Stat 5301 – Intermediate Data Analysis I   Autumn 2023

Instructor R. Scott Linder (linder.5)

Office Hours
Tuesday 3:30–5pm   Wednesday 5–6:30pm.   Thursday 2–3:30pm   Friday 5–6pm

Grader/TA
Yuxuan Xin (xin.155)   Office hours to be determined

Course Description
Stat 5301 is the first course in a two-semester sequence on data analysis methods. This class covers descriptive methods (numerical and graphical), methods of good experimental and survey sampling design, probability and sampling distributions, as well as inferential methods, including methods of estimation (confidence intervals), significance testing for one and two sample problems. In addition, methods for analyzing categorical data and methods for comparing several means (ANOVA) are discussed. A specific list of topics to be covered can be found in the tentative schedule below.

Course prerequisite
This class is not a calculus-based treatment of Statistics. A strong background in algebra is sufficient mathematical preparation. That said, the class presents deep, conceptually challenging ideas in Statistics (learning to think statistically is challenging).

Course delivery
This course will be delivered asynchronously and online. Course material will be distributed through approximately 4 hours of recorded lessons each week.

Online meeting obligations for scheduled exam sessions
There are three times during which you are **obligated** to be in attendance in a Zoom meeting session:
- Exam 1 (Thursday, October 5 4–5:30pm),
- Exam 2 (Thursday, November 16, 4–5:30pm), and
- Final Exam (Tuesday, December 12, 7:30pm–9:30pm).
These exam times and dates will not change so that all students can plan ahead.

Makeup exams will be scheduled <u>only</u> for (1) official OSU conflicts (e.g. you have a class scheduled at the time of an exam, you're travelling with an OSU sports team, etc.); or (2) documented emergency that prevents you from taking the exam (e.g. a medical emergency). For either scenario, documentation will be required.

Non-OSU conflicts (e.g. a job shift) will not merit rescheduling an exam. Again, these exam dates are scheduled at the beginning of class so that you can plan around them.

Course learning outcomes

By end of semester, students should successfully be able to

- recognize sources of potential bias that might arise when collecting data in an experimental or survey sampling context;
- outline the components of an experiment by applying randomization, blinding, replication and blocking principles effectively;
- describe a list of numbers (a sample or a population) graphically and numerically;
- interpret commonly used metrics (mean, median, proportion, odds ratio, relative risk, standard deviation, range, coefficient of variation, etc.) and understand their limitations and caveats;
- recognize a statistical problem from a non-statistical problem, so that the student knows when statistical methods may be helpful or necessary to achieve an understanding;
- understand the role that probability plays in statistical reasoning and in stataistical inference;
- recognize which kind of statistical analysis might be appropriate for a given context;
- know how to carry out a statistical analysis, and how to interpret the results of the analysis in context of the research problem;
- understand that statistical analyses require that certain assumptions are valid – know what the assumptions are for various analyses, and how to look for evidence that they are not valid.

GE Course Information

This course satisfies the General Education foundation requirement in *Mathematical and Quantitative Reasoning or Data Analysis* which has the following goals and expected learning outcomes:

Goals: Successful students will be able to apply quantitative or logical reasoning and/or mathematical/ statistical methods to understand and solve problems and will be able to communicate their results.

Expected Learning Outcomes (ELOs): Successful students are able to:

1.1 Use logical, mathematical and/or statistical concepts and methods to represent real-world situations.

1.2 Use diverse logical, mathematical and/or statistical approaches, technologies and tools to communicate about data symbolically, visually, numerically and verbally.

1.3 Draw appropriate inferences from data based on quantitative analysis and/or logical reasoning.

1.4 Make and evaluate important assumptions in estimation, modeling, logical argumentation and/or data analysis.

1.5 Evaluate social and ethical implications in mathematical and quantitative reasoning.

This course also satisfies the Legacy General Education requirement in *Data Analysis*. which has the following goals and expected learning outcomes:

Goals: Students develop skills in drawing conclusions and critically evaluating results based on data.

Expected Learning Outcomes:
1. Students understand basic concepts of statistics and probability.
2. Students comprehend methods needed to analyze and critically evaluate statistical arguments.
3. Students recognize the importance of statistical ideas.

