



Week 3

Univariate

Central tendency

- Mean
- Median
- Mode

Multivariate

→ Mean Vectors

Dispersion

Range
IOSR

Sbd. Dublor
(variance)

 Covariance matrix
 Correlation matrix

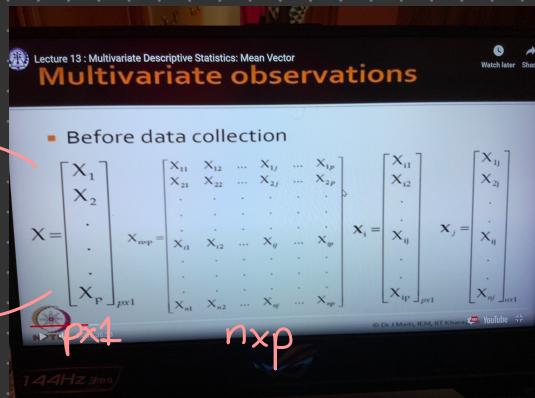
Simultaneous
Occurrence

$\chi_i = i^{\text{th}}$ Multivariate
observation
 $\mathbf{px 1}$

Rows are no of observations
 $= n$

Columns are number
of variables

J-1 P



x_{ij} = i^{th} observation of j^{th} variable
to be collected

e → Random Observation
→ Random Variable

$$x \rightarrow E(x) \rightarrow \mu$$

$$x_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{bmatrix}_{n \times 1}$$

D
looks like fixed value unlike the previous case

- After data collection

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_p \end{bmatrix}_{px1} \quad \mathbf{x}_{wp} = \begin{bmatrix} \mathbf{x}_{11} & \mathbf{x}_{12} & \dots & \mathbf{x}_{1p} & \dots & \mathbf{x}_{2p} \\ \mathbf{x}_{21} & \mathbf{x}_{22} & \dots & \mathbf{x}_{2p} & \dots & \mathbf{x}_{3p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{x}_{n1} & \mathbf{x}_{n2} & \dots & \mathbf{x}_{np} & \dots & \mathbf{x}_{np} \end{bmatrix} \quad \mathbf{x}_i = \begin{bmatrix} \mathbf{x}_{i1} \\ \mathbf{x}_{i2} \\ \vdots \\ \mathbf{x}_{ip} \end{bmatrix}_{px1} \quad \mathbf{x}_j = \begin{bmatrix} \mathbf{x}_{j1} \\ \mathbf{x}_{j2} \\ \vdots \\ \mathbf{x}_{jp} \end{bmatrix}_{px1}$$

What is the difference between data matrices before and after data collection?

18:09 / 36:53

Dr J Maiti, IEM, IIT Kharagpur

Lecture 13 : Multivariate Descriptive Statistics: Mean Vector Mean vector

$$\mathbf{X}_{12 \times 5} = \begin{pmatrix} 10 & 100 & 9 & 62 & 1 \\ 12 & 110 & 8 & 58 & 1.3 \\ 11 & 105 & 7 & 64 & 1.2 \\ 9 & 94 & 14 & 60 & 0.8 \\ 9 & 95 & 12 & 63 & 0.8 \\ 10 & 99 & 10 & 57 & 0.9 \\ 11 & 104 & 7 & 55 & 1 \\ 12 & 108 & 4 & 56 & 1.2 \\ 11 & 105 & 6 & 59 & 1.1 \\ 10 & 98 & 5 & 61 & 1 \\ 11 & 105 & 7 & 57 & 1.2 \\ 12 & 110 & 6 & 60 & 1.2 \end{pmatrix}$$

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} \quad \bar{\mathbf{X}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}_{px1} = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_{i1} \\ \frac{1}{n} \sum_{i=1}^n x_{i2} \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{ip} \end{bmatrix}_{px1}$$

$$\mu_j = E(X_j) = \begin{cases} \sum_{i=1}^n x_{ij} f(x_{ij}) & \text{for discrete } X_j \\ \int x_j f(x_j) dx_j & \text{for continuous } X_j \end{cases}$$

