ENGR 2600: Modeling and Analysis of Uncertainty
Section 4: Tuesday/Friday 12:00-1:20 PM in DCC 324

Course Objectives: 1) convey an appreciation for uncertainty in the application of engineering models. 2) introduce methods of probability and statistics that engineers should know.

Instructor: Charles J. Malmborg, Professor of Decision Sciences and Engineering Systems
Office: CII 5015, Telephone: 518-276-2935, E-mail: **malmbc@rpi.edu**
Office Hours: Tuesday 3:30-4:30 PM and Friday: 3:30-4:30 PM

TA: William Lalonde, Office: CII 5208, Telephone: 518-276-2957, E-mail: lalonw@rpi.edu
Office Hours: Monday 10:00-11:30 AM and 12:00-1:30 PM

Text: DeVore J.L., *Probability and Statistics for Engineering and the Sciences*, Thomson, 7th edition, 2008. **Software:** The student version 14 of MINITAB that comes bundled with the text will be used.

Learning Outcomes: It is anticipated that this course will contribute to varying degrees to the following learning outcomes: 1.) an ability to apply knowledge of mathematics, science and engineering, 2.) an ability to design and conduct experiments as well as to analyze and interpret data, 3.) an ability to design a system, component or process to meet stated performance constraints, 4.) an ability to identify, formulate and solve engineering problems, and 5.) an ability to use the techniques, skills and modern engineering tools necessary for engineering practice.

Attendance Policy: While class attendance is not mandatory, it is critically important to performing well in this course. All exams and homework assignments are discussed in detail during class. Class attendance significantly improves learning effectiveness and performance on graded course deliverables.

Course Web Page: This section of the course will use a WebCT site for distribution of course materials. Once you are registered for this section, you can access the site using your RCS username and password.

Homework Problems must be submitted by the end of class on the due date. Many homework problems will be done during regular class meetings. Students are allowed and encouraged to work together on homework problems but every student must submit their own solutions for each problem set.
Assignment 1 – due March 2, 2010
Chapter 1: 19, 21, 23, 33, 81
Chapter 2: 21, 26, 34, 105, 110
Chapter 3: 13, 23, 30, 36, 49, 69, 75, 87
Chapter 4: 5, 13, 40, 41, 47, 53, 61
Chapter 5: 1, 3, 22, 37, 49, 51
Assignment 2 – due April 8, 2010
Chapter 7: 3, 15, 37, 43, 45
Chapter 8: 5, 11, 16, 23, 35, 37, 47, 48, 53
Chapter 9: 7, 19, 29, 38, 41, 47, 51, 57, 62
Chapter 10: 7, 9, 41
Chapter 11: 5, 8, 10, 17
Assignment 3 – due May 11, 2010
Chapter 12: 35, 37, 59
Chapter 13: 37, 41
Chapter 14: 3, 7, 17, 31
Chapter 15: 3, 4, 23
Chapter 16: 5, 6, 9, 15, 20, 23

Course Grades: Final grades in the lecture section of *ENGR-2600* will be assigned in accordance with a "rank ordering of cumulative scores" representing performance on exams and the homework problems. Your performance on any individual deliverable is evaluated relative to overall class performance. For example, suppose that your score on an exam representing 35% (a=0.35) of the final course grade is 68 (x=68), and the class average on this exam is 63 (μ=63) with a standard deviation of 7 (σ=7). To determine the contribution of this exam to your final cumulative score, (which is used to determine the rank ordering of cumulative scores), you would use:

$a(x-\mu)/\sigma = 0.35(68-63)/7 = +0.2500$

At the end of the semester, these weighted "Z-scores" from your exams and the homework assignments are totaled to determine your cumulative score. Grades are then assigned in accordance with the rank ordering of the cumulative weighted Z-scores. This technique eliminates scaling inconsistencies between individual deliverables and yields a rank ordering reflecting the intended grade distribution. Given a rank ordering, it's the responsibility of the instructor to define intervals on the rank ordering associated with letter grades. To appeal a grade on any course deliverable, you must do so in writing within one week of receiving that grade. The procedure is to write a brief note describing what you believe to be the error and then submit it with the graded exam (or homework) to the instructor within the one-week deadline.

