



**Calhoun: The NPS Institutional Archive**  
**DSpace Repository**

---

Theses and Dissertations

1. Thesis and Dissertation Collection, all items

---

1998-03-01

# An analysis of aviation test scores to characterize Student Naval Aviator disqualification

Wahl, Erich J.

Monterey, California. Naval Postgraduate School

---

<http://hdl.handle.net/10945/7887>

---

This publication is a work of the U.S. Government as defined in Title 17, United States Code, Section 101. Copyright protection is not available for this work in the United States.

*Downloaded from NPS Archive: Calhoun*



Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

**Dudley Knox Library / Naval Postgraduate School**  
**411 Dyer Road / 1 University Circle**  
**Monterey, California USA 93943**

<http://www.nps.edu/library>

**NPS ARCHIVE**  
**1998.03**  
**WAHL, E.**

DUDLEY KNOX LIBRARY  
NAVAL POSTGRADUATE SCHOOL  
MONTEREY CA 93943-5101





NAVAL POSTGRADUATE SCHOOL  
Monterey, California



THESIS

AN ANALYSIS OF AVIATION TEST  
SCORES TO CHARACTERIZE STUDENT  
NAVAL AVIATOR DISQUALIFICATION

by

Erich J. Wahl

March, 1998

Thesis Advisor:  
Second Reader:

Robert R. Read  
Samuel E. Buttrey

Approved for public release; distribution is unlimited.

DUDLEY KNOX LIBRARY  
NAVAL POSTGRADUATE SCHOOL  
MONTEREY CA 93943-5101

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.

|  |   |  |  |  |
|--|---|--|--|--|
| <b>1. AGENCY USE ONLY (Leave blank)</b>  |   | <b>2. REPORT DATE</b><br>March 1998                            | <b>3. REPORT TYPE AND DATES COVERED</b><br>Master's Thesis |  |
| <b>4. TITLE AND SUBTITLE</b> An Analysis of Aviation Test Scores To Characterize Student Naval Aviator Disqualification  |   |  | <b>5. FUNDING NUMBERS</b>                                  |  |
| <b>6. AUTHOR(S)</b> Erich J. Wahl  |   |  |  |  |
| <b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b><br>Naval Postgraduate School<br>Monterey, CA 93943-5000  |   |  | <b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>            |  |
| <b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b><br>Chief of Naval Education and Training, Pensacola, FL   |   |  | <b>10. SPONSORING / MONITORING AGENCY REPORT NUMBER</b>    |  |
| <b>11. SUPPLEMENTARY NOTES</b><br>The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.   |   |  |  |  |
| <b>12a. DISTRIBUTION / AVAILABILITY STATEMENT</b><br>Approved for public release; distribution is unlimited.   |   |  | <b>12b. DISTRIBUTION CODE</b>                              |  |
| <b>13. ABSTRACT (maximum 200 words).</b><br>The U.S. Navy uses the Aviation Selection Test Battery (ASTB) to identify those Student Naval Aviator (SNA) applicants most likely to succeed in flight training. Using classification and regression trees, this thesis concludes that individual answers to an ASTB subtest, the Biographical Inventory, are not good predictors of SNA primary flight grades. It also concludes that those SNA who score less than a 6 on the Pilot Biographical Inventory have a significantly higher disqualification rate in primary flight training than those SNA who score a 6 or higher. Those SNA who repeat the taking of the ASTB are more likely to disqualify from primary flight training than those SNA who pass it on the first attempt. Incidentally, significant differences exist in SNA performance and disqualification rates in Aviation Preflight Indoctrination among different racial groups. However, neither race nor gender is a significant factor in primary flight-training disqualification. Recommendations are provided to reduce the number of SNA entering the flight-training pipeline, if necessary, while significantly reducing the disqualification rate. Additionally, a method is given to identify those SNA most at risk of disqualifying from primary flight training. |   |  |  |  |
| <b>14. SUBJECT TERMS</b> Attributes, Attrition, CART, Cross-Validation, Disqualification, Partition, Tree  |   |  | <b>15. NUMBER OF PAGES</b><br>86                           |  |
|  |   |  | <b>16. PRICE CODE</b>                                      |  |
| <b>17. SECURITY CLASSIFICATION OF REPORT</b><br>Unclassified   | <b>18. SECURITY CLASSIFICATION OF THIS PAGE</b><br>Unclassified | <b>19. SECURITY CLASSIFICATION OF ABSTRACT</b><br>Unclassified | <b>20. LIMITATION OF ABSTRACT</b><br>UL                    |  |





Approved for public release; distribution is unlimited

AN ANALYSIS OF AVIATION TEST SCORES TO  
CHARACTERIZE STUDENT NAVAL AVIATOR DISQUALIFICATION

Erich J. Wahl  
Lieutenant, United States Navy  
B.S., United States Naval Academy, 1988

Submitted in partial fulfillment of the  
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL  
March 1998



**ABSTRACT**

The U.S. Navy uses the Aviation Selection Test Battery (ASTB) to identify those Student Naval Aviator (SNA) applicants most likely to succeed in flight training. Using classification and regression trees, this thesis concludes that individual answers to an ASTB subtest, the Biographical Inventory, are not good predictors of SNA primary flight grades. It also concludes that those SNA who score less than a 6 on the Pilot Biographical Inventory have a significantly higher disqualification rate in primary flight training than those SNA who score a 6 or higher. Those SNA who repeat the taking of the ASTB are more likely to disqualify from primary flight training than those SNA who pass it on the first attempt. Incidentally, significant differences exist in SNA performance and disqualification rates in Aviation Preflight Indoctrination among different racial groups. However, neither race nor gender is a significant factor in primary flight-training disqualification. Recommendations are provided to reduce the number of SNA entering the flight-training pipeline, if necessary, while significantly reducing the disqualification rate. Additionally, a method is given to identify those SNA most at risk of disqualifying from primary flight training.

STATE OF CALIFORNIA  
COUNTY OF MONTEREY  
SUPERIOR COURT

IN RE: [Illegible]

[The following text is extremely faint and illegible, appearing to be a legal document or court order.]

## TABLE OF CONTENTS

|      |  |    |
|------|--|----|
| I.   | INTRODUCTION.....  | 1  |
| A.   | OBJECTIVE.....   | 1  |
| B.   | METHODOLOGY.....   | 1  |
| C.   | ORGANIZATION.....  | 7  |
| II.  | BACKGROUND.....  | 9  |
| A.   | HISTORY OF THE NAVAL AVIATION SELECTION TEST.....  | 9  |
| B.   | OVERVIEW OF THE NAVAL FLIGHT-TRAINING PROGRAM.....                                       | 10 |
| C.   | DISQUALIFICATION.....  | 12 |
| III. | DATA.....  | 17 |
| A.   | DATA COLLECTION AND LOCATION.....  | 17 |
| B.   | DATA AVAILABLE FOR ANALYSIS.....   | 17 |
| IV.  | ANALYSIS.....  | 19 |
| A.   | DETERMINING COMPOSITION OF API DATA.....   | 19 |
| B.   | ANALYSIS OF EFFECT OF TIME ON DISQUALIFICATION.....                                      | 23 |
| C.   | PREDICTING FLIGHT GRADES AND DISQUALIFICATION<br>RATES FROM TEST ANSWERS AND SCORES..... | 25 |
| D.   | REPEAT TESTING.....  | 47 |
| V.   | COST CONSIDERATIONS.....   | 57 |
| VI.  | CONCLUSIONS AND RECOMMENDATIONS.....   | 59 |
| A.   | CONCLUSIONS.....   | 59 |
| B.   | RECOMMENDATIONS.....   | 60 |
|      | APPENDIX. LIST OF S-PLUS COMMANDS.....   | 65 |
|      | LIST OF REFERENCES.....  | 69 |
|      | INITIAL DISTRIBUTION LIST.....   | 71 |



## LIST OF SYMBOLS, ACRONYMS, AND/OR ABBREVIATIONS

|          |   |
|----------|---|
| $\alpha$ | Level of significance for a hypothesis test                           |
| AI       | Aviation Interest   |
| ANIT     | Aviation and Nautical Information Test                                |
| API      | Aviation Preflight Indoctrination                                     |
| AQR      | Academic Qualification Rating   |
| AQT      | Academic Qualification Test   |
| ASTB     | Aviation Selection Test Battery                                       |
| ATJ      | Aviation Training Jacket  |
| BI       | Biographical Inventory  |
| BUPERS   | Bureau of Naval Personnel   |
| CART     | Classification and Regression Trees                                   |
| CNATRA   | Chief of Naval Air Training   |
| CNET     | Chief of Naval Education and Training                                 |
| DOR      | Drop On Request   |
| ETS      | Educational Testing Services  |
| FAERO    | Predictor Variable for Final API Aeronautical Test                    |
| FENGINE  | Predictor Variable for Final API Jet-Engine Test                      |
| FFRR     | Predictor Variable for Final API Flight Rules and Regulations<br>Test |
| FMET     | Predictor Variable for Final API Meteorology Test                     |
| FNAV     | Predictor Variable for Final API Navigation Test                      |
| FAR      | Flight Aptitude Rating  |
| FOBI     | Flight Officer Biographical Inventory                                 |
| FOFAR    | Flight Officer Flight Aptitude Rating                                 |
| GLRT     | Generalized-Likelihood-Ratio Test                                     |
| MCT      | Mechanical Comprehension Test   |



|         |   |
|---------|---|
| MEPS    | Medical Examination Processing Station  |
| MVT     | Math-Verbal Test  |
| NAA     | Not Aeronautically Adaptable  |
| NASC    | Naval Aviation Schools Command  |
| NOM     | Not Officer Material  |
| NOMI    | Naval Aerospace and Operational Medical Institute                                       |
| NPQ     | Not Physically Qualified  |
| OPD     | Operational Psychology Department   |
| PAERO   | Predictor Variable for First API Aeronautical Test                                      |
| PENGINE | Predictor Variable for First API Jet-Engine Test  |
| PBI     | Pilot Biographical Inventory  |
| PFAR    | Pilot Flight Aptitude Rating  |
| PRI.A.G | Response Variable for Disqualification or Graduation from Primary Flight Training       |
| SAT     | Spatial Apperception Test   |
| SNA     | Student Naval Aviator   |
| SNFO    | Student Naval Flight Officer  |
| z       | Number of standard deviations above or below the mean of a standard normal distribution |

## EXECUTIVE SUMMARY

The U.S. Navy uses the Aviation Selection Test Battery (ASTB) to identify those Student Naval Aviator (SNA) applicants most likely to succeed in flight training.

This thesis examines two questions: Can individual answers to the Biographical Inventory (BI), a subtest of the ASTB, be used to predict SNA performance in primary flight training? Also, does repeat taking of the ASTB overpredict SNA success in primary flight training? Using classification and regression trees, this thesis analyzes flight-training data from September 1993 to March 1997, obtained from the Operational Psychology Department of the Naval Aerospace and Operational Medical Institute, Pensacola, Florida.

This thesis concludes that individual answers to the BI are not good predictors of SNA flight grades. Instead, BI scores serve as accurate indicators of flight-training disqualification.

It also concludes that those SNA who score less than a 6 on the Pilot Biographical Inventory (PBI) have a significantly higher disqualification rate in primary flight training than those who score a 6 or higher. Those SNA who repeat the taking of the ASTB also have a significantly higher disqualification rate in primary flight training than those SNA who pass the ASTB on the first taking.

Incidentally, significant differences exist in SNA performance and disqualification rates in Aviation Preflight Indoctrination (API) among different racial groups. This may be attributed to varying technical backgrounds among ethnic groups.

However, neither race nor gender is a significant factor in primary flight-training disqualification.

If annual reductions are required, then the following two options may reduce the number of SNA entering the flight-training pipeline while significantly decreasing the disqualification rate. The first option would raise the required PBI score for SNA from 4 to 6. The second option would allow candidates to take the ASTB only once.

A method is given to identify those SNA most at risk of disqualification from primary flight training. These SNA have repeated the ASTB, scored a 4 or 5 on the PBI and have an overall API score that is less than the group average.

If no annual reductions in the number of SNA entering the flight-training pipeline are required, then this recommendation may assist. This thesis found no reason to alter the current qualification criteria. The Navy allows extra flights and a longer time for training to those SNA who are having difficulty in primary flight training. It could be wise to allow those SNA in the preceding paragraph extra flights and a longer time for training at the beginning of primary flight training, before problems become apparent. This group of SNA has demonstrated borderline motivation for aviation training and weak academic skills. They are at a disproportionately high risk for disqualification.

The average taxpayer cost of an SNA disqualification from primary flight training in Fiscal Year 1996 was \$82,541. Approximately \$1,000,000 a year could be saved by this recommendation.

## ACKNOWLEDGEMENTS

The author would like to thank LCDR Philip Holmes, LT Henry Williams and Annette Baisden from the Operational Psychology Department, Naval Aerospace and Operational Medical Institute, Pensacola, Florida, for their invaluable assistance. LCDR Chris Schreier from the Aviation Warfare Training Division, Chief of Naval Education and Training, Pensacola, Florida; Captain Brian Lambert, USMC, from Aviation Preflight Indoctrination, Naval Aviation Schools Command, Pensacola, Florida; and Pat Smith from Naval Education and Training Professional Development and Technology Center, Pensacola, Florida, all went out of their way to provide data support.



# I. INTRODUCTION

Once you have tasted flight, you will always walk the Earth with your eyes turned skyward; for there you have been and there you will always be

*Leonardo da Vinci*

## A. OBJECTIVE

The U.S. Navy uses the Aviation Selection Test Battery (ASTB) to screen applicants for flight training. Applicants accepted as Student Naval Aviators (SNA) complete a six-week Aviation Preflight Indoctrination (API) before proceeding to primary flight training. Primary flight training is approximately six months in duration.

The objective of this thesis is to identify those ASTB and API component test scores that may help to predict success or failure for SNA in primary flight training. Specifically, this thesis will address two questions:

1. Can individual questions on the Biographical Inventory (BI), a subtest of the ASTB, be used to predict flight grades for SNA with the same standardized Pilot Biographical Inventory (PBI) score?
2. Do those SNA who repeat the ASTB to obtain a higher score have less success in primary flight training?

## B. METHODOLOGY

This thesis uses classification and regression trees (CART) to determine test scores that may assist in forecasting those SNA most likely to disqualify from primary flight training. A summary follows of CART methodology and terms needed to understand the plots presented in Chapter IV. The fundamental reference source for CART is Breiman et al [Ref. 1]. Purcell [Ref. 2] provides a brief tutorial suitable for the purposes of this thesis. He demonstrates the use of CART to

characterize loss rates for Army manpower models. Breiman's text is a seminal work on CART.

### 1. CART Description

Classification and regression tree techniques are non-parametric, computer-based systems used to uncover structure in a data set [Ref. 1, p. viii]. The purpose of such tree-based models is to predict the value of a response (dependent) variable based on the values of a set of predictor (explanatory) variables. Classification trees are used when the response variable is categorical in nature and regression trees are used when the response variable is continuous.

The root node of the tree contains all the data points, or cases, of the data set. The model splits the data set in two at the root node after examining all values, or attributes, of each predictor variable. The algorithm assigns each case into one of the two nodes by selecting the split that results in the highest node purity. The purity of a node is measured in terms of misclassification rates when working with classification trees and as deviance when working with regression trees. The purity is calculated from the split that maximizes the reduction in misclassification rate (or deviance). Each node that is formed from a split is based on one or more attributes of a single predictor variable.

