A Comment on Dye and McManus' Use of Discriminant Function Analysis

by Bernard Grofman

In a recent article, Dye and McManus (1976) use discriminant function analysis to test the capacity of demographic and regional variables to predict city governmental structure: mayor versus manager government, partisan versus nonpartisan elections, and ward versus at-large council constituencies. Dye and McManus claim that "discriminant function analysis provides a better estimate of how well demographic and regional variables discriminate between cities with different structural forms, and a better understanding of which independent variables are the best estimators of each of these structural arrangements" (1976:257). I agree that the "test of the utility of any methodological tool is its ability to improve upon our understanding of substantive problems confronting political scientists" (Dye and McManus, 1976:256-266), and I do not wish to debate in the abstract the relative merits and suitabilities of techniques such as contingency tables, regression, discriminant function analysis, probit analysis, etc. (See Aldrich and Cnudde, 1975.) Rather, I wish to point out that their choices about categorization in using discriminant function analysis require clarification of their conclusions (1976:269, especially Conclusions 1-3). Moreover, they overstate the predictive power of their analysis because they fail to compare their results with those that could be reached by a simple all-or-none prediction based on marginal analysis of crosstabulation tables.

Dye and McManus use as their dependent variables three

_____

*Professor Joseph Tanenhaus provided invaluable comments on an earlier draft of this article. I would also like to thank Lorraine Nilsen at Stony Brook and the Word Processing Center at Irvine for transcribing my scribblings into finished copy. I take full responsibility for any errors of interpretation or data analysis that remain.

dichotomies: mayor versus manager, partisan versus nonpartisan, and ward versus at-large. For their dependent variables, with one exception, they use quantitative variables, e.g., size of city, percent nonwhite population, age of city, etc. The exception is the nominal variable REGION (categorized as Northeast, South, Midwest, and West), which they dichotomize by grouping together Northeast with South and Midwest with West. Thus, differences among the four regions not captured by their dichotomization are necessarily overlooked in their discriminant analysis. They provide no justification for this choice of dichotomization and I find no compelling argument for it, or for any other alternative dichotomization of region, in the literature on city politics.

The limitation of their dichotomization is apparent in looking at the relationship between form of government and region, shown in Table 1a (adapted from Dye and McManus, 1976:266, Table 4).[1] To predict best which regions will have mayors and which managers, we should predict mayors for the Northeast and Midwest and managers for the South and West. Doing so, we obtain 69 percent $\left(\frac{37 + 51 + 34 + 26}{216}\right)$ correct predictions, which may be compared with the 51 percent $\left(\frac{110}{216}\right)$ correct predictions we obtain by simply predicting that all cities will have mayors, or with the 51 percent $\left(\frac{111}{216}\right)$ correct predictions we obtain by dichotomizing region as Dye and McManus do. If, however, we look for what might be called east-west differences by grouping together South, Midwest and West vs. Northeast, and if we predict managers for the former regions and mayors in the Northeast, we obtain 62.5 percent $\left(\frac{37 + 51 + 21 + 26}{216}\right)$ correct predictions. Thus, dichotomizing region as Dye and McManus do <u>reduces</u> the power of regional differences to predict city government structure. This is, I believe, a central problem for their analysis. (See Tables 1b and 1c, adapted from Dye and McManus, 1976:267-268, Tables 5 and 6.)

From Table 1b, we see that the optimal prediction again occurs when we group South, Midwest and West on the one hand versus Northeast on the other—predicting partisan elections for the latter and nonpartisan elections for the former. Doing so results in 72 percent $\left(\frac{34 + 53 + 41 + 34}{224}\right)$ correct predictions, compared with the 63 percent $\left(\frac{142}{224}\right)$ correct predictions obtainable merely by predicting all cities to have

TABLE 1a

FORM OF GOVERNMENT BY REGION[a]

	Northeast	South	Midwest	West	Total
Mayor	37	27	34	12	110
Manager	8	51	21	26	106

TABLE 1b

TYPE OF ELECTION BY REGION[a]

	Northeast	South	Midwest	West	Total
Partisan	34	25	19	4	82
Nonpartisan	14	53	41	34	142

TABLE 1c

	Northeast	South	Midwest	West	Total
Ward	21	67	24	28	140
At-Large	17	9	19	9	54

[a]Source: adapted from data in Dye and McManus (1976:266-268, Tables 4, 5, and 6).

nonpartisan elections, and with the 60 percent $\left(\frac{134}{224}\right)$ correct predictions obtainable by dichotomizing region a la Dye and McManus.

From Table 1c it is easy to see that although significant differences in predilection for ward vs at-large elections exist across region, region is useless for improving

predictions of ward versus at-large elections, since the optimal prediction will predict cities in <u>all</u> four regions to have ward elections--achieving a predictive accuracy of 72 percent $\left(\frac{140}{994}\right)$. Predicting the Northeast and South to have ward elections and the Midwest and West to have at-large elections actually leads to a predictive accuracy of below 50 percent $\frac{78}{194} = 40.2\%$.

I do not claim that predictive power is the sole desideratum for judging an analysis (compare the views of Friedman, 1953, versus those of Koopmans, 1957), but merely assert that, in this case, in the absence of a theoretically based prediction as to exactly how region should affect form of city government, I see no reason not to make use of whichever breakdown of region maximizes predictive power.

For purposes of comparison, I have reproduced as Table 2a the contingency tables in Dye and McManus (1976:264, Table 3) that purport to give the predictive relationships between regions (Northeast and South vs. Midwest and West) and the three independent variables. In Table 2b I show the best predictions, based on the data in my Table 1 (derived in turn from the raw data in Dye and McManus, 1976, Tables 4-6). The figures on predictive accuracy of region as a lone predictor variable given by Dye and McManus (1976, Table 3), reproduced in Table 2a, misstate the case. Merely predicting all cities in all regions to be mayoral, nonpartisan, and ward would be more accurate than their predictions.

