The Hong Kong University of Science and Technology

UG Course Syllabus

Quantitative Data Analysis for Social Research II

SOSC 2400

3 Credit

Pre-requisite: SOSC 1100.

Time: Tuesday & Thursday, 10:30AM – 11:50AM

Venue: Room 4580, Academic Building (Lift 27-28)

Name: Titi ZHOU

Email: ttz@ust.hk

Office Hours: Thursday, 12:00PM – 1:00PM

Course Description

Building upon the foundational data analysis and R skills acquired in SOSC 1100, this course provides students with the advanced knowledge and skills to conduct more sophisticated quantitative research and critically assess statistical claims in various contexts. Students will master core principles of statistical inference, enhance their R programming capabilities to address complex social questions, and learn to rigorously conduct analyses, interpret nuanced results, and identify the strengths and limitations of different analytical approaches, empowering them to engage confidently with data-driven discussions on social issues.

Intended Learning Outcomes (ILOs)

By the end of this course, students should be able to:

ILO1: Demonstrate proficiency in and apply core principles of statistical inference to conduct sophisticated quantitative analyses of social science data.

ILO2: Effectively employ R programming techniques for data manipulation, statistical analysis, and compelling visualization to address and interpret social questions.

ILO3: Formulate and substantiate rigorous, data-driven arguments to address complex social questions, critically evaluating the statistical methodologies and data sources employed.

ILO4: Critically evaluate, quantify and interpret the uncertainty inherent in statistical analyses.

Canvas

Canvas is the primary learning platform by which the course is delivered. If you encounter any problems with Canvas, please visit https://cei.hkust.edu.hk/canvas/faqs-students for assistance.

You are responsible for all information posted on Canvas for this course, so please check it regularly for announcements, class assignments, and any schedule changes.

Textbook

Llaudet, Elena and Kosuke Imai. 2022. *Data Analysis for Social Science: A Friendly and Practical Introduction*. Princeton: Princeton University Press.

Agresti, Alan, Christine Franklin and Bernhard Klingenberg. 2023. *Statistics: The Art and Science of Learning from Data*. Harlow; Hoboken, NJ: Pearson.

Software

We will be conducting data analysis using R, an open-source statistical software known for its power and flexibility. R is widely utilized by data analysts in both corporate and academic settings. You can download it and access helpful documentation at http://www.r-project.org.

To enhance your experience with R, we highly recommend using RStudio (https://posit.co/downloads), a free user interface that simplifies many common operations.

Assessment and Grading

This course will be assessed using criterion-referencing and grades will not be assigned using a curve. Detailed rubrics for each assignment are provided below, outlining the criteria used for evaluation.

1. **Weekly Question/Comment Submission (10%):**

- To promote class discussion, you need to submit a question or comment about the lecture *EACH* week. Evaluation will follow the rubric provided below.
- Submission deadlines are listed in the Summary Table on page 8 and again on the last page.
- Late submission without a valid reason will not be accepted.

2. **Problem Sets (25%):**

- To practice data analysis skills using R and learn key statistical concepts, there will be TWO take-home problem sets, with each carrying equal weight toward the course grade.
- Problem set will be available on Canvas. Check the "Schedule" on the last page for the posting time and the submission deadline.
- Late submission without a valid reason will cause a grade deduction. Six hours late submission will not be accepted.

3. **In-class Quizzes (40%):**

- To evaluate your understanding of the key concepts discussed in the class meeting, there will be TWO closed-book quizzes during class meetings. Each will contribute 20% to the course grade.
- The dates for each in-class quiz are listed in the Summary Table on page 8 and again on the last page.

4. Group Project (25%):

- To practice using R to investigate a social phenomenon with real-world datasets, students will be assigned a topic and corresponding dataset, and will conduct a quantitative analysis using the analytical approaches learned in this course.
- Each group will consist of 3 members and should be self-formed. Individuals not in a group by **Oct 30** will be assigned to groups randomly. A Google sheet will be provided to record group memberships. The finalized group information will be posted on Canvas.
- Group projects will be presented in class. The deadline to submit presentation slides is **Nov 23**. Evaluation will follow the rubric provided below.
- **Peer evaluation**: individual grades for the group project will be adjusted based on peer evaluations conducted with the Group Member Evaluation (GME) tool. Students will receive full marks if their group members fail to submit their peer evaluations by **Nov 30**.

