Week 10
Lecture - 1

Principal Component Analysis: [PCA]

↳ Data Reduction Technique ⟶ Lower Dimension

↳ Developed by Hotellings

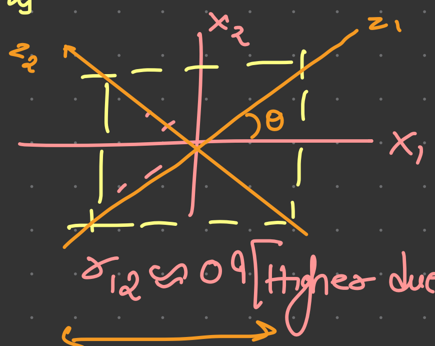
↳ Orthogonality of new dimensions

Contents:
- Intro
- Conceptual Model
- Extraction
- Sampling distn of eigen
- Model Adequacy Tests,
- Case Study

$$\begin{array}{c c} & \begin{array}{cc} x_1 & x_2 \end{array} \\ X = & \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{bmatrix} \end{array}$$

$$Cov = \begin{bmatrix} s_{11} & s_{21} \\ s_{12} & s_{22} \end{bmatrix}_{2\times 2} \xrightarrow{\text{If pop}} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

$$Cor(X) = \begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix}_{2\times 2}$$



Variability Range (x_2)

$r_{12} \approx 0.9$ [Higher due to +ve correlation]

Variability Range(x_1)

When we rotate by angle θ, then we calculate variability along z_1, z_2

Variability along z_2

Variability along z_1

$V(z_1) > V(z_2)$ ⟶ If $V(z_1) >> V(z_2)$, info along z_2 is very less, we can even ignore z_2 and capture variability by z_1 along [only 1 dimension]

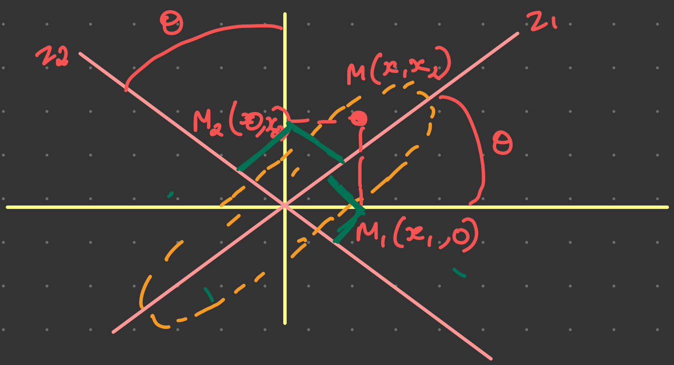
This shows z_1, z_2 are independent

$\therefore$ Orthogonality is preserved [Unlike x_1, x_2]

$$x = \begin{bmatrix} & & \\ & & \end{bmatrix}_{n\times p} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_p \end{bmatrix}$$

$$\boxed{\underset{p\times 1}{x} \xrightarrow[\text{Conversion}]{} \underset{m\times 1}{z} \quad \text{where } \underbrace{m \le p}_{\text{Main requirement}}}$$

In MLR, if IV's are correlated then estimation may have issues. So PCA can be used to make them independent



$Z_1 : X_1\cos\theta + X_2\sin\theta$

$Z_2 : -X_1\sin\theta + X_2\cos\theta$

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$Z = A^T X$$

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \qquad A^T = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

$$A = \begin{array}{cc} a_{11} & a_{21} \\ \left[\begin{array}{c|c} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{array}\right] \\ a_{12} & a_{22} \end{array} = [a_1 \quad a_2]$$

$$a_1 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \qquad a_1^T a_1 = [\cos\theta \ \sin\theta] \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$$

$$\boxed{\begin{array}{l} a_1^T a_1 = 1 \\ a_2^T a_2 = 1 \end{array}}$$

For generalised case,

$$\underset{p\times 1}{Z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_p \end{bmatrix}} = \underset{p\times p}{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & & & \\ a_{pp} & \cdots & & a_{pp} \end{bmatrix}} \underset{p\times 1}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}}$$

