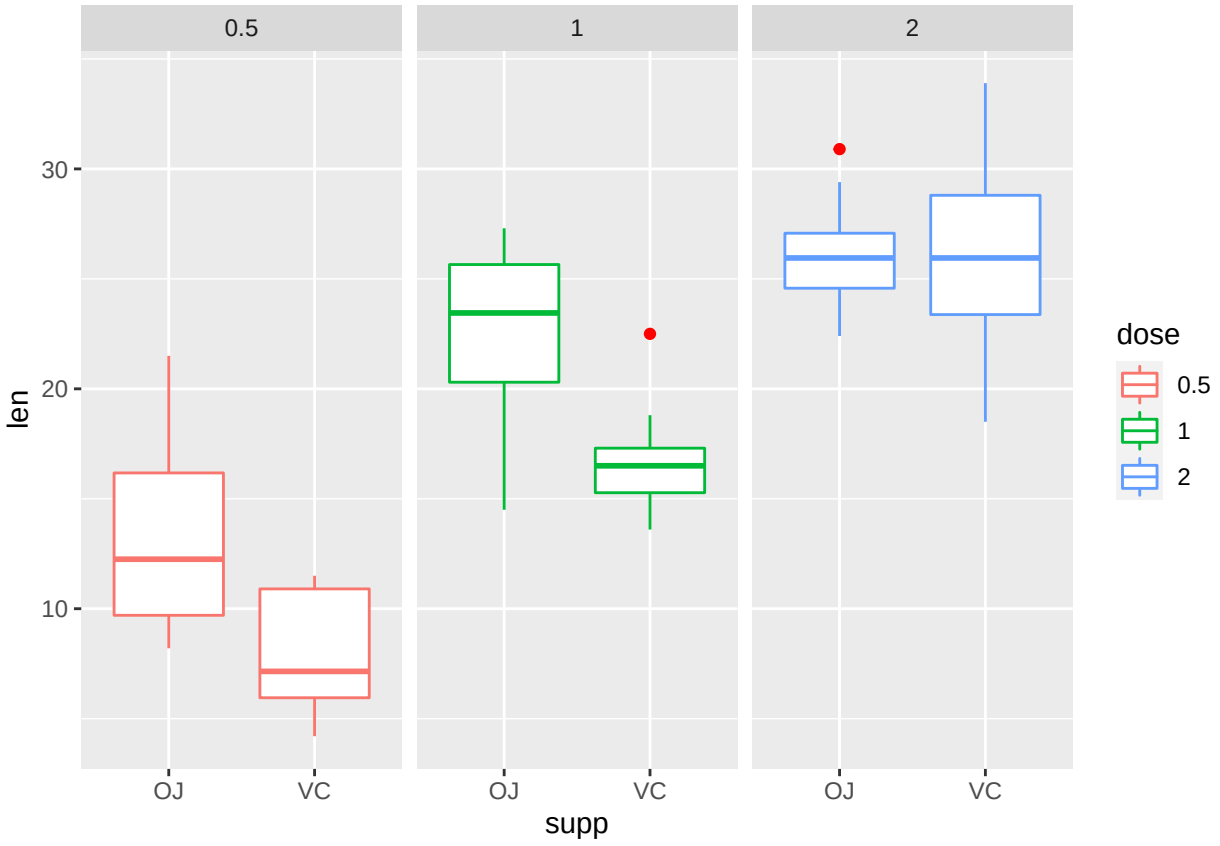
BoxPlot that shows relation between lenght and dose grouped by supplement type.

```
ggplot(ToothGrowth, aes(x = dose, y = len,color = dose)) +
 geom_boxplot(outlier.colour = "red") +
 facet_grid(.~supp)
```



