Statistical Inference Project Part 2

Wednesday, April 22, 2015

Synopsis

This is the report of the second part of Peer assignment of Statistical inference of coursera. The ToothGrowth dataset is the database collected based on the the Effect of Vitamin C on Tooth Growth in Guinea Pigs. As described in R documentation, this ToothGrowth dataset recorded the response of the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

Result

1. Load the ToothGrowth data and perform some basic exploratory data analyses.

Answer :

```
# load the neccessary libraries and dataset.
library(ggplot2)
library(datasets)
data(ToothGrowth)

# Take a glance at the loaded ToothGrowth dataset.
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 ...
```

```
# Verify the number of row as mentioned in R documentation.
nrow(ToothGrowth)
```

```
## [1] 60
```

```
# Processing the variable dose of the dataset to factor.
ToothGrowth$dose <- as.factor(ToothGrowth$dose)

# re-look the processed dataset.
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 ...
```

2. Show a basic summary of the data.

Answer :

```
# Show the summary statistics for the processed dataset.
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
```

There are 60 samples in this dataset and it consists of **len**, **supp** and **dose**. The variable **len** is the length of odontoblasts (teeth) in each of 10 guinea pigs. **supp** are the two delivery methods: orange juice (OJ) or ascorbic acid (VC). **dose** is the three dose levels of Vitamin C (0.5, 1, and 2 mg) conducted for this test.
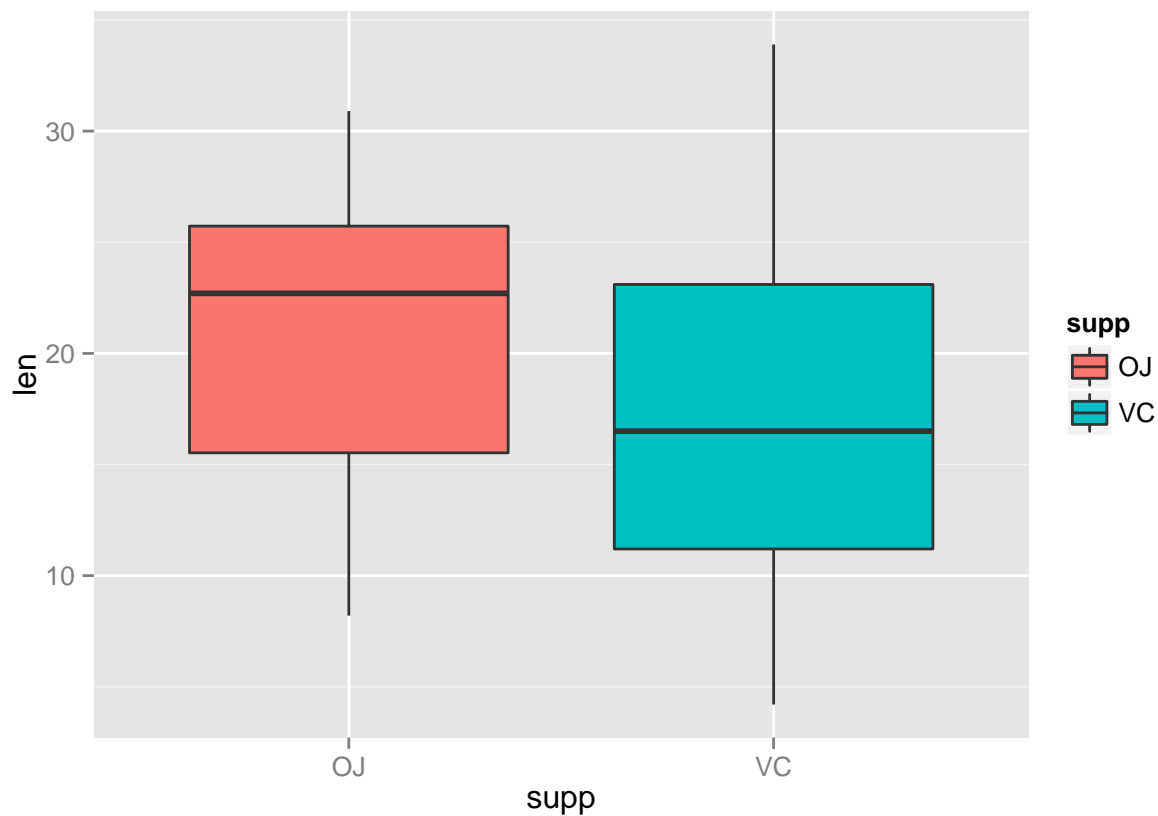
```
# To use hypothesis tests to compare tooth growth by supp and dose, firstly split the cases into differ
table(ToothGrowth$dose, ToothGrowth$supp)
```

