

TESTING OF HYPOTHESES ON A MULTIVARIATE
POPULATION, SOME OF THE VARIATES BEING CONTINUOUS
AND THE REST CATEGORICAL

by

M. D. Moustafa

This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 18(600) - 83. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Institute of Statistics
Mimeograph Series No. 179
July, 1957

3. AFOSR-TN 57-440
4. ASTIA-AD 136 430

UNCLASSIFIED

SECURITY INFORMATION

BIBLIOGRAPHICAL CONTROL SHEET

1. University of North Carolina, Chapel Hill, N. C.
2. Mathematics Division, Air Force Office of Scientific Research
5. TESTING OF HYPOTHESES ON A MULTIVARIATE POPULATION, SOME OF
THE VARIATES BEING CONTINUOUS AND THE REST CATEGORICAL
6. M. D. Moustafa
7. July, 1957
8. AF 18(600)-83
9. File 3.3

TESTING OF HYPOTHESES ON A MULTIVARIATE
POPULATION; SOME OF THE VARIATES BEING CONTINUOUS
AND THE REST CATEGORICAL *

by

M. D. Moustafa

Summary: We consider a $(k+1)$ -variate distribution in which k variates are continuous and 1 variates are categorical. The k variates are assumed to have a conditional multivariate normal distribution with respect to the 1 categorical variates which are assumed to have a multinomial distribution. Appropriate hypotheses are framed in this situation, analogous to the customary hypotheses on a single multivariate normal distribution, large sample tests of such hypotheses are developed and some of their properties studied. Next, instead of assuming a single multinomial distribution on the 1 categorical variates, a product of multinomial distributions is assumed (in case some of the 1 -categorical are ways of classification) and hypotheses are framed in this situation analogous to the customary ones for several multivariate normal distributions, and large sample tests of such hypotheses and some of their properties are studied.

* This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command.

1. Introduction:

Many contributions have been made to the multivariate analysis of variance and covariance of large masses of data on either continuous or categorical variates.

In this paper, we are applying the technique of the multivariate analysis of variance and covariance to a multi-way table such that certain ways refer to continuous variates and the other ways are categorical. For certain problems all the categorical ways refer to variates; for certain other problems all of them refer to ways of classification; and for some problems some of the ways refer to variates and the rest to ways of classification.

We use the fact that under certain broad conditions, the $-2 \log \lambda$ statistic, for large samples and on the null hypothesis, has asymptotically the χ^2 -distribution with certain degrees of freedom (Wilkes, Wald). Also we use the fact that, for categorical data, the $-2 \log \lambda$ and the χ^2 -statistics are both equivalent (Anderson, Ogawa).

In section 2, we consider the two way or (X,Z) table in which we take X continuous and Z categorical. Z may be a random variate or a way of classification.

In section 3, we consider the case of a three way (X,Y,Z) table, in which X is a continuous variate, Y and Z are both categorical variates or one of them a categorical variate and the other a way of classification, or both are ways of classification.

In section 4, we consider the case of a three way (X,Y,Z) table in which X and Y are continuous variates and Z a categorical variate or a way of classification.

Section 5 points out the possibility of an extension to a $(k+1)$ -way table where the k 's are continuous and the 1 's categorical.

In all these cases we formulate all possible hypotheses to be tested concerning conditional independence, joint independence, and total independence; and we give the $-2 \log \lambda$ statistic used, together with its distribution. The likelihood ratio λ is defined as:

$$\lambda = \frac{\max P(\text{sample} \mid H_0)}{\max P(\text{sample} \mid H)}$$

$$= \frac{P(\underline{x}_1, \dots, \underline{x}_n \mid p_1^0, p_2^0, \dots, p_v^0, \hat{p}_{v+1}^0, \dots, \hat{p}_m^0)}{P(\underline{x}_1, \dots, \underline{x}_n \mid \hat{p}_1, \hat{p}_2, \dots, \hat{p}_v, \hat{p}_{v+1}, \dots, \hat{p}_m)}$$

where, if the null hypothesis is such that $p_i = p_i^0$ for $i = 1, 2, \dots, v$, and $H \neq H^0$, then the numerator of λ is the likelihood function after substituting for p_i , $i = v+1, \dots, m$, the maximum likelihood estimates \hat{p}_i^0 subject to the null hypothesis H_0 ; and the denominator of λ is the likelihood function after substituting the maximum likelihood estimates \hat{p}_i , $i = 1, 2, \dots, m$, of p_i subject to the alternative H .

In all these cases, we assume that the conditional distribution of the continuous variates, given the categorical variates, is a multivariate normal whose parameters might depend upon the values of the categorical variates (at which the conditional probability is being considered). In case some of the directions along the categorical part of the table refer to "ways of classification," the parameters of the conditional normal distribution might also depend upon the particular cell of the part of the table constituted

by the ways of classification. In every case it can be shown (and has, in fact, been shown in another paper) that the conditional distribution of the continuous variates, given the categorical ones, satisfy Doob's conditions, from which, by using certain other theorems, it has been proved in the other paper that $-2 \log \lambda$ for the joint distribution of the continuous and categorical variates has asymptotically a χ^2 -distribution with proper degrees of freedom. Actually, however, in each case, it is not $-2 \log \lambda$ but another algebraically simple and more convenient statistic, proved to be equivalent in probability to $-2 \log \lambda$, that is used. A method of obtaining such a statistic is outlined in section 6, the mathematical proof being reserved for another paper.

2. The case of the two way (X,Z) table, X continuous and Z categorical:

Z may be a random variate or a way of classification, which may belong to r exhaustive and exclusive categories $i = 1, 2, \dots, r$, say. We assume that the distribution function of X is normal for the different categories of Z but that the parameters, viz., the mean and the variance of the distribution function of X depend on the category.

Suppose we have a sample of n individuals; every one belongs to one or another of the r categories, and also has a measurement x .

Let n_i , $i = 1, 2, \dots, r$, be the number of individuals belonging to the i^{th} category such that $\sum_{i=1}^r n_i = n$ is fixed from sample to sample.

