Statistics for Social and Behavioral Sciences

Fall 2026

Course Information

Course Title	Statistics for Social and Behavioral Sciences
Course Code	SOCSC-UH 1010Q
Credits	4
Lecture	Campus Center Room W004 | Tuesdays & Thursdays, 3:20 PM – 4:35 PM
Recitation	Campus Center Room 306 | Fridays, 10:40 AM – 11:55 AM
Prerequisites	MATH-UH 1000A, or math proficiency test
Cross-lists	Core Curriculum > Quantitative Reasoning Majors > Business, Organizations and Society Majors > Business, Organizations and Society > Social Science Courses Majors > Economics Majors > Political Science Majors > Social Research and Public Policy Minors > Economics Minors > Social Research and Public Policy

Faculty Information

Faculty	Dr. Yanwen Wang
Office	A5 1186-C
Email	yanwen.wang@nyu.edu
Office Hours	Wednesdays 2:00 PM – 4:00 PM, or by appointment
Website	`http://www.yanwenwang.com`

Instructor Information

Instructor	Bishnoi Shivangi
Email	shivangi.bishnoi@nyu.edu
Office Hours	By appointment

1 Course Description

This course introduces students to the fundamentals of statistics and probability as tools for reasoning with data in the social and behavioral sciences. Students learn how data are collected, organized, visualized, modeled, and used to support claims under uncertainty. The course emphasizes statistical thinking: choosing appropriate methods, interpreting results in context, communicating uncertainty, and recognizing the limits of statistical evidence.

The course begins with data, study design, and exploratory data analysis before turning to regression as a framework for describing and predicting relationships between variables. Students then study the foundations of statistical inference, including randomization, bootstrapping, mathematical models, confidence intervals, hypothesis tests, errors, and power. Later units apply these ideas to practical problems involving group comparisons, association, regression inference, causality, and model choice for different data structures.

R Labs provide hands-on experience with data analysis, visualization, modeling, and reproducible research using R and Quarto. Students will also learn how to use AI tools responsibly to support coding, interpretation, and communication while maintaining ownership of their analysis. The course culminates in a Research Hackathon, where students work in teams to formulate research questions, analyze a real-world dataset, communicate statistical conclusions, and reflect on uncertainty and limitations.

By the end of the semester, students will have developed the conceptual, computational, and interpretive skills needed to conduct introductory quantitative research and to evaluate statistical claims in academic, professional, and public contexts.

2 Course Learning Outcomes

Upon successful completion of this course, students will be able to:

CLO1: Understand data, study design, and statistical evidence. Students will be able to distinguish between populations, samples, variables, and observations; evaluate sampling and study design choices; and explain how data collection affects the scope of statistical conclusions.

CLO2: Describe and visualize empirical patterns. Students will be able to summarize and visualize categorical and numerical data, compare distributions across groups, and identify patterns, outliers, and associations using appropriate descriptive statistics and graphics.

CLO3: Apply and interpret statistical models and inference. Students will be able to apply regression models, confidence intervals, hypothesis tests, simulation-based methods, and formula-based inference to answer research questions and quantify uncertainty.

CLO4: Use R for reproducible statistical analysis. Students will be able to use R and Quarto to import, inspect, visualize, model, and analyze data, while producing reproducible code and written interpretations of statistical results.

CLO5: Evaluate claims and communicate statistical reasoning. Students will be able to distinguish descriptive, predictive, inferential, and causal claims; assess the limits of statistical evidence; and communicate results, uncertainty, and limitations clearly in written and oral formats.

The course contributes to several program-specific learning outcomes (PLOs), including:

- Critical thinking (Econ PLO1, high)
- Project management (Econ PLO3, high)
- Proficiency in empirical analysis (Econ PLO6, high)
- Capacity to engage with the professional literature (Political Science PLO1, medium)
- Information technology skills (Political Science PLO5, medium)
- Critical thinking, writing, and analysis (SRPP PLO3, high)
- Critical evaluation of methodological approaches (SRPP PLO6, high)
- Abilities and skills necessary to design, plan, and carry out a research project independently (SRPP PLO8, medium)

A complete list of PLOs can be found here: Economics, Political Science, Business, Organizations, and Society, and Social Research and Public Policy.

3 Teaching and Learning Methodologies

This course is designed to build statistical reasoning through a combination of conceptual instruction, hands-on data analysis, regular practice, and integrative application. The goal is not only to learn statistical procedures, but also to understand when they are appropriate, how they connect to research questions and study designs, and how results should be interpreted and communicated.

