SOC 1029

305 12110

Advanced Social Statistics

2026 Spring

National Taiwan University

Class: Friday 2, 3, 4, 6, 7 (6 & 7 are practicum)

Social Science Building Room 305 for 2, 3, 4; Social Science Building Room 302 for 6, 7 (Note: NOT Sociology and Social Work Building)

Expected weekly time commitment (outside of class): 8 hours/week

Instructor: Meng-Jung Lin 林孟瑢 mjlinmj@ntu.edu.tw

Office hours:

Mondays 4:00-5:00 PM (only by appointment)

Fridays 3:30-5:00 PM in R417 or online via Google Meet (please sign-up using Calendly: https://calendly.com/mjlinmj/15min)

Teaching Assistants:

Yu-Shen Lin 林譽神 r11626022@ntu.edu.tw

TBA

Does social inequality transmit from parents to children? Does the effect of education on income differ by gender? How do adverse childhood experiences affect adulthood outcomes? Extending the materials we covered last semester, we will use statistics and programming to answer more advanced and complex questions this semester. This course introduces tools to test the assumptions of the regression model, take into account categorical variables and interaction effects in regression analysis, deal with multicollinearity, treat categorical variables as outcomes, and analyze panel data. Applying statistics, analyzing data, and interpreting results are the focuses of our class. Although very basic calculation skills (e.g., +, -, ×, ÷, √) and knowledge of hypothesis testing and bivariate linear regression are required, you do not need further mathematics knowledge to be successful in this class. The prerequisite of this course is SOC 1028 Social Statistics or equivalent.

Goals of this course

After taking this course, you are expected to be able to:

1. Explain key statistical concepts in your own words.
2. Analyze real life data, including cross-sectional and longitudinal data, using R

programming.
3. Interpret results from the statistical models covered in class.
4. Apply statistical methods and computer skills to address daily and social issues.
5. Evaluate statistics and statistical method used in academic research.
We will try to accomplish these together. Sometimes it can be hard to catch up with the class if you miss a class or misunderstand just one concept. You have to let me know if you encounter problems along the way. I will enroll you to the Piazza platform for this class (https://piazza.com/ntu.edu.tw/spring2026/soc1029) soon after the first class. Please feel free to ask questions on the platform (and yes, you can do it anonymously).

Who should take this course?
This is a required course for SOCI majors. For those of other majors, I encourage you to join us if you are:

1. Familiar with bivariate regression analysis and would like to enrich your statistics toolkits.
2. Interested in applying appropriate statistic methods to understand social phenomena.
3. Willing to spend time outside of classroom to figure out how statistics and R work.
4. Considering to be a data analyst or go to graduate school.

Course Requirements
Practicum Assignments and Discussion (50%): There will be 9 assignments distributed on Fridays during our practicum session and due on the following **Friday at 9:10 AM**. In the NTU COOL calendar, you can find the exact distribution dates and due dates of the assignments. Usually, two or three problems are included in each assignment. Textbook and R examples in class are the main sources of the problems. For us to help, you must submit your **hand calculating procedures for by-hand problems**, and **programming codes and outputs for R problems**. You can collaborate with others to understand the concepts and work through the conceptual programming procedures, but the submitted assignments must be your own works. **Each assignment is worth 50 points, except for Assignment 9, which is worth 100 points.** You are required to contact me beforehand if you want to hand in an assignment late, or I may not accept it. To support learning and to ensure you understand your own work, I may occasionally ask you to submit a **short audio reflection** with an assignment. This reflection will typically be **3-5 minutes** and should explain what you did, what you found, and how you interpreted the key output **in your own words**.

As you are required to do a research poster by the end of the semester, **you must meet with me at least once during our practicum or/and my office hours** to discuss the progress of your research and data analysis. 10 points will be deducted if you failed to meet with me. Please develop an empirically testable research question and apply statistical methods

learned in this class to a public dataset (or a dataset of your own) to answer your question. **A few questions in the assignments will help you develop your research.** You are encouraged to ask for assistance regarding any aspects (e.g., literature, data acquisition, data analysis, etc.) of your study.