If Z is a way of classification, then n_i , $i = 1, 2, \dots, r$, is fixed from sample to sample; but if Z is a random variate, then suppose that p_i is the probability that an individual belongs to the i^{th} category.

2.1 The case when Z is a variate: According to our assumption, the model of the conditional likelihood function is

$$(2.1.1) \quad P \{ x_1, \dots, x_n | \underline{n} \} = \prod_{i=1}^r \left(\frac{1}{2\pi v(i)} \right)^{n_i} \times \exp - \sum_{i=1}^r \sum_{j=1}^{n_i} \frac{(x_{ij} - \mu(i))^2}{2v(i)},$$

where x_{ij} is the j^{th} measurement on X of an observation belonging to the i^{th} category, and $\underline{n}' = (n_1, n_2, \dots, n_r)$. Therefore, the likelihood function is:

$$(2.1.2) \quad P \{ x_1, \dots, x_n; \underline{n} \} = P \{ x_1, \dots, x_n | \underline{n} \} \frac{n!}{\prod_{i=1}^r n_i!} \prod_{i=1}^r p_i^{n_i}.$$

We are interested in testing the composite hypothesis that the distribution of X is the same in the r-different categories, i.e.,

$$H_0: \mu(r) = \mu(r-1) = \dots = \mu(1) = \mu, \\ \text{and} \quad v(r) = v(r-1) = \dots = v(1) = v,$$

against the alternative $H \neq H_0$, where μ and v are nuisance parameters.

For n very large, and $\frac{n_i}{n} = a$ constant, the $-2 \log \lambda$ statistic is

$$(2.1.3) \quad -2 \log \lambda = \sum_{i=1}^r n_i \log \frac{\hat{v}}{\hat{v}(i)}.$$

We shall prove in another paper, that this statistic is asymptotically equivalent, in probability, to

$$(2.1.4) \quad \sum_{i=1}^r \frac{n_i (\hat{\mu}(i) - \hat{\mu})^2}{\hat{v}} + \sum_{i=1}^r \frac{n_i (\hat{v}(i) - \hat{v})^2}{2\hat{v}^2},$$

where
$$\hat{\mu}(i) = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \quad , \quad \hat{\mu} = \frac{1}{n} \sum_{i=1}^r n_i \hat{\mu}(i) \quad ,$$

$$\hat{v}(i) = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \hat{\mu}(i))^2 \quad ,$$

$$\text{and } \hat{v} = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} (x_{ij} - \hat{\mu})^2 \quad ;$$

and that it has the χ^2 -distribution with $2(r-1)$ degrees of freedom.

2.2 The case when Z is a way of classification:

Here, we have r independent normal populations; and the hypothesis to be tested is that $X(i)$, $i = 1, 2, \dots, r$, have identical distributions.

If n_i is very large, for $i = 1, 2, \dots, r$, then the statistics (2.1.3) and (2.1.4) hold and have the same χ^2 -distribution with $2(r-1)$ degrees of freedom, though the asymptotic power of this test is different from that of the previous section.

3. The case of a three way (X,Y,Z) table, X continuous, and Y, Z categorical:

Suppose that Y can belong to the categories $i = 1, 2, \dots, r$, and Z can belong to the categories $k = 1, 2, \dots, s$.

We assume that, for given cell (i,k) , the probability density function of X is normal with mean $\mu(i,k)$ and variance $v(i,k)$.

Suppose we have a random sample of size n , where n is fixed from sample to sample, such that n_{ik} individuals belong to the (i,k) th cell and $\sum_{i=1}^r \sum_{k=1}^s n_{ik} = n$. Also every individual observation has the measurement x_{ikj} , $j = 1, 2, \dots, n_{ik}$ for $i = 1, 2, \dots, r$ and $k =$

$1, 2, \dots, s$. We have $\sum_{k=1}^s n_{ik} = n_{i0}$, and $\sum_{i=1}^r n_{ik} = n_{0k}$. We

could have n_{i0} , $i = 1, 2, \dots, r$, or n_{ok} , $k = 1, 2, \dots, s$, or n_{ik} fixed in advance in which case either Y or Z or both would be ways of classification.

3.1 Both Y and Z are random variates:

Let $\underline{n}' = (n_{11} \ n_{12} \ \dots \ n_{1s}, \ n_{21} \ n_{22} \ \dots \ n_{2s}, \ \dots, \ n_{r1} \ n_{r2} \ \dots \ n_{rs})$, and p_{ik} is the probability that an individual observation belongs to the $(i,k)^{\text{th}}$ cell, where

$$\sum_{k=1}^s p_{ik} = p_{i0} \quad , \quad \sum_{i=1}^r p_{ik} = p_{ok} \quad , \quad \text{and} \quad \sum_{i=1}^r \sum_{k=1}^s p_{ik} = 1 \quad .$$

The conditional likelihood function will be

$$(3.1.1) \quad P \{ x_1, \dots, x_n | \underline{n} \} = \prod_{i,k} \left(\frac{1}{2\pi v(i,k)} \right)^{\frac{n_{ik}}{2}} \\ \times \exp - \sum_{i=1}^r \sum_{k=1}^s \sum_{j=1}^{n_{ik}} \frac{(x_{ikj} - \mu(i,k))^2}{2v(i,k)} \quad ,$$

and therefore the likelihood function in this case is

$$(3.1.2) \quad P \{ x_1, \dots, x_n, \underline{n} \} = P \{ x_1, \dots, x_n | \underline{n} \} \frac{n!}{\prod_{i,k} n_{ik}!} \prod_{i,k} p_{ik}^{n_{ik}} \quad .$$

We consider the following kinds of composite hypotheses:

(3.1,a) Conditional independence between X and Y, given Z:

Given that Z belongs to the k^{th} category, then ..

$$H_0: \mu(i,k) = \mu(o,k) \quad , \quad \text{and} \quad v(i,k) = v(o,k)$$

$$\text{for } i = 1, 2, \dots, r \quad , \quad \text{and } k = 1, 2, \dots, s$$

against $H \neq H_0$,

where $\mu(o,k)$ and $v(o,k)$ are arbitrary nuisance parameters.

