Design and Analysis of Experiments

Lecture 7-Factorial Designs

Ahmed A.(Msc)

DIRE DAWA UNIVERSITY
COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES
DEPARTMENT OF SATATISTICS

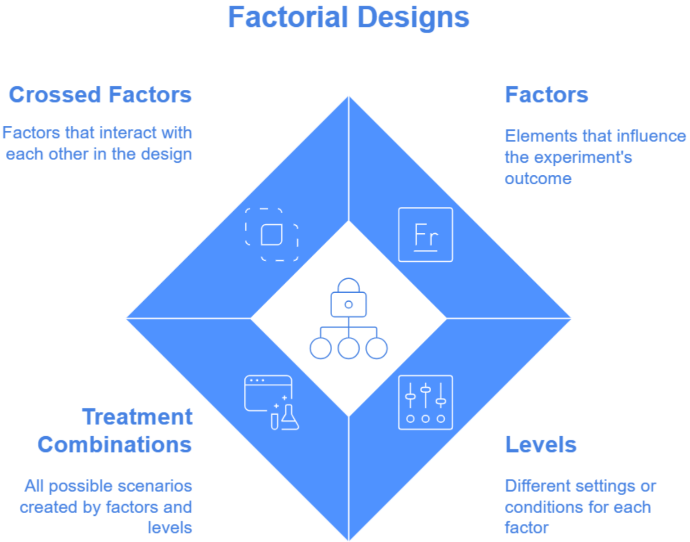
December 10, 2024

Outline

Factorial Designs:Basic definitions and principles

- Many experiments involve the study of the effects of two or more factors.
- **Factorial designs** are the most efficient for this type of experiment.
- In each complete trial or replicate, all possible combinations of factor levels are investigated.
  - Example: If there are a levels of factor A and b levels of factor B, each replicate contains all ab treatment combinations.
- Factors in factorial designs are often said to be **crossed**.

Basic definitions and principles



Advantage of Factorial Design

The Advantage of Factorials

1 Efficiency

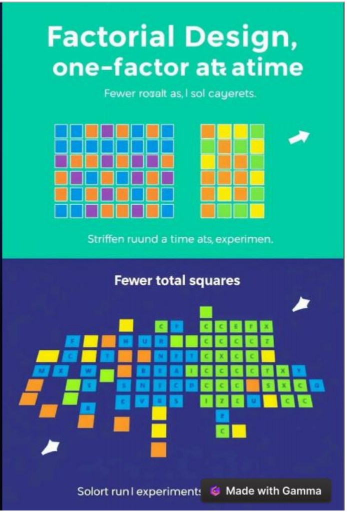
Factorial designs are more efficient than one-factor-at-a-time experiments, requiring fewer experimental runs to obtain the same level of information.

2 Interaction Detection

Factorial designs are essential for detecting interactions between factors, which can be masked in one-factor-at-a-time experiments.

3 Range of Conditions

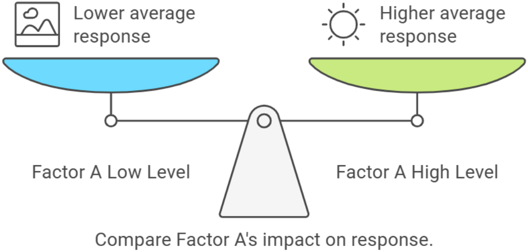
Factorial designs allow the effects of a factor to be estimated at multiple levels of other factors, providing conclusions valid over a wider range of experimental conditions.



Main Effects in Factorial Designs

- The **effect of a factor** is defined as the change in response produced by a change in the level of the factor.
- Called a **main effect** as it refers to the primary factors of interest.
- Example: In a two-factor factorial experiment (Figure 5.1):
  - Both factors have two levels (*low* and *high*, denoted as $-$ and $+$, respectively).
  - Main effect of factor A: The difference between the average response at the low level of A and the high level of A.

