

A Course in Statistical Analysis for Engineers in an Integrated Engineering Curriculum

John Kinney
Rose-Hulman Institute of Technology

Rose-Hulman Institute of Technology is one of a number of institutions comprising the Foundation Coalition which is funded by the National Science Foundation. The Coalition seeks to create model programs in engineering for national use which are tested at the Coalition institutions.

With a freshman integrated program in place, a team at Rose-Hulman created a sophomore program in engineering during the summer of 1995. The philosophy of the program and some of the details regarding the engineering courses have been discussed by Froyd (1) and Richards (2). Courses in mathematics are coordinated with the courses in engineering and comprise primary mathematical threads for the curriculum. Courses in differential equations are of course included; recognizing the increased importance of data analysis for engineering as well as the current emphasis on quality methods and statistical process control in manufacturing, a course in statistics also comprises a main thread of the curriculum. The purpose of this paper is to discuss the reasons for including this course, its content and plans for future development.

Need for the course

Engineers, in research and in creating manufacturing processes, by necessity deal with samples or a portion of all the possible observations of a process, so it is recognized that statistical analysis of the data gathered is appropriate; hence we must produce engineers who are not only literate in a verbal sense but literate in a quantitative sense as well.

Statistics, relative to parts of mathematics that have been known for about 5000 years, is brand new mathematics. Although some facts were known before this, the substance of statistics has been developed since approximately 1920. Tremendous strides in areas such as

regression, fitting curves to data, and the design of experiments have been made in the past 10 years. It is not surprising then that statistics has only recently been recognized as important for the engineering curriculum.

Engineers in industry are increasingly called upon to design efficient experiments and to perform statistical analyses of the data gathered. Short courses by universities as well as private vendors are easily available, attesting to the need for retraining existing industrial engineers. It is then most sensible to produce engineers who are trained in statistics in the first place, rather than to produce a product that must be re-tooled early in its career.

Some manufacturing processes demand that the resulting product be inspected for quality before delivery and that the results of this sampling be delivered along with the product; engineers involved with such processes must then understand some principles of sampling and the consequences of that sampling.

Production methods themselves are currently under heavy scrutiny so that the resulting product meet or exceed customer specifications; observation of the production process often involves sampling at various stages in order to detect causes of variability. Control chart techniques in particular are popular devices for this purpose. Process capability studies are also of importance.

These reasons were more than sufficient for the team at Rose-Hulman to include statistics as a necessary thread in the sophomore program we devised.

A course philosophy and an outline

With the goal of providing industry with statistically literate engineers, we decided to emphasize the applications of statistics throughout we have no interest in the usual dice and coin games often played with beginning statistics texts. It also is a fact that probability by itself plays only a small role in statistical analysis at this level, so we included only as much of probability theory as we absolutely needed.

Engineers use regression techniques and designed experiments frequently so we started with some principles of sampling and the basis of designed experiments such as factorial designs, giving many examples of experiments of this sort. It is important from the outset for the student to recognize the use of the subject and its prevalence in the engineering world.



Drawing plots of data that are linear or curvilinear in nature is also important and so we showed various sorts of devices for exhibiting data visually and deciding, in a very informal way, whether or not one variable is a linear function of another so that we could determine whether factors in an experiment were linearly related.

We established a linear model for some experiments and showed how the effects of various factors can be determined using some simple geometry. We then used the sizes of these effects and the linear model to explain both average and individual observations from an experiment. This allows a discussion of random errors and their inclusion in a linear model.

The nature of these errors and the fact that they arise in a random fashion leads to a discussion of their probability distribution. The prevalence of normally distributed quantities is surprising here to most beginning students; this fact was illustrated, connected to the linear model, and a visual method for assessing normality, a normal plot of data, was given.

We have now established many statistical ideas without the burden of a formal statistical structure. This has the effect of interesting the student (although most know the importance of statistics) in our subject. I think it is vital to show only real and practical examples to our students.

