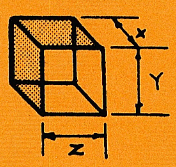
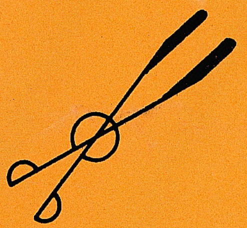
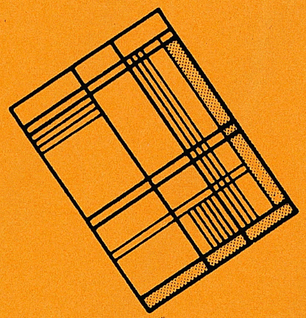
Geraldine D. Nelson

A TECHNICAL REPORT ON ENROLMENT FORECASTS OF REGISTERED INDIAN STUDENTS IN CANADA

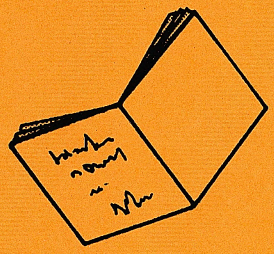
1975-76 — 1979-80

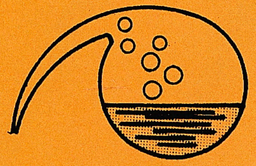






12 x 2 - 7 = 17





DEPARTMENTAL STATISTICS DIVISION
ARTMENT OF INDIAN AFFAIRS AND NORTHERN DEVELOPMENT

A TECHNICAL REPORT ON ENROLMENT PROJECTIONS OF REGISTERED INDIAN STUDENTS IN CANADA 1975-76 - 1979-80

D.E. Stewart

D.E. Stewart,
Statistician.

D.G. Saigaonkar,
Senior Statistician.

The views expressed in this report are those of the authors and do not necessarily reflect the official views of the Department of Indian Affairs and Northern Development.

DEPARTMENTAL STATISTICS DIVISION

DEPARTMENT OF INDIAN AFFAIRS & NORTHERN DEVELOPMENT

APRIL, 1976

ACKNOWLEDGEMENTS

The authors wish to gratefully acknowledge the constant co-operation received during the preparation of this report from the Education Branch, the Computer Information Systems Division and from Mr. W. Zayachkowski and Dr. S. Kumar of the Departmental Statistics Division.

Special mention need be made of Mr. S. Deering for the computational assistance, Mr. T. Vella for the cover design, graphs and charts and of Mrs. P. Murphy for the diligent typing. The authors are extremely thankful to them.

D.E. Stewart,
Statistician.

D.G. Saigaonkar,
Senior Statistician.

TABLE OF CONTENTS

TABLE OF CONTENTS

(cont'd)

Page

1. Introduction

This technical report is being prepared as a complement to the publication entitled, "Enrolment of Registered Indian Students - Trend Analysis and Historical Review, 1949-50 to 1974-75".

The analytical techniques which were employed in order to arrive at the enrolment projections for the school year 1979-80 (page 2 of the said report) are documented in this report.

Table I gives enrolment figures of registered Indian students for the school years 1949-50 to 1974-75 inclusive, at the Canada level. The projections for the next five years have been arrived at on the basis of this historical data.

2. Concepts, Restrictions and Assumptions on the Various Methods Attempted

Six different estimating procedures presented in this report were used to project the registered Indian student enrolment for the next five school years.

The first approach is known as the "Method of Least Squares". Methods I and II although differing with respect to their individual models are all based on the concept of least squares fit to the data. This approach is explained as follows. Consider Figure 2.1 in which the data points are given by (X_1, Y_1), (X_2, Y_2),.....,(X_n, Y_n). For a given value of X, say X_1, there will be a difference between the value Y_1 and the corresponding value as determined from the curve C. As shown in figure 2.1 this difference is denoted by D_1, which is referred to as a deviation, error or residual and may be positive, negative or zero.

Similarly, we obtain the deviations $D_2, \ldots\ldots, D_n$ corresponding to the values $X_2 \ldots, X_n$. A measure of the "goodness of fit" of the curve C to the given data is provided by the quantity $D_1^2 + D_2^2 + \ldots + D_n^2$. If this is small the fit is reasonable, if it is large the fit is not reasonable. As a result, we make the following statement.

Definition: Of all curves fitting a given set of data points, the curve having the property that $D_1^2 + D_2^2 + \ldots + D_n^2$ is a minimum is called a best fitting curve.

A curve having this property is said to fit the data in the least square sense and is called a least square curve. Thus a line having this property is called a least square line, a parabola with this property is called a least square parabola, etc.

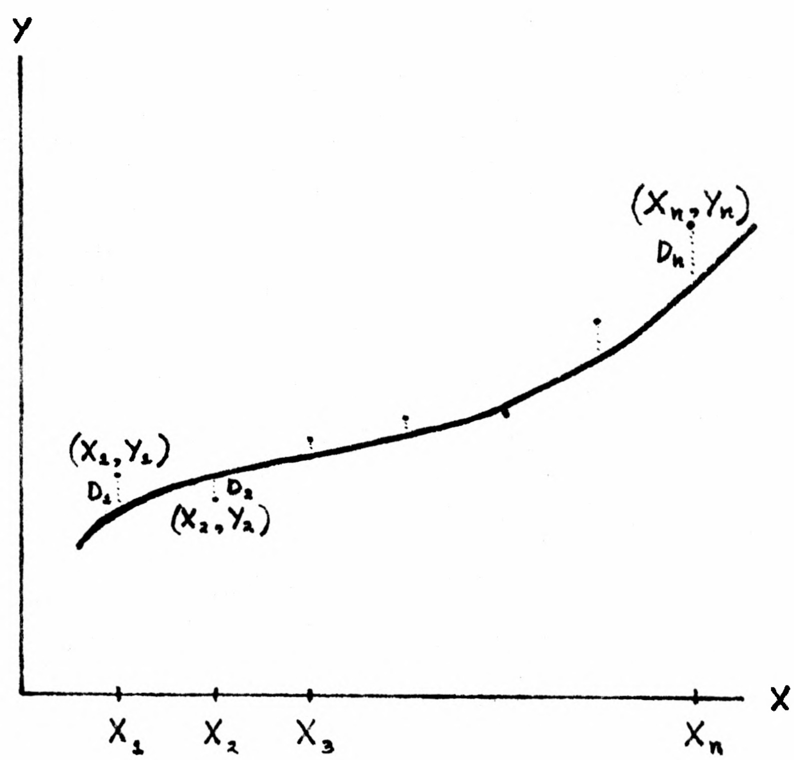


Figure 2.1

This method of analysis only takes into account the overall trend evident in the data. Any major fluctuations or deviations from this general trend in the observed data are not compensated for in this method of analysis. Consequently, the forecast figures produced are only based on the best fitting curve with respect to the past trend established and assume that the same trend will continue in the future.

The second approach (see Method III) was based on calculating first differences between successive years of the actual enrolment figures and then computing the percentage increase or decrease of these first differences with respect to the actual enrolment figures. With the percentage increases or decreases computed for all successive years, the Method of Least Squares was applied to these resulting percentages in order to forecast percentage changes in enrolment for 1975-76 to 1979-80. Based on these projected percentage changes, the enrolment projections were computed. The basic assumption is that the computed percentage changes are adequately estimated by a least square linear fit. The restrictions and limitations on this method are similar to those of the least square estimation.

The third approach (Method IV) involved the fitting of an exponential curve to the enrolment figures for the school years 1967-68 to 1974-75. The basis for using this estimation procedure is that the enrolment figures displayed an increasing trend with a noticeable exponential decay from the 1967-68 school year. The assumption here is that the above-mentioned trend will continue for the next five years. Although this method of analysis produced estimates very close to the actual

...4

enrolment figures for the school years 1967-68 to 1974-75, all of the past historical data (e.g. 1949-50 to 1966-67) is not considered in the analysis. This is the only drawback of this method.

The last approach implemented utilizes "Time Series Analysis" to produce forecast estimates. This method of analysis can be mathematically defined by the observed values $Y_1, Y_2, \ldots, Y_n$ of a variable Y (temperature, enrolment, etc.) at times $t_1, t_2, \ldots, t_n$. Thus Y is a function of t, symbolized by Y = F(t). This function F(t) is composed of characteristic movements which are usually classified into four main types, often called components of a time series as follows:

(a) Long term or secular movements, T

(b) Cyclical movements, C

(c) Seasonal movements, S

(d) Irregular or random movements, I.

A more detailed explanation of these components and their effects on the analysis is presented later. Let it suffice to say for our purpose here that a time series variable Y is a product of the variables T, C, S and I which produce the trend, cyclical, seasonal and irregular movements in the series respectively. The analysis of the time series comprises an investigation of these movements and is referred to as a decomposition of a time series into its basic component movements. The assumption inherent in this method of analysis is that the basic long term movements identified will continue in the future. The only restriction or limitation on this technique is that the forecast values of T, C, S and I must be estimated independently of each other.

...5

Having discussed the basic concepts, restrictions and assumptions relevant to the various methods considered, the detailed analysis and forecast values are presented for each of the methods.

3. Method I: Least Square Linear Fit

Assuming that the data may be adequately fitted to a straight line, the least square line approximating the set of points (X_1, Y_1), (X_2, Y_2),....., (X_n, Y_n) will be used and is given by the equation,

$$Y_i = B_o + B_1 X_i + E_i \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where,

(a) X_i is a given year and X_i takes the 26 integral values 0, 1, 2,...,25 corresponding to the years 1949-50, 1950-51, 1951-52,...,1974-75 respectively,

(b) Y_i is the observed number of registered Indian students for the corresponding year X_i,

(c) According to Least Squares theory, B_o and B_1 are constants which are estimated by solving the following normal equations,

$$\left.\begin{aligned} \sum_{i=0}^{25} Y_i &= B_o 26 + B_1 \sum_{i=0}^{25} X_i \\ \sum_{i=0}^{25} X_i Y_i &= B_o \sum_{i=0}^{25} X_i + B_1 \sum_{i=0}^{25} X_i^2 \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots (2)$$

(d) E_i is said to be the regression error. For a derivation of the normal equations see Appendix I. Least Square fitting of the line can be mathematically simplified by transforming the data so that $x_i = X_i - \bar{X}$ and $y_i = Y_i - \bar{Y}$ where $\bar{X}$ and $\bar{Y}$ are the arithmetic means of the X_i's and Y_i's respectively. The equation of the least square line in (x_i, y_i) is,

$$y_i = \frac{\sum_{i=0}^{25} x_i y_i}{\sum_{i=0}^{25} x_i^2} \cdot x_i \quad \text{.............................(3)}$$

We will use equation (3) for our purpose. The data in Table I is transformed into (x_i, y_i) and using it we have the following relevant sums and sums of squares:

$$\sum_{i=0}^{25} X_i = 325$$

$$\bar{X} = \sum_{i=0}^{25} X_i/26 = 12.5$$

$$\sum_{i=0}^{25} Y_i = 1,239,966$$

$$\bar{Y} = \sum_{i=0}^{25} Y_i/26 = 47,690.77 \simeq 47,691$$

$$\sum_{i=0}^{25} x_i = \sum_{i=0}^{25} (X_i - \bar{X}) = 0$$

$$\sum_{i=0}^{25} x_i^2 = \sum_{i=0}^{25} (X_i - \bar{X})^2 = 1,462.5 \simeq 1,463$$

$$\sum_{i=0}^{25} x_i y_i = \sum_{i=0}^{25} (X_i - \bar{X})(Y_i - \bar{Y}) = 3,079,698.5 \simeq 3,079,699$$

Hence the least square line is

$$y = \frac{\sum_{i=0}^{25} x_i y_i}{\sum_{i=0}^{25} x_i^2} .x = \frac{3,079,699}{1,463} .x = 2,105x$$

which can be rewritten in original variables (X_i, Y_i) as,

$$Y_i - 47,691 = 2,105 (X_i - 12.5)$$

or

$$Y_i = 21,379 + 2,105 X_i \text{.............................(4)}$$

where the origin $X_i = 0$ corresponds to the school year 1949-50, $X_i = 1$ corresponds to 1950-51, and so on. Using equation (4), the forecast figures of registered Indian student enrolment for the years 1975-76 to 1979-80 are computed by substituting the values 26, 27, 28, 29 and 30 for X_i in equation (4).

