DOCUMENT RESUME

ED 047 674 JC 710 062

AUTHOR Wasik, John L.
TITLE The Development of a Mathematical Model to Project Enrollments in a Community College System.
PUB DATE [70]
NOTE 16p.; Paper presented at the 1971 Annual Meeting cf the American Educational Research Association, February 4-7, 1971, New York, New York

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Enrollment Projections, Enrollment Rate, *Enrollment Trends, *Junior Colleges, *Mathematical Models, Operations Research, *Statistics, Systems Analysis

ABSTRACT

Because of increased competition for funds, educational leaders are now recommending use of systematic planning techniques such as Planning-Programming-Budgeting (PPB). Effective implementation of such a procedure calls for a data base of enrollment of students in a program, retention rates, and costs per instructional hour. The intent of this report is to indicate how mathematical modeling procedures are used in the development of enrollment predictions in North Carolina community colleges. Three general methods used in enrollment projection are extrapolation, structural models, and Markov models. An evaluation of mathematical models for estimating enrollment in community college curricula is offered. Such mathematical models as CC Flow I and CC Flow II are discussed in detail and their four expected uses are presented. (CA)

The Development of a Mathematical Model To Project Enrollments in a Community College System[1]

John L. Wasik

North Carolina State University

INTRODUCTION

The rapid increase in enrollment in community colleges during the decade of the sixties has been cited as one of the real success stories of an educational concept. In view of wide public acceptance of the idea of a universal educational opportunity for all individuals regardless of race or creed, a continuing growth in two year college enrollments is almost guaranteed.

At the same time, the increasing competition for the tax dollar has led leaders at the national education level to recommend the use of systematic planning techniques in education. The Planning-Programming-Budgeting (PPB) system has been pointed out to the educational community as appropriate for use at state level as well as at the national in the setting of institutional goals. Research carried out in North Carolina and Florida has shown that costs per student instructional hour vary within the various instructional programs of a community college. In addition, informal surveys carried out

---

[1]This paper is based on research conducted part of the research program of the Center for Occupational Education located at North Carolina State University at Raleigh, Raleigh, North Carolina, in cooperation with the Division of Comprehensive and Vocational Education Research, National Center for Educational Research and Development, U. S. Office of Education. The research reported herein was performed pursuant to a grant with the Office of Education, U. S. Department of Health, Education, and Welfare.

in North Carolina also indicate little thought is given to planning for future enrollment growth in the various programs offered by the community colleges. These findings resulted in a recommendation to the North Carolina Community College System that a rational planning procedure such as PPB be implemented at the state (system-wide) and local (institution-wide) levels.

An effective implementation of a systematic planning procedure such as PPB at either the local or state level requires an accurate data base. For a community college, such data would include 1) enrollments of students by program, 2) retention rates, and 3) costs per instructional hour.

The intent of this report is to indicate how mathematical modeling procedures can be used in the development of future enrollment projections in North Carolina community colleges. The projections can be expected to provide direct estimates of the first two data requirements and can be used to indirectly estimate the third data requirement for a PPB system.

METHODS FOR ENROLLMENT PROJECTION

A review of the literature reveals several different approaches to the development of enrollment estimates. In attempting to categorize the several available enrollment projection studies, it became apparent that a three level classification scheme would adequately account for all models reviewed.

Extrapolation Methods

The first level of classification includes the various extrapolation procedures used to develop future enrollment projections. Two general methods

