

# EEOS601 SUMMER 2011 SYLLABUS

## INTRODUCTION TO PROBABILITY AND APPLIED STATISTICS, UMASS/ONLINE

### *TABLE OF CONTENTS*

	Page:
Web Links. ....	2
Goals of the Course. ....	3
Examples of problems to be covered. ....	4
Who should take the course?. ....	7
Who should <i>not</i> take the course?. ....	7
Prerequisites & Requirements. ....	8
Math background. ....	8
Computer requirements. ....	8
Assessment. ....	8
Understanding by Design & Differentiated Instruction. ....	8
UBD Templates. ....	9
UBD & Quality Matters. ....	10
Texts on Understanding by Design and Differentiated Instruction. ....	10
Weekly online discussions. ....	10
Weekly problem sets. ....	10
Differentiated Instruction on Homework Problems. ....	10
Collaboration encouraged on assigned problems. ....	11
Problem solutions & Discussion posts due Wednesday nights at 10 pm. ....	12
Grading of problem sets. ....	12
Midterm project & paper. ....	12
Final Examination. ....	13
Grading. ....	13
Textbooks & Software. ....	14
Required. ....	14
Larsen, R. J. and M. L. Marx. 2006. ....	14
MATLAB. ....	14
Recommended books. ....	14
Larsen, R. J. and M. L. Marx. 2006b. ....	14
Nahin, P. J. 2002. ....	14
Nahin, P. J. 2008. ....	14
Matlab toolboxes & Texts. ....	15
Toolboxes. ....	15
Gilat, A. 2005. ....	16

Hanselman, D and B. Littlefield. 2004. . . . .	16
Pratap, R. 2005. . . . .	16
Other Recommended Probability & Statistics Books. . . . .	16
Bevington, P. R. and D. K. Robinson, and P. Bevington. 2002. . . . .	16
Cleveland, W. S. 1993. . . . .	16
Denny, M and S. Gaines. 2000. . . . .	16
Gonick, L. and W. Smith. 1993. . . . .	16
Hollander, M. and D. A. Wolfe. 1999. . . . .	16
Legendre, P. and L. Legendre. 1998. . . . .	16
Manly, B. F. J. 1997. . . . .	17
Mosteller, F. 1965. . . . .	17
Ramsey, F. L. and D. W. Schafer. 2002. . . . .	17
Salsburg, D. 2001. . . . .	17
Taylor, J. R. 1997. . . . .	17
A calculus text. . . . .	17
Neuhauser, C. 2004. . . . .	17
Course Logistics. . . . .	17
Time & Location. . . . .	17
Interactive & Accordent videos & Distance Learning. . . . .	17
Office Hours and E-mail. . . . .	18
Virtual Office Hours. . . . .	18
Reading Assignments. . . . .	18
Textbook. . . . .	18
Handouts. . . . .	18
Reading the handouts. . . . .	18
Hypertext links. . . . .	19
Blackboard Vista. . . . .	19
Handout Distribution on Blackboard. . . . .	19
Matlab programs. . . . .	20
Class Format. . . . .	20
Healey Library. . . . .	21
Outline of the Course. . . . .	21
Lecture Schedule. . . . .	24
Web Links. . . . .	26
Matlab Tutorials and other online Resources. . . . .	26
Mathworks' Matlab tutorial. . . . .	26
Matlab manuals/documentation. . . . .	26
The Matlab getting started page. . . . .	26
Cleve Moler, Matlab's founder, Numerical computing in Matlab. . . . .	26
Frequently asked Matlab questions from the Matlab user's group. . . . .	26
Matlab's guide to debugging Matlab programs. . . . .	26
Other Matlab tutorials online. . . . .	26
Matlab statistics programs online. . . . .	27
Casino games websites. . . . .	27
Index. . . . .	27

---

## Web Links

Gallagher's web page:

<http://alpha.es.umb.edu/faculty/edg/files/edgwebp.htm>

EEOS, the Department of Environmental, Earth & Ocean Sciences:

<http://www.es.umb.edu/index.html>

UMASS/Boston Home page:

<http://www.umb.edu/>

UMASS Online & CCDE:

<http://www.conted.umb.edu/dl/>

and Blackboard Vista

<http://www.umassonline.net>

## Goals of the Course

The course is designed to:

- ! Introduce probability theory, the foundation of statistical inference,
- ! Apply probability models to environmental problems, including calculation of diversity indices, analysis of species-abundance patterns, analysis of population size, disease frequency
- ! Cover basic inferential statistics including
  - " One sample and two-sample tests & their nonparametric analogues
    - Student's  $t$  tests (independent samples & paired) and Welch's unequal variance  $t$  test
    - Wilcoxon signed rank and rank sum tests
    - Fisher's Sign test
  - " Goodness of fit tests and tests of  $r \times c$  contingency tables
  - " Least-squares regression including with a brief overview of nonlinear models
  - " Correlation, including Pearson product-moment, Spearman's  $\rho$  and Kendall's  $\tau$ .
  - " One-way and randomized block ANOVA (and the nonparametric Kruskal-Wallis & Friedman's ANOVA)
- ! Teach basic algorithm development and data analysis with Matlab
  - " Working with matrices and creating Matlab data files
  - " Plotting data in Matlab
  - " Introduce Matlab-based Monte Carlo simulation methods for analyzing problems in probability and statistics
  - " Analysis simple integrals of probability density functions with Matlab's symbolic math toolbox
  - " Enable students to write their own Matlab programs to analyze linear models and as a foundation for multivariate statistics

I've programmed all of the text's case studies and examples in Matlab and will provide these programs that you can run as you read each section of the textbook. I'll distribute outlines of the reading material with Matlab programs that you can cut & paste and run as you read the text.

The course will focus on solving probability and elementary statistics problems with Matlab™. My approach to the course is to provide a broad overview of basic probability and statistics, with an emphasis on problem solving. If the problem can't be done in Matlab, I won't focus much attention on it. Along the way, we'll apply probability theory to a variety of interesting problems. I'll show how Monte Carlo simulations using Matlab can be used to solve problems for which analytical solutions are difficult if not impossible. The second half of the course will cover basic statistical tests, such as Student's t tests, goodness-of-fit tests, correlation & least squares regression, ANOVA and nonparametric statistics.

The textbook for the course is thorough on probability models and somewhat demanding, but we'll only be covering a subset of the book's content on probability theory. I chose the book because it does a superb job of presenting the mathematics behind each of the probability topics, introduces the history of the topic and supplements the material with a remarkable set of innovative and challenging problems. The textbook is calculus-based and provides proofs for every theorem. I will focus on the applications of the theorems not their proofs. Most of the theorems with proofs in the book correspond to Matlab programs that will be distributed in a zipped file on the course website. We need the mathematics in order to program the distributions and problems in Matlab. This book is a phenomenal reference for the mathematical underpinnings of probability and statistics.

This course will require computer programming using the Matlab computer language. Matlab is fairly easy to learn. Matlab is perhaps the best program on the market for dealing with probability problems, Monte Carlo simulations, matrices, and graphics (the other programs with comparable abilities are Mathematica and S-plus/R). I don't know of a problem in statistics or probability that can't be solved in Matlab. In my follow-up course, EEOS611 dealing primarily with ANOVA and regression, I have students solve problems in PASW (formerly called SPSS). The only reason for using PASW in that course is that PASW has a rich array of statistical programs with diagnostic statistics. PASW is far inferior to Matlab for the types of probability-based analyses we'll be doing in EEOS601, and it is virtually worthless for doing Monte Carlo simulations.

