Topic 7. Double grouping. Latin Squares [ST&D 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 removing a known source of variation (blocks) from experimental errors. When this idea is extended to remove two known sources of variations by blocking in two ways, a resulting design, among others, is a **Latin square**. In this design the randomization of treatments is restricted further by grouping them into columns as well as rows. Each row and each column receives each treatment once.

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.** Fig. 1 shows the standard squares for 2 x 2, 3 x 3 and 4 x 4 designs.

A	B
B	A

A	B	C
B	C	A
C	A	B

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

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

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

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

Figure 2. 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. The number of possible different Latin squares is equal to:

(# of standard squares) (K!) (K - 1)!

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

7. 2. Examples

At times there are reasons to believe that the experimental units will respond to a treatment differently due to more than one factor. For example, shading and soil texture may vary in some systematic way within a field. In comparing three cultivars, these sources of variation can be separated from the random variation of experimental error by arranging the degree of shading and soil texture into blocks as shown in Table 1.

Table1 1. 3 x 3 Latin square with cultivars (C1, C2 and C3) arranged in blocks according to degree of shading (low, medium, and high) and gradient of soil texture (1, 2 and 3).

	Degree of shading		
Soil Texture	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 and columns as the number of treatments. However, the arrangement does not necessarily have to be a physical square. If trees in an orchard can be classified as to size and distance from a windbreak, a Latin square may be used in treatment assignment.

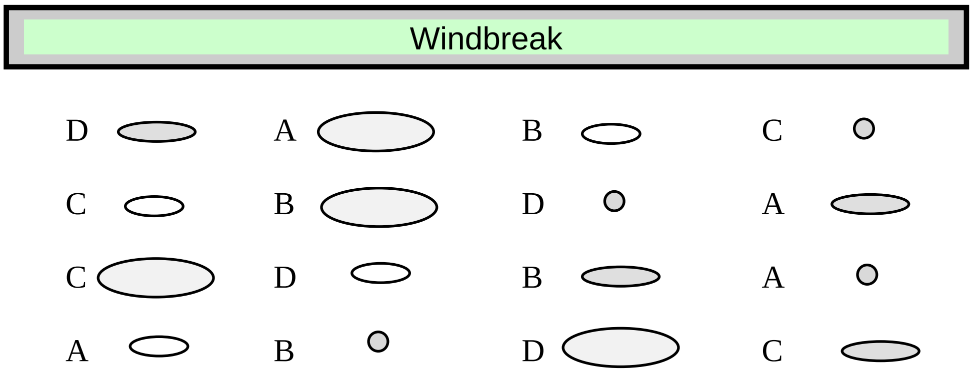


Figure 3. Sixteen orchard trees classified by size and distance from a windbreak to which 4 treatments (A, B, C and D) are assigned. The size of a circle symbolizes a size of tree.

Rows and columns do not necessarily refer to the spatial distribution of the experimental units. They can also refer to the 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 not exclusive of agricultural research. In sensory testing programs, the materials or treatment to be tested can be blocked by evaluators (judges) and time periods so that differences in judges and time periods can be removed from experimental error. Marketing studies sometimes use Latin Squares with days being rows and stores being columns.

Consider an experiment to test the effect of 4 hormones on a particular enzyme in the blood of young cows. Assume that there is a large animal to animal variation in the blood constituent as well as considerable variation in the same animal at different periods. With four animals to experiment with, the four treatments can be randomly assigned to the four animals in each of four time periods such that each animal receives all 4 treatments but in different time periods (Table 2). After each time period, the animals are rested to allow the blood enzyme under study to return to a normal level.

Table 2. 4 x 4 Latin square with hormone treatments (H1, H2, H3, and H4) assigned to each of the four animals in four different time periods.

	Time period (weeks)			
Animal	1	2	3	4
1	H2	H3	H4	H1
2	H3	H4	H1	H2
3	H4	H1	H2	H3
4	H1	H2	H3	H4

7. 3. Randomization

Proper randomization is crucial for the validity of any conclusions to be drawn from an experiment. Randomization is used both to neutralize the effects of any systematic biases as well as to provide the basis for the assumptions underlying the analysis.

7. 3. 1. Manual method

The only restriction on the Latin square arrangement is that each treatment must appear in every row and every column of the table. Since the number of possible Latin squares increases rapidly as the size of the square increases it is necessary to choose one 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. e.g. 4 treatments are involved and the following standard square has been 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, e.g.

	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 columns of the standard square (Step 1). Generate a new square according to the rankings of the rows (Step 2). Generate a new square according to the rankings of the columns (Step 3).

