What to do today ?

Part I. Introduction and Preparation

Part I.1. General Introduction

Part I.2. Review on Matrix Algebra (Chp 2.1-4, Supplement 2A)

Part I.3. Introduction to R (More at the 1st tutorial)

Part I.4. Multivariate Random Variables and Distributions (Chp 1, 2.5-6, 3)

Part I.4.1 Random Vectors and Matrices (Chp 2.5)
Part I.4.3 Descriptive Multivariate Analysis (Chp 1)
Part I.4.4 More on Descriptive Multivariate Analysis (Chp 3)

Part II. Inference under Multivariate Normal Distribution (Textbook Chp 4-7)

I.2.4. Matrix Operations

- The **determinant** of a square matrix $\mathbf{A} = \left(a_{ij}\right)_{k \times k}$ is denoted by $|\mathbf{A}|$ or $det(\mathbf{A})$:

$$\begin{array}{ll} |\mathbf{A}| = a_{11}, & k = 1 \\ |\mathbf{A}| = \sum_{j=1}^{k} (-1)^{1+j} a_{1j} |\mathbf{A}_{1j}|, & k > 1 \end{array}$$

- The **trace** of a square matrix $\mathbf{A} = \left(a_{ij}\right)_{k \times k}$ is $tr(\mathbf{A}) = \sum_{i=1}^{k} a_{ii}$.
- **partitioned matrices:** block representation and block multiplication.
- A square matrix $\mathbf{A}$ is **positive (semi-positive) definitive** if $\mathbf{x}^{'}\mathbf{A}\mathbf{x} > 0 (\geq 0)$ for any $\mathbf{x} \neq \mathbf{0}$.
- **eigenvalue** and **eigenvector**: for a square matrix $\mathbf{A}$, if there are λ and nonzero $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, λ and $\mathbf{x}$ are $\mathbf{A}$'s eigenvalue and eigenvector, respectively.

- **random variable.** $X : \mathcal{S} \longrightarrow \mathcal{R}$. cdf $F(x) = P(X \leq x)$; discrete r.v.; continuous r.v.; expectation $E(X)$; variance $V(X)$.
- **covariance/correlation (two r.v.s):** $Cov(X, Y) = E[(X - EX)(Y - EY)] = E(XY) - (EX)(EY)$; $corr(X, Y) = Cov(X, Y)/\sqrt{V(X)V(Y)}$.
- **random vector:** $\mathbf{X} = (X_1, \ldots, X_p)'$.
- **joint distribution:** joint cdf $F(\mathbf{x}) = P(X_1 \leq x_1, \ldots, X_p \leq x_p)$. If $X_1, \ldots, X_p$ mutually independent then $F(\mathbf{x}) = \prod_{j=1}^{p} F_{X_j}(x_j)$.
- **mean vector:** $E(\mathbf{X}) = \boldsymbol{\mu} = (E(X_1), \ldots, E(X_p))'$.
- **covariance matrix:** $\Sigma = V(\mathbf{X}) = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] = E(\mathbf{X}\mathbf{X}') - \boldsymbol{\mu}\boldsymbol{\mu}'$ with entries $\sigma_{ij} = Cov(X_i, X_j)$; symmetric; (semi-)positive definite
- **Key properties of Σ:** symmetric; (semi-)positive definite.
- **correlation matrix:** $\rho = (\rho_{ij})$ with $\rho_{ij} = corr(X_i, X_j) = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$; symmetric; SPD; Let $\boldsymbol{V} = \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{pp})$, $\rho = \boldsymbol{V}^{-1/2}\Sigma\boldsymbol{V}^{-1/2}$ and $\Sigma = \boldsymbol{V}^{1/2}\rho\boldsymbol{V}^{1/2}$.

Linear Combination of Random Variables

Suppose $Y = c_1X_1 + c_2X_2 + \ldots + c_pX_p$.

- $Y = \mathbf{c}'\mathbf{X}$.
- $E(Y) = \mathbf{c}'E(\mathbf{X}) = \mathbf{c}'\boldsymbol{\mu}$.
- $V(Y) = \mathbf{c}'V(\mathbf{X})\mathbf{c} = \mathbf{c}'\Sigma\mathbf{c}$.

