4
Teaching of Statistics

Fisher to D.J. Finney: 27 June 1957

For further guidance I am wanting to say distinctly that opportunities for research are the attraction.[1]

As you say, I should have no difficulty at Caius or at Rothamsted of going on with the work, in which more and more extensive fields continually open up, of pointing out faulty and misleading methods now taught to students. There is surely a lifetime's work in front of me there, but it is less agreeable than doing a bit towards clarifying the situation in one of the natural sciences, where the application of statistical methods is not bogged down with all the pedantic verbiage that seems to be necessary in mathematical departments, where it seems to amount to no more than finding excuses for replacing known competent methods by less adequate ones. If you want to forget that a test of significance is liable to be invalid if based on inefficient estimation, it seems to be sufficient to speak of it as a non-parametric test.

[1] Fisher was considering where and with whom he might work after his retirement from the Chair of Genetics, University of Cambridge, in 1957.

Fisher to H. Hotelling: 22 February 1946

. . . One side of your problem,[1] which I have no doubt has not escaped your attention, is the difficulty of combining in the same man, mathematical skill with the kind of width of logical outlook only, I think, easily acquired by long contact with the atmosphere of experimental research. In statistical methods, as it seems to me, we are developing the mathematics appropriate to a logical attitude which has not in formal pure mathematics ever found coherent expression, and personally I have often felt exasperated at the obtuseness of perfectly good mathematicians when faced with any logical situation involving gaining new knowledge by experience or experimental observation. I am sure that this is simply because the logical presuppositions with which they have for years been familiar in deductive analysis are by this time taken for granted, not only as axiomatic, but as universally sufficient. . . .

[1] Hotelling had sought Fisher's help in obtaining suitable personnel for the new Department of Mathematical Statistics at the University of North Carolina.

Fisher to T. Koopmans: 14 *January* 1936

I have at last been able to look through your interesting lectures on Modern Sampling Theory. I have annotated the typescript with a few suggested amendments, mostly quite trifling.

The main point which I think deserves consideration is that of order of development and presentation. After an introduction to the general notions of population and sample, parameter and statistic, specification, estimation and distribution, making the point that we must judge of the value of different possible methods of estimation by means of the distribution of the estimates, I believe it would be wise to give some considerable amount of time to problems of distribution, which, though full of mathematical interest, have the advantage of being purely deductive, and of involving only the classical concept of probability representing an ideal frequency.

I think at this stage I should illustrate three methods by which exact sampling distributions may be obtained: the method of generalized space, of the use of which the solution of the distribution of the correlation coefficient is a good example, with, possibly, extension to the multiple correlation coefficient. The procedure here developed includes, as by-products, the simultaneous distribution of the mean and variance of a single normal variate. Another good example of the geometrical method is that of the test of significance for the largest Fourier component in harmonic analysis. The two other methods which should be illustrated are the use of mathematical induction for samples of increasing size, and the use of the characteristic function. Of these, the former works well for demonstrating the χ^2 distribution for the sums of squares of independent normal deviates having the same variance, and leads naturally to the three principal distributions, χ^2, t and z, used in tests of significance. The use of the characteristic function, though beautiful in its generality, has not often led to new distributions which have not been found in other ways. There are, I think, some examples in my paper 'Two new properties' [*CP* 108]. A special case of importance, with a very direct method of solution, is that of an estimated percentile, such as a median of an odd number of observations. Under the heading of 'problems of distribution' also should be included the approximate large sample formulae for the variance of any chosen statistic, best developed perhaps from the general multinomial distribution.

