# Primer on Analysis of Experimental Data and Design of Experiments

## *Lecture 8. Statistical Design of Experiments*

Muhammad A. Alam
alam@purdue.edu

# Outline

1. Context and background
2. Full factorial DOE explained by a toy model
3. Taguchi and fractional factorial experiments to reduce the number of experiments
4. Conclusions

# Design of Experiments

- Set of guidelines for designing, conducting and analyzing experiments for system optimization

- Foundations of DOE were laid by Sir. R.A. Fisher in early 1920s (Analysis of Farm data, output as good as input).

- Concepts of Orthogonal arrays were introduced by Taguchi in 1950s. (Formalized the whole analysis)

- DOE has revolutionized quality control/reliability in all fields of science and technology (Toyota was one of the early adopter, most semiconductor companies use the method).

# Philosophical shift with DOE

## Before Fisher …

Experimentalist determine what experiments to do

Results →

Statisticians/ Theorists/Expt collaborate to interpret results

## After Fisher …

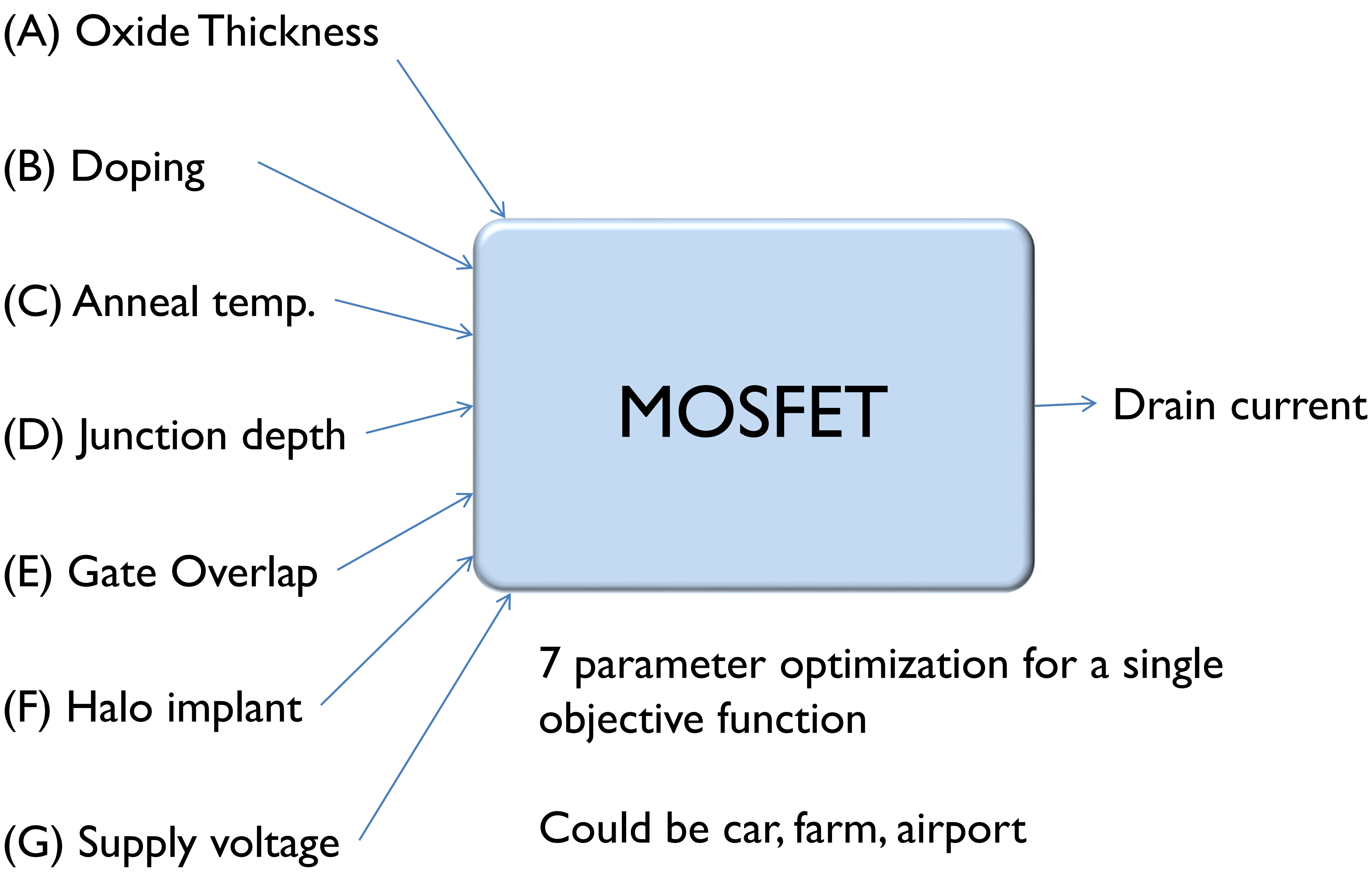
Statisticians/ Theorists/Expt plan what experiments to do

Results →

Statisticians/ Theorists/Expt collaborate to interpret results

Output cannot be greater than input …..

# Problem definition (7 factor problem)




7 parameter optimization for a single objective function

Could be car, farm, airport

# Definition of terms

| Factor | Level | Run/trial/replicate |
|---|---|---|
| Tox | 1, 2, 3 nm | (2 nm, $10^{17}$ cm$^{-3}$, 4 μm)$_{rep}$ |
| Doping | $10^{16}$, $10^{17}$ cm$^{-3}$ | |
| Lch | 2, 3, 4 μm | |

- 1 factor, 3 level, 4 replicate experiment
- 2 factor, 2 level, 3 replicate experiment
- 8 factor, 2 level, 1 replicate experiment

# 7 Factor, 2 level: One factor at a time

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \epsilon$$

| | A | B | C | D | E | F | G | Output |
|---|---|---|---|---|---|---|---|---|
| Run 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 |
| Run 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 15 |
| Run 3 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 12 |
| Run 4 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 9 |
| Run 5 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 18 |
| Run 6 | **2** | **1** | **1** | **2** | **2** | **1** | **1** | **19** |
| Run 7 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 17 |
| Run 8 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 13 |
| **Final** | **2** | **1** | **1** | **2** | **2** | **1** | **1** | **19** |

Simple, widely used, but non-optimum solutions

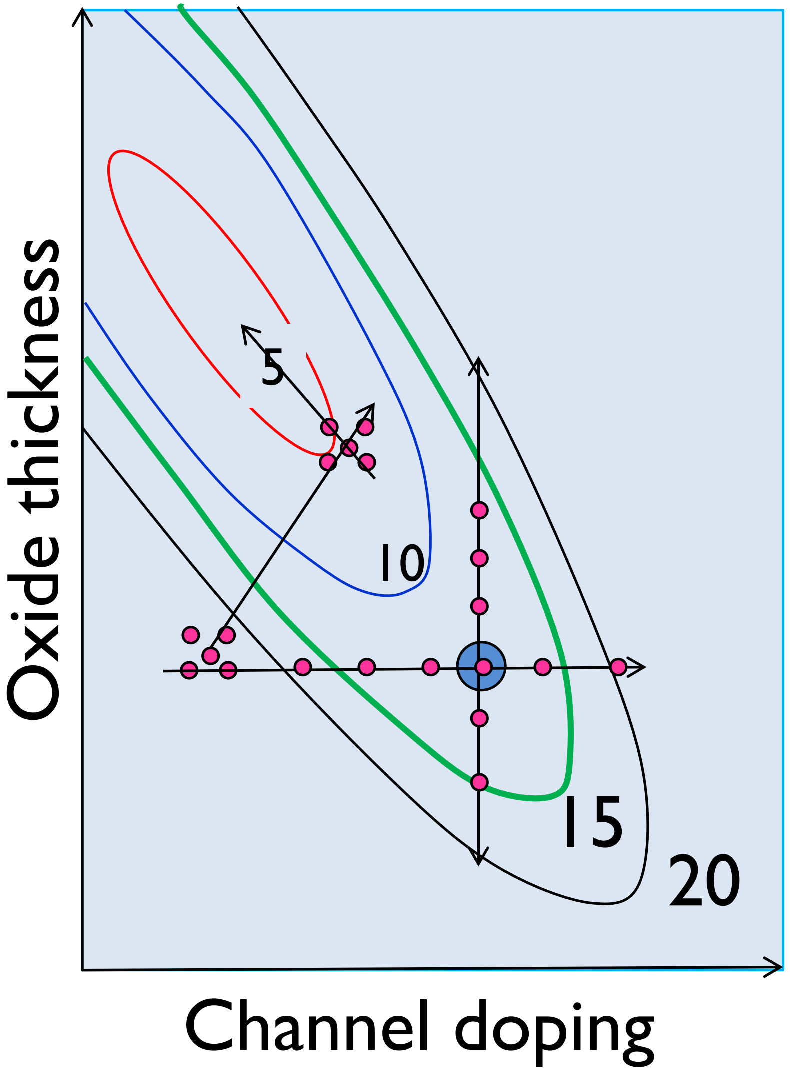
# 7 Factor, 2 Level: Full factorial analysis

$$Level^{factor} = L^F = 2^7 = 128$$

| | | | | $A_1$ | | | | $A_2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $B_1$ | | $B_2$ | | $B_1$ | | $B_2$ | |
| | | | | $C_1$ | $C_2$ | $C_1$ | $C_2$ | $C_1$ | $C_2$ | C1 | $C_2$ |
| $D_1$ | $E_1$ | $F_1$ | $G_1$ | R-1 (10) | | | | R-2 (15) | | R-3 (12) | |
| | | | $G_2$ | | | | | | | | |
| | | $F_2$ | $G_1$ | | | | | | R-4 (9) | | |
| | | | $G_2$ | | | | | | | | |
| | $E_2$ | $F_1$ | $G_1$ | | | | | | | | |
| | | | $G_2$ | | | | | | | | |
| | | $F_2$ | $G_1$ | | | | | | | | |
| | | | $G_2$ | | | | | | | | |
| $D_2$ | $E_1$ | $F_1$ | $G_1$ | | | | | R-5 (18) | | | |
| | | | $G_2$ | | | | | | | | |
| | | $F_2$ | $G_1$ | | | | | | | | |
| | | | $G_2$ | | | | | | | | |
| | $E_2$ | $F_1$ | $G_1$ | | | | | R-6 (19) | | | |
| | | | $G_2$ | | | | | R-8 (13) | | | |
| | | $F_2$ | $G_1$ | | | | | R-7 (17) | | | |
| | | | $G_2$ | | | | | | | | |

Single parameter method is a fractional non-optimal factorial method: After A2 win, will never visit A1. After B2 loss, will never visit B2. Same for C2 Column, etc.