### **Examples of problems to be covered**

- ! Probability
  - " In a class of 25, how likely is it that two students will have the same birthday?
  - " Bayes theorem
    - When Monty Hall in "The Price is Right" offers to switch doors with you, should you take his offer?
    - Calculate the predictive value of a PSA test given the probability of probability of a false positive on a prostate-specific antigen assay (PSA test) and the prevalence of prostate cancer in the US
    - Given the standard error in MCAS scores (Massachusetts' high-stakes high school competency exam), what is the probability that a student is erroneously labeled a failure after one MCAS test? After three MCAS tests?

- " How many games should the 7-game world series or Stanley-Cup finals take if the teams are evenly matched? If the teams aren't evenly matched, what is the maximum likelihood estimate for the probability of the better team winning?
- " What are your odds of making your point if your first roll in craps is an 8? What are the odds of making your point of 8 'the hard way'?
- " What is the probability of
- " How many generations before all genetic diversity is lost due to genetic drift in a population of 100 Northern Right whales?
- " How many species would you expect if 100 individuals are drawn at random from the deep sea? From Boston's Inner Harbor?
- ! Estimation & hypothesis testing
  - " How frequently can you expect Richter magnitude 8 earthquakes in California?
  - " How should the margin of error be interpreted in a election poll?
  - " How do you perform an independent samples  $t$  test and how can the  $p$  values be simulated using permutation tests?
  - " When should you use a nonparametric test (e.g., Wilcoxon rank sum test) instead of a parametric test (Student's  $t$  test)?
  - " How can you test for independence and homogeneity with chi-square and Fisher's exact tests?
  - " How can Monte Carlo simulations be used to test hypotheses? How can problems be solved with the bootstrap and jackknife statistic?
- ! Regression & ANOVA
  - " How can all regression and ANOVA problems be solved with this Matlab one-line program:  $B=X\backslash Y$  ?
  - " What is "regression to the mean," how can it be simulated using dice throwing in Matlab, how does it affect decisions on whether men are better statisticians than women or whether schools have met the "No child left behind act"?

EEOS601 is designed to help students understand the basis in probability for statistical inference. This course will prepare students to understand the foundations for the complex regression and ANOVA models in advanced applied statistics courses, like **EEOS611** offered in the spring. Through analysis of probability problems, I hope to convey the probabilistic foundation of all statistical inference. I'll also introduce students to MATLAB®, a software package that can be used for a diverse array of research problems.

I've found that statistics and probability can only be learned through doing problems. I'm also convinced that an understanding of probability levels is best achieved through doing Monte Carlo simulations of probability and statistics problems. That's why I've chosen Matlab as the computer program for this course.

Page 21 provides a full outline of the course. Here are some of the highlights:

- ! Matlab programming (We'll use the latest student edition, Release 2010a, of Matlab)
  - " Entering data in Matlab
  - " Writing Matlab m.files
  - " Basic linear algebra: multiplying and adding matrices and vectors
  - " Graphing data with Matlab
- ! Probability theory
  - " Sets, sample spaces, Venn diagrams, meanings of probability
  - " Conditional probability, Bayes' theorem
  - " Permutations & combinations, combinatorial probability
- ! Random variables and expected values
- ! Distributions (all will be programmed or simulated in Matlab)
  - " Binomial, Poisson, Normal, Geometric, Hypergeometric,  $F$  &  $t$  distributions
  - " Central limit theorem
- ! Estimation & confidence intervals
- ! Hypothesis testing
  - " P values & confidence limits
  - " Type I & Type II error, Power
  - " One-sample t tests, Paired t test, two-sample t tests
  - " Sign test, Wilcoxon ranks sum and signed rank tests
  - " F test for the equality of variances
  - " Tests for the difference in proportions
  - " Monte Carlo simulations and bootstrapped confidence intervals
- ! Goodness of fit tests
  - "  $\chi^2$  tests, including log-likelihood tests and Fisher's exact test
  - " Contingency tables
- ! Regression & correlation
  - " Least squares regression in Matlab:  $B=X\backslash Y$ ;
  - " Confidence limits for the slope & intercept
  - " Multiple regression analysis with Matlab
  - " Covariation and correlation
  - " Regression to the mean
- ! ANOVA
  - " The F test
  - " Tukey & Bonferroni multiple comparisons & the studentized range
  - " Kruskal-Wallis & Friedman, Nonparametric ANOVA

## Who should take the course?

This course is designed for three groups of students:

- ! Students who need an introductory graduate-level course in probability and statistics to prepare them for their research or to prepare them for advanced graduate statistics classes.
  - " Note, that EEOS601 covers the same topics as many undergraduate statistics courses, but EEOS601 is a graduate course. Probability theory is covered in more depth than in most undergraduate courses and statistical models are covered in considerably more depth, mathematical sophistication, and detail than in an undergraduate statistics class.
  - " The text (**Larsen & Marx 2006**) is a calculus-based text. I would expect all students to have had calculus. I don't base much of the course content on calculus.
- ! Students who want a refresher course in probability and statistics to prepare them for EEOS611, or students who want to learn how Matlab can be used to analyze problems in statistics and probability.
- ! Professional Science Masters (PSM) candidates who need a single-course broad overview of probability and statistics. PSM students who've completed EEOS601 can also proceed to EEOS611 and EEOS612 (Multivariate Statistics).
- ! High school teachers who want background for teaching an AP-statistics course. The course content in EEOS601 exceeds the year-long suggested curriculum for AP high school statistics courses. However, this course covers the statistical tests in a far more rigorous fashion than a high school or introductory college class. This course uses the high-level program Matlab, but most AP stats courses use programmable calculators, such as the TI83. Because of the structure of the AP Stats curriculum, it would be difficult to directly apply the calculus-based and Matlab-based content of this course directly to an AP course. The AP curriculum was designed to match an undergraduate course, not a graduate course.

## Who should not take the course?

If you don't want to learn how to solve problems with Matlab, please don't take this course. One student dropped an earlier version of this course with the complaint that, "Dr. Gallagher expects us to learn how to program in Matlab!" I did, and I do.

If you are looking for an introductory course, without calculus, equations or programming, don't take this course. Math 125 uses **Freedman et al. (2007)** and is offered each semester and through UMASS/Online. Statistics for earth & geographic sciences (EEOS261) is an introductory statistics course which is not calculus-based and uses PASW (formerly called SPSS). It is offered every spring (but not online).

Don't take this course if you want to **immediately** begin analyzing the statistical tests produced by your computer software (e.g., Microsoft Excel, SAS, or PASW). Formal hypothesis testing isn't introduced in this course until **Week 6**.