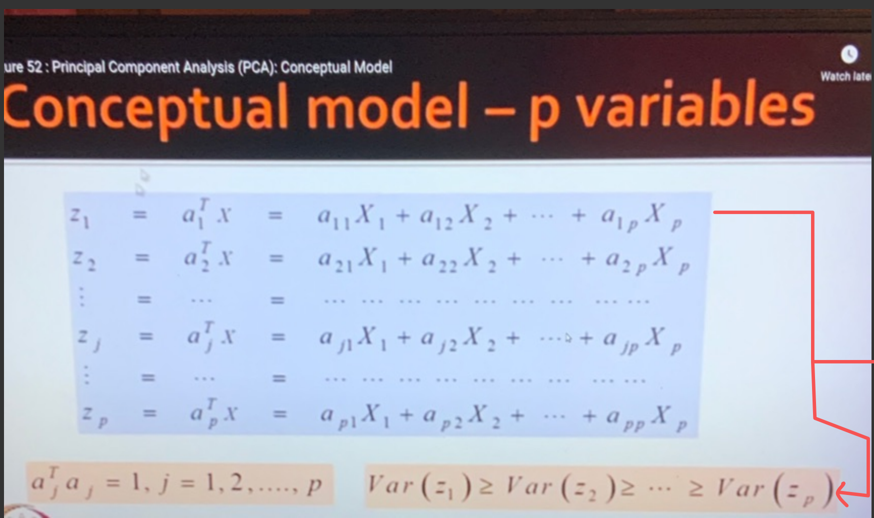
$$z = A^T x$$

$$A^T A = I = AA^T = A^{-1}A$$

Now as $A'A = I$

↓

It satisfies orthogonality criteria



ure 52 : Principal Component Analysis (PCA): Conceptual Model

Watch later

Conceptual model – p variables

$$\begin{aligned} z_1 &= a_1^T X = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\ z_2 &= a_2^T X = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \\ \vdots &= \cdots = \cdots \\ z_j &= a_j^T X = a_{j1}X_1 + a_{j2}X_2 + \cdots + a_{jp}X_p \\ \vdots &= \cdots = \cdots \\ z_p &= a_p^T X = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p \end{aligned}$$

$a_j^T a_j = 1, j = 1, 2, \ldots, p \qquad Var(z_1) \ge Var(z_2) \ge \cdots \ge Var(z_p)$

→ First Principal Component will give the maximum variability [explains most variability]

↳ Similar to case taken in ecclipse

Week-2

$z_j = a_j^T x$

$v(z_j) = v(a_j^T x) = a_j^T v(x) a_j = a_j^T \Sigma a_j$

$$\mathrm{Cov}(x) = \Sigma = \begin{bmatrix} \sigma_{11} & & \sigma_{p1} \\ & \sigma_{22} & \\ \sigma_{1p} & & \sigma_{pp} \end{bmatrix}$$

$E(z_j) = E[a_j^T x] = a_j^T E(x)$

$E(z_j) = a_j^T \mu$

$\boxed{a_j^T x \sim N(a_j^T \mu, a_j^T \Sigma a_j)}$ → Gives a univariate case

All the PCA are linear transformations

If we know $\Sigma, \mu \Rightarrow$ Population PCA

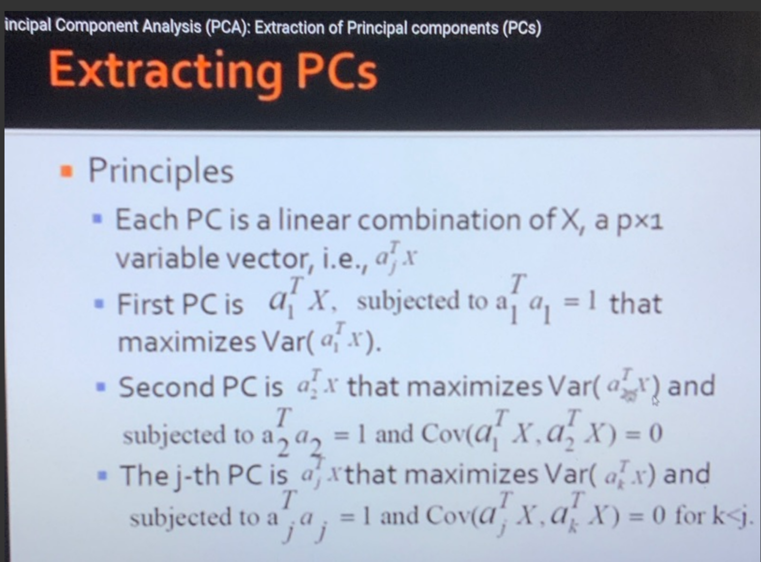
If we don't know $\Sigma, \mu \Rightarrow$ Sample PCA [use $\bar{x}, S$]

