NORTHWESTERN UNIVERSITY | SCHOOL OF PROFESSIONAL STUDIES

MSDSP 401-DL Applied Statistics with R

Contact Information

Course Lead Faculty: William T. Mickelson, Ph.D.

DSP Instructor: Srabashi Basu, Ph. D.

Email: william.mickelson@northwestern.edu
srabashi.basu@northwestern.edu

Office Hours: By appointment only

Response Time: All efforts will be made to respond to emails within 24-48 hours only if the course related emails are sent within Canvas.

Course Description

This course teaches fundamentals of statistical analysis, which is the basis of all subsequent courses. This includes evaluating statistical information, performing data analyses, and interpreting and communicating analytical results. Students will learn to use the R language for statistical analysis, data visualization, and report generation. Topics covered include descriptive statistics, central tendency, exploratory data analysis, probability theory, discrete and continuous distributions, statistical inference, correlation, multiple linear regression, contingency tables, and chi-square tests.

Course Objectives

At the conclusion of the course, students will be able to:

- Perform statistical analyses.
- Interpret and evaluate statistical information.
- Prepare technical reports.
- Use the language R for data analysis.

Required Resources

Required Readings

- Black, K. (2013). *Business Statistics.* 8th edition, NJ: John Wiley & Sons, Inc. [ISBN-13: 978-1-118-80048-5] - Available on O'Reilly online for FREE
- Kabacoff, R. I. (2015) *R in Action, 2nd* ed. Shelter Island, NY: Manning Co. [ISBN-13: 978-1617291388] – Available on O'Reilly online for FREE https://www.manning.com/books/r-in-action-second-edition.
- Course Reserves from Wilcox and others – Details are in Canvas

MSDSP 401 will use the high-level language R (http://cran.r-project.org/). Instructions are given on the course site how to download. RStudio is not required but is highly recommended for new users of R. Everything students need to do in this course can be accomplished using the standard R console with a plain text editor.

RStudio http://www.rstudio.com/provides an integrated environment, and installer packages are easily available online. The installation process is straightforward. Tutorials dealing with R are available on the course site. Students are encouraged to review the *R Tutorial Materials* module on the course site, and to start studying R early in the quarter. There are tests during the quarter dealing with R programming. Check the syllabus and course site for details on date.

Assignment Overview and Grading Breakdown

The students' final grade will be based on the following:

Discussion Board	100 possible points	20%
4 self-administered Tests	100 possible points	20%
2 R Programming Assignments	125 (50+75) possible points	25%
2 Data Analysis Assignments	125 (50+75) possible points	25%
Final Exam (self-proctored)	20 possible points	
Take home Final Exam	30 possible points	10%

There will be four programming assignments requiring free responses (i.e.) not multiple choice questions. No need to include the codes in the responses.

Assignment Preparations:

- The self-administered tests are due at the end of the weeks of 2, 4, 6, and 8.
- The 4 assignments (2 on R, 2 on Data Analysis) are due at the end of the weeks of 3, 5, 7, and 9.
- All tests and assignments have a time limit. Tests 1 – 4 must be completed within 48 hours, once you have first accessed the test, inside the window period when the tests are available. The Data Analyses (1 & 2) and R Assignments (1 & 2) must be completed within 72 hours inside the window period given. Once you access the quizzes and assignments for the first time, your clock starts ticking. Pay careful attention to the instructions

Sync Sessions

There are seven scheduled sync sessions that will be given by the Course Instructors. Dates and times are listed in the Course Schedule and links are announced well ahead in time. Please note that attendance at any scheduled synchronous or "live" meetings is considered optional, though your attendance is highly encouraged. All sync sessions will be recorded and made available to students for your use. In addition, there will be 4 help sessions on R and data analysis that are tailored to the course programming assignments. These sessions will be led by the course TAs and held on Saturdays (10/8, 10/15, 11/5, 11/19). Again, these sessions are considered optional and will be recorded and made available for student use.

Final Exam Proctoring

Panopto is used to do the self-proctoring for the final exam. Students are responsible for testing their system ahead of time. Instructions are provided in Canvas. The exam is taken within Canvas using the course site.

Final exam consists of conceptual questions only and hence no need for any computations. Final exam is open book, open notes. **Access to the internet (outside of the approved resources) during the exam is not permitted. Separate portable devices such as iPads and Kindles are not permitted.** Since only one computer screen may be used for the exam, any student who uses a device such as an iPad or Kindle or multiple screens must plan ahead and be prepared to migrate the necessary files to the computer used for the final exam.

