Randomized Blocked Designs

Hammou ElBarmi

Columbia University

Introduction

- For randomized block designs, there are some factors that are of primary interest. However, there are also several other nuisance factors.
- Nuisance factors are those that may affect the measured result, but are not of primary interest.
- For example, in applying a treatment, nuisance factors might be the specific operator who prepared the treatment, the time of day the experiment was run, and the room temperature.

Introduction

- All experiments have nuisance factors. The experimenter will typically need to spend some time deciding which nuisance factors are important enough to keep track of or control, if possible, during the experiment.
- Many times there are nuisance factors that are unknown and uncontrollable (sometimes called a "lurking" variable).
- We use randomization to balance out their impact. We always randomize so that every experimental unit has an equal chance of being assigned to a given treatment.
- Randomization is our insurance against a systematic bias due to a nuisance factor.
- When we can control nuisance factors, blocking can be used to reduce or eliminate the contribution to experimental error contributed by nuisance factors.

Introduction

- The basic concept is to create homogeneous blocks in which the nuisance factors are held constant and the factor of interest is allowed to vary.
- Within blocks, it is possible to assess the effect of different levels of the factor of interest without having to worry about variations due to changes of the block factors, which are accounted for in the analysis.
- Definition of blocking factors: A nuisance factor is used as a blocking factor if every level of the primary factor occurs the same number of times within each level of the nuisance factor.

Introduction

- The analysis of the experiment will focus on the effect of varying levels of the factor that is of primary interest
- Suppose a researcher is interested in how several treatments affect a continuous response variable (Y).
- The treatments may be the levels of a single factor or they may be the combinations of levels of several factors. Suppose we have a total of k
- Suppose we have available to us a total of kn experimental units to which we are going to apply the different treatments.
- The Completely Randomized (CR) design randomly divides the experimental units into k groups of size n and randomly assigns a treatment to each group.

Example 1:

- Suppose we are interested in how weight gain (Y) in rats is affected by Source of protein (Beef, Cereal, and Pork) and by Level of Protein (High or Low).
- There are a total of $k = 6$ treatment combinations of the two factors (Beef -High Protein, Cereal-High Protein, Pork-High Protein, Beef -Low Protein, Cereal-Low Protein, and Pork-Low Protein) .
- Suppose we have available to us a total of 60 experimental rats to which we are going to apply the different diets based on the $k = 6$ treatment combinations.
- Prior to the experimentation the rats were divided into $b = 10$ homogeneous groups (blocks) of size 6
- The grouping was based on factors that had previously been ignored (Example - Initial weight size, appetite size etc.)
- Within each of the 10 blocks a rat is randomly assigned a treatment combination (diet).
- The weight gain after a fixed period is measured for each of the test animals.

Example

Block	T1	T2	T3	T4	T5	T6
1	107	96	112	83	87	90
2	102	72	100	82	70	94
3	102	76	102	85	95	86
4	93	70	93	63	71	63
5	111	79	101	72	75	81
6	128	89	104	85	84	89
7	56	70	72	64	62	63
8	97	91	92	80	72	82
9	89	61	87	82	81	61
10	103	102	112	83	93	81

Example 2:

- The following experiment is interested in comparing the effect of four different chemicals (A, B, C and D) in producing water resistance (y) in textiles.
- A strip of material, randomly selected from each bolt, is cut into four pieces (samples) the pieces are randomly assigned to receive one of the four chemical treatments.
- This process is replicated three times producing a Randomized Block (RB) design. Moisture resistance (y) were measured for each of the samples. (Low readings indicate low moisture penetration).
- the data is

Block	A	B	C	D
1	10.1	11.4	9.9	12.1
2	12.2	12.9	12.3	13.4
3	11.9	12.7	11.4	12.9

Model

- The model for the randomized block design is

$$y_{ij} = \mu + \alpha_i + b_j + \epsilon_{ij}, i = 1, 2, \ldots, k, \ j = 1, 2, \ldots, b.$$

  where
  - y_{ij} = the observation in the jth block receiving the ith treatment
  - μ = the over all mean
  - α_i = the effect of the ith treatment
  - b_j the effect of the jth block
  - ϵ_{ij} = random error

Analysis of Variance Table

The ANOVA table is given by

Source	df	SS	MS	F
Treatment	k-1	SSTR	MSTR	MSTR/MSE
Block	b-1	SSBL	MSBL	MSBL/MSE
Error	(k-1)(b-1)	SSE	MSE	
Total	bk-1	SST		

Analysis of Variance Table

- A randomized block experiment is assumed to be a two-factor experiment. The factors are blocks and treatments.
- The is one observation per cell. It is assumed that there is no interaction between blocks and treatments.
- The degrees of freedom for the interaction is used to estimate error.

Example 1 (continued

The ANOVA for the diet experiment (example 1) is

Source	df	SS	MS	F	p-value
Diet	5	4572.883	914.577	13.077	0.00000
Block	9	5992.417	665.824	9.52	0.00000
Error	45	3147.283	69.400		
Total	49	13712.58			

Example 2 (continued)

The ANOVA for the textile experiment (example 2) is

Source	df	SS	MS	F	p-value
Chem	3	5.20	1.73	19.44	0.0017
Block	2	7.17	3.58	40.21	0.0003
Error	6	0.54	0.09		
Total	11	12.91			

Example 1 (continued

- If the treatments are defined in terms of two or more factors, the treatment Sum of Squares can be split (partitioned) into main effects and interactions
- Example (Diet continued): here we can assume that we have two treatments (Source= Beef, Cereal or Pork and Level= High or Low and fit the model

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + b_k + \epsilon_{ijk}, i = 1, 2, 3, j = 1, 2, k = 1, 2, \ldots, 10.$$

  where
  - y_{ijk} = the observation in the kth block receiving the (i,j) treatment
  - μ = the over all mean
  - α_i = the effect of the ith Source
  - β_j = the effect of the jth level
  - b_k the effect of the kth block
  - ϵ_{ijk} = random error
  - $(\alpha\beta)_{ij}$ = interaction terms.

