ON TESTS AND ESTIMATORS IN TWO-WAY DATA

By

A. P. Gore

Maharashtra Association for Cultivation of Science, Pune

(Received : March, 1977)

Introduction

In this paper we consider some problems in the analysis of either quantitative or qualitative (categorical) data laid out in rows and columns. This distinction between quantitative and qualitative data is by no means watertight. It is a common practice to convert quantitative observations into frequencies of classes for goodness of fit test and conversely to associate numerical values with categories. However, the models and procedures appropriate for the two cases do turn out to be different. We shall make rather weak assumptions about the underlying distributions and in that sense offer non-parametric procedures.

1. The Problem (Quantitative data)

Literature on non-parametric tests for one-way layouts is quite extensive (see *e.g.* Hodges and Lehmann [13] Kruskal [16], Sugiura [18] etc.). The results available for two-way data are relatively fewer. Perhaps the only well known test for the problem of main effects is due to Friedman [8]. We propose to describe in the sequel some new tests for the same problem.

Consider the linear model

$$X_{ijk}=\mu+A_i+B_j+E_{ijk}$$

where μ is the general mean, A_i $(i=1, 2,..., r)$ effect of ith row, B_j $(j=1, 2..., c)$ effect of jth column and E_{ijk} $(k=1, 2,..., n_{ij})$ is a random error component. These error components are assumed to be all independent and identically distributed with a common continuous distribution F with median zero. In addition, we make the usual assumptions that A_i summed over i and B_j summed over j are equal to zero. The hypothesis to be tested is

$$H_0 : A_1=A_2=...=A_r$$

i.e. the hypothesis of no row effect.

2. Test Statistics

(*i*) Let $W_{ii'j}$ denote the number of pairs of observations (x_{ijk}, $x_{i'jk'}$) such that the observation from ijth cell is greater than the observation from $i'j$th cell. Note that these two cells are in the same column (jth) and hence given H_0 either observation could be greater than the other with probability 1/2. Further, due to the assumption of continuity of F, ties occur with probability zero and can be ignored. Of course ties do occur in practice, but could be resolved by randomisation. It is clear that $W_{ii'j}$ is the Mann-Whitney statistic for the two cells in question. Let

$$U_{ii'j} = W_{ii'j}/n_{ij}\, n_{i'j}$$

$$U_i = \sum_j \sum_{i \neq 1} U_{ii'j}$$

It can be shown easily that if H_0 is true, expectation of U_i is $(r-1)\, c/1$. For the case $n_{ij}=n$ for all i and j we have the statistic

$$S_1 = \frac{12\, n}{r^2 c} \sum_{i=1}^{r} (U_i - (r-1)\, c/2)^2$$

More generally if the number of observations per cell is not the same, let $n_{ij} = Np_{ij}$, $O < p_{ij} < 1$, where sum of p_{ij} over all i and j is unity. Further let $q_{ij} = 1/p_{ij}$, $q_{i\cdot} = \sum_j (1/p_{ij})$.

$$q_{\cdot\cdot} = \sum_i q_{i\cdot} \text{ and } q^*_{\cdot\cdot} = \sum_i (1/q_{i\cdot})$$

Then for testing H_0 we can use the statistic

$$S_2 = \frac{12\, N}{r^2} \Big[\sum_1 (U_i - (r-1)\, c/2)^2 / q_{i\cdot}$$

$$- (\sum (U_i - (r-1)\, c/2)/q_{i\cdot})^2 / q^*_{\cdot\cdot} \Big]$$

For the case $r=2$ and $c=1$ (two-sample problem) S_2 is proportional to the square of the Mann-Whitney statistic. In this sense the latter is generalized here. Further, for the case $n_{ij}=n$ and $c=1$ (any r) let R_{i1k} denote and rank of x_{i1k} among all observations in the r samples and let $I(t)$ be a function which equals unity if argument is positive and zero otherwise.

Then

$$R_{ilk}=\sum_{k'} I\ (x_{ilk}-x_{ilk'})+\sum_{i\neq i'}\ \sum_{k'} I\ (x_{ilk}=x_{i'lk'})$$

while

$$n^2\,U_{ii'}\,l=\sum_{k}\ \sum_{k'} I(x_{ilk}-x_{ilk'})$$

Thus $n^2\ U_i$ differs from sum of rank of all observations in the ith cell by a constant $(n(n-1)/2)$. In this sense S_1 is a generalization of the Kruskal-Wallis H statistics to the case of two-way data.

