RMSC4002 Data Analysis in Finance and Risk Management Science

Tutorial 1 Supplement

Chan Chun Ho, Benjamin[*,†,‡,§]

September 11, 2018

1 Matrix Algebra and Random Vectors

1.1 Definitions and Properties

- Suppose $\boldsymbol{Y}' = [Y_1, Y_2, \ldots, Y_p]$ is a $p \times 1$ random vector. The $p \times 1$ mean vector is

$$E(\boldsymbol{Y}) = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \underset{(p\times 1)}{\boldsymbol{\mu}}.$$

- The variance-covariance matrix (or simply covariance matrix) is

$$\begin{aligned}
\underset{(p\times p)}{\boldsymbol{\Sigma}} &= \mathrm{Cov}(\boldsymbol{Y}) = E(\boldsymbol{Y} - \boldsymbol{\mu})(\boldsymbol{Y} - \boldsymbol{\mu})' \\
&= E \begin{bmatrix} Y_1 - \mu_1 \\ Y_2 - \mu_2 \\ \vdots \\ Y_p - \mu_p \end{bmatrix} \begin{bmatrix} Y_1 - \mu_1 & Y_2 - \mu_2 & \cdots & Y_p - \mu_p \end{bmatrix} \\
&= \begin{bmatrix} E(Y_1 - \mu_1)^2 & E(Y_1 - \mu_1)(Y_2 - \mu_2) & \cdots & E(Y_1 - \mu_1)(Y_p - \mu_p) \\ E(Y_2 - \mu_2)(Y_1 - \mu_1) & E(Y_2 - \mu_2)^2 & \cdots & E(Y_2 - \mu_2)(Y_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ E(Y_p - \mu_p)(Y_1 - \mu_1) & E(Y_p - \mu_p)(Y_2 - \mu_2) & \cdots & E(Y_p - \mu_p)^2 \end{bmatrix} \\
&= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix},
\end{aligned}$$

  where $\sigma_{ij} = \mathrm{Cov}(Y_i, Y_j) = E(Y_i - \mu_i)(Y_j - \mu_j)$ for $i, j = 1, \ldots, p$. In fact, $\sigma_{ii} = \mathrm{Var}(Y_i)$ for $i = 1, \ldots, p$.

[*] For enquiry, please email to 1155049861@link.cuhk.edu.hk.
[†] Personal profile: www.linkedin.com/in/benjamin-chan-chun-ho
[‡] GitHub repository: https://github.com/BenjaminChanChunHo/CUHK-STAT-or-RMSC-Tutorial-Note
[§] RPubs: http://rpubs.com/Benjamin_Chan_Chun_Ho

1.2 Random Samples, Sample Mean and Covariance Matrix

- The $n \times p$ data matrix is[1]

$$\underset{(n\times p)}{\boldsymbol{X}} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_n' \end{bmatrix},$$

  where the row vector $\boldsymbol{x}_j'$ represents the jth observation, i.e.

$$\boldsymbol{x}_j = \begin{bmatrix} x_{j1} \\ x_{j2} \\ \vdots \\ x_{jp} \end{bmatrix}.$$

- Let the (j,k)-th entry in the data matrix be the random variable X_{jk}. Each set of measurements $\boldsymbol{X}_j$ on p variables is a random vector, and we have the random matrix

$$\underset{(n\times p)}{\boldsymbol{X}} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1' \\ \boldsymbol{X}_2' \\ \vdots \\ \boldsymbol{X}_n' \end{bmatrix}. \tag{1}$$

- If the row vectors $\boldsymbol{X}_1', \boldsymbol{X}_2', \ldots, \boldsymbol{X}_n'$ in Equation 1 represent independent observations from a common joint distribution with density function $f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_p)$, then $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ are said to form a random sample from $f(\boldsymbol{x})$.

