



NORTHWESTERN
UNIVERSITY

SCHOOL OF
CONTINUING
STUDIES

PREDICT 401: Introduction to Statistical Analysis – Section 56 Summer 2012

Matthew S McCallum, PhD

matthew-mccallum@northwestern.edu

Course Description

Students learn to apply statistical techniques to the processing and interpretation of data from various industries and disciplines. Topics covered include probability, descriptive statistics, study design and linear regression. Emphasis will be placed on the application of the data across these industries and disciplines, in order to develop a core thought process through the entire Predictive Analytics curriculum.

Required Text

Sirkin, R.M. (2006). *Statistics for the social sciences* (3rd ed.). Thousand Oaks, CA: Sage.
[ISBN-13: 978-1412905466]

Software

IBM SPSS Statistics Premium GradPack 19 or 20 for Windows OR Mac OS-X

Prerequisites

NONE

Learning Goals

The goals of this course are to:

- Conduct statistical analysis.
- Interpret statistical information.
- Evaluate statistical information.

Evaluation

The student's final grade will be determined as follows:

- Participation 25% (100 possible points)
- Midterm 25% (100 possible points)
- Final 50% (100 possible points online non-proctored, 100 possible points proctored)

Grading Scale

A = 93–100% (372–400 points)
A- = 90–92% (360–371 points)
B+ = 87–89% (348–359 points)
B = 83–86% (332–347 points)
B- = 80–82% (320–331 points)
C+ = 77–79% (308–319 points)
C = 73–76% (292–307 points)
C- = 70–72% (280–291 points)
F = 0–69% (0–279 points)

Discussion Board Etiquette & Procedure

The purpose of the discussion board groups 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 is not unimportant, but content of the message is paramount. Please remember to cite all sources (when relevant) in order to avoid plagiarism.

Attendance and participation is a critical part in forming a dynamic learning experience, in school as well as on the job. As a result, you are expected to participate in activities and discussions within your discussion board group. The discussion questions will be answered each week in threaded topics under the weekly 'discussion' tabs. The expectation is several quality posts per week per person. I expect these to be polished, well-structured and APA-compliant posts. Even though they are only discussions, you still must include references and check for spelling and grammar. Since you have a spell-checker built in, this should not be an onerous task.

I will be the facilitator of this process and will be assessing your contributions to the topic-related discussions. 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/web site you come across. If it is the latter, please do not forget to cite it properly.

Problem Sets

In sessions two through nine you will be given a set of problems to do in order to test your understanding of the concepts being taught. In addition to the problem sets, there will be solution sets posted. It is my strong suggestion that you do all of the problems before looking at the solutions. If you are having problems any step of the way, we can discuss them in the 'Q & A' discussion forum.

Proctored Assessment

There is a proctored assessment requirement for this class. Please see the Assignments section in Blackboard for more information.

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 class. Please note that any scheduled synchronous or "live" meetings are considered supplemental and optional. While your attendance is highly encouraged, it is not required and you will not be graded on your attendance or participation.

Late Work

Unless otherwise noted, all work is due on the assigned day by 11:55 PM (Central Time). This includes exams and participation in the discussions. Late work is not accepted.

One more piece of advice—do not fall behind. We cover a lot of material in this course, and falling behind is the primary reason why folks fail. To that end, you have below the due dates for the entire course. It is much, much better to be ahead than behind.

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 www.scs.northwestern.edu/student/issues/academic_integrity.cfm.

Plagiarism is one form of academic dishonesty. Students can familiarize themselves with the definition and examples of plagiarism, by visiting [<www.northwestern.edu/uacc/plagiar.html>](http://www.northwestern.edu/uacc/plagiar.html). A myriad of other sources can be found online.

Some assignments in SCS courses may be required to be submitted through SafeAssign, a plagiarism detection and education tool. You can find an explanation of the tool at <http://wiki.safeassign.com/display/SAFE/How+Does+SafeAssign+Work>. In brief, SafeAssign compares the submitted assignment to millions of documents in large databases. It then generates a report showing the extent to which text within a paper is similar to pre-existing sources. The user can see how or whether the flagged text is appropriately cited. SafeAssign also returns a percentage score, indicating the percentage of the submitted paper that is similar or identical to pre-existing sources. High scores are not necessarily bad, nor do they necessarily indicate plagiarism, since the score does not take into account how or whether material is cited. If a paper consisted of one long quote that was cited appropriately, it would score 100%. This would not be plagiarism, due to the appropriate citation. However, submitting one long quote would probably be a poor paper. Low scores are not necessarily good, nor do they necessarily indicate a lack of plagiarism. If a 50-page paper contained all original material, except for one short quote that was not cited, it might score around 1%. But, not citing a quotation is still plagiarism, as is repurposing of one's own work without citation.

