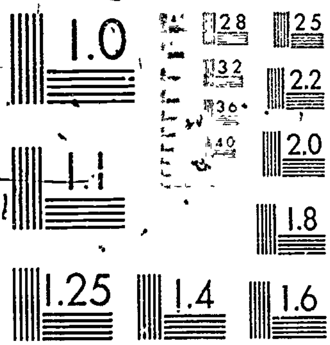

MICROCOPY RESOLUTION TEST CHART

DOCUMENT RESUME

ED 106 930 EA 007 122

TITLE Projection Techniques for the Non-Statistically Inclined. Research Report No. 113.

INSTITUTION Florida State Dept. of Education, Tallahassee. Div. of Elementary and Secondary Education.

REPORT NO RR-113

PUB DATE Sep 74

NOTE 25p.; Reprint

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE

DESCRIPTORS Data Analysis; Graphs; *Predictive Measurement; *Research Methodology; *School Statistics; *Statistical Analysis; *Statistical Data

ABSTRACT

This report briefly describes and analyzes several of the most frequently used techniques for depicting trends and making projections of education statistics. Each technique is described simply and nontechnically, with its uses and shortcomings, and a step-by-step analytical and graphic example and explanation. Projection techniques described include such methods as the freehand, semiaverage, average of period, moving average, least squares, ratio, and cohort-survival. (Author/JG)

Research Report

Number 113 – REPRINT

DIVISION OF ELEMENTARY AND SECONDARY EDUCATION

PROJECTION TECHNIQUES for the NON-STATISTICALLY INCLINED

SEPTEMBER 1974

STATE OF FLORIDA DEPARTMENT OF EDUCATION

DEPARTMENT OF EDUCATION
Tallahassee, Florida
RALPH D. TURLINGTON, COMMISSIONER

Research Report 113 is a new concept in the report series and is designed to provide districts and community colleges with methods for extrapolating base line data. A companion report, Research Report 114 will provide the historical data, where available, to facilitate the projections for each.

This report was designed and prepared by the Research Information and Surveys Section of the Bureau of Research and Information, Division of Elementary and Secondary Education, Department of Education. Inquiries regarding the Research Report should be addressed to James A. Kemp, Educational Consultant, Research Information and Surveys, 409 Knott Building, Tallahassee, Florida 32304 (450)

This public document was promulgated at an annual cost of $149.44 or $.23 per copy to assist school districts in developing base line data for needs identification and planning.

State of Florida
Department of Education
Tallahassee, Florida
Ralph D. Turlington, Commissioner

PROJECTION TECHNIQUES FOR THE NON-STATISTICALLY INCLINED

I. Introduction

The value of interpolating the future of some element in the universe based on the assessment of past and existing conditions, is obvious. Too often, however, the potential for insight into a problem is not attained due, in part, to a lack of understanding of basic projection techniques.

Three serious misconceptions regarding projection techniques are prevalent. First, it should not be assumed that all methods are difficult. Although there are many which are best handled by a computer, several require little more than paper and pencil, and the rudiments of basic algebra. Some of the more useful of these methods will be delineated later.

Second, it should not be assumed that because some standard technique has been utilized, that all or any conclusions derived therefrom will be infallable. Projections are merely estimates and as such can never be more accurate than the data from which they were obtained. Environmental conditions impinging upon the variable to be forecast can significantly alter the degree of accuracy. The most accurate predictions occur when these outside conditions vary little from the expected or the norm over a selected period of time.

Third, it should not be assumed that all methods of projection will generate identical information concerning a specified event, even though all raw data may have been identical. These discrepancies may be linked to the degree of rigorousness of the technique used. In general, the more rigorous the projection technique, the higher the probability that the resultant information will be less contaminated.

The simple methods of projections outlined in this report, although not difficult to compute, are not without merit. They are calculated quite rapidly and under fairly stable conditions serve quite adequately.

II. Time Series

Introductory discussions about projection techniques must also address time series upon which data the computations will be made. A time series is a representation of some variable over any given length of time. When this variable is represented statistically, its analysis is possible.

In general, there are four basic patterns which influence time series: (1) long-term/or basic trends, (2) seasonal fluctuations, (3) cyclical variations, and (4) irregular fluctuations. The characteristics of each of these must be inspected in order to understand the nature of possible discrepancies.

Long-term or basic trends involve relatively lengthy periods of time relative to the duration of the phenomenon under study. Such statistical data plotted on a graph would reveal a comparatively smooth pattern with no sudden reversals or changes. Depending upon the type of graph used, the trend line may be relatively straight or may gradually curve. In projecting variables based on long-term trends it is assumed that the environmental elements which effect changes in the specified variable will remain stable.

Seasonal fluctuations are controlled by two primary factors: climatic variations and local customs. That climatic fluctuations influence trends is easily comprehendible. The latter factor is less obvious. Customs vary from nation to nation, and from region to region. Included in the term "customs" would be holidays and religious influences, among others.

Cyclical variations are those which follow a definite pattern but which are not bound by a calendar. Such cycles may be several years in duration. Ideally these cycles should be of (near) identical length, but in reality external forces often influence it, causing consecutive cycles of uneven length or magnitude. The erratic length in cyclical variations is not as acute, however, when the variations are viewed from the perspective of the much larger long-term trend. A number of cyclical patterns of consequence have been identified such as the Julliard (10-year) and the 37-month business cycles, and various weather cycles.

Irregular fluctuations are single or multiple, unique deviations from that which has been identified as normal. Although usually isolated both in time and in space from one another, a seccession of unique elements can contribute significantly to any trend, especially as the paramenters controlling time and space are increasingly restricted.

