Ashwini Hebbar
Research Scholar, Department of
Statistics, Jain University,
Bengaluru, Karnataka, India

Dr. Suma AP
Research Guide, Department of
Statistics, Jain University,
Bengaluru, Karnataka, India

R codes for experimental designs in agricultural research with special reference to CRD, RCBD, LSD, split-plot and strip-plot

Ashwini Hebbar and Dr. Suma AP

Abstract
Agriculture is an applied science that involves all features of crop production. Crop production is an agricultural practice followed by worldwide farmers to grow and produce crops to use as food and fiber. To obtain the optimization in crop production, the technique of designs of experiment is very useful which comes under the branch of applied statistics. Designs of experiment are efficient method to deal with problems in agriculture using different resources in crop production technique. The objective of this article is to exhibit the researchers to some basic, sophisticated designs utilized by agricultural researchers in performing their investigation and also instigate R codes for some selected complete and incomplete block designs. In the present article an attempt has been taken to give R codes by taking real life research problems for each design. Selected designs have been described in short followed by the R codes.

Keywords: Agricultural research, designs, R codes

Introduction

Agricultural research is a specialized kind of research system which can be carried out using laboratory and field facilities as well as by interacting with farmers as critical informants for their betterment and raising their day – to – day level of livelihood. Agricultural experimentation is based on crops, livestock, fisheries, forests and the environment. Prof. R.A. Fisher has developed modern concepts of experimental designs in planning and conduct of agricultural field experiments. In agricultural field experiments, an experiment may be conducted to see the effect of different fertilizers. There are many types of complete and incomplete block designs. In this study, only a few designs have been discussed depending on its applicability in the real-life situations. An effort has been made to solve some examples related to experimental designs using R codes.

It is very essential to know the following terminologies about experimental designs.

Experimental unit: A subject or a group of objects or the total material to which a treatment is applied in a trial in a single replication is known as an experimental unit.

Treatment: A treatment is a substance or known factor that is administered or allocated to one or more experimental units to estimate the effect pertaining to certain characteristics or for comparing it with others.

Experimental error: Response from all experimental units which are given similar treatment may not be same even under identical situations. These differences in responses may be due to various reasons. Other factors like diverseness of soil, climatic factors and genetic differences also may cause variations. The variations in response caused by extraneous factors are known as experimental errors.

Basic principles of design of experiment

To reduce the experimental error, one needs to adopt certain principles known as basic principles of experimental design.

The basic principles are replication, randomization and local control.

Corresponding Author:
Ashwini Hebbar
Research Scholar, Department of
Statistics, Jain University,
Bengaluru, Karnataka, India

1. **Replication:** When a treatment appears more than once in an experiment, it is said to be replicated. In other words, repeated application of the treatments is known as replication.
2. **Randomization:** The allotment of the treatments to the experimental units by random procedure is known as randomization. Here all the treatments have equal chance of being allocated to different experimental units. Randomization assures unbiased estimates of treatment means and differences among them. Also, there is an unbiased estimate of experimental error and thereby valid test of significance.
3. **Local Control:** The operation of reducing experimental errors by providing relatively heterogeneous experimental areas into homogeneous units. The local control will increase the efficiency of the experimental designs.

Types of Experimental Designs
- Completely randomized designs (CRD)
- Randomized block design (RCBD)
- Latin square design (LSD)
- Split plot design
- Strip plot design

Completely Randomized Design (CRD)
CRD is the basic single factor design. In this design the treatments are assigned completely at random so that an experimental unit has the same chance of receiving any treatment. But CRD is suitable only when the experimental material is homogeneous. CRD is not preferred in field experiments.
The statistical model for CRD with one observation per unit is

$$Y_{ij} = \mu + t_i + e_{ij}$$

μ = overall mean effect
t_i = true effect of the ith treatment
e_{ij} = error term of the jth unit receiving i^{th} treatment

The design is very flexible and statistical analysis is simple compared to other designs. In CRD loss of information due to missing data is small compared to other designs due to the larger number of degrees of freedom for the error source of variation.
CRD is less accurate than other designs. It is difficult to find homogeneous experimental units in all aspects and hence CRD is seldom suitable for field experiments as compared to other experimental designs.
Completely randomized design is most useful in laboratory technique and methodological studies where either the experimental material is homogeneous or the intrinsic variability between units can be reduced. CRD is also recommended in circumstances where an appreciable fraction of units is likely to be destroyed or fail to respond.