Lectures. Lectures introduce the core concepts of statistics and probability, including data collection, exploratory data analysis, regression, inference, and statistical reasoning. Class sessions combine conceptual explanation, worked examples, short exercises, and discussion. Students are expected to complete the assigned readings before class so that lecture time can be used to clarify ideas, connect concepts, and practice interpretation.

R Labs. R Labs provide hands-on experience with data analysis, visualization, modeling, inference, and reproducible research. In these sessions, students apply ideas from lecture to real or realistic datasets, write and run code in R, interpret output, and produce short explanations of their results. The labs are designed to build both technical skill and statistical judgment.

Quizzes and Exams. Quizzes and exams assess students' understanding of core concepts, methods, and interpretations. Quizzes provide regular checks on recent material, while the midterm and final exam ask students to synthesize ideas across larger parts of the course.

Research Hackathon. The course culminates in a Research Hackathon, where students work in teams to apply the full statistical workflow to a real-world dataset. Teams develop research questions and hypotheses, conduct exploratory analysis, select appropriate methods, interpret results, communicate uncertainty, and discuss limitations. The hackathon emphasizes reproducible code, clear statistical reasoning, and appropriately scoped conclusions.

Discussion Board and Office Hours. Students will have access to support through an online discussion board for peer-to-peer exchange and instructor feedback. Additional guidance will be available through email and during scheduled office hours. Students are encouraged to ask questions early and often, especially when working through R code, statistical interpretation, or project decisions.

3.1 Instructional Time

The course is organized into (i) **lectures** covering statistical concepts, methods, and examples, (ii) **R Labs** focused on performing data analysis in R, and (iii) **graded activities** including in-class quizzes, exams, and the Research Hackathon.

See Table 1 for a breakdown of instructional time.

Table 1: Instructional Activities

Activities	Format	Minutes	Frequency	Total Minutes
Lectures	in-person	75	20	1500
R Labs	in-person	75	12	900
Quizzes	in-person	75	3	225
Exams	in-person	75	2	150
Research Hackathon	in-person	75	3	225
Total				**3000**

3.2 Textbook and Course Materials

The course uses the following resources:

- Çetinkaya-Rundel, Mine, and Johanna Hardin. 2024. *Introduction to Modern Statistics*. Second edition. OpenIntro, Inc.
- Diez, David, Mine Çetinkaya-Rundel, and Christopher D. Barr. 2019. *OpenIntro Statistics*. Fourth edition. OpenIntro, Inc.
- Hanck, Christoph, Martin Arnold, Alexander Gerber, and Martin Schmelzer. *Introduction to Econometrics with R*.
- Wickham, Hadley, Mine Çetinkaya-Rundel, and Garrett Grolemund. 2023. *R for Data Science*. Second edition. O'Reilly Media. (Optional)

Assigned readings are listed in the detailed course schedule. All other materials, including lecture slides, lab exercises, datasets, assignments, and additional resources, will be posted on Brightspace.

3.3 Hardware and Software

A **laptop** is required. Students will also need a **calculator** with basic functions for quizzes and exams.

The course relies on **R**, a free, open-source programming language used widely in data science and academia. We will use **Positron**, a free Integrated Development Environment (IDE) designed for data science, with built-in support for R, interactive data exploration, and AI-assisted coding workflows.

Note: We will not use Stata, SPSS, or Excel for data analysis in this course. All assignments must be completed in R.

Please complete the following steps to set up your environment before the first R Lab:

1. **Install R:** Download and install the latest version of R from CRAN.
2. **Install Positron:** Download and install the latest version of Positron.
3. **Install packages:** Open Positron and run the following command in the R console: `install.packages("tidyverse")`

---

4 Graded Activities

The graded activities in this course are designed to combine regular practice, conceptual checks, and integrative application. R Lab problem sets provide repeated opportunities to practice coding, data analysis, and interpretation. Quizzes and exams assess understanding of core statistical concepts and methods. The Research Hackathon brings these skills together in a collaborative project that asks students to move from a research question to data analysis, interpretation, and communication.

Grades are not curved. Each activity is assessed according to the criteria described below, not in comparison to other students.

4.1 R Lab Problem Sets

Each R Lab centers on an in-class problem set designed to build skills in R coding, statistical reasoning, and reproducible analysis. Students will submit their completed code by the end of the lab session. Each submission must also include a brief reflection in the form of a short conversation with an AI assistant about what they did, what difficulties they encountered, and what they learned.