These basic goals established, we then began to lay a mathematical foundation for statistical tests and procedures. This involved the definition of random variable and the idea of a probability distribution. We only had time to discuss some very basic distributions such as the binomial, Poisson, hypergeometric, normal, and exponential. We gave applied examples of these in every case, including some acceptance sampling using the hypergeometric distribution.

After the central limit theorem was noted, control charts for a process mean were used as examples.

Next the notions of estimation and hypothesis testing were discussed. This led to the creation of hypothesis tests in production processes involving means and variances. We established tests for comparing two samples which may have arisen from different manufacturing processes. This raises the question of comparing several samples and leads to an analysis of variance test. If the data arose from a factorial experiment we then determined the statistical significance of the factors whose sizes were found in the beginning of the course through geometry.

We also examined the probability distributions of regression coefficients and tests of hypotheses concerning them.

The course thus came full circle; we started and ended talking about designed experiments and regression techniques.

The role of the computer

The statistical analysis of data cannot be carried out without considerable computer assistance. Computers not only aid in the calculation necessary; they can also serve to make the subject visual.

Spending too much time on the computer in this course (and in many others as well) is a trap to be avoided. We used Minitab largely since it is extremely easy to use for our students who have already had considerable computer experience in calculus and differential equations where a computer algebra system is used extensively. We used a Windows version of Minitab which is completely a point-and-click program; the student needs to learn no line commands of any sort.

I also made some use of Mathematic, especially to draw graphs. The use of Mathematical or some other computer algebra system in courses in probability and statistics is discussed in Kinney (3).

In the future, our students will all have laptop computers in the classroom so that we can make classroom use of the computer.

Some observations and what to do next time

Although I had taught courses in probability and statistics for engineers for some time, I regarded this course as a new one, primarily because it was an integral part of a new overall curriculum and because of the non-traditional nature of the order of the topics and the emphasis given them.

Since we operate on the quarter system I have only 2000 minutes of the students' time. So I am keenly aware that each and every minute is of importance and cannot be squandered. With that in mind the next time I give the course I probably will spend less time on estimation; for one, confidence intervals and hypothesis tests are closely related (and thus reinforce each other), but it is probably not necessary to do both in a formal way for good students .

There are many more topics that could, and perhaps should, be included. I would like much more time on the analysis of designed experiments; we should spend more time on regression, especially with recent developments. One must caution, however, that our students are primarily deterministic thinkers and have only rarely been introduced to probabilistic models and random variation. They must shift intellectual gears to become comfortable with probabilistic thinking.

It is obvious that employers would like entering engineers to be more acquainted with statistical process control and quality methods in general.

The students are increasingly interested in hands-on experiences. It would be nice to include some projects for individuals or groups to carry out; data could be gathered and analyzed making the subject more alive in the student's mind.

There is increasing research evidence that students come to us with certain assumptions about the physical world and especially about the probabilistic world. One way to overcome some of these misconceptions is to perform experiments and illustrate that things do not behave in the way the student may think. This suggests simulation experiments with the computer for various topics, especially in determining the probability distributions of random variables.

All of these desirable require either more time, and even more sharply focused course, or both.



Bibliography

- 1] Froyd Jeff, Integrated Engineering Curricula, these Proceedings, Session 1230.
- 2] Richards, Don E., A New Sophomore Engineering Curriculum - The Rose-Hulman Experience, these Proceedings, Session 1230.
- 3] Kinney, John J., The Use of Computer Algebra Systems in Courses in Probability and Statistics for Engineers, these Proceedings, Session 3520.

Biographical Information

John Kinney is Professor of Mathematics at Rose-Hulman Institute of Technology where he has taught courses in probability and statistics for engineers and scientists since 1974. He holds a Ph.D. degree in Statistics from Iowa State University. He is the author of Probability: An Introduction With Statistical Applications, John Wiley&Sons, 1997.

