# A TEST OF SIGNIFICANCE FOR MULTIPLE OBSERVATIONS

BY

PROF. D. D. KOSAMBI

(*Fergusson College, Poona*)

1. A test of significant discrimination between two sample-groups of multivariate observations can be made by Hotelling's extension[1] of Student's $t$-test; by R. A. Fisher's discriminating function[2] based on the multiple correlation coefficient; the generalized distance[3] of Mahalanobis, Bose and Roy. In addition to these closely related $T^2$, $R^2$, $D^2$ tests, Wilks[4] has suggested others which would not involve the group means entering into the first three; but these last, as well as $D^2$ necessitate new sets of tables. For the case of two variates, however, it has been shown[5] that the usual analysis of variance can be carried out exactly, using the $z$-tables of Fisher, provided the degrees of freedom are suitably readjusted.

Here, I propose to extend the $z$-test partially to samples drawn from a normally distributed population in $p > 2$ linearly independent variates. I also consider briefly the limiting case in which the number of variates increases beyond any limit, which leads us to discrimination between samples consisting of sets of whole curves. This has the advantage of theoretical simplicity, in that all finite dimensional normal distributions are special cases, in much the same way as polygonal area rules like Simpson's come under the general $\int y dx$ formula. If accepted, the method would extend analysis of variance to such material as electrocardiograms, cranial shapes, temperature curves and the like. It is emphasized that the discrimination is performed by the best linear combination of the old variates, and not by the characteristic roots as such that appear in the process.

The contents of the opening chapters of Courant-Hilbert: *Methoden der Mathematischen Physik I* (1931) are taken for granted in the deduction.

2. We use the tensor summation convention: a repeated index denotes summation over all possible values of the index. The variates 1, 2, ..., $p$ are indicated by Greek indices; sampling values 1, 2, ..., $n$ of each variate by an additional Latin index. Thus $x_{\nu i}$ is the $i$th sample value of the $\nu$th variate. Without loss of generality, the population mean for each variate is taken as zero. The multivariate normal distribution has then the probability density $c\,exp\text{-}\phi/2$ where $\phi$ is a positive definite quadratic form in the $p$ variates, $c$ a constant so chosen as to make the total probability over the whole $p$-space equal to unity.

There exist infinitely many linear homogeneous transformations of the variates reducing $\phi$ to a sum of squares:

$$\phi = \sigma^{\alpha\beta} x_\alpha x_\beta = \delta^{\alpha\beta} y_\alpha y_\beta;$$
$$\delta^{\alpha\beta} = 0,\ \alpha \neq \beta;\ = 1,\ \alpha = \beta. \tag{2.1}$$
$$y_\alpha = a^\nu_\alpha x_\nu,\ |a^\mu_\nu| \neq 0;\ \sigma^{\alpha\beta} = \mathcal{E}^{\mu\nu} a^\alpha_\mu a^\beta_\nu.$$

The new variates $y$ are therefore uncorrelated, each with unit variance. The method of discrimination proposed is that of applying the $z$-test in that particular one of the hypothetical $y$ variates for which the observed samples give a maximum value of $z$. Let this be $y_\lambda$. For a sample of $n$ observations, we have:

$$\frac{1}{n}\sum_{i=1}^{n} y_{\lambda i} = \bar{y}_\lambda = \bar{x}_\nu a^\nu_\lambda, \text{ where } \bar{x}_\nu = \frac{1}{n}\sum_{i=1}^{n} x_{\nu i};$$
$$\frac{1}{n-1}\sum_{i=1}^{n} (y_{\lambda i} - \bar{y}_\lambda)^2 = \frac{1}{n-1}\sum_{i=1}^{n} \{a^\nu_\lambda (x_{\nu i} - \bar{x}_\nu)\}^2 \tag{2.2}$$
$$= a^\nu_\lambda a^\mu_\nu s_{\mu\nu};$$
$$\text{where } s_{\mu\nu} = s_{\nu\mu} = \frac{1}{n-1}\sum_{i=1}^{n} (x_{\nu i} - \bar{x}_\nu)(x_{\mu i} - \bar{x}_\mu).$$

The tensors $s_{\mu\nu}$, $s'_{\mu\nu}$ are unbiassed estimates of the normalized cofactors of the population tensor $\sigma^{\alpha\beta}$, calculated from $n$, $n'$

random multiple observations respectively. Nothing is to be assumed known as to the actual values of $\sigma^{\alpha\beta}$ or of the normalizing transformation coefficients $a^{\mu}_{\lambda}$.

