Brit. J. prev. soc. Med. (1958), **12**, 141–146

A NOTE ON THE INTERPRETATION OF $n \times 2$ TABLES

BY

J. H. EDWARDS

From the Department of Social Medicine, University of Birmingham

Data are frequently acquired or extracted in the form of a contingency table of two rows and several columns, and it is sometimes necessary to consider the distribution of entries into the various cells in terms of consistency with some hypothesis.

Where the ranking order of the columns is of no consequence, the conventional and widely-recommended χ^2 tests may be appropriate, provided the totals in the two rows are not grossly disproportionate in number. If they are grossly disproportionate in number tests based on the ratio of the observed to the expected variance on the assumption of a Poisson distribution are more efficient (Berkson, 1940; Cochran, 1954). These cases in which the ranking order of the columns is inconsequential are comparatively rare, except in relation to genetic segregation.

Almost always the order of the columns is meaningful, and any test which fails to appreciate this is clearly inefficient. The conventional χ^2 test of contingency gives a result which is uninfluenced by arranging the columns in any of $n!$ ways, and this result is related to $n-1$ degrees of freedom, only one or two of which are likely to be related to any biologically meaningful variation. In addition, it is unrelated to any intuitively meaningful estimator.

Nevertheless, cases in which these tests are used on $n \times 2$ Tables in which the order of the columns is obviously meaningful, as in degrees of exposure to therapies, toxins, etc., or in degrees of variation by such attributes bounded by heights, weights, or blood pressures, are very common.

Frequently sufficient information is available to do a "t" test, if necessary after some normalizing transformation, and this will always be preferable to grouping bounded variates as attributes and applying statistical methods based on the sampling of attributes.

The only occasion on which $n \times 2$ Tables can be regarded as an efficient representation of data for analysis is when the ranking order of groups of variates can be specified but there are no estimates of their relative value, or when, as in the case of the distribution of birth order or family size, the variates are limited to integral values.

If data are in the form of an $n \times 2$ Table, and a simple inefficient test for trend is required, it is preferable to lose information by removing partitions, converting the table into a 2×2 Table, than to lose precision by indulging in a multiplicity of degrees of freedom.

This advice need not permit a choice of grouping after examination of the data: the most efficient arrangement, that which maximizes the harmonic mean of the class frequencies or minimizes the sum of their reciprocals (Edwards, 1957b), may form the basis of a simple objective rule. The proportion of information lost cannot exceed

$$1-\frac{2}{\pi} \text{ or } 36\cdot 3 \text{ per cent.}$$

Many methods have been advanced to overcome the obvious inefficiency of either ignoring ranking order or losing the information provided by some of the partitions. Pearson's biserial correlation coefficient is now rarely used, as it is complicated, the relevant estimator is not always meaningful, and there is no exact method of determining its standard error. Simple regression methods, based on the logistic approximation or on Berkson's Normit Tables (Berkson, 1957), will overcome some of these difficulties. The problem, however, seems rare in practice, although its logistic approximation may be simulated when a disease is predisposed to by a logarithmic dose response and comparable numbers of control and diseased persons are interrogated in relation to their dosage (Edwards, 1957a).

Yates (1948) presented a simple scoring method for the analysis of trend, and suggested that the score should be based on integral differences. Similar methods have been advanced by Cochran (1954) and Armitage (1955). These have, as their exponents point out, the disadvantage that the scoring is not objective, and that to decide on a scoring system before seeing the data is in practice unrealistic. This can be a very serious objection as when the trend in proportion is constant in direction it is always possible to choose a scoring system which will lead to an infinite variance ratio, and will implicate a single degree of freedom as responsible for the whole of the contribution to the conventional χ^2 test. A consistent trend will be expected from chance alone with the probability of $\frac{2}{n!}$.

A further difficulty is that the investigator has apparently the choice of either basing a test of significance on the variance ratio or on the contribution to linear regression (both of which, incidentally, are negatively correlated with the remaining components of variance which are also included in the conventional χ^2 test if there really is a trend). In addition, it is not clear what meaning should be attached to heterogeneity relevant to an arbitrary scoring system even in a non-biological framework.

Tests of trend in $n \times 2$ Tables are usually considered in terms of regression. It is perhaps simpler to consider the problem in terms of mean scores, which, as Yates (1948) has shown, lead to formally identical tests of significance.

