## Roy L. Sanford, Ph.D.

roy.sanford@northwestern.edu

## Course Description

This course teaches students the fundamentals of conducting statistical analyses.  This includes interpreting and evaluating statistical information. Topics covered include descriptive statistics, including central tendency, dispersion, and data display; probability; binomial and normal distributions; inference, including confidence intervals and hypothesis testing; correlation; bivariate regression; contingency tables and chi-square.  Selected contemporary statistical concepts are introduced to supplement traditional statistical methods.

## Learning Goals

At the conclusion of the course, students will be able to: 1) perform statistical analyses, 2) interpret statistical information, 3) evaluate statistical information, and perform data analysis using the language R.

## Prerequisites

PREDICT 400 is a prerequisite for MSPA students who begin graduate studies fall term 2014 and thereafter.

## Texts

Required:

Black, K. (2013). *Business Statistics.* 8th ed. Hoboken, NJ: John Wiley & Sons, Inc. [ISBN-13: 978-1118876824]

Wilcox, R. R. (2009). *Basic Statistics*  New York, NY: Oxford University Press  [ISBN-13: 978-0195315103]

Recommended:

Chang, Winston (2013).  *R Graphics Cookbook*  1st ed.  Sebastopol, CA:  O'Reilly Media, Inc. [ISBN-13:  978-1449316952]

Lander, J. P. (2014). *R for everyone: Advanced analytics and graphics.* Upper Saddle River, N.J.: Addison-Wesley/Pearson Education. [ISBN-13 978-0321888037]

Verzani, J. (2014). *Using R for Introductory Statistics* 2nd ed. Boca Raton, FL:  CRC Press (Taylor and Francis Group)  [ISBN-13 978-1466590731]

## Library Reserves

Chihara, L. and Hesterberg, T. (2011).  *Mathematical Statistics with Resampling and R.*  Hoboken, NJ:  John Wiley & Sons, Inc. [ISBN-13: 978-1118029855]

Downey, A. B. (2013).  *Think Bayes.*  Sebastopol, CA:  O'Reilly Media, Inc. [ISBN-13:  978-1449370787]

(A pdf copy of the book by Downey may be downloaded from the internet at www.greeteapress.com/thinkbayes/. )

**Texts Available for Download from Course Site**

Stowell, S. (2014). *Using R for* statistics. New York: Apress (distributed by Springer Science+Business Media) [ISBN-13: 978-1484201404]

Wilcox, R. R. (2010). *Fundamentals of Modern Statistical Methods*. 2nd ed. New York: Apress (distributed by Springer Science+Business Media) [ISBN-13: 978-1441955241]

The books by Stowell and Wilcox are posted on the course site in the module titled "Statistics Texts". They may be downloaded from there. The book by Stowell is referenced during the course and has valuable material regarding the use of R for statistical analysis. The book by Wilcox supplements the Wilcox text required for this course. These books, and other selected titles, may be downloaded from the Springer library. Access this library at http://link.springer.com.turing.library.northwestern.edu/ with your NetID and password.

## Software

Predict 401 will use the high-level language R. R can be obtained at *http://cran.r-project.org/*. Instructions are given in course shell as well as in the texts by Lander or Verzani. RStudio is recommended for new users of R. It is not required. Everything students need to do in this course can be accomplished using the standard R console with a plain text editor. RStudio is an integrated development environment. It is described in the text by Verzani. Installer packages for RStudio are located at *www.**rstudio**.com/*. The installation process is straightforward.

## Evaluation

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

- Participation                  20% (100 possible points)
- Four Quizzes                   20% (100 possible points online, non-proctored)
- Two Data Analysis Assignments  40% (100 possible points)
- Final exam                     20% (100 possible points online, proctored)

**Reports:** The data analysis assignments entail use of R. Instructions appear in the course shell along with the data.

## Proctored Final Exam

Students need to be aware of the proctoring procedure used within Predictive Analytics. ProctorU, an independent organization, does the final exam proctoring. The proctor must be able to monitor the student using a webcam, and also view the screen images seen by the student. The exam is taken within Canvas using the course site. Access to the internet during the exam is not permitted. Access to and use of WileyPlus, texts, printed materials, notes and all files stored on the computer is permitted during the exam. Computational applications such as R, Excel or equivalent are permitted during the exam including calculators, such as a TI 84, Casio or comparable. Separate portable devices such as iPads and Kindles are not permitted since the proctor cannot monitor the display on those devices. This means a student who uses a device such as an iPad or Kindle must plan ahead and be prepared to migrate the necessary files to the computer used during 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.

