Chapter 3
Multiple Linear Regression

3.1 Multiple Regression Models

- Suppose that the yield in pounds of conversion in a chemical process depends on temperature and the catalyst concentration. A **multiple regression model** that might describe this relationship is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \tag{3.1}$$

- This is a multiple linear regression model in two variables.

3.1 Multiple Regression Models

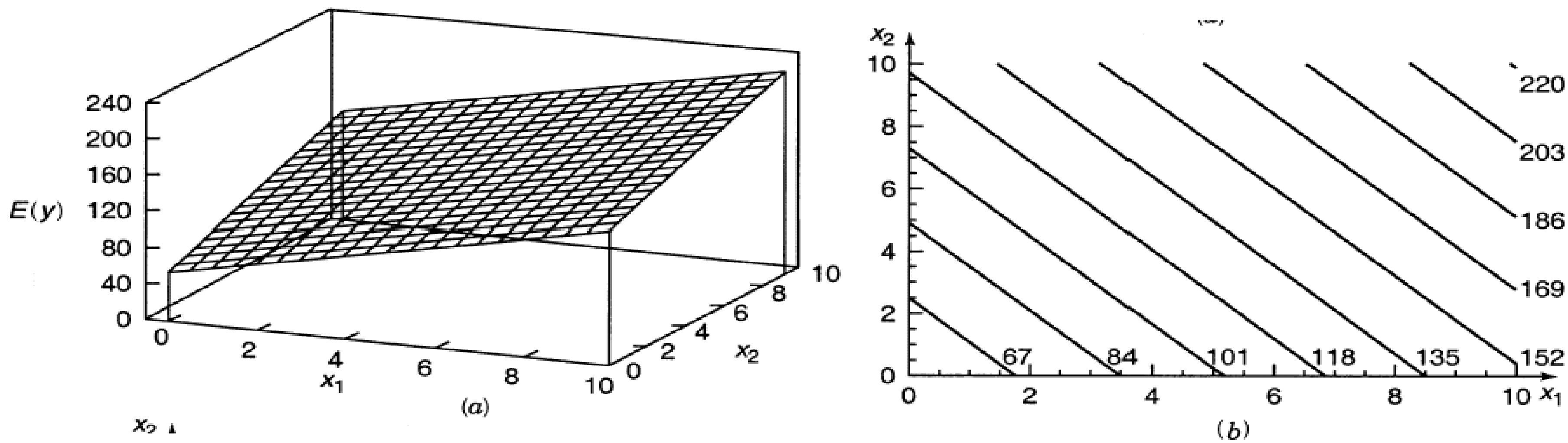


Figure 3.1 (a) The regression plane for the model E(y)= $50+10x_1+7x_2$. (b) The contour plot.

3.1 Multiple Regression Models

In general, the multiple linear regression model with *k* regressors is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

3.1 Multiple Regression Models

Models that are more complex in structure than Eq. (3.2) may often still be analyzed by multiple linear regression techniques. For example, consider the cubic polynomial model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon \tag{3.3}$$

If we let $x_1 = x$, $x_2 = x^2$, and $x_3 = x^3$, then Eq. (3.3) can be written as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \tag{3.4}$$

3.1 Multiple Regression Models

Linear regression models may also contain **interaction** effects:

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

If we let $x_3 = x_1 x_2$ and $\beta_3 = \beta_{12}$, then the model can be written in the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

Example of interaction between two variables
Time to run a maze (min)

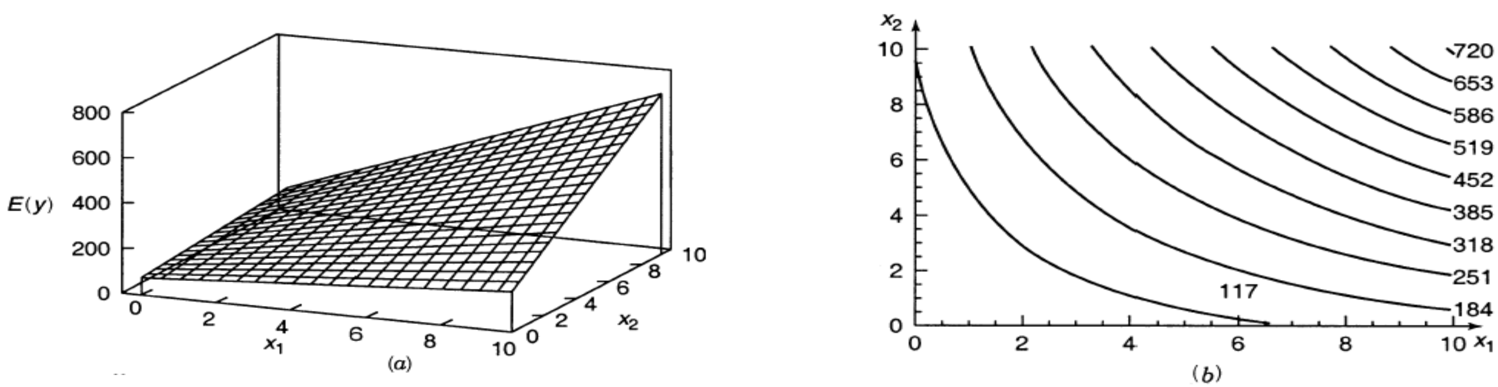
DOSE OF DRUG	LOW	HIGH
AGE <65 yo	1.2	0.8
>= 65 yo	1.0	1.3

3.1 Multiple Regression Models

Figure 3.2a shows the three-dimensional plot of the regression model

$$y = 50 + 10x_1 + 7x_2 + 5x_1x_2$$

and Figure 3.2b the corresponding two-dimensional contour plot. Notice that, although this model is a linear regression model, the shape of the surface that is generated by the model is not linear. In general, **any regression model that is linear in the parameters (the β's) is a linear regression model, regardless of the shape of the surface that it generates.**



