Advanced Statistical Methods for Engineering Research

EMEN 5610 / APPM 7400a

Course Syllabus & Tentative Content Outline

Dr. Jeffrey T. Luftig
W. Edwards Deming Professor of Management

1. **Course Description**

Combines intermediate and advanced statistical methods with practical research applications and computer software. Develops commonly used statistical models such as Two and Three-Way Analysis of Variance as well as Multiple Linear Regression for the solution of common business and industrial research problems. The statistical models are implemented and interpreted in the context of actual data sets using available statistical software. Prerequisites: EMEN 5005 and EMEN 5900; OR APPM 4570/5570 and APPM 4580/5580; or the equivalent as assessed and solely approved by the instructor.

2. **Course Goals and Objectives**

As a result of successfully completing this course, the student will be capable of performing a vast array of activities in business and industry utilizing the principles and tools of applied research, statistical methods, and experimental design techniques. Specifically, when faced with a *statement of a problem* in either a production or administrative setting, in the context of a product or a service, the student will be capable of:

(a) properly classifying the problem/issue as a candidate for the *Problem-Solving* (defect elimination) *Strategy* versus the *Quality Improvement* (variability reduction) *Strategy*;

(b) describing an appropriate (most efficient in balancing power and cost) *research design* to respond to the statement of the problem;

(c) developing a high-power, cost-effective *experimental design* to test the required hypotheses, which will guard against threats to External and Internal Validity;

(d) designing the appropriate *sampling plan* in relation to the experiment, including the calculation of a defensible sample size, and selecting replications, or repetitions, or both in the conduct of the design;

(e) explaining the methods which must be utilized to *conduct and control the execution of the study* in the business and industrial environment;

and subsequently, incorporating the major thrust of this course,

(f) designing an appropriate *plan for the statistical analysis of the data*, taking internal and external customer issues, as well as cost and profitability, into account; then

(e) *executing the analytical plan* and correctly answering the research questions posed; and finally

(g) *preparing and presenting the results of the study* in a professionally acceptable format for decision-making purposes at the highest level of an organization.

The analyses taught by the instructor, and utilized by the student, in this course will require the assessment of criterion measures evaluated on nominal, ordinal, and continuous data scales (in some cases, all within a single research study). Further, the underlying assumptions related to each statistical test and its interpretation will be thoroughly reviewed.

To properly learn these techniques, a series of increasingly complex research studies drawn from contemporary business and industrial applications will be employed. These examples have been extracted from actual studies conducted for firms such as Anheuser-Busch; the Earthgrains company; ALCOA, and ALCOA of Australia; Molex, and Molex-Singapore; the Applied Magnetics Corporation; the Inland Steel company; Ford Motor Company; Motorola - Semi-Conductor Division; and the Oregon Saw Chain company (Omark Industries).

For all of the statistical analyses presented and conducted, the student will make use of an extensive array of computer software. The primary software programs utilized in this course will be SPSSPc for Windows (available on a 6 month rental basis at: http://estore.e-academy.com/) and MVPStats (which all students should already own from the prerequisite courses). Both of these programs are also resident and available for use by students in the computer labs on the second floor of the Engineering Building.

3. **Rationale**

The application and use of applied statistical methods in business and industry today does not take place within a vacuum. Applied research in contemporary business organizations plays a vital role in the enhancement of profitability through focused cost reduction, quality improvement, problem-solving, and increasing asset utilization efforts.

Within this context, statistics, and the statistical analysis of data, is primarily a tool; not an objective. In this regard, applied research and statistics should be learned in the same way; that is, as a tool employed within a larger strategy for the purposes that will be found in the business environment. The student will find that all of the following research questions:

1. What is the quality of the telephone response rate among the inside sales group, and is there a day-of-the-week or time-of-day effect(s)?

2. What temperature setting, feed rate, and atmosphere condition should be used for our new heat treat furnace to maximize quality, and yet minimize cost?

3. What is the optimum shelf stacking/display design to use in Kroger's in order to maximize sales and minimize stale for our baked goods?