## Prerequisites & Requirements

### Math background

There are no formal prerequisites for the course. I don't assume that you've had a prior course in either probability or statistics. However, the text uses calculus throughout, and Matlab programming requires that you become familiar with some of the basics of linear algebra. I would expect that you have had a 1- or 2-semester calculus course which provided an overview of integral calculus. I will not assign core problems requiring calculus, nor will I require calculus on any exams. I will expect you to understand the meaning of the many integral equations used throughout the text. I will assign optional **advanced** and **master problems** involving calculus for students who want to gain familiarity with advanced probability and statistical models. I treat calculus in this course in the same fashion that I handle calculus in my graduate ecology and biological oceanography courses: you can excel in the course without it, but there are portions of the course that you won't understand properly without being able to understand the calculus-based equations. The student edition of Matlab includes the symbolic math toolbox, and I'll show you that integrating probability distribution functions in Matlab is nearly as easy as calculating the square root of a number with a pocket calculator.

Every assigned problem will require analyzing problems using Matlab. Matlab is an easy-to-learn programming language, but it assumes a basic familiarity with matrices and vectors. Students do NOT need a course in linear algebra to work with Matlab, but there may be a few hours of review or background reading required to become reacquainted with basic linear algebra.

### Computer requirements

If you are taking this course, you should have ready access to a computer capable of running the Matlab student version (see below on the roughly \$100 cost). If money is a real problem, the full version of Matlab is available in the Healey Library's Graduate Computing Center which is open whenever the library is open.

## Assessment

### Understanding by Design & Differentiated Instruction

In my three graduate courses, I use a pedagogical technique described as 'differentiated instruction' integrated with 'Understanding by Design (UBD)' (**Tomlinson 2001**, **McTighe and Wiggins 2004**, and **Tomlinson & McTighe 2006**). Understanding by design has one basic rule: if a student's understanding of a topic can't be assessed, don't devote class time to that topic. Moreover, the instruction should be geared to help students understand the material.



Differentiated instruction is a technique that provides students of different abilities the opportunity to excel. It arose in K-12 education as an alternative to ‘tracking,’ in which students were slotted into accelerated, ‘normal’ and decelerated groups. Tracking has a major flaw. Following the Pygmalion principle, the gifted students become even more gifted, and the students in the lower tracks tend to fall further behind. Differentiated instruction is designed to meet the needs of mixed-ability classrooms, offering projects and assessment tailored to the strengths of students with different interests and capabilities. Students are challenged and given opportunities to work on topics that interest them. While the same general concepts are covered, they are covered in different ways depending on different styles of learning and student backgrounds.

This course draws graduate students with remarkably diverse backgrounds. Some will earn doctorates in Environmental Sciences with dissertations comparable to those of students attending the top oceanography programs, while others may do research in hydrology, aquatic chemistry, marine policy, environmental law, GIS or microbiology.

Over the last several years, in this class and in Biological Oceanographic Processes (EEOS630), I’ve seen students who have earned very low grades on formal classroom exams excel on advanced homework and term papers. The best predictor of student success in this course is performance on the homework assignments.

### UBD Templates

For each weekly session, I provide a lengthy handout. At the start of that handout are two UBD Templates. **McTighe & Wiggins (2004)** template is shown in Figure 1.

The Stage 1 template — Desired Results — provides a roadmap of the goals for the weekly session, and the Stage 2 Template — Assessment Evidence — provides a listing of methods to assess whether those goals have been met. The second template will include the weekly discussion topic and homework problems.

The Stage 1 template has 5 categories. The **Established goals** include all of those items that students are expected to have mastered at the end of the course. If there were national graduate education standards in statistics, they would be listed here. For example, at the end of every first course in probability and statistics, every student will have mastered the one- and two- sample Student’s *t* tests. In some cases, I

Stage 1 – Desired Results	
Established Goals: <span style="float: right;">G</span>	
Understandings: Students will understand that... <span style="float: right;">U</span>	Essential Questions: <span style="float: right;">EQ</span>
Students will know... <span style="float: right;">K</span>	Students will be able to... <span style="float: right;">S</span>
Stage 2 – Assessment Evidence	
Performance Tasks: <span style="float: right;">T</span>	Other Evidence: <span style="float: right;">OE</span>
Stage 3 – Learning Plan	
Learning Activities: <span style="float: right;">L</span>	

have listed the high school standards in probability and statistics from the MA Dept. of Education. **Understandings** include the big ideas for the topic, understandings desired, and misunderstandings anticipated. **Essential questions** pose provocative questions that will foster inquiry and promote learning. The final two topics cover the essential topics that students will **know** at the end of the week and the **tasks students can accomplish** at the end of the unit.

The Stage 2 template lists the discussion topic for the week and the assigned homework problems with hints for their solution.

## UBD & Quality Matters

Classes designed to meet the principles of understanding by design and differentiated instruction, will tend to be closely aligned with the 'Quality Matters' principles for the design of online instruction <http://www.qmprogram.org/>

## Texts on Understanding by Design and Differentiated Instruction

McTighe, J. and G. Wiggins. 2004. Understanding by design: professional development workbook. ASCD Alexandria VA. 295 pp. [8, 9]

Tomlinson, C. A. 2001. How to differentiate instruction in mixed-ability classrooms, 2<sup>nd</sup> Ed. ASCD Alexandria VA. 117 pp. [8]

Tomlinson, C. A and J. McTighe. 2006. Integrating differentiated instruction and understanding by design. ASCD Alexandria VA. 199 pp. [8]

## Weekly online discussions

There will be discussion topics posted for each class. In addition to the weekly discussion topic, I encourage students to post about questions that they have on the text or Matlab programming. I strongly encourage students to participate in these threads and to help out their fellow students.

## Weekly problem sets

A big part of the learning experience in this course is learning how to solve and interpret probability and statistics problems. In this course, my philosophy is that you'll master the concepts by programming the algorithms in Matlab. Nearly all of the assigned problems will involve writing Matlab programs and interpreting the results.

## Differentiated Instruction on Homework Problems

I expect that student backgrounds will range from those with very little prior background in statistics to some students with superb mathematical backgrounds. All should be able to succeed in the course.

In order to meet the needs of students with diverse backgrounds, I will pose three types of problems on each weekly topic, all involving Matlab programming. These will be:

- Core problems:** Everyone is expected to solve these problems (10 points possible for 4 problems)
- Advanced problems:** These problems cover the same material, but require a bit more programming skill and require a more thorough understanding of the subject. These may involve using more sophisticated Matlab programming to solve. (2.5 points possible for each advanced problem solved)
- Master problems:** These problems would be of the type described in **Nahin's (2000) "Duelling idiots"** or **Nahin's (2008) Digital Dice**, or **Mosteller's (1965) 50 challenging problems in probability**. In order to solve these problems, you'll need a thorough grasp of the mathematical principles, Matlab programming and, for some problems, you'll use the calculus for their solution. (3 points possible each master problem solved)

The core problems count for 10 points, and the advanced and masters problems count for an additional 10 points. Since 40% of the grade is based on homeworks, earning all points possible including the Advanced and Master problems would enable a student to attain a score of 140% in the course. A student with just 80% (B-) grade based on the core problems, midterm paper, and exams could raise their grade to a 100% by solving just half the advanced and master problems.

### Collaboration encouraged on assigned problems

One of the nice things about Matlab users is that they often collaborate in solving problems. Matlab users pose problems and get other users to help them with their solutions. There is a Matlab users group to facilitate these interactions. You can go onto the Matlab users discussion group through the Mathworks web site or through Google groups:

`comp.soft-sys.matlab`

Messages from `comp.soft-sys.matlab` are archived on Google groups. In this course, feel free to pose your questions about how to solve the problems on the discussion boards in the course web sites or in Wimba sessions. I'll try to reward students helping students through the course discussion pages.

