23rd. May 1950.

My dear Rob,

Somewhere in your book you give the result of testing a few hundred people all with a large number of different reagents, summarised in a list of the number of genotypes which appear once, twice, three times and so on in these series. When you were here I forgot to take up the point with you, but you make the point that the usefulness of a blood group locus is measured in an inverse sense by the sum of the squares of the phenotypic frequencies. It is relevant that this quantity is directly estimated from such a record as

a_1 genotypes appearing once
a_2 " " twice
a_3 " " three times

out of

$$a_1 + 2a_2 + 3a_3 + \ldots\ldots = N$$ people tested.

For the expected value of

$$a_2 + 3a_3 + 6a_4 + 10a_5 + \ldots\ldots$$ divided by $\frac{1}{2}N(N-1)$

is just this sum of squares. I think I jotted the empirical values in the margin in these two cases, and, if I remember right, both came to less than 1%. I fancy the second one ought to have

been smaller, but was really larger, which suggests that the estimate, though convincingly direct and emperical, is not of very high precision.

Yours sincerely,