The $-2 \log \lambda$ statistic is

$$(3.1.3) \quad \sum_{k=1}^s \sum_{i=1}^r n_{ik} \log \frac{\hat{v}(o,k)}{\hat{v}(i,k)} ;$$

and this, except for a quantity that converges to zero in probability, is equal to

$$(3.1.4) \quad \sum_{k=1}^s \sum_{r=1}^r \left[\frac{n_{ik} (\hat{\mu}(i,k) - \hat{\mu}(o,k))^2}{\hat{v}(o,k)} + \frac{n_{ik} (\hat{v}(i,k) - \hat{v}(o,k))^2}{2\hat{v}^2(o,k)} \right] .$$

Further, it has the χ^2 -distribution with $2s(r-1)$ degrees of freedom.

We note that

$$\hat{\mu}(i,k) = \frac{1}{n_{ik}} \sum_{j=1}^{n_{ik}} x_{ikj} , \quad \hat{\mu}(o,k) = \frac{1}{n_{ok}} \sum_{i=1}^r n_{ik} \hat{\mu}(i,k) ,$$

$$\hat{v}(i,k) = \frac{1}{n_{ik}} \sum_{j=1}^{n_{ik}} (x_{ikj} - \hat{\mu}(i,k))^2 ,$$

$$\hat{v}(o,k) = \frac{1}{n_{ok}} \sum_{i=1}^r \sum_{j=1}^{n_{ik}} (x_{ikj} - \hat{\mu}(o,k))^2 .$$

Likewise, we can test the conditional independence between X and Z , given Y , in which case the $-2 \log \lambda$ statistic has a χ^2 -distribution with $2r(s-1)$ degrees of freedom.

(3.1,b) Independence between (X,Y) jointly and Z:

The likelihood function in this case can be written in the form

$$(3.1.5) \quad P \{ x_1, \dots, x_n, \underline{n} \} = P \{ x_1, \dots, x_n, \underline{n} \mid \underline{n}_2 \} P \{ \underline{n}_2 \} ,$$

where $\underline{n}'_2 = (n_{o1} \ n_{o2} \ \dots \ n_{os})$,

$$(3.1.6) \quad P \{ x_1, \dots, x_n, n \mid n_2 \} = \prod_{i,k} \left(\frac{1}{2\pi v(i,k)} \right)^{\frac{n_{ik}}{2}}$$

$$\times \exp - \sum_{k=1}^s \sum_{r=1}^r \sum_{j=1}^{n_{ik}} \frac{(x_{ijk} - \mu(i,k))^2}{2v(i,k)} \prod_{k=1}^s \left\{ \frac{n_{ok}!}{r} \prod_{i=1}^r p_{ik}^{*n_{ik}} \right\} \prod_{i=1}^r n_{ik}!$$

$$p_{ik}^* = \frac{p_{ik}}{p_{ok}},$$

$$\text{and} \quad P \{ n_2 \} = \frac{n!}{\prod_{k=1}^s n_{ok}!} \prod_{k=1}^s p_{ok}^{n_{ok}}.$$

From (3.1.8), we can write the null hypothesis as:

$$H_0: \quad \mu(i,k) = \mu(i,o) \quad , \quad v(i,k) = v(i,o) \quad ,$$

$$\text{and } p_{ik} = p_{io} p_{ok} \quad , \quad \text{for } i = 1, \dots, r, \quad k = 1, \dots, 2 \quad ,$$

$$\text{against } H \neq H_0 \quad .$$

The $-2 \log \lambda$ statistic in this case is

$$(3.1.7) \quad \sum_{i=1}^r \sum_{k=1}^s n_{ik} \log \frac{\hat{v}(i,o)}{\hat{v}(i,k)} + 2 \sum_{i=1}^r \sum_{k=1}^s n_{ik} \left[\log \frac{n_{ik}}{n} - \log \frac{n_{io} n_{ok}}{n^2} \right];$$

and it is asymptotically equivalent, in probability, to:

$$(3.1.8) \quad \sum_{i=1}^r \sum_{k=1}^s \left[\frac{n_{ik} (\hat{\mu}(i,k) - \hat{\mu}(i,o))^2}{\hat{v}(i,o)} + \frac{n_{ik} (\hat{v}(i,k) - \hat{v}(i,o))^2}{2\hat{v}^2(i,o)} \right. \\ \left. + \frac{\left(n_{ik} - \frac{n_{io} n_{ok}}{n} \right)^2}{\frac{n_{io} n_{ok}}{n}} \right]$$

Further, this statistic has the χ^2 -distribution with $2r(s-1) + (r-1)(s-1) = (s-1)(3r-1)$ degrees of freedom.

Likewise, we can test the independence between (X,Z) jointly and Y . Here, the $-2 \log \lambda$ statistic has asymptotically the χ^2 -distribution with $(r-1)(3s-1)$ degrees of freedom; and it will be asymptotically equivalent, in probability, to:

$$(3.1.9) \quad \sum_{k=1}^s \sum_{i=1}^r \left[\frac{n_{ik}(\hat{\mu}(i,k) - \hat{\mu}(o,k))^2}{\hat{v}(o,k)} + \frac{n_{ik}(\hat{v}(i,k) - \hat{v}(o,k))^2}{2\hat{v}^2(o,k)} \right. \\ \left. + \frac{\left(n_{ik} - \frac{n_{io}n_{ok}}{n} \right)^2}{\frac{n_{io}n_{ok}}{n}} \right] .$$

(3.1.c) Independence between X and (Y,Z) jointly:

The hypothesis to be tested is

$$H_0: \mu(i,k) = \mu, \quad v(i,k) = v, \quad \text{for all } i, k,$$

where μ, v are nuisance parameters, against $H \neq H_0$.