Lecture 13 : Multivariate Descriptive Statistics: Mean Vector

$$E(x) = \sum_{\text{all } x} x f(x) \quad \left| \begin{array}{l} \text{for } x \in \{x_1, x_2, \dots, x_n\} \\ -\infty \leq x \leq +\infty \end{array} \right.$$

$$E(x_j) = \sum_{\text{all } x_j} x_j f(x_j) \quad \left| \begin{array}{l} \text{for } x_j \in \{x_{j1}, x_{j2}, \dots, x_{jn}\} \\ -\infty \leq x_j \leq +\infty \end{array} \right. \quad \text{discrete}$$

$$M = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}_{px1} = \begin{bmatrix} E(x_1) \\ E(x_2) \\ \vdots \\ E(x_p) \end{bmatrix}_{px1} = \int_{-\infty}^{+\infty} x_j f(x_j) dx_j \quad \left| \begin{array}{l} j=1, 2, \dots, p \\ -\infty \leq x_j \leq +\infty \end{array} \right.$$

23:32 / 36:53

144Hz 3ms

Lecture 13 : Multivariate Descriptive Statistics: Mean Vector

$$\bar{x} = \hat{M} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}_{px1} = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_{i1} \\ \frac{1}{n} \sum_{i=1}^n x_{i2} \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_{ip} \end{bmatrix}_{px1}$$

Lecture 13 : Multivariate Descriptive Statistics: Mean Vector

$$\bar{x}_{px1} = \begin{bmatrix} \bar{x}_{11} & \bar{x}_{12} & \dots & \bar{x}_{1p} \\ \bar{x}_{21} & \bar{x}_{22} & \dots & \bar{x}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_{n1} & \bar{x}_{n2} & \dots & \bar{x}_{np} \end{bmatrix}_{pxp} \Rightarrow \bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}_{px1} \quad \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

30:22 / 36:53

Covariance Matrix [lec 2]

From population point of view,

Population Covariance Matrix

$$x \rightarrow V(x) = E(x - \mu)^2 = \sum (x - \mu)^2 f(x)$$

$$= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \sigma^2$$

$$\text{Var}_j = V(x_j) = E[(x_j - \mu_j)^2] = \sum_{\text{all } x_j} (x_j - \mu_j)^2 f(x_j) \leftarrow \text{Discrete}$$

Continuous

$$= \int_{-\infty}^{\infty} (x_j - \mu_j)^2 f(x_j) dx_j \quad \checkmark$$

$$\text{For } j=1 \rightarrow \sigma_1^2$$

$$j=2 \rightarrow \sigma_2^2$$

$$\vdots$$
$$j=P \rightarrow \sigma_P^2$$

If x_1, x_2 are not independent,

When x_1 varies, x_2 covaries

$$V(x_j) = E(x_j - \mu_j)^2$$



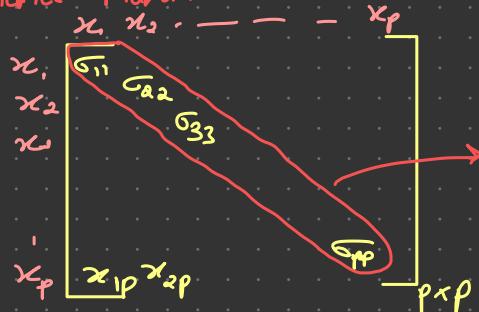
$$\sigma_{jk} = \text{Cov}(x_j, x_k) = E((x_j - \mu_j)(x_k - \mu_k))$$
$$= \sum_{\text{all } j, k} (x_j - \mu_j)(x_k - \mu_k) f(x_j, x_k)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_j - \mu_j) (x_k - \mu_k) f(x_j, x_k) dx_j dx_k$$

