Topic 7: Incomplete, double-blocked designs: Latin Squares

[ST&D sections 9.10 – 9.15]

7.1. Introduction

The Randomized Complete Block Design is commonly used to improve the ability of an experiment to detect real treatment differences by partitioning a known source of variation (blocks) from the experimental error. When this idea is extended to remove two different known sources of variations (i.e. two-way blocking), one resulting design is the **Latin square**. In this design, the randomization of treatments is further restricted, relative to an RCBD, due to orthogonality requirements with *both* blocking variables. In this discussion, we will refer to the two blocking variables in a general Latin Square as *rows* and *columns*. The experiment is organized such that each treatment appears exactly once within each row and within each column. Similar to RCBD's with one replication per block-treatment combination, Latin Squares have one replication per row-column-treatment combination. Unlike RCBD's, Latin Squares are not complete designs; they are an example of an incomplete blocked design.

A Latin Square with treatments assigned to the first row and the first column in an alphabetical or numerical sequence is called a **standard square.** Figure 1 shows the standard squares for 2 x 2, 3 x 3 and 4 x 4 designs. Treatment levels are indicated by A, B, C, etc.

2x2

A	**B**
B	A

3x3

A	**B**	**C**
B	C	A
C	A	B

4x4

A	**B**	**C**	**D**
B	A	D	C
C	D	B	A
D	C	A	B

A	**B**	**C**	**D**
B	C	D	A
C	D	A	B
D	A	B	C

A	**B**	**C**	**D**
B	D	A	C
C	A	D	B
D	C	B	A

A	**B**	**C**	**D**
B	A	D	C
C	D	A	B
D	C	B	A

Figure 1. Standard Latin squares for 2 x 2, 3 x 3 and 4 x 4.

As the size of the square increases, the number of possible standard squares increases rapidly. For any given square size (t x t), the number of possible different Latin squares is:

$$(\text{\# of standard squares})\ (t!)\ (t - 1)!$$

where t is the number of treatments. For example, for t = 4, the number of total possible unique squares is: (4) (4!) (3!) = 576.

7.2. Examples

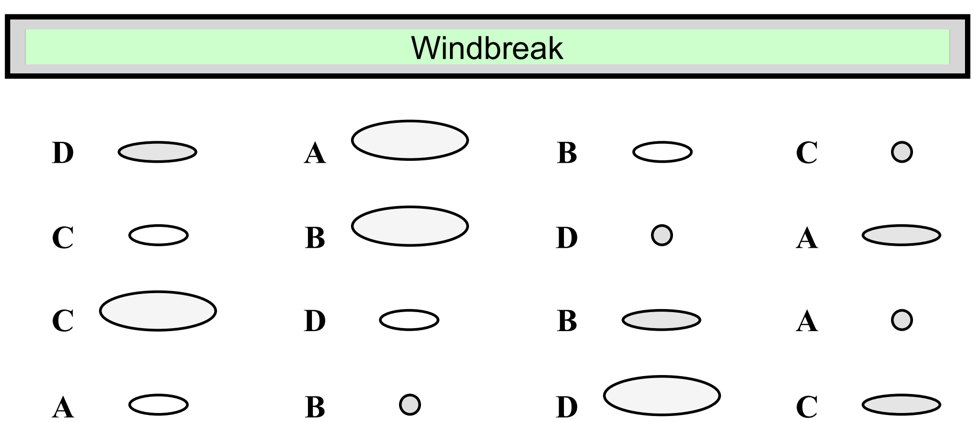
Sometimes, part of the native variability among EU's within an experiment can be attributed to two distinct factors. For example, natural fertility and soil texture may vary in independent, systematic ways within a field. When comparing three cultivars, these two sources of variation can be separated from the random variation of experimental error by arranging the levels of fertility and soil texture into blocks, as shown in Table 1.

Table 1 A 3 x 3 Latin Square with treatments (cultivars C1, C2 and C3) randomized according to two blocking variables: level of fertility (low, medium, high = columns) and gradient of soil texture (1, 2, 3 = rows).

Soil Texture	Fertility		
	L	**M**	**H**
1	C3	C1	C2
2	C2	C3	C1
3	C1	C2	C3

Note that Latin Squares require the same number of rows, columns, and treatments. Requiring the number of treatment levels to equal the number of levels in each of two separate blocking variables is a severe design constraint indeed. Recognize, however, that this arrangement does not necessarily have to be a physical square. For example, if trees in an orchard can be classified in terms of both size and distance from a windbreak, a Latin Square may be appropriate. The diagram below is the field map of such an experiment, where the four treatment levels (A – D) were assigned to EU's according to a Latin Square:

Figure 2 Sixteen orchard trees classified (i.e. blocked) by size and distance from a windbreak. 4 treatment levels (A - D) are assigned according to a Latin Square. The size of a circle indicates tree size.



