Exercises

7.1. Given the data

z_1	10	5	7	19	11	8
y	15	9	3	25	7	13

fit the linear regression model $Y_j = \beta_0 + \beta_1 z_{j1} + \varepsilon_j$, $j = 1, 2, \dots, 6$. Specifically, calculate the least squares estimates $\hat{\boldsymbol{\beta}}$, the fitted values $\hat{\mathbf{y}}$, the residuals $\hat{\boldsymbol{\varepsilon}}$, and the residual sum of squares, $\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$.

√**7.2.** Given the data

z_1	10	5	7	19	11	18
z_2	2	3	3	6	7	9
y	15	9	3	25	7	13

fit the regression model

$$Y_j = \beta_1 z_{j1} + \beta_2 z_{j2} + \varepsilon_j, \qquad j = 1, 2, \dots, 6.$$

to the *standardized* form (see page 412) of the variables y, z_1, and z_2. From this fit, deduce the corresponding fitted regression equation for the original (not standardized) variables.

√**7.3.** (*Weighted least squares estimators.*) Let

$$\underset{(n\times 1)}{\mathbf{Y}} = \underset{(n\times(r+1))}{\mathbf{Z}} \; \underset{((r+1)\times 1)}{\boldsymbol{\beta}} + \underset{(n\times 1)}{\boldsymbol{\varepsilon}}$$

where $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ but $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\mathbf{V}$, with $\mathbf{V}(n \times n)$ known and positive definite. For $\mathbf{V}$ of full rank, show that the *weighted least squares* estimator is

$$\hat{\boldsymbol{\beta}}_W = (\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Y}$$

If σ^2 is unknown, it may be estimated, unbiasedly, by

$$(n - r - 1)^{-1} \times (\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}_W)'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}_W).$$

Hint: $\mathbf{V}^{-1/2}\mathbf{Y} = (\mathbf{V}^{-1/2}\mathbf{Z})\boldsymbol{\beta} + \mathbf{V}^{-1/2}\boldsymbol{\varepsilon}$ is of the classical linear regression form $\mathbf{Y}^* = \mathbf{Z}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*$, with $E(\boldsymbol{\varepsilon}^*) = \mathbf{0}$ and $E(\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*\prime}) = \sigma^2\mathbf{I}$. Thus, $\hat{\boldsymbol{\beta}}_W = \hat{\boldsymbol{\beta}}^* = (\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime}\mathbf{Y}^*$.

√**7.4.** Use the weighted least squares estimator in Exercise 7.3 to derive an expression for the estimate of the slope β in the model $Y_j = \beta z_j + \varepsilon_j, j = 1, 2, \dots, n$, when (a) $\text{Var}(\varepsilon_j) = \sigma^2$, (b) $\text{Var}(\varepsilon_j) = \sigma^2 z_j$, and (c) $\text{Var}(\varepsilon_j) = \sigma^2 z_j^2$. Comment on the manner in which the unequal variances for the errors influence the optimal choice of $\hat{\beta}_W$.

7.5. Establish (7-50): $\rho^2_{Y(Z)} = 1 - 1/\rho^{YY}$.
Hint: From (7-49) and Exercise 4.11

$$1 - \rho^2_{Y(\mathbf{Z})} = \frac{\sigma_{YY} - \boldsymbol{\sigma}'_{\mathbf{Z}Y}\boldsymbol{\Sigma}^{-1}_{\mathbf{ZZ}}\boldsymbol{\sigma}_{\mathbf{Z}Y}}{\sigma_{YY}} = \frac{|\boldsymbol{\Sigma}_{\mathbf{ZZ}}|}{|\boldsymbol{\Sigma}_{\mathbf{ZZ}}|}\frac{(\sigma_{YY} - \boldsymbol{\sigma}'_{\mathbf{Z}Y}\boldsymbol{\Sigma}^{-1}_{\mathbf{ZZ}}\boldsymbol{\sigma}_{\mathbf{Z}Y})}{\sigma_{YY}} = \frac{|\boldsymbol{\Sigma}|}{|\boldsymbol{\Sigma}_{\mathbf{ZZ}}|\sigma_{YY}}$$

From Result 2A.8(c), $\sigma^{YY} = |\boldsymbol{\Sigma}_{\mathbf{ZZ}}|/|\boldsymbol{\Sigma}|$, where σ^{YY} is the entry of $\boldsymbol{\Sigma}^{-1}$ in the first row and first column. Since (see Exercise 2.23) $\boldsymbol{\rho} = \mathbf{V}^{-1/2}\boldsymbol{\Sigma}\mathbf{V}^{-1/2}$ and $\boldsymbol{\rho}^{-1} = (\mathbf{V}^{-1/2}\boldsymbol{\Sigma}\mathbf{V}^{-1/2})^{-1} = \mathbf{V}^{1/2}\boldsymbol{\Sigma}^{-1}\mathbf{V}^{1/2}$, the entry in the $(1, 1)$ position of $\boldsymbol{\rho}^{-1}$ is $\rho^{YY} = \sigma^{YY}\sigma_{YY}$.