The $-2 \log \lambda$ statistic is

$$(3.1.10) \quad \sum_{i=1}^r \sum_{k=1}^s n_{ik} \log \frac{\hat{v}}{\hat{v}(i,k)},$$

where

$$\hat{v} = \frac{1}{n} \sum_{i=1}^r \sum_{k=1}^s \sum_{j=1}^n n_{ik} (x_{ikj} - \hat{\mu})^2,$$

and

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^r \sum_{k=1}^s \sum_{j=1}^n n_{ik} x_{ikj} = \frac{1}{n} \sum_{i=1}^r \sum_{k=1}^s n_{ik} \hat{\mu}(i,k).$$

This statistic is asymptotically equivalent, in probability, to

$$(3.1.11) \quad \sum_{i=1}^r \sum_{k=1}^s \left[\frac{n_{ik}(\hat{\mu}(i,k) - \hat{\mu})^2}{\hat{v}} + \frac{n_{ik}(\hat{v}(i,k) - \hat{v})^2}{2\hat{v}^2} \right];$$

and it has the χ^2 -distribution with $2(rs-1)$ degrees of freedom.

(3.1,d) Total independence between the three variates:

In view of (3.1.2), the hypothesis to be tested is

$$H_0: \mu(i,k) = \mu, \quad v(i,k) = v,$$

$$\text{and } p_{ik} = p_{i0} p_{0k} \quad \text{for all } i, k,$$

against $H \neq H_0$, where μ , v , p_{i0} and p_{0k} are all

nuisance parameters.

The $-2 \log \lambda$ statistic is

$$(3.1.12) \quad \sum_{i=1}^r \sum_{k=1}^s \left[n_{ik} \log \frac{\hat{v}}{\hat{v}(i,k)} + 2n_{ik} \left\{ \log \frac{n_{ik}}{n} - \log \frac{n_{i0} n_{0k}}{n^2} \right\} \right];$$

and this is asymptotically equivalent, in probability, to

$$(3.1.13) \quad \sum_{i=1}^r \sum_{k=1}^s \left[\frac{n_{ik}(\hat{\mu}(i,k) - \hat{\mu})^2}{\hat{v}} + \frac{n_{ik}(\hat{v}(i,k) - \hat{v})^2}{2\hat{v}^2} + \frac{(n_{ik} - \frac{n_{i0} n_{0k}}{n})^2}{\frac{n_{i0} n_{0k}}{n}} \right].$$

It has the χ^2 -distribution with $2(rs-1) + (r-1)(s-1) = 3rs - r - s - 1$ degrees of freedom.

3.2 Y is a random variate, and Z a way of classification:

In this case, the hypothesis of (3.1,a) would be replaced by the hypothesis that X and Y are independent for each Z separately, that of (3.1,b) would be replaced by the hypothesis that the joint distribution of (X,Y) is the same for all Z, and that of (3.1,c) does not seem to have an immediate analogue. The hypothesis of (3.1,d) would be replaced by the hypothesis that X and Y are independent, each having a distribution which is the same for all Z.

For each of these problems the statistic and the distribution (on the null hypothesis) are the same as for the corresponding problem of the previous cases, although the asymptotic power of the test for any problem in this case would be quite different from that for the corresponding problem of the previous case.

3.3 Both Y and Z are ways of classification:

Here n_{ik} , for each i, k , could be fixed or the marginals $\underline{n}'_1 = (n_{10} \ n_{20} \ \dots \ n_{r0})$ and $\underline{n}''_2 = (n_{o1} \ n_{o2} \ \dots \ n_{os})$ could be fixed.

(i) If n_{ik} , $i = 1, \dots, r$, $k = 1, \dots, s$, are fixed from sample to sample, we should have rs independent variates $X(i, k)$; on each we have a sample of n_{ik} observations.

The hypothesis (3.1, a) should be replaced by the hypothesis that the distribution of $X(i, k)$ will be independent of Z for each Z .

The hypothesis (3.1, b) and (3.1, d) have no analogues. But the hypothesis (3.1, c) should be replaced by the hypothesis that $X(i, k)$ will have the same distribution for each Y and Z . For each of these two problems, the statistic and the distribution (on the null hypothesis) are the same as for the corresponding problems in section 3.1, except that the asymptotic power of the test for any problem here would be quite different from that for the corresponding problem in section 3.1.

(ii) If the marginals $\underline{n}'_1 = (n_{10} \ n_{20} \ \dots \ n_{r0})$ and $\underline{n}''_2 = (n_{o1} \ n_{o2} \ \dots \ n_{os})$ are fixed, the likelihood function of the sample will be

$$(3.3.1) \quad P \{ x_1, x_2, \dots, x_n, \underline{n} \mid \underline{n}_1 \text{ and } \underline{n}_2 \text{ fixed} \} \\ = P \{ x_1, x_2, \dots, x_n \mid \underline{n}, \text{ such that } \underline{n}_1 \text{ and } \underline{n}_2 \text{ fixed} \} \\ \times P \{ \underline{n} \mid \underline{n}_1 \text{ and } \underline{n}_2 \text{ fixed} \} \quad ,$$

where

$$(3.3.2) \quad P \left\{ \underline{n} \mid \underline{n}_1 \text{ and } \underline{n}_2 \text{ fixed} \right\} = \frac{\prod_{i=1}^r n_{i0}! \prod_{k=1}^s n_{ok}!}{n! \prod_{i,k} n_{ik}!},$$

in which case $P \left\{ \underline{n} \mid \underline{n}_1 \text{ and } \underline{n}_2 \text{ fixed} \right\}$ could have been obtained from

$$P \left\{ \underline{n} \right\} = \frac{n!}{\prod_{i,k} n_{ik}!} \prod_{i,k} p_{ik}^{n_{ik}}$$

by putting $p_{ik} = p_{i0}p_{ok}$ and, under this condition of independence, finding $P \left\{ \underline{n} \right\}$ subject to \underline{n}_1 and \underline{n}_2 being fixed. This makes it difficult to write down $P \left\{ \underline{n} \mid \underline{n}_1 \text{ and } \underline{n}_2 \text{ fixed} \right\}$ under a distinct assumption other than $p_{ik} = p_{i0}p_{ok}$ (Roy and Mitra). If the n_{ik} 's are very large, the hypotheses (3.1,a) and (3.1,c) would be the same in this case and each will have the same statistic with the same χ^2 -distribution under the null hypothesis. For the hypotheses corresponding to (3.1,b) and (3.1,d), the statistics could not be obtained directly, due to the form of $P \left\{ \underline{n} \mid \underline{n}_1 \text{ and } \underline{n}_2 \text{ fixed} \right\}$; but may be taken over from (3.1,b) and (3.1,d) by analogy. The power of each of these two tests cannot be obtained.

