# MULTI-PURPOSE VIDEO-TAPED COURSE IN DATA ANALYSIS

R. A. GREENKORN and D. P. KESSLER
*Purdue University*
*West Lafayette, Indiana 47807*

*The iterative process of formulating a mathematical model, design of experiments to test that model, analysis of the data from these experiments, the use of the experimental results to modify the hypothesized model, and the in corporation of the model in larger systems is one which is fundamental to all branches of engineering. Although this process is basic to the engineering analysis of problems and design procedures, there exist few courses in which the complete cycle is treated. The difficulty with teaching the complete loop by usual methods is that typically the background of the students is relatively disparate, therefore one is seldom able to teach to a body of students with uniform backgrounds. Nonetheless, we feel that such a course is important to engineering graduates so we have attempted to approach the problem using video tape.*

*A significant advantage in treating this type of subject, where backgrounds may not all be the same, is offered by video-tape and video-tape cassette capabilities. These tools permit different students to use different portions of the same course, and also permit these students to progress at varying rates as they so desire. We attempted to design a course which, for educational efficiency, we tried to fit to the needs of both continuing education students and full-time students on the campus, and, in addition, created the course in modular form so that it could be taught as various series of self-contained mini-courses to students who wanted only a portion of the overall material.*

*The course is multi-purpose basically in two different ways. One, there is a combination of sub-units which can be selected to accomplish the educational objective of each student, and two, the course applies to a variety of educational situations. The variety of path choices was accomplished by designing the course as a series of self-contained mini-courses which can be assembled to form a maxi-course in a variety of ways, depending on the education background of the particular student and his particular educational objective. The variety of educational situations which the course can be applied are: a) normal or self-paced classroom use, b) continuing education use, c) broadcast TV to larger segments of the community. An important but not primary purpose of this course is also to furnish a pilot effort toward a video correspondence Master's degree program which would permit a student at a remote location to complete requirements for a Master's degree by selection of an appropriate series of video courses. The course consists of 43 thirty-minute video tapes which are as shown in Table I.*

One is, in effect competing with a program like Sesame Street (with a 6 million dollar budget) in production and entertainment value while trying to present a much more sophisticated level of concepts to a much more critical and discriminating audience.

**Table 1—Course Content**

| Unit | | Parts |
|---|---|---|
| 1 | Introduction | 1 |
| 2 | Curve Fitting | 4 |
| 2-4 | Nomography | |
| 3 | Statistical and Numerical Errors | 3 |
| 4 | Differences and Lagrangian Methods | 3 |
| 5 | Least Squares | 3 |
| 6 | Population Characteristics | 2 |
| 7 | Probability | 2 |
| 8 | Sample Characteristics | 3 |
| 9 | Analysis of Variance | 3 |
| 10 | Regression | 3 |
| 11 | Matrix Regression | 3 |
| 12 | Dimensional Analysis | 2 |
| 13 | Model Building | 2 |
| 14 | Time Series | 2 |
| 15 | Inference | 2 |
| 16 | Factorial Designs | 3 |
| 17 | Systems/Networks | 2 |

## COURSE CONTENT

THE COURSE IS ORGANIZED in 17 units. Each of these units has from two to four parts with the exception of Unit 1 which is a single part introduction. Each part represents a 30-minute lecture. At Purdue the remaining part of a 50-minute period is used for discussion. In each of the units about half of the material presented is actual examples taken from practice.

UNIT 1 is the introduction and it sets the objective for the course, which is to interface theory and data. The use for the interface is to build models, plan experiments, process data, interpret data, and design data systems.

UNIT 2 is concerned with curve fitting and nomography, to permit summarizing data so that it can be interpolated and extrapolated, to check theory, and for empirical prediction of new data. The two parts to the curve fitting problem are: First, to determine the form of curve. This is usually accomplished by plotting the data in various ways until a straight line results. Second, to determine the parameters by fitting a straight line to the rectified data using the method of selected points, method of least squares. One of the parts of this unit discusses nomography, a graphical representation of the functional relationship among variables. We give a brief introduction to methods of constructing nomographs emphasizing addition, subtraction, multiplication and division.

