

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: Monday & Wednesday, 12:00PM – 1:20PM

Venue: Room 2404, Academic Building (Lift 17-18)

Name: Titi ZHOU

Email: ttz@ust.hk

Office Hours: Monday, 1:30PM – 2:30PM

Course Description

Data analysis has profoundly changed our understanding of social issues and decision-making processes. This course is designed to provide students with the essential knowledge and skills to conduct data analysis in their own research and critically assess statistical claims made in news media, policy reports, and academic studies.

Students will learn the core principles of statistical inference and develop programming skills to tackle various social questions through data analysis. We will demonstrate how to conduct analyses, interpret results, and identify the strengths and potential limitations of each analysis.

Intended Learning Outcomes (ILOs)

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

ILO1: Recognize fundamental statistical concepts and methods in data analysis.

ILO2: Use R programming to summarize data both numerically and visually.

ILO3: Develop preliminary data-backed arguments based on the analysis of real-world data.

ILO4: Understand and quantify uncertainty in data analysis.

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

Lludet, 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. **Attendance (6%):** Attendance at all class meetings is mandatory. There will be THREE randomly conducted attendance checks, with each contributing 2% to the overall course grade.
2. **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.
 - The questions/comments need to be posted in the *Discussion* section on CANVAS. Refer to the "Summary Table" on the next page for the submission deadlines.
 - Late submission without a valid reason will not be accepted.
 - If you miss three or more submissions, you will not receive any grade for this component.
3. **Problem Sets (18%):**
 - To practice data analysis skills using R and learn key statistical concepts, there will be THREE take-home problem sets, with each set accounting for 6% of the course grade.
 - Problem set will be available on CANVAS. Check the "Schedule" at the end for the posting time and the submission deadline.
 - Late submission without a valid reason will cause a grade deduction by half. Six hours late submission will not be accepted.
4. **In-class Quizzes (16%):**
 - 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 8% to the course grade.
 - Check the "Schedule" at the end for the time of the in-class quiz.

5. Group Project (15%):

- To better practice how to use R to investigate a social phenomenon with real-world datasets, students will choose a topic and conduct a quantitative data analysis using the analytical approaches learned in this course.
- Students will form their own groups, each consisting of 3-5 members.
- Topics and corresponding datasets will be provided later.
- Group projects will be presented in class. The deadline to submit presentation slides is **Nov 15**. Evaluation will follow the rubric provided below.

6. Group Project: Peer Evaluation (5%):

- To ensure individual accountability and promote fair contribution, group members will conduct peer evaluations to assess each other's efforts.
- The peer evaluation form, which is also provided below, will be accessible on CANVAS, and the submission deadline is **Nov 22**.

7. **Final Exam (30%)**: The closed-book final exam will cover the content learned throughout the semester. Specific arrangements will be announced later.

Summary Table:

Assessment Task	Contribution to Overall Course grade (%)	Due date
Attendance	6%	Week 1 to 13
Weekly Question/Comment Submission	10%	Sep 4, Sep 11, and Mondays from Sep 23 to Nov 11
Problem Sets	18%	Oct 3, Oct 31, Nov 14
In-class Quizzes	16%	Oct 9, Nov 6
Group Project	15%	Nov 15
Group Project: Peer Evaluation	5%	Nov 22
Final Exam	30%	Nov 27

Mapping of Course ILOs to Assessment Tasks:

Assessed Task	Mapped ILOs	Explanation
Attendance	ILO1	Regular attendance helps students grasp fundamental statistical concepts (ILO1) through active engagement in class discussions.
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).
Group Project: Peer Evaluation	ILO2, ILO3	This task promotes accountability and helps students reflect on the contributions of their group members in doing data analysis using R (ILO2) and developing data-backed arguments (ILO3).
Final Exam	ILO1, ILO2, ILO3, ILO4	This task evaluates students' comprehensive understanding of statistical concepts (ILO1), their ability to analyze data using R (ILO2), their capacity to develop data-backed arguments (ILO3), and their understanding of uncertainty in data analysis (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 Research	Contribution to Analysis	Contribution to Presentation	Overall Contribution	Comments
1.					
2.					
3.					

Evaluation Criteria:

- **Contribution to Research:** Active involvement in selecting the topic and designing the analysis.
- **Contribution to Analysis:** Engagement in using R for data analysis and interpretation of results.
- **Contribution to Presentation:** Participation in preparing and delivering the group presentation, including creating visual aids.

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 submission will *NOT* be accepted unless a valid reason is given and prior special permission is obtained.

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	PROBLEM SET		IN-CLASS QUIZ
			Posted	Submitted	
2-SEP	0	Introduction			
4-SEP	1	Data and Descriptive Analysis			
9-SEP					
11-SEP	2	Relationship between Two Variables	1st (Sep 11)		
16-SEP					
18-SEP		<i>(Public Holiday)</i>			
23-SEP	3	Linear Regression: Coefficients and Prediction			
25-SEP					
30-SEP	4	Multiple Linear Regression			
2-OCT				1st (Oct 3)	
7-OCT	5	Probability	2nd (Oct 7)		
9-OCT					1st (Oct 9)
14-OCT	6	Probability Distributions			
16-OCT					
21-OCT	7	Statistical Inference: Confidence Intervals			
23-OCT					
28-OCT	8	Statistical Inference: Hypothesis Testing			
30-OCT				2nd (Oct 31)	
4-NOV	9	Linear Regression with Uncertainty	3rd (Nov 4)		
6-NOV					2nd (Nov 6)
11-NOV	10	Linear Regression: Practical Issues			
13-NOV				3rd (Nov 14)	
18-NOV		Group Project Presentation*			
20-NOV					
25-NOV		Review: Q & A**			
27-NOV		Final Exam			

* Presentation slides are required. The submission deadline is Nov 15.

** The course review lecture notes will be available on Nov 18.