Consider the case in which an event has a probability $p_a, p_b \ldots p_n$ of falling in exclusive and exhaustive categories $a, b, \ldots n$, and let these categories be arbitrarily given numerical scores $s_a, s_b \ldots s_n$.

Then, after N events, the expected mean score will be $\sum ps$, with variance

$$\frac{1}{N}\left(\sum ps^2 - \left(\sum ps\right)^2\right).$$

In the case of an $n \times 2$ Table, as in Table I, the method allows a simple solution to the problem of whether the two rows could reasonably be regarded as samples from the most likely parent population from which the marginal totals may be regarded as samples.

TABLE I

SYMBOLISM DESIGNATING FREQUENCIES IN AN $n \times 2$ TABLE

						Total
	p_1	p_2	p_3		p_n	P
	q_1	q_2	q_3		q_n	Q
Total	n_1	n_2	n_3		n_n	N
Score	s_1	s_2	s_3		s_n	

The expectations of the mean scores of each row will be:

$$\frac{1}{N}\sum ns,$$

and writing

$$d_p = \frac{1}{P}\sum ps - \frac{1}{N}\sum ns$$

$$d_q = \frac{1}{Q}\sum qs - \frac{1}{N}\sum ns$$

$$V_p = \frac{1}{P}\left(\frac{1}{N}\sum ns^2 - \left(\frac{1}{N}\sum ns\right)^2\right)$$

$$V_q = \frac{1}{Q}\left(\frac{1}{N}\sum ns^2 - \left(\frac{1}{N}\sum ns\right)^2\right),$$

$$\frac{d_p^2}{V_p} + \frac{d_q^2}{V_q}$$

will be distributed as a χ^2 with one degree of freedom, and this distribution will be true of any scoring system devised *in ignorance* of the body of the Table.

We may take as an example the numbers of infants of birth rank 2, 3, and 4 in a group of infants suffering from pyloric stenosis (P) and a control group (Q) from the data of McKeown, MacMahon, and Record (1951). While a primiparity effect is common knowledge, it is not unreasonable to ask whether this can be explained as exclusively due to the obvious peculiarities of first births, or whether a trend persists in the 2nd, 3rd, and 4th births. We may do this by considering the mean birth rank in each group (the mean scores) and the significance of its difference (Table II, opposite).

Where the scores are obviously suggested by the data, as in the case above, no problem will arise. In the far commoner case in which the scores refer to such ranked adjectives as "worse", "same", "better", or some other largely arbitrary subdivision, a simple objective rule is required.

By transforming the scores so that they represent the abscissa of a rectangular distribution specified by the marginal totals, this objectivity can be achieved with considerably less computation than is

TABLE II

BIRTH RANK DISTRIBUTIONS OF CASES OF PYLORIC STENOSIS AND CONTROLS

Pyloric Stenosis	Con-trols		Birth Rank				
p	q	n	s	ps	qs	ns	ns^2
139 57 19	271 123 74	410 180 93	2 3 4	278 171 76	542 369 296	820 540 372	1640 1620 1468
215	468	683		525	1207	1732	4748
$\sum p$	$\sum q$	$\sum n$		$\sum ps$	$\sum qs$	$\sum ns$	$\sum ns^2$
Means				2·442	2·579	2·535	6·952

$d_p=-0\cdot 093 \qquad d_q=+0\cdot 044$

$V_p=\frac{1}{215}(6\cdot 952-2\cdot 535^2)=0\cdot 00244$

$V_q=\frac{1}{468}(6\cdot 952-2\cdot 535^2)=0\cdot 00112$

$\chi^2=\frac{d_p^2}{V_p}+\frac{d_q^2}{V_q}=5\cdot 3$

Difference in mean birth rank $=0\cdot 14$ in Ranks 2, 3, and 4.

required for the conventional $n\times 2$ χ^2 test. The logical basis is strictly analogous to using the "t" test after applying a normalizing transformation to the total set of observations. As well as being simple to perform, the test involves no reference to tables other than those of χ^2, or its square root, the normal deviate, and provides a very simple estimator of any linear trend. Its use as a test of significance is, of course, only appropriate in those rare circumstances in which a null hypothesis of zero trend is meaningful.