# The problem with one-at-a-time approach

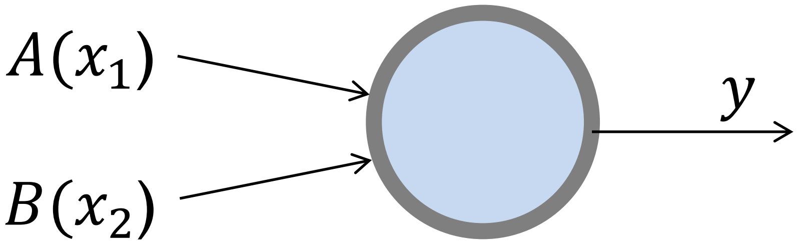


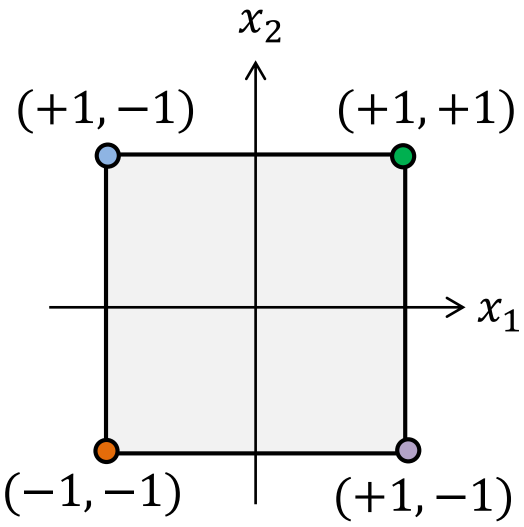

Response surface
Orthogonal sampling

# Outline

1. Context and background
2. Full factorial DOE explained by a toy model
3. Taguchi and fractional factorial experiments to reduce the number of experiments
4. Conclusions

# A Toy Problem with two factors and two levels




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

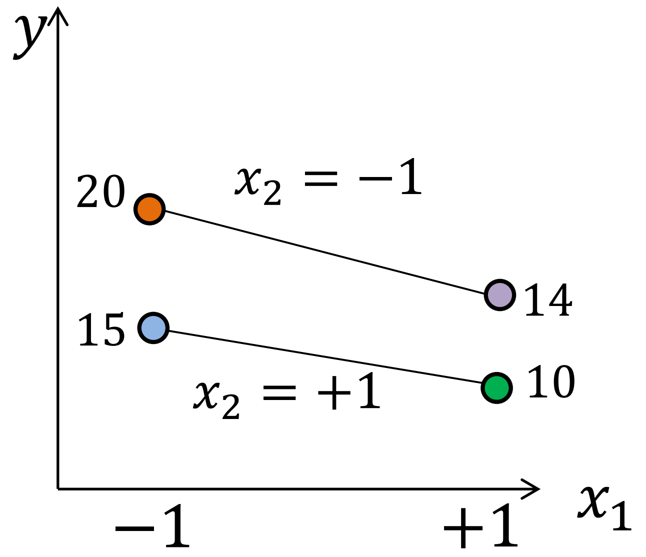
| | | |
|---|---|---|
| $A$ | $200 \ (x_1 = -1)$ | $300 \ (x_1 = +1)$ |
| $B$ | $0 \ (x_2 = -1)$ | $> 1 \ (x_2 = +1)$ |

| $x_1$ | $x_2$ | $x_1 x_2$ | | $y$ |
|---|---|---|---|---|
| $-1$ | $-1$ | $+1$ | (1) | 20 |
| $+1$ | $-1$ | $-1$ | $a$ | 14 |
| $-1$ | $+1$ | $-1$ | $b$ | 15 |
| $+1$ | $+1$ | $+1$ | $ab$ | 10 |

# A toy problem with two factors

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

| $x_1$ | $x_2$ | $x_1 x_2$ | | $y$ |
|---|---|---|---|---|
| $-1$ | $-1$ | $+1$ | $(1)$ | 20 |
| $+1$ | $-1$ | $-1$ | $a$ | 14 |
| $-1$ | $+1$ | $-1$ | $b$ | 15 |
| $+1$ | $+1$ | $+1$ | $ab$ | 10 |




Effect of A=$[(a + ab) - ((1) + b))]/2 = -5.5$

Effect of B=$[(b + ab) - ((1) + a))]/2 = -4.5$

Regression coefficient
=Effect/(+1-(-1))=Effect/2=slope

AB interaction= $[((1) + ab) - (a + b))]/2 = 0.5$

# Randomization: Importance of the error term

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

Sample selection is critical

For the existing problem, randomize over unknown variables such as sub-areas, gender, fall/winter session, etc.

For dielectric breakdown, choose samples from multiple tools, days of the week, randomize test equipment, etc.

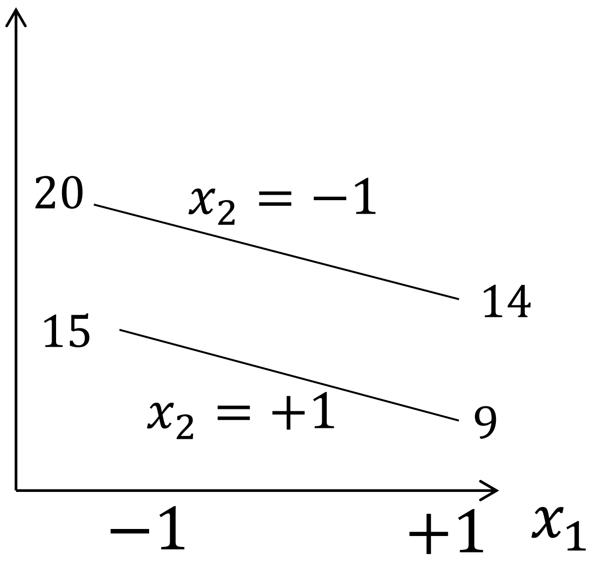
Don't trust yourself: use a random number generator.

# A Toy Problem with two factors: Fractional factorial experiments

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

Choose $x_1 x_2 = +1$ OR $x_1 x_2 = -1$

| | $x_1$ | $x_2$ | $x_1 x_2$ | | $y$ | |
|---|---|---|---|---|---|---|
| | $-1$ | $-1$ | $+1$ | $(1)$ | 20 | 0 |
| | $+1$ | $-1$ | $\boxed{-1}$ | $a$ | 14 | $-6$ |
| | $-1$ | $+1$ | $\boxed{-1}$ | $b$ | 15 | $-5$ |
| | $+1$ | $+1$ | $+1$ | $ab$ | 10 | |




Effect of A=$[a]/1 = -6$

Effect of B=$[b]/1 = -5$

Regression coefficient
=Effect/(+1-(-1))=Effect/2=slope

# A Toy Problem with 3 factors, 2 levels

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{23} x_2 x_3 + \beta_{13} x_1 x_3 + \beta_{123} x_1 x_2 x_3 + \epsilon$$

| $x_1$ | $x_2$ | $x_3$ | $x_1x_2$ | $x_2x_3$ | $x_1x_3$ | $x_1x_2x_3$ | | $y$ |
|---|---|---|---|---|---|---|---|---|
| $-1$ | $-1$ | $-1$ | $+1$ | $+1$ | $+1$ | $-1$ | $(1)$ | 67 |
| $+1$ | $-1$ | $-1$ | $-1$ | $+1$ | $-1$ | $+1$ | $a$ | 79 |
| $-1$ | $+1$ | $-1$ | $-1$ | $-1$ | $+1$ | $+1$ | $b$ | 61 |
| $+1$ | $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $-1$ | $ab$ | 75 |
| $-1$ | $-1$ | $+1$ | $+1$ | $-1$ | $-1$ | $+1$ | $c$ | 59 |
| $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $+1$ | $-1$ | $ac$ | 90 |
| $-1$ | $+1$ | $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $bc$ | 52 |
| $+1$ | $+1$ | $+1$ | $+1$ | $+1$ | $+1$ | $+1$ | $abc$ | 87 |

Step 1: Choose full-factorial and randomize order,

Step 3: For data analysis, prepare the signs for cross-terms

Step 2: Do the experiments and collect the results.

# A Toy Problem with 3 factors, 2 levels

| $x_1$ | $x_2$ | $x_3$ | $x_1x_2$ | $x_2x_3$ | $x_1x_3$ | $x_1x_2x_3$ | | $y$ |
|---|---|---|---|---|---|---|---|---|
| $-1$ | $-1$ | $-1$ | $+1$ | $+1$ | $+1$ | $-1$ | $(1)$ | 67 |
| $+1$ | $-1$ | $-1$ | $-1$ | $+1$ | $-1$ | $+1$ | $a$ | 79 |
| $-1$ | $+1$ | $-1$ | $-1$ | $-1$ | $+1$ | $+1$ | $b$ | 61 |
| $+1$ | $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $-1$ | $ab$ | 75 |
| $-1$ | $-1$ | $+1$ | $+1$ | $-1$ | $-1$ | $+1$ | $c$ | 59 |
| $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $+1$ | $-1$ | $ac$ | 90 |
| $-1$ | $+1$ | $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $bc$ | 52 |
| $+1$ | $+1$ | $+1$ | $+1$ | $+1$ | $+1$ | $+1$ | $abc$ | 87 |

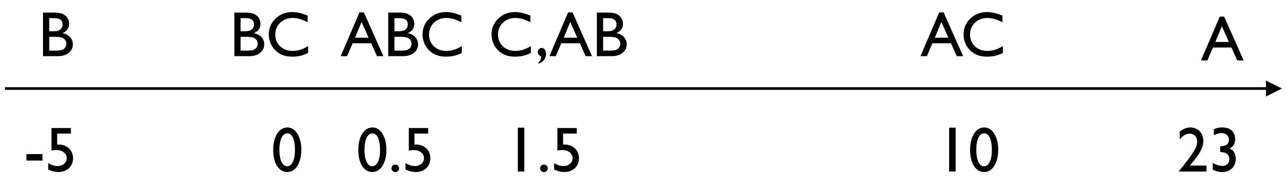
Effect of C=1.5
Effect of AC = 10.0
Effect of AB = 1.5
Effect of BC = 0.0
Effect of ABC=0.5

Effect of A=$[(a + ab + ac + abc)$ – $((1) + b + c + bc))]/4 = 23$

Effect of B=$[(b + ab + bc + abc)$ – $((1) + a + c + ac))]/4 = -5$

Effect of AB = $[((1) + ab + c + abc)$ – $(a + b + bc + ac))]/4 = 1.5$

# Importance of factors by Normal Q-Q plot: Step 1




Step 1: Order effect of A, B, C, AC, AB, BC, ABC = 23, -5, 1.5, 10, 1.5, 0.0, 0.5