Course Textbooks
The required textbook for this course is provided via CarmenBooks. Through which students obtain publisher materials electronically through CarmenCanvas, saving money. The fee for this material is included as part of tuition and is listed as CarmenBooks fee on your Statement of Account. In addition to cost-savings, materials provided through CarmenBooks are available immediately on or before the first day of class. Unless you chose to opt-out of the program, you do **NOT** need to purchase any materials for this course at the bookstore or online. For more information on the program or information on how to opt out, please visit the CarmenBooks website.

Required textbook:
The Statistical Sleuth: A Course in Methods of Data Analysis, 3rd Edition, by Ramsey and Schafer (2012), Cengage Learning • ISBN-13:978-1-133-49067-8

Access this eBook through the CarmenBooks reader link in the course navigation.
Note: This textbook is also required for Stat 5302.

Optional and recommended textbook:
Introduction to the Practice of Statistics (any edition 5 or higher), by David Moore and George McCabe.

Instructor note about these textbooks:
The optional book (*IPS*) is an easy to read, straightforward introductory textbook that should be very helpful for about the first half of the class. The required text (*Sleuth*) is a little denser, but contains a lot of useful information. *Sleuth* will be more useful the second half of class.
Reading suggestions from both texts will be provided.

Course Technology

For help with your password, university e-mail, Carmen, or any other technology issues, contact the OSU IT Service Desk. Standard support hours are available at https://ocio.osu.edu/help, and support for urgent issues is available 24x7:

- Self-Service and Chat support: http://ocio.osu.edu/selfservice
- Phone: 614-688-HELP (4357)
- Email: 8help@osu.edu
- TDD: 614-688-8743

Baseline technical skills necessary for online courses

- Basic computer and web-browsing skills
- Navigating Carmen
- CarmenZoom

Necessary hardware

For day-to-day work, you need a computer (Mac running OSX or PC running Windows 10+) with a high-speed internet connection. For the exam meetings and for office hour meetings, your computer must have a webcam. For office hour meetings, your computer must have a microphone.

Necessary software

This class requires use of the statistical software package R. This software is available for free download. (The R Project for Statistical Computing; https://cran.r-project.org or https://cran.rstudio.com/).

An in-depth introduction to R is available at https://cran.r-project.org/doc/manuals/R-intro.pdf

Hands-on tutorials are available in the Swirl system, which you can learn about at http://swirlstats.com/ In particular, "R Programming: The basics of programming in R" is an appropriate first tutorial for students who have never used R.

A (arguably more) user-friendly interface to R is available via the software package RStudio (recently re-branded as Posit). This package is available for Windows, Mac, and Linux and can be downloaded for free from https://posit.co/downloads/ . Note that RStudio requires R to be installed first.

More details about R will be provided through class lectures and handouts.

Microsoft Office 365 ProPlus

All Ohio State students are now eligible for free Microsoft Office 365 ProPlus through Microsoft's Student Advantage program. Each student can install Office on five PCs or Macs, five tablets (Windows, iPad® and AndroidTM) and five phones. Students are able to access Word, Excel, PowerPoint, Outlook and other programs, depending on platform. Users will also receive 1 TB of OneDrive for Business storage. Office 365 is installed within your BuckeyeMail account. Full instructions for downloading and installation can be found https://ocio.osu.edu/kb04733 .

Course Grade

Your course grade will be determined by your total course average (TCA), which will be computed according to the following weights:

Homework average	30%
Exam 1	20%
Exam 2	20%
Final Exam	30%

You are guaranteed a course grade at least as high as that determined by your TCA according to the following scale:

92% ≤ TCA ≤ 100%	A	78% ≤ TCA < 80%	C+
90% ≤ TCA < 92%	A–	72% ≤ TCA < 78%	C
88% ≤ TCA < 90%	B+	70% ≤ TCA < 72%	C–
82% ≤ TCA < 88%	B	60% ≤ TCA < 70%	D
80% ≤ TCA < 82%	B–	TCA < 60%	E

Hypothetical example A student's class record follows
In class there were 8 assignments: 85, 80, 92, 90, 94, 90, 88 and 95
  ➔ The lowest assignment score (80) gets dropped, so this student's average assignment score is (85+92+90+94+90+88+95)/7 = 90.57%

Exam 1: 85%   Exam 2: 89%   Final Exam: 84%

TCA = 90.57(.3) + 85(.2) + 89(.2) + 84(.3) = 87.17%
This student will earn a grade of **at least** B. Depending on how the distribution of class TCA values looks at the end of the term, this TCA **might** earn a B+ or an A–.

