The Annals of Mathematical Statistics
1969, Vol. 40, No. 1, 290–298

MULTIVARIATE TWO SAMPLE TESTS WITH DICHOTOMOUS AND CONTINUOUS VARIABLES. I. THE LOCATION MODEL[1]

BY A. A. AFIFI AND R. M. ELASHOFF[2]

University of California at Los Angeles and Harvard University

1. Introduction and definition of the model. In this paper we study a testing problem that arises with vector random variables having both dichotomous and continuous components. We consider a probability model for the dichotomous and continuous variables called the location model (see Olkin and Tate (1961)). Alternative models will be defined in later work. In the location model, the set of dichotomous random variables is represented by a random variable $\mathbf{x}$ with the multinomial distribution $m(\mathbf{x}; \mathbf{p}) = \prod_{j=1}^{d} p_j^{x_j}$ where $\mathbf{x} = (x_1, x_2, \cdots, x_d)'$, x_j is one or zero, $\sum_{j=1}^{d} x_j = 1$, $p_j = \Pr(x_j = 1)$, $\sum_{j=1}^{d} p_j = 1$ and $\mathbf{p} = (p_1, p_2, \cdots, p_d)'$. We denote the continuous variables by $\mathbf{y} = y_1, y_2, \cdots, y_c)'$, and assume that the conditional density of $\mathbf{y}$ given $\mathbf{x}$ is multinormal with mean vector depending upon $\mathbf{x}$ and denoted by $\boldsymbol{\mu}_j = (\mu_{1j}, \mu_{2j}, \cdots, \mu_{cj})'$ when $x_j = 1$, and covariance matrix $\boldsymbol{\Sigma}$ independent of $\mathbf{x}$. The unconditional distribution of $\mathbf{y}$, then, is a mixture of multinormal distributions. For brevity, we call the dichotomous and continuous variables the response variables.

Our problem is to test the null hypothesis, known as the location hypothesis, that the parameters $(\mathbf{p}, \boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_d)$ are equal in two different populations. This testing program grew out of the authors' consulting experience with medical investigators. The following are three examples on which the methods described in this paper were used: (1) Greenblatt et al. (1962) compared several treatments for depression using for the response variables total scores on questionnaires designed to measure intensity of depression, and an overall assessment of psychiatric improvement (yes or no); (2) Elashoff and his colleagues at the Kaiser Multiphasic Screening Program obtained significant differences between non-diabetics and drug diabetics on the vector variable composed of blood pressure, cardiac enlargement (yes or no), serum cholesterol and used this and other results to develop a diabetic profile in the upper age groups; (3) Shubin and Weil (1967) successfully discriminated between shock patients with favorable prognosis and those with unfavorable prognosis using such variables as cardiac output, arterial blood pressure and cyanosis (yes or no). In all three examples the distributional assumption of the location model regarding $\mathbf{y}$ was consistent with the data. The

Received 8 July 1966; revised 16 July 1968.

[1] This work has been facilitated by grants from the Ford Foundation and from the National Science Foundation NSF G-13040, a contract with the Office of Naval Research Nonr 1866(37), a Final Year Fellowship from the Graduate School of Arts and Sciences, Harvard University, by the Clinical Research Center of the Massachusetts Mental Health Center, Public Health Grant GM-10525, University of California, Berkeley, and by U.S.P.H.S.-FR-00122-04.

[2] Now at University of California, San Francisco Medical Center.

behavior of systolic blood pressure and cardiac enlargement is typical. The presence or absence of cardiac enlargement separates individuals into two subpopulations. When the distributions of systolic blood pressure in these subpopulations are plotted side by side, we find that individuals with cardiac enlargement generally have a higher mean systolic blood pressure than those with no cardiac enlargement. The two distributions are approximately symmetrical and have approximately the same variance. After the two distributions are combined, we find a bimodal curve with a long right tail.

In subsequent sections we discuss related work on such testing problems, and then derive an information-theoretic test and a likelihood ratio test for location hypotheses. We derive the small sample distribution theory for these tests and make some brief comparisons between them.

2. Related work. A large literature exists for mixed continuous and dichotomous variable problems for $d = 2$, $c = 1$, but only a few papers consider the problem for c larger than one or d larger than two. Moustafa (1957) studies a multifactor experiment where the response variables follow the distribution specified in Section 1, except that $\mathbf{\Sigma}$ depends upon $\mathbf{x}$ in an arbitrary way. He constructs likelihood ratio tests for location hypotheses and for equality of covariance matrices using, exclusively, the asymptotic theory of these tests that Ogawa, Moustafa and Roy (1957) show to be valid in this situation. Roy and Bhapkar (1959) derive asymptotically nonparametric tests for several location hypotheses by categorizing the continuous variables and representing the resulting variables by the product multinomial. Such tests are consistent against the location hypotheses considered in this paper and utilize the limiting χ^2 approximation to the cumulative distribution function of the test criteria.

