STATS 101 - Introduction to Applied Statistical Methods
Fall 2025, Session 2

click here for latest version
Section 1434 (lecture)
Section 1435 (lab)

Instructor:	Qian Wu
Office Location:	IB2017
E-mail:	qian.wu2@dukekunshan.edu.cn

Class Dates:	10/20/2025 – 12/04/2025
Lecture Time:	MoWe 12:00PM - 2:30PM
Lecture Location:	AB 1087
Office Hours:	MoWe 9:30AM – 11:30AM and by appointment

Lab Manager:	Jingyu Wang
E-mail:	jingyu.wang@dukekunshan.edu.cn

Lab Time:	Tu 1:15PM - 2:30PM
Lab Location:	IB 2028
Office Hours:	Fr 10:00AM - 12:00PM in WDR 2134

Course Description:

How can we use data to shed light on age-old and new human problems such as pollution, discrimination, and economic growth? How can we be "sure" that the evidence we have points us in the right direction? How meaningful are our findings? Do our results suggest the relationships we find between factors such smoking and cancer are meaningful or meaningless? How would we know? How should one properly display and explain your statistical results to these important issues?

This class introduces you to the tools and concepts that begin to tackle these questions. We will cover topics such as data summaries, sampling, data analysis, production of graphical displays, and regression techniques. The goal at the end of the course is that you will be able to conduct basic data manipulation, know how to properly summarize and display data, and make basic statistical inferences using real datasets.

A third set of goals for the course is that you will also be able to read more fluently research literature that employs statistics. During this course I will reference, in class, a number of historically important academic articles and we will analyze the data from those articles. Doing so should help you understand how data is used (and misused) to construct social science arguments.

Required Background:

The emphasis in the course will not be on learning mathematical formulas related to statistics but rather to develop an intuitive understanding of statistical inference and measures of uncertainty. For those interested in developing this intuitive understanding of statistics in a more rigorous way, you may also consider taking the following courses:

- MATH 205: the mathematical foundations of statistics.
- ECON 203: advanced study of modern regression techniques.
- SOSC 320: advanced statistical techniques applied to real-world problems.

Course Objectives:

Upon completing the course, you will develop the following abilities:

- Intuitively interpret statistics in course materials and in the larger world.
- Become a statistics results producer in addition to a statistics consumer.
- Assess when and how to use statistics to answer specific questions in the social sciences.
- Analyze how previously learned problems can be answered with statistical methods.
- Apply statistical methods to future social science coursework and capstone project.
- Judge how appropriately statistics are used in everyday life when reading the news, business reports, and other real-world applications.

In support of this you will be able to:

- Understand and interpret basic statistical properties of data (confidence intervals, t-tests, etc).
- Identify when various statistical tests are appropriate given a specific dataset.
- Formulate testable hypotheses in the data and learn how to execute those tests.
- Interpret statistical results to understand both significance of the results and their substantive impact.
- Illustrate statistical results with appropriate and clear graphical displays that provide meaning to the reader.
- Evaluate critically other, published, statistical work with the skills and techniques learned in class.
- Propose an independent research project that integrates statistical methods with their research interest for their capstone project.

Course Structure:

In general, each week will proceed roughly as follows:

- Monday-Thursday: Read the textbook chapter, make progress on the lab(s) and homework.

- Monday&Wednesday: Class session with a conceptual review of the chapter material.
- Tuesday: Lab activities designed to help gain familiarity with the technical aspects of using R, and RStudio.
- Monday&Wednesday: Office hours.
- Friday: Complete DataCamp lab before midnight on Friday.
- Sunday: Complete the homework that is due before midnight on Sunday.

How to Prepare:

We will be using three different online tools for this class:

- Class Canvas Websites: https://canvas.duke.edu/courses/64773. I will use Canvas for posting assignments, class announcements, grades, and other materials, so please make sure to read these messages carefully and check Canvas regularly.
- Programming Tools: We will use R statistical programming language for conducting statistical investigations.
  - Language: R. Please download R from here.
  - Environment: RStudio. After installing R, please download RStudio from here.
  - Publishing System: Quarto. Fter installing RStudio, you can install Quarto by typing "quarto install tinytex" in the terminal.
- Online Learning Platform: DataCamp. DataCamp has a lot of very useful tutorials that will help you learn how to code in R. **Note: you must sign up with your @dukekunshan.edu.cn email otherwise you will not receive the free access provided to us. You can access DataCamp from here.

Required Textbook:

Intro Stats, 6th Edition by Richard D. De Veaux, Paul F. Velleman, and David E. Bock.