↳ Sample Covariance Matrix

Sample PCA Case:

$E(z_j) = E(a_j^T x) = a_j^T E(x) = a_j^T \bar{x}$

$V(z_j) = V(a_j^T x) = a_j^T \mathrm{Cov}(x) a_j = a_j^T S a_j$



While extracting, we have 2 conditions

① $v(z_j) = a_j^T S a_j$ ② $a_j^T a_j = 1$

$\downarrow$ Maximise $\quad$ ↳ $a_j^T a_j - 1 = 0$

So we define a fn:

$$\boxed{L = a_j^T S a_j - \lambda(a_j^T a_j - 1)}$$

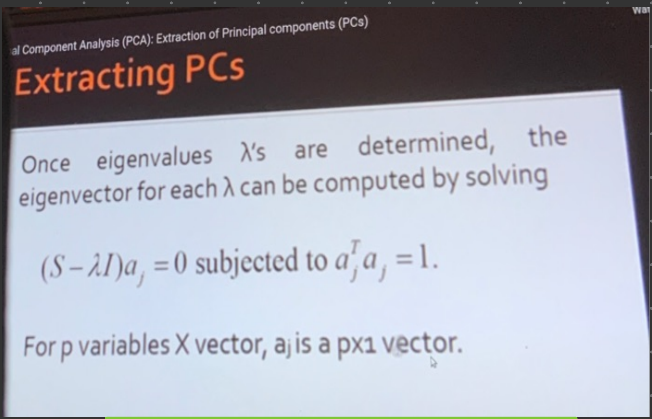
↳ Lagrange Multiplier

$$\frac{\partial L}{\partial a_j} = 0$$

$$\boxed{(S - \lambda I)a_j = 0}$$

pxp $\quad$ pxp

→ Analogous to Characteristic Eqn has p-roots when $\lambda_1 > \lambda_2 > \lambda_3 \quad \lambda_p$ for $|S - \lambda I| = 0$



Example ab20 00

Lecture-3

PCA Model Adequacy & Interpretation

Extraction on PC:

$|S-\lambda I|=0$

$\lambda_1 > \lambda_2 > \lambda_3 \quad > \lambda_p$

$(S-\lambda I)a_1=0$

$a_1^T a_1 = 1$

Example JMP

$V(a_j^T x) = a_j^T S a_j$

$V(z_1) = V(a_1^T x) = a_1^T S a_1 = \lambda_1$

$E(\lambda_j) = \theta_j$

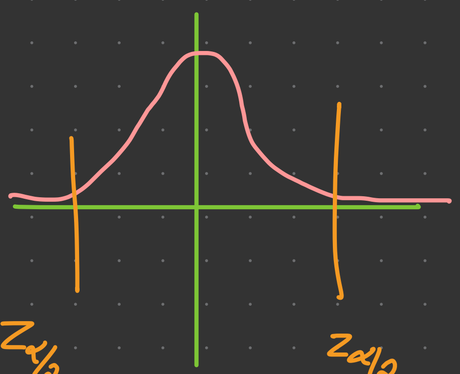
$\rightarrow$ Variance component

$$Cov(\lambda_j, \lambda_u) = \begin{cases} \dfrac{2\theta_j^2}{n-1}, & \text{for } j=u \\ 0, & \text{for } j \neq u \end{cases}$$

$\hookrightarrow$ Covariance component

$$\boxed{\lambda_j \sim N\left(\theta_j, \frac{2\theta_j^2}{n-1}\right)} \rightarrow n \text{ is large}$$

Finding CI of λ_j



$$\frac{\lambda_j - E(\lambda_j)}{SE(\lambda_j)} \sim Z(0,1)$$

$$-Z_{\alpha/2} < \frac{\lambda_j - \theta_j}{\theta_j\sqrt{2/n-1}} < Z_{\alpha/2}$$

$\Downarrow$

$$\frac{\lambda_j}{1+Z_{\alpha/2}\sqrt{\frac{2}{n-1}}} < \theta_j < \frac{\lambda_j}{1-Z_{\alpha/2}\sqrt{\frac{2}{n-1}}}$$