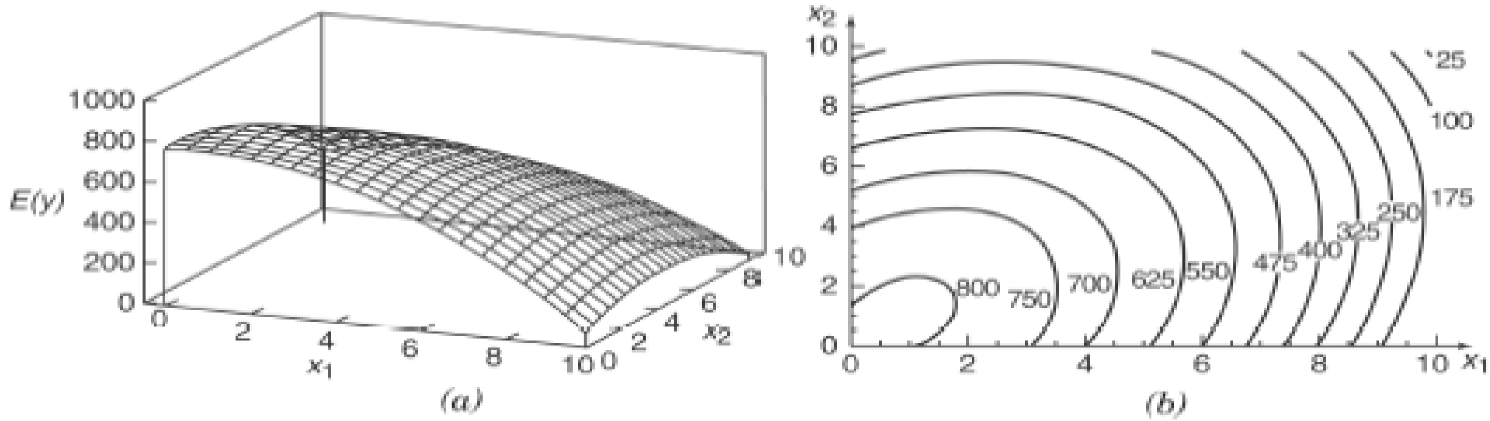


Figure 3.3 (*a*) Three-dimensional plot of the regression model $E(y) = 800 + 10x_1 + 7x_2 - 8.5x_1^2 - 5x_2^2 + 4x_1x_2$. (*b*) The contour plot.

3.2 Estimation of the Model Parameters

3.2.1 Least Squares Estimation of the Regression Coefficients

Notation

$$E(\varepsilon) = 0, \operatorname{Var}(\varepsilon) = \sigma^2,$$

n – number of observations available

k – number of regressor variables

y – response or dependent variable

x_{ij} – ith observation on regressor j.

3.2.1 Least Squares Estimation of Regression Coefficients

TABLE 3.1 Data for Multiple Linear Regression

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

3.2.1 Least Squares Estimation of the Regression Coefficients

The sample regression model can be written as

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i \\ &= \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i, \quad i = 1, 2, \ldots, n \end{aligned}$$

3.2.1 Least Squares Estimation of the Regression Coefficients

The least squares function is

$$S(\beta_0, \beta_1, \ldots, \beta_k) = \sum_{i=1}^{n} \varepsilon_i^2$$

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

The function S must be minimized with respect to the coefficients.

3.2.1 Least Squares Estimation of the Regression Coefficients

The least squares estimates of the coefficients must satisfy

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

and

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

3.2.1 Least Squares Estimation of the Regression Coefficients

Simplifying, we obtain the least squares normal equations:

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

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

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

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

The ordinary least squares estimators are the solutions to the normal equations.

3.2.1 Least Squares Estimation of the Regression Coefficients

Matrix notation is typically used:

Vector, matrix, add, multiply, transpose, inverse, diagonal

Let

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

3.2.1 Least Squares Estimation of the Regression Coefficients

We wish to find the vector of least-squares estimators, $\hat{\boldsymbol{\beta}}$, that minimizes

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Note that $S(\boldsymbol{\beta})$ may be expressed as

$$\begin{aligned} S(\boldsymbol{\beta}) &= \mathbf{y}'\mathbf{y} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \end{aligned}$$

since $\boldsymbol{\beta}'\mathbf{X}'\mathbf{y}$ is a 1×1 matrix, or a scalar, and its transpose $(\boldsymbol{\beta}'\mathbf{X}'\mathbf{y})' = \mathbf{y}'\mathbf{X}\boldsymbol{\beta}$ is the same scalar. The least-squares estimators must satisfy

3.2.1 Least Squares Estimation of the Regression Coefficients

See page 579 of text for derivatives involving matrices

$$\left.\frac{\partial S}{\partial \boldsymbol{\beta}}\right|_{\hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0}$$

which simplifies to

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \tag{3.12}$$

These are the **least-squares normal equations**. The solution is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad \boxed{=\mathrm{sum}(c_i y_i)}$$

3.2.1 Least Squares Estimation of the Regression Coefficients

The fitted regression model corresponding to the levels of the regressor variables $\mathbf{x}' = [1, x_1, x_2, \ldots, x_k]$ is

$$\hat{y} = \mathbf{x}'\hat{\boldsymbol{\beta}} = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j x_j$$

The vector of fitted values $\hat{y}_i$ corresponding to the observed values y_i is

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y} \qquad (3.14)$$

The $n \times n$ matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is usually called the **hat matrix.**

H x H = H ➔ H is idempotent

3.2.1 Least Squares Estimation of the Regression Coefficients

The *n* residuals can be written in matrix form as

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$$

There will be some situations where an alternative form will prove useful

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{y} - \mathbf{H}\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y}$$

see matrix_simple.pdf for matrix approach to simple linear model

Example 3-1. The Delivery Time Data

The model of interest is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

TABLE 3.2 Delivery Time Data for Example 3.1

Observation Number	Delivery Time (Minutes) y	Number of Cases x_1	Distance (Feet) x_2
1	16.68	7	560
2	11.50	3	220
3	12.03	3	340
4	14.88	4	80
5	13.75	6	150
6	18.11	7	330
7	8.00	2	110
8	17.83	7	210
9	79.24	30	1460
10	21.50	5	605
11	40.33	16	688
12	21.00	10	215
13	13.50	4	255
14	19.75	6	462
15	24.00	9	448
16	29.00	10	776
17	15.35	6	200
18	19.00	7	132
19	9.50	3	36
20	35.10	17	770
21	17.90	10	140
22	52.32	26	810
23	18.75	9	450
24	19.83	8	635
25	10.75	4	150

