Introduction to Multivariate Statistics
Lecture 2: Basics

Lexin Li

University of California, Berkeley

Outline

- Review of matrix algebra:
  - Vector
  - Matrix
- Multivariate statistics
- Reading:
  - Chapter 2: Supplement
  - Chapter 2: 2.1, 2.2, 2.5, 2.6

Vector

- Vector:

$$\boldsymbol{a} = \begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \quad \boldsymbol{e} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

- Vector operations:

$$\begin{aligned} \boldsymbol{a} + \boldsymbol{b} &= \begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ -1 \end{pmatrix}, \\ \frac{1}{3}\boldsymbol{a} &= \begin{pmatrix} \frac{1}{3} \\ 1 \\ -\frac{2}{3} \end{pmatrix}, \\ \boldsymbol{a}^\mathsf{T}\boldsymbol{b} &= \begin{pmatrix} 1 & 3 & -2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = 1 \times 1 + 3 \times (-1) + (-2) \times 1 = -4. \end{aligned}$$

Vector

- Vector in abstract form:

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}, \quad \boldsymbol{a}_1 = \begin{pmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1p} \end{pmatrix}, \quad \boldsymbol{a}_2 = \begin{pmatrix} a_{21} \\ a_{22} \\ \vdots \\ a_{2p} \end{pmatrix}, \ldots, \boldsymbol{a}_n = \begin{pmatrix} a_{n1} \\ a_{n2} \\ \vdots \\ a_{np} \end{pmatrix},$$

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}, \quad \boldsymbol{x}_1 = \begin{pmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1p} \end{pmatrix}, \quad \boldsymbol{x}_2 = \begin{pmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2p} \end{pmatrix}, \ldots, \boldsymbol{x}_n = \begin{pmatrix} x_{n1} \\ x_{n2} \\ \vdots \\ x_{np} \end{pmatrix}.$$

Vector

- Vector operations:

$$
\begin{aligned}
\sum_{i=1}^{n} \boldsymbol{x}_i &= \boldsymbol{x}_1 + \boldsymbol{x}_2 + \ldots + \boldsymbol{x}_n \\
&= \begin{pmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1p} \end{pmatrix} + \begin{pmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2p} \end{pmatrix} + \ldots + \begin{pmatrix} x_{n1} \\ x_{n2} \\ \vdots \\ x_{np} \end{pmatrix} \\
&= \begin{pmatrix} x_{11} + x_{21} + \ldots + x_{n1} \\ x_{12} + x_{22} + \ldots + x_{n2} \\ \vdots \\ x_{n1} + x_{n2} + \ldots + x_{np} \end{pmatrix}, \\
\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i &= \frac{1}{n} \{ \boldsymbol{x}_1 + \boldsymbol{x}_2 + \ldots + \boldsymbol{x}_n \}.
\end{aligned}
$$

Vector

- Vector operations:

$$
\begin{aligned}
\boldsymbol{a}^{\mathsf{T}}\boldsymbol{x} &= \begin{pmatrix} a_1 & a_2 & \ldots & a_p \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_p \end{pmatrix} = a_1 x_1 + a_2 x_2 + \ldots + a_p x_p, \\
\boldsymbol{a}_1^{\mathsf{T}}\boldsymbol{x}_1 &= \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1p} \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \\ \ldots \\ x_{1p} \end{pmatrix} = a_{11} x_{11} + a_{12} x_{12} + \ldots + a_{1p} x_{1p}, \\
&\vdots \\
\boldsymbol{a}_n^{\mathsf{T}}\boldsymbol{x}_n &= \begin{pmatrix} a_{n1} & a_{n2} & \ldots & a_{np} \end{pmatrix} \begin{pmatrix} x_{n1} \\ x_{n2} \\ \ldots \\ x_{np} \end{pmatrix} = a_{n1} x_{n1} + a_{n2} x_{n2} + \ldots + a_{np} x_{np}, \\
\boldsymbol{a}_1^{\mathsf{T}}\boldsymbol{x}_1 + \boldsymbol{a}_2^{\mathsf{T}}\boldsymbol{x}_2 &= a_{11} x_{11} + a_{12} x_{12} + \ldots + a_{1p} x_{1p} + \\
&\qquad a_{21} x_{21} + a_{22} x_{22} + \ldots + a_{2p} x_{2p}, \\
\sum_{i=1}^{n} \boldsymbol{a}_i^{\mathsf{T}}\boldsymbol{x}_i &= \boldsymbol{a}_1^{\mathsf{T}}\boldsymbol{x}_1 + \boldsymbol{a}_2^{\mathsf{T}}\boldsymbol{x}_2 + \ldots + \boldsymbol{a}_n^{\mathsf{T}}\boldsymbol{x}_n,
\end{aligned}
$$