R Codes
Example: The following table gives the yields in pounds per plot, of five varieties of Wheat after being applied to each of 4 plots, completely randomized. (Gupta S.C. and Kapoor V.K., 2014) [3]

Varieties	Yield in lbs.			
A	8	8	6	10
B	10	12	13	9
C	18	17	13	16
D	12	10	15	11
E	8	11	9	8

```
> treat <-rep(c("A","B","C","D","E"), each = 4)
>fac<- factor(rep(c(1, 2, 3, 4, 5), each = 4))
> treat

OUTPUT
[1] "A" "A" "A" "A" "B" "B" "B" "B" "C" "C" "C" "C" "D"
"D" "D" "D" "E" "E" "E" "E"
> yield <- c(8, 8, 6, 10, 10, 12, 13, 9, 18, 17, 13, 16, 12, 10,
15, 11, 8, 11, 9, 8)
>exp<- data.frame(treat, treatment = fac, response = yield)
> mod <-aov(response ~ treatment, data = exp)
> summary(mod)

OUTPUT
DfSum SqMean SqF value Pr(>F)
treatment 4 155.2 38.80 11.19 0.000208 ***
Residuals 15 52.0 3.47
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Inference: There is significant difference between the varieties.

Randomized Complete Block Design (RCBD)
In field trial, if the whole of the experimental area is not homogeneous and the fertility gradient is only in one direction, then a simple method of controlling the variability of the experimental material consists in stratifying or grouping the whole area into relatively homogeneous strata or sub-groups perpendicular to the direction of the fertility gradient. Now if the treatments are applied at random to relatively homogeneous units within each strata or block and replicated over all the blocks, the design is a Randomised Block Design.
A Randomized Complete Block Design (RCBD) is defined by an experiment whose treatment combinations are assigned randomly to the experimental units within a block.
Generally, blocks cannot be randomized as the blocks represent factors with restrictions in randomization such as location, place, time, gender etc.
In this design the experimental blocks divided into homogeneous blocks. Within the blocks it will be homogeneous and between the blocks it is heterogeneous.
Completely randomize block design means that every block has all treatments and the treatments are randomizing with all blocks.

The linear statistical model is

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$$

Where
τ_i = i^{th} treatment effect
β_j = jth block effect
ϵ_{ij} = Error

Model includes additional additive block effect
In RBD, no restrictions are placed on the number of treatments or the number of replicates. Here statistical analysis is simple, rapid and straight forward. Moreover, the error of any treatment can be segregated and any number of treatments may be omitted from the analysis without

complicating it. RBD has been shown to be more efficient or accurate than CRD for most types of experimental works.

If blocks are not homogeneous RBD may give misleading results. It is not suitable for large number of treatments because in that case the block size will increase and it may not be possible to keep large blocks homogeneous. In RBD if the data on more than two plots is missing, the statistical analysis becomes quite difficult and complicated.

Randomized block design is most useful in conditions in which the experimental material is heterogeneous and it is possible to divide the experimental material into homogeneous groups of units or plots called blocks or replications.

It is mostly used in agricultural field experiments, where soil heterogeneity may be present due to soil fertility gradient, and in clinical trials on animals, where the animals (experimental units) vary in age, breed, initial body weight, and so on.

Example: The yields of 6 varieties of crop in pounds along with the plan of the experiment are given below. The number of blocks is 5, plot size is 1/20 acre and the varieties have been represented by A, B, C, D, E and F (Gupta S.C. and Kapoor V.K., 2014) [3]

Block I	D 17	F 70	A 20	C 12	B 9	E 28
Block II	B 12	E 26	D 10	C 15	A 26	F 62
Block III	E 23	C 15	F 56	A 30	D 20	B 10
Block IV	A 28	B 9	E 35	F 64	D 23	C 14
Block V	F 75	D 2	E 3	C 14	B 7	A 23

R codes

```
>block <- factor(rep(c("Block I", "Block II", "Block III",
"Block IV", "Block V"), each = 6))
> treat <- factor(rep(c("A", "B", "C", "D", "E", "F"), each =
5))
> block
OUT PUT
[1] Block I Block I Block I Block I Block I Block I Block II
Block II Block II Block II
[11] Block II Block II Block III Block III Block III Block III
Block III Block III Block IV Block IV
[21] Block IV Block IV Block IV Block IV Block V Block V
Block V Block V Block V Block V
Levels: Block I Block II Block III Block IV Block V
>y <- c(17, 70, 20, 12, 9, 28, 12, 26, 10, 15, 26, 62, 23, 15, 56,
30, 20, 10, 28, 9, 35, 64, 23, 14, 75, 2, 3, 14, 7, 23)
> # y is the yield
> results <- data.frame(y, block, treat)
>
> fit <- aov(y ~ treat+block, data = results)
> summary(fit)

OUTPUT
DfSum SqMean SqF value Pr(>F)
treat 5 3672 734.4 2.298 0.0837 .
block 4 1660 415.0 1.299 0.3043
Residuals 20 6392 319.6
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Inference: There is no significant difference between the blocks and treatments.