<u>Grades for the course will be determined as follows:</u>

Exam 1-March 2, 2010	35%*
Exam 2-April 8, 2010	35%*
Exam 3-Finals Week	35%*
Homework	15%

*The exam with the least favorable impact on your grade counts for only 15%.

Midterm Grading Assessment: After each exam, the class average and standard deviation will be announced in class. As described in the "Course Grades" section above, this information will enable students to approximate their class standing and current course grade after each exam.

Statement on Academic Integrity: Student-teacher relationships are built on trust. Acts that violate this trust undermine the educational process. Students should familiarize themselves with the portions of the <u>Rensselaer Handbook</u> dealing with academic integrity and should note that penalties for plagiarism and other forms of academic dishonesty can be quite harsh. In this course, students are required to work individually on exams but may discuss the homework assignments. However, each student must submit their own individual homework solutions. All exams will be given in class. Students found to be cheating on a course deliverable such as an exam will receive a grade of zero for that deliverable. Any student found to be cheating more than one time will receive a grade of F for the course.

Exams: There will be two in-semester exams and a final exam. All will be open-book/open-notes. The final exam will not be cumulative.

Late Submission Policy: All homework assignments are due by the end of class on the specified due date. A 10% reduction in score may be imposed if assignments are turned in after the end of a class. An additional 10% reduction may be imposed for every additional day late following the due date.

Exam/Homework Policy for Snow Days or Class Cancelation for Other Reasons:

If there is a snow day or class cancelation for other reasons on the class prior to an exam – that exam and corresponding homework due date will be postponed until the class following the originally scheduled exam date. If there is a snow day or class cancelation for other reasons on a scheduled exam day, that exam will be given and the corresponding assignment will be due during the next class.

Reading Assignments (Topic coverage may vary - be prepared to adjust reading ahead of class periods).

Lecture 1:	Lecture 1. Definitions: population, variable, census, random sample, sampling error, parameter, statistic, biased sample, data types, probability, frequency distribution. Graphical representations of data: bar chart, histogram, pie chart, dot plot, stem and leaf diagram, shape, variability, skew, bimodal data. Numerical descriptors of data: mean median, mode.	Pages 1-31
Lecture 2:	Numerical descriptors of data: range, standard deviation, empirical rule, z-score, percentiles, percentile rank, trimmed mean, quartiles. Box plots. Problems 1-56, 1-77, 1-81. Definitions: probability, experiment, sample space, event, complement, union, intersection, mutually exclusive, contingency table, random variable, combinations, permutation. Problem 2-34.	Pages 31-67
Lecture 3:	Problem 2-36. Conditional probability, Bayes rule, multiplication rule, independence. Problems 2-105, 2-108, 2-109. Random variable, Bernoulli random variable, discrete random variable, continuous random variable, probability distribution, cumulative distribution function. Problems 3-11, 3-12. Expected value, expectation.	Pages 67-108
Lecture 4:	Expectation, estimator, biased estimator. Problems 3-29, 3-30, 3-25. Binomial distribution. Problems 3-49, 3-60. Hypergeometric distribution.	Pages 108-121
Lecture 5:	Hypergeometric distribution, Problem 3-69, negative binomial distribution, Problem 3-76, Poisson distribution, Problems 3-86, 3-87. Some additional discrete random variables, geometric distribution, continuous random variables, density functions, uniform distribution, Problem 4-5, normal distribution, empirical rule, standardization, approximations. Problem 4-32.	Pages 121-157
Lecture 6:	Problems 4-32, 4-47, 4-58. Exponential distribution, Problem 4-59. Probability modeling application.	Notes, Pages 157-162
Lecture 7:	Probability modeling application. Joint random variables, marginal probability mass function, Problem 5-1. Covariance, correlation coefficient, Problems 5-22, 5-28, 5-33. Sampling distributions, Problem 5-41.	Pages 184-212
Lecture 8:	Central limit theorem Problems 5-46, 5-47. Normal approximation to the Poisson, Problems 5-49, 5-55. Independence rule, Problem 5-64. Monte Carlo simulation, random number generation, applications in error propagation.	Pages 213-224, Notes
Lecture 9:	Monte Carlo simulation homework example. Point estimation, standard error, interval estimation. Problems 7-3, 7-6, 7-7.	Notes, Pages 254-270
Exam 1	Exam will be open book and open notes. You may use a calculator but not a computer.	
Lecture 10:	Problem 7-11. Large sample CI's, Problem 7-12, CI's on proportions, Problems 7-15, 7-23, 7-25. Student's t-distribution, small sample CI's, prediction and tolerance intervals, Problem 7-30, 7-35, 7-38.	Pages 270-278
Lecture 11:	Chi-square distribution, CI on variances, Problem 7-42, 7-43, 7-45. Single population tests of hypothesis, hypothesis formulation, rejection region, test statistic, test procedure, p-value, type I and type II errors, two-tail and one-tail tests, tests on means, sample size determination, computing type II errors. Problems 8-2.	Pages 278-294

