February 13, 1941

Dear Mr. Lyle,

I think your letter makes the position clearer by getting practical a little nearer to the actual needs of the case, which, in my experience, is always necessary before a theoretical problem can be properly stated.

The case as you state it is still not so near the real application as I should desire, but it is beginning to take shape. If you know (1) that the errors of determination of $g_1$ and $g_2$ are equal; (2) that these errors are independent and this is usually the easiest thing to be sure of; (3) that, apart from these errors, there would be some linear relationship between the observations, when the slope of this line is easily determined. From the available estimates of the two variances equal quantities are subtracted, as in Sheppard's correction; but in this case to a sufficient extent to make the remaining adjusted variances lie on either side of the covariance with the latter as the geometric mean of the two. In fact, knowing the form, but not the absolute magnitude, of the simultaneous distribution of the two errors, it is possible to resolve the observed distributions into that portion
ascorable to error and a remainder concerned with the true values.

Some questions do, however, still remain. You say you must be interested in being able to predict the probable value of $g_2$ at the end of a process having the initial value $g_1$.

Now, of course, if you were given the true value of the initial reading $g_1$, the corresponding value of the final reading $g_2$ could be read off the chart constructed as in the preceding paragraph; but do you expect to be given the true initial value $g_1$? If, on the other hand, you are only given $g_1$ as read, subject to error, will you not certainly be exaggerating the importance of any deviation this reading shows from the average by treating it as though it were a reading without error. In fact, if it is demonstrable that if $b$ is the regression coefficient of erroneous $g_2$ on erroneous $g_1$, then

$$g_2 = bg_1$$

has a smaller mean square than that of any similar formula, it would, therefore, be more accurate to use the ordinary regression for predicting from an erroneous reading than to use the true relationship between true readings.

Yours sincerely