Statistics Projects – Three Examples to Relate Theory and Application

Robin Lovgren, Michael Racer
University of Memphis

Abstract

The application of statistical concepts can play an important role in an engineering analysis or design. These concepts and their applications are sometimes difficult to convey to engineering students in a typical classroom setting. To aid the students in making the connection between lecture and real world applications, a series of projects was developed, and assigned to the students. This paper provides three examples that are used in an undergraduate engineering class. Sample student responses are provided and general results are examined.

I. Introduction

Recent assessments of the skills required by engineers have revealed that current engineering curricula need to be revised to show students the techniques and value of statistical analyses. In particular, the recent ABET2000 initiative highlighted the need for fundamental statistics, data analysis, and design of experiments.

The authors of this paper have been involved in the instruction of statistics courses for engineering students in various areas, and have recently instituted a series of projects within the classroom to enhance the students' awareness of the need for statistics.

This paper presents a brief synopsis of this effort, providing detail for a few of the projects, and an assessment of the outcomes to date.

II. Objectives

The fundamental objective in developing these projects was to motivate interest in statistical analysis and applications. A statistics course was recently added to the civil engineering curricula at The University of Memphis, and there seemed to be much resistance by the students to the learning of this material. The course was originally taught in a traditional lecture format with limited results. Upon reevaluation of the situation, the following problems were identified:

- the students were slow to learn, understand, and assimilate statistical definitions, terms, and concepts;
- they had a hard time seeing applications for statistical methods. Working problems out of the book did not automatically convince the student that tools could actually be applied to that situation in the real world;

- the students were not convinced that there would be a true need for statistical analysis in their future careers as engineers.

The projects were designed to teach the students the technical concepts as well as the potential for application in a variety of settings in the real world – thus motivating interest in the topic.

III. Sample Projects and Outcomes

The three projects presented here were designed to foster the students' understanding of the value of statistics, to acquire hands-on experience in statistical analysis, and to develop the capacity to learn new concepts on their own. Before presenting the three projects, we identify the primary goals of each.

Project	Progressive Expectations
#1 - Food Court Visit	- Data Collection - Data Analysis
#2 - Work Experience	- Recognition of Opportunities at Work
#3 - Markov Chains	- Self-Learning

Project #1 – Food Court Visit

Task:

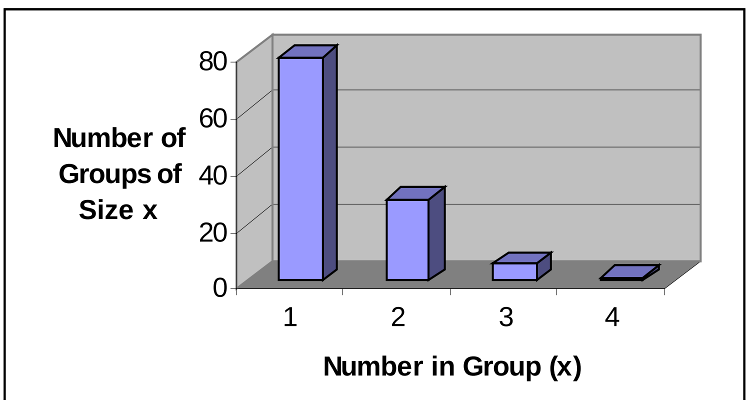
Visit a mall food court. Your task is to collect data such that the following questions can be answered: How many people enter the food court but do not eat? What sizes of groups of people arrive at the food court together? How long is the service time at Chick-fil-A(or similar restaurant)? How many of the people that eat at Chick-fil-A (or similar restaurant) order fries? Which restaurant is the most popular? How do they rank?

Discuss the following. What parameters influence the answers to each question? How did you determine your approach to data collection? How did you limit your study? What is the impact? Sketch the sample space for each of the questions posed. Which questions can be captured as discrete distributions? As continuous? Which questions involve permutations? Combinations? What other outcomes might we be interested in investigating?

