STAT537: Statistics for Research I: HW#8

Due on Sep. 20, 2016

Dr. Schmidhammer TR 11:10am – 12:25pm

Wenqiang Feng

Contents

Problem 1

Homework on Correlation Coefficients

Solution. 1. **Perform an analysis of variance on these data, and test the hypothesis:** From the summary of the ANOVA test,we get the p-vaule is $5.85e-05 < 0.05$, hence reject H_0. Therefore, there is at least one of the means is different from the others.

```
            Df Sum Sq Mean Sq F value   Pr(>F)
group        3  6.621  2.2070   11.05 5.85e-05 ***
Residuals   28  5.594  0.1998
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2. **Use Levene's Test to test the assumption that the variances in the 4 groups are equal. State your conclusions.** From the following Levene's Test, we can see that the p-vaule is $0.6428 > 0.05$, hence there is no enough information to reject H_0. Therefore, the assumption that the variances in the 4 groups are equal is valid.

```
> leveneTest(yields~group, data)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  0.5647 0.6428
      28
```

3. **Check to see if the residuals are modeled well by a normal distribution. State your conclusions.** From the Normal QQ plot of the residuals in Figure.1, we can conclude that residuals are modeled well by a normal distribution.

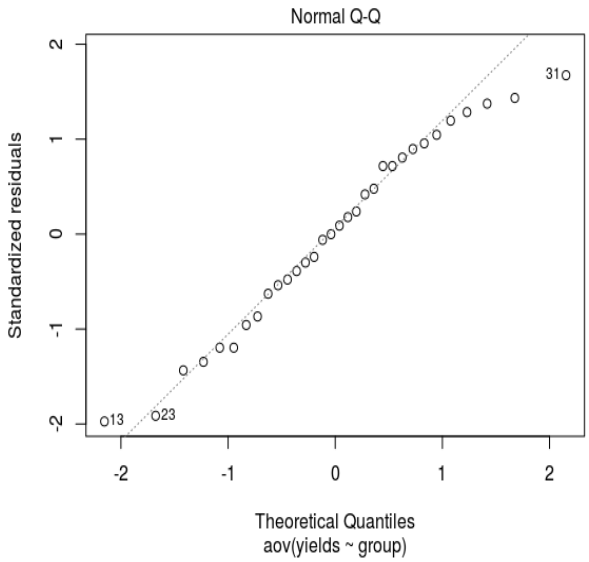

Figure 1: Normal QQ plot of the residuals.

4. **If you determine that the means are not equal in part (1), use Tukey's HSD procedure to determine which means are different. State your conclusions.** From the p_{adj} value we can see that the means of B-A, C-A, D-B and D-C are significantly different from each other at 95% level. Moreover, we can see the difference from the figure.2

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = yields ~ group, data = data)

$group
       diff         lwr        upr     p adj
B-A  0.6625  0.05232456  1.2726754 0.0294781
C-A  0.9375  0.32732456  1.5476754 0.0013447
D-A -0.1625 -0.77267544  0.4476754 0.8854051
C-B  0.2750 -0.33517544  0.8851754 0.6132192
D-B -0.8250 -1.43517544 -0.2148246 0.0049860
D-C -1.1000 -1.71017544 -0.4898246 0.0001916
```

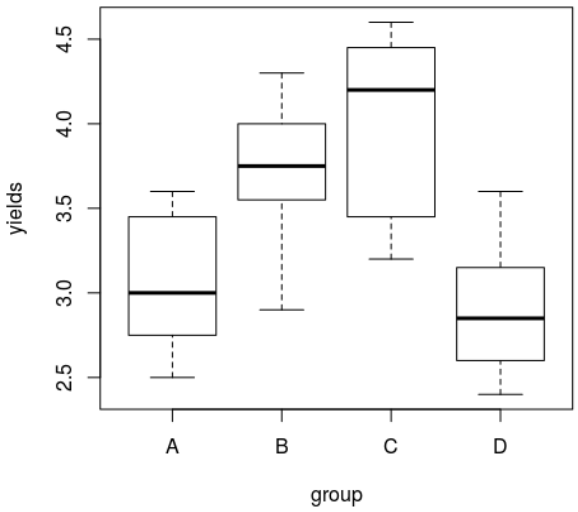

Figure 2: boxplots.

5. **To determine whether the distributions are the same in the four groups, perform a Kruskal-Wallis test on the ranks. State your conclusions.** The null hypothesis is that the distributions are the same in the four groups. From the summary of the Kruskal-Wallis rank sum test, we can see that the p-value is $0.0008698 < 0.05$, hence reject H_0. Therefore, at least two of these four groups have different distributions.

```
> kruskal.test(yields~group, data)
```

```
    Kruskal-Wallis rank sum test

    data:  yields by group
    Kruskal-Wallis chi-squared = 16.561, df = 3, p-value = 0.0008698
```

□

Appendix

R code for HW#8

Listing 1: Source code for problem 1

```
# reference: http://www.stat.columbia.edu/~martin/W2024/R3.pdf
rm(list = ls())
# set the path or enverionment
setwd("/home/feng/Dropbox/UTK_Course/Stat537/Excel/CH08")


#install.packages("readxl") # CRAN version
library(readxl)
#install.packages("moments")
library(moments)
rawdata = read_excel("ex8-32.xls",sheet = 1)
#attach(rawdata)

numData =as.matrix(rawdata)
data = data.frame(yields=c(A,B,C,D), group=rep(LETTERS[1:4], each=length(A)))

# problem 1
onewayAOV=aov(yields~group, data)
summary(onewayAOV)
#
library(car)
leveneTest(yields~group, data)
#
plot(onewayAOV, which=2)

#
TukeyHSD(onewayAOV,conf.level=0.95)
plot(onewayAOV)
plot(yields~group, data)
```