can be classified under the heading of extrapolation procedures. The first uses information obtained on survival of cohort groups to make straight-line extrapolation estimates. The second approach requires the construction of regression equations based upon cohort data to generate future enrollment estimates. Among the procedures previously reported, the simple linear regression equation based upon cohort information or the more complex multiple regression equation modified by lagged values of variables appear to be the most efficient procedures for predicting total enrollment. In an empirical comparison, Webster (1970) found the linear regression procedure based upon cohort information to provide more precise estimates a significantly greater number of times than did the survival cohort method when predicting elementary and secondary school enrollments in five Michigan school districts over a five year period of time. Studies conducted at the University of Minnesota indicated that a multiple regression equation utilizing numbers of Minnesota high school graduates estimated from a derived growth curve for the year of interest and the previous year plus estimates of the national military demands gave the best estimate of freshman enrollment at the University of Minnesota (Brown and Savage, 1960). Haggstrom (1969) found that a regression equation utilizing the estimates of members of high school graduates sufficed to predict total national female university enrollment; an expanded model including terms representing numbers of veterans from World War II and the number of draftees for the Korean and Vietnam hostilities provided the most precise estimates of total national university enrollment of male students.

Structional Model Methods

A second class of enrollment projection procedures have developed through the use of structural flow models. By the use of differential or

difference equations it is possible to model the flow of individuals through a particular educational system. By the application of these equations to the knowledge or estimates of individuals within the various system states, it is possible to derive aggregated estimates of total enrollments of that particular system. Structural models have been used to provide estimates of graduate degrees awards (Bolt, Koltan, and Levine, 1965; Hamond, 1968; and Reisman and Taft, 1969) and for elementary and secondary class sizes (Correa, 1966; and Durstire, 1969).

Markov Type Models

A third class of methods have utilized the theory of Markov chain processes to develop models for use in estimating future enrollment at different levels within the system. The basic idea underlying the application of this type of model is that knowledge of the probability of individuals with certain characteristics to move from one educational state to another state can be utilized to predict future movement of individuals within the same characteristics within the same system. Thus, if the distribution of the students within the educational system of interest among specific categories (e.g., curricula in a community college), F_t, is known for a specific time interval, then the transition proportions, P, may be employed to predict the distribution of students for the succeeding time period, F_{t+1}. Thus, if new entries are given for a sequence of time intervals, then the population in each of the educational system states for successive time intervals may be predicted by the iteration of the above described process.

Markov based models have been used to generate estimates of (1) total U. S. student and teacher populations (Zabrowski, 1968); (2) total number of

British students classified by level (Stone, 1965); (3) numbers of enrollments and degrees awarded in Australian universities (Gani, 1963); (4) numbers of students electing various curricula at the University of Minnesota (Brown and Savage, 1960) and (5) the number of personnel in various career fields and service grades in the U. S. Air Force (Merck, 1965). Three of the above reported studies (i.e., Zabrowski, Gani, and Brown and Savage) compared their results to available data and/or independent projections and found their estimates to provide fairly good fits.

One of the difficulties in determining the effectiveness of these models lies in the fact that there are no statistical tests available to provide a basis for determining the goodness of fit of estimates developed from the types of models presented above.

EVALUATION OF MATHEMATICAL MODELS FOR ESTIMATING ENROLLMENT IN COMMUNITY COLLEGE CURRICULA

One major consideration related to the development of any model is whether or not it can utilized by relatively untrained personnel. The major outcome of this project from which this paper developed is to be a set of procedures that can be used by personnel with the responsibility for developing budgeting and/or policy decisions at either the local, regional (i.e., county or groups of counties) and state levels. It is expected that data will need to be collected periodically to update the models in order to take into account the social and/or economic environmental changes resulting in college attendance and curriculum election tendencies. For this reason, it would appear that the Markov model which is based upon the use of a transition matrix of student movement in, out, and between curricula

offered by an institution is most appropriate. It is assumed that few administrators will have administrative assistants with the technical ability to use previous class enrollments to predict the future as in the case of regression procedures or solving for the new parameters required by structural flow models, but that relatively untrained personnel could follow procedures for developing estimates of transition probabilities. Thus, an administrator would set up data processing procedures that could match students at the beginning of two successive semesters or years in order to provide the empirical transition proportions required for the Markove type models. It is expected that with some training, the responsibility for updating the transition matrix can delegate to a staff member.

A second major consideration is related to the question of whether the estimation procedure requires a deterministic or probabilistic mathematical model. In other words, the model builder has to concern himself with whether or not the model equations will include chance or probabilistic components or not.

