Design of Experiments: A gentle introduction

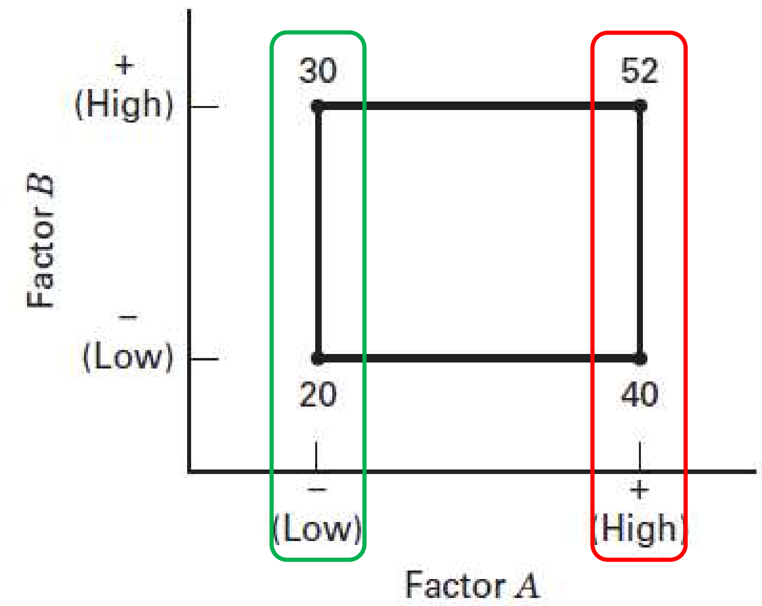
FACTORIAL DESIGN

Fabio Stella
University of Milano-Bicocca
fabio.stella@unimib.it

LECTURE LEARNING OBJECTIVES

1) Learn the definitions of **main effects** and **interactions**.
2) Learn about **two-factor factorial experiments**.
3) Learn how the **analysis of variance** can be extended to factorial experiments.
4) Know how to **check model assumptions in a factorial experiment**.
5) Know how factorial experiments can be used for **more than two factors**.
6) Know how to analyze factorial experiments by **fitting response curves and surfaces**.

Factorial Design; in each complete trial or replicate of the experiment, **all possible combinations of the levels of the factors are investigated**.



When **factors** are arranged in a **factorial design**, they are often said to be **crossed**.

The **effect of a factor** is defined to be the **change in response produced by a change in the level of the factor**. This is also called a **main effect** because it refers to the primary factors of interest in the experiment.

Main Effect of A; *difference between the average response at the High level of A and the average response at the Low level of A*.

$$A = \frac{40 + 52}{2} - \frac{20 + 30}{2} = 21$$

Increasing factor *A* from the Low level to the High level **causes an average response increase** of 21 units.

$$B = \frac{30 + 52}{2} - \frac{20 + 40}{2} = 11$$

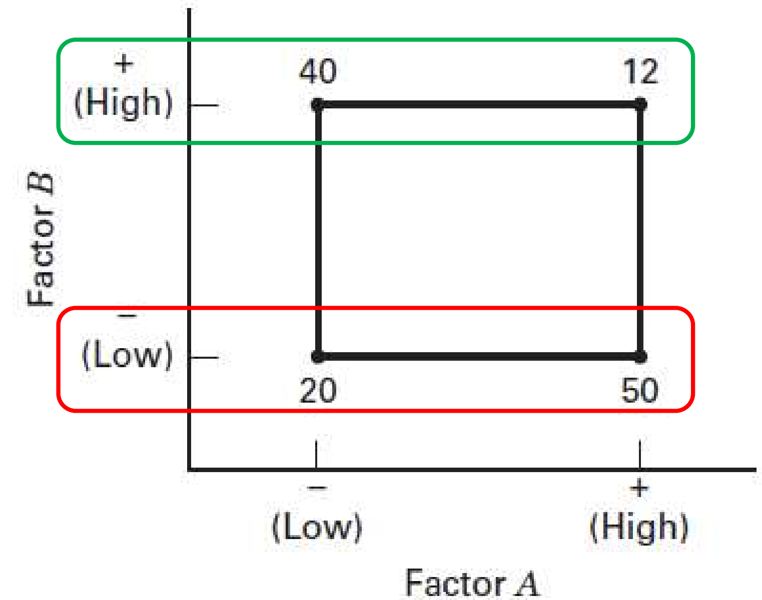
In some experiments, we may find that the difference in response between the levels of one factor is not the same at all levels of the other factors. When this occurs, there is an **interaction between the factors**.

Because the **effect of A depends on the level chosen for factor B**, we see that there is **interaction between A and B**.

The **magnitude of the interaction** effect is the **average difference in these two A effects**, or

$$AB = \frac{(-28-3\ \ )}{2} = -29$$

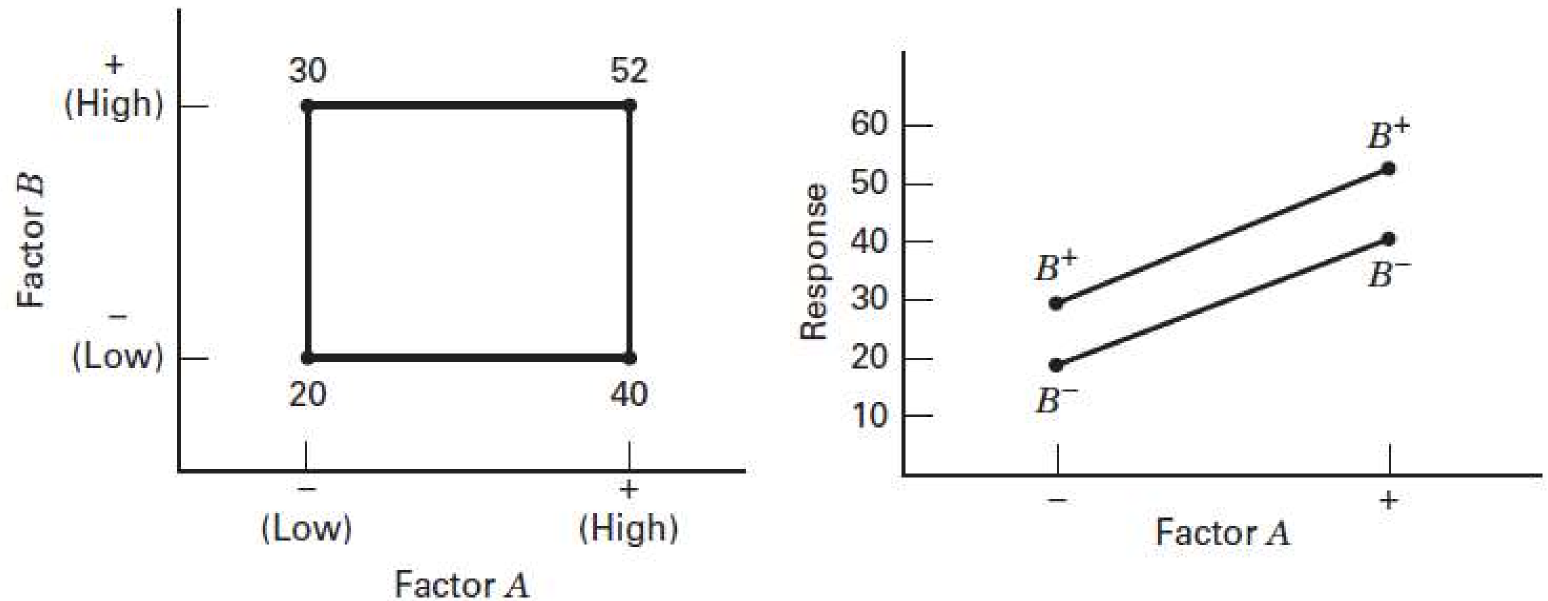
Clearly, the interaction is large in this experiment.



