



THE OHIO STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

STAT 5301: Intermediate Data Analysis I

Spring 2022

Instructor: Nasser Sadeghkhan	Time: M W 8:00–9:50 AM
Email: sadeghkhan.1@osu.edu	Place: SH0235

Office Hours: F 4:00-5:00 pm or by appointment. My office is located at 205B Cockins Hall.

TA: Ghosh Amartya (ghosh.147@osu.edu)

Course Description:

- STAT5301 is the first course in a two-semester pre-calculus sequence in data analysis covering descriptive statistics, design of experiments, probability and statistical inference including one-sample and two-sample problems, goodness of fit and one-way ANOVA. The goals are for students to develop skills in drawing conclusions and critically evaluating results based on data. This course satisfies the General Education (GE) requirement in Data Analysis.

Course learning outcomes:

- Students understand basic concepts of statistics and probability, comprehend methods needed to analyze and critically evaluate statistical arguments, and recognize the importance of statistical ideas.

Course delivery:

- This course is the **four credit hours** which meets **Mondays, and Wednesdays from 8:00 am to 9:50 am at Stillman Hall 235**. Lectures will be delivered **in-person class** during the scheduled class meeting times.
- Students are expected to attend and participate in these **in-person class** meetings. **Please arrive on-time**. Class meetings will be used to provide in-depth investigation of the topics for the week using a lecture format. Students will participate in these class sessions by engaging in discussions prompted by the instructor and by asking and answering questions. Students should plan to take notes during class. **No recorded version** of lectures are provided.

Main References:

- Introduction to the Practice of Statistics (Eighth Edition)
by D.S. Moore, G.P. McCabe, and B.A. Craig.
This book is not required, but may be useful as a reference for the first half of the course. We will only cover Chapters 1-9 of this book.

- The Statistical Sleuth: A Course in Methods of Data Analysis (Third Edition) by F.L. Ramsey and D.W. Shafer.

This text is required for the second half of the course and is also used in STAT 5302, the second course in the Intermediate Data Analysis sequence. STAT 5301 will cover Chapters 1-6 of the book.

- Access this eBook through the CarmenBooks reader link in the course navigation.

Prerequisites: Math 1075 (104) or equivalent, or Math Placement Level of R, or permission of instructor. The sequence is intended for students with limited formal mathematics background. However, in terms of data analysis and interpretation, the conceptual level of the course is high.

Necessary software

- This class requires you to use the statistical software package called R (The R Project for Statistical Computing; <http://www.r-project.org/>) to illustrate certain aspects. Here is the information for obtaining R.
 - You can download R for Windows, Mac, and Linux, from the CRAN archive at <https://cran.r-project.org>.
 - An in-depth introduction to R is available at <http://cran.r-project.org/doc/manuals/R-intro.pdf>
- An easier to use interface to R is available in the software package RStudio. This package is available for Windows, Mac, and Linux and can be downloaded for free from <http://rstudio.org>. Note that RStudio requires R to be installed.

Course Requirements:

You are responsible for all material covered in class; this includes derivation, proofs, computational techniques, etc. This is an applied course and the emphasis will be on applying concepts learned in class to real-world datasets. This is not a purely computational course although you will learn how to use software to analyze data and apply concepts learned during the lectures. I will use the statistical computing language R in class to demonstrate ideas and examples.

Evaluation: The evaluation will be determined based on **six** assignments, **two** midterm exams, and **one** final exam.

Assignments 20%, First Midterm 25%, Second Midterm 25%, Final 30%.

Grades will be recorded on [Carmen](#)

Important Dates:

Assignments	Almost biweekly (on Carmen)
Midterm #1	W–Feb 16 (in class)
Midterm #2	W–Mar 23 (in class)
Final Exam	Th Apr 28, 8:00-9:45 AM

Grading Scale:

93–100	A
90–92.9	A-
87–89.9	B+
83–86.9	B
80–82.9	B-
77–79.9	C+
73–76.9	C
70 –72.9	C-
67 –69.9	D+
60 –66.9	D
Below 60	E

Course Policy:

- Lecture notes are partial, and you need to add some material during classes. So try not to miss classes!
- The lecture notes will be released one week in advance of every session on [Carmen](#).

Assignment Policy:

- The assignments will be posted on [Carmen](#).
- You answers must be uploaded electronically through Carmen in a single PDF file.
- Pay attention to the due date, it is usually one week after the assignment is posted. No late assignment will be accepted. If you are unable to complete an assignment on time, please get in touch with me asap, so we can discuss your situation.
- You are encouraged to work together on the homework, but **do not** copy any part of a assignment.
- For the Homework assignments that require R, Make sure that the computer output and discussion are placed together (do not put the computer output at the end of homework). Raw computer output is not acceptable. Make it clear what parts of the output are relevant and show how they answer the questions posed in the homework.
- Homework assignments that do not require R may be handwritten (electronically, or on paper and scanned) and uploaded.