# Importance of factors by Normal Plot: Step 2

Step 2 : Determine $F_i$

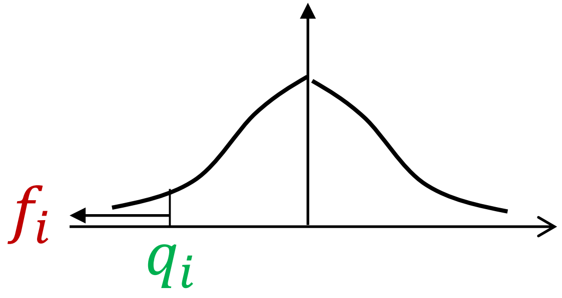
| $i$ | $f_i = (i - 0.5)/7$ | $Z_i$ |
|---|---|---|
| 1 | 0.071428 | |
| 2 | 0.214286 | |
| 3 | 0.357143 | |
| 4 | 0.500000 | |
| 5 | 0.642857 | |
| 6 | 0.785714 | |
| 7 | 0.928571 | |

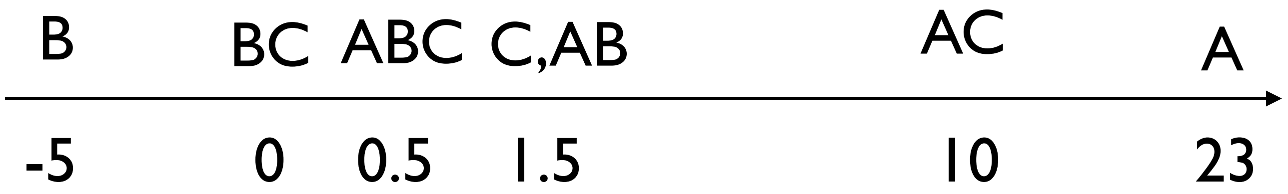
# Importance of factors by Normal plots: Step 3

Step 3 : Find $P(q_i) < f_i$

Normal table $q_i = 4.91\,[f_i^{0.14} - (1 - f_i)^{0.14})$

| i \| | | $f_i$ | $q_i$ |
|---|---|---|---|
| 1 \| | $-5.0$ | 0.071428 | $-1.47$ |
| 2 \| | $+0.0$ | 0.214286 | $-0.79$ |
| 3 \| | $+0.5$ | 0.357143 | $-0.36$ |
| 4 \| | $+1.5$ | 0.500000 | $+0.00$ |
| 5 \| | $+1.5$ | 0.642857 | $+0.36$ |
| 6 \| | $+10$ | 0.785714 | $+0.79$ |
| 7 \| | $+23$ | 0.928571 | $+1.47$ |

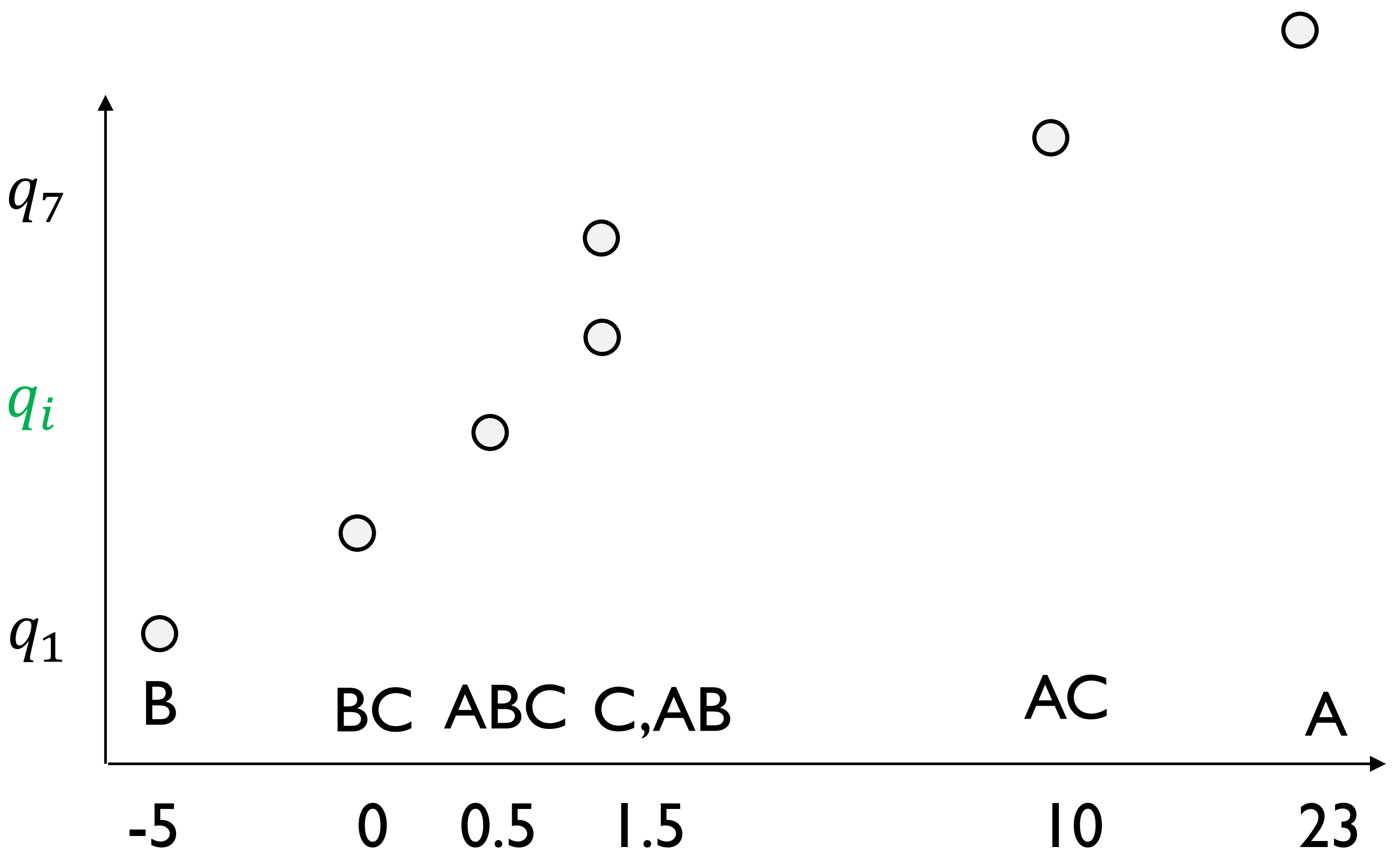


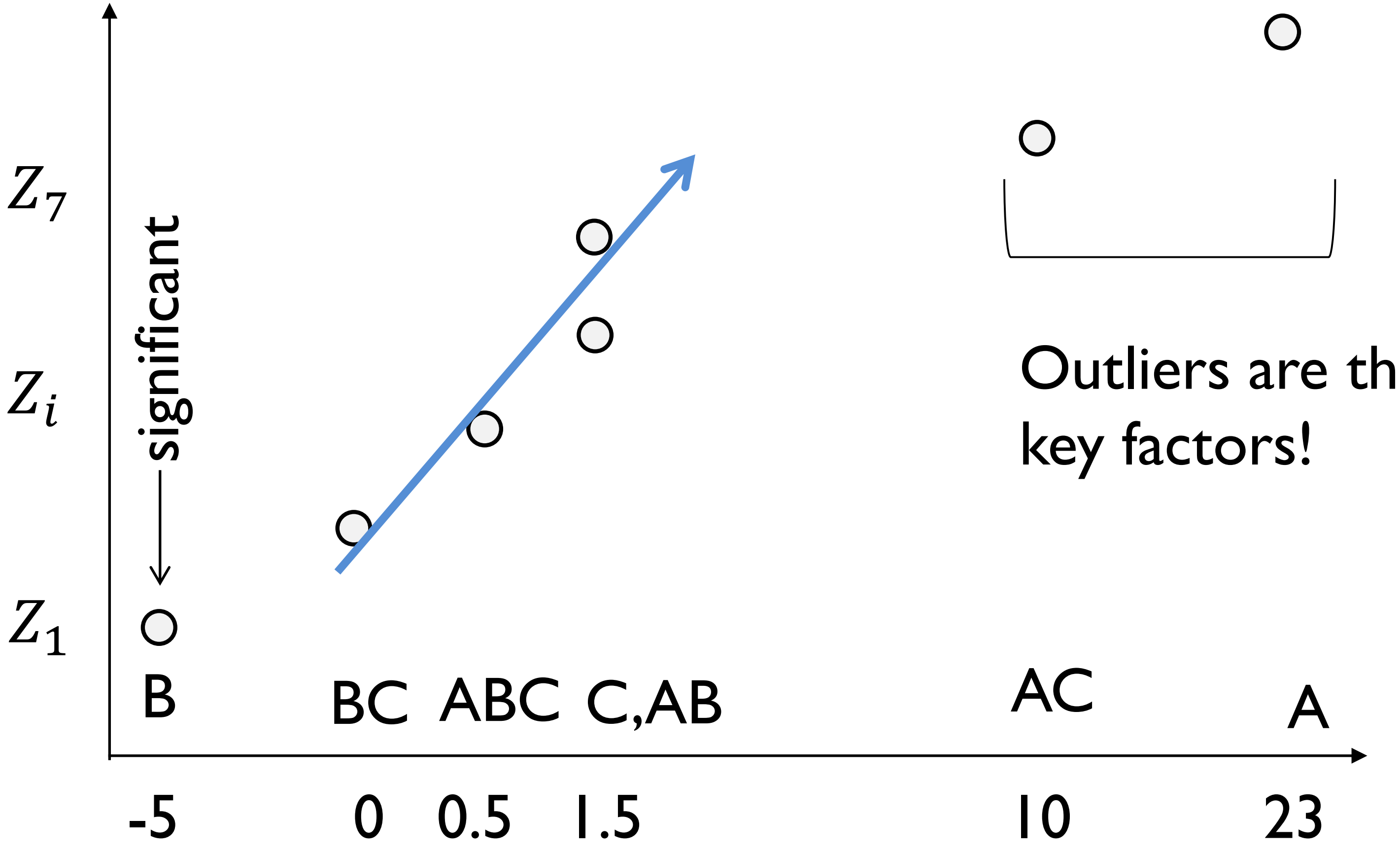
# Importance of factors by Normal plots: Step 4

