

AD-A158 994

USE OF A BINARY CHOICE MODEL TO DETERMINE MARINE
OFFICER ATTRITION(U) NAVAL POSTGRADUATE SCHOOL MONTEREY
CA S F HURST ET AL. JUN 85

1/1

UNCLASSIFIED

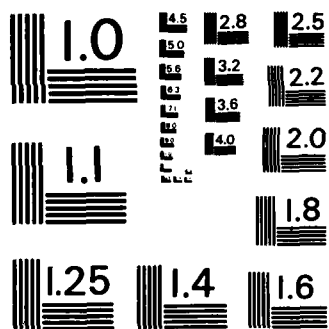
F/G 5/9

NL

END

FILMED

DTM:



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

2

NAVAL POSTGRADUATE SCHOOL

Monterey, California

AD-A158 994



THESIS

USE OF A BINARY CHOICE MODEL TO DETERMINE
MARINE OFFICER ATTRITION

by

Stephen Frederick Hurst

and

Thomas Edward Manion

June 1985

Thesis Advisor:

Dan C. Boger

DTIC
ELECTE
SEP 12 1985
S A D

DTIC FILE COPY

Approved for Public Release; Distribution Unlimited

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A158994	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Use of a Binary Choice Model to Determine Marine Officer Attrition		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis: June, 1985
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Stephen F. Hurst and Thomas E. Manion		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93943-5100		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93943-5100		12. REPORT DATE June, 1985
		13. NUMBER OF PAGES 92
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Marine Officer Attrition, Logit Analysis, Binary Choice		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The ability to predict Marine officer attrition continues to be a significant obstacle to accurate manpower planning at Headquarters Marine Corps. This paper expands on a previous study which used unemployment in a linear regression model to predict attrition. Through the use of a binary choice model, this study incorporates a number of additional variables, principal of which		

DD FORM 1473
1 JAN 73EDITION OF 1 NOV 65 IS OBSOLETE
S N 0102-LF-014-6601

1

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

is a performance index score to measure promotion potential. The logit function was employed to evaluate the inherent causal relationships and to further define the ability of statistical analysis techniques to predict officer attrition.

S N 0102- LF- 014- 6601

UNCLASSIFIED

2 SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

Approved for public release; distribution is unlimited.

**Use of a Binary Choice Model to Determine
Marine Officer Attrition**

by

Stephen F. Hurst
Lieutenant Colonel, United States Marine Corps
BBA, Wichita State University, 1969

and

Thomas E. Manion
Captain, United States Marine Corps
BS, Widener University, 1977

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
June 1985

Authors:

Stephen F. Hurst
Stephen F. Hurst

Thomas E. Manion
Thomas E. Manion

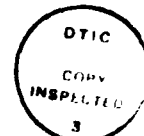
Approved by:

Dan C. Boger
Dan C. Boger, Thesis Advisor

David E. Melchar
David E. Melchar, Second Reader

Willis R. Greer Jr.
Willis R. Greer Jr, Chairman, Department of
Administrative Sciences

Kneale T. Marshall
Kneale T. Marshall, Dean of
Information and Policy Sciences



Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

ABSTRACT

The ability to predict Marine officer attrition continues to be a significant obstacle to accurate manpower planning at Headquarters Marine Corps. This paper expands on a previous study which used unemployment in a linear regression model to predict attrition. Through the use of a binary choice model, this study incorporates a number of additional variables, principal of which is a performance index score to measure promotion potential. The logit function was employed to evaluate the inherent causal relationships and to further define the ability of statistical analysis techniques to predict officer attrition.

TABLE OF CONTENTS

I.	INTRODUCTION	9
	A. GENERAL	9
	B. BACKGROUND	9
	1. Historical	9
	2. Application and results	12
	C. PLANNED IMPROVEMENTS	14
II.	DATA BASE	16
	A. SOURCE	16
	B. TRANSFORMATIONS	21
III.	MODEL SELECTION	26
	A. GENERAL	26
	1. Linear Multiple Regression	26
	2. Discriminant Analysis	28
	3. Logit	31
	B. MODEL SELECTION	34
IV.	ANALYSIS	40
	A. GENERAL	40
	B. INDEPENDENT VARIABLES	40
	C. DEPENDENT VARIABLE	42
	D. SCOPE OF STUDY	42
	1. General	42
	2. Biased Data Elimination	43
	3. Model Population Groups	43

E.	OUTPUT	46
1.	General	46
2.	First Lieutenants (Ground)	48
3.	Captains (Ground)	54
4.	Captains (Air)	60
5.	Majors (Air and Ground)	65
6.	Lieutenant Colonels (Ground)	67
7.	Lieutenant Colonels (Air)	73
8.	Colonels (All)	78
F.	SUMMARY	79
V.	SUMMARY	81
A.	GENERAL	81
B.	PREDICTION RATES	81
C.	SIGNIFICANCE OF VARIABLES	84
D.	PERFORMANCE INDEX INTERVALS	88
E.	CONCLUSION	90
	LIST OF REFERENCES	91
	INITIAL DISTRIBUTION LIST	92

LIST OF TABLES

I.	COMPARISON OF MODEL RESULTS	14
II.	RANK REPRESENTATION IN DATA BASE	19
III.	MILITARY PAY COMPARABILITY	21
IV.	LAGGED UNEMPLOYMENT RATES	22
V.	SIGNIFICANT YEARS BY RANK	45
VI.	GROUP/VARIABLE CORRELATION	46
VII.	VARIABLE SIGNIFICANCE, FIRST LIEUTENANT	50
VIII.	VARIABLE SIGNIFICANCE, CAPTAINS (GND)	59
IX.	VARIABLE SIGNIFICANCE, CAPTAINS (AIR)	63
X.	VARIABLE SIGNIFICANCE, LIEUTENANT COLONELS (GND)	70
XI.	VARIABLE SIGNIFICANCE, LIEUTENANT COLONELS (AIR)	75
XII.	VARIABLE SIGNIFICANCE, COLONELS (ALL)	79
XIII.	CONSOLIDATED PERFORMANCE INDEX INTERVALS	89

LIST OF FIGURES

2.1	Sample of the Basic Data	18
2.2	Example of USMC Fitness Report	20
2.3	Sample of the Revised Data Base	25
3.1	Sample SAS Output	36
3.2	Sample SAS Output	37
3.3	Sample SAS Output	38
3.4	Sample SAS Output	39
4.1	First Lieutenant (Ground)	49
4.2	First Lieutenant (Ground)	51
4.3	First Lieuten. : (Ground)	55
4.4	Captains (Ground)	57
4.5	Captains (Ground)	58
4.6	Captains (Ground)	61
4.7	Captains (Air)	62
4.8	Captains (Air)	64
4.9	Captains (Air)	66
4.10	Lieutenant Colonels (Ground)	68
4.11	Lieutenant Colonels (Ground)	69
4.12	Lieutenant Colonels (Ground)	72
4.13	Lieutenant Colonels (Air)	74
4.14	Lieutenant Colonels (Air)	76
4.15	Lieutenant Colonels (Air)	77
4.16	Colonels (All)	80
5.1	Consolidated Results	82

I. INTRODUCTION

A. GENERAL

This thesis is a follow-on to a previous study conducted at the Naval Postgraduate School entitled, "Marine Officer Attrition Model". That study, using economic factors and a linear regression model attempted to predict officer attrition by rank. The ability to correctly predict officer attrition has been in the past and continues to be a significant obstacle to accurate promotion planning and budget preparation at Headquarters, U. S. Marine Corps (HQMC). Previous studies have attempted to improve the unreliable "running average" method through the use of multiple variable linear regression. However, their limited predictive ability in determining an individual's decision to leave the Marine Corps suggested that better results could be achieved through the use of additional variables in a binary choice model.

B. BACKGROUND

1. Historical

The ability to accurately predict officer attrition contributes to the efficient management of the budget, promotion boards, and recruiting requirements. HQMC currently

uses a running average method to forecast attrition. This method uses the average annual attrition for the last six years adjusted by the intuitive assumptions of the officer responsible for determining the officer loss rate. This method is reliable only to a certain degree, with some years proving more accurate than others. The "off years" create significant complications for the efficient management of manpower planning. Because of this problem, HQMC requested a statistical study be undertaken to improve their predicting capabilities. The request was referred to the Naval Postgraduate School, and Major W. J. ESMANN, USMC, conducted an analysis which resulted in his thesis entitled, "Marine Officer Attrition Model".

Major Esmann's initial research seemed to indicate a dramatic improvement in the ability to predict total attrition, but he concluded his thesis with a recommendation that further research be conducted using additional variables in a binary choice model.

The first problem encountered by Major Esmann was to narrow the choice of variables which would be used to predict attrition to a workable number. The initial choice of variables was made through interviews and the intuitive decisions of the officer conducting the analysis. At first, three categories of variables were identified as factors which could determine whether an officer would stay on active duty or leave the Marine Corps. These variables were

military pay, the economy (specifically unemployment), and promotion potential. Of these three variables, promotion potential was immediately eliminated because a method to measure that variable had not yet been developed. However, it is significant to note that throughout the study, those officers interviewed felt promotion potential would be a major factor in determining officer attrition.

Next, military pay was studied using factors which showed its comparability with the private sector. A linear regression model was used to provide an insight to their predictive value. The results of this analysis indicated that comparability between military and civilian pay was not a significant factor in an officer's decision to leave the Marine Corps. These results contradicted a study conducted by the Center for Naval Analysis (CNA) which showed that pay comparability was a factor in an officer's decision to remain on active duty. However, the CNA study included all branches of the Armed Forces while Major Esmann's model considered Marine officers alone. The conclusion was that a Marine officer was a different breed of serviceman, not motivated by pay when making his decision to leave the Marine Corps [Ref. 1].

Once pay comparability and promotion potential were eliminated, only the economy remained as a primary factor. To determine which aspects were significant, the possible economic impact on attrition was broken into three areas:

colonels, MOS equalled their first additional MOS (FSTAMOS), or the MOS they had been assigned prior to promotion.

Finally, a determination was required to accurately present the unemployment statistics, or the UNEMP variable. Like Major Esmann, the authors assigned Naval Aviators and Naval Flight Officers the unemployment figures associated with technical executives while the other officers were assigned the managerial statistics. This was clearly an subjective decision, but one which appeared to most correctly reflect the employability of the population.

An example of the revised data appears in Figure 2.3.

alphanumeric code for education major was converted to a binary (1 and 2) choice that could be used with the LOGIST procedure of SAS. This was accomplished in a somewhat subjective manner based on the marketability of the individual's level and type of degree. As an example, those individuals possessing degrees in the soft skills, such as business administration and education were given an EDSPEC of 1 while those with more technical degrees such as computer science or electrical engineering were given an EDSPEC of 2.

The variable OUT was derived from the DATESEPY element of the basic data. If an individual had a entry for this element, it indicated he had been released from active duty and was assigned an OUT variable value of 1. An individual remaining on active duty received a value of 0.

YEARS, or the number of years each officer had been on active duty, was simply his active duty base date (ADBDY) minus the data cycle year (CYC_YR), a date used by the MMS to mark the annual update of individual's record.

Because colonels are assigned a 99XX series PMOS, or numerical designator for an officer's primary military occupational specialty, upon promotion, it was necessary to create a new variable entitled MOS that would correlate the actions of these officers with those of lesser grades. For lieutenant colonel and below, MOS equalled PMOS, but for

TABLE IV
LAGGED UNEMPLOYMENT RATES

Year	Managerial	Tech. Executive
1977	.0295	.0310
1978	.0245	.0280
1979	.0210	.0250
1980	.0225	.0245
1981	.0255	.0265
1982	.0315	.0305
1983	.0355	.0320

The inflation of fitness reports during the period also required the INDEX be normalized to a common mean. This was accomplished using the following formula:

$$\text{INDEX}_n = (\text{INDEX} - \text{INDEX}_{rm}) / \text{Std Dev}$$

where:

INDEX_n = Normalized Index

INDEX = given

INDEX_{rm} = Index mean by rank and years of service

Std Dev = Standard Deviation corresponding to the INDEX_{rm}

To determine if an individual's education major played a significant role in his decision to remain on active duty, the Marine Corps Manpower Management System (MMS)

TABLE III
MILITARY PAY COMPARABILITY

Year	Difference
1977	-4.8
1978	-4.3
1979	-6.5
1980	-7.3
1981	-4.8
1982	-0.4
1983	-4.2

Last, managerial and technical executive unemployment figures were obtained from the Bureau of Labor and Statistics. These figures [Ref. 4], which were lagged by six months to bring them in line with a date closer to the time an officer made the decision to resign, are displayed in TABLE IV.

B. TRANSFORMATIONS

Although some of the data could be used directly, many of the required elements needed to be derived from the basic data as supplied.

An individual's performance INDEX was determined by dividing his SCORE by his POS_SCR. Due to the inflation of fitness reports, the INDEX covered a very narrow range.