3. We now take a new vector variable $u^{\alpha} = a^{\alpha}_{\lambda}$, since $\lambda$ is to be fixed for the problem in hand. The two quadratic forms $s_{\alpha\beta}\, u^{\alpha}u^{\beta}$, $s'_{\alpha\beta}\, u^{\alpha}u^{\beta}$ are positive definite because all principal determinants in any sampling matrix $||s_{\alpha\beta}||$ calculated as in (2.2) are Gram determinants, which are positive whenever the $p$ variates are linearly independent. Our special discrimination problem is thus reduced to finding the maximum of $F = s'_{\alpha\beta}u^{\alpha}u^{\beta}/s_{\mu\nu}u^{\mu}u^{\nu}$ or of its reciprocal.

The answer to this is well known. All we need here is the greatest relative characteristic root of the two forms, *i.e.*, of the determinantal equation

$$(3\cdot1) \qquad \text{det.}\ |s_{\alpha\beta} - \vartheta s'_{\alpha\beta}| = 0,$$

or of the reciprocal equation, interchanging $s, s'$. These roots are all positive. If arranged in descending order of magnitude, they have the minimax property: $\vartheta^{\nu}$, $1 < \nu \leqslant p$, is the smallest value assumed by the maximum of F when the $u$ are subjected to $\nu - 1$ independent linear homogeneous restrictions. Thus, all we have to do here is to put $z = \frac{1}{2}|\log\vartheta|$ for the extreme root, using the $z$-tables of Fisher with degrees of freedom based on the samples alone, as for the single variate. The distribution of the greatest or of any other characteristic root does not enter into the argument, the ratio of the two hypothetically transformed quadratic forms being always that of two sample-variances. What we have obtained is essentially an existence theorem to the effect that the change by means of a suitable linear transformation of co-ordinates (variates) can give a $z$-value as great as but no greater than the greatest relative characteristic root of the two sampling tensor matrices. So, the $z$-tables are to be entered with degrees of freedom one less than the number in the samples, in the absence of any other linear restriction on the variates than that incurred in measuring from the sample mean. It might be possible to use the other roots by compounding probabilities, but it must be kept in mind that the minimax property requires that our transformation coefficients, not the variates, be sufficiently unrestricted. For example, our method of deduction cannot be called valid for $p = 1$, $p = 2$, as there are then not enough of the $a^{\mu}_{\nu}$ left free, for a maximum to exist necessarily, after reducing the population form to a sum of squares. Of course, this is immaterial in view of the fact that $p = 1$ is trivial and $p = 2$ settled by means of a special device.[5] In each of these cases it is true that no *greater* $z$-discrimination is possible with linear combinations than is indicated by our test.

4. One advantage of the extension is that it holds for any $p > 2$. The ordinary analysis of variance is to be carried out exactly, in view of the fact that any sampling matrix may be broken up into various additive components due to the sources between which one wishes to discriminate. There is the further advantage that in case significant discrimination has been shown, the residual matrix of $||s_{\mu\nu}||$ may be used as the fundamental matrix in Hotelling's $T^2$ in the same way that the residual estimate of variance is used for Student's $t$ after analysis of variance in a single dimension. The disadvantage is that our test would not be so powerful as others in rejecting $H_0$ when it is false; $H_0$ here being the null hypothesis that the various sampling tensors are pairwise compatible estimates of the same population tensor.

One method of calculating the extreme root has been given by Fisher (*SMRW* ex. 46.2) who uses divided differences. But equation (3.1) also lends itself to approximation for the greatest root by means of root-squaring. Where the greatest root is not multiple, the rule can be stated immediately, without going into the very simple proof. We define: $\Delta = |s_{\alpha\beta}|$; $\Delta' = |s'_{\alpha\beta}|$; $\Theta$ is the sum of the $p$ determinants formed by substituting in rotation a single row in $||s_{\alpha\beta}||$ by the corresponding row of $||s'_{\alpha\beta}||$, and $\Theta'$ the same function interchanging $s, s'$. Finally, let $\Delta_m$, $\Delta'_m$, $\Theta_m$, $\Theta'_m$ be the corresponding functions constructed by squaring (iteration) $m$ times, according to the rule for matrix multiplication, each of the two matrices. Then an approximate value of $z$ for maximal significance is the greater of

$$(4.1) \qquad \frac{1}{2^{m+1}} \log\left(\frac{\Theta_m}{\Delta_m}\right) \text{ or } \frac{1}{2^{m+1}} \log\left(\frac{\Theta'_m}{\Delta'_m}\right).$$

Approximation is quite rapid when the greatest root is isolated. For a multiple root the ratio $\Theta/\Delta$ must be divided by a factor corresponding to the multiplicity; a similar precaution should also be taken for roots very close together.

