L23: Principal components

1. Predictors in regression

   (1) Two regressions

   For regression with intercept $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{q-1}x_{q-1} + \epsilon = B'\begin{pmatrix}1\\x\end{pmatrix} + \epsilon$ where $\mathbf{x} = \begin{pmatrix}x_1\\ \vdots \\ x_{q-1}\end{pmatrix}$,
   $E(y) = B'\begin{pmatrix}1\\x\end{pmatrix}$ is the regression function. For normal vector $\begin{pmatrix}y\\x\end{pmatrix} \sim N\left(\begin{pmatrix}\mu_x\\ \mu_y\end{pmatrix}, \begin{pmatrix}\Sigma_y & \Sigma_{yx}\\ \Sigma_{xy} & \Sigma_x\end{pmatrix}\right)$,
   $E(y|x) = \mu_y + \Sigma_{yx}\Sigma_x^{-1}(x-\mu_x)$ is also called the regression function of y on x.

   (2) Equal prediction functions
   With data $Y \in R^{n\times p}$ and $X_* = (1_n,\, X) \in R^{n\times q}$, $E(y)$ and $E(y|x)$ in two models are estimated by
   $$\widehat{y}(x) = \widehat{B}'\begin{pmatrix}1\\x\end{pmatrix} \text{ where } \widehat{B} = (X_*'X_*)^{-1}X_*'Y \text{ and}$$
   $$\widehat{E}(y|x) = \overline{y} + \left[Y'\left(I_n - \tfrac{1_n1_n'}{n}\right)X\right]\left[X'\left(I_n - \tfrac{1_n1_n'}{n}\right)X\right]^{-1}(x-\overline{x}).$$
   It can be shown that $\widehat{y}(x) = \widehat{E}(y|x)$.

   (3) Random predictors
   Thus one can treat the predictor vector $x \in R^{q-1}$ as random with $(\mu_x,\, \Sigma_x)$. By intuition
   The information provided by x_i is $\le$ that by $x_j \iff \mathrm{var}(x_i) \le \mathrm{var}(x_j)$
   The information provided by x_i and x_j are highly overlapped $\iff |\rho(x_i,\, x_j)|$ is close to 1.

2. Principal components

   (1) Definition
   For $\mathbf{x} \sim (\mu,\, \Sigma)$ with $\mu \in R^q$ and $\Sigma > 0$, $y_1, ..., y_q$ are ordered principal components of $\mathbf{x}$ if $y_i = l_i'\mathbf{x}$ where $l_i \in \mathcal{C}_i$ and $\mathrm{var}(l_i'x) \ge \mathrm{var}(l'x)$ for all $l \in \mathcal{C}_i$. Here
   $$\mathcal{C}_i = \{l \in R^q :\, \|l\| = 1 \text{ and } \mathrm{cov}(l'\mathbf{x},\, y_j) = 0 \text{ for all } j < i\}.$$
   Comments: $\mathcal{C}_1 \supset \mathcal{C}_2 \supset \cdots \supset \mathcal{C}_q$.
   $y_1, .., y_q$, if exist, may not be unique. But $\mathrm{var}(y_i)$, $i = 1, .., q$, are unique.
   $\mathrm{var}(y_1) \ge \mathrm{var}(y_2) \ge \cdots \ge \mathrm{var}(y_q) > 0$.

   (2) Existence
   If by EVD $\Sigma = P\Lambda P'$ where $\Lambda = \mathrm{diag}(\lambda_1, .., \lambda_q)$ with $\lambda_1 \ge \cdots \ge \lambda_q > 0$, then the components of $\mathbf{y} = P'\mathbf{x}$ are principal components of $\mathbf{x}$ and $\mathrm{var}(y_i) = \lambda_i$, $i = 1, ..., p$.

