A GENERALIZED T TEST AND MEASURE OF MULTIVARIATE DISPERSION

HAROLD HOTELLING
UNIVERSITY OF NORTH CAROLINA

CONTENTS

1. Introduction

A set of distributions advanced here generalizes to more than one dimension the analysis of variance of R. A. Fisher, which is itself a generalization including several simpler tests as special cases. The first of these, a generalization of the Student distribution to p dimensions, was introduced in 1931, and provides a test of significance for discriminant functions. The others developed initially in connection with a wartime problem of air testing sample bombsights, which led to a reconsideration of the whole problem of measuring ballistic dispersion. The statistical tests found efficient for these purposes were then seen to be suitable for a variety of industrial quality control and sampling inspection problems, particularly those involving multiple measurements, as on complicated and expensive assemblies and performance tests. These statistical methods are also related to certain procedures developed in India in connection with anthropometric investigations.

Of the bombsight report [9] the larger and less mathematical part has now been published [11]; this published work gives details regarding certain applications of the sampling distributions for which the fundamental mathematics, originally written as appendices to the bombsight report, appears publicly for the first time in the last three sections of this paper.

The multivariate generalization of the analysis of variance is not uniquely confined to the distributions here considered. In passing from one to a plurality of dimensions the variety of null hypotheses of interest is greatly multiplied, and the

Completion of this investigation was aided by the Office of Naval Research under Project NR-042 031 for research in multivariate analysis at Chapel Hill. Preliminary versions were embodied in two appendices of a report [9] prepared during the war, in a paper presented at a meeting of the Institute of Mathematical Statistics at Atlantic City, January 25, 1947, and in an abstract [10].

relevant families of alternatives are multiplied still more. The statistics considered in this paper are proportional to sums of roots of determinantal equations of the form

$$|S_1 - \lambda S_0| = 0,$$

where S_1 and S_0 are independent sample matrices estimating the same covariance matrix in a normal population. As such they generalize the well known ratio F of independent sample variances. They are invariants under arbitrary linear transformations of the variates, and this fact is of great significance, both in actually determining their distribution in readily calculable form in certain cases, and in showing, in all cases, just what parameters enter into the distributions. But the roots themselves and all functions of them are also invariants, and offer possibilities for statistical tests of still other varieties, whose properties need further exploration. The statistics here denoted generically by T are, for reasons that will appear in the course of the paper, seemingly the most suitable for certain classes of applications. The elimination from use for comparisons of ballistic accuracy of one important statistic, the ratio of determinants $|S_1|/|S_0|$, which equals the product of the roots, has already been discussed [11, p. 166].

2. Generalized Student ratios and discriminant functions

The Fisher-Student t-tests for the significance of means, differences of means, and other linear functions of observations on a single variate, were generalized in 1931 [6], [26], [15, p. 235 ff.] to make possible combined tests based on two or more variates which may or may not be independent. An important feature of these tests is that they yield exact probabilities independent of unknown parameters. A large category of such tests, corresponding to the many uses of t introduced by Fisher as extensions of Student's work, may be summarized as follows:

Let $x_1, \ldots, x_p$ be normally correlated variates, and in a random sample of N individuals let $x_{i\alpha}$ be the value of x_i for the α-th individual. Let $[s_{ij}]$ be a matrix of unbiased estimates of the covariances among the x_i with n degrees of freedom, and having essentially the Wishart distribution. The conditions of the particular test to be made lead to the values of constants $a_1, \ldots, a_p, c_1, \ldots, c_N$ such that the p statistics

$$z_i = Sc_\alpha (x_{i\alpha} - a_i)$$

have the same distribution as $x_1, \ldots, x_p$ independently of the s_{ij}. (Following Fisher, we use S for summation over the sample, here from 1 to N, and $\sum$ for summation over the p variates.) Thus for testing the collective deviations of a set of sample means from assigned numbers a_i we have

$$c_\alpha = N^{-1/2}, \qquad n = N - 1, \qquad s_{ij} = \frac{1}{n} S (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j).$$

For testing whether the means of two independent samples of N_1 and N_2 individuals respectively differ significantly from each other we put N_1 of the c_α corresponding to one sample and N_2 corresponding to the other equal respectively to

$N_2^{1/2}N_1^{-1/2}(N_1 + N_2)^{-1/2}$ and $-N_1^{1/2}N_2^{-1/2}(N_1 + N_2)^{-1/2}$, while $n = N_1 + N_2 - 2$,

$$ns_{ij} = S'(x_{ia} - \bar{x}'_i)(x_{ja} - \bar{x}'_j) + S''(x_{ia} - \bar{x}''_i)(x_{ja} - \bar{x}''_j),$$

where the sums are over the two samples and refer to products of deviations from the respective sample means. The a_i are all zero in this test.

Letting in general

$$[l_{ij}] = [s_{ij}]^{-1},$$

we define T as the positive square root of

$$T^2 = \sum \sum l_{ij} z_i z_j ,$$

and have for its distribution

$$\frac{\left(\frac{T^2}{n}\right)^{(p-2)/2} d\,\frac{T^2}{n}}{B\left(\frac{n-p+1}{2}, \frac{p}{2}\right)\left(1+\frac{T^2}{n}\right)^{(n+1)/2}}.$$

The probability P that T should on the null hypothesis tested exceed that actually found from the observations is the integral of this expression to infinity, and may with the notation

$$x = \frac{T^2}{n+T^2}$$

be written in terms of the incomplete beta function [19]

$$P = I_x\left(\frac{n-p+1}{2}, \frac{p}{2}\right).$$

The now familiar percentage points of the variance ratio distribution may also be used to test T by putting

$$F = \frac{(n-p+1)T^2}{np}, \qquad n_1 = p, \qquad n_2 = n - p + 1.$$

The case $p = 1$ yields the Student test.

