8-1950

Statistical Analysis and Forecast of the University of Tennessee Enrollment

Edward J. Boling
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a thesis written by Edward J. Boling entitled "Statistical Analysis and Forecast of the University of Tennessee Enrollment." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Statistics.

R. M. Deepree, Major Professor

We have read this thesis and recommend its acceptance:

David B. Chambers, O. M. Suskind

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

August 4, 1950

To the Committee on Graduate Study:

I am submitting to you a thesis written by Edward J. Boling entitled "Statistical Analysis and Forecast of The University of Tennessee Enrollments." I recommend that it be accepted for nine quarter hours of credit in partial fulfillment of the requirements for the degree of Master of Science, with a major in Statistics.

Major Professor

We have read this thesis
and recommend its acceptance:

David B Chambers

Accepted for the Committee

E. A. Waters

Dean of the Graduate School

STATISTICAL ANALYSIS AND FORECAST OF THE UNIVERSITY OF TENNESSEE ENROLLMENT

A THESIS

Submitted to
The Committee on Graduate Study
of
The University of Tennessee
in
Partial Fulfillment of the Requirements
for the degree of
Master of Science

by

Edward J. Boling

August 1950

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

THE PROBLEM

Introduction

In recent years there has been a pronounced interest exhibited in higher education. This interest reached an unusually high peak in the United States during the years immediately following World War II. A great portion of the concern evidenced resulted from the mass enrollment in universities and colleges of veterans who through acts of Congress received free tuition and subsistence. Never before in the history of higher education were there so many persons enrolled, and never before were there so many problems facing university and college officials. Most of these problems involved attempts on the part of college officials to provide facilities for many more students than the school could normally accommodate.

With the influx of veterans plus the normal enrollments, it became quite evident that there was a drastic need for additional facilities in the colleges of the United States.[1] Not only was there a need for class room, but there was a distinct need for more residential buildings, larger administrative staffs and enlarged faculties. In short, universities and colleges were going to have to expand if they intended to

---

[1]Francis J. Brown, "The President's Commission on Higher Education--A View of Its Findings and Recommendations," American Association of University Professors Bulletin, 34:27, Spring 1948.

accommodate those who sought higher education; and the nation was aware of this acute situation.

The President of the United States added to this wave of interest when in July 1946 he appointed a President's Commission on Higher Education to consider specific problems in higher education. Some of these problems included:

> . . . ways and means of expanding educational opportunities for all able young people; the adequacy of curricula, particularly in the fields of international affairs and social understanding; the desirability of establishing a series of intermediate technical institutes; the financial structure of higher education with particular reference to the requirements for the rapid expansion of physical facilities.[2]

The Commission recognized that consideration of such problems "made necessary a determination of the potential needs of the Nation and the ability of our adult population to profit from higher education."[3] Accordingly the members of this Commission considered two approaches to the problem. One was an attempt to predict possible future enrollment on the basis of long-term trends or trends immediately following World War I. The second approach was that of endeavoring to determine national needs and the potential ability of the adult population. The Commission chose the latter alternative.[4]

More of the activities and reports of this Commission will be cited at a later stage in this study. Suffice it to say now that the vigor with

---

[2]George F. Zook, "The President's Commission on Higher Education Reports," Higher Education, 4:133, February 15, 1948.

[3]*Loc. cit.*

[4]*Loc. cit.*

which this investigating board went about its job brought about much action in higher education circles. Colleges became very interested in the history of their enrollments, with particular emphasis on the present and future enrollments. At least eight estimates of future college enrollments in the United States have appeared in professional literature since the war.[5] Many colleges have analyzed their past enrollment records with some hope of using knowledge gained to an advantage in estimating and preparing for future enrollments. Others have blandly avoided using any objectifying techniques, and have subjectively predicted enrollment figures without weighting or in some cases even considering the many variables that influence college enrollment.

Statement of the Problem

It is the problem of this investigation to analyze by the use of statistical methods the number of students enrolled at The University of Tennessee from 1899 to 1950, and with the aid of such an analysis to forecast possible future enrollments.

The data used in this investigation will be limited in that only the resident students of that portion of The University of Tennessee which is located at Knoxville, Tennessee, will be used. Thus, the Colleges, Schools and Experiment Stations located in Memphis, Martin, Jackson, Columbia, Greeneville, Nashville, Crossville and Springfield, and the

---

[5] J. Harold Goldthorpe, "Estimates of Future College and University Enrollments," Higher Education, 4:157, March 15, 1948.

University extension work carried on throughout the counties of the state are not to furnish data for this study. The analysis of the enrollment data will cover the period beginning 1899 and ending 1950 with particular emphasis being given to the data in the past two decades. With the aid of the results of the analysis, the possible future enrollment will be forecast through 1970. It is essential that, continuously, recognition be given to the vast number of unknown variables entering into a forecast of future college enrollment.

Definition of Terms

By the term "statistical methods" is meant "the use of the principles of scientific investigation in the study of aggregates of numerical information."[6] It may be recognized that a problem such as this is complicated in that conditions cannot be controlled as they are in the field of physics or chemistry. However, the statistician has developed methods of investigation which may be adjusted to the type of data with which he deals and to the uncontrolled conditions under which he must use them. In this study the methods followed will make use of many statistical devices such as averages, time-series analyses, charts, graphs, tables and correlation analyses. The devices will be used as tools to objectify the enrollment data much as a physicist and chemist use scales and thermometers.[7]

---

[6]Martin A. Brumbaugh and Lester S. Kellogg, Business Statistics (Chicago: Richard D. Irwin, Inc., 1946), p. 43.

[7]Frederick C. Mills, Statistical Methods (New York: Henry Holt and Company, 1924), pp. 3-7.

The use of the statistical approach in this study will not preclude the use of the case study method which does not deal with masses of data. It is essential that the enrollment data be studied intensively so that historical facts may be used and given weight in the analysis.

"Forecast" will be used in this study to mean a description of the future situation as regards enrollment. Since the forecast will be expressed in figures, it may be termed a definition of the future situation. A forecast making use of precise figures may be a source of misunderstanding because of the danger that those who use the forecast will be prone to think that it necessarily will be absolutely accurate since it is expressed in exact figures. Every forecast is wrong in the sense that actual events will not be identical with those described.[8] This forecast of enrollment will prove to be no exception; however, every attempt will be made to control the magnitude and direction of the error.

Importance of the Study

Knowledge of past, present and particularly the future enrollment at The University of Tennessee should prove most valuable to administrative heads who have the problem of preparing for the future. Some of the needs that may be anticipated with the knowledge of the number of students to be enrolled in future years are: size of faculty, class room facilities, housing facilities for both students and faculty, and financial needs in general.

---

[8]Wilson Wright, Forecasting for Profit (New York: John Wiley and Sons, Inc., 1947), p. 12.

Business men in the Knoxville area would be highly interested in an unbiased analysis of the enrollment figures and a scientific estimate of the number of students to be registered at the University in years to come. Such knowledge would influence to a great extent the building and expansion plans of merchants in the immediate vicinity of the school. Churches, amusement places and many other establishments in and around Knoxville are affected greatly by the number of students enrolled at the University.

This study should provide the Board of Trustees and the State Legislature with objective evidence of what has happened in the past as concerns the number of students enrolled at the University, and also an estimate of the number of students to be enrolled in the future. Such figures, though probably in error, should be welcomed as an unbiased attempt to arrive at approximate enrollment figures. Objectively determined estimates would aid state officials in determining the future needs of the University whereas in the past they have had no such estimate but have depended upon subjective estimates by school officials.

A study of this type should prove important as a basis for consideration of the present plan of the federal government to aid in higher education and the extent to which it should aid if at all. Certainly if federal aid is granted to colleges and universities, the forecast of enrollment will be greatly affected. Recent trends in enrollment may show, however, that higher education is becoming popular and available to the extent that federal aid is not desired. It will not be proper within the scope of this study to take issue for or against federal aid to higher education; nevertheless, the

importance of such a factor in making a corecast cannot be ignored.

If the value of a forecast is limited, the study should still prove of definite use in depicting what has happened in the past as concerns enrollment data. Such a picture of past enrollment history analyzed statistically is needed by school officials.

Review of Related Studies

Investigation reveals that there has never been a statistical analysis of the number of students enrolled at The University of Tennessee. Various attempts have been made to forecast enrollment from one year to the next, and many guesses have been made concerning the possible enrollment in the distant future. However, it is the problem of this study to analyze the data statistically and organize it into a scientific work.

Earlier mention was made of eight estimates of future college enrollment in the United States. It is not considered necessary to discuss the methods used in each of these studies. It is desirable, however, to discuss briefly the estimate arrived at by the President's Commission on Higher Education. This estimate has received much publicity and has been widely quoted as though it were a fact. In the introduction it was emphasized that the Commission had a choice of two courses to take in attempting to arrive at an estimate. First, it could predict possible future enrollment on the basis of trends, or secondly, it could try to determine national needs and the potential ability of the adult population. The latter course which resulted in a "National Inventory of Talent" was chosen.

Mr. G. F. Zook, Chairman of the Commission, briefly summarized the steps taken in the pursuit of this course:

> Through data supplied by the Bureau of Labor Statistics, the professional organizations and other groups, it became clear that present enrollment in colleges and universities could not meet the Nation's needs in the years ahead. This fact became all the more apparent in terms of the greater needs for understanding of our national and international problems on the part of an increasing number of our population.
>
> The second variable in the formulation of the statistical framework was the ability and interest in higher education of potential students. The factor of interest could not be measured but it was possible to determine potential ability to profit from higher education. With the invaluable assistance of the Army and Navy a correlation was made between Army General Classification Test scores of 11 million men and years of schooling which they had completed. Through such a statistical analysis correlated still further with the results of college entrance examinations, it was found that 49 percent of young people 18 and 19 years of age had a reasonable expectation of completing at least the first 2 years of college; 32 percent might be expected to complete at least a full 4-year college course. When these percentages were applied to the group that will be of college age in 1960, the total number for whom college education is a reasonable risk is 4 million undergraduates.
>
> The number of graduate students was determined largely on the basis of national need and estimated by the Commission to be 600,000. Thus the numerical goal which the Commission established was the enrollment of 4,600,000 by 1960.[9]

This approach to the problem of estimating has been cited merely to show one of the methods used in arriving at probable future enrollments. It should be emphasized that this estimate was for the college and university enrollments for the whole nation and is definitely dependent on many factors. One of these factors is obviously the passage of a "federal aid to higher education" bill by Congress.

---

[9] Zook, op. cit., pp. 12-13.

Analysis of the Problem and Methods of Procedure

It would be a worthy task to try to estimate the future enrollment at the University of Tennessee by using the technique devised by the Commission and described above. It is believed, however, that a much more objective estimate will evolve from analysing the past enrollment data statistically, i. e., studying trends, seeking relationships, and with the aid of this analysis making a forecast.

Any attempt to break this investigation down into its elements must of necessity be general because of the nature of the problem. It is true that this problem offers many opportunities for the use of statistical methods in bringing out relationships, trends, averages, etc; however, the determination of what methods to use and when to use them can be made only after the analysis is under way. Essentially this thesis will entail three projects: collection of data, analysis of data and forecast of future enrollment based on the results of the analysis.

The collection of the data, though of prime importance, should not prove to be a difficult problem. The enrollment figures needed are on file in the Office of the Dean of Admissions and are available for reproduction. However, it is essential that the gathering of this data be preceded by a detailed study of the history of enrollment at the University. A study of this type tends to clarify terms and concepts which may have one or more meanings.

At this stage it is imperative that the different enrollment figures gathered by the Dean of Admissions Office which are pertinent to this study

be discussed and properly defined. First, records are kept which show the total number of students enrolled each quarter. As an example, the total number of students enrolled during the Summer quarter of 1948 was 4,916. For the Fall quarter of 1948 the enrollment was 7,790. The 1949 Winter quarter enrollment was 7,332 and for the Spring quarter the enrollment dropped to 6,987. These "total enrollment per quarter" figures will be used in an analysis of quarterly variation in Chapter III.

This concept of the total number of students enrolled per quarter is not to be confused with the total number of students enrolled in a regular session school year. The latter is a gross concept and actually means the total number of individuals enrolled at the University during the regular session where regular session means Fall, Winter, and Spring quarters. The regular session gross totals were used without inclusion of Summer session students not enrolled during regular session because:

(1) Regular session gross totals were the only homogeneous enrollment figures that could be obtained for years extending as far back as 1899.[10]

(2) Summer session students not enrolled during regular session are predominately special students, public school teachers taking special short courses, or similar persons whose presence during the Summer session from year to year is not indicative of University enrollment growth.

---

[10] In this thesis, the year given indicates the school year ending in June. For example, 1899 means the school year 1898-1899.

As an example of the difference between the two types of enrollment figures, the gross enrollment for the regular session 1948-1949 was 8,914. A comparison of this total enrollment figure with the total enrollment per quarter as discussed above for the same school year shows that never in any one of the quarters did the enrollment reach the gross figure, 8,914.

In the chapters to follow, the two types of enrollment figures will be distinctly labeled as above, i.e., "total enrollment per quarter," pertaining to students enrolled by quarters, and "gross enrollment," pertaining to the total number of individuals enrolled in the regular session.

Statistical texts and research manuals will serve as reference material when the statistical techniques are applied to the data. Such references may be obtained from library sources as may other data to be used in this study. State publications revealing data on educational statistics will be used when available. Data gathered by the Bureau of Census branch of the Department of Commerce will be used.

The analysis of the data will prove to be the major task in this investigation. In this phase all possible statistical methods of objectifying data will be brought into play with the intention of discovering hidden relationships between enrollment figures and other variables.

The tabulation of the data involves the orderly arrangement of the figures in vertical columns and horizontal rows. A great portion of the tabulation in this study is confined to classifying the enrollment figures by quarters into a time series. Graphs are useful methods of displaying numerical relationships instantly. When appropriately used,

graphic methods sometimes clarify relationships that remained obscure even after careful study of the numerical data. Particularly are graphs a popular method of presenting data to others. Hence, tables and graphs will be used in an effort to analyze the enrollment figures.

One of the major phases of the analyses will be an attempt to use known methods in scrutinizing the changes that have occurred in the series of enrollment figures with the passing of time. Such a project is known as a time series analysis. It is not enough to know merely that the enrollment at the University today is larger than at any other period in the history of the institution. It is not enough to know that the enrollment has increased over the period of years covered. There must also be the realization that there are many factors interacting, the net effect of which has been an increase in the series. In order to recognize the individual importance of such factors, it is necessary to segregate them and inspect them singly.[11]

The factors to be analyzed and measured are trend, season, cycle and irregular forces. All these are forms of variation which may be measured and viewed separately, but when synthesized make up the original data. The trend, for instance, is the tendency of the enrollment figures to increase or decrease over a long period of time. The recognition that trend does exist gives rise to two problems--(1) location of trend, and (2) measurement of trend. Neither of these problems may be solved adequately until a thorough knowledge of the background of the enrollment figures is gained.

---

[11] Brumbaugh and Kellog, op. cit., pp. 537-542.

Once the analyst is familiar with the data, there remains the task of fitting a trend line. In accomplishing this a subjective method such as "inspection" may be used. Equally feasible is the use of a method described by a mathematical equation. If one of the methods described by an equation, e. g., Gomperts, Pearl-Reed, appear to fit the data, there could be an argument for the projection of such a curve into the future and thus having a mathematical basis from which predictions could be made.

An effort will be made to identify cyclical movements, i. e., movements representing constantly recurring rises and declines in enrollment data. A further attempt will be made to obtain the periodicity of the cyclical movements with the intention of fitting a periodic curve to the data. If such a curve is fitted, it may be possible to project the cyclical movement into the picture and predict turning points. Such an extrapolation when combined with the trend extrapolation would provide an estimate of enrollment superior to one based on trend alone.

When the enrollment figures have been plotted against the quarters, it is probable that year after year the fall quarter will contain the largest enrollment and the summer quarter will contain the smallest enrollment. The time-series analysis will separate out and measure this more or less regular and recurring seasonal fluctuation.

As concerns irregular variation or those movements of a random or episodic character, it may well prove that such variations are the most important factor in this time-series analysis. An example of an irregular movement might be the decline in students during the war. This decline

was in no way accounted for by trend and season while cycle explains only a portion of it; consequently, it may be considered irregular and of a residual nature. Needless to say, such variations may not be predicted and can be accounted for only if the investigator is familiar with the history of the data.[12]

Another phase of the statistical analysis of this problem will involve the use of simple, and multiple correlation. These techniques will be used in an attempt to find a relationship between the number of students enrolled at the University and some other variable or variables, such as the number of students graduating from high schools in Tennessee the previous year. If such a relationship does appear to exist, it will be subjected to a regression test (F test) to see if the relationship is real or if there was a high probability that it could have happened by chance. If the relationship is of a functional nature and is of a degree sufficiently high, such knowledge should prove of great service in making the forecast. At this stage it is impossible to cite all the variables which may be considered in finding a relationship with the enrollment number. Although many of the variables are not known, an attempt will be made to use the more important ones in an effort to find past and present relationships which may prove helpful in forecasting future enrollment figures.

As concerns the interpretation of the results of the analysis and the forecast, it is sufficient to say that these elements will be dependent

---

[12] Frederick E. Croxton and Dudley J. Cowden, Practical Business Statistics (New York: Prentice-Hall, Inc., 1948), p. 207.

upon what the analysis produces. If a growth curve gives a reasonable fit to the time-series, the data may be projected into the future. If a high degree of relationship exists between enrollment figures and some other variable or group of variables, such knowledge will be used in predicting future enrollment. Finally, an attempt will be made to synthesize all the pertinent results of the analysis plus a cognizance of those non-numerical factors into a forecast which will prove of value.

CHAPTER II

TREND ANALYSIS

The first major phase of this analysis is concerned with an attempt to explain the changes that have occurred in The University of Tennessee student enrollment figures with the passage of time. Such an analysis is a dynamic problem and may be called a time series analysis.[1]

The data of the time series used in this study, i.e., The University of Tennessee student enrollment, came directly from the Office of the Dean of Admissions. The figures are available and may be used without preliminary adjustment. By preliminary adjustment is meant the attempt to express the enrollment figures in terms that will be the most significant for purposes of analysis. Such adjustment is often necessary before attempting to discover what is the normal year to year growth and the normal quarter to quarter variation in the activity of the enrollment.[2] An example of a preliminary adjustment would be one for population changes. In this case no such adjustment is made because the method of trend measurement considers such changes. It is important to note, however, that at times the gross enrollment is used, while at other times the total enrollment per quarter is used. The first will be used in a study of the upward growth over a period of years while the latter will be used in a study of the variation within the period of one year. A definite statement will indicate which of these figures is being used so that confusion should not

---

[1] Frederick E. Croxton and Dudley J. Cowden, Applied General Statistics (New York: Prentice-Hall, Inc., 1939), p. 363.

[2] Ibid., pp. 378-79.

exist.

There is no set approach to the analysis for a time series, but generally the object is to isolate the effects of the individual forces affecting the series in question. No method will isolate the effects of the individual forces perfectly, but the methods used here have been developed to the point that such an analysis permits understanding of past behavior and even permits predictions of future behavior of the series in some instances.[3]

The question arises as to just what the different forces are that affect time series. The answer to such a question lies in the particular series being analyzed; however, a certain characteristic of time series are apparent upon very brief inspection and may be placed in a limited number of categories. Generally four forces or factors are recognized: trend, seasonal, cyclical, and irregular variations. As concerns this study of The University of Tennessee enrollment, all of these factors will receive intensive study. This chapter is concerned with the location and measurement of the trend.

Trend Characteristics

There are general characteristics of trend analysis that should be stated and recognized before an attempt is made to fit a trend line to a time series. First, trend is the tendency of data to move gradually in

---

[3]William Addison Neiswanger, Elementary Statistical Methods (New York: The Macmillan Company, 1948), pp. 475-76.

direction, either an increase or a decrease, as distinguished from the tendency to fluctuate periodically with the season or to follow cyclical swings. It is then an increase or decrease over a long period of time; hence, before a trend analysis may prove of value, a relatively large number of observations must be used. Second, the trend value at any date is defined as being the normal value at that date if the data covers a period of a year or longer. This concept of a normal value gives a base in judging the effects of forces other than the growth factor. Such a normal figure, then, should not be assumed to possess any special normative significance.[4] Third, the trend line should be objective and logical. This point is summarized by Croxton and Cowden in the passage below:

> If . . . the object is to make comparisons, generalizations, or forecasts, the survey should not be only logical, but also of a nature that can readily be expressed by a mathematical formula. By so doing, a person can, for instance, say that at a given time a series shows a certain rate or a certain amount of growth per annum, and that, if this tendency continues, the trend will reach a certain value at some specific time in the future. Fitting a trend line does not, however, remove the subjective element from trend fitting. The statistician can vary somewhat the shape or the curve by selection of the type of formula he employs, or the years to which he fits the curve. It remains true, therefore, that the statistician decides in advance, upon as objective and logical a basis as possible, what he thinks the trend ought to look like, and then selects the mathematical method that will closely approximate this result.[5]

---

[4]Croxton and Cowden, *op. cit.*, p. 285.

[5]*Loc. cit.*

Gompertz Curve

The gross enrollment figures have been obtained from the Dean of Admissions office for each year from 1899 through 1949. These figures are shown in Table I and are pictured graphically on arithmetic and semi-logarithmic paper in Figures 1 and 2 respectively. A thorough study indicated that the mathematical method which best describes the data is the Gompertz growth curve, represented by the equation $y=abc^x$.[6] In this equation y represents the ordinates of trend of the original series; a, b, and c are constants; and x represents time in years with the origin set at the date of the first observation.[7] The decision to use the Gompertz curve was made

---

[6] Before the decision was made to use the Gompertz growth curve to describe the growth of the University enrollment, the following curves were fitted to the data:

1. Straight line
2. Second degree parabola
3. Logistic
4. Exponential
5. Modified Exponential
6. Gompertz

The results of fitting the above curves may be summarized as follows: Curves 1 and 2 were discarded because they follow no growth pattern since extrapolation produces illogical results; curve 3 is theoretically sound but could not be fitted to the enrollment figures because of the pattern in which University enrollment is growing; curves 4 and 5 fitted the enrollment data fairly well; however, the Gompertz equation yielded the results which best describe the enrollment data. Again it is essential to stress that curve fitting is a very subjective technique.

[7] Frederick C. Mills, Statistical Methods (New York: Henry Holt and Company, 1924), p. 671.

only after much experimentation and fitting of other types of growth curves, i.e., the Logistic and the Exponential, etc.

Justification for the use of the Gompertz curve as it is used in this study may be found by analyzing the theory supporting the method. The Gompertz curve, which has important uses in actuarial science and which has had many applications in the study of economic and business trends, has received its greatest support from Raymond D. Prescott. Prescott concludes that since the general shape of the curve is so common to many industries, it describes a law of growth.[8] The term "growth curve" does seem applicable to it when the general shape of the curve is noted. The amount of growth is small at first, then becomes larger until it reaches a point of inflection after which the growth continues but by decreasing absolute increments to the end without retrogression. Prescott insists that this trend is a function of population growth.

With the above general statements concerning the shape of the Gompertz curve as criteria, it is possible to observe the fitting of the Gompertz curve to The University of Tennessee enrollment figures. As has been stated, these figures were plotted on both arithmetic and semi-logarithm graph paper so that after the proper computations were made the calculated trend values could be plotted and the trend line drawn.

Gompertz Curve Fitted to Enrollment

For the purpose of fitting, the equation for the Gompertz curve was transformed from the natural form, $y = ab^{c^x}$, to the logarithmic form,

---

[8]Croxton and Cowden, op. cit., pp. 447-48.

$\log y = \log a + (\log b)\, c^x$.[9]

As fitted to the series of enrollment figures measuring the growth of The University of Tennessee, log a is the logarithm of the maximum value. Practically, this means that the value of a is an asymptote (ceiling) and that if we extended the trend line indefinitely, it would approach closer to the value but never reach it. The term $(\log b)\, c^x$ measures the amount by which the trend value at a given time falls short of the asymptote. This amount decreases with the passage of time.

A brief explanation will be given of the formulaw used in fitting the Gompertz curve, but the reader can find all the important computations necessary to obtain the trend equation for the University enrollment by referring to Table I.

The fifty-one year period from 1899 through 1949 was chosen as the period of observation because it covers practically the first half of the twentieth century, which represents the modern history of the University, and also because it may be conveniently broken into three equal portions of seventeen years each. This break-down is essential to the fitting of the curve by the method of selected points. Since the transformation of the trend equation requires the problem to be worked in logarithms, it is necessary to obtain the logarithms of each of the fifty-one enrollment figures. The logarithms of the enrollment figures in each of the three segments must be summed and from these sums, and the difference between them, the constants can be computed.

The number of terms entering into each of the three sub-totals in this problem is 17 and is represented by the letter n. The sub-totals are

---

[9] Mills, op. cit., p. 671.

TABLE I

GOMPERTZ CURVE FITTED TO THE UNIVERSITY OF TENNESSEE ENROLLMENT 1899-1949

Year (1)	y Students (2)	Log y (3)	(4)	(5)
1899	266	2.424882		
1900	343	535294		
1901	364	561101		
1902	335	525045		
1903	412	614897		
1904	364	561101		
1905	353	547775		
1906	388	588832		
1907	428	631444		
1908	509	706718	S_1=44.733238	
1909	616	789581		
1910	514	710963		
1911	414	617000		
1912	489	689309		
1913	486	686636		
1914	531	725095		d_1=S_2-S_1
1915	657	817565		
				8.859725
1916	752	876218		
1917	781	892651		
1918	679	831870		
1919	619	791691		
1920	962	983175		
1921	1,020	3.008600	S_2=53.592963	
1922	1,178	071145		
1923	1,313	118265		
1924	1,459	164055		
1925	1,620	209515		
1926	1,675	224015		
1927	2,064	314710		
1928	2,479	394277		
1929	2,670	426511		
1930	2,824	450865		
1931	2,642	421933		
1932	2,591	413467		
				d_2=S_3-S_2
				6.837184

TABLE I (continued)

GOMPERTZ CURVE FITTED TO THE UNIVERSITY OF TENNESSEE ENROLLMENT 1899-1949

Year (1)	$\bar{y}$ Students (2)	Log $\bar{y}$ (3)	(4)
1933	2,259	3.353913	
1934	2,030	307496	
1935	2,336	368473	
1936	2,731	436322	
1937	3,019	479863	
1938	3,239	510411	
1939	3,574	553155	
1940	3,741	572988	
1941	3,850	585461	
1942	3,569	552547	S_3 = 60.430147
1943	3,360	526339	
1944	1,894	277380	
1945	2,237	349666	
1946	5,401	732474	
1947	8,405	924538	
1948	8,914	950073	
1949	8,893	949048	

Source: Dean of Admissions Office.

represented by S_1, S_2, and S_3, and the first differences between the subtotals are represented by d_1 and d_2. These quantities are essential in solving for the three constants, log a, log b and c. These constants are obtained by using the above quantities in the following formulae:[10]

$$C^n = C^{17} = \frac{d_2}{d_1} = \frac{6.837184}{8.859725} = .771715$$

$$C = \sqrt[17]{.7717154} = .984872$$

$$\log b = \frac{d_1\,(c-1)}{(c^n - 1)^2} = \frac{8.859725\ (.984872 - 1)}{(.771715 - 1)^2} = -2.570846$$

$$\log a = \frac{1}{n}\left(S_1 - \frac{d_1}{C^n - 1}\right)$$

$$= \frac{1}{17}\left(44.733238 - \frac{8.859725}{.771715 - 1}\right)$$

$$= 4.914308$$

$$a = 82{,}093$$

The Gompertz trend equation, $\log y = \log a + (\log b)\, c^x$, then becomes $\log y = 4.914308 - 2.571869\ (.98487)^x$, where x represents the deviation in years from the origin, 1899.