On this groundwork of knowledge of the distributions of estimates the criteria which I have developed are easily intelligible. As an illustration of an inconsistent statistic I should not use one which tends to no limit at all, but rather one which tends to the wrong limit, as, for example, in using Sheppard's correction with grouped normal data. If a statistic tends to no limit at all, I should not know whether it were consistent or not. The notion of efficiency can be introduced with easy examples, such as the median of a normal sample, and is consolidated by proving that there is a lower limit to the limiting value of nV as the sample tends to infinity, i.e. that such a limit exists whenever the distribution of the statistic tends to normality. This proof needs supplementing by the proof that it is possible to find a statistic having this limiting value for nV, as can be satisfactorily done by considering all statistics which are solutions of equations linear in the frequencies and showing that, when the variance is minimised, we are led to the equation of maximum likelihood with a limiting variance as small as possible. Consequently, the equation of maximum likelihood leads to one efficient estimate. The maximum precision obtainable in large sample estimation leads, by one of the most important steps of the whole theory, to the concept of the intrinsic accuracy of an error curve definable for all curves for which it has a finite value, e.g. for 'Student's' distribution when $n=1$, and to the recognition of the intrinsic accuracy of the error curve of a statistic as the amount of information which that statistic extracts from the data. The proof of the existence of a maximum precision is then easily adapted to show that estimation may involve a loss, but cannot introduce a gain of information, and that the loss can only be zero when the sets of sample for which $d(\log L)/d\theta$ is constant are the same for all values of θ. This is the essential property of sufficient statistics, and from it the fact that no additional information can be supplied by other statistics can be demonstrated. It should also be shown at this stage that, in general, a finite amount of information is lost by maximum likelihood estimates, but that, in the limit for large samples, the loss is less for such estimates than for other efficient statistics. At this stage also I should show that the loss of information may be further reduced by the use, in conjunction with the estimate, of one or more ancillary statistics which themselves provide no information, and that in special and easy cases, as in location and scaling, the whole of the information may be recovered in this way.

Turning now to tests of significance, I should show that when there is a variable parameter any test of significance, whatever level of significance be chosen, will generally reject some of the possible values. If the parametric values form a continuum, a portion may be defined for each test of significance, within which the values are not rejected, but outside which all values are rejected, the probability of rejecting the right value being known with exactitude. If, therefore, the test is based upon statistics which contain the whole of the information supplied by the sample, we may make a probability statement respecting the unknown parameter which is entirely well defined. Its logical content differs entirely from a statement of inverse probability. It is independent of all prior knowledge respecting the frequency distribution of the parameter. To distinguish it, it is called a statement of fiducial probability. The aggregate of such statements possible define the fiducial frequency distribution of the parameter as inferred from the observations. Such statements derived from a portion only of the observations would necessarily be erroneous, as would be, for the same reason, statements of fiducial probability based on non-sufficient estimates. No such fiducial distribution is available from discontinuous observations, though the observational programme may sometimes be modified deliberately to lead to conclusions of this type. It is, in my view, most important that the limited applicability of the notion of fiducial probability be stressed.

I am returning your typescript herewith.

Fisher to J. Maclean: 3 March 1930

. . . I ought to say that it is undoubtedly the practitioners rather than the teachers who are absorbing the book [*SMRW*]. I can understand this as since I wrote it I have had to do some lecturing and examining in the subject, and am now convinced that, apart from really first class mathematicians, statistics should be taught by the direct handling of numerical data in the laboratory.

Fisher to P.C. Mahalanobis: 9 October 1954

. . . I think the arrangement for the seminar on statistical inference would suit me perfectly, but though I may have to appear sometimes before a large crowd, I have no illusions as to how little teaching can be effected in this way. In a large crowd with important names present, young and unknown men who might really have something to gain are completely inhibited. If, however, you have anyone with a keen mind among your younger students, I hope you will not hesitate to encourage him to set his wits to work and make of my seminar a genuine discussion rather than a monologue by a super-Pundit whose word may not be doubted.

The world has enough dictators anyway. . . .

Fisher to K. Mather: 12 March 1943

. . . I should like to impose as a necessary condition for being accredited as competent in business statistics the ability to test the significance of a 2 × 2 table. After all, when business men can do that, it may spread anywhere — into the Insurance Offices or even into the higher ranks of the Civil Service. . . .