Student Work:

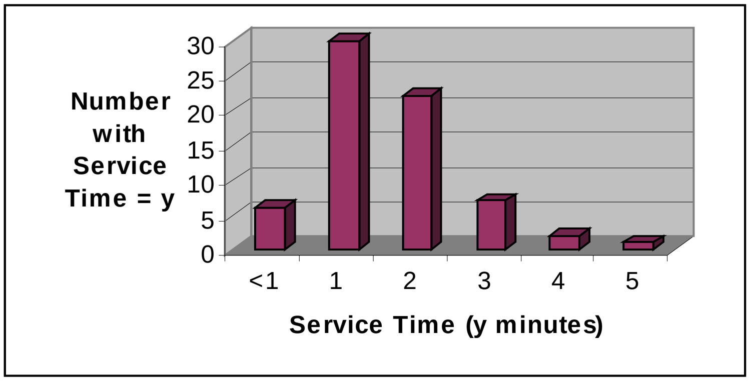
Typical sample results and report excerpts are given for a few of the questions.

“The number of people who enter the food court and the



number of people who do not eat could be very useful in deciding whether or not the franchise wishes to take a risk and place a restaurant there."

"This data could also be used by the mall to evaluate whether or not customers are satisfied with both the conditions of the food court and the selection of restaurants in the food court."



"The time data could be useful in order to evaluate how efficient the store is running and decide if any measures need to be taken to ensure customer satisfaction."

"The area selected to study was quite large … contributed to some loss of a representative sample."

Instructor Observations:

- The students were overly concerned with collecting the "right" data and the right amount of data, relative to the time constraints. They found that they sometimes had incomplete observations.
- The project helped to solidify in the students' minds the relationship between data analysis and application of methods (e.g. understanding that the "probability of success" for the binomial must be identified from some prior observations)
- The students recognized the applicability of statistics in this setting, which made later applications easier to accept

The educational value of this project was that it showed the students first-hand the role of data collection in creating statistical models. The students also learned how common activities could be modeled using some fundamental probability distributions (most notably binomial, hypergeometric, and exponential).

Project #2 - Work Experience

Task:

Consider the current jobs of at least 2 individuals in your group, and describe three problems. One must be an application utilizing the binomial, hypergeometric, negative binomial, or geometric distribution. One must be an application of some distribution that has not yet been presented in CIVL3113. One must be a potential application that would require more sophisticated modeling (e.g. compound distributions)

Presentations will be short - 5 minutes. The presenter will not be the expert. Discuss the business. Discuss the application.

In-depth discussion will follow. Within each discussion group, do the following: for one company, pose a set of at least 5 questions that could be addressed probabilistically; which involve discrete distributions? Continuous? For one of the problems presented, consider the following in more depth: how/whether the problem is simplified; how the problem could be more completely defined; data collection requirements; costs of collecting data - identify, but do not collect.

Student Work:

Typical sample results and report excerpts are given for a few of the questions.

- Cut & Fill for Road Design – group evaluated the impact of measurement spacing on the comparison of design vs. as-built specifications. "What is the probability the as-built elevation will equal design elevation?"- Proposed using the binomial distribution.
- GIS Digitizing – group evaluated the quality of an area digitized as a function of the amount of time the digitizer spent on the task. "How does the length of time spent digitizing affect the number of errors?"- Proposed using a non-homogeneous Poisson distribution.
- Chemical Decay – group evaluated "product spiking", in which chemicals that have been on the shelf for awhile are mixed with a newer product to prolong sellability, at a lesser value. "After one year on the shelf twenty-five percent of the barrels with a particular chemical makeup are known to not meet the requirements for sellability. If a company put in an order for 20 barrels, what is the expected number of barrels that need spiking?" - Proposed using the hypergeometric distribution.
- Bar Inventory – group evaluated the expected lifetime of liquor, based on customer requests. "Given the expected demand for the number of drinks per day, the quantity of alcohol per drink, and the amount of alcohol available, what is the probability that the bar runs out of alcohol that day?"- Proposed using the normal distribution.

Instructor Observations:

- The students, most of whom work, had no trouble identifying probabilistic problems at work.
- The students spent a reasonable amount of time on this project.
- One group missed opportunities in the work environments of its members.
- Groups were reluctant to have a non-expert be the presenter.

The primary outcome of this experiment was that the students learned to statistically analyze situations within their own technical job environments. In previous classes, the students had been able to learn the statistical concepts such as the idea of "trials and successes" in the binomial distribution, and viewing data through histograms, etc. They also learned to recognize instances of uncertainty in their field of engineering such as beam failures and construction specification tolerance limits. This project helped the students to make the connection between the statistical tools and the engineering applications.