R Lab problem sets account for 20% of the final grade and are graded using two completion-based marks: **Completion** and **Completion+**. A submission earns **Completion** when it shows a good-faith effort to complete the assigned tasks. A submission earns **Completion+** when it shows especially thoughtful engagement, such as clear code, careful interpretation, meaningful troubleshooting, or well-explained use of AI assistance.

4.2 Quizzes

There will be three in-class quizzes covering material from the preceding course units. These quizzes are designed to provide regular checks on students' understanding of core concepts, methods, and interpretations. Together, the quizzes count for 20% of the final grade.

4.3 Exams

The course includes a midterm exam and a final exam. The midterm exam covers the first half of the course, with emphasis on data, exploratory data analysis, and regression. The final exam is comprehensive and asks students to synthesize ideas across the full semester. Each exam contributes 20% to the final grade.

4.4 Research Hackathon

The course culminates in a Research Hackathon, where students apply the concepts and tools developed throughout the semester to a real-world data analysis task. During the final two weeks of the course, students work in teams with a dataset provided in class and develop a focused statistical analysis that integrates the full statistical workflow: formulating a research question, exploring and visualizing data, selecting appropriate methods, interpreting results, communicating uncertainty, and discussing limitations.

The Research Hackathon accounts for 20% of the final grade and consists of the following deliverables:

1. Active participation in the hackathon launch and work sessions.
2. A reproducible Quarto document that includes code, results, interpretation, and discussion of limitations.
3. A short Hackathon Showcase presentation.

The Quarto document and presentation should be developed collaboratively, but each team member must submit the final `.qmd` and PDF files individually on Brightspace. Students should be prepared to explain the code, analysis choices, results, and any AI-assisted components of the project.

Please refer to Tables 2 and 3 for the breakdown of graded activities and the grading scale.

---

5 Course Policies

5.1 Participation, Punctuality, and Attendance

Regular attendance and punctual participation are essential for this course. Many class meetings include in-class activities, quizzes, R Lab problem sets, or collaborative work that cannot be

Table 2: Graded Activities

Evaluation	Percentage	Date(s)
R Lab problem sets	20%	Ongoing
Quizzes	20%	Sep 14, Sep 30, Nov 9
Midterm exam	20%	Oct 21
Final exam	20%	Dec 17
Research Hackathon	20%	Nov 30, Dec 7, Dec 9

Table 3: Grading Scale

A	A-	B+	B	B-	C+	C	C-	D+	D	F
[100 – 93]	(93 – 90]	(90 – 87]	(87 – 83]	(83 – 80]	(80 – 77]	(77 – 73]	(73 – 70]	(70 – 67]	(67 – 60]	(60 – 0]

fully replicated outside of class. Students are expected to arrive on time, bring the required materials, and participate actively in individual, small-group, and class-wide activities.

Quizzes take place during the scheduled class session and begin at the start of class. Students who arrive late may still take the quiz, but they will not receive additional time, except in cases of documented accommodation or excused absence.

R Labs are designed as guided work sessions. Students are expected to use lab time productively, ask questions when they encounter difficulties, and submit their work by the end of the session unless an extension has been approved.

5.2 Feedback and Exit Tickets

Students will have regular opportunities to give feedback on their learning. Short exit tickets may be used at the end of class to identify unclear concepts, remaining questions, or topics that would benefit from additional review. These responses are not graded for correctness; they are used to guide future explanations, examples, and review activities.

Students are encouraged to use the discussion board for questions about course concepts, R code, assignments, or statistical interpretation. The discussion board will be monitored by the instructors. Course-related emails should be directed to the instructors, while students may attend office hours with either the instructors or the professor.

5.3 Make-up and Late Work Policy

Make-up quizzes and exams, as well as extensions for the Research Hackathon, are granted only for documented medical reasons, religious observances, or other approved obligations. Students should contact the instructor as early as possible, and preferably before the scheduled assessment or deadline, to arrange an approved make-up or extension.

5.4 Regrade and Grade Appeal Policy

If a student believes that an assignment, quiz, exam, or project was graded incorrectly, they may request a regrade. To do so, the student should submit a brief written statement explaining the concern and identifying the specific part of the work they would like reviewed.

Regrade requests must be submitted within one week of receiving the grade and feedback. The work will be reviewed again in light of the student's statement. A regrade may result in a grade increase, a grade decrease, or no change.