The results of these computations are presented in Table V and in graph 1.

4. Method II: Least Square Parabolic Fit

This method uses the assumption that the data can be adequately fitted to a parabolic curve (i.e. a non-linear relationship). Examination of Graph 2 shows that there are at least two main points of inflection or trend changes occurring between 1949-50 and 1974-75 and there is an increasing trend in enrolment. For this reason the least square parabolic fit was considered.

The least square parabola approximating the set of points (X_1, Y_1) $(X_2, Y_2), \ldots, (X_n, Y_n)$ is represented by the equation,

$$Y_i = B_o + B_1X_1 + B_2X_i^2 + E_i \ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

where,

(a) X_i is a given school year and has the integral values 0, 1, 2,..., 25 corresponding to the years 1949-50, 1950-51, 1951-52,..., 1974-75 respectively,

(b) Y_i is the observed number of registered Indian students for the corresponding year X_i,

(c) According to Least Squares theory, B_o, B_1 and B_2 are constants which are estimated by solving the following normal equations,

$$\left.\begin{array}{l}
\sum_{i=0}^{25} Y_i = B_o N + B_1 \sum_{i=0}^{25} X_i + B_2 \sum_{i=0}^{25} X_i^2 \\[2ex]
\sum_{i=0}^{25} X_i Y_i = B_o \sum_{i=0}^{25} X_i + B_1 \sum_{i=0}^{25} X_i^2 + B_2 \sum_{i=0}^{25} X_i^3 \\[2ex]
\sum_{i=0}^{25} X_i^2 Y_i = B_o \sum_{i=0}^{25} X_i^2 + B_1 \sum_{i=0}^{25} X_i^3 + B_1 \sum_{i=0}^{25} X_i^4
\end{array}\right\} \qquad \text{(6)}$$

(b) E_i is the regression error.

For a derivation of the normal equations (6), see Appendix II.

As before in Section 3, it is mathematically convenient to use the transformed variables (x_i, y_i) where $x_i = X_i - \bar{X}$ and $y_i = Y_i - \bar{Y}$.

Note that $\Sigma x_i = \Sigma x_i^3 = 0$. Using this we have the following sums and sums of squares,

$$\sum_{i=0}^{25} x_i = 0$$

$$\sum_{i=0}^{25} y_i = 1,239,966$$

$$\sum_{i=0}^{25} x_i^2 = 5,850$$

$$\sum_{i=0}^{25} x_i^3 = 0$$

$$\sum_{i=0}^{25} x_i^4 = 2,364,570$$

$$\sum_{i=0}^{25} x_i y_i = 6,158,314$$

$$\sum_{i=0}^{25} x_i^2 \, y_i = 285{,}191{,}574$$

Substituting these values in (6), the normal equations are:

$$\left.\begin{aligned} 26B_o + 5{,}850\ B_2 &= 1{,}239{,}966 \\ 5{,}850\ B_1 &= 6{,}158{,}314 \\ 5{,}850\ B_o + 2{,}364{,}570\ B_2 &= 285{,}191{,}574 \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots (7)$$

Solving these equations in (7), we have $B_o = 46{,}361.25$, $B_1 = 1{,}051.22$ and $B_2 = 5.91$. Hence $E(y_i) = 46{,}361.25 + 1{,}051.22x_i + 5.91x_i^2$ ....(7*). Least Square Parabola fit in original variables (X_i, Y_i) is obtained from (7*) by using $x_i = X_i - 12.5$. Therefore,

$$Y_i = 46{,}361.25 + 2{,}102.44\ (X_i - 12.5) + 23.64\ (X_i - 12.5)^2$$

or

$$Y_i = 23{,}774.5 + 1{,}511.44\ X_i + 23.64\ X_i^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

Using equation (8), forecasts of registered Indian student enrolment for the years 1975-76 to 1979-80 are computed by substituting the values 26, 27, 28, 29 and 30 for X_i respectively. The results of these computations are presented in Table V and in graph 2.

5. Comparative Evaluation of Projections using Methods I and II

Table II gives, (i) the actual enrolment figures, (ii) the estimated enrolment figures produced by the equations (4) and (8) of methods I and II respectively, and (iii) the amount of deviation of estimated figures from actual figures for both methods. Examining this table more closely, we observe the following:

(a) Deviations: Deviation is defined as the difference between actual enrolment and estimates. In general, (i) the estimates produced by method II show less deviation from the actual enrolment figures than those of method I in 16 of the 26 school years; (ii) for 8 school years, method I gives estimates with less deviation from the actual enrolment than the estimates by method II.

(b) Trends: Further examination of Table II shows the following:

(i) From 1949-50 to 1956-57, both estimating procedures tend to under-estimate the actual enrolment figures;

(ii) From 1957-58 to 1964-65, both estimating procedures tend to over-estimate the actual enrolment figures;

(iii) From 1965-66 to 1972-73, both estimating procedures tend to under-estimate the actual enrolment figures, and

(iv) For the school years 1973-74 and 1974-75 both estimating procedures tend to over-estimate.

(v) For the school years 1954-55 and 1969-70 estimates of enrolment by both methods are almost identical. There appears to be a unique pattern developed by these estimates which may be identified. Both methods appear to go through a period of 8-10 years of over-estimation and under-estimation.

Use of either estimating procedure would appear to produce estimates which are in excess of the actual enrolment figures for the next 4-5 school years.

6. Method III: First Differences & Their Percentage Increase With Respect to Observed Enrolment

Definition:

The method of first differences and their percentage increase may be explained as follows:

Let $\hat{P}_{26}, \hat{P}_{27}, \ldots, \hat{P}_{30}$ be the estimated percentage increases in student enrolment for the school years 1975-76, 1976-77,..., 1979-80. Then the projected student enrolment for these school years is given by,

$$\hat{Y}_i = \hat{Y}_{i-1} (\hat{p}_i/100 + 1), \; i = 26, 27, \ldots, 30.$$

The assumption that the percentage increase or decrease of the first differences with respect to the actual enrolment figures are adequately estimated by a linear trend is employed here. Table III reveals these first differences and their percentages of the actual enrolment figures. Based on these percentage values, a least square linear fit of the form

$$Y_i = B_o + B_1 X_i + E_i \quad \ldots\ldots(12)$$

was applied. The estimates of B_o and B_1 were obtained and the expected percentages were computed by substituting the values X_i = 0, 1, 2,..., 25 for the years 1949-50, 1950-51, 1951-52,... , 1974-75 respectively. The expected percentages for 1975-76 to 1979-80 were obtained by substituting the values X_i = 26, 27, 28, 29 and 30 into equation (12) above. The forecast figures of registered Indian student enrolment for the school years 1975-76 to 1979-80 were computed by multiplying each percentage by the corresponding observed enrolment figure to obtain the enrolment increase, then adding this enrolment increase to the observed enrolment figure in order to obtain the expected enrolment for the next school year. Now we illustrate this

...13

method by the following example.

Example: Expected percentage increase for 1975-76 is 2.556. Multiplying the 1974-75 actual enrolment of 72,249 by 1.02556 gives 74,096 which is the expected enrolment for the 1975-76 school year. In the same manner, forecast figures for each of the school years 1976-77 through to 1979-80 were obtained by using the preceding years estimate and are presented in Table V and in graph 3.

7. Method IV: Exponential Fit

Examining the plot of the enrolment figures in graph 4, the enrolment is increasing and appears to have a noticeable exponential decay from the 1967-68 school year. For this reason the following general exponentially-increasing model was fitted to the actual enrolment data for the school years 1967-68 to 1974-75 inclusive:

$$E = A + Bq^{Y} \qquad (13)$$

where E is the number of Indian students enrolled for a given school year Y, B and q are regression co-efficients, and A is an asymptote to the function. It should be pointed out that $Y = 0, 1, \ldots, 7$ corresponds to the school years 1967-68, 1968-69,..., 1974-75. An iterative procedure in order to obtain the minimum sum of square estimates for the parameters A, B and q was implemented on the computer, yielding

$$E = 74{,}253.38 - 15{,}267.55 \times 0.73^{Y} \qquad (14)$$

The details of this estimation procedure and the computer results are described in Appendix III.

By substituting the values Y = 8, 9, 10, 11 and 12 respectively in equation (14), the forecast figures of registered Indian student enrolment for the school years 1975-76 to 1979-80 were computed. The results of these forecasts are given in Table V and are also shown in graph 4.

8. Method V: Time Series Analysis

The analysis of data by time series generally consists of a mathematical investigation of component movements present in it. To introduce these concepts consider Figure 10.1 which refers to ideal time series. We have the following:

Definitions: (i) Long-Term Trend: It refers to the general direction in in which the graph of a time series appears to be going over a long interval of time. It is also known as a secular trend.

(ii) Cyclical Movements: They refer to the long-term oscillations or swings about a trend line or curve. These cycles may or may not be periodic.

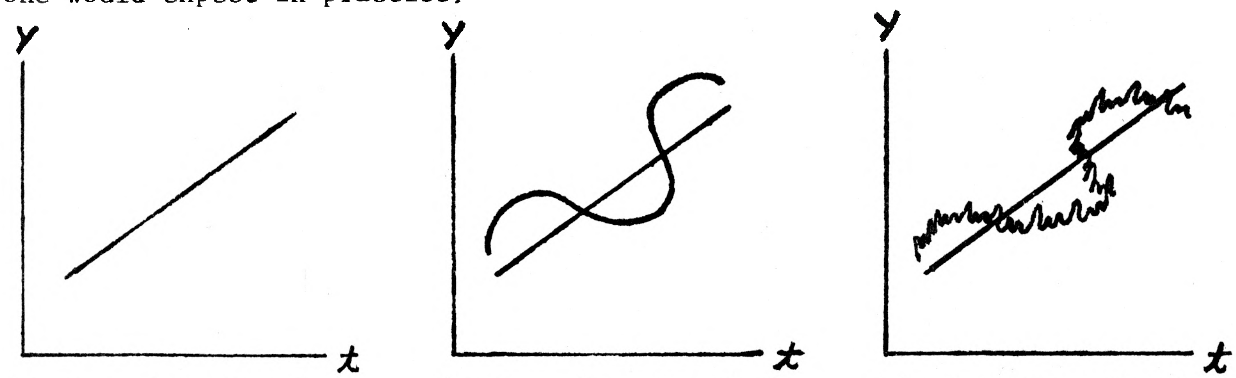
(iii) Seasonal Movements: They refer to the identical, or almost identical, patterns which a time series appears to follow during corresponding periods of successive time intervals.

(iv) Irregular or Random Movements: These refer to sporadic motions of time series due to chance events. Although it is ordinarily assumed that such events produce variations lasting only a short time, it is conceivable that they may

...15

be so intense as to result in new cyclical or other movements.

These trends are exhibited graphically in Figure 10.1. Figure (a) displays a graph of a long-term or secular trend line. Figure (b) shows the long-term line with a super-imposed cyclical movement (assumed to be periodic). Figure (c) shows a super-position of a seasonal movement on the graph of Figure (b). If one were to super-impose on the graph (c) some random or irregular movements, the result would look more like a time series that one would expect in practice.



(a) Long-Term Trend

(b) Long-Term Trend & Cyclical Movement

(c) Long-Term Trend, Cyclical & Seasonal Movements

Figure 10.1

For our analysis we assume that the time series variable Y is a product of the variables T, C, S and I which denote respectively the trend, cyclical, seasonal and irregular movements. Mathematically,

$$Y_i = T_i \times C_i \times S_i \times I_i \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (15)$$

The analysis of a time series consists of an investigation of these components T, C, S and I and is referred to as a "decomposition" of a time series into its basic component movements.

Now, we apply the time series analysis to project the student enrolment for the school years 1975-76 to 1979-80.

The enrolment data in Table III is used for estimating the various components.

The projected values for student enrolment are given by $E_i = T_i \times C_i \times S_i \times I_i$, where T_i, C_i, S_i and I_i must be estimated and E_i is the student enrolment.

(a) Estimation of the Trend, T_i: To estimate the trend T_i, many possible methods can be used including the Method of Least Squares; The Freehand Method; The Moving Average Method; or the Method of Semi-Averages. The Method of Least Squares is used to estimate the trend component, T_i, here. In fact, Least Square Linear Fit is used and the $\hat{T}_i$'s computed are identical to the $\hat{Y}_i$'s estimated in Method I.

