

# Experimental Design

## Note 14-1

### Split-plot Design

Keunbaik Lee

Sungkyunkwan University

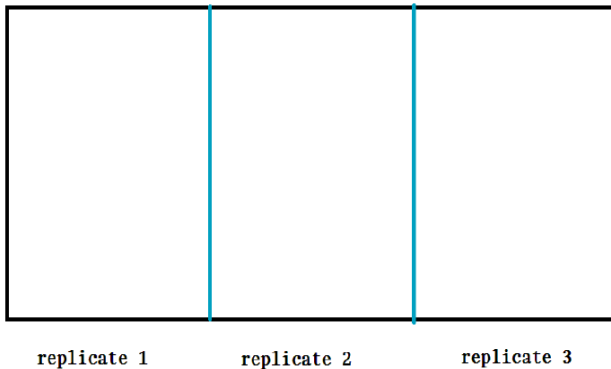
## Example 1 I

**Example:** Study six corn varieties and four fertilizers and yield is the response. Three replicates are needed.

**Method 1:** Completely randomized full factorial design, 24 level combinations of variety and fertilizer are applied to  $24 \times 3 = 72$  pieces of land (each to three).

**Method 2:** Select three fields of large area. Each field is divided into four areas (four whole-plots), four fertilizers are randomly assigned to the four whole-plots. Each area is further divided into six subareas (sub-plots), and the six varieties are randomly planted in these sub-plots.

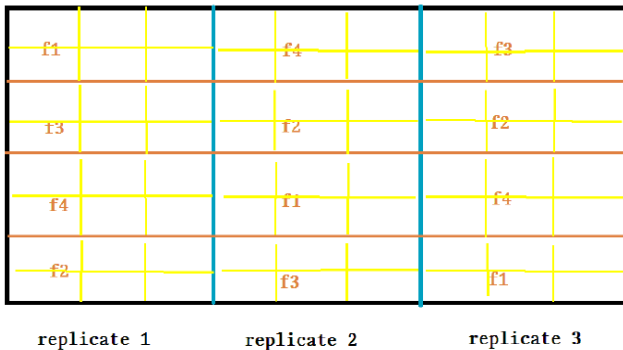
## Example 1 II



## Example 1 III

|             |             |             |
|-------------|-------------|-------------|
| f1          | f4          | f3          |
| f3          | f2          | f2          |
| f4          | f1          | f4          |
| f2          | f3          | f1          |
| replicate 1 | replicate 2 | replicate 3 |

## Example 1 IV



## Example 1 V

This leads to a **split-plot design**:

- whole-plot (treatment) factor: fertilizer
- sub-plot (treatment) factor: corn variety

## Example 2 I

- Example 2: A paper manufacturer is investigating three different pulp preparation methods and four different cooking temperatures for the pulp and study their effect on the tensile strength of the paper. Three replicates are needed.
- Because the pilot plant is only capable of making 12 runs per day, so the experimenter decides to run one replicate on each of the three days and to consider the days as blocks.

## Example 2 II

On any day, a batch of pulp is produced by one of the three methods (a whole-plot). Then the batch is divided into four samples (four sub-plots), and each sample is cooked at one of the four temperatures. Then a second batch of pulp is made up using another of the three methods. This second batch is also divided into four samples that are tested at the four temperatures. The process is then repeated for the third method.

- **The split-plot is a multifactor experiment where it is not possible to completely randomize the order of the runs:**
  - In replicate 1, select a pulp preparation method, prepare a batch
  - Divide the batch into four sections or samples, and assign one of the temperature levels to each
  - Repeat for each pulp preparation method



## Example 2 III

- Conduct replicates 2 and 3 similarly
- Each replicate (sometimes called **blocks**) has been divided into three parts, called the **whole plots**
- Pulp preparation methods is the **whole plot treatment**
- Each whole plot has been divided into four **subplots** or **split-plots**
- Temperature is the **subplot treatment**
- Generally, the hard-to-change factor is assigned to the whole plots
- This design requires only 9 batches of pulp (assuming three replicates)

The data is given below.

