The Philippine Statistician
Volume XVI, Nos. 3-4, 1967

A UNIFIED TABLE FOR TESTING UNIVARIATE AND MULTIVARIATE HYPOTHESES OF VARIOUS TYPES

By T. A. MIJARES[1]

1. **The Null Distribution of the Roots of an Equation.** A useful statistic for testing various types of hypotheses in univariate and multivariate analyses may be expresed as a function of the root of a (determinantal) equation

$$|V_1 - \lambda V_2| = 0 , \tag{1}$$

where V_1 and V_2 are independent Wishart matrices of certain d.f. By a Wishart matrix is meant a symmetric matrix that has elements with joint distribution following the well-known distribution derived by Wishart (1928); that is,

$$\frac{|\sigma^{ij}|^{\frac{1}{2}n} \, |v_{ij}|^{\frac{1}{2}(n-k-1)} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{k} \sigma^{ij} v_{ij}\right]}{2^{\frac{1}{2}kn} \, \pi^{k(k-1)/4} \, \Gamma(\frac{n}{2}) \, \Gamma(\frac{n-1}{2}) \cdots \Gamma(\frac{n-k+1}{2})} \tag{2}$$

where (σ_{ij}) is the population covariance matrix with inverse matrix (σ^{ij}), (v_{ij}) is the sample value corresponding to the population value (σ_{ij}), n is the size of the sample and $k \leq n$ is the number of variates.

Another convenient form of (1) is

$$\begin{aligned} 0 = |V_1 - \lambda V_2| &= |(1+\lambda)V_1 - \lambda(V_1 + V_2)| \\ &= \left|V_1 - \frac{\lambda}{1+\lambda}(V_1 + V_2)\right| \\ &= |V_1 - \theta(V_1 + V_2)| , \end{aligned} \tag{3}$$

[1] Director, Bureau of the Census & Statistics and Professor of Statistics, Statistical Center, U.P.

where $\theta = \lambda/(1+\lambda)$. It is about the ($\theta$ or θ's) that we will be concerned with later.

When k = 1, the univariate case, the distribution, (2) reduces to

$$(4) \qquad \frac{\left(\frac{V}{2\sigma}\right)^{\frac{1}{2}n-1} e^{-\frac{V}{2\sigma}}}{\Gamma\left(\frac{n}{2}\right)}$$

which is the χ^2-distribution with n d.f. Then, for two independent chi-squares with v_1 and v_2 d.f., (1) also reduces to

$$(5) \qquad \lambda = \frac{\sigma_1 \ \chi^2_1/v_1}{\sigma_2 \ \chi^2_2/v_2},$$

which is a function of the F-ratio that has the central F-distribution (named after Fisher) obtained by Snedecor when $\sigma^2_1 = \sigma^2_2$; that is, $\lambda = (v_1/v_2)\,F$. Hence, (3) becomes

$$(6) \qquad \theta = \frac{(v_1/v_2)\,F}{1+(v_1/v_2)\,F}$$

when $\sigma^2_1 = \sigma^2_2$ The distribution of θ in (6) is an incomplete beta distribution, or simply, a beta distribution. For the more general case when k = s, the distribution of the non-zero roots of (3) is

$$(7) \qquad \frac{\pi^{s/2} \prod_{i=1}^{s} \Gamma\left(\frac{Q+R+s+i+2}{2}\right)}{\Gamma\left(\frac{Q+i+1}{2}\right)\Gamma\left(\frac{R+i+1}{2}\right)\Gamma\left(\frac{i}{2}\right)} \prod_{i}^{s} \theta_i^{Q/2}\,(1-\theta_i)^{R/2} \prod_{i<j} (\theta_i - \theta_j) \prod_{i=1}^{s} d\theta_i\,,$$

$$0 < \theta_1 \leq \ldots \leq \theta_s < 1\,.$$

This is sometimes called the generalized beta distribution. It has been independently obtained by Fisher (1939), Roy (1939), Girshick (1939) for the case s =2 and Mood (1951). (The derivation of Mood was said to have been done also in 1939.