Definition 1.1. (Sample Covariance Matrix). The symmetric matrix of sample variances and covariances is called the sample covariance matrix:[2]

$$\boldsymbol{S} = (s_{ij}) = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix}.$$

The jth diagonal element s_{jj} is the sample variance of the jth variable:

$$s_{jj} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_{ij}^2 - n\bar{x}_j^2\right).$$

The off-diagonal element s_{ij} is the sample covariance of the ith and jth variables:

$$s_{jk} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_{ij}x_{ik} - n\bar{x}_j\bar{x}_k\right).$$

---

[1] "Applied Multivariate Statistical Analysis" 6th ed. (Johnson and Wichern) Ch3 p.111-123
[2] "Multivariate Statistical Inference and Applications" (Rencher) Ch1 p.8

1.3 Sample Mean, Covariance and Correlation as Matrix Operations

- The descriptive statistics $\bar{\boldsymbol{x}}$ and $\boldsymbol{S}$ can be calculated by the data matrix $\boldsymbol{X}$ using matrix operations. The calculation can then be easily programmed on computers.[3]

- First of all,

$$\bar{\boldsymbol{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \frac{1}{n} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \frac{1}{n}\boldsymbol{X}'\mathbf{1}_n,$$

where $\mathbf{1}_n$ is an $n \times 1$ vector whose n elements are all 1.

- Next, an $n \times p$ matrix of means can be created as follows:

$$\underset{(n\times 1)}{\mathbf{1}_n}\underset{(1\times p)}{\bar{\boldsymbol{x}}'} = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\boldsymbol{X} = \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_p \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_p \end{bmatrix}_{(n\times p)}$$

- Then, the $n \times p$ matrix of deviations (residuals) can be obtained as follows:

$$\boldsymbol{X}_c = \boldsymbol{X} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\boldsymbol{X} = \begin{bmatrix} x_{11}-\bar{x}_1 & x_{12}-\bar{x}_2 & \cdots & x_{1p}-\bar{x}_p \\ x_{21}-\bar{x}_1 & x_{22}-\bar{x}_2 & \cdots & x_{2p}-\bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1}-\bar{x}_1 & x_{n2}-\bar{x}_2 & \cdots & x_{np}-\bar{x}_p \end{bmatrix}_{(n\times p)}$$

- In fact, $\boldsymbol{X}_c = (\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n')\boldsymbol{X}$ is the centered form of the data matrix $\boldsymbol{X}$. From above, the centering matrix $\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$ centers $\boldsymbol{X}$.[4]

- Note that $(n-1)\boldsymbol{S} = \boldsymbol{A} = \sum_{i=1}^{n}(\boldsymbol{X_i} - \bar{\boldsymbol{X}})(\boldsymbol{X_i} - \bar{\boldsymbol{X}})'$. So we have

$$(n-1)\underset{(p\times p)}{\boldsymbol{S}} = \begin{bmatrix} x_{11}-\bar{x}_1 & x_{21}-\bar{x}_1 & \cdots & x_{n1}-\bar{x}_1 \\ x_{12}-\bar{x}_2 & x_{22}-\bar{x}_2 & \cdots & x_{n2}-\bar{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p}-\bar{x}_p & x_{2p}-\bar{x}_p & \cdots & x_{np}-\bar{x}_p \end{bmatrix}_{(p\times n)} \times \begin{bmatrix} x_{11}-\bar{x}_1 & x_{12}-\bar{x}_2 & \cdots & x_{1p}-\bar{x}_p \\ x_{21}-\bar{x}_1 & x_{22}-\bar{x}_2 & \cdots & x_{2p}-\bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1}-\bar{x}_1 & x_{n2}-\bar{x}_2 & \cdots & x_{np}-\bar{x}_p \end{bmatrix}_{(n\times p)}$$

$$= \left(\boldsymbol{X} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\boldsymbol{X}\right)'\left(\boldsymbol{X} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\boldsymbol{X}\right) = \boldsymbol{X}'\left(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\right)'\left(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\right)\boldsymbol{X}$$

$$= \boldsymbol{X}_c'\boldsymbol{X}_c = \boldsymbol{X}'\left(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\right)\boldsymbol{X}.$$

- In summary, matrix operations on the data matrix $\boldsymbol{X}$ lead to $\bar{\boldsymbol{x}}$ and $\boldsymbol{S}$.

$$\bar{\boldsymbol{x}} = \frac{1}{n}\boldsymbol{X}'\mathbf{1}_n \tag{2}$$

$$\boldsymbol{S} = \frac{1}{n-1}\boldsymbol{X}'\left(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\right)\boldsymbol{X} = \frac{1}{n-1}\boldsymbol{X}_c'\boldsymbol{X}_c \tag{3}$$

[3] "Applied Multivariate Statistical Analysis" 6th ed. (Johnson and Wichern) Ch3 p.137-140
[4] "Multivariate Statistical Inference and Applications" (Rencher) Ch1 p.9

Definition 1.2. The sample correlation between the jth and kth variables is given by[5]

$$r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}s_{kk}}}.$$

The sample correlation matrix is defined as

$$\boldsymbol{R} = (r_{jk}) = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix},$$

which is symmetric, since $r_{jk} = r_{kj}$.