The significance of the set of differences between corresponding means of two samples is often judged by calculating a discriminant function $\sum k_i x_i$, where $k_i = \sum l_{ij}(\bar{x}'_j - \bar{x}''_j)$, substituting for each x_i the difference of its means in the two samples, and comparing with a standard error in a way equivalent to an analysis of variance with p and $n - p + 1$ degrees of freedom. When carried out accurately, this procedure gives exactly the same final result as the T-test, and is therefore correct, since the T distribution has been derived rigorously in full detail. Arguments of dubious validity are often given in support of this kind of "analysis of variance," but the only real proof of its correctness is its equivalence with the T-distribution, or by an argument paralleling that leading to it.

Discriminant functions are useful in assigning new individuals to one or another of populations already distinguished from each other by the T-test or otherwise, but are unnecessary where the sole object is to decide whether real differences exist. This point was clarified by R. A. Fisher as early as 1938 [4], but has been overlooked by some recent writers who seem to have been unaware of the T test and to have gone to unnecessary trouble with discriminant functions.

The probability of misclassification of a new individual by a discriminant function based on samples from two populations has been studied by Wald [24], who employed a bilinear form with the same matrix $[l_{ij}]$ as T^2 and reduced its distribution function to a triple integral.

Theorems on the efficiency of the T tests have been established by P. L. Hsu [13], J. B. Simaika [23] and A. Wald [25], who have shown under various conditions that no better test is possible. Daly [2] has shown these tests to be unbiased in the Neyman-Pearson type A sense.

David Durand applied T^2 and discriminant functions in a study of criteria for distinguishing good from bad applications for loans [3]. Durand not only gives instructive applications but develops additional details of technique, one of which uses a theorem that for a particular set of observations T is never less when all the variates are used than when only a subset of them is employed.

3. Figurative distance and coefficient of racial likeness

T may be employed not only for testing but also to estimate a population parameter which may be called the "figurative distance." In a Euclidean space of p dimensions each individual defines a point, and the normal population is represented by a swarm of points having ellipsoidal symmetry. Two species may be represented by two such ellipsoidal swarms. If the covariance matrix is the same for the two species, a transformation of axes, usually oblique, exists which changes the ellipsoids to spheres and transforms the covariance matrix into the identity matrix. The squared distance between the centers of gravity of the two swarms is then $\sum\sum \lambda_{ij}\delta_i\delta_j$, where δ_i is the difference between the two population means of the i-th variate and $[\lambda_{ij}]$ is the inverse of the covariance matrix. If δ_i and λ_{ij} are replaced by their sample estimates $\bar{x}_i' - \bar{x}_i''$ and l_{ij} respectively, the result is $T^2(N_1^{-1} + N_2^{-1})$, which is thus a sufficient statistic for estimating the squared figurative distance. A distribution generalizing that above to take account of values of this parameter possibly different from zero was derived by R. C. Bose and S. N. Roy [1]. This distribution may be used to obtain confidence limits for the figurative distance, to test hypotheses about it, and as a power function for the original T-test.

The use of a multiple of T as a measure of "distance" was introduced by Mahalanobis [16], who was concerned about its probable variation when the number p of variates measured is changed, and particularly about its limiting properties as p increases. This is a difficult problem and strongly resembles that of psychologists interested in the variations resulting from changes in a battery of tests designed to measure a particular set of characters. No solution of these problems appears possible without reference to a theory of variates regarded as samples of a larger aggregate of variates that might have been measured on the same population of individuals. Only a beginning has been made on such a theory [7, pp. 504–514]. Mahalanobis used the "distance" in connection with anthropometric measurements to demonstrate that, contrary to the accepted rules, much intermarriage has taken place among castes in Bengal.

Karl Pearson's "coefficient of racial likeness" [18], the sum of squares of the ratios of the differences of means to the respective standard deviations, was designed as a measure of what might be called distance between two populations, but is unsuitable for practical use because its sampling distribution, which is still undetermined, appears to involve a large number of nuisance parameters in such a way as to make exact probability tests impossible. The sample value of this coefficient appears also to be an inefficient estimate of its population value because it takes no account of correlations between the variates. Pearson was indeed aware that without correlation terms his coefficient could not be ideal, and alluded to a possible χ^2 test as a substitute which he considered impractical for computational reasons. It may perhaps be inferred from this remark that Pearson thought of using T, the distribution of which, though then unknown, approaches that of χ for large samples, but was deterred by the labor of computing the inverse of the covariance matrix. This labor has now been greatly lightened by new iterative methods, better direct methods, and improved machines, and up to rather high values of p may in the future be regarded as almost negligible in comparison with the cost of obtaining the data.

In an endeavor to avoid the weakness of Pearson's coefficient, Romanovsky [21] introduced certain substitute statistics and derived exact distributions on the basis of assumed independence among the variates. These sampling distributions may be very much affected by correlations among the variates, and such correlations are in fact very likely to exist. In using T, on the other hand, one does not need to know whether there are real correlations or not.

4. Approximate T distribution for large samples

The approach of the T distribution discussed above to that of χ is easy to demonstrate, either directly or through the fact that if $[l_{ij}]$ is replaced by $[\lambda_{ij}]$, to which it converges in probability, the resulting statistic has the χ^2 distribution. A method of approximation to the probability that T should be exceeded, using the χ^2 distribution and an asymptotic series in n^{-1}, has been given [12]. Writing

$$\chi^2 \sim T^2 \left\{ 1 - \frac{p+T^2}{2n} + \frac{4-p^2+(2+5p)T^2+8T^4}{24n^2} + \cdots \right\},$$

we have for large values of n an approximation by the first few terms to a variate having the χ^2 distribution with p degrees of freedom.

Substituting for T^2 in this formula a series of powers of n^{-1} with undetermined coefficients and then making the total coefficient of each power of n^{-1} zero serves to fix the coefficients and yield the series

$$T^2 \sim \chi^2 \left\{ 1 + \frac{p+\chi^2}{2n} + \frac{7p^2-4+(13p-2)\chi^2+4\chi^4}{24n^2} + \cdots \right\}.$$

Substituting on the right any percentage point of the χ^2 distribution with p degrees of freedom (for example, the value exceeded with probability .05) gives an approximation to the corresponding percentage point of the T^2 distribution. It must be remembered, however, that the series apparently diverge and that this method is good only for large values of n.

