20. Suppose that $A\mathbf{x} = \mathbf{0}$. Then $A^TA\mathbf{x} = A^T\mathbf{0} = \mathbf{0}$. Since A^TA is invertible, $\mathbf{x}$ must be $\mathbf{0}$. Hence the columns of A are linearly independent.

21. a. If A has linearly independent columns, then the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. By Exercise 19, the equation $A^TA\mathbf{x} = \mathbf{0}$ also has only the trivial solution. Since A^TA is a square matrix, it must be invertible by the Invertible Matrix Theorem.

b. Since the n linearly independent columns of A belong to $\mathbb{R}^m$, m could not be less than n.

c. The n linearly independent columns of A form a basis for Col A, so the rank of A is n.

22. Note that A^TA has n columns because A does. Then by the Rank Theorem and Exercise 19,

$$\text{rank}\, A^TA = n - \dim \text{Nul}\, A^TA = n - \dim \text{Nul}\, A = \text{rank}\, A$$

23. By Theorem 14, $\hat{\mathbf{b}} = A\hat{\mathbf{x}} = A(A^TA)^{-1}A^T\mathbf{b}$. The matrix $A(A^TA)^{-1}A^T$ is sometimes called the *hat-matrix* in statistics.

24. Since in this case $A^TA = I$, the normal equations give $\hat{\mathbf{x}} = A^T\mathbf{b}$.

25. The normal equations are $\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \end{bmatrix}$, whose solution is the set of all (x, y) such that $x + y = 3$. The solutions correspond to the points on the line midway between the lines $x + y = 2$ and $x + y = 4$.

26. **[M]** Using .7 as an approximation for $\sqrt{2}/2$, $a_0 = a_2 \approx .353535$ and $a_1 = .5$. Using .707 as an approximation for $\sqrt{2}/2$, $a_0 = a_2 \approx .35355339$, $a_1 = .5$.

6.6 SOLUTIONS

Notes: This section is a valuable reference for any person who works with data that requires statistical analysis. Many graduate fields require such work. Science students in particular will benefit from Example 1. The general linear model and the subsequent examples are aimed at students who may take a multivariate statistics course. That may include more students than one might expect.

1. The design matrix X and the observation vector $\mathbf{y}$ are

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \end{bmatrix},$$

and one can compute

$$X^TX = \begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix}, X^T\mathbf{y} = \begin{bmatrix} 6 \\ 11 \end{bmatrix}, \hat{\beta} = (X^TX)^{-1}X^T\mathbf{y} = \begin{bmatrix} .9 \\ .4 \end{bmatrix}$$

The least-squares line $y = \beta_0 + \beta_1 x$ is thus $y = .9 + .4x$.

2. The design matrix X and the observation vector $\mathbf{y}$ are

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 5 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix},$$

and one can compute

$$X^TX = \begin{bmatrix} 4 & 12 \\ 12 & 46 \end{bmatrix}, X^T\mathbf{y} = \begin{bmatrix} 6 \\ 25 \end{bmatrix}, \hat{\beta} = (X^TX)^{-1}X^T\mathbf{y} = \begin{bmatrix} -.6 \\ .7 \end{bmatrix}$$

The least-squares line $y = \beta_0 + \beta_1 x$ is thus $y = -.6 + .7x$.

3. The design matrix X and the observation vector $\mathbf{y}$ are

$$X = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 4 \end{bmatrix},$$

and one can compute

$$X^TX = \begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix}, X^T\mathbf{y} = \begin{bmatrix} 7 \\ 10 \end{bmatrix}, \hat{\beta} = (X^TX)^{-1}X^T\mathbf{y} = \begin{bmatrix} 1.1 \\ 1.3 \end{bmatrix}$$

The least-squares line $y = \beta_0 + \beta_1 x$ is thus $y = 1.1 + 1.3x$.

