Design of Engineering Experiments Part 5 – Introduction to Factorials

- Text reference, Chapter 5
- **General principles** of factorial experiments
- The **two-factor factorial** with fixed effects
- The **ANOVA** for factorials
- Extensions to more than two factors
- **Quantitative** and **qualitative** factors – response curves and surfaces

A Motivated Example

- A supermarket manager wants to know the **simultaneous effect** of selling price (Factor A, 55 cents, 60 cents) and Promotional campaigns (Factor B, radio advertising, newspaper adverting) on the sales of a product.

  How to design an experiment to address this question?

Treatment	Description
($a \times b = 4$)	
1	55 price, radio ad.
2	60 price, radio ad.
3	55 price, newspaper ad.
4	60 price, newspaper ad

Twenty communities throughout the United States, ***of approximately equal size and similar socioeconomic characteristics,*** were selected and the treatments were assigned to them ***at random***, such that each treatment was given to five experimental units.

Some Basic Definitions

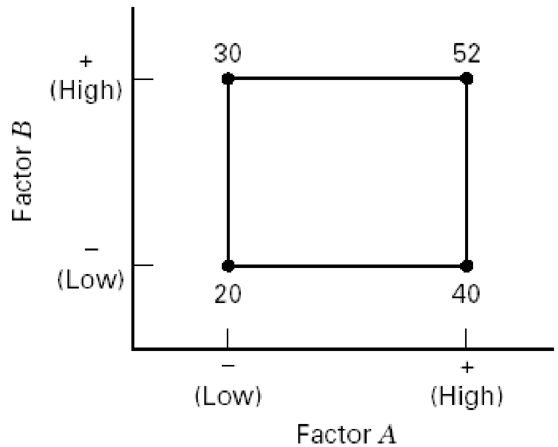


Figure 5-1 A two-factor factorial experiment, with the response (y) shown at the corners.

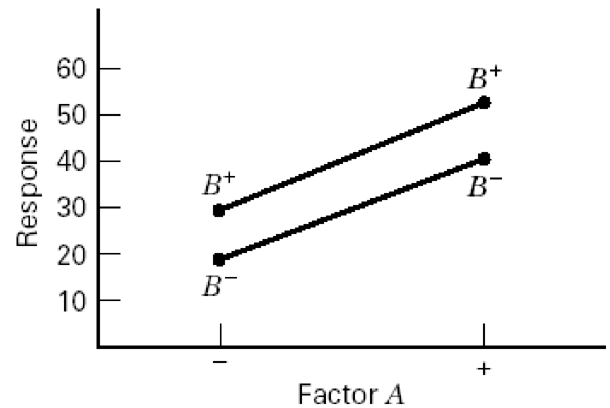


Figure 5-3 A factorial experiment without interaction.

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:

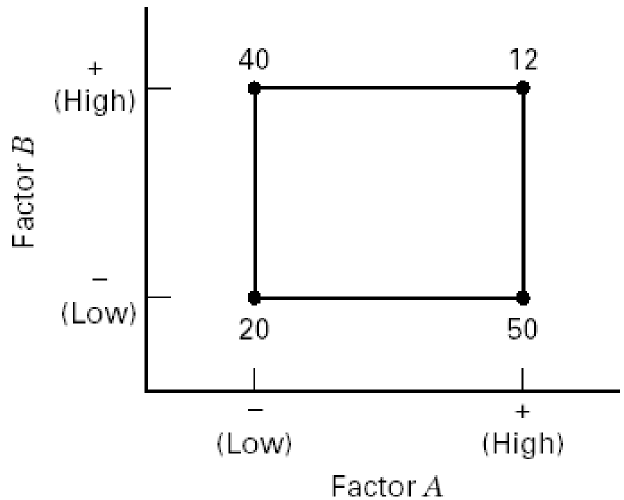


Figure 5-2 A two-factor factorial experiment with interaction.

