9/9

1. A vector (linear) space is a set V such that:

a) $\forall\, u, v \in V \Rightarrow u+v \in V$

b) if c is a scalar, $\forall\, u \in V \Rightarrow cu \in V$

c) $u+(v+w) = (u+v)+w \qquad \forall\, u, v, w \in V$

$u+0 = 0+u = u$

$u+v = v+u$

d) a, b scalars $\Rightarrow a(bu) = (ab)u$

$(a+b)u = au + bu$

e) $1 \cdot u = u$

2. $A+B = B+A$

$(A+B)+C = A+(B+C) \qquad A, B, C$ are matrices

$\alpha(A+B) = \alpha A + \alpha B \qquad \alpha, \beta$ scalars

$(\alpha+\beta)A = \alpha A + \beta A$

$(\alpha\beta)A = \alpha(\beta A)$

$\alpha(AB) = (\alpha A)B$

$A(BC) = (AB)C$

$A(B+C) = AB + AC$

$(A+B)C = AC + BC$

$(A+B)^T = A^T + B^T$

$(\alpha A)^T = \alpha A^T$

$(AB)^T = B^T \cdot A^T$

$AB \neq BA$ (in general)

3. $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ $A_{m\times n} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$

$AX = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ $X^TA = (z_1, \cdots, z_n)$

$X^TX = \sum_{i=1}^{n} x_i^2$

$(XX^T)_{n\times n} =$ $r(XX^T) = 1$

4. Inner product

Let V be a vector (linear) space

$\forall u, v \in V$. assign a scalar such that

$\langle u, v\rangle$:

a) $\forall w \in V$, $\forall a, b \in \mathbb{C}$ ← complex.

$\langle au + bw, v\rangle = a\langle u, v\rangle + b\langle w, v\rangle$

$\langle u, av + bw\rangle = \bar{a}\langle u, v\rangle + \bar{b}\langle u, w\rangle$ $\left(\overline{x+iy} = x - iy \quad z = re^{i\varphi} \quad \bar{z} = re^{-i\varphi} \right)$

$= r(\cos\varphi + i\sin\varphi)$

bilinearity and antilinearity in the 2nd slot.

也可以定义为： $\langle au + bw, v\rangle = \bar{a}\langle u, v\rangle + \bar{b}\langle w, v\rangle$

$\langle u, aw + bv\rangle = a\langle u, w\rangle + b\langle u, v\rangle$

b) $\langle u, v\rangle = \overline{\langle v, u\rangle}$ under $\mathbb{C}$.

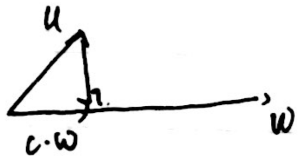
quisi reflexifility

c) $\langle u, u\rangle \geq 0$

$\langle u, u\rangle = 0 \iff u = 0$

If V is equipped with an inner product, V is an inner product space.

$u, v \in V,\ u, v \neq 0 \Rightarrow \langle u, v\rangle = 0 \Rightarrow$ u, v are orthorgonal.



$$\text{projection}_w\, v = \frac{\langle v, w\rangle}{\langle w, w\rangle} w$$

c

三維：$\text{project}_w\, v = \frac{\langle v, w\rangle}{\langle w, w\rangle} w + \frac{\langle v, u\rangle}{\langle u, u\rangle} u$

5. Distance

a) $d(x, y) = d(y, x)$

b) $d(x, y) \geq 0$

$d(x, y) = 0 \Leftrightarrow x = y$

c) $d(x, y) \leq d(x, z) + d(z, y)$ triangular inequality

$d(x, y) = |\langle x-y, x-y\rangle|^{\frac{1}{2}}$

$\|x\| = \sqrt{\langle x, x\rangle}$

6. Norm

1) $\|av\| = |a|\|v\|$

2) $\|u+v\| \leq \|u\| + \|v\|$ triangular inequality

3) $\|u\| = 0 \Leftrightarrow u = 0$

from 2) and 3) $\|u + (-u)\| = \|0\| = 0 \leq 2\|u\| \Rightarrow \|u\| \geq 0 \leftarrow$ result, not def.

$\|u\| = \sqrt{\langle u, u\rangle}$.

$u, v \in V,\ u, v \neq 0 \quad \langle u, v\rangle = 0 \quad \|u\| = \|v\| = 1$

Orthonormal.

A set of vectors $x_1, \cdots x_k \in V$ is said to be linearly dependent if $\exists\, c_1, \cdots c_k$ not all 0. s.t. $c_1 x_1 + \cdots + c_k x_k = 0$

$\dim(V) = \#$ of basis vectors

$A^{-1}_{n\times n}$ exists. $\Rightarrow$ The columns/rows are linearly indep.

