BRITISH STATISTICS AND STATISTICIANS TODAY

By Harold Hotelling, *Stanford University*

Modern statistical theory originated in England, and is today advancing faster there than in any other country. The validity of methods in everyday use in a multitude of varied inquiries rests upon results obtained by British mathematical statisticians; for example the correlation coefficient and, much later, its interpretation accurately in terms of probability, are English discoveries. A brief survey of the present activities in the fields of most central statistical importance there may be of interest.

The oldest and best known institution for the study of statistics is the Galton Biometric Laboratory, presided over by Karl Pearson, and a part of the University of London. L. N. G. Filon, who coöperated with Pearson in the famous memoirs of the nineties which first gave probable error formulas for the correlation coefficient, is still professor of mathematics at University College, which includes the Galton Laboratory. Following the interests of its founder, the laboratory devotes much attention to heredity and to anthropometric and mental measurements, as well as to the pure theory of statistics. Its walls are lined with selections from such writers with statistical interests as Galton, Darwin, Pasteur, and Florence Nightingale, and with memorabilia, such as cartoons from the London papers, of those associated with the institution and with science in bygone years. Advanced students come here for short or long periods from all over the world, and it is in this way that such Americans as Raymond Pearl, Truman Kelley, Henry Schultz, and B. H. Camp have come to know these halls. Each student is provided with a desk and a calculating machine, and spends his entire working time in the laboratory. The staff includes Miss Ethel M. Elderton—joint author with her brother of a primer of statistics—Dr. George Stocks, a medical investigator, and Egon S. Pearson, son of Karl Pearson. The pioneer work of Karl Pearson is, of course, familiar. He is now devoting himself largely to the history of statistics. His great services to statistical theory, of which no account need be given here, are being ably continued by E. S. Pearson.

A newer but extremely active center is the statistical department of the Rothamsted Agricultural Experimental Station, at Harpenden, a village 25 miles north of London. The station was founded by Sir John Lawes to carry on the experiments which he and Gilbert had been conducting since about 1839. In some of the fields wheat has been

grown continuously under uniform treatment since 1844, providing a series of yields much more extensive than has been recorded at the younger experiment stations. It is noteworthy that institutions founded for such practical purposes as vocational training and increasing crop yields tend, as experience is accumulated, to give more and more attention to pure science. At Rothamsted many departments of biology, physics, and chemistry have been advanced. The long series of data evidently embodied valuable information, but it was so involved with confusing factors, such as varying weather and soil heterogeneity, that the existing mathematical methods of treating statistics seemed inadequate to disentangle it. To assign to each of the numerous factors its due portion of the responsibility for the variability of yields, and to judge the significance of the result so as to decide whether something really had been discovered or whether the apparently valuable agricultural conclusions were merely accidental, required in fact a number of new discoveries in pure mathematics.

Sir John Russell, the director of the Rothamsted Station, in 1919 persuaded R. A. Fisher to tackle the mathematical problems. Fisher had studied astronomy at Cambridge, had there become interested in least squares and probable errors, and had, while still a student, published a short paper setting forth the "maximum likelihood" idea which he has developed further in recent years. Leaving Cambridge he went to work for a London investment house as statistician. In 1915 he published in *Biometrika* the fundamental equation for testing the significance of correlation coefficients, obtaining it by an ingenious application of n-dimensional geometry. After coming to Rothamsted, Fisher spent several years purely on the mathematics of statistical inference, publishing his novel ideas from time to time in various journals. It was not until 1924 that his great memoir appeared, *On the Influence of Rainfall on the Yield of Wheat at Rothamsted*, in which the effect of each kind of season in conjunction with each kind of fertilizer used was definitely measured, and in which, moreover, there appeared important new ways of using regression equations, orthogonal functions, and multiple correlation coefficients, and of interpreting the results in terms of probability.

Fisher has continued in recent years to make fundamental and important contributions to statistical theory, as well as to advise the workers in other departments at Rothamsted regarding their problems. Recently he published in the *Philosophical Transactions of the Royal Society* the exact distribution in samples of the multiple correlation coefficient, a result which will supersede various corrections and probable error formulas which are now being applied as approximations.

In the same journal he has just shown how to test the significance of a periodogram when, as is always the case in practice, the most favorable period is selected and the standard error is judged from the observed data themselves. This solves a long-standing puzzle, and will make it possible to discriminate among the many and conflicting results of seekers after cycles. In the *Proceedings of the London Mathematical Society* he has just given the solution of another problem on which many people have worked with only partial success, that of the sampling errors of moments.

Fisher was joined in 1927 by John Wishart, descendant of a famous Scotch family, who had studied with Whittaker at Edinburgh, then with Pearson at London, had taught mathematics at the Imperial College of Science, and had worked with Spearman on the theory of the "tetrad differences" among correlations which are now agitating psychologists. Wishart succeeded in generalizing Fisher's paper of 1915 so as to give joint distributions of variances and covariances among several variables. J. O. Irwin has since joined the staff, and has done valuable work with problems of sampling. Miss Frances Allen of Australia, who has worked with Yule at Cambridge and is now at Rothamsted, is a promising young investigator. Students and voluntary workers, agricultural college men and Indian Civil Servants, are coming to Harpenden in increasing numbers from all parts of the world to learn the mathematics of crop experimentation and the theory of statistics.