4. The case of a three way (X,Y,Z) table, X and Y continuous, and Z categorical:

Suppose that Z belongs to the i^{th} category, where $i = 1, 2, \dots, r$; and for given i , we assume X and Y to have bivariate normal distribution with parameters given by the vector $\underline{\mu}'(i) = (\mu_1(i) \quad \mu_2(i))$ for means, and

$$V(i) = \begin{pmatrix} v_{11}(i) & v_{12}(i) \\ v_{12}(i) & v_{22}(i) \end{pmatrix}$$

for variance-covariance matrix.

Suppose we have a sample of n observations, where n is fixed from sample to sample, such that n_i individual observations belong to the i^{th} category, $i = 1, 2, \dots, r$, and $\sum_{i=1}^r n_i = n$. Also every individual observation has two measurements (x_{ij}, y_{ij}) , $j = 1, 2, \dots, n_i$.

We could have n_i as random numbers subject to $\sum_{i=1}^r n_i = n$ being fixed in which case Z would be a random variate, or n_i fixed for $i = 1, 2, \dots, r$, in which case Z would be a way of classification.

4.1 Z a categorical variate:

Let p_i be the probability that an observation belongs to the i^{th} category, $\sum_{i=1}^r p_i = 1$; and put $\underline{x}' = (x \ y)$.

The likelihood function, in this case, is

$$\begin{aligned}
 (4.1.1) \quad P\{\underline{x}_1, \dots, \underline{x}_n, \underline{n}\} &= P\{\underline{x}_1, \dots, \underline{x}_n \mid \underline{n}\} P\{\underline{n}\} \\
 &= \prod_{i=1}^r \left(\frac{1}{2\pi |V(i)|^{1/2}} \right)^{n_i} \exp - \sum_{i=1}^r \sum_{j=1}^{n_i} \frac{1}{2} (\underline{x}'_{ij} \\
 &\quad - \underline{\mu}'(i)) V(i)^{-1} (\underline{x}'_{ij} - \underline{\mu}(i)) \\
 &\quad \times \frac{n!}{\prod_{i=1}^r n_i!} \prod_{i=1}^r p_i^{n_i}.
 \end{aligned}$$

We consider the following problems in testing of hypotheses.

(4.1, a) Conditional independence between X and Y, given Z:

If Z belongs to the i^{th} category, we set $H_0: v_{12}(i) = 0$, for $i = 1, 2, \dots, r$, against $H \neq H_0$. The $-2 \log \lambda$ statistic is

$$(4.1.2) \quad \sum_{i=1}^r n_i \log \frac{1}{(1 - \hat{r}^2(i))},$$

where $\hat{r}(i)$ is the estimate of the ordinary correlation coefficient calculated from the i^{th} category; and, except for a quantity that converges to zero in probability, is asymptotically equivalent to

$$(4.1.3) \quad \sum_{i=1}^r n_i \hat{r}^2(i)$$

Further, it has the χ^2 -distribution with r degrees of freedom.

(4.1,b) Independence between (X,Y) jointly and Z:

The hypothesis to be tested is

$$H_0: \underline{\mu}'(i) = \underline{\mu}' = (\mu_1 \mu_2) , \\ V(i) = V = \begin{pmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{pmatrix} \quad \text{for } i = 1, 2, \dots, r,$$

against $H \neq H_0$, where $\underline{\mu}$ and v consist of five arbitrary parameters.

The $-2 \log \lambda$ statistic is

$$(4.1.4) \quad \sum_{i=1}^r n_i \log \frac{|V|}{|V(i)|} = \sum_{i=1}^r n_i \log \frac{\hat{v}_{11} \hat{v}_{22} (1 - \hat{r}^2)}{\hat{v}_{11}(i) \hat{v}_{22}(i) (1 - \hat{r}^2(i))} ;$$

and this is asymptotically equivalent, in probability, to

$$(4.1.5) \quad \sum_{i=1}^r n_i (\hat{\underline{\mu}}(i) - \hat{\underline{\mu}})' U_1 (\hat{\underline{\mu}}(i) - \hat{\underline{\mu}}) + \sum_{i=1}^r n_i (\hat{\underline{v}}(i) - \hat{\underline{v}})' \\ \times U_2 (\hat{\underline{v}}(i) - \hat{\underline{v}}) ,$$

where $\underline{v}' = (v_{11} \quad v_{22} \quad v_{12})$,

$$U_1 = \begin{pmatrix} \hat{v}_{11}^{-1} & \hat{v}_{12}^{-1} \\ \hat{v}_{12}^{-1} & \hat{v}_{22}^{-1} \end{pmatrix} ,$$

$$2 = \begin{bmatrix} \frac{1}{2}(\hat{v}_{11}^{-1})^2 & \frac{1}{2}(\hat{v}_{12}^{-1})^2 & \hat{v}_{11}^{-1} \hat{v}_{12}^{-1} \\ \frac{1}{2}(\hat{v}_{12}^{-1})^2 & \frac{1}{2}(\hat{v}_{22}^{-1})^2 & \hat{v}_{22}^{-1} \hat{v}_{12}^{-1} \\ \hat{v}_{11}^{-1} \hat{v}_{12}^{-1} & \hat{v}_{22}^{-1} \hat{v}_{12}^{-1} & \hat{v}_{11}^{-1} \hat{v}_{22}^{-1} + (\hat{v}_{12}^{-1})^2 \end{bmatrix},$$

$$\hat{v}_{11}^{-1} = \frac{\hat{v}_{22}}{|\mathbf{V}|}, \quad \hat{v}_{22}^{-1} = \frac{\hat{v}_{12}}{|\mathbf{V}|}, \quad \hat{v}_{12}^{-1} = \frac{-\hat{v}_{12}}{|\mathbf{V}|},$$

and \hat{r} is the correlation coefficient estimated from all the categories pooled. This statistic has asymptotically the χ^2 -distribution with $5(r-1)$ degrees of freedom.