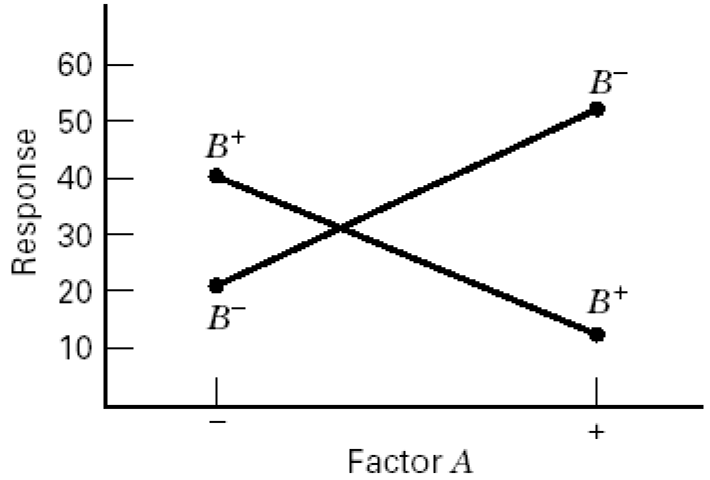


Figure 5-4 A factorial experiment with interaction.

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

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

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

Regression Model & The Associated Response Surface

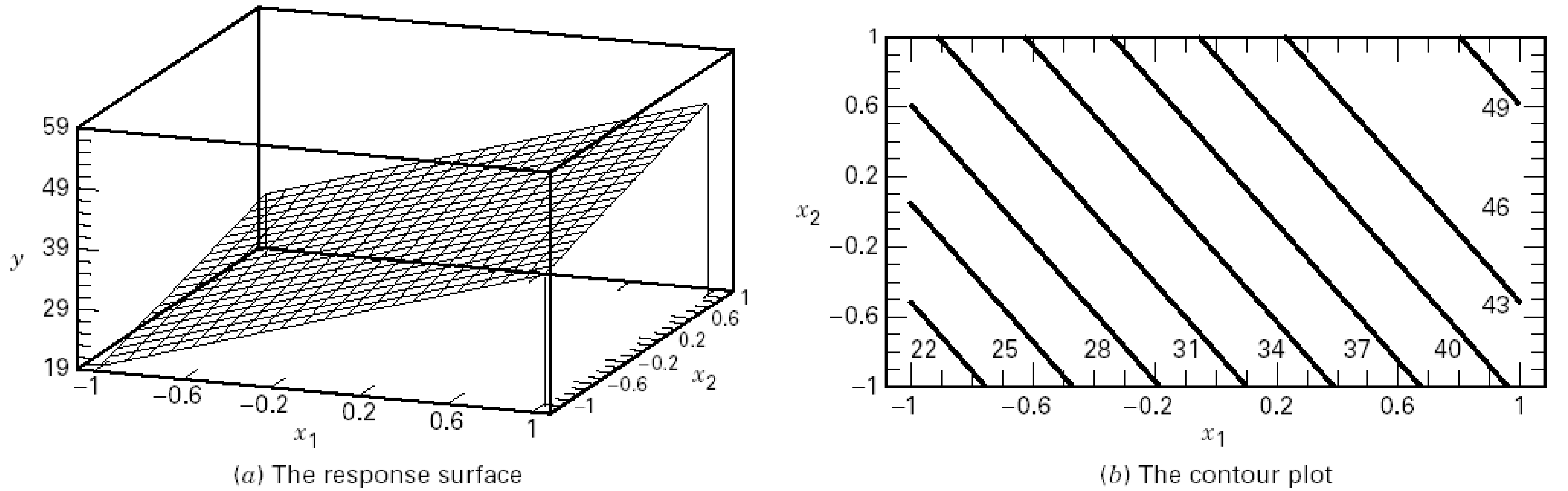


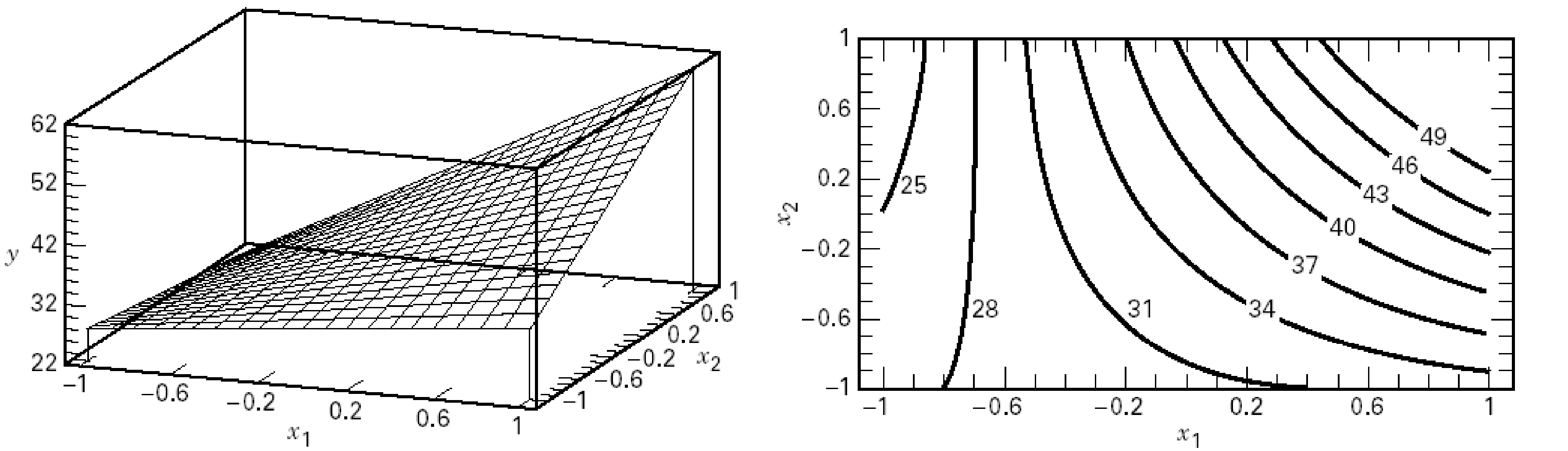
Figure 5-5 Response surface and contour plot for the model $\hat{y} = 35.5 + 10.5x_1 + 5.5x_2$.

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

The least squares fit is

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

The Effect of Interaction on the Response Surface



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

Figure 5-6 Response surface and contour plot for the model $\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + 8x_1x_2$.

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**

The Advantage of Factorials (1)

- More efficient than one-factor-at-a-time experiments.
  - One-factor-at-a time:

    effect of factor A: $A^+B^- - A^-B^-$ (2 obs)

    effect of factor B: $A^-B^+ - A^-B^-$ (2 obs)

    total observation needed:

    $2(A^+B^-)+2(A^-B^-)+2(A^-B^+)=6$
  - Two factorial:

    effect of factor A: $A^+B^- - A^-B^-$, $A^+B^+ - A^-B^+$

    effect of factor B: $A^-B^+ - A^-B^-$, $A^+B^+ - A^+B^-$

    total observation needed:

    $1(A^+B^-)+1(A^-B^-)+1(A^-B^+)+1(A^+B^+)=4$

Efficiency =6/4=1.5, increase as the number of factors increases

The Advantage of Factorials (2)

