ARAHOOD, Dale Arnold, 1933-
A CREDIT SCORING APPROACH TO THE
COMMERCIAL LENDING CREDIT DECISION
PROCESS.

The Ohio State University, Ph.D.,
1971
Statistics

University Microfilms, A XEROX Company, Ann Arbor, Michigan

© Copyright by
Dale Arnold Arahood
1971

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED
A CREDIT SCORING APPROACH TO THE
COMMERCIAL LENDING CREDIT DECISION PROCESS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the
Degree Doctor of Philosophy in the Graduate School of
The Ohio State University

By

Dale Arnold Arahood, B. A.

*****

The Ohio State University
1971

Approved by

[Signature]
Adviser
Department of Industrial Engineering
PLEASE NOTE:

Some Pages have indistinct print. Filmed as received.

UNIVERSITY MICROFILMS
ACKNOWLEDGMENTS

Bringing an undertaking such as this dissertation to a successful conclusion requires the support and encouragement of many people. I am indebted to Dr. William T. Morris for suggesting a bank as a laboratory for data collection. Dr. John B. Neuhardt contributed to the data collection, data analysis and statistical procedures. Dr. Walter C. Giffin deserves thanks for suggesting ideas regarding models.

The late Dr. David F. Baker was instrumental in my attending The Ohio State University. Working with him was truly an inspiration. I am grateful for the opportunity to have known him if only for a brief time.

Special recognition should go to Mr. Stuart C. MacIntire, vice president, and Charles F. Knight, president, of Lester B. Knight & Associates, Inc. for their continuing encouragement and support of this dissertation over the past several years. Thanks are due Mr. Richard E. Dooley, vice president, and Mr. Rudolph E. Palluck, senior vice president, both of the First National Bank of Chicago, for their assistance in granting access to a laboratory environment and for supporting this research effort.

Perhaps most important of all, to my wife, Ardis, I owe thanks for enduring my years of study with patience and understanding and for always providing encouragement and optimism when most needed.

Columbus, Ohio
July, 1971

Dale A. Arahood
VITA

August 11, 1933 .................. Born - Flanagan, Illinois

1955 ............................. B.A., Knox College, Galesburg, Illinois

1956–1959 ......................... U. S. Army

1959–1967 ......................... IBM Corporation

1967–1969 ......................... Teaching Associate, Department of Engineering Graphics, Department of Computer Sciences, The Ohio State University


FIELDS OF STUDY

Major Field: Industrial Engineering/Operations Research

Studies in Industrial Engineering. Professors David F. Baker and William T. Morris

Studies in Decision Theory. Professor William T. Morris


Studies in Management Science. Professor William T. Morris

Studies in Organization Theory. Professor Robert F. Miller

Studies in Queueing Theory. Professor Walter C. Giffin

Studies in Inventory Theory. Professor Walter C. Giffin

Studies in Statistics. Professor John B. Neuhardt

iii
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>VITA</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
</tbody>
</table>

## Chapter

### I. INTRODUCTION AND REVIEW OF THE LITERATURE

- Background of the Problem
- Research Objectives
- General Statement of Hypotheses
- Motivations for a Study of the Lending Decision
- Previous Analyses of Consumer Lending
- Previous Analyses of Commercial Lending
- Research on Credit Scoring Related to Lending
- Order of Presentation

### II. THE COMMERCIAL LENDING CREDIT DECISION

- The Present Multistage Decision Process
- The Credit Analysis Stage
- Management Loan Policy
- Commercial Lending Support Functions and Information Sources
- The Credit Evaluation Process
- Summary

### III. RESEARCH DESIGN

- Summary Description of the Variables
- Pilot Study: Stratified Sample of Loans and Turndowns
- Research Questions and Hypotheses
- Data Collection: The Pilot Study
- Data Collection: The Larger Sample
- Limitations of the Data
LIST OF TABLES

Table | Page
---|---
1 | Summary of Related Research to Credit Scoring of Loans | 18
2 | Classification of Cases by Two Variables in Order of Entry | 53
3 | Normal Minimax Cost Curve Values | 61
4 | Exponential Minimax Cost Curve Values | 62
5 | Summary of Normal and Exponential Minimax Cost Curves | 63
6 | Classification of Cases by Variables in Order of Entry | 97
7 | Discriminant Model Using Ratios | 102
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Multistage Organization Process</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>Histograms of Variables</td>
<td>43</td>
</tr>
<tr>
<td>3</td>
<td>Minimax Cost Curves - Normal Distributions</td>
<td>61</td>
</tr>
<tr>
<td>4</td>
<td>Exponential Minimax Cost Curves</td>
<td>63</td>
</tr>
<tr>
<td>5</td>
<td>Present and Proposed System</td>
<td>71</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION AND REVIEW OF THE LITERATURE

Background of the Problem

The making of commercial loans is the most crucial of banking functions, accounting for the major portion of income to banks and comprising most bank resources in terms of assets. This dissertation is a study of the reliability of methods used by lending officers in determining the creditworthiness of loan applicants. The research was carried out in a large metropolitan bank having approximately $7,000,000,000 in assets. A sample was drawn from a population of 5,000 loans with a total asset value of $4,000,000,000. These were primarily corporate loans of amounts greater than $25,000 and for various time periods exceeding 90 days.

Commercial lending is a complex and involved organizational process. Commercial lending officers tend to think of their work as a nonquantitative activity and rely primarily upon their experience and judgment (54). The objective of the commercial bank is to supply funds to as many applicants as qualify. The lending officer must enable the bank to achieve this objective by setting parameters of qualification for the loan applicant. He does this by arranging the terms and conditions of the loan for the mutual benefit of the bank and the customer.

In times of easier money, the lender fills an important marketing function by recruiting potential loan customers. Whatever the supply of potential customers, however, the loan officer must carefully evaluate the qualifications of the applicant and must arrange the terms and conditions of the loan to ensure that the loan will be mutually beneficial for the bank and the customer. The commercial lending officer has the dual objectives of maximizing the number of applications for loans and making sure that all recipients of
these loans are properly qualified. If the loan officer achieves his objectives fully, he will submit to the management of his bank a full quota of creditworthy applicants.

The loan officer must be familiar with the peculiarities of the customer's industry and past and present economic conditions. He must understand the influence of economic conditions on the availability and cost of money and on other aspects of the money market. He must know how to handle a financial arrangement for a customer which may include a "package" of various credit instruments such as a loan, sale of stock and/or sale of bonds. He must be ready to tap new sources of funds and find new ways of satisfying a customer's need for funds.

The banking industry stands to gain tremendous benefit from study and research leading to improvements in credit analysis, data collection and retrieval, and optimal loan portfolio management. Innovation in this area has been held back by the conservative nature of the banking business. This conservation is understandable in the light of the special responsibility a bank has for the funds of depositors, but this responsibility does not obviate the need to study and research new ways of handling these problems.

The commercial lending activity has not been the subject of extensive reporting by internal banking operations research departments. It is only recently that systematic research, using statistical techniques such as regression analysis and discriminant analysis, has been applied to the credit scoring problem in commercial loans (35). Consumer lending, on the other hand, has been studied quite thoroughly and discriminant analysis and regression analysis have been successfully applied (9, Chapter 6).

The discussion in this chapter will be devoted to a statement of the research objectives, a general statement of the hypotheses to be tested in the research, and some of the motivations for a study of the lending decision. Previous research in consumer lending is reviewed because it is a principal part of the background for subsequent developments in research on commercial lending. Finally, research on commercial lending is summarized, with special emphasis on work in credit scoring of commercial and small business loans.
Research Objectives

In this dissertation, determination of creditworthiness is viewed as a decision-making process requiring two types of evaluation by the lender:

1. Objective evaluation of quantitative information.
2. Subjective evaluation of qualitative factors.

The emphasis in this research is on the first of these types of evaluation, including the evaluation of policy and credit variables.

The primary goal of this study is to achieve a sufficiently thorough understanding of the credit evaluation process to permit the lender to make an accurate decision to lend or to turn down an application. If the lending decisions are to be made accurately, it will be necessary to:

1. Identify the variables actually involved in the decision process, together with their relative weights.
2. Achieve high reliability in predicting the lender's decision to lend or turn down an application.

To achieve the above-mentioned goal, this research tests several hypotheses about the credit analysis stage of the commercial lending decision process. These hypotheses are stated in the section entitled General Statement of Hypotheses on page 4.

In discussing the quantitative factors in the decision-making process, it must be pointed out that past emphasis in commercial credit analysis has been on studies of financial ratios and financial analysis. Evidence presented in this paper, however, indicates that most actual lending decisions may, in fact, be based on considerations other than financial ratios. It may be that ratios are frequently analyzed and prepared for reporting to management after the lender’s decision has been made. The analysis of the ratios, then, is of secondary importance in supporting the decision. Evidence is also presented in this dissertation to refute the allegation, that has been made so frequently in jest; that a person can obtain a loan only if it can be proven conclusively that he does not need the money.
This research sought to identify those ingredients of the evaluation process which had the greatest influence on lending decisions. To this end, a pilot study was conducted in which a large number of variables considered important to the credit decision were applied to a small sample of cases. This study showed, among other things, the central importance of variables other than financial ratios to credit analysis. In the second stage of the research, the most significant variables identified in the pilot study were applied to a large number of cases. The critical variables in credit analysis were divided into quantitative and qualitative sets with the emphasis on using the quantitative variables in a revised model of the decision process.

General Statement of Hypotheses

Past research on credit scoring of loans has been concerned mainly with whether a loan was "good" or "bad" on the basis of whether or not it was defaulted. This research has concentrated on the credit variables that affect the decision process. Instead of examining the outcome of loans "after the fact," this dissertation concerns the lending decision which is made after a loan application is processed. We shall examine the credit variables involved in the decision process, namely, balance sheet variables and financial ratios.

This research has been designed to test three hypotheses about the loan decision. The hypotheses have been designed to test the importance of credit variables, such as balance sheet variables, which do not have established policy guidelines, as contrasted with credit variables for which bank management has established policy guidelines. Those credit variables for which policy criteria have been established will be referred to in this dissertation as policy variables. Thus, policy variables are a subset of credit variables.

The first hypothesis to be tested is the following:

\[ H_1 \text{ Management policy variables are significant in classifying applications as loans or turn downs.} \]
As an example, the policy variable balances/amount is governed by a bank policy which states that an applicant must maintain a balances/amount ratio of 15% when a loan is outstanding and 10% when no loan is outstanding. This policy rule applies to the particular bank under study. (See Appendix A for a definition of the variable balances/amount.)

The second hypothesis follows:

\[ H_2 \text{ There are no significant differences in the key policy variables among industries (i.e., among lending divisions of the bank). } \]

This hypothesis is applied to test whether all industries may be grouped together and the variables used as discriminators which can be treated as if they belonged to one homogeneous distribution.

The third hypothesis is the following:

\[ H_3 \text{ Policy variables are more significant than credit variables in the lending decisions. } \]

This hypothesis is applied to test the relative importance of policy versus credit variables.

In summary, the first hypothesis is applied to test whether policy guidelines established by bank management can be used to predict the lending decision; the second hypothesis is applied to test whether the policy variables apply across industry lines; the third hypothesis is applied to test whether policy variables are more significant in predicting the lending decision than are ordinary credit variables.

Motivations for a Study of the Lending Decision

Of the many factors that influence the overall profitability of a bank, loan activities are clearly the most important. A calculation of the ratio of loan income to total operating income for the top ten banks in the United States in terms of assets as listed in *Fortune*
magazine, shows that an average of 70% of income comes directly from the loan assets created by the lending function. Income ratios of smaller banks are sufficiently similar that it seems valid to infer that loan income is also quite significant for smaller banks.

All of the major factors that influence management decision making are worthy of empirical analysis, and certainly those factors that make the heaviest contributions to profit are most worthy of all. Clarkson, in Managerial Economics (7, pages 40-41), proposes that, if we are to understand decision processes and use the knowledge to assist in decision making, much effort must be devoted to empirical research. This research, he suggests, will have to focus on decision-making processes of individuals and organizations and be accomplished in a reasonably systematic way.

Aside from the profit potential of commercial loans, their potential for incurring loss is certainly significant enough to motivate this study. Actual losses from default of commercial loans are a tightly guarded secret in most banks. The Robert Morris Associates has documented some of the reasons for loan default which indicate that analysis of financial data may be useful in predicting that default (44).

This research also provides an opportunity to improve upon the methodology of research in banking. Studies by Altman (2), Beaver (5), Ewert (16), and others examined financial data on publicly held companies and small businesses in order to find a set of financial variables and financial ratios which would predict failure of a given company. Most of the data was gathered after a company, in fact, failed. These studies are typical of the tendency to study "good" and "bad" loans and compare them after the bad loans are known to be bad.

Improvement in the quality of research concerning commercial loans can be achieved only if attention is shifted to new loan applications. Orgler (35), in discussing the results to be expected from credit scoring models and directions for future research, suggests that

_The lack of standard review systems in many banks and the time pressure on examiners of bank regulatory agencies are two important reasons why such a model (credit scoring model) is necessary. Further developments in the credit scoring area should be directed to the evaluation of new loan applications. Such models could be developed, however, only for small borrowers in specific industries._
**Previous Analyses of Consumer Lending**

The problem of determining credit risk is as old as banking. It has long been recognized that the ability to extend credit profitably requires judgment and experience based on past credit decisions. A number of attempts have been made to develop methods of strengthening the reliability of the lender's decisions. These attempts have met with the greatest success in the consumer lending area, where quantitative rating systems have been widely used.

The earliest study dealing with the development of numerical rating systems, or "credit scoring" as it is sometimes called, was that by Durand (14) in 1941. Durand employed discriminant analysis to develop weighting systems to analyze good and bad personal loan accounts in commercial banks. Later studies were carried out in department store chains and finance companies, oil companies and mail order companies.

Weingartner (52) has presented a basic summary of the objectives and use of credit rating systems. The objective of consumer credit rating as outlined by Weingartner is to forecast the ability and willingness of a borrower to repay a loan. He reports that studies have usually been carried out on good and bad loans, using a technique such as discriminant analysis to arrive at weights to be assigned to the critical variables in order to categorize a new application as good or bad. The two main results of the analysis are:

1. To determine the critical variables and their weights.
2. To set the cutoff limits which assign a case to a good or bad category.

A study by Myers and Forgy (9, Chapter 6) was carried out at a finance company which already used a simplified rating system. The purpose of the study was to develop a new scoring system for grading conditional sales contracts (i.e., title to the property does not pass to the buyer until all payments are made). Several alternative methods of developing proper weights were tried. The statistical techniques used included discriminant analysis and stepwise regression. The results of the various approaches were compared. In several of the methods used, it was found that as few as 12 variables did as good a job of discriminating between good and bad cases as did 21 variables. The study used
discriminant analysis on all the cases to rank 300 good and bad cases. The 250 cases with the lowest discriminant score (125 good and 125 bad) were used to recompute the discriminant weights. The same was done for the lowest 200, 150, 100, and 50 cases. This system was developed to improve the discriminative power of the critical variables at lower score levels.

Other discussions of the consumer lending problem include those by McGrath (30) and Myers and Cordner (33). The study by McGrath, conducted for an automobile dealer, showed that about 20% of credit losses could be eliminated at the cost of only 1% of good business. Myers and Cordner studied credit accounts in a California chain dealing in personal loans. Results showed that approximately 6% of the losses in a single branch could be eliminated with no loss in the volume of good business; about 24% of losses could be eliminated if 3% of good business could be sacrificed and almost 50% of losses could be eliminated at the cost of 7% of good business.

The variables considered in consumer/lending credit scoring which have been found to be useful in classifying cases as potentially good or bad loans generally include personal data such as age of applicant, time on present job, possession of phone, years at present address, years with present employer, etc.

Orgler has discussed the difficulty of applying the methodology used in credit scoring consumer loans to commercial loans. The difficulty arises from:

1. The lack of standardization among customers in the borrowing population, which presents a problem of obtaining enough data for a statistically significant study.

2. The presence of substantial variations among commercial loans with respect to size, terms, and payback procedures (all of which are relatively uniform in consumer loans).

3. The lack of up-to-date financial data on commercial borrowers and particularly on those who default.
The review of the available literature which follows illustrates the qualitative nature of approaches to the commercial lending credit decision problem, as well as some of the more recent efforts to apply to commercial lending the techniques successfully used in consumer lending.

**Previous Analyses of Commercial Lending**

The decision of the commercial loan officer must be made within the context of a larger framework determined by policies, rules, and procedures of the bank. A number of works are available on general guidelines to bank management policy and practice, such as those by Hodgman (21), Reed (41) and Robinson (46).

Robinson discusses the organization of the lending function and related loan policy and emphasizes the importance to the lending function's management of the activities of credit analysis, budgeting, and supervision.

These references will afford the reader or researcher insight into the impact of management policy on the credit decision process. They are also references which will familiarize the researcher with the important policy variables which should be included in any study purporting to determine the effects of management policy on decision processes.

