Statistical Inference Part 2: Basic Inferential Data Analysis

JackHo

29 October 2016

Table of Contents

Before Cooking 1
Overview 1
Part 2: Basic Inferential Data Analysis Instructions 1
2.1 Basic EDA and Summary of the Toothgrowth 1
2.2 Basic Hypothesis/ Confidence Intervals testing 3
2.3 Summary 3

Before Cooking

To keep the article is shorter than 3 pages, some codes have been shaded by setting `echo=FALSE' in the chunk option.

```
# Loading necessary packages
require(ggplot2)
require(dplyr)
```

Overview

Analyze the ToothGrowth data in the R datasets package by applying Exporatory Data Analysis (EDA) and basic statistical tests on data. Try to find out the tooth growth effects about different supplement types at each daily dose level.

Part 2: Basic Inferential Data Analysis Instructions

The dataset `ToothGrowth` describes the Effect of Vitamin C on Tooth Growth in Guinea Pigs. For its variables, here are the description below:

- len: the numeric Tooth length
- supp: the factor Supplement type (VC or OJ)
- dose: the numeric Dose in milligrams/day

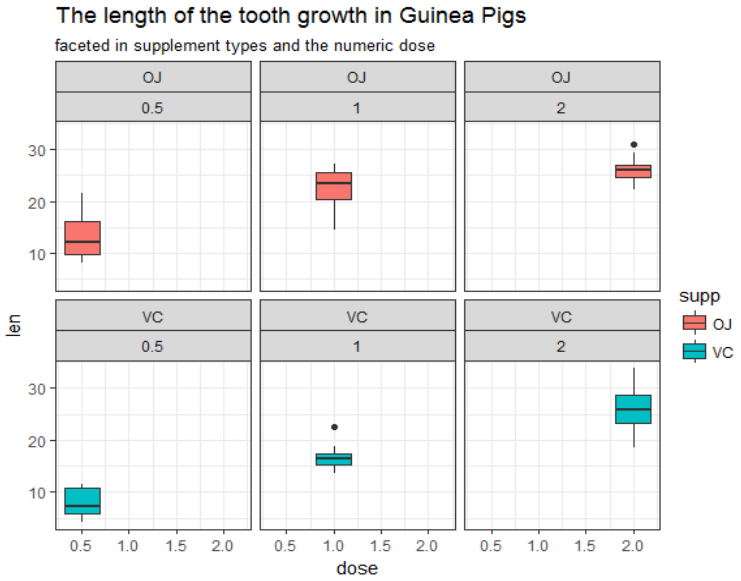
2.1 Basic EDA and Summary of the Toothgrowth

```
data(ToothGrowth)
dataTooth <- ToothGrowth
summary(dataTooth)
```

```
##       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
```

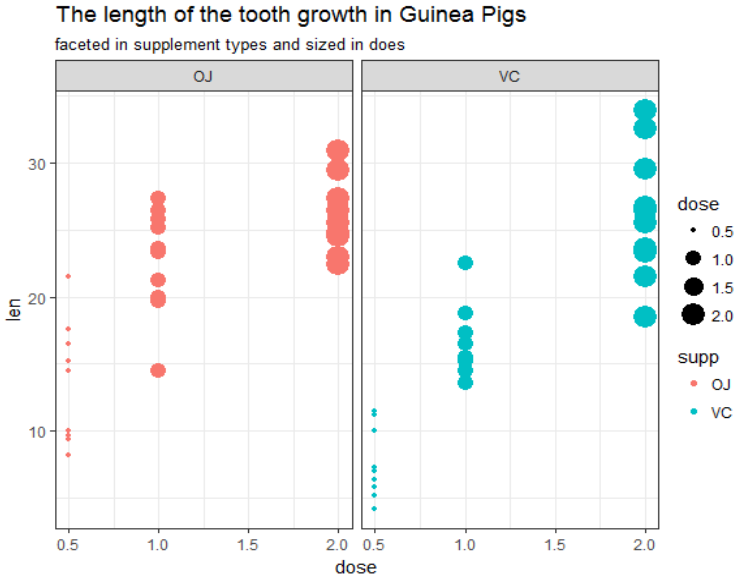
From the summary above, obviously, there are two groups of data in the dataset, the `OJ` group and the `VC` group. Each group has the `len` and `dose` variables, and these variables describes the effects of the OJ and VC on the tooth growth in Guinea Pigs.

```
ggplot(data = dataTooth, aes(y=len, x= dose,fill = supp))+ geom_boxplot()+facet_wrap(~supp+dose)+theme_bw() + labs(title="The length of the tooth growth in Guinea Pigs",subtitle = "faceted in supplement types and the numeric dose")
```