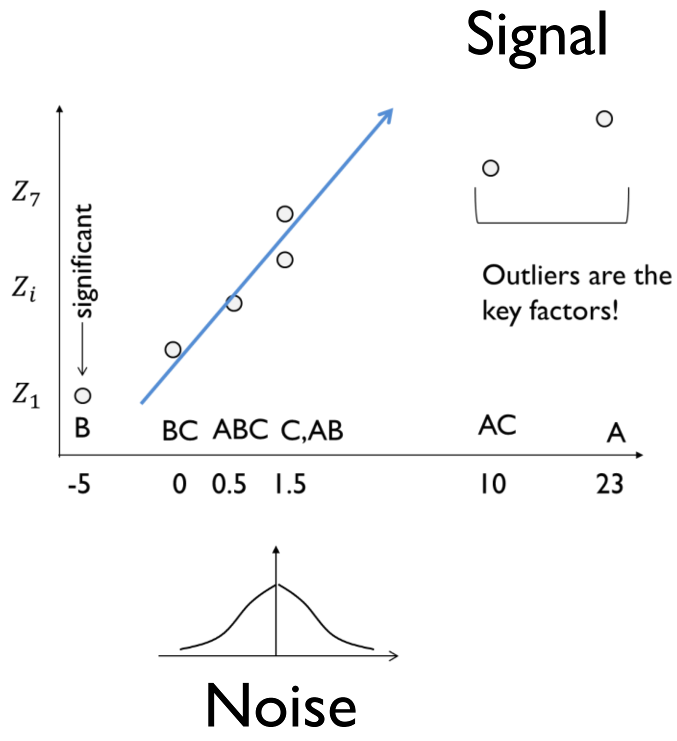
Step 3 : Find $P(q_i) < f_i$

| $i$ \| | | $F_i$ | $q_i$ |
|---|---|---|---|
| 1 \| | $-5.0$ | 0.071428 | $-1.468$ |
| 2 \| | $+0.0$ | 0.214286 | $-0.793$ |
| 3 \| | $+0.5$ | 0.357143 | $-0.366$ |
| 4 \| | $+1.5$ | 0.500000 | $+0.000$ |
| 5 \| | $+1.5$ | 0.642857 | $+0.366$ |
| 6 \| | $+10$ | 0.785714 | $+0.793$ |
| 7 \| | $+23$ | 0.928571 | $+1.468$ |



# Importance of factors by Normal plots: Step 5



# Outline

1. Context and background
2. Full factorial DOE explained by a toy model
3. Taguchi and fractional factorial experiments to reduce the number of experiments
4. Conclusions

# A toy problem with 3 factors, 2 levels

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{23} x_2 x_3 + \beta_{13} x_1 x_3 + \beta_{123} x_1 x_2 x_3 + \epsilon$$

| $x_1$ | $x_2$ | $x_3$ | $x_1x_2$ | $x_2x_3$ | $x_1x_3$ | $x_1x_2x_3$ | | $y$ |
|---|---|---|---|---|---|---|---|---|
| $-1$ | $-1$ | $-1$ | $+1$ | $+1$ | $+1$ | $-1$ | $(1)$ | 67 |
| $+1$ | $-1$ | $-1$ | $-1$ | $+1$ | $-1$ | $+1$ | $a$ | 79 |
| $-1$ | $+1$ | $-1$ | $-1$ | $-1$ | $+1$ | $+1$ | $b$ | 61 |
| $+1$ | $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $-1$ | $ab$ | 75 |
| $-1$ | $-1$ | $+1$ | $+1$ | $-1$ | $-1$ | $+1$ | $c$ | 59 |
| $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $+1$ | $-1$ | $ac$ | 90 |
| $-1$ | $+1$ | $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $bc$ | 52 |
| $+1$ | $+1$ | $+1$ | $+1$ | $+1$ | $+1$ | $+1$ | $abc$ | 87 |

Step 1: Choose full-factorial and randomize order,

Step 3: For data analysis, prepare the signs for cross-terms

Step 2: Do the experiments and collect the results.

# Two approaches to reduce the number of experiments

Determine the number of variables and experiments you have time to do.

Fractional Factorial Approach: 4, 8, 16, 32, 64

Taguchi Orthogonal Array: 4, 8, 12, 20, 24, 28

# Full factorial: 3 factors, 2 levels

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{23} x_2 x_3 + \beta_{13} x_1 x_3 + \beta_{123} x_1 x_2 x_3 + \epsilon$$

| $x_1$ | $x_2$ | $x_3$ | $x_1x_2$ | $x_2x_3$ | $x_1x_3$ | $x_1x_2x_3$ | | $y$ |
|---|---|---|---|---|---|---|---|---|
| $-1$ | $-1$ | $-1$ | $+1$ | $+1$ | $+1$ | $-1$ | $(1)$ | 67 |
| $+1$ | $-1$ | $-1$ | $-1$ | $+1$ | $-1$ | $+1$ | $a$ | 79 |
| $-1$ | $+1$ | $-1$ | $-1$ | $-1$ | $+1$ | $+1$ | $b$ | 61 |
| $+1$ | $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $-1$ | $ab$ | 75 |
| $-1$ | $-1$ | $+1$ | $+1$ | $-1$ | $-1$ | $+1$ | $c$ | 59 |
| $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $+1$ | $-1$ | $ac$ | 90 |
| $-1$ | $+1$ | $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $bc$ | 52 |
| $+1$ | $+1$ | $+1$ | $+1$ | $+1$ | $+1$ | $+1$ | $abc$ | 87 |

Effect of A=$[(ab + ac) - ((1) + bc))]/2 = ?$

Effect of B=$[(ab + bc) - ((1) + ac))]/2 = ?$

# Half-factorial experiments

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{23} x_2 x_3 + \beta_{13} x_1 x_3 + \beta_{123} x_1 x_2 x_3 + \epsilon$$

| $x_1$ | $x_2$ | $x_3$ | $x_1x_2$ | $x_1x_2x_3$ | | $y$ |
|---|---|---|---|---|---|---|
| $-1$ | $-1$ | $-1$ | $+1$ | $-1$ | $(1)$ | 67 |
| $+1$ | $-1$ | $-1$ | $-1$ | $+1$ | $a$ | 79 |
| $-1$ | $+1$ | $-1$ | $-1$ | $+1$ | $b$ | 61 |
| $+1$ | $+1$ | $-1$ | $+1$ | $-1$ | $ab$ | 75 |
| $-1$ | $-1$ | $+1$ | $+1$ | $+1$ | $c$ | 59 |
| $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $ac$ | 90 |
| $-1$ | $+1$ | $+1$ | $-1$ | $-1$ | $bc$ | 52 |
| $+1$ | $+1$ | $+1$ | $+1$ | $+1$ | $abc$ | 87 |

Effect of A=$[(ab + ac)$ – $((1) + bc))]/2 =?$

Effect of B=$[(ab + bc)$ – $((1) + ac))]/2 =?$

The number of experiments must be greater than the number of variables.

The technique allows to determine the correlation experiments.

Fractional factorial tables are balanced. We have the same need for randomization.

# Taguchi approach: 3 factors, 2 levels (P-B Matrix)

Generating vector

| $x_1$ | $x_2$ | $x_3$ |
|---|---|---|
| $+1$ | $-1$ | $+1$ |
| $+1$ | $+1$ | $-1$ |
| $-1$ | $\mp 1$ | $+1$ |
| $-1$ | $-1$ | $-1$ |

Plackett-Burman design

Use the col 1 generating vector

Set bottom row -1

Roll the vector from the first column to fill other columns.

# Another way: 7 factors,  2 levels (P-B Matrix)

Vector

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ |
|---|---|---|---|---|---|---|
| $+1$ | $-1$ | $-1$ | $+1$ | $-1$ | $+1$ | $+1$ |
| $+1$ | $+1$ | $-1$ | $-1$ | $+1$ | $-1$ | $+1$ |
| $+1$ | $+1$ | $+1$ | $-1$ | $-1$ | $+1$ | $-1$ |
| $-1$ | $+1$ | $+1$ | $+1$ | $-1$ | $-1$ | $+1$ |
| $+1$ | $-1$ | $+1$ | $+1$ | $+1$ | $-1$ | $-1$ |
| $-1$ | $+1$ | $-1$ | $+1$ | $+1$ | $+1$ | $-1$ |
| $-1$ | $-1$ | $+1$ | $-1$ | $+1$ | $+1$ | $+1$ |
| $-1$ | $-1$ | $-1$ | $-1$ | $-1$ | $-1$ | $-1$ |

Plackett-Burman design

Use the col 1 generating vector

Set bottom row -1

Roll the vector from the first column to fill other columns.

# Another way: 11 factors, 2 levels (P-B Matrix)

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | ……. | $x_{11}$ |
|---|---|---|---|---|---|---|---|---|
| $\boxed{+1}$ | $\boxed{-1}$ | $+1$ | $-1$ | $-1$ | | | | |
| $\boxed{+1}$ | $\boxed{+1}$ | $-1$ | | | | | | |
| $-1$ | $\boxed{+1}$ | $+1$ | | | | | | |
| $+1$ | $-1$ | $+1$ | | | | | | |
| $+1$ | $+1$ | $+1$ | | | | | | |
| $+1$ | $+1$ | $+1$ | | | | | | |
| $-1$ | $+1$ | $+1$ | | | | | | |
| $-1$ | $-1$ | $+1$ | | | | | | |
| $-1$ | $-1$ | $-1$ | | | | | | |
| $+1$ | $-1$ | $-1$ | | | | | | |
| $\boxed{-1}$ | $+1$ | $-1$ | $-1$ | $-1$ | | | | |
| $-1$ | $-1$ | $-1$ | $-1$ | $-1$ | $-1$ | $-1$ | $-1$ | $-1$ |

Plackett-Burman design

Use the col 1 generating vector

Set bottom row -1

Roll the vector from the first column to fill other columns.