## 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)

## Discussion Board Etiquette

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. Content of the message is paramount. Please remember to cite all sources (when relevant) in order to avoid plagiarism.

You are expected to log on and participate actively on a regular basis each week. You are expected to respond to three discussion threads each week. The first two questions are directly related to the week's topics and the last will be a reflection question, where you will be asked to tie the ideas from the week's materials to your profession, hobbies, etc. Additionally, you are expected to participate in discussions with others.  I expect several polished, well-structured and APA-compliant posts each week. Please add  references and check spelling and grammar.

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, cite it properly.

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.  The discussion topics will become unavailable at 8:00 pm CST each Sunday.  I monitor the discussions and do my best to provide feedback in the *Reflections*  discussions. Discussion grades will be provided no later than Tuesday after each session ends.

## Attendance

This course will not meet at a particular time each week. All course goals, session learning objectives, and assessments are supported through course site elements that can be accessed at any time.  Your participation in threaded discussion boards is required, graded, and paramount to your success in this class.

There will be five scheduled sync sessions.  They are listed in the Course Schedule.  They will also be announced on the course site.  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.  Recordings are made an d made available along with handouts.

## Due Dates and Late Work

When stating due dates for work, the abbreviation "CST" is used.  "CST" is to always be taken to mean 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 8 pm CST deadline is a 6 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. Late work is not accepted without prior arrangement.  Without prior arrangement, an assignment turned in more than 24 hours late automatically receives a 25% deduction.  Communication with the instructor is essential in these matters.

One more piece of advice—**do not fall behind**. We cover a great deal of material in this course, and falling behind is the primary reason why folks have difficulty particularly toward the end of the course.  To that end, the Course Schedule gives you 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. For more information, visit:  www.scs.northwestern.edu/student/issues/academic_integrity.cfm .

## 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

## Session 1 – Complete by Sunday, September 27, 2015

### Learning Objectives

After this session, the student will be able to:
- List examples of statistical applications in business.
- Describe different types of data displays.
- Define important statistical terms.
- Explain the difference between variables, measurement and data.
- Define and compare four different levels of data.
- Construct a frequency distribution.
- Construct different types of data graphs.
- Construct and interpret two-variable tables and scatter plots.
- Write simple programs using the language R.

### Course Content
**Reading**
Black, K. *Business Statistics* Chapters 1 & 2 pages 2-49.

**Videos**
Levels of Data Measurement
Stem-and-Leaf Plot

### Assignments –

- Install R
- Read *The Quick Start Guide to R*

This session requires installation of R and completion of *The Quick Start Guide to R*. Completion of the tutorials in lynda.com is encouraged. Future sessions will have one or more optional practice lessons using R for data analysis. Each lesson is numbered corresponding to a chapter in *Business Statistics*. These lessons, the associated data, solution code and answers are posted within the course shell. These lessons carry no point value.

Instructions on how to access the WileyPlus materials are posted in the module *Business Statistics Videos + WileyPLUS*. Students should be aware that the videos listed above and in other sessions are available in WileyPlus and can be downloaded and retained for future viewing. Instructions on how to do this appear in the course shell.

Students should be aware that there is a required self-administered test to be completed by the end of Session #2. This test will cover material from Sessions 1 and 2. This is open book, open notes. R may be used for computation. This test has 25 points toward the final grade.

Students should be aware the first Data Analysis Assignment is due the end of Session #5. This assignment should be started early in the quarter. Do not wait until the last minute. The assignment instructions and data are available in a separate module on the course site. This report carries 100 points toward the final grade.

### Sync Session

There will be a sync session the first week of class Thursday, September 24, 2015 from 7 PM to 9 PM CST. Attendance is optional. A recording of the session will be posted in class the following day. The first session will address course requirements, schedule and some course content.

## Session 2 – Complete by Sunday, October 4, 2015

### Learning Objectives
After this session, 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 by two methods.
- Trim data, winsorize data and perform calculations.

## Course Content

**Reading**
Black, K. *Business Statistics* Chapter 3 pages 52-96.

Wilcox R. R. *Basic Statistics* Chapters 2 & 3 pages 9-45.

**Videos**
Computing Variance and Standard Deviation
Understanding and Using the Empirical Rule

## Assignments

Complete the self-administered test by 11:55 pm CST Sunday, October 4, 2015. This test will cover material from Sessions 1 and 2. It carries 25 points toward the final grade.

Students should be aware the first Data Analysis Assignment is due the end of Session #5. This assignment should be started early in the quarter. Do not wait until the last minute. The assignment instructions and data are available in a separate module on the course site. This report carries 100 points toward the final grade.