In an article in *The Journal of Commercial Bank Lending* (45), "Loan Decision Making and the Critical Assumption," G. L. Work states that the commercial lending decision is not whether to make a loan but rather what interest rate and other terms are required in order to make a particular loan. The two major problems of making a loan are said to be:

1. The effect on the lending decision of the terms of the loan.
2. The effect on the lending decision of the characteristics of the borrower. Characteristics of the borrower include the industry he is in and his financial condition.
Characteristics of the borrower include the industry he is in and his financial condition. Terms of the loan include the loan amount, interest rate, security, balances/amount ratio which is to be maintained and repayment procedures. The problem of making the commercial lending decision should be defined as a problem of deciding whether or not to make a loan, given a particular set of credit terms and a particular set of borrower characteristics.

G. L. Work goes on to outline the steps commonly leading to lending decisions:

1. Data relevant to the decision must be gathered and structured.

2. Data must be analyzed and transformed into useful credit information.

3. Decision makers must forecast future possible events resulting from a decision.

4. Alternatives available to the lender must be outlined.

5. Alternatives must be compared and a choice made, as in any decision.

A study made by the research committee of the Michigan Chapter of The Robert Morris Associates on "The Evaluation of Loan Officer Performance" suggests three criteria for performance evaluation:

1. The officer's work methods
2. His results
3. His personality.
The prime personal characteristic was reported to be judgment. The main performance characteristics are the ability to:

1. Assess customer requirements
2. Attract and retain business
3. Handle complex credits
4. Minimize loss experience
5. Understand bank services.

The studies discussed up to this point are relevant for their emphasis on the qualitative aspects of measuring the loan officer's performance. No attempt is made in these studies to emphasize the direct contribution of a loan officer to a bank's profit position by making optimum loans, since most of the studies are of a qualitative nature.

The following studies on credit analysis suggest the additional critical variables which should be included in a study of this subject. They also illustrate the qualitative nature of that analysis, as well as more recent attempts to apply scientific analysis to widely accepted practices. A great deal of literature deals with and emphasizes the importance of the credit evaluation function in the process of lending.

Training in the credit department is considered good preparation for a loan officer at many banks. A number of books have been written on the subject of credit analysis, such as those by Williams (53) and Troy (51), and The Robert Morris Associates "Annual Statement Studies" (42).

Articles such as that by Zimmerman (54) emphasize the qualitative nature of the decision process as it is now carried out. The article stresses the importance of the five C's of credit:

1. **Character** - includes such qualitative factors as responsibility, honesty, integrity, industry, trustworthiness, morality and reputation.
2. **Capacity** - the ability to perform business functions satisfactorily in order to generate funds and use them to repay a loan according to its terms. Capacity also includes the legal or technical ability to enter the contractual agreement represented by the loan documentation.

3. **Capital** - refers to assets, working capital and net worth - the borrower's commitment to the firm prior to the loan.

4. **Collateral** - refers to assets that can be pledged to secure a loan, that can be made available for liquidation in the event of default, and that are sufficient to repay the loan at liquidated value.

5. **Conditions** - include economic conditions of the industry and the national economy.

These C's of credit are, at best, crude reminders of the important facets of a complex and involved process, credit analysis.

The origin of the C's of credit has been traced to a book by William Post, *The Five C's of Credit*, published in Philadelphia in 1910 and are a form of rule of thumb that has persisted for 60 years. Perhaps it is characteristic of people responsible for any complex decision process to try to simplify and search for guidelines and rules of thumb. This is evident in the fact that lenders often resort to the use of ratio comparisons of individual companies to industry ratios and emphasize simple guidelines when available.

Attempts have been made to discover analytical bases for concluding that different institutions have basic differences in the kinds of loans they make. This is reported by Hester (9, Chapter 9). He presents evidence that banks trade off among various terms of loans such as rate, maturity date, and security when granting loans. Suggestions are made to explain the way in which these trade-offs are affected by characteristics of the bank and the potential customer. Hester found that the terms upon which a bank will make a business loan are influenced by borrower characteristics such as profits, current
assets, deposit balances, and total assets; bank characteristics such as the size of the lending bank; and external variables such as the prime interest rate.

Hester defines a "loan offer function" as a relationship which specifies the terms at which a bank with particular characteristics is willing to lend to a borrower with a known profit, balance sheet, and credit history and with particular presumed prospects for the future. This loan function is in reality a supply function; a set of loan terms, determined by external variables. Multiple regression and canonical correlation are employed to arrive at the relationship between internal bank variables and external variables.

The lending officer as a decision maker is discussed by Cohen, Gilmore, and Singer (9, Chapter 10). The objective, as stated by the authors, is to provide an understanding of the types of analyses which bankers undertake and the key factors which influence their decisions to grant or reject loans. The vehicle used is a simulation model which has as its ultimate goal the duplication of the analyses and decisions of the lender with a view to making loans which are as good as or better than his. A basic premise stated by the authors is that the loan decision-making process is not a straightforward optimizing process and that nothing remotely resembling a utility function incorporating bank objectives of minimum risk, maximum profit, and maximum service is apparent in the work of a lending officer.

A central premise is that the loan evaluating process, rather than being an optimizing process, seems to be handled by "sets of Heuristics" which lead to the "satisficing" behavior discussed by H. A. Simon (47). As part of the overall simulation model of loan officer activity, use is made of:

1. Subjective probabilities of potential profits of a new customer relationship, including other bank areas such as trust.

2. Subjective probabilities of the extent to which a new customer relationship will "build the bank." This reflects the emphasis in the banking industry on the size of a bank in terms of assets.

3. Analysis of creditworthiness of an applicant.
Much of the article is devoted to the portion of the simulation model which performs a credit evaluation of bank loan applications. This analysis considers the bank's share of rate, the sufficiency and liquidity of the firm's assets and potential profitability of the customer, to yield a final credit rating. Also considered are legal and policy requirements and appraisal of the loan purpose, amount, maturity date, payback, and security. Simon's article is significant because it attempts to describe an entire decision process with a single simulation model. Also, important variables of the process are identified and the significance of major portions of the process, such as credit evaluation, is emphasized. The variables used in the simulation of credit evaluation include current assets, ratios which measure liquidity of assets, and profitability, as well as outside credit ratings.

The article by Cohen, Gilmore and Singer describes the lending process and is therefore primarily descriptive rather than prescriptive.

**Research on Credit Scoring Related to Lending**

The reports of research on credit scoring in the next few paragraphs indicate the present state of the art in applying quantitative scoring techniques to determine the creditworthiness of commercial and small business loans.

The most recent attempt to predict default of commercial loans appears in a study by Orgler (35). The author used discriminant analysis to find six ratios which would discriminate between "good" and "bad" commercial loans, where bad loans were defined as those which had been criticized by a bank examiner. A classification of "bad" by a bank examiner was assumed to be an indicator of a potentially bad loan.

The objective of Orgler's study was to develop a general credit scoring model for evaluating existing commercial loans. The six variables found useful in predicting bad loans were:

1. Secured versus unsecured loans
2. Past due data
3. Whether or not a loan had been audited
4. Net profit greater or less than zero
5. Ratio of working capital to current assets
6. Whether or not the loan had been criticized on its last examination by a bank examiner.

The study of "bad" loans is the one most directly related to the subject of this dissertation. Prediction of bad loans prior to default is a promising approach to a difficult problem. The data are based on subjective evaluation of "bad" loans by a number of examiners. No attempt is made to ascertain the validity of the bad prediction, that is, to determine if a loan predicted to be bad actually went bad.

In *Developments in Credit Scoring for Commercial Loans* (20), Hammer and Orgler review briefly the preliminary results of research by the Federal Deposit Insurance Corporation in trying to identify the elements which would help differentiate loans of doubtful quality from those of good quality. The variables found to be important included:

- Customer financial statement data
- Customer financial statement ratios
- General financial information concerning the borrower, such as whether his financial statements were audited
- Past performance of a loan, including past evaluations by a bank examiner.

Ewert (16) developed a regression analysis model for reviewing new customers for trade credit. The model used three ratios to screen trade credit applications for small firms, mostly one-owner retail stores. The model combined information on the owner with data on the financial position of the store as listed in Dun & Bradstreet reports.

In an unpublished dissertation at The Ohio State University, Edmister (15) viewed the problem of discriminating between good and bad loans as a problem in ratio analysis. A number of financial ratios were used to discriminate between good and defaulted loans.
made by the Small Business Administration. Discriminant analysis was used to study financial ratios for their usefulness in predicting small business failure. The author concluded that ratio analysis appears to be sensitive to either the purpose of the analysis or the population studied or both. Forecasting functions may be reliable only if based on data from a similar population and failure event.

A study by Beaver (5), carried out in 1966 on firms with high assets, was an attempt to identify a small critical number of ratios, by means of discriminant analysis, which could be used to predict failure of a company. Of some 30 ratios calculated on 79 pairs of failed and nonfailed companies, five ratios were chosen as most useful in predicting failure.

The emphasis in Beaver's study is on financial ratios as predictors of important events, one of which is failure of a firm, and the underlying predictive ability of the financial statements themselves. The conclusions drawn from the study were that some ratios are better predictors than others, e.g., the ratio of cash flow to total debt discriminated between firms that failed and those which were successful. Also, firms which did not fail can be correctly classified to a greater extent than firms which did. Ratio analysis can be useful in predicting failure as far as five years in advance. The Beaver study suggested variables and ratios which were incorporated into the analysis reported in Chapter IV of this dissertation.

Altman (2) used discriminant analysis to determine five ratios which would predict corporate bankruptcy of firms ranging in size from $700,000 to $25,900,000 in assets. In the study, firms which remained solvent were compared with those that went bankrupt. An attempt was made to predict bankruptcy one year prior to its occurrence.

The difficulty with applying the results of Altman's study to commercial loans is that default may result in complete or partial loss of a loan prior to a company's going bankrupt. That is, predicting bankruptcy is not particularly relevant to the commercial loan problem, because the loan may be in default long before bankruptcy occurs.
The study of failure and bankruptcy of a firm are not directly related to this dissertation. Although a loan that is outstanding at the time a firm went bankrupt or failed would probably default, knowing the time of bankruptcy is too late to be useful to the lending officer. These studies are relevant because of the methodology and analyses used to predict an important event in the life of a firm. The studies on failure and bankruptcy indicate that certain financial ratios may be useful to predict failure up to five years before that failure. The variables are, for the most part, different in each study and also differ from those that are useful in predicting bad or defaulted loans.

All of these studies represent attempts to develop empirical bases for statements about criteria for judging what will happen to a firm. They are scientifically designed attempts to assess the validity of commonly accepted rules of thumb or judgmental processes.

The references reviewed here are intended to represent a survey of the work that has been done on topics which bear on the lending decision process. Once again, these topics include management policy which sets the boundaries for the process and the important elements of the lending process such as data search and analysis and credit analysis and evaluation. These references have been used in this chapter to suggest the major facets of the decision process and policy and credit analysis and to aid in determining the important variables related to management policy and credit analysis. These references were also useful in suggesting research approaches and analyses of the variables determined to be part of the decision process.

Table 1, following this page, summarizes studies which illustrate the scope of research on credit scoring of commercial and small business loans. The studies are shown in the first four columns together with a representative example of consumers lending in the last column. "Group" 1 and 2 are the good and bad cases. The criterion generally used to indicate a bad loan has been bankruptcy or default, except in the study by Orgler. The last row shows the Type I and Type II errors of classification, where applicable. Type I is the error of classifying a good loan as bad; Type II is classifying a bad loan as good.
Table I

SUMMARY OF RESEARCH RELATED TO CREDIT SCORING OF LOANS

<table>
<thead>
<tr>
<th></th>
<th>All Items</th>
<th>Brewer</th>
<th>Edminster</th>
<th>Event</th>
<th>Opinion</th>
<th>Consumers</th>
<th>Lending Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 1</td>
<td>11</td>
<td>70</td>
<td>200</td>
<td>100</td>
<td>30</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Group 2</td>
<td>33</td>
<td>70</td>
<td>200</td>
<td>150</td>
<td>50</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Asset Size (n)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Varies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject of Study [16]</td>
<td>Manufacturing</td>
<td>Commercial</td>
<td>Small</td>
<td>Trade-Credit</td>
<td>Commercial</td>
<td>Consumer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Firm</td>
<td>Firm</td>
<td>Business Loans</td>
<td>Loans</td>
<td>Loans</td>
<td>Loans</td>
<td></td>
</tr>
<tr>
<td>Criteria used for</td>
<td>Bankruptcy</td>
<td>Failure</td>
<td>Default</td>
<td>D &amp; B Rating</td>
<td>Bad</td>
<td>Default</td>
<td></td>
</tr>
<tr>
<td>Group 2 (d)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Variables</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Classification Errors:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type I</td>
<td>0.06</td>
<td>0.22</td>
<td>0.01</td>
<td>None</td>
<td>0.43</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>Type II</td>
<td>0.03</td>
<td>0.05</td>
<td>0.01</td>
<td>Available</td>
<td>0.12</td>
<td>0.13</td>
<td></td>
</tr>
</tbody>
</table>

(a) Study by Myers and Forgy (9, Chapter VII)

(b) Firms of small asset size were those whose assets ranged from 1 to 25 million dollars. Firms of large asset size were those whose assets totaled more than 25 million dollars.

(c) Commercial firms - include manufacturing firms as well as others. Small business loans - loans made by the Small Business Administration. Trade credit - credit extended to small store owners.

(d) Bankruptcy - this term is applied to firms that have formally declared bankruptcy. Failure - defined as the inability to pay financial obligations when due and includes bankruptcy, bond default, overdrawn bank account, or nonpayment of stock dividend; default - failure to repay a loan. D & B rating - a good or poor rating by Dun & Bradstreet; bad - defined as a loan that was criticized by a bank examiner.

Notes:
- All items include both consumer and commercial loans.
- Brewer and Edminster refer to specific studies.
- Event and Opinion columns indicate the number of occurrences of each category.
The Type I and Type II errors shown in the table indicate success in predicting the desired results. The costs of the Type I and Type II errors are difficult to determine; for this reason, most studies illustrate the effect on one of these errors of changes in the other error. This will be discussed in detail later in this paper.

Considering the various studies of the firm and loans as being on a time scale, as in the diagram on this page, may help to interpret the results of the studies. The loan decision is made first, then a loan becomes bad and then a default occurs. The input to the loan decision is a set of raw loan applications which have not been prescreened. In the loan decision, some applications for loans are accepted and others rejected. Only those that are accepted for loans can subsequently be examined for "good" or "bad" qualities. Those that are classed as "bad" should have the highest probability of later default, although this has not been verified empirically.

```
Studies of Loans

<table>
<thead>
<tr>
<th>Loan Decision</th>
<th>Bad Loan</th>
<th>Default of a Loan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

```
Studies of the Firm

<table>
<thead>
<tr>
<th>&quot;Failure&quot; of the Firm</th>
<th>Bankruptcy of the Firm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

Predicting default is more successful than predicting bad loans. It appears from the research done to date on failure and bankruptcy that the errors of classification increase the farther away in time from bankruptcy the study is carried out.

The study on bankruptcy by Altman achieved a very low level of classification errors. This could be explained by considering that, by the time a firm reaches bankruptcy, the variables used to predict that bankruptcy clearly indicate a bankrupt state. Similarly, for failed firms, failure may be relatively easy to predict by the time a company reaches a state of failure.
The results for defaulted small business loans are based on analysis after the loans have been defaulted. The errors made in classifying loans that have been defaulted are larger than the errors made in predicting failed and bankrupt firms. Default may be harder to predict than bankruptcy.

Orgler studied his sample of bad loans prior to the actual default of the loans. These loans were prescreened by a lender; that is, those applicants who were bad credit risks, according to his judgement, were not granted loans, and all loans were active loans at the time they were classed as bad by bank examiners.

The classification errors may be large because the obviously bad loans have previously been screened out and marginally bad loans remain, making accurate classification more difficult.

The Type II errors, which represent the classifying of a bad loan or defaulted loan as good, are consistently less than the Type I errors, good loans classed as bad, in order to avoid lending to truly bad applicants; that is, it is believed that the cost of a Type II error exceeds the cost of a Type I error.

It is anticipated that the errors of classification which result from predicting the loan decision, the subject of this dissertation, will be less than those resulting from a study of bad or defaulted loans. Predicting the loan decision is concerned with screening unscreened loan applications, and the results should be better, in terms of fewer classification errors, than predicting prescreened loans. It should be emphasized that the data summarized in Table 1 are the results of analysis of different kinds of loans - that is, trade credit, and small business as well as commercial loans. Edmister (15) has concluded that the results of various studies are sensitive to the population studied and the variables used. Only very rough comparisons can be made among the results of the various analyses.
Order of Presentation

Chapter II discusses some of the important elements of the commercial lending decision process. Chapter III poses research questions and hypotheses and reports on the data-gathering techniques used to collect data which is analyzed in Chapter IV. Chapter IV reports the results of a pilot study on a few cases with a large number of variables, as well as a larger study of a large number of cases with only a few variables. Chapter V proposes a descriptive model of the present lending process and suggests a prescriptive model of that process. Finally, Chapter VI summarizes the results of the analysis and modeling, discusses the hypotheses posed in Chapter III, and suggests areas for further study related to the loan decision process.
Which businesses receive commercial credit and how much they receive is, to a large extent, determined by the commercial banking industry. The individual within the bank organization who plays the most important role in determining how credit is to be allocated is the commercial loan officer. He makes his decisions within the framework of the bank organization, following the policies and procedures set forth by management. But these policies and procedures only partially determine his actions. Very little information is available concerning the total range of factors that determine how he makes the individual credit and lending decisions which are reflected in the loan portfolio of the bank.