Example 3-1. The Delivery Time Data

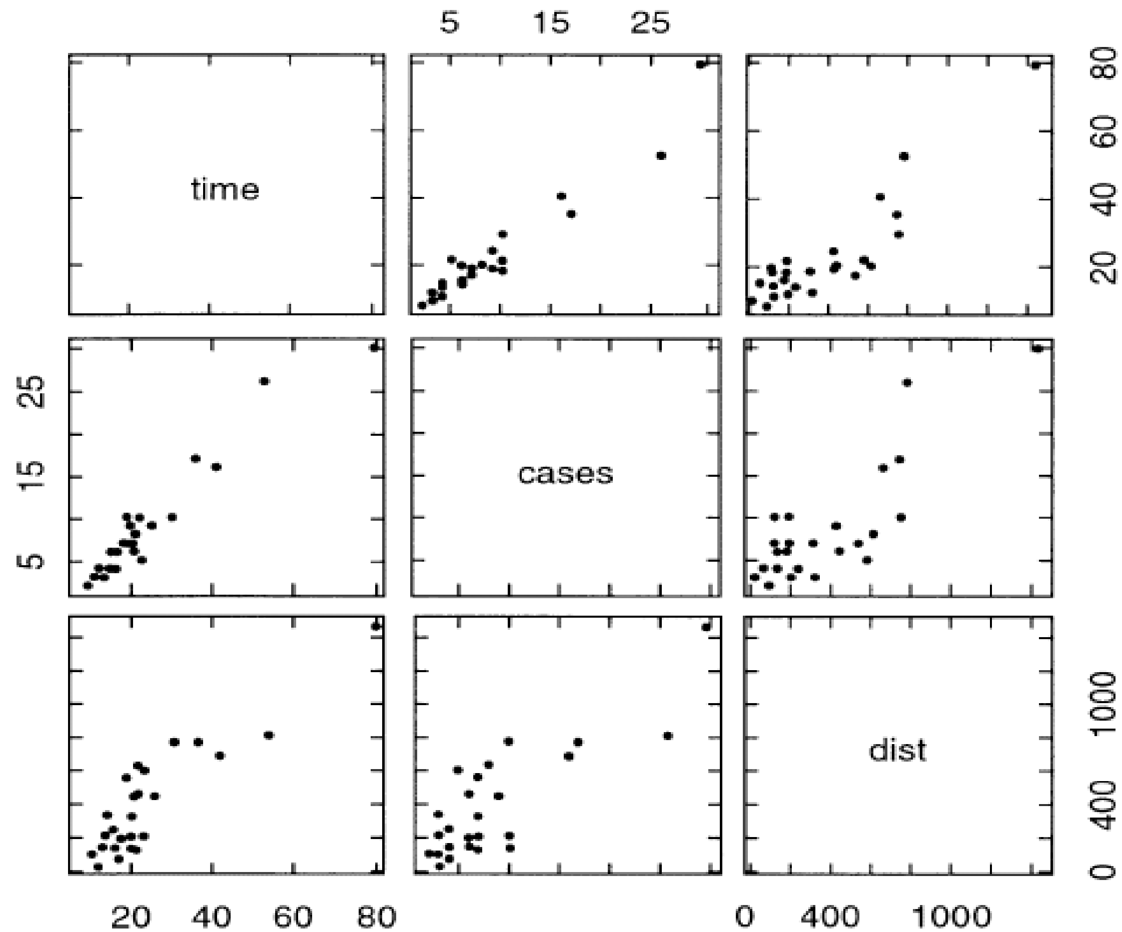


Figure 3.4
Scatterplot matrix for the delivery time data from Example 3.1.

Example 3-1 The Delivery Time Data

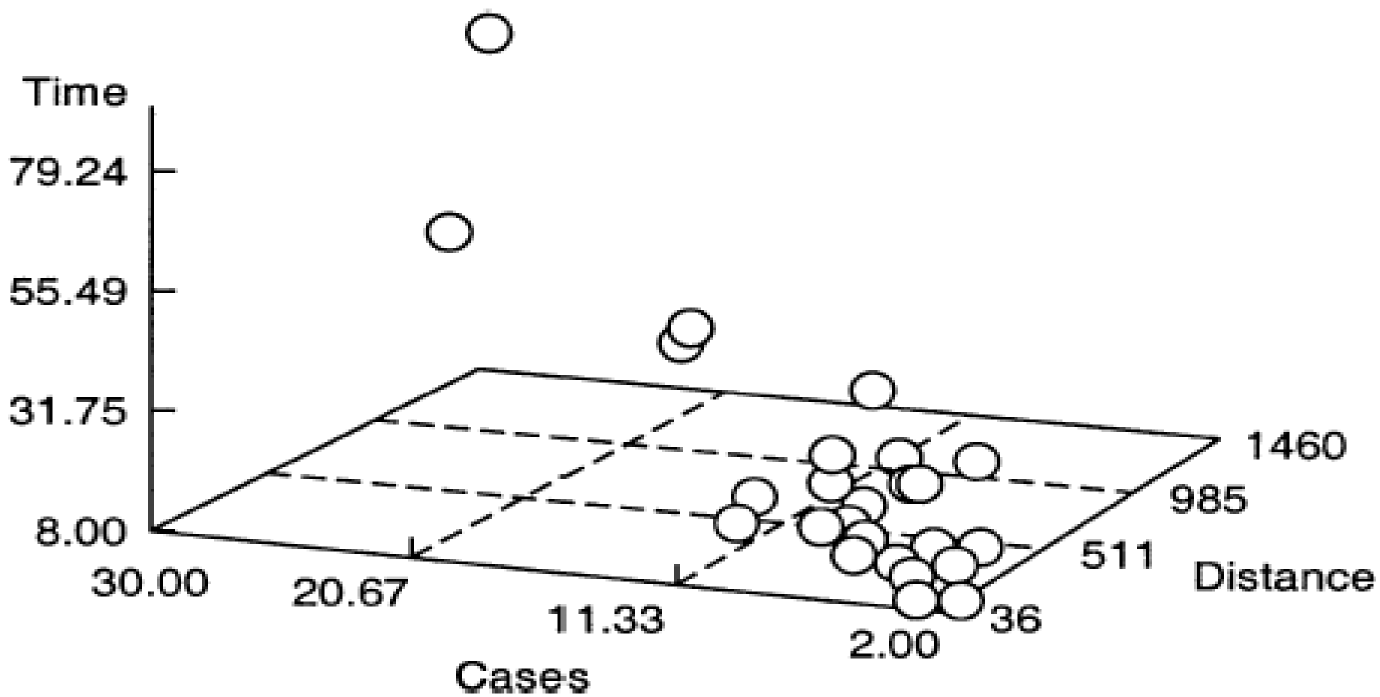


Figure 3.5 Three-dimensional scatterplot of the delivery time data from Example 3.1.