UNITED STATES DEPARTMENT OF THE ARMY
FORM 10-835-4

— See reverse for instructions —

PROGRAM		ORGANIZATION		DESCRIPTIVE TITLE (Approved on request)	
1. MARINE REPORTED ON		2. POST NAME		3. GRADE	
4. OCCASION AND PERIOD COVERED		5. TYPE		6. PERIODS OF NONAVAILABILITY (If more than one, explain)	
7. FIRST REGULAR DUTY		8. MONTHS		9. LINE NO.	
10. SECOND REGULAR DUTY		11. MONTHS		12. LINE NO.	
13. DEPENDENTS REQUIRING TRANSPORTATION		14. NO.		15. LOCATION	
16. DUTY PREFERENCE (Code)		17. DUTY PREFERENCE (Descriptive Title)		18. ADDRESS (on request)	
19. REPORTING SENIOR		20. SENIOR NO.		21. NAME AND DUTY ASSIGNMENT	
22. SPECIAL USE (Mark if applicable)		23. ATTENTION TO DUTY		24. YOUR ESTIMATE OF MARINE'S GENERAL VALUE TO THE SERVICE	
25. PERFORMANCE		26. COOPERATION		27. DISTRIBUTION OF RATING	
28. ADDITIONAL DUTIES		29. INITIATIVE		30. ALL BOXES SO THAT THE SUM OF EACH COLUMN CORRESPONDS TO ITEM 15b	
31. ADMINISTRATIVE DUTIES		32. DOCUMENT		33. ALL MARINES OF THIS GRADE	
34. HANDLING OFFICERS (Name, No., No.)		35. PRESENCE OF MIND		36. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
37. HANDLING OFFICERS (Name, No., No.)		38. PRESENCE OF MIND		39. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
40. HANDLING OFFICERS (Name, No., No.)		41. PRESENCE OF MIND		42. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
43. HANDLING OFFICERS (Name, No., No.)		44. PRESENCE OF MIND		45. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
46. HANDLING OFFICERS (Name, No., No.)		47. PRESENCE OF MIND		48. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
49. HANDLING OFFICERS (Name, No., No.)		50. PRESENCE OF MIND		51. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
52. HANDLING OFFICERS (Name, No., No.)		53. PRESENCE OF MIND		54. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
55. HANDLING OFFICERS (Name, No., No.)		56. PRESENCE OF MIND		57. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
58. HANDLING OFFICERS (Name, No., No.)		59. PRESENCE OF MIND		60. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
61. HANDLING OFFICERS (Name, No., No.)		62. PRESENCE OF MIND		63. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
64. HANDLING OFFICERS (Name, No., No.)		65. PRESENCE OF MIND		66. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
67. HANDLING OFFICERS (Name, No., No.)		68. PRESENCE OF MIND		69. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
70. HANDLING OFFICERS (Name, No., No.)		71. PRESENCE OF MIND		72. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
73. HANDLING OFFICERS (Name, No., No.)		74. PRESENCE OF MIND		75. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
76. HANDLING OFFICERS (Name, No., No.)		77. PRESENCE OF MIND		78. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
79. HANDLING OFFICERS (Name, No., No.)		80. PRESENCE OF MIND		81. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
82. HANDLING OFFICERS (Name, No., No.)		83. PRESENCE OF MIND		84. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
85. HANDLING OFFICERS (Name, No., No.)		86. PRESENCE OF MIND		87. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
88. HANDLING OFFICERS (Name, No., No.)		89. PRESENCE OF MIND		90. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
91. HANDLING OFFICERS (Name, No., No.)		92. PRESENCE OF MIND		93. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
94. HANDLING OFFICERS (Name, No., No.)		95. PRESENCE OF MIND		96. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
97. HANDLING OFFICERS (Name, No., No.)		98. PRESENCE OF MIND		99. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	
100. HANDLING OFFICERS (Name, No., No.)		101. PRESENCE OF MIND		102. ONLY DURING THE PERIODS OF SERVICE IN WHICH YOU INDICATE YOUR	

SECTION D. SIGNATURES

22. CERTIFY the information in section A is correct to the best of my knowledge. I have been counseled immediately prior to the submission of this report (AWM MCO 11610.7).

Signature of Marine Reported On: _____ Date: _____

24. Check one when required I HAVE SEEN THIS COMPLETED REPORT AND: ☐ HAVE NO STATEMENT TO MAKE ☐ HAVE ATTACHED A STATEMENT

Signature of Reporting Senior: _____ Date: _____

23. I CERTIFY that to the best of my knowledge and belief all entries made herein are true and without prejudice or partiality. I HAVE (NOT) counseled this Marine concerning his overall performance of duty.

Signature of Reporting Senior: _____ Date: _____

25. REVIEWING OFFICER (Name, Grade, Duty Assignment): _____

25a. INITIALS: _____

25b. DATE: _____

0565920

Figure 2.2 Example of USMC Fitness Report

TABLE II

RANK REPRESENTATION IN DATA BASE

Colonel	4,072
Lieutenant Colonel	10,327
Major	20,532
Captain	26,750
First Lieutenant	<u>26,250</u>
TOTAL	87,931

First, from the officer's fitness reports, HQMC provided a performance SCORE which was determined by weighting each of the "Performance" and "Qualities" blocks (see Figure 2.2) of the report and taking this value times a numeric value assigned to the mark received (Outstanding, Excellent, etc.). The maximum score assigned, or POS_SCR, is the score that would have resulted had the officer received a mark of "Outstanding" in each of the observed areas. Both SCORE and POS_SCR are cumulative, with senior officers having a higher value for these variables than a more junior officer.

Second, a relationship between military and civilian pay, expressed as a percentage difference of the former from the latter, was obtained from the government's Employment Cost Index [Ref. 3] and added to the data base for each year of the period covered. The figures used in our analysis are displayed in Table III.

18

- (7) FCOMMY, FCOMMM, FCOMMD. This is the date the officer was first commissioned.
- (8) SCORE. This is the officer's Performance Index Score, to be described later in this chapter.
- (9) POS SCR. This is the maximum, or Possible Score, that an officer could have attained for his length of service.
- (10) DATESEPY, DATESEPM, DATESEPD. This is the date the officer was separated from active service. A "." in these columns means the officer is still on active duty.
- (11) CYC YR. This is Manpower Management System date indicating the end of the record year this data was compiled from.

The first change made to the basic data supplied was to eliminate those categories of officers whose presence could bias the results or that would have negligible impact on the outcome. To this end, all second lieutenants were eliminated because they would be promoted prior to reaching a decision point in their career. Those officers with more than four years of enlisted service, warrant officers, and limited duty officers were also eliminated because a preliminary statistical analysis indicated they tended to be career oriented with virtually none leaving the service prior to their twentieth year. The number of officers, by rank, remaining in the data base are displayed in Table II.

In addition to the personal data provided through the MMS, data pertinent to an individual's decision to remain on active duty were obtained from a number of other sources.

II. DATA BASE

A. SOURCE

The data for this thesis were obtained from the Marine Corps Manpower Management System (MMS) and fitness report input. It consisted of 132,903 records containing nineteen data elements on every Marine officers on active duty from 1977 through 1984. A record was created for each officer for each year, so that of the 132,903 records, somewhat less were actually in the data base due to multiple entries for those officers that remained on active duty during the period covered. A detailed example of the data supplied by HQMC is contained in Figure 2.1. The variables included were:

- (1) ID CODE. This is an encoded identification of the officer whose data appears on this line.
- (2) PMOS. This numeric descriptor indicates an officer's Primary Military Occupational Speciality.
- (3) FSTAMOS. This is the First Additional Military Occupational Speciality, or secondary skill the officer is qualified for.
- (4) ED MAJ. This a numeric designator describing the officer's civilian education major.
- (5) RANK. This alphanumeric variable represents the pay grade equivalent of the officer's rank.
- (6) ADBDY, ADBDM, ADBDD. Combined, these variables indicate the Active Duty Base Date. This date is often not the actual date an officer first came on active duty, but has been adjusted for time spent in the reserves or as an enlisted person.

averaging method which was proving more accurate. HQMC felt that although 1984 was not a typical year in terms of attrition, the averaging method provided a more accurate prediction than the Esmann model. These discussions gave some initial concerns but the authors believed the use of the performance index would prove to be critical factor in improving attrition prediction.

In the subsequent chapters, the authors will describe the basic data as supplied by HQMC and the transformations required to create the working data base for this study; a description of the statistical functions available for this type of research, their strengths and weaknesses, and the eventual selection of the model that was used; the methodology employed during the data manipulation phase; and an analysis of the results with concluding comments.

TABLE 1
COMPARISON OF MODEL RESULTS

Rank	Avg Method 1983	Regress Model 1983	Actual 1983	% Error Avg 1983	% Error Reg 1983	Regress Model 1984
LtCol	177	152	127	39.4	19.7	152
Major	186	149	146	27.4	2.1	164
Capt	452	377	392	15.3	-3.8	407
1stLt	525	416	396	32.6	5.1	425

C. PLANNED IMPROVEMENTS

With this initial background, the authors set out to expand the previous study and provide HQMC with an improved method for predicting officer attrition. The problem of establishing a variable that could measure promotion potential, one of the key factors missing from Major Esmann's model, was solved using fitness report data provided by HQMC. An explanation of this process is covered in Chapter II of this thesis.

Next, a number of interviews were conducted with officers at HQMC responsible for predicting officer attrition. These interviews provided insight into the effectiveness of previous attempts to use Major Esmann's model. In essence, HQMC had achieved negative results using the Marine Officer Attrition Model during FY1984 and had reverted to the

requirements and, therefore, recruiting goals. From these figures, the manpower budget is more accurately prepared and the budget call improved. Second, with a more accurate prediction of attrition, realistic eligibility zones can be determined, thus eliminating the mismatch between the number of vacancies and the number of officers selected for promotion.

The results of the model, using unemployment as the key variable, produced significant improvements over the old averaging method. Again, the averaging method took the sum of the last six years attrition, by rank, and divided it by six. The value derived was then modified depending on the intuition of responsible officer. TABLE 1 shows the results of the averaging method, Major Esmann's model, the actual figures for FY83, and a prediction (using his model) for FY84. As depicted in TABLE I, a clear improvement can be seen over the averaging method.

Major Esmann's study concluded that unemployment is a significant factor in predicting officer attrition. However, his concluding recommendation centered around adding a promotion potential variable to the unemployment statistics in order to improve the model's forecasting ability [Ref. 2].

unemployment, consumer price index (CPI), and the Gross National Product (GNP). Using this economic data, initial tests were again run using a linear regression model with both the CPI and GNP proving insignificant as factors in determining officer attrition.

Finally, unemployment was evaluated and segmented into professional-technical and managerial unemployment. These factors were assigned to groups within the data base, aviation officers being classified as professional-technical and ground officers as managerial. Again, a linear regression model was used to determine the significance of this variable.

The data base used for Major Esmann's study was limited for several reasons. First, the data included information from the years 1976 through 1983 only. This was a result of the Marine Corps not being fully automated prior to this period. Data collected previous to 1976 was incomplete and unreliable due to system complications. Second, attrition resulting from the Viet Nam war made it undesirable to use data from that period.

2. Application and results

As previously mentioned, Major Esmann's model was formulated to assist planners in two ways. First, it was designed to predict future manpower levels, an annual process whereby officer attrition determines accession

III. MODEL SELECTION

A. GENERAL

To assist in the analysis and prediction of Marine officer attrition, it was necessary to select a computer-assisted model that would most appropriately handle the unique aspects of this particular study. Previous studies of the subject had used multivariate regression analysis. Although this procedure was somewhat successful, it was apparent that other models were available that might prove more adept at accurately predicting an officer's decision to remain in the Marine Corps or leave the service. Therefore, following the collection and refinement of the data base, a thorough analysis of the available models was conducted to determine the algorithm that would most accurately predict the binary "stay / leave" decision.

1. Linear Multiple Regression

Linear multiple regression analysis is probably the most commonly used tool of the statistician who wishes to develop a model that will predict a dependent output from a series of known, independent inputs. This form of regression analysis is concerned with modeling the relationship among the independent variables. It quantifies how a response, the dependent variable, is related to a set of explanatory, independent variables. If the true relationship among the variables were known precisely, the

statistician would be able to accurately predict the outcome. However, the true relationship is rarely known and approximations must be developed from empirical evidence.

To accomplish this approximation, the concept of least squares is used to fit the regression equation to the observed data. The least squares method requires that a line be chosen to fit the data so that the sum of the squares of the vertical deviations separating the points from the line will be a minimum. These deviations are represented by the lengths of the vertical line segments that connect the points to the estimated regression line. In the case of multiple regression, the line is replaced by a hyperplane with dimensions equal to the number of independent variables. The resulting estimated multiple regression equation would be:

$$Y = a + b_1 X_1 + b_2 X_2 . . . + b_n X_n + e$$

where:

- Y = The value of Y calculated from the regression equation.
- a = The value of the constant or "y-intercept".
- b_n = The estimated regression coefficient of a particular independent variable.
- X_n = An independent variable.
- e = Residual error associated with the model

As can be seen, the major goal of linear multiple regression is to predict Y from the regression plane for given levels of X. This is an excellent tool, with some reservations, if the independent variables are continuous and randomly distributed (weight, speed, or payload) and the independent variable will also be continuous, such as cost [Ref. 5]. But, if the investigator wishes to predict the behavior of an individual or group based on the observed past performance of a similar group, then the traditional forms of regression analysis are not able to accurately predict their response or action because the independent variable is qualitative, or at least observed to be qualitative, and not quantitative. Voting, the decision to marry, have children, or go to college are examples of qualitative decisions that are not adequately handled by linear multiple regression. What is needed, therefore, is a set of statistical techniques that can do the work of multiple regression but with qualitative choices as the independent variables. Following will be a discussion of two of the more commonly used procedures to accomplish this task.