Definition 1.3. The population correlation matrix is defined as

$$\boldsymbol{P}_\rho = (\rho_{jk}) = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix},$$

where $\rho_{jk} = \sigma_{jk}/\sqrt{\sigma_{jj}\sigma_{kk}}$.

- The sample covariance matrix $\boldsymbol{S}$ and the sample correlation matrix $\boldsymbol{R}$ are related to each other through matrix operations. Define a $p \times p$ diagonal matrix $\boldsymbol{D} = \mathrm{diag}(s_{11}, \ldots, s_{pp})$, where $\boldsymbol{S} = (s_{ij})$. Then,

$$\underset{(p\times p)}{\boldsymbol{D}^{1/2}} = \begin{bmatrix} \sqrt{s_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{s_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{s_{pp}} \end{bmatrix} \quad \text{and} \quad \underset{(p\times p)}{\boldsymbol{D}^{-1/2}} = \begin{bmatrix} \frac{1}{\sqrt{s_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{s_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{s_{pp}}} \end{bmatrix}.$$

  Thus, we have

$$\boldsymbol{R} = \boldsymbol{D}^{-1/2}\boldsymbol{S}\boldsymbol{D}^{-1/2}$$

  and

$$\boldsymbol{S} = \boldsymbol{D}^{1/2}\boldsymbol{R}\boldsymbol{D}^{1/2}.$$

- The population covariance matrix $\boldsymbol{\Sigma}$ and the population correlation matrix $\boldsymbol{P}_\rho$ are related to each other through matrix operations. Define a $p \times p$ diagonal matrix $\boldsymbol{D}_\sigma = \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{pp})$, where $\boldsymbol{\Sigma} = (\sigma_{ij})$. Then

$$\boldsymbol{P}_\rho = \boldsymbol{D}_\sigma^{-1/2}\boldsymbol{\Sigma}\boldsymbol{D}_\sigma^{-1/2}$$

  and

$$\boldsymbol{\Sigma} = \boldsymbol{D}_\sigma^{1/2}\boldsymbol{P}_\rho\boldsymbol{D}_\sigma^{1/2}.$$

[5] "Multivariate Statistical Inference and Applications" (Rencher) Ch1 p.11-12

1.4 Eigenvalues and Eigenvectors

Definition 1.4. (Eigenvalue and Eigenvector). Let $\boldsymbol{A}$ be a $p \times p$ matrix and let λ be an eigenvalue of $\boldsymbol{A}$. If $\boldsymbol{x}$ is a $p \times 1$ nonzero vector ($\boldsymbol{x} \neq \boldsymbol{0}$) such that

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$$

then $\boldsymbol{x}$ is said to be an **eigenvector** of the matrix $\boldsymbol{A}$ associated with the **eigenvalue** λ. An equivalent condition for λ to be a solution of the eigenvalue-eigenvector pair is $|\boldsymbol{A} - \lambda \boldsymbol{I}_p| = 0$.[6]

Definition 1.5. A **symmetric matrix** has $\boldsymbol{S}' = \boldsymbol{S}$. It means that $s_{ji} = s_{ij}$.

Theorem 1.1. (Real Eigenvalues and Eigenvectors). All the eigenvalues and eigenvectors of a real symmetric matrix are real.[7]