Olkin and Tate (1961) have attacked another type of problem with dichotomous and continuous variables. They derive a test for independence between $\mathbf{x}$ and $\mathbf{y}$ by use of canonical correlation theory. Das Gupta (1960) has also studied tests for independence with d arbitrary and has given sound advice on some practical aspects of such tests. Hannan and Tate (1965) consider multivariate biserial problems.

3. Tests for location-hypotheses. We wish to test the null hypothesis

$$H_0: (\mathbf{p}^{(1)}, \boldsymbol{\mu}_j^{(1)}) = (\mathbf{p}^{(2)}, \boldsymbol{\mu}_j^{(2)}), \qquad j = 1, 2, \cdots, d, \tag{3.1}$$

against the alternative hypothesis

$$H_1: (\mathbf{p}^{(1)}, \boldsymbol{\mu}_j^{(1)}) \neq (\mathbf{p}^{(2)}, \boldsymbol{\mu}_j^{(2)}) \qquad \text{for some } j, \tag{3.2}$$

when the location model holds. We discuss the use of Hotelling's T^2 for testing (3.1) against (3.2) with mixed dichotomous and continuous variables. Then in Sections 4 and 5 we derive an information-theoretic test and a likelihood ratio test for this location hypothesis and obtain the exact distributions of the test statistics.

Let

$$n_i = \text{sample size from population } i, \; i = 1, 2;$$

$$r_j^{(i)} = \text{the number of observations from population } i \text{ with } x_j = 1;$$

$$\mathbf{r}^{(i)} = (r_1^{(i)}, r_2^{(i)}, \cdots, r_d^{(i)})';$$

and

$$\bar{r}_j^{(i)} = r_j^{(i)}/n_i; \quad \bar{\mathbf{r}}_{d-1}^{(i)} = (\bar{r}_1^{(i)}, \bar{r}_2^{(i)}, \cdots, \bar{r}_{d-1}^{(i)})'. \tag{3.3}$$

Define

$$\bar{\mathbf{y}}_j^{(i)} = \text{sample mean of } \mathbf{y} \text{ based on observations with } x_j = 1 \text{ from population } i; \tag{3.4}$$

$$\bar{\mathbf{y}}^{(i)} = \textstyle\sum_{j=1}^{d} \bar{r}_j^{(i)} \mathbf{y}_j^{(i)} = \text{sample mean of } \mathbf{y} \text{ for population } i. \tag{3.5}$$

We write T^2 as

$$T^2 = n_1 n_2 (n_1 + n_2)^{-1} (\bar{\mathbf{z}}^{(1)} - \bar{\mathbf{z}}^{(2)})' \mathbf{S}^{-1} (\bar{\mathbf{z}}^{(1)} - \bar{\mathbf{z}}^{(2)}) \tag{3.6}$$

where

$$\bar{\mathbf{z}}^{(i)} = \begin{pmatrix} \bar{\mathbf{r}}_{d-1}^{(i)} \\ \bar{\mathbf{y}}^{(i)} \end{pmatrix}, \qquad \mathbf{S} = \begin{pmatrix} \mathbf{S}_{xx} & \mathbf{S}_{xy} \\ \mathbf{S}_{yx} & \mathbf{S}_{yy} \end{pmatrix},$$

$\mathbf{S}_{xx}$, $\mathbf{S}_{xy}$, $\mathbf{S}_{yy}$ are the covariances matrices of $\mathbf{x}^* = (x_1, x_2, \cdots, x_{d-1})'$, $\mathbf{x}^*$ and $\mathbf{y}$, and $\mathbf{y}$ respectively.

