Syllabus
AMS 315
Data Analysis
Spring Semester, 2023

Introduction and Overview

Welcome to AMS 315. This is a second course that has AMS 310 or its equivalent as a prerequisite. The course requires that you attend using a relatively high level of computer and computer network. Each of these points will be discussed in more detail in this syllabus.

This course will use an in-person lecture setting, supplemented by distance learning resources. The Blackboard learning management system will be used. See the Technical Requirements section for the minimal specification requirements of the computer and internet systems that you need. You will access course materials and resources from the course Blackboard.

Examinations and Quizzes

Examinations and quizzes will be given through the Respondus Lock-down Browser. Examination and quiz papers will only be accepted through the Respondus browser.

Evaluation

I will use examinations (quizzes, three midterm examinations, and a final), projects (two computer projects), and class participation (as recorded on the Class Blackboard Learning Management System) to evaluate you. I will calculate your total points, which is the sum of the points you earned on your examinations and quizzes, the points that you earned on your projects, and your Blackboard participation points. The maximum total points that you can earn will be about 2,400. Examinations and quizzes will account for about 1700 of the total points possible; projects about 385 points; and Blackboard participation about 300 points.

Resources available to you

The Class Blackboard contains links to the recorded lectures, the pdf files used in the lectures, material summarizing the text (chapter guides and chapter lecture notes for each chapter), supplemental readings, and pre-recorded Zoom videos discussing the major problem types covered in this course. It will also contain the data sets that you must use in your projects, instructions for accessing online tools and components, and information about examinations and projects. I will hold 3 hours of Zoom contact weekly: on Mondays from 2:30 to 4:00 pm and on Wednesdays from 3:30 to 5:00 pm Stony Brook time. The TAs will also hold 2 hours of contact (either in person or on Zoom) at the times

specified on the class Blackboard. The Student Accessibility Services Center may also be able to provide you with additional resources.

Free academic support services including one-on-one and small group course-based tutoring, one-on-one skill-based tutoring, peer assisted learning (Supplemental Instruction), and public speaking courses are available for you. Learn more about these services by visiting www.stonybrook.edu/tutoring.

Complete Details

Examinations and Quizzes

Syllabus Quiz

There is a quiz worth 25 points on the class syllabus available during the first week of class. To get credit for this quiz, you must download the Respondus browser and use it to take this quiz. You will not get any points if you do not use the Respondus browser. You will be penalized additional points if you do not use the Respondus browser for the first midterm. There will be additional penalties if you do not use the Respondus browser in subsequent examinations.

Examinations:

There will be three midterm examinations and a final. You must use the Respondus browser to take the examinations. The total maximum grades on these examinations and quizzes will be about 1,700 points or about 71% of the total points possible. The tentative dates of the midterm examinations are below. All times in this syllabus are Stony Brook times. Midterm examinations will be held on-line during the scheduled class time, which is 4:45 to 6:05 pm, Stony Brook time. If you miss one of the midterms, there will be an 80-minute make-up examination on Friday, May 5, at 1 pm Stony Brook time. The make-up examination will be comprehensive. If you miss two or more midterms, my academic advice is to withdraw from the course. The date and time of the final are set by the University. The final will be on Monday, May 15, from 5:30 pm to 8:00 pm, Stony Brook time.

I will allow you an additional 15 minutes to scan your midterm examination paper and submit it through Blackboard. A submission before 6:20 (6:05 plus 15 minutes) will receive a 25-point bonus. Papers submitted between 6:20 and 6:30 will not receive the bonus but will not be penalized. Papers submitted after 6:31 will be penalized 2 points for each minute after 6:31 up to a maximum of 60 points.

For the final examination, I will allow you 20 minutes to scan and submit your final examination paper. Final examinations submitted before 8:20 pm will receive a 50-point bonus. Papers submitted between 8:21 pm and 8:30 pm will not receive the bonus but will not be penalized. Papers submitted after 8:31 pm will be penalized 2 points for each minute after 8:31 up to a maximum of 60 points. When making your reservation for a

flight at the end of the semester, please do not make a reservation for a flight leaving New York before 12:20 am on May 16. Remember that your final lasts to 8:20 pm, that it can take 2 hours to get to JFK or other NYC airport, and that it can take 2 hours to clear security.

There will be no review classes prior to the examinations, as the lectures are videorecorded. The make-up examination is given for students who were ill on the day of the examination or who had to miss an examination for a job interview or other professional reason. Please do not miss the opportunity to check your progress by taking each midterm examination as scheduled. Do not give in to the temptation to skip an exam because you do not feel prepared. Only about 33% of eligible students take the make-up. However poorly prepared you feel you are for a midterm, taking the exam will likely be a net benefit to you.