Assignments:

- Pop Quizzes (10%): There will be 10 equally weighted unannounced quizzes throughout the semester. **Quizzes may occur in both lectures and labs.** They are designed to encourage consistent engagement and help you stay current with the material. You earn full credit if you answer $\geq$60% of questions correctly. Your two lowest quiz grades will be dropped.
- Unit 1&2 homework (35%): At the end of the first two groups of content, a homework will be assigned that will ask you to analyze a dataset and answer questions related to that concept group. These assignments are always due the Sunday at 11:59 pm China time.
  - Unit 1 homework: 15%.
  - Unit 2 homework: 20%.

- Unit 3 exam (20%): Unit 3 comprehension is better checked through an in-class exam. The exam will be open notes but closed book and will take place during the normal lab session that week.
- DataCamp Labs (10%): There will be 5 equally weighted online DataCamp labs introducing you to key components and packages in R. These assignments are always due Friday at 11:59 pm China time.
- Individual Final Project (25%): The final project is a chance to show that you can apply the statistical skills we have learned to real-world questions. You will be given a dataset and submit a research report, which has 8-12 pages, double-spaced, including tables and figures. It is due on December 9th (Tuesday) at 11:59 pm China time.
- Extra credit maximum 5%: Extra credit opportunities may be offered occasionally. These are optional and designed to reward participation, but they will not significantly affect your final grade.

Late Submissions:

- Pop Quizzes & Individual Final Project: No late submission is allowed.
- Other Assignments: Late submission is accepted within 48 hours of the deadline with a penalty. For each 24-hour period after the deadline, a 10% flat deduction will be applied to the total possible points for that assignment, regardless of your earned score. After 48 hours, the submission will not be accepted. [1]

Attendance Policy:

I do not take attendance. However, if you are not in class for the pop quiz, there is no opportunity to make it up. I urge you not to abuse the opportunity to miss a class without penalty unless your situation absolutely requires to. Topics in this course are interrelated and build on each other. So, missing one class has a potential to set you back in future classes. I strongly encourage you to review the lecture slides and read the textbook if you missed a class.

Contact Policy:

Please contact me via email or Canvas message. I generally respond within 48 hours. However, during busy times, especially when deadlines are approaching, I may receive a high volume of messages, so please allow extra time for a reply.

AI Policy:

You may use AI tools (such as ChatGPT, DeepSeek, or similar) to support your learning in this course. However, all submitted work must be your own original writing and analysis.

[1] For example, the total possible points for unit 2 homework are 20%. If you submit it 10 hours after the deadline, you will receive a 20%*10%=2% deduction. If you submit it 28 hours after the deadline, you will receive a 20%*(2*10%)=4% deduction. If you submit it 48 hours after the deadline, your grade is 0.

You may not copy and paste AI-generated content into your final project report, quizzes, or other graded assignments. If you use AI tools to brainstorm or check your work, you are responsible for verifying the accuracy and originality of your final submission.

Academic Integrity:

In addition to skills and knowledge, DKU aims to teach you appropriate ethical and professional standards. You will find extensive information on academic integrity here.

In line with these policies, dishonesty of any kind will not be tolerated in this course. Dishonesty includes, but is not limited to, cheating, plagiarizing, fabricating information, facilitating acts of academic dishonesty by others, having unauthorized possession of examinations, submitting work of another person or work previously used without informing the instructor, or tampering with academic work of other students. Whenever in doubt, ask me about appropriateness of your actions.

Disability Policy:

If you need an accommodation due to a disability, you should not hesitate to request one. The process is that requests should be sent to the Dean of Undergraduate Studies, who will contact me with recommended type of accommodation that is needed. You do not need to disclose your reason for requesting an accommodation with me, and asking through the Dean of Undergraduate Studies helps make things official for both you and me.

Tentative Schedule[2]

Unit 1: Distributions

Week 1

- Lecture 1.1: Intro to the course (Monday, Oct 20)
  - Reading: Chapter 1, 2.1 and 2.2, and 3
  - Topics:
    - Syllabus
    - What are data and variables?
    - How to display quantitative and qualitative variables?
    - Contingency tables
- Lab 1: R and Quarto familiarization (Tuesday, Oct 21)
- Lecture 1.2: Characteristics of distributions (Wednesday, Oct 22)
  - Reading: Chapter 2.3-5 and 4
  - Topics:
    - How to describe the shape, center, and spread of a distribution?
    - How to compare distributions?
    - Dealing with problems (outliers, reexpression)
- To-dos:
  - DataCamp lab 1: Introduction to R and Tidyverse (due Friday, Oct 24 at 11:59 pm)

Week 2

- Lecture 2.1: Normal distribution (Monday, Oct 27)
  - Reading: Chapter 5
  - Topics:
    - Standard deviation and standardizing values
    - Normal models
    - Normal percentiles
- Lab 2: Advanced Quarto editing (Tuesday, Oct 28)

Unit 2: Relationships between variables

Week 2

[2]Please notice that this schedule is tentative and subject to changes. All changes will be announced via Canvas.