## Example 2 IV

|        | Day 1 |    |    | Day 2 |    |    | Day 3 |    |    |
|--------|-------|----|----|-------|----|----|-------|----|----|
| Method | 1     | 2  | 3  | 1     | 2  | 3  | 1     | 2  | 3  |
| Temp   |       |    |    |       |    |    |       |    |    |
| 200    | 30    | 34 | 29 | 28    | 31 | 31 | 31    | 35 | 32 |
| 225    | 35    | 41 | 26 | 32    | 36 | 30 | 27    | 40 | 34 |
| 250    | 27    | 38 | 33 | 40    | 42 | 32 | 41    | 39 | 39 |
| 275    | 36    | 42 | 36 | 41    | 40 | 40 | 40    | 44 | 45 |

# Split-Plot Structure

- factors are crossed (different from nested)
- randomization restriction (different from completely randomized)
- information on factor effects from two levels (or strata)
- split-plot can be considered as two superimposed blocked designs:
  - $A$ : whole-plot factor ( $a$ );  $B$ : sub-plot factor ( $b$ );  $r$  replicates
  - $RCBD_A$ : number of treatment:  $a$ , number of blocks:  $r$ .
  - $RCBD_B$ : number of treatment:  $b$ , number of blocks:  $ra$ . for whole-plots, subdivision to smaller sub-plots are ignored. For sub-plots, whole-plots considered blocks.
- More power for main subplot effect and interaction
- Should use design only for practical reasons
- Randomized factorial design more powerful if feasible

# A Typical Data Layout for Split-Plot Design

|             | Block 1   |           |   | Block 2 |   |   | Block 3 |   |           |
|-------------|-----------|-----------|---|---------|---|---|---------|---|-----------|
| WP-Factor A | 1         | 2         | 3 | 1       | 2 | 3 | 1       | 2 | 3         |
| SP-Factor B |           |           |   |         |   |   |         |   |           |
| 1           | $y_{111}$ | $y_{121}$ | . | .       | . | . | .       | . | $y_{331}$ |
| 2           | $y_{112}$ | $y_{122}$ | . | .       | . | . | .       | . | $y_{332}$ |
| 3           | $y_{113}$ | $y_{123}$ | . | .       | . | . | .       | . | $y_{333}$ |
| 4           | $y_{114}$ | $y_{124}$ | . | .       | . | . | .       | . | $y_{334}$ |

In general:

$y_{ijk}$  where  $i$  denotes Block  $i$ ,  $j$  denotes the  $j$ th level of the whole-plot factor  $A$ , and  $k$  denotes the  $k$ th level of the sub-plot factor  $B$ .

# Statistical Model I I

$$y_{ijk} = \mu + r_i + \alpha_j + (r\alpha)_{ij} + \beta_k + (r\beta)_{ik} + (\alpha\beta)_{jk} + (r\alpha\beta)_{ijk} + \epsilon_{ijk}$$

for  $i = 1, \dots, r$ ;  $j = 1, \dots, a$ ;  $k = 1, \dots, b$

- $r_i$ : block effects (random)  $\sim N(0, \sigma_r^2)$
- $\alpha_j$ : whole-plot factor ( $A$ ) main effects (fixed)
- $(r\alpha)_{ij}$ : whole-plot error (random)  $\sim$  normal with  $\sigma_{r\alpha}^2$
- $\beta_k$ : sub-plot factor ( $B$ ) main effects (fixed)
- $(r\beta)_{ik}$ : block- $B$  interaction (random)  $\sim$  normal with  $\sigma_{r\beta}^2$
- $(\alpha\beta)_{jk}$ : interaction between  $A$  and  $B$  (fixed)
- $(r\alpha\beta)_{ijk}$ : sub-plot error (random)  $\sim$  normal with  $\sigma_{r\alpha\beta}^2$
- $\epsilon_{ijk}$ : random error  $\sim N(0, \sigma^2)$

# Statistical Model I II

## Sum of Squares

- $SS_r = ab \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2, df = r - 1$
- $SS_A = rb \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2, df = a - 1$
- $SS_{rA} = b \sum_i \sum_j (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2, df = (r - 1)(a - 1)$
- $SS_B = ar \sum_k (\bar{y}_{..k} - \bar{y}_{...})^2, df = b - 1$
- $SS_{rB} = a \sum_i \sum_k (\bar{y}_{i.k} - \bar{y}_{i..} - \bar{y}_{..k} + \bar{y}_{...})^2, df = (r - 1)(b - 1)$
- $SS_{AB} = r \sum_j \sum_k (\bar{y}_{.jk} - \bar{y}_{.j.} - \bar{y}_{..k} + \bar{y}_{...})^2, df = (a - 1)(b - 1)$
- $SS_{rAB} = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.} - \bar{y}_{i.k} - \bar{y}_{.jk} + \bar{y}_{i..} + \bar{y}_{.j.} + \bar{y}_{..k} - \bar{y}_{...})^2, df = (r - 1)(a - 1)(b - 1).$
- $SSE =$

