Chapter-3

- Given my data matrix is

$$\underset{n\times p}{\underline{X}} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & & \ddots & \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

- $E(S_n) = \frac{n-1}{n} \Sigma$

$$= \Sigma - \frac{1}{n}\Sigma$$

So $\frac{n}{n-1} S_n$ is an unbiased estimator of Σ

Sample Mean Vector, Covariance and Correlation as matrix Operation:

Let $\underset{n\times p}{\underline{X}}$ is a given data matrix, then

1) Sample Mean Matrix: $\underset{p\times 1}{\underset{\sim}{\bar{X}}} = \frac{1}{n} \underline{X}^T \mathbb{1}$

2) Sample Variance matrix: $S = \frac{1}{n-1} \underline{X}'\left(I - \frac{1}{n}\mathbb{1}\mathbb{1}'\right)\underline{X}$

$$S_n = \frac{1}{n} \underline{X}'\left(I - \frac{1}{n}\mathbb{1}\mathbb{1}'\right)\underline{X}$$

3.> Sample Standard deviation Matrix :

$$D^{1/2} = \begin{bmatrix} \sqrt{S_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{S_{22}} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{S_{pp}} \end{bmatrix}$$

4.> Sample correlation Matrix

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

5.> Sample variance matrix, If R is given

$$S = D^{1/2} R D^{1/2}$$

- $A^T A = A \implies A$ is Idempotent matrix.

<u>Sample Values of Linear Combinations of Variable</u> :

Consider a linear combination of vectors

$$\underset{\sim}{C}^T \underset{\sim}{X} = c_1 \underset{\sim}{X}_1 + c_2 \underset{\sim}{X}_2 + \cdots + c_p \underset{\sim}{X}_p$$

where $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_p$ are independent realizations from the random vector.

Recall : $E(\underset{\sim}{C}^T \underset{\sim}{X}) = \underset{\sim}{C}^T \underset{\sim}{\mu}$ & $Var(\underset{\sim}{C}^T \underset{\sim}{X}) = \underset{\sim}{C}^T \Sigma C$

1.) An estimate of the mean $\underset{\sim}{c}^T \underset{\sim}{\mu}$ is $\underset{\sim}{c}^T \underset{\sim p\times 1}{\bar{X}}$

2.) An estimate of the variance $\underset{\sim}{c}^T \Sigma \underset{\sim}{c}$ is $\underset{\sim}{c}^T S_n \underset{\sim}{c}$

3.) An estimate for the population covariance matrix of $\underset{\sim}{b}'\underset{\sim}{X}, \underset{\sim}{c}'\underset{\sim}{X}$
Which is $\operatorname{Cov}(\underset{\sim}{b}^T \underset{\sim}{X}, \underset{\sim}{c}^T \underset{\sim}{X}) = \underset{\sim}{b}^T \Sigma (\underset{\sim}{c}^T)^T$

$$= \underset{\sim}{b}^T \Sigma \underset{\sim}{c} \text{ is } \underset{\sim}{b}^T S_n \underset{\sim}{c}.$$

Chapter - 3 Idempotent matrix

From book

→ Random sampling implies

1.) measurment taken on different items (or trials) are unrelated to one another and

2.) The joint distribution of all p variables remains the same for all items

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{bmatrix} \begin{array}{l} \leftarrow \text{1st (multivariate) observation} \\ \\ \\ \leftarrow n^{th} \text{ (multivariate) observation} \end{array}$$

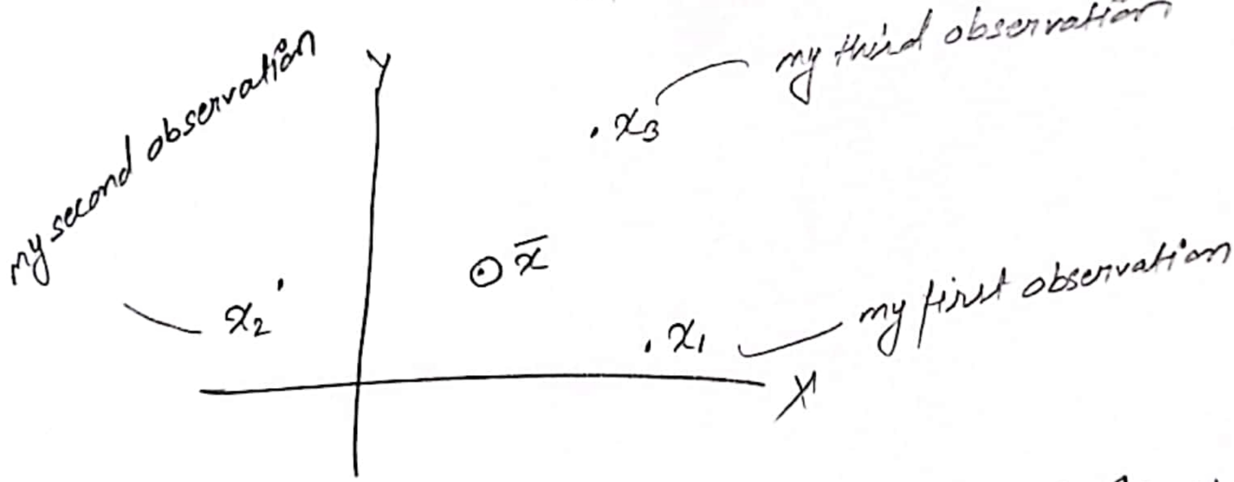
→ The rows of X represent n points in p-dimensional space (because n rows $\exists n$)

→ If the points are regarded as solid spheres, the sample mean vector $\bar{x}$, is the center of balance.

