

1970

# The use of multiple regression and path analysis in analyzing success in journalism at Iowa State University

Richard Lee Byerly  
*Iowa State University*

Follow this and additional works at: <https://lib.dr.iastate.edu/rtd>

 Part of the [Education Commons](#)

## Recommended Citation

Byerly, Richard Lee, "The use of multiple regression and path analysis in analyzing success in journalism at Iowa State University " (1970). *Retrospective Theses and Dissertations*. 4292.  
<https://lib.dr.iastate.edu/rtd/4292>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

71-7253

BYERLY, Richard Lee, 1938-  
THE USE OF MULTIPLE REGRESSION AND PATH  
ANALYSIS IN ANALYZING SUCCESS IN JOURNALISM  
AT IOWA STATE UNIVERSITY.

Iowa State University, Ph.D., 1970  
Education, theory and practice

University Microfilms, Inc., Ann Arbor, Michigan

THE USE OF MULTIPLE REGRESSION AND PATH ANALYSIS IN  
ANALYZING SUCCESS IN JOURNALISM AT IOWA STATE UNIVERSITY

by

Richard Lee Byerly

A Dissertation Submitted to the  
Graduate Faculty in Partial Fulfillment of  
The Requirements for the Degree of  
DOCTOR OF PHILOSOPHY

Major Subject: Education

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Area

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University  
Ames, Iowa

1970

## TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
A. Background and Setting	1
B. Statement of the Problem	2
C. Objectives of the Study	6
II. REVIEW OF LITERATURE	8
A. Theory Linking Causation and Educational Research	9
B. Causation	10
C. A Review of Empirical Studies That Relate to the Methodology or Factors of Student Achievement in Journalism	14
D. A Review of Multivariate Statistical Concepts and Empirical Results of These Conceptualizations	18
E. Multiple Regression Analysis	20
F. Path Analysis	26
G. Theoretical Concerns of Path Analysis	28
H. Summary	30
III. METHOD OF PROCEDURE	32
A. Selection of the Population and Collection of Data	33a
B. Criterion of Success in Journalism at Iowa State University	34
C. Prediction Variables	34
D. Basic Assumptions	38
E. Hypothesis to Be Tested	39
F. Treatment of Data	40
G. Variable Selection	40
H. Regression Techniques	41
I. Path Analysis	49
J. Summary	59
IV. FINDINGS	62
A. Chapter Contents	62
B. Backward Solution Regression	62
C. Forward Regression Solution	74
D. Stepwise Regression Solution	78
E. Single Stage Path or Network Analysis	81
F. Analyzing Relative Variable Strengths	85
G. Path Analysis	86
H. Direct Effects	87
I. Computation of Residuals for Multi-Stage Path Model	90
J. Interpretation of the Multi-Stage Model	93

TABLE OF CONTENTS  
(Continued)

	Page
V. DISCUSSION	95
A. Comparison of the Population of the Study to General University Norms	96
B. Comparison of the Three Multiple Regression Techniques	97
C. Implications of the Regression Models	98
D. Relative Importance of Variables	99a
E. Utility of the Regression Findings	100
F. Implications from Path Analysis	102
G. Limitations of the Study	103
H. Recommendations for Further Study	104
VI. SUMMARY	106
VII. LITERATURE CITED	110
VIII. ACKNOWLEDGEMENTS	116

## LIST OF TABLES

Table	Page
1 Mean scores and standard deviations of the criterion and predictor variables	63
2 Product-moment correlation coefficient matrix for predictor and criterion variables	65
3 Summary of analyses of multiple regression for all combinations of the predictor variables ( $X_1$ , $X_2$ , $X_3$ , $X_4$ , and $X_6$ )	66
4 Analysis of multiple regression using five of the variables ( $X_1$ , $X_2$ , $X_3$ , $X_4$ , $X_6$ )	68
5 Analysis of multiple regression using the four predictor variables ( $X_2$ , $X_3$ , $X_4$ , $X_6$ )	70
6 Test for loss in predictive ability due to the elimination of the English writing ability variable ( $X_1$ ) from the five variable regression equation	71
7 Analysis of multiple regression using the three predictor variables ( $X_2$ , $X_3$ , $X_4$ )	72
8 Test for loss in predictive ability of the criterion due to the elimination of the written English ( $X_1$ ) and physical science variables ( $X_6$ ) from the five variable regression equation	74
9 Rank ordering of predictor variables by size of correlation with criteria	76
10 Forward solution: partial regression coefficients	76
11 Forward solution: changes in characteristics of the regression equations	78
12 Stepwise solution: partial regression coefficients	79
13 Stepwise solution: changes in characteristics of the regression equation	81
14 Summary table for single stage path analysis	83
15 Partial regression coefficients associated with the regression equations in the recursive set	88

Table	LIST OF TABLES (Continued)	Page
16	Summary of multi-stage path analysis	91
17	Residual path coefficients	93
18	A comparison of backward, forward, and stepwise regression solutions	97
19	Standardized beta coefficients for the three multiple regression techniques	99b

## LIST OF FIGURES

Figure		Page
1	Path Model I (single stage path or network model)	53
2	Path Model II (multi-stage model)	56
3	Final single stage path Model I	84
4	Final multistage path Model II	89



## I. INTRODUCTION

This chapter contains: (1) an introduction providing the background and the factors for prediction of student success in journalism at Iowa State University; (2) a statement of the problems including the basic assumption, purposes, problems, and working hypothesis; (3) the sources of data for the study.

### A. Background and Setting

The 1960's ushered in an ever increasing wave of students to the college campuses. However, this influx of students was dispersed over all academic departments. Price (53) wrote that journalism enrollment soared to record highs in the mid 1960's and universities across the nation showed journalism enrollment gains at every level. A national total of 19,229 journalism students were enrolled in 1965. Consequently, with ever expanding enrollments, college journalism departments found that more thorough research of the academic behavior of their graduates was necessary. Academic selection or screening practices may become a necessity in the future.

Typical of the national scene, the Iowa State journalism department announced sharply increased enrollments for the years 1967-69. While the entire university enrollment predictions for 1970 indicated an approximate 10 percent decrease, the Department of Journalism and Mass Communication estimated an approximate 34 percent increase in enrollment. Therefore, for practical reasons, a need to investigate the academic

behavior and traits of journalism students and subsequent graduates became evident.

One other event further accentuated the need for study of student's success in journalism. The 1970 Iowa General Assembly, reduced the budget requested by the Iowa Board of Regents. The results were decreased funding for the state supported universities. Faced with the increasing enrollments and possible shortage of funds the need for further delineation of skills necessary for student academic success was apparent. Thus, the basis of this study was initiated.

To obtain a representative population, journalism graduates from a five year period (1965-1970) were studied. The frame or population of these years was 217 graduates. Final data included the 215 students whose permanent records were found.

Using multiple regression techniques and path analysis, the final prediction equations and path models were reached.

#### B. Statement of the Problem

The problem of the study was to develop, investigate, and analyze the academic patterns of the Iowa State journalism graduates from 1965-1970 and attempt to ascertain and determine possible insights or inferences regarding their academic success. For this study, student success in journalism was determined by the students' completion of the requirements necessary for graduation from Iowa State Department of Journalism. The student grade point averages, then, were considered as a quantitative measure of the students' academic success or achieve-

ment. This study was further designed to develop, investigate, and analyze the data in terms of developing a more thorough pattern of investigatory procedures relevant to the statistical methodology used.

Two general purposes were defined: (1) to determine the influence of student achievement in university academic areas, previous achievement, and aptitude on student's academic success in journalism, and (2) to compare and contrast insights and inferences regarding student's academic success resulting from various multivariate statistical approaches such as multiple regression and path analysis. More specifically, the problems were:

1. To develop a methodological and theoretical approach to analyze the data using three multiple regression techniques.  
Specifically:
  - a. backward solution
  - b. forward solution
  - c. stepwise solution
2. To develop a theoretical and conceptual framework for the methodological approach of the path analysis and causal investigation on predictive or explanatory data.
3. To compare and contrast the insights and inferences gained from the different multivariate statistical methods.
4. To determine what combination of academic predictor variables might be used to predict student academic success in journalism at Iowa State University.

The rationale underlying this study is stated in the following basic assumption: Student success in journalism at Iowa State University, as indicated by graduation and measured by grade point average, is dependent upon a combination of behavioral factors unique to each individual. This rationale generated the null hypothesis which was tested in the study: No combination of academic predictor variables can be used to predict students' academic journalism success at Iowa State University as measured by the graduates' records from 1965-1970.

The stated hypotheses necessitated the selection and delineation of predictor variables in these areas:

1. academic achievement at Iowa State University
2. academic achievement at high school
3. test aptitude prior to college entrance.

The implications of this study could be helpful for guidance of prospective and presently enrolled journalism students. Professors and instructors might also be able to imply what specific academic skills or criteria appear to be related to student achievement.

For the initial development of this study and the appropriateness of the data for conceptual model building, the usefulness for prediction of students' first quarter grade-point-averages at Iowa State University was investigated by first viewing these predictor variables.

1. English Placement Test (EPT)
2. Minnesota Scholastic Aptitude Test (MSAT)
3. High school rank (HSR)

Student achievement or success in academic endeavors at Iowa State University was evaluated by the grade point average for all

areas of study at the termination of the first quarter.

In further developing this predictive study, the utility for prediction of student journalism success, as measured by the student's journalism grade point at Iowa State University, was investigated by then viewing these additional predictor variables.

1. English and speech success at Iowa State University (ES)
2. Humanities success at Iowa State University (HUM)
3. Social science success at Iowa State University (SS)
4. Life science success at Iowa State University (LS)
5. Mathematics, statistics, and computer science success at Iowa State University (MATH)
6. Physical science success at Iowa State University (PS)
7. High school rank (HSR)
8. Minnesota Scholastic Aptitude Test (MSAT)
9. English Placement Test (EPT)

However, since it was theorized that success in journalism was, in part, a measure of writing and communication skills, the English training courses were partitioned into two areas and listed as:

1. Student's English writing achievement at Iowa State University which remained as variable  $X_1$ .
2. The English literature courses, which were listed under the classification of humanities.

The classification of college courses into categories was determined by using the indices given by the Iowa State University general catalog (38).

### C. Objectives of the Study

The objectives of the study were:

1. To determine what academic areas could be used as predictors in determining student academic success in journalism at Iowa State University.
2. To attempt to ascertain whether or not certain multiple regression applications were capable of yielding additional insights when employed in the analysis of the data.
3. To attempt to ascertain whether or not additional insights were gained when the causal framework was utilized in the analysis of the study.
4. To provide methodological information which might be utilized in applying statistical applications to future predictive studies.
5. To attempt to illustrate the additional insights and inferences gained by utilizing the causal approach in the analysis of the data of this study.
6. To demonstrate the applicability and relevance of conceptual model building in predictive and educational research.

To accomplish these purposes or objectives, the dissertation is divided into six chapters. The second chapter will be a discussion of the review of literature relevant to the appropriate segments of the study. The methodology for data analyzation and design for the study are discussed in the third chapter. The fourth chapter will include

all findings and analyses results of the study. The fifth chapter includes a presentation and discussion of the data collected. The final chapter will summarize the findings and provide a summary of the findings of the study, conclusions, and recommendations for further study.

## II. REVIEW OF LITERATURE

The primary purposes for the review of literature are: 1) to determine what theoretical work has been done in the area of concern, 2) to review what empirical work has been accomplished in the area of student achievement in college journalism, 3) to provide a sound basis for the theoretical framework of the study; 4) to provide a basis for analyzing and interpreting the data of the study; 5) to provide insights as to what operational techniques, methods, and procedures might assist in conceptual model building relevant to this particular journalism study.

Basically, the review of literature attempted to set the foundation for this study. The review of literature can best be used by integrating some of its relevant materials into the appropriate subdivisions of this chapter. These categories were essential in prefacing the later sections of this study.

The review of literature is presented under the following headings:

- A. Theory linking causation and educational research
- B. Causation
- C. A review of empirical studies that relate to the methodology or factors on student achievement in college journalism programs.
- D. A review of multivariate statistical concepts and empirical studies of the following statistical conceptualizations.
  1. General regression comments
  2. Multiple regression analysis



- a. Backward solution
  - b. Forward solution
  - c. Stepwise solution
3. Path analysis
- a. Historical review
  - b. Insights of path analysis
  - c. Objectives of path analysis
  - d. Theoretical concerns of path analysis
- A. Theory Linking Causation and Educational Research

Probably a more inclusive view of the methodological approaches to the problem will be gained by scanning the relevant areas of conceptual and theoretical concern. This will be done separately in more detail later in the study, but will be generalized in this section.

First of all, multivariate techniques were not a recent innovation of statistical investigation (63). More specifically, causation was not a recent adaptation in statistical thinking. Indeed, the debate on cause as such is as old as recorded history (28). Philosophically, the views regarding statistical causation were widely diversified. To engage in a lengthy philosophical debate of statistical causation would be, however, beyond the scope of this study. However, the theoretical basis for causal model approaches certainly was considered and time invested therein.

As causal theory became more inclusive, from genetic research

(63) to social sciences (6, 9), the relevant theory changed as well. Thus, statements representing causal or explanatory relationships related to more and more areas of educational concern. Social scientists, concerned with the component aspects of such areas as the decision-making processes, were beginning to explore new and vitally more inclusive methods of investigation and analysis. Causal theory, which will be developed more thoroughly in a later section, came into prominence with a decision-making orientation (14).

Educational research, while focusing its attention on a vast area of study, adapted various multivariate techniques as a means to an end.

#### B. Causation

The theory of causation has been discussed and studied for many years. The important concern for this paper, however, is focused on the needed theoretical approach for the proposed methodology.

How has cause been defined? Nagel (47) defined the cause of an event as anything which is thought to be, to some varying degree, responsible for that event. One point which grew from this theoretical view was that cause was not said to exist but merely assumed to exist.

Causation, as we know it today, evolved from a natural science background. The cause-and-effect theory was begun by natural scientists looking for curative and generative techniques and variables (13). Physicists, chemists, and geneticists were the scientists most involved

in the initial stages. The theory which the natural scientist usually proposed was the deterministic theory.

The deterministic theory stated: (30)

If X then Y or  $X \rightarrow Y$ , which says that if X occurs, then a Y will follow or, if Y is found, then X is within the causes that preceded the event.

Therefore, X was an example of the true cause-effect theory of that day. Essentially, it agreed with John Stuart Mills' method of agreement, or as Cohen stated:

If two or more instances of phenomena under investigation have only one circumstance in common, the circumstance in which alone all the instances agree is the cause of the given phenomena (18, p. 251).

The arguments and extensions of these developments increased until causation theory had been expanded to include the following theories:

Method of difference:

If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring in the former, the circumstances in which alone the two instances differ is the effect, or the cause, or an indispensable part of the cause, of the phenomenon (18, p. 256).

Method of residues:

Subduct from any phenomenon such part as is known by previous inductions to be the effect of certain antecedents, and the residue of the phenomenon is the effect of the remaining antecedents (18, p. 264).

Then, the utilization and extensions of these theories by the social scientist moved the theory to a more all-inclusive proposition of causation and explanation. Consequently, the theory of concomitant variation, the basic theoretical consideration of this study, came into

observance. Essentially, concomitant variation is stated as:

Whatever phenomenon varies in any manner, whenever another phenomenon varies in some particular manner, is either a cause or an effect of that phenomenon or is connected with it through some fact of causation (18, pp. 261-262).

This outlook on causation was referred to as multiple causation. Nagel (47) and others of the same theoretical beliefs have forwarded the present approach to cause and effect. Basically, they state that a cause might have been partially responsible in the "causing" of the event taking place. Probably, a more important aspect of this theoretical conceptualization can be derived. That is, a set of causal factors cannot completely claim absoluteness. Blalock (2) has argued on somewhat the same basis. He suggested that special attention should be given to the delineation of causation, mainly noting cause and effect of the nomological networks. However, Blalock has pointed out that the issue of causation rests solely on the theoretical basis. He stated that causal thinking belongs completely on the theoretical level and that causal laws can never be completely demonstrated empirically.

What this entire causation trend did was to attempt to better understand the universe as a continuous, changing environment. Blalock said:

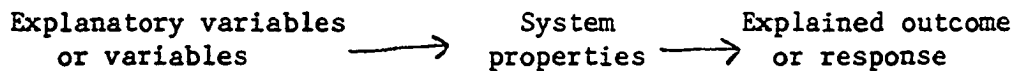
If we ever wish to understand the nature of the real world, we have to act and think as though events are repeated and as if objects do have properties that remain constant for some period of time, however short.... One way of dealing with the problem is to make use of theoretical models of reality (2, p. 7).

Theory, then, was merely a picture of reality, snapped at a particular time and focusing on a changing set of events and causes (33).

The focal point of the theory purported by this study will be derived from the interventionist model theory. According to Feuer (32), the interventionist model of causal theory has become more prominent in the past twenty years.

The intervention theory relates that an individual or group of individuals can manipulate a situation and the existing state of affairs so that what follows would not have occurred without the intervention (43).

Theorists, at the beginning of the 1960 decade, began to take a more inclusive view. Mario Bunge (14) stated that causation on causal models were component relationships of changing situations. The notion of causation which this study used was to study a causal system in terms of explanatory magnitude. The theoretical model for this explanatory relationship became:



Since the boundary frame of this explanatory system can be governed only by the researcher's choice, Blalock (2) suggested that all variables of this model be considered as causal variables.

Simon (56) listed asymmetry as one component part of causation. Asymmetry, unlike functional relationships, is listed as irreversible.

The function of empirical research in the causal approach developed and assisted the conceptual model theorists with practicum results. Causal methodology, then, can be used to demonstrate the goodness-of-fit of the conceptual model building stage. To further the view, Blalock states,

The fact that causal inferences are made with considerable risk of error does not, of course, mean that they should not be made at all. For it is difficult to imagine the development and testing of social science theory without such inferences. Since they are in fact being made in practical research....(5, p. 5).

Finally, the function of an empirical study and the subsequent data generated, are to "assist" the preliminary causal model.

### C. A Review of Empirical Studies That Relate to the Methodology or Factors of Student Achievement in Journalism

DeFotis (24) investigated and analyzed the working functions of the writer and attempted to suggest the educational means to achieve these goals. The investigation attempted, by reviewing the history of technical writing, to study the characteristics of the curricula. Viewing the discipline from sources such as technical writers, journalism instructors, educators, and science journalists, the study was able to present various sides.