UNIT 3 is concerned with statistical and numerical errors. The object here is to identify and separate statistical error, those random errors that are associated with measurement; and systematic error, those that are not random errors; and further errors that result from operation on the data numerically. We end the unit with a discussion of the meaning of accuracy and precision both in the statistical sense and in the sense of relating these concepts directly to the numbers involved in experiments.

UNIT 4 treats differences and Lagrangian methods. One often has to interpolate between data points, especially when data is in tabular form, and it is also often necessary either to differentiate or integrate tabulated data. We discuss the divided differences, backward, forward and central finite differences. We end with a discussion of Lagrangian methods specifically applied to numerical differentiation and numerical integration.

In UNIT 5 the principle of least squares is considered in detail. Also we begin an early discussion of how least squares and linear regression are related, since we use linear regression to predict statistical behavior. The principle of least squares is usually used to fit the data in regression analysis. We give a discussion of the use of least squares to identify important variables and consider the more complex polynomial least squares and nonlinear least squares.

In UNIT 6 population characteristics are discussed so that we can use statistical models of the various distribution functions to describe sample spaces. We discuss some of the simple distributions—the uniform distribution, the normal distribution—and the meaning of these distributions in a probability sense.

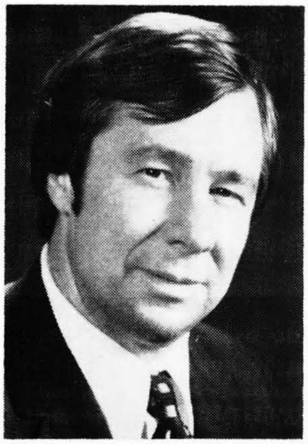
UNIT 7 is more detailed discussion of probability and investigates the meaning of experiments, outcomes, sample spaces and elements of sample spaces, and how these various concepts are utilized in probability formulations. A short discussion of probability in terms of logic and Venn diagrams is included. Marginal and conditional probabilities and the Bayes theorem are also discussed.

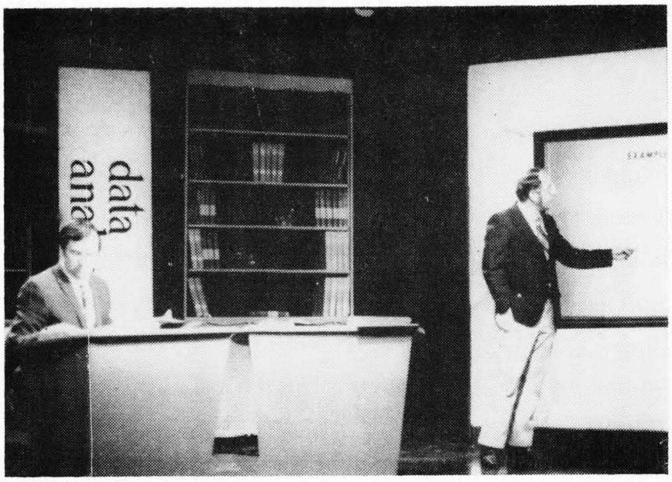
In UNIT 8 we discuss sample characteristics, concentrating on utilizing the normal distribution from a designed experiment, look at the probability meaning of distribution functions in terms of the normalization of these distribution functions and the relationship to probability. We discuss the use of various kinds of tabulated probability distribution functions and the distribution of sample characteristics. The unit ends with a discussion of confidence intervals and a preliminary treatment of hypothesis testing and type I and type II errors. These last topics are repeated in more depth in Unit 15.

In UNIT 9 we begin our discussion of experimental design by introducing the analysis of variance technique—dissecting total variation in such a way that various kinds of experimental effects are eliminated. The analysis of variance allows us to show how experiments may be de-

Robert Greenkorn, after five years as a naval aviator, returned to Wisconsin to obtain B.S., M.S., and Ph.D. in Chemical Engineering from the University of Wisconsin, 1954, 1955, and 1957 respectively. He spent the academic year 1957-58 at the Norwegian Technical Institute in Norway as a past-doctoral fellow. From 1958 to 1963 he was research engineer with Jersey Production Company in Tulsa and lecturer in evening division of University of Tulsa. From 1963 to 1965 was Associate Professor of theoretical and applied mechanics at Marquette University. Associate Professor in the School of Chemical Engineering at Purdue University from 1965 to 1967. Professor and Head of the School of Chemical Engineering 1967-1973; Acting Head of Aeronautical and Astronautical Engineering, June 1973-Aug. 1973. Asst. Dean for Research, Director, Institute for Interdisciplinary Engineering Studies, and Associate Director of the Engineering Exp. Station, Aug. 31, 1972-present. (LEFT)