2. The Beta-Function Parameters for the Distribution of the Mean Sum of the Roots. A function of the roots in θ which are presently being tabulated for percentage points at different levels is the sum of the roots. Sometimes this is known as the "trace" of the associated matrix. We shall designate this by $V_1^{(s)} = \theta_1 + \ldots + \theta_s$, with the subscript 1 in $V_1^{(s)}$ to denote the first elementary symmetric function of the roots in θ. For $s = 1$, $V_1^{(1)} = \theta_1$ which is easily seen to be a beta distribution given by

$$(8) \quad \frac{\Gamma\left(\frac{Q+R+4}{2}\right)}{\Gamma\left(\frac{Q+2}{2}\right)\Gamma\left(\frac{R+2}{2}\right)} \cdot \theta_1^{\frac{Q+2}{2}-1} (1-\theta_1)^{\frac{R+2}{2}-1} d\theta_1, \quad 0 < \theta_1 < 1,$$

with parameters $\frac{Q+2}{2}$, $\frac{R+2}{2} > -1$. Studies on the distribution of $V_1^{(s)}$ indicate that the moments of $V_1^{(s)}/s$ follow very closely a beta distribution with parameters as functions of Q, R and s. If a dan b are the parameters of this beta distribution, the general form of these parameters can be put into

$$(9) \quad 2a = \frac{v_1(v_1 - v_2)}{v_2 - v_1^2}, \qquad 2b = \frac{(1 - v_1)(v_1 - v_2)}{v_2 - v_1^2},$$

where

$$v_1 = \frac{Q + s + 1}{Q + R + 2s + 2}$$

$$\upsilon_2 = \frac{Q+s+1}{s(Q+R+2s+1)(Q+R+2s+2)(Q+R+2s+4)}$$

$$.[R\{Qs + (s^2+s+2)\} + Q^2s + Q(3s^2+4s) + (2s^3+5s^2+3s+2)].$$

When $s=1$, (10) becomes exactly the **1st** and **2nd** moment about the origin of the beta distribution and (9) reduces exactly to (8). As a matter of fact the first four raw moments of the beta distribution (8) can be obtained also by utilizing (9) for $s=1$. More details of the generalization may be found in Mijares (1958).

When $s>1$ numerical sets of values of 2a and 2b have been computed for $Q= -1(1)10(10)60(20)120$ and $R=10(10)20$ and $s=2$ to 50. Since this is a two-moment fitting the first two moments between the true distribution of $V_1^{(s)}/s$ and the approximate beta distribution with parameters given by (9) are in exact agreement. As a measure of accuracy of the approximation, the approximate third and fourth moments have been computed, compared with the true distribution and the differences obtained are found to be practically zero in all cases. Some sets of values mav be found in Mijares (1962).

3. **Pearson Curves and Normal Approximation.** One main difficulty in obtaining percentage points with the use of the beta approximation is that the parameters obtained for 2a and 2b are often fractional. However, fitting by making use of Pearsons's system of curves appears to be more fruitful in obtaining percentage points. The true distribution of the sum of the roots have moment ratios $\beta_1 = \mu_3^2/\mu_2^3$ and $\beta_2 = \mu_4/\mu_2^2$ within the range of values tabulated in Pearson and Hartley (1958) eds., "Percentage points of Pearson curves for given β_1, β_2 expressed in standardized measure," **Biometrika Tables for Statisticians Volume 1.** The table gives upper and lower percentage point at different levels for values of $\beta_1 = 0\,(.01)$ and $\beta_2 = 1.8\,(.2)\,5$.

For values of the moment ratios beyond the table, it was found that the beta-function parameters a, b are quite large.

Normal approximations to $V_1^{(s)}$ are convenient in these cases for obtaining percentage points.

4. **The Unified Table.** The unified table is entitled "Percentage points of the sum $V_1^{(s)}$ of s roots ($s = 1 - 50$)." For $s = 1$, the lower percentage points for levels .5%, 1%, 2.5%, and 5% computed by C. M. Thompson in "Lower percentage points of the incomplete beta function," **Biometrika** vol. 32 (1941) can be utilized. The upper percentage points are easily computed from the table given. For $s = 2$ to 50, upper and lower percentage points have been computed for levels of only 1% and 5%, however. If levels other than those given are required an auxiliary table for obtaining the beta-funotion parameters a, b is given and then the percentage points of $V_1^{(1)}$ can be used to obtain the percentage points which may require a two-way interpolation.