Consider Table I, which shows an $n\times 2$ Table of pairs of numbers $p_1, q_1 \ldots p_n, q_n$ and their marginal totals. On the null hypothesis of zero trend, the most likely proportions falling into the classes in the body of the Table, specified by a symbol in brackets, will be determined by the marginal frequencies, so that:

$$(p_1) = \frac{Pn_1}{N},\ (q_3) = \frac{Qn_3}{N},\ \text{etc.}$$

using brackets to specify expectations.

Consider now a series of scores linearly related to N and ranging between 0 and 1, so that:

$$s_1=\frac{\frac{1}{2}n_1}{N},\ s_2=\frac{n_1+\frac{1}{2}n_2}{N},\ s_k=\frac{n_1+n_2+\ldots\frac{1}{2}n_k}{N}.$$

Then, on the null hypothesis, the expectation of

$$\frac{1}{P}\sum ps=\frac{1}{Q}\sum qs=\frac{1}{2},$$

with sampling variance

$$\frac{1}{P}\left(\frac{1}{P}\sum(p)s^2-\left[\frac{1}{P}\sum(p)s\right]^2\right)=\frac{1}{P}\left(\frac{1}{N}\sum ns^2-\frac{1}{4}\right)$$

and

$$\frac{1}{Q}\left(\frac{1}{Q}\sum(q)s^2-\left[\frac{1}{Q}\sum(q)s\right]^2\right)=\frac{1}{Q}\left(\frac{1}{N}\sum ns^2-\frac{1}{4}\right)$$

respectively.

Call these variances V_p and V_q.

Then, writing

$$d_p=\frac{1}{P}\sum ps-\frac{1}{P}\sum(p)s=\frac{1}{P}\sum ps-\frac{1}{2}$$

$$d_q=\frac{1}{Q}\sum qs-\frac{1}{Q}\sum(q)s=\frac{1}{Q}\sum qs-\frac{1}{2},$$

$$\frac{d_p^2}{V_p}+\frac{d_q^2}{V_q}$$

is distributed as a χ^2 with one degree of freedom.

In the limiting case when $n=2$, this test of trend reduces to the conventional χ^2 test:

$$\frac{(p_1q_2-q_1p_2)^2}{(p_1+q_1)(p_2+q_2)(p_1+p_2)}+\frac{(p_1q_2-q_1p_2)^2}{(p_1+q_1)(p_2+q_2)(q_1+q_2)}$$

$$=\frac{(p_1q_2-q_1p_2)^2N}{(p_1+q_1)(p_2+q_2)(p_1+p_2)(q_1+q_2)}.$$

The calculation of $\frac{1}{P}\sum ps$, $\frac{1}{Q}\sum qs$, and V_p and V_q is extremely simple, and may be made easier by presenting the data in the vertical form as in Table III (overleaf).

An example is given in Table IV (overleaf) from date published by Record and McKeown (1949) on the work done by the mother during first pregnancies resulting in malformations of the central nervous system.

In the limiting form when one row total is far less than the other, so that the two with the smaller frequencies may be assumed to follow the Poisson distribution, and the expectations in each group are equal, so that the data present as a single series of numbers, as, for example, deaths per year from some disease for several years in an equal population at risk*, a test for trend is extremely simple.

*Provided the variation in population did not exceed a few per cent the numbers could be "standardized" with advantage; although the numbers might be non-integral the slight bias in the estimates of variance would be trivial compared to the other biases eliminated.

TABLE III

SYMBOLISM USED IN ANALYSIS OF TREND

p	q	n			s*	ps	qs	ns^2
p_1	q_1	n_1	n_1	n_1	s_1	s_1p_1	s_1q_1	$n_1s_1^2$
p_2	q_2	n_2	n_1+n_2	$2n_1+n_2$	s_2	s_2p_2	s_2q_2	$n_2s_2^2$
p_3	q_3	n_3	n_2+n_3	$2n_1+2n_2+n_3$	s_3	s_3p_3	s_3q_3	$n_3s_3^2$
p_4	q_4	n_4	n_3+n_4	$2n_1+2n_2+2n_3+n_4$	s_4	s_4p_4	s_4q_4	$n_4s_4^2$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
P	Q	N				$\sum ps$	$\sum qs$	$\sum ns^2$

$$d_p = \frac{1}{P}\sum ps - \frac{1}{2} \qquad \chi^2_1 = \frac{d_p^2}{V_p} + \frac{d_q^2}{V_q}$$

$$d_q = \frac{1}{Q}\sum qs - \frac{1}{2} \qquad \Delta_p = 12d_p\left(\frac{P}{N}\right) = \Delta_q = 12d_q\left(\frac{Q}{N}\right)$$

$$V_p = \frac{1}{P}\left[\frac{1}{N}\sum ns^2 - \frac{1}{4}\right] \qquad \{P_k\} = \{P_o\} + s_k\,\Delta p$$

$$V_q = \frac{1}{Q}\left[\frac{1}{N}\sum ns^2 - \frac{1}{4}\right] \qquad \{P_o\} = \frac{P}{N} - 6d_p\frac{P}{N}$$

*Note: $s_1 = \frac{n_1}{2N}$, $s_2 = \frac{2n_1+n_2}{2N}$, etc.