Main Effects in Factorial Designs



The two factor factorial design

let y_{ijk} be the observed response when factor A is at the ith level $(i = 1, 2, \ldots, a)$ and factor B is at the jth level $(j = 1, 2, \ldots, b)$ for the kth replicate $(k = 1, 2, \ldots, n)$.

- Observations in a factorial experiment can be described by a model.
- There are several ways to write the model for a factorial experiment.
- **Effects Model:**

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \\ k = 1, 2, \ldots, n \end{cases}$$

  where:

  - μ: Overall mean effect.
  - α_i: Effect of the ith level of the row factor A.
  - β_j: Effect of the jth level of the column factor B.

Factorial Experiment Model: Details

- In the effects model:
  - $(\alpha\beta)_{ij}$: Interaction effect between α_i and β_j.
  - ε_{ijk}: Random error component.
- Assumptions:

$$\sum_{i=1}^{a} \alpha_i = 0, \qquad \sum_{j=1}^{b} \beta_j = 0, \qquad \sum_{i=1}^{a} \sum_{j=1}^{b} (\alpha\beta)_{ij} = 0.$$

- Both factors are assumed to be fixed.
- With n replicates, there are abn total observations.

Alternative Models for Factorial Experiments

- Another possible model is the **means model**,

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \\ k = 1, 2, \ldots, n \end{cases}$$

  where the mean of the ijth cell is:

$$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}.$$

- A **regression model** can also be used, especially when one or more factors are quantitative.
- Throughout most of this chapter, we use the **effects model** (Equation 5.1) with an illustration of the regression model in Section 5.5.

Hypotheses in a Two-Factor Factorial

- In a two-factor factorial, both row and column factors (A and B) are of equal interest.
- Hypotheses to test:
  - Equality of row treatment effects:

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0,$$
$$H_1 : \text{At least one } \alpha_i \neq 0.$$

  - Equality of column treatment effects:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_b = 0,$$
$$H_1 : \text{At least one } \beta_j \neq 0.$$

  - Interaction between row and column treatments:

$$H_0 : (\alpha\beta)_{ij} = 0 \text{ for all } i, j,$$
$$H_1 : \text{At least one } (\alpha\beta)_{ij} \neq 0.$$

Statistical Analysis of the Fixed Effects Model

- Let $y_{i..}$ denote the total of all observations under the ith level of factor A.
- Let $y_{.j.}$ denote the total of all observations under the jth level of factor B.
- Let $y_{ij.}$ denote the total of all observations in the ijth cell.
- Hypotheses are tested using a **two-factor analysis of variance**.

Summation Notation for Factorial Experiments

$$y_{i..} = \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{i..} = \frac{y_{i..}}{bn}, \quad i = 1, 2, \ldots, a$$

$$y_{.j.} = \sum_{i=1}^{a} \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{.j.} = \frac{y_{.j.}}{an}, \quad j = 1, 2, \ldots, b$$

$$y_{ij.} = \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{ij.} = \frac{y_{ij.}}{n}, \quad i = 1, 2, \ldots, a, \; j = 1, 2, \ldots, b$$

$$y_{...} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{...} = \frac{y_{...}}{abn}$$

ANOVA Equation for Two-Factor Factorial

- The total corrected sum of squares is expressed as:

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{...})^2$$

- This can be partitioned as:

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E,$$

where:

- SS_A: Sum of squares for factor A (rows).
- SS_B: Sum of squares for factor B (columns).
- SS_{AB}: Sum of squares for interaction.
- SS_E: Sum of squares for error.

Degrees of Freedom for ANOVA

- The number of degrees of freedom (DF) for each component:

Effect	Degrees of Freedom (DF)
Factor A (SS_A)	$a-1$
Factor B (SS_B)	$b-1$
Interaction (SS_{AB})	$(a-1)(b-1)$
Error (SS_E)	$ab(n-1)$
Total (SS_T)	$abn-1$

- Justification for the degrees of freedom:
  - Total DF: $abn-1$.
  - Main effects DF: $a-1$ for A and $b-1$ for B.
  - Interaction DF: $ab-1-(a-1)-(b-1)$.

