

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 **your course and section numbers, 6359.5w1**, in every email to me.

My Spring 2024 office hours, in person at JSOM 3.420:

Mondays: 2:00pm to 5:00pm, no appointment needed.

Wednesdays and Thursdays by appointment (request by email)

No office hours on holidays, school closures, or after classes end.

CLASSROOM LOCATION & TIMES

Online in Microsoft Teams Mondays 7:00pm to 9:45pm

See the UTD Academic Calendar for holidays and registration deadlines. If UTD cancels classes for winter weather, our class is canceled, too.

HARDWARE & SOFTWARE

>Webcam, to be activated and displaying your face during class. From coursebook: **“Enrollment in the section necessitates the student is online and participating in class with their cameras on.”**

>MS Teams for online class.

>Regular and reliable access to **elearning**.

>**Microsoft Word** or similar software to compose solutions to the mini projects and paste the visuals created in R.

>We use **R** throughout the course. R is a powerful statistics and data software that is free (monetarily) to download and install: <https://cran.r-project.org/>

>After installing R, you will probably want to download and install **RStudio** Desktop. RStudio provides a more user-friendly interface on an already-installed R: <https://posit.co/download/rstudio-desktop/>

COURSE MATERIALS

Whalen Statistics: My own suite of notes (PDF files), lecture slides (PowerPoint files), and R scripts, the result of many, many hours of work, I will share with you in elearning. I will also provide, via elearning, several data files (ready to use in R, so you don't need Excel) for the various examples.

Statistics, Freedman, Pisani, Purves, 4th edition. “Freedman”

The only statistics textbook I like. Entirely optional, but I do recommend reading this along with the corresponding topics (listed below). If you're scared of statistics, you really ought to start here.

Freedman explains the fundamentals and discusses deep concepts better than any other book I've seen. Lots of practice exercises with solutions, especially for the concepts. The only downside is that the book does not cover every topic that our course requires.

PREREQUISITES

>Absolutely NO prior experience with R

We begin R “from scratch.” With that said, I assume you have experience downloading, installing, and using software, typing, preparing documents, saving and loading files in folders on your computer, using elearning, copying and pasting graphs into a document, and things of that sort.

>Proficiency with routine mathematics.

Arithmetic with sums, products, powers, roots, fractions, formulas. Plotting and reading graphs. Reading tables of numbers. Equalities, inequalities, intervals. Functions.

>Acquainted with the ideas and concepts of calculus.

Primarily the concepts of “area under a curve,” derivatives, and limits; hopefully you have seen The Fundamental Theorem of Calculus. Multivariable calculus will help (but is not necessary) for multiple regression. You will not have to do any calculus by hand—R will do that for you.

>Familiarity with undergraduate probability and statistics.

We cover all the following topics, but it will help if you already know about them: averages, standard deviations, histograms, medians, and boxplots. Basic probability calculations with multiplication, addition, and complements. (Think dice, coins, marbles, and cards.) Mutual exclusiveness and independence. Random variables and probability distributions, including some named distributions like the Binomial distributions and the Normal distributions.

TOPIC LIST

“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*)

