STAT-445/645: Applied Multivariate Analysis

K. Ken Peng

Department of Statistics and Actuarial Science
Simon Fraser University

Spring 2026

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)

I.2.1. Why do we need matrix/vector algebra in STAT445/645?
I.2.2. Notation and Basic Definitions
I.2.3. Vector Operations
I.2.4. Matrix Operations

Part I.3. Introduction to R

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.4.2 Mean Vectors and Covariance Matrices (Chp 2.6)

I.2.2. Notation and Basic Definitions

- a real number; a **scalar**; a physical quantity
- A **vector** is a group of p numbers/elements arranged in a *column*: a p-dim vector.
- A $p \times q$ **matrix** is a group of mk numbers/elements arranged into a rectangular array with p *rows* and q *columns*.

I.2.3. Vector Operations

- **scalar multiplication**, **addition**, **subtraction**;
- **inner product**, **length of a vector**, **angle between two vectors**

I.2.4. Matrix Operations

- **transpose** , **scalar multiplication**, **addition**, **subtraction**.

  Properties. Associative property; distributive property; commutative property

- **matrix multiplication**. The **product** of $p \times q$ matrix $\mathbf{A} = \left(a_{ij}\right)$ and $q \times k$ matrix $\mathbf{B} = \left(b_{ij}\right)$ is $\mathbf{AB} = \mathbf{C} = \left(c_{ij}\right)$, a $p \times k$ matrix with $c_{ij} = \sum_{l=1}^{q} a_{il} b_{lj}$.

  Properties. Associative; distributive over addition; not commutative (!); $(\mathbf{AB})^{'} = \mathbf{B}^{'}\mathbf{A}^{'}$

- **inverse** of a square matrix $\mathbf{A} = \left(a_{ij}\right)$ is $\mathbf{B} = \mathbf{A}^{-1}$ such that $\mathbf{BA} = \mathbf{AB} = \mathbf{I}$, the identity matrix.
  - If $\mathbf{A}^{-1}$ exists, $\mathbf{A}$ is *invertible (nonsingular, full rank)*.
  - $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ if well-defined.
  - $(\mathbf{A}^{'})^{-1} = \left(\mathbf{A}^{-1}\right)^{'}$ if well-defined.
  - If $\mathbf{A}^{'}\mathbf{A} = \mathbf{I}$, $\mathbf{A}$ is **orthogonal**. Iff $\mathbf{A}^{'} = \mathbf{A}^{-1}$, $\mathbf{A}$ is **orthogonal**.

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

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

where $\mathbf{A}_{1j}$ is the $(k-1)\times(k-1)$ matrix obtained from $\mathbf{A}$ after deleting its first row and jth column.

e.g.

$$\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = 1 * |4|(-1)^2 + 2 * |3|(-1)^3 = -2$$

$$\begin{vmatrix} 3 & 1 & 6 \\ 7 & 4 & 5 \\ 2 & -7 & 1 \end{vmatrix} = 3*\begin{vmatrix} 4 & 5 \\ -7 & 1 \end{vmatrix}(-1)^2+1*\begin{vmatrix} 7 & 5 \\ 2 & 1 \end{vmatrix}(-1)^3+6*\begin{vmatrix} 7 & 4 \\ 2 & -7 \end{vmatrix}(-1)^4 = -222.$$

Property. $\left|\mathbf{A}\right| = \left|\mathbf{A}'\right|$; $\left|\mathbf{A}^{-1}\right| = \left|\mathbf{A}\right|^{-1}$; $\left|\mathbf{AB}\right| = \left|\mathbf{A}\right|\left|\mathbf{B}\right|$; $\left|c\mathbf{A}\right| = c^k\left|\mathbf{A}\right|$.

- The **trace** of a square matrix $\mathbf{A} = \left(a_{ij}\right)_{k \times k}$ is $tr(\mathbf{A}) = \sum_{i=1}^{k} a_{ii}$.
  - $tr(c\mathbf{A}) = c * tr(\mathbf{A})$.
  - $tr(\mathbf{A} + \mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B})$.
  - $tr(\mathbf{AB}) = tr(\mathbf{BA})$.
  - $tr(\mathbf{AA}') = \sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij}^2$.