One thing I don't want is to see pairs or larger groups of students submitting identical Matlab algorithms as homework solutions. Remember, I'll be running these programs as soon as they are submitted. If you are relying on an algorithm posted on the course website, please acknowledge that in your submission using Matlab comment statements:

```
% Algorithm based on hints posted by Jane Smith & Professor Gallagher
```

```
% Algorithms based on Wimba session on 9/20/2010 or
```

```
% Blackboard Problem 4 discussion board, 7/15/2011
```

## Problem solutions & Discussion posts due Wednesday nights at 10 pm

There will be about 10 problem sets, due Wednesday night at midnight, starting in the 2<sup>nd</sup> week. The problem solutions will usually be submitted in two parts:

- ! A brief writeup of what you did, submitted as a Word, WordPerfect or rtf file. I expect no more than a few paragraphs to describe your solution to each problem, with the total length of the problem solution not to exceed 3 pages. Matlab can produce outstanding graphics which can be inserted into your documents.
- ! The Matlab m.files that you used to solve the problem
  - " Often a Matlab m.file will call other Matlab m.files that you've written.
  - " You must make sure that you've included all of the necessary m.files in your submission
  - " If you are submitting a function that calls a function, the subsidiary functions can be submitted as part of the major function (saving me a lot of hassle in keeping m.files straight)

## Grading of problem sets

I'll grade each problem set on a 10-point scale. There is no TA for the class, so I'll be grading your assignments. I'll award extra points (up to 10 per problem set) if you have solved any advanced or master problems in addition to the core problems. Handling these challenging problems can make up for missed points on the core problems, raising your grade to the 10-point maximum. The extra points, over and above the 10-point maximum, will be kept in a bonus pool, which could add 40% to your final grade. If you've relied on a Matlab tip posted on the course website, acknowledge it in your program. If more than one student has used the tip, I'll award the original poster 1 or more points to his or her homework. I want to encourage valid collaboration while discouraging plagiarism.

At the end of the semester, I'll grade the students based on total points BEFORE taking into account the bonus problem solution points. After your grade has been determined, I'll then add your bonus points (to a maximum of 40% of total course points).

## Midterm project & paper

On 7/27/11 you will hand in a 5-15 page double-spaced paper describing the results of applying one of the probability distributions covered in the first 4 weeks of the class to a problem that interests you. On 7/6/11, you should submit a 1-page description of the problem that you are tackling. I'm pretty open on the topics and I'm willing to help. One requirement is that you must also submit a Matlab program that applies the probability distribution to real or simulated data. Here are some of the term paper topics students submitted for Fall 2010:

- An analysis of animal data after an oilspill, including a fit of the lognormal distribution to the abundance data
- Analysis of winter snowfall data in an assessment of climate change with data downloaded from the National Weather Service
- Application of Mark-Recapture sampling to estimate the population size of brown trout

- Analysis of river-flow data for three New England Rivers with fits to the lognormal and Pearson probability distribution functions in order to estimate flood-of-century statistics
- Analysis of beach profiles with an assessment of long-term erosional patterns
- Analysis of infection patterns in American hemlock trees

### Final Examination

During the last 3 days of the class, 8/23/11 (Tu)- 8/25/11 (Th), there will be a 3-h proctored final exam. You can either come to UMASS/Boston to take the exam at a time to be scheduled or make arrangements with a local college to have someone proctor an exam for you. The final exam will be open book.

We cover a large amount of material in the course. In order to help you prepare for the final exam, I will hand out a list of five final exam topics about 10 days before the exam. This list of topics will often describe probability models or statistical analyses on real data or papers from the scientific literature. During the final, I'll ask you questions about these problems.

### Grading

Table 1 shows the relative weight given to items in calculating grades.

<b>Table 1. Grading for the course</b>	
<b>AREA</b>	<b>% of grade</b>
<b>Assigned problem solutions</b>	40%
<b>Discussion &amp; Class Participation</b>	15%
<b>Midterm Paper</b>	20%
<b>Final Exam</b>	25%
<b>TOTAL</b>	<b>100%</b>

As a rough guide, I award A's for total points above 93; A- for 88 to 92; B+ for 85 to 88, B for 80 to 84. The mean homework and midterm project scores tend to be high (over 90%), but students tend to have much lower final exam scores (mean < 80%).

## Textbooks & Software

### Required

- Larsen, R. J. and M. L. Marx. 2006. An introduction to mathematical statistics and its applications, 4<sup>th</sup> edition. Prentice Hall, Upper Saddle River, NJ. 928 p. {ISBN-13: 9780131867932}[*This book sells for \$149.93 from **Prentice Hall**, \$121.76 at <http://Amazon.com>, with used copies available for \$80.*]
- MATLAB<sup>TM</sup>. Student Version. The current release is 2010a and includes the statistics toolbox. The cost is \$110. There are at least two options available.
  - Buy directly from the Mathworks
    - [http://www.mathworks.com/academia/student\\_version/](http://www.mathworks.com/academia/student_version/)
    - Most university bookstores, but not the UMASS/Boston bookstore, carry copies (the MIT COOP has copies)
  - I would strongly recommend that you had your own copy of Matlab. Much of the course will involve running problems and simulations in Matlab. It is something that you should do at the same computer that you are using for Blackboard Vista. There are copies of Matlab in the 5<sup>th</sup> floor graduate computing lab at UMASS/Boston, but there are no toolboxes available. Matlab, without toolboxes, is also available in the undergraduate teaching labs. It might be possible to complete the course using these programs. There are free versions of many of the statistical functions in the statistics toolbox, but, as the instructor, I don't have the time to rewrite programs accessing these user-contributed free versions of the statistics toolbox.

### Recommended books

- ! Larsen, R. J. and M. L. Marx. 2006b. Student Solutions manual to an introduction to mathematical statistics and its applications. Prentice Hall, Upper Saddle River, NJ. {ISBN-13: 9780131867963} [*This book is available through the UMASS bookstore and online from **Prentice Hall** for \$36.80 I would recommend this book if you want to cover more of the material in Larsen & Marx's text than we'll be covering in the course. Most of the problems I'll be assigning will have to be solved in Matlab, and this book won't help you with those solutions.*]
- ! Nahin, P. J. 2002. Duelling idiots and other probability puzzlers. Princeton University Press, Princeton N.J. {ISBN: 0-691-10286-4}[*Princeton University Press reissued this 2000 book in paperback in 2002. This book costs \$21.80 new at <http://Amazon.com> and \$22.95 at <http://pup.princeton.edu> Copies can be found for under \$10 at <http://www.campusi.com/> This book contains Matlab code that we will use to solve problems from **Larsen & Marx (2006)**. If you have the money, I'd strongly recommend buying this book. It is available now in the UMASS/Boston bookstore.*]
- ! Nahin, P. J. 2008. Digital Dice: Computational solutions to practical probability problems. Princeton, University Press, Princeton NJ {ISBN 13: 978-0-691-12698-2 } [*More Matlab solutions to challenging problems*]

