6-3 Multiple Regression

6-3.1 Estimation of Parameters in Multiple Regression

The method of least squares may be used to estimate the regression coefficents in the multiple regression model, equation 6-3. Suppose that $n > k$ observations are available, and let x_{ij} denote the ith observation or level of variable x_j. The observations are

$$(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i) \quad i = 1, 2, \ldots, n > k$$

It is customary to present the data for multiple regression in a table such as Table 6-4.

Table 6-4 Data for Multiple Linear Regression

y	x_1	x_2	$\cdots$	x_k
y_1	x_{11}	x_{12}	$\cdots$	x_{1k}
y_2	x_{21}	x_{22}	$\cdots$	x_{2k}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
y_n	x_{n1}	x_{n2}	$\cdots$	x_{nk}

6-3 Multiple Regression

6-3.1 Estimation of Parameters in Multiple Regression

- The least squares function is given by

$$L = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2$$

- The least squares estimates must satisfy

$$\left. \frac{\partial L}{\partial \beta_0} \right|_{\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k} = -2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \sum_{j=1}^{k} \hat{\beta}_j x_{ij} \right) = 0$$

and

$$\left. \frac{\partial L}{\partial \beta_j} \right|_{\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k} = -2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \sum_{j=1}^{k} \hat{\beta}_j x_{ij} \right) x_{ij} = 0 \quad j = 1, 2, \ldots, k$$

6-3 Multiple Regression

6-3.1 Estimation of Parameters in Multiple Regression

- The **least squares normal equations** are

$$
\begin{array}{rcccccccl}
n\hat{\beta}_0 & + & \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} & + & \hat{\beta}_2 \sum_{i=1}^{n} x_{i2} & + \cdots + & \hat{\beta}_k \sum_{i=1}^{n} x_{ik} & = & \sum_{i=1}^{n} y_i \\
\hat{\beta}_0 \sum_{i=1}^{n} x_{i1} & + & \hat{\beta}_1 \sum_{i=1}^{n} x_{i1}^2 & + & \hat{\beta}_2 \sum_{i=1}^{n} x_{i1}x_{i2} & + \cdots + & \hat{\beta}_k \sum_{i=1}^{n} x_{i1}x_{ik} & = & \sum_{i=1}^{n} x_{i1}y_i \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
\hat{\beta}_0 \sum_{i=1}^{n} x_{ik} & + & \hat{\beta}_1 \sum_{i=1}^{n} x_{ik}x_{i1} & + & \hat{\beta}_2 \sum_{i=1}^{n} x_{ik}x_{i2} & + \cdots + & \hat{\beta}_k \sum_{i=1}^{n} x_{ik}^2 & = & \sum_{i=1}^{n} x_{ik}y_i
\end{array}
$$

- The solution to the normal equations are the **least squares estimators** of the regression coefficients.

6-3 Multiple Regression

EXAMPLE 6-7
Wire Bond Pull Strength

In Chapter 1, we used data on pull strength of a wire bond in a semiconductor manufacturing process, wire length, and die height to illustrate building an empirical model. We will use the same data, repeated for convenience in Table 6-5, and show the details of estimating the model parameters. Scatter plots of the data are presented in Figs. 1-11*a* and 1-11*b*. Figure 6-17 shows a matrix of two-dimensional scatter plots of the data. These displays can be helpful in visualizing the relationships among variables in a multivariable data set.

Table 6-5 Wire Bond Pull Strength Data for Example 6-7

Observation Number	Pull Strength y	Wire Length x_1	Die Height x_2	Observation Number	Pull Strength y	Wire Length x_1	Die Height x_2
1	9.95	2	50	14	11.66	2	360
2	24.45	8	110	15	21.65	4	205
3	31.75	11	120	16	17.89	4	400
4	35.00	10	550	17	69.00	20	600
5	25.02	8	295	18	10.30	1	585
6	16.86	4	200	19	34.93	10	540
7	14.38	2	375	20	46.59	15	250
8	9.60	2	52	21	44.88	15	290
9	24.35	9	100	22	54.12	16	510
10	27.50	8	300	23	56.63	17	590
11	17.08	4	412	24	22.13	6	100
12	37.00	11	400	25	21.15	5	400
13	41.95	12	500				

6-3 Multiple Regression

EXAMPLE 6-7

Fit the multiple linear regression model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

where Y = pull strength, x_1 = wire length, and x_2 = die height.

