SUMMARY
MULTIVARIATE METHODS IN TESTING COMPLEX EQUIPMENT

by

Harold Hotelling
University of North Carolina, Chapel Hill

Institute of Statistics
Mimeograph Series No. 106
May 1954

SUMMARY
MULTIVARIATE METHODS IN TESTING COMPLEX EQUIPMENT
By Harold Hotelling, University of North Carolina, Chapel Hill
May 1954

        This paper deals with certain methods of comparing groups of objects on
each of which a definite set of p quantitative measurements are made. The ob-
jects may be bombsights (References $[2]$, $[3]$) each of which was used to
drop four bombs on which range and deflection errors were measured; or samples
of powder, with muzzle velocity and average pressure in the barrel as measures;*
or persons for whom two or more physical, mental or biographical characters are
to be considered together; or wheat plants (Ref. $[5]$) each of which provided
five different measures of moisture content, one for each of the first five
leaves. In each case the purpose is to compare two or more samples, differen-
tiated from each other by some difference in material content, experimental
treatment, or other known conditions, while at the same time deciding whether
the differences are big enough so that they must be ascribed to the differentia
or are only such as can be ascribed to the kind of random chance represented
by the variations within a supposedly homogeneous group. For manufactured
articles coming from an assembly line, on each of which p predetermined measure-
ments are taken, the differentia is either between a new batch and an initial
lot used as the basis for setting up the system, or between the new batch and
some specified combination of means and tolerances. For the wheat plants the
moisture content was compared between plants that had and had not flowered.
Other possible applications are to agricultural experimentation on plants such
as cotton and flax, for which both seed and fibre are valuable and for which
the variety, cultivation method and fertilizer most suitable to particular
localities are to be determined; and to analogous chemical processes giving
more than one product.

        The problem is to generalize the analysis of variance to two or more
dimensions. In the present state of knowledge, and apparently for a long time
to come, any exact solution must be based on an assumption of a normal distri-
bution. This is predicted by certain mathematical theorems where there are
numerous independent causes of variation, and verified, usually rather roughly,
in many fields, but quite closely in test bombing, in errors of skilled and
careful physical measurements, and in much biological variation. If normality
is untenable, a satisfactory normalizing transformation can sometimes be found;
otherwise non-parametric methods must be used though these waste information
if the distribution is known to be practically normal.

        All recommended methods involve use of roots or functions of roots of a
determinantal equation in λ,

$$|S^* - \lambda S| = 0 ,$$

where S and S* are p-rowed square matrices. The greatest and least roots provide
one type of multivariate analysis of variance, which has been studied by
S. N. Roy, M. S. Bartlett and others. The product of the roots equals the
ratio of the determinant of S* to that of S; its distribution was obtained by
S. S. Wilks $[6]$ in 1932. The present paper is however concerned with methods
using the sum of the roots.
*J. S. Hunter has applied methods described in this paper to compare samples of
powder at Aberdeen Proving Ground.

Student's t, to whose use the analysis of variance reduces in the comparison of two groups with respect to the mean of a single variate, or of the mean of one group with a hypothetical value, was generalized in $\underline{/}1\underline{/}$ to a statistic T suitable for such comparisons where two or more variates are measured in each case. Corresponding to an estimate of error variance s^2, with the number of degrees of freedom n in the denominator, covariance estimates s_{ij} between the ith and jth variates may be formed. The square p-rowed symmetric matrix $S = \underline{/}s_{ij}\underline{/}$ will generally have an inverse, whose elements we denote by ℓ_{ij}: $S^{-1} = \underline{/}\ell_{ij}\underline{/}$. We define linear functions $z_1, \ldots, z_p$ according to the purpose in view. If the object is to test whether there is any real difference in means between two independent samples of N with means $\bar{x}_i$ and of N* with means $\bar{x}_i*$, we have as in the one-dimensional case treated by Sir Ronald Fisher,

$$z_1 = \frac{\bar{x}_i - \bar{x}_i*}{\sqrt{\frac{1}{N} + \frac{1}{N*}}} \quad ,$$

while the sums of products of deviations from the means, like the sums of squares, are usually to be pooled to obtain S, which then has n = N + N* - 2 degrees of freedom. Exact distribution of $T^2 = \Sigma \Sigma \ell_{ij} z_i z_j$ were found in $\underline{/}1\underline{/}$, $\underline{/}7\underline{/}$ and $\underline{/}8\underline{/}$; whether the probability of a greater value by chance is less than .01 or .05 may be ascertained by various methods, of which the most convenient at present is to use a table of 1 per cent or 5 per cent points of the variance ratio F, with