## Matlab toolboxes & Texts

### Toolboxes

One of the major virtues of using Matlab is that there is a huge and growing group of Matlab users around the world, and they've written and made available thousands of free Matlab programs. I've written Matlab programs that solve most of the case studies in the textbook, and you'll write your own Matlab programs to solve problems from the book and other problems related to concepts in the text. The Mathworks now ships the statistics toolbox with the latest student release of the program. It will contain most of the routines needed for performing the statistical analyses in the latter portions of the text. There are some other toolboxes that I've found very useful in the course:

- ! Nahin's m.files to accompany "**Duelling Idiots**" These m.files show how challenging probability problems can be solved through Monte Carlo simulations.
  - " <http://www.pup.princeton.edu/titles/6914.html> &
  - " <http://press.princeton.edu/matlab/duel.zip>
- ! The Data visualization toolbox, for publication quality plots described in **Cleveland's visualizing data**, on the recommended reference list. This free toolbox includes files to produce a variety of plots including boxplots, QQ plots, Loess plots, and scattermatrix plots
  - " <http://www.datatool.com/prod01.htm>
- ! Anders Holtzberg's STIXBOX of statistical functions
  - " <http://www.maths.lth.se/matstat/stibox/>
- ! The Mathtools website, including dozens of links to free statistical toolboxes
  - " <http://www.mathtools.net/>
- ! Richard Strauss's Matlab functions, including most of the basic (and advanced) statistical tests and probability distributions
  - " <http://www.biol.ttu.edu/Strauss/Matlab/matlab.htm>
- ! Gordon Smyth's Statbox 4.2, includes all of the major probability distributions and includes a nice routine for Poisson regression.
  - " <http://www.statsci.org/matlab/statbox.html>
- ! James LeSage's econometrics toolbox, includes all of the major probability distributions as part of the toolbox.
  - " <http://www.spatial-econometrics.com/>



There are some texts that I've found worthwhile:

Gilat, A. 2005. Matlab: An introduction with applications, 2<sup>nd</sup> edition. Wiley. *[This is a very nice, concise guide to Matlab and Matlab programming. \$55.24 at Amazon, \$45 used]*

Hanselman, D and B. Littlefield. 2004. Mastering MATLAB® 7: a comprehensive tutorial and reference. Prentice-Hall, Upper Saddle River NJ. *[This is a superb supplement to the extensive Matlab manuals. It is much more detailed (and expensive) than Gilat's simple guide \$57 at Amazon.com, \$45 used.]*

Pratap, R. 2005. Getting Started with MATLAB 7: A Quick Introduction for Scientists and Engineers (The Oxford Series in Electrical and Computer Engineering). Oxford University Press. [\$26.37 new at Amazon.com]

## Other Recommended Probability & Statistics Books

Note: If I had an extra \$65 to spend on background books for this course, I'd buy the books marked in green. I'd especially recommend **Gonick & Smith's (1993)** cartoon guide to statistics for a fun overview of the basic probability and statistics concepts we'll be covering.

Bevington, P. R. and D. K. Robinson, and P. Bevington. 2002. Data reduction and error analysis for the physical sciences, 3<sup>rd</sup> Edition. McGraw-Hill, Boston. 336 pp. *[One of the main questions students in chemistry ask is how to propagate error in their calculations. This is not a topic that is dealt with in **Larsen & Marx (2006)** or any major graduate statistics text. This book is the best source on this topic that I know about. It covers the same topics as **Taylor (1997)** but goes into considerably more detail about how to propagate error with complicated functions. It includes Fortran and C++ code to perform Monte Carlo and maximum likelihood estimation. \$50 at Amazon]*

Cleveland, W. S. 1993. Visualizing data. AT&T Bell Labs, Murray Hill NJ. 360 pp. *[A wonderful book describing how to plot scientific data. \$45 at Amazon. All of the plotting techniques described in this book are available, for free, in the **dataviz toolbox for Matlab**]*

Freedman, D., R. Pisani, and R. Purves. 2007. Statistics, 4<sup>th</sup> ed. WW Norton & Col., New York. *[This is a WONDERFUL statistics book, full of tremendous examples and historical background. However, one of the book's strengths is its main weakness. It is non-mathematical, lacking discussion of the equations underlying the statistics, and it doesn't provide many numerical examples to solve]*

Denny, M and S. Gaines. 2000. Chance in biology: using probability to explore nature. Princeton University Press, 291 pp. *[The first four chapters provide a quick treatment of the probability material in **Larsen & Marx (2006)**. Chapters 5 & 6 provide some biological examples of random walks and diffusion processes. Chapter 7 deals with the probability of some extreme events and the final chapter deals with the perception of noise. \$30 at Amazon].*

Gonick, L. and W. Smith. 1993. The cartoon guide to statistics. HarperPerrenial, New York. 230 pp. *[This is a dynamite little book and a real bargain at \$12 at Amazon. Wollcott Smith is a top-notch statistician, formerly at WHOI, now at Temple.]*

Hollander, M. and D. A. Wolfe. 1999. Nonparametric Statistical Methods, 2<sup>nd</sup> edition. John Wiley & Sons, New York. 787 p. *[This book provides an authoritative and well-written description of nonparametric (distribution-free) methods. \$110 at Amazon. ]*

Legendre, P. and L. Legendre. 1998. Numerical Ecology, 2<sup>nd</sup> English Edition, Elsevier, Amsterdam. 853 pp. *[This is the text used for EEOS612, multivariate statistics. The text also provides a good summary of univariate statistics and is strong on regression and Monte Carlo simulations to test hypotheses]*



- Manly, B. F. J. 1997. Randomization, bootstrap and Monte Carlo methods in biology. Chapman & Hall, London. 399 p. *[This remarkable book shows how with a high speed computer (and a program like MATLAB™), you can generate confidence intervals and test hypotheses for a wide range of biological problems without having to assume a parametric distribution. At \$75, I'd recommend only if you are a biologist with plans to use statistics in your research]*
- Mosteller, F. 1965. Fifty challenging problems in probability with solutions. Dover, New York. 88 p. *[This is a remarkable small book describing probability puzzlers. It includes solutions to finding the odds of making a point in craps and the classic birthday problem. It is an incredible bargain at under \$10 at Amazon.]*
- Ramsey, F. L. and D. W. Schafer. 2002. The statistical sleuth: a course in methods of data analysis. Second edition. Duxbury Press, Belmont CA. 742 pp. *[This textbook is required for **EEOS611**. It presents a clear-cut exposition of advanced ANOVA and regression analyses ]*
- Salsburg, D. 2001. The lady tasting tea: how statistics revolutionized science in the twentieth century. W. H. Freeman & Co., New York. 340 pp. *[This book describes some of the personalities, such as Fisher, Wilcoxon, Tukey & Neymann that led to the statistical revolution in the 20<sup>th</sup> century. It's a bargain at under \$12]*
- Taylor, J. R. 1997. An introduction to error analysis: the study of uncertainties in physical measurements, 2<sup>nd</sup> edition. University Science Books, Sausalito CA 327 pp. *[This is a superb book on the propagation of error, an important topic for any natural scientist. \$34 at Amazon]*

## **A calculus text**

The following textbook is now being used by the UMB Mathematics dept for its 2-semester 8-credit course on differential and integral calculus for environmental scientists (Math 145 & 146). It includes a nice chapter near the end on calculus in probability and statistics.

