RECENT ADVANCES IN MATHEMATICAL STATISTICS, II[1]

By Cecil C. Craig

University of Michigan

The statistical theory of the linear relationship between a dependent variable x_1, and a set of independent variables x_2, x_3, $\cdots$, x_{t+1}, is by now quite generally understood. Supposing that the x_i's are measured from their respective means, we determine the coefficients, b_2, b_3, $\cdots$, b_{t+1}, in such a way as to maximize the coefficient of correlation $r_{1.23\ldots t+1}$ between x_1 and $\sum_{i=2} b_i x_i$. This coefficient of correlation, usually called the multiple correlation coefficient, measures the exactness of the linear relationship that exists, and it has the property of being quite unchanged if the origins or the scales for the separate x_i's are changed in any way or even if the set x_2, x_3, $\cdots$, x_{t+1} should be replaced by any equivalent set of linear combinations of them. That is, e.g., if $t = 3$, the new variables, $v_2 = x_2 + x_3 + x_4$, $v_3 = 2x_1 - x_3 + 3x_4$, $v_4 = x_1 + 2x_3 - 2x_4$ are equivalent to x_2, x_3, x_4, since the latter can be found if the v_i's are known, and the multiple correlation between x_1 and the v_i's is exactly the same as that between x_1 and x_2, x_3, x_4. Moreover, the requisite sampling theory if the variables involved are normally distributed is well established.

I want to discuss briefly an important generalization of this kind of situation that has been the subject of recent research. In particular, in his paper, "Relations between two sets of variables," published in *Biometrika* in 1936 [1] H. Hotelling set forth these ideas in excellent fashion and contributed much to the mathematical theory required for their practical application. We now suppose that we have two sets of measurements, x_1, $\cdots$, x_s, and x_{s+1}, $\cdots$, x_{s+t}, made on the same object and that we are interested in the linear relations that may exist between the members of one set and the members of the other. As an example, x_1, $\cdots$, x_s might be the prices of s more or less related commodities at a given time, and x_{s+1}, $\cdots$, x_{s+t} measures of factors which may be thought to be effective in the price situation.

In the more special case I began with, $s = 1$, and a single equation fully expressed the linear statistical relationship of x_1 with x_2, $\cdots$, x_{t+1}. Now there are s dependent variables and now with $s \leq t$, not one but s distinct linear relations will exist and will be required to fully describe the linear connections between the two sets of variables. We may assume that there is no mere duplication among the variables we are using, i.e., no one of the s x_i's is always exactly given by a linear combination of the others in the set and the same is

---

also true of the set $x_{s+1}, \cdots, x_{s+t}$. Now there is no logical or mathematical necessity for the way in which we are so far using our measurements. Suppose $s = 2$ and $t = 3$. We can find the best linear regression equation for x_1 on x_3, x_4, x_5 and then find the like equation for x_2 on x_3, x_4, x_5. But we could very possibly get more meaning out of the situation if we began by replacing x_1 and x_2 by, say, $u_1 = x_1 + x_2$ and $u_2 = x_1 - x_2$ and similarly replacing x_3, x_4, x_5 by three v_i's formed from these three x's in a similar fashion. We have really been making a quite arbitrary choice among the u's and v's that could be used and the question presents itself: What significance is there in the way we choose our u's and v's?

It turns out to be much more than a merely reasonable beginning to try to determine a u from the first set and a v from the second in such a way that they will be more closely correlated than any other u and v formed in this linear fashion from the s x's in the first set and the t x's in the second. That is, we set,

$$u = \sum_{\alpha=1}^{s} a_\alpha x_\alpha \quad \text{and} \quad v = \sum_{i=s+1}^{s+t} b_i x_i,$$

and determine the a_α's and the b_i's which will maximize r_{uv}. We may say that this u and v will account for more of the linear dependence of $x_1, \cdots, x_s$ upon $x_{s+1}, \cdots, x_{s+t}$ than will any other u and v. To the mathematician familiar things begin to appear, though, as Hotelling remarks, in its purely mathematical form this problem seems to be new. A very important observation is the fact that this maximum r_{uv} would be quite unaffected by any change in origin or scale on any of the x's; it is even unaffected if we should begin by replacing the first s x's by any equivalent set of s linear combinations of them as new variables to work with and by doing the same thing on the second set of t x's. Hotelling makes use of this circumstance to greatly simplify his mathematical developments.