# Statistical Model I III

Expected mean squares (restricted)

|            |                        |     |     |     |         | There are two error structures; the whole-plot error and the subplot error                  |
|------------|------------------------|-----|-----|-----|---------|---|
|            |                        |     |     |     |         |   |
|            | $r$                    | $a$ | $b$ | 1   |         |   |
|            | $R$                    | $F$ | $F$ | $R$ |         |   |
| term       | $i$                    | $j$ | $k$ | $h$ | $E(MS)$ |   |
| whole plot | $r_i$                  | 1   | $a$ | $b$ | 1       | $\sigma^2 + ab\sigma_r^2$   |
|            | $\alpha_j$             | $r$ | 0   | $b$ | 1       | $\sigma^2 + b\sigma_{r\alpha}^2 + \frac{rb\Sigma\alpha_j^2}{a-1}$                           |
|            | $(r\alpha)_{ij}$       | 1   | 0   | $b$ | 1       | $\sigma^2 + b\sigma_{r\alpha}^2$  |
| subplot    | $\beta_k$              | $r$ | $a$ | 0   | 1       | $\sigma^2 + a\sigma_{r\beta}^2 + \frac{ra\Sigma\beta_k^2}{b-1}$                             |
|            | $(r\beta)_{ik}$        | 1   | $a$ | 0   | 1       | $\sigma^2 + a\sigma_{r\beta}^2$   |
|            | $(\alpha\beta)_{jk}$   | $r$ | 0   | 0   | 1       | $\sigma^2 + \sigma_{r\alpha\beta}^2 + \frac{r\Sigma\Sigma(\alpha\beta)_{jk}^2}{(a-1)(b-1)}$ |
|            | $(r\alpha\beta)_{ijk}$ | 1   | 0   | 0   | 1       | $\sigma^2 + \sigma_{r\alpha\beta}^2$  |
|            | $\epsilon_{ijk}$       | 1   | 1   | 1   | 1       | $\sigma^2$ (not estimable)  |

16

16

# Statistical Model I IV

## Estimates and tests of fixed effects

- $\hat{\alpha}_j = \bar{y}_{\cdot j} - \bar{y}_{\dots}$  for  $j = 1, \dots, a$
- $\hat{\beta}_k = \bar{y}_{..k} - \bar{y}_{\dots}$  for  $k = 1, \dots, b$
- $(\hat{\alpha}\hat{\beta})_{jk} = \bar{y}_{\cdot jk} - \bar{y}_{\cdot j} - \bar{y}_{..k} + \hat{y}_{\dots}$
- Test  $\alpha_j = 0$ ,  $F_0 = MS_A / MS_{rA}$
- Test  $\beta_k = 0$ ,  $F_0 = MS_B / MS_{rB}$
- Test  $(\alpha\beta)_{jk} = 0$ ,  $F_0 = MS_{AB} / MS_{rAB}$



# Statistical Model I V

**Example** See split\_plot.SAS.

■ **TABLE 14.16**

Analysis of Variance for the Split-Plot Design Using the Tensile Strength Data from Table 14.14

| Source of Variation                                   | Sum of Squares | Degrees of Freedom | Mean Square | $F_0$ | $P$ -Value |
|---|----------------|--------------------|-------------|-------|------------|
| Replicates (or blocks)                                | 77.55          | 2                  | 38.78       |       |            |
| Preparation method ( $A$ )                            | 128.39         | 2                  | 64.20       | 7.08  | 0.05       |
| Whole plot error (replicates (or blocks) $\times A$ ) | 36.28          | 4                  | 9.07        |       |            |
| Temperature ( $B$ )                                   | 434.08         | 3                  | 144.69      | 41.94 | <0.01      |
| Replicates (or blocks) $\times B$                     | 20.67          | 6                  | 3.45        |       |            |
| $AB$  | 75.17          | 6                  | 12.53       | 2.96  | 0.05       |
| Subplot error (replicates (or blocks) $\times AB$ )   | 50.83          | 12                 | 4.24        |       |            |
| Total   | 822.97         | 35                 |             |       |            |