All four influences co-exist under most circumstances. In those situations in which one or more irregular features dominate the contributions of the other factors, the trend will become increasingly less reliable with the frequency and magnitude of the fluctuations.

III. Techniques of Projection

Presented here are simple methods of predicting future values of a desired variable. The description of these techniques have been kept as basic as possible. In general, the techniques are presented in increasing order of difficulty.

Freehand Method. Like the other methods described below, this technique is applicable only in comparing one variable with one other (i.e., it is two dimensional). Data must first be arranged in some specified order, e.g. chronologically. Next, this must be plotted on a graph and the consecutive points connected by straight lines. A smooth curve may then be drawn along that imaginary line which the eye perceives as fitting the data the best. (See figures 1 & 2.) One definite advantage of the freehand method is that the line of interpolation may be a curve; the other methods to be outlined will necessarily be straight line methods. The extention of the curve past the last data point represents future predicted values.

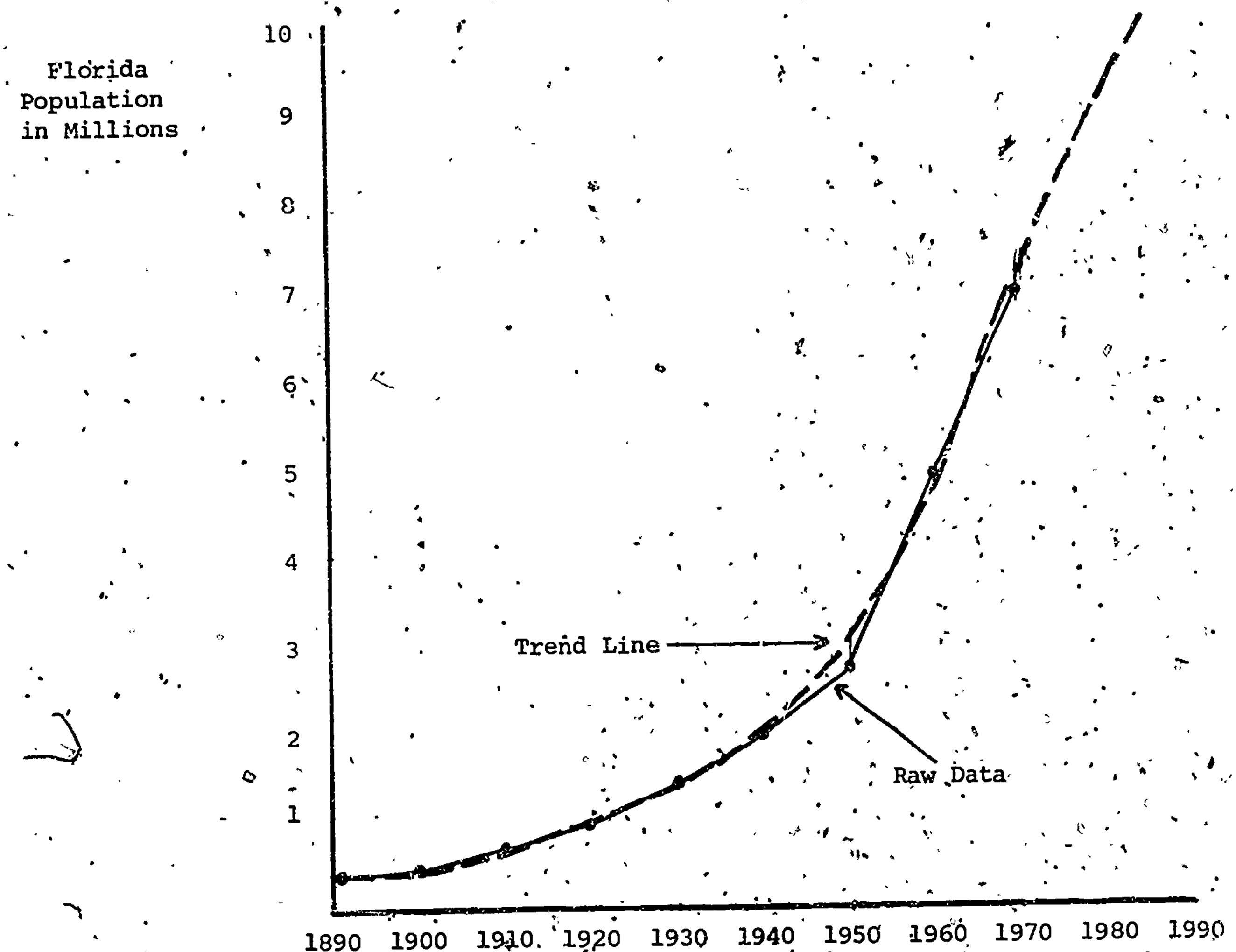

FIGURE 1. FREEHAND METHOD

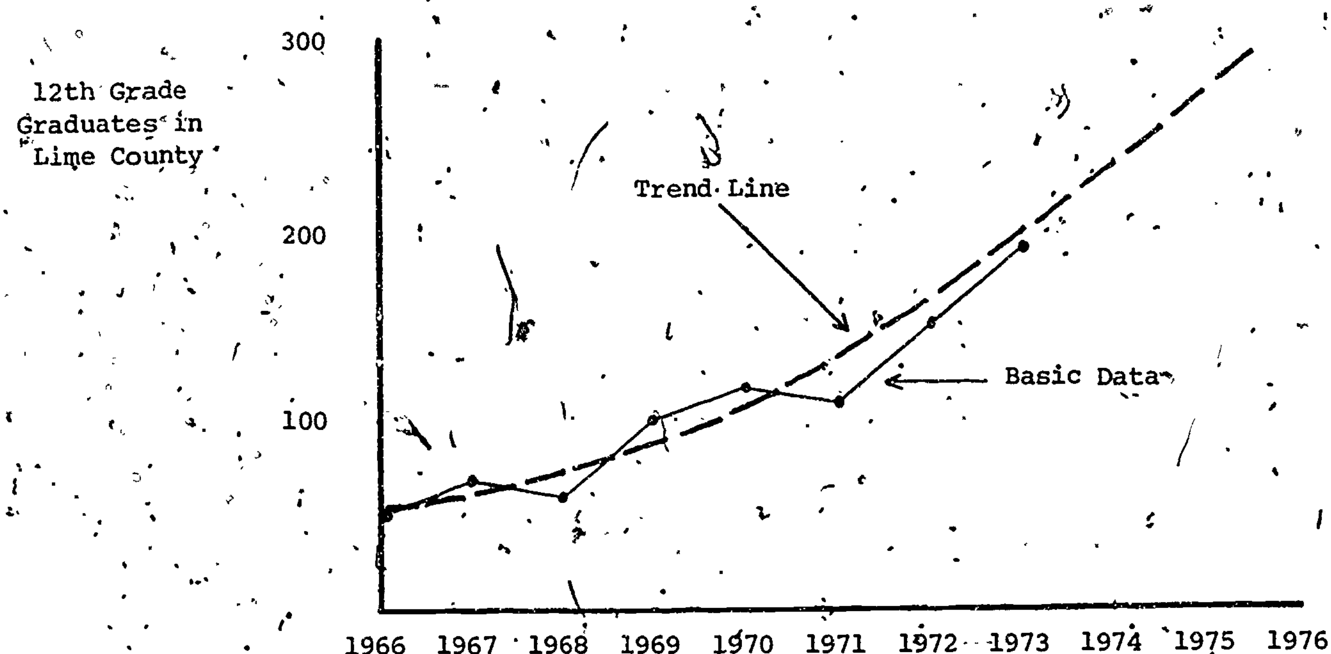

FIGURE 2. FREEHAND METHOD

Semi-Average. This method of projecting a trend involves basic mathematics. It is extremely fast to calculate and is quite satisfactory when it has been determined that the trend is linear.

The original data must first be arranged in some specified order and then plotted on a graph, consecutive points connected by straight lines. The trend period (horizonal ordinate) is divided into two equal parts and the arithmetic mean value of the variable (verticle ordinate) is calculated for each. Any extremely divergent values of the variable may be omitted from this computation. This trend line will be more representative of the long-term trend than it would have had the erratic data been included. The two average values are then plotted at the midpoints of each period (see "X's" in fig. 3 at coordinates 1942½,800 and 1962½, 1600) and a straight line is drawn between them and extending to either side. The line extending to the right extrapolates the predicted, future values of the variable. (See Figure 3.)

Example. Assume that it is desirable to predict the number of new residents in a particular county.

Step 1. Arrange the data chronologically.

Year	Number of New Residents
1935	400
1940	900
1945	1100
1950	2500
1955	1200
1960	1500
1965	1800
1970	1900