Examinations will be open book and open notes. You may use any calculator that you wish. The pdf file of statistical tables is available on the class Blackboard in the Assignments section. I recommend that you print it out for use in the examinations. This pdf file will also be attached to your examination file. However, in past semesters some students had trouble accessing this file during an examination.

Students with SASC accommodations will start their mid-term examinations at 4:45 pm Eastern time and must submit their papers to Blackboard within the time allocated by SASC for them. They will start their final examination at 5:30 pm Stony Brook time and will have the time SASC allocated for them.

The dates and tentative content of the examinations and other important dates are:

Feb 23:	***Examination One: Chapters 3, 4, 5, 6, and 7***
Mar 9:	*Project 1 Data Posted*
Mar 30:	***Examination Two: Chapters 11 and 12, in addition to Chapters 3, 4, 5, 6, and 7.***
Apr 11:	*Project 1 Due at 11:59:00 pm Eastern U.S. Time*
Apr 18:	*Project 2 Data Posted*
Apr 27:	***Examination Three: Chapters 8, and 9 in addition to Chapters 3, 4, 5, 6, 7, and 12.***
May 2	*Project 2 Due at 11:59:00 pm Eastern U.S. Time*
May 5:	***Make-up examination for students who missed at least one mid-term, 1:00 pm.***
May 15:	***5:30 pm-8:00 pm Eastern US Time. Final examination: Chapters 3-12.***

Projects

There will be two projects with a combined maximum value of 385 points, which will be about 16% of the possible total points in the course. The individual data sets for your projects will be posted on the class Blackboard in a Zip file. The file containing your data

set will have the last six digits of your Stony Brook identification number in the name of the file. ***You must analyze your assigned data set. If you do not, your score will be zero.*** If your report is plagiarized from another student, both students will be charged with academic dishonesty—the enabler is more guilty than the plagiarizer.

The data for each project will be generated synthetically according to a model that differs for each student. The models are selected because they occur commonly, as noted in the scientific papers posted on the class Blackboard. You may use the statistical program of your choice to analyze your data.

The first project will be to analyze two different sets of synthetic data using single predictor linear regression. One part of this project will require you to merge two data files; this part focuses on data processing issues and dealing with missing data and will be worth 40 points. The second part of the project will be to analyze data that may not fully meet the assumptions of Chapter 11 and will be worth 60 points. The first project is due on April 11 at 11:59 pm. A paper submitted before that time will receive a 20 point bonus. I will liberally grant time extensions for project 1. On on-time submission may earn as much as 120 points; late submissions may earn as much as 100 points.

The second project will be to analyze synthetic data with multiple predictors and to estimate the model that was used to generate the data. It will be worth a maximum of 250 points. The model generating the data may include transformations of the outcome variable, non-linear predictors, and interactions of predictors. The second project is due at 11:59 pm of May 2. An on-time submission will earn a 25-point bonus so that the paper may earn as much as 275 points. The project 2 due date can be extended a bit, but not much. Please note that the project 2 due date is near the end of the semester.

Each student will submit a report in scientific paper format for each project. The grade of a student who does not submit computer projects will be severely impacted. A minimally satisfactory computer score is a sum of 175 points or more (out of 350) in the grades of project 1 and project 2.

The dates of the projects are below.

Mar 9:   *Project 1 Data Posted*
Apr 11:   *Project 1 Due at 11:59:00 pm Eastern U.S. Time*
Apr 18:   *Project 2 Data Posted*
May 2:   *Project 2 Due at 11:59:00 pm Eastern U.S. Time*

Course Blackboard

Course materials posted to Blackboard will include:
- The syllabus.
- Videos of class lectures.
- Pre-recorded Zoom videos of specific problem types.
- Supplemental readings.

- Instructions for accessing online tools and components.
- Information on assignments and assessments.

Supplemental Readings

Scientific papers and other material will be posted on the class Blackboard. These papers provide case studies and more detailed discussion and illustration of the application of statistical techniques than your text or I can give. They are an important supplement to your studies. You should study them to enhance your understanding of the applicability of the material in the course and how the material of the course is part of quantitative research.

Communication via Blackboard

For personal/private issues, please send me an email (Stephen.finch@stonybrook.edu). If you use Blackboard's Email Tool, it will automatically include your full name, course name, and section when you send emails. The TAs and I will work to respond to your emails as soon as possible, but please allow 2 business days for a reply or an announcement in Blackboard on the issue. Your Stony Brook University email must be used for all University related communications. All correspondence will be sent to your SBU email account. To log in to Stony Brook Google Mail, go to http://www.stonybrook.edu/mycloud and sign in with your NetID and password.