Optional R Lessons 1, 2 & 3 are posted within the course session module. These are practice problems and carry no point value. Students are encouraged to complete these lessons as they provide a way to apply course concepts and learn about R.

## Sync Session

**None**

# Session 3 -- Complete by Sunday, October 11, 2015

## Learning Objectives

After this session, 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.

## Course Content

**Reading**
Black, K. *Business Statistics* Chapters 4 & 5 pages 98-182.

Downey, A. B. *Think Bayes* Chapter 1 pages 1-10.

**Videos**
Constructing and Solving Joint Probability Tables
Solving Probability Word Problems
Solving Binomial Distribution Problems, Part I
Solving Binomial Distribution Problems, Part II

## Assignments

Students should be aware a self-administered test is required by the end of Session 4.  This test will cover Sessions 3 and 4 and count for 25 points toward the final grade.

Students should be aware the first Data Analysis Assignment is due the end of Session #5.  This assignment should be started early in the quarter.  Do not wait until the last minute. The assignment instructions and data are available in a separate module on the course site.  This report carries 100 points toward the final grade.

Optional R Lessons 4 and 5 are posted within the course shell.  These are practice problems and carry no point value.  Students are encouraged to complete these lessons as they provide a way to apply course concepts and learn about R.


## Sync Session

There will be a sync session Thursday, October 8, 2015 from 7 PM to 9 PM CST. Attendance is optional.  A recording of the session will be posted in class the following day.


# Session 4 -- Complete by Sunday, October 18, 2015

## Learning Objectives
After this session, 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, and convert distributions to standard normal.
- Use the normal distribution as an approximation to the binomial distribution.
- Explain what are jointly distributed random variables.
- Calculate expected values for jointly distributed random variables.
- Explain different types of sampling plans.
- Calculate statistics using sampling distributions.
- Explain the central limit theorem.

## Course Content
**Reading**
Black, K. *Business Statistics* Chapters 6 & 7 pages 184-255.

Wilcox R. R. *Basic Statistics* Chapter 4 pages 70-76 and Chapter 5 pages 77-101.

**Videos**
Solving Problems Using the Normal Curve
Solving for Probabilities of Sample Means using the z Statistic


## Assignments

Complete the self-administered test by 11:55 pm CST Sunday, October 18, 2015.  This test will cover material from Sessions 3 and 4.  It carries 25 points toward the final grade.

Students should be aware the first Data Analysis Assignment is due the end of Session #5.  This assignment should be started early in the quarter.  Do not wait until the last minute. The assignment instructions and data are available in a separate module on the course site.  This report carries 100 points toward the final grade.

Optional R Lessons 6 and 7 are posted within the course session module.


## Sync Session
None

## Session 5 -- Complete by Sunday, October 25, 2015

### 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.
- Discuss the implications of non-normality and outliers when constructing confidence intervals.

### Course Content
**Reading**
Black, K. *Business Statistics* Chapter 8 pages 260-296.

Wilcox R. R. *Basic Statistics* Chapter 6 pages 121-126 and Chapter 7 pages 146-147.

**Videos**
Confidence Intervals
Determining Which Inferential Technique to Use: Confidence Intervals

### Assignments

Complete and submit the first Data Analysis Assignment by 11:55 pm CST Sunday, October 25, 2015. This assignment carries 100 points toward the final grade. The assignment instructions and data are available in a separate module on the course site.

Students should be aware a self-administered test is due the end of Session 6. This test will cover Sessions 5 and 6. It carries 25 points toward the final grade.

Optional R Lesson 8 is posted within the course shell. These are practice problems and carry no point value.

### Sync Session

There will be a sync session Thursday, October 22, 2015 from 7 PM to 9 PM CST. Attendance is optional. A recording of the session will be posted in class the following day.


## Session 6 -- Complete by Sunday, November 1, 2015

### 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.


### Course Content
**Reading**
Black, K. *Business Statistics* Chapter 9 pages 298-352 and Chapter 16 pages 664-686.

**Videos**
   Hypothesis Testing Using the z Statistic
   Establishing Hypotheses
   Understanding p-values
   Type I and Type II Errors
   Two-Tailed Tests

## Assignments

Complete the self-administered test by 11:55 pm CST Sunday, November 1, 2015.  This test covers material from Sessions 5 and 6.  It carries 25 points to the final grade.

Students should be aware the second Data Analysis Assignment is due the end of Session 10.  Do not wait until the last minute to get started.  The assignment instructions and data are available in a separate module on the course site.  This report carries 100 points toward the final grade.

