Statistics for the Sciences

Introduction

Xuemao Zhang
East Stroudsburg University

January 18, 2025

What's covered in this lecture?

- Math 210 Course Outline
  - Course Objectives
  - Tentative Contents
  - Assessments
  - References
- Introduction to IBM SPSS

Math 210 Course Outline

Course Admin Information

- Instructor: Dr. Xuemao Zhang
  - Office: SciTech Rm 128
  - Email: xzhang2@esu.edu
- Lecture Hours:
  - MWF: 10:00–10:50Am
- Department Secretary: Christine Getz
  - Office: SciTech Rm 118
  - Email: cgetz@esu.edu
  - Telephone: 570-422-3447

Course Objectives

- This course helps you develop critical statistical literacy skills essential for interpreting and analyzing scientific research findings.
- **Software** IBM SPSS Statistics
- **You will learn:**
  - Probability Distributions
  - Hypothesis Testing and Confidence Intervals
  - Design of Experiments and ANOVA
  - Regression Models
  - Generalized Linear Models
  - Multivariate Analysis and Dimension Reduction
- **Prerequisites:** Elementary Statistics (Math 110 or similar courses).

Tentative Contentes

Topics covered include:

- Probability Distributions:
  - Some typical discrete probability distributions like discrete uniform, Binomial, Poisson, Geometric and hyper-geometric distributions
  - Some typical continuous probability distributions like uniform, normal, student-t, Chi-square and F distributions
- Hypothesis Testing and Confidence Intervals
  - Confidence interval estimation of a population mean
  - t-test and z-test of a population mean
- Design of Experiments and ANOVA
  - ANOVA for experiments with one factor
  - ANOVA for experiments with two factors including block design, factorial design, nested design and split-plot design.
  - Analysis of hierarchical or clustered data with Linear Mixed Models

Tentative Contentes

- Simple Linear Regression Models
  - Visualization of bivariate data and correlation analysis
  - Model fit and inferences for bivariate data using simple linear regression models
- Multiple Linear Regression Models
  - Model fit and inferences using multiple linear regression models
  - Criteria for Model Selection: Adjusted R-squared, Mallows' Cp, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) etc.
  - Automatic search for model selection using software
  - Generalized Additive Models (GAM) and Generalized Additive Mixed Models (GAMM)

Tentative Contentes

- Generalized Linear Models
  - Logistic regression for binary/binomial data
  - Multinomial logistic regression models
  - Poisson regression for count data
  - Negative binomial regression for over-dispersed count data
- Classification using
  - Logistic regression and multinomial logistic regression
  - Linear Discriminant Analysis (LDA)
  - k-Nearest Neighbors (kNN)

Tentative Contentes

- Multivariate Analysis and Dimension Reduction Methods
  - Principal Component Analysis (PCA)
  - Factor analysis
  - Multidimensional Scaling (MDS) and Nonmetric Multidimensional Scaling (NDMS) to explore similarities among a set of objects
  - Correspondence Analysis (CA) to explore the relationships between two sets of variables
  - Canonical Correlation Analysis (CCA) to explore the relationships between two sets of variables
  - Principal Component Regression (PCR)
  - Partial Least Squares (PLS)

Assessments

- 5% Class attendance
- 30% Homework assignments (6 sets)
- 45% Exams (3 sets)
- 20% Final Exam

References

- *Textbook*: Experimental Design and Data Analysis for Biologists (2nd Edition), Gerry P. Quinn and Michael J. Keough, 2023. Cambridge University Press. https://qkstats.com/
  - Software used is programming language R.
  - SPSS will be used for data analysis in this course.

- John H. McDonald (2014). *Handbook of Biological Statistics (3rd Edition)*, https://www.biostathandbook.com/HandbookBioStatThird.pdf
- Josafhat Salinas Ruíz (2023). *Generalized Linear Mixed Models with Applications in Agriculture and Biology*, https://library.oapen.org/handle/20.500.12657/76233
- Lang WU and Jin QIU (2021). Applied Multivariate Statistical Analysis and Related Topics with R,Available at Kemp ESU.
- James, G., Witten, D., Hastie, T. and Tibshirani, R.(2013). *An Introduction to Statistical Learning: with Applications in R*. Springer. https://www.statlearning.com/

SPSS

- IBM SPSS is used for statistical analysis, data management, and data mining.
- IBM SPSS products
  - IBM SPSS Statistics
  - IBM SPSS Modeler: A data mining and text analytics software application with machine learning models
- IBM SPSS Statistics is a comprehensive software for managing, analyzing, and visualizing data.
  - Descriptive statistics
  - Bivariate statistics
  - Predictive analytics including confidence intervals and hypothesis testing
  - Advanced statistical procedures perform complex analyses such as regression, ANOVA, factor analysis and other mutivariate analysis.

SPSS

- Installation: go to spss statistics
  https://www.ibm.com/products/spss-statistics

License