(4.1,c) Independence between (X,Z) jointly and Y:

A necessary and sufficient set of conditions for this is that

$$H_0: \mu_2(i) = \mu_2, \quad v_{22}(i) = v_{22}, \quad v_{12}(i) = 0, \\ \text{for } i = 1, 2, \dots, r,$$

against $H \neq H_0$.

That this set of conditions is sufficient is obvious; but the necessity we prove in the appendix.

The $-2 \log \lambda$ statistic is

$$(4.1.6) \quad \sum_{i=1}^r n_i \log \frac{\hat{v}_{22}}{\left(\hat{v}_{22}(i) - \hat{v}_{12}(i) \frac{\hat{v}_{12}(i)}{\hat{v}_{11}(i)} \right)} = \sum_{i=1}^r n_i \log \frac{\hat{v}_{22}}{\hat{v}_{22}(i)(1-\hat{r}^2(i))}.$$

This statistic, except for a quantity that converges to zero in probability, is equivalent to

$$(4.1.7) \quad \sum_{i=1}^r \left[\frac{n_i(\hat{\mu}_2(i) - \hat{\mu}_2)^2}{\hat{v}_{22}} + \frac{n_i(\hat{v}_{22}(i) - \hat{v}_{22})^2}{2\hat{v}_{22}^2} + n_i \hat{r}^2(i) \right],$$

and has the χ^2 -distribution with $(3r-2)$ degrees of freedom.

4.1,d Total independence:

The hypothesis we are interested in is

$$H_0: \mu(i) = \mu, \quad V(i) = V, \quad \text{and } v_{12} = 0,$$

for $i = 1, 2, \dots, r$, against $H \neq H_0$.

Here we shall have four arbitrary nuisance parameters, viz.,

$$\mu_1, \mu_2, v_{11}, v_{22}.$$

The $-2 \log \lambda$ statistic is given by

$$(4.1.8) \quad \sum_{i=1}^r n_i \log \frac{\hat{v}_{11} \hat{v}_{22}}{\hat{v}_{11}(i) \hat{v}_{22}(i) (1 - \hat{r}^2(i))};$$

and this is asymptotically equivalent, in probability, to

$$(4.1.9) \quad \sum_{k=1}^2 \sum_{i=1}^r \left\{ \frac{n_i (\hat{\mu}_k(i) - \hat{\mu}_k)^2}{\hat{v}_{kk}} + \frac{n_i (\hat{v}_{kk}(i) - \hat{v}_{kk})^2}{2\hat{v}_{kk}} \right\} + \sum_{i=1}^r \frac{n_i (\hat{v}_{12}(i))^2}{\hat{v}_{11} \hat{v}_{22}}.$$

Further, it has the χ^2 -distribution with $4(r-1) + r = 5r - 4$ degrees of freedom.

There is a case which we could not test, viz.,

4.1,e Conditional independence between X and Z, given Y:

Owing to some difficulties about establishing for this situation the asymptotic χ^2 -distribution that has been established for the other situations by using a line of proof discussed in another paper, we could not test this case.

4.2 Z, a way of classification:

If Z may belong to r categories, then we shall be dealing with r independent bivariate normal populations. The hypothesis in (4.1,a) will be that X and Y are independent in all the r different bivariate populations, that in (4.1,b) will be that (X,Y) will have the same distribution in all different r populations, that in (4.1,c) will have no analogue here, and finally that of (4.1,d) will be that X and Y are independent and have the same distribution in all the r different populations.

For each of these problems, the statistic and the distribution (on the null hypotheses) are the same for the corresponding problems in section 4.1, although the asymptotic power of the test for any problem here would differ from the corresponding problem in section 4.1.

5. From the line of proof of these results, which will be given in another place, similar test problems to these could be considered generally for a $(k+l)$ -way table in which k ways refer to continuous variates and the l ways are categorical. The l categorical ways may refer to all random variates or all ways of classification or some of them refer to random variates and the remaining to ways of classification; and the hypotheses have to be phrased accordingly.

6. In every case we consider, we give, besides the $-2 \log \lambda$ statistic which we calculate directly from (1.1) according to the null hypothesis H_0 and the alternative H , a statistic that differs from the $-2 \log \lambda$ by a quantity that converges to zero in probability. Such a statistic will be called one which is equivalent in probability to

$-2 \log \lambda$. This statistic has the same pattern for all the cases, and could be written down easily once the null hypothesis for any particular case is laid down.

We give here an outline, without mathematical proofs, of the method we used to obtain such a statistic (which is asymptotically equivalent, in probability, to the $-2 \log \lambda$ statistic). The mathematical proofs are left to be given in another paper.

Having satisfied ourselves that the probability density function of the variates considered, in every case, comply with Doob's requirements, we can use the fact that, for large samples, the $-2 \log \lambda$ has the χ^2 -distribution with certain degrees of freedom. If the null hypothesis H_0 is not a simple hypothesis, as happens in all the cases we consider, we define a simple hypothesis H_0^* and consider a statistic equivalent in probability to $-2 \log \lambda_1$, where

$$\lambda_1 = \frac{\max P(\text{sample} \mid H_0^*)}{\max P(\text{sample} \mid H_0)} .$$

Also, together with the alternative hypothesis H , we define a simple hypothesis $H^* \equiv H_0^*$ and consider a statistic equivalent, in probability, to $-2 \log \lambda_2$, where

$$\lambda_2 = \frac{\max P(\text{sample} \mid H^*)}{\max P(\text{sample} \mid H)} .$$

