RANDOMISED BLOCK DESIGN (RBD)

Amore careful examination of design 2 in table 2 will reveal some disadvantages of CRD in this problem. One thing to be noted is that brand A is never used on car 3 or brand B on car 1 or D on car 2. Any variation within brand A may reflect variation between cars 1, 2 and 4. Thus the random error may not be merely an experimental error but may include variation between cars. Since the chief objective of experimental design is to reduce the experimental error, a better design might be one in which the car variation is removed from the error and a design that requires that each brand be used once on each car is called a Randomised Block Design.

Table 3: Design 3 – RBD for the tyre- type example.

Car

	I	**II**	**III**	**IV**
Brand distribution and loss in thickness	B(14) C(12) A(17) D(13)	D(11) C(12) B(14) A(14)	A(13) B(13) D(11) C(10)	C(9) D(9) B(8) A(13)

In this design the order in which the 4 brands are placed on a car is random and each car gets one tyre of each brand. In this way, better comparisons can be made between brands since they are all driven over the same terrain using the same drivers and so on.
This provides a more homogeneous environment in which to test the 4 brands. In general these groupings for homogeneity are called blocks and the randomisation is now restricted within blocks. The design also allows the car (block) variation to be independently assessed and removed from the error term and the model for the design is;

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

Where y_{ij} - observation from the i^{th} tyre type and the j^{th} car type

μ - general mean

α_i - i^{th} tyre type effect

β_j - j^{th} car type (block) effect

ϵ_{ij} - error term

Advantages

i) The chief advantage of the RBD over the CRD is that it makes an attempt to control the error by identifying a portion of the total variation with the block means.
ii) This design enables the research to use different techniques to different blocks though the techniques should be the same within the block.

iii) The analysis is straight forward and remains so unless due to accident, data (on an entire block or treatment) gets missing. If data from individual units is missing, then we use the Yates missing plot technique to estimate that missing value and then continue with the test and analysis.

iv) Each experimental unit receives each of the treatments assigned in a random sequence. This cuts down considerably on the number of experimental units or subjects needed for the experiment (point of view, cost) => one experimental unit can appear in more than one treatment.

Disadvantages

1. The chief disadvantage is that if the blocks are not internally homogeneous then a large error term will result.

I	**II**
A	D
B	C
C	A
D	B

   If A in I is different from A in II, there will be a large error term. Therefore it is important to make sure that the 4 tyres for each tyre type are homogeneous. Internally homogeneous may also mean same conditions for mostly from the point of view of the environment.

2. With the increase in the number of treatments, the block size increases and therefore one has less control over error and this will increase the probability of including material of a heterogeneous nature.

Layout of RBD

In general, consider a rectangular array of **r** rows and **c** columns. i.e we may have a factor A as the rows with **r** levels and factor B as the columns with **c** levels.

Rows (factor A)	**Columns (factor B)**					**Totals**	**Means**
	1	**2**	**3 …….. j ……… c**				
1	y_{11}	y_{12}	$y_{13} \cdots \cdots y_{1j} \cdot \cdots \cdot \cdot y_{1c}$			y_{1o}	$\bar{y}_{1o}$
2	y_{21}	y_{22}	$y_{23} \cdots \cdots y_{2j} \cdots \cdot \cdot y_{2c}$			y_{2o}	$\bar{y}_{2o}$
3	y_{31}	y_{32}	$y_{33} \cdots \cdots y_{3j} \cdots \cdot \cdot y_{3c}$			y_{3o}	$\bar{y}_{3o}$
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
I	y_{i1}	y_{i2}	$y_{i3} \cdots \cdots y_{ij} \cdots \cdot \cdot y_{ic}$			y_{io}	$\bar{y}_{io}$
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
R	y_{r1}	y_{r2}	$y_{r3} \cdots \cdots y_{rj} \cdots \cdot \cdot y_{rc}$			y_{ro}	$\bar{y}_{ro}$