Matrix

- Matrix: row and column

$$\boldsymbol{A} = \begin{pmatrix} -7 & 2 \\ 0 & 1 \\ 3 & 4 \end{pmatrix}, \quad \boldsymbol{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.7 & -0.3 \\ 0.7 & 2 & 1 \\ -0.3 & 1 & 8 \end{pmatrix}.$$

- Matrix operations:

$$\begin{aligned}
\boldsymbol{A}^\mathsf{T} &= \begin{pmatrix} -7 & 0 & 3 \\ 2 & 1 & 4 \end{pmatrix}, \qquad 10\boldsymbol{A} = \begin{pmatrix} -70 & 20 \\ 0 & 10 \\ 30 & 40 \end{pmatrix}, \\
\boldsymbol{A}^\mathsf{T}\boldsymbol{a} &= \begin{pmatrix} -7 & 0 & 3 \\ 2 & 1 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix} = \begin{pmatrix} -13 \\ -3 \end{pmatrix} \\
\boldsymbol{A}^\mathsf{T}\boldsymbol{A} &= \begin{pmatrix} -7 & 0 & 3 \\ 2 & 1 & 4 \end{pmatrix} \begin{pmatrix} -7 & 2 \\ 0 & 1 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 58 & -2 \\ -2 & 21 \end{pmatrix}.
\end{aligned}$$

$\leftarrow$ like "square" a^2

Matrix

- Matrix operations:

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} 1.430 & -0.563 & 0.124 \\ -0.563 & 0.755 & -0.115 \\ 0.124 & -0.115 & 0.144 \end{pmatrix} \text{ such that } \boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma} = \boldsymbol{I},$$

$$\leftarrow \text{like "inverse" } \frac{1}{a}$$

$$\begin{aligned} \boldsymbol{a}^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{a} &= \begin{pmatrix} 1 & 3 & -2 \end{pmatrix} \begin{pmatrix} 1 & 0.7 & -0.3 \\ 0.7 & 2 & 1 \\ -0.3 & 1 & 8 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix} \\ &= \begin{pmatrix} 3.7 & 4.7 & -13.3 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix} = 44.4, \\ \boldsymbol{a}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{a} &= \begin{pmatrix} 1 & 3 & -2 \end{pmatrix} \begin{pmatrix} 1.430 & -0.563 & 0.124 \\ -0.563 & 0.755 & -0.115 \\ 0.124 & -0.115 & 0.144 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix} \\ &= \begin{pmatrix} -0.507 & 1.932 & -0.509 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix} = 6.307. \end{aligned}$$

$$\leftarrow \text{like "square and inverse" } \frac{a^2}{b}$$

Matrix

- Matrix in abstract form:

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \cdots \\ a_p \end{pmatrix}_{p\times 1} \quad \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{n\times 1} \quad \boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}_{n\times p}$$

$$\begin{aligned} \boldsymbol{X}^{\mathsf{T}}\boldsymbol{y} &= \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \\ &= \begin{pmatrix} x_{11}y_1 + x_{21}y_2 + \ldots + x_{n1}y_n \\ x_{12}y_2 + x_{22}y_2 + \ldots + x_{n2}y_n \\ \vdots \\ x_{1p}y_1 + x_{2p}y_2 + \ldots + x_{np}y_n \end{pmatrix}_{n\times 1} \end{aligned}$$