5.5 Communication

Please use the subject line "Stats: [Your Topic]" when sending emails about the course. Course-related emails should be directed to the instructors. For office hours, students should use the booking links listed in the instructor information section rather than scheduling by email. Students may attend office hours with either the instructors or the professor.

5.6 Generative AI

Students are encouraged to make responsible use of generative AI tools, such as ChatGPT, Claude, Gemini, or similar systems, to support their learning in this course. When used critically, these tools can help clarify concepts, troubleshoot code, improve writing, and support project planning. However, they should complement, not replace, students' own reasoning, analysis, and engagement with the material.

You may use generative AI to help you:

- Clarify or elaborate on key concepts.
- Brainstorm ideas.
- Proofread or edit writing.
- Troubleshoot code or understand R syntax.
- Improve the structure or presentation of project materials.

However, you may not use generative AI to:

- Complete any part of quizzes or exams.
- Substitute for completing required readings.
- Generate responses for class discussions or in-class activities.
- Produce work that you cannot explain or defend as your own.
- Replace your own authorship of the core analysis, interpretation, or argument in graded assignments.

Students are fully responsible for any text, code, output, or interpretation they submit. The instructor may ask students to explain how key parts of their text or code were produced. If a student cannot demonstrate understanding of their submission, or if the submitted code does not reproduce the reported results, this may be treated as a potential violation of academic integrity.

Students must include a brief AI-use statement in the Research Hackathon submission declaring whether generative AI was used and, if so, for what purpose. Honest disclosure of permitted AI use will not negatively affect the grade. Failure to disclose AI use, if detected, will be treated as a potential violation of academic integrity.

5.7 Integrity

At NYU Abu Dhabi, a commitment to excellence, fairness, honesty, and respect within and outside the classroom is essential to maintaining the integrity of our community. By accepting membership in this community, students, faculty, and staff take responsibility for demonstrating these values in their own conduct and for recognizing and supporting these values in others. In turn, these values create a campus climate that encourages the free exchange of ideas, promotes scholarly excellence through active and creative thought, and allows community members to achieve and be recognized for achieving their highest potential.

Students should be aware that those who engage in behaviors that violate the standards of academic integrity will be subject to review and may face the imposition of penalties in accordance with the procedures set out in the NYUAD policy.

6 Course Schedule

Week	Date	Day	Topic
			Topic 1: Introduction to Data
1	Aug 31	M	Hello Data
	Sep 2	W	Knowing the Data
	Sep 4	F	R Lab 1: Getting Started with R
2	Sep 7	M	Sampling from Populations
	Sep 9	W	Designing Studies
	Sep 11	F	R Lab 2: Importing and Inspecting Data
			Topic 2: Exploratory Data Analysis
3	Sep 14	M	**Quiz 1 on Topic 1**
	Sep 16	W	Exploring Categorical Variables
	Sep 18	F	R Lab 3: Visualization of Categorical Data
4	Sep 21	M	Exploring Numerical Variables
	Sep 23	W	Comparing Numerical Distributions Across Groups
	Sep 25	F	R Lab 4: Visualization of Numerical Data and Groups
5	Sep 28	M	Exploring Relationships Between Variables
	Sep 30	W	**Quiz 2 on Topic 2**
	Oct 2	F	R Lab 5: Data Wrangling for Exploratory Analysis
			Topic 3: Regression
6	Oct 5	M	Simple Linear Regression: Prediction and Association
	Oct 7	W	Multiple Regression: Adjusted Associations and Categorical Predictors
	Oct 9	F	R Lab 6: Simple and Multiple Regression
7	Oct 12	M	Interactions and Marginal Effects
	Oct 14	W	Model Diagnostics: Residuals, Outliers, and Assumptions
	Oct 16	F	R Lab 7: Interaction and Diagnostics
			Midterm Exam and Fall Break
8	Oct 19	M	No class
	Oct 21	W	**Midterm Exam on Topics 1–3**
	Oct 23	F	No class
			Topic 4: Foundations of Inference
9	Oct 26	M	Randomization and Hypothesis Testing
	Oct 28	W	Confidence Intervals and Bootstrap Distributions
	Oct 30	F	R Lab 8: Simulation-Based Inference
10	Nov 2	M	Mathematical Models for Inference
	Nov 4	W	Errors, Power, and Practical Significance
	Nov 6	F	R Lab 9: CLT and Formula-Based Inference
			Topic 5: Inference in Practice
11	Nov 9	M	**Quiz 3 on Topic 4**
	Nov 11	W	Inference for Estimates and Comparisons

