Experimental Design
Note 4-1
Latin Square Design

Keunbaik Lee

Sungkyunkwan University

block이 2개

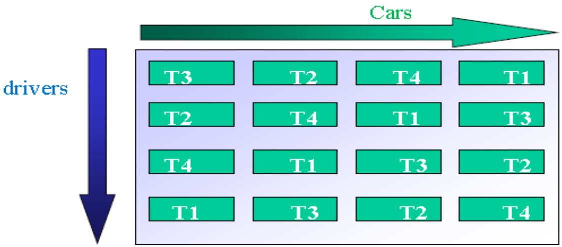
Automobile Emission Experiment

Four **cars** and **four drivers** are employed in a study of **four gasoline additives** (A, B, C, D). Even though cars can be identical models, systematic differences are likely to occur in their performance, and even though each driver may do his best to drive the car in the manner required by the test, systematic differences can occur from driver to driver. It would be desirable to eliminate both the car-to-car and driver-to-driver variations when comparing the additives.

	cars			
drivers	1	2	3	4
1	A=24	B=26	D=20	C=25
2	D=23	C=26	A=20	B=27
3	B=15	D=13	C=16	A=16
4	C=17	A=15	B=20	D=20

Latin Square Design I

- Test each additive exactly once by each car.
- Test each additive exactly once by each driver.
- For four treatments (T1, T2, T3 and T4)
- Randomization? Sudoku puzzle???



Latin Square Design II

- Design is represented in $p \times p$ grid, rows and columns are blocks and Latin letters are treatments.
  - Every row contains all the Latin letters and every column contains all the Latin letters.
- Standard Latin Square: letters in first row and first column are in alphabetic order.

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

Latin Square Design III

- Properties:
  - Block on two nuisance factors
  - Two restrictions on randomization
- Model and Assumptions

$$y_{ijk} = \underset{\text{grand mean}}{\mu} + \overset{i^{th}\text{ block 1 effect}}{\alpha_i} + \underset{j^{th}\text{ treatment effect}}{\tau_j} + \overset{k^{th}\text{ block 2 effect}}{\beta_k} + \epsilon_{ijk}, \quad i, j, k = 1, 2, \cdots, p$$

where

$$\mu = \text{grand mean}$$

$$\alpha_i = i\text{th block 1 effect } (i\text{th row effect}) : \quad \sum_{i=1}^{p} \alpha_i = 0$$

Latin Square Design IV

$$\tau_j = j\text{th treatment effect}: \sum_{j=1}^{p} \tau_j = 0$$

$$\beta_k = k\text{th block 2 effect } (k\text{th column effect}): \sum_{k=1}^{p} \beta_k = 0$$

$$\epsilon_{ijk} \sim N(0, \sigma^2): \text{(Normality, Independence, Constant Variance).}$$

Completely additive model (no interaction)

- The advantages of Latin square designs are:
  - They handle the case when we have several nuisance factors and we either cannot combine them into a single factor or we wish to keep them separate.
  - They allow experiments with a relatively small number of runs.

Latin Square Design V

- 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.

Estimation and SS partition

- Rewrite observation y_{ijk} as:

$$\begin{aligned} & y_{ijk} \\ = \;& \bar{y}_{\cdots} + (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}) + (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}) + (\bar{y}_{\cdot\cdot k} - \bar{y}_{\cdots}) + (y_{ijk} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot k} + 2\bar{y}_{\cdots}) \\ = \;& \hat{\mu} + \hat{\alpha}_i + \hat{\tau}_j + \hat{\beta}_k + \hat{\epsilon}_{ijk} \end{aligned}$$

- Partition $SS_T = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{\cdots})^2$ into

$$\begin{array}{llll} p\sum_i(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots})^2 + p\sum_j(\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots})^2 + p\sum_k(\bar{y}_{\cdot\cdot k} - \bar{y}_{\cdots})^2 + \sum_i\sum_j\sum_k \hat{\epsilon}_{ijk}^2 & & & \\ SS_{Row} & + SS_{Treatment} & + SS_{Col} & + SSE \\ (p-1) & (p-1) & (p-1) & (p-1)(p-2) \end{array}$$

Dividing SS by df gives MS: $MS_{Treatment}$, MS_{Row}, MS_{Col}, and MSE.

Basic Hypotheses Testing

- Basic hypotheses: $H_0 : \tau_1 = \tau_2 = \cdots = \tau_0 = 0$ vs H_1 : at least one is not
- Test Statistic: $F_0 = MS_{Treatment}/MSE \sim F_{p-1;(p-1)(p-2)}$ under H_0.
- Caution testing row effects ($\alpha_i = 0$) and column effects ($\beta_k = 0$).