If the effect of any change in the system can be predicted with certainty, it is said to be deterministic. If not, then a chance component may be required to account for the discrepancies between the predictions made by the model and actual behavioral outcomes. When the model is concerned with a sequence of events where the outcome on each particular event depends upon some chance element, then the sequence is called a stochastic process. Thus, it can be seen that the regression model approach would be considered probabilistic, the structural flow model approach as deterministic, and the Markov chain approach as stochastic.

It would seem that the career plans of community college students are not completely fixed, thus leading to a situation where student progression within an institution cannot be predicted with complete accuracy. While it may be that all information relating the elements necessary for accurate prediction of a student's future educational or work plans cannot be obtained, it would seem that a stochastic model based upon a Markov chain type process would be the appropriate method to utilize in generating estimates of student enrollment in various curricula over time. The attempt to predict total institutional enrollment would seem to best be accomplished by use of a probabilistic model. Thus, regression models would appear to be appropriate for the estimation of a single enrollment as would be required for one institution or for the state for a particular year.

In commenting upon an earlier proposal to model student flow through a community college Wong (1969) noted that structural flow models tend to over simply the movements of student within the various levels of the educational system. He further pointed out that the use of continuous distributions of movements over time but independent with respect to time over a wide range of values are necessary for the development of a structural flow model. Since students "travel" in discrete through the present educational system, he concluded that the use of a Markov type process would provide a more accurate model of student movement within a community college system than would a structural flow model.

The theory of Markov chains assumes that there is a constant transition probabilities matrix for the population from which each of the observed empirical transition matrices were sampled under the constraint that they may be subject to random sampling errors. A second assumption is that the

probabilities of the various outcomes for a subject at any transition is based upon his status at the prior time period and not at all on his history more than one time period removed. Lohnes and Gribbons (1970) seriously questioned this one-state dependency assumption for the development of career aspiration.

It has been noted by Billingsley (1968) that no natural (e.g., social) process exactly satisfies the Markov chain condition, however, many will come close enough to make a Markov chain model useful. Further, as Creager (1970) comments, when the requirements for a Markov chain model are relaxed, greater flexibility and realism in reflecting the educational process accrue. He also notes that while this means the extensive mathematical development associable with the classical Markov model is no longer applicable, the multiplicative relationship between input and output distributions of students in the educational states still hold. That is, $F_{t+1} = F_t^T P$, can be still calculated since these are simple transition matrix calculations not dependent upon Markov chain assumptions. The case of $P_{13} = P_{12} P_{23}$ is an example of where several probability transition matrices can be reduced from multi-panel data to a single one-stage overall transition matrix.

In this case, a stationary probability matrix between stages does not need to be assumed. While many stochastic models do search for a stationary transition matrix as a "steady-state," it does not appear that this is a relevant question to ask of the movement of students through a postsecondary education system.

Models for Projecting Institutional Enrollments

The development of a model to generate estimates of total student enrollment in a community college can be approached in two ways. One way

would be to simply develop a regression model to estimate the total enrollment by institution. A second approach would involve the use of a Markove type model which would allow intermediate estimates of enrollment in curricula which, in turn, would be aggregated to develop institutional or system enrollment totals. Since studies have shown differential costs associated with per student instruction at the program level (trade and vocational program have been found to be more expensive than college parallel--C.f., Parry, 1969), it would appear that the usefulness of the procedure to an administrator would be substantially greater if individual program enrollment estimates are available. For this reason, it has been decided to develop a Markov type model as an integral part of the system for enrollment estimation under development.

Two models are proposed here for use in an enrollment projection system for community colleges. Both methods will require estimates of enrollments at the college level for each year into the future it is desired to have estimates for. Thus, it will be necessary to develop a regression model specifically for the purposes of estimating the numbers of students to be served by the community college.

