

Course Syllabus: MSDS 6372
Summer 2021

Live Session	Mondays and Wednesdays 8:30-10pm Central
Instructor:	Dr. Jacob Turner
Office Hours:	Sundays at 8:30pm Central.
Course Text:	Ramsey, F. L., and Schafer, D. W. (2013), <i>The Statistical Sleuth: A Course in Methods of Data Analysis</i> (3rd ed.), Boston, MA: Brooks/Cole, with associated website www.statisticsleuth.com .

Helpful Free Online Resource: *Introduction to Statistical Learning with Applications in R* by James, Witten, Hastie, and Tibshirani. Springer.

Prerequisites: A passing grade in MSDS 6371.

The text is available in electronic version from CengageBrain.com and is much less expensive this way!

Learning Objectives: The student will

- Gain understanding and confidence statistical methods and concepts that complement their knowledge from MSDS 6371.
- Demonstrate understanding of the advantages and disadvantages of a given experimental design, particularly with respect to the type of conclusions that can be made.
- Appropriately apply the methods discussed in the course to numerical and categorical data, including multivariate responses, correlated responses, and explanatory variables.
- Communicate the findings of a statistical analysis from these new methods in a clear, concise, and scientific manner. This includes learning various structures of professional journal and statistical report writing.

Course Expectations

- As a Data Scientist in training, you need to learn to search out answers to questions for yourself before asking your instructor. It's amazing what five minutes on the Internet can turn up! (Cross Validated, Stack Exchange, etc.) Please, train yourself to try to find an answer for yourself before asking someone else. Give yourself a time limit (i.e. I will search for one hour and if I can't find anything, I will ask). It's better for you in the long run!!
- Download SAS prior to the first live session meeting. Mac users must use the version at apps.smu.edu or SAS On Demand (or download a virtual machine and run the local version). If you have questions on obtaining SAS, please contact help.smu.edu. The live session instructors cannot help you with questions regarding downloading and installation of SAS.
- Expect to see more R this semester. HWs will be structured in R although folks who want to use SAS should still be able to complete all tasks in SAS if they want to.
- Watch all asynchronous material for the course. If you do not view required materials for a given week, the instructor will redirect you to watch the videos before answering any questions about the material.

Course Coverage

The course follows the textbook fairly closely. We will start with chapter 23 in order to get grounding in various research designs. Then we will move on to chapter 12, which was partially covered in MSDS 6371. This will not be entirely review. We will introduce more detail on regression that was not covered in 6371. More detail is provided in the unit breakdown below.

End of Chapter Quizzes (0%): For some units, there are also comprehension quizzes at the end of the unit, you **DO NOT** need to do these. Our HW will replace these quiz questions.

Grading: Homework (20%), Participation (10%) Two Course Projects (20% each), Midterm (15%), Final(15%)

Homework (20%): A common request among most students taking this course is to learn how to apply the methods we learn in R. In the past, HW's have been completed in SAS and R scripts are provided as supplement. This semester we will be emphasizing conducting most of the analysis in R and SAS will be provided as supplement in other parts of the course. HW assignments will be provided in a document that will serve two purposes. The first is a tutorial of how things are done in R. We will use examples that we used in live session to help drive the concept sense you are familiar with it already. The second purpose will be to assess your high level understanding of the topic currently under discussion. This will comprise of short answers along with some rinsing and repeating of some the R scripts provided.

Note: This approach to HW will be the third or fourth time through. Most of the errors and issues have been caught the first two times through, but I' appreciate any and all feedback in regards to the HW assignments. New versions of R.4.0 or higher may have a few things we still need to sort out. I will try to proactively mention them. If you figure something out, post to the wall. We can all benefit.

Homework's should be produced in a separate word document addressing each question. Ambitious students who would like to use R markdown may do so. I highly encourage you try to do this if you can. There is a learning curve, but saves mass amounts of time later. You will have examples of R markdown when I start assigning homework. Solutions to the HW will be posted in 2DS roughly one week after submission.