Example 3-1 The Delivery Time Data

See matrix_simple.pdf in Lecture 2 folder in Sakai for another example.

$$\mathbf{X} = \begin{bmatrix} 1 & 7 & 560 \\ 1 & 3 & 220 \\ 1 & 3 & 340 \\ 1 & 4 & 80 \\ 1 & 6 & 150 \\ 1 & 7 & 330 \\ 1 & 2 & 110 \\ 1 & 7 & 210 \\ 1 & 30 & 1460 \\ 1 & 5 & 605 \\ 1 & 16 & 688 \\ 1 & 10 & 215 \\ 1 & 4 & 255 \\ 1 & 6 & 462 \\ 1 & 9 & 448 \\ 1 & 10 & 776 \\ 1 & 6 & 200 \\ 1 & 7 & 132 \\ 1 & 3 & 36 \\ 1 & 17 & 770 \\ 1 & 10 & 140 \\ 1 & 26 & 810 \\ 1 & 9 & 450 \\ 1 & 8 & 635 \\ 1 & 4 & 150 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} 16.68 \\ 11.50 \\ 12.03 \\ 14.88 \\ 13.75 \\ 18.11 \\ 8.00 \\ 17.83 \\ 79.24 \\ 21.50 \\ 40.33 \\ 21.00 \\ 13.50 \\ 19.75 \\ 24.00 \\ 29.00 \\ 15.35 \\ 19.00 \\ 9.50 \\ 35.10 \\ 17.90 \\ 52.32 \\ 18.75 \\ 19.83 \\ 10.75 \end{bmatrix}$$

Example 3-1 The Delivery Time Data

The $\mathbf{X}'\mathbf{X}$ matrix is

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 7 & 3 & \cdots & 4 \\ 560 & 220 & \cdots & 150 \end{bmatrix} \begin{bmatrix} 1 & 7 & 560 \\ 1 & 3 & 220 \\ \vdots & \vdots & \vdots \\ 1 & 4 & 150 \end{bmatrix}$$

$$= \begin{bmatrix} 25 & 219 & 10{,}232 \\ 219 & 3{,}055 & 133{,}899 \\ 10{,}232 & 133{,}899 & 6{,}725{,}688 \end{bmatrix}$$

the $\mathbf{X}'\mathbf{y}$ vector is

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 7 & 3 & \cdots & 4 \\ 560 & 220 & \cdots & 150 \end{bmatrix} \begin{bmatrix} 16.68 \\ 11.50 \\ \vdots \\ 10.75 \end{bmatrix} = \begin{bmatrix} 559.60 \\ 7{,}375.44 \\ 337{,}072.00 \end{bmatrix}$$

Example 3-1 The Delivery Time Data

The least-squares estimator of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 25 & 219 & 10{,}232 \\ 219 & 3{,}055 & 133{,}899 \\ 10{,}232 & 133{,}899 & 6{,}725{,}688 \end{bmatrix}^{-1} \begin{bmatrix} 559.60 \\ 7{,}375.44 \\ 337{,}072.00 \end{bmatrix}$$

$$= \begin{bmatrix} .11321518 & -.00444859 & -.00008367 \\ -.00444859 & .00274378 & -.00004786 \\ -.00008367 & -.00004786 & .00000123 \end{bmatrix} \begin{bmatrix} 559.60 \\ 7{,}375.44 \\ 337{,}072.00 \end{bmatrix}$$

$$= \begin{bmatrix} 2.34123115 \\ 1.61590712 \\ 0.01438483 \end{bmatrix} \qquad \hat{y} = 2.34123 + 1.61591x_1 + 0.01438x_2$$

TABLE 3.3 Observations, Fitted Values, and Residuals for Example 3.1

Observation Number	y_i	$\hat{y}_i$	$e_i = y_i - \bar{y}_i$
1	16.68	21.7081	−5.0281
2	11.50	10.3536	1.1464
3	12.03	12.0798	−0.0498
4	14.88	9.9556	4.9244
5	13.75	14.1944	−0.4444
6	18.11	18.3996	−0.2896
7	8.00	7.1554	0.8446
8	17.83	16.6734	1.1566
9	79.24	71.8203	7.4197
10	21.50	19.1236	2.3764
11	40.33	38.0925	2.2375
12	21.00	21.5930	−0.5930
13	13.50	12.4730	1.0270
14	19.75	18.6825	1.0675
15	24.00	23.3288	0.6712
16	29.00	29.6629	−0.6629
17	15.35	14.9136	0.4364
18	19.00	15.5514	3.4486
19	9.50	7.7068	1.7932
20	35.10	40.8880	−5.7880
21	17.90	20.5142	−2.6142
22	52.32	56.0065	−3.6865
23	18.75	23.3576	−4.6076
24	19.83	24.4028	−4.5728
25	10.75	10.9626	−0.2126

MINITAB Output

TABLE 3.4 MINITAB Output for Soft Drink Time Data

Regression Analysis: Time versus Cases, Distance

The regression equation is
Time = 2.34 + 1.62 cases + 0.0144 Distance

Predictor	Coef	SE Coef	T	P
Constant	2.341	1.097	2.13	0.044
Cases	1.6159	0.1707	9.46	0.000
Distance	0.014385	0.003613	3.98	0.001

S = 3.25947   R- Sq = 96.0%   R- Sq(adj) = 95.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	5550.8	2775.4	261.24	0.000
Residual Error	22	233.7	10.6		
Total	24	5784.5			

Source	DF	Seq SS
Cases	1	5382.4
Distance	1	168.4

3.2.3 Properties of Least-Squares Estimators

- Statistical Properties

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

$$Cov(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

Var(Ay)=AVar(y)A' where y is a vector of random variables

- Variances/Covariances

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_j) = \sigma^2 \mathrm{C}_{jj}$$

$$\mathrm{Cov}(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 \mathrm{C}_{ij}$$

3.2.4 Estimation of σ^2

(y-yhat)'(y-yhat)=y'y-yhat'y-y'yhat+yhat'yhat

- The residual sum of squares can be shown to be:

=y'y-(Xb)'y-y'Xb+(Xb)'(Xb)

=y'y -b'X'y - b'X'y + b'X'X(X'X)$^{-1}$ X'y

$$SS_{Res} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$$

=y'y -b'X'y

- The residual mean square for the model with *p* parameters is:

$$MS_{Res} = \frac{SS_{Res}}{n-p} = \hat{\sigma}^2$$

3.2.4 Estimation of σ^2

- Recall that the estimator of σ^2 is **model dependent** - that is, change the form of the model and the estimate of σ^2 will invariably change.
  - Note that the variance estimate is a function of the errors; "unexplained noise about the fitted regression line"

see matrix_simple.pdf in Sakai.
see deliverytimematrix.sas and deliverymatrix.pdf in Sakai.