As the above diagram shows, rows and columns do not necessarily refer to the spatial distribution of the experimental units. They can also refer to the temporal order in which treatments are performed, to different pieces of equipment used in the experiment, to different technicians taking the measurements, etc.

Latin Squares are commonly used in sensory panels. In these situations, the materials or treatments to be tested (e.g. different wines, fruits, etc.) can be blocked by evaluators (judges) and days so that differences among judges and differences among days of testing can be removed from experimental error.

Table 2 A 3 x 3 Latin Square with treatments (Wines A, B, C) randomized according to two blocking variables: Judge (Joe, Laura, Rose = columns) and day of testing (M, W, F = rows).

Day	**Judge**		
	Joe	**Laura**	**Rose**
Monday	B	C	A
Wednesday	C	A	B
Friday	A	B	C

If the *order* in which judges taste the different products is thought to affect their evaluations, sequence of tasting could be a blocking variable:

Table 3 A 3 x 3 Latin Square with treatments (Wines A, B, C) randomized according to two blocking variables: Judge (Joe, Laura, Rose = columns) and *sequence* of tasting (1st, 2nd, 3rd = rows). In this design, each wine has the chance to be the first, second, and third in the tasting sequence, thereby removing variation due to sequence.

Sequence	**Judge**		
	Joe	**Laura**	**Rose**
First	B	C	A
Second	C	A	B
Third	A	B	C

Marketing studies sometimes use Latin Squares with days and stores as the two blocking variables. Indeed, absolutely anything that can function as a block in an RCBD can function as a block in a Latin Square, assuming you can satisfy the strict requirements of the design (rows = columns = treatments; and each treatment appears exactly once in each row and exactly once in each column).

Consider an experiment to test the effect of 4 different computer keyboard key spacings (S1 – S4) on typing speed. Four people are recruited to use the four new keyboard designs. Because different people type at different speeds, independent of keyboard design, the researchers decide to block by people. For each test, the volunteer is provided a printed page of text which he or she must type. Because the content of the text (e.g. word size, sentence complexity, etc.) may affect typing speed, independent of keyboard design, the researchers decide to use four different pages of text in the experiment and to block by text as well.

	Volunteer			
Text	**1**	**2**	**3**	**4**
1	S2	S3	S4	S1
2	S3	S4	S1	S2
3	S4	S1	S2	S3
4	S1	S2	S3	S4

7.3 Randomization

Proper randomization is crucial to the validity of conclusions to be drawn from any experiment, Latin Squares included. Randomization is used both to neutralize the effects of any systematic biases as well as to meet the assumption of independence underlying the analysis.

7.3.1 Manual method of randomization

The only restriction on the Latin square arrangement is that each treatment must appear in every row and every column of the table. The number of possible Latin squares increases rapidly as the size of the square increases, and it is necessary to choose one of these possible arrangements at random. A procedure that gives a satisfactorily randomized square is the following:

1. Arbitrarily select a standard square for the number of treatments involved. For example, if 4 treatments are involved, the following standard square could be selected:

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

2. From a table of random numbers or by some other procedure, select two sets of random numbers with size equal to the treatments involved. Then assign ranks to these sets of random numbers. For example:

	Set 1 (for columns)	Set 2 (for rows)
Random number	9 1 6 4	8 5 3 7
Rank	4 1 3 2	4 2 1 3

3. Assign the ranks to the rows and column headers of the standard square chosen in Step 1 (a). Order the rows according to rank (b). Order the columns according to rank (c).

a. Assign ranks

	4	1	3	2
4	A	B	C	D
2	B	C	D	A
1	C	D	A	B
3	D	A	B	C

b. Order rows

	4	1	3	2
1	C	D	A	B
2	B	C	D	A
3	D	A	B	C
4	A	B	C	D

c. Order columns

	1	2	3	4
1	D	B	A	C
2	C	A	D	B
3	A	C	B	D
4	B	D	C	A

4. Finally, once this new Latin Square has been generated, the generic treatment ID's (A – D) are randomly assigned to the 4 treatments. The square thus shows the how the treatments should be assigned to the experimental units.