TABLE IV

EXAMPLE OF COMPUTATION OF ANALYSIS OF TREND

Work of Mother	Mal-formations	Con-trols	Sum	See Table III					
	p	q	n			s	ps	qs	ns^2
No work	8	4	12	12	12	0·0102	0·0816	0·0408	0·0013
Own housework with help	27	15	42	54	66	0·0563	1·5201	0·8445	0·1331
Own housework without help	95	104	199	241	307	0·2619	24·8805	27·2376	13·6498
Remunerated employment only ..	59	17	76	275	582	0·4966	29·2994	8·4422	18·7424
Remunerated employment and own housework with help	37	27	64	140	722	0·6160	22·7920	16·6320	24·2854
Remunerated employment and own housework without help	108	85	193	257	979	0·8353	90·2124	71·0005	134·6611
Total	334	252	586				168·79	124·20	191·47
	P	Q	N				$\sum ps$	$\sum qs$	$\sum ns^2$

$$d_p = \frac{1}{P}\sum ps - \frac{1}{2} = +0{\cdot}00536 \qquad \chi^2 = \frac{d_p^2}{V_p} + \frac{d_q^2}{V_q} = 0{\cdot}293$$

$$d_q = \frac{1}{Q}\sum qs - \frac{1}{2} = -0{\cdot}007215 \qquad \Delta_p = 12d_p\left(\frac{P}{N}\right) = +0{\cdot}037$$

$$V_p = \frac{1}{P}\left[\frac{1}{N}\sum ns^2 - \frac{1}{4}\right] = 0{\cdot}000230 \qquad \Delta_q = 12d_q\left(\frac{Q}{N}\right) = -0{\cdot}037$$

$$V_q = \frac{1}{Q}\left[\frac{1}{N}\sum ns^2 - \frac{1}{4}\right] = 0{\cdot}000304 \qquad \{P_o\} = \frac{P}{N} - 6\,d_p\,\frac{P}{N} = 0{\cdot}552$$

Consider a series $p_1, p_2, p_3 \ldots p_n$, of total P. Then the sum of the scored frequencies will be

$$\frac{1}{2P}(1p_1+3p_2+5p_3+\ldots\ldots(2n-1)\,p_n),$$

with an expectation of ½ and sampling variance of

$$\frac{1}{P}\left(\frac{1}{4n^3}(1^2+3^2+5^2+\ldots(2n-1)^2)-\frac{1}{4}\right),$$

which varies between $\frac{1}{16P}$ when $n=2$ and $\frac{1}{12P}$ when $n=\infty$.

Some values are given in Table V. Algebraically these results are identical with those of Armitage (1955), Formula 8.

TABLE V

VARIANCE OF THE MEAN AND RELATIVE EFFICIENCY OF VARIOUS PARTITIONS OF A RECTANGULAR UNIVERSE

n	Variance of Mean	Efficiency in Equally Divided Rectangular Distribution (per cent.)
2	$\frac{1}{16P}$	75·0
3	$\frac{1}{13 \cdot 500P}$	88·9
4	$\frac{1}{12 \cdot 800P}$	93·8
5	$\frac{1}{12 \cdot 500P}$	96·0
6	$\frac{1}{12 \cdot 343P}$	97·2
7	$\frac{1}{12 \cdot 250P}$	98·0
8	$\frac{1}{12 \cdot 191P}$	98·4
9	$\frac{1}{12 \cdot 150P}$	98·8
10	$\frac{1}{12 \cdot 121P}$	99·0
∞	$\frac{1}{12P}$	100·0

Estimates of linear trend may be obtained directly from the discrepancies d_p and d_q.