2. Discriminant Analysis

Discriminant Analysis deals with a specific class of statistical problem that focuses on the study of groups of populations or data sets. In general it is assumed that the groups being analyzed are discrete and identifiable, that each observation in the group can be described by a set of

measurable variables, and that the variables have a normal distribution within each group. With this in mind, the purpose of discriminant analysis is to test for mean group differences and to describe the overlaps among the groups (the decision to leave the Marine Corps for instance, is a different one than the decision to stay so that the probabilities may add to more than one due to this overlap). Furthermore, discriminant analysis will construct a classification scheme based on the variable set and thereby create a predictive model with which to determine the probable actions of similar groups.

To use discriminant analysis as a predictive tool, a concept similar to regression analysis called the discriminant function is employed. However, unlike the linear regression problem, the dependent variable is dichotomous. This function uses a weighted combination of the independent variables to classify a response into a variable group. Furthermore, using a weighted sum of values derived from the individual predictor variables, a discriminant score is calculated. A cutoff score is then determined that, used in conjunction with the discriminant score, assigns those observations with discriminant scores higher than the cutoff score to one of the criterion groups while those observations with discriminant scores less than the cutoff are assigned to the other [Ref. 6]. The discriminant function is expressed as follows:

$$L = b_1 X_1 + b_2 X_2 + \dots + b_n X_n$$

where:

L = Discriminant score

b_n = Weights associated with the predictors

X_n = Values of the predictor variables

The default cutoff score in most automated systems, such as SAS or SPSS, is .5 with the user being able to modify the breakpoint, if desired, to weight the outcome. This would be done to reduce the number of misclassifications in a particular direction.

In general, discriminant analysis allows the statistician to classify objects into groups based on their values as predictor variables and, much like multiple regression, creates a model that can predict individual or group behavior based on the actions of a similar population base from the past. One of the key factors in accurately predicting behavior, however, is that there be a large difference between the criterion groups. This was certainly not the case with one of the primary variables used in the study of Marine officer prediction, namely the performance INDEX. Due to the inflation of the fitness report marks during the period covered by the data base, it was found that the vast majority of the INDEX values fell into a very narrow range. Thus the difference between an outstanding officer and a less capable one was numerically very small. This led to classification errors and therefore less than

acceptable results when trying to determine what percentage of a particular group would leave the Marine Corps.

3. Logit

Logit is a binary-choice model that assumes that individuals faced with a choice between two alternatives will make a decision based on their respective characteristics. Unlike discriminant analysis, which postulates that the observed values of individual characteristics and attributes of alternatives are drawings from posterior distributions conditioned on actual responses, logit proposes that the actual responses are drawings from multinomial distributions with selection probabilities conditioned on the observed values of individual characteristics and attributes of alternatives. With this in mind, it can be seen that the logit model can be used to calculate the probability of an individual making a particular choice, such as to get out of the Marine Corps, conditioned on an observation of some specific elements of information about that particular officer. To accomplish that process, the function takes the general form:

$$Z_j = \text{Log} \left(\frac{P_j}{1 - P_j} \right) = b_1 X_{1j} + b_2 X_{2j} + \dots b_n X_{nj} + e$$

where:

Z_j = Probability that a particular decision will be made given a particular variable grouping

P_j = Probability that an individual will make a

particular decision given observation "j" of the variables X

X_{ij} = A variable such that "i" = 1,2,3, etc. representing the 1st, 2nd, and 3rd variables and "j" being the jth observation of a particular variable

b_n = Coefficient associated with each independent variable

To assist in the explanation of this process, the following example of the logit function is provided.

SAMPLE DATA BASE

Name	Sex	Income	Status	P
A	M	H	I	.62
B	M	M	I	.47
C	F	M	I	.69
.	.	.	.	.
.	.	.	.	.
X	M	L	O	.47
Y	M	L	I	.53
Z	M	H	O	.38

P = The percentage of the data base having a particular profile compared to the total number in the independent variable group. For example, males(M) with high(H) income that stayed in(I) divided by males with high income

Applying this percentage to the logit function,

$$Z_j = \text{Log} \left(\frac{P_j}{1 - P_j} \right)$$

the probabilities associated with all (in this example six) of the possible combinations of the independent variables are derived. These are:

- Z_1 = Male, high
- Z_2 = Male, medium
- Z_3 = Male, low
- Z_4 = Female, high
- Z_5 = Female, medium
- Z_6 = Female, low

Once these probabilities have been calculated in conjunction with a particular choice, for instance to get out, the model is run against the population to be tested and the individual probabilities summed to determine the number predicted to leave the Marine Corps. For example, using the following test data,

TEST DATA			
Obs	Sex	Income	Z
1	M	H	.34
2	F	M	.69
3	F	H	.52
4	M	L	.23
5	M	L	.23
6	M	M	.38
			<u>2.39</u>

the logit function predicted that 2.39 officers would be leaving the Marine Corps.

Since the logit function makes far fewer assumptions than the discriminant model, such as no multivariate normality for covariates, this form of regression analysis is often preferred to discriminant analysis. In addition, the theory that an officer would leave the Marine Corps

because of pay differences, his educational background, or perceived probability of promotion suggests a causal relationship between the explanatory variables and the probability of his separation. Here again, in side-by-side comparisons, the logit function has proved to be somewhat more accurate in forecasting the direction the dependent variable will take [Ref. 7].

B. MODEL SELECTION

Based on the foregoing analysis, and the availability of the appropriate models at the Naval Postgraduate School, the authors elected to use the logit function to conduct this study. As implemented by the Statistical Analysis System (SAS), the logit function is entitled LOGIST. Supporting the basic program are a number of standard statistical tools to assist the user in the stepwise elimination of insignificant variables. As a guide to the reader in evaluating the SAS LOGIST output, a sample printout is included in Figures 3.1 - 3.4, with the critical points explained below.

1. This matrix depicts the percentile distribution of a particular group based on their individual performance INDEX. This matrix, although not part of the LOGIST procedure, was included to assist in the analysis by displaying the number of officers falling into each of the ten categories. This breakout by INDEX gives the investigator a detailed view of why the model is, or is not, performing as expected. In this example, there is a wide difference between those staying in (the column headed "0") and those leaving (the column headed "1"). Although this difference slowly fades as the INDEX

scores increase, this model should perform very well in predicting officer attrition because of the distinct classification process that has occurred.

2. This is the number of observations in the group considered, less those with missing data.
3. Displayed in this area is the basic statistical data on the independent variables.
4. The P-value for testing the joint association of all variables in the model with the dependent variable is displayed at this point. In this example, the P-value of 0 indicates the model is totally significant.
5. These are the P-values of the individual variables deleted from the model as insignificant.
6. Displayed here are the estimated regression coefficients for each variable in the model.
7. This column contains the estimated standard errors of the regression coefficient estimates.
8. This column contains the P-value of the variables retained by the model as being significant.
9. This column depicts the lower confidence limit of the predicted probability.
10. Displayed in this column is the predicted probability of this particular officer leaving the Marine Corps given the values of the independent variables associated with this observation.
11. This column contains the upper confidence limit of the predicted probability.
12. This is the summation of the predicted probabilities for this group, which equals the total number of officers the model predicts will be getting out of the Marine Corps [Ref. 8].

This concludes a description of the model selection procedures. In Chapter IV, the authors will analyze the output resulting from the application of the logit model to each of the groups studied.

LT COLS (GND) WITH 20 YEARS

TABLE OF INTERVAL BY OUT

①

INTERVAL	OUT				
FREQUENCY					
PERCENT					
ROW PCT					
COL PCT					
	.	0	1	TOTAL	
1	11	54	11	65	
	:	8.33	1.70	10.03	
	:	83.03	18.92		
	:	11.66	5.95		
2	12	55	10	65	
	:	3.49	1.54	10.03	
	:	84.62	15.38		
	:	11.88	5.41		
3	21	55	10	65	
	:	8.49	1.54	10.03	
	:	84.62	15.38		
	:	11.88	5.41		
4	22	69	15	84	
	:	10.65	2.31	12.96	
	:	22.14	17.96		
	:	14.90	3.11		
5	11	33	14	47	
	:	5.09	2.13	7.25	
	:	70.21	29.79		
	:	7.13	7.57		
6	13	44	21	65	
	:	6.77	3.24	10.03	
	:	67.69	33.31		
	:	9.50	11.35		
7	19	43	17	60	
	:	7.41	2.62	10.03	
	:	73.32	26.15		
	:	10.37	9.19		
8	21	33	25	58	
	:	6.02	3.86	9.88	
	:	60.94	39.06		
	:	8.42	13.31		
9	16	36	29	65	
	:	5.50	4.43	10.03	
	:	55.38	44.62		
	:	7.78	15.58		
10	11	30	33	63	
	:	4.63	5.09	9.72	
	:	47.62	52.38		
	:	6.46	17.34		
TOTAL	.	463	185	648	
	:	71.45	28.55	100.00	

Figure 3.1 Sample SAS Output

LT COLS (CMD) WITH 20 YEARS
STEPWISE LOGISTIC REGRESSION PROCEDURE
DEPENDENT VARIABLE: OUT

548 OBSERVATIONS
403 OUT = 0
135 OUT = 1
157 OBSERVATIONS DELETED DUE TO MISSING VALUES

VARIABLE	MEAN	MINIMUM	MAXIMUM	RANGE
INDXN	0.884749	0.789441	0.933742	0.194301
PAY	3.77731	-7.3	-4.5	2.3
UNEMP	0.0244909	0.021	0.0235	0.0025
EDSPEC	1.16321	1	2	1
MOS	1.60494	1	2	1

-2 LOG LIKELIHOOD FOR MODEL CONTAINING INTERCEPT ONLY= 775.10

STEP 0. THE FOLLOWING VARIABLES ARE ENTERED:

INTERCEP INDXN PAY UNEMP EDSPEC MOS

MODEL CHI-SQUARE= 73.46 WITH 5 D.F. (SCORE STAT.) P=0.0
CONVERGENCE OBTAINED IN 5 ITERATIONS. R=0.297.
MAX ABSOLUTE DERIVATIVE=0.79610-08. -2 LOG L= 703.81.
MODEL CHI-SQUARE= 70.36 WITH 5 D.F. (-2 LOG L.R.) P=0.0

FRACTION OF CONCORDANT PAIRS OF PREDICTED PROBABILITIES AND RESPONSES 0.707
RANK CORRELATION BETWEEN PREDICTED PROBABILITY AND RESPONSE 0.430

FAST BACKWARD ELIMINATION

VARIABLE DELETED	CHI-SQUARE	P	RESIDUAL CHI-SQUARE	D.F.	P
EDSPEC	0.02	0.8674	0.03	1	0.8674
PAY	1.05	0.3093	3.03	2	0.2143
UNEMP	3.84	0.0500	0.92	2	0.6345

STEP 1.

MODEL CHI-SQUARE= 67.37 WITH 2 D.F. (SCORE STAT.) P=0.0
CONVERGENCE OBTAINED IN 5 ITERATIONS. R=0.295.
MAX ABSOLUTE DERIVATIVE=0.70750-07. -2 LOG L= 703.81.
MODEL CHI-SQUARE= 71.29 WITH 2 D.F. (-2 LOG L.R.) P=0.0

FRACTION OF CONCORDANT PAIRS OF PREDICTED PROBABILITIES AND RESPONSES 0.598
RANK CORRELATION BETWEEN PREDICTED PROBABILITY AND RESPONSE 0.413

TEST FOR ALL VARIABLES NOT IN THE MODEL
RESIDUAL CHI-SQUARE= 0.99 WITH 3 D.F. P=0.0723

STEP 2. VARIABLE UNEMP ADDED WITH CHI-SQUARE= 3.93 P=0.0473 R=0.050.

MODEL CHI-SQUARE= 70.34 WITH 3 D.F. (SCORE STAT.) P=0.0
CONVERGENCE OBTAINED IN 5 ITERATIONS. R=0.299.
MAX ABSOLUTE DERIVATIVE=0.23170-08. -2 LOG L= 698.90.
MODEL CHI-SQUARE= 73.29 WITH 3 D.F. (-2 LOG L.R.) P=0.0

FRACTION OF CONCORDANT PAIRS OF PREDICTED PROBABILITIES AND RESPONSES 0.701
RANK CORRELATION BETWEEN PREDICTED PROBABILITY AND RESPONSE 0.413

Figure 3.2 Sample SAS Output

LT COLS (END) WITH 20 YEARS
STEPWISE LOGISTIC REGRESSION PROCEDURE

DEPENDENT VARIABLE: OUT

TEST FOR ALL VARIABLES NOT IN THE MODEL
RESIDUAL CHI-SQUARE= 3.09 WITH 2 D.F. P=0.2135

NO ADDITIONAL VARIABLES MET THE 0.0500 SIGNIFICANCE LEVEL FOR ENTRY.