Course Issues

Prerequisites

The prerequisite for this course is AMS 310 or equivalent. For example, a student who has completed AMS 102, AMS 110 or PSY 201 and received a grade of B or better should be reasonably well prepared for this course. Specifically, you should have some knowledge of probability theory, a basic understanding of the central limit theorem, knowledge of the single sample t-test, and an understanding of the concepts of testing hypotheses and confidence intervals. I will also use basic matrix calculations (such as matrix multiplication and inversion) in the second half of the course and some calculus (differentiation, optimization of a function, and Taylor series approximations). I do expect that you have some familiarity with computing and that you wish to expand your expertise to being able to work with statistical computing packages.

Tentative Schedule Spring 2023 AMS 315
Lectures are from 4:45 to 6:05 pm Eastern Time

Jan 24:	Introduction, Chapter Three, Data Description
Jan 26:	Chapter Four, Probability and Probability Distributions (simulation issues)
Jan 31:	Chapter Five, Inferences about Population Central Values
Feb 2:	Chapter Five, Inferences about Population Central Values
Feb 7:	Chapter Six, Inferences Comparing Two Population Central Values
Feb 9:	Chapter Six, Inferences Comparing Two Population Central Values
Feb 14:	Chapter Seven, Inferences about Population Variances
Feb 16:	Chapter Seven, Inferences about Population Variances
Feb 21:	Chapter Eleven, Linear Regression and Correlation
Feb 23:	***Examination One: Chapters 3, 4, 5, 6, and 7***
Feb 28:	Chapter Eleven, Linear Regression and Correlation
Mar 2:	Chapter Eleven, Linear Regression and Correlation
Mar 7:	Chapter Eleven, Linear Regression and Correlation
Mar 9:	Chapter Twelve, Multiple Regression and the General Linear Model
	Project 1 Data Posted
Mar 14:	Spring recess, no class
Mar 16:	Spring recess, no class
Mar 21:	Chapter Twelve, Multiple Regression and the General Linear Model
Mar 23:	Chapter Twelve, Multiple Regression and the General Linear Model
Mar 28:	Chapter Twelve, Multiple Regression and the General Linear Model
Mar 30:	***Examination Two: Chapters 11 and 12, in addition to Chapters 3, 4, 5, 6, and 7.***
Apr 4:	Chapter Eight, Inferences about More than Two Population Central Values
Apr 6:	Chapter Eight, Inferences about More than Two Population Central Values
Apr 11:	Chapter Eight, Inferences about More than Two Population Central Values
	Project 1 Due at 11:59:00 pm Eastern U.S. Time
Apr 13:	Chapter Nine, Multiple Comparisons
Apr 18:	Chapter Nine, Multiple Comparisons
	Project 2 Data Posted
Apr 20:	Chapter Nine, Multiple Comparisons
Apr 25:	Chapter Nine, Multiple Comparisons
Apr 27:	***Examination Three: Chapters 8, and 9 in addition to Chapters 3, 4, 5, 6, 7, and 12.***
May 2:	Chapter Ten, Categorical Data
	Project 2 Due at 11:59:00 pm Eastern U.S. Time
May 4:	Chapter Ten, Categorical Data
May 5:	***Make-up examination for students who missed one mid-term, 1:00 pm.***
May 15:	***5:30 pm-8:00 pm Eastern US Time. Final examination: Chapters 3-12.***

Recommended text

Ott, R.L., Longnecker, M. (2016). *Statistical Methods and Data Analysis, latest edition*. Pacific Grove, CA: Duxbury. The tables that you will need in your examinations are available on the class Blackboard. You may use earlier editions of this text or another text if you prefer. If you use another text, please make sure that it has the statistical tables that you will need.

Assignments

There are no homework assignments. Past examination problems are in the study guides for each chapter that are posted on the class Blackboard. Past examinations and their solutions are also posted. You may turn in your solutions for comments from the TAs or me. The TAs will read your solutions and offer suggestions and instruction. They are instructed *not* to solve the problem for you. Because of the rampant sharing of homework solutions, homework papers will not be graded or included in the grading process. Example problems will be discussed in lectures. Zoom videos solving specific examination problems are posted on the class blackboard. In lieu of homework, your usage of the class Blackboard will be tracked and is a component in your evaluation.

Office Hours

My Zoom contact hours are on Mondays from 2:30 to 4:00 pm and on Wednesdays from 3:30 to 5:00 pm Stony Brook time. There will be extra office hours the week before each examination. These will be announced on the class Blackboard. My e-mail is Stephen.Finch@stonybrook.edu. If you have confidential issues that you wish to discuss with me, please be aware that I will not have a waiting room for my contact hours. Please arrange an individual Zoom meeting with me to discuss private issues.