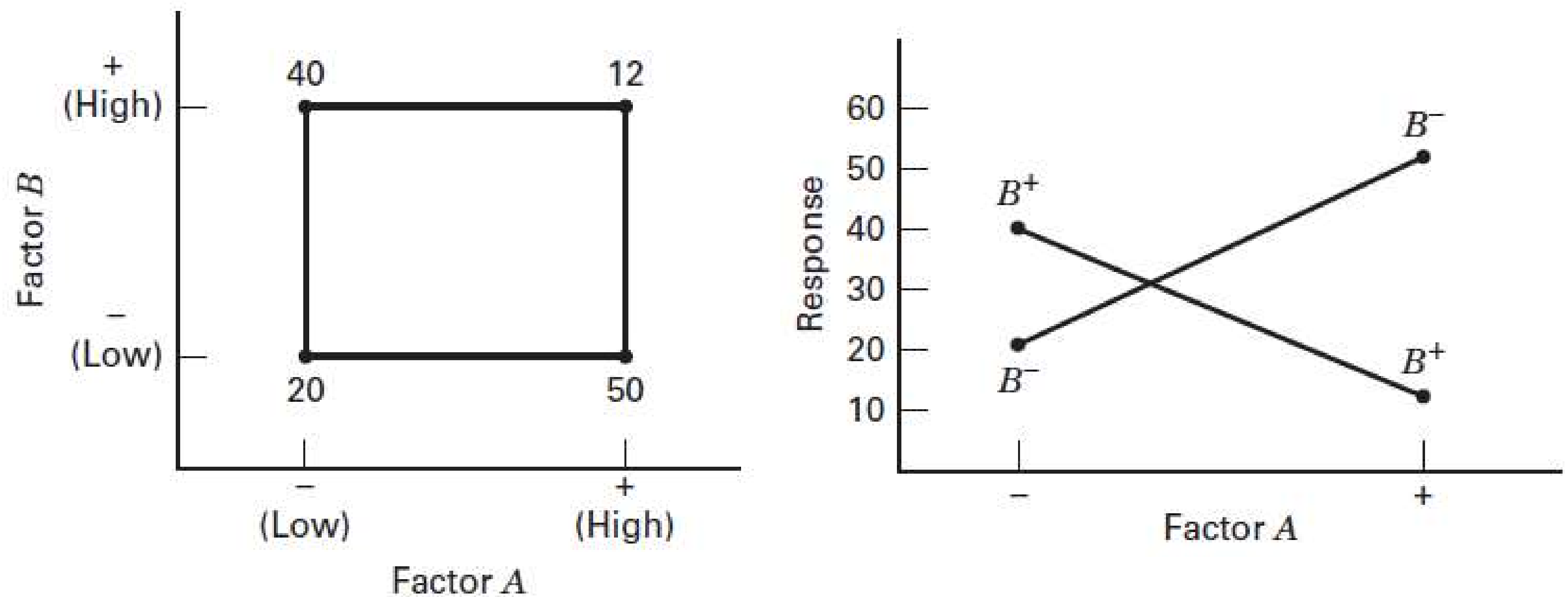
$A = 50 - 20 = 30$

$A = 12 - 40 = -28$

Plots the response data against factor A for both levels of factor B. Note that the **B− and B+ lines are approximately parallel**, indicating a **lack of interaction between factors A and B**.



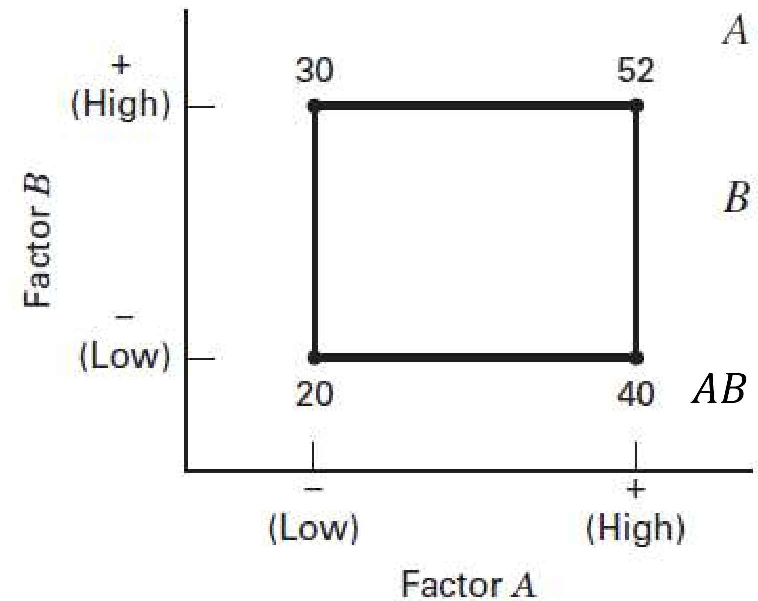
Plots the response data against factor A for both levels of factor B. Here we see that the **B− and B+ lines are not parallel**. This indicates an **interaction between factors A and B**.



Factorial Design: Basic definitions and principles

There is another way to illustrate the concept of interaction.

Suppose that both of our **design factors** are **quantitative** (such as temperature, pressure, time). Then a **regression model representation of the two-factor factorial experiment** could be written as



$$A = \frac{40+52}{2} - \frac{20+30}{2} = 21$$

$$B = \frac{30+52}{2} - \frac{20+40}{2} = 11$$

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

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

response (y), Factor A (x_1), Factor B (x_2), Interaction AB ($x_1 x_2$), random error term (ε)

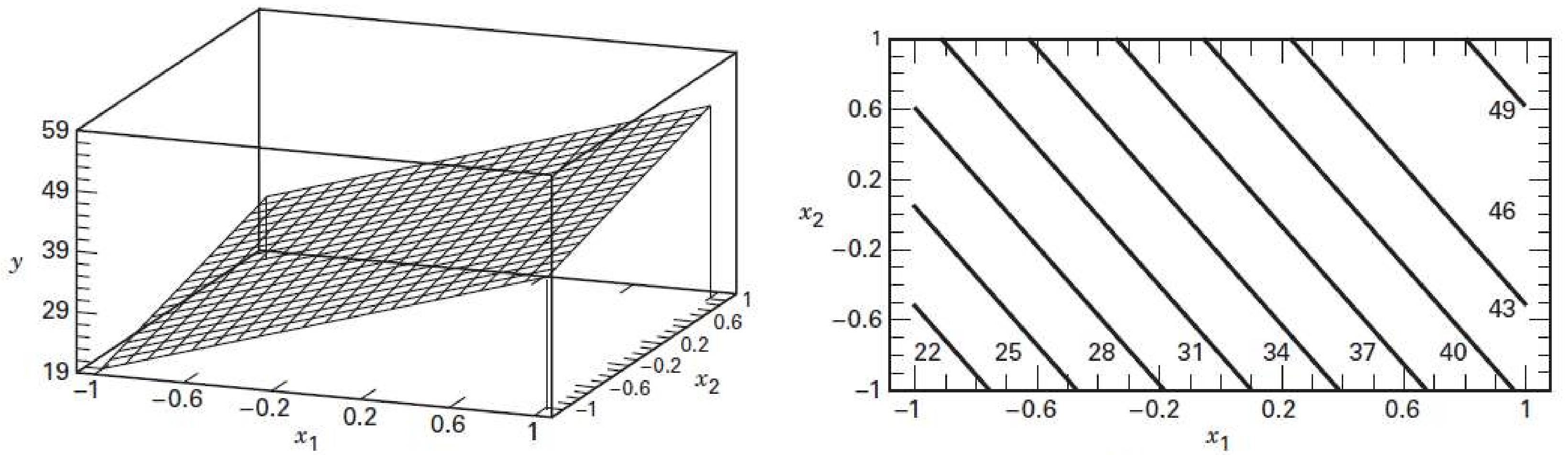
$$\beta_0 = \frac{(20+40+30+52)}{4} = 35.5$$

$$\beta_1 = \frac{21}{2} = 11.5 \qquad \beta_2 = \frac{11}{2} = 5.5$$

$$\beta_{12} = \frac{1}{2} = 0.5$$

Least Squares Estimate

$$y = 35.5 + 11.5x_1 + 5.5x_2 + 0.5x_1x_2 + \varepsilon$$



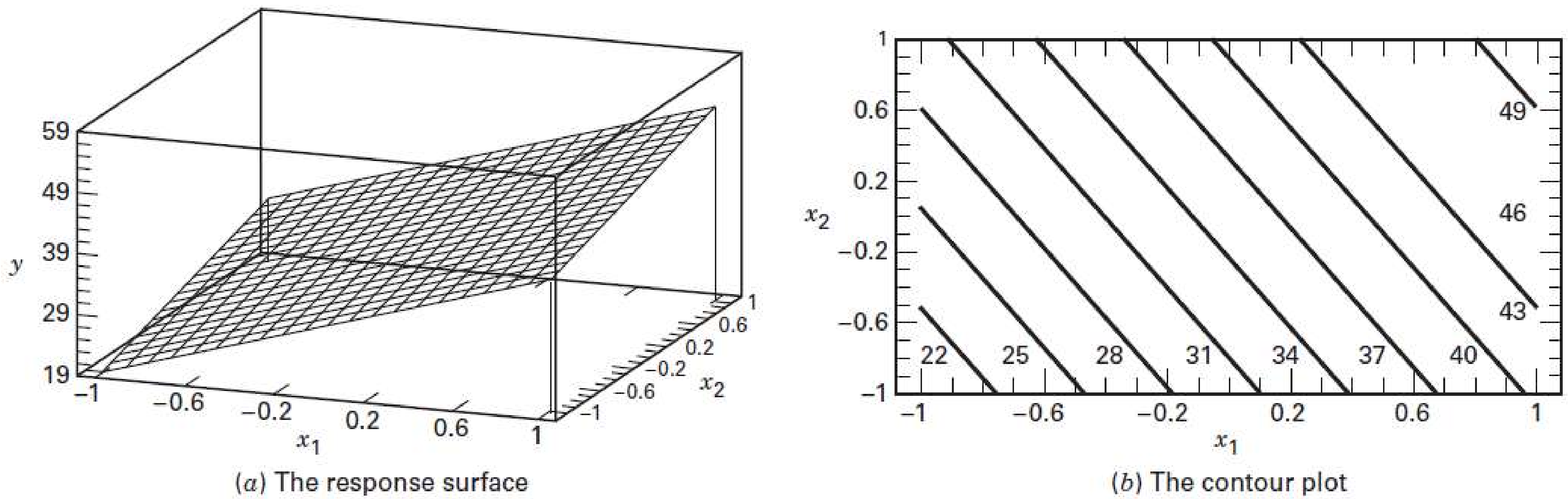
(*a*) The response surface (*b*) The contour plot