Step 2. Plot the data on a graph and interconnect the points by short straight lines (Figure 3).

Step 3. Divide the data into two equal chronological periods. Period 1: 1935, 1940, 1945, 1950; Period 2: 1955, 1960, 1965, 1970.

Step 4. Determine the average of the variable for each of the two periods, eliminating from the computations any data which is extremely high or extremely low. Period 1: $\frac{400 + 900 + 1100}{3} = 800$, Period 2: $\frac{1200 + 1500 + 1800 + 1900}{4} = 1600$.

Step 5. Plot the two averages at the midpoint of each half. Point 1: (1942½, 800); Point 2: (1962½, 1600).

Step 6. Draw a straight line between these two points and extending to either side. This is the trend line. The extension of the trend line beyond the last data point gives the predicted values of the variable.

Average of Period. This method is very similar to the last, the main difference being the number of periods to be averaged to establish the trend. The Average of Period method is slightly more sensitive than the semi-average method, but like the semi-average is useful only for linear trends. (See Figure 4; again computed averages for each period are marked by "X's").

As with the semi-average, this method of extropolation has little to recommend it over the free-hand method.

Example. Assume that it is desirable to predict the number of new residents in a particular county. (Compare this method with the semi-average, above.)

Step 1. Arrange the raw data chronologically.

Year	Number of New Residents
1935	400
1938	900
1941	800
1944	900
1947	1200
1950	2500
1953	1500
1956	1300
1959	1500
1962	1500
1965	1800
1968	1800

Step 2. Plot the data on a graph, connecting each point with the next by a straight line.

Step 3. Divide the data into several periods of equal duration, e.g., 9 years. Period 1: 1935, 1938, 1941; Period 2: 1944, 1947, 1950; Period 3: 1953, 1958, 1959; Period 4: 1962, 1965, 1968. (Note that each period begins 1½ years before the first date given and extends 1½ years beyond the last date given).

Step 4. Compute the mean value of the variable for each period eliminating any extremely high or extremely low value. Period 1: $\frac{400 + 900 + 800}{3} = 700$; Period 2: $\frac{900 + 1200 \quad 2500}{3} = 1533$; Period 3: $\frac{1500 + 1300 + 1500}{3} = 1433$; Period 4: $\frac{1500 + 1800 + 1800}{3} = 1700$.

Step 5. Plot each average at the midpoint of the period. Point 1: 1938, Period 2: 1947; Period 3: 1956; Period 4: 1965.

Step 6. Connect each point by a short straight line. Use the straight line between the last two averages (i.e., the last two "X's" in Figure 4) as the line of extrapolation for future values.