7.3.2 Automated method of randomization

The following R script creates a function Latin() that generates a template for an nxn Latin square design.

```
#This script defines a function Latin() that can be used to generate LS randomizations

Latin <- function (x)
{
  LS = matrix(LETTERS[1:x], x, x)
  LS = t(LS)
  for (i in 2:x) LS[i, ] = LS[i, c(i:x, 1:(i - 1))]
  for (i in 1:20) {
      LS = LS[sample(x), ]
      LS = LS[, sample(x)]
  }
  LS
}

#Example: For a 5x5 LS:
Latin(5)
```

Output:

```
> Latin(5)
     [,1] [,2] [,3] [,4] [,5]
[1,] "B"  "E"  "A"  "C"  "D"
[2,] "A"  "D"  "E"  "B"  "C"
[3,] "C"  "A"  "B"  "D"  "E"
[4,] "E"  "C"  "D"  "A"  "B"
[5,] "D"  "B"  "C"  "E"  "A"
```

7.4 The linear model

The linear model for the Latin Square:

$$Y_{(i)jk} = \mu + \tau_{(i)} + \rho_j + \kappa_k + \varepsilon_{(i)jk}$$

where $Y_{(i)jk}$ represent the observation in the j^{th} row and k^{th} column, ρ_j represents j^{th} row effect, κ_k represents k^{th} column effect, and $\tau_{(i)}$ represents the i^{th} treatment effect. The parentheses around the index "i" is due to the incomplete nature of this design; in any given row-column combination, only one treatment level occurs. In dot notation:

$$\sum_{i,j}^{r}(\overline{Y}_{ij}-\overline{Y}_{..})^2 = r\sum_{i=1}^{r}(\overline{Y}_{i}-\overline{Y}_{..})^2 + r\sum_{j=1}^{r}(\overline{Y}_{j.}-\overline{Y}_{..})^2 + r\sum_{k=1}^{r}(\overline{Y}_{.k}-\overline{Y}_{..})^2 + \sum_{i,j}^{t}(Y_{ij}-\overline{Y}_{i}-\overline{Y}_{j.}-\overline{Y}_{.k}+2\overline{Y}_{..})^2$$

$$\text{TSS} \quad = \quad \text{SST} \quad + \quad \text{SS}_{\text{rows}} \quad + \quad \text{SS}_{\text{columns}} \quad + \quad \text{SSE}$$

7.5 ANOVA

The generic ANOVA table for a Latin Square looks like this:

Source	df	SS	MS	F
Rows	r - 1	SSR	SSR/(r-1)	MSR/MSE
Columns	r - 1	SSC	SSC/(r-1)	MSC/MSE
Treatments	r - 1	SST	SST/(r-1)	MST/MSE
Error	(r-1)(r-2)	SSE	SSE/(r-1)(r-2)	
Total	r^2-1	TSS		

The ANOVA from for the example provided on ST&D page 230 is included below. This 4 x 4 Latin Square for yields of four wheat varieties shows highly significant differences among varieties, no significant difference among rows, and marginally significant differences among columns.

Source	DF	SS	MS	F Value	Pr > F
ROW	3	1.955	0.652	1.44	0.322
COL	3	6.800	2.267	5.00	0.045
TRTMT	3	78.925	26.308	58.03	0.000
Error	6	2.720	0.453		
Total	15	90.400			

There are 4 identified sources of variation in this example, three due to the design (rows, columns, and treatments, each with 3 df) and one due to error (6 df). For testing the hypotheses that there are no column, row, or treatment differences, the mean squares for each of these main effects are divided by the MSE.

7.6 Advantages and disadvantages of the Latin square

When two kinds of heterogeneity that are either due to the nature or arrangement of experimental units can be identified, the Latin square is an effective design for partitioning those sources of variation from the experimental error. Some disadvantages of Latin squares:

1. Severe design constraints. The number of treatments must be equal to the number of rows and columns. This restriction imposes an inconvenience for actual experimental work, particularly for experiments with a large number of treatments.

2. The design is only valid if there are no interactions among rows, columns, and treatments. **When there is significant interaction between any two of these factors, no valid tests of significance are possible**.

3. Squares smaller than 4 x 4 generally have too few replications for a desirable level of precision. This is reflected in the few degrees of freedom available for experimental error in the ANOVA.

ANOVA for the completely randomized design (CRD), randomized complete block design (RCBD), and the Latin Square design (LS) are shown schematically below. Assume 5 treatments are assigned to 25 experimental units, with equal replication per treatment.