6/9 1. Determinants

$$A_{n\times n} = \begin{cases} a_{11} & n=1 \\ \sum_{j=1}^{n} a_{1j} |A_{1j}| (-1)^{1+j} & n>1. \end{cases}$$

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

$\det(A) = |A|$

a) $n=2$

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11} \cdot a_{22} (-1)^{1+1} + a_{12}\, a_{21} \cdot (-1)^{1+2} = a_{11} \cdot a_{22} - a_{12} \cdot a_{21}$$

b) $n=3$

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{21}a_{32}a_{13} - a_{13}a_{22}a_{31} - a_{12}a_{33}a_{21} - a_{11}a_{23}a_{32}$$

$$\begin{pmatrix} 6 & 1 & 1 \\ 4 & -2 & 5 \\ 2 & 8 & 7 \end{pmatrix} = -84 + 10 + 32 + 4 - 240 - 28 = -320 + 14 = -306.$$

2. Properties:

1) c is a scalar

$|c \cdot A_{n\times n}| = c^n |A_{n\times n}|$

2) $|A^T| = |A|$

3) $|AB| = |A| \cdot |B|$ matrix do not commute

4) $D = \begin{pmatrix} d_{11} & & & 0 \\ & d_{22} & & \\ & & \ddots & \\ 0 & & & d_{nn} \end{pmatrix}$

$|D| = \prod_{i=1}^{n} d_{ii}$

3. $|A| = 0$ iff. columns (rows) of A are lin. dep.

4. $I_{n\times n} = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix}$

$IA = AI = A$

5. $A^{-1}_{n\times n} \Rightarrow A \cdot A^{-1} = I = A^{-1}A$

$\exists A^{-1}$ iff $|A| \neq 0$ iff all col (rows) of A lin. indep.

Properties

(1) $(A^{-1})^T = (A^T)^{-1}$

(2) $(AB)^{-1} = B^{-1}A^{-1}$

(3) $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$

6. Eigenvalues and eigenvectors

$A_{n\times n}$, λ is an eigenvalue of A if $\exists x$. $|x| \neq 0$. $Ax = \lambda x$

(λ, x) eigenpair.

Every square matrix has eigenvalues

$|A - \lambda I| = 0$

polynomial $(\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n) = 0$

eigenvalues are unique, eigenvectors are not.

$|A| = 0$ iff A has 0 eigenvalues.

7. $A = A^T$, λ_i and λ_j are eigenvalues of A, $\lambda_i \neq \lambda_j$
$\Rightarrow x_i \perp x_j$ eigenvectors orthorgonal.

8. $A_{n\times n}$, A is diagonalizable i.e. $A = P \Lambda P^{-1}$ iff A has n linearly indep. eigenvectors. 可对角化.
$P = [e_1, e_2, \cdots e_n]$
$\Lambda = diag(\lambda_1, \lambda_2, \cdots \lambda_n)$.

9. P is called orthorgonal if $P^T = P^{-1}$; $P^T P = P P^T = I$
$A^T_{n\times n} = A \Rightarrow A = \underset{(e_1, \cdots, e_n)}{E} \underset{diag\{\lambda_1, \cdots, \lambda_n\}}{\Lambda} E^T$

10. $\overset{\text{symetric}}{A_{n\times n}}$. $A^{\frac{1}{2}} = E \Lambda^{\frac{1}{2}} E^T$ $\quad A^{\frac{1}{2}}$ not unique
$\Lambda^{\frac{1}{2}} = \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix}$

11. Quadratic forms
$A_{n\times n} = A^T$; $X \in R^n$. $X^T A X$ is a scalar
$A \geq 0$. $\forall X$, $|x| \neq 0$, $X^T A X \geq 0 \Leftrightarrow \min \lambda_i \geq 0$

12. $Tr(A) = Sp(A) = \sum_{i=1}^{n} a_{ii}$
a) $Tr(A) = Tr(A^T)$
b) $Tr(\lambda A) = \lambda Tr(A)$
c) $Tr(A+B) = Tr(A) + Tr(B)$
d) $Tr(AB) = Tr(BA)$ commute

e) $\mathrm{Tr}(A) = \sum_{i=1}^{n} \lambda_i$

1. SVD

$A_{m\times n} = \underset{m\times m,\ \text{orthorgonal}}{U}\Sigma V^T_{n\times n}$, orthorgonal

$\Sigma_{m\times n} = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$, D is diagonal