About homework

There will be 7 to 10 assignments during the semester. Assignments will be posted around the same time that new recorded lessons are posted. You should start the assignments only after you've learned material by watching the lessons and reading the textbook… but you should leave yourself enough time to complete the assignment while having time to get help, if it's needed. Solutions must be written neatly, and your work should be well organized. Many assignment problems require use of R, and for these you need to provide output documenting your work. **All assignments must be submitted to the Canvas portal as a single file in pdf format**. Assignments submitted after solutions have been posted will not be accepted at all. Otherwise, late assignments will be accepted only with instructor permission. Your lowest assignment score will be dropped.

About exams

During exams you'll work through problems that do not require use of R, but which may require you to interpret R output that is provided.

There are three times during which you are **obligated** to be in attendance in a Zoom meeting session:
- Exam 1 (Thursday, October 5 4–5:30pm),
- Exam 2 (Thursday, November 16, 4–5:30pm), and
- Final Exam (Tuesday, December 12, 7:30pm–9:30pm).

These exam times and dates will not change so that all students can plan ahead.

Makeup exams will be scheduled only for (1) official OSU conflicts (e.g. you have a class scheduled at the time of an exam, you're travelling with an OSU sports team, etc.); or (2) documented emergency that prevents you from taking the exam (e.g. a medical emergency). For either scenario, documentation will be required.

Early exams will never be given. Any makeup exam must occur within 96 hours of the regularly scheduled exam period. If a student has a documented reason for being unable to take an exam within 96 hours of the regularly scheduled exam period (e.g. the student has been medically incapacitated for an extended period), then the missed exam will be vacated, and the student's TCA will be computed by weighting the final exam for 50%.

Copyright Disclaimer and Fair Use

The instructor holds the rights to all recorded lessons, assignment files, solution set files, exam files, and practice exam files. You may download them for your own use while you're a student in the class. At no time are you allowed to share any of these materials with others, or post them at a location that others can access. In short, these materials are created by the instructor for your use only, and for your use only while you're a student in the class.

Other Policies and resources

<u>Health and Safety</u>

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. Guidelines and requirements for campus safety from the University's Reactivation Task Force are published on the Safe and Healthy website (https://safeandhealthy.osu.edu).

<u>Student Academic and Support Services</u>

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

Student support services offered on OSU main campus: https://artsandsciences.osu.edu/academics/current-students/resources .

Mental health services offered on OSU main campus:

Office of Student Life's Counseling and Consultation Service (CCS): ccs.osu.edu ccs.osu.edu or calling 614- 292-5766. CCS is located on the 4th Floor of the Younkin Success Center and 10th Floor of Lincoln Tower. You can reach an on-call counselor when CCS is closed at 614-292-5766, and 24 hour emergency help is also available through the 24/7 National Suicide Prevention Hotline at 1-800-273-TALK or at suicidepreventionlifeline.org .

Accessibility Accommodations for Students

The university strives to make all learning experiences as accessible as possible. If you anticipate or experience academic barriers based on a disability (including mental health, chronic, or temporary medical conditions), please let the instructor know immediately so that we can privately discuss options. To establish reasonable accommodations, you will be asked to register with Student Life Disability Services. After this registration, arrangements will be made with you quickly in order to provide all appropriate accommodations in a timely fashion. SLDS contact information: slds@osu.edu ; 614-292-3307; http://slds.osu.edu ; 098 Baker Hall, 113 W. 12th Avenue.

Religious Accommodations

It is Ohio State's policy to reasonably accommodate the sincerely held religious beliefs and practices of all students. The policy permits a student to be absent for up to three days each academic semester for reasons of faith or religious or spiritual belief.

Students planning to use religious beliefs or practices accommodations for course requirements must inform the instructor in writing no later than 14 days after the course begins. The instructor is then responsible for scheduling an alternative time and date for the course requirement, which may be before or after the original time and date of the course requirement. These alternative accommodations will remain confidential. It is the student's responsibility to ensure that all course assignments are completed.

