THE ESTIMATION OF LINKAGE WITH INCOMPLETE PENETRANCE

ENRIQUE SÁNCHEZ-MONGE

Estación Experimental de Aula Dei, Zaragoza, Spain

Received 11.ix.51

PENETRANCE of a gene is measured as the proportion of the individuals carrying this gene which are phenotypically different from the individuals of the population carrying the allele.

The problem discussed in the following is that of the estimation of linkage between two genes, one of which has incomplete penetrance. Tests of significance of the estimates of linkage and penetrance are also given.

We will treat separately the case of backcross data and the case of double heterozygotes intercross or selfing data.

1. BACKCROSS DATA

The offspring of a backcross $AaBb \times aabb$, supposing linkage in coupling phase and p being the recombination fraction, falls in four phenotypic classes: AB, Ab, aB, ab, the expected frequencies being those given in table 1 under $u = 1$. The total of individuals is n .

Supposing now that a fraction u of the Aa genotype is classified phenotypically as A -, the expected frequencies will be those given in table 1 under $u \neq 1$.

TABLE 1

Phenotype	Expected frequencies $\times \frac{2}{n}$	
	$u = 1$	$u \neq 1$
AB	$1-p$	$(1-p)u$
Ab	p	pu
aB	p	$p + (1-p)(1-u)$
ab	$1-p$	$1-p + p(1-u)$

Writing $(1-p)u = x$ and $pu = y$, the relative frequencies of the phenotypes AB, Ab, aB, ab will be respectively $x, y, 1-x, 1-y$, and if a, b, c, d are the observed numbers ($a+b+c+d = n$), the likelihood of obtaining the observed family is represented by (see Fisher, 1921)

$$\frac{n!}{a!b!c!d!} x^a y^b (1-x)^c (1-y)^d \quad . \quad . \quad . \quad (1)$$

The method of maximum likelihood depends on the maximization of this expression with respect to x and y . Its differentiation being difficult we use its logarithm that will be made maximum by the same values of x and y .

Differentiating (1) logarithmically we obtain :

$$\frac{\delta L}{\delta x} = \frac{a}{x} - \frac{c}{1-x} \quad . \quad . \quad . \quad (2)$$

$$\frac{\delta L}{\delta y} = \frac{b}{y} - \frac{d}{1-y} \quad . \quad . \quad . \quad (3)$$

Equating (2) and (3) to zero, the values of x and y will be :

$$x = \frac{a}{a+c} \quad , \quad y = \frac{b}{b+d} \quad . \quad . \quad . \quad (4)$$

Therefore :

$$u = x+y = \frac{a}{a+c} + \frac{b}{b+d} \quad . \quad . \quad . \quad (5)$$

$$p = \frac{y}{u} = \frac{b(a+c)}{a(b+d)+b(a+c)} \quad . \quad . \quad . \quad (6)$$

If we wish to test the significance of the departure of u from unity, admitting the linkage, we have that the maximum likelihood equation for $u = 1$ gives $p = \frac{1}{n}(b+c)$ and the expectations will be :

$$\frac{1}{2}(a+b) \quad , \quad \frac{1}{2}(b+c) \quad , \quad \frac{1}{2}(b+c) \quad , \quad \frac{1}{2}(c+d)$$

which compared with a, b, c, d give :

$$\chi^2 = \frac{(a-d)^2}{a+d} + \frac{(b-c)^2}{b+d} \quad . \quad . \quad . \quad (7)$$

Of the three available degrees of freedom one has been used in the estimation of p , leaving two for testing the significance of $u \neq 1$.

Conversely, to detect the existence of linkage, admitting incomplete penetrance, the value of u for $p = \frac{1}{2}$ is $u = \frac{2}{n}(a+b)$, and the expectations will be :

$$\frac{1}{2}(a+b) \quad , \quad \frac{1}{2}(a+b) \quad , \quad \frac{1}{2}(c+d) \quad , \quad \frac{1}{2}(c+d)$$

which compared with a, b, c, d give :

$$\chi^2 = \frac{(a-b)^2}{a+b} + \frac{(c-d)^2}{c+d} \quad . \quad . \quad . \quad (8)$$

with two degrees of freedom, as one is lost in the estimation of u .

For the calculations of the variance of u and p , we can use the formula (see Fisher, 1946) :

$$V_u = \Sigma \left[m \left(\frac{\delta u}{\delta a} \right)^2 \right] - n \left(\frac{\delta u}{\delta n} \right)^2 \quad . \quad . \quad . \quad (9)$$

where m are the expectations corresponding to the observed values a, b, c, d .