(b) Estimation of Seasonal Component, S_i: Seasonal variations refer to identical, or almost identical patterns which a time series appears to follow during corresponding months, weeks, days or hours of successive years. Since the enrolment data itself is on a yearly basis, there is no justification for including a seasonal component, S_i, in the projection model. Due to this fact, our model becomes,

$$E_i = T_i \times C_i \times I_i \quad \ldots\ldots (16)$$

(c) Estimation of Cyclic Component, C_i: Now consider the cyclical component C and the irregular variations I which may be present in the data. From (16), one can see that division of E by T gives C x I; i.e. cyclical and irregular variations. In practice it is found that irregular movements tend to be of small magnitude and that they often tend to follow the pattern of a normal distribution, i.e. small deviations occur with large frequency, large deviations occur with small frequency. As a result of this, the $C_i \times I_i$ component may be assumed to contain basically cyclical effects present in the data. Based on this, we proceeded with the analysis by dividing the observed enrolment figures, E_i by the estimated trend values, $\hat{T}_i$, in order to adjust for trend. The results, given as percentages, are given in Table IV.

With this adjustement for trend, we have $E_i/\hat{T}_i = C_i \hat{\times} I_i$, subtraction of 100(%) gives $E_i/\hat{T}_i - 100 = (C_i \hat{\times} I_i) - 100$ and is given in Table V. Thus, in Graph **7** the independent variable is time t_i.

Graph 7 is theoretically composed of only the cyclical and irregular movements contained in the enrolment data represented by the corresponding components C_i and I_i respectively. From the graph, it is evident that the cyclical component undergoes a decreasing linear trend for the first 11 school years, an increasing linear trend for the next 11 school years and is displaying a decreasing linear trend over the past four school years. Under the assumption that this decreasing linear trend will continue for at least the next five school years, a least square linear fit was applied to the last five cyclical figures in order to project cyclic and irregular components for the school years 1975-76 to 1979-80 inclusive. The results are given in Table IV.

Now we may apply these cyclical and irregular components to the time series model (16) as follows. To project the student enrolment we use equation (16), i.e.

$$\hat{E}_i = \hat{T}_i \times \hat{C}_i \times \hat{I}_i$$

$$= \hat{T}_i \times (C \hat{\times} I)_i$$

where,

$\hat{T}_i$'s are given in Table II and $(C \hat{\times} I)_i$'s are given in Table IV for 1975-76 to 1979-80 inclusive. The results of these computations are presented in Table V and are shown in graph 5.

*Note: A comparison of enrolment projections by all methods for the school years 1975-76 to 1979-80 is shown in graph 6.

9. Discussion on the Methods of Analysis Presented

(a) Method I:

This method is the simplest to apply in order to produce forecast figures. The amount of time and computations involved is minimal in comparison with other methods attempted. Therefore rough estimates may be obtained by the use of this method. However, the basic underlying linear trend in the data is accounted for with no consideration being given towards unusual deviations or trend changes occurring in the observed enrolment figures.

(b) Method II:

A least square parabola was fitted to the enrolment figures. Computations involved here are a little more complex and time consuming than those of method I. An added feature of estimation is gained through the application of this method, i.e. a parabolic fit allows one point of inflection to be identified in the observed enrolment figures. Examining the enrolment figures, there is at least one point of inflection evident. The application of this method resulted in projection figures for the school years 1975-76 to 1979-80 that deviated very much from the natural trend displayed since the 1967-68 school year. Hence, these estimates were not intuitively regarded as feasible.

(c) Method III:

This method is advantageous in that it is simple to apply and requires minimal computations in order to produce the forecast figures. However, since the number of estimates computed in arriving at the projections increases with time, therefore the errors compounded over time as well. As can be seen from Table III the expected percentage increase figures are not compatible with the actuals. As a result, the expected percentage increase figures produced do not appear to have any basis on which to believe that they may be representative projections of what the true values may be. Therefore, the resulting projection estimates of enrolment for the school years 1975-76 to 1979-80 are based on very little statistical support.

(d) Method IV:

Since the enrolment data display a noticeable exponential decay since the 1967-68 school year (see Graph 4), a general exponentially-increasing model appeared to be a suitable fit. The expected enrolments computed under this model for the school years 1967-68 to 1974-75 compared with the observed enrolments are extremely close and within a maximum deviation of 1 per cent in the 1968-69 school year. On the average, the expected and observed enrolment figures do not deviate from one another by more than 0.4 per cent. As a result of this it appeared very plausible to compute forecast estimates for student enrolment during the school years 1975-76 to 1979-80. It is assumed that the increasing trend would continue along with the exponentially decaying trend in the future. The approach has a limitation, i.e. enrolment figures for the past eight school years 1967-68 to 1974-75 are used in this approach, while the information available for the earlier years is not used.

...20

(e) Method V:

This particular approach employs the most sensitive analysis techniques in order to produce efficient forecast estimates. Not only are long-term or secular movements accounted for, but the effects of other characteristic movements or variations that may be present in the time series data (some or all of which are present to varying degrees) are also computed. The forecast figures predicted by this method have a strong statistical basis and are therefore recommended for future reference.

10. Enrolment Projections by Grade-Groups

The estimates for student enrolment at the national level for the years 1975-76 to 1979-80 are used as a basis for preparation of projections of student enrolments by suitable grade groups at the national level.

10.1 The Necessity of Grade-Groups

The necessity for using grade-groups may be supported by the following factors:

a) K4 and K5 are grouped together to form the "kindergarten" or "pre-school" student group. This differs from the others because the students only attend school for half-days, the program is becoming more increasingly band-administered year by year, and the nature of the cirriculum for K4 and K5 students is quite different from the other grades in the school system.

b) Grades I-VI are grouped together because they are generally found in Public Schools resulting in separate management and buildings for these grades, and the single teacher class system is employed for this elementary school system.

...21

c) As may be observed from the grade system by province (Chart I), the patterns of junior high and senior high schools are different depending on the province. Hence, separate estimates for grades VII and VIII are required.

d) Grades IX - XI are grouped together since these three years of secondary education are more or less similar in nature.

e) Grades XII and XIII form a grade-group since the final secondary graduating year is either XII or XIII depending on the region. Also, credits for University Education entrance are given in either of these graduating years.

f) The assistance program of the DIAND is different for grades I-VIII as compared to IX-XI; e.g., all students over 14 years of age or those in grade 9 or higher get pocket money amounting to $10.00 per month, or so.

g) Students' Honour System - some students are allowed to manage their lodging, boarding, etc. financed by the Department depending upon criteria such as continuation in school, achievement, behaviour, etc.

10.2 Forecasts of Student Enrolments in Schools by Grade-Groups at the National Level 1975-76 to 1979-80

Two methods for forecasting student enrolments in schools by grade-groups were considered. The methods and concepts employed in each will be described briefly.

Method A:

This method uses the group proportions of the total student enrolment for 1974-75 in order to estimate the enrolment for the school years 1975-76 to 1979-80. The assumption involved here is that the grade-group proportions that exist in 1974-75 will continue to remain constant over the next five

...22

years. The advantage of this method is the minimum amount of computation involved allowing one to obtain estimates quickly, whereas, the major disadvantage appears to be the powerful assumption stated above. For the actual and rounded estimated figures for the next five years see Table VI.

Method B:

A second method of estimation is based upon the trends of the proportions of student enrolments in the respective grade-groups to the total student enrolment (i.e. proportions follow the trend established over the past years).

Example -

For the K4 and K5 grade-group enrolment, the proportions of this group to the total enrolment for the past years were computed. Having obtained these proportions, a least square linear fit was applied to them in order to estimate the K4 and K5 proportions for the next five years. Based on these proportions, the forecasts for the K4 and K5 enrolment for the next five years were computed. In the same fashion, the estimated proportions and subsequent enrolments for the other grade-groups were computed.

The assumption here is that the proportions for the respective grade-groups over the next five years will follow the trend which has been established over the past 15 years or so.

Therefore, the advantage of this method over Method A is the more realistic and less stringent assumption involved. The major disadvantage, of course, is the increased number of computations involved resulting in a much greater amount of time required in order to produce the forecast figures. The actual and rounded estimated figures for the next five years computed via Method B

...23

may be observed in Table VII..

Note: (1) No effort was made to compare the efficiencies of Methods A and B with respect to their forecasting abilities because both are forecasts only.

(2) Bar Charts depicting the estimated student enrolment for the different grade-groups over the next five years may be observed for Method B in Chart II. The estimates computed via Method B are recommended for future reference.

11. Enrolment Forecasts by School Type

Initially, the responsibility of providing educational facilities to registered Indian students was mainly limited to the Federal Government. A number of elementary schools were built on the Reserves for this purpose. Over a number of years, the participation of provincial schools and parochial schools was also inevitable. In some of these schools, joint responsibility for capital expenditure is also borne by the Federal Government, while in others tuition and overhead expenses are paid for every Indian student studying in the school. Over the past 26 years considerable responsibility for educating Indian Students has been undertaken by these non-federal agencies. In the 1949-50 school year, only 6.4% of the Indian students attended non-federal schools. This percentage has substantially increased to 57.4% during the school year 1974-75. For the purpose of enrolment forecasts, the proportion of Indian students in non-federal schools over the past eight years, during which time it seems to have stabilized considerably, has been averaged. It is assumed that this average proportion will remain constant over the next five years. The resulting enrolment forecasts in federal and non-federal schools are given below:-

...24

Enrolment Forecasts of Registered Indian Students
1975-76 to 1979-80

Year	Federal	Non-Federal	Total
1975-76	30,716	43,566	74,282
1976-77	31,080	44,084	75,164
1977-78	31,419	44,563	75,982
1978-79	31,765	45,054	76,819
1979-80	32,052	45,461	77,513

Note: Band-operated schools are included under federal.

