Principal Components Analysis

Nathaniel E. Helwig
Assistant Professor of Psychology and Statistics
University of Minnesota (Twin Cities)

Updated 16-Mar-2017
Outline of Notes

1) Background
   • Overview
   • Data, Cov, Cor
   • Orthogonal Rotation

2) Population PCs
   • Definition
   • Calculation
   • Properties

3) Sample PCs
   • Definition
   • Calculation
   • Properties

4) PCA in Practice
   • Covariance or Correlation?
   • Decathlon Example
   • Number of Components
Background
Definition and Purposes of PCA

Principal Components Analysis (PCA) finds linear combinations of variables that best explain the covariation structure of the variables.

There are two typical purposes of PCA:

1. Data reduction: explain covariation between $p$ variables using $r < p$ linear combinations
2. Data interpretation: find features (i.e., components) that are important for explaining covariation
Data Matrix

The data matrix refers to the array of numbers

\[
X = \begin{pmatrix}
  x_{11} & x_{12} & \cdots & x_{1p} \\
  x_{21} & x_{22} & \cdots & x_{2p} \\
  x_{31} & x_{32} & \cdots & x_{3p} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{n1} & x_{n2} & \cdots & x_{np}
\end{pmatrix}
\]

where \( x_{ij} \) is the \( j \)-th variable collected from the \( i \)-th item (e.g., subject).

- items/subjects are rows
- variables are columns

\( X \) is a data matrix of order \( n \times p \) (# items by # variables).
Population Covariance Matrix

The **population covariance matrix** refers to the symmetric array

\[
\Sigma = \begin{pmatrix}
\sigma_{11} & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1p} \\
\sigma_{21} & \sigma_{22} & \sigma_{23} & \cdots & \sigma_{2p} \\
\sigma_{31} & \sigma_{32} & \sigma_{33} & \cdots & \sigma_{3p} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\sigma_{p1} & \sigma_{p2} & \sigma_{p3} & \cdots & \sigma_{pp}
\end{pmatrix}
\]

where

- \( \sigma_{jj} = E([X_j - \mu_j]^2) \) is the **population variance** of the \( j \)-th variable
- \( \sigma_{jk} = E([X_j - \mu_j][X_k - \mu_k]) \) is the **population covariance** between the \( j \)-th and \( k \)-th variables
- \( \mu_j = E(X_j) \) is the **population mean** of the \( j \)-th variable
Sample Covariance Matrix

The sample covariance matrix refers to the symmetric array

\[
S = \begin{pmatrix}
    s_{11} & s_{12} & s_{13} & \cdots & s_{1p} \\
    s_{21} & s_{22} & s_{23} & \cdots & s_{2p} \\
    s_{31} & s_{32} & s_{33} & \cdots & s_{3p} \\
      & \vdots & \vdots & \ddots & \vdots \\
    s_{p1} & s_{p2} & s_{p3} & \cdots & s_{pp}
\end{pmatrix}
\]

where

- \( s_{jj} = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2 \) is the sample variance of the \( j \)-th variable
- \( s_{jk} = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \) is the sample covariance between the \( j \)-th and \( k \)-th variables
- \( \bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij} \) is the sample mean of the \( j \)-th variable
**Covariance Matrix from Data Matrix**

We can calculate the (sample) covariance matrix such as

\[
S = \frac{1}{n-1} X'_c X_c
\]

where \( X_c = X - 1_n \bar{x}' = CX \) with
- \( \bar{x}' = (\bar{x}_1, \ldots, \bar{x}_p) \) denoting the vector of variable means
- \( C = I_n - n^{-1}1_n1'_n \) denoting a centering matrix