(ii) Consider r-plets of observations formed by selecting one observation from each of the r cells in a column. There are $\prod_i n_{ij}$ such r-plets that can be formed from observations in the jth column. Let V_{ij1} (V_{ijr}) denote the number of such r-plets in which the observation selected from the ijth cell is smallest (largest).

Let $U_{ijt}=V_{ijt}/\prod_i n_{ij},\ t=1,\ r,$

$U_{it}=\sum_j\ U_{ijt}$ and $a_i=U_{i1}-c/r$

Under the null hypothesis all orderings among the components of an r-plet are equally likely and expectation of a_i is zero. This fact is used in constructing the statistic B which generalizes Bhapkar's v-test (see Bhapkar, 1961)

$$B=N\ (2r-1)\left[\sum_i\ (a_i^2\ /q_{i\cdot}\ \ -\left(\sum_i a_i/q_{i\cdot}\right)^2/q^*_{\cdot\cdot}\right]$$

(iii) Let $d_{ij}=U_{ijr}-U_{ij1}$ and $d_i=\sum_j\ d_{ij}$.

Again, under the null hypothesis expectation of d_i is zero. The statistic T based on d_is is

$$T=\frac{N\ (2r-1)\ (r-1)^2\binom{2r-2}{r-1}}{2r^2\left[\binom{2r-2}{r-1}-1\right]}\left[\sum_i\ d_i^2\ /q_{i\cdot}\ -\left(\sum d_i/q_{i\cdot}\right)^2/q^*_{\cdot\cdot}\right]$$

For the case of 1-column (or essentially the r-sample problem) $T(B)$ reduces to the statistic $L\ (V)$. (see Deshpande, [7]).

The tests consist in rejecting H_0 at a chosen level if the observed value of the statistic exceeds the appropriate upper value of the chi-square distribution with $r-1$ degrees of freedom. It has been shown (see Gore [9], Tan and Gore [19] and Gore [10]) that each of these statistics has a limiting chi-square distribution with $r-1$ degrees of freedom as N tends to infinity such that n_{ij}/N (positive) remains fixed for all i and j.

It may be noted that while using these statistics no distinction need be made between orthogonal and non-orthogonal data.

3. Efficiency

Efficiency characteristics of a test do not always remain unchanged after generalization. Thus Kruskal-Wallis H test, Bhapkar's V-test and Deshpande's L-test are all generalizations of the Mann-Whitney (Wilcoxon) test and still do not have identical asymptotic relative efficiency (A.R.E.). H test is usually recommended in preference to L test to detect differences in locations of symmetric distributions with thin tails while L is suggested for distributions with thick tails. V test is especially useful for asymmetric distributions. Further, A.R.E. of V or L test compared to ANOVA F-test is not the same as A.R.E. of the Mann-Whitney U test when compared to t test. However the generalizations offered here have the same A.R.E. as the original tests.

Consider the sequence of alternatives

$$H : A_i = N^{-1/2}\,\theta_i,\ i=1, 2, \ldots, r$$

where θ_is (real constants) add upto zero. It has been shown (in Gore [9], Tan and Gore [19], Gore [10] that under H each of S_2, B and T has a limiting non-central chi-square distribution with $r-1$ degrees of freedom and appropriate non-centrality parameter as listed below.

$$S_2 : 12\,C^2\,(\phi\,(F))^2\,D$$

where

$$\phi\,(F)=\int_{-\infty}^{\infty} f^2\,(x)\,dx$$

and

$$D=\Big[\sum_i \theta_i^2/q_{i\cdot} - \Big(\sum_i \theta_i^2/q_{i\cdot}\Big)^2/q^*_{\cdot\cdot}\Big]$$

$$B : c^2r^2\,(2r-1)\,DM_2^2$$

where

$$M_2(F)=\int_{-\infty}^{\infty}(1-F(x))^{r-2}f^2(x)\,dx$$

and

$$T: \frac{c^2 \ (2r-1) \ (r-1)^2\binom{2r-2}{r-1}}{2\left[\binom{2r-2}{r-1}-1\right]}(M_1+M_2)^2 \ D$$

