MULTIPLICATIVE EFFECTS AND THE LOGARITHMIC TRANSFORMATION IN THE ANALYSIS OF BALANCED MULTI-POINT LINKAGE TESTS

WALTER F. BODMER

Department of Genetics, University of Cambridge

1. INTRODUCTION

IN a previous paper Bodmer and Parsons (1959) have outlined a comprehensive analysis of χ^2 for data from balanced multi-point linkage tests. By drawing an analogy with factorial experimentation, it was possible to obtain components measuring genic viability effects and interactions, recombination effects, parental heterozygote effects and possible interactions between these three sources of variation. These components were obtained from a corresponding analysis of variance on the assumption that, at least to a good approximation, the various effects were additive, whereas the usual treatment assumes a multiplicative set of effects (Fisher, 1949; Parsons, 1957). If these are large, departure from additivity may be severe and should be detectable in a difference between the interactions as measured on an additive or a multiplicative scale.

If we take logarithms, we can turn a multiplicative system into an additive system involving the logarithms of the original effects. It will be shown that a similar analysis to that given by Bodmer and Parsons, but based on the logarithms of the observed data provides, to a good approximation, measures of the various effects and interactions on a multiplicative scale. This makes it possible to detect differences in the mode of gene action, at least on viability, from data obtained in a balanced multi-point linkage test.

2. THE LOGARITHMIC TRANSFORMATION

The expectation of the observed frequency in a balanced three-point test of the kth genotype from the ith parental heterozygote and the jth mode of gamete formation may be taken as $a_i\, p_j\, v_k\, e_{ijk}$ (see table 1, Bodmer and Parsons (1959)). The parental heterozygote, recombination and genotypic effects are a_i, p_j and v_k respectively and e_{ijk} is the interaction between these effects. The logarithm of the expectation is therefore the sum of the logarithms of the multiplicative effects. In order to be able to interpret an analysis of variance and so of χ^2 on the logarithms of the observed data, the logarithms of these parameters must satisfy restrictions of the form considered by Bodmer and Parsons for the additive effects.

Consider the transformation of the proportions a_i given by

$$b_i = a_i \mid [\pi a_i]^{\frac{1}{i}}.$$

It is clear that $\pi b_i = 1$ and so $\underset{i}{\Sigma} \log b_i = 0$. Furthermore, when all the a_i are equal $b_i = 1$ and so $\log b_i = 0$. Hence, if we transform p_j to a_j and v_k to w_k by similar transformations, the sums of the logarithms of these new parameters are zero. Since e_{ijk} are not proportions which must add up to unity, the form of restriction imposed on them is immaterial. The restrictions appropriate for the multiplicative model are

$$\pi e_{ijk} = 1, \quad \pi e_{ijk} = 1, \quad \pi e_{ijk} = 1, \quad \pi e_{ijk} = 1.$$
$$\text{given } i \qquad \text{given } j \qquad \text{given } k \qquad \text{all } i,j,k$$

The condition that the total observed should equal the total expected, is replaced by the condition that the sum of the logarithms of the observed frequencies should equal the sum of the logarithms of the expected frequencies. It may be noted that these conditions are analogous to calculating a S.S. in an analysis of variance, given the mean of the observations, and provide ancillary statistics indicating a measure of the amount of information contained in the sample.

The component S.S. for the analysis of variance on the logarithmic data will be of the form

$$\left[\sum_{r=1}^{n} k_{ir} \log a_r \right]^2 \Big| \sum_{r=1}^{n} k_{ir}^2$$

where $a_1 \ldots a_n$ are the observed frequencies and $\sum_{r=1}^{n} k_{ir} \log a_r$ is approximately the score for the ith multiplicative effect. The corresponding information is given approximately by the variance of the score (see e.g. Fisher (1956)), i.e. by

$$v\left[\sum_{r=1}^{n} k_{ir} \log a_r \right] \simeq \sum_{r=1}^{n} k_{ir}^2 \, v[\log a_r] \simeq \sum_{r=1}^{n} k_{ir}^2 \frac{1}{m^2} v(a_r) \simeq \frac{1}{m} \sum_{r=1}^{n} k_{ir}^2$$

where scoring is at equal expectations of the observed classes so that $v(a_r) \simeq m$ the overall mean. Hence if we multiply the components in the analysis of variance by m, or $m \, (\log_e 10)^2$ if the calculations were done on logarithms to the base ten, we obtain the corresponding χ^2s. The score for any effect is unaltered by variation in the other effects, and the information may be expected to vary only by a small amount. This analysis should therefore provide reliable measures for multiplicative effects of any magnitude.