Note that the centered matrix \( X_c \) has the form

\[
X_c = \begin{pmatrix}
x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{1p} - \bar{x}_p \\
x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{2p} - \bar{x}_p \\
x_{31} - \bar{x}_1 & x_{32} - \bar{x}_2 & \cdots & x_{3p} - \bar{x}_p \\
\vdots & \vdots & \ddots & \vdots \\
x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \cdots & x_{np} - \bar{x}_p
\end{pmatrix}
\]
The population correlation matrix refers to the symmetric array

\[
P = \begin{pmatrix}
1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1p} \\
\rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2p} \\
\rho_{31} & \rho_{32} & 1 & \cdots & \rho_{3p} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho_{p1} & \rho_{p2} & \rho_{p3} & \cdots & 1
\end{pmatrix}
\]

where

\[
\rho_{jk} = \frac{\sigma_{jk}}{\sqrt{\sigma_{jj}\sigma_{kk}}}
\]

is the Pearson correlation coefficient between variables \(X_j\) and \(X_k\).
The sample correlation matrix refers to the symmetric array

$$R = \begin{pmatrix}
1 & r_{12} & r_{13} & \cdots & r_{1p} \\
 r_{21} & 1 & r_{23} & \cdots & r_{2p} \\
 r_{31} & r_{32} & 1 & \cdots & r_{3p} \\
 \vdots & \vdots & \vdots & \ddots & \vdots \\
 r_{p1} & r_{p2} & r_{p3} & \cdots & 1
\end{pmatrix}$$

where

$$r_{jk} = \frac{s_{jk}}{s_j s_k} = \frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2} \sqrt{\sum_{i=1}^{n} (x_{ik} - \bar{x}_k)^2}}$$

is the Pearson correlation coefficient between variables $x_j$ and $x_k$. 

Correlation Matrix from Data Matrix

We can calculate the (sample) correlation matrix such as

\[ R = \frac{1}{n-1} X'_s X_s \]

where \( X_s = C X D^{-1} \) with

- \( C = I_n - n^{-1} 1_n 1'_n \) denoting a centering matrix
- \( D = \text{diag}(s_1, \ldots, s_p) \) denoting a diagonal scaling matrix

Note that the standardized matrix \( X_s \) has the form

\[
X_s = \begin{pmatrix}
\frac{(x_{11} - \bar{x}_1)}{s_1} & \frac{(x_{12} - \bar{x}_2)}{s_2} & \cdots & \frac{(x_{1p} - \bar{x}_p)}{s_p} \\
\frac{(x_{21} - \bar{x}_1)}{s_1} & \frac{(x_{22} - \bar{x}_2)}{s_2} & \cdots & \frac{(x_{2p} - \bar{x}_p)}{s_p} \\
\frac{(x_{31} - \bar{x}_1)}{s_1} & \frac{(x_{32} - \bar{x}_2)}{s_2} & \cdots & \frac{(x_{3p} - \bar{x}_p)}{s_p} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{(x_{n1} - \bar{x}_1)}{s_1} & \frac{(x_{n2} - \bar{x}_2)}{s_2} & \cdots & \frac{(x_{np} - \bar{x}_p)}{s_p}
\end{pmatrix}
\]
Suppose we have \( \mathbf{z} = (x, y)' \in \mathbb{R}^2 \), i.e., points in 2D Euclidean space.

A 2 \times 2 \text{ orthogonal rotation of } (x, y) \text{ of the form}

\[
\begin{pmatrix}
  x^* \\
  y^*
\end{pmatrix} = \begin{pmatrix}
  \cos(\theta) & -\sin(\theta) \\
  \sin(\theta) & \cos(\theta)
\end{pmatrix}
\begin{pmatrix}
  x \\
  y
\end{pmatrix}
\]

rotates \((x, y)\) counter-clockwise around the origin by an angle of \(\theta\) and

\[
\begin{pmatrix}
  x^* \\
  y^*
\end{pmatrix} = \begin{pmatrix}
  \cos(\theta) & \sin(\theta) \\
  -\sin(\theta) & \cos(\theta)
\end{pmatrix}
\begin{pmatrix}
  x \\
  y
\end{pmatrix}
\]

rotates \((x, y)\) clockwise around the origin by an angle of \(\theta\).
Visualization of 2D Clockwise Rotation

