b. Obtain the eigenvalues and the eigenvectors of the matrix A.

c. Determine the spectral representation of the matrix A.

d. Determine whether the matrix A can be a covariance matrix. If so, determine the corresponding correlation matrix.

3.5. Consider the matrix

$$A = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 4 & 2 \\ 1 & 2 & 2 \end{bmatrix}$$

a. Calculate the determinant and the inverse of the matrix A.

b. Obtain the eigenvalues and the eigenvectors of the matrix A.

c. Determine the spectral representation of the matrix A.

d. Determine whether the matrix A can be a covariance matrix. If so, determine the corresponding correlation matrix.

3.6. Consider the matrix

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 4 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

a. Calculate the determinant and the inverse of the matrix A.

b. Obtain the eigenvalues and the eigenvectors of the matrix A.

c. Determine whether or not the matrix A can be a covariance matrix. If so, determine the corresponding correlation matrix.

3.7. Consider the matrix

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 2 & 3 \\ 3 & 3 & 6 \end{bmatrix}$$

a. Calculate the determinant and the inverse of the matrix A.

b. Obtain the eigenvalues and the eigenvectors of the matrix A.

c. Determine whether the matrix A can be a covariance matrix. If so, determine the corresponding correlation matrix. Identify the linear combination that has variance zero.

3.8. Consider the matrices

$$A = \begin{bmatrix} 1 & 4 & 2 \\ 3 & 1 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 4 & 1 \\ 2 & 2 \\ 2 & 4 \end{bmatrix}$$

a. Determine the matrix product AB and obtain its rank.

b. Determine the matrix product BA and obtain its rank.

3.9. Obtain a 3×3 orthogonal matrix other than the trivial case of the identity matrix.

3.10. Specify a linear regression model for the hardness data in Table 1.1. Specify the 14×2 matrix X and the 14×1 vector of responses $\mathbf{y}$. Determine the 2×2 matrix $X'X$ and its inverse $(X'X)^{-1}$. Using matrix algebra, write down the expression for the least squares estimates in $\hat{\beta} = (X'X)^{-1}X'\mathbf{y}$.

3.11. Consider a trivariate normal distribution $\mathbf{y} = (y_1, y_2, y_3)'$ with mean vector $E(\mathbf{y}) = (2, 6, 4)'$ and covariance matrix

$$V(\mathbf{y}) = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & -1 \\ 1 & -1 & 3 \end{bmatrix}$$

a. Determine the marginal bivariate distribution of $(y_1, y_2)'$.

b. Determine the conditional bivariate distribution of $(y_1, y_2)'$, given that $y_3 = 5$.

3.12. Let z_1, z_2, z_3 be random variables with mean vector and covariance matrix

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}; \quad V = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Define the new variables

$$y_1 = z_1 + 2z_3; \quad y_2 = z_1 + z_2 - z_3;$$

$$y_3 = 2z_1 + z_2 + z_3 - 7$$

a. Find the mean vector and the covariance matrix of (y_1, y_2, y_3).

b. Find the mean and variance of $y = \frac{1}{3}(y_1 + y_2 + y_3)$.

3.13. Let X be an $n \times p$ matrix. Assume that the inverse $(X'X)^{-1}$ exists, and define

$A = (X'X)^{-1}X'$ and $H = XA$.

a. Show that (i) $HH = H$;
   (ii) $(I - H)(I - H) = (I - H)$; and
   (iii) $HX = X$.

b. Find (i) $A(I - H)$; (ii) $(I - H)A'$;
   (iii) $H(I - H)$; and (iv) $(I - H)'H'$.

3.14. Suppose that the covariance matrix of a vector y is $\sigma^2 I$, where I is an $n \times n$ identity matrix. Using the matrices A and H in Exercise 3.13, find the covariance matrix of

a. Ay

b. Hy

c. $(I - H)y$

d. $\begin{bmatrix} A \\ -- \\ I - H \end{bmatrix} y$

3.15. Consider the bivariate random vector with covariance matrix

$$A = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix};$$

$|\rho| < 1$ is the correlation coefficient

a. Show that the eigenvalues of the matrix A are given by $1 + \rho$ and $1 - \rho$.

b. Show that the normalized eigenvectors that correspond to these two eigenvalues are given by

$$p_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \quad \text{and} \quad p_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

c. Confirm the spectral representation of the covariance matrix A. That is, show that

$$P\Lambda P' = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{bmatrix}$$

$$\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} = A = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

d. Generate $n = 20$ independent random vectors from a bivariate normal distribution with mean vector zero and covariance matrix A.