5. Still more interesting is the passage to the limit. Suppose we have to deal with silhouttes taken on the profiloscope. One method would be to take some well-defined point such as the ear orifice for the origin, some well-defined line such as that from the origin to the base of the nose as prime vector, and to expand the distance from the origin to the general point of the profile as a Fourier series in terms of the angle from the prime vector. The co-ordinates would then be the Fourier coefficients; if enough were determined to permit the reproduction of any profile to within the original limits of observation, our test or any suitable multivariate test could be applied directly. Yet this is clearly unsatisfactory in that we are using a finite number of co-ordinates in an indefinite number of dimensions without knowing anything of those discarded. The argument that professional anthropometrists do this or worse in using a finite number of characters instead of our harmonic analyser, without proving normality of the distribution, does not suffice. So, we take the other form of the passage to the limit represented by integral equations.

We keep the original quadratic form, extended to infinitely many dimensions; take the co-ordinates as "Fourier" coefficients associated with expansion in some given set of orthonormal functions defined over $0 \leqslant x \leqslant 1$, which is also to be taken hereafter as the range for all undefined integrations. The probability density will again be represented by $c \exp -\phi/2$, with

$$\phi = \int\int \mathrm{K}(s,t) f(s) f(t)\, ds\, dt\,;\ \bar{f}(s) = \frac{1}{n} \sum_1^n f_i(s) \tag{5·1}$$

$$\mathrm{S}(s,t) = \frac{1}{n-1} \sum_1^n \{f_i(s) - \bar{f}(s)\}\{f_i(t) - \bar{f}(t)\}.$$

These now replace (2.1), (2.2) in the function-space, each multiple observation on the variates being taken to define a function $f(x)$ over $0 \leqslant x \leqslant 1$. For significance tests, the reciprocal to $\mathrm{S}(s,t)$ is the best estimate of the population kernel $\mathrm{K}(s,t)$. An alternative simultaneous visualization of the space is, as before, the Hilbert space of the coefficients in the orthogonal-function expansion of $f_i(x)$ Naturally, it is essential to take the population kernel $\mathrm{K}(s,t)$ as positive, semi-definite or definite; its characteristic functions form the most convenient orthogonal functions to use for theoretical purposes, which amounts to using a quadratic form with diagonal matrix. If the characteristic orthonormal functions do not form a closed set, as many more are to be adjoined as are necessary for closure, taking the additional co-ordinates associated with these extra functions to constitute the orthocomplement to the function manifold of $\mathrm{K}(s,t)$. In probability integrations, these extra co-ordinates will be undetermined, hence to be integrated over the whole of the orthocomplement. This allows all kernels to be considered in a proper function-space, even the degenerate kernels that actually include the ordinary $p$-variate normal distribution; conversely, the $p$-variate case may be considered as associated with a degenerate $\mathrm{K}(s,t)$, by ascribing one function of an arbitrary orthonormal set to each co-ordinate as coefficient. For limits of integration, we use the convenient as well as fashionable terminology of lattice theory, taking $f \smile g$, $f \frown g$ respectively as the functions whose "Fourier" coefficients are the greater and the lesser of the corresponding coefficients in the expansions of $f$ and $g$. Thus, the integration can extend from $f \frown g$ to $f \smile g$, and over the whole of the orthocomplement whenever integration "between" two function-limits $f$, $g$ is to be performed.

6. The trouble with all this is that it has only an appearance of verisimilitude. In a space of infinitely many dimensions, we have as yet failed to define the volume element. If we take the multiple integral over infinitely many dimensions as evaluated by successive iterated integrals in the usual manner, it will be seen that any consistent evaluation making the total probability unity leads in general to zero probability in integrating over any proper sub-manifold of the whole space. One must go much deeper than the intuitive methods of 5. It is seen that if we merely take limits increasing the number of dimensions, the "volume" of a hypercube is 0, 1, or $\infty$; of a hypersphere zero, as the $n$-dimensional

sphere has the volume $2\pi^2 r^n/n\,\Gamma(n/2) \to 0$ as $n \to \infty$.