SafeAssign includes an option in which the student can submit a paper and see the resultant report before submitting a final copy to the instructor. This ideally will help students better understand and avoid plagiarism.

Other Processes and Policies

Please refer to your SCS student handbook at www.scs.northwestern.edu/grad/information/handbook.cfm for additional course and program processes and policies.

Course Schedule

Important Note: Changes may occur to the syllabus at the instructor's discretion. When changes are made, students will be notified via an announcement in Blackboard and by email.

Session 1

Learning Objectives

After this session, the student will be able to:

- Describe the relationships between prediction, co-incidence (correlation), and causation.
- Describe different forms of data and levels of measurement.
- Explain the importance of defining variables in statistics.

Course Content

Textbook Reading

Sirkin, Chapters 1–3

Discussion Board

This session you are required to participate in two threads of the session-specific discussion board forum. Your participation in both posting and responding to other students' comments is graded. For this session's discussion topics, visit the discussion board in Blackboard.

Assignments

NONE

Sync Session

Saturday, June 23, 2012 - 1pm Central US Time

Session 2**Learning Objectives**

After this session, the student will be able to:

- Identify measures of dispersion and their purposes.
- Identify the components of tabular and visual statistics.
- Explain the components of tabular and visual statistics.
- Define mean, median, and mode.
- Apply mean, median, and mode.
- Calculate the mean, median, and mode for individual and grouped data.
- Calculate measures of dispersion.
- Use SPSS to calculate mean, median, mode, and measures of dispersion.
- Define a frequency distribution.
- Describe the shapes of frequency distributions.
- Articulate the definition, advantages, and disadvantages of measures of dispersion.

Course Content**Reading**

Sirkin, Chapters 4 and 5

Problem Set 1**Discussion Board**

This session you are required to participate in two threads of the session-specific discussion board forum. Your participation in both posting and responding to other students' comments is graded. For this session's discussion topics, visit the discussion board in Blackboard.

Assignments

NONE

Sync Session

NONE

Session 3**Learning Objectives**

After this session, the student will be able to:

- Define probability.
- Articulate the properties of random variables.
- Apply the rule of addition for probabilities for mutually exclusive events.
- Apply the rule of multiplication of probabilities for independent events.
- Define the central limit theorem (CLT).

Course Content**Reading**

Sirkin, Chapters 7 and 8

Multimedia

Probability

Normal Probability Distribution

Problem Set 2**Discussion Board**

This session you are required to participate in two threads of the session-specific discussion board forum. Your participation in both posting and responding to other students' comments is graded. For this session's discussion topics, visit the discussion board in Blackboard.

Assignments

NONE

Sync Session

NONE

Session 4**Learning Objectives**

After this session, the student will be able to:

- Differentiate samples and populations.
- Describe the use of the CLT to determine the properties of the probability distribution for a sampling distribution.
- Compute the standard error for a sample mean based on n observations.
- Compute the confidence interval for a sample mean based on n observations.
- Describe when the t -distribution should be used.
- Determine the critical t -value for a stated significance level and sample size.

Course Content**Reading**

Sirkin, Chapters 7 and 8

Multimedia

Confidence Intervals

Problem Set 3**Discussion Board**

This session you are required to participate in two threads of the session-specific discussion board forum. Your participation in both posting and responding to other students' comments is graded. For this session's discussion topics, visit the discussion board in Blackboard.

Assignments

NONE

Sync Session

NONE

Session 5**Learning Objectives**

After this session, the student will be able to:

- State the null and alternative hypotheses for a general research hypothesis.
- Conduct a one-sample test of a sample mean.
- Distinguish between Type I and Type II errors in hypothesis testing.

Course Content**Reading**

Meier, *Applied Statistics*, pages 189–209 and 551

Handout

Hypothesis Testing

Problem Set 4**Discussion Board**

This session you are required to participate in two threads of the session-specific discussion board forum. Your participation in both posting and responding to other students' comments is graded. For this session's discussion topics, visit the discussion board in Blackboard.

Assignments

The “take-home” Midterm Examination will be posted Monday, July 16, 2012 by 12pm Noon (Central US Time) for downloading and is due Sunday, July 22, 2012 at 11:55 p.m. (Central US Time). For more information, click Assignments on the left navigation panel in Blackboard, and scroll to this assignment's item.