The result in (c) helps you with the generation (simulation) of correlated random variables. Assume that you want to generate bivariate normal random variables (y_1, y_2) with covariance matrix A given previously. You can achieve this by generating independent random variables (x_1, x_2) with variances $1 + \rho$ and $1 - \rho$ and applying the transformation

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

The resulting random vector (y_1, y_2) has a bivariate normal distribution with covariance matrix A.

Most computer programs make it easy to generate univariate normal random variables, but they lack routines for simulating correlated random vectors. For generating multivariate normal random variables with a certain specified covariance matrix, one can use the spectral representation of the covariance matrix to determine the matrix that transforms independent random variables into correlated random variables.

3.16. The linear regression model in Chapter 2, $y = \beta_0 + \beta_1 x + \varepsilon$, assumes that the settings of the regressor variable are fixed (nonstochastic). In Section 2.9.2, we showed that the standard regression results still apply in the random x case, provided that the random components on the right-hand side of the model (the regressor x and the error ε) are independent. As this exercise now shows, difficulties arise if the error and the regressor are dependent.

Assume that the vector (ε, x) in the linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$ follows a bivariate normal distribution with mean vector $(0, \mu_x)$ and covariance matrix

$$V(\varepsilon, x) = \begin{bmatrix} \sigma_\varepsilon^2 & \rho_{\varepsilon x}\sigma_\varepsilon\sigma_x \\ \rho_{\varepsilon x}\sigma_\varepsilon\sigma_x & \sigma_x^2 \end{bmatrix}$$

$\rho_{\varepsilon x}$ is the correlation between the error and the random regressor.

a. Use the result on linear transformations of (jointly) normal random variables in Section 3.4 and show that the distribution

Consider the orthogonal transformation $z = P'y$ and its inverse

$$
\begin{aligned}
y = Pz &= \sum_{i=1}^{n} c_i z_i \\
&= \sum_{i=1}^{p+1-l} c_i z_i + \sum_{i=p+2-l}^{p+1} c_i z_i + \sum_{i=p+2}^{n} c_i z_i \\
&= \hat{\mu}_A + (\hat{\mu} - \hat{\mu}_A) + (y - \hat{\mu})
\end{aligned}
$$

where $\hat{\mu}_A$ is the projection of y on $L_A(X)$, and $\hat{\mu}$ is the projection of y on $L(X)$; $\hat{\mu} - \hat{\mu}_A$ is in $L(X)$ and perpendicular to $L_A(X)$; $\|y - \hat{\mu}\|^2 = \sum_{i=p+2}^{n} z_i^2 = S(\hat{\beta})$ and $\|\hat{\mu} - \hat{\mu}_A\|^2 = \sum_{i=p+2-l}^{p+1} z_i^2$.

Since $y \sim N(\mu, \sigma^2 I)$, it follows that $z = P'y \sim N(P'\mu, \sigma^2 I)$. Under the null hypothesis $A\beta = \mathbf{0}$, the mean vector

$$
P'\mu = \begin{bmatrix} P_1'\mu \\ P_2'\mu \\ P_3'\mu \end{bmatrix} = \begin{bmatrix} P_1'\mu \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}
$$

This is because under the null hypothesis $\hat{\mu} = \hat{\mu}_A$ is in $L_A(X)$ and the columns of P_2 are perpendicular to $L_A(X)$. In addition, $P_3'\mu = \mathbf{0}$ since the columns of P_3 are perpendicular to $L(X)$. Hence,

i. $z_1, z_2, \ldots, z_n$ are independent normal random variables with variance σ^2.

ii. $z_{p+2-l}, \ldots, z_{p+1}$ have zero means under the null hypothesis $A\beta = \mathbf{0}$.

iii. $z_{p+2}, \ldots, z_n$ have zero means under the original model, even if the null hypothesis is false.

Thus,

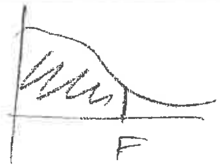
i. $\|\hat{\mu} - \hat{\mu}_A\|^2 / \sigma^2 = \sum_{p+2-l}^{p+1} z_i^2$ is the sum of l independent χ_1^2 random variables. It has a χ_l^2 distribution.

ii. $S(\hat{\beta})$ is a function of $z_{p+2}, \ldots, z_n$, whereas $\|\hat{\mu} - \hat{\mu}_A\|^2$ is a function of $z_{p+2-l}, \ldots, z_{p+1}$. Furthermore, $z_1, z_2, \ldots, z_n$ are independent. This shows that $S(\hat{\beta})$ and $\|\hat{\mu} - \hat{\mu}_A\|^2$ are independent. ■

EXERCISES

4.1. Consider the regression on time, $y_t = \beta_0 + \beta_1 t + \epsilon_t$, with $t = 1, 2, \ldots, n$. Here, the regressor vector is $x' = (1, 2, \ldots, n)$. Take $n = 10$. Write down the matrices $X'X$, $(X'X)^{-1}$, $V(\hat{\beta})$, and the variances of $\hat{\beta}_0$ and $\hat{\beta}_1$.