The discussion in Chapter II has two objectives:

1. To describe the lending decision process in general terms.

2. To define certain sources of information which may be of use to future researchers who are studying the credit decision process.

The lending officer performs his function within the organizational setting of the bank. The two most important parts of this environment are:

1. Management policy, which sets the rules under which he operates.

2. Supporting departments and groups which provide him with essential information he uses for credit analysis and evaluation.
The focus in this chapter will be on the elements of "organization process" which support the individual decisions made by the loan officer. The term "process" is used in the sense that it was used by Simon (47) (48). Simon considers a decision process a cooperative effort among a number of individuals who act together to reach a decision. This concept of a process aids in considering the entire lending situation to be one in which the lending officer is working in concert with other individuals. All parts of the process contribute to the ultimate decision of whether to allocate funds to a particular creditworthy applicant.

The concept of a decision process is discussed in the literature of operations research, economics, and psychology; this process generally includes these steps:

1. Identification of a problem
2. Obtaining necessary information
3. Determination of possible solutions
4. Evaluation of possible alternative solutions
5. Selection of a strategy of performance
6. Actual performance of an action
7. Subsequent experience, learning and revision

The characteristics of the entire lending process may be studied within this theoretical outline of decision steps.

Step 1, identification of a problem, in times of tight money, is the receiving of a request for funds from a potential borrower or, in times of easier money, it is the result of a marketing effort which seeks out potential borrowers.

Step 2, obtaining necessary information, is an important step for the lender. In practice, this step consists of obtaining sufficient information within reasonable limits of time and cost to determine creditworthiness and decide whether to make a bank loan.

Step 3, determination of possible solutions, comprises, for the lender, an enumeration of all the ways in which the loan could be made, i.e., the various feasible combinations of terms and conditions of the loan.
Step 4, *evaluation of possible alternative solutions*, entails comparison of alternatives on a common basis, in preparation for selecting the terms of the loan that are most favorable for the bank and at the same time for the customer.

Step 5, *selection of a strategy of performance*, is the actual selection of the terms of the loan, and Step 6, *actual performance of an action*, is the formal setting up of the loan relationship.

Step 7, *subsequent experience, learning and revision*, is the subsequent monitoring of customer performance by the lending officer to ensure that the terms and conditions are met. This step also involves an occasional audit by an internal bank loan review function.

For the purposes of this dissertation, the important action takes place during Step 2, obtaining necessary information for making a credit decision. The credit decision process itself will be viewed as a multiple-phase process. The first phase of the process is the researching of information to serve as a basis for making a preliminary judgment as to whether to grant a loan. If an applicant is not rejected, then an in-depth credit analysis is carried out and a decision is made as to creditworthiness. If the decision is affirmative, further decisions must be made concerning the exact terms: the amount, the interest rate, the maturity date, security requirements, in terms of collateral, and methods and schedules of repayment.

One of the most potentially useful types of data pertinent to the credit decision is the available information regarding past financial performance of a loan applicant. This information can be quantitative, as in balance sheets and income statements, or qualitative, as might be reflected in credit files or general information on the applicant's business record.

Preliminary evidence indicates that the experienced lender tends to place greater weight on the "quality of management," a more subjective factor. A new loan officer is likely to spend more time analyzing data on quantitative factors such as working capital, net worth, financial ratios, and the like. The difference between beginners and veterans may
result from the fact that a new officer has not yet learned how to recognize the important qualities of a reliable management. Also, the newer lender probably has not known the customer long enough to make a wise judgment based on qualitative observation of the customer's managerial abilities.

The preliminary evidence referred to above is based on the author's personal observations of the lending decision process. Unfortunately, there is not a great body of empirical data concerning the commercial credit process. Although the credit decisions are important and have a great impact on the United States economy, there have been very few systematic studies of the way these credit decisions are made. Reasons for this lack of research include the attitude held by many lenders that the decision process involves mostly art and skill, and the belief that each lending situation is unique and therefore it is impossible to find universal quantitative variables which apply to the many different situations.

A purely optimal strategy of making loan decisions may not suffice to explain the commercial lending function if it is true, as Nadler has said (34), that banks should make a loan if "humanly possible." Nevertheless, an attempt to maximize the number of creditworthy applications and ultimately an attempt to maximize the return from a loan portfolio, which succeeded in increasing profits on individual loans by even a small percentage would have a significant impact on a bank's profit performance.

Another objective of the commercial lending decision process is to increase a bank's profit performance by reducing the costs of making loans to applicants that are not creditworthy. Considerations such as preventing loan defaults and maintaining the quality of the portfolio are important to reducing these costs.

The Present Multistage Decision Process

The credit decision process is a part of the bank's overall organization process. This relationship is shown schematically in Figure 1 on the following page. This concept of the decision process is based on the author's close observations of the lending function in a single bank and his familiarity with the industry in general.
The organization process includes the marketing function of finding loan applicants and screening their applications. Some applications are rejected after a preliminary check by a loan officer. Those that survive are analyzed for creditworthiness and potential as a satisfactory customer. The applications whose terms have been arranged to a loan officer's satisfaction are submitted to management for a decision on fund allocation. Those loans which are approved for allocation of funds are reviewed by a top management committee and, if approved, these become valid claims on the bank's assets. Loan review is an auditing function which is on the alert for deteriorating credit situations, generally in an advisory capacity. This dissertation is concerned mainly with the credit analysis and evaluation stage of the process.

The Credit Analysis Stage

The analysis needed to determine creditworthiness is controlled by guidelines set by management policy. In carrying out policy, the lending officer receives the support of a number of bank departments and has access to information relating to the credit decision. The following discussion outlines some of the important aspects of loan policy and the supporting information sources.
Management Loan Policy

The management loan policy framework applies during the credit analysis stage. The results of actual decisions should reflect the policy of management regarding such fundamental conditions of a loan as the amount of deposit balances or the ratio of balances to loan amount that a customer must maintain when borrowing. Policy may govern acceptable loan purpose, security and creditworthiness, maturities, interest rate, lines of credit, and conditions of the loan.

Policy may be stated in general terms, without elaboration in detail, such as "any officer may commit the bank in any legal amount up to the legal limit," at one extreme, to specified written rules governing the ranges of the amounts of loans that can be made of a particular type or in a particular industry. For example, full lending authority may be reserved for the senior officers. Arbitrary limits may be specified, for example, at $1,000,000 for senior vice presidents, $500,000 for a vice president when joined by another officer, and $100,000 for a junior lending officer when concurred in by another lending officer. These are simply rule-of-thumb cutoff points for lending authority developed over the years and they differ among banks.

General policy guidelines may be adjusted during times of tight or easy money. In times of tight money the initial credit decision and screening are the responsibility of the loan officer, who then must make a recommendation to a senior officer who has responsibility for allocating limited funds. In times of easy money, the guidelines governing the loan an officer can make may be changed and broadened.

Policy limits may include restrictions as to the types of loans made. Types of loans are defined in Appendix A under "Loan Type," variable $X_{\text{j}}$. A bank may also put a limit on the amount that it will lend to any one industry group, such as the securities/stock brokerage industry or finance companies. This places limitations on the lender which restrict him as well as the customer.
Management policy may regulate, in general terms, the allocation of funds to long-term versus short-term loans. R. E. Palluck (36) has outlined the relevant criteria for term lending policies and practices based on considerable practical experience in making term loans.

It is sometimes difficult for management to determine if policies are carried out according to its dictates. For example, if management establishes a policy guideline which stipulates that any loan over $1,000,000 must be approved by higher level officers, a great many loans may be made at less than $1,000,000, thus effectively circumventing the policy.

Policy governing the acceptable purposes for which money will be loaned may be more or less formally set. Loan purpose, used here, includes:

- Working capital, including accounts receivable, inventory, and tax payments
- Fixed asset acquisition
- Business acquisition
- Other miscellaneous purposes.

Policy regulations on loan purpose, such as not granting loans for mergers or acquisitions, restrict availability of funds for that purpose.

Two elements of policy which are particularly important to this research are the policy relating to:

1. The ratio of balances to amount which a borrower must maintain.
2. The maximum amount a lender has authority to lend.

The policy governing the ratio of deposit balances to loan amount in the bank studied and reported on here was that a corporation desiring to borrow must keep 15% of the borrowed amount in a demand deposit (checking) account. A corporation which was not currently borrowing but decided to do so later was expected to maintain a deposit balance of 10% of the amount to be requested at a later date.
In the bank under study here the policy governing loan amount was simply that a lender must receive clearance from the manager of fund allocation prior to credit analysis on any loan request exceeding $1,000,000. One additional policy constraint was that money would not be lent for the purpose of acquiring another firm. The analysis in later chapters will show the importance of these informal but nevertheless well understood policies.

**Commercial Lending Support Functions and Information Sources**

The role of the lending officer becomes somewhat akin to that of the expediter in a manufacturing operation. The trigger that initiates activity on a loan is a request by a new applicant for money or a request from an old customer who wishes to make use of his bank relationship or who has caused some action to be taken because of failure to fulfill his part of the lending arrangement. The first technical problem encountered on loan requests is obtaining sufficient credit information.

J. B. Pipkin (37) emphasized the part played by the credit department in the lending process. The credit department collects and analyzes customer data and initiates and replies to credit inquiries. It is therefore a staff support function and may at times perform a training function for new lenders. It may also supply specialized staff assistance in specific credit situations.

Numerous sources of information are available to the lender, including bank-maintained credit files containing historical financial data, records of past dealings, and perhaps evaluations of special situations. Another information source is collateral files which provide information on collateral pledged to back up a loan. Loan ledgers contain historical data about past loans as well as data on loans currently outstanding. Demand deposit account information includes historical and current deposit balances and may provide a basis for calculation of charges for the checking activity within the account.

A request from a loan officer for "credit" data generally means a recapitulation of the financial statements of the client. Financial statements are maintained with the credit file information. An important function of maintaining this financial data is to "spread" the
contents of the financial statements into a somewhat standard format. From the bank lender's point of view, this serves to consolidate certain categories of information, to change the classification of other elements, in short, to facilitate comparison to other financial statements.

Standardizing the entries from the balance sheet, income statements, and statement of sources and uses of funds permits comparison of important ratios and data elements. Such standardization also permits period-to-period and year-to-year comparisons of important data, as well as comparison of firms in the same industry classification.

Capture and maintenance of collateral information is a complex, cumbersome, and expensive task. Collateral is requested from certain customers as part of the loan arrangement, normally as a substitute for the capital position of the borrower. Some forms of securities such as stock may be valued at a percentage of the market price; the adequacy of others, such as securities of closely held companies for coverage of a loan, may be subjectively evaluated by the loan officer.

In periods of stock market fluctuation regular review of the entire collateral package is necessary to assure that coverage is sufficient and if not, to request more collateral or a loan payment.

An important historical record is the average balance record or its equivalent. This type of record presents a history of average checking account deposit balances by month for several years. Where this record is hand posted, it comes to be used for entry of other data, such as the date the account opened and the lending officer assigned responsibility for the account.

The importance of the average balance record stems from the historical reliance on "supporting balances" required of borrowers. Reference to a deposit balance history will indicate how well the customer has maintained the deposit part of the bargain over the past several years.
Colquitt (10) has discussed briefly the role of deposit balances in commercial lending. Compensating balances have the effect of raising the interest rate charged on a loan and they also provide a continuing relationship with a bank when a customer is not borrowing.

The loan ledger ranks in importance with the balance card in supplying historical data. The ledger provides the conditions and terms and a summary of the promissory note or notes outstanding, high and low yearly comparison of loan balances, and a history of income, rates and maturity dates. The ledger may also supply historical data on running balances of net borrowing and payments. The summary of the detail notes provides a convenient source for the history of the borrowing relationship of an account. This record may be referred to for data on past borrowing and to determine if an account has "cleaned up", - i.e., dropped to a zero borrowing balance.

Another significant data source is the summary of current deposit balances versus charges for account activity such as charges for check handling. The calculation of the bank's costs for handling an individual or entity account is the weakest single element in the data information system, in most banks. It is only now being recognized that cost accounting, as it is known in manufacturing industries, has implications for the future of banking. A very elementary cost analysis of charges for check handling is about all that is done.

An important gap in the decision maker's information base is knowledge about other relationships a client has with this same lending institution. For example, the customer may have a large trust relationship or may purchase bonds through the bank. A particularly serious gap is the lack of knowledge of trust department relationships. This is important because the historical separation of the trust and commercial banking relationships which in turn is based on legal requirements which assure this separation.

Another information gap is in knowledge about other accounts with which this customer should be affiliated in considering the value of a relationship to the bank. For example, assuming Corporation ABC maintains a relationship as do several of its subsidiaries. One of the subsidiaries may borrow but not maintain balances. The important fact is that
the total corporate relationship is profitable, not just one particular account within the relationship. Information on deposit and loan relationships is needed if the loan application is to be supported with basic account information.

Except for currently outstanding borrowing and deposit balances, much historical data on loans is in terms of averages - for example, average loan balances and average deposit balances. This is generally due to the difficulty of collection, maintenance, and assimilation of large volumes of data on a current basis and the cost of maintaining historical data for a long time.

A source which helped to make this research possible was a weekly report to bank loan management which listed all tumdowns of loan requests, including the company name and amount as well as the reason. From this information, it was then possible to obtain further deposit and financial data on loan tumdowns.

Other miscellaneous but important sources of data are the law department, which performs the usual legal services, and the "discount" department, which handles the clerical maintenance of loan records.

The Credit Evaluation Process

The credit evaluation activity forms the analytical or technical backbone of the commercial lending credit decision process. The support functions and data sources discussed in the previous pages supply the credit data used in credit evaluation of a loan application. The literature on credit analysis in its relationship to banking (18) has a long history.

The scope of a credit investigation depends on a number of factors related to the customer of the loan. The investigation ranges from a complete investigation of a new customer for whom little information is on hand to a cursory check of a blue chip customer. This investigation assembles data on the new customer and ensures that there have been no significant changes in the financial condition of the blue chip customer.
A complete investigation of a loan application includes obtaining information on the history of the firm, including mergers and acquisitions, its form of organization (e.g., by-laws, article of incorporation, etc.), its operating record, financial difficulties, if any, and market and growth prospects. Information on the executives, their experience in the business, their financial expertise, and the depth of management is obtained. Inquiries are made into outside affiliations of individuals, outside investments of the firm, and opinions of impartial individuals as to the integrity and capability of management. Information on the labor situation, purchasing and distribution methods, and unusual hazards or seasonal influences is gathered, and also data on the company's markets.

Most important of all is information on the concern's financial condition, including a review of balance sheets, income statements, statements of source and use of funds, and other miscellaneous data such as lease agreements. Information is gathered through personal visits of an officer, visits by the bank's own internal field auditors, and collection of statements and data from independent sources. Financial data are available from sources such as Moody's Investors Service, Dun & Bradstreet and The Robert Morris Associates (e.g., Annual Statement Studies [42]).

The technical problem in credit analysis is most often viewed as a problem in financial ratio analysis. Ratios such as current assets/current liabilities, net profit/sales, net worth/total liabilities, and sales/inventories are used as indicators of the financial stability of a corporation. Each of the studies referred to earlier by Beaver (5), Edmister (15), Orgler (35) and Altman (2) was an attempt to gain useful information for predicting some particular future event in a company, such as default or bankruptcy, by means of analysis of historical ratio data. Ratio analysis has long been regarded as an art or a skill. Edmister and Orgler both attempted to develop scientifically designed studies to test assertions concerning which financial ratios should be used and the proper values of these ratios.

Various rules of thumb for determining acceptable ratios are known to exist. For example, a two-to-one ratio of current assets to current liabilities and a one-to-one ratio of debt to net worth are commonly accepted ratios. In actual practice, these standards are used merely as guidelines. Different lenders may use different ratios and different levels or
values for those ratios. Evidence presented in this dissertation indicates that, at least in
times of tight money, the historical reliance on analysis of financial ratios as a basis for
making a loan is too simplistic. Factors other than commonly accepted ratios must be
taken into account. These factors will be discussed later.

Summary

This discussion of the commercial bank lending decision process has sketched some of
the important elements of that process. The lender is charged with responsibility for making
decisions which allocate the bank's assets to borrowing customers in the form of loans.
Interest on these loans supplies the bank's major source of income. The organization process
is governed by management policies which are restated from time to time to take into
account changing conditions in the marketplace, such as tight money.

Of all the traditional steps taken in making a loan decision, the most important are:

1. Credit analysis and evaluation.
2. Fund allocation.

Of these two steps, credit analysis and evaluation have historically received more attention
from lenders. They have, however, not received much attention from researchers. Very
often a good lender is considered to be one who understood credit analysis and could
apply the techniques of credit evaluation. The allocation decision depended simply on
whether or not funds were available. Thus, emphasis in the remainder of this study will
be on the credit analysis portion of the lending decision.
CHAPTER III

RESEARCH DESIGN

The purpose of this chapter is to define the variables used in subsequent analyses, pose research questions, state the hypotheses to be tested by subsequent research and discuss the data collection procedures.