Summary Table:

Assessment Task	Contribution to Overall Course grade	Due date
Weekly Question/Comment Submission	10%	**Tuesday** (Sep 9 – Sep 30, Nov 18), **Thursday** (Oct 9 – Nov 6)
Problem Sets	25%	Sep 30, Oct 28
In-class Quizzes	40%	Oct 2, Oct 28
Group Project	25%	Nov 23

Mapping of Course ILOs to Assessment Tasks:

Assessed Task	Mapped ILOs	Explanation
Weekly Question/Comment Submission	ILO1, ILO2, ILO3	This task encourages students to deepen their understanding of statistical concepts (ILO1), practice data analysis using R (ILO2), and develop data-backed arguments through critical thinking (ILO3).
Problem Sets	ILO1, ILO2, ILO3	This task provides opportunities for students to practice data analysis with R (ILO2), enhance their understanding of statistical concepts (ILO1), and develop preliminary data-backed arguments (ILO3).
In-class Quizzes	ILO1, ILO3	This task assesses students' understanding of key statistical concepts (ILO1) and their ability to develop data-backed arguments based on analysis (ILO3).
Group Project	ILO2, ILO3, ILO4	This task allows students to apply R programming skills (ILO2) to analyze real-world datasets, develop data-backed arguments (ILO3), and understand and quantify uncertainty in their analyses (ILO4).

Final Grade Descriptors:

Grades	Short Description	Elaboration on subject grading description
A	Excellent Performance	Demonstrates a comprehensive understanding of statistical concepts and methods, exceptional ability to summarize and analyze data using R, and outstanding skill in developing insightful, data-backed arguments based on real-world datasets. Shows a strong grasp of how to quantify uncertainty in data analysis.
B	Good Performance	Shows a good understanding of statistical concepts, strong skills in data analysis using R, and a solid ability to create well-structured data-backed arguments. Can adequately interpret results and discuss the strengths and limitations of analyses, with some understanding of uncertainty.
C	Satisfactory Performance	Meets basic expectations with an adequate understanding of fundamental statistical concepts and methods. Demonstrates basic data analysis skills using R and is able to develop satisfactory data-backed arguments, though interpretations may lack depth. Shows some awareness of uncertainty in data analysis.
D	Marginal Pass	Displays minimal understanding of statistical concepts, with limited ability to summarize and analyze data using R. Struggles to develop coherent data-backed arguments and offers weak interpretation of results. Limited awareness of uncertainty in data analysis.
F	Fail	Fails to demonstrate an adequate understanding of statistical concepts and methods, does not effectively summarize or analyze data using R, and lacks the ability to develop coherent data-backed arguments. Shows no understanding of uncertainty in data analysis.

Student Rubrics

The following rubrics will be used to assess the tasks that you submit or achieve.

Weekly Question/Comment Submission Rubric:

Criteria	Excellent	Good	Average	Unsatisfactory
Relevance to Course Content	Highly relevant, demonstrating a thorough understanding of the course content.	Mostly relevant with clear connections to course content.	Somewhat relevant but may include off-topic elements.	No relevance.
Contribution to Class Discussion	Engages deeply and stimulates discussion.	Enhances class engagement.	Provides basic input but lacks depth.	No contributions.

Group Project Rubric:

Criteria	Excellent	Good	Satisfactory	Needs Improvement
Content Understanding	Demonstrates a thorough understanding of statistical concepts and methods; effectively applies them to the chosen topic.	Shows a good understanding of statistical concepts; applies them adequately to the topic.	Displays a basic understanding of statistical concepts; some application to the topic is present but lacks depth.	Lacks understanding of statistical concepts; minimal or incorrect application to the topic.
Use of R Programming	Effectively utilizes R to analyze data, providing clear and insightful visualizations and summaries.	Uses R adequately to analyze data with some visualizations and summaries that are mostly clear.	Uses R minimally; visualizations and summaries are present but unclear or poorly executed.	Fails to effectively use R for data analysis; visualizations and summaries are missing or irrelevant.
Argument Development	Develops compelling, data-backed arguments that are well-structured and logically presented.	Presents solid data-backed arguments; structure is mostly logical but may lack some clarity.	Arguments are present but lack depth or clarity; some evidence is provided but not well integrated.	Arguments are unclear or poorly structured; lacks data backing or relevance.
Presentation Skills	Engages the audience effectively; speaks clearly and confidently; uses visuals to enhance understanding.	Engages the audience; speaks clearly; uses some visuals, though they may not fully enhance understanding.	Limited audience engagement; speech may lack clarity; visuals are present but not effectively used.	Fails to engage the audience; speech is unclear; visuals are lacking or ineffective.

Group Project Peer Evaluation Form:

Group Member Name	Contribution to Workload (effort, responsibility)	Quality of work	Reliability (met deadlines, preparedness)	Communication (clarity, responsiveness)	Collaboration (respect, flexibility)
1.					
2.					
3.					

Course AI Policy

The use of generative AI tools is permitted to assist students with understanding the course materials, but overreliance on AI tools is discouraged, and students must verify the information from reliable sources and not rely solely on AI, ensuring all work submitted is original and properly cited to adhere to academic integrity.

Communication and Feedback

Assessment marks for individual assessed tasks will be communicated via Canvas within two weeks of submission. Feedback on assignments will include comments on strengths and areas for improvement. Students who have further questions about the feedback including marks should consult the instructor within five working days after the feedback is received.

Late submission Policy

Late submissions for individual assignments and requests for in-class quiz postponements are generally not permitted. Exceptions will only be considered for *documented* extenuating circumstances (e.g., medical emergencies) and require *prior approval* from the instructor. Students are responsible for communicating any potential delays as soon as possible and providing the necessary documentation.

Group assignments are tied to specific in-class presentation dates and cannot be submitted late. Failure to present on the scheduled date will result in a zero for the assignment.

Academic Integrity

Students are expected to adhere to the university's academic integrity policy. Students are expected to uphold HKUST's Academic Honor Code and to maintain the highest standards of academic integrity. The University has zero tolerance of academic misconduct. Please refer to Academic Integrity | HKUST – Academic Registry for the University's definition of plagiarism and ways to avoid cheating and plagiarism.

Course Outline and Tentative Schedule

DATE	#	TOPIC	WEEKLY SUBMISSION	PROBLEM SET Posted	PROBLEM SET Submitted	IN-CLASS QUIZ
2-SEP	0	Course Introduction				
4-SEP	1	Introduction to R and RStudio				
9-SEP 11-SEP	2	Data and Descriptive Analysis	1st (Sep 9)			
16-SEP 18-SEP	3	Exploring the Relationship between Two Variables	2nd (Sep 16)			
23-SEP 25-SEP	4	Predicting Outcomes Using Linear Regression	3rd (Sep 23)	1st (Sep 25)		
30-SEP 2-OCT	5	Estimating Causal Effects with Randomized Experiments	4th (Sep 30)		1st (Sep 30)	1st (Oct 2)
7-OCT		*No Class (Public Holiday)*				
9-OCT 14-OCT	6	Estimating Causal Effects with Observational Data	5th (Oct 9)			
16-OCT 21-OCT	7	Probability	6th (Oct 16)	2nd (Oct 21)		
23-OCT 28-OCT	8	Statistical Inference: Confidence Intervals	7th (Oct 23)		2nd (Oct 28)	2nd (Oct 28)
30-OCT 4-NOV	9	Statistical Inference: Hypothesis Testing	8th (Oct 30)			
6-NOV 11-NOV	10	Linear Regression with Uncertainty	9th (Nov 6)			
13-NOV		Group Project Consultation*				
18-NOV 20-NOV	11	Linear Regression: Practical Issues	10th (Nov 18)			
25-NOV 27-NOV		Group Project Presentation**				

* Finalized group information will be posted on Oct 30; topics and datasets will be assigned on Nov 6.

** Presentation slides must be submitted by Nov 23.