231

NOTE ON THE TEST OF SIGNIFICANCE FOR DIFFERENTIAL VIABILITY IN FREQUENCY DATA FROM A COMPLETE THREE-POINT TEST

R. A. FISHER

If, in a three-point linkage test, all four triply heterozygous genotypes are used in parallel, the data, after summing complementary genotypes from each cross, take the form of a Latin Square,

TABLE 1

Latin square for three-point backcrosses

A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A,

in which the columns may represent the modes of gamete formation (cross-over classes), the relative frequencies of which it is desired to estimate, the rows may be the four parallel tests, while the letters stand for the four pairs of complementary genotypes obtained in each test.

An overall χ^2 test with nine degrees of freedom is, of course, available to detect non-proportionality of the entries in the rows and columns, such as unequal viability of the pairs of genotypes might be expected to produce. We may, however, aim at a more specific test of the three degrees of freedom represented by possibly different viabilities of the four classes of offspring. A simple, though crude, test of this kind was suggested by the author to M. E. Wright, for the test involving sex, wavy and shaker in the sex chromosome of the house mouse. It is the object of the present note to set out an accurate χ^2 test for such a case.

Wright's data, including 453 mice bred, are shown in this form below :—

TABLE 2

Data for wavy, shaker and sex

	Recombination involving				
	None	Sex	wv_2	sh_2	Total
All in coupling . .	23	44	18	21	106
wv_2-sh_2 in coupling . .	36	55	17	12	120
sh_2-sex in coupling . .	39	46	19	23	127
wv_2-sex in coupling . .	29	40	15	16	100
Total . . .	127	185	69	72	453

Heredity, 3: 215-219, (1949).

From the marginal totals proportionate expectations are calculated for each cell, as in the table following :—

TABLE 3

Expectations for equal viability

29·7174	43·2892	16·1457	16·8477	106·0000
33·6424	49·0066	18·2781	19·0729	120·0000
35·6049	51·8653	19·3444	20·1854	127·0000
28·0353	40·8389	15·2318	15·8940	100·0000
127·0000	185·0000	69·0000	72·0000	453·0000

From these, comparing each frequency separately with its expectation we find

$$\chi^2 = 7{\cdot}8239$$

for nine degrees of freedom, giving no indication of disturbances due to unequal viability.

The simple test referred to above consists in comparing the total expectations for each pair of genotypes derived from this table, with the totals observed, thus :—

TABLE 4

Crude test for viability

Pair of genotypes	Expected	Observed	Deviation
A	113·9624	113	−0·9624
B	112·3488	118	+5·6512
C	111·6624	109	−2·6624
D	115·0264	113	−2·0264
Total	453·0000	453	0·0000

giving a value

$$\chi^2 = 0{\cdot}39157$$

for three degrees of freedom. This is exceptionally small, and leaves no ground for suspecting any viability disturbance. Its very smallness, however, arouses some suspicion that it may be theoretically invalid, and indeed on reflection it appears likely that some part of the chance variations in numbers of the different pairs of genotypes may have been compensated in adjusting the expectations from the totals of the rows and columns.

The exact treatment of the problem in which the frequencies are affected by arbitrary viability differences, and in which the expectations are therefore the products of three unknown factors, offers exceptional algebraic difficulty. An exact test of significance for disturbance of viability, which *a priori* is always to be suspected in such data, is, however, a real need, so that without elaborate algebraic discussion it may be useful to set out such a test, and to illustrate it on these data.

Representing the column totals by p, q, r, s and the row totals by a, b, c, d out of a grand total of N observations we may calculate the matrix

$$\frac{16}{N^2}\begin{Bmatrix} (a+b)(c+d)(p+q)(r+s) & (ad-bc)(ps-qr) & (ac-bd)(pr-qs) \\ (ad-bc)(ps-qr) & (a+c)(b+d)(p+r)(q+s) & (ab-cd)(pq-rs) \\ (ac-bd)(pr-qs) & (ab-cd)(pq-rs) & (a+d)(b+c)(p+s)(q+r) \end{Bmatrix}$$

having the numerical values :—

TABLE 5

Covariance matrix of viability scores

388·44844	2·89182	−1·14671
2·89182	444·41957	·06378
−1·14671	·06378	442·66618

This is the covariance matrix, subject to fixed marginal totals, of scores for parameters of differential viability derivable from the deviations in table 4, as follows :—

A+B−C−D	9·3776
A−B+C−D	−7·2496
A−B−C+D	−5·9776

Inverting the matrix to supply multipliers, and for convenience multiplying by 1000, we have :—

TABLE 6

Covariance matrix for estimating

2·5744887	−·0167530	·0066715
−·0167530	2·2502350	−·0003676
·0066715	−·0003676	2·2590560

The sums of the products of the scores with the corresponding rows of any column now give estimates of viability differences corresponding with the scores, and the sum of the products of scores and estimates supplies the χ^2 needed for testing significance

Estimate of viability deviation		Score
λ	·024224	9·3776
μ	−·016468	−7·2496
ν	−·013439	−5·9776

with

$$\chi^2 = 0{\cdot}42688$$

for three degrees of freedom.

