Experiment – 11

Performing ANOVA for real dataset for Completely Randomized Design, Randomized Block design, Latin square Design

Aim: To understand the one way, two way and three way analysis of variances using R

Completely Randomized Design

Introduction

Responses among experimental units vary due to many different causes, known and unknown. The process of the separation and comparison of sources of variation is called the Analysis of Variance (AOV).
The AOV can be used for this purpose. It involves:
1. The partitioning of the total sum of squares of the experiment into each specified source of variation.
2. The estimation of the variance per experimental unit from these sources of variation.
3. The comparison of these variances by F-tests, which will lead to conclusions concerning the equality of the means.
The completely randomized design (CRD) refers to the random assignment of experimental units to a set of treatments. It is essential to have more than one experimental unit per treatment to estimate the magnitude of experimental error and to make probability statements concerning treatment effects.

Analysis of variance of a CRD

Source	df	Sum of squares (SS)	Mean squares (MS)	Observed F
Total	kr - 1	TSS		
Between treatments	k - 1	SST	MST	MST/MSE
Within treatments (experimental error)	k(r - 1)	SSE	MSE	

where r is the replication number per treatment.

Procedure:

- Import the data set
- Determine the summary and ANOVA using R functions
- Visualize the problem using R functions

Problem:

A car rental agency, which uses 5 different brands of tyres in the process of deciding the brand of tyre to purchase as standard equipment for its fleet, finds that each of 5 tyres of each brand last the following number of kilometres (in thousands):

A	B	C	D	E
36	46	35	45	41
37	39	42	36	39
42	35	37	39	37
38	37	43	35	35
47	43	38	32	38

Test the hypothesis that the five brands have almost the same average life.

Code and Results:

```
#One-way ANOVA
# Types of tyres
A=c(36,37,42,38,47)
B=c(46,39,35,37,43)
C=c(35,42,37,43,38)
D=c(45,36,39,35,32)
E=c(41,39,37,35,38)
group<-data.frame(cbind(A,B,C,D,E))
group

##    A  B  C  D  E
## 1 36 46 35 45 41
## 2 37 39 42 36 39
## 3 42 35 37 39 37
## 4 38 37 43 35 35
## 5 47 43 38 32 38

summary(group)

##        A             B             C              D              E
##  Min.   :36   Min.   :35   Min.   :35   Min.   :32.0   Min.   :35
##  1st Qu.:37   1st Qu.:37   1st Qu.:37   1st Qu.:35.0   1st Qu.:37
##  Median :38   Median :39   Median :38   Median :36.0   Median :38
```

```
##  Mean   :40   Mean   :40   Mean   :39   Mean   :37.4   Mean   :38
##  3rd Qu.:42   3rd Qu.:43   3rd Qu.:42   3rd Qu.:39.0   3rd Qu.:39
##  Max.   :47   Max.   :46   Max.   :43   Max.   :45.0   Max.   :41
```

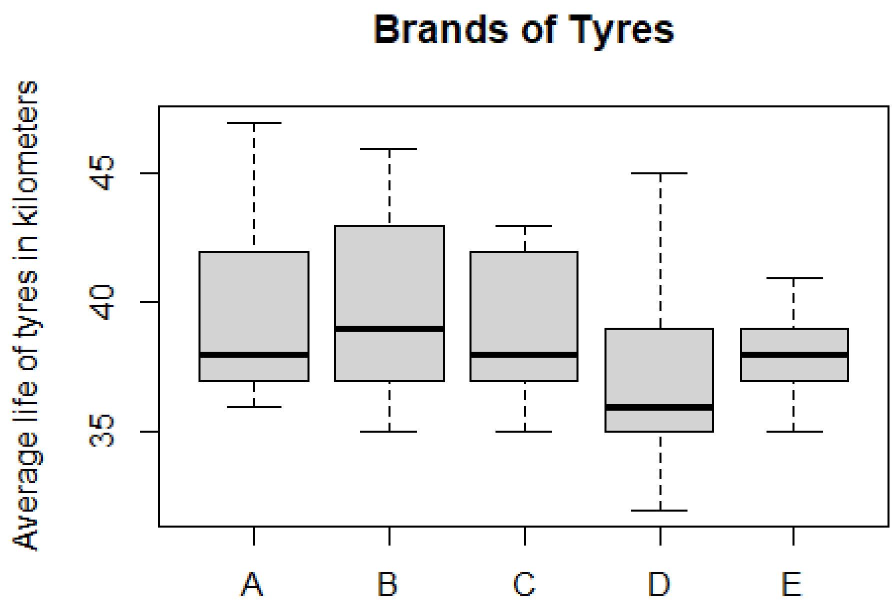
```
# stack vector from data frame
stgr<-stack(group);stgr
```

```
##    values ind
## 1      36   A
## 2      37   A
## 3      42   A
## 4      38   A
## 5      47   A
## 6      46   B
## 7      39   B
## 8      35   B
## 9      37   B
## 10     43   B
## 11     35   C
## 12     42   C
## 13     37   C
## 14     43   C
## 15     38   C
## 16     45   D
## 17     36   D
## 18     39   D
## 19     35   D
## 20     32   D
## 21     41   E
## 22     39   E
## 23     37   E
## 24     35   E
## 25     38   E
```

```
# completely randomized design
crd<-aov(values~ind,data=stgr)
# ANOVA table
summary(crd)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## ind          4   27.4    6.86   0.422  0.791
## Residuals   20  325.2   16.26
```

```
# Visualization of data
boxplot(group, ylab="Average life of tyres in kilometers",main="Brands of
Tyres")
```