$(A_{m\times n}\cdot A^T_{n\times m})_{m\times m}$ symmetric $\sim PLP^T$ $= U\Sigma V^T(U\Sigma V^T)^T = U\Sigma \underbrace{V^T V}_{I}\Sigma U^T$

$= U\begin{pmatrix} D^2 & 0 \\ 0 & 0 \end{pmatrix}_{m\times m} U^T = U\begin{pmatrix} d_1^2 \ddots d_n^2 & 0 \\ 0 & 0 \end{pmatrix} U^T$ d_is are eigenvalues of $A\cdot A^T$

$(A^T\cdot A)_{n\times n} = (U\Sigma V^T)^T U\Sigma V^T = V_{n\times n}\Sigma^T\cdot\Sigma V_{n\times n}$

$d_1, \cdots d_k$ are singular values of A.

$d_1 \geq d_2 \geq \cdots \geq d_k$

2. A random vector Y

$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$, x_i is a random variable, $i = 1, \cdots, n$

3. x outcome of X $= \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

The joint CDF of a continous r.v. X is $F_X(x) = F_{x_1\cdots x_n}(x_1, \cdots x_n)$

$= P(X \leq x) = P\{x_1 \leq x_1, \cdots, x_n \leq x_n)$

4. $\mu_X = EX = \begin{pmatrix} EX_1 \\ \vdots \\ EX_n \end{pmatrix}$ joint pdf

$f_X(x) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{x_1 \cdots x_n}(x_1, \cdots x_n)$

5. $E[(X-EX)(X-EX)^T] = C_x$

$C_x(i,j) = Var(x_i, x_j) = E[(x_i-\mu_i)(x_j-\mu_j)]$

6. $C_x \geq 0 \iff \forall a \in R^n, \|a\| \neq 0. \quad a^T C_x a \geq 0$

Proof: C_x be a cov. matrix.

$a^T C a = a^T E[(x-\mu)(x-\mu)^T] a = E[\overbrace{a^T(x-\mu)}^{w^T}\overbrace{(x-\mu)^T a}^{w}] = E(w^T w) = E(\sum_{i=1}^n w_i^2) > 0$

n observations

p variables

$p >> n$

1. Moore - Penrose
2. Dimension Reduction (e.g. SVD)
3. Thresholding

estimate of Σ: $S = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & & & \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{pmatrix}$

take $\varepsilon > 0$, $\hat{S}_T = \begin{cases} s_{ij} & \text{if } |s_{ij}| > \varepsilon \\ 0 & \text{otherwise} \end{cases}$

4. $\hat{S}_L = \begin{cases} s_{ij} & \text{if } |i-j| < k \\ 0 & \text{otherwise} \end{cases}$

$\hat{S}_L = \begin{pmatrix} s_{11} & s_{12} & s_{13} & 0 & \cdots \\ s_{21} & s_{22} & s_{23} & s_{24} & 0 \cdots \\ 0 & & & & \end{pmatrix}$

5. $S + \begin{pmatrix} \alpha_1 & & & \\ & \alpha_2 & & \\ & & \ddots & \\ & & & \alpha_p \end{pmatrix}$ ridge operator

$\varphi_z(t) = Ee^{tz} = e^{t\mu + \frac{1}{2}t^2\sigma^2}$ $\quad Z \overset{d}{=} N(\mu, \sigma^2)$ univariate

$\forall a$, take $t=a$, $\psi_X(t) = e^{t'\mu + \frac{1}{2}t'\Sigma t}$

Kac's theorem:

$X=(x_1,\cdots,x_n)'$. Components of X. i.e. $x_1 \cdots x_n$ are independent iff

$\phi_X(s) = E[e^{is'X}] = \prod_{i=1}^{n} \phi_{x_i}(s_i)$

$\begin{pmatrix}0\\0\\ \vdots\\k\\0\\0\end{pmatrix}$ ← kth ort

Slutsky theorem

Normality test.

Shapiro-Wilk $\quad W = \frac{\sum a_i x_{[i]}}{\sum x_i} \approx 1.$ ← order

Omnibus: test whether the explained variance in a set of data is significantly greater than unexplained variance.

