Two Way Anova in R

Jeevan Koshy - 1740256

June 25,2019

**Question 1**

In milk filling plant, four machines are filling milk in the cans to a level of 12.0 liters;

random samples of the cans from each of the four machines were taken. Compare the

4 different machines with respect to mean level of filling.

**Aim**

H0: There is no significant difference between the mean level of filling in the four machines.

H1: There is a significant difference between the mean level of filling in the four machines.

Alpha level=0.01

**Procedure**

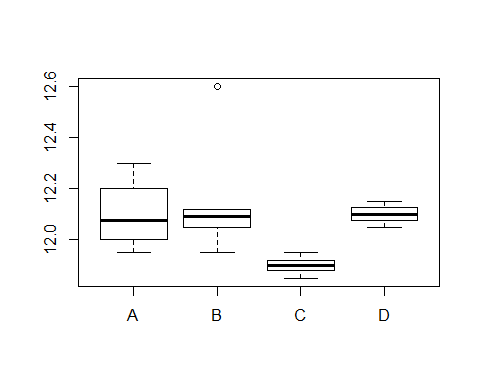
Milk <- read.csv("C:/Users/Jeevan/Desktop/Christ University/Statistics/DOE/Milk.csv")  
View(Milk)  
library(agricolae)

## Warning: package 'agricolae' was built under R version 3.5.3

str(Milk)

## 'data.frame': 17 obs. of 2 variables:  
## $ Machine: Factor w/ 4 levels "A","B","C","D": 1 1 1 1 2 2 2 2 2 3 ...  
## $ Level : num 11.9 12.1 12.1 12.3 12.6 ...

attach(Milk)  
boxplot(Level~Machine)



model <- aov(Level~Machine)  
summary(model)

## Df Sum Sq Mean Sq F value Pr(>F)  
## Machine 3 0.1912 0.06374 2.495 0.106  
## Residuals 13 0.3321 0.02554

**Conclusion**

The calculated probability value is greater than 0.01. Therefore, H0 is accepted. Hence, we can conclude that there is no significant difference between the mean level of fillings in the four machines.

**Question 2**

Four batches of paneer were prepared by taking three levels of fat content of milk.

Moisture content of the samples were recorded. Test for the significant difference

between batches of paneer and levels of fat content of milk.

**Aim**

Factor 1

H0: There is no significant difference between the means.

H1: There is a significant difference between atleast one pair of means.

Factor 2

H0: There is no significant difference between the means.

H1: There is a significant difference between atleast one pair of means.

Alpha level=0.01

**Procedure**

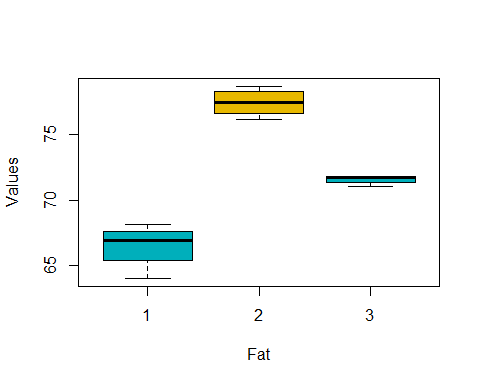
Paneer <- read.csv("C:/Users/Jeevan/Desktop/Christ University/Statistics/DOE/Paneer.csv")  
View(Paneer)  
str(Paneer)

## 'data.frame': 12 obs. of 3 variables:  
## $ Fat : int 1 2 3 1 2 3 1 2 3 1 ...  
## $ Batch : int 1 1 1 2 2 2 3 3 3 4 ...  
## $ Values: num 66.8 77.9 71.7 68.1 78.7 ...

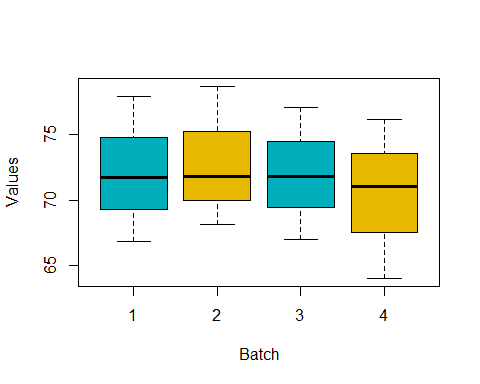
Paneer$Batch = as.factor(Paneer$Batch)  
Paneer$Fat = as.factor(Paneer$Fat)  
str(Paneer)