In the univariate case the formula

$$x \sim t \left\{ 1 - \frac{1+t^2}{4n} + \ldots \right\},$$

obtained in [12] as a means of relating t to a normally distributed variate x, and thus obviating to some extent tables of percentage points of t, may be obtained from the first of the asymptotic formulae above by extracting the square root and putting $p = 1$. Upon extracting the square root of the second asymptotic series and putting $p = 1$ we have a series for percentage points of the Student distribution, the first two terms of which were obtained by A. M. Peiser [20] in a different way.

5. Ballistic and quality control applications

A new use for T and certain generalizations of it arose during the war in the study of ballistic dispersion, initially in relation to the measurement of accuracy of test bombing for quality control purposes [9], [11]. Such dispersion is usually of a type close to the normal distribution in two dimensions, and the dispersion of anti-aircraft shells fired from a gun held in a fixed position can doubtless be approximated by the trivariate normal distribution.

Test bombing errors are expressed in terms of *range error* x_1, measured in the direction of flight of the airplane, and *deflection error* x_2 measured perpendicular to this direction. On the basis of these measures a particular bomb—or the personnel, equipment or procedure used in dropping it—might be classified as defective. Such a decision would be based on extraordinarily large values (positive or negative) of x_1 or x_2 or both. The question is, what combinations of x_1 and x_2 should be taken as indicating that a bomb is defective? An answer to this question is equivalent to delimiting an area around the target and regarding all bombs falling within this area as of standard quality, while rejecting as defective all those falling outside it. The problem is to find an area of suitable shape and size.

A first thought is that the appropriate region is a circle centered at the target. Consideration must, however, be given to the fact that deflection errors are as a rule distinctly larger than range errors, since the former include a substantial component absent from the latter in that the course of the airplane may be too far to the right or left and is not easily corrected in the last few seconds before the bomb is dropped. There is also some indication of a small correlation between range and deflection errors. The distribution of actual bombs dropped has, therefore, an elliptical rather than a circular symmetry. If the covariance matrix, and therefore its inverse λ, are known, an ellipse of constant probability density is defined by

$$\sum \sum \lambda_{ij} x_i x_j = \text{constant}, \qquad i, j = 1, 2.$$

If an ellipse of this family is chosen as the boundary of a region of acceptance such that a definite proportion, say .99, of bombs will fall within it, it will overlap a circle containing the same proportion of the bombs. The part of the elliptical region outside the circle has the same probability, and in the long run will receive the same proportion of normal bombs, as the part of the circular region outside the ellipse.

The *density* of probability for normal bombs is, however, greater in the portion of the ellipse outside the circle than in the portion of the circle outside the ellipse, because of the way the ellipse was chosen. The density is indeed greater at every point within the ellipse than at any point outside it. Since for any region

$$\text{probability} = (\text{mean probability density}) \times (\text{area}) \,,$$

the area of the part of the circle outside the ellipse must be greater than that of the part of the ellipse outside the circle. Consequently, the whole circular region has a greater area than the elliptical region.

Now if a bomb, or the bombsight used in dropping it, is for some reason abnormal, while the chance factors remain the same as usual, the probability of the bomb falling within either the circle or the ellipse is decreased. We cannot say exactly how much it is decreased without knowing details regarding the nature of its abnormality which are not ordinarily available. The normal distribution with matrix λ no longer holds, and we do not know precisely what is the distribution of abnormal bombs. In this situation, a certain presumption exists that the larger area will receive more of the abnormal bombs than the smaller area. Thus we may expect the circular area, since it is the larger, to receive more abnormal bombs than the elliptical area, though the proportions of normal bombs are equal. Since we *call* only those bombs abnormal which fall outside the chosen region, the chance of detecting abnormal bombs is likely to be greater for the elliptical than for the circular region.

An illustration of this situation would be an error creeping into the manufacture of certain bombs which adds to each coordinate of the dropped bomb a normally distributed error of zero mean and variance large in comparison with the variances otherwise existing. To make the illustration more specific, though this is not essential to the argument, we may further suppose that the new errors introduced by the manufacturing defect are independent of each other as well as of the old errors. Moreover, let us consider them of equal variance in the two dimensions. Then if these manufacturing errors were the only source of variance, the distribution of the dropped bombs would have circular symmetry about the target. But it would be quite wrong to use a circular region of acceptance. Since the variances are now large, the probability density near the target of the abnormal bombs is substantially constant, and the probability of an abnormal bomb falling into any region is practically proportional to the area of that region. Since the circular acceptance region has greater area than the elliptical region defined above, a greater proportion of defective bombs will erroneously be accepted as normal when the circular than when the elliptical region is used.

To apply this argument it is not necessary to confine oneself to defects producing equal and independent variances in the two directions, nor having normal distributions, nor to circles as acceptance regions alternative to the ellipses defined above. It is clear that as against a very extensive class of abnormal bombs (though not all kinds of abnormality) the particular elliptical regions considered are the most efficient possible means of discrimination between normal and abnormal bombs. For precisely defined distributions of bombing deviations corresponding to defects for which it is desired to reject bombs, it may be possible to devise better acceptance regions, but in the absence of exact knowledge of such distributions the elliptical regions are to be preferred. The plausible circular regions are clearly not the most efficient discriminators, even where as in the illustration above the defects would by themselves and in the absence of other variation produce a circular distribution.

The statistic

$$\chi^2 = \sum_{j=1}^{p} \sum_{i=1}^{p} \lambda_{ij} x_i x_j$$

was introduced by Karl Pearson [17] in 1900 as a criterion for deciding whether an outlying individual belongs to a specified population, and was shown to have the now well known χ^2 distribution with p degrees of freedom for individuals actually belonging to a normal population with zero means and covariance matrix λ^{-1}. For $p = 2$, the case appropriate to bombing, the distribution is particularly simple: the probability that the particular value χ^2 should be exceeded is $e^{-\chi^2/2}$. It is this statistic that commends itself, in accordance with the discussion above, as a means of picking out abnormal bombs when the standard covariance matrix λ^{-1} is known.