Randomized Block Design

Introduction

A randomized block design is a type of experiment where participants who share certain characteristics are grouped together to form blocks, and then the treatment (or intervention) gets randomly assigned within each block. The objective of the randomized block design is to form groups where participants are similar, and therefore can be compared with each other.

ANOVA Table for a Randomized Block Design

Source	df	SS	MS	F
Treatments	$k-1$	SST	$MST = SST/(k-1)$	MST/MSE
Blocks	$b-1$	SSB	$MSB = SSB/(b-1)$	MSB/MSE
Error	$n-k-b+1$	SSE	$MSE = SSE/(n-k-b+1)$	
Total	$n-1$	$TotalSS$		

Procedure:

- Import the data set
- Determine the summary and ANOVA using R functions
- Visualize the problem using R functions

Problem:

The following table gives monthly sales (in thousand rupees) of a certain firm in the 3 states by its four salesmen.

States	**Salesmen**			
	I	II	III	IV
A	6	5	3	8
B	8	9	6	5
C	10	7	8	7

Setup the analysis of variance table and test whether there is any significant difference (i) between the salesmen (ii) between sales in the states.

Code and Results:

```
#Monthly sales of States
StateA=c(6,5,3,8)
StateA

## [1] 6 5 3 8

StateB=c(8,9,6,5)
StateB

## [1] 8 9 6 5

StateC=c(10,7,8,7)
StateC

## [1] 10  7  8  7

#frame the data set
Group<-data.frame(cbind(StateA,StateB,StateC))
Group

##   StateA StateB StateC
## 1      6      8     10
## 2      5      9      7
## 3      3      6      8
## 4      8      5      7

Sales=c(t(as.matrix(Group))); Sales

##  [1]  6  8 10  5  9  7  3  6  8  8  5  7

f=c("State A","State B","State C")
f

## [1] "State A" "State B" "State C"
```

```r
g=c("Salesman1","Salesman2","Salesman3","Salesman4")
g

## [1] "Salesman1" "Salesman2" "Salesman3" "Salesman4"

# number of columns
k=ncol(Group)
k

## [1] 3

# number of rows
n=nrow(Group)
n

## [1] 4

# Generate factor levels of States
States=gl(k,1,n*k,factor(f))
States

##  [1] State A State B State C State A State B State C State A State B State
C
## [10] State A State B State C
## Levels: State A State B State C

# Generate factor levels of Salesmen
Salesmen=gl(n,k,n*k,factor(g))
Salesmen

##  [1] Salesman1 Salesman1 Salesman1 Salesman2 Salesman2 Salesman2 Salesman3
##  [8] Salesman3 Salesman3 Salesman4 Salesman4 Salesman4
## Levels: Salesman1 Salesman2 Salesman3 Salesman4

# ANOVA table
anova=aov(Sales ~ States + Salesmen)
summary(anova)

##             Df Sum Sq Mean Sq F value Pr(>F)
## States       2 12.667   6.333   1.839  0.238
## Salesmen     3  8.333   2.778   0.806  0.535
## Residuals    6 20.667   3.444
```

Latin square Design

The Latin square design applies when there are repeated exposures/treatments and two other factors. This design avoids the excessive numbers required for full three way ANOVA.