Please check the class blackboard to determine whether a TA contact hour is via Zoom or in person.

Incompletes

Professor Arkin, who is the undergraduate program director in the Applied Mathematics Department, is the only person who can authorize an incomplete grade for an upper division AMS course. If Professor Arkin has not authorized an incomplete, absence from the final may result in failure of this class.

Since this class can be offered as part of the training requirements of the Society of Actuaries, academic integrity issues are fundamental. Academic integrity standards will be enforced to the best of our ability.

Grading

The target grade distribution is roughly 25% A, 25% B, and 25% C+, and the remainder C or lower. That is, the target course GPA is 2.75. An examination score in the upper quartile is roughly an A grade; a score in the second quartile is roughly a B grade; a score in the third quartile is roughly a C+ grade; and a score in the lowest quartile is a C or lower grade.

Your computer project component is the sum of the scores received for the two projects, $CP_1 + CP_2$.

Your examination component E is the sum of your examination and quiz scores:
$$E = E_1 + E_2 + E_3 + E_F + Q.$$
Each in-class examination will have 6 or 7 questions, each worth about 50 points. Each midterm examination will be worth approximately 300 points. The final examination will have about 14 problems and will be worth approximately 650 points.

Your class participation component P is based upon your usage of the class Blackboard. It is the sum of four components:
$$P = P_1 + P_2 + P_3 + P_{Cumulative},$$
where P_i is your class participation score for the time covered in the ith examination and $P_{Cumulative}$ is your class participation for the whole semester. The component participation scores satisfy $0 \le P_i \le 60$ and $0 \le P_{Cumulative} \le 120$. Participation scores are based on the amount of active time spent in the AMS 315 course Blackboard page over the period in question. Any time spent within the subpages of the AMS 315 Blackboard will count towards this time (*e.g.,* course documents, videos, assignments). In terms of maximizing participation score, it is recommended that you view any course documents from within the Blackboard page itself, rather than viewing downloads of these files from a separate application on your computer. The score itself is calculated based on your rank of time spent in the AMS 315 Blackboard page among all students enrolled in AMS 315. For example, the student with the most time spent in the AMS 315 Blackboard page between the specified dates will earn a participation score of 60, whereas the student with the least amount of time spent in the AMS 315 Blackboard page will earn a participation score near 0.

Your total point score is the sum of your examination scores, project scores, and class participation score:
$$TP = E_1 + E_2 + E_3 + E_F + Q + CP_1 + CP_2 + P.$$

Your final grade will be determined using your final examination grade and your total point score. To calculate your final course grade, total point boundaries for an A, B, C+, C, D, and F grades will be set. Similarly, final examination boundaries for A, B, C+, C, D, and F grades will be set. Each student will have a total point grade and a final examination grade. A student with a satisfactory computer project score (150 or more computer project points) will have a course grade that is the higher of the final examination grade and the total point grade. The grade of a student without satisfactory

computer projects scores will be the total point grade. That is, if you have at least minimally satisfactory work on the computer reports, a strong performance on the final will be a major factor in the final grading decision. A student with at least minimally satisfactory computer performance who gets an *A* on the final gets an *A* on the course; a student with satisfactory computer performance who gets a *B* in the final gets at least a *B* in the course; and so on.

Computer Software Issues

Technical Requirements

This course uses Blackboard (https://blackboard.stonybrook.edu) to facilitate communications between faculty and students, submission of assignments, and posting grades. If you are unsure of your NetID, visit https://it.stonybrook.edu/help/kb/finding-your-netid-and-password for more information.

You are responsible for having a reliable computer and Internet connection throughout the term.

Caution! You may be at a disadvantage if you attempt to complete all coursework on a smart phone or tablet. It may not be possible to submit the required files. The following list details a minimum recommended computer set-up and the software packages you will need to access and use:

• *Hardware*:
– PC (Windows 7, 8, or 10) or Macintosh (OS X/macOS 10.10 or higher).
– 4 GB RAM.
– A high-speed internet connection. Note that public WiFi (e.g., Starbucks) and internet service provider hotspots (e.g., optimumwifi or xfinitywifi) are not recommended.
– Printer and scanner. A cell phone or tablet camera can work in some cases for the latter with apps such as CamScanner (there are myriad others).
– Speakers (either internal or external) or headphones. Headphones are strongly recommended to reduce the risk of feedback during communications. – Microphone (either internal or external).
– WebCam or other camera (interfacing with your computer) for producing video.

