



# Week 10

## Lecture - 1

### Principal Component Analysis: [PCA]

#### Contents:

Intro  
Conceptual Model  
Extraction  
Sampling distribution of eigen  
Model Adequacy Tests  
Case Study

→ Data Reduction

Technique

→ Developed by Hotellings

→ Lower Dimension

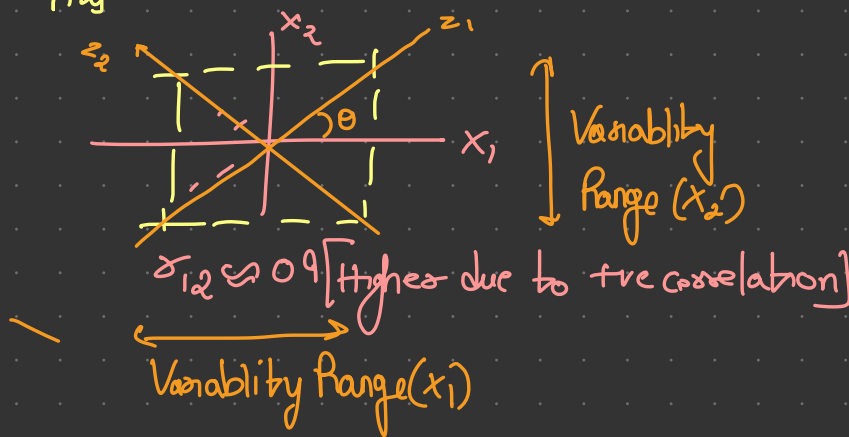
→ Orthogonality of new dimensions

IF pop  $\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$

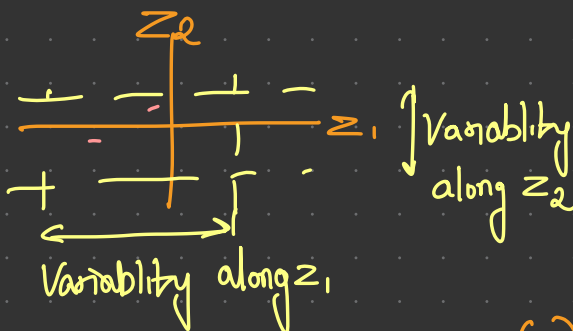
$$X = \begin{bmatrix} x_1 & x_2 \\ x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{bmatrix}$$

$$\text{Cov} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}_{2 \times 2}$$

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



When we rotate by angle  $\theta$ , then we calculate variability along  $z_1, z_2$



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

This shows  $z_1, z_2$  are independent

∴ Orthogonality is preserved [unlike  $x_1, x_2$ ]

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

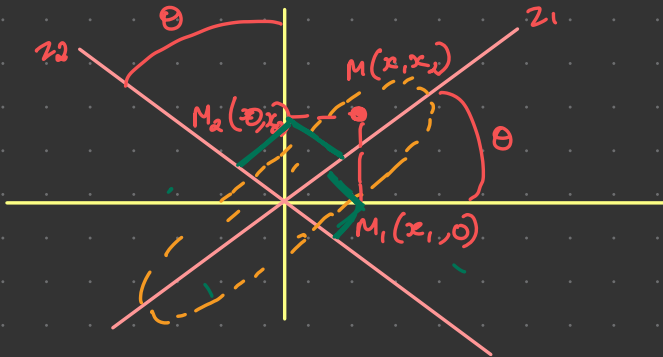
$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_p \end{bmatrix}$$

$$X \rightarrow Z \quad \text{where } m \leq p$$

Conversion

Main requirement

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



$$Z_1 = x_1 \cos \theta + x_2 \sin \theta$$

$$Z_2 = -x_1 \sin \theta + x_2 \cos \theta$$

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

$$Z = A^T X$$

$$Z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \quad A^T = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$$

For generalised case,

$$Z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_p \end{bmatrix}_{p \times 1} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pp} \end{bmatrix}_{p \times p} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}_{p \times 1}$$

$$Z = A^T X$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pp} \end{bmatrix} = [a_1 \ a_2 \ \dots \ a_p]$$

$$a_1 = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} \quad a_1^T a_1 = \begin{bmatrix} \cos \theta & \sin \theta \end{bmatrix} \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} = 1$$

$$a_2^T a_2 = 1$$

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

Now as  $A^T A = I$

It satisfies orthogonality criteria

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

### Conceptual model – p variables

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

$a_j^T a_j = 1, j = 1, 2, \dots, p$        $Var(z_1) \geq Var(z_2) \geq \dots \geq Var(z_p)$