Totals	$\boldsymbol{y_{o1}}$ $\boldsymbol{y_{o2}}$ $\boldsymbol{y_{o3} \dots \dots y_{oj} . \dots . . y_{oc}}$	$\boldsymbol{y_{oo}}$	
Means	$\boldsymbol{\bar{y}_{o1}}$ $\boldsymbol{\bar{y}_{o2}}$ $\boldsymbol{\bar{y}_{o3} \dots \dots \bar{y}_{o4} \dots \dots \bar{y}_{oc}}$		$\boldsymbol{\bar{y}_{oo}}$

Model

$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$

Where α_i - i^{th} row effect

β_j - j^{th} column effect

y_{ij} is the observation in the i^{th} row and j^{th} column i.e the ij^{th} cell

e.g $y_{23=}$ observation in the 2^{nd} row and 3^{rd} column

y_{io} and $\bar{y}_{io}$ are the total and mean respectively of all the observations in the i^{th} row.

y_{oj} and $\bar{y}_{oj}$ are the total and mean respectively of all observations in the j^{th} column.

y_{oo} and $\bar{y}_{oo}$ are the grand total and mean respectively of all the **rc** observations.

N = rc

Hypotheses

To determine if part of the variation in the observations is due to differences among the rows or due to factor A, we test the following hypothesis

H_{01}: $\mu_{1o} = \mu_{2o} = \mu_{3o} = \dots\dots = \mu_{ro} = \mu$(say) Vs H_{A1}: Some $\mu_{io} \neq \mu$

OR

H_{o1}: $\alpha_1 = \alpha_2 = \alpha_3 = \dots\dots = \alpha_r = 0$ Assuming the α_i are row effects

H_{A1}: Some $\alpha_i \neq 0$

Similarly to determine if part of the variation is due to differences among the columns or due to Factor B

H_{o2}: $\mu_{o1} = \mu_{o2} = \mu_{o3} = \dots\dots = \mu_{oc} = \mu$(say) Vs H_{A2}: Some $\mu_{oj} \neq \mu$

OR

H_{o2}: $\beta_1 = \beta_2 = \beta_3 = \dots\dots = \beta_c = 0$ Assuming the β_j are the column effects

H_{o2}: Some $\beta_j \neq 0$

Rejection Criteria

Reject H_{o1} if $F_{1c} \geq F_{T\alpha\,,\,(r-1),\,(r-1)\,(c-1)}$

Reject H_{o2} if $F_{2c} \geq F_{T\alpha\,,\,(c-1),\,(r-1)\,(c-1)}$

ANALYSIS OF RBD

Note that SST = SSR +SSC +SSE

SST = $\sum_i^r \sum_j^c y_{ij}^2 - \frac{y_{oo}^2}{N}$ from $\sum_i^r \sum_j^c \left(y_{ij} - \bar{y}_{oo}\right)^2$

SSR = $c \sum_{i=i}^{r} \left(\bar{y}_{io} - \bar{y}_{oo}\right)^2$ = $\sum \frac{y_{io}^2}{C} - \frac{y_{oo}^2}{N}$

SSC = $r \sum_{j=i}^{c} \left(\bar{y}_{oj} - \bar{y}_{oo}\right)^2$ $\Rightarrow$ $\sum \frac{y_{oj}^2}{r} - \frac{y_{oo}^2}{N}$

SSE = SST – SSR – SSC = $\sum_i \sum_j y_{ij}^2 - \frac{\sum y_{io}^2}{c} - \frac{\sum y_{oj}^2}{r} + \frac{y_{oo}^2}{N}$

Analysis of Variance table

Source of variation	**Df**	**SS**	**MS**	**F-ratio**
Rows (factor A)	r –1	SSR	MSR = $\frac{SSR}{r-1}$	$F_{1c} = \frac{MSR}{MSE}$
Columns (factor B)	c –1	SSC	MSC = $\frac{SSC}{c-1}$	$F_{2c} = \frac{MSC}{MSE}$
Error	(r-1) (c-1)	SSE	MSE = $\frac{SSE}{(r-1)\,(c-1)}$	–
Total	**N-1 or (rc)-1**	**SST**	–	–

Critical region

Reject H_{o1} if $F_{1c} \geq F_{T\alpha\,,\,(r-1),\,(r-1)\,(c-1)}$

Reject H_{o2} if $F_{2c} \geq F_{T\alpha\,,\,(c-1),\,(r-1)\,(c-1)}$

<u>Example</u>

Test for the significance of the car type effect and the tyre type effect on tread loss.