Proof. Suppose that $\boldsymbol{S}\boldsymbol{x} = \lambda\boldsymbol{x}$ and λ is a complex number $a + ib$, where $i = \sqrt{-1}$ and a and b are real. Its complex conjugate is $\bar{\lambda} = a - ib$. Similarly suppose that the components of $\boldsymbol{x}$ are complex, and switching the signs of their imaginary parts gives $\bar{\boldsymbol{x}}$. Since $\boldsymbol{S}$ is real symmetric, taking conjugates of

$$\boldsymbol{S}\boldsymbol{x} = \lambda\boldsymbol{x}$$

leads to $\boldsymbol{S}\bar{\boldsymbol{x}} = \bar{\lambda}\bar{\boldsymbol{x}}$. Transpose to give

$$\bar{\boldsymbol{x}}'\boldsymbol{S} = \bar{\boldsymbol{x}}'\bar{\lambda}.$$

Take the dot product of the first equation with $\bar{\boldsymbol{x}}$ and the last equation with $\boldsymbol{x}$:

$$\bar{\boldsymbol{x}}'\boldsymbol{S}\boldsymbol{x} = \bar{\boldsymbol{x}}'\lambda\boldsymbol{x} \quad \text{and} \quad \bar{\boldsymbol{x}}'\boldsymbol{S}\boldsymbol{x} = \bar{\boldsymbol{x}}'\bar{\lambda}\boldsymbol{x}.$$

The left sides are the same so the right sides are equal. One equation has λ, the other has $\bar{\lambda}$. They multiply $\bar{\boldsymbol{x}}'\boldsymbol{x} = |x_1|^2 + |x_2|^2 + \cdots$, which is length squared and not zero. Therefore λ must equal $\bar{\lambda}$, and $a + ib$ equals $a - ib$. So $b = 0$ and $\lambda = a$, which is real.

The eigenvectors come from solving the real equation $(\boldsymbol{S} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$. So the $\boldsymbol{x}$'s are real. □

Theorem 1.2. (Orthogonal Eigenvectors). Eigenvectors of a real symmetric matrix (when they correspond to distinct eigenvalues) are always perpendicular.[8]

Proof. Suppose $\boldsymbol{S}\boldsymbol{x} = \lambda_1\boldsymbol{x}$ and $\boldsymbol{S}\boldsymbol{y} = \lambda_2\boldsymbol{y}$. Assume that $\lambda_1 \neq \lambda_2$. Use $\boldsymbol{S}' = \boldsymbol{S}$, take dot product of the first equation with $\boldsymbol{y}$:

$$(\boldsymbol{S}\boldsymbol{x})'\boldsymbol{y} = (\lambda_1\boldsymbol{x})'\boldsymbol{y} \quad \Rightarrow \quad \boldsymbol{x}'\boldsymbol{S}\boldsymbol{y} = \lambda_1\boldsymbol{x}'\boldsymbol{y}.$$

Take dot product of the second equation with $\boldsymbol{x}$:

$$\boldsymbol{x}'(\boldsymbol{S}\boldsymbol{y}) = \boldsymbol{x}'(\lambda_2\boldsymbol{y}) \quad \Rightarrow \quad \boldsymbol{x}'\boldsymbol{S}\boldsymbol{y} = \lambda_2\boldsymbol{x}'\boldsymbol{y}.$$

The left sides are the same so the right sides are equal, i.e. $\lambda_1\boldsymbol{x}'\boldsymbol{y} = \lambda_2\boldsymbol{x}'\boldsymbol{y}$. Since $\lambda_1 \neq \lambda_2$, it means that $\boldsymbol{x}'\boldsymbol{y} = 0$. The eigenvector $\boldsymbol{x}$ (for λ_1) is perpendicular to the eigenvector $\boldsymbol{y}$ (for λ_2). □

[6] "Applied Multivariate Statistical Analysis" 6th ed. (Johnson and Wichern) Ch2 p.98
[7] "Introduction to Linear Algebra" 5th ed. (Strang) Ch6 p.339-340
[8] "Introduction to Linear Algebra" 5th ed. (Strang) Ch6 p.340

1.5 Positive Definite Matrices

Definition 1.6. (Quadratic Form). A **quadratic form** $Q(\boldsymbol{x})$ in the p variables $x_1, x_2, \ldots, x_p$ is $Q(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$, which has only squared terms x_i^2 and product terms $x_i x_k$, and where $\boldsymbol{x}' = [x_1, x_2, \ldots, x_p]$ and $\boldsymbol{A}$ is a $p \times p$ symmetric matrix.