Matrix

- Matrix in abstract form:

$$
\mathbf{X}^{\mathsf{T}}\mathbf{X} = \begin{pmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} =
$$

$$
\begin{pmatrix} x_{11}^2 + x_{21}^2 + \ldots + x_{n1}^2 & \dots & x_{11}x_{1p} + x_{21}x_{2p} + \ldots + x_{n1}x_{np} \\ \vdots & \ddots & \vdots \\ x_{11}x_{1p} + x_{21}x_{2p} + \ldots + x_{n1}x_{np} & \dots & x_{1p}^2 + x_{2p}^2 + \ldots + x_{np}^2 \end{pmatrix}_{p\times p}
$$

$$
\begin{aligned}
(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1} &= \\
\mathbf{a}^{\mathsf{T}}(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{a} &=
\end{aligned}
$$

Matrix

- Spectral decomposition of a square matrix $\boldsymbol{A}_{p\times p} = \boldsymbol{V}_{p\times p}\boldsymbol{\Lambda}_{p\times p}\boldsymbol{V}^{\mathsf{T}}_{p\times p}$:

$$\boldsymbol{A} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_p) \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_1^{\mathsf{T}} \\ \boldsymbol{v}_2^{\mathsf{T}} \\ \vdots \\ \boldsymbol{v}_p^{\mathsf{T}} \end{pmatrix}$$

  - $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_p$ are called eigenvectors, and $\boldsymbol{v}_1^{\mathsf{T}}\boldsymbol{v}_1 = 1$, and so on; $\lambda_1, \lambda_2, \dots, \lambda_p$ are called eigenvalues, and $\lambda_1 \geq 0$, and so on.
  - matrix determinant: $|\boldsymbol{A}| = \lambda_1\lambda_2\dots\lambda_p$.
  - example:

$$\begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} -0.851 & 0.526 \\ -0.526 & -0.851 \end{pmatrix} \begin{pmatrix} 3.618 & 0 \\ 0 & 1.382 \end{pmatrix} \begin{pmatrix} -0.851 & -0.526 \\ 0.526 & -0.851 \end{pmatrix}$$

Matrix

- Singular value decomposition (SVD) of a rectangular matrix $\boldsymbol{B}_{n\times p} = \boldsymbol{U}_{n\times p}\ \boldsymbol{D}_{p\times p}\ \boldsymbol{V}^{\mathsf{T}}_{p\times p}$:

$$\boldsymbol{B} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_p) \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_p \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_1^{\mathsf{T}} \\ \boldsymbol{v}_2^{\mathsf{T}} \\ \vdots \\ \boldsymbol{v}_p^{\mathsf{T}} \end{pmatrix}$$

  - $\boldsymbol{u}_1, \boldsymbol{v}_2, \dots, \boldsymbol{u}_p$ are called left singular vectors, and $\boldsymbol{u}_1^{\mathsf{T}}\boldsymbol{u}_1 = 1$, and so on; $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_p$ are called right singular vectors, and $\boldsymbol{v}_1^{\mathsf{T}}\boldsymbol{v}_1 = 1$, and so on; $d_1, d_2, \dots, d_p$ are called singular values, and they can be $\pm$.
  - example:

$$\begin{pmatrix} 3 & 1 \\ 1 & 2 \\ -2 & 4 \end{pmatrix} = \begin{pmatrix} 0.022 & 0.880 \\ -0.325 & 0.454 \\ -0.945 & -0.136 \end{pmatrix} \begin{pmatrix} 4.702 & 0 \\ 0 & 3.590 \end{pmatrix} \begin{pmatrix} 0.347 & 0.938 \\ -0.938 & 0.347 \end{pmatrix}$$

Multivariate statistics

- A **univariate** random variable: X
  - Sample observations: $x_1, x_2, \ldots, x_n$
  - Mean: $\mu = E(X)$
    Sample mean:
    $$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$
  - Variance: $\sigma = \mathrm{var}(X)$
    Sample variance:
    $$s = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$
  - Covariance: $\sigma_{X,Y} = \mathrm{cov}(X, Y)$
    Sample covariance:
    $$s_{X,Y} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$
  - Correlation: $\rho = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}}$