The only source of knowledge of the covariance matrix ordinarily available is a record of observations, which in the case of acceptance bombing would mean a record of coordinates of bombs dropped under conditions of altitude, visibility, turbulence, etc., of a somewhat standardized character. Quality control plans start in general from the record of some initial period, and undertake to determine whether later production is as good as in the initial period. Let $x_{i\alpha}$ be the i-th coordinate ($i = 1, 2$) of the α-th among N bombs dropped in the initial period. Then the covariance σ_{ij} of x_i and x_j (where i and j may be the same or different; if they are the same, the covariance is a variance) may be estimated by $Sx_{i\alpha}x_{j\alpha}/N$, or by $S(x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)/(N - 1)$, or by sums of products of deviations from corresponding least-square regression values adjusted to take account of various extraneous conditions. The first of these estimates is appropriate if the distribution is clearly symmetrical about the target and the coordinates are measured from the target as origin, but the second is safer because of the strong possibility of some bias, and more refined adjustments may sometimes be needed. Let s_{ij} be the estimate adopted, and let n be the corresponding number of degrees of freedom, which for the first of the estimates above is N, and for the second is $N - 1$. If the choice is made in a way suitably corresponding to the actual situation, the set of s_{ij}'s will have the Wishart distribution. We shall speak of this initial set of bombs as the *old sample*. In selecting the old sample, care should be taken to eliminate any bombs known to be defective, and to insure that the bombing technique was of the same standard type used later.

Let $[l_{ij}]$ be the inverse of the matrix $[s_{ij}]$ of covariance estimates derived from the old sample, and therefore an estimate of the unknown matrix $\lambda = [\sigma_{ij}]^{-1}$. Using l_{ij} in place of λ_{ij} in the χ^2 statistic for dealing with a new bomb yields the statistic

$$T^2 = \sum \sum l_{ij} x_i x_j \,.$$

This has the T^2-distribution described in section 2 with $p = 2$, a case for which the distribution simplifies to

$$\frac{(n-1)\, d\, \frac{T^2}{n}}{2\left(1+\frac{T^2}{n}\right)^{(n+1)/2}} .$$

This is readily integrated to give as the probability that, for a bomb of standard quality, the statistic should exceed the particular value,

$$\left(1+\frac{T^2}{n}\right)^{-(n-1)/2} .$$

Suppose now that a sample of M bombs from a new lot are dropped, and that the i-th coordinate ($i = 1, 2$) of the B-th bomb is x_{iB}. The foregoing statistic for testing the B-th bomb may be written

$$T_B^2 = \sum \sum l_{ij} x_{iB} x_{jB} .$$

A combined over all test of the quality of the new lot is supplied by the statistic

$$T_0^2 = T_1^2 + T_2^2 + \ldots + T_M^2 .$$

If the fundamental covariance matrix were known exactly we should instead use for the single bomb

$$\chi_B^2 = \sum \sum \lambda_{ij} x_{iB} x_{jB} ,$$

and for the whole new sample,

$$\chi_0^2 = \chi_1^2 + \ldots + \chi_M^2 ,$$

with the knowledge that, because of the well known additive property, χ_0^2 has the standard χ^2 distribution with $2M$ degrees of freedom.

Similarly, if p dimensions are measured in quality control and these have the joint normal distribution and are measured from their means as origin, we have for an individual article χ_B^2 as given by the formula above, with the summations from 1 to p, and χ_0^2 as the sum of M values of χ_B^2. These statistics will have the standard χ^2 distribution with p and Mp degrees of freedom respectively. They must, however, in practice be replaced ordinarily by the statistics T_B^2 and T_0^2, whose distribution is different.

The centroid or mean point of impact (MPI) of the bombs in the new sample is of considerable interest in itself, and so is the dispersion about the MPI. The coordinates of the MPI are $\bar{x}_i = S'x_{iB}/M$, where S' denotes summation over the new sample. Because of the identity

$$S'x_{iB}x_{jB} = S'(x_{iB} - \bar{x}_i)(x_{jB} - \bar{x}_j) + M\bar{x}_i\bar{x}_j ,$$

we may resolve χ_0^2 into components

$$\chi_D^2 = \sum \sum S'\lambda_{ij}(x_{iB} - \bar{x}_i)(x_{jB} - \bar{x}_j) ,$$

$$\chi_M^2 = M \sum \sum \lambda_{ij}\bar{x}_i\bar{x}_j ,$$

which have independent χ^2 distributions with $Mp - p = 2(M - 1)$ and $p = 2$ degrees of freedom respectively; and

$$\chi_0^2 = \chi_D^2 + \chi_M^2 .$$

In the same way the over all quality measure T_0^2 may be resolved into components

$$T_D^2 = \sum \sum S' l_{ij}(x_{iB} - \bar{x}_i)(x_{jB} - \bar{x}_j) ,$$

$$T_M^2 = M \sum \sum S' l_{ij} \bar{x}_i \bar{x}_j ,$$

of which the last has the T^2 distribution described above, and the former has a distribution which, with that of T_0^2, will be investigated in the next section. We have

$$T_0^2 = T_M^2 + T_D^2 .$$

Since T_M and T_D depend on the same old sample they are not independent. However, their conditional distributions for a particular old sample are independent.

Control charts have been recommended and illustrated [9], [11] for T_M as a means of detecting faults leading to abnormal mean points of impact, for T_D as revealing dispersion about the *MPI*, for the means and standard deviations of range and deflection separately, and for the over all measure T_0 combining all these into one number measuring the general accuracy. When the charts illustrating the method were first prepared the exact distributions of T_0 and T_D had not been discovered, so that the control lines corresponding to the .01 and .05 probabilities were fixed on the basis of the χ^2 approximation. The resulting error in the positions of these lines has been found to be appreciable since the discovery of the exact distribution, though the number of degrees of freedom n for the old sample was as large as 256. This emphasizes the importance of using the exact distribution, which will next be derived for the case $p = 2$.

The distributions for $p > 2$ are of a distinctly more complicated mathematical type than for $p = 2$. Dorothy Morrow Gilford is preparing a study of the distributions for $p > 2$.

Further partitions of T_0^2 or of χ_0^2, corresponding to the subdivision of sums of squares into more than two parts in the familiar analysis of variance tables, may yield much information, and call for no additional distributions. Thus these T statistics may readily be adapted [9], [11] to examine variations due to day of flight, personnel, bombing equipment, etc.