No Rotation

30 degrees

45 degrees

60 degrees

90 degrees

180 degrees

Nathaniel E. Helwig (U of Minnesota)
Visualization of 2D Clockwise Rotation (R Code)

```r
rotmat2d <- function(theta) {
    matrix(c(cos(theta), sin(theta), -sin(theta), cos(theta)), 2, 2)
}

x <- seq(-2, 2, length=11)
y <- 4*exp(-x^2) - 2
xy <- cbind(x, y)
rang <- c(30, 45, 60, 90, 180)
dev.new(width=12, height=8, noRStudioGD=TRUE)
par(mfrow=c(2, 3))
plot(x, y, xlim=c(-3, 3), ylim=c(-3, 3), main="No Rotation")
text(x, y, labels=letters[1:11], cex=1.5)
for(j in 1:5) {
    rmat <- rotmat2d(rang[j]*2*pi/360)
    xyrot <- xy %*% rmat
    plot(xyrot, xlim=c(-3, 3), ylim=c(-3, 3))
text(xyrot, labels=letters[1:11], cex=1.5)
title(paste(rang[j], " degrees"))
}
```
Orthogonal Rotation in Two Dimensions

Note that the $2 \times 2$ rotation matrix

$$
R = \begin{pmatrix}
\cos(\theta) & -\sin(\theta) \\
\sin(\theta) & \cos(\theta)
\end{pmatrix}
$$

is an orthogonal matrix for all $\theta$:

$$
R' R = \begin{pmatrix}
\cos(\theta) & \sin(\theta) \\
-\sin(\theta) & \cos(\theta)
\end{pmatrix}
\begin{pmatrix}
\cos(\theta) & -\sin(\theta) \\
\sin(\theta) & \cos(\theta)
\end{pmatrix}
$$

$$
= \begin{pmatrix}
\cos^2(\theta) + \sin^2(\theta) & \cos(\theta)\sin(\theta) - \cos(\theta)\sin(\theta) \\
\cos(\theta)\sin(\theta) - \cos(\theta)\sin(\theta) & \cos^2(\theta) + \sin^2(\theta)
\end{pmatrix}
$$

$$
= \begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix}
$$
Orthogonal Rotation in Higher Dimensions

Suppose we have a data matrix $X$ with $p$ columns.

- Rows of $X$ are coordinates of points in $p$-dimensional space
- Note: when $p = 2$ we have situation on previous slides

A $p \times p$ orthogonal rotation is an orthogonal linear transformation.

- $R'R = RR' = I_p$ where $I_p$ is $p \times p$ identity matrix
- If $\tilde{X} = XR$ is rotated data matrix, then $\tilde{X}\tilde{X}' = XX'$
- Orthogonal rotation preserves relationships between points
Population Principal Components
**Linear Combinations of Random Variables**

\[ \mathbf{X} = (X_1, \ldots, X_p)' \] is a random vector with covariance matrix \( \mathbf{\Sigma} \), where \( \lambda_1 \geq \cdots \geq \lambda_p \geq 0 \) are the eigenvalues of \( \mathbf{\Sigma} \).

Consider forming new variables \( Y_1, \ldots, Y_p \) by taking \( p \) different linear combinations of the \( X_j \) variables:

\[
\begin{align*}
Y_1 &= \mathbf{b}_1' \mathbf{X} = b_{11} X_1 + b_{21} X_2 + \cdots + b_{p1} X_p \\
Y_2 &= \mathbf{b}_2' \mathbf{X} = b_{12} X_1 + b_{22} X_2 + \cdots + b_{p2} X_p \\
& \quad \vdots \\
Y_p &= \mathbf{b}_p' \mathbf{X} = b_{1p} X_1 + b_{2p} X_2 + \cdots + b_{pp} X_p
\end{align*}
\]

where \( \mathbf{b}_k = (b_{1k}, \ldots, b_{pk}) \) is the \( k \)-th linear combination vector.

- \( \mathbf{b}_k \) are called the **loadings** for the \( k \)-th principal component.
Defining Principal Components in the Population

Note that the random variable $Y_k = b'_k X$ has the properties:

$$\text{Var}(Y_k) = b'_k \Sigma b_k$$
$$\text{Cov}(Y_k, Y_\ell) = b'_k \Sigma b_\ell$$

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

$$b_1 = \arg\max \{ \text{Var}(b'_1 X) \} \quad ||b_1||=1$$
$$b_2 = \arg\max \{ \text{Var}(b'_2 X) \} \quad \text{subject to} \quad \text{Cov}(b'_1 X, b'_2 X) = 0 \quad ||b_2||=1$$
$$b_\ell = \arg\max \{ \text{Var}(b'_\ell X) \} \quad \text{subject to} \quad \text{Cov}(b'_k X, b'_\ell X) = 0 \quad \forall \ k < \ell$$
Figure 8.1 The constant density ellipse $x' \Sigma^{-1} x = c^2$ and the principal components $y_1, y_2$ for a bivariate normal random vector $X$ having mean $\mu$.