Example 3.2 Delivery Time Data

$$\mathbf{y}'\mathbf{y} = \sum_{i=1}^{25} y_i^2 = 18{,}310.6290$$

$$\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = \begin{bmatrix} 2.34123115 & 1.61590721 & 0.01438483 \end{bmatrix} \begin{bmatrix} 559.60 \\ 7{,}375.44 \\ 337{,}072.00 \end{bmatrix}$$

$$= 18{,}076.90304$$

Example 3.2 Delivery Time Data

$$SS_{\text{Res}} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$$
$$= 18{,}310.6290 - 18{,}076.9030 = 233.7260$$

$$\hat{\sigma}^2 = \frac{SS_{\text{Res}}}{n-p} = \frac{233.7260}{25-3} = 10.6239$$

3.3 Hypothesis Testing in Multiple Linear Regression

Once we have estimated the parameters in the model, we face two immediate questions:

1. What is the overall adequacy of the model?

2. Which specific regressors seem important?

3.3 Hypothesis Testing in Multiple Linear Regression

This section considers four cases:

- Test for Significance of Regression (sometimes called the global test of model adequacy)
- Tests on Individual Regression Coefficients (or groups of coefficients)
- Special Case of Hypothesis Testing with Orthogonal Columns in **X**
- Testing the General Linear Hypothesis

3.3.1 Test for Significance of Regression

- The test for significance is a test to determine if there is a linear relationship between the response and **any** of the regressor variables
- The hypotheses are

$$H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0$$
$$H_1: \beta_j \neq 0 \text{ for at least one } j$$

3.3.1 Test for Significance of Regression

- As in Chapter 2, the total sum of squares can be partitioned in two parts:

$$SS_T = SS_R + SS_{Res}$$

- This leads to an ANOVA procedure with the test (*F*) statistic

$$F_0 = \frac{SS_R / k}{SS_{Res} / (n - k - 1)} = \frac{MS_R}{MS_{Res}}$$

3.3.1 Test for Significance of Regression

- The standard ANOVA is conducted with

$$SS_R = SS_T - SS_{Res}$$

$$SS_R = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \qquad SS_{Res} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$$

$$SS_T = \mathbf{y}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

3.3.1 Test for Significance of Regression

ANOVA Table:

TABLE 3.4 Analysis of Variance for Significance of Regression in Multiple Regression

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Regression	SS_R	k	MS_R	MS_R/MS_{Res}
Residual	SS_{Res}	$n-k-1$	MS_{Res}	
Total	SS_T	$n-1$		

Reject H_0 if $F_0 > F_{\alpha,k,n-k-1}$

$$F_0 = \frac{SS_R/k}{SS_{Res}/(n-k-1)} = \frac{MS_R}{MS_{Res}}$$

follows the $F_{k,\,n-k-1}$ distribution. Appendix C.3 shows that

$$E(MS_{Res}) = \sigma^2$$

$$E(MS_R) = \sigma^2 + \frac{\boldsymbol{\beta}^{*\prime}\mathbf{X}_c'\mathbf{X}_c\boldsymbol{\beta}^*}{k\sigma^2}$$

where $\boldsymbol{\beta}^* = (\beta_1, \beta_2, \ldots, \beta_k)'$ and $\mathbf{X}_c$ is the "centered" model matrix given by

$$\mathbf{X}_c = \begin{bmatrix} x_{11}-\bar{x}_1 & x_{12}-\bar{x}_2 & \cdots & x_{1k}-\bar{x}_k \\ x_{21}-\bar{x}_1 & x_{22}-\bar{x}_2 & \cdots & x_{2k}-\bar{x}_k \\ \vdots & \vdots & & \vdots \\ x_{i1}-\bar{x}_1 & x_{i2}-\bar{x}_2 & \cdots & x_{ik}-\bar{x}_k \\ \vdots & \vdots & & \vdots \\ x_{n1}-\bar{x}_1 & x_{n2}-\bar{x}_2 & \cdots & x_{nk}-\bar{x}_k \end{bmatrix}$$

These expected mean squares indicate that if the observed value of F_0 is large, then it is likely that at least one $\beta_j \neq 0$. Appendix C.3 also shows that if at least one $\beta_j \neq 0$, then F_0 follows a noncentral F distribution with k and $n-k-1$ degrees of freedom and a noncentrality parameter of

$$\lambda = \frac{\boldsymbol{\beta}^{*\prime}\mathbf{X}_c'\mathbf{X}_c\boldsymbol{\beta}^*}{\sigma^2}$$

This noncentrality parameter also indicates that the observed value of F_0 should be large if at least one $\beta_j \neq 0$. Therefore, to test the hypothesis H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$, compute the test statistic F_0 and reject H_0 if

$$F_0 > F_{\alpha,k,n-k-1}$$

Example 3.3 Delivery Time Data

$$SS_{\mathrm{T}} = \mathbf{y}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

$$= 18{,}310.6290 - \frac{(559.60)^2}{25} = 5784.5426$$

$$SS_{\mathrm{R}} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

$$= 18{,}076.9030 - \frac{(559.60)^2}{25} = 5550.8166$$

Example 3.3 Delivery Time Data

$$\begin{aligned} SS_{\text{Res}} &= SS_{\text{T}} - SS_{\text{R}} \\ &= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = 233.7260 \end{aligned}$$

To test H_0: $\beta_1 = \beta_2 = 0$, we calculate the F–statistic:

$$F_0 = \frac{MS_{\text{R}}}{MS_{\text{Res}}} = \frac{2775.4083}{10.6239} = 261.24$$

Example 3.3 Delivery Time Data

TABLE 3.6 Test for Significance of Regression for Example 3.3

Source Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P Value
Regression	5550.8166	2	2775.4083	261.24	4.7×10^{-16}
Residual	233.7260	22	10.6239		
Total	5784.5426	24			

3.3.1 Test for Significance of Regression

- R^2
  - R^2 is calculated exactly as in simple linear regression
  - R^2 can be inflated simply by adding more terms to the model (even insignificant terms)
- Adjusted R^2
  - Penalizes you for added terms to the model that are not significant

p=#B's = k+1

$$R^2_{adj} = 1 - \frac{SS_{\mathrm{Res}}/(n-p)}{SS_T/(n-1)}$$

3.3.1 Test for Significance of Regression

- Adjusted R^2 Example
  - Say SS_T = 245.00, n = 15
  - Suppose that for a model with three regressors, the SS_{res} = 90.00, then

$$R^2_{adj} = 1 - \frac{90/(15-4)}{245/(15-1)} = 0.53$$

  - Now suppose that a fourth regressor has been added, and the SS_{Res} = 88.00

$$R^2_{adj} = 1 - \frac{88/(15-5)}{245/(15-1)} = 0.49$$

3.3.2 Tests on Individual Regression Coefficients

- Hypothesis test on any single regression coefficient:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

- Test Statistic:

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

  – Reject H_0 if $|t_0| > t_{\alpha/2, n-k-1}$
  – This is a **partial** or **marginal** test!