From the above boxplot, seperating the data into six groups will show that at the 0.5 dose level, the range of the length will be wider in OJ group than that in VC group, and so do the means and other quantile numbers. For at the 1 dose level and 2 dose level can be applied the similiar analysis pattern.

The codes of plotting the scatterplot below have been shaded.



From the scatterplot above, inside the OJ group, the larger dose applied, the larger length of the tooth growth will be, so do inside the VC group. Bewteen the OJ and VC groups, the data in the sub-groups (0.5 dose, 1 dose and 2 dose) has been analyzed in the boxplot above.

2.2 Basic Hypothesis/ Confidence Intervals testing

Based on the EDA, the data can be seperated into 3 groups:

Dose	OJ	VC
Dose(0.5)	OJ-0.5	VC-0.5
Dose(1.0)	OJ-1.0	VC-1.0
Dose(2.0)	OJ-2.0	VC-2.0

Assume the mean in OJ05 and VC05 are equal to each other, and OJ05 and VC05 are indenpendent groups, so use unpaired t- test. For OJ10- VC10 and OJ20- VC20, apply the same methods.

- Ho in OJ05- VC05 is 0 ==> Ho_05 = 0;
- Ho in OJ10- VC10 is 0 ==> Ho_10 = 0;
- Ho in OJ20- VC20 is 0 ==> Ho_20 = 0;

```
Logic_05 <- t.test(OJ05$len,VC05$len,paired = F,conf.level = .05)$p.value < .05
conf.int_05 <-t.test(OJ05$len,VC05$len,paired = F,conf.level = .05)$conf.int
print(sprintf("The logic value of 'p.value of < 0.05' is: %d (1 for TRUE, 0 for FALSE)
,and the confidence interval is from %f to %f",Logic_05,conf.int_05[1], conf.int_05[2]
))
```

```
## [1] "The logic value of 'p.value of < 0.05' is: 1 (1 for TRUE, 0 for FALSE) ,and the
 confidence interval is from 5.144384 to 5.355616"
```

The codes of OJ10- VC10 and OJ20- VC20 have been shaded.

```
## [1] "The logic value of 'p.value of < 0.05' is: 1 (1 for TRUE, 0 for FALSE) ,and the
 confidence interval is from 5.836274 to 6.023726"

## [1] "The logic value of 'p.value of < 0.05' is: 0 (1 for TRUE, 0 for FALSE) ,and the
 confidence interval is from -0.190693 to 0.030693"
```

2.3 Summary

From the retuslts of t-tests, we could concluded that:

- In Group OJ05- VC05 and OJ10- VC10, the population of OJ brings better effects to tooth growth than VC does.
    - In Group OJ05- VC05: Ho is rejected, because the p-value is smaller than 0.05, which means the means of population OJ and VC at .5 dose level are `different` with each other.
    - In Group OJ10- VC10: Ho is rejected, because the p-value is smaller than 0.05, which means the means of population OJ and VC at 1.0 dose level are `different` with each other.
- In Group OJ20- VC20, the population of OJ and VC have the similiar (not very different when alpha = 0.05) means.
    - In Group OJ20- VC20: Ho is accepted, because the p-value is smaller than 0.05, which means the means of population OJ and VC at 2.0 dose level are `not different` with each other.