

**Tema: PCA**

1. Determine the population principal components  $Y_1$  and  $Y_2$  for the covariance matrix

$$\Sigma = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$$

Also, calculate the proportion of the total population variance explained by the first principal component.

2. Convert the covariance matrix in Exercise 1 to a correlation matrix  $\rho$ .
  - (a) Determine the principal components  $Y_1$  and  $Y_2$  from  $\rho$  and compute the proportion of total population variance explained by  $Y_1$ .
  - (b) Compare the components calculated in Part a with those obtained in Exercise 1. Are they the same? Should they be?
  - (c) Compute the correlations  $\rho_{Y_1, Z_1}$ ,  $\rho_{Y_1, Z_2}$ , and  $\rho_{Y_2, Z_1}$
3. Data on  $x_1$  = sales and  $x_2$  = profits for the 10 largest companies in the world were already listed. We know that

$$\bar{\mathbf{x}} = \begin{bmatrix} 155.60 \\ 14.70 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 7476.45 & 303.62 \\ 303.62 & 26.19 \end{bmatrix}$$

- (a) Determine the sample principal components and their variances for these data. (You may need the quadratic formula to solve for the eigenvalues of  $\mathbf{S}$ .)
  - (b) Find the proportion of the total sample variance explained by  $\hat{y}_1$ .
  - (c) Compute the correlation coefficients  $r_{\hat{y}_1, x_k}$ ,  $k = 1, 2$ . What interpretation, if any, can you give to the first principal component?
4. Convert the covariance matrix  $\mathbf{S}$  in Exercise 3 to a sample correlation matrix  $\mathbf{R}$ .
    - (a) Find the sample principal components  $\hat{y}_1, \hat{y}_2$  and their variances.
    - (b) Compute the proportion of the total sample variance explained by  $\hat{y}_1$ .
    - (c) Compute the correlation coefficients  $r_{\hat{y}_1, z_k}$ ,  $k = 1, 2$ . Interpret  $\hat{y}_1$
    - (d) Compare the components obtained in Part a with those obtained in Exercise 3 (a) do you feel that it is better to determine principal components from the sample covariance matrix or sample correlation matrix? Explain.
  5. The weekly rates of return for five stocks listed on the New York Stock Exchange are given in Table 1 ( `stock_data.DAT` ).
    - (a) Construct the sample covariance matrix  $\mathbf{S}$ , and find the sample principal components. (Note that the sample mean vector  $\bar{\mathbf{x}}$  is displayed in Example 8.5.)

	J P		Wells	Royal	Exxon
Week	Morgan	Citibank	Fargo	Dutch Shell	Mobil
1	0.01303	-0.00784	-0.00319	-0.04477	0.00522
2	0.00849	0.01669	-0.00621	0.01196	0.01349
3	-0.01792	-0.00864	0.01004	0	-0.00614
4	0.02156	-0.00349	0.01744	-0.02859	-0.00695
5	0.01082	0.00372	-0.01013	0.02919	0.04098
6	0.01017	-0.01220	-0.00838	0.01371	0.00299
7	0.01113	0.02800	0.00807	0.03054	0.00323
8	0.04848	-0.00515	0.01825	0.00633	0.00768
9	-0.03449	-0.01380	-0.00805	-0.02990	-0.01081
10	-0.00466	0.02099	-0.00608	-0.02039	-0.01267
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Table 1: Stock-Price Data (Weekly Rate of Return)

- (b) Determine the proportion of the total sample variance explained by the first three principal components. Interpret these components.
- (c) Given the results in Parts a-b, do you feel that the stock rates-of-return data can be summarized in fewer than five dimensions? Explain.
6. Consider the census-tract data listed in Table 2 (`census_tract.DAT`). Suppose the observations on  $X_5$  = median value home were recorded in ten thousands, rather than hundred thousands, of dollars; that is, multiply all the numbers listed in the sixth column of the table by 10.
- (a) Construct the sample covariance matrix  $S$  for the census-tract data when  $X_5$  = median value home is recorded in ten thousands of dollars.
- (b) Obtain the eigenvalue-eigenvector pairs and the first two sample principal components for the covariance matrix in Part a.
- (c) Compute the proportion of total variance explained by the first two principal components obtained in Part b. Calculate the correlation coefficients,  $r_{y_i, x_k}$ , and interpret these components if possible.

	Total population (thousands)	Professional degree (percent)	Employed age over 16 (percent)	Government employment (percent)	Median home value (\$100,000)
1	2.67	5.71	69.02	30.3	1.48
2	2.25	4.37	72.98	43.3	1.44
3	3.12	10.27	64.94	32.0	2.11
4	5.14	7.44	71.29	24.5	1.85
5	5.54	9.25	74.94	31.0	2.23
6	5.04	4.84	53.61	48.2	1.60
7	3.14	4.82	67.00	37.6	1.52
8	2.43	2.40	67.20	36.8	1.40
9	5.38	4.30	83.03	19.7	2.07
10	7.34	2.73	72.60	24.5	1.42
	⋮	⋮	⋮	⋮	⋮

Table 2: Census-track Data