PRINCIPAL COMPONENTS

In multivariate analysis the dimension of $\underline{X}$ causes problems in obtaining suitable statistical analysis to analyze a set of observations (data) on $\underline{X}$. It is natural to look for method for rearranging the data, so that with as little loss of information as possible, the dimension of problem is considerably reduced. This reduction is possible by transforming the original variable to a new set of variables which are uncorrelated. These new variables are known as principal components.

Principal components are normalized linear combination's of original variables which has specified properties interms of variance. They are uncorrelated and are ordered, so that the first component displays the largest amount of variation, the second components displays the second largest amount of variation, and so on.

If there are $p-$variables, then p components are required to reproduce (rearrange) the total variability present in the data, this variability can be accounted for by a small number, k $(k < p)$ of the components. If this is so, there is almost as much information in the k components as there is in original p variables and then k components can be replace the original p variables. . That is why; this is considered as linear reduction technique. This technique produces best results when the original variables are highly correlated, positively or negatively.

For example, suppose we are interested in finding the level of performance in mathematics of the tenth grade students of a certain school. We may then record their scores in mathematics: i.e., we consider just one characteristic of each student. Now suppose we are instead interested in overall performance and select some p characteristics, such as mathematics, English, history, science, etc. These characteristics, although related to each other, may not all contain the same amount of information, and in fact some characteristics may be completely redundant. Obviously, this will result in a loss of information and waste of resources in analyzing the data. Thus, we should select only those characteristics that will truly discriminate one student from another, while those least discriminatory should be discarded.

Determination

Let $\underline{X}$ be a $p-$component vector with covariance matrix Σ be positive definite (all the characteristic roots are positive and distinct). Since we are interested only in the variances and covariances, we shall assume that $E(\underline{X}) = \underline{0}$. Let $\underline{\beta}$ is a $p-$component column vector such that $\underline{\beta}'\underline{\beta} = 1$ (in order to obtain a unique solution).

Define,

$$U = \underline{\beta}'\underline{X}, \text{ then } EU = E(\underline{\beta}'\underline{X}) = 0, \text{ and } V(\underline{\beta}'\underline{X}) = \underline{\beta}'E(\underline{X}\,\underline{X}')\underline{\beta} = \underline{\beta}'\Sigma\underline{\beta} \tag{8.1}$$

The normalized linear combination $\underline{\beta}'\underline{X}$ with maximum variance is therefore obtained by maximizing equation (8.1) subject to condition $\underline{\beta}'\underline{\beta} = 1$. We shall use technique of Lagrange multiplier and maximize

$$\phi = \underline{\beta}'\Sigma\underline{\beta} - \lambda(\underline{\beta}'\underline{\beta} - 1) = \sum_{i,\,j}\beta_i\sigma_{ij}\beta_j - \lambda\left(\sum_i \beta_i^2 - 1\right),$$

where λ is a Lagrange multiplier, the vector of partial derivatives $\left(\dfrac{\partial \phi}{\partial \beta_i}\right)$ is

$$\frac{\partial \phi}{\partial \underline{\beta}} = 2\Sigma\underline{\beta} - 2\lambda\,\underline{\beta}\text{, equating this to zero, gives}$$

$$\Sigma\underline{\beta} - \lambda\,\underline{\beta} = \underline{0},$$

$$\Rightarrow\ (\Sigma - \lambda I)\,\underline{\beta} = \underline{0}. \tag{8.2}$$

The equation (8.2) admits a non-zero solution in $\underline{\beta}$ if only,

$$\left|\, \Sigma - \lambda I \,\right| = \underline{0}. \tag{8.3}$$

i.e. $(\Sigma - \lambda I)$ is a singular matrix.

The function $\left|\Sigma - \lambda I\right|$ is a polynomial in λ of degree p (i.e. $\lambda^p + a_1\lambda^{p-1} + \cdots + a_{p-1}\lambda + a_p$). Therefore, the equation (8.3) will give p solutions in λ, let these be $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p$.

Pre-multiplying (8.2) with $\underline{\beta}'$, gives

$$\underline{\beta}'\Sigma\,\underline{\beta} - \lambda\,\underline{\beta}'\,\underline{\beta} = 0$$

$$\Rightarrow\ \underline{\beta}'\Sigma\underline{\beta} = \lambda. \tag{8.4}$$

This shows if $\underline{\beta}$ satisfies equation (8.2) and $\underline{\beta}'\,\underline{\beta} = 1$, then the variance of $\underline{\beta}'\,\underline{X}$ is λ.