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

↳ Similar to case taken in eclipse

$$z_j = a_j^T x$$

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

$$\downarrow$$

$$\text{Cov}(x) = \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix}$$

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

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

$$a_j^T x \sim N(a_j^T \mu, a_j^T \Sigma a_j)$$

→ 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:

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

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

Principal Component Analysis (PCA): Extraction of Principal components (PCs)

## Extracting PCs

### Principles

- Each PC is a linear combination of  $X$ , a  $p \times 1$  variable vector, i.e.,  $a_j^T x$
- First PC is  $a_1^T X$ , subjected to  $a_1^T a_1 = 1$  that maximizes  $\text{Var}(a_1^T x)$ .
- Second PC is  $a_2^T x$  that maximizes  $\text{Var}(a_2^T x)$  and subjected to  $a_2^T a_2 = 1$  and  $\text{Cov}(a_1^T X, a_2^T X) = 0$
- The  $j$ -th PC is  $a_j^T x$  that maximizes  $\text{Var}(a_j^T x)$  and subjected to  $a_j^T a_j = 1$  and  $\text{Cov}(a_j^T X, a_k^T X) = 0$  for  $k < j$ .

While extracting, we have 2 conditns

$$\textcircled{1} v(z_j) = a_j^T S a_j \quad \textcircled{2} a_j^T a_j = 1$$

↓  
Maximize

↳  $a_j^T a_j - 1 = 0$

So we define a fn:

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

↳ Lagrange Multiplier

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

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

pxp                  pxp

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

Principal Component Analysis (PCA): Extraction of Principal components (PCs)

### Extracting PCs

Once eigenvalues  $\lambda$ 's are determined, the eigenvector for each  $\lambda$  can be computed by solving

$$(S - \lambda I) a_j = 0 \text{ subjected to } a_j^T a_j = 1.$$

For p variables X vector,  $a_j$  is a px1 vector.

Example at 20 00

# Lecture-3

## PCA Model Adequacy & Interpretation

Extraction on PC:

$$|S - \lambda I| = 0$$

$$\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_p$$

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

$$a_i^T a_i = 1$$

Example JMP

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

$$V(z_j) = V(a_j^T x) = a_j^T S a_j = \lambda_j$$

$$E(\lambda_j) = \theta_j$$

→ Variance component

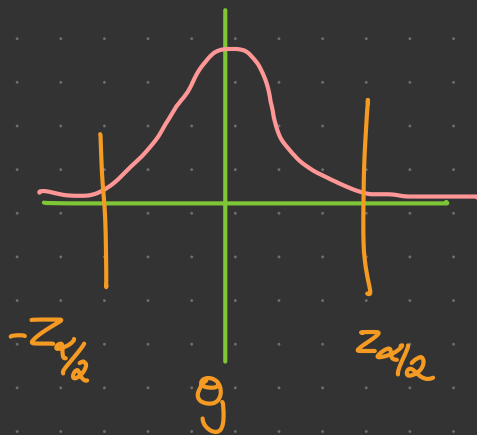
$$\text{Cov}(\lambda_j, \lambda_u) = \begin{cases} \frac{2\theta_j^2}{n-1}, & \text{for } j=u \\ 0, & \text{for } j \neq u \end{cases}$$

→ Covariance component

$$\lambda_j \sim N\left(\theta_j, \frac{2\theta_j^2}{n-1}\right)$$

→ n is large

Finding CI of  $\lambda_j$



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

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

⇓

$$z_j = a_j^T x$$

Eigenvectors from sample

$\alpha_j \rightarrow$  population  $j^{\text{th}}$  eigenvector

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

$$a_j = Np(a_j, T_j) \cdot$$

$$T_j = \frac{\Theta_j}{n-1} \sum_{\substack{k=1 \\ j \neq k}}^p \frac{\Theta_j}{(\Theta_k - \Theta_j)^2} \tau_k \tau_k^T$$