3. MULTIPLICATIVE EFFECTS AND INTERACTIONS

It is possible to define effects and interactions for a multiplicative system which are analogous to those defined for an additive system. Thus, in terms of logarithms, the three-factor interaction, for example, is

$$\tfrac{1}{4}[\log abc + \log a + \log b + \log c - \log 1 - \log bc - \log ac - \log ab]$$

$$= \log 4\sqrt{\frac{abc \,.\, a \,.\, b \,.\, c}{1 \,.\, bc \,.\, ac \,.\, ab}}.$$

Hence we may define $4\sqrt{\dfrac{abc \,.\, a_{\cdot} \,.\, b \,.\, c}{1 \,.\, bc \,.\, ac \,.\, ab}}$ as a consistent measure of three-factor interaction on a multiplicative scale. The principles of confounding may also be applied, the ratios of the sets of confounded treatments being indistinguishable from the block effects instead of the differences.

It may be noted that for a multiplicative system estimation by geometric means replaces estimation by arithmetic means. This implies, for example, that in the model for a three-point test considered above, when all the interactions $e_{ijk} = 1$, consistent estimates of the various effects can be obtained by equating observed to expected. The estimates will be simple functions of the observed frequencies, whose variances can be calculated by Fisher's approximate formula (Fisher, 1925-54), and may be expected to be fairly efficient.

When performing an analysis on the logarithms of the observed data it is the multiplicative effects and interactions, as considered above, which are being tested. There is thus a duality between additive and multiplicative systems, the experimental techniques for the one being applicable to the other.

4. NUMERICAL ILLUSTRATIONS OF THE ANALYSIS OF χ^2 WITH THE LOGARITHMIC TRANSFORMATION

In order to illustrate the method of analysis and the possibility of detecting differences between additive and multiplicative gene action an example is given of the analysis applied to an artificially constructed set of data for a two-point experiment.

The form of the analysis of variance for a two-point experiment can be obtained exactly as that for a three-point experiment was obtained by Bodmer and Parsons (1959). The data consist of observed frequencies from four pairs of complementary genotypes, which can be arranged in a 2×2 Latin square. From these are obtained the " sums " and " differences " squares, giving the scheme for the analysis of a two-point test shown in table 1, where a and b are the two factors concerned.

A recombination fraction of 20 per cent. was assumed and the relative contributions of coupling and repulsion heterozygotes were taken to be 9 : 8. Viabilities were assumed which showed a competition effect, and the expected proportions for each class calculated as the product of the relevant parameters. The set of data was obtained by multiplying the expected proportions by a convenient observed total. The assumed values of the parameters and resulting set of data are given in table 2.

The competition effect is such as might be expected with additive viabilities and one genotype considerably deficient in the coupling phase when there is severe competition from the non-recombinant normals. The viability of the b genotype is considerably improved

in the repulsion phase when the normals are in the less frequent recombinant category, and that of the severely deficient genotype *ab*

TABLE 1

Scheme for the analysis of a two-point experiment

$\frac{1}{2}$ " sums " Latin square   d.f.

