

**SOC 1029**  
**305 12110**  
**Advanced Social Statistics**  
2024 Spring  
National Taiwan University

**Class: Friday 2, 3, 4, 6, 7 (6 & 7 are practicum)**  
**R101 for 2, 3, 4; R501 in College of Social Science for 6 & 7**

**Instructor: Meng-Jung Lin 林孟瑤 [mjlinmj@ntu.edu.tw](mailto:mjlinmj@ntu.edu.tw)**

**Office hours: Wednesdays 2:00-5:00 PM in R417 or online via Google Meet (please sign-up using Calendly: <https://calendly.com/mjlinmj/15min>) and by appointment**

**Teaching Assistants:**

**Yuan-Wen Li 李元文 [r10325010@ntu.edu.tw](mailto:r10325010@ntu.edu.tw)**

**Sheng-Wen Ke 柯盛文 [r12325001@ntu.edu.tw](mailto:r12325001@ntu.edu.tw)**

**Sheng-Pin Huang 黃聖斌 [b08302151@ntu.edu.tw](mailto:b08302151@ntu.edu.tw)**

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 ([piazza.com/ntu.edu.tw/spring2023/soc1029/home](https://piazza.com/ntu.edu.tw/spring2023/soc1029/home)) 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 10 assignments distributed on each Friday during our practicum session and be 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. You are required to contact me beforehand if you wanted to hand in an assignment late, or I may not accept it.

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 (15%): 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 (27%): You are required to **come to class** on the exam day. You will be asked to **submit an electronic poster for a research proposal on any topic of your choice and an accompanying 3-5 minute recorded presentation by the first hour of the exam time**. You will then **finish a worksheet/questionnaire** asking you to comment on at least two posters/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. 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 the 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 several classes. I will announce if you have to turn in the worksheet of the day at the beginning of the class period. You may miss 1 week of classes. **Please contact me for additional absences.** For those who have time conflicts, please let me know beforehand. Recording of each class will be posted in NTU COOL. You have to watch the recording before the next class. The video manager in NTU COOL will document your name and the amount of time you watched the recordings. The record will be taken as your

attendance. **MJ doesn't really care if you attend or not.** However, **MJ does care if you:** 1. **skip classes due to mental or physical illness or/and emergency.** 2. **say MJ didn't cover things show up on the assignments or exams.** 3. **do not realize you may not pass or get a good grade until almost the end of the semester.** So, if you understand all the above but don't want to come someday, you can send me an email saying that you are tired/lazy and want to get more sleep at home. I can totally understand that and will not count it as an absence. 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*. 11<sup>th</sup> Edition. Cengage Learning.**

**Studenmund, Arnold H. 2016. *Using Econometrics: A Practical Guide*. 7<sup>th</sup> Edition. Pearson.**

**Gordon, Rachel A. 2012. *Applied Statistics for the Social and Health Sciences*. 1<sup>st</sup> 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. 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.

5. Check your email account daily: When assignments or quizzes are available in NTU COOL, notification will be sent to your email address. Any changes to the course schedule will also be announced in COOL and through notification.

6. 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.