where

$$M_1(F)=\int_{-\infty}^{\infty}F^{r-2}(x)\,f^2(x)\,dx$$

It follows immediately that A.R.E. (which can be expressed as the ratio of the corresponding non-centrality parameters) of B test relative to S_2 is

$$r^2(2r-1)\,M_2^2\,/12\,(\phi(\mathrm{F}))^2$$

which is the same as A.R.E. of V test relative to H test. Similarly A.R.E. of T test relative to B test is

$$\frac{(r-1)^2\binom{2r-2}{r-1}}{2r^2\left[\binom{2r-2}{r-1}-1\right]}\;\frac{(M_1+M_2)^2}{M_2^2}$$

which is the same as A.R.E. of L test relative to V test. It follows that earlier remarks about comparison of H, V, and L tests apply to these generalizations as well. Hence S_2 should be used for underlying distribution F which is symmetric with thin tails while T is more useful for symmetric F with thick tails. B test is preferable for asymmetric distributions.

For comparison of these tests with parametric ANOVA F-test the cases of equal and unequal number of observations per cell have to be considered separately.

Case 1. $n_{ij}=n$ for all i and j. In this case the F statistic has, under the sequence of alternatives H, a limiting non-central chi-square distribution with $r-1$ degrees of freedom and non-centrality porameter.

$$\sum \theta_i^2 \,/\, r\sigma^2$$

where σ^2 is the variance of F. (*see* Graybill, [11]). Using this result it can be immediately verified that A.R.E. of S_1, B and T tests relative to F test are the same as A.R.E. of H, V and L tests respectively relative to the one-way ANOVA F-test.

Case 2. Unequal number of observations per cell. Here we encounter a technical difficulty in the comparisons. The ratio of the non-centrality parameters now becomes dependent on the particular sequence of alternatives H (or on θ_i's) and hence cannot be interpreted easily. We, therefore, consider another parametric test, namely the weighted-squares of means test (Y) due to Yates [20] (*see* also Bancroft, [1], p. 24). It can be easily shown that the Yates' statistic has under H, a limiting non-central chi-square distribution with $r-1$ degrees of freedom and non-centrality parameter

$$c^2 D/\sigma^2$$

It follows that A.R.E. of B test relative to Y test is

$$r^2 (2r-1) M_2^2 \sigma^2$$

which is precisely the same as A.R.E. of V-test relative to F-test. Analogous results are immediate for S_2 and T.

4. *The Problem* (Qualitative or Categorical Data)

Analysis of an $r \times c$ contingency table for testing the hypothesis of either homogenity of populations or independence of attributes is by now common enough to have found a place in almost all elementary text books of statistics. A two-way contingency table with number of rows necessarily equal to the number of columns is of special interest though encountered less frequently. Such a table may arise if a character is measured by two methods or by the same method on each of the twins. Here the hypothesis of interest is symmetry of the data. If P_{ij} denotes the probability of an individual belonging to the ijth cell (sum of all P_{ij} s $i, j=1, 2, \ldots, r$ must be unity), the hypothesis of symmetry across the diagonal may be defined as

$$H_o : P_{ij} = P_{ji} \text{ for all } i \text{ and } j$$

A test for this hypotheis has been proposed by Bowkar [5].

A hypothesis related to H_o is the hypothesis of symmetry of marginal distributions given by

$$H_1 : P_{io} = P_{oi}, \text{ for all } i$$

where the zero denotes sum over the subscript. Stuart [17], Bhapkar [4], Grizzle *et al.* [12], Ireland *et al.* [14] Koch and

Reinfurt [15] have studies the problem of testing H_1. Recently Belle and Cronell [2] proposed a statistic to test "whether the degree of asymmetry is the same in all categories". This notion can be made explicit in the model in terms of a third hypothesis.

$$H_2 : P_{ij} = \theta P_{ji}, \text{ for all } i, j \ (i \text{ less than } j).$$

Since for the symmetric case $\theta = 1$ and deviation of θ from 1 indicates asymmetry, θ may be called "degree of asymmetry". It is obvious that both H_1 and H_2 are weaker than H_0 in the sense that whenever H_0 holds H_1 (and H_2) holds as well while the converse is not necessarily true. It is estimation of θ and testing of H_2 which needs systematic exploration. Observe that H_1 and H_2 are by no means equivalent. In fact H_1 and H_2 are both true if and only if H_0 is true.