Parental heterozygotes (rows)	1
Recombination (columns)	1
Two-factor interaction (diagonals)	1
	3

$\frac{1}{2}$ " differences " Latin square

Viability × Parental heterozygote (rows)	1
Viability × Recombination (columns)	1
Main viability effect of *a* (diagonals)	1
	3

$\frac{1}{2}$ total of " differences " square

Main viability effect of *b*	1
	7

TABLE 2

Artificially constructed set of data for a two-point experiment

Viabilities in	Genotypes				
	1	*ab*	*a*	*b*	
Coupling . .	10	2	8	5	*25*
Repulsion . .	10	3	9	8	*30*

Recombination 20 per cent.
Representation of heterozygotes : Coupling : Repulsion = 9 : 8

	Non-Recombinants		Recombinants		Total
Coupling	1	432	*a*	87	
	ab	87	*b*	54	
Repulsion	*a*	288	1	81	
	b	255	*ab*	24	
					1308

is also improved a little. It should be noticed that both sets of viabilities for coupling and repulsion show considerable multiplicative interaction but only mild additive interaction.

The χ^2 analysis on the original and the logarithmic data is given in table 3. As expected, the two-factor interaction is significant only on the logarithmic data, showing multiplicative but not additive interaction of the two gene effects on viability. The analysis on the original data shows a highly significant viability × heterozygote interaction or competition effect, but indicates a viability × recombination interaction of the same magnitude. It should be pointed out that in the analysis

TABLE 3

Analysis of χ^2 for the two-point data of table 2

Latin square of sums	d.f.	χ^2s	
		Original data	Logarithmic data
Heterozygotes	1	0·11	0·95
Recombination . . .	1	509·1	666·6
Two-factor interaction . .	1	2·75	49·2
	3		
Viability × Heterozygote (competition effect) . .	1	63·4	11·2
Viability × Recombination . .	1	63·4	0·09
Main viability effect of a . .	1	86·3	100·7
	3		
Main viability effect of b . .	1	167·4	238·7
	7		

of a two-point experiment a competition effect may inflate the viability × recombination component, but that the reverse is unlikely to occur. It is interesting to note that on neither analysis is the slight departure from orthogonality detected. The exceedingly low value for the viability × recombination effect on the logarithmic analysis is because the actual expected values were used for the data and so no random variation has been allowed for. The known situation is seen to be clearly and accurately represented by the analysis on the logarithmic data.

The usual heterogeneity χ^2_1, calculated from the 2 × 2 Latin square of sums to detect viability interactions, is 5·7011. This considerably underestimates the significance of the multiplicative interaction, and it is to be expected, in general, that the accuracy of such a χ^2_1 in detecting multiplicative interactions lies between those obtained from an analysis of χ^2 on the original and the logarithmic data. The estimate of recombination obtained, using the product formula of estimation

L

(Fisher, 1925-54), is 0.1865 ± 0.01. Thus the effect of viability competition on the usual method of estimation is appreciable though the estimate obtained is not significantly different from the true value of 20 per cent. This effect may be expected to decrease as the number of loci used increases.

An analysis on the logarithms of Wallace's (1947) data for a balanced three-point test involving the factors *wavy-2* (*wv-2*), *shaker-2*

TABLE 4

Analysis of χ^2 on Wallace's (1947) data after taking logarithms

Latin square of sums	d.f.	χ^2	P in %
Parental heterozygote . . .	3	5·63	20 – 10
Recombination . . .	3	79·34	
Two-factor interactions . . .	3	0·98	70 – 50
Error	6	8·20	30 – 20
	15		
Latin square of differences			
Viability × parental heterozygotes .	3	2·36	50
Viability × recombination . . .	3	4·88	20 – 10
Main viability effects . . .	3	0·13	99 – 98
Error	6	16·11	2 – 1
	15		
Three-factor interaction . . .	1	2·30	20 – 10
	31		

(*sh-2*) and sex in the house mouse is given in table 4. The analysis is on the whole very similar to that given by Bodmer and Parsons (1959) on the original data. None of the interaction components show significant increases, showing that the effects are too small for a difference between additive and multiplicative scales to be detected. The only striking difference is the significant component for the error term from the differences square. This is, however, the largest of eight χ^2s, not including the component for recombination, but taking this into account the significance level is still less than 10 per cent. This supports the suspicion mentioned by Bodmer and Parsons (1959) that there exists some disturbance not taken into account by an analysis of the above form. A closer examination of the original data indicates a disturbance of the ratio of the complementary genotypes $+sh_2 Y$. and $wv_2 +X$. from the four parental heterozygotes, the other complementary pairs being unaffected. A χ^2_3 of 11·6 is obtained with a

probability of less than 1 per cent. This seems to confirm the above suspicion, as such an effect cannot be explained simply in terms of a competition effect or straightforward viability interaction. As the disturbance is only indicated by the differences square it cannot affect the estimation of recombination.