Definition 1.7. (Positive Definite Matrix). A $p \times p$ symmetric matrix $\boldsymbol{A}$ is called **positive definite** if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$ for all vectors $\boldsymbol{x} \neq \mathbf{0}$, denoted by $\boldsymbol{A} \succ 0$. It is called **positive semidefinite** if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \neq \mathbf{0}$, denoted by $\boldsymbol{A} \succeq 0$. It is called **non-negative definite** if $\boldsymbol{A} \succ 0$ or $\boldsymbol{A} \succeq 0$, i.e., if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$ for all $\boldsymbol{x}$.[9]

Theorem 1.3. (Positive Eigenvalues). A symmetric matrix $\boldsymbol{A}$ is positive definite if and only if its eigenvalues are positive.[10]

Proof. Let λ and $\boldsymbol{x}$ be an eigenvalue and corresponding eigenvector of $\boldsymbol{A}$. If $\boldsymbol{A}$ is positive definite, then $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}'\boldsymbol{x}$. Since $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$, then $\lambda > 0$. Vice versa, if the eigenvalues of $\boldsymbol{A}$ are positive, then by the spectral decomposition, $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'\boldsymbol{x}$. Since $\boldsymbol{\Lambda}$ is a diagonal matrix with all the diagonal elements being positive, then $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{\Lambda}^{1/2}\boldsymbol{P}'\boldsymbol{x})'(\boldsymbol{\Lambda}^{1/2}\boldsymbol{P}'\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \neq \mathbf{0}$. ☐

Lemma 1.4. For any matrix $\boldsymbol{B}$, $\boldsymbol{B}'\boldsymbol{B}$ is non-negative definite.

Proof. Let $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{x}$. Assume that $\boldsymbol{y}$ is a $p \times 1$ vector. $\boldsymbol{x}'\boldsymbol{B}'\boldsymbol{B}\boldsymbol{x} = \boldsymbol{y}'\boldsymbol{y} = \sum_{i=1}^{p} y_i^2 \geq 0$. ☐

Theorem 1.5. The sample covariance matrix $\boldsymbol{S}$ is non-negative definite.

Proof. It suffices to show that the sums of squares and cross products matrix $\boldsymbol{A}$ is non-negative definite since $\boldsymbol{S} = \frac{1}{n-1}\boldsymbol{A}$. Note that we can write

$$\boldsymbol{A} = \sum_{i=1}^{n} (\boldsymbol{X_i} - \bar{\boldsymbol{X}})(\boldsymbol{X_i} - \bar{\boldsymbol{X}})' = \left(\boldsymbol{X} - \frac{1}{n}\mathbf{1_n}\mathbf{1_n'}\boldsymbol{X}\right)'\left(\boldsymbol{X} - \frac{1}{n}\mathbf{1_n}\mathbf{1_n'}\boldsymbol{X}\right).$$

By Lemma 1.4, $\boldsymbol{A}$ is non-negative definite and hence $\boldsymbol{S}$ is non-negative definite. ☐

Theorem 1.6. (Inverse of Positive Definite Matrix). If $\boldsymbol{\Sigma}$ is positive definite, so that $\boldsymbol{\Sigma}^{-1}$ exists, then[11]

$$\boldsymbol{\Sigma}\boldsymbol{e} = \lambda\boldsymbol{e} \ \text{ implies } \ \boldsymbol{\Sigma^{-1}}\boldsymbol{e} = \left(\frac{1}{\lambda}\right)\boldsymbol{e}$$

so $(\lambda, \boldsymbol{e})$ is an eigenvalue-eigenvector pair for $\boldsymbol{\Sigma}$ corresponding to the pair $(1/\lambda, \boldsymbol{e})$ for $\boldsymbol{\Sigma}^{-1}$. Also, $\boldsymbol{\Sigma}^{-1}$ is positive definite.

Proof. For positive definite $\boldsymbol{\Sigma}$ and eigenvector $\boldsymbol{e} \neq \mathbf{0}$, we have $0 < \boldsymbol{e}'\boldsymbol{\Sigma}\boldsymbol{e} = \boldsymbol{e}'(\lambda\boldsymbol{e}) = \lambda\boldsymbol{e}'\boldsymbol{e} = \lambda$. Moreover, $\boldsymbol{e} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}\boldsymbol{e}) = \boldsymbol{\Sigma}^{-1}(\lambda\boldsymbol{e})$, or $\boldsymbol{e} = \lambda\boldsymbol{\Sigma}^{-1}\boldsymbol{e}$, and division by $\lambda > 0$ gives $\boldsymbol{\Sigma}^{-1}\boldsymbol{e} = (1/\lambda)\boldsymbol{e}$. Thus, $(1/\lambda, \boldsymbol{e})$ is an eigenvalue-eigenvector pair for $\boldsymbol{\Sigma}^{-1}$. For any $p \times 1$ vector $\boldsymbol{x}$,