4. The design matrix X and the observation vector $\mathbf{y}$ are

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 5 \\ 1 & 6 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 3 \\ 2 \\ 1 \\ 0 \end{bmatrix},$$

and one can compute

$$X^TX = \begin{bmatrix} 4 & 16 \\ 16 & 74 \end{bmatrix}, X^T\mathbf{y} = \begin{bmatrix} 6 \\ 17 \end{bmatrix}, \hat{\beta} = (X^TX)^{-1}X^T\mathbf{y} = \begin{bmatrix} 4.3 \\ -.7 \end{bmatrix}$$

The least-squares line $y = \beta_0 + \beta_1 x$ is thus $y = 4.3 - .7x$.

5. If two data points have different x-coordinates, then the two columns of the design matrix X cannot be multiples of each other and hence are linearly independent. By Theorem 14 in Section 6.5, the normal equations have a unique solution.

6. If the columns of X were linearly dependent, then the same dependence relation would hold for the vectors in $\mathbb{R}^3$ formed from the top three entries in each column. That is, the columns of the matrix $\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{bmatrix}$ would also be linearly dependent, and so this matrix (called a Vandermonde matrix) would be noninvertible. Note that the determinant of this matrix is $(x_2 - x_1)(x_3 - x_1)(x_3 - x_2) \neq 0$ since x_1, x_2, and x_3 are distinct. Thus this matrix is invertible, which means that the columns of X

are in fact linearly independent. By Theorem 14 in Section 6.5, the normal equations have a unique solution.

7. a. The model that produces the correct least-squares fit is $\mathbf{y} = X\beta + \epsilon$ where

$$X = \begin{bmatrix} 1 & 1 \\ 2 & 4 \\ 3 & 9 \\ 4 & 16 \\ 5 & 25 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1.8 \\ 2.7 \\ 3.4 \\ 3.8 \\ 3.9 \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \text{and } \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}$$

b. [M] One computes that (to two decimal places) $\hat{\beta} = \begin{bmatrix} 1.76 \\ -.20 \end{bmatrix}$, so the desired least-squares equation is $y = 1.76x - .20x^2$.

8. a. The model that produces the correct least-squares fit is $\mathbf{y} = X\beta + \epsilon$ where

$$X = \begin{bmatrix} x_1 & x_1^2 & x_1^3 \\ \vdots & \vdots & \vdots \\ x_n & x_n^2 & x_n^3 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, \text{and } \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

b. [M] For the given data,

$$X = \begin{bmatrix} 4 & 16 & 64 \\ 6 & 36 & 216 \\ 8 & 64 & 512 \\ 10 & 100 & 1000 \\ 12 & 144 & 1728 \\ 14 & 196 & 2744 \\ 16 & 256 & 4096 \\ 18 & 324 & 5832 \end{bmatrix} \text{and } \mathbf{y} = \begin{bmatrix} 1.58 \\ 2.08 \\ 2.5 \\ 2.8 \\ 3.1 \\ 3.4 \\ 3.8 \\ 4.32 \end{bmatrix}$$

so $\hat{\beta} = (X^TX)^{-1}X^T\mathbf{y} = \begin{bmatrix} .5132 \\ -.03348 \\ .001016 \end{bmatrix}$, and the least-squares curve is

$y = .5132x - .03348x^2 + .001016x^3$.

9. The model that produces the correct least-squares fit is $\mathbf{y} = X\beta + \epsilon$ where

$$X = \begin{bmatrix} \cos 1 & \sin 1 \\ \cos 2 & \sin 2 \\ \cos 3 & \sin 3 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 7.9 \\ 5.4 \\ -.9 \end{bmatrix}, \beta = \begin{bmatrix} A \\ B \end{bmatrix}, \text{and } \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}$$

10. a. The model that produces the correct least-squares fit is $\mathbf{y} = X\beta + \epsilon$ where

$$X = \begin{bmatrix} e^{-.02(10)} & e^{-.07(10)} \\ e^{-.02(11)} & e^{-.07(11)} \\ e^{-.02(12)} & e^{-.07(12)} \\ e^{-.02(14)} & e^{-.07(14)} \\ e^{-.02(15)} & e^{-.07(15)} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 21.34 \\ 20.68 \\ 20.05 \\ 18.87 \\ 18.30 \end{bmatrix}, \beta = \begin{bmatrix} M_A \\ M_B \end{bmatrix}, \text{and } \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix},$$

b. [M] One computes that (to two decimal places) $\hat{\beta} = \begin{bmatrix} 19.94 \\ 10.10 \end{bmatrix}$, so the desired least-squares equation is $y = 19.94e^{-.02t} + 10.10e^{-.07t}$.