A terminal node is a node that is split no further. The objective of the algorithm is to select a set of terminal nodes which are as pure as possible. For example, a classification tree would be used if the response variable were graduation or disqualification from primary flight training. If half of all cases in a particular terminal node disqualified from primary flight training, then the misclassification rate for that node would be 0.5. A regression tree would be used if the response variable were primary flight grade (on a 4.0 scale). If all

cases in a particular terminal node had the same flight grade, then the deviance of that node would be zero. The total misclassification rate (or deviance) of a tree is measured at the root node and is the sum of the misclassification rates (or deviance) of all terminal nodes. [Ref. 2:pp. 13-15, 23]

If no user-imposed constraints are placed on the algorithm, then the resulting tree may have the same number of terminal nodes as there are cases in the data set. In this instance, the misclassification rate (or deviance) for each terminal node would be zero. Such trees are regarded as overgrown and may have little predictive power if their number of terminal nodes makes them too complex. A method to reduce the complexity of the overgrown tree and another method to ascertain its predictive reliability are outlined below. [Ref. 2:p. 17]

It is useful at this point to define the terms *validity* and *reliability* with respect to CART. Validity refers to the issue of selecting the correct variables in order to keep bias low in the analysis and ensure no significant factors are overlooked. For example, the response variable must be dependent on at least one of the predictor variables or else the analysis will be useless. Pruning ensures predictive validity by selecting those predictor variables that are the most important in affecting the response variable.

Reliability is the measure of the stability of the selected variables. It is the ability to achieve the same results after a repetition of the analysis. Cross-validation ensures predictive reliability by repeating the results obtained from the pruned tree with the data set that was not used to grow the tree. [Ref. 4]

The *pruning* algorithm increases the predictive accuracy of the tree by decreasing the number of terminal nodes. It successively



deletes the least important splits, creating a sequence of subtrees. The importance of a subtree is determined by the cost-complexity measure,  $D_k(T')$ :

$$D_k(T') = D(T') + k * \text{size}(T'),$$

where  $D(T')$  is the deviance of subtree  $T'$ ,  $k$  is a variable cost-complexity parameter and  $\text{size}(T')$  is the number of terminal nodes of  $T'$ . Pruning identifies the  $T'$  that minimizes  $D_k(T')$ . The deviance,  $D(T')$ , is a function of the cost-complexity parameter,  $k$ , and the number of terminal nodes,  $\text{size}(T')$ . The deviance decreases as the cost-complexity parameter decreases and the number of terminal nodes increases. [Ref. 3:p. 264]

*Cross-validation* is a method used to determine the predictive reliability of a tree. The data are divided randomly into  $x$  sets of roughly equal size. Each of the  $x$  sets is held out in turn while a tree is grown and pruned. Then the set that was held out is used to measure the predictive reliability of the tree. The total misclassification rate (or deviance) of the  $x$  sets is then plotted versus tree size in terms of number of terminal nodes. The tree size with the lowest misclassification rate (or deviance) has the best predictive reliability. [Ref 1:p. 19]

## **2. Reasons for Using CART**

Classification and regression trees are recommended for large multivariate data sets and for their ability to handle both categorical and continuous predictor variables simultaneously [Ref. 2:p.14]. CART provides a method of organizing the predictor variables and the resulting values for the response variable in an easy-to-understand

format. The most important consideration is that a tree has predictive validity and reliability in order to serve as a useful tool.

### 3. Creating a Tree

A fictitious data set will be used to illustrate the creation of a classification tree. The simplicity of the data will nullify the need to use the pruning and cross-validation methods to simplify the tree and verify its predictive reliability. The set has 100 cases with data on whether the individual graduated or disqualified from a school, IQ, hair color, eye color, height and weight.

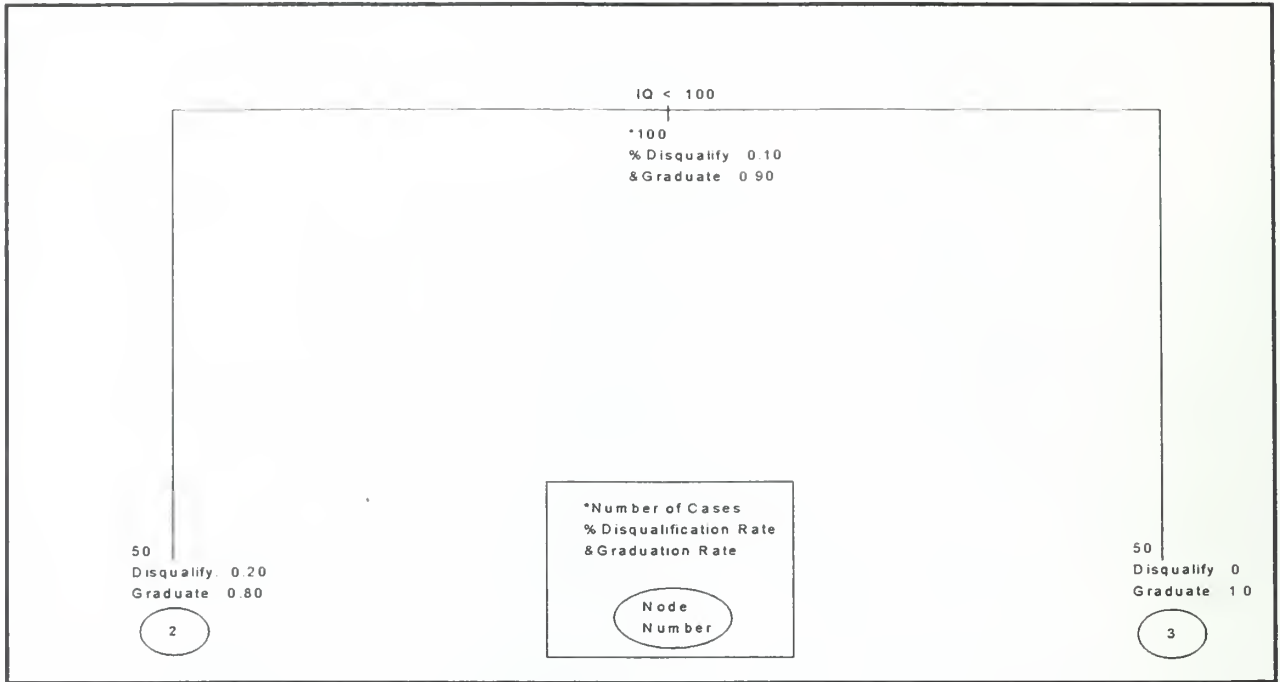
The goal of this analysis will be to determine what factors are the most important in predicting graduation from a school and the specific value of each important factor selected from the data set. The response variable will be graduation or disqualification from a school. The predictor variables will be IQ, hair color, eye color, height and weight. It is evident that the response variable and hair and eye color are categorical in nature. Height, weight, and IQ use continuous measurements. One of the advantages of CART is its ability to handle a mix of categorical and continuous explanatory variables.

*S-Plus* is a statistical-analysis software application system. The following is a command in *S-Plus* format to create a classification tree:

```
tree(GRAD.OR.DISQUALIFY ~ IQ + HAIR.COLOR + EYE.COLOR + HEIGHT +  
WEIGHT, data=Fictitious.set, na.action=na.omit)
```

The argument *na.action=na.omit* deletes all cases that have missing data.

Figure 1.1 is the resulting classification tree. The figure shows that the root node contains all the cases and that the disqualification rate is 0.10. Ten of the 100 cases in the data set disqualified from school. The root node splits on the IQ variable. Nodes 2 and 3 each contain 50 cases. The disqualification rate in Node 2 is 0.20. Node 2



**Figure 1.1** Classification Tree with Two Terminal Nodes for Fictitious Data Set. The Response Variable is Graduation or Disqualification. The Predictor Variables Are IQ, Hair Color, Eye Color, Height and Weight.

contains all ten individuals who disqualified. This is verified by Node 3 in which all 50 cases graduated.

The root node indicates that IQ was the only important predictor variable with respect to graduation or disqualification from school. The other predictor variables were of lesser importance and discarded. The left split from the root node is for all cases that had an IQ less than 100. The right split is for all cases that had an IQ greater than 100. This tree shows that all individuals who disqualified from school had an IQ of less than 100. The misclassification rate is 0.20 for Node 2 and 0.00 for Node 3. The sum produces a misclassification rate of 0.10 for the root node. The sum of the terminal nodes will be weighted for more complicated trees.

## C. ORGANIZATION

The objectives of this thesis and the methodology to achieve it have been presented in this introduction. Chapter II provides background information, including the history of naval aviation selection tests, an overview of the naval flight-training program and an explanation of disqualification. Chapter III details the data collection process used for this thesis and the type of data available for analysis. Chapter IV contains the analysis of the data. Chapter V considers costs and the effects of disqualification on them. Chapter VI presents the conclusions and recommendations. The Appendix contains the *S-Plus* commands used to create all figures in Chapter IV.



## II. BACKGROUND

### A. HISTORY OF THE NAVAL AVIATION SELECTION TEST

The demands of the Second World War produced a need for a large number of naval aviators in a short period of time. The high cost of training required that the loss of aviation candidates due to poor or unsatisfactory proficiency be minimized [Ref. 5]. The first naval aviation selection test was implemented in 1942. This test was revised in 1953 and again in 1971. It was composed of two parts: An Academic Qualification Test (AQT) and a Flight Aptitude Rating (FAR). A more recent initiative to revamp the test was begun in 1984 because of changes in the demographics of the applicant population, changes in training (e.g., the increased use of simulators) and operational aircraft (e.g., the introduction of glass cockpits), possible compromises in test security, decreased predictive validity and changes in federal law regarding employee selection procedures.

The Navy awarded the contract to develop the ASTB to Educational Testing Services (ETS) of Princeton, New Jersey [Ref. 6:p. 1]. During development of the test, 16,000 individuals were administered the experimental version. ETS identified predictive items, performed sensitivity analysis on them and conducted statistical analyses for item bias. As a result, the Naval Aerospace and Operational Medical Institute (NOMI), the controlling authority for the ASTB, is confident the ASTB has improved predictive validity over the 1971 version of the AQT/FAR.

The ASTB was introduced in 1992. It comprises the Math-Verbal Test (MVT) of general intelligence; the Mechanical Comprehension Test (MCT) of ability to perceive physical relationships and solve practical problems in mechanics; the Spatial Apperception Test (SAT) of ability to

perceive spatial relationships from differing orientations; the Aviation and Nautical Information Test (ANIT) of aviation and nautical knowledge showing an interest in naval aviation; the BI, a questionnaire of personal history and interest; and the Aviation Interest (AI), a questionnaire of aviation-related items showing early interest in aviation. Weighted combinations of these subtests are used to produce a number of scores; specifically, the Academic Qualification Rating (AQR), the Pilot Flight Aptitude Rating (PFAR), the Flight Officer Flight Aptitude Rating (FOFAR), the PBI and the Flight Officer Biographical Inventory (FOBI).

Both the AQT and FAR were used to predict disqualification from training. In the ASTB, however, only the PBI and FOBI are intended to predict disqualification. The AQR predicts academic performance and the PFAR and FOFAR predict flight performance. The ASTB was designed to be bias-free for gender and race and was separately validated for SNA and Student Naval Flight Officers (SNFO). [Ref. 6:pp. 5-6]

The Bureau of Naval Personnel (BUPERS) sets the minimum qualifying scores for naval and Coast Guard applicants as 3 for the AQR, 3 for the FOBI and 4 for the PFAR, FOFAR and PBI. The Marine Corps Order, MCO P1100.73B, sets the minimum qualifying scores for Marine applicants as 4 for the AQT and 6 for the FAR. It does not set minimum scores for the AQR, FOBI, FOFAR, PBI and PFAR. It was written before the ASTB was introduced in 1992. [Ref. 7]

## **B. OVERVIEW OF THE NAVAL FLIGHT-TRAINING PROGRAM**

The Navy, Marine Corps and Coast Guard select their respective SNA and SNFO from newly commissioned officers who have taken either the AQT/FAR or the ASTB. Prior to flight training, the officers complete the six-week API at the Naval Aviation Schools Command (NASC) in

Pensacola, Florida. In API, SNA and SNFO undergo courses in water survival, physical fitness and outdoor survival. They also take classes in meteorology, aerodynamics, engineering and other subjects in aviation. Progression to one of the primary flight-training squadrons in Milton, Florida or Corpus Christi, Texas for the SNA follows successful completion of API. [Ref. 8]



Primary flight training takes place in the T-34C Turbo Mentor, shown above, or the T-34C simulator. The flight syllabus consists of fourteen familiarization flights, ten basic instrument and fifteen radio instrument flights flown in the aircraft or simulator, five precision aerobatics flights, six formation flights and two night familiarization flights. The same instructor flies with a particular student for nine familiarization flights. An instructor is randomly assigned from the pool of available instructors to fly with the student for each of his or her remaining graded flights. For each graded flight, the instructor grades the SNA on various items including flight brief preparation, preflight knowledge, emergency procedures knowledge, ability to think and act under stress, airwork and items related to that particular flight.

The possible grades are unsatisfactory, below average, average and above average with each grade being assigned a value of 1.0, 2.0, 3.0 and 4.0, respectively. The total number of graded items in primary





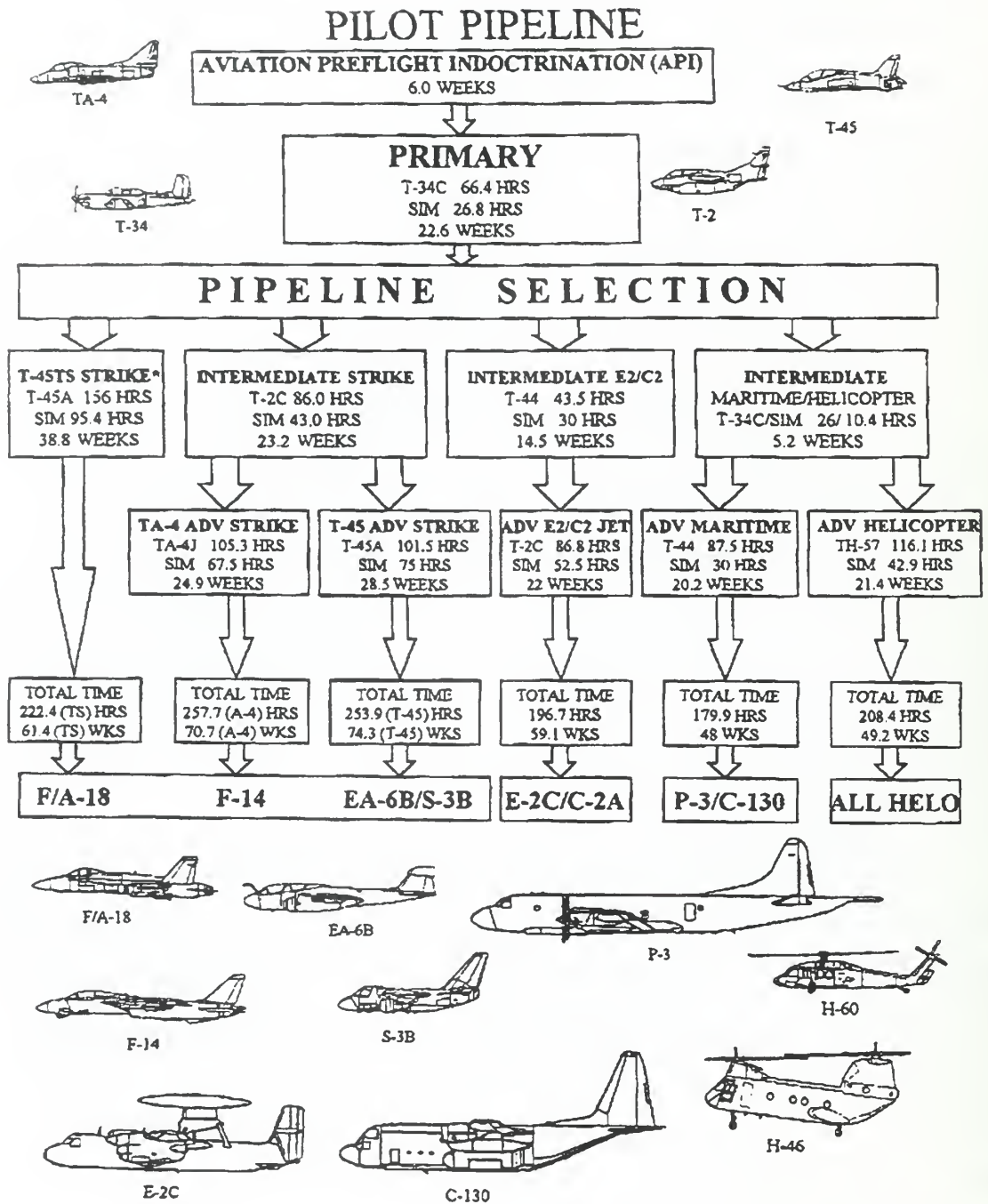
flight training is 530. The final grade for the flight portion of primary flight training is computed by summing the product of each graded item and the corresponding numerical value assigned by the instructor. This result is divided by the total number of graded items. The range of the final grade for the flight portion of primary flight training is from 1.0 for unsatisfactory to 4.0 for above average. In addition to flying, each SNA takes academic classes similar to those in API. [Ref. 9]