The ANOVA table is

Source	df	SS	MS	F	p-value
Source	2	882.223	441.117.577	6.31	0.00380
Level	1	2680.017	2680.017	38.32	0.00000
SourcexLevele	2	1010.633	505.317	7.23	0.00190
Block	9	5992.417	665.824	9.52	0.00000
Error	45	3147.283	69.400		
Total	49	13712.58			

- The latin square design generalizes the randomized block design by having two blocking variables in addition to the treatment variable of interest.
- The latin square design has a three way layout, making it similar to a three way ANOVA.
- The latin square design has only one treatment for block.

- The latin square design only permits three main effects to be estimated because the design is incomplete.
- It is an "incomplete" factorial design because not all cells are represented.
- For example, in a 4x4x4 factorial design there are 64 possible cells. But, a latin square design where the treatment variable has 4 levels and each of the two blocking variables have 4 levels, there are only 16 cells.
- This type of design is useful when it is not feasible to test all the cells required in a three-way ANOVA, say, because of financial or time constraints in conducting a study with many cells.

- The number of levels for each blocking factor is determined by the number of treatments.
- A Latin Square plan (or layout) starts with a systematic arrangement of the treatments in rows (levels of one blocking variable) and columns (levels of the second blocking variable).
- Latin Square designs are similar to randomized block designs, except that instead of the removal of one blocking variable, these designs are carefully constructed to allow the removal of two blocking factors.
- They accomplish this while reducing the number of experimental units needed to conduct the experiment.

- The advantages of Latin square designs are: They allow experiments with a relatively small number of runs.
- The disadvantages are: The number of levels of each blocking variable must equal the number of levels of the treatment factor.
- The Latin square model assumes that there are no interactions between the blocking variables or between the treatment variable and the blocking variable.
- Example with three treatments A, B, C

A	B	C
B	C	A
C	A	B

- Note that in a Latin Square plan that each treatment appears once in each row and once in each column

A plant biologist conducted an experiment to compare the yields of four varieties of peanuts (A, B, C, D). A plot of land was divided into 16 subplots (4 columns and 4 rows) and the following latin square was run

	Column			
Row	E	EC	WC	W
N	C	A	B	D
NC	A	B	D	C
SC	B	D	C	A
S	D	C	A	B

The data is

	Column			
Row	E	EC	WC	W
N	26.7	19.7	29.0	29.8
NC	23.1	21.7	24.9	29.0
SC	29.3	20.1	29.0	27.3
S	25.1	17.4	28.7	35.1

- The model is

$$y_{ijk} = \mu + \rho_i + \gamma_j + \tau_k + \epsilon_{ijk}, i = 1, 2, \ldots, p, j = 1, 2 \ldots, p, k = 1, 2, \ldots k.$$

  with row effect
  - ρ_i = row effect
  - γ_j = column effect
  - τ_k = treatment effect
  - ϵ_{ijk} = the error term
- The error terms are assumed to be independent with mean 0 and variance σ^2
- In addition

$$\sum_{i=1}^{p} \rho_i = \sum_{i=1}^{p} \gamma_i = \sum_{i=1}^{p} \tau_i = 0.$$

- The sums of squares decomposition for the latin square design is

$$SST = SS_{trt} + SS_{row} + SS_{col} + SSE$$

The sums of squares are

$$
\begin{aligned}
SSTR &= \sum_{k=1}^{p} (Y_{\bullet\bullet k} - \bar{Y}_{\bullet\bullet\bullet})^2 \\
SS_{row} &= \sum_{i=1}^{p} (Y_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 \\
SS_{col} &= \sum_{j=1}^{p} (Y_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 \\
SSE &= \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} (Y_{ijk} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet k} + 2\bar{Y}_{\bullet\bullet\bullet})^2 \\
SST &= \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} (Y_{ijk} - \bar{Y}_{\bullet\bullet\bullet})^2
\end{aligned}
$$

The Analysis of Variance Table:

Source	df	SS	MS	F	p-value
Treatment	p-1	$SSTR$	MSTR= SSTR/(p-1)	MSTR/MSE	
Row	p-1	SS_{row}	$MS_{row} = SS_{row}/(p-1)$		
Column	p-1	SS_{col}	$MS_{col} = SS_{col}/(p-1)$		
Error	(p-1)(p-2)	SSE	$MSE = SSE/(p-1)(p-2)$		
Total	p^2-1	SST			

Latin Square Designs-Example

```
> row<-c(rep("N",4), rep("NC",4), rep("SC",4), rep("S",4))
> row
 ["N"  "N"  "N"  "N"  "NC" "NC" "NC" "NC" "SC" "SC" "SC" "SC"
 "S"  "S"  "S"  "S"
> col<-rep(c("E","EC","WC","W"),4)
> col
  "E"  "EC" "WC" "W"  "E"  "EC" "WC" "W"  "E"  "EC" "WC" "W"
    "E"  "EC" "WC" "W"
> yield<-c(26.7,19.7,29.0, 29.8, 23.1,21.7, 24.9, 29.0, 29.3,20.1,
29.0,27.3, 25.1, 17.4,28.7,35.1 )
> fit<-lm(y~trt+row+col)
> Anova(fit)
Anova Table (Type II tests)

Response: y
           Sum Sq Df F value   Pr(>F)
trt        42.667  3  3.5580 0.086997 .
row         9.427  3  0.7861 0.543940
col       245.912  3 20.5065 0.001483 **
Residuals  23.984  6
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
```