The research questions and hypotheses set forth here will be tested in the next chapter on the basis of analysis of the results of, first, a pilot study and, second, a larger sample of key variables in the credit decision process.

The period covered by the data was a time of relatively uncertain economic conditions. The years 1968 - 1970 were generally recognized as a period of "tight money," that is, money for commercial lending was scarce, as reflected in the high prime interest rate of 8% to 8.5% and the high rediscoun rate.

Summary Description of the Variables

Appendix A contains a detailed explanation of the variables used in the analysis. The $n$ variables can be put into several major categories, including the following:

- $X_1$ - a loan officer variable, years of experience
- $X_2 - X_8$ loan variables - e.g., interest rate, loan amount
- $X_9 - X_{12}$ borrower variables - e.g., years a customer
- $X_{13} - X_{26}$ balance sheet variables - e.g., cash, inventory
- $X_{27} - X_{30}$ income statement variables - e.g., sales, profits
- $X_{34}$ - ratio variable - balances/amount
These broad categories that are useful for data collection are further subdivided into subsets of variables for purposes of analysis. These subsets include:

- **Policy variables**
  - $X_4$ - Loan amount
  - $X_{31}$ - Type of loan
  - $X_{32}$ - Purpose of loan
  - $X_{34}$ - Deposit balances/amount

- **Credit variables**
  - $X_{14}$ - Accounts receivable
  - $X_{16}$ - Total current assets
  - $X_{22}$ - Long-term debt
  - $X_{24}$ - Net worth
  - $X_{25}$ - Working capital
  - $X_{26}$ - Total liabilities

Furthermore, policy variables are a subset of all credit variables.

**Pilot Study: Stratified Sample of Loans and Turndowns**

One difficulty in taking a random sample arises because 5,000 active loans are contained within a file of some 10,000 to 12,000 accounts. A stratified sample is more satisfactory for selecting a small sample from this large population, since it results in a sample of active cases only and ignores the inactive.

The data collected for the pilot study on loans and turndowns is a stratified sample from the loan application population. The sample was chosen from cases suggested by loan
officers as being representative of difficult credit decisions. The reasoning behind the approach of having the lender screen the cases to be used is presented in the following paragraphs.

The proportion of loans made is perhaps 70%. Turndowns make up a smaller portion of the applications, roughly 30%. In the total population of loan applications, it is likely that, for a variable such as deposit balances, \( X_7 \), the means of the distribution of discriminant function values of the two populations of loans and turndowns are significantly different. There is assumed to be only a small "overlap" of the two populations. This overlap indicates the misclassification of loans as turndowns and turndowns as loans, as discussed in the next few paragraphs.

For example, consider two normal subpopulations, loans, \( \Pi_1 \), and turndowns, \( \Pi_2 \), which represent routine credit decisions and have distributions on discriminant function values as a result of discriminant analysis as shown below:

\[
\text{turndowns} = \Pi_2 \quad || \quad \Pi_1 = \text{loans}
\]

where the means \( \mu_1 \) and \( \mu_2 \) of the distribution of the discriminant functions are some significant distance apart. A critical value of the discriminant function at \( K \) would distinguish between the populations with very little misclassification. That is, the probability of misclassifying observations from \( \Pi_1 \) as being from \( \Pi_2 \) and vice versa is very small, less than \( \epsilon_i, i=1,2 \).

\[
\begin{align*}
P(\Pi_1 \mid \Pi_2) & \leq \epsilon_1 \\
P(\Pi_2 \mid \Pi_1) & \leq \epsilon_2
\end{align*}
\]
Consider next a sample which represents difficult credit decisions in the population of loan applications. If, in fact, the sample cases require that a decision for either loan or turndown be made, it is likely that there will be considerable overlap between the two subpopulations; i.e., the means of the distributions of the discriminant functions for the subpopulations are close together and the probability of misclassification is large. Analysis of the two groups of sample loans and turndowns which is intended to distinguish between the groups on the basis of several variables must allow for the probabilities of misclassification of the two groups. In some cases on the borderline between the two groups, it will be difficult to distinguish the group to which they properly belong.

The subpopulations for difficult credit decisions are now as shown below:

![Diagram showing distributions of discriminant functions]

The distributions of the discriminant functions now overlap and it is possible to have a large number of misclassifications of cases taken from the two groups.

The probabilities of misclassification are larger than in those cases previously discussed. That is,

\[
P(\tau_1 | \tau_2) \leq \xi_3
\]

and

\[
P(\tau_2 | \tau_1) \leq \xi_4
\]

where

\[
\xi_3 > \xi_1 \text{ and } \xi_4 > \xi_2
\]
It is conceivable that credit decisions requiring lending officer judgment would present the greatest difficulty to quantitative analysis techniques. If meaningful conclusions can be gained from analysis of the difficult cases, this would provide encouragement to further exploration of the easier credit cases. For this reason, the observations chosen for the pilot study were selected on the basis of the lender's judgment that the observations involved hard credit decisions.

Research Questions and Hypotheses

The outline and discussion of the decision process in Chapters I and II and the subsequent analysis and modeling described in Chapters IV and V are oriented toward testing or amplification of the following research questions and hypotheses:

1. *Can management policy variables* $X_a \ldots X_b$ *be used to distinguish between loans and turndowns?*

   $H_1$: Management policy variables will be significant in classifying applications as loans or turndowns.

   $H_2$: There are no significant differences between lending divisions (industries) in the key policy variables such as $X_{34}$, balances/amount.

2. *Are policy variables* $X_a \ldots X_b$ *better at predicting loans and turndowns than are credit variables* $X_{c \ldots d}$?

   $H_3$: Policy variables such as balances/amount, $X_{34}$, will be more significant in classifying cases as loans or turndowns than credit variables, such as assets, $X_{16}$, or customer long-term debt load, $X_{22}$.

It is assumed that analysis of loans made and loans turned down will indicate the key variables which enter the decision process, even though certain qualitative information, such as subjective evaluation of the competence and financial expertise of customer management, is omitted. In addition, other quantitative information on customer profitability to the bank is not available and is therefore also omitted.
Data Collection: The Pilot Study

One of the most significant aspects of the data collection effort is the selection of the variables on which data will be gathered and which will be used in subsequent analysis. The variables used in this analysis were chosen because they appear most often in discussions of the lending activity both in the literature reviewed in Chapter I and in conversations with individual lenders.

Some of the variables were the same as those used by Hester (9, Chapter 20), and Cohen, Gilmore, and Singer (9, Chapter 10) as well as Financial Statements for Bank Credit Purposes, published by The Robert Morris Associates (44). The variables finally selected to enter the analysis were those actual variables used by the lender rather than derivatives of those variables. For example, the latest year's profit, which is available to the lender, is used rather than some hypothetical variable such as the variance of the applicant's profits for five years prior to the loan, as used by Hester (9, Chapter 20).

The data represent a cross section of commercial loan customers, divisions, and officers. The analysis to be discussed in the next chapter is concerned with determining the variables which appear to govern lending decisions when viewing the lending activity in an aggregate sense, that is, across several divisions, lenders, and industries.

For the pilot study, data on commercial loans were collected from credit files, loan ledgers, and demand deposit balance books, as well as from interviews with lending officers (25) and discussions with lending division and credit department personnel. Data were also collected from standardized financial statements for companies representing eight different industry classifications.

Some 120 cases were screened to get the sample of 51 cases; 34 cases for loans made and 17 for applications turned down are included. This does not represent the true ratio of loans to turn-downs, because the computer program used to analyze the data requires that the number of cases be greater than the number of variables examined. For each case or company, 33 variables were originally tabulated, 18 of these representing the loan officer, loan terms and conditions, and customer characteristics. Additional variables such
as financial ratios can be calculated from the variables gathered. For example, the current ratio is a new variable calculated from the ratio of current assets to current liabilities.

Data were included for variables which would facilitate calculation of the most commonly used ratios as found on bank standardized financial statement "spread sheets," The Robert Morris Associates Annual Statement Studies (42), and the ratios found in books by Reed (41) and a study by Beaver (5).

The companies whose data were used in the pilot study data analysis were chosen as a stratified sample from a large population of loans made during 1968 and 1969, a period of generally "tight money" conditions. Most of the cases were selected to represent an instance where a credit decision was required. None of the cases was a simple or routine credit decision, such as a simple renewal of a ninety-day note.

Individual loan officers were interviewed to identify cases which represented "interesting" credit decisions (e.g., other than a simple renewal of a ninety-day note), whether accepted or turned down. After the loan officer had suggested an interesting credit situation, it was necessary to determine if sufficient data were available to collect data on the 33 variables chosen for analysis. An attempt was made to gather data on cases of both loans made and turned down by the same lender.

Fourteen new loan customers are included in the data. Although not many automatic credits were included, there were several cases of prestige loans or loans made as "accommodations," as well as several loans for which collection was doubtful but eventually paid out. Also included are several cases of customers who had a long-time relationship and finally requested money. Some were turned down and some were granted. Data are included on several situations which did not conform to management policy regarding the purpose for which funds would be loaned; for example, loans for acquisitions.

As well as representing a cross section of loan divisions and industries, the data represent a cross section of lending officer experience. Loan officers cross lending division lines during their careers. Some will also have experience in term loans, the credit department, and other financial areas such as the trust department or investment banking. The average experience of loan officers represented in the 34 cases of loans made was seven years.
Figure 2, following this page, shows histograms of some of the important variables of the cases represented by the data. These variables are used to test the hypotheses presented at the beginning of this chapter.

Among the 34 cases, 16 were made at prime rate and 18 at greater than prime. Fifteen were for loans greater than $1,000,000 and 19 were less. Twelve cases represent the loan type, line of credit. Fourteen loans were for working capital and 20 for other purposes. Finally, 14 were made to customers one year or less and 20 were made to those who had been customers longer than one year. The average number of years was twelve and one-half years.

Seven of the applications that were turned down were for less than $1,000,000 and ten were for more than that sum. Ten were lines of credit, new or increases, and seven were other types of requests. Five requests were for working capital, and 12 were for other reasons. Turndowns averaged ten years as customers. Eight of seventeen turndowns had been customers for less than one year.

A fundamental assumption which is made prior to data analysis is that the decision maker, the loan officer, is rational in the sense that, given the same data repeatedly on the same loan, he would make the same decision. This assumption is significant in that, if the lender is assumed to be rational, then any differences in classifying cases by some other means should be attributed to that classification system and not the lender. This assumption will apply until after the data are analyzed and reasons are found to indicate why the assumption may be false.

Data Collection - The Larger Sample

As a result of analysis of the data collected for the pilot study, a small number of variables were selected for use in distinguishing between loans made and applications turned down. The variables were balances, $X_7$, amount, $X_4$, and years a customer, $X_{12}$. The additional data were collected as a random sample from loans made and turndowns. The data included 200 cases of loans made and 60 turndowns. The results of the analysis of the data from the pilot study and the larger study are reported in the next chapter.
Figure 2. Histograms of Variables.
Limitations of the Data

Limitations of this study originate from several sources. One limitation is the difficulty of collecting data on a large number of variables for even a few cases. This is the main reason that more work has not been done on the lending decision prior to default of a loan. The difficulty arises because the data are confidential and not readily available. An advantage for this study was that the author had access to "spread sheets" containing standardized financial data on a number of firms as well as to the loan officer who had made the loan. Financial data were obtained for the year just prior to the making of the lending decision. Altman (2) and Beaver (5) have shown that one financial statement contained sufficient information for a discriminant function of large businesses in predicting financial failure.

A second limitation of this study is that it deals with only one large bank. The policy and credit variables may be peculiar to this particular bank and may not be generally applicable to a large number of banks.

A final limitation is that this study was carried out during a period of tight money. This has the advantage that economic conditions were fairly uniform throughout the country during the time of the study. An individual institution could be particular about lending to an individual company, at least until companies discovered commercial paper. A disadvantage of studying the lending problem during a tight money period is that it was difficult to determine the real reason a loan was turned down. The reason most often given was "tight money," rather than a more specific reason.
CHAPTER IV

DATA ANALYSIS AND EVALUATION OF RESULTS

Data Analysis Methodology

This chapter describes the various analyses of the data for both the pilot study and the larger sample. The correlation matrix obtained for all variables is discussed briefly and is exhibited in Appendix D. Multivariate discriminant analysis is used as an analysis technique to determine the variables and the coefficients of a linear combination of those variables which distinguish between loans and turn downs. A discussion of the mathematical basis for discriminant analysis is contained in Appendix B. Multivariate normal discriminant analysis is a well developed technique (32) and is used here as a first approximation. Different results, in terms of fewer classification errors, are obtained from the use of an exponential discriminant function. Exponential distributions are used to determine the results for one and two independent variables. These results are compared with the results of the normal approximations.

The data analysis was carried out on an IBM System/360, Model 75 using a number of statistical routines available in the library of programs of the Instruction and Research Computer Center, The Ohio State University.

The Discriminant Analysis

The statistical problem in this research is one of separating individual applicants into one of two classes, loans or turn downs on the basis of a number of variables. Multiple discriminant analysis is a method for forming a linear model which can be used to classify individual cases on the basis of observed characteristics.
Data for the pilot study on 34 cases of loans made and 17 cases for turndowns were analyzed using discriminant analysis on various combinations of variables to yield the most satisfactory discriminant function which will distinguish between the two groups, loans made and turndowns.

The concepts of discriminant analysis are best understood by first assuming \( t \) measurements on two groups of cases of sizes \( N_1 \) and \( N_2 \). We seek some linear combination of the variables which will maximize the "difference between groups" relative to the "differences within groups" (17). One approach might be to compare the two groups on each of the \( t \) variables one at a time, but this ignores the interrelationships among the variables and, in addition, does not allow for assessment of the relative power of each of the \( t \) variables.

Cohen and Hammer (9, page 128) discuss the essential similarities and differences between discriminant analysis and regression analysis. The statistical techniques of the two are similar in several ways; the computational techniques are similar but not identical. They are both multivariate techniques which can be used to summarize the simultaneous effects of several interacting variables.

The main differences between the two techniques is that discriminant analysis assumes a special measurement scale for the dependent variable; that is, in discriminant analysis, the dependent variable is assumed to be 0 or 1 for observations from the two populations. Regression analysis attempts to estimate the coefficients of a single linear relationship which best describes the interrelationships among the variables on observations from a single population. Discriminant analysis attempts to estimate coefficients of a different linear relationship which best categorizes the observations into one or another of the separate populations.

A regression function of the form \( y = a + bx \) relates \( x \) to \( y \). If observations are taken from two separate populations such as loans and turndowns, and the observations from one population - say turndowns - are assigned a \( y \) value of 0 and the observations from the other population are assigned a \( y \) value of 1, a regression analysis on the data will yield results equivalent to those of discriminant analysis and will contain additional statistical information (e.g., the standard error of the individual coefficients will allow
for tests of significance of those coefficients. Furthermore, Anderson (4, pages 140-141) has shown that the relationship between the discriminant coefficients and the regression coefficients is one of proportionality to the variance-covariance matrix. Also, Ladd (29) has discussed the significance of 0 and 1 values for \( y \).

The use of discriminant analysis yields a set of weights \( a_i \) which, when applied to the independent variables \( x_{ti} \) of the cases in a group, yield a value for each case:

\[
d_i = a_1 x_{1i} + a_2 x_{2i} + \ldots + a_t x_{ti}
\]

In order to classify \( d_i \) as belonging in group one, loans, or group two, turn downs, the set of weights \( a_i \), are applied to the means \( \bar{x}_{ti} \) of the variables to arrive at a discriminant function value for each group, that is \( \bar{d}_1 \) and \( \bar{d}_2 \). For group one:

\[
\bar{d}_1 = a_1 \bar{x}_{11} + a_2 \bar{x}_{21} + \ldots + a_t \bar{x}_{t1}
\]

and for group two:

\[
\bar{d}_2 = a_1 \bar{x}_{12} + a_2 \bar{x}_{22} + \ldots + a_t \bar{x}_{t2}
\]

Assign a case having a value for \( d_i \) close to \( \bar{d}_1 \) to group one and a case having a value for \( d_i \) close to \( \bar{d}_2 \) to group two. Assuming equal size groups, the value \( k \) for classifying \( d_i \) is determined to be midway between the two groups as

\[
k = \frac{\bar{d}_1 + \bar{d}_2}{2}
\]
When the covariances of the populations are assumed to be equal and the proportions of cases in each population are identical, the above expression holds. When the proportion of cases in population one, $\Pi_1$, differs from the proportion in population two, $\Pi_2$, the discriminant critical value is given by

$$ k = \frac{d_1 + d_2}{2} + \text{adjustment} $$

The adjustment factor will be discussed later.

Morrison (32, pages 84-85) points out that the usual probabilities of misclassification computed from the assumptions of normality of the discriminant population are true only if the parameters of the population, the mean and the variance, are known or are at least estimated from substantially large samples.

**Using Costs To Determine Cutoff Values**

The sample data were taken across several divisions and industries to determine if information about the population of loans and turn-downs could be gained from such a sample. Many lending situations are unique, but in spite of these difficulties it is also recognized that people have limited cognitive ability. People tend to simplify unusual situations to fit a pattern familiar to them. Morris (31, Chapter 2) has discussed the problem of simplification and gives a number of examples of ways in which decision makers attempt to simplify complex decision situations. As a first attempt to predict the lender's decision, multivariate normal discriminant analysis is used to find patterns in the data.