5. DISCUSSION

As with any χ^2 test, small numbers of observations render the test inaccurate. Moreover large disturbances can arise from the differences of the logarithms of small numbers so that an analysis on the logarithmic data is more sensitive to disturbances due to a small number of observations. In order to take this into account some form of weighted analysis, using the number of observations, or amount of information, as weights, would be required.

Clearly, for it to be possible in practice to detect differences between additive and multiplicative action of genes, the effects must be fairly large or otherwise the two modes of action are approximately equivalent. In general three- and higher-point experiments will prove very difficult to run when there are severe viability disturbances. Hence differences in the mode of gene action are most likely to be detected with two-point experiments where more severe disturbances can be tolerated. This is well illustrated by preliminary results from a two-point back-cross linkage experiment with the factors *pallid* and *fidget* in the house mouse, being carried out by the author. Both these factors cause severe reduction in viability, and the results so far obtained show clearly multiplicative, as opposed to additive, action of the genes concerned on viability. Similar results have been observed by Parsons (1959) with two-point experiments in *Drosophila melanogaster*.

Whether viability effects of genes are additive or multiplicative will in general depend on the detailed physiological and biochemical situation. For example, genes with some sort of dosage effect may be expected to act additively. On the whole it seems more reasonable to suppose that the viability effects of genes act in a multiplicative way, the effects on the Malthusian parameter, as defined by Fisher (1930), being therefore additive. However, there is, at present, little data which could justify such an assumption.

The transformation to logarithms employed above is an example of one of the many transformations that are used in statistical analysis for diverse reasons. In this situation it is in fact employed to turn the multiplicative system into an additive system to which the standard techniques of analysis can be applied. A distinct meaning can be attached to the transformation as used here and a similar process could be carried out to investigate the possibility of gene action being a mixture of the multiplicative and the additive as might, in general, be expected to occur. It would seem, however, that in general due consideration must be given to the meaning of a transformation before

results from an analysis of transformed data can be legitimately interpreted.

6. SUMMARY

1. The use of the logarithmic transformation in the analysis of a multiplicative system of effects is described.

2. Consistent measures of multiplicative effects and interactions are defined and the duality between multiplicative and additive systems is discussed.

3. The methods of Bodmer and Parsons (1959) are applied, in conjunction with the logarithmic transformation, to the analysis of data from backcross linkage experiments.

4. The possibility of detecting differences between multiplicative and additive gene action on viability in data from such experiments is discussed.

7. REFERENCES

BODMER, W. F., AND PARSONS, P. A. 1959. The analogy between factorial experimentation and balanced multi-point linkage tests. *Heredity, 13*, 145-156.

FISHER, R. A. 1925-54. *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh.

FISHER, R. A. 1930. *The Genetical Theory of Natural Selection.* University Press, Oxford.

FISHER, R. A. 1949. Note on the test of significance for differential viability in frequency data from a complete three-point test. *Heredity, 3*, 215-219.

FISHER, R. A. 1956. *Statistical Methods and Scientific Inference.* Oliver and Boyd, Edinburgh.

PARSONS, P. A. 1957. An effect of gene arrangement on the recombination fraction of *Drosophila melanogaster. Heredity, 11*, 117-127.

PARSONS, P. A. 1959. The dependence of genotypic viabilities on the proportion of co-existing genotypes in *Drosophila melanogaster. Heredity,* in press.

WRIGHT, M. E. (= WALLACE). 1947. Two sex linkages in the house mouse with unusual recombination values. *Heredity, 1*, 349-354.