R Codes for Randomized Complete Block Design

M. Iqbal Jeelani*, Nageena Nazir**, Manish Kr Sharma*, Anil Bhat*, Mansha Gul** and Banti Kumar*

*Sher-e-Kashmir University of Agricultural Sciences & Technology of Jammu, India

**Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir, India

* Author to whom correspondence should be addressed; E-Mail: jeelani.miqbal@gmail.com.

Abstract: The present work is an attempt to build some new functions of Randomized Complete Block Design in R-software. Different computer programs are developed using R-software. All these functions are run on real data set .The present work is an attempt to show the flexibility of R.

Keywords: Randomized Complete Block Design, Grain Yield, R-software.

1. Introduction

Any scientific investigation involves formulation of certain hypothesis whose validity is examined through the data generated from an experiment conducted for the purpose. Designing an experiment is an integrated component of every research programme. The three basic principles of designing an experiment are replication, local control (blocking) and randomization. Whereas the first two help in increasing precision of the experiment and the last one is used to decrease bias. The data generated through designed experiments exhibit a lot of variability. Even experimental units (plots) subjected to same treatment give rise to different observations thus creating variability. Howsoever careful planning is made for the experimentation the error component is always present. A popular measure of the experimental error is the Coefficient of Variation (CV). Treatment is one component in every designed experiment that causes variability. If the experimental material is homogeneous and does not exhibit any variability then the treatments are applied randomly to the experimental units. Such designs are known completely randomized design (CRD). Besides the variability arising because of the

application of treatments, the variability present in the experimental material (plots) is the other major known source of variability. Forming groups called blocks containing homogeneous experimental material is in one direction only. Contrary to the random allotment to all the experimental units in a CRD, the treatments are allotted randomly to the experimental units within each block. Such designs are termed as randomized complete block (RCB) design. Further, the variability in the experimental material may be in two directions forming rows and columns which can control the variability in two directions and the treatments are assigned to each cell. For the randomization purpose, first the rows are randomized and then the columns are randomized. Such a design is known as two-way elimination of heterogeneity and is termed as Latin square design (LSD). The other row-column designs are the Youden square designs, Youden type designs, Generalized Youden designs, Pseudo Youden designs, etc. In some experimental situations, some factors require large plot sizes and the effect of these factors is obvious, the experimenter is interested in the main effects of others factors and interaction with high precision. Split plot designs are used for such experimental situations. If the experimenter is interested only in interaction of the two factors and both factors require large plot sizes, the strip plot designs may be used.

1.1. Principles of a Good Design

The three basic principles of a good design are as follows;

A. **Replication:** The repetition of the treatment under comparison is called the replication. The purpose of replication is:

i) to reduce the experimental error and increase the precision of the design

ii) to estimate the experimental error, given by $\sqrt{\sigma^2/r}$, where σ is the S.D and r is the number of replications without which the significance of the difference between the two treatments cannot be studied.

B. **Randomization:** It is random allotment of treatments to various experimental units (plots) so that each treatment gets an equal chance of selection which will reduce the bias in the experiment. Thus, the purpose of randomization is:

i) to reduce the bias in the experiment and ensure that the different treatments are subjected to the same environmental effects and

ii) to guarantee the validity of the test of significance.

C. **Local Control:** If the experimental material is not homogeneous and the treatments are scattered randomly over the whole area, the soil heterogeneity enters the chance factor and thus increases the experimental error. It is, therefore, desirable to reduce the experimental error as far as possible. In order to remove the soil fertility effect from the experimental error, the whole experimental

area is divided into homogeneous groups (blocks) in such a way that the variation between the blocks is large and within the blocks is least. This process of reducing the experimental error by dividing the experimental area into more homogeneous blocks is known as local control.

The precision of a design is the ability with which it detects the smaller real differences between the treatments which can be increased by decreasing the experimental error. The lesser, the experimental error the greatest is the precision. Further, the precision can also be increased by increasing the number of replications and use of local control.