   Proof: With orthogonal $P = (P_1, .., P_q) \in R^{q\times q}$, we do induction on $i = 1, .., q$.
   When $i = 1$, $P_1 \in C_1 = \{l \in R^q :\, \|l\| = 1\}$ since $\|P_1\| = 1$.
   $$\mathrm{var}(P_1'x) = P_1'\Sigma P_1 = P_1'P\Lambda P'P_1 = e_1'\Lambda e_1 = \lambda_1.$$
   For $l \in C_1$, $P'l \in R^q$ and $\|P'l\| = 1$. Moreover,
   $$\mathrm{var}(l'x) = l'\Sigma l = l'P\Lambda P'l = \textstyle\sum_{i=1}^q (P'l)_i^2\lambda_i \le \lambda_1 = \mathrm{var}(P_1'x).$$
   So $y_1 = P_1'\mathbf{x}$ is the first PC for $\mathbf{x}$ and $\mathrm{var}(y_1) = \lambda_1$.
   Suppose $y_i = P_i'\mathbf{x}$, $i = 1, .., k < q$ are the first k PCs for $\mathbf{x}$ with $\mathrm{var}(y_i) = \lambda_i$, $i = 1, .., k$. Then for $i \le k$, $0 = \mathrm{cov}(l'\mathbf{x},\, y_i) = \mathrm{cov}(l'\mathbf{x},\, P_i'\mathbf{x}) = l'P\Lambda P'P_i = (P'l)_i\lambda_i \iff (P'l)_i = 0$. Thus
   $$C_{k+1} = \{l \in R^q :\, \|l\| = 1 \text{ and } (P'l)_i = 0 \text{ for all } i = 1, ..., k\}.$$
   Now $P_{k+1} \in \mathcal{C}_{k+1}$ since $P_{k+1} \in R^q$, $\|P_{k+1}\| = 1$, and $(P'P_{k+1})_i = (e_{k+1})_i = 0\ \forall\, i \le k$.
   Let $y_{k+1} = P_{k+1}'\mathbf{x}$. Then $\mathrm{var}(y_{k+1}) = e_{k+1}'\Lambda e_{k+1} = \lambda_{k+1}$. For $l \in \mathcal{C}_{k+1}$,
   $$\mathrm{var}(l'\mathbf{x}) = l'P\Lambda P'l = \textstyle\sum_{i=1}^q (P'l)_i^2\lambda_i = \sum_{i=k+1}^q (P'l)_i^2\lambda_i \le \lambda_{k+1} = \mathrm{var}(P_{k+1}'\mathbf{x}).$$
   Thus $y_{k+1} = P_{k+1}'\mathbf{x}$ is the $(k+1)$th PC of $\mathbf{x}$ with $\mathrm{var}(y_{k+1}) = \lambda_{k+1}$.

(3) Properties

From $\mathbf{x}$ to $\mathbf{y} = P'\mathbf{x}$, the total variance keeps unchanged.

Proof: $\mathrm{var}(x_1) + \cdots + \mathrm{var}(x_q) = \mathrm{tr}(\Sigma) = \mathrm{tr}(\Lambda) = \mathrm{var}(y_1) + \cdots + \mathrm{var}(y_q)$.

Comments: Among total variance in $\mathbf{x}$, $\sum \mathrm{var}(x_i) = \lambda_1 + \cdots + \lambda_q$, the amount explained by $y_1, .., y_k$ is $\sum_{i=1}^k \mathrm{var}(y_i) = \sum_{i=1}^k \lambda_i$. Thus the proportion explained by $y_1, .., y_i$ is $\frac{\lambda_1 + \cdots + \lambda_k}{\lambda_1 + \cdots + \lambda_q}$.

3. Principal components of standardized $\mathbf{x}$ and SAS

(1) Standardized $\mathbf{x}$

For $\mathbf{x} \sim (\mu, \Sigma)$, $V = \mathrm{diag}(\Sigma)$, $V^{1/2}$ and $\rho = V^{-1/2}\Sigma V^{-1/2}$ are the variance, the standard deviation and the correlation matrix. $\mathbf{z} = V^{-1/2}(\mathbf{x} - \mu) \sim (0, \rho)$ is the standardized $\mathbf{x}$.

(2) Principal components of $\mathbf{x}$ and $\mathbf{z}$.