Now things fall out in a very interesting way. One actually solves not for the a's and b's at first but instead for the maximized r_{uv}. Having this the corresponding a's and b's can then be found. But generally the equation for r_{uv} gives not one but s different values for r_{uv}! What is the meaning of the s different r_{uv}'s? Well, you remember that I said that s relations ($s \leqq t$) would appear to exist between the two sets of variables. These s r_{uv}'s correspond to those s linear relations which are picked out in a unique way. We now have s u, v pairs which are independent of each other in the sense that no u or v is correlated with any other u or v with the exception of the other member of its pair, and of course this correlation is precisely the r_{uv} by which the pair was determined. Further, the largest r_{uv} gives the maximum u and v we set out to find; the second largest r_{uv} determines the pair u, v of maximum correlation among those independent, in the sense just described, of the first pair; the third largest r_{uv} leads to the u, v of maximum correlation among those independent of the first two pairs, and so on. The s independent linear relations among them completely describe the linear statistical dependence of the one set of variables upon the other. The relations are essentially those between the u, v pairs and

the closeness of these are measured by r_1, r_2, $\cdots$, r_s, which I write for the s r_{uv}'s. The new variables are called canonical variables and the correlations between them canonical correlations. We may say that the maximum pair, u, v, gives both the best linear predictor that can be formed from x_{s+1}, $\cdots$, x_{s+t} and also the linear combination of x_1, $\cdots$, x_s that can be best predicted.

I have to try to deal briefly with the numerous ideas and results in this paper which is not unrelated to earlier work by the author and by S. S. Wilks. First, what about an over all measure of the linear connection between the two sets of variables? It is shown that

$$q = \pm r_1 r_2 \cdots r_s \quad \text{and} \quad z = (1 - r_1^2)(1 - r_2^2) \cdots (1 - r_s^2),$$

have properties that make it appropriate to call the first the (vector) correlation coefficient between the two sets and the second the coefficient of alienation. Both are simply expressed by means of determinants of the covariances (product moments) among the st x's. For example, if $s = 1$, q is simply $r_{1\cdot2\,3\ldots t+1}$. If $s = t = 2$,

$$q = \frac{r_{13}r_{24} - r_{14}r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{34}^2)}},$$

the numerator of which is the tetrad difference of the psychologists. Further, if it should happen that x_2 and x_4 are identical, this q becomes $r_{13\cdot2}$.

In an application, of course, the various quantities appearing above will have to be calculated from an observed set of values of x_1, $\cdots$, x_s, x_{s+1}, $\cdots$, x_{s+t}. Hotelling adapts an iterative process he had previously given to calculating the canonical r_1, $\cdots$, r_s, from which the canonical variables can be found, and he numerically illustrates the whole procedure. But what is more difficult is to solve the sampling problems that arise. It is very helpful to assume that all the x's obey a multiple normal frequency law.

First, Hotelling derives expressions for the standard errors of the r's and of q and z which are approximations useful for large samples. But for small samples exact sampling distributions are needed. Wilks [2] had earlier studied the exact sampling distribution of z in the case in which we are interested, that in the population the set x_1, $\cdots$, x_s, is completely independent of the set x_{s+1}, $\cdots$, x_{s+t}, though he did not leave his general result in a form suitable for calculation. Hotelling now finds the distribution function for q for $s = 2$. The result is not in all cases simple in form but numerical values can be obtained from it. The relations between these two possible tests, one based on z and the other based on q, are discussed at length.

An obvious undertaking would be to try to find the exact joint sampling distribution of the canonical correlations for any s and t, and I will say something about the very interesting papers in which this problem was solved. But some of this later work arose in a different though related setting which I want to discuss briefly first.

