Serdica J. Computing **8** (2014), No 3, 227–232

Serdica
Journal of Computing

Bulgarian Academy of Sciences
Institute of Mathematics and Informatics

TEACHING STATISTICS TO ENGINEERS: LEARNING FROM EXPERIENTIAL DATA

V. Mandrekar

Abstract. The purpose of the work is to claim that engineers can be motivated to study statistical concepts by using the applications in their experience connected with Statistical ideas.The main idea is to choose a data from the manufacturing factility (for example, output from CMM machine) and explain that even if the parts used do not meet exact specifications they are used in production. By graphing the data one can show that the error is random but follows a distribution, that is, there is regularily in the data in statistical sense. As the error distribution is continuous, we advocate that the concept of randomness be introducted starting with continuous random variables with probabilities connected with areas under the density. The discrete random variables are then introduced in terms of decision connected with size of the errors before generalizing to abstract concept of probability.Using software, they can then be motivated to study statistical analysis of the data they encounter and the use of this analysis to make engineering and management decisions.

ACM Computing Classification System (1998): A.0, G.3, K.3.2.

Key words: Statistics for Engineers, Experiential data, quality control, statistical concepts, engineering decisions.

Introduction. In manufacturing machines one needs parts. Parts can be produced by using exact measurements required. Costs of producing such parts will be extremely high as maintaining quality will require constant monitoring and one has to discard several parts with cost of material. Deming advocated that goods should be acceptable within tolerance limits so as not to sacrifice quality. However, errors occurring in the production are random and the errors being within limits is an event with a certain probability. Of course, to be safe this probability is taken close to one or the probability of accepting bad parts is low. As randomness was connected with decision making in manufacturing, it was felt that engineers are to be trained by statisticians to make these decisions with some error and to compute the cost connected with it. Once this was realized a three-day working conference was held in July 1984 at the University of Iowa to develop recommendations for the statistical education of engineers [1]. "The report addressed the following questions: who are the customers? What are their statistical needs? What type of in house courses are desirable? How do we approach statistical education for engineering management? What is the best first course in statistics for engineering students? How do we convince the engineering profession that statistical ideas and methods can help to improve quality and productivity in American industry and its competitive position in international market?" [1] The basic idea of the report was to provide statistical tools to engineers. As stated in [3], "it is imperative that a dialogue between engineering and statistical professions take place if major gains are to be made in quality and productivity." At this stage the basic idea was to convince engineers that they should be customers of statisticians who will provide them with tools.

However, this being a preliminary approach, there were some problems one had to deal with, such as, what data to choose for motivating engineers in the study of statistics and how the study of this nature will impact their manufacturing quality. The interaction will be one way between engineers and statistics: mainly, learning the statistical tools. The statistical concepts were taught based on the textbook and artificial data given in the problems in the textbook was used to illustrate the tools. Conclusions failed to motivate the engineers as they could not relate to the data. This approach of the use of artificial data has the following disadvantages as described in the general setting in [2]. "We believe that the practice of using artificial data sets to demonstrate statistical methods in applied statistics courses removes much of the intrinsic interest in learning to do good data analysis and contributes to the myth that statistics is dry and dull." One should therefore eliminate artificial data sets. Further, "real data supplemented by suitable background material enable students to acquire analytic skills in an authentic research context and enable instructors to demonstrate how statistical analysis is used to model real world phenomena. In addition, "Good data analysis requires diverse skills if they are to conduct analysis of high methodological quality. They must be able to formulate interesting research questions, select appropriate statistical techniques, conduct necessary calculations, interpret the analytic results … manner" (see [2]).

An Approach to Statistical Education for Engineers. We note that engineers are not motivated to study data not related to their job (as they want to improve the quality of production), the data is to be such that it affects their daily life. In view of the practice of statisticians to present tools without real data, engineers decided to use the data they encounter and use various statistical concepts like t and z tests and p-values which they were exposed to in their in house training in Statistics. Obviously, in view of the fact that the engineers were not trained in Statistics, other concepts like reliability of structures, design of manufacturing, dependence of quality (specification) on variables which interact were not included. The main purpose of this article is to use the data and methods experienced by manufacturing engineers at their work which contain statistical ideas. These are used to convince them that they already have used statistical ideas to work to improve quality of manufacturing. A systematic understanding of statistics experienced by them will lead to improved decision making to further the quality dynamically. We shall start by describing in detail how a manufacturing engineer encounters statistical concepts in the production facility. These ideas originated as a motivation to develop [3].