$z_j = a_j^T X$

$\downarrow$

Eigenvectors from sample $\longrightarrow$ $\alpha_j \rightarrow$ population j^{th} eigenvector

$$a_j = N_p(\alpha_j, T_j)$$

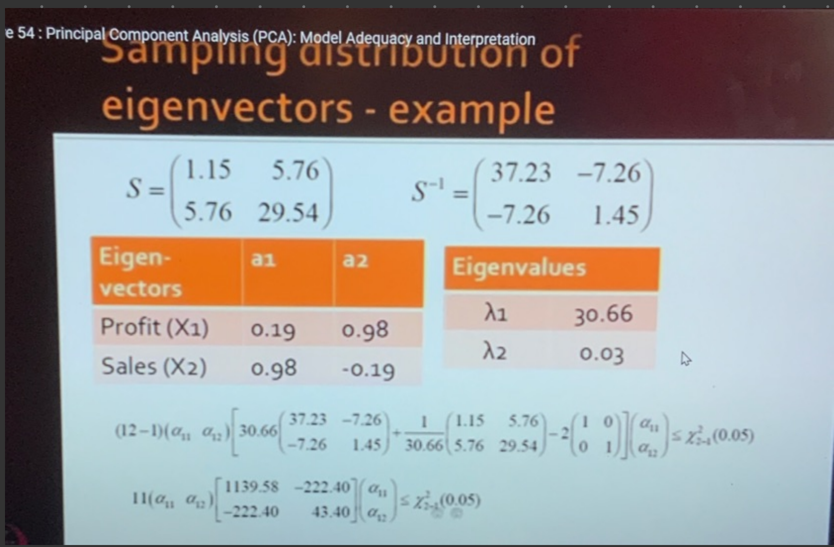
$$T_j = \frac{\theta_j}{n-1} \sum_{\substack{k=1 \\ j \neq k}}^{p} \frac{\theta_j}{(\theta_k - \theta_j)^2} \alpha_k \alpha_k^T$$

$100(1-\alpha)\%$ CR

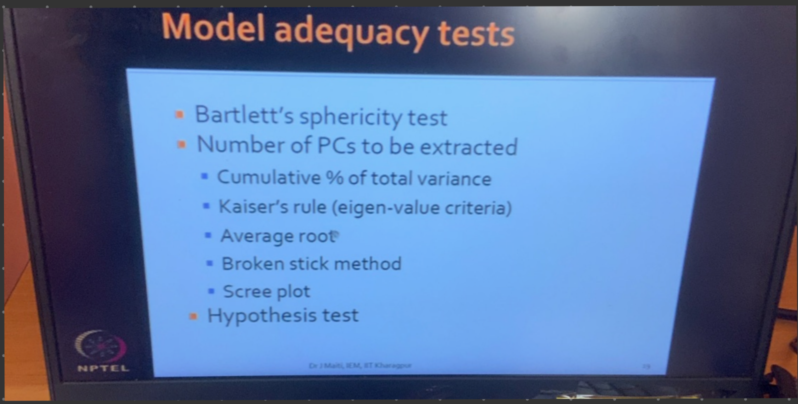
$$(n-1)\alpha_j^T(\lambda_j S^{-1} + \lambda_j^{-1} S - 2I)\alpha_j \leq \chi^2_{p-1}(\alpha)$$