The study concluded that the most important facet of technical writing was the "scientific and technical information" to be presented and not the specific "form" of writing. The study further concluded that the technical writer should take his primary course work in his major area, and then develop his communication skills. In this way, technical writing would be classified as an integral part of any science or curriculum, and not just an extra dimension.

Feinberg (31) attempted to determine entrance practices and criteria for admission of foreign journalism students. Utilizing data from a 1965-66 field survey of 13 colleges and universities, a great variety

of practices and policies regarding foreign students was illustrated. A degree of competence in English grammar was termed as desirable for student success in journalism.

Shaner (55) looked for significant differences between college students with high school journalism background and those without a high school journalism background. The study attempted to compare these differences with regard to their performance in college journalism classes and exposure to and attitudes towards the mass media. High school journalism programs were rated by writer's opinion "above average", "average", "below average".

Students were compared in: (1) grades in college journalism and (2) English and speech classes.

Skills such as ability to write and ability to determine a publication's position upon current issues were examined. The attitudes measured were as follows:

1. Attitude toward high school journalism
2. Development of self-confidence from journalism courses.

The study concluded that attitudes are extremely important to success in all journalism programs.

A study by Spencer (58) investigated the effects on readers of grammatical errors in written English. The dependent variables were:

1. Assessment of quality writing
2. Attitude toward writing
3. Comprehension.

The purposes were to test the theory that incorrect grammar didn't affect comprehension as much as it affected the attitude of the reader toward the writing. In support, language theorists believe that, since the English language is highly redundant, then errors in grammar might be overcome by the repetition. However, it was also perceived that these errors might have an effect on the reader. The sample of 350 college freshmen were given essays to read. Tests of quality assessment and comprehension followed. The results of the study illustrated:

1. Attitude toward written message was affected by grammatical error.
2. Comprehension appeared not to be affected by grammatical error.
3. Level of verbal aptitude had an affect on perception of error.
4. Verbal aptitude did not appear to have an effect on comprehension.

The study concluded that the concept of "incorrectness" as used by many English writing teachers needs further study and the area of response to written messages needs even further study.

Studies have been conducted in English composition. However, since it is considered a similar conceptual area, the results of these investigations were of concern.

Using grades and standardized tests, Kunhart and Olsen (39) attempted to predict the success of students in freshman college English at Hartwell College, Salinas, California. The sample consisted of only students who had scored between the 30th and 75th percentiles on the Cooperative English Test. Students in this range were required to pass a review of high school (English A) with a C or better. If these criteria



were met, students could be admitted into a college English composition course. The English composition grade was the dependent or criterion variable.

Basically, the investigation had two groups of students. In the first section (N=163) the predictor variables were English A grade, high school intelligence quotient score, and a high school English grade.

The correlations with the composition grade were as follows: English A grade  $+0.303$ , I.Q. score  $-0.052$ , and high school English twelve grade  $+0.306$ . The multiple correlation (R) was  $.418$ .

From the original sample (N=163) a second strata of fifty was selected. This group included the following predictor variables: English A grade, high school intelligence grade, the cooperative English Test score, high school English grade, and the American Council of Educational Psychological Test Linguistic Score. The correlations with the English Composition grade were: English A grade  $+0.333$ , I.Q. score  $+0.277$ , Cooperative English Test Score  $+0.307$ , and the American Council of Psychological Test  $+0.174$ . A multiple correlation (R) from this group of fifty was  $.531$ .

Kunhart and Olsen concluded that an improvement of predictions for composition was accomplished as additional information was added to the study. Thus, the addition of more variables that relate to writing increased predictive ability.

D. A Review of Multivariate Statistical Concepts and Empirical Results of These Conceptualizations

1. General multiple regression comments

In the conventional framework of regression analysis certain problematic processes are encountered. For example, a research problem may relate to how well a variable or a set of variables can predict a particular model of variable relationships. In this conceptual framework, to determine the predictive ability of a set of variables to explain another variable, multiple linear regression and correlation are the techniques utilized. Some key to variations in the dependent variable might be formulated by subtracting each score from the calculated mean for that variable. These deviations, called deviations from the mean, are then squared and summed to obtain the corrected sum of squares. With this foundation, the two component parts of the corrected sum of squares for the dependent variable can then be analyzed. The two parts, the sum of squares due to deviations from regression and sum of squares of regression, are partitioned to determine their relative contribution to the total sum of squares. The sum of squares due to regression represents the combined effects of the independent variables. This is referred to as explained variation.

What happens in this step is indicative of the entire predictive process. Thus, this reduction in total sum of squares attributable to the regression of the dependent variable on the group of independent variables is an index of the amount gained by using these predicted dependent variable scores rather than the mean of the dependent variable

for prediction. A statistical "F" test is employed to evaluate if the reduction in the sum of squares due to regression is significant. This test is actually a ratio of the sum of squares due to regression divided by the degrees of freedom involved to the sum of squares due to deviations from regression divided by the correct degrees of freedom (50).

From this point another important relationship is developed. Using this multiple regression equation with a given set of independent variables, a degree of the closeness between the predicted dependent variable scores and the observed scores is the multiple correlation coefficient,  $R$ . The larger the multiple correlation coefficient the closer the association. Using this concept, the square of the multiple correlation,  $R^2$ , represents the ratio of the explained variation to the total variation (44). The researcher is able to interpret the proportion of total variation in the dependent variable which is explained by the set of independent or predictive variables (44).

Essentially, if  $R^2$  is small, most of the dependent variable is unexplained (44). A large  $R^2$  indicates that the regression method accounted for much of the variation in the dependent variable. Various regression techniques such as forward solutions, stepwise procedure, and backward solution are sometimes employed by the researcher to develop the regression equation (22).

## E. Multiple Regression Analysis

### 1. Backward solution

In the backward solution model, given the model in which the possibility of all independent variables not contributing to the explained variance is possible, the problem of elimination of those variables not contributing is encountered. As surmised, the statistical and empirical objective of this process is revolved around the systematic removal of all non-contributing variables. Backward solution was one systematic attempt. Draper and Smith (25) related the procedural sequence (1966:167-168) which involves the steps further developed in Chapter III of this study.

Edwards (29) used multiple regression and path analysis statistical techniques in a study to predict innovativeness of Iowa farmers. Demographic and social psychological factors were utilized in order to define the conceptual area.

Innovativeness was the dependent or criterion variable and the predictor variables were:

1. Age ( $X_1$ )
2. Education ( $X_2$ )
3. Scale of operation ( $X_3$ )
4. Cosmopolite behavior ( $X_4$ )
5. Media use ( $X_5$ )
6. Scientific orientation ( $X_6$ )
7. Risk orientation ( $X_7$ )

All possible combinations of predictor variables were explored before the final equation was determined. Using multiple regression techniques, the best predictor variables were found to be ( $X_1$ ) scale of operation, ( $X_4$ ) cosmopolite behavior, ( $X_6$ ) scientific orientation, and ( $X_7$ ) risk orientation, and the final predictor equation, in corrected notation, became:

$$Y = b_3X_3 + b_4X_4 + b_6X_6 + b_7X_7 .$$

In an educational study, Byerly (15) used the backward solution multiple regression technique in a linguistics study to derive prediction equations for two groups:

1. A group of students having a high school linguistics background
2. A group of students not having a high school linguistics background.

This study was further developed and a paper, "A Case for Network Analysis" (16) was presented at the 1970 American Educational Research Association (AERA) Convention in Minneapolis, Minnesota. That paper illustrated the applicability of the backward elimination multiple regression solution to conversion of standardized regression coefficients in the explanatory path analysis framework. The backward elimination, multiple regression method is determined by the variables included in the initial conceptual model.

In another educational study, Netusil (48) utilized backward multiple regression techniques to determine what factors of salary changes for Iowa superintendents, secondary school principals, and

beginning teachers were instrumental in changes. Twenty-two descriptive variables were considered as predictor variables. Regression analyses were performed on the criterion variables and final equations were determined after applying the statistical test for loss.

## 2. Forward solution

Forward solution is predicated upon the premise that a predetermined order for the independent variables has been established (25). --Those variables exhibiting the highest degree of correlation with the dependent variable are normally thought to be of the highest degree for inclusion into the model. Variables are then added to the model in order of descending importance.

The forward solution begins with a given, selected number of variables and proceeds to test each successive variable by means of the sequential procedures. A more detailed statistical discussion will be undertaken in Chapter III.

## 3. Stepwise regression

The stepwise regression model involves the opposite procedural operations from the backward solution. Whereas the backward solution is an elimination process, the stepwise is a forward additive system whereby variables are combined with those variables already given. Similarly with the backward regression technique, the stepwise regression technique has outlined an explicit objective -- to retain only those variables which contribute to the explained variance (35).

Simplified, this regression solution is based upon a partial correlation procedure which minimizes the residual sum of squares. Initially, the process begins with the "most important" predictor variable and continues until any or all of three termination procedures on conditions occur.

Essentially, then the process involves entering each successive variable and then checking that variable for each of three termination conditions. If the variable fails on any, the process stops and the equation is finalized. If, however, all three provisional conditions are not met, the entire cycle is repeated.

A more intensified discussion of the statistical components of the stepwise regression method, including the termination criteria, will be presented in Chapter III.

Numerous studies involving the utilization of several multivariate statistical techniques have been done in education, but few of the studies have employed and contrasted the uses of these techniques.

Lee (43) compared the techniques when using multiple regression and path analysis in a methodological study at Iowa State University to study role behavior. The emphasis of the study centered upon the determinants of role behavior. Primarily, the role behavior chosen was that of managers of farm cooperatives.

The major objectives of the thesis were:

1. To delineate theoretical determinants of role behavior and construct causal models of role behavior.
2. To apply path analysis to use in the adequacy of models and modifications of the model after data analysis.

3. To determine recommendations of improvements in measures used in testing the model.

The study stated that use of path analysis was beneficial in making relationships among independent variables explicit as well as identifying and quantifying the magnitudes and relative strengths of each independent variable to the dependent variable.

Coward (22) used two forms of multivariate regression techniques and contrasted and compared them to predict and generate models dealing with the subsistence of commercial transition in agricultural development. The two regression techniques that were used were backward multiple regression and stepwise multiple regression solution.

Coward's study focused its comparison of the two techniques on: (1) the variable included in each of the two solutions and (2) the importance of the variables in each solution. From the analyses 17 different variables were included in the final two solutions. Nine of the 17 variables were common to both solutions:

1. Risk preference score ( $X_5$ )
2. Cosmopolitaness score ( $X_8$ )
3. Farm size ( $X_9$ )
4. Information orientation score and market score ( $X_{10}$ )
5. Information orientation score and risk preference score ( $X_{13}$ )
6. Market orientation score and risk preference score ( $X_{18}$ )
7. Credit orientation score and risk preference score ( $X_{22}$ )
8. Credit orientation score and cosmopolitaness score ( $X_{23}$ )
9. Control-over-nature score and risk preference score ( $X_{25}$ ).



In the backward solution the top four variables in terms of largest standard beta coefficients are:

1. Risk preference score ( $X_5$ )
2. Control-over-nature score and risk preference score, ( $X_{25}$ )
3. Credit orientation score and control-over-nature score ( $X_{21}$ )
4. Information orientation score and risk preference score ( $X_{13}$ ).

This technique yielded an important insight: The importance of risk preference score, both alone and in interaction, as an important variable was apparent.

In stepwise, as in backward, the actual variable names were important only to illustrate the similarities of the two techniques.

In the stepwise solution the four variables with the variables listed as the top four were:

1. Risk preference score ( $X_5$ )
2. Control-over-nature and risk preference score ( $X_{25}$ )
3. Cosmopolitanness score ( $X_8$ )
4. Information orientation score and risk preference score ( $X_{13}$ ).

The common variables of the four given for both solutions were given. These variables were listed as:

1. Risk preference score ( $X_5$ )
2. Control-over-nature score and risk preference score ( $X_{25}$ )
3. Information orientation score and risk preference score ( $X_{13}$ ).

By utilizing these two techniques and comparing them, the top variable strengths were given. The variable which exhibited the most importance was the risk preference score.

## F. Path Analysis

### 1. Historical review

The statistical multivariate method of path analysis was developed shortly before 1920 by the well-known geneticist Sewall Wright (65). It wasn't until the 1920's, however, that the technique appeared in publications (64, 65). Although the technique was published in the 1920's, it was not widely accepted. The main application of the technique was used only within the discipline of genetics. The reason for the limited reception of the technique may be found in its theoretical dimension.

The theoretical basis in causal terms limited the model conceptualization to the cause and effect notion. Thus, the causal relationship was stated: A causes B.

In recent years, changes in theory regarding causation has moved the theory into the explanatory framework. That is, the model was now stated: A explains B. Increased uses of path analysis in the field of social sciences followed (26). This date climb was in sharp contrast with prior use.

To illustrate, the Educational Index shows how little causal notation had been used. Only three articles before 1965 appeared in journal listings. However, the 1969 Index listed over fifty articles.

A derivative of the path analysis technique, called the Simon-Blalock approach, helped to stimulate further interest in the technique (1). Boudon (12) in 1965 termed the Simon-Blalock technique a "weak"

path analysis. Actually, the Simon-Blalock technique is formed by using only the first three steps of the path analysis approach (60).

From 1966 until 1968 Duncan (26) demonstrated the utility of the path approach in studies dealing with demographic variables. The first entrance of the path technique to educational data was accomplished by Duncan and Blau (27). In this study they applied the path technique to a study involving occupational status with educational variables.

Several additional factors have prompted the utilization of path analysis into the research of the social sciences. Most prominent, however, was the theoretical concern regarding causation. What in 1920 began as a pattern of "cause and effect" (65) was later redefined to be inclusive of social science research.

In summary, the present causal theory states, as viewed by Mario Bunge (14), that an independent variable may "explain" a portion of the variance of the dependent variable.

## 2. Insights gained from the use of path analysis

1. Path analysis regards and defines the magnitude of the various predictor variables and gives a method for the assessment of these unidirectional strengths.
2. Path analysis facilitates the use of the calculated direct magnitudes (paths) in connection with the correlation coefficients to calculate the indirect components of the model.
3. A measurement of the error component for the model is calculated and is represented by the residual path component.

4. It provides standardized indices upon which the researcher can easily assess relativeness of measurement.

### 3. Objectives of path analysis

Essentially, two basic objectives of path analysis were aptly classified by Coward. The objectives were:

First to identify the form of the network of relationships that exist and second to analyze this network into the direct and indirect relationships of which it is composed (22, p. 14).

### 4. The application of path analysis

Path analysis has a role in research that Duncan defined as:

The technique of path analysis is not a method for discovering causal laws but a procedure for giving a quantitative interpretation to the manifestation of an unknown or assumed causal system as it operates in a particular population (26, p. 177).

With regard to role, Wright stated:

Path analysis is an extension of the usual verbal interpretation of statistics not of the statistics themselves. It is usually easy to give a plausible interpretation of any significant statistic taken by itself. The purpose of path analysis is to determine whether a proposed set of interpretations is consistent throughout (66, p. 444).

### G. Theoretical Concerns of Path Analysis

The theory which precedes any causal or path analysis is considered to be equally as important, if not more important, than the methodological analysis. The theoretical concerns are governed by these assumptions or criteria: (43)

1. Causal ordering must be assumed before constructing the path diagram: Causal ordering is external or a priori with respect to the data analysis. From the data analysis, an investigator merely obtains the information as to the convariation of variables. Such information will only support an investigator to assess the adequacy of his models.

As far as causal orderings are concerned in this thesis, the essential element is that change in one variable is assumed, to a degree, to contribute to a change in another variable but not that one is a "cause of another's existence".

2. To adequately cope with sampling error, a large sample size is required. Also, measurement errors in all variables should be small.
3. The relationships among the variables are assumed to be additive, linear, and asymmetric. Also, the variables are measurable on interval scales. This assumption is necessary since path analysis uses recursive regression models (43, p. 117).
4. Each criterion variable, or any dependent variable is assumed to be determined or explained by some combination of variables in the diagram. However, where complete determination does not hold, a residual variable uncorrelated with other variables must be introduced. These residual variables are assumed by definition to be uncorrelated with any of the immediate predictors of the dependent variable (26).

Under the utilization of path analysis in this study, any feedback (direct or indirect) will not be considered as appropriate to the model conceptualization. Therefore, this study excludes non-recursive systems where instant reciprocal action might be considered.

A more detailed statistical analysis of the path technique will be discussed in Chapter III of this study.

#### H. Summary

Much of the previous literature served as a foundation for the presentation of techniques in Chapter 3. In brief, the following points appear to emerge from the literature:

1. Adaptation of explanatory theoretical conceptualizations have allowed educational studies to be inclusive of causal theory.
2. Causation evolved from a natural science background using a deterministic approach.
3. The theory of concomitant variation is the basis for this study.
4. What the entire causation movement did was to better understand the universe.
5. The focal point of the theory of this study is the interventionist theory.
6. DeFotis (24) concluded that the most important facet of technical writing was the "scientific and technical information" to be presented.
7. Shaner (55) concluded attitudes were important in journalism prediction.

8. Various regression solution methods are frequently employed to develop sound methodological studies.
9. Prior to 1965 path analysis was used very little.
10. Adaptation of causal theory to include explanatory theory prompted the use of path analysis in social science research.
11. Path analysis is a technique for analyzing and quantifying the relative strengths of variables.

### III. METHOD OF PROCEDURE

This study was conducted with several basic purposes in mind. Initially, the investigation attempted to establish a predictor-criterion relationship between the introductory tests (MSAT, EPT, and HSR) and first quarter grade point at Iowa State University. This yielded the appropriateness of the data for conceptual model building in predicting college success for journalism students.

Second, the study attempted to distinguish what inferences could be made about student success in journalism at Iowa State University relevant to the student's prior record and achievement in other college academic areas.

Third, the inquiry attempted to ascertain whether statistical techniques were capable of yielding additional insights and inferences regarding student's academic success by utilizing and contrasting multi-variate techniques such as multiple regression and path analysis.