When the trend equation is used and the values of x from 1899 through 1949 are substituted, the logarithms of the trend values are obtained. The anti-log is then obtained for the log of each year, and when the result for each year is plotted, a curve is formed which is defined graphically along

---

[10]Mills, op. cit., p. 448.

with the original data in Figure 1. The methods of calculation of the trend values are shown in Table II.

In Figure 1 the enrollment figures and the Gompertz curve fitted to them are shown plotted on arithmetic graph paper. Remembering that two other known factors, cycle and irregular forces, cause the major portion of the variations in the original data, observation reveals that the Gompertz curve adequately describes the growth in the enrollment figures. Perhaps a better picture of the closeness of fit may be obtained from Figure 2, where the original data and the Gompertz curve are shown on a chart with a logarithm vertical scale. The extrapolation of the curve into the future illustrates the general shape. The amount of growth is small at first, then it becomes larger until it reaches an inflection point where the amount of growth begins to decrease. Such an inflection point is not obvious during the years covered by the original data which indicates that the University enrollment is still growing by increasing absolute amounts.

Prescott Theory Applied to Enrollment

Prescott, in fitting the Gompertz curve to depict the growth of industry, believes that industrial growth may be divided into four stages:[11]

(1) Period of experimentation
(2) Period of growth into the social fabric
(3) Through the point where growth increases but at a diminishing rate
(4) Period of stability.

---

[11] Croxton and Cowden, *op. cit.*, p. 448.

TABLE II

TREND ORDINATES OF GOMPERTZ CURVE FITTED TO THE UNIVERSITY OF TENNESSEE ENROLLMENT 1899-1949 AND EXTRAPOLATED TO 1970

Year (1)	x (2)	c^x (3)	(log b) c^x (4)	log y (4) ≠ log a (5)	Anti-log of (5) (in students) (6)
1899	0	1.000000	-2.571869	2.342439	220
1900	1	.984872	-2,532962	2.381346	241
1901	2	.969973	-2.494643	2.419665	263
1902	3	.955299	-2.456904	2.457404	287
1903	4	.940847	-2.419735	2.494573	313
1904	5	.926614	-2.383130	2,531178	340
1905	6	.912596	-2.347077	2.567231	369
1906	7	.898790	-2.311570	2.602738	401
1907	8	.885193	-2.276600	2.637708	434
1908	9	.871802	-2.242161	2.672147	470
1909	10	.858613	-2.208240	2.706068	508
1910	11	.845624	-2.174834	2.739474	548
1911	12	.832832	-2.141935	2.772373	592
1912	13	.820233	-2.109532	2.804776	638
1913	14	.807824	-2.077618	2.836690	686
1914	15	.795603	-2.046187	2.868121	738
1915	16	.783567	-2.015232	2.899076	792
1916	17	.771714	-1.984747	2.929561	850
1917	18	.760039	-1.954721	2.959587	911
1918	19	.748541	-1.925149	2.989159	975
1919	20	.737217	-1.896026	3.018282	1043
1920	21	.726064	-1.867341	3.046967	1114
1921	22	.715081	-1.839095	3.075213	1189
1922	23	.704263	-1.811272	3.103036	1268
1923	24	.693609	-1.783871	3.130437	1350
1924	25	.683116	-1.756885	3.157423	1437
1925	26	.672781	-1.730305	3.184003	1528
1926	27	.662639	-1.704221	3.210087	1622
1927	28	.652580	-1.678350	3.235958	1722
1928	29	.642708	-1.652961	3.261347	1825

TABLE II (continued)

TREND ORDINATES OF GOMPERTZ CURVE FITTED TO THE UNIVERSITY OF TENNESSEE ENROLLMENT 1899-1949 AND EXTRAPOLATED TO 1970

Year (1)	x (2)	c^x (3)	(log b) c^x (4)	log y (4) + log a (5)	Anti-log of (5) (in students) (6)
1929	30	.632985	-1.627954	3.286354	1933
1930	31	.623409	-1.603326	3.310982	2046
1931	32	.613978	-1.579071	3.335238	2164
1932	33	.604688	-1.555178	3.359130	2286
1933	34	.595542	-1.531656	3.382652	2414
1934	35	.586533	-1.508486	3.405822	2546
1935	36	.577659	-1.485663	3.428645	2693
1936	37	.568921	-1.463190	3.451118	2826
1937	38	.560314	-1.441054	3.473254	2973
1938	39	.551838	-1.419255	3.495053	3126
1939	40	.543489	-1.397783	3.516525	3285
1940	41	.535267	-1.376637	3.537671	3449
1941	42	.527170	-1.355812	3.558496	3618
1942	43	.519195	-1.335302	3.579006	3793
1943	44	.511340	-1.315099	3.599209	3974
1944	45	.503605	-1.295206	3.619102	4160
1945	46	.495986	-1.275611	3.638697	4352
1946	47	.488483	-1.256324	3.657984	4550
1947	48	.481093	-1.237318	3.676990	4752
1948	49	.473815	-1.218600	3.695708	4963
1949	50	.466647	-1.200154	3.714154	5178
1950	51	.459588	-1.182001	3.732307	5399
1951	52	.452635	-1.164118	3.750190	5626
1952	53	.445787	-1.146506	3.767802	5859
1953	54	.439043	-1.129161	3.785147	6097
1954	55	.432402	-1.112081	3.802227	6342
1955	56	.425860	-1.095256	3.819052	6593
1960	61	.394608	-1.014880	3.899428	7933
1965	66	.365649	- .940401	3.973907	9417
1970	71	.338825	- .871414	4.042894	11038

Source: Table I.

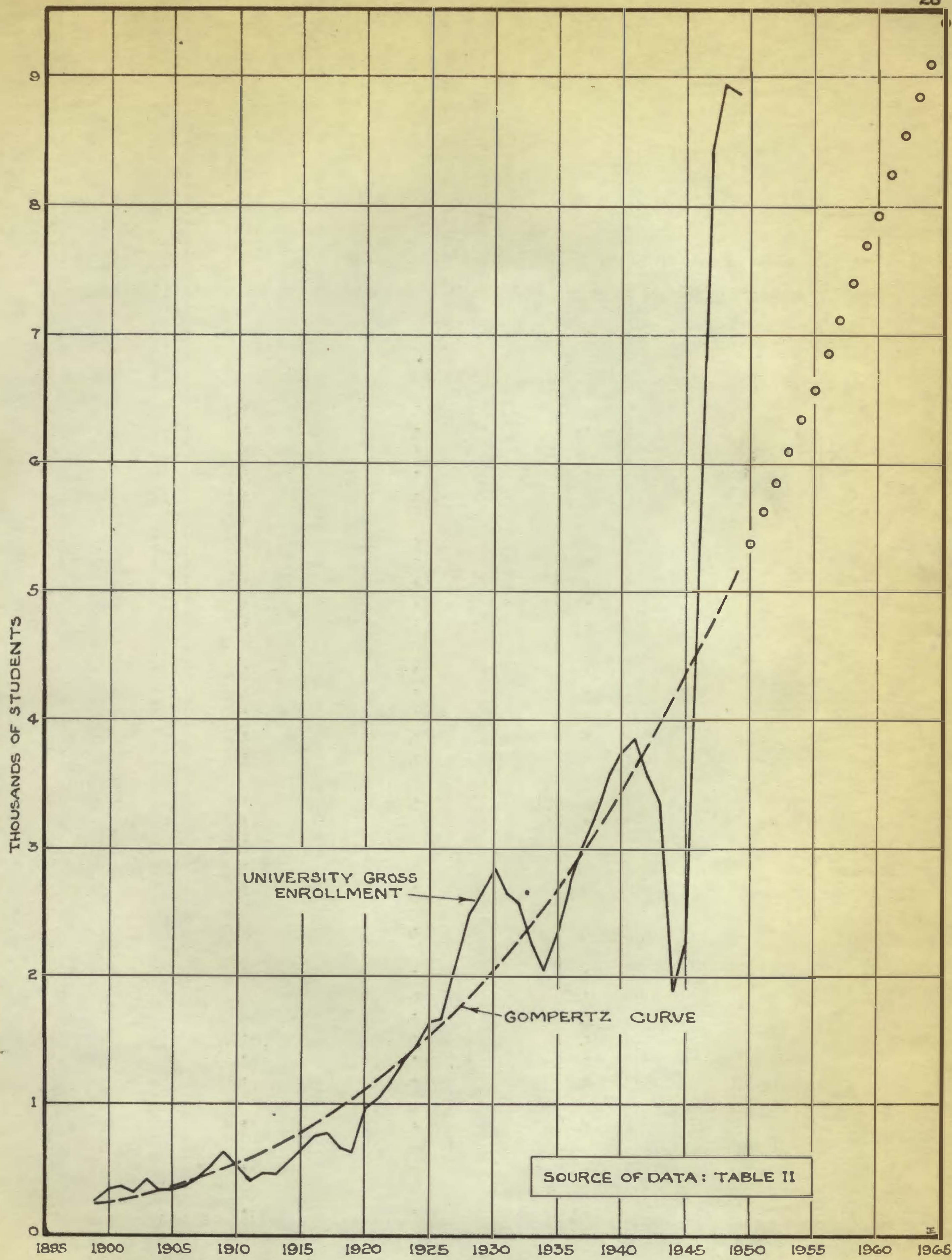

Figure 1. Gompertz curve fitted to The University of Tennessee gross enrollment, 1899 – 1949, with trend line extrapolated

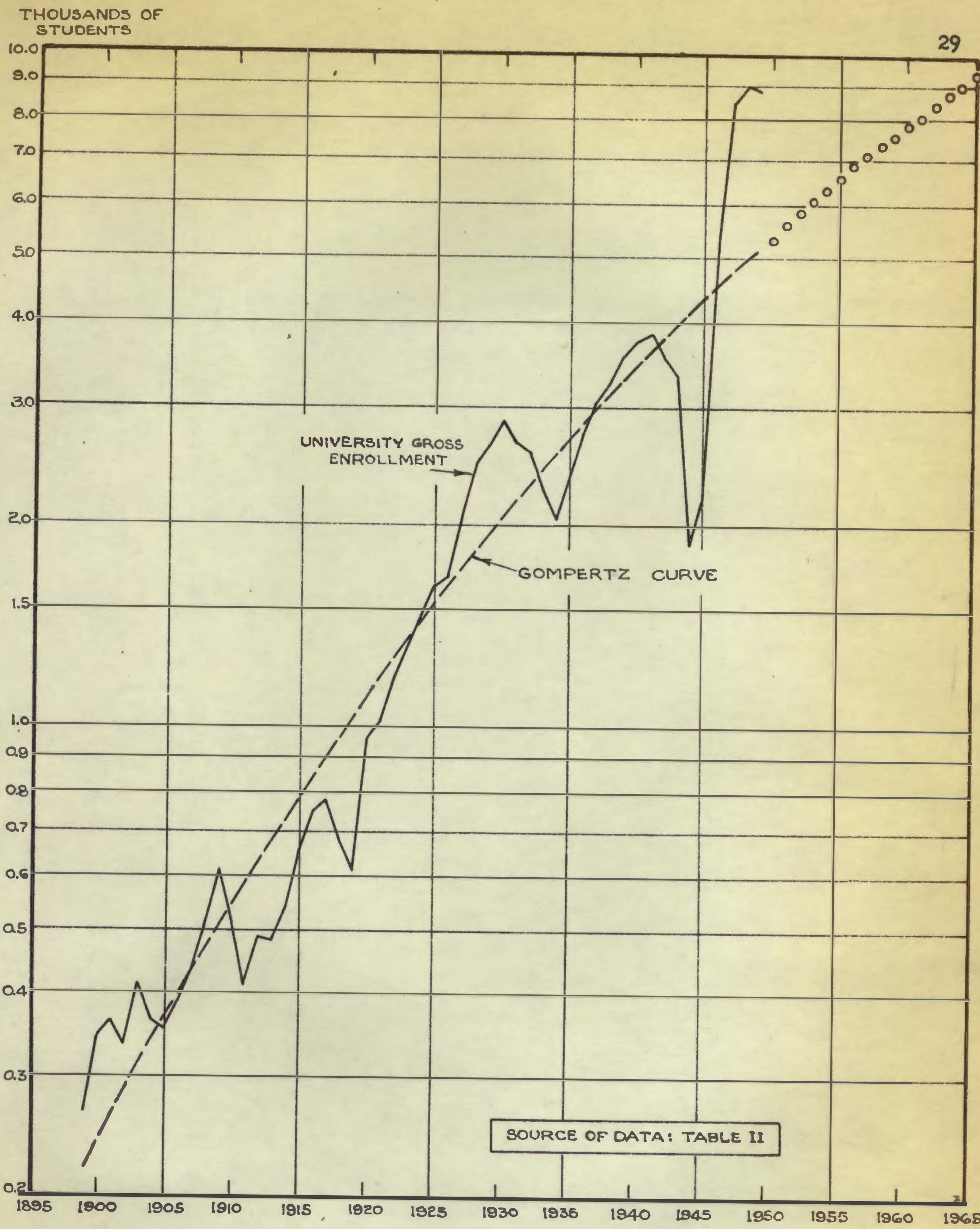

Figure 2. Gompertz curve fitted to The University of Tennessee gross enrollment, 1899-1949, with trend line extrapolated (Logarithm vertical scale)

Although The University of Tennessee is not an industry, the increase in enrollment is dependent to an extent on population growth and many other factors that are common to industrial growth. It may be possible to say then that the growth of The University of Tennessee is subject to division into four comparable stages. Analysis of the modern history of the institution reveals that the "period of experimentation" may have existed until approximately 1920. During this period the School of Commerce, School of Home Economics, Graduate School and the Pre-Med program came into existence and became a part of the present organization. During these same years, the Summer sessions began, while Knoxville and Fisk Colleges ceased being a part of the University.[12] During this period of experimentation, the growth was slow; however, it began to increase much faster during the 1920's and the University entered its second stage of growth, "the period of growth into the social fabric."

It is possible that the University is at present in this second stage of growth. The growth in enrollment continued to increase until the war years of the 1940's. The effect of the war years will be discussed in Chapter IV. Suffice it to say all appearances now indicate that The University of Tennessee with its program of higher education is growing into the social fabric and becoming a part of every day life in this state.

Two stages of growth would then remain for The University of Tennessee, i.e., through the point where enrollment continues to grow but at a diminishing rate and through the period of stability. Although there may be violent

---

[12] This information was obtained from records in the Dean of Admissions Office compiled by F. Johnston.

fluctuation in enrollment during the next few years, it is virtually impossible to conceive of an inflection point in growth being reached before 1960. Many factors, including the favorable attitude of the people toward higher education, the growing interest of the government, and the increased population resulting from excessive births during the war, point toward greater increases in enrollment figures.[13]

The asymptote (ceiling) of the curve, which was set by the constant a as 82,093, is not indicative of an enrollment that will be approached in the near future. Instead, it is an upper limit which will never be reached and which will be closely approached in theory only in the distant future (about 5000 A.D.). If the growth of The University of Tennessee does continue as the Gompertz curve indicates, it is highly probable that steps will be taken to decentralize the institution before it would be allowed to reach an enrollment greatly in excess of, for example, 20,000. Many reasons both physical and economic serve as evidence that the enrollment may never grow to 82,093 students.

Projection of Gompertz Curve

Projection of the Gompertz into the more immediate future gives the following enrollment figures:

---

[13] George F. Zook, "The President's Commission on Higher Education Reports", Higher Education, 4:133, Feb. 15, 1948.

Year	Estimated enrollment
1955	6,593
1960	7,933
1965	9,414
1970	11,038

It is essential that the reader realizes that these figures are not purported to be exact enrollments in the years cited. As has been stated earlier, these figures are taken to be normal values as of those dates, assuming that the conditions underlying the growth of the University remain approximately the same as they have been in the past. There are many factors which will cause the actual enrollment figures to be higher or lower than the above mentioned estimates. As an example, assume that another war becomes a reality around 1960; certainly the enrollment will be much below the normal value as determined by the Gompertz curve. Again assume the Federal government subsidizes higher education; the enrollment estimates might be on the low side. Even if such events of an irregular nature fail to materialize, future enrollment will not coincide with these normal trend estimates because of cyclical movements. Some of the causes for fluctuations around these normal values will be discussed in Chapter IV.

CHAPTER III

QUARTERLY VARIATION

In a further application of the methods of time series analysis to The University of Tennessee enrollment data, this chapter is devoted to the study of the periodic fluctuation which has a duration of one year.[1] Having discussed the measurement of the trend component, one of the remaining tasks is to explain the nature of the quarterly variation and the method of measuring it.

In the preceding chapter the development was in terms of gross enrollment. In periodic analysis several observations are needed during each year; and since the University uses the four quarters per year system, quarterly data are used. Since the determination of the quarterly pattern requires a fairly long period of years, the total number of students enrolled per quarter has been obtained for every quarter beginning with the Summer of 1937 and ending with the Spring of 1949.[2] The four quarters of the year will be designated as follows: Summer, Fall, Winter and Spring.

Concept of Quarterly Pattern

Before a measurement is made of the quarterly pattern in the enrollment data, it is necessary to summarize the concept of the quarterly

---

[1]Periodic fluctuations are referred to by the term "seasonal" variation which is used interchangably with the term quarterly variation in this thesis.

[2]Summer quarter enrollment does not include persons enrolled in special short courses such as banking seminars, etc.

pattern. A quarterly variation is a rythmic movement which recurs annually with approximately the same relative intensity.[3] Such a definition places limitations on the concept of a quarterly variation in three directions and at the same time points the way to methods by which such variations can be measured. First, it recurs annually: all other rhythmic components of a series are eliminated except the regular one with a period of four quarters. Second, quarterly variation recurs from year to year with similar intensity. Finally, relative intensity means that a particular quarter, i. e., Fall, is expected to vary from a normal quarter in the same direction by the same percent year after year because of quarterly influence.[4]

The concept of quarterly variation then may be summarized as a pattern which is assumed to be typical of any year of a series. As will be noted later such a pattern is obtained in the form of an index composed of four quarterly percentages whose average is 100, and which represents the quarter variation as a percent of trend or "normal."[5]

---

[3]This definition is restricted purposely to what is commonly called a stable seasonal pattern or one in which the pattern is constant from year to year. This is justified in view of evidence from test and by inspection that the pattern is not changing or moving.

[4]Martin A. Brumbaugh and Lester S. Kellog, Business Statistics (Chicago: Richard D. Irwin, Inc., 1946), p. 594.

[5]Frederick E. Croxton and Dudley J. Cowden, Practical Business Statistics (New York: Prentice-Hall, Inc., 1948), p. 207.

It is pertinent at this stage to discuss the reasons for measuring quarterly movements. Actually, the quarterly movements may be the most important of all the factors entering into a time series because these movements are indicative of what is happening to the enrollment figures in the short run. The knowledge provided by such a measurement may be very helpful in the short run, i. e., over the period of a year. As an example, if it is found that more students are enrolled in the Fall quarter than any other quarter, such information may be helpful in determining the minimum needs in instructional staff, administrative staff, housing, class rooms , and many other things. If it is found that the Summer contains the smallest number of students, this may be the time to repair and build. It is entirely possible that if a definite pattern is found where enrollment in one quarter differs greatly from the preceding or following quarter, such fluctuations might be reduced by finding the reasons why students prefer one quarter to another and changing their preferences. Again, as it is possible to measure the individual quarters in a quantitative manner, it is possible to predict the number of students to be enrolled in a future quarter from the number of students enrolled in the present quarter, e. g., if the Summer quarter's enrollment were known, it would be possible to predict enrollments for the Fall, Winter and Spring quarters, given the quarterly index.

As has been pointed out earlier in the chapter, a series must have quarterly fluctuations recurring regularly from year to year before it can be said to contain a quarterly pattern. Unless it does have such a pattern no index of quarterly variation can be computed. It is important then to make certain that a quarterly pattern does exist before an attempt is made

to measure it.[6] In the case of the enrollment data, a study of the enrollment figures in Table III or a glance at the graphical presentation of these data in Figure 3 reveals a definite pattern. Generally, this pattern shows that during the Summer quarter the smallest number of students are registered, then the enrollment increases considerably in the Fall and the most students are enrolled in this quarter. The Winter quarter enrollment is a little less than the Fall enrollment, and the Spring enrollment is consistently a little less than the Winter enrollment. This same pattern occurred year after year with the exception of a short period in 1945 and 1946 when the post-war readjustment was being made, i. e., veterans released from the services were enrolling in school the first quarter after they were discharged irrespective of quarter. In an attempt to measure the quarterly variation, the method selected will provide for the exclusion of the enrollment figures which do not conform to the pattern since such deviation from the pattern may be explained.

Moving Average Method

Many methods of measuring periodic movements are explained in statistical textbooks and periodicals. One method, the use of the moving average to separate periodic movements from trend and cycle has become the most used of those available.[7] This method in the form used in this analysis was originally published in the Federal Reserve Bulletin in

---

[6]Brumbaugh and Kellog, op. cit., pp. 594-95.

[7]John R. Stockton, An Introduction to Business Statistics (New York: D. C. Heath and Company, 1938), p. 174.

TABLE III

PERCENTAGES OF CENTERED FOUR-QUARTER MOVING AVERAGE OF THE UNIVERSITY OF TENNESSEE QUARTERLY ENROLLMENT, 1937-1949

Year and quarter (1)	Number of students enrolled (2)	Four-quarter moving total (3)	Centered eight-quarter moving total (4)	Centered four-quarter moving average (Col. 4 ÷ 8) (5)	Percent of centered four-quarter moving average (Col. 2 ÷ Col. 5) (6)
1937: Summer	1,473				
38 Fall	2,929	9,738			
Winter	2,688	10,107	19,845	2,481	108.3
Spring	2,648	10,351	20,458	2,557	103.6
1938: Summer	1,842	10,717	21,068	2,633	69.9
39 Fall	3,173	11,058	21,775	2,722	116.6
Winter	3,054	11,218	22,276	2,784	109.7
Spring	2,989	11,470	22,688	2,836	105.4
1939: Summer	2,002	11,623	23,093	2,887	69.4
40 Fall	3,425	11,748	23,371	2,921	117.3
Winter	3,207	11,739	23,847	2,936	109.2
Spring	3,114	11,792	23,531	2,941	105.9
1940: Summer	1,933	11,852	23,644	2,956	67.4
41 Fall	3,478	11,749	23,601	2,950	117.9
Winter	3,267	11,661	23,410	2,926	111.7
Spring	3,011	11,483	23,144	2,889	104.2
1941: Summer	1,905	11,193	22,676	2,835	67.2
42 Fall	3,300	11,007	22,270	2,784	118.5
Winter	2,977	11,085	22,092	2,762	107.8
Spring	2,825	10,876	21,961	2,745	102.9
1942: Summer	1,983	10,596	21,472	2,684	73.9
43 Fall	3,091	10,087	20,683	2,585	119.6
Winter	2,697	9,568	19,655	2,457	109.8
Spring	2,316	8,128	17,696	2,212	104.7
1943: Summer	1,464	6,962	15,090	1,886	77.6
44 Fall	1,651	5,984	12,946	1,618	102.0
Winter	1,531	5,705	11,689	1,461	104.8
Spring	1,338	5,905	11,510	1,451	92.2
1944: Summer	1,185	6,101	12,006	1,501	78.9
45 Fall	1,851	6,397	12,498	1,515	122.1
Winter	1,727	6,471	12,868	1,609	107.3
Spring	1,634		13,568	1,696	96.3

TABLE III (continued)

PERCENTAGES OF CENTERED FOUR-QUARTER MOVING AVERAGE OF THE UNIVERSITY OF TENNESSEE QUARTERLY ENROLLMENT, 1937-1949

Year and quarter	Number of students enrolled	Four-quarter moving total	Centered eight-quarter moving total	Centered four-quarter moving average (Col. 4 ÷ 8)	Percent of centered four-quarter moving average (Col. 2 ÷ Col. 5)
(1)	(2)	(3)	(4)	(5)	(6)
1945: Summer	1,259	7,097	15,933	1,922	63.2
46 Fall	2,477	8,836	20,424	2,553	97.0
Winter	3,466	11,588	26,218	3,277	105.8
Spring	4,386	14,630	34,225	4,278	102.5
1946: Summer	4,301	19,595	42,678	5,335	80.6
47 Fall	7,442	23,083	48,465	6,058	122.8
Winter	6,954	25,382	51,386	6,423	108.3
Spring	6,685	26,004	52,256	6,532	102.3
1947: Summer	4,923	26,252	52,917	6,615	74.4
48 Fall	7,690	26,665	53,512	6,689	115.0
Winter	7,367	26,847	53,687	6,711	109.8
Spring	6,867	26,840	53,780	6,723	102.11
1948: Summer	4,916	26,940	53,845	6,856	71.7
49 Fall	7,790	26,905	53,840	6,730	115.8
Winter	7,332	26,935	53,852	6,732	108.8
Spring	6,897	26,917			
1949: Summer	4,898				
50 Fall	7,506				
Winter	6,668				
Spring	6,330				

Source: Quarterly enrollment figures taken from Dean of Admissions Office.

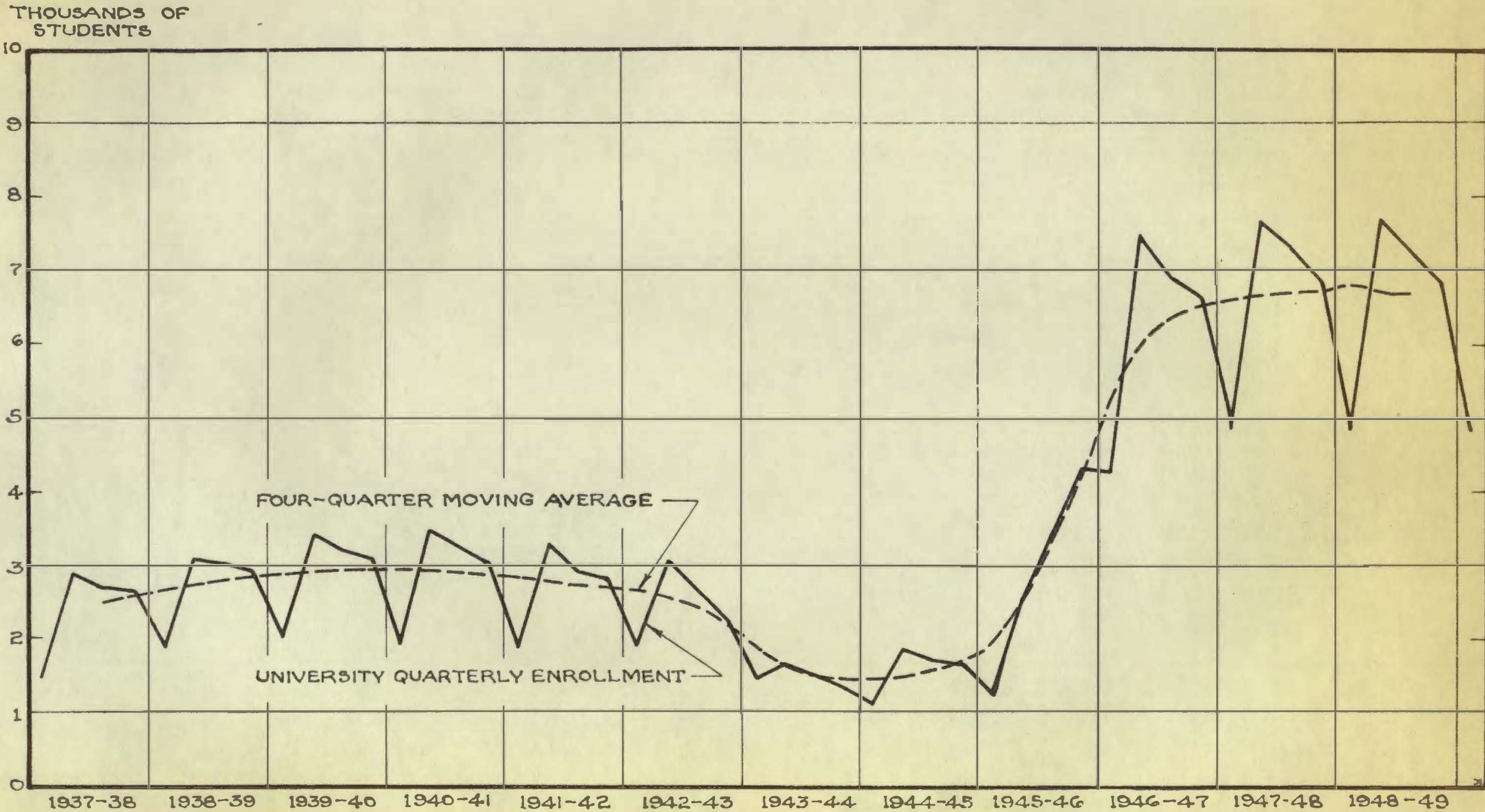

Figure 3. The University of Tennessee quarterly enrollment and centered four-quarter moving average, 1937 - 1949

Source of Data: Table III

December 1922 as a part of an article, "Index of Production in Selected Basic Industries."[8] Since that time it has been acclaimed the most satisfactory method of computing a typical quarterly index. It eliminates from the final expression of quarterly variation the influences of trend, cycle and random movements to a large degree. The steps in the calculation of the quarterly index by this method will be examined in order to point out and explain why in the process of computation such desirable results are obtained.