- Necessary when interactions may be present to avoid misleading conclusion

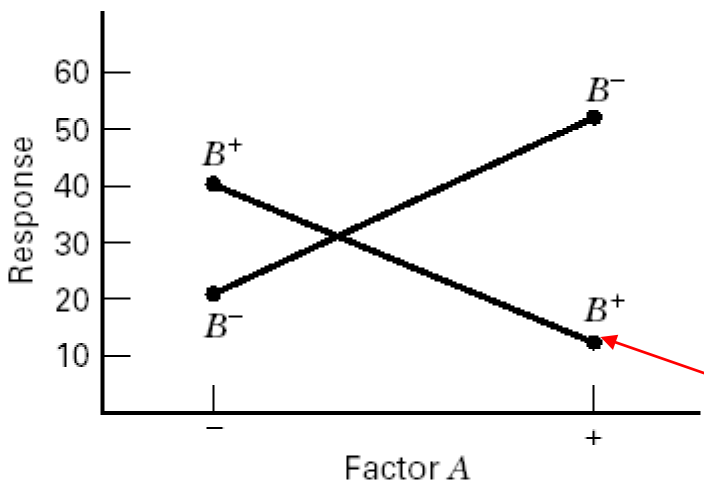


Figure 5-4 A factorial experiment with interaction.

The one-factor-at-a-time design, would indicate that A^-B^+ and A^+B^- gave better responses than A^-B^-. A logical conclusion would be that A^+B^+ would be even better. However ?

- allow the effects of a factor to be estimated at several levels of the other factors, yielding conclusions that are valid over a range of experimental conditions.

Example 5-1 The Battery Life Experiment
Text reference pg. 165

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)?

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 (effects) model:

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

– Identifiability constraints: $\mu_{..}, \ \alpha_i, \ \ \beta_j, \ (\alpha\beta)_{ij}$ are unknown parameters (fixed effects) subject to:

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

– Model assumptions:

$$\{\epsilon_{ijk}\} \quad \text{i.i.d.} \sim N(0, \sigma^2).$$

- Estimation by sample means:

$$\overline{Y}_{ij\cdot} = \frac{1}{n}\sum_{k=1}^{n} Y_{ijk} \longrightarrow \mu_{ij} = \mu_{\cdot\cdot} + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

$$\overline{Y}_{i\cdot\cdot} = \frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk} \longrightarrow \mu_{i\cdot} = \mu_{\cdot\cdot} + \alpha_i$$

$$\overline{Y}_{\cdot j\cdot} = \frac{1}{an}\sum_{i=1}^{a}\sum_{k=1}^{n} Y_{ijk} \longrightarrow \mu_{\cdot j} = \mu_{\cdot\cdot} + \beta_j$$

$$\overline{Y}_{\cdots} = \frac{1}{abn}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk} \longrightarrow \mu_{\cdot\cdot}$$

Thus

$$\hat{\alpha}_i = \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdots} \longrightarrow \mu_{i\cdot} - \mu_{\cdot\cdot} = \alpha_i$$

$$\hat{\beta}_j = \overline{Y}_{\cdot j\cdot} - \overline{Y}_{\cdots} \longrightarrow \mu_{\cdot j} - \mu_{\cdot\cdot} = \beta_j$$

$$\widehat{(\alpha\beta)}_{ij} = \overline{Y}_{ij\cdot} - \overline{Y}_{\cdots} - (\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdots}) - (\overline{Y}_{\cdot j\cdot} - \overline{Y}_{\cdots})$$

$$\longrightarrow \mu_{ij} - \mu_{\cdot\cdot} - \alpha_i - \beta_j = (\alpha\beta)_{ij}.$$

- Residuals:

$$e_{ijk} = Y_{ijk} - \overline{Y}_{ij.} \longrightarrow \epsilon_{ijk}.$$

- Partitioning of total sum of squares (SSTO)
  - Decompose of total deviations

$$\begin{aligned} Y_{ijk} - \overline{Y}_{...} &= (\overline{Y}_{ij.} - \overline{Y}_{...}) + (Y_{ijk} - \overline{Y}_{ij.}) \\ &= (\overline{Y}_{ij.} - \overline{Y}_{...}) + e_{ijk}. \end{aligned}$$

  - Decompose of SSTO

$$\underbrace{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk} - \overline{Y}_{...})^2}_{SSTO} = \underbrace{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{Y}_{ij.} - \overline{Y}_{...})^2}_{SSTR} + \underbrace{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk} - \overline{Y}_{ij.})^2}_{SSE}$$