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

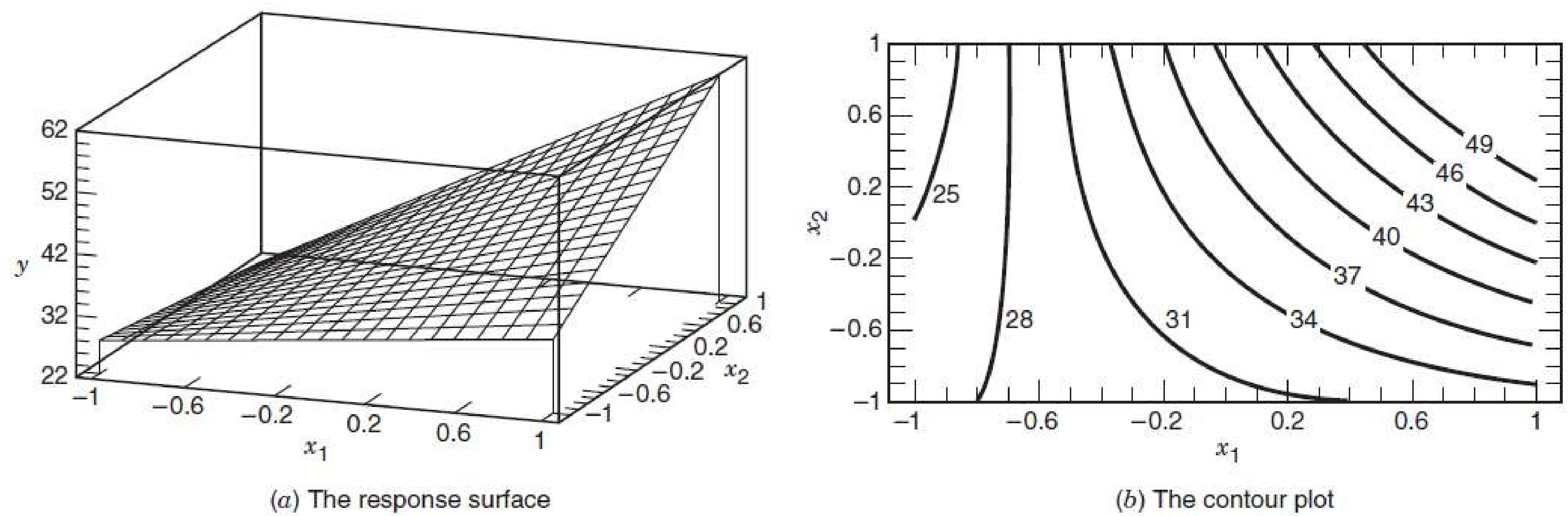
Small compared to main effects

$$y = 35.5 + 11.5x_1 + 5.5x_2 + 0.5x_1x_2 + \varepsilon$$

$$y = 35.5 + 11.5x_1 + 5.5x_2 + \varepsilon$$



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

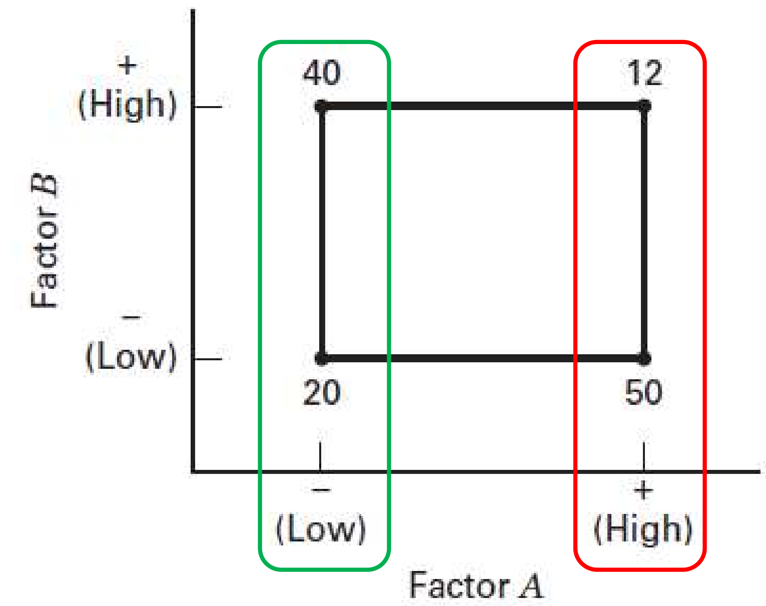


■ **FIGURE 5.6** Response surface and contour plot for the model $\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + 8x_1x_2$

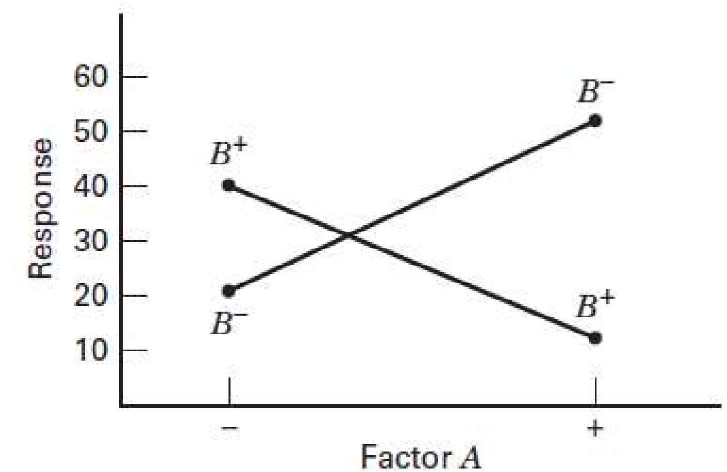
Generally, **when an interaction is large, the corresponding main effects have little practical meaning**.

For the experiment in figure to the right, we would estimate the main effect of A to be

$$A = \frac{50+12}{2} - \frac{20+40}{2} = 1$$



The **main effect is very small**, and we are tempted to conclude that there is no effect due to A. However, when we examine the effects of A at different levels of factor B, we see that this is not the case.



Factor A has an effect, but it depends on the level of factor B. That is, **knowledge of the AB interaction is more useful than knowledge of the main effect.** A significant interaction will often mask the significance of main effects.

In the presence of significant interaction, the experimenter must usually examine the levels of one factor, say A, with levels of the other factors fixed to draw conclusions about the main effect of A.

The simplest types of factorial designs involve only two factors or sets of treatments.

There are “***a” levels of factor A*** and “***b” levels of factor B***, and these are arranged in a **factorial design**; that is, **each replicate of the experiment** contains **all “*ab*” treatment combinations**. In general, there are “***n*” replicates**.

An engineer is designing a **battery for use in a device** that will be subjected to some **extreme variations in temperature**.

The only **design parameter** that she/he can select at this point is the **plate material (controlled factor) for the battery**, and she/he has **three possible choices**.

When the device is manufactured and is shipped to the field, the engineer has *no control over the temperature* **(uncontrolled factor)** *extremes that the device will encounter*, and she/he knows from experience that *temperature will probably affect the effective battery life* (**response**).

However, **temperature can be controlled in the product development laboratory** for the purposes of a test.

The engineer decides to test all **three plate materials** at **three temperature levels**—15, 70, and 125∘F—because these temperature levels are consistent with the product end-use environment.

Because there are **two factors (plate material & temperature) at three levels**, this design is sometimes called a $\mathbf{3^2}$ **factorial design** (*Two-Factor factorial*).

Four batteries are tested at **each combination of plate material and temperature**, and all **36 tests are run in random order**.

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

In this problem, the engineer wants to **answer the following questions**:

- What effects do material type and temperature have on the life of the battery?
- Is there a choice of material that would give *uniformly long life regardless of temperature*?

To pass to the general case, let y_{ijk} be the **observed response** when

- **factor A** is at the i^{th} level ($i = 1, 2, . . . , a$) and
- **factor B** is at the j^{th} level ($j = 1, 2, . . . , b$)
- for the k^{th} **replicate** ($k = 1, 2, . . . , n$).

In general, a **two-factor factorial experiment** will appear as in table right. The order in which the *abn* observations are taken is selected at random so that this design is a **completely randomized design**.

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

The observations in a factorial experiment can be described by a model. There are several ways to write the model for a factorial experiment.

The **effects model** is

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

- μ: Overall mean effect
- τ_i: effect of the *i*th level of the row factor A
- β_j: effect of the *j*th level of the column factor B
- $(\tau\beta)_{ij}$: effect of the interaction between A and B at each pair *ij*
- ϵ_{ijk}: random error component

Factor *B*

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

Another possible model for a factorial experiment is the **means model.**

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

$$\mu_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij}$$

We could also use a **regression model** as shown before.

Regression models are particularly useful when **one or more of the factors in the experiment are quantitative**.

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

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

In the two-factor factorial, both row and column factors (or treatments), A and B, are of equal interest.

Specifically, we are interested in **testing hypotheses about the**:

- **equality of row treatment effects**

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0$$
$$H_1: \text{at least one } \tau_i \neq 0$$

- **equality of column treatment effects**

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_b = 0$$
$$H_1: \text{at least one } \beta_i \neq 0$$

- **determining whether row and column treatments interact**

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

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

$$\mu_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij}$$

Factor B (columns); Factor A (rows)

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

Total Corrected Sum of Squares $$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{...})^2 \qquad \bar{y}_{...}=\frac{y_{...}}{abn} \qquad y_{...}=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijk}$$

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

Fundamental ANOVA equation for the two-factor factorial

The **total sum of squares is partitioned into** a

- sum of squares due to "rows," or factor A, (SS_A);
- sum of squares due to "columns," or factor B, (SS_B);
- sum of squares due to the interaction between A and B, (SS_{AB});
- sum of squares due to error, (SS_E).