- Lecture 2.2: Association and correlation (Wednesdat, Oct 29)
  - Reading: Chapter 6
  - Topics:
    * Scatterplots
    * Correlations
    * Does correlation imply causation?
- To-dos:
  - DataCamp lab 2: Introduction to Data Visualization with ggplot2 (due Friday, Oct 31 at 11:59 pm)
  - Homework assignment 1 (due Sunday, Nov 2 at 11:59 pm)

Week 3

- Lecture 3.1: Simple linear regression (Monday, Nov 3)
  - Reading: Chapter 7
  - Topics:
    * Line of best fit: least squares
    * The linear model
    * What are residuals?
    * Regression assumptions
- Lab 3: Working with regressions using dplyr (Tuesday, Nov 4)
- Lecture 3.2: Regression wisdom (Wednesday, Nov 5)
  - Reading: Chapter 8
  - Topics:
    * Beware extrapolation
    * Outliers and leverage
    * Lurking variables
    * Straightening scatterplots
- To-dos:
  - DataCamp lab 3: Intermediate Data Visualization with ggplot2 (due Friday, Nov 7 at 11:59 pm)

Week 4

- Lecture 4.1: Multiple regression (Monday, Nov 10)
  - Reading: Chapter 9
  - Topics:
    * What is multiple regression?
    * Interpreting multiple regression coefficients
    * Partial regression plots

        * Indicator variables

- Lab 4: Interpreting coefficients (Tuesday, Nov 11)
- Lecture 4.2: Confidence intervals for proportions (Wednesday, Nov 12)
  - Reading: Chapter 13
  - Topics:
    * What is a sampling distribution?
    * When does the normal model apply?
    * Constructing a confidence interval
    * Interpreting a confidence interval
- To-dos:
  - DataCamp lab 4: Modeling with Data in the Tidyverse (due Friday, Nov 14 at 11:59 pm)
  - Homework assignment 2 (due Sunday, Nov 16 at 11:59 pm)

Unit 3: Measuring uncertainty

Week 5

- Lecture 5.1: Confidence interval for means (Monday, Nov 17)
  - Reading: Chapter 14
  - Topics:
    * The Central Limit Theorem
    * Confidence interval for means
    * Interpreting a confidence interval
    * Final thoughts on confidence intervals
- Lab 5: Bootstrapping (Tuesday, Nov 18)
- Lecture 5.2: Hypothesis testing (Wednesday, Nov 19)
  - Reading: Chapter 15
  - Topics:
    * What are hypotheses?
    * P-values
    * How to make decisions based on p-values?
- To-dos:
  - DataCamp lab 5: Your choice of any DataCamp course (due Friday, Nov 21 at 11:59 pm)

Week 6

- Lecture 6.1: Hypothesis testing wisdom (Monday, Nov 24)

  - Reading: Chapter 16
  - Topics:
    * Interpreting p-values
    * Alpha and critical values
    * Practical vs. statistical significance
    * Type I and II errors
    * Power of a test
    * Ethical issues

- Lab 6: Unit 3 exam (Tuesday, Nov 25)

Unit 4: Statistical inference

Week 6

- Lecture 6.2: Comparing groups (Wednesday, Nov 26)
  - Reading: Chapter 17
  - Topics:
    * Confidence intervals for comparing two samples
    * Assumptions and conditions for two-sample hypothesis tests
    * Two-sample z test
    * Two-sample t test

Week 7

- Lecture 7.1: Returning to regression (Monday, Dec 1)
  - Reading: Chapter 20
  - Topics:
    * Regression inference and intuition
    * The regression table
    * Confidence and prediction intervals
- Lab 7: Independent learning for individual final project (Tuesday, Dec 2)
- Lecture 7.2: Interpretation activity & model building practice (Wednesday, Dec 3)
  - Topics:
    * How to read academic statistical results
    * Locating the model
    * Interpreting the test
    * Determining possible weaknesses of the model
- To-dos:
  - Individual final project (due Tuesday, Dec 9 at 11:59 pm)