Figure: Figure 8.1 from Applied Multivariate Statistical Analysis, 6th Ed (Johnson & Wichern). Note that $e_1$ and $e_2$ denote the eigenvectors of $\Sigma$. 
We can write the population covariance matrix $\Sigma$ such as

$$\Sigma = V \Lambda V' = \sum_{k=1}^{p} \lambda_k v_k v_k'$$

where

- $V = [v_1, \ldots, v_p]$ contains the eigenvectors ($V'V = VV' = I_p$)
- $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$ contains the (non-negative) eigenvalues

The PCA solution is obtained by setting $b_k = v_k$ for $k = 1, \ldots, p$:

- $\text{Var}(Y_k) = \text{Var}(v_k'X) = v_k' \Sigma v_k = v_k' V \Lambda V' v_k = \lambda_k$
- $\text{Cov}(Y_k, Y_\ell) = \text{Cov}(v_k'X, v_\ell'X) = v_k' \Sigma v_\ell = v_k' V \Lambda V' v_\ell = 0$ if $k \neq \ell$
The proportion of the total variance accounted for by the $k$-th PC is

$$R_k^2 = \frac{\lambda_k}{\sum_{\ell=1}^{p} \lambda_{\ell}}$$

If $\sum_{k=1}^{r} R_k^2 \approx 1$ for some $r < p$, we do not lose much transforming the original variables into fewer new (principal component) variables.
The covariance between $X_j$ and $Y_k$ has the form

$$\text{Cov}(X_j, Y_k) = \text{Cov}(e'_j X, v'_k X) = e'_j \Sigma v_k = e'_j (V \Lambda V') v_k = e'_j v_k \lambda_k = v_{jk} \lambda_k$$

where

- $e'_j$ is a vector of zeros with a one in the $j$-th position
- $v_k = (v_{1k}, \ldots, v_{pk})'$ is the $k$-th eigenvector of $\Sigma$

This implies that the correlation between $X_j$ and $Y_k$ has the form

$$\text{Cor}(X_j, Y_k) = \frac{\text{Cov}(X_j, Y_k)}{\sqrt{\text{Var}(X_j)} \sqrt{\text{Var}(Y_k)}} = \frac{v_{jk} \lambda_k}{\sqrt{\sigma_{jj}} \sqrt{\lambda_k}} = \frac{v_{jk} \sqrt{\lambda_k}}{\sqrt{\sigma_{jj}}}$$
Sample Principal Components
Linear Combinations of Observed Random Variables

\( x_i = (x_{i1}, \ldots, x_{ip})' \) is an observed random vector and \( x_i \overset{iid}{\sim} (\mu, \Sigma) \).

Consider forming new variables \( y_{i1}, \ldots, y_{ip} \) by taking \( p \) different linear combinations of the \( x_{ij} \) variables:

\[
\begin{align*}
y_{i1} &= b_1' x_i = b_{11} x_{i1} + b_{21} x_{i2} + \cdots + b_{p1} x_{ip} \\
y_{i2} &= b_2' x_i = b_{12} x_{i1} + b_{22} x_{i2} + \cdots + b_{p2} x_{ip} \\
&\vdots \\cdots \\vdots \\
y_{ip} &= b_p' x_i = b_{1p} x_{i1} + b_{2p} x_{i2} + \cdots + b_{pp} x_{ip}
\end{align*}
\]

where \( b_k' = (b_{1k}, \ldots, b_{pk}) \) is the \( k \)-th linear combination vector.
Sample Properties of Linear Combinations

