Statistical Inference - Basic Inferential Data Analysis

Abdul Rasheed Narejo

18/09/2018

```r
# load libraries
library(ggplot2) # for visualization
```

Basic inferential data analysis

In the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

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. (Only use the techniques from class, even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

About the data

The database contains 60 observations to check the impact on response variabl tooth growth (len variable) from Vitamin C delivered by one of two delivery methods - orange juice (OJ), ascorbic acid (CV). There are 30 observations for each method of delivery with 10 observations for each dosage (0.5, 1, 2).

Load data and perform some basic exploratory data analysis

```r
data("ToothGrowth")
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 ...
```

Provide basic summary of data

```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
```
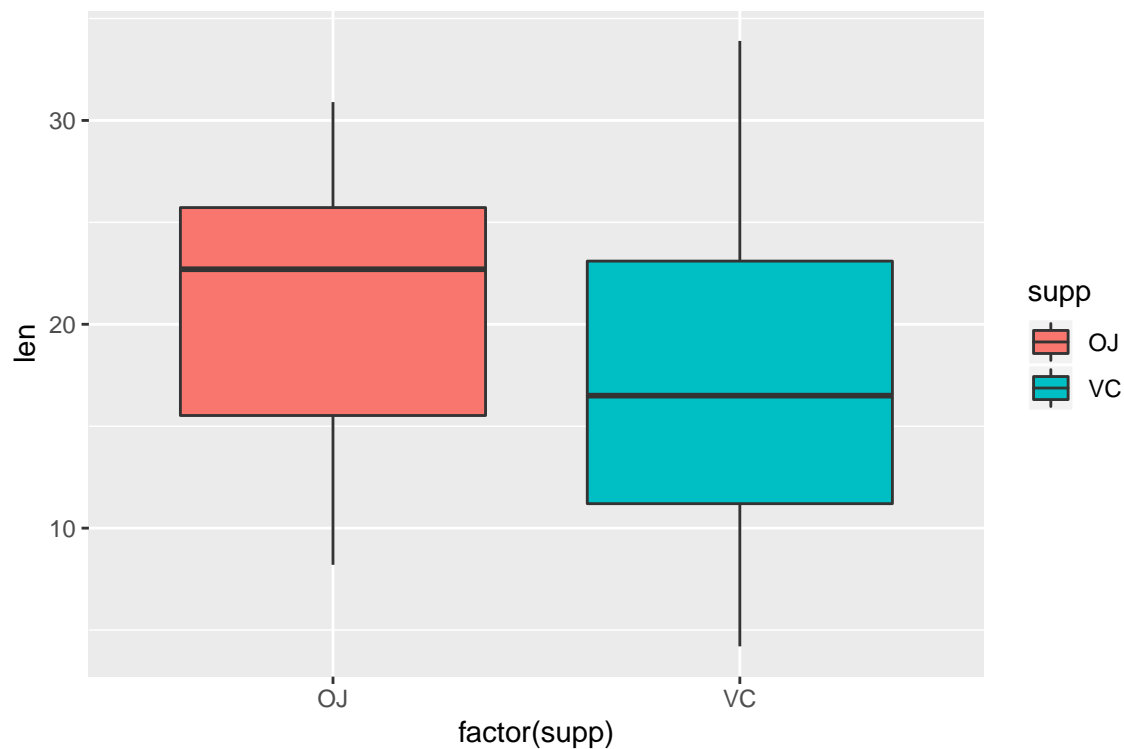
Hypothesis Testing

Hypothesis testing 1: Does supplement type affect tooth growth?

Null Hypothesis: Supplement type (OC, VJ) does not affect delivery method.
Alternative Hypothesis: Supplement type affects delivery method.