Week	Date	Day	Topic
	Nov 13	F	R Lab 10: Inference for Estimates and Comparisons
12	Nov 16	M	Inference for Association and Regression
	Nov 18	W	Causality Revisited
	Nov 20	F	R Lab 11: Association, Regression, and Causality
13	Nov 23	M	Models for Different Data Structures
	Nov 25	W	Statistical Reasoning in Practice
	Nov 27	F	R Lab 12: Modeling Different Data Structures
14	Nov 30	M	Hackathon Launch: Teams, Data, and Hypotheses
	Dec 2	W	No class (National Day)
	Dec 4	F	No class (National Day)
15	Dec 7	M	Hackathon Work Session: Analysis
	Dec 9	W	Hackathon Work Session: Analysis
	Dec 11	F	Hackathon Showcase and Submission
Final Assessments			
16	Dec 14	M	Exam Review Session
	Dec 17	Thu	**Final exam**

7 Course Schedule (Detailed)

Session	Topics and Reading
Topic 1: Introduction to Data	
Aug 31 **Hello Data**	• Course structure, objectives, and expectations. • Why statistical thinking matters: how data ground reasoning, evidence, and decisions under uncertainty. • Observations, variables, values, and data matrices. • Reading: *IMS, Chapter 1: Hello data, 1.1, 1.2.1*
Sep 2 **Knowing the Data**	• Making messy data ready for analysis: tidy structure, codebooks, units, and missing values. • Types of variables: numerical (discrete, continuous), categorical (ordinal, nominal). • Relating two variables: explanatory and response, and the distinction between association and independence. • Reading: *IMS, Chapter 1: Hello data, 1.2.2–1.2.4*
Sep 4 **R Lab 1: Getting Started with R**	• Getting set up in Positron: opening a Quarto document and running your first code chunks. • Installing and loading packages, and what it means for an analysis to be reproducible. • Taking a first look at a dataset with `glimpse()`, `count()`, and `summarize()`.
Sep 7 **Sampling from Populations**	• Population, sampling frame, sample, and what each gap means for generalization. • Four sampling methods: simple random, stratified, cluster, and multistage sampling. • How non-response, self-selection, and uneven coverage introduce bias, and the role of survey weights in adjusting for it. • Reading: *IMS, Chapter 2: Study design, 2.1*
Sep 9 **Designing Studies**	• Observational studies vs. experiments, and the kinds of claims each can support. • Random sampling supports generalization; random assignment supports causal inference. • Experimental design: how control, randomization, replication, and blocking work together to isolate an effect. • Confounding and the challenge of drawing causal conclusions from observational data. • Reading: *IMS, Chapter 1: Hello data, 1.2.5; Chapter 2: Study design, 2.2–2.3*
Sep 11 **R Lab 2: Importing and Inspecting Data**	• Importing data, inspecting variable types, and creating a data dictionary. • Running basic summaries and checking data. • Applying descriptive, inferential, and causal claims in context.
Topic 2: Exploratory Data Analysis	
📌 Sep 14 **Quiz 1 on Topic 1**	• Quiz on data basics, study design, sampling, experiments, and observational studies.

Session	**Topics and Reading**
Sep 16 **Exploring Categorical Variables**	• Frequency tables, proportions, and bar plots as ways of seeing the distribution of a single categorical variable. • Contingency tables and conditional proportions as tools for examining how two categorical variables relate. • Comparing categorical distributions across groups. • Reading: *IMS, Chapter 4: Exploring categorical data, 4.1–4.3*
Sep 18 **R Lab 3: Visualizing Categorical Data**	• Getting started with `ggplot2`; building your first bar plots. • Counts vs. proportions, grouped vs. stacked • Making categorical comparisons easy to read: categorical ordering, labels.
Sep 21 **Exploring Numerical Variables**	• Seeing a numerical distribution: center, spread, shape, and outliers. • Describing a numerical distribution: the mean, median, and standard deviation. • Visualizing the distribution: dot plots, histograms, density curves, and box plots. • Transforming a variable when its distribution is strongly skewed. • Reading: *IMS, Chapter 5: Exploring numerical data, 5.2–5.5, 5.7*
Sep 23 **Comparing Numerical Distributions Across Groups**	• Comparing distributions across groups: differences in center, spread, shape, and outliers across groups. • Visualizing comparisons: side-by-side box plots, faceted visualizations. • Summarizing groups: grouped means, medians, and standard deviations. • Reading: *IMS, Chapter 4: Exploring categorical data, 4.6*
Sep 25 **R Lab 4: Visualizing Numerical Data by Group**	• Rendering grouped figures: histograms, box plots, density plots, and faceted layouts. • Computing summary statistics: means, medians, and standard deviations. • Writing short interpretations of numerical comparisons.
Sep 28 **Exploring Relationships Between Variables**	• Visualizing relationships between two numerical variables: scatterplots, (non-)linear patterns, and outliers. • Measuring linear association with the correlation coefficient: what it captures and what it misses. • Simpson's paradox and spurious correlation: how an aggregate relationship can mislead about what is actually going on. • Reading: *IMS, Chapter 3: Applications, 3.2; Chapter 5: Exploring numerical data, 5.1; Chapter 6: Applications*
📌Sep 30 **Quiz 2 on Topic 2**	• Quiz on categorical summaries, numerical summaries, group comparisons, and relationships between variables.
Oct 2 **R Lab 5: Data Wrangling for Exploratory Analysis**	• Reshaping data with `dplyr`: filtering rows, selecting columns, creating new variables, grouping observations, and summarizing. • Building a small exploratory analysis from raw data to interpretation.
Topic 3: Regression	

