STAT 371 Discussion 11

Diagnostics of ANOVA Model and Multiple Comparison

(Review of lecture examples and R function learning)

1. Steps of Working on ANOVA Model

1.1 Load data into R

1) Create your dataset using notepad(save as .txt) or excel(save as .csv)
2) Set the working directory using **setwd()**
3) Import data using **read.table()** for .txt file and **read.csv()** for .csv file

1.2 Fit an ANOVA model use **aov(y~x, data=)** function

1) y is the response. For example, you want to compare the mean weight of several groups of rats, then the weight is y.
2) x is the variable indicating groups

1.3 Checking the model assumptions

1) Underlying populations normally distributed: QQ plot of residuals
2) Underlying populations have equal variance: residual vs. fitted
3) Observations are independent: residuals against index of observation or time order of taking observations and so on.

1.4 Transformation

1) Log transformation
   When the plot of residuals vs. fitted values show a right-open-mouth horn shape. Use **log()** in R.
2) Square transformation
   When data are counts. Use is **sqrt()** in R.

1.5 Multiple comparisons

1) Unplanned comparisons and planned comparisons
2) Adjustment of significance level for multiple comparisons – Bonferroni adjustment and Tukey adjustment

2. Example for model assumption checking

2.1 Load data into R

il10 <- read.csv("http://www.biostat.wisc.edu/~kbroman/teaching/stat371/il10.csv")

The first 15 observations of dataset il10:

```
Strain  IL10
     7   985.425
     7  1057.254
     7    989.48
     7   816.093
     7  1410.767
    25  1159.141
    25   453.573
    25   548.546
    25    422.12
    25    178.64
```

2.2 Fit an ANOVA model use **aov(y~x, data=)** function

1) Fit model: **aov.il10 <- aov(IL10 ~ Strain, data=il10)**
2) Use anova() to get the ANOVA table: **anova(aov.il10)**

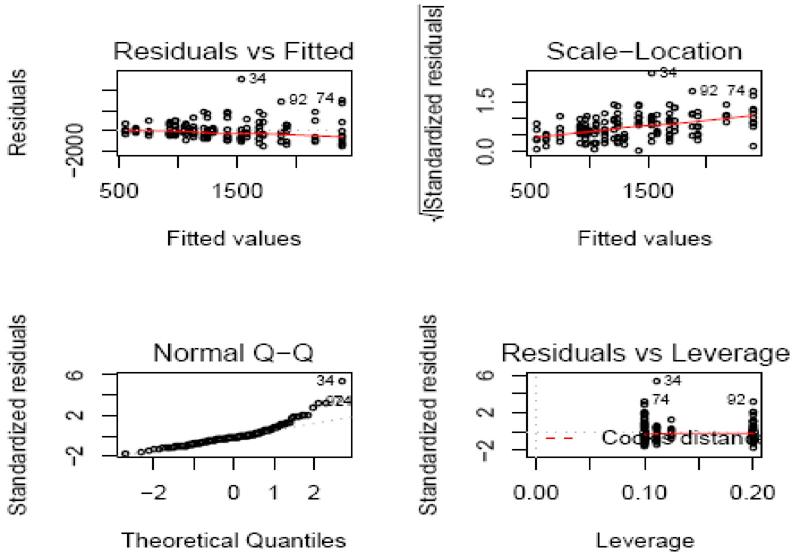
```
Response: IL10
           Df    Sum Sq   Mean Sq F value  Pr(>F)
Strain     20  33757424   1687871  1.7001 0.04154 *
Residuals 125 124099169    992793
```

2.3 Checking the model assumptions

1)Underlying populations normally distributed: QQ plot

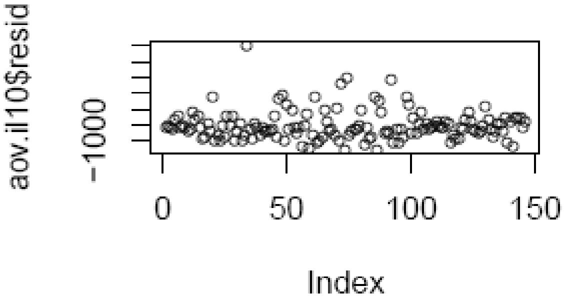
par(mfcol=c(2,2))   # specify the layout of plots
plot(aov.il10, cex=0.7)

Four plots, use the ‘Normal Q-Q’ plot.



2) Underlying populations have equal variance: residual plots
Use plot ‘Residuals vs Fitted’ given in 1).

3) Observations are independent within and between groups
Residuals against index of observation: **plot(aov.il10$resid)**



The residuals are evenly distributed, not correlated, thus independent.

2.4 Transformation

Right-open-mouth horn shape: log transformation

2.5 Refit the ANOVA model using log(IL10) instead of IL10 for response

```
aov.logil10 <- aov(log(IL10) ~ Strain, data=il10)
```

2.6 Check model assumptions for model aov.logil10

3. Example for Multiple Comparisons (Unplanned)

3.1 Use dataset in last Friday’s lecture (better use excel to create a dataset)

```
rsp <- c(75,67,70,75,65,71,67,67,76,68,
       57,58,60,59,62,60,60,57,59,61,
       58,61,56,58,57,56,61,60,57,58,
       58,59,58,61,57,56,58,57,57,59,
       62,66,65,63,64,62,65,65,62,67)
ttt <- factor(rep(c("C","G","F","G+F","S"),rep(10,5)))
sugar <- data.frame(rsp, ttt)
```

3.2 Fit ANOVA model

```
aov.out <- aov(rsp ~ ttt, data=sugar)
anova.out <- anova(aov.out)
```

3.2 Unplanned pair wise comparisons, compare mean response of each pair of groups

1) Bonferroni adjustment

```
source("http://www.biostat.wisc.edu/~kbroman/teaching/stat371/func29.R")
sugar.bonf <- ci.bonf(sugar$rsp, sugar$ttt)
sugar.bonf
```

```
               est    lower    upper
     C - F   11.9  9.09056 14.70944
     C - G   10.8  7.99056 13.60944
     C - G+F 12.1  9.29056 14.90944
     C - S    6.0  3.19056  8.80944
     F - G   -1.1 -3.90944  1.70944
     F - G+F  0.2 -2.60944  3.00944
     F - S   -5.9 -8.70944 -3.09056
     G - G+F  1.3 -1.50944  4.10944
     G - S   -4.8 -7.60944 -1.99056
     G+F - S -6.1 -8.90944 -3.29056
```

No significant difference if confidence interval covers 0.

2) Tukey adjustment

```
sugar.tuk <- TukeyHSD(aov.out)
sugar.tuk$ttt
```

```
       diff        lwr       upr        p adj
F-C   -11.9 -14.868072 -8.931928 3.370637e-13
G-C   -10.8 -13.768072 -7.831928 2.079004e-12
G+F-C -12.1 -15.068072 -9.131928 3.010925e-13
S-C    -6.0  -8.968072 -3.031928 7.223105e-06
G-F     1.1  -1.868072  4.068072 8.291029e-01
G+F-F  -0.2  -3.168072  2.768072 9.996878e-01
S-F     5.9   2.931928  8.868072 9.983469e-06
G+F-G  -1.3  -4.268072  1.668072 7.256157e-01
S-G     4.8   1.831928  7.768072 3.242398e-04
S-G+F   6.1   3.131928  9.068072 5.222269e-06
```