Exercises

7.1. Given the data

z_1	10	5	7	19	11	8
y	15	9	3	25	7	13

fit the linear regression model $Y_j = \beta_0 + \beta_1 z_{j1} + \varepsilon_j$, $j = 1, 2, \ldots, 6$. Specifically, calculate the least squares estimates $\hat{\boldsymbol{\beta}}$, the fitted values $\hat{\mathbf{y}}$, the residuals $\hat{\boldsymbol{\varepsilon}}$, and the residual sum of squares, $\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$.

✓**7.2.** Given the data

z_1	10	5	7	19	11	18
z_2	2	3	3	6	7	9
y	15	9	3	25	7	13

fit the regression model

$$Y_j = \beta_1 z_{j1} + \beta_2 z_{j2} + \varepsilon_j, \qquad j = 1, 2, \ldots, 6.$$

to the *standardized* form (see page 412) of the variables y, z_1, and z_2. From this fit, deduce the corresponding fitted regression equation for the original (not standardized) variables.

✓**7.3.** (*Weighted least squares estimators.*) Let

$$\underset{(n\times 1)}{\mathbf{Y}} = \underset{(n\times(r+1))}{\mathbf{Z}}\ \underset{((r+1)\times 1)}{\boldsymbol{\beta}} + \underset{(n\times 1)}{\boldsymbol{\varepsilon}}$$

where $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ but $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\mathbf{V}$, with $\mathbf{V}(n \times n)$ known and positive definite. For $\mathbf{V}$ of full rank, show that the *weighted least squares* estimator is

$$\hat{\boldsymbol{\beta}}_W = (\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Y}$$

If σ^2 is unknown, it may be estimated, unbiasedly, by

$$(n - r - 1)^{-1} \times (\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}_W)'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}_W).$$

Hint: $\mathbf{V}^{-1/2}\mathbf{Y} = (\mathbf{V}^{-1/2}\mathbf{Z})\boldsymbol{\beta} + \mathbf{V}^{-1/2}\boldsymbol{\varepsilon}$ is of the classical linear regression form $\mathbf{Y}^* = \mathbf{Z}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*$, with $E(\boldsymbol{\varepsilon}^*) = \mathbf{0}$ and $E(\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*\prime}) = \sigma^2\mathbf{I}$. Thus, $\hat{\boldsymbol{\beta}}_W = \hat{\boldsymbol{\beta}}^* = (\mathbf{Z}^*\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime}\mathbf{Y}^*$.

✓**7.4.** Use the weighted least squares estimator in Exercise 7.3 to derive an expression for the estimate of the slope β in the model $Y_j = \beta z_j + \varepsilon_j$, $j = 1, 2, \ldots, n$, when (a) $\mathrm{Var}(\varepsilon_j) = \sigma^2$, (b) $\mathrm{Var}(\varepsilon_j) = \sigma^2 z_j$, and (c) $\mathrm{Var}(\varepsilon_j) = \sigma^2 z_j^2$. Comment on the manner in which the unequal variances for the errors influence the optimal choice of $\hat{\beta}_W$.

7.5. Establish (7-50): $\rho^2_{Y(\mathbf{Z})} = 1 - 1/\rho^{YY}$.
Hint: From (7-49) and Exercise 4.11

$$1 - \rho^2_{Y(\mathbf{Z})} = \frac{\sigma_{YY} - \boldsymbol{\sigma}'_{\mathbf{Z}Y}\boldsymbol{\Sigma}^{-1}_{\mathbf{Z}\mathbf{Z}}\boldsymbol{\sigma}_{\mathbf{Z}Y}}{\sigma_{YY}} = \frac{|\boldsymbol{\Sigma}_{\mathbf{Z}\mathbf{Z}}|\,(\sigma_{YY} - \boldsymbol{\sigma}'_{\mathbf{Z}Y}\boldsymbol{\Sigma}^{-1}_{\mathbf{Z}\mathbf{Z}}\boldsymbol{\sigma}_{\mathbf{Z}Y})}{|\boldsymbol{\Sigma}_{\mathbf{Z}\mathbf{Z}}|\,\sigma_{YY}} = \frac{|\boldsymbol{\Sigma}|}{|\boldsymbol{\Sigma}_{\mathbf{Z}\mathbf{Z}}|\,\sigma_{YY}}$$