Wright's conclusion that the viability disturbances are negligible is thus confirmed by a more exact analysis. The value of χ^2 is only a trifle greater than before.

The estimates obtainable by this method evidently supply material for a first correction of the column factors appropriate to small but significant differences in viability. I am not satisfied how this can best be done, and the present example, with no sign of genuine viability disturbance, is not suitable for exhibiting the methods that might be used.

As a first step, however, we may use as natural logarithms of the estimated viability factors, the expressions

A	$\lambda+\mu+\nu$	$-\cdot005683$
B	$\lambda-\mu-\nu$	$+\cdot054131$
C	$-\lambda+\mu-\nu$	$-\cdot027253$
D	$-\lambda-\mu+\nu$	$-\cdot021195$

with corresponding increments of the natural logarithms of the column factors

Column	
1	$-\lambda(a+b-c-d)-\mu(a-b+c-d)-\nu(a-b-c+d)\div N$
2	$-\lambda(a+b-c-d)+\mu(a-b+c-d)+\nu(a-b-c+d)\div N$
3	$\lambda(a+b-c-d)-\mu(a-b+c-d)+\nu(a-b-c+d)\div N$
4	$\lambda(a+b-c-d)+\mu(a-b+c-d)-\nu(a-b-c+d)\div N$

and for the row factors

Row	
1	$-\lambda(p+q-r-s)-\mu(p-q+r-s)-\nu(p-q-r+s)\div N$
2	$-\lambda(p+q-r-s)+\mu(p-q+r-s)+\nu(p-q-r+s)\div N$
3	$-\lambda(p+q-r-s)-\mu(p-q+r-s)+\nu(p-q-r+s)\div N$
4	$-\lambda(p+q-r-s)+\mu(p-q+r-s)-\nu(p-q-r+s)\div N$

The adjustments in the column factors vanish if the row totals a, b, c and d are all equal, the ideal condition in the absence of differential viability, and are small if they are nearly equal, as in the example. Numerically, for the four columns they are :—

TABLE 7

Logarithmic adjustments of column factors

1	2	3	4
$-\cdot000690$	$+\cdot000797$	$+\cdot001635$	$-\cdot001742$

Increasing the natural logarithms of the observed totals by these amounts, adding and dividing by the total, we have percentage estimates of the frequencies of the four modes of gamete formation, agreeing closely with those obtained without adjustment, *i.e.* :—

TABLE 8

Frequencies of modes of gamete formation

Recombination involving	Without adjustment	Adjusted for viability
None	28·0353	28·0130
Sex	40·8389	40·8671
wv_2	15·2318	15·2551
sh_2	15·8940	15·8647
	100·0000	99·9999

as must equally be the case with the recombination estimates :—

TABLE 9

Recombination fractions

	Without adjustment percentage	Adjusted for viability percentage
wv_2-sh_2	31·1258	31·1198
wv_2-sex	56·0707	56·1222
sh_2-sex	56·7329	56·7318

The only perceptible change is a slight rise of about one-twentieth of 1 per cent. in the estimated recombination fraction of wavy with sex. It is evident that the excess of this value above 50 per cent. is not to be explained by anything in the nature of viability differences among the genotypes.

Finally, that this method, while not providing a perfect fit, goes some way towards doing so may be seen by reconstructing the whole table of expectations, using the logarithmic adjustments explained above. We find then :—

TABLE 10

Adjusted expectations

29·1310	45·1176	15·5254	16·2439	106·0179
35·2822	48·4833	17·8195	18·4197	120·0047
34·9003	51·2239	19·4201	21·4410	126·9853
27·6801	40·1376	16·2532	15·9212	99·9921
126·9936	184·9624	69·0182	72·0258	453·0000

The aim of the viability adjustments has been to bring the expected totals for genotypes into agreement with the numbers observed. The extent to which this has been accomplished is shown below :—

TABLE 11

Frequencies of pairs of complementary genotypes

	Unadjusted expectation	Observed frequency	Adjusted expectation
A	113·9624	113	112·9556
B	112·3488	118	118·0940
C	111·6624	109	108·9830
D	115·0264	113	112·9674

The largest discrepancy has been reduced about sixty-fold, and this without introducing any crying discrepancy in the totals for rows and columns. Generally speaking the viability corrections appear to have been slightly in excess of what was needed.

Obviously, in this example, the correction of viability disturbances, which are far from significant, is of no practical value. The method, however, seems to be serviceable, in the absence of a more exact approach, in cases in which small though real viability differences are suspected.