Use of Experiential Data. In a manufacturing facility we worked with, parts are delivered which are to be used in production. They were measured with a coordinate measuring machine and the data was sent to a computer. It is easily observed that the three coordinates (length, breadth and height) for each part were not the same. Yet almost all of them were used in the production. Only a few which were discarded can be seen to be outliers in the graph of the data sent to the computer. Naturally, this leads to the conclusion that the parts with slight error in the coordinates were acceptable and the error is different for different parts used. Thus one can conclude from the data that error is not constant but varies from part to part. In other words the error is random and thus cannot be measured exactly for different parts. Yet based on our experience, decision was made to accept some parts and reject others. Note that this is not a random decision. Can we learn how this is done so that we can use this to make other decisions in the manufacturing context?

Understanding the regularity in data through distribution. In order to examine the behaviour of error within the data, we look at the data from CMM calculations graphically for each co-ordinate by drawing Histogram of the data. This can be done by using simple MINITAB commands. This is a graph of the frequency table prepared with class intervals in which minimum and maximum of data points are broken and associate with each interval the number of data points fallen in it. The chance of a point falling in the interval i is the frequency corresponding to it, *fi*, divided by the total number of points n. The relative frequency histogram is obtained by drawing a graph with rectangles base of each interval corresponds to interval and the height equals the associated relative frequency. If one reduces interval size one can see that the tops of rectangles will form a curve and we are getting approximate area under the curve corresponding to each interval. Thus the part sizes can be associated with a curve. In short, there is regularity in the data.

Random Variables and Underlying Idea of Quality Maintenance. One can use the idea of Histogram to consider a curve with area under the curve being one. Then one can describe the chance of a part having measurements between two numbers on the real line as the area under the curve. In particular case of items produced with error the curve can be a bell-shaped curve with center at the specification (call it m) and variation (s) described by the spread of the bell. In production one rejects the parts with large ($2s$, $3s$) variation from m to maintain quality. Thus acceptable parts have large probability if s is small. This connects the idea of probability with decisions made in production. One can then introduce the concept of discrete random variables Y by saying that we accept the part ($Y = 1$) if it is good and reject it ($Y = 0$) if it is bad. Thus the decisions in quality maintenance have a chance associated with them.

One can then describe the situation where a lot of parts with above process are delivered to a manufacturing facility with acceptance probability p and you want to check if this lot is acceptable. One can choose a sample of n parts and accept the lot if 0 or 1 item is defective. One can then introduce Binomial random variable and operating characteristic curves (OCC) as functions of p. The Poisson can be introduced as limit of Binomial (for n large, p small). Geometric can be introduced as a waiting time for first defective in binomial and exponential as waiting time for Poisson process. We now give the outline of the topics one can cover once the students understand the concept of associating probability with events.

An Outline of Topics to be Covered.

1. Concepts of independence of events: Machine construction with redundancy to improve reliability.

2. Bayes Formula: Use of observed data to improve probability of correct decisions.

3. Evaluation of quality of received parts and supplies: Operating Characteristic Curves.

4. Meeting specifications with high probability using samples: Central Limit Theorem, Normal Random Variable and Creation of control charts.

5. Control Charts and testing hypotheses.

6. Dependence of various factors in the machinery on production specification: Simple and multiple regression.

7. Eliminating unnecessary variables in multiple regression using correlation.

8. Bayesian decision theory: Making decisions using marketing surveys. Most of the statistical data analysis done by the students will be based on MINITAB for which they will be given instructions. The data will be from the facility.

9. The output of the Minitab program would be interpreted by them for decision making.

The examination structure will be as follows. Use 9 to create projects for each of the previous topics 1–8. The students would be asked to carry out the project, interpret the results and write reports.

Acknowledgements. I want to thank my colleague Professor Vince Melfi for providing me ideas and references to Statistical Education Literature.

REFERENCES

[1] Hogg R. et al. Statistical Education for Engineers: An Initial Task Force Report. *The American Statistician*, **39** (1985), No 3, 168–175.

[2] Singer J., J. Willett. Improving the Teaching of Applied Statistics: Putting the Data Back Into Data Analysis. *The American Statistician*, **44** (1990), 223–230.

[3] Mandrekar V., K. Morris, R. L. Tummala. Statistics for engineers: Experiential Learning Approach. In: Proceedings of the Annual Conference Amer. Soc. Eng. Education, 2002. 6723–6728.

V. Mandrekar
Michigan State University
Department of Statistics and Probability
East Lansing
MI 48824, USA
e-mail: `atmah@stt.msu.edu`

Received December 4, 2014
Final Accepted February 23, 2015