DISCUSSION

George A. Schnack, National Center for Health Statistics

I found Mr. Cohen's paper particularly interesting because he wrote about a system that I developed for the National Center for Health Statistics about 15 years ago. The two-parameter model and fitting criterion came from the Bureau of the Census, from Tom Jabin, I believe. My task, as I recall, was to develop the normal equations and program them for use in the IBM System 360. The program was for the use of the staff of the Division of Health Interview Statistics (DHIS). The three-parameter model has been used less extensively by NCHS.

Mr. Cohen's paper is about how one can determine which of two curve-fitting criteria is better. That is, for different samples he used the same predicting model, the same input, but different fitting criteria to produce two curves. The fitting criterion for each is to minimize a weighted sum of squares. In one case the weight is the measure of the statistic and in the other a measure of the statistic's relative variance. The question is: Which is better?

The answer obviously rests on what is meant by "better." Better can be, according to Mr. Cohen, that criterion which yields the lower average absolute residuals. To NCHS, the criterion is that which yields the lower average squared relative residuals. We chose the latter criterion because we wanted variance curves to approach more closely those statistics having the greater precision. In the Health Interview Survey, for example, when we are plotting relative variance against aggregate value for a statistic, we want the curve to fit the larger aggregate values more closely than the smaller ones. Since the DHIS design is essentially self-weighting, we want a better curve estimate of variance for those statistics having larger sample sizes.

If the criterion used by NCHS does not fit the larger statistics more closely than the smaller ones, then the statistics are perhaps not from a self-weighting design; or possibly the input is not homogeneous. In my experience, the model and criterion used by NCHS are for situations in which the statistics are from self-weighting designs and the statistics are homogeneous with respect to variance.

I am pleased, however, that Mr. Cohen has investigated some of the properties of our model and fitting criterion. For too long we have assumed their worth and I hope that from his work we at NCHS will rethink our procedures. I am especially interested in looking into the variability of the estimated model parameters (a and b) found by Mr. Cohen.

As I said initially, the program discussed above was written more than 10 years ago for a very specific use: To generate variance curves for the DHIS publications. Over the years this program became somewhat "generalized," that is, more people began using it in different situations. We all know this is risky. But generalized software always--or usually--presents this problem. That is, often the user compromises his model to fit the software; or he doesn't fully understand the assumptions used in the software; or, worst of all, the software is faulty and gives wrong values. There is not much to say about the programs reviewed by Mr. Sedransk--what can one say about programs that are faulty? I do have a few comments, however, on generalized programs and their uses.

One must ask, when using generalized software, who is at fault when errors occur--the user or the developer of the program? Both obviously are. The user who is not well trained in mathematical statistics or sample survey techniques can easily use the wrong package; and the developers are at fault for poor documentation and computer-generated error messages. I think that the cases presented today illustrate both poor documentation and computer-produced error messages. I think that the cases presented today illustrate both poor documentation and computer-produced error messages. Surely it would not be asking too much of the developers to provide messages when user-supplied data are not compatible with the model. When we developed the DHIS curve-fitting program, we required the following features:

(1) With the proper input, the program would calculate the appropriate curve.

(2) The input would be checked for appropriate sizes, and messages would be printed for input values that were thought to be too small or too large. (For example, the input into the program consists of aggregate values and relative variances. If the relative variance is greater than 1, a message is printed next to the output asking the user to check this value. The program doesn't stop but simply prints the message at the time of output.

(3) Messages would be printed for input that are 3 or more standard errors from the curve. (Here we are checking for homogeneity.)

(4) And, lastly, the current date and input name and tape serial number would appear on all output tables.

I believe these elements should be incorporated in all generalized programming.