Total Sum of Squares
df = 24

CRD SS Treatment df = 4 | SS Experimental Error **df = 20**

RCBD SS treatment df = 4 | SS block df = 4 | SS Experimental error **df =16**

LS SS treatment df = 4 | SS Column df = 4 | SS Rows df = 4 | SS Experimental Error **df = 12**

This diagram illustrates the loss of degrees of freedom from the experimental error as the chosen design becomes more complex. Unless there is enough variation in the experimental error of the CRD that can be allocated to the chosen blocking variable(s) to compensate for the loss of degrees of freedom from the experimental error, the resulting design will be less efficient.

A numerical example of the effect of the missing df_e on the critical F value:

3x3 Latin Square	**5x5 Latin Square**
LS $df_e = 2$ (RCBD $df_e = 4$)	LS $df_e = 12$ (RCBD $df_e = 16$)
LS $F_{0.05,2,2} = 19.0$	LS $F_{0.05,2,12} = 3.89$
RCBD $F_{0.05,2,4} = 6.94$	RCBD $F_{0.05,2,16} = 3.63$

This loss of df_e has a huge effect on the sensitivity of small Latin Squares!

7. 7. Relative Efficiency

Section 9.3 of ST&D (p237) provides a good discussion for the relative efficiencies of Latin Squares compared to RCBD. Once an experiment is conducted as a Latin Square, the researcher can ask the question of how efficient it was relative to an RCBD for that particular set of data. The Latin Square can be converted into an RCBD (i.e. simplified) in one of two ways, either by removing the rows as a blocking variable or by removing the columns as a blocking variable. In either case, the estimated MSE for the alternative RCBD needs to be calculated. In the example from ST&D page 230, $MSE_{LS} = 0.45$, $MSR_{LS} = 0.65$, and $MSC_{LS} = 2.27$.

If **columns** are removed as a blocking variable, the estimated MSE for the alternative RCBD is:

$$M\hat{S}E_{RCBD} \cong \frac{df_{col} * MSC_{LS} + (df_{trt} + df_{error})MSE_{LS}}{df_{col} + df_{trt} + df_{error}} = \frac{3*2.27 + (3+6)\ 0.45}{3+3+6} = 0.91$$

and the relative efficiency of the two designs is:

$$RE_{LS:RCBD} = \frac{(df_{e(LS)} + 1)(df_{e(RCBD)} + 3)MSE_{RCBD}}{(df_{e(RCBD)} + 1)(df_{e(LS)} + 3)MSE_{Ls}} = \frac{(6+1)(9+3)0.91}{(9+1)(6+3)0.45} = 1.89$$

This indicates that the column grouping increased the precision of the model by an estimated 89%. This parallels the significant effect of columns identified in the F tests (see previous ANOVA table).

If **rows** are removed as a blocking variable, the estimated MSE for the alternative RCBD is:

$$M\hat{S}E_{RCBD} \cong \frac{df_{row} * MSR_{LS} + (df_{trt} + df_{error})MSE_{LS}}{df_{row} + df_{trt} + df_{error}} = \frac{3*0.65 + (3+6)\ 0.45}{3+3+6} = 0.50$$

and the relative efficiency of the two designs is:

$$RE_{LS:RCBD} = \frac{(6+1)(9+3)0.50}{(9+1)(6+3)0.45} = 1.04$$

This indicates that the row blocking increased the precision of the model only by an estimated 4%. This parallels the non-significant effect of rows identified in the F tests (see previous ANOVA table).

7.7 Repeated Latin squares

In cases where the number of treatments is small but considerable variability is expected to enter the experiment from two factors (two blocking variables), two or more Latin squares may be used to increase the degrees of freedom for experimental error. This strategy is known as repeated Latin Squares, where each individual Latin Square can be thought of as an independent replication of the entire experiment.