Students with disabilities working through AccessibleNU must discuss reasonable accommodations, including use of non-approved technology, with the instructor and the proctors well before their exam. Please see the Canvas course site for more information.

There is a data analysis part to the Final exam, which must be completed within 72 hours from the time the exam is first accessed within the window period.

Grading Scale

A	**93-100%**	**465-500 points**
A-	**90-92%**	**450-464 points**
B+	**87-89%**	**435-449 points**
B	**83-86%**	**415-434 points**
B-	**80-82%**	**400-414 points**
C+	**77-79%**	**385-399 points**
C	**73-76%**	**365-384 points**
C-	**70-72%**	**350-364 points**
F	**0-69%**	**000-349 points**

Late Work Policy

When stating due dates and times for work, the abbreviation "CST" is used. "CST" means Chicago, IL clock time. This defines "course time". Canvas will adjust what students see as deadlines according to the time zone specified by the student in personal settings for Canvas. (Thus an 12 pm CST deadline is a 10 pm deadline on the West Coast, and so forth depending on time zone.) Deadlines for all work are stated in this syllabus and posted on the Course Site. This includes exams, reports and participation in the discussions. Assignments are to be submitted prior to the deadlines. **Without prior arrangement, any late assignment will receive a 1% point deduction for**

each hour late, totaling to a maximum 50% deduction. For example, ten hours late will result in a 10% point deduction from the total possible assignment points. Prior communication with the instructor is essential, if you foresee unavoidable delay.

Do not fall behind. We cover a great deal of material in this course, and falling behind is the primary reason why students have difficulty, particularly toward the end of the course. To that end, the syllabus and course site give due dates for the entire course.

Online Communication and Interaction Expectations

Discussion Forums

The purpose of the discussion boards is to allow students to freely exchange ideas. It is imperative to remain respectful of all viewpoints and positions and, when necessary, agree to respectfully disagree. While active and frequent participation is encouraged, cluttering a discussion board with inappropriate, irrelevant, or insignificant material will not earn additional points and may result in receiving less than full credit. Frequency matters, but contributing content that adds value is paramount. Please remember to cite all sources—when relevant—in order to avoid plagiarism. Please post your viewpoints first and then discuss others' viewpoints.

The quality of your posts and how others view and respond to them are the most valued. A single statement mostly implying "I agree" or "I do not agree" is not considered to be a post. Explain, clarify, politely ask for details, provide details, persuade, and enrich communications for a great discussion experience. Please note, there is a requirement to respond to at least one post from a class member. I'm looking for insightful analysis, probing questions, and *constructive* suggestions to each other. Keep thinking from the perspective—how can I *add something useful?* It may be an experience you have had professionally or a quote from an article/website you come across. If it is the latter, cite it properly.

You are expected to participate actively with other students in one discussion each week. This discussion will be focused on a topic is related to the course assignments. You are also expected to post a comment to a second review and reflections question. This latter question provides for comments about the course, and how your studies are progressing. You are expected to participate in both forums with polished, well-structured and APA- compliant posts each week, adding references as needed. Be sure to check spelling and grammar.

It is highly desirable that your initial comments be posted Thursday so that follow-up comments can be made. The discussion forum is intended for exchange of ideas between students. Discussion Board responses are DUE by 11:59 pm CST on every Sunday. **Five points are available for the first discussion topic, two points for class participation and three points for the reflections topic making ten points total per week. Only one grade is entered each week for the Discussion Board.**

Participation and Attendance

This course will not meet at a particular time each week. All course goals, session learning objectives, and assessments are supported through classroom elements that can be accessed at any time. To measure class participation (or attendance), your participation in threaded discussion boards is required, graded, and paramount to your success in this course. Please note that any scheduled synchronous meetings are optional. While your attendance is highly encouraged, it is not required and you will not be graded on your attendance or participation.

Academic Integrity at Northwestern

Students are required to comply with University regulations regarding academic integrity. If you are in doubt about what constitutes academic dishonesty, speak with your instructor or graduate coordinator before the assignment is due and/or examine the University Web site. Academic dishonesty includes, but is not limited to, cheating on an exam, obtaining an unfair advantage, and plagiarism (e.g., using material from readings without citing or copying another student's paper). Failure to maintain academic integrity will result in a grade sanction, possibly as severe as failing and being required to retake the course, and could lead to a suspension or expulsion from the program. Further penalties may apply. For more information, visit The Office of the Provost's Academic Integrity page.

Instructional Observations

Please note that this course may have a faculty observer for this term. The observer is present exclusively for training purposes, and will not be responsible for, or engage in, any in-class interactions, student assessment or grading, or any other aspect of course delivery. Please contact the instructor if you have any questions.