$$\sigma_j^2 = \sigma_{jj} \rightarrow \text{Variance of } x_j$$

σ_{jk}^2 : Covariance between x_j, x_k

Population Covariance Matrix



Diagonal: Describe variance
Rest: Describe covariance

$$\sigma_{jk} = \sigma_{kj}$$

\sum
↓
population covariance matrix

Lecture 14: Multivariate Descriptive Statistics: Covariance Matrix

Population covariance matrix

$\sigma_{ij}^2 = \sigma_{ii} = E[(x_i - \mu_i)^2]$ = $\begin{cases} \sum_{\text{all } x_i} (x_i - \mu_i)^2 f(x_i) & \text{for discrete } X_i \\ \int_{-\infty}^{\infty} (x_i - \mu_i)^2 f(x_i) dx_i & \text{for continuous } X_i \end{cases}$

$\text{Cov}(X_i, X_k) = \sigma_{ik} = E[(X_i - \mu_i)(X_k - \mu_k)]$
 $= \sum_{\text{all } x_i} \sum_{\text{all } x_k} (x_i - \mu_i)(x_k - \mu_k) f_{pk}(x_i, x_k)$

$\text{Cov}(X_j, X_k) = \sigma_{jk} = E[(X_j - \mu_j)(X_k - \mu_k)]$
 $= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_j - \mu_j)(x_k - \mu_k) f_{jk}(x_j, x_k) dX_j dX_k$

$\sum_{\text{all } x_i} \sum_{\text{all } x_k} \dots \sum_{\text{all } x_p}$ $\begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1j} & \dots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \dots & \sigma_{2j} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma_{1j} & \sigma_{2j} & \dots & \sigma_{jj} & \dots & \sigma_{jp} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \dots & \sigma_{jp} & \dots & \sigma_{pp} \end{bmatrix}_{p \times p}$

Deej Mani, IIM-BT Kharagpur

144Hz 3ms

Sample Covariance Matrix

Size: $p \times p$

- Methods
- Why to appro

Lecture 14 : Multivariate Descriptive Statistics: Covariance Matrix

Sample covariance matrix

$$S = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1j} & \dots & s_{1p} \\ s_{12} & s_{22} & \dots & s_{2j} & \dots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ s_{ij} & s_{2j} & \dots & s_{jj} & \dots & s_{jp} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ s_{1p} & s_{2p} & \dots & s_{jp} & \dots & s_{pp} \end{bmatrix}_{p \times p}$$

$$s_{jj} = s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)^2$$

$$s_{ji} = s_{ij} = \frac{1}{n-1} (\mathbf{x}_j - \bar{\mathbf{x}}_j)^T (\mathbf{x}_i - \bar{\mathbf{x}}_i)$$

$$s_{jk} = s_{kj} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j) (\mathbf{x}_{ik} - \bar{\mathbf{x}}_k)$$

$$s_{jk} = \frac{1}{n-1} (\mathbf{x}_j - \bar{\mathbf{x}}_j)^T (\mathbf{x}_k - \bar{\mathbf{x}}_k)$$

$$s_{\mu\nu} = \frac{1}{n-1} \left[(\mathbf{x} - \bar{\mathbf{x}})^T \right]_{p \times n} \left[(\mathbf{x} - \bar{\mathbf{x}})^T \right]_{n \times p}$$

144Hz 3ms

Dr J Maiti, IIT Kharagpur

Can be modified and written
as

$$s_{ij} = \frac{1}{n-1} \sum_{k=1}^n \mathbf{x}_{ik} \mathbf{x}_{jk}^T$$

Lecture 14 : Multivariate Descriptive Statistics: Covariance Matrix

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{np} \end{bmatrix}_{n \times p}$$