4.2. For the regression model $y_t = \beta_0 + \epsilon_t$ with $n = 2$ and $y' = (2, 4)$, draw the data in two-dimensional space. Identify the orthogonal projection of y onto $L(X) = L(\mathbf{1})$. Explain geometrically $\hat{\beta}_0$, $\hat{\mu}$, and e.



pf(F, df1, df2)

4.3. Consider the regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, 2, 3$. With

$$x = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \quad y = \begin{bmatrix} 2.2 \\ 3.9 \\ 3.1 \end{bmatrix}$$

draw the data in three-dimensional space and identify the orthogonal projection of y onto $L(X) = L(\mathbf{1}, x)$. Explain geometrically $\hat{\beta}$, $\hat{\mu}$, and e.

4.4. Consider the regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, 2, 3$. With

$$x = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \quad y = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$$

draw the data in three-dimensional space and identify the orthogonal projection of y onto $L(X) = L(\mathbf{1}, x)$. Explain geometrically $\hat{\beta}$, $\hat{\mu}$, and e.

4.5. After fitting the regression model,

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

on 15 cases, it is found that the mean square error $s^2 = 3$ and

$$(X'X)^{-1} = \begin{bmatrix} 0.5 & 0.3 & 0.2 & 0.6 \\ 0.3 & 6.0 & 0.5 & 0.4 \\ 0.2 & 0.5 & 0.2 & 0.7 \\ 0.6 & 0.4 & 0.7 & 3.0 \end{bmatrix}$$

Find

a. The estimate of $V(\hat{\beta}_1)$.

b. The estimate of $\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_3)$.

c. The estimate of $\mathrm{Corr}(\hat{\beta}_1, \hat{\beta}_3)$.

d. The estimate of $V(\hat{\beta}_1 - \hat{\beta}_3)$.

4.6. When fitting the model

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

to a set of $n = 15$ cases, we obtained the least squares estimates $\hat{\beta}_0 = 10$, $\hat{\beta}_1 = 12$, $\hat{\beta}_2 = 15$, and $s^2 = 2$. It is also known that

$$(X'X)^{-1} = \begin{bmatrix} 1 & 0.25 & 0.25 \\ 0.25 & 0.5 & -0.25 \\ 0.25 & -0.25 & 2 \end{bmatrix}$$

a. Estimate $V(\hat{\beta}_2)$.

b. Test the hypothesis that $\beta_2 = 0$.

c. Estimate the covariance between $\hat{\beta}_1$ and $\hat{\beta}_2$.

d. Test the hypothesis that $\beta_1 = \beta_2$, using both the t ratio and the 95% confidence interval.

e. The corrected total sum of squares, SST = 120. Construct the ANOVA table and test the hypothesis that $\beta_1 = \beta_2 = 0$. Obtain the percentage of variation in y that is explained by the model.

4.7. Consider a multiple regression model of the price of houses (y) on three explanatory variables: taxes paid (x_1), number of bathrooms (x_2), and square feet (x_3). The incomplete (Minitab) output from a regression on $n = 28$ houses is given as follows:

The regression equation is price = −10.7 + 0.190 taxes + 81.9 baths + 0.101 sqft

Predictor	Coef	SE Coef	t	p
Constant	−10.65	24.02		
taxes	0.18966	0.05623		
baths	81.87	47.82		
sqft	0.10063	0.03125		

Analysis of variance

Source	DF	SS	MS	F	p
Regression	3	504541			
Residual Error					
Total	27	541119			

a. Calculate the coefficient of determination R^2.

b. Test the null hypothesis that all three regression coefficients are zero (H_0: $\beta_1 = \beta_2 = \beta_3 = 0$). Use significance level 0.05.

c. Obtain a 95% confidence interval of the regression coefficient for "taxes." Can you simplify the model by dropping "taxes"? Obtain a 95% confidence interval of the regression coefficient for "baths." Can you simplify the model by dropping "baths"?

4.8. Continuation of Exercise 4.7. The incomplete (Minitab) output from a multiple regression