Thus

$$SSTO = SSTR + SSE$$

- Partitioning of treatment sum of squares (SSTR)
  - Decompose of the treatment deviations

$$\begin{array}{ccccccc} \overline{Y}_{ij\cdot} - \overline{Y}_{\cdots} & = & (\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdots}) & + & (\overline{Y}_{\cdot j\cdot} - \overline{Y}_{\cdots}) & + & (\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y}_{\cdots}) \\ & = & \hat{\alpha}_i & + & \hat{\beta}_j & + & \widehat{(\alpha\beta)}_{ij}. \end{array}$$

  - Decompose of SSTR

$$n\sum_{i,j}(\overline{Y}_{ij\cdot} - \overline{Y}_{\cdots})^2 = nb\sum_{i}(\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdots})^2 + na\sum_{j}(\overline{Y}_{\cdot j\cdot} - \overline{Y}_{\cdots})^2$$

$$+n\sum_{i,j}(\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y}_{\cdots})^2.$$

Thus

$$SSTR = SSA + SSB + SSAB,$$

where SSA——factor A sum of squares, SSB——factor B sum of squares, SSAB——AB interaction sum of squares.

- Final partitioning of SSTO is

$$SSTO = SSA + SSB + SSAB + SSE$$

- Partitioning of degrees of freedom

  –

$$\begin{array}{rcccc} SSTO & = & SSTR & + & SSE \\ df: n_T - 1 & = & (ab-1) & + & (n_T - ab) \end{array}$$

  where $n_T = nab$, $n_T - ab = (n-1)ab$.

  –

$$\begin{array}{rcccccc} SSTR & = & SSA & + & SSB & + & SSAB \\ df: ab - 1 & = & a-1 & + & b-1 & + & (a-1)(b-1 \end{array}$$

- Mean squares:= $\frac{\text{sum of squares}}{\text{degrees of freedom}}$

$$MSE = \frac{SSE}{(n-1)ab}, MSA = \frac{SSA}{a-1}, MSB = \frac{SSB}{b-1},$$

$$MSAB = \frac{SSAB}{(a-1)(b-1)}$$

- Expected mean squares:

$$\mathrm{E}(MSE) = \sigma^2,$$

$$\mathrm{E}(MSA) = \sigma^2 + nb\frac{\sum_{i=1}^{a} \alpha_i^2}{a-1},\ \mathrm{E}(MSB) = \sigma^2 + na\frac{\sum_{j=1}^{b} \beta_j^2}{b-1},$$

$$\mathrm{E}(MSAB) = \sigma^2 + n\frac{\sum_{i=1}^{a}\sum_{j=1}^{b} (\alpha\beta)_{ij}^2}{(a-1)(b-1)}.$$

ANOVA Table – Fixed Effects Case

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	$abn-1$		

- F tests for main effects and interactions: F ratio based on the ratio of corresponding mean squares and follows an F distribution under H_0.
  - First to test whether the two factors interact: H_0: all $(\alpha\beta)_{ij} = 0$; vs. H_a : not all $(\alpha\beta)_{ij}$ equal zero. The F ratio

$$F^* = \frac{MSAB}{MSE} \sim F_{(a-1)(b-1),(n-1)ab} \quad \text{under } H_0.$$

Hence compare F^* with $F(1-\alpha; (a-1)(b-1), (n-1)ab)$.

- Test for Factor A main effects: $H_0 : \alpha_{\mathbf{1}} = \cdots = \boldsymbol{\alpha}_a = 0$ vs. H_a: not all α_i equal zero. Use F ratio

$$F^* = \frac{MSA}{MSE} \sim F_{a-1,(n-1)ab} \quad \text{under } H_0.$$

Hence compare F^* with $F(1-\alpha; a-1, (n-1)ab)$.