Quizzes (3%): Quiz with 1 or 2 short question(s) for the readings of that week will be available in NTU COOL on Wednesday at 9:10 AM. The due time is **Friday at 9:10 AM**. Quizzes are open book and open notes, but you are not allowed to discuss with others. Each quiz is worth 5 points.

Exams (12%): There will be 2 **in-class** exams throughout the semester. You will be asked to answer multiple-choice questions, summarize key concepts in your own words, and use R to answer programming questions. You will have to answer the multiple-choice questions in NTU COOL. The rest of the questions will be available in a Word (and PDF) file in the same NTU COOL exam session. Be sure to attach your answer file when you submit your exams. You have to take the exams at the designated time. All of them are open book and open notes, but you have to finish the exam **within the class period**. Remember to have your laptop ready for the exam. Makeup exams are available if you can show me the official proof, but you will have a different version of the exam which may or may not be harder.

Final Research Poster and Presentation (30%): You are required to **come to class** on the exam day. The default format is an in-class presentation. You will prepare a presentation for a research study on any topic of your choice and give a **10-minute presentation in class**. Showing a recording of your presentation during the exam is not allowed. You will then **finish a worksheet/questionnaire** asking you to comment on at least two presentations and write a reflection of the posters/presentations you saw. We will use several class periods throughout the semester to discuss your research (or research poster) in groups. If the class size makes in-class presentation infeasible, I will switch to an online submission format. In that case, you will be asked to **submit an electronic poster for a research study on any topic of your choice and an accompanying 3-5 minute recorded presentation by the end of the day before the exam time,** and you will complete the same **peer-review worksheet/questionnaire**. You have to have your **laptop and headphones** ready for the final exam which will also be available in COOL. The exam will be open book and open notes, but you have to finish it on time. Makeup exam is available if you can show me the official proof, but you will have a different version of the exam which may or may not be harder.

Attendance (5%): Every class (including practicum) counts. For in-person sessions, I will use **Zuvio** (https://www.zuvio.com.tw/) to take attendance, so please install the App and check-

in every time you come to class. For online sessions (if necessary), I will use the Google Meet Attendance List to take attendance. Your **participation with Zuvio questions** and **worksheet submissions** will also be used to verify your attendance. Points will not be deducted if you answered the questions wrong. I will only use them to see how well you understood the materials. Worksheets will be distributed in almost every class, but you will only be asked to submit them in select classes. I will announce at the beginning of class whether the worksheet for that day needs to be submitted. You may miss 1 week of classes without penalty. **Please contact me for additional absences.** For those who have time conflicts, please let me know beforehand. Class recordings will be posted on NTU COOL. If you miss a class, you must watch the recording before the next class. The NTU COOL video manager will document your name and the time spent watching the recording, which will be counted toward your attendance.

Attendance in the first class is required, because adding a course but not showing up can limit other students' opportunities to enroll, and the first class covers core course policies and setup that you are expected to follow throughout the semester. If you do not appear in the first class and do not contact me in advance, I will **deduct 20% of your final course grade**.

MJ doesn't mind if you occasionally miss class, but MJ does care if you:

1. **Skip classes due to mental or physical illness or emergencies.**
2. **Claim that MJ didn't cover material that appears on assignments or exams.**
3. **Realize too late in the semester that you might not pass or achieve a good grade.**

If you understand all of the above but decide not to attend class on a particular day, you can email me, saying you're tired or just need more sleep. I completely understand and won't count it as an absence, <u>provided your absentee rate stays within the school policy limits</u>. Thank you.

Statistical Software

We will set the open-source statistical software **R** up in our first class. You might have heard researchers using SPSS, STATA, or SAS. These are convenient but expensive software packages. **R** is free, flexible, powerful, and widely used in many fields (from medical, data science, to social sciences). Although R can be challenging, it opens the door to many career options when you master it.