Neuhauser, C. 2004. Calculus for biology and medicine, 2<sup>nd</sup> ed. Pearson Education, Upper Saddle River, NJ.

## **Course Logistics**

### **Time & Location**

During Summer 2011, this course will be offered exclusively through UMASS/Online. During the Fall, the course is offered in-person at UMASS/Boston.

### **Interactive & Accordent videos & Distance Learning**

Each fall in-person class was taped using the **Accordent** video system. These videos will be available as a supplement for the Summer 2011 course. Watch them if you want some additional background on the topics.

## Office Hours and E-mail

My email is [Eugene.Gallagher@umb.edu](mailto:Eugene.Gallagher@umb.edu) My office is in the Science building 1<sup>st</sup> floor, room 55. I'll often hold office hours in my lab, Wheatley 3<sup>rd</sup> floor, Rooms 41-45 (opposite biology main office & mailroom).

## Virtual Office Hours

I'll have office hours from 8-9 pm Tu & Th and 8-9 am M & W using the WIMBA interactive conferencing system within the course shell on Blackboard.

## Reading Assignments

### Textbook

The readings will be drawn almost entirely from **Larsen & Marx (2006)**. Readings from the textbook will be supplemented with pdf handouts, Matlab programs, streaming videos, and slide presentations on Blackboard Vista. I'll be posting a large number of Matlab m.files, which analyze the case studies from the textbook. I'd like you to run these m.files as you read the material from the text. Problem assignments will be based on the book and these Matlab m.files.

### Handouts

I've prepared 16 handouts for the course, including Handout 1, the syllabus, Handout 2, terms and definitions, and Handout 3, Matlab basics. Handouts 4 through 15 are the handouts describing the 12 weekly modules for the course. Each handout has nearly the same format. Each has the Understanding by Design templates for the weekly module describing what is to be covered and what is to be assessed. Then, I present my comments on the weekly module. With only a few exceptions, each weekly module will cover one chapter in the text. I often include a section on the important theorems and definitions from the chapter, followed by a somewhat detailed discussion of the key examples and case studies from the chapter. Finally, each weekly handout will have an annotated outline of the chapter. I've programmed almost all of the examples and case studies from the textbook, and I put the Matlab code into the annotated outline. I provide all of these m.files in two large zip files, so you will never need to copy these Matlab programs.

### Reading the handouts

I designed the handouts to be read electronically. You should be able to move from text item to text item by clicking on the hypertexted links (see note below). I try to provide a hypertext link for every reference in the handout and most of the terms and definitions. The terms and definitions are a continual work in progress. I try to identify all of the key terms with their definitions from the text and put them in the handout. For really important terms, there will also be an entry in my Handout 2 which has a listing of terms and concepts from mathematics and statistics.

## Hypertext links

As you read the handouts, you'll come across highlighted words, when I wrote and edited these handouts, I created hypertext links for all of the references and many of the terms. Click on the term to find the definition. To return to your original position in the pdf document from within a pdf reader, use the **alt-left arrow** combination of keystrokes, that is the pdf reader equivalent of the back arrow on a web browser. **Alt-right arrow** is the equivalent of a web browser's forward arrow.

## Blackboard Vista

Almost all of the course material will be distributed on Blackboard Vista, the UMASS/Online software. It can be accessed here:

<http://boston.umassonline.net/>

If you have any technical support questions – please direct them, as they occur, via email or live chat to the WebCT support staff at the above link.

When you register for the course, you will be provided with instructions for accessing the Blackboard Vista server. Blackboard Vista will contain handouts with the problems assigned each week, data and Matlab m.files.

In the outline section for Blackboard Vista will be the topic of each class, a handout to supplement the material in the text, and Matlab m.files for solving the case studies, and the slides to supplement the material for the class.

## Handout Distribution on Blackboard

The handouts will be available in pdf format on Blackboard Vista. These pdf documents can be viewed and printed with the free Adobe Acrobat reader. For each week's course content, I'll provide:

- ! A handout to supplement the text's material
  - " UBD Templates
  - " Outline of the chapter being covered with
    - Matlab source code
    - Matlab graphics
- ! Matlab solutions to most of the case studies and examples in **Larsen & Marx's text**.
- ! Problems to be solved.
- ! Streaming videos showing how to solve the case studies in Matlab.

## Matlab programs

I've programmed several hundred Matlab programs for the course. These are distributed in two zipped files, one for general files, EEOS601files.zip, and another specific to Larsen & Marx. Lmfiles.zip. I'd recommend the following format for unzipping the files. I created the following two directories for this course and added them to my Matlab path.

C:\Users\Eugene Gallagher\Documents\MATLAB\myfiles\EEOS601

C:\Users\Eugene Gallagher\Documents\MATLAB\myfiles\EEOS601\LarsenM

I unzipped the EEOS601 files into the 1<sup>st</sup> directory and all of the LarsenM files into the second. All of the Larsen-Marx specific m.files created for the summer course will have the following format: LMcs010201\_4th.m for the 4<sup>th</sup> edition version of Larsen & Marx's case study 1.2.1. All of the m.files for examples begin with Lmex. I've also programmed most of the figures in the text. If they aren't contained within an example or case study, they'll be denoted LMFig...

I'll be continually updating these programs as I use them. I'll post updated versions of the zip files throughout the semester. I always keep my programs backwardly compatible, so I won't change the way a program is called. I'll correct errors and make the graphs look nicer each time I work with a program.

## Class Format

There are 12 weekly sections for this class. I'll post readings for each session, a handout describing in detail what you should know for each topic, Matlab solutions to try on your home computer, and assigned problems from the book to be solved with Matlab.

I recommend the following order for moving through a sessions's topics:

- ! Go to the outline section in Blackboard Vista and read my handout on that sessions's topic
- ! Read the material in **Larsen & Marx (2006)**. I'll have this material outlined in my handout. As you encounter a case study, check my handout. I have most of the case studies programmed in Matlab. You can download the Matlab m.file to your computer and run the analysis to see how the result was obtained.
- ! Post your questions about the reading material from **Larsen & Marx** in the discussion section for the course OR bring your questions to class.
- ! **Homework:** I'll usually request that you analyze four problems in Matlab for each week. These problems will be taken from **Larsen & Marx (2006)**. The problems almost always follow the paradigm presented in the case studies. **Larsen & Marx (2006b), the student solutions manual** provides solutions to the odd numbered problems, but in this course I'll expect you to know how to write simple Matlab programs to solve the problems and the student solutions manual won't provide much help.
- ! Post questions about the Matlab m.files on the discussion section or have your questions ready for the class.

## Healey Library

The sixth floor of the library contains the bound copies of most of the top statistical journals. The current periodical room on the 5th floor contains the latest issues of each journal. The graduate computing lab on the 5<sup>th</sup> floor of the library has copies of Matlab<sup>®</sup> on the PCs.

