January 12, 1939

Dear Jeffreys,

I always think that the best interests of mathematical notation are served by each man using the convention which suits him best. Consequently, I am not inclined to argue about using n or $n - m$. I don't quite follow what your letter says about least squares. However, since I always find it convenient to develop the theory in terms of undivided sums, e.g., in the coefficients and right-hand sides of the equations for regression coefficients, and in the formulae for obtaining the sums of squares of residuals. As I see it, the only question arises when one wants a mean square derived from this sum of squares, and then I am glad you agree that the old procedure of dividing by n - m is the one to use. Of course, there is no great harm in doing what the biometricians do, and dividing all sums of squares by n automatically at any other point in the work, for such divisors cancel out in estimating the regressions. In fact, I regard regression work from another point of view, as a good example of ancillary information, in that the precision of the regression does not really depend on the number in the sample, but only on the sum of squares of the independent variate, or, in general, on the dispersion

sums of squares and products of the set of independent variates. Since the actual values of these are provided by the sample, there is no need to consider estimation in respect of the variances and covariances of these independent variates. In fact the whole work is completely independent of how they may be distributed in the population sampled.

I showed your letter to Finney, who is now working in my Department, but he does not seem to have time to take up the computation you suggest.

Yours sincerely,