In 1936 R. A. Fisher published "The use of multiple measurements in taxonomic problems," [3] which was the introduction of linear discriminant functions

to the statistical world. Suppose that N_1 random individuals of one race (species, variety, etc.) have been measured with respect to each of k characteristics and that N_2 random individuals of another race have been similarly measured. What linear combination of these measurements would serve best to distinguish members of one race from those of the other? An example used by Fisher in this paper was that of two samples of 50 plants each of two varieties of iris found growing together in the same colony. In the flower on each plant there was measured the sepal length, x_1, the sepal width, x_2, the petal length, x_3, and the petal width, x_4. What linear function,

$$X = \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 + \lambda_4 x_4,$$

would enable one to most surely identify the variety to which each single plant belongs? To choose such an X Fisher proposed the mathematical principle that the coefficients, λ_i, $i = 1, 2, 3, 4$, be determined so that the difference in the average value of X in the one variety and the average value in the other divided by the sum of squares of the X's taken about the two group means shall be a maximum. Then quite simple mathematics leads to the required numerical values of the λ_i's.

But now that we have set up such an instrument as X, there is a more interesting use to which it can be put. Suppose that the question were to establish that the N_1 individuals from the one group and the N_2 individuals from the other really belong to different races distinguishable with respect to the complex of characters we have chosen to measure in each. We are on the old question of racial likeness or unlikeness and obviously the word "race" may have a meaning broad enough to give this work of Fisher's wide application indeed. Subject to the principle according to which the coefficients λ_i are determined from sample sets of measurements, X is the best possible linear discriminant function. We are now faced with the question of the statistical significance of the difference between the means of X for each group compared to the above mentioned internal sum of squares.

It is generally useful and enlightening in a problem of this general nature turning on the use of linear and quadratic forms to consider its interpretation as an analysis of variance or covariance. Fisher readily provides such a set-up in this case by assigning to the quality of belonging to race A a numerical value, y_1, the same for all members of that race, and by assigning in like fashion a different numerical value, y_2, to the quality of belonging to race B. It is mathematically convenient if we have samples of N_1 and N_2 from races A and B respectively, to let

$$y_1 = \frac{N_2}{N_1 + N_2} \quad \text{and} \quad y_2 = -\frac{N_1}{N_1 + N_2},$$

for then over the combined sample of $N_1 + N_2$, we have,

$$S(y) = 0 \quad \text{and} \quad S(y^2) = \frac{N_1 N_2}{N_1 + N_2}.$$

This may seem somewhat arbitrary at first glance, but let us start anew by writing the linear regression equation,

$$y = \sum_{i=1}^{k} b_i(x_i - \bar{x}_i),$$

in which y takes on one of the two values above and in which $\bar{x}_i$ is the mean of x_i in the combined sample, and then proceeding to determine the b_i's in the usual least squares fashion. The b_i's turn out to be proportional to the λ_i's previously found. Now the total variance of the y's is analyzed into that within groups and that between groups and it is immediately suggested that the usual z-test with k and $N - k - 1$ degrees of freedom is the appropriate one. But, as Fisher remarks, ordinarily for the application of this test one postulates a population in which the y's have a normal distribution for each fixed set of values of $x_1, x_2, \cdots, x_k$. Here, however, the y remains fixed and one postulates a normal distribution of the x's associated with a given value of y. Not to leave this matter in doubt, though I shall return to it, I may remark that Fisher noted that earlier work by Hotelling [4] showed that the z-test is nevertheless the proper one to use.

I have to be brief indeed concerning linear discriminant functions. Fisher wrote further papers dealing with them in 1938 [5], 1939 [6], and 1940 [7] and among others, Mahalanobis [8], Bose [9, 10], and Roy [10], of the "Calcutta School" have made relevant contributions. In particular, Mahalanobis [8] introduced the concept of the generalized distance by which two sets of multiple measurements differ, which has an obvious connection with the present subject. Fisher also discussed a test for the direction in k-space in which two such samples differ most and in case we have three such samples from three different races provided a test for their collinearity.