A. L. Bowley, whose work is well known in this country, is a resident of Harpenden. He is professor of statistics in the London School of Economics and Political Science, where E. C. Rhodes is also active.

British universities have on the whole been slow to introduce statistics as a subject of instruction. The notable exceptions, besides those just mentioned, are Edinburgh, where E. T. Whittaker has developed an important statistical center, with a laboratory, in the department of mathematics, and Cambridge, which has so important a statistician as G. Udny Yule as lecturer and fellow of St. John's College. Yule is alone in his work. There is talk at Cambridge of expanding it, but what progress will be made is uncertain. Whittaker is most widely known, perhaps, on account of the *Calculus of Observations* which he and G. R. Robinson wrote, but his most important accomplishments are in pure mathematics. He has contributed extensively to the theory of functions, and I heard him give a presidential address in November, 1929, before the London Mathematical Society on a subject connected with relativity. Many actuarial students are trained in his laboratory. Outside of London, Edinburgh and Cambridge, statistics

seems to have found no place as a subject of university instruction, except as a very minor phase of economics, psychology or astronomy.

Among British actuarial statisticians several names have become familiar in addition to Whittaker's and Robinson's. Thus W. F. Sheppard, whose corrections for grouping have been used for years, is still a busy actuary, as are also David Heron and W. Palin Elderton, author of *Frequency Curves and Correlation*, and brother of Ethel M. Elderton. In a related field is Dr. Major Greenwood, professor of vital statistics in the London School of Hygiene and Tropical Medicine, who must be distinguished from his relative, Dr. Arthur Greenwood, minister of health in the Labor cabinet.

Sir Gilbert Walker, of the Imperial College of Science in London, has been working for many years over astronomical and meteorological statistics, with particular reference to cycles, and has in the process developed statistical methods of value in other fields. The same is true of David Brunt, author of the well known treatise on the *Combination of Observations*. Chiefly to study the methods of Walker and the Pearsons, Dinsmore Alter, of the University of Kansas, is in London this year.

American students of statistics have long speculated as to the identity of "Student," whose papers of 1908 and 1912 in *Biometrika* inspired R. A. Fisher, and who was the first to take the fundamental step of examining the distribution in random samples of the ratio of mean to standard error, thus opening the way for escape from the haze of hypothetical standard errors and inverse probability which have obfuscated the theory of statistics. I have heard guesses in this country identifying "Student" with Egon S. Pearson and with the Prince of Wales. He is now so well known in Great Britain that no confidence is violated in revealing that he is W. S. Gosset, a research chemist employed by a large Dublin brewery. This concern years ago adopted a rule forbidding its chemists to publish their findings. Gosset pleaded that his mathematical and philosophical conclusions were of no possible practical use to competing brewers, and finally was allowed to publish them, but under a pseudonym, to avoid difficulties with the rest of the staff. Following his example "Sophister" and "Mathetes," younger chemists employed by the same brewery, have published contributions to the theory of statistics. This same firm has made large grants to the Rothamsted station for research on barley, which incidentally has stimulated the work in statistics.

British books on statistics are few and of high quality. Frequently the educated Britisher, like the educated continental European, has gone through an elementary course in calculus before reaching the age of

eighteen, and has no use for long-winded discourses which get nowhere in order to avoid elementary mathematics. Thus even those English books on statistics written by economists for economists, such as the treatises of Bowley and D. Caradog Jones, and of course such works as those of Brunt, Elderton, and Whittaker and Robinson, use the calculus, as well as the kind of algebra commonly taught to college freshmen in this country but taught in Europe at a much earlier stage.

British economists and economic statisticians are in general too well known in this country to need discussion here, but mention should be made of J. R. Stevenson and Arthur Newsholme, who have written on population and therefore used statistics, of J. M. Keynes and P. Sargent Florence. Apart from Yule and Bowley, British economists have not shown any strong tendency to introduce new statistical methods.

Facilities for publication of important results in statistical theory have always been somewhat limited, though *Biometrika* publishes some work in addition to that of the Galton Laboratory, and the *Journal of the Royal Statistical Society* contains many short notes and a few longer papers of a theoretical nature. The pressure has of late been somewhat relieved by sending papers abroad and by the more hospitable attitude of the mathematical and semi-mathematical journals to papers of a statistical nature. However, a good deal of important work is greatly delayed in publication. It is partly due to this fact that it has been so little understood in other countries. The Scandinavians have indeed always followed it, but on account of the language difficulty nobody else could follow the Scandinavians. Of late the English work has received some attention in France in the treatises of Darmois, Borel, and others.

There is an active interest in unsolved problems in statistical theory in England. For example the problem of generalizing "Student's" distribution to non-normal populations was the subject of a lively correspondence in the semi-popular journal *Nature* last autumn. This and related problems had been brought to the front in two papers by E. S. Pearson and J. S. Neyman, a Pole. The general level of activity in the development of statistics is high, throughout the country, and the results are proving valuable.