Statistics for Social and Behavioral Sciences

Fall 2026

Course Information

Course Title	Statistics for Social and Behavioral Sciences
Course Code	SOCSC-UH 1010 003
Credits	4
Lecture	Tue & Thu, 3:20 PM – 4:35 PM | Campus Center Room W004
R Lab	Fri, 10:40 AM – 11:55 AM | Campus Center Room 306
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	Tue, 1:00 PM – 3:00 PM; schedule here.
Website	yanwenwang.com

Instructor Information

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

1 Course Description

This course introduces students to the basics of statistics and probability, the building blocks for much of the research done in the social sciences and essential tools for understanding, interpreting, and shaping the world around us. It offers an overview of descriptive and inferential statistics, emphasizing conceptual understanding and statistical thinking, with particular attention to the big picture: how to pose precise questions and draw sound conclusions from data. Throughout, real data from across the social sciences are studied using modern software tools. By the end of the semester, students should understand the difference between populations and samples, what a standard error and a confidence interval mean, how to summarize and analyze a statistical experiment, how to test basic hypotheses, and how to communicate statistical insights.

Expanded 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.
2. **Install Positron:** Download and install the latest version of Positron.
3. **Install Quarto:** Download and install the latest version of Quarto.
4. **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.

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 of three with datasets 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. Students will be provided with several options of real-world datasets.

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

1. Active participation in the hackathon 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.

Table 2: Graded Activities

Evaluation	Percentage	Date(s)
R Lab problem sets	20%	Ongoing
Quizzes	20%	Sep 15, Oct 1, Nov 10
Midterm exam	20%	Oct 22
Final exam	20%	Dec 17
Research Hackathon	20%	Nov 26, Dec 1, Dec 8

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]

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 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 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.

5.3 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.

5.4 Communication

Please use the subject line "Stats: [Your Topic]" when sending emails about the course. For office hours, students should schedule using the booking links, or by email. Students may attend office hours with either the instructor or the professor.

5.5 Generative AI

Students may use generative AI tools to support their learning in this course. These tools can be useful for clarifying concepts, brainstorming ideas, proofreading writing, troubleshooting code, or improving the organization of project materials. However, AI should complement, not replace, students' own reasoning, analysis, and engagement with the course material.

Students may not use generative AI to complete quizzes or exams, substitute for required readings, or generate responses for class discussions or in-class activities. Students are fully responsible for understanding and verifying any material they submit, and should not submit work that they cannot explain, defend, or reproduce. This includes, but is not limited to, the framing of the research question, selection and use of evidence, data analysis, interpretation of results, written argument, code, visualizations, tables, and conclusions. The instructor may ask students to explain key parts of their submission. If a student cannot demonstrate understanding of their work, or if submitted code does not reproduce the reported results, this will affect the evaluation of the work.

5.6 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	Sep 1	Tue	Hello Data
	Sep 3	Thu	Knowing the Data
	Sep 4	Fri	R Lab 1: Getting Started with R
2	Sep 8	Tue	Sampling from Populations
	Sep 10	Thu	Designing Studies
	Sep 11	Fri	R Lab 2: Importing and Inspecting Data
			Topic 2: Exploratory Data Analysis
3	Sep 15	Tue	**Quiz 1 on Topic 1**
	Sep 17	Thu	Exploring Categorical Variables
	Sep 18	Fri	R Lab 3: Visualizing Categorical Data
4	Sep 22	Tue	Exploring Numerical Variables
	Sep 24	Thu	Comparing Numerical Distributions Across Groups
	Sep 25	Fri	R Lab 4: Visualizing Numerical Data by Group
5	Sep 29	Tue	Exploring Relationships Between Variables
	Oct 1	Thu	**Quiz 2 on Topic 2**
	Oct 2	Fri	R Lab 5: Data Wrangling for Exploratory Analysis
			Topic 3: Regression
6	Oct 6	Tue	Simple Linear Regression: Prediction and Association
	Oct 8	Thu	Multiple Regression: Adjustment and Categorical Predictors
	Oct 9	Fri	R Lab 6: Simple and Multiple Regression
7	Oct 13	Tue	Nonlinear Terms, Interactions and Marginal Effects
	Oct 15	Thu	Model Diagnostics: Assumptions, Fit, and Influence
	Oct 16	Fri	R Lab 7: Flexible Regression Models and Diagnostics
			Midterm Exam and Fall Break
8	Oct 20	Tue	No class
	Oct 22	Thu	**Midterm Exam on Topics 1–3**
	Oct 23	Fri	No class
			Topic 4: Foundations of Inference
9	Oct 27	Tue	Randomization and Hypothesis Testing
	Oct 29	Thu	Confidence Intervals and Bootstrap Distributions
	Oct 30	Fri	R Lab 8: Simulation-Based Inference
10	Nov 3	Tue	Inference with Mathematical Models
	Nov 5	Thu	Errors, Power, and Practical Significance
	Nov 6	Fri	R Lab 9: CLT and Formula-Based Inference
			Topic 5: Inference in Practice
11	Nov 10	Tue	**Quiz 3 on Topic 4**
	Nov 12	Thu	Choosing Inference Methods for Estimates and Comparisons