Thus for the maximum variance we should use in (8.2) the largest root λ_1. Let $\underline{\beta}^{(1)}$ be a normalized solution of

$$(\Sigma - \lambda_1 I)\,\underline{\beta} = \underline{0} \quad \text{i.e. } \Sigma\underline{\beta}^{(1)} = \lambda_1\underline{\beta}^{(1)}\text{, then}$$

$$U_1 = \underline{\beta}^{(1)'}\,\underline{X}$$

is a normalized linear combination (first principal component) with maximum variance i.e.

$$V(U_1) = \underline{\beta}^{(1)'}\,\Sigma\underline{\beta}^{(1)} = \lambda_1 \text{ [from equation (9.4)].}$$

We have next to find a normalized linear combination $\underline{\beta}'\,\underline{X}$ that has maximum variance of all linear combinations un-correlated with U_1. Non-correlation with U_1 is same as

$$0 = \operatorname{Cov}(\underline{\beta}'\,\underline{X}, U_1) = E\ (\underline{\beta}'\,\underline{X} - E\,\underline{\beta}'\,\underline{X})(\underline{\beta}^{(1)'}\,\underline{X} - E\,\underline{\beta}^{(1)'}\,\underline{X})'$$

$$= E\ (\underline{\beta}'\,\underline{X}\,\underline{X}'\,\underline{\beta}^{(1)}) = \underline{\beta}'\Sigma\underline{\beta}^{(1)}.$$

Since $\Sigma\underline{\beta}^{(1)} = \lambda_1\underline{\beta}^{(1)}$, then

$$\underline{\beta}'\Sigma\underline{\beta}^{(1)} = \lambda_1\underline{\beta}'\,\underline{\beta}^{(1)} = 0. \tag{8.5}$$

Thus we have to determine $\underline{\beta}$ as to maximize the function

$$\phi_2 = \underline{\beta}'\Sigma\underline{\beta} - \lambda(\underline{\beta}'\,\underline{\beta} - 1) - 2\upsilon_1(\underline{\beta}'\Sigma\underline{\beta}^{(1)}),$$

where λ and υ_1 are Lagrange multipliers. The vector of partial derivatives is

$$\frac{\partial\phi_2}{\partial\underline{\beta}} = 2\Sigma\underline{\beta} - 2\lambda\,\underline{\beta} - 2\,\upsilon_1\,\Sigma\underline{\beta}^{(1)}, \text{ and we set this to } \underline{0}.$$

$$\Rightarrow\ \Sigma\underline{\beta} - \lambda\,\underline{\beta} - \upsilon_1\,\Sigma\underline{\beta}^{(1)} = \underline{0}.$$

Premultiplying by $\underline{\beta}^{(1)'}$

$$\underline{\beta}^{(1)'}\Sigma\underline{\beta} - \lambda\underline{\beta}^{(1)'}\,\underline{\beta} - \upsilon_1\,\underline{\beta}^{(1)'}\Sigma\underline{\beta}^{(1)} = 0.$$

Since $\underline{\beta}'\Sigma\underline{\beta}^{(1)} = \lambda_1\underline{\beta}'\,\underline{\beta}^{(1)} = 0$, so that

$$-\upsilon_1\,\underline{\beta}^{(1)'}\Sigma\underline{\beta}^{(1)} = 0 \quad \text{or} \quad \upsilon_1\lambda_1 = 0, \text{ because } \underline{\beta}'\Sigma\underline{\beta} = \lambda \quad \Rightarrow \quad \underline{\beta}^{(1)'}\Sigma\underline{\beta}^{(1)} = \lambda_1.$$

$\Rightarrow\ \upsilon_1 = 0$, because $\lambda_1 \neq 0$.

This means that the condition of uncorrelated ness is itself satisfied when $\underline{\beta}$ satisfied (8.2) and λ satisfied (8.3).

So the procedure to find the second principal component is to solve (8.2) for $\lambda = \lambda_2$, and call the vector as $\underline{\beta}^{(2)}$ and the corresponding linear combination

$$U_2 = \underline{\beta}^{(2)'}\,\underline{X}, \text{ clearly,}$$

$$\mathrm{V}(U_2) = \underline{\beta}^{(2)'}\Sigma\underline{\beta}^{(2)} = \lambda_2$$

Similarly, $\underline{\beta}^{(3)}, \underline{\beta}^{(4)}, \cdots$, and their linear combinations

$$\underline{\beta}^{(3)'}\,\underline{X}, \underline{\beta}^{(4)'}\,\underline{X}, \cdots$$

and their variances with $\lambda_3, \lambda_4, \cdots$

As a general case, we want to find a vector $\underline{\beta}$ such that $\underline{\beta}'\,\underline{X}$ has maximum variance of all the normalized linear combinations which are uncorrelated with $U_1, U_2, \cdots, U_r$, that is