Sync Session

Saturday, July 21, 2012 - 1pm Central US Time

Session 6**Learning Objectives**

After this session, the student will be able to:

- State the basic assumptions for a t-test of the difference between two sample means.
- State the hypotheses for the t-test.
- Calculate the combined standard error for the two populations with unequal variances.
- Explain the analysis of variance (ANOVA) method for comparing three or more sample means in terms of between groups and within groups sum of squares.

Course Content**Reading**

Sirkin, Chapters 9 and 10

Problem Set 5**Discussion Board**

This session you are required to participate in two threads of the session-specific discussion board forum. Your participation in both posting and responding to other students' comments is graded. For this session's discussion topics, visit the discussion board in Blackboard.

Assignments

NONE

Sync Session

NONE

Session 7

Learning Objectives

After this session, the student will be able to:

- Specify the two parameters of a straight line.
- Define the “best equation” based on minimization of the residual sum of squares.
- Identify the key assumptions of linear regression.
- Interpret output generated by SPSS for linear regression.
- Define coefficient of determination .
- Explain the null and alternative hypotheses that are tested by the F-statistic.
- Explain the null and alternative hypotheses that are tested by the t-statistic.
- Use SPSS to create a least squares equation based on dummy (categorical) variables.
- Describe the properties, assumptions, and limitations of the Pearson linear correlation.
- Calculate a Pearson linear correlation.

Course Content

Reading

Sirkin, Chapter 13 and Chapter 14 up to page 508.

Multimedia

Linear Regression Analysis

Problem Set 6

Discussion Board

This session you are required to participate in two threads of the session-specific discussion board forum. Your participation in both posting and responding to other students' comments is graded. For this session's discussion topics, visit the discussion board in Blackboard.

Assignments

NONE

Sync Session

NONE

Session 8**Learning Objectives**

After this session, the student will be able to:

- Describe the Bernoulli process.
- Calculate the binomial probabilities given sample size, number of successes, and probability of success.
- Estimate binomial probabilities with the Normal probability distribution.
- Use the binomial probability to estimate the mean and standard deviation of a sample proportion.
- Calculate the confidence interval for a proportion using estimates based on the binomial probability distribution.

Course Content**Reading**

Meier, *Applied Statistics*, pages 151–161 and 550

Handout

Binomial Probability Distribution

Problem Set 7**Discussion Board**

This session you are required to participate in two threads of the session-specific discussion board forum. Your participation in both posting and responding to other students' comments is graded. For this session's discussion topics, visit the discussion board in Blackboard.

Assignments

NONE

Sync Session

NONE

Session 9**Learning Objectives**

After this session, the student will be able to:

- Describe the components of a contingency table.
- Use a contingency table to estimate the relationship between dependent and independent variables.
- State the null and alternative hypothesis for the chi-square test for independence and the chi square test for goodness of fit.
- Calculate the chi-square statistic for a contingency table.
- Calculate the degrees of freedom of a chi-square statistic.
- Test the significance of a chi-square statistic
- Identify the properties of measures of association (MOA) for ordinal and nominal variables.
- Calculate the gamma MOA for a contingency table.
- Calculate the lambda MOA for a contingency table.

Course Content**Reading**

Sirkin, Chapters 6, 11, and 12

Problem Set 8**Discussion Board**

This session you are required to participate in two threads of the session-specific discussion board forum. Your participation in both posting and responding to other students' comments is graded. For this session's discussion topics, visit the discussion board in Blackboard.

Assignments

NONE

Sync Session

NONE

Session 10**Learning Objectives**

After this session, the student will be able to:

- No new learning objectives.

Course Content

NONE

Discussion Board

This session you are required to participate in the session-specific discussion board forum. Your participation in both posting and responding to other students' comments is graded. For this session's discussion topic, visit the discussion board in Blackboard.

Assignments

The Final Examination is due Sunday, August 26, 2012 at 11:55 pm (Central US Time). For more information, click Assignments on the left navigation panel in Blackboard, and scroll to this assignment's item.

The *online non-proctored* "take-home" portion of the Final Examination will be posted Monday, August 20, 2012 by 12pm Noon (Central US Time) for downloading. You are responsible for scheduling the *online proctored* portion of the exam sometime during Session 10 as well. Both take-home and proctored portions of the exam are due Sunday, August 26, 2012 at 11:55 p.m. (Central US Time).

Sync Session

NONE