

INDIAN STATISTICAL INSTITUTE

TWENTIETH CONVOCATION ADDRESS

Some Thoughts on Statistics

by

T. W. Anderson

Stanford University



10th December 1985

203 BARRACKPORE TRUNK ROAD
CALCUTTA 700 035

INDIAN STATISTICAL INSTITUTE

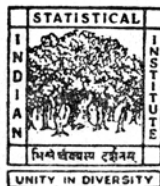
TWENTIETH CONVOCATION ADDRESS

Some Thoughts on Statistics

by

T. W. Anderson

Stanford University



10th December 1985

203 BARRACKPORE TRUNK ROAD
CALCUTTA 700 035

SOME THOUGHTS ON STATISTICS

By T. W. ANDERSON

Stanford University

Congratulations on your graduation from the Indian Statistical Institute! You are entering into careers in a discipline with a long history and at a time of yet increasing activity and interest.

Last year the Royal Statistical Society celebrated its 150th Anniversary. It was founded in 1834 as the Statistical Society of London, an outgrowth of the section on statistics of the British Association, formed the previous year. When the Statistical Society of London was given chartered status in 1887, its name was changed to the Royal Statistical Society. Its membership of 378 Fellows at the end of 1834 has grown to a membership of 4,840 Fellows in 1984, a factor of 12. Of these, somewhat more than one-quarter are overseas Fellows at the present time.

The American Statistical Association shall soon celebrate its 150th Anniversary, being founded in 1839. Its membership is now greater than 15,000 with 62 local chapters and 10 sections and subsections; these sections and subsections have all originated since I became a member in the early 1940's.

The International Statistical Institute, another I.S.I., has just concluded its centenary session in Amsterdam, the capital of its administrative base. The statistical organisation with which I have been most closely affiliated, the Institute of Mathematical Statistics, celebrated its 50th Anniversary this past summer—in fact, the week before the International Statistical Institute.

Nowhere has the progress of statistics been more striking than in Education, that aspect reflected in the present occasion. Until the end of World War II in the United States Statistics was taught in Departments of Mathematics—with some notable exceptions. At Columbia University Harold Hotelling and Abraham Wald were members of the Department

of Economics. World War II witnessed a great increase in interest and activity in statistics. Special needs, such as military operations research, the control of quality of industrial output, and allocation of manpower and resources, forced government, industry, and universities to become more aware of the many functions of statistics and allied disciplines.

In 1946 Gertrude Cox, Harold Hotelling and William Cochran initiated an Institute of Statistics in North Carolina with a Department of Mathematical Statistics at Chapel Hill and a Department of Applied Statistics at Raleigh. At the same time Abraham Wald set up the Department of Mathematical Statistics at Columbia University; it happened that Jacob Wolfowitz and I were members of the original faculty. Now nearly every major university in the United States has a Department of Statistics. Several years ago there were 67 Departments of Statistics and of Biostatistics in the United States. Unfortunately, one of those Departments has since been dismantled—namely at Princeton University.

I am not fully aware of the development of the Indian Statistical Institute under the leadership of Professor Mahalanobis, and more recently Professor Rao. In several respects the I.S.I. was probably ahead of events in the United States.

Along with the expansion of personnel and training facilities, has gone an increase in the literature of statistical theory and methodology. In 1840 the *Journal of the Statistical Society of London* published about 400 pages with 17 papers. In 1980 there appeared 100 papers on over 1200 pages in three series of the *Journal of the Royal Statistical Society*. The Institute of Mathematical Statistics in its brief existence has gone from the single *Annals of Mathematical Statistics* to the two *Annals of Statistics and Probability* and the forthcoming journal *Statistical Science*, concentrating on exposition and matters of general interest. The major expansion of the literature, however, has come about in the proliferation of journals. Some of these new journals, such as the *Canadian Journal of Statistics*, represent societies, while others, such as the *Journal of Multivariate Analysis*, the *Journal of Time Series Analysis*, *Econometric Theory*, and *Statistical Computation*, develop special aspects of our field.

Of course, this represents a part of the over-all information explosion. It has been estimated that at the present rate of increase at the end of a

life-time 97% of one's current knowledge would have been created during that lifetime.