#	Whalen Statistics Topic	Freedman Reference
01	Experiments vs. Observations <i>RCTs, blinding, observational studies, confounding factors</i>	Freedman Ch.01-02
02	Shape, Center, and Spread <i>Histograms, the average of a list, the SD of a list, z-scores, empirical rule, Cheybshev inequality, medians, boxplots</i>	Freedman Ch.03-04
03	Normal Approximation to Data <i>Properties of normal curves, compare to data histograms, qq-plots, skew, kurtosis</i>	Freedman Ch.05
04	Analyzing Variance with Sums of Squares! <i>Descriptive ANOVA: within-group, between-group, and total sum of squares; coefficient of determination; comparing one-average and multiple-average summaries</i>	n/a
05	Scatterplots, Correlation, and Regression <i>Summarizing the relationship (or lack thereof) between two variables, using and interpreting a regression line, regression effect, least-squares method, RMS-residual, coefficient of determination as r-squared, log-transforms</i>	Freedman Ch.08-12
06	Regression Planes and Fallacies <i>Multiple regression introduced with regression planes, interpreting regression coefficients, indicator and interaction terms, regression parabola; correlation does not imply causation, beware of various types of extrapolation</i>	Freedman Ch.08-12 (for the fallacies; for correlation vs. causation, see esp. ch.09, for the dangers of extrapolation, see esp. ch.11)
07	Multiple Regression Model Building <i>Residual plots, stepwise variable selection, model selection criteria (such as Akaike)</i>	Freedman Ch.11-12 (for residual plots, see ch.11; for a brief but helpful discussion of multiple regression, see end of ch.12)
08	Probability: from Basics to Bayes’ Rule <i>Interpretations, multiplication principle and independence, addition rule and mutual exclusiveness, complement rule; conditional probability, Bayes’ rule</i>	Freedman Ch.13-14
09	Random Variables <i>Probability distributions and probability histograms, determining a distribution using direct calculation, using simulation, and using prefab templates (such as Binomial)</i>	Freedman Ch.15-16

10	Expected Value, Standard Error, and Normal Approximations to Probability <i>Definition and mathematical properties of Expectation and SE; the Central Limit Theorem (CLT)</i>	Freedman Ch.17-18
11	Statistical Inference <i>History lessons about sampling, random sampling model, confidence intervals, z-intervals, interpreting confidence</i>	Freedman Ch.19-21, 23-24
12	Probability Tests <i>Basic probability tests, null hypothesis and chance model, z-test, t-test</i>	Freedman Ch.26
13	More Probability Tests <i>Two-sample z-test, two-sample t-test, multi-sample F-test, F-test and model comparison, issues and jargon</i>	Freedman Ch.27, 29 (for two-sample z-test and concepts; the very important ch.29 discusses issues with tests)

ABOUT THE ORDER OF TOPICS

Lesson 01 establishes the difference between descriptive and inferential statistics.

Lessons 02-07 are about descriptive statistics and apply to practically any data set.

- *We couch analysis in terms of “lists of numbers” and do not distinguish between population and sample until after probability. In particular, we defined the “SD of a list” to be the root-mean-square deviation from the average of the list. These are correct both mathematically and practically for descriptive statistics.*
- *We include ANOVA and regression before and without probability for theoretical, practical, and pedagogical reasons—*
 - *Theoretical: sums of squares and the least squares method (with all the related calculations) do not require any probability assumptions; it’s just geometry and calculus. The corresponding tests, like t-tests and F-tests, usually do not apply, and often do not contribute any additional information anyway; these ideas are discussed in my notes.*
 - *Practical: regression models give useful insight into observational data and powerful descriptive summaries that don’t have anything to do with probability.*
 - *Pedagogical: there’s plenty to cover in regression before getting bogged down with probability.*

Lessons 08-10 are about probability, necessary for understanding statistical inference.

- *We adhere to this convention: expected value and standard error apply to random variables. By contrast, average and standard deviation apply to lists of numbers (data).*
- *The relationship: take a list of numbers. Draw one at random. The average of the list equals the expected value of the number you will draw, and the SD of the list equals the SE of the number you will draw.*

Lessons 11-13 are about inferential statistics and apply only to data gathered by a probability method.

ADVICE FOR SUCCESS

>Attend every class.

This is true for any college course, regardless of the presentation skill of the instructor. If you care about your grade and GPA, at a minimum you should never skip class. Attend every session, put away your smartphone, and actively participate (even just listening). These actions train your brain to take the course seriously, and they encourage the presenter, thus improving the quality of the class.

>Build time into your weekly schedule for the course outside of class time.

This is true for any college course. Set aside at least 3 hours per week outside of class session, depending on your skill level. As a starting point, I suggest 1.5 hours per week for reading the textbooks, 1.5 hours per week for reviewing notes and practicing R, and 2.5 hours per week for working the homework. You can adjust based on your individual abilities. If study and practice are not planned into your schedule, then they will not happen, and the exams will be very hard.