$S_{p\times p}$ $\quad$ $I_{p\times p}$

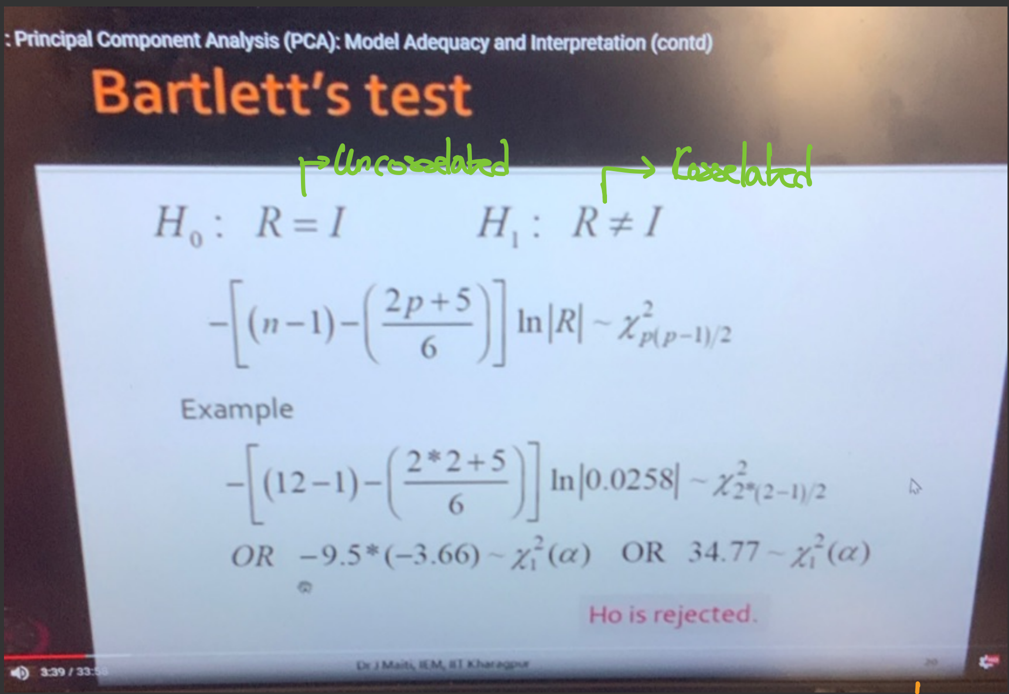
To get SCI,



Lecture-4



Bartletts Sphericity Test:



$p=2 \Rightarrow$ Circle
$p>2 \Rightarrow$ sphere

Means that they are random

↳ Even if rotated, there will be no effect as equidistant

↳ No correlation

Correlation Matrix

$$R = \begin{bmatrix} 1, & 0 \\ 0 & 1, \end{bmatrix} = I$$

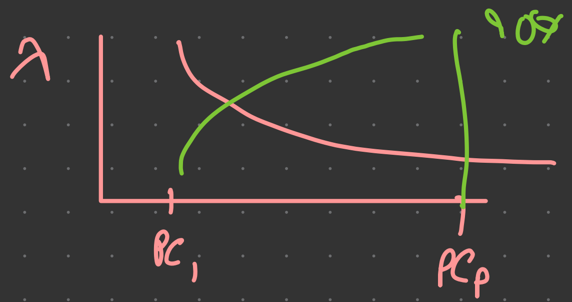
If H_0, then no need of PCA as they are uncorrelated anyways

No of PC's to be retained,

① Cumulative % Variance explained

$$\sum_{j=1}^{p} s_j^2 = \sum_{j=1}^{p} \lambda_j$$

λ_j	Value	Cumulative	% Cumulative
λ_1			$\lambda_1 / \Sigma \lambda_j$
λ_2			$\frac{\lambda_1+\lambda_2}{\Sigma \lambda_j}$
⋮			
λ_p			



Take λ_{j+1} λ_p

② Average Root

$$\bar{\lambda} = \frac{1}{p} \sum_{j=1}^{p} \lambda_j$$

Then take all $\lambda_j > \bar{\lambda}$ and reject rest

③ Kaiser's Rule:

$S \rightarrow R$ correlation matrix to extract PC's

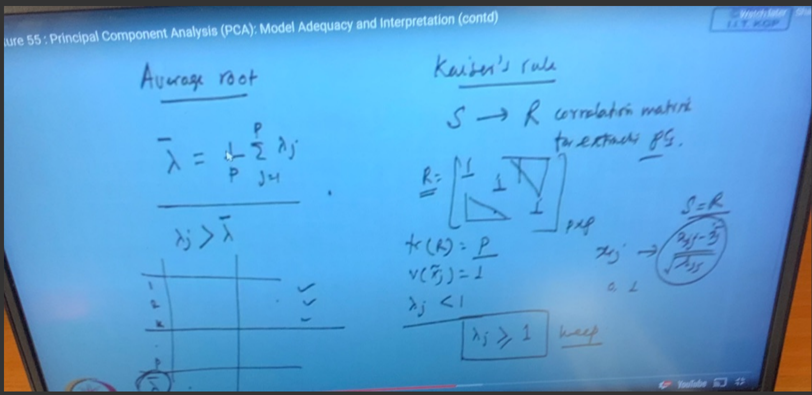
Keep values with $\lambda_j \geq 1$

$$R = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}_{p \times p}$$

trace(R) = p = No of variables

↳ $r(x_j) = 1$ → If $\lambda_j < 1$ we can ignore

We use transform $x_{ij} \rightarrow \frac{x_{ij} - \bar{x}_j}{\sqrt{S_{jj}}}$ $\Rightarrow S = R$