A four quarter moving average is first calculated from the enrollment data. Since one complete quarterly wave is included in each computation, the four quarter moving average contains no quarterly variation. Theoretically, this four quarter moving average represents the data if the quarterly variations were not present. In other words, the quarterly expansion and contractions tend to cancel each other in a four quarter moving average and it is possible to say that the moving average now contains trend and cycle represented by the letters TC.

The next step in obtaining the quarterly index is to divide each quarterly enrollment figure which contains trend, cycle, season and irregular movements (TCSI) by the corresponding value of the four quarter moving average (TC). Since the four quarter moving average represents trend and cycle, the ratio, original data (TCSI) divided by four quarter moving average (TC), gives the original data as a percentage of trend and cycle combined and may be said to contain seasonal and irregular movements represented by the letters SI.[9]

---

[8]Brumbaugh and Kellog, op. cit., p. 599.

[9]Ibid., pp. 599, 600.

Moving Average Method Applied to Total Enrollment Per Quarter

The detailed computation of the moving average for the enrollment data as depicted in Table III is started by making a total of the enrollment figures for four quarters, for example, Summer 1937 through Spring 1938 to obtain the first total, 9,738, column 2. This total represents the cumulation of the four quarters and should be posted after the middle of the period which is January 1, 1938. Then the Summer quarter of 1937 is dropped and the Summer quarter of 1938 is added. This total which represents the cumulation of the four quarters, Fall 1937 through Summer 1938 is posted at the end of the Winter quarter and is 10,107. This process of dropping a quarter and adding a new one continues as the totals move forward in time.

When an even number of items are employed in the computation of a moving average, it is necessary to perform another computation to center the average. In this instance, when the four quarter total is taken for the period Summer 1937 through Spring 1938, the average would fall at the center of the one year period or at the end of the Winter quarter, but data which are typical of a quarter must be plotted at the middle of the quarter. Therefore it is necessary to find the value of the moving average for the middle of the Winter quarter 1938. Using a short-cut method, the value is obtained by adding the total as of the end of the Fall quarter to the total as of the end of the Winter quarter. This gives a total of 19,845 students, which represents eight quarters. The total as of the end of the Winter quarter (9,738) is then dropped and the total at the end of the Spring

quarter is added to give a total of 20,458, which represents eight quarters and is posted as of the middle of the Spring quarter. This process of a two item or an eight quarter centered moving total is carried through in column 3. When each of the numbers in column 3 is divided by eight, as is done in column 4, the result is a four-quarter centered moving average which is posted at the middle of the respective quarters. The plotting of these values may be seen as the dotted line in Figure 3 which represents the trend and the cycle of the data.

As was stated before, each four quarter moving average (TC) is divided into the corresponding actual enrollment figure (TCSI) and the results as calculated in column 5 are the original enrollment figures as a percentage of trend and cycle combined. These ratios contain seasonal and irregular movements and tend to fluctuate around 100. When these ratios are above 100 percent, positive quarterly movement is the cause; when they are below 100 percent, negative quarterly movement is the cause.[10]

Typical Quarterly Index Calculated

Now, it may be observed, there is present a rough measurement of the quarterly pattern in the form of the ratios. These ratios, however, contain not only quarterly but also irregular movements. They also are specific in that each ratio describes a particular quarter. The next step in the calculation has, therefore, a twofold purpose: (1) to eliminate from these ratios the influence of irregular movements and (2) to take an average of

---

[10]Stockton, _op. cit._, p. 17.

these ratios in order to get a "typical" quarterly index. The process of averaging will cancel out much of the error of random nature, but at the same time an average will be sensitive to cyclical or other effects not eliminated from the ratios. It is advisable in such a situation to use a modified mean, which is merely a method by which the advantages of both the mean and the median are obtained. To compute this average it is necessary first to make an array of the ratios in column 5 of Table III for each quarter. The ratios of the original data to the centered four quarter moving average for the Fall quarter are arrayed in the first column of Table IV, the ratios for the Winter quarter are arrayed in the second column, and so on for each quarter.[11]

Once the ratios are arrayed in four columns labeled Summer, Fall, Winter and Spring, it becomes necessary to decide how many of the central items will go into the make-up of the modified mean. To eliminate the tendency of the median to be erratic, it is advisable to compute the mean of the four or five central items of an array as the best measure of typical size. Just how many of the central items to be used in finding the typical value is difficult to decide; however, it usually does not make a great deal of difference since a certain amount of subjectivity is inherent in the method of securing the best typical value. In this case the modified mean was taken of the five middle items by obtaining the total of the five and dividing by five for each of the four quarters. The four figures obtained in this manner may be called crude seasonal indexes. The final step,

---

[11] William Addison Neiswanger, Elementary Statistical Methods (New York: The Macmillan Company, 1948), pp. 557-571.

TABLE IV

ARRAYED PERCENTAGES OF CENTERED FOUR-QUARTER MOVING AVERAGE OF THE UNIVERSITY OF TENNESSEE ENROLLMENT, WINTER QUARTER 1937-38-FALL 1948-49 AND COMPUTATION OF QUARTERLY INDEX

	Summer	Fall	Winter	Spring	Total
	63.2*	97.0*	104.8*	92.2*	
	67.2*	102.0*	105.8*	96.3*	
	67.4*	115.0*	107.3*	102.1*	
	69.4	115.8	107.8	102.3	
	69.9	116.6	108.3	102.5	
	71.7	117.3	108.3	102.9	
	73.9	117.9	109.2	103.6	
	74.4	118.5	109.7	104.2	
	77.6*	119.6*	109.8*	104.7*	
	78.9*	122.1*	109.8*	105.4*	
	80.6*	122.8*	111.7*	105.9*	
Total of middle five	359.3	586.1	543.3	515.5	
Average of middle five	71.9	117.2	108.7	103.1	400.9
Refined index	71.7	117.0	108.4	102.9	400.0
Correction factor	400 ÷ 400.9 = .99776				

*Eliminated in calculating modified mean.

Source: Table III.

obtaining a refined quarterly index, involves adjusting the values of the four quarterly indexes so that their sum will be 400.9. The step is necessary in order to obtain a quarterly index in which the percents for the several quarters are balanced above and below 100 percent.

From Table IV it may be noted that the four crude quarterly index figures give a total of 400.84. Since this figure should have been 400, it is necessary to decrease each of the quarterly index figures so that their sum will be 400. This is done by dividing 400 by 400.9 and obtaining the correction factor of .99776 which is multiplied by each crude quarterly index to obtain the following refined quarterly indexes:

Quarter	Percent
Summer	71.7
Fall	117.0
Winter	108.4
Spring	102.9

The total of the four is 400.0.

Uses of an index of quarterly variation have been cited earlier in this chapter; however, it may prove advantageous at this time to reiterate that such an index may be very useful in making the quarterly plans for the University. The extent of the use of a measure of quarterly variation depends upon the importance of the variation in the operation of the University. The administration no doubt knows with considerable accuracy the quarterly pattern in the fluctuations of enrollment. This knowledge is limited, however, to a general impression and has never been computed and expressed quantitatively.

Quarterly Index Use in Cycle Analysis

Another use for information concerning quarterly variation is to show more clearly the clyclical fluctuations in the enrollment series. While quarterly fluctuations complicate the planning of the University, the fact that they do recur regularly makes it possible to forecast them and allow for the changes. Cyclical fluctuations, on the other hand, are not so regular as quarterly fluctuation, and therefore may not be predicted with as great a degree of accuracy. Also, the cyclical swings may be greater in amplitude and duration than the quarterly swings, and consequently their effects are much more damaging. Since the effects of the cyclical swing in enrollment is evident, it is necessary that notice be taken of the cyclical effect on enrollment data.[12]

The index of quarterly variation is the tool by which significant enrollment changes are measured. In the University enrollment series which has definite quarterly fluctuations, it would be difficult to tell from quarter to quarter what cyclical changes are taking place. However, with an index of quarterly variation available which shows the effect of quarterly variation on each quarter, it is possible to determine how much of the increase or decrease is due to quarterly variation and whether any cyclical change has been experienced. As an example, according to Table IV, the typical decrease from Fall to Winter quarter is from 117.0 percent to 108.4 percent or 8.6 percentage points, which represents a decrease of 7.4 percent. Similarly, it is computed that Spring is normally 14.1 percentage points

---

[12] *Ibid.*, pp. 188-192.

lower than Fall, or a decrease of 12.1 percent; and Summer, 45.3 percentage points lower, or a decrease of 38.7 percent. When these typical quarterly decreases are compared with the decreases actually taking place in any one year, it may be determined whether the decrease was more or less than normal. If Winter enrollment decreased 15 percent from Fall as compared with a normal decrease of 7.4 percent between Fall and Winter, this is indicative of a decrease greater than the normal seasonal decrease. If the same type of thing happened for three or four quarters, it might not be possible to forecast that decrease in enrollment would continue; it would be possible, however, to state that the decrease in enrollment for four consecutive quarters was in excess of the normal quarterly decrease.

The manner in which the effects of the quarterly variation are removed from the enrollment series by the index of quarterly variation in order to show clearly the cyclical movements is perhaps the most important use of the index.[13] The cyclical effects will be discussed in the next chapter for gross enrollments.

Quarterly Index Computed for Gross Annual Enrollment

It is highly significant to note that the index of quarterly variations which has been the basis of the discussion to this point may be used in forecasting only when quarterly data are available. An example may serve to clarify this point. If the Fall quarter enrollment were given as 8,000 (which would mean that 8,000 persons were enrolled in the Fall quarter),

---

[13] Ibid., pp. 194-196.

the forecast as to persons enrolled in the Winter quarter would be computed as follows:

$$\frac{117.0}{8000} = \frac{108.4}{x}$$ where: 117.0 = Fall quarterly index

108.4 = Winter quarterly index

x = Winter quarterly enrollment

$$117.0 = 867,200$$

$$x = 7,412$$

An actual example may illustrate the mechanics and results of this procedure better than the hypothetical situation cited. The problem may be set up as follows:

$$\frac{108.4}{6,668} = \frac{102.9}{x}$$

$$108.4x = 686,137$$

$$x = 6,330$$

where:
108.4 = Winter quarterly index
102.9 = Spring quarterly index
6,668 = 1950 Winter quarter enrollment
x = 1950 Spring quarter enrollment

Using the quarterly index figures and setting up the proper ratios, the calculated estimate of enrollment for the Spring quarter would be 6,330 students. When the actual number of students for the Spring quarter was tabulated, it was 6,330 which indicates that the calculated estimate using the quarterly index gave the exact number enrolled for this quarter. The fact that this estimate was exactly the same as the actual enrollment is not evidence that actual results may always be estimated. The calculated result is the most probable enrollment, but it is very improbable that such an estimate will coincide with the actual enrollment. This is true because the basis of the quarterly index is a moving average.

However, if as in Chapter II an enrollment figure pertains to the total number of different persons enrolled during the regular session, i. e.,

gross enrollment, rather than the quarterly total, it will be impossible to use the index of quarterly variation discussed above. Instead, a new index must be devised where each of the quarterly index figures will be less than 100 percent since in no one quarter will all the different persons be registered. This index of quarterly variation to be applied to gross enrollment may be computed as follows:

Total (1)	Total enrollment Fall quarter (2)	Gross enrollment (3)	Percent of gross enrollment enrolled Fall quarter (Col. 2 ÷ Col. 3) (4)
1937	2,929	3,239	90.43
1938	3,173	3,574	88.78
1939	3,425	3,741	91.55*
1940	3,478	3,850	90.33
1941	3,300	3,569	92.46*
1942	3,091	3,360	91.99*
1943	1,651	1,894	87.17*
1944	1,851	2,237	82.74*
1945	2,477	5,401	45.86*
1946	7,442	8,405	88.54
1947	7,690	8,914	86.26*
1948	7,790	8,893	87.59
		Total of middle five =	445.67
		Modified mean 445.67 ÷ 5 =	89.1

*Eliminated in calculating modified mean.

The typical Fall quarter enrollment is 89.1 percent of the gross enrollment. Thus by taking this calculated figure and relating it to the index of quarterly variation calculated by the moving average method it is possible to find the index figures for the Winter, Spring and Summer quarters. This may be done as follows:

$$\frac{89.1}{117.0} = \frac{w}{108.4}$$

$$117.0w = 9658.4$$

$$w = 82.6$$

where: 117.0 = Fall quarterly index
108.4 = Winter quarterly index
89.1 = Fall index based on gross enrollment
w = Winter index based on gross enrollment

$$\frac{89.1}{117.0} = \frac{y}{102.9}$$

$$117.0y = 9168.4$$

$$y = 78.4$$

102.9 = Spring quarterly index
y = Spring index based on gross enrollment

$$\frac{89.1}{117.0} = \frac{z}{71.7}$$

$$117.0z = 6387.5$$

$$z = 54.6$$

71.7 = Summer quarterly index
z = Summer index based on gross enrollment

The computations above produce the following indexes of quarterly variation based on gross enrollment:

Quarter	Percent
Summer	54.6
Fall	89.1
Winter	82.6
Spring	78.4

Again it is necessary to mention that these index numbers may be used only when they are applied to gross enrollment. In such a situation the practical meaning of the index is this: 89.1 percent of the gross enrollment will be enrolled in the Fall quarter, and so on for the other two quarters.

As an example, suppose that the problem is this: The estimates of gross enrollment at The University of Tennessee in 1970 is 11,038 students based on the trend analysis alone (See page 32.). What will be the number of persons enrolled in the Fall quarter? The solution may be found by taking the Fall index as above, 89.1 percent, and multiplying it by 11,038, i. e., finding 89.1 percent of 11,038. The answer reveals that approximately 9,835 students will be enrolled during the Fall quarter of 1970. This estimate is based only on trend and quarterly variation calculations. Cyclical and irregular effects have not been taken into account.

CHAPTER IV

CYCLICAL AND IRREGULAR VARIATION ANALYSIS

Two types of variation in the University enrollment remain to be measured and explained before the time series analysis is complete. These variations are called cyclical and irregular and are designated by the symbols C and I respectively.

One of the chief purposes of a time series analysis is the isolation of the effects of a cycle on an individual series. It is not possible to isolate cyclical movements completely because of the subjectivity of the measurement of trend, seasonal, and particularly the irregular movements. An inspection of most time series charts, whether they pertain to data concerning the natural sciences, business, or any other field, will reveal wave-like fluctuations recurring at irregular periods and with varying degrees of severity. These movements are called cycles, and the variation is termed cyclical rather than periodic or quarterly because it does not occur with complete regularity as to duration. Also, the movements are cyclical rather than irregular or random because the transition from a low point to a high point, or vice versa, is a progressive development.[1]

Irregular movements, as suggested above, do not show a progressive development but rather are of a random or episodic character. In effect, this type of variation is of a residual nature since it contains the variation not accounted for by trend, season, and cycle. Such movements are

---

[1]John R. Stockton, An Introduction to Business Statistics. (New York: D. C. Heath and Company, 1938), p. 231.

made up of inconsequential fluctuations of a random nature which perhaps have little or no effect on a series, or they may consist of events of an episodic nature which are very important, e. g., war.[2]

It is proper to identify the movements discussed above as they pertain to the University enrollment. A cyclical analysis makes it possible to present adjusted enrollment figures influenced chiefly by cyclical fluctuations. Such an adjusted series makes it possible to see the cyclical fluctuations in University enrollment data expressed more definitely than would be possible without the analysis. The analysis is called the residual method and consists of eliminating from the gross enrollment data, trend and irregular variations. The gross enrollment figures contain only trend, cycle, and irregular movements; consequently, when the trend and irregular movements are removed, the residuals when plotted give a picture of the enrollment cycles.[3]

Cyclical Analysis Used in Forecasting

The knowledge that a series follows a definite cycle is valuable to persons attempting to forecast, particularly if it is possible to predict the turning points of such a cycle and know its average length. Assume that a forecast is being made on the basis of a trend analysis. Obviously since the trend is the "normal" line, the actual data fluctuate above and

---

[2]Frederick E. Croxton and Dudley J. Cowden, Applied General Statistics (New York: Prentice-Hall, Inc., 1939), p. 372.

[3]*Ibid.*, pp. 540-553.

below it. If it is noted that the most recent observations are in the declining phase of the cycle, it would not be logical to use the estimated figure based on trend alone, since the cyclical movement might change the estimate considerably. The estimates for several observations might be considerably lower than the trend estimate, but as the cycle turns upward and crosses the trend line, exactly the converse reasoning would be used in forecasting. Obviously, if cycles had constant lengths and constant amplitudes, it would be possible to predict turning points and make fairly accurate forecasts; however, very few series present cycles that are uniform in every respect. It is necessary then to use the results of cyclical analysis subjectively as well as objectively in forecasting.

The Residual Method Applied to Enrollment Data

The residual method used in this problem to present the cyclical pattern of University enrollment is shown in Table V, and a description is presented below of the step-by-step process.

1. Column 2 lists the gross enrollment for the years 1899 through 1950. Since these are annual figures, they do not contain quarterly (periodic) fluctuations but are made up of trend, cycle, and irregular movements.

2. Column 3 consists of the trend values obtained by fitting a Gomperts growth curve to the gross enrollment. The values are obtained from Table II.

3. Column 4 represents cyclical-irregular movements and is obtained by dividing column 2 (TCI) by column 3 (T). The cyclical-irregular movements are expressed in percentage form since the original data, expressed

TABLE V

THE UNIVERSITY OF TENNESSEE GROSS ENROLLMENT CYCLES AND IRREGULAR VARIATIONS CALCULATED FOR CONTROL CHART TECHNIQUE

Year (1)	Gross enrollment (students) (2)	Trend value (students) (3)	Cyclical irregular movements (percent) (4)	Cyclical relatives three-year binomial moving average of column 4 (5)	Irregular variations (Col. 4 - Col.5) (6)	Three-year moving range of column 6 (7)
1899	266	220	120.9			
1900	343	241	142.3	135.9	6.4	
1901	364	263	138.4	133.8	4.6	11.8
1902	335	287	116.2	121.6	-5.4	12.9
1903	412	313	131.6	124.1	7.5	12.9
1904	364	340	107.0	110.3	-3.3	10.8
1905	353	369	95.6	98.7	-3.1	3.5
1906	388	401	96.7	96.9	.2	3.3
1907	428	434	98.6	100.5	-1.9	2.1
1908	509	470	108.2	109.0	-8.0	12.1
1909	616	508	121.2	111.0	10.2	11.1
1910	514	548	93.7	94.6	- .9	17.8
1911	414	592	69.9	77.5	-7.6	10.8
1912	489	638	76.6	73.4	3.2	10.8
1913	486	686	70.8	72.5	-1.7	5.6
1914	531	738	71.9	74.3	-2.4	3.8
1915	657	792	82.9	81.5	1.4	4.5
1916	752	850	88.4	86.3	2.1	2.0
1917	781	911	85.7	82.3	3.4	3.5
1918	679	975	69.6	71.0	-1.4	12.7

TABLE V (continued)

THE UNIVERSITY OF TENNESSEE GROSS ENROLLMENT CYCLES AND IRREGULAR VARIATIONS CALCULATED FOR CONTROL CHART TECHNIQUE

Year (1)	Gross enrollment (students) (2)	Trend value (students) (3)	Cyclical irregular movements (percent) (4)	Cyclical relatives three-year binomial moving average of column 4 (5)	Irregular variations (Col. 4 - Col.5) (6)	Three-year moving range of column 6 (7)
1919	619	1,043	59.3	68.6	-9.3	16.3
1920	962	1,114	86.4	79.4	7.0	16.3
1921	1,020	1,189	85.7	87.6	-1.9	8.9
1922	1,178	1,268	92.9	92.1	.8	2.7
1923	1,313	1,350	97.2	97.2	.0	.8
1924	1,459	1,437	101.5	101.5	.0	1.9
1925	1,620	1,528	106.0	104.1	1.9	6.8
1926	1,675	1,622	103.2	108.1	-4.9	6.8
1927	2,064	1,722	119.8	119.6	.2	8.4
1928	2,479	1,825	135.8	132.3	3.5	3.3
1929	2,670	1,933	138.1	137.5	.6	3.4
1930	2,824	2,046	138.0	134.0	4.0	5.8
1931	2,642	2,164	122.0	123.8	-1.8	5.8
1932	2,591	2,286	113.3	110.5	2.8	4.6
1933	2,259	2,414	93.5	95.0	-1.5	8.0
1934	2,030	2,546	79.7	84.9	-5.2	4.5
1935	2,336	2,693	86.7	87.4	- .7	6.5
1936	2,731	2,826	96.6	95.3	1.3	2.0
1937	3,019	2,973	101.5	100.8	.7	2.0
1938	3,239	3,126	103.6	104.3	- .7	2.1

TABLE V (continued)

THE UNIVERSITY OF TENNESSEE GROSS ENROLLMENT CYCLES AND IRREGULAR VARIATIONS CALCULATED FOR CONTROL CHART TECHNIQUE

Year (1)	Gross enrollment (students) (2)	Trend value (students) (3)	Cyclical irregular movements (percent) (4)	Cyclical relatives three-year binomial moving average of column 4 (5)	Irregular variations (Col. 4 - Col.5) (6)	Three-year moving range of column 6 (7)
1939	3,574	3,285	108.7	107.3	1.4	2.1
1940	3,741	3,449	108.4	107.9	.5	2.1
1941	3,850	3,618	106.4	103.8	2.6	3.2
1942	3,569	3,793	94.1	94.7	-.6	10.6
1943	3,360	3,974	84.5	77.1	7.4	18.6
1944	1,894	4,160	45.4	56.6	-11.2	22.7
1945	2,237	4,352	51.4	66.7	-15.3	17.6
1946	5,401	4,550	118.7	116.4	2.3	29.2
1947	8,405	4,752	176.8	162.9	13.9	11.6
1948	8,914	4,963	179.6	176.9	2.7	11.6
1949	8,893	5,178	171.7	169.4	2.3	
1950	8,360	5,399	154.8			
Totals					13.3*	398.2*

*Used in control chart technique

Source: Tables I and II.

in students, are divided by trend, expressed in students, i.e., $\frac{TCI}{T} = CI$. Figure 4 shows by the dashed line the cyclical-irregular movements.

4. Although the dashed line in Figure 4 is regular enough to show the wave-like movements, there are still irregularities that should be eliminated if the cyclical movement is to be represented adequately. Although no completely satisfactory method has been found to eliminate irregularities, it is possible by the use of a binomially-weighted moving average to smooth the curve so as to bring the cyclical movements into clearer relief. Column 5 then, is obtained by calculating a three year moving average using weights of 1 - 2 - 1. This tends to weight the central item heaviest but also gives weight to the years on either side of the middle year. By this method the effect of an unusually low or high enrollment, the result of some unusual occurence, is reduced. Results are shown as percentage figures in column 5 and by the solid line in Figure 4.[4]

Harmonic Analysis Applied to Enrollment Cycles

After obtaining the enrollment cyclical movement by the residual method an alternative procedure is to fit a mathematical curve to the data. Such a procedure is seldom appropriate since the following conditions are desirable before the harmonic analysis is used:

1. Cycles to be a variety of periodic movement
2. Cycles to be similar in pattern
3. The pattern capable of description by a mathematical equation

---

[4]*Ibid.*, pp. 540-553.

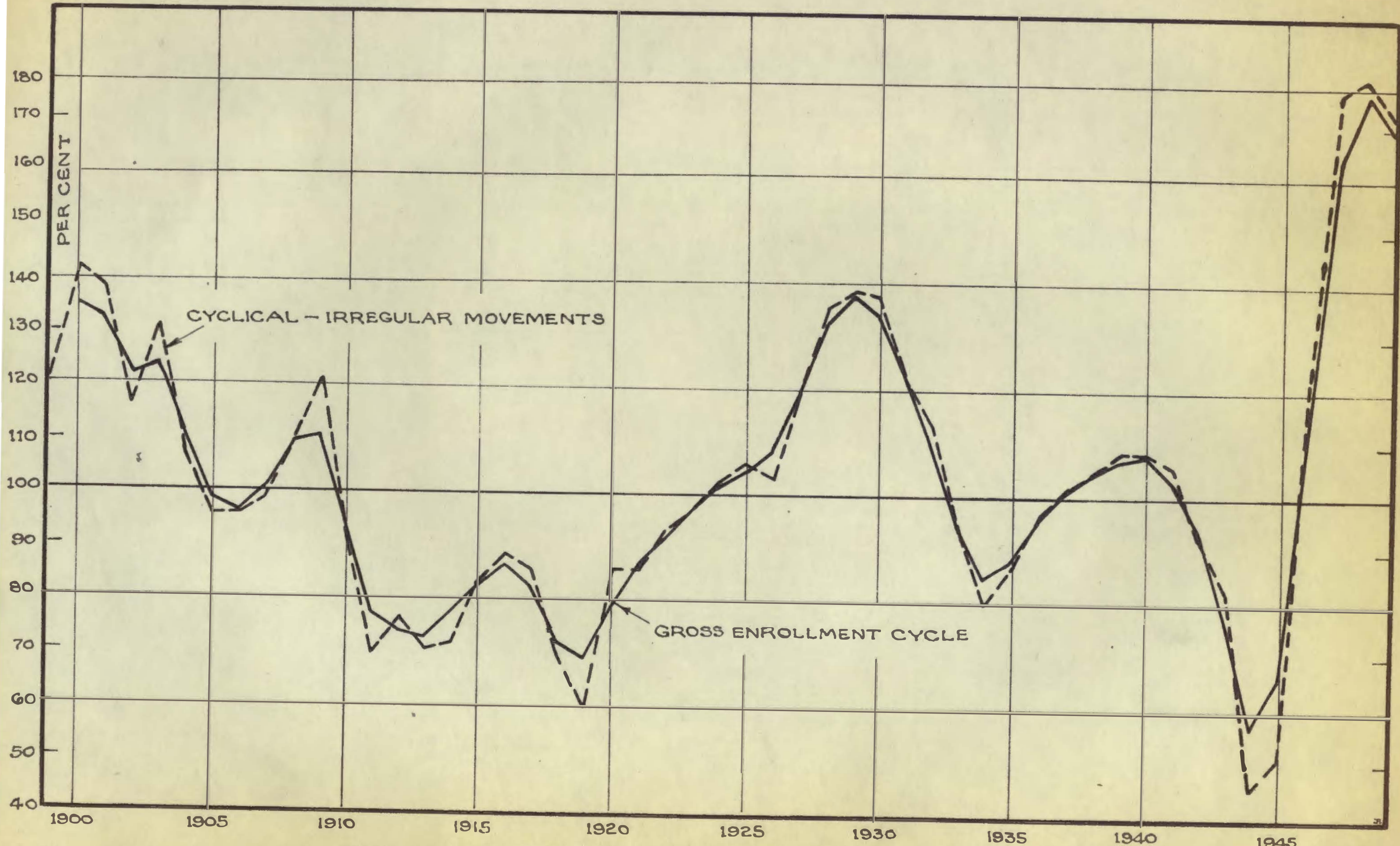

Figure 4. Cycles of The University of Tennessee gross enrollment, 1899 – 1949

Source of Data: Table V

Most economic series are not periodic, and are often described inadequately by mathematical curves; however, University enrollment seems to exhibit a fairly regular ten-year cycle within a thirty-year cycle and will be subjected to harmonic analysis.[5]

The procedure may be divided into two steps. First, an attempt is made to discover the periodicity of the data and its average cyclical pattern by means of <u>periodogram</u> analysis. Second, a <u>periodic curve</u> is fitted to the average pattern found in periodogram analysis and applied to the enrollment series.