In his 1939 paper mentioned above [6], Fisher called attention to the connections between the theory of linear discriminant functions and Hotelling's canonical correlations. Of course it can be said at once that a linear discriminant function arises as the very special case of investigating the linear relationship between the artificially introduced y and $x_1, x_2, \cdots, x_k$. And the test of significance based on the analysis of variance turns on the ratio of the sum of squares due to regression, i.e., among the predicted values, to the total sum of squares for the regression and for the residuals. This analysis is quite general in form and can equally well be set up if one is predicting linear forms formed from N_1 variables from linear forms made up from N_2 other variables. If one sets up the condition that this ratio, ϑ, be a maximum one is led, as Fisher shows, to a determinantal equation in ϑ, the roots of which are the squares of Hotelling's canonical correlations.

Mathematically the general problem we are interested in is equivalent to the following: We have a sample of $N_1 + N_2$ observed values of p normally distributed variables. If a_{ij} is the covariance of the i-th and j-th variables in the sample of N_1 and b_{ij} the like covariance in the sample of N_2 we want the sampling distribution of the roots of the determinantal equation:

$$| a_{ij} - \vartheta(a_{ij} + b_{ij}) | = 0,$$

under the hypothesis that the first sample is independent of the second. This problem Fisher solved in his 1939 paper though in his characteristically concise and intuitive manner. But in the same number of the *Annals of Eugenics*, P. L. Hsu [11], at Fisher's suggestion, gave a complete analytical solution. Hsu also showed more in detail how the result applies to Hotelling's case of N observations on $s + t$ normally distributed variables in which the set of s is independent of the second set of t. In his 1936 paper Hotelling gave the result for $s = t = 2$ and in 1939, Girschick [12] gave the solution for $s = 2$ and $t > 2$. Hsu showed, too, the striking fact, mentioned by Fisher, that it is sufficient that only one of the two sets of s and of t variables be normally distributed in order that the distribution function found apply. This provides the explanation of why the test of significance applied by Fisher for linear discriminant functions is valid even though the y introduced had an arbitrary distribution of values.

The simultaneous distribution of the canonical correlations is fundamental but on finding it not all difficulties are thereby resolved. As mentioned above, either of the quantities, z or q, as they appear in Hotelling's paper, furnish over all tests, or rather they would if their distribution functions were obtained in a satisfactory form. The form of the distribution of z for complete independence was given by Wilks as early as 1932 [2] but that of q for $s > 2$ is still lacking. For $s > 2$ there are difficulties in applications even with z and in 1938 [13] M. S. Bartlett proposed a more convenient approximate test. Ordinarily, however, one would want to test the largest canonical correlation alone for significance. There are two kinds of trouble here. First, there is no assurance that the largest observed canonical correlation corresponds to the largest one in the population. Second, it is quite important to know whether the remaining population correlations are zero or not. Bartlett in 1941 [14] discussed these points.

Now I make an abrupt change in subject. Some interesting work has been done on the theory of runs and its applications during the last five years.

First, I want to try to convey some idea of the contents of three papers by W. D. Kermack and A. G. McKendrick published in 1937 [15, 16] and 1938 [17]. Suppose we have an unlimited set of numbers, no two of which are equal, and start drawing from them at random, recording the numbers in sequence as they come. Within the sequence drawn there will occur runs up and runs down of varying lengths. Thus in the sequence of 10 numbers, 2, 5, 11, 8, 9, 4, 3, 7, 14, 12, there are 3 runs up, one of length 2 and 2 of length 3, and 3 runs down, 2 of length 2 and one of length 3. Both ends of a run are counted in finding its length; no run can have a length less than 2. The total number of runs is 6 of which 3 are of length 2 and 3 are of length 3. We can also count the *gaps* which extend from crest to crest or from trough to trough and note their lengths with the convention that again both ends are counted in determining a length, so that no gap length is less than 3. Thus in the sequence of 10 numbers above there is one gap of length 3, 3 of length 4, and one of length 5.

