October 6, 1942

Dear Jackson,

In your last letter you wrote about the question of the estimation of fly numbers from recapture figures, and especially about the precision of such estimates as can be made. This led me to try out on the material published for your Kakoma square, Table 2, p. 352-353 of your <u>Annals</u> article, a fairly systematic method, using as few assumptions as seemed possible.

Starting with the date ~~between~~ of your ~~first and~~ second collection ~~dates~~ of August 1938, I have tried making three different estimates of y_o, which is, I think, 10,000 divided by the number of flies in the square, including a fringe due to ambits as we have discussed before. The three estimates I make are:

$$x = \frac{.821 \times .824}{.370} \quad \text{or } 1.82839;$$

$$y = \frac{.501 + .248 + .178 + .114}{.301 + .157 + .146 + .106} \times .824, \text{ or } 1.20815; \text{ and}$$

$$z = \frac{.201 + .170 + .111 + .054}{.192 + .203 + .104 + .078} \times .821, \text{ or } 0.76266 .$$

These were calculated for the 73 dates which your table makes ~~poss~~ possible. The sum of squares for each of the variates, x, y and z, and the sum of products for each pair of these are found to be:

	x	y	z
x	7.98514	4.28193	4.18929
y		7.39631	2.85096
z			6.30300 .

Now, if x, y and z are all unbiassed estimates of the same quantity, all these sums of squares and products will contain as one component the true sum of squares of the quantity estimated. This is the component due to real variation in fly frequency. But y and z have been calculated from independent entries in your table; if we may take it that their errors are uncorrelated, or that the correlation between their sampling errors is zero, then the sums of squares and products of The errors of x, y and z can be estimated by subtracting from each entry 2.85096, so as to give the table

	x	y	z
x	5.13418	1.43097	1.33833
y		4.54535	0.
z			3.45204 .

Now, if x, y and z have these sums of squares and products due to error, one may ask what linear function $\lambda x + \mu y + \nu z = w$ shall have the least possible sampling variance when $\lambda + \mu + \nu = 1$. What comes of this is that

$$\lambda = .13453$$
$$\mu = .37202$$
$$\nu = .49345$$

and the sum of squares of the compound w is 1.88346, giving, on dividing by 70, .0269066 as the sampling variance of ~~a single weighted observation.~~ the weighted mean, the values of which I send in the enclosed table.

As you supposed in your letter, it seems to be inevitable that adjacent estimates shall be negatively correlated. I think this comes from the way in which the same entries appear first in the numerator and then in the denominator of corresponding fractions. For the 73 values given, the sum of the squared differences is 5.45744, which, on dividing by 144 gives .0378989, larger than the estimated variance of a single observation, which was .0269066, in the ratio 1.40854, showing that adjacent estimates have a negative correlation as high as .40854.

I do not know that this hurts anyone. It may, in fact, give some encouragement to the process of smoothing the ~~combined~~ adjacent values; for example, the formula with coefficients

$$-3 \; +12 \; +17 \; +12 \; -3 \; \div \; 35$$

when applied to uncorrelated series reduces the variance in the ratio $\frac{17}{35}$, or to just less than one half, but introduces a correlation $+\frac{48}{85}$ between successive terms. If it is applied to a series like the estimates w, having correlation .408535 between successive terms, I make the variance of a single smoothed estimate only .2815, and the correlation between successive smoothed terms positive, but lower than before, being about .33.

Notice that such a smoothing formula as this does not tend to even out smooth rises and hollows, i.e., when applied to such a series as 1, 6, 9, 10, 9, it reproduces the middle term, as a simple moving average does not. When you have time I should like to know how you feel about the series of estimates obtained in this way, in particular as to whether it ever seems to give less reliable results than those you have previously been using.

Sincerely yours,

1938			W
August	2nd	week	1.07176
	3rd	"	1.27375
	4th	"	0.92957
	5th	"	0.76921
Sept	1st	"	0.99473
	2nd	"	0.75468
	3rd	"	1.01431
	4th	"	1.20382
Oct.	1st	"	0.59156
	2nd	"	0.40798
	3rd	"	0.56765
	4th	"	0.97930
Nov.	1st	"	0.59341
	2nd	"	0.64442
	3rd	"	0.77007
	4th	"	1.17541
	5th	"	0.84085
Dec.	1st	"	0.57943
	2nd	"	0.99684
	3rd	"	0.52863
	4th	"	0.93497
1939			
Jan.	1st	"	0.91985
	2nd	"	0.47616
	3rd	"	0.52232
	4th	"	0.57335
	5th	"	0.67838
Feb.	1st	"	0.62968
	2nd	"	0.59102
	3rd	"	0.70439
	4th	"	0.64964
March	1st	"	0.62699
	2nd	"	0.59670
	3rd	"	0.69207
	4th	"	0.73064
April	1st	"	0.49921
	2nd	"	0.80124
	3rd	"	0.54729
	4th	"	0.41997

1939			W	
May	1st	week	0.42706	
	2nd	"	0.55825	
	3rd	"	0.44371	
	4th	"	0.33602	
	5th	"	0.45581	
June	1st	"	0.71848	
	2nd	"	0.96970	
	3rd	"	0.49760	
	4th	"	0.63982	
July	1st	"	0.55649	
	2nd	"	0.31431	
	3rd	"	0.49454	
	4th	"	0.38321	
Aug.	1st	"	0.76476	
	2nd	"	0.62015	0·62015
	3rd	"	0.34358	
	4th	"	0.35692	
	5th	"	0.99733	
Sept.	1st	"	0.66804	
	2nd	"	0.61387	
	3rd	"	0.88233	
	4th	"	0.52117	
Oct.	1st	"	0.61866	
	2nd	"	0.43455	
	3rd	"	0.41897	
	4th	"	0.75477	
	5th	"	0.34695	
Nov.	1st	"	0.74042	
	2nd	"	1.46927	
	3rd	"	1.04428	
	4th	"	1.18859	
Dec.	1st	"	0.82961	
	2nd	"	0.96244	
	3rd	"	1.13515	
	4th	"	0.99048	