Periodogram Analysis of Gross Enrollment Cycles

Discovering if the enrollment cycles possess the quality of periodicity is the first objective. Since the statement has been made above that a fairly regular ten-year cycle seems to exist, the data should be tested for such periodicity. In Table VI, the data of Table V, column 5, the cyclical relatives, are arranged in rows of ten successive items with column 1 containing the years 1909, 1919, 1929, 1939, and 1949 reading down the column. In the same manner, column 2 contains the years 1900, 1910, 1920, 1930, 1940, and the other eight columns are arranged in the same order in such a way that it is possible to obtain the average of each column. The lowest arithmetic average of 85.5 percent when subtracted from the highest, 118.8 gives a range from low year to high year of 33.3 percentage points.

---

[5] <u>Ibid.</u>, pp. 560-561

TABLE VI

PERIODOGRAM ANALYSIS OF THE UNIVERSITY OF TENNESSEE GROSS ENROLLMENT CYCLES
(Cycle movements in percent)

Measure	1	2	3	4	5	6	7	8	9	10	Total
		135.9	133.8	121.6	124.1	110.3	98.7	96.9	100.5	109.0	
	111.0	94.6	77.5	73.4	72.5	74.3	81.5	86.3	82.3	71.0	
	68.6	79.4	87.6	92.1	97.2	101.5	104.1	108.1	119.6	132.3	
	137.5	134.0	123.8	110.5	95.0	84.9	87.4	95.3	100.8	104.3	
	107.3	107.9	103.8	94.7	77.1	56.6	66.7	116.4	162.9	176.9	
	169.4										
Average pattern	118.8 H	110.3	105.3	98.4	93.1	85.5 L	87.6	100.6	113.2	118.7	1,031.5
Adjusted pattern (y values)*	115.2	106.9	102.1	95.4	90.3	92.9	84.9	97.5	109.7	115.1	1,000.0

*Average values multiplied by .96946 = 1000.0 ÷ 1031.5
H = high. L = low.

Source: Table V.

The procedure is repeated with an arrangement of nine successive items in a row. The table for this and the other four arrangements used is not presented here, but the procedure is easily followed. For nine items the last number in the first row is 100.5 which is the cyclical relative for the year 1907. Computation gives a range of means of 28.5 in this arrangement. Repeating again with different assumptions as to periodicity, the above procedure is followed until the true periodicity is discovered. This is considered to be the periodicity that gives the greatest range of column means. From six different trials the following results are recorded and are presented graphically as a periodogram in Figure 5.

Assumed periodicity (years)	Range of column means (percent)
8	21.0
9	28.5
10	33.3
11	23.4
12	20.4
13	12.4

The column means of Table VI are now accepted as numerical values for the average cyclical behavior of the University enrollment as concerns the ten-year cycle. Conforming with the refinement method used in the computation of quarterly indexes, page 45, the column means are adjusted to average 100 percent. These refined averages are in the last row of Table VI and are shown by the solid line of Figure 6.[6]

Sufficient years are not available to do a periodogram analysis of the suggested 30 year cycle; however, close observation of the enrollment

---

[6]Ibid., pp. 554-558.

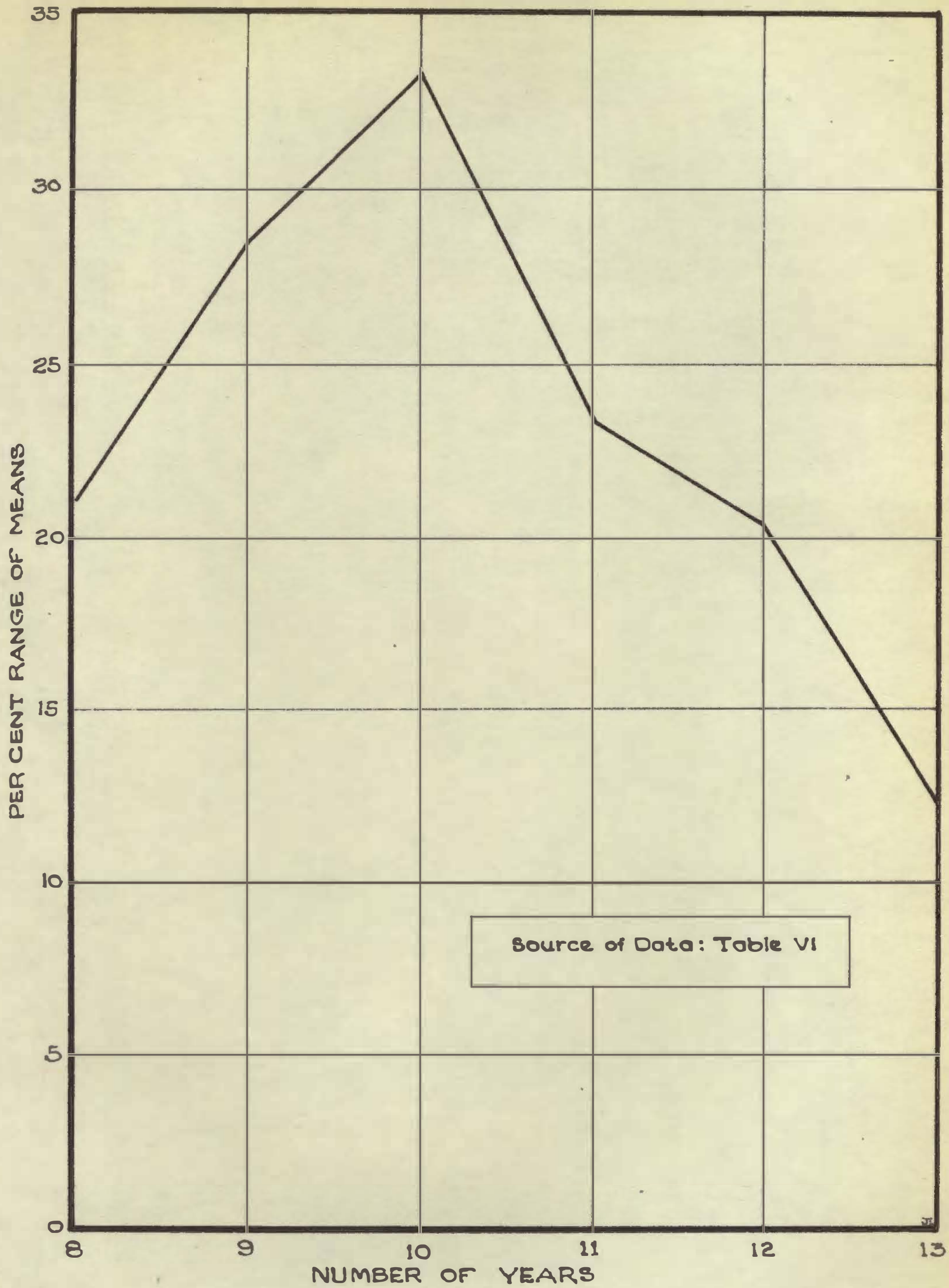

Figure 5. Periodogram of cyclical movements of The University of Tennessee gross enrollment, 1899 — 1949

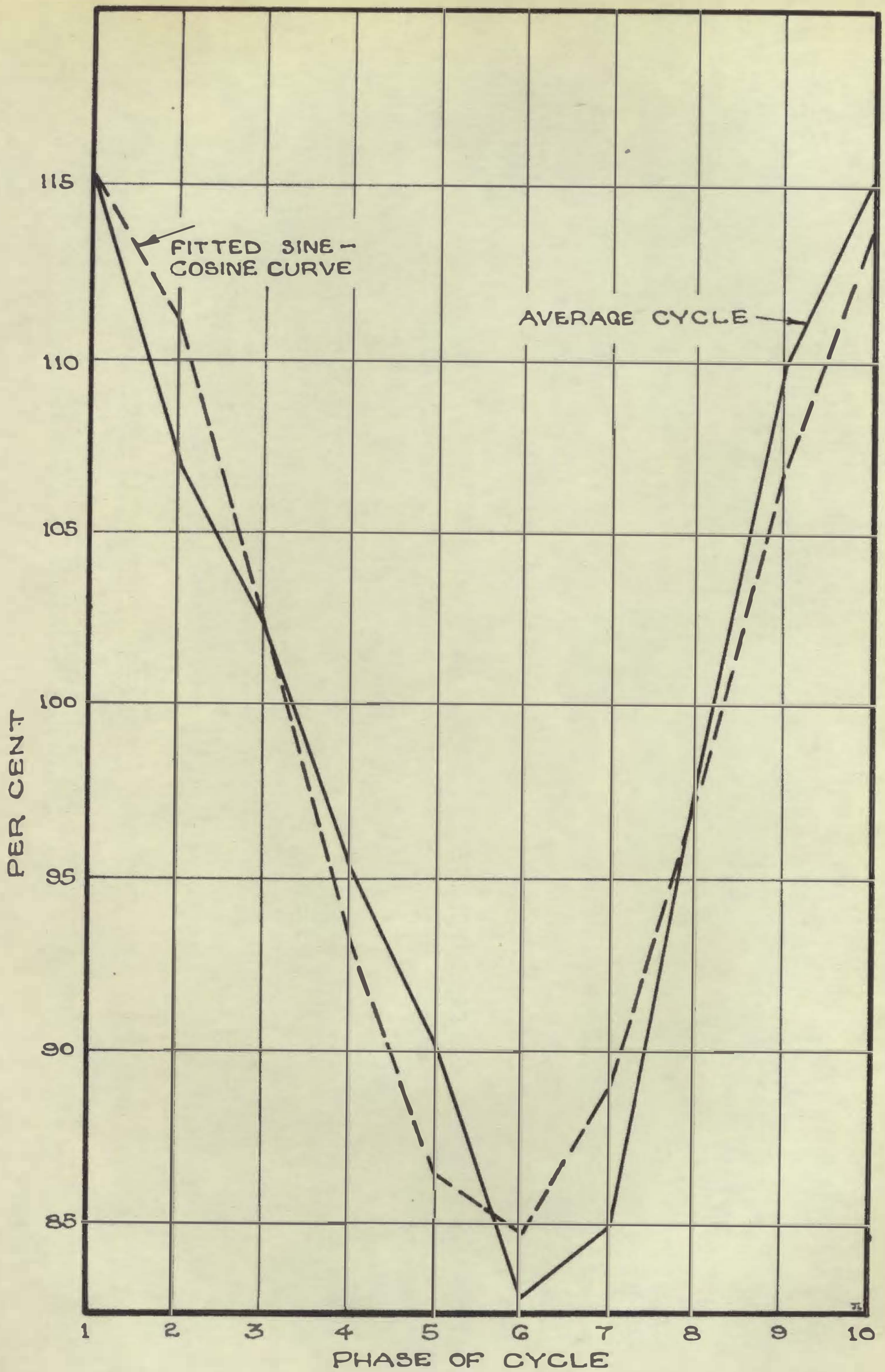

Figure 6. Average cyclical pattern of The University of Tennessee ten-year gross enrollment, 1889-1949, and sine-cosine curve

Source of Data: Table VII

cycles plotted in Figure 4, indicates that such a cycle does exist. From a peak in 1899 to the next major peak in 1929, thirty years later, the cycle reaches a trough about 1914 which shows approximately 15 years of decline and then 15 years of upswing. This thirty-year cycle contains three of the ten-year cycles analyzed above. From the 1929 peak, the curve follows substantially the same patterns as in the first thirty-year cycle, i. e., it reaches the trough in 1944 which shows 15 years in the negative phase. From 1944 the up-swing in the actual cyclical relatives far exceeds expectations based on the thirty-year cycle; however, the effect of the episodic irregular variation, World War II (discussed on page 75 is apparently responsible for this difference.

Periodic Curve Fitted to Ten-Year Enrollment Cycle

It is now possible to fit a sine-cosine curve to the average ten year cyclical pattern which is represented by the Y_c values in Table VII and were obtained from Table VI.

The curve is of the type $Y_c = \bar{Y} + A \sin \left(\frac{360}{T} X\right)^{\circ} + B_{\cos} \left(\frac{360}{T} X\right)^{\circ}$

Y_c = Ordinate values of fitted sine-cosine curve

X = Phase of cycle in years

T = Periodicity in years

$$A = \frac{2}{T} \Sigma \left[Y \sin \left(\frac{360}{T} X\right)^{\circ} \right]$$

$$B = \frac{2}{T} \Sigma \left[Y \cos \left(\frac{360}{T} X\right)^{\circ} \right]$$

TABLE VII

PERIODIC CURVE FITTED TO THE UNIVERSITY OF TENNESSEE TEN-YEAR GROSS ENROLLMENT CYCLES (T = 10)

X	36X	sin(36X)°	cos(36X)°	Y	Y sin(36X)°	Y cos(36X)°	A sin(36X)°	B cos(36X)°	Y_c [100 + Col. B + Col. 9]
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	36	.5878	.8090	115.2	67.71	93.20	4.22	10.99	115.21
2	72	.9511	.3090	106.9	101.67	33.03	6.83	4.20	111.03
3	108	.9511	-.3090	102.1	97.11	-31.55	6.83	- 4.20	102.63
4	144	.5878	-.8090	95.4	56.08	-77.18	4.22	-10.99	93.23
5	180	.0000	-1.0000	90.3	0.00	-90.30	0.00	-13.60	86.40
6	216	-.5878	-.8090	82.9	-48.73	-67.01	-4.22	-10.99	84.79
7	252	-.9511	-.3090	84.9	-80.75	-26.23	-6.83	- 4.20	88.97
8	288	-.9511	.3090	97.5	-92.73	30.13	-6.83	4.20	97.37
9	324	-.5878	.8090	109.7	-64.48	88.75	-4.22	10.99	106.77
10	360	-.0000	1.0000	115.1	- 0.00	115.10	-0.00	13.60	113.60
				1000.0	35.88	67.94			1000.00

Source: Y values from Table VI.

From the periodogram analysis it is known that $T = 10$ and $\bar{Y} = 100$, consequently the above formula becomes

$$Y_c = 100 + A \sin (36X)^\circ + B \cos (36X)^\circ$$

$$A = \frac{\Sigma \left[Y \sin (36X)^\circ \right]}{5}$$

$$B = \frac{\Sigma \left[Y \cos (36X)^\circ \right]}{5}$$

Table VII shows the calculations involved in fitting this curve. The X values in column 1 start with one and continue through ten and refer to the stages or phases of the cycle. The sines and cosines used in columns 3 and 4 may be obtained from any table of sines and cosines. The Y values, as mentioned above, are the average values taken from the last row of Table VI. Columns 6 and 7 are added to obtain the numerator of the A and B formulae respectively and the values, when substituted, become:

$$A = \frac{35.88}{5} = 7.176$$

$$B = \frac{67.94}{5} = 13.588$$

The equation then is:

$$Y_c = 100 + 7.176 \sin (36X)^\circ + 13.588 \cos (36X)^\circ$$

Columns 8, 9, and 10 are self-explanatory.[7]

In order to ascertain to what extent the mathematical values calculated in column 10 describe the average cyclical relatives in column 5,

---

[7] *Ibid.*, pp. 559-560.

the two series were plotted as Figure 6, where the broken line represents the fitted curve. The fit appears to be very good and leaves little to be desired; however, in Figure 7, where the fitted curve, again represented by the broken line, is fitted to the cyclical relatives for the 49 year period, the result is not as desirable. The peaks and troughs of the fitted curve correspond adequately enough with the peaks and troughs of the cyclical relatives as concerns periodicity; however, the amplitude of the fitted curve does not reach high enough nor dip low enough in certain years. The peak of the ten-year cycle shows an amplitude of 15.21 percent as does the trough, or a range of 30.42 percent (115.21 - 84.79).

Periodic Curve Fitted to Thirty-Year Cycle

The above criticism as to amplitude may be rectified by fitting a cosine curve of amplitude equal to the amplitude of the ten year sin-cosine curve to the thirty year cycle. The procedure is described briefly below and may be easily followed by reference to Figure 8.

In this figure the value of the ordinate at 1899 and 1929 is 115.21 percent, and at 1914 and 1944 the value is 84.79 percent. To these points of equal amplitude on either side of the 100 percent line, a cosine curve is fitted by the use of the following equation:

$$Y'_c = \bar{Y} + 15.21 \cos \frac{360}{T} X$$

where

Y'_c = Ordinate value for fitted cosine curve

X = Phase of cycle in years

T = Periodicity in years

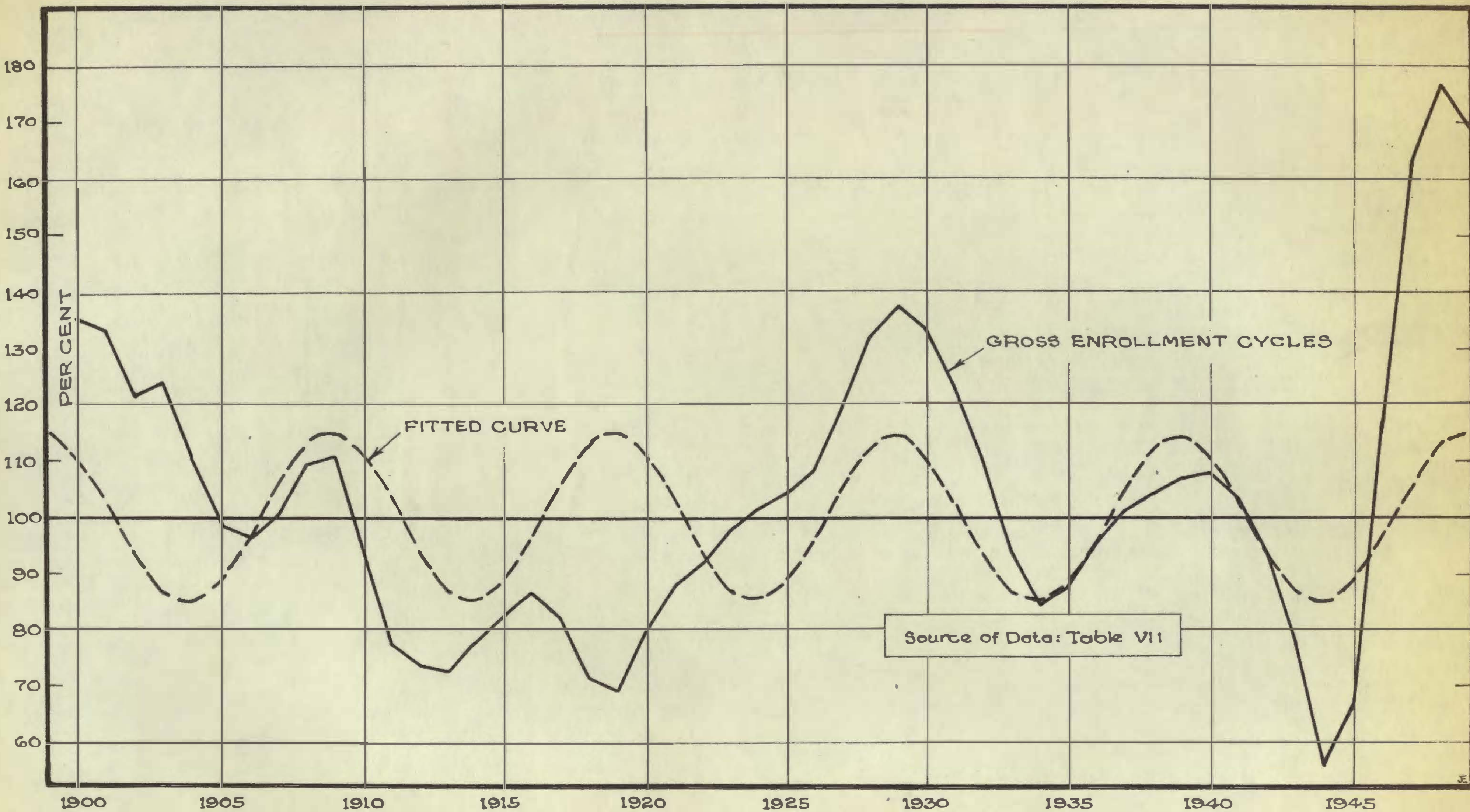

Figure 7. Ten-year sine-cosine curve fitted to The University of Tennessee gross enrollment cycles, 1899-1949

$\frac{360}{T}$ = Period in degrees per year

15.21 = Amplitude in percent

When the appropriate values are substituted in the above equation, it becomes:

$$Y'_c = 100 + 15.21 \cos 12X$$

Substituting for X any phase of the cycle, the above equation may be solved for Y'_c, the ordinate value. In Figure 8, a sufficient number of ordinate values were plotted to produce the dashed line which represents the cosine curve fitted to the thirty year cycle.

Synthesis of the Ten and Thirty-Year Cycles

It is now possible to obtain a curve that fits the data both as to periodicity and amplitude by synthesizing the two fitted periodic curves. This procedure consists of adding the deviations of the ordinates from the baseline (Y_c - 100) of the ten-year sine-cosine curve to the ordinate values (Y'_c) of the thirty-year cosine curve for each of the thirty years in a cycle. This procedure is followed and the fitted cyclical relatives are shown in Table VIII for the resulting curve. Only one complete cycle (1899-1928) is shown in Table VIII since the pattern produced by the two curves will repeat. Column 2 is obtained by taking the algebraic difference between the ordinate values of the ten year sine-cosine curve and the baseline of one hundred percent (Y_c - 100). These ordinate values came from column 10, Table VII. The ordinate values (Y'_c) of the thirty-year cosine curve are obtained by solving the equation in the previous section for

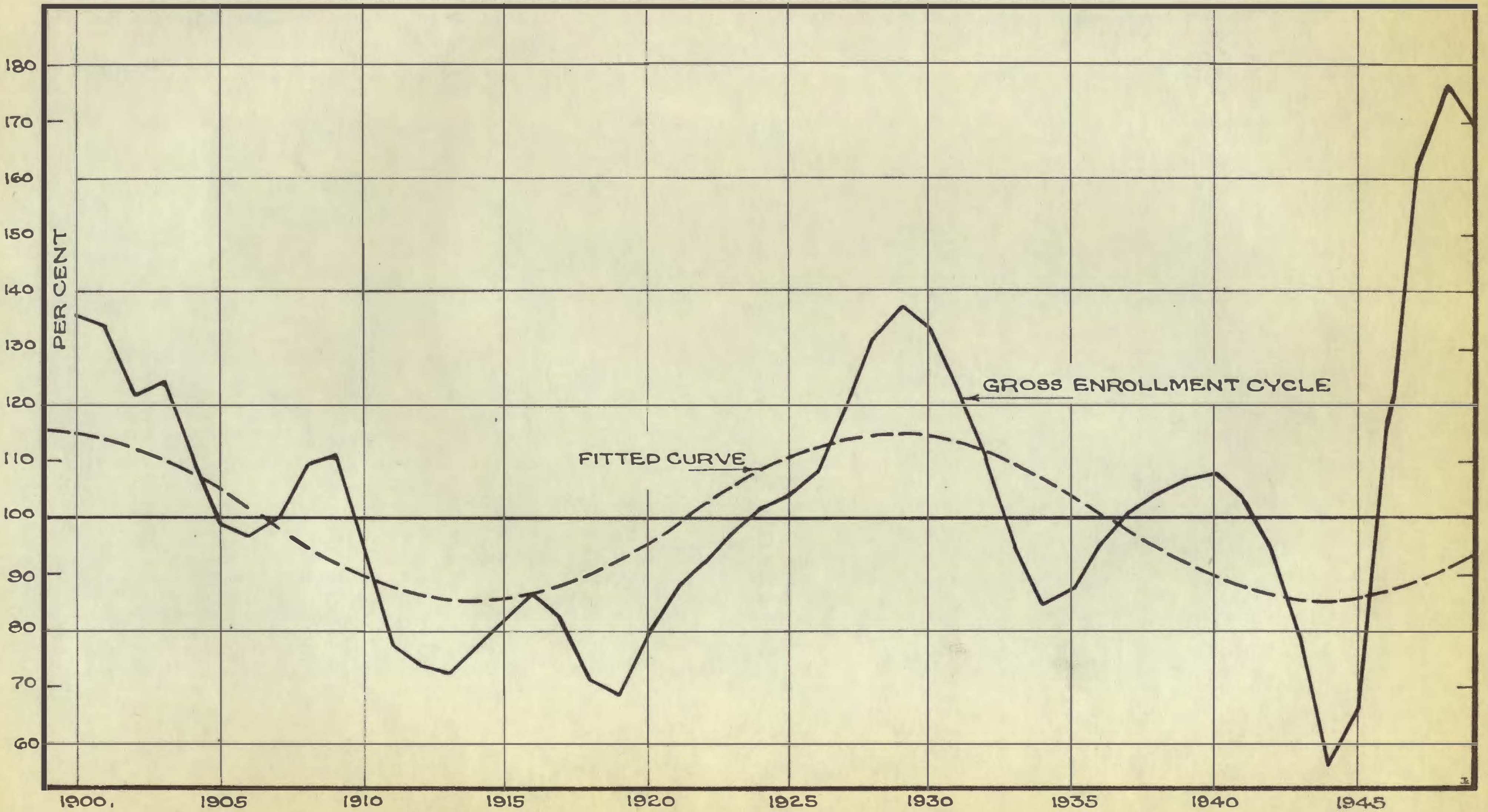

Figure 8. Thirty-year cosine curve fitted to The University of Tennessee gross enrollment cycles, 1899 – 1949

Source of Data: Adjacent pages and Table V

TABLE VIII

SYNTHESIZED TEN AND THIRTY-YEAR PERIODIC CURVE FITTED TO FIRST THIRTY YEARS OF THE UNIVERSITY OF TENNESSEE GROSS ENROLLMENT CYCLES
(Y_c and Y_c' as defined on pp. 65 and 72)

Year (1)	$Y_c - 100$ (2)	Y_c' (3)	Fitted cyclical relatives (Col. 2 + Col. 3) (4)
1899	15.21	115.21	130.42
1900	11.03	115.00	126.03
1901	2.63	114.00	116.63
1902	- 6.77	112.50	105.73
1903	-13.60	110.30	96.70
1904	-15.21	107.60	92.39
1905	-11.03	105.00	93.97
1906	- 2.63	102.00	99.37
1907	6.77	98.00	104.77
1908	13.60	95.00	108.60
1909	15.21	82.40	107.61
1910	11.03	89.70	100.73
1911	2.63	87.50	90.13
1912	- 6.77	86.00	79.23
1913	-13.60	85.00	71.40
1914	-15.21	84.79	69.58
1915	-11.03	85.00	73.97
1916	- 2.63	86.00	83.37
1917	6.77	87.50	94.27
1918	13.60	89.70	103.30
1919	15.21	82.40	107.61
1920	11.03	95.00	106.03
1921	2.63	98.00	100.63
1922	- 6.77	102.00	95.23
1923	-13.60	105.00	91.40
1924	-15.21	107.60	92.39
1925	-11.03	110.30	99.27
1926	- 2.63	112.50	109.87
1927	6.77	114.00	120.77
1928	13.60	115.00	128.60

Source: Table VII.

values of X, one through thirty. The fitted cyclical relatives in column 4 result from an algebraic addition of columns 2 and 3 and represent the cyclical factor which will be applied to the enrollment estimates.