This chapter describes the methods and procedures that were used to gather and analyze the data necessary to conduct the study. The chapter has been subdivided into categories or parts as follows:

1. Selection of population and collection of the data
2. Criterion of student success in journalism at Iowa State University
3. Prediction variables
4. Basic assumptions
5. Hypothesis to be tested
6. Treatment of data



7. Variable-selection
8. Regressions
9. Path analysis
10. Summary

A. Selection of the Population and Collection of Data

It was decided to include in this study the 217 journalism graduates of Iowa State University between the years of 1965 to 1970. Through the assistance of Mr. James Schwartz, Head of Department, and Mr. Jerome Nelson of the Iowa State University Journalism Department, the records of academic achievement of the frame were gathered. With the assistance of the Department of Journalism and Mass Communication and the listings given by the Iowa State University course catalog (38), the college courses of the graduates were considered.

The population began with those students who were the Spring 1965 journalism graduates and continued until it encompassed the most recent graduates. A quarter by quarter division yielded the following:

<u>Quarter</u>	<u>Number</u>	<u>Quarter</u>	<u>Number</u>	<u>Quarter</u>	<u>Number</u>
Spring 1965	18	Winter 1967	3	Fall 1968	9
Summer 1965	6	Spring 1967	22	Winter 1969	7
Fall 1965	6	Summer 1967	8	Spring 1969	26
Winter 1966	7	Fall 1967	6	Summer 1969	13
Spring 1966	16	Winter 1968	7	Fall 1969	9
Summer 1966	5	Spring 1968	21	Winter 1970	7
Fall 1966	11	Summer 1968	10		

The entire frame or population would be utilized for further analysis. Of the 217 graduates, the records of 215 students were accessible and found. In accord with many regression studies (50, 54), the population of 215 was considered adequate for analysis.

In addition to the data gathered from the student record transcripts, additional data on each journalism graduate were then obtained from the Iowa State University Testing Service. The data were then analyzed and placed on code sheets.

Utilization of the computer center services at Iowa State University was responsible for punching, verification, and computational procedures. The coded data were recorded on code sheets and forwarded to the Iowa State University Computation Center. The coded information was then punched on International Business Machine (IBM) cards and computed on the 360/65 IBM computer. Using various correlation, multiple linear regression, and path analysis techniques, combinations of predictor variables were analyzed.

## B. Criterion of Success in Journalism at Iowa State University

One of the most important goals or objectives of journalism at Iowa State University is to increase a student's ability to write and speak clearly, concisely, and accurately. Completion of the required program in journalism, then, would indicate that the desired objectives had been met. Success in journalism at Iowa State University, therefore, can be assumed to be the completion of the prescribed course work and consequent graduation from Iowa State University Department of Journalism.

The criterion of journalism success for this study, then, was the average of a combination of the grades achieved in journalism classes during the student's college training. This was referred to as the criterion variable (Y). The grades received by the students at the end of each quarter were used as measures of student success.

The investigation adhered to the following literal and numerical equalities:

A = 400

B = 300

C = 200

D = 100

F = 000

## C. Prediction Variables

The following measures were designated as predictor variables and were used to attempt to predict and explain student success in journalism

at Iowa State University.

Two variables, English Placement Test (EPT) and Minnesota Scholastic Aptitude Test (MSAT) were recorded from the testing records of Iowa State University. A description of those two variables was listed as follows:

Minnesota Scholastic Aptitude Test (MSAT): This test was divided into the following areas: reading comprehension, vocabulary, and verbal analogies. The primary purpose was to appraise the student's scholastic aptitude or general intelligence with consideration given primarily to the requirements of most college curricula. Mathematics was not involved in this instrument. This variable will, hereafter, be referred to as MSAT and will be designated as the  $X_{12}$  variable.

English Placement Test: This instrument gave evaluation of a student's background in various phases of English such as grammatical usage, active vocabulary, and organization. In general a low score dictated some possibility of difficulty with English. English Placement Test was designated as the  $X_{13}$  variable.

The remainder of the variables were recorded from the college transcripts of the journalism graduates. All courses listed on the student's transcript, with the exception of those listed in the assumptions, were classified and placed under one of the seven academic headings:

1. Student's academic success in social sciences
2. Student's academic success in written and spoken English

3. Student's academic success in ~~humanities~~
4. Student's academic success in life science
5. Student's academic success in physical sciences
6. Student's academic success in mathematics
7. Student's academic success in journalism.

In addition, the student's first quarter grade point average was recorded from the transcript and called variable  $X_8$ . These categorizations into these headings were accomplished by utilizing the divisions proposed by the Iowa State University General Catalog (38). A description of the predictor variables follows:

Written English and Speech Achievement: Measured by the student's grade point average in this area, this variable gave an index to the success the ISU journalism graduate had in writing and speaking courses in college. This variable also gave an index as to the student's grammatical usage and general writing skill. This variable was called the  $X_1$  variable (ES).

Humanities Achievement: Measured by the student's grade point average in this area, this variable gave the measure of college academic success that the student had in the areas of history, literature, philosophy, or foreign language. Related conceptual disciplines were also categorized in this division. Student humanities achievement (HUM) was referred to as the  $X_2$  variable.

Social Science Achievement: This variable as measured by the cumulative college grade point average, indicated the student's academic success in college in the broad area called social science. Included were psychology, sociology, anthropology, economics, political

science, geography, industrial administration (except accounting) and other related courses in the social science core. The social science variable was called  $X_3$ .

Life Science Success: Measured by the student's grade point average in this area, this variable related the student's college success in the living sciences such as botany, zoology, biology, genetics, bacteriology and other courses dealing with similar concepts. This life science variable was called  $X_4$ .

Mathematics, Computer Science, and Statistics Success: Measured by the student's grade point average in this area, this indice gave a degree of measure of the student's college competence and success in mathematics, computer science, or statistics. Other similar courses, such as accounting were also included. Hereafter, the student's academic success in mathematics will be referred to as the  $X_5$  variable.

Physical Science Success: Measured by the student's grade point average in this area, this variable included the graduate's college achievement in disciplines such as: physics, chemistry, geology, metallurgy, biochemistry, and biophysics. Also, any related topical courses were listed under physical sciences. Student physical science success was listed as the  $X_6$  variable.

The following additional data were obtained from the transcripts:

High School Rank: Academic ranking denoted the process of ordering, by cumulative grade point average, the students in a given high school class. The highest rank was 1. To compute each student's high school rank, the individual's rank in the total population was

divided by the total number of students in the class. Iowa State University usually accepts only those students within the top half of their high school class. High school rank was referred to as the  $X_{11}$  variable.

#### D. Basic Assumptions

From the objectives of this study, the following assumptions were made:

1. The grades received in the student's course work at Iowa State University were a valid and satisfactory measure of achievement at the college level.
2. A linear relationship does exist between the criterion measure and the prediction variables.
3. The test scores the student received upon taking the MSAT and English Placement Test were satisfactory measures.
4. By definition, all of those individuals in the population of the study were assumed to be "successful" in journalism if they met the requirements for graduation in journalism at Iowa State University.
5. By definition, the relative success of a subject is measured by the student's grade point average as recorded on the student records.
6. The study assumed that divisions of college courses given by the various colleges in the college catalog: (a) written English and speech, (b) humanities, (c) social sciences, (d)

life sciences, (e) physical sciences, and (f) mathematics student achievement at Iowa State University are accurate and indicative of the six major categories used in the analysis.

7. The study assumed that all of the graduates in the study are journalism majors, and that there is a catalog description of courses, which will be used to assign "professional" courses to one of the six areas, (i.e. Rationale for course assignment. If catalog description is  $X_i$ , then course is assigned to appropriate  $Y_i$  category.)
8. The study assumed that courses in physical education, ROTC, performance music courses, seminars, and workshops, if omitted, would have no significant effect on the study.

#### E. Hypothesis to Be Tested

The purpose of stating the null hypotheses is to formulate the empirical problem into testable statistical terms. The null hypothesis was assumed for the purposes of statistical inferences. It was postulated in the following manner:

The Minnesota Scholastic Aptitude Test, high school rank, English Placement Test, written English success, physical science success, social sciences success, humanities success, life science success, and mathematics student achievement were of no significant value when used as predictors of journalism success for 1965 - 1970 journalism graduates at Iowa State University.



#### F. Treatment of Data

The information collected from the student's transcript and from the permanent records of the student was coded and placed on International Business Machine data processing cards. Results were then processed by the Computation Center at Iowa State University. Programs written by the statistical laboratory and adapted for the International Business Machine 360/65 computer were used for the analyses.

#### G. Variable Selection

Just as the criteria for all research designs insist, the researcher must be extremely critical of the variables that are selected for use in this study. Certain a priori assumptions, which coincide with the present theory and the researcher's common sense, determine the final variable selection. The "best" regression evaluation is not necessarily judged to yield the most useful set of variables for path analysis (23).

Variables are frequently selected for educational studies by the use of four criteria: (16)

1. The correlational matrix among all variables in the study is computed and the variables which correlate the highest with the dependent variable are noted.
2. The intercorrelation between the independent variables is also noted. However, if two variables have an extremely high

intercorrelation, it is possible that they are measuring similar characteristics.

3. The variables with the lowest standard error are considered.
4. The variables which, by virtue of the researcher's "theoretical basis and common sense," the researcher chooses to include.

Combinations of these four situations are one basis for variable selection. Much of the following review of the regression techniques is based on the work of Draper (25).

#### H. Regression Techniques

##### 1. Preliminary considerations for multiple regression analysis

Correlations, Pearson product moment coefficients, were computed for all predictor variables and interrelationships of criterion and predictor variables. The magnitude of these correlations was one criteria for selection of predictor variables. For the purposes of selection, a high correlation with the criterion (Y variable) was desired. However, while the predictor exhibited a high degree of relationship with the criterion, it was theoretically desirable for that variable to exhibit low intercorrelations with the other predictor variables.

Therefore, those variables which exhibit high correlation coefficients with criterion variable indicated a strong relationship. A low intercorrelation among predictor variables indicated that they were measuring different or separate conceptual areas or different attributes.

Since the multiple correlation,  $R$ , was a measure of the degree of association between the predictor and criterion variables, the greater the magnitude of the coefficient the closer the association. Regression equations with a corresponding low standard error of estimate were desirable in order to indicate the reduction of error in the prediction.

## 2. Backward solution

Essentially, this solution or technique is predicated on the assumption that the best set of predictor variables can be selected from a larger set of variables by using certain criteria for variable determination (25).

In the backward solution method, starting with the five variable prediction equation, one variable elimination was attempted. In this framework, the significance of loss in prediction due to the elimination was measured by the  $F$  test (62). This method of eliminating one variable at a time was continued until the dropping resulted in a loss of predictive capability. In this case, the process was repeated until a significant  $F$  indicated that further removal resulted in too great a predictive loss. From the variables that remained, the final backward prediction equation was formed. The actual computational procedure (25) is as follows:

1. A regression equation containing all variables is computed.
2. The partial  $F$  test is calculated for each variable. This variable is then treated as though it were the last to enter the regression equation.

3. The lowest partial F test value,  $F_L$ , is compared with a pre-selected significant level,  $F_O$ :

Two conditional situations arise from this statement:

- a. If  $F_L < F_O$ ; remove the variable which gave rise to  $F_L$ . Then recompute the regression equation with the remaining variables. Last, in this part of the process, re-enter stage 2.
- b. If  $F_L > F_O$ ; adopt the regression equation as it now stands and include all variables now entered.

In raw score form, the following regression equation with five variables was used to predict the criterion (journalism success):

$$Y = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_6X_6 + b_0$$

where

$X_7$  or  $Y$  = journalism success at Iowa State University

$X_1$  = English writing achievement

$X_2$  = humanities achievement

$X_3$  = social science achievement

$X_4$  = life science achievement

$X_6$  = physical science achievement.

$b_0, b_1, b_2, b_3, b_4,$  and  $b_6$  or the regression coefficients were obtained from the multiple regression program.

In order to evaluate the usefulness of each variable for predicting students' journalism success at Iowa State University, analysis of regression tables was computed for all regression combinations. Those values were entered and recorded in regression tables.

### 3. Forward solution

Whereas the backward solution method begins with the complete set or largest set of prediction variables and eliminates the number of variables until the final equation is derived, the forward solution regression method is predicated on the premise that there was a predetermined set or order of independent variables (25). Thus, by ordering variables by a set of predetermined criteria, those independent variables which were assessed as having the highest degree of "importance" were placed first in the regression equation. The order of insertion of variables is determined by utilizing the partial correlation coefficient as an evaluation of variables not yet entered in the equation. In this way, the last variable to enter the regression equation was judged to be the least important.

This solution (25) is initiated with the basic set of predictor-criterion variable relationships and proceeds to test each successive addition to that independent set. When an additional variable no longer accounts for significant reduction in the sum of squares attributed to residual factors, the process is terminated. The F level of significance can be set. The procedure is as follows (25):

1. The variable  $X_i$  which exhibited the greatest degree of relationship with Y (criterion) is selected.
2. The second step is to find the first-order, linear regression equation:  $\hat{Y} = f(X_i)$ .

3. Next, the partial correlation coefficient of  $X_j (j \neq 1)$  and  $Y$  (minus  $X_1$  effects) is computed.

The mathematical equivalency of this step is to find the correlation between (1) the residuals from the initial regression [ $Y = f(X_1)$ ] and (2) the residuals from the regression equation (not yet performed)  $\hat{X}_j = f_j(X_1)$ .

4. At this point the  $X_j$  with the highest partial correlation coefficient with the criterion ( $Y$ ) is chosen and a second regression equation  $Y = f(X_1, X_{j+1})$  is fitted.

5. The process is then repeated and the equation becomes:

$$\hat{Y} = f(X_1, X_2, \dots, X_k)$$

When these variables are entered, the correlations are between

(1) residuals from  $\hat{Y} = f(X_1, X_2, \dots, X_k)$  and

(2) the residuals from the equation  $\hat{X}_j = f_j(X_1, X_2, \dots, X_k)$

where  $(j > k)$ .

As each variable is entered into the regression equation, these calculations are done:

1.  $R^2$ , multiple correlation coefficient, indicating the percentage of explained variance.
2. The partial F test for the variable most recently entered indicates whether that variable has taken a significantly greater amount of the variance over the amount removed by previous variables in other regression analysis.

When the partial F value for the newly entered variable becomes nonsignificant, the process is terminated.

The hypotheses and subsequent steps are as follows (35):

Step 1: Test  $H_1: B_{n+1} = 0$

If this hypothesis is accepted (fail to reject), the process is terminated and the basic set of n variables is used to predict Y.

Step 2: Test  $H_2: B_{n+2} = 0$

If this hypothesis is accepted (fail to reject), the procedure is terminated and the variables  $X_1, X_2, X_3, \dots, X_{n+1}$  are used to predict Y.

Step 3: Test  $H_3: B_{n+3} = 0$

The basic test of the hypothesis  $H_i: B_{n+i} = 0$  is to reject if that ratio  $V > F_\alpha$  where  $F_\alpha$  is a specified significance level.

$$V = \frac{\hat{B}_i^2 / C_{ii}}{\text{RMS}}$$

where  $\hat{B}_i^2 / C_{ii}$  = attributable reduction in the regression sum of squares when  $i^{\text{th}}$  variable is removed from the regression equation. The estimator of  $B_i$  is  $\hat{B}_i$  and  $C_{ii}$  is the  $i^{\text{th}}$  diagonal element on the inverse matrix  $(X'X)^{-1}$ . RMS refers to the statistical quantity, residual mean square.

The generative regression technique, then, was as follows:

$$Y = b_1 X_1$$

which became:

$$Y = b_1 X_1 + b_2 X_2 + \dots + b_i X_i$$

where  $i = 1, \dots, n$  and  $n$  continues until the variable being added does not contribute significantly to the reduction in residual mean squares and  $b$  is equal to the corresponding regression coefficients from 1 to  $n$ .

#### 4. Stepwise regression

While the name, stepwise regression technique, does not imply a relationship, this procedure is statistically similar to the forward solution technique (25).

One basic difference is, however, that the variables included in the regression equation are re-examined at every stage. Assuming that differences exist between theoretical variable selection and empirical analysis (16), the stepwise procedure accounts for these predictor variable relationships. That is, a variable which may have been classified as the 'best' single variable predictor may, at a later stage, be dropped from the regression equation due to relationships between it and other variables now entered.

Draper and Smith (25) stated:

A variable which may have been the best single variable to enter at an early stage may, at a later stage, be superfluous because of the relationships between it and other variables now in the regression. To check on this, the partial F criterion for each variable in the regression at any stage of calculation is evaluated and compared with a preselected percentage point of the appropriate F distribution (25, p. 171).

This process provides a judgement on the contribution of each variable as though it had been the last variable to enter. In the event of a non-significant F ratio, the variable is removed from the equation.



The process is continued until no variables are entered in the equation and no further rejections are made.

The steps of the stepwise procedure are as follows (25):

Step 1: The procedure begins by examining the correlation matrix and enters into the equation the variable ( $X_i$ ) that is most highly correlated with the dependent variable ( $Y$ ).

Step 2: Using partial correlations, the next variable to enter is the variable exhibiting the second highest correlation coefficient with the criterion variable.

Step 3: With the regression equation  $\hat{Y} = f(X_i, X_j)$ , the method examines the contribution given by variable  $X_j$  if it had been entered first and variable  $X_i$  second. If the partial F is significant at the desired significance level, the variable  $X_j$  is retained. With variables ( $X_i, X_j$ ) already in the equation, the variable with the next highest correlation is selected.

Step 4: The regression equation now becomes  $\hat{Y} = f(X_i, X_j, X_k)$  and is determined by least squares. At this stage, partial F tests for  $X_i$  and  $X_j$  are made to determine if they should remain.

Hanson (35) gives the termination procedures for stepwise regression as when:

1. there is no variable to enter and no variable to remove;
2. the procedure dictates that the same variable be entered and removed successively; this can be corrected by changing the F levels if the user so wishes;
3. the total number of steps executed reaches the maximum number

of steps specified by the user.

Operationally, however, the procedure appears in two basic phases (35).

Phase 1: Enter a variable  $X_i$  into the regression framework if it satisfied

$$r_{p+i}^2 \geq r_{p+j}^2$$

where  $j = 1, 2, \dots, K-p$  and  $(t_c^2)_{p+i} = \frac{(n-p-2)r_{p+i}^2}{(k-r^2)_{p+i}} > F_{\alpha, n}$

where  $p$  = number of variables in regression and  $F_{\alpha, n}$  is the significance value chosen.