In the case of two populations where the a priori probabilities of the two populations are known, Anderson (4, Chapter 6) discusses a procedure for determining the critical values which minimize the expected cost of misclassification. Let the a priori probability that an observation comes from population $\Pi_1$ be $q_1$ and that it comes from population $\Pi_2$ be $q_2$. The density function of population $\Pi_1$ is $p_1(x)$ and of $\Pi_2$, $p_2(x)$. Assuming a region of classification, $R_1$, to be from $\Pi_1$ and $R_2$, from $\Pi_2$, the following probabilities are specified.
The probability of correctly classifying an observation actually drawn from population $\Pi_1$ is, for $R$, the critical region for both populations,

$$P(1 \mid 1, R) = \int_{R_1} p_1(x) dx$$

The probability of misclassification of an observation from $\Pi_1$ is

$$P(2 \mid 1, R) = \int_{R_2} p_1(x) dx$$

The probability of correctly classifying an observation from $\Pi_2$ is

$$P(2 \mid 2, R) = \int_{R_2} p_2(x) dx$$

The probability of misclassifying an observation from $\Pi_2$ is

$$P(1 \mid 2, R) = \int_{R_1} p_2(x) dx$$

The probability of drawing an observation from $\Pi_1$ is $q_1$ and therefore the joint probability of drawing an observation from $\Pi_1$ and correctly classifying it is $q_1 \cdot P(1 \mid 1, R)$. This is the probability of the situation in the upper lefthand corner of the contingency table below. The zero is the cost of correct classification. The columns labeled $\Pi_1$ and $\Pi_2$ under "population" represent the actual observations in the two populations. The rows labeled "classification decision" represent the observations as classified into groups, loan and refuse. The lower left and upper right corners thus represent the errors of classification, the Type I and Type II errors, of statistical hypothesis testing.

<table>
<thead>
<tr>
<th>Population</th>
<th>$\Pi_1$</th>
<th>$\Pi_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loans</td>
<td>0</td>
<td>$C(1 \mid 2)$</td>
</tr>
<tr>
<td>Turndowns</td>
<td>$C(2 \mid 1)$</td>
<td>0</td>
</tr>
</tbody>
</table>
In statistical hypothesis testing, the size of the Type I error is the probability that observation will fall within the critical region when the hypothesis is true, that is, the probability of rejecting the hypothesis when it is true. The Type II error is the probability that an observation will fall within the noncritical region when the alternative hypothesis is true, that is, the probability of accepting the hypothesis when it is false (22, page 33).

The joint probability of drawing an observation from $\Pi_1$ and misclassifying it is $q_1 \cdot P(2 \mid 1, R)$, the situation in the lower lefthand corner of the diagram, corresponding to the Type I error in statistical hypothesis testing. $C(2 \mid 1)$ is the cost of improper classification. The joint probability of drawing from $\Pi_2$ and misclassifying is the situation shown in the upper righthand corner, the Type II error, $q_2 \cdot P(1 \mid 2, R)$. The cost or loss associated with the Type II error is $C(1 \mid 2)$. The joint probability in the lower righthand corner is $q_2 \cdot P(2 \mid 1, R)$.

The expected cost of misclassification is the sum of the products of each misclassification cost multiplied by the probability of its occurrence:

$$C(2 \mid 1) \cdot q_1 \cdot P(2 \mid 1, R) + C(1 \mid 2) \cdot q_2 \cdot P(1 \mid 2, R)$$

The purpose is to minimize this expected cost.

The expected cost of misclassification is also given by the expression:

$$q_1 \cdot C(2 \mid 1) \cdot \int_{R_2} p_1(x) dx + q_2 \cdot C(1 \mid 2) \cdot \int_{R_1} p_2(x) dx$$

For a given observed $x$, Anderson (3, Chapter 6) has used Bayesian analysis to show that the cost of misclassification is minimized by assigning the point to the population which has the higher expected cost. If

$$C(2 \mid 1) \cdot q_1 \cdot P(2 \mid 1) \geq C(1 \mid 2) \cdot q_2 \cdot P(1 \mid 2)$$

choose population $\Pi_1$; otherwise choose $\Pi_2$. 

Anderson (3, page 130) develops the solution in terms of density functions. The rule is, assign an observation to $R_1$ if
\[
C(2 \mid 1) \cdot q_1 \cdot p_1(x) \geq C(1 \mid 2) \cdot q_2 \cdot p_2(x)
\]
and to $R_2$ if
\[
C(2 \mid 1) \cdot q_1 \cdot p_1(x) < C(1 \mid 2) \cdot q_2 \cdot p_2(x)
\]
or alternatively to $R_1$ if
\[
\frac{p_1(x)}{p_2(x)} \geq \frac{C(1 \mid 2) \cdot q_2}{C(2 \mid 1) \cdot q_1}
\]
or to $R_2$ if the expression on the left is less than the expression on the right. The cutoff value is the value on the right.

If the costs are assumed to be equal for Type I and Type II errors and the proportions from the populations are equal in the two populations, the expression for classification in region $R_1$ is obtained from
\[
k = \frac{d_1 + d_2}{2} + \ln s
\]
where
\[
s = \frac{C(1 \mid 2) \cdot q_1}{C(2 \mid 1) \cdot q_2} = 1
\]
For populations of equal size having equal costs, ln s = 0 and

$$k = \frac{\bar{d}_1 + \bar{d}_2}{2}$$

In the case of normal distributions, the cutoff value we seek, $k$, is midway between the means of the distributions of the two populations.

**Multivariate Normal Discriminant Analysis**

The discriminant function can be used to classify an application within the accepted or turned down category. In the case of the lending decision problem, it is the Type II error that is the more serious error. Classifying as a loan an application that should have been a turndown may be more serious because the lender might be less inclined to review an application classed as a loan by a computer model.

Table 2 gives the results of the discriminant analysis on two variables and shows the variable and the order in which it was entered into the analysis. A matrix shows how many cases have been given the same classification as that given them by a loan officer on the basis of the variables used up to that point. Also shown is the value of the $F$ statistic calculated to test for significance, together with the degrees of freedom. The heading "loan officer decision" indicates the way cases were classed as loans or turndowns by the lender. The heading "program decision" indicates the classification made by the program.
Table 2

CLASSIFICATION OF CASES BY TWO VARIABLES
IN ORDER OF ENTRY

<table>
<thead>
<tr>
<th>Variable</th>
<th>Classification</th>
<th>F value</th>
<th>d.f.</th>
<th>Table F</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_{34}$ Balances/Amount</td>
<td>Loan Officer Decision</td>
<td>Loan Refuse</td>
<td>5.22</td>
<td>1.49 4.03</td>
</tr>
<tr>
<td></td>
<td>Loan</td>
<td>12</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Refuse</td>
<td>22</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>$X_7$ Balances</td>
<td>Loan Officer Decision</td>
<td>Loan Refuse</td>
<td>4.65</td>
<td>2.48 3.19</td>
</tr>
<tr>
<td></td>
<td>Loan</td>
<td>16</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Refuse</td>
<td>18</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 illustrates the results of a stepwise, two variable vector discriminant analysis. The first variable, $X_{34}$, is used by itself to classify cases as "loan" or "refuse" as shown. The second variable, $X_7$, is added and the two together form the discriminant function used to classify cases.

These results suggest that this preliminary discriminant analysis could be used to carry out a test on a new loan application and, if the result indicated a loan, it might be a reasonably accurate classification. If the results indicated a turndown, the case would be reviewed by a lending officer. Further discussion will refine this first approximation to a decision model.
Appendix E shows the results of a stepwise discriminant analysis, documented by Dixon (13), carried out on five variables. Variables are added and deleted one at a time from a nucleus of variables which best discriminate between groups under consideration.

**Exponential Discriminant Analysis**

In a reasonably large sample of data, a linear combination of several variables, such as is used in discriminant analysis, would tend to be normally distributed. This is true whether or not the individual variables are normally distributed. The discriminant function for the groups of loans and turn-down cases would be normally distributed. However, in the data gathered from actual loan files, the distributions of the key variable balances/amount is believed to be exponential. The following discussion develops the equivalent of the discriminant function for exponential variables. Later, the results of the larger sample will substantiate the finding that the key variables in the samples are exponentially distributed.

The density function, \( f(x) \), for the exponential distribution is

\[
f(x) = \frac{1}{\Theta} e^{-\frac{x}{\Theta}}
\]

and for the cumulative distribution is

\[
F(x) = 1 - e^{-\frac{x}{\Theta}}
\]

where \( \Theta \) is the mean of the distribution.
For the single variable $X_{34}$, balances/amount, the sample density functions for the two groups of loans and turndowns appears as follows:

![Graph showing density functions for loans and turndowns.]

The purpose is to find a value of $x$ which distinguishes between the two groups, loans and turndowns. The procedure described by Anderson (4) and discussed earlier is used to determine the ratio:

$$\frac{p_1(x)}{p_2(x)} \geq \frac{C(1 \mid 2) \cdot q_2}{C(2 \mid 1) \cdot q_1}$$

$p_1(x)$ and $p_2(x)$ are the density functions for loans and turndowns, $C(1 \mid 2)$ is the cost of misclassification of a turndown as a loan and $C(2 \mid 1)$ is the cost of misclassification of a loan as a turndown; $q_1$ and $q_2$ are the proportions in the populations of loans and turndowns, $\theta_{x1}$ and $\theta_{x2}$ are means of the populations.

The ratio of densities for the exponential distributions is:

$$\frac{\frac{1}{\theta_{x1}} \cdot e^{-\frac{x}{\theta_{x1}}}}{\frac{1}{\theta_{x2}} \cdot e^{-\frac{x}{\theta_{x2}}}} \geq \frac{C(1 \mid 2) \cdot q_2}{C(2 \mid 1) \cdot q_1}$$
It is assumed that $q_1 = q_2$ and that the costs of misclassification are equal, so the righthand side is one. Taking logarithms and rearranging terms, the expression becomes:

$$\ln \left( \frac{\Theta_{x2}}{\Theta_{x1}} \right) - \frac{x_1}{\Theta_{x1}} + \frac{x_2}{\Theta_{x2}} \geq 0$$

For the exponential distribution with one variable, the value of $x$ is the same for both populations and

$$- \frac{x}{\Theta_{x1}} + \frac{x}{\Theta_{x2}} + \ln \left( \frac{\Theta_{x2}}{\Theta_{x1}} \right) \geq 0$$

The classification matrix for the new exponential distribution on variable $X_{34}$, balances/amount, is:

<table>
<thead>
<tr>
<th>Loan Officer Decision</th>
<th>L</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>21</td>
<td>2</td>
</tr>
</tbody>
</table>

Exponential Classification

<table>
<thead>
<tr>
<th></th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>13</td>
</tr>
<tr>
<td>T</td>
<td>15</td>
</tr>
</tbody>
</table>

$L$ indicates a loan made and $T$ indicates a turndown. In the above calculations, $\Theta_{x1} = .6$, $\Theta_{x2} = .1$, and $X_{34} = .2$. $x$ will vary according to the means of the distributions of the two populations.
An example of the calculations used to arrive at the foregoing table is shown below. An $x$ must be found for which the following expression applies:

$$f(x) = -\frac{x}{\Theta_{x1}} + \frac{x}{\Theta_{x2}} \geq -\ln \frac{\Theta_{x2}}{\Theta_{x1}}$$

or

$$f(x) = -\frac{x}{.6} + \frac{x}{.1} \geq -\ln \frac{.1}{.6}$$

A table of various values illustrates the correct value:

<table>
<thead>
<tr>
<th>$x$</th>
<th>$f(x)$</th>
<th>$-\ln \frac{1}{.6}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.83</td>
<td>1.83</td>
</tr>
<tr>
<td>.19</td>
<td>1.58</td>
<td>1.83</td>
</tr>
<tr>
<td>.20</td>
<td>1.67</td>
<td>1.83</td>
</tr>
<tr>
<td>.21</td>
<td>1.85</td>
<td>1.83</td>
</tr>
<tr>
<td>.25</td>
<td>2.09</td>
<td>1.83</td>
</tr>
<tr>
<td>.30</td>
<td>4.50</td>
<td>1.83</td>
</tr>
</tbody>
</table>

The value for which the above expression holds is $x = .21$.

Assuming the cost of misclassification of loans as turndowns, $C(2 \mid 1) = 1$, to be equal to the cost of classifying turndowns as loans, $C(1 \mid 2) = 1$, the cost of misclassification for the normal case is $22 + 0 = 22$ and for the exponential case, it is $13 + 2 = 15$, the sum of the costs of misclassification.
For the case of two variables $X_{34}$, balances/amount, and $X_7$, balances only, which are both exponentially distributed variables, the normal discriminant analysis as reported previously, classifies cases as follows:

<table>
<thead>
<tr>
<th>Loan Officer Decision</th>
<th>L</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>Program Decision</td>
<td>T</td>
<td>18</td>
</tr>
</tbody>
</table>

For the case of two variables with independent exponential distributions, the density functions are:

$$f(x_1,y_1) = \frac{1}{\Theta_{x_1}} \frac{1}{\Theta_{y_1}} \exp \left( -\frac{x_1}{\Theta_{x_1}} - \frac{y_1}{\Theta_{y_1}} \right)$$

and

$$f(x_2,y_2) = \frac{1}{\Theta_{x_2}} \frac{1}{\Theta_{y_2}} \exp \left( -\frac{x_2}{\Theta_{x_2}} - \frac{y_2}{\Theta_{y_2}} \right)$$

$x_1$ and $y_1$ are the loans for the two variables; $x_2$ and $y_2$ are the turndowns. The ratio of these densities yields:

$$\ln \frac{\Theta_{x_2}}{\Theta_{x_1}} - \frac{x}{\Theta_{x_1}} + \frac{x}{\Theta_{x_2}} + \ln \frac{\Theta_{y_2}}{\Theta_{y_1}} - \frac{y}{\Theta_{y_1}} + \frac{y}{\Theta_{y_2}} \geq \ln k = 0$$

Addition of more independent variables increases the number of terms in the expression on the left.
The value of $x = y$ which satisfies this relationship is $x = y = .22$, and the classification matrix for the exponential discriminant function is the following:

<table>
<thead>
<tr>
<th>Loan Officer Decision</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L</td>
<td>T</td>
</tr>
<tr>
<td>L</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td>Exponential Classification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>17</td>
<td>15</td>
</tr>
</tbody>
</table>

Returning to the diagram for variable $X_{34}$, balances/amount, for loans and turnovers, as shown below in the diagram indicates that a number of cases were classed as loans which would be classed as turnovers by discriminant analysis. It is possible that there are other variables which are used by the lender in his analysis but are not included in the discriminant function.

The cases having $X_{34} \geq .30$ were removed from the group of loans and a discriminant analysis, with the cases having the lowest ratio of balances to loan amount from the loan group and the turnover group, indicated that the variable which aided in distinguishing between the groups was $X_{12}$, number of years a customer.
For loans made, eight cases have variable $X_{34}$, balances/amount, $\leq 10\%$ and for turndown nine cases out of seventeen. Of the eight loans with $X_{34} \leq 10\%$, seven out of eight were new customers. Of the nine turndowns with $X_{34} \leq 10\%$, eight were new customers. It appears that the lender made a decision as to whether the potential customer relationship would be a good one for the bank. In those cases, where the lender believed the firm had good potential, loans were granted.

**Costs of Misclassification - Multivariate Normal Case**

The procedure discussed earlier is followed when the a priori probabilities of a case being from population one or two is known. The solution which minimizes the maximum loss, the minimax solution, is given for properly selected $P$ functions by:

$$q_2 C(1 | 2) P(1 | 2) = C(2 | 1) P(2 | 1) q_1$$

Without loss of generality, the quantity $C(1 | 2)$ can be taken as unity and the expression becomes, where $q_2 = 1/3$ and $q_1 = 2/3$ -

$$1/3 \cdot P(1 | 2) = C(2 | 1) P(2 | 1) \cdot 2/3$$

or

$$P(1 | 2) = 2 \cdot C(2 | 1) P(2 | 1)$$

For normal distributions, the probabilities, expressed as integrals in the following expression, appear as -

$$\int_{\frac{c+a/2}{a}}^{\infty} N(0,1) = 2 \cdot C(2 | 1) \int_{-\infty}^{\frac{c-a/2}{a}} N(0,1)$$

and $c$ is selected so that this expression holds. The mean of group one is $a/2$ and of group two, $-a/2$. The variance is taken to be $a$, the variance of the discriminant function (3, page 135).
Table 3 shows calculations for various values of $C(2 \mid 1)$ when $C(1 \mid 2) = 1$. The variables for the group of loans are the mean, $u_1 = .6$, and the variance $\sigma_1^2 = .5$. For the group of turndowns $u_2 = .1$ and $\sigma_2^2 = .5$. The variable is $X_{34}$, balances/amount.