Course Schedule

Sync Session

During the term there will be a sync sessions designed to: a) discuss statistical content, b) help with R programming and assignments, c) to convey knowledge wisdom about the profession of Data Science, and d) to just hold open office hours in case you need help. Attendance is optional. Presentations will be recorded and posted for your use by the next class period. Please check the Canvas Course site calendar via the SYLLABUS option for the schedule of sessions.

(All times mentioned below are in CST, Central Standard Time in Chicago, USA. Please be aware that after Nov 7 CST shifts 1 hour earlier. The Time in Calcutta is 10 and ½ hours later into the day than CST. To check your time zone go to https://canvas.northwestern.edu/profile/settings**.)**

Week 1 – Complete by Sunday

Learning Objectives

After this week, the student will be able to:

- List examples of statistical applications in business.
- Explain the difference between variables, measurement and data.
- Define and compare four different levels of data.
- Construct a frequency distribution and different data displays.
- Construct and interpret two-variable tables and scatter plots.
- Write simple programs using the language R.

Reading

- Black, K. *Business Statistics* (8th edition): Chapter 1 & Chapter 2.
- Course reserves – Chapters 1 & 2 from 10th edition of Business Statistics by Ken Black

R Installation:

- Installation of R and completion of *The Quick Start Guide to R* is expected this week.

Videos:

- Chapter 1 Lecture (Panopto)
- Chapter 2 Lecture (Panopto)
- R Lesson Video

Assignments

- Install R
- Study *The Quick Start Guide to R* and complete the exercises.

Week 2 – Complete by Sunday

Learning Objectives

After this week, the student will be able to:

- Calculate and apply measures of central tendency and variability.
- Describe a data distribution using a box-and-whisker plot.
- Interpret graphical data displays.
- Detect outliers using box plots.
- Perform calculations to trim data.

Reading

- Black, K. *Business Statistics* (8th edition): Chapter 3.

  Course Reserves:

- Wilcox R. R. *Basic Statistics* Chapter 2: Pages 22-29, Chapter3: Pages 34-45
- Chihara & Hetersberg: *Mathematical Statistics with Resampling and R* Chapter 2 (EDA)

Videos

- Chapter 3 Lecture (Panopto)
- R Lesson Video

Assignments

- Test 1

Week 3 -- Complete by Sunday

Learning Objectives

After this week, the student will be able to:

- Describe probability.
- Articulate the different methods of assigning probabilities.
- Understand and apply axioms and properties of probability.
- Compute probabilities under different conditions.
- Understand conditional probability and Bayes' theorem.
- Determine the mean, variance and standard deviation for a discrete variable.
- Solve problems using binomial and Poisson probability distributions.

Reading

- Black, K. *Business Statistics* (8th edition): Chapter 4 & Chapter 5.
- Downey, A. B. *Think Bayes* Chapter 1 pages 1-10 (Course Reserves on the course site.) Here is the link: http://www.greenteapress.com/thinkbayes/thinkbayes.pdf

Videos

- Chapter 4 Lecture (Panopto)
- Chapter 5 Lecture (Panopto)
- R Lesson Video

Assignments

- R Programming Assignment #1

Week 4 -- Complete by Sunday

Learning Objectives

After this week, the student will be able to:

- Explain what is a probability density function for a continuous variable.
- Compute the expected mean value and variance.
- Describe a standard normal distribution and its properties
- Use the standard normal distribution to find z-scores
- Convert distributions to standard normal.
- Use the normal distribution as an approximation to the binomial distribution.
- Explain different types of sampling plans.
- Explain the central limit theorem.

Reading

- Black, K. *Business Statistics* (8th edition): Chapter 6 & Chapter 7.

Course Reserves:

- Wilcox R. R. *Basic Statistics* Chapter 4 Section 7 (Pages 70-76)
- Wilcox R.R. *Fundamental of Modern Statistical Methods* – Chapter 3

Videos

- Chapter 6 Lecture (Panopto)
- Chapter 7 Lecture (Panopto)
- R Lesson Video

Assignments

- Test 2

Week 5 -- Complete by Sunday

Learning Objectives

After this session, the student will be able to:

- Estimate a population mean and a proportion.
- Define the t-distribution and determine probabilities given degrees of freedom.
- Use the chi-square distribution to estimate a population variance.
- Determine the sample size needed to estimate a population mean and a proportion.
- State what is a confidence interval and how it is used for statistical inference.
- Compute confidence intervals for a mean and a proportion.

