

Princomp function in R

Qingyang Li

The PC scores that using "princomp" command does not match the PC scores that are calculated by formula. The reason is that the help file for "princomp()" says: the default calculation uses divisor N for the covariance matrix. This means the estimated covariance matrix is calculated as

$$\frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T$$

This is the usual "biased" estimate rather than "unbiased" estimate.

$$\frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T$$

However, the eigenvalues and eigenvectors are matched with our manual calculation by using the "princomp()" command. That is when we specify "cor=TRUE" in "princomp()", the correlation matrix is used for almost all of the computations.

$$r_{jk} = \frac{\hat{\sigma}_{jk}}{\sqrt{\hat{\sigma}_{jj}\hat{\sigma}_{kk}}} \quad (1)$$

$$= \frac{\frac{1}{n} \sum (x_j - \mu_j)(x_k - \mu_k)}{\sqrt{\frac{1}{n} \sum (x_j - \mu_j)^2 \times \frac{1}{n} \sum (x_k - \mu_k)^2}} \quad (2)$$

Therefore, the $1/n$ vs. $1/(n-1)$ divisor is not going to matter based on the above formula. In conclusion, the only place where it does matter is when finding the PC scores.