Note that the sample mean and variance of the $y_{ik}$ variables are:

$$\bar{y}_k = \frac{1}{n} \sum_{i=1}^{n} y_{ik} = b'_k \bar{x}$$

$$s^2_{y_k} = \frac{1}{n-1} \sum_{i=1}^{n} (y_{ik} - \bar{y}_k)^2 = \frac{1}{n-1} \sum_{i=1}^{n} (b'_k x_i - b'_k \bar{x})(b'_k x_i - b'_k \bar{x})'$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} b'_k (x_i - \bar{x})(x_i - \bar{x})' b_k = b'_k S b_k$$

and the sample covariance between $y_{ik}$ and $y_{i\ell}$ is given by

$$s_{y_k y_\ell} = \frac{1}{n-1} \sum_{i=1}^{n} (y_{ik} - \bar{y}_k)(y_{i\ell} - \bar{y}_\ell) = b'_k S b_\ell$$
The **principal components** are the *uncorrelated* linear combinations \( y_{i1}, \ldots, y_{ip} \) whose sample variances are as large as possible.

\[
\mathbf{b}_1 = \arg\max \{ \mathbf{b}'_1 \mathbf{S} \mathbf{b}_1 \} \\
\| \mathbf{b}_1 \| = 1
\]

\[
\mathbf{b}_2 = \arg\max \{ \mathbf{b}'_2 \mathbf{S} \mathbf{b}_2 \} \quad \text{subject to} \quad \mathbf{b}'_1 \mathbf{S} \mathbf{b}_2 = 0 \\
\| \mathbf{b}_2 \| = 1
\]

\[
\vdots
\]

\[
\mathbf{b}_\ell = \arg\max \{ \mathbf{b}'_\ell \mathbf{S} \mathbf{b}_\ell \} \quad \text{subject to} \quad \mathbf{b}'_k \mathbf{S} \mathbf{b}_\ell = 0 \, \forall \, k < \ell \\
\| \mathbf{b}_\ell \| = 1
\]
PCA Solution via the Eigenvalue Decomposition

We can write the sample covariance matrix $\mathbf{S}$ such as

$$\mathbf{S} = \hat{\mathbf{V}} \hat{\mathbf{\Lambda}} \hat{\mathbf{V}}' = \sum_{k=1}^{p} \hat{\lambda}_k \hat{\mathbf{v}}_k \hat{\mathbf{v}}_k'$$

where

- $\hat{\mathbf{V}} = [\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_p]$ contains the eigenvectors ($\hat{\mathbf{V}}' \hat{\mathbf{V}} = \hat{\mathbf{V}} \hat{\mathbf{V}}' = \mathbf{I}_p$)
- $\hat{\mathbf{\Lambda}} = \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_p)$ contains the (non-negative) eigenvalues

The PCA solution is obtained by setting $\mathbf{b}_k = \hat{\mathbf{v}}_k$ for $k = 1, \ldots, p$:

- $s_{y_k}^2 = \hat{\mathbf{v}}_k' \mathbf{S} \hat{\mathbf{v}}_k = \hat{\mathbf{v}}_k' \hat{\mathbf{V}} \hat{\mathbf{\Lambda}} \hat{\mathbf{V}}' \hat{\mathbf{v}}_k = \hat{\lambda}_k$
- $s_{y_k y_\ell} = \hat{\mathbf{v}}_k' \mathbf{S} \hat{\mathbf{v}}_\ell = \hat{\mathbf{v}}_k' \hat{\mathbf{V}} \hat{\mathbf{\Lambda}} \hat{\mathbf{V}}' \hat{\mathbf{v}}_\ell = 0$ if $k \neq \ell$
\{y_{i1}, \ldots, y_{ip}\}_{i=1}^n \text{ has the same total variance as } \{x_{i1}, \ldots, x_{ip}\}_{i=1}^n:\n
\sum_{j=1}^{p} s_k^2 = \text{tr}(S) = \text{tr}(\hat{V}\hat{\Lambda}\hat{V}') = \text{tr}(\hat{\Lambda}) = \sum_{j=1}^{p} s_{y_k}^2

The proportion of the total variance accounted for by the \( k \)-th PC is

\[ \hat{R}_k^2 = \frac{\hat{\lambda}_k}{\sum_{\ell=1}^{p} \hat{\lambda}_\ell} \]