If there is a linear trend varying between $\frac{P}{N} - \frac{1}{2}\triangle_p$ and $\frac{P}{N} + \frac{1}{2}\triangle_p$ between the extremes of the distribution, writing $\{p_o\}$ for the proportion at the margin bordered by the class frequencies p_1 and q_1, which we may specify by the score s_o, the proportion associated with the score s_k, say $\{p_k\}$, will be:

$$\{p_k\} = \{p_o\} + s_k\triangle_p, \text{ when } \{p_o\} = \frac{P}{N} - \frac{1}{2}\triangle_p,$$

and a consideration of the first moments of a trapezoid show that:

$$\triangle_p = 12\frac{P}{N}d_p \text{ and } \triangle_q = 12\frac{Q}{N}d_q.$$

Obviously $\triangle_p = \triangle_q$ and this provides a simple check of the arithmetic. Where heterogeneity is meaningful, the computation of expected frequencies from this linear trend will allow a test to be made along the lines of the conventional χ^2 test, and will have $n-2$ degrees of freedom.

Table VI shows a test of trend in frequencies applied to data on the number of men aged 30 to 34 years certified as dying from carcinoma of the lung in England and Wales. The mean increment per year as a percentage of the mean is, when small, approximately equal to the rate of change per year. For computational simplicity a twenty-fold multiple of the scores has been used.

TABLE VI

NUMBER OF DEATHS EACH YEAR DUE TO CARCINOMA OF LUNG AMONG MEN AGE 30–34

Year	p	$s \times 20$	$s \times p \times 20$
1946	50	1	50
1947	67	3	201
1948	69	5	345
1949	57	7	399
1950	59	9	531
1951	47	11	517
1952	56	13	728
1953	66	15	990
1954	65	17	1105
1955	54	19	1026
Total	590 P		5892 $20\sum ps$
Mean	59·0		0·4993

$$\frac{1}{P}\sum ps = 0 \cdot 4993$$

$$d_p = -0 \cdot 0007$$

$$V_p = \frac{1}{12 \cdot 12P} = 0 \cdot 000140$$

$$\chi^2 = \frac{d^2_p}{V_p} = 0 \cdot 004$$

$$\Delta_p = 12d_p \pm 12\sqrt{V_p}$$
$$= -0 \cdot 009 \pm 0 \cdot 14$$

Mean increment per year as percentage of mean
= $-0 \cdot 1$ per cent. $\pm 1 \cdot 4$ per cent.

It is of interest to observe the very limited advantages of increasing the subdivision beyond 4 or so. The relative efficiency of different numbers of equal divisions in a continuous, and, in principle, infinitely divisible rectangular distribution is given in Table V.

It cannot, however, be concluded from this that the use of scores based on a rectangular transformation allows justifiable simplification of the "t" test for data approximating to the normal distribution or to one of its transformations. The use of a 2×2 Table as a simple test for the difference of means of large normal populations cannot have a higher efficiency than $\frac{2}{\pi}$.

It would appear that the breaking up of two normal distributions into the form of an $n \times 2$ Table cannot have a higher efficiency than

$$\frac{2}{\pi} \times \frac{16}{12} \text{ or } 84 \cdot 9 \text{ per cent.}$$

Summary

The widely-used method of testing for trend by using χ^2 for several degrees of freedom is a less powerful method of detecting trend than the use of a 2×2 test after removing some of the partitions. A rule for deciding which partitions to remove so as to maximize the residual information is advanced.

The loss of the information provided by these subdivisions, though less serious than the loss involved with methods using several degrees of freedom, may be reduced by various well known scoring methods. A system of scoring which is simple to compute, fully objective, and which leads to an intelligible estimate in the case of a rectangular distribution with a linear change in proportion, is advanced.

References

Armitage, P. (1955). *Biometrics*, **11**, 375.
Berkson, J. (1940). *J. Amer. stat. Ass.*, **35**, 362.
—— (1957). *Biometrika*, **44**, 411.
Cochran, W. G. (1954). *Biometrics*, **10**, 417.
Edwards, J. H. (1957a). *Brit. J. prev. soc. Med.*, **11**, 10.
—— (1957b). *Ibid.*, **11**, 73.
McKeown, T., MacMahon, B., and Record, R. G. (1951). *Ann. Eugen. (Lond.)*, **16**, 249.
Record, R. G., and McKeown, T. (1949). *Brit. J. soc. Med.*, **3**, 183.
Yates, F. (1948). *Biometrika*, **35**, 176.