Effect	Degrees of Freedom
A	$a-1$
B	$b-1$
AB interaction	$(a-1)(b-1)$
Error	$ab(n-1)$
Total	$abn-1$

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	$abn-1$		

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

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

$$- 9613.78 = 18{,}230.75$$

Life Data (in hours) for the Battery Design Experiment

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

Factorial Design: The two-factor factorial design

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			

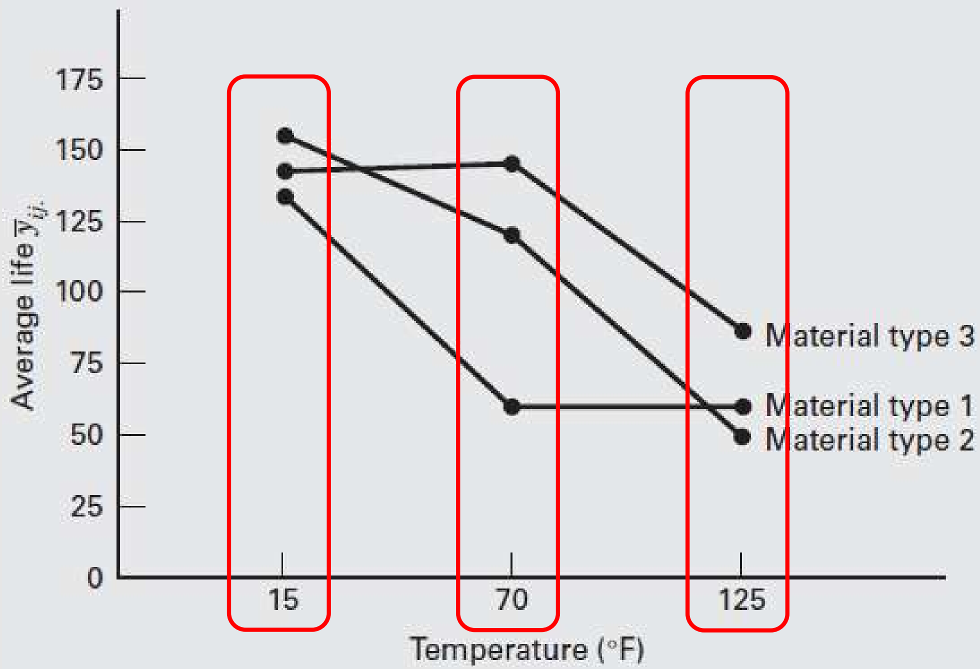
The Analysis of Variance Table for the Two-Factor Factorial, Fixed Effects Model

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	$abn-1$		

Factorial Design: The two-factor factorial design

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			



The **significant interaction is indicated by the lack of parallelism of the lines**. In general, longer life is attained at low temperature, regardless of material type.

Changing from low to intermediate temperature, battery life with material type 3 may actually increase, 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 results if we want less loss of effective life as the temperature changes.

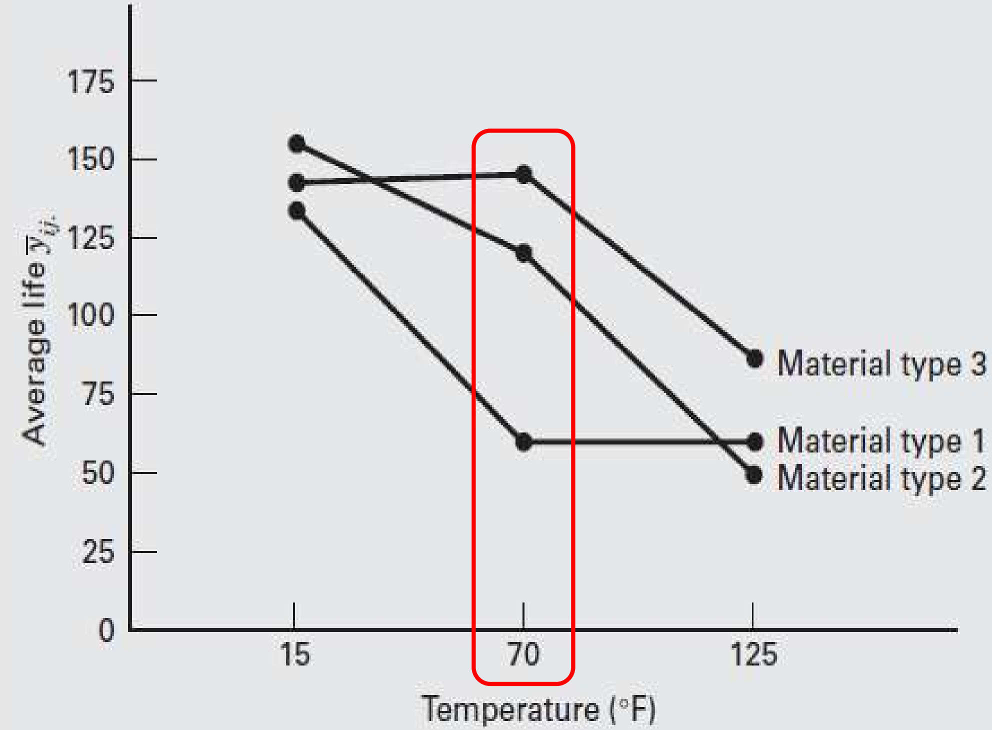
When interaction is significant, comparisons between the means of one factor (e.g., A) may be obscured by the AB interaction.

One approach to this situation is to **fix factor B at a specific level and apply Tukey's test to the means of factor A at that level**.

To illustrate, suppose that we are interested in **detecting differences among the means of the three material types**.

Because interaction is significant, we make this comparison at just one level of temperature, say level 2 (70∘F).

We assume that the best estimate of the error variance is the MS_E from the ANOVA table, utilizing the assumption that the experimental error variance is the same over all treatment combinations (**homoscedastic**).



$\bar{y}_{12.} = 57.25$ (material type 1)

$\bar{y}_{22.} = 119.75$ (material type 2)

$\bar{y}_{32.} = 145.75$ (material type 3)

$$T_{0.05} = 45.47$$

3 vs. 1: $145.75 - 57.25 = 88.50 > T_{0.05} = 45.47$

3 vs. 2: $145.75 - 119.75 = 26.00 < T_{0.05} = 45.47$

2 vs. 1: $119.75 - 57.25 = 62.50 > T_{0.05} = 45.47$

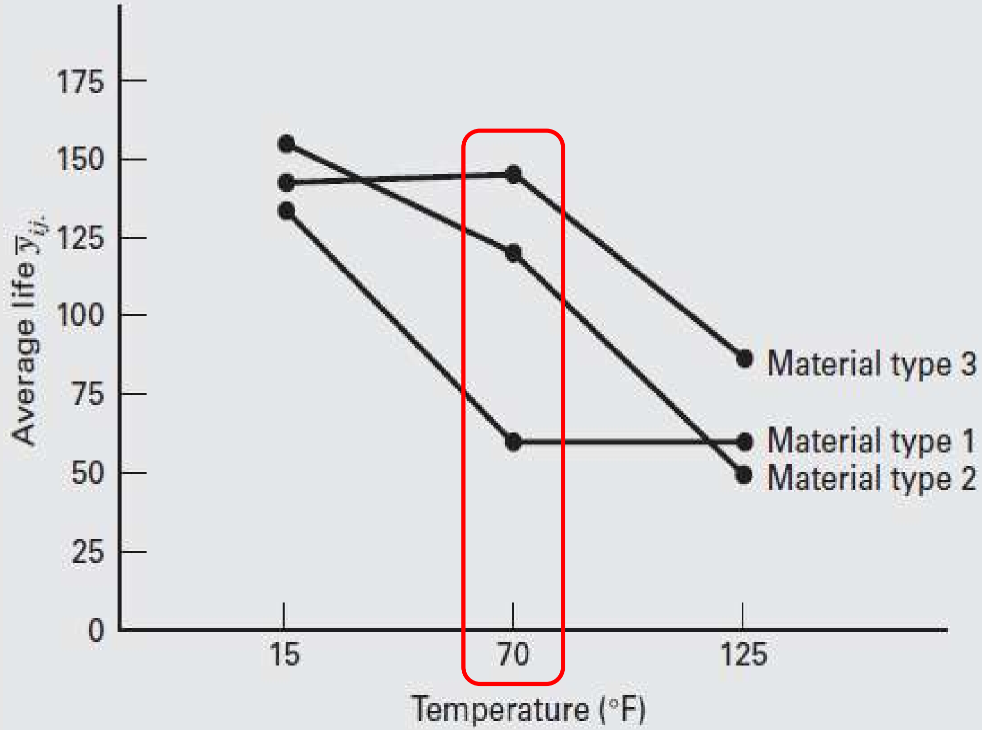
When interaction is significant, comparisons between the means of one factor (e.g., A) may be obscured by the AB interaction.

One approach to this situation is to **fix factor B at a specific level and apply Tukey's test to the means of factor A at that level**.

To illustrate, suppose that we are interested in **detecting differences among the means of the three material types**.

