The Philippine Statistician
Vol. xxvi, Nos. 1 & 2 1977

THE STATISTICIAN'S ROLE IN INDUSTRY

MANUEL PERLAS, Ph,D.*

First let me say that I welcome this opportunity to address fellow statisticians and my thanks to the Philippine Statistical Association for this honor. When Ms. Buenaventura asked me to talk on "Statistics in Industry", I pondered all possible different ways of approaching the topic — or should I say, all possible combinations and permutations. I finally decided on a very straight forward approach — a survey of statistical methodology applicable to industry, and an opinion on the statistician's role in industry. Of course, the term "statistician" includes probabilists.

Historically, statistics had been associated with numbers gathered for governments in the form of sample surveys and censuses. Many a time laymen have focused only on the descriptive aspect of statistics. Even now the notion of a statistician is equated by the layman with a lifetime devoted to taming a horde of numbers. I'm sure each and everyone here has at one time or another been confronted with the question "What do you do for a living?" to which you may have answered "I'm a statistician". Your acquaintance's response usually varies from a blank stare and "Oh, I see", to a stale joke. "The only statistics I ever like were vital statistics". After a while, they usually get back to the same question "What do you *really* do?"

Well, what do we *really* do? Most of us may be in industry in a management capacity. When was the last time we had really sat down, rolled up our sleeves, and done some honest-to-goodness statistical application. Those of us who are in teaching and consulting may be in a better position to answer that question. Or may be the question is "What *should*

* Head, System and Operative Support Division; Computer Information System, MERALCO, Philippines

we be doing?"... as statisticians in industry? In formulating an answer to this question, let us first take a brief look at the development and methodology of the field.

Statisticians, along with applied mathematicians, engineers, economist and other scientists have stimulated theoretical and practical work. For example, trying to improve the accuracy of physical measurements led the 19th-century German mathematician Karl Friedrich Gauss to develop the m ethod of least squares, which is now the bastion of many a quantitative economist and which has today evolved to a general flexible theory adapted to nearby every quantitative field. Gauss had also dabbled in mortality tables and actuarial applications, which of course is now the lifeblood of the insurance industry and an important tool in the banking industry.

For the manufacturing industry, methods for controlling and monitoring the quality of manufactured products have been essential. And to this end the statistical foundations of quality control were laid out.

The 20th century British Statistician, Sir Ronald Fisher, developed the design of experiments which is widely used to improve precision of results obtained in agriculture, medical, biological, and industrial applications. These deas of effective design for data gathering are also basic to the construction of a sample survey.

In economics, the idea of using a time series for a prediction world is not at all far-fetched. Also, regression and the analysis of variance are accepted tools of the trade.

Yet with this whole gamut of applications, the tree of influence of statistics and statisticians is but a sapling. The aforementioned applications are ripe for improvements and modifications. For instance, quality control is especially adaptable to a Bayesian statistical approach. The applicability of sequential analysis in lot acceptance has been firmly established by the works of Abraham Weld and should be entrenched as an analytical tool in industry.

As another example, let us take time series. With the advent of the computer, the sphere of influence of time series has expanded greatly. It is possible to deal with mounds of data and formulate models of time structures by means of spectral analysis and deconvolution. In a technically advanced

environment, an oil exploration geophysicist may record and interpret seismic information in the form of time series. Until recently this process was done by fixed analog methods dependent on analog recording and filtering equipment. With the versality of the digital computer, the geophysicst (hopefully with the aid of a statistician) can filter and refilter his information, alter the design of a filter in a matter of minutes by a programming change, incorporate his own ideas into the analysis and see the results immediately. In the biological sciences, the state of health of internal organisms of an individual may be monitored by equipment such as the EKG and EEG and interpreted by means of a time series approach. Thanks to the work of scientists like Wiener and Kolmogorov and also the development of the fast Fourier transform method by Cooley and Tukey, an industrial statistician may add another weapon to his data analysis arsenal — the spectral analysis of single and multi-channel time series.

Another expanding field of industrial statistical applications, more on the probability modelling side, has been elaborated upon earlier in this conference by Dr. Velasco of AIM the field of OR. In my own short experience with Computer Information System, Inc., I have encountered, and will no doubt encounter in the future, various queveing and network applications and simulation. Ranging from a performance analysis of our newly installed IBM 370/758 to optimum retrieval of information from a data base to fuel oil studies to manpower movement studies for our affiliate companies, operations research and applied probability applications are in great quantity and ripe for the taking.

Newer developments in other branches of statistical methodology have arisen and should slowly establish their roots among us and consequently we should foster their growth in industry. Reliability theory and life testing is one such field. Researchers are slowly churning out meaningful methodology concerning order statistics with applications to reliability theory. It has been found that apart from the exponential and Erlang distributions, the Weibull and the lognormal distributions may serve as more appropriate models for some lifetime distributions. For instance, the parameters of a Weibull distribution may be varied to model lifetimes with either an increasing or decreasing failure rate. Present methodology based on such distributions, however, are quite unwieldy and numerical and iterative approaches are more the rule than the exception. Computer generated tables are necessary in this type of envi-

ronment. Well, then, you might say, reliability theory applications may not be too practical. Let us open ourselves though to the fact that the normal distribution, until it is tabulated, was quite an unwieldy monster itself.