Lecture 12:	Problem 8-11. Computing type II errors, large/small sample tests, sample size determination, single population hypothesis testing case examples, application of the p-value approach, Problems 8-15, 8-16.	Pages 294-306
Lecture 13:	Problems 8-19, 8-19, 8-30, 8-31, 8-53, 8-57.	Page 306-321
Lecture 14:	Two population tests, tests on differences in means, variance known vs. unknown, sample size determination, large sample cases. Problems 9-1, 9-5, 9-7. Small sample cases, standard error, pooled variance. Problems 9-19, 9-29. Paired observations, Problem 9-37. Paired t-test versus direct tests on means, two-population tests on proportions.	Pages 325-364
Lecture 15:	Problem 9-47, calculation of type II errors, two-population tests on variances, Problems 9-57, 9-61, sequential testing application. Single factor ANOVA, one-way randomized designs, definitions, ANOVA assumptions, sums of squares, ANOVA procedure, ANOVA table.	Pages 369-379
Lecture 16:	Problems 10-3, 10-6, multiple comparisons procedure, Problem 10-11. ANOVA example testing assumptions.	Pages 379-394
Lecture 17:	Unequal sample sizes, missing observations, two-factor ANOVA, randomized block designs, efficiency of blocking, Problem 11-3, example using the p-value approach. Two-factor analysis with replication, sums of squares calculations, ANOVA procedure for a sample problem, estimating interaction effects without replication, Latin square designs, Latin square example.	Pages 397-410-419 and 424-429
Exam 2	Exam (open book and notes, may use calculator but not computer).	
Lecture 18:	Simple Linear Regression, least squares method, estimating model parameters, tests of model adequacy, t-test on the slope, ANOVA approach, coefficient of multiple determination.	Pages 446-477
Lecture 19:	Sample problem, confidence intervals, prediction intervals, sample problems, regression forecasting example, Problem 12-35.Correlation coefficient, tests on the correlation coefficient, Problem 12-59.	Pages 477-494
Lecture 20:	Multiple linear regression, standardized residuals, tests of model adequacy, sample problems. Problems 13-37, 13-41. Multiple linear regression model building, adjusted R^2, residual analysis, multicolinearity, sample problems. Goodness of fit testing for continuous and discrete distributions. Problems 14-1, 14-3.	Pages 528-562 and 568-576
Lecture 21:	Problems , 14-9, 14-17. Tests of homogeneity, tests of independence. Problems 14-29, 14-31. Distribution-free tests on means, Wilcoxon signed rank test, paired observations, large sample case, Problem 15-1.	Pages 576-595 and 599-613
Lecture 22:	Problems 15-4, 15-7. Wilcoxon rank sum test, Problems 15-11, 15-13. Distribution-free ANOVA, Kruskal-Wallis procedure, Problem 15-23, dist.-free ANOVA with blocking, Friedman's test, Problem 15-27.	Pages 618-624
Lecture 23:	Quality Control Methods for Variables. Xbar, R and s charts and their applications. Problems 16-9, 16-14, 16-15.	Pages 625-640
Lecture 24:	Control charts for attributes, p-charts, np-charts, c-charts, u-charts, Problems 16-19, 16-21, 16-23, 16-25.	Pages 641-654
Lecture 25:	Acceptance sampling, basic definitions, OC curves, single, double, multiple sampling plans, AOQL, ATI, ASN, lot by lot sampling, Problems 16-32, 16-35, 16-38.	Pages 654-661
Lecture 26:	Topics Review	

ENGR-2600 Modeling and Analysis of Uncertainty – Fall 2009 Exam 1

Instructions: This is a 75 minute exam. It is open book and open notes and you may use a non-programmable calculator. Please read all questions carefully. Good luck.