(The format of these tables will be shown during the lecture.)

5. **Some Examples.** 1. For tests on sample correlation squared. r^2:

$$V_1^{(1)} = r^2, \; Q + 2 = 1, \; R + 2 = n - 2,$$

where n = sample size.

2. Forf tests on multiple correlation squared, R^{*2}

$$V_1^{(1)} = R^{*2}, \; Q + 2 = q, \; R + 2 = n - q - 1,$$

where q = number of independent variables and n = sample size.

3. For tests based on Student's-t:

$$V_1^{(1)} = \frac{1}{1 + t^2/n'}, \; Q + 2 = n', \; R + 2 = 1,$$

where n' is the degree of freedom.

4. For tests based on variance ratio—F:

$$V_1^{(1)} = \frac{(\nu_1/\nu_2)F}{1 + (\nu_1/\nu_2)F}, \quad Q + 2 = \nu_1, \quad R + 2 = \nu_2,$$

where ν_1 and ν_2 are the degrees of freedom.

5. For tests based on Hotelling's generalized T^2:

a. One sample problem.

$$V_1^{(1)} = \frac{T^2/n'}{1 + T^2/n'}, \quad Q + 2 = n' - p + 1, \quad R + 2 = p,$$

where $T^2 = Y' S^{-1} Y$; Y is the p-dimensional mean vector, S is the sample covariance matrix and $n' = n - 1$, where n is the sample size.

b. Two-sample problem.

$$V_1^{(1)} = \frac{1}{1 + T^2/(n_1 + n_2 - 2)}, \quad Q + 2 = n_1 + n_2 - 1,$$
$$R + 2 = p,$$

the sample sizes, $\bar{Y}_1$ and $\bar{Y}_2$ are the respective p-dimensional

where $T = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} (\bar{Y}_1 - \bar{Y}_2)' S^{-1} (\bar{Y}_1 - \bar{Y}_2)$, n_1 and n_2 are

mean vectors and S is pooled covariance matrix given by

$$S = \frac{1}{n_1 + n_2 - 2} \left[\sum_{a=1}^{n_1} (Y_{1a} - \bar{Y}_1)(Y_{1a} - \bar{Y}_1)' + \sum_{a=1}^{n_1} (Y_{2a} - Y_2)(Y_{2a} - \bar{Y}_2)' \right]$$

6. Tests on canonical correlation:

$$V_1^{(s)} = \mathrm{tr}\,(S_{12} S_{22}^{-1} S_{21} S_{11}^{-1}), \quad s = \min\,(p.\ q)$$

$$Q = |q - p| - 1, \quad R = n - q - p - 2,$$

where S_{11} = sample covariance matrix among the p-set of variates

$S_{12} = S'_{-21}$ = sample covariance matrix between the p and y-sets of variates

S_{22} = sample covariance matrix among the q-set of variates

7. Tests on equality of two covariance matrices:

$$V_1^{(s)} = \mathrm{tr}\left[S_1 (S_1 + S_2)^{-1}\right], \quad s = \min(p, n_i - 1), \; i = 1, 2,$$

$$Q = n_1 - p - 2, \quad R = n_2 - p - 2 \qquad p \leq n_i,$$

where S = sample covariance matrix for ist sample.

8. Tests on equality of mean vectors in k-samples:

$$V_1^{(s)} = \mathrm{tr}\left[B(B + W)^{-1}\right],$$

$$s = \min(k - 1, p), \quad Q = |k - p - 1| - 1$$

$$R = \Sigma n_i - k - p - 1,$$

where $B = \sum_{j=1}^{k} n_j (\bar{x}^{(j)} - \bar{x})(\bar{x}^{(j)} - \bar{x})'$,

$\bar{x} = \sum_{j=1}^{k} n_j \bar{x}^{(j)} / (\sum_j n_j)$, n_j = sample size of the jth sample,

$$W = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_i^j - \bar{x}^{(j)})(x_i^j - \bar{x}^{(j)})'.$$