It is clear that if we know the distribution for runs or for gaps of different

lengths we can compare an observed sequence, or rather an observed distribution of runs or gaps by lengths, with the frequencies calculated on the hypothesis of randomness and be by way of acquiring a test for this hypothesis. To be brief, in these papers these theoretical distributions are found together with their means and variances. There are some interesting applications. Tippett's random sampling numbers and a series of reversed telephone numbers both passed the χ^2-test as random and also passed the test based on the departure of the mean from its expected value compared with its standard deviation. On the other hand, the series of Swedish death rates for the period 1740–1930 could not conceivably be random. This investigation was prompted in the first place by the fluctuations of the death rate from ectromelia in mice in an experimentally induced epidemic.

The problems here dealt with had been only partially solved by earlier writers. There is much interesting material in these papers I have no space for. The authors readily include the case in which the numbers composing the population are not all different. They also studied series of limited length, series arranged in a cycle or ring and even what may be termed a Möbius cycle.

A. M. Mood in 1940 [18] in an interesting paper investigated a different form of the problem of runs. Suppose we have n elements of two kinds, say n_1 a's and $n_2 = n - n_1$ b's, and that these are arranged at random in a row. For example, if $n_1 = 5$ and $n_2 = 7$, and if a random arrangement of the 12 a's and b's is $babbbabbaaab$, the a's occur in 2 runs of one and in one run of 3 and the b's come in 2 runs of one, in one run of 2 and in one run of 3. If r_{ij} $(i = 1, 2)$ is the number of runs of j of elements of variety i, Mood finds the probability of obtaining a given set of values of r_{ij} such that $\sum_j j r_{ij} = n_i$ $(i = 1, 2)$, i.e., of obtaining a given pattern of runs in the two kinds of objects. Besides this basic distribution function he obtains certain marginal distributions such as that for the occurrence of a given set of runs in the a's regardless of how the b's fall (except that they must provide the necessary points of division), or that for r_1 and r_2 if these are respectively the total number of runs of a's and of b's, or that for r_1 or r_2 alone. He finds the factorial moments of these variables and then their means, variances and covariances. Similar results are obtained in case there are more than two kinds of elements. In the second part of the paper, Mood turns to the case of drawings from an infinite population in which articles of two or more kinds occur in fixed proportions. Finally, in both of the two kinds of drawings considered he derives the limiting forms of the distributions studied as the sample size increases. As Mood notes, here, too, a few of the results had previously been found, but this paper is the first really thoroughgoing investigation of its subject.

In a paper antedating Mood's by some six months, A. Wald and J. Wolfowitz [19] used the distribution function for the total number of runs (irrespective of length) for arrangements of fixed numbers of two kinds of elements to provide a test of the hypothesis that two samples have come from the same population with a continuous distribution law. If the observations in the two samples

combined are arranged in order of magnitude and if then the observations from the first sample are each replaced by a zero and those from the second are each replaced by a one, we have a situation to which this distribution function for runs applies. W. L. Stevens in 1939 [20] also discussed an application of this distribution.

The third principal topic I have chosen for my remarks is developments in the use of the probability integral transformation. The use of this device at all seems to be quite recent, appearing in a paper by H. Cramer in 1928 [21] who invented a test of goodness of fit which reappeared as the "ω^2-test" in apparently independent work of R. von Mises in 1931 [22]. In 1932 in a section new in the fourth edition of "Statistical Methods for Research Workers," [23] Fisher showed the usefulness of this transformation in combining independent tests of significance and in 1933 and 1934 Karl Pearson [24, 25] had papers in *Biometrika* on the subject.