FINAL PARAMETER ESTIMATES					
VARIABLE	(6) BETA	(7) STD. ERROR	CHI-SQUARE	(8) P	R
INTERCEPT	-16.21361579	2.59137240	39.15	0.0000	
INDXN	20.52179014	2.91017724	49.74	0.0000	0.248
MOS	-1.7446711	0.19242113	25.68	0.0000	-0.175
UNEMP	-53.57797422	29.62733416	3.91	0.0480	-0.050

Figure 3.3 Sample SAS Output

PARAMETER ESTIMATES AND 95% CONFIDENCE LIMITS										
VARIABLE	LABEL	PREDICTED PROBABILITY	N	MEAN	STANDARD DEVIATION	MINIMUM VALUE	MAXIMUM VALUE	STD ERROR OF MEAN	SUM	VARIANCE
-P-			157	0.27374870	0.13164282	0.06777126	0.65397102	0.01050624	42.97854519	0.01732983
Q35	123456789									
Q36	123456789									
Q37	123456789									
Q38	123456789									
Q39	123456789									
Q40	123456789									
Q41	123456789									
Q42	123456789									
Q43	123456789									
Q44	123456789									
Q45	123456789									
Q46	123456789									
Q47	123456789									
Q48	123456789									
Q49	123456789									
Q50	123456789									
Q51	123456789									
Q52	123456789									
Q53	123456789									
Q54	123456789									
Q55	123456789									
Q56	123456789									
Q57	123456789									
Q58	123456789									
Q59	123456789									
Q60	123456789									
Q61	123456789									
Q62	123456789									
Q63	123456789									
Q64	123456789									
Q65	123456789									
Q66	123456789									
Q67	123456789									
Q68	123456789									
Q69	123456789									
Q70	123456789									
Q71	123456789									
Q72	123456789									
Q73	123456789									
Q74	123456789									
Q75	123456789									
Q76	123456789									
Q77	123456789									
Q78	123456789									
Q79	123456789									
Q80	123456789									
Q81	123456789									
Q82	123456789									
Q83	123456789									
Q84	123456789									
Q85	123456789									
Q86	123456789									
Q87	123456789									
Q88	123456789									
Q89	123456789									
Q90	123456789									
Q91	123456789									
Q92	123456789									
Q93	123456789									
Q94	123456789									
Q95	123456789									
Q96	123456789									
Q97	123456789									
Q98	123456789									
Q99	123456789									
Q100	123456789									

Figure 3.4 Sample SAS Output

indicate increased employability, the expected sign of this coefficient would be positive. However, in this case the negative coefficient assigned to the EDSPEC variable indicates that those officers with the more technical, and therefore more saleable, degrees are remaining in the Marine Corps.

The remaining variables, MOS and PAY, were less significant than the above variables but not enough to be deleted from the model. While MOS and PAY are factors in this model it is understandable why they are less significant. In the case of MOS, the junior officer is at a point in his career that the significance of a combat arms MOS has little effect on his outlook regarding career aspirations in the Marine Corps. If it has, the officer is still at a point in his career to allow him the flexibility to change his MOS to one of the combat arms. Nevertheless, the positive coefficient associated with this variable indicates, as expected, that those officers with the combat arms MOS's are remaining in the Corps. The least significant of all of the variables was PAY. Interestingly enough, the positive coefficient associated with this variable indicates that as pay becomes less comparable with the private sector, attrition decreases. This apparent anomaly is unexplained. Officers in this data group, for the most part, are experiencing their first full-time job. The military offers these junior officers a comparable

continue to make the Marine Corps their career, the distinction between those with the greater promotion potential and those who will fail selection or get out is much less later in their careers. Since the dependent variable is the probability of separating from active service, one would expect a negative relationship between the INDEX score and the dependent variable. This negative coefficient clearly indicates a strong negative correlation between a higher INDEX score and an officer's decision to leave active duty.

Following INDEX is the UNEMPLOYMENT variable. This would seem to indicate that many of these young officers initially made the military their choice of employment because of the high unemployment rates they faced upon graduation from college. Three years later, these same officers, faced with a similar decision, opt for the security of a military career, making the high unemployment rate more profound. Supporting this view is the negative coefficient associated with the UNEMPLOYMENT variable, which indicates that as the unemployment rate increases, the probability that an officer will leave active duty decreases.

The educational background of this data group also proved to be highly significant. Again, since the dependent variable is the probability of separating from the Marine Corps and a higher education level would seem to

LIEUTENANT (GND) WITH 3 YEARS
STEPWISE LOGISTIC REGRESSION PROCEDURE

DEPENDENT VARIABLE: OUT

4108 OBSERVATIONS
3179 OUT = 0
929 OUT = 1
1052 OBSERVATIONS DELETED DUE TO MISSING VALUES

VARIABLE	MEAN	MINIMUM	MAXIMUM	RANGE
INDXN	0.881927	0.325017	1	0.674383
PAY	-5.88715	-7.3	-4.5	2.8
UNEMP	0.0238633	0.021	0.0295	0.0085
EDSPEC	1.15287	1	2	1
MOS	1.3313	1	2	1

-2 LOG LIKELIHOOD FOR MODEL CONTAINING INTERCEPT ONLY= 4392.07

STEP 0. THE FOLLOWING VARIABLES ARE ENTERED:

INTERCLP INDXN PAY UNEMP EDSPEC MOS

MODEL CHI-SQUARE= 335.68 WITH 5 D.F. (SCORE STAT.) P=0.0
CONVERGENCE OBTAINED IN 5 ITERATIONS. R= 0.265.
MAX ABSOLUTE DERIVATIVE=0.98760-07. -2 LOG L= 4074.31.
MODEL CHI-SQUARE= 317.76 WITH 5 D.F. (-2 LOG L.R.) P=0.0

FRACTION OF CONJUGANT PAIRS OF PREDICTED PROBABILITIES AND RESPONSES :0.654
RANK CORRELATION BETWEEN PREDICTED PROBABILITY AND RESPONSE :0.335

FINAL PARAMETER ESTIMATES

VARIABLE	BETA	STD. ERROR	CHI-SQUARE	P	R
INTERCEPT	10.05289643	0.86809310	134.11	.	
INDXN	-9.20897329	0.59249968	241.65	.	-0.234
PAY	0.09546295	0.04702005	4.12	0.0423	0.022
UNEMP	-99.43144785	18.62692274	28.49	0.0000	-0.078
EDSPEC	-0.48775572	0.12139434	16.14	0.0001	-0.057
MOS	0.17945947	0.08319732	4.65	0.0310	0.025

CLASSIFICATION TABLE

		PREDICTED		TOTAL
		NEGATIVE	POSITIVE	
TRUE	NEGATIVE	3131	48	3179
	POSITIVE	821	108	929
TOTAL		3952	156	4108

SENSITIVITY: 11.6% SPECIFICITY: 98.5% CORRECT: 73.8%
FALSE POSITIVE RATE: 30.8% FALSE NEGATIVE RATE: 20.8%

FRACTION OF CONJUGANT PAIRS OF PREDICTED PROBABILITIES AND RESPONSES :0.654
RANK CORRELATION BETWEEN PREDICTED PROBABILITY AND RESPONSE :0.335

Figure 4.2 First Lieutenants (Ground)

shows, by their order of significance, each variable and their corresponding P-value and the Chi Square statistic.

Figure 4.2 shows the final parameter estimates of each variable with its associated P - statistic and regression coefficients of each of the significant variables. These results would seem to indicate that the junior officer is most concerned with his level of success during his initial military contract agreement. Yet another, probably more important factor, is the officer's realization that the Marine Corps is not what he expected. This happens quite early in a young officer's career and his performance and attitude drop off quickly relative to that of his peers. This difference in outlook, and therefore performance, is reflected in the officer's INDEX

TABLE VII		
VARIABLE SIGNIFICANCE, FIRST LIEUTENANTS		
Variable	P - Value	Chi - Square
INDEX	.	241.65
UNEMP	0.0000	28.49
EDSPEC	0.0001	16.14
MOS	0.0310	4.65
PAY	0.0423	4.12

score and a clear distinction is made between the two groups. Because only those officers with the higher scores

LOGIST, interval tables displaying the performance index by percentile were addressed.

2. First Lieutenants (Ground)

a. Prediction Rates

Prediction rates for this group showed varying results. For FY81 actual attrition for ground officers in this data group was 183. LOGIST predicted a total of 186 officers would attrit during FY81 for a prediction rate of 98 percent. Encouraged by these results the authors decided to evaluate FY82 for a more conclusive study. For FY82 the results of the model were far less accurate. Actual attrition for FY82 was 169 while LOGIST provided a forecast of 140. These figures equated to an 83 percent prediction rate or far below the rate achieved with the FY81 model. Figure 4.1 displays 43 observations from the FY81 data group, along with the predicted probability (column P) of each officer leaving the Marine Corps. Below the observations, in the "SUM" column, is a summation of the predicted probabilities, or the total number LOGIST predicts will leave the Marine Corps in FY81 and FY82

b. Significance of Variables

All variables within this population group proved significant at the .05 level. Performance INDEX showed the lowest P-value indicating its high level of significance to the model while PAY was the least significant with a P-value closest to .05. Table VII

areas. First, the authors fit the models to the FY77 through FY80 data using the LOGIST function. This was accomplished by the LOGIST procedure, using the stepwise mode, which provided statistical data on those variables remaining in the model as well as those deleted because they failed at the .05 level. The estimated models resulting from this procedure provided the output that was then used to produce a probability, or P-value, for each individual to be tested in the FY81 and FY82 forecast years. The summation of these individual probabilities gave the predicted attrition for the population group being analyzed. Comparing the predicted number with the actual figures provided by HQMC, the authors were able to calculate a prediction rate for each population group. For consistency, this prediction rate was calculated as an absolute value, always using the lower of the two numbers (actual or predicted) as the numerator.

Because the primary purpose of this analysis was to evaluate the predictive capabilities of the logit model of Marine officer attrition, prediction rates are presented first. Then the significance of the variables in the estimated models are discussed. To assist the reader, the authors have included in the discussion that follows for each population group, a table ranking the variables in order of significance by P-value of their associated Chi Square tests. In addition, using a procedure unrelated to

TABLE VI
GROUP/VARIABLE CORRELATION

Data Group	Variables Used
1st Lt (Gnd)	INDEX, PAY, UNEMP(Mgr) MOS, EDSPEC
Capt (Gnd)	same as 1st Lt (Gnd)
Capt (Air)	INDEX, PAY, UNEMP(Tech) EDSPEC
Maj (Gnd)	same as Capt (Gnd)
Maj (Air)	same as Capt (Air)
Lt Col (Gnd)	same as Capt (Gnd)
Lt Col (Air)	same as Capt (Air)
Col	INDEX, PAY, UNEMP(Mgr) EDSPEC

(Infantry and Artillery) and the remainder of the population. For colonels, the authors eliminated any MOS distinction because these officers have made a critical transition to an executive level where MOS loses its significance. Because the aviation population groups were tested independent of the ground officers, MOS was not a factor and therefore eliminated.

E. OUTPUT

1. General

Once the data groups were established, the authors, using the LOGIST procedure, analyzed the data groups in two

ground communities were a different breed, separated not only geographically but also by career goals.

Summarizing, Table VI shows the population groups that were evaluated along with the independent variables associated with each group.

TABLE V					
SIGNIFICANT YEARS BY RANK					
Data Group	3	5	11	20	26
1st Lt	183	-	-	-	-
Capt	-	118	-	-	-
Maj	-	-	4	-	-
Lt Col	-	-	-	47	-
Col	-	-	-	-	16

As indicated above, a different variable set was utilized for each group. In the case of UNEMPLOYMENT, the complexity of assigning a technical unemployment rate among the ground community required that all ground officers be given the managerial unemployment rate. On the other hand, aviators with their known technical background were assigned the technical unemployment rate. Similarly, MOS was assigned only within the ground community from first lieutenant to lieutenant colonel, because a distinction could be made in these data groups between the primary combat arms MOS's

Through this transformation, the variable YEARS provided the authors with the ability to further separate the RANK populations by the number of years an officer had been on active duty. This procedure greatly expanded the number of population groups which could be evaluated. To reduce the complexity of this problem, a program was written to show the frequency of attrition by year and rank. These data proved useful in evaluating the final results, as well as indicating which years by rank were the critical years for attrition.

The authors, using the results of this analysis, determined that a satisfactory evaluation of Marine officer attrition could be accomplished by studying five years by rank groupings. The significant year is that point in an officer's career when the greatest number of his peers leave the Marine Corps. As shown in Table V, the significant years for each rank were: First lieutenants, year 3; captains, year 5; majors, year 11; lieutenant colonels, year 20; colonels, year 26. Also depicted in Table V are the actual attrition figures for FY 81, the year used to establish the validity of the LOGIST model.

Once this decision was reached, the only remaining factor was to determine whether ground and aviation officers should be separated and tested independently. The decision was, like the Esmann model, to test these groups independently. The authors realized that the aviation and

analyzed in the time allotted and still provide a accurate projection of officer attrition.

