Question

Suppose that 40% of the students in a large population are freshman, 30% are sophomores, 20% are juniors and 10% are seniors. Suppose that 10 students are selected at random from te population; and let X_1, X_2, X_3, X_4 denote, respectively, the numbers of freshmen, sophomores, juniors, and seniors that are obtained.

(a) Determine $\rho(X_i, X_j)$ for each pair of values i and j $(i < j)$.

(b) For what values of i and j $(i < j)$ is $\rho(X_i, X_j)$ most negative?

(c) For what values of i and j $(i < j)$ is $\rho(X_i, X_j)$ closest to 0?

Answer

	Prob.	Number
Freshman	$p_1 = 0.4$	x_1
Sophomore	$p_2 = 0.3$	x_2
Junior	$p_3 = 0.2$	x_3
Senior	$p_4 = 0.1$	x_4

n=10. $\text{var}(X_i) = np_i(1 - p_i) \quad \text{cov}(X_i, X_j) = -np_ip_j$

$\text{var}(X_1) = 2.4, \quad \text{var}(X_2) = 2.1, \quad \text{var}(X_3) = 1.6, \quad \text{var}(X_4) = 0.9$

We can form a matrix of variances and covariances.

$$\begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_3 \end{array} \begin{array}{c} \begin{array}{cccc} X_1 & X_2 & X_3 & X_4 \end{array} \\ \begin{pmatrix} 2.4 & -10(0.4)(0.3) & -10(0.4)(0.2) & -10(0.4)(0.1) \\ -1.2 & 2.1 & -10(0.3)(0.2) & -10(0.3)(0.1) \\ -0.8 & -0.6 & 1.6 & -10(0.2)(0.1) \\ -0.4 & -0.3 & -0.2 & 0.9 \end{pmatrix} \end{array}$$

Then we can find the correlations $\rho_{X_i,X_j} = \dfrac{\text{cov}\,(X_i, X_j)}{\sqrt{\text{var}\,(X_i)\,\text{var}\,(X_j)}}$

(a) $\begin{pmatrix} 1 & -0.534 & -0.408 & -0.272 \\ -0.534 & 1 & -0.327 & -0.218 \\ -0.408 & -0.327 & 1 & -0.167 \\ -0.272 & -0.218 & -0.167 & 1 \end{pmatrix}$ = correlation matrix

(b) $i = 1, \ j = 2$

(c) $i = 3, \ j = 4$