<table>
<thead>
<tr>
<th>$c$</th>
<th>$C(1 \mid 2)$</th>
<th>$C(2 \mid 1)$</th>
<th>$P(1 \mid 2)$</th>
<th>$P(2 \mid 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.10</td>
<td>1</td>
<td>1</td>
<td>.50</td>
<td>.25</td>
</tr>
<tr>
<td>.15</td>
<td>1</td>
<td>3/4</td>
<td>.47</td>
<td>.31</td>
</tr>
<tr>
<td>.35</td>
<td>1</td>
<td>1/2</td>
<td>.31</td>
<td>.31</td>
</tr>
<tr>
<td>.58</td>
<td>1</td>
<td>1/4</td>
<td>.25</td>
<td>.50</td>
</tr>
<tr>
<td>.82</td>
<td>1</td>
<td>1/8</td>
<td>.15</td>
<td>.60</td>
</tr>
</tbody>
</table>

The curves in Figure 3 show the minimax relation of the probabilities of misclassification for various ratios of the cost functions. For example, if $C(2 \mid 1) = 1/2$, given that $C(1 \mid 2) = 1$, the probabilities of misclassification are equal at .31. This indicates that, for example, if the cost of classifying a loan as a turndown is $100$, and the loss of classifying a turndown as a loan is $200$, the point $c = .15$ classifies cases for equal Type I and Type II errors of .31.

This result will now be compared to the exponential case.
Costs of Misclassification - Exponential Case

For the exponential distribution, the following integrals express equality of errors of misclassification:

\[
\int_c^\infty \exp(\theta_2=0.1) = 2 \cdot C(2 \mid 1) \cdot \int_0^c \exp(\theta_1=0.6)
\]

and \(c\) is selected so this expression holds. For the exponential distribution, \(c\) is a value on the abscissa.

Table 4 shows the values for various probabilities given \(C(2 \mid 1)\) and \(C(1 \mid 2) = 1\). The calculations are for a single variable \(X_{34}\), balances/amount.

Table 4

<table>
<thead>
<tr>
<th>(c)</th>
<th>(C(1 \mid 2))</th>
<th>(C(2 \mid 1))</th>
<th>(P(1 \mid 2))</th>
<th>(P(2 \mid 1))</th>
</tr>
</thead>
<tbody>
<tr>
<td>.11</td>
<td>1</td>
<td>1</td>
<td>.33</td>
<td>.17</td>
</tr>
<tr>
<td>.13</td>
<td>1</td>
<td>3/4</td>
<td>.28</td>
<td>.19</td>
</tr>
<tr>
<td>.15</td>
<td>1</td>
<td>1/2</td>
<td>.22</td>
<td>.22</td>
</tr>
<tr>
<td>.20</td>
<td>1</td>
<td>1/4</td>
<td>.14</td>
<td>.28</td>
</tr>
<tr>
<td>.25</td>
<td>1</td>
<td>1/8</td>
<td>.08</td>
<td>.33</td>
</tr>
</tbody>
</table>
The curves in Figure 4 show the minimax relation of the probability of misclassification for the exponential distribution. If $C(2 \mid 1) = 1/2$, given that $C(1 \mid 2) = 1$, the probabilities of misclassification are equal at .22, where $c = .15$.

![Figure 4. Exponential Minimax Cost Curves](image)

By comparing $P(1 \mid 2)$ and $P(2 \mid 1)$ of Tables 3 and 4, it is apparent that the probabilities of misclassification for the exponential distribution are consistently less than those for the normal approximation. The results are summarized in Table 5.

<table>
<thead>
<tr>
<th>C(1 \mid 2)</th>
<th>C(2 \mid 1)</th>
<th>Normal P(1 \mid 2)</th>
<th>Normal P(2 \mid 1)</th>
<th>Exponential P(1 \mid 2)</th>
<th>Exponential P(2 \mid 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>.50</td>
<td>.25</td>
<td>.33</td>
<td>.17</td>
</tr>
<tr>
<td>1</td>
<td>3/4</td>
<td>.47</td>
<td>.31</td>
<td>.28</td>
<td>.19</td>
</tr>
<tr>
<td>1</td>
<td>1/2</td>
<td>.31</td>
<td>.31</td>
<td>.22</td>
<td>.22</td>
</tr>
<tr>
<td>1</td>
<td>1/4</td>
<td>.25</td>
<td>.50</td>
<td>.14</td>
<td>.28</td>
</tr>
<tr>
<td>1</td>
<td>1/8</td>
<td>.15</td>
<td>.60</td>
<td>.08</td>
<td>.33</td>
</tr>
</tbody>
</table>
The point of interest is the adequacy of the normal approximation to the exponential discriminant function, especially when costs of misclassification differ greatly. Assume, for example, that the costs of classifying a turndown as a loan, $C(1/2)$, are eight times greater than $C(2 | 1)$, the cost of classifying a loan as a turndown. For the exponential case, from Table 4, $C = .25$, $P(1 | 2) = .08$ and $P(2 | 1) = .33$. When $c = .82$, the normal approximation with mean and variance as in Table 3 gives a value of .15 for $P(1 | 2)$ and .60 for $P(2 | 1)$. When the costs of misclassification are very different, the normal approximation leads one to believe that errors of classification are twice as large as when the exponential distribution is used.

In discriminant analysis, the variables are generally assumed to belong to multivariate normal populations. Gilbert (19) has shown that functions with dichotomous variables - i.e., 0 and 1 - can be used efficiently for discriminant analysis.

The variables upon which this analysis is based are believed to be exponential. A search of the literature did not reveal work on discriminant analysis with exponential variables. The literature on "life testing" has concentrated on studies of exponential distributions. Specific applications of discriminant analysis with exponential variables was not found in this literature. The function which represents exponential discriminant analysis is fairly simple in form and was developed here for the case of one and two independent exponential variables.

Several articles by John (26, 27) have used the method of moments and maximum likelihood to identify the population of origin of observations from two groups. The parameters which are estimated are $N_1$ and $N_2$, the numbers in the two groups, $\mu_1$ and $\mu_2$, the means of the populations, and $\sigma^2$, the variance common to the groups.

Validation of the results of the data analysis from a small sample is the next important objective. A new approach to small sample validation is discussed by Hutchins (24) in an unpublished dissertation at The Ohio State University. The approach, called the "Selection Battery" approach, uses multiple regression runs to select one or two significant variables on each run and then a combination of these variables from independent trials is used as a better regression estimate.
In the analysis reported here, the results of several different runs indicate that the variables balances/amount and balances each appear as the most significant variables in separate runs. This result, together with a large sample, are used to validate the results of the small sample.

One final point in regard to validation of the classification results deals with the problem of misclassification in the original sample. If the original sample contained misclassifications, this may affect the variables which enter the discriminant function. Research reported by Lachenbruch (28) on the impact of incorrect classifications on the multivariate normal discriminant function indicates that the utility of the discriminant function may not be seriously affected. The distribution of the discriminant function is shown to be normal with means of two groups which are closer together and has a smaller variance.

**Results of Analysis on the Larger Sample**

The pilot study was successful in finding a single variable from the group of original variables which did the best job of approximating the decisions made by the lender. The most important variable was determined to be \( X_{34} \), balances/amount. The pilot study also gave evidence that the underlying distributions are other than normal. The variable \( X_{34} \), balances/amount, was found to be exponentially distributed.

On the basis of these results, additional data were collected on 200 loan cases and 60 turndown cases. A frequency distribution on both groups of data shows the distributions to be exponential.

The classification of cases on the basis of the variable \( X_{34} \), using parameters for the exponential discriminant function that are identical to the pilot sample, resulted in the classification matrix shown on the following page.
Of the loans made, 66% are correctly classed as loans and 80% of the turndowns are classed as turndowns.

For the sample of 200 loans and 60 turndowns, half the cases may be kept as a "hold out" sample and the critical values calculated for the remaining cases. The results are then used on the "hold out" sample to validate that sample.

For a sample of 100 loans and 30 turndowns, \( \theta_{x1} = .51 \), \( \theta_{x2} = .095 \), and \( x = .2 \). The classification results are:

<table>
<thead>
<tr>
<th>Lender</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>132</td>
</tr>
<tr>
<td>T</td>
<td>68</td>
</tr>
</tbody>
</table>

The discriminant function is the same as in the pilot sample. Sixty-eight percent of loans are classed correctly and 83% of turndowns.

Using the values for \( \theta \) and \( x \) on the hold-out sample results in a classification matrix as shown on the following page.
Sixty-four percent of loans are classified correctly and 80% of the turndowns.

An analysis of variance was performed in order to test the hypothesis $H_2$ under research question one of Chapter III. This hypothesis states that there are no significant differences between lending division (industries) with respect to the key variable $X_{34}$, balances/amount. The calculations are summarized below:

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Divisions</td>
<td>6</td>
<td>1.04</td>
<td>.17</td>
</tr>
<tr>
<td>Error</td>
<td>57</td>
<td>7.94</td>
<td>.14</td>
</tr>
<tr>
<td>Total</td>
<td>63</td>
<td>8.98</td>
<td></td>
</tr>
</tbody>
</table>

The analysis of variance model used to test hypothesis $H_2$ is a single-factor model as follows:

$$X_{ij} = \mu + T_j + \epsilon_{ij}$$
$X_{ij}$ represents the average of the sum of each of three random observations, $u$ is a common effect for variable $X_{34}$ for the whole experiment, $T_j$ is the effect for the $j$th lending division, and $(-y_i)$ represents the random error of the $ith$ observation on the $j$th division. The null hypothesis tested here is that

$$H_0: T_j = 0 \text{ for all } j$$

The results of the analysis indicated that there was no significant difference among divisions with respect to the variable $X_{34}$. Hypothesis $H_2$ is accepted. This result indicates that while there are probably important differences in the way lenders analyze loans, in terms of key variables where management policy has been made explicit, the lenders are consistent in implementing that policy. The analysis which combines the divisions and treats them together is thus validated, at least for the variable $X_{34}$.

**Other Analyses**

Other analyses were performed on the data using a number of financial ratios as well as financial statement variables. Appendix E summarizes the results obtained from the use of several ratios to predict the lending decision.

The analyses discussed in this chapter are proposed as possible methods for examining quantitative data from an actual decision process, a lending decision process, to aid in a study of that process and to formulate a model of that process. Discriminant analysis was used in this chapter as an analysis tool and in the next chapter it will be used as a modeling tool. The next chapter discusses a quantitative decision model, a prescriptive model, to aid the decision maker in the lending decision.
CHAPTER V
THE COMMERCIAL LOAN CREDIT DECISION MODEL

It is apparent from the analysis in the preceding chapter that a single variable may be useful in differentiating between applications that should be loans and those that should be turndowns. These useful variables have been found not to be the historically standard financial ratios which are commonly accepted as measures of creditworthiness.

The discussion in Chapter II shows the lending decision process to be composed of several phases of loan application analysis. For purposes of this modeling effort, the key phase is assured to be credit analysis.

In the existing lending system as it functioned prior to the advent of "tight money" in 1968, the credit decision was also the allocation decision, much as in consumer lending. That is, generally, a favorable credit rating resulted in the granting of a loan, since ample funds were available. Since the advent of tight money, there is a distinctly separate fund allocation step in the lending process.

When and if economic conditions return to a state of ample funds for commercial lending it is anticipated that benefits still would accrue to a bank which followed some objective means of allocating funds.

Some of the reasons for attempting to rationalize or model the lending process are to improve:

• Profit
• The quality of the loan portfolio
• Responsiveness of the lending process to management policy
• Evaluation of individual loan officer performance
• Loan division performance
• The competitive position of the bank
• The operation of the loan review or audit function
• Provide a training vehicle for new lenders.
A bank is a business which has profit as its foremost objective and has secondary objectives of customer service, community service, employee satisfaction, and the like. Improving the quality of the loan portfolio implies improving the quality of the individual loans which make up that portfolio. The quality of a loan is related primarily to its creditworthiness.

A model of the lending process would presumably allow management to monitor the workings of that process to see that policy directives are carried out. The performance of the individual lender should determine, in some fashion, the rewards that accrue to a satisfactory performance of his job. The performance of the individual lenders makes up the performance of a division as reflected in the number of bad loans, the deposit balances of a division, and the profit of a division, and that performance should be measurable in terms that would make the performance comparable to that of other divisions. The competitive position of the bank should be enhanced by the ability to rationalize the loan function and measure its performance.

The loan review, an audit function, would be helped significantly by a rational and consistent method of considering the tremendous volume of loans which must be monitored for trouble spots. The ability to replicate past decisions would be a significant advantage to loan review and a model which explicitly states the workings of the loan function would assist in training new lenders. A model would make explicit the variables that must be considered in the lending process. It would also focus on the variables which past experience has proven to be the most important and would suggest the variables which in reality should be used.

The Decision Process

The commercial loan decision process will be modeled as a process which has as its significant phase the analysis of creditworthiness of an application. The first part of the discussion will suggest a model which represents the way in which a lender processes a loan application under the present system. This model considers the variables used in processing an application as being made up of quantitative and qualitative sets. The second part of the discussion will propose that a screening process be used to analyze the quantitative portion of the critical variables. The third part will suggest criteria for use of the proposed model. A diagram of the lending process is shown in Figure 5, following this page.
Figure 5. Present and Proposed Systems.
The Present Credit Analysis System

The first approach to a lending decision model will attempt to duplicate the decision process of the lender, that is, to make a decision to loan or turn down an application. The decision process of the lender is assumed to be analogous to the process represented by discriminant analysis. It is assumed that the lender possesses a critical value $K$ based on his analysis of key lending variables, which separates a group of applications into loan or turndown categories. $K$ is the lender's critical discriminant value, analogous to the discriminant value discussed in the previous chapter.

In analyzing an application, the lender gathers data on a vector of variables, $M$, which help him make a lending decision. $M$ is composed of two parts, $m$, and $\overline{m}$, i.e., $M = (m, \overline{m})$. The variables in $m$ are the quantitative, measurable variables used to classify an application (e.g., balances/amount) and the variables in $\overline{m}$ are quantitative variables not included in $m$, as well as other factors of a qualitative nature used by the lender to make that classification (e.g., customer reputation, geographical location, industry comparisons, etc.).

It is assumed that the lender possesses the critical value $K$ which enables him to discriminate between loans and turndowns. The analysis of the variables in $M$ results in the lender's assigning a discriminant value $D$ to the particular application. To be accepted, an application must have:

\[ D > K \]

$D$ is therefore a function of $M$

\[ D = f(M) \]

That is, the discriminant value $D$ is determined from analysis of the variables in $M$. Furthermore, $f(M)$ is a function of those variables $m$ which are quantitative and measured and those variables $\overline{m}$ which are qualitative and, for the purposes of this model, are not measured.

Thus, the discriminant value $D$ is a function of $m$ and $\overline{m}$ expressed by

\[ D = f(m, \overline{m}) \]
Under the present decision process, the lender may, for some cases, examine only the variables in $m$ and accept or reject the application based on this review. For other cases, the lender may examine only the variables in $m$. This occurs, for example, in reviewing a loan where the customer has a good relationship and the lender finds it necessary to review the variables in $m$ before a decision is made.

The following diagram illustrates the nature of the lending problem in the present system. This diagram includes an allowance for default in the lender's decision to make the loan or turn down applications.

```
<table>
<thead>
<tr>
<th>Loan Potential</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Good Loan</td>
</tr>
<tr>
<td>L=1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Default</td>
</tr>
<tr>
<td>L=0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lender's Discriminant Value</th>
<th>Good Loan</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D &gt; K$</td>
<td>$n_1$</td>
<td>$n_3$</td>
</tr>
<tr>
<td>$D \leq K$</td>
<td>$n_2$</td>
<td>$n_4$</td>
</tr>
</tbody>
</table>
```

$L = 1$ and $L = 0$ are results of decisions by the loan officer to make the loan or turn down an application. A good loan is one whose terms are fulfilled; a default is one whose terms are not met - e.g., late pay, bankruptcy. $D$ represents the discriminant value which was assigned by the lender in his own evaluation of an application. $D > K$ represents a case where the discriminant value exceeds the critical value and the application was accepted as a loan, and $D \leq K$ represents a case that was a turn down. The $n_1$ are the numbers of cases that fall into each classification. $n_2$ and $n_3$ are the errors of classification actually made by the lender. The error in $n_3$ is the error which may lead to default on a loan. The error in $n_3$ is the error which may lead to default on a loan. The error in $n_2$ is the error of refusing a potentially good loan.

The probability of making the error in $n_2$, which is refusing a potentially good loan, is expressed by:

$$P(L=1 \mid D \leq K)$$

and for the error in $n_3$, the probability of default is:

$$P(L=0 \mid D > K)$$
The probability of correctly classifying an application as a loan in \( n_1 \) is expressed as

\[
P(L=1 \mid D>K)
\]

One of the objectives of the present system is minimizing the errors in \( n_2 \) and \( n_3 \).

This discussion outlines the theoretical way in which the decision process of the lender may be viewed and is concerned with the decision to grant or not to grant a loan.

**The Proposed Credit Analysis System**

The data actually gathered in Chapter IV were gathered after the lending decision had already been made. The model which illustrates the situation represented by the data is to consider first that the lender made a decision to loan or turn down; then discriminant analysis is used to try to match that decision.

The results of the pilot study of a number of variables which might be used to discriminate between loans and turn-downs indicated that the classification of cases based on the single variable balances/amount were reasonably successful. The data collection and analysis of an additional 260 cases and classification in terms of a single variable, balances/amount, were also successful.