The following theorem gives the null and nonnull distribution of T^2:

THEOREM 3.1. *Suppose the location model obtains. Then, the exact distribution of* T^2 *in* (3.6) *depends upon nuisance parameters. Furthermore, if* $n_1, n_2 \to \infty$ *such that* $n_1/(n_1 + n_2) \to \lambda_1 > 0$, *and* $p_j^{(i)} > 0$, *all* i *and* j, *then the limiting nonnull distribution of* T^2 *when*

$$\mathbf{p}^{(1)} = \mathbf{p}^{(2)} = \mathbf{p} = (p_1, p_2, \cdots, p_d)', \tag{3.7}$$
$$\textstyle\sum_{j=1}^{d} p_j \boldsymbol{\mu}_j^{(1)} = \sum_{j=1}^{d} p_j \boldsymbol{\mu}_j^{(2)}$$

is a central χ^2 *with* $c + d - 1$ *degrees of freedom.*

PROOF. As $n_i \to \infty$, $\mathbf{S}^{-1}$ converges with probability one to

$$\begin{pmatrix} \boldsymbol{\Sigma}_{xx} \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_{yy} \end{pmatrix}^{-1}$$

which always exists for our model. A typical element in $\boldsymbol{\Sigma}_{xy}$, say σ_{jv}, is equal to $p_j[\mu_{jv}(1 - p_j) - \sum_{k \neq j} \mu_{kv} p_k]$. Also, the $\bar{\mathbf{z}}^{(i)}$ is the sample mean vector of independent and identically distributed vector random variables. Hence, by the multivariate central limit theorem given in Anderson (1958), page 74, the asymptotic distribution of T^2 is a noncentral χ^2 with $c + d - 1$ degrees of freedom and noncentrality parameter $(\boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \cdot n_1 n_2 (n_1 + n_2)^{-1}$, where the first $(d - 1)$

components of $\boldsymbol{\mu}$ are $(p_j^{(1)} - p_j^{(2)})$, $j = 1, 2, \cdots, d-1$, and the remaining c components are given by

$$[\textstyle\sum_{j=1}^{d} p_j^{(1)} \boldsymbol{\mu}_j^{(1)} - \sum_{j=1}^{d} p_j^{(2)} \boldsymbol{\mu}_j^{(2)}]. \tag{3.8}$$

Thus, if condition (3.7) holds, T^2 is distributed asymptotically as central χ^2 with $c + d - 1$ degrees of freedom. But, this is also the null asymptotic distribution of T^2. Hence, T^2 is not a consistent test of H_0 vs. H_1. It is clear that the exact distribution of T^2 depends upon the unknown parameters.

4. An information-theoretic test for location hypothesis. We derive a test of the null hypothesis (3.1) against the alternative hypothesis (3.2) using an information-theoretic approach. The information-theoretic method is based upon a particular estimate of the directed divergence (see Kullback (1959)).

$$I(1, 0) = \textstyle\int_{\mathbf{z}} \ln[f_1(\mathbf{z}; \boldsymbol{\theta})/f_0(\mathbf{z}; \boldsymbol{\theta})] \cdot f_1(\mathbf{z}; \boldsymbol{\theta})\, d\mathbf{z} \tag{4.1}$$

between the null-and-alternative-hypothesis distributions of $\mathbf{z}$, $f_0(\mathbf{z}; \boldsymbol{\theta})$ and $f_1(\mathbf{z}; \boldsymbol{\theta})$ respectively. Unbiased, sufficient estimators are substituted for the parameters $\boldsymbol{\theta}$ in (4.1) and the derived expression becomes the test statistic. For our model such estimators always exist. The set of unbiased, sufficient statistics is given by the estimators defined in (3.3) and (3.4) and by

$$\mathbf{S}_i = (s_{iuv});\ u, v = 1, 2, \cdots, c \text{ denote subscripts for the continuous variables} \tag{4.2}$$

$$s_{iuv} = \textstyle\sum_{j=1}^{g_i} \sum_{l\varepsilon j} (y_{iul} - \bar{y}_{ju}^{(i)})(y_{ivl} - \bar{y}_{jv}^{(i)})(n_i - g_i)^{-1}$$

where $j = 1, 2, \cdots, g_i$ denote those j such that $r_j^{(i)} \geqq 1$, $\sum_{l\varepsilon j}$ denotes the summation over all individuals with $x_j = 1$, y_{iul} denotes the observations on the lth individual on variable u in population i, $\bar{y}_{ju}^{(i)}$ is the uth component of (3.4).

NOTE. It might not be possible to estimate all the vectors $\boldsymbol{\mu}_j^{(i)}$ in a given sample, i.e., $x_j = 0$ for all individuals l in the sample from population i. This situation does occur frequently in problems where d is large and the n_i are relatively small. One effect is the non-testability of the null hypothesis $\boldsymbol{\mu}_j^{(1)} = \boldsymbol{\mu}_j^{(2)}$.