Subtract mean

$$\mathbf{X}^* = \begin{bmatrix} x_{11}-\bar{x}_1 & x_{12}-\bar{x}_2 & \dots & x_{1j}-\bar{x}_j & \dots & x_{1p}-\bar{x}_p \\ x_{21}-\bar{x}_1 & x_{22}-\bar{x}_2 & \dots & x_{2j}-\bar{x}_j & \dots & x_{2p}-\bar{x}_p \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1}-\bar{x}_1 & x_{n2}-\bar{x}_2 & \dots & x_{nj}-\bar{x}_j & \dots & x_{np}-\bar{x}_p \end{bmatrix}_{n \times p}$$

where $x_{ij}^* = x_{ij} - \bar{x}_j$

144Hz 3ms

Lec 3

Lecture 15 : Multivariate Descriptive Statistics: Correlation Matrix

Correlation matrix

Let, $s_{ij} = \frac{x_{ij} - \bar{x}_i}{\sqrt{s_{ii}}} = \frac{x_{ij} - \bar{x}_i}{\sqrt{s_{ii}}}$

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_{11} & \mathbf{x}_{12} & \dots & \mathbf{x}_{1p} \\ \mathbf{x}_{21} & \mathbf{x}_{22} & \dots & \mathbf{x}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{n1} & \mathbf{x}_{n2} & \dots & \mathbf{x}_{np} \end{bmatrix}_{n \times p}$$

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1p} & \dots & r_{1p} \\ r_{21} & r_{22} & \dots & r_{2p} & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & r_{pp} & \dots & r_{pp} \end{bmatrix}_{p \times p}$$

$$(n-1)\mathbf{R} = \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}$$

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}} = \frac{s_{ij}}{s_{ii}s_{jj}}$$

144Hz 3ms

Lecture 15 : Multivariate Descriptive Statistics: Correlation Matrix

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & \dots & x_p \\ x_{11} & x_{22} & \dots & x_{pp} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}_{n \times p}$$

$$x_{ij}^* = x_{ij} - \bar{x}_j$$

$$\mathbf{X}^* = \begin{bmatrix} x_{11}^* & x_{21}^* & \dots & x_{p1}^* \\ x_{21}^* & x_{22}^* & \dots & x_{2p}^* \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1}^* & x_{n2}^* & \dots & x_{np}^* \end{bmatrix}_{n \times p}$$

$$\mathbf{X}^* \mathbf{X}^{*T} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & \ddots \end{bmatrix}_{p \times p}$$

$$= (n-1) \mathbf{S}^*$$

$$\mathbf{S}^* = \mathbf{Cov}(\mathbf{X})$$

144Hz 3ms

Example:

Lecture 15: Multivariate Descriptive Statistics: Correlation Matrix

$$X = \begin{bmatrix} 10 & 100 \\ 12 & 110 \\ 11 & 105 \end{bmatrix}_{3 \times 2}$$

$$\bar{X} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = \begin{bmatrix} 11 \\ 105 \end{bmatrix}$$

$$X^T X^* = \begin{bmatrix} -1 & 1 & 0 \\ -5 & 5 & 0 \\ 2 & 2 & 0 \end{bmatrix}_{(2 \times 3) \times 3} \begin{bmatrix} 1 & -1 & -5 \\ 1 & 1 & 5 \\ 0 & 0 & 0 \end{bmatrix}_{3 \times 2}$$

$$X^* = \begin{bmatrix} 10-11 & 100-105 \\ 12-11 & 110-105 \\ 11-11 & 105-105 \end{bmatrix} = \begin{bmatrix} 2 & 10 \\ 10 & 50 \end{bmatrix}_{2 \times 2}$$

$$(M-1)S = X^* X^* = \begin{bmatrix} 2 & 10 \\ 10 & 50 \end{bmatrix}_{2 \times 2}$$

$$S = \frac{1}{2} \begin{bmatrix} 2 & 10 \\ 10 & 50 \end{bmatrix}$$

YouTube

924 / 3737

44Hz 3ms



Lecture 15: Multivariate Descriptive Statistics: Correlation Matrix

$$S = \begin{bmatrix} 1 & 5 \\ 5 & 25 \end{bmatrix} \quad \bar{X} = \begin{bmatrix} 11 \\ 105 \end{bmatrix}$$