From Result 2A.8(c), $\sigma^{YY} = |\boldsymbol{\Sigma}_{\mathbf{Z}\mathbf{Z}}|/|\boldsymbol{\Sigma}|$, where σ^{YY} is the entry of $\boldsymbol{\Sigma}^{-1}$ in the first row and first column. Since (see Exercise 2.23) $\boldsymbol{\rho} = \mathbf{V}^{-1/2}\boldsymbol{\Sigma}\mathbf{V}^{-1/2}$ and $\boldsymbol{\rho}^{-1} = (\mathbf{V}^{-1/2}\boldsymbol{\Sigma}\mathbf{V}^{-1/2})^{-1} = \mathbf{V}^{1/2}\boldsymbol{\Sigma}^{-1}\mathbf{V}^{1/2}$, the entry in the $(1,1)$ position of $\boldsymbol{\rho}^{-1}$ is $\rho^{YY} = \sigma^{YY}\sigma_{YY}$.

✓**7.6.** (*Generalized inverse of* $\mathbf{Z}'\mathbf{Z}$) A matrix $(\mathbf{Z}'\mathbf{Z})^-$ is called a generalized inverse of $\mathbf{Z}'\mathbf{Z}$ if $\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^-\mathbf{Z}'\mathbf{Z} = \mathbf{Z}'\mathbf{Z}$. Let $r_1 + 1 = \mathrm{rank}(\mathbf{Z})$ and suppose $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{r_1+1} > 0$ are the nonzero eigenvalues of $\mathbf{Z}'\mathbf{Z}$ with corresponding eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{r_1+1}$.

(a) Show that

$$(\mathbf{Z}'\mathbf{Z})^- = \sum_{i=1}^{r_1+1} \lambda_i^{-1}\mathbf{e}_i\mathbf{e}_i'$$

is a generalized inverse of $\mathbf{Z}'\mathbf{Z}$.

(b) The coefficients $\hat{\boldsymbol{\beta}}$ that minimize the sum of squared errors $(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})$ satisfy the *normal equations* $(\mathbf{Z}'\mathbf{Z})\hat{\boldsymbol{\beta}} = \mathbf{Z}'\mathbf{y}$. Show that these equations are satisfied for any $\hat{\boldsymbol{\beta}}$ such that $\mathbf{Z}\hat{\boldsymbol{\beta}}$ is the projection of $\mathbf{y}$ on the columns of $\mathbf{Z}$.

(c) Show that $\mathbf{Z}\hat{\boldsymbol{\beta}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^-\mathbf{Z}'\mathbf{y}$ is the projection of $\mathbf{y}$ on the columns of $\mathbf{Z}$. (See Footnote 2 in this chapter.)

(d) Show directly that $\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^-\mathbf{Z}'\mathbf{y}$ is a solution to the normal equations $(\mathbf{Z}'\mathbf{Z})[(\mathbf{Z}'\mathbf{Z})^-\mathbf{Z}'\mathbf{y}] = \mathbf{Z}'\mathbf{y}$.

Hint: (b) If $\mathbf{Z}\hat{\boldsymbol{\beta}}$ is the projection, then $\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\beta}}$ is perpendicular to the columns of $\mathbf{Z}$.

(d) The eigenvalue–eigenvector requirement implies that $(\mathbf{Z}'\mathbf{Z})(\lambda_i^{-1}\mathbf{e}_i) = \mathbf{e}_i$ for $i \leq r_1 + 1$ and $0 = \mathbf{e}_i'(\mathbf{Z}'\mathbf{Z})\mathbf{e}_i$ for $i > r_1 + 1$. Therefore, $(\mathbf{Z}'\mathbf{Z})(\lambda_i^{-1}\mathbf{e}_i)\mathbf{e}_i'\mathbf{Z}' = \mathbf{e}_i\mathbf{e}_i'\mathbf{Z}'$. Summing over i gives

$$(\mathbf{Z}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z})^-\mathbf{Z}' = \mathbf{Z}'\mathbf{Z}\left(\sum_{i=1}^{r_1+1} \lambda_i^{-1}\mathbf{e}_i\mathbf{e}_i'\right)\mathbf{Z}'$$

$$= \left(\sum_{i=1}^{r_1+1} \mathbf{e}_i\mathbf{e}_i'\right)\mathbf{Z}' = \left(\sum_{i=1}^{r+1} \mathbf{e}_i\mathbf{e}_i'\right)\mathbf{Z}' = \mathbf{I}\mathbf{Z}' = \mathbf{Z}'$$

since $\mathbf{e}_i'\mathbf{Z}' = \mathbf{0}$ for $i > r_1 + 1$.