If \( \sum_{k=1}^{r} \hat{R}_k^2 \approx 1 \) for some \( r < p \), we do not lose much transforming the original variables into fewer new (principal component) variables.
The sample covariance between $x_{ij}$ and $y_{ik}$ has the form

$$
\text{Cov}(x_{ij}, y_{ik}) = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(y_{ik} - \bar{y}_k)
$$

$$
= \frac{1}{n-1} \sum_{i=1}^{n} e'_j(x_i - \bar{x})(x_i - \bar{x})'\hat{v}_k
$$

$$
= e'_j S \hat{v}_k = e'_j \hat{v}_k \hat{\lambda}_k = \hat{v}_{jk} \hat{\lambda}_k
$$

where
- $e_j$ is a vector of zeros with a one in the $j$-th position
- $\hat{v}_k = (\hat{v}_{1k}, \ldots, \hat{v}_{pk})'$ is the $k$-th eigenvector of $S$

This implies that the (sample) correlation between $x_{ij}$ and $y_{ik}$ is

$$
\text{Cor}(x_{ij}, y_{ik}) = \frac{\text{Cov}(x_{ij}, y_{ik})}{\sqrt{\text{Var}(x_{ij})} \sqrt{\text{Var}(y_{ik})}} = \frac{\hat{v}_{jk} \hat{\lambda}_k}{s_j \hat{\lambda}_k^{1/2}} = \frac{\hat{v}_{jk} \hat{\lambda}_k^{1/2}}{s_j}
$$
Assume that $\mathbf{x}_i \overset{iid}{\sim} \mathcal{N}(\mu, \Sigma)$ and that the eigenvalues of $\Sigma$ are strictly positive and unique: $\lambda_1 > \cdots > \lambda_p > 0$.

As $n \to \infty$, we have that

$$\sqrt{n}(\hat{\lambda} - \lambda) \approx \mathcal{N}(0, 2\Lambda^2)$$
$$\sqrt{n}(\hat{v}_k - v_k) \approx \mathcal{N}(0, V_k)$$

where $V_k = \lambda_k \sum_{\ell \neq k} \frac{\lambda_{\ell}}{(\lambda_{\ell} - \lambda_k)^2} \mathbf{v}_{\ell} \mathbf{v}'_{\ell}$

Furthermore, as $n \to \infty$, we have that $\hat{\lambda}_k$ and $\hat{v}_k$ are independent.
Principal Components Analysis in Practice
Let \( \hat{U}\hat{D}\hat{V}' \) denote the SVD of \( X_c = CX \).

The PCA (covariance) solution is directly related to the SVD of \( X_c \):

\[
Y = \hat{U}\hat{D} \quad \text{and} \quad B = \hat{V}
\]

Note that columns of \( \hat{V} \) are the . . .

- Right singular vectors of the mean-centered data matrix \( X_c \)
- Eigenvectors of the covariance matrix \( S = \frac{1}{n-1} X_c' X_c = \hat{V}\hat{\Lambda}\hat{V}' \) where \( \hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_p) \) with \( \hat{\lambda}_k = \frac{\hat{d}_{kk}}{n-1} \) and \( \hat{D} = \text{diag}(\hat{d}_{11}, \ldots, \hat{d}_{pp}) \).
Let $\tilde{U}\tilde{D}\tilde{V}'$ denote the SVD of $X_s = CXD^{-1}$ with $D = \text{diag}(s_1, \ldots, s_p)$.

The PCA (correlation) solution is directly related to the SVD of $X_{cs}$:

$$Y = \tilde{U}\tilde{D} \quad \text{and} \quad B = \tilde{V}$$

Note that columns of $\tilde{V}$ are the...

- Right singular vectors of the centered and scaled data matrix $X_s$
- Eigenvectors of the correlation matrix $R = \frac{1}{n-1}X_s'X_s = \tilde{V}\tilde{\Lambda}\tilde{V}'$ where $\tilde{\Lambda} = \text{diag}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_p)$ with $\tilde{\lambda}_k = \frac{\tilde{d}_{kk}^2}{n-1}$ and $\tilde{D} = \text{diag}(\tilde{d}_{11}, \ldots, \tilde{d}_{pp})$. 
Covariance versus Correlation Considerations

Problem: there is no simple relationship between SVDs of $X_c$ and $X_s$.