Reading

- Black, K. *Business Statistics* (8th edition): Chapter 8.

Course Reserves:

- Wilcox R. R. *Basic Statistics* Chapter 6 Section 5 (Pages 121-129)
- Wilcox R.R. *Fundamental of Modern Statistical Methods* – Chapter 6

Videos

- Chapter 8 Lecture (Panopto)
- R Lesson Video

Assignments

- Data Analysis Assignment #1

Week 6 -- Complete by Sunday

Learning Objectives

After this session, the student will be able to:

- Develop one- and two-tailed hypotheses that can be tested.
- Develop test critical regions.
- Reach conclusions based on hypothesis tests
- Explain Type I and Type II errors.
- Perform hypothesis tests on means and proportions.
- Use p-values for hypothesis testing.
- Discuss statistical significance versus practical significance.

Reading

- Black, K. *Business Statistics* (8th edition): Chapter 9 and Chapter 16.

  Course Reserves:

- Wilcox R.R. *Fundamental of Modern Statistical Methods* – Chapter 5

Videos

- Chapter 9 Lecture (Panopto)
- Chapter 16 Lecture (Panopto)
- R Lesson Video

Assignments

- Test 3

Week 7 -- Complete by Sunday

Learning Objectives

After this session, the student will be able to:

- Develop hypotheses for testing the difference in means or proportions of two populations.
- Use the z-statistic to develop confidence intervals for the difference in two means.
- Perform two-sample t-tests, paired t-tests and construct confidence intervals.
- Develop confidence intervals for the difference in two population proportions.
- Test hypotheses about the difference in variance between two populations.

Reading

- Black, K. *Business Statistics* (8th edition): Chapter 10.

  Course Reserves:

- Wilcox R. R. *Basic Statistics* pages Chapter 9 Section 9.1 pages 184-193.

Videos

- Chapter 10 Lecture (Panopto)
- R Lesson Video

Assignments

- R Programming Assignment #2

Week 8 – Complete by Sunday

Learning Objectives

After this session, the student will be able to:

- Describe what is a designed experiment.
- Use a single factor AOV model for analysis.
- Recognize a randomized block design.
- Explain the advantages of a two-way AOV.
- Compute sums of squares and mean squares
- Use multiple comparison tests.
- Explain what is an interaction.
- Calculate correlations.
- Fit a simple linear regression equation.

Reading

- Black, K. *Business Statistics* (8th edition): Chapter 11 & Chapter 12 Sections 12.1-12.3.

  Course Reserves:

- Wilcox R. R. *Basic Statistics* Chapter 10 Section 10.1 pages 210-217.

Videos

- Chapter 11 Lecture (Panopto)
- Chapter 12 Lecture (Panopto)
- R Lesson Video

Assignments

- Test 4
- Practice Proctored Survey

Week 9 – Complete by Sunday

Learning Objectives

After this session, the student will be able to:

- Explain a simple linear regression model.
- Determine the equation of a simple linear regression line.
- Specify the two parameters of a straight line.
- Discuss the risks of extrapolation.
- Perform inference about regression coefficients.
- Calculate the Pearson product-moment correlation coefficient.
- Calculate standard errors and confidence intervals for regression coefficients.
- Test the overall model.
- Assess Model Adequacy.

Reading

- Black, K. *Business Statistics* (8th edition): Chapter 12 Sections 12.4-12.7 & Chapter 13.

  Course Reserves:

- Wilcox R. R. *Basic Statistics* Chapter 8 Section 8.3 pages 172-176.

Videos

- Chapter 12 Lecture (Panopto)
- Chapter 13 Lecture (Panopto)
- R Lesson Video

Assignments

- Data Analysis Assignment #2

Sync Session

This will be the Q&A session in preparation for the final exam.

Week 10 – Complete Sunday

Learning Objectives

- No new learning objectives.

Assignments

- Final Proctored Exam
- Final Take-home Exam: COVID-19

The Final Examination opens at 12:01 am CST Monday of Week 10. The Final Proctored Exam & the Take-home Final Exam are due by 11:59 pm CST Sunday of Week 10. For Summer Quarters only, due data is 11:59 pm CST Saturday of Week 10. **You are responsible for preparing for the proctored exam with Panopto. Please be aware access to the internet during the exam is not permitted. Separate portable devices such as iPads and Kindles are not permitted. See the course site for instructions.**

There is a practice test with solutions and a sample data analysis posted in Canvas under Week 10. ***Practice Problems*** are posted in the weekly modules. These are practice problems and carry no point value. They may be attempted at any time during the course.