7.7. Suppose the classical regression model is, with rank $(\mathbf{Z}) = r + 1$, written as

$$\underset{(n\times 1)}{\mathbf{Y}} = \underset{(n\times(q+1))}{\mathbf{Z}_1}\ \underset{((q+1)\times 1)}{\boldsymbol{\beta}_{(1)}} + \underset{(n\times(r-q))}{\mathbf{Z}_2}\ \underset{((r-q)\times 1)}{\boldsymbol{\beta}_{(2)}} + \underset{(n\times 1)}{\boldsymbol{\varepsilon}}$$

where $\mathrm{rank}(\mathbf{Z}_1) = q + 1$ and $\mathrm{rank}(\mathbf{Z}_2) = r - q$. If the parameters $\boldsymbol{\beta}_{(2)}$ are identified beforehand as being of primary interest, show that a $100(1 - \alpha)\%$ confidence region for $\boldsymbol{\beta}_{(2)}$ is given by

$$(\hat{\boldsymbol{\beta}}_{(2)} - \boldsymbol{\beta}_{(2)})'[\mathbf{Z}_2'\mathbf{Z}_2 - \mathbf{Z}_2'\mathbf{Z}_1(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{Z}_2](\hat{\boldsymbol{\beta}}_{(2)} - \boldsymbol{\beta}_{(2)}) \leq s^2(r - q)F_{r-q,n-r-1}(\alpha)$$

Hint: By Exercise 4.12, with 1's and 2's interchanged,

$$\mathbf{C}^{22} = [\mathbf{Z}_2'\mathbf{Z}_2 - \mathbf{Z}_2'\mathbf{Z}_1(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{Z}_2]^{-1}, \qquad \text{where } (\mathbf{Z}'\mathbf{Z})^{-1} = \begin{bmatrix} \mathbf{C}^{11} & \mathbf{C}^{12} \\ \mathbf{C}^{21} & \mathbf{C}^{22} \end{bmatrix}$$

Multiply by the square-root matrix $(\mathbf{C}^{22})^{-1/2}$, and conclude that $(\mathbf{C}^{22})^{-1/2}(\hat{\boldsymbol{\beta}}_{(2)} - \boldsymbol{\beta}_{(2)})/\sigma^2$ is $N(\mathbf{0}, \mathbf{I})$, so that

$$(\hat{\boldsymbol{\beta}}_{(2)} - \boldsymbol{\beta}_{(2)})'(\mathbf{C}^{22})^{-1}(\hat{\boldsymbol{\beta}}_{(2)} - \boldsymbol{\beta}_{(2)})\ is\ \sigma^2\chi^2_{r-q}.$$

✓**7.8.** Recall that the hat matrix is defined by $\mathbf{H} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ with diagonal elements h_{jj}.

(a) Show that $\mathbf{H}$ is an idempotent matrix. [See Result 7.1 and (7-6).]

(b) Show that $0 < h_{jj} < 1$, $j = 1, 2, \ldots, n$, and that $\sum_{j=1}^{n} h_{jj} = r + 1$, where r is the number of independent variables in the regression model. (In fact, $(1/n) \leq h_{jj} < 1$.)

(c) Verify, for the simple linear regression model with one independent variable z, that the leverage, h_{jj}, is given by

$$h_{jj} = \frac{1}{n} + \frac{(z_j - \bar{z})^2}{\sum_{j=1}^{n} (z_j - \bar{z})^2}$$

✓ **7.9.** Consider the following data on one predictor variable z_1 and two responses Y_1 and Y_2:

z_1	-2	-1	0	1	2
y_1	5	3	4	2	1
y_2	-3	-1	-1	2	3

Determine the least squares estimates of the parameters in the bivariate straight-line regression model

$$\begin{aligned} Y_{j1} &= \beta_{01} + \beta_{11} z_{j1} + \varepsilon_{j1} \\ Y_{j2} &= \beta_{02} + \beta_{12} z_{j1} + \varepsilon_{j2}, \qquad j = 1,2,3,4,5 \end{aligned}$$

Also, calculate the matrices of fitted values $\hat{\mathbf{Y}}$ and residuals $\hat{\boldsymbol{\varepsilon}}$ with $\mathbf{Y} = [\mathbf{y}_1 \mid \mathbf{y}_2]$. Verify the sum of squares and cross-products decomposition

$$\mathbf{Y}'\mathbf{Y} = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} + \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$$

✓**7.10.** Using the results from Exercise 7.9, calculate each of the following.