REFERENCES

1. Daniel & Wood (1971), Fitting Equations to Data, John Wiley & Sons, Inc.

2. Dixon, W.J. (1970), BMD (Biomedical Computer Programs), University of California Press.

3. Spiegel, Murray R. (1961), Theory and Problems of Statistics, McGraw-Hill, Inc.

4. Stevens, W.L. (1951), Asymptotic Regression, Biometrics, Vol. 7, No.3.

5. Winer, B.J. (1971), Statistical Principles in Experimental Design, McGraw-Hill, Inc.

TABLE I

ENROLMENT OF REGISTERED INDIAN STUDENTS IN FEDERAL AND NON-FEDERAL SCHOOLS, BY GRADE

INDIENS INSCRITS - INSCRIPTIONS DANS LES ECOLES FEDERALES ET NON-FEDERALES, SELON LE NIVEAU

CANADA

Year - Année	K_4* M_4	K_5* M_5	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII-XIII	Spec.	Total
1949-50	-	-	8,129	3,793	3,401	2,638	2,235	1,610	1,077	685	443	136	98	50	12	24,307
1950-51	-	-	8,373	4,156	3,694	2,909	2,416	1,710	1,176	727	564	178	126	79	14	26,122
1951-52	-	-	8,739	4,358	3,675	3,079	2,505	1,830	1,271	831	622	231	126	112	16	27,395
1952-53	-	-	8,289	4,411	3,868	3,359	2,712	2,015	1,406	928	613	303	172	105	43	28,224
1953-54	-	-	8,740	4,775	4,290	3,585	2,925	2,325	1,651	1,076	757	349	220	93	75	30,861
1954-55	-	-	8,875	4,821	4,471	3,622	3,033	2,391	1,722	1,076	718	431	243	145	69	31,617
1955-56	-	-	9,535	5,159	4,639	3,990	3,284	2,612	1,937	1,326	742	459	273	179	90	34,225
1956-57	-	2,247	7,578	5,401	4,917	4,367	3,487	2,876	2,040	1,451	978	478	306	195	36	36,357
1957-58	-	2,562	6,645	5,219	4,629	4,205	3,743	2,801	2,077	1,367	1,002	462	284	186	36	35,218
1958-59	-	2,516	7,125	5,337	4,954	4,420	3,890	3,147	2,331	1,516	1,015	638	275	200	68	37,432
1959-60	-	2,687	6,905	5,688	5,197	4,717	4,048	3,337	2,606	1,700	1,109	596	380	182	75	39,227
1960-61	-	2,911	6,646	5,654	5,401	4,747	4,231	3,633	2,750	1,933	1,286	688	415	280	1,096	41,671
1961-62	-	3,560	6,650	6,005	5,726	5,293	4,594	3,817	3,340	2,292	1,663	812	472	362	858	45,444
1962-63	-	3,678	6,811	6,235	5,923	5,227	4,959	3,964	3,277	2,411	1,704	1,021	513	376	875	46,974
1963-64	-	3,858	7,269	6,493	6,171	5,625	4,748	4,254	3,489	2,577	1,931	1,122	611	343	770	49,261
1964-65	-	3,874	7,457	6,600	6,089	5,756	5,189	4,303	3,818	2,818	2,276	1,195	709	505	804	51,393
1965-66	-	3,538	8,280	7,078	6,520	5,901	5,595	4,684	3,892	3,095	2,436	1,392	764	535	960	54,670
1966-67	-	3,694	8,697	7,303	6,766	6,045	5,676	5,049	4,169	3,170	2,544	1,482	868	490	1,205	57,158
1967-68	-	4,409	8,702	7,200	6,668	6,424	5,776	4,969	4,501	3,447	2,750	1,750	929	570	1,232	59,327
1968-69	152	5,610	8,485	7,321	6,789	6,344	5,852	5,215	4,613	3,692	3,032	1,915	1,223	706	1,500	62,449
1969-70	427	6,380	8,851	7,392	6,813	6,512	6,075	5,706	5,030	4,090	3,401	2,208	1,245	896	1,207	66,233
1970-71	1,330	5,506	8,604	7,300	6,856	6,561	6,227	5,835	5,462	4,320	3,724	2,650	1,493	1,006	1,575	68,449
1971-72	1,749	5,958	8,370	7,605	6,841	6,538	6,402	5,823	5,420	4,661	4,043	2,850	1,681	1,040	1,480	70,461
1972-73	2,371	5,439	8,430	7,161	7,175	6,522	6,238	5,990	5,716	4,849	4,219	2,910	1,769	1,076	1,274	71,139
1973-74	2,898	5,768	8,137	7,174	6,855	6,788	6,104	6,053	5,586	5,186	4,169	3,071	1,791	1,185	1,338	72,103
1974-75	3,698	5,575	8,287	6,792	6,903	6,500	6,496	5,763	6,074	4,914	4,262	2,952	1,774	992	1,267	72,249

* Kindergarten, four- and five-year olds -- Maternelles, les enfants âgés de 4 et 5 ans.

Table II

Comparison of the Enrolment Projection Formulae for the Least Square Linear Fit and the Least Square Parabolic Fit with Respect to Actual Enrolment Figures

School Year	Actual Enrolment	Method I Least Square — Linear Fit		Method II Least Square — Parabolic Fit	
		Estimates	Deviation From Actual Enrolment	Estimates	Deviation From Actual Enrolment
1949-50	24,307	21,379	2,928*	23,775**	532*
1950-51	26,122	23,484	2,638*	25,310**	812*
1951-52	27,395	25,589	1,806*	26,892**	503*
1952-53	28,224	27,694	530*	28,522**	298
1953-54	30,861	29,799	1,062*	30,199**	662*
1954-55	31,617	31,904	287	31,923	306
1955-56	34,225	34,009**	216*	33,694	531*
1956-57	36,357	36,114**	243*	35,513	844*
1957-58	35,218	38,219	3,001	37,379**	2,161
1958-59	37,432	40,324	2,892	39,292**	1,860
1959-60	39,227	42,429	3,202	41,253**	2,026
1960-61	41,671	44,534	2,863	43,261**	1,590
1961-62	45,444	46,639	1,195	45,316**	128*
1962-63	46,974	48,744	1,770	47,418**	444
1963-64	49,261	50,849	1,588	49,568**	307
1964-65	51,393	52,954	1,561	51,765**	372
1965-66	54,670	55,059**	389	54,009	661*
1966-67	57,158	57,164**	6	56,301	857*
1967-68	59,327	59,269**	58*	58,640	687*
1968-69	62,449	61,374**	1,075*	61,026	1,423*
1969-70	66,233	63,479	2,754*	63,459	2,774*
1970-71	68,449	65,584	2,865*	65,940**	2,509*
1971-72	70,461	67,689	2,772*	68,468**	1,993*
1972-73	71,139	69,794	1,345*	71,043**	96*
1973-74	72,103	71,899**	204*	73,666	1,563
1974-75	72,249	74,004**	1,755	76,336	4,087

Note: (1) * - An estimate less than the actual enrolment figure.
(2) ** - An estimate reasonably closer to the actual enrolment.

Table III

First Differences, Their Percentages of the

Actual Enrolment Figures and Projections

School Year	Actual Enrolment (E_i)	First Differences ($E_i - E_{i-1}$)	Increase/Decrease (%) $(E_i - E_{i-1})/(E_{i-1}) \times 100$
1949-50	24,307	1,815	7.47
1950-51	26,122	1,273	4.87
1951-52	27,395	829	3.03
1952-53	28,224	2,637	9.34
1953-54	30,861	756	2.45
1954-55	31,617	2,608	8.25
1955-56	34,225	2,132	6.23
1956-57	36,357	-1,139	-3.13
1957-58	35,218	2,214	6.29
1958-59	37,432	1,795	4.80
1959-60	39,227	2,444	6.23
1960-61	41,671	3,773	9.05
1961-62	45,444	1,530	3.37
1962-63	46,974	2,287	4.87
1963-64	49,261	2,132	4.33
1964-65	51,393	3,277	6.38
1965-66	54,670	2,488	4.55
1966-67	57,158	2,169	3.79
1967-68	59,327	3,122	5.26
1968-69	62,449	3,784	6.06
1969-70	66,233	2,216	3.35
1970-71	68,449	2,012	2.94
1971-72	70,461	678	0.96
1972-73	71,139	964	1.36
1973-74	72,103	146	0.20
1974-75	72,249		
1975-76	Projected Values		2.56
1976-77			2.41
1977-78			2.26
1978-79			2.11
1979-80			1.96

Table IV

Estimation of the Cyclical Component, C

School Year	$(C \hat{x} I)_i = \left[\frac{\hat{E}_i}{T_i} \times 100\right]\%$	Deviation from 100% $\left[(C \hat{x} I_i) - 100\right]$
1949-50	113.7	13.7
1950-51	111.2	11.2
1951-52	107.1	7.1
1952-53	101.9	1.9
1953-54	103.6	3.6
1954-55	99.1	-0.9
1955-56	100.6	0.6
1956-57	100.7	0.7
1957-58	92.1	-7.9
1958-59	92.8	-7.2
1959-60	92.5	-7.5
1960-61	93.6	-6.4
1961-62	97.4	-2.6
1962-63	96.4	-3.6
1963-64	96.9	-3.1
1964-65	97.1	-2.9
1965-66	99.3	-0.7
1966-67	100.0	0.0
1967-68	100.1	0.1
1968-69	101.8	1.8
1969-70	104.3	4.3
1970-71	104.4	4.4
1971-72	104.1	4.1
1972-73	101.9	1.9
1973-74	100.3	0.3
1974-75	99.0	-1.0
1975-76	97.6*	-2.4
1976-77	96.1*	-3.9
1977-78	94.6*	-5.4
1978-79	93.2*	-6.8
1979-80	91.7*	-8.3

Note: * - Estimates based on Linear Regression Analysis.

Table V

Enrolment Projections of Registered Indian Students at the National Level

1975-76 to 1979-80

School Year	Method				
	I	II	III	IV	V
1975-76	76,109 (76,100)	79,053 (79,100)	74,095 (74,100)	73,056 (73,100)	74,282 (74,300)
1976-77	78,214 (79,200)	81,817 (81,800)	75,878 (75,900)	73,383 (73,400)	75,164 (75,200)
1977-78	80,319 (80,300)	84,629 (84,600)	77,591 (77,600)	73,620 (73,600)	75,982 (76,000)
1978-79	82,424 (82,400)	87,488 (87,500)	79,227 (79,200)	73,792 (73,800)	76,819 (76,800)
1979-80	84,529 (84,500)	90,394 (90,400)	80,779 (80,800)	73,919 (73,900)	77,513 (77,500)

Note: (1) Figures indicated in brackets are the values of the expected enrolment rounded to the nearest hundred.

(2) The forecasts of Method V are recommended for future reference.

Table VI

Enrolment Forecasts for Registered Indian Students in Federal and Non-Federal Schools

By Grade-Groups

1975-76 to 1979-80

Method A								
YEARS	Grade-Groups							
	K4 + K5	I-VI	VII	VIII	IX-XI	XII/XIII	SPL.	Total Enrolment
	Proportions of Total Enrolment (%)							
1974-75	12.83	56.40	8.41	6.80	12.44	1.37	1.75	100.00
	Forecasts (No. of Students)							
1975-76	9,533 (9,500)	41,905 (41,900)	6,249 (6,300)	5,052 (5,100)	9,243 (9,200)	1,018 (1,000)	1,300 (1,300)	74,300 (74,300)
1976-77	9,648 (9,700)	42,413 (42,400)	6,324 (6,300)	5,114 (5,100)	9,355 (9,400)	1,030 (1,000)	1,316 (1,300)	75,200 (75,200)
1977-78	9,751 (9,700)	42,864 (42,900)	6,392 (6,400)	5,168 (5,200)	9,454 (9,500)	1,041 (1,000)	1,330 (1,300)	76,000 (76,000)
1978-79	9,854 (9,800)	43,315 (43,300)	6,459 (6,500)	5,222 (5,200)	9,554 (9,600)	1,052 (1,100)	1,344 (1,300)	76,800 (76,800)
1979-80	9,943 (9,900)	43,710 (43,700)	6,518 (6,500)	5,270 (5,300)	9,641 (9,600)	1,062 (1,100)	1,356 (1,400)	77,500 (77,500)

Note: Figures in brackets indicate enrolment forecasts rounded to the nearest hundred.

Table VII

Enrolment Forecasts for Registered Indian Students in Federal and Non-Federal Schools