$$\boldsymbol{x}'\boldsymbol{\Sigma^{-1}}\boldsymbol{x} = \boldsymbol{x}'\left(\sum_{i=1}^{p}\left(\frac{1}{\lambda_i}\right)\boldsymbol{e}_i\boldsymbol{e}_i'\right)\boldsymbol{x} = \sum_{i=1}^{p}\left(\frac{1}{\lambda_i}\right)(\boldsymbol{x}'\boldsymbol{e}_i)^2 \geq 0$$

since each term $\lambda_i^{-1}(\boldsymbol{x}'\boldsymbol{e}_i)^2$ is nonnegative. In addition, $\boldsymbol{x}'\boldsymbol{e}_i = 0$ for all i only if $\boldsymbol{x} = \mathbf{0}$. So $\boldsymbol{x} \neq \mathbf{0}$ implies that $\sum_{i=1}^{p}(1/\lambda_i)(\boldsymbol{x}'\boldsymbol{e}_i)^2 > 0$, and it follows that $\boldsymbol{\Sigma}^{-1}$ is positive definite. ☐

[9] "Aspects of Multivariate Statistical Theory" (Muirhead) Appendix p.585
[10] "Matrix Algebra Useful for Statistics" 2nd ed. (Searle and Khuri) Ch6 p.136
[11] "Applied Multivariate Statistical Analysis" 6th ed. (Johnson and Wichern) Ch4 p.153

1.6 Orthogonality

- Two vectors are orthogonal when their dot product is zero: $\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{v}'\boldsymbol{w} = 0$.

Definition 1.8. A set of vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ is called **orthogonal** if $\boldsymbol{v}_i'\boldsymbol{v}_j = 0$, for all $i \neq j$. If, in addition, $\boldsymbol{v}_i'\boldsymbol{v}_i = 1$, for all $i = 1, 2, \ldots, n$, then the set is called **orthonormal**.[12]

Theorem 1.7. A matrix $\boldsymbol{V}$ with orthonormal columns $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ satisfies $\boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}$.[13]

Proof.

$$\boldsymbol{V}'\boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_1' \\ \boldsymbol{v}_2' \\ \vdots \\ \boldsymbol{v}_n' \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix} = \boldsymbol{I}.$$

When row i of $\boldsymbol{V}'$ multiplies column j of $\boldsymbol{V}$, the dot product is $\boldsymbol{v}_i'\boldsymbol{v}_j$. Off the diagonal $(i \neq j)$ that dot product is zero by orthogonality. On the diagonal $(i = j)$ the unit vectors give $\boldsymbol{v}_i'\boldsymbol{v}_i = \|\boldsymbol{v}_i\|^2 = 1$. □

- If the columns are only orthogonal (not unit vectors), dot products give a diagonal matrix (not the identity matrix).
- To have $\boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}$, the columns $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ are orthonormal. However, $\boldsymbol{V}$ is not required to be square. When $\boldsymbol{V}$ is rectangular, $\boldsymbol{V}'$ is only an inverse from the left.

Definition 1.9. (Orthogonal Matrix). A square matrix $\boldsymbol{Q}$ is said to be **orthogonal** if its columns $\{\boldsymbol{q}_1', \boldsymbol{q}_2', \ldots, \boldsymbol{q}_n'\}$ form an orthonormal set in $\mathbb{R}^n$.

Theorem 1.8. A square matrix $\boldsymbol{Q}$ is an orthogonal matrix if and only if $\boldsymbol{Q}$ is invertible with $\boldsymbol{Q}^{-1} = \boldsymbol{Q}'$. Then $\boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}' = \boldsymbol{I}$.

- When $\boldsymbol{Q}$ is orthogonal, $\boldsymbol{Q}'$ is the two-sided inverse. The rows of a square $\boldsymbol{Q}$ are orthonormal like the columns.