At the end of primary flight training, which takes approximately six months to complete, the SNA enters one of four aircraft pipelines for intermediate flight training. Selection is based on the SNA's flight grades and current needs of the Navy, Marine Corps or Coast Guard. Generally, those with the highest grades are selected for jets, followed by carrier-based propeller aircraft, maritime propeller aircraft and helicopters, respectively. Figure 2.1 shows the pilot training pipeline. [Ref. 10]

### **C. DISQUALIFICATION**

Loss of potential pilots due to a multitude of factors has always been a major concern for the naval service. Table 2.1 illustrates the enormous recruiting, selection and training effort undertaken by the Navy to produce 1,000 fleet-qualified naval aviators during a typical fiscal year. The greatest loss in potential naval aviators occurs during the recruitment and selection phases. The greatest cost, however, occurs because of the training losses.

According to Table 2.1, the Navy will disqualify 279, or approximately 18 percent, of the estimated 1,537 students upon arrival to NASC. Another 91 students, approximately seven percent of the remainder, will be disqualified from further training during API. The



\* T-45 TRAINING SYSTEM (TS): BOTH INTERMEDIATE AND ADVANCED IN T-45a

Figure 2.1 Student Naval Aviator Training Pipeline.  
From Ref. [11].

|   | Total  | Percent of<br>Contacted | Number<br>of Total<br>who fail<br>or decline | Percent<br>of Total<br>who fail<br>or decline | Time from<br>Contacted<br>(months) |
|---|--------|-------------------------|--|---|------------------------------------|
| Contacted   | 70,621 | 100.0                   | 0  | 0.0   | 0                                  |
| Take aviation<br>selection test                             | 24,364 | 34.5                    | 12,182                                       | 50.0  | 3                                  |
| Take military<br>physical exam                              | 12,182 | 17.2                    | 6,822  | 56.0  | 6                                  |
| Submit application<br>to Chief, Naval<br>Recruiting Command | 5,360  | 7.6                     | 2,755  | 51.4  | 9                                  |
| Qualify for aviation<br>officer training                    | 2,605  | 3.7                     | 1,068  | 41.0  | 12                                 |
| Enter Naval Aviation<br>Schools Command                     | 1,537  | 2.2                     | 279*<br>91#                                  | 18.2<br>5.9                                   | 15                                 |
| Enter Primary Flight<br>Training                            | 1,167  | 1.7                     | 103  | 8.8   | 18                                 |
| Enter Intermediate<br>Flight Training                       | 1,064  | 1.5                     | 12   | 1.1   | 24                                 |
| Enter Advanced Flight<br>Training                           | 1,052  | 1.5                     | 22   | 2.1   | 30                                 |
| Enter Readiness<br>Training                                 | 1,030  | 1.5                     | 30   | 2.9   | 36                                 |
| Fleet Qualified<br>Aviators                                 | 1,000  | 1.4                     | 0  | 0.0   | 42                                 |

\* Upon arrival  
# During training

**Table 2.1** U.S. Navy Projections of Recruiting, Selection and Training Requirements to Produce 1,000 Fleet Qualified Aviators for Fiscal Year 1992. The Time from Contacted column is an approximation for a Student Naval Aviator to complete Aviation Officer Candidate School and F-14 Readiness Training. Derived from Ref. [12].

Navy's forecasts for disqualification in primary, intermediate and advanced flight training for Fiscal Year 1992 were fairly accurate. The actual percentages were 9.3, 1.3 and 1.8 compared to the forecasted percentages of 8.8, 1.1 and 2.1, respectively. [Ref. 13]

The manner in which a SNA is disqualified from training is an important consideration. The categories of disqualification include Drop On Request (DOR), Flight Failure, Not Physically Qualified (NPQ), Not Officer Material (NOM), Not Aeronautically Adaptable (NAA), Academic Failure and Other (misconduct, etc.). Since flying is considered voluntary, a student may request to be dropped from training, DOR, at any time. Those SNA who display a lack of leadership potential are considered NOM. If a SNA has a difficult time adjusting to flying (e.g., he or she becomes airsick frequently) they are classified as NAA.

Despite an expected 56 percent medical disqualification rate from the Medical Examination Processing Station (MEPS) examination, another 18 percent will disqualify upon arrival to NASC. Most of these latter disqualifications are due to pre-existing medical conditions that were not identified during the MEPS examination. The rate of NPQ disqualification after API decreases throughout the training pipeline. About 1.5 percent were found to be NPQ during primary flight training for Fiscal Year 1992. This compares to 0.17 percent and 0.42 percent obtained during intermediate and advanced flight training, respectively. The jump in disqualification between intermediate and advanced flight training occurs primarily in the jet pipeline. [Ref. 13]

The Navy considers disqualification due to NPQ as unavoidable and does not include it when accounting for total disqualification due to preventable factors. The primary concern is for those SNA who leave the flight-training program due to academic or flight failure and those who

DOR. The Navy has a continuing objective to minimize the number of SNA who disqualify from further training for these reasons.



### III. DATA

#### A. DATA COLLECTION AND LOCATION

Data on SNA performance and test scores have been confined to written records until very recently. Within the past several years, some of the data have been entered into computer-based spreadsheets. Another problem is lack of data centralization. Data are recorded in individual Aviation Training Jackets (ATJ) at the various commands to which each SNA is assigned during his or her flight training. The NASC records API performance and the individual training squadrons record primary, intermediate and advanced flight training performance. Statistics are compiled at NASC and the training wings to which the training squadrons belong. The ATJ is sent to the Chief of Naval Air Training (CNATRA) in Corpus Christi and placed in storage. The Operational Psychology Department (OPD) of NOMI in Pensacola performs data analysis on the statistics compiled by the various commands mentioned above. Both CNATRA and NOMI report directly to the Chief of Naval Education and Training (CNET) in Pensacola.

#### B. DATA AVAILABLE FOR ANALYSIS

The data available at the OPD are in the file format of the *SPSS for Windows* system. The data include SNA and SNFO who graduated from API and primary flight training between September 1993 and March 1997. Three *SPSS* files were pertinent to this thesis and were converted to three *Excel 2.1* files. The first file contained API data, with records for 2,556 individuals. This file was filtered by excluding 30 individuals who were disqualified from training due to NPQ. The resulting total was 2,526 SNA and SNFO, 64 of whom disqualified due to reasons other than NPQ. SNFO were included in the file because the



requirements for acceptance into and training during API are the same for SNA and SNFO.

The second file contained primary flight-training data, with records for 756 individuals. The file was filtered to exclude three SNFO and further filtered to exclude 10 SNA who were disqualified due to NPQ or NAA. The result was primary flight training data for 743 SNA, 33 of whom were disqualified for reasons other than NPQ or NAA.

The third file contained answers for each of the 76 questions in the BI. There were 1,230 entries organized by Social Security number; month, day and year the BI was taken; sex; and race. It was filtered to include only those SNA who had data in the first two files. It was further filtered to exclude subsequent sets of BI answers for those who took the test more than once. This was done to negate any advantage from a "learning effect." The result was 795 SNA with BI answers and API or primary flight-training information or both. Nineteen individuals were excluded because of incomplete test scores. The resulting file contained 776 SNA. This included 659 SNA with API and primary flight-training data and 117 SNA with only API data. All 659 SNA with both sets of data graduated from primary flight training. Of the 117 SNA with only API data, 13 disqualified from training due to reasons other than NPQ.

## IV. ANALYSIS

### A. DETERMINING COMPOSITION OF API DATA

The goal of the analysis was to discover characteristics that may be useful for predicting SNA performance in API and primary flight training. In that regard a closer look at the API data was necessary to ascertain whether the disqualification rates among SNA and SNFO were significantly different. If different, the SNFO would be excluded so as to leave no doubt of the predictors of performance among SNA in API.

The disqualification rate difference between SNA and SNFO was examined by comparing proportions. Let the group of SNA be considered one set of trials and the group of SNFO be another set. Then for each set of trials  $i$ ,  $\{i = 1, 2\}$ , if each SNA and each SNFO is treated as an independent trial, disqualification is considered a success for this analysis. Then the probability of success for each SNA is  $p_1$  and for each SNFO is  $p_2$ . These probabilities are the *proportion parameters*. The two models both give rise to the binomial distribution. The goal is to determine any significant difference between  $p_1$  and  $p_2$ . [Ref. 3:p. 89]

According to Fleiss [Ref. 14:p. 19], the "simplest and most frequently applied statistical test of the significance of the association indicated by the data is the classic chi-square test." A hypothesis test was used to detect any difference. The null hypothesis was that the proportions of disqualifications were the same for the sets of trials. The alternative hypothesis was that the proportions of disqualifications were not the same for the sets of trials. This statistic is

$$\chi^2 = \frac{n_{++} (|n_{11}n_{22} - n_{12}n_{21}| - \frac{1}{2}n_{++})^2}{n_{1+}n_{2+}n_{+1}n_{+2}},$$

where the values for the numerator and denominator are derived from Table 4.1.

|              | Graduates | Disqualified* | Total    |
|--------------|-----------|---------------|----------|
| SNA          | $n_{11}$  | $n_{12}$      | $n_{1+}$ |
| SNFO         | $n_{21}$  | $n_{22}$      | $n_{2+}$ |
| <b>Total</b> | $n_{+1}$  | $n_{+2}$      | $n_{++}$ |

\*Disqualification other than by Not Physically Qualified

**Table 4.1** Relative Placement of Values for Chi-Square Statistic.

For the null hypothesis, the statistic has an asymptotic chi-square distribution with one degree of freedom. The  $\frac{1}{2}n_{++}$  subtracted in the numerator is Yates' correction for continuity. Fleiss [Ref. 14:p. 27] recommends that the correction always be used because it "brings probabilities associated with  $\chi^2$  and z into closer agreement with the exact probabilities than when it is not incorporated..." Applying the statistic to Table 4.2 gave

$$\chi^2 = \frac{2526(|1725 \times 32 - 32 \times 737| - \frac{1}{2}(2526))^2}{1757 \times 769 \times 2462 \times 64} = 10.9313.$$

|              | Graduates | Disqualified* | Total |
|--------------|-----------|---------------|-------|
| SNA          | 1725      | 32            | 1757  |
| SNFO         | 737       | 32            | 769   |
| <b>Total</b> | 2462      | 64            | 2526  |

\*Disqualification other than by Not Physically Qualified

**Table 4.2** Graduation and Disqualification from API for SNA and SNFO.

Referring to a  $\chi^2$  table reveals that for one degree of freedom, the  $P(\chi^2 > 10.9313) = 0.0009$ . The null hypothesis that the proportions of disqualification were the same for the sets of trials may be rejected at a 0.05 level of significance.

Performing a one-sided test can increase the power of the test. Let  $x$  and  $y$  denote the numbers of disqualifications observed in  $n$  SNA and  $m$  SNFO, respectively. Then each SNA and each SNFO is considered an independent Bernoulli trial where the outcome is either graduation or disqualification. The null hypothesis is that the proportions of disqualifications are equal. The alternative hypothesis is that the proportion of disqualifications for SNA is less than the proportion of disqualifications for SNFO. An approximate generalized-likelihood-ratio test (GLRT) is

$$\frac{\frac{x}{n} - \frac{y}{m}}{\sqrt{\frac{\left(\frac{x+y}{n+m}\right)\left(1 - \frac{x+y}{n+m}\right)(n+m)}{nm}}}$$

If this test statistic is less than or equal to  $-z_\alpha$  or greater than or equal to  $z_\alpha$ , where  $z$  is the number of standard deviations above or below the mean of a standard normal distribution and  $\alpha$  is the level of significance, then the null hypothesis may be rejected. [Ref. 15] Applying this statistic to Table 4.2 gives

$$\frac{\frac{32}{1757} - \frac{32}{769}}{\sqrt{\frac{\left(\frac{32+32}{1757+769}\right)\left(1 - \frac{32+32}{1757+769}\right)(1757+769)}{(1757)(769)}}} = -3.4438.$$

This statistic should be negative since the alternative hypothesis is that the proportion of disqualifications among SNA is less than the proportion of disqualifications among SNFO. At a 0.05 level of significance,  $-z_{0.05} = -1.6449$ . A  $z$  of  $-3.4438$  corresponds to a  $p$ -value of 0.0003. The null hypothesis may be rejected.

This confirms the existence and direction of a significant difference between the two proportions. This led to the exclusion of the SNFO from the first file. This left 1,757 SNA, 32 of whom disqualified from API due to reasons other than NPQ.

## B. ANALYSIS OF EFFECT OF TIME ON DISQUALIFICATION

Another factor that may influence possible predictors of performance is time. If disqualification is affected by a change in policy or another reason due to time of year, then this aberration may affect the analysis. To determine if the disqualification rates in API and primary flight training varied over time, the data were sorted chronologically into cohorts. Tables 4.3 and 4.4 are the API and primary flight-training graduation and disqualification data, respectively, sorted into ten cohorts of approximately equal numbers of cases. Performing chi-square proportion tests on Table 4.4 indicated no differences in disqualification existed between the time periods at a 0.05 level of significance.

For Table 4.3, asterisks indicate the time periods that are significantly different from the September-to-November 1996 period. During these three months, the head of the Aviation Training School at NASC disqualified a number of SNA from API per CNATRAINST 1500.4E [Ref. 16]. This instruction directs those responsible for training of SNA to "ensure that resources are not expended on those individuals who clearly demonstrate an inability to achieve curriculum criteria within normal time limitations" [Ref. 17]. A significant number of students had exceeded the time allotted for training during this period.