Solution. From the data in Table 6-5 we calculate

$$n = 25,\ \sum_{i=1}^{25} y_i = 725.82,\ \sum_{i=1}^{25} x_{i1} = 206,\ \sum_{i=1}^{25} x_{i2} = 8{,}294$$

$$\sum_{i=1}^{25} x_{i1}^2 = 2{,}396,\ \sum_{i=1}^{25} x_{i2}^2 = 3{,}531{,}848$$

$$\sum_{i=1}^{25} x_{i1}x_{i2} = 77{,}177,\ \sum_{i=1}^{25} x_{i1}y_i = 8{,}008.47,\ \sum_{i=1}^{25} x_{i2}y_i = 274{,}816.71$$

6-3 Multiple Regression

EXAMPLE 6-7

For the model $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$, the normal equations 6-43 are

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2} = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^{n} x_{i1}x_{i2} = \sum_{i=1}^{n} x_{i1}y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{i2} + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1}x_{i2} + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2}^2 = \sum_{i=1}^{n} x_{i2}y_i$$

Inserting the computed summations into the normal equations, we obtain

$$25\hat{\beta}_0 + 206\hat{\beta}_1 + 8{,}294\hat{\beta}_2 = 725.82$$

$$206\hat{\beta}_0 + 2{,}396\hat{\beta}_1 + 77{,}177\hat{\beta}_2 = 8{,}008.47$$

$$8{,}294\hat{\beta}_0 + 77{,}177\hat{\beta}_1 + 3{,}531{,}848\hat{\beta}_2 = 274{,}816.71$$

6-3 Multiple Regression

EXAMPLE 6-7

The solution to this set of equations is

$$\hat{\beta}_0 = 2.26379,\ \hat{\beta}_1 = 2.74427,\ \hat{\beta}_2 = 0.01253$$

Using these estimated model parameters, the fitted regression equation is

$$\hat{y} = 2.26379 + 2.74427x_1 + 0.01253x_2$$

Practical interpretation: This equation can be used to predict pull strength for pairs of values of the regressor variables wire length (x_1) and die height (x_2). This is essentially the same regression model given in equation 1-6, Section 1-3. Figure 1-13 shows a three-dimentional plot of the plane of predicted values $\hat{y}$ generated from this equation.

6-3 Multiple Regression

6-3.1 Estimation of Parameters in Multiple Regression

Variance Estimate

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n - p} = \frac{SS_E}{n - p} \qquad (6\text{-}45)$$

Adjusted Coefficient of Multiple Determination (R^2_{Adjusted})

The **adjusted coefficient of multiple determination** for a multiple regression model with k regressors is

$$R^2_{\text{Adjusted}} = 1 - \frac{SS_E/(n - p)}{SS_T/(n - 1)} = \frac{(n - 1)R^2 - k}{n - p} \qquad (6\text{-}46)$$

6-3 Multiple Regression

6-3.2 Inferences in Multiple Regression

Test for Significance of Regression

Testing for Significance of Regression in Multiple Regression

$$MS_R = \frac{SS_R}{k} \qquad MS_E = \frac{SS_E}{n-p}$$

Null hypothesis:	$H_0\colon \beta_1 = \beta_2 = \cdots = \beta_k = 0$	
Alternative hypothesis:	H_1: At least one $\beta_j \neq 0$	
Test statistic:	$F_0 = \dfrac{MS_R}{MS_E}$	(6-47)
P-value:	Probability above f_0 in the $F_{k,n-p}$ distribution	
Rejection criterion for a fixed-level test:	$f_0 > f_{\alpha,k,n-p}$	

6-3 Multiple Regression

6-3.2 Inferences in Multiple Regression

Inference on Individual Regression Coefficients

Inferences on the Model Parameters in Multiple Regression

1. The test for H_0: $\beta_j = \beta_{j,0}$ versus H_1: $\beta_j \neq \beta_{j,0}$ employs the **test statistic**

$$T_0 = \frac{\hat{\beta}_j - \beta_{j,0}}{se(\hat{\beta}_j)} \tag{6-48}$$

and the null hypothesis is rejected if $|t_0| > t_{\alpha/2,n-p}$. A P-value approach can also be used. One-sided alternative hypotheses can also be tested.

2. A $100(1-\alpha)\%$ CI for an individual regression coefficient is given by

$$\hat{\beta}_j - t_{\alpha/2,n-p}se(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + t_{\alpha/2,n-p}se(\hat{\beta}_j) \tag{6-49}$$

- This is called a **partial** or marginal test

6-3 Multiple Regression

6-3.2 Inferences in Multiple Regression

Confidence Intervals on the Mean Response and Prediction Intervals

Confidence Interval on the Mean Response in Multiple Regression

A $100(1-\alpha)\%$ CI on the mean response at the point $(x_1 = x_{10}, x_2 = x_{20}, \ldots, x_k = x_{k0})$ in a multiple regression model is given by

$$\hat{\mu}_{Y|x_{10},x_{20},\ldots,x_{k0}} - t_{\alpha/2,n-p}\, se(\hat{\mu}_{Y|x_{10},x_{20},\ldots,x_{k0}}) \le \mu_{Y|x_{10},x_{20},\ldots,x_{k0}} \le \hat{\mu}_{Y|x_{10},x_{20},\ldots,x_{k0}} + t_{\alpha/2,n-p}\, se(\hat{\mu}_{Y|x_{10},x_{20},\ldots,x_{k0}}) \tag{6-52}$$

where $\hat{\mu}_{Y|x_{10},x_{20},\ldots,x_{k0}}$ is computed from equation 6-51.