$$\lambda_{11} = 1 \quad \lambda_{22} = 25^{-1} \quad X_{2n} \sim N_2(\mu, \Sigma)$$

$$\lambda_{12} = 1 \quad \lambda_{12}^{-1} = 25 \quad \hat{\mu} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = \bar{X}$$

$$\lambda_{21} = 1 \quad \lambda_{21}^{-1} = 5 \quad \hat{\Sigma} = \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} \\ \hat{\sigma}_{12} & \hat{\sigma}_{22} \end{bmatrix} = S$$

* Mean vector
* Covariance matrix
* Correlation matrix

144Hz 3ms

Correlation Matrix

Population Correlation Matrix $\rho_{p \times p}$

$$\rho_{p \times p} = \begin{bmatrix} 1 & \dots & \rho_{1p} \\ \vdots & \ddots & \vdots \\ \rho_{p1} & \dots & 1 \end{bmatrix}$$

Covariance

$$\sum = \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \dots & \sigma_{pp} \end{bmatrix}$$

$$\text{Correlation}(x_j, x_k) = \frac{\text{Covariance}(x_j, x_k)}{\text{Std}(x_j, x_k)}$$

$$\text{Correlation}(x_j, x_k) = \frac{\text{Cov}(x_j, x_k)}{\sqrt{\sigma_j} \cdot \sqrt{\sigma_k}}$$

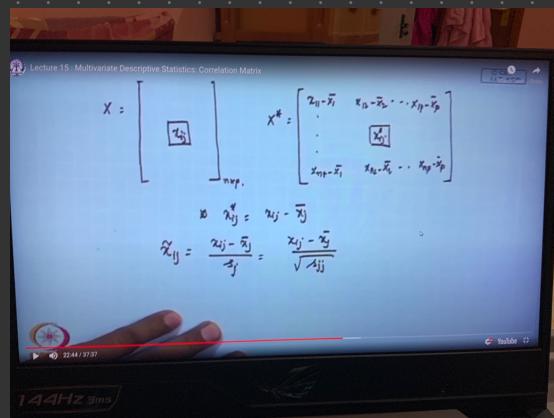
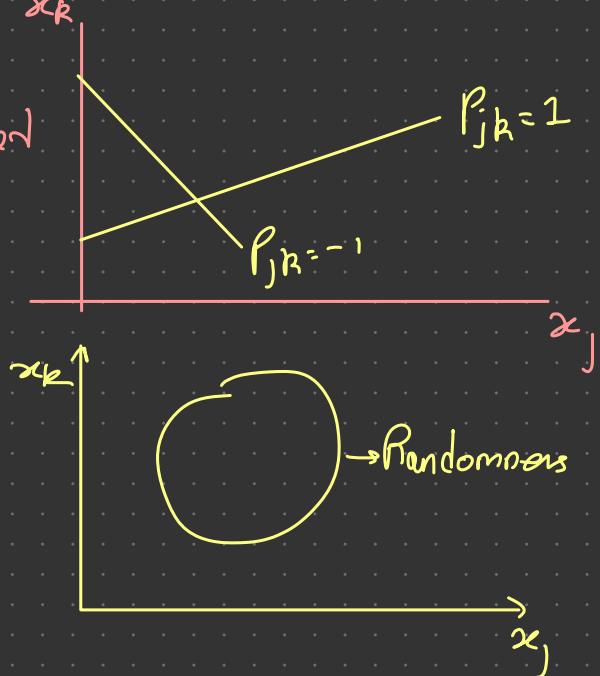
$$\rho_{jk} = \frac{\sigma_{jk}}{\sqrt{\sigma_j} \sqrt{\sigma_k}} \leftarrow \text{For } j=k,$$