→ 100(1-α)% CR

$$(n-1) \tau_j^T (\underbrace{\lambda_j S^{-1}}_{S_{p \times p}} + \underbrace{\lambda_j^{-1} S}_{T_{p \times p}} - \alpha I) \tau_j \leq \chi^2_{p-1}(\alpha)$$

To get SCI,

e 54 : Principal Component Analysis (PCA): Model Adequacy and Interpretation

## Sampling distribution of eigenvectors - example

$$S = \begin{pmatrix} 1.15 & 5.76 \\ 5.76 & 29.54 \end{pmatrix}$$

$$S^{-1} = \begin{pmatrix} 37.23 & -7.26 \\ -7.26 & 1.45 \end{pmatrix}$$

Eigen-vectors	a1	a2	Eigenvalues	
Profit (X1)	0.19	0.98	λ1	30.66
Sales (X2)	0.98	-0.19	λ2	0.03

$$(12-1)(a_{11} \ a_{12}) \left[ 30.66 \begin{pmatrix} 37.23 & -7.26 \\ -7.26 & 1.45 \end{pmatrix} + \frac{1}{30.66} \begin{pmatrix} 1.15 & 5.76 \\ 5.76 & 29.54 \end{pmatrix} - 2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] \begin{pmatrix} a_{11} \\ a_{12} \end{pmatrix} \leq \chi^2_{2-1}(0.05)$$

$$11(a_{11} \ a_{12}) \begin{bmatrix} 1139.58 & -222.40 \\ -222.40 & 43.40 \end{bmatrix} \begin{pmatrix} a_{11} \\ a_{12} \end{pmatrix} \leq \chi^2_{2-1}(0.05)$$



# Lecture-4

## Model adequacy tests

- Bartlett's sphericity test
- Number of PCs to be extracted
  - Cumulative % of total variance
  - Kaiser's rule (eigen-value criteria)
  - Average root
  - Broken stick method
  - Scree plot
- Hypothesis test

## Bartlett's Sphericity Test

Principal Component Analysis (PCA): Model Adequacy and Interpretation (contd)

### Bartlett's test

$$H_0: R = I \quad H_1: R \neq I$$

$$-\left[(n-1) - \left(\frac{2p+5}{6}\right)\right] \ln|R| \sim \chi^2_{p(p-1)/2}$$

Example

$$-\left[(12-1) - \left(\frac{2*2+5}{6}\right)\right] \ln|0.0258| \sim \chi^2_{2*(2-1)/2}$$

$$\text{OR } -9.5 * (-3.66) \sim \chi^2_1(\alpha) \quad \text{OR } 34.77 \sim \chi^2_1(\alpha)$$

Ho is rejected.

$p=2 \Rightarrow$  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



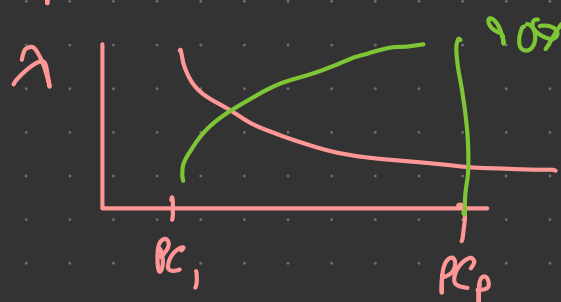
No of PC's to be retained,

① Cumulative % Variance explained

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

$\lambda_j$	Value	Cumulative	% Cumulative
$\lambda_1$			$\lambda_1 / \sum \lambda_j$
$\lambda_2$			$\frac{\lambda_1 + \lambda_2}{\sum \lambda_j}$
$\vdots$			
$\lambda_p$			

Take  $\lambda_{j+1} \approx \lambda_j$



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

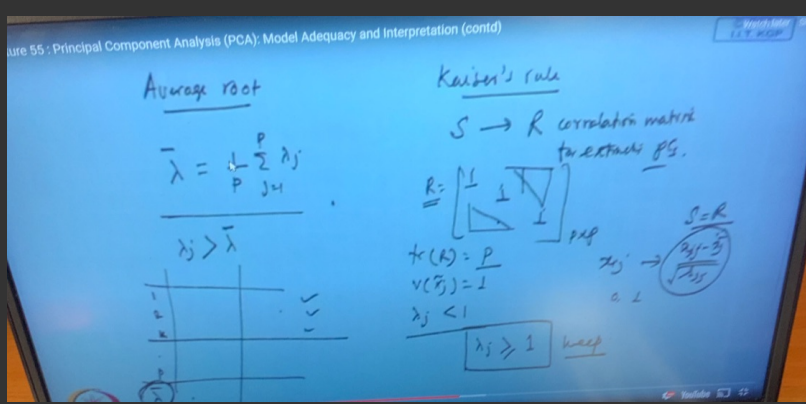
Keep values  
with  
 $\lambda_j \geq 1$