>Email the instructor early and often.

Do not let your first email to the instructor be about a low grade, and certainly do not let your first contact with the instructor be after the last exam. Instead, email the instructor as soon as you have any question, concern, or confusion. Make your first email about a course topic and not about a grade or what will be on a test.

>Make emails useful, brief, and courteous.

Read Dr. Whalen's email guide, posted in elearning alongside the syllabus.

>Expect to make mistakes, get frustrated, and try repeatedly.

We learn as much when things go wrong as when they go right. ("Get messy! Make mistakes!" as Ms. Frizzle says.) All of us (students and instructors) will encounter errors and mistakes both in and out of R.

>This is primarily a *statistics* class.

Stay focused on the provided materials and avoid getting lost in R scripting tutorials. Learn the concepts and not just the calculations. I choose examples to explain the statistics concepts, and I do not restrict myself to a business setting. (That said, there are lots of examples that should be of interest to business majors.)

>Take care of yourself.

This is true for any work. Your physical, mental, emotional, and spiritual health are interrelated. Get plenty of sleep each night on a regular schedule. Maintain a regular diet. Limit your time in social media. (I guarantee that, if you go at least 4 weeks without using any social media, you will notice an improvement in your physical and mental wellbeing. How often do you come away from social media sessions feeling refreshed, rejuvenated, and enlightened?) Spend time with friends and/or family in person. Engage in wholesome activities and entertainment.

GRADE CALCULATION

30% Online homework average	A	$X \geq 93\%$
10% Histogram mini project	A-	$90\% \leq X < 93\%$
10% Regression mini project	B+	$87\% \leq X < 90\%$
10% Probability mini project	B	$83\% \leq X < 87\%$
10% Online Quiz 1 (Sums of Squares)	B-	$80\% \leq X < 83\%$
10% Online Quiz 2 (Regression)	C+	$75\% \leq X < 80\%$
10% Online Quiz 3 (Inference)	C	$70\% \leq X < 75\%$
10% Online Quiz 4 (F-tests)	F	$X < 70$

ASSIGNMENT POLICIES

Online Homework: Sets of problems to be completed and submitted via elearning, usually within about one week from the date assigned.

- These may require the use of R.
- When more than one attempt is allowed, we use your highest score.
- There are no homework deadline extensions.
- Your attempt must be submitted before the elearning deadline. Late submissions will get a score of 0%, even if they're only a minute late, and even if they're because of technical issues, so allow at least 12 hours in advance.

Mini Projects: Each project is one long homework problem (usually about a single data set) with multiple parts. You submit a one-page PDF document (by upload into elearning) that includes a nice plot or graph that you constructed in R, along with your R script.

- Each individual student must submit his or her own work.
- Everything you need to do has an example R script in my notes, that you may copy and adapt to the project.
- Other sources are not allowed except by permission.
- If you are caught cheating (such as copying from an online source or another student), you will get a score of 0% on the project.
- Do not email your project. If your submission is not uploaded to elearning by the deadline, then you get a score of 0% on the project, even if it's because of technical issues.
- Make sure to submit the correct file, your own file, in the proper format (PDF), with all the required information, else you may get a score as low as 0%.

Online Quizzes: Take online in elearning. Dates and time frames to be announced in class (such as, “this Saturday between 10:00AM and 10:00PM.”). Begin the quiz at home during the assigned time frame.

Once you start, your timer starts, and the time limit will be tight. No backtracking.

- You are allowed to use R, your own personal notes, and any of Dr. Whalen’s notes/slides/scripts.
- No collaborating and no other sources allowed. If you are caught cheating, copying, or collaborating, you will get a score of 0% on the quiz.
- You are responsible for technical issues. If we do not have a submission from you (in elearning) before the deadline, you will get a score of 0% for the quiz.
- **Online Quiz 01:** Lesson 04, sums of squares analysis, requires R.
- **Online Quiz 02:** Lessons 05 and 06, regression, including topics from Lessons 01-04 as they apply to regression.
- **Online Quiz 03:** Lessons 11-12, including topics from Lessons 09-10 as they apply to inference.
- **Online Quiz 04:** Lesson 13, multi-sample probability tests (esp. F-test), including topics from previous lessons as they apply to Lesson 13. Requires R.

