PRINCIPAL COMPONENTS

Explaining the variance-covariance structure of a set of variables through linear combinations of the variables is the goal of principal components analysis. The two primary purposes of this technique are (a) data reduction and (b) interpretation. This procedure is frequently used because there are (a) too many independent variables relative to the number of observations and (b) some variables are highly correlated, producing unstable estimates.

P components are required to reproduce all of the variability in the system; however, most of the variability can usually be accounted for by fewer than *p* components. If there are *k* components that account for most of the variability of the *p* variables, then we say that the data set has been reduced to *k* principal components.

Principal components consist of linear combinations of the *p* variables in the data set. Geometrically, the linear combinations represent a new coordinate system that was obtained by rotating the original system to maximize the variability of a simpler representation of the *p* coordinate axes (from the original *p* variables in the system).

Let $\underline{x}' = (x_1, x_2, \ldots, x_3)$ with the variance-covariance matrix Σ (sigma) and eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$.

Consider linear combinations of the $\underline{x}$s:

$$Y_1 = \underline{a}'_1 \mathbf{X} = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p$$
$$Y_2 = \underline{a}'_2 \mathbf{X} = a_{21}X_1 + a_{22}X_2 + \ldots + a_{2p}X_p$$
$$\vdots$$
$$Y_p = \underline{a}'_p \mathbf{X} = a_{p1}X_1 + a_{p2}X_2 + \ldots + a_{pp}X_p$$

Then we can obtain

$$\text{Var}(Y_i) = \underline{a}'_i \Sigma \underline{a}_i \text{ and } \text{Cov}(Y_i, Y_k) = \underline{a}'_i \Sigma \underline{a}_k$$

The principal components are the uncorrelated linear combinations $Y_1, Y_2, \ldots, Y_p$ with variances as large as possible.

The first principal component is the linear combination with the maximum variance; it maximizes $\text{Var}(Y_1) = \underline{a}'_1 \Sigma \underline{a}_1$. Of course, the variance can be increased by multiplying any $\underline{a}$ by a constant. To avoid this indeterminacy, coefficient vectors are constrained to be unit length.

First principal component = linear combination $\underline{a}'_1 \mathbf{X}$ that maximizes $\text{Var}(\underline{a}_1\mathbf{X})$ so that $\underline{a}'_1\underline{a}_1 = 1$.
Second principal component = linear comb. $\underline{a}'_2 \mathbf{X}$ that maximizes $\text{Var}(\underline{a}_2\mathbf{X})$ so that $\underline{a}'_2\underline{a}_2 = 1$ and $\text{Cov}(\underline{a}'_1 \mathbf{X}, \underline{a}'_2 \mathbf{X}) = 0$.

Etc…