4. What is the superior inside spray formulation to use if we wish to control off-taste in beer?

5. Which of the 67 possible variables affecting plugging on the caster must be controlled, and at what targets, in order to realize an additional profit of $33,000,000.00 per year?

may all be answered through controlled and designed experimentation followed by appropriate statistical analyses (in some of the cases above, the same statistical tests), but that the focus and purpose of the activity is **not** to obtain t, F, X^2 and p-values, but to correctly answer the research question(s) at the lowest possible cost. Additionally, some of the studies required to answer these research questions may cost a company tens or hundreds of thousands of dollars to execute ..... you get one chance at the experiment, and there are no 'do-overs'. Therefore, the applied statistician (researcher) must understand threats to the internal validity, and subsequent external validity, of the design employed. Statistical methods may often be robust against certain violations of underlying assumptions, but no test statistic known can save the researcher from a flawed experimental design. This is of particular concern when one notes that performing research in many business and industrial situations is nothing like conducting experiments in a laboratory, classroom, or at the end of a chapter in a textbook. Most continuous distributions encountered are not normally distributed (regardless of what

some so-called practitioners / instructors would have you believe); many planned sample sizes are not realized; subjects or test units are often lost; data turns up missing; gauges are modified in mid-study. Knowing how to adjust and control for all of these types of occurrences can make the difference between a successful research effort and a catastrophic event.

Finally, contemporary parameters require that in the conduct of research and statistical analyses, obtaining a correct answer is not enough; if that answer is not associated with a set of conditions or outcomes that may be achieved at the lowest possible cost, and highest profit, to the organization implementing the solution. Concurrent with this observation is the realization that no solution will be implemented by upper-level management (particularly the high risk variety) if the results of the research and its associated statistical analysis are not presented in such a way so as to convince those who do not understand the difference between alpha and beta (i.e. Type I and II error), much less the revelations offered by a Bartlett-Kendall-Box procedure, why a particular solution is the right one. It is the researcher's responsibility to translate the analysis conducted from the universal language used to discuss variability (i.e. statistics) into the language of business (dollars).

Given the rationale presented, this course has been organized by the instructor to maximize the probability that the student successfully learning this content will be able to operate in the manner described above. Each case study presented will require the student to select an appropriate investigative strategy, select an appropriate statistical analysis plan (mindful of the nature of the data available, and the underlying assumptions associated with the test involved), correctly execute the analytical plan, and make recommendations for further research / action in a format consistent with the presentation of such data in a business environment. Each case study has been carefully selected so as to make certain that the student will have covered a significant number of diverse statistical analyses by the end of the course.

4. **Instructor**

Dr. Jeffrey T. Luftig
ECOT 414 – College of Engineering & Applied Sciences
Office Hours : 9:00-11:00 P.M., Tuesday & Thursday
e-mail : Jeffrey.Luftig@colorado.edu

5**. The Course Website**

All students must register on the web at XXXXXXXXXXXXXXXXXXX for access to eCompanion (look for the *register* link). ALL of the lectures and materials used in this course are contained on this website. Additionally, all communication associated with

this course will take place via the e-mail function on the course website; students will be responsible for checking their e-mail frequently enough to be knowledgeable about the information sent by the instructor to them using this utility throughout the semester.

A Note on e-Mail Addresses

When you register on ECollege, you can set your e-mail address to be sent from the course website to any address you choose. I know it can be a hassle to check multiple e-mail locations on a daily / regular basis, but may I suggest that you use you CU e-mail address for this purpose (e.g. firstname.lastname@colorado.edu). If you do not know your assigned CU address, contact ITS (dial 5-HELP from any on-campus phone) and they will provide it to you. The reason I mention this is because in the past, I have sent e-mails to students with attachments of interest as the course proceeds. In some cases, where students have used their work e-mail address, the e-mail goes through but the attachment is blocked. In some rare cases, fire-walled work systems, when blocking an attachment, block the associated e-mail as well, and provide no indication to the student that anything was sent out by me and subsequently blocked. It is *your* responsibility to be cognizant of any information sent via the course website by me, so it might be useful to use either a CU or at least personal e-mail address that will not present this sort of possibility.

Additional information associated with the use of the ECollege website and accessing videos of the lectures are available on the ECollege website, under the **Course Home** section, in the Content Item labeled '**Instructions**'.

6. **Textbook / References**

After careful review, the instructor has determined that no single text can adequately comprise a resource for **all** of the topics we will review in this course. The primary text for this course is *Design of Experiments in Quality Engineering*, by Jeffrey Luftig & Victoria Jordan, McGraw-Hill Publishing Company, 1998. The material in this text was provided in large part to students in EMEN 5900. Additional excerpts from *Experimental Design and Industrial Statistics – Level 3,* by Jeffrey Luftig, Luftig & Warren International, 1993 will also be used to support some of the lecture material, and will be provided to students on the course website. For content associated particularly with Simple and Multiple Regression, the textbook which will provide some of the background material for this course is the text employed in APPM 4580/5580: *A Second Course in Statistics: Regression Analysis (5th Edition)* by Mendenhall & Sincich, Prentice-Hall Publishing Company. Additional support material in this area will be provided to students on the course website.