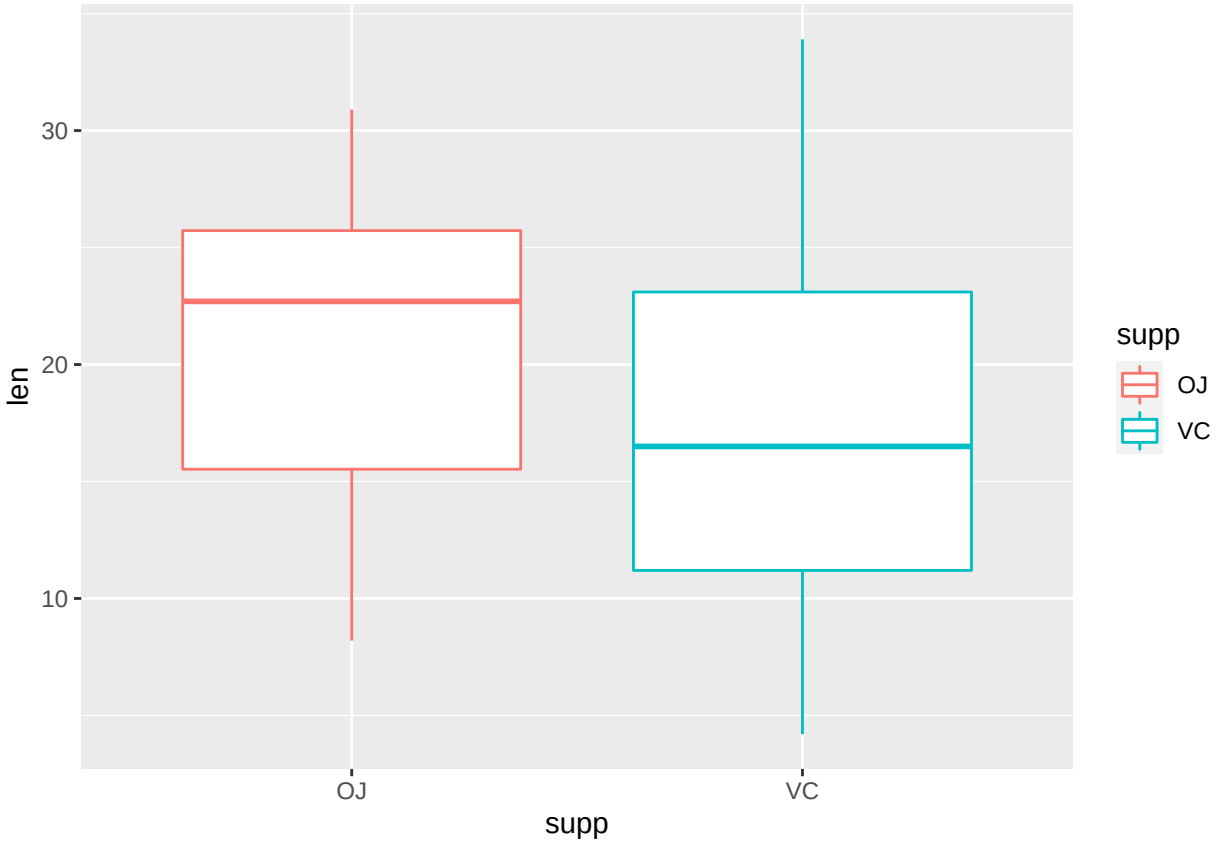
BoxPlot that shows relation between lenght and supplement type grouped by dose.

```
ggplot(ToothGrowth, aes(x = supp, y = len,color = dose)) +
 geom_boxplot(outlier.colour = "red") +
 facet_grid(.~dose)
```