Example

An Example

An engineer is designing a battery for a device that will experience extreme temperature variations. The engineer wants to investigate the effect of plate material on battery life under different temperature conditions. This scenario presents a two-factor factorial design, with plate material as one factor and temperature as the other factor.

Interaction Effect: Material and Temperature

Example: Battery Life

■ **TABLE 5.1**

Life (in hours) Data for the Battery Design Example

Material Type	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

Example: Battery Life

■ **TABLE 5.4**

Life Data (in hours) for the Battery Design Experiment

Material Type	Temperature (°F) 15			70			125			$y_{i..}$
	130	155	(539)	34	40	(229)	20	70	(230)	
1	74	180		80	75		82	58		998
	150	188	(623)	136	122	(479)	25	70	(198)	
2	159	126		106	115		58	45		1300
	138	110	(576)	174	120	(583)	96	104	(342)	
3	168	160		150	139		82	60		1501
$y_{.j.}$	1738			1291			770			$3799 = y_{...}$

Example: Battery Life

Using Equations 5.6 through 5.10, the sums of squares are computed as follows:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{abn}$$

$$= (130)^2 + (155)^2 + (74)^2 + \cdots$$

$$+ (60)^2 - \frac{(3799)^2}{36} = 77,646.97$$

$$SS_{\text{Material}} = \frac{1}{bn}\sum_{i=1}^{a} y_{i..}^2 - \frac{y_{...}^2}{abn}$$

$$= \frac{1}{(3)(4)}[(998)^2 + (1300)^2 + (1501)^2]$$

$$- \frac{(3799)^2}{36} = 10,683.72$$

Example: Battery Life

$$
\begin{aligned}
SS_{\text{Temperature}} &= \frac{1}{an}\sum_{j=1}^{b} y_{.j.}^2 - \frac{y_{...}^2}{abn} \\
&= \frac{1}{(3)(4)}\left[(1738)^2 + (1291)^2 + (770)^2\right] \\
&\quad - \frac{(3799)^2}{36} = 39{,}118.72
\end{aligned}
$$

Example: Battery Life

$$SS_{\text{Interaction}} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij.}^2 - \frac{y_{...}^2}{abn} - SS_{\text{Material}}$$

$$- SS_{\text{Temperature}}$$

$$= \frac{1}{4}[(539)^2 + (229)^2 + \cdots + (342)^2]$$

$$- \frac{(3799)^2}{36} - 10,683.72$$

$$- 39,118.72 = 9613.78$$

and

$$SS_E = SS_T - SS_{\text{Material}} - SS_{\text{Temperature}} - SS_{\text{Interaction}}$$

$$= 77,646.97 - 10,683.72 - 39,118.72$$

$$- 9613.78 = 18,230.75$$

Example: Battery Life

■ **TABLE 5.5**

Analysis of Variance for Battery Life Data

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Material types	10,683.72	2	5,341.86	7.91	0.0020
Temperature	39,118.72	2	19,559.36	28.97	< 0.0001
Interaction	9,613.78	4	2,403.44	3.56	0.0186
Error	18,230.75	27	675.21		
Total	77,646.97	35			

ANOVA Results Overview

- The ANOVA is shown in **Table 5.5**.
- Because $F_{0.05,4,27} = 2.73$, we conclude that there is a significant interaction between material types and temperature.
- Furthermore, $F_{0.05,2,27} = 3.35$, so the main effects of material type and temperature are also significant.
- Table 5.5 also shows the *P-values* for the test statistics.
- To assist in interpreting the results of this experiment, it is helpful to construct a graph of the average responses at each treatment combination.