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

There are also many other uses of T, generalizing those of t in testing means and regression coefficients, and in finding confidence regions with specified exact probabilities of covering the unknown parameter point. Certain optimum properties of T were found by Hsu $\underline{/}8\underline{/}$.

In quality control a new article or a small group of new articles is ordinarily compared with norms such as mean, variance, or proportion defective in some earlier experience such as a lengthy series of such articles recorded at or before the time of establishment of the quality control system for the particular article. This earlier series we shall refer to as the "old sample." For one normally distributed measurement on each article, the t distribution may be used to obtain (say) 5 percent control limits for comparing a new measurement with the mean of those in the old sample. A new sample of one is then being compared with an old sample of, say, N, so that the estimate of variance is

derived entirely from the old sample, with n=N-1 degrees of freedom. For several newly produced articles the variance may also be compared with that in the old sample, using the variance ratio distribution, and this may be done by means of control limits with known probability.

In a wartime study of acceptance sampling of bombsights by dropping four test bombs with each, records of a preliminary run under what were taken to be standard conditions were available for estimating the variances in range and deflection, and the covariance of range and deflection. A T test with $p = 2$ is thus possible for the range and deflection error of each new bomb. With 260 bombs in the old sample, n was so large that the measure

$$T_B^2 = \Sigma\Sigma \, \ell_{ij} x_{iB} x_{jB}$$

(where x_{1B} = range error and x_{2B} = deflection error for the new bomb and the summations are over the values 1 and 2 of i and j) has very nearly the χ^2 distribution with 2 degrees of freedom, which would be exact if $\left[\ell_{ij}\right]^{-1}$ were the true covariance matrix.

A combined measure of accuracy for the four bombs dropped with a bombsight is obtained by adding together the four values of T_B^2. The result T_o^2 is an over-all measure of accuracy (or rather inaccuracy) which may also be written

$$T_o^2 = 4 \, \Sigma\Sigma \, \ell_{ij} \, s_{ij}^*$$

by putting

$$s_{ij}^* = \frac{1}{4} \sum_{B=1}^{4} x_{iB} x_{jB} \, ,$$

which is an estimate of covariance of the standard type, with 4 degrees of freedom, if the range and deflection errors are unbiased, i.e. have zero expectations. For large n, T_o^2 has approximately the χ^2 distribution with 8 degrees of freedom. This over-all measure of dispersion may be separated into a part depending only on the mean point of impact (centroid) of the four bombs,

$$T_M^2 = 4 \, \Sigma\Sigma \, \ell_{ij} \, \bar{x}_i \, \bar{x}_j \, ,$$

(approximately χ^2 with 2 d. f.), and a part depending only on the deviations from the mean point of impact,

$$T_D{}^2 = 3 \ \Sigma\Sigma \ \ell_{ij} \ s_{ij}{}^{**} \ ,$$

(approximately χ^2 with 6 d. f.), where $s_{ij}{}^{**}$ is a standard-type estimate of covariance with 3 d. f. The identity

$$T_o{}^2 = T_M{}^2 + T_D{}^2$$

holds. Excessive values of $T_M{}^2$ suggest certain kinds of defects in the bomb-sight, excessive values of $T_D{}^2$ certain other kinds. To combine evidence from several flights, for example those made with a particular pilot-bombardier team, it is possible to add together the corresponding values of $T_o{}^2$, or of $T_M{}^2$ or $T_D{}^2$.