The combination ten-year and thirty-year curve is shown as a dashed line in Figure 9 where it may be compared with the enrollment cycles, a solid line. Such a comparison of the two curves reveals that the fitted curve is strikingly similar to the enrollment cycles with but two exceptions. In 1919, the fitted curve reaches a peak while enrollment cycles dropped to a trough. This is the result of the effect of World War I on student enrollment. (See pages 74 and 75.) From about 1913 to 1916, the enrollment cycle turned upward and was following the ten-year cycle; however, in 1917 a decline set in which lasted until the war ended in 1919. World War I was not as long nor did it affect University enrollment as much as did World War II; however, enrollment was not normal for a period extending from 1917 to about 1924.[8]

Again in the years from 1943 to 1950, and possibly until 1955, a difference may be noticed in the two curves. This situation has resulted from the effects of World War II, and may be considered a period during which an irregular variation of an episodic nature is the controlling factor. Irregular variations are discussed later in this chapter. It may be noted that even in this war period, the fitted curve compares favorably with the enrollment cycles as concerned periodicity; however, the 1944 trough and

---

[8]Edward R. Dewey and Edwin F. Dakin, Cycles, the Science of Prediction (New York: Henry Holt and Company, Inc., 1947), pp. 160-172.

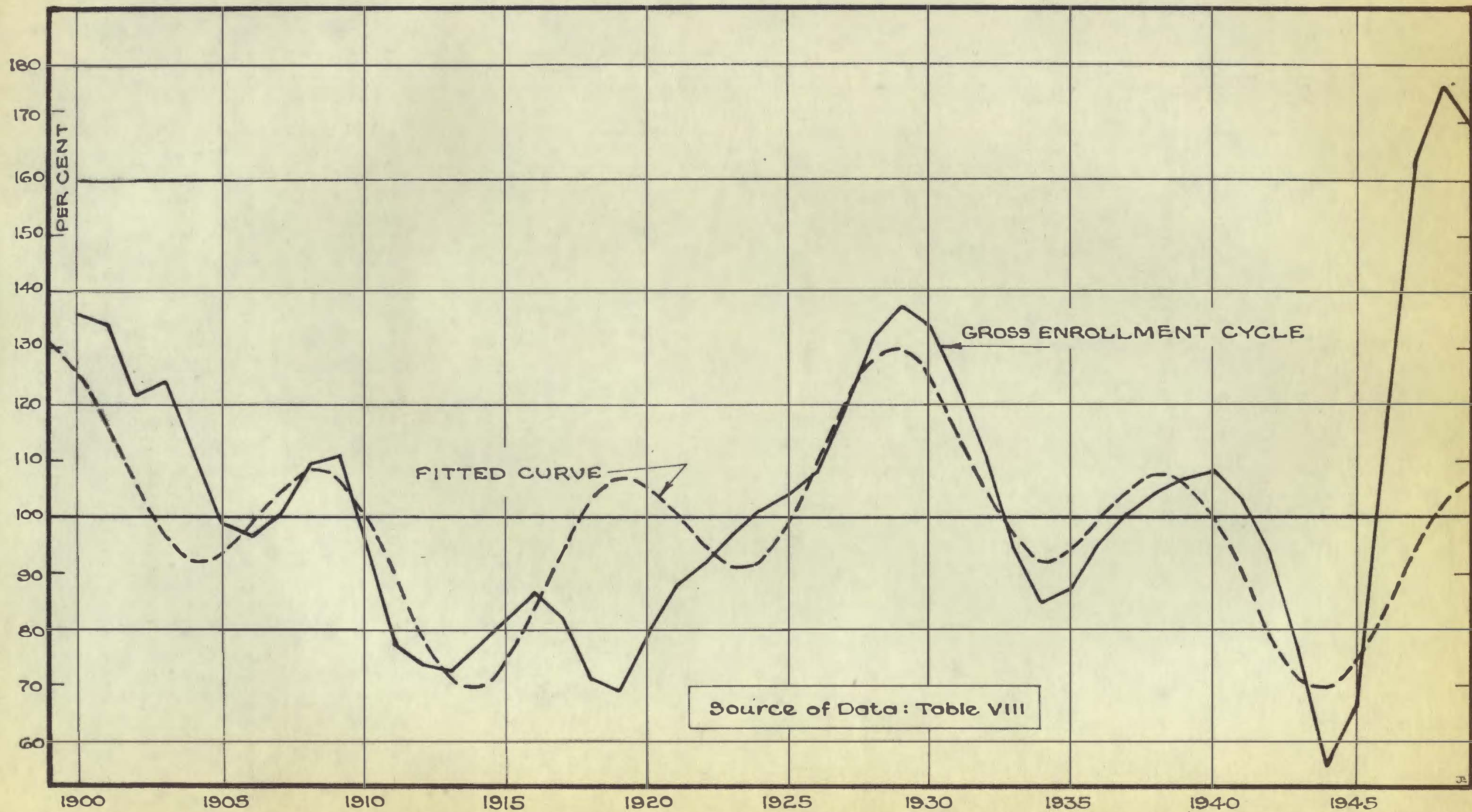

Figure 9. Synthesized ten and thirty-year curve fitted to The University of Tennessee gross enrollment cycles, 1899 – 1949

the 1948 peak far exceed the expectations set by the fitted curve. It is entirely possible that the effects of World War II, i. e., increased enrollment, will be evident until the G. I. benefits expire in 1955. If enrollment under G. I. benefits is not accelerated by unforseen events, it is probable that the number of veterans attending the University after 1953 or 1954 will be negligible.[9]

It is essential that the condition brought about by World War II be considered since it is an irregular variation which will have an effect on the future enrollment estimates. A discussion of irregular movements in the University enrollment data is presented briefly in the pages to follow, while the last part of the chapter will be devoted to a forecast of the University enrollment which will be based on a synthesis of the time series analysis.

Control Chart Technique Applied to Irregular Variation

The residual method of cyclical analysis ended with the smoothed line in Figure 4, representing the cyclical movement of the enrollment data; however, further techniques may be applied to classify the irregular movements as either random or episodic. Random movements are usually minor fluctuations while episodic movements consist of a set of events that stand out from others as a particular time and whose effects are cumulative over a period of time.

---

[9]A veteran is defined as a civilian who served as a member of the armed forces of the United States on active duty at any time between September 16, 1940, and July 25, 1947.

By the use of a dot chart, a chart for individual observations, it is possible to segregate the irregularities in the enrollment data which are normally expected, i. e., minor random fluctuations, from those irregular movements of an episodic nature. The process will be described briefly in this thesis, but full explanation of this technique may be obtained from E. L. Grant's Statistical Quality Control.

Unbiased Estimate of Universe Standard Deviation

The computations for much of this phase of the problem are shown in columns 6 and 7 of Table V. Column 6 contains the irregular variations which were obtained when the cyclical variations were subtracted from cyclical-irregular variations. The purpose of this phase of the analysis will be to segregate these irregular variation measurements into episodic and random classifications. The segregation technique is based upon the deviation of each of these measurements from the average measurement. Practically, this means that the magnitude of the measurement in relation to the magnitude of the average measurement determines whether it is an episodic or random irregularity. This concept necessitates the use of some measure of scatter or deviation.

The measure of scatter or deviation used in the dot chart is the universe standard deviation (σ). If an unbiased estimate of the universe standard deviation is obtained and the values $\bar{X}$ (average measurement) $\pm$ σ are computed, 68 percent of the measurements will be expected to fall in

this range, assuming normality. Consequently 32 percent of the measurements are expected to be outside this range. If now $\bar{X} \pm 2\sigma$ limits are calculated, approximately five percent of the measurements are expected to lie outside these limits. If a measurement actually falls outside the limits, an attempt is made to find an assignable reason, since the probability is so small (.05) that it occurred as the result of chance alone.

Applying the above concept to the enrollment irregularities, it is first essential to find the mean ($\bar{X}$) for the irregular variation. Column 6, Table V, when summed and divided by the number of observations gives an average of .27. This is plotted on Figure 10 as the central line. The next step is the securing of an unbiased estimate of the universe standard deviation. It is not possible in this thesis to discuss the theory supporting the methods and formulae used in obtaining this estimate; however, the procedure will be explained briefly.

In column 7 of Table V, a three year moving range is obtained. The sum of the ranges, when divided by the number of ranges $\left(\frac{398.2}{48}\right)$ yielded an average range ($\bar{R}$) of 8.295. Using the formula UCL_R (upper control limit) = $D_4\bar{R}$, and inserting the appropriate values, the result is $UCL_R = 2.57$ (8.259) = 21.32 where D_4 is a factor for obtaining control limits based on the number of observations in the moving range.[10] The lower control limit for sub-groups of size three is zero. It is possible now by comparing the individual ranges in column 7 to the upper control limit (21.32) to observe

---

[10]E. L. Grant, Statistical Quality Control (New York: McGraw-Hill Book Company, Inc., 1946), p. 536.

that two measurements fall outside this control limit. They are 22.7 in 1944 and 29.2 in 1946. These two ranges are considered heterogeneous and a new range value excluding them, is calculated.

This yields a revised $\bar{R}$ of 7.53, and a revised upper control limit 19.35, i. e. (2.75) (7.53). The 46 ranges used in the revision are all contained within the revised limits; consequently, the new $\bar{R}$ values may be used in the formula, $\frac{\bar{R}}{d_2}$, to obtain an unbiased estimate of the universe standard deviation (σ), where d_2 is a factor based on subgroup size and may be obtained from tables in quality control texts.[11] Substituting in the above formula the unbiased estimate of the universe standard deviation is 4.46, i. e., $\left(\frac{7.53}{1.69}\right)$.

The Dot Chart Applied to Irregular Variations

Using the mean, .27, and the estimate of σ calculated above, a dot chart with 2 σ limits is constructed in Figure 10. The limits of $\bar{X} \pm 2\sigma$ becomes $.27 \pm 2\ (4.46)$ and when solved give an upper limit of 9.19 and a lower limit of -8.65. The measurements of irregular variation (column 6 Table V) for each year are plotted on the dot chart in Figure 10. The use of 2 σ limits in charting economic data implies the assumption that a measurement falling outside these limits is different from those which fall inside. The practical meaning of the dot chart then as applied to measurements of irregular variation is that dots falling outside the limits in

[11] *Ibid.*, p. 157.

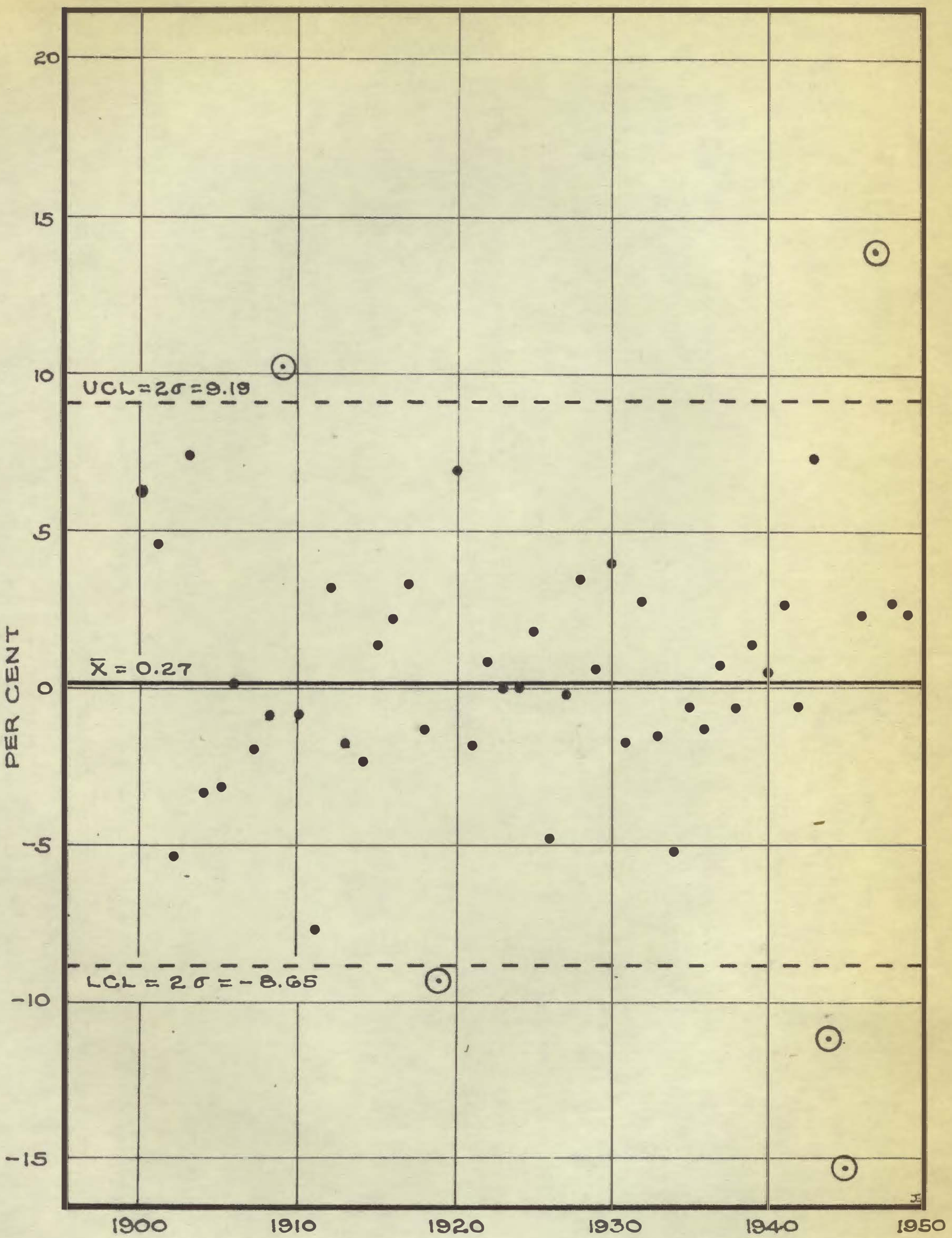

Figure 10. Irregular variations of The University of Tennessee gross enrollment data, 1899-1949

Source of Data: Adjacent pages and Table V

Figure 10 would fall there only five times in one hundred due to the chance factor alone; consequently there must be an assignable cause for such a situation. The measurements of irregular variation falling outside the limits in 1909, 1919, 1944, 1945, and 1947 are then termed episodic.

A historical analysis of enrollment at the University shows that there were assignable causes for the "out of control" conditions evidenced by the five measurements of irregular variation. In 1909 the episodic irregularity may be explained by the College of Agriculture which had an enrollment of 197 students as compared with 139 in 1908 and 40 in 1910. President Emeritus of the University, James D. Hoskins, states that the College of Agriculture gave a series of short courses which resulted in the violent fluctuation of the enrollment figures during the years cited. The records substantiate this explanation and show that during this year P. P. Claxton made an active canvass of the entire state for University students. As was mentioned earlier, World War I was the event that caused the episodic irregularity in 1919. In 1944, 1945 and 1947 again a war, World War II, is credited with causing the episodic irregularities.

The measurements that fall within the limits of the dot chart may be considered random; however, no statistical test gives the positive assurance that no assignable causes of variation are present.[12] Except for the irregular variation of an episodic nature explained above, the dot chart shows a general "in-control" condition. Consequently, the binomially-weighted moving average which produced the cyclical relatives (column 5 of Table V)

---

[12] *Ibid.*, pp. 148-180.

may be considered an adequate representation of the gross enrollment cycles as pictured graphically in Figure 4 and analyzed earlier in the chapter.

University Enrollment Estimates Through 1970

Trend, seasonal, cyclical and irregular variations have been analyzed as factors influencing The University of Tennessee enrollment. The knowledge gained from each analysis may be synthesized into an objective forecast of University enrollment. In Table IX a forecast has been made covering the period from 1951 through 1970. It must be noted, however, that the estimates of enrollment for the year 1951 through 1955 do not account for veterans who may be enrolled during those years. In other words the postwar displacement, though recognized as an episodic irregularity which may and probably will prolong the positive phase of the cycle, has not been considered in the estimates. It is probable that by 1955 all of the effects of World War II as concerns veteran enrollment at The University of Tennessee will have disappeared. Assuming that farther episodic irregularities do not occur, it will be possible to use the estimates as approximations subject to inconsequential irregular fluctuations of a random nature after 1955.

A brief description may serve to clarify the calculations in Table IX. In column 1 the years to be forecast are listed through 1970. Column 2 contains the extrapolated values of the Gompertz growth curve. This curve was fitted to the gross enrollment data in Chapter II. In column 3 the synthesized ten and thirty-year periodic curve fitted to the gross enrollment cycles in Chapter IV is projected into the future by placing the appropriate

TABLE IX

FALL QUARTER ENROLLMENT ESTIMATES FOR THE UNIVERSITY OF TENNESSEE, 1951-1970

Year (1)	Extrapolated trend (Gross enrollment) (2)	Fitted cyclical relatives (3)	Trend and cycle (Gross enrollment) (Col. 2 x Col. 3) (4)	Estimate for Fall quarter (Col. 4 x 89.1) (5)
1951	5,622	100.63	5,657	5,040
1952	5,859	95.23	5,580	4,972
1953	6,097	91.40	5,573	4,966
1954	6,217	92.39	5,744	5,118
1955	6,593	99.27	6,545	5,832
1956	6,849	109.87	7,525	6,705
1957	7,111	120.77	8,588	7,652
1958	7,379	128.60	9,489	8,455
1959	7,653	130.42	9,981	8,893
1960	7,933	126.03	10,109	8,927
1961	8,218	116.23	9,552	8,511
1962	8,518	105.73	9,006	8,024
1963	8,806	96.70	8,515	7,587
1964	9,109	92.39	8,416	7,499
1965	9,417	93.75	8,828	7,866
1966	9,730	99.39	9,669	8,615
1967	10,049	104.77	10,528	9,380
1968	10,372	108.60	11,264	10,036
1969	10,702	107.61	11,516	10,261
1970	11,038	100.73	11,119	9,907

Source: Table VIII and Chapters II and III.

cyclical relative after the proper year. When the trend value is multiplied by the cyclical relative (column 2 multiplied by column 3), a gross enrollment estimate is obtained which is based on trend and cycle. Column 4, then, contains the gross enrollment estimates which are comparable in meaning to the term Net Total Regular Session which is found in the general summary section of The University of Tennessee Record.

Persons interested in future enrollment are primarily concerned with the greatest number of students enrolled in any one quarter of a year; consequently column 5 presents the estimated Fall enrollment for each of the years listed. Fall estimates were obtained by multiplying the gross enrollment estimates in column 4 by the Fall index (89.1) prepared for such a purpose in Chapter III. It should be emphasized again that the estimates for 1951 through 1955 are probably on the low side because veteran enrollment has not been considered in any of these years.

Although the estimates in column 5 are based on a combination of trend cyclical and quarterly variation, it is essential to stress again that irregular variations of a random nature will cause deviations between the estimated and the actual enrollment. Emphasis must also be given to the fact that this time series analysis has been based on past enrollment data; and no allowance has been made in the estimates for fluctuations caused by irregularities of an episodic nature such as another major war or "federal aid to education". The estimates in column 5 should then be used as expected enrollments subject to revision from year to year as episodic irregularities are noted or anticipated.

CHAPTER V

VARIABLES IN REGRESSION ANALYSIS

The change in the enrollment with the passage of time is the only characteristic of The University of Tennessee enrollment that has been considered thus far. As yet, no attempt has been made to determine the relationship between two or more characteristics. For example, what is the relationship between University enrollment and Tennessee high school graduates and per capita income in Tennessee? In other words, would it be possible to determine the enrollment in the University more accurately if the high school graduates from the previous school year and per capita income for the previous calendar year were known? If a numerical relationship is found between University enrollment on the one hand, and Tennessee high school graduates and per capita income in Tennessee on the other hand--a relationship that yields enrollment estimates very close to the observed figure--the accuracy of the enrollment estimates for particular years may be increased considerably.[1]

The derivation of such numerical relationships is known as regression analysis, and the measurement of the degree of relationship between the variables under consideration is commonly known as correlation analysis. In this thesis both of these subjects will be combined under the single heading of correlation and will be presented in conjunction with each other.

In this chapter a very general introduction will be given concerning the practical use of correlation in a problem of this sort. Also, an attempt

---

[1]Robert Ferber, Statistical Techniques in Market Research (New York: McGraw-Hill Book Company, Inc., 1949), p. 301.

will be made to explain the variables used in this regression analyses, the manner in which they were collected, and the justification for their use. In Chapter VI the technical aspects of correlation will be discussed along with the application of such technique to the problem concerned in this thesis.

Correlation Objectives

Generally there may be two objectives in a correlation problem: the measurement of the degree of relationship irrespective of the quantitative nature of the relationship, and a derivation of a numerical (or graphic) relationship between the variables in question.[2]

A numerical relationship, the objective of this correlation analysis, is necessary for purposes of forecasting. If such a relationship is found, it may be used to forecast enrollment and to substantiate the forecase made which was based on the time series analysis. In any case, the variable being estimated is denoted as the dependent variable and should theoretically be explained by the values of the independent variables which are known or predicted. Actually, in a correlation analysis it is possible to place more faith in results if there is a causal relationship between the dependent and the independent variables. By causal relationship is meant the effect on one series actually caused by the change in one or more other series. As an example, there is a causal relationship between the dependent variable, crop

---

[2] *Ibid.*, p. 302.

growth, and the two independent variables, moisture and temperature.[3] Conversely, there is no causal relationship between the percentage of the population in the several states of the United States that are members of Masonic societies and the average horsepower produced per electric power plant in the same states; still, a moderate negative correlation exists between the two variables.[4]

Examples such as these indicate that the first question as to what variables may be correlated must be answered by the investigator after a thorough study of the variables. Certainly if a cause and effect relation exists, correlation analysis can be employed. Correlation is also justified when a close connection exists between the correlated variables and when the correlated variables are both dependent upon a third underlying variable.[5] The independent variables used in the regression analysis of this enrollment problem, i. e., Tennessee high school graduates and per capita income in Tennessee, certainly are justified in view of the above statement. This will be discussed and pictured graphically at a later stage in this chapter.

Finding a relationship between series of data may be of little or no use unless a measure is made of the closeness of the relationship. The perfect relationship, as regards estimating, is obtained when the estimates of the dependent variable obtained from the relationship exactly coincide

---

[3] Mordecai Ezekiel, Methods of Correlation Analysis (New York: John Wiley and Sons, Inc., 1941), pp. 220-225.

[4] Martin A. Brumbaugh and Lester S. Kellog, Business Statistics (Chicago: Richard K. Irwin, Inc., 1946), p. 705.

[5] Ibid., p. 705.

with the corresponding observed values.[6] In such a case the multiple correlation coefficient is plus or minus one. Obviously, the farther the estimates of the dependent variable deviate from the actual value of the dependent variable, the closer to zero will be the value of the multiple correlation coefficient. The coefficient of multiple correlation is then just a pure number varying from plus one through zero to minus one. The sign indicates whether the correlation is positive or negative, and the magnitude of the corefficient indicates the degree of closeness or goodness of fit.[7] As a consequence, the higher the absolute value of the coefficient, the closer the relationship between the variables; and conversely, if the multiple correlation coefficient is zero, no relationship exists between the dependent variable and the independent variables. This means that the independent variables are useless for estimating the values of the dependent variable.

Oftentimes the purpose of a correlation analysis is to ascertain the degree of relationship irrespective of its exact numerical nature. For example, in this study there were many variables that could be associated with University gross enrollment and used in a multiple correlation problem. It was necessary then to run a simple correlation analysis between enrollment and these various variables before a decision could be made as to which variables were more closely related to enrollment and as a consequence should be used in the multiple correlation problem. Specifically, simple correlation

---

[6] Ferber, op. cit., p. 303.

[7] Frederick E. Croxton and Dudley J. Cowden, Practical Business Statistics (New York: Prentice-Hall, Inc., 1948), p. 293.

analysis were run on the following series:

Dependent variable	Independent variable
1. University enrollment	Tennessee high school enrollment
2. University enrollment	Tennessee county high school graduates
3. University enrollment	Tennessee county high school and selected Tennessee city high school graduates
4. University enrollment	Tennessee per capita income payments
5. University enrollment	Tennessee total income

Examination of the coefficients of correlation reveals that the highest degree of correlation exists in the case of numbers 3 and 4 above. The exact numerical nature of the relationship is not so important as is the degree which indicates the two variables that may best be used in a multiple correlation analysis to estimate enrollment.

Limitation on Number of Observations

In making this multiple correlation analysis, many unusual factors must be taken into consideration. One of these factors is the number of observations entering into the regression analysis. Obviously, since the analysis is based on observations which will represent what has happened in the past, it is essential that if the analysis be of use in generalizing or forecasting, an adequate number of observations must be taken. At the same time it is

important to note that there are restrictions which, in effect, make it impossible to use certain observations. As an example, it is possible to obtain the number of persons enrolled in The University of Tennessee each year from 1899 to the present year. Necessary data pertaining to high school graduates may be obtained for many years in the past. However, the third variable, per capita income in Tennessee (Department of Commerce estimates), can only be obtained from 1929 to the present, thus eliminating the possibility of correlating observations earlier than that year.[8]

Again, common knowledge of the distortion caused by the war years from 1940 to 1946 and the post-war years following, makes the use of these observations impossible. In the war years, per capita income payments in Tennessee grew quickly, but University enrollment could not increase because a great portion of the persons who would have attended college were in the armed service. This situation made it necessary, then, to use as observations, those years from 1930 to 1940—years which, though not to be considered normal, may be considered periods during which there were no physical restriction of the higher education population. The number of sets of observations (ten) is not as great as is desired in a correlation problem; but due to the reasons cited above, it is impossible to obtain more meaningful observations.

University Enrollment--Dependent Variable

At this stage it is pertinent that the variables entering into this multiple correlation analysis be discussed separately. The dependent variable,

---

[8]Robert E. Graham, Jr. and Charles F. Schwartz, "State Income Payments in 1947", Survey of Current Business, Vol. 28, No. 8 (August 1948), p. 19.