From $-2 \log \lambda_1$ based on H_0^* against H_0 and $-2 \log \lambda_2$ based on H^* against H , we get $-2 \log \lambda$ based on H_0 against H by the following form

$$(6.1) \quad -2 \log \lambda = -2 \log \lambda_2 + 2 \log \lambda_1 .$$

Now, if on the right side of (6.1), we replace, as indicated above, $-2 \log \lambda_2$ and $-2 \log \lambda_1$ by statistics equivalent to them in probability, the right side of (6.1) becomes now a statistic equivalent in probability to $-2 \log \lambda$, and involving H_0^* . It is possible to replace this by another statistic, equivalent in probability, and to show that this latter equivalent statistic does not involve the unknown true values of the parameters occurring in H_0^* . Furthermore, since this is equivalent in probability to $-2 \log \lambda$, therefore this has asymptotically the central χ^2 -distribution with appropriate degrees of freedom, if: (a) H_0^* is true and (b) the maximum likelihood estimates of the parameters are in the neighborhood of H_0^* . Combining these two results we observe that this equivalent statistic involves only H_0 (and H) and has a limiting central χ^2 -distribution (a) for all H_0^* contained in H_0 and (b) for all maximum likelihood estimates of the parameters that lie in the neighborhood of the unknown true H_0^* contained in H_0 . The feature (b), otherwise called the property of consistency, is proved for each case separately in the other paper.

We give as an example, the problem in section 2.1 in which we have:

$$H_0: \mu(i) = \mu \quad \text{and} \quad v(i) = v \quad \text{for } i = 1, 2, \dots, r,$$

where μ and v are arbitrary nuisance parameters, against $H \neq H_0$.

We define H_0^* as:

$$H_0^*: \mu(i) = \mu = \mu^0 \quad \text{and} \quad v(i) = v = v^0, \quad i = 1, 2, \dots, r,$$

where μ^0 and v^0 are the true values of the parameters μ , v , and we consider H_0^* against H_0 . We get:

$$-2 \log \lambda_1 \approx \frac{n(\hat{\mu} - \mu^0)^2}{v^0} + \frac{n(\hat{v} - v^0)^2}{2v^{02}}.$$

Also, considering $H^* \equiv H_0^*$ against H , we get:

$$-2 \log \lambda_2 \approx \sum_{i=1}^r \frac{n_i(\hat{\mu}(i) - \mu^0)^2}{v^0} + \sum_{i=1}^r \frac{n_i(\hat{v}(i) - v^0)^2}{2v^{02}}$$

From λ_1 and λ_2 , we get:

$$(6.2) \quad -2 \log \lambda \equiv -2 \log \lambda_2 + 2 \log \lambda_1 \\ \approx \left\{ \sum_{i=1}^r \frac{n_i(\hat{\mu}(i) - \mu^0)^2}{v^0} - \frac{n(\hat{\mu} - \mu^0)^2}{v^0} \right\} \\ + \left\{ \sum_{i=1}^r \frac{n_i(\hat{v}(i) - v^0)^2}{2v^{02}} - \frac{n(\hat{v} - v^0)^2}{2v^{02}} \right\}.$$

Now, if $\hat{\mu}$ and \hat{v} are in the neighborhood of the true values μ^0 and v^0 (irrespective of these values), — and this follows from the consistency of $\hat{\mu}$ and \hat{v} which is proved in the other paper — then we can show that, except for a quantity that converges to zero in probability, the $-2 \log \lambda$ given in (6.2) is equal to

$$\sum_{i=1}^r \frac{n_i(\hat{\mu}(i) - \hat{\mu})^2}{\hat{v}} + \sum_{i=1}^r \frac{n_i(\hat{v}(i) - \hat{v})^2}{2\hat{v}^2},$$

which is the statistic we suggest for this case; it does not involve the true values μ^0 and v^0 , and has a limiting χ^2 -distribution with $2(r-1)$ degrees of freedom, no matter what μ^0 , v^0 might be.

APPENDIX

To prove that for the independence of (x and n) jointly and y, a set of necessary and sufficient conditions is that

$$\begin{aligned} \mu_2(i) = \mu_2 \quad , \quad v_{22}(i) = v_{22} \quad , \quad v_{12}(i) = 0 \\ \text{for } i = 1, 2, \dots, r \quad . \end{aligned}$$

We proceed as follows

$$\begin{aligned} P \{ \underline{x}, \underline{y}, \underline{n} \} &= \prod_{i=1}^r \left(\frac{1}{2\pi |V(i)|^{1/2}} \right)^{n_i} \exp - \sum_{i=1}^r \sum_{j=1}^{n_i} \frac{1}{2} \left\{ v_{11}^{-1}(i)(x_{ij} - \mu_1(i))^2 \right. \\ &\quad \left. + 2v_{12}^{-1}(i)(x_{ij} - \mu_1(i))(y_{ij} - \mu_2(i)) + v_{22}^{-1}(i)(y_{ij} - \mu_2(i))^2 \right\} . \\ &= \frac{n!}{\prod_{i=1}^r n_i!} \prod_{i=1}^r p_i^{n_i} \\ &= P(\underline{x}, \underline{y} | \underline{n}) P(\underline{n}) . \end{aligned}$$

If (x and n) jointly are to be independent of y, then n and y should be independent, and x and y should also be independent. For necessity therefore, we prove that (i) y and n are independent, and next that (ii) x and y are independent.

$$(i) \quad P(\underline{y} | \underline{n}) = P(\underline{y}, \underline{n}) \div P(\underline{n}) \quad ,$$

and we require this to be independent of n.