This difficulty is surmounted under the hypotheses that the abstract space under consideration has a distance relationship obeying the usual postulates; is separable, locally compact, with a congruence relation. The two middle ones have to remain assumptions, distance $r$ being defined by $r^2=\phi(f-g)$, for any two elements $f$, $g$. The space has to be restricted to elements for which $K(s,t)$ is a positive definite kernel. Congruence of two regions may be taken as transformability of one region into the other by some member of a suitably restricted (linear) transformation group, preserving $\phi(f-g)$ and transforming the entire manifold into the entire manifold. Then a Haar measure[7] and a Lebesgue-Stieltjes integral exist. Unrestricted Hilbert space is not locally compact because no infinite sequence of orthogonal functions can converge in $L^2$.

It follows that all classical results can be stated and proved again in general abstract spaces, though it is better for our purpose to take kernels of the second (Fredholm) kind for some theorems, which means only the addition of a term $\int f^2 ds$ to the $\phi$ of (5.1). We may then state such results as: *The sum of two normally distributed variates is also normally distributed with mean the sum of the two means and kernel whose (formal) reciprocal is the sum of the two (formal) reciprocals of the given kernels.*

Many fundamental procedures and distributions may be generalized to this space, including some of the more powerful tests considered by P. L. Hsu.[6] Not only can the Hotelling-Fisher formulæ[2] be derived from a degenerate population kernel of p degrees of freedom, but a space of sufficiently large (or infinite) number of dimensions would lead to corresponding formulæ with $p=n$, the degrees of freedom within groups. It is clear, however, that the nature of the fundamental abstract space associated with a given population will not be revealed in general by means of the sample taken by a practising statistician; here, I regret my inability to demonstrate with a practical example, for which there is data enough but no access to the necessary machines: ordinary or cinema integraph, differential analyser, etc. In any case, it is clear that a test which applies independently of dimentionality,[8] without new tables, becomes of importance whether or not more efficient and powerful tests could be devised for the particular unknown population in question. This test is the analogue of (3.1); taking limits, we state it as the problem of locating the extreme characteristic root of $\int \{S[s,t] - \vartheta S'[s,t]\} f dt = 0$. *By* noting that the sample kernels S, S′ are degenerate, this can be reduced to a set of linear equations in a finite number of unknowns, whence the existence of a finite number of positive determinate roots follows at once. It is proposed that the extreme root be used as before for the $z$-tests; the estimating kernels may still be broken up into additive components, permitting analysis of variance. It would, of course, be convenient to have the distribution of certain sampling functions, as for example of $\iint S^{-1}S'\,ds dt$, where $S^{-1}$ is the reciprocal to $S(s,t)$.

---


1 H. Hotelling, "The Generalization of Student's Ratio," *Annals of Mathematical Statistics*, **1931**, **2**, 360–78.

2 R. A. Fisher, *Statistical Methods for Research Workers*, 1938, 7th ed., 294–98.

3 P. C. Mahalanobis, *Proc. Nat. Inst. Sci. India*, **1936**, **2**, 49–55; R. C. Bose., *Sankhyā*, **1936**, **2**, 143–54, 379–84; S. N. Roy, *Ibid.*, 385–96.

4 S S. Wilks, "Certain Generalizations in the Analysis of Variance," *Biometrika*, **1932**, **24**, 471–94.

5 D. D. Kosambi, "A Bivariate Extension of Fisher's $z$-Test," *Cur. Sci.*, **1941**, **10**, 191–92.

6 P. L. Hsu, *Biometrika*, **1940**, **31**, 221–37; *Annals of Mathematical Statistics*, **1939**, **9**, 231–43; *J. London Math. Soc.*, **16**, **1941**, 183–94.

7 Stefan Banach, in S. Saks, *Thèorie de l'Intègrale*, **1933**, 2(4–72.


8 If the Haar volume of the sphere $\phi(f) \leqslant r^2$ is $cr^k$, we have the usual $k$-dimensional space or its equivalent. But we also get fractional dimensionality when $k$ is non-integral. So, the degenerate kernel need not necessarily lead to the ordinary $p$-dimensional case. For the existence and construction of point-sets with fractional dimension, see F. Hansdorff, *Math. Annalen*, **1919**, **79**, 157–79.