→ variability occurs in more than one direction and it is quantified by the sample variance-covariance matrix S_n.

→ A single numerical measure of variability is provided by the determinant of the sample variance-covariance matrix.

→ $\bar{X}$ लाई centre of balance भन्दछौं जुन $\bar{x}$ लाई पनि उही dimensional space मा लेख्न सकिन्छ पर्छ (मतलब no. of variable for $\bar{x}$ should be same as data point) $\frac{\downarrow + \downarrow + \downarrow + \downarrow}{\#}$



→ Elements of columns of the data matrix are the co-ordinate of vector (Another approach)

$$\begin{array}{cc} X & Y \\ \end{array}$$
$$\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}$$

→ If the two vectors are nearly perpendicular, the sample correlation will be approximately zero. If the two vectors are oriented in nearly opposite directions, the sample correlation will be close to −1.

→ Let $X_1, X_2, \ldots, X_n$ be a random sample from a joint distribution that has mean vector μ and covariance matrix Σ. Then $\bar{X}$ is an unbiased estimator of μ, and its covariance matrix is $\frac{1}{n}\Sigma$ that is

$$E(\bar{X}) = \mu$$

$$\mathrm{cov}(\bar{X}) = \frac{1}{n}\Sigma$$ — Remember.

For the covariance matrix S_n

$$E(S_n) = \frac{n-1}{n}\Sigma$$

$$= \Sigma - \frac{1}{n}\Sigma$$

Thus

$$E\left(\frac{n}{n-1} S_n\right) = \Sigma$$

So $[n/(n-1)]S_n$ is an unbiased estimator of Σ, while S_n is biased estimator with $\text{bias} = E(S_n) - \Sigma$

$$= \Sigma - \frac{1}{n}\Sigma - \Sigma$$

$$= -\left(\frac{1}{n}\right)\Sigma.$$

→ Proof given $X_1, X_2, \ldots, X_n$ are random sample

Now, $\overline{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$

so ① $E(\overline{X}) = E\left(\dfrac{X_1 + X_2 + \cdots + X_n}{n}\right)$

$= \dfrac{1}{n}(\underbrace{\mu + \mu + \cdots + \mu}_{n \text{ times}})$

$= \mu$

② $\mathrm{cov}(\overline{X}) = \dfrac{1}{n}\Sigma$ (How?)

For this

$(\overline{X} - \mu)(\overline{X} - \mu)'$

$= \left(\dfrac{1}{n}\sum_{j=1}^{n} x_j - \dfrac{n\mu}{n}\right)\left(\dfrac{1}{n}\sum_{l=1}^{n} x_l - \dfrac{n\mu}{n}\right)$

$= \left(\dfrac{1}{n}\sum_{j=1}^{n}(x_j - \mu)\right)\left(\dfrac{1}{n}\sum_{l=1}^{n}(x_l - \mu)'\right)$

$= \dfrac{1}{n^2}\sum_{j=1}^{n}\sum_{l=1}^{n}(x_j - \mu)(x_l - \mu)' \quad — *$

Now by defn

$\mathrm{cov}(\overline{X}) = E\left[(\overline{X} - \mu)(\overline{X} - \mu)'\right]$

using *

$$\operatorname{cov}(\bar{X}) = E\left[\frac{1}{n^2}\sum_{j=1}^{n}\sum_{l=1}^{n}(x_j-\mu)(x_l-\mu)'\right]$$

for $j \neq l$, all covariance are zero, these random variables are independent

For $j=l$

$$\operatorname{cov}(\bar{X}) = E\left[\frac{1}{n^2}\sum_{j=1}^{n}(x_j-\mu)(x_j-\mu)'\right]$$

$$= \frac{1}{n^2}\left(\sum_{j=1}^{n} \underbrace{E(x_j-\mu)(x_j-\mu)'}_{\text{by def}^n\ \Sigma}\right)$$

$$= \frac{1}{n^2}\left(\sum_{j=1}^{n} * \Sigma\right)$$

$$= \frac{n\Sigma}{n^2}$$

$$= \frac{\Sigma}{n}$$

→ sample variance covariance matrix

$$S_{ik} = \frac{1}{n-1}\sum_{j=1}^{n}(x_{ji}-\bar{x}_i)(x_{jk}-\bar{x}_k)$$

→ Sometimes it is desirable to assign a single numerical value for the variation expressed by S. one choice for a value is the determinant of S, which reduces to the usual sample variance of a single characteristic when $p=1$. This determinant is called the generalized sample variance.

$\therefore$ Generalized sample variance $= |S|$

Sample Mean, Covariance and Correlation as matrix Operations.

→ $\underset{\sim}{\bar{X}} = \frac{1}{n} \underline{X}' \mathbb{1}$ (sample mean vector)

→ $(n-1) S = \underline{X}^T \left(I - \frac{1}{n} \mathbb{1} \mathbb{1}^T \right) \underline{X}$ (sample covariance matrix)

→ $D^{1/2} = \begin{bmatrix} \sqrt{S_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{S_{22}} & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \sqrt{S_{pp}} \end{bmatrix}$ (sample standard deviation matrix)

→ $R = D^{-1/2} S D^{-1/2}$ (sample correlation matrix

⇒ $\boxed{S = D^{1/2} R D^{1/2}}$

→ sample mean of $c'x = c'\bar{X}$

$\xrightarrow{\text{observation}}$ पाउन (एका observation पाउन)

From note:

→ $\operatorname{cov}(\bar{X}) = \frac{1}{n} \Sigma_x$

Watch CLT in home for Linear combination.

To prove $A^T A = A$, then A is idempotent matrix

LHS