SOC 1028

305 12100

Social Statistics

2022 Fall

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 (Class A) & 7 (Class B)

Instructor: Meng-Jung Lin mjlinmj@ntu.edu.tw
Office hours: Tuesdays 1:00-5:00 PM in R320C or online via Google Meet (please sign-up using Calendly: https://calendly.com/mjlinmj/15min) and by appointment

Does income inequality lead to political polarization? Is social class related to parenting? What accounts for income inequality between ethnic groups in the labor market? We will use statistics and programming to answer these questions throughout the semester. This course introduces tools to summarize the characteristics of data, and offers methods to draw conclusions about population from samples. Applying statistics, analyzing data, and interpreting results are the focuses of our class. Although very basic calculation skills are required (e.g., +, -, ×, ÷, √), you do not need further mathematic knowledge to be successful in this class.

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 using R programming.
3. Interpret results of statistical tests covered in class.
4. Apply statistical methods and computer skills to address daily and social issues.
5. Evaluate statistics and statistical method used in news, reports, and even 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/fall2022/soc1028/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. Interested in using statistics to understand social phenomena.
2. Willing to spend time outside of classroom to figure out how statistics and R work.
3. Considering to be a data analyst or go to graduate school.

Course Requirements

Practicum Assignments and Discussion (60%): There will be 12 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 me to help, you must submit your **hand calculating procedures for textbook 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.

Since my plan is to have you submit an electronic research poster by the end of next semester, **you are required to meet with me at least once during our practicum or/and my office hours** this semester to discuss your ideas. 10 points will be deducted if you failed to meet with me. Please develop at least one empirically testable research question and apply statistical methods learned in this class to a public dataset 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 (5%): Quiz with 1 or 2 short question(s) for the readings of that week will be available in NTU COOL on Monday 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. You can miss 1 of the quizzes.

Exams (15%): There will be 3 **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 Exam (10%): The final exam will be a cumulative exam. Similar to the previous 3 exams, there will be multiple choice, short answers, and R programming questions. You have to have your laptop ready for the final exam which will also be available in NTU COOL. You are required to take the final exam at the designated time. 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 (10%): Every class (including practicum) counts. For the in-person sessions, I will use the **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.

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 textbook **BEFORE** each class. The older versions of the textbook are allowed because the key concepts of statistics haven't been changed and the main organization of the book is almost the same.

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

The following books are recommended:

Agresti, Alan. 2018. Statistical Methods for the Social Sciences. Pearson.

Imai, Kosuke. 2018. Quantitative Social Science: An Introduction. Princeton University Press.

Wiley, Matt and Joshua F. Wiley. 2020. Beginning R. Apress: Imprint: Apress. (ebook available: https://ntu.primo.exlibrisgroup.com/permalink/886NTU_INST/14poklj/alma991038689242404786)