We have

$$P(\underline{y}, \underline{n}) = \int \dots \int_{\underline{x}} P(\underline{x}, \underline{y}, \underline{n}) \, d\underline{x}$$

$$= \prod_{i=1}^r \left(\frac{1}{2\pi v_{22}(i)} \right)^{\frac{n_i}{2}} \exp - \sum_{i=1}^r \sum_{j=1}^{n_i} \frac{(y_{ij} - \mu_2(i))^2}{2v_{22}(i)}$$

$$\times \frac{n!}{\prod_{i=1}^r n_i!} \prod_{i=1}^r p_i^{n_i} ;$$

therefore

$$P(\underline{y} | \underline{n}) = \prod_{i=1}^r \left(\frac{1}{2\pi v_{22}(i)} \right)^{\frac{n_i}{2}} \exp - \sum_{i=1}^r \sum_{j=1}^{n_i} \frac{(y_{ij} - \mu_2(i))^2}{2v_{22}(i)}$$

In order that this should be independent of categories i , given \underline{n} (a vector of random numbers subject to $\sum_{i=1}^r n_i = n$ fixed), we require that :

$$P(\underline{y} | \underline{n}) = f(\underline{y}) \quad \text{a function of } \underline{y} \text{ only for any } \underline{n}$$

Suppose that

$$\underline{n}^{(1)'} = (n_1 + 1, n_2, \dots, n_i - 1, \dots, n_r) ,$$

$$\text{and let } \underline{n}^{(0)'} = (n_1, n_2, \dots, n_i, \dots, n_r) ,$$

$$\text{then } \frac{P(\underline{y} | \underline{n}^{(1)'})}{P(\underline{y} | \underline{n}^{(0)'})} = \sqrt{\frac{v_{22}(i)}{v_{11}(1)}} \exp - \left\{ \frac{(y_{1, n_1+1} - \mu_2(1))^2}{2v_{22}(1)} - \frac{(y_{i, n_i} - \mu_2(i))^2}{2v_{22}(i)} \right\}$$

$$= 1$$

This is true for any y_{1, n_1+1}, y_{i, n_i} .

Next we note that if $y_{1, n_1+1} = \mu_2(1)$ and $y_{i, n_i} = \mu_2(i)$,

then $v_{22}(i) = v_{22}(1)$; and if $y_{1, n_1+1} = y_{i, n_i} = 0$, then

$\mu_2(i) = \mu_2(1)$. This shows that $v_{22}(i) = v_{22}$ and $\mu_2(i) = \mu_2$, for $i = 1, 2, \dots, r$, are necessary conditions for the independence

of $(\underline{x}$ and $\underline{n})$ jointly and \underline{y} .

(ii) For \underline{x} and \underline{y} to be independent, we have that

$$\begin{aligned}
 P(\underline{x}, \underline{n} | \underline{y}) &= \frac{P(\underline{x} | \underline{y}, \underline{n}) P(\underline{y} | \underline{n}) P(\underline{n})}{P(\underline{y})} \\
 &= P(\underline{x} | \underline{y}, \underline{n}) P(\underline{n}) \quad , \text{ using (i) } . \\
 &= \prod_{i=1}^r \left(\frac{v_{22}}{2\pi |V(i)|} \right)^{\frac{n_i}{2}} \exp - \sum_{i=1}^r \sum_{j=1}^{n_i} \frac{v_{22}}{2 |V(i)|} \\
 &\quad \times \left\{ x_{ij} - \mu_1(i) - \frac{v_{12}(i)}{v_{22}} (y_{ij} - \mu_2) \right\}^2 \\
 &\quad \times \frac{n!}{\prod_{i=1}^r n_i!} \prod_{i=1}^r p_i^{n_i} \quad ,
 \end{aligned}$$

and
$$P(\underline{x} | \underline{y}) = \sum_{\underline{n}} P(\underline{x}, \underline{n} | \underline{y}) \quad ;$$

this should be independent of \underline{y} , i.e., $= \phi(\underline{x})$, (say) ,

where

$$\begin{aligned}
 \phi(\underline{x}) &= \sum_{\underline{n}} \left(\frac{n!}{\prod_{i=1}^r n_i!} \prod_{i=1}^r p_i^{n_i} \right) \prod_{i=1}^r \left(\frac{1}{2\pi v_{11}(i)} \right)^{\frac{n_i}{2}} \\
 &\quad \times \exp - \sum_{i=1}^r \sum_{j=1}^{n_i} \frac{(x_{ij} - \mu_2(i))^2}{2v_{11}(i)} \quad ,
 \end{aligned}$$

i.e., equals the unconditional distribution of \underline{x} .

If this is to be true for all x_{ij} and y_{ij} (which are random variates), then $v_{12}(i) = 0$ for $i = 1, 2, \dots, r$. This therefore, is another necessary condition for $(\underline{x}$ and $\underline{n})$ to be jointly independent of \underline{n} . That this set of conditions is sufficient for the independence of $(\underline{x}$ and $\underline{n})$

jointly with y is evident.

In conclusion, the author expresses his thanks to Professor S. N. Roy for suggesting the problem and the general line of approach, and to Professor J. Ogawa for his many valuable comments and suggestions.

REFERENCES

1. Anderson, T. W. and Goodman, L. A. (1957), "Statistical Inference about Markov Chains," Ann. Math. Statist. 28, pp. 89 - 110 .
2. Cramér, H., "Mathematical Methods of Statistics," Princeton University Press, 1946.
3. Doob, J. L. (1934), "Probability and Statistics," Trans. Amer. Math. Soc., 34, pp. 759 - 775.
4. Mitra, S. K. (1955), "Contributions to the Statistical Analysis of Categorical data," North Carolina Institute of Statistics, Mimeo. Series 142.
5. Ogawa, J. (1957), "A limit theorem of Cramér and its generalizations," North Carolina Institute of Statistics, Mimeo Series 163.
6. Ogawa, J. (1957), "On the mathematical principles underlying the theory of the χ^2 -test," North Carolina Institute of Statistics, Mimeo. Series 162.
7. Ogawa, J., Moustafa, M. D. and Roy, S. N., (1957), "On the Asymptotic Distribution of the Likelihood Ratio in Some Mixed Variates Populations," North Carolina Institute of Statistics, Mimeo Series (to be published) .
8. Roy, S. N. and Mitra, S. K. (1956), "An introduction to some non-parametric generalizations of analysis of variance and multivariate

- analysis," Biometrika, 43, pp. 361 - 376.
9. Wald, A., (1943), "Tests of Statistical Hypotheses concerning several parameters when the number of observations is large," Trans. Amer. Math. Soc., 54, 426 - 482.
 10. Wilks, S. S., (1938), "The large-sample distribution of the likelihood ratio for testing composite hypotheses," Ann. Math. Statist., IX, 60 - 62.