Session	Topics and Reading
Oct 5 **Simple Linear Regression: Prediction and Association**	• Regression is introduced as a tool for describing and predicting relationships between two numerical variables. • Fitted lines, residuals, least squares, slope, intercept, prediction, and R^2 are used to describe association and assess model fit. • Scalar notation for the simple linear model is introduced, with a brief conceptual preview of outcomes, fitted values, and residuals as vectors. • Regression is previewed as a broader modeling framework for relating predictors to different kinds of outcomes (e.g., binary outcomes). • **Reading:** *IMS, Chapter 7: Linear regression with a single predictor*.
Oct 7 **Multiple Regression: Adjusted Associations and Categorical Predictors**	• Regression is extended to models with multiple predictors, with coefficients interpreted as adjusted associations within the fitted model. • Comparisons between simple and multiple regression are used to clarify the meaning and limits of "holding other variables constant." • Categorical predictors are introduced by showing how regression models compare categories using indicator variables and a baseline category. • Matrix notation is introduced conceptually through design matrices, coefficient vectors, and the linear predictor, with emphasis on interpretation rather than calculation. • **Reading:** *IMS, Chapter 8: Linear regression with multiple predictors; Hanck et al., Chapter 6: Regression Models with Multiple Regressors*.
Oct 9 **R Lab 6: Simple and Multiple Regression**	• Simple and multiple regression models are estimated in R using `lm()`. • Model output, fitted values, residuals, and coefficient estimates are connected to graphical and substantive interpretation. • Continuous and categorical predictors are interpreted in fitted regression models. • Model matrices are inspected to connect R formulas to the structure of the fitted model.
Oct 12 **Nonlinear Terms, Interactions, and Marginal Effects**	• Nonlinear terms, such as polynomial and transformed predictors, are introduced as ways to model curved relationships within the regression framework. • Interactions are introduced as a way to allow regression relationships to differ across groups or across values of another predictor. • Group-specific regression functions are used to show how intercepts, slopes, and fitted curves can vary across groups. • Fitted values, simple slopes, marginal effects, and interaction plots are used to interpret and communicate conditional relationships. • **Reading:** *Hanck et al., Chapter 8: Nonlinear Regression Functions*

Session	Topics and Reading
Oct 14 **Model Diagnostics: Assumptions, Fit, and Influence**	• Regression models are evaluated through assumptions, residual patterns, and model fit. • Linearity, independence, residual variation, and model error are examined using plots and substantive reasoning. • Outliers, leverage, and influential observations are distinguished as different ways observations can affect a model. • Multicollinearity is introduced as a challenge for interpreting individual coefficients in models with related predictors. • Diagnostic results are connected to responsible interpretation and communication of regression findings. • **Reading:** *OpenIntro Statistics, Section 9.3: Checking model conditions using graphs*.
Oct 16 **R Lab 7: Flexible Regression Models and Diagnostics**	• Regression models with nonlinear terms and interactions are estimated and interpreted in R. • Fitted values, group-specific regression functions, marginal effects, and interaction plots are computed and visualized. • Residual plots and diagnostic measures are used to assess model fit, nonlinearity, and influential observations. • A sample midterm is released for practice and review.
Midterm Exam and Fall Break	
Oct 19	• No Class
📌Oct 21 **Midterm Exam on Topics 1–3**	• Covers data, study design, exploratory data analysis, and regression.
Oct 23	• No Class
Topic 4: Foundations of Inference	
Oct 26 **Randomization and Hypothesis Testing**	• Framing a question as a null (no effect) and an alternative hypothesis. • Computing an observed statistic that captures the pattern of interest. • Building a randomization distribution: what the statistic would look like under the null. • Reading the p-value: how unusual the observed result is under the null, and what significance does and doesn't tell us. • Reading: *IMS, Chapter 11: Hypothesis testing with randomization*
Oct 28 **Confidence Intervals and Bootstrap Distributions**	• Computing a point estimate from the sample. • Resampling with replacement to gauge how much the estimate would vary from sample to sample. • Building a bootstrap distribution and reading the standard error and confidence interval. • Interpreting confidence intervals, and spotting common misinterpretations. • Reading: *IMS, Chapter 12: Confidence intervals with bootstrapping*