(a) A 95% confidence interval for the mean response $E(Y_{01}) = \beta_{01} + \beta_{11} z_{01}$ corresponding to $z_{01} = 0.5$

(b) A 95% prediction interval for the response Y_{01} corresponding to $z_{01} = 0.5$

(c) A 95% prediction region for the responses Y_{01} and Y_{02} corresponding to $z_{01} = 0.5$

7.11. *(Generalized least squares for multivariate multiple regression.)* Let $\mathbf{A}$ be a positive definite matrix, so that $d_j^2(\mathbf{B}) = (\mathbf{y}_j - \mathbf{B}'\mathbf{z}_j)'\mathbf{A}(\mathbf{y}_j - \mathbf{B}'\mathbf{z}_j)$ is a squared statistical distance from the jth observation $\mathbf{y}_j$ to its regression $\mathbf{B}'\mathbf{z}_j$. Show that the choice $\mathbf{B} = \hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$ minimizes the sum of squared statistical distances, $\sum_{j=1}^{n} d_j^2(\mathbf{B})$, for any choice of positive definite $\mathbf{A}$. Choices for $\mathbf{A}$ include $\boldsymbol{\Sigma}^{-1}$ and $\mathbf{I}$.
Hint: Repeat the steps in the proof of Result 7.10 with $\boldsymbol{\Sigma}^{-1}$ replaced by $\mathbf{A}$.

7.12. Given the mean vector and covariance matrix of Y, Z_1, and Z_2,

$$\boldsymbol{\mu} = \left[\begin{array}{c} \mu_Y \\ \hline \boldsymbol{\mu}_{\mathbf{Z}} \end{array}\right] = \left[\begin{array}{c} 4 \\ \hline 3 \\ -2 \end{array}\right] \quad \text{and} \quad \boldsymbol{\Sigma} = \left[\begin{array}{c|c} \sigma_{YY} & \boldsymbol{\sigma}'_{\mathbf{Z}Y} \\ \hline \boldsymbol{\sigma}_{\mathbf{Z}Y} & \boldsymbol{\Sigma}_{\mathbf{Z}\mathbf{Z}} \end{array}\right] = \left[\begin{array}{c|cc} 9 & 3 & 1 \\ \hline 3 & 2 & 1 \\ 1 & 1 & 1 \end{array}\right]$$

determine each of the following.

(a) The best linear predictor $\beta_0 + \beta_1 Z_1 + \beta_2 Z_2$ of Y

(b) The mean square error of the best linear predictor

(c) The population multiple correlation coefficient

(d) The partial correlation coefficient $\rho_{YZ_1 \cdot Z_2}$

(c) Verify, for the simple linear regression model with one independent variable z, that the leverage, h_{jj}, is given by

$$h_{jj} = \frac{1}{n} + \frac{(z_j - \bar{z})^2}{\sum_{j=1}^{n} (z_j - \bar{z})^2}$$

√ **7.9.** Consider the following data on one predictor variable z_1 and two responses Y_1 and Y_2:

z_1	−2	−1	0	1	2
y_1	5	3	4	2	1
y_2	−3	−1	−1	2	3

Determine the least squares estimates of the parameters in the bivariate straight-line regression model

$$\begin{aligned} Y_{j1} &= \beta_{01} + \beta_{11} z_{j1} + \varepsilon_{j1} \\ Y_{j2} &= \beta_{02} + \beta_{12} z_{j1} + \varepsilon_{j2}, \qquad j = 1, 2, 3, 4, 5 \end{aligned}$$

Also, calculate the matrices of fitted values $\hat{\mathbf{Y}}$ and residuals $\hat{\boldsymbol{\varepsilon}}$ with $\mathbf{Y} = [\mathbf{y}_1 \mid \mathbf{y}_2]$. Verify the sum of squares and cross-products decomposition

$$\mathbf{Y}'\mathbf{Y} = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} + \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$$

√**7.10.** Using the results from Exercise 7.9, calculate each of the following.