Jarque-Bera test

$r_n \frac{(b_1-3)^2}{c_1} + r_n \frac{b_2^2}{c_2} \sim \chi^2_{(2)}$

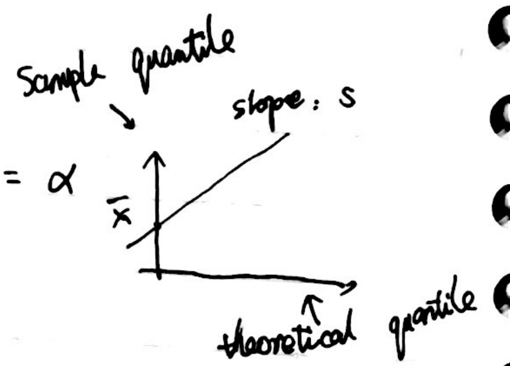
kurtosis, skewness; $\chi^2_{(1)}$, $\chi^2_{(1)}$

Q-Q (continous distribution)

$X \sim N(\mu, \sigma^2)$

$P\{X \le X_\alpha\} = P\{\frac{X-\mu}{\sigma} \le \frac{X_\alpha-\mu}{\sigma}\} = P(Z \le Z_\alpha) = \alpha$

$X_\alpha = S \cdot Z_\alpha + \bar{X}$



The multivariate test controls type I error rate.

$\forall a$, take $t = a$, $\varphi_x(t) = e^{t'\mu + \frac{1}{2}t'\Sigma t}$

Kac's theorem:

$X = (x_1, \cdots, x_n)'$. Components of X. i.e. $x_1 \cdots x_n$ are independent iff
$\phi_X(s) = E[e^{is'x}] = \prod_{i=1}^{n} \phi_{x_i}(s_i)$

$\begin{pmatrix} 0 \\ 0 \\ k \\ 0 \\ 0 \end{pmatrix} \leftarrow$ kth ont

Slutsky theorem

Normality test.

Shapiro - Wilk $\quad W = \frac{\sum a_i X_{[i]}}{\sum X_i} \approx 1$. order

Omnibus : test whether the explained variance in a set of data is significantly greater than unexplained variance.

Jarque - Bera test

$r_n \frac{(b_1 - 3)^2}{C_1} + r_n \frac{b_2^2}{C_2} \sim \chi^2_{(2)}$

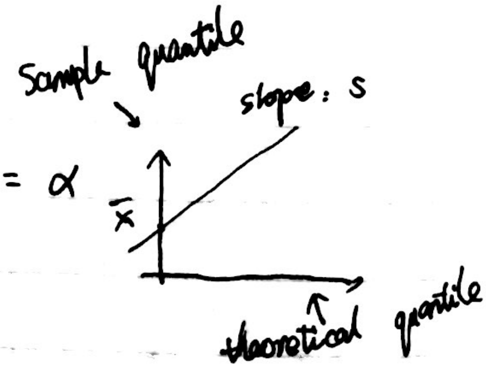
kurtosis $\quad \chi^2_{(1)} \quad$ skewner $\quad \chi^2_{(1)}$

Q-Q (continuous distribution)

$X \sim N(\mu, \sigma^2)$

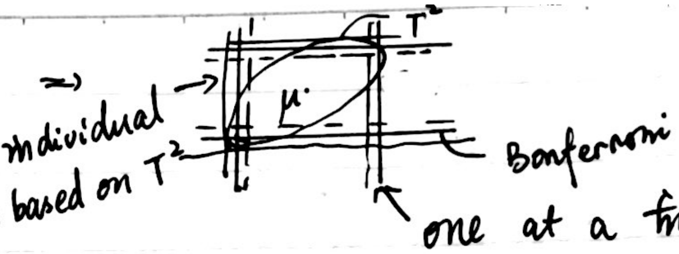
$P\{X \le X_\alpha\} = P\{\frac{X-\mu}{\sigma} \le \frac{X_\alpha - \mu}{\sigma}\} = P(Z \le Z_\alpha) = \alpha$

$X_\alpha = S \cdot Z_\alpha + \bar{X}$



The multivariate test controls type I error rate

$n(\bar{x}-\mu)' S^{-1} (\bar{x}-\mu) \Rightarrow$



One-at-a-time

1. Narrower (Shaper)
2. More powerful : "reject H_0 when H_0 is true." more likely.
3. liberal.
4. Coverage rate $< 100(1-\alpha)\%$
5. Coverage rate depends on p and S.

T^2 intervals

1. wider
2. less powerful
3. conservative
4. C. R $\geq 100(1-\alpha)\%$

Bonferroni method

C_i is a confidence statement about $a_i^T \mu$ and $P(C_i \text{ is true}) = 1-\alpha$

$i = 1, 2, \cdots m$.

$$P(\text{all } C_i \text{ true}) = 1 - P(\text{at least one } C_i \text{ is false})$$
$$\geq 1 - \sum_{i=1}^{m} P(C_i \text{ false}) = 1 - \sum_{i=1}^{m} (1 - P(C_i \text{ true})) = 1 - \sum_{i=1}^{m} \alpha_i$$

$$Z = a^T X \sim N(a^T \mu, a^T \Sigma a)$$
$$\bar{Z} \pm t_{n-1}\left(\frac{\alpha}{2m}\right) \sqrt{\frac{a^T S a}{n}}$$