Session	Topics and Reading
Oct 30 **R Lab 8: Simulation-Based Inference**	• Running a randomization test. • Constructing a bootstrap confidence interval. • Comparing testing and estimation workflows.
Nov 2 **Mathematical Models for Inference**	• Thinking of a statistic as a random variable with a sampling distribution. • Introducing the Central Limit Theorem (CLT) and using it to predict the shape of that distribution. • Locating an observed result on the normal distribution using standard errors and z-scores. • Building confidence intervals and margins of error from the normal distribution. • Reading: *IMS, Chapter 13: Inference with mathematical models*
Nov 4 **Errors, Power, and Practical Significance**	• Distinguishing the two errors a test can make: Type I (false positive) and Type II (false negative). • Why we cannot make both error rates small at once without a larger sample. • Reading power as the probability of detecting a real effect, and what determines it. • Distinguishing statistical significance from substantive importance. • Reading: *IMS, Chapter 14: Decision Errors*
Nov 6 **R Lab 9: CLT and Formula-Based Inference**	• Simulating sampling distributions. • Connecting simulated variability to standard errors and normal approximations. • Formula-based confidence intervals and tests.
Topic 5: Inference in Practice	
📌Nov 9 **Quiz 3 on Topic 4**	• Quiz on randomization, bootstrapping, the CLT, mathematical models for inference, decision errors, and power.
Nov 11 **Choosing Inference Methods for Estimates and Comparisons**	• Inference methods are organized around the research question, study design, and quantity of interest. • Proportions, means, and group differences are analyzed using one-sample, two-sample, and paired designs. • Confidence intervals and hypothesis tests are interpreted as complementary summaries of uncertainty and evidence. • **Reading:** Selected sections from *IMS, Chapters 16–21: Statistical inference.*
Nov 13 **R Lab 10: Inference for Categorical and Numerical Responses**	• Case studies are used to practice inference for categorical and numerical responses in R. • Confidence intervals and hypothesis tests are computed for proportions, means, and group comparisons. • Results are interpreted in relation to the research question, study design, uncertainty, and scope of conclusion.

Session	Topics and Reading
Nov 16 **Inference for Association and Regression**	• Inference for association is considered for both categorical relationships and regression models. • Chi-square methods are used to assess evidence of association between categorical variables in two-way tables. • Regression coefficients are revisited as estimated associations, now with attention to standard errors, confidence intervals, and p-values. • Single-predictor and multiple-predictor regression models are used to distinguish interpretation of coefficients from uncertainty about those estimates. • **Reading:** *IMS, Chapter 18: Inference for two-way tables*; *IMS, Chapter 24: Inference for linear regression with a single predictor*; *IMS, Chapter 25: Inference for linear regression with multiple predictors*.
Nov 18 **Causality Revisited**	• Causal claims are revisited through the distinction between association, prediction, and causal effects. • Confounding, treatment effects, and internal and external validity are discussed as limits on what can be concluded from data. • Potential outcomes are introduced conceptually as a way to express the fundamental problem of causal inference. • **Reading:** *Hanck et al., Chapter 9: Assessing Studies Based on Multiple Regression*; *Hanck et al., Chapter 13: Experiments and Quasi-Experiments*.
Nov 20 **R Lab 11: Association, Regression, and Causal Language**	• Chi-square and regression inference are practiced in R as tools for assessing association. • Results are interpreted with attention to uncertainty, study design, and confounding. • Written conclusions distinguish descriptive, predictive, inferential, and causal claims.
Nov 23 **Regression Extensions for Different Data Structures**	• Regression is revisited as a flexible modeling framework whose extensions depend on the outcome type, research question, and structure of the data. • Logistic regression is introduced as a regression model for binary outcomes • Panel, time-indexed, count, and clustered data are discussed as examples where the regression equation, dependence structure, and interpretation may differ from standard linear regression. • **Reading:** *IMS, Chapter 9: Logistic regression*; *Hanck et al., Chapter 10: Regression with Panel Data; Chapter 11: Regression with a Binary Dependent Variable; Chapter 14: Introduction to Time Series Regression and Forecasting*.
Nov 25 **Statistical Reasoning in Practice**	• Statistical analysis is framed as a sequence from research question to study design, visualization, model choice, inference, and communication. • Method choice, uncertainty, validity, and appropriate scope of conclusion are emphasized. • Prediction, association, and causal interpretation are distinguished as different uses of statistical models. • **Reading:** *IMS, Chapter 23: Applications: Infer*; *Hanck et al., Chapter 9: Assessing Studies Based on Multiple Regression*.