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

```
# Display the tooth growth in delivery methods (supplement type).
ggplot(aes(x=dose, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=dose))
```
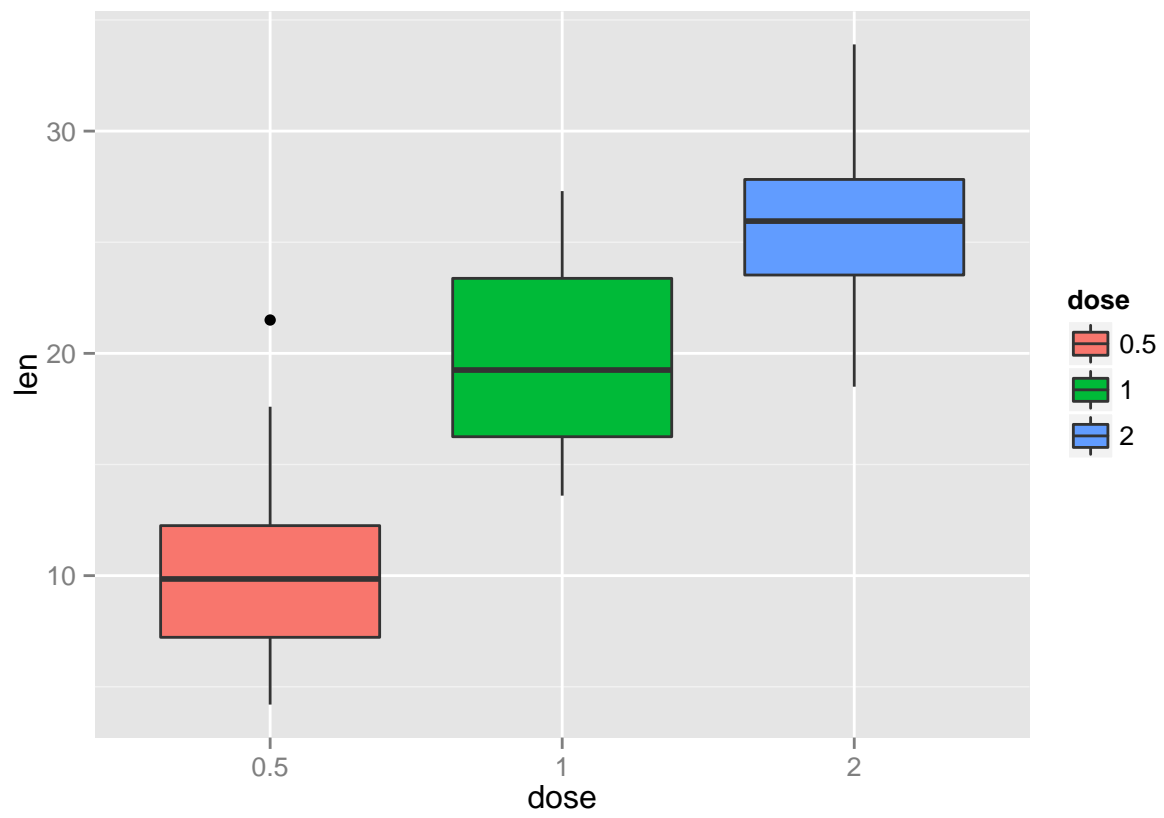
```
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp))
```

```
# Display the tooth growth in dose levels of Vitamin C.
#ggplot(aes(x=dose, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp))
```