$$0 = \mathrm{Cov}(\underline{\beta}'\,\underline{X}, U_i) = E(\underline{\beta}'\,\underline{X}\,U_i) = E(\underline{\beta}'\,\underline{X}\,\underline{X}'\,\underline{\beta}^{(i)}) = \underline{\beta}'\Sigma\underline{\beta}^{(i)}, \quad i = 1, 2, \cdots, r$$

$$= \lambda_i\,\underline{\beta}'\,\underline{\beta}^{(i)}. \tag{8.6}$$

We want to maximize

$$\phi_{r+1} = \underline{\beta}'\Sigma\underline{\beta} - \lambda(\underline{\beta}'\,\underline{\beta} - 1) - 2\sum_{i=1}^{r}\upsilon_i\,\underline{\beta}'\Sigma\underline{\beta}^{(i)},$$

where λ and υ_i are Lagrange multipliers, the vector of partial derivatives is

$$\frac{\partial \phi_{r+1}}{\partial \underline{\beta}} = \underline{0} = 2\Sigma\underline{\beta} - 2\lambda\underline{\beta} - 2\sum_{i=1}^{r} \upsilon_i \, \Sigma\underline{\beta}^{(i)} \, . \tag{8.7}$$

Premultiplying by $\underline{\beta}^{(j)'}$, we obtain

$$\underline{\beta}^{(j)'} \Sigma\underline{\beta} - \lambda\underline{\beta}^{(j)'} \underline{\beta} - \upsilon_j \, \underline{\beta}^{(j)'} \Sigma\underline{\beta}^{(j)} = 0, \text{ using equation (8.5)}$$

$\Rightarrow \; -\upsilon_j \, \underline{\beta}^{(j)'} \Sigma\underline{\beta}^{(j)} = 0$

$\Rightarrow \; \upsilon_j \lambda_j = 0$, because $\underline{\beta}' \Sigma \underline{\beta} = \lambda$

$\Rightarrow \; \upsilon_j = 0$, because $\lambda_j \neq 0$.

Thus equation (8.7) becomes

$$\Sigma\underline{\beta} - \lambda\,\underline{\beta} = \underline{0}$$

$\Rightarrow \; (\Sigma - \lambda \, I)\underline{\beta} = \underline{0}$

and therefore, $\underline{\beta}$ satisfies (8.2) and λ must satisfies (8.3).

Let λ_{r+1} be the maximum of $\lambda_1, \lambda_2, \cdots, \lambda_p$, such that there exists a vector $\underline{\beta}$ satisfying $(\Sigma - \lambda_{r+1} \, I)\underline{\beta} = \underline{0}$, $\underline{\beta}' \underline{\beta} = 1$ and equation (8.6), call this vector $\underline{\beta}^{(r+1)}$ and corresponding linear combination $U_{r+1} = \underline{\beta}^{(r+1)'} \underline{X}$.

We shall write,

$$\mathrm{B}_{p\times p} = (\underline{\beta}^{(1)} \quad \underline{\beta}^{(2)} \quad \cdots \quad \underline{\beta}^{(p)})$$

and

$$\Lambda = diag\,(\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_p) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix}$$

$$\underline{U} = \mathrm{B}' \underline{X} \, .$$

Since $\Sigma\underline{\beta}^{(r)} = \lambda_r \underline{\beta}^{(r)}$ for $r = 1, 2, \cdots, p$ can be written in matrix form as

$$\Sigma \mathrm{B} = \mathrm{B}\Lambda \tag{8.8}$$

and the equations $\underline{\beta}^{(r)'} \underline{\beta}^{(r)} = 1$ and $\underline{\beta}^{(r)'} \underline{\beta}^{(s)} = 0$, $r \neq s$ can be written as

$$\mathrm{B}'\mathrm{B} = \mathrm{I} \tag{8.9}$$

i.e. B is an orthogonal matrix

From equations (8.8) and (8.9) we obtain

$$\mathrm{B}'\Sigma\mathrm{B} = \mathrm{B}'\mathrm{B}\Lambda = \Lambda \, .$$

Thus given a positive definite matrix Σ, there exist an orthogonal matrix B such that $\mathrm{B}'\Sigma\mathrm{B}$ is in diagonal form and the diagonal elements of $\mathrm{B}'\Sigma\mathrm{B}$ are the characteristic roots of Σ.

Therefore, we proceed as follows:

Solve $|\Sigma - \lambda I| = 0$. Let the roots be $\lambda_1 > \lambda_2 > \cdots > \lambda_p$, then solve

$(\Sigma - \lambda_1 \mathrm{I})\underline{\beta} = \underline{0}$, get a solution $\underline{\beta}^{(1)}$ i.e. $\Sigma\underline{\beta}^{(1)} = \lambda_1 \underline{\beta}^{(1)}$.