$$\rho_{jj} = \frac{\sigma_{jj}}{\sqrt{\sigma_j} \sqrt{\sigma_j}} = 1$$

$\rho_{jk} = 1 \rightarrow$ Positively Correlated x_k

$\rho_{jk} = -1 \rightarrow$ Negatively correlated

$\rho_{jk} = 0 \rightarrow$ No correlation



$$\tilde{X} = \begin{bmatrix} \frac{x_1 - \bar{x}_1}{\sqrt{s_{11}}} & \frac{x_2 - \bar{x}_2}{\sqrt{s_{22}}} & \dots & \frac{x_p - \bar{x}_p}{\sqrt{s_{pp}}} \\ \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_{11}}} & \frac{\bar{x}_2 - \bar{x}_3}{\sqrt{s_{22}}} & \dots & \frac{\bar{x}_{p-1} - \bar{x}_p}{\sqrt{s_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\bar{x}_{n1} - \bar{x}_1}{\sqrt{s_{11}}} & \frac{\bar{x}_{n2} - \bar{x}_2}{\sqrt{s_{22}}} & \dots & \frac{\bar{x}_{np} - \bar{x}_p}{\sqrt{s_{pp}}} \end{bmatrix}_{p \times p}$$

$$\tilde{X}_{p \times n}^T \tilde{X}_{n \times p} = (n-1) R$$

\checkmark
 $p \times p$

Multivariate Descriptive Statistics: Correlation Matrix

$$X_{\text{P.M.}}^T \cdot X_{\text{P.M.}} = (n-1) R$$

$$\tilde{X} = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} \\ \tilde{x}_{21} & \tilde{x}_{22} \\ \tilde{x}_{31} & \tilde{x}_{32} \\ \tilde{x}_{41} & \tilde{x}_{42} \end{bmatrix}$$

$$X = \begin{bmatrix} \bar{x}_{11} - \tilde{x}_{11} \\ \bar{x}_{21} - \tilde{x}_{21} \\ \bar{x}_{31} - \tilde{x}_{31} \\ \bar{x}_{41} - \tilde{x}_{41} \\ \bar{x}_{12} - \tilde{x}_{12} \\ \bar{x}_{22} - \tilde{x}_{22} \\ \bar{x}_{32} - \tilde{x}_{32} \\ \bar{x}_{42} - \tilde{x}_{42} \end{bmatrix}$$

Figure 15: Multivariate Descriptive Statistics: Correlation Matrix

$$\tilde{X}^T \cdot \tilde{X} = \begin{bmatrix} \frac{\sum_{i=1}^n (\tilde{x}_{11} - \tilde{\bar{x}}_1)^2}{n-1} & \frac{\sum_{i=1}^n (\tilde{x}_{11} - \tilde{\bar{x}}_1)(\tilde{x}_{12} - \tilde{\bar{x}}_2)}{n-1} & \dots & \frac{\sum_{i=1}^n (\tilde{x}_{11} - \tilde{\bar{x}}_1)(\tilde{x}_{41} - \tilde{\bar{x}}_4)}{n-1} \\ \frac{\sum_{i=1}^n (\tilde{x}_{12} - \tilde{\bar{x}}_2)(\tilde{x}_{11} - \tilde{\bar{x}}_1)}{n-1} & \frac{\sum_{i=1}^n (\tilde{x}_{12} - \tilde{\bar{x}}_2)^2}{n-1} & \dots & \frac{\sum_{i=1}^n (\tilde{x}_{12} - \tilde{\bar{x}}_2)(\tilde{x}_{42} - \tilde{\bar{x}}_4)}{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sum_{i=1}^n (\tilde{x}_{41} - \tilde{\bar{x}}_4)(\tilde{x}_{11} - \tilde{\bar{x}}_1)}{n-1} & \frac{\sum_{i=1}^n (\tilde{x}_{41} - \tilde{\bar{x}}_4)(\tilde{x}_{12} - \tilde{\bar{x}}_2)}{n-1} & \dots & \frac{\sum_{i=1}^n (\tilde{x}_{41} - \tilde{\bar{x}}_4)^2}{n-1} \end{bmatrix}$$