A textbook that is recommended highly as a compendium of statistical tests for any student serious about utilizing statistical methods in research is the *Handbook of Parametric and Non-Parametric Statistical Procedures* by David Sheskin, published by

CRC Press.

Some of the other textbooks which may be used as references for some or most of the material reviewed in this course (these are only sample references and are not intended to constitute a collectively exhaustive list) include the following publications:

Box, Hunter, and Hunter	*Statistics for Experimenters*, John Wiley & Sons
Campbell & Stanley	*Experimental & Quasi-Experimental Designs for Research*, Rand McNally College Publishing Co.
Daniel, C.	*Applications of Statistics to Experimental Design*, John Wiley & Sons
Draper & Smith	*Applied Regression Analysis*, John Wiley & Sons
Dowdey & Wearden	*Statistics for Research*, John Wiley & Sons
Gibbons, J.	*Nonparametric Methods for Quantitative Analysis*, American Sciences Press
Hicks, C.	*Fundamental Concepts in the Design of Experiments*, Holt, Rinehart, and Winston
Siegel & Castellan	*Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill

7. **Course Structure / Approach**

The course topic outline which follows identifies the proposed and tentative lecture topics which will be presented throughout the semester in order to allow the student to achieve the goals and objectives of the course.

For each topic covered, it will be the student's responsibility to:

(a) have read any assigned material identified by the instructor ***before*** the corresponding lecture; and

(b) attend the lecture, and participate in classroom discussions of the material presented (on-campus students)

OR

review the lecture, and post questions associated with the content on the course website (distance students); and

(c) complete all homework assigned following each lecture. These assignments are designed to allow the student to confirm that the content presented in any given lecture has been learned and mastered, before material of increased complexity is presented in subsequent lectures.

The course requirements are as follows:

(1) Each student will complete and electronically submit all homework Assignments. The N assignments in this course are equally weighted, and the number of assignments submitted will be a function of how much content is ultimately covered.

(2) Students will be assigned to teams, and each team will be provided with a research problem and associated data set from an actual business application (i.e. Research Study). It will be the team's responsibility to:

(a) Perform all correct and appropriate statistical analyses required to answer the research question(s);

(b) Prepare a report suitable for presentation to a Chief Executive Officer and management team in business or industry; and

(c) Provide a presentation to the class on the methods utilized, techniques employed, and the results obtained. The members of each team will receive the same grade for the final project.

8. **Grading & Final Course Grade Calculation**

The final grade for the course will be assigned on the basis of the following weighting:

- Homework Assignments	75%
- Team Project	25%

For purposes of illustration, the final grade will be calculated by assigning the weighted averages of the scores received

e.g.	Homework Assignment 1:	B +	=	3.30
	Homework Assignment 2:	A	=	4.00
	Homework Assignment …….	….	=	…..
	Homework Assignment N:	A-	=	3.70
	Homework Assignment Sub-Total Grade: (For All N Assignments)	B+	=	3.30
	Research Project :	A-	=	3.70

Final Grade = { (75 * 3.30) + (25 * 3.70) } / 100 = **3.40 = B+**

When compared to the following scale:

3.85 - 4.3000	A	
3.50 - 3.8499	A-	
3.15 - 3.4999	**B+**	(the category within which the example above falls)
2.85 - 3.1499	B	
2.50 - 2.8499	B-	
2.15 - 2.4999	C+	
1.85 - 2.1499	C	
1.50 - 1.8499	C-	
1.15 - 1.4999	D+	
0.85 - 1.1499	D	
0.70 - 0.8499	D-	
< 0.70	F	

The ECollege Gradebook:

Each submission is assigned a letter grade and GPA value (A+ is possible), which is converted by the ECollege Gradebook into a percentage which is equivalent to the GPA value earned against the value of an 'A'. For example, a B is equivalent to a GPA value of 3.0. This value is divided by the value of an 'A' of 4.0, and converted into a value of 75%. B+ equals 3.3, and is converted into an 82.5%. An A+ would be 4.3/4 or 107.5%. To generate your final grade, ECollege multiplies each percentage by its weight, and generates a final cumulative percentage (fcp). You can determine your final grade by comparing your final cumulative percentage to the following table:

Final Weighted Percentage Earned - Low	Final Weighted Percentage Earned - High	Final Course Grade Equivalent
96.250	107.500	A
87.500	96.249	A-
78.750	87.499	B+
71.250	78.749	B
62.500	71.249	B-
53.750	62.499	C+
46.250	53.749	C
37.500	46.249	C-
28.750	37.499	D+
21.250	28.749	D
17.500	21.249	D-
< 17.50		F

For example, using the same grades as previously described:

Homework Assignments	B+ = 3.30
Term Project	A- = 3.70

The final cumulative percentage (fcp) generated by ECollege would be:

fcp = ((75)(3.3/4.00)) + ((25)(3.7/4.00))
= 61.875 + 23.125
= 85

Converting the fcp to a letter grade using the table immediately above, the final grade earned would be a B+ (exactly the same as previously described).