We substitute these statistics for the parameters in the expression (4.1) computed under (3.1) and (3.2), and derive the information test statistic

$$\begin{aligned} \hat{I} = \hat{I}(1, 0) &= \textstyle\sum_{j=1}^{g} r_j^{(1)} r_j^{(2)} (2t_j)^{-1} \\ &\quad \cdot (\bar{\mathbf{y}}_j^{(1)} - \bar{\mathbf{y}}_j^{(2)})' \mathbf{S}^{-1} (\bar{\mathbf{y}}_j^{(1)} - \bar{\mathbf{y}}_j^{(2)}) + A(\mathbf{r}^{(1)}) \\ &\quad \text{where both } r_j^{(1)} \text{ and } r_j^{(2)} \geqq 1 \text{ only for } j = 1, 2, \cdots, g \\ &= A(\mathbf{r}^{(1)}) \quad \text{if at least one of } r_j^{(1)}, r_j^{(2)} = 0 \text{ for all } j, \end{aligned} \tag{4.3}$$

where

$$\begin{aligned} A(\mathbf{r}^{(1)}) &= \textstyle\sum_{i=1}^{2} \sum_{j=1}^{d} (r_j^{(i)}) \ln r_j^{(i)}) \\ &= \textstyle\sum_{i=1}^{2} n_i \ln n_i - \sum_{j=1}^{d} (r_j^{(1)} + r_j^{(2)}) \ln (r_j^{(1)} + r_j^{(2)}) \end{aligned}$$

$$+ (n_1 + n_2) \ln (n_1 + n_2)$$

$$t_j = r_j^{(1)} + r_j^{(2)},$$

and

$$\mathbf{S} = (s_{uv}); \qquad u \cdot v = 1, 2, \cdots, c;$$

$$s_{uv} = \sum_{i=1}^{2} (n_i - g_i) s_{iuv} / (n_1 + n_2 - g_1 - g_2).$$

We study the distribution theory of the test statistic in two parts. First, we assume that $d = 2$ and c is arbitrary. Second, we suppose that $c > 1$ and $d > 2$. This breakdown is required to facilitate the use of these statistics.

To proceed with the distribution theory of $\hat{I}$ when $d = 2$, we use this notation: $r^{(i)}$ = number of observation with $x_1 = 1$ in population i, $t = r^{(1)} + r^{(2)}$. Let $f(r^{(1)} | t)$ denote the general term in the hypergeometric distribution with parameter $\rho = p^{(1)}(1 - p^{(2)})/(1 - p^{(1)})p^{(2)}$. Also, let $dF_{k,h,\tau}$ denote the density and differential element of a noncentral F random variable with degrees of freedom k and h and noncentrality parameter τ, and let $dU^{(2)}$ denote the density and differential element for the sum of the latent roots of the determinantal equation

$$|r^{(1)}r^{(2)}t^{-1} \cdot (\bar{\mathbf{y}}_1^{(1)} - \bar{\mathbf{y}}_1^{(2)})(\bar{\mathbf{y}}_1^{(1)} - \bar{\mathbf{y}}_1^{(2)})'$$
$$+ (n_1 - r^{(1)})(n_2 - r^{(2)})(n_1 + n_2 - t)^{-1}(\bar{\mathbf{y}}_2^{(1)} - \bar{\mathbf{y}}_2^{(2)})(\bar{\mathbf{y}}_2^{(1)} - \bar{\mathbf{y}}_2^{(2)})' - \theta \mathbf{S}| = 0,$$

(given by Hotelling (1951) when the null hypothesis holds; otherwise, no explicit density has been derived). We summarize our results in Theorem 4.1.

THEOREM 4.1. *Suppose that the location model obtains and that* $d = 2$. *Then, the cumulative distribution of* $\hat{I}$ *conditional upon* $t = (r^{(1)} + r^{(2)})$, *when* $\mathbf{S}$ *exists and* $p^{(1)}, p^{(2)} > 0$, *takes the following forms*:

(a) *If*

$$2 \leqq t \leqq n_1 \leqq n_2 - 1; \quad \text{or} \quad 2 \leqq n_1 < t < n_2, \tag{4.4}$$

hen

$$\begin{aligned} \Pr (\hat{I} \geqq b \mid t) = {} & f(r^{(1)} = 0 \mid t) \int_{2(b-A(0))/M}^{\infty} dF_{c,(n_1+n_2+c-2)} \\ & + \sum_{r^{(1)}=1}^{t-1} f(r_1 \mid t) \int_{2(b-A(r^{(1)}))}^{\infty} dU^{(2)} \\ & + f(r^{(1)} = t \mid t) \int_{2(b-A(t))/M}^{\infty} dF_{c,(n_1+n_2-c-2)} \end{aligned} \tag{4.5}$$

where $M = (n_1 + n_2 - 3)c/(n_1 + n_2 - c - 2)$.