- No simple relationship between PCs obtained from $S$ and $R$
- Rescaling variables can fundamentally change our results

Note that PCA is trying to explain the variation in $S$ or $R$

- If units of $p$ variables are comparable, covariance PCA may be more informative (because units of measurement are retained)
- If units of $p$ variables are incomparable, correlation PCA may be more appropriate (because units of measurement are removed)
Men’s Olympic Decathlon Data from 1988

Data from men’s 1988 Olympic decathlon

- Total of $n = 34$ athletes
- Have $p = 10$ variables giving score for each decathlon event
- Have overall decathlon score also ($\text{score}$)

```r
> decathlon[1:9,]
     run100 long.jump shot high.jump run400 hurdle discus pole.vault javelin run1500 score
Schenk  11.25    7.43  15.48     2.27   48.90  15.13  49.28    4.7  61.32  268.95  8488
Voss    10.87    7.45  14.97     1.97   47.71  14.46  44.36    5.1  61.76  273.02  8399
Steen   11.18    7.44  14.20     1.97   48.29  14.81  43.66    5.2  64.16  263.20  8328
Thompson 10.62    7.38  15.02     2.03   49.06  14.72  44.80    4.9  64.04  285.11  8306
Blondel 11.02    7.43  12.92     1.97   47.44  14.40  41.20    5.2  57.46  256.64  8286
Plaziat  10.83    7.72  13.58     2.12   48.34  14.18  43.06    4.9  52.18  274.07  8272
Bright   11.18    7.05  14.12     2.06   49.34  14.39  41.68    5.7  61.60  291.20  8216
De.Wit   11.05    6.95  15.34     2.00   48.21  14.36  41.32    4.8  63.00  265.86  8189
Johnson  11.15    7.12  14.52     2.03   49.15  14.66  42.36    4.9  66.46  269.62  8180
```
For the running events (\texttt{run100}, \texttt{run400}, \texttt{run1500}, and \texttt{hurdle}), lower scores correspond to better performance, whereas higher scores represent better performance for other events.

To make interpretation simpler, we will resign the running events:

```r
> decathlon[,c(1,5,6,10)] <- (-1)*decathlon[,c(1,5,6,10)]
> decathlon[1:9,]
```

<table>
<thead>
<tr>
<th></th>
<th>run100</th>
<th>long.jump</th>
<th>shot</th>
<th>high.jump</th>
<th>run400</th>
<th>hurdle</th>
<th>discus</th>
<th>pole.vault</th>
<th>javelin</th>
<th>run1500</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schenk</td>
<td>-11.25</td>
<td>7.43</td>
<td>15.48</td>
<td>2.27</td>
<td>-48.90</td>
<td>-15.13</td>
<td>49.28</td>
<td>4.7</td>
<td>61.32</td>
<td>-268.95</td>
<td>8488</td>
</tr>
<tr>
<td>Voss</td>
<td>-10.87</td>
<td>7.45</td>
<td>14.97</td>
<td>1.97</td>
<td>-47.71</td>
<td>-14.46</td>
<td>44.36</td>
<td>5.1</td>
<td>61.76</td>
<td>-273.02</td>
<td>8399</td>
</tr>
<tr>
<td>Steen</td>
<td>-11.18</td>
<td>7.44</td>
<td>14.20</td>
<td>1.97</td>
<td>-48.29</td>
<td>-14.81</td>
<td>43.66</td>
<td>5.2</td>
<td>64.16</td>
<td>-263.20</td>
<td>8328</td>
</tr>
<tr>
<td>Thompson</td>
<td>-10.62</td>
<td>7.38</td>
<td>15.02</td>
<td>2.03</td>
<td>-49.06</td>
<td>-14.72</td>
<td>44.80</td>
<td>4.9</td>
<td>64.04</td>
<td>-285.11</td>
<td>8306</td>
</tr>
<tr>
<td>Blondel</td>
<td>-11.02</td>
<td>7.43</td>
<td>12.92</td>
<td>1.97</td>
<td>-47.44</td>
<td>-14.40</td>
<td>41.20</td>
<td>5.2</td>
<td>57.46</td>
<td>-256.64</td>
<td>8286</td>
</tr>
<tr>
<td>Plaziat</td>
<td>-10.83</td>
<td>7.72</td>
<td>13.58</td>
<td>2.12</td>
<td>-48.34</td>
<td>-14.18</td>
<td>43.06</td>
<td>4.9</td>
<td>52.18</td>
<td>-274.07</td>
<td>8272</td>
</tr>
<tr>
<td>Bright</td>
<td>-11.18</td>
<td>7.05</td>
<td>14.12</td>
<td>2.06</td>
<td>-49.34</td>
<td>-14.39</td>
<td>41.68</td>
<td>5.7</td>
<td>61.60</td>
<td>-291.20</td>
<td>8216</td>
</tr>
<tr>
<td>De.Wit</td>
<td>-11.05</td>
<td>6.95</td>
<td>15.34</td>
<td>2.00</td>
<td>-48.21</td>
<td>-14.36</td>
<td>41.32</td>
<td>4.8</td>
<td>63.00</td>
<td>-265.86</td>
<td>8189</td>
</tr>
<tr>
<td>Johnson</td>
<td>-11.15</td>
<td>7.12</td>
<td>14.52</td>
<td>2.03</td>
<td>-49.15</td>
<td>-14.66</td>
<td>42.36</td>
<td>4.9</td>
<td>66.46</td>
<td>-269.62</td>
<td>8180</td>
</tr>
</tbody>
</table>
# PCA on covariance matrix (default)
> pcaCOV <- princomp(x=decathlon[,1:10])
> names(pcaCOV)
[1] "sdev" "loadings" "center" "scale" "n.obs"
[6] "scores" "call"

