Dr Jay L. Lush,
Texas Agric. Expt. Stat.,
Texas, U.S.A.

Dear Sir,

I was very much interested in your letter, which shows, if I may say so, just the rational and critical attitude in which statistical methods ought to be examined; I wish I could name half a dozen professed statisticians who would have seen the difficulty so clearly.

(i) For the individual samples you have a rigid mathematical relationship like

\[ \arccos \gamma_1 + \arccos \gamma_2 = \arccos \gamma_3 \]

where \( \arccos \gamma \) is the angle of which the cosine is \( \gamma \). This relationship will be retained in making estimates from combined samples, if we simply average the angles. This method, however, simply shirks the sampling problem, for the averaged angle is not the \( \arccos \) of the best estimate of \( \gamma \).

(ii) Your general problem, ignoring the general relationship, \( x + y = z \), is that of obtaining a simultaneous estimate of \( \gamma_1, \gamma_2, \gamma_3 \) from a number of sets of values \( \gamma_1, \gamma_2, \gamma_3 \), obtained from different samples. Any method of treating \( \gamma_1, \gamma_2, \gamma_3 \) separately and ignoring the values of the other two with which it occurs will be inexact. So the solution does not lie in the exact treatment of the set of values \( \gamma_1 \), to which the \( \gamma \)-method is an approximation, but in an examination of the simultaneous sampling distribution of \( \gamma_1, \gamma_2, \gamma_3 \).

Wishart (Biometrika, XX, p. 38) has given the simultaneous distribution of the 3 variances and 3 co-variances for 3 variates, but has not integrated out the variances to obtain that of the 3 correlations only. I doubt if the expression would be manageable if this were done.

(iii) In your special problem Wishart's expression degenerate, and a solution is quite simple, since

\[
\begin{align*}
S(xz) &= S(x^2) + S(xy) \\
S(yz) &= S(xy) + S(y^2) \\
S(z^2) &= S(x^2) + 2S(xy) + S(y^2)
\end{align*}
\]
the three correlations in the population must be determined by the ratio of the mean values of \( x^2, xy, y^2 \).

The three correlations to be estimated therefore have a simultaneous meaning if the different treatments not only do not disturb the ratio of \( \bar{x} \) to the geometric mean of \( \overline{x^2} \) and \( \overline{y^2} \), but if they also do not disturb the ratio of the latter two. If this ratio differs much in your different samples, the discrepancy is to be expected, if not take the sums for all samples

\[
\sum S(x^2), \sum S(xy), \sum S(y^2)
\]

as obviously the right method of estimating the ratios.

Of course \( \sum S(x^2) \) etc. can be built from these, and three consistent correlations obtained.

Yours sincerely,

[Signature]

C. A. F.