At this point the termination checks are employed. In the event none of the criteria are satisfied, continue.

Phase 2: Remove the variable  $X_i$  if that variable

$$\frac{(n-p-2)r_i^2}{1-r_i^2} \leq \frac{(n-p-2)r_j^2}{1-r_j^2}$$

where  $j = 1, 2, \dots, p$  and  $(t_c^2)_i < F$  remove.  $F$  remove is the set significance level for variable removable. Now, the termination procedures are checked. If one of the criteria is satisfied, termination is done. If these criteria are not satisfied, the process is continued and enters Phase 1 again.

## I. Path Analysis

### 1. Theoretical concern for the path models of the study

The actual computational procedures for this technique will be discussed later in this chapter. However, conceptually, the framework

and subsequent model building is important.

Research, even in its most theoretically conceived designs, is merely a snapshot on a sample of the actual happening (43). The conceptual framework relates to the research as capturing a bit of time and portraying those replications and implications of that moment.

The difficulty, then, in evaluating these time frequency snapshots, in terms of simultaneous multivariate equations and designs, is evident. The supposition that an "instant and all" relationships occurs at this time is not feasible (43).

Therefore, path analysis is one statistical attempt to overcome this deficiency. They do, however, have some similarities. Listed, they are:

1. Path analysis techniques determine simultaneous relationships between the criterion and predictor variables.
2. Path analysis uses conventional analysis for developmental and applied technique. But, path analysis is able to go beyond conventional statistical techniques by measuring the magnitudes of direct and indirect influences on the dependent variable. To illustrate, this technique evaluates not only the predictor variables on the criterion but also the predictors on predictors. To word that in causal or explanatory framework: The technique explains causes of the phenomena, and it reflects as to the causes of the causes (43).

The single stage path analysis technique or network analysis, as it is sometimes called in single stage models (14), pertains to a set

of variable relationships which are linear, additive, and causal and whose variable measurements are on an interval scale (41). The variables in this set may be called either exogenous or endogenous variables. Endogenous variables are those variables of the models which are determined or explained by a combination of variables that the theory suggests. The exogenous variables are those which are not explained. In other words, no other variables in the model appear to be causing or explaining these exogenous variables.

The purpose of this section of the chapter is to apply the path analysis technique to the present research study. The objectives will be to:

1. Identify the form of the paths of relationships among variables that exist.
2. Conceptualize these relationships in path models.

2. Path models, path coefficients, and path diagrams

The hypothesized relationship between a set of variables can be represented by a set of recursive (regression) equations. This is appropriate only if the relationships in the set are assumed to be unidirectional. For each endogenous variable a regression equation is written. This is cast with one variable as the dependent variable and the remainder of the variables as the independent variables.

Land (41) illustrated how a path coefficient ( $P_{ji}$ ) is equal to the least squares estimator of the standardized partial regression coefficient ( $b_{ij}^*$ ). Thus, by solving the appropriate set of regression

equations for the hypothesized model and standardizing the partial regression coefficients, an estimate of the path coefficient is reached.

The relationships that are statistically represented as a set of regression equations can be shown visually as a path model. The diagram for the single stage path or network model is illustrated in

Figure 1. The path diagrams are drawn according to the conventions established by Land (41). These conventions are as follows:

1. An hypothesized causal relationship is shown by a unidirectional arrow extending from a determining or predictor variable to a dependent variable.
2. An hypothesized noncausal relationship (correlation) between exogenous variables is shown by a two-headed curved arrow.
3. Each residual variable is related to its respective dependent variable by a unidirectional arrow. Literal subscripts,  $Z_j$ , are attached to the residual variables to indicate that they are unmeasured variables.
4. The numerical value of the path coefficient is entered beside the unidirectional arrow to which it corresponds. The value of the correlation coefficient may also be added to its corresponding two-headed curved arrow.

### 3. Preliminary steps to path analysis. . . .

a. Selecting the variable set      The first step in path analysis is to select the operational set of variables. Based on the theoretical considerations presented earlier, a set of ten variables (nine predictor

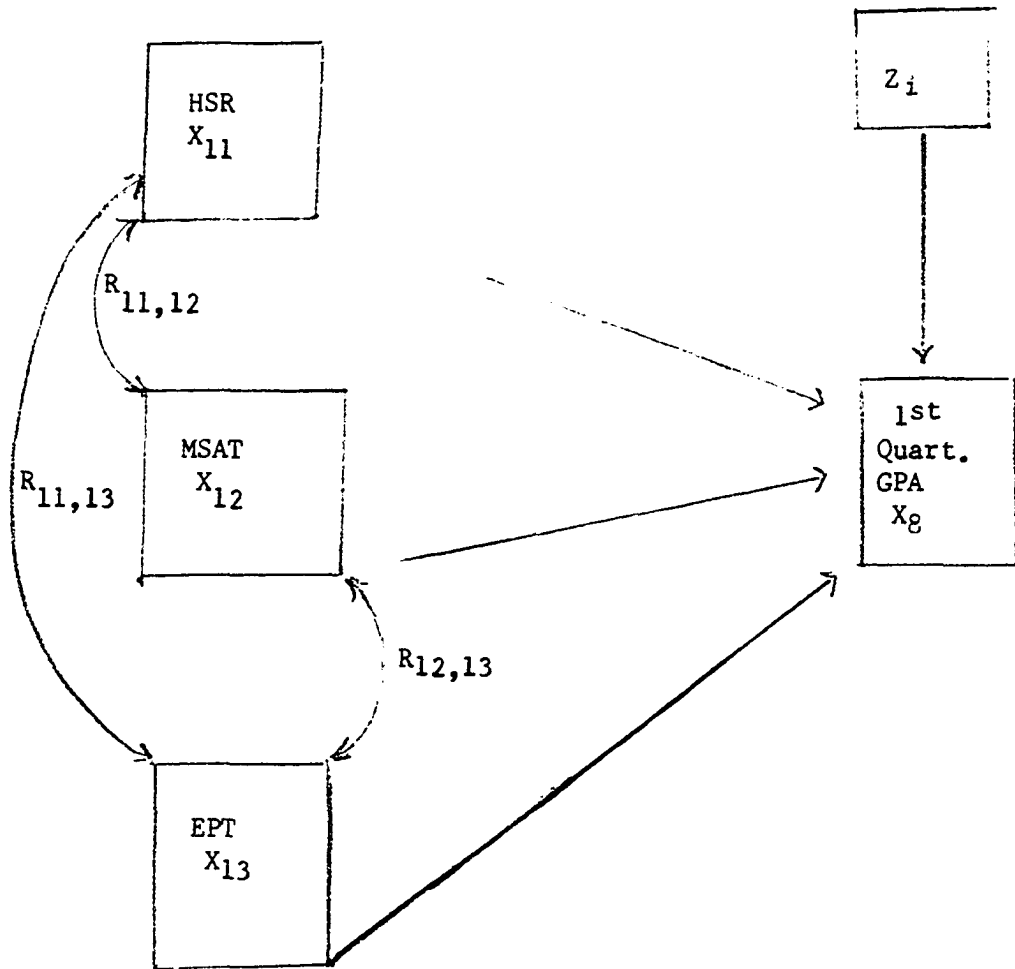


Figure 1. Path Model I (single stage path or network model)

or independent variables and one criterion variable) was selected. The selected predictor variables included the following: written English achievement, humanities achievement, social science achievement, life science achievement, mathematics achievement, physical science achievement, MSAT, HSR, and EPT. As mentioned previously, the dependent variable was student's journalism success.

One final consideration revolves on the merit of consideration of conceptual model building techniques as instrumental in achieving the objective of prediction. However, path analysis is concerned with achieving the objective of explaining of a set of variable relationships.

b. Assumption regarding the variables As was mentioned in Chapter II of the study, the path analysis technique is predicated on the assumption that the variables are linear and additive. The variables were examined to determine the existence of linear and curvilinear relationships with the criterion measure. However, all relationships were determined to be linear and transformations into linear forms were not employed. Labovitz (40) sets an argument in support for this view.

c. Variable ordering and model building with path analysis The next procedural phase in path analysis necessitates the explanatory ordering of the variables in the set. This ordering is necessary because of the assumption of unidirectional explanatory relationships and also because multistage models were somewhat made easier by the time dimension involved within the sets. For example, MSAT, high school

rank, and EPT are all measures of student behavior prior to official college entrance. Therefore, these predictors form an explanatory network or model to the variable (first quarter college grade point average). Then, if after viewing the predictive and explanatory values of these on student college academic success, the three became representative of the six extensions of that criterion measure. Likewise, since the six measures of college academic success are measures of student behavioral patterns, it is logical that they might explain a similar behavioral phenomena, student success in journalism.

The final ordering of the predictor and criterion variables is given in the two path models (I and II) represented in Figures 1 and 2.

Due to the problem of space and clarity, all correlations and the curvedline segments representing these lines were omitted from Path Model II.

#### 4. Listing the recursive equations

For each endogenous variable in the model a regression equation is written. The independent variables are those variables that were hypothesized in the model to be directly related to the variable under consideration. For example, the regression equation for variable  $X_2$  will include variables  $X_{11}$ ,  $X_{12}$ , and  $X_{13}$ . Following the procedure given by Land (41), the recursive equations for Path Model II are:

$$X_1 = b_{1,11.12,13} X_{11} + b_{1,12.11,13} X_{12} + b_{1,13.11,12} X_{13} + e_1 \quad (1)$$

$$X_2 = b_{2,11.12,13} X_{11} + b_{2,12.11,13} X_{12} + b_{2,13.11,12} X_{13} + e_2 \quad (2)$$

$$X_3 = b_{3,11.12,13} X_{11} + b_{3,12.11,13} X_{12} + b_{3,13.11,12} X_{13} + e_3 \quad (3)$$



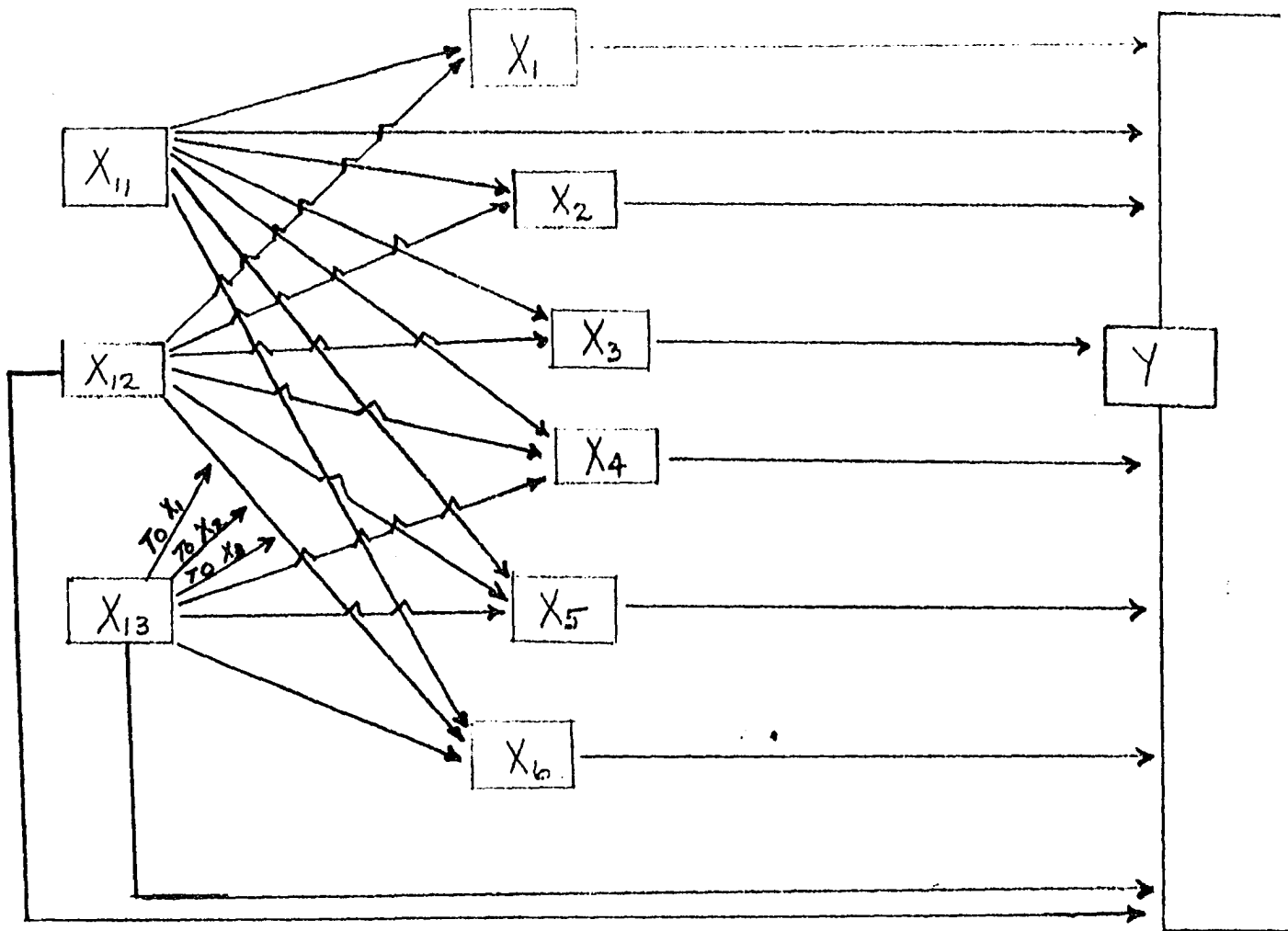


Figure 2. Path Model II (multistage model)

$$X_4 = b_{4,11,12,13} X_{11} + b_{4,12,11,13} X_{12} + b_{4,13,11,12} X_{13} + e_4 \quad (4)$$

$$X_5 = b_{5,11,12,13} X_{11} + b_{5,12,11,13} X_{12} + b_{5,13,11,12} X_{13} + e_5 \quad (5)$$

$$X_6 = b_{6,11,12,13} X_{11} + b_{6,12,11,13} X_{12} + b_{6,13,11,12} X_{13} + e_6 \quad (6)$$

$$\begin{aligned} X_7 = & b_{7,11,123456,12,13} X_{11} + b_{7,12,123456,11,13} X_{12} + \\ & b_{7,13,123456,11,12} X_{13} + b_{71,23456,11,12,13} X_1 + \\ & b_{72,13456,11,12,13} X_2 + b_{73,12456,11,12,13} X_3 + \\ & b_{74,12356,11,12,13} X_4 + b_{75,12346,11,12,13} X_5 + \\ & b_{76,12345,11,12,13} X_6 + e_7 \end{aligned} \quad (7)$$

##### 5. Determining path coefficient values

The initial step in the calculation of the path coefficients is to solve for the value of the partial regression equations in the recursive set of equations. After calculating these values a "t" test for significance was applied to each appropriate beta regression coefficient. The coefficients which exhibited a calculated "t" value less than the accepted tabular "t" level were removed from the model. After the removal of the direct arrows or paths of the non-significant regression coefficients, the equation was recalculated. This process was repeated until all remaining regression coefficients were significant at the .05 level.

However, just as the literal symbols are extracted from the recursive equation, the lines representing these relationships are removed.

The next step was the conversion of the partial regression coefficients into standardized regression coefficients, or path coefficients, and print these values on the appropriate lines of the diagram. Transfer

of the regression coefficients ( $b_i$ 's) into standardized path coefficients is accomplished by standardizing the appropriate regression coefficient for each variable set.

The formula was:

$$b_{yx}^* = b_i \left[ \frac{\sum x_i^2}{\sum y^2} \right]^{1/2} .$$

#### 6. Determination of residual paths

Residual effects are entered into the model when the endogenous variable is not completely determined by the explanatory variables in the model. The residual path coefficient is interpreted as the proportion of unexplained variance not accounted for in the model. Land (41) claimed it was caused by variables outside the model under consideration. The residual path coefficient is measured by the expression:

$$P_{zi} = \sqrt{1 - R^2} \quad (41, p. 19).$$

#### 7. Quantifying indirect effects of the path model

The path method of investigation can show the direct strengths of the independent variables in the model or can differentiate as to the relative strengths of each variable. Moreover, unlike other multivariate techniques, the path method can yield the extremely important indirect effects of the independent variables. As stated earlier in this study, theoretically the independent variables should be totally unrelated and exhibit a low intercorrelation. However, in the empirical setting this is virtually impossible. Thus, the variable may work indirectly to cause or explain the dependent variable (16).

Therefore, an explicitly outlined procedure for determining the indirect effects of one variable or another in a path network has been outlined (41).

The total indirect effects of an independent variable on the criterion are indicated by the formula:

$$\text{Total indirect effects (TIE) of } X_i \text{ on } X_j = r_{ij} - P_{ji} .$$

Since the indirect effects of any independent variable may involve paths through more than one variable, the separate components of the indirect effects can be calculated by using a formula given by Land (41):

$$r_{ij} = \sum_K P_{ik} r_{jk}$$

where  $i$  and  $j$  denote two variables in the model and  $K$  includes all variables from which paths lead directly to the particular variable. The sum of all the indirect effects of an exogenous variable is equal to the total indirect effect of the variable.

### 8. Summary of path analysis

Summarizing then, path analysis and model building involve: (1) selecting and ordering variables, (2) testing relative strengths of direct relationships, (3) finalizing model, and (4) determining indirect effects.

### 3. Summary

This chapter served as a methods and procedural foundation for the data analyzation to be used later in this study. In brief, the following points appeared to emerge from the procedural processes:

1. The population of the study were 217 journalism graduates of Iowa State University between the years 1965 to 1970.
2. The criterion of students' journalism success was the grade point average achieved in journalism classes during the student's college training.
3. The following variables were designated as predictor variables: ES ( $X_1$ ), HUM( $X_2$ ), SS( $X_3$ ), LS( $X_4$ ), MATH( $X_5$ ), PS( $X_6$ ), HSR( $X_{11}$ ), MSAT( $X_{12}$ ), and EPT( $X_{13}$ ).
4. The null hypothesis was postulated in the following manner: The Minnesota Scholastic Aptitude Test, high school rank, English Placement Test, written English, humanities, life science, mathematics, and physical science student achievement, as measured by the grade point average, were of no significant value when used as a predictor of academic success for 1965-1970 journalism graduates of Iowa State University.
5. Variables are frequently selected by the use of four criteria.
6. Backward solution regression techniques are predicated upon the assumption that the "best" set of predictor variables can be selected from a larger set.
7. Forward solution regression techniques are predicated upon the premise that a predetermined set or order of independent variables is possible.
8. Stepwise regression is based on a similar assumption with the forward method. Like the forward solution, variables are

ordered previously. The basic difference is, however, that the variable included in the stepwise equation are re-examined at every stage.

