Akshay Kumar

End Sem Syllabus

Date:

Page No.

SDA

$$\begin{bmatrix} x_{11} & ---- & x_{1p} \\ x_{21} & -- & x_{2p} \\ x_{31} & & \\ \vdots & & \\ x_{n1} & & x_{np} \end{bmatrix}$$

n items, p random variables.

Sample mean for first variable:

$$\bar{x}_{\cancel{1}} = \frac{1}{n} \sum_{j=1}^{n} x_{j1}$$

$$\bar{x}_k = \frac{1}{n} \sum_{j=1}^{n} x_{jk} \qquad k = 1, 2, 3, \ldots p$$

Sample variance on first variable

$$s_1^2 = \frac{1}{n} \sum_{j=1}^{n} (x_{j1} - \bar{x}_1)^2$$

$$s_k^2 = \frac{1}{n} \sum_{j=1}^{n} (x_{jk} - \bar{x}_k)^2 \qquad k = 1, 2, \ldots p$$

$s_k^2 = s_{kk}$ ⇒ Sample variance. just representation

Sample standard deviation $\sqrt{s_{kk}}$

Linear association between the measurement of variables 1 and 2 is provided by <u>Sample covariance.</u>

$$s_{12} = \frac{1}{n}\sum_{j=1}^{n}(x_{j1}-\bar{x}_1)(x_{j2}-\bar{x}_2)$$

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

Covariance = Sample variance
when $i = k$

$$s_{ik} = s_{ki} \quad \forall\ i, k$$

Sample Correlation Coefficient (does not depend on the unit of measurement)

$$r_{ik} = \frac{s_{ik}}{\sqrt{s_{ii}}\sqrt{s_{kk}}} = \frac{\frac{1}{n}\sum_{j=1}^{n}(x_{ji}-\bar{x}_i)(x_{jk}-\bar{x}_k)}{\sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_{ji}-\bar{x}_i)^2}\sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_{jk}-\bar{x}_k)^2}}$$

$$r_{ik} = r_{ki} \quad \forall\ k, i \qquad k, i \in [1, p]$$

Scanned with CamScanner

Standardized $x_{ji} = \dfrac{(x_{ji} - \bar{x}_i)}{\sqrt{s_{ii}}}$

value of $r \in [-1, 1]$

r measure strength of linear association

① $r = 0$, lack of linear association b/w components

② $r < 0$, tendency of one value in pair to be larger than its avg and tendency of other to be smaller than its avg

③ $r > 0$, either both are larger or both are smaller.

④ r_{ik} remains unchanged if measurement of i^{th} variable changed to $y_{ji} = a x_{ji} + b$. $\forall j \in [1, n]$

and k^{th} variable changed to $y_{jk} = c x_{jk} + d$. $j \in [1, n]$

Provided constants a and c should have same sign

Date:
Page No.

$$w_{kk} = \sum_{j=1}^{n} (x_{jk} - \bar{x}_k)^2$$

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

$$i, k \in [1, p]$$

Arrays of descriptive statistics

Sample means:- $\bar{X} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}$

Sample variances and covariances. $S_n = \begin{bmatrix} s_{11} & s_{12} & & s_{1p} \\ \vdots & s_{22} & & s_{2p} \\ s_{p1} & \vdots & & s_{pp} \end{bmatrix}$.

n here represent, n is deployed as divisor

Sample Correlation $R = \begin{bmatrix} 1 & r_{12} & & r_{1p} \\ r_{21} & 1 & & \\ & & \ddots & \\ & & & 1 \\ r_{p1} & & & r_{pp} \end{bmatrix}$