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

Answer :

```
# Firstly, assume that there is a unequal variance between the two group of testing, supp and dose, in
#t.test(len ~ supp, data = ToothGrowth)
len  <- ToothGrowth$len
supp <- ToothGrowth$supp
dose <- ToothGrowth$dose

t.test(len[supp=="OJ"], len[supp=="VC"], paired = FALSE, var.equal = FALSE)
```

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

Based on the computed result, the p-value of this test is 0.0603 at a significance level of 5%. This data gives us less evidence to neglect null hypothesis and to conclude that the effectiveness of the delivery of the vitamin on the tooth length.The confidence interval of this test is (-0.1710156 7.5710156). Obviously, one of the reading shows that the confidence interval contains zero.

Next, steps required to find out the effect on the mean of the tooth length when the test conducted on the vitamin C dosage of 2mg and 1mg.

Before the tests conduct, let the Ho be the Null hypotheses of equal means of the the two groups, versus the alternative hypothesis that the two means are different. Hence a two sided test can be carried out.

```
t.test(len[dose == 1], len[dose == 2], paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  len[dose == 1] and len[dose == 2]
## t = -4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.994387 -3.735613
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

The obtained result shows that the p-value of this test is very close to 0 (1.811e-05) at a significance level of 5%. With this informative data, sufficient evidence allows us to reject th null hypothesis, consequently, it is concluded that the means of dosage change from 1mg to 2mg lead to tooth length increses.

The confidence interval of the test is (-8.994387 -3.735613) and it gives direct information of no zero in confidence interval.

Test will continue on the mean tooth length of the group with vitamin C dose of 1mg versus the vitamin C dose of 0.5mg.

```
t.test(len[dose == 0.5], len[dose == 1], paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  len[dose == 0.5] and len[dose == 1]
## t = -6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983748  -6.276252
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

The obtained result shows that the p-value of this test is very close to 0 (1.266e-07) at a significance level of 5%. Similar to previous test, there is sufficient evidence allows us to reject th null hypothesis, consequently, it is concluded that the means of dosage change from 0.5mg to 1mg lead to tooth length increses.

The confidence interval of the test is (-11.983748 -6.276252) and it gives direct information of no zero in confidence interval.

Find out the 4 different pvalues for 4 tests.

```
# Store pvalues in a vector.
pValues <- c(t.test(len[supp=="OJ"], len[supp=="VC"], paired = FALSE, var.equal = FALSE)$p.value,
             t.test(len[dose==2], len[dose==1], paired = FALSE, var.equal = TRUE)$p.value,
             t.test(len[dose==1], len[dose==0.5], paired = FALSE, var.equal = TRUE)$p.value,
             t.test(len[dose==2], len[dose==0.5], paired = FALSE, var.equal = TRUE)$p.value)

# Manipulates Family-wise error rate (FWER)
sum(p.adjust(pValues, method = "bonferroni") < 0.05)
```

```
## [1] 3
```

```
round(p.adjust(pValues, method = "bonferroni"),3)
```

```
## [1] 0.243 0.000 0.000 0.000
```

```
# Manipulates False discovery rate (FDR)
sum(p.adjust(pValues, method = "BH") < 0.05)
```

```
## [1] 3
```

```
round(p.adjust(pValues, method = "BH"),3)
```

```
## [1] 0.061 0.000 0.000 0.000
```

The test result of pvalues test shows the value of 1st pvalues test is the biggest and it is considered sufficient evident to accept the Null hypothesis.

4. Conclusion.

- Increase the dose level from 0.5mg to 1mg or from 1mg to 2mg lead to increased tooth growth.
- Supplement type, either orange juice (OJ) or ascorbic acid (VC), brings no effect on tooth growth.

5. Assumption for the Conclusion.

- The sample of 60 Guinea pigs is assumed to be the representative of the majority so that the conclusion can be considered valid for this population.
- This experiment is conducted by assuming that each Guinea pig is randomly pick for the tests.