9. Path analysis and model building involve: (1) selecting and ordering variables, (2) testing relative strengths of direct relationships, (3) finalizing model, and (4) determining indirect effects.

#### IV. FINDINGS

##### A. Chapter Contents

This chapter contains sections devoted to the backward, forward, and stepwise multiple regression analysis techniques to determine prediction equations for student academic success in journalism at Iowa State University. Also, it contains sections of multi-stage path analysis used to determine additional insights about the problem.

For educators the problem dealing with a multiple regression model involves, not just determining an estimate of parameters, but also to determine the variables to be included in the final prediction model. Also, the nature of relationships between these predictor variables and the criterion is considered. Therefore, given the objectives, the problem centers upon choosing the "best" regression model. The problem to be solved relates, then, to the elimination of variables that are not contributing to the explained variance,  $R^2$ , of the model. To accomplish this, three regression models were utilized: backward, forward, and stepwise. On the other hand, to methodologically determine the magnitude of direct and indirect relationships of the configuration of variables, path analysis techniques were employed.

##### B. Backward Solution Regression

The mean scores and standard deviations for all the independent variables and the dependent variable were recorded in Table 1.

From Table 1 the following information regarding the population of the study was concluded:

Table 1. Mean scores and standard deviations of the criterion and predictor variables

Criterion and variables	Mean score	Standard deviation
X <sub>1</sub> Written English	283.2837	57.8276
X <sub>2</sub> Humanities	266.5767	53.8846
X <sub>3</sub> Social Science	260.9302	56.1433
X <sub>4</sub> Life Sciences	246.9581	68.0459
X <sub>5</sub> Mathematics	210.1302	68.7519
X <sub>6</sub> Physical Sciences	223.4047	69.5742
X <sub>8</sub> 1st Qtr. GPA	244.9907	66.8932
X <sub>11</sub> High School Rank	30.0837	51.3560
X <sub>12</sub> MSAT	57.5767	10.4364
X <sub>13</sub> EPT	58.0233	11.2650
Y Journalism Success	296.0930	44.2087

1) the mean high school rank for journalism graduates was 30.08 percent. The highest ranking percent was 1.

2) The mean performance of the journalism graduates on the Minnesota Scholastic Aptitude Test was 57.57.

3) The mean performance on the journalism graduates on the English Placement Test was 58.02.

With regard to student academic performance, information on Table 1 further illustrated that the journalism graduates had the highest grade point average in journalism (2.96), followed by the mean performance in



written English courses (2.83) and humanities grade point average (2.66). The lowest mean grade point recorded for the population of the study was mathematics success (2.10).

However, in order to initially establish the relationships between the criterion variable and the predictor set of variables a correlation matrix was computed. The results of this computation were recorded in Table 2. The relationships among the predictor variables (intercorrelations) were also listed and recorded in this tabular summary. Then, from considerations of the criteria in Chapter 3, the "best" five predictor variables were selected. The remaining four variables were not judged, by the criteria, to be of value in the prediction equation.

The hypothesis to be tested in this section was that written English ( $X_1$ ), humanities ( $X_2$ ), social science ( $X_3$ ), life science ( $X_4$ ), and physical science ( $X_6$ ) were of no significant value when used as predictors of journalism success of students at Iowa State University.

Upon inspection of Table 2, the variable with the highest correlation with the criterion was humanities success ( $X_2$ ) with a correlation coefficient of .6590. Social science success ( $X_3$ ) was of the second highest magnitude and exhibited a correlation coefficient of .6486. High school rank was judged as the variable having the least amount of relationship (-.1611) with the criterion variable.

The summary analysis of multiple regression for the "best five variable selection" was listed in Table 3. A closer inspection of Table 3 revealed the summary of multiple regression for predicting students' success in journalism at Iowa State University. This table contained information obtained from the analysis of multiple regression for all

Table 2. Product-moment correlation coefficient matrix for predictor and criterion variables

Variables	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>8</sub>	X <sub>11</sub>	X <sub>12</sub>	X <sub>13</sub>	X <sub>7</sub>
X <sub>1</sub> Written English	1.00										
X <sub>2</sub> Humanities	.464	1.00									
X <sub>3</sub> Social Science	.477	.682	1.00								
X <sub>4</sub> Life Science	.384	.539	.576	1.00							
X <sub>5</sub> Mathematics	.387	.460	.389	.330	1.00						
X <sub>6</sub> Phys. Science	.373	.455	.479	.417	.363	1.00					
X <sub>8</sub> 1st Qtr GPA	.530	.518	.550	.475	.580	.440	1.00				
X <sub>11</sub> HSR	-.194	-.338	-.258	-.228	-.237	-.150	-.234	1.00			
X <sub>12</sub> MSAT	.419	.327	.421	.342	.266	.360	.392	-.113	1.00		
X <sub>13</sub> EPT	.348	.310	.300	.306	.301	.163	.277	-.240	.574	1.00	
Y Journalism GPA	.453	.659	.649	.530	.340	.465	.443	-.161	.384	.314	1.00

Table 3. Summary of analyses of multiple regression for all combinations of the predictor variables ( $X_1$ ,  $X_2$ ,  $X_3$ ,  $X_4$ , and  $X_6$ )

Variables used for predictions		Variables eliminated		F value sig. at		Ry	Ry <sup>2</sup>	St. error of estimate
No.	Var.	No.	Var.	1%	5%			
5	$X_1X_2X_3X_4X_6^a$	-	-	48.9241		.7343	.5393	30.3647
4	$X_1X_2X_3X_4$	1	$X_6$	59.4703		.7288	.5311	30.5586
4	$X_1X_2X_3X_6$	1	$X_4$	59.1172		.7278	.5297	30.6069
4	$X_1X_2X_4X_6$	1	$X_3$	54.4054		.7134	.5089	31.2741
4	$X_1X_3X_4X_6$	1	$X_2$	50.4257		.6999	.4899	31.8730
4	$X_2X_3X_4X_6^a$	1	$X_1$	60.0456		.7304	.5335	30.4804
3	$X_1X_2X_3$	2	$X_4X_6$	75.7643		.7201	.5186	30.8911
3	$X_1X_2X_4$	2	$X_3X_6$	68.7112		.7030	.4942	31.6648
3	$X_1X_2X_6$	2	$X_3X_4$	66.4357		.6970	.4858	31.9272
3	$X_1X_3X_4$	2	$X_2X_6$	63.7064		.6894	.4753	32.2506
3	$X_1X_3X_6$	2	$X_2X_4$	62.1222		.6848	.4690	32.4429
3	$X_1X_4X_6$	2	$X_2X_3$	46.1577		.6295	.3962	34.5946
3	$X_2X_3X_4^a$	2	$X_1X_6$	77.1805		.7233	.5232	30.7424
3	$X_2X_3X_6$	2	$X_1X_4$	76.9791		.7229	.5226	30.7634
3	$X_2X_4X_6$	2	$X_1X_3$	69.5656		.7052	.4973	31.568.0
3	$X_3X_4X_6$	2	$X_1X_2$	64.1468		.6907	.4770	32.1978
2	$X_1X_2$	3	$X_3X_4X_6$	90.9508		.6796	.4618	32.5853
2	$X_1X_3$	3	$X_2X_4X_6$	85.6926		.6686	.4970	33.0291
2	$X_1X_4$	3	$X_2X_3X_6$	57.9108		.5944	.3533	35.7187
2	$X_1X_6$	3	$X_2X_3X_4$	46.9132		.5539	.3068	36.9809
2	$X_2X_3^a$	3	$X_1X_4X_6$	109.5976		.7130	.5083	31.1443
2	$X_2X_4$	3	$X_1X_3X_6$	96.7010		.6907	.4771	32.1197
2	$X_2X_6$	3	$X_1X_3X_4$	93.5336		.6847	.4688	32.3737
2	$X_3X_4$	3	$X_1X_2X_6$	89.2136		.6760	.4570	32.7299
2	$X_3X_6$	3	$X_1X_2X_4$	87.2973		.6720	.4516	32.8918
2	$X_4X_6$	3	$X_1X_2X_3$	57.7825		.5940	.3528	35.7327
1	$X_1$	4	$X_2X_3X_4X_6$	54.8665		.4526	.2048	39.5144
1	$X_2^a$	4	$X_1X_3X_4X_6$	163.5488		.6590	.4343	33.3276
1	$X_3$	4	$X_1X_2X_4X_6$	154.6361		.6486	.4206	33.7292
1	$X_4$	4	$X_1X_2X_3X_6$	83.0126		.5296	.2804	37.5889
1	$X_6$	4	$X_1X_2X_3X_4$	58.8305		.4652	.2164	39.2253

<sup>a</sup>Indicates the best combination in each set.

possible combinations of the five selected predictor variables. Also included in the table were computed F values, standard errors of estimate, multiple correlation coefficients,  $R_y$ , and the multiple correlation coefficient squared  $R_y^2$ . Data in this table were then used to select the "best" possible combination of independent variables for predicting student success in Journalism at Iowa State University using five, four, three, two, and one variable combinations for the prediction equation.

#### 1. Five variable prediction equation

The regression equation used to predict the dependent variable with five variables ( $X_1, X_2, X_3, X_4, X_6$ ) was:

$$Y = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_6X_6 + b_0$$

where

$Y$  = student success in journalism at Iowa State University

$X_1$  = written English success

$X_2$  = humanities success

$X_3$  = social science success

$X_4$  = life science success

$X_6$  = physical science success.

The regression coefficient values of the five variables were as follows:

$$b_1 = .0685$$

$$b_4 = .0817$$

$$b_2 = .2640$$

$$b_6 = .0681$$

$$b_3 = .2066$$

$$b_0 = 117.0019$$

From these values, the raw score prediction equation in symbolic mathematical form became:

$$Y = .0685 X_1 + .2640 X_2 + .2066 X_3 + .0817 X_4 + .0681 X_6 + 117.0019$$

The conclusions and results of the five variable multiple regression analysis were listed in Table 4. A multiple correlation squared  $R_y^2$  of .5393, a multiple correlation of .7343, and a standard error of 30.36 were recorded from the five variable analysis. The multiple correlation coefficient squared,  $R_y^2$ , indicated that the five variable model accounted for over 50 percent of the variance involved. The F ratio that was obtained, 48.92, was highly significant beyond the one percent significance level.

Table 4. Analysis of multiple regression using five of the variables ( $X_1, X_2, X_3, X_4, X_6$ )

Source of variation	d.f.	Sum of squares	MS	F
Regression	5	225543.281	45108.656	48.924**
Residual	209	192700.859	922.014	
Total	214	418244.140		
Standard Error = 30.36				
	$R_y (1,2,3,4,6) = .734$			.01 = 3.02
	$R_y^2 (1,2,3,4,6) = .539$	Table F <sub>5,209</sub>		.05 = 2.21

\*Denotes significance at the .05 level, here and throughout.

\*\*Denotes significance at the .01 level, here and throughout.

Therefore, the hypothesis that written English ( $X_1$ ), humanities ( $X_2$ ), social science ( $X_3$ ), life science ( $X_4$ ), and physical sciences ( $X_6$ ) were of no significant value in predicting success in journalism at Iowa State University was rejected. From the data, then, the results indicated that prediction of students' success in journalism at Iowa State University was possible.

At this point in the analyses, it was imperative to see if any of the five predictor variables could be eliminated from the regression equation without causing a significant loss in the predictive ability of the remainder of variables.

Referring to the summary of regression analyses (Table 3), the best four variable combination was sought. In this fashion, the loss of a variable could then be observed by comparing the five and four variable regression analysis results.

## 2. Four variable prediction equation

Inspection of Table 3, the summary table, revealed the best combination of four predictor variables. They were:

1.  $X_2$  = humanities success
2.  $X_3$  = social science success
3.  $X_4$  = life science success
4.  $X_6$  = physical science success.

The backward solution multiple regression analysis for this particular four variable set was computed and recorded in Table 5. As in the five variable case, a highly significant value of F, (60.0456),

Table 5. Analysis of multiple regression using the four predictor variables ( $X_2, X_3, X_4, X_6$ )

	d.f.	SS	MS	F
Regression	4	223142.686	55785.671	60.046**
Residual	210	195101.454	929.054	
Total	214	418244.140		

.01 = 3.32

Table  $F_{4,210}$

Standard error = 30.480

$R_y(2,3,4, \text{ and } 6) = .730$

$R_y^2(2,3,4, \text{ and } 6) = .534$

.05 = 2.37

was found. This reinforced the earlier conclusion that combinations of these variables (in this case  $X_2, X_3, X_4, X_6$ ) could be used for prediction of the criterion variable. The following regression coefficients for the four variable set were computed:

$$b_2 = .2788$$

$$b_3 = .2224$$

$$b_4 = .0869$$

$$b_6 = .0759$$

$$b_0 = 125.3278$$

Thus, in raw score form, the prediction equation for four variables became:

$$Y = .2788 X_2 + .2224 X_3 + .0869 X_4 + .0759 X_6 + 125.3278$$

Considering the loss of the written English variable, a test to determine if a significant decline in predictive ability had occurred, was then

computed. This tested the hypothesis that written English ( $X_1$ ) could be dropped from the backward prediction equation without a significant loss of predictive ability. Results of that test were printed in Table 6.

Table 6. Test for loss in predictive ability due to the elimination of the English writing ability variable ( $X_1$ ) from the five variable regression equation

Source of variation	d.f.	Sum of squares	MS	F
Regression (5 var.)	5	225543.281		
Regression (4 var.)	4	223142.686		
Loss ( $X_1$ )	1	2400.595	2400.595	2.6036
Residual (5 var.)	209	192700.859	922.014	
Total	214	418244.140		

.01 = 6.63

Tabled values

$F_{1,209} =$

.05 = 3.84

The F value of 2.6036, calculated in Table 6, was not significant. This indicated that the loss of the written English variable ( $X_1$ ) could be dropped from the prediction equation without a significant loss of predictive ability in predicting students' journalism success at Iowa State University.

The analysis then proceeded to the best three variable set, as determined by the summary of multiple regression, Table 3. The results of that three variable analysis were summarized in Table 7.



Table 7. Analysis of multiple regression using the three predictor variables ( $X_2$ ,  $X_3$ ,  $X_4$ )

Source of variation	d.f.	Sum of squares	MS	F
Regression	3	218828.961	72942.987	77.181**
Residual	211	199415.179	945.0956	
Total	214	418244.140		

Standard Error = 30.742

$R_y(2,3,4) = .723$   
 $R_y^2 = .523$

Table F<sub>3,211</sub>  
 .01 = 3.78  
 .05 = 2.60

### 3. Three variable prediction equation

The best three variable prediction, as dictated by inspection of data in Table 3, the summary table, included the humanities variable ( $X_2$ ), the social science variable ( $X_3$ ), and the life science variable ( $X_4$ ). Using only these three predictor variables in combinations, a prediction of journalism success for Iowa State University could be attempted.

From the multiple regression analysis the following regression coefficients were obtained:

$$b_2 = .2978$$

$$b_3 = .2459$$

$$b_4 = .1000$$

$$b_0 = 127.8637$$

In raw score form, the three variable backward regression equation became:

$$Y = .2978 X_2 + .2459 X_3 + .1000 X_4 + 127.8637$$

The three variable analysis of multiple regression, given in Table 7, yielded an F value of 77.1805. When compared to the tabled value ( $F_{3,211} = 3.78$  at the 1 percent level), the result is highly significant beyond that level. This F ratio again substantiated the prior assertion that students' journalism success could, indeed, be predicted using the three predictor variables ( $X_2, X_3, X_4$ ). The amount of total variance accounted for by these three variables ( $R_y^2$ ) was found to be .5232. The multiple correlation coefficient  $R_y$  was .7233.

At this stage, since one variable ( $X_1$ ) could be dropped without significant loss, a test to evaluate if the loss of the physical science variable ( $X_6$ ) in addition to the loss of written English ( $X_1$ ) would result in a significant reduction of predictive power was computed. The test for loss for two variables,  $X_1$  and  $X_6$ , was recorded in Table 8. The results of Table 8 indicated an F value of 3.641, and the hypothesis was rejected.

Significant at the five percent level, this F value indicated that the prediction process would be hampered and less efficient if the variable ( $X_6$ ) were removed. Therefore, the physical science variable ( $X_6$ ) was essential to the regression equation in order to predict journalism success at Iowa State University.

From the preceding data analyses, it was determined by the backward solution regression technique that the most accurate equation contained the humanities variable ( $X_2$ ), the social science variable ( $X_3$ ), the life science variable ( $X_4$ ), and the physical science variable ( $X_6$ ).

Table 8. Test for loss in predictive ability of the criterion due to the elimination of the written English ( $X_1$ ) and physical science variables ( $X_6$ ) from the five variable regression equation

Source of variation	d.f.	Sum of squares	MS	F
Regression (5 var.)	5	225543.281		
Regression (3 var.)	3	218828.961		
Loss ( $X_1, X_6$ )	2	6714.320	3357.160	3.641*
Residual (5 var.)	209	192700.859	922.014	
Total	214	418244.140	$F_{2,210}$	.01 = 4.61 .05 = 3.00

The final regression equation in raw score form, for the backward solution method was as follows:

$$Y = .2788 X_2 + .2224 X_3 + .0869X_4 + .0759 X_6 + 125.3278$$

### C. Forward Regression Solution

The hypothesis to be tested in this section was similar to the hypothesis listed in the backward solution.

It stated that no combination of these variables: ES ( $X_1$ ), HUM ( $X_2$ ), SS ( $X_3$ ), LS ( $X_4$ ), MATH ( $X_5$ ), PS ( $X_6$ ), HSR ( $X_{11}$ ), MSAT ( $X_{12}$ ), and EPT ( $X_{13}$ ) were of no significant value when used as predictors of students' academic success in journalism at Iowa State University.