During the century and a half that statistics has been considered an independent discipline worthy of organisation, its nature has changed and its scope broadened. The establishment of the Statistical Society of London had as its object "the collection and classification of all facts illustrative of the present condition and prospects of society," reflecting opinions of such men as Malthus and Quetelet. In time, however, the Society's business and papers were tending more towards economics, trade, and finance instead of social conditions. Nevertheless, it mainly dealt with "facts which can be stated numerically and arranged in tables." In fact, the original Prospectus of the Society said "its first and most essential rule of its conduct to exclude carefully all opinions from its transactions and publications."

The name "Statistics" is defined to be the observations necessary to the sciences of state—the Government. A report of the International Statistical Congress in Paris in 1856 stated "statistics are to Politics and to the art of governing what the observation of the stars is to astronomy."

It took about 50 years until the mathematical ideas of median, bivariate normal distribution, correlation coefficient, and standard deviation were introduced into British Statistics by Edgeworth, Galton, Karl Pearson, and others. Of course the arithmetic mean had been indispensable in the days of collection of numerical facts. There is a saying that a statistician with his head in the icebox and his feet in the fireplace would feel all right on the *average*.

As far as our English tradition of statistical influence goes, its significant beginnings took place around the turn of the century with asymptotic standard error of Pearson and the methods of maximum likelihood of Edgeworth and Fisher. Student obtained the distributions of what is now known as student's t and of the correlation coefficient under independence on the basis of some moments, calculated algebraically, some simulation, and a great deal of intuition. Fisher derived these distributions and a host of other exact distributions by mathematical methods—rigorous, but sketchy—not well understood because the proofs were usually geometric. Neyman and Pearson, Wald, Rao and others put statistical theory on a

firm mathematical foundation. Currently the theory of inference is being broadened and deepened at a great rate.

Of perhaps greater importance is the widening scope of applications. The original activity of describing social and political conditions is carried out on a large scale. Data collection has been made more accurate and more timely by use of modern organization and communications—especially electronic computers; new areas have been opened up by new methodology, such as sample surveys of various kinds. Incidentally, in the United States we have concern that the current federal administration is reducing the collection of statistics in order to economize. Someone has facetiously said “A Statistician is a person who passes as an expert on the grounds of being able to turn out with prolific fortitude vast quantities of incomprehensible numeric figures calculated with micromathematical precision from vague assumptions based on debated results taken from incomplete data carried out through the form of problematical intelligence for the avowed purpose of overwhelming a hopelessly befuddled lay person who was never interested in statistics anyway.”

Agriculture and biology have been important in statistics, both in the sense of stimulating new methodology such as the analysis of variance, and in the sense of receiving the benefit of informative statistical procedures. The rapid increase in yields of various crops would not have been possible without proper statistical analysis of the data on effects of treatments and varieties—now familiar statistical terms.

As I indicated earlier, statistics became important in the industrial sphere during World War II. Since then, that emphasis in the United States declined, but with increasing competition from abroad, particularly from Japan, new interest has been aroused. The approach is now termed “quality improvement.”

I do not have time to review many other areas in which statistics has made great progress : medicine, environment, communication, meteorology and others. My greatest interest in a field of application has been economics, especially the use of multivariate analysis in simultaneous equations models and time series analysis, including forecasting. We have not been as successful as we had hoped, but we are able to control fluctuations and avoid depressions to a considerable extent.

It might be mentioned in passing that for the average American newspaper reader statistics refers to financial markets and sports. I am reminded of the cartoon strip called "Peanuts", which is popular in the United States. In one strip the little boy who is captain of his baseball team says "In the last game we had zero runs, two hits, 25 errors and 11 walks. What shall we do about these terrible statistics?" The entire team shouts back "get a new statistician!"

Possibly the biggest impact on statistics has been made by the development of high speed computing. The acquisition, storage, and manipulation of data has been increased by orders of magnitude. Computers make feasible the application of complex statistical procedures, such as the estimation of models consisting of hundreds of simultaneous equations. It, in fact, is modifying our methodology itself, techniques that are effectively based on economic computing, such as the jackknife, the bootstrap, and cross-validation, are now of practical significance.

The statistical world you are entering today is a far cry from the one I entered about forty years ago. You have great opportunities to develop and use your knowledge.