Thus we derive :

$$V_u = \frac{n}{2} \left[\frac{ac}{(a+c)^4} + \frac{bd}{(b+d)^4} \right] \quad . \quad . \quad . \quad (10)$$

An easy way to calculate V_p is to write

$$z = L \frac{1-p}{p} = L \frac{a(b+d)}{b(a+c)}$$

and then to use (9) to evaluate V_z . We then have

$$V_z = \frac{n}{2} \left[\frac{c}{a(a+c)^2} + \frac{d}{b(b+d)^2} \right]$$

and V_p being (see Fisher, 1947)

$$V_p = V_z \frac{1}{\left(\frac{dz}{dp} \right)^2} \quad . \quad . \quad . \quad (11)$$

we have :

$$V_p = \frac{np^2(1-p)^2}{2} \left[\frac{c}{a(a+c)^2} + \frac{d}{b(b+d)^2} \right] \quad . \quad . \quad (12)$$

2. INTERCROSS DATA

In a cross of the type $AaBb \times AaBb$ with complete penetrance of both genes the frequencies of the four phenotypes in the offspring are those given in table 2 under $v = 1$; p and p' are the recombination fractions for male and female gametogenesis, giving $P = pp'$ for matings in repulsion and $P = (1-p)(1-p')$ for matings in coupling. But if penetrance of gene A is incomplete, so that a fraction v of the AA and Aa genotypes are classified phenotypically as aa , the frequencies will be those given in table 2 under $v \neq 1$.

TABLE 2

Phenotypes	Frequencies $\times \frac{1}{n}$	
	$v = 1$	$v \neq 1$
AB	$\frac{2+P}{4}$	$\frac{(2+P)v}{4}$
Ab	$\frac{1-P}{4}$	$\frac{(1-P)v}{4}$
aB	$\frac{1-P}{4}$	$\frac{1-P+(2+P)(1-v)}{4}$
ab	$\frac{P}{4}$	$\frac{P+(1-P)(1-v)}{4}$

Writing $(2+P)v = x$, and $(1-P)v = y$, the relative frequencies will be $x, y, 3-x, 1-y$, and if a, b, c, d , are the corresponding observed numbers ($a+b+c+d = n$), the logarithm of the likelihood expression will be :

$$L = K + aLx + bLy + cL(3-x) + dL(1-y)$$

Differentiating and equating to zero we obtain the maximum likelihood equations :

$$\frac{\partial L}{\partial x} = \frac{a}{x} - \frac{c}{3-x} = 0 \quad . \quad . \quad . \quad (13)$$

$$\frac{\partial L}{\partial y} = \frac{b}{y} - \frac{d}{1-y} = 0 \quad . \quad . \quad . \quad (14)$$

Then :

$$x = \frac{3a}{a+c} ; \quad y = \frac{b}{b+d}$$

Therefore :

$$v = \frac{1}{3}(x+y) = \frac{a}{a+c} + \frac{b}{3(b+d)} \quad . \quad . \quad . \quad (15)$$

$$P = 1 - \frac{y}{v} = \frac{ab+3ad-2bc}{4ab+3ad+bc} \quad . \quad . \quad . \quad (16)$$

The significance of the discrepancy between v and unity, admitting linkage, can be tested by a χ^2 obtained by comparing the actual numbers a, b, c, d with the frequencies obtained supposing $v = 1$. This χ^2 has two degrees of freedom as one is used for the calculation of P .

Conversely, in order to detect the existence of linkage, admitting $v \neq 1$, we make $P = \frac{1}{4}$, $\left(p = p' = \frac{1}{2}\right)$, and the value of v will be :

$$v = \frac{4(a+b)}{3n}$$

The expected numbers are then :

$$\frac{3}{4}(a+b) \quad , \quad \frac{1}{4}(a+b) \quad , \quad \frac{3}{4}(c+d) \quad , \quad \frac{1}{4}(c+d)$$

which compared with the observed numbers a, b, c, d , give :

$$\chi^2 = \frac{(3b-a)^2}{3(a+b)} + \frac{(3d-c)^2}{3(c+d)} \quad . \quad . \quad . \quad (17)$$

For the calculation of V_v we can use (9), and then we have :

$$V_v = \frac{n}{4} \left[\frac{3ac}{(a+c)^4} + \frac{bd}{9(b+d)^4} \right] \quad . \quad . \quad . \quad (18)$$

Now if we write

$$z = L \frac{1-P}{2+P} = L \frac{b(a+c)}{3a(b+d)}$$

we have :

$$V_z = \frac{n}{4} \left[\frac{3c}{a(a+c)^2} + \frac{d}{b(b+d)^2} \right] \quad . \quad . \quad . \quad (19)$$

and then

$$V_P = V_z \frac{1}{\left(\frac{dz}{dP}\right)^2} = \frac{n(1-P)^2(2+P)^2}{36} \left[\frac{3c}{a(a+c)^2} + \frac{d}{b(b+d)^2} \right] \quad . \quad (20)$$

If the recombination fraction is the same in both sexes, $p = p'$, we obtain p from $P = (1-p)^2$, and V_p from :

$$V_p = V_P \frac{1}{\left(\frac{dP}{dp}\right)^2} \quad ; \quad \frac{dP}{dp} = 2(p-1)$$

thus :

$$V_p = \frac{V_P}{4(p-1)^2} = \frac{n(2p-p^2)^2(3-2p+p^2)^2}{144(p-1)^2} \left[\frac{3c}{a(a+c)^2} + \frac{d}{b(b+d)^2} \right] \quad . \quad (21)$$

3. REFERENCES

- FISHER, R. A. 1921. On the mathematical foundations of theoretical statistics.
Phil. Trans. Roy. Soc. A., 122, 309-368.
 FISHER, R. A. 1946. *Statistical Methods for Research Workers*. 10th ed. London.
 FISHER, R. A. 1947. *The Design of Experiments*. 4th ed. London.