# Taguchi table: Continued

$L_4(2^3)$

| Run | Columns | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| 1 | 1 | 1 | 1 |
| 2 | 1 | 2 | 2 |
| 3 | 2 | 1 | 2 |
| 4 | 2 | 2 | 1 |

$L_8(2^7)$

| Run | Columns | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 |

$L_{12}(2^{11})$

| Run | Columns | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 |
| 4 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 2 |
| 5 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 |
| 6 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 |
| 7 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 |
| 8 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 |
| 9 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 |
| 10 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 2 |
| 11 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 2 | 2 |
| 12 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 |

$L_N(S^M)$

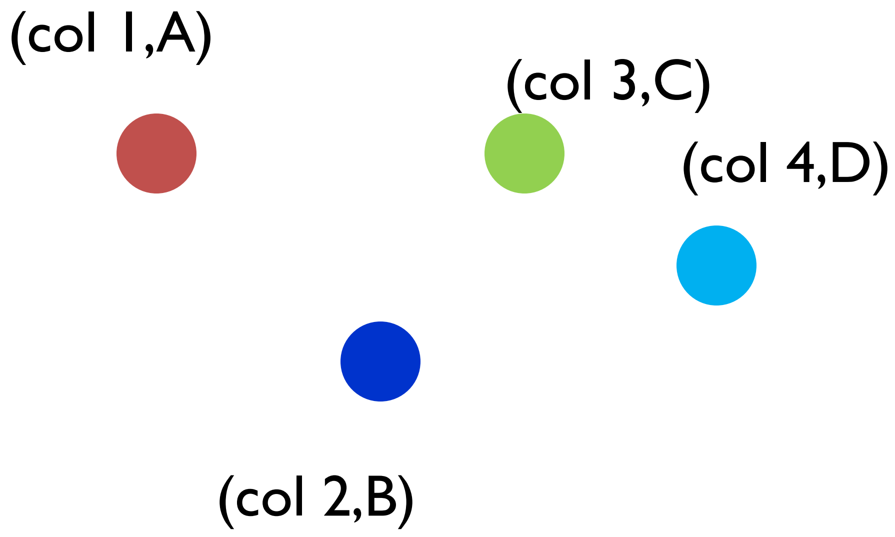
M … Factors

S … Levels

N … experiment > DOF = 1 + M(S-1)

# A 4-Factor, 2-Level experiment

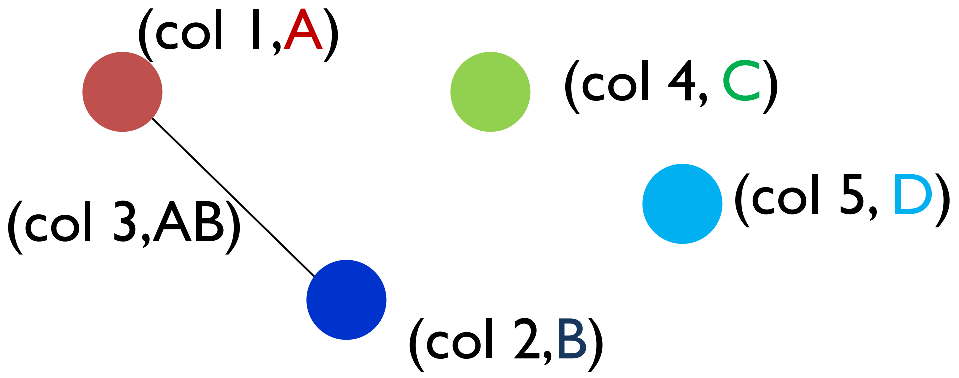
| Run | Columns | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 |




Only (AB) pair correlation found, no other correlation ….

# Main effect and interactions

| | A | B | AB | C | D | | |
|---|---|---|---|---|---|---|---|
| Run | Columns | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 |




(AB) is a dummy column, without it the C and D would have different arrangements …

Still need L8 array (4-7), other two-level arrays L4 (1-3) and L12(8-11)

# Conclusions

1. Design of experiment is a powerful technique universally used in industry and in large scale field trials.

2. Taguchi/Fisher methods replace the older one-factor-at-a-time experiments with experiments based on orthogonal arrays; In this approach, only the effect of main factors remain; others are cancelled.

3. Understanding and analyzing correlation is important in design of experiments. Unless the correlation is well understood and incorporated through dummy variables, the analysis may lead to faulty conclusions.

# Primer on Analysis of Experimental Data and Design of Experiments

## *Lecture 9. DOE Analysis by ANOVA*

Muhammad A. Alam
alam@purdue.edu

# Outline

1. Introduction to Analysis of Variance (Anova)

2. Single factor Analysis of Variance

3. Two factor Anova

4. Generalized Anova

5. Conclusions

# Recall: A Toy Problem with 3 factors, 2 levels

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{23} x_2 x_3 + \beta_{13} x_1 x_3 + \beta_{123} x_1 x_2 x_3 + \epsilon$$

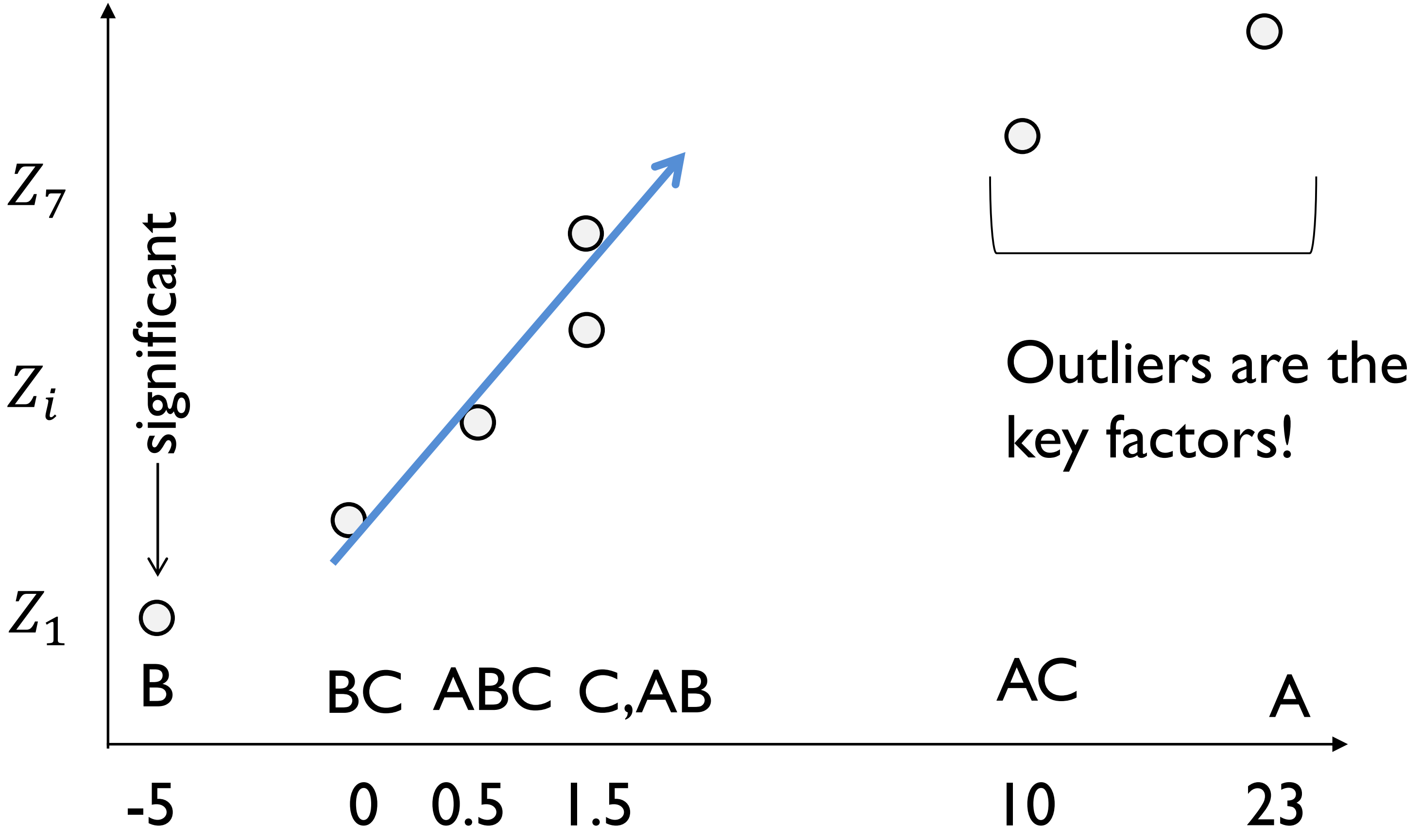
| $x_1$ | $x_2$ | $x_3$ | $x_1 x_2$ | $x_2 x_3$ | $x_1 x_3$ | $x_1 x_2 x_3$ | | $y$ |
|---|---|---|---|---|---|---|---|---|
| $-1$ | $-1$ | $-1$ | $+1$ | $+1$ | $+1$ | $-1$ | $(1)$ | 67 |
| $+1$ | $-1$ | $-1$ | $-1$ | $+1$ | $-1$ | $+1$ | $a$ | 79 |
| $-1$ | $+1$ | $-1$ | $-1$ | $-1$ | $+1$ | $+1$ | $b$ | 61 |
| $+1$ | $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $-1$ | $ab$ | 75 |
| $-1$ | $-1$ | $+1$ | $+1$ | $-1$ | $-1$ | $+1$ | $c$ | 59 |
| $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $+1$ | $-1$ | $ac$ | 90 |
| $-1$ | $+1$ | $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $bc$ | 52 |
| $+1$ | $+1$ | $+1$ | $+1$ | $+1$ | $+1$ | $+1$ | $abc$ | 87 |

Step 1: Choose full-factorial and randomize order,

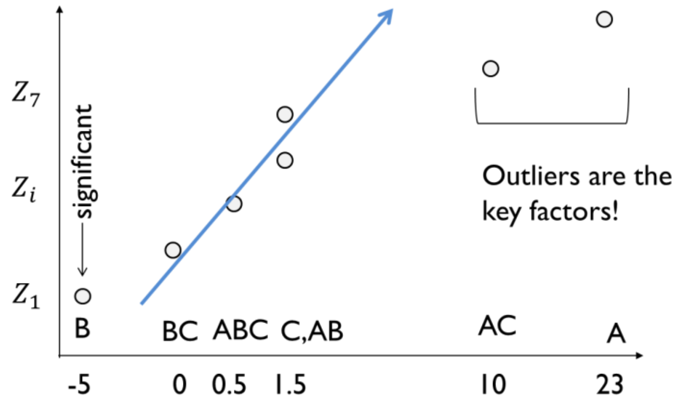
Step 3: For data analysis, prepare the signs for cross-terms

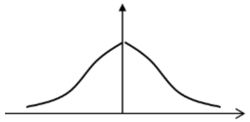
Step 2: Do the experiments and collect the results.