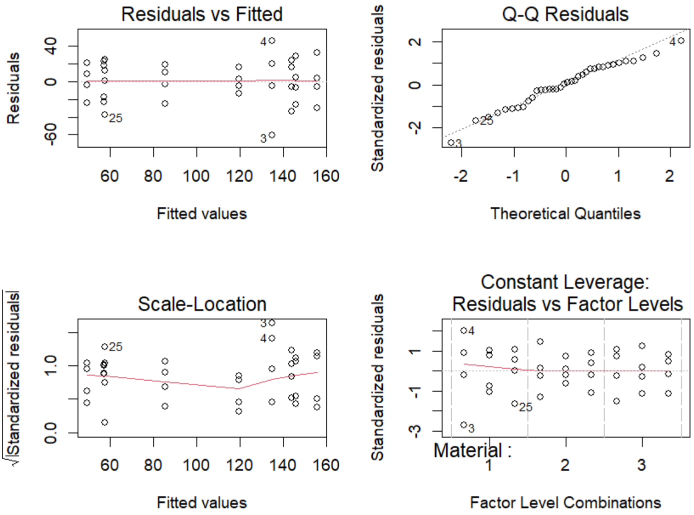
Factorial Design Model Summary Using R

```
# Fit the factorial design model
> factorial_model <- aov(Life ~ Material * Temp, data = data.b
> # Display the model summary
> summary(factorial_model)
Df Sum Sq Mean Sq F value   Pr(>F)
Material        2  10684    5342   7.911  0.00198 **
Temp            2  39119   19559  28.968 1.91e-07 ***
Material:Temp   4   9614    2403   3.560  0.01861 *
Residuals      27  18231     675
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model Adequacy Checking

- Before the conclusions from the ANOVA are adopted, the adequacy of the underlying model should be checked.
- As before, the primary diagnostic tool is residual analysis.

Model Adequacy Checking



Model Adequacy Checking

```
#1 Shapiro-Wilk test for normality of residuals

shapiro.test(residuals(factorial_model))

Shapiro-Wilk normality test

data:  residuals(factorial_model)
W = 0.97606, p-value = 0.6117
```

- the P-value greater than $\alpha = 0.05$, then we conclude that the Residuals is Normally Distributed.

Model Adequacy Checking

2. Homogeneity Test

```
fligner.test(Life ~ interaction(Material , Temp), data = data.

Fligner-Killeen test of homogeneity of
variances

data:  Life by interaction(Material, Temp)
Fligner-Killeen:med chi-squared = 5.667, df
= 8, p-value = 0.6845
```

Fligner-Killeen Test Interpretation

Test Description:

- `Life   interaction(Material, Temp)`
- Assesses whether the variances of the response variable (`Life`) are equal across the groups formed by the interaction of `Material` and `Temp`.

Output Summary:

- **Test Statistic:** Fligner-Killeen:med chi-squared = 5.667
- **Degrees of Freedom:** df = 8
- **p-value:** 0.6845

Fligner-Killeen Test Interpretation

Conclusion:

- Null Hypothesis (H_0): Variances across groups are equal.
- Since **p-value (0.6845)** $> \alpha = 0.05$, we **fail to reject the null hypothesis**.
- There is no evidence to suggest significant differences in variances across the groups.

Practical Implications:

- Homogeneity of variances is satisfied.
- Suitable to proceed with ANOVA.

Model Adequacy Checking

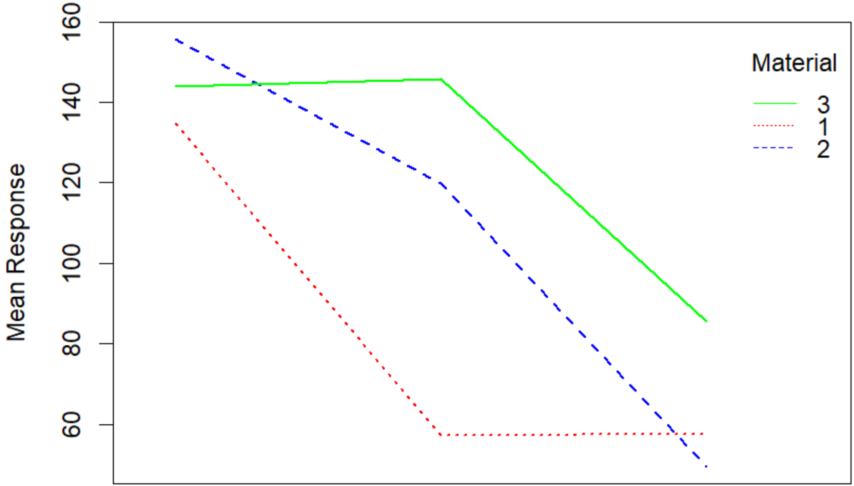
```
# Levene's Test for homogeneity of variances
> leveneTest(Life ~ Material * Temp, data = data.battery)
Levene's Test for Homogeneity of Variance (center = median)
Df F value Pr(>F)
group  8  0.7996 0.6081
27
```

Model Adequacy Checking

Interaction plots for visualization