5. Estimation and Goodness of Fit Test

Under the hypothesis H_2 one can obtain the following m.l. estimators θ and P_{ij} by differentiating the log likelihood in the usual way.

$$\hat{P}_{ii} = n_{ii}/N, \qquad i = 1, 2, \ldots, r$$

$$\hat{P}_{ij} = \hat{\theta}\,\hat{P}_{ji} \quad i<j, \ \hat{P}_{ji} = (n_{ij} + n_{ji})/N\,(1+\hat{\theta}) \ i<j$$

and

$$\hat{\theta} = \sum_{i<j} n_{ij} \Big/ \sum_{i<j} n_{ji} = N_1/N_2 \text{ (say)}$$

Note that θ is the ratio of the probability of an observation belonging to a cell above the diagonal (P_{ij}, $i<j$) and the corresponding cell below the diagonal. The m.l. estimator of θ given here namely sum of all observed frequencies above the diagonal divided by sum of all observed frequencies below the diagonal, appears heuristically attractive.

Using these estimates, expected frequencies and usual Pearson chi-square statistic can be calculated to test H_2. The degrees of freedom for the asymptotic distribution of this statistic will be $[r(r-1)/2]-1$. Compared to the Bowker statistic for H_0, one additional degree of freedom is lost here. This is because an extra parameter θ is estimated from the data.

The pearson chi-square statistic is nothing but

$$X^2 = \sum_{i<j} \frac{[n_{ij} - \hat{\theta}\,(n_{ij}+n_{ji})/(1+\hat{\theta})]^2}{\hat{\theta}\,(n_{ij}+n_{ji})/(1+\hat{\theta})} + \sum_{i<j} \frac{[n_{ji} - (n_{ij}+n_{ji})/(1+\hat{\theta})]^2}{(n_{ij}+n_{ji})/(1+\hat{\theta})}$$

which after a little algebra reduces to

$$\sum_{i<j} (N_1\, n_{ji} - N_2\, n_{ij})^2 / N_1 N_2\, (n_{ij}+n_{ji})$$

which is precisely the ad-hoc statistic proposed by Belle and Cornel. We only claim to have justified its use. Note that the diagonal frequencies do not play any role here as is to be expected since the model H_2 is not related to the corresponding probabilities.

6. Application

A model can be claimed to be useful only if it helps us understand some data better than other available models. This turns out to be the case for the data from case records of the eye testing of 7477 employees in Royal Ordnance Factories (see Table 1). The

TABLE 1

Unaided Distance Vision of Women
(with expected frequencies under H_2 in brackets

	Right Eye	*Left Eye*			
		Highest grade	*Second grade*	*Third grade*	*Lowest grade*
RIGHT EYE	Highest grade	1520	266 (268.4542)	124 (129.3950)	56 (54.7646)
	Second grade	234 (221.5458)	1512	432 (426.3052)	78 (85.9054)
	Third grade	117 (111.6051)	362 (267.6947)	1772	205 (206.1729)
	Lowest grade	36 (47.2353)	82 (74.0947)	179 (177.8272)	492

hypothesis H_1 was tested for the data by several authors using different tests and in each case it was rejected. Thus we know that the data do not exhibit symmetry of marginl distributions. It is therefore of interest to check whether the data exhibit homogeneity of degree of asymmetry. Table 1 includes expected frequencies under H_2 in brackets in off diagonal locations. For diagonal locations, expected frequencies are the same as observed. Here estimates of θ is 1.1594 and the Pearson chi-square (5 d.f.) is 6.2611. This value is less than the upper 20% point of the chi-square distribution (7.289). Thus H_2 is accepted.