(a) A 95% confidence interval for the mean response $E(Y_{01}) = \beta_{01} + \beta_{11} z_{01}$ corresponding to $z_{01} = 0.5$

(b) A 95% prediction interval for the response Y_{01} corresponding to $z_{01} = 0.5$

(c) A 95% prediction region for the responses Y_{01} and Y_{02} corresponding to $z_{01} = 0.5$

7.11. *(Generalized least squares for multivariate multiple regression.)* Let $\mathbf{A}$ be a positive definite matrix, so that $d_j^2(\mathbf{B}) = (\mathbf{y}_j - \mathbf{B}'\mathbf{z}_j)'\mathbf{A}(\mathbf{y}_j - \mathbf{B}'\mathbf{z}_j)$ is a squared statistical distance from the jth observation $\mathbf{y}_j$ to its regression $\mathbf{B}'\mathbf{z}_j$. Show that the choice $\mathbf{B} = \hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$ minimizes the sum of squared statistical distances, $\sum_{j=1}^{n} d_j^2(\mathbf{B})$, for any choice of positive definite $\mathbf{A}$. Choices for $\mathbf{A}$ include $\boldsymbol{\Sigma}^{-1}$ and $\mathbf{I}$.
Hint: Repeat the steps in the proof of Result 7.10 with $\boldsymbol{\Sigma}^{-1}$ replaced by $\mathbf{A}$.

7.12. Given the mean vector and covariance matrix of Y, Z_1, and Z_2,

$$\boldsymbol{\mu} = \left[\begin{array}{c} \mu_Y \\ \hline \boldsymbol{\mu}_{\mathbf{Z}} \end{array}\right] = \left[\begin{array}{c} 4 \\ \hline 3 \\ -2 \end{array}\right] \quad \text{and} \quad \boldsymbol{\Sigma} = \left[\begin{array}{c:c} \sigma_{YY} & \boldsymbol{\sigma}'_{\mathbf{Z}Y} \\ \hline \boldsymbol{\sigma}_{\mathbf{Z}Y} & \boldsymbol{\Sigma}_{\mathbf{ZZ}} \end{array}\right] = \left[\begin{array}{c:cc} 9 & 3 & 1 \\ \hline 3 & 2 & 1 \\ 1 & 1 & 1 \end{array}\right]$$

determine each of the following.

(a) The best linear predictor $\beta_0 + \beta_1 Z_1 + \beta_2 Z_2$ of Y

(b) The mean square error of the best linear predictor

(c) The population multiple correlation coefficient

(d) The partial correlation coefficient $\rho_{YZ_1 \cdot Z_2}$

7.13. The test scores for college students described in Example 5.5 have

$$\bar{\mathbf{z}} = \begin{bmatrix} \bar{z}_1 \\ \bar{z}_2 \\ \bar{z}_3 \end{bmatrix} = \begin{bmatrix} 527.74 \\ 54.69 \\ 25.13 \end{bmatrix}, \qquad \mathbf{S} = \begin{bmatrix} 5691.34 & & \\ 600.51 & 126.05 & \\ 217.25 & 23.37 & 23.11 \end{bmatrix}$$

Assume joint normality.

(a) Obtain the maximum likelihood estimates of the parameters for predicting Z_1 from Z_2 and Z_3.

(b) Evaluate the estimated multiple correlation coefficient $R_{Z_1(Z_2,Z_3)}$.

(c) Determine the estimated partial correlation coefficient $R_{Z_1,Z_2 \cdot Z_3}$.

7.14. Twenty-five portfolio managers were evaluated in terms of their performance. Suppose Y represents the rate of return achieved over a period of time, Z_1 is the manager's attitude toward risk measured on a five-point scale from "very conservative" to "very risky," and Z_2 is years of experience in the investment business. The observed correlation coefficients between pairs of variables are

$$\mathbf{R} = \begin{array}{c} \begin{array}{ccc} Y & Z_1 & Z_2 \end{array} \\ \begin{bmatrix} 1.0 & -.35 & .82 \\ -.35 & 1.0 & -.60 \\ .82 & -.60 & 1.0 \end{bmatrix} \end{array}$$

(a) Interpret the sample correlation coefficients $r_{YZ_1} = -.35$ and $r_{YZ_2} = -.82$.

(b) Calculate the partial correlation coefficient $r_{YZ_1 \cdot Z_2}$ and interpret this quantity with respect to the interpretation provided for r_{YZ_1} in Part a.

The following exercises may require the use of a computer.

√**7.15.** Use the real-estate data in Table 7.1 and the linear regression model in Example 7.4.

