STAT-445/645: Applied Multivariate Analysis

K. Ken Peng

Department of Statistics and Actuarial Science
Simon Fraser University

Spring 2026

What to do today?

A: Course Syllabus

B: I. Introduction and Preparation

I.1. General Introduction

STAT-445/645: Applied Multivariate Analysis

Instructor: K. Ken Peng (kangyi_peng@sfu.ca; K9516)

Lecture Time and Location:
Tue, 4:30-6:20 pm DFA300 (Jan 5 - Jan 23, 2026);
Tue, 4:30-6:20 pm K9500 (Jan 26 - Apr 10, 2026);
Thu, 4:30-5:20 pm WMC3520 (Jan 5 - Apr 10, 2026)

Office Hour: Tue/Thu 3:00-4:00 pm, or by appointment
location: K10504

Teaching Assistants: Muthukuda Arachchilage, Niwanthi for tutorials; Nguyen, Chelsy for marking

Tutorials: (starting from the week of Jan 12)
STAT445-E101: Th 6:30-7:20PM; BLU9402
STAT445-E102: Th 7:30-8:20PM; BLU9402
STAT445-E103: Mo 4:30-5:20PM; BLU9402

STAT-445/645: Applied Multivariate Analysis

Textbook:

- "*Applied Multivariate Statistical Analysis*", by R.A. Johnson and D.W. Wichern. Publisher: Prentice Hall

References:

- "Applied Multivariate Statistical Analysis" (2nd Edition), by W. Hardle and L. Simar
- "Methods of Multivariate Analysis", by A.C. Rencher

Computer Software: *R* will be used in class (URL http://www.r-project.org/)

STAT-445/645: COURSE OUTLINES

Part I. Introduction (Textbook Chp 1-3), week 1–2.

- I.1. General Introduction
- I.2. Review on Matrix Algebra
- I.3. Introduction to R
- I.4. Multivariate Random Variables and Distributions

Part II. Inference under Multivariate Normal Distribution (Textbook Chp 4-7), week 3–7.

- II.1. Multivariate Normal Distribution
- II.2. Inference on mean vector
- II.3. Comparisons of Several Mean Vector
- II.4. Multivariate Linear Regression

Part III. Commonly-Used Multivariate Analysis Methods (Textbook Chp 8-11), week 8–13.

- III.1. Discrimination and Classification
- III.2. Principal Components Analysis
- III.3. Factor Analysis
- III.4. Clustering Analysis

Part IV. Other Topics

STAT-445/645: COURSE EVALUATION

- Homework - 40% (7 assignments: 8% per homework, the 5 highest marks are used in the final evaluation)
- Midterm - 20% (one midterm in class; 15% for STAT 645)
- Final - 40% (25% for STAT 645)
- Project - 20% (for STAT 645 only)
- Class Participation - 3-5% bonus (bonus questions during exams)

STAT-445/645: Assignments

- **Late policy:** Late assignments are not accepted. The best 5 out of 7 assignments are counted.
- **Presentation:** Clarity of presentation matters. Work that is difficult to follow may receive reduced credit.
- **Content:** Assignments include both written calculations and applied questions.
- **Regrading:** Requests must be submitted in writing within one week, clearly indicating the specific parts in question. Regrading may result in a higher, lower, or unchanged score.
- **Collaboration:** Discussion of general ideas is encouraged, but all submitted solutions and code must be your own.
- **Use of AI:** Allowed for exploration and understanding; not allowed for producing assignment solutions. Over-reliance on AI is likely to result in poor exam performance.
  AI is not your teacher— teach your own AI.

STAT 445 / STAT 645: Midterm

- **Schedule:** During class time on March 5.
- **Cheat sheet:** One A4 cheat sheet and a calculator are permitted. Cheat sheets will be collected; please write your name and student ID.
- **Make-up:** No make-up midterm! Students who miss the midterm due to documented illness may be required to complete an oral examination to recover the missing credit.
- **Content:** Similar to those in assignments; written calculations; interpretation of `R` output; analysis planning based on a given question and data description. Coverage includes Topics I.1–II.4.

STAT-445/645: Final

- **Schedule**: TBD
- **Cheat sheet**: One A4 cheat sheet and a calculator are permitted.
- **Content**: Similar to those in assignments; written calculations; interpretation of R output; analysis planning based on a given question and data description. Cover I.1. - III.4.

STAT-445/645: Final Project (645 only)

- Topic to be assigned in Week 11.
- Report to be collected by the end of the term.

STAT-445/645: Class Participation (3 - 5% bonus)

- Bonus questions in midterm/final exams. Will be something closely related to examples during class time.

Why to study STAT-445/645?

- Why study statistics?
  "*Statistics is the science of learning from data.*"
  - There are a plenty of data.
  - Data: observations on variables with **uncertainty**.
  - Analysis on **uncertainty** matters.
  - Statistical analyses provide not only results, but also the confidence of the results.

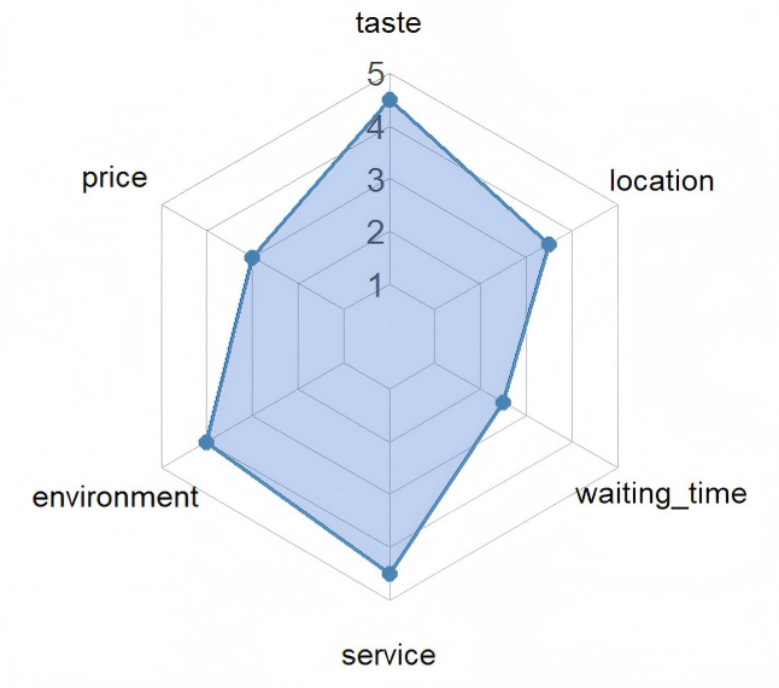
- Different types of variables:
  - categorical vs quantitative variables
  - discrete vs continuous variables
  - single (univariate) vs **multiple variables (multivariate)**

- Tons of *multivariate data*: how to analyze them? $\Longrightarrow$ **STAT-445/645**

Example - rating a restaurant

How would you rate a restaurant from 1-5?

- Rate different aspects, and combine them into one?

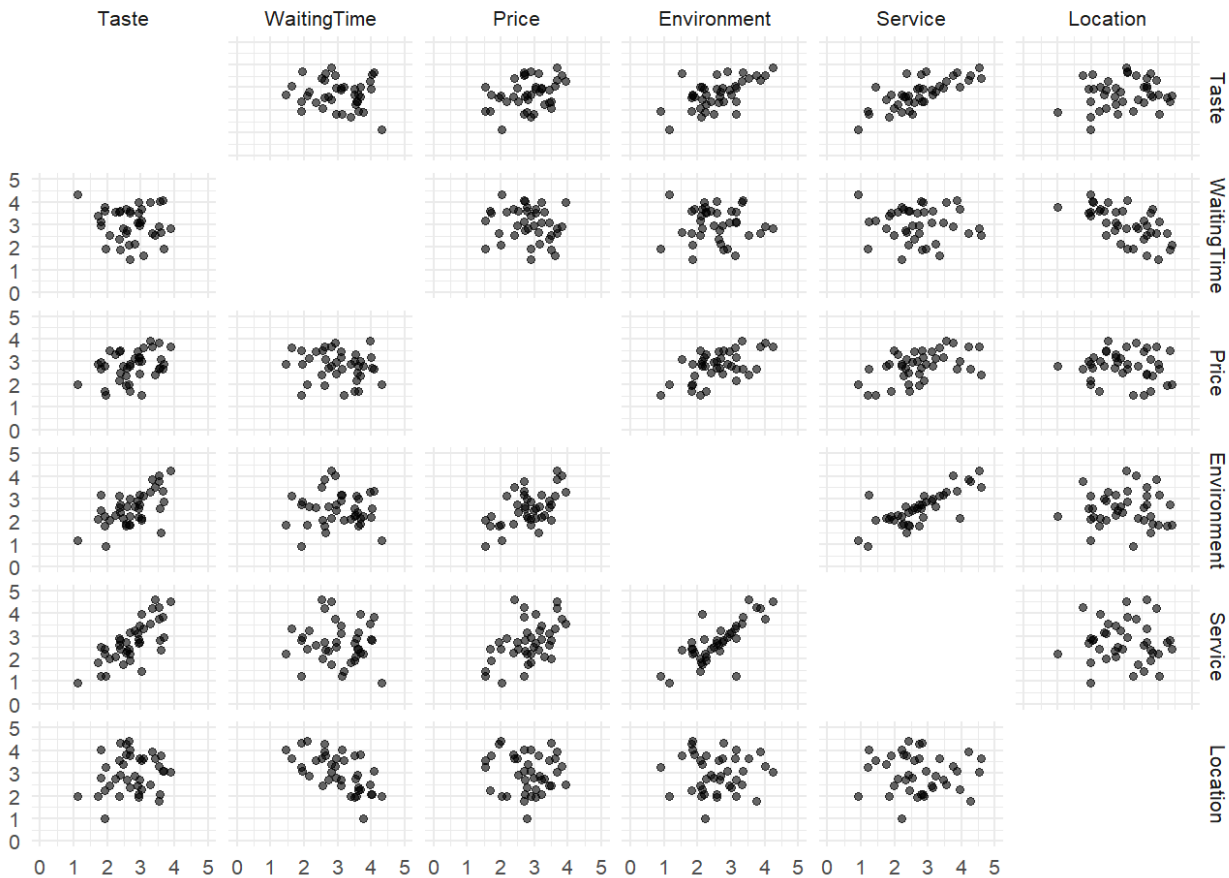


- Average of scores $\sim$ 3.8?
- *Data reduction*

Example – rating a restaurant

If you want to design a scoring algorithm, how would you weight different aspects?

- Environment might be highly correlated with service; location might be highly correlated with waiting time

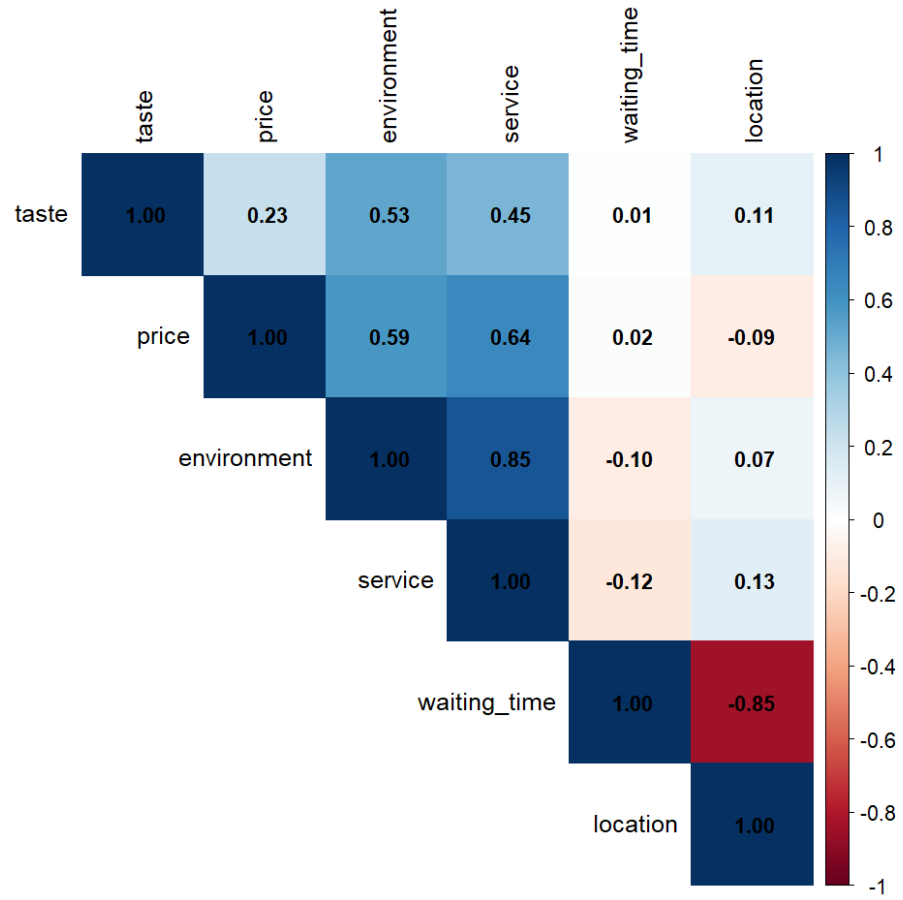


- *Dependence among variables can be complex*

Example – rating a restaurant

If you want to design a scoring algorithm, how would you weight different aspects?

- Environment might be highly correlated with service; location might be highly correlated with waiting time

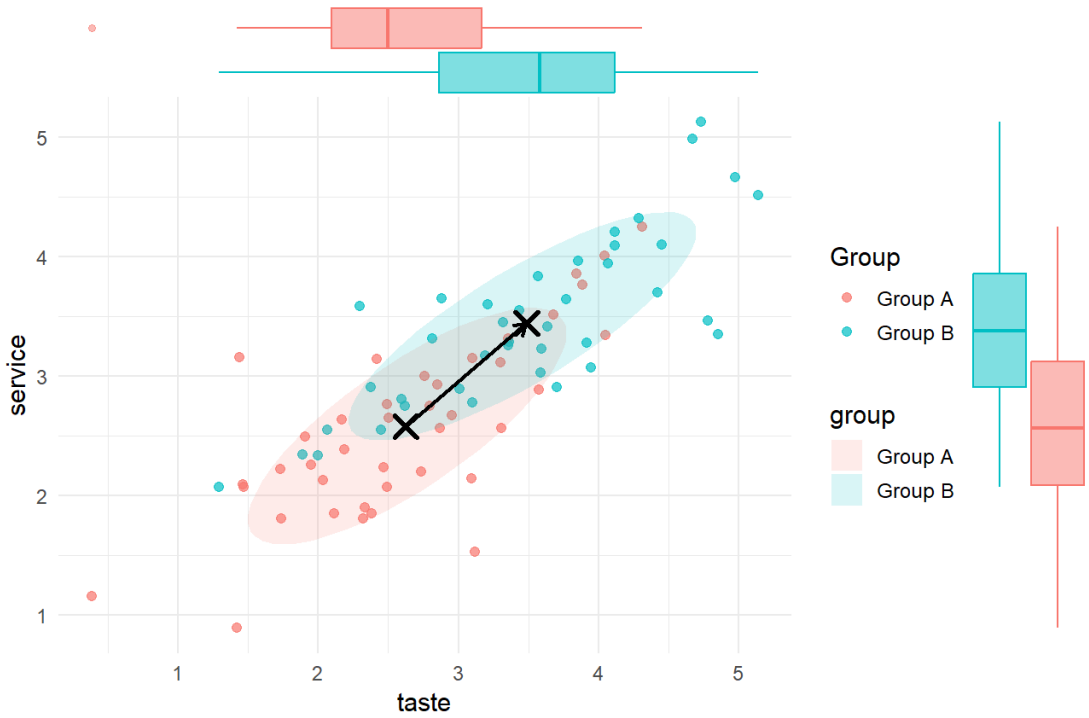


- *Dependence among variables can be complex*

Example – rating a restaurant

Compare restaurant ratings between two groups? (e.g. Chain vs Independent restaurants)

- For univariate case: two-sample t test.
- What if multivariate case? What if more than two groups?

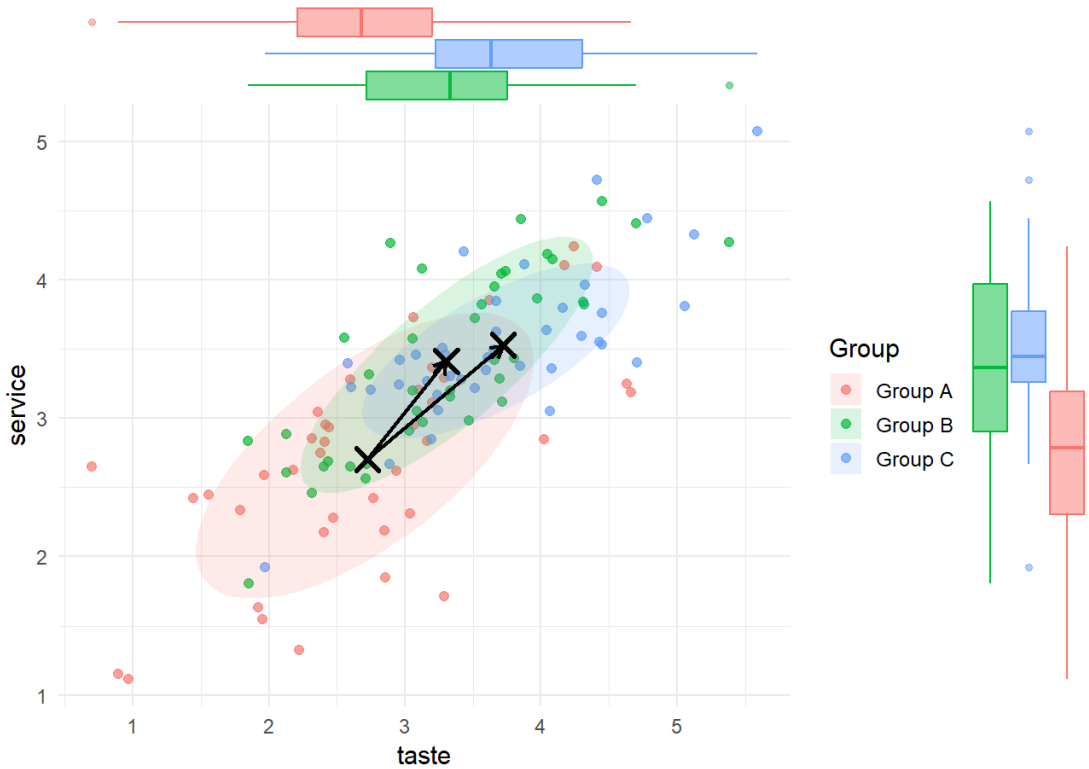


- *Hypothesis testing*

Example – rating a restaurant

Compare restaurant ratings between two groups? (e.g. Chain vs Independent restaurants)

- For univariate case: two-sample t test.
- What if multivariate case? What if more than two groups?

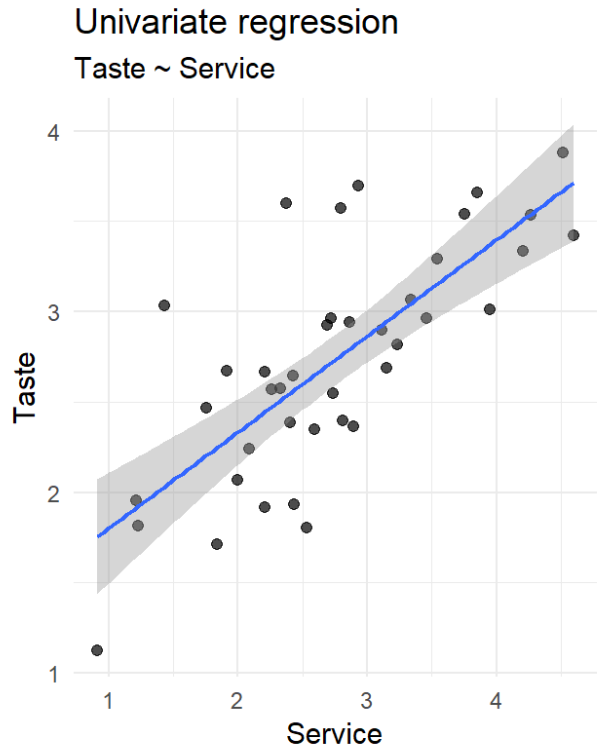


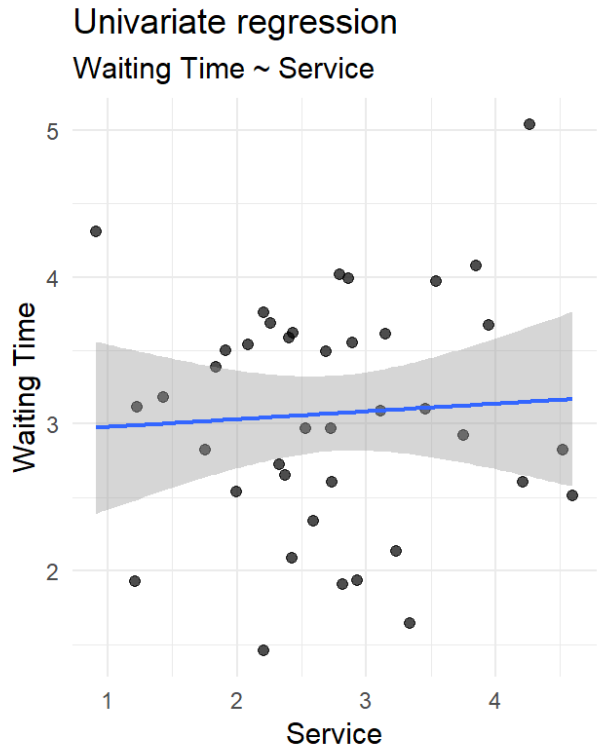
- *Hypothesis testing*

Example - rating a restaurant

Predict taste and waiting time using other variables?

- For univariate case: simple linear regression, multiple linear regression.



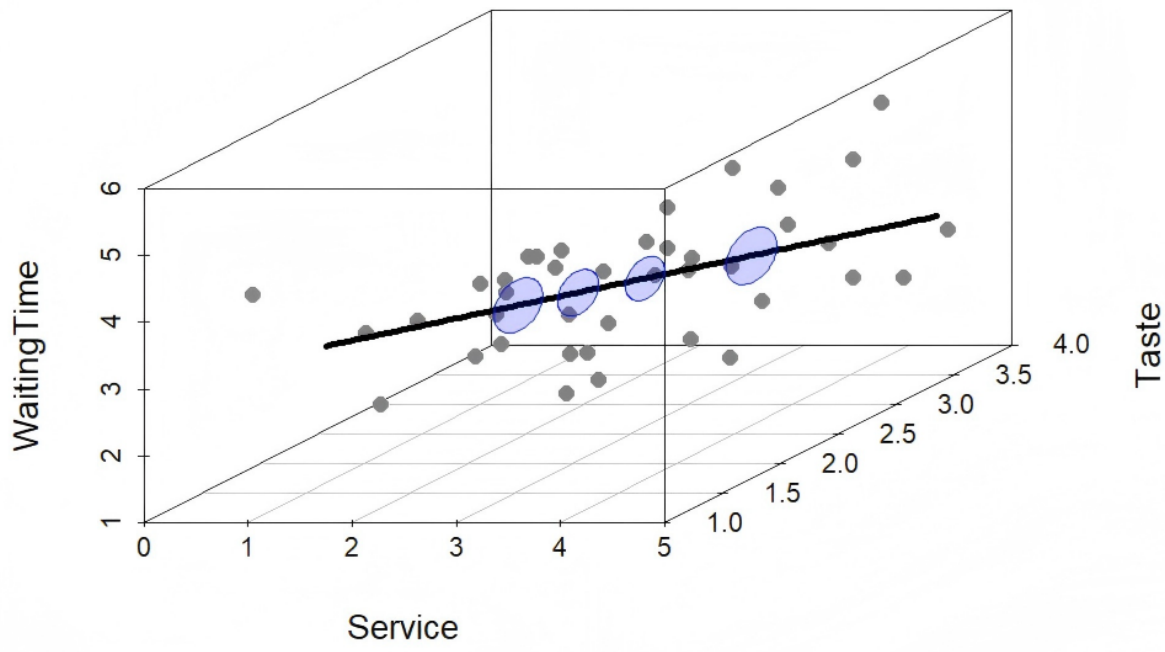


- What if we are interested in more than one response variable?
- *Regression & prediction*

Example - rating a restaurant

Predict taste and waiting time using other variables?

- For univariate case: simple linear regression, multiple linear regression.
- What if we are interested in more than one response variable?



- *Regression & prediction*

How to study STAT-445/645?

- Understand the basic ideas of the methods
  - know what we are doing
  - know when to use what procedures
  - able to implement the statistics methods via R
  - able to interpret the analysis results
  - know what conditions are needed for success, when the methods can fail
- Master the basic inference procedures
  - *review and master* the basic univariate statistics procedures studied in STAT-285 and STAT-270, or STAT-201 and STAT-302/305
  - such as: two-sample t test, ANOVA, simple linear regression

Don't fall behind.

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
- *Part II. Inference under Multivariate Normal Distribution (Textbook Chp 4-7)*
- *Part III. Commonly-Used Multivariate Analysis Methods (Textbook Chp 8-11)*
- *Part IV. Other Topics (Textbook Chp 12)*