The forward solution regression methods required a ranking, by partial correlation coefficients, of the predictor variables. The independent variables were ranked by their correlation with the criterion and these results were recorded in Table 9.

Close inspection of Table 9 indicated that the humanities success was the strongest variable (.6590) and that social science success followed with (.6486). Ranked third was the life science variable (.5296). The last two variables were EPT (.3138) and HSR (-.1611).

As mentioned in Chapter 3, by entering first the variable exhibiting the highest correlation with the criterion and proceeding until termination (as designed in Chapter 3) occurs, the regression coefficients listed in Table 10 were found. Table 10 listed the variables in the order of their entry into the equations and their respective regression equations. In the final computer run, termination occurred. Examination of the corresponding *t* values for these regression coefficients, and the variables not yet entered in the equation, yielded the variables that were judged to be non-significant at the .05 level of significance. From the data presented on Table 10, variable  $X_1$ , written English success (ES) was judged to be non-significant.

The forward regression equation, in raw score form, became:

$$Y = b_2X_2 + b_3X_3 + b_4X_4 + b_6X_6 + b_0$$

where

$Y$  = student success in journalism at Iowa State University

$X_2$  = humanities success

$X_3$  = social science success

Table 9. Rank ordering of predictor variables by size of correlation with criteria

Rank	Variable	Correlation $R_{xy}$
1	$X_2$ (HS)	.6590
2	$X_3$ (SS)	.6486
3	$X_4$ (LS)	.5296
4	$X_6$ (PS)	.4652
5	$X_1$ (ES)	.4526
6	$X_{12}$ (MSAT)	.3843
7	$X_5$ (MATH)	.3400
8	$X_{13}$ (EPT)	.3138
9	$X_{11}$ (HSR)	-.1611

Table 10. Forward solution: partial regression coefficients

Variables	Regression equations in order				
	1	2	3	4	5
$X_2$	.5407	.3323	.2978	.2787*	.2640*
$X_3$		.2930	.2458	.2223*	.2066*
$X_4$			.0999	.0869*	.0816*
$X_6$				.0759*	.0681*
$X_1$					.0685
$X_{12}$					
$X_5$					
$X_{13}$					
$X_{11}$					

$X_4$  = life science success

$X_6$  = physical science success

The regression coefficient values of the four variables were as follows:

$$b_2 = .2787$$

$$b_3 = .2223$$

$$b_4 = .0869$$

$$b_5 = .0759$$

$$b_0 = 117.002$$

From these values, the final forward prediction equation became:

$$Y = .2787 X_2 + .2223 X_3 + .0869 X_4 + .0759 X_6 + 117.002$$

Changes in other parameters of the regression equations, calculated in the forward solution, are shown in Table 11. In this table the changes in the cumulative size of the  $R^2$ ,  $\bar{R}^2$ , and the additional  $R^2$  for the last variable were entered.

The final result of the forward model is, by inspection of Table 11, a predictive model with four variables ( $X_2$ ,  $X_3$ ,  $X_4$ , and  $X_6$ ) which yield a multiple R squared of .5335. A total of five variables ( $X_1$ ,  $X_5$ ,  $X_{11}$ ,  $X_{12}$ , and  $X_{13}$ ) were excluded from the final prediction equation of forward regression analyses.

Thus, the hypothesis that no combination of variables: ES ( $X_1$ ), Hum ( $X_2$ ), SS ( $X_3$ ), LS ( $X_4$ ), Math ( $X_5$ ), PS ( $X_6$ ), HSR ( $X_{11}$ ), and EPT ( $X_{13}$ ) were of no significant value when used as predictors of students' academic success in journalism at I.S.U, is rejected. Therefore, it is

Table 11. Forward solution: changes in characteristics of the regression equations

Variable entered	Cumulative $R^2$	Additional $R^2$ for last variable	Cumulative $\bar{R}^2$
$X_2$	.434	.4343	.425
$X_3$	.508	.0740	.502
$X_4$	.523	.0149	.514
$X_6$	.533	.0303	.523

concluded that the predictor variables do form a measure of students' academic success in journalism at Iowa State University.

#### D. Stepwise Regression Solution

An alternative procedure for data analyzation is the stepwise regression technique. Like the forward technique, but unlike the backward technique, stepwise regression is a forward solution procedure where variables are added rather than eliminated from the analyses.

The objective, as was the case in the previous two, was to keep only those variables which contribute to the explained variation,  $R^2$ , of the model. Essentially, the same hypothesis will be tested for this technique as was tested in the forward solution technique.

The solution procedures discussed in Chapter III were applied to the study data. The procedure was to add variables until the partial F test of the last variable to be entered was less than 2.26, the table

value of F at .05 level of significance. Previous variables were also dropped from the equation if their corresponding F ratio was less than 2.26.

Utilizing this procedure, a stepwise solution of six variables was obtained. These variables, their order of entry, and partial regression coefficients were listed in Table 12.

Table 12. Stepwise solution: partial regression coefficients

Variables	Regression equations in order					
	1	2	3	4	5	6
$X_2$	.541	.332	.298	.279	.299	.291*
$X_3$		.293	.246	.222	.225	.218*
$X_4$			.099	.087	.090	.081*
$X_6$				.076	.074	.076*
$X_{11}$					.073	.084*
$X_{13}$						.340*

Viewing Table 12, the information yielded the stepwise regression equation, in raw score form, that follows:

$$Y = b_2X_2 + b_3X_3 + b_4X_4 + b_6X_6 + b_{11}X_{11} + b_{13}X_{13} + b_0$$

where

Y = student success in journalism at Iowa State University

$X_2$  = humanities success

$X_3$  = social science success



$X_4$  = life science success

$X_6$  = physical science success

$X_{11}$  = HSR

$X_{13}$  = EPT

The regression coefficient values of the six variables were as follows:

$$b_2 = .2909$$

$$b_3 = .2176$$

$$b_4 = .0811$$

$$b_6 = .0760$$

$$b_{11} = .0836$$

$$b_{13} = .3402$$

$$b_0 = 102.457$$

From these values, then, this final stepwise regression equation evolved:

$$Y = .2909X_2 + .2176X_3 + .0811X_4 + .0760X_6 + .0836X_{11} + .3402X_{13} + b_0$$

Changes in the parameters of the regression equations calculated in the backward solution were exhibited in Table 13. The final six variable prediction equation is, by viewing Table 13, a predictive model with six variables ( $X_2$ ,  $X_3$ ,  $X_4$ ,  $X_6$ ,  $X_{11}$ , and  $X_{13}$ ) which yielded a multiple R squared of .5463. A total of three variables ( $X_1$ ,  $X_5$ , and  $X_{12}$ ) were excluded from the final equation of stepwise regression analyses. In Table 13 the changes in the cumulative size of  $R^2$ ,  $\bar{R}^2$ , and the additional  $R^2$  for the last variable were entered.

Thus, the hypothesis that no combination of these variables:

Table 13. Stepwise solution: changes in characteristics of the regression equation

Variable entered	Cumulative $R^2$	Additional $R^2$ for last variable	Cumulative $\bar{R}^2$
$X_2$	.434	.434	.425
$X_3$	.508	.074	.502
$X_4$	.523	.015	.514
$X_6$	.534	.011	.523
$X_{11}$	.540	.006	.525
$X_{13}$	.546	.006	.528

E S ( $X_1$ ), HUM ( $X_2$ ), SS ( $X_3$ ), LS ( $X_4$ ), MATH ( $X_5$ ), PS ( $X_6$ ), HSR ( $X_{11}$ ), MSAT ( $X_{12}$ ), and EPT ( $X_{13}$ ) were of any significant value when used as predictors of student journalism success at Iowa State University was rejected.

Therefore, the six variable combination, listed previously, did form a predictive measure of the expressed criterion.

#### E. Single Stage Path or Network Analysis

In order to check the appropriateness of the hypothesized single-stage path model, (Model I, shown in Chapter III), it was desirable to examine the background of the population prior to their academic entrance at Iowa State University. In this framework, possible relationships to college achievement or success might be gained. Three records were available for examination:

1. High school rank ( $X_{11}$ )
2. MSAT ( $X_{12}$ )
3. EPT ( $X_{13}$ )

The model, at this point, involved the relationships which these variables might have on initial college success, journalism graduates' first quarter grade point. In other words, what magnitude of explanation could be assessed to initial college achievement by the use of these three variables.

The correlation matrix for the predictor and criterion variables was recorded in Table 2. Inspection of the correlation matrix indicated that MSAT had the highest correlation (.3920) with the criterion (GPA) followed by EPT with .2767 and HSR with .2343.

In the conceptual Model I, the computation of network coefficients was accomplished by the standardization of regression coefficients (14). In order to obtain those coefficients, the appropriate computer regression runs were accomplished.

Using the formula the computations of path coefficients were accomplished as follows (14):

$$P_{yx} = b_{yx}^* = [byx] \left[ \frac{(\sum X^2)}{(\sum y^2)} \right]^{\frac{1}{2}}$$

The summary of the network or standardized beta coefficients was presented in column 2 of Table 14.

Table 14. Summary table for single stage path analysis

Variable	Total effect $R_y$	Total direct effect $P_{ji}$	Total indirect effect $(R_y - P_{ji})$	Component indirect effects	Indirect effect not explained
$X_{11}$	.234	.193	.042	$X_{11}$ on $X_8$ through $X_{12} = .042$	0.00
$X_{12}$	.392*	.370*	.022	$X_{12}$ on $X_8$ through $X_{11} = .022$	0.00
$X_{13}$	.277	.000	.277*	$X_{13}$ on $X_8$ through $X_{11} = .046$  $X_{13}$ on $X_8$ through $X_{12} = \frac{.213}{.259}$	0.018

### 1. Computation of residual effects for single stage path model

The residual path effects, the amount not accounted for in the model, were solved by the use of the formula (41):

$$Z_i = \sqrt{1 - R^2}$$

Following this procedure the network residual was computed and added to the final single stage model, Diagram 3.

### 2. Interpretation of the network analysis

Essentially, the pertinent question was what relationships concerning introductory college achievement ( $X_8$ ) and subsequent later success ( $X_7$ ) can be explained by network analysis. Or, in other words, what

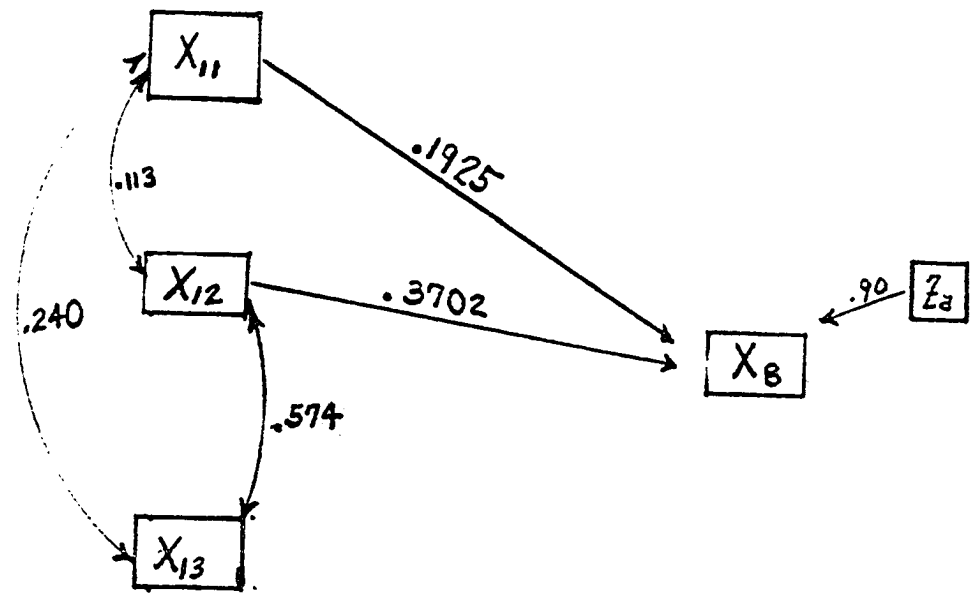


Figure 3. Final single stage path Model I

insights did network analysis offer?

In answer to the question two aspects will be considered. One aspect is called the theoretical dimension and the other is called the relative variable strengths.

### 3. Theoretical dimension

Based on the conceptual framework, the relationship of each of the variables  $X_{11}$ ,  $X_{12}$ , and  $X_{13}$  to  $X_8$  was examined using the network or standardized beta approach. On the basis of the empirical data, one variable ( $X_{13}$ ) lost its network path to the dependent variable ( $X_8$ ). Therefore, only the remaining two,  $X_{11}$  and  $X_{12}$ , were directly related to  $X_8$ .

#### F. Analyzing Relative Variable Strengths

In analyzing the final network model the variable which exhibited the strongest network path (.3702) was MSAT ( $X_{12}$ ). High school rank exhibits a direct path of .1925. Therefore, when considering direct relationships, MSAT was a much more important selection.

When considering the indirect effects listed in column 3 of Table 14, English Placement Test (.277) was ranked the highest. Thus, while MSAT ( $X_{12}$ ) had a strong direct effect on students' first quarter grade point average ( $X_8$ ), the English Placement Test ( $X_{13}$ ) exhibited the greatest magnitude of indirect strengths. Expressing this another way, the EPT, while not exhibiting a strong direct link, exhibited a strong indirect influence through the other predictor variables. The

total effects added another dimension. By summing the total indirect and direct effects, the grand total of network recorded in column 1 of Table 14 is calculated. MSAT (.392) would be selected as the most important variable.

Summarizing, then, the final model for the single stage path analysis would be the model illustrated in Figure 3. Viewing Diagram 3, the two variables (MSAT and HSR) have significant direct paths to college prediction of student success. While EPT does not exhibit a significant direct path, its indirect strengths are of the greatest magnitude.

#### G. Path Analysis

The analysis of educational data has been given new dimensions with the introduction of path analysis. This technique allows analysis of a complex network of direct and indirect configurations of variables and considering all possible relationships among the variables included in the set (17, p. 140).

The hypothesized multistage model was shown in Figure 2 of Chapter III. The hypothesized relationships, represented by the unidirectional arrows of the diagram, will be tested.

The purpose of this section of this chapter is to apply the path analysis technique to the research data. The analyses will differ because of the effects of predictor variables on other predictor variables will be considered in the indirect relationships of the model.

In order to test the hypothesized model the computational procedures listed in Chapter III will be followed.

#### H. Direct Effects

The first step in the computation of a path coefficient is to calculate the value of the partial regression coefficients.

At this point, it is wise to equate the fact that the number of computer runs will be governed by the actual number of unidirectional arrows in the hypothesized model (14, p. 11). A listing of the partial regression coefficients derived from the computer printouts is given in Table 15.

After the calculation of these regression coefficients ( $b_i$ ), a "t" test was applied to each coefficient. Those coefficients with a calculated t value lower than the table t value at .05 level, were eliminated and the equation was re-calculated without that variable (17, p. 149).

This procedure was repeated until all paths were significant. All these significant paths are recorded in Table 15. The arrows representing these significant regression coefficients are the straight line arrows remaining in Figure 4. The lines now deleted were those regression coefficients that were non-significant. In studying Diagram 4, only three direct routes to the dependent variable still exist ( $X_2$  to  $X_7$ ), ( $X_3$  to  $X_7$ ), and ( $X_4$  to  $X_7$ ). The remainder have been eliminated from consideration.



Table 15. Partial regression coefficients associated with the regression equations in the recursive set

Dependent	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>11</sub>	X <sub>12</sub>	X <sub>13</sub>
E.S (X <sub>1</sub> )							-.1422*	1.838*	.654
							-.1676*	6.5002*	
HUM (X <sub>2</sub> )							-.3002*	1.203*	.0512
							-.3202*	1.510*	
SS (X <sub>3</sub> )							-.2266*	2.0401*	.1613
							-.1489*	2.3151*	
LS (X <sub>4</sub> )							-.2271*	1.6741*	.7072
							-.2546*	2.0972*	
MATH (X <sub>5</sub> )							-.2391*	.9651	1.0604*
							-.1323		1.6433*
									1.7879*
Phy. Sci. (X <sub>6</sub> )							-.1725	2.6786*	-.6075
								2.3977*	
Journ. (X <sub>7</sub> )	.0587	.2897*	.2008*	.0781*			.0794	.1471	.2345
GPA		.2978*	.2458*	.0999*	-.0230	.0686			

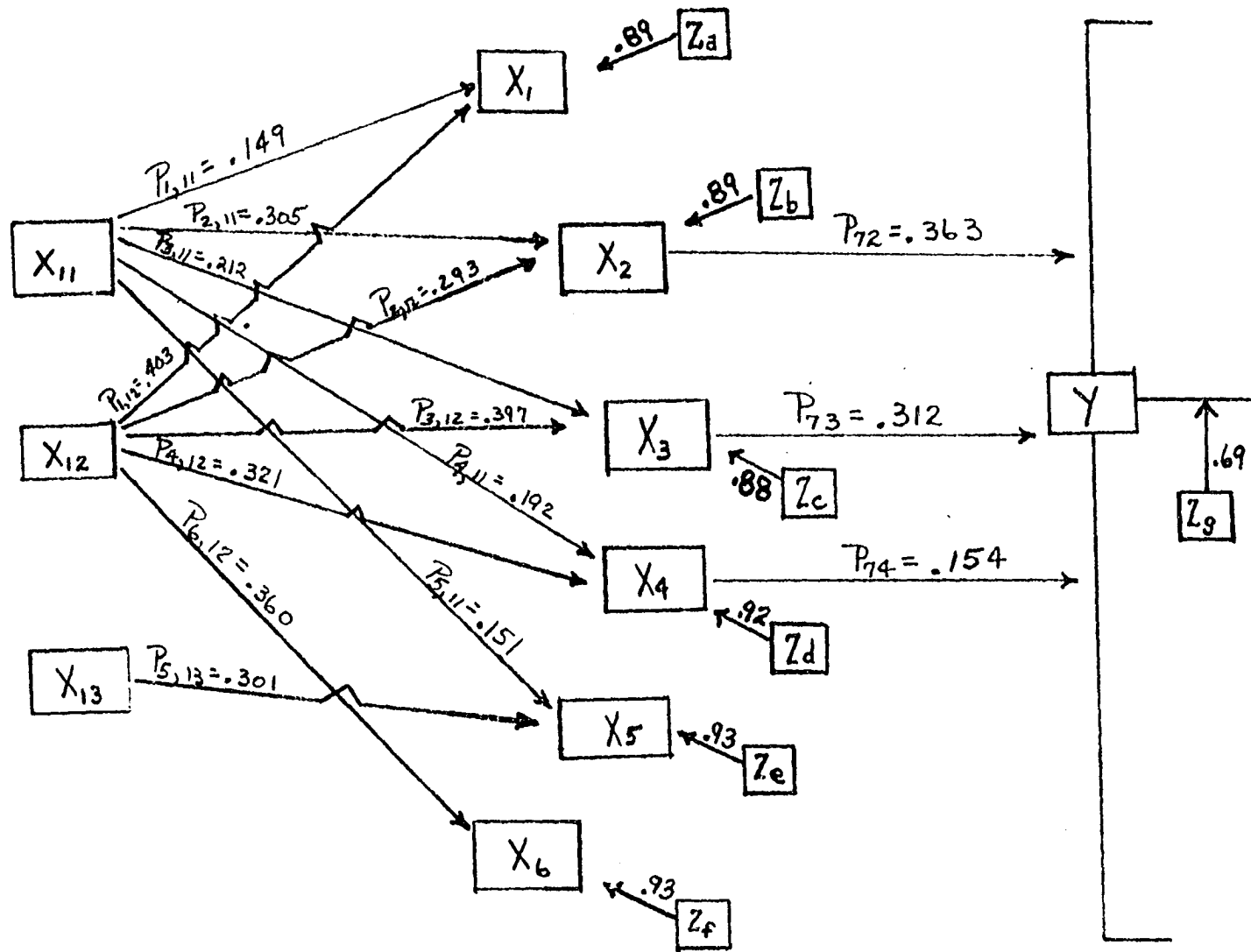


Figure 4. Final multistage path Model II

The next step, operationally, was to convert these significant regression weights, listed in Table 15, into standardized regression coefficients ( $b_{yx}^*$ ) and add them to the final path model in Figure 4. As stated in Chapter III, the formula for these standardized betas or path coefficients was:

$$P_{ji} = b_{yx}^* = b_i \left[ \frac{\sum x_i^2}{\sum y^2} \right]^{1/2}$$

The summary of the direct effects for Model II is recorded in the path summary table, Table 16. The path coefficients or direct effects are given in the second column.