- Test for Factor B main effects: $H_0 : \beta_1 = \cdots = \beta_b = 0$ vs. H_a: not all β_j equal zero. Use F ratio

$$F^* = \frac{MSB}{MSE} \sim F_{b-1,(n-1)ab} \quad \text{under } H_0.$$

Hence compare F^* with $F(1-\alpha; b-1, (n-\mathbf{1})ab)$.

Example 5-1

```
>battery.fit=lm(Life~Mtype*Temp)
>anova(battery.fit)
```

Analysis of Variance Table

Response: Life

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Mtype	2	10684	5342	7.9114	0.001976	**
Temp	2	39119	19559	28.9677	1.909e-07	***
Mtype:Temp	4	9614	2403	3.5595	0.018611	*
Residuals	27	18231	675			

- **First, we begin by testing whether or not interaction effects are present:**
  - **H_0:$(\alpha\beta)_{ij}$=0 for all i,j**

    H_a: at least one $\alpha\beta)_{ij} \neq 0$

    Test Statistics F=3.56, p-value=0.019

    since p-value<0.05, we reject H_0 and conclude that there is a significant interaction between material types and temperature

- **Second, we turn to test for material type main effect and temperature main effect**
  - **H_0 : α_i=0 for all i**

    $H_{a:}$ at least one $\alpha_i \neq 0$

    Test Statistics F=7.91, p-value=0.001976

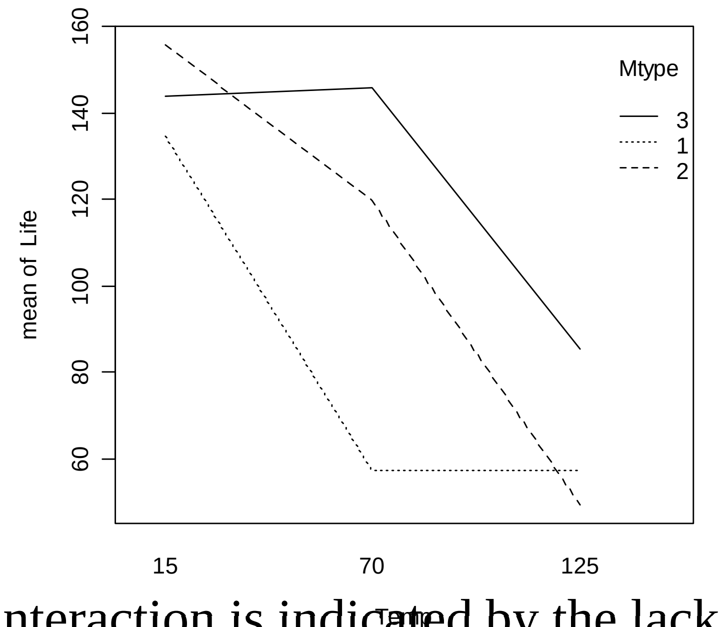
    since p-value<0.05, we reject H_0 and conclude that there is a significant main effect of material type
  - **H_0 : β_j=0 for all j**

    H_a: at least one $\beta_j \neq 0$

    Test Statistics F=28.97, p-value=0.00000019

    since p-value<0.05, we reject H_0 and conclude that there is a significant main effect of temperature.

Interpreting Interaction



- The significant interaction is indicated by the lack of parallelism of the lines.
  - Longer life is attained at low temperature, regardless of material type.
  - Changing from low to intermediate temperature, battery life with material type 3 actually increases, whereas it decreases for types 1 and 2
  - From intermediate to high temperature, battery life decreases for material types 2 and 3 and is essentially unchanged for type 1.
  - Material type 3 seems to give the best result if want less loss of effective life as the temperature changes.