# Recall: Importance of factors by Normal plots




Signal



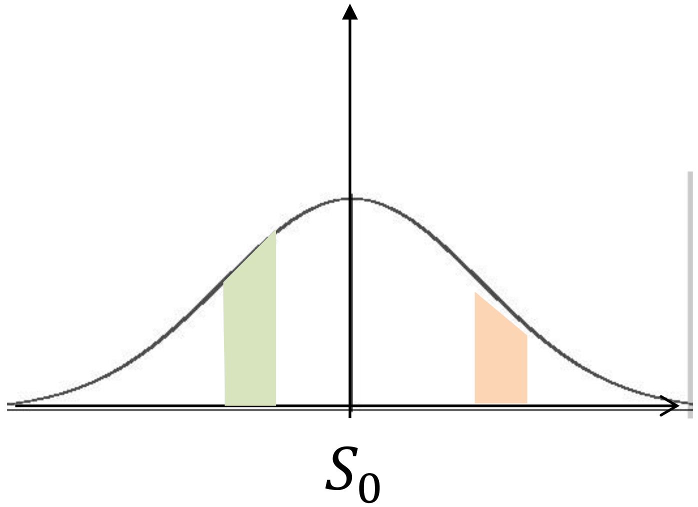
Noise

# Single factor ANOVA: Treatment

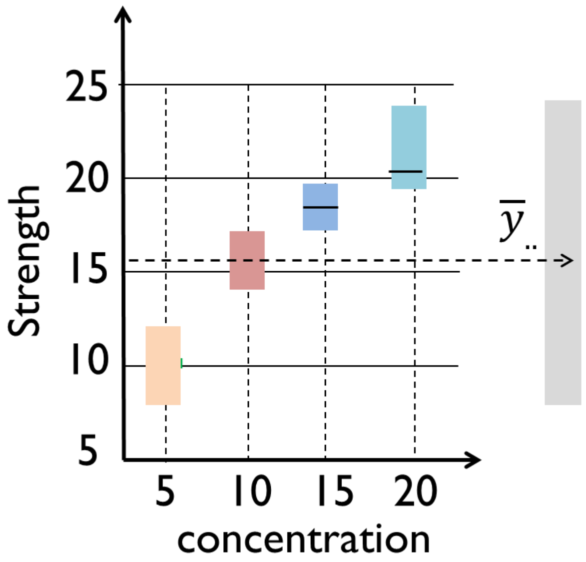
replicates

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 5 | 7 | 8 | 15 | 11 | 9 | 10 |
| 10 | 12 | 17 | 13 | 18 | 19 | 15 |
| 15 | 14 | 18 | 19 | 17 | 16 | 18 |
| 20 | 19 | 25 | 22 | 23 | 18 | 20 |

Treatments (levels)




In essence, no effect

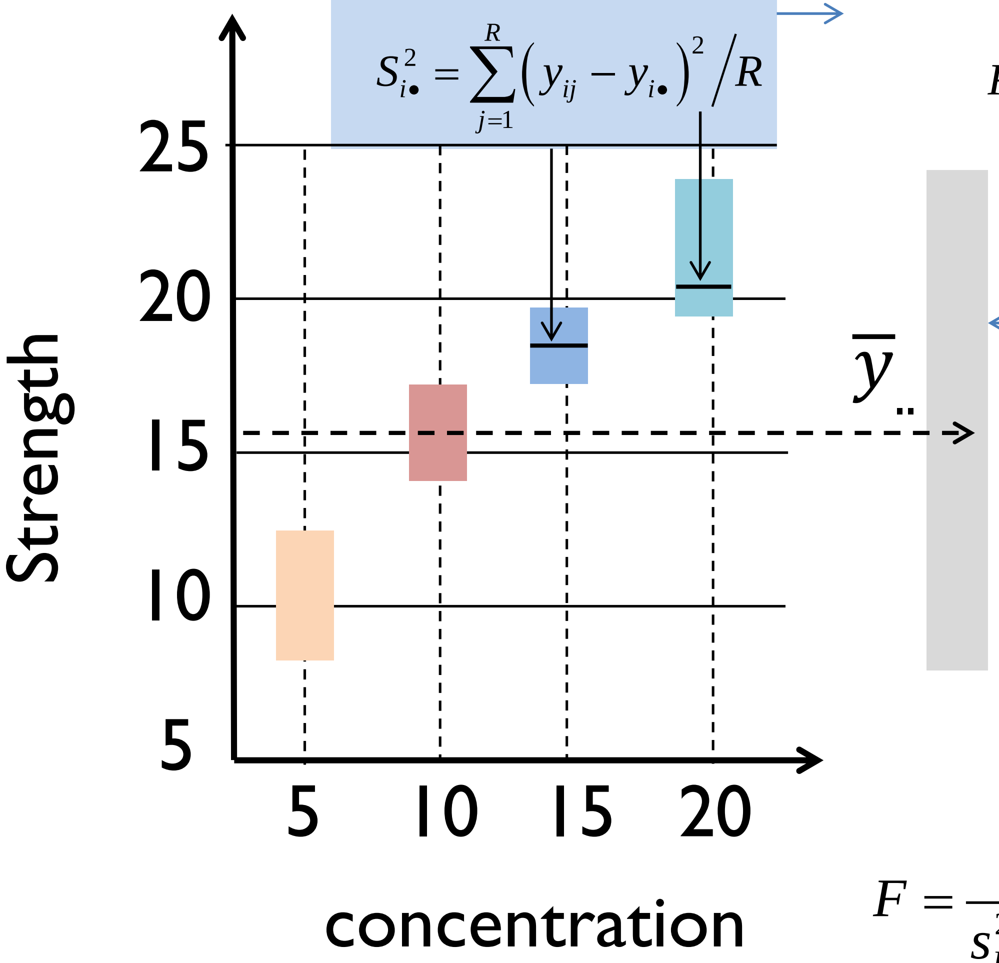


# Single factor Anova: Treatment Analysis

R=6 (replicate),
L=4 (level)

$$\bar{y}_{i\bullet} = \sum_{j=1}^{R} y_{ij} \big/ R$$

$$S_{i\bullet}^2 = \sum_{j=1}^{R} \left(y_{ij} - y_{i\bullet}\right)^2 \big/ R$$




$$SS_T^2 = \sum_{j=1}^{L} \sum_{i=1}^{R} (y_{ij} - \bar{y}_{\bullet\bullet})^2$$

$$RL \times s_E^2 = SS_T^2 - SS_{Treatment}^2$$

$$\bar{y}_{\bullet\bullet} = \sum_{i=1}^{L} \sum_{j=1}^{R} y_{ij} \big/ RL$$

$$SS_{Treatment} = L\sigma^2 = R\sum_{i=1}^{L} \left(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}\right)^2$$

Recall: $\chi^2$, KS, etc.

$$F = \frac{\sigma^2 \times L/(L-1)}{s_E^2 \times RL/L(R-1)} : F_{crit}(L-1, L(R-1))$$

# Single factor ANOVA (continued)

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 5 | 7 | 8 | 15 | 11 | 9 | 10 |
| 10 | 12 | 17 | 13 | 18 | 19 | 15 |
| 15 | 14 | 18 | 19 | 17 | 16 | 18 |
| 20 | 19 | 25 | 22 | 23 | 18 | 20 |

1. Treatment number, $a = 4;, \mathrm{dof}_a = 3;$ Sample number: $n = 6$
2. Global sample number: $a \times n = 24, \ \mathrm{dof}_n = 23,$ AVG (global)=15.96
3. Total "Sum of square" $SS_T = \sum_{24}(\mathrm{data} - AVG)^2 = 512.96$
4. Treatment sum: $SS_{treatment} = n \times \sum_4(\mathrm{treat.avg} - AVG)^2$=382.
5. $SS_{error} = SS_T - SS_{treatment}$=130.62
6. $ME_{treatment} = SS_{treatment}/\mathrm{dof}_a, \ ME_{Error} = SS_{error}/(\mathrm{dof}_n - \mathrm{dof}_a)$
7. Finally, $F = (ME_{treatment})/(ME_{error}) = \frac{382/3}{130.62/20} = 19.6$
8. Compare: $f(0.01, \mathrm{dof}_a, \mathrm{dof}_n)$, or $P\left(F_{3,20} > 19.6\right) = 3.59 \times 10^{-6}$

# Single factor ANOVA: Wood Treatment

replicates

treatments

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 5 | 7 | 8 | 15 | 11 | 9 | 10 |
| 10 | 12 | 17 | 13 | 18 | 19 | 15 |
| 15 | 14 | 18 | 19 | 17 | 16 | 18 |
| 20 | 19 | 25 | 22 | 23 | 18 | 20 |

$\sum(\text{data} - \text{AVG})^2 = 512$

| t.avg | $(\text{t.avg} - \text{AVG})^2$ |
|---|---|
| 10.00 | 35.50174 |
| 15.67 | 0.085069 |
| 17.00 | 1.085069 |
| 21.17 | 27.12674 |
| 15.96 | 63.79861 |

$6 \times 63.8 = 382.8$

| Variation | SS | df | MS | F | P-value | F crit |
|---|---|---|---|---|---|---|
| Between Groups | 382.7917 | 3 | 127.60 | 19.605 | 3.59E-06 | 4.94 |
| Within Groups | 130.1667 | 20 | 6.51 | | | |
| Total | 512.9583 | 23 | | | | |

1 3 4 2 5 6

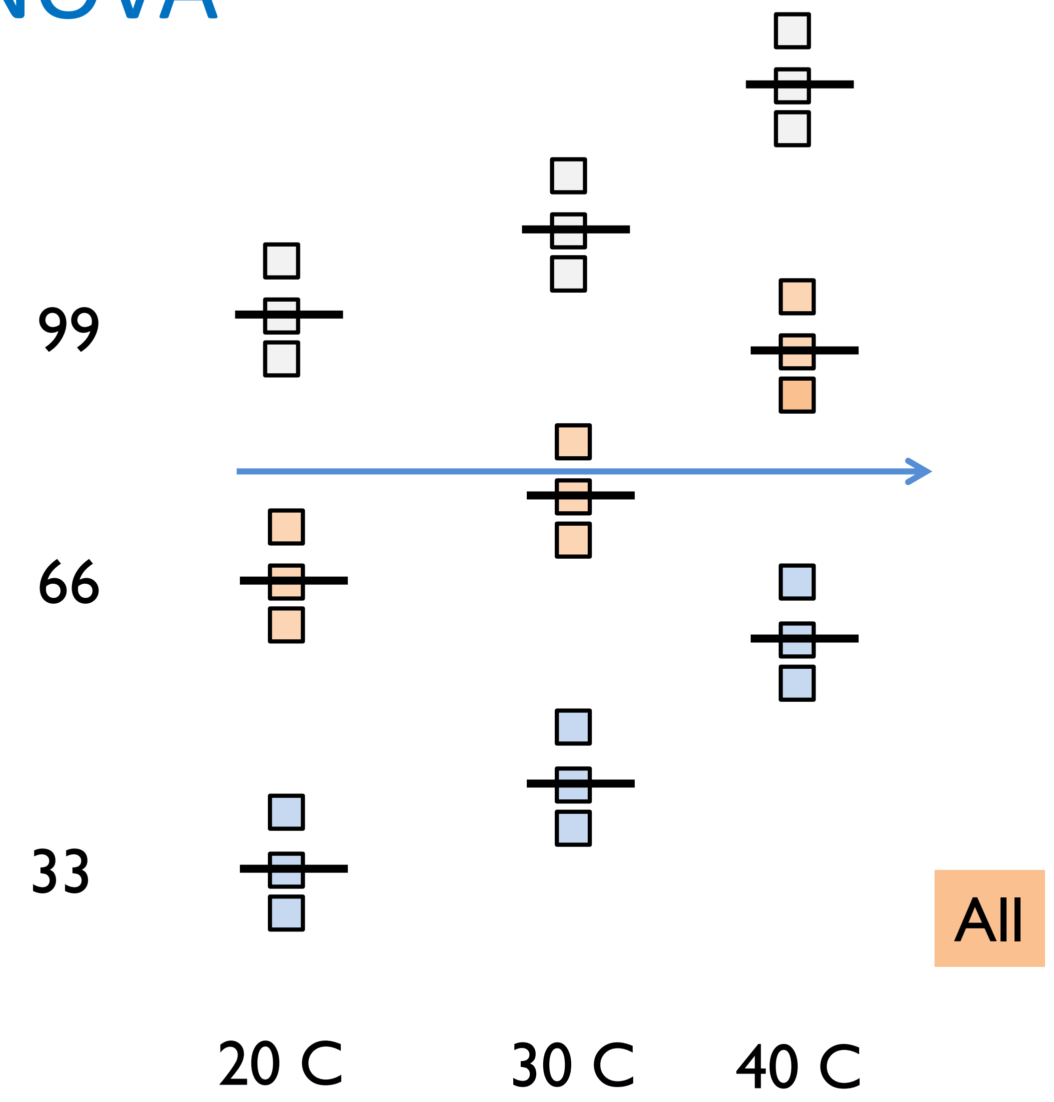
# Outline

1. Introduction to Analysis of Variance (Anova)
2. Single factor Analysis of Variance
3. **Generalized Anova**
4. Conclusions

