# Experimental Design
## Note 6
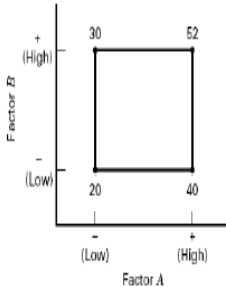## Introduction to factorial design

Keunbaik Lee

Sungkyunkwan University

# Introduction to Factorial Design of Engineering Experiments

- Many experiments involve the study of the effects of two or more factors
  - Factorial design are most efficient
  - All possible combinations of the levels of the factors are investigated
    - crossed �
  - Effect of a factor: change in response produced by a change in the level of the factor
    - ie. main effect (primary factor)

■ FIGURE 5.1 A two-factor factorial experiment, with the response (y) shown at the corners
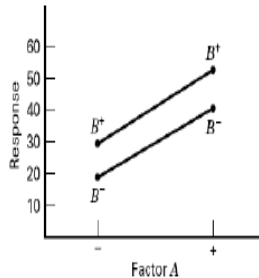


■ FIGURE 5.3 A factorial experiment without interaction
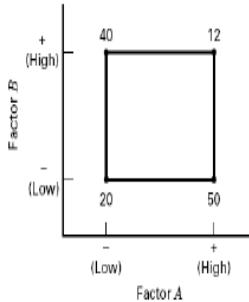
## Some Basic Definitions II

Definition of a factor effect: <mark>The change in the mean response when the factor is changed from low to high</mark>

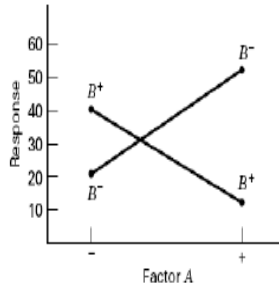$$A = \bar{y}_{A^+} - \bar{y}_{A^-} = \frac{40 + 52}{2} - \frac{20 + 30}{2} = 21$$

$$B = \bar{y}_{B^+} - \bar{y}_{B^-} = \frac{30 + 52}{2} - \frac{20 + 40}{2} = 11$$

$$AB = \frac{\overset{++}{52 + 20}}{2} - \frac{\overset{+-}{30 + 40}}{2} = -1$$

**FIGURE 5.2** A two-factor factorial experiment with interaction

**FIGURE 5.4** A factorial experiment with interaction

## The Case of Interaction II

$$A = \bar{y}_{A^+} - \bar{y}_{A^-} = \frac{50 + 12}{2} - \frac{20 + 40}{2} = 1$$

$$B = \bar{y}_{B^+} - \bar{y}_{B^-} = \frac{40 + 12}{2} - \frac{20 + 50}{2} = -9$$

$$AB = \frac{\overset{++}{12} + \overset{--}{20}}{2} - \frac{\overset{+-}{40} + \overset{-+}{50}}{2} = -29$$

Assume design factors are quantitative (e.g. temperature, time)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon$$

ex) let $x_2 = \begin{cases} 1 \\ 0 \end{cases}$ .

For $x_2 = 1$, $y = \beta_0 + \beta_1 x_1 + \beta_2 + \beta_{12} x_1 + \epsilon$
$$= (\beta_0 + \beta_2) + (\beta_1 + \beta_{12}) x_1 + \epsilon$$

For $x_2 = 0$, $y = \beta_0 + \beta_1 x_1 + \epsilon$

The least squares fit is

$$\hat{y} = 35.5 + 10.5 x_1 + 5.5 x_2 + 0.5 x_1 x_2$$

We ignore the interaction. Then $\hat{y} = 35.5 + 10.5 x_1 + 5.5 x_2$.

(a) The response surface   (b) The contour plot
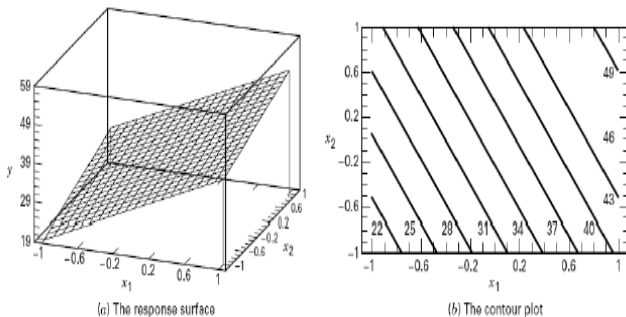
■ FIGURE 5.5   Response surface and contour plot for the model $\hat{y} = 35.5 + 10.5x_1 + 5.5x_2$