2. Biased Data Elimination

In keeping with the Esmann study, the authors eliminated all ground second lieutenants from the analysis because their initial contracts would require them to remain on active duty until after they would have been promoted to first lieutenant. By the same criterion, aviators were also not evaluated until they had reached the rank of captain.

Additionally, the authors eliminated all limited duty officers and warrant officers. As explained in Chapter II, these officers have normally amassed tenure from prior enlisted service and their committment to remain on active duty is strengthened by their selection for promotion to officer status.

3. Model Population Groups

The initial intention was to run the independent variables against the dependent variable "OUT" by rank. However, further study clearly showed that to run this model against a population categorized by rank only would give a very biased result. A captain with 6 years of service faced with the decision to leave the Marine Corps is influenced by different factors than a captain with 9 years of service. As discussed in Chapter II, the variable YEARS was created from the basic data by subtracting an individual's active duty base date (ADBDY) from the data cycle year (CYC YR).

indication of his career potential in the Marine Corps. Those officers with a PMOS of 0302 (infantry) and 0802 (artillery) would seem to be taking the "ideal" career path of a Marine ground officer. Both authors have observed that officers without the above specialties did perceive themselves to be less promotable and therefore more likely to attrit when the opportunity arose.

C. DEPENDENT VARIABLE

The dependent variable for this model was entitled "OUT". This variable positioned a 1 in every record which had a separation date and a 0 in those records for which there was no separation date.

D. SCOPE OF STUDY

1. General

In order to remove data that would bias the outcome and to limit the scope of the study to a reasonable number of population groups, two decisions were required prior to the application of the logit function to the data. First, it was determined that a large portion of the population, for which the decision to attrit was either insignificant or not a factor at all, would have to be eliminated from the data base. Second, once this group was deleted from the data, a decision was required to determine which of the many possible rank by years-of-service population groups could be

variables to the model so that their impact could be evaluated. Several variables were selected from the initial study conducted by Major Esmann. He found the unemployment variable, both technical and managerial, to be highly significant in predicting attrition, and the authors felt it could only strengthen the binary choice model. While unemployment was the only strong predictor in the Esmann study, it was believed that one other variable, PAY comparability, might prove significant when combined with INDEX and UNEMPLOYMENT.

Along with performance index, unemployment (both managerial and technical), and pay comparability, two additional variables were identified as possible factors in predicting officer attrition. First, education major was selected because of its ability to determine the "employability" of an individual in the civilian sector. The authors felt that those officers with a technical background would have a better opportunity for gaining outside employment while those with non-technical backgrounds would seek the security of the service. Initial perceptions were that this variable would prove applicable only to the junior officers on their initial service commitment. The significance of a technical degree would lose its viability as the officer moved further away from an entry level position in a civilian firm. Second, the authors felt that the PMOS of an officer would be an

IV. ANALYSIS

A. GENERAL

The initial step of the analysis was to perform a preliminary study of the basic data. Through this review, the authors selected and, if necessary, transformed the independent variables which they felt would prove useful in predicting attrition. Along with this procedure, the dependent variables were transformed from the basic data to reflect a binary choice decision. With this accomplished, the basic data were purged to eliminate biased information. Through these procedures, a data base was created from which eight population groups were selected and evaluated using the SAS LOGIST procedure.

B. INDEPENDENT VARIABLES

As indicated in Chapter III, manipulation of the data was required to create a file which would contain the significant factors to conduct the analysis. Once the data base was transformed into a workable file, the independent variables to be used in the study were selected on the basis of intuitive decisions by the authors.

The first variable selected was INDEX. This variable was the primary reason for conducting this follow-on analysis of officer attrition. However, the binary choice model enabled the authors to add additional independent

salary to their peers in the civilian sector and therefore they perceive their pay to be adequate.

c. Performance Index Interval Table

Figure 4.3 shows ten equal sub-group intervals within the first lieutenant population group. These sub-groups show actual attrition rates for each index interval during the period FY77 through FY80. For this data group only one interval shows an "out" (1) percentage which is greater than the "in" (0) percentage. This is interval 1, or all those officers with a performance index less than or equal to .796. At the other end of the scale these percentages change, showing that officers with the higher index have the greatest retention rates. This table clearly indicates that for this data group, the Marine Corps is retaining those officers with the higher performance index while those at the extreme bottom are attriting at a significant rate.

3. Captains (Ground)

a. Prediction Rates

Like the previous data group, the prediction rate for captains in the ground community for FY81 was extremely high. Actual attrition for FY81 was 51 officers while LOGIST predicted 50 officers would attrit for a 98 percent prediction rate. As with the previous data group, the model was also evaluated with FY82 data in order to confirm its validity. The FY82 results were far less

LIEUTENANT (GND) WITH 3 YEARS				
TABLE OF INTERVAL BY OUT				
INTERVAL	OUT	0	1	TOTAL
FREQUENCY				
PERCENT				
ROW PCT				
COL PCT				
INTERVAL 1	0	182	204	386
		4.41	4.95	9.36
		47.15	52.35	
		5.71	21.82	
INTERVAL 2	.796879	306	141	447
		7.42	3.42	10.84
		68.46	31.54	
		9.59	15.08	
INTERVAL 3	.840549	332	104	436
		8.05	2.52	10.57
		76.15	23.85	
		10.41	11.12	
INTERVAL 4	.862793	342	92	434
		8.29	2.23	10.52
		78.80	21.20	
		10.72	9.84	
INTERVAL 5	.879596	361	70	431
		8.75	1.70	10.45
		83.76	16.24	
		11.32	7.49	
INTERVAL 6	.893814	332	83	415
		8.05	2.01	10.06
		80.00	20.00	
		10.41	8.88	
INTERVAL 7	.908022	340	68	408
		8.24	1.65	9.89
		83.33	16.67	
		10.66	7.27	
INTERVAL 8	.921425	324	64	388
		7.95	1.55	9.41
		83.51	16.49	
		10.16	6.84	
INTERVAL 9	.935459	354	54	408
		8.58	1.31	9.89
		86.76	13.24	
		11.10	5.78	
INTERVAL 10	.955157	317	55	372
		7.68	1.33	9.02
		85.22	14.78	
		9.94	5.88	
TOTAL	1.0	3190	935	4125
		77.33	22.67	100.00

Figure 4.3 First Lieutenants (Ground)

encouraging than those for FY81. Attrition during FY82 was 43 officers while LOGIST forecasted that 32 officers would leave the Corps. This equated to a 74 percent prediction rate. Figure 4.4 shows the results of this analysis.

b. Significance of Variables

With this particular data group, only two variables remain in the model at the .05 significance level. As discussed below, these variables, INDEX and EDSPEC, play a significant role in an officer's decision to attrit from the Marine Corps. Here again, the negative coefficient associated with the INDEX variable indicates that officers with higher performance index scores are much more likely to remain on active duty. The remaining variables used (PAY, UNEMPLOYMENT, and MOS) were deleted from the model because they proved to be insignificant at the .05 level. Table VIII shows the variables used to predict attrition in this group ranked from those proving to be the most significant to those having no impact on the model's ability to predict.

The final parameter estimates along with the variables deleted from the model are shown in Figure 4.5. As in the previous data group, INDEX is again the most significant variable. The output from these two data groups would seem to indicate that an officer's performance level plays a major role in his decision to either leave or stay in the Marine Corps. The other variable in the model, EDSPEC, may remain a factor because these officers, with

CAPTAINS (GND) WITH 5 YEARS
STEPWISE LOGISTIC REGRESSION PROCEDURE

DEPENDENT VARIABLE: OUT

1371 OBSERVATIONS
1190 INLT = 0
181 OUT = 1
434 OBSERVATIONS DELETED DUE TO MISSING VALUES

VARIABLE	MEAN	MINIMUM	MAXIMUM	RANGE
INDXN	0.903236	0.656302	0.994171	0.33789
PAY	-6.07644	-7.3	-4.5	2.8
UNEMP	0.024051	0.021	0.0295	0.0085
EDSPEC	1.18891	1	2	1
MOS	1.40481	1	2	1

-2 LOG LIKELIHOOD FOR MODEL CONTAINING INTERCEPT ONLY= 1069.95

STEP 0. THE FOLLOWING VARIABLES ARE ENTERED:

INTERCEPT INDXN PAY UNEMP EDSPEC MOS

MODEL CHI-SQUARE= 58.97 WITH 5 D.F. (SCORE STAT.) P=0.0
CONVERGENCE OBTAINED IN 5 ITERATIONS. R= 0.205.
MAX ABSOLUTE DERIVATIVE=0.12650-06. -2 LOG L= 1014.83.
MODEL CHI-SQUARE= 55.12 WITH 5 D.F. (-2 LOG L.R.) P=0.0

FRACTION OF CONCORDANT PAIRS OF PREDICTED PROBABILITIES AND RESPONSES :0.636
RANK CORRELATION BETWEEN PREDICTED PROBABILITY AND RESPONSE :0.316

FAST BACKWARD ELIMINATION

VARIABLE DELETED	CHI-SQUARE	P	RESIDUAL CHI-SQUARE	D.F.	P
PAY	0.36	0.5458	0.36	1	0.2453
UNEMP	2.67	0.1020	3.04	2	0.2139
MOS	3.24	0.0720	6.27	3	0.0990

STEP 1.

MODEL CHI-SQUARE= 52.95 WITH 2 D.F. (SCORE STAT.) P=0.0
CONVERGENCE OBTAINED IN 5 ITERATIONS. R= 0.204.
MAX ABSOLUTE DERIVATIVE=0.19030-07. -2 LOG L= 1021.30.
MODEL CHI-SQUARE= 48.66 WITH 2 D.F. (-2 LOG L.R.) P=0.0

FRACTION OF CONCORDANT PAIRS OF PREDICTED PROBABILITIES AND RESPONSES :0.624
RANK CORRELATION BETWEEN PREDICTED PROBABILITY AND RESPONSE :0.296

TEST FOR ALL VARIABLES NOT IN THE MODEL
RESIDUAL CHI-SQUARE= 6.32 WITH 3 D.F. P=0.0970

RESIDUAL CHI-SQUARE IS NOT SIGNIFICANT AT THE 0.0500 LEVEL.

CAPTAINS (GND) WITH 5 YEARS
STEPWISE LOGISTIC REGRESSION PROCEDURE

DEPENDENT VARIABLE: OUT

FINAL PARAMETER ESTIMATES

VARIABLE	BETA	STD. ERROR	CHI-SQUARE	P	R
INTERCEPT	5.38333053	1.63809552	10.80	0.0010	
INDXN	-9.30583703	1.84010268	26.14	0.0000	-0.150
EDSPEC	0.87254874	0.17913430	23.71	0.0000	0.142

Figure 4.5 Captains (Ground)

TABLE VIII
VARIABLE SIGNIFICANCE, CAPTAINS (GND)

Variable	P - Value	Chi - Square
INDEX	0.0000	26.14
EDSPEC	0.0000	23.71
MOS	0.0720	6.27
UNEMP	0.1020	3.04
PAY	0.5458	0.36

five years of service, are very employable in the civilian market based on their educational background and job experience. Supporting this idea is the positive coefficient for the EDSPEC variable. Unlike the first lieutenant population group, this positive relationship indicates that those officers with the more technical educations have a higher probability of leaving the service.

A number of variables were deleted from the model. MOS and PAY may have proved insignificant for the same reasons they were less than totally significant in the previous data group. The deletion of the UNEMPLOYMENT variable may explain the reason EDSPEC becomes significant to the model. A technical education, along with five years of experience as a military manager, may make civilian employment more available for this group than the national unemployment rate indicates.

c. Performance Index Interval Table

This table again shows the highest attrition rate occurring at interval 1, or those officers having a performance index less than or equal to .843051. Indicated is that out of 111 officers in this sub-group, 40 have left the service for a 36 percent attrition rate. The remaining intervals are generally within the same range with interval 2 showing an "in" rate of 87.5 percent with a 12.5 percent "out" rate and interval 10 having an 88 percent "in" rate with an 11.9 percent "out" rate. This table clearly shows that those officers with the lowest performance index, interval 1, attrit at the greatest rate. Figure 4.6 shows the interval table with corresponding attrition rates.

4. Captains (Air)

a. Prediction Rates

Unlike the previous data groups analyzed, LOGIST was unsuccessful in predicting attrition for both FY81 and FY82. Actual attrition for FY81 was 67 officers. The LOGIST sum predicted 29 officers would attrit for a 43 percent prediction rate. For FY82 the actual attrition was 26 officers while LOGIST forecasted 6 officers would attrit for a 23 percent prediction rate. Figure 4.7 depicts the predicted attrition rates provided by LOGIST. The reasons for the model's poor performance are discussed following Table IX.

