

Statistical Inference Project_Part 1 Simulation Exercise

Luoning

5/3/2020

1.Synopsis

In the second portion of the project, the ToothGrowth data in the R datasets package was analyzed and the following questions were answered.

1. The ToothGrowth data was loaded and some basic exploratory data analyses were performed
2. A basic summary of the data was provided
3. The tooth growth by supp and doseUse was compared using confidence intervals and/or hypothesis tests

2.Data Analysis

2.1 Load Data

The toothgrowth data was loaded from the datasets.

```
data<-ToothGrowth
```

2.2 Data analysis

Look at the observations and variables of the data and the class of the variables.

```
str(data)
```

```
## '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 ...
```

Look at the first 6 lines of the data

```
head(data)
```

```
##      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
```

Look at the summary of the data

```
summary(data)
```

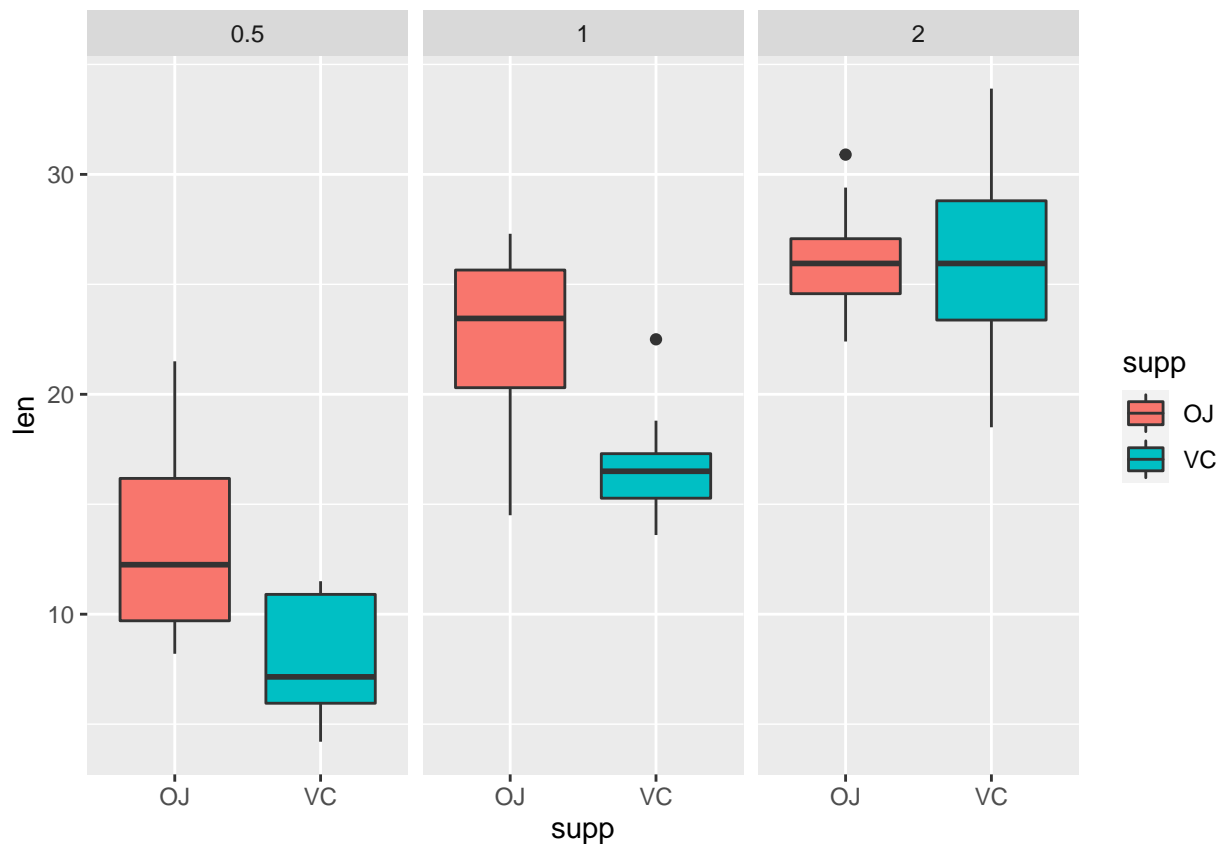
```
##      len      supp      dose
## Min.   : 4.20   OJ:30   Min.    :0.500
## 1st Qu.:13.07   VC:30   1st Qu.:0.500
```

```
## 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
```

This dataset has 60 observations and three variables, tooth length(len), delivery method (supp) and dose. The 60 observations of tooth growth of Guinea pigs shows the effect of dose of vitamin C on the tooth growth. The delivery method is a two-values factor variable : “OJ” (*Orange Juice*) and VC (*Ascorbi Acid*, a form of Vitamin C). The dose given contains three different values : 0.5, 1, and 2 mg/day. The measures are done on 10 Guinea pigs.

The influence of supply delivery and dose on tooth growth was plotted using botplot.

```
library(ggplot2)
data$dose<-as.factor(data$dose)
data$supp<-as.factor(data$supp)
p<-ggplot(data,aes(x=supp,y=len))
p<-p+geom_boxplot(aes(fill=supp))
p<-p+facet_wrap(~dose)
p
```



Intuitively from observation of the boxplots, the tooth length genergally increases with dose, regardless of the type of Vitamine C supply.

3. Compare tooth growth by supply and dose

3.1 Compare tooth growth with supply

T-test was applied for comparison with the two groups unpaired.

```
t.test(data$len[data$supp=='VC'],data$len[data$supp=='OJ'],paired=FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: data$len[data$supp == "VC"] and data$len[data$supp == "OJ"]
## 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:
## -7.5710156 0.1710156
## sample estimates:
## mean of x mean of y
## 16.96333 20.66333
```

From the t-test of tooth length and Vitamine C supply, we learn that 1. The p-value is 0.06, which is close to the significance level (0.05). So we do not reject the null hypothesis. 2. The confidence interval contains 0, so the test is not really significant. This means there is no significant difference between the *Ascorbic Acid* and *Orange Juice* with respect to the influence on tooth growth.

3.2 Compare tooth growth with dose

Now let's try to test the influence of the dose on the tooth growth, with two intervals, lower level (0.5-1) and higher level (1-2) and unpaired.

```
t.test(data$len[data$dose==1],data$len[data$dose==0.5],paired=FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: data$len[data$dose == 1] and data$len[data$dose == 0.5]
## 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:
## 6.276219 11.983781
## sample estimates:
## mean of x mean of y
## 19.735 10.605
```

```
t.test(data$len[data$dose==2],data$len[data$dose==1],paired=FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: data$len[data$dose == 2] and data$len[data$dose == 1]
## 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:
## 3.733519 8.996481
## sample estimates:
## mean of x mean of y
## 26.100 19.735
```

From the t-test of tooth length and supply dose, we learn that:

1. The p-value is very small, nearly 0, which means either the sample is an extreme or the null hypothesis (change in dose has no effect on tooth growth) is false, and the latter makes more sense.

2. the confidence interval of both tests on low and high level of doses are significantly positive. This means that regardless *Ascorbic Acid* or *Orange Juice*, more doses of Vitamine C can increase the rate of tooth growth.

4. Conclusions

Vitamine C has positive effect on the tooth growth of Guinea pigs. Both *Ascorbic Acid* and *Orange Juice* are supply of Vitamine C and show insignificant difference in the effect.