# resign PCA solution
> pcsign <- sign(colSums(pcaCOV$loadings^3))
> pcaCOV$loadings <- pcaCOV$loadings %*% diag(pcsign)
> pcaCOV$scores <- pcaCOV$scores %*% diag(pcsign)

# Note: R uses MLE of covariance matrix
> n <- nrow(decathlon)
> sum((pcaCOV$sdev - sqrt(eigen((n-1)/n*cov(decathlon[,1:10]))$values))^2)
[1] 8.861398e-28
# PCA on correlation matrix

```r
> pcaCOR <- princomp(x=decathlon[,1:10], cor=TRUE)
> names(pcaCOR)
[1] "sdev"    "loadings" "center"  "scale"  "n.obs"
[6] "scores"  "call"

> pcsign <- sign(colSums(pcaCOR$loadings^3))
> pcaCOR$loadings <- pcaCOR$loadings %*% diag(pcsign)
> pcaCOR$scores <- pcaCOR$scores %*% diag(pcsign)

# Note: PC standard deviations are sqrts correlation matrix eigenvalues
> sum((pcaCOR$sdev - sqrt(eigen(cor(decathlon[,1:10]))$values))^2)
[1] 2.249486e-30
```
Plot Covariance and Correlation PCA Results

> dev.new(width=10, height=5, noRStudioGD=TRUE)
> par(mfrow=c(1,2))
> plot(pcaCOV$loadings[,1:2], xlab="PC1 Loadings", ylab="PC2 Loadings",
+ type="n", main="PCA of Covariance Matrix", xlim=c(-0.15, 1.15), ylim=c(-0.15, 1.15))
> text(pcaCOV$loadings[,1:2], labels=colnames(decathlon))
> plot(pcaCOR$loadings[,1:2], xlab="PC1 Loadings", ylab="PC2 Loadings",
+ type="n", main="PCA of Correlation Matrix", xlim=c(0.05, 0.45), ylim=c(-0.5,0.6))
> text(pcaCOR$loadings[,1:2], labels=colnames(decathlon))
Correlation of PC Scores and Overall Decathlon Score

> round(cor(decathlon$score, pcaCOV$scores),3)

[1,] 0.214 0.792 0.297 0.395 0.113 0.207 0.06 0.061 0.016 0.127

> round(cor(decathlon$score, pcaCOR$scores),3)

[1,] 0.991 0.017 0.079 0.064 -0.04 -0.025 0.013 -0.009 0.011 -0.002
In practice, the optimal number of components is often unknown.

In some cases, possible/feasible values may be known a priori from theory and/or past research.

In other cases, we need to use some data-driven approach to select a reasonable number of components.
Scree Plots

A scree plot displays the variance explained by each component.

We look for the “elbow” of the plot, i.e., point where line bends.

Could do formal test on derivative of scree line, but common sense approach often works fine.
Scree Plots For Decathlon PCA

Principal Components Analysis in Practice
Choosing the Number of Components

PCA of Covariance Matrix

PCA of Correlation Matrix

Nathaniel E. Helwig (U of Minnesota)