To provide homogeneity with respect to time, the September-November 1996 cohort was excluded from the first file. This resulted in 1,581 SNA, with 23 disqualified from training due to reasons other than NPQ. Excluding this cohort from the third file resulted in 763 SNA with nine disqualified from API for reasons other than NPQ. An effective analysis was then performed because the data sets were homogeneous with respect to composition over time.

| <u>Time Period</u>   | <u>Graduates</u> | <u>Disqualified</u> | <u>Total</u> |
|----------------------|------------------|---------------------|--------------|
| Jan 94-Apr 95        | 174              | 2                   | 176          |
| Apr 95-Aug 95        | 176              | 0 *                 | 176          |
| Aug 95-Oct 95        | 174              | 2                   | 176          |
| Oct 95-Feb 96        | 174              | 2                   | 176          |
| Feb 96-Apr 96        | 173              | 3                   | 176          |
| Apr 96-Jun 96        | 175              | 1 *                 | 176          |
| Jun 96-Sep 96        | 171              | 5                   | 176          |
| <b>Sep 96-Nov 96</b> | 167              | <b>9</b>            | 176          |
| Nov 96-Jan 97        | 171              | 5                   | 176          |
| Jan 97-Mar 97        | 170              | 3                   | 173          |
| <b>Total</b>         | 1725             | 32                  | 1757         |

\*Disqualification rate significantly less than Sep 96-Nov 96 rate

**Table 4.3** Student Naval Aviators Who Graduated or Disqualified from Aviation Preflight Indoctrination Between January 1994 and March 1997. The SNA Are Divided into Ten Groups and Sorted Chronologically.

| <u>Time Period</u> | <u>Graduates</u> | <u>Disqualified*</u> | <u>Total</u> |
|--------------------|------------------|----------------------|--------------|
| Sep 93-Dec 94      | 70               | 4                    | 74           |
| Dec 94-Apr 95      | 72               | 2                    | 74           |
| Apr 95-Jun 95      | 70               | 4                    | 74           |
| Jun 95-Aug 95      | 73               | 1                    | 74           |
| Aug 95-Sep 95      | 72               | 2                    | 74           |
| Sep 95-Sep 95      | 71               | 3                    | 74           |
| Sep 95-Nov 95      | 72               | 2                    | 74           |
| Nov 95-Jan 96      | 70               | 4                    | 74           |
| Jan 96-Mar 96      | 70               | 4                    | 74           |
| Mar 96-Jun 96      | 70               | 7                    | 77           |
| <b>Total</b>       | 710              | 33                   | 743          |

\*No significant differences between disqualification rates for each time period

**Table 4.4** Student Naval Aviators Who Graduated or Disqualified from Primary Flight Training Between September 1993 and June 1996. The SNA Are Divided into Ten Groups and Sorted Chronologically.

## C. PREDICTING FLIGHT GRADES AND DISQUALIFICATION RATES FROM TEST ANSWERS AND SCORES

The BI is graded by two different scoring sheets to obtain a negative PBI score and a positive PBI score. The negative score is subtracted from 47 and the positive score is added to the difference to obtain the total raw score. This raw score is converted to a numeral between 1 and 9 with 9 being the highest grade possible. A similar process is performed on the BI to obtain the FOBI grade. Analysis done by ETS resulted in the PBI and FOBI scoring methods. The PBI and FOBI are used to predict disqualification among SNA and SNFO, respectively.

The first question posed by the OPD was whether individual questions on the BI could be used to predict flight grades for SNA with the same standardized PBI score.

### 1. Analysis of BI Test Answers and Primary Flight Grades

Several factors were taken into consideration to determine the best method for analyzing OPD's first question. The third file described above had a large sample set (763 individuals), each individual had 76 answers and the answers were categorical while the primary flight-training grade was a continuous value. Item analysis was initially considered but was dropped due to lack of available software. In his thesis, Purcell [Ref. 2] demonstrates the use of CART with *S-PLUS*. The availability of software and the advantages of CART, as discussed in Chapter I, led to its use in the following analyses.

Since the Navy regards the BI primarily as a predictor of disqualification and not necessarily of flight performance, it was important to first establish whether the BI had any predictive validity with regard to flight grades. If it did, then the question of which BI



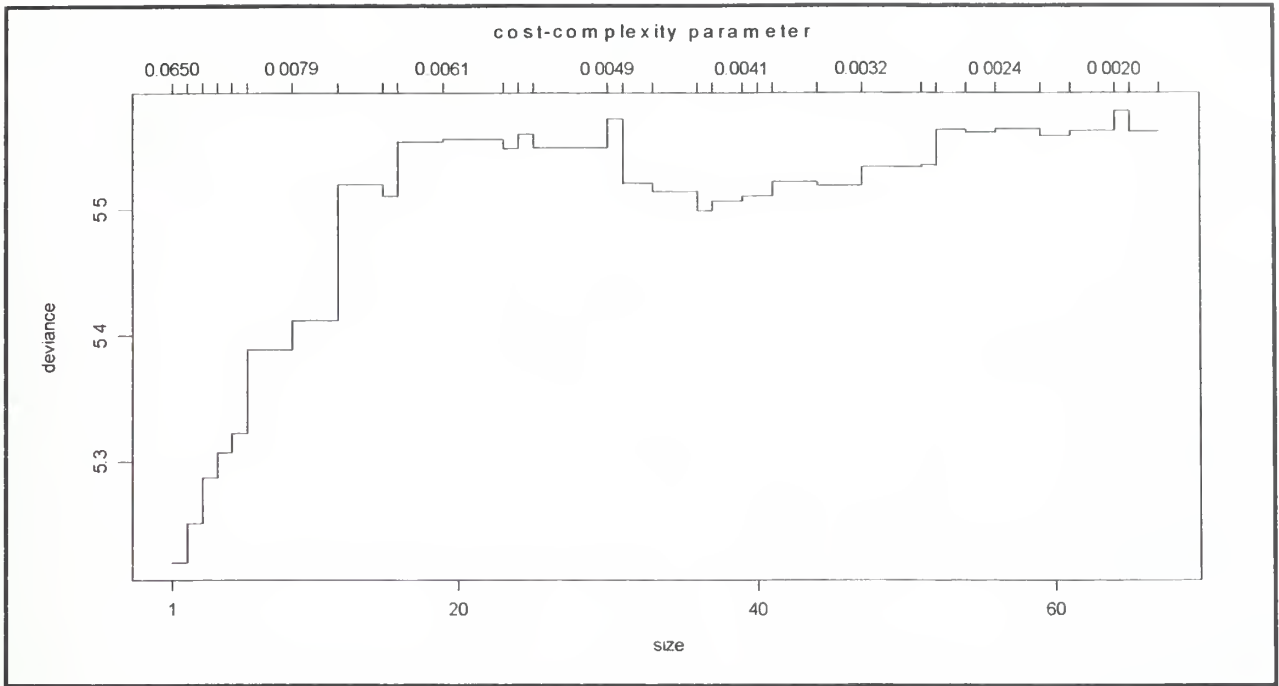
questions are good predictors of performance for those with the same PBI score may be studied.

The third file with the first set of BI answers was analyzed using a regression tree. The response variable was the primary flight grade *PRIGRADE* and the predictor variables were the answers *BI1, BI2, ..., BI76*. The *S-Plus* command, listed under Figure 4.1 in the Appendix, created the overgrown tree. Since cross validation would be important in determining the predictive reliability of the pruned tree, it is useful to discuss the cross-validation methodology in *S-Plus*.

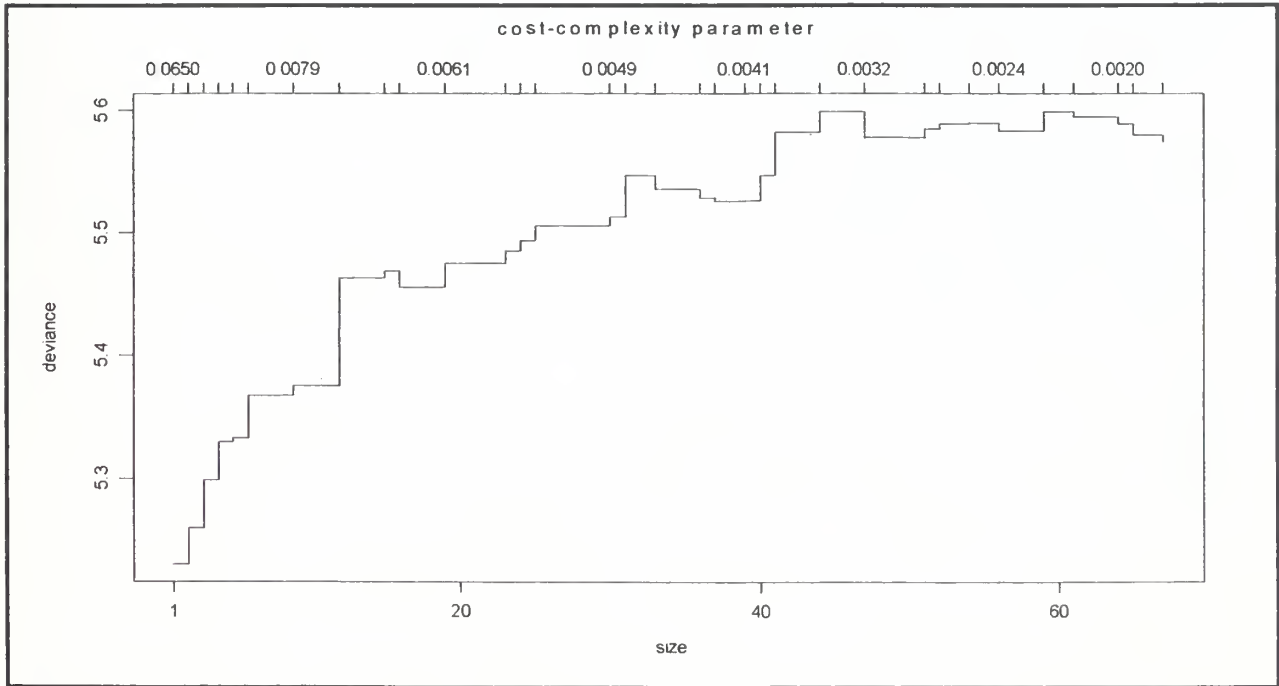
Ten-fold cross-validation is the default option for CART in *S-PLUS*. As described in Chapter I, ten-fold cross-validation randomly divides the data into ten sets, each of the ten sets is held out in turn while a tree is grown and pruned and then the set which was held out is used to measure the predictive reliability of the tree. The total misclassification rate (or deviance) of the ten sets is then plotted versus tree size in terms of number of terminal nodes. The tree size that has the lowest misclassification rate (or deviance) has the best predictive reliability.

Figures 4.1 and 4.2 are replicated cross-validation plots for the same set of commands executed consecutively in *S-PLUS*. The variation in the plots is great and demonstrates that the bin size for a ten-fold cross-validation is too small. The cross-validation method analyzed 659 out of the 763 cases, rejecting 104 cases because of missing data. It divided the data set into ten groups of approximately 66 cases each. The ten sets, each held out in turn, were not predicted well. This is due to the small numbers of cases.

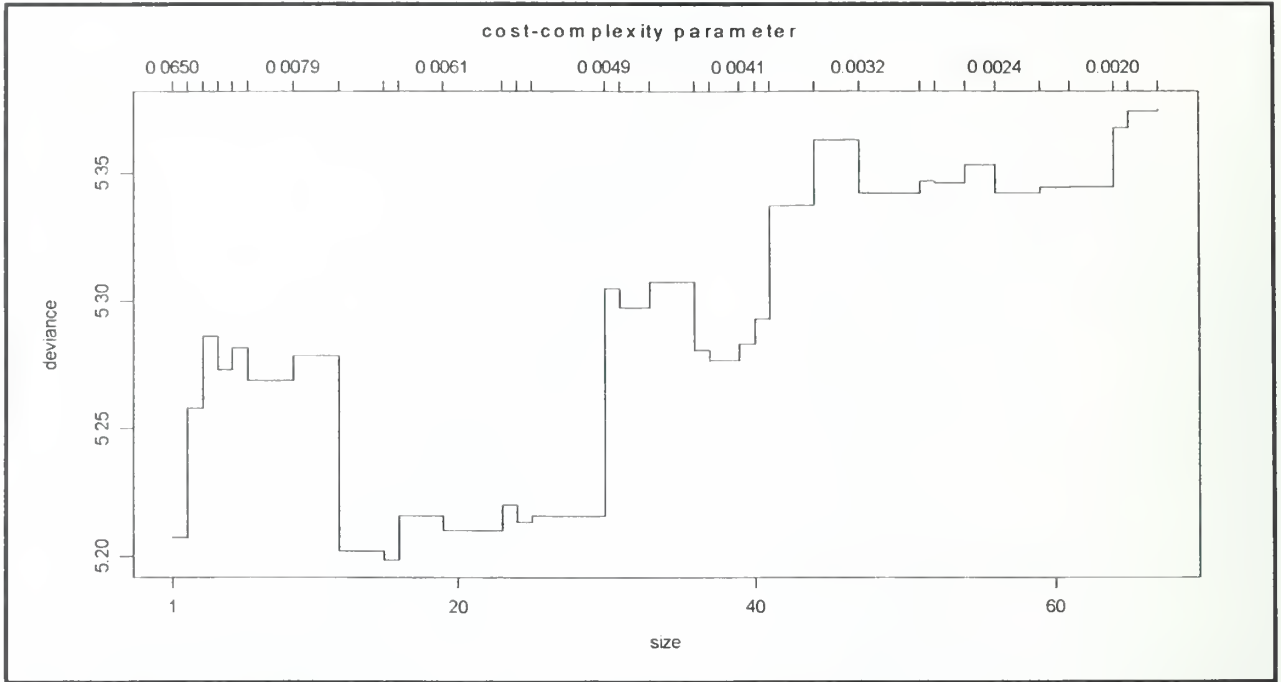
Figure 4.3 shows a five-fold cross-validation plot (twice the bin size) that proved to be more stable than the ten-fold variant.



**Figure 4.1** First Ten-Fold Cross-Validation Plot for First Set of BI Answers as Predictors of Primary Flight Grade for SNA.



**Figure 4.2** Second Ten-Fold Cross-Validation Plot for First Set of BI Answers as Predictors of Primary Flight Grade for SNA.

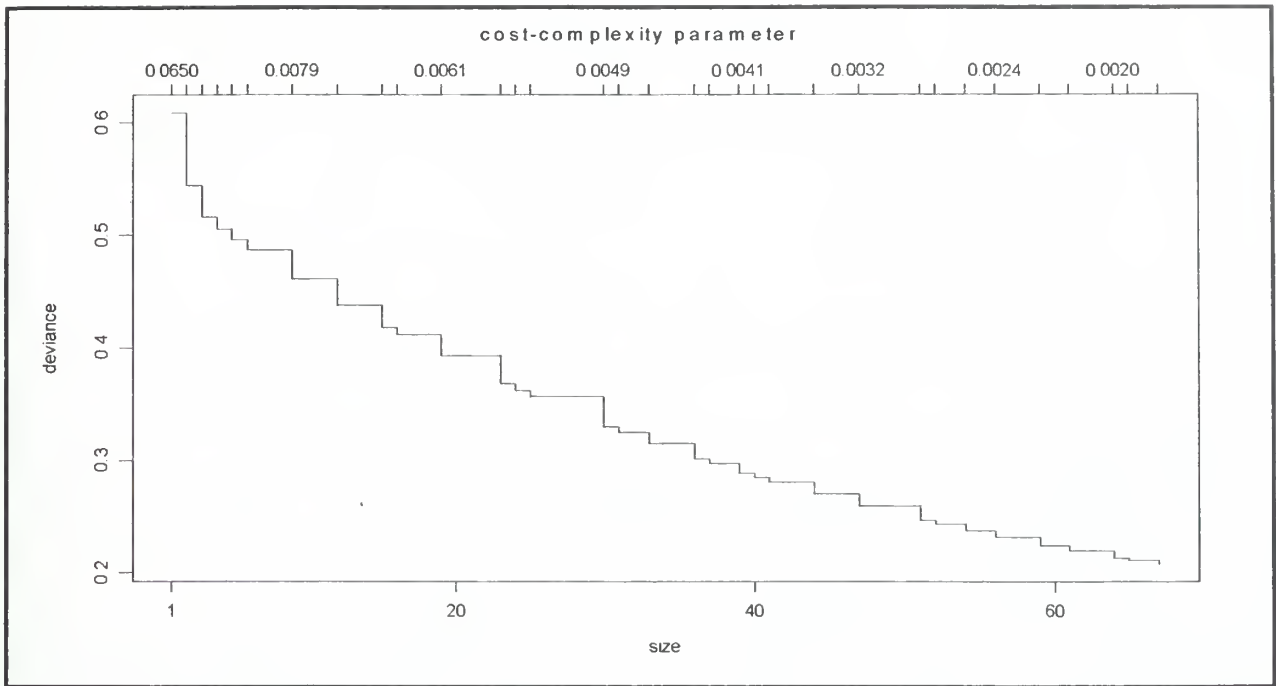


**Figure 4.3** Five-Fold Cross-Validation Plot for First Set of BI Answers as Predictors of Primary Flight Grade for SNA.