The University of Tennessee student enrollment, consists of the series of gross enrollment figures used in Chapter II. Gross enrollment is then the dependent variable which will be forecast on the basis of the independent variables, and these figures are shown in Table XI for the years 1929 through 1941.

Per Capita Income Payments in Tennessee--Independent Variable

One of the independent variables is per capita income payments in Tennessee. Per capita income payments are derived by division of total income payments in Tennessee by total population in Tennessee.[9] These figures were compiled by the Bureau of the Census of the Department of Commerce. Table XI lists these figures 1929 through 1940.

A comparison of The University of Tennessee enrollment and the per capita income payments in Tennessee, and a subsequent graphic presentation of the two series in Figure II show that there is a lead-lag relationship between the two series. In other words a turning point in the series of per capita income payments precedes the turning point in University enrollment figures in this manner: a low point is reached in per capita income in 1932 while the low point in the University enrollment is not reached until the school year 1933-34. This lag of enrollment behind income payments, indicates that economic effects do not effect University enrollment immediately, and may possibly be explained as follows: the parents' decision to sent his son or

[9] *Ibid.*, p. 19.

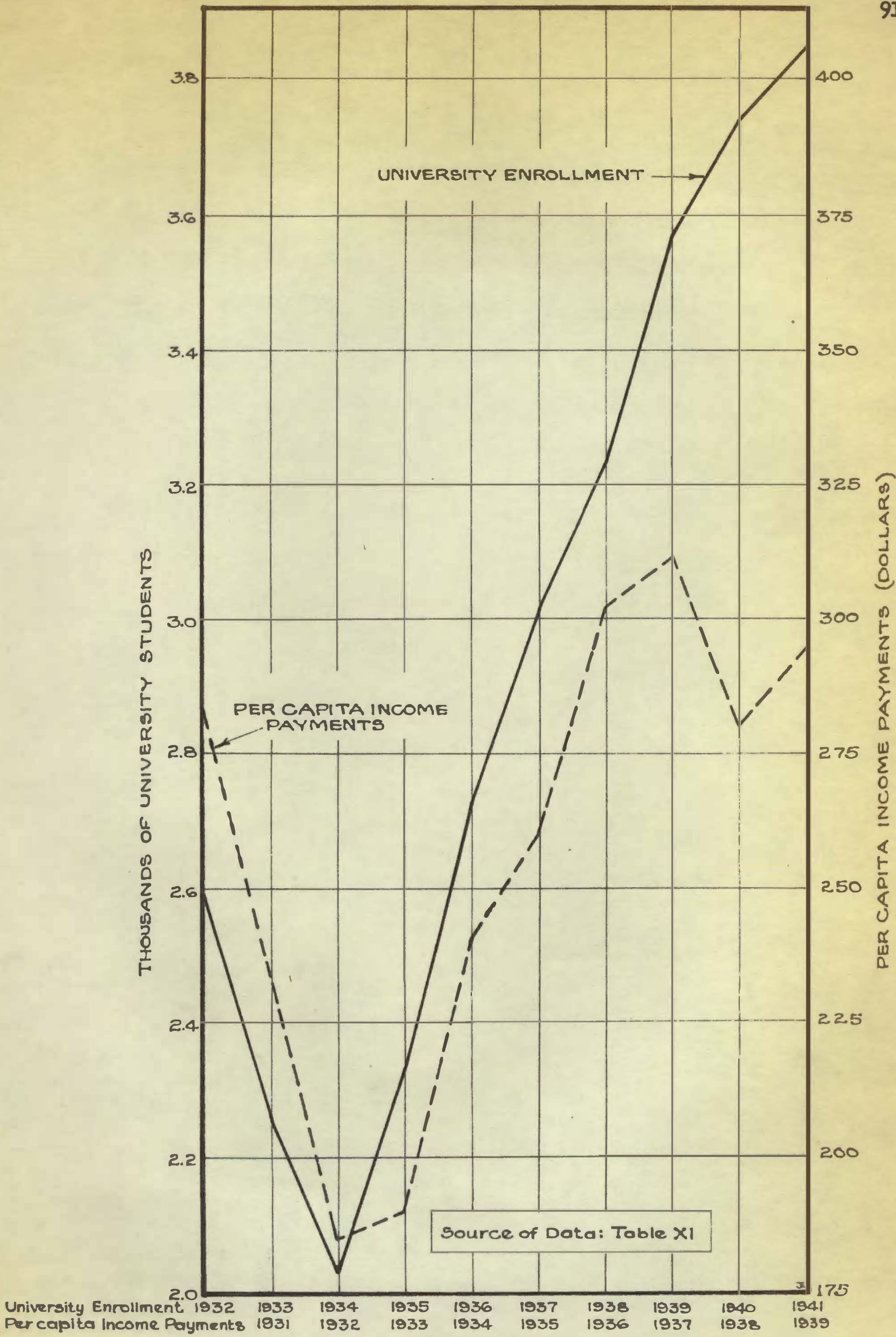

Figure 11. University gross enrollment and per capita income payments adjusted for lead-lag relationship, 1930-1941

daughter to The University is dependent more on what pecuniary gains were made in the previous year than what is made in the year of enrollment. It is a definite aid to the comprehension of economic processes to measure the period of time by which one series precedes another; so, in this instance it is helpful to know that per capita income payments lead enrollment by a half year.[10]

In using the series of per capita income payment figures, consideration has been given to the question of deflation. Deflation is the division of a value series by a price series in order to obtain a quantity series.[11] In this problem the process would be the division of the per capita income figures by a general cost of living index in order to see what the actual income figures were once the fluctuation in the value of the dollar had been removed. The decision was made to use the figures without deflating them because it is assumed that the deflated income is not as important to the parent who would send his son to college as is the dollar amount he receives. As an example, if a parent commands an income of $3600 per year, he might send his son to school notwithstanding the fact that his income if converted to 1939 figures might be $2000. Another argument against deflation is the fact that some deflator or price index for the state of Tennessee would have to be used in the deflation process, and there is no such price index available.

As a result of the above discussion the decision was made to relate the actual 1930 per capita income payment as taken from the August 1948

---

[10]Frederick E. Croxton and Dudley J. Cowden, Applied General Statistics (New York: Prentice-Hall, Inc., 1939), p. 805.

[11]Croxton and Cowden, Practical Business Statistics, p. 214.

edition of *The Survey of Current Business* to the University of Tennessee student enrollment figure for the 1931-32 school year which is labeled 1932. This combination of observations is continued, i. e., 1931 per capita income payments with 1933 University enrollment, and is set up for problem work in Table XIII.

High School Graduates in Tennessee--Independent Variable

As concerns the other independent variable used in this study, high school graduates, many points must be clarified. Since The University of Tennessee is dependent upon high school graduates and primarily Tennessee high school graduates for its enrollment, a functional relationship should exist between these two series. Actually the Tennessee high school graduates make up a large percent of the population available for the freshman class at The University of Tennessee. Since the percentage of persons passing from one class to another is fairly constant from year to year, the number of persons entering the University in the freshman year is the most important factor in determining any one year's enrollment.

To obtain figures pertaining to public school populations it was necessary to turn to secondary data. From the *Annual Report of the Department of Education* which is submitted each year to the Governor of Tennessee, it is possible to obtain many significant series relating to the Tennessee public school system. These volumes are available in The University of Tennessee library and date from the year 1869. The data in these volumes are compiled and tabulated by the Commissioner of Education's Office, and the information

used is supplied by the various county school superintendents and principals.[12]

Much experimentation was made with the various series of data available before the decision was made to use county high school graduates plus graduates from selected city high schools. For example, charts were made in an attempt to relate total high school enrollment with University enrollment. Other data were used similarly and relationships were found; however, the best relationship seemed to exist between county high school graduates and University enrollment. In order to check the correlation it was deemed advisable at this stage to follow through a graphic multiple correlation problem using the dependent variable, University enrollment and the two independent variables, per capita income payments and county high school graduates. It is not feasible to include this preliminary problem in this thesis. Suffice it to say the coefficient of multiple correlation in this case was .92 which indicates that there is a good relationship and that a multiple correlation problem worked by the mathematical method would be fruitful.

In an attempt to gain a better relationship between University enrollment and high school graduates, the decision was made to supplement the county high school graduate figures by adding the graduates of eight selected city high school systems. Since the Annual Report of the Department of Education did not provide adequate information on city high school graduates for the years 1930 through 1940, it became necessary to appeal directly to the eight city high school systems concerned. Consequently a letter of transmittal

---

[12]Tennessee Department of Education, Annual Report of the Department of Education. Volumes for the scholastic years ending 1928 through 1946.

and an accompanying schedule as per Figure 12 and Figure 13, were sent to each of the following city high school systems:

Alcoa	Harriman	Morristown
Bristol	Johnson City	Nashville
Chattanooga	Knoxville	

All of the cities listed above have at the present time a city school system and are able to furnish a record of high school graduates for the years 1930 to 1941. Data are also available for the present enrollment in these city schools making possible a forecast of future graduates. The same cannot be said of the remaining nine city school systems in the state; thus, the decision was made to use the graduates of the above listed city high schools to supplement the county high school graduates.

The eight city high school systems cooperated by returning the completed schedules. The graduate figures listed on the schedules represented white students who graduated from the high school or schools operated by the city. Table X gives a summary of the results of the schedules, and a combination of the county high school graduates and the selected high school graduates.

The enrollment figures in the last column of Table X which will be labeled "high school graduates" became the second independent variable in the multiple correlation problem. Again the importance of finding a lead-lag relationship became evident, and again it was necessary to plot the two series in question on graph paper to see if one series leads the other. The graphic picture of University enrollment plotted against the high school graduates indicates and logic would substantiate that the high school graduates for the school year 1940-31 would have their prime effect on University enrollment during the 1931-32 school year (See Figure 14). This is true

THE UNIVERSITY OF TENNESSEE
Knoxville
College of Business Administration

March 22, 1950

Dear Mr. ______:

In order to complete a study pertaining to the student enrollment at The University of Tennessee we must have an estimate of the population available for higher education. This population consists primarily, we think, of Tennessee high school graduates. In our efforts to obtain this population estimate we find that the Educational Commissioner's Annual Report to the Governor failed to list the number of white students graduating from your city high school, or schools, prior to 1940. We will appreciate your cooperation in supplying the information requested on the enclosed form.

If such information is not available, or for some reason cannot be supplied, would you please make a notation to such effect on the form and return it. A self-addressed stamped envelope is enclosed for your convenience.

Very truly yours,

s/ E. J. Boling
Bureau of Research

EFB:mfb
Enclosures - 2

Figure 12. A letter of transmittal mailed to eight Tennessee city high school systems.

White High School Graduates of the City High School
or Schools in ______________________

Note: The figures listed should represent white students only who graduated from the high school or schools which are operated by the city. This information has already been obtained from county operated high schools.

School year	White high school graduates
1948-49	______
1947-48	______
1946-47	______
1945-46	______
1944-45	______
1943-44	______
1942-43	______
1941-42	______
1940-41	______
1939-40	______
1938-39	______
1937-38	______
1936-37	______
1935-36	______
1934-35	______
1933-34	______
1932-33	______
1931-32	______
1930-31	______
1929-30	______
1928-29	______
1927-28	______
1926-27	______
1925-26	______
1924-25	______
1923-24	______
1922-23	______
1921-22	______

Figure 13. A schedule mailed to eight Tennessee city high school systems.

TABLE X

GRADUATES OF COUNTY HIGH SCHOOL SYSTEMS IN TENNESSEE COMBINED WITH GRADUATES FROM EIGHT SELECTED CITY HIGH SCHOOL SYSTEMS, 1930-31 to 1939-40

School year (White graduates only) (1)	County high school graduates in Tennessee (2)	Graduates of eight city high school systems in Tennessee (3)	High school graduates (col. 2 + col. 3) (4)
1931	7,902	1,128	9,030
1932	7,913	970	8,883
1933	7,887	1,464	9,351
1934	8,008	1,584	9,592
1935	8,379	1,616	9,995
1936	9,081	1,641	10,722
1937	8,939	1,653	10,592
1938	9,647	1,915	11,562
1939	10,191	2,017	12,208
1940	11,924	1,827	13,751

*White graduates only were included in this study.

Source: Annual Report of the Department of Education, 1931-1940 and results of poll.

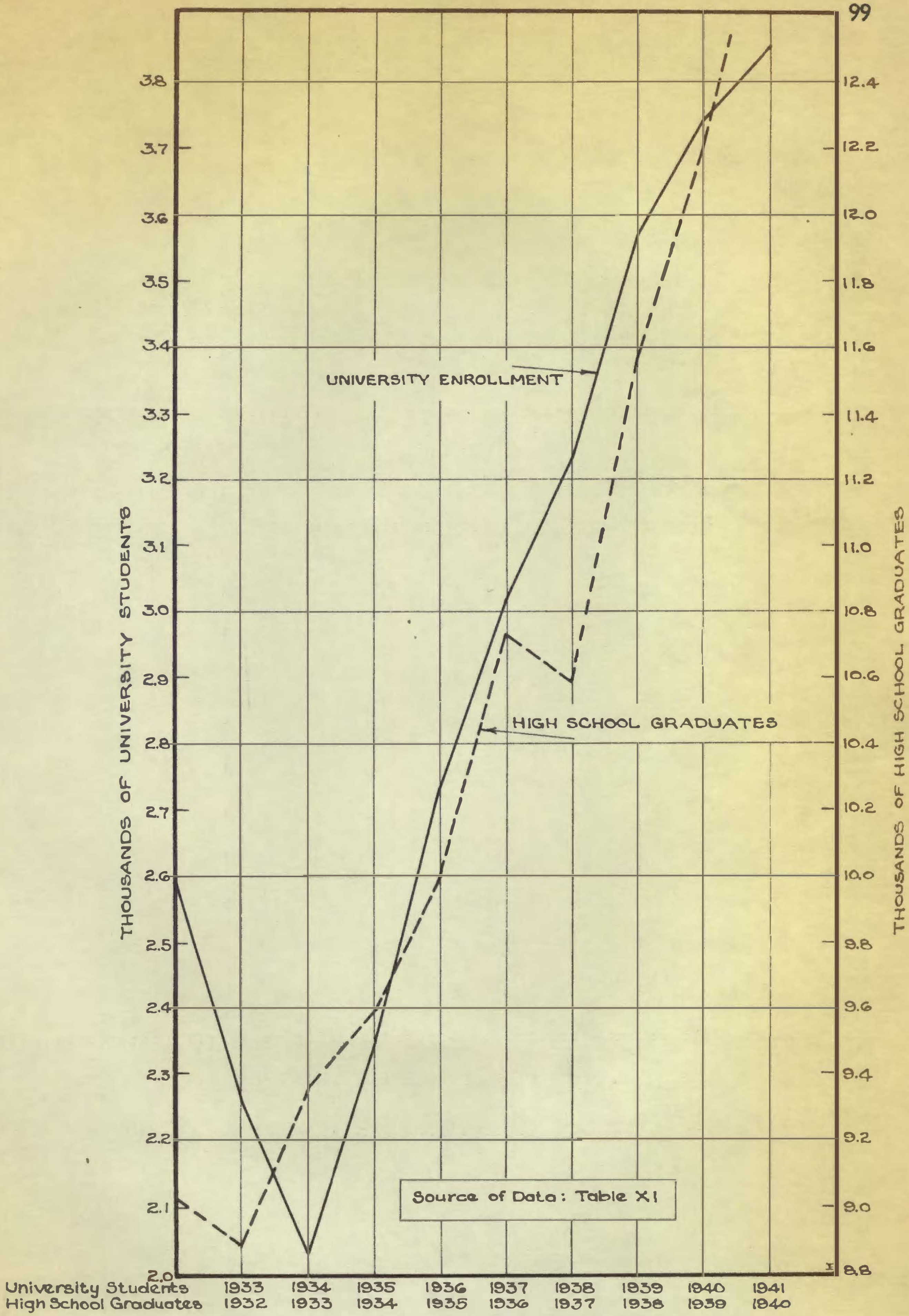

Figure 14. University gross enrollment and high school graduates adjusted for lead-lag relationship, 1931-1941

because it is usually the practice for students to graduate from high school in June and enroll in college in the following September. Table XIII is set up for the multiple correlation problem in such a way that the combination high school graduate figure for 1930-31, hereafter called 1931, is related to University enrollment figure for 1932. This lead-lag relationship continues through the ten observations in a similar manner.

Out-of-state students are not included in this study because there is no practical manner of determining the potential out-of-state enrollment. The failure to include this last element has no detrimental effects on the multiple correlation problem, however, since the series of high school graduate figures are not purported to be the universe from which University freshman enrollment could come, but rather a related series. Further discussion of the variables will be limited to a mention of the method by which estimates of future high school graduates and per capita income payments are obtained.

Estimates of Future High School Graduates

The number of persons graduating from high schools in the state of Tennessee can be expected to decline annually for the five year period from 1951 through 1955. The number of high school graduates has decreased since 1942 and will probably reach a trough in 1955 when an estimated 8,073 will graduate from county high schools and 1,538 will graduate from the eight selected city high school systems. This gives an aggregate of 9,611 graduates from city and county high schools in 1955. A decrease in high school graduates during the period to 1955 is expected largely because of the decreased number of births during the "depression" years from 1930 to 1935.

In Tennessee the ratio of the number of persons graduating from high school to those enrolled in the first grade eleven years earlier provides a means of forecasting high school graduates from enrollment statistics. It is

TABLE XI

DATA USED IN MULTIPLE CORRELATION ANALYSIS OF THE UNIVERSITY OF TENNESSEE GROSS ENROLLMENT

Year	University gross enrollment[a] (students)	High school graduates[b] (students)	Per capita income payments[c] (dollars)
1929	2,670	7,354	349
1930	2,824	7,403	283
1931	2,642	9,030	234
1932	2,591	8,883	185
1933	2,259	9,351	190
1934	2,030	9,592	241
1935	2,336	9,995	260
1936	2,731	10,722	302
1937	3,019	10,592	311
1938	3,239	11,562	280
1939	3,574	12,208	295
1940	3,741	13,751	317
1941	3,850		

[a]Table I.

[b]All county high school graduates plus graduates from eight city high school systems as per Table X.

[c]Graham and Schwartz, op. cit., p. 19.

the relatively constant ratio between the enrollment in the first grade and the number of high school graduates eleven years later that serves as the basis for the estimates presented in this section.[13] It is not deemed advisable to follow through the calculations involved in obtaining the estimates of high school graduates; however, the procedure will be explained in order that the method may be duplicated and revision may be made.

The enrollment figures used in this forecast came from the statistical section of the Annual Report of the Department of Education which is available for the necessary years. From each report it is possible to obtain the number of students graduating from both county schools and the selected eight city school systems used in this thesis. It is also possible to obtain from this same report the number of students enrolled in the first grade for both county and city schools. By securing the first grade enrollments from 1928 through 1936 and the comparable high school graduate figures from 1939 through 1945 it is possible to obtain eight ratios of high school graduates to first grade enrollment eleven years earlier. This technique was applied to both county and the eight selected city schools. The ratios when averaged for county schools show that approximately 12.6 percent of the persons in the first grade complete twelve years of scholastic work. The fact that the ratio is relatively constant is shown by the range of the eight ratios, i.e., 2.9 percent. For the eight selected city schools where the ratios were based on the city school system is Davidson, Hamilton, Knox and Sullivan counties, the average figure is 16.9 percent with a range of 3.8 percent which indicates that the average adequately represents the ratios.

---

[13] J. F. Thaden, "Forecast of Future Public School Enrollment by Grades in Michigan," Michigan Agricultural Experiment Station Quarterly Bulletin, Vol. 31, No. 4, May 1949.

By obtaining the first grade enrollments in both the county and city schools for the year 1938 through 1945, it is possible to apply the average percentage figures calculated above and obtain estimates of high school graduates through 1956. Table XII shows in column 1 the county school first grade enrollment for the years 1938 through 1945. In column 2 the 12.6 percent is multiplied by each figure in column 1 to obtain the estimated county high school graduates for the years 1949 through 1955. Column 3 shows the total of the eight selected cities first grade enrollment for 1938 through 1945, and in column 4 the 16.9 percent is applied to the figures in column 3 to obtain an estimate of city high school graduates during the years 1949 through 1955. In column 5 estimated county high school graduates are added to estimated city high school graduates to obtain the aggregate figure which will be labeled X_2', an estimate of Tennessee high school graduates to be used in the multiple correlation equation developed in the chapter to follow.

Estimates of Per Capita Income Payments in Tennessee

Estimates of per capita income payments in the United States are available, or sufficient data to make such estimates are plentiful. These figures are released each quarter of the year by the Department of Commerce. It is also possible to get national estimates of per capita income payments for future years; however, the same is not true of individual state estimates. State estimates are made annually but when released they are too late to be of use in a problem of this type. It is imperative then that some reasonably accurate estimate of future per capita income be obtained if this correlation technique is to be used. David Chambers, Associate Professor of Statistics at The University of Tennessee, has been very successful in forecasting Tennessee

TABLE XII

ESTIMATES OF HIGH SCHOOL GRADUATES AND PER CAPITA INCOME PAYMENTS IN TENNESSEE

First grade enrollment county	Estimated high school graduates county (Col. 1 x 12.6)	First grade enrollment city	Estimated high school graduates city (Col. 3 x 16.9)	Estimated high school graduates[a] county and city (Col.2 + Col.4) X_2'	Year	Estimated per capita income payments[b] (dollars) X_3'
(1)	(2)	(3)	(4)	(5)	(6)	(7)
80,537	10,147	10,159	1,717	11,864	1949	956
79,094	9,966	9,623	1,626	11,592	1950	960
73,962	9,319	9,001	1,521	10,834	1951	958
71,777	9,044	9,115	1,540	10,584	1952	952
68,602	8,644	8,815	1,490	10,134	1953	960
64,775	8,162	8,891	1,503	9,665	1954	980
63,116	7,953	9,119	1,541	9,494	1955	995

[a]Table XI.

[b]Mr. David Chambers, Associate Professor of Statistics, University of Tennessee.

income figures. He has furnished the estimates of per capita income payments in Tennessee shown in Table XII and these estimates which are labeled X_3 will be used in the multiple correlation equation.

CHAPTER VI

MULTIPLE CORRELATION ANALYSIS OF UNIVERSITY ENROLLMENT

The measurement of the relationship between the dependent variable, University enrollment, and the two independent variables, per capita income payments and high school graduates, is the subject of this chapter. The assumption is made that the elementary concepts and procedures concerning correlation have been reviewed before this multiple correlation technique is developed. Certainly, this thesis, which is concerned with a particular problem, is not the place for a discussion of simple correlation technique. The knowledge of such a technique is essential, however, because with slight modifications the methods employed in simple correlation problems are carried over to multiple correlation problems.[1]

Multiple correlation differs from simple correlation in that it extends the subject to the consideration of the relationship among three or more variables. Consideration of more than two variables is necessary in a problem of the type being discussed in this thesis where the differences in the dependent variable may be due to a number of other variables, all acting simultaneously. For example, the differences in the enrollment figures from year to year are the combined result of differences in high school enrollment and per capita income payments.[2]

---

[1]Robert Ferber, Statistical Techniques in Market Research (New York: McGraw-Hill Book Company, Inc., 1949), p. 379.

[2]Mordecai Ezekiel, Methods of Correlation Analysis (New York: John Wiley and Sons, Inc., 1941), p. 163.

Multiple Correlation Estimating Equation

The variables used in this correlation problem will be designated in the following manner:[3]

Dependent variable:
University enrollment . . . . X_1

Independent variables:
High school graduates . . . . X_2
Per capita income payments . . . . X_3

The equation as applied to this problem becomes:

$$X_1' = a + b_2 X_2 + b_3 X_3$$

which includes both independent variables and also provides a means of estimating enrollment. The estimate is labeled X_1' and is an estimate of variable X_1 computed from variables X_2 and X_3, hence the term multiple regression equation.[4] Since there are two independent variables, there are also two coefficients, b_2 and b_3. These are labeled net regression coefficients with the "net" connotation added to indicate that the coefficients show the relation of X_1 to X_2 and X_3 respectively, excluding the influences of the other independent variable. Thus b_2 is an estimate of the variation in enrollment associated with a variation in high school graduates, independent of variation in per capita income payments.[5] In other words, as between

---

[3]Ferber, op. cit., p. 346.

[4]Frederick E. Croxton and Dudley J. Cowden, Applied General Statistics (New York: Prentice-Hall, Inc., 1939), p. 740.

[5]Ezekiel, op. cit., p. 167.

years that have the same per capita income payments, but differ with respect to high school graduates, each variation of graduates between years will be accompanied by a variation of b_2 in University enrollment. The other net regression coefficient, b_3, is interpreted in the same manner, the variable high school graduates being held constant. In order to know the effect on enrollment of high school graduates along, all other variables, not just per capita income payments, should be held constant. As more and more independent variables are introduced, the b_2 value becomes better, but obviously, the multiple correlation procedure is becoming so complicated that it is not practical.[6]

As was mentioned in Chapter I, the social scientist cannot use laboratory methods to hold variables constant; but he often uses the term <u>ceteris paribus</u> ("other things being equal"). Statistically the multiple correlation technique is one of the methods used to keep "other things equal."

The term <u>a</u> is a constant in the equation and is the hypothetical enrollment when the two indpendent variables, high school graduates and per capita income payments are zero. The value of <u>a</u> when added to the sum of the net amounts associated with each independent variable produces the enrollment estimate for any given year.

---

[6]Croxton and Cowden, <u>op. cit.</u>, p. 167.

Deviation Formulae in Multiple Regression Analysis

An equation to express the average relation between University enrollment, high school graduates, and per capita income payments, $X_1 = a + b_2 X_2 + b_3 X_3$, is the linear multiple regression equation used to determine the values of the constants, <u>a</u>, <u>b_2</u> and <u>b_3</u>. The direct methematical process for obtaining the "best" set of values for these constants is called the "method of lease squares," and the normal equation for obtaining the constants by this process are produced below:[7]

$$\text{I} \quad \Sigma(x_2^2)b_2 + \Sigma(x_2x_3)b_3 = \Sigma(x_1x_2)$$

$$\text{II} \quad \Sigma(x_2x_3)b_2 + \Sigma(x_3^2)b_3 = \Sigma(x_1x_3)$$

$$\text{III} \quad a = M_1 - b_2M_2 - b_3M_3$$

In the above equations the symbol M represents the mean value of each variable while the particular variable is indicated by the subscript. The symbols $\Sigma(x_1x_2)$, $\Sigma(x_1x_3)$ and $\Sigma(x_2x_3)$ represents sums of the product of the variables, corrected to adjust them to deviations from the mean, e.g., $x_1 = X_1 - M_1$ or $\Sigma(x_1x_2) = \Sigma[(X_1 - M_1) \quad (X_2 - M_2)]$. Similarly, the symbols $\Sigma(x_2^2)$, etc., represents the sums of the squares of the variables adjusted to deviations from the mean.[8]

It will be noted that all the variables, X_1, X_2, and X_3 have been stated in the two normal equations in terms of their deviation from their

---

[7]Ezekiel, <u>op</u>. <u>cit</u>., p. 191.