Car types	**Tyre types**				
	A	**B**	**C**	**D**	y_{io}
I	17	14	12	13	56
II	14	14	12	11	51
III	13	13	10	11	47

IV	13	8	9	9	39
$\boldsymbol{y_{ij}}$	**57**	**49**	**43**	**44**	$\boldsymbol{y_{oo}}$ **= 193**

Model

y_{ij} = μ + α_i+ β_j +$\in_{ij}$

Columns – tyre types

Rows – car types

Where y_{ij} – observation for the i^{th} car type and the j^{th} tyre type

μ - general mean

α_i - i^{th} car type effect

β_j - j^{th} tyre type effect

$\in_{ij}$ – experimental error

Hypotheses

H_{o1}: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ Vs H_{AI}: Some $\alpha_i \neq 0$

OR

H_{o2}: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ Vs H_{A2}: Some $\beta_j \neq 0$

Critical region

Reject H_{o1} if $F_{1c} \geq F_{0.05, 3, 9} = 3.86$ $F_{0.01, 3, 9} = 6.99$

Reject H_{o2} if $F_{2c} \geq F_{0.05, 3, 9} = 3.86$

Computations

Correction factor = $\frac{y_{oo}^2}{N} = \frac{193^2}{16}$ = 2328.06

SST = $\Sigma_i\Sigma_j\, y_{ij}^2 - \frac{y_{oo}^2}{N}$ = [$17^2 + 14^2$ + ... + 9^2] – C.F = 2409 – 2328.06 = 80.94

SSR = $\sum\frac{y_{io}^2}{C} - \frac{y_{oo}^2}{N} = \frac{[56^2+51^2+47^2+39^2]}{4}$ − C.F = 2366.75−C.F = 38.69

SSC = $\sum\frac{y_{oj}^2}{r} - \frac{y_{oo}^2}{N} = \frac{[57^2+49^2+43^2+44^2]}{4}$ − C.F = 2358.75− C.F = 30.69

SSE = SST – SSR – SSC = 80.94 − 38.69 − 30.69 = 11.56

ANOVA Table

Source of variation	**df**	**ss**	**ms**	**F-ratio**

Rows (car type)	3	38.69	12.9	$F_{1c} = 9.9$
Columns (tyre type)	3	30.69	10.2	$F_{2c} = 7.8$
Error	9	11.56	MSE = 1.3	
Total	**15**	**80.94**		

Conclusions

To test the hypothesis that the average tread loss of all the four car types is the same, the computed F obtained was 9.9 which is also significant at the 1% level of significance. This means that the car to car variation is significant since we reject H_{o1}.

For the case of the tyre types $F_{2c} = 7.8$ which is also significantly larger than the corresponding critical F given. i.e The hypothesis of equal tyre type means H_{o2} is also rejected at both 1% and 5% level of significance.

Note that this hypothesis of equal tyre type means could not be rejected using a CRD.

The RBD allows for the removal of the car effect which has reduced the common variance significantly from 4.2 to 1.3.

Remark

When data are presented in a tabular form, there is usually no way to determine how the data were collected. Was the randomisation complete over all N observations or was the experiment run in blocks with randomisation restricted to within the blocks?

To help in signifying the design of the experiment, it is suggested that in the case of a CRD, no vertical or horizontal lines be drawn as in table 2. When randomisation has been restricted, either vertical or horizontal lines (as shown in table 3) can be used to indicate this order of restriction on the randomisation.