X가 증가함에 따라 Y도 증가한다

## The Effect of Interaction on the Response Surface I

Suppose that we add an interaction term to the model:

$$\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + 8x_1x_2$$

Interaction is actually a form of curvature

(a) The response surface
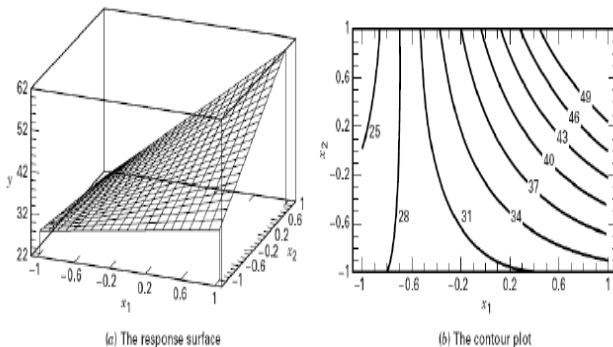
(b) The contour plot

■ FIGURE 5.6  Response surface and contour plot for the model $\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + 8x_1x_2$

(Interaction Term이 추가되었을 때) : $X_1$ 이 작을 때 $X_2$ 가 천천히 증가 하다가 $X_1$ 이 커질수록 가파르게 증가

**One-factor-a-time design** as the opposite of factorial design.

Advantages of factorial over one-factor-a-time

- more efficient (runsize and estimation precision)
- able to accommodate interactions
- results are valid over a wider range of experimental conditions

## Example I: Battery life experiment I

An engineer is studying the effective life of a certain type of battery. Two factors, plate material and temperature, are involved. There are three types of plate materials (1, 2, 3) and three temperature levels (15, 70, 125). Four batteries are tested at each combination of plate material and temperature, and all 36 tests are run in random order. The experiment and the resulting observed battery life data are given as following:

$3 \times 3 = 9$ treatments

# Example I: Battery life experiment II

■ **TABLE 5.1**
Life (in hours) Data for the Battery Design Example

| Material Type | $\bar{y}_{i1\cdot}$ | Temperature (°F) | | | | | |
|---|---|---|---|---|---|---|---|
| | | 15 | | 70 | | 125 | |
| 1 | | 130 | 155 | 34 | 40 | 20 | 70 |
| | | 74 | 180 | 80 | 75 | 82 | 58 |
| 2 | | 150 | 188 | 136 | 122 | 25 | 70 |
| | | 159 | 126 | 106 | 115 | 58 | 45 |
| 3 | | 138 | 110 | 174 | 120 | 96 | 104 |
| | | 168 | 160 | 150 | 139 | 82 | 60 |

$$\bar{y}_{\cdot1\cdot} \qquad \bar{y}_{\cdot2\cdot} \qquad \bar{y}_{\cdot3\cdot}$$

A = Material type; B = Temperature (a quantitative variable)

- What effects do material type & temperature have on life?
- Is there a choice of material that would give long life regardless of temperature (a robust product)?

## Example II: Bottling Experiment

A soft drink bottler is interested in obtaining more uniform fill heights in the bottles produced by his manufacturing process. An experiment is conducted to study three factors of the process, which are

| | |
|---|---|
| the percent carbonation (A): | 10, 12, 14 percent |
| the operating pressure (B): | 25, 30 psi |
| the line speed (C): | 200, 250 bpm |

The response is the deviation from the target fill height. Each combination of the three factors has two replicates and all 24 runs are performed in a random order. The experiment and data are shown below.

| | Pressure (B) | | | |
|---|---|---|---|---|
| | 25 psi | | 30 psi | |
| | Line Speed (C) | | Line Speed (C) | |
| Carbonation (A) | 200 | 250 | 200 | 250 |
| 10 | -3, -1 | -1, 0 | -1, 0 | 1, 1 |
| 12 | 0, 1 | 2, 1 | 2, 3 | 6, 5 |
| 14 | 5, 4 | 7, 6 | 7, 9 | 10,11 |

# Factorial Design

- a number of factors: $F_1, F_2, \cdots, F_r$.
- each with a number of levels: $l_1, l_2, \cdots, l_r$
- number of all possible level combinations (treatments): $l_1 \times l_2 \times \cdots \times l_r$
- interested in (main) effects, 2-factor interactions (2 fi), 3-factor interactions (3 fi), etc

# The General Two-Factor Factorial Experiment

**■ TABLE 5.2**
General Arrangement for a Two-Factor Factorial Design

| | | Factor $B$ | | |
|---|---|---|---|---|
| | | 1 | 2 | . . . | $b$ |
| Factor $A$ | 1 | $y_{111}, y_{112},$ $\ldots, y_{11n}$ | $y_{121}, y_{122},$ $\ldots, y_{12n}$ | | $y_{1b1}, y_{1b2},$ $\ldots, y_{1bn}$ |
| | 2 | $y_{211}, y_{212},$ $\ldots, y_{21n}$ | $y_{221}, y_{222},$ $\ldots, y_{22n}$ | | $y_{2b1}, y_{2b2},$ $\ldots, y_{2bn}$ |
| | $\vdots$ | | | | |
| | $a$ | $y_{a11}, y_{a12},$ $\ldots, y_{a1n}$ | $y_{a21}, y_{a22},$ $\ldots, y_{a2n}$ | | $y_{ab1}, y_{ab2},$ $\ldots, y_{abn}$ |

$a$ levels of factor $A$; $b$ levels of factor $B$; $n$ replicates
This is a **completely randomized design**.