Suppose $Z_j = c_{j1}X_1 + c_{j2}X_2 + \ldots + c_{jp}X_p$ and $\mathbf{Z} = (Z_1, \ldots, Z_q)'$.

- $\mathbf{Z} = \mathbf{C}_{q\times p}\mathbf{X}_{p\times 1}$.
- $E(\mathbf{Z}) = \mathbf{C}E(\mathbf{X}) = \mathbf{C}\boldsymbol{\mu}$.
- $V(\mathbf{Z}) = \mathbf{C}V(\mathbf{X})\mathbf{C}' = \mathbf{C}\Sigma\mathbf{C}'$.

The **covariance** of two random vectors $\mathbf{X}$ and $\mathbf{Y}$ is defined as

$$Cov(\mathbf{X}, \mathbf{Y}) = E\left[\mathbf{X} - E(\mathbf{X})\right]\left[\mathbf{Y} - E(\mathbf{Y})\right]'.$$

Suppose $\mathbf{U} = \mathbf{AX}$ and $\mathbf{W} = \mathbf{BY}$.

- $Cov(\mathbf{U}, \mathbf{W}) = \mathbf{A}\,Cov(\mathbf{X}, \mathbf{Y})\mathbf{B}'$

Part I.4.3 Descriptive Multivariate Analysis

Summary Statistics:

Suppose a study has n iid observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ on a p-dim r.v. $\mathbf{X} = (X_1, \ldots, X_p)'$ with $E(\mathbf{X}) = \boldsymbol{\mu}$ and $V(\mathbf{X}) = \Sigma$.

- $E(\mathbf{X}_i) = \boldsymbol{\mu}$ and $V(\mathbf{X}_i) = \Sigma$ for $i = 1, \ldots, n$.
- **sample mean** $\bar{\mathbf{x}} = (\bar{x}_1, \ldots, \bar{x}_p)'$ with $\bar{x}_j = \sum_{i=1}^n x_{ij}/n$.
- **sample variance**
$$\mathbf{S}_n = \begin{pmatrix} s_{11} & \ldots & s_{1p} \\ s_{21} & \ldots & s_{2p} \\ \vdots & \ldots & \vdots \\ s_{p1} & \ldots & s_{pp} \end{pmatrix}$$
with $s_{jk} = \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)/n$.
- **sample correlations**
$$\mathbf{R}_n = \begin{pmatrix} 1 & r_{12} & \ldots & r_{1p} \\ r_{21} & 1 & \ldots & r_{2p} \\ \vdots & \ldots & \vdots & \\ r_{p1} & r_{p2} & \ldots & 1 \end{pmatrix}$$
with $r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}}\sqrt{s_{kk}}}$.

Part I.4.3 Descriptive Multivariate Analysis with R

What will we study in the next class?

- **Part I. Introduction and Preparation**
  - *I.1. General Introduction*
  - *I.2. Review on Matrix Algebra*
  - *I.3. Introduction to R*
  - **I.4. Multivariate Random Variables and Distributions (Chp 1, 2.5-6, 3)**
    - *I.4.1 Random Vectors and Matrices (Chp 2.5)*
    - *I.4.2 Mean Vectors and Covariance Matrices (Chp 2.6)*
    - *I.4.3 Descriptive Multivariate Analysis (Chp 1)*
    - **I.4.4 More on Descriptive Multivariate Analysis (Chp 3)**

- **Part II. Inference under Multivariate Normal Distribution (Textbook Chp 4-7)**
  - **II.1 Multivariate Normal Distribution (Chp 4)**
  - *II.2 Inferences on Mean Vector (Chp 5)*
  - *II.3 Comparisons of Several Mean Vector (Chp 6)*
  - *II.4 Multivariate Linear Regression (Chp 7)*

- *Part III. Commonly-Used Multivariate Analysis Methods (Textbook Chp 8-11)*
- *Part IV. Other Topics (Textbook Chp 12)*