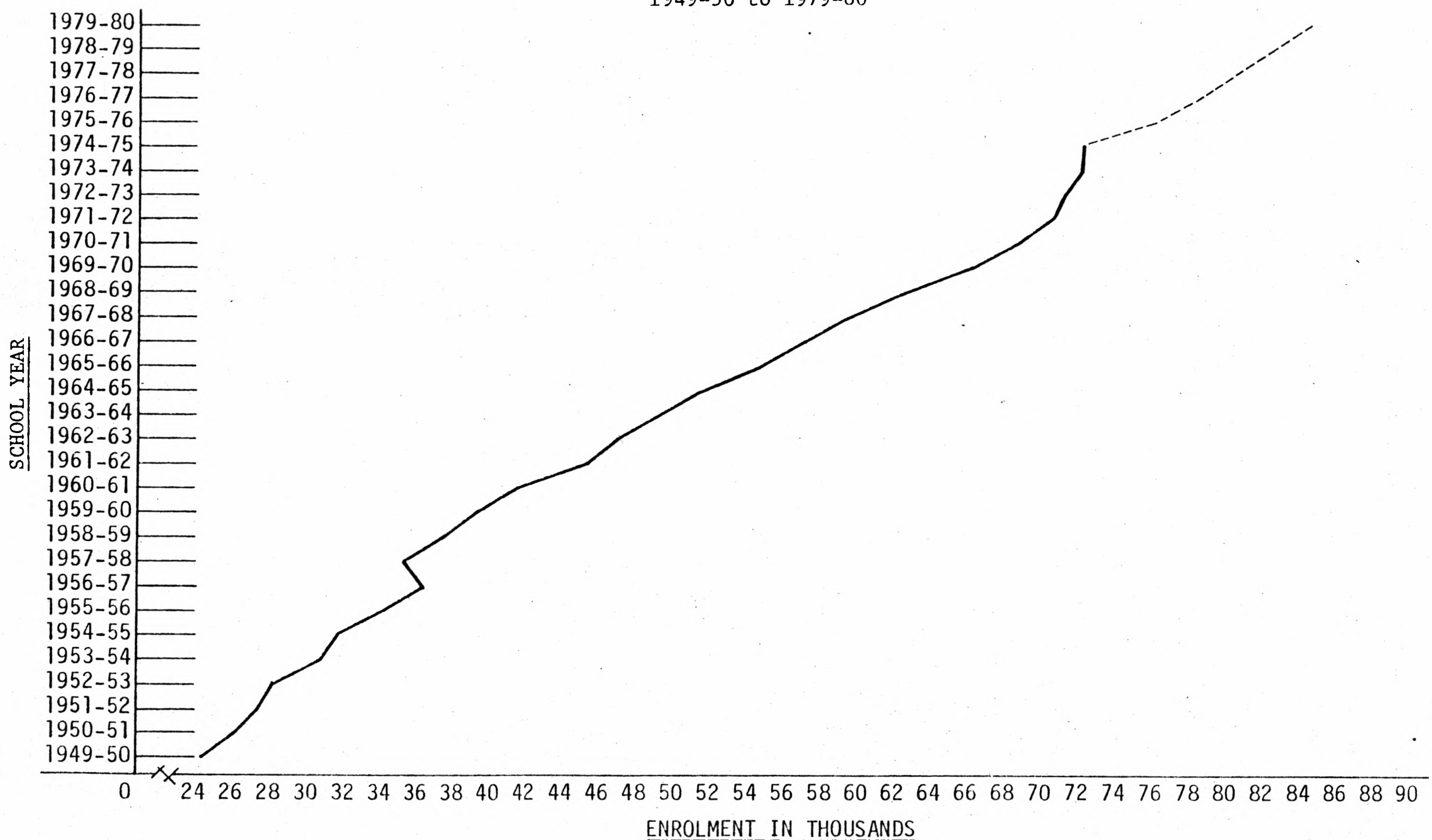
By Grade-Groups

1975-76 to 1979-80

METHOD B																
	Grade-Groups															
Years	K4 + K5		I - VI		VII		VIII		IX - XI		XII/XIII		SPL.		Total Enrolment	
	Enrolment Percentages (%) and Forecasts (F)															
	%	F	%	F	%	F	%	F	%	F	%	F	%	F	%	F
1975-76	12.57	9,339 (9,400)	55.48	41,222 (41,200)	8.65	6,427 (6,400)	7.28	5,409 (5,400)	12.83	9,533 (9,500)	1.54	1,144 (1,200)	1.65	1,226 (1,200)	100.00	74,300 (74,300)
1976-77	12.60	9,475 (9,500)	54.66	41,105 (41,100)	8.80	6,618 (6,600)	7.45	5,602 (5,600)	13.24	9,956 (10,000)	1.60	1,203 (1,200)	1.65	1,241 (1,200)	100.00	75,200 (75,200)
1977-78	12.64	9,606 (9,600)	53.81	40,896 (40,900)	8.95	6,802 (6,800)	7.63	5,799 (5,800)	13.66	10,382 (10,400)	1.65	1,254 (1,200)	1.66	1,261 (1,300)	100.00	76,000 (76,000)
1978-79	12.71	9,761 (9,700)	52.95	40,666 (40,700)	9.10	6,989 (7,000)	7.80	5,990 (6,000)	14.07	10,806 (10,800)	1.71	1,313 (1,300)	1.66	1,275 (1,300)	100.00	76,800 (76,800)
1979-80	12.78	9,904 (9,900)	51.96	40,269 (40,300)	9.25	7,169 (7,200)	7.98	6,185 (6,200)	14.48	11,222 (11,200)	1.76	1,364 (1,300)	1.79	1,387 (1,400)	100.00	77,500 (77,500)

Note: Figures in brackets indicate enrolment forecasts rounded to the nearest hundred.

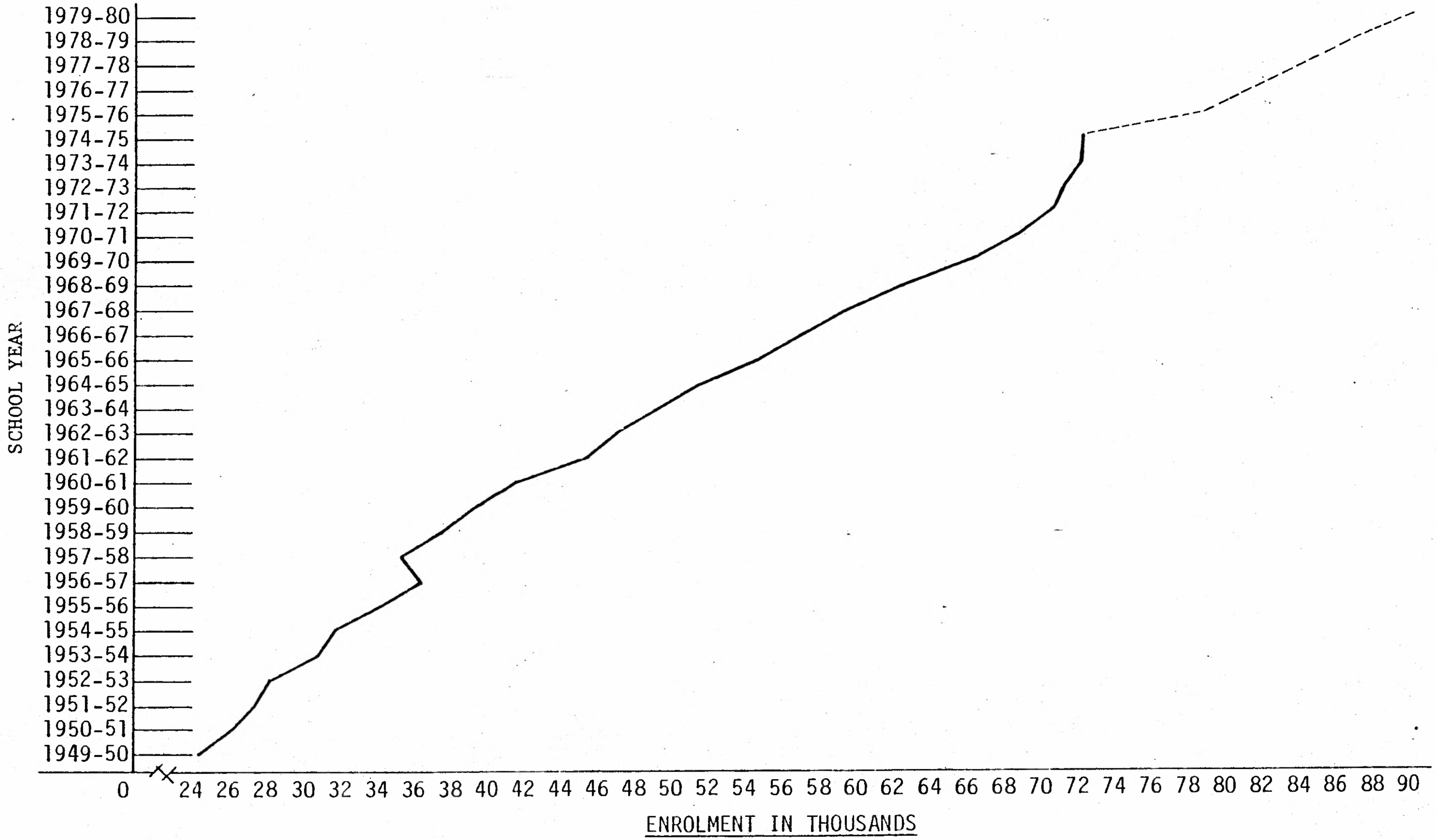
GRAPH 1
Registered Indian Student Enrolment
in Federal and Non-Federal Schools
1949-50 to 1979-80



Notes: ———— observed enrolment
- - - - - - - - enrolment forecasts
Graph on page 37 compares enrolment forecasts by all methods.

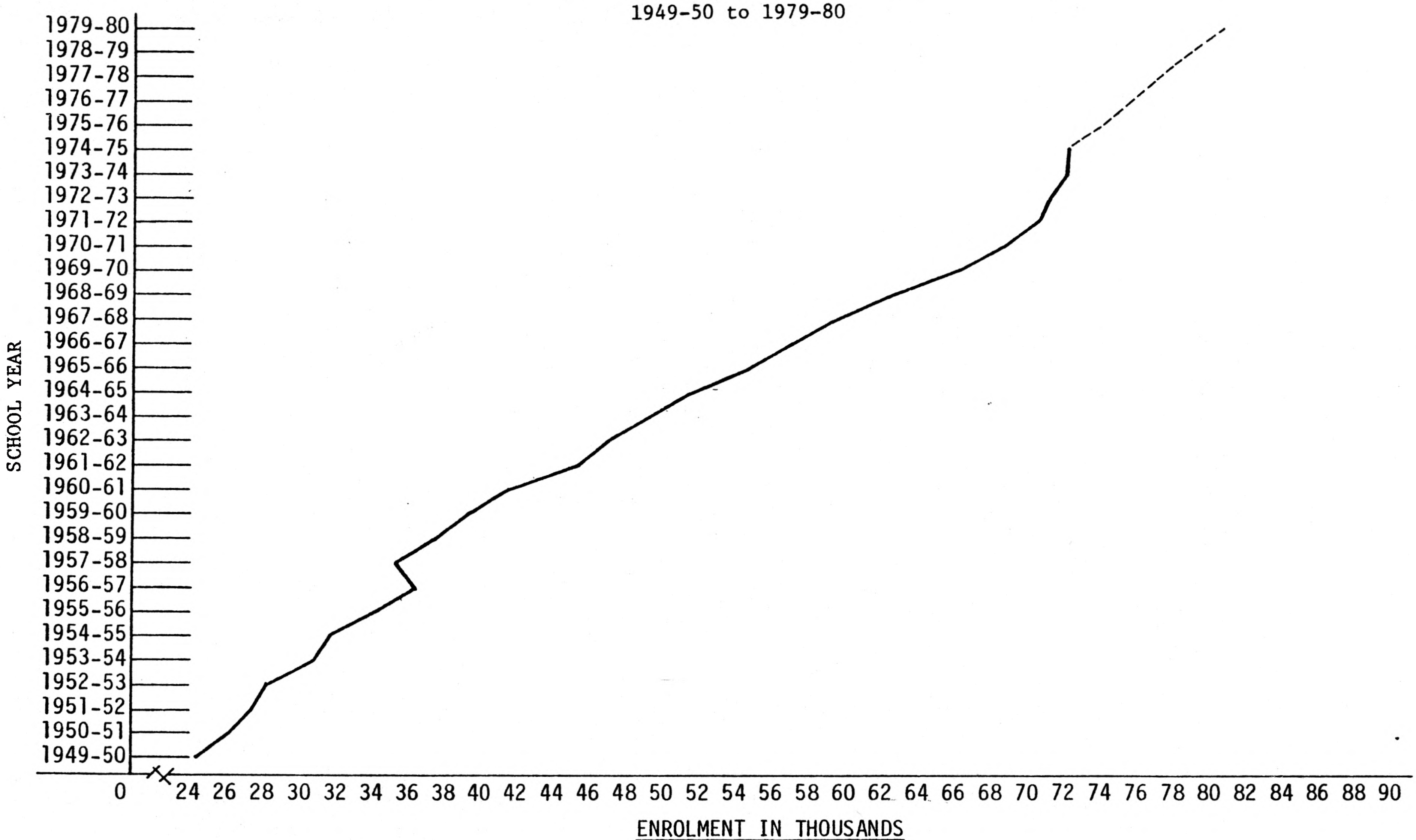
GRAPH 2

Registered Indian Student Enrolment in Federal and Non-Federal Schools 1949-50 to 1979-80



Notes: ———— observed enrolment
------ enrolment forecasts by Method II
Graph on page 37 compares enrolment forecasts by all methods.