(a) Verify the results in Example 7.4.

(b) Analyze the residuals to check the adequacy of the model. (See Section 7.6.)

(c) Generate a 95% prediction interval for the selling price (Y_0) corresponding to total dwelling size $z_1 = 17$ and assessed value $z_2 = 46$.

(d) Carry out a likelihood ratio test of $H_0: \beta_2 = 0$ with a significance level of $\alpha = .05$. Should the original model be modified? Discuss.

√**7.16.** Calculate a C_p plot corresponding to the possible linear regressions involving the real-estate data in Table 7.1.

7.17. Consider the *Forbes* data in Exercise 1.4.

(a) Fit a linear regression model to these data using profits as the dependent variable and sales and assets as the independent variables.

(b) Analyze the residuals to check the adequacy of the model. Compute the leverages associated with the data points. Does one (or more) of these companies stand out as an outlier in the set of independent variable data points?

(c) Generate a 95% prediction interval for profits corresponding to sales of 100 (billions of dollars) and assets of 500 (billions of dollars).

(d) Carry out a likelihood ratio test of $H_0: \beta_2 = 0$ with a significance level of $\alpha = .05$. Should the original model be modified? Discuss.

7.25. Amitriptyline is prescribed by some physicians as an antidepressant. However, there are also conjectured side effects that seem to be related to the use of the drug: irregular heartbeat, abnormal blood pressures, and irregular waves on the electrocardiogram, among other things. Data gathered on 17 patients who were admitted to the hospital after an amitriptyline overdose are given in Table 7.6. The two response variables are

Y_1 = Total TCAD plasma level (TOT)
Y_2 = Amount of amitriptyline present in TCAD plasma level (AMI)

The five predictor variables are

Z_1 = Gender: 1 if female, 0 if male (GEN)
Z_2 = Amount of antidepressants taken at time of overdose (AMT)
Z_3 = PR wave measurement (PR)
Z_4 = Diastolic blood pressure (DIAP)
Z_5 = QRS wave measurement (QRS)

Table 7.6 Amitriptyline Data

y_1 TOT	y_2 AMI	z_1 GEN	z_2 AMT	z_3 PR	z_4 DIAP	z_5 QRS
3389	3149	1	7500	220	0	140
1101	653	1	1975	200	0	100
1131	810	0	3600	205	60	111
596	448	1	675	160	60	120
896	844	1	750	185	70	83
1767	1450	1	2500	180	60	80
807	493	1	350	154	80	98
1111	941	0	1500	200	70	93
645	547	1	375	137	60	105
628	392	1	1050	167	60	74
1360	1283	1	3000	180	60	80
652	458	1	450	160	64	60
860	722	1	1750	135	90	79
500	384	0	2000	160	60	80
781	501	0	4500	180	0	100
1070	405	0	1500	170	90	120
1754	1520	1	3000	180	0	129

Source: See [24].

(a) Perform a regression analysis using only the first response Y_1.
  (i) Suggest and fit appropriate linear regression models.
  (ii) Analyze the residuals.
  (iii) Construct a 95% prediction interval for Total TCAD for $z_1 = 1$, $z_2 = 1200$, $z_3 = 140$, $z_4 = 70$, and $z_5 = 85$.

(b) Repeat Part a using the second response Y_2.

✓ **7.25.** Amitriptyline is prescribed by some physicians as an antidepressant. However, there are also conjectured side effects that seem to be related to the use of the drug: irregular heartbeat, abnormal blood pressures, and irregular waves on the electrocardiogram, among other things. Data gathered on 17 patients who were admitted to the hospital after an amitriptyline overdose are given in Table 7.6. The two response variables are

$$Y_1 = \text{Total TCAD plasma level (TOT)}$$
$$Y_2 = \text{Amount of amitriptyline present in TCAD plasma level (AMI)}$$

The five predictor variables are

$$Z_1 = \text{Gender: 1 if female, 0 if male (GEN)}$$
$$Z_2 = \text{Amount of antidepressants taken at time of overdose (AMT)}$$
$$Z_3 = \text{PR wave measurement (PR)}$$
$$Z_4 = \text{Diastolic blood pressure (DIAP)}$$
$$Z_5 = \text{QRS wave measurement (QRS)}$$

Table 7.6 Amitriptyline Data

y_1 TOT	y_2 AMI	z_1 GEN	z_2 AMT	z_3 PR	z_4 DIAP	z_5 QRS
3389	3149	1	7500	220	0	140
1101	653	1	1975	200	0	100
1131	810	0	3600	205	60	111
596	448	1	675	160	60	120
896	844	1	750	185	70	83
1767	1450	1	2500	180	60	80
807	493	1	350	154	80	98
1111	941	0	1500	200	70	93
645	547	1	375	137	60	105
628	392	1	1050	167	60	74
1360	1283	1	3000	180	60	80
652	458	1	450	160	64	60
860	722	1	1750	135	90	79
500	384	0	2000	160	60	80
781	501	0	4500	180	0	100
1070	405	0	1500	170	90	120
1754	1520	1	3000	180	0	129

Source: See [24].

