

The Hong Kong University of Science and Technology

UG Course Syllabus

**Quantitative Data Analysis for Social Research I**

SOSC 1100

3 Credit

**Time:** Tuesday & Thursday, 12:00PM – 1:20PM

**Venue:** Room 5620, Academic Building (Lift 31-32)

**Name:** Titi ZHOU

**Email:** [ttz@ust.hk](mailto:ttz@ust.hk)

**Office Hours:** Tuesday, 10:30AM – 11:30AM

**Course Description**

This course offers an introduction to data analysis in the social sciences, with no prior knowledge required. Using plain language, the course introduces basic statistical concepts and shows how to analyze real-world data using the statistical program R. Specifically, the course will (1) cover essential practices for analyzing data in social science research, including data management, descriptive analysis, and correlation; (2) introduce basic ideas and intuition about causality; and (3) demonstrate how these practices are applied to real-world data using R.

**Intended Learning Outcomes (ILOs)**

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

ILO1: Understand basic statistical concepts relevant to social science research.

ILO2: Summarize data numerically and visually using basic functions in R.

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

**Course Format**

This course will be taught in *blended mode*. You are required to watch lecture videos on CANVAS and participate in learning activities in class meetings. Specifically, each week:

- **Tuesday (in the classroom):** Attend the in-person class meeting, participate in discussions and activities, and engage in R practice. Don't forget to bring your laptop.
- **Thursday (online learning):** Lecture notes and videos will be available on CANVAS by 12:00 PM. Please watch the videos at your own pace before next Tuesday's class meeting.

## 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.

## 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 (12%):**
  - To promote discussion in class meeting, you need to submit a question or comment about the lecture notes/videos for *EACH* week. Evaluation will follow the rubric provided below.
  - The questions/comments need to be posted in the *Discussion* section on CANVAS by **Friday**, except Sep 27.
  - 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 (24%):**
  - To practice data analysis skills using R and deepen the understanding of key statistical concepts, there will be THREE take-home problem sets, with each set accounting for 8% 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 (18%):

- To evaluate your understanding of the key concepts presented in the lecture videos, there will be THREE closed-book quizzes during class meetings.
- Check the "Schedule" at the end for the time of the in-class quiz.
- Only the TWO highest grades will be considered for your final grade, with each contributing 9% to the overall course grade.

5. **Final Exam (40%):** 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	12%	Fridays (except Sep 27)
Problem Sets	24%	Sep 25, Oct 23, Nov 13
In-class Quizzes	18%	Sep 26, Oct 29, Nov 19
Final Exam	40%	Nov 28

#### Mapping of Course ILOs to Assessment Tasks:

Assessed Task	Mapped ILOs	Explanation
Attendance	ILO1	Regular attendance helps students understand basic statistical concepts (ILO1) through engagement with class meetings.
Weekly Question/Comment Submission	ILO1, ILO2, ILO3	This task helps students deepen their understanding of statistical concepts (ILO1), perform data analysis with R (ILO2) and develop data-backed arguments through critical thinking (ILO3).
Problem Sets	ILO1, ILO2, ILO3	This task allows students to practice data analysis using R (ILO2) and develop data-backed arguments (ILO3). It also enhances their understanding of statistical concepts (ILO1).
In-class Quizzes	ILO1, ILO3	This task assesses students' understanding of key statistical concepts (ILO1) and data analysis skills (ILO3).
Final Exam	ILO1, ILO2, ILO3	This task evaluates students' understanding of statistical concepts (ILO1), the ability to analyze data using R (ILO2) and the ability to develop data-based arguments (ILO3).

**Final Grade Descriptors:**

<b>Grades</b>	<b>Short Description</b>	<b>Elaboration on subject grading description</b>
A	Excellent Performance	Demonstrates a comprehensive understanding of statistical concepts, excellent ability to summarize and analyze data using R, and outstanding skill in developing insightful data-backed arguments.
B	Good Performance	Shows a good grasp of statistical concepts, strong skills in data analysis using R, and good ability to create solid data-backed arguments.
C	Satisfactory Performance	Meets basic expectations with an adequate understanding of statistical concepts, basic data analysis skills using R, and satisfactory ability to develop data-backed arguments.
D	Marginal Pass	Shows minimal understanding of statistical concepts, limited ability to summarize and analyze data using R, and weak skill in developing data-backed arguments.
F	Fail	Fails to demonstrate adequate understanding of statistical concepts, does not effectively summarize and analyze data using R, and lacks the ability to develop data-backed arguments.

**Student Rubrics**

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

**Weekly Question/Comment Submission Rubric:**

<b>Criteria</b>	<b>Excellent</b>	<b>Good</b>	<b>Average</b>	<b>Unsatisfactory</b>
<b>Relevance to Course Content</b>	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.
<b>Contribution to Class Discussion</b>	Engages deeply and stimulates discussion.	Enhances class engagement.	Provides basic input but lacks depth.	No contributions.

### **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

Video (Thursday)	Meeting (Tuesday*)	#	Topic	Problem Set		In-class Quiz
				Posted	Submitted	
	Sep 3	0	Course Introduction			
			<b>Data and Data in R</b>			
Sep 5	Sep 10	1	R and RStudio	1st (Sep 5)		
Sep 12	Sep 17	2	Data, Observations, and Variables			
Sep 19	Sep 24	3	Computing and Interpreting Means		1st (Sep 25)	
	Sep 26					1st (Sep 26)
			<b>Inferring Population Characteristics via Survey Research</b>			
Oct 3	Oct 8	4	Sampling	2nd (Oct 3)		
Oct 10	Oct 15	5	Descriptive Statistics			
Oct 17	Oct 22	6	Tables and Graphs		2nd (Oct 23)	
Oct 24	Oct 29	7	Scatter Plot	3rd (Oct 24)		2nd (Oct 29)
Oct 31	Nov 5	8	Correlation			
			<b>Interpreting Causal Effects with Randomized Experiments: Preliminary</b>			
Nov 7	Nov 12	9	Causal Effects with Randomized Experiments		3rd (Nov 13)	
Nov 14	Nov 19	10	Difference-in-means Estimator			3rd (Nov 19)
Nov 21	Nov 26	11	Examples of Randomized Experiments and Course Review			
	<b>Nov 28</b>		<b>Final Exam</b>			

\* All class meetings are scheduled for Tuesdays, except for Sep 26 (Thursday).