(b) *If*

$$t = \min (n_1, n_2) = n_1 \geqq 2; \qquad n_1 < n_2, \tag{4.6}$$

then

$$\begin{aligned} \Pr (\hat{I} \geqq b \mid t) = {} & f(r^{(1)} = 0 \mid t) \int_{2(b-A(0))/M}^{\infty} dF_{c,(n_1+n_2-c-2)} \\ & + \sum_{r^{(1)}=1}^{t-1} f(r^{(1)} \mid t) \int_{2(b-A(r^{(1)}))}^{\infty} dU^{(2)} + f^{(1)} = n_1 \mid t)\delta_b^{A(r^{(1)})} \end{aligned} \tag{4.7}$$

where

$$\delta_b^{A(r^{(1)})} = 1 \quad \text{if} \quad A(r^{(1)}) \geqq b$$
$$= 0 \quad \text{otherwise.}$$

The situation for $2 \leqq n_1 < t = n_2$ *follows from* (4.6) *very easily.*

(*c*) *If*

$$2 \leqq n_1 \leqq n_2 < t; \qquad t < n_1 + n_2 , \tag{4.8}$$

then

$$\begin{aligned}\Pr(\hat{I} \geqq b \mid t) = {} & f(r^{(1)} = t - n_2)\cdot \int_{2(b-A(r^{(1)}))/M}^{\infty} dF_{c,(n_1+n_2-c-2)} \\ & + \sum_{r^{(1)}=t-n_2+1}^{n_1-1} \int_{2(b-A(r^{(1)}))}^{\infty} dU^{(2)} \\ & + f(n_1 \mid t) \int_{2(b-A(n_1))/M}^{\infty} dF_{c,n_1+n_2-c-2} .\end{aligned} \tag{4.9}$$

(d) *If*

$$t = n_1 + n_2 ; \qquad 2 \leqq t, \tag{4.10}$$

then

$$\Pr(\hat{I} \geqq b \mid t) = \int_{2(b-A(n_1))/M^*}^{\infty} dF_{c,(n_1+n_2-c-3} \tag{4.11}$$

where $M^* = (n_1 + n_2 - 4)c/(n_1 + n_2 - c - 3)$. *If, in place of* (4.10) *we have*

$$t = n_1 = n_2 ; \qquad 2 \leqq t, \tag{4.12}$$

then

$$\begin{aligned}\Pr(\hat{I} \geqq b \mid t) = {} & 2f(r^{(1)} = 0 \mid t)\cdot \delta_b^{A(0)} \\ & + \sum_{r^{(1)}=1}^{t-1} f(r^{(1)} \mid t)\cdot \int_{2(b-A(r^{(1)}))}^{\infty} dU^{(2)}.\end{aligned} \tag{4.13}$$

Other combinations of t, n_1, n_2 *may be found by a relabeling of populations* 1 *and* 2.

PROOF. We outline the proof of (4.5). Define the following statistics:

$$T_1^2 = r^{(1)} r^{(2)} t^{-1} (\bar{\mathbf{y}}_1^{(1)} - \bar{\mathbf{y}}_1^{(2)})' \mathbf{S}^{-1} (\bar{\mathbf{y}}_1^{(1)} - \bar{\mathbf{y}}_1^{(2)}); \tag{4.14}$$

$$T_2^2 = (n_1 - r^{(1)})(n_2 - r^{(2)})(n_1 + n_2 - t)^{-1} (\bar{\mathbf{y}}_2^{(1)} - \bar{\mathbf{y}}_2^{(2)})' \mathbf{S}^{-1} (\bar{\mathbf{y}}_2^{(1)} - \bar{\mathbf{y}}_2^{(2)}); \tag{4.15}$$

$$T_3^2 = T_1^2 + T_2^2. \tag{4.16}$$

Then $\hat{I} = \frac{1}{2} T_3^2 + A(r^{(1)})$. Since the distribution functions of T_1^2, T_2^2 and T_3^2 are known, we determine the $\Pr(\hat{I} > b)$ by summing the probabilities for the three events

(1) If $0 < r^{(1)} < t$, then, $\hat{I} \geqq b$ if and only if $T_3^2 \geqq 2(b - A(r^{(1)}))$;
(2) If $r^{(1)} = 0$, then $\hat{I} \geqq b$ if and only if $T_2^2 \geqq 2(b - A(0))$;
(3) If $r^{(1)} = t$, then $\hat{I} \geqq b$ if and only if $T_1^2 \geqq 2(b - A(t))$.