Transformable Interaction

- The interpretation of interactions can be quite difficult when the interacting effects are complex. Only when the interactions have a simple structure, the joint factor effects can be described in a straightforward manner. Try to transform interaction whenever possible

- Example 1: $\mu_{ij} = \mu_{..}\alpha_i\beta_j$ (Multiplicative interactions)

$$\log\mu_{ij} = \log\mu_{..} + \log\alpha_i + \log\beta_j$$
$$\downarrow \qquad \downarrow \qquad \downarrow \qquad \downarrow$$
$$\mu'_{ij} \qquad \mu'_{..} \qquad \alpha'_i \qquad \beta'_j$$

- Example 2: $\mu_{ij} = \alpha_i + \beta_j + 2\sqrt{\alpha_i}\sqrt{\beta_j}$

$$\sqrt{\mu_{ij}} = \sqrt{\alpha_i} + \sqrt{\beta_j}$$
$$\downarrow \qquad \downarrow \qquad \downarrow$$
$$\mu'_{ij} \qquad \alpha'_i \qquad \beta'_j$$

Multiple Comparisons of Factor Level Means

- When interaction is not significant, apply Tukey's test to the means of factor A (or B) across all the levels of factor B (A).

$$H_0 : \mu_{i.} = \mu_{i'.}$$

$$H_a : \mu_{i.} \neq \mu_{i'.}$$

Tukey critical value

$$T_\alpha = \frac{q_\alpha(a, ab(n-1))}{\sqrt{2}} \sqrt{MSE\left(\frac{1}{bn_i} + \frac{1}{bn_j}\right)}$$

```
#Tukey multiple comparison when interation is not significant
#(should not be used in the battery example, I used here just for illustration)
>TukeyHSD(aov(battery.fit),"Mtype",ordered=TRUE)
>TukeyHSD(aov(battery.fit),"Temp",ordered=TRUE)
```

Tukey multiple comparisons of means
95% family-wise confidence level
factor levels have been ordered

Fit: aov(formula = battery.fit)

$Mtype

	diff	lwr	upr	p adj
2-1	25.16667	-1.135677	51.46901	0.0627571
3-1	41.91667	15.614323	68.21901	0.0014162
3-2	16.75000	-9.552344	43.05234	0.2717815

$Temp

	diff	lwr	upr	p adj
70-125	43.41667	17.11432	69.71901	0.0009787
15-125	80.66667	54.36432	106.96901	0.0000001
15-70	37.25000	10.94766	63.55234	0.0043788

Multiple Comparisons of Factor Level Means

- When interaction is significant, apply Tukey's test to the means of factor A (or B) at each level of factor B(A)

$$H_0 : \mu_{ij} = \mu_{i'j'}$$

$$H_a : \mu_{ij} \neq \mu_{i'j'}$$

- Tukey critical value $T_\alpha = \frac{q_\alpha(a, ab(n-1))}{\sqrt{2}} \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$