All assignments are expected to be completed in conformance with generally accepted standards associated with Academic Honesty. The CU Honor Code website provides and explanation of these standards; as well as the Pledge each student will be asked to sign for each major course requirement submitted. If you at any time have any questions regarding what is and is not appropriate in this area, make certain to speak with the Instructor.

The JIC Provision

Any student who (1) completes all N homework assignments; AND (2) is carrying an average gpa equivalent score less than an A into the final week of class may elect to complete an open-book, open-notes, open-computer final examination. The grade received on the final examination will substitute for the homework assignment grade (whether higher or lower) in the calculation of the final grade for the course.

9. **Additional Course Clarifications Recommended for Inclusion by the Office of Undergraduate Education, University of Colorado – Boulder**

a. If you qualify for accommodations because of a disability, please submit a letter to me from Disability Services in a timely manner so that your needs may be addressed. Disability Services determines accommodations based on documented disabilities. Contact: 303-492-8671, Willard 322, or http://www.Colorado.EDU/disabilityservices

b. Campus policy regarding religious observances requires that faculty make every effort to reasonably and fairly deal with all students who, because of religious obligations, have conflicts with scheduled exams, assignments or required attendance. In this class, my expectation is that each student will review the policy details http://www.colorado.edu/policies/fac_relig.html) and religious holiday calendar (http://www.interfaithcalendar.org/) during the first week of class, and by the end of the second week of class notify the instructor via e-mail as to what dates the student anticipates they will not be in class, and the religious holiday they will be observing.

c. Students and faculty each have responsibility for maintaining an appropriate learning environment. Students who fail to adhere to behavioral standards may be subject to discipline. Faculty have the professional responsibility to treat students with understanding, dignity and respect, to guide classroom discussion and to set reasonable limits on the manner in which students express opinions. Policies regarding classroom behavior may be reviewed at:

http://www.colorado.edu/policies/classbehavior.html

http://www.colorado.edu/studentaffairs/judicialaffairs/code.html#student_code

d. All students of the University of Colorado at Boulder are responsible for knowing and adhering to the academic integrity policy of this institution. Violations of this policy may include: cheating, plagiarism, aid of academic dishonesty, fabrication, lying, bribery, and threatening behavior. All incidents of academic misconduct shall be reported to the Honor Code Council

(honor@colorado.edu; 303-725-2273). Students who are found to be in violation of the academic integrity policy will be subject to both academic sanctions from the faculty member and non-academic sanctions (including but not limited to university probation, suspension, or expulsion). Additional information on the Honor Code can be found at:

http://www.colorado.edu/policies/honor.html

and at

http://www.colorado.edu/academics/honorcode/

e. The University of Colorado Policy on Sexual Harassment applies to all students, staff and faculty. Sexual harassment is unwelcome sexual attention. It can involve intimidation, threats, coercion, or promises or create an environment that is hostile or offensive. Harassment may occur between members of the same or opposite gender and between any combination of members in the campus community: students, faculty, staff, and administrators. Harassment can occur anywhere on campus, including the classroom, the workplace, or a residence hall. Any student, staff or faculty member who believes s/he has been sexually harassed should contact the Office of Sexual Harassment (OSH) at 303-492-2127 or the Office of Judicial Affairs at 303-492-5550. Information about the OSH and the campus resources available to assist individuals who believe they have been sexually harassed can be obtained at:

http://www.colorado.edu/sexualharassment/

f. Appropriate Classroom Laptop Use

Although having a laptop in class opens up new learning possibilities for students, sometimes students utilize it in ways that are inappropriate. It is easy for your laptop to become a distraction to you and to those around you. Therefore, please refrain from instant messaging, e-mailing, surfing the Internet, playing games, writing papers, doing homework, etc. during class time. Acceptable uses include taking notes, following along with the instructor on PowerPoint, and other directed class activities, as well as working on assigned in-class activities, projects, and discussions that require laptop use.