[12] "Numerical Analysis" 9th ed. (Burden and Faires) Ch9 p.566-570
[13] "Introduction to Linear Algebra" 5th ed. (Strang) Ch4 p.233-238

1.7 Spectral Decomposition

- The spectral decomposition of a $p \times p$ symmetric matrix $\boldsymbol{A}$ is given by[14]

$$\underset{(p\times p)}{\boldsymbol{A}} = \lambda_1 \underset{(p\times 1)}{\boldsymbol{e}_1} \underset{(1\times p)}{\boldsymbol{e}_1'} + \lambda_2 \underset{(p\times 1)}{\boldsymbol{e}_2} \underset{(1\times p)}{\boldsymbol{e}_2'} + \cdots + \lambda_p \underset{(p\times 1)}{\boldsymbol{e}_p} \underset{(1\times p)}{\boldsymbol{e}_p'} = \sum_{i=1}^{p} \lambda_i \boldsymbol{e}_i \boldsymbol{e}_i', \tag{4}$$

  where $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_p$ are the associated normalized eigenvectors. Thus, $\boldsymbol{e}_i'\boldsymbol{e}_i = 1$ for $i = 1, 2, \ldots, p$, and $\boldsymbol{e}_i'\boldsymbol{e}_j = 0$ for $i \neq j$.

- Let the spectral decomposition of $\boldsymbol{A}$ be defined in Equation 4. Let the normalized eigenvectors be the columns of a $p \times p$ matrix $\boldsymbol{P} = [\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_p]$. Then,

$$\underset{(p\times p)}{\boldsymbol{A}} = \underset{(p\times p)}{\boldsymbol{P}} \underset{(p\times p)}{\boldsymbol{\Lambda}} \underset{(p\times p)}{\boldsymbol{P}'}, \tag{5}$$

  where $\boldsymbol{\Lambda}$ is the diagonal matrix

$$\underset{(p\times p)}{\boldsymbol{\Lambda}} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$$

  and $\boldsymbol{P}$ is orthogonal. Hence, $\boldsymbol{P}^{-1} = \boldsymbol{P}'$ such that $\boldsymbol{P}\boldsymbol{P}' = \boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}_p$.

- Let $\boldsymbol{A}$ be a $p \times p$ positive definite matrix with $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'$ defined as above.

  The inverse of $\boldsymbol{A}$ is given by

$$\boldsymbol{A}^{-1} = \boldsymbol{P}\boldsymbol{\Lambda}^{-1}\boldsymbol{P}' = \sum_{i=1}^{p} \frac{1}{\lambda_i} \boldsymbol{e}_i \boldsymbol{e}_i' \quad \text{with } \lambda_i > 0 \tag{6}$$

  since $(\boldsymbol{P}\boldsymbol{\Lambda}^{-1}\boldsymbol{P}')\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}' = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'(\boldsymbol{P}\boldsymbol{\Lambda}^{-1}\boldsymbol{P}') = \boldsymbol{P}\boldsymbol{P}' = \boldsymbol{I}_p$.

- Let $\boldsymbol{A}$ be a $p \times p$ non-negative definite matrix with $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'$ defined as above.

  The (symmetric) square-root of $\boldsymbol{A}$ is given by

$$\boldsymbol{A}^{1/2} = \boldsymbol{P}\boldsymbol{\Lambda}^{1/2}\boldsymbol{P}' = \sum_{i=1}^{p} \sqrt{\lambda_i} \boldsymbol{e}_i \boldsymbol{e}_i' \quad \text{with } \lambda_i \geq 0 \tag{7}$$

  since $\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2} = \boldsymbol{\Lambda}$ and $(\boldsymbol{P}\boldsymbol{\Lambda}^{1/2}\boldsymbol{P}')\boldsymbol{P}\boldsymbol{\Lambda}^{1/2}\boldsymbol{P}' = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}' = \boldsymbol{A}$. Note $(\boldsymbol{A}^{1/2})' = \boldsymbol{A}^{1/2}$ so $\boldsymbol{A}^{1/2}$ is symmetric.