Textbook

You are required to read the designated chapters/sections of the following textbooks **BEFORE** each class. No new purchase is required if you were with us last semester as I will

post the readings other than the Healey one in COOL at least a week ahead before class.

Healey, Joseph F. 2020. Statistics: A Tool for Social Research & Data Analysis. 11th Edition. Cengage Learning.

Studenmund, Arnold H. 2016. Using Econometrics: A Practical Guide. 7th Edition. Pearson.

Gordon, Rachel A. 2012. Applied Statistics for the Social and Health Sciences. 1st Edition. Routledge.

Allison, Paul D. 2001. Missing Data. Sage Publications.

Luke, Douglas A. 2004. Multilevel Modeling. Sage Publications.

Wooldridge, Jeffrey M. 2019. Introductory Econometrics: A Modern Approach. Cengage Learning.

Class Policy

1. We will be spending much time on doing exercises and analyzing data in class. You **MUST read the assigned readings BEFORE** class so we can use the concepts and interpret the outputs from R. **Both quizzes and assignments are based upon the readings**. In addition, only will you know what **you and R** are doing after finished the assigned reading.
2. **Respect others and be responsible.**
3. Use your laptop during class, so we can use R to analyze data together. You can use smartphones to answer Zuvio questions if that is easier.
4. You are welcome to use AI tools to check concepts and aid your learning. However, relying on AI-generated content without critical evaluation may negatively impact your learning and performance. Submitting AI-generated responses without modification or using AI to bypass required coursework is not acceptable. If you use AI in your coursework, you must **disclose the AI platform/tool and briefly describe how you used it**. If the AI use affected your writing or analysis beyond simple proofreading, you must also include the prompts you used (or a short prompt log) in an appendix. You remain responsible for the accuracy of your code, results, and interpretation, and you should not input personal or sensitive data into AI tools.
5. Check your NTU COOL site daily: check the assignments and quizzes pages. Assignments and quizzes will be available in NTU COOL and should be submitted through COOL.
6. Check your email account daily: Any changes to the course schedule will also be announced in COOL and through notification.
7. I will try to reply to your emails within 24 hours during weekdays (Monday thru Friday). I usually answer them between 9 AM to 5 PM, so please arrange your time accordingly to ensure that I have enough time to get back to you before the deadlines.
8. Please be prepared before coming to the office hours.

Grading Policy

NTU uses letter grades that reflect how well course learning goals are achieved. In this course, A+ indicates that course goals are achieved and performance clearly exceeds expectations, A indicates that course goals are achieved, and A- indicates that course goals are achieved but there remains room for improvement. Below are the reference ranges for raw scores and letter grades for this course:

Letter grade	Raw score range
A+	90 - 100
A	85 - 89
A-	80 - 84
B+	77 - 79
B	73 - 76
B-	70 - 72
C+	67 - 69
C	63 - 66
C-	60 - 62
F	Below 59
X	0

If more than 30% of enrolled students have a final course score greater than 90, then only the top 30% (by final course score) will receive A+.

Honor Code

You have to complete all assignments, quizzes, and exams independently. I will make a checkbox available to you to indicate whether you do the work by yourself when you submit you works. You can familiarize yourself with the **NTU Honor Code here**. The Honor Code of the University is in effect at all times, and the submission of work signifies understanding and acceptance of those requirements. Plagiarism and the use of online educational support platform (e.g., Chegg) will not be tolerated.

While AI tools can support learning, using them to generate assignments, bypass required readings, or replace original thought violates academic integrity. Misrepresentation of AI work as your own may be considered academic misconduct.

Please consult with me if you have any questions about the Honor Code.

Accessibility Resources

Please contact me if you need accommodations due to disabilities, chronic medical conditions, a temporary disability or pregnancy complications resulting in barriers to fully accessing the course. You may receive extensions to your exams or/and assignments, and you may be allowed to make-up your absences by watching recordings. We can negotiate

about your accommodations depending on your circumstances.