## Statistical Model (Two Factors: A and B) I

- Statistical model is

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$$

for $i = 1, 2, \cdots, a$; $j = 1, 2, \cdots, b$; $k = 1, 2, \cdots, n$, where

$\mu =$ grand mean

$\tau_i = i$th level effect of factor $A$ (controls $B$) (main effects of $A$)

$\beta_j = j$th level effect of factor $B$ (controls $A$) (main effects of $B$)

$(\tau\beta)_{ij} =$ interaction effect of combination $ij$

(Explain variation not explained by main effects)

$\epsilon_{ijk} \sim N(0, \sigma^2)$

## Statistical Model (Two Factors: A and B) II

- Over-parameterized model: must include certain parameter constraints. Typically

$$\sum_i \tau_i = 0, \quad \sum_j \beta_j = 0, \quad \sum_i (\tau\beta)_{ij} = 0, \quad \sum_j (\tau\beta)_{ij} = 0$$

## Estimates

- Rewrite observation as:

$$
\begin{aligned}
y_{ijk} &= \bar{y}_{\cdots} + (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}) + (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}) \\
&\quad + (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}) + \underset{residuals}{(y_{ijk} - \bar{y}_{ij\cdot})}
\end{aligned}
$$

- Results in estimates

$$
\hat{\mu} = \bar{y}_{\cdots}, \quad \hat{\tau}_i = \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}, \hat{\beta}_j = \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots},
$$
$$
(\hat{\tau\beta})_{ij} = \bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}
$$

- Predictive value at level combination $ij$ is

$$
\hat{y}_{ijk} = \bar{y}_{ij\cdot}
$$

- Residuals are

$$
\hat{\epsilon}_{ijk} = y_{ijk} - \bar{y}_{ij\cdot}
$$

## Battery Example I

**Effects Estimation (Battery Experiment)**

0. $\hat{\mu} = \bar{y}_{...} = 105.5278$

1. Treatment mean response, or cell mean, or predicted value,

$$\hat{y}_{ij} = \hat{\mu}_{ij} = \bar{y}_{ij.} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j + \hat{(\tau\beta)}_{ij}$$

|          | temperature |        |       |
|----------|-------------|--------|-------|
| material | 1           | 2      | 3     |
| 1        | 134.75      | 57.25  | 57.50 |
| 2        | 155.75      | 119.75 | 49.50 |
| 3        | 144.00      | 145.75 | 85.50 |

## Battery Example II

2. Factor level means
   row means $\bar{y}_{i..}$ for $A$; column means $\bar{y}_{.j.}$ for $B$
   material: $\bar{y}_{1..} = 83.166$, $\bar{y}_{2..} = 108.3333$, $\bar{y}_{3..} = 125.0833$;
   temperature: $\bar{y}_{.1.} = 144.8333$, $\bar{y}_{.2.} = 107.5833$,
   $\bar{y}_{.3.} = 64.1666$

3. Main effects estimates
   $$\hat{\tau}_1 \overset{\bar{y}_{1..} - \bar{y}_{...}}{=} -22.3612, \quad \hat{\tau}_2 \overset{\bar{y}_{2..} - \bar{y}_{...}}{=} 2.8055, \quad \hat{\tau}_3 = 19.555$$
   $$\hat{\beta}_1 = 39.3055, \quad \hat{\beta}_2 = 2.0555, \quad \hat{\beta}_3 = -41.3611$$

4. Interactions $((\hat{\tau\beta})_{ij})$

| | $\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$ temperature | | |
|----------|-----------|-----------|-----------|
| material | 1 | 2 | 3 |
| 1 | 12.2779 | -27.9721 | 15.6946 |
| 2 | 8.1112 | 9.3612 | -17.4722 |
| 3 | -20.3888 | 18.6112 | 1.7779 |

## Battery Example III

**Partitioning the Sum of Squares**

- Based on

$$
\begin{aligned}
y_{ijk} &= \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) \\
&\quad + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) + (y_{ijk} - \bar{y}_{ij.})
\end{aligned}
$$

- Calculate

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

## Battery Example IV

- Right hand side simplifies to

$$
\begin{aligned}
SS_T &= bn \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 + an \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2 \\
&+ n \sum_i \sum_j (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.})^2 \\
&= SS_A + SS_B + SS_{AB} + SSE \\
df &= (a-1) + (b-1) + (a-1)(b-1) + ab(n-1)
\end{aligned}
$$

- Using $SS/df$ leads to $MS_A$, $MS_B$, $MS_{AB}$, and $MSE$.