CAPTAINS (GND) WITH 5 YEARS				
TABLE OF INTERVAL BY OUT				
INTERVAL	OUT			
FREQUENCY				
PERCENT				
NOM PCT				
CCL PCT				
INTERVAL	0	0	1	TOTAL
INTERVAL 1		71	40	111
		5.18	2.92	8.10
		63.96	36.04	
		5.97	22.10	
INTERVAL 2	.343051	91	13	104
		6.64	0.93	7.59
		87.50	12.50	
		7.65	7.18	
INTERVAL 3	.862186	115	20	135
		8.39	1.46	9.85
		85.19	14.81	
		9.66	11.05	
INTERVAL 4	.878630	125	16	141
		9.12	1.17	10.28
		88.65	11.35	
		10.50	8.84	
INTERVAL 5	.891161	98	13	111
		7.15	0.95	8.10
		88.29	11.71	
		8.24	7.18	
INTERVAL 6	.900515	132	8	140
		9.63	0.58	10.21
		94.29	5.71	
		11.09	4.42	
INTERVAL 7	.911001	117	11	128
		8.53	0.80	9.34
		91.41	8.59	
		9.83	6.08	
INTERVAL 8	.920912	140	19	159
		10.21	1.39	11.60
		88.05	11.95	
		11.76	10.50	
INTERVAL 9	.934047	154	21	175
		11.23	1.53	12.76
		88.00	12.00	
		12.94	11.60	
INTERVAL 10	.950598	147	20	167
		10.72	1.46	12.18
		88.02	11.98	
		12.35	11.05	
TOTAL	1.0	1190	181	1371
		86.80	13.20	100.00

Figure 4.6 Captains (Ground)

PARAMETER ESTIMATES AND 95% CONFIDENCE LIMITS													
14:21 THURSDAY, MAY 23, 1985													
PARAMETER	ESTIMATE	95% CL	95% CL	STD. ERROR	MAXIMUM	MINIMUM	PREDICTION	STD. ERROR	SUM	VARIANCE	INTERVAL	LOWER	UPPER
MEAN	0.12237594	0.06245207	0.06415844	0.00403126	0.40057957	0.00403126	25.44346520	0.00390026					
MEAN	0.03261768	0.01651340	0.01013743	0.00114701	0.09514230	0.00114701	0.00020613						

CAPTAINS (AIR) FY81 PREDICTION RATE RESULTS													
N	MEAN	STD. DEVIATION	MINIMUM	MAXIMUM	STD. ERROR	SUM	VARIANCE	INTERVAL	LOWER	UPPER	INTERVAL	LOWER	UPPER
240	0.12237594	0.06245207	0.06415844	0.40057957	0.00403126	25.44346520	0.00390026						

CAPTAINS (AIR) FY82 PREDICTION RATE RESULTS													
N	MEAN	STD. DEVIATION	MINIMUM	MAXIMUM	STD. ERROR	SUM	VARIANCE	INTERVAL	LOWER	UPPER	INTERVAL	LOWER	UPPER
272	0.03261768	0.01651340	0.01013743	0.09514230	0.00114701	0.00020613							

Figure 4.7 Captains (Air)

b. Significance of Variables

In the case of captains (Air), all variables remain in the model with INDEX again proving the most significant. The variables are shown in Table IX with the P-value and the associated Chi Squares.

TABLE IX		
VARIABLE SIGNIFICANCE, CAPTAINS (AIR)		
Variable	P - Value	Chi - Square
INDEX	0.0000	26.14
PAY	0.0016	10.01
UNEMP	0.0054	7.74
EDSPEC	0.0315	4.62

Figure 4.8 shows the variables and pertinent statistics provided through the LOGIST procedure. While the LOGIST model proved unsuccessful in predicting attrition for this group, an analysis of the variables may give some insight as to its failure. Like the ground community, the aviation element places a high level of significance on performance. However, PAY becomes more significant within this population group. This is reflected by the negative coefficient associated with the PAY variable which shows a direct connection between higher military pay and retention. This may explain the need of the Marine Corps to pay bonus dollars in order to retain these officers. The UNEMPloyment

CAPTAINS (AIR) WITH 5 YEARS
STEPWISE LOGISTIC REGRESSION PROCEDURE

DEPENDENT VARIABLE: OUT

793 OBSERVATIONS
583 OUT = 0
210 OUT = 1
244 OBSERVATIONS DELETED DUE TO MISSING VALUES

VARIABLE	MEAN	MINIMUM	MAXIMUM	RANGE
INDXN	0.892373	0.694634	0.993156	0.298522
PAY	-5.95183	-7.3	-4.5	2.8
UNEMP	0.0269552	0.0245	0.031	0.0065
EUSPEC	1.25221	1	2	1

-2 LOG LIKELIHOOD FOR MODEL CONTAINING INTERCEPT ONLY= 916.76

STEP 0. THE FOLLOWING VARIABLES ARE ENTERED:

INTERCEPT INDXN PAY UNEMP EUSPEC

MODEL CHI-SQUARE= 59.30 WITH 4 D.F. (SCORE STAT.) P=0.0
CONVERGENCE OBTAINED IN 5 ITERATIONS.
MAX ABSOLUTE DERIVATIVE=0.35510-08. -2 LOG L= 851.12.
MODEL CHI-SQUARE= 59.64 WITH 4 D.F. (-2 LOG L.R.) P=0.0

PERCENTAGE OF CONCORDANT PAIRS OF PREDICTED PROBABILITIES AND RESPONSES :0.651
SPEARMAN CORRELATION BETWEEN PREDICTED PROBABILITY AND RESPONSE :0.325

FINAL PARAMETER ESTIMATES

VARIABLE	BETA	STD. ERROR	CHI-SQUARE	P	R
INTERCEPT	1.71863564	3.38581687	0.26	0.6117	
INDXN	-12.75030682	2.06195064	38.24	0.0000	-0.199
PAY	-0.49904236	0.15774012	10.01	0.0016	-0.093
UNEMP	188.28116708	67.67093328	7.74	0.0054	0.079
EUSPEC	0.40481351	0.18824171	4.62	0.0315	0.054

CLASSIFICATION TABLE

		PREDICTED		TOTAL
		NEGATIVE	POSITIVE	
TRUE	NEGATIVE	572	11	583
	POSITIVE	184	26	210
TOTAL		756	37	793

SENSITIVITY: 12.4% SPECIFICITY: 98.1% CORRECT: 75.4%
FALSE POSITIVE RATE: 29.7% FALSE NEGATIVE RATE: 24.3%

PERCENTAGE OF CONCORDANT PAIRS OF PREDICTED PROBABILITIES AND RESPONSES :0.651
SPEARMAN CORRELATION BETWEEN PREDICTED PROBABILITY AND RESPONSE :0.325

Figure 4.8 Captains (Air)

variable remains significant while the variable closest to .05 is EDSPEC. Although EDSPEC remains in the model, it has the least significance compared to the captain (Ground) model. This may be the result of all of the officers in this group having a technical education. One interesting observation is that in this population group, all variables remain in the model. Similar results were achieved within the first lieutenant model. Both models show officers making a decision on whether to remain in the service after their initial contract obligation.

c. Performance Index Interval Table

Figure 4.9 shows the interval table with the corresponding percentages. This interval table, like the ones that precede it, indicates that attrition is greatest in the first interval. Interval 1, depicting those officers with an INDEX score less than or equal to .840127, shows a 53 percent attrition rate in this population sub-group. While attrition is greatest at interval 1, the next three intervals also show a significant attrition rate when compared with the remaining intervals in the table. The combined attrition rate for these three intervals is 31 percent. Again, this table shows attrition is greatest for those with the lower performance index scores.

5. Majors (Air and Ground)

Although the LOGIST procedure was utilized to analyze these two data groups, the attrition rates were too small to

CAPTAINS (AIR) WITH 5 YEARS				
TABLE OF INTERVAL BY OUT				
INTERVAL	OUT	0	1	TOTAL
FREQUENCY				
PERCENT				
ROW PCT				
COL PCT				
INTERVAL 1	0	43	50	93
		5.40	6.27	11.67
		46.24	53.76	
		7.35	23.58	
INTERVAL 2	.840127	55	24	79
		6.90	3.01	9.91
		69.62	30.38	
		9.40	11.32	
INTERVAL 3	.859267	56	24	80
		7.03	3.01	10.04
		70.00	30.00	
		9.57	11.32	
INTERVAL 4	.872385	48	25	73
		6.02	3.14	9.16
		65.75	34.25	
		8.21	11.79	
INTERVAL 5	.884252	61	13	74
		7.65	1.63	9.28
		82.43	17.57	
		10.43	6.15	
INTERVAL 6	.893682	57	10	67
		7.15	1.25	8.41
		85.07	14.93	
		9.74	4.72	
INTERVAL 7	.904269	61	21	82
		7.65	2.63	10.29
		74.39	25.61	
		10.43	9.91	
INTERVAL 8	.915116	54	14	68
		6.7	1.76	8.53
		79.	20.59	
		9.	6.60	
INTERVAL 9	.925926	9.	17	26
		81.	2.13	11.7
		12.99	18.38	
			8.02	
INTERVAL 10	.944003	74	14	88
		9.28	1.76	11.04
		84.09	15.91	
		12.65	6.60	
TOTAL	1.0	585	212	797
		73.40	26.60	100.00

Figure 4.9 Captains (Air)

provide a worthwhile analysis. Actual attrition for FY81 for both the air and ground communities was four officers! During FY82 a total of five officers within these data groups left the Marine Corps. A review of majors with 11 years of service showed similar attrition rates from FY77 through FY83. With this level of attrition, the authors believed a worthwhile analysis of the data was impossible and discontinued any further study of this group.

6. Lieutenant Colonels (Ground)

a. Prediction Rates

The LOGIST procedure provided poor FY81 and FY82 prediction rates for this data group. For FY81, actual attrition was 34 Marine officers while LOGIST predicted 45 officers would leave the Marine Corps, or a 76 percent prediction rate. The results for FY82 were even less promising. Actual attrition was 15 officers compared to a LOGIST prediction of 44, or a 34 percent prediction rate. Figure 4.10 shows the results of this model for both FY81 and FY82.

b. Significance of Variables

Figure 4.11 shows the final parameter estimates for this model. This data group retained two variables in the construction of the model. As with all of the other data groups studied, INDEX was the most significant. The other variable retained in the model was MOS, continuing the trend (discussed below) that in the ground community at

V. SUMMARY

A. GENERAL

This chapter will consolidate the results of the output discussed in the previous chapter and offer some comments and recommendations regarding further research in the area of predicting Marine officer attrition. Figure 5.1 depicts, by data group, the prediction sums developed by the SAS LOGIST function along with actual attrition for FY81 and FY82. These figures were transformed into a prediction rate which is also included in Figure 5.1. Additionally, variables which were retained in the model are listed in order of significance while those deleted are presented in a similar format.

B. PREDICTION RATES

The overall results of this binary choice model, using the SAS LOGIST procedure, were not encouraging. For FY81, three of the five data groups analyzed gave a prediction rate over 90 percent. At first glance, an ability to predict attrition of Marine Officers at a 90 percent or better rate would seem to validate the models and thus provide a useful tool for manpower planning. But, in all three cases, when these same population groups were used to forecast FY82 attrition, the prediction rates dropped well below the 90 percent benchmark achieved for the FY81 data.

COLONELS (ALL) WITH 26 YEARS

TABLE OF INTERVAL BY OUT

INTERVAL	OUT	0	1	TOTAL
FREQUENCY				
PERCENT				
ROW PCT				
CUL PCT				
INTERVAL 1	0	27 7.63 77.14 9.82	8 2.26 22.86 10.13	35 9.89
INTERVAL 2	.857835	30 8.47 83.33 10.91	6 1.69 16.67 7.59	36 10.17
INTERVAL 3	.868889	27 7.63 77.14 9.82	8 2.26 22.86 10.13	35 9.89
INTERVAL 4	.877647	27 7.63 77.14 9.82	8 2.26 22.86 10.13	35 9.89
INTERVAL 5	.836073	26 7.34 72.22 9.43	10 2.82 27.78 12.66	36 10.17
INTERVAL 6	.894349	23 7.91 80.00 10.13	7 1.98 20.00 8.36	35 9.89
INTERVAL 7	.903686	31 8.76 86.11 11.27	5 1.41 13.89 5.33	36 10.17
INTERVAL 8	.911153	30 8.47 85.71 10.91	5 1.41 14.29 5.33	35 9.89
INTERVAL 9	.919439	31 8.76 86.11 11.27	5 1.41 13.89 5.33	36 10.17
INTERVAL 10	.933774	13 5.00 51.43 6.55	17 4.80 43.57 21.52	35 9.89
TOTAL	1.0	275 77.68	79 22.32	354 100.00

Figure 4.16 Colonels (All)

TABLE XII

VARIABLE SIGNIFICANCE, COLONELS (ALL)

Variable	P - Value	Chi - Square
PAY	0.0855	2.96
INDEX	0.2001	1.64
UNEMP	0.4870	0.48
EDSPEC	0.5257	0.40

attrition rate of 48 percent for those officers with a score greater than or equal to .933774. The remaining intervals show a substantial decrease in their attrition rates when compared with interval 10. Figure 4.16 shows the results of the interval tables by performance INDEX.

F. SUMMARY

This concludes the data analysis phase of the authors' study of Marine officer attrition. In the following chapter, these results will be consolidated for clarity and the authors will provide comments and recommendations for future studies in this area.

officers with a performance index score less than or equal to .847943, shows retention at 91 percent and attrition at 8.7 percent. Intervals 9 and 10, or those officers with the highest performance index, show attrition rates of 56 and 41 percent respectively. Again, contrary to the earlier groups analyzed, this population group shows that officers with the highest performance indexes are attriting at the greatest rate.