7. Please be prepared before coming to the office hours.

### **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 your 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. 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	2/23	Descriptive Statistics, Hypothesis Testing, & Association/Correlation Review	Healey Ch.1-Ch.13		1dis		
2	3/01	Elaborating Bivariate Tables	Healey Ch.14	SP	1due/2dis	1due	partial tables, direct relationships, spurious or intervening relationships, interaction, gamma
3	3/08	Multiple Regression	Healey Ch.15	PA	2due/3dis	2due	control variables, partial correlation coefficient (r), partial slopes, standardized partial slopes (beta-weights) multiple correlation coefficient (R), multiple regression, coefficient of multiple determination (R2)
4	3/15	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 $\beta$ , $SE(\beta)$ , t-test of single coefficient, F-test (joint hypotheses testing)
5	3/22	<b>Exam 1</b>	Healey Ch.1-15				
6	3/29	Dummy Variables, Interaction terms, & Nonlinear Relationships	(Gordon Ch.10, 11, 12)	DM	4due/5dis	4due	dummy variable, reference category, (linear combination), Interaction, the magnitude of the effect, curvilinear, logarithmic transformation, log-lin, quadratic term
7	4/05	<b>(NO CLASS) Spring Break</b>					
8	4/12	Outliers, Heteroskedasticity, & Multicollinearity	Gordon Ch.14		5due/6dis	5due	outlier, influential observation, hat values (leverage), studentized residual (standardized residual), Cook's distance, DFBETA, DFFITS, heteroskedasticity, the Breusch-Pagan Test, robust standard errors, multicollinearity, variance inflation factor (VIF)
9	4/19	Models for Binary Outcomes: Logistic (and Probit) Models	Studenmund Ch.13		6due/7dis	6due	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
10	4/26	Models for Binary Outcomes: Logistic (and Probit) Models Multinomial Models & Ordered Logit Models	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)
11	5/03	<b>Exam 2</b>	Week 1 to Week 10				
12	5/10	Multinomial Models & Ordered Logit Models	Agresti pgs. 480-492		8due/9dis		multinomial logit model, baseline-category logit
13	5/17	Specification Error and Omitted Variable Bias,  (Instrumental Variables and Causality)	Studenmund Ch.6 Gordon Ch.13.3 (Wooldridge Ch.15.1-3)		9due/10dis		specification, omitted variable bias instrumental variable regression (IV), two-stage least squares (TSL or 2SLS)
14	5/24	Models for Panel Data: First difference models, Fixed and Random Effects	Studenmund Ch.16		10due		panel data first difference model, fixed effects model, random effects model
15	5/31	<b>(NO CLASS) Review and Final Research Poster Preparation</b>					
16	6/07	<b>Final Research Poster Presentation and Peer Review</b>	Cumulative Final				

\*QOC: Question of the Class.

## Practicum Schedule

Week	Date	Topic	Readings	QOC*	Assignment	Keywords
1	2/23	Intro to Research Method and Survey Design	Ch.1-Ch.13		1dis	
2	3/01	Data Management	Ch.14		1due/2dis	partial tables, direct relationships, spurious or intervening relationships, interaction, gamma
3	3/08	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 (R <sup>2</sup> )
4	3/15	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 $\beta$ , $SE(\beta)$ , t-test of single coefficient, F-test (joint hypotheses testing)
5	3/22	<b>Project discussion 1</b>				
6	3/29	Dummy Variables, Interaction terms, & Nonlinear Relationships	(Gordon Ch.10, 11, 12)	DM	4due/5dis	dummy variable, reference category, (linear combination), Interaction, the magnitude of the effect, curvilinear, logarithmic transformation, log-lin, quadratic term
7	4/05	<b>(NO CLASS) Spring Break</b>				
8	4/12	Outliers, Heteroskedasticity, & Multicollinearity	Gordon Ch.14	DM	5due/6dis	outlier, influential observation, hat values (leverage), studentized residual (standardized residual), Cook's distance, DFBETA, DFFITS, heteroskedasticity, the Breusch-Pagan Test, robust standard errors, multicollinearity, variance inflation factor (VIF)
9	4/19	<b>Project discussion 2</b> Models for Binary Outcomes: Logistic (and Probit) Models	Studenmund Ch.13	GC	6due/7dis	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
10	4/26	<b>Project discussion 3</b> Multinomial Models & Ordered Logit Models	Gordon Ch.17.1-17.2	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)
11	5/03	<b>NO CLASS</b>				
12	5/10	Multinomial Models & Ordered Logit Models	Agresti pgs. 480-492	ML	8due/9dis	multinomial logit model, baseline-category logit
13	5/17	Specification Error and Omitted Variable Bias,  (Instrumental Variables and Causality)	Studenmund Ch.6 Gordon Ch.13.3 (Wooldridge Ch.15.1-3)	DEI	9due/10dis	specification, omitted variable bias instrumental variable regression (IV), two-stage least squares (TSLS or 2SLS)
14	5/24	First difference models, Fixed and Random Effects	Studenmund Ch.16	WW	10due	panel data first difference model, fixed effects model, random effects model
15	5/31	<b>(NO CLASS) Review and Final Research Poster Preparation</b>				
16	6/07	<b>Final Presentation and Peer Review (No Practicum)</b>	Cumulative Final			

\*QOC: Question of the Class.

### \*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)
9. WW: Work experience and Wages

**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.