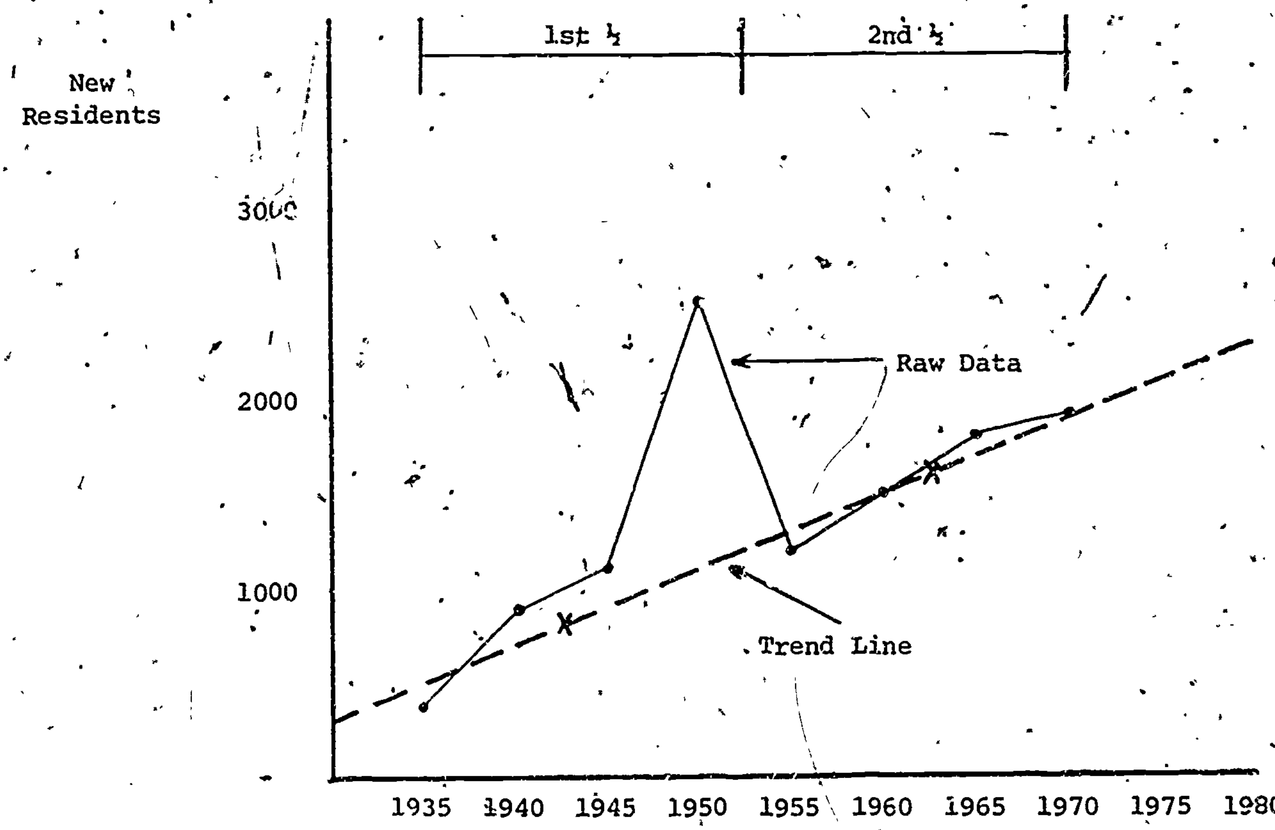

FIGURE 3. SEMI-AVERAGE

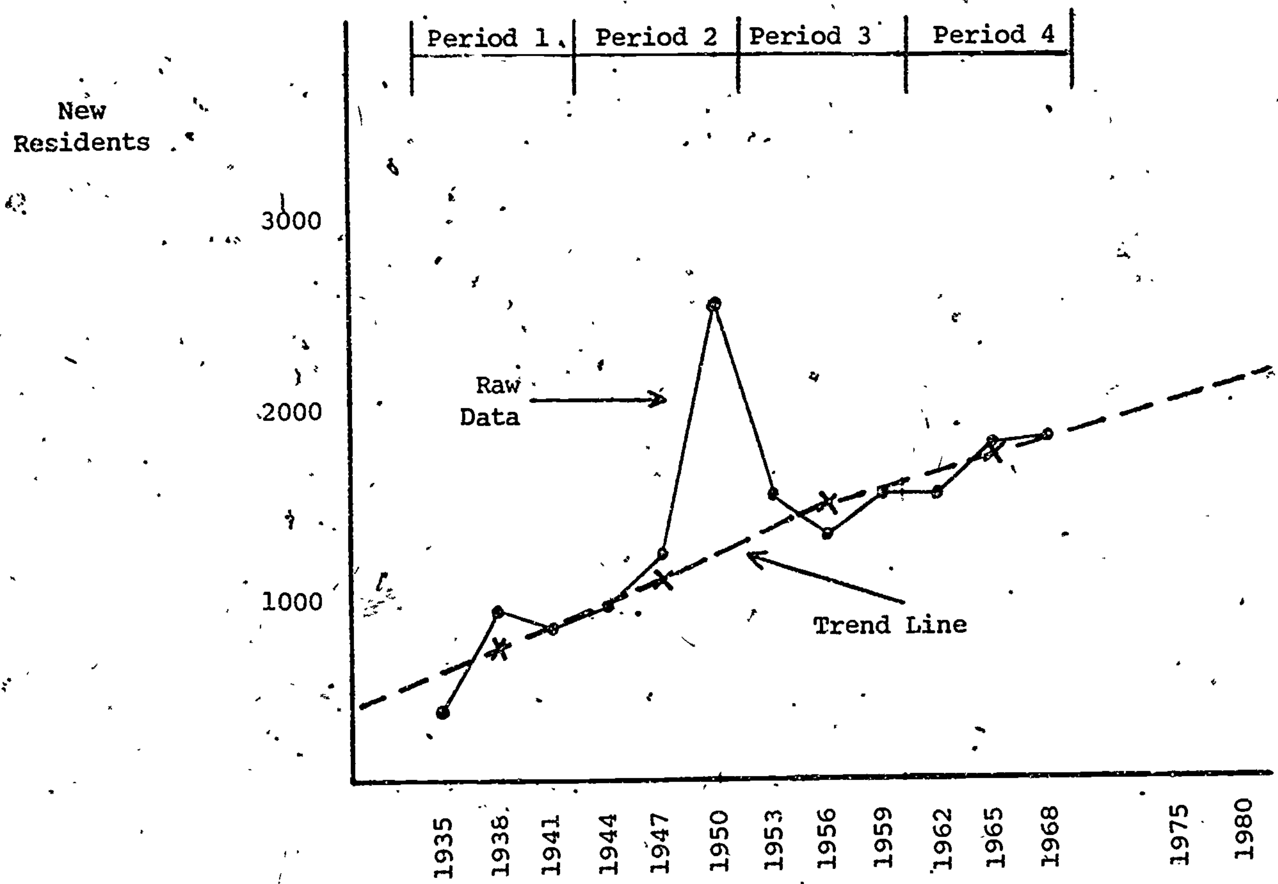

FIGURE 4. AVERAGE OF PERIOD

Moving Average. Like the semi-average and average of period, this method utilizes a series of averages to establish a trend and to extrapolate future values of a given variable. However, unlike the previous two methods, the Moving Average exmploys overlapping periods and averages. This allows the trend to be more sensitive to change.

In this method, data is again arranged in a specified order and plotted on a graph. The total duration of the variable being studied must then be divided into smaller components which will be grouped into overlapping sets and averaged. For example, assume it is discovered that annual enrollment is increasing. This data is then graphically plotted. The components are years and the chosen

-1- Fiscal Year	-2- Increase in Enrollment	-3- 3 Year Average
1960	55	-
1961	50	-