If you have P intervals, then the correlation factor $\frac{\alpha}{p}$

$$\frac{\text{length of B CI}}{\text{length of } T^2} = \frac{t_{n-1}(\alpha/2p)}{\sqrt{\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)}}$$

Principal Component Analysis (PCA)
Singular Value Decomposition (SVD)

1) $B_{p\times p}$: $\lambda_1, \cdots, \lambda_p$ are eigenvalues of B.
$Be_i = \lambda e_i$, $|e_i| \neq 0$, e_i is an eigenvector corresponding to λ_i.
iff $\lambda_i, \cdots \lambda_p$ is a solution for $|B_{p\times p} - \lambda I_{p\times p}| = 0$
$\|e_i\| = 1$

1) if $B = B^T$, then all eigenvalues are real.
2) if $B = B^T$, $e_i \perp e_j$ if $\lambda_i \neq \lambda_j$.
3) $|B| = \prod^{p} \lambda_i$
4) $tr(B) = \sum_{i=1}^{p} \lambda_i$
5) if $\Sigma_{p\times p} = diag\{\alpha_1, \cdots \alpha_p\}$ eigenvalues
6) $\# \lambda_i$, $\lambda_i \neq 0$, is a rank.
7) $B = B^T$, $B > 0$ iff $\lambda_i > 0$ $\forall i = 1 \cdots p$.
8) Non-zero eigenvalues of B and B^T are identical, but eigenvectors are not. in general.
9) if B^{-1} exists, then $\lambda_1^{-1} \cdots \lambda_p^{-1}$ are e.v of B^{-1}

$X_{n\times p}$, $Y = XA$
$A_{p\times p}$ orthorgonal matrix. $A \cdot A^T = A^T \cdot A = I$. My goal is to find such A that $Y^T Y$ is diagonal.

$$Y^T Y = (XA)^T XA = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{pmatrix} \rightarrow L \qquad A_{p\times p} = \begin{pmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{p1} & \cdots & a_{pp} \end{pmatrix}$$

$$A^T X^T X A = L \qquad = [a_1, a_2, \cdots a_p]$$

$$A \cdot A^T X^T X A = AL \quad \Rightarrow \quad X^T X [a_1, \cdots, a_p] = [a_1, \cdots a_p] \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{pmatrix}$$

$$X^T X A = AL \qquad X^T X a_j = \lambda_j a_j. \quad \|a_j\| = 1 \quad y_j = X a_j$$

$\boxed{y_i^T y_j = \lambda_j}$

1. $S_{y_i}^2 = \lambda_i = \dfrac{y_i^T y_i}{n-1}$

$S_{y_1}^2 \geq S_{y_2}^2 \geq \cdots \geq S_{y_p}^2$

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \Rightarrow$ All $\lambda_i \geq 0 \quad i = 1 \cdots p$.

2. The principal components are orthorgonal and ordered $\Rightarrow$ uncorrelated

$y_i^T y_j = 0 \quad i \neq j$

3. The emp. var-cov matrix of $Y = (Y_1 \cdots Y_p)$

$$S_Y = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_p \end{pmatrix} = \frac{Y^T Y}{n-1} = L.$$

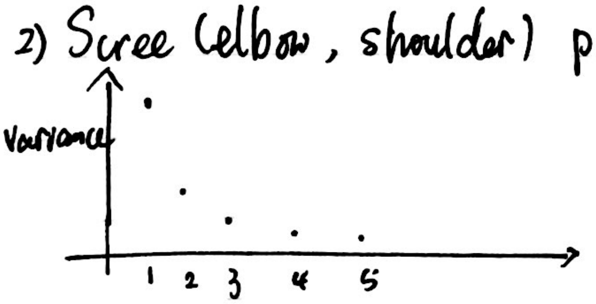
9/10 Total variance $\sum_{j=1}^{p} s_j^2 = tr(S) = \sum_{j=1}^{p} \lambda_j$

the j-component $\dfrac{\lambda_j}{\sum_{j=1}^{p} \lambda_j}$

Truncation Rule

1) λ_k s.t $\dfrac{\sum_{i=1}^{k} \lambda_i}{\sum_{j=1}^{p} \lambda_j} \geq 80 \qquad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$

2) Scree (elbow, shoulder) plot.



3) Kaiser criterion

$\bar{\lambda} = \sum_{i=1}^{p} \dfrac{\lambda_i}{p}$; $\lambda_i > \lambda$

If run PCA on $R \Rightarrow \bar{\lambda} = 1$