(a) Perform a regression analysis using only the first response Y_1.
  - (i) Suggest and fit appropriate linear regression models.
  - (ii) Analyze the residuals.
  - (iii) Construct a 95% prediction interval for Total TCAD for $z_1 = 1$, $z_2 = 1200$, $z_3 = 140$, $z_4 = 70$, and $z_5 = 85$.

(b) Repeat Part a using the second response Y_2.

(c) Perform a multivariate multiple regression analysis using both responses Y_1 and Y_2.
  - (i) Suggest and fit appropriate linear regression models.
  - (ii) Analyze the residuals.
  - (iii) Construct a 95% prediction ellipse for both Total TCAD and Amount of amitriptyline for $z_1 = 1$, $z_2 = 1200$, $z_3 = 140$, $z_4 = 70$, and $z_5 = 85$. Compare this ellipse with the prediction intervals in Parts a and b. Comment.

7.26. Measurements of properties of pulp fibers and the paper made from them are contained in Table 7.7 (see also [19] and website: www.prenhall.com/statistics). There are $n = 62$ observations of the pulp fiber characteristics, z_1 = arithmetic fiber length, z_2 = long fiber fraction, z_3 = fine fiber fraction, z_4 = zero span tensile, and the paper properties, y_1 = breaking length, y_2 = elastic modulus, y_3 = stress at failure, y_4 = burst strength.

Table 7.7 Pulp and Paper Properites Data

y_1 BL	y_2 EM	y_3 SF	y_4 BS	z_1 AFL	z_2 LFF	z_3 FFF	z_4 ZST
21.312	7.039	5.326	.932	−.030	35.239	36.991	1.057
21.206	6.979	5.237	.871	.015	35.713	36.851	1.064
20.709	6.779	5.060	.742	.025	39.220	30.586	1.053
19.542	6.601	4.479	.513	.030	39.756	21.072	1.050
20.449	6.795	4.912	.577	−.070	32.991	36.570	1.049
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
16.441	6.315	2.997	−.400	−.605	2.845	84.554	1.008
16.294	6.572	3.017	−.478	−.694	1.515	81.988	.998
20.289	7.719	4.866	.239	−.559	2.054	8.786	1.081
17.163	7.086	3.396	−.236	−.415	3.018	5.855	1.033
20.289	7.437	4.859	.470	−.324	17.639	28.934	1.070

Source: See Lee [19].

(a) Perform a regression analysis using each of the response variables Y_1, Y_2, Y_3 and Y_4.
  - (i) Suggest and fit appropriate linear regression models.
  - (ii) Analyze the residuals. Check for outliers or observations with high leverage.
  - (iii) Construct a 95% prediction interval for SF (Y_3) for $z_1 = .330$, $z_2 = 45.500$, $z_3 = 20.375$, $z_4 = 1.010$.

(b) Perform a multivariate multiple regression analysis using all four response variables, Y_1, Y_2, Y_3 and Y_4, and the four independent variables, Z_1, Z_2, Z_3 and Z_4.
  - (i) Suggest and fit an appropriate linear regression model. Specify the matrix of estimated coefficients $\hat{\boldsymbol{\beta}}$ and estimated error covariance matrix $\hat{\boldsymbol{\Sigma}}$.
  - (ii) Analyze the residuals. Check for outliers.
  - (iii) Construct simultaneous 95% prediction intervals for the individual responses $Y_{0i}, i = 1, 2, 3, 4$, for the same settings of the independent variables given in part a (iii) above. Compare the simultaneous prediction interval for Y_{03} with the prediction interval in part a (iii). Comment.

7.27. Refer to the data on fixing breakdowns in cell phone relay towers in Table 6.20. In the initial design, experience level was coded as Novice or Guru. Now consider three levels of experience: Novice, Guru and Experienced. Some additional runs for an experienced engineer are given below. Also, in the original data set, reclassify Guru in run 3 as