6. Distributions of the *T* statistics in two dimensions

If each coordinate of each of the M bombs in the new sample has zero expectation, the $p(p + 1)/2$ (= 3) statistics

$$s'_{ij} = \frac{1}{M} S' x_{iB} x_{jB}$$

have jointly the Wishart distribution with M degrees of freedom. If these expectations have any values, the same for all the bombs in the new sample, the statistics

$$s''_{ij} = \frac{1}{M-1} S' (x_{iB} - \bar{x}_i)(x_{jB} - \bar{x}_j)$$

have jointly the Wishart distribution with $M - 1$ degrees of freedom. To cover simultaneously these and other cases, we use s^*_{ij} to denote the sum over the new sample of products of deviations of x_{iB} and x_{jB} from their respective regression values upon a common set of independent variables, divided by the number of degrees of freedom, which we shall call m. If there is only a single set of independent variables, and these always equal unity, the regression value is the mean, and the last formula gives s^*_{ij} with $m = M - 1$. If the common set of independent variables is the null set, we have the first formula for s'_{ij}, with $m = M$.

Upon referring to the definitions of T_0 and T_D in the preceding section we find that in terms of the notation above,

$$T_0^2 = M \sum \sum l_{ij} s'_{ij}, \qquad T_D^2 = (M-1) \sum \sum l_{ij} s''_{ij}.$$

From these relations it is clear that if the new sample is drawn independently from the same normal population as the old, the distribution of T_D^2 is of exactly the same form as that of T_0^2 with M replaced by $M - 1$. We shall, therefore, investigate, for a general value of m, the distribution of the general statistic

$$T^2 = m \sum \sum l_{ij} s^*_{ij},$$

where the variates s^*_{ij} have the Wishart distribution with m degrees of freedom. For $m = 1$ the distribution reduces to that originally found for T^2, which is the same as that of T_B^2 and T_M^2. The family of distributions involves no parameters but the three integers n, m and p specifying respectively the numbers of degrees of freedom in the old and the new samples and the dimensionality. We obtain here the distributions for $p = 2$.

$$\text{Let } S_0 = [s_{ij}], \text{ whence } S_0^{-1} = [l_{ij}], \text{ and let } S_1 = [s^*_{ij}].$$

The trace of the matrix $S_0^{-1}S_1$, that is, the sum of its principal diagonal elements, is then T^2/m, and we write

$$u = \frac{T^2}{m} = \text{trace}\,(S_0^{-1}S_1).$$

Now the trace is the sum of the roots of the characteristic equation

$$|S_0^{-1}S_1 - \lambda| = 0,$$

and therefore of the equation, which has the same roots,

$$|S_1 - \lambda S_0| = 0.$$

The distribution of roots of such equations has been the subject of various investigations, with papers by S. N. Roy [22], P. L. Hsu [14] and R. A. Fisher [5] appearing simultaneously in 1939 with a joint distribution which may be written in terms of the roots λ_1, λ_2 of the equation above when $p = 2$, $m > 1$ and $n > 2$, in the form

$$C m^m n^n \frac{(\lambda_1\lambda_2)^{(m-3)/2}(\lambda_1 - \lambda_2)\, d\lambda_1 d\lambda_2}{[(n + m\lambda_1)(n + m\lambda_2)]^{(m+n)/2}},$$

where

$$C = \frac{(m+n-2)!}{4(m-2)!(n-2)!}$$

and $\lambda_1 \geqq \lambda_2 \geqq 0$. From this the joint distribution of the sum and product of the roots is found by putting

$$u = \lambda_1 + \lambda_2 = \frac{T^2}{m}, \qquad v = \lambda_1\lambda_2,$$

so that

$$du dv = (\lambda_1 - \lambda_2) d\lambda_1 d\lambda_2.$$

This gives

$$\frac{C m^m n^n v^{(m-3)/2} du\, dv}{(n^2 + mnu + m^2 v)^{(m+n)/2}}.$$

The distribution of u is found by integrating this with respect to v from 0 to $u^2/4$. Changing the variable of integration to

$$z = \frac{m^2 v}{n^2 + mnu + m^2 v},$$

so that

$$v = \frac{n(n + mu) z}{m^2 (1 - z)}, \qquad dv = \frac{n(n + mu)\, dz}{m^2 (1 - z)^2},$$

and replacing u by T^2/m gives the distribution of T^2 in the form

$$\frac{C d \frac{T^2}{n}}{\left(1 + \frac{T^2}{n}\right)^{(n+1)/2}} \int_0^{T^4/(2n+T^2)^2} z^{(m-3)/2}(1 - z)^{(n-1)/2} dz.$$

The incomplete beta function

$$B_x(p, q) = \int_0^x z^{p-1}(1 - z)^{q-1} dz$$

has been tabulated in detail [19]. The integral appearing in the element of probability is of this form. Apart from this integral and a constant factor, the distribution is the same as that obtained in 1931 for T^2.

The probability integral for T^2 may be evaluated in the following manner. We first observe that the factor of the element of probability before the integral sign is the differential of

$$-\frac{2C}{n - 1}\left(1 + \frac{T^2}{n}\right)^{-(n-1)/2}.$$

Hence, if we put $T^2 = 2nw/(1 - w)$, this factor may be written

$$-\frac{2C}{n - 1}\frac{d}{dw}\left(\frac{1 - w}{1 + w}\right)^{(n-1)/2} dw.$$

If P denotes the probability that T exceeds a positive value T', then $1 - P$ is the probability that $T \leqq T'$, and is also the probability that w lies between 0 and

$$w' = \frac{T'^2}{2n + T'^2}.$$

Hence

$$1 - P = -\frac{2C}{n-1} \int_0^{w'} \int_0^{w^2} z^{(m-3)/2} (1-z)^{(n-1)/2} \frac{d}{dw} \left(\frac{1-w}{1+w}\right)^{(n-1)/2} dz\, dw.$$