#### I. Computation of Residuals for Multi-Stage Path Model

The next computation is the residual effect of the final multi-stage path model. As stated earlier in this chapter, the effects can be estimated by:  $\sqrt{1 - R^2}$ . Following this procedure, the residual path effects were computed and added to the model. The basis for these calculations are shown in Table 17.

##### 1. Indirect effects

The next calculations involved a procedure for determining the relative magnitudes of the indirect influences of the multi-stage path diagram.

The method of calculation is proposed by Land, (41, p. 10):

If the total effect of an exogenous variable  $Z_1$  on an endogenous variable  $Z_3$  is defined as the bivariate correlation of the two variables, and if the direct effect is estimated by  $P_{31}$  (the path coefficient), then the indirect effect must be estimated by  $r_{12}P_{32}$ , or in a more generally applicable form: Total Indirect Effect (TIE) of  $Z_1$  on  $Z_3 = r_{31} - P_{31}$ .

Table 16. Summary of multi-stage path analysis

Variable	Total effect $R_y$	TDE $P_{ji}$	TIE $(R_y - P_{ji})$	Component indirect effects	Indirect effect not explained
$X_1$	.4526	0.00	.4526	$X_1$ on $X_7$ thru $X_2 = .168$ $X_1$ on $X_7$ thru $X_3 = .149$ $X_1$ on $X_7$ thru $X_4 = .059$ <u>.376</u>	.077
$X_2$	.6590*	.363*	.2960	$X_2$ on $X_7$ thru $X_3 = .213$ $X_2$ on $X_7$ thru $X_4 = .083$ <u>.296</u>	.000
$X_3$	.6486	.312	.3370	$X_3$ on $X_7$ thru $X_2 = .248$ $X_3$ on $X_7$ thru $X_4 = .089$ <u>.337</u>	.000
$X_4$	.5296	.154	.3756	$X_4$ on $X_7$ thru $X_2 = .196$ $X_4$ on $X_7$ thru $X_3 = .180$ <u>.376</u>	.000
$X_5$	.3400	.000	.3400	$X_5$ on $X_7$ thru $X_2 = .167$ $X_5$ on $X_7$ thru $X_3 = .121$ $X_5$ on $X_7$ thru $X_4 = .051$ <u>.339</u>	.001
$X_6$	.4652	.000	.4652*	$X_6$ on $X_7$ thru $X_2 = .165$ $X_6$ on $X_7$ thru $X_3 = .149$ $X_6$ on $X_7$ thru $X_4 = .064$ <u>.378</u>	.047

Table 16. (Continued)

Variable	Total effect $R_y$	TDE $P_{ji}$	TIE $(R_y - P_{ji})$	Component indirect effects	Indirect effect not explained
$X_{11}$	.1611	0.00	.1611	$X_{11}$ on $X_7$ thru $X_2 = .111$ $X_{11}$ on $X_7$ thru $X_3 = .067$ $X_{11}$ on $X_7$ thru $X_4 = .030$ <u>.208</u>	.047
$X_{12}$	.3843	0.00	.3843	$X_{12}$ on $X_7$ thru $X_2 = .106$ $X_{12}$ on $X_7$ thru $X_3 = .124$ $X_{12}$ on $X_7$ thru $X_4 = .049$ <u>.279</u>	.105
$X_{13}$	.3138	0.00	.3143	$X_{13}$ on $X_7$ thru $X_4 = .010$ $X_{13}$ on $X_7$ thru $X_3 = .037$ $X_{13}$ on $X_7$ thru $X_2 = .050$ <u>.097</u>	.217

Table 17. Residual path coefficients

Endogenous variables	Regression equation no.	R <sup>2</sup>	Residual path coefficient	Estimate of residual path coefficient
X <sub>1</sub>	2	.20	Z <sub>a</sub>	.89
X <sub>2</sub>	3	.20	Z <sub>b</sub>	.89
X <sub>3</sub>	4	.22	Z <sub>c</sub>	.88
X <sub>4</sub>	5	.15	Z <sub>d</sub>	.92
X <sub>5</sub>	6	.12	Z <sub>e</sub>	.93
X <sub>6</sub>	7	.13	Z <sub>f</sub>	.93
Y	1	.52	Z <sub>g</sub>	.69

Using this statistical definition, the indirect effects were calculated and reported in the path analysis summary, Table 16.

## 2. Component indirect effects

The separate components were calculated by means of the formula listed in Chapter III and recorded in the summary table, Table 16.

### J. Interpretation of the Multi-Stage Model

The relationships illustrated in Figure 4 are viewed as a statistically satisfactory set of relationships.

Thus, academic success in the humanities has the strongest direct link or path (.363) with the criterion measure (journalism success). This path is followed closely by X<sub>3</sub> (social science success) with .312 and X<sub>4</sub>

(life science success) with .154.

In contrast, the remainder of the predictor variables do exhibit indirect influence on the criterion measure through one of the three direct variables. Thus, in examining Table 16 and Figure 4, it can be seen that the life science variable ( $X_6$ ) has the strongest indirect force. However, the English writing variable ( $X_1$ ) and the life science variable ( $X_4$ ) exhibit almost as great an influence on the criterion through other variables.

An important outlet for the indirect influences of the model is noted by  $X_2$ . In almost every case, the component indirect effects traveling through  $X_2$  have the greatest magnitude.

As indicated by Table 16, the paths in the model do not explain all the indirect effects associated with the variables. Therefore, as is the case with the variable  $X_1$ , the total indirect effect (TIE) is listed as .4526. However, component additions yielded a total of .376. Thus, as was stated, the amount of explained component is subtracted from the total indirect effect (.4526), leaving .077 unexplained.

The previous regression techniques of analyses indicated that a number of variables were related to the criterion measure. However, when considered together as the path technique emphasized, only three variables ( $X_2$ ,  $X_2$ , and  $X_4$ ) are directly related to the criterion (Y) measure.

## V. DISCUSSION

As initially stated, this dissertation was concerned with the prediction and identification of academic measures which would predict a students' success in journalism at Iowa State University. Secondly, the study attempted to compare and contrast multivariate statistical techniques and, in so doing, illustrate insights or inferences that were gained.

The general objectives of the study were to identify, theoretically and empirically, variables which were related to students' success in journalism. It was hoped that in the pursuit of these objectives the goals for the methodology could also be achieved.

Frequently, the implications of an investigation are enshrouded in the mathematical nomenclature of the statistical expressions. For this study, however, the implications were not quite so difficult to interpret.

This study contains the following points of discussion:

- (1) Comparison of the population of the study to general Iowa State University norms
- (2) Comparison of the three multiple regression techniques
- (3) Implications of the regression models
- (4) Relative importance of variables
- (5) Utility of multiple regression findings
- (6) Implications of the path analysis
- (7) Limitations of the study



(8) Recommendations for further study.

A. Comparison of the Population of the Study to General University Norms

One further concern to the study was some measure of the comparison of the population of the study to the general student body. Before any inferences were made, examination of the relationship of the population of the study to the general college entrance standards was considered. The mean performance levels for the journalism graduates of the study and the norms generated by the students who entered in 1968 were considered. A review of the records of the Iowa State University Testing Service indicated that the 1968 mean scores on the measures of HSR, EPT, and MSAT were representative of any one year. Therefore, the year 1968 was selected for the comparison. With regard to HSR, the 3,739 Iowa State University students had a mean of 22.49 and a standard deviation of 16.53. The journalism graduates ranked in the 30.08 percent, approximately one half standard deviation from the mean.

On MSAT, the journalism graduates were slightly over the norm: 57.57 to 50.59. The standard deviation was 12.54.

On the English Placement Test the journalists exceeded their all-college counterparts 58.02 to 52.66. A standard deviation of 10.49 was recorded.

Concluding, the population of the study was slightly higher on EPT and MSAT, but slightly below the mean on HSR.

Since the population of the study was, however, well within the bounds of one standard deviation, the population of journalism students

did constitute a normally distributed pattern of measure.

### B. Comparison of the Three Multiple Regression Techniques

Since a major goal or objective of this study is the identification of variables related to student success in journalism, a comparison is centered upon: (1) the actual variables included in each solution, and (2) the relative degree of importance of the variables that were included in the three solutions. These results are listed in Table 18.

Table 18. A comparison of backward, forward, and stepwise regression solutions

	Backward	Forward	Stepwise
Variables included	X <sub>2</sub> (HS) X <sub>3</sub> (SS) X <sub>4</sub> (LS) X <sub>6</sub> (PS)	X <sub>2</sub> (HS) X <sub>3</sub> (SS) X <sub>4</sub> (LS) X <sub>6</sub> (PS)	X <sub>2</sub> (HS) X <sub>3</sub> (SS) X <sub>4</sub> (LS) X <sub>6</sub> (PS) X <sub>11</sub> (HSR) X <sub>13</sub> (EPT)
Amount of explained variation, R <sup>2</sup>	.534	.534	.546
Percent of total variables entered	44	44	67

From the examination of Table 18 a total of six different variables were included in the three equations. Four variables (X<sub>2</sub>, X<sub>3</sub>, X<sub>4</sub>, and X<sub>6</sub>) were common to all three equations. In addition, the equations for forward and backward solutions were inclusive of the same variables HUM (X<sub>2</sub>),

SS ( $X_3$ ), LS ( $X_4$ ), and PS ( $X_6$ ).

From this, then, emerged four strong predictor variables. In the stepwise equation, however, high school rank ( $X_{11}$ ) and English Placement Test ( $X_{13}$ ) were added to the final prediction equation.

It was also interesting to note that the variable (EPT), which was included in the stepwise solution, constituted, to some degree, a measure of the students' ability in performing one set of the fundamental operations of a journalist, English skills.

Consideration must also be given to the predictive accuracy of the three final equations. As might be expected, the backward and forward solutions, since they employed the same predictor variables, had the same reliability figures. Both, as seen in Table 18, accounted for 53 percent of the variance. The stepwise equation raised the figure to 54 percent. Thus, if cost were no factor, the stepwise equation provided for a slightly elevated figure of reliability.

In terms of the variables entering the regression equations, 44 percent of the predictor variables were included in the backward and forward solutions whereas 67 percent were included in stepwise procedure.

### C. Implications of the Regression Models

After viewing the results of the three regression techniques, it would appear that a wide general academic background would enhance a student's chance of success in journalism. This may appear to be in sharp contrast to some present thinking on the matter. Theory might have

indicated that the written English skills would have been the most important variable. The regression analyses, however, contradicted this theoretical approach.

However, this general view regarding journalism may not be as divergent as it might seem. A wide knowledge of material should certainly be beneficial to the journalist. Moreover, these results appear to indicate that the philosophic and logical approach to journalist problems, eminent in the humanities, could be quite an important aspect of journalism.

The relative strengths of a variable could be compared when each predictor variable was considered as a single independent variable (see Table 3).

Using the regression, single variable criterion, the best single predictor was humanities success ( $X_2$ ).

#### D. Relative Importance of Variables

Using the regression coefficients obtained in the three regression analyses, the relative importance of the predictor variables was considered. A comparison of the standardized regression coefficients is presented in Table 19.

Inspection of Table 19, revealed that the most important variable was judged to be the humanities success ( $X_2$ ) with a standardized coefficient of .339 in the forward and backward techniques and a coefficient of .355 with the stepwise technique.

Table 19. Standardized beta coefficients for the three multiple regression techniques

Variables	Standardized beta coefficients		
	Backward	Forward	Stepwise
X <sub>1</sub>			
X <sub>2</sub>	.339*	.339*	.355*
X <sub>3</sub>	.282*	.282*	.276*
X <sub>4</sub>	.125*	.125*	.125*
X <sub>5</sub>			
X <sub>6</sub>	.118*	.118*	.120*
X <sub>11</sub>			.096*
X <sub>12</sub>			
X <sub>13</sub>			.087*

The second ranking standardized beta weight (.282) was associated with the social science variable. The third ranking variable was  $X_4$  (life science) with .125 followed by  $X_6$  (physical science) with .118. The standardized beta weights listed for stepwise were slightly higher in each case.

The conclusion, then, again substantiated that variable  $X_2$  (humanities success),  $X_3$  (social science),  $X_4$  (life science) and  $X_6$  (physical science) were the key variables.

#### E. Utility of the Regression Findings

Use of the final regression equations could be of future predictive value to the Iowa State Department of Journalism and Mass Communications.

From the comparison of the three regression prediction models, it could be concluded that the variables  $X_2$  (HUM),  $X_3$  (SS),  $X_4$  (LS) and  $X_6$  (PS) constituted a reliable set of predictor variables. Their use, then, in an empirical setting would be of value.

It must be remembered, at this point, that this study involved an ex post facto regression model. Therefore, an important assumption must be considered before future prediction utility would be feasible. The assumption, that this journalism population is representative of future journalism graduates, must be made.

With this assumption in mind, prediction for future groups of students could be made within the predictive confidence limits of the study ( $\alpha = .05$ ).

Along similar lines, individual prediction would involve utilizing the prediction equation and determining the values within the "best" confidence intervals that were desired by the user.

For example, one way to use the equation would be to formulate the normative measures for the population of this study. Hypothetically, in this way, a level of performance for the top journalism graduates could be ascertained. Similarly, the user of the prediction equation could establish the minimum standard necessary for completion of the program. These standards could be arbitrarily set by the user. Likewise, if desired, a mean standard of performance could also be obtained from the population of this study.

Using the three criteria (low, middle, and high points), utility could then be made to evaluate or guide incoming students. In this way, an evaluation stage for incoming or prospective journalists could be developed at the completion of a year's study. As a group or as individuals, the academic grade point averages for the predictors could be calculated. Consequently, an equivalent predicted score,  $\hat{Y}$ , would be given. Comparing this score with the normative population scales could give, within the confidence limits, an indication of success or failure in the journalism program of a student or group.

Besides being an introductory screening criteria, another function of the regression equations might be their use in evaluating students that were presently enrolled in the curriculum. In this way, the students' performance could be evaluated at the end of each year; and, again, a comparison with normative measures might yield some insights as to the

student's future chances.

Therefore, in summarizing this section, the regression equations could be used for: (1) evaluation of progress of students presently enrolled, and (2) determination of probability of student success for the journalism course of study.

#### F. Implications from Path Analysis

The results of the path analysis revealed several additional insights that were not apparent with the regression models.

Similar to the regression analyses, humanities success ( $X_2$ ) was considered as having the strongest direct path to the dependent variable. All three of the multiple regression methods gave the prediction equation as including variables  $X_2$ ,  $X_3$ ,  $X_4$ , and  $X_6$ .

However, the application of the path analysis technique illustrated that the direct path from variable  $X_6$  (physical science success) was dropped.

The indirect effects offered even more interesting inferences. The regression models indicated that some indication of performance on variables  $X_2$ ,  $X_3$ ,  $X_4$ , and  $X_6$  were necessary for prediction of student journalism success at Iowa State University.

Written English ( $X_1$ ), which theory might support as the top variable, became more important when considering indirect effects. In terms of explained indirect influences, it ranked as the second most important variable (.376). This suggested that additional consideration of



students' performance might be gained by viewing written English performance as a secondary indication of student performance. In other words, as might be expected, written English achievement appeared to have a positive effect on performance in other academic areas. Therefore, written English skills could give additional insights into student achievement in journalism.

Another variable which had relatively no predictive value in regression analyses was mathematic success ( $X_5$ ). Viewed in the indirect framework, it became the fourth ranking variable. Again, although this variable had not yielded strong direct strengths, it had strongly interacted indirectly with the other predictors.

Path analysis, then, was able to offer some interpretation of relationships which the regression techniques had not featured. Commonly applied, these indirect influences allow the practitioner additional insights upon which decisions may be based.

