A NOTE ON FITTING OF STRAIGHT LINES IF BOTH VARIABLES ARE SUBJECT TO ERROR

By K. R. NAIR AND K. S. BANERJEE

The problem of fitting straight lines if both variables are subject to error, has been considered by many workers. But a recent study of this problem by A. Wald (1940) in which he also reviews all past work on the subject is mainly responsible for the present note. Wald has given a simple method of fitting a straight line to a set of n observed pairs of values (x_i, y_i) when both x and y are subject to mutually uncorrelated errors.

The method which Wald has advocated can be easily recognized as the Method of Averages for fitting curves. Nair and Shrivastava (1942) have recently studied the efficiency of this method, in comparison to the method of least squares, for fitting polynomials in x, when y alone is subject to error. They have found that a slightly modified method which they called the 'Method of Group Averages' will give more efficient results than the 'Method of Averages'. For fitting a straight line they found that by plotting the points of mean values of x and y for the first one-third and the last one-third of the whole set of observations, arranged in order of magnitude of x, and by joining these mean points we get a better estimate of the straight line than any other two group means.

By model sampling we have collected evidence that when both x and y are subject to error, then also the method of Group Averages gives better estimates for a and b, of the line $y = ax + b$, than the method of averages which Wald has put forward.

We started with the line $y = x + 1$, that is, with the hypothetical value 1, for a and b; and gave x integral values from 1 to 30 and calculated corresponding values of y, namely 2 to 31. On each pair of hypothetical values of x and y thus obtained we superimposed normally distributed random errors with mean zero and S.D. 1/10 using Mahalanobis' table of normal deviates (Mahalanobis and others, 1934). We then get an (x_i, y_i) sample of size 30, with x and y both subject to uncorrelated errors. Since the S.D. of the errors was one-tenth the interval between consecutive true values of x and y the observed and true values of both variables will have identical ranks in most cases. The values of a and b were calculated by the method of averages, that is, keeping 15 in each group and again by the method of maximal-group averages, that is, keeping 10 in each group. 100 Samples of 30 paired observations were obtained, giving 100 values of a and b, whose expectations are unity. The mean and S.D. of these 100 values of a and b were calculated and are presented below :—

Method of Fitting	a		b	
	Mean	S.D.	Mean	S.D.
Method of Averages (*Wald*)	·998583	0·0104	1·010063	0·0991
Method of Group Averages (*Nair and Shrivastava*)	·999703	0·0035	1·003634	0·0767

It is clear from the above table that while both the methods give consistent estimates for a and b the method of Group Averages gives more accurate estimates than the Method of Averages.

REFERENCES

1. P. C. MAHALANOBIS and others (1934): Tables of random samples from a normal population, *Sankhyā*, 1(2), 289-328.

2. NAIR, K. R. and SHRIVASTAVA, M. P. (1942): On a simple method of curve fitting, *Sankhya*, 6(2), 121-132.

3. Wald, A. (1940): The fitting of straight lines if both variables are subject to error, *Annals of Math. Stat.* 10, 284-300.

An interesting situation where both x and y are subject to correlated errors has been investigated by P. C. Mahalanobis in his paper "On errors of observation and upper air relationship": *Memoirs of the Indian Meteorological Department*, 24, 1923, 11-19.

Paper received : 28 December, 1942.