The model which is desired is one which first analyzes the applications using a screening calculation based on discriminant analysis and then submits some portion of the output from the screen to the lender. The underlying assumption here is that of the variables in \( M \), that portion which are quantitative and measurable will be analyzed to yield a measure of those variables which will determine the screening decision. Furthermore, it is assumed that the variables in \( m \) are not measured. The assumption here is that some linear combination of the variables in \( m \) represented by

\[
d = \sum a_i m_i
\]

may be used in classifying an application. The results of the analysis of the linear combination is compared to a critical value \( k \) as determined by discriminant analysis. For an application to be accepted, \( d \) must be greater than \( k \), i.e.,

\[
d > k
\]
Furthermore, the hypothesis was tested in Chapter IV that the linear combination of variables was the same for all lending divisions.

The proposed model of the lending decision process is then first to screen applications, which can be based on minimizing the cost of misclassification, and then to submit some portion of those applications to the lender for his final analysis and decision.

The Lender's Role in the Decision Process

The remainder of this discussion focuses on ways in which the output from the screening process could be presented to the lender. The credit analysis could be set up in a number of ways to control the flow of applications through the system. Four of these are:

1. Give the lender applications accepted by the screen.
2. Give the lender applications rejected by the screen.
3. Give the lender all applications after they pass through the screen.
4. Give the lender some selected portion of the applications processed by the screen.

If the lender is given only the cases which are rated as loans by the screen, the applicants which rank low on critical variables such as balances/amount would not receive loans. New customers who might have a zero deposit balance would never be reviewed by the lender and would not receive loans. Cases which appear in the Type I error would not be reviewed when, in fact, some should be.

When the lender reviews only the turndowns, a certain number of marginal loans would be made which perhaps should not be. That is, the Type II error would not be reviewed by the lender and these applicants would receive loans.

If the lender is given all cases reviewed by the screen this would simply be more information for the lender and the final decision would be his. This strategy would not reduce the workload of the lender. One difficulty with this approach is that the lender might be inclined to accept the results of discriminant analysis at face value and might not be motivated to innovate at all.

The last alternative is to select a portion of cases analyzed by the screen to pass to the lender for review. The selected cases would first be classified by the screen. This alternative is discussed in the following paragraphs.
It is desirable to adjust the screen values so that the lender receives an acceptable workload. The lender would concentrate on the marginal cases; at the same time, the errors of classification would be minimized. The contingency table following illustrates the problem.

<table>
<thead>
<tr>
<th>Lender Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>( D &gt; K )</td>
</tr>
<tr>
<td>( d &gt; k )</td>
</tr>
<tr>
<td>( Type \ II )</td>
</tr>
<tr>
<td>( d \leq k )</td>
</tr>
</tbody>
</table>

Under "computer screen" \( d > k \) and \( d \leq k \) are loans and turndowns by a program and under "Lender Classification" \( D > K \) and \( D \leq K \) are loans and turndowns by the lender. The \( r_i, i=1,...,4 \) are the numbers of cases and the Type I and Type II are differences of classification.

The examination by the computer screen results in classification distributions as shown in the diagram below.

The distribution of turndowns includes those designated as \( r_2 \) and \( r_4 \) in the previous contingency table. The distribution of cases designated as "loans" includes the cases in \( r_1 \) and \( r_3 \). Subsequent classification by the lender results in some classifications which are the same as the screen and some which differ. As was illustrated in Chapter IV, most of the differences in classification occur at the area of overlap between the two groups of cases.
The applications are first classified by the screening process and then by the lender. The classification by the screen is illustrated by the following diagram:

The curve shown as (a) is the distribution of turn downs based on the screen alone. The curve in (b) is the distribution of loans based on the screen alone. The curve designated by (c) is the combined distributions of both loans and turn downs illustrating the way in which the critical value \( k \) separates the two groups of cases. The classification is based on only the variables in \( m \) as analyzed by the screen.
The next step in the system is to submit cases to the lender. The lender uses his knowledge of the variables in $m$ to make the final decision to make the loan or to turn down the application.

The distributions in (d) and (e) are again classified by the screen without knowledge of $m$. The cases in the area shown as $m$ are correctly classified on the basis of knowledge $m$. The distribution in (f) illustrates as $w$ the cases with which the lender will be concerned.

The "window," $w$, through which the lender views the cases submitted to him may be adjusted to achieve a trade-off between error cost and processing cost. The width of the window is $\Delta_1 + \Delta_2$. The "window" is shown in the diagram on the following page.
The "error" is the error of classification remaining after the lender has made a decision. In the case of applications classed as turndowns which should have been approved for loans the error occurs in refusing a loan to a potentially good customer. With a turndown case approved as a loan, the error is in lending to a customer who eventually defaults. Orgler (35) discusses the difficulty of estimating the cost of eliminating a bad loan from consideration as well as the evaluation costs.

Adjustment of the width of the window results in cost curves as shown in the following diagram:

For a window of width zero, the lender would not review any cases and the only costs would be costs of errors of classification made by the screen using the computer. For a window of width $\infty$, the lender would review all cases and the cost would be the cost of processing all cases.
Considerations of Exponential Distributions

The variables which determine the loans reviewed by the screen and by the lender after the screening process may be viewed as coming from exponential distributions. After classification by the screen, the loans and turndowns are as follows:

The curves in (g) and (h) are classified by the screening variables only. The curves in (i) show the classification after the lender considers the variables in \( m \). The situation existing after analysis by the lender is as shown in the following diagram. The window \( w \) includes the cases having a zero value for the variables. For example, the lender would review cases having a zero ratio of balances/amount.
The shaded portion includes the cases reviewed by the lender after the screening. Those outside the shaded area are cases which are not reviewed by the lender.

The model proposed here is thus an exponential discriminant analysis model with a single variable, balances/amount. This model of the commercial lending decision process is the first evidence to be presented which is based on empirical analysis which indicates that the lending decision is perhaps not based on the commonly accepted variables but rather is determined by management emphasis on balances in determining loan recipients.

Rao (40) has discussed the use of a grading concept to classify cases as good, bad, and marginal. Orgler (35) develops an expression which is intended to minimize the evaluation cost for processing good and bad loans. Orgler notes that cost information is not readily available and the cutoff values which determine the size of the window should be determined subjectively or by use of rough cost estimates. Cohen and Hammer (9) develop an expression to maximize expected profits from classifying good and bad loans where the amount of loans extended depends on the cutoff value for the window.

The next chapter summarizes the important results of this research and suggests other research approaches which might prove fruitful.
CHAPTER VI

SUMMARY AND CONCLUSIONS

The analysis of the data as discussed in Chapter IV and the model proposed in Chapter V indicate that it may be possible to support the decisions made by the lender. The remaining tasks are to propose ways in which this support could be implemented and to suggest other approaches to studies of the decision process.

The discriminant analysis model is proposed, not to replace the experience and judgment of the lender, but to augment that judgment. Very good reliability for predicting good loans is obtained from a model based on a single variable, e.g., balances/amount. If an affirmative answer is received when this model is used, the amount of data collection and analysis can be reduced. If a negative evaluation is made, the assistance given by the model regarding the applicant may be simply to rank a case by a weight which indicates its relative position with respect to other loans and then turn it over to the lender for study.

This model represents a departure from the commonly accepted approaches to the credit decision which require an analysis of financial ratios. Financial ratios are presently analyzed, but the results may not be used in the way they are generally believed to be used. Other variables which are emphasized by management appear to outweigh in importance the standard financial ratios.

The discussion in the remainder of this chapter will review the validation of the hypotheses posed in Chapter III, present suggestions for implementation of the new proposed credit analysis model, and suggest further avenues for research related to the loan decision problem.
Validation of Hypotheses

The hypotheses posed in Chapter III are reviewed and discussed in light of the analysis of the sample data collected and analyzed, as reported in Chapter IV.

\[ H_1: \text{Management policy variables will be significant in classifying applications as loans or turndowns} \]

The first research question of Chapter III dealt with the feasibility of using guidelines of management policy to distinguish between loans and turndowns. This question is answered in the discriminant analysis section. In the stepwise discriminant analysis, the first variable chosen as yielding the greatest discrimination between groups was deposit balances/amount. Management emphasizes the balances requirement at the credit analysis and fund allocation stages.

The first hypothesis is validated by the results of the pilot study and the larger sample. The policy variable, balances/amount, is found to be the most significant in the credit decision.

\[ H_2: \text{There are no significant differences between lending divisions in the key policy variables}. \]

The second hypothesis was that no significant differences would be found between divisions on the key variables such as \( X_{34} \), balances/amount. The analysis of variance of Chapter IV shows that this hypothesis is accepted. This particular variable is consistent across divisions, indicating uniformity of policy enforcement regarding balances.

\[ H_3: \text{Policy variables will be more significant in classifying cases as loans or turndowns than credit variables}. \]

The second research question dealt with the comparative importance of policy variables and credit variables in distinguishing between loans and turndowns. That question is also answered by the results of the discriminant analysis.
Hypothesis 3 is accepted on the basis of analysis of the pilot study data. The discriminant function does a good job of classifying loans and turndowns solely on the basis of policy variables.

A number of cases, both loans and turndowns, cannot be correctly classified by discriminant analysis. In the case of new customers, the variable balances/amount cannot be used to screen applications because a new customer would not have a deposit balance. This points up the importance of the other credit variables, both quantitative ones and those involving experience and judgment. It is possible that two lenders would not agree on borderline cases; one officer might grant the loan and another turn it down. The discriminant model is less accurate at classifying cases the same way as the lender when they are on the borderline between loan and turndown. This tends to verify that the solid credits are probably easy to decide on. It is the marginal credits that require the experience and judgment.

This is still another justification for development of decision aids to relieve the lender of the time-consuming detailed analysis of credit data. A certain amount of this work is necessary to train a loan officer. One of the primary objectives is to relieve the lender of some of the unnecessary detail of the job and let the rewards for the job, including financial rewards, be based on exercising judgment and calling on experience.

This result - that is, the significance of the balances in the credit decision - suggests that perhaps much of the work which stresses financial ratio analysis may be somewhat misdirected. Loans are made for many reasons other than those commonly accepted as the basis for a loan. Not enough research has been done to validate empirically many of the commonly accepted rules for credit and loan decision making.

In addition to hypotheses relating to the loan decision, other evidence is available concerning the credit evaluation situation. Table 7 in Appendix E shows the results of a discriminant analysis on some ratios commonly used in credit analysis. The signs of these variables differ from what is believed to be necessary for a sound credit. For example, the current ratio is shown with a negative sign rather than a positive sign which would be found with an acceptable 2:1 ratio. The ratio net worth/total liabilities also has a
minus sign. This would also be expected to be plus. The signs of these variables are evidence that perhaps the loans were granted because the borrower needed money and not because his financial ratios met commonly accepted standards.

**Suggestions for Implementation**

Implementation of any automated credit checking aid program would have to be carried out under carefully controlled and monitored conditions. The possibilities for incorrect or uninformed use of a credit evaluation aid are endless. The first step in implementation of a decision aid program could be improvement of the data retrieval and analysis procedures currently available to the lender. This first step could include retrieval of loan ledger information, deposit balances, and other bank relationship data. Most effort toward implementing bank management information subsystems in the commercial loan area has been devoted to data retrieval.

The second step in implementation would be to use an evaluation model in a loan auditing area such as loan review. Orgler (35) has proposed that a credit scoring model which discriminates between good loans and those that have been criticized by a bank examiner be implemented first in the loan review area. A loan review department is generally made up of senior lenders with a great deal of experience. After a loan is made, it could be analyzed to compare the results of the analysis with the actual decision. When results differed, the loan review function could determine if the differences were due to errors in the model, deviations from management policy by the lender or deviations from sound credit practices. Monitoring of the construction of a model would help to build confidence in that model as well as provide much needed expert advice on lending practices.

The next step in implementation could be for management of a particular division to use the model to monitor individual decisions made by loan officers in a division. This would further help to refine the model and allow division-level people to gain confidence in the model.
The last step in implementation would be the use of the model by a select group of lenders who were willing to match their judgment against the results of a model. This step would determine if a credit analysis model would be useful in actual lending decisions.

Another possibility for use of such a credit analysis model would be as a training tool for loan division lending officer trainees. Such a model would help trainees to gain insight into the important variables and their influence on credit situations.

**Suggestions for Further Research**

The results of the analysis of the sample of data suggest that a great many applications classified as either approved loans or turn downs could readily have been classified differently if some other lender had been responsible for making the decision. This is inevitable where human judgment is involved. Nevertheless, experimentation should be conducted to verify the premise that much lending decision making is inconsistent, in a statistical sense.

A study which would perform similar analyses within divisions could be undertaken to determine variables which are more important in a particular industry or geographical area. The first data collection and analysis effort reported in this study was carried out among divisions and among lenders, and covered short-term lending and long-term lending, loans at prime rate and not at prime, and lenders having different amounts of experience.

Further research should be carried out on actual decisions to lend or turn down, with subsequent analysis of those loans which were made and then were defaulted for some reason. Research which followed up on the default would require a long period of time to complete. This is perhaps one reason it has not been done.

Further studies should include qualitative factors such as the attitude of an individual lender toward risk in relation to the loans he makes. Some lenders find a challenge in making difficult credits out of easy ones.
Other qualitative factors such as character judgments of customer management, financial expertise, and depth of management should be included in future analysis of the lending process.

This study afforded a unique opportunity to study the credit decision process during a period of tight money. Similar studies could be conducted when and if funds for lending again become plentiful. The results of these studies could be compared, to draw conclusions concerning the impact of tight money. It may be that the results of this study are distorted because of special economic situations prevailing in 1968 - 1970.

Corroboration of the results of this study can be inferred from an editorial in the March, 1971 issue of American Banker. The magazine quoted a bank executive summarizing the aftermath of a default on several loans to a large company in which a number of banks suffered losses. He said,

\[ \text{We will pay far less attention to balances in our future lending decisions and far more to the fundamentals of a loan.} \]

As another avenue for possible research, it should be possible to construct a "policy" discriminant function, that is, a function which would screen loans on the basis of policy guidelines laid down by management rather than on the basis of averages of loans already made. Such a function would permit a loan under consideration to be compared to major management policy directives such as those relating to balances, security, maturity, loan type, and purpose.

Another potential area for exploration would be the use of sensitivity analysis applied to individual coefficients of the discriminant function. This analysis would indicate which variables in the discriminant function caused the screening process to place a case in a group as a loan or another group as a turndown. For example, if it were possible to indicate that an application had been classified as a turndown but an increase in deposit balances would shift it to the loan category, this would be useful information.
Final Analysis

The successful application of lending decision aids rests on selection of the correct analysis tools and use of the correct variables to classify cases.

One conclusion which might be drawn from this study is that a great deal of work would be required to gather data and analyze it before any analytical technique could be applied to a task as subjective as making loans. It also bears emphasizing that other judgmental tasks such as inventory control in manufacturing and gasoline blending in the petroleum industry were based primarily on judgment and experience until they yielded to analytical techniques. It is somewhat surprising that, in spite of the difficulties of data collection, more effort has not been devoted to decision models in the loan area in a key industry such as banking. Much work and a great deal of challenge lies ahead.

The stratified sample taken for this analysis proved to be useful because of the wealth of information gained from that analysis. The sample permitted the gathering of information about the population which might have required a much larger sample and a much larger data collection effort if it had been a random sample of loan cases. A random sample from the population of applications would have yielded a larger number of cases that were obvious loans or turn-downs. The stratified sample concentrated on the "interesting" credit cases.

Edmister (15) points out the sensitivity of the results of past studies to both:

- The data
- The event or result measured.