>Read <https://www.utdallas.edu/conduct/dishonesty>

>Also see the student code of conduct: <https://policy.utdallas.edu/utdsp5003>

FREQUENTLY ASKED QUESTIONS

Do you curve grades? When we finish grading a quiz or project, I always look at the score distribution and summary, including how many people got an A. If the median or average are low (depending on the course), I always adjust the scores by a constant shift in the positive direction. I do not call this “curving.” This is a statistics class, so note two things: if you do ask about curves, you ought to specify what curve you have in mind (there are lots: bell curves, exponential curves, logarithmic curves, Pareto curves,...). Also, you don’t *really* want me to *curve* grades, because this might imply lowering your score, along with some others, to get the histogram to fit under a curve.

Is there extra credit? I build some bonus points into many of my assignments, sometimes not revealed until the end of the semester, so make sure to complete all the work I assign. Otherwise, please do not ask for extra credit, bonus projects, extra attempts, and the like. These requests will be ignored to ensure fairness to the whole class. Instead, do the assigned work.

What if I have to miss a quiz? Notify the instructor *immediately* if an emergency prevents you from taking a quiz. (For example, you wake up too sick to leave home and send me an email.) We can usually work out a make-up with prompt notice.

- No make-ups will be offered for your own personal matters (such as family trips or weddings), nor for other non-emergency reasons.
- No make ups “after the fact.” If you take a quiz, and end up with a low score, don’t say, “Well, I wasn’t feeling well that day. Can I retake it?” The answer is No.
- If you miss a quiz and do not contact the instructor before the deadline, you will receive a grade of 0% for the missed quiz.

Do you record lecture?

No. This course is “synchronous.” Live attendance is expected, just like with a traditional in-person class. It is your responsibility to read the material and find out about announcements if you have to miss class.

I took the quiz or submitted the assignment, but I don’t have a score in the grade book!

Give us time to finish grading and/or review the grade distribution before releasing quiz and project scores. We email and announce when you can see your grade in this case.

You said we could see our grades. I did the thing, but I got a 0.

Read the policies above. Submissions must be completed before the deadline. If we don’t have yours, we have to assume you missed or skipped it. If you believe there was a technical issue, you’ll have to email tech support and, if applicable, have them contact me:

- elarning@utdallas.edu – if you think you submitted a quiz or project but it didn’t “go through.”
- assist@utdallas.edu – for issues related to your UTD accounts (NetID, sign-in, etc.)

Can I have a homework deadline extension?

No.

I inadvertently missed a homework/quiz/project. Can I have the opportunity submit it late, maybe for partial credit?

No.

My quiz score is low. I was having a really hard day and dealing with lots of family and personal issues. Can I take the quiz again?

No.

I uploaded the wrong file! Can you accept my file by email, or let me change my submitted file?

No.

What can I do to improve my grade in class?

Come to every class and pay attention. Take note of announcements, complete everything assigned, never copy from anyone (except Dr. Whalen), email Dr. Whalen your questions and concerns well before the critical deadlines. Focus on the provided materials to keep from getting overwhelmed. Ask the instructor questions instead of losers like stackexchange or ChatGPT. Read the lessons and supplements as listed above in the topic list.

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. Give the instructor your attention, focus on the lecture, and put away smartphones and other distractions.

>Avoid leaving early and avoid arriving late.

>At the instructor's discretion, you may be asked to leave the session and/or receive a grade penalty for behavior that interferes with class.

>You are expected to attend class with your webcam on. Skipping class and not participating will bring you lots of stress and difficulties. Instead, always come to class.

Various lawyers have suggested that we mention that the content of this syllabus may change at the instructor's discretion.

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*)