Because interaction is significant, we make this comparison at just one level of temperature, say level 2 (70∘F).

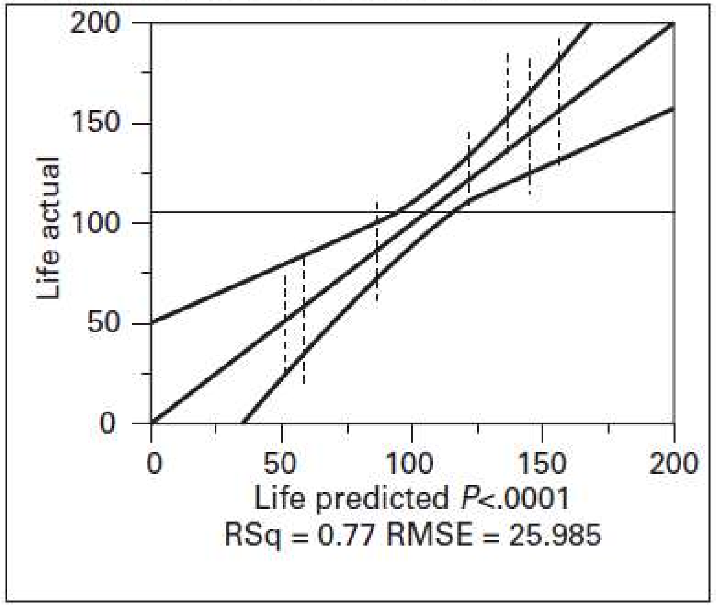
We assume that the best estimate of the error variance is the MS_E from the ANOVA table, utilizing the assumption that the experimental error variance is the same over all treatment combinations.



This analysis indicates that at the temperature level 70∘F, the mean battery life is the same for material types 2 and 3 and that the mean battery life for material type 1 is significantly lower in comparison to both types 2 and 3.

3 vs. 1:	$145.75 - 57.25 = 88.50 > T_{0.05} = 45.47$
3 vs. 2:	$145.75 - 119.75 = 26.00 < T_{0.05} = 45.47$
2 vs. 1:	$119.75 - 57.25 = 62.50 > T_{0.05} = 45.47$

Response Life
Whole Model
Actual by Predicted Plot



Summary of Fit

RSquare	0.76521
RSquare Adj	0.695642
Root Mean Square Error	25.98486
Mean of Response	105.5278
Observations (or Sum Wgts)	36

77 percent of the variability in the battery life is explained by the plate material in the battery, the temperature, and the material type–temperature interaction,

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	*F* Ratio
Model	8	59416.222	7427.03	10.9995
Error	27	18230.750	675.21	Prob > F
C.Total	35	77646.972		<.001

Effect Tests

Source	Nparm	DF	Sum of Squares	*F* Ratio	Prob > *F*
Material Type	2	2	10683.722	7.9114	0.0020
Temperature	2	2	39118.722	28.9677	<.0001
Material Type Temperature	4	4	9613.778	3.5595	0.0186

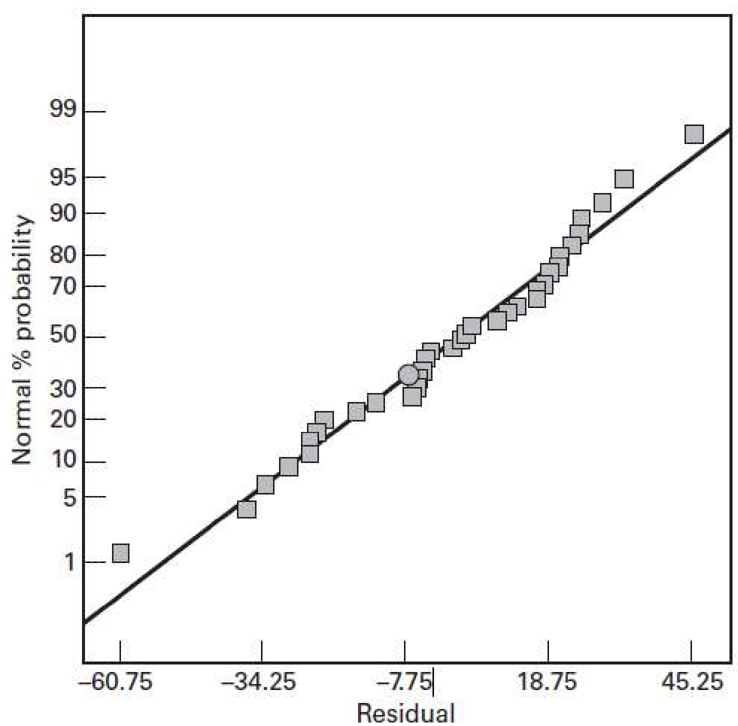
at least one of the three terms in the model is significant.

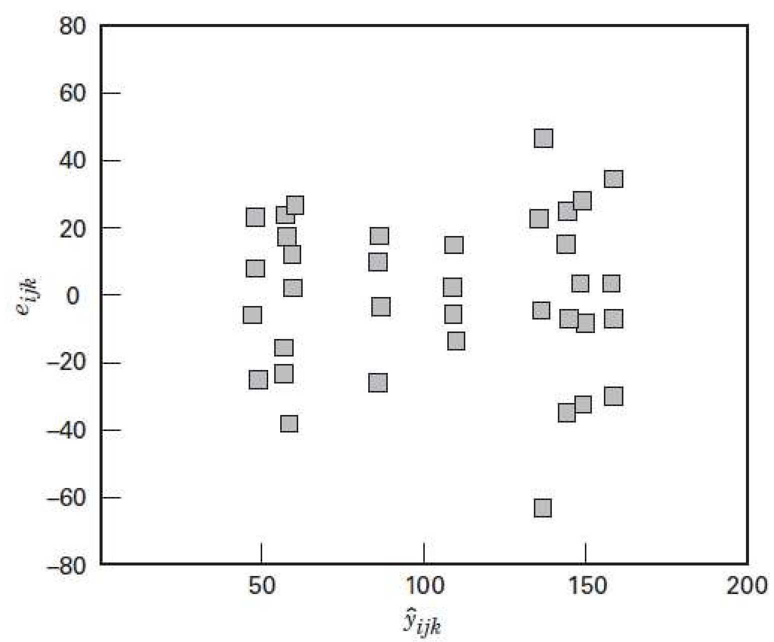
$$SS_{\text{Model}} = SS_{\text{Material}} + SS_{\text{Temperature}} + SS_{\text{Interaction}}$$
$$= 10{,}683.72 + 39{,}118.72 + 9613.78$$
$$= 59{,}416.22$$

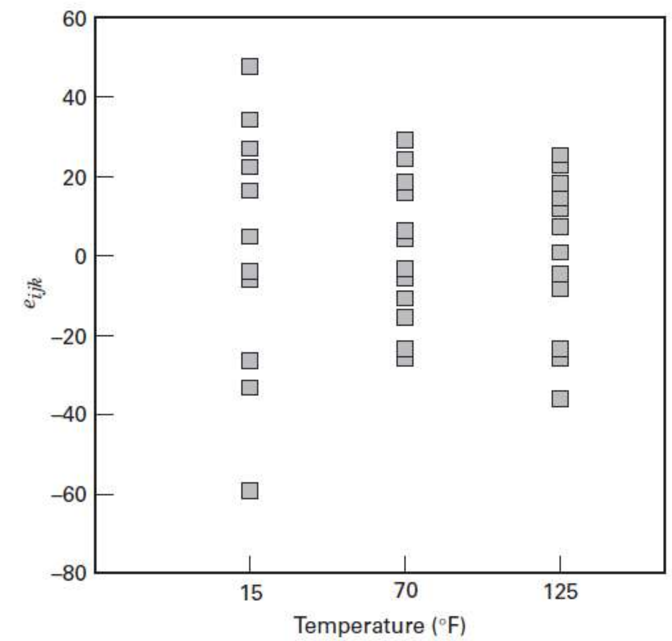
$$R^2 = \frac{SS_{\text{Model}}}{SS_{\text{Total}}} = \frac{59{,}416.22}{77{,}646.97} = 0.7652$$

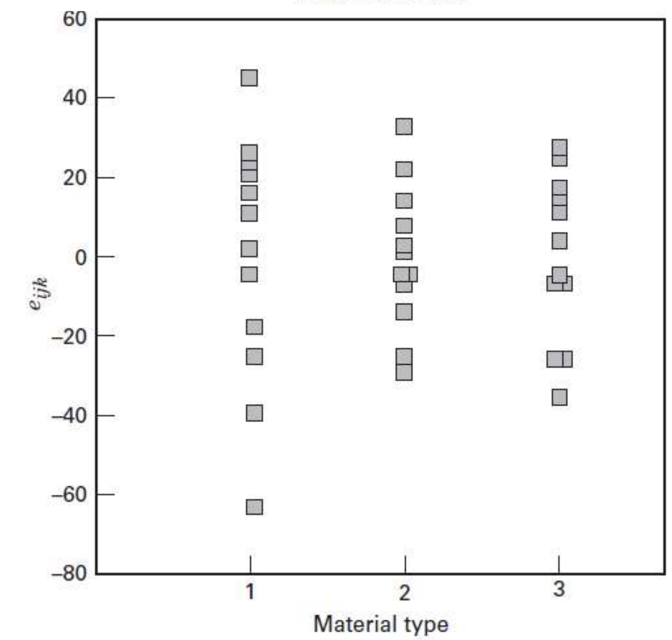
Residuals for Example 5.1

Material Type	Temperature (°F) 15		70		125	
1	−4.75	20.25	−23.25	−17.25	−37.50	12.50
	−60.75	45.25	22.75	17.75	24.50	0.50
2	−5.75	32.25	16.25	2.25	−24.50	20.50
	3.25	−29.75	−13.75	−4.75	8.50	−4.50
3	−6.00	−34.00	28.25	−25.75	10.50	18.50
	24.00	16.00	4.25	−6.75	−3.50	−25.50









The results for the two-factor factorial design may be extended to the general case where there are "***a*** **levels of factor A**", "***b*** **levels of factor B**", "***c*** **levels of factor C**, and so on, **arranged in a factorial experiment**.

