



MA370 Introduction to Data Analysis (Online)

| | | | |
|------------------------------------|---|-----------------|---------|
| Instructor Information: | Puyu San Home Institution: Bridgewater State University Email: psan@bridgew.edu Office Hours: TBD | | |
| Term: | June 27, 2022 - July 22, 2022 | Credits: | 4 units |
| Course Delivery | The class will be delivered in the format of online. Other than recorded lecture videos, the instructor will arrange 3 hours' real-time interactions with students per week (via discussion forum, zoom meeting, and WeChat). The workload students are expected to complete to properly pass this course is about 15 hours per week. | | |
| Required Texts (with ISBN): | Introduction to the Practice of Statistics 10 th Edition, Moore, McCabe, Craig | | |
| Prerequisite: | Elementary Statistics | | |



Course Overview

This course is an introductory level course to data analysis. Course topics cover a variety of basic statistical analysis and data science, including analyzing the association between categorical variables, such as chi-square independence and goodness of fit tests and logistic regression, sampling and experimental design, linear and multiple regression and the corresponding statistical reference of the model parameters, One-way, two-way, and more-way ANOVA, and nonparametric statistics. In addition, students will learn to use the software R to produce, process, and clean data, and implement all the topics above with real-life examples.

Course Goals

Upon successful completion of this course, students will be conversant with

- inference about means, proportions, and categorical data
- general linear regression analysis, including simple and multiple regression
- inference about model parameters
- analysis of variance for one factor
- analysis of variance for two-factor
- introduction of the nonparametric statistics
- statistical software to analyze the real-life data

Software: R and RStudio

R is a commonly used statistical programming language, it is widely used in data analysis and data visualization. R is also a free, open-source software, which can be downloaded through <http://cran.r-project.org>.

We encourage R beginners to use RStudio, which is an Integrated Development Environment (IDE) for R. RStudio is more user friendly and will be taught for this course. The free version of RStudio can be downloaded from <http://www.rstudio.com/products/rstudio/download/>.

Course Structure

1. Asynchronous Hours: Monday through Friday, total 25 hours Pre-recorded videos will be posted on SJTU SCE online learning platform.
2. Synchronous Hours (Beijing Time):
 - a) Required Zoom meeting: Tuesday: 8 :00 – 9:00 pm
 - b) Optional Zoom office hour for this course: Friday: 9:00 – 10:00 am
 - c) Optional open office hour: Tuesday: 11:00 am – 12:00 pm
Tuesday: 11:00 am – 12 pm



Grading Policy

| | |
|---------------------|-----|
| Attendance | 20% |
| Assignment 1 | 15% |
| Assignment 2 | 15% |
| Midterm Exam | 25% |
| Final Group Project | 25% |

Grading Scale is as follows

| Number grade | Letter grade | GPA |
|--------------|--------------|-----|
| 90-100 | A | 4 |
| 85-89 | A- | 3.7 |
| 80-84 | B+ | 3.3 |
| 75-79 | B | 3 |
| 70-74 | B- | 2.7 |
| 67-69 | C+ | 2.3 |
| 65-66 | C | 2 |
| 62-64 | C- | 1.7 |
| 60-61 | D | 1 |
| ≤59 | F (Failure) | 0 |



Class Schedule

| Date | Lecture | Online Teaching Arrangement |
|--------|---|--|
| Day 1 | Review of Chapter 1 – 6 | Approximately 80 minutes pre-recorded video lectures |
| Day 2 | Chapter 7. Inference for Means | Approximately 30 minutes pre-recorded video lectures plus 60 minutes online interaction via Zoom |
| Day 3 | Chapter 7. Inference for Means | Approximately 80 minutes pre-recorded video lectures |
| Day 4 | Chapter 8. Inference for Proportions | Approximately 80 minutes pre-recorded video lectures |
| Day 5 | Chapter 9. Inference for Categorical Data | Approximately 80 minutes pre-recorded video lectures |
| Day 6 | Chapter 9. Inference for Categorical Data | Approximately 80 minutes pre-recorded video lectures |
| Day 7 | Chapter 10. Inference for Regression | Approximately 30 minutes pre-recorded video lectures plus 60 minutes online interaction via Zoom |
| Day 8 | Chapter 10. Inference for Regression | Approximately 80 minutes pre-recorded video lectures |
| Day 9 | Chapter 11. Multiple Regression | Approximately 80 minutes pre-recorded video lectures |
| Day 10 | Chapter 11. Multiple Regression & Midterm Exam Review | Approximately 80 minutes pre-recorded video lectures |
| Day 11 | Midterm Exam | Midterm Exam on Zoom |
| Day 12 | Chapter 12. One-Way Analysis of Variance | Approximately 30 minutes pre-recorded video lectures plus 60 minutes online interaction via Zoom |
| Day 13 | Chapter 12. One-Way Analysis of Variance | Approximately 80 minutes pre-recorded video lectures |
| Day 14 | Chapter 13. Two-Way Analysis of Variance | Approximately 80 minutes pre-recorded video lectures |



| | | |
|--------|---|--|
| Day 15 | Chapter 13. Two-Way Analysis of Variance | Approximately 80 minutes pre-recorded video lectures |
| Day 16 | Chapter 14. Logistic Regression | Approximately 80 minutes pre-recorded video lectures |
| Day 17 | Chapter 14. Logistic Regression | Approximately 30 minutes pre-recorded video lectures plus 60 minutes online interaction via Zoom |
| Day 18 | Chapter 15: Nonparametric Rank Tests | Approximately 80 minutes pre-recorded video lectures |
| Day 19 | Chapter 16: Bootstrap Methods and Permutation Tests | Approximately 80 minutes pre-recorded video lectures |
| Day 20 | Final Group Presentation | Group Presentation on Zoom |