## Testing Hypotheses I

1. Main effects of $A$:
   $H_0 : \tau_1 = \cdots = \tau_a = 0$ vs $H_1$ : at least one $\tau \neq 0$.

2. Main effects of $B$:
   $H_0 : \beta_1 = \cdots = \beta_b = 0$ vs $H_1$ : at least one $\beta_j \neq 0$.

3. Interaction effects of $AB$:
   $H_0 : (\tau\beta)_{ij} = 0$ for all $i, j$ vs
   $H_1$ : at least one $(\tau\beta)_{ij} \neq 0$.

Use $F$-statistic for testing the hypotheses above:

1: $F_0 = \frac{SS_A/(a-1)}{SSE/(ab(n-1))}$  2: $F_0 = \frac{SS_B/(b-1)}{SSE/(ab(n-1))}$  3: $F_0 = \frac{SS_{AB}/(a-1)(b-1)}{SSE/(ab(n-1))}$.

Note that

$$
\begin{aligned}
E(MSE) &= \sigma^2 \\
E(MS_A) &= \sigma^2 + bn \sum_i \tau_i^2/(a-1) \\
E(MS_B) &= \sigma^2 + an \sum_j \beta_j^2/(b-1) \\
E(MS_{AB}) &= \sigma^2 + n \sum_i \sum_j (\tau\beta)_{ij}^2/(a-1)(b-1)
\end{aligned}
$$

■ **TABLE 5.3**

**The Analysis of Variance Table for the Two-Factor Factorial, Fixed Effects Model**

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | $F_0$ |
|---|---|---|---|---|
| A treatments | $SS_A$ | $a-1$ | $MS_A = \dfrac{SS_A}{a-1}$ | $F_0 = \dfrac{MS_A}{MS_E}$ |
| B treatments | $SS_B$ | $b-1$ | $MS_B = \dfrac{SS_B}{b-1}$ | $F_0 = \dfrac{MS_B}{MS_E}$ |
| Interaction | $SS_{AB}$ | $(a-1)(b-1)$ | $MS_{AB} = \dfrac{SS_{AB}}{(a-1)(b-1)}$ | $F_0 = \dfrac{MS_{AB}}{MS_E}$ |
| Error | $SS_E$ | $ab(n-1)$ | $MS_E = \dfrac{SS_E}{ab(n-1)}$ | |
| Total | $SS_T$ | $abn-1$ | | |

만약 $n$, 즉 반복수가 1일 경우
$SS_E$의 df가 0이 되어버리기
때문에 Interaction을 Error로
간주하고 진행한다.

$$SS_T = \sum\sum y_{ijk}^2 - y_{...}^2/abn; \quad SS_A = \frac{1}{bn}\sum y_{i..}^2 - y_{...}^2/abn$$

$$SS_B = \frac{1}{an}\sum y_{.j.}^2 - y_{...}^2/abn; \quad SS_{\text{subtotal}} = \frac{1}{n}\sum\sum y_{ij.}^2 - y_{...}^2/abn$$

$$SS_{AB} = SS_{\text{subtotal}} - SS_A - SS_B; \quad SS_E = \text{Subtraction}$$

## Analysis of Variance Table II

$df_E > 0$ only if $n > 1$. When $n = 1$, no $SSE$ is available.
So we cannot test the effects.

- If we can assume that the interactions are negligible
  $((\tau\beta)_{ij} = 0)$, $MS_{AB}$ becomes a good estimate of $\sigma$ and it can
  be used as $MSE$
- Caution: if the assumption is wrong, then error and
  interaction are confounded and testing results can go wrong.

See Battery-Life.SAS.