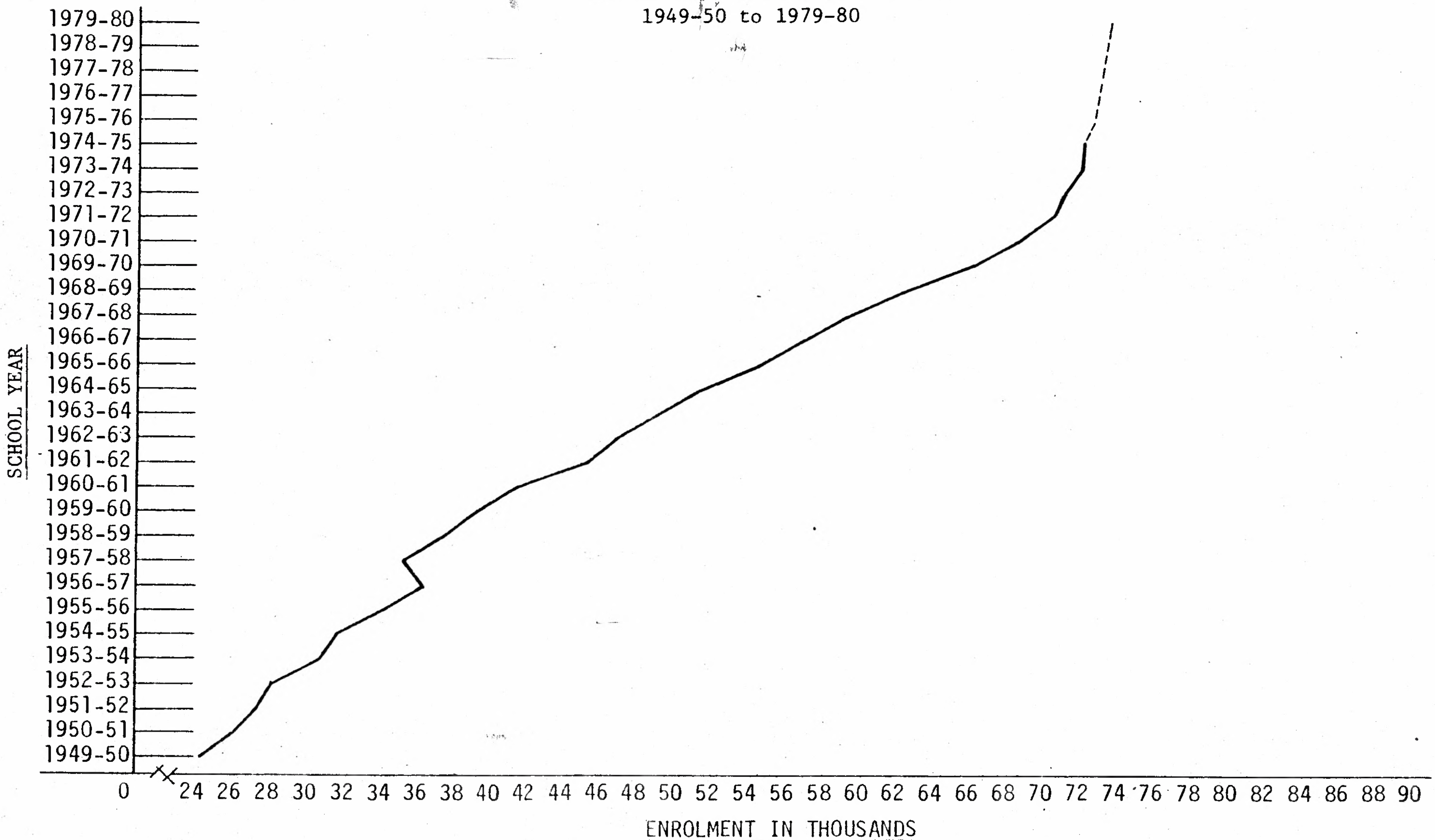
GRAPH 3
Registered Indian Student Enrolment in
Federal and Non-Federal Schools
1949-50 to 1979-80



Notes: ——— observed enrolment
------ enrolment forecasts by Method II
Graph on page 37 compares enrolment forecasts
by all methods.

GRAPH 4

Registered Indian Student Enrolment in
Federal and Non-Federal Schools
1949-50 to 1979-80

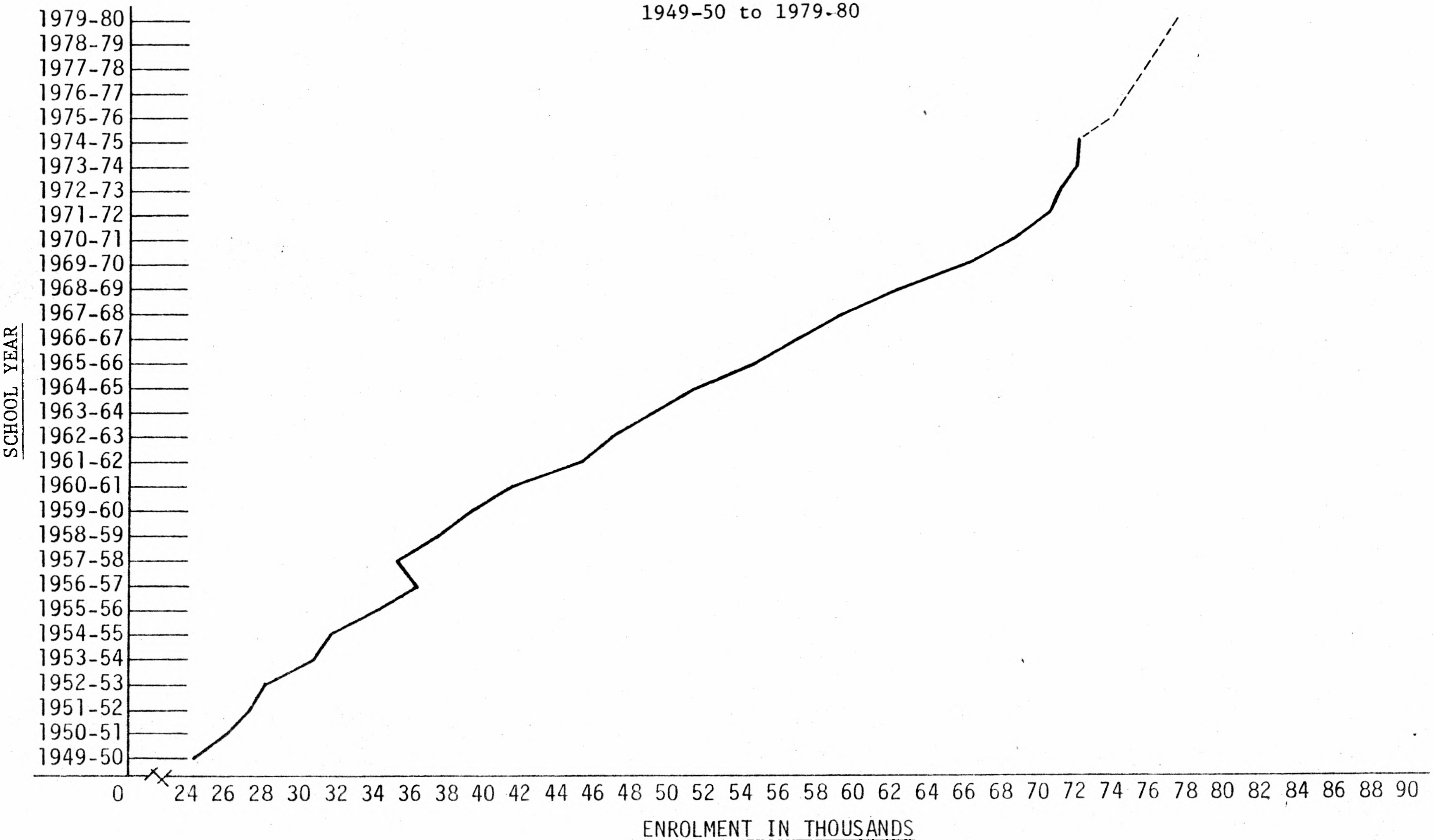


Notes: ———— observed enrolment

-------enrolment forecasts by Method IV

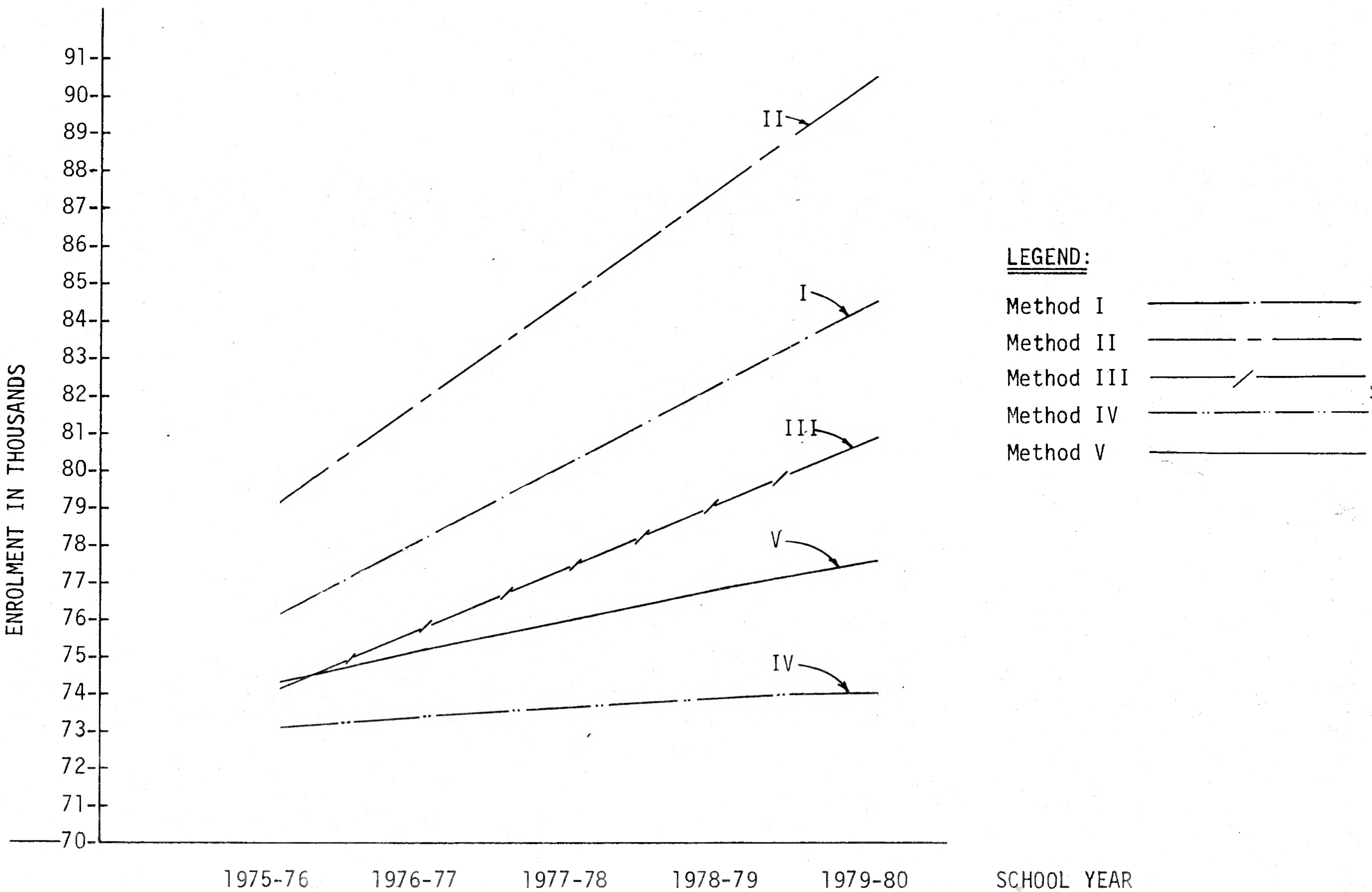
Graph on page 37 compares enrolment forecasts by all methods.

GRAPH 5
Registered Indian Student Enrolment in
Federal and Non-Federal Schools
1949-50 to 1979-80



Notes: ———— observed enrolment
-------enrolment forecasts by Method V
Graph on page 37 compares enrolment forecasts
by all methods.