11. [M] The model that produces the correct least-squares fit is $\mathbf{y} = X\beta + \epsilon$ where

$$X = \begin{bmatrix} 1 & 3\cos .88 \\ 1 & 2.3\cos 1.1 \\ 1 & 1.65\cos 1.42 \\ 1 & 1.25\cos 1.77 \\ 1 & 1.01\cos 2.14 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 3 \\ 2.3 \\ 1.65 \\ 1.25 \\ 1.01 \end{bmatrix}, \beta = \begin{bmatrix} \beta \\ e \end{bmatrix}, \text{and } \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}$$

One computes that (to two decimal places) $\hat{\beta} = \begin{bmatrix} 1.45 \\ .811 \end{bmatrix}$. Since $e = .811 < 1$ the orbit is an ellipse. The equation $r = \beta / (1 - e \cos \vartheta)$ produces $r = 1.33$ when $\vartheta = 4.6$.

12. [M] The model that produces the correct least-squares fit is $\mathbf{y} = X\beta + \epsilon$, where

$$X = \begin{bmatrix} 1 & 3.78 \\ 1 & 4.11 \\ 1 & 4.39 \\ 1 & 4.73 \\ 1 & 4.88 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 91 \\ 98 \\ 103 \\ 110 \\ 112 \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \text{and } \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}$$

One computes that (to two decimal places) $\hat{\beta} = \begin{bmatrix} 18.56 \\ 19.24 \end{bmatrix}$, so the desired least-squares equation is $p = 18.56 + 19.24 \ln w$. When $w = 100$, $p \approx 107$ millimeters of mercury.

13. [M]

a. The model that produces the correct least-squares fit is $\mathbf{y} = X\boldsymbol{\beta} + \epsilon$ where

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 2^2 & 2^3 \\ 1 & 3 & 3^2 & 3^3 \\ 1 & 4 & 4^2 & 4^3 \\ 1 & 5 & 5^2 & 5^3 \\ 1 & 6 & 6^2 & 6^3 \\ 1 & 7 & 7^2 & 7^3 \\ 1 & 8 & 8^2 & 8^3 \\ 1 & 9 & 9^2 & 9^3 \\ 1 & 10 & 10^2 & 10^3 \\ 1 & 11 & 11^2 & 11^3 \\ 1 & 12 & 12^2 & 12^3 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 0 \\ 8.8 \\ 29.9 \\ 62.0 \\ 104.7 \\ 159.1 \\ 222.0 \\ 294.5 \\ 380.4 \\ 471.1 \\ 571.7 \\ 686.8 \\ 809.2 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, \text{ and } \epsilon = \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \\ \epsilon_{10} \\ \epsilon_{11} \\ \epsilon_{12} \end{bmatrix}$$

One computes that (to four decimal places) $\hat{\boldsymbol{\beta}} = \begin{bmatrix} -.8558 \\ 4.7025 \\ 5.5554 \\ -.0274 \end{bmatrix}$, so the desired least-squares polynomial is $y(t) = -.8558 + 4.7025t + 5.5554t^2 - .0274t^3$.

b. The velocity $v(t)$ is the derivative of the position function $y(t)$, so $v(t) = 4.7025 + 11.1108t - .0822t^2$, and $v(4.5) = 53.0$ ft/sec.

14. Write the design matrix as $[\mathbf{1} \quad \mathbf{x}]$. Since the residual vector $\epsilon = \mathbf{y} - X\hat{\boldsymbol{\beta}}$ is orthogonal to Col X,

$$0 = \mathbf{1} \cdot \epsilon = \mathbf{1} \cdot (\mathbf{y} - X\hat{\boldsymbol{\beta}}) = \mathbf{1}^T \mathbf{y} - (\mathbf{1}^T X)\hat{\boldsymbol{\beta}}$$

$$= (y_1 + \dots + y_n) - \begin{bmatrix} n & \sum x \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \sum y - n\hat{\beta}_0 - \hat{\beta}_1 \sum x = n\overline{y} - n\hat{\beta}_0 - n\hat{\beta}_1 \overline{x}$$

This equation may be solved for $\overline{y}$ to find $\overline{y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{x}$.