Class Policy

1. We will be spending much time on doing exercises and analyzing data in class. You **MUST read the textbook BEFORE** class so we can use the concepts and interpret the outputs from R. **Both quizzes and assignments are based upon the textbook**. In addition, only will you know what **you and R** are doing after reading the textbook.
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 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 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	Lesson	Date	Topic	Readings	QOC*	Assignment	Quiz	Keywords
1		9/09	Mid-Autumn Festival Holiday (No class)					
2	1	9/16	Introduction	Ch.1: pg.xxiii-7	R Setup	1dis	1due	data, hypothesis, variable
	2		Statistics and Social Research	Ch.1: pg.7-19	Data			descriptive statistics, inferential statistics, level of measurement
3	3	9/23	Descriptive statistics	Ch.2: pg.25-43	HSB	1due/2dis	2due	frequency distribution, midpoints, cumulative frequency
	4		Descriptive statistics	Ch.2: pg.43-59	HSB			ratios, percentage change, charts
	5		Measures of Central Tendency	Ch.3	War			mode, median, mean, skew
4	6	9/30	Measures of Dispersion	Ch.4	WVS	2due/3dis	3due	range, standard deviation, variance
	7		Normal Distribution	Ch.5: pg.129-135	LLN			normal curve, Z scores, normal curve table
	8		Normal Distribution	Ch.5: pg.135-144	LLN			probability and Z score
5	9	10/07	Exam 1	Ch.1-Ch.4		3due	4due	level of measurement, median, standard deviation
	10		Sampling and Sampling Distribution	Ch.6: pg.153-161	LLN			representative sample, sampling methods, sampling distribution, Central Limit Theorem
	11		Sampling and Sampling Distribution	Ch.6: pg.162-170	LLN			representative sample, sampling methods, sampling distribution, Central Limit Theorem
6	12	10/14	Estimation I	Ch.7: pg.173-182	RD	4dis	5due	bias, efficiency, confidence interval (interval-ratio)
	13		Estimation II	Ch.7: pg.179-189	RD			confidence interval (interval-ratio), confidence interval (proportion)
	14		Estimation III	Ch.7: pg.189-193	RD			width of interval, confidence level, sample size
7	15	10/21	Exam 2	Ch.5-Ch.7		4due/5dis		Z scores, sampling distribution, confidence interval
	16		Hypothesis Testing I: One-sample	Ch.8: pg.199-204	Obama			null hypothesis, alternative hypothesis
	17		Hypothesis Testing I: One-sample	Ch.8: pg.205-208	Obama			five-step model
8	18	10/28	Hypothesis Testing I: One-sample	Ch.8: pg.208-224	Obama	5due/6dis	6due	alpha level, student's t distribution, five-step model
	19		Hypothesis Testing II: Two-sample	Ch.9: pg.229-240	Mental			five-step model (interval-ratio), sample size, independent random samples
	20		Hypothesis Testing II: Two-sample	Ch.9: pg.241-250	Mental			five-step model (proportions), statistical significance, substantial significance
9	21	11/04	Hypothesis Testing III: ANOVA	Ch.10: pg.260-266	SWB	6due/7dis	7due	analysis of variance, total sum of squares, sum of squares within/between
	22		Hypothesis Testing III: ANOVA	Ch.10: pg.266-275	SWB			five-step model, interpretation, post hoc test
	23		Hypothesis Testing III: ANOVA	Ch.10: pg.275-279	SWB			five-step model, interpretation, post hoc test
10	24	11/11	Hypothesis Testing IV: Nonparametric Tests	Ch.11: pg.290-303	Parent	7due/8dis	8due	chi-square test, independence, five-step model
	25		Hypothesis Testing IV: Nonparametric Tests	Ch.11: pg.303-310	Parent			chi-square test, goodness-of-fit test, five-step model
	26		Hypothesis Testing (Review)	Ch.8-Ch.11				hypothesis testing, five-step model
11	27	11/18	Exam 3	Ch.8-Ch.11		8due/9dis		association, conditional distribution of y, positive/negative association
	28		Association/Correlation	Ch.12: pg.321-331				hypothesis testing, five-step model
	29		Association-Nominal Variables	Ch.12: pg.331-338	Fertility			phi, Cramer's V, lambda
12	30	11/25	Association-Ordinal Variables	Ch.12: pg.338-348	Ethnic	9due/10dis	9due	gamma, Spearman's rho
	31		Correlation-Interval-Ratio Variables	Ch.13: pg.357-368	SP			regression line, intercept (a), slope (b)
	33		Correlation-Interval-Ratio Variables	Ch.13: pg.369-377	SP			Pearson's r, r2 (coefficient of determination), correlation matrix
13		12/02	NTU Sports Meeting (No class)					
14	34	12/09	Correlation-Interval-Ratio Variables	Ch.13: pg.377-384	SP	10due/11dis	10due	dummy variable, testing Pearson's r
	35		Elaborating Bivariate Tables	Ch.14: pg.397-405	SP			partial tables, direct relationships, spurious or intervening relationships
	36		Elaborating Bivariate Tables	Ch.14: pg.405-415	SP			interaction, gamma
15	37	12/16	Multiple Regression	Ch.15: pg.424-435	PA	11due/12dis	11due	control variables, partial correlation coefficient (r)
	38		Multiple Regression	Ch.15: pg.436-438	PA			partial slopes, standardized partial slopes (beta-weights)