Graph 6
A COMPARISON OF ENROLMENT PROJECTIONS
BY ALL METHODS
1975-76 -- 1979-80



CYCLIC AND IRREGULAR VARIATIONS

GRAPH 7

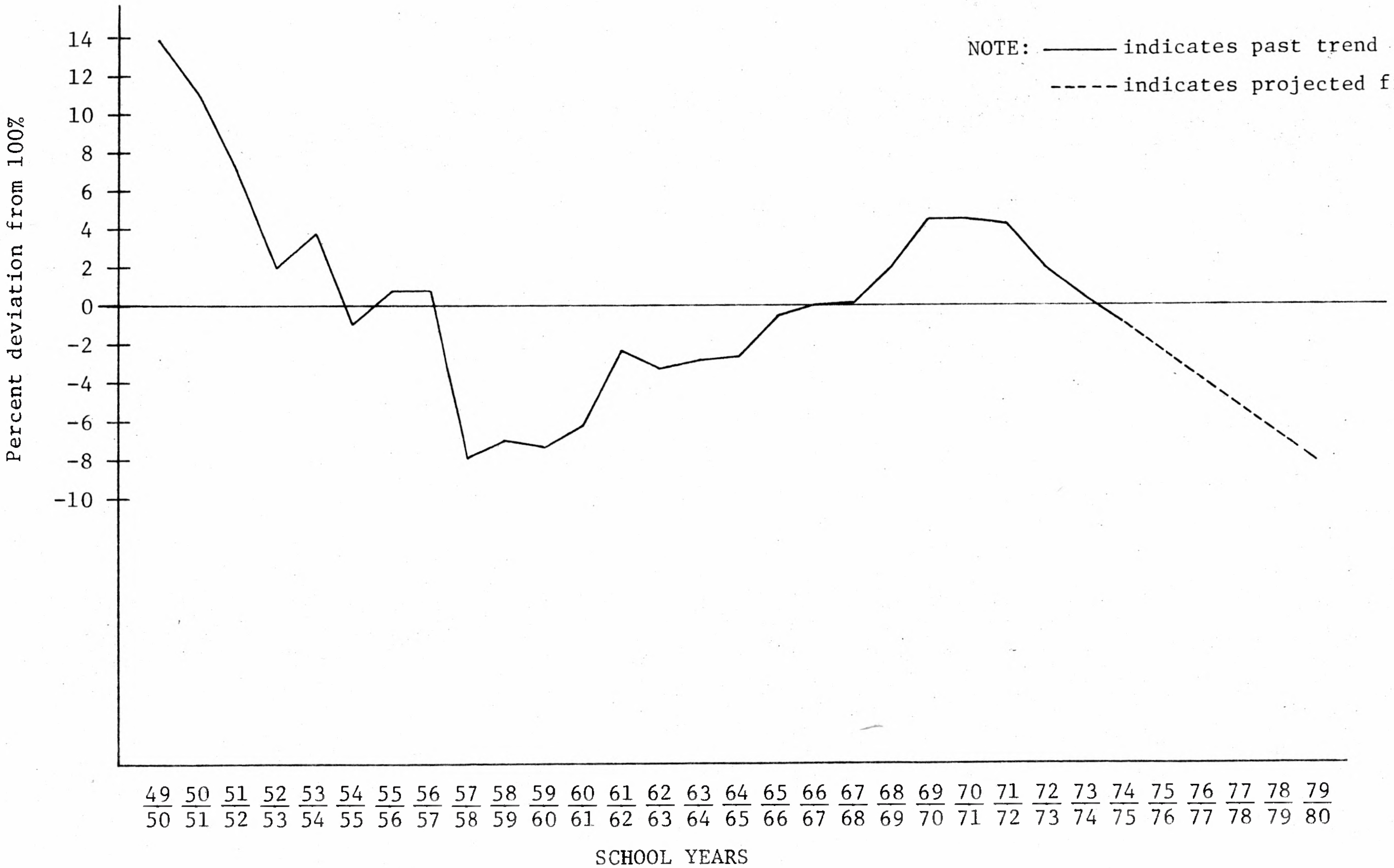
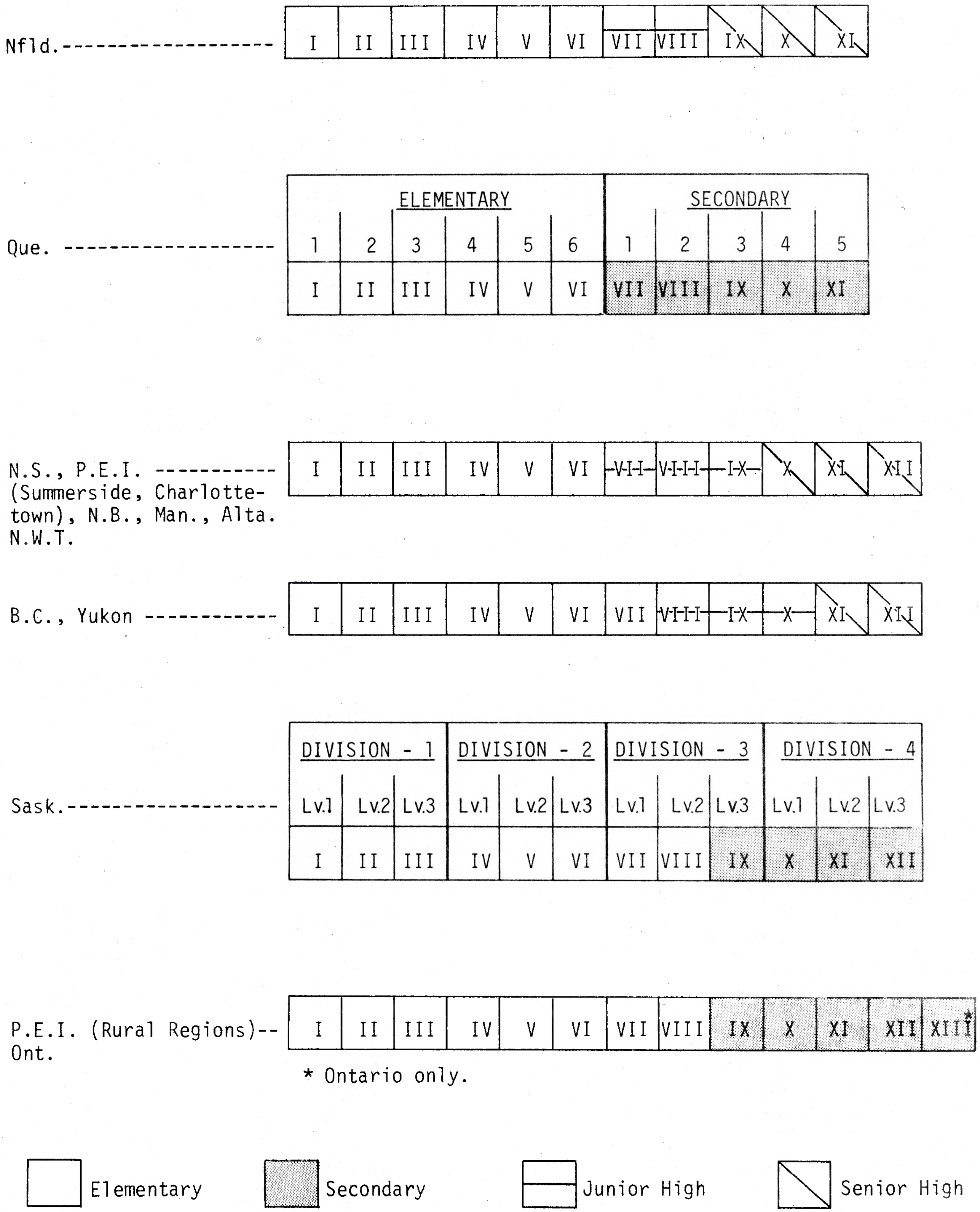


CHART I

Educational Grade Systems By Province

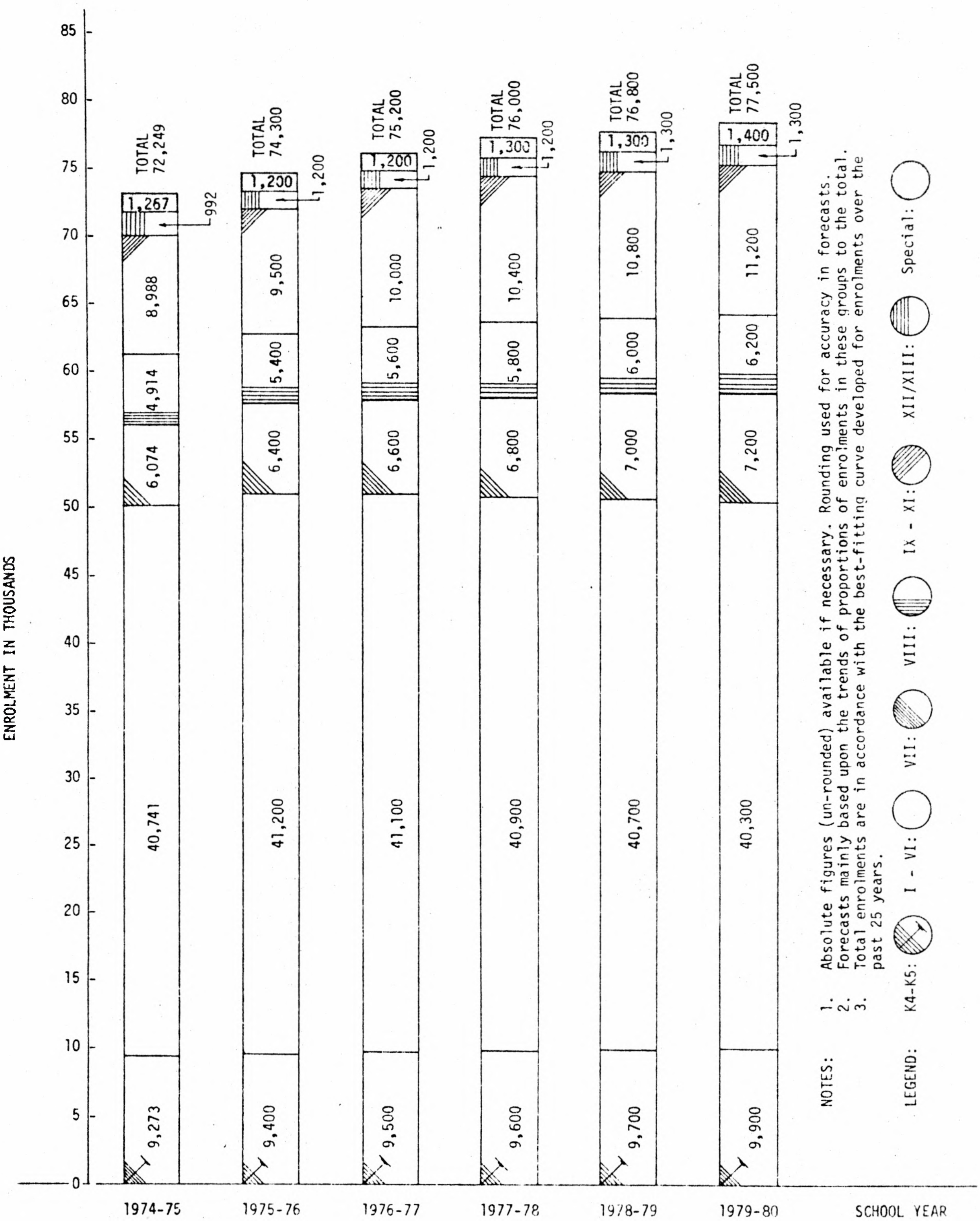


Source: Statistics Canada

FORECASTS OF STUDENT ENROLMENTS IN SCHOOLS - CANADA

1975-76 TO 1979-80

CHART II



Appendix I

Derivation of Normal Equations for Least Square Line Fit

Let the equation of the required least square line be $Y_i = B_o + B_1 X_i + E_i$.

