CORRESPONDENCE

This department is for the publication of informal communications that are of interest because they are informative and stimulating, and for the discussion of controversial matters.

Electron Microscopic Study of the Morphologic Index

TO THE EDITOR:

An article appeared in the INTERNATIONAL JOURNAL OF LEPROSY (**41** [1973] 1-6) entitled "Electron Microscopic Study of the Morphologic Index" by K. Sugiyama and S. Izumi, which calls for some comments from the statistical point of view. In fact, the authors have attempted, for a very useful tool, to measure the electron microscopic values of the Morphologic Index (MI). The methodology involved in the prediction of the (EM-MI) values from the (LM-MI) values is basically wrong despite its own quota of limitations.

In order to predict the (EM-MI) values from the (LM-MI) values it is necessary to keep the former one as the dependent variable and the latter, which is under control, as the independent variable (1). Such an equation can be obtained in a similar way as equation (A). The equation thus obtained is (EM-MI) % = 2.28 (LM-MI) % + 6.398 . . . . . (C) [or more correctly (EM-MI) % = 2.2835 (LM-MI) % + 6.3975]. The correlation coefficient r (=0.72) is responsible for the difference between equations (A) and (C). The equation (C) only can be used to predict the (EM-MI) values from the (LM-MI) values. The validity of equation (C) will be discussed in the following section.

TABLE 1. *Analysis of variance.*

Source	Degrees of freedom	Sum of squares (SS)	Mean of SS	F ratio
Total	29	8205 . 2104		
Regression	1	4282 . 5245	4282 . 5245	30 . 568
Residual	28	3922 . 7029	140 . 0965	
Lack of fit	22	3023 . 4841	137 . 4311	0 . 917
Pure error	6	899 . 2188	149 . 8698	

The basic mistake committed by the authors is the methodology itself in the prediction equation. The authors have chosen first the (LM-MI) values as the dependent variable and the (EM-MI) values as the independent variable and obtained a regression equation viz.:

(LM-MI) % = 0.23 (EM-MI) % + 1.84 . . . . . (A) and divided equation (A) by 0.23 and obtained an equation to predict the (EM-MI) values, the equation so obtained is (EM-MI) % = 4.38 (LM-MI) % - 8.08 . . . . . (B) which is not the correct way of obtaining the prediction equation. The equation (B) can be obtained from equation (A) if, and only if, the r, the correlation coefficient between (EM-MI) and (LM-MI), is either -1 or +1, which is not true with present data, but is only 0.72. Hence, the prediction equation (B) is not correct.

After obtaining any prediction equation the next thing to be done is to test its efficiency of prediction. As equation (B), the authors' prediction equation, is basically wrong it does not deserve any sort of tests for its accuracy. So attention will be focused only on equation (C). The initial test is to test whether the regression coefficient (i.e., 2.28) is significant or not. The following is the analysis of variance of the regression coefficient and the lack of fit.

From the above table it can be seen that the regression is significant ($p=<0.01$). This test does not give any guarantee regarding the accuracy of prediction. Also the

proportion of the variability explained by the regression equation (^{2})

$R^2 \left(=\frac{\text{Sum of squares due to regression}}{\text{Sum of squares of the total}}\right)$ is only 0.522 (i.e., 52%). So only 52% of the total variability is explained by the regression. The remaining 48% is still to be explained. This itself indicates that the above prediction equation is quite inadequate to explain the total variability. Finally the value of the X^2 to test the goodness of fit is 192 for 29 degrees of freedom, which is highly significant and reveals that the regression is not a good fit.

An attempt has also been made to find out the pure error from the repeated observations of the (LM-MI) values and hence the lack of fit. It has been found that the lack of fit is not significant (see analysis of variance table). But the important thing to be noted with the data is that about 79% of the total sum of squares of pure error occurred due to the repeated observations of (LM-MI) values of 4.25 (observation Nos. 12 and 13). By deleting these particular values it can be seen that the lack of fit is significant and still the X^2 is significant. So the lack of fit and pure errors are not sensitive enough to detect the adequacy of the regression equation so obtained due to very high pure errors.

In the light of the above three indices viz. the R^2, the pure errors and the X^2 value, one can easily see that regression equation (C) is not a good predictor of the (EM-MI) values from the (LM-MI) values.

—Mr. P. R. Raveendran Nair, M.Sc., D.P.S.
Statistician
Acworth Leprosy Hospital
Wadala, Bombay-400 031
India

REFERENCES

1. Berkson, J. Are there two regressions? JASA **45** (1950) 164-180.
2. Draper, N. R. and Smith, H. *Applied Regression Analysis*, New York: John Wiley and Sons, 1966, p 26.

Reply: *I am very much obliged to Mr. P. R. R. Nair for his kind comment to our paper entitled "Electron Microscopic Study of the Morphologic Index" (Int. J. Lepr.* **41** *[1973] 1-6). I have read his criticism very carefully and analyzed our data again. Since I did the statistical analysis in the paper, I would like to answer his criticism which is composed of two parts. In the first part he pointed out that our equation for predicting EM-MI value from LM-MI value is wrong. In the second part he tested the adequacy of the corrected equation (i.e., equation C) by using three statistical methods and concluded that this equation is quite inadequate to predict EM-MI from LM-MI.*

We totally agree with his pointing out in the first part of the critique, and with apology to the readers, we want to correct our equation to read: EM-MI(%) = LM-MI(%) × 2.28 + 6.39.

However, I cannot agree with Mr. Nair's conclusion that the prediction equation is quite inadequate for predicting EM-MI from LM-MI. The reasons are as follows:

1) The regression coefficient 2.28 is highly significant (more than 99%).

2) As pointed out by Mr. Nair, the percentage variation explained (R^2) is only 52%. Of course we are pleased if we get higher R^2 value, but 52% is still highly significant (more than 99%).

3) As one can know from Mr. Nair's analysis-of-variance table, lack of fit is not significant. This itself means that the equation is adequate for prediction. Based on his suggestion, I calculated the variance ratio of lack of fit and pure error after deleting cases 12 and 13. As the calculated variance ratio 3.518 is smaller than $F(21, 5; 0.05) = 4.55$, the lack of fit is still not significant. So, we have no reason to assert that there is a significant lack of fit in the "residual variation."

4) I also tried curvilinear regression $y = a + bX + cX^2$, but the regression coefficient of term X^2 is only 0.0047 and statistically not significant. This result shows the curvilinear regression is meaningless.

5) I do not think X^2-test is a suitable method to test the adequacy of our regression equation.

Thus, we have no doubt about the adequacy of the prediction equation from the statistical point of view. But the standard error of estimated EM-MI value by this equation is 11.8% and this is the limitation of our data. Because of the limited biological meanings of the Morphologic Index, and despite this statistical limitation, I believe that this prediction equation is still useful both in clinical and experimental researches of leprosy.

—Shinzo Izumi, M.D.
The JALMA Leprosy Center
Agra, Uttar Pradesh, India