Let $\Sigma = P\Lambda P'$ and $\rho = P_* \Lambda_* P_*'$ be EVDs with $\Lambda = \mathrm{diag}(\lambda_1, .., \lambda_q)$, $\Lambda_* = \mathrm{diag}(\lambda_{1*}, .., \lambda_{q*})$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q > 0$ and $\lambda_{1*} \geq \lambda_{2*} \geq \cdots \geq \lambda_{q*} > 0$. Then the components of $y = P'\mathbf{x}$ and $y_* = P_*'\mathbf{z}$ are the PCs of $\mathbf{x}$ and $\mathbf{z}$ respectively.

(3) SAS for principal component analysis

(i) Enter Σ into SAS

$$\Sigma = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

```
data a (type='cov');
    _TYPE_='COV';
    input _NAME_ $ x1 x2 x3;
    datalines;
    x1  1 -2 0
    x2 -2  5 0
    x3  0  0 2
    ;
```

```
x1 1 -2 0
x2 .  5 0
x3 .  . 2

x1 1  .  .
x2 -2 5  .
x3 0  0  2
```

(ii) principal components for $\mathbf{x}$

```
proc princomp cov;
     var x1 x2 x3;
     run;
```

Total Variables: 3

Total Variances: $\lambda_1 + \lambda_2 + \lambda_3$

EVs of Covariance Matrix

EVs	Difference	Proportion	Cumulative
λ_1	$\lambda_1 - \lambda_2$	$\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3}$	$\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3}$
λ_2	$\lambda_2 - \lambda_3$	$\frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_3}$	$\frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+\lambda_3}$
λ_3		$\frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3}$	$\frac{\lambda_1+\lambda_2+\lambda_3}{\lambda_1+\lambda_2+\lambda_3}$

Eigenvectors

	Prin1	Prin2	Prin3
x1	p_{11}	p_{12}	p_{13}
x2	p_{21}	p_{22}	p_{23}
x3	p_{31}	p_{32}	p_{33}

(iii) Principal components for standardized x

```
proc princomp;
     var x1 x2 x3;
     run;
```

Total Variables: 3

EVs of Correlation Matrix

EVs	Difference	Proportion	Cumulative
λ_1	$\lambda_1 - \lambda_2$	$\frac{\lambda_1}{3}$	$\frac{\lambda_1}{3}$
λ_2	$\lambda_2 - \lambda_3$	$\frac{\lambda_2}{3}$	$\frac{\lambda_1+\lambda_2}{3}$
λ_3		$\frac{\lambda_3}{3}$	$\frac{\lambda_1+\lambda_2+\lambda_3}{3}$

Eigenvectors

	Prin1	Prin2	Prin3
x1	p_{11}	p_{12}	p_{13}
x2	p_{21}	p_{22}	p_{23}
x3	p_{31}	p_{32}	p_{33}

L24: Sample principal components

1. Sample principal components

   (1) Population principal components
   For $\mathbf{x} \sim (\mu, \Sigma)$ with EVD $\Sigma = P\Lambda P'$, PC vector $\mathbf{y} = P'\mathbf{x} \sim (P'\mu, \Lambda)$ with $\text{var}(y_i) = \lambda_i$, $i = 1, .., q$.
   $$\sum_i \text{var}(x_i) = \text{tr}(\Sigma) = \text{tr}(\Lambda) = \sum_i \lambda_i = \sum_i \text{var}(y_i).$$
   Thus $\frac{\lambda_1}{\sum_i \lambda_i} = \frac{\text{var}(y_1)}{\sum_i \text{var}(x_i)}$ is the proportion of total variance in $\mathbf{x}$ explained by y_1, and $\frac{\lambda_1+\lambda_2}{\sum_i \lambda_i}$ is the proportion of total variance in $\mathbf{x}$ explained by y_1 and y_2. For $\mathbf{x} = V^{-1/2}(\mathbf{x} - \mu) \sim (0, \rho)$ with EVD $\rho = P_*\Lambda_*P_*'$, PC vector $\mathbf{y}_* = P_*'\mathbf{z} \sim (0, \Lambda_*)$ with $\text{var}(y_{i*}) = \lambda_{i*}$, $i = 1, ..., q$. So
   $$q = \sum_i \text{var}(z_i) = \text{tr}(\rho) = \text{tr}(\Lambda_*) = \sum_i \text{var}(y_{i*}).$$
   Thus $\frac{\lambda_{1*}+\lambda_{1*}}{\sum_i \lambda_{i*}} = \frac{\text{var}(y_{1*})+\text{var}(y_{2*})}{q}$ is the proportion of total variance in $\mathbf{z}$ explained by y_{1*} and y_{2*}.