Rearranging this equation, we have

$$E_i = Y - B_o - B_1 X$$

$$\therefore \Sigma E_i^2 = \Sigma (Y - B_o - B_1 X)^2$$

Now taking the partial derivatives

$$\frac{\partial(\Sigma E_i^2)}{\partial B_o}, \quad \frac{\partial(\Sigma E_i^2)}{\partial B_1}$$

and equating them to zero gives the "minimum least square" or "Maximum Likelihood" normal equations for estimators B_o and B_1.

$$\frac{\partial(\Sigma E_i^2)}{\partial B_o} = \frac{\partial(\Sigma\ Y - B_o - B_1 X^2)}{\partial B_o}$$

$$= 2\ \Sigma(Y - B_o - B_1 X)\cdot(-1)$$

Now equating this expression to zero, we have:

$$(Y - B_o - B_1 X) = 0$$

$$\therefore \Sigma Y - NB_o - B_1\ \Sigma X = 0$$

or,

$$\Sigma Y = B_o N + B_1\ \Sigma X \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

$$\frac{\partial(\Sigma E_i^2)}{\partial B_1} = \frac{\partial(\Sigma\ Y - B_o - B_1\ X\ \ ^2)}{\partial B_1}$$

$$= 2\ \Sigma(Y - B_o - B_1 X)\cdot(-X)$$

Now equating to zero, we have:

$$\Sigma(XY - B_oX - B_1X^2) = 0$$

$$\therefore \quad \Sigma XY - B_o \Sigma X - B_1 \Sigma X^2 = 0$$

or,

$$\Sigma XY = B_o \Sigma X + B_1 \Sigma X^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Appendix II

Derivation of Normal Equations for Least Square Parabolic Fit

Let the equation of the required least square parabolic fit be,

$$Y_i = B_o + B_1 X_i + B_2 X_i^2 + E_i.$$

Rearranging this equation, we have

$$E_i = Y - B_o - B_1 X - B_2 X^2$$

$$\therefore\ \Sigma E_i^2 = \Sigma(Y - B_o - B_1 X - B_2 X^2)^2$$

Now taking the partial derivatives

$$\frac{\partial(\Sigma E_i^2)}{\partial B_o},\quad \frac{\partial(\Sigma E_i^2)}{\partial B_1} \text{ and } \frac{\partial(\Sigma E_i^2)}{\partial B_2},$$

equating them to zero yields the "minimum least square" or "Maximum Likelihood" normal equations for estimating B_o, B_1 and B_2.

$$\frac{\partial(\Sigma E_i^2)}{\partial B_o} = \frac{\partial(\Sigma\,[Y - B_o - B_1 X - B_2 X^2]^2)}{\partial B_o}$$

$$= 2\ \Sigma(Y - B_o - B_1 X - B_2 X^2)\cdot(-1)$$

Now equating to zero, we have,

$$\Sigma(Y - B_o - B_1 X - B_2 X^2) = 0$$

$$\therefore\ \Sigma Y - B_o N - B_1\ \Sigma X - B_2\ \Sigma X^2 = 0$$

or,

$$\Sigma Y = B_o N + B_1\ \Sigma X + B_2\ \Sigma X^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

$$\frac{\partial(\Sigma E_i^2)}{\partial B_1} = \frac{\partial(\Sigma[Y - B_o - B_1 X - B_2 X^2]^2)}{\partial B_1}$$

$$= 2\ \Sigma\ (Y - B_o - B_1 X - B_2 X^2)\cdot(-X)^2$$

Now equating to zero, we have,

$$\Sigma(XY - B_oX - B_1X^2 - B_2X^3) = 0$$

$$\therefore \Sigma XY - B_o \Sigma X - B_1 \Sigma X^2 - B_2 \Sigma X^3 = 0$$

or,

$$\Sigma XY = B_o\Sigma X + B_1\Sigma X^2 + B_2 \Sigma X^3 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

$$\frac{\partial(\Sigma E_i^2)}{\partial B_2} = \frac{\partial(\Sigma[Y - B_o - B_1X - B_2X^2]^2)}{\partial B_2}$$

$$= 2\Sigma(Y - B_o - B_1X - B_2X^2)\cdot(-X^2)$$

Now equating to zero, we have,

$$\Sigma(X^2Y - B_oX^2 - B_1X^3 - B_2X^4) = 0$$

$$\therefore \Sigma X^2Y - B_o\Sigma X^2 - B_1\Sigma X^3 - B_2\Sigma X^4 = 0$$

or,

$$\Sigma X^2Y = B_o\Sigma X^2 + B_1\Sigma X^3 + B_2\Sigma X^4 \quad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Appendix III

Asymptotic Regression

1. General Description:

The program used performs asymptotic regression analysis using the following modified exponential function,

$$E = A + Bq^{Y}$$

where, i) E is the enrolment for a particular year Y,

ii) A is the asymptote to the function,

iii) B and q are regression coefficients, the values of which determine the degree of dampening on the enrolment increase as the school years progress.

2. Computational Procedure:

By the least squares approach, this program fits the regression function,

$$E = A + Bq^{Y}$$

to a set of data values,

$$(Y_i,\ E_{ij}),\ (i = 1, 2, \ldots, k),\ (j = 1, 2, \ldots, m_i)$$

where $Y_1 < Y_2 < \ldots < Y_k$, $m_i \geq 1$ and $(Y_i,\ E_{ij})$ are the actual or observed values.

Let a, b and r be estimates of A, B and q.

Step 1

Sums, averages, and estimate of q

$$\text{Set } E_{i.} = \frac{1}{m_i} \sum_{j=1}^{m_i} E_{ij},\quad \hat{E}_o = \sum_{i=1}^{k} m_i E_{i.}$$

A trial value of the least-square estimate of q is defined by,

$$r_o = \left[E_{2.} - E_{k.}\right] / \left[E_{1.} - E_{(k-1).}\right]$$

if this constant is not specified.

Step 2

Information matrix Y* and sums $\hat{E}_1$ and $\hat{E}_2$

$$Y^* = \begin{bmatrix} \sum_{i=1}^{k} m_i & \Sigma m_i (r_o)^{Y_i} & \Sigma m_i Y_i (r_o)^{Y_i - 1} \\ \Sigma m_i (r_o)^{Y_i} & \Sigma m_i (r_o)^{2Y_i} & \Sigma m_i Y_i (r_o)^{2Y_i - 1} \\ \Sigma m_i Y_i (r_o)^{Y_i - 1} & \Sigma m_i Y_i (r_o)^{2Y_i - 1} & \Sigma m_i Y_i^2 (r_o)^{2Y_i - 2} \end{bmatrix}$$

$$\hat{E}_1 = \sum_{i=1}^{k} m_i (r_o)^{y_i} E_{i.} \quad , \quad \hat{E}_2 = \sum_{i=1}^{k} m_i Y_i E_{i.} (r_o)^{Y_i - 1}$$

Step 3

The information matrix Y* is inverted, giving the covariance matrix F.

$$F = (Y^*)^{-1} = \begin{bmatrix} F_{aa} & F_{ab} & F_{ar} \\ F_{ab} & F_{bb} & F_{br} \\ F_{ar} & F_{br} & F_{rr} \end{bmatrix}$$

Using matrix F, estimates of the parameters are found as follows:

$$a = F_{aa}\hat{E}_o + F_{ab}\hat{E}_1 + F_{ar}\hat{E}_2$$

$$b = F_{ab}\hat{E}_o + F_{bb}\hat{E}_1 + F_{br}\hat{E}_2$$

$$r = r_o + (F_{ar}\hat{E}_o + F_{br}\hat{E}_1 + F_{rr}\hat{E}_2)/b$$

Step 4

A test of the "goodness of fit" of the iterated solution is performed.

Let,

$\tilde{E}_i = a + b\,(r)^{Y_i}$ and Sum $= \Sigma(\tilde{E}_i - E_{i.})^2$

If $\left|\dfrac{r_o - r}{r}\right| > 10^{-5}$, set $r_o = r$ and go to Step 2; otherwise go to Step 5.

Step 5

Standard deviations and the analysis of variance.

(i) Estimated standard error of a

$$S_a = Se\sqrt{F_{aa}}$$

(ii) Estimated standard error of b

$$S_b = Se\sqrt{F_{bb}}$$

(iii) Estimated standard error of r

$$S_r = Se\sqrt{F_{rr}}/\,|b|$$

where Se is the estimated standard deviation away from the regression curve that is fitted to the $(Y_i, E_{i.})$ pairs.

$$Se = \left[\Sigma\Sigma\,(E_{ij} - \tilde{E}_i)^2/(N - 3)\right]^{\frac{1}{2}},$$

where,

$$N = \sum_{i=1}^{k} m_i .$$

ANALYSIS OF VARIANCE

Deviation	Sum of Squares
(i) From Mean	$\Sigma\Sigma(E_{ij} - E_{..})^2$
(ii) From Y_i Means	$\Sigma\Sigma(E_{ij} - E_{i.})^2$
(iii) From Curve	$\Sigma\Sigma(E_{ij} - \tilde{E}_i)^2$
(iv) Of Means from Curve	$\Sigma\Sigma(E_{i.} - \tilde{E}_i)^2 = \Sigma m_i(E_i - \tilde{E}_i)^2$

BMD06R - ASYMPTOTIC REGRESSION - REVISED JANUARY 12, 1971
HEALTH SCIENCES COMPUTING FACILITY,UCLA

REGRESSION EQUATION

$$Y = A + B.R^{X}$$

PROBLEM CARD

```
PROBLEM CODE        QUICKY
NO. OF X VALUES          8          INPUT PATTERN           XY
X RE-SCALED                           SORT
PRINT RESIDUALS        YES
X TRANS CODE             0          OUTPUT DATA            YES
X CONSTANT             0.0          VARIABLE FORMAT          1
THE VARIABLE FORMAT IS  (F2.0,12F6.0)
```

ORIGINAL DATA

NO.	X VALUE	Y VALUE
1	0.0	59327.0000
2	1.0000	62449.0000
3	2.0000	66233.0000
4	3.0000	68449.0000
5	4.0000	70461.0000
6	5.0000	71139.0000
7	6.0000	72103.0000
8	7.0000	72249.0000

FIT NO. 1

TRANSFORMATION CARD

CODE	CONSTANT	PASS NO.
0	0.0	1

INITIAL ESTIMATE OF R= 0.7177

ITERATION NO.	A	B	R	SUM(E(Y)-MEAN(Y))**2
1	74264.562500	-15290.085938	0.727780	987381.062500
2	74251.625000	-15267.148438	0.727397	986923.187500
3	74253.812500	-15267.964844	0.727482	986928.562500
4	74253.562500	-15265.289063	0.727439	986689.812500
5	74253.062500	-15268.152344	0.727471	986876.[illegible]
6	74252.125000	-15265.660156	0.727416	986860.312500
7	74253.812500	-15266.246094	0.727450	986886.625000
8	74253.500000	-15267.101563	0.727429	986910.125000
9	74252.687500	-15262.839844	0.727450	986913.437500
10	74253.375000	-15267.550781	0.727450	986921.750000

```
INFORMATION MATRIX

*
*         8.00000000        3.38132381        9.24204063  *  Y(0)=     542410.00000000
*                                                         *
*         3.38132381        2.11090660        3.11787796  *  Y(1)=     218851.31250000
*                                                         *
*         9.24204063        3.11787796       12.53656754  *  Y(2)=     638656.81250000
*                                                         *
INVERSE OF INFORMATION MATRIX

*
*         3.23262119       -2.62164307       -1.73082352  *  Y(0)=     542410.00000000
*                                                         *
*        -2.62164307        2.87486553        1.21751118  *  Y(1)=     218851.31250000
*                                                         *
*        -1.73082352        1.21751022        1.05277443  *  Y(2)=     638656.81250000
*                                                         *

    FINAL       STANDARD
    ESTIMATES   DEVIATIONS

A= 74253.3750     798.7920

B=-15267.5508     753.2949

R=      0.7274       0.0299

                         ANALYSIS OF VARIANCE

DEVIATIONS                            SUM OF SQUARES
*******************************************
FROM MEAN                             *************
FROM X(I) MEANS                              0.0

FROM CURVE                              986921.7500
OF X(I) MEANS FROM CURVE                986921.7500
*******************************************
```

TABLE OF RESIDUALS

NO.	X VALUE	Y VALUE	Y PREDICTED	RESIDUAL	(RESIDUAL)^2/(Y PREDICTED)
1	0.0	59327.0000	58985.8242	341.1758	1.9734
2	1.0000	62449.0000	63146.9961	-697.9961	7.7153
3	2.0000	66233.0000	66174.0000	59.0000	0.0526
4	3.0000	68449.0000	68376.0625	72.9375	0.0778
5	4.0000	70461.0000	69977.8750	483.1250	3.3855
6	5.0000	71139.0000	71143.1875	-4.1875	0.0002
7	6.0000	72103.0000	71990.8750	112.1250	0.1746
8	7.0000	72249.0000	72607.5000	-358.5000	1.7701

$\chi^2 = 15.0995$

GRAPH CODES A=AVERAGED Y P=PREDICTED Y B=BOTH

X AND Y VALUES ARE PLOTTED IN TRANSFORMED UNITS