ANOVA table for Latin Square design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F_0
Row	SS_{Row}	$p-1$	MS_{Row}	
Treatments	$SS_{Treatment}$	$p-1$	$MS_{Treatment}$	$MS_{Treatment}/MSE$
Column	SS_{Column}	$p-1$	MS_{Column}	
Error	SSE	$(p-2)(p-1)$	MSE	
Total	SS_T	p^2-1		

$$SS_T = \sum_i \sum_j \sum_k y_{ijk}^2 - y_{...}^2/p^2; \quad SS_{Row} = \frac{1}{p}\sum_i y_{i..}^2 - y_{...}^2/p^2;$$

$$SS_{Treatment} = \frac{1}{p}\sum_j y_{.j.}^2 - y_{...}^2/p^2; \quad SS_{Column} = \frac{1}{p}\sum_k y_{..k}^2 - y_{...}^2/p^2;$$

$$SSE = SS_T - SS_{Row} - SS_{Treatment} - SS_{Col}$$

Decision Rule: If $F_0 > F_{\alpha,p-1,(p-2)(p-1)}$, then reject H_0.

Another example

Consider an experiment to investigate the effect of 4 diets on milk production. There are 4 cows. Each lactation period the cows receive a different diet. Assume there is a washout period so previous diet does not affect future results. Will block on lactation period and cow. A 4 by 4 Latin square is used.

	Cows			
Periods	1	2	3	4
1	A=38	B=39	C=45	D=41
2	B=32	C=37	D=38	A=30
3	C=35	D=36	A=37	B=32
4	D=33	A=30	B=35	C=33

See Latin.SAS.

Replicating Latin Squares

- Latin Squares result in small degree of freedom for SSE: $df = (p-1)(p-2)$.
  - If 3 treatments, $df_E = 2$
  - If 4 treatments, $df_E = 6$
  - If 5 treatments, $df_E = 12$

  Use replication to increase df_E.
- Different way for replicating Latin squres:
  - Same rows and same columns
  - New rows and same columns
  - Same rows and new columns
  - New rows and new columns

  Degree of freedom for SSE depends on which method is used; Often need to include an additional blocking factor for "replicate" effect.

Model and ANOVA Table for Method 1 I

Usually includes replicate (e.g., time) effects $(\delta_1, \cdots, \delta_n)$

$$y_{ijkl} = \mu + \alpha_i + \tau_j + \beta_k + \delta_l + \epsilon_{ijkl}$$

for $i = 1, 2, \cdots, p$; $j = 1, 2, \cdots, p$; $k = 1, 2, \cdots, p$; $l = 1, 2, \cdots, n$

Method 1: same rows and same columns in additional squares

```
Example:
  1  2  3        data         replication
1 A  B  C      7.0 8.0 9.0
2 B  C  A      4.0 5.0 6.0         1
3 C  A  B      6.0 3.0 4.0

  1  2  3         data
1 C  B  A       8.0 4.0 7.0
2 B  A  C       6.0 3.0 6.0        2
3 A  C  B       5.0 8.0 7.0

  1  2  3         data
1 B  A  C        9.0 6.0 8.0
2 A  C  B        5.0 7.0 6.0       3
3 C  B  A        9.0 3.0 7.0
```

Model and ANOVA Table for Method 1 II

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Row	SS_{Row}	$p-1$		
Column	SS_{Column}	$p-1$		
Replicate	$SS_{Replicate}$	$n-1$		
Treatment	$SS_{Treatment}$	$p-1$	$MS_{Treatment}$	$MS_{Treatment}/MSE$
Error	SSE	$(p-1)(n(p+1)-3)$	MSE	
Total	SS_T	np^2-1		

See Latin-Method1.SAS.

Model and ANOVA Table for Method 2 I

- Row effects are nested within square
- α_i can be different for different squares, so they are denoted $\alpha_{i(l)}$ for $i = 1, \cdots, p$ and $l = 1, \cdots, n$, and satisfy by $\sum_{i=1}^{p} \alpha_{i(l)} = 0$ for any fixed l.

$$y_{ijkl} = \mu + \alpha_{i(l)} + \tau_j + \beta_k + \delta_l + \epsilon_{ijkl}$$

for $i = 1, 2, \cdots, p$; $j = 1, 2, \cdots, p$; $k = 1, 2, \cdots, p$; $l = 1, 2, \cdots, n$.