#### G. Limitations of the Study

Several problems evolved as a result of the investigation. One preliminary problem was related to transfer students. Some students transferred to Iowa State University after previously completing a number of the academic requirements elsewhere.

Including these measures would have necessitated making additional assumptions. Deletion of these records allowed the study to have comparable measurements but forced the study to subtract these courses from the appropriate academic measures.

Another limitation of the study involved the fact that the total population was not sufficiently large to appraise possible changing patterns from 1965 to 1970. If numbers had been larger, regression analyses of a time sequence design could have been accomplished.

One further point of discourse was concerned with the population including only those who had completed the program. More meaningful inferences might have been found with the inclusion of people who had failed to complete the program. Possibly a contrast of the graduates and those who experienced failure in the program might suggest further information relevant to the problem.

#### H. Recommendations for Further Study

Several possible follow-up studies could be developed using the present study as a foundation. One study would be to partition the criterion variable (student success in journalism) into two conceptual areas. One segment of the criterion could be a measure of the student's academic success in journalism courses where writing was the primary consideration. The second segment of the criterion or the second partition could become the student's performance in journalism courses where writing skills were not involved. In this way, it might be possible that more accurate insights regarding student performance in journalism might be assessed. The prediction equation for the written journalism might possibly include a different predictive frame than the non-writing criterion.

Another possible study would involve introducing additional independent variables that measure student performance in other areas. For

example, this study predicted behavior based on past behavior. Possibly a more inclusive view could be obtained by viewing additional variables representing certain specified behavioral skills or performance criteria. With these additions, the possibility of encompassing skills and traits, other than those in academic performance, might be gained.

One plan for further study could focus upon the comparison of this population with those students who failed to meet journalism graduation standards. These comparisons might provide further areas for research.

Another research possibility would involve analyzing the data of the study and making certain assumptions relevant to student's academic standing. In this way profile levels of the graduates would be possible. In addition, data on drop-outs could be compiled and normative patterns given. The comparison of drop-outs and graduates might prove fruitful in future journalism prediction studies.

Still another possibility for further study would involve the continuation of several of the suggestions. Using the partitioning of grades as to a (1) high, (2) medium and (3) low degree of success, residuals could be plotted for each score within those groups. In this way a plot of the predicted scores ( $\hat{Y}$ ) could be compared with the observed score ( $Y$ ). Insights into the components of the error variance might be gained by this residual plotting.

## VI. SUMMARY

The 1960's ushered in an ever increasing wave of journalism students to the college campuses of the nation. Iowa State University was typical of the national scene and announced sharply increased enrollments for the years 1967 - 1969. Therefore, for practical reasons, journalism departments found that more thorough research of the academic behaviors and traits of their graduates was necessary. Thus, the basis for this study was initiated.

There were two general purposes in this study. First, the investigation attempted to determine the influence of student achievement in university academic areas, previous achievement, and aptitude on student's academic success in journalism. Second, the investigation compared and contrasted insights regarding student academic success in journalism resulting from various multivariate statistical approaches.

The study included 217 journalism graduates of Iowa State University who completed the requirements for graduation during the years 1965-1970. Data were obtained from student transcripts and the records of the Iowa State Testing Service. All information was tabulated on International Business Machine Cards and processed by the Iowa State Computational Center. Programs written for the IBM 360/65 were used for the statistical analyses.

The criterion variable for the study was the student's journalism grade point average achieved during college. The variables that were used for prediction of student's journalism success were as follows:

1. written English success ( $X_1$ )
2. humanities success ( $X_2$ )
3. Social science success ( $X_3$ )
4. life science success ( $X_4$ )
5. mathematics success ( $X_5$ )
6. physical science success ( $X_6$ )
7. high school rank ( $X_{11}$ )
8. Minnesota Scholastic Aptitude Test ( $X_{12}$ )
9. English Placement Test ( $X_{13}$ )

Backward, forward, and stepwise solution were the three forms of multiple regression analyses applied to the data analyzation of the study. The objective of the regression analyses was to derive an equation which would yield a relatively high  $R^2$  value with as few variables as possible. Each of the variables was significantly related to the criterion or dependent variable.

The testable null hypothesis was as follows:

No combination of academic predictor variables can be used to predict students' academic success in journalism at Iowa State University as measured by grade point averages achieved by the journalism graduates from 1965 to 1970.

Analysis of the backward-solution, multiple regression technique showed that the best set of predictor variables included:  $X_2$  (humanities success),  $X_3$  (social science success),  $X_4$  (life science success), and  $X_6$

(physical science success). Using this four variable regression combination produced a multiple correlation coefficient squared,  $R^2$ , of .54.

Results of the forward solution multiple regression analysis were identical to the backward solution. Along with the four variables determined by forward and backward methods, the stepwise regression technique yielded an additional two predictor variables into the final equation. The variables were  $X_1$  (HSR) and  $X_{13}$  (EPT).

All three regression techniques explained a similar amount of variance,  $R^2$ : backward .53, forward .53, and stepwise solution .54.

The application of path analysis and conceptual model building yielded several additional insights into the determination of variables. First of all, the variable which exhibited the greatest degree of relationship with the criterion was variable ( $X_2$ ) humanities student success with a total effect of (.6590). The humanities variable, ( $X_2$ ), with a standardized beta of .363, also ranked as the most important variable of the model, as estimated by the magnitude of that direct path. The second ranking variable in path magnitude was  $X_3$  (social science success) which exhibited a direct path of .312.

In analyzing the indirect paths, several other variables were considered. The physical science variable ( $X_6$ ) recorded the highest component indirect effect (.378). The written English variable ( $X_1$ ), which was not present in any regression models, became an important indirect effect. A total component indirect effect of .376 was recorded for that variable.

Likewise, the life science variable ( $X_4$ ) exhibited the same component indirect effect (.376) as the written English. Moreover, the life science

variable was responsible for the third highest direct effect, .154.

From the results of this study, then, it was concluded that the best predictor variable for academic success of journalism students at Iowa State University was the humanities variable ( $X_2$ ).

However, the best combination of predictor variables used to predict the criterion were inclusive of four variables:  $X_2$  (humanities success),  $X_3$  (social science success),  $X_4$  (life science success), and  $X_6$  (physical science success).

## VII. LITERATURE CITED

1. Best, John W. Research in education. Englewood Cliffs, New Jersey, Prentice-Hall, Inc. 1959.
2. Blalock, Hubert M., Jr. Causal inferences, closed populations and measures of association. American Political Science Review 61: 130-136. 1967.
3. Blalock, Hubert M., Jr. Causal inferences in natural experiments: some complications in matching designs. Sociometry 30: 300-315. 1967.
4. Blalock, Hubert M., Jr. Causal inferences in on-experimental search. Chapel Hill, North Carolina, The University of North Carolina Press. c1961.
5. Blalock, Hubert M., Jr. Correlational analysis and causal inferences. American Anthropologist 62: 624-631. 1960.
6. Blalock, Hubert M., Jr. Correlation and causality; the multivariate case. Social Forces 39: 246-251. 1961.
7. Blalock, Hubert M., Jr. Evaluating the relative importance of variables. American Sociological Review 26: 866-874. 1961.
8. Blalock, Hubert M., Jr. Social Statistics. New York, New York, McGraw-Hill Book Co., Inc. 1960.
9. Blalock, Hubert M., Jr. and Blalock, A. B. Methodology in social research. New York, New York, McGraw-Hill. 1968.
10. Borg, Walter R. Educational research an introduction. New York, New York, David McKay Co., Inc. c1963.
11. Borgatta, Edgar F., ed. Sociological methodology 1969. San Francisco, California, Jossey-Bass, Inc. 1969.
12. Boudon, Raymond. A method of linear causal analysis: dependence analysis. American Sociological Review 30: 365-374. 1969.
13. Brodbeck, May. The nature and function of the philosophy of science. In Feigl, Herbert and Brodeck, May. Readings in the philosophy of science. Pp. 3-7. New York, New York, Appleton-Century-Crofts, Inc. c1953.
14. Bunge, Mario. Causality: the place of the causal principle in modern science. Cambridge, Massachusetts, The Harvard University Press. 1959.



15. Byerly, Richard L. Relationship between high school English background in linguistics and achievement in English 101 at Iowa State University. Unpublished M.S. thesis. Ames, Iowa, Library, Iowa State University. 1969.
16. Byerly, Richard L., Howe, Trevor G., and Warren, Richard. A case for network (path) analysis. Unpublished mimeographed paper presented at the American Educational Research Association, Minneapolis, Minnesota, March 1970. Ames, Iowa, Departments of Education and Statistics, Iowa State University. 1970.
17. Cohen, Morris R. Reason in social science. In Feigl, Herbert and Brodbeck, May. Readings in the philosophy of science. Pp. 663-673. New York, New York, Appleton-Century-Crofts, Inc. c1953.
18. Cohen, Morris R. and Nagel, Ernest. An introduction to logic and the scientific method. In Phillips, Bernard S. Social research strategy and tactics. P. 87. New York, New York, The Macmillan Co. 1966.
19. Cohen, Morris R. and Nagel, Ernest. The nature of a logical or mathematical system. In Feigl, Herbert and Brodbeck, May. Readings in the philosophy of science. Pp. 129-147. New York, Appleton-Century-Crofts, Inc. c1953.
20. Conant, James B. The American high school today. New York, New York, McGraw-Hill Book Co. 1959.
21. Coward, E. Walter, Jr. Factors related to different degrees of commercialization among farmers in Guatemala. Unpublished M.S. thesis. Ames, Iowa, Library, Iowa State University. 1966.
22. Coward, E. Walter, Jr. The subsistence to commercial transition in agricultural development. Unpublished Ph.D. dissertation. Ames, Iowa, Library, Iowa State University. 1969.
23. Coward, E. Walter, Jr., Bashor, Allan L., Warren, Richard D., and Klonglan, Gerald E. Social psychological aspects of market interactions in a subsistence economy: a case of path analysis. Unpublished mimeographed paper presented at a Rural Sociological Society Meeting, Boston, Massachusetts. August, 1968. Ames, Iowa, Department of Sociology and Anthropology, Iowa State University. 1968.
24. DeFotis, William. Technical writing and the education of the technical writer. Unpublished M.S. thesis. Evanston, Illinois, Library, Northwestern University. 1963.
25. Draper, N. R., and Smith, H. Applied regression analysis. New York, New York, John Wiley and Sons, Inc. 1966.

26. Duncan, Otis Dudley. Path analysis: sociological examples. *American Journal of Sociology* 72: 1-16. 1966.
27. Duncan, Otis Dudley and Blau, Peter. American occupational structure. In Lee, Motoko. A study of managerial behavior. Unpublished Ph.D. dissertation. P. 112. Ames, Iowa, Library, Iowa State University. 1969.
28. Durant, Will. The story of philosophy. Pp. 62-63. New York, New York, Washington Square Press, Inc. c1961.
29. Edwards, Selwyn Eugene. Predicting innovativeness of Iowa corn-hog-beef farmers through social psychological and demographic factors by means of path analysis. Unpublished M.S. thesis. Ames, Iowa, Library, Iowa State University. 1969.
30. Feigl, Herbert. The scientific outlook: naturalism and humanism. In Feigl, Herbert and Brodbeck, May. Readings in the philosophy of science. Pp. 408-418. New York, New York, Appleton-Century-Crofts, Inc. c1953.
31. Feinberg, Lilian Okner. A study of some selection and admission criteria for foreign students of journalism in United States colleges and universities. Unpublished M.S. thesis, Ames, Iowa, Library, Iowa State University. 1967.
32. Feuer, Lewis S. Causality in the social sciences. In Lerner, Daniel, ed. Cause and effect. Pp. 191-208. New York, New York, The Free Press. 1965.
33. Frank, Phillip. Modern science and its philosophy. In Lee, Motoko. A study of managerial behavior. Unpublished Ph.D. dissertation. P. 101. Ames, Iowa, Library, Iowa State University. 1969.
34. Gibbs, Gordon Lynn. Use of computer simulation to examine the validity of Getzel's and Guba's model in terms of its ability to predict administrative behavior. Unpublished Ph.D. dissertation. Ames, Iowa, Library, Iowa State University. 1968.
35. Hanson, Bonnie L. Mouflon: Reference Manual. Iowa State University, Statistical Laboratory, Statistical Numerical Analysis and Data Processing Series, No. 12. 1969.
36. Heise, David R. Problems in path analysis and causal inference. In Borgatta, Edgar F., ed. Sociological methodology. Pp. 3-37. San Francisco, California, Jossey-Boss, Inc. 1969.

37. Hemmerle, William J. Statistical computations on a digital computer. Waltham, Massachusetts, Blaisdell Publishing Co. c1967.
38. Iowa State University of Science and Technology general catalog 1967/1969. Iowa State University Bulletin 65, No. 15. Ames, Iowa, Registrar's Office, Iowa State University. 1967.
39. Kunhart, William E. and Olsen, Lionel R. An analysis of test scores and grades for predicting success of college students in English composition. Journal of Educational Research 53: 79-81. 1959.
40. Labovitz, Sanford. Some observations on measurement and statistics. Social Forces 46: 151-160. 1967.
41. Land, Kenneth C. Principles of path analysis. In Borgatta, Edgar F. Sociological methodology. Pp. 3-36. San Francisco, California, Jossey-Bass, Inc., Publishers. 1969.
42. Lauer, Alvah R. and Evans, J. E. The relative predictive value of different high school subjects on college grades. School and Society 31: 159-160. 1930.
43. Lee, Motako. A study of managerial behavior. Unpublished Ph.D. dissertation. Ames, Iowa, Library, Iowa State University. 1969.
44. Mueller, Charles W. Causal models and their use by sociologists. Monogram. Ames, Iowa, Department of Sociology and Statistics, Iowa State University. 1967.
45. MacBride, George Dexter. Prediction of first year achievement for Ames High School graduates at Iowa State University of Science and Technology. Unpublished M.S. thesis. Ames, Iowa, Library, Iowa State University. 1961.
46. McNally, Mary E. Relationship between achievement in communication skills and achievement in English 101 at Iowa State University. Unpublished M.S. thesis. Ames, Iowa, Library, Iowa State University. 1963.
47. Nagel, Ernest. Types of causal explanation in science. In Lerner, Daniel, ed. Cause and effect. Pp. 11-32. New York, New York, The Free Press. 1965.
48. Netusil, Anton Josef, Jr. Influences of Iowa school district characteristics on salary increases for administrations and teachers. Unpublished Ph.D. dissertation. Ames, Iowa, Library, Iowa State University. 1967.

49. Nunnally, Jum C. Psychometric theory. New York, New York, McGraw-Hill Book Co. c1967.
50. Ostle, Bernard. Statistics in research. Second edition. Ames, Iowa, Iowa State University Press. 1963.
51. Phillips, Bernard S. Social research strategy and tactics. New York, New York, The Macmillan Co. 1968.
52. Pophen, W. James. Educational statistics use and interpretation. New York, New York, Harper and Row, Publishers. 1967.
53. Price, Frank James. Journalism enrollments up 21.5%, at record high. Journal Quarterly 42: 712-715. 1965.
54. Sampford, M. R. An introduction to sampling theory with applications to agriculture. London, Oliver and Boyd. c1962.
55. Shaner, James Gerald. Scholastic journalism's influence in college. Unpublished M.S. Thesis. Columbia, Missouri, Library, University of Missouri. 1969.
56. Simon, Herbert A. Causal ordering and identifiability. In Lerner, Daniel, ed. Cause and effect. Pp. 157-190. New York, New York, The Free Press. 1965.
57. Simon, Herbert A. Models of man. New York, New York, John Wiley and Sons, Inc. 1957.
58. Spencer, Robert A. An investigation of the effect of incorrect grammar on attitude and comprehension of written English messages. Unpublished Ph.D. dissertation. East Lansing, Michigan, Library, Michigan State University. 1965.
59. Thorndike, Robert L. and Hagen, Elizabeth. Measurement and evaluation in psychology and education. Third edition. New York, New York, John Wiley and Sons, Inc. c1969.
60. Warren, Richard D. Field experiments in sociology methodological and statistical implications. Unpublished mimeographed paper presented at the annual meeting of the Rural Sociological Society, Miami Beach, Florida, August 1966. Ames, Iowa, Department of Sociology and Anthropology, Iowa State University. 1966.
61. Warren, Richard D., Klonglan, Gerald E., and Mueller, Charles W. Selected approaches to model building and testing in sociology--discussions and illustrations with emphasis on regression, partial correlation, and path analysis. Unpublished mimeographed working paper. Ames, Iowa, Department of Sociology and Anthropology, Iowa State University. 1968.

62. Wert, J. E., Neidt, C. O. and Ahmann, J. Stanley. Statistical methods in educational and psychological research. New York, New York, Appleton-Century-Crofts, Inc. 1954.
63. Wright, Sewall. Correlations and causation. *Journal of Agricultural Research* 20: 557-585. 1921.
64. Wright, Sewall. The method of path coefficients. *Annals of Mathematical Statistics* 5: 161-215. 1934.
65. Wright, Sewall. The theory of path coefficients. *Genetics* 8: 239-255. 1923.
66. Wright, Sewall. The treatment of reciprocal interaction with or without lag in path analysis. *Biometrika* 16: 423-445. 1960.
67. Yoesting, Dean R. and Bohlen, Joe M. A longitudinal study of occupational aspirations and attainments of Iowa young adults. Unpublished mimeographed paper. Ames, Iowa, Department of Sociology and Anthropology, Iowa State University. ca. 1968.

## VIII. ACKNOWLEDGEMENTS

The writing of a dissertation is the culmination of the formal educational process. There have been many participants who have contributed to the final product. I would like to thank these people.

First, my thanks go to Dr. Trevor Howe and Dr. Ross Engel for a large part in my educational success. Their help was instrumental in the conceptualizing, writing, and editing of this dissertation.

Thanks also to Mr. James Schwartz and Mr. Jerome Nelson for their contributions to this work. Largely, through their guidance, the data collection was accomplished. Dr. Ray Bryan has given generously of his time in the reading and discussing of the manuscript.

In addition to the above, I would particularly like to thank Dr. Richard Warren for his contributions in effectively organizing the sections of this dissertation. It was his teaching enthusiasm which generated this project.

Finally, although not least, I would like to acknowledge the efforts made by my wife, Elna. Without her, the dissertation, and all that it represents, would not have even begun.