1962	57	54.0
1963	50	52.3
1964	62	56.3
1965	75	62.3
1966	100	79.0
1967	140	105.0
1968	97	112.3
1969	90	109.0
1970	91	92.7
1971	87	89.3
1972	85	87.7
1973	88	86.7

set is three years. (See figure 5.)

Obtain the average for each overlapping, three-year period (See Column 3, above). Plot this data at the midpoint of each period and connect the points by short straight lines. To determine future values extend the last short line beyond the period of average.

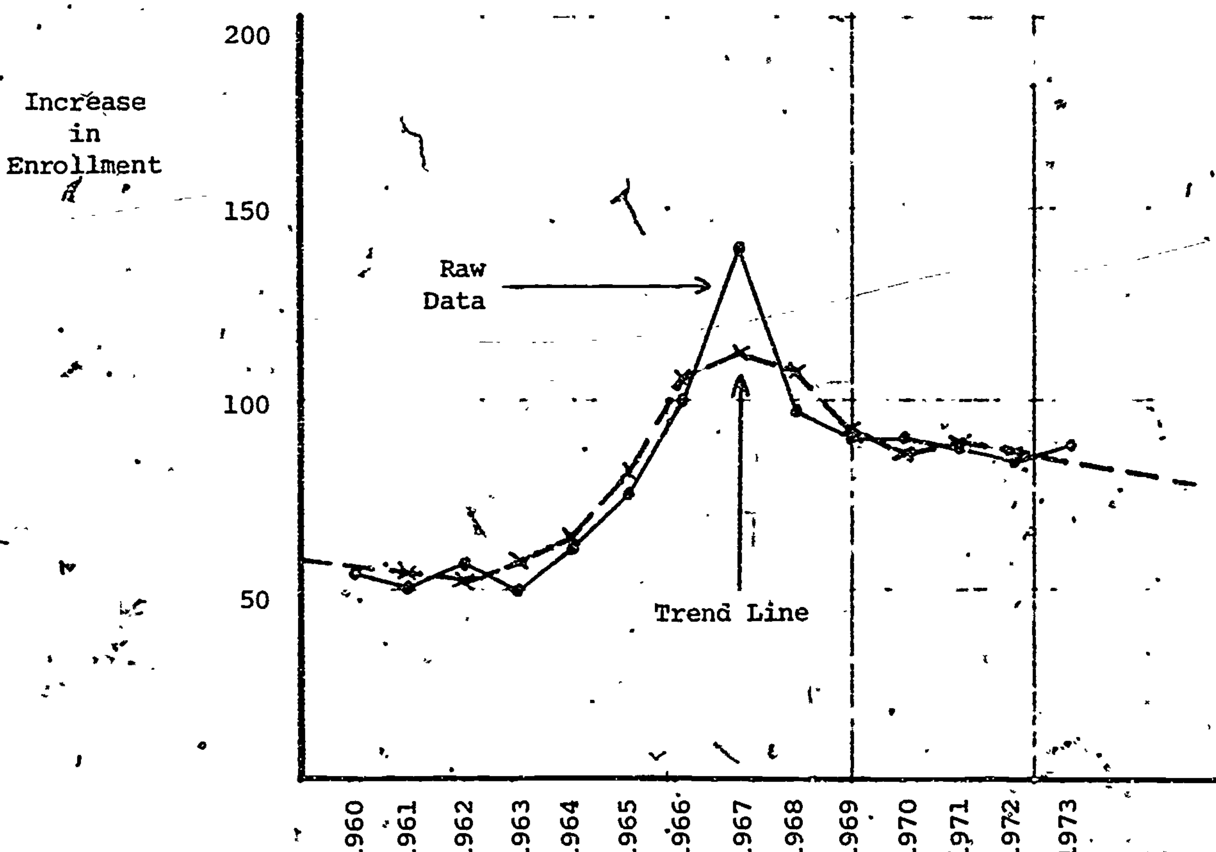

FIGURE 5. MOVING AVERAGE

Least Squares Method. This method may be used for both straight and curved trends. It also forms the basis for linear regression. The least squares technique is a method for fitting a line so that the sum of the squares of the deviations of the variable above and below the line will be a minimum. The general process for least squares is outlined below. A detailed explanation of each step is found in the example.