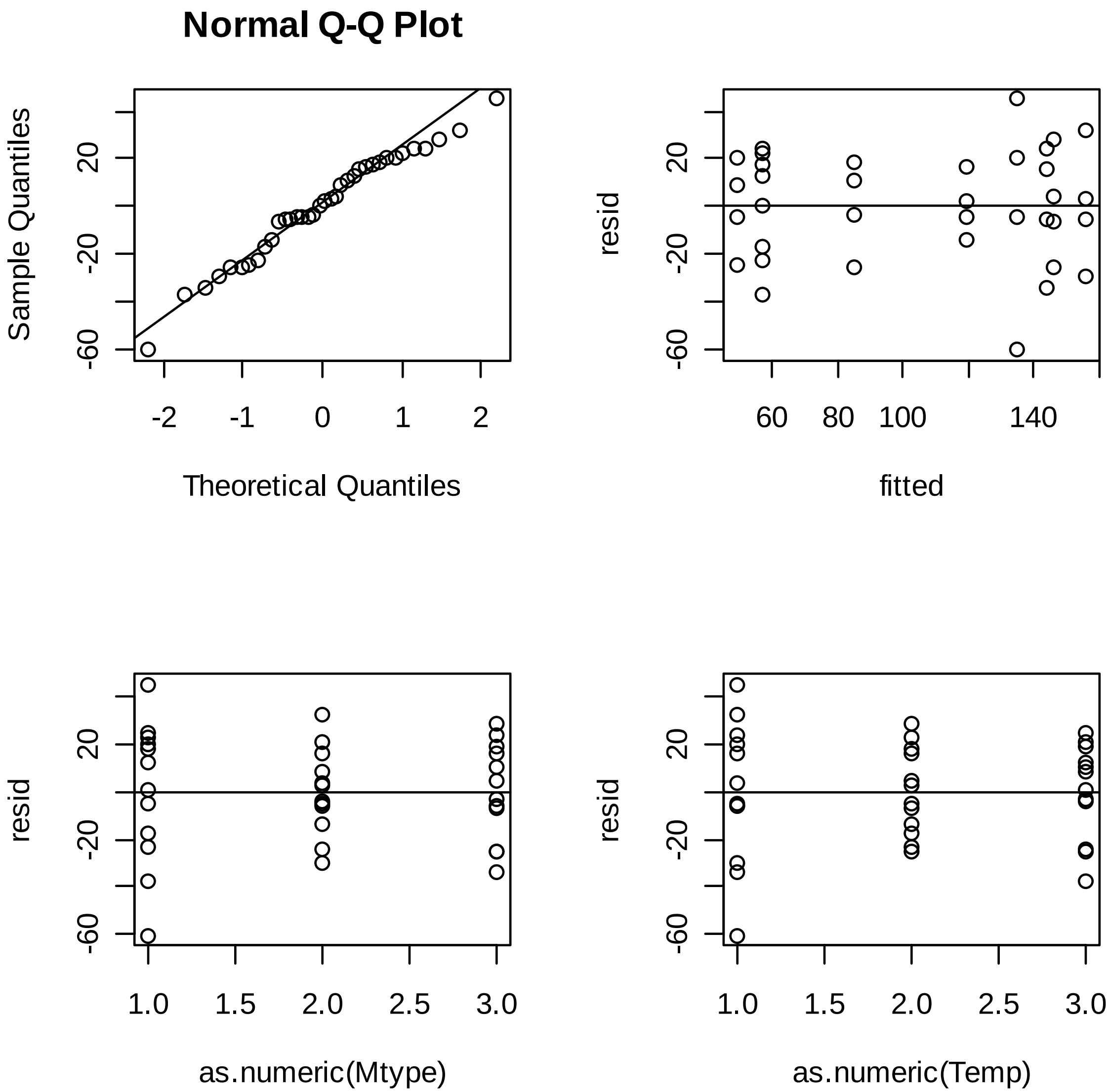
```
#Tukey's significance value Talpha
>qtukey(0.95,3,27)*sqrt(675*(1/4+1/4))/sqrt(2)
#make a table for the material differencs at Temperature 70
>scoefs<-
c(0,unlist(battery.fit$coef[6])+unlist(battery.fit$coef[2]),unlist(battery.fit$coef[7])+
unlist(battery.fit$coef[3]))
>outer(scoefs,scoefs, "-")
```

		Mtype2:Temp70	Mtype3:Temp70
	0.0	-62.5	-88.5
Mtype2:Temp70	62.5	0.0	-26.0
Mtype3:Temp70	88.5	26.0	0.0

```
> qtukey(0.95,3,27)*sqrt(675*(1/4+1/4))/sqrt(2)
[1] 45.54981
```

Model Checking

```
#######Check Assumptions
resid<-battery.fit$resid
fitted<-battery.fit$fitted
par(mfrow=c(2,2))
#normality checking
qqnorm(resid)
qqline(resid)
#constant variances checking
plot(fitted, resid)
abline(h=0)
plot(as.numeric(Mtype),resid)
abline(h=0)
plot(as.numeric(Temp),resid)
abline(h=0)
```


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-60
-20
20
-2
-1
0
1
2
resid
fitted
60
80
100
140
as.numeric(Mtype)
as.numeric(Temp)
1.0
1.5
2.0
2.5
3.0