Latin Square Design (LSD)

The major feature of the Latin square (LS) design is its capacity to simultaneously handle two known sources of variation among experimental units. It treats the sources as two independent blocking criteria, instead of only one as in the RCB design. The two-directional blocking in a LS design, commonly referred to as row-blocking and column-blocking, is accomplished by ensuring that every treatment occurs only once in each row-block and once in each column-block. This procedure makes it possible to estimate variation among row-blocks as well as among column-blocks and to remove them from experimental error.

When the heterogeneity is existing in all the direction this design is appropriate. The block can be made in two directions along the rows and columns. In the Latin Square matrix each 'treatment' appears once in each row and once in each column. Randomisation of the treatments, however, is basic to this design. Latin square design is said to be standard latin square if first row and first column is in natural order.

Linear model for Latin Square Design is

$$Yijk = \mu + \alpha_i + \beta_j + \tau_k + \epsilon_{ijk}$$

Where, μ = overall mean

α_i = effect of ith row

β_j = effect of jth row

τ_k = effect of kth column

ϵ_{ijk} = Error term

Some of the cases where the LS design can be appropriately used are
1. Field trials in which the experimental area has two fertility gradients running perpendicular to each other, or has a unidirectional fertility gradient but also has residual effects from previous trials.
2. Laboratory trials with replication over time such that the difference among experimental units conducted at the same time and among those conducted over time constitute the two known sources of variability.
3. Insecticide field trials where the insect migration has a predictable direction that is perpendicular to the dominant fertility gradient of the experimental field.

Example: Below are given the plan and yields in pounds per plot of a 5 × 5 Latin square experiment on the wheat crop carried out for testing effects of five manorial treatments A, B, C, D and E. A denoting control (or no treatment).

B 15	A 8	E 17	D 20	C 17
A 9	D 21	C 19	E 16	B 13
C 18	B 12	D 23	A 8	E 17
E 18	C 16	A 10	B 15	D 23
D 22	E 15	B 13	C 18	A 10