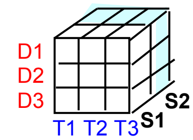
Repeated Latin Squares can relate to one another in four possible ways, depending on how the levels of the blocking variables relate across squares (i.e. experiments):

1. They share the same rows but not the same columns (commons rows, independent cols)
2. They share the same columns but not the same rows (commons cols, independent rows)
3. They share the same rows and the same columns (commons rows and cols)
4. They share neither the same rows nor the same columns (independent rows and cols)

Example: Three gasoline additives (TREATMENTS: A,B,C) were tested for gas efficiency by three drivers (ROWS: 1,2,3) using three different tractors (COLUMNS: 1,2,3). The variable measured was the yield of carbon monoxide in a trap. The experiment was repeated on two separate days (SQUARES: 1,2). The possible replication scenarios:

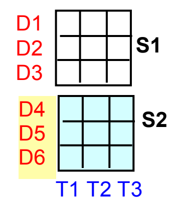
Case 1: Same tractors (col) and same drivers (row)

```
LS1_mod<-lm(CO ~ Day + Tractor + Driver + Trtmt, LS_dat)
anova(LS1_mod)
```



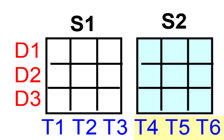
Case 2: Same tractors (col) but different drivers (row)

```
LS4_mod<-lm(CO ~ Day + Tractor + Day:Driver + Trtmt, LS_dat)
anova(LS4_mod)
```



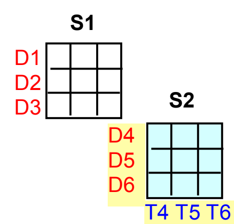
Case 3: Different tractors (col) but same drivers (row)

```
LS3_mod<-lm(CO ~ Day + Day:Tractor + Driver + Trtmt, LS_dat)
anova(LS3_mod)
```



Case 4: Different tractors (col) and different drivers (row)

```
LS4_mod<-lm(CO ~ Day + Day:Tractor + Day:Driver + Trtmt, LS_dat)
anova(LS4_mod)
```



The complete R programs for each of these different cases will be covered in lab. In the meantime, to convince yourself that R is doing the correct calculations for each of the above scenarios, take the following dataset:

Square 1		
B 26.0	C 25.0	A 21.3
C 28.7	A 23.6	B 28.5
A 25.3	B 28.4	C 30.1

Square 2		
C 32.4	B 28.7	A 25.8
B 31.7	A 24.3	C 30.5
A 24.9	C 29.3	B 29.2

Analyze this data assuming Case 2 (shared tractors, independent drivers). Keep your output.

Next, replace the IDs of the rows/drivers (1, 2, 3) in the second square by new IDs (4, 5, 6) to manually differentiate them from the drivers in square one. In the lm() statement, change "square/row" simply to "row".

You will obtain identical results with both analyses. The syntax in the first approach [square/row] tells R that the ID "row" only has meaning once the square has been specified (i.e. Driver 1 in Square 1 is not the same as Driver 1 in Square 2). In the second approach, changing the row IDs of the second square accomplishes the same thing. Clearly, Driver 1 is not the same as Driver 4.

One consequence of replicating a Latin Square is that an additional classification variable enters the analysis (Square). In essence, this new variable functions as yet a *third* blocking variable, partitioning from the experimental error any variation due to overall differences among experiments. For example, if the first round of tests (Square 1) were performed on Monday and the second round of tests (Square 2) were performed on Tuesday, the "Square" variable would capture any variability in trapped CO due to differences between the days (e.g. different weather, ambient pressure, etc.).

7.8 More complicated Latin squares (Box, Hunter, & Hunter Chapter 8)

To partition more than two independent sources of variability from the experimental error, more complicated designs like the Graeco-Latin Square (GLS) or the Hyper-Graeco-Latin Square (HGLS) may be useful. A GLS is a highly organized, incomplete t x t pattern that permits the study of t treatments simultaneously while accounting for three separate blocking variables. If even more blocking variables are considered, an HGLS results.

An example of a 4 x 4 Graeco-Latin square could be the comparison of 4 different oil additives (Treatments A, B, C, D), blocked by 4 different drivers (I, II, III, IV), 4 different cars (1, 2, 3, 4), and 4 different days (α, β, γ, δ).

	Car			
Driver	**1**	**2**	**3**	**4**
I	Aα	Bβ	Cγ	Dδ
II	Bδ	Aγ	Dβ	Cα
III	Cβ	Dα	Aδ	Bγ
IV	Dγ	Cδ	Bα	Aβ

Note in this particular case that a quadruple-replicated 4x4 Latin Square could accomplish a similar objective (Squares = Days) but would require four times the resources (64 driving tests, 16 on each of four days, vs. 16 driving tests, 4 on each of four days). On the other hand, replicated Latin Squares with shared cars and drivers would provide **much more power** (Replicated LS $df_e = 51$ vs. GLS $df_e = 3$). Here again we see the large penalty of small squares on power.