The number of replications required is to provide at least 12 degrees of freedom for error as the values of F do not decrease rapidly beyond 12degrees of freedom. For the validity of the experiments, it is necessary that the different treatments should be randomly allocated to experimental unit (plots). The procedure of randomization differs for different experimental designs. The basic principles of a good design play an important role in planning of experiments and analysis of basic designs. The three basic designs are:

    i)   Completely Randomized Design (CRD)

    ii)   Randomized Complete Block Design (RCBD) and

    iii) Latin Square Design (LSD). In this article emphasis is given on RCBD.

2. Materials and Methods

It has been seen that when the experimental units are homogeneous then CRD should be adopted. The experimental material is a major source of variability in the data. When the experimental units are heterogeneous, a part of the variability is accounted for by grouping the experimental units. If the fertility gradient is in one direction only, then the experimental units are divided into the homogeneous blocks perpendicular to the direction of fertility gradient and each block constitutes a single replication. The treatments are then allotted randomly to the experimental units within each group (or blocks). Such a design is called a Randomized Complete Block Design. Since all the treatments are applied in each block, the blocks are divided into as many units as many number of treatments. The variability of the whole experimental material is divided into two parts: (i) between blocks and (ii) within blocks (error). If, however, variation between the blocks is not significantly large, this type of grouping of the units does not lead to any advantages. In such situations it is not desirable to adopt randomized complete block designs in preference to completely randomized designs.

2.1. Procedure of Randomization

The treatments are assigned to the units (plots) within each group (block) entirely at random with the help of random numbers as explained in the case of CRD. It is important that for every group the

randomization should be a fresh. The same set of random numbers should not be used for all the groups. For example, the layout of RCBD with 5 treatments each replicated 5 times will be given as follows.

Blocks/Replications

D	A	B	C	D
A	B	C	D	E
E	C	A	E	B
B	D	E	A	C
C	E	D	B	A

2.2. Analysis of Randomized Complete Block Design

The model used for analysis is $Y_{ij} = \mu + T_i + B_j + e_{ij}$ where, Y_{ij} is the observation on i^{th} treatment in j^{th} block, μ is the general mean, T_i (i = 1,2,….,t) is the effect of ith treatment, B_j (j = 1,2,….,r) is the effect of jth block, and e_{ij} is the random error distributed normally, independently with zero mean and constant variance ie $N(0, \sigma^2)$.The data collected from experiments with randomized block design are arranged in a two way classification according to treatments and blocks. Let $\Sigma Y_{ij} = G$ be the grand total of n (= rt) observations, ΣT_i (i = 1,2,….,t) and ΣB_j (j = 1,2,….,r) be the treatment and block totals respectively. Then calculate the following:

i) Correction Factor (C.F) = G^2/n,

ii) ii) Total Sum of Squares (TSS) = $\sum_i \sum_j Y_{ij}^2 - C.F$

iii) Sum of Squares due to treatments (SS$_t$) = $\Sigma T_i^2 / r - CF$

iv) Sum of .Squares due to blocks (SS$_b$) = $\Sigma B_j^2 / t - CF$

v) Sum of Squares due to error (SS$_e$) = TSS - SS$_t$ -SS$_b$

vi) Prepare ANOVA table as follows:

ANOVA TABLE

Source of variation	df	SS	MSS	F
Blocks	(t-1)	SS$_t$	SS$_t$/(t-1) = s_t^2	
Treatments	(r-1)	SS$_b$	SS$_b$/(r-1) = s_b^2	s_t^2 / s_e^2
Error	(t-1)(r-1)	SS$_e$	SS$_e$ /(r-1)(t-1) = s_e^2	
Total	(rt-1)	TSS	-	

To test the hypothesis that the treatments have equal effects F- test is used. If the treatments differ significantly then calculate critical difference to find out which of the treatments have same effect. Some of the recent work done in R-software has been given by Iqbal *et.al* (2014) in context to sampling. In this paper some new functions `RCBD<-function(x,r,t)`is developed for carrying out analysis of a Randomized Complete Block Design. A complete description of R-software is given in Pinheiro and Bates (2007).