To provide base points, I show the predictive accuracy of such all-or-none predictions as Table 2c. These base point data are quite important. Dye and McManus, throughout their analysis, appear oblivious to the importance of unequal marginals. <u>To evaluate the power of a discriminant function analysis, we must compare it, at minimum, with the accuracy of the simplest of all predictions--an all-or-none prediction.</u>[2] Comparing Tables 2b and 2c, note that the apparent high predictive power (e.g., ward versus at-large) is quite illusory, since it offers no substantial improvement (indeed, no improvement at all) over the accuracy of an all-or-none prediction, which simply predicts all cases to be of the most common sort. Thus, in only two of the three cases Dye and McManus consider does knowledge of region improve the power to predict form of government--a point which they completely overlook.

Dye and McManus' claim that a discriminant function analysis involving twelve variables (population characteristics, socioeconomic composition, ethnicity and age and region

TABLE 2a

REGION AS LONE PREDICTOR OF CITY GOVERNMENT IN A DISCRIMINANT FUNCTION ANALYSIS: ACCORDING TO DYE AND MCMANUS (1976)

| | | Predicted | |
		Mayor	Manager
Actual	Mayor	64	46
	Manager	59	46
Correctly Predicted		51%	

| | | Predicted | |
		Partisan	Nonpartisan
	Partisan	59	23
	Nonpartisan	67	75
Correctly Predicted		60%	

| | | Predicted | |
		Ward	At-Large
	Ward	52	88
	At-Large	26	26
Correctly Predicted		40%	

TABLE 2b

REGION AS LONE PREDICTOR OF CITY GOVERNMENT IN A DISCRIMINANT FUNCTION ANALYSIS: ACCORDING TO GROFMAN

| | | Predicted[1] | |
		Mayor	Manager
Actual	Mayor	37	73
	Manager	8	98
Correctly Predicted		63%	

| | | Predicted | |
		Partisan	Nonpartisan
	Partisan	34	48
	Nonpartisan	14	128
Correctly Predicted		72%	

| | | Predicted | |
		Ward	At-Large
	Ward	140	0
	At-Large	54	0
Correctly Predicted		72%	

TABLE 2^c

ALL-OR-NONE PREDICTIONS OF CITY GOVERNMENT BASED ON UNEQUAL MARGINALS[1]

		Predicted	
		Mayor	Manager
Actual	Mayor	110	0
	Manager	106	0
Correctly Predicted		51%	

		Predicted	
		Partisan	Nonpartisan
	Partisan	0	82
	Nonpartisan	0	142
Correctly Predicted		63%	

		Predicted	
		Ward	At-Large
	Ward	140	0
	At-Large	54	0
Correctly Predicted		72%	

aSource: Dye and McManus (1975: 264, Table 3)

bSource: Adapted from data in Dye and McManus (1976: 266-268, Tables 4-6)

cSource: Adapted from data in Dye and McManus (1976: 264, Table 3)

[1] To increase comparability among the tables we have treated region as dichotomized as Northeast vs. South, Midwest and West. By grouping South and West versus Midwest and Northeast instead of South, West and Midwest versus Northeast, we could have improved our predictive accuracy to 69% (see discussion in text).

of cities) is "able to correctly classify only 67 percent of the cities by the form of government, 72 percent by type of election, and 71 percent by type of constituency...(1976: 269)," is quite misleading, since it is the dichotomy into which they force region that reduces its predictive power, and thus the overall predictive power, of their discriminant analysis. I have shown above that region alone can do as well or better than that—correctly classifying 69 percent of the cities by form of government, 72 percent by type of election and 72 percent by type of constituency.[3] Similarly, their assertions about the relative predictive power of region versus other variables in their data set (1976:169, Conclusions 2 and 3) are misleading in the same way.[4] These points are ones which are lost in relying too heavily on a particular analytic technique.

NOTES

1. The seeming inconsistencies between the data in my Table 1 and in Dye and McManus result from two typographical errors in the leftmost contingency table of their Table 2. The cells in that table, whose values are 27 and 44, have apparently inadvertently been interchanged.

2. This point is also made forcefully, and with a number of interesting examples, by Weisberg (1976), who also provides statistical measures of the improvement in accuracy of a model over that of baseline predictions based on marginals. I am indebted to an anonymous referee for calling this paper to my attention.

3. If region is confined to the Dye-McManus dichotomization, the 69 percent figure must, of course, be reduced to 63 percent.

4. However, some of their other conclusions (particularly Conclusions 4 and 5 on within-region differences (Dye and McManus, 1976:269) are, I believe, unaffected by the categorization scheme for Region they use.

REFERENCES

Aldrich, John and Charles F. Cnudde.
      1975   "Probing the Bounds of Conventional Wisdom:
              A Comparison of Regression Probit, and
              Discriminant Analysis." American Journal of
              Political Science 19:571-608.

Dye, Thomas, R. and Susan A. McManus.
      1976   "Predicting City Government Structure."
              American Journal of Political Science 20:
              257-271.

Friedman, Milton.
      1953   "The Methodology of Positive Economics." In
              Essays in Positive Economics. Chicago:
              University of Chicago Press:3-46.

Koopmans, Tjalling C.
      1957   "The Construction of Economic Knowledge."
              In Three Essays on the State of Economic
              Science. New York: McGraw-Hill:127-166.

Weisberg, Herbert C.
      1976   "The inherent predictability of Legislative
              Votes: The Perils of Successful Prediction."
              Delivered at the Annual meeting of the Midwest
              Political Science Association.