Dr. Tristan Whalen
OPRE/BUAN 6359.010 Fall 2022
Advanced Statistics for Data Science

The country is hungry for information; everything of a statistical character, or even a statistical appearance, is taken up with an eagerness that is almost pathetic; the community have not yet learned to be half skeptical and critical enough in respect to such statements.

—General Francis A. Walker, Superintendent of the 1870 census
(Quoted in Freedman, Pisani, and Purves *Statistics,* 4ed)

INSTRUCTOR EMAIL & OFFICE

Dr. Tristan Whalen tristan.whalen@utdallas.edu JSOM 3.420

I encourage you to email me throughout the semester with questions or concerns about the class. Please include **6359.010** in every email to me.

My Fall 2022 office hours at JSOM 3.420

Monday 1:00pm to 3:45pm
Tuesday, Thursday, Friday: 11:30am to 12:45pm
No appointment needed; just come by during listed time. No office hours on holidays.

CLASSROOM LOCATION & TIMES

JSOM 1.217 Fridays 1:00PM – 3:45PM

Class meetings will consist of lecture, examples, and discussion. Most homework and exam materials are based on what we cover in class, so regular attendance and participation are expected.

CALENDAR & HOLIDAYS

Last day of classes: Thursday, Dec 8 Withdrawal dates: Sept 7 (without a W), Nov 8
Labor Day: Monday, Sept 5 Fall Break: Monday, Nov 21 to Wednesday, Nov 23
Thanksgiving Break: Thursday, Nov 24 to Sunday, Nov 27

TA: Sourabh Vitthal Shelke svs200004@utdallas.edu

COURSE MATERIAL, TEXTBOOKS & SOFTWARE

>You need a **laptop computer** to bring to class on exam days.
>This course currently requires *The Statistical Sleuth* (**Ramsey**, Schafer, 3rd edition). Data files for the examples and exercises can be found at www.statisticalsleuth.com
>The recommended supplemental textbooks are *Statistics for Management and Economics* (**Keller**, 11th edition) and *The Book of R* (**Davies**, 1st edition). Reference these during Module 1 as needed.
>My favorite book is *Statistics*, **Freedman**, Pisani, Purves, 4th edition. It is an undergraduate, introductory-level text, but the core concepts still apply, and few other sources explain statistical thinking so well.
>This course uses R and **RStudio**. Both are free to download and install. First download and install R. Then, download and install RStudio Desktop. RStudio provides a more beginner-friendly interface to work with an already-installed R.
https://cran.r-project.org/
https://www.rstudio.com/products/rstudio/download/

"Education never ends, Watson. It is a series of lessons with the greatest for the last."
—Sherlock Holmes (from Arthur Conan Doyle's *The Red Circle*)

MODULE 1: Fundamentals of Probability and Statistics

Aug 26	Syllabus and course overview, installing R Descriptive statistics fundamentals
Sept 2	Probability interpretations, axioms, basic calculations Conditional probability and Bayes' Rule
Sept 9	Probability distributions: Binomial, Poisson, Uniform, Normal, Exponential, Gamma
Sept 16	The Central Limit Theorem and sampling distributions
Sept 23	Confidence intervals for unknown averages and proportions Tests of a null hypothesis
Sept 30	Exam for Module 1 in class – bring your laptop

MODULE 2: Advanced Statistical Inference

Oct 7	[Ch.1] Controlled experiments, observational studies, and statistical conclusions [Ch.2] Using Student *t*-distributions
Oct 14	[Ch.3.1-4] Pros, cons, and underlying assumptions of Student t-tools [Ch.3.5-6] Logarithmic and other transformations of data
Oct 21	[Ch.5] Comparing multiple samples with Student and F-distributions [Ch6.4-5] Other comparison methods
Oct 28	Exam for Module 2 in class – bring your laptop

MODULE 3: Simple and Multiple Regression

Nov 4	[Ch.7] Simple linear regression and inferences (intervals and tests) [Ch.8] Underlying assumptions of simple linear regression models
Nov 11	[Ch.9.2-4] Multiple regression [Ch.10] Inferential tools for multiple regression, especially with F-distributions
Nov 18	[Ch.11] Model checking and refinement
Nov 25	Thanksgiving break
Dec 2	[Ch.12] Strategies to narrow down the set of explanatory variables
TBA	Exam for Module 3 during final exam time – bring your laptop

**Note: dates are tentative and may be adjusted if necessary. Chapters refer to Ramsey text.*

COURSEBOOK DESCRIPTION

This course uses statistical methods to analyze data from observational studies and experimental designs to communicate results to a business audience. The course mandates prior knowledge of fundamental statistical concepts such as measures of central location, standard deviations, histograms, the normal and t-distributions (knowledge of calculus is not required). The course also emphasizes interpretation and inference, as well as computation using a statistical software package such as R or STATA.

COMMENTS

>Active and informed participation is expected from every student. Class sessions will consist primarily of lectures, with some discussions and in-class exercises as appropriate to the covered topic. Students are expected to read the assigned readings in preparation for class and exams.
>Data analysis course involves plenty of both concepts and computation, and you will learn as much when things go wrong as when they go right. ("Get messy! Make mistakes!" as Ms. Frizzle says.)
>This is not an R course. We will use R or RStudio as needed, but the main goals of the course are the statistics concepts and calculations.

LEARNING OBJECTIVES

1. Develop and test hypotheses using multiple statistical methods.
2. Understand differences between observational and experimental studies.
3. Learn how randomization and sampling influence scope of inference.
4. Explore experimental and observational designs that compare multiple populations when the response is continuous or binary.
5. Communicate the findings of a statistical analysis from these new methods in a clear, concise, and scientific manner.
6. Integrate and analyze real-world datasets using common software packages.