See Example 3.4, pg. 88, text

The **Extra Sum of Squares** method can also be used to test hypotheses on individual model parameters or groups of parameters

Consider the regression model with k regressors

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where y is $n \times 1$, $\mathbf{X}$ is $n \times p$, $\boldsymbol{\beta}$ is $p \times 1$, $\boldsymbol{\varepsilon}$ is $n \times 1$, and $p = k + 1$. We would like to determine if some subset of $r < k$ regressors contributes significantly to the regression model. Let the vector of regression coefficients be partitioned as follows:

$$\boldsymbol{\beta} = \left[\frac{\boldsymbol{\beta}_1}{\boldsymbol{\beta}_2}\right]$$

where $\boldsymbol{\beta}_1$ is $(p - r) \times 1$ and $\boldsymbol{\beta}_2$ is $r \times 1$. We wish to test the hypotheses

$$H_0\colon \boldsymbol{\beta}_2 = \mathbf{0}$$

$$H_1\colon \boldsymbol{\beta}_2 \neq \mathbf{0} \tag{3.30}$$

The model may be written as

Full model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \tag{3.31}$$

For the full model, we know that $\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. The regression sum of squares for this model is

$$SS_{\mathrm{R}}(\boldsymbol{\beta}) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \text{ (} p \text{ degrees of freedom)}$$

and

$$MS_{\mathrm{Res}} = \frac{\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}}{n - p}$$

To find the contribution of the terms in $\boldsymbol{\beta}_2$ to the regression, fit the model assuming that the null hypothesis H_0: $\boldsymbol{\beta}_2 = \mathbf{0}$ is true. This **reduced model** is

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon} \tag{3.32}$$

The least-squares estimator of $\boldsymbol{\beta}_1$ in the reduced model is $\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}$. The regression sum of squares is

$$SS_{\mathrm{R}}(\boldsymbol{\beta}_1) = \hat{\boldsymbol{\beta}}_1\mathbf{X}_1'\mathbf{y} \text{ (} p - r \text{ degrees of freedom)} \tag{3.33}$$

The regression sum of squares due to $\boldsymbol{\beta}_2$ given that $\boldsymbol{\beta}_1$ is already in the model is

$$SS_R(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = SS_R(\boldsymbol{\beta}) - SS_R(\boldsymbol{\beta}_1) \tag{3.34}$$

with $p - (p - r) = r$ degrees of freedom. This sum of squares is called the **extra sum of squares due to $\boldsymbol{\beta}_2$** because it measures the increase in the regression sum of squares that results from adding the regressors $x_{k-r+1}, x_{k-r+2}, \ldots, x_k$ to a model that already contains $x_1, x_2, \ldots, x_{k-r}$. Now $SS_R(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$ is independent of MS_{Res}, and the null hypothesis $\boldsymbol{\beta}_2 = \mathbf{0}$ may be tested by the statistic

$$F_0 = \frac{SS_R(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)/r}{MS_{\mathrm{Res}}} \tag{3.35}$$

If $F_0 > F_{\alpha,r,n-p}$, we reject H_0, concluding that at least one of the parameters in $\boldsymbol{\beta}_2$ is not zero, and consequently at least one of the regressors $x_{k-r+1}, x_{k-r+2}, \ldots, x_k$ in $\mathbf{X}_2$ contribute significantly to the regression model. Some authors call the test in (3.35) a **partial F test** because it measures the contribution of the regressors in $\mathbf{X}_2$ given that the other regressors in $\mathbf{X}_1$ are in the model.

Cochran's Theorem is useful for justification of F ratios

3.4.1. Confidence Intervals on the Regression Coefficients

A 100(1-α) percent C.I. for the regression coefficient, β_j is:

$$\hat{\beta}_j - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 C_{jj}} \le \beta_j \le \hat{\beta}_j + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 C_{jj}}$$

Or,

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

Example 3.8 The Delivery Time Data

We now find a 95% CI for the parameter β_1 in Example 3.1. The point estimate of β_1 is $\hat{\beta}_1 = 1.61591$, the diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$ corresponding to β_1 is $C_{11} = 0.00274378$, and $\hat{\sigma}^2 = 10.6239$ (from Example 3.2). Using Eq. (3.45), we find that

$$\hat{\beta}_1 - t_{0.025,22}\sqrt{\hat{\sigma}^2 C_{11}} \le \beta_1 \le \hat{\beta}_1 + t_{0.025,22}\sqrt{\hat{\sigma}^2 C_{11}}$$

$$1.61591 - (2.074)\sqrt{(10.6239)(0.00274378)}$$
$$\le \beta_1 \le 1.61591 + (2.074)\sqrt{(10.6239)(0.00274378)}$$

$$1.61591 - (2.074)(0.17073) \le \beta_1 \le 1.61591 + (2.074)(0.17073)$$

and the 95% CI on β_1 is

$$1.26181 \le \beta_1 \le 1.97001$$

Notice that the Minitab output in Table 3.4 gives the standard error of each regression coefficient. This makes the construction of these intervals very easy in practice. ■

3.4.2. Confidence Interval Estimation of the Mean Response

- 100(1-α) percent CI on the mean response at the point $x_{01}, x_{02}, \ldots, x_{0k}$ is

Var(yhat) = Var(x_0'b) = sigma**2 x_0'(X'X)$^{-1}$ x_0

$$\hat{y}_0 - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 \mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0} \le E(y \mid \mathbf{x}_0)$$

$$\le \hat{y}_0 + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 \mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}$$

- See Example 3-9 on page 95 and the discussion that follows

Example 3.9 The Delivery Time Data

The soft drink bottler in Example 3.1 would like to construct a 95% CI on the mean delivery time for an outlet requiring $x_1 = 8$ cases and where the distance $x_2 = 275$ feet. Therefore,

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 8 \\ 275 \end{bmatrix}$$

The fitted value at this point is found from Eq. (3.47) as

$$\hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}} = [1 \quad 8 \quad 275] \begin{bmatrix} 2.34123 \\ 1.61591 \\ 0.01438 \end{bmatrix} = 19.22 \text{ minutes}$$