10. **Lecture Topics & Associated Discussion Points**

Primary Lecture Topic	Case Studies	Statistical Tests Reviewed & Discussion Points
Introduction to the Course & Course Requirements * Syllabus * Content Outline * Assignments / Papers * Examination * Suggested Support References	N.A.	N.A.
Review of EMEN 5900 / APPM 4580/5580 Final Examination Solutions	N.A.	* One Way ANOVA & Blocking * Post-Hoc Analyses and Procedures * Non-Parametric Analyses and their Post-Hoc Procedures

Primary Lecture Topic	Case Studies	Statistical Tests Reviewed & Discussion Points
Two Factor Analyses * The Design & Analysis of Factorial Experiments for 2 Factors : The Basics	The Case of the Flagging Labels : Maximizing Label Adhesion for Anheuser-Busch (J=3, K=3) Setting Up the Lincoln, Nebraska Injection Molder (J=2, K=4) Designing a Humidity-Resistant Packaging Method for Baked Goods (J=5, K=4) The Case of the Saw Blade Suppliers : Make Us More Money (J=2, K=2) Rolling Can Stock : Where's the Earing Coming From ? (J=3, K=3)	* Two Way ANOVA - Model I * Two Way ANOVA - Model II * Two Way ANOVA - Model III * Underlying Assumptions and Determining Expected Mean Squares / Appropriate Error Terms (AETs) * Crossed and Nested Designs * Approaches for Unequal Cell Counts (Disproportionate Frequency Analysis) * Interaction Analysis : Making Interactions Your Friend * Common & Recommended Transformations for Nominal and Ordinal Data

Primary Lecture Topic	Case Studies	Statistical Tests Reviewed & Discussion Points
The Design & Analysis of Factorial Experiments with Multiple (>2) Treatments * The Design & Analysis of 2^3 Factorial Experiments * The Design & Analysis of Factorial Experiments : 3 or More Factors at 2 or More Levels	How Should I Run My Heat Treat Furnace ? (J=2, K=2, L=2; Model I, Fully Crossed) Can I Control Density on the Molded Preforms ? (J=2, K=3, L=2; Model I, Fully Crossed) Let's Eliminate Earing Once and for All ! (J=3, K=3, L=15; Model III, Fully Crossed) Should We Purchase the New Bubble Gauge ? (J=3, K=3, L=3, M=5; Model III, Partially Nested)	* Three Way ANOVA - Model I * Three Way ANOVA - Model II * Three Way ANOVA - Model III * Underlying Assumptions and Determining Expected Mean Squares / Appropriate Error Terms (AETs) * When There's No AET : The Quasi-F Ratio & Satterthwaite's Formula * Fully Crossed, Partially and Fully Nested Designs * Advanced Interaction Analysis: Making Interactions Your Client's / Company's Friend

Primary Lecture Topic	Case Studies	Statistical Tests Reviewed & Discussion Points
Review of Content from EMEN 5005/5830 & APPM 4570/5570 – Simple Regression and Correlation	Sample Data Sets & Problems for Review Provided by Instructor (Course Website)	* Using Simple Regression to Describe a Linear Relationship * Testing Inferences About the Population Regression Line, the Intercept (β_0), and Slope (β_1), and the Underlying Assumptions of the Model * Assessing the Fit of the Regression Line – Using the ANOVA Table * The Coefficients of Correlation, Determination, and Alienation * Prediction and Forecasting; Confidence and Prediction Limits * Using SPSSPc for Correlation & Regression Analyses. * Generating Output & Testing Assumptions Using MVPStats and SPSSPc
Introduction to Multiple Regression Analysis	Series of Review Problems Supplied by Instructor (Course Website)	* Underlying Theory & Assumptions – Linear Models * Testing Inferences About the Regression Coefficients * Assessing the Fit of the Regression Line; the ANOVA Table, the Coefficient of Determination and the Multiple Correlation Coefficient * Full & Reduced Models; Forward and Backward Regression Approaches * Prediction Using the Multiple Regression Model * Lagging Variables in Time Series Analyses * Generating Output with SPSSPc

Primary Lecture Topic	Case Studies	Statistical Tests Reviewed & Discussion Points
Assessing the Assumptions of the Multiple Regression Model	Series of Review Problems Supplied by Instructor (Course Website)	* Testing The 5 Basic Assumptions of the Model - The Relationship is Linear - Variance is Constant - The Residuals are Normally Distributed - The Residuals are Independent - There is No Multicollinearity * Residual Analysis and Corrections for Model Violations * Utilizing Other Influence Statistics (Durbin-Watson, Cook's D) * Handling 'Outliers' * Generating Response Surfaces * Multiple Regression and Model-Fitting Utilizing TableCurve3D.