Design of experiments with several factors

Objectives of the topic:

- Designing and conducting engineering experiments involving several factors.
- Analyzing experiments using the analysis of variance.
- Assessing model adequacy.
- Using the two level series of factorial designs.

- A factorial experiment consists of experimental trials performed at all combinations of levels of factors.
- The industrial engineer could be interested to investigate the effect of the machine model and the feed rate on the number of defective parts.
- He wants to find the best levels of the two factors (machine model and feed rate) that will yield the least number of defectives.
- The analysis of variance will be used in the analysis.

Factorial experiment

- When there are two factors A and B with ‘a’ levels for factor A and ‘b’ levels for factor B, each replicate of the experiment contains ‘ab’ treatment combinations.
- The effect of a factor is taken as the change in response (called main effect) produced by a change in the level of the factor.
- Consider the factorial experiment:

	Factor B	
Factor A	B_{low}	B_{high}
A_{low}	10	20
A_{high}	30	40

- The main effect of factor A is the difference between the average response at the high level of A and the average response at the low level of A:

$$A = \frac{30+40}{2} - \frac{10+20}{2} = 20$$

- Hence, changing the factor A from the low level to the high level causes an average response increase of 20.
- When the difference in response between the levels of one factor is not the same at all levels of the other factors, an interaction between the factors exists.

- Consider the factorial experiment:

	Factor B	
Factor A	B_{low}	B_{high}
A_{low}	10	20
A_{high}	30	0

- At the low level of B, the A effect is 30 – 10 = 20.
- At the high level of B, the A effect is 0 – 20 = -20.
- The A effect depends on the level chosen for factor B.
- The main effect of A is

$$A = \frac{30+0}{2} - \frac{10+20}{2} = 0$$

- To say factor A has no effect may not be correct.
- In this case, the AB interaction is more useful than the main effect.
- The estimate of the AB interaction effect in factorial experiments is the difference in the diagonal averages:

Factor A	Factor B: B_{low}	B_{high}
A_{low}	10	20
A_{high}	30	40

$$AB = \frac{30+20}{2} - \frac{10+40}{2} = 0$$

Factor A	Factor B: B_{low}	B_{high}
A_{low}	10	20
A_{high}	30	0

$$AB = \frac{20+30}{2} - \frac{10+0}{2} = 20$$

Two factor factorial experiments

- A two factor factorial experiment with n replicates is represented as:

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

- The yijk observation under level i of factor A, level j of factor B, and replicate k, is represented by

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

where μ is the overall average, τ_i is the effect of the ith level of factor A, β_j is the effect of the jth level of factor B, $(\tau\beta)_{ij}$ is the effect of the interaction between A and B, and $\in_{ijk}$ is a random error.

- The error is assumed to be normal with mean zero and variance σ^2.
- The effects τ_i, β_j, and $(\tau\beta)_{ij}$ are defined as deviations from the overall mean such that:

$$\sum_{i=1}^{a} \tau_i = 0, \sum_{j=1}^{b} \beta_j = 0, \sum_{i=1}^{a} (\tau\beta)_{ij} = 0, \text{and} \sum_{j=1}^{b} (\tau\beta)_{ij} = 0$$

Tests of significance of effects

- Take two fixed factors A and B, with ‘a’ levels and ‘b’ levels respectively.
- The analysis of variance can be used to test the hypothesis about the main factor effects of A and B and AB interaction:
  1. Hypothesis on no main effect of factor A:

     $H_0: \tau_1 = \tau_2 = \ldots = \tau_a = 0$
  2. Hypothesis on no main effect of factor B:

     $H_0: \beta_1 = \beta_2 = \ldots = \beta_b = 0$
  3. Hypothesis on no interaction:

     $H_0: (\tau\beta)_{11} = (\tau\beta)_{12} = \ldots = (\tau\beta)_{ab} = 0$

- The ANOVA identity is given by

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

$$+an\sum_{j=1}^{b}(\bar{y}_{.j.}-\bar{y}_{...})^2$$

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

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

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

$$(abn-1) = (a-1)+(b-1)+(a-1)(b-1)+ab(n-1)$$

- When the factors are fixed factors, the expected values of the means squares are

$$E(MS_A) = \sigma^2 + \frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1}$$

$$E(MS_B) = \sigma^2 + \frac{an\sum_{j=1}^{b}\beta_j^2}{b-1}$$

$$E(MS_{AB}) = \sigma^2 + \frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{(a-1)(b-1)}$$

$$E(MS_E) = \sigma^2$$

- When the three hypotheses on the main effects are true, the four mean squares are unbiased estimators of the variance.
- To test the hypothesis on no effect of factor A, the test statistics is

$$F_0 = \frac{MS_A}{MS_E}$$

which has an F distribution with a-1 and ab(n-1) d.f.

- To test the hypothesis on no effect of factor B, the test statistic is

$$F_0 = \frac{MS_B}{MS_E}$$

which has an F distribution with b-1 and ab(n-1) d.f.

- To test the hypothesis on no effect of the interaction, the test statistic is

$$F_0 = \frac{MS_{AB}}{MS_E}$$

which has an F distribution with (a-1)(b-1) and ab(n-1) d.f.

- If a significant interaction effect exists between the factors, then examining the main effects of the factors has no practical value.
- Therefore, it is recommended to test for the significance of the interaction first.

- The summations can easily computed as

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

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

$$SS_B = \sum_{j=1}^{b} \frac{y_{.j.}^2}{an} - \frac{y_{...}^2}{abn}$$

$$SS_{AB} = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{y_{ij.}^2}{n} - \frac{y_{...}^2}{abn} - SS_A - SS_B$$

$$SS_E = SS_T - SS_{AB} - SS_A - SS_B$$

- The computations are organized in the ANOVA table:

Source of Variation	Sum of squares	Degrees of freedom	Mean squares	F_0
A treatments	SS_A	a-1	MS_A	MS_A/MS_E
B treatments	SS_B	b-1	MS_B	MS_B/MS_E
Interaction	SS_{AB}	(a-1)(b-1)	MS_{AB}	MS_{AB}/MS_E
Error	SS_E	ab(n-1)	MS_E	
Total	SS_T	abn-1		

Example

- The industrial engineer is interested to know what kind of metal type will lead to a strong adhesion of color using two application methods: dipping or spraying.
- The plant he works with can produce three types of metal.
- Under each type of metal, he used both coloring methods and made three observations, completely random:

	Application method					
Metal type	Dipping			Spraying		
1	4	4.5	4.3	5.4	4.9	5.6
2	5.6	4.9	5.4	5.8	6.1	6.3
3	3.8	3.7	4	5.5	5	5

- How significant the interaction between the metal type and application method is?
- How significant the main effects of the factors are?

- The ANOVA table is:

Source of Variation	SS	d.f.	MS	f_0
Metal type	4.58	2	2.29	28.63
Application method	4.91	1	4.91	61.38
Interaction	0.24	2	0.12	1.50
Error	0.99	12	0.08	
Total	10.72	17		

- Since $1.50 < f_{0.05,2,12} = 3.88$, the interaction between factors is not significant.
- Since $28.63 > f_{0.05,2,12} = 3.88$ and $61.38 > f_{0.05,1,12} = 4.75$, it is concluded that the metal type and the application method affect color adhesion.

Hypothesis on the difference between levels means

- If either of the hypotheses on the factor levels is rejected, then testing for significant difference between two levels of a specific factor is of interest.
- Considering the factor A, the hypothesis on the difference between two levels means is:

$$H_0: \mu_{i.} = \mu_{j.}$$
$$H_1: \mu_{i.} \neq \mu_{j.}$$

- The test statistic is

$$T_0 = \frac{\bar{y}_{i..} - \bar{y}_{j..}}{\sqrt{2\frac{MS_E}{bn}}}$$

which has a t distribution with ab(n-1) degrees of freedom.