GRADE CALCULATION
30% Homework
22% Exam 1
23% Exam 2
25% Exam 3
To keep grading fair, requests for special treatment, including extra credit or rounding exceptions, will be ignored.

A	$X \geq 93\%$
A-	$90\% \leq X < 93\%$
B+	$87\% \leq X < 90\%$
B	$83\% \leq X < 87\%$
B-	$80\% \leq X < 83\%$
C+	$75\% \leq X < 80\%$
C	$70\% \leq X < 75\%$
F	$X < 70$

The content of this syllabus may change at the instructor's discretion.

EXAMS
>The three exams are noncumulative except where topics necessarily build on each other.
>On the exam dates listed above, bring your laptop computer to the classroom. You will take the exam in elearning.
>Open notes and open book, but individual work only. No collaboration with anyone else about the exam. No communication about the exam with anyone before, during, or after, except the instructor or the TA.
>If an emergency (such as an illness) prevents you from taking the exam, notify the instructor immediately.
>No make-up exams will be offered for your own personal matters (such as family trips or weddings) or other non-emergency reasons.
>If you miss the exam and do not contact the instructor before the end of the exam day, you will receive a grade of zero for the missed exam.

HOMEWORK
>Around 7 homework sets will be posted and submitted via elearning.
>Late submissions are not accepted for any reason.
>Homework sets will be labeled “Group” or “Individual.”
>Individual homework sets must be completed on your own.
>Group homework sets may be completed individually or in a group of up to 4 people total.
>Each member of the group will upload and submit the same document, and this document must have the names of all group members clearly typed on the cover page.

EXAM & HOMEWORK GENERAL POLICIES
>Check your UTD email regularly for announcements sent via elearning.
>No extra credit or bonus projects will be given. Complete the given work, plan ahead, and be vigilant!
>Submit only your own work or, where allowed, your group’s own work.
>Evidence of academic dishonesty (such as copying work, sharing or getting exam material, or communicating about the exam in online platforms) will be referred to the appropriate dean’s office. If you are found guilty, as far as this course is concerned, you will receive a score of zero on your submission.
>Read https://www.utdallas.edu/conduct/dishonesty
>Also see the student code of conduct: https://policy.utdallas.edu/utdsp5003

CLASSROOM POLICIES
>In general, please treat others the way you want them to treat you.
>Please show courtesy and charity to other students and the instructor. Focus on the lecture and raise your hand to contribute.
>Avoid leaving early or arriving late (except for emergencies).
>At the instructor’s discretion, you may be asked to leave the classroom and/or receive a grade penalty for behavior that interferes with class.
>You are expected to attend class and participate regularly.
>Skipping class and not participating will bring you lots of stress and difficulties. Instead, always come to class, sit as close to the front as possible, and participate regularly (in class, in email, or in office hours).

GENERAL SYLLABUS STUFF

>>University restrictions about class recordings

Students are expected to follow appropriate University policies and maintain the security of passwords used to access recorded lectures. Unless the Office of Student AccessAbility has approved the student to record the instruction, **students are expressly prohibited from recording any part of this course**. The instructor's recordings may not be published, reproduced, or shared with those not in the class, or uploaded to other online environments except to implement an approved Office of Student AccessAbility accommodation. If the instructor or a UTD school/department/office plans any other uses for the recordings, consent of the students identifiable in the recordings is required prior to such use unless an exception is allowed by law. Failure to comply with these University requirements is a violation of the Student Code of Conduct.

>>University restrictions about class materials (lecture note files, instructor notes, solutions, etc.)

The materials posted by the instructor may be downloaded during the course; however, **these materials are for registered students' use only**. Classroom materials may not be reproduced or shared with anyone not in the class, or uploaded to other online environments except to implement an approved Office of Student AccessAbility accommodation. Failure to comply with these University requirements is a violation of the Student Code of Conduct.

>>University technical requirements and help

In addition to a confident level of computer and Internet literacy, certain minimum technical requirements must be met to enable a successful learning experience. Please review the important technical requirements on the Getting Started with eLearning webpage: https://ets.utdallas.edu/elearning/students/current/getting-started

This course can be accessed using your UT Dallas NetID account on the eLearning website. Please see the course access and navigation section of the Getting Started with eLearning webpage for more information. To become familiar with the eLearning tool, please see the Student eLearning Tutorials webpage.

UT Dallas provides eLearning technical support 24 hours a day, 7 days a week. The eLearning Support Center includes a toll-free telephone number for immediate assistance (1-866-588-3192), email request service, and an online chat service.

The University is committed to providing a reliable learning management system to all users. However, in the event of any unexpected server outage or any unusual technical difficulty which prevents students from completing a time sensitive assessment activity, the instructor will provide an appropriate accommodation based on the situation. Students should immediately report any problems to the instructor and contact the online eLearning Help Desk. The instructor and the eLearning Help Desk will work with the student to resolve any issues at the earliest possible time.

>>UT Dallas syllabus policies and procedures:

The information contained in the following link constitutes the university's policies and procedures segment of the course syllabus.

https://go.utdallas.edu/syllabus-policies

>>Comet Creed

This creed was voted on by the UT Dallas student body in 2014. It is a standard that Comets choose to live by and encourage others to do the same:

"As a Comet, I pledge honesty, integrity, and service in all that I do."

>>Academic Support Resources

The information contained in the following link lists the University's academic support resources for all students. Please go to Academic Support Resources webpage for these policies:
https://provost.utdallas.edu/syllabus-policies/#academic-support-resources

"That is the one eternal education: to be sure enough that something is true that you dare to tell it to a child."

—G. K. Chesterton (from the book *What's Wrong with the World*)