Another growing application is multivariate analysis. The concept is not new. Industrial statisticians have been dealing with multiple regression and analysis of variance before. However, the utilization of matrix algebra and geometric interpretation have lit up a new side of multivariate analysis. With the help of matrix-oriented programming languages like interative APL, multivariate statistical methods are now more within reach. One method that deserved mention, as it is frequently overlooked, is that of principal components. Often, the statistician faces a problem with a gut feeling of the variables involved, a feeling he may have derived from his or some other scientist's experience with the problem. And also quite often, neither the statistician nor the scientist have a good idea of which variables are influencing a particular problem. In both cases the method of principal components may clear the air somewhat and point to a direction of attack. A major disadvantage of the method is that it presupposes the existence of information concerning all such variables which may be begging the question. Nevertheless, the method under the right circumstances may bail us out of a real jam.

Non-parametric methods have been in used in psychological and sociological applications for some time now. However it is not hard to produce an industrial problem that may need a distribution — free approach, especially in ranking situations.

In fact, results in reliability theory have been duplicated by a non-parametric approach. However, there still exists the trade off — less constricting assumptions lead to less efficiency in results. Courses in non-parametric statistics tacitly accept this trade-off when they introduce the notions of Pitman and Bahadur efficiency and investigate the asymptotic efficiencies of various non-parametric tests.

Probably the most promising, and the most intriguing, approach to statistical applications is the subjective or Bayesian approach. Most of us are classically trained and the concept of a statistician's prior belief may lead us into a controversy whhich is still ongoing among top statisticians. However, seen in the context of the real world, the Bayesian approach and the concept of utility functions may well be the apt model.

Each statistician has his own individual beliefs and utility functions and, with the growth of acceptance of Bayesian concepts, statistical analysis may develop into a more personalized service much like the medical field where we have conservative and revolutionary approaches to the same situation. A statistician may then be conceived of as either a risk-taker or a risk-averter depending on the convexity of his utility function. Though several questions remain unanswered, some applications have gained a quick foothold due to the appropriateness of the methodology, for example, sequential analysis. A procedure that samples sequentially and formulates a stopping rule while one samples is more economically aesthetic to most enterprises.

What then is the role of the statistician in industry? In this paper, I have briefly presented some possible applications for your scrutiny. These are the tools and we are the carpenters and industry is the builder.

We are a bridge between industry and its goal towards a more rational decision-making process. And we must face the task of making this decision process not only acceptable but also palatable to management's manner of thinking. This task may require us to tread a fine line between oversimplifying and overcomplicating a given problem, a fine line between the overly theoretical and the watered-down approach. Surely we should not chisel away essential components of a problem and mold it to fit a certain set of assumptions and techniques. Yet, we must also keep in mind that an overwhelmingly elegant solution may be incomprehensible to the non-statistician.

So I suggest we temper our consultancies with moderation. And if we cannot find a happy middle road? What then? I will leave that question open for you to resolve in the light of your individual utility functions, which of course include your conscience.

In making available high-powered statistical tools to the non-rigorous client, we may obtain considerable help in the use of the computer and some standard statistical packages like the Statistical Package for the Social Sciences (SPSS) and BIOMED, a statistical library of subroutines developed at UCLA. Such packages allow the user to be fairly detached from the technical details of a problem and still invoke the use of sophisticated methodology.

Furthermore, our role is not only to make available these tools to industry but also to provide information concerning new developments. We may very well need two basic ingredients to achieve this. First, we must continuously search for increased personal development. Several statistical societies and institutes can serve as excellent vehicles for self development, among them, the IMS, the ASA, and the ORSA. The journals published by these associations provide a wealth of information in recent developments. Other sources of information are monograph series in probability and statistics published by different textbook publication companies. Of course, rather than having various individuals amassing their own private libraries, a centralized bank or library of such information is preferable and ideal and it is my hope, as well as the hope of most of you I'm sure, that such a library be established and maintained here in Metro-Manila by the PSA in conjunction with the UP Statistical Center.

The second ingredient for us to fulfill our roles in industry is acceptance by top management. Many a project, especially statistical consulting projects has died on some executive's "in" box. This problem has been faced and will be faced by us and people in like professions. Although we can never eliminate it, we may actually face such prospects in the future with a higher probability of success by means of constant statistical seminars and management training. We can attune industry and its top management to the fact that probability and statistics are not quite as far-fetched as they seem to the laymen. A more direct and basic way of gaining acceptance by top management is by showing them the costsaving advantages of a statistical approach. What could be a more convincing argument than favorable return-on-investment figures?

In closing, let me state that as individuals we are but isolated voices in the wilderness. Together, in an association such as the PSA, we may be heard as one large and clear voice and I submit that this voice foster teamwork and brotherhood among statisticians and other members of industry.

Thank you.