MidSemLab 1740256

Jeevan

17/09/2019

1. 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 using two way ANOVA and Do the post hoc analysis.

		Batch			
		1	2	3	4
Fat level	1	66.8	68.12	67	64
	2	77.9	78.7	77.1	76.2
	3	71.7	71.8	71.8	71

Aim

Factor 1 : Paneer
H0: There is no significant difference between the means.
H1: There is a significant difference between atleast one pair of means.
Factor 2 : Fat Content
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)
```

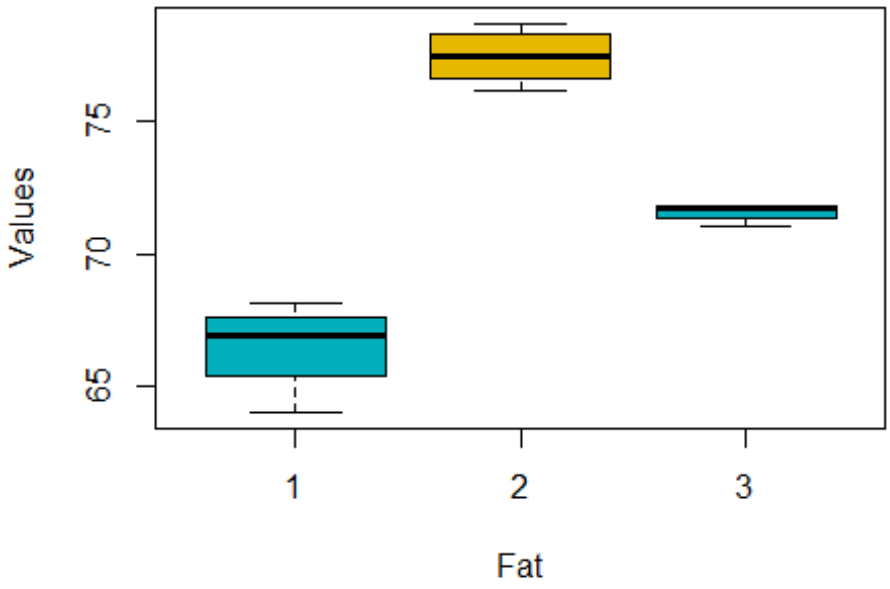
```
## 'data.frame':    12 obs. of  3 variables:
##  $ Fat   : int  1 2 3 1 2 3 1 2 3 1 ...
##  $ Batch : int  1 1 1 2 2 2 3 3 3 4 ...
##  $ Values: num  66.8 77.9 71.7 68.1 78.7 ...
```

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

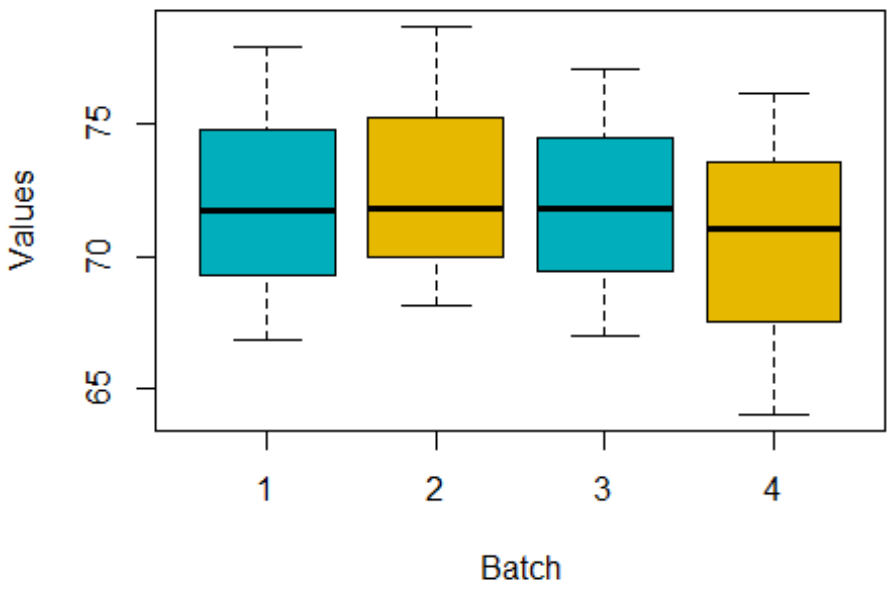
```
## 'data.frame':    12 obs. of  3 variables:
##  $ Fat   : Factor w/ 3 levels "1","2","3": 1 2 3 1 2 3 1 2 3 1 ...
```

```
##  $ 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(lm1, "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.