This result led to the use of five-fold cross validation for the remaining analyses because of the larger bin size.

Figure 4.3 indicates that a tree with 15 terminal nodes is the best predictor of flight grades. However, all three figures cast doubt on the validity of any size tree to predict flight grades. The range of deviance for Figures 4.1 and 4.2 is approximately 5.20 to 5.60. The range of deviance for Figure 4.3 is approximately 5.20 to 5.40. The range of the cost-complexity parameter for all three figures is approximately 0.0015 to 0.0650. This means that an overgrown tree with 67 terminal nodes has only a slightly higher cross-validated deviance than one with only 15 terminal nodes. The predictive validity of any size tree is questionable.

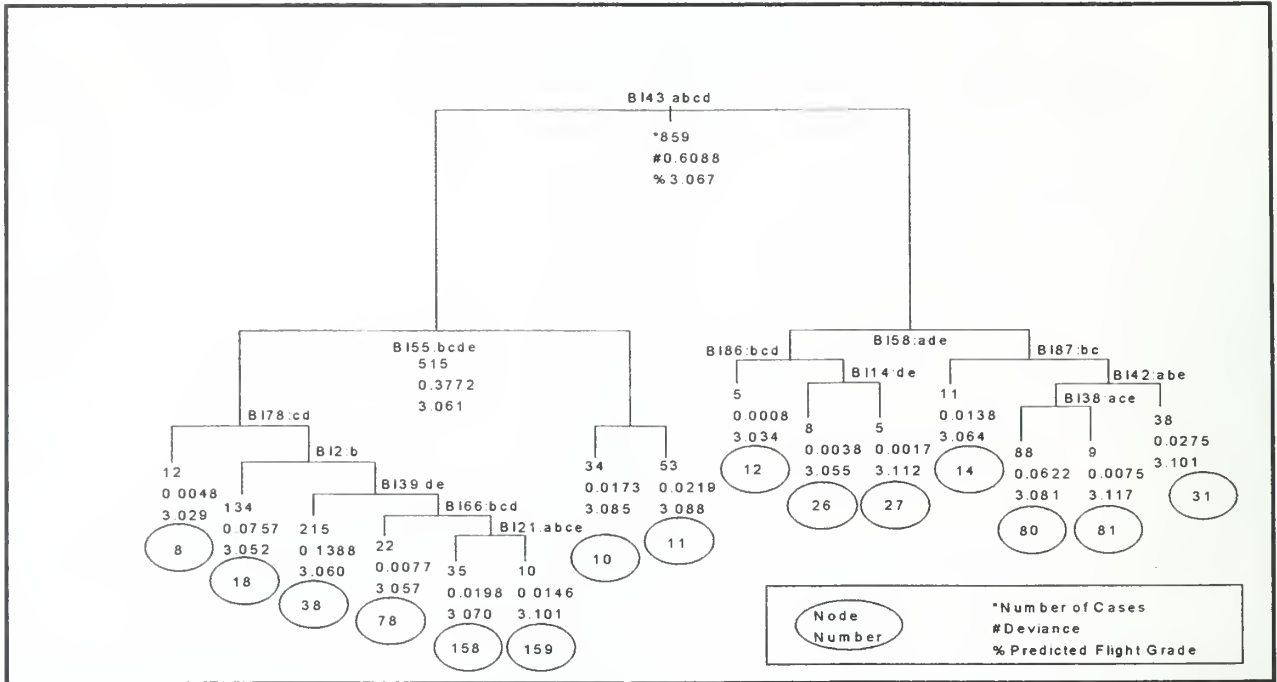
Figure 4.4 shows the pruning plot for this problem. As should happen, the total deviance decreases as the number of terminal nodes



**Figure 4.4** Pruning Plot for First Set of BI Answers as Predictors of Primary Flight Grade for SNA.

increases. However, no sharp "knee" exists in the curve that would indicate good predictive validity for a specific number of terminal nodes.

To demonstrate the complexity of the pruned tree indicated by Figure 4.3, a regression tree with 15 terminal nodes is presented in Figure 4.5. The terminal node of most interest is the one with the largest number of cases. Node 38 contains 215, or about one-third, of the 659 total cases. Follow the splits beginning at the root node. The left split at BI43 is for those who a) had never been in the air; b) had flown in large transport or passenger planes; c) had ridden in a light plane with friends occasionally; or d) had had some formal instruction in a light plane. The right split from the root node is for those who e) had soloed. The left split at BI55 is for those whose average grade in college engineering courses was a b) B- to B+; c) C- to C+;



**Figure 4.5** Regression Tree Pruned to 15 Terminal Nodes for First Set of BI Answers as Predictors of Primary Flight Grade for SNA.

d) below C-; or e) did not take engineering courses. The right split from BI55 is for those whose average grade in college engineering courses was an a) A- to A+. The right split from BI76 is for those who had learned to swim when they were a) under 6 years old; b) 6 to 9 years old; or e) had never learned to swim. An examination of the data revealed that everyone had learned to swim. The left split from BI76 is for those who had learned to swim when they were c) 10 to 14 years old or d) 15 years old or older. The right split from BI2 is for those who a) had skied on other than a beginners' slope. The left split from BI2 is for those who b) had not skied on other than a beginners' slope. The left split from BI39 is for those whose college major was one of the d) social sciences or e) none of these applies. The right split from BI39 is for those whose college major was one of the a) physical sciences;

b) biological sciences; or c) behavioral sciences. The predicted flight grade for those in node 38 is 3.060.

In summary, the predicted flight grade is 3.060 for those whose college major was not one of the physical, biological or behavioral sciences; who averaged less than an A- in or did not take any engineering courses in college; who learned to swim early in life; who was an intermediate or advanced skier; and who had not soloed in an aircraft. It is evident this node describes the traits of a plurality of the cases in that the criteria are so broad and the predicted flight grade of 3.060 is very close to the overall average of 3.067.

The next largest bin, node 18, has the same criteria except for the type of college major. Its final criterion is those who had not skied on other than beginners' slopes. Node 18 contains 134 cases, about one-fifth of the total, and the predicted flight grade is 3.052.

One could predict flight grades almost as well as this pruned tree by using the overall average as the prediction. In order to have a classification or regression tree with predictive validity, it is necessary that the range of the deviance be relatively large. Further, the number of terminal nodes corresponding to the tree with the lowest deviance on the cross-validation plot should be relatively small to reduce the complexity of the splitting criteria. Such a tree would be useful in determining the characteristics of the average SNA.

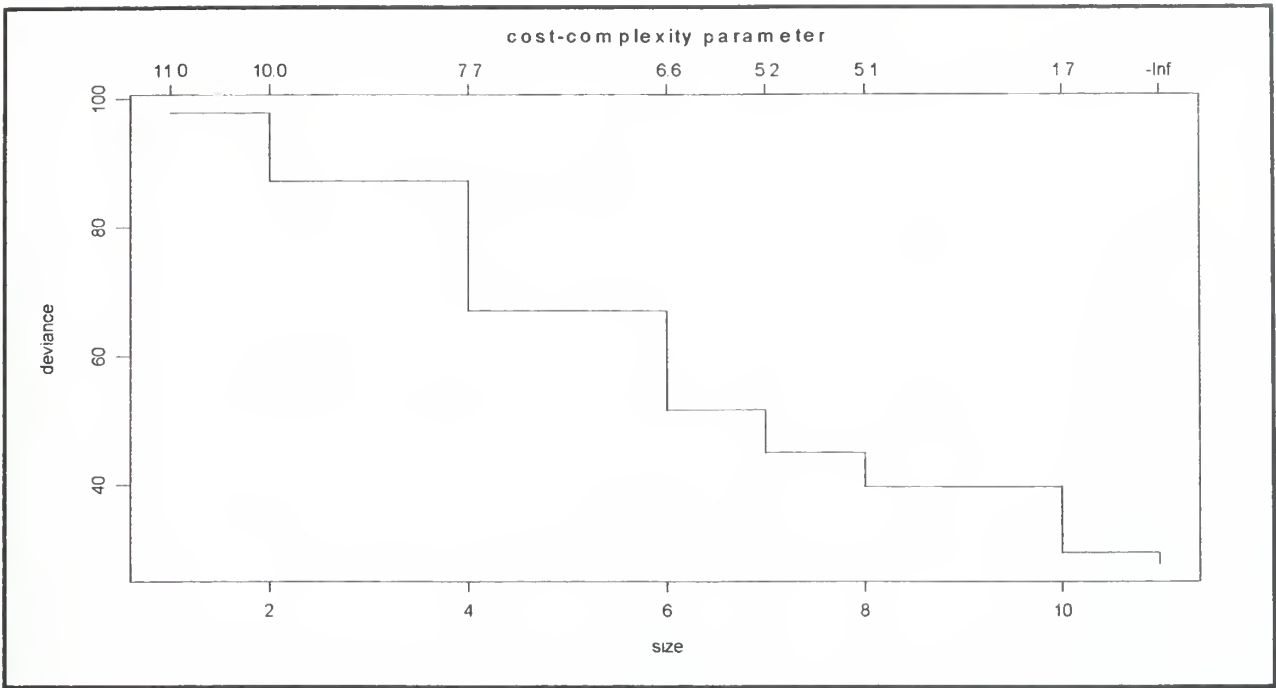
Figures 4.3, 4.4 and 4.5 show that BI questions are not very useful in predicting flight grades. The cross-validation and pruning plots' ranges of deviance are relatively small and the complexity of the pruned tree provides little, if any, predictive validity. Further analysis follows to ascertain whether the BI questions can serve as good predictors of disqualification.

## 2. Analysis of BI Test Answers and API Disqualification

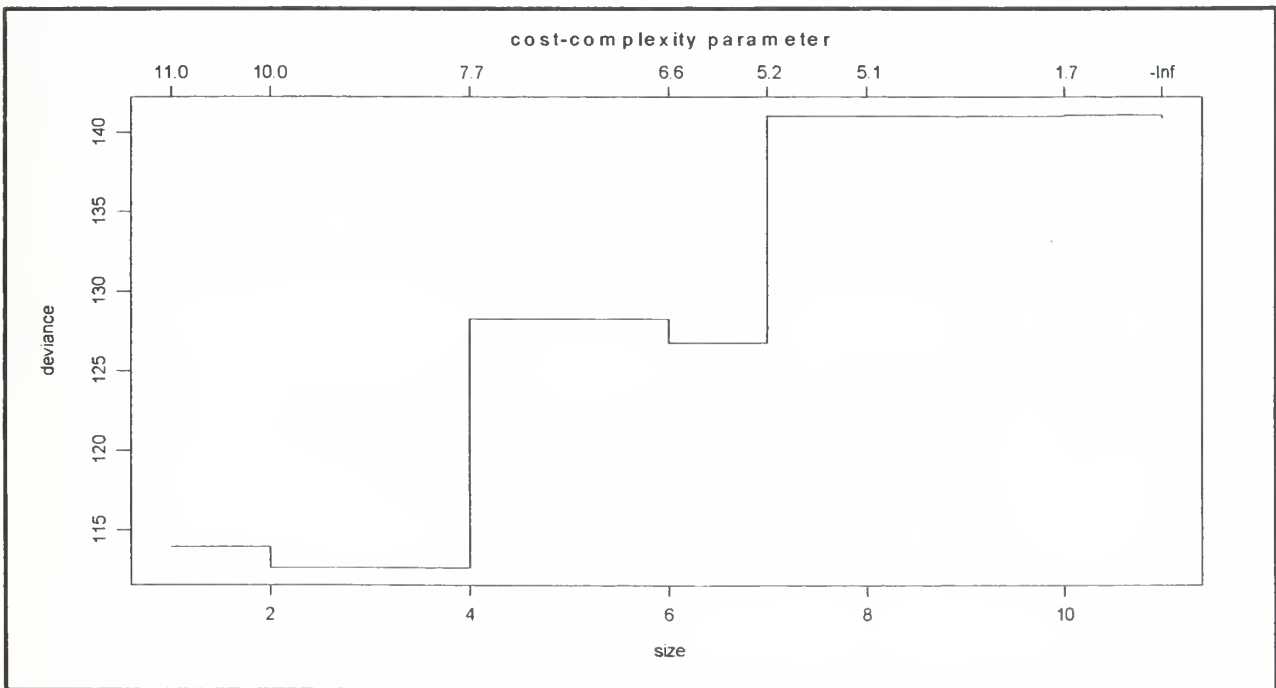
Graduation or disqualification from API was examined because all SNA for whom we have BI test answers graduated from primary flight training. The third file with the first set of BI answers was analyzed again. The response variable was *API.STAT*, whether the SNA graduated or disqualified from API, and the predictor variables were the answers *BI1*, *BI2*, ..., *BI76*.

Figures 4.6 and 4.7 are the pruning and cross-validation plots, respectively. While the pruning plot shows two large drops in deviance at four and six terminal nodes, the cross-validation plot indicates the greatest predictability is with two terminal nodes. Also, the third largest drop in deviance in the pruning plot occurs at two terminal nodes. The ranges of the deviance and the cost-complexity parameter are relatively large for both the pruning and cross-validation plots and indicate good reliability in a pruned tree of two terminal nodes. Two terminal nodes were adopted.

Figure 4.8 is the pruned classification tree. The splitting criterion is amount of flying experience. The left split is for those who have flown in large transport or passenger planes or have ridden occasionally in a light plane with friends. It contains 426 SNA with a disqualification rate of two percent. This node has all nine of those who disqualified from training in the data set. The right split contains those who have never been in the air or who have had some formal instruction in a light plane or who have soloed. Examination of the data indicated that this group of 337 included three SNA who had never been in the air. This classification tree determined that the nine SNA who disqualified had never had any formal flight training.