## Checking Assumptions I

- Errors are normally distributed
  Histogram or QQ plot of residuals

- Constant variance
  Residuals vs $\hat{y}_{ij}$ plot, Residuals vs factor A plot, and Residuals vs factor B plot

## Checking Assumptions II

- If $n = 1$, no interaction.
  Tukey's Test of Nonadditivity Assumption $(\tau\beta)_{ij} = \gamma\tau_i\beta_j$.
  $H_0 : \gamma = 0$

  $$
  SS_N = \frac{\left[\sum_i \sum_j y_{ij} y_{i\cdot} y_{\cdot j} - y_{\cdot\cdot}(SS_A + SS_B + y_{\cdot\cdot}^2/ab)\right]^2}{ab SS_A SS_B}
  $$
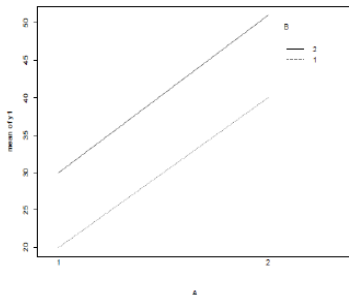
  $$
  F_0 = \frac{SS_N/1}{(SSE - SS_N)/((a-1)(b-1)-1)} \sim F_{1,(a-1)(b-1)-1}
  $$

  the convenient procedure used for RCBD can be employed.
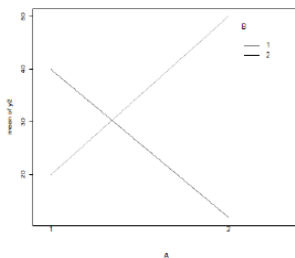
See Battery-Life.SAS.

## Interaction I

**Interaction plot for $A$ and $B$ (No Interaction)**



Difference between level means of $B$ (with $A$ fixed at a level) does not depend on the level of $A$; demonstrated by two parallel lines.

# Interaction II
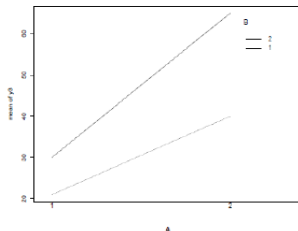
**Interaction Plot for $A$ and $B$ (Antagonistic Interaction from B to A)**



Difference between level means of $B$ (with $A$ fixed at a level) depends on the level of $A$. If the trend of mean response over $A$ reverses itself when $B$ changes from one level to another, the interaction is said to be antagonistic from $B$ to $A$. Demonstrated by two lines with slopes of opposite signs.

# Interaction III

**Interaction Plot for $A$ and $B$ (Synergistic Interaction from $B$ to $A$)**



Difference between level means of $B$ (with $A$ fixed at a level) depends on the level of $A$. If the trend of mean response over $A$ do not change when $B$ changes from one level to another, the interaction is said to be synergistic; demonstrated by two unparalleled lines with slopes of the same sign.

See Battery-Life.SAS.

## Multiple comparison when factors do not interact I

When factors do not interact, i.e., the $F$ test for interaction is not significant in the ANOVA, factor level means can be compared to draw conclusions regarding their effects on response.

- $var(\bar{y}_{i\cdot\cdot}) = \frac{\sigma^2}{nb}$, $var(\bar{y}_{\cdot j\cdot}) = \frac{\sigma^2}{na}$
- For $A$ or rows: $var(\bar{y}_{i\cdot\cdot} - \bar{y}_{i'\cdot\cdot}) = \frac{2\sigma^2}{nb}$;
  For $B$ or columns: $var(\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot j'\cdot}) = \frac{2\sigma^2}{na}$;
- Tukey's method
  For rows: $CD = \frac{q_\alpha(a, ab(n-1))}{\sqrt{2}} \sqrt{MSE \frac{2}{nb}}$
  For columns: $CD = \frac{q_\alpha(b, ab(n-1))}{\sqrt{2}} \sqrt{MSE \frac{2}{na}}$
- Bonferroni method: $CD = t_{\alpha/2m, ab(n-1)} SE$ where $SE$ depends on whether for rows or columns.

## Multiple comparison when factors do not interact II

However, when interactions are present, be careful interpreting factor level means (row or column) comparisons because it can be misleading. Usually we will directly compare treatment means (or cell means) instead.

# Multiple comparison when factors interact: treatment (cell) mean comparison I

When factors interact, multiple comparison is usually directly applied to treatment means

$$\mu_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} \text{ vs } \mu_{i'j'} = \mu + \tau_{i'} + \beta_{j'} + (\tau\beta)_{i'j'}$$

- $\hat{\mu}_{ij} = \bar{y}_{ij\cdot}$ and $\hat{\mu}_{i'j'} = \bar{y}_{i'j'\cdot}$.
- $var(\bar{y}_{ij\cdot} - \bar{y}_{i'j'\cdot}) = \frac{2\sigma^2}{n}$
- there are $ab$ treatment means and $m_0 = \frac{ab(ab-1)}{2}$ pairs.
- Tukey's method: $CD = \frac{q_\alpha(ab, ab(n-1))}{\sqrt{2}} \sqrt{MSE \frac{2}{n}}$

## Multiple comparison when factors interact: treatment (cell) mean comparison II

- Bonferroni's method: $CD = t_{\alpha/2m, ab(n-1)} \sqrt{MSE \frac{2}{n}}$

See Battery-Life.SAS.