The analysis of variance table for LSD is as follows:

Sources of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F-ratio
Rows	t-1	RSS	RMS	RMS/EMS
Columns	t-1	CSS	CMS	CMS/EMS
Treatments	t-1	TrSS	TrMS	TrMS/EMS
Error	(t-1)(t-2)	ESS	EMS	
Total	t^2-1	TSS		

Problem:

Perform Latin Square Design for the following.

Consider analyzing the productivity of five kinds of manure, five kinds of cultivation, and five kinds of crops. As follows, the data are organized in a Latin Square format:

	cultP	cultQ	cultR	cultS	cultT
manure1	"P42"	"R47"	"Q55"	"S51"	"T44"
manure2	"T45"	"Q54"	"R52"	"P44"	"S50"
manure3	"R41"	"P46"	"DS7"	"T47"	"Q48"
manure4	"Q56"	"S52"	"T49"	"R50"	"P43"
manure5	"S47"	"T49"	"P45"	"Q54"	"R46"

The three factors are: manure (manure1:5), cultivation (cultP:T), crop(P:T).

Codes and Results:

```
#creating dataframes in R
manure=c(rep("manure1",1), rep("manure2",1), rep("manure3",1),
rep("manure4",1), rep("manure5",1))
cultivation=c(rep("cultP",5), rep("cultQ",5), rep("cultR",5), rep("cultS",5),
rep("cultT",5))
```

```r
crop=c("P","T","R","Q","S", "R","Q","P","S","T", "Q","R","S","T","P",
"S","P","T","R","Q", "T","S","Q","P","R")
freq=c(42,45,41,56,47, 47,54,46,52,49, 55,52,57,49,45, 51,44,47,50,54,
44,50,48,43,46)
data=data.frame(cultivation,manure,crop,freq)
data

##    cultivation  manure crop freq
## 1        cultP manure1    P   42
## 2        cultP manure2    T   45
## 3        cultP manure3    R   41
## 4        cultP manure4    Q   56
## 5        cultP manure5    S   47
## 6        cultQ manure1    R   47
## 7        cultQ manure2    Q   54
## 8        cultQ manure3    P   46
## 9        cultQ manure4    S   52
## 10       cultQ manure5    T   49
## 11       cultR manure1    Q   55
## 12       cultR manure2    R   52
## 13       cultR manure3    S   57
## 14       cultR manure4    T   49
## 15       cultR manure5    P   45
## 16       cultS manure1    S   51
## 17       cultS manure2    P   44
## 18       cultS manure3    T   47
## 19       cultS manure4    R   50
## 20       cultS manure5    Q   54
## 21       cultT manure1    T   44
## 22       cultT manure2    S   50
## 23       cultT manure3    Q   48
## 24       cultT manure4    P   43
## 25       cultT manure5    R   46

#recreating the original table, using the matrix function
matrix(data$crop,5,5)

##      [,1] [,2] [,3] [,4] [,5]
## [1,] "P"  "R"  "Q"  "S"  "T"
## [2,] "T"  "Q"  "R"  "P"  "S"
## [3,] "R"  "P"  "S"  "T"  "Q"
## [4,] "Q"  "S"  "T"  "R"  "P"
## [5,] "S"  "T"  "P"  "Q"  "R"

matrix(data$freq,5,5)

##      [,1] [,2] [,3] [,4] [,5]
## [1,]   42   47   55   51   44
## [2,]   45   54   52   44   50
## [3,]   41   46   57   47   48
```

```
## [4,]   56   52   49   50   43
## [5,]   47   49   45   54   46
```

```
#creating the anova table
fit=lm(freq~manure+cultivation+crop,data)
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: freq
##             Df Sum Sq Mean Sq F value   Pr(>F)
## manure       4  17.76   4.440  0.7967 0.549839
## cultivation  4 109.36  27.340  4.9055 0.014105 *
## crop         4 286.16  71.540 12.8361 0.000271 ***
## Residuals   12  66.88   5.573
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion: The problems on ANOVA have been executed using R