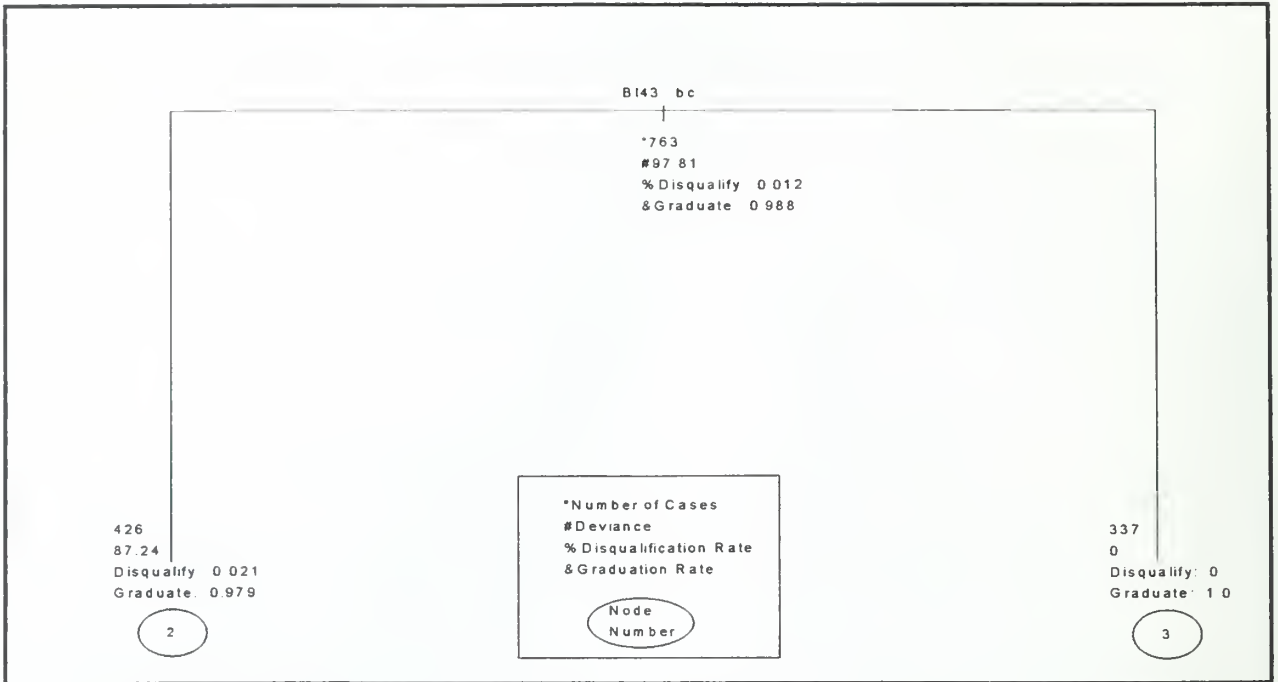


**Figure 4.6** Pruning Plot for First Set of BI Answers as Predictors of Disqualification Among SNA in API.



**Figure 4.7** Cross-Validation Plot for First Set of BI Answers as Predictors of Disqualification Among SNA in API.





**Figure 4.8** Classification Tree Pruned to Two Terminal Nodes for First Set of BI Answers as Predictors of Attrition Among SNA in API.

The GLRT statistic was computed from Table 4.5 to test the null hypothesis that the proportion of disqualifications in API for those SNA who had had at least some formal flight training is equal to the proportion of disqualifications in API for those SNA who had never had any formal flight training:

$$\frac{\frac{0}{337} - \frac{9}{426}}{\sqrt{\frac{\left(\frac{0+9}{337+426}\right)\left(1 - \frac{0+9}{337+426}\right)(337+426)}{(337)(426)}}} = -2.6842.$$

|                           | Graduates | Disqualified* | Total |
|---------------------------|-----------|---------------|-------|
| Formal Flight Training    | 337       | 0             | 337   |
| No Formal Flight Training | 417       | 9             | 426   |
| <b>Total</b>              | 754       | 9             | 763   |

\*Disqualification other than by Not Physically Qualified

**Table 4.5** Graduation and Disqualification from API for SNA With and Without Formal Flight Training.

This value of  $-2.6842$  is less than  $-z_{0.05} = -1.6449$ . A  $z$  of  $-2.6842$  corresponds to a  $p$ -value of  $0.0036$ . The null hypothesis may be rejected at a  $0.05$  level of significance in favor of the alternative that the proportion of disqualifications in API for those SNA who had had at least some formal flight training is less than the proportion of disqualifications in API for those SNA who had never had any formal flight training.

The rate of disqualification in API was significantly lower among those SNA who had had at least some formal flight training. This is not a surprising result and sheds no new light on characteristics of successful SNA.

### 3. Analysis of Primary Flight Grades

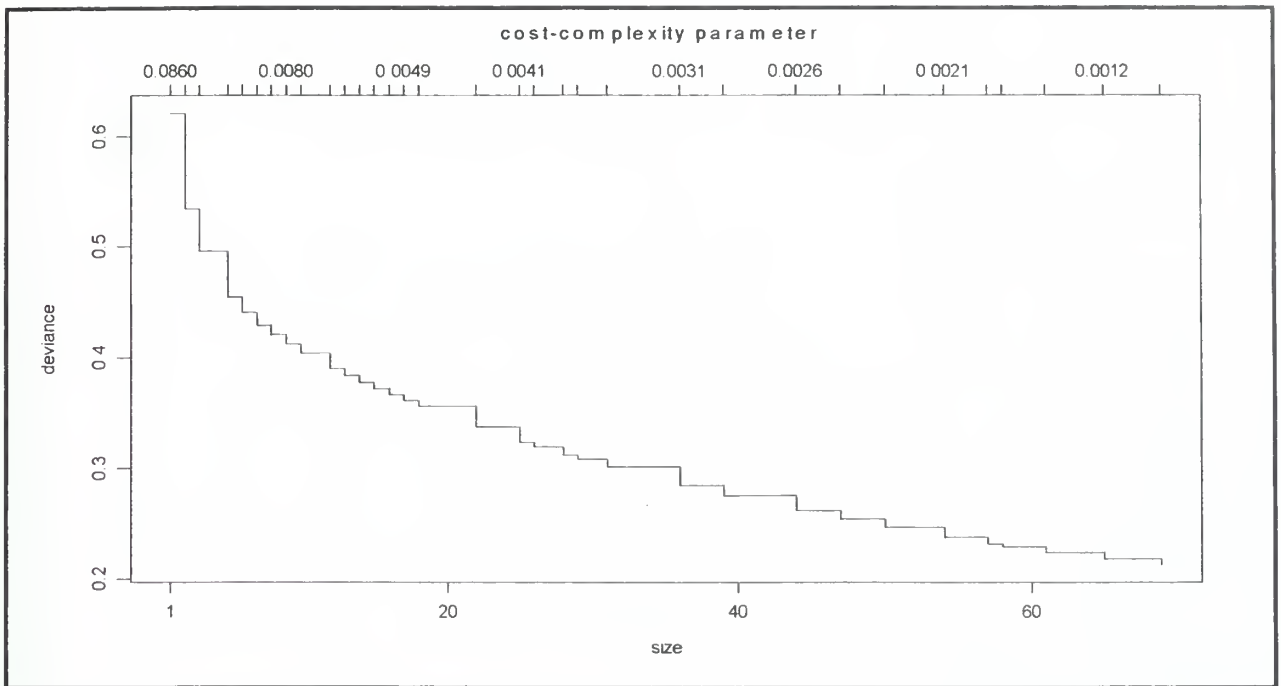
The next goal was to determine if primary flight-training grades could be predicted from test scores. The second file, with primary

flight-training data for 743 SNA, was analyzed. The response variable was *PRIGRADE*, the final primary flight grade of each SNA. The predictor variables were *TEST*, whether the SNA took the ASTB once or more than once; *PBI*, the raw PBI score; *MVT*, the raw MVT score; *MCT*, the raw MCT score; *ANIT*, the raw ANIT score; *SAT*, the raw SAT score; *PAERO* and *FAERO*, the raw scores of the first and final aerodynamics tests in API, respectively; *PENGINE* and *FENGINE*, the raw scores of the first and final jet-engine tests in API, respectively; *FNAV*, the raw score of the final navigation test in API; *FMET*, the raw score of the final meteorology test in API; *FFRR*, the raw score of the final Flight Rules and Regulations test in API; *RACE*, comprised of Asian, Black, Hispanic, Indian and White; *SEX*; *PRIACAD*, the total academic raw score from primary flight training; and *FLTHRS*, the number of flight hours each SNA had before beginning primary flight training.

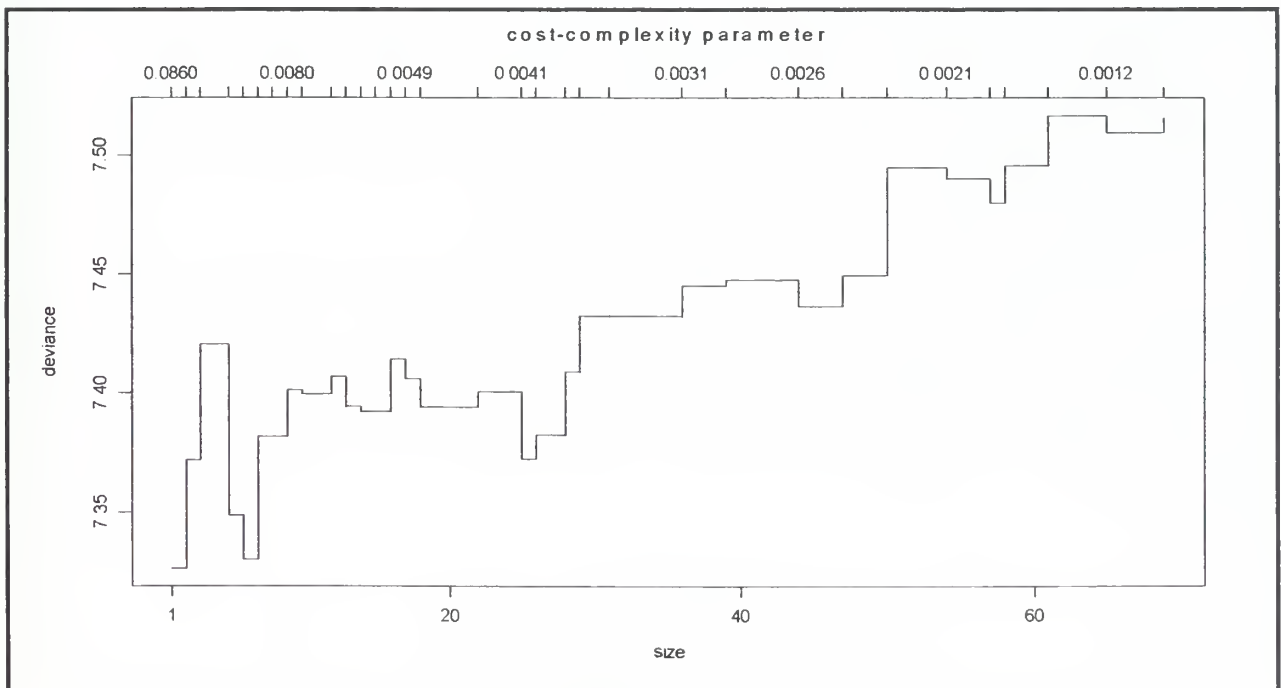
Figures 4.9 and 4.10 show the pruning and cross-validation plots, respectively. The pruning plot shows a "knee" in an initial large decrease in deviance at five terminal nodes and then a general tapering off as the number of terminal nodes increases. The cross-validation plot indicates that one terminal node has the lowest deviance; however, five terminal nodes does about as well. The ranges of the deviance and the cost-complexity parameter, on the other hand, are relatively small and confirm that no combinations of predictor variables are good predictors of primary flight grades. Therefore, no tree was produced.

#### **4. Analysis of Primary Flight-Training Disqualification**

The next goal was to determine if primary flight-training disqualification could be predicted from test scores. The response variable was *PRI.A.G*, whether the SNA graduated or disqualified from



**Figure 4.9** Pruning Plot for ASTB Scores, API Test Scores, Primary Academic Total, Previous Flight Hours, Race and Sex as Predictors of Primary Flight Grades for SNA.



**Figure 4.10** Cross-Validation Plot for ASTB Scores, API Test Scores, Primary Academic Total, Previous Flight Hours, Race and Sex as Predictors of Primary Flight Grades for SNA.

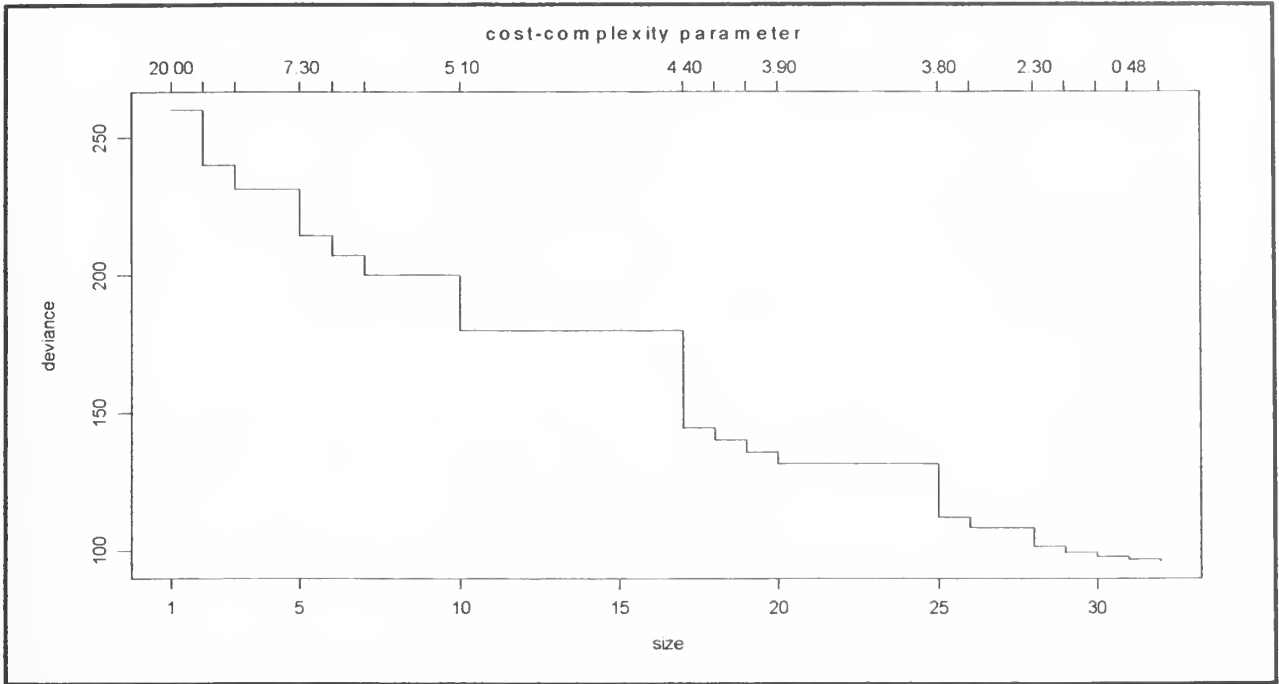
primary flight training. The predictor variables were *TEST*, *PBI*, *MVT*, *MCT*, *ANIT*, *SAT*, *PAERO*, *FAERO*, *PENGINE*, *FENGINE*, *FNAV*, *FMET*, *FFRR*, *RACE* and *SEX*. *PRIACAD* and *FLTHRS* were dropped because none of the disqualified had data in these areas.

Figures 4.11 and 4.12 are the pruning and cross-validation plots, respectively. The cross-validation plot indicates two terminal nodes have the greatest predictability. The ranges of the deviance and the cost-complexity parameter are relatively large and indicate good reliability in a pruned tree of two terminal nodes. While the pruning plot shows large deviance reductions at 10, 17 and 25 terminal nodes, two terminal nodes also have a relatively large drop. Two terminal nodes were adopted.