Step 1: 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

Step 2: 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

Step 3 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. Codes A, B, C, and D are then randomly assigned to the 4 treatments and the square shows the assignment of the treatments to the experimental units.

7. 3. 2. SAS procedure PROC PLAN

The following simple SAS program using PROC PLAN generates a template for a 5x5 Latin square design. The row and column FACTORS are identified as ORDERED and the TREATMENTS as CYCLIC. The last two factor statements generate the randomized ranks to reorganize the rows and columns by hand as indicated in point 3 above (this can also be generated by more complicated programming using PROC TABULATE).

```
proc plan;
  factors rows=5 ordered cols=5 ordered;
  treatments tmts=5 cyclic;
  factor r=5;
  factor c=5;
run;
```

7. 4. The linear model

The linear model for the Latin Square is,

$$y_{ij(t)} = \mu + \beta_i + \gamma_j + \tau_{(t)} + \varepsilon_{ij}$$

where Y_{ij} represent the observation in the i^{th} row and j^{th} column, β_i represents i^{th} row effect, γ_j represents j^{th} column effect, and $\tau_{(t)}$ represents the t^{th} treatment effect. The parenthesis around the t indicates that each t appears only once in each row and each column.

The Sum of Squares equation becomes

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

or, $SS_{total} = SS_{rows} + SS_{col.} + SS_{trtmnt.} + SS_{error}$

7. 5. ANOVA

The ANOVA table looks like this:

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

The ANOVA from the SAS output for the example provided by ST&D p230 is included below. This 4 x 4 Latin Square for yields of 4 wheat varieties shows highly significant differences for varieties and marginal differences for the 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		
Corrected Total	15	90.400			

There are 4 sources of variation in this example; 3 due to the design: columns, rows and varieties, each with 3 degrees of freedoms. The remaining source of variation is the experimental error which has degrees of freedom (4-1)(4-2) = 6. For testing the hypotheses that there are no column, or row, or treatment differences, the corresponding mean squares are divided by the MSE.

The following SAS program was used to produce the previous table.. Note that the only change to the statements in PROC ANOVA compared to a RCBD is the inclusion of one extra classification variable

```
data STp230ls;
  do row = 1 to 4;
    do col = 1 to 4;
      input trtmt $ yield @;
      output;
    end;
  end;
cards;
C  10.5  D   7.7  B  12.0  A  13.2
B  11.1  A  12.0  C  10.3  D   7.5
D   5.8  C  12.2  A  11.2  B  13.7
A  11.6  B  12.3  D   5.9  C  10.2

proc GLM;
  class row col trtmt;
  model yield=row col trtmt;
  means trtmt / lsd;
  output out=resplot p=predyiel r=resiyiel;
* Visual checking of the residuals;
proc plot data=resplot;
  plot resiyiel*predyiel=trtmt;

* Normality of the residuals;
proc univariate data=resplot normal;
  var resiyiel;

* Absence of multiplicative effects;
* This should be tested for the 3 possible interactions by three independent PROC GLM for col*row, col*trt, and row*trt;
proc GLM data= STp230ls;
  class row trtmt;
  model yield = row trtmt;
  output out=tukrt p= prt   r= rrt;
proc glm data=tukrt;
       class row trtmt;
       model yield = row trtmt prt*prt;

* Homogeneity of variances;
* This should be tested for the 3 one-way ANOVAS;
proc GLM data= STp230ls;
  class trtmt;
  model yield = trtmt;
  means trtmt / hovtest = levene;
run;quit;
```

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 appropriate design to remove these variations from experimental error. The disadvantages of Latin squares are:

1. 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 interaction between any two among all of the criteria no valid tests of significance are possible**. To check this assumption there is a version of Tukey's test for nonadditivity for Latin squares (Snedecor & Cochran1980, Chapter 15)
3. Squares smaller than 4 x 4 generally have too few replications for a desirable level of precision that is shown in the few degrees of freedom available for experimental error in the analysis of variance.

The analysis of variance for the completely randomized designs, randomized complete block designs, and Latin squares are shown schematically as follows. Assume 5 treatments to be 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 are appreciable amounts of variation in the experimental error of the CRD that can be removed by the various ways of blocking to compensate for the less of the degrees of the freedom in the experimental error, the resulting design will be less efficient.