2.3. R-Codes for RCBD

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. One of the important feature of R- software is that it is an Open Source and freely available on website *http://cran-project.org*.Verzani (2002), Khan and Mir (2005) , Venables*et.al* (2009) and Jeelani *et.al* (2014) have discussed in detail the application of R- software in agricultural data analysis. R language is essentially a functional language for all practical purposes of data analysis and graphics. However, in case some specific situations data analyst is forced to develop his own functions according to his requirements. Consequently, few functions have been developed according to the requirements of this study.

```
RCBD<-function(d,r,t)
```

This is function developed in R-software. It takes the arguments d (name of the data), r (total number of replications), t (total number of treatments and returns treatment means, standard error of difference of means and least significant difference at 95% confidence .The codes of the function follow

```
RCBD<-function(x,r,t)
{
d=data.frame(d)
r=as.numeric(r)
t=as.numeric(t)
response=d[,3]
treatment=d[,2]
replication=d[,1]
treatment=factor(treatment)
replication=factor(replication)
fit1=aov(response~treatment+replication)
result=summary(fit1)
meantables=model.tables(fit1,type="means",se=T)
sed=meantables$se$treatment
edf=(r-1)*(t-1)
cd=qt(.975,edf)*sed
```

```
list(result=result,meantables=meantables,sed=sed,cd=cd)
}
```

3. Results and Discussion

The data taken for running the above function was taken from Gomez (1984), page number 26. To illustrate the steps involved in carrying out the analysis for data from a RCBD, we use the data from an experiment that compared six rates of seeding of rice variety IR8, given in Table.No.1 below.