SUMMARY

It is often possible to generalise a test for two-samples or several samples problem to the problem of two-way data. Such generalised versions do not necessarily have the efficiency characteristics of the parent test. We have constructed tests for two-way data by simply combining statistics for every row (or every pair of cells). The following tests are generalised. Mann-Whitney U-test, Bhapkar's V-test, Kruskal-Wallis H-test, Deshpande's L-test. In all the cases, the generalised version has the same efficiency as the parent in comparison with appropriate parametric counterpart.

In two-way contingency tables with number of rows equal to number of columns, a hypothesis of interest is that of symmetry. Various versions of this (symmetry across the diagonal, symmetry of marginal distributions) have been studied in literature. Recently Belle and Cornell (1971) proposed the possibility of homogeneity of the degree of asymmetry and suggested an adhoc test. We offer m.l. estimator of the degree of asymmetry, associated Pearson chi-square test and show that the model fits a well known data used by Stuart.

ACKNOWLEDGMENT

I wish to thank Prof. O.P. Bagai and Dr. J.V. Deshpande for inviting me to participate in the seminar in Statistical Inference (21 to 25 December, 1976) at the Department of Statistics, Panjab University, Chandigarh, for which this paper was prepared.

REFERENCES

[1] Bancroft, T. A. (1968) : *Topics in Intermediate Statistical Methods*, Vol. 1, the Iowa State University Press, Ames.

[2] Van Belle, G. and Cornel, R.G. (1971) : Strengthening tests of symmetry in contingency tables, *Biometrics*, **27**, 1074-78.

[3] Bhapkar, V.P. (1961) : A non-parametric test for the problem of several samples, *Ann. Math. Statist*, **32**, 1108-17.

[4] Bhapkar, V.P. (1966) : A note on the equivalence of two test criteria for hypotheses in categorical data, *J. Amer. Statist. Ass.*, **61**, 228-35.

[5] Bowker, A.H. (1948) : A test for symmetry in contingency tables., *J. Amer. Statist. Ass.* **43**, 572-574.

[6] Cochran, W.G. (1954) : Some methods for strengthening the common chi-square tests, *Biometrics*, **10**, 417-51.

[7] Deshpande, J.V. (1965) : A non-parametric test based on U-statistic for problem of several samples, J. *Indian Statist. Ass.*, **3**, 20-29.

[8] Friedman, M. (1937) : The use of ranks to avoid the assumption of normality implicit in the analysis of variance, *J. Amer. Statist. Asso.*, **32**, 657-701.

[9] Gore, A.P. (1975) : Some non-parametric tests and selection procedures formain effects in two-way layouts, *Ann. Inst. Stat. Math.* (Tokyo), **27**, 487-500.

[10] Gore, A.P. (1976) : A distribution free test for randomised blocks, Unpublished.

[11] Graybill, F.A. (1961) : *An Introduction to Linear Statistical Models*, Vol **1**, McGraw-Hill, New York.

[12] Grizzle, J.E., Starmer, C.F. and Koch, G.G. (1969) : Analysis of categorical data by linear models, *Biometrics*, **25** 480-504.

[13] Hodges, J.L. and Lehmann, E.L. (1962) : Rank methods for combination of independent experiments in analysis of variance, *Ann. Math. Statist*, **37**, 482-97.

[14] Ireland, C.T., Ku, H.H. and Kullback, S. (1969) : Symmetry and marginal homogeneity of an $r \times r$ contingency table *J. Amer. Statist. Ass.*, **64**, 1323-41.

[15] Koch, G.G. and Reinfurt, D.W. (1971) : The analysis of categorical data from mixed models, *Biometrics*, **27**, 157-73.

[16] Kruskal, W.H. (1952) : A non-parametric test for the several sample problem, *Ann. Math. Statist*. **23**, 525-40.

[17] Stuart, A. (1955) : A test for homogeneity of marginal distributions in a two-way classification, *Biometrika*, **42**, 412-16.

[18] Sugiura, N. (1965) : Multisample and multivariate nonparametric tests based on U Statistics and their asymptotic efficiencies, *Osaka J.* Math, **2**, 385-426.

[19] Tan, K.C. and Gore, A.P.(1976) : A distribution free test for two-way layout. *Australian J Statist*. **18**, 151-57.

[20] Yates, F. (1934) : The analysis of multisample classifications with unequal numbers in the different classes, *J. Amer. Statist. Ass.* **29**, 51-66.