JOHN INNES HORTICULTURAL INSTITUTION,
MOSTYN ROAD, MERTON PARK,
LONDON, S.W. 19.

Dear Fisher

Here is a problem for you.

In a group of linkage experiments the cross-over values are p and q in the two sexes. We have the following data from different families:–

From back-crosses. $p = P$, $\sigma^{-2} = I_1$.

" " " $q = Q$, $\sigma^{-2} = I_2$.

" coupling F_2, $(1-p)(1-q) = C$, $\sigma^{-2} = I_3$.

" repulsion F_2, $pq = R$, $\sigma^{-2} = I_4$.

Problem. What are the likeliest values of p, q, with their standard error?

At one time I thought I had obtained the former. I never had this illusion about the latter. I proceeded as follows:–

The likelihood of a pair of values p, q, is proportional to:–

$$\text{Exp} - \tfrac{1}{2}\left[I_1(p-P)^2 + I_2(q-Q)^2 + I_3(pq-R)^2 + I_4\left[(1-p)(1-q)-C\right]^2\right].$$

I believe this to be false because as far as I can see p and q are correlated, the coefficient being a function of the I_r's. I minimized the expression in the bracket, and obtained a pair of ~~exp~~ equations for $x = p - P$ and $y = q - Q$. These lead to quintics for x or y. But if we remove small quantities of the second order they reduce to simple equations, and we have

$$p - P = \frac{I_3\left[I_2 Q + I_4 (1-P)(P-Q)\right](R-PQ) + I_4\left[I_2(1-Q) + I_3 P(P-Q)\right]\left[C-(1-P)(1-Q)\right]}{I_1I_2 + I_1I_3P^2 + I_2I_3Q^2 - I_1I_4(1-P)^2 - I_2I_4(1-Q)^2 - I_3I_4(P-Q)^2}$$

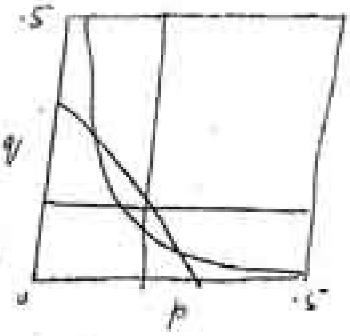
and a similar expression for Q.

In Primula sinensis I_3 and I_4 are much smaller than I_1 and I_2 e.g. for S and G.

$P = \cdot 3371$, $I_1 = 20,140$
$Q = \cdot 4038$, $I_2 = 20,040$
$C = \cdot 4578$ ~~·3217~~, $I_3 = 0,701$
$R = \cdot 1403$; $I_4 = 1764$

But in Pisum I_3 is generally larger than I_1 and I_2, I_4 often so.

My instinct is to make a graph as follows:—



The 4 lines represent the maximum likelihoods of (p, q) on the 4 sets of experimental data. I f one took the tangent to the 2 curves where they approach the straight lines one could (or indeed I think even I could) deal with the matter. But I would vastly sooner you did it. The problem will become urgent for us about June–July of this year. Do you feel you can deal with it? It may be trivial to you, but is not so to me. ~~M~~

Meanwhile, for our Primula data, I am assuming, rightly I hope, that p and q are only slightly correlated, and as I_3, I_4 are relatively small, I use the approximations.

$$p = P + \frac{I_3 Q(R-PQ) + I_4(1-Q)\left[C-(1-P)(1-Q)\right]}{I_1}$$

$$q = Q + \frac{I_3 P(R-PQ) + I_4(1-P)\left[C-(1-P)(1-Q)\right]}{I_2}.$$

as a guide to the sort of correction wanted. But it is a rough guide. And in spite of my appointment by Prof. Macbride (see Nature, Jan. 1931) as statistician to this institution, I am not satisfied with it. I hope you see the urgency of the problem, as I want to tidy up our Primula linkage data, and a good deal depends on the significance of certain statistics.

I think it would be an excellent thing to present your results about eugenics in a more popular form. I hope you will refer to the fact that Berlin, as well as Stockholm, has now got a definite differential fertility in favour of the rich. However I take it the Malthusian parameter for all classes is negative.

Yrs sincerely.

JBS Haldane

P.S. I am aware that in the example given $C^2+K^2>1$, hence the pair of p, q, values deduced from them are complex, i.e. the hyperbolae do not intersect. This may be due to miscalculation. Anyway, don't use those numerical data. We have a lot more this year, not yet added up.