6-3 Multiple Regression

6-3.2 Inferences in Multiple Regression

Confidence Intervals on the Mean Response and Prediction Intervals

Prediction Interval on a Future Observation in Multiple Regression

A $100(1-\alpha)\%$ PI on a future observation at the point $(x_1 = x_{10}, x_2 = x_{20}, \ldots, x_k = x_{k0})$ in a multiple regression model is given by

$$\hat{y}_0 - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 + [se(\hat{\mu}_{Y|x_{10},x_{20},\ldots,x_{k0}})]^2} \le Y_0$$
$$\le \hat{y}_0 + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 + [se(\hat{\mu}_{Y|x_{10},x_{20},\ldots,x_{k0}})]^2} \qquad (6\text{-}54)$$

where $\hat{y}_0 = \hat{\mu}_{Y|x_{10},x_{20},\ldots,x_{k0}}$ is computed from equation 6-53.

6-3 Multiple Regression

6-3.2 Inferences in Multiple Regression

Confidence Intervals on the Mean Response and Prediction Intervals

Table 6-7 Minitab Output

Predicted Values for New Observations

New Obs	Fit	SE Fit	95.0% CI	95.0% PI
1	32.889	1.062	(30.687, 35.092)	(27.658, 38.121)
New Obs	Fit	SE Fit	95.0% CI	95.0% PI
2	16.236	0.929	(14.310, 18.161)	(11.115, 21.357)

Values of Predictors for New Observations

New Obs	Wire Ln	Die Ht
1	11.0	35.0
New Obs	Wire Ln	Die Ht
2	5.00	20.0

6-3 Multiple Regression

6-3.2 Inferences in Multiple Regression

A Test for the Significance of a Group of Regressors

$$H_0\colon \beta_{r+1} = \beta_{r+2} = \cdots = \beta_k = 0$$

$$H_1\colon \text{At least one of the } \beta\text{'s} \neq 0$$

we would use the test statistic

$$F_0 = \frac{[SS_E(RM) - SS_E(FM)]/(k-r)}{SS_E(FM)/(n-p)}$$

6-3 Multiple Regression

6-3.3 Checking Model Adequacy

Residual Analysis

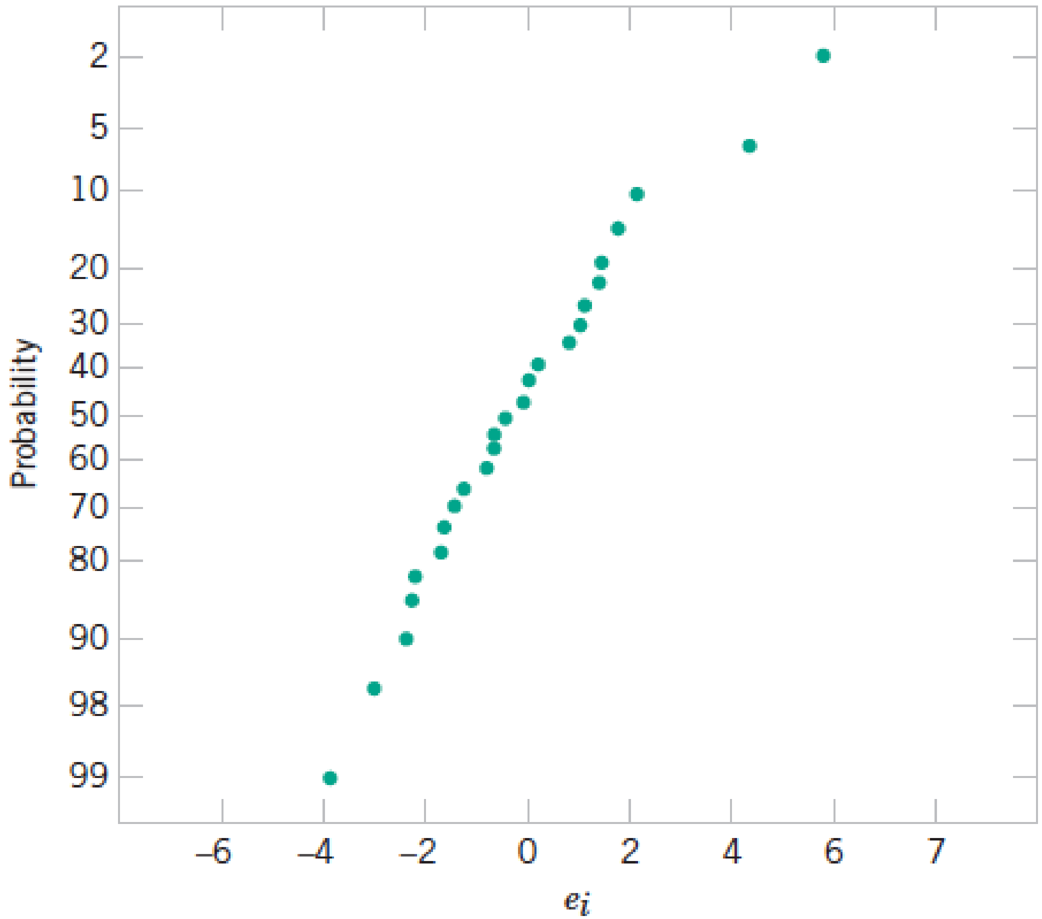

Figure 6-18 Normal probability plot of residuals for wire bond empirical model.