Week	Date	Day	Topic
	Nov 13	Fri	R Lab 10: Inference for Categorical and Numerical Responses
12	Nov 17	Tue	Inference for Categorical Association
	Nov 19	Thu	Inference for Regression
	Nov 20	Fri	R Lab 11: Association, Regression, and Causal Language
13	Nov 24	Tue	Regression Extensions for Different Data Structures
	Nov 26	Thu	Statistical Reasoning in Practice & Hackathon Launch
	Nov 27	Fri	R Lab 12: Modeling Different Data Structures
			Research Hackathon
14	Dec 1	Tue	Hackathon Work Session
	Dec 3	Thu	No class (National Day)
	Dec 4	Fri	No class (National Day)
15	Dec 8	Tue	Hackathon Showcase and Submission
	Dec 10	Thu	No class
	Dec 11	Fri	No class
			Final Assessments
16	Dec 17	Thu	**Final exam**

7 Course Schedule (Detailed)

Session	Topics and Reading
	Topic 1: Introduction to Data
Sep 1 **Hello Data**	• Course structure, objectives, and expectations. • Why statistical thinking matters: how data ground reasoning, evidence, and decisions under uncertainty. • Language of the data: observations, variables, values, and data matrices. • Reading: *IMS, Chapter 1: Hello data, 1.1, 1.2.1*
Sep 3 **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 8 **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 10 **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. • 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 15 **Quiz 1 on Topic 1**	• Quiz on data basics, study design, sampling, experiments, and observational studies.

Session	Topics and Reading
Sep 17 **Exploring Categorical Variables**	• Seeing the distribution of a single categorical variable with frequency tables, proportions, and bar plots. • Examining the relationship between two categorical variables with contingency tables and conditional proportions. • 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 22 **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. • Reading: *IMS, Chapter 5: Exploring numerical data, 5.2–5.5, 5.7*
Sep 24 **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 29 **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*
Oct 1 **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 6 **Simple Linear Regression: Prediction and Association**	• Using regression to describe and predict the relationship between two numerical variables. • Fitting a line by least squares: slope, intercept, residuals, prediction, and what R^2 tells us about fit. • Writing the simple linear model in scalar notation, with a conceptual look at outcomes, fitted values, and residuals as vectors. • Reading: *IMS, Chapter 7: Linear regression with a single predictor*
Oct 8 **Multiple Regression: Adjustment and Categorical Predictors**	• Extending regression to multiple predictors, reading each coefficient as an adjusted association within the fitted model. • Comparing simple and multiple regression to see what "holding other variables constant" does and doesn't mean. • Bringing categorical predictors into a model with indicator variables and a baseline category. • 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**	• Fitting simple and multiple regression models in R with `lm()`. • Connecting model output, fitted values, residuals, and coefficient estimates to graphical and substantive interpretation. • Interpreting continuous and categorical predictors in a fitted model.
Oct 13 **Nonlinear Terms, Interactions and Marginal Effects**	• Modeling curved relationships with nonlinear terms, such as polynomial predictors. • Using interactions to let a relationship differ across groups or across values of another predictor. • Reading group-specific regression functions to see how intercepts, slopes, and fitted curves vary across groups. • Interpreting and communicating conditional relationships with fitted values, simple slopes, marginal effects, and interaction plots. • Reading: *Hanck et al., Chapter 8: Nonlinear Regression Functions*
Oct 15 **Model Diagnostics: Assumptions, Fit, and Influence**	• Checking whether a model's assumptions hold: linearity, constant variance, independence, and roughly normal residuals. • Distinguishing outliers, leverage, and influential observations as different ways a single point can affect a model. • Recognizing multicollinearity as the limit of "holding other variables constant." • Turning diagnostic results into responsible interpretation and communication of regression findings. • Reading: *OpenIntro Statistics, Section 8.3: Types of outliers in linear regression; Section 9.3: Checking model conditions using graphs*
Oct 16 **R Lab 7: Flexible Regression Models and Diagnostics**	• Fitting and interpreting regression models with and without nonlinear terms and interactions in R. • Computing and visualizing fitted values, group-specific regression functions, marginal effects, and interaction plots. • Using residual plots and diagnostic measures to assess fit, nonlinearity, and influential observations.
Midterm Exam and Fall Break	