Table 1. Grain Yield of Rice variety IR8 with six different rates of seeding

Treatment Kg seed/ha	Grain Yield, kg/ha				Treatment Total	Treatment Mean
	R1	R2	R3	R4		
25	5113.00	5398.00	5307.00	4678.00	20496.00	5124.00
50	5346.00	5952.00	4719.00	4264.00	20281.00	5070.25
75	5272.00	5713.00	5483.00	4749.00	21217.00	5304.25
100	5164.00	4831.00	4986.00	4410.00	19391.00	4847.75
125	4804.00	4848.00	4432.00	4748.00	18832.00	4708.00
150	5254.00	4542.00	4919.00	4098.00	18813.00	4703.25
Replication Total	30953.00	31284.00	29846.00	26947.00	119030.00	
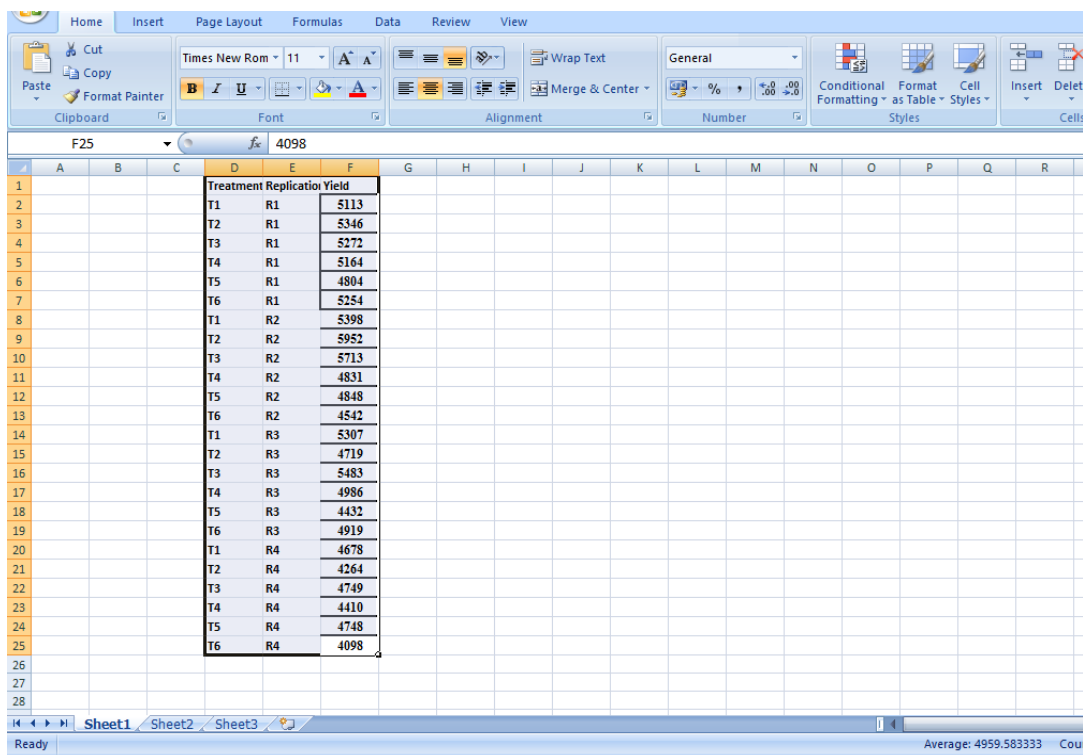
Figure 1. Grain data with six treatments and 4 replications in MS Excel

```
>RCBD=read.table("clipboard",header=T)
# data recalled from MS Excel
>names(RCBD)
[1] "Treatment"  "Replication" "Yield"
>RCBD $Treatment<-factor(RCBD $Treatment)
>RCBD $Replication<-factor(RCBD$Replication)
>summary(RCBD)
 Treatment Replication     Yield
 T1:4      R1:6        Min.   :4098
 T2:4      R2:6        1st Qu.:4709
 T3:4      R3:6        Median :4884
 T4:4      R4:6        Mean   :4960
 T5:4                  3rd Qu.:5281
 T6:4                  Max.   :5952
     Out put
     RCBD(rbd,4,6)
     $result
     Df  SumSq Mean Sq F value    Pr(>F)
     treatment    5 1198331  239666  2.1678 0.112809
     replication  3 1944361  648120  5.8622 0.007416 **
     Residuals   15 1658376  110558
     ---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     meantables

     Tables of means
     Grand mean
     4959.583

     treatment
      t1   t2   t3   t4   t5   t6
     5124 5070 5304 4848 4708 4703
     replication
      r1   r2   r3   r4
     5159 5214 4974 4491

     Standard errors for differences of means
     treatment replication
              235.1        192.0

     sed # standard error of difference of mean
     235.1153

     cd # critical  difference
     501.1364
```

4. Conclusions

RCBD is the widely used design of experiment in agriculture. It has been established that RCBD

has its practical implications. It can be applied to analyze data generated in a scientific investigation.

Finally R-software is flexible and can be used to any kind of data set, and due its flexibility and object

oriented nature this software can be used for developing new functions for solving the problems which are not solved by standard statistical packages.

Acknowledgments

References

[1] Gomez. K. A. & Gomez. A.A.; *Statistical Procedure for Agricultural Research*. New York: John Willey and sons.pp. 26-32, 1984

[2] Jeelani. M.I., M. Nazir, N. Mir, S. A., Jeealni, F., Dar, N. A., Haq, S., Maqbool, S., Wani, S. Application of Simple Random Sampling in Agriculture using R-software. *Indian Journal of Science and Technology*. 7(5)(2014): 706-709.

[3] Khan, A. A. and Mir, A. H.; Applications of R- software in Agricultural Data Analysis. *SKUAST Journal of Research*. 7(1)(2005): 36-64.

[4] Pinheiro, J.C. and Bates, M.D.; *Mixed-Effects Models in S and S-PLUS*. Springer-Verlag, New York. 2007

[5] Venables W. N., Smith D. M. and R Development Core Team. *An Introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics*, Version 2.9.0. (2009) URL http://cran.cnr.berkeley.edu/doc/manuals/R-intro.pdf

[6] Verzani, J.; *Simple R – Using R for introductory Statistics*. (2002) URL http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf