SOC 6003: Advanced Statistical Analysis
2nd term, 2023-24
Mondays 11:30 – 2:15 pm
Sino Building 429

Instructor: Jin Lei
e-mail: ljin@cuhk.edu.hk

Course Assistant: TBC
e-mail: lanyaxin@link.cuhk.edu.hk

Assessment: Assignments 30%, Quiz 30%, term paper 40%

Plagiarism

Attention is drawn to University policy and regulations on honesty in academic work, and to the disciplinary guidelines and procedures applicable to breaches of such policy and regulations. Details may be found at http://www.cuhk.edu.hk/policy/academichonesty/ . With each assignment, students will be required to submit a statement that they are aware of these policies, regulations, guidelines and procedures.

Course Outline

1   Basic concepts
    1.1   The research process
    1.2   Measurement
    1.3   Descriptive statistics
    1.4   Inferential statistics

Knoke et al. 2002. Chapters 1-3; Freund, Wilson & Sa. 2006. Chapter 1: 5-17.

2   Multiple regression: Basics
    2.1   Model specification
    2.2   Partial effects
    2.3   Standardized coefficients
    2.4   Estimation method
    2.5   Model evaluation

Allison. 1999. Chapters 2, 4 and 5; Freund, Wilson & Sa. 2006. Chapters 2 – 3

3   Multiple regression: Techniques
    3.1   Modeling group differences
    3.2   Modeling interaction effects
    3.3   Modeling nonlinear effects
    3.4   Mediation analysis

Allison. 1999. Chapter 8; Kahane. 2001. Chapter 5; MacKinnon. 2008. Chapters 1 – 3

4   Multiple regression: Diagnostics
   4.1   Regression assumptions
   4.2   Detection of assumption violation
   4.3   Remedies of assumption violation
   4.4   Multicolinearity
   4.5   Robustness to influential cases

Alison. 1998. Chapters 6 – 7; Freund, Wilson & Sa. 2006. Chapters 4 – 5

5   Logistic regression
   5.1   Binary dependent variables
   5.2   The logic of logistic regressions
   5.3   Probability, odds, and log odds
   5.4   Interpreting coefficients
   5.5   Estimation method
   5.6   Model evaluation
   5.8   Extensions

Pampel. 2000. Chapters 1-3; O'Connell. 2005. Chapter 4

6   Poisson regression
   6.1   Analyzing count data
   6.2   Estimation method
   6.3   Goodness of fit measures

Kleinbaum et al. 1998. Chapter 24: 687-709.

7   Fixed and random-effect models
   7.1   Pooled data and panel data
   7.2   Fixed-effect model for linear outcomes: two-period data
   7.3   Fixed-effect model for linear outcomes: multiple repeated measures
   7.3   Random-effect model for linear outcomes

Allison 2009. Chapters 1-3; Fitzmaurice et al. 2011. Chapter 14.

References

Allison, P.D. 1998. *Multiple Regression: A Primer*. Newbury Park, CA: Sage.

Allison, P.D. 2009. *Fixed Effects Regression Models*. Los Angeles: Sage.

Fitzmaurice, GM., NM. Laird, and JH. Ware. 2011. *Applied Longitudinal Analysis*. Hoboken, N.J: Wiley. (Available online)

Freund, R.J., Wilson, W.J., and Sa, P. 2006. *Regression Analysis: Statistical Modeling of a Response Variable*, 2nd edition. Burlington, MA: Academic Press. (Available online)

Kahane, L.H. 2001. *Regression Basics*. CA: Sage.

Kleinbaum, D.G., Kupper, L.L., Muller, K.E., and Nizam, A. 1998. *Applied Regression Analysis and Other Multivariable Methods*, 3rd edition. Duxbury Press.

Knoke, D, GW. Bohrnstedt, and AP Mee. 2002. *Statistics for Social Data Analysis*. Itasca, Ill: Wadsworth Publishing.

MacKinnon, D.P. 2008. *Introduction to Statistical Mediation Analysis*. New York: Lawrence Erlbaum Associates. (Available online)

O'Connell, AA. 2005. *Logistic Regression Models for Ordinal Response Variables*. 1 edition. Thousand Oaks, Calif: SAGE Publications.

Pampel, FC. 2000. *Logistic Regression: A Primer.* Thousand Oaks, CA: Sage.

Some books available online through the UL; other books are on reserve at the UL.

Class schedule

Date		Topics	Assignments due
January	10	1. Basic concepts	
	17	2. Multiple regression: Basics	
	24	2. Multiple regression: Basics	Exercise 1
	31	No class: Chinese New Year	
February	7	3. Multiple regression: Techniques	
	14	3. Multiple regression: Techniques	
	21	3. Multiple regression: Techniques	Exercise 2
	28	4. Multiple regression: Diagnostics	
March	7	5. Logistic regression	
	14	5. Logistic regression	Exercise 3
	21	6. Poisson regression	
	28	7. Fixed and random-effect models	Exercise 4
April	4	7. Fixed and random-effect models	
	11	Quiz	
	30		**Term paper**

Assessment

Assignments 30%, Quiz 30%, Term paper 40%

Grade	Descriptors
A	Outstanding performance on all learning outcomes.
A-	Generally outstanding performance on all (or almost all) learning outcomes.
B	Substantial performance on all learning outcomes, OR high performance on some learning outcomes which compensates for less satisfactory performance on others, resulting in overall substantial performance.
C	Satisfactory performance on the majority of learning outcomes, possibly with a few weaknesses.
D	Barely satisfactory performance on a number of learning outcomes.
F	Unsatisfactory performance on a number of learning outcomes, OR failure to meet specified assessment requirement.

Term Paper
Due: Apr. 30, 2021

Students are required to conduct a statistical analysis and write a research paper similar to the format of articles published in the *American Sociological Review (ASR)*. You are encouraged to read a few *ASR* articles in order to familiar yourself with conducting statistical analyses and writing quantitative research papers. You have to choose your own data for this exercise. The structure of the paper and the length of each section depend on the nature of the research topic. The following format provides a rough guideline for your reference.

1. **Introduction** (1 to 2 pages): Motivate your readers
   What is your research question?
   Is the question important either in terms of theoretical significance or policy implications?
   What is the nature of your study: A replication or a test of new theories?

2. **Literature review** (2 pages): Discuss the conceptual framework
   Review at least 5 articles related to your research topic.
   Identify research tradition(s) as the starting point of your own research.
   What are the unanswered questions in previous research?
   Does your research extend analyses of previous studies?

3. **Model and hypotheses** (1 page): Focus your study on specific issues
   Construct your own model based on previous research evidence and your own speculations. A model is simply a set of statements describing how theoretical concepts are linked together (include path diagrams if appropriate).
   Formally state the hypotheses that you are going to test in the paper.

4. **Data and variables** (1 page)
   Describe data source: when it was collected, method of data collection, no. of cases ...
   Define each variable used in the analysis (discuss measurement issues, e.g. scale, coding ...)
   What are the concepts that these variables are supposed to measure? Give justification.

5. **Method** (1 to 2 pages)
   Describe the method of analysis.
   Is the method appropriate?
   Are there any methodological issues that we need to pay attention to (i.e. outliers, linearity, and multicollinearity)? Describe the steps that you have taken to check and correct these potential problems?

6. **Findings** (2 pages)
   Report the results from your analysis. Report descriptive statistics first and then the results from your statistical models.

7. **Summary and conclusion** (1 to 2 pages)
   Briefly summarize key findings.
   Discuss implications of the findings to your model and theory.
   Are there any unanswered questions?
   Discuss implications for future research or policy recommendation?