Session	Topics and Reading
Oct 22 **Midterm Exam on Topics 1–3**	• Covers data, study design, exploratory data analysis, and regression.
	Topic 4: Foundations of Inference
Oct 27 **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 29 **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*
Oct 30 **R Lab 8: Simulation-Based Inference**	• Running a randomization test. • Constructing a bootstrap confidence interval. • Comparing testing and estimation workflows.
Nov 3 **Inference with Mathematical Models**	• 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. • Reading confidence intervals and margins of error from the normal distribution. • Reading: *IMS, Chapter 13: Inference with mathematical models*
Nov 5 **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 10 **Quiz 3 on Topic 4**	• Quiz on randomization, bootstrapping, the CLT, mathematical models for inference, decision errors, and power.

Session	Topics and Reading
Nov 12 **Choosing Inference Methods for Estimates and Comparisons**	• Choosing an inference method from the research question, study design, and quantity of interest. • Introducing the t-distribution for means, with the standard deviation estimated from the sample. • Analyzing proportions, means, and group differences with one-sample, two-sample, and paired designs. • Reading confidence intervals and hypothesis tests as complementary summaries of uncertainty and evidence. • Reading: *IMS, Chapter 16: Inference for a single proportion, 16.2; Chapter 17: Inference for comparing two proportions, 17.3; Chapter 19: Inference for a single mean, 19.2; Chapter 20: Inference for comparing two independent means, 20.3–20.4; Chapter 21: Inference for comparing paired means, 21.3*
Nov 13 **R Lab 10: Inference for Categorical and Numerical Responses**	• Practicing inference for categorical and numerical responses in R through case studies. • Computing confidence intervals and hypothesis tests for proportions, means, and group comparisons. • Interpreting results in relation to the research question, study design, uncertainty, and scope of conclusion.
Nov 17 **Inference for Categorical Association**	• Extending inference to association between two categorical variables. • Using the chi-square test for two-way tables: expected counts under independence, the chi-square statistic, and degrees of freedom. • Reading the result as evidence for or against independence, and what it does and doesn't establish. • Reading: *IMS, Chapter 18: Inference for two-way tables*
Nov 19 **Inference for Regression**	• Revisiting regression coefficients as estimated associations, now with standard errors, confidence intervals, and p-values. • Comparing single- and multiple-predictor models to separate interpreting a coefficient from quantifying uncertainty about it. • Reading: *IMS, Chapter 24: Inference for linear regression with a single predictor; Chapter 25: Inference for linear regression with multiple predictors*
Nov 20 **R Lab 11: Association, Regression, and Causal Language**	• Practicing chi-square and regression inference in R as tools for assessing association. • Interpreting results with attention to uncertainty, study design, and confounding. • Writing conclusions that distinguish descriptive, predictive, inferential, and causal claims.
Nov 24 **Regression Extensions for Different Data Structures**	• Revisiting regression as a flexible framework whose extensions depend on the outcome type, research question, and structure of the data. • Introducing logistic regression as a model for binary outcomes. • Touring further extensions: panel, time-indexed, count, and clustered data as cases where the regression equation, dependence structure, and interpretation 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*

Session	**Topics and Reading**
Nov 26 **Statistical Reasoning in Practice & Hackathon Launch**	• Statistical reasoning in practice: from research question to study design, visualization, model choice, inference, scope of conclusion, and communication. • **Hackathon launch:** students receive the dataset, form teams, and develop research questions, hypotheses, and analysis plans. • Reading: *IMS, Chapter 23: Applications: Infer*; *Hanck et al., Chapter 9: Assessing Studies Based on Multiple Regression*
Nov 27 **R Lab 12: Modeling Different Data Structures**	• Fitting and interpreting logistic regression in R as a model for binary outcomes. • Using predicted probabilities, model output, and classification-style summaries to interpret a fitted model. • Comparing model choices and limitations across examples with different outcome types and data structures. • Communicating results with attention to prediction, uncertainty, and appropriate scope of conclusion.
	Research Hackathon
Dec 1 **Hackathon Work Session**	• Teams conduct exploratory analyses, fit models, and interpret results with attention to uncertainty and limitations. • Students prepare results for presentation with reproducible code, clear visualizations, and appropriately scoped conclusions.
Dec 8 **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 17 **Final Exam**	• Comprehensive final exam.

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 The Writing Center

Located in the NYUAD Library, the Writing Center is an excellent resource for you to use throughout your university career. Consultants will meet with you to discuss your writing for any writing project. They don't do the work for you—they don't edit or proofread your work or give you ideas—but they can help you figure out what you need to do in order to improve your writing. To register as a client and to schedule an appointment, go to https://nyuad.mywconline.com. It is wise to schedule appointments well in advance because appointment slots fill up.

8.4 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.