# Two factor ANOVA

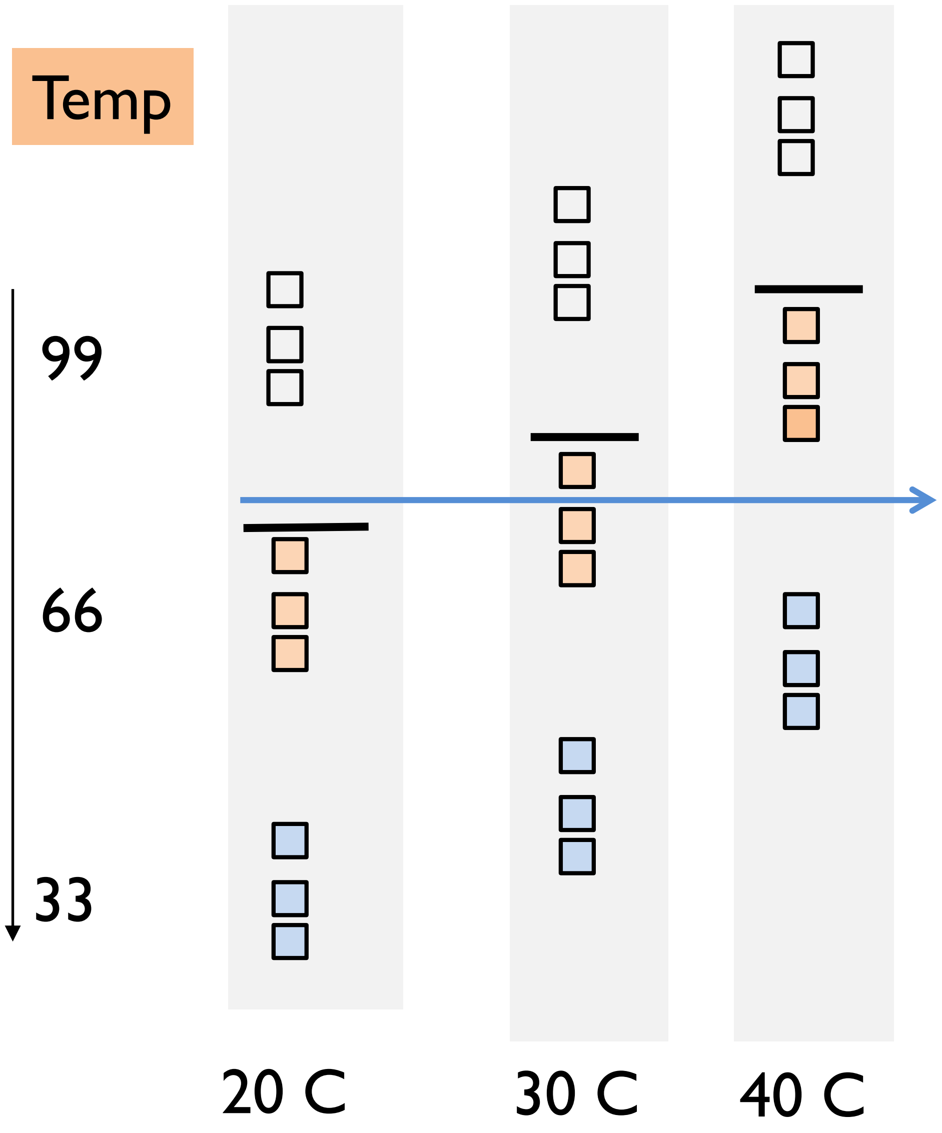
Full factorial:
2 factor, 3 level,
3 replicate experiment

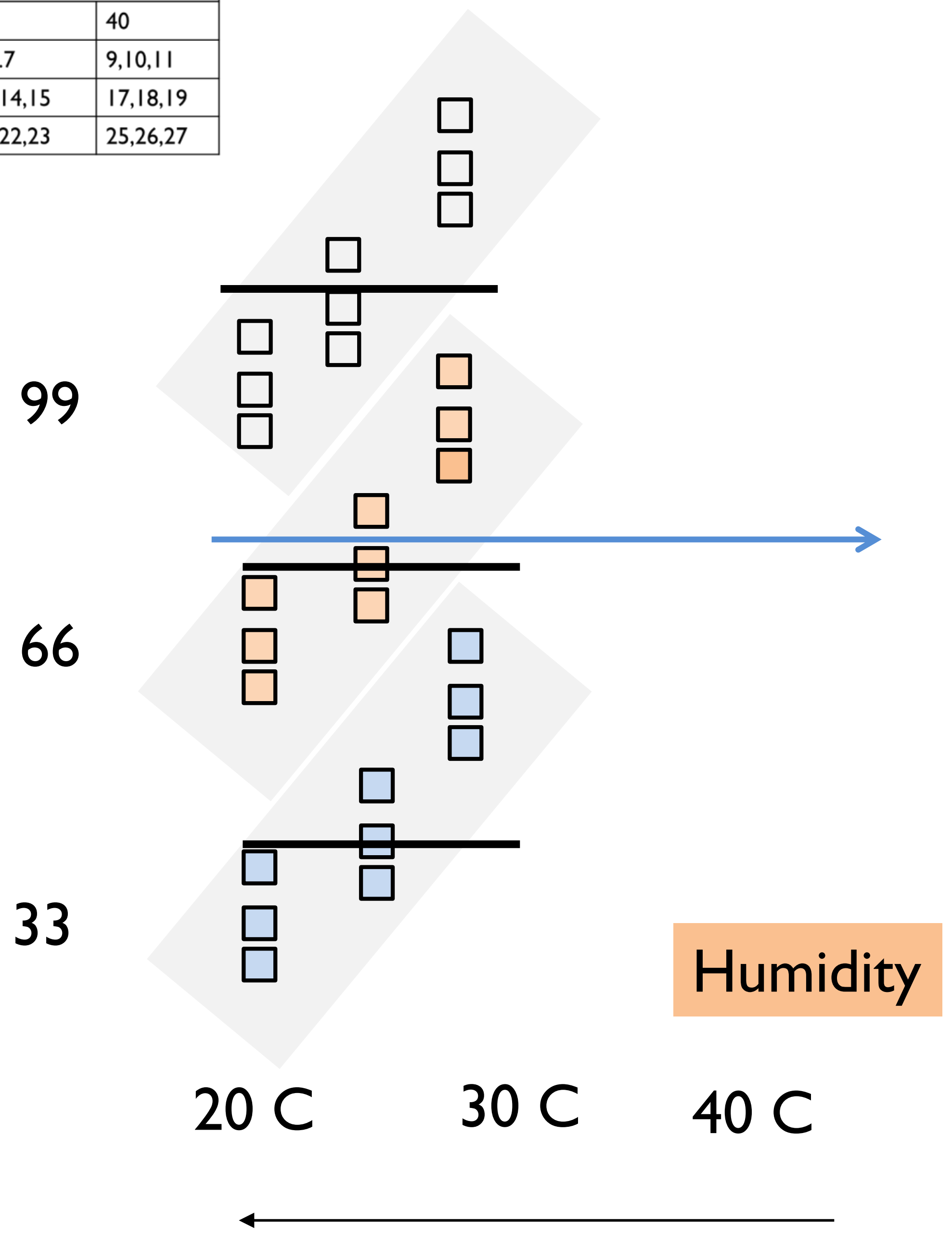
| | Temperature (C) | | |
|---|---|---|---|
| Humidity (%) | 20 | 30 | 40 |
| 33 | 1,2,3 | 5.6.7 | 9,10,11 |
| 66 | 9,10,11 | 13,14,15 | 17,18,19 |
| 99 | 17,18,19 | 21,22,23 | 25,26,27 |



# Two factor ANOVA

| | Temperature (C) | | |
|---|---|---|---|
| Humidity (%) | 20 | 30 | 40 |
| 33 | 1,2,3 | 5.6.7 | 9,10,11 |
| 66 | 9,10,11 | 13,14,15 | 17,18,19 |
| 99 | 17,18,19 | 21,22,23 | 25,26,27 |



# Two factor ANOVA

Full factorial:
2 factor, 3 level,
3 replicate experiment

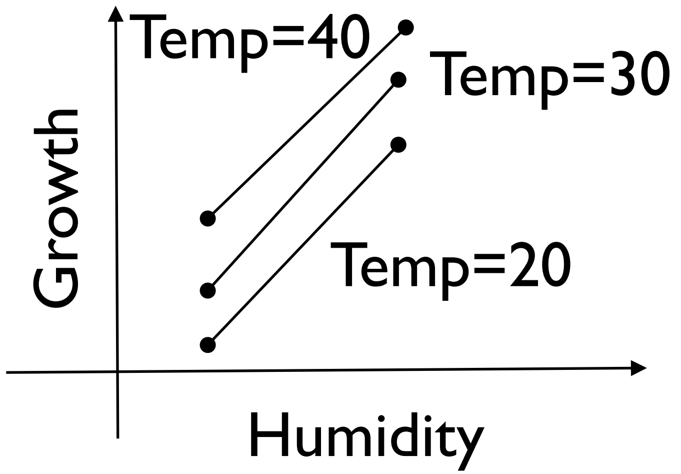
| | Temperature (C) | | |
|---|---|---|---|
| Humidity (%) | 20 | 30 | 40 |
| 33 | 1,2,3 | 5.6.7 | 9,10,11 |
| 66 | 9,10,11 | 13,14,15 | 17,18,19 |
| 99 | 17,18,19 | 21,22,23 | 25,26,27 |