Optional R Lesson 9 is posted within the course session module of the course shell.  These are practice problems and carry no point value.

## Sync Session

None.

# Session 7 -- Complete by Sunday, November 8, 2015

## Learning Objectives
   After this session, the student will be able to:

* Develop hypotheses for testing the difference between two populations.
* Develop and test hypotheses for testing the difference in means for two independent populations.
* Develop and test hypotheses for testing the difference in means between two dependent populations.
* Calculate confidence intervals for the difference in means.
* Test hypotheses about the difference in variance between two populations.
* Determine sample sizes.
* Perform a two-sample t-test and construct a confidence interval.
* Perform a paired t-test and construct a confidence interval.
* Make inferences about two population variances.

## Course Content

**Reading**
Black, K. *Business Statistics* Chapter 10 pages 354-411.

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

**Videos**
Hypothesis Tests of the Difference in Means of Two Independent Populations Using the t Statistic

## Assignments

Students should be aware that a self-administered test is due the end of Session 8.  This test will cover Sessions 7 and 8.  It carries 25 points toward the final grade.

Students should be aware the second Data Analysis Assignment is due the end of Session 10.  Do not wait until the last minute to get started.  The assignment instructions and data are available in a separate module on the course site.  This report carries 100 points toward the final grade.

Optional R Lesson 10 is posted within the course session module.  These are practice problems and carry no point value.

**Sync Session**

There will be a sync session Thursday, November 5, 2015 from 7 PM to 9 PM CST. Attendance is optional. A recording of the session will be posted in class the following day.


# Session 8 – Complete by Sunday, November 15, 2015

## Learning Objectives

After this session, the student will be able to:

- Describe an experimental design.
- 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.


## Course Content

**Reading**

Black, K. *Business Statistics* pages Chapter 11 pages 415-472.

Wilcox R. R. *Basic Statistics* Chapter 10 pages 210-219.

**Video**

Computing and Interpreting a One-Way ANOVA


## Assignments

Complete the self-administered test by 11:55 pm CST Sunday, November 15, 2015. This quiz covers material from Sessions 7 and 8. It carries 25 points toward the final grade.

Students should be aware the second Data Analysis Assignment due at the end of Session 10. Do not wait until the last minute to get started. The assignment instructions and data are available in a separate module on the course site. This carries 100 points toward the final grade.

Optional R Lesson 11 is posted within the course session module. These are practice problems and carry no point value.


## Sync Session

None.


# Session 9 – Complete by Sunday, November 22, 2015

## 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.
- Define the "best equation" based on minimization of the residual sum of squares.
- 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.

## Course Content
**Reading**

Black, K. *Business Statistics* Chapter 12 pages 476-527 .

Wilcox R. R. *Basic Statistics* Chapter 8 pages 172-176 and Chapter 13 pages 283-286.

**Videos**

Testing the Regression Model I—Predicted Values, Residuals, and Sum of Squares of Error
Testing the Regression Model II—Standard Error of the Estimate and $r^2$

## Assignments

Students should be aware that the proctored final exam opens at 12:01 am CST Monday, November 15, 2015. The Final Examination is due by 11:55 pm CST Sunday, August 30, 2015. This is an open book, open notes Proctored exam. For calculations Excel, R or some other computational aid may be used. **You are responsible for scheduling this proctored exam.**

Students should be aware the second Data Analysis Assignment due at the end of Session 10. Do not wait until the last minute to get started. The assignment instructions and data are available in a separate module on the course site. This carries 100 points toward the final grade.

Optional R Lesson 12 is posted within the course session module. These are practice problems and carry no point value.

## Sync Session

There will be a sync session Thursday, November 19, 2015 from 7 PM to 9 PM CST. Attendance is optional. A recording of the session will be posted in class the following day. This session will be a Q&A session in preparation for the final.

# Session 10 – Complete Sunday, December 6, 2015

## Learning Objectives
- No new learning objectives.

## Course Content
None

## Assignments

Complete and submit the second Data Analysis Assignment by 11:55 pm CST Sunday, December 6, 2015. The assignment instructions and data are available in a separate module on the course site. This assignment carries 100 points toward the final grade.

The Final Examination opens at 12:01 am CST Monday, November 15, 2015. It is due by 11:55 pm CST Sunday, December 6, 2015. This is an open book, open notes Proctored exam. For calculations Excel, R or some other computational aid may be used. For more information, click Assignments on the left navigation panel in Blackboard, and scroll to this assignment's item. **You are responsible for scheduling this proctored exam.**

## Sync Session

None