   (2) Sample principal components
   In reality Σ and ρ may not be known. But with a sample from $\mathbf{x} \in R^q$, Σ has unbiased estimator $\mathbf{S}_u$ and ρ can be estimated by sample correlation matrix $\mathbf{R}$. By EVDs $S_u = \widehat{P}\widehat{\Lambda}\widehat{P}'$ and $\mathbf{R} = \widehat{P}_*\widehat{\Lambda}_*\widehat{P}_*'$ where $\widehat{\Lambda} = \text{diag}(\widehat{\lambda}_1, ..., \widehat{\lambda}_q)$ and $\widehat{\Lambda}_* = \text{diag}(\widehat{\lambda}_{1*}, ..., \widehat{\lambda}_{q*})$ with $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq \widehat{\lambda}_q > 0$ and $\widehat{\lambda}_{1*} \geq \widehat{\lambda}_{2*} \geq \cdots \geq \widehat{\lambda}_{q*} > 0$. Then the components of $\mathbf{y} = \widehat{P}'\mathbf{x}$ and $\mathbf{y}_* = \widehat{P}_*'\mathbf{z}$ are the sample principal components of $\mathbf{x}$ and $\mathbf{z}$ respectively.

   (3) SAS for sample principal components

   (i) Enter sample

```
data a;
     infile "D:\ex1.txt";
     input x1 x2 x3 x4 @@;
```

   (ii) Request sample PC for $\mathbf{x}$

```
proc princomp cov;
     var x1 x2 x3 x4;
     run;
```

   (3) Request sample PC for $\mathbf{z}$

```
proc princomp;
     var x1 x2 x3 x4;
     run;
```

2. Principal component scores

   (1) Principal component scores for $\mathbf{x}$
   Based on sample $\mathbf{x}_i \in R^q$, $i = 1, ..., n$, on $\mathbf{x} \sim (\mu, \Sigma)$, PC vector $\widehat{P}'\mathbf{x}$ is obtained. Then $\widehat{P}'\mathbf{x}_i$, $i = 1, ..., n$, are PC scores for $\mathbf{x}$. It contains n values for each of q sample PCts.

```
data a;
     infile "D:\ex2.txt";
     input x1 x2 x3 @@;
proc princomp cov out=b;
     var x1 x2 x3;
     rn;
proc print;
     run;
```

displays x1, x2, x3, prin1, prin2, prin3.

(2) Principal components scores for $\mathbf{z}$
Based on sample $\mathbf{x}_i \in R^q$, $i = 1, ..., n$, on $\mathbf{x} \sim (\mu, \Sigma)$, PC vector $\widehat{P}_*'\mathbf{z}$ is obtained. Then $\widehat{P}_*'\widehat{\mathbf{z}}_i = \widehat{V}^{-1/2}(\mathbf{x}_i - \overline{x})$, $i = 1, ..., n$, are PC scores for $\mathbf{z}$. It contains n values for each of q sample PCts.

```
data a;
     infile "D:\ex2.txt";
     input x1 x2 x3 @@;
proc princomp out=b;
     var x1 x2 x3;
     rn;
proc print;
     run;
```

displays x1, x2, x3, prin1, prin2, prin3.

3. Make use of principal components

(1) Factor model
Study on principal components leads to factor model.

(2) Select predictors
When selecting k predictors in a pool of q predictors for regression, the first k principal components is one of the choice.

```
data a;
     infile "D:\ex3.txt";
     input y1 y2 x1 x2 x3 @@;
proc princomp out=b;
     var x1 x2 x3;
     rn;
proc reg;
     model y1 y2=prin1 prin2;
     run;
```

Caution: In univariate regression we have many ways to select predictors, largest R^2-criterion for example. All those methods look at the relations of response and predictors. But using principal components is only based on the behavior of $\mathbf{x}$. The response did not participate in the selection process.