1.) (25 points) An analysis of historical data on the number of night emergency calls on a doctor's pager shows that it varies between 0 and 2 according to the following probability distribution:

Number of calls:	0	1	2
Probability:	.2	.4	.4

According to the hospital contract, the doctor must wear the pager three nights per week. Compute the standard deviation of the number of night emergency calls per week for this doctor.

2.) (25 points) The number of bridal flower arrangements demanded over the past 20 weddings served by a florist is summarized below:

Wedding:	1	2	3	4	5	6	7	8	9	10
Demand:	4	5	6	6	4	7	5	4	4	5
Wedding:	11	12	13	14	15	16	17	18	19	20
Demand:	5	7	4	4	4	6	4	5	5	5

When contracting with the florist, wedding parties are generally unwilling to commit to a fixed number of flower arrangements even though the florist must pre-assemble them in advance. The profit from selling a pre-assembled floral arrangement is $75. If the florist does not pre-assemble a sufficient number of arrangements, the demand for the excess is met by a competitor. Any of the pre-assembled but unsold arrangements can be reused but they must be disassembled at a cost of $20. How many floral arrangements should be pre-assembled in advance of a wedding to maximize expected profit?

3.) (25 points) The 22nd congressional district has 65% Democrats and 35% Republicans. It is comprised of four counties *with this same distribution of registered voters* including Saratoga, Rensselaer, Albany and Columbia which respectively represent 26%, 17%, 45% and 12% of the 22nd congressional district population. Assume that the large number of registered voters in the district justifies probability calculations that assume sampling with replacement. Approximate the probability that a random sample of three voters consists of three Republicans, all from different counties.

4.) (25 points) The battery life of a disposable penlight follows an exponential distribution with a mean value of 3.6 hours. The manufacturer guarantees that fewer than 40% of penlights will experience battery failure before three hours of use. A consumer protection agency is planning to take a sample of 50 penlights and test them to see if they meet the manufacturer's warranty claim.

a.) (15 points) Compute the probability that the sample of 50 will meet the manufacturer's guarantee.

b.) (10 points) What would be the probability of meeting the manufacturer's guarantee if battery life were normally distributed with a mean of 3.6 hours and a variance of 9 hours?

ENGR-2600 Modeling and Analysis of Uncertainty – Fall 2009 Exam 2

Instructions: This is a 75 minute exam. It is open book and open notes and you may use a non-programmable calculator. Good luck.

1.) (25 points) A bank manager is investigating the utilization of an ATM located in the lobby of a branch office. Over 220 daily observations made at random times, the machine was observed to be idle 61 times.

a.) (15 points) Determine the number of observations that would be needed to estimate the utilization of the ATM with a precision of +/-3% with a confidence level of 98%.
b.) (10 points) If the manager wanted to generate an estimate with a precision of +/-5% using just the observations taken so far, what's the corresponding confidence level?

2.) (25 points) A quality systems analyst at the Educational Testing Service (ETS) suspects fraud in GRE test grading at one particular test site in a foreign country. Average GRE verbal test scores reported by this site over the past 7 on-site exam offerings and the number of students taking the exam are summarized below:

Exam session:	1	2	3	4	5	6	7
Number of Students:	12	8	14	17	8	6	15

Over the observed period, the average exam score was 610 from the suspected test site. Other exam sites in the same country serving the same population of students have consistently reported average GRE verbal test scores of 587 with a standard deviation 121.

a.) (10 points) Using $\alpha=0.10$, formulate and test the appropriate hypothesis to prove that the suspected sites test scores are higher than normal. Draw the appropriate conclusions.
b.) (5 points) Compute the p-value for the test results.
c.) (10 points) What would be the probability of incorrectly failing to reject the null hypothesis from the test in part a.) if the true mean test score was 598 instead of 587?