Using *lsmeans* pair 1-3 is significantly better than other pairs

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.

Using *lsmeans pair 1-3 is significantly better than other pairs*

2. The R &D manager of a manufacturing firm is in a state of dilemma whether the sales revenue (Crores of rupees) is affected by sales region. Because there might be variability from one period to another period, he decides to use the RCBD by taking the period as block. The corresponding data are presented below.

		Sales region					
		A	**B**	**C**	**D**	**E**	**F**
Period	**1**	18	9	15	22	9	10
	2	25	7	14	18	28	13
	3	20	8	12	9	15	17
	4	11	13	30	12	20	23
	5	18	11	25	15	16	8
	6	24	30	17	16	20	30

Aim

Factor 1 : Region
H0: There is no significant effect of region on sales revenue.
H1: There is a significant effect of region on sales revenue.
Factor 2 : Period
H0: There is no significant effect of period on sales revenue.
H1: There is a significant effect of period on sales revenue.

Procedure

```
library(readxl) # importing the package for reading data

## Warning: package 'readxl' was built under R version 3.5.2

library(lsmeans)# for performing CLD

## 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'.

library(multcompView)

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

MidSem_Q2 <- read_excel("C:/Users/Jeevan/Desktop/Christ
University/Statistics/DOE/MidSem_Q2.xlsx") # importing the data
```

```r
# View(MidSem_Q2) # viewing the data
str(MidSem_Q2) # getting the structure of the data
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    36 obs. of  3 variables:
##  $ Period: num  1 1 1 1 1 1 2 2 2 2 ...
##  $ Region: chr  "A" "B" "C" "D" ...
##  $ Sales : num  18 9 15 22 9 10 25 7 14 18 ...
```

```r
attach(MidSem_Q2) # attaching the data for ease
model<-aov(Sales~Region+Period) # making a 1 way ANOVA model
summary(model) # getting a summary of the model
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Region       5  171.2   34.24    0.80 0.5586
## Period       1  163.4  163.44    3.82 0.0604 .
## Residuals   29 1240.9   42.79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
lml = lm(Sales ~ Region+Period, data = MidSem_Q2)
lsm1 = lsmeans(lml, "Region")
lsm1
```

```
##  Region lsmean   SE df lower.CL upper.CL
##  A        19.3 2.67 29    13.87     24.8
##  B        13.0 2.67 29     7.54     18.5
##  C        18.8 2.67 29    13.37     24.3
##  D        15.3 2.67 29     9.87     20.8
##  E        18.0 2.67 29    12.54     23.5
##  F        16.8 2.67 29    11.37     22.3
##
## Confidence level used: 0.95
```

```r
pairs(lsm1)
```

```
##  contrast estimate   SE df t.ratio p.value
##  A - B       6.333 3.78 29  1.677  0.5570
##  A - C       0.500 3.78 29  0.132  1.0000
##  A - D       4.000 3.78 29  1.059  0.8933
##  A - E       1.333 3.78 29  0.353  0.9992
##  A - F       2.500 3.78 29  0.662  0.9847
##  B - C      -5.833 3.78 29 -1.545  0.6395
##  B - D      -2.333 3.78 29 -0.618  0.9888
##  B - E      -5.000 3.78 29 -1.324  0.7697
##  B - F      -3.833 3.78 29 -1.015  0.9091
##  C - D       3.500 3.78 29  0.927  0.9363
##  C - E       0.833 3.78 29  0.221  0.9999
##  C - F       2.000 3.78 29  0.530  0.9945
##  D - E      -2.667 3.78 29 -0.706  0.9797
##  D - F      -1.500 3.78 29 -0.397  0.9986
##  E - F       1.167 3.78 29  0.309  0.9996
```

```
##
## P value adjustment: tukey method for comparing a family of 6 estimates
```

Conclusion

Factor 1- The calculated probability value is greater than 0.05, therefore H0 is accepted. Hence, we can conclude that there is no significant effect on the region on sales revenue.

Using *lsmeans* region pair C-E is significantly better than other region pairs.

Factor 2- The calculated probability value is lesser than 0.05, therefore H0 is rejected. Hence, we can conclude that there is no significant effect of Period on sales revenue.