Total Institutional Estimates

In attempting to develop an alternative procedure for estimating institutional enrollment totals, it will be necessary to analyze the factors that would be associated with enrollment in a community college. Information available on the characteristics of students enrolled in community colleges indicate that the majority immediately enrolled in college upon graduation from high school. As noted before, national military manpower needs as

manifested by the number of young men who are drafted or enlist will likely influence the estimates of individuals enrolled in a community college as will the level of economic activity in the selected unit of study. For if there is a large number of jobs available requiring minimal skills, it is possible that many young adults will forego further educational training--this would be an example of the thinking "I will go to school only if I cannot get a job or get drafted." Another economic influence on the numbers of individuals seeking additional education that may be of consequence is concerned with the extent of technological advancement and job skill obsolescence. As in the case of labor intensive industry, manufacturers may find it more profitable to move towards a less intensive labor orientation via automation or some similar procedure. Thus, the individuals who lose their jobs may be interested in returning to a community college to either upgrade their present skills or to learn a new trade.

It is suggested that a regression equation for developing estimates of community college enrollment (E) would be based upon information available on the following demographic variables:

H. Numbers of high school graduates by year

M. Required local draft board needs by year

A. Estimate of economic activity by year

C. County population

A model for predicting institutional enrollment based upon the above independent variables would have the following structural form:

$$E = f(H, M, A, C)$$

<u>The Mathematical Models for Estimating Enrollments by Curricula</u>

1. CC FLOW I

The following terms will first need to be defined.

Let X_t be a column vector of students in the ith curricula in year t.

Let P be the transition matrix describing the conditional probabilities that a student in curricula i at time t will be in curricula j at time t + 1.

Let C_t be a column vector of proportions of students in the various curricula averaged over a certain number of years; E_t is the total institutional enrollment for year t.

By utilizing the following matrix multiplication:

(1) $X_t = E_t C_t$

We obtain a distribution of individuals expected to be enrolled in each program at time t. By then carrying out the matrix equation corresponding to the classical Markov chain approach, we leave

(2) $F_{t+1} = X_t^T P$

which gives the expected distributions of students in the various curricula at the beginning of the next time interval.

It is also possible to think of P as a non-symmetric matrix--unlike the square matrix required by Markov chain theory. This is because the states of dropout and graduate from a community college cannot be observed to have individuals in these cells at time t, but can have at time t + 1. That is, you do not input dropouts or graduates, but you can output individuals falling into these classifications.

2. CC FLOW II

The second model will differ from the first in that the input each year will include estimates of the numbers newly enrolled students in an institution. This model is similar to the DYNAMOD II model developed by Zabrowski (1969) in that the number of newly enrolled students will be

superimposed on the total number of students expected back to the institution in succeeding years much like births were superimposed on the total numbers of children found in each age by the Zabrowski model.

Again we begin the specification of the model by defining the terms. E_t will be the total new enrollment by year t in an institution estimated by a regression model. Let F_t be a column vector of the number of students in the ith curricula in year t.

Let P be the transition matrix describing the conditional probabilities that a student in curricula i at time t will be in curricula j at time t + 1.

Let C_t be a column vector of proportions of new students who enroll in the curricula averaged over a certain number of years.

Let N_t be a column vector of the frequency of new students who enroll in various curricula.

Then we can say

(1) $N_t = E_t C_t$

We obtain a distribution of new students expected to be enrolled in each curricula at time t. By adding this value to what was obtained from the previous matrix multiplication of $F_t^T P$, we get a new estimate of total enrollment at time t + 1. Mathematically, this relationship is stated as follows:

(2) $F_{t+1} = F_t^T P + N_t$

Thus, the major differences in the two formulated models arise from the type of input estimates are to be obtained, total institutional enrollment or newly admitted students. Evaluating the two models for superiority, it would appear that CC FLOW I is a simpler model, conceptically speaking, however, with the availability of computers, there does not appear to be any real difference in computational efficiency.