R Codes

```
> row<-c(rep("A",5), rep("B",5), rep("C",5), rep("D",5), rep("E",5))
> col<-c(rep(c("01","02", "03", "04", "05"), 5))
> rep<-c("B", "A", "E", "D", "C", "A", "D", "C", "E", "B", "C", "B",
"D", "A", "E", "E", "C", "A", "B", "D", "D", "E", "B", "C", "A")
> treat<-c(15, 8, 17, 20, 17, 9, 21, 19, 16, 13, 18, 12, 23, 8, 17, 18,
16, 10, 15, 23, 22, 15, 13, 18, 10)
> y<-data.frame(col,row,rep,treat)
>matrix(y$rep,5,5,byrow = T)
OUTPUT
[,1] [,2] [,3] [,4] [,5]
[1,] "B" "A" "E" "D" "C"
[2,] "A" "D" "C" "E" "B"
[3,] "C" "B" "D" "A" "E"
[4,] "E" "C" "A" "B" "D"
[5,] "D" "E" "B" "C" "A"
> matrix(y$treat,5,5,byrow = T)
OUTPUT
[,1] [,2] [,3] [,4] [,5]
[1,] 15 8 17 20 17
[2,] 9 21 19 16 13
[3,] 18 12 23 8 17
[4,] 18 16 10 15 23
[5,] 22 15 13 18 10
>yfit<-lm(treat~row+rep+col,y)
>anova(yfit)

OUTPUT
Analysis of Variance Table

Response: treat
DfSum SqMean SqF value Pr(>F)
row 4 3.04 0.760 0.8201 0.53670
rep 4 454.64 113.660 122.6547 1.27e-09 ***
col 4 14.24 3.560 3.8417 0.03105 *
Residuals 12 11.12 0.927
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Inference:
1. There is no significant difference between the rows
2. There is significant difference between replications
3. There is significant difference between columns

Split – plot Design

In field experiments certain factors may require larger plots than for others. For example, experiments on irrigation, tillage etc. requires larger areas. On the other hand, experiments on fertilizers etc. may not require larger areas. To accommodate factors which require different sizes of experimental plots in the same experiment, split plot design has been evolved.

In this design larger plots are taken for the factor which requires larger plots. Next each of the larger plots is split into smaller plots to accommodate the other factor. The different treatments are allotted at random to their respective plots. Such arrangement is called split plot design. In split plot design the larger plots are called main plots and smaller plots within the larger plots are called as sub plots. The factor levels allotted to the main plots are main plot treatments and the factor levels allotted to sub plots are called as sub plot trearments.

Layout and analysis of variance table

First the main plot treatment and sub plot treatment are usually decided based on the needed precision. The factor for which greater precision is required is assigned to the sub plots.

The replication is then divided into number of main plots equivalent to main plot treatments. Each main plot is divided into subplots depending on the number of sub plot treatments.

The main plot treatments are allocated at random to the main plots as in the case of RCBD. Within each main plot the sub plot treatments are allocated at random as in the case of RCBD. Thus randomization is done in two stages. The same procedure is followed for all the replications independently.

The analysis of variance will have two parts, which correspond to the main plots and sub-plots. For the main plot analysis, replication × main plot treatments table is formed. From this two-way table sum of squares for replication, main plot treatments and error (a) are computed. For the analysis of sub-plot treatments, main plot × sub-plot treatments table is formed. From this table the sums of squares for sub-plot treatments and interaction between main plot and sub-plot treatments are computed. Error (b) sum of squares is found out by residual method.

Linear Statistical model for Split Plot Design is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ij} + \gamma_k + (\alpha\gamma)_{ik} + \delta_{ijk}$$

Where, μ = overall mean

α_i = effect of ith main plot

β_j = effect of jth main plot

ϵ_{ij} = Error term related to main plot

γ_k = Effect of kth sub plot

$(\alpha\gamma)_{ik}$ = effect of ith main k^{th} sub plot (Interaction)

δ_{ijk} = error associated with sub plot

In split plot design, an increased precision is attained on the sub plot treatments and the interaction between sub plot and main plot treatments. It is statistically more efficient.

Split plot design is used when the factorial treatment structure has two levels of experimental units. In this case, two levels of randomization will be applied to assign experimental units to treatments.

R codes

Example: A researcher was interested to study the spacing and seedling (4 types of spacing, 5 kinds of seedlings with three replications). Spacing were kept in the main plot and seedling in subplots. (Kailasam C. and GangaiSelvi R., 2010) [4]

```
install. Packages ("agricolae") library (agricolae) library
(tidyverse)
>mod.aov<- aov(Yield ~ Spacing*Seedling + Error
(Seedling/Block), data = read.csv ("SPLIT PLOT. csv",
header = TRUE)) > summary (mod.aov)

Out Put
Error: Seedling Df Sum Sq Mean Sq
Seedling 4 369.7 92.44
Error: Seedling: Block
Df Sum Sq Mean SqF value Pr(>F)
Residuals 10 7.817 0.7817
Error: Within
DfSum SqMean SqF value Pr(>F)
Spacing 3 7.653 2.5510 2.943 0.0489 *
Spacing:Seedling12 1.201 0.1001 0.115 0.9998
Residuals 30 25.999 0.8666---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Inference: Spacing shows significant difference in the yield level.

Strip – plot design

The strip plot design is especially suited for a two – factor experiment in which the desired precision for measuring the interaction effect between the two factors is higher than that for measuring the main effect of either one of the two factors. This is accomplished with the use of three plot sizes:

1. Vertical strip plot for the first factor – the vertical factor.
2. Horizontal strip plot for the second factor – the horizontal factor.
3. Intersection plot for the interaction between the two factors.

The vertical – strip plot and horizontal – strip plot are always perpendicular to each other. However, there is no relationship between their sizes, unlike the case of main plot and sub plot of the split – plot design. The interaction plot is the smallest plot.

In a strip plot design, the degrees of precision associated with the main effects of both factors are sacrificed in order to improve the precision of the interaction effect.

The analysis is carried out in 3 parts
• Vertical strip analysis
• Horizontal strip analysis
• Interaction analysis

Linear statistical model for Strip Plot Design is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ij} + \gamma_k + \delta_{jk} + (\alpha\gamma)_{ik} + \theta_{ijk}$$

Where, μ = Over all mean

α_i = effect of ith main plot

β_j = effect of jth main plot

ϵ_{ij} = Error term (1)

γ_k = effect of kth sub plot

δ_{jk} = Error term (2)

$(\alpha\gamma)_{ik}$ = effect of jth main kth sub plot (Interaction)

θ_{ijk} = Error (3)

The strip plot design is specifically suited for two factor experiment in which the desired precision for measuring the interaction effects between the two factors is higher than that of the main effect of two factors.