- **partitioned matrices** e.g.

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$$

$$\mathbf{AB} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}$$

- A square matrix $\mathbf{A}$ is **positive (semi-positive) definitive** if $\mathbf{x}'\mathbf{A}\mathbf{x} > 0 (\geq 0)$ for any $\mathbf{x} \neq \mathbf{0}$.
  e.g. If $\mathbf{A} = \mathbf{B}'\mathbf{B}$, $\mathbf{A}$ is (semi -) positvie definite.

- **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.

  - If $\lambda_1, \ldots, \lambda_k$ are all the eigenvalues of $\mathbf{A}_{k\times k}$, $tr(\mathbf{A}) = \sum_{j=1}^{k} \lambda_j$ and $det(\mathbf{A}) = \prod_{j=1}^{k} \lambda_j$.

  - $\mathbf{A}_{k\times k}$ is positive definite (semi-positive definite) iff all its eigenvalues are positive (non-negative).

Recall ... ...

I.2.1. Why do we need matrix/vector algebra in STAT445/645?

For example, consider

$$Y = \beta_0 + \beta X + \epsilon, \quad E(\epsilon) = 0, \; V(\epsilon) = \sigma^2$$

with observations $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ from n independent units. That is, for $i = 1, \ldots, n$,

$$y_i = \beta_0 + \beta x_i + \epsilon_i, \quad E(\epsilon_i) = 0, \; V(\epsilon_i) = \sigma^2$$

with $\epsilon_1, \ldots, \epsilon_n$ indpt.

- LSE: $\hat{\beta} = \frac{S_{XY}}{S_{XX}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$; $\hat{\beta}_0 = \bar{y} - \hat{\beta}\bar{x}$.
- $V(\hat{\beta}) = \sigma^2 / S_{XX}$ and $V(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{XX}}\right]$;
- $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ with $e_i = y_i - \hat{y}_i = y_i - [\hat{\beta}_0 + \hat{\beta} x_i]$.

What if to consider

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p + \epsilon, \quad E(\epsilon) = 0, \; V(\epsilon) = \sigma^2$$

with observations $\{(x_{11}, \ldots, x_{p1}, y_1), \ldots, (x_{1n}, \ldots, x_{pn}, y_n)\}$ from n independent units?

What if to study how $(Y_1, \ldots, Y_k)$ depends on $(X_1, \ldots, X_p)$?

$\Longrightarrow$ **vector/matrix algebra** as a tool for communication in general, together with software packages such as R and SAS to conduct the required computing.

Multiple linear regression model. $Y = \boldsymbol{\beta}^{'}\mathbf{x} + \epsilon$ with $E(\epsilon) = 0,\ V(\epsilon) = \sigma^2$, where

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_p \end{pmatrix}.$$

with observations $\{(x_{11}, \ldots, x_{p1}, y_1), \ldots, (x_{1n}, \ldots, x_{pn}, y_n)\}$ from independent units $i = 1, \ldots, n$, $\mathbf{y} = \mathbf{X}^{'}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ x_{11} & x_{12} & \ldots & x_{1n} \\ \vdots & \vdots & \ldots & \vdots \\ x_{p1} & x_{p2} & \ldots & x_{pn} \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

with $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $V(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$.

- LSE for $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.
- $V(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$
- $\hat{\sigma}^2 = \frac{1}{n-(p+1)}\mathbf{e}'\mathbf{e}$ with

$$\mathbf{e} = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} y_1 - \hat{y}_1 \\ \vdots \\ y_n - \hat{y}_n \end{pmatrix} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}'\hat{\beta}$$

as $e_i = y_i - \hat{y}_i = y_i - \hat{\beta}'\mathbf{x}_i$ for $i = 1, \ldots, n$.

Part I.3. Introduction to R

Part I.4.1 Random Vectors and Matrices (Chp 2.5)

Review on (univariate) random variable:

- **random variable.** $X : \mathcal{S} \longrightarrow \mathcal{R}$
- **distribution.** The cumulative distribution function (cdf) of X: $F(x) = P(X \leq x)$
  - If X is discrete and all its possible values are $a_1, a_2, \ldots$, $F(x) = \sum_{a_j \leq x} p(a_j)$ with its probability mass function (pmf) $p(x) = P(X = x)$ for $x = a_1, a_2, \ldots$
  - If X is continuous, $F(x) = \int_{-\infty}^{x} f(u)du$ with its probability density function (pdf) $f(x)$ for $-\infty < x < \infty$.
- Examples
  - e.g. "tossing a coin": $\mathcal{S} = \{H, T\}$; $X = 1/0$ if getting H/T.
  - e.g. "waiting time for a bus": X with $0 < X < a$.

The commonly-used (univariate) distributions, discrete/continuous?

Part I.4.1 Random Vectors and Matrices (Chp 2.5)

- **random vector (multivariate random variable).**

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} = (X_1, X_2, \ldots, X_p)'$$

  is a p-dim random vector if $X_1, \ldots, X_p$ are r.v.s.

- **distribution.** The cdf of **X** is the joint cumulative distribution function (joint cdf) of $X_1, \ldots, X_p$: for $\mathbf{x} = (x_1, x_2, \ldots, x_p)'$,

$$F(\mathbf{x}) = P(\mathbf{X} \le \mathbf{x}) = P(X_1 \le x_1, \ldots, X_p \le x_p).$$

  - Iff $X_1, \ldots, X_p$ are multually independent from each other,

$$F(\mathbf{x}) = P(X_1 \le x_1) \ldots P(X_p \le x_p) = \prod_{j=1}^{p} F_{X_j}(x_j),$$

    the product of $X_1, \ldots, X_p$'s (*marginal*) cdfs.

Part I.4.1 Random Vectors and Matrices (Chp 2.5)

Examples

- e.g. "tossing a coin twice": $\mathcal{S} = \{(H, H), (H, T), (T, H), (T, T)\}$; $X_1 = 1$ or 0 if the 1st toss gets H or T and $X_2 = 1$ or 0 if the 2nd toss gets H or T.
  - distribution?

- e.g. "two housemates' waiting time for a bus": $\mathbf{X} = (X_1, X_2)'$ with $0 < X_j < a_j$.
  - distribution?

Part I.4.1 Random Vectors and Matrices (Chp 2.5)

Example of Multivariate Data. Consider variables X_1=tumor stage, X_2=age, X_3=sex, X_4=blood pressure, ... in a study. Suppose the study has n subjects with $\mathbf{X}_1, \ldots, \mathbf{X}_n$ their observations on the p-dim r.v. $\mathbf{X} = (X_1, \ldots, X_p)^{'}$.

$$\begin{pmatrix} X_{11} & X_{12} & \ldots & X_{1p} \\ X_{21} & X_{22} & \ldots & X_{2p} \\ \vdots & \vdots & \ldots & \vdots \\ X_{n1} & X_{n2} & \ldots & X_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1^{'} \\ \mathbf{X}_2^{'} \\ \vdots \\ \mathbf{X}_n^{'} \end{pmatrix}$$

is a $n \times p$ **random matrix.**

- Often assume the n observations *independent and identically distributed* (iid).
  That is, $\mathbf{X}_1, \ldots, \mathbf{X}_n$ form a *random sample of size* n from the population.

Part I.4.2 Mean Vectors and Covariance Matrices (Chp 2.6)

Review on (univariate) random variable:

- **expectation (population mean).** r.v. X's expectation, denoted by $E(X)$.
  - If X is discrete and all its possible values are $a_1, a_2, \ldots$ with its probability mass function (pmf) $p(x) = P(X = x)$ for $x = a_1, a_2, \ldots$,

$$E(X) = \sum_{all\ \ a_j} a_j p(a_j) = a_1 p(a_1) + a_2 p(a_2) + \ldots$$

  - If X is continuous with its probability density function (pdf) $f(x)$ for $-\infty < x < \infty$,

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

- **(population) variance.** r.v. X's variance, denoted by $V(X)$, is $E\big[X - E(X)\big]^2 = E(X^2) - (EX)^2$.

- **Examples.**
  - e.g. "tossing a coin": $\mathcal{S} = \{H, T\}$; $X = 1/0$ if getting H/T: $E(X) = 1/2$ and $V(X) = 1/4$ for an even coin.
  - e.g. "waiting time for a bus": X with $0 < X < a$: $E(X) = a/2$ and $V(X) = a^2/12$ for $X \sim U(0, a)$.
- **covariance.** Two r.v.. X and Y's covariance is $Cov(X, Y) = E\big[X - E(X)\big]\big[Y - E(Y)\big] = E(XY) - (EX)(EY)$.
  - If X and Y are indpt, $Cov(X, Y) = 0$.
  - If $Cov(X, Y) = 0$, are X and Y indpt?
  - **(population) correlation coefficient.**

$$corr(X, Y) = Cov(X, Y)/\sqrt{V(X)V(Y)}$$

Part I.4.2 Mean Vectors and Covariance Matrices (Chp 2.6)

- **expectation (population mean).** p-dim random vector $\mathbf{X}$'s expectation:

$$E(\mathbf{X}) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{pmatrix},$$

denoted by $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$.

- **(population) variance.** p-dim r.v. $\mathbf{X}$'s variance:

$$V(\mathbf{X}) = E[\mathbf{X} - E(\mathbf{X})][\mathbf{X} - E(\mathbf{X})]' = E(\mathbf{X}\mathbf{X}') - (E\mathbf{X})(E\mathbf{X})',$$

denoted by

$$\Sigma = E[\mathbf{X} - E(\mathbf{X})][\mathbf{X} - E(\mathbf{X})]' = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2p} \\ \vdots & \vdots & \ldots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \ldots & \sigma_{pp} \end{pmatrix} = (\sigma_{ij})$$

with $\sigma_{ij} = \sigma_{ji} = Cov(X_i, X_j)$.

Part I.4.2 Mean Vectors and Covariance Matrices (Chp 2.6)

Remarks on Σ.

- symmetric, (semi-)positive definite and thus invertible.
- the mean of the *outer product* of the *centered* random vector and itself.
- If $X_1, \ldots, X_p$ are indpt, $\Sigma = diag(\sigma_{11}, \ldots, \sigma_{pp})$. How about vice versa?

population correlation matrix. A standardized variance-covariance matrix:

$$\boldsymbol{\rho} = \begin{pmatrix} 1 & \rho_{12} & \ldots & \rho_{1p} \\ \rho_{21} & 1 & \ldots & \rho_{2p} \\ \vdots & \vdots & \ldots & \vdots \\ \rho_{p1} & \rho_{p2} & \ldots & 1 \end{pmatrix} = (\rho_{ij})$$

with $\rho_{ij} = corr(X_i, X_j)$.

- symmetric, positive definite and thus invertible.
- If $X_1, \ldots, X_p$ are indpt, $\boldsymbol{\rho} = \mathbf{I}$. How about vice versa?

$$\mathbf{V} == \begin{pmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \sigma_{pp} \end{pmatrix} = diag\big(\sigma_{11}, \dots, \sigma_{pp}\big),$$

$$\mathbf{V}^{1/2} = diag\big(\sqrt{\sigma_{11}}, \dots, \sqrt{\sigma_{pp}}\big), \quad \mathbf{V}^{-1/2} = diag\big(\frac{1}{\sqrt{\sigma_{11}}}, \dots, \frac{1}{\sqrt{\sigma_{pp}}}\big).$$

$$\boldsymbol{\rho} = \mathbf{V}^{-1/2}\Sigma\mathbf{V}^{-1/2}, \quad \Sigma = \mathbf{V}^{1/2}\boldsymbol{\rho}\mathbf{V}^{1/2}$$

$\rho_{ij} = \sigma_{ij} \big/ \sqrt{\sigma_{ii}\sigma_{jj}}$.

Example. "two housemates' waiting time for a bus": $\mathbf{X} = (X_1, X_2)'$ with X_1, X_2 indpt and following $U(0, a_1), U(0, a_2)$.

- $\boldsymbol{\mu} = E(\mathbf{X}) = (a_1/2, a_2/2)'$.
- $V(\mathbf{X}) = diag(a_1^2, a_2^2)/12$.

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)**