Homework will be assigned every week but it will not necessarily be picked up every time. Due to midterms and projects, I realize the crunch there and HW around those deadlines tend to get a little rough. Tentatively, homeworks for Unit 2,3,4,8,9,11, and 12 will be picked up and graded officially. HWs that are not picked up are still fair game from a concept point of view for examinations. Solutions to all homework's, including those not taken up, will be provided.

Regardless of what the 2DS system says, all HWs will count towards your grade. It is imperative that you complete these assignments so we can assess your individual knowledge and work. If an occasion extension is needed. I can work with you, but you need to be proactive and communicate your situation with me early and often.

Participation (10%):This component has two essential parts: viewing of asynchronous material and participation during live sessions. Watching all the content and being present, both in body and mind, is expected. The only way this can negatively impact your grade is if you are simply not showing up

and its clear your not going through the material. Your final participation grade will be primarily based on the completion and submission of the “pre live” assignments discussed below. Again, you DO NOT have to participate in BLTs on the 2DS system or complete any quizzes.

In order to participate fully in the synchronous sessions, you must complete ALL of the material for the asynchronous sessions each week. To verify that you have done so, we will open each Live Session with a 10 to 30 minute discussion on the “pre live” assignment. I will post this assignment within 2 days of your last live session and you will submit your discussion points into the 2DS system before the next live session. Students will prepare a 3 to 7 slide (no more than 7) presentation on what they found for the given problem. This is a completion grade and the student must simply attempt the problem ... a correct answer is not necessary. Each power point should be an attempt at the analysis which may include running and interpreting SAS code. I will take volunteers and randomly assign students to present so everyone is participating. Again, a correct analysis is not required or even really expected; however, a solid and well thought out attempt is. This will ensure a great discussion and maximize learning!

You may miss 2 Live Sessions without a grade penalty. However, you are still required to watch the video and submit the pre live assignment. Your learning depends on this. If you know of conflicts in advance or something crazy has happened, talk to me so we can sort out an appropriate course of action.

2 Projects (20% each): There are two course projects, due in Unit 6 (Due **June 6th** midnight) and Unit 15 (Due **August 8th** midnight). There is no asynchronous content for these weeks. You are expected to take the week to *finish* (as opposed to start and finish) your project. Project descriptions will be available the first week of class. I will discuss strategies and what makes a good project a good project. The main goals of the projects are to produce reasonable models using the strategies discussed, demonstrate the ability to think statistically, and effectively communicate output, graphics, and results. Projects will be submitted on 2DS along with a peer review form. I will also be checking in, starting Week 3, with the groups to make sure there are no last minute “hiccups” or issues among group members.

Midterm and Final (15% each): There will be a midterm in Unit 7 and a final in Unit 14 this semester. The midterm exams will each be 3 hour exams. The midterm will be from 11am – 2pm CST on Saturday, **June 12th**. The midterm will be a combination of multiple choice and free response. Nearly all questions will be related to either live session questions, homework questions, or high level discussions in class. If you are actively working on homework, engaged during live session, and participating in your projects, you should be well prepared. The midterm will be submitted through the 2DS system before time has expired.

The final exam will be an individual semester wide project this semester. I refer to this as the “Cheat sheet project”. The project consists of students filling out a cheat sheet that answer 4-5 key questions about each topic covered throughout the semester. The goals of this project are to help students grow good study habits, to learn how to ask the right questions when learning new predictive tools, and to create a document by the end of the term that the student can take with them in the future to have as a short reference....a “cheat sheet”. The final document will be submitted through the 2DS system tentatively on **August 1st**. Note: we still have another week of class after your final assignment is submitted.

Honor Code

The Honor Code

When you signed your letter of intent to enroll in the program, you initialed the following statement:

“I have read and agree to abide by the SMU Honor Code available online at:
<http://www.smu.edu/StudentAffairs/StudentLife/StudentHandbook/HonorCode>”

Please know that the Honor Code is taken very seriously.

Honor Code Violations Include: Academic Sabotage, Cheating, Fabrication, Facilitating academic dishonesty, Plagiarism among others.

Collaborative Work

Data science is a collaborative subject

Most professors encourage collaborative work except when explicitly prohibited.

Collaboration means helping one other, not copying answers from one another.

