Topic 11: Introduction to factorial design (I)

Montgomery: chapter 5

Prof. Lingling An
University of Arizona

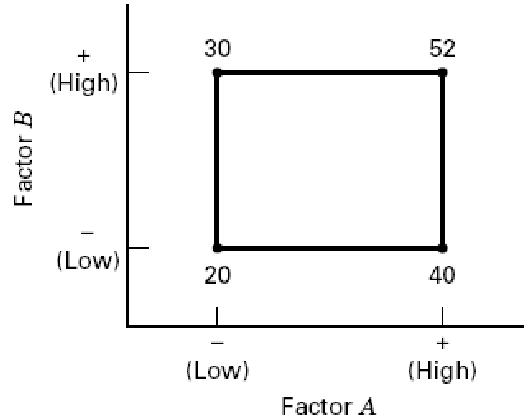
Outline

- Introduction factorial design
- General two-factor factorial design and its statistical model

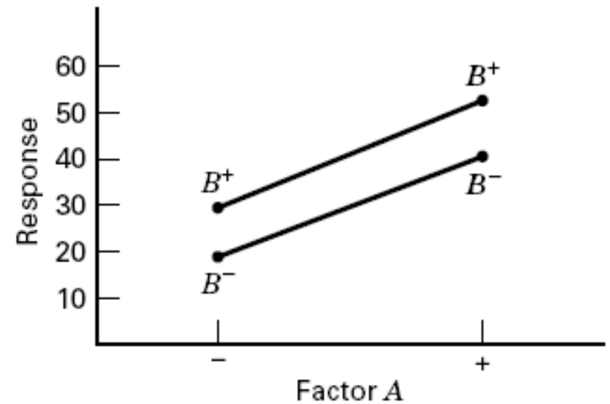
Introduction to Factorial Design of Engineering Experiments

- 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)

Some Basic Definitions



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



■ FIGURE 5.3 A factorial experiment without interaction

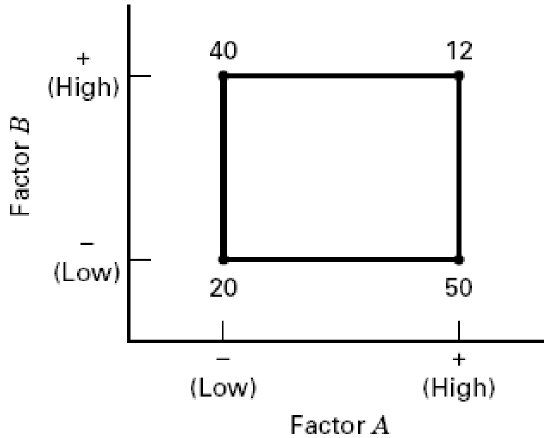
Definition of a factor effect: The change in the mean response when the factor is changed from low to high

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

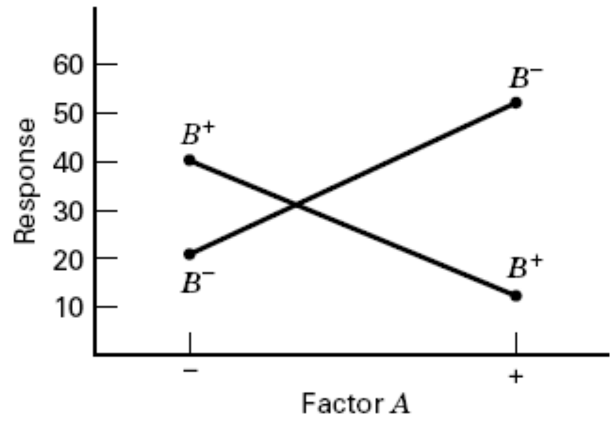
$$B = \bar{y}_{B^+} - \bar{y}_{B^-} = \frac{30+52}{2} - \frac{20+40}{2} = 11$$

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

The Case of Interaction



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



■ **FIGURE 5.4** A factorial experiment with interaction

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

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

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

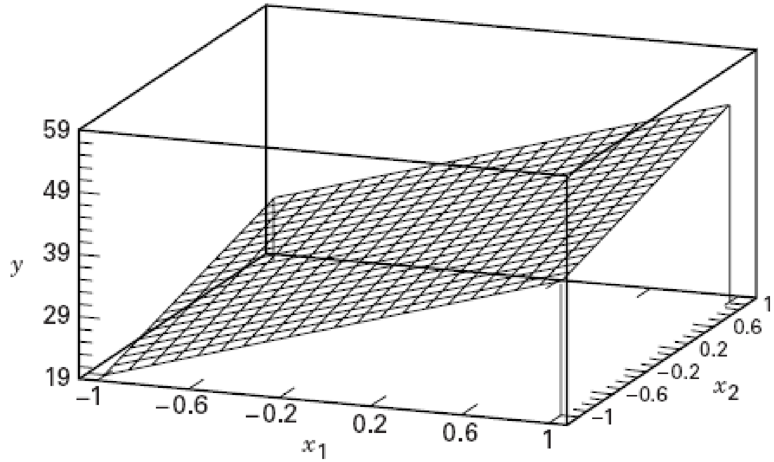
Another way for interaction:
Regression Model & Associated Response Surface