Exam Policy:

- All exams are closed book/closed notes. A basic calculator is required – tablets, laptops, and cellphones are not.
- Midterm 1 covers the material up to and including Mon 14 Feb. Midterm 2 covers the material up to and including Mon 21 Mar.
- All exams do not test on R programming.
- The exams will be cumulative, but will emphasize the more recent material. There will be **no make-up** exams. If exceptional circumstances (sudden onset of illness, unexpected family situations, etc.) arise, contact me asap, so we can discuss your situation.

Class Policy:

- **Arrive to class on time.**

- Be courteous when using mobile devices. Make sure your cell phone is turned fully off, or silent. No texting, reading emails, playing games, or whatever else it is that people do with those wretched gizmos.
- If you must use a laptop in class, then turn off the sound and do not type on laptop keyboards which is really distracting.
- Missing one class could easily lead to a disastrous domino effect. If you have to miss a lecture, then I strongly recommend you study the material you missed before you return to class. Do not hesitate to come by and ask me. I require that you know all material covered in class. You are responsible for making up anything that was covered in lectures you missed.

Health and safety:

The Ohio State University Wexner Medical Center's Coronavirus Outbreak site (<https://wexnermedical.osu.edu/features/coronavirus>) includes the latest information about COVID-19 as well as guidance for students, faculty and staff.

Student academic services:

Student academic services offered on the OSU main campus <http://advising.osu.edu/welcome.shtml>.

Student support services:

Student support services offered on the OSU main campus <http://ssc.osu.edu>.

Ohio State's academic integrity policy:

It is the responsibility of the Committee on Academic Misconduct to investigate or establish procedures for the investigation of all reported cases of student academic misconduct. The term "academic misconduct" includes all forms of student academic misconduct wherever committed; illustrated by, but not limited to, cases of plagiarism and dishonest practices in connection with examinations. Instructors shall report all instances of alleged academic misconduct to the committee (Faculty Rule 3335-5-487). For additional information, see the Code of Student Conduct <http://studentlife.osu.edu/csc/>.

Copyright disclaimer:

The materials used in connection with this course may be subject to copyright protection and are only for the use of students officially enrolled in the course for the educational purposes associated with the course. Copyright law must be considered before copying, retaining, or disseminating materials outside of the course.

Accessibility accommodations for students with disabilities:

- The University strives to make all learning experiences as accessible as possible. If you anticipate or experience academic barriers based on your disability (including mental health, chronic or temporary medical conditions), please let me know immediately so that we can privately discuss options. To establish reasonable accommodations, I may request that you register with Student Life Disability Services. After registration, make arrangements with me as soon as possible to discuss your accommodations so that they may be implemented in a timely fashion. SLDS contact information: slds@osu.edu; 614-292-3307; <https://slds.osu.edu>; 098 Baker Hall, 113 W. 12th Avenue.

Statement on title IX:

- Title IX makes it clear that violence and harassment based on sex and gender are Civil Rights offenses subject to the same kinds of accountability and the same kinds of support applied to offenses against other protected categories (e.g., race). If you or someone you know has been sexually harassed or assaulted, you may find the appropriate resources at <http://titleix.osu.edu> or by contacting the Ohio State Title IX Coordinator at titleix@osu.edu.

Course topics/Schedule

Below are tentative topics to be covered. The schedule is one session on each topic. Note that they are tentative and may be subject to change.

1. Exploratory data analysis (1): graphical summary tools; R basics (1)
2. Exploratory data analysis (2): numerical summary tools; R basics (2)
3. Simple linear regression
4. Data collection: sampling and experimental design
5. Probability (1): basic concepts and rules; conditional probability
6. Probability (2): random variables (types, distributions); expectation; variance
7. Probability (3): sampling distributions; large sample distribution; CLT
8. Confidence interval; concepts and interpretation
9. One sample Z test; concepts and steps; approx. Z test; use CI to test
10. One sample t-test; power; abuse of tests; Bonferroni correction
11. Two independent sample Z and t-tests (pooled);
12. Unequal population variances: non-pooled t-test; F test
13. Robustness of the t-procedure; assumption assessment; data transformation
14. Inference for proportions: two sample tests and CI; power calculation
15. Nonparametric tests (one sample): sign/signed rank/permutation tests
16. Nonparametric tests (two sample): Wilcoxon rank sum/permutation tests
17. Pearson's χ^2 test for goodness-of-fit; contingency table
18. Separate population mean model; one-way ANOVA
19. Multiple comparison: Bonferroni; Tukey-Kramer; Contrasts; Scheffé
20. Two-way ANOVA; interaction effect
21. General model comparison; summary of one-way multiple comparison