Shepherd University
Department of Computer Sciences,
Mathematics, and Engineering
P.O. Box 5000
Shepherdstown, WV
www.shepherd.edu

# Math 329 Mathematical Modeling: Course Introduction

— Fall 2016 —

Credit hours : 3
Course text : *An Introduction to Statistical Learning* by James, Witten, Hastie and Tibshirani. 2014.
www-bcf.usc.edu/∼gareth/ISL
Instructor : Dr. Ralph Wojtowicz
Office : 204C Stutzman-Slonaker
Email : rwojtowi@shepherd.edu
Office Hours : Mondays 1:00 – 4:00, Wednesdays 1:00 – 2:00, Fridays 1:00 – 3:00 or by appointment

**Prerequisites**. Math 318, 321, and 310 (or permission of instructor)

**Course Overview**. In this course we study the interface between mathematics and the environment through which we navigate. We will typically start with a dataset: a collection of repeated measurements or observations about some phenomenon. Each measurement involves a vector $X = (X_1, \ldots, X_p)$ of *independent variables*, *inputs*, *features* or *predictors*. In *supervised learning* problems, we also have a value $Y$ called the *dependent variable*, *output* or *response*. We seek to predict $Y$ from $X$ or, in some cases, infer properties of the relationship between $X$ and $Y$. In *unsupervised learning* problems, we do not have an output $Y$. Instead, we try to organize or cluster the inputs $X$ from repeated experiments into cohesive groups. In this course, we will focus on supervised learning.

Mathematically, we represent the relationship between inputs and outputs by

$$Y = f(X_1, \ldots, X_p) \; + \; \epsilon$$

where the function $f$ represents the *systematic* information that $X$ provides about $Y$ and $\epsilon$ is a random error term. There are many challenges involved in using this representation:

- We don't know $f$ so, we must estimate it by constructing a *model* $\widehat{f}$. Then we use the model to compute an estimate $\widehat{Y} = \widehat{f}(X)$ of $Y$. Linear regression, logistic regression, linear discriminant analysis and $K$-nearest neighbors are methods that we will study for estimating $f$.
- We may not have measured the inputs $X$ that we need to get a good estimate of $Y$.
- We may not have enough data to get a good estimate of $f$.
- The model $\widehat{f}$ may be too *flexible*. That is, it may change a lot if the input data changes.
- The model $\widehat{f}$ may not be flexible enough to approxiate the $f$. That is, $\widehat{f}$ may have high *bias*.
- The random (or *irreducible*) error $\epsilon$ may be large or may itself depend on $X$.

**Software**. We will make daily use of two freely-available software packages:

- The `R` environment for statistical computing: www.r-project.org
- The LaTeX document preparation system: www.tug.org

They are available for Linux, MacOSX and MS Windows. Please download, install and test these on your personal computer as soon as possible.

**Assignments and Grading**. We will cover Chapters 1–5 very thoroughly. That is 200 pages in 14 weeks. In other words, $200/(14 * 7) \approx 2$ pages to read per day or $200/42 \approx 4.8$ pages to cover in class each day. One ongoing homework assignment is to carefully study your text. Each chapter ends with a set of Labs, a set of Exercises on concepts, and a set of applied Exercises. The Labs are `R` examples and demonstrations. We will do most of these during class. They often illustrate `R` commands that can be used to solve the applied Exercises. The concept exercises typically ask you to prove statements from the book, do manual calculations using equations from the text, or write essays related to text discussions.

Semester grades will be based on the final exam, weekly technical reports, and occasional quizzes. The reports will be solutions to 1–3 exercises from the text. See the course syllabus for the exercises and due dates. Reports must be written using LaTeX and emailed to me as a on the due date. Reports written using some other word processing system will not be graded. Late reports will be penalized. An incomplete solution is better than a late one. The idea to keep in mind is that real world research problems can be very complex so, an initial goal is often simply to get a rough, reasonable answer. If time allows, one then develops improved solutions.

The final exam will be held at the officially-scheduled time in our usual classroom. For the final exams schedule, see www.shepherd.edu/register/finals_schedule.html.

**Grading Policy**. Final grades will computed using the following components.

Quizzes 10% Reports 70% Final Exam 20%

Mid-term and semester grades will be calculated according to the following scale.

A = (90–100) B = (80–89) C = (70–79) D = (60–69) F = (below 60)

**Attendance**. You are responsible for the material discussed or distributed in class. Cell phones should be turned off in class. Please be on time and respectful of other students while in class.

**Late Work Policy**. Missed exams will not be graded except in the case of an emergency or other exceptional event. Written documentation such as a doctor's note must be provided.

**Students in Athletics**. If you will miss an assignment, quiz or exam due to a school athletic event, it is your responsibility to provide a letter signed by your coach at least one week ahead of time.

**Academic Integrity**. Student in this course are expected to abide by the Academic Integrity Procedures on page 154 of the Student Handbook — www.shepherd.edu/students/studenthandbook.pdf Academic dishonesty will not be tolerated and is grounds for dismissal from and failure in this course.

**Academic Support Center**. You are encouraged to make use of the free services (including tutoring) available through the Academic Support Center. See www.shepherd.edu/ascweb.

**Disability Support Services**. Disability Support Services at Shepherd University believes that every student should succeed, and works closely with students to meet their needs. Students requesting any disability related accommodation should contact the Disability Coordinator at 304-876-5453. This includes students with learning disabilities needing classroom accommodations, students requesting specific housing accommodations for health-related reasons, and all other disability accommodations. Accommodations need to be documented and provided to instructors. Please see www.shepherd.edu/mcssweb/dss.

**RAVE Alerts**. Students are encouraged to sign up for RAVE alerts and check the Shepherd website in order to be informed of campus closures. See www.shepherd.edu/university/rave.

**LEAP Goals**. Shepherd University has adopted goals from the American Association of Universities and Colleges (AAC&U) Liberal Education and America's Promise (LEAP) initiative with minor modifications. Goals 1 and 2 are relevant to this course. See www.aacu.org/leap.

**Core Curriculum Competencies**. This course addresses the following competencies: critical thinking, life-long learning, creative thinking, quantitative literacy, and problem solving.

**Learning Outcomes**. Students in this course will:

- perform basic computational/arithmetic operations;
- perform basic algebraic and/or logical operations that involve levels of abstraction, including use of words, tables, graphs, mathematical equations, etc. as appropriate;
- demonstrate basic problem-solving skills, applied and analytical skills;
- express abstract calculations and mathematical concepts as well as demonstrate skills at estimating and approximating results;
- identify discipline-specific courses that include real-life application scenarios, or develop discipline-specific courses, as needed, which include real-life application scenarios.

**University Required Alert**: You must attend your classes regularly and engage in the requirements for each course; otherwise, your financial aid may be revoked either partially or in full. This would result in an amount due by you to the University immediately. Please refer to www.shepherd.edu/faoweb/ for details.