Numerical example of the effect of the missing df in the critical ***F***:

LS3x3= df**LS**$_e$=2 df**RCBD**$_e$=4	LS5x5= df**LS**$_e$=12 df**RCBD**$_e$=16
LS ***F*** df 2, **2**, 0.05= **19.0**	LS ***F*** df 2, **12**, 0.05= **3.89**
RCBD ***F*** df 2, **4**, 0.05= **6.94**	RCBD ***F*** df 2, **16**, 0.05= **4.49**

The loss of df has a huge effect on the sensitivity of small LS!

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.

If one of the blocking variables is not included, the MSE needs to be recalculated. In the previous example (MSE_{LS} = 0.45, MSR_{LS} = 0.65, and MSC_{LS}= 2.27) if **columns** are not included the MS for columns is included in the MS for error

$$MSE_{RCBD} \cong \frac{f_{col} * MSC_{LS} + (f_{trt} + f_{error})\ MSE_{LS}}{f_{row} + f_{trt} + f_{error}} = \frac{3 * 2.27 + (3+6)\ 0.45}{3+3+6} = 0.91$$

and the relative efficiency is

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

This indicates that the column grouping increased the precision of the model by an estimated 89%. This parallels the Significant F for columns (see previous ANOVA table)

If **rows** are not included in the model the MS for rows is included in the MS for error the estimated MSE is:

$$MSE_{RCBD} \cong \frac{f_{row} * MSR_{LS} + (f_{trt} + f_{error})\ MSE_{LS}}{f_{row} + f_{trt} + f_{error}} = \frac{3 * 0.65 + (3+6)\ 0.45}{3+3+.6} = 0.50$$

and the relative efficiency is

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

This indicates that the row grouping increased the precision of the model only by an estimated 4%. This parallels the NS F for rows (see previous ANOVA table).

7. 7. Repeated Latin squares

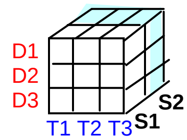
In cases where the number of treatments is small, but considerable variability is expected in two directions (two sources), two or more Latin squares may be used to increase the degrees of freedom for experimental error. Latin squares may be replicated over time or space. There are several variations of Latin squares replications: Latin squares that share common rows and columns (case-1); share common columns but not rows (case-2); share rows but not columns (case-3); or have totally independent rows and columns (case-4).

Example:

Three gasoline additives (TREATMENTS, A B & C) were tested for gas efficiency by three drivers (ROWS) using three different tractors (COLUMNS). The variable measured was the yield of carbon monoxide in a trap. The experiment was repeated twice. These replications can be:

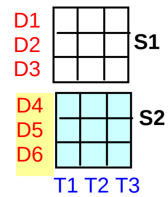
Case 1: **same** **tractors (col)** and **drivers (col)**

```
PROC GLM;
   class square col row treat;
   model yield= square col row treat;
```



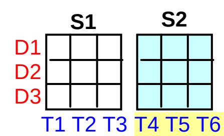
Case 2: Using **different drivers (row)** but the **same tractors (col)** .

```
PROC GLM;
   class square col row treat;
   model yield= square col row(square) treat;
```



Case 3: Using **different tractors (col)** but the **same drivers (row)**

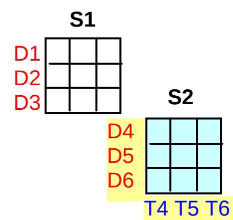
```
PROC GLM;
   class square col row treat;
   model yield= square col(square) row treat;
```



Case 4: Using **different** **drivers (row)** and **tractors (col)**.

Columns and rows are within each square

```
PROC GLM;
  class square col row treat;
  model yield= square col(square) row(square) treat;
```



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

The complete SAS program for the different cases will be discussed at the Lab class.

To convince yourself that SAS is doing the correct calculations when you indicate a nested row or column within a square, do the following exercise.

Run the previous example using row(square). Keep your output.

Then, replace the numbers of rows (1, 2, 3) in the second square by 4, 5, 6 in the data to differentiate them from the rows in square one; and change row(square) to row.

You will get identical results in both analyses. The change of row numbers in the second square indicates SAS that there are 6 different tractors. The effect of the row(square) statement is exactly the same. It tells SAS that the numbers of the rows are only meaningful within each square.

7. 8. Other complicated Latin squares (See Chapter 8 of Box, Hunter & Hunter, Appendix 8 included)

To eliminate more than 2 sources of variability a Graeco-Latin square (GLS) or a hyper- Graeco-Latin square (HGLS) is sometimes useful. A GLS is a r x r pattern that permits the study of r treatments simultaneously with three different blocking variables. If more blocking variables are included the design is called HGLS.

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

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