The probability that (3) obtains, for example, is given by the last term on the right hand side of (4.5).

The conditions on n_1, n_2, and t define the set of integers over which $r^{(1)}$ and $r^{(2)}$ may vary and, thus, define when the statistics T_1^2, T_2^2 exist.

The details of the distribution theory for the information statistic are too long to present for $d > 2$ and $c > 1$. Due to this fact, we do not give full details; rather, we consider two cases. Before we state the next theorem, let us define $\mathbf{t} = (t_1, t_2, \cdots, t_d)$, and $f(\mathbf{r}^{(1)} \mid \mathbf{t})$ to be the multivariate hypergeometric function given by

$$\begin{aligned}[c_t(\rho)]^{-1} &= \textstyle\prod_{j=1}^{d} \binom{t_j}{r_j^{(1)}} \rho_j^{r_j(1)}, \\ \rho_j &= p_j^{(1)}\,(1 - p_j^{(2)})/(1 - p_j^{(1)})p_j^{(2)}, \\ c_t(\rho) &= \textstyle\prod_{j=1}^{d} \sum_{r_j(1)=0}^{t_j} \binom{t_j}{r_j^{(1)}} \rho_j^{r_j(1)},\end{aligned} \tag{4.17}$$

$dU^{(d-k)}$ to be the density and differential element for the sum of the $(d - k)$ nonzero latent roots of the determinantal equation

$$\left| \textstyle\sum_j [r_j^{(1)} r_j^{(2)}/t_j](\bar{\mathbf{y}}_j^{(1)} - \bar{\mathbf{y}}_j^{(2)})(\bar{\mathbf{y}}_j^{(1)} - \bar{\mathbf{y}}_j^{(2)})' - \theta \mathbf{S} \right| = 0 \tag{4.18}$$

where $\mathbf{S}$ has $(n_1 + n_2 - 2d + k)$ degrees of freedom.

THEOREM 4.2. *If the location model obtains, then the cumulative distribution of* $\hat{I}$ *conditional upon* $\mathbf{t}$ *when* $p_j^{(1)} > 0$ *all* j *may take the following forms*:

(1) *If* $2 \leqq t_j$, *all* j, $\sum t_j \leqq \min(n_1, n_2)$ *where the summation extends over any* $(d - 1)$ *j's, then*

$$\begin{aligned}\Pr(\hat{I} \geqq b \mid t_N) = &\textstyle\sum_{d-1}^{d-1} f(\mathbf{r}^{(1)} \mid \mathbf{t}) \cdot \delta_b^{A(\mathbf{r}^{(1)})} \\ &+ \textstyle\sum_{k=0}^{d-1} \sum_k^d f(\mathbf{r}^{(1)} \mid t) \int_{2(b - A(\mathbf{r}^{(1)})/M_{d-k})}^{\infty} dU^{(d-k)}\end{aligned} \tag{4.19}$$

where $\sum_k^d$ *means the summation over all* $\mathbf{r}^{(j)}$ *such that* d *of the* $r_j^{(1)}$ *are greater than zero and* k *of the* $r_j^{(2)} = 0$, *and*

$$M_{d-k} = c(n_1 + n_2 - 2d + k)/(n_1 + n_2 - 2d + k + 1 - c).$$

(2) *If* $2 \leqq t_j$, *all* j, $n_1 < \min \sum t_j < \max \sum t_j < n_2$ *or* $n_1 \leqq n_2 < \min \sum t_j < \max \sum t_j$, *then*

$$\begin{aligned}\Pr(\hat{I} \geqq b \mid t) = &\textstyle\sum_{h=1}^{d} \sum_h^h f(\mathbf{r}^{(1)} \mid \mathbf{t}) \cdot \delta_b^{A(\mathbf{r}^{(1)})} \\ &+ \textstyle\sum_{k=0}^{d-1} \sum_{h=1}^{d} \sum_{k,h>k}^{h} f(\mathbf{r}^{(1)} \mid t) \int_{2(b - A(\mathbf{r}^{(1)}))/M_{h-k}}^{\infty} dU^{(h-k)}\end{aligned} \tag{4.20}$$

and $M_{h-k} = c(n_1 + n_2 - d - (h - k))/(n_1 + n_2 - d - h + k + 1 - c)$.