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

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

$\text{trace}(R) = p = \text{No of variables}$   
transform

We use  $x_j \rightarrow \frac{x_j - \bar{x}_j}{\sqrt{s_{jj}}} \Rightarrow S = R$



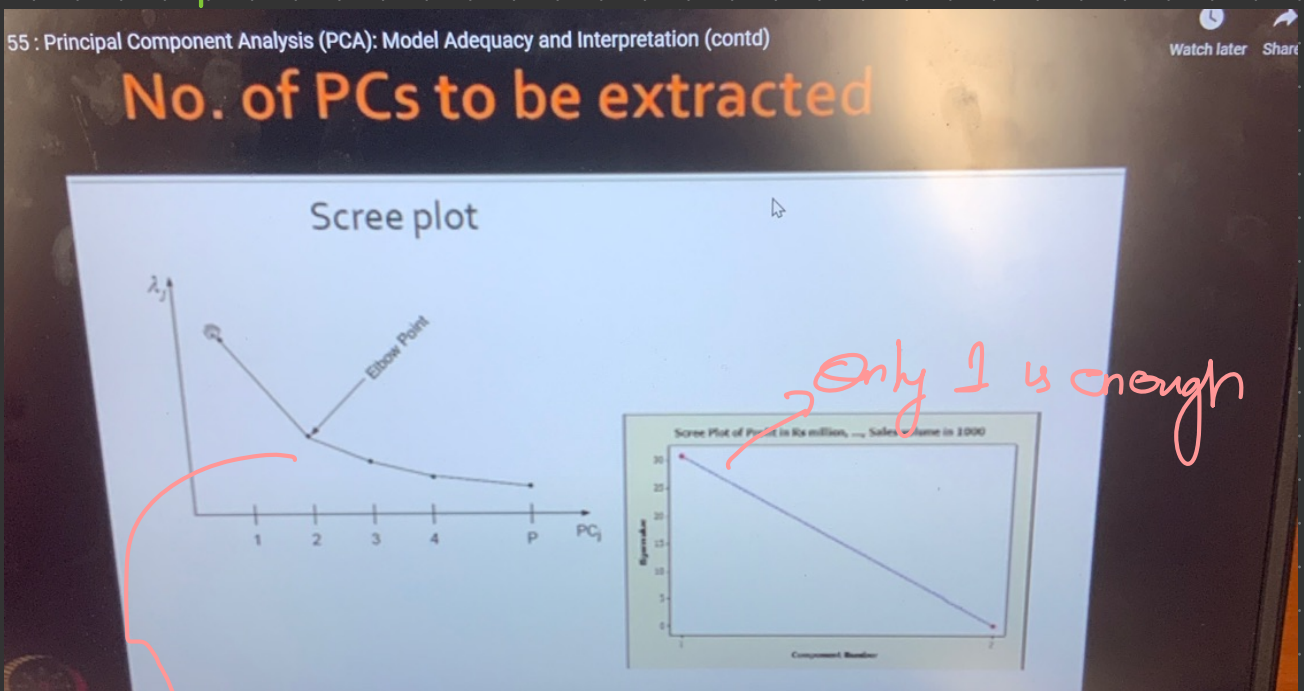
#### ④ Broken Stick Method:

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

$\rightarrow$  Broken into several components

$\rightarrow$  Keep  $\frac{\lambda_j}{\sum_{j=1}^p \lambda_j} > l_j$

#### ⑤ Scree plot:



Elbow should make roughly 1<sup>st</sup> [take everything before]

# Hypothesis test

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

Bartlett 1950

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

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

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

Reject  $H_0$  when  $D \geq \chi^2_{\frac{1}{2}(p-m-1)(p-m+2), \alpha}$

ARAVIND

NARAYANAN