6-3 Multiple Regression

6-3.3 Checking Model Adequacy

Residual Analysis

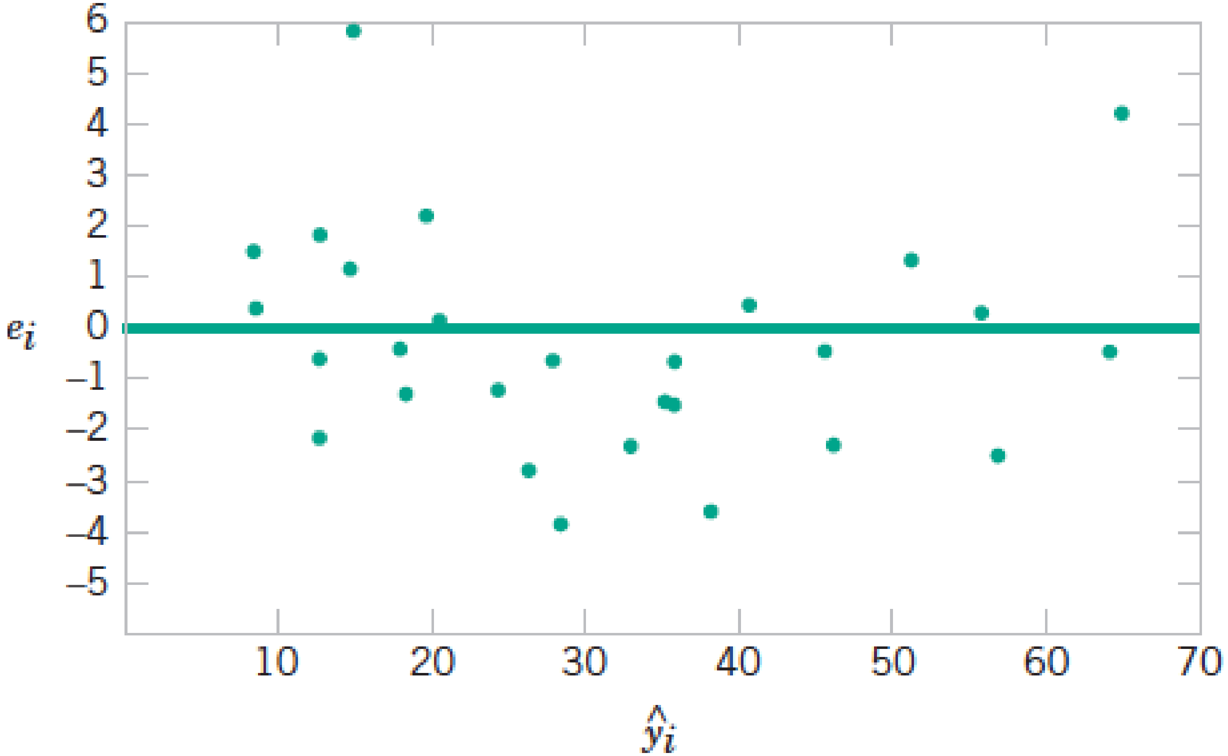

Figure 6-19 Plot of residuals against $\hat{y}$ for wire bond empirical model.