CONCLUSION

The development of a transition matrix Markov type process model for developing estimating enrollment is expected to provide realistic estimates of future enrollments for the use of administrators concerned with efficient allocation of all resources. The exceptional growth of postsecondary education provided by community colleges/technical institutes over the last decade is expected to continue at least until 1975 when it is expected that over one half of all individuals receiving postsecondary education and/or occupational training will be provided by a two year institution.

Some of the expected uses of the output from the CC FLOW model are presented below:

(1) As noted before, the development of an institutional budget could utilize the product of estimated enrollments in a curriculum and the costs per student-instructional hour to give total costs of existing or proposed curricula or programs.

(2) Industrial development of a state is tied directly to the amount of trained manpower available. By the development of expected numbers of students electing various trade and technical programs and the resultant expected graduation rates, industrial development personnel could give prospective industries a rough estimate of the numbers of persons with various skills in a particular region.

(3) The proposed model can also be used to estimate the numbers of teachers required in the different curricula sponsored by an institution.

(4) The development of a model also presents the possibility of simulating student flow through a community college and/or a system.

Alternative educational policies can then be implemented and the simulated student flow can be studied to determine the specific effects of the various policy changes.

When one considers the increasing competition for tax dollars at the local and state level, it is quite apparent that not all two year or four year institutions will continue to receive support for improving and expanding educational services without a well developed plan documenting expected outcomes from the educational system.

For these reasons, it is expected that a project to develop two year institutional enrollment estimates by curriculum will be of substantial help to local and state community college administrators.

REFERENCES

Billingsley, P. Markov Chains, In David L. Sills (Ed.), International Encyclopedia of the Social Science, New York: MacMillan and Free Press, 1968, 14, 581-585.

Bolt, R. H., Koltan, W. L. and Levine, O. H. Doctoral feedback into higher education. Science, 1965, 148, 918-928.

Brown, B. W. and Savage, J. R. Methodological Studies in Educational Attendance Prediction. Department of Statistics, University of Minnesota, Minneapolis, Minnesota, 1960.

Cornea, H. Basis for the quantitative analysis of the educational system. Journal of Experimental Education, 1966, 35, 11-18.

Creager, J. A. Use of empirical transition matrices in educational research. Paper read at the meeting of the American Educational Research Association, Minneapolis, Minnesota, March, 1970.

Durstine, R. M. In quest of useful models for educational systems. Socio-Economic Planning Sciences, 1969, 2, 417-437.

Gani, J. Formulae for projecting enrollments and degrees awarded in universities. Journal of the Royal Statistical Society, Series A, 1963, 126, Part 3, 400-409.

Haggstrom, G. W. On analyzing and predicting enrollments and costs in higher education. Paper read at the meeting of the American Statistical Association, New York, August, 1969.

Hammond, A. A Flow Model for Higher Education, Rand Corporation, Santa Monica, California, August, 1968.

Lohnes, P. R. and Gribbons, W. D. The Markov Chain as a null hypothesis in a development survey. Journal of Educational Measurement, 1970, 7, 25-32.

Merck, J. W. A Markovian model for projecting movements of personnel through a system. Technical Documentary Report PRL-TDR-62-17. 6570th. Personnel Research Laboratory, Aerospace Medical Division, Lackland Air Force Base, Texas, March, 1965.

Parry, E. P. Research and development study of state funding and business management procedures in community colleges and technical institutes in the state of North Carolina. An investigation of cost differentials between trade, technical, and college parallel curriculums. Unpublished doctoral dissertation, University of North Carolina, Chapel Hill, North Carolina, 1968.

Reisman, A. and Taft, M. I. The generation of doctorates and their feedback into higher education. Socio-Economic Planning Sciences, 1969, 2, 473-486.

Stone, R. A model of the educational system. Minerva, 1965, 3, 172-186.

Webster, W. J. The cohort-survival ratio method in the projection of school attendance. Journal of Experimental Education, 1920, 39, 89-96.

Wong, Y. Personnel Communication, 1969.

Zabrowski, E. K. The dynamod model of student and teacher population growth. Socio-Economic Planning Science, 1969, 2, 455-464.