Of the three statistics T_o, T_M and T_D only T_M has the original T distribution. The others have a new distribution. Generalizing the meaning assigned above to T_o we may write a class of such statistics in the form

$$T_o{}^2 = m \ \Sigma\Sigma \ \ell_{ij} \ s_{ij}{}^* \ ,$$

where the summations are with respect to i and j from 1 to p, and $\underline{/} s_{ij}{}^* \underline{}\underline{/} = S^*$ is a standard-type matrix of unbiased covariance estimates with m degrees of freedom. Three integers enter into the definition and distribution of this generalized T statistic: the number p of dimensions or variates, the number n of degrees of freedom in the old sample, and the number m of degrees of freedom in the new sample. For the over-all measure of the bombsight study, m = 4, while m = 3 for T_D and m = 1 for T_M.

The exact distribution of T_o for p = 2 when the old and the new samples come from the same bivariate normal population was obtained in 1944; cf $\underline{/}2\underline{}\underline{/}$ or $\underline{/}4\underline{}\underline{/}$. The probability of a greater value is

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

where $w = T_o{}^2 / (2n + T_o{}^2)$ and I_w is the incomplete beta function tabulated by Karl Pearson. I understand that Dr. Frank E. Grubbs and his staff have now

completed a table of the .01 and .05 points of this distribution, using the ENIAC, and I hope this will soon be published. The range of possible applications is very large.

No usable general form of the distribution is available for p greater than 2, though we have been studying this problem for several years at Chapel Hill and hope to have something before long. One result now in hand supplies upper and lower bounds for the error of approximation in ascribing to T_o^2 the χ^2 distribution with mp degrees of freedom, and makes it clear that these errors are quite small over an important range of values.

Since T_o^2/m is by its definition the trace of the matrix $S*S^{-1}$, it equals the sum of the characteristic roots of this matrix, which are also the roots of the determinantal equation $|S* - \lambda S| = 0$. The joint distribution of these roots was discovered in 1939 by Hsu, Fisher, Roy and Girshick, and was used in obtaining the distribution of T_o^2 for p = 2. The use of the t statistics thus involves choice of one particular function of these roots, the sum. It must not be forgotten that other functions, such as the product, the maximum and the minimum may be most appropriate for certain purposes; but this paper cannot enter into them. Also, it will be observed that T_0 is not symmetrical as between the "old" and "new" samples; if we were to interchange them we should be dealing not with the sum of the roots but with the sum of their reciprocals. The choice of method must depend on the ultimate purpose, and take account of the types of deviation to which it is most important to make the test sensitive. If, for example, the purposes are such that definite relative prices can be assigned to two or more of the variates, these should be used to reduce the number of variates. The bombsight example and certain others, together with general considerations of the kind discussed in [4] and [8], suggest that T_o may have an extensive usefulness.

ABBREVIATED BIBLIOGRAPHY

I. Papers by the author

1. "The generalization of Student's ratio," Ann. Math. Stat., Vol. 2 (1951), pp. 360-378.

2. "Air testing sample bombsights," Applied Mathematics Panel, Statistical Research Group. SRG Report 328. Confidential, later declassified. November, 1944.

3. "Multivariate quality control, illustrated by the air testing of sample bombsights." Chapter 3, pp. 111-184 of Selected Techniques of Statistical Analysis, McGraw-Hill, 1947.

4. "A generalized T test and measure of multivariate dispersion," Proceedings of the second Berkeley symposium on mathematical statistics and probability. Pp. 23-42. Univ. of Cal. Press. 1951.

5. "Multivariate analysis." To be published in a volume of 1952 symposium lectures by the Collegiate Press, Ames, Iowa.

II. Papers by other authors (partial list only)

6. S. S. Wilks, "Certain generalisations in the analysis of variance," Biometrika, Vol. 24, pp. 471ff.

7. R. C. Bose and S. N. Roy, "The distribution of the Studentised D^2-statistic," Sankhya, Vol. 4 (1938), pp. 19ff.

8. P. L. Hsu, "Notes on Hotelling's generalised T," Ann. Math. Stat. Vol. 9 (1938), pp. 231ff.