3.) (25 points) An engineer is attempting to calibrate a batch mixer at a drum filling workstation where the goal is to fill drums with exactly 44 gallons of an emulsion. Nine filled drums were carefully measured with the following results (the sample standard deviation is about 0.48):

Drum:	1	2	3	4	5	6	7	8	9
Volume (gallons):	44.6	44.1	43.2	44.0	44.4	44.6	43.7	43.4	44.0

a.) (10 points) Assuming that fill volumes follow a normal distribution, compute a 95% confidence interval on the average filling volume.
b.) (10 points) Estimate a 99% upper limit on the fill volume of the next drum.
c.) (5 points) Over what range can you be 99% confident that 90% of individual drum volumes will fall within it?

4.) (25 points) Two Chevrolet dealers are offering the same exact models at different prices. The data (in $1000's) is summarized below:

Model:	1	2	3	4	5	6
Dealer 1:	12.9	13.5	16.5	17.1	20.0	20.4
Dealer 2:	15.5	17.1	19.3	20.0	21.7	24.0

What percentage reduction in the test p-value is achievable using a paired procedure versus a direct test on the difference in mean selling prices between the two dealers? (Hint: The standard deviation for both dealers' data is equal to 3.1.)

ENGR-2600 Modeling and Analysis of Uncertainty – Fall 2009 Exam 3

Instructions: This is a 75 minute exam. It is open book and open notes and you may use a non-programmable calculator. Unless specified otherwise, assume that α=.05. Good luck.

1.) (20 points) The dean of students wishes to determine if there is a relationship between class year and gender in the usage of the campus counseling center. The number of counseling center visits reported by these different groups over the past several years is summarized below:

	1st Yr	2nd Yr	3rd Yr	4th Yr
Male	27	63	29	35
Female	41	55	17	44

a.) (15 points) Does the data suggest that gender is independent of class year?
b.) (5 points) Estimate the p-value for the corresponding test of independence.

2.) (40 points) The School of Engineering Curriculum Committee has questioned whether different majors have different grades in ENGR-2600. While it has been understood for a long time that the class year (1st, 2nd, 3rd, 4th) makes a difference, the committee wishes to specifically investigate the impact of the major. The following historical average GPA data on five engineering majors is shown below. The total sum of squares is 1.588.

	AE	ChE	CE	EE	ME
1st Year:	2.8	3.0	2.6	2.6	3.2
2nd Year:	3.3	3.0	2.8	2.9	2.8
3rd Year:	3.3	3.4	3.2	3.2	3.3
4th Year:	3.6	3.2	2.7	2.9	3.4

a.) (20 points) Perform the appropriate parametric test to determine if there is a difference among the majors and compute the efficiency of blocking for the test procedure.
c.) (20 points) Demonstrate that your test results from above are either consistent or inconsistent with analogous results obtainable using nonparametric statistical methods.

3.) (20 points) An audiologist theorizes that the correlation between hearing loss and age is over 95%. Average hearing loss data at ten year intervals is shown below:

Age:	30	40	50	60	70	80	90
Hearing Loss:	6%	6%	8%	14%	20%	21%	21%

Use an appropriate test on correlation to investigate the theory.

4.) (20 points) A traveler is trying to decide which major frequent flyer program to commit to. She wants to select a major airline with low fares but is not sure whether there is really any difference between airlines in their fares. The major airlines serving her hometown airport are: American (AA), Continental (CA), Northwest (NA), Southwest (SA), and USAir (UA). Although it's difficult to sort through all the differences introduced by booking agents and destinations, she was able to collect the following data on fares for her most frequent destinations:

	Austin	Denver	Detroit	Seattle	Atlanta
Expedia:	AA-390	CA-250	NA-150	SA-210	UA-120
Hotwire:	CA-300	NA-400	SA-220	UA-450	AA-280
Orbitz:	NA-510	SA-310	UA-280	AA-380	CA-320
Travelocity:	SA-230	UA-510	AA-300	CA-380	NA-300
Priceline:	UA-380	AA-360	CA-180	NA-470	SA-210

The total sum of squares is 275,016. Apply the appropriate procedure to determine whether there is a difference in airline fares using this data.