Data must first be arranged in some specified order, e.g., chronologically. (See Columns 1 & 2, page 16.) Compute the mean of the variable (See Column 2). Next the deviations (or differences) from the midpoint are determined. In this instance the midpoint is a year since time is the independent variable and the variable to be predicted depends upon the passage of time. (Column 3, page 16). Square the deviations (Column 4, page 16). Multiply the variables in Column 2 by the deviations in Column 3. Obtain the totals of the squared deviations and the variable multiplied by the deviations. Divide the second total by the first. This number gives the amount by which the variable in Column 2 increases - on the average - from year to year. The graphic ordinate of the dependent variable (Column 6, page 16) is computed by adding to the mean of Column 2 (for each year), the product of the deviation (for each year) and the average annual increment. See Figure 6 for a graphic representation.

Example. Assume that it is desirable to predict the total number of blind students in the district.

Step 1. Collect data for the last few years showing the total number of blind students in the district in each year. Data for at least four (4) years should be used.

Step 2. Arrange this data chronologically.

Year	Number of Blind Students (Variable)
1960	16
1961	40
1962	30
1963	47
1964	55
1965	23
1966	41
1967	69
1968	60
1969	73

Step 3. If the number of years for which you have collected data is even, leave a space between the middle two years and insert a small "dash" in the year column and the variable column. Since the example includes a ten-year period 1960-1969, these "dash" marks are inserted between 1964 and 1965.

Year	Number of Blind Students
1964	55
-	-
1965	23

Step 4. Add the number of blind students for each year (16 + 40 + 30 + 47 + 55 + 23 + 41 + 69 + 60 + 73 = 454) and divide this total by the number of years in the sample ($\frac{454}{10}$ = 45.4). This gives the average number of blind students in the district over the ten-year period.

Step 5. Determine the middle of the time period of the sample. If the number of years in the sample is even, then this point will fall between the middle two years (1964 and 1965). If the number of years in the sample is odd, then the middle year would be chosen.

Step 6. The deviation is the distance in time each year is from the "middle". In this example this middle lies between two years, so each year will deviate by some number plus or minus .5. (This example is true for any sample containing an even number of years. If the example had an odd number of years, then each deviation would be a whole number.) The deviations of the years prior to the midpoint are preceeded by a minus sign, while those following the midpoint are preceeded by a plus sign.

Step 7. Make a third column labeled "deviation". Place an "0" at the midpoint and insert the deviation for all other years.

Year	(Variable) Number of Blind Students	Deviation
1960	16	-4.5
1961	40	-3.5
1962	30	-2.6
1963	47	-1.5
1964	55	- .5
-	-	0
1965	23	+ .5
1966	41	+1.5
1967	69	+2.5
1968	60	+3.5
1969	73	+4.5

Step 8. Square each deviation in column 3 and enter these numbers in a fourth column.

Year	(Variable) Number of Blind Students	Deviation	Squared Deviation
1960	16	-4.5	20.25
1961	40	-3.5	12.25
1962	30	-2.5	6.25
1963	47	-1.5	2.25
1964	55	- .5	.25
-	-	0	0
1965	23	+ .5	.25
1966	41	+1.5	2.25
1967	69	+2.5	6.25
1968	60	+3.5	12.25
1969	73	+4.5	20.25

Step 9. Add the "squared deviations column". (20.25 + 12.25 + 6.25 + 2.25 + .25 + .25 + 2.25 + 6.25 + 12.25 + 20.25 = 82.50).

Step 10. For each year multiply the number of blind students in the district by the deviation. Enter the answer in a new column.

Year	(Variable) Number of Blind Students	Deviation	Squared Deviation	Col. 2 x Col. 3
1960	16	-4.5	20.25	- 72.0
1961	40	-3.5	12.25	-140.0
1962	30	-2.5	6.25	- 75.0
1963	47	-1.5	2.25	- 70.0
1964	55	- .5	.25	- 27.0
-	-	0	0	0
1965	23	+ .5	.25	+ 11.5
1966	41	+1.5	2.25	+ 61.5
1967	69	+2.5	6.25	+172.5
1968	60	+3.5	12.25	+210.0
1969	73	+4.5	20.25	+328.5

73 (blind students) x 4.5 (deviation) = 328.5

Step 11. Total all values obtained in Step 10 (column 5). Divide this number by that obtained in Step 9: $\frac{+399.0}{82.5} = 4.84$. The value 4.84 is the average annual increment of the variable.

Step 12. Label the next column "graphic ordinates". When plotting the information on a graph, the data in this column along with that in column 1 will mark the points through which the trend line will pass. To obtain the values in this column, for each year in this example, multiply the increment (Step 11) by the deviation and add this product to the average number of blind students (Step 4). For the year 1960 we would have: (4.84) (-4.5) + 45.4 = 23.62.