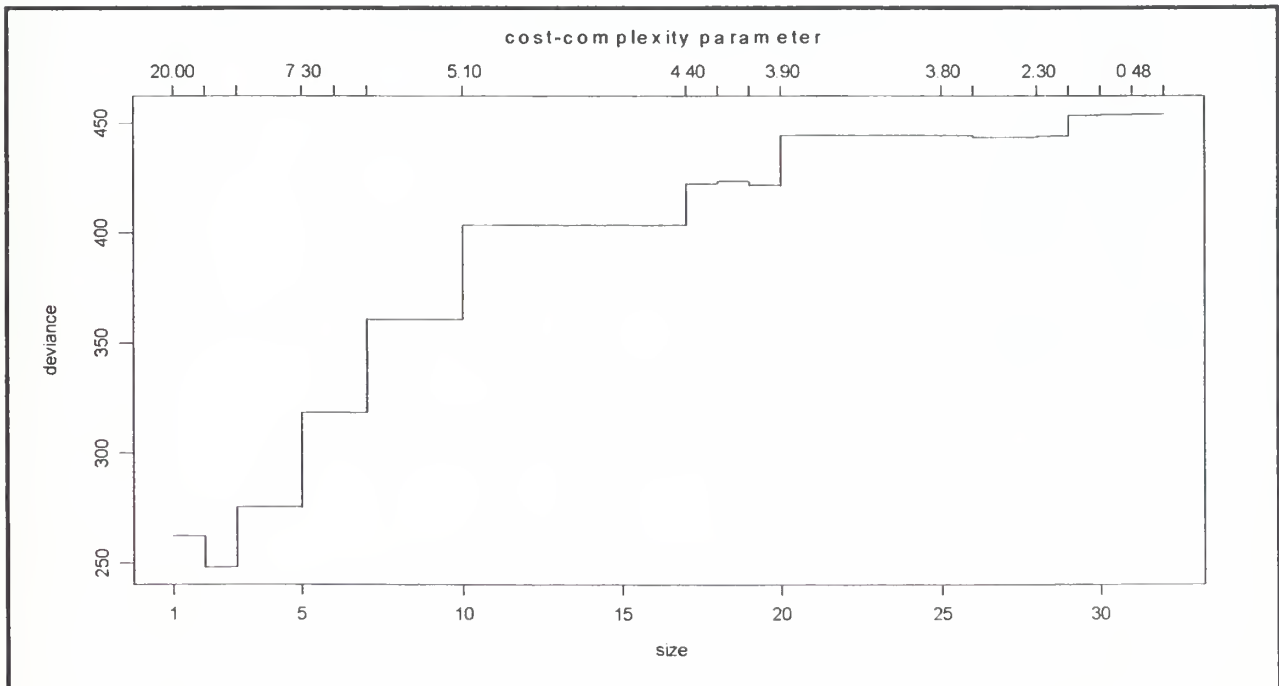
Figure 4.13 is the resulting classification tree pruned to two terminal nodes. Forty-one SNA were dropped from the analysis because of incomplete data. The root node shows that the disqualification rate among the 702 remaining SNA was 4.6 percent. Node 2 shows that the disqualification rate was 9.2 percent among the 260 SNA whose raw PBI score was less than 56.5. The disqualification rate for the 442 who scored better than 56.5 was 1.8 percent.

The GLRT statistic was computed from Table 4.6 to test the null hypothesis that the proportion of disqualifications in primary flight training for those SNA whose raw PBI score was higher than 56.5 is equal to the proportion of disqualifications for those SNA who scored lower than 56.5:

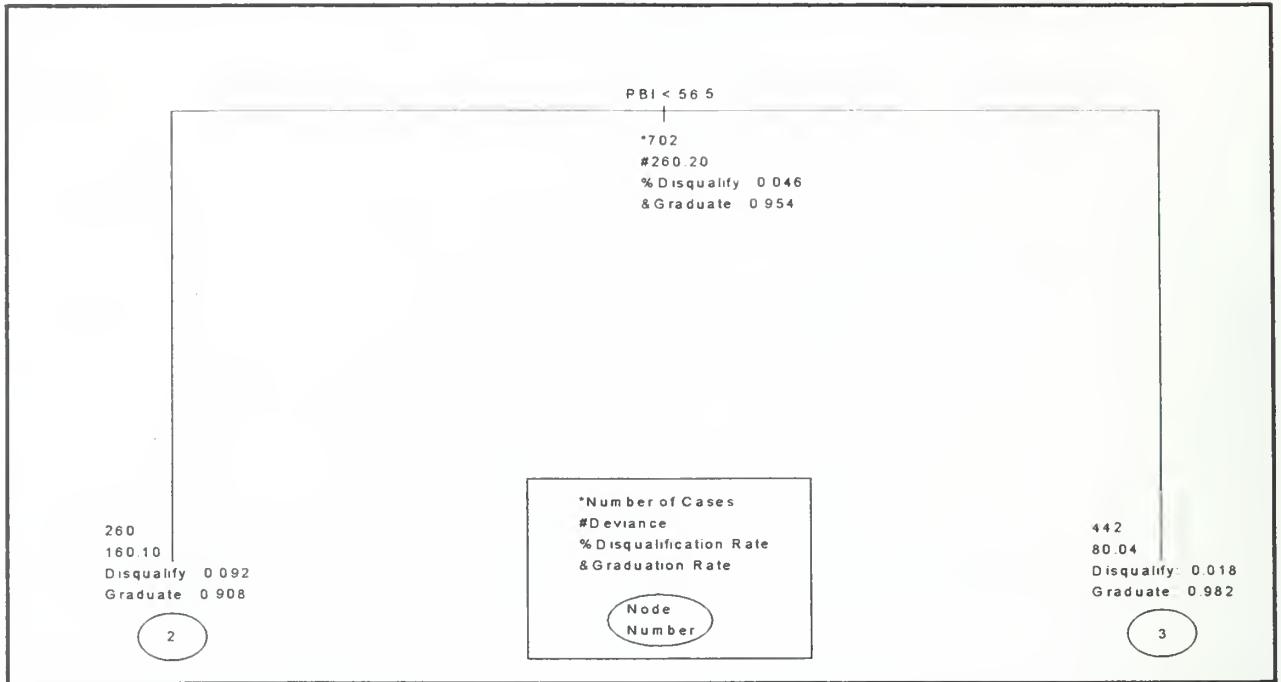
$$\frac{\frac{8}{442} - \frac{24}{260}}{\sqrt{\frac{\left(\frac{8+24}{442+260}\right)\left(1 - \frac{8+24}{442+260}\right)(442+260)}{(442)(260)}}} = -4.5520.$$



**Figure 4.11** Pruning Plot for ASTB Scores, API Test Scores, Race and Sex as Predictors of Disqualification in Among SNA Primary Flight Training.



**Figure 4.12** Cross-Validation Plot for ASTB Scores, API Test Scores, Race and Sex as Predictors of Disqualification Among SNA in Primary Flight Training.



**Figure 4.13** Classification Tree Pruned to Two Terminal Nodes for ASTB Scores, API Test Scores, Race and Sex as Predictors of Disqualification Among SNA in Primary Flight Training.

|                        | Graduates | Disqualified* | Total |
|------------------------|-----------|---------------|-------|
| PBI<br>Score ><br>56.5 | 434       | 8             | 442   |
| PBI<br>Score <<br>56.5 | 236       | 24            | 260   |
| <b>Total</b>           | 670       | 32            | 702   |

\*Disqualification other than by Not Physically Qualified

**Table 4.6** Graduation and Disqualification from Primary Flight Training for SNA By PBI Score.

This value of  $-4.5520$  is less than  $-z_{0.05} = -1.6449$ . A  $z$  of  $-4.5520$  corresponds to a  $p$ -value of  $0.0000$ . The null hypothesis may be rejected at a  $0.05$  level of significance in favor of the alternative that the proportion of disqualifications in primary flight training for those SNA whose raw PBI score was higher than  $56.5$  is less than the proportion of disqualifications for those SNA whose score was lower than  $56.5$ .

The rate of disqualification in primary flight training was significantly lower among SNA who scored higher than  $56.5$  on the PBI.

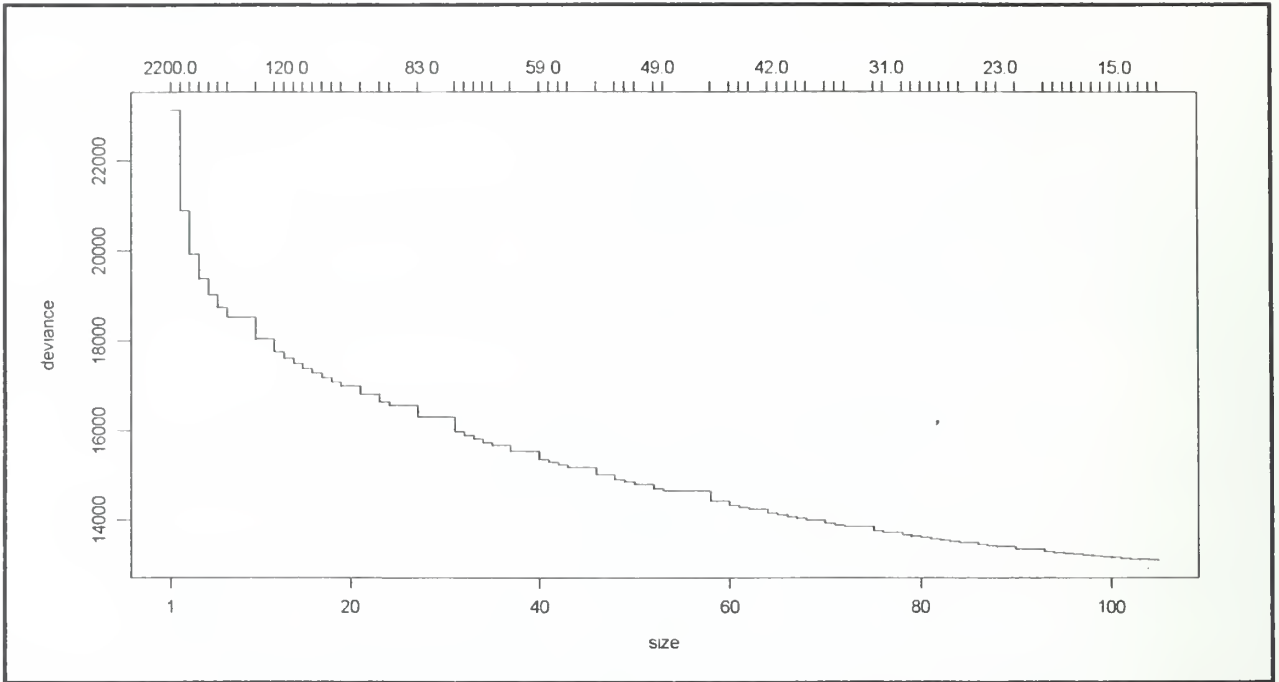
### **5. Analysis of API Final Grade**

After analyzing flight grades and disqualification in primary flight training, the next area to examine was final grades and disqualification in API. The overall API grade each SNA is assigned is comprised of two aeronautical exams, two jet-engine exams, a navigation exam, a meteorology exam and a flight rules and regulations exam. A passing grade of  $80$  percent on each exam is required.

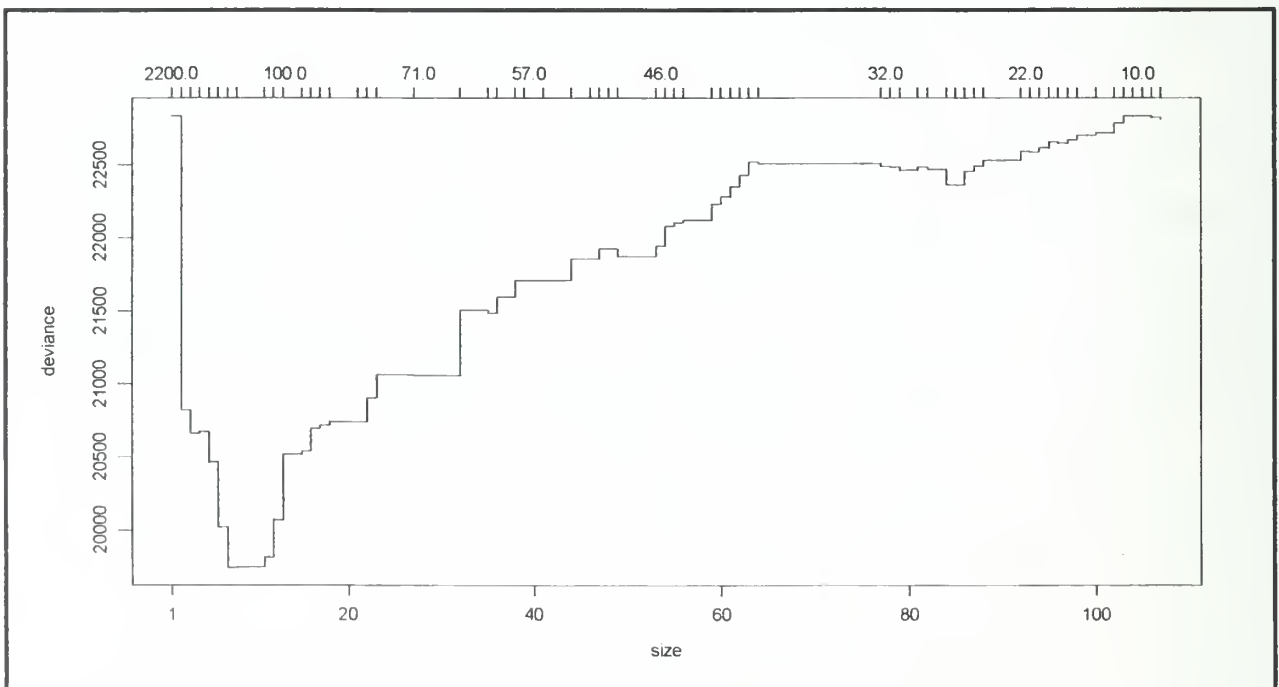
For this analysis, the first file with the  $1,581$  SNA was used. The response variable was *OVERALL*, the overall API grade. The predictor variables were *TEST*, whether the SNA took the ASTB once or more than once; *PBI*, the raw PBI score; *MVT*, the raw MVT score; *MCT*, the raw MCT score; *ANIT*, the raw ANIT score; *SAT*, the raw SAT score; *RACE*, composed of Asian, Black, Hispanic, Indian and White; and *SEX*.

Figures 4.14 and 4.15 are the pruning and cross-validation plots, respectively. The pruning plot shows a large drop in deviance at two terminal nodes followed by a steady reduction in deviance. After  $60$  terminal nodes the rate of reduction becomes quite small. The cross-validation plot shows seven to be the number of terminal nodes with the greatest predictability. The ranges of the deviance and cost-complexity





**Figure 4.14** Pruning Plot for ASTB Scores, Race and Sex as Predictors of Overall Grade Among SNA in API.



**Figure 4.15** Cross-Validation Plot for ASTB Scores, Race and Sex as Predictors of Overall Grade Among SNA in API.

parameter are large with a significant difference in deviance between seven nodes and the size of the unpruned tree. Seven nodes were adopted.

Figure 4.16 is the regression tree pruned to seven terminal nodes. The single most important criterion for splitting the data is the MVT. The MCT, race and the ANIT are the other significant criteria. Retaking the test, the PBI, the SAT and sex were not important as predictors of API grades. The root node contains 1,552 out of the 1,581 cases in the file. *S-PLUS* deleted 29 cases because of missing data.

Node 15 contains 501 cases or 32 percent of the total. It contains those Asian and Caucasian SNA who scored more than 26.5 on the MVT and more than 18.5 on the ANIT for a predicted grade of 94.85.