Students who turn in exactly the same answers to the same homework will share the grade assigned (i.e. if two students have the same answers, and the grade on the assignment is a 90, then each student will receive a 45).

Some instructors may impose stricter penalties

Okay	Not Okay
Discussing problems (or even solutions) with other students Telling another student how you did something in SAS (which PROC and how to use it, etc.) or sending “snippets” of code (i.e. an example of how to use a PROC) to help them out.	Sending your answers to a student or receiving them from another student Sharing anything, including ideas, when expressly forbidden, such as on exams Passing off anyone else’s work or explanations as your own Using prior semester solutions as your own

Plagiarism

Original sentence from a the easycalculation.com website: *A regression is a statistical analysis assessing the association between two variables. It is used to find the relationship between two variables.*

Here is an example of plagiarism:

A regression is a statistical analysis assessing the association between two variables. It is used to find the relationship between two variables.

The following is NOT plagiarism:

“A regression is a statistical analysis assessing the association between two variables. It is used to find the relationship between two variables.” (<https://www.easycalculation.com/statistics/learn-regression.php>).

The difference is in the punctuation and the attribution.

Note that one can self-plagiarize. If you are using something that you wrote (e.g. a blog or a previously published article), please reference yourself.

Consequences

Plagiarism, sabotage, fabrication, and cheating carry high penalties.

Instructors may choose to fail the student on the assignment, give a 0 for the assignment, fail the student for the course, and/or bring the student before the Honor Council at which point expulsion is an option.

The worst penalty from breaking the honor code is the fact that either the person or their peers (or both) miss out on learning the material; they miss out on growth.

The obvious takeaway is:

Have a blast in this course learning the material. We learn through making mistakes and then talking through and correcting those mistakes.

Weekly Breakdown of Course

Unit/Week	Chapter	Topic
1	23	Elements of Research Design
2	12	Strategies for Variable Selection in Multiple Regression
3	13	The Analysis of Variance for Two-Way Factors
4	15	Adjustment for Serial Correlation
5	16	Analysis of Repeated Measures
6		Project 1 Work Week: Due: Sunday Jun 6th midnight Live Session is additional office hours.
7		Midterm Review during Live Session. Midterm (Saturday Jun 12th 11am-2pm CST)
8	17	MANOVA and Linear Discriminant Analysis (LDA Focus)
9	17	Principal Components
10	18	Comparison of Proportions or Odds
11	19	Chi Square Tests and Other Tools for Tables of Counts
12	20	Logistic Regression for Binary Responses, ROC Curves
13	ISLR Book/Notes	Unsupervised Analysis / Cluster Analysis / Heatmap Visuals
14	ISLR Book	Final "Cheat sheet" due. Due: Sunday August 1st midnight Live Session is additional office hours. Turner will cover RF and Tree's for students interested.
15		Live Session is additional office hours. Project 2 Work Week: Due: Sunday August 8th midnight

Summary of Assignments Due Each Week that should be submitted to 2DS.

Prelives are due before live session starts that week. HW assignment are due the day of your live session by midnight.

Unit/Week	Pre Live Assignment	Homework Assignment
1	PPT Slides from wall (No deliv)	None
2	Unit 2 Pre Live (PPT)	Read the syllabus and see this table!!!
3	Unit 3 Pre Live (PPT)	Unit 2 HW Assignment (Word/Rmark)
4	Unit 4 Pre Live (PPT)	Unit 3 HW Assignment (Word/Rmark)
5	Unit 5 Pre Live (PPT)	Unit 4 HW Assignment (Word/Rmark)
6	None	No Unit 5 HW due. HW is available for practice though.
7	None	None
8	Unit 8 Pre Live (PPT)	None
9	Unit 9 Pre Live (PPT)	Unit 8 HW Assignment (Word/Rmark)
10	Unit 10 Pre Live (PPT)	Unit 9 HW
11	Unit 11 Pre Live (PPT)	No HW due.
12	Unit 12 Pre Live (PPT)	Unit 10/11 HW due
13	No Pre Live	Unit 12 HW due
14	No Pre live	No HW
15	No Pre Live	No HW

Best Practices for Success in MSDS 6372 (and other courses also)

Attendance. Take responsibility for your commitment. Attendance means not only being there for synchronous sessions but also participating in asynchronous work.