Project #3 - Markov Chains

Task:

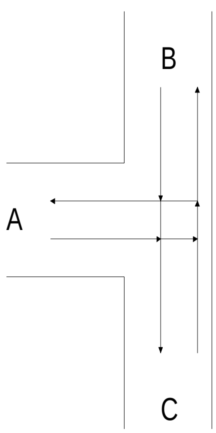
Choose one of the following environments for a Markov Chain analysis. (If you wish to select your own example, have it approved within the next week.) You may choose either the level of undissolved solids entering a wastewater treatment facility; the number of cars passing through some portion of a road, or intersection; rainfall levels of major rains; river level near a lock.

For your problem, collect a small amount of data, only enough to create a very basic model, and recognize transition probabilities. For your environment, discuss what costs are involved in the system, given the conditions of the period. Characterize these costs; you need not collect actual data.

For your problem, set up the model. Discuss the major limitations of the model. Discuss how transition and steady-state probabilities can be used in decision making. Provide an example of the use of transition and steady-state probabilities that can be used in decision making; provide an example of how sensitivity analysis with respect to costs and transition probabilities can be employed.

Student Work:

Traffic analysis:



The figure to the left shows the intersection evaluated by the students. They identified the six feasible transitions. Data was collected within a fifteen minute time period to determine the number of cars traversing to and from each direction. This group initially defined the three states as being the three road segments A, B, and C. Through discussions in class the table below was developed. The students were able to recognize that the state description of the vehicles must also include direction of travel.

The transition matrix in this table shows that cars traveling southbound on segment A (A/S) can either turn and travel eastbound on segment B (B/E) or westbound on segment C (C/W), etc. Cars traveling away from the intersection enter the Regeneration State. This state returns proportion R_A of the cars to the system on A/S etc.

Traffic Intersection Transition Matrix

States	A/N	A/S	B/E	B/W	C/E	C/W	Regeneration
A/N	0	0	0	0	0	0	1
A/S	0	0	0.72	0	0	0.28	0
B/E	0	0	0	0	0	0	1
B/W	0.15	0	0	0	0	0.85	0
C/E	0.45	0	0.55	0	0	0	0
C/W	0	0	0	0	0	0	1
Regeneration	0	R_A	0	R_B	R_C	0	0

"Should the intersection be redesigned to better accommodate traffic patterns?" "Costs of the system include pollution and fuel of cars waiting at the intersection, personal delay costs in terms of time, and possible costs of redesigning the intersection."

Instructor Observations:

- One group presented a good application, but failed to complete the task, because of poor instruction from the responsible group member.
- All students quickly understood the applicability of Markov Chains.
- All groups had easy access to large amounts of data that could have been used to enhance their models.
- The students learned about a number of modeling issues: assumptions, control variables, objectives, tradeoff, etc.

The focus of this project was on self-learning. The students were presented with a problem for which they had not been taught the appropriate statistical tool for analyzing – Markov Chains in this case. It was the students' responsibility to learn the material sufficiently on their own, and apply it to the problem. In particular, this introduced the students to the concepts of problem definition and modeling.

IV. Applications in Mechanical Engineering

While this course was implemented within a civil engineering curriculum, the concepts are just as applicable to mechanical engineering. Project #1 – Food Court Visit – focuses solely on the issue of data collection and analysis. Project #2 poses the greatest challenge, in guiding the students to identify opportunities within their own discipline. Much of the success of being able to accomplish this rests on any previous exposure to statistical aspects within the discipline.

One item that we should note here is that the students were simply asked to recognize the opportunities for employing statistical analyses, and not required to carry out the analysis. Most notably in the examples cited the Cut & Fill and GIS analyses are beyond the level of detail taught in this introductory course. The fact that the students could recognize an element of uncertainty in the environment was the key issue in this exercise.

One certain area of applicability would be in experimental design. For instance, in manufacturing a part, a mechanical engineer would be interested in evaluating the factors that influence the quality of the production, as well as the relative influence of these factors. The quality of turbine production, for example, might be related to properties of the raw material, skill of the manufacturer, quality of the tools, and other factors.

Quality control is also an issue within a manufacturing environment. On the design side, the engineer needs to be able to adequately assess reasonable, attainable tolerances. On the production side, the engineer needs to be aware of the quality of the part relative to established tolerances. The engineer must be able to recognize natural variability in production vs. significant disruptions.

One issue closely related to quality control, and certainly of importance to the mechanical engineer, is reliability. The engineer must ascertain how well and for how long some system component will be able to function. This most often involves probabilistic analyses.