Session	Topics and Reading
Nov 27 **R Lab 12: Modeling Different Data Structures**	• Logistic regression is estimated and interpreted in R as a model for binary outcomes. • Predicted probabilities, model output, and classification-style summaries are used to interpret fitted models. • Model choices and limitations are compared across examples with different outcome types and data structures. • Results are communicated with attention to prediction, uncertainty, and appropriate scope of conclusion.
Research Hackathon	
📌 Nov 30 **Hackathon Launch: Teams, Data, and Hypotheses**	• Students receive the hackathon dataset, form teams, and develop research questions and hypotheses. • Teams draft analysis plans that connect the data context to exploratory analysis, modeling, inference, and communication.
Dec 2	• No Class (National Day)
Dec 4	• No Class (National Day)
📌 Dec 7 **Hackathon Work Session: Analysis**	• Teams conduct exploratory analyses, fit models, conduct inference, and interpret results. • Students focus on reproducible code, appropriate method choice, and clear connections between hypotheses, analyses, and statistical evidence.
📌 Dec 9 **Hackathon Work Session: Analysis**	• Teams refine their analyses through robustness checks and attention to uncertainty and limitations. • Students prepare results for presentation using clear visualizations, defensible conclusions, and appropriately scoped claims.
📌 Dec 11 **Hackathon Showcase and Submission**	• Teams present their research questions, analyses, results, and conclusions. • Teams submit final reports, code, and reflections that emphasize statistical reasoning, uncertainty, and limitations.
Final Assessments	
📌 Dec 14 **Exam Review Session**	• Course concepts are reviewed through an interactive exam review activity (*Jeopardy!*).
📌 Dec 17 **Final Exam**	• Comprehensive final exam. • Covers data, exploratory data analysis, regression, foundations of inference, inference in practice, and statistical reasoning.

8 Appendix: University Resources and Policies

8.1 NYU Moses Center for Student Accessibility

New York University is committed to providing equal educational opportunity and participation for students with disabilities. The Moses Center works with NYU students to determine appropriate and reasonable accommodations that support equal access to a world-class education. Confidentiality is of the utmost importance. Disability-related information is never disclosed without student permission.

Please find further information at the Moses Center for Accessibility and Inclusive Culture, or email mosescenter@nyu.edu.

8.2 Mental Health Resources

As a university student, you may experience a range of issues that can interfere with your ability to perform academically or impact your daily functioning, such as heightened stress, anxiety, difficulty concentrating, sleep disturbance, strained relationships, grief and loss, or personal struggles. If you have any well-being or mental health concerns, please visit the Counseling Center on the ground floor of the Campus Center from 9 AM – 5 PM Abu Dhabi time, Sunday – Thursday. You can also schedule an appointment to meet with a counselor by calling +971 2-628-8100 or emailing nyuad.healthcenter@nyu.edu.

If you require mental health support outside of these hours, call NYU's Wellness Exchange hotline at +971 2-628-5555, which is available 24 hours a day, 7 days a week. You can also utilize the Wellness Exchange mobile chat feature, details of which you can find on the student portal.

8.3 Copyright and Course Materials

All course materials—including slides, recordings, lecture notes, handouts, assignments, and exam questions—remain the intellectual property of the faculty. You may use these materials solely for your own learning and research purposes, with proper citation where appropriate.

You are not permitted to disseminate, post, or share these materials in any form or medium, such as uploading them to external websites or sharing them with students outside the course. Doing so violates intellectual property rights and is subject to disciplinary action by the University under the Code of Student Conduct.