- For factor B, the hypothesis on the difference between two levels means is:

$$H_0: \mu_{.i} = \mu_{.j}$$
$$H_1: \mu_{.i} \neq \mu_{.j}$$

- The test statistic is

$$T_0 = \frac{\bar{y}_{.i.} - \bar{y}_{.j.}}{\sqrt{2\frac{MS_E}{an}}}$$

Example

- Consider the data:

	Application method					
Metal type	Dipping			Spraying		
1	4	4.5	4.3	5.4	4.9	5.6
2	5.6	4.9	5.4	5.8	6.1	6.3
3	3.8	3.7	4	5.5	5	5

- It has been proven that there is no significant interaction between the two factors, and also there are significant effects of the factors.
- Are the levels means of 1 and 2 of factor A equal?

- The following hypothesis has to be tested:

$$H_0: \mu_{1.} = \mu_{2.}$$
$$H_1: \mu_{1.} \neq \mu_{2.}$$

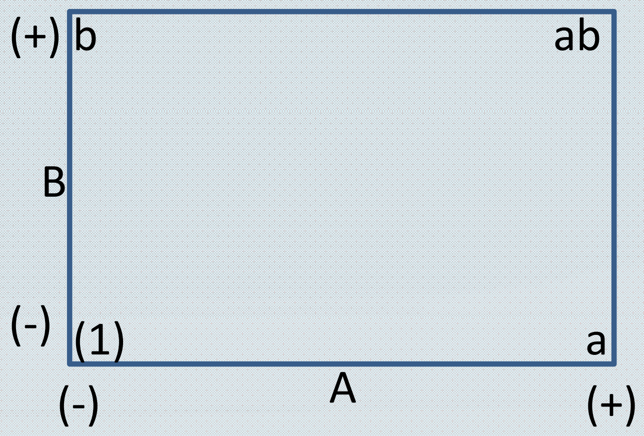
- The value of the test statistic is

$$t_0 = \frac{4.78 - 5.68}{\sqrt{2\frac{0.08}{(2)(3)}}} = -5.51$$

- From the t tables, $t_{0.025,12} = 2.18$.
- Since $|t_0| > t_{0.025,12}$, H_0 is rejected.
- There is a significant difference between levels means of 1 and 2 of factor A.

2^k factorial designs

- One special case of factorial designs is that of k factors, each at only two levels.
- A complete replicate of such a design requires $2 \times 2 \times \ldots \times 2 = 2^k$ combinations of levels; hence, it is called 2^k factorial design.
- The simplest type of 2^k design is the 2^2: two factors A and B, each at two levels.
- Usually, the two levels are labeled as low (denoted by -) and high (denoted by +).



Labeling

- A special notation is used to label the treatment combinations.
- A treatment combination is represented by a series of lowercase letters.
- If a letter is present, the corresponding factor is run at the high level.
- It a letter is absent, the factor is run at its low level.
- The treatment combination with both factors at the low level is represented by (1).

Effects estimation

- The effects in the 2^2 design are the main effects A and B and the two factor interaction AB.
- Let the letters (1), a, b, and ab represent the totals of n observations taken at the design points.
- The main effect of A is estimated as the difference of the average of the observations at the high level of A and the average of the observation at the low level of A:

$$
\begin{aligned}
A &= \bar{y}_{A+} - \bar{y}_{A-} \\
&= \frac{a+ab}{2n} - \frac{b+(1)}{2n} \\
&= \frac{1}{2n}[a+ab-b-(1)]
\end{aligned}
$$

- Similarly, the main effect of B is

$$B = \bar{y}_{B+} - \bar{y}_{B-}$$

$$= \frac{b + ab}{2n} - \frac{a + (1)}{2n}$$

$$= \frac{1}{2n}[b + ab - a - (1)]$$

- The interaction effect is estimated as

$$AB = \frac{ab + (1)}{2n} - \frac{a + b}{2n}$$

$$= \frac{1}{2n}[ab + (1) - a - b]$$

- The quantities between the brackets are called contrasts.
- The A contrast is

$$\text{contrast}_A = a + ab - b - (1)$$

- The contrast coefficients are either +1 or -1.
- For convenience, the contrasts are usually organized in the following table:

Treatment combination	Factorial effect A	B	AB	Total
(1)	-	-	+	+
a	+	-	-	+
b	-	+	-	+
ab	+	+	+	+

- In the contrast table, the AB column is the result of the multiplication of columns A and B.
- The total is the result of the multiplication of columns A, B, and AB.
- From the contrast table, the sum of squares can be computed:

$$SS_A = \frac{[a + ab - b - (1)]^2}{4n}$$

$$SS_B = \frac{[b + ab - a - (1)]^2}{4n}$$

$$SS_{AB} = \frac{[ab + (1) - a - b]^2}{4n}$$

with one degree of freedom for each.

- When a linear regression model is used, the total sum of squares is

$$SS_T = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n} y_{ijk}^2 - \frac{\left(\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n} y_{ijk}\right)^2}{4n}$$

with 4n – 1 degrees of freedom.

- The error sum of squares is found by subtraction:

$$SS_E = SS_T - SS_A - SS_B - SS_{AB}$$

with 4(n – 1) degrees of freedom.

- The ANOVA table is

Source of variation	Sum of squares	Degrees of freedom	Mean Squares	F_0
A	SS_A	1	MS_A	MS_A/MS_E
B	SS_B	1	MS_B	MS_B/MS_E
AB	SS_{AB}	1	MS_{AB}	MS_{AB}/MS_E
Error	SS_E	4(n - 1)	MS_E	
Total	SS_T	4n - 1		

Example

- In a heater production line, an insulator is applied to the inside surface of the heater to prevent corrosion.
- The thickness of the insulator in mm is of special interest to the quality engineer.
- It is thought that the thickness is affected by two factors: application time (A) and application flow rate (B).
- The levels of factor A are 2 seconds (-) and 5 seconds (+).
- The levels of factor B are 1 cubic cm / second (-) and 1.3 cubic cm / second.

- He made 4 observations under each combination of the factors:

Factor A	Factor B: -				Factor B: +			
-	14.037	14.165	13.972	13.907	13.880	13.860	14.032	13.914
+	14.821	14.757	14.843	14.878	14.888	14.921	14.415	14.932

- The quality engineer is interested in assessing the significance of individual factors and their interaction.

- The treatment combinations are

Treatment combination	Thickness				Total
(1)	14.037	14.165	13.972	13.907	56.081
a	14.821	14.757	14.843	14.878	59.299
b	13.88	13.86	14.032	13.914	55.686
ab	14.888	14.921	14.415	14.932	59.156

- The estimates of the effects are

$$A = \frac{1}{2(4)}[59.299 + 59.156 - 55.686 - 56.081] = 0.836$$

$$B = \frac{1}{2(4)}[55.686 + 59.156 - 59.299 - 56.081] = -0.067$$

$$AB = \frac{1}{2(4)}[59.156 + 56.081 - 59.299 - 55.686] = 0.032$$

- From the estimates of the main effects of the factors:
  - Application time (factor A) has a large effect on the thickness, and increasing the application time is going to increase the thickness.
  - Effects of the application flow rate (factor B) and the interaction (AB) have small effect on the thickness.
- ANOVA can be applied to further assess the effects of the factors:

Source of variation	Sum of squares	Degrees of freedom	Mean Squares	f_0
A	2.7956	1	2.7956	134.40
B	0.0181	1	0.0181	0.87
AB	0.0040	1	0.0040	0.19
Error	0.2495	12	0.0208	
Total	3.0672	15		

- From the F tables, $f_{0.05, 1, 12} = 4.747$.
- Since $134.40 > f_{0.05, 1, 12}$, the effect of factor A is significant.
- Since 0.87 and $0.19 < f_{0.05, 1, 12}$, the effects of factor B and the interaction AB are not significant.