Expectations play an important role within any environment considered by a mechanical engineer. For instance, an engineer designing an HVAC system must be able to assess expected demands on a system. In order to determine the expected lifetime of an air conditioning system, for instance, the mechanical engineer must consider the uncertainties inherent in the weather as well as personal usage preferences.

Another example of the need for an evaluation of expected demands would be within the area of power generation. On a daily basis, the demand for power fluctuates constantly, as it does on a day-to-day and seasonal basis. The impact of inadequately planning for expected demands – and being unable to accommodate fluctuations – is significant. Overestimating power demands results in excess costs, whether it's design costs on the front end, or production costs during the day. Underestimating demands are similarly hurtful, and most strongly felt by the consumer. Brownouts or blackouts on high demand days are obvious effects. Systems that are under-designed will often have to be redesigned or have services augmented by other facilities. Both of these are costly propositions.

These three areas – quality control, reliability, and uncertainties of demand – pose a wide range of potential applications within the realm of mechanical engineering. All could be well-suited for the Project #2 environment.

As with Project #1, the goal of Project #3 truly fell outside the realm of any specific engineering discipline. In Project #1, the students were asked to focus on the issues involved in data collection and analysis – something that they could already relate to on some level. The goal of Project #3, on the other hand, was to investigate a new topic – Markov Chains (MC) – and recognize the usefulness of that tool. The major task is self-learning.

Simplified civil engineering environments were posed to the students. Likewise, in a mechincal engineering course, the same could be done. A simple MC problem related to HVAC might be to ask the student to develop a model showing how AC demand might vary from day to day, and how the MC analysis could be useful in assessing the lifetime system demand. In the area of quality control, a Markov Chain analysis could be done in order to track deviations from tolerances, and flag the occurrence of likely significant problems. A Markov Chain analysis of historical reliabilities could be used to help a mechanical engineer determine the most cost-effective strategy for replacing machine components. As suggested earlier, early replacement of a part could be considered somewhat wasteful, while failure to replace parts in a timely fashion could result in serious production defects.

V. Conclusions

The projects have been a successful mechanism for improving the statistics course in terms of technical capabilities, as well as motivational needs. The quality of the work has improved in the class. The students accept and understand the terminology and concepts more readily than with the traditional lecture format. There is an increased awareness of the presence of statistics in our everyday lives as well as the possible use in a professional engineering career. As was hoped, the students' comfort level increased after the implementation of projects and their communication skills improved throughout the semester with each oral and written presentation. The questions and issues posed in class were more sophisticated and well thought out. One noted improvement that was unexpected was that after the first few projects, the students were much more open to the introduction and discussion of new statistical concepts and extensions (e.g. design of experiments). A survey given after completion of four such projects in one semester showed that 33% of the students felt that the course exceeded their expectations and 47% felt that the course met their expectations. These results were a great improvement from the previous semesters.

The advantages of these and similar projects in the ME curriculum is twofold. First, using such projects can teach the students real world applications of statistics, which is a valuable addition to their education, as well as satisfying a requirement that appears to be solidly a part of the ABET 2000 requirements. Second, an independent statistics course does not have to be added to existing curricula. Instead, statistics-type projects can be made a part of a course in Engineering Analysis, which would make the course much more interesting.

One fundamental lesson learned from this course and these projects is the nature of statistics, relative to the nature of traditional engineering. Statistics requires an abstraction of thought, a consideration of non-physical concepts; this is the antithesis of engineering, which is very solidly rooted in physical concepts. The manner in which students must address a statistics issue is different. This must be understood and accepted by the instructor. It is the recognition of this distinction that is at the core of these projects. For the engineer to solidly accept and apply statistics there must be some bridging of the gap between the abstract and physical notions. A project oriented hands-on approach helps engineering students learn and understand statistics.

ROBIN LOVGREN
Robin Lovgren is an Assistant Professor of Industrial and Systems Engineering in the Civil Engineering Department at The University of Memphis. Dr. Lovgren received a Bachelor of Industrial Engineering from the Georgia Institute of Technology in 1986, and an M.S. degree in

Statistics, and a Ph.D. in Management Science from The University of Tennessee at Knoxville in 1996.

MICHAEL RACER

Michael Racer is an Associate Professor of Industrial and Systems Engineering in the Civil Engineering Department at The University of Memphis. Dr. Racer received a B.S. degree in Mathematical Sciences in 1984, and an M.S. degree and a Ph.D. in Operations Research from the University of California at Berkeley in 1986 and 1990 respectively.