• *Software* (additional tools may be needed). Remember to use your Stony Brook email or NetID when configuring specialized software:
– The ability to download and install software applications and plug-ins. Note that you may need administrator access to install some applications and plug-ins.
– An up-to-date Internet browser, such as Chrome, Firefox, Explorer/Edge (Windows), or Safari (macOS). A complete list of supported browsers and operating systems can be found on the My Institution page when you log in to Blackboard.
– Word processing software (e.g., Microsoft Word, Pages).
– PDF viewer, such as Adobe Reader.

– Zoom. Stony Brook has a site license for Zoom; you can find information on downloading, installing, and using Zoom at https://it.stonybrook.edu/services/zoom/students.

Respondus LockDown Browser
Stony Brook has a site license for these packages; see https://download.respondus.com/lockdown/download.php?id=772113517.

If you need technical assistance at any time during the course or to report a problem with Blackboard you can:
• Visit the Stony Brook University Student Help Desk Page, http://www.stonybrook.edu/helpme
• Phone: – (631) 632-2358 (technical support and Blackboard issues) – (631) 632-9800 (client support, wifi, software and hardware)
• Create a ticket at http://service.stonybrook.edu. Students who need assistance with their personal devices can contact DoIT's service desk at (631) 632- 9800 or submit an online request. For more information, visit: https://it.stonybrook.edu/students

Methods of Student Evaluation

There will be four examinations with a total with between 30 and 35 questions. Approximately seven will be theoretically oriented, such as proving that the expected value of the sample variance using a random sample is equal to the variance of the sampled random variable. The final examination will be cumulative and focused on the material taught in the course beyond the content of AMS 310, the prerequisite for this class. There will be two projects requiring statistical computation. The goal for each objective is "executive mastery." For example, executive mastery of a test of a hypothesis question is making the correct decision. Executive mastery of a sample size question is finding the correct size. Executive mastery of a multiple regression project is identifying the variables truly associated with the dependent variable.

Course Objectives

1. Review topics from the prerequisite course (AMS 310 or any one semester college level introduction to statistics.
   a. Basic probability distributions; i.e., binomial, Poisson, normal, and exponential distributions (Chapter Four of text).
   b. Probability calculations; e.g., probability of an event for a random variable following one of the basic distributions, use of Bayes' Theorem, finding expected value and variance of a random variable (Chapter Four of text).
   c. Vocabulary of statistical procedures; e.g., confidence intervals, tests of hypotheses, Type I and Type II error rates (Chapter Five of text).
   d. One sample Student t procedures (Chapter Five of text).
2. Extend knowledge of probability theory.
   a. Central chi-square and central F-distributions (Chapter Seven of text).

b. Bonferroni's inequality applied to multiple test of hypotheses (Chapter Nine of text).
c. Logic of multiple comparison procedures (Chapter Nine of text).
d. Decomposing chi-square sums of squares (Chapter Nine of text).
e. Expected value and variance of multiple linear combinations of random variables (Chapter Eleven of text).

3. Learn intermediate level statistical procedures.
   a. Two sample tests and confidence intervals (Chapter Six of text).
   b. Tests and confidence intervals for the variance of a normally distributed random variable (Chapter Seven of text).
   c. Tests comparing the variances of two independent normal random variables (Chapter Seven of text).
   d. Tests and confidence intervals for the one way analysis of variance (Chapters Eight and Nine of text).
   e. Statistical procedures for multiple comparisons (Chapter Nine of text).
   f. Categorical variable tests using the chi-squared distribution (Chapter Ten of text).
   g. One predictor linear regression (Chapter Eleven of text).
   h. Multiple predictor linear regression (Chapters Twelve and Thirteen of text).

4. Review scientific studies that use the techniques of the course and introduce the basic ethical principles governing scientific research.
   a. Read papers posted on class Blackboard.
   b. Reference to papers as techniques studied in lecture.

5. Learn the statistical computing package of the student's choice and apply it to obtain the statistical model that generated a set of synthetic data.
   a. One predictor linear regression group project using synthetic data that requires students to merge separate files.
   b. Multiple predictor linear regression group project using synthetic data to recreate statistical model that generated the data. Model includes non-linear predictors and interactions of two predictors.

Method of Evaluating Each Objective
1. Review of prerequisite material: three or four examination questions.
2. Extend knowledge of probability theory: approximately five examination questions.
3. Learn intermediate statistical procedures: approximately 15 questions.
4. Review of scientific studies: in-class discussion.
5. Learn and use statistical computing package: two project reports.

End of Syllabus for AMS 315, Spring Semester, 2023