21 February 1944

Dear Lyle,

I am a little puzzled by the algebra in your letter of February 15th and so have set out on a seperate sheet what I think is the basis of the point that puzzles you. I imagine that Method B(1) is that applicable to all or most of your problems.

Yours sincerely,

A. Large sample method

$$s_1^2 = \frac{1}{n_1} S(x_1 - \bar{x}_1)^2 \qquad \text{Sample of } (n_1 + 1) \text{ values}$$

$$s_2^2 = \frac{1}{n_2} S(x_2 - \bar{x}_2)^2 \qquad " \quad " \quad (n_2 + 1) \quad "$$

Estimated variance of difference of means

$$V(\bar{x}_1 - \bar{x}_2) = \frac{s_1^2}{n_1+1} \quad \frac{s_2^2}{n_2+1}$$

No allowance can here be made for the errors of estimation of s_1 and s_2.

B. Small sample method

(i) If the sources of error in the two samples are analogous, so that an observation is expected to have the precision, whichever sample it belongs to,

$$s^2 = \frac{1}{n_1 + n_2}\left\{S(x_1 - \bar{x}_1)^2 + S(x_2 - \bar{x}_2)^2\right\}$$

$$V(\bar{x}_1 - \bar{x}_2) = \frac{s^2}{n_1 + 1} + \frac{s^2}{n_2 + 1},$$

the same estimate based on $(n_1 + n_2)$ degrees of freedom being used for the precision of both means. Student's t table is then applicable.

In terms of s_1 and s_2 of A above this is rather complicated, i.e.

$$\left(\frac{1}{n_1 + 1} + \frac{1}{n_2 + 1}\right)\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$$

The larger sample contributes most of the information about precision, but the smaller sample contributes most of the error of the comparison. Consequently when n_1 is the larger, the part that matters comes from

$$\frac{n_1 s_1^2}{n_2 + 1}.$$

B. Small sample method (continued)

(ii) If the sources of error of the two samples are different, as when an astronomer derives estimates of some value, such as the distance of the Sun, by two totally different methods, the small sample problem is more complicated. We give some workable tables (V.1 and V.2) in the Second Edition of Statistical Tables, with a worked example in the Introduction explaining the logic of the thing.