In strip plot design the interaction is tested with more precision.

R codes

The analysis of variance of a strip plot design is divided into three parts. They are the horizontal factor analysis, vertical factor analysis and the interaction analysis.

Example: The researcher wants to find the efficacy of 3 different sources of organic manure at 4 irrigation schedule. An experiment was conducted in strip plot with 3 replications. Yield in ton per hectare has given. Find the best irrigation level and best manure and its combination. (Kailasam C. and Gangai Selvi R., 2010) [4]

```
>library(agricolae)
> library(datasets)
> data <- read.csv("K:/D
drive/PhD/ASHWINI_ARTICLE/STRIP
PLOT/Stripplot.csv")
> View(data)
> head(data)
Replication Manure Irrigation Yield
1 1 M1 I1 11.2
2 1 M1 I2 10.2
3 1 M1 I3 14.5
4 1 M1 I4 12.3
5 1 M2 I1 11.8
6 1 M2 I2 10.9
>str(data)
'data.frame':  36 obs. of 4 variables:
$ Replication: int 1 1 1 1 1 1 1 1 1 1 ...
$ Manure : chr "M1" "M1" "M1" "M1" ...
$ Irrigation : chr "I1" "I2" "I3" "I4" ...
$ Yield : num 11.2 10.2 14.5 12.3 11.8 10.9 16.2 13.5 11 9.5
...
>as.factor(data$Manure)
[1] M1 M1M1M1 M2 M2M2M2 M3 M3M3M3 M1
M1M1M1 M2 M2M2M2 M3 M3M3M3 M1 M1M1M1 M2
M2M2M2 M3 M3M3M3
Levels: M1 M2 M3
>as.factor(data$Irrigation)
[1] I1 I2 I3 I4 I1 I2 I3 I4 I1 I2 I3 I4 I1 I2 I3 I4 I1 I2 I3 I4 I1 I2
I3 I4 I1 I2 I3 I4 I1 I2 I3 I4 I1 I2 I3 I4
Levels: I1 I2 I3 I4
> model <- with(data, strip.plot(replication, Manure,
Irrigation, Yield))
```

Out put
ANALYSIS STRIP PLOT: Yield
Class level information

Manure : M1 M2 M3
Irrigation   : I1 I2 I3 I4
Replication : 1 2 3
Number of observations: 36
model Y: Yield ~ Replication + Manure + Ea + Irrigation + Eb + Irrigation:Manure + Ec

Analysis of Variance Table
Response: Yield

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Replication	2	0.065	0.0325		
Manure	2	32.382	16.1908	733.17	7.401e-06 ***
Ea	4	0.088	0.0221		
Irrigation	3	83.263	27.7544	715.39	4.735e-08 ***
Eb	6	0.233	0.0388		
Irrigation:Manure	6	10.669	1.7782	37.62	4.241e-07 ***
Ec	12	0.567	0.0473	---	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
cv(a) = 1.2%, cv(b) = 1.6%, cv(c) = 1.8%, Mean = 12.075

Inference: There is highly significant difference between the manures, irrigation schedule and combination of manure + irrigation.

Conclusion

Investigation or experimentation is fundamental part of agricultural research. Here, designs of experiment play salient role to realize the reaction of inputs on output of an experiment. This article provides R code for evaluating

different experimental designs to ease the researcher to continue with excellence in agricultural research. Many agricultural research problems have been incorporated to establish analysis of designs with R codes.

References

1. Zoric AB, *et al*., Applied statistics: Basic principles and application, International Journal of Innovation and Economics Development. 2021;7(3):27-33.
2. Gomez KA, Gomez AA., Statistical procedures for Agricultural Research, Second edition, A Wiley-inter sclence Publication; c1983.
3. Gupta SC, Kapoor VK. Fundamentals of Applied Statistics, Sultan Chand & Sons Publication; c2014.
4. Kailasam C, Gangai Selvi R. Applied Statistics, Kalyani Publishers; c2010.
5. SARE Technical bulletin. How to Conduct Research on Your Farm or Ranch. 2017;(301):779-1007. www.sare.org/research.
6. Sehgal DK. Split plot and Strip plot designs, Split-split Design Macro (iasri.res.in)
7. www.r-bloggers.com
8. www.rdocumentation.org.