```
interaction.plot(data.battery$Temp, data.battery$Material,
data.battery$Life,
col = c("red", "blue", "green"), lwd = 2,
ylab = "Mean Response", xlab = "Temperature",
trace.label = "Material")
```

Example: Battery Life



The Assumption of No Interaction in a Two-Factor Model

Overview:

- Occasionally, an experimenter assumes a two-factor model without interaction is appropriate:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \\ k = 1, 2, \ldots, n \end{cases}$$

- This simplifies the model and the analysis.

The Assumption of No Interaction in a Two-Factor Model

Key Considerations:

- Be cautious when excluding interaction terms.
- Significant interaction can dramatically affect data interpretation.

Analysis Without Interaction:

- Straightforward statistical analysis is possible.
- Example: Table 5.8 presents the analysis of the battery life data (Example 5.1) under the assumption of no interaction.

General Factorial Design

- Extends the two-factor factorial design to a general case with:
  - a levels of factor A
  - b levels of factor B
  - c levels of factor C, and so on.
- Total observations: $abc \ldots n$, where n is the number of replicates.
- At least two replicates ($n \geq 2$) are required to calculate the sum of squares due to error.

Fixed Effects Model

- Hypotheses for main effects and interactions can be tested using ANOVA.
- For a fixed effects model:
  - Test statistics are computed by dividing the mean square for the effect or interaction by the mean square error.
  - All F-tests are upper-tail, one-tailed tests.
- Degrees of freedom:
  - Main effect: levels of factor -1.
  - Interaction: Product of the degrees of freedom for individual components.

Three-Factor Analysis of Variance Model

- Assuming A, B, and C are fixed factors:
  - Analysis of variance table is constructed.
  - F-tests for main effects and interactions are based on expected mean squares.
- Manual formulas for sums of squares are occasionally useful.

Total Sum of Squares

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} y_{ijkl}^2 - \frac{y_{....}^2}{abcn} \quad (1)$$

Main Effects Sums of Squares

$$SS_A = \frac{1}{bcn}\sum_{i=1}^{a} y_{i\ldots}^2 - \frac{y_{\ldots}^2}{abcn} \tag{2}$$

$$SS_B = \frac{1}{acn}\sum_{j=1}^{b} y_{.j\ldots}^2 - \frac{y_{\ldots}^2}{abcn} \tag{3}$$

$$SS_C = \frac{1}{abn}\sum_{k=1}^{c} y_{..k\ldots}^2 - \frac{y_{\ldots}^2}{abcn} \tag{4}$$

Conclusion

- ANOVA computations are often done using statistical software.
- Manual computations of sums of squares can aid in understanding.
- The general factorial design provides a framework for analyzing complex experiments with multiple factors and levels.