(3) *The expression for* $\Pr(\hat{I} \geqq b \mid \mathbf{t})$ *may be obtained for any other condition by arguing along the lines suggested in the derivation to Theorem* 4.1.

An examination of equations (4.19) and (4.20) clearly shows that unless d is small experimenters will not compute the $\Pr(\hat{I} \geqq b \mid t)$ from the right hand side of (4.19) or (4.20); instead, experimenters will want to employ some computationally simple distribution theory for the $\hat{I}$ statistic.

If sample sizes are "large", d is "small," and the $p_j^{(1)}$ are not "extreme," the asymptotic distribution of $\hat{I}$ may give a usable approximation to the exact distribution. It is straightforward to show that $\hat{I}$ is asymptotically distributed as $(\chi^2/2)$ under (3.1) where χ^2 has $(cd + d - 1)$ degrees of freedom; a rigorous proof may be constructed along the lines of Ogawa, Moustafa and Roy (1957).

Another approach to simplifying the computation of the percentage points of $\hat{I}$ in the general situation is to use the device suggested in studying Theorem 4.1 and 5.1 in Olkin and Tate (1961). For our problem, we condition not only on $\mathbf{t}$ but on the relations $r_j^{(1)} \geqq 1$ for all i and all j. Thus, all of the components of (3.1) are testable and we state the distribution theory as follows:

THEOREM 4.3. *If the location model obtains, then the cumulative distribution of* $\hat{I}$ *condition upon* $\mathbf{t}$ *and such that* $r_j^{(i)} \geqq 1$ *all* i *and* j *is given by*

$$\text{(4.21)} \quad \Pr(I \geqq b \mid \mathbf{t}, \mathbf{r}^{(1)} \geqq \mathbf{1}, \mathbf{r}^{(2)} \geqq \mathbf{1}) = \sum_{\mathbf{r}^{(1)}} f^*(\mathbf{r}^{(1)} \mid t) \cdot \int_{2(b-A(\mathbf{r}^{(1)}))/M_d}^{\infty} dU^{(d)}$$

where

$$f^*(\mathbf{r}^{(1)} \mid \mathbf{t}) = f(\mathbf{r}^{(1)} \mid \mathbf{t})/(1 - \sum_{r_j^{(i)}} f(\mathbf{r}^{(1)} \mid \mathbf{t}))$$

$\sum_{r_j^{(i)}}$ *is the sum over all* $r_j^{(i)}$ *such that* $r_j^{(1)}, r_j^{(2)} \geqq 1$ *all* j.

The right hand side of equation (4.21) may be evaluated approximately by methods due to Pillai and Samson (1959). The $\hat{I}$ test is a consistent test of (3.1) against (3.2). Questions of admissibility and asymptotically most powerful unbiasedness are open.

5. The likelihood ratio test. We now apply the likelihood ratio method to test (3.1) against (3.2). The likelihood ratio test statistic is

$$\begin{aligned} L &= \ln \lambda \\ \text{(5.1)} \quad &= ((n_1 + n_2)/2) \ln [1 + \textstyle\sum_{j=1}^{g} r_j^{(1)} r_j^{(2)} t_j^{-1} (\bar{\mathbf{y}}_j^{(1)} - \bar{\mathbf{y}}_j^{(2)})' \mathbf{S}^{-1} (\bar{\mathbf{y}}_j^{(1)} - \bar{\mathbf{y}}_j^{(2)}) \\ &\quad \cdot (n_1 + n_2 - g_1 - g_2)^{-1}] + A(\mathbf{r}^{(1)}) \end{aligned}$$

if at least one $r_j^{(i)} = 0$ for $j = (g + 1), \cdots, d$ in at least one population; g_1, g_2 are the number of categories j in which there is at least one observation in population 1, population 2 respectively; or

$$\text{(5.2)} \qquad L = \ln \lambda = A(\mathbf{r}^{(1)})$$

if at least one of $r_j^{(1)}, r_j^{(2)} = 0$ for all j.

If we observe $L = a$, $T^2 = \sum_{j=1}^{g} T_j^2 (n_1 + n_2 - d_1 - d_2)^{-1}$, $A(\mathbf{r}^{(1)}) = A^0$, then

$$\text{(5.3)} \quad \Pr(L \geqq a \mid \mathbf{t}) = \Pr[T^2 \geqq \exp((a - A(\mathbf{r}^{(1)}))2/(n_1 + n_2)) - 1 \mid \mathbf{t}].$$

Hence, the distribution theory of the likelihood ratio statistic can be obtained in a similar way to that of the $\hat{I}$ statistic.