(We follow the convention that the inner differential goes with the inner integral sign and corresponds to the first integration.) Reversing the order of integration gives

$$1 - P = -\frac{2C}{n-1} \int_0^{w'^2} \int_{\sqrt{z}}^{w'} z^{(m-3)/2} (1-z)^{(n-1)/2} \frac{d}{dw} \left(\frac{1-w}{1+w}\right)^{(n-1)/2} dw\, dz$$

$$= \frac{2C}{n-1} \int_0^{w'^2} z^{(m-3)/2} (1-z)^{(n-1)/2} \left[\left(\frac{1-\sqrt{z}}{1+\sqrt{z}}\right)^{(n-1)/2} - \left(\frac{1-w'}{1+w'}\right)^{(n-1)/2}\right] dz.$$

Separating this into two integrals, and putting $z = x^2$ in the first, and then for simplicity replacing w' by w, gives finally[1]

$$P = 1 - I_w(m-1, n) + \frac{(m+n-2)!}{2(m-2)!(n-1)!} \left(\frac{1-w}{1+w}\right)^{(n-1)/2} B_{w^2}\left(\frac{m-1}{2}, \frac{n+1}{2}\right).$$

Replacing the factorials by gamma functions with the help of the formula

$$q! = \frac{2^q}{\sqrt{\pi}} \Gamma\left(\frac{q+2}{2}\right) \Gamma\left(\frac{q+1}{2}\right)$$

we find also

$$P = 1 - I_w(m-1, n) + \sqrt{\pi} \frac{\Gamma\left(\frac{m+n-1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{1-w}{1+w}\right)^{(n-1)/2} I_{w^2}\left(\frac{m-1}{2}, \frac{n+1}{2}\right).$$

The .01 and .05 control limits could be ascertained by Newton's method from this formula. Indeed, if the foregoing probability be represented by $P(w)$, the Newtonian iteration starting with any value T assumed to correspond to probability ϵ of being exceeded, may be applied by first calculating

$$w = \frac{T^2}{2n + T^2},$$

then $P(w)$, then

$$P'(w) = -2C \frac{(1-w)^{(n-3)/2}}{(1+w)^{(n+1)/2}} B_{w^2}\left(\frac{m-1}{2}, \frac{n+1}{2}\right),$$

and next the corrected estimate

$$w_1 = w + \frac{\epsilon - P(w)}{P'(w)}.$$

When repetitions of this process give adequately stationary values, the final value

[1] A table of P for values of m and n varying from 1 to 50 is planned for computation on the ENIAC at Aberdeen Proving Ground under the direction of Dr. Frank E. Grubbs.

of w is to be substituted in

$$T^2 = \frac{2nw}{1-w}$$

to give the desired control limit.

Values of T^2, T_0^2, etc., may be added together in much the same way as independent values of χ^2 to obtain a variate having a distribution of like form, which may be used for a combined test of significance. The condition for such additivity is that the several functions T^2 added shall be *conditionally independent*, in the sense that they shall be mutually independent for each fixed set of values for the "old" sample. To prove this, consider q "new" sample covariance matrices $S_1, \ldots, S_q$ corresponding respectively to degrees of freedom $m_1, \ldots, m_q$ in each covariance estimate, and to values $T_1^2, \ldots, T_q^2$. In case of independence among the matrices S_i, the matrix

$$S = \frac{\sum m_i S_i}{\sum m_i}, \qquad i = 1, \ldots, q,$$

has the same distribution as each of the S_i, with $m = m_1 + \ldots + m_q$ degrees of freedom. Putting

$$\begin{aligned} T^2 &= \sum T_i^2 \\ &= \sum m_i \operatorname{tr}(S_0^{-1} S_i) = \sum \operatorname{tr}(S_0^{-1} m_i S_i) = \operatorname{tr}\left(S_0^{-1} \sum m_i S_i\right) \\ &= m \operatorname{tr}(S_0^{-1} S), \end{aligned}$$

we have an expression of the same form as before, and the proposition is proved.

An additional problem of considerable importance is to work out in usable form the exact distribution of the ratio of two values of T^2 which, for a fixed "old" sample, are independent. This generalization of the F distribution would make it possible to test with accuracy the effects of differing conditions, such as changed auxiliary equipment, personnel, type of sight, etc., by comparing the sum of values of T^2 under one set of conditions with the corresponding sum under another set. Until this important distribution can be studied further, this kind of comparison can be made only approximately.

7. The general degree of dispersion of a multivariate normal distribution

Suppose that the correlations and the ratios among the variances of a normally distributed set of variates are known, but that a common factor fixing the values of the variances remains to be estimated from a sample. Thus if σ_{ij} denote the covariance of x_i and x_j, where i and j take values from 1 to p, the number of variates, it is supposed that $p(p+1)/2$ numbers $t_{ij} (= t_{ji})$ are known such that

$$\sigma_{ij} = \gamma t_{ij}.$$

The multiplier γ may be called "the general degree of dispersion" of the distribution. We consider its estimation from a sample. Without loss of generality we shall suppose that both γ and any estimates of it are positive. Since the matrix $\sigma = [\sigma_{ij}]$ is positive definite, the matrix $t = [t_{ij}]$ must then also be positive definite.

We consider the case in which the expectations of the variates are all zero. The general case in which the expectations are unknown may be reduced to this by methods now familiar. In this more general, and in practice more common, situation, we should use the estimate of σ_{ij},

$$\frac{S\,(x_i - \bar{x}_i)\,(x_j - \bar{x}_j)}{N - 1}$$

where N is the sample number, S stands for summation over the sample, and $\bar{x}_i$ for the sample mean of x_i. The distribution of these estimates in samples of N from a population with arbitrary means is exactly the same as the distribution in samples of $n = N - 1$ from a population with zero means of the estimates

$$s_{ij} = \frac{1}{n}\,S x_i x_j\,.$$

These estimates are *unbiased*, that is, $Es_{ij} = \sigma_{ij}$, where E stands for the expectation.