  As $\boldsymbol{A}$ is non-negative definite, then $\boldsymbol{A}^{1/2}$ is non-negative definite and is called a non-negative definite square-root of $\boldsymbol{A}$. If $\boldsymbol{A}$ is positive definite, $\boldsymbol{A}^{1/2}$ is positive definite and is called the positive definite square-root of $\boldsymbol{A}$.[15]

[14] "Applied Multivariate Statistical Analysis" 6th ed. (Johnson and Wichern) Ch2 p.60-66
[15] "Aspects of Multivariate Statistical Theory" (Muirhead) Appendix p.588

1.8 Trace and Determinant

Lemma 1.9. (Trace is invariant under cyclic permutations). Let $\boldsymbol{B}$ and $\boldsymbol{C}$ be $m \times k$ and $k \times m$ matrices, respectively. Then

$$\text{tr}(\boldsymbol{BC}) = \text{tr}(\boldsymbol{CB}),$$

where the trace of a matrix is the sum of its diagonal elements.

Proof. $\boldsymbol{BC}$ has $\sum_{j=1}^{k} b_{ij}c_{ji}$ as its ith diagonal element, so $\text{tr}(\boldsymbol{BC}) = \sum_{i=1}^{m}(\sum_{j=1}^{k} b_{ij}c_{ji})$. Similarly, the jth diagonal element of $\boldsymbol{CB}$ is $\sum_{i=1}^{m} c_{ji}b_{ij}$, so $\text{tr}(\boldsymbol{CB}) = \sum_{j=1}^{k}(\sum_{i=1}^{m} c_{ji}b_{ij}) = \sum_{i=1}^{m}(\sum_{j=1}^{k} b_{ij}c_{ji}) = \text{tr}(\boldsymbol{BC})$. $\square$

Lemma 1.10. (Trace and Eigenvalues). Let $\boldsymbol{D}$ be a $p \times p$ symmetric matrix. Then, the trace of $\boldsymbol{D}$ is

$$\text{tr}(\boldsymbol{D}) = \sum_{i=1}^{p} \lambda_i,$$

where the λ_i are the eigenvalues of $\boldsymbol{D}$.

Proof. Using the spectral decomposition in Equation 5, $\boldsymbol{D} = \boldsymbol{P\Lambda P'}$, where $\boldsymbol{P'P} = \boldsymbol{I_p}$ and $\boldsymbol{\Lambda}$ is a diagonal matrix with entries $\lambda_1, \lambda_2, \ldots, \lambda_p$. Therefore, $\text{tr}(\boldsymbol{D}) = \text{tr}(\boldsymbol{P\Lambda P'}) = \text{tr}(\boldsymbol{\Lambda P'P}) = \text{tr}(\boldsymbol{\Lambda}) = \lambda_1 + \lambda_2 + \cdots + \lambda_p$, using Lemma 1.9. $\square$

Lemma 1.11. (Determinant and Eigenvalues). Let $\boldsymbol{D}$ be a $p \times p$ symmetric matrix. Then, the determinant of $\boldsymbol{D}$ is

$$|\boldsymbol{D}| = \prod_{i=1}^{p} \lambda_i,$$

where the λ_i are the eigenvalues of $\boldsymbol{D}$.

Proof. Using the spectral decomposition in Equation 5, $\boldsymbol{D} = \boldsymbol{P\Lambda P'}$, where $\boldsymbol{P'P} = \boldsymbol{I_p}$ and $\boldsymbol{\Lambda}$ is a diagonal matrix with entries $\lambda_1, \lambda_2, \ldots, \lambda_p$. Therefore, $|\boldsymbol{D}| = |\boldsymbol{P\Lambda P'}| = |\boldsymbol{P}||\boldsymbol{\Lambda}||\boldsymbol{P'}| = |\boldsymbol{\Lambda}||\boldsymbol{P'P}| = |\boldsymbol{\Lambda}||\boldsymbol{I_p}| = |\boldsymbol{\Lambda}| = \lambda_1\lambda_2\cdots\lambda_p$. $\square$

---

2 Reference Books

1. Rencher, A.C. & Christensen, W.F. (2012). *Methods of Multivariate Analysis* (3rd ed.).
2. Rencher, A.C. (1998). *Multivariate Statistical Inference and Applications.*
3. Johnson, R.A. & Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis* (6th ed.).
4. Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory.*
5. Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis* (3rd ed.).
6. Burden, R.L. & Faires, J.D. (2011). *Numerical Analysis* (9th ed.).
7. Strang, G. (2009 & 2016). *Introduction to Linear Algebra* (4th & 5th ed.).
8. Searle, S.R. & Khuri, A.I. (2017). *Matrix Algebra Useful for Statistics* (2nd ed.).