Multivariate statistics

- Example:
  - Researchers have suggested that a change in skull size over time is evidence of the interbreeding of a resident population with immigrant populations.
  - Measurements:
    X_1 = base height of skull (mm)
    X_2 = base length of skull (mm)
  - $n = 5$ data points:
    $X_1 : 138, 131, 132, 143, 137$
    $X_2 : 89, 92, 99, 100, 89$
  -

$$
\begin{aligned}
\bar{x}_1 &= \frac{1}{5}(138 + 131 + 132 + 143 + 137) = 136.2 \\
s_1 &= \frac{1}{5}\{(138 - 136.2)^2 + (131 - 136.2)^2 + (132 - 136.2)^2 + \\
&\quad (143 - 136.2)^2 + (137 - 136.2)^2\} = 18.96 \\
s_{x_1, x_2} &= \frac{1}{5}\{(138 - 136.2)(89 - 93.8) + (131 - 136.2)(92 - 93.8) + \\
&\quad (132 - 136.2)(99 - 93.8) + (143 - 136.2)(100 - 93.8) + \\
&\quad (137 - 136.2)(89 - 93.8)\} = 4.3
\end{aligned}
$$

Multivariate statistics

- A multivariate random vector of random variables:

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}_{p\times 1}$$

  - Sample observations: $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$

$$\begin{pmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1p} \end{pmatrix}, \quad \begin{pmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2p} \end{pmatrix}, \quad \begin{pmatrix} x_{n1} \\ x_{n2} \\ \vdots \\ x_{np} \end{pmatrix}.$$

  - Mean vector:

$$\boldsymbol{\mu} = E(\boldsymbol{X}) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}_{p\times 1}$$

Multivariate statistics

- A multivariate random vector of random variables:
  - Covariance matrix:

$$
\begin{aligned}
\boldsymbol{\Sigma} = \mathrm{cov}(\boldsymbol{X}) &= \begin{pmatrix} \mathrm{var}(X_1) & \mathrm{cov}(X_1, X_2) & \dots & \mathrm{cov}(X_1, X_p) \\ \mathrm{cov}(X_2, X_1) & \mathrm{var}(X_2) & \dots & \mathrm{cov}(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(X_p, X_1) & \mathrm{cov}(X_p, X_2) & \dots & \mathrm{var}(X_p) \end{pmatrix} \\
&= \begin{pmatrix} \sigma_1 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{12} & \sigma_2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \dots & \sigma_p \end{pmatrix}_{p \times p}
\end{aligned}
$$

  - Correlation matrix:

$$
\boldsymbol{\rho} = \begin{pmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \dots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} \\ \frac{\sigma_{21}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{11}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \dots & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\sigma_{p1}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{11}}} & \frac{\sigma_{p2}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{22}}} & \dots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{pp}}} \end{pmatrix} = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \dots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{pmatrix}_{p \times p}
$$

Multivariate statistics

- Example:

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \text{base height of skull} \\ \text{base length of skull} \end{pmatrix}$$

$$\begin{aligned}
\hat{\boldsymbol{\mu}} &= \begin{pmatrix} 136.2 \\ 93.8 \end{pmatrix} \\
\hat{\boldsymbol{\Sigma}} &= \begin{pmatrix} 18.96 & 4.3 \\ 4.3 & 22.96 \end{pmatrix} \quad \text{or} \quad \hat{\boldsymbol{\Sigma}} = \begin{pmatrix} 23.7 & 4.3 \\ 4.3 & 28.7 \end{pmatrix} \\
\hat{\rho} &= \begin{pmatrix} 1 & 0.206 \\ 0.206 & 1 \end{pmatrix} \quad \text{or} \quad \hat{\rho} = \begin{pmatrix} 1 & 0.165 \\ 0.165 & 1 \end{pmatrix}
\end{aligned}$$