David Kessler has taught at Purdue University since 1964. Prior to his academic career, he was employed in process engineering and statistical quality control by the Dow Chemical Company and in process and product development by the Proctor and Gamble Company. He did his undergraduate work at Purdue and received his graduate degrees from the University of Michigan. His current research interests are flow in heterogeneous, non-uniform and anisotropic porous media, momentum transfer in multiphase flow, and bioengineering (artificial blood, cardiac contractility, and hemorrhapic shock). He is co-author with Professor Greenkorn of the undergraduate text "Transfer Operations" (McGraw-Hill, 1972). (Right)






Professors Greenkorn (left) and Kessler (right) on the set for filming a unit of their multi-purpose video tape.

signed so that we can get the most information from the data. We discuss the one-way classification and two-way classification (and randomized complete block designs). The linear models associated with these kinds of designs are discussed as are the short-cut methods of calculating the analysis of variance table.

REGRESSION IS DISCUSSED in UNIT 10 based on the units on least squares and analysis of variance. Analysis of variance is used to interpret the meaning of regression coefficients in the various kinds of regression models. The "extra sum of squares" principle is introduced and methods for analyzing the meaning of the various regression coefficients in models that have more than one independent variable are considered.

In UNIT 11 regression analysis is viewed from the standpoint of matrix manipulations. There is a short review of linear algebra and matrix theory and then the matrix approach to regression is discussed with use of the Doolittle method for determining regression coefficients.

In UNIT 12 we enter a discussion of dimensional enalysis, a systematic way in which the number of variables required to describe a given experimental situation is reduced, since normal model building uses dimensionless forms. We also investigate the relationship between dimensional analysis and the differential equations which are the models for various experiments.

Model building is considered in UNIT 13 in a philosophical sense and we try to answer the questions: What is a model? How does it relate to the real world? How do we build models? Mathematical and physical analogs are discussed. Example models are formulated through use of an entity balance.

In UNIT 14 we treat time-dependent stochastic processes, that is, processes where the parameters of the probability density and distribution functions are time-dependent. Much of what we do in engineering is time-dependent and we cannot ignore this time-dependence. Ways and means of investigating the statistical properties of systems that do depend on time are considered. The ergodic assumption is also discussed.

THE PROBLEM OF INFERENCE and the estimating of population parameters from experiments in a detailed manner is discussed in UNIT 15. The meaning of inference is investigated in terms of the various kinds of distribution functions. The meaning of hypothesis testing and multiple-hypothesis testing are discussed and the operating characteristic curve for various kinds of hypothesis tests is introduced.

In UNIT 16 we consider factorials which are posed as experimental designs—randomized block and Latin square. The meaning of factors in experiments is analyzed using the linear hypothesis and is based on the discussion of inference and hypothesis testing in the previous unit. We consider multi-factor experiments and how one confounds data in a factorial experiment. The use of aliases in designing fractional factorial experiments is also discussed.

In UNIT 17 we look at the total data acquisition and analysis system. Network models and graph theory are discussed. Information flow as related to executive programming is also considered.

## USE OF COURSE

WE PRESENTLY TEACH the course in its entirety over the Purdue closed-circuit video facilities. As can be seen from the network diagram, a number of ways to trace out either the total course or selected sub-sets are available. Typical mini-courses might be Units 9, 10, and 11 in Regression or Units 9 and 16 in Experimental Design. Most of the individual Units also stand alone without reference to other units.

In the future we hope to incorporate all segments of this course on video cassettes which can be played over monitors equipped so that the tape may be stopped without erasing the picture from the screen. This will permit much greater economy in presenting graphical material, in that the student can simply stop the monitor and hold the picture on the screen rather than wasting several minutes of tape for a static display.