First principal component

$$U_1 = \underline{\beta}^{(1)'} \underline{X} .$$

Again solve

$(\Sigma - \lambda_2 \mathrm{I})\underline{\beta} = \underline{0}$, get a solution $\underline{\beta}^{(2)}$ i.e. $\Sigma\underline{\beta}^{(2)} = \lambda_2 \underline{\beta}^{(2)}$.

Second principal component

$$U_2 = \underline{\beta}^{(2)'} \underline{X}$$

$$\vdots$$

$r-$th principal component

$$U_r = \underline{\beta}^{(r)'} \underline{X}$$

Stop at m $(\le p)$ for which λ_m is negligible as compared to $\lambda_1, \cdots, \lambda_{m-1}$. Thus we get a fewer (countable less) linear combinations or at most as many linear combinations as the original variable (p).

Contribution of first principal component $= \dfrac{\lambda_1}{\sum_{i=1}^{p} \lambda_i}$

Contribution of $(I + II)$ principal component $= \dfrac{\lambda_1 + \lambda_2}{\sum_{i=1}^{p} \lambda_i}$

Contribution of $(I + II + III)$ principal component $= \dfrac{\lambda_1 + \lambda_2 + \lambda_3}{\sum_{i=1}^{p} \lambda_i}$.

Theorem: An orthogonal transformation $\underline{V} = C\underline{X}$ of a random vector $\underline{X}$ leaves invariant the generalized variance and the sum of the variances of the components.

Proof: Let $E\,\underline{X} = \underline{0}$, and $E\,\underline{X}\,\underline{X}' = \Sigma$, then $E(\underline{V}) = \underline{0}$ and $\Sigma_V = E(\underline{V}\,\underline{V}') = E(C\underline{X}\,\underline{X}'C') = C\Sigma C'$.

The generalized variance of $\underline{V}$ is

$|C\Sigma C'| = |C||\Sigma||C'| = |\Sigma||CC'| = |\Sigma|$, because C is orthogonal matrix, $\Rightarrow$ $CC' = I$

$=$ generalized variance of $\underline{X}$.

The sum of the variances of the components of $\underline{V}$ is

$$\sum_{i=1}^{p} E(V_i)^2 = tr\,(C\Sigma C') = tr\,\Sigma C'C = tr\,\Sigma I = tr\,\Sigma = \sum_{i=1}^{p} E(X_i)^2 .$$

Corollary: The generalized variance of the vector of principal components is the generalized variance of the original vector.

Proof: We have,

$\underline{U} = \mathrm{B}'\,\underline{X}$, and $\Sigma_{\mathrm{U}} = \mathrm{B}'\Sigma\mathrm{B}$, then

$|\Sigma_{\mathrm{U}}| = |\mathrm{B}'\Sigma\mathrm{B}| = |\Sigma||\mathrm{B}'\mathrm{B}| = |\Sigma|\,|\mathrm{I}| = |\Sigma|$.

Corollary: The sum of the variances of the principal components is equal to the sum of the variances of the original variates.

Proof: We have, $\underline{U} = \mathrm{B}'\,\underline{X}$, and $\Sigma_{\mathrm{U}} = \mathrm{B}'\Sigma\mathrm{B}$, then

$tr\,\Sigma_U = tr\,\mathrm{B}'\Sigma\mathrm{B} = tr\,\Sigma\mathrm{B}'\mathrm{B} = tr\,\Sigma\mathrm{I} = tr\,\Sigma$.

But

$tr\,\Sigma_U = V(u_1) + \cdots V(u_p)$, and $tr\,\Sigma_X = V(x_1) + \cdots V(x_p)$.

Exercise: Let $R = \begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix}$. Solve

$|R - \lambda I| = 0$, or $\begin{vmatrix} 1-\lambda & r_{12} \\ r_{12} & 1-\lambda \end{vmatrix}$, or $(1-\lambda)^2 - r_{12}^2 = 0$

or $\lambda^2 - 2\lambda + 1 - r^2 = 0$, $r_{12} = r$

or $\lambda = \dfrac{2 \pm \sqrt{4 - 4(1-r^2)}}{2} = 1 \pm r$

If $r > 0$ $\quad\lambda_1 = 1 + r$, $\lambda_2 = 1 - r$

If $r < 0$ $\quad\lambda_1 = 1 - r$, $\lambda_2 = 1 + r$

If $r = 0$ $\quad\lambda_1 = 1 = \lambda_2$,

i.e. in case of perfect correlation, we need one principal component which explains fully but in case of zero correlation, no principal component.

Exercise: Find the variance of the first principle component of the covariance matrix Σ defined by

$\Sigma = (1-\rho)\mathrm{I} + \rho\,\underline{e}\,\underline{e}'$, where $\underline{e}' = (1 \;\; 1 \;\; \cdots \;\; 1)$.

Exercise: Find the characteristic vector of $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ corresponding to the characteristic roots $1+\rho$ and $1-\rho$.