	39	Multiple Regression	Ch.15: pg.438-448	PA		multiple correlation coefficient (R), multiple regression, coefficient of multiple determination (R2)
16	12/23	Final Exam	Cumulative Final	12due		

*QOC: Question of the Class.

Practicum Schedule

Week	Date	Topic	Readings	QOC*	Assignment	Keywords
1	9/09	Mid-Autumn Festival Holiday (No class)				
2	9/16	Introduction	Ch.1	R Setup	1dis	R studio interface, Import datasets
		Statistics and Social Research		Data		
3	9/23	Descriptive statistics	Ch.2	HSB	1due/2dis	Histogram, plots, measures of central tendency; Public datasets overview
		Measures of Central Tendency	Ch.3	War		
4	9/30	Measures of Dispersion	Ch.4	WVS	2due/3dis	Z table look-up, simulation; Idea discussion (4~5 people/class)
		Normal Distribution	Ch.5	LLN		
5	10/07	Sampling and Sampling Distribution	Ch.6	LLN	3due	Sampling method demo, simulation
6	10/14	Estimation	Ch.7	RD	4dis	Confidence interval; Idea discussion (4~5 people/class)
7	10/21	Hypothesis Testing I: One-sample	Ch.8	Obama	4due/5dis	One-sample hypothesis testing; Idea discussion (4~5 people/class)
8	10/28	Hypothesis Testing II: Two-sample	Ch.9	Mental	5due/6dis	Two-sample hypothesis testing
9	11/04	Hypothesis Testing III: ANOVA	Ch.10	SWB	6due/7dis	ANOVA, post hoc test; Idea discussion (4~5 people/class)
10	11/11	Hypothesis Testing IV: Nonparametric Tests	Ch.11	Parent	7due/8dis	Chi-square test; Idea discussion (4~5 people/class)
11	11/18	Association-Nominal Variables	Ch.12: pg.321-331	Fertility	8due/9dis	Phi, Cramer's V, lambda; Idea discussion (4~5 people/class)
12	11/25	Association-Ordinal Variables	Ch.12: pg.338-348	Ethnic	9due/10dis	Gamma, Spearman's rho; Idea discussion (4~5 people/class)
13	12/02	NTU Sports Meeting (No class)				
14	12/09	Correlation-Interval-Ratio Variables	Ch.13	SP	10due/11dis	Regression line, Pearson's r, correlation matrix, dummy variable; Idea discussion (4~5 people/class)
15	12/16	Multiple Regression	Ch.15	PA	11due/12dis	Partial slopes, standardized partial slopes (beta-weights), multiple regression, coefficient of multiple determination (R2)
16	12/23	Final Exam (No Practicum)	Cumulative Final		12due	

*QOC: Question of the Class.

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

1. HSB: Gender and Math Ability
2. War: Measuring Civilian Victimization during Wartime
3. WVS: Which Country is More Conservative?
4. LLN: Simulation for the law of large numbers
5. RD: Racial Discrimination in the Labor Market
6. Obama: Obama's Supports Rate
7. Mental: Gender and Mental Health
8. SWB: Social Class and Subjective Well-being
9. Parent: Social Class and Parenting
10. Fertility: Level of Development and Fertility
11. Ethnic: Ethnic Diversity and Economic Inequality
12. SP: Subjective Social Position and Attitudes on Economic Inequality
13. PA: Education, Income, and Physical Attractiveness

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.