-1- Year	-2- Variable	-3- Deviation	-4- Squared Deviation	-5- Col.2 x Col.3	-6- Graphic Ordinates
1960	16	-4.5	20.25	- 72.0	23.62
1961	40	-3.5	12.25	-140.0	28:46
1962	30	-2.5	6.25	- 75.0	33.30
1963	47	-1.5	2.25	- 70.5	38.14
1964	55	- .5	.25	- 27.5	42.98
-	-	0	0	0	45.40
1965	23	+ .5	.25	+ 11.5	47.82
1966	41	+1.5	2.25	+ 61.5	52.66
1967	69	+2.5	6.25	+172.5	57.50
1968	60	+3.5	12.25	+210.0	62.34
1969	73	+4.5	20.25	+328.5	67.18
			82.50	+399.0	

MEAN = $10\overline{)454}$ = 45.4

INCREMENT = $\frac{399.0}{82.5}$ = 4.84

Step 13. Plot the raw data on a graph and connect the points by solid, straight lines. Enter the coordinates obtained in Steps 1 through 12 (column 1 and 6) on the graph and connect all these points by a broken line. This line should be perfectly straight. If this broken line is extended beyond the last data point, it then represents the line of predicted values.

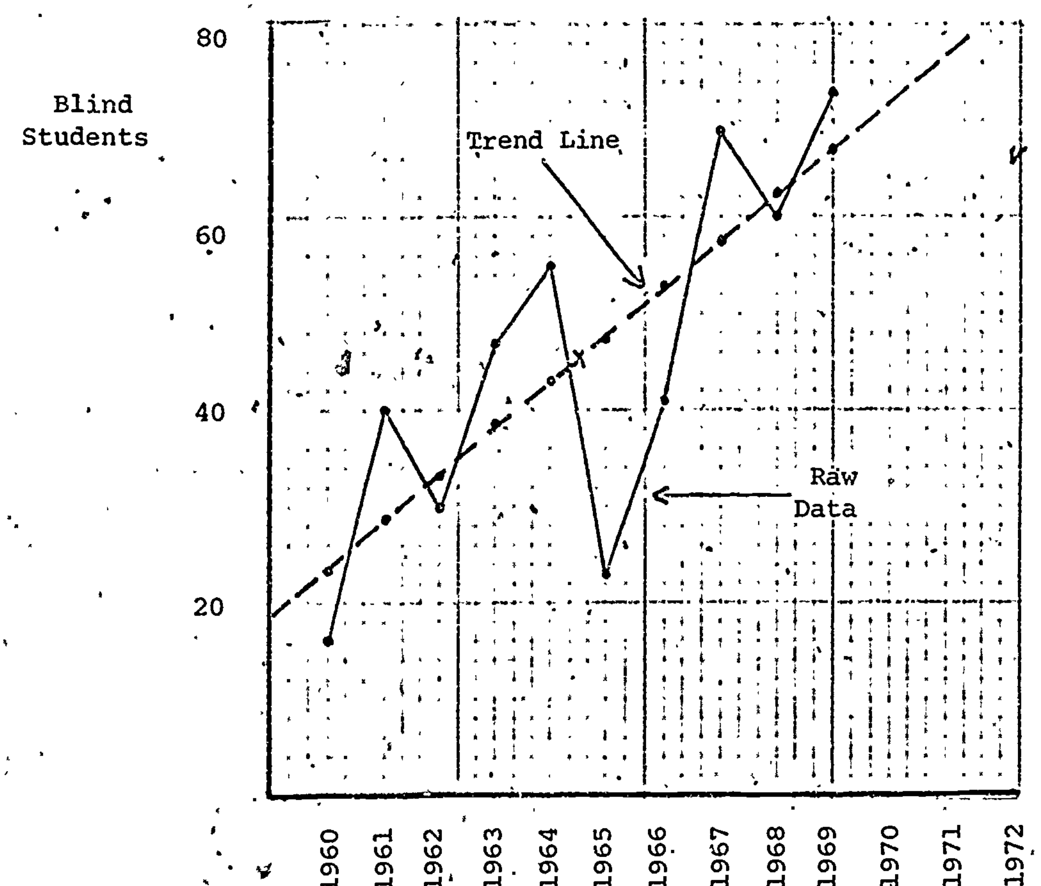

FIGURE 6. LEAST SQUARES

Ratio Method. This method is widely used, but often is inferior to the last method to be outlined in this paper, the Cohort Survival Technique. Its utility, however, is that it allows for a very rapid calculation of an approximate future value of a given variable.

In its most basic form a predicted value of a variable may be calculated by dividing past values of the same variable by the total population from which it was taken or by some other variable which has been shown to correlate very highly and multiplying this by the total population (or related variable) of future years.

For example, assume that it is desirable to know the number of new students a district can anticipate in the fall. Past records have shown that for every 100 new residential telephones installed between May and August 15th, 34 children will enter the public school system in the fall. The ratio method assumes that this relationship will continue unchanged. Therefore, if the local telephone company records show that 275 new residential connections have been made, the estimated number of new students would be: $\frac{34}{100} \times 275 = 93.5$.

In problems dealing with the school population as a fraction of the age pool, each age would be weighted. Although this makes the estimate more accurate, it increases the complexity of the calculations to the point that this method has nothing to offer that the Cohort-Survival Technique cannot offer more accurately.

Cohort-Survival Techniques. This group of closely related methods is based upon the extent to which a particular phenomenon or groups of individuals can survive through a sequence of pre-determined steps (e.g., grades 1, 2, 3, etc.). This method, as opposed to several of the previous ones, does not lend itself to graphic prediction, but rather is a succession of mathematical ratios.

The easiest way to explain the method is through an example. Assume that it is desirable to predict the future public school average daily membership by grade in Lime County.

