ANÁLISIS DE COMPONENTES PRINCIPALES

Resultados importantes

1

Sea el vector aleatorio $x=(x_1, x_2, ..., x_p)$ con matriz de covarianzas Σ cuyos valores y vectores característicos $(\lambda_1, e_1), (\lambda_2, e_2), ...,(\lambda_p, e_p)$ donde $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \geq 0$. Sean $Y_1 = e_1'x$, $Y_2 = e_2'x$, ..., $Y_p = e_p'x$ los componentes principales, entonces:

$$\sigma_{11} + \sigma_{22} + ... + \sigma_{pp} = \sum_{i=1}^{p} Var(X_i) = \lambda_1 + \lambda_2 + ... + \lambda_p = \quad var(y_1)=var(y_2)=...=var(y_p)$$

Sol.-

Dado $\Sigma = (\sigma_{ij}) \Rightarrow Tr(\Sigma) = \sigma_{11} + \sigma_{22} + \ldots + \sigma_{pp}$ --------- (i)

Pero $\Sigma = P\Lambda P'$ donde $P = (e_1 \; e_2 \ldots e_p)$

$\Rightarrow Tr(\Sigma) = Tr(P\Lambda P') = Tr(\Lambda) = \lambda_1 + \lambda_1 + \ldots + \lambda_p$ --------- (ii)

$\Rightarrow Var(y_i) = e_i'\Sigma e_i = \lambda_i e_i' e_i = \lambda_i$ ------------------------- (iii)

Luego,

De (i), (ii) y (iii):

$$\sigma_{11} + \sigma_{22} + \ldots + \sigma_{pp} = \sum_{i=1}^{p} var(y_i) = Tr(\Sigma) = \lambda_1 + \lambda_1 + \ldots + \lambda_p$$

Si $Y_1=e_1x$, $Y_2=e_2x$, …, $Y_p= e_px$ son los componentes principales obtenidos a partir de Σ, entonces,

$$\rho_{Yi,Xk} = \frac{e_{ki}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}$$

son los coeficientes de correlación entre las componentes Y_i y las variables X_k. Los pares (λ_1, e_1), (λ_2, e_2), …,(λ_p, e_p) son los valores característicos y sus vectores característicos asociados.

Sol.-

Haciendo: $l_k = \left[\, 0, \; \dots, 0, 1, 0, \dots, 0 \,\right]$

$\uparrow$ k-ésimo

$x_k = l_k'x$ y $Cov(x_k, y_i) = cov(l_k'x\,,\, e_i'x) = l_k'\Sigma e_i$

Dado que: $\Sigma e_i = \lambda_i e_i$

, entonces, $Cov(x_k, y_i) = l_k'\Sigma e_i = \lambda_i\, l_k'\, e_i = \lambda_i\, e_{ki}$

$$\rho_{Yi,Xk} = \frac{\lambda_i\, e_{ki}}{\sqrt{\lambda_i}\,\sqrt{\sigma_{kk}}} = \frac{e_{ki}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}$$

3

Si la matriz de covarianzas de **x** tiene rango “**r**” (r<p), entonces, la variación de **x** puede ser totalmente explicada por las primeras “**r**” componentes principales.

Sol.-

Si Σ tiene rango **r**, entonces,

$$\Sigma = P\Lambda P' = P \left(\begin{array}{ccc|ccc} \lambda_1 & & & & & \\ & \ddots & & & 0 & \\ & & \lambda_r & & & \\ \hline & & & \lambda_{r+1} & & \\ & 0 & & & \ddots & \\ & & & & & \lambda_p \end{array}\right) P'$$

donde $\lambda_{r+1}=\lambda_{r+2}= \ldots = \lambda_p = 0$

Luego, $$r_r = \frac{\lambda_1 + \lambda_1 + \ldots + \lambda_r}{Tr(\Lambda)} = 1$$

Ninguna combinación lineal estandarizada de $x_{(px1)}$ tiene varianza mayor que λ_1, la varianza del primer componente principal

Sol.-

Sea L´x la combinación lineal estandarizada tal que L´L=1.
Los vectores característicos de Σ: e_1, e_2, …, e_p constituyen una base de R^p

De allí que, hacemos:
L= $c_1e_1 + c_2e_2 + \ldots + c_pe_p$ y
α=L´x la combinación lineal, tal que:
$V(\alpha)=L'\Sigma L= L'(\sum_{i=1}^{p} \lambda_i e_i e'_i)L$
$\Rightarrow V(\alpha) = (c_1e_1 + c_2e_2 + \ldots + c_pe_p)' (\sum_{i=1}^{p} \lambda_i e_i e'_i) (c_1e_1 + c_2e_2 + \ldots + c_pe_p)$
$\Rightarrow V(\alpha) = \sum_{i=1}^{p} \lambda_i \, c_i^2$ (i=j).

El máximo cuando $\sum_{i=1}^{p} c_i^2 = 1$ se da para λ_1.
La $V(\alpha)$ es máxima cuando L=e_1.

Calcular la media y varianza de una muestra aleatoria de observaciones ***p*** variadas a las que se les aplica una transformación de escala.

Sol.-

Sea $\mathbf{X}_{(nxp)}$ la muestra aleatoria **p**-variada de tamaño **n**

$$\bar{y} = \frac{1}{n} Y'1_n = \frac{1}{n}(HXD^{-1})'1_n = 0 \qquad ; \quad D = diag(\sigma_{ii})$$

$$\Rightarrow \qquad \bar{y} = 0$$

$$S_y = \frac{1}{n} Y'HY = \frac{1}{n}(HXD^{-1})'H\,(HXD^{-1}) = D^{-1}S_xD^{-1}$$

$$\Rightarrow \qquad S_y = R_x$$

Gracias!!!