ToothGrowth data analysis

Yinyan Guo

October 9, 2015

Overview

This is a report (part II) for "Statistical Inference Course Project. hypothesis tests on ToothGrowth data (len in each supp, and len in each dose).
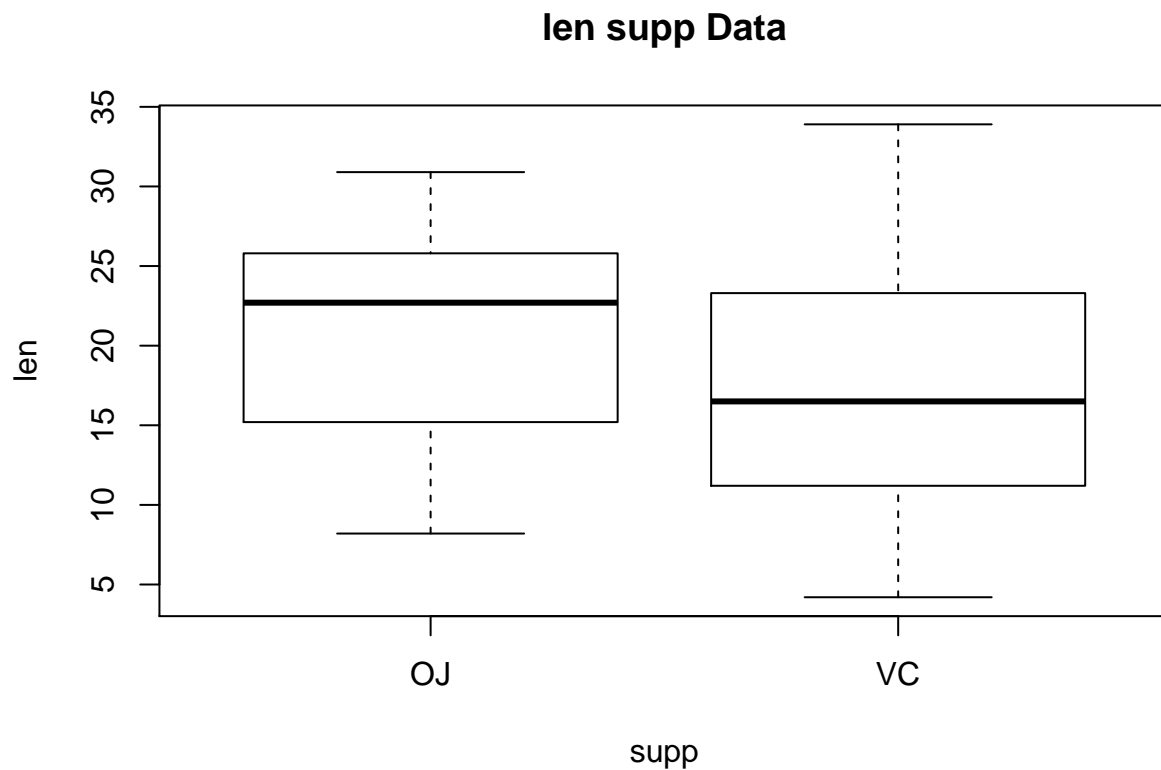
Get data and summerize data

```
library(datasets)
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
```

explore tooth growth by supp

```
boxplot(len~supp,data=ToothGrowth, main="len supp Data",
        xlab="supp", ylab="len")
```

len supp Data

Hypothesis tests and difference interval

H0: Mu(OJ)=Mu(VC) Ha: Mu(OJ)<>Mu(VC), alpha=0.05

```
## assume equal variance
OJ<- ToothGrowth[ToothGrowth$supp=="OJ",1]
VC<- ToothGrowth[ToothGrowth$supp=="VC",1]
t.test(OJ,VC, var.equal=TRUE, paired=FALSE)
```

```
##
##  Two Sample t-test
##
## data:  OJ and VC
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

```
## assume variance are not equal
t.test(ToothGrowth$len~ToothGrowth$supp)
```

```
## 
##  Welch Two Sample t-test
## 
## data:  ToothGrowth$len by ToothGrowth$supp
## 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 in group OJ mean in group VC
##         20.66333         16.96333
```

conclusion:

Tooth len is significant different between supp(OJ) and sup(VC) with given alpha.

explore tooth growth by dose