## Quantitative and Qualitative Factors

- The basic ANOVA procedure treats every factor as if it were qualitative

- Sometimes an experiment will involve both quantitative and qualitative factors, such as in Example 5.1

- This can be accounted for in the analysis to produce regression models for the quantitative factors at each level (or combination of levels) of the qualitative factors

- These response curves and/or response surfaces are often a considerable aid in practical interpretation of the results

## Fitting Response Curves or Surfaces I

Goal: Model the functional relationship between lifetime and temperature at every material level. In Battery Experiment:

15, 70, 125

- Material is qualitative while temperature is quantitative.
- Want to fit the response using effects of material, temperature, and their interactions.
- Temperature has quadratic effect. Here we will simply $t$ and $t^2$.
- Levels of material need to be converted to indicator variables denoted by $x_1$ and $x_2$ as follows

| material | $x_1$ | $x_2$ |
|:--------:|:-----:|:-----:|
| 1        | 1     | 0     |
| 2        | 0     | 1     |
| 3        | -1    | -1    |

# Fitting Response Curves or Surfaces II

- For convenience, convert temperature to -1, 0, and 1 using $t = \frac{\text{temperature} - 70}{55}$

- The following model is used:

$$
\begin{aligned}
y_{ijk} &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 t + \beta_4 x_1 t + \beta_5 x_2 t + \beta_6 t^2 \\
&\quad + \beta_7 x_1 t^2 + \beta_8 x_2 t^2 + \epsilon_{ijk}
\end{aligned}
$$

Want to estimate the coefficients: $\beta_0, \beta_1, \cdots, \beta_8$ using regression.

$\sum_{i=1}^{3} T_i = 0$
restriction anova

In regression:

| material | 1 | 2 | 3 |
|----------|---|---|---|
| $X_1$ | 1 | 0 | 0 |
| $x_2$ | 0 | 1 | 0 |

In Experimental Design;

$$
\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} T_1, & T_2 \end{bmatrix} = \begin{bmatrix} T_1 \\ T_2 \\ -T_1 - T_2 \end{bmatrix}
$$

| material | 1 | 2 | 3 |
|----------|---|---|----|
| $X_1$ | 1 | 0 | -1 |
| $x_2$ | 0 | 1 | -1 |

## Fitting Response Curves or Surfaces III

See Battery-Life-Quantity.SAS.

Results: From the SAS output, the fitted model is

$$
\begin{aligned}
\hat{y} = {} & 107.58 - 50.33x_1 + 12.17x_2 - 40.33t + 1.71x_1t - 12.79x_2t \\
& -3.08t^2 + 41.96x_1t^2 - 14.04x_2t^2
\end{aligned}
$$

where $x_1 \begin{cases} 1, & \text{material 1} \\ 0, & \text{material 2} \\ -1, & \text{material 3} \end{cases}$ $x_2 \begin{cases} 1, & \text{material 1} \\ 0, & \text{material 2} \\ -1, & \text{material 3} \end{cases}$

Note that terms with insignificant coefficients are still kept in the fitted model here. In practice, model selection may be employed to remove unimportant terms and choose the best fitted model. But we will not pursue it in this course.

The model above are in terms of both $x_1$, $x_2$ and $t$. We can specify the level of material, that is, the values of dummy variable $x_1$ and $x_2$, to derive fitted response curves for material at different levels. Three response curves:

- Material at level 1 ($x_1 = 1$, $x_2 = 0$)

$$\hat{y}_{1t} = 57.25 - 38.62t + 38.88t^2$$

- Material at level 2 ($x_1 = 0$, $x_2 = 1$)

$$\hat{y}_{2t} = 119.75 - 53.12t - 17.12t^2$$

- Material at level 3 ($x_1 = -1$, $x_2 = -1$)

$$\hat{y}_{3t} = 145.74 - 29.25t - 31t^2$$

These curves can be used to predict lifetime of battery at any temperature between 15 and 125 degree. But one needs to be careful about extrapolation. For example, the fitted curve at Material level 1 suggests that lifetime of a battery can be infinity when temperature goes to infinity, which is clearly false.

