El-Nathan
August 7, 1937 Walcot, Happisburgh
Norfolk

Dear Deming,

I have your letter of July 27. It is curious that the point you raise is one of the first that attracted my attention to Statistics. Indeed, I wrote a juvenile paper upon it in 1912 or 1913. It has no particular merit for your purpose.

I notice that with the definition $s'^2 = \frac{1}{n} S(x-\bar{x})^2$

you call s' <u>the</u> S.D. of the sample. I think this is a very arbitrary nomenclature, since, whereas the standard deviation of the population sampled is unequivocally defined, a sample can provide innumerable different estimates of varying merit, and, without discussing the relative advantages of these, any one of these might equally be called the S.D. of the sample. When the theory of estimation was developed one of the points which I found most surprising is that bias in an estimate, at least bias of the order of 1/n, where n is the size of the sample, is of no practical importance, whereas its variance and, in the theory of small samples, the form of its distribution curve, is all important. The reason is that, in estimating the value of an unknown parameter θ, you are equally estimating the value of any given function of θ, and if one such estimate is chosen to be unbiassed, the others will generally be found to have positive or negative biases of order $1/n$.

Now, there is no denying the fact that, when there is more than one unknown parameter, maximising the likelihood provides simultaneous equations of estimation, and these are solved by taking as the estimated variance the statistic s'^2, as defined above. This is not the estimate I use, but it might be used, at some later algebraic inconvenience, to lead to identically the same tests of significance. In fact, in these tests of significance we deliberately make exact allowance for the sampling distribution of s' or s, and avoid the older practices of assuming the true standard deviation to be equal to s' or s; and since s' is a known function of s, to make exact allowance for the sampling distribution of one is to do so equally for the other.

My reason for using an unbiassed estimate, s^2, of the variance, apart from the fact that it simplifies the algebra, of testing the significance of the differences of two variances drawn from samples of different sizes, is that variances are things which one often wants to sum or to average. If this were equally true of standard deviations, or of invariances, it would be equally desirable to use unbiassed estimates of these; and for these, of course, s and $1/s^2$ are not unbiassed estimates. The indifference of bias is brought out most clearly by the property of sufficiency. The fact that the likelihood function involves, apart from a constant factor peculiar to the sample, only the statistics s^2 and $\bar{x}$, shows that these are jointly sufficient for the estimation of the mean and variance of the population. The relation is essentially a joint one. We cannot infer from it in general that $\bar{x}$ is a sufficient estimate for μ and s^2 for the variance. Indeed, as you note, if μ is known, the variance is properly estimated by $s'^2+(\bar{x}-\mu)^2$. This happens also to be a sufficient estimate. But if, instead of μ being known, there was known only some more general, functional relationship between the mean and the variance, the maximum likelihood estimates of these will generally involve both statistics, and not generally be sufficient. The joint sufficiency, however, of $\bar{x}$ and s^2 implies equally the joint sufficiency of any two independent functions of n, $\bar{x}$, and s^2, if such functions can be regarded as estimates at all.

In fact, I think the distinction you are drawing is one without an essential difference, one's choice of an unbiassed estimate being arbitrary in the sense that it is only justifiable by the use to which the estimate is intended to be put.

Yours sincerely,

R.A.FISHER.