Citizenship. You need to be actively engaged to succeed in this class. Talking on cell phones, texting, “facebooking,” tweeting, or leisure web browsing are prohibited in class. I consider these to be a disruption (not to mention rude).

Integrity. A lot of the graded work occurs outside of class, so I expect honesty and integrity in what you submit for evaluation. Evidence of academic dishonesty will minimally result in zeros for all involved parties, and perhaps University-level disciplinary action. Don’t risk your academic career.

Humility. Don’t get lost! Ask questions in class. If something isn’t clear to you, it probably isn’t clear to others either. Questions may arise because I haven’t made a connection clear or have inadvertently left out an important point. Your question gives me a chance to explain more clearly. Don’t be proud or shy.

Organization. Don’t procrastinate! This is a technology-driven course. Count on your computer failing or your wireless connection breaking the night before a due date. Start early and give yourself a chance to succeed.

Deadlines. You will generally have a week to complete an assignment. Due dates and times will be clearly indicated. Late submissions will be penalized, but it is much better to turn in work late than not at all (or to turn in incomplete/sloppy work). ***Work turned in after solutions have been posted to the course website will receive no credit.***

Getting help. If questions arise while doing assignments/exams, do your best to resolve these questions before the assignment is due, first by taking time to seek answers yourself, next by asking questions on the wall, and finally via e-mail to your instructor or other students. **I encourage you and expect you to seek help.** For questions during exams, please e-mail the live session instructor directly.

Collaboration. I encourage the formation of study groups and collaboration with your fellow students in tackling the assignments. Working together in groups on homework is permitted, even encouraged. **However, every student should write up and complete his or her homework independently. Students who chose to turn in exactly the same work will share the grade assigned.** Talking about problems with other people does help in learning, but just copying the solutions from one another doesn't help! See Plagiarism discussion above.

Looks do matter! All assignments must be NEATLY executed and organized. You risk a zero on any assignment submitted in a sloppy manner. See submission guidelines for more detail.

Submission guidelines for assignments

- Your name must be at the top of the first page and on each successive page.
- Submit solutions in problem order.
- Use an easy-to-read variable-width font (I like Ariel, Helvetica, and Geneva fonts—this document is in Helvetica 11 point) with a minimum of 11 point font.
- Relevant SAS code and output from the SAS console must be included in-line at the appropriate point using Courier New (or other fixed width) font, in 10 point size. **Inclusion of irrelevant code or output will be penalized.**
- Any graphics from SAS must be electronically cut and pasted in-line at the appropriate point of the write-up. You can use Word to resize the graphics appropriately.
- Any mathematical notation must be provided with appropriate use of subscripts, superscripts, and symbols. Use MS Equation or another equation editor if you submit your work in Word.

University Policies

Incompletes will be given only in the case of extraordinary circumstances that prevent you from finishing the semester. You must have completed at least 50% of the course with a passing grade to be eligible for an incomplete.

Religious Observance: Religiously observant students wishing to be absent on holidays that require missing class should notify the live session instructor via e-mail, and should discuss with the instructor, in advance, acceptable ways of making up any work missed because of the absence. (See University Policy No. 1.9.)

Excused Absences for University Extracurricular Activities: Students participating in an officially sanctioned, scheduled University extracurricular activity will be given the opportunity to make up class assignments or other graded assignments missed as a result of their participation. It is the responsibility of the student to make arrangements with the instructor prior to any missed scheduled examination or other missed assignment for making up the work (University Undergraduate Catalogue).

CALENDAR (All times Central)

Summer TERM 2021 (April 27th, 2021 – August 8th, 2021)

Project 1: Due midnight Jun 6th, 2021

Midterm: 11am-2pm Jun 12th, 2021

Final Exam Project: Due by Aug 1st at midnight

Project 2: Tentatively August 8th at midnight

More academic calendar details:

<http://www.smu.edu/EnrollmentServices/Registrar/AcademicCalendarsCourseCatalogs/AcademicCalendars>