

# Statistics 4310: Statistical Inference

Fall 2022

---

**Classes:** Section 001, Tue/Thu 10:15–11:45 a.m., JMHH F70  
Section 002, Tue/Thu 12:00–1:30 p.m., JMHH F70

**Instructor:** Zongming Ma  
**Email:** zongming@wharton.upenn.edu  
**Office hours:** Thu 3–4 p.m. or by appointment  
**Location:** ARB 421

**Teaching assistant:** Sagnik Nandy  
**Email:** sagnik@wharton.upenn.edu  
**Office hours:** Fri 2–4 p.m.  
**Location:** ARB 421

## Course Overview

The course aims to equip students with ideas and tools in statistics which range from the very beginning of the subject to an intermediate level. Together, we will examine a collection of basic concepts and commonly used methods, with an emphasis on the understanding of when and how to apply them, and why. Students will experiment the ideas on data examples using the statistical software R.

Topics include (1) collection, summary and visualization of data, (2) estimation, hypothesis testing, and confidence statements, and (3) simple and multiple linear regression. If time permits, we will also discuss likelihood based inference and other more advanced topics.

## Prerequisites

The official prerequisite of the course is STAT 4300. The effective prerequisite is *fluency* with basic probabilistic reasoning and analysis (e.g., probability distributions and densities; joint distributions; conditional probability, independence, correlation, and covariance; moment generating functions; law of large numbers; central limit theorem; etc.)

It would be helpful to have previous exposure to linear algebra, but it is *not required*. Previous exposure to the statistical computing software R is *not required*, either.

## Textbook

*Statistics and Data Analysis: from Elementary to Intermediate*, by A. C. Tamhane and D. D. Dunlop, Prentice Hall, 2000.

A few copies of the textbook is on reserve at the Lippincott Library.

## Course Website

The course website uses the **Canvas** platform. Please check the course website for announcements, handouts, sample codes, assignments, solutions, and other materials, etc.

## Statistical Computing Software

The statistical computing software R (version 4.1.0 or higher) will be used in the course. It is free, and can be downloaded at the R-project website:

`www.r-project.org`.

The above website also contains a list of manuals for using the software. **RStudio** (available at `www.rstudio.com`) can be a helpful IDE for R, the basic version of which is also free. Basic usage of R will be illustrated in class and through sample codes posted on course website. *No previous exposure to the software is required.*

## Homework Assignments

- There will be six assignments in total. Each assignment will be graded. *The best five assignment scores will be counted toward your final grade.*
- Homework assignments will be posted on course website. After the due dates, solutions will be posted.
- *No late homework will be accepted. No exception will be made.*
- Students can help each other on solving the problems, but are expected to prepare the final writeup individually with acknowledgment of the help received.

## Quizzes and Exams

- There will be four 10-minute quizzes given in class. The exact date of each quiz will be announced in class one week before each quiz takes place.
- Midterm exam: Thursday, October 20, in class.
- Final exam: TBA.
- Both exams will be open-book.

## Grading Policy

- Homework assignments: 20% (with the lowest score dropped);
- Quizzes: 10%
- Midterm exam: 35%;
- Final exam: 35%.

## **A Tentative List of Topics**

1. Introduction
2. Summarizing data – one variable
3. Summarizing data – multiple variables
4. Basic concepts of inference: point estimation and confidence intervals
5. Basic concepts of inference: hypothesis testing
6. Sampling distributions: normal and chi-square distributions
7. Sampling distributions: Student's t and F-distributions
8. One sample inferences: mean with known variance
9. One sample inferences: mean with unknown variance, and inference for variance
10. Two sample inferences: independent samples design
11. Two sample inferences: matched pair design
12. Simple regression: the least square regression line
13. Simple regression: statistical model and basic inferences
14. Simple regression: analysis of variance and prediction
15. Simple regression: regression diagnostics I
16. Simple regression: regression diagnostics II
17. Multiple regression: statistical model and basic inferences
18. Multiple regression: analysis of variance and prediction
19. Multiple regression: specially constructed predictor variables
20. Multiple regression: collinearity
21. Multiple regression diagnostics
22. Variable selection
23. Maximum likelihood estimation