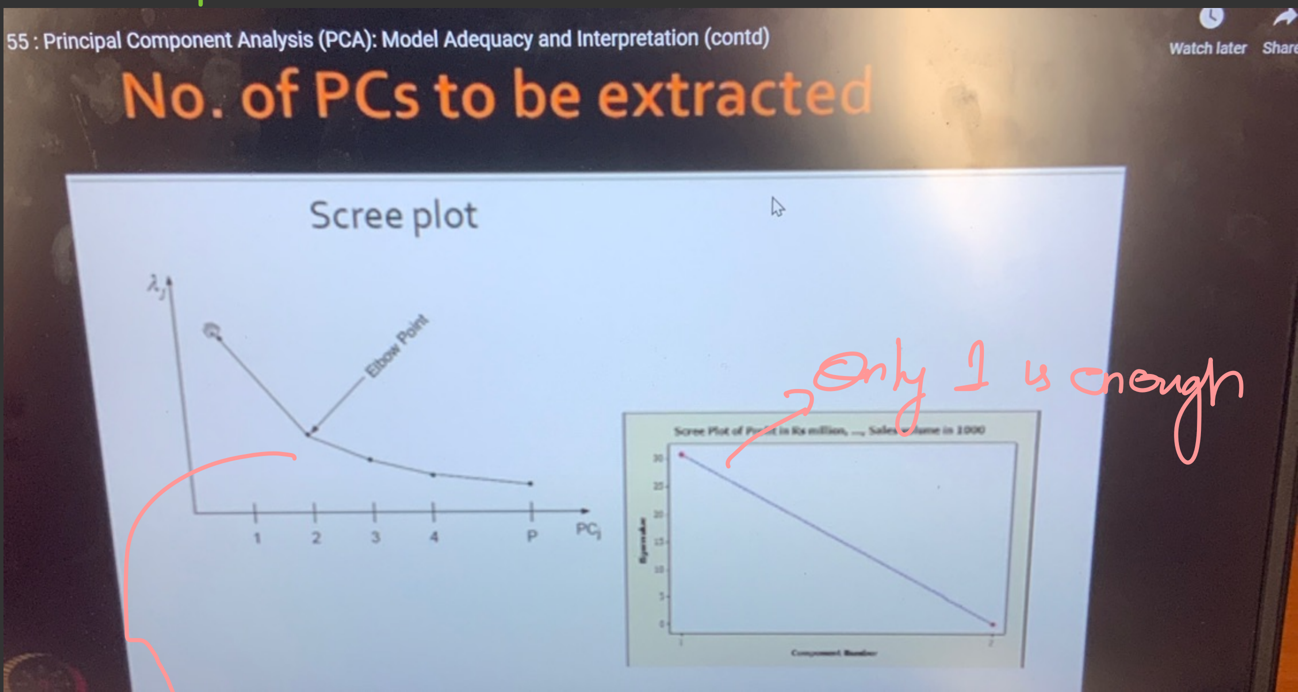
④ Broken Stick Method:

$$l_j = \frac{1}{p}\sum_{k=j}^{p}\frac{1}{k}$$

↳ Broken into several components

λ_j : Keep $\dfrac{\lambda_j}{\sum_{j=1}^{p}\lambda_j} > l_j$

⑤ Screeplot:



Only 1 is enough

Elbow should make roughly 180° [take everything before]

Hypothesis test

$H_0:\ \lambda_{m+1} = \lambda_{m+2} = \cdots = \lambda_p$

Bartlett 1950

$H_1:\ \lambda_{m+j} \neq \lambda_{m+k}\quad j \neq k,$ for at least one pair of λ from last p-m λ's.

$$D = n\left[(p-m)\ln\left(\bar{\lambda}_m\right) - \sum_{j=m+1}^{p} \ln\left(\lambda_j\right)\right] \sim \chi^2_{\frac{1}{2}(p-m-1)(p-m+2)}$$

$$\bar{\lambda}_m = \sum_{j=m+1}^{p} \frac{\lambda_j}{p-m}$$

Reject Ho when D≥ $\chi^2_{\frac{1}{2}(p-m-1)(p-m+2),\alpha}$

ARAVIND

NARAYANAN