6-3 Multiple Regression

6-3.3 Checking Model Adequacy

Residual Analysis

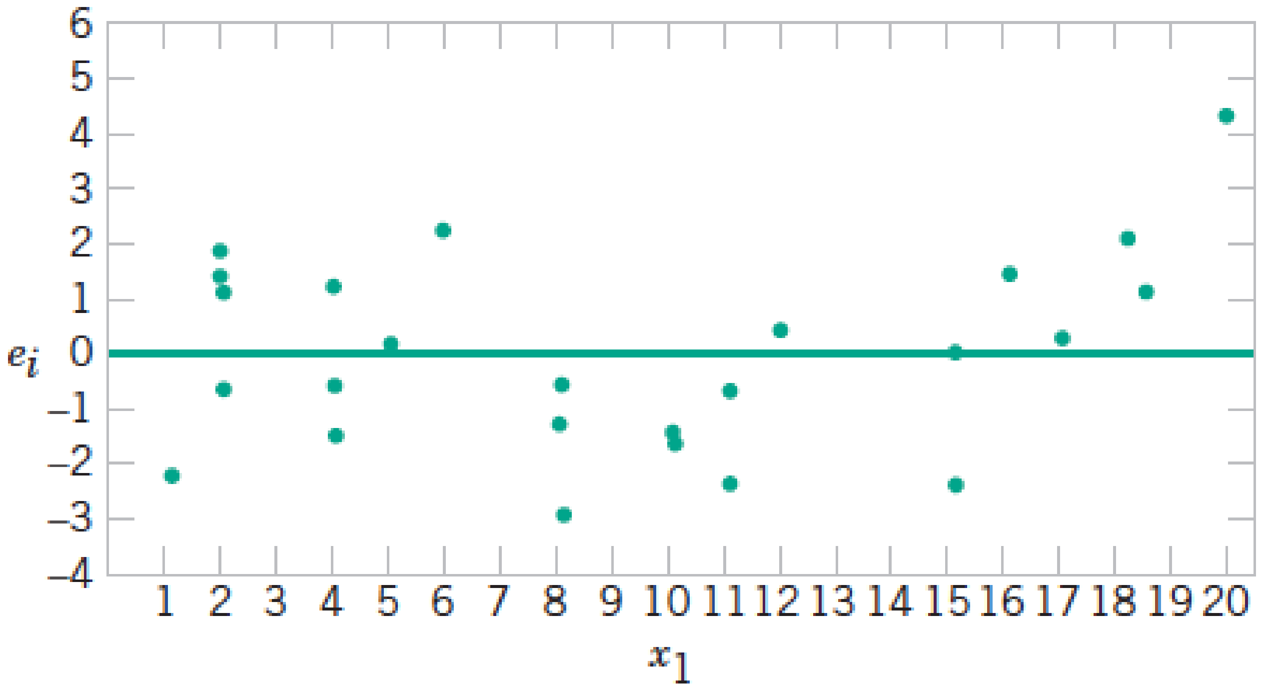

Figure 6-20 Plot of residuals against x_1 (wire length) for wire bond empirical model.

6-3 Multiple Regression

6-3.3 Checking Model Adequacy

Residual Analysis

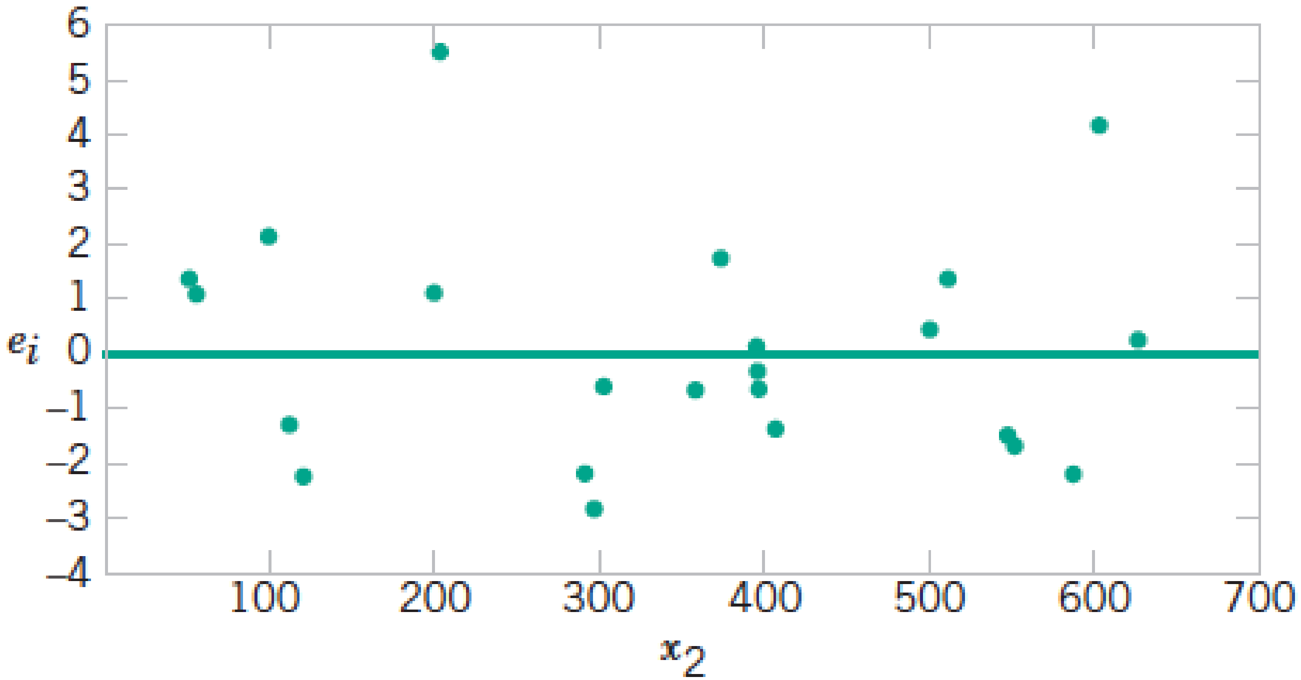

Figure 6-21 Plot of residuals against x_2 (die height) for wire bond empirical model.

6-3 Multiple Regression

6-3.3 Checking Model Adequacy

Residual Analysis

Studentized Residuals

The studentized residuals are defined as

$$r_i = \frac{e_i}{se(e_i)} = \frac{e_i}{\sqrt{\hat{\sigma}^2(1 - h_{ii})}}, i = 1, 2, \ldots, n \qquad (6\text{-}58)$$

6-3 Multiple Regression

6-3.3 Checking Model Adequacy

Influential Observations

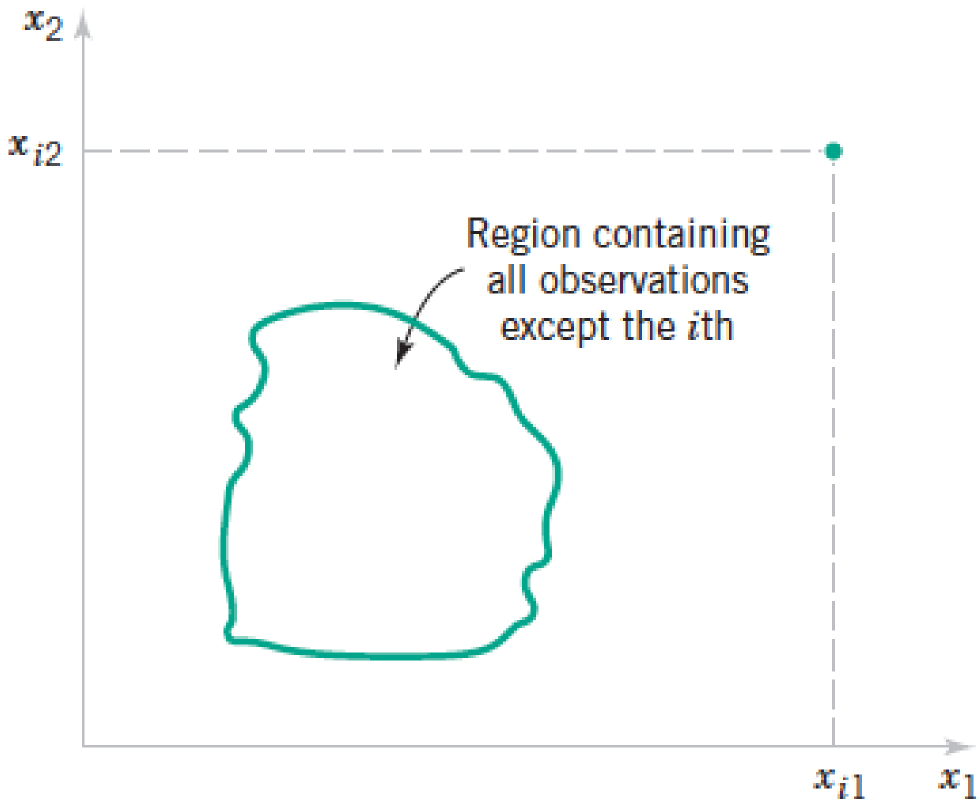

Figure 6-22 A point that is remote in x-space.

6-3 Multiple Regression

6-3.3 Checking Model Adequacy

Influential Observations

Cook's Distance Measure

$$D_i = \frac{r_i^2}{p} \frac{h_{ii}}{(1 - h_{ii})} \qquad i = 1, 2, \ldots, n \tag{6-59}$$

6-3 Multiple Regression

6-3.3 Checking Model Adequacy

Table 6-8 Influence Diagnostics for the Wire Bond Pull Strength Data

Observations i	h_{ii}	Cook's Distance Measure D_i	Observations i	h_{ii}	Cook's Distance Measure D_i
1	0.1573	0.035	14	0.1129	0.003
2	0.1116	0.012	15	0.0737	0.187
3	0.1419	0.060	16	0.0879	0.001
4	0.1019	0.021	17	0.2593	0.565
5	0.0418	0.024	18	0.2929	0.155
6	0.0749	0.007	19	0.0962	0.018
7	0.1181	0.036	20	0.1473	0.000
8	0.1561	0.020	21	0.1296	0.052
9	0.1280	0.160	22	0.1358	0.028
10	0.0413	0.001	23	0.1824	0.002
11	0.0925	0.013	24	0.1091	0.040
12	0.0526	0.001	25	0.0729	0.000
13	0.0820	0.001			

6-4 Other Aspects of Regression

6-4.1 Polynomial Models

In Section 6-1 we observed that models with polynomial terms in the regressors, such as the second-order model

$$Y = \beta_0 + \beta_1 x_1 + \beta_{11} x_1^2 + \epsilon$$

are really linear regression models and can be fit and analyzed using the methods discussed in Section 6-3. Polynomial models arise frequently in engineering and the sciences, and this contributes greatly to the widespread use of linear regression in these fields.

Table 6-9 The Acetylene Data

Observation	Yield, Y	Temp., T	Ratio, R	Observation	Yield, Y	Temp., T	Ratio, R
1	49.0	1300	7.5	9	34.5	1200	11.0
2	50.2	1300	9.0	10	35.0	1200	13.5
3	50.5	1300	11.0	11	38.0	1200	17.0
4	48.5	1300	13.5	12	38.5	1200	23.0
5	47.5	1300	17.0	13	15.0	1100	5.3
6	44.5	1300	23.0	14	17.0	1100	7.5
7	28.0	1200	5.3	15	20.5	1100	11.0
8	31.5	1200	7.5	16	29.5	1100	17.0

6-4 Other Aspects of Regression

6-4.1 Polynomial Models

The second-order model in two regressors is

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

$$\begin{aligned} Y = {} & \beta_0 + \beta_1 (T - 1212.5) + \beta_2 (R - 12.444) \\ & + \beta_{12} (T - 1212.5)(R - 12.444) + \beta_{11} (T - 1212.5)^2 + \beta_{22} (R - 12.444)^2 + \epsilon \end{aligned}$$

6-4 Other Aspects of Regression

6-4.1 Polynomial Models

The regression equation is

Yield = 36.1 + 0.134 Temp + 0.351 Ratio

Predictor	Coef	SE Coef	T	P	VIF
Constant	36.1063	0.9060	39.85	0.000	
Temp	0.13396	0.01191	11.25	0.000	1.1
Ratio	0.3511	0.1696	2.07	0.059	1.1

S = 3.624 R-Sq = 92.0% R-Sq(adj) = 90.7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	1952.98	976.49	74.35	0.000
Residual Error	13	170.73	13.13		
Total	15	2123.71			

6-4 Other Aspects of Regression

6-4.1 Polynomial Models

The regression equation is

Yield = 36.1 + 0.134 Temp + 0.351 Ratio

Predictor	Coef	SE Coef	T	P	VIF
Constant	36.1063	0.9060	39.85	0.000	
Temp	0.13396	0.01191	11.25	0.000	1.1
Ratio	0.3511	0.1696	2.07	0.059	1.1

S = 3.624 R-Sq = 92.0% R-Sq(adj) = 90.7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	1952.98	976.49	74.35	0.000
Residual Error	13	170.73	13.13		
Total	15	2123.71			

6-4 Other Aspects of Regression

6-4.1 Polynomial Models

$$H_0\colon \beta_{r+1} = \beta_{r+2} = \cdots = \beta_k = 0$$

$$H_1\colon \text{At least one of the } \beta\text{'s} \neq 0$$

The test statistic for the above hypotheses was originally given in equation 6-56, repeated below for convenience:

$$F_0 = \frac{[SS_E(RM) - SS_E(FM)]/(k - r)}{SS_E(FM)/(n - p)}$$

6-4 Other Aspects of Regression

6-4.2 Categorical Regressors

- Many problems may involve **qualitative** or **categorical** variables.
- The usual method for the different levels of a qualitative variable is to use **indicator** variables.
- For example, to introduce the effect of two different operators into a regression model, we could define an indicator variable as follows:

$$x_3 = \begin{cases} 0 \text{ if the car has an automatic transmission} \\ 1 \text{ if the car has a manual transmission} \end{cases}$$

6-4 Other Aspects of Regression

6-4.2 Categorical Regressors

The regression equation is

Quality = 89.8 + 1.82 Foam − 3.38 Residue − 3.41 Region

Predictor	Coef	SE Coef	T	P
Constant	89.806	2.990	30.03	0.000
Foam	1.8192	0.3260	5.58	0.000
Residue	−3.3795	0.6858	−4.93	0.000
Region	−3.4062	0.9194	−3.70	0.001

S = 2.21643 R-Sq = 77.6% R-Sq (adj) = 74.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	339.75	113.25	23.05	0.000
Residual Error	20	98.25	4.91		
Total	23	438.00			

6-4 Other Aspects of Regression

6-4.2 Categorical Regressors

The regression equation is

$$\text{Quality} = 88.3 + 1.98\ \text{Foam} - 3.22\ \text{Residue} - 1.71\ \text{Region} - 0.642\ \text{F} \times \text{R} + 0.43\ \text{R} \times \text{Res}$$

Predictor	Coef	SE Coef	T	P
Constant	88.257	4.840	18.24	0.000
Foam	1.9825	0.4292	4.62	0.000
Residue	−3.2153	0.9525	−3.38	0.003
Region	−1.707	6.572	−0.26	0.798
F × R	−0.6419	0.9434	−0.68	0.505
R × Res	0.430	1.894	0.23	0.823

S = 2.30499 R-Sq = 78.2% R-Sq (adj) = 72.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	342.366	68.473	12.89	0.000
Residual Error	18	95.634	5.313		
Total	23	438.000			

6-4 Other Aspects of Regression

6-4.3 Variable Selection Procedures

Best Subsets Regressions

Table 6-10 Minitab Best Subsets Regression for Shampoo Data

Response is Quality

Vars	R-Sq	R-Sq(adj)	Mallows C-p	S	Foam	Scent	Color	Residue	Region
1	26.2	22.9	46.4	3.8321	X				
1	25.7	22.3	46.9	3.8455					X
1	23.9	20.5	48.5	3.8915				X	
1	6.3	2.1	64.3	4.3184		X			
1	3.8	0.0	66.7	4.3773			X		
2	62.2	58.6	16.1	2.8088	X			X	
2	50.3	45.6	26.7	3.2185	X				X
2	42.6	37.2	33.6	3.4589				X	X
2	40.9	35.3	35.2	3.5098	X		X		
2	32.6	26.2	42.7	3.7486	X	X			
3	77.6	74.2	4.2	2.2164	X			X	X
3	63.1	57.6	17.2	2.8411	X		X	X	
3	62.5	56.9	17.7	2.8641	X	X		X	
3	52.9	45.9	26.4	3.2107	X		X		X
3	51.8	44.6	27.4	3.2491	X	X			X
4	79.9	75.7	4.1	2.1532	X	X		X	X
4	78.6	74.1	5.3	2.2205	X		X	X	X
4	64.8	57.4	17.7	2.8487	X	X	X	X	
4	53.0	43.1	28.3	3.2907	X	X	X		X
4	51.4	41.2	29.7	3.3460		X	X	X	X
5	80.0	74.5	6.0	2.2056	X	X	X	X	X

6-4.3 Variable Selection Procedures

Backward Elimination

Table 6-11 Stepwise Regression Backward Elimination for Shampoo Data: Quality versus Foam, Scent, Color, Residue, Region

Backward elimination. Alpha-to-Remove: 0.1

Response is Quality on 5 predictors, with N = 24

Step	1	2	3
Constant	86.33	86.14	89.81
Foam	1.82	1.87	1.82
T-Value	5.07	5.86	5.58
P-Value	0.000	0.000	0.000
Scent	1.03	1.18	
T-Value	1.12	1.48	
P-Value	0.277	0.155	
Color	0.23		
T-Value	0.33		
P-Value	0.746		
Residue	−4.00	−3.93	−3.38
T-Value	−4.93	−5.15	−4.93
P-Value	0.000	0.000	0.000
Region	−3.86	−3.71	−3.41
T-Value	−3.70	−4.05	−3.70
P-Value	0.002	0.001	0.001
S	2.21	2.15	2.22
R-Sq	80.01	79.89	77.57
R-Sq (adj)	74.45	75.65	74.20
Mallows C-p	6.0	4.1	4.2

6-4 Other Aspects of Regression

6-4.3 Variable Selection Procedures

Forward Selection

Table 6-12 Stepwise Regression Forward Selection for Shampoo Data: Quality versus Foam, Scent, Color, Residue, Region

Forward selection. Alpha-to-Enter: 0.25

Response is Quality on 5 predictors, with N = 24

Step	1	2	3	4
Constant	76.00	89.45	89.81	86.14
Foam	1.54	1.90	1.82	1.87
T-Value	2.80	4.61	5.58	5.86
P-Value	0.010	0.000	0.000	0.000
Residue		−3.82	−3.38	−3.93
T-Value		−4.47	−4.93	−5.15
P-Value		0.000	0.000	0.000
Region			−3.41	−3.71
T-Value			−3.70	−4.05
P-Value			0.001	0.001
Scent				1.18
T-Value				1.48
P-Value				0.155
S	3.83	2.81	2.22	2.15
R-Sq	26.24	62.17	77.57	79.89
R-Sq (adj)	22.89	58.57	74.20	75.65
Mallows C-p	46.4	16.1	4.2	4.1

6-4 Other Aspects of Regression

6-4.3 Variable Selection Procedures

Stepwise Regression

Table 6-13 Stepwise Regression Combined Forward and Backward Elimination: Quality versus Foam, Scent, Color, Residue, Region

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is Quality on 5 predictors, with N = 24

Step	1	2	3
Constant	76.00	89.45	89.81
Foam	1.54	1.90	1.82
T-Value	2.80	4.61	5.58
P-Value	0.010	0.000	0.000
Residue		−3.82	−3.38
T-Value		−4.47	−4.93
P-Value		0.000	0.000
Region			−3.41
T-Value			−3.70
P-Value			0.001
S	3.83	2.81	2.22
R-Sq	26.24	62.17	77.57
R-Sq (adj)	22.89	58.57	74.20
Mallows C-p	46.4	16.1	4.2

IMPORTANT TERMS AND CONCEPTS

Adjusted R^2
All possible regressions
Analysis of variance (ANOVA)
Backward elimination
Coefficient of determination, R^2
Confidence interval on mean response
Confidence interval on regression coefficients
Contour plot
Cook's distance measure, D_i
C_p statistic
Empirical model
Forward selection
Indicator variables
Influential observations
Interaction
Intercept
Least squares normal equations
Mechanistic model
Method of least squares
Model
Model adequacy
Multicollinearity
Multiple regression
Outliers
Polynomial regression
Population correlation coefficient, ρ
Prediction interval
Regression analysis
Regression coefficients
Regression model
Regression sum of squares
Regressor variable
Residual analysis
Residual sum of squares
Residuals
Response variable
Sample correlation coefficient, r
Significance of regression
Simple linear regression
Standard errors of model coefficients
Standardized residuals
Stepwise regression
Studentized residuals
t-tests on regression coefficients
Unbiased estimators
Variance inflation factor