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

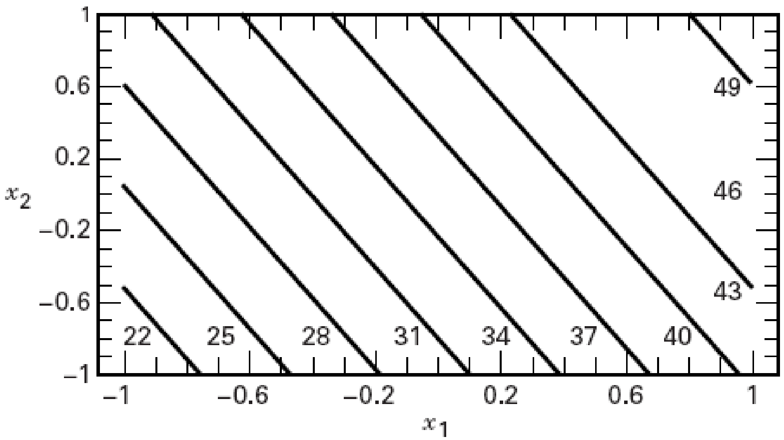
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

The least squares fit is

$$\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + 0.5x_1x_2 \cong 35.5 + 10.5x_1 + 5.5x_2$$



(*a*) The response surface



(*b*) The contour plot

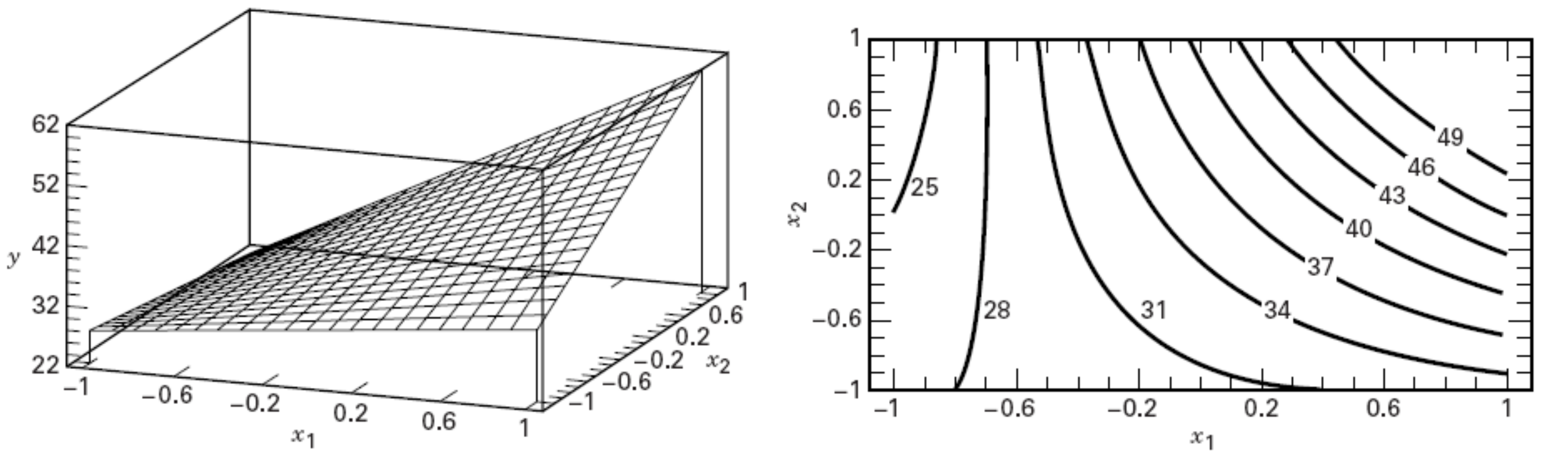
■ **FIGURE 5.5** Response surface and contour plot for the model $\hat{y} = 35.5 + 10.5x_1 + 5.5x_2$

The Effect of Interaction on the Response Surface

Suppose that we add an interaction term to the model:

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

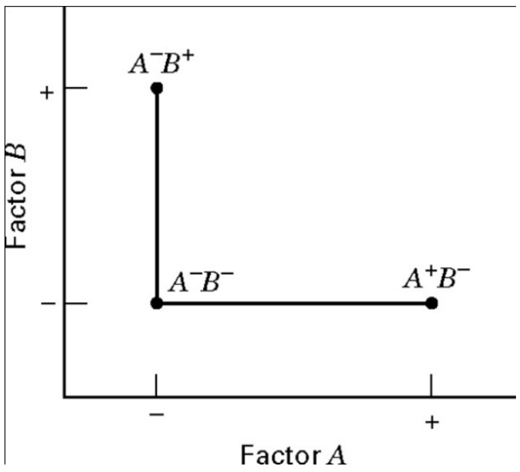
Interaction is actually a form of **curvature**



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

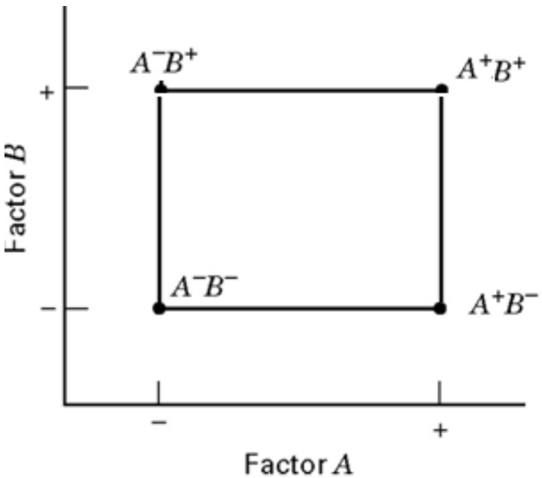
■ FIGURE 5.6 Response surface and contour plot for the model $\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + 8x_1x_2$

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: Battery life experiment

- 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:

■ **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

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

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

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

- Statistical model is

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

μ - grand mean

τ_i - ith level effect of factor A (ignores B) (main effects of A)

β_j - jth level effect of factor B (ignores A) (main effects of B)

$(\tau\beta)_{ij}$ - interaction effect of combination ij (Explain variation not explained by main effects)

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

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

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

Estimates

- Rewrite observation as:

$$y_{ijk} = \overline{y}_{...} + (\overline{y}_{i..} - \overline{y}_{...}) + (\overline{y}_{.j.} - \overline{y}_{...}) + (\overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...}) + (y_{ijk} - \overline{y}_{ij.})$$

- result in estimates

$$\widehat{\mu} = \overline{y}_{...}$$

$$\widehat{\tau}_i = \overline{y}_{i..} - \overline{y}_{...}$$

$$\widehat{\beta}_j = \overline{y}_{.j.} - \overline{y}_{...}$$

$$\widehat{(\tau\beta)}_{ij} = \overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...}$$

- predicted value at level combination ij is

$$\widehat{y}_{ijk} = \overline{y}_{ij.}$$

- Residuals are

$$\hat{\epsilon}_{ijk} = y_{ijk} - \overline{y}_{ij.}$$

Battery Example

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

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

$$\hat{\tau}_1 = -22.3612, \hat{\tau}_2 = 2.8055, \hat{\tau}_3 = 19.555$$

$$\hat{\beta}_1 = 39.3055, \hat{\beta}_2 = 2.0555, \hat{\beta}_3 = -41.3611$$

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

	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

Partitioning the Sum of Squares

- Based on

$$y_{ijk} = \overline{y}_{...} + (\overline{y}_{i..} - \overline{y}_{...}) + (\overline{y}_{.j.} - \overline{y}_{...}) + (\overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...}) + (y_{ijk} - \overline{y}_{ij.})$$

- Calculate $\mathrm{SS}_{\mathrm{T}} = \sum (y_{ijk} - \overline{y}_{...})^2$

- Right hand side simplifies to

$$
\begin{array}{rcl}
\mathrm{SS_A}: & bn\sum_i (\overline{y}_{i..} - \overline{y}_{...})^2 + & df = a-1 \\
\mathrm{SS_B}: & an\sum_j (\overline{y}_{.j.} - \overline{y}_{...})^2 + & df = b-1 \\
\mathrm{SS_{AB}}: & n\sum_i \sum_j (\overline{y}_{ij.} - \overline{y}_{i.} - \overline{y}_{.j} + \overline{y}_{..})^2 + & df = (a-1)(b-1) \\
\mathrm{SS_E}: & \sum_i \sum_j \sum_k (y_{ijk} - \overline{y}_{ij.})^2 & df = ab(n-1)
\end{array}
$$

- $\mathrm{SS_T}$=$\mathrm{SS_A}$ + $\mathrm{SS_B}$ + $\mathrm{SS_{AB}}$ + $\mathrm{SS_E}$
- Using $\mathrm{SS/df}$ leads to $\mathrm{MS_A}$,$\mathrm{MS_B}$, $\mathrm{MS_{AB}}$ and $\mathrm{MS_E}$.