[8]<u>Ibid</u>., p. 191.

mean values and have been denoted by the symbols, small x_1, x_2, and x_3. Formulae using correction factors eliminate much of the labor necessary in obtaining these deviations. The formulae listed below facilitate the application of multiple correlation to the enrollment problem:[9]

$$\Sigma(x_1 x_2) = \Sigma(X_1 X_2) - n M_1 M_2$$

$$\Sigma(x_1 x_3) = \Sigma(X_1 X_3) - n M_1 M_3$$

$$\Sigma(x_2 x_3) = \Sigma(X_2 X_3) - n M_2 M_3$$

$$\Sigma(x_2^2) = \Sigma(X_2^2) - n (M_2^2)$$

$$\Sigma(x_3^2) = \Sigma(X_3^2) - n (M_3^2)$$

Adjustment and Coding of Original Observations

Before applying the multiple correlation equations to the enrollment problem, it was necessary to check the variables and see that they were properly set up. Table XIII shows that the three variables have been listed with lead-lag relationships as discussed in Chapter V. It may also be noted that to make the computations smaller, two of the variables, University enrollment and high school graduates, have been coded. Coding is a mathematical manipulation whereby the size of the number is decreased for purposes of computation. Both variables in this case have been divided by ten and the decimals have been rounded to the nearest unit.

---

[9] *Ibid.*, p. 192.

TABLE XIII

VALUES FOR MULTIPLE REGRESSION EQUATION

	University gross enrollment X_1 (Hundreds of students) (1)	High school graduates X_2 (Hundreds of students) (2)	Per capita income payments X_3 (Dollars) (3)	X_2^2 (4)	X_2X_3 (5)	X_1X_2 (6)	X_3^2 (7)	X_1X_x (8)	X_1^2 (9)
	259	903	283	815409	255549	233877	80089	73297	67081
	226	888	234	788544	207792	200688	54756	52884	51076
	203	935	185	874225	172975	189805	34225	37555	41209
	234	959	190	919681	182210	224406	36100	44460	54756
	273	1000	241	1000000	241000	273000	58081	65793	74529
	302	1072	260	1149184	278720	323744	67600	78520	91204
	324	1059	302	1121481	319818	343116	91204	97848	104976
	357	1156	311	1336336	359516	412692	96721	111027	127449
	374	1221	280	1490841	341880	456654	78400	104720	139876
	385	1375	295	1890625	405625	529375	87025	113575	148225
Sums	2937	10568	2581	11386326	2765085	3187357	684201	779679	900381
Means	293.7	1056.8	258.1						
Correction Item				11168262	2727601	3103822	666156	758040	862597
Corrected Sum				218100	37484	83535	18045	21639	37784

Values for Deviation Formulae

In order to work out the necessary values to substitute in the normal equations, it is necessary to compute eith arithmetic values from the original data. These are $\Sigma X_1, \Sigma X_2, \Sigma X_3$, $\Sigma(X_2^2)$, $\Sigma(X_3^2)$, $\Sigma(X_1 X_3)$, and $\Sigma(X_2 X_3)$. The computation of the values for the University enrollment problem is shown in Table XIII along with an additional computation, $\Sigma(X_1^2)$, to be used later.

After the above sums have been obtained, it is necessary to compute the values M_1, M_2, and M_3 by dividing the sums of the first three columns by the number of observations (ten). The correction values for each of the columns is then calculated and placed below the value from which it is to be subtracted. As an example, the value beneath the sum of column 5, $\Sigma(X_2 X_3)$, is its correction factor, $n\, M_2\, M_3$. In numbers, this correction factor becomes (10) (1056.8) (258.1) or 2,727,601, the value entered. The other correction factors are worked in an analagous fashion and entered. Then subtracting each correction factor from the value above it produces values to be placed in the normal equations. Then these values are placed in the two normal equations they become:

$$\text{I} \quad \Sigma(x_2^2)\, b_2 + \Sigma(x_2\, x_3)\, b_3 = \Sigma x_1\, x_2$$

$$218{,}100 + 37{,}484 \qquad = 83{,}535$$

$$\text{II} \quad \Sigma(x_2\, x_3)\, b_2 + \Sigma(x_3^2)\, b_3 = \Sigma x_1\, x_3$$

$$37{,}484\, b_2 + 18{,}045\, b_3 = 21{,}639$$

Solving Simultaneous Equations by the Doolittle Method

Obviously the next step is concerned with solving the two equations simultaneously in order to arrive at values for the two net regression coefficients. The Doolittle method offers the simplest manner of solving these two equations. The first equation is divided through by the coefficient of b_2 with the sign changed. This gives the first derived equation Ia:

$$\text{I} \quad 218{,}100b_2 + 37{,}484b_3 = 83{,}535$$

$$\text{Ia} \quad -b_2 - .171866b_3 = -.383012$$

At this time equation II is entered and unter it is written equation I multiplied by the coefficient of b_3 in equation Ia (-.171866). These two equations are then summed, eliminating the values in b_2:

$$\text{II} \quad 37{,}484b_2 + 18{,}045b_3 = 21{,}639$$

$$(-\ .171866) \quad \underline{\text{I}\ -37{,}484b_2 - \ \ 6{,}442b_3 = -14{,}357}$$

$$11{,}603b_3 = \ \ 7{,}282$$

$$\text{IIa} \qquad b_3 = \ .627596$$

The value of b_3 as given in IIa is substituted in Ia to give the value for b_2:

$$-b_2 - .171866\ (.627596) = .383012$$

$$-b_2 - .107862 = .383012$$

$$b_2 = .275150$$

With the values for b_2 and b_3 it is possible to find the value for a by substituting in equation III:

III $a = M_1 - b_2 M_2 - b_3 M_3$

$= 293.7 - .27515\ (1056.8) - .627596\ (258.1)$

$= 293.7 - 290.77 - 161.98$

$a = -159.05$

Assuming that multiple linear correlation is the method to be used, constants a, b₂, and b₃, the "best" values possible are used in the regression equation. With this equation it is possible to estimate X_1 from X_2 and X_3 by substituting the values for the two independent variables. The regression equation, with consideration being given to the coding manupulation, becomes:[10]

$$\frac{X_1}{10} = -159.05 + .27515\ \frac{X_2}{10} + .627596X_3$$

$$\text{or} \quad X_1 = -1590.5 + .27515X_2 + 6.27596X_3$$

Correlation Devices Necessary in Forecasting

The methods applied to the enrollment problem to this point have been used to estimate the enrollment figures when the values for high school graduates and per capita income payments were known. They also reveal the average change in value of the dependent variable for each individual change in the value of the given independent variables. If, however, the regression equation is to be used in forecasting, it is essential to know:

---

[10] *Ibid.*, pp. 190-195.

(1) The accuracy of the equation as a basis for estimating enrollment from high school graduates and per capita income, or how close enrollment values can be estimated from the values of high school graduates and per capita income.

(2) The relative importance of the relationship, or the importance of the relationship between the dependent variable, University enrollment, and the two independent variables, high school graduates and per capita income payments.[11]

Knowledge of the above is given by three correlation devices: the standard error of estimate, the coefficient of multiple determination.

The Standard Error of Estimate

The standard error of estimate is the measuring device determining the accuracy of the estimating equation. This measure of accuracy may be obtained by computing estimated enrollment for each of the ten observations, using the regression equation, then comparing such estimated values with the actual values. In Table XIV the necessary computations for obtaining the standard error of estimate are shown. These computations are explained mathematically below.

An estimated enrollment figure was worked out in Table XIV by substituting in the multiple regression equation the values of X_2 and X_3 for each successive observation. If the symbol X_1' is used to represent this estimated

---

[11] *Ibid.*, p. 128.

TABLE XIV

ACTUAL ENROLLMENT AND ENROLLMENT ESTIMATED FROM HIGH SCHOOL GRADUATES AND PER CAPITA INCOME PAYMENTS ON BASIS OF MATHEMATICALLY DETERMINED RELATIVES

High school graduates X_2 (1)	Per capita income payments X_3 (2)	Estimates for high school graduates b_2X_2 (3)	Estimates for per capita income payments b_3X_3 (4)	Constant a (5)	Estimated enrollment X_1' (6)	Actual enrollment X_1 (7)	Actual minus estimated enrollment $X_1 - X_1'$ (8)
9,030	283	2,485	1,776	- 1,590	2,671	2,590 1932	-81
8,880	234	2,443	1,469	- 1,590	2,322	2,260	-62
9,350	185	2,573	1,161	- 1,590	2,144	2,030	-114
9,590	190	2,639	1,192	- 1,590	2,241	2,340	99
10,000	24	2,752	1,513	- 1,590	2,675	2,730	55
10,720	260	2,950	1,632	- 1,590	2,992	3,020	28
10,590	302	2,914	1,895	- 1,590	3,219	3,240	21
11,560	311	3,181	1,952	- 1,590	3,543	3,570	27
12,210	280	3,360	1,757	- 1,590	3,527	3,740	213
13,750	295	3,783	1,851	- 1,590	4,044	3,850 1941	-194

value of X_1 it may be written:

$$X_1' = a + b_2X_2 + b_3X_3$$

In column 8 each estimated enrollment has been subtracted from the corresponding actual enrollment. The residual is represented by $X_1 - X_1'$ and shows by how many students the estimate missed the actual enrollment. These residuals could be squared and summed, divided by the number of observations, and when the square root of this figure was obtained, it would be called the standard error of estimate. Actually, it is the standard deviation around the regression line, a changing quantity, and may be represented by the formula, $S = \sqrt{\frac{(X_1 - X_1')^2}{N}}$ where N = the number of observations. This device measures the scatter of the observations around the regression plane and may be interpreted exactly by the standard deviation concept, i.e., the regression plane plus and minus one standard deviation gives a range which will include the middle two-thirds of the observations.

The above calculations have not been carried through because for linear multiple regression equations a much simpler process can be used. An advantage of this simpler method is that it tends to adjust the standard deviation of residuals in such a way as to give an unbiased estimate of the standard error of estimate. Such an adjustment is needed since where the size of the sample is small in proportion to the number of variables involved, the standard deviation of the residual s for the observations included in the sample tend to have a downward bias. In other words such a standard error would tend to be smaller than it would have been if more observations had been made.[12]

---

[13] *Ibid.*, p. 208.

The formula actually used to obtain the standard error of estimate of the enrollment problem makes use of the b values plus $\Sigma(x_1^2)$ which was computed in Table XIII. The formula[13] is given below:

$$\bar{S}^2 = \frac{\Sigma(x_1^2) - [b_2(\Sigma x_1 x_2) + b_3(\Sigma x_1 x_3)]}{n-m}$$

where: n = number of sets of observations in the sample
m = number of constants in the regression equation, including a, b_2, and b_3.

When the values from Table XIII are substituted into the above formula the equation becomes in terms of coded values for X_1:

$$\frac{\bar{S}^2}{10} = 37,784 - \frac{[.27515\ (83,535) + .627596\ (21,639)]}{10-3}$$

$$\frac{\bar{S}}{10} = \sqrt{1220}$$

$$\bar{S} = 132$$

The standard error of estimate, since it is expressed in the same units as the dependent variables, becomes 132 students. In agreement with the standard deviation concept, two-thirds of the residuals calculated in column 8 in Table XIV, should be expected to come between + 132 and -132. Of the ten observations, eight came within the range of the line which represents 80 percent of all the observations. Similarly, only 4.55 percent of the cases would be expected to fall outside the range $\pm 2\bar{S}$, i. e. below -264 or above + 264. Actually none come outside this range.

If in years exclusive of observation years, essentially the same set of conditions exist as those under which the enrollment data were selected,

---

[13] Ibid., pp. 209-212.

and only the independent variables, high school graduates and per capita income payments, are known, it is possible to estimate the probable enrollment from the known values of the independent variables. If the estimates are to be made for years to come using new observations, a knowledge of the standard error of estimate provides a basis for judging how closely the new estimates are likely to approximate the true, but unknown enrollment. Using a purely hypothetical case, an enrollment estimate could be made as follows:

$$X_1' = a + b_2X_2' + b_3X_3'$$
$$= -1590.5 + .27515\ (20,000) + 6.27596\ (1,000)$$
$$= 10,188$$

where:

X_1' = estimated enrollment 19?

X_2' = estimated high school enrollment 19?

X_3' = estimated per capita income payments 19?

The most probable estimate of enrollment would be 10,188 students. However, at this stage the standard error concept would be used in this manner: $X_1' \pm 2\bar{S}$ will include approximately 95 percent of the observations, or about 95 times out of 100 a range will be established which will contain the true value, in this case the enrollment. Using the values calculated above, $10,188 \pm 2(132) = 10,188 \pm 264$, which means the range within which the actual enrollment would probably fall extends from 9,924 to 10,452.

The Coefficient of Multiple Correlation

The relative importance of the relationship between enrollment and the two independent variables, high school graduates and per capita income payments, is measured by the proportion of the variation in the enrollment data which can be accounted for by variation in the two independent variables. The square root of this proportion is called the coefficient of multiple correlation and may be calculated in the following manner:[14]

(1) Calculate the standard deviation of the estimated enrollment for the ten observations.

(2) Calculate the standard deviation of the ten original enrollment observations.

(3) Divide the result of Step 1 by the result of Step 2, and the ratio, the coefficient of multiple correlation, measures the combined importance of the two indpendent variables as a means of explaining differences in the dependent variable.

It is important to understand the method given above, but practical to use a short formula involving only values already determined. Such a formula is presented below, and the proper values are substitued to produce the coefficient of multiple correlation (R):

$$R = \frac{b_2\,\Sigma(x_1x_2) + b_3\,(\Sigma(x_1x_3)}{\Sigma x_1^2}$$

$$R = \frac{.27515\ (83{,}535) + .627596\ (21{,}6391)}{37{,}7894}$$

$$R = .9677 = .983$$

---

[14] *Ibid.*, p. 210.

Again an adjustment for the small sample size must be made, since there is a tendency for the multiple correlation shown by the sample to be in excess of the correlation existing in the universe from which the sample was drawn. Consequently, the coefficient R computed above has to be adjusted to give $\bar{R}$, the unbiased estimate of the correlation most probably existing in the whole universe.[15] The formula is:

$$\bar{R} = 1-(1-R^2)\left(\frac{n-1}{n-m}\right)$$

where: n and m have the same meaning as above.

$$\bar{R} = 1-(1-.9677)\left(\frac{10-1}{10-3}\right)$$

$$\bar{R} = .9585 = .979$$

The adjusted coefficient of multiple correlation, $\bar{R}$, is $+$ .979 in this problem. Since this coefficient is an abstract measure of correlation, it is only meaningful in its absolute value as it fluctuates between plus one and minus one. The sign of $\bar{R}$ in this case is positive, since the sign of the coefficients of net regression were positive and because an increase in the dependent series is associated with an increase in the independent series. Actually, the coefficient of multiple correlation, $\bar{R}$, could never exceed one; and because the square root of a fraction is always larger (in absolute terms) than the fraction itself, such a measure tends to exaggerate the degree of correlation unless this fact is recognized.

---

[15] Ferber, op. cit., pp. 356-358.

The Coefficient of Multiple Determination

Inasmuch as $\bar{R}$ is an abstract measure of correlation, it is clearer to interpret the results of a correlation analysis in terms of the proportion of the total variance that has been explained by the regression, i.e., in terms of the coefficient of determination.[16] Where the dependent variable X_1 and the two independent variables X_2 and X_3 are assumed to be made up of elements of variability, all of which are present in X_1, but some of which are lacking in X_2 and X_c, it may usually be stated that $\bar{R}^2$ measures that portion of all the elements in X_1 which are also present in X_2 and X_3. Because of this, where the dependent variable is causally related to the independent variables, $\bar{R}^2$ is called the coefficient of multiple determination. This measure is and must always be positive since it measures the percentage of the variance in the values of the dependent variable that can be accounted for or explained by the variation in the values of the independent variable.[17] It may be noted that this measure is calculated before the coefficient of multiple correlation is determined; hence, there is no extra work in obtaining the measure.

The value of $\bar{R}^2$, the coefficient of multiple determination, in the problem under analysis is .9585 or 95.85 percent after correction is made for the size of the sample. The practical meaning is that approximately 96 percent of the variation in University enrollment may be attributed to the variation

---

[16] *Ibid.*, p. 314.

[17] Ezekiel, *op. cit.*, p. 139.

in high school enrollment and per capita income payments.

Summary of Correlation Technique

It is possible at this stage to set up a table of measures calculated and to summarize the correlation analysis to this point:

Dependent variable	Independent variables	$\bar{S}$ Standard error of estimate	$\bar{R}$ Coefficient of multiple correlation	$\bar{R}^2$ Coefficient of multiple determination
University enrollment (X_1)	High school graduates (X_2) Per capita income payments (X_3)	132 (students)	.979	95.85%

A linear multiple regression equation has yielded estimates of University enrollment from values of high school graduates and per capita income payments. The accuracy of these estimates is measured by the standard error of estimate of 132 students. The coefficient of multiple correlation ($\bar{R}$ = .98) and the coefficient of multiple determination ($\bar{R}^2$ = .96) show that about 96 percent of the variance in enrollment for those years sampled could be attributed to the differences in high school graduates and per capita income payments. Since this leaves only about four percent of the variance to be accounted for by all other factors, it would appear that the two variables, high school graduates and per capita income payments were the most important factors associated with the enrollment for the years sampled.[18] Figure 15

[18] *Ibid.*, p. 181.

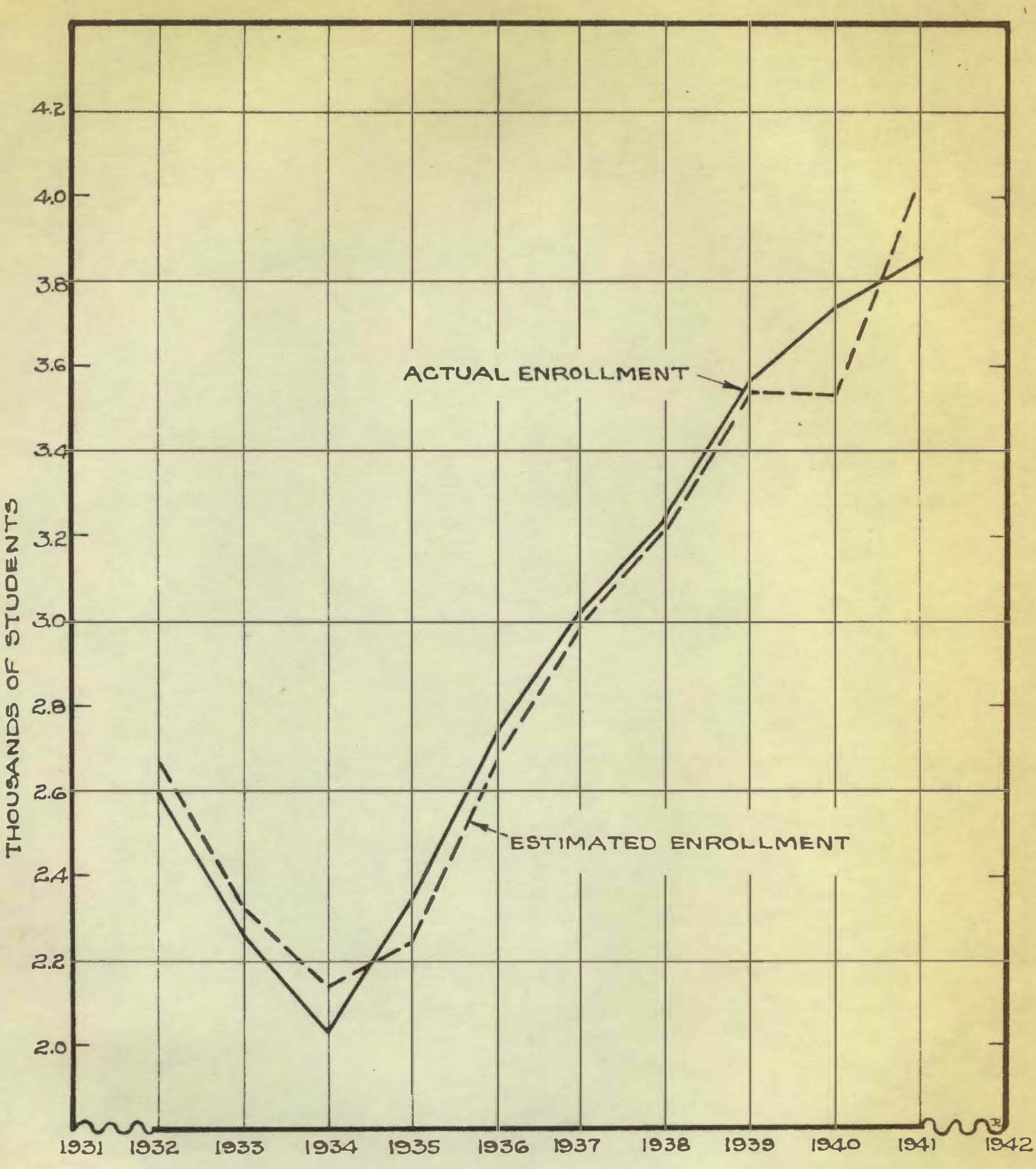

Figure 15. Estimated and actual enrollment of The University of Tennessee, 1932 – 1941

Source of Data: Table XIV

shows the actual enrollment and estimated enrollment from Table XIV plotted against time.

Significance of Regression

From the above summary it is evident that the relationship among the variables concerned is excellent for the years sampled; however, the primary purpose of this analysis was to find a relationship that could be used in forecasting. In other words, the problem now becomes one of estimating true values of correlation and regression parameters from sample data. This problem entails finding how far the regression plane and the correlation measures are likely to depart from the true values for the universe from which the sample was drawn.

Practically, the above paragraph suggests this problem: Will it be possible to forecast enrollment for years to come if the relationship among University enrollment, high school graduates, and per capita income (for the ten year period) are known? If the conditions of sampling were ideal, the above question could be answered by the use of error formulas without hesitation. However, it is necessary to recognize that the ten sets of observations used in this problem formed a time series where one of the essential elements of sampling, randomness, is not present as such. Forecasts worked out by extrapolating an earlier formula to subsequent years have given results which agreed remarkably well with the standard error of estimate; and as a consequence, many staticians agree that sampling equations have a wider applicability than their basic assumptions seem to warrant.[19]

[19] Ibid., p. 352.

It is not feasible in this thesis to consider the arguments for and against the use of error formulas in the correlation of time series; however, it is deemed necessary to justify the use of such formulas by using a quotation from Ezekiel:

> Where there is clear indication of lagging effects from period to period which cannot be specifically allowed for, or where the serial correlations in the data are so high as to make the several observations not really independent observations at all, then the sampling formulas simply do not apply, because the assumptions on which they are based are not fulfilled in the given problem. In such cases the error formulas may still be calculated, in the hope that they will indicate the minimum possible reliability of the results instead of the maximum possible unreliability.[20]

In the preceding chapter, explanation was given for the allowance made for lagging effects from period to period as related to the problem under consideration. Also, independence of observations could be justified on many counts.

Analysis of Variance

Irrespective of the status of the variables, the decision was made to subject the relationship to a regression test. Variance analysis has proven to be an extremely effective method for analyzing the significance of correlation results; consequently, the decision was made to subject the relationship to the F test. Ferber explains variance analysis as follows:

> This method is based on the fact that a coefficient of (simple or multiple) determination is essentially the ratio of the sum of squares accounted for by the correlation to the

---

[20] *Ibid.*, pp. 355-357.

> total sum of squares. The difference between these two sums of squares is the sum of squares remaining after correlation, which presumably measures the random sampling variations in the variable under study. Hence, the significance of a correlation coefficient may be gauged by the extent to which the sum of squares explained by correlation exceeds the unexplained (sampling) sum of squares, both terms being divided by their appropriate degrees of freedom. The more significant is a correlation coefficient, the more will this ratio, exceed 1. . . . the probability of a particular F ratio arising as a result of chance is determined with reference to Appendix Table 12.[21]

To test the significance of the multiple correlation coefficient in the problem under consideration, it is necessary to obtain the sum of the squares of the enrollment variable which was computed to be 37,784 in Table XIII. This figure, when multiplied by the proportion of total variance explained by the multiple regression, $\bar{R}^2$ (coefficient of multiple determination), gives the explained sum of squares:

$$(37{,}784)\ (.985) = 36{,}197$$

The unexplained sum of squares may be ascertained simply by finding the difference between the total sum of squares and the explained sum of squares.

It is now possible to construct an "analysis of variance" table as shown below:

TABLE XV

SIGNIFICANCE OF CORRELATION BY ANALYSIS OF VARIANCE

(1) Type of variance	(2) Sum of squares	(3) Degrees of freedom	(4) Estimate of σ^2
Explained by correlation	36,197	2	18,088
Unexplained	413	7	59
Total	37,784	9	

[21] Ferber, *op. cit.*, p. 396.

The number of degrees of freedom (in column 3 of the above table) associated with the unexplained sum of squares is the number of observations less the parameters in the regression equation, seven in this case. The number of degrees of freedom associated with the total sum of squares is the total number of observations less one, the grand mean, or nine degrees of freedom. If the seven degrees of freedom for the unexplained variation are subtracted from the nine degrees of freedom for the total variation, the remainder for the unexplained variance must be two degrees of freedom. At this point the sums of squares in column 2 are divided by the degrees of freedom in column 3, and column 4 gives the estimates of σ^2.

The hypothesis in an F test is that there is no significant difference between the two estimates of σ^2. This would mean that if the two estimates of σ^2 are not significantly different then there is no real linear multiple relationship between X_1, X_2, and X_3. The remaining task then is to find the value for F, i.e., to divide the explained estimate of σ^2 by the unexplained estimate of σ^2:

$$F = \frac{18,088}{59} = 306$$

It is highly improbable that the two estimates of σ^2 will be the same; so the actual test is whether the difference between the two estimates, as found by the ratio above and considering the degrees of freedom, would have happened frequently by chance or whether such a difference would have happened very infrequently by the chance factor alone. Using an F table with n_1 (two) and n_2 (seven) degrees of freedom, the value at the one percent level is 9.55. Since the F value computed above, 306, exceeds 9.55, the conclusion is that a difference so large as exists would have happened due to the chance factor

alone much less than one time in one hundred, or the probability of such an occurrence by the chance factor is very remote.[22] The hypothesis that the two estimates of σ^2 are the same or that there is no real regression is rejected, and the conclusion is made that the multiple linear correlation found in this problem is real and is significant.

Estimating Future University Enrollment

The multiple regression equation and standard error of estimate will be used then in an attempt to forecast the enrollment for The University of Tennessee.

It is important to stress again that the number of observations used in this analysis (ten) is not as large as is desired; and as a consequence, the use of the estimating equation in forecasting future enrollment is limited to a short period of time. The multiple correlation technique provides estimates of University gross enrollment for the five year period from 1951 through 1955. If this method is to be used in the future, the suggestion is made that more observations should be included as they become available.

In Table XVI, estimates of Fall quarter enrollment are presented for the five year period from 1951 through 1955. These figures are obtained by solving the equation for gross enrollment estimates and multiplying each result by the Fall quarterly index (89.1 percent). In solving the equation,

---

[22] H. A. Freeman, Industrial Statistics (New York: John Wiley & Sons, Inc., 1942), pp. 96-99.

TABLE XVI

ESTIMATES OF THE UNIVERSITY OF TENNESSEE FALL ENROLLMENT BASED ON MULTIPLE CORRELATION EQUATION, 1951-1955

Year (1)	Estimated high school graduates X'_2 (2)	Estimated per capita income payments X'_3 (3)	$b_2 = .27515$ $b_2 X'_2$ (4)	$b_3 = 6.27596$ $b_3 X'_3$ (5)	$a = -1590$ a (6)	Estimated university gross enrollment X'_1 (7)	Estimated university fall enrollment ($X'_1 \times 89.1$) (8)
1951	11,592	956	3,190	6,000	-1,590	7,600	6,772
1952	10,834	960	3,021	6,025	-1,590	7,456	6,643
1953	10,584	958	2,912	6,012	-1,590	7,334	6,535
1954	10,134	952	2,788	5,975	-1,590	7,173	6,391
1955	9,665	960	2,659	6,025	-1,590	7,094	6,321

Source: Tables XII and X.

it may be noted that X_2' (estimated high school graduates) and X_3' (estimated per capita income payments) taken from Table XII are used. Even though an estimate (X_1') is being made which is based on two other estimates (X_2' and X_3'), this does not impair the significance of the results because the methods used in estimating the two independent variables have produced reasonably accurate results in the past.