The likelihood ratio test is not equivalent to the $\hat{I}$ test, since we can write (5.3) in the form

$$\begin{aligned} &\Pr(L \geqq a \mid \mathbf{t}) \\ &\quad = \Pr[\tfrac{1}{2} \textstyle\sum_{j=1}^{g} T_j^2 \geqq \tfrac{1}{2}(n_1 + n_2 - g_1 - g_2)(\exp(a - A(\mathbf{r}^{(1)})) \\ \text{(5.4)} \quad &\qquad \cdot 2/(n_1 + n_2) - 1) \mid \mathbf{t}] \\ &\quad = \Pr[\tfrac{1}{2} \textstyle\sum_{j=1}^{g} T_j^2 \geqq \tfrac{1}{2}(n_1 + n_2 - g_1 - g_2)\{[\exp(b - A(\mathbf{r}^{(1)}) - \tfrac{1}{2} T_0^2 \\ &\qquad + \tfrac{1}{2}(n_1 + n_2) \ln(1 + T_0^2/(n_1 + n_2 - g_1 - g_2))2/(n_1 + n_2))] - 1\} \mid \mathbf{t}] \end{aligned}$$

where $\sum_{j=1}^{g} T_j^2$ has the observed value T_0^2 and the observed value of $\hat{I}$ is b.

Both $\hat{I}$ and L have the same limiting null and non-null distributions. Bahadur efficiency may be a possible way to choose between these tests in large samples. In addition, it may be of interest to compute the Pitman efficiency of the Roy and Bhapkar test with respect to the $\hat{I}$ or L test. The limiting power function of $\hat{I}$ or L is easy to compute for Pitman alternatives. Since the likelihood ratio test requires a longer computation than the $\hat{I}$ statistic, one would ordinarily prefer to use the $\hat{I}$ test unless small sample power studies indicated otherwise.

6. Acknowledgments. The second author acknowledges a great debt of gratitude to his thesis advisor, Professor Frederick Mosteller. He thanks Professor A. P. Dempster for his extremely helpful discussions on a number of points in this work. He expresses his gratitude to Professor W. G. Cochran for suggesting this problem. He thanks Professors H. Scheffé and O. J. Dunn for their detailed comments on this paper and to the referee and associate editor for their critical comments.

REFERENCES

ANDERSON, T.W. (1958). *Introduction to Multivariate Statistical Analysis.* Wiley, New York.

DAS GUPTA, S. (1960). Point biserial correlation and its generalization. *Psychometrika* **25** 393–408.

GREENBLATT, M., GROSSER, G. H. and WECHSLER, H. (1962). A comparative study of selected anti-depressant medications and electric shock therapy. *Amer. J. Psychiatry* **119** 144–153.

HANNAN, J. F. and TATE, R. F. (1965). Estimation of the parameters for a multivariate normal distribution when one variable is dichotomized. *Biometrika* **52** 664–668.

HOTELLING, H. (1951). A generalized T test and measures of multivariate dispersion. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 23–42. University of California Press.

ITO, K. (1956). Asymptotic formulae for the distribution of Hotelling's generalized T_0^2 statistic. II. *Ann. Math. Statist.* **31** 1148–1153.

KULLBACK, S. (1959). *Information Theory and Statistics.* Wiley, New York.

MOUSTAFA, M. D. (1957). Tests of hypotheses on a multivariate population, some of the variables being continuous and the rest categorical. Institute of Statistics Mimeograph Series No. 179, Chapel Hill, North Carolina.

OGAWA, J., MOUSTAFA, M. D., and ROY, S. N. (1957). On the asymptotic distribution of the likelihood ratio in some problems on mixed variate populations. Institute of Statistics Mimeograph Series No. 180, Chapel Hill, North Carolina.

OLKIN, I. and TATE, R. F. (1961). Multivariate correlation models with mixed discrete and continuous variables. *Ann. Math. Statist.* **32** 448–465.

PILLAI, K. C. S. and SAMSON, P., JR. (1959). On Hotelling's generalization of T^2. *Biometrika.* **46** 160–168.

ROY, S. N. and BHAPKAR, V. P. (1959). Some nonparametric analogues of normal ANOVA and MANOVA. *Ann. Math. Statist.* **30** 617.

SHUBIN, H. and WEIL, M. (1967). *Diagnosis and Treatment of Shock.* Williams and Wilkins, Baltimore, Md.