Statistics 251: Lab 8 Pre-Reading Document 1

Analysis of Variance (ANOVA)

Objectives:

- Understand
  1. Analysis of Variance - ANOVA
  2. Pairwise Comparison of Means
- Prepare for lab 8 by trying a pre-lab exercise

Introduction:

Analysis of Variance (ANOVA) is a statistical method that tests the equality of three or more population means by analyzing sample variances or variation in the data.

1.0 Analysis of Variance - ANOVA

Pre-reading document 2 (pages 1-3) contains a detailed description of *analysis of variance,* an example and R commands.

2.0 Pairwise Comparison of Means

Pre-reading document 2 (pages 4-6) contains a detailed description of *pairwise comparison of means,* an example and R commands.

3.0 R Commands for Analysis of Variance

3.1 R Commands for Pre-Lab Exercise in Pre-Reading Document 2

Note 1: R might give an error if you copy and paste the commands that include quotation marks (‘). You can either type these commands directly in the R console or after you copy and paste commands in the R console, delete the quotation marks and type them again.

Note 2: It is best to briefly examine the raw data by printing it out in the R console. Using a graphical method to visually compare the data will give you a preliminary idea about the similarities and differences.

Note 3: The *HockeyData.txt* file contains a random sample of hockey players for each team. Assume the population variances are equal for all three teams.

```
## First download the data from Connect to your computer. Save the data file in your working directory.
## For example, save it to '~/Stat251/lab8'.

hockey = read.table('~/Stat251/lab8/HockeyData.txt', head = T)
attach(hockey) # attach data file so that you can call them by vector names
names(hockey) # see the column vector names
hockey   # print the raw data (examine briefly)
boxplot(height ~ team, xlab = 'Team', ylab = 'Height', main = 'Heights of the Players') #Examine/compare data
model = lm(height ~ team) # fit linear model
fit = aov(model) # ANOVA
summary(fit) # summary ANOVA table

u1 = mean(height[which(team == 'Can')]) #mean of height for just 'can' group
u2 = mean(height[which(team == 'Oil')]) #mean of height for just 'oil' group
u3 = mean(height[which(team == 'Fla')]) #mean of height for just 'fla' group
n1 = length(which(team == 'Can')) #size of 'can' group
n2 = length(which(team == 'Oil')) #size of 'oil' group
n3 = length(which(team == 'Fla')) #size of 'fla' group
mse <- sum(fit$residuals^2)/fit$df.residual #MSE
se12 <- sqrt(mse)*sqrt((n1+n1)/(n1*n2)) #SE for can & oil
se13 <- sqrt(mse)*sqrt((n1+n3)/(n1*n3)) #SE for can & fla
se23 <- sqrt(mse)*sqrt((n2+n3)/(n2*n3)) #SE for oil & fla
k<-length(unique(team))  #number of groups
C<- choose(k,2)   #compute kC2
t <- qt((1-(0.05/(2*C))),(length(team)-k)) # t-statistic
c(u1-u2-t*se12,u1-u2+t*se12) #CI for can &  oil
c(u1-u3-t*se13,u1-u3+t*se13) # CI  for can  &  fla
c(u2-u3-t*se23,u2-u3+t*se23) # CI  for oil  &  fla
```

3.2 Explanation of R Commands used in the Pre-Lab Exercise

Please see **Pre-Reading Document 2** for a detailed explanation of R commands used above.

3.3 Food for thought

Try the following:

1. Write the appropriate hypotheses to test whether the players in the three teams have the same population mean weight.
2. Use an appropriate plot to visually compare the weights of the three teams. Is it plausible that the population variances are the same?
3. Use analysis of variance techniques to test whether the population mean weights of the three teams are statistically significantly different.
4. If the ANOVA test is significant, i.e. the data provide sufficient evidence to conclude that a difference exists in the players' weights among the three teams, do a Bonferroni pairwise comparison.