## Statistical Model II I

$$y_{ijk} = \mu + r_i + \alpha_j + (r\alpha)_{ij} + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk}$$

for  $i = 1, \dots, r; j = 1, \dots, a; k = 1, \dots, b$

- $r_i$ : block effects (random)  $\sim N(0, \sigma_r^2)$
- $\alpha_j$ : whole-plot factor ( $A$ ) main effects (fixed)
- $(r\alpha)_{ij}$ : whole plot error  $\sim$  normal with  $\sigma_{r\alpha}^2$
- $\beta_k$ : sub-plot factor ( $B$ ) main effects (fixed)
- $(\alpha\beta)_{jk}$ :  $A$  and  $B$  interaction (fixed)
- $\epsilon_{ijk}$ : sub-plot error  $N(0, \sigma_\epsilon^2)$

## Statistical Model II II

### Expected mean square

| Term                      | $E(MS)$   |
|---------------------------|---|
| $r_i$                     | $\sigma_\epsilon^2 + ab\sigma_r^2$  |
| $\alpha_j(A)$             | $\sigma_\epsilon^2 + b\sigma_{r\alpha}^2 + \frac{rb \sum_j \alpha_j^2}{a-1}$  |
| $(r\alpha)_{ij}$          | $\sigma_\epsilon^2 + b\sigma_{r\alpha}^2$ (whole plot error)                  |
| $\beta_k (B)$             | $\sigma_\epsilon^2 + \frac{ra \sum_j \alpha_j^2}{b-1}$                        |
| $(\alpha\beta)_{jk} (AB)$ | $\sigma_\epsilon^2 + \frac{r \sum_j \sum_k (\alpha\beta)_{jk}^2}{(a-1)(b-1)}$ |
| $\epsilon_{ijk}$          | $\sigma_\epsilon^2$ (subplot error)   |

# General Split-Plot Designs I

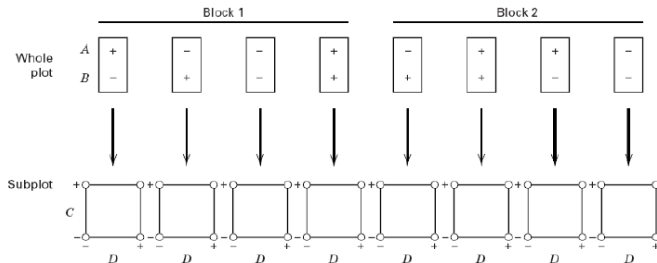
- Can have  $>$  one whole-plot factor and  $>$  one subplot factor with various blocking schemes
- split-plot design consists of two superimposed block design
  - Whole plot
    - CRD, RCBD, Factorial D, BIBD, etc.
  - Subplot
    - RCBD, BIBD, Factorial Design, etc.
- Analysis of Covariance

# Variations of the Basic Split-Plot Design

More than two factors - see page 627

*A* & *B* (gas flow & temperature) are hard to change;

*C* & *D* (time & wafer position) are easy to change.



■ **FIGURE 14.7** A split-plot design with four design factors, two in the whole plot and two in the subplot

# Other Variations I

- Split-split-plot design

- 1 randomization restriction can occur at any number of levels within the experiment
- 2 two-level: split-split-plot design

Example: Absorption times of antibiotic capsule

- 3 technicians
- 4 dosage strengths
- 4 capsule wall thickness
- 4 replicates/days
- A split-split-plot design
- Two randomization restrictions present within each replicate

## Other Variations II

- Strip-split-plot design (or Criss cross design, or Split-block design)

Example: we want to compare the yield of a certain crop under different systems of soil preparation ( $A : a_1, a_2, a_3, a_4$ ) and different density of seeding ( $B : b_1, b_2, b_3, b_4, b_5$ ). Both operations (tilling and seeding) are done mechanically and it is impossible to perform both on small pieces of land. The arrangement shown below (strip-split-plot design) is then replicated  $r$  times, each time using different randomizations for  $A$  and  $B$ .

## Other Variations III

|      | $b1$     | $b4$     | $b2$     | $b3$     | $b5$     | Strip plots ↓ |
|------|----------|----------|----------|----------|----------|---------------|
| $a4$ | $a_4b_1$ | $a_4b_4$ | $a_4b_2$ | $a_4b_3$ | $a_4b_5$ |               |
| $a1$ | $a_1b_1$ | $a_1b_4$ | $a_1b_2$ | $a_1b_3$ | $a_1b_5$ | Whole plots ← |
| $a2$ | $a_2b_1$ | $a_2b_4$ | $a_2b_2$ | $a_2b_3$ | $a_2b_5$ |               |
| $a3$ | $a_3b_1$ | $a_3b_4$ | $a_3b_2$ | $a_3b_3$ | $a_3b_5$ |               |