Fisher to H.F. Smith: 29 May 1946

. . . About Indian and indeed most foreign mathematicians, I seem to feel much as you do that the choice between what has to be elaborately explained and what can be assumed to be spontaneously evident differs a great deal in different schools of mathematical thought. I recently received a very highbrow treatise of a first class Swede, namely Cramer, purporting to deal with mathematical statistics, of which I think a full half was a comprehensive introduction to a theory of point sets, which I should not feel spontaneously had any real bearing in reasoning from frequencies or the inter-relation of distribution functions, but I suppose in view of the didactic manner in which the theory of probability is approached in Russia, for example, and in France, and in recent years in the United States, this sort of thing must seem a quite essential clarification of the subject.

In Paris recently I found an interesting and perhaps useful distinction being made between *statisticians* and *probabilists*, broadly speaking putting me in the first class and Cramer in the second. I also learned that a certain, I believe very learned, Russian called Kolmogoroff has postulated an axiom, which purports to justify axiomatically certain types of argument for which I am responsible. It runs something like this. If the probability of an event dependent on a number of parameters, $\theta_1, \theta_2, \ldots \theta_s$, has one and the same value ω for all possible combinations of the parametric values, then the absolute probability of the event exists and is equal to ω. So now when I am in any difficulty I can insert with confidence 'by Kolmogoroff's axiom'. . . .

Fisher to H.C.S. Thom: 8 October 1941

I have just received your note of August 30th about the new 8th edition of *Statistical Methods*. The edition had, however, been completed before I saw your letter, so that my consideration of its suggestion has been what is called 'academic'.

I think there are advantages in the idea,[1] and really only one serious disadvantage. Your reader wants, perhaps, a concise little elucidation of the mathematical connection stated and assumed in the text. Then he wants, and may expect to be referred to, a technically competent source for this particular mathematical result, from which he can easily pick out exactly what he wants. His interests, in fact, are centripetal. Having the haystick he wants to find the needle.

Now the chief use for a student of looking up papers, either in my own bibliography or in the general list of sources, is centrifugal. When a man is ready to widen his association with a particular method, to consider more fully some of the difficulties of its particular application, and the marginal applications in which its appropriateness is doubtful, or if he is inclined to browse in the history of the subject and detect how this idea led to that, I think that foundation papers may make useful reading; but they are nearly all mixed in respect of theory and practice, theoretical statements being put in *ad hoc,* when they are immediately needed, perhaps on discovering that the point had not previously been made explicitly clear, and by no means in the orderly succession of a connected theoretical treatise.

In my view a statistician ought to strive above all to acquire versatility and resourcefulness, based on a repertoire of tried procedures, always keeping aware that the next case he wants to deal with seriously may not fit any particular recipe. Of course I know that my book has been described as though it supplied such cut and dried recipes, and has been criticised for not supplying the mathematical background by which they can be better understood; but you will see from the above that, from my point of view, this is a misapprehension, based on the belief that the understanding required can be obtained from the mathematical background rather than, as I think, from the particular peculiarities of the actual body of data to be examined. This kind of understanding seems to be ignored to a frightful extent in mathematical discussions.

[1] Thom had suggested that if Fisher's papers listed in the bibliography of the end of *SMRW* had specific references to them inserted at relevant places in the book, mathematically inclined readers would be greatly helped and the book would then be a useful reference for the applied mathematician interested in theoretical developments.

Fisher to G.T. Walker: 27 *February* 1942

. . . I need not say how glad I am that you have been looking at some of my own work. I have resisted all temptation so far to overload my book on Statistical Methods with mathematical demonstrations for I should far rather that most workers should give their minds to thinking of the nature and origin of the data before them in relation to the appropriateness of any contemplated procedure than to what is usually rather irrelevant, the algebraic manipulation by which results are formally demonstrated. Consequently I have usually given formal analysis, not in this book but in other papers, such as the enclosed, written from time to time as the occasion seemed to require. . . .

Click here for next section