visualize data

```
ggplot(ToothGrowth, aes(x=factor(supp), y=len,fill=supp)) +
    geom_boxplot()
```


break data by supplement type

```
OJ <- subset(ToothGrowth, supp == "OJ", select = len)
VC <- subset(ToothGrowth, supp == "VC", select = len)
```

t.test

```
t <- t.test(OJ, VC, paired = FALSE)$p.value
t
```
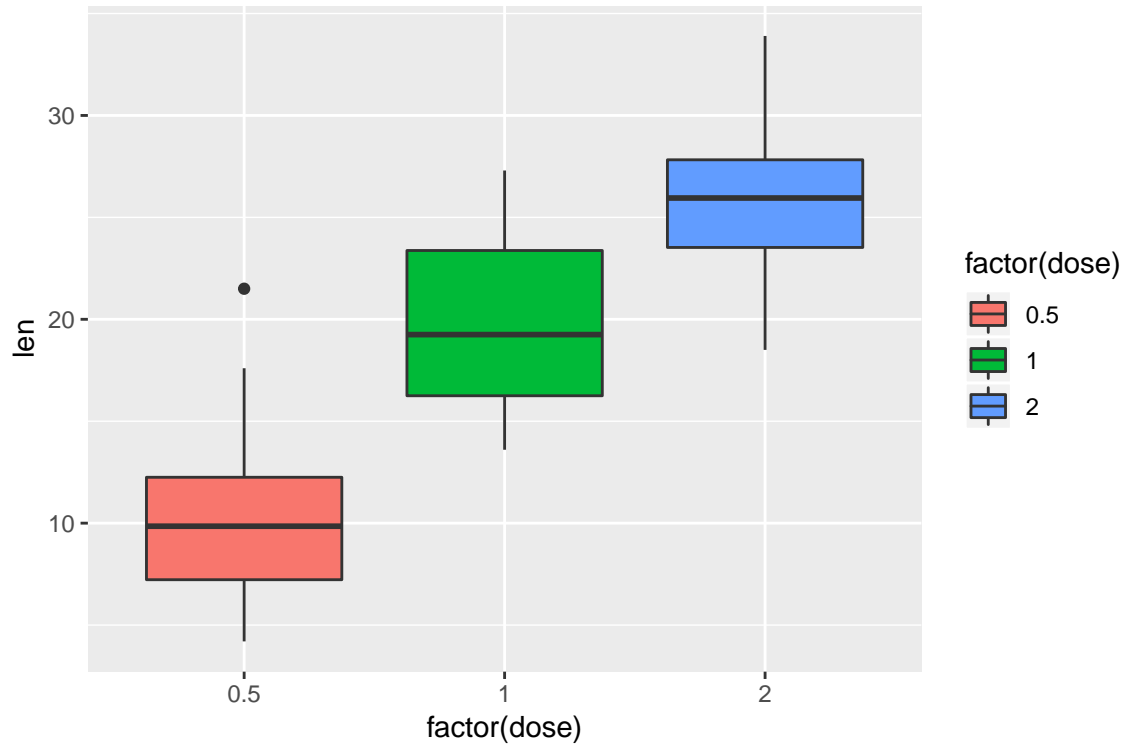
```
## [1] 0.06063451
```

Hypothesis testing 2: Does dose affect tooth growth?

Null Hypothesis: Supplement dose (0.5, 1, 2) doesnot affect tooth growth
Alternative Hypothesis: Supplement dose affects tooth growth

visualize data

```
ggplot(ToothGrowth, aes(x=factor(dose), y=len,fill=factor(dose))) +
    geom_boxplot()
```


break data by dose level

```
dose1 <- subset(ToothGrowth, dose == 0.5)$len
dose2 <- subset(ToothGrowth, dose == 1)$len
dose3 <- subset(ToothGrowth, dose == 2)$len
```

t.test

```
t1 <- t.test(dose1, dose2, paired = FALSE)$p.value
t2 <- t.test(dose1, dose3, paired = FALSE)$p.value
t3 <- t.test(dose2, dose3, paired = FALSE)$p.value
t1; t2; t3
```

```
## [1] 1.268301e-07
```

```
## [1] 4.397525e-14
```

```
## [1] 1.90643e-05
```

Conclusion

For hypothesis 1 - does delivery method (irrespective of dose) affect tooth growth ? we have failed to reject null hypothesis as p-values 0.0606345 is greater than 0.05. Hence, delivery method does not have significant impact on tooth growth.

For hypothesis 2 - does dose affect tooth growth ? we have conducted three separate t.tests to compare effect of different doses on tooth growth. Based on p-values, we reject null hypothesis. There is significant difference in tooth growth for various supplement doses.