Excel/Minitab Analysis

| | Sum of Squares | dof | Mean-square | F Ratio | Significance |
|---|---|---|---|---|---|
| Temp | 312.66 | 3-1=2 | 312.66/2=156.33 | 156.33/1.0=156.33 | 0.000 (significant) |
| Humidity | 1200.66 | 3-1=2 | 1200.66/2=600.33 | 600.66/1=600.33 | 0.000 (Significant) |
| Temp*Humidity | 1.33 | 2x2=4 | 1.33/4=0.33 | 0.33/1.0=0.33 | 0.853 (insignificant) |
| Error | 18.00 | 27-2-2-4=19 | 18.00/18=1 | | |

# Two factor ANOVA (Excel/Minitab Analysis)

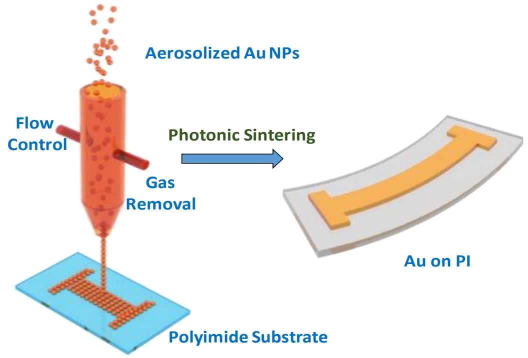
| | Sum of Squares | dof | Mean-square | F Ratio | p-value |
|---|---|---|---|---|---|
| Temp | 312.66 | 3-1=2 | 312.66/2=156.33 | 156.33/1.0=156.33 | 0.000 (significant) |
| Humidity | 1200.66 | 3-1=2 | 1200.66/2=600.33 | 600.66/1=600.33 | 0.000 (Significant) |
| Temp*Humidity | 1.33 | 2x2=4 | 1.33/4=0.33 | 0.33/1.0=0.33 | 0.853 (insignificant) |
| Error | 18.00 | 27-2-2-4=19 | 18.00/18=1 | | |

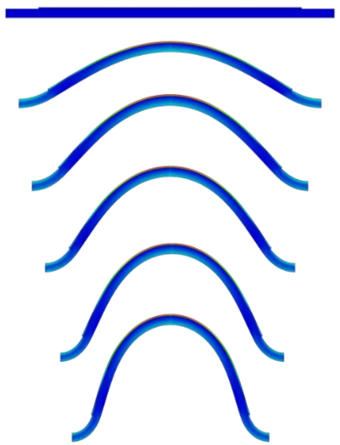



No mutual interaction

Ref. Statistics Explained,
S. Mckillup, Cambridge Press

# Multiple Factor ANOVA



| Thickness (µm) | BR | BF | Resistance Change (Percent) | | | | |
|---|---|---|---|---|---|---|---|
| | | | 50 | 100 | 150 | 200 | 250 |
| 10 | Half | Slow | 46.5 | 52.61 | 66.3 | 91.58 | 149.6 |
| 10 | Half | Fast | 67.7 | 69.4 | 97.34 | 124.4 | 139.3 |
| 10 | Full | Slow | 92.975 | 174.45 | 232.45 | 252.64 | 275.3 |
| 10 | Full | Fast | 85.6 | 135.47 | 162.59 | 157.53 | 208.6 |
| 25 | Half | Slow | 22.4 | 23.57 | 24.87 | 29.14 | 28.9 |
| 25 | Half | Fast | 25.2 | 35.22 | 22.14 | 24.5 | 29.65 |
| 25 | Full | Slow | 24.7 | 32.89 | 54.68 | 78.23 | 95.63 |
| 25 | Full | Fast | 45.23 | 51.29 | 65.26 | 61.4 | 78.95 |

Field view

Thickness=2 (i.e. 10, 15), BR=2 (i.e. slow, fast)
BF=2 (Half and full), Cycles=5 (i.e. 50,100,150, 200, 250)

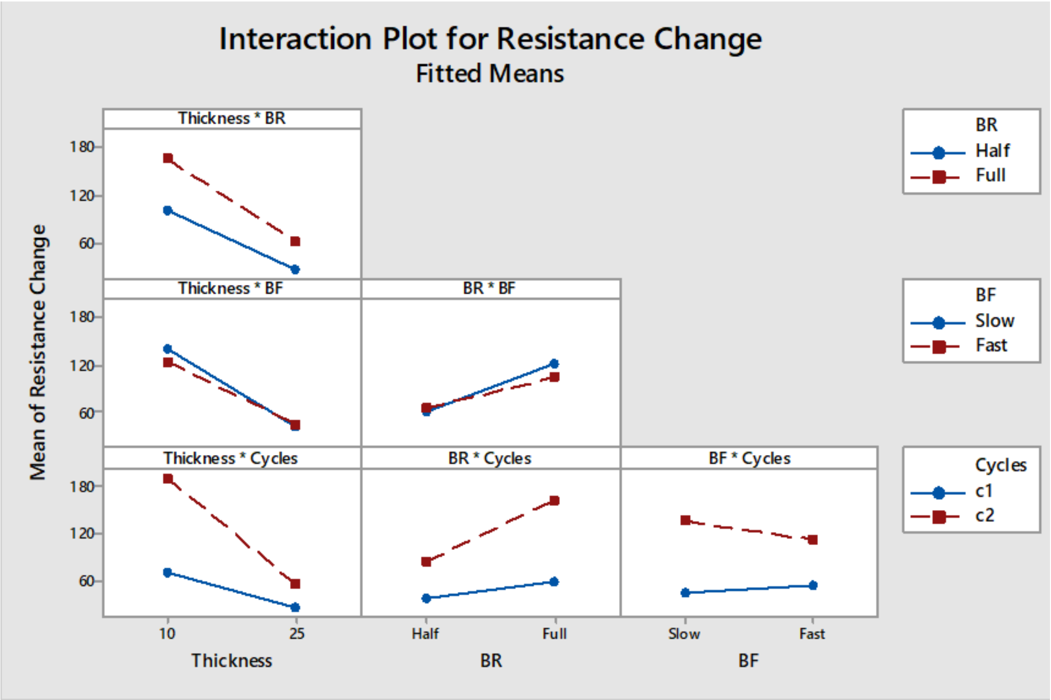
5 factors, (2,2,2,5) levels, single replicate DOE experiments

40 experiments (Need to use statistical package, e.g. Minitab)

# Multiple Factor ANOVA (Continued)

Analysis of Variance

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Model | 14 | 79330.2 | 5666.4 | 1685.12 | 0.019 |
| Linear | 4 | 64198.5 | 16049.6 | 4772.93 | 0.011 |
| Thickness | 1 | 31944.0 | 31944.0 | 9499.67 | 0.007 |
| BR | 1 | 9887.1 | 9887.1 | 2940.27 | 0.012 |
| BF | 1 | 194.4 | 194.4 | 57.82 | 0.083 |
| Cycles | 1 | 22173.1 | 22173.1 | 6593.95 | 0.008 |
| 2-Way Interactions | 6 | 14177.8 | 2363.0 | 702.71 | 0.029 |
| Thickness*BR | 1 | 915.3 | 915.3 | 272.19 | 0.039 |
| Thickness*BF | 1 | 311.3 | 311.3 | 92.58 | 0.066 |
| Thickness*Cycles | 1 | 8300.3 | 8300.3 | 2468.40 | 0.013 |
| BR*BF | 1 | 448.1 | 448.1 | 133.26 | 0.055 |
| BR*Cycles | 1 | 3145.1 | 3145.1 | 935.31 | 0.021 |
| BF*Cycles | 1 | 1057.6 | 1057.6 | 314.52 | 0.036 |
| 3-Way Interactions | 4 | 953.8 | 238.5 | 70.91 | 0.089 |
| Thickness*BR*BF | 1 | 454.5 | 454.5 | 135.16 | 0.055 |
| Thickness*BR*Cycles | 1 | 85.2 | 85.2 | 25.34 | 0.125 |
| Thickness*BF*Cycles | 1 | 166.2 | 166.2 | 49.42 | 0.090 |
| BR*BF*Cycles | 1 | 247.9 | 247.9 | 73.74 | 0.074 |
| Error | 1 | 3.4 | 3.4 | | |
| Total | 15 | 79333.5 | | | |




Check for p-values < 0.05

One way: Thickness, BR, and Cycles (BF does not matter on its own)

Two-way: Thickness/BR, Thickness/cycles, BR/cycles, BF/cycles

Three-way: Thickness, BR, BF (may be, close to 0.05, more experiments)

# Conclusions

1. Design of experiment results are analyzed by ANOVA test to make sure that the effect of variables on the final result is statistically significant. Insignificant variables can be dropped to simplify analysis.

2. ANOVA generalizes hypothesis testing to continuous variables.

3. A positive test from Anova says one of the treatment is different from others – it does not say which one. With a positive result, one can do pair-wise comparison.

4. Simple Anova tests are easily done by calculator. More complicated Anova tests are best done by statistical packages, such as S or Minitab, etc.

5. With results of Anova at hand, the new design of experiments based on new Taguchi table must be performed.

# Review Questions

1. What does the word ANOVA stand for?  Who developed the technique?
2. How does ANOVA compare with standard hypothesis testing?
3. If an ANOVA test identifies correlation among the variables, how should one redo the Taguchi tables?
4. Can ANOVA analysis include discrete variables?
5. If there are 7 replicates and 5 treatments, how many samples are tested?
6. For 7 replicates and 5 treatments, what is the degree of freedom for the treatments? What about the samples?
7. An experiment involving single factor ANOVA can be analyzed by Excel. Is this correct?

# ReferenceS

The classical AVONA method is discussed in great detail in Chapter 13 and 14 of "Applied Statistics and Probability for Engineers, 3rd Edison, D.C. Montgomery and G. C. Runger, Wiley, 2003.

Hunter's lectures on AVONA is also very enjoybale

http://www.youtube.com/watch?v=k3n9iSB6Cns

http://www.youtube.com/watch?v=F05zZL3uyRo

A slightly different approach that also reduces the number of experiments greatly is based on the response surface approach. It uses Newton-like algorithm to find the peaks/valleys of the response surface, see R. H. Myers and D.C. Montgomery, "Response Surface Methodology", Wiley Interscience, 2002. This book discusses design of experiment in great detail.

For general reference see

Joan Fisher Box, "R. A. Fisher and the Design of Experiments, 1922-1926", *The American Statistician*, vol. 34, no. 1, pp. 1-7, Feb. 1980.

F. Yates, "Sir Ronald Fisher and the Design of Experiments", *Biometrics*, vol. 20, no. 2, In Memoriam: Ronald Aylmer Fisher, 1890-1962., pp. 307-321, (Jun. 1964.