# General Factorial Design and Model I

- Factorial Design - including all possible level combinations
- $a$ levels of Factor $A$, $b$ levels of Factor $B$, $\cdots$
- Straightforward ANOVA if all fixed effects
- In 3 factor mdoel $\rightarrow$ $nabc$ observations
- Need $n > 1$ to test for all possible interactions
- Statistical Model (3 factors)

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\beta\gamma)_{jk} + (\tau\gamma)_{ik} + (\tau\beta\gamma)_{ijk} + \epsilon_{ijkl}$$

where $i = 1, 2, \cdots, a$; $j = 1, 2, \cdots, b$; $k = 1, 2, \cdots, c$;
$l = 1, 2, \cdots, n$.
number of replicates

## General Factorial Design and Model II

- ANOVA Table

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | F |
|---|---|---|---|---|
| Factor A | $SS_A$ | $a-1$ | $MS_A$ | $MS_A/MSE$ |
| Factor B | $SS_B$ | $b-1$ | $MS_B$ | $MS_B/MSE$ |
| Factor C | $SS_C$ | $c-1$ | $MS_C$ | $MS_C/MSE$ |
| AB | $SS_{AB}$ | $(a-1)(b-1)$ | $MS_{AB}$ | $MS_{AB}/MSE$ |
| AC | $SS_{AC}$ | $(a-1)(c-1)$ | $MS_{AC}$ | $MS_{AC}/MSE$ |
| BC | $SS_{BC}$ | $(b-1)(c-1)$ | $MS_{BC}$ | $MS_{BC}/MSE$ |
| ABC | $SS_{ABC}$ | $(a-1)(b-1)(c-1)$ | $MS_{ABC}$ | $MS_{ABC}/MSE$ |
| Error | $SSE$ | $abc(n-1)$ | $MSE$ | |
| Total | $SS_T$ | $abcn-1$ | | |

- See Bottling.SAS.

## General Factorial Model-2

- Usual assumptions and diagnostics.
- Multiple comparisons: simple extensions of the two-factor case.
- Often higher order interactions are negligible.
- Beyond three-way interactions difficult to picture
- Pooled together with error (increase $df_E$).

# Blocked Factorial Experiment I

**Revisit Battery Life Experiment**

If we assume further that four operators (1,2,3,4) were hired to conduct the experiment. It is known that different operators can cause systematic difference in battery lifetime. Hence operators should be treated as blocks.

For each treatment (treatment combination), the observations were in the order of the operators 1, 2, 3, and 4.

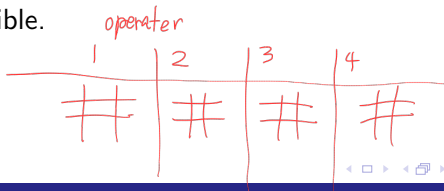This is a **blocked factorial design**.

**Statistical Model**:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \delta_k + \epsilon_{ijk}$$

for $i = 1, 2, \cdots, a$; $j = 1, 2, \cdots, b$; and $k = 1, 2, \cdots, n$; $\delta_k$ is the effect of the $k$the block.

- randomization restriction is imposed (complete block factorial design).
- interactions between blocks and treatment effects are assumed to be negligible.

## Blocked Factorial Experiment III

- The previous ANOVA table for the experiment should be modified as follows:
  Add: Block Sum of Square

$$SS_{Blocks} = \frac{1}{ab} \sum_k y_{..k}^2 - \frac{y_{...}^2}{abn}, \quad df = n - 1$$

  Modify: Error Sum of Squares:

$$(new)SSE = (old)SSE - SS_{Blocks}, \quad df = (ab - 1)(n - 1)$$

- other inferences should be modified accordingly.

**ANOVA Table**

■ TABLE 5.20
Analysis of Variance for a Two-Factor Factorial in a Randomized Complete Block

| Source of Variation | Sum of Squares | Degrees of Freedom | Expected Mean Square | $F_0$ |
|---|---|---|---|---|
| Blocks | $\frac{1}{ab}\sum_k y_{..k}^2 - \frac{y_{...}^2}{abn}$ | $n-1$ | $\sigma^2 + ab\sigma_\delta^2$ | |
| A | $\frac{1}{bn}\sum_i y_{i..}^2 - \frac{y_{...}^2}{abn}$ | $a-1$ | $\sigma^2 + \frac{bn\sum\tau_i^2}{a-1}$ | $\frac{MS_A}{MS_E}$ |
| B | $\frac{1}{an}\sum_j y_{.j.}^2 - \frac{y_{...}^2}{abn}$ | $b-1$ | $\sigma^2 + \frac{an\sum\beta_j^2}{b-1}$ | $\frac{MS_B}{MS_E}$ |
| AB | $\frac{1}{n}\sum_i\sum_j y_{ij.}^2 - \frac{y_{...}^2}{abn} - SS_A - SS_B$ | $(a-1)(b-1)$ | $\sigma^2 + \frac{n\sum\sum(\tau\beta)_{ij}^2}{(a-1)(b-1)}$ | $\frac{MS_{AB}}{MS_E}$ |
| Error | Subtraction | $(ab-1)(n-1)$ | $\sigma^2$ | |
| Total | $\sum_i\sum_j\sum_k y_{ijk}^2 - \frac{y_{...}^2}{abn}$ | $abn-1$ | | |

