HOTELLING'S T^2 APPROXIMATION FOR BIVARIATE MIXED (DICHOTOMOUS & CONTINUOUS) DATA

Imad Khamis

Pradeep Singh

James Higgins

Conference on Applied Statistics in Agriculture
Kansas State University

HOTELLING'S T^2 APPROXIMATION FOR BIVARIATE MIXED (DICHOTOMOUS & CONTINUOUS) DATA

Imad Khamis
Department of Mathematics
SE Missouri State University
Cape Girardeau, MO 63701
Email: ikhamis@semo.edu

Pradeep Singh
Department of Mathematics
SE Missouri State University
Cape Girardeau. MO 63701
Email: psingh@semo.edu

James Higgins
Department of Statistics
Kansas State University
Manhattan, KS 66506
Email: higgins@stat.ksu.edu

ABSTRACT

The comparison of the means of two treatments or populations when more than one variable is measured may be done using Hotelling's T^2 statistic. In many real world situations the data obtained are mixed, i.e. one variable is dichotomous and the other variable is continuous. The assumption of multivariate normality upon which Hotelling's T^2 is based is no longer valid. In this paper, an approximate Hotelling T^2 test is proposed for bivariate mixed data and empirically evaluated in terms of Type I error rate. It is shown that the approximation does a good job of controlling the Type I error rate for a range of bivariate mixed parameters even for relatively small sample sizes.

Key Words and Phrases: bivariate mixed data, dichotomous response, Hotelling T^2, multivariate analysis

1. Introduction

It is very common to have multivariate data in which the individual variates are both dichotomous and continuous. In bivariate mixed data one variable takes one of just two possible values, 0 or 1 and the other variable can take continuous values. Multivariate models with mixed data have found extensive application in reliability and biostatistics. In meat sciences this type of data may arise in comparing the contamination of beef carcasses under two methods of decontamination where bivariate responses are presence or absence of one type of bacteria on the carcasses and the number of bacteria on the beef carcasses as a continuous type of response.

If observations are selected randomly from multivariate normal populations, a common multivariate statistic for comparing two populations is Hotelling T^2 [Anderson (1984)]. A permutation test that is based on the computation of the t-statistic for each of the response variables is also appropriate for multivariate data. Blair et. al. [1994] showed that one sided multivariate tests can enjoy substantial power advantages over Hotelling T^2 test under certain conditions.

In a two group experiment with mixed responses, the central problem is to describe the joint distribution of a set of binary and continuous variables. The oldest approach to multivariate binary data is to define indices of association following essentially Yule. Goodman and Kruskal [1954, 1959, and 1963] have reviewed and extended this work. Singh et. al. [2003] proposed an approximate Hotelling T^2 test for bivariate dichotomous data. The approximation does a good job of controlling the Type I error rate.

In this paper, an adaptation of Hotelling T^2 is proposed for comparison of two populations having bivariate mixed responses. An empirical study is done to examine the Type I error rate.

2. Bivariate mixed Data

We consider the problem of comparing two treatments in experiments in which one dichotomous response variable and one continuous response variable are measured on each experimental unit. For example, suppose we want to test two varieties of tomato with respect to their resistance to a pest and the average fruit size. On every sampled plant, two measurements are made. One, whether the pest is absent or present on the plant and secondly, the average fruit size of the plant.

For variety I, the response would be X_1 and X_2

$$X_1 = \begin{cases} 1 \text{ when pest is present on a plant} \\ 0 \text{ otherwise} \end{cases}$$

X_2 = the size of fruit.

For variety II, the response would be Y_1 and Y_2

$$Y_1 = \begin{cases} 1 \text{ when pest is present on a plant} \\ 0 \text{ otherwise} \end{cases}$$

Y_2 = the size of fruit.

In matrix notation we can represent the data as

$$\begin{matrix} & X_1 & X_2 \\ X = & \begin{pmatrix} 0 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 0 & x_4 \\ 0 & x_5 \end{pmatrix} \end{matrix} \text{ and } \begin{matrix} & Y_1 & Y_2 \\ Y = & \begin{pmatrix} 0 & y_1 \\ 1 & y_2 \\ 0 & y_3 \\ 1 & y_4 \\ 0 & y_5 \end{pmatrix} \end{matrix} \qquad (1)$$

The hypotheses to be tested are H_0: $\mu_X = \mu_Y$, H_a: $\mu_X \neq \mu_Y$ where $\mu_X = (p_{x_1} \quad \mu_{x_2})$ and $\mu_Y = (p_{y_1} \quad \mu_{y_2})$ are vectors of expected proportion of pests and expected fruit size of the plants.

3. Permutation Test

A two sample permutation test is carried out by randomly assigning experimental units or subjects to one of two treatments. All possible two-sample data sets are obtained by permuting $m + n$ observations among two groups. There are $\binom{m+n}{m}$ such data sets.

The permutation principle states that the permutation distribution is an appropriate reference distribution for determining the p-value of a test and deciding whether or not a test is statistically significant. One may extend permutation tests to the multivariate setting. Here one permutes observed vectors among the groups, keeping the vectors intact in doing the permutations. See Higgins [2003] Chapter 6 for more details.

A multivariate permutation test for this problem may be carried out using PROC MULTTEST in SAS®. The permutation test is based on the computation of a t-statistic for each of the response variables. Let t_j denote the two sample t-statistic for testing the difference between the means of treatments 1 and 2 on response variable j, $j = 1,2,\ldots.k$. The statistic computed in MULTTEST is maximum of the absolute values of the t-statistics

$$T_{max\ abs} = max(|t_1|,|t_2|,..,|t_k|).$$

The permutation p-value for the jth variate is the proportion of the permutation distribution of $T_{max\ abs}$ greater than or equal to the observed value of $|t_j|$. Because the permutation distribution is used as the reference distribution, the statistic may be applied to dichotomous data, as well as continuous data without concern about the violation of the normality assumption associated with the parametric test. One may also use bootstrap sampling, or sampling with replacement from the set of multivariate vectors, instead of permutation sampling. Bilder [2000] considered bootstrap sampling for a similar problem. One may also carry out a one-sided multivariate permutation test although this is not implemented in SAS®.

4. Hotelling T^2 Approximation

Suppose we have *n* observations from population 1 and *m* observations from population 2. There are k response variables for each population. The response matrices are represented by

$$X = \begin{bmatrix} x_{11} \;\dots\; x_{1k} \\ x_{21} \;\dots\; x_{2k} \\ . \qquad . \\ . \qquad . \\ . \qquad . \\ x_{n1} \;\dots\; x_{nk} \end{bmatrix} \qquad Y = \begin{bmatrix} y_{11} \;\dots\; y_{1k} \\ y_{21} \;\dots\; y_{2k} \\ . \qquad . \\ . \qquad . \\ . \qquad . \\ y_{m1} \;\dots\; y_{mk} \end{bmatrix} .$$

We assume that responses are distributed as multivariate normal with mean and covariance as shown below.

$X = [X_1 X_2 ......X_k] \sim MVN(\mu_X, \Sigma_X)$, where $X_i = [x_{i1}........x_{in}]'$, $i = 1, ...,n$.

$Y = [Y_1 Y_2 ......Y_k] \sim MVN(\mu_Y, \Sigma_Y)$, where $Y_i = [\,y_{i1}........y_{im}]'$, $i = 1, ...,m$.

Hotelling's T^2 statistic**,** which assumes that $\Sigma_X = \Sigma_Y = \Sigma$, is given by

$$T^2 = \frac{mn}{n+m}\left(\bar{X} - \bar{Y}\right)'\left(S_{pooled}\right)^{-1}\left(\bar{X} - \bar{Y}\right), \qquad (2)$$

where $S_{pooled} = \dfrac{(n-1)S_X + (m-1)S_Y}{n+m-2}$ and S_X, S_Y are sample variance–covariance matrices of X and Y respectively. Under the null hypothesis $F = \dfrac{m+n-k-1}{(n+m-2)k}T^2$ has an F-distribution with degrees of freedom k and $m+n-k-1$.

Now suppose the data have dichotomous as well as continuous responses. Because of the Central Limit Theorem, the analysis of univariate dichotomous data may be done with normal approximations for large samples. The approximations are generally good even for moderate sample sizes if the population proportions are not too close to 0 or 1. The question of interest here is the possible use of multivariate normal methods to analyze bivariate mixed data.

The suggested approach is to apply (2) directly to the mixed data. The dichotomous variable is assumed to be quantitative. The expected values are given by

$$E(X) = \mu_x = \left[p_{x1} \;\; \mu_{x2}\right] \text{ and } E(Y) = \mu_y = \left[p_{Y1} \;\; \mu_{Y2}\right],$$

where $p_{x1} = P(X_1 = 1)$, $\mu_{x2} = E(X_2)$, $p_{y1} = P(Y_1 = 1)$, and $\mu_{y2} = E(Y_2)$.
The covariance between X_1 and X_2 can be computed by

$$Cov(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2).$$

First, let us find $E(X_1 X_2)$.

$E(X_1 X_2) = E(X_1 X_2 | X_1 = 1)P(X_1=1) + E(X_1 X_2 | X_1 = 0)P(X_1=0)$
$= (\mu_{21}) p_{x1},$ where $\mu_{21} = E(X_2 | X_1=1)$

The covariance is

$$Cov(X_1, X_2) = (\mu_{21}) p_{x1} - p_{x1} \mu_{x2}$$
$$= p_{x1}(\mu_{21} - \mu_{x2}).$$

We can write $\mu_{x2} = (p_{x1})\mu_{21} + (1- p_{x1})\mu_{20}$, where $\mu_{20} = E(X_2 | X_1=0)$.

Substituting the above equation,

$$Cov(X_1, X_2) = p_{x1}(1- p_{x1}) (\mu_{21} - \mu_{20}).$$

Similarly the covariance between Y_1 and Y_2 is given by

$Cov(Y_1, Y_2) = E(Y_1Y_2) – E(Y_1)E(Y_2).$

First, let us find $E(Y_1Y_2)$.

$E(Y_1 Y_2) = E(Y_1 Y_2 | Y_1 = 1)P(Y_1=1) + E(Y_1 Y_2 | Y_1 = 0)P(Y_1=0)$
$= (\mu'_{21}) p_{y1},$ where $\mu'_{21} = E(Y_2 | Y_1=1)$

The covariance is

$$Cov(Y_1, Y_2) = (\mu'_{21}) p_{y1} - p_{y1} \mu_{y2}$$
$$= p_{y1}(\mu'_{21} - \mu_{y2}).$$

We can write $\mu_{y2} = (p_{y1}) \mu'_{21} + (1- p_{y1}) \mu'_{20}$, where $\mu'_{20} = E(Y_2 | Y_1=0)$.

Substituting the above equation,

$$Cov(Y_1, Y_2) = p_{y1}(1- p_{y1}) (\mu'_{21} - \mu'_{20}).$$

Let $\sigma^2_{21} = V(X_2 | X_1 = 1)$ and $\sigma^2_{20} = V(X_2 | X_1 = 0)$, and similarly $\sigma'^2_{21} = V(X_2 | X_1 = 1)$ and $\sigma'^2_{20} = V(X_2 | X_1 = 0)$. The variance-covariance matrix for X can be written as

$$V(X) = \Sigma_x = \begin{bmatrix} p_{x1}(1-p_{x1}) & p_{x1}(1-p_{x1})(\mu_{21}-\mu_{20}) \\ p_{x1}(1-p_{x1})(\mu_{21}-\mu_{20}) & \sigma_{x_2}{}^2 \end{bmatrix}, \quad (3)$$

where $\sigma^2_{X2} = p_{x1}(\sigma^2_{21}) + (1-p_{x1})\sigma^2_{20} + p_{x1}(1-p_{x1})(\mu_{21}-\mu_{20})^2$.

The variance-covariance matrix for Y can be written as

$$V(Y) = \Sigma_y = \begin{bmatrix} p_{y1}(1-p_{y1}) & p_{y1}(1-p_{y1})(\mu'_{21}-\mu'_{20}) \\ p_{y1}(1-p_{y1})(\mu'_{21}-\mu'_{20}) & \sigma_{y_2}{}^2 \end{bmatrix}, \quad (4)$$

where $\sigma_{Y2}^2 = p_{y1}\sigma_{21}'^2 + (1-p_{y1})\sigma_{20}'^2 + p_{y1}(1-p_{y1})(\mu_{21}' - \mu_{20}')^2$

The sample statistics are given by

$$\hat{p}_{x1} = \frac{\sum_{i=1}^{n} x_{i1}}{n},\ \overline{x}_2 = \frac{\sum_{i=1}^{n} x_{i2}}{n} \text{ and } \hat{p}_{y1} = \frac{\sum_{i=1}^{m} y_{i1}}{m},\ \overline{y}_2 = \frac{\sum_{i=1}^{m} y_{i2}}{m},$$

$$\hat{\Sigma}_x = \begin{bmatrix} \hat{p}_{x1}(1-\hat{p}_{x1}) & \hat{p}_{x1}(1-\hat{p}_{x1})(\overline{x}_{21} - \overline{x}_{20}) \\ \hat{p}_{x1}(1-\hat{p}_{x1})(\overline{x}_{21} - \overline{x}_{20}) & \hat{\sigma}_{x2}^2 \end{bmatrix},$$

$$\hat{\Sigma}_y = \begin{bmatrix} \hat{p}_{y1}(1-\hat{p}_{y1}) & \hat{p}_{y1}(1-\hat{p}_{y1})(\overline{y}_{21} - \overline{y}_{20}) \\ \hat{p}_{y1}(1-\hat{p}_{y1})(\overline{y}_{21} - \overline{y}_{20}) & \hat{\sigma}_{y2}^2 \end{bmatrix} \quad (5)$$

The unbiased estimates of the variance-covariance matrices for the two groups are $S_X = \frac{n}{n-1}\hat{\Sigma}_x$ and $S_Y = \frac{m}{m-1}\hat{\Sigma}_y$. Under the assumption that $\Sigma_x = \Sigma_y = \Sigma$, we use the usual pooled estimate S_{pooled} to estimate Σ and then apply the formula for T^2 defined in (2) to the bivariate mixed data.
A modification of this procedure is to use the estimated variance-covariance matrix defined by

$$\hat{\Sigma} = \frac{n\hat{\Sigma}_x + m\hat{\Sigma}_y}{n+m-2}. \quad (6)$$

The test statistic for testing the hypothesis H_0: $\mu_x = \mu_y$ is given by

$$\hat{T}^2 = \frac{mn}{m+n}\left(\hat{\mu}_x - \hat{\mu}_y\right)' \hat{\Sigma}^{-1}\left(\hat{\mu}_x - \hat{\mu}_y\right), \quad (7)$$

where $\hat{\mu}_x = \left[\hat{p}_{x1}\ \hat{\mu}_{x2}\right]$ and $\hat{\mu}_y = \left[\hat{p}_{y1}\ \hat{\mu}_{y_2}\right]$. As with the use of T^2 we assume that the distribution of $F = \frac{m+n-2-1}{(n+m-2)2}\hat{T}^2$ approximately follows an F-distribution with numerator degrees of freedom 2 and denominator degrees of freedom $n + m - 3$. This test statistic and the test statistic in (2) applied directly to dichotomous data are evaluated with respect to probability of Type I error rate.

5. Simulation

In order to study the performance of these tests, random samples were generated from bivariate distribution. We generated Bernoulli distribution for one variable and a normal distribution for the other. The two variables have to be correlated for bivariate analysis to be relevant. We used SAS® for our simulation study.

The first variable X_1 was generated from Bernoulli(p_{x1}). The second variable X_2 was generated conditional on X_1. Let X_{21} be the second variable X_2 corresponding to X_1 =1 and X_{20} be the second variable X_2 corresponding to X_1 =0. Therefore, a $X_{21} \sim N(\mu_{21}, \sigma_{21}^2)$ and $X_{20} \sim N(\mu_{20}, \sigma_{20}^2)$, where

$$\mu_{21} = E(X_2 \mid X_1 = 1),\ \sigma_{21}^2 = V(X_2 \mid X_1 = 1),$$
$$\mu_{20} = E(X_2 \mid X_1 = 0), \text{ and } \sigma_{20}^2 = V(X_2 \mid X_1 = 0).$$

The values of μ_{21} and μ_{20} were chosen in such a way so that the difference μ_{21} - μ_{20} is 0, 5, 10, 20, and 30. Each of these mean differences were taken in combination with $\frac{\sigma_{21}}{\sigma_{20}} = 1, 2, \text{and } 3$. The variable 1 ($X_1$) was generated using different values of p_{x1}. The values of p_{x1} considered are 0 .3, 0.4, 0.6, and 0.7 for n = 20, and for n = 40 the values of p_{x1} considered are 0 .3 and 0.6. The same procedure was repeated for the other population. The variance-covariance matrix was the same for both cases under H_0. For this study n and m were equal.

The two statistics given in (2) and (7) were computed. Each combination of p_{x1}, μ_{21}, μ_{20}, σ_{21}, σ_{20}, n, and m was repeated 5000 times and the test statistics were computed at each repetition. The Type I error rate is taken to be the relative frequency with which the test statistics given by (2) and (7) exceeded the critical value in 5000 replications. The critical value is computed at 5% significance level.

6. Conclusions

Probability of Type I errors is given in Table 1 through 6. The sample sizes considered are 20 and 40. There are two values of Type1 error rates- Alpha_UB using the test statistic given in (2) and Alpha_B by using the test statistic given in (7).

The statistic T^2 defined in (2) does a good job of controlling the Type I error rate for the cases considered. The elements of the variance-covariance matrix S_{pooled} in (2) are slightly larger than those of $\hat{\Sigma}$ for (7). Because the statistics involve the inverse of the variance-covariance matrix, it follows that T^2 is smaller than $\hat{T}^2$, and so the Type I error rate of $\hat{T}^2$ will be higher as is evident from the table. However, the difference is particularly large for higher values of $\mu_{21} - \mu_{20}$. The variance-covariance matrices Σ_x and Σ_y contain the square of the difference term $\mu_{21} - \mu_{20}$.

The ratio of standard deviation also affects the Type I error rate. For N=20, Type I error rate decreases as the ratio increases for all the values of ‘p’. But when N= 40, the Type I error rate remains relatively constant. The simulation results indicate that for the cases considered the probability of Type I error is reasonably close to the nominal level of .05. For moderately large differences $\mu_{21} - \mu_{20}$, both test statistics have reasonably good control over Type I error rate. Thus, in terms of controlling Type I error, the direct application of Hotelling's T^2 appears to be an acceptable methodology for analyzing bivariate mixed data.

7. Future Investigation

The proposed test for bivariate mixed data was studied in terms of controlling the Type I error rate. The power of this test deserves further investigation.
We are also studying that the corresponding F statistics indeed have an approximate F-distribution.

Table 1

N=20 P=0.30

	$\mu_{21} - \mu_{20}$	Alpha_UB	Alpha_B
$\frac{\sigma_{21}}{\sigma_{20}} = 1$	0	.046	.049
	5	.048	.050
	10	.052	.037
	20	.054	.037
	30	.047	.023
$\frac{\sigma_{21}}{\sigma_{20}} = 2$	0	.052	.056
	5	.049	.049
	10	.048	.042
	20	.045	.034
	30	.051	.031
$\frac{\sigma_{21}}{\sigma_{20}} = 3$	0	.046	.050
	5	.039	.039
	10	.049	.047
	20	.041	.028
	30	.042	.020

Table 2

N=20 P=0.40

	$\mu_{21}-\mu_{20}$	Alpha_UB	Alpha_B
$\frac{\sigma_{21}}{\sigma_{20}}=1$	0	.046	.049
	5	.049	.050
	10	.050	.045
	20	.057	.037
	30	.047	.031
$\frac{\sigma_{21}}{\sigma_{20}}=2$	0	.047	.049
	5	.048	.048
	10	.056	.045
	20	.045	.034
	30	.053	.034
$\frac{\sigma_{21}}{\sigma_{20}}=3$	0	.049	.052
	5	.044	.045
	10	.053	.048
	20	.044	.032
	30	.045	.024

Table 3

N=20 P=0.60

	$\mu_{21}-\mu_{20}$	Alpha_UB	Alpha_B
$\frac{\sigma_{21}}{\sigma_{20}}=1$	0	.047	.048
	5	.049	.051
	10	.049	.049
	20	.048	.043
	30	.049	.034
$\frac{\sigma_{21}}{\sigma_{20}}=2$	0	.049	.051
	5	.049	.049
	10	.052	.044
	20	.054	.033
	30	.049	.026
$\frac{\sigma_{21}}{\sigma_{20}}=3$	0	.053	.055
	5	.058	.057
	10	.050	.046
	20	.050	.041
	30	.041	.024

Table 4

N=20 P=0.70

	$\mu_{21}-\mu_{20}$	Alpha_UB	Alpha_B
$\frac{\sigma_{21}}{\sigma_{20}}=1$	0	.055	.057
	5	.057	.057
	10	.040	.036
	20	.061	.038
	30	.054	.025
$\frac{\sigma_{21}}{\sigma_{20}}=2$	0	.046	.050
	5	.045	.046
	10	.049	.047
	20	.055	.046
	30	.056	.042
$\frac{\sigma_{21}}{\sigma_{20}}=3$	0	.045	.048
	5	.049	.051
	10	.049	.049
	20	.048	.043
	30	.049	.034

Table 5

N = 40   P = 0.60

	$\mu_{21}-\mu_{20}$	Alpha_UB	Alpha_B
$\frac{\sigma_{21}}{\sigma_{20}}=1$	0	.049	.051
	5	.048	.048
	10	.046	.043
	20	.052	.037
	30	.054	.028
$\frac{\sigma_{21}}{\sigma_{20}}=2$	0	.049	.051
	5	.050	.050
	10	.049	.047
	20	.051	.042
	30	.054	.031
$\frac{\sigma_{21}}{\sigma_{20}}=3$	0	.058	.060
	5	.053	.054
	10	.052	.052
	20	.049	.044
	30	.050	.038

Table 6

N = 40 P = 0.30

	$\mu_{21}-\mu_{20}$	Alpha_UB	Alpha_B
$\frac{\sigma_{21}}{\sigma_{20}}=1$	0	.051	.052
	5	.049	.051
	10	.051	.048
	20	.048	.033
	30	.050	.025
$\frac{\sigma_{21}}{\sigma_{20}}=2$	0	.044	.048
	5	.049	.049
	10	.045	.042
	20	.044	.030
	30	.044	.024
$\frac{\sigma_{21}}{\sigma_{20}}=3$	0	.043	.043
	5	.044	.046
	10	.050	.050
	20	.043	.034
	30	.055	.035

Summary

In this paper, we consider the problem of comparing two treatments in experiments in which bivariate mixed response variables are measured on each experimental unit. A multivariate permutation test for this problem may be carried out using PROC MULTTEST in SAS®. The question of interest here is the possible use of multivariate normal methods to analyze bivariate mixed data. An approximate Hotelling T^2 test is proposed for bivariate mixed data and empirically evaluated in terms of Type I error rate. It is shown that the approximation does a good job of controlling the Type I error rate. Thus, Hotelling's T^2 can be used to compare the mean vectors of two populations having bivariate mixed data.

Acknowledgements

The authors wish to thank the referee(s) for their valuable comments and suggestions in improving the manuscript.

References

Agresti, A. and Liu, I.-M. (1999) *Modeling a Categorical Variable Allowing Arbitrary Many Category Choices*. Biometrics, 55, 935-943.

Anderson, T.W. (1984) *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, New York.

Bilder, C.R., Loughin, T.M., and Nettleton, D. (2000) *Multiple Marginal Independence Testing for pick any/c Variables*. Communications in Statistics: Simulation and Computation, 29(4), 1285-1316.

Blair, R.C., Higgins, J.J., Karniski, W., and Kromery, J.D. (1994) *A Study of Multivariate Permutation Tests which may repalce Hotelling's T-square Test in Prescribed Circumstances*. Multivariate Behavioral Research, 29, 141-163.

Cox, D.R. (1972) *The Analysis of Multivariate Binary Data*. Applied Statistics, 21, 113-120.

Goodman, L.A. and Kruskal, W.H. (1954) *Measure of Association for Cross Classifications*. Journal of the American Statistical Association, 49, 732-764.

Goodman, L.A. and Kruskal, W.H. (1959) *Measure of Association for Cross Classifications*. Journal of the American Statistical Association, 54, 123-163.

Goodman, L.A. and Kruskal, W.H. (1963) *Measure of Association for Cross Classifications*. Journal of the American Statistical Association, 58, 310-364.

Higgins, J.J. (2004) An *Introduction to Modern Nonparametric Statistics*. Thompson Brooks/Cole, Duxbury Advanced Series, Pacific Grove, CA.

Knocke, J.D. (1976) *Multiple Comparisons with Dichotomous Data*. Journal of the American Statistical Association, 71, 849-853.

Singh, P., Khamis, I., and Higgins J. (2003) *Hotelling T^2 Approximation for Bivariate Dichotomous Data*. Proceedings of 15th annual Kansas State University Conference on Applied Statistics in Agriculture, 218-226.