## DIFFERENCES FROM CONVENTIONAL COURSES

IT IS INTERESTING to observe the reactions of students when viewing a course on what looks like a conventional television set. They react to the course much as one observes groups of people reacting to television programming in

In the future we hope to incorporate all segments of the course in video cassettes which can be played over monitors equipped so that the tape may be stopped without erasing the picture.

their home—that is, there is far less reluctance to create a disturbance, much as one will carry on a conversation in one's own living room while the TV set is on. There also is a much greater need for entertainment value to hold the student's attention than in an ordinary class room lecture, because the students, seeing the material on the television set, expect a far more professional degree of treatment than is true in the ordinary lecture. One is, in effect, competing with a program like Sesame Street (with a six million dollar

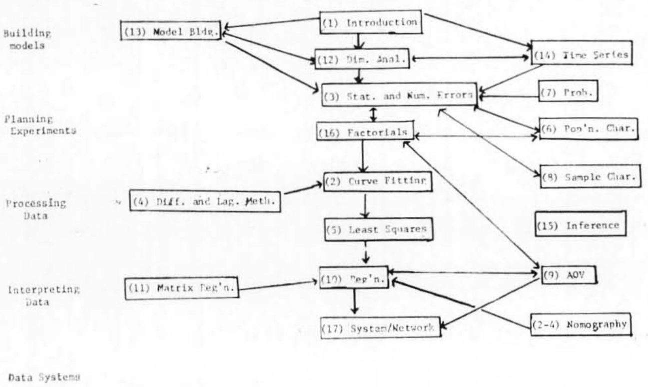



FIGURE 1. Network Diagram

budget) in production and entertainment value, while at the same time attempting to present a much more sophisticated level of concepts to a much more critical and discriminating audience.

It is also interesting that the students do not perceive the pace at which the course is going. At times they feel that the material is coming quite slowly when, in fact, because of the compactness of the presentation, material is being presented at a far greater rate than was ever possible in an ordinary classroom lecture. Students are also far more critical of mistakes that appear on a television tape than mistakes that appear in an ordinary classroom lecture. (The preparers of the tape, of course, should also be extremely critical of such mistakes because these mistakes will be perpetuated from year to year.)

It is interesting that the television tape prompts a far greater need on the student's part to be supplied with **All** the material than does an ordinary classroom lecture—students appear to feel that since a course is taught on TV there should be no need to consult outside references. Again, this seems to be a psychological set induced by commercial TV viewing. In the future we may attempt to remedy this by calling for more response from the class during the television taping via short questions, etc. This, perhaps, is one of the strongest reasons why television tapes must be entertaining—the student cannot partipate by talking back to a television screen in the same way that a good lecturer can stop and ask questions at a pertinent point in the presentation, and listen to feedback from students. At present there is no practical possibility of branching or changing pace in a television presentation as there is in the ordinary classroom lecture. We hope to circumvent this difficulty to some extent by keeping individual presentations short and thus permitting the student to select among a variety of short presentations so that if the pace becomes too slow or too fast he can alter the pace to suit himself. In the future we also hope to tape a greater variety of example problems so that the student can go directly to an example problem if he has difficulty with the theoretical concept which has been presented on the tape.

**In taping the course we used a producer/director and three cameramen, with visual material on rear-projection slides and newsprint. The set is shown in the photo. One of the major difficulties is the preparation of visual material (about 1000 items for this course). We hope to do some work soon on automating much of this with the computer. Our current production costs (exclusive of authors) is about $300 per Unit.** □

**TO DEPARTMENT CHAIRMEN:**

**The staff of CHEMICAL ENGINEERING EDUCATION wishes to thank the 72 departments whose advertisements appear in this sixth graduate issue. We also appreciate the excellent response you gave to our request for names of prospective authors. We regret that, because of space limitations, we were not able to include some outstanding papers and that certain areas are not represented. In part our selection of papers was based on a desire to complement this issue with those of the previous years. As indicated in our letter we are sending automatically to each department for distribution to seniors interested in graduate school at least sufficient free copies of this issue for 20% of the number of bachelor's degrees reported in "ChE Faculties." Because there was a large response to our offer in that letter to supply copies above this basic allocation, we were not able to fully honor all such requests. However, if you have definite need for more copies than you received, we may be able to furnish these if you write us. We also still have some copies of previous Fall issues available.**

**We would like to thank the departments not only for their support of CEE through advertising, but also through bulk subscriptions. We hope that you will be able to continue or increase your support next year.**

**Ray Fahien Editor**