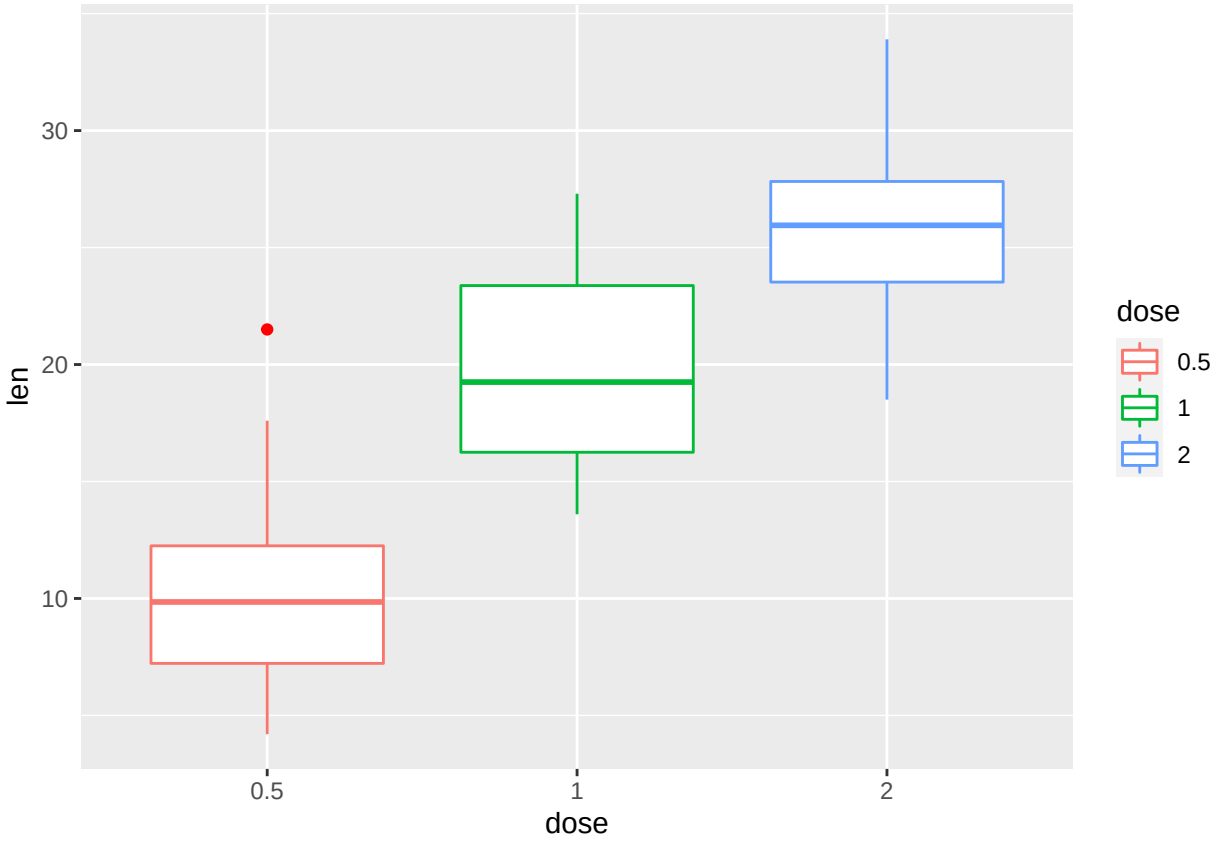
BoxPlot that shows relation between lenght by supplement type in general.

```
ggplot(ToothGrowth, aes(x = supp, y = len, color = supp)) +
 geom_boxplot(outlier.colour = "red")
```



BoxPlot that shows relation between lenght by dose in general.

```
ggplot(ToothGrowth, aes(x = dose, y = len, color = dose)) +
 geom_boxplot(outlier.colour = "red")
```



conclusion

It seams that there is strong correlation between tooth growth and dose and little or not difference between supplements.

Hypothesis testing

Three hypothesis are tested:

- Test if there is a difference in tooth growth effect between supplements.
- Test if there is a difference in tooth growth effect between dose 0.5 and 1.
- Test if there is a difference in tooth growth effect between dose 1 and 2.

```
test1 <- t.test(len ~ supp, data = ToothGrowth, var.equal = FALSE, paired=FALSE ,conf.level = .95)
pander(test1)
```

Table 1: Welch Two Sample t-test: `len` by `supp` (continued below)

Test statistic	df	P value	Alternative hypothesis	mean in group OJ
1.915	55.31	0.06063	two.sided	20.66

mean in group VC
16.96

```
test2 <- t.test(len ~ dose, data = subset(ToothGrowth, dose %in% c(0.5,1)), var.equal = FALSE, paired=F
pander(test2)
```

Table 3: Welch Two Sample t-test: `len` by `dose` (continued below)

Test statistic	df	P value	Alternative hypothesis
-6.477	37.99	1.268e-07 * * *	two.sided

mean in group 0.5	mean in group 1
10.61	19.73

```
test3 <- t.test(len ~ dose, data = subset(ToothGrowth, dose %in% c(1,2)), var.equal = FALSE, paired=FALS
pander(test3)
```

Table 5: Welch Two Sample t-test: `len` by `dose` (continued below)

Test statistic	df	P value	Alternative hypothesis
-4.9	37.1	1.906e-05 * * *	two.sided

mean in group 1	mean in group 2
19.73	26.1

conclusion

For first test p value is greater than significance level so null hypothesis cannot by rejected, which means that there is not difference between supplements in their effects in tooth growth. But second and third tests shows that supplement dose have different effects for tooth growth.