Counseling and Psychological Services

The NTU Student Counseling Center is strongly committed to help students with mental health problems and psychological well-being needs through consultation and connection to clinically appropriate services. Go to their website: https://scc_osa.ntu.edu.tw/ or visit their facilities at Downtown Campus College of Medicine Area C (R204-1) to learn more.

Class Schedule

Week	Date	Topic	Readings	QOC*	Assignment	Quiz	Keywords
1	2/27	(NO CLASS) 228 Peace Memorial Day					
2	3/06	Descriptive Statistics, Hypothesis Testing, & Association/Correlation Review	Healey Ch.1-Ch.13		1dis		
3	3/13	Elaborating Bivariate Tables	Healey Ch.14	SP	1due/2dis	1due	partial tables, direct relationships, spurious or intervening relationships, interaction, gamma
4	3/20	Multiple Regression	Healey Ch.15	PA	2due/3dis	2due	control variables, partial correlation coefficient (r), partial slopes, standardized partial slopes (beta-weights)
5	3/27	Assumptions of OLS and the Central Limit Theorem	Studenmund Ch.1, 4, 5 (Gordon Ch.8.2-8.4.6)	CT	3due/4dis	3due	theoretical regression equation, estimated regression equation, stochastic error term, residual, estimator, BLUE, unbiased, efficiency, Gauss-Markov Theorem, sampling distribution of $\hat{\beta}$, SE($\hat{\beta}$), t-test of single coefficient, F-test (joint hypotheses testing)
6	4/03	(NO CLASS) Spring Break					
7	4/10	Exam 1	Healey Ch.1-15		4due		
8	4/17	Hypothesis Testing Dummy Variables	Studenmund Ch.5 Gordon Ch.10	DM	5dis	4due	sampling distribution of $\hat{\beta}$, SE($\hat{\beta}$), t-test of single coefficient, F-test (joint hypotheses testing) dummy variable, reference category, (linear combination)
9	4/24	Interaction terms & Nonlinear Relationships	Gordon Ch.11, 12		5due/6dis	5due	Interaction, the magnitude of the effect, curvilinear, logarithmic transformation, log-lin, quadratic term
10	5/01	(NO CLASS) Labor Day					
11	5/08	Outliers, Heteroskedasticity, & Multicollinearity Models for Binary Outcomes: Logistic (and Probit) Models	Gordon Ch.14 Studenmund Ch.13		6due/7dis	6due	outlier, influential observation, hat values (leverage), studentized residual (standardized residual), Cook's distance, DFBETA, DFFITS, heteroskedasticity, the Breusch-Pagan Test, robust standard errors, heteroskedasticity, the Breusch-Pagan Test, robust standard errors multicollinearity, variance inflation factor (VIF) linear probability model, logistic regression, log odds, logit, logit link, maximum likelihood
12	5/15	Models for Binary Outcomes: Logistic (and Probit) Models Multinomial Models & Ordered Logit Models	Studenmund Ch.13 Agresti pgs. 480-492		7due/8dis		linear probability model, logistic regression, log odds, logit, logit link, maximum likelihood fit statistics (likelihood ratio test, AIC, BIC), (optional: ROC curve and AUC, sensitivity and specificity)
13	5/22	Exam 2	Week 1 to Week 11				
14	5/29	Multinomial Models & Ordered Logit Models Specification Error and Omitted Variable Bias (Instrumental Variables and Causality)	Agresti pgs. 480-492 Studenmund Ch.6 Gordon Ch.13.3 (Wooldridge Ch.15.1-3)		8due/9dis		multinomial logit model, baseline-category logit specification, omitted variable bias instrumental variable regression (IV), two-stage least squares (TSLS or 2SLS)
15	6/05	(NO CLASS) Review and Final Research Poster Preparation Or Final Research Poster Presentation and Peer Review			9due		
16	6/12	Final Research Poster Presentation and Peer Review					

*QOC: Question of the Class.