The variance of $\hat{y}_0$ is estimated by

$$\hat{\sigma}^2 \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 = 10.6239[1 \quad 8 \quad 275]$$
$$\times \begin{bmatrix} 0.11321518 & -0.00444859 & -0.00008367 \\ -0.00444859 & 0.00274378 & -0.00004786 \\ -0.00008367 & -0.00004786 & 0.00000123 \end{bmatrix} \begin{bmatrix} 1 \\ 8 \\ 275 \end{bmatrix}$$
$$= 10.6239(0.05346) = 0.56794$$

Therefore, a 95% CI on the mean delivery time at this point is found from Eq. (3.49) as

$$19.22 - 2.074\sqrt{0.56794} \le E(y|\mathbf{x}_0) \le 19.22 + 2.074\sqrt{0.56794}$$

which reduces to

$$17.66 \le E(y|\mathbf{x}_0) \le 20.78$$

Ninety-five percent of such intervals will contain the true delivery time. ■

The length of the CI or the mean response is a useful measure of the quality of the regression model. It can also be used to compare competing models. To illustrate, consider the 95% CI on the the mean delivery time when $x_1 = 8$ cases and $x_2 = 275$ feet. In Example 3.9 this CI is found to be (17.66, 20.78), and the length of this interval is 20.78 – 17.16 = 3.12 minutes. If we consider the simple linear regression model with x_1 = cases as the only regressor, the 95% CI on the mean delivery time with $x_1 = 8$ cases is (18.99, 22.97). The length of this interval is 22.47 – 18.99 = 3.45 minutes. Clearly, adding cases to the model has improved the precision of estimation. However, the change in the length of the interval depends on the location of the point in the x space. Consider the point $x_1 = 16$ cases and $x_2 = 688$ feet. The 95% CI for the multiple regression model is (36.11, 40.08) with length 3.97 minutes, and for the simple linear regression model the 95% CI at $x_1 = 16$ cases is (35.60, 40.68) with length 5.08 minutes. The improvement from the multiple regression model is even better at this point. Generally, the further the point is from the centroid of the x space, the greater the difference will be in the lengths of the two CIs.

3.4.3. Simultaneous Confidence Intervals on Regression Coefficients

It can be shown that

$$\frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{pMS_{\mathrm{Re}s}} \sim F_{p,n-p}$$

From this result, the joint confidence region for all parameters in β is

$$\frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{pMS_{\mathrm{Re}s}} \le F_{\alpha,p,n-p}$$

3.5 Prediction of New Observations

- A 100(1-α) percent prediction interval for a future observation is

$$\hat{y}_0 - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2(1+\mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)} \le y_0$$
$$\le \hat{y}_0 + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2(1+\mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)}$$

Example 3.12 The Delivery Time Data

Suppose that the soft drink bottler in Example 3.1 wishes to construct a 95% prediction interval on the delivery time at an outlet where $x_1 = 8$ cases are delivered and the distance walked by the deliveryman is $x_2 = 275$ feet. Note that $\mathbf{x}_0' = [1, \quad 8, \quad 275]$, and the point estimate of the delivery time is $\hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}} = 19.22$ minutes. Also, in Example 3.9 we calculated $\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 = 0.05346$. Therefore, from (3.54) we have

$$19.22 - 2.074\sqrt{10.6239(1 + 0.05346)} \le y_0 \le 19.22 + 2.074\sqrt{10.6239(1 + 0.05346)}$$

and the 95% prediction interval is

$$12.28 \le y_0 \le 26.16 \qquad \blacksquare$$

3.9 Standardized Regression Coefficients

- It is often difficult to directly compare regression coefficients due to possible varying dimensions.
- It may be beneficial to work with dimensionless regression coefficients.
- Dimensionless regression coefficients are often referred to as **standardized regression coefficients.**
- Two common methods of scaling:

1. Unit normal scaling
2. Unit length scaling

3.9 Standardized Regression Coefficients

Unit Normal Scaling

The first approach employs **unit normal scaling** for the regressors and the response variable. That is,

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, k$$

$$y_i^* = \frac{y_i - \bar{y}}{s_y}, \quad i = 1, 2, \ldots, n$$

This is what SAS does – ex3_1stdb.sas

see ex3_1stdb.R for R package that provides standardized coefficients

3.9 Standardized Regression Coefficients

Unit Normal Scaling

where

$$s_j^2 = \frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}{n - 1}$$

$$s_y^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n - 1}$$

3.9 Standardized Regression Coefficients

Unit Normal Scaling

- All of the scaled regressors and the scaled response have sample mean equal to zero and sample variance equal to 1.
- The model becomes

$$y_i^* = b_1 z_{i1} + b_2 z_{i2} + \cdots + b_k z_{ik} + \varepsilon_i, \quad i = 1, 2, \ldots, n$$

- The least squares estimator: $\hat{\mathbf{b}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}^*$

3.9 Standardized Regression Coefficients

Unit Length Scaling

- In unit length scaling:

$$w_{ij} = \frac{x_{ij} - \bar{x}_j}{S_{jj}^{1/2}}, \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, k$$

$$y_i^0 = \frac{y_i - \bar{y}}{SS_T^{1/2}}, \quad i = 1, 2, \ldots, n$$

$$\boxed{S_{jj} = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}$$

3.9 Standardized Regression Coefficients

Unit Length Scaling

- Each regressor has mean 0 and length

$$\sqrt{\Sigma_{i=1}^{n}(w_{ij} - \overline{w}_j)^2} = 1$$

- The regression model becomes

$$y_i^0 = b_q w_{i1} + b_2 w_{i2} + \cdots + b_k w_{ik} + \varepsilon_i, \quad i = 1, 2, \ldots, n$$

- The vector of least squares regression coefficients: $\hat{\mathbf{b}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y}^0$

3.9 Standardized Regression Coefficients

Unit Length Scaling

- In unit length scaling, the **W'W** matrix is in the form of a **correlation matrix**:

$$\mathbf{W'W} = \begin{bmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1k} \\ r_{12} & 1 & r_{23} & \cdots & r_{2k} \\ r_{13} & r_{23} & 1 & \cdots & r_{34} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{1k} & r_{2k} & r_{3k} & \cdots & 1 \end{bmatrix}$$

where r_{ij} is the simple correlation between x_i and x_j.