Step 1. Obtain the birth statistics for the preceeding 10 years. Obtain the average daily membership statistics for the current and preceeding five years for grades 1 - 12 and arrange this in chronological order.

BIRTH DATA

1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
253	247	229	179	204	189	201	163	216	181	219

ADM FOR GRADE – 12

Year	67-68	68-69	69-70	70-71	71-72	72-73
Grade						
1	252	268	238	195	149	173
2	230	226	256	196	219	174
3	278	239	224	224	179	223
4	250	266	239	196	227	186
5	279	270	263	197	197	193
6	207	249	260	239	204	203
7	246	195	267	246	230	217
8	260	245	196	225	233	236
9	192	243	227	160	216	250
10	172	166	217	188	157	180
11	175	167	151	169	176	141
12	152	156	146	131	91	146
Total K-12	2693	2690	2684	2366	2278	2322

Step 2. To calculate the survival ratio for first grade, total the number of resident births for the five-year period 1962-66. Now find the total ADM for first grade from 1968-1969 through 1972-1973. These are the students who were enrolled in first grade six years later. Divide the total number of first grade students by the total number of births: $\frac{1023}{1112} = .92$. The figure .92 is the average survival ratio of resident births to 1st graders.

Step 3. To estimate the future enrollment in first grade, multiply the number of resident births for a given year by .92. This will give the approximated first grade enrollment six years later.

BIRTH DATA

1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
253	247	229	179	204	189	201	163	216	181	219

ADM FOR GRADES 1-12

67-68	68-69	69-70	70-71	71-72	72-73	73-74	74-75	75-76	76-77	77-78	78-79

SURVIVAL RATIO

	Known							Predicted					
Grade 1	252	268	238	195	149	173	.92	174	185	150	199	167	201

x =

201 (births in 1968) x .92 (survival ratio) = 185 (predicted 1st grades in 1974-75).

Step 4. To calculate the survival ratio for any two consecutive grades, add the ADM for 5 consecutive years (e.g., 1967-68 through 1971-72, inclusive) for the lower of the two grades. Add the ADM for 5 consecutive years for the upper of the two consecutive grades beginning 1 year later (e.g., 1968-69 through 1972-73, inclusive). Divide the second total by the first. In this example the survival ratio for second grade would be

$$\frac{226 + 256 + 196 + 219 + 174}{253 + 268 + 238 + 195 + 149} = \frac{1071}{1102} = .97.$$

Step 5. To estimate the future enrollment for any grade, multiply the survival ratio for the grade by the number of students in the next lower grade one year before.

ADM FOR GRADES 1-12

Year	67-68	68-69	69-70	70-71	71-72	72-73	SURVIVAL RATIO	73-74	74-75	75-76	76-77	77-78	78-79
Grade													
1	252	268	238	195	149	173	.92	174	185	150	199	167	201
2	230	226	256	196	219	174	.97	168	169	180	146	193	162
3	278	239	224	224	179	223	.97	168	162	163	174	141	187
4	250	266	239	196	227	186	.97	217	164	158	159	169	137
5	279	270	263	197	197	193	.95	177	206	156	150	151	161
6	207	249	260	239	204	203	.96	185	169	198	149	144	145
7	246	195	267	246	230	217	1.00	202	184	169	197	149	144
8	260	245	196	225	233	236	.96	208	194	177	162	189	142
9	192	243	227	160	216	250	.95	223	197	183	167	153	179
10	172	166	217	188	157	180	.87	219	195	172	160	146	134
11	175	187	151	169	176	141	.89	161	195	174	154	143	130
12	152	156	146	131	91	146	.80	113	129	158	139	123	115
							X	=					

194 (1st grades in 1973-74) x .97 (survival ratio) = 169. (2nd grades in 1974-75).

In the Cohort-Survival method, errors appear to be cyclical which will necessitate the yearly revision of the ratios. The following table gives the complete data for the Lime County example. Note that in this projection the survival ratio was computed to four decimal places and rounded to two (2) in this table. This accounts for all discrepancies which may be encountered.

COHORT SURVIVAL PROJECTION
Lime County

BIRTH DATA

1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
253	247	229	179	204	189	201	163	216	181	219

ADM FOR GRADES 1-12

YEAR	67-68	68-69	69-70	70-71	71-72	72-73	Survival Ratio	73-74	74-75	75-76	76-77	77-78	78-79
GRADE													
1	252	268	238	195	149	173	.92	174	185	150	199	167	201
2	230	226	256	196	219	174	.97	168	169	180	146	193	162
3	278	239	224	224	179	223	.97	168	162	163	174	141	187
4	250	266	239	196	227	186	.97	217	164	158	159	169	137
5	279	270	263	197	197	193	.95	177	206	156	150	151	161
6	207	249	260	239	204	203	.96	185	169	198	149	144	145
TOTAL 1-6	1496	1518	1480	1247	1175	1152		1089	1056	1005	977	965	993
7	246	195	267	246	230	217	1.00	202	184	169	197	149	144
8	260	245	196	225	233	236	.96	208	194	177	162	189	142
9	192	243	227	160	216	250	.95	223	197	183	167	153	179
TOTAL 7-9	698	683	690	631	679	703		633	575	529	526	490	465
10	172	166	217	188	157	180	.87	219	195	172	160	146	134
11	175	167	151	169	176	141	.89	161	195	174	154	143	130
12	152	156	146	131	91	146	.80	113	129	156	139	123	115
TOTAL 10-12	499	489	514	488	424	467		492	519	503	454	412	379
TOTAL 1-12	2693	2690	2684	2366	2278	2322		2215	2150	2036	1956	1868	1836
TOTAL K-12	2869	2885	2880	2498	2425	2472		2363	2270	2195	2089	2029	