Practicum Schedule

Week	Date	Topic	Readings	QOC*	Assignment	Keywords
1	2/27	(NO CLASS) 228 Peace Memorial Day				
2	3/06	Intro to Research Method and Survey Design	Ch.1-Ch.13		1dis	
3	3/13	Data Management[a]	Ch.14		1due/2dis	partial tables, direct relationships, spurious or intervening relationships, interaction, gamma
4	3/20	Multiple Regression	Ch.15	PA	2due/3dis	control variables, partial correlation coefficient (r), partial slopes, standardized partial slopes (beta-weights) multiple correlation coefficient (R), multiple regression, coefficient of multiple determination (R2)
5	3/27	Assumptions of OLS and the Central Limit Theorem	Studenmund Ch.1, 4, 5 (Gordon Ch.8.2-8.4.6)	CT	3due/4dis	theoretical regression equation, estimated regression equation, stochastic error term, residual, estimator, BLUE, unbiased, efficiency, Gauss-Markov Theorem, sampling distribution of $\hat{\beta}$, SE($\hat{\beta}$), t-test of single coefficient, F-test (joint hypotheses testing)
6	4/03	(NO CLASS) Spring Break				
7	4/10	Project discussion 1[a]				
8	4/17	Hypothesis Testing Dummy Variables	Studenmund Ch.5 Gordon Ch.10	DM	4due/5dis	sampling distribution of $\hat{\beta}$, SE($\hat{\beta}$), t-test of single coefficient, F-test (joint hypotheses testing) dummy variable, reference category, (linear combination)
9	4/24	Interaction terms & Nonlinear Relationships	Gordon Ch.11, 12	DM	5due/6dis	Interaction, the magnitude of the effect, curvilinear, logarithmic transformation, log-lin, quadratic term
10	5/01	NO CLASS				
11	5/08	Project discussion 2[a] Outliers, Heteroskedasticity, & Multicollinearity Models for Binary Outcomes: Logistic (and Probit) Models	 Gordon Ch.14 Studenmund Ch.13	GC	6due/7dis	outlier, influential observation, hat values (leverage), studentized residual (standardized residual), Cook's distance, DFBETA, DFFITS, heteroskedasticity, the Breusch-Pagan Test, robust standard errors, heteroskedasticity, the Breusch-Pagan Test, robust standard errors multicollinearity, variance inflation factor (VIF) linear probability model, logistic regression, log odds, logit, logit link, maximum likelihood
12	5/15	Project discussion 3[a] Multinomial Models & Ordered Logit Models	 Gordon Ch.17.1-17.2 Agresti pgs. 480-492	 PM	7due/8dis	linear probability model, logistic regression, log odds, logit, logit link, maximum likelihood fit statistics (likelihood ratio test, AIC, BIC), (optional: ROC curve and AUC, sensitivity and specificity)
13	5/22	Exam 2				
14	5/29	Multinomial Models & Ordered Logit Models Specification Error and Omitted Variable Bias, (Instrumental Variables and Causality)	Agresti pgs. 480-492 Studenmund Ch.6 Gordon Ch.13.3 (Wooldridge Ch.15.1-3)	ML DEI	8due/9dis	multinomial logit model, baseline-category logit specification, omitted variable bias instrumental variable regression (IV), two-stage least squares (TSLS or 2SLS)
15	6/05	(NO CLASS) Review and Final Research Poster Preparation Or Final Research Poster Presentation and Peer Review			9due	
16	6/12	Final Presentation and Peer Review (No Practicum)	Cumulative Final			

*QOC: Question of the Class. a: TA will present related works.

***Question of the Class (QOC) (i.e., R_lab):**

1. SP: Subjective Social Position and Attitudes on Economic Inequality
2. PA: Education, Income, and Physical Attractiveness
3. CT: Class size and Test Score
4. DM: Distance to Mother
5. GC: Who Graduated from College?
6. PM: Parental Education and College Major
7. ML: College Major and Importance of Becoming a Leader
8. DEI: Distance to College, Education, and Income (an IV approach)

QOCs are subject to change. More examples will be added if time permitted.

*The professor reserves the right to make changes to the syllabus, including due dates and test dates. These changes will be announced as early as possible.