Testing Hypotheses

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

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

3 Interaction effects of AB:

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

- Use F-statistics for testing the hypotheses above:

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

- E($\mathrm{MS_E}$)=σ^2

  E($\mathrm{MS_A}$) = $\sigma^2 + bn \sum \tau_i^2/(a-1)$

  E($\mathrm{MS_B}$) = $\sigma^2 + an \sum \beta_j^2/(b-1)$

  E($\mathrm{MS_{AB}}$) = $\sigma^2 + n \sum (\tau\beta)_{ij}^2/(a-1)(b-1)$

Analysis of Variance Table

■ **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	$a-1$	$MS_A = \frac{SS_A}{a-1}$	$F_0 = \frac{MS_A}{MS_E}$
B treatments	SS_B	$b-1$	$MS_B = \frac{SS_B}{b-1}$	$F_0 = \frac{MS_B}{MS_E}$
Interaction	SS_{AB}	$(a-1)(b-1)$	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$F_0 = \frac{MS_{AB}}{MS_E}$
Error	SS_E	$ab(n-1)$	$MS_E = \frac{SS_E}{ab(n-1)}$	
Total	SS_T	$abn-1$		

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

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

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

- $df_E > 0$ only if $n > 1$. When $n = 1$, no SS_E 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 σ^2 and it can be used as MS_E.
  - Caution: if the assumption is wrong, then error and interaction are confounded and testing results can go wrong.

SAS code

```
data battery;
infile 'D:\TEACHING\T_STAT571B\datasets_files
\BATTERY-LIFE.csv' delimiter=',';
input temp mat life;
;

proc print data=battery;
run;

proc glm;
class mat temp;
model life=mat temp mat*temp;
output out=batnew r=res p=pred;
means mat temp mat*temp;
run;
```

Obs	temp	mat	life
1	15	1	130
2	15	1	74
3	15	1	155
4	15	1	180
5	70	1	34
6	70	1	80
7	70	1	40
8	70	1	75
9	125	1	20
10	125	1	82
11	125	1	70
12	125	1	58
..	…	…	..
35	125	3	104
36	125	3	60

Dependent Variable: life

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	59416.22222	7427.02778	11.00	<.0001
Error	27	18230.75000	675.21296		
Corrected Total	35	77646.97222			

R-Square	Coeff Var	Root MSE	life Mean
0.765210	24.62372	25.98486	105.5278

Source	DF	Type III SS	Mean Square	F Value	Pr > F
mat	2	39118.72222	19559.36111	28.97	<.0001
temp	2	10683.72222	5341.86111	7.91	0.0020
mat*temp	4	9613.77778	2403.44444	3.56	0.0186

Means at different **material** levels:

Level of mat	N	life	
		Mean	Std Dev
1	12	83.166667	48.5888751
2	12	108.333333	49.4723676
3	12	125.083333	35.7655455

Means at different **temperature** levels:

Level of temp	N	life	
		Mean	Std Dev
15	12	144.833333	31.6940870
70	12	107.583333	42.8834750
125	12	64.166667	25.6721757

Means at different combinations of **temperature** and **material** levels:

Level of mat	Level of temp	N	life	
			Mean	Std Dev
1	15	4	134.750000	45.3532432
1	70	4	57.250000	23.5990819
1	125	4	57.500000	26.8514432
2	15	4	155.750000	25.6173769
2	70	4	119.750000	12.6589889
2	125	4	49.500000	19.2613603
3	15	4	144.000000	25.9743463
3	70	4	145.750000	22.5444006
3	125	4	85.500000	19.2786583

Checking Assumptions

1 Errors are normally distributed

Histogram or QQplot of residuals

2 Constant variance

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

3 If n=1, no interaction.

Tukey's Test of Nonadditivity Assume $(\tau\beta)_{ij} = \gamma\tau_i\beta_j$. $H_0 : \gamma = 0$.

$$\mathrm{SS}_N = \frac{\left[\sum\sum y_{ij}y_{i.}y_{.j} - y_{..}(\mathrm{SS}_A + \mathrm{SS}_B + y_{..}^2/ab)\right]^2}{ab\mathrm{SS}_A\mathrm{SS}_B}$$

$$F_0 = \frac{\mathrm{SS}_N/1}{(\mathrm{SS}_E - \mathrm{SS}_N)/((a-1)(b-1)-1)} \sim F_{1,(a-1)(b-a)-1}$$