```
boxplot(len~dose,data=ToothGrowth, main="len dose Data",
        xlab="dose", ylab="len")
```
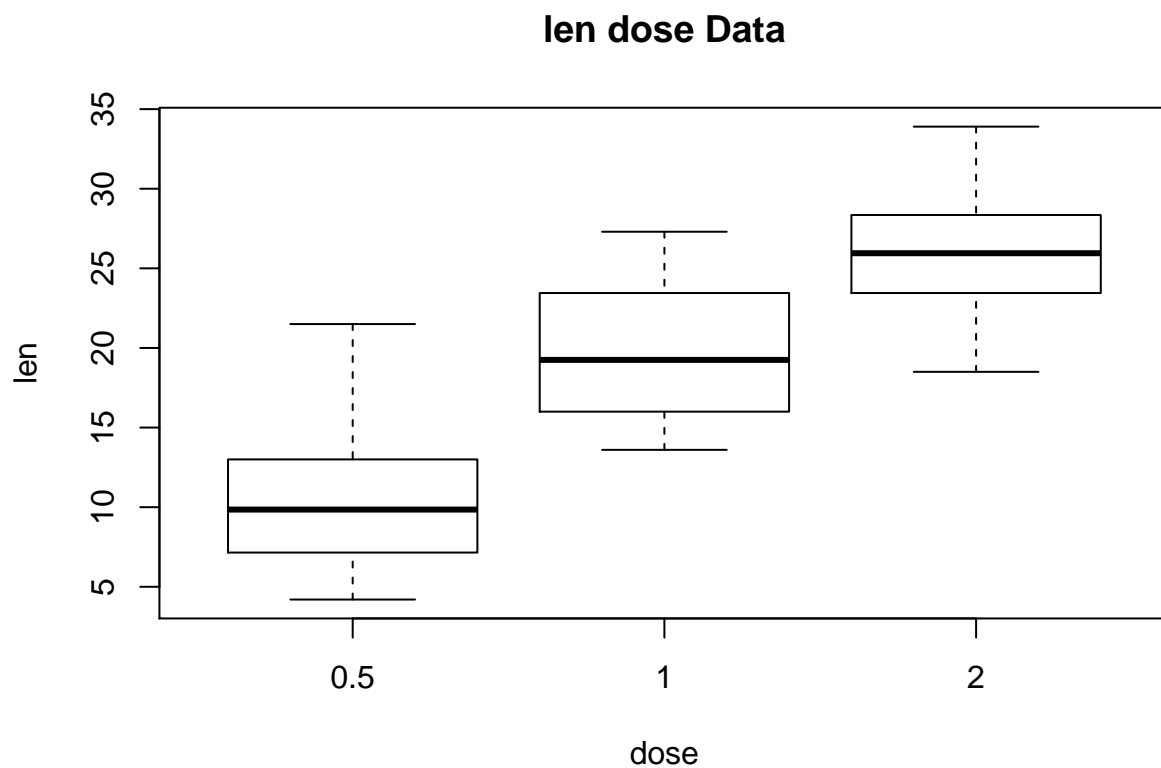

len dose Data

```
tapply(ToothGrowth$len, ToothGrowth$dose, mean)
```

```
##    0.5      1      2
## 10.605 19.735 26.100
```

ANOVA test since there are > 2 levels in dose

```
Tooth.df =ToothGrowth
Tooth.df$Dosegroup = factor(Tooth.df$dose)
Tooth.mod1 = lm(len ~ Dosegroup, data = Tooth.df)
anova(Tooth.mod1)
```

```
## Analysis of Variance Table
##
## Response: len
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Dosegroup  2 2426.4  1213.2  67.416 9.533e-16 ***
## Residuals 57 1025.8    18.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(Tooth.mod1)
```

```
##
## Call:
## lm(formula = len ~ Dosegroup, data = Tooth.df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.6000 -3.2350 -0.6025  3.3250 10.8950
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.6050     0.9486  11.180 5.39e-16 ***
## Dosegroup1    9.1300     1.3415   6.806 6.70e-09 ***
## Dosegroup2   15.4950     1.3415  11.551  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.242 on 57 degrees of freedom
## Multiple R-squared:  0.7029, Adjusted R-squared:  0.6924
## F-statistic: 67.42 on 2 and 57 DF,  p-value: 9.533e-16
```

Conclusion1:

- ANOVA comparison results showed at least two doses their tooth len are different

- In 3 doses comparison. Dose 0.5 is significant different with the other 2 doses (in default, lm uses level1 dose=0.5 as baseline)

```
## compare  dose(1) vs dose (2)
mat<- cbind(rep(1/3, 3), " 1 vs 2"=c(0,1,-1)," 0.5 vs 1"=c(-1,1,0))
mymat <- solve(t(mat))
my.contrast <- mymat[,2:3]
contrasts(Tooth.df$Dosegroup) <- my.contrast
summary.lm(aov(len ~ Dosegroup, data = Tooth.df))
```

```
##
## Call:
## aov(formula = len ~ Dosegroup, data = Tooth.df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.6000 -3.2350 -0.6025  3.3250 10.8950
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        18.8133     0.5477  34.352  < 2e-16 ***
## Dosegroup 1 vs 2   -6.3650     1.3415  -4.745 1.44e-05 ***
## Dosegroup 0.5 vs 1  9.1300     1.3415   6.806 6.70e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.242 on 57 degrees of freedom
## Multiple R-squared:  0.7029, Adjusted R-squared:  0.6924
## F-statistic: 67.42 on 2 and 57 DF,  p-value: 9.533e-16
```

Conclusion2:

- Comparison results showed there is significant different in tooth len in dose (1) with dose(2).

Final conclusion

There are significant difference among three doses on tooth len.

Notes: the significant tests above is based on the assumption that the len data are normal distribution, and variance in three doses are the same.