## Outline of the Course

Section numbers in **Larsen & Marx (2006)** indicated (e.g., 2.3 The probability function)

1. Chapter 1: Introduction to the course (**Week 1** 5/31-6/6)
  - a. Readings
    - i. **Larsen & Marx (2006) Chapter 1 Introduction**
      - (1) A brief history [of probability & statistics]
      - (2) Some examples
      - (3) Chapter summary
2. Probability (**Week 2** 6/7-6/13)
  - a. Homework 1 due 6/8/11 W
  - b. Readings
    - i. **Larsen & Marx (2006) Chapter 2 Probability**
      - (1) 2.1 Introduction
      - (2) 2.2 Sample spaces and the algebra of sets, Venn diagrams
      - (3) 2.3 The Probability Function
      - (4) 2.4 Conditional probability & Bayes theorem
      - (5) 2.5 Independence
      - (6) 2.6 Combinatorics
      - (7) 2.7 Combinatorial probability
      - (8) 2.8 Taking a second look at statistics
3. Random variables (**Week 3** 6/14-6/20)
  - a. Homework 2 on chapter 2 due 6/15/11 W
  - b. Readings
    - i. **Larsen & Marx (2006) Chapter 3 Random Variables**
      - (1) 3.1 Introduction
      - (2) 3.2 Binomial and hypergeometric distributions
      - (3) 3.3 Discrete random variables
      - (4) 3.4 Continuous random variables
      - (5) 3.5 Expected value
      - (6) 3.6 The variance
4. Special distributions (**Week 4** 6/21-6/27)
  - a. Homeworks
    - i. HW3 on Chapter 3 due 6/22/11 W
  - b. Discussion postings
    - i. Post by 6/29/11 W 10 pm
  - c. Readings: **Larsen & Marx (2006) Chapter 4 Special Distributions**
    - i. Week 4
      - (1) 4.1 Introduction

- (2) 4.2 The Poisson distribution
- (3) 4.3 the Normal distribution
- (4) Skip 4.4-4.7 (p 317-342)
  - (a) Gamma distribution
  - (b) Negative binomial distribution
- 5. Estimation (**Week 5** 6/28-7/4)
  - a. Homework 4 on Chapter 4 due 6/29/11 W
  - b. Readings
    - i. **Larsen & Marx (2006) Chapter 5 Estimation**
      - (1) 5.1 Introduction
      - (2) 5.2 Estimating parameters: the method of maximum likelihood
      - (3) 5.3 Interval estimation
        - (a) Confidence interval for the mean
        - (b) Confidence intervals for the binomial parameter
        - (c) Margin of error in surveys
        - (d) Choosing sample sizes for surveys
        - (e) Estimating confidence intervals using bootstrap sampling
- 6. Hypothesis testing (**Week 6** 7/5-7/11)
  - a. Homework 5 on Chapter 5 due 7/6/11 W
  - b. Get midterm paper topics approved by 7/6/11 W
  - c. Readings from **Larsen & Marx (2006) Chapter 6 Hypothesis Testing**
    - i. 6.1 Introduction
    - ii. 6.2 The Decision Rule
    - iii. 6.3 Testing binomial data,  $H_0: p=p_0$
    - iv. 6.4 Type I and Type II error, Power
- 7. The Normal distribution (**Week 7** 7/12-7/18)
  - a. Homework 6 on Chapter 6 due 7/19/11 W
  - b. Readings from **Larsen & Marx (2006) Chapter 7 The Normal Distribution and Chapter 14.1 Nonparametric Statistics** (1-sample problems)
    - i. Introduction
    - ii. Z ratio & t ratio
    - iii. Deriving the t distribution
    - iv. Drawing inferences about  $\mu$
    - v. Drawing inferences about  $\sigma^2$
    - vi. Taking a second look at statistics ("Bad" estimators)
    - vii. Non-parametric 1-sample tests
      - (1) Sign test
      - (2) Wilcoxon one-sample signed rank test
- 8. Two Sample problems (**Week 8** 7/19-7/25 & **Week 9** 7/26-8/1)
  - a. Midterm papers due 7/27/11 W 10 pm
  - b. Homework 7 on Chapter 7 due 8/3/11 W
  - c. Discussion postings due

- d. Readings
  - i. **Larsen Marx (2006) Chapter 9 (9.1-9.2) Two-Sample Problems & Chapter 14.2 & 14.3 Nonparametric Statistics (Two-sample problems)**
    - (1) 9.1 Introduction
    - (2) 9.2 Independent sample t test
      - Welch's t test (unequal variance t test)
    - (3) 14.3 Wilcoxon rank sum test
    - (4) 13.3 The paired t test
    - (5) 14.2 The Sign test for paired data
    - (6) 14.3 Wilcoxon signed-rank test for paired data
    - (7) 9.3 The F test
    - (8) 9.4 Two-sample binomial test
    - (9) 9.5 Confidence intervals for 2-sample problems
- 9. Goodness of fit tests (**Week 10** 8/2-8/8)
  - a. Homework 8 due 8/3/11 W
  - b. Readings from **Larsen & Marx (2006) Chapter 10**
    - i. 10.1 Introduction
    - ii. 10.4 Goodness of fit tests: parameters unknown, estimating fits of data to distributions (e.g., Poisson)
    - iii. 10.5 Contingency tables
      - (1) Chi-square & Fisher's exact tests
      - (2) Odds ratios
- 10. Regression (**Week 11** 8/9-8/15)
  - a. Homework 9 due 8/10/11 W
  - b. Readings
    - i. Larsen & Marx (2006) Chapter 11
      - (1) 11.1 Introduction
      - (2) 11.2 The method of least squares,  $B=X\backslash Y$
      - (3) 11.3 The linear model, standard errors for regression coefficients
      - (4) 11.4 Covariance & correlation
    - ii. Regression to the mean & dice throwing (from **Campbell & Kenny 1999**)
- 11. ANOVA (**Week 12** 8/16-8/22)
  - a. Homework 9 due 8/10/11 W
  - b. Readings
    - i. **Larsen & Marx (2006) Chapter 12 ANOVA**
    - ii. Chapter 13 Randomized Block Designs
      - (1) F test for Randomized Blocks
      - (2) Paired vs. Independent Samples t test
  - c. Note, **Larsen & Marx (2006) Chapter 14** on nonparametric statistics will be covered in parts when the analogous parametric tests are introduced.
- 12. Final Exam, at UMB or proctored off-campus

## Lecture Schedule

<b>Table 1 Summer EEOS601 Lecture Schedule</b>				
Week	Date	Topic	Readings	Problems
1	Tu 5/31/11- M 6/6/11	Overview of the Course; Introduction to Probability, Statistics & Matlab	Introduction to Matlab and Larsen & Marx Chapter 1 & Handouts 2-4	
2	Tu 6/7/11- M 6/13/11	Probability	Larsen & Marx Chapter 2 & Handout 5	HW1 on Chapter 1 due 6/8/11 W 10 PM
3	Tu 6/14/11- M 6/20/11	Random Variables	Larsen & Marx Chapter 3.1-3.6 & Handout 6	HW2 on Chapter 2 due 6/15/11 W 10 PM
4	Tu 6/21/11- M 6/27/11	Special Distributions	Larsen & Marx Chapter 4.1-4.3 p 274-317 & Handout 7	HW3 on Chapter 3 due 6/22/11 W 10 PM
5	Tu 6/28/11- M 7/4/11	Estimation	Larsen & Marx Chapter 5 & Handout 8	HW4 on Chapter 4 due 6/29/11 W 10 PM
6	Tu 7/5/11- M 7/11/11	Hypothesis Testing	Larsen & Marx Chapter 6 & Handout 9	HW5 on Chapter 5 due 7/6/11 W 10 PM <b>7/6/11 Get Topic for Midterm Paper Approved</b>