It is apparent that a technique such as discriminant analysis will select from a set of variables a subset which best distinguishes between several populations. The results are mathematically correct, but there is little guarantee that the same results will be achieved with another subset or sample. This is a reason for emphasis on sample validation. More research on larger samples is needed to explore these problems.
It is hoped that the work begun here can be pursued as a fruitful avenue for research. Many interesting and important aspects of decision processes remain to be explored. The author hopes to contribute to future research in this area.
APPENDIX A

LIST OF VARIABLES

Variables used in the pilot study are listed and briefly explained in this appendix.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$ Officer Experience</td>
<td>Years of lending experience, credit department or discount department experience.</td>
</tr>
<tr>
<td>$X_2$ Interest Rate</td>
<td>Rates are adjusted to maintain a relationship with prime rate changes over times.</td>
</tr>
<tr>
<td>$X_3$ Loan Maturity</td>
<td>Period of loan, in months</td>
</tr>
<tr>
<td>$X_4$ Amount</td>
<td>In thousands of dollars</td>
</tr>
<tr>
<td>$X_5$ Security</td>
<td>Collateral value of loan in dollars</td>
</tr>
<tr>
<td>$X_6$ Commitment</td>
<td>Amount of funds available under formal written agreement</td>
</tr>
<tr>
<td>$X_7$ Balances - 1969</td>
<td>Demand deposit balances kept on deposit in checking accounts</td>
</tr>
<tr>
<td>$X_8$ Balances - 1968</td>
<td>Demand deposit balances kept on deposit in checking accounts</td>
</tr>
<tr>
<td>$X_9$ Clean-Up</td>
<td>0 - Customer did not reduce his account balance to zero in last 12 months</td>
</tr>
<tr>
<td></td>
<td>1 - Account was cleaned up</td>
</tr>
<tr>
<td>$X_{10}$ Year-Ago Borrowing</td>
<td>Outstanding borrowing 12 months previous to the present loan request</td>
</tr>
<tr>
<td>$X_{11}$ Outstanding Borrowing</td>
<td>Dollar amount outstanding at the time of this request</td>
</tr>
<tr>
<td>$X_{12}$ Years a Customer</td>
<td>Years a customer has had a lending division relationship</td>
</tr>
</tbody>
</table>
The following are the customer balance sheet and income statement variables:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X_{13} ) Cash</td>
<td>Cash plus marketable securities</td>
</tr>
<tr>
<td>( X_{14} ) Accounts Receivable</td>
<td>Average for year</td>
</tr>
<tr>
<td>( X_{15} ) Inventory</td>
<td>Average inventory for last full year prior to loan request</td>
</tr>
<tr>
<td>( X_{16} ) Current Assets</td>
<td>Includes cash, plus accounts receivable, plus inventory</td>
</tr>
<tr>
<td>( X_{17} ) Fixed Assets</td>
<td>Fixed assets minus depreciation, depletion, amortization</td>
</tr>
<tr>
<td>( X_{18} ) Depreciation, depletion, amortization</td>
<td>Self-explanatory</td>
</tr>
<tr>
<td>( X_{19} ) Total Assets</td>
<td>Total current and fixed assets</td>
</tr>
<tr>
<td>( X_{20} ) Current Liabilities</td>
<td>Includes short-term debt and other liabilities</td>
</tr>
<tr>
<td>( X_{21} ) Short-Term Debt</td>
<td>Debt due in less than one year</td>
</tr>
<tr>
<td>( X_{22} ) Long-Term Debt</td>
<td>Debt due more than one year from date of financial statement</td>
</tr>
<tr>
<td>( X_{23} ) Preferred Stock</td>
<td>Total assets minus total liabilities</td>
</tr>
<tr>
<td>( X_{24} ) Net Worth</td>
<td>Current assets minus current liabilities</td>
</tr>
<tr>
<td>( X_{25} ) Working Capital</td>
<td>Current plus long-term liabilities</td>
</tr>
<tr>
<td>( X_{26} ) Total Liabilities</td>
<td>Total revenue</td>
</tr>
<tr>
<td>( X_{27} ) Sales</td>
<td>Cost of sales or operations</td>
</tr>
<tr>
<td>( X_{28} ) Cost of Sales</td>
<td>Sales, less cost of goods sold less operating expenses, plus other incomes</td>
</tr>
<tr>
<td>( X_{29} ) Profits Before Taxes</td>
<td>Profits after federal income taxes</td>
</tr>
<tr>
<td>Variable</td>
<td>Explanation</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>$X_{31}$ Loan Type</td>
<td>Classification based on characteristics of the loan;</td>
</tr>
<tr>
<td></td>
<td>1 - seasonal</td>
</tr>
<tr>
<td></td>
<td>2 - term</td>
</tr>
<tr>
<td></td>
<td>3 - line of credit</td>
</tr>
<tr>
<td></td>
<td>4 - transaction for specific purpose</td>
</tr>
<tr>
<td></td>
<td>5 - revolving credit</td>
</tr>
<tr>
<td>$X_{32}$ Loan Purpose</td>
<td>Reason for application;</td>
</tr>
<tr>
<td></td>
<td>1 - accounts receivable</td>
</tr>
<tr>
<td></td>
<td>2 - plant</td>
</tr>
<tr>
<td></td>
<td>3 - other fixed assets</td>
</tr>
<tr>
<td></td>
<td>4 - working capital</td>
</tr>
<tr>
<td></td>
<td>5 - inventory</td>
</tr>
<tr>
<td></td>
<td>6 - refinancing</td>
</tr>
<tr>
<td></td>
<td>7 - acquisition or merger</td>
</tr>
<tr>
<td>$X_{33}$ Other Bank Relations</td>
<td>Other customer account relationships;</td>
</tr>
<tr>
<td></td>
<td>1 - venture capital company</td>
</tr>
<tr>
<td></td>
<td>2 - trust</td>
</tr>
<tr>
<td></td>
<td>3 - no other relationship</td>
</tr>
<tr>
<td></td>
<td>4 - international banking</td>
</tr>
<tr>
<td></td>
<td>5 - other lending divisions</td>
</tr>
<tr>
<td>$X_{34}$ Balances/Amount</td>
<td>Average monthly balances divided by average monthly loan amount for year prior to loan request</td>
</tr>
</tbody>
</table>

From the financial statement data, the following ratios are among those which may be calculated:

**Current Ratio (Current Assets/Current Liabilities)** - Cash plus marketable securities, plus inventory, plus net receivables, divided by notes and accounts payable, plus current maturity debt.

**Net Worth/Total Liabilities** - Total assets minus total liabilities divided by (total current liabilities plus long-term debt).

**Cost of Sales/Inventory** - Net income before taxes, divided by (total assets minus total liabilities).

**Profits/Net Worth** - Net income before taxes, divided by (total assets minus total liabilities).
Sales/Working Capital - Sales divided by (current assets minus current liabilities).

Sales/Net Worth - Sales divided by (total assets minus total liabilities).

Current Liabilities/Net Worth - Accounts and amounts payable, plus current debt, divided by (total assets minus total liabilities).

Profits/Total Liabilities - Net income divided by current liabilities and long-term assets.

Profits/Total Assets - Net income divided by total current plus long-term assets.

Working Capital/Total Assets - Current assets, minus current liabilities divided by total assets.

Operating Ratio - (Selling and Operating Expenses/Net Sales)

Quick Ratio - (Cash plus marketable securities plus net accounts receivable)/Current Liabilities

Cash Flow - Earnings before taxes, plus depreciation, plus amortization.
R. A. Fisher (17) introduced the method of discriminant analysis into statistics. Anderson (4) presents the mathematical foundation for discriminant analysis for known multivariate normal populations with equal covariance matrices, that is, for populations distributed $N(\mu_1, \Sigma)$ and $N(\mu_2, \Sigma)$ where $\mu_1$ and $\mu_2$ are the vector means of the populations and $\Sigma$ is the common variance-covariance matrix.

The basic multivariate normal density function is given by:

$$p_1(x) = \frac{1}{(2\pi)^{p/2}} \exp \left[ -\frac{1}{2} (x - \mu_1)' \Sigma^{-1} (x - \mu_1) \right]$$

The likelihood ratio of densities is given for two populations as:

$$\frac{p_1(x)}{p_2(x)} = \frac{\exp \left[ -\frac{1}{2} (x - \mu_1)' \Sigma^{-1} (x - \mu_1) \right]}{\exp \left[ -\frac{1}{2} (x - \mu_2)' \Sigma^{-1} (x - \mu_2) \right]} = \exp \left[ -\frac{1}{2} (x - \mu_1)' \Sigma^{-1} (x - \mu_1) - (x - \mu_2)' \Sigma^{-1} (x - \mu_2) \right]$$

The region of classification $R_1$ into population $\Pi_1$ is the set of $x$'s for which this expression is $\geq$ a suitably chosen $s$, a constant. Taking logarithms, expanding and rearranging, the expression becomes -

$$x' \Sigma^{-1} (\mu_1 - \mu_2) - 1/2 (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2) \geq \ln s$$

The first term, $x' \Sigma^{-1} (\mu_1 - \mu_2)$, is the discriminant function and the second term is simply the point midway between the means of the discriminant function as computed for each population.
The variables used in Chapter IV are defined as follows:

- \( a_t \) = The coefficients of the discriminant function.
- \( X_{ti} \) = The variables in the cases for which the discriminant function is computed.
- \( d_i \) = The value of the discriminant calculated for case in population \( i \).
- \( \overline{d}_i \) = The mean value of the discriminant calculated for population \( i \).

With these definitions, the above expression for the discriminant function can be written as:

\[
a_tX_{ti} \geq \frac{\overline{d}_1 + \overline{d}_2}{2} + \ln s
\]

This expression yields the test which places cases into population one or two depending on whether the expression on the left is greater than or equal to the value on the right. For populations of equal size costs of misclassification for both populations, the expression reduces to:

\[
a_tX_{ti} \geq \frac{\overline{d}_1 + \overline{d}_2}{2}
\]

a value midway between the two populations.
APPENDIX C

Table 6 displays the results of discriminant analysis on five variables. The "classification" column indicates the way the program classified the cases as compared to the way a loan officer classed them. Also shown are a calculated $F$ statistic value, the degrees of freedom and a Table $F$ value.

A word of caution is needed regarding interpretation of as many as five variables with 51 cases. The results of this analysis should not be used as a working model, but are presented to illustrate a way in which discriminate analysis might be used. Only the first several coefficients are statistically significant.

The signs of accounts receivable and loan amount are negative and the signs for balances/amount, balances and experience are positive. The larger the accounts receivable and loan amount, the less is the possibility of getting a loan. The larger the balances/amount, balances, and officer experience, the more likely is an application to receive a loan. The coefficients are in accord with what one would anticipate.
Table 6
CLASSIFICATION OF CASES BY VARIABLES
IN ORDER OF ENTRY

<table>
<thead>
<tr>
<th>Variables</th>
<th>Classification</th>
<th>Sign</th>
<th>P Value</th>
<th>d.f.</th>
<th>Table F</th>
</tr>
</thead>
<tbody>
<tr>
<td>34 Balances/ Amount</td>
<td></td>
<td>+</td>
<td>5.22</td>
<td>1.49</td>
<td>4.04</td>
</tr>
<tr>
<td>Loan Officer Decision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loan</td>
<td>12</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refuse</td>
<td>22</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Balances</td>
<td>L R</td>
<td>+</td>
<td>4.65</td>
<td>2.48</td>
<td>3.19</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>16</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>18</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 Accounts Receivable</td>
<td>L R</td>
<td>-</td>
<td>2.77</td>
<td>3.47</td>
<td>2.80</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>17</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>17</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Officer Experience</td>
<td>L T</td>
<td>+</td>
<td>2.32</td>
<td>4.46</td>
<td>2.57*</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>19</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>15</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Loan Amount</td>
<td>L R</td>
<td>-</td>
<td>1.97</td>
<td>5.45</td>
<td>2.42*</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>21</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>13</td>
<td>16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Indicates value not significant
APPENDIX D

A correlation matrix for the 34 original variables in the pilot study sample is shown in this appendix. It is apparent from a study of this table that there is considerable redundancy in the data. For example, variables seven and eight have correlation of .92. These variables represent the demand deposit balances for 1969 and 1968. Variable seven is used in subsequent analysis as a proxy for balances. Variables 27 and 28 represent sales and cost of sales, variables 29 and 30 represent pre-tax profits and profit after taxes. Variables 27 and 28 are highly correlated, as are 29 and 30. For this reason, variables 28, cost of sales and 29, pre-tax profits are eliminated.

Perhaps a less obvious example of high correlation is the correlation of .80 between variable 4, amount, and variable 6, commitment. This result is not surprising, but rather indicates consistency with what one would expect. Also, year-ago borrowing, variable 10, is correlated .88 with outstanding amount of borrowing at the time of the loan, variable 11. In this manner, the number of variables can be reduced prior to subsequent analysis. Finally, variable 34, balances/amount, is correlated with the commitment, variable 6, and also with variables 12 and 19, years a customer and total assets.

Assuming normality in the data, the individual correlations of the table may be examined for significance. Any correlation in the table may be assumed to be a transformation of a variable with zero mean and standard deviation \( \sigma \) where

\[
\frac{1}{\sqrt{n-3}} = .15
\]

where \( n \) is the number of items in the random sample. (Refer to Hoel [22, Chapter 7] for an explanation of the reliability of correlation coefficients.)
If a significance level of .05 is desired for \( n = 51 \), an estimate of a correlation coefficient \( r \) will be significant if \( r \geq 0.30 \). It is also possible to determine an interval of values within which one could reasonably be expected to fall for a sample size of 51 cases. "Reasonable" is understood here to mean with a probability of 0.95. The interval is, for a correlation of \( r = 0.9 \), \( 0.82 < r < 0.94 \). An interval which will contain \( r = 0.9 \) with a probability of 0.90 is \( 0.84 < r < 0.94 \). Finally, an interval which will contain a correlation of 0.5 with a probability of 0.95 is \( 0.25 < r < 0.69 \). A correlation of \( r = 0.5 \) has a large interval requirement for a probability of 0.95 in a sample of 51 cases.

In summary, correlations below 0.30 probably can be ignored. Any other correlations should be interpreted carefully in light of the large intervals needed for significance for a sample as small as 51 cases.
<table>
<thead>
<tr>
<th>Category</th>
<th>Correlation Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Officer Experience</td>
<td>0.77</td>
</tr>
<tr>
<td>2. Interest Rate</td>
<td>0.09</td>
</tr>
<tr>
<td>3. Maturity</td>
<td>0.12</td>
</tr>
<tr>
<td>4. Age</td>
<td>0.80</td>
</tr>
<tr>
<td>5. Security</td>
<td>0.36</td>
</tr>
<tr>
<td>6. Commitment</td>
<td>0.19</td>
</tr>
<tr>
<td>7. Balance-1969</td>
<td>0.28</td>
</tr>
<tr>
<td>8. Balance-1978</td>
<td>0.55</td>
</tr>
<tr>
<td>9. Loan &quot;Close-up&quot;</td>
<td>0.32</td>
</tr>
<tr>
<td>10. Ten-year Borrowing</td>
<td>0.84</td>
</tr>
<tr>
<td>11. Outstanding Amount</td>
<td>0.24</td>
</tr>
<tr>
<td>12. Years as a Customer</td>
<td>0.30</td>
</tr>
<tr>
<td>13. GPR</td>
<td>0.13</td>
</tr>
<tr>
<td>14. Accounts Receivable</td>
<td>0.35</td>
</tr>
<tr>
<td>15. Inventory</td>
<td>0.09</td>
</tr>
<tr>
<td>16. Current Assets</td>
<td>0.33</td>
</tr>
<tr>
<td>17. Fixed Assets</td>
<td>0.11</td>
</tr>
<tr>
<td>18. Deferred</td>
<td>0.09</td>
</tr>
<tr>
<td>19. Total Assets</td>
<td>0.19</td>
</tr>
<tr>
<td>20. Current Liabilities</td>
<td>0.39</td>
</tr>
<tr>
<td>21. Short Term Debt</td>
<td>0.40</td>
</tr>
<tr>
<td>22. Long Term Debt</td>
<td>0.14</td>
</tr>
<tr>
<td>23. Preferred Stock</td>
<td>0.08</td>
</tr>
<tr>
<td>24. Net Worth</td>
<td>0.12</td>
</tr>
<tr>
<td>25. Working Capital</td>
<td>0.20</td>
</tr>
<tr>
<td>26. Total Liabilities</td>
<td>0.26</td>
</tr>
<tr>
<td>27. Sales</td>
<td>0.07</td>
</tr>
<tr>
<td>28. Cost of Sales</td>
<td>0.13</td>
</tr>
<tr>
<td>29. Pre-tax Profits</td>
<td>0.12</td>
</tr>
<tr>
<td>30. Profits</td>
<td>0.11</td>
</tr>
<tr>
<td>31. Loan Type</td>
<td>0.09</td>
</tr>
<tr>
<td>32. Loan Purpose</td>
<td>0.08</td>
</tr>
<tr>
<td>33. Other Relationships</td>
<td>0.27</td>
</tr>
<tr>
<td>34. Balance/Amount</td>
<td>0.04</td>
</tr>
</tbody>
</table>
APPENDIX E

Additional computer runs were made using certain financial ratios, in an attempt to determine if these ratios would be useful in discriminating between loans and turn-downs. The variables current ratio, net worth/total liabilities, CGS/inventory, cash flow, quick ratio, and operating ratio were included. The results are displayed in Table 7.
**Table 7**

**DISCRIMINANT MODEL USING RATIOS**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Classification</th>
<th>Coefficient</th>
<th>d.f.</th>
<th>F Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balances/Amount</td>
<td></td>
<td>12.14</td>
<td>1,40</td>
<td>5.13</td>
</tr>
<tr>
<td>Balances</td>
<td></td>
<td>9.12</td>
<td>2,39</td>
<td>4.35</td>
</tr>
<tr>
<td>CGS/Inventory</td>
<td></td>
<td>-1.19</td>
<td>3,38</td>
<td>3.28</td>
</tr>
<tr>
<td>Net Worth/Total Liabilities</td>
<td></td>
<td>-0.31</td>
<td>4,37</td>
<td>2.91</td>
</tr>
</tbody>
</table>

* Not significant

<table>
<thead>
<tr>
<th>L</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>14</td>
</tr>
</tbody>
</table>

| Cash Flow                      | .02  | 5,36 | 2.59 |
| Current Ratio                  | -1.32| 6,35 | 2.18 |

Explanation of the signs of the coefficients requires some imagination. A negative coefficient on CGS/Inventory indicates that a company has a requirement for funds for inventory as demonstrated by a low inventory turnover ratio. A low ratio of net worth total liabilities indicates a need for capital. The cash flow with negative sign indicates a need for cash. The current ratio with a minus sign also shows a low ratio of current assets to current liabilities and may indicate a need for input of borrowed funds.