## 'data.frame': 12 obs. of 3 variables:  
## $ Fat : Factor w/ 3 levels "1","2","3": 1 2 3 1 2 3 1 2 3 1 ...  
## $ Batch : Factor w/ 4 levels "1","2","3","4": 1 1 1 2 2 2 3 3 3 4 ...  
## $ Values: num 66.8 77.9 71.7 68.1 78.7 ...

attach(Paneer)  
boxplot(Values~Fat,data = Paneer,xlab = "Fat",ylab = "Values",col = c("#00AFBB","#E7B800"))



boxplot(Values~Batch,data = Paneer,xlab = "Batch",ylab = "Values",col = c("#00AFBB","#E7B800"))



model <- aov(Values ~ Batch + Fat, data = Paneer)  
summary(model)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Batch 3 9.73 3.24 5.762 0.0336 \*   
## Fat 2 242.21 121.11 215.138 2.6e-06 \*\*\*  
## Residuals 6 3.38 0.56   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

library(lsmeans)

## Warning: package 'lsmeans' was built under R version 3.5.3

## Loading required package: emmeans

## Warning: package 'emmeans' was built under R version 3.5.3

## The 'lsmeans' package is now basically a front end for 'emmeans'.  
## Users are encouraged to switch the rest of the way.  
## See help('transition') for more information, including how to  
## convert old 'lsmeans' objects and scripts to work with 'emmeans'.

lml = lm(Values ~ Batch+Fat, data = Paneer)  
lsm1 = lsmeans(lml, "Fat")  
lsm1

## Fat lsmean SE df lower.CL upper.CL  
## 1 66.5 0.375 6 65.6 67.4  
## 2 77.5 0.375 6 76.6 78.4  
## 3 71.6 0.375 6 70.7 72.5  
##   
## Results are averaged over the levels of: Batch   
## Confidence level used: 0.95

pairs(lsm1)

## contrast estimate SE df t.ratio p.value  
## 1 - 2 -10.99 0.531 6 -20.725 <.0001   
## 1 - 3 -5.09 0.531 6 -9.604 0.0002   
## 2 - 3 5.90 0.531 6 11.121 0.0001   
##   
## Results are averaged over the levels of: Batch   
## P value adjustment: tukey method for comparing a family of 3 estimates

library(multcompView)

## Warning: package 'multcompView' was built under R version 3.5.3

CLD(lsm1,Letters = "abc")

## Warning: 'CLD' will be deprecated. Its use is discouraged.  
## See '? CLD' for an explanation. Use 'pwpp' or 'multcomp::cld' instead.

## Fat lsmean SE df lower.CL upper.CL .group  
## 1 66.5 0.375 6 65.6 67.4 a   
## 3 71.6 0.375 6 70.7 72.5 b   
## 2 77.5 0.375 6 76.6 78.4 ca   
##   
## Results are averaged over the levels of: Batch   
## Confidence level used: 0.95   
## P value adjustment: tukey method for comparing a family of 3 estimates   
## significance level used: alpha = 0.05

lsm2 = lsmeans(lml, "Batch")  
lsm2

## Batch lsmean SE df lower.CL upper.CL  
## 1 72.1 0.433 6 71.1 73.2  
## 2 72.9 0.433 6 71.8 73.9  
## 3 72.0 0.433 6 70.9 73.0  
## 4 70.4 0.433 6 69.3 71.5  
##   
## Results are averaged over the levels of: Fat   
## Confidence level used: 0.95

pairs(lsm2)

## contrast estimate SE df t.ratio p.value  
## 1 - 2 -0.740 0.613 6 -1.208 0.6444   
## 1 - 3 0.167 0.613 6 0.272 0.9922   
## 1 - 4 1.733 0.613 6 2.829 0.1056   
## 2 - 3 0.907 0.613 6 1.480 0.5018   
## 2 - 4 2.473 0.613 6 4.037 0.0262   
## 3 - 4 1.567 0.613 6 2.557 0.1466   
##   
## Results are averaged over the levels of: Fat   
## P value adjustment: tukey method for comparing a family of 4 estimates

**Conclusion**

Factor 1- The calculated probability value is greater than 0.01, therefore H0 is accepted. Hence, we can conclude that there is no significant difference between the means of the different batches of paneer.

Factor 2- The calculated probability value is lesser than 0.01, therefore H0 is rejected. Hence, we can conclude that there is a significant difference between the means of the different levels of fat content in the paneer.