8. Colonels (All)

a. Prediction Rates

With this last data group, there were no variables which passed the significance test. Table XII shows the P-value and Chi-Square statistics of the variables deleted from this model. Because these variables did not enter the model, LOGIST assigned the entire population group a uniform probability of leaving the service. This was due in large part to the relatively small size of the population group. The results of this procedure were discarded because they merely equated to a random chance probability of attrition. Accordingly, further analysis of the variables and prediction rates was not conducted for this group.

b. Performance Index Interval Table

The colonels displayed similar results when compared with the lieutenant colonel's ground and air data groups. Attrition is greatest at the highest performance INDEX interval. That interval, interval 10, shows an

LT COLS (AIR) WITH 20 YEARS				
TABLE OF INTERVAL BY OUT				
INTERVAL	OUT	0	1	TOTAL
FREQUENCY				
PERCENT				
ROW PCT				
COL PCT				
INTERVAL 1	0	21 8.79 91.30 11.17	2 0.84 8.70 3.92	23 9.62
INTERVAL 2	.847943	19 7.95 82.61 10.11	4 1.67 17.39 7.84	23 9.62
INTERVAL 3	.862274	22 9.21 91.67 11.70	2 0.84 8.33 3.92	24 10.04
INTERVAL 4	.872510	20 8.37 90.91 10.64	2 0.84 9.09 3.92	22 9.21
INTERVAL 5	.881697	21 8.79 87.50 11.17	3 1.26 12.50 5.88	24 10.04
INTERVAL 6	.890303	19 7.95 79.17 10.11	5 2.09 20.83 9.80	24 10.04
INTERVAL 7	.900135	23 9.62 92.00 12.23	2 0.84 8.00 3.92	25 10.46
INTERVAL 8	.910602	18 7.53 72.00 9.57	7 2.93 28.00 13.73	25 10.46
INTERVAL 9	.921015	11 4.60 44.00 5.85	14 5.86 56.00 27.45	25 10.46
INTERVAL 10	.932495	14 5.86 58.33 7.45	10 4.18 41.67 19.61	24 10.04
TOTAL	1.0	188 78.66	51 21.34	239 100.00

Figure 4.15 Lieutenant Colonels (Air)

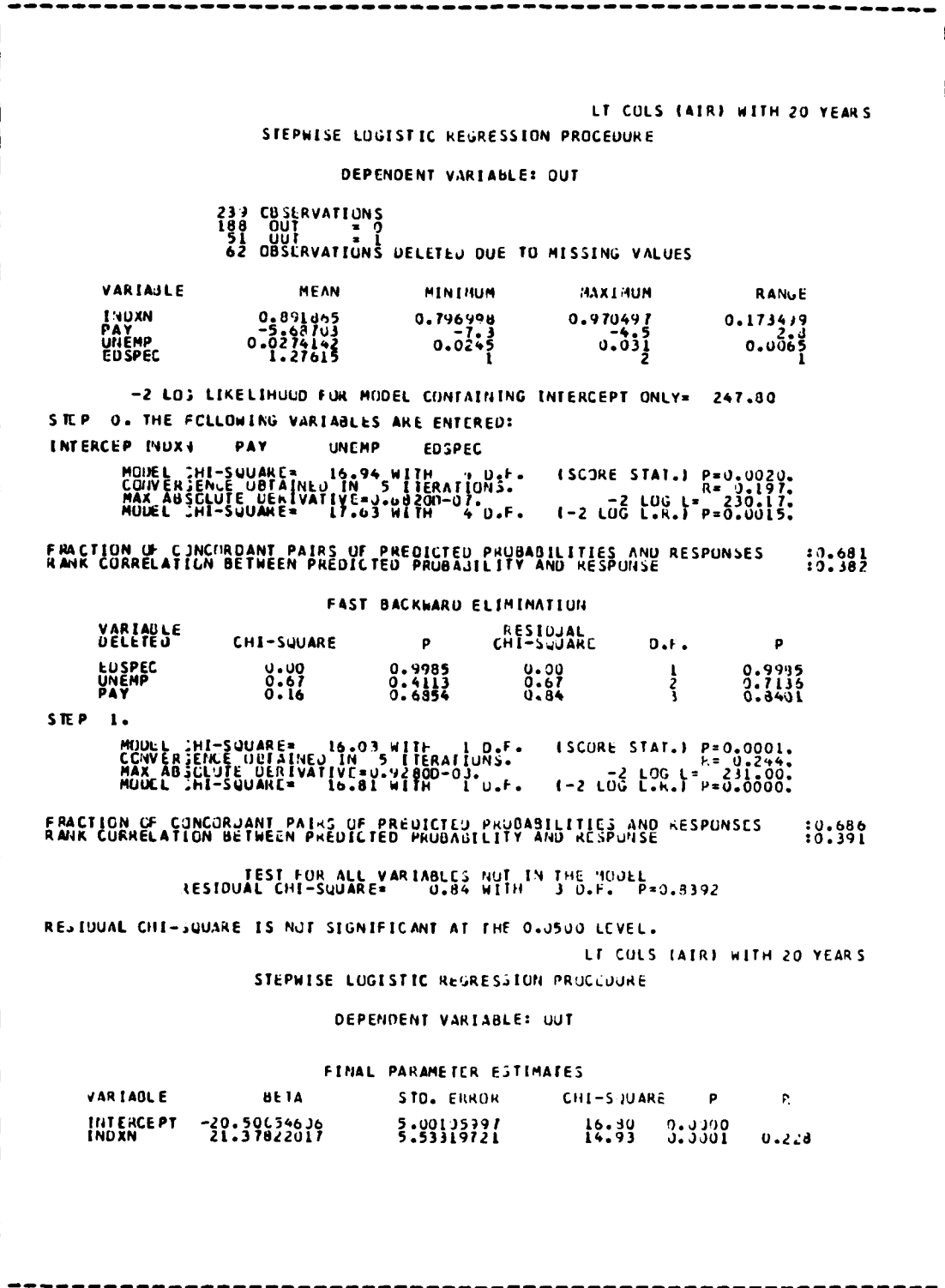


Figure 4.14 Lieutenant Colonels (Air)

The variables evaluated in this model are shown below in Table XI.

INDEX again emerges as the primary variable for determining attrition for this group and, like the lieutenant colonel (Ground) population group, the coefficient is positive. The other variables analyzed by the model were deleted for the same reasons given during the discussion of the previous data group. Figure 4.14 shows

TABLE XI		
VARIABLE SIGNIFICANCE, LIEUTENANT COLONELS (AIR)		
Variable	P - Value	Chi - Square
INDEX	0.0001	14.93
UNEMP	0.4113	0.67
PAY	0.6854	0.16
EDSPEC	0.9985	0.00

the final parameter estimates of this data group along with pertinent statistics on the variables deleted from the model.

c. Performance Index Interval Table

Figure 4.15 shows the table for this population group with attrition and retention percentages by interval. Although the total number of officers appearing in this table is much less than the lieutenant colonel ground table, 239 versus 648, the results are similar. Interval 1,

[illegible]

less than or equal to .838258, have an "in" percentage of 83 percent and an "out" percentage of 16.9. On the other hand, interval 10 has an "out" rate of 52 percent and an "in" rate of 47.6 percent. A review of this table clearly shows that those officers in the higher interval subgroupings attrited at the greatest rate. This would indicate the Marine Corps is losing those officers with the higher performance INDEX scores from this data group.

7. Lieutenant Colonel (Air)

a. Prediction Rates

The application of LOGIST to this population group produced mixed results for FY81 and FY82. The prediction rate for FY81 was 93 percent, with LOGIST forecasting 14 officers against an actual attrition of 13 officers. For FY82, LOGIST predicted 13 officers would leave the Marine Corps when actually 16 officers resigned their commissions. This resulted in a predictability rate of only 81 percent for FY82. Figure 4.13 displays the LOGIST output for this data group.

b. Significance of Variables

INDEX was the only variable that remained in this model. The input variables for this model were similar to those used to analyze the lieutenant colonel (Ground) population group with the exception of MOS. MOS was not used with this group because, like the other groups involving pilots, they had been arranged by MOS initially.

LT COLS (GND) WITH 20 YEARS				
TABLE OF INTERVAL BY OUT				
INTERVAL	OUT	0	1	TOTAL
FREQUENCY				
PERCENT				
ROW PCT				
CUL PCT				
INTERVAL 1	0	54	11	65
		8.33	1.70	10.03
		83.08	16.92	
		11.66	5.95	
INTERVAL 2	.838258	55	10	65
		8.49	1.54	10.03
		84.62	15.38	
		11.88	5.41	
INTERVAL 3	.855389	55	10	65
		8.49	1.54	10.03
		84.62	15.38	
		11.88	5.41	
INTERVAL 4	.866850	69	15	84
		10.65	2.31	12.96
		82.14	17.86	
		14.90	8.11	
INTERVAL 5	.873684	33	14	47
		5.07	2.16	7.23
		70.21	29.79	
		7.13	7.57	
INTERVAL 6	.884532	44	21	65
		6.79	3.24	10.03
		67.69	32.31	
		9.50	11.35	
INTERVAL 7	.892454	48	17	65
		7.41	2.62	10.03
		73.82	26.18	
		10.37	9.19	
INTERVAL 8	.905655	39	25	64
		6.02	3.86	9.88
		60.94	39.06	
		8.42	13.51	
INTERVAL 9	.916338	36	29	65
		5.56	4.48	10.03
		55.38	44.62	
		7.78	15.68	
INTERVAL 10	.929695	30	33	63
		4.63	5.07	9.72
		47.62	52.38	
		6.48	17.84	
TOTAL	1.0	463	185	648
		71.45	28.55	100.00

Figure 4.12 Lieutenant Colonels (Ground)

promotion. Marine officers in this data group with combat arms MOS's have had more command assignments, making their future promotion potential greater than those who have never served in this type of job due to their restrictive MOS's.

Addressing the variables deleted, EDSPEC seems to become less significant as the officer moves further away from his undergraduate years. This supports the authors' original impressions about this variable, believing it would prove significant only for the younger officers. PAY was also deleted. This is because these officers have spent twenty years in the military, and any real concerns with pay would have forced them to attrit at an earlier stage in their careers. UNEMPLOYMENT is just insignificant at a P-value of .05, which may indicate that the civilian positions this group are seeking are not related to the national unemployment rate.

c. Performance Index Interval Table

Figure 4.12 shows the interval table with the corresponding attrition percentages. With this data group, the performance INDEX does a complete reversal from the previous data groups. In the previous populations, attrition was highest among the intervals representing the lowest index scores. A percentile breakdown of the performance INDEX scores of this data group reveals that those officers with the highest scores attrit at the greatest rate. Interval 1, or those officers with a score

least, having a combat rms MOS is perceived to be important to an officer's future in the Marine Corps. Variables deleted from the model were UNEMP, PAY and EDSPEC. Table X shows the variables for this data group.

While INDEX remained the most significant variable throughout this study, the positive coefficient assigned to INDEX shows a distinct break from the preceding

TABLE X		
VARIABLE SIGNIFICANCE, LIEUTENANT COLONELS (GND)		
Variable	P - Value	Chi - Square
INDEX	0.0000	47.72
MOS	0.0000	25.71
UNEMP	0.0500	3.84
PAY	0.0808	3.05
EDSPEC	0.8678	0.03

population groups. This indicates, for the first time in this study, that those officers with the higher performance index scores have a greater probability of attrition. This output supports the data displayed in the Performance Index Interval Table discussed below. Additionally this data group placed a high level of significance on MOS. This would seem to indicate that after reaching the rank of lieutenant colonel, officers begin to regard their job assignments as a means of enhancing their future chances for

LT COLS (GND) WITH 20 YEARS
STEPWISE LOGISTIC REGRESSION PROCEDURE

DEPENDENT VARIABLE: OUT

648 OBSERVATIONS
463 OUT = 0
185 OUT = 1
157 OBSERVATIONS DELETED DUE TO MISSING VALUES

VARIABLE	MEAN	MINIMUM	MAXIMUM	RANGE
INDXN	0.884749	0.789441	0.983742	0.194301
PAY	-5.77731	-7.3	-4.5	2.8
UNEMP	0.0244069	0.021	0.0295	0.0085
EDSPEC	1.16821	1	2	1
MOS	1.60494	1	2	1

-2 LOG LIKELIHOOD FOR MODEL CONTAINING INTERCEPT ONLY= 775.10

STEP 0. THE FOLLOWING VARIABLES ARE ENTERED:

INTERCEP INDXN PAY UNEMP EDSPEC MOS

MODEL CHI-SQUARE= 73.46 WITH 5 D.F. (SCORE STAT.) P=0.0
CONVERGENCE OBTAINED IN 5 ITERATIONS. R= 0.297.
MAX ABSOLUTE DERIVATIVE=0.79010-06. -2 LOG L= 696.74.
MODEL CHI-SQUARE= 78.36 WITH 5 D.F. (-2 LOG L.R.) P=0.0

FRACTION OF CONCORDANT PAIRS OF PREDICTED PROBABILITIES AND RESPONSES :0.707
RANK CORRELATION BETWEEN PREDICTED PROBABILITY AND RESPONSE :0.430

FAST BACKWARD ELIMINATION

VARIABLE DELETED	CHI-SQUARE	P	RESIDUAL CHI-SQUARE	D.F.	P
EDSPEC	0.03	0.8674	0.03	1	0.8674
PAY	3.05	0.0803	3.03	2	0.2148
UNEMP	3.84	0.0500	6.92	3	0.0746

STEP 1.