Academic Integrity Policy

Policies for this course
Exams: You must complete all exams yourself, without any external help or communication. As described above, exams are to be taken during scheduled Zoom sessions during which your web camera is to be on. Your phone must be out of sight and out of reach during these sessions. You must remain in the exam session until the instructor verifies that your exam has been submitted. You may not communicate with anybody except the instructor (through Zoom 'chat') while the exam is in session.

Homework: Homework may be discussed with classmate or TAs but submitted homework assignments should represent your own effort. Copying the work of another (even if you rewrite it as your own work) is not acceptable. A conversation with a classmate that leads to your successful solution is.

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 https://trustees.osu.edu/bylaws-and-rules/code .

All incidents of academic misconduct will be reported to the Committee on Academic Misconduct.

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, Melissa Mayhan, at titleix@osu.edu

Disclaimer
This syllabus should be taken as a fairly reliable guide for course content and policy. However, you cannot claim any rights from it where dates and assessments are concerned. In particular, the instructor reserves the right to change due dates or methods of grading and/or assessing. Policies listed under Course Grade (above) will not change. Any other changes made will be communicated through Canvas.

Tentative Course Schedule

Week 1 T 8.22 – Su 8.27	Induction vs. deduction – what makes a problem statistical? Basic terminology (population vs. sample; parameter vs. statistic); Random vs. nonrandom error (precision vs. accuracy); Types of bias.
Week 2 M 8.28 – Su 9.3	Observational study vs. experiments; Elements of basic experiments (randomization, blinding, blocking); Methods for selecting a random sample (SRS, stratified, cluster, systematic); What makes for a good blocking or stratification variable.
Week 3 M 9.4 – Su 9.10	Histograms and distribution shapes; Measures of center (mean and median); Boxplots; Measures of spread (range, standard deviation); Interpreting the standard deviation with the mean; Notion of degrees of freedom (why n–1?); Linear transformations and impact on mean and standard deviation.
Week 4 M 9.11 – Su 9.17	Means and standard deviations of linear combinations of independent RVs; Mean and standard deviation of a sample mean; Notion of a sampling distribution.
Week 5 M 9.18 – Su 9.24	The Normal distribution model; Sampling distributions; Central Limit Theorem for sample mean and sample proportion.
Week 6 M 9.25 – Su 10.1	Confidence interval for population mean and population proportion based on CLT; Sample size required to estimate a mean or proportion with specified precision and confidence.
Week 7 M 10.2 – Su 10.8	Confidence interval for population mean based on t-distribution; Robustness of one-sample t-distribution method; Normal probability plot; Transformation; **Exam 1 (R 10.5 4:00pm–5:30pm)**
Week 8 M 10.9 – Su 10.15	Introduction to significance tests – p-value for testing a mean. **Fall Break R 10.12 and F 10.13**

Week 9 M 10.16 – Su 10.22	One-sample significance tests for mean and proportion – specifying hypotheses; Two error types; Selecting a level of significance based on which type of error is more problematic; Computing the p-vale, and reaching a decision in context.
Week 10 10.23 – Su 10.29	Power of significance test; Introduction to two-sample methods for comparing means – pooled and unpooled methods.
Week 11 M 10.30 – Su 11.5	Pooled and unpooled two-sample t-based confidence intervals and significance tests for comparing means; Nonparametric alternative method – Wilcoxon rank sum test.
Week 12 M 11.6 – Su 11.12	Inference and issues for categorical data: Marginal vs. conditional associations and Simpson's paradox; Chi-square test for homogeneity or independence; Inference for relative risk and odds ratio.
Week 13 M 11.13 – Su 11.19	Comparing several means – ANOVA; Sums of squares and degrees of freedom; The F-statistic and the F-test. **Exam 2 (R 11.16 4:00–5:30pm)**
Week 14 M 11.20 – Su 11.26	F-test for several means; The problem of multiple comparisons – the Bonferroni and Tukey methods. **Break around Thanksgiving W 11.23 – F 11.25**
Week 15 M 11.27 – Su 12.3	ANOVA (continued); The problem of multiple comparisons – Bonferroni and Tukey methods; If time allows, Introduction to Bayesian estimation.
Week 15 M 12.4 – R 12.7	Bays estimation; Exam Review. Note: Material on Bayes estimation will be covered if time allows. If we need this week to get caught up, then Bayes estimation will be omitted.
T 12.12	**Final Exam 7:30pm – 9:30pm (Covers course material through ANOVA)**