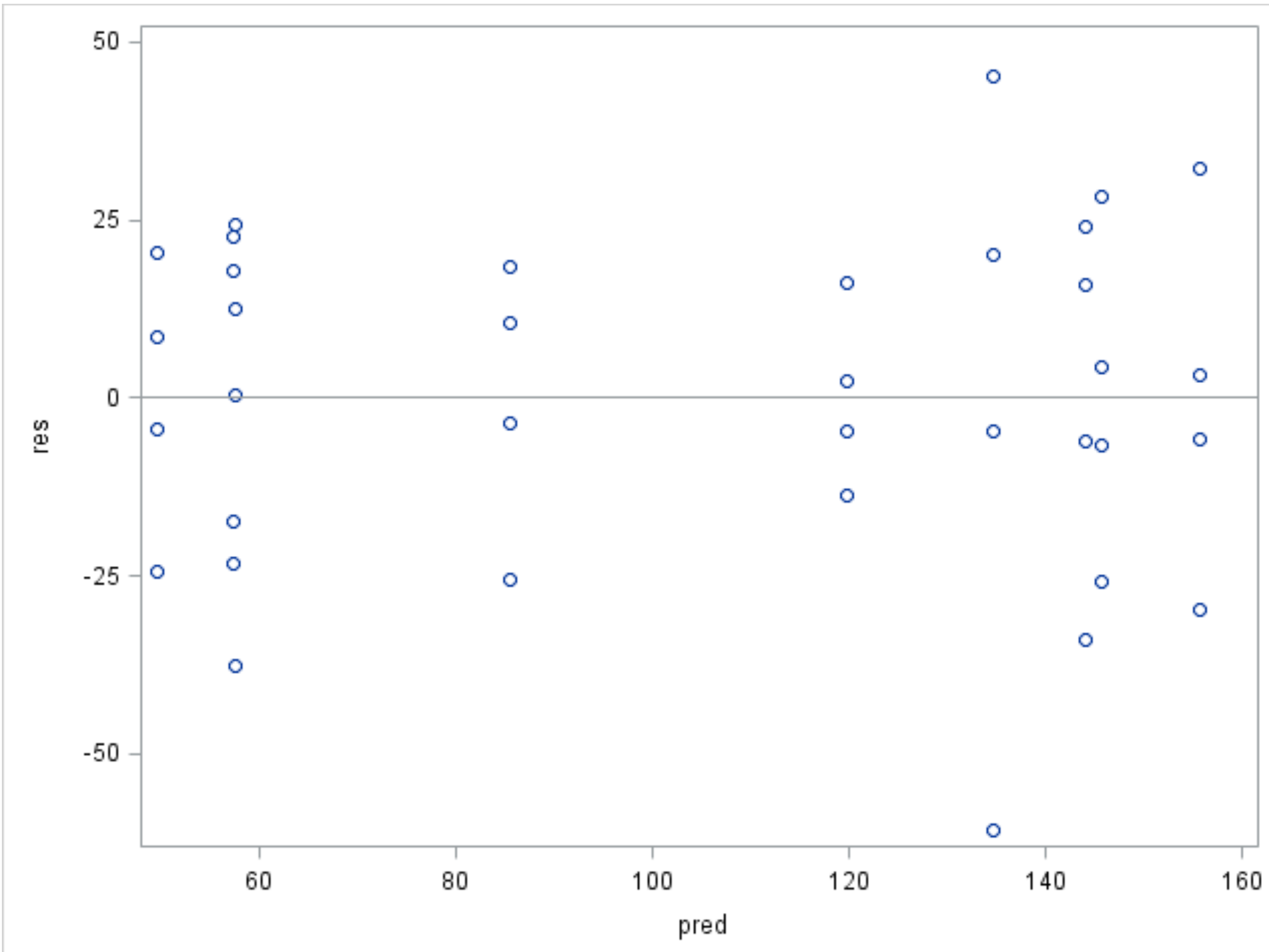
– the convenient procedure used for RCBD can be employed.

```
PROC univariate data=batnew normal;
var res;
qqplot res /normal(MU=0 SIGMA=EST COLOR=RED L=1);
run;

proc sgplot;
scatter x=pred y=res;
refline 0;
run;
```

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.976057	Pr < W	0.6117
Kolmogorov-Smirnov	D	0.10593	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.054415	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.340337	Pr > A-Sq	>0.2500

Residual plot: residual vs. predicted values



Last slide

- Read Sections: 5.1-5.3.4