Assuming that the basic economic and social conditions which existed from 1930 through 1940 are still present, it is possible to use the standard error concept in forecasting. In other words, if the standard error of estimate is used with the estimates (X_1'), a range ($X_1' \pm \bar{S}$) will be established which should contain the actual enrollment with a set degree of probability.

Applying the standard error to the estimates ($X_1' \pm 2\bar{S}$) shown below, there is a 95.45 percent probability that the actual enrollment figures will lie between the two limits:

Year	Estimated enrollment X_1'	Lower limit $X_1' - \bar{S}$	Upper limit $X_1' + \bar{S}$
1951	6,772	6,508	7,036
1952	6,643	6,379	6,907
1953	6,535	6,271	6,799
1954	6,391	6,127	6,655
1955	6,321	6,057	6,585

It should be emphasized again that the above estimates are subject to change if episodic conditions such as a long war or "federal aid to education" become realities.

CHAPTER VII

SUMMARY AND CONCLUSIONS

A statistical analysis of The University of Tennessee enrollment and a forecast of future enrollment having been completed, this chapter will be devoted to a summary of the results of the study and conclusions concerning factors affecting enrollment which have not been measured.

Time Series Analysis and Synthesis--Summary

The first major phase of the study was a time series analysis, in which the variations in University enrollment were separated into four factors--trend, season, cycle, irregular--and inspected singly. The first factor to be measured was the trend, which appeared to be represented adequately by a Gompertz growth curve. This curve describes how University enrollment has grown from 1899 to 1949, while the use of the curve in forecasting implies the assumption that basically the same pattern of behavior will continue.

In order to project this curve into the future, it is necessary to solve the equation, logarithm of enrollment estimate = 4.914308 - 2.571869 $(.98487)^x$ where x represents the deviation in years from the origin, 1899. In other words, the value of x for years progresses as follows: 1899 = 0, 1900 = 1, 1901 = 2 . . . . 1949 = 50 . . . . 1970 = 71. Once x is replaced by the appropriate value and the equation is solved, the remaining task is to take the anti-logarithm of the result in order to obtain

the projected trend value for the year represented by x. Table XVIII shows the trend extrapolations for 1955, 1960, 1965, 1970.

When the quarterly enrollment figures were plotted against the quarters, it was obvious that the Fall quarter consistently contained the largest enrollment with the other quarters ranking behind in the order of their occurrence, i. e., Winter, Spring, Summer. Not only was it obvious that such a pattern existed, but it was also possible to measure this apparent pattern by known methods. One method, the twelve month moving average, yielded a quarterly index number for each quarter which represented the position of that quarter relative to a normal quarter (where a normal quarter equals 100 percent). The quarterly indexes calculated by this method are:

Summer - 71.7 percent

Fall - 117.0 percent

Winter - 108.4 percent

Spring - 102.9 percent

When quarterly data are available, the above percentages may be used in forecasting University enrollment from one quarter to another as is illustrated on page 48.

Because gross enrollment figures are used in trend, cyclical, irregular, and correlation analyses and forecasts were based on them, a different type of index was calculated which applies to these gross enrollment figures. Obviously, if the gross enrollment is forecast, the number to be enrolled in a particular quarter may only be estimated from this by

TABLE XVII

ESTIMATES OF THE UNIVERSITY OF TENNESSEE FALL ENROLLMENT BY SYNTHESIS OF TREND, SEASON AND CYCLE, 1955, 1960, 1965, 1970

Year (1)	Trend estimates (2)	Adjustment to quarterly data (Trend X Index) (Col. 2 x 89.1) (3)	Cyclical relative applicable (percent) (4)	Fall estimates (Col. 3 x Col. 4) (5)
1955	6,593	5,874	99.27	5,831
1960	7,933	7,068	126.03	8,908
1965	9,417	8,391	93.97	7,885
1970	11,038	9,835	100.73	9,907

Source: Chapters II, III, and IV.

applying an index if, in the past, the ratio of enrollment in that particular quarter to gross enrollment has been fairly constant. In Chapter III, such a ratio was evolved for each quarter, and the results are shown below as percentage figures: Summer 54.6, Fall 89.1, Winter 82.6, Spring 78.4. In Table XVII the trend projections through 1970 are multiplied by the Fall percentage, 89.1, to obtain gross enrollment estimates adjusted to Fall quarterly enrollment estimates. Any other quarter may be found by multiplying the trend estimate by the appropriate index. It should be emphasized that these percentages may, if desired, be applied after trend has been adjusted for cycle.

The cyclical factor was the next to be analyzed. First, by dividing the trend estimates into the actual enrollment figures which contained trend, cyclical, and irregular forces, it was possible to remove trend. The remaining percentage figures contained cyclical and irregular forces but a binomially-weighted moving average eliminated the irregular forces and left a percentage figure for each year which represented the cyclical effect on the enrollment data. These percentage figures fluctuated around 100 percent. After intensive study and a periodogram analysis, a thirty-year cycle containing three ten-year cycles was discovered in the cyclical relatives. A sine-cosine curve was fitted to the ten-year cycle, and a cosine curve was fitted to the thirty-year cycle. When these two fitted periodic curves were synthesized, they were found to follow the cyclical pattern very closely. A percentage figure for each year of the thirty-year synthesized cycle is available in Chapter IV.

In order to take into account the cyclical effect in forecasting, it is necessary to multiply the gross enrollment estimate or Fall quarter enrollment estimate by the appropriate cyclical relative taken from Chapter IV. In applying these relatives to future years, it is necessary to note that the thirty relatives repeat. In other words, the relative 130.42 for 1899 applies also to 1929, 1959, 1989, etc. The appropriate relatives are applied to the Fall enrollment estimates in Table XVII for the years 1955, 1960, 1965, and 1970. The figures in column 5 represent the completed estimates of University enrollment based on trend, season, and cycle.

To analyze irregular variations, an attempt was made to segregate the variations into either random or episodic irregularities by control chart technique. This technique involved obtaining measurements of irregular movements by subtracting the cyclical relatives from the cyclical-irregular relatives. The procedure then was to test, on the basis of probability, to see if the irregular measurements could deviate as much as they did from their mean value because of chance alone. When the chart was constructed and the individual measurements of irregular variation were plotted, it was found that five measurements fell outside the control limits. The limits were so set that a point would not fall outside by chance alone more than five times in one hundred; therefore, it was assumed that any point that fell outside the limits did so because of an assignable reason. These five points were then deemed to be the results of episodic conditions. The measurements falling outside the limits pertained to: 1909, 1919, 1944, 1945, and 1947. Each of these have been explained by episodic conditions. In 1909 the enrollment was increased by special short courses in Agriculture,

while the remaining years showed enrollments badly displaced because of the episodic events—World War I and World War II.

Time Series Technique in Long-Range Forecasting

Upon summarizing the time series analysis and synthesis, it became obvious that this method could be used successfully only as a means of making a long-range forecast. In order to use the annual estimates of enrollment based on a synthesis of trend, season, and cycle, it becomes necessary to adjust the estimates until the time when the episodic condition, World War II, ceases to be effective, since no attempt was made to account for the increased enrollment brought about by this condition. This means that until the effects of World War II on University enrollment have disappeared, which presumably will be around 1955, figures based on trend, season, and cycle will be too low. However, from 1955 to 1970 the estimates based on a synthesis of these three factors are the most probable enrollment figures, since irregular variations may not be anticipated for those years. The estimates from 1955 to 1970 are shown as Table XVII.

Multiple Correlation Technique—Summary

The last major phase of this thesis presented a method of making a short-run forecast. This method involved the use of simple and multiple correlation techniques in analyzing gross enrollment data. The simple correlation technique was used in finding a relationship between the dependent

variable, University enrollment, and each of two other variables, high school graduates and per capita income payments. These two variables were then used as independent variables in a multiple correlation analysis which showed how mych of the change in enrollment was associated with changes in the number of high school graduates and in per capita income. The result of this analysis, covering observations from 1930 to 1940, showed that approximately 96 percent of the change in enrollment was associated with a change in the two independent variables.

A further correlation formula provided a measure, the standard error of the mean ($\bar{S}$), that served to provide limits within which the estimates should fall. The standard error in this problem, when corrected for the small number of observations used, was 132 students. Practically, the meaning of the standard error is this: if by using the estimating equation, an enrollment is obtained, and two standard errors are added to the estimate to obtain an upper limit and subtract from the estimate to obtain a lower limit, approximately ninety-five times out of one hundred the range will contain the actual enrollment figure.

If the regression equation produced such accurate estimates for past years, it could be used in estimating future enrollments. Before using it as an estimating devise, it was subjected to a regression test (F test) to see if the relationship was real or if there was a high probability that it could have happened by chance. The F test showed that the relationship was real; and since it was of a degree sufficiently high, the decision was made to use the equation in making a short-range forecast.

Multiple Correlation Technique in Short-Range Forecasting

Since the independent variables were estimates and since the size of the sample was smaller than desired, it was not considered feasible to forecast University enrollment figures past 1955. For each of the five years, 1951 through 1955, an estimate was derived from the estimating equation. These estimates are shown below with an upper and lower estimate on either side of the expected figures. The chances are ninety-five out of one hundred that the actual enrollments will fall within these limits, assuming that underlying economic and social conditions do not change radically.

	Lower estimate $X_1' - 2S$	Estimate X_1'	Upper estimate $X_1' + 2S$
1951	6,508	6,772	7,036
1952	6,379	6,643	6,907
1953	6,271	6,535	6,799
1954	6,127	6,391	6,655
1955	6,057	6,321	6,585

It may be noticed that these estimates have been adjusted so that each estimate pertains not to gross enrollment of that year but to Fall quarter enrollment. This was done by multiplying each X_1' by 89.1 percent. Since these figures partially account for the effects of World War II, they are in excess of the estimates based on the synthesis of trend, season and cycle where no such effects were considered.

In order to use the multiple regression equation to forecast University enrollment for a given year the following equation is solved:

$X_1' = -1590 + .27515\ X_2' + 6.27596\ X_3'$ where X_2' is an estimate of high school graduates for the school year preceding the given year. X_3' represents an estimate of per capita income payments in Tennessee for the calendar year preceding the given school year. When solved this equation produces an estimate of gross enrollment. To convert the gross figure to the enrollment estimate for any quarter of the given year, it is necessary to multiply the estimate by the appropriate index given on page 50.

Unmeasureable Factors in University Enrollment

The use of statistical analysis in analyzing the enrollment data is comparable to the use of any other technical aid that may be employed. It makes no attempt to explain the fluctuations but rather furnishes a means of measuring the factors involved and of revealing the manner in which they are related. Statistical methods provide excellent tools for investigating the enrollment data but such methods are not substitutes for full knowledge of a problem which comes from research, clear-cut thinking, and logical analysis.[1] In fact, statistical analysis if employed carelessly may well give results worse than no analysis at all since some persons are prone to accept the results of mathematical methods as fact.[2]

In view of the above statements it is essential that recognition also be given to the vast number of variables which cannot be measured in forecasting University enrollment. Many of these variables are recognized as episodic or potential episodic irregular conditions while many are not

---

[1]Mordecai Ezekiel, Methods of Correlation Analysis (New York: John Wiley and Sons, Inc., 1941), p. 453.

[2]*Loc. cit.*

known and consequently can not be considered. In the pages that follow some of the known variables which may not be measured objectively will be discussed.

Effects of a Major War

The effects of a major war on college enrollment may well be depicted by the conditions existing at the University during and following World War II. At the beginning of the war in 1941 the enrollment began to drop from an all-time peak of 3,850 to a trough in 1944 when the gross enrollment was 1,894. A part of this drop may be attributed to the cyclical fluctuation measured and discussed in Chapter IV; however, the major cause was the fact that over three million youth in the United States who would normally have been in school entered the armed forces or the labor force.[2] The University of Tennessee shared with the rest of the nation in this loss of students and potential students.

This condition resulted in greatly increased University enrollment in the years immediately following the war when it was estimated that more than ten thousand students would have been in attendance if they could have been accomodated.[3] Such increased enrollment was due in part to the fact that the Federal government furnished a tuition and subsistence allowance to large numbers of veterans who wished to obtain higher education. A summary of the effect of veterans on The University of Tennessee enrollment may be obtained

---

[3] In a conference July 18, 1950 President Emeritus James D. Hoskins provided this estimate with many suggestions and facts concerning factors entering into University enrollment.

from Table XVIII or by observing the graphic picture of this table in Figure 16. The top veteran enrollment figure was reached in the Fall quarter of the 1947 school year, i. e., Fall, 1946. From this enrollment of 4,829, veteran enrollment dropped to 2,490 in the Spring quarter of 1950. Since Public Law 346 provides that G. I. benefits will be operative until 1955 if the veteran has established an educational goal by July 25, 1951, it is probable that the effect of the veteran enrollment will be felt until 1955. In the years from 1953 to 1955 the major load from the veteran enrollment should fall on the graduate school since the trend of veterans working toward advanced degrees is becoming apparent.

Those alarmed by the rapid disappearance of veterans from the University have only to observe the solid line on Figure 16 for encouragement. This line shows a steady increase in non-veteran enrollment. Table XVIII shows that whereas in the Winter quarter of 1947 the non-veterans made up only 34.04 percent of the total enrollment with 2,367 students, three years later in the Winter quarter of 1950 the percentage was 57.58 with 3,840 students. This represented an increase of more than 1,500 non-veteran students.

Although the rapid rise in non-veteran enrollment may be explained partially by the fact that many of these persons occupied positions in the war-time labor force and in the post-war years obtained "deferred education", there remain other explanations. As an example, the transition from war-time to peace-time production brought with it special retraining for adult workers. Many of these adults who had never before considered attending college found themselves in a position where they had to obtain higher

TABLE XVIII

ENROLLMENT IN THE UNIVERSITY OF TENNESSEE, CLASSIFIED BY VETERANS AND NON-VETERANS, WINTER 1946-SPRING 1950

Year and quarter		Enrollment			Percentage of total	
		Total	Veterans	Non-veterans	Veterans	Non-veterans
1945-46	Winter	3,466	1,401	2,065	40.42	59.58
	Spring	4,386	2,381	2,005	54.29	45.71
1946-47	Summer	4,301	2,648	1,653	61.57	38.43
	Fall	7,442	4,829	2,613	64.89	35.11
	Winter	6,954	4,587	2,367	65.96	34.04
	Spring	6,685	4,355	2,330	65.15	34.85
1947-48	Summer	4,923	2,604	2,319	52.89	47.11
	Fall	7,690	4,418	3,272	57.45	42.55
	Winter	7,367	4,187	3,180	56.83	43.17
	Spring	6,867	3,896	2,971	56.74	43.26
1948-49	Summer	4,916	2,778	2,138	56.51	43.49
	Fall	7,790	4,034	3,756	51.78	48.22
	Winter	7,332	3,651	3,681	49.80	50.20
	Spring	6,897	3,410	3,487	49.44	50.56
1949-50	Summer	4,898	2,507	2,391	51.18	48.82
	Fall	7,212	3,109	4,103	43.10	56.90
	Winter	6,668	2,788	3,840	41.82	57.58
	Spring	6,382	2,490	3,892	39.02	60.98

Source: Registrar's Office.

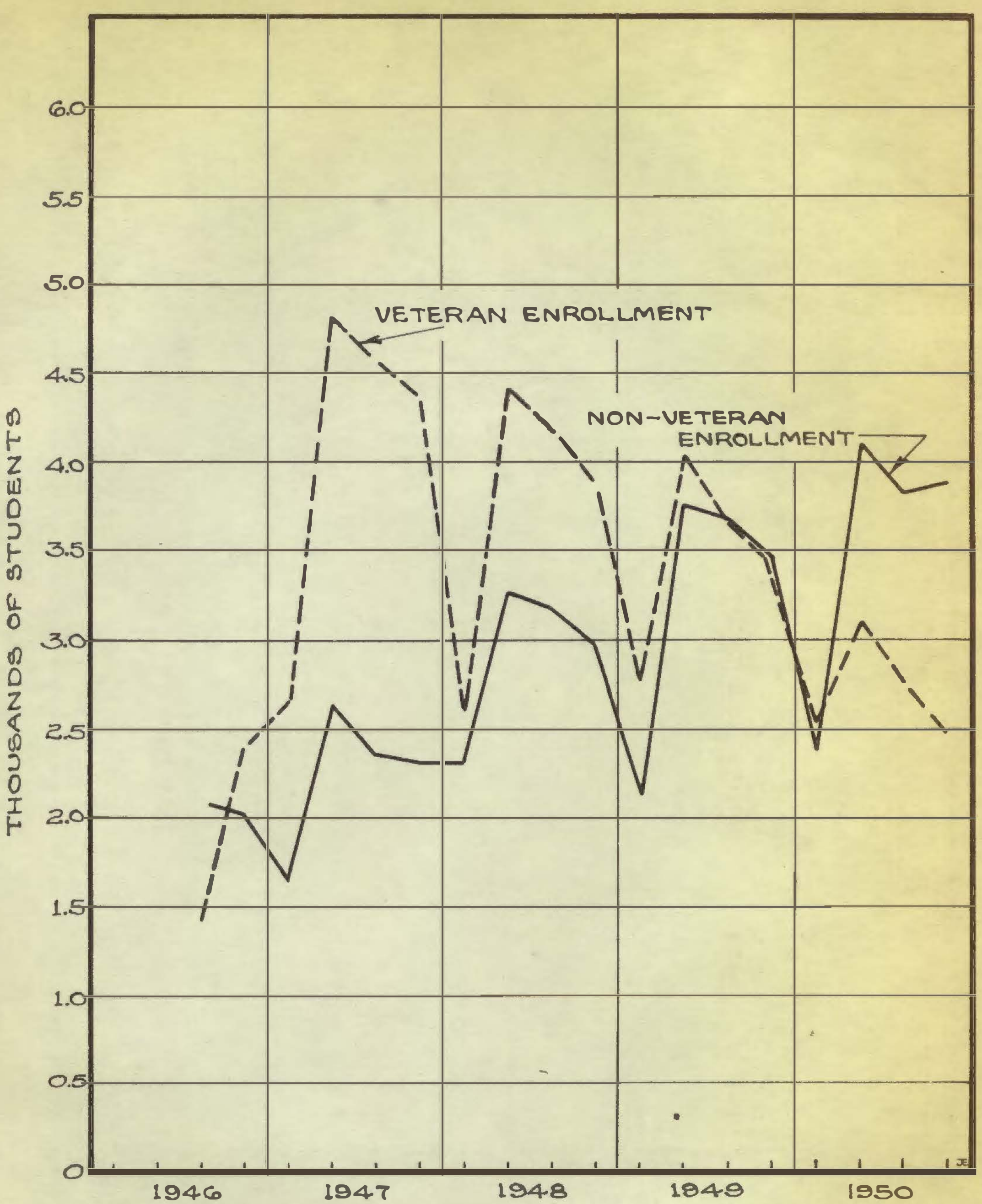

Figure 16. Veteran and non-veteran enrollment in The University of Tennessee, Winter 1946 — Spring 1950

Source of Data: Table XVIII

education if they were to compete for the better jobs.

There was also a social consciousness of education brought about by the disqualification of numerous men from service in the armed forces because of illiteracy. Knowledge of such a condition as this plus the fact that in 1940 there were eleven million adults who had not progressed beyond the fourth grade impressed the public with the feeling that such persons were hinderances not only to themselves but to society.[4]

Effects of Federal Aid to Education

The facts mentioned above have also served to focus the attention of the public on the "inequalities" in educational opportunity. In order to erase such inequalities, an attempt is being made to pass a "federal aid to education" bill which would in effect subsidize a person desiring higher education. Such a proposal is based on the idea that American people require the amount of education that will enable all citizens to manage their own political and economic affairs efficiently, maintain a high standard of living and realize their potentialities for a full and happy life. The nation would also require a supply of persons possessing superior abilities fully cultivated and socially directed by advanced education, e. g., teachers, ministers, artists, etc. Finally the advocates of the bill think that our system of social values requires the opportunity to pursue education in terms

[4]Dewhurst and Associates, op. cit., p. 316.

of ability and industry.[5]

Certainly the fact that such a bill is pending should be mentioned in any discussion of college enrollment. As to what effect federal subsidization would have on The University of Tennessee enrollment, no answer is available. The fact remains that the objectively determined forecasts of enrollment which were obtained in this thesis would need revision if such a bill is passed.

Effects of Specific Factors

The effects of a major war and "federal aid to education" have both been discussed as generalized causes for fluctuation in The University of Tennessee enrollment. Specifically, there are many variables which have a part in determining enrollment. One of these concerns the reputation of the faculty. Certainly if enough competent staff members cannot be employed to maintain the scholastic reputation of the University, the enrollment will drop. On the other hand, if the faculty be improved, increased enrollment might be expected.

Along the same line of reasoning, one of the factors which may affect the enrollment at the University is the availability of facilities. The word facilities may be construed to mean supplies, equipment, and particularly buildings. The forecast was made in Chapter IV that in the Fall quarter of

---

[5] Ibid., pp. 316-317.

[6] See footnote 3.

1970 the enrollment would be approximately ten thousand students. Certainly if at that time the University will accomodate, at the most, eight thousand students, the availability of facilities will be a limiting factor and will necessitate a revision in enrollment estimates.

Abnormal changes in population are factors which are not always measurable but which have and will influence University enrollment. One such change is the thousands of persons brought into this area to live and work in Oak Ridge, the atomic energy headquarters. Potential population growth in this area may be seen in the developments springing from further atomic research, e. g., the new cancer hospital.

Changes in curriculum, addition of new courses and departments, parents' attitude toward the University, physical location of the University in the state, and many other factors could be mentioned which influence in some manner the enrollment of The University of Tennessee.[7]

Conclusion

Any forecast is hazardous because knowledge of the future is imperfect. Nevertheless, forecasting is an essential part of planning since, constantly, judgments are being made as to the future; and commitments are being made based on these judgments. Past forecasts of University enrollment have been based on opinions, hunches, the most recent happenings, rule-of-thumb analyses, and in some instances blind guess-work. While any forecast is

---

[7] See footnote 3.

better than none, the forecast in this thesis which is based on a statistical analysis of past relations should help to provide a surer guide into the future. Its value as a guide will rest upon the average accuracy of the various enrollment estimates rather than upon the lucky naming of any one individual estimate.[8]

Forecasting enrollment in an organized and objective manner becomes more necessary as the University increases in size. Such forecasting is a part of a broad movement in America in the direction of more and better planning, analysis and control of operations. As an example, successful planning of long-term construction programs depends on reasonably accurate long-term forecasting. The quality of planning will improve as the quality of forecasting is improved, and forecasting will improve in some proportion to the number of persons making use of objective and sound techniques.[9]

During the first half of the 20th century, a rapidly expanding program of secondary education has undoubtedly been the major characteristic of the school system in the United States. The last half of the century is expected to produce similar results in higher education. If The University of Tennessee is to assume a proportional share of the responsibility for educating American youth, successful planning is necessary; and there can be no successful planning without the preliminary step of successful forecasting.

---

[8]Ezekiel, op. cit., p. 357.

[9]Research Arm of Controllers Institute, Business Forecasting (New York: Controllership Foundation, Inc., 1950), pp. 1-6.

BIBLIOGRAPHY

BIBLIOGRAPHY

American Association of University Professors. "The President's Commission on Higher Education," Bulletin, 34 (Spring 1948).

Bratt, Elmer Clark. Business Cycles and Forecasting. Chicago: Richard D. Irwin, Inc., 1948.

Brumbaugh, Martin A., and Kellog, Lester S. Business Statistics. Chicago: Richard D. Irwin, Inc., 1946.

Bureau of the Census, Department of Commerce. "Educational Attainment of the Civilian Population: April, 1947," Current Population Reports, Population Characteristics, p. 20, (May 4, 1948).

Bureau of the Census, Department of Commerce. "School Enrollment of the Civilian Population: October, 1949," Current Population Reports, Population Characteristics, p. 20, (April 26, 1950).

Bureau of the Census, Department of Commerce. "School Enrollment of the Civilian Population: April, 1947," Current Population Reports, Population Characteristics, p. 20, November 21, 1947).

Bureau of the Census, Department of Commerce. "School Enrollment of the Civilian Population: April, 1947," Current Population Reports, Population Characteristics, p. 20, (February 16, 1948).

Bureau of the Census, Department of Commerce. "School Enrollment of the Civilian Population: October, 1947," Current Population Reports, Population Characteristics, p. 20, (February 23, 1948).

Bureau of Economic and Business Research, University of Utah. "Enrollment in Institutions of Higher Learning, Fall Term, 1949," Utah Economic and Business Review, 10:4-5, (February 1950).

Burns, Arthur F., and Mitchell, Wesley C. Measuring Business Cycles. New York: National Bureau of Economic Research, 1946.

Controllership Foundation. Business Forecasting. New York: Controllership Foundation, Inc., 1950.

Croxton, Frederick E., and Cowden, Dudley J. Applied General Statistics. New York: Prentice-Hall, Inc., 1940.

Croxton, Frederick E., and Cowden, Dudley J. Practical Business Statistics. New York: Prentice-Hall, Inc., 1948.

Davies, George R., and Yoder, Dale. Business Statistics. New York: John Wiley & Sons, Inc., 1937.

Dewey, Edward R. and Dakin, Edwin F. Cycles. New York: Henry Holt and Company, 1947.

Engelhardt, Fred. Forecasting School Population. New York: Teachers College, Columbia University, 1925.

Ezekiel, Mordecai. Methods of Correlation Analysis. New York: John Wiley & Sons, Inc., 1941.

Ferber, Robert. Statistical Techniques in Market Research. New York: McGraw-Hill Book Company, Inc., 1949.

Freeman, H. A. Industrial Statistics. New York: John Wiley & Sons, Inc., 1942.

Grant, Eugene L. Statistical Quality Control. New York: McGraw-Hill Book Company, Inc., 1946.

Mills, Frederick Cecil. Statistical Methods. New York: Henry Holt and Company, 1938.

Mitchell, Wesley Clair. Business Cycles and Their Causes. Los Angeles: University of California Press, 1941.

Neiswanger, William A., Elementary Statistical Methods. New York: The MacMillan Company, 1948.

Schwartz, Charles F., and Graham, Robert E., Jr. "State Income Payments in 1947," Survey of Current Business, 28:10-21, (August 1948).

Stockton, John R. An Introduction to Business Statistics. Atlanta: D. C. Heath and Company, 1938.

Thaden, J. F. "Forecast of Future Public School Enrollments, by Grades, in Michigan," Michigan Agricultural Experiment Station Quarterly Bulletin. 21: 378-386, (May 1949).

United States Office of Education, Higher Education Division. "College and University Building Needs," Higher Education, IV (May 15, 1948.

United States Office of Education, Higher Education Division, "Estimates of Future College and University Enrollment," Higher Education, IV (March 15, 1948).

United States Office of Education, Higher Education Division, "Federal and State Cooperation in Education and Research," Higher Education, V (January 15, 1949).

United States Office of Education, Higher Education Division, "President's Commission on Higher Education Reports Further," Higher Education, IV (March 1, 1948).

Wright, Wilson. Forecasting for Profit. New York: John Wiley & Sons, Inc., 1947.