In general, there will be ***abc . . . n*** **total observations** if there are ***n*** **replicates of the complete experiment**.

Once again, note that we must have at least two replicates (***n* ≥ 2**) to determine a sum of squares due to error if all possible interactions are included in the model.

For example, consider the **three-factor analysis of variance model**:

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \epsilon_{ijkl} \qquad \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \\ k = 1, 2, \ldots, c \\ l = 1, 2, \ldots, n \end{cases}$$

Assuming that A, B, and C are fixed, the **analysis of variance table** is shown in the table to the right.

The **F-tests on main effects and interactions** follow directly from the expected mean squares.

The Analysis of Variance Table for the Three-Factor Fixed Effects Model

Source of Variation	Sum of Square	Degrees of Freedom	Mean Squares	Expected Mean Square	F_0
A	SS_A	$a-1$	MS_A	$\sigma^2 + \frac{bcn\sum \tau_i^2}{a-1}$	$F_0 = \frac{MS_A}{MS_E}$
B	SS_B	$b-1$	MS_B	$\sigma^2 + \frac{acn\sum \beta_j^2}{b-1}$	$F_0 = \frac{MS_B}{MS_E}$
C	SS_C	$c-1$	MS_C	$\sigma^2 + \frac{abn\sum \gamma_k^2}{c-1}$	$F_0 = \frac{MS_C}{MS_E}$
AB	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	$\sigma^2 + \frac{cn\sum\sum(\tau\beta)_{ij}^2}{(a-1)(b-1)}$	$F_0 = \frac{MS_{AB}}{MS_E}$
AC	SS_{AC}	$(a-1)(c-1)$	MS_{AC}	$\sigma^2 + \frac{bn\sum\sum(\tau\gamma)_{ik}^2}{(a-1)(c-1)}$	$F_0 = \frac{MS_{AC}}{MS_E}$
BC	SS_{BC}	$(b-1)(c-1)$	MS_{BC}	$\sigma^2 + \frac{an\sum\sum(\beta\gamma)_{jk}^2}{(b-1)(c-1)}$	$F_0 = \frac{MS_{BC}}{MS_E}$
ABC	SS_{ABC}	$(a-1)(b-1)(c-1)$	MS_{ABC}	$\sigma^2 + \frac{n\sum\sum\sum(\tau\beta\gamma)_{ijk}^2}{(a-1)(b-1)(c-1)}$	$F_0 = \frac{MS_{ABC}}{MS_E}$
Error	SS_E	$abc(n-1)$	MS_E	σ^2	
Total	SS_T	$abcn-1$			

The **ANOVA** always treats all of the **factors** in the experiment as if they were **qualitative or categorical**.

However, many experiments involve at least one quantitative factor.

It can be useful to **fit a response curve to the levels of a quantitative factor** so that the experimenter has an equation that relates the response to the factor.

This equation might be used for **interpolation**, that is, for **predicting the response at factor levels between those actually used in the experiment**.

When **at least two factors are quantitative**, we can **fit a response surface for predicting *y* at various combinations of the design factors**.

In general, **linear regression methods are used** to fit these models to the experimental data.

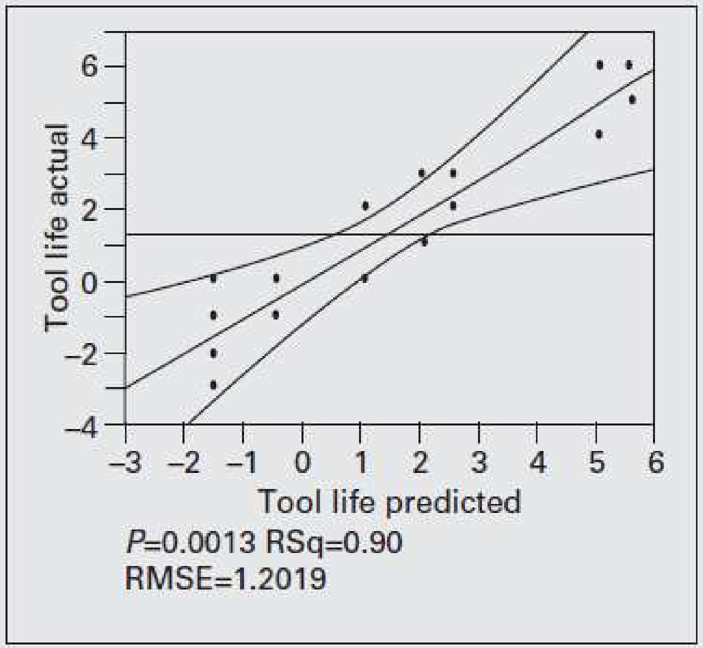
The **effective life of a cutting tool (response)** installed in a numerically controlled machine is thought to be affected by the **cutting speed (factor)** and the **tool angle (factor)**.

Three speeds (levels) and **three angles (levels)** are selected, and a **3^2 factorial experiment with two replicates (*n*=2)** is performed. The circled numbers in the cells are the cell totals $\{y_{ij\cdot}\}$.

Data for Tool Life Experiment

Total Angle (degrees)	Cutting Speed (in/min) 125		150		175		$y_{i..}$
15	−2, −1	(−3)	−3, 0	(−3)	2, 3	(5)	−1
20	0, 2	(2)	1, 3	(4)	4, 6	(10)	16
25	−1, 0	(−1)	5, 6	(11)	0, −1	(−1)	9
$y_{.j.}$	−2		12		14		$24 = y_{...}$

Response Tool Life
Whole Model
Actual by Predicted Plot

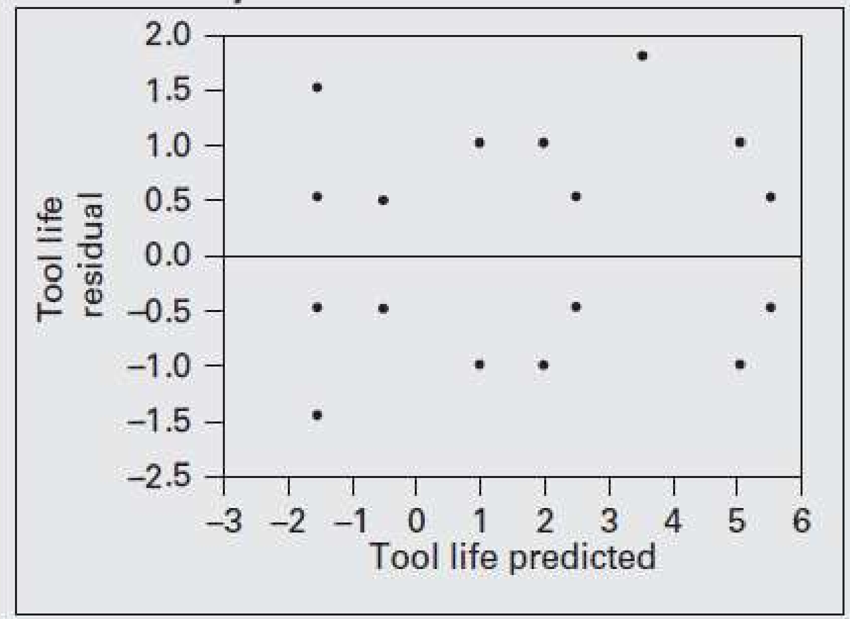


$$R^2 = \frac{SS_{\text{Model}}}{SS_{\text{Total}}}$$

Summary of Fit

RSquare	0.895161
RSquare Adj	0.801971
Root Mean Square Error	1.20185
Mean of Response	1.333333
Observations (or Sum Wgts)	18

Residual by Predicted Plot



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	*F* Ratio
Model	8	111.00000	13.8750	9.6058
Error	9	13.00000	1.4444	Prob > *F*
C. Total	17	124.00000		0.0013

Effect Tests

Source	Nparm	DF	Sum of Squares	*F* Ratio	Prob > *F*
Angle	2	2	24.333333	8.4231	0.0087
Speed	2	2	25.333333	8.7692	0.0077
Angle*Speed	4	4	61.333333	10.6154	0.0018

This is a classical ANOVA, treating both factors as categorical.

Notice that design factors tool angle and speed as well as the angle–speed interaction are significant.