MODEL CHI-SQUARE= 67.37 WITH 2 D.F. (SCORE STAT.) P=0.0
CONVERGENCE OBTAINED IN 5 ITERATIONS. R= 0.295.
MAX ABSOLUTE DERIVATIVE=0.70750-07. -2 LOG L= 703.81.
MODEL CHI-SQUARE= 71.29 WITH 2 D.F. (-2 LOG L.R.) P=0.0

FRACTION OF CONCORDANT PAIRS OF PREDICTED PROBABILITIES AND RESPONSES :0.698
RANK CORRELATION BETWEEN PREDICTED PROBABILITY AND RESPONSE :0.413

TEST FOR ALL VARIABLES NOT IN THE MODEL
RESIDUAL CHI-SQUARE= 6.99 WITH 3 D.F. P=0.0723

RESIDUAL CHI-SQUARE IS NOT SIGNIFICANT AT THE 0.0500 LEVEL.

LT COLS (GND) WITH 20 YEARS

DEPENDENT VARIABLE: OUT

FINAL PARAMETER ESTIMATES

VARIABLE	BETA	STD. ERROR	CHI-SQUARE	P	R
INTERCEPT	-17.42502086	2.52255510	47.72	0.0000	
INDXN	20.27835271	2.89204992	49.16	0.0000	0.247
MOS	-0.97168114	0.19163512	25.71	0.0000	-0.175

Figure 4.11 Lieutenant Colonels (Ground)

POPULATION GROUP	FY81 LOGIST	FY81 ACTUAL	FY81 RATE	FY82 LOGIST	FY82 ACTUAL	FY82 RATE	RETAINED VARIABLES (by significance)	DELETED VARIABLES
1st Lt (GND)	186	183	98%	140	169	83%	INDEX, UNEMP, EDSPEC MOS, PAY	NONE
Capt (GND)	50	51	98%	32	43	74%	INDEX, EDSPEC	MOS, UNEMP, PAY
Capt (AIR)	29	67	43%	7	26	27%	INDEX, PA EDSPEC	NONE
Maj (ALL)	*** Deleted from analysis due to insufficient data ***							
Lt Col (GND)	45	34	76%	44	15	34%	INDEX, MOS	UNEMP, PAY, EDSPEC
Lt Col (AIR)	14	13	93%	13	16	81%	INDEX	UNEMP, PAY, EDSPEC
Col (ALL)	*** Deleted from analysis because no variable proved significant ***							

Figure 5.1 Consolidated Results

Additionally, the rates obtained for the remaining data groups were far below the 90 percent level for both FY81 and FY82. Since the results of the tests conducted varied from highly successful to totally unacceptable, the logit procedure would appear to contain sufficient inaccuracy to preclude its use.

This analysis however, has provided further insight into the difficult question of predicting officer attrition. This problem is very difficult to solve using statistical models. Surely, there are a number of variables which cannot be easily quantified yet play a major role in an officer's decision to leave the service. Some of those variables, such as family situation, next duty assignment, personal contacts in the civilian job market and self perception, would have to be individually modeled for this type of analysis to prove statistically relevant to the user. Through personal experiences, the authors have consistently heard the above reasons given as to why an officer has chosen to leave the service. This study seems to confirm the diverse nature of this decision.

Future studies in this area should concentrate on evaluating more population groups with additional variables over a longer period of time to provide a wider range for analysis. Only in this manner can the investigator hope to achieve relevant results for this problem through the use of standard statistical techniques.

C. SIGNIFICANCE OF VARIABLES

As shown in Figure 5.1, INDEX was the most significant variable in each of the five population groups analyzed. These results confirmed the perceptions provided by Major Esmann during his initial research. Officers interviewed for that study felt that promotion potential was a major factor when determining whether to remain on active duty or leave the service. Further, they felt that a high promotion potential would indicate satisfaction with the service while a low promotion potential would cause an officer to resign his commission [Ref. 9]. While the performance INDEX proved significant with all of the population groups, it is of particular note that for the three most senior groups analyzed (lieutenant colonels, Air and Ground, and colonels), the variable had a positive coefficient. The inverse relationship of this positive coefficient means that as an individual's INDEX score increases at these higher ranks, so does the probability that he will leave the Marine Corps. Additionally there is the question of whether the performance INDEX equates to promotion potential. One major shortcoming of this variable in relation to promotion potential was that it failed to reflect an officer's level of experience or his past job assignments. The assignment variable in particular is one of the most significant factors considered by promotion boards and is continually

stressed throughout the Marine Corps as a vital ingredient when planning the "ideal" career. In essence, those Marine officers who have been assigned the "better jobs", may in fact have lower INDEX scores because of the competitive nature of those billets. Nevertheless, these officers will still have a greater promotion potential than their contemporaries who received a higher INDEX score from a less career-enhancing billet. Since the INDEX score has such a narrow range, job assignment may in fact be the vital ingredient for determining promotion potential. A review of the GAO study on quality retention of Marine officers shows that a job assignment factor was numerically transformed and combined with a performance indicator in order to provide a more realistic presentation of promotion potential [Ref. 10]. Future studies in this area should include a job assignment factor to weight the performance INDEX when evaluating the effect of promotion potential on attrition.

While the Esmann study concluded that the unemployment variable(UNEMP) was significant in predicting attrition, the SAS LOGIST procedure provided quite the opposite results. In three of the five data groups analyzed, the unemployment variable did not meet the .05 significance level and was deleted from the model. In those cases where it remained, the population groups were completing their initial contracts and reviewing their employment opportunities in the civilian community. These groups, first lieutenant

(Ground) and captain (Air), were facing similiar decisions and the unemployment rate appears to have been an important factor in their decision to leave the Marine Corps. For the remaining three data groups, the effects of the employment situation were of little or no significance.

During the initial phases of this study it was felt that the educational specialty (EDSPEC) variable would be significant only to the more junior population groups whose recently acquired skills remained a marketable asset. This assumption proved correct, with the first lieutenant through captain data groups retaining EDSPEC and the lieutenant colonel data groups deleting this variable. This would seem to indicate that a college major is more important to a young officer entering the civilian market and less important to the career officer who must sell his managerial background rather than what he studied in college twenty years ago.

The MOS variable was input as a independent variable in only three of the five population groups analyzed. Of those three groups, it was retained by only two as being significant, first lieutenants and lieutenant colonels (Ground). The retention of this variable and its degree of significance in the lieutenant colonels (Ground) model should be noted. It may be that MOS is directly correlated to the job assignment factor previously discussed. It would appear that those officers with combat arms MOS's (0302 and

0802) are in a better position to serve in those assignments considered career enhancing while those officers without these MOS's would not have the opportunity to assume the limited number of command billets. While an individual's past assignments were not introduced into this study, the presence of MOS in this particular model may indicate its relationship with promotion potential and therefore attrition.

The last independent variable to be included in this study was pay comparability with the civilian sector. For the most part this variable proved insignificant, being retained in only two of the eight data groups analyzed. However, in the case of the captain (Air), PAY was second to INDEX in its level of significance to the model. Although not a principle factor in this population group's decision to leave the Marine Corps, the appearance of PAY in the captain (Air) model still suggests that money is important to this group and may explain the payment of bonuses to retain the younger naval aviator.

To summarize, the viability of the performance INDEX score as a predictor of promotion potential was not determined during this study. However, its value in predicting attrition was confirmed. The same can not be stated of the other independent variables which nevertheless provided some interesting insights into the various elements affecting an officer's decision to leave the Marine Corps.

D. PERFORMANCE INDEX INTERVALS

The performance index interval tables were originally constructed to assist the authors in determining at what INDEX level the greatest attrition was occurring for each population group. Using this information in conjunction with the SAS LOGIST output, the authors would be able to ascertain at which intervals SAS was having the most success in predicting attrition. From the results provided, the indication is that LOGIST predicts quite well at the extreme ends of the tables while the prediction capabilities in the middle intervals are far less accurate.

Through this process the authors discovered that the senior officers, lieutenant colonels - year 20, and colonels - year 26, with the highest performance index scores were attriting at the greatest rate. Table XII shows the retention and attrition percentages at the interval extremes for each of the population groups analyzed. As this table indicates, the percentage of officers staying in or leaving the Marine Corps compared to their performance index interval is completely reversed between the junior and senior officers. If a high performance index equates to being a good officer, then the significance of this information would be that the Marine Corps is losing its best officers when they reach the senior ranks. Of course, as discussed previously, the performance index may not

TABLE XIII

CONSOLIDATED PERFORMANCE INDEX INTERVALS

Population Group	Interval	Index Value	Percentage In	Percentage Out
1st Lt (GND)	1	<.796879	47%	52%
	10	>.955157	85%	14%
Capt's (GND)	1	<.843051	83%	16%
	10	>.950598	89%	10%
Capt's (AIR)	1	<.840127	59%	40%
	10	>.944003	91%	8%
Lt Col's (GND)	1	<.838258	83%	16%
	10	>.929695	47%	52%
Lt Col's (AIR)	1	<.847943	91%	8%
	10	>.932495	58%	41%
Col's (ALL)	1	<.857835	77%	22%
	10	>.933774	51%	48%

indicate the total professional ability of Marine Corps senior officers because it fails to consider past assignments when evaluating a Marine's worth to the service. Nevertheless, an officer with twenty years of service has had a variety of assignments with different reporting seniors. If he has been able to maintain an above average performance index score throughout that period, it should be a true indicator of his greater ability and increased promotion potential.

The GAO report used a job experience factor along with a performance indicator which showed that the Marine Corps was

retaining its better senior officers. The opposing results achieved by this study and the GAO report indicates that different methods of evaluating promotion potential will yield varying results. Any future studies in this area should begin by identifying the value of the performance index to determine if it is a true indicator of an officer's performance and, therefore, promotion potential.

E. CONCLUSION

The overall success rate of the logit function to predict Marine officer attrition was less than the authors had expected. A possible explanation may lie with the variables selected to describe the individuals in the population groups. The performance INDEX score is clearly a starting point for determining the promotion potential of an officer, but his past assignments cannot be overlooked as a significant factor in this regard. While the remaining variables provide some interesting insights into the differences between the population groups, their usefulness in predicting attrition was generally minimal. In retrospect, the complexity of this decision, to remain on active duty or leave the Corps, may make it impossible to accurately model using currently available data. Only with the inclusion of a number of additional, difficult to obtain or measure variables, can an analyst hope to achieve accurate results across all population groups and years of service.

LIST OF REFERENCES

1. Center for Naval Analysis 83-3116/22, CNA Memorandum: "Marine Officer Loss Rates for 1984," by Kathleen P. Utgoff, July, 1983.
2. Esmann, William J., Marine Officer Attrition Model, M.S. Thesis, Naval Postgraduate School, Monterey, 1984.
3. "The Pay Gap," Navy Times, 34th Year No. 13, p. 3, 14 January 1985.
4. Esmann, William J., Marine Officer Attrition Model, M.S. Thesis, Naval Postgraduate School, Monterey, 1984.
5. Ott, Lyman, and Hildebrand, David K., Statistical Thinking for Managers, pp. 401-493, Duxbury Press, 1983.
6. Eisenbeis, Robert A., and Avery, Robert B., Discriminant Analysis and Classification Procedures, pp. 1-3, Lexington Books, 1972.
7. Annals of Economic and Social Measurement, "A Comment on Discriminant Analysis "Versus" Logit Analysis," by Danial McFadden, May 1976.
8. SAS Institute, SUGI Supplemental Library User's Guide, SAS Institute Inc., 1983.
9. Esmann, William J., Marine Officer Attrition Model, M.S. Thesis, Naval Postgraduate School, Monterey, 1984.
10. Report to the Secretary of the Navy, "High-Quality Senior Marine Corps Officers: How Many Stay Beyond 20 Years of Service?," General Accounting Office, November 1984.

INITIAL DISTRIBUTION LIST

	No. Copies
1. Defense Technical Information Center Cameron Station Alexandria, Virginia 22304-6145	2
2. Library, Code 0142 Naval Postgraduate School Monterey, California 93943-5100	2
3. Professor Dan Boger, Code 54Bk Naval Postgraduate School Monterey, California 93943-5100	1
4. LtCol David Melchar, Code 0309 Naval Postgraduate School Monterey, California 93943-5100	1
5. Commandant of the Marine Corps, Code MPP Headquarters, U. S. Marine Corps Washington, D. C. 20380-0001	1
6. LtCol S. F. Hurst, USMC 9029 Rosewall Court Springfield, Virginia 22152	1
7. Capt T. E. Manion, USMC 9029 Rosewall Court Springfield, Virginia 22152	1
8. Maj W. J. Esmann, USMC 4509 Banf Street Annandale, Virginia 22314	1

END

FILMED

11-85

DTIC