Node 10 has 479 cases, or 31 percent of the total, and contains those American Indian and Caucasian SNA who scored less than 26.5 on the

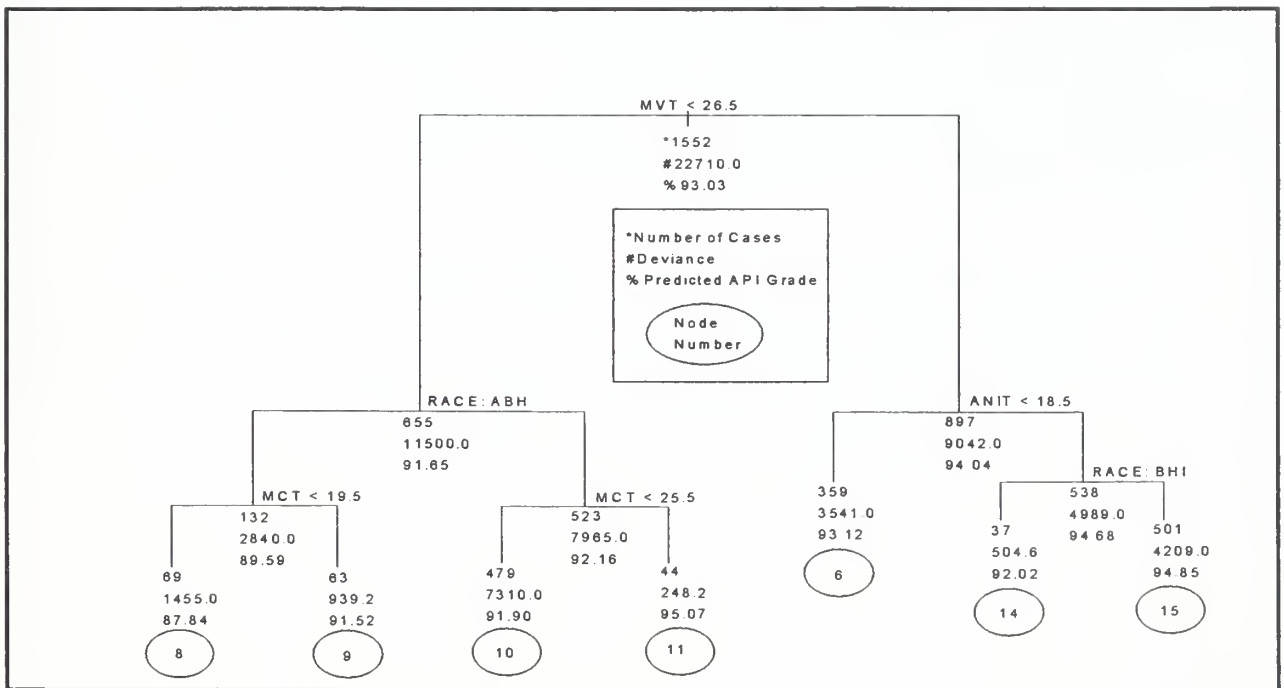


Figure 4.16 Regression Tree Pruned to Seven Terminal Nodes for ASTB Scores, Race and Sex as Predictors of Overall Grade Among SNA in API.

MVT and less than 25.5 on the MCT for a predicted grade of 91.90.

Node 6 has 359 cases, or 23 percent of the total, and contains those SNA who scored greater than 26.5 on the MVT and less than 18.5 on the ANIT for a predicted score of 93.12.

In summary, the pruned tree shows that the better the SNA did on the knowledge and problem-solving portions of the ASTB (the MVT, the MCT and the ANIT), the better he or she performed in API.

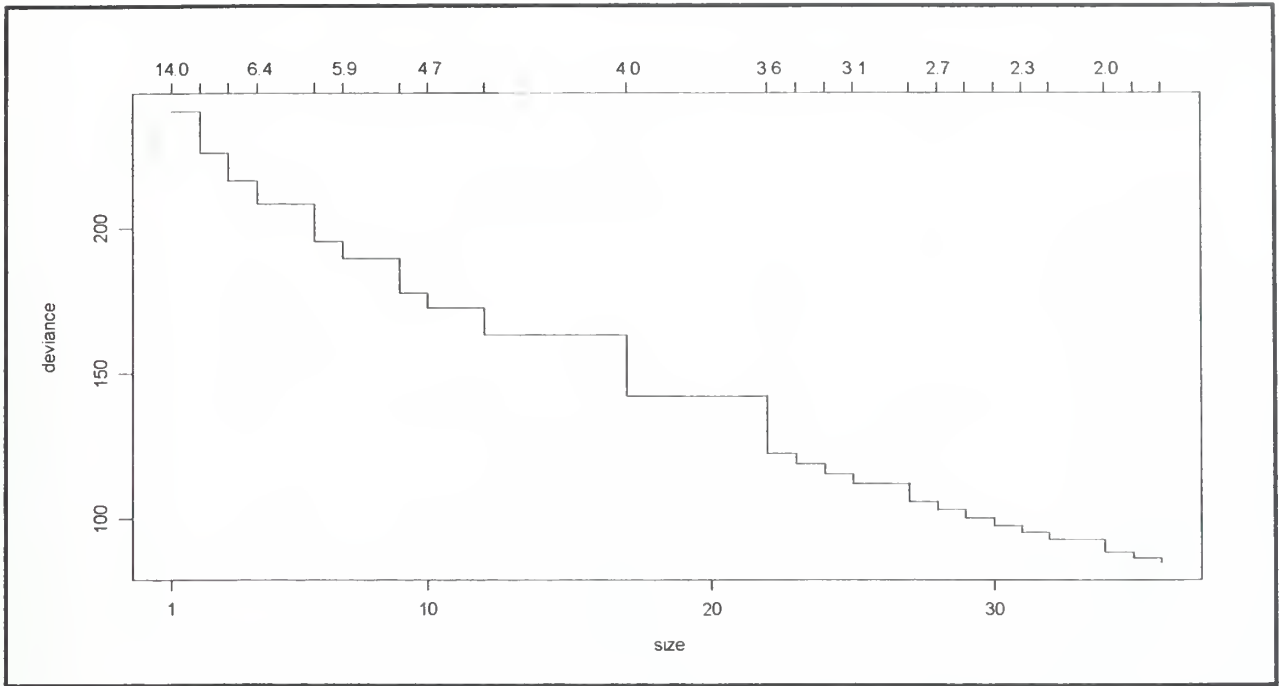
## 6. Analysis of API Disqualification

The last aspect to study about API was the possible predictors of disqualification. The response variable was *ATTRITE*, whether the SNA graduated or disqualified from API. The predictor variables were the same ones used previously: *TEST*, *PBI*, *FOBI*, *MVT*, *MCT*, *ANIT*, *SAT*, *RACE* and *SEX*.

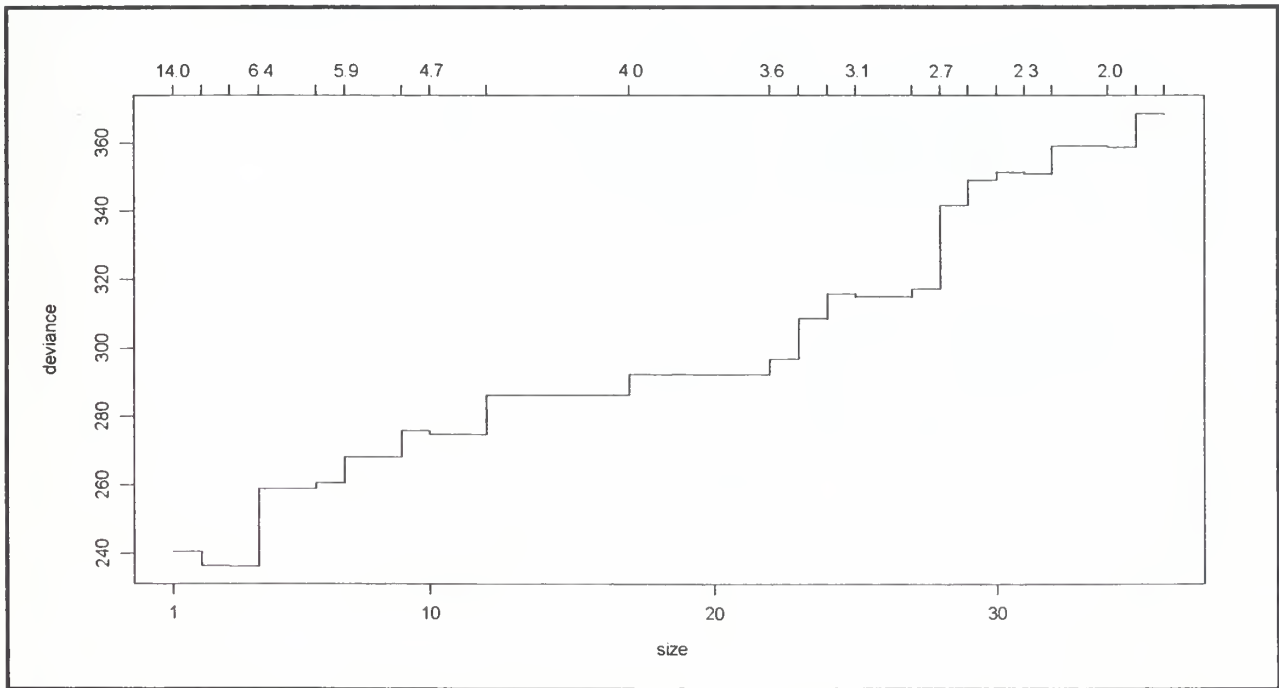
Figures 4.17 and 4.18 are the pruning and cross-validation plots, respectively. The pruning plot shows large decreases in deviance at 17 and 22 terminal nodes with a steady decrease elsewhere. The ranges of deviance and the cost-complexity parameter in the cross-validation plot are relatively large. Two terminal nodes were chosen based on Figure 4.18.

Figure 4.19 is the classification tree pruned to two terminal nodes. The most important splitting criterion is race. The disqualification rate for African-Americans, Asians and Hispanics is 4.7 percent while the disqualification rate for American Indians and Caucasians is about one percent.

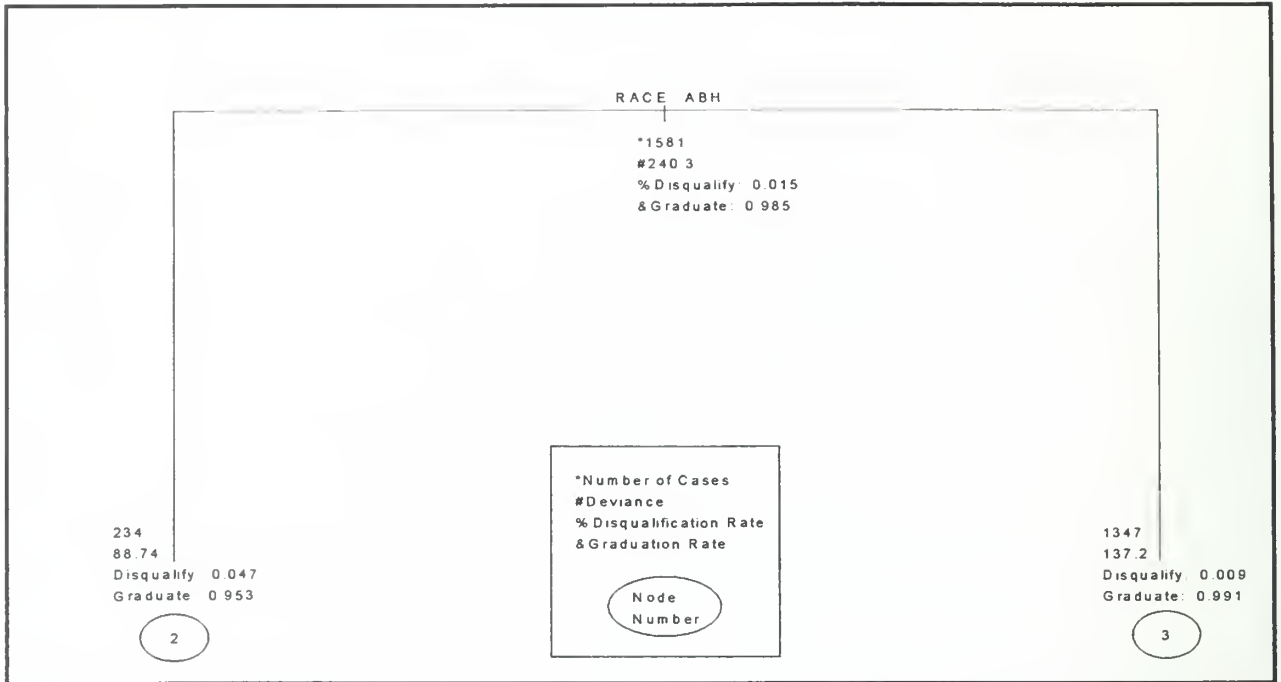
The GLRT statistic was computed from Table 4.7 to test the null hypothesis that the proportion of disqualifications in API for American Indian and Caucasian SNA is equal to the proportion of disqualifications for African-American, Asian and Hispanic SNA:



**Figure 4.17** Pruning Plot for ASTB Scores, Race and Sex as Predictors of Disqualification Among SNA in API.



**Figure 4.18** Cross-Validation Plot for ASTB Scores, Race and Sex as Predictors of Disqualification Among SNA in API.



**Figure 4.19** Classification Tree Pruned to Two Terminal Nodes for ASTB Scores, Race and Sex as Predictors of Attrition Among SNA in API.

$$\frac{\frac{12}{1347} - \frac{11}{234}}{\sqrt{\frac{\left(\frac{12+11}{1347+234}\right)\left(1 - \frac{12+11}{1347+234}\right)(1347+234)}{(1347)(234)}}} = -4.4930.$$

This value of -4.4930 is less than  $-z_{0.05} = -1.6449$ . A z of -4.4930 corresponds to a p-value of 0.0000. The null hypothesis may be rejected at a 0.05 level of significance in favor of the alternative that the proportion of disqualifications in API for American Indian and Caucasian SNA is less than the proportion of disqualifications for African-American, Asian and Hispanic SNA.

The rate of disqualification in API was significantly lower among American Indian and Caucasian SNA.

|                                     | Graduates | Disqualified* | Total |
|-------------------------------------|-----------|---------------|-------|
| Indian,<br>White<br>SNA             | 1335      | 12            | 1347  |
| Asian,<br>Black,<br>Hispanic<br>SNA | 223       | 11            | 234   |
| <b>Total</b>                        | 1558      | 23            | 1581  |

\*Disqualification other than by Not Physically Qualified

**Table 4.7** Graduation and Disqualification from Primary Flight Training for SNA By Race.

#### D. REPEAT TESTING

Current Navy regulations allow potential SNA to repeat the ASTB as many times as they desire. The retaking of the ASTB is subject only to a 180-day waiting period between tests with the most recent test scores replacing the previous ones [Ref. 18].

The second question posed by the OPD was whether repeating the ASTB to obtain a higher score overpredicts success in the flight-training program, no doubt due to a "learning effect".

##### 1. Analysis of Disqualification Rates

The two files with API and primary flight-training data contain information on whether the individual took the test once or more than once and whether he or she graduated or disqualified from API or primary flight training. Tables 4.8 and 4.9 show disqualification rates for API

|                   | <u>Graduates</u> | <u>Disqualified*</u> | <u>Total</u> | <u>Percent Disqualified</u> |
|-------------------|------------------|----------------------|--------------|-----------------------------|
| Once              | 1020             | 11                   | 1031         | 1.07                        |
| More than<br>once | <u>538</u>       | <u>