<b>Table 1 Summer EEOS601 Lecture Schedule</b>				
Week	Date	Topic	Readings	Problems
7	Tu 7/12/11- M 7/18/11	The Normal Distribution & One-sample Nonparametric Tests	Larsen & Marx Chapter 7 & Handout 10	HW6 on Chapter 6 due 7/13/11 W 10 PM
8	Tu 7/19/11- M 7/25/11	Two-Sample Problems (1 of 2) Independent $t$ tests, Rank Sum test, 2-sample binomial, F test	Larsen & Marx Chapter 9 & Chapter 14 (part) & Handout 11	HW7 on Chapter 7 due 7/20/11 W 10 PM
9	Tu 7/26/11- M 8/1/11	Paired two-Sample Problems (2 of 2) Paired $t$ tests Sign Test for paired data Signed Rank Test for Paired data	Larsen & Marx Chapter 13 (part), Chapter 14 (part) Handout 12	<b>Midterm Term papers due 7/27/11 W 10 PM</b>
10	Tu 8/2/11- M 8/8/11	Goodness-of-Fit Tests	Larsen & Marx Chapter 10 & Handout 13	HW8 due on Chapter 9 8/3/11 W 10 PM
11	Tu 8/9/11- M 8/15/11	Regression & Correlation	Larsen & Marx Chapter 11 & Handout 14	HW9 on Chapter 10 due 8/10/11 W 10 PM
12	Tu 8/16/11- M 8/22/11	ANOVA & Randomized Block Designs	Larsen & Marx Chapters 12 & 13; Chapter 14 (part) & Handout 15	HW10 on Chapter 11 due 8/10/11 W 10 PM
Th 8/25/11 Th		Last Day of Summer Session		
8/23-25/11		In-person or proctored Final Exam		

## Web Links

See **links** below for free Matlab statistical toolboxes.

### Matlab Tutorials and other online Resources

#### Mathworks' Matlab tutorial

[http://www.mathworks.com/academia/student\\_center/tutorials/launchpad.html](http://www.mathworks.com/academia/student_center/tutorials/launchpad.html)

#### Matlab manuals/documentation

[http://www.mathworks.com/access/helpdesk/help/techdoc/index.html?/access/helpdesk/help/techdoc/matlab\\_product\\_page2.html&http://www.mathworks.com/access/helpdesk/help/techdoc/matlab\\_product\\_page.html](http://www.mathworks.com/access/helpdesk/help/techdoc/index.html?/access/helpdesk/help/techdoc/matlab_product_page2.html&http://www.mathworks.com/access/helpdesk/help/techdoc/matlab_product_page.html)

#### The Matlab getting started page

[http://www.mathworks.com/access/helpdesk/help/techdoc/index.html?/access/helpdesk/help/techdoc/learn\\_matlab/bqr\\_2pl.html](http://www.mathworks.com/access/helpdesk/help/techdoc/index.html?/access/helpdesk/help/techdoc/learn_matlab/bqr_2pl.html)

#### Cleve Moler, Matlab's founder, Numerical computing in Matlab

<http://www.mathworks.com/moler/>

#### Frequently asked Matlab questions from the Matlab user's group

<http://www.mit.edu/~pwb/cssm/matlab-faq.html>

#### Matlab's guide to debugging Matlab programs

<http://www.mathworks.com/support/tech-notes/1200/1207.shtml>

### Other Matlab tutorials online

13. **Matlab tutorial from the Mathworks** (also available with the student version. This tutorial links to much of the functionality in Matlab. You can cut & paste the code to produce some stunning graphics in a few minutes to see the graphic power of Matlab).
14. **Matlab tutorial from U. Michigan Carnegie/Mellon** (This is a very basic tutorial that will get you started in a hurry).
15. **Matlab tutorials, videos & links from the U of Cambridge Dept. of Engineering**
16. **Matlab primer by Kermit Sigmon (Written for Matlab 3.5 and 4)**

## Matlab statistics programs online

Compilation from Mathtools (including commercial & free programs):

<http://www.mathtools.net/MATLAB/Statistics/index.html>

See **links** below for free Matlab statistical toolboxes.

Rajeev Raizada's wonderful Matlab simulations of the normal distribution, one-sample t tests, two-sample t tests, and the correlation coefficient:

<http://www.dartmouth.edu/~raj/intro-stats.html>

## Casino games websites

Craps table with interactive guide to the rules

[http://www.learning-craps.com/table\\_layout.htm#top](http://www.learning-craps.com/table_layout.htm#top)

Loaded dice

<http://www.gamblingcollectibles.com/equipment.html>

## Index

Accordent.....	17
ANOVA.....	3-6, 17, 23, 25
Bayes theorem.....	4, 21
Binomial distribution.....	22
Bonferroni multiple comparison procedure.....	6
Bootstrap.....	5, 17, 22
Central limit theorem.....	6
combinations.....	6
combinatorics.....	21
Conditional probability.....	6, 21
Confidence interval.....	22
correlation.....	3, 4, 6, 23, 25, 27
covariance.....	23
Distributions	
gamma.....	22
lognormal.....	12
negative binomial.....	22
normal.....	22, 25, 27
Poisson.....	22
Expected value.....	21
Fisher.....	17
Fisher's exact test.....	6
Goodness of fit.....	3, 4, 6, 23, 25
independence.....	5, 21
Jackknife.....	5

Kruskal-Wallis. . . . .	3, 6
least squares. . . . .	4, 6, 23
Legendre. . . . .	16
likelihood. . . . .	5, 6, 16, 22
Linear model. . . . .	23
Margin of error. . . . .	5, 22
Matlab. . . . .	3-8, 10-12, 14-21, 24, 26, 27
Symbolic Math Toolbox. . . . .	3, 8
Maximum likelihood. . . . .	5, 16, 22
Maximum likelihood estimation. . . . .	16
Method of least squares. . . . .	23
Multiple regression. . . . .	6
multivariate. . . . .	3, 7, 16
nonparametric. . . . .	3-6, 16, 22, 23, 25
Odds. . . . .	5, 17, 23
Office Hours. . . . .	18
paired data. . . . .	23, 25
Parameter. . . . .	22
Pearson. . . . .	3, 13, 17
Poisson. . . . .	6, 15, 22, 23
population. . . . .	3, 5, 12
Power. . . . .	6, 22, 26
Probability. . . . .	1, 3-17, 21, 24
Probability function. . . . .	21
Propagation of error. . . . .	17
Regression. . . . .	3-6, 15-17, 23, 25
Regression to the mean. . . . .	6, 23
sample. . . . .	3, 6, 9, 21-23, 25, 27
Sign test. . . . .	3, 6, 22, 23, 25
Standard error. . . . .	4
Statistic. . . . .	1, 3-5, 7-10, 13-18, 21-24, 26, 27
Student's t distribution. . . . .	22
Studentized range. . . . .	6
Student's t. . . . .	3-5, 9
Type II error. . . . .	6, 22
univariate. . . . .	16
variance. . . . .	3, 21, 23
Virtual Office Hours. . . . .	18
Wilcoxon rank sum test. . . . .	5, 23, 25
Wilcoxon signed rank test. . . . .	22, 25
WIMBA. . . . .	11, 18