A Multivariate Analysis of Public Housing Residents: Comment

JOSEPH G. VAN MATRE*

In a recent issue of this journal, Kau and Floyd [1] employed the multivariate statistical technique of discriminant analysis in a study concerned with occupant and dwelling characteristics of public versus substandard housing in Savannah, Georgia. Their stated goal was:

> "to determine which of the poor have lived in public housing. In other words, given the limited supply of public housing, to what extent is the selection process of the public housing authority biased? The paper also compares the structural and environmental qualities of public housing with those of substandard housing."

The subsequently derived discriminant function correctly classified 86.8 per cent of the sample observations. K and F conclude that "the Public Housing Authority does discriminate and that in general the model has discriminatory power." However, an examination of their analysis reveals several methodological flaws which weaken, if not negate, their conclusion.

First, the use of the words "bias" and "discriminate" in conjunction with the P.H.A. is unfortunate. Both terms, according to the dictionary [4], connote prejudice. Even if the entire study had no other flaws, such a conclusion would be unwarranted; one must be very careful to distinguish between cause and effect. The establishment of differences between the two groups does not show cause. Further, the assumption that "the Public Housing Authority makes the selection of who shall live in public housing from the available people in substandard housing" weakens still more such a conclusion. Should not differences be established between those selected and the actual pool of applicants instead of the potential pool? Lastly, what, if any, are the P.H.A.'s stated criteria for selection: need, a FIFO queuing discipline, what? K and F state that "there is no theoretical model to explain how the Public Housing Authority in Savannah decides who . . . will be allowed to occupy . . . public housing units."

As concerns the authors' utilization of multivariate discriminant analysis, a few observations might be noted. First, data concerning nineteen independent variables were obtained on the 196 residents in the sample; the data were used to construct the discriminant function in order to determine which of the poor

*Associate Professor, School of Business, University of Alabama in Birmingham, Birmingham, Alabama 35294.

live in public housing. Unfortunately, six of the nineteen variables measure characteristics of the dwelling and not the occupant:

X_1 : No. of Rooms
X_2 : Rent
X_3 : Plumbing and Heating Problems
X_{10} : Pests Problems
X_{11} : Waste Disposal Problems
X_{12} : Fire Hazards

A priori, one expects such differences, but more importantly, their use in determining which of the poor live in public housing is inappropriate. One should not attempt to determine characteristics of the substandard vs. public housing dweller while mixing in characteristics of the dwelling. One observes in a comparison of means, four of the above six differ significantly at the .05 level (K & F, Table 1), and X_2 is ranked second in importance in the discriminant analysis (K & F, Table 2). Of course, how much these variables contributed toward the 86.8 percent correct classification would be speculation.

The inclusion of all nineteen variables in the discriminant function is also somewhat surprising. Very seldom does one encounter a variable list of this magnitude where all possess statistical significance. It appears that a stepwise technique should have been employed so that insignificant variables could be culled. Further, the inclusion of all variables magnifies the problem of multicollinearity making the interpretation of the individual coefficients of the discriminant function most difficult. To quote Morrison:

> High degrees of collinearity (high correlation) among the dependent variables should be avoided. The resulting discriminant coefficients will be unstable, and it will be more difficult to interpret the contribution of each independent variable. Hence, if two independent variables are highly correlated, e.g. $r = .95$, only one of these variables should be included in the analysis. Otherwise the variances of the b's (the discriminant coefficients) will be unnecessarily large. [3, p. 162]

Multicollinearity would obviously exist among such variables as:

X_7 : Income
X_{17} : Net Income
X_{18} : Per-Capita Income
X_{19} : Per-Capita Net Income

This problem is manifested in five sign reversals in the discriminant function: X_7, X_8, X_{14}, X_{15}, X_{16}. For example, in a comparison of means for X_7, substandard housing is the higher, yet the discriminant coefficient for X_7 indicates higher income is associated with public housing.

In the validation of the discriminant function, K & F employ the C_{max} method suggested by Morrison [3]. However, Morrison cautions "that our dis-

cussion on chance models applies to individuals not used in calculating the discriminant function. If the individuals were used in calculating it, then some upward adjustment must be made in C_{pro} or C_{max}." [3, p. 158] Perhaps, in this case, a better validation approach is the split sample method suggested by Frank, Massy, and Morrison. [2]

The actual size of the sample employed in the analysis is also not clear. On p. 43, "The data consist of a random sample of 149 substandard housing dwellers and 71 public housing dwellers." On p. 48, "A random sample of 125 substandard housing residents and 71 public housing residents was selected. . ." Finally, the classification table on p. 48 shows 121 substandard housing residents and 75 public housing residents comprise the sample. A possible explanation is that not all of the 149 substandard were usable, only 125, and that Table 3 has been erroneously transposed; a reconciliation of these differences would be most helpful.

One final note concerns variables X_4, X_5, and X_6 (children under 6, children six through eighteen, and total family size) for public housing. Since the means of X_4, X_5, and X_6 are 1.21, 2.06, and 4.04 respectively, this indicates that 0.77 is the mean number of household members that are not children eighteen or less. Now since household members must be an integer value, to realize a mean of 0.77 indicates that several observations were zero for number of household members not children under eighteen. Then the question naturally arises, is head of household counted in X_6? If so, either a misprint or measurement error is in existence. (Or possibly, if head of household were less than eighteen, he or she was counted as a child.)

In summary, K and F attempted a worthwhile endeavor but fell somewhere short of their goal. It would appear that the authors possess data to investigate a modification of the two separate questions posed in the first quotation: (1) the determination of what socio-economic variables differ significantly between the two groups of dwellers, and the development of a discriminant function for classifying a dweller, and (2) the determination of what structural and environmental variables differ significantly between substandard and public housing, and perhaps a discriminant function for the classification of a dwelling. Unfortunately, pooling the two into a single discriminant analysis obscures the answer to either question. One is impressed with the significant amount of data that the authors have accumulated and trusts subsequent analyses will yield valuable insights to the regional scientist.

REFERENCES

1. Kau, James B. and Floyd, Charles F. "A Multivariate Analysis of Public Housing Residents," *The Review of Regional Studies* Vol. 4, No. 1, Spring, 1974, pp. 42-49.
2. Frank, Ronald E., Massy, William F., and Morrison, Donald G. "Bias in Multiple Discriminant Analysis," *Journal of Marketing Research,* Vol. II, August, 1965, pp. 250-258.
3. Morrison, Donald G. "On the Interpretation of Discriminant Analysis," *Journal of Marketing Research,* Vol. VI, May, 1969, pp. 156-163.
4. _____, *The American Heritage Dictionary of the English Language,* Morris, William (ed.), New York: American Heritage Publishing Company, 1973.