Putting $[q_{ij}] = t^{-1}$ we may write the element of probability for the p variates

$$\frac{\sqrt{|\,q\,|}}{(2\pi\gamma)^{p/2}}\,e^{-(1/2\gamma)\Sigma\Sigma q_{ij}x_ix_j}dx_1 \ldots dx_p\,.$$

The likelihood L is defined, apart from an arbitrary constant factor, as the product of n such elements of probability with x_i replaced in each such element by the observed value in that individual of the sample corresponding to the particular element. With a constant factor chosen in the interests of simplicity, and with the help of the definition of s_{ij} above, this gives

$$\log L = \frac{n}{2}\left(\log|\,q\,| - p\log\gamma - \frac{1}{\gamma}\sum\sum q_{ij}s_{ij}\right).$$

Since no function of the observations other than the s_{ij}'s is involved in this expression, these sample covariances are *sufficient statistics*. This means essentially that all the information inherent in the sample and relevant to the estimation of γ is contained in these statistics, and that no use of any other statistics can supply an improvement of a suitable statistic which is a function of the s_{ij}'s alone.

A statistic whose variance for large samples, considered asymptotically, is a minimum, may be obtained as the value of the parameter making the likelihood a maximum. The asymptotic variance of this statistic is the negative of the reciprocal of the expectation of the second derivative of log L. We have:

$$\frac{\partial \log L}{\partial \gamma} = -\frac{np}{2\gamma} + \frac{n\sum\sum q_{ij}s_{ij}}{2\gamma^2},$$

$$\frac{\partial^2 \log L}{\partial \gamma^2} = \frac{np}{2\gamma^2} - \frac{n\sum\sum q_{ij}s_{ij}}{\gamma^3},$$

$$E\frac{\partial^2 \log L}{\partial \gamma^2}=\frac{np}{2\gamma^2}-\frac{n\sum\sum q_{ij}\sigma_{ij}}{\gamma^3}=\frac{np}{2\gamma^2}-\frac{n\sum\sum q_{ij}t_{ij}}{\gamma^2}=-\frac{np}{2\gamma^2},$$

since $[q_{ij}] = [t_{ij}]^{-1}$ and therefore $\sum\sum q_{ij}t_{ij} = p$.

From the first of these equations it is evident that the maximum of L will be obtained if we substitute for γ the statistic

$$\hat{g}=\frac{1}{p}\sum\sum q_{ij}s_{ij}.$$

From the last of the equations it is found that the variance of this statistic, at least approximately for large samples, is

$$\sigma_{\hat{g}}^2=\frac{2\gamma^2}{np}.$$

Turning from approximate large sample theory to exact results, we shall prove that the exact variance of $\hat{g}$ is given by the formula above. We shall also prove that no other statistic of a certain class has so small a variance. This class consists of those unbiased statistics g which are sums over the sample of the values of a quadratic function of the p observations on each individual. We shall assume independent random sampling of the n individuals from a multivariate normal population with zero means. The results are directly applicable in a broader class of cases, provided we use for s_{ij} an unbiased estimate of variance based on deviations from means, and provided we mean by n, not the sample number, but the number of degrees of freedom among the deviations in each variate that enter into the sums of products.

For a multivariate normal distribution of zero means the fourth moments are given, for example, in [8, section 5], by

$$E\,x_i x_j x_k x_m = \sigma_{ij}\sigma_{km} + \sigma_{ik}\sigma_{jm} + \sigma_{im}\sigma_{jk}\,.$$

This formula may be obtained by differentiating the characteristic function. From it is obtained the exact covariance of two of the sample covariances in the form

$$E\,d\,s_{ij}d\,s_{km}=\frac{1}{n}(\sigma_{ik}\sigma_{jm}+\sigma_{im}\sigma_{jk}),$$

where ds_{ij} denotes $s_{ij} - \sigma_{ij}$. In all these formulae the subscripts vary independently from 1 to p.

We shall use the convention that the repetition of any Latin subscript in a term shall denote summation with respect to that subscript from 1 to p. We shall also use the Kronecker delta δ_{ij} defined as equal to unity if the two subscripts are equal, and otherwise equal to zero. The matrix relation $q = t^{-1}$, or $[q_{ij}] = [t_{ij}]^{-1}$, may thus be written

$$q_{ij}t_{ik} = \delta_{jk}\,.$$

We shall put $x_{i\alpha}$ for the value taken by the i-th variate in the α-th individual of the sample. Repetition of the Greek subscript in any term will denote summation with respect to this subscript from 1 to n. Thus we may write the sample covariance

$$s_{ij} = \frac{1}{n} x_{i\alpha} x_{j\alpha} .$$

The class of statistics g defined above consists of all those given by the formula

$$g = \frac{1}{n} u_{ij} x_{i\alpha} x_{j\alpha}$$

for arbitrary values of the coefficients. There is, however, no loss of generality in supposing that these coefficients satisfy $u_{ij} = u_{ji}$, and this we shall do. Evidently

$$g = u_{ij} s_{ij} .$$

Furthermore,

$$Eg = u_{ij} \sigma_{ij} ,$$

and since absence of bias in the statistic means that its expectation must equal the parameter estimated, it follows that

$$u_{ij} \sigma_{ij} = \gamma , \qquad \text{or} \qquad u_{ij} t_{ij} = 1 .$$

Subject to this condition, we seek the values of the u_{ij} that will minimize the variance of g, which is

$$\begin{aligned} \sigma_g^2 &= u_{ij} u_{km} E \, d s_{ij} d s_{km} \\ &= \frac{1}{n} u_{ij} u_{km} (\sigma_{ik} \sigma_{jm} + \sigma_{im} \sigma_{jk}) \\ &= \frac{\gamma^2}{n} u_{ij} u_{km} (t_{ik} t_{jm} + t_{im} t_{jk}) . \end{aligned}$$

In the last term the indices of summation k and m may be interchanged, since $u_{km} = u_{mk}$, and, therefore,

$$\sigma_g^2 = \frac{2\gamma^2}{n} u_{ij} u_{km} t_{ik} t_{jm} .$$

Differentiating with respect to u_{ij} to obtain the minimum under the assigned condition gives

$$u_{km} t_{ik} t_{jm} = a t_{ij} ,$$

where a is a Lagrange multiplier. Multiplying by $q_{ih} q_{jr}$ and summing with respect to i and j gives

$$u_{km} \delta_{hk} \delta_{mr} = a \delta_{hj} q_{jr} ,$$

that is,

$$u_{hr} = a q_{hr} .$$

The condition that g shall be unbiased leads to the evaluation of a as $1/p$, whence $u_{ij} = q_{ij}/p$, and finally

$$g = \frac{1}{p} q_{ij} s_{ij} = \hat{g} .$$

That this procedure yields an actual minimum follows from the facts that σ_g^2 is a quadratic form in the variables u_{ij}; and that, because it is a variance, it is essentially positive, and is, therefore, a positive definite quadratic form, which, with or without a linear constraint, must necessarily be a minimum where the first order conditions are satisfied.