As for the transformation itself, suppose that $p(x)$ is the probability density function of a continuous variable x defined on the range (a, b) such that,

$$\int_a^b p(x)\, dx = 1.$$

Then let us introduce the variable,

$$y = \int_a^x p(x)\, dx,$$

which is the probability that a value of the variable at random will be less than x. It will be seen that since x is a random variable, the proportion of population values less than an x drawn at random is itself a random variable. Perhaps this will be clearer if I use a simple example of J. Neyman's to show how a sample of x's also determines a sample of y's for a given $p(x)$. Suppose that,

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2},$$

and that a sample of 5 values of x arranged in order of magnitude is: -1.5, -1.1, -0.5, 0.6, 1.6. Then by reference to a table of areas under the normal curve of error, we find that the corresponding observed y's are: 0.067, 0.136, 0.309, 0.726, 0.945. It is obvious that the range for y is always, for any $p(x)$, $(0, 1)$. Further if $f(y)$ is the probability density function for y, of course,

$$f(y)\, dy = p(x)\, dx.$$

But from the definition of y,

$$dy = p(x)\, dx,$$

so that $f(y) = 1$. Thus, quite independently of $p(x)$, y obeys a rectangular distribution law on the range $(0, 1)$.

This simplicity of the distribution of the quantity y and its independence of

$p(x)$ are most attractive properties. I shall note briefly some of the applications that have been made in recent years.

In 1936 W. R. Thompson [26] denoted by p_k the probability that in a sample of N a randomly chosen x will be less than x_k, the k-th value observed. Then the probability that $p' \leq p_k \leq p''$ is just $p'' - p'$. The probability that exactly r other members of the sample will be less than x_k is then,

$$\binom{N-1}{r} p_k^r (1 - p_k)^{N-1-r}.$$

Further for all samples in which just r values occur less than x_k, the proportion of occasions on which $p' \leq p_k \leq p''$ is given by

$$\int_{p'}^{p''} p^r (1 - p)^{N-r-1} dp \Big/ \beta(r + 1, N - r),$$

the difference of two incomplete β-functions. But that there are exactly r observed x's less than x_k is equivalent to saying that x_k is the $(r + 1)$-st observation in order of magnitude, so that in the above we may as well replace r by $k - 1$. It is easy to find that the expected value of p_k in such samples is $\dfrac{k}{N + 1}$ and that the variance is $\dfrac{k(N - k + 1)}{(N + 1)^2(N + 2)}$. It follows from the first of these two expressions that the proportion of occasions on which $x_k < x < x_{N-k+1}$ is $\dfrac{k + 1 - 2k}{N + 1}$, $(N + 1 > 2k)$. Statements of this kind establish confidence limits. Thus if one says that in a sample of N, an observation at random will fall between the k-th and the $(N - k + 1)$-st observations in order of magnitude, such a statement has a probability of $\dfrac{N + 1 - 2k}{N + 1}$ of being true. Or, the integral just above is the fiducial probability of the truth of $p' \leq p_k \leq p''$ if in a sample of N the k-th observation is the $(r + 1)$-st in order of magnitude. Thompson went on to obtain confidence limits for the median in a sample of N from any population.

In 1939 Wald and Wolfowitz [27] studied the problem of obtaining confidence limits for $\varphi(x)$, the proportion of observations in a sample of N with values less than a given x, the population obeying any continuous distribution law. Their arguments are too complicated to attempt to sketch them here, but they are based on the fact that the transformed variable, y, as defined above, is rectangularly distributed on the interval $(0, 1)$. With their exact solution they gave a more convenient approximate method for calculation in applications.

In 1938 (I am not being strictly chronological) E. S. Pearson [28] published a study of test criteria based on this probability integral transformation. Suppose that we have n independently observed y's, $y_1, y_2, \cdots, y_n$. How should the y's be used to test the hypothesis that the observations from which the y's were calculated all came from the same population? K. Pearson [24] had

already suggested the use of $Q = y_1 y_2 \cdots y_n$ or $Q' = (1 - y_1)(1 - y_2) \cdots (1 - y_n)$. It is known that a simple function of Q or of Q' obeys a χ^2-distribution with $2n$ degrees of freedom so that we have a ready means of combining independent tests based on Q or Q'. But how is one to choose among Q, Q', or other functions of the y's that might be suggested? E. S. Pearson emphasized the role that the hypotheses conceived as alternate to the one being tested should play in making such a choice. He illustrates this in a case of testing the hypothesis that a sample came from a normal population of zero mean and unit variance and in which the alternate populations, from one of which the sample might have been drawn, are such that the corresponding y's calculated on the hypothesis being tested would follow a Pearson type I distribution law. Using the likelihood principle he was led in this case to Q or Q', which are then concluded to be "best possible tests."