Factorial Design: Fitting response curves and surfaces

Since the **factors are quantitative**, and **both factors have three levels**, we fit a **second-order model** such as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$$

intercept (β_0), angle (x_1), speed (x_2)

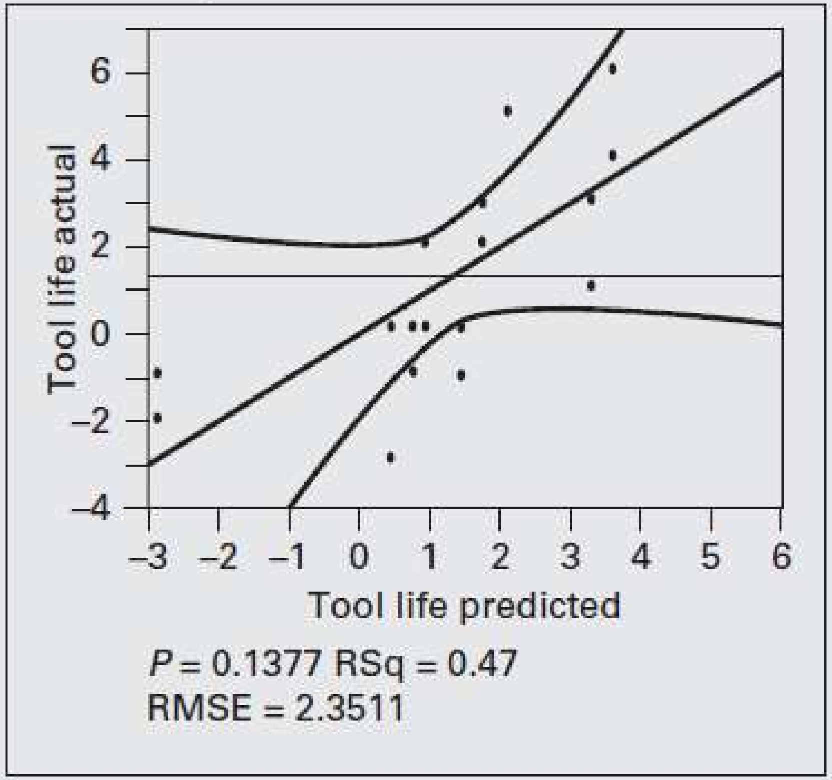
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	57.66667	11.5333	2.0864
Error	12	66.33333	5.5278	Prob > F
C. Total	17	124.00000		0.1377

Parameter Estimates

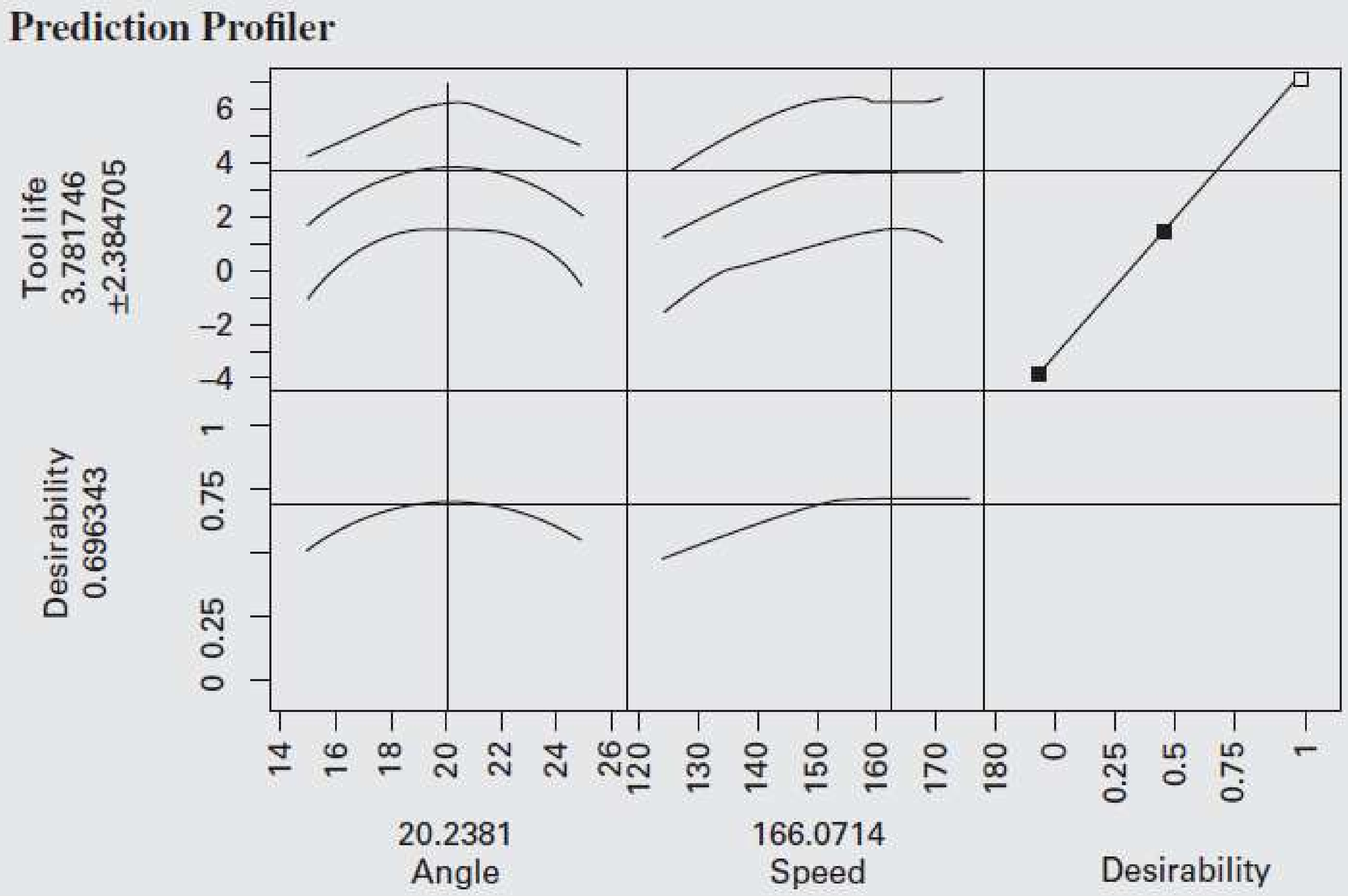
Term	Estimate	Std. Error	t Ratio	Prob > |t|
Intercept	−8	5.048683	−1.58	0.1390
Angle	0.1666667	0.135742	1.23	0.2431
Speed	0.0533333	0.027148	1.96	0.0731
(Angle-20)*(Speed-150)	−0.008	0.00665	−1.20	0.2522
(Angle-20)*(Angle-20)	−0.08	0.047022	−1.70	0.1146
(Speed-150)*(Speed-150)	−0.0016	0.001881	−0.85	0.4116

Response Tool Life
Actual by Predicted Plot



Summary of Fit

RSquare	0.465054
RSquare Adj	0.242159
Root Mean Square Error	2.351123
Mean of Response	1.333333
Observations (or Sum Wgts)	18



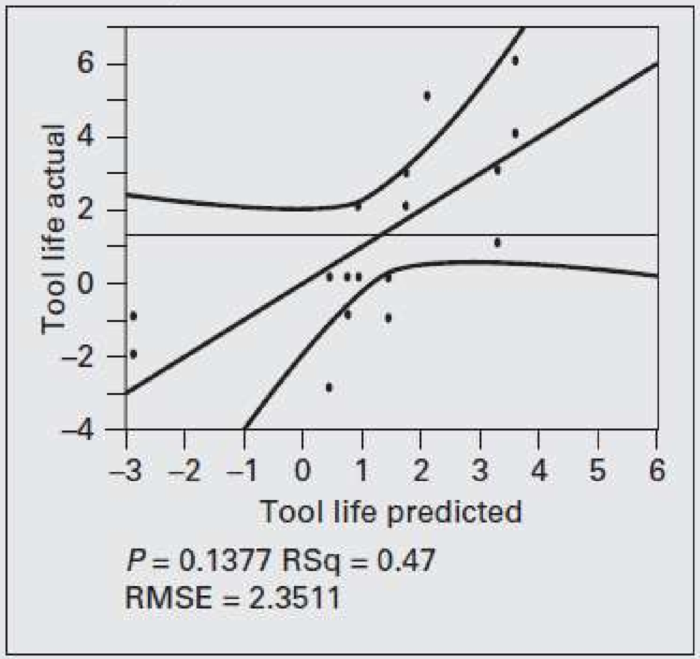
A graphical display showing the response variable life as a function of each design factor, angle and speed.

The prediction profiler is very useful for optimization.

Here it has been set to the levels of angle and speed that result in maximum predicted life.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$$

Response Tool Life
Actual by Predicted Plot



Summary of Fit

RSquare	0.465054
RSquare Adj	0.242159
Root Mean Square Error	2.351123
Mean of Response	1.333333
Observations (or Sum Wgts)	18

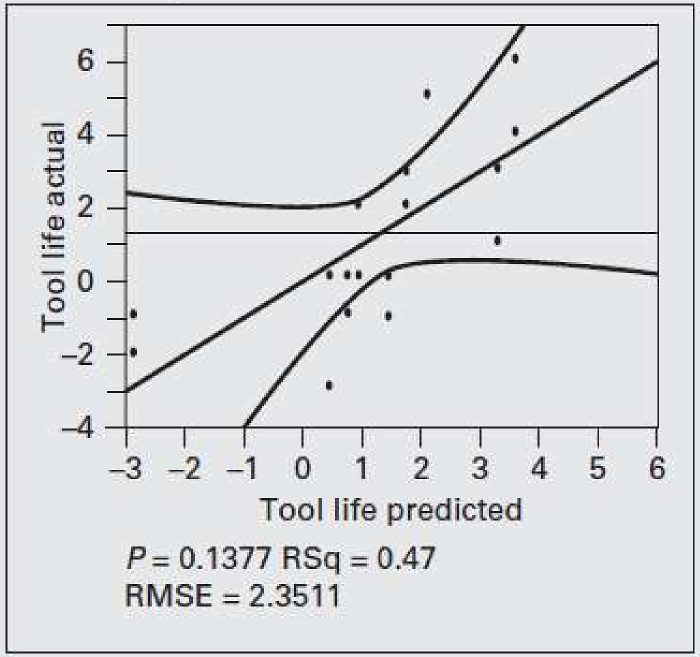
Part of the reason for the relatively poor fit of the **second order model** is that **only one of the four degrees of freedom for interaction are accounted for in this model**.