Method 2: New (different) rows and same columns

```
Example:
   1   2   3          data            replication
1  A   B   C      7.0  8.0  9.0
2  B   C   A      4.0  5.0  6.0            1
3  C   A   B      6.0  3.0  4.0

   1   2   3          data
4  C   B   A      8.0  4.0  7.0
5  B   A   C      6.0  3.0  6.0            2
6  A   C   B      5.0  8.0  7.0

   1   2   3          data
7  B   A   C      9.0  6.0  8.0
8  A   C   B      5.0  7.0  6.0            3
9  C   B   A      9.0  3.0  7.0
```

Model and ANOVA Table for Method 2 II

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Row	SS_{Row}	$n(p-1)$		
Column	SS_{Column}	$p-1$		
Replicate	$SS_{Replicate}$	$n-1$		
Treatment	$SS_{Treatment}$	$p-1$	$MS_{Treatment}$	$MS_{Treatment}/MSE$
Error	SSE	$(p-1)(np-2)$	MSE	
Total	SS_T	np^2-1		

See Latin-Method2.SAS.

Latin Rectangle I

- If there do not exist "replicate effects", the distinction between the squares can be neglected, so the replicated Latin squares form a Latin Rectangle with np separate rows.
- Row effects are $\alpha_1, \alpha_2, \cdots, \alpha_{np}$ satisfying $\sum_{i=1}^{np} \alpha_i = 0$.
- Model:

$$y_{ijk} = \mu + \alpha_i + \tau_j + \beta_k + \epsilon_{ijk}$$

for $i = 1, 2, \cdots, np$; $j = 1, 2, \cdots, p$; $k = 1, 2, \cdots, p$.

Latin Rectangle II

- ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F
Row	SS_{Row}	$np-1$		
Column	SS_{Column}	$p-1$		
Treatment	$SS_{Treatment}$	$p-1$	$MS_{Treatment}$	$MS_{Treatment}/MSE$
Error	SSE	$(p-1)(np-2)$	MSE	
Total	SS_T	np^2-1		

Method 3: Same rows and new (different) columns, is similar to Method 2. Details are omitted.

Latin Rectangle

- If there do not exist "replicate effects", the distinction between the squares can be neglected, so the replicated Latin squares form a Latin Rectangle with np separate columns.

Model and ANOVA Table for Method 4 I

- Usually both row effects ($\alpha_{i(l)}$) and column effects ($\beta_{k(l)}$) are nested with squares.

$$y_{ijkl} = \mu + \alpha_{i(l)} + \tau_j + \beta_{k(l)} + \delta_l + \epsilon_{ijkl}$$

for $i = 1, 2, \cdots, p$; $j = 1, 2, \cdots, p$; $k = 1, 2, \cdots, p$; $l = 1, 2, \cdots, n$.

```
Example:
   1   2   3          data           replication
1  A   B   C       7.0 8.0 9.0
2  B   C   A       4.0 5.0 6.0            1
3  C   A   B       6.0 3.0 4.0

   4   5   6          data
4  C   B   A       8.0 4.0 7.0
5  B   A   C       6.0 3.0 6.0            2
6  A   C   B       5.0 8.0 7.0

   7   8   9          data
7  B   A   C       9.0 6.0 8.0
8  A   C   B       5.0 7.0 6.0            3
9  C   B   A       9.0 3.0 7.0
```

Model and ANOVA Table for Method 4 II

- ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Row	SS_{Row}	$n(p-1)$		
Column	SS_{Column}	$n(p-1)$		
Replicate	$SS_{Replicate}$	$n-1$		
Treatment	$SS_{Treatment}$	$p-1$	$MS_{Treatment}$	$MS_{Treatment}/MSE$
Error	SSE	$(p-1)(n(p-1)-1)$	MSE	
Total	SS_T	np^2-1		

See Latin-Method4.SAS.

Crossover (Changeover) Design I

- Consider an experiment for investigating the effects of 3 diets (A, B, C) on milk production. Suppose the experiment involves 3 lactation periods. Cows take different diets in different periods, that is, a cow will not take the same diet more than once (crossover).
- Case 1: 3 cows are used:

	cow 1	cow 2	cow 3
period 1	A	B	C
period 2	B	C	A
period 3	C	A	B

- Case 2: 6 cows are employed:

	cow 1	cow 2	cow 3	cow 4	cow 5	cow 6
period 1	A	B	C	A	B	C
period 2	B	C	A	C	A	B
period 3	C	A	B	B	C	A

Crossover (Changeover) Design II

- In general, there are p treatments, np experimental units, and p periods. n Latin squares ($p \times p$) are needed to form a rectangle ($p \times np$). So that
  - Each unit has each treatment for one period
  - In each period, each treatment is used on n units

Source of Variation	df
units	$np - 1$
periods	$p - 1$
treatment	$p - 1$
error	$np^2 - (n+2)p + 2$
total	$np^2 - 1$