$$A_{ij} = \frac{1}{n-1} \sum_{i=1}^n (\tilde{x}_{ij} - \tilde{\bar{x}}_j)^2$$

$$R = \begin{bmatrix} 1 & \frac{\sum_{i=1}^n (\tilde{x}_{11} - \tilde{\bar{x}}_1)(\tilde{x}_{12} - \tilde{\bar{x}}_2)}{n-1} & \dots & \frac{\sum_{i=1}^n (\tilde{x}_{11} - \tilde{\bar{x}}_1)(\tilde{x}_{41} - \tilde{\bar{x}}_4)}{n-1} \\ \frac{\sum_{i=1}^n (\tilde{x}_{12} - \tilde{\bar{x}}_2)(\tilde{x}_{11} - \tilde{\bar{x}}_1)}{n-1} & 1 & \dots & \frac{\sum_{i=1}^n (\tilde{x}_{12} - \tilde{\bar{x}}_2)(\tilde{x}_{42} - \tilde{\bar{x}}_4)}{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sum_{i=1}^n (\tilde{x}_{41} - \tilde{\bar{x}}_4)(\tilde{x}_{11} - \tilde{\bar{x}}_1)}{n-1} & \frac{\sum_{i=1}^n (\tilde{x}_{41} - \tilde{\bar{x}}_4)(\tilde{x}_{12} - \tilde{\bar{x}}_2)}{n-1} & \dots & 1 \end{bmatrix}$$

$$\boxed{\gamma_{ij} = \frac{s_{ij}}{s_i s_j}} \rightarrow \gamma_{11} = 1$$

Lecture - 4

$$S = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \dots & s_{pp} \end{bmatrix} \quad R = \begin{bmatrix} 1 & \gamma_{12} & \dots & \gamma_{1p} \\ \gamma_{12} & 1 & \dots & \gamma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{1p} & \gamma_{2p} & \dots & 1 \end{bmatrix}$$

$$D_S = \begin{bmatrix} s_{11} & 0 & 0 & \dots & 0 \\ r & s_{22} & & & \\ 0 & & s_{33} & & \\ \vdots & & & \ddots & \\ 0 & & & & s_{pp} \end{bmatrix}_{p \times p}$$

$$R = D_S^{-1/2} S D^{-1/2} \quad S = D_S^{1/2} R D_S^{1/2}$$

Sum Square & Cross Product Matrix [SSCP]

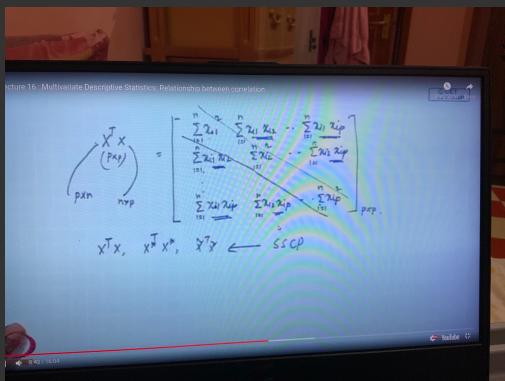
$$(n-1) S = \mathbf{X}^* \mathbf{X}^*$$

$$(n-1) R = \mathbf{X}^T \mathbf{X}$$

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix}$$

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} \sum_{i=1}^3 x_{11}^2 & \sum_{i=1}^3 x_{12} x_{12} \\ \sum_{i=1}^3 x_{11} x_{12} & \sum_{i=1}^3 x_{12}^2 \end{bmatrix}_{p \times p}$$

Can be expanded for $p \times p$



Sum products for variance
Cross products for covariance