The final paper I want to discuss is an important one by J. Neyman on the "Smooth test of goodness of fit," published in 1937 [29]. Suppose again that a random sample of N values of x gives the set, y_1, y_2, $\cdots y_N$ on the hypothesis H_0 that the population distribution law is $p(x \mid H_0)$. If H_0 is true the y's in random samples do follow a rectangular distribution on $(0, 1)$. But what would be the distribution of the y's if the distribution law for the population were actually $p(x \mid H_1)$? We have for the y's as calculated,

$$y = \int_a^x p(x \mid H_0) \, dx.$$

But to find $f(y)$,

$$f(y) \, dy = p(x \mid H_1) \, dx,$$

so that,

$$f(y) = \frac{p(x \mid H_1)}{p(x \mid H_0)} \neq 1.$$

Therefore if H_0 is not true, the y's calculated on the assumption that it is may be expected to exhibit a statistically significant set of deviations from a rectangular distribution.

As Neyman remarks, it is a defect of the χ^2-test of goodness of fit that the information one has of the algebraic signs of the differences between calculated and observed frequencies, particularly of the way in which positive and negative differences succeed each other, is completely unused. And in forming a test of a statistical hypothesis it is now well understood, thanks to Neyman and Pearson, that due account should be taken of the alternate hypotheses conceivably true.

Neyman begins by specifying a wide class of alternate hypotheses in a form that lends itself to mathematical treatment. This is done by assuming that the distribution of y's calculated for H_0 will, if an alternate H_1 is true, be given by a function of the form,

$$p(y \mid \theta_1, \theta_2, \cdots, \theta_k) = c e^{\sum_1^k \theta_i \pi_i(y)}$$

in which $\pi_i(y)$ is a polynomial of degree i (a transformed Legendre polynomial) with convenient properties. For low values of k, such as will ordinarily be used, this permits alternate distribution curves to deviate in a smooth manner from the distribution tested, with a limited number of intersections with it.

Now the problem is to determine the function of the observed y's which will provide a suitable test of H_0 with respect to the alternate hypotheses of order or class k, k having been decided upon in advance of making the test. The mathematics, proceeding along Neyman and Pearson lines, shows that the appropriate function, for large samples at least, is simply $\sum_1^k u_i^2$ in which,

$$u_i = \frac{1}{\sqrt{N}} \sum_{j=1}^N \pi_i(y_i')$$

the y_j''s being calculated from the sample. Moreover, the probability that the sum $\sum_1^k u_i^2$ exceeds a given value is at once obtained from a table of the incomplete Γ-function, i.e., this sum is proportional to a χ^2.

This is a very fine piece of work but, as Neyman points out, there are still questions to be settled concerning the general utility of this "smooth test." F. N. David in 1939 [30] further discussed this test. In particular, it may be pointed out that the parameters in $p(x \mid H_0)$ must be assumed known; what would be the effect on the test of estimating these parameters is unknown. A reasonably large sample seems to be required to make the developments on the assumption of large samples applicable but a y must be calculated for each observation. This makes for a good deal of computing but it is not known how grouping of observations might be effected. And the matter of the choice of the order of the test to be applied, i.e., of a value of k, is still somewhat in doubt.

I will not debate the proposition that there are papers completely omitted from this discussion as important as those I have included however inadequately. The limitations of space forced me to choose and it is quite possible that my personal tastes and interests had more weight than they should.

REFERENCES

[1] H. HOTELLING, "Relations between two sets of variates," *Biometrika*, 28 (1936), 328–377.

[2] S. S. WILKS, "Certain generalizations in the analysis of variance," *Biometrika*, 24 (1932), 471–494.

[3] R. A. FISHER, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, 7 (1936), 179–188.

[4] H. HOTELLING, "The generalization of Student's ratio," *Annals of Math. Stat.*, 2 (1931), 360–378.