In addition to the term $\beta_{12} x_1 x_2$, there are three other terms that could be fit to completely account for the four degrees of freedom for interaction, namely

$$\beta_{112} x_1^2 x_2 \qquad \beta_{122} x_1 x_2^2 \qquad \beta_{1122} x_1^2 x_2^2$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{112} x_1^2 x_2 + \beta_{122} x_1 x_2^2 + \beta_{1122} x_1^2 x_2^2 + \varepsilon$$

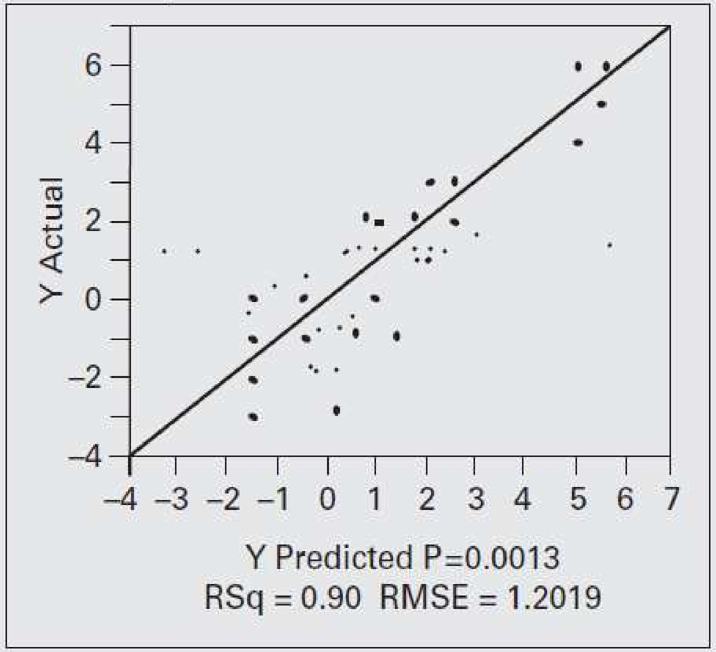
Response Tool Life
Actual by Predicted Plot



Summary of Fit

RSquare	0.465054
RSquare Adj	0.242159
Root Mean Square Error	2.351123
Mean of Response	1.333333
Observations (or Sum Wgts)	18

Response Y
Actual by Predicted Plot



Summary of Fit

RSquare	0.895161
RSquare Adj	0.801971
Root Mean Square Error	1.20185
Mean of Response	1.333333
Observations (or Sum Wgts)	18

Analysis of Variance

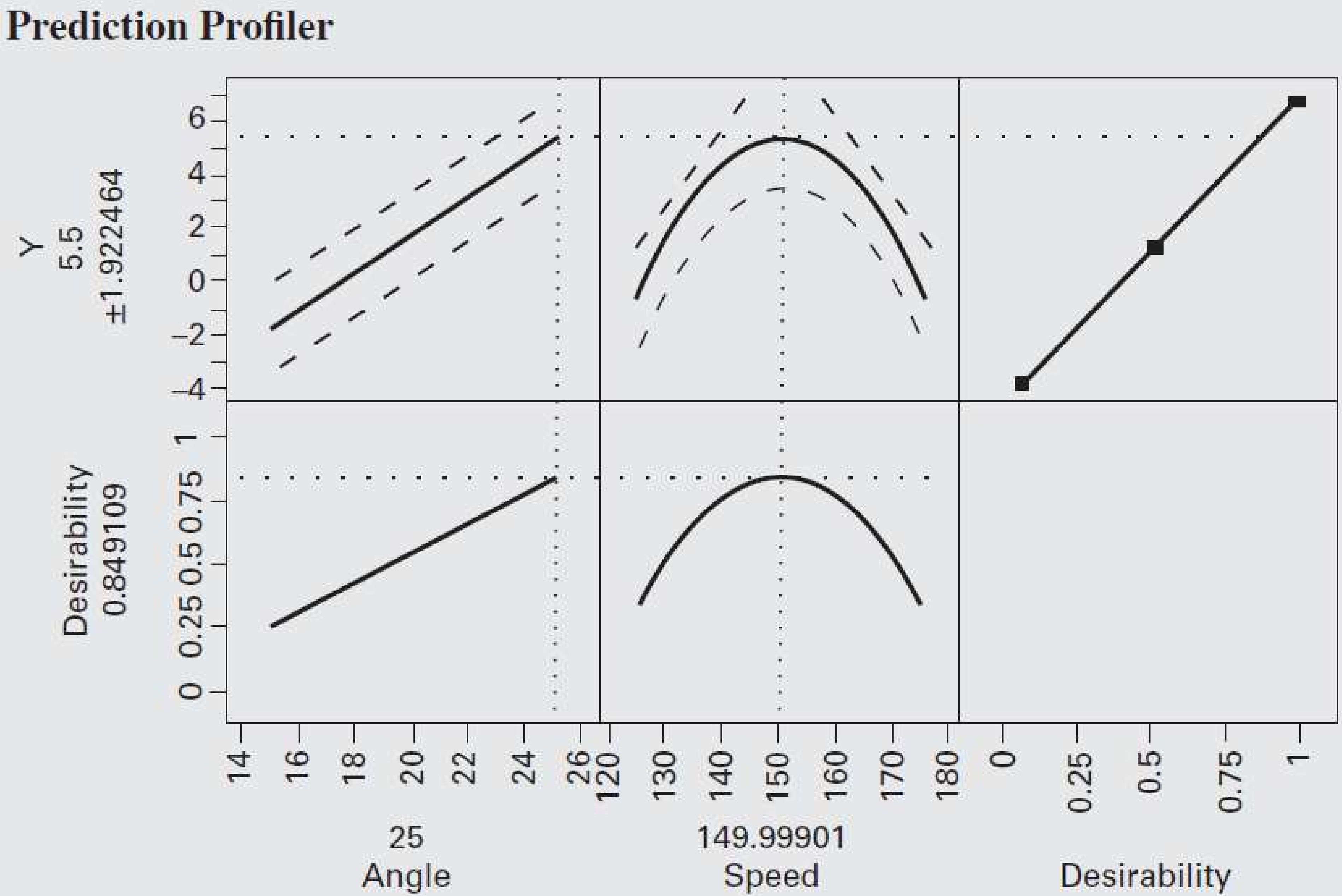
Source	DF	Sum of Squares	Mean Square	*F* Ratio
Model	8	111.00000	13.8750	9.6058
Error	9	13.00000	1.4444	**Prob > *F***
C. Total	17	124.00000		0.0013*

Parameter Estimates

Term	Estimate	Std Error	*t* Ratio	Prob > |*t*|
Intercept	−24	4.41588	−5.43	0.0004*
Angle	0.7	0.120185	5.82	0.0003*
Speed	0.08	0.024037	3.33	0.0088*
(Angle-20)*(Speed-150)	−0.008	0.003399	−2.35	0.0431*
(Angle-20)*(Angle-20)	2.776e-17	0.041633	0.00	1.0000
(Speed-150)*(Speed-150)	0.0016	0.001665	0.96	0.3618
(Angle-20)*(Speed-150)*(Angle-20)	−0.0016	0.001178	−1.36	0.2073
(Speed-150)*(Speed-150)*(Angle-20)	−0.00128	0.000236	−5.43	0.0004*
(Angle-20)*(Speed-150)*(Angle-20)*(Speed-150)	−0.000192	8.158a-5	−2.35	0.0431*

Effect Tests

Source	Nparm	DF	Sum of Squares	*F* Ratio	Prob > *F*
Angle	1	1	49.000000	33.9231	0.0003*
Speed	1	1	16.000000	11.0769	0.0088*
Angle*Speed	1	1	8.000000	5.5385	0.0431*
Angle*Angle	1	1	6.4198e-31	0.0000	1.0000
Speed*Speed	1	1	1.333333	0.9231	0.3618
Angle*Speed*Angle	1	1	2.666667	1.8462	0.2073
Speed*Speed*Angle	1	1	42.666667	29.5385	0.0004*
Angle*Speed*Angle*Speed	1	1	8.000000	5.5385	0.0431*


Prediction Profiler
Y
5.5
±1.922464
Desirability
0.849109
6
4
2
0
−2
−4
0
0.25
0.5
0.75
1
14
16
18
20
22
24
26
120
130
140
150
160
170
180
0
0.25
0.5
0.75
1
25
Angle
149.99901
Speed
Desirability

Factorial Design: Fitting response curves and surfaces