The exact result

$$\sigma_{\hat{g}}^2 = \frac{2\gamma^2}{np}$$

follows by direct substitution from the formulae just used.

8. Efficiency of mean square circular error

The mean square deviation from the target of n bombs yields an estimate g for which $u_{ij} = \delta_{ij}$. The efficiency of any unbiased estimate g is

$$\frac{\sigma_{\hat{g}}^2}{\sigma_g^2} = \frac{1}{p u_{ij} u_{km} t_{ik} t_{jm}} .$$

Putting $u_{ij} = \delta_{ij}$ and $p = 2$ gives for the efficiency of the mean square circular error

$$(2t_{jk}t_{jk})^{-1} ,$$

a quantity which in accordance with the preceding theorem cannot exceed unity. Since the absence of bias implies that $u_{ij}t_{ij} = 1$, we have in this case that $\delta_{ij}t_{ij} = 1$. Now if we have numbers t'_{ij} proportional to the covariances, so that $t'_{ij} = ct_{ij}$, it follows that $c = \delta_{ij}t'_{ij} = \sum_i t'_{ii}$. Therefore, the efficiency of the mean square circular error is

$$\frac{\left(\sum t'_{ii}\right)^2}{2\sum\sum t'^2_{ij}} = \frac{(\sigma_{11} + \sigma_{22})^2}{2(\sigma_{11}^2 + 2\sigma_{12}^2 + \sigma_{22}^2)} .$$

We may estimate the efficiency of the mean square circular error for determining the general degree of dispersion of bombing by using for the t'_{ij} the covariance estimates obtained from the 260 bombs discussed in [11]. Substituting these in the formula just obtained for efficiency gives .9550. The value .82 given in [11, p. 165] is an arithmetical error, kindly pointed out by J. W. Tukey.

REFERENCES

[1] R. C. Bose and S. N. Roy, "The distribution of the Studentized D²-statistic," *Sankhya*, Vol. 4 (1938), pp. 19–38.

[2] J. F. Daly, "On the unbiased character of likelihood ratio tests for independence in normal systems," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 1–32.

[3] David Durand, *Risk Elements in Consumer Installment Financing*, technical ed., National Bureau of Economic Research, New York, 1941. Appendices A, B, C.

[4] R. A. Fisher, "The statistical utilization of multiple measurements," *Annals of Eugenics*, Vol. 8 (1938), pp. 376–386.

[5] ———, "The sampling distribution of some statistics obtained from non-linear equations," *Annals of Eugenics*, Vol. 9 (1939), pp. 238–249.

[6] Harold Hotelling, "The generalization of Student's ratio." *Annals of Math. Stat.*, Vol. 2 (1931), pp. 360–378.

[7] HAROLD HOTELLING, "Analysis of a complex of statistical variables into principal components," *Jour. Educational Psychology*, Vol. 24 (1933), pp. 417–441 and 498–520.

[8] ———, "Relations between two sets of variables," *Biometrika*, Vol. 28 (1936), pp. 321–377.

[9] ———, "Air testing sample bombsights," prepared for the Applied Mathematics Panel, National Defense Research Committee, by the Statistical Research Group, Columbia University. SRG Report No. 328, AMP Report No. 5.1R. November, 1944. Originally classified "Confidential" but later declassified. (Believed to be out of print.)

[10] ———, "A generalized T measure of multivariate dispersion," abstract, *Annals of Math. Stat.*, Vol. 18 (1947), p. 298.

[11] ———, "Multivariate quality control, illustrated by the air testing of sample bombsights," *Selected Techniques of Statistical Analysis*, edited by C. Eisenhart, M. W. Hastay and W. A. Wallis, Chap. 3, McGraw-Hill, New York, 1947.

[12] ——— and L. R. FRANKEL, "The transformation of statistics to simplify their distribution," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 87–96.

[13] P. L. HSU, "Notes on Hotelling's generalized T," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 231–243.

[14] ———, "On the distribution of roots of certain determinantal equations," *Annals of Eugenics*, Vol. 9 (1939), pp. 250–258.

[15] M. G. KENDALL, *The Advanced Theory of Statistics*, Vol. 2, Griffin, London, 1948.

[16] P. C. MAHALANOBIS, "On tests and measures of group divergence," *Jour. Asiatic Soc. of Bengal*, Vol. 26 (1930*a*), pp. 541–588.

[17] KARL PEARSON, "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *Phil. Mag.*, Vol. 50 (1900), pp. 157–175.

[18] ———, "On the coefficient of racial likeness," *Biometrika*, Vol. 18 (1926), pp. 105–118.

[19] ———, (ed.), *Tables of the Incomplete Beta Function*, Biometrika Office, London, 1934.

[20] A. M. PEISER, "Asymptotic formulae for significance levels of certain distributions," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 56–62.

[21] V. ROMANOVSKY, "On the criteria that two given samples belong to the same normal population," *Metron*, Vol. 7, No. 3 (1928), pp. 3–46.

[22] S. N. ROY, "p-statistics, or some generalizations in analysis of variance appropriate to multivariate problems," *Sankhya*, Vol. 4 (1939), pp. 381–396.

[23] J. B. SIMAIKA, "On an optimum property of 2 important statistical tests," *Biometrika*, Vol. 32 (1941), pp. 70–80.

[24] A. WALD, "On a statistical problem arising in the classification of an individual into one of two groups," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 145–162.

[25] ———, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 426–482.

[26] S. S. WILKS, "Certain generalisations in the analysis of variance," *Biometrika*, Vol. 24 (1932), pp. 471–494.