[5] R. A. FISHER, "The statistical utilization of multiple measurements," *Annals of Eugenics*, 8 (1938), 376–386.

[6] R. A. FISHER, "The sampling distribution of some statistics obtained from non-linear equations," *Annals of Eugenics*, 9 (1939), 238–249.

[7] R. A. FISHER, "The precision of discriminant functions," *Annals of Eugenics*, 10 (1940), 422–429.

[8] P. C. MAHALANOBIS, "On the generalized distance in statistics," *Proc. Nat. Inst. Sci. Ind.*, 12 (1936), 49–55.

[9] R. C. BOSE, "On the exact distribution of the D^2 statistic," *Sankhya*, 2 (1936), 143–154.

[10] R. C. BOSE AND S. N. ROY, "The exact distribution of the Studentized D^2 statistic," *Sankhya*, 3 (1938), part 4.

[11] P. L. HSU, "On the distribution of roots of certain determinantal equations," *Annals of Eugenics*, 9 (1939), 250–258.

[12] M. A. GIRSCHICK, "On the sampling theory of the roots of determinantal equations," *Annals of Math. Stat.*, 10 (1939), 203–224.

[13] M. S. BARTLETT, "Further aspects of the theory of multiple regression," *Proc. Camb. Phil. Soc.*, 34 (1938), 33–40.

[14] M. S. BARTLETT, "The statistical significance of canonical correlations," *Biometrika*, 32 (1941), 29–37.

[15] W. D. KERMACK AND A. G. McKENDRICK, "Tests for randomness in a series of numerical observations," *Roy. Soc. Edin. Proc.*, 57 (1937), 228–240.

[16] W. D. KERMACK AND A. G. McKENDRICK, "Some distributions associated with a randomly arranged set of numbers," *Ibid.*, 332–376.

[17] W. D. KERMACK AND A. G. McKENDRICK, "Some properties of points arranged at random on a Möbius surface," *Mathematical Gazette*, 22 (1938), 66–72.

[18] A. M. MOOD, "The distribution theory of runs," *Annals of Math. Stat.*, 11 (1940), 367–392.

[19] A. WALD AND J. WOLFOWITZ, "On a test whether two samples are from the same population," *Annals of Math. Stat.*, 11 (1940), 147–162.

[20] W. L. STEVENS, "Distribution of groups in a sequence of alternatives," *Annals of Eugenics*, 9 (1939), 10–17.

[21] H. CRAMER, "On the composition of elementary errors, Second paper, Statistical applications," *Skandinavisk Aktuarietidskrift*, 11 (1928), 141–180.

[22] R. VON MISES, "Vorlesungen aus dem Gebiete der angewandten Matematik," Bd. 1: *Wahrscheinlichkeitsrechnung*, (Leipzig, 1931), 316–335.

[23] R. A. FISHER, *Statistical Methods for Research Workers*, (Edinburgh, 4th edition, 1932), Article 21.1.

[24] K. PEARSON, "On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random," *Biometrika*, 25 (1933), 379–410.

[25] K. PEARSON, "On a new method of determining 'goodness of fit,'" *Biometrika*, 26 (1934), 425–442.

[26] W. R. THOMPSON, "On confidence ranges for the median and other expectation distributions for populations of unknown distribution form," *Annals of Math. Stat.*, 7 (1936), 122–128.

[27] A. WALD AND J. WOLFOWITZ, "Confidence limits for continuous distribution functions," *Annals of Math. Stat.*, 10 (1939), 105–118.

[28] E. S. PEARSON, "The probability integral transformation for testing goodness of fit and combining independent tests of significance." *Biometrika*, 30 (1938), 134–148.

[29] J. NEYMAN, "A smooth test of goodness of fit," *Skandinavisk Aktuarietidskrift*, 20 (1937), 149–199.

[30] F. N. DAVID, "On Neyman's 'smooth' test of goodness of fit. I. Distribution of the criterion ψ^2 when the hypothesis tested is true," *Biometrika*, 31 (1939), 191–199.