**Revisit Battery Example** See Battery-Life-Blocked-Factorial.SAS.

## Factorial Experiment with Two blocking factors I

- Use Latin square as blocking scheme
  Suppose the experimental factors are $F_1$ and $F_2$. $F_1$ has three
  levels (1,2,3) and $F_2$ has 2 levels. There are $3 \times 2 = 6$
  treatment combinations. These treatments can be represented
  by Latin letters

  | $F_1$ | $F_2$ | Treatment |
  |-------|-------|-----------|
  | 1     | 1     | A         |
  | 1     | 2     | B         |
  | 2     | 1     | C         |
  | 2     | 2     | D         |
  | 3     | 1     | E         |
  | 3     | 2     | F         |

## Factorial Experiment with Two blocking factors II

- Two blocking factors are Block 1 and Block 2, each with 6 blocks

  A $6 \times 6$ Latin square can be used as the blocking scheme:

|         |   |   | Block 1 |   |   |   |
|---------|---|---|---------|---|---|---|
| Block 2 | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | A | B | C | D | E | F |
| 2 | B | C | D | E | F | A |
| 3 | C | D | E | F | A | B |
| 4 | D | E | F | A | B | C |
| 5 | E | F | A | B | C | D |
| 6 | F | A | B | C | D | E |

- Statistical Model

$$y_{ijkl} = \mu + \alpha_i + \tau_j + \beta_k + (\tau\beta)_{jk} + \theta_l + \epsilon_{ijkl}$$

where $\alpha_i$ and $\theta_l$ are blocking effects, $\tau_j$, $\beta_k$, and $(\tau\beta)_{jk}$ are the treatment main effects and interactions.

- Radar Detection Experiment Run in a $6 \times 6$ Latin Square

Radar Detection Experiment Run in a 6 × 6 Latin Square

| Day | Operator | | | | | |
|-----|----------|----------|----------|----------|----------|----------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | $A(f_1g_1 = 90)$ | $B(f_1g_1 = 106)$ | $C(f_1g_1 = 108)$ | $D(f_1g_1 = 81)$ | $F(f_2g_2 = 90)$ | $E(f_2g_2 = 88)$ |
| 2 | $C(f_1g_1 = 114)$ | $A(f_1g_1 = 96)$ | $B(f_1g_1 = 105)$ | $F(f_2g_2 = 83)$ | $E(f_2g_2 = 86)$ | $D(f_2g_2 = 84)$ |
| 3 | $B(f_1g_2 = 102)$ | $E(f_2g_2 = 90)$ | $G(f_2g_2 = 95)$ | $A(f_1g_1 = 92)$ | $D(f_2g_2 = 85)$ | $C(f_1g_1 = 104)$ |
| 4 | $E(f_2g_2 = 87)$ | $D(f_2g_2 = 84)$ | $A(f_1g_1 = 100)$ | $B(f_1g_2 = 96)$ | $C(f_1g_1 = 110)$ | $F(f_2g_2 = 91)$ |
| 5 | $F(f_2g_2 = 93)$ | $C(f_1g_2 = 112)$ | $D(f_2g_1 = 92)$ | $E(f_2g_2 = 80)$ | $A(f_1g_1 = 90)$ | $B(f_1g_2 = 98)$ |
| 6 | $D(f_2g_1 = 86)$ | $F(f_2g_2 = 91)$ | $E(f_2g_2 = 97)$ | $C(f_2g_2 = 98)$ | $B(f_1g_2 = 100)$ | $A(f_1g_1 = 92)$ |

- Analysis of Variance for Radar Detection Experiment Run as a $3 \times 2$

| Source of Variation | Sum of Squares | Degrees of Freedom | General Formula for Degrees of Freedom | Mean Square | $F_0$ | P-Value |
|---|---|---|---|---|---|---|
| Ground clutter, $G$ | 571.50 | 2 | $a - 1$ | 285.75 | 28.86 | <0.0001 |
| Filter type, $F$ | 1469.44 | 1 | $b - 1$ | 1469.44 | 148.43 | <0.0001 |
| $GF$ | 126.73 | 2 | $(a - 1)(b - 1)$ | 63.37 | 6.40 | 0.0071 |
| Days (rows) | 4.33 | 5 | $ab - 1$ | 0.87 | | |
| Operators (columns) | 428.00 | 5 | $ab - 1$ | 85.60 | | |
| Error | 198.00 | 20 | $(ab - 1)(ab - 2)$ | 9.90 | | |
| Total | 2798.00 | 35 | $(ab)^2 - 1$ | | | |