15. From equation (1) on page 369,

$$X^T X = \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix}$$

$$X^T \mathbf{y} = \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}$$

The equations (7) in the text follow immediately from the normal equations $X^T X \boldsymbol{\beta} = X^T \mathbf{y}$.

16. The determinant of the coefficient matrix of the equations in (7) is $n\sum x^2 - (\sum x)^2$. Using the 2×2 formula for the inverse of the coefficient matrix,

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \frac{1}{n\sum x^2 - (\sum x)^2} \begin{bmatrix} \sum x^2 & -\sum x \\ -\sum x & n \end{bmatrix} \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}$$

Hence

$$\hat{\beta}_0 = \frac{(\sum x^2)(\sum y) - (\sum x)(\sum xy)}{n\sum x^2 - (\sum x)^2}, \quad \hat{\beta}_1 = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$$

Note: A simple algebraic calculation shows that $\sum y - (\sum x)\hat{\beta}_1 = n\hat{\beta}_0$, which provides a simple formula for $\hat{\beta}_0$ once $\hat{\beta}_1$ is known

17. a. The mean of the data in Example 1 is $\overline{x} = 5.5$, so the data in mean-deviation form are (–3.5, 1), (–.5, 2), (1.5, 3), (2.5, 3), and the associated design matrix is $X = \begin{bmatrix} 1 & -3.5 \\ 1 & -.5 \\ 1 & 1.5 \\ 1 & 2.5 \end{bmatrix}$. The columns of X are orthogonal because the entries in the second column sum to 0.

b. The normal equations are $X^T X \boldsymbol{\beta} = X^T \mathbf{y}$, or $\begin{bmatrix} 4 & 0 \\ 0 & 21 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 7.5 \end{bmatrix}$. One computes that $\hat{\boldsymbol{\beta}} = \begin{bmatrix} 9/4 \\ 5/14 \end{bmatrix}$, so the desired least-squares line is $y = (9/4) + (5/14)x^* = (9/4) + (5/14)(x - 5.5)$.

18. Since

$$X^T X = \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix}$$

$X^T X$ is a diagonal matrix when $\sum x = 0$.

19. The residual vector $\epsilon = \mathbf{y} - X\hat{\boldsymbol{\beta}}$ is orthogonal to Col X, while $\hat{\mathbf{y}} = X\hat{\boldsymbol{\beta}}$ is in Col X. Since ϵ and $\hat{\mathbf{y}}$ are thus orthogonal, apply the Pythagorean Theorem to these vectors to obtain

$$\text{SS(T)} = \|\mathbf{y}\|^2 = \|\hat{\mathbf{y}} + \epsilon\|^2 = \|\hat{\mathbf{y}}\|^2 + \|\epsilon\|^2 = \|X\hat{\boldsymbol{\beta}}\|^2 + \|\mathbf{y} - X\hat{\boldsymbol{\beta}}\|^2 = \text{SS(R)} + \text{SS(E)}$$

20. Since $\hat{\boldsymbol{\beta}}$ satisfies the normal equations, $X^T X\hat{\boldsymbol{\beta}} = X^T \mathbf{y}$, and

$$\|X\hat{\boldsymbol{\beta}}\|^2 = (X\hat{\boldsymbol{\beta}})^T (X\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\beta}}^T X^T X\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^T X^T \mathbf{y}$$

Since $\|X\hat{\boldsymbol{\beta}}\|^2 = \text{SS(R)}$ and $\mathbf{y}^T \mathbf{y} = \|\mathbf{y}\|^2 = \text{SS(T)}$, Exercise 19 shows that

$$\text{SS(E)} = \text{SS(T)} - \text{SS(R)} = \mathbf{y}^T \mathbf{y} - \hat{\boldsymbol{\beta}}^T X^T \mathbf{y}$$