3.9 Standardized Regression Coefficients

The regression coefficients $\hat{\mathbf{b}}$ are usually called **standardized regression coefficients.** The relationship between the original and standardized regression coefficients is

$$\hat{\beta}_j = \hat{b}_j \left(\frac{SS_T}{S_{jj}} \right)^{1/2}, \quad j = 1, 2, \ldots, k \tag{3.63}$$

and

$$\hat{\beta}_0 = \bar{y} - \sum_{j=1}^{k} \hat{\beta}_j \bar{x}_j \tag{3.64}$$

Example 3.14

We will find the standardized regression coefficients for the delivery time data in Example 3.1. Since

$$
\begin{aligned}
SS_{\mathrm{T}} &= 5784.5426 & S_{11} &= 1136.5600 \\
S_{1y} &= 2473.3440 & S_{22} &= 2{,}537{,}935.0330 \\
S_{2y} &= 108{,}038.6019 & S_{12} &= 44{,}266.6800
\end{aligned}
$$

Example 3.14

we find (using the unit length scaling) that

$$r_{12} = \frac{S_{12}}{(S_{11}S_{22})^{1/2}} = \frac{44{,}266.6800}{\sqrt{(1136.5600)(2{,}537{,}935.0303)}} = 0.824215$$

$$r_{1y} = \frac{S_{1y}}{(S_{11}SS_{\mathrm{T}})^{1/2}} = \frac{2473.3440}{\sqrt{(1136.5600)(5784.53426)}} = 0.964615$$

$$r_{2y} = \frac{S_{2y}}{(S_{22}SS_{\mathrm{T}})^{1/2}} = \frac{108{,}038.6019}{\sqrt{(2{,}537{,}935.0330)(5784.5426)}} = 0.891670$$

Example 3.14

the correlation matrix for this problem is

$$\mathbf{W'W} = \begin{bmatrix} 1 & 0.824215 \\ 0.824215 & 1 \end{bmatrix}$$

The normal equations in terms of the standardized regression coefficients are

$$\begin{bmatrix} 1 & 0.824215 \\ 0.824215 & 1 \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} = \begin{bmatrix} 0.964615 \\ 0.891670 \end{bmatrix}$$

Example 3.14

the standardized regression coefficients are

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0.824215 \\ 0.824215 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.964615 \\ 0.891670 \end{bmatrix}$$

$$= \begin{bmatrix} 3.11841 & -2.57023 \\ -2.57023 & 3.11841 \end{bmatrix} \begin{bmatrix} 0.964615 \\ 0.891670 \end{bmatrix}$$

$$= \begin{bmatrix} 0.716267 \\ 0.301311 \end{bmatrix}$$

The fitted model is

$$\hat{y}^0 = 0.716267 w_1 + 0.301311 w_2$$

3.8 Multicollinearity

- A serious problem that may dramatically impact the usefulness of a regression model is **multicollinearity**, or **near-linear dependence** among the regression variables.
- Multicollinearity implies near-linear dependence among the regressors. The regressors are the columns of the **X** matrix, so clearly an **exact linear dependence** would result in a **singular X’X**.
- The presence of multicollinearity can dramatically impact the ability to estimate regression coefficients and other uses of the regression model.

3.8 Multicollinearity

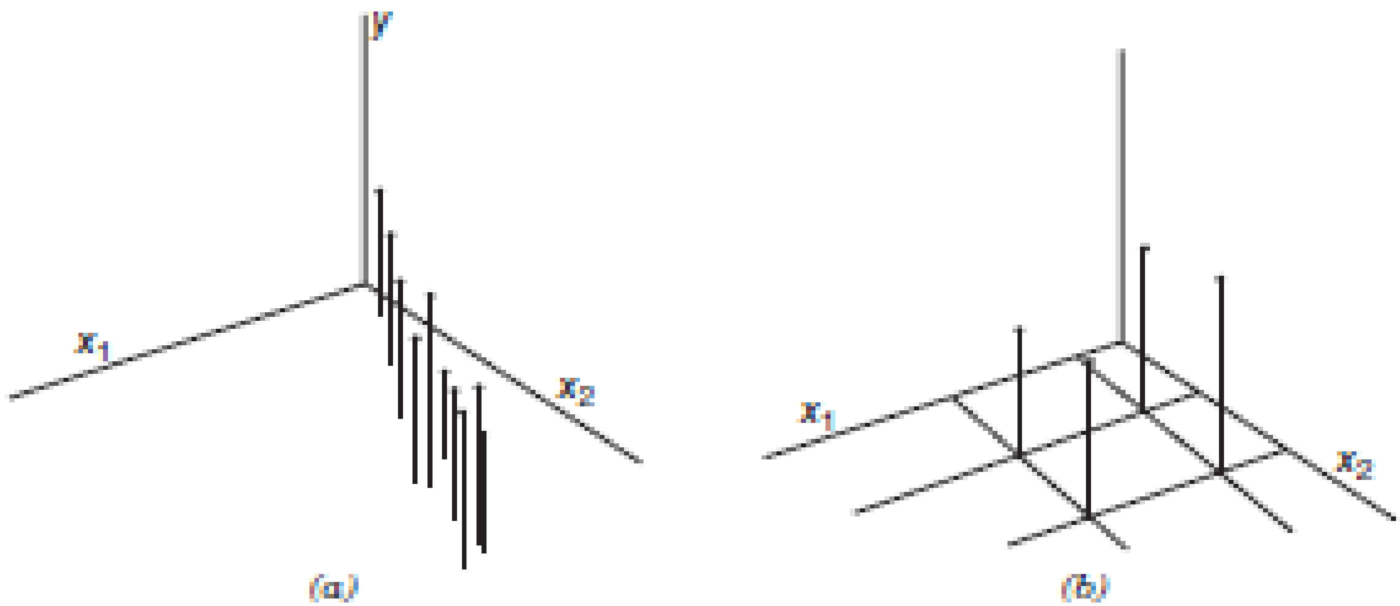


Figure 3.13 (*a*) A data set with multicollinearity. (*b*) Orthogonal regressors.

3.8 Multicollinearity

- The main diagonal elements of the inverse of the **X'X** matrix in correlation form $(\mathbf{W'W})^{-1}$ are often called **variance inflation factors** VIFs, and they are an important multicollinearity diagnostic.
- For the soft drink delivery data,

$$\mathrm{VIF}_1 = \mathrm{VIF}_2 = 3.11841$$

3.8 Multicollinearity

Look at Exercise 3.7

- The variance inflation factors can also be written as:

$$\mathrm{VIF}_j = \frac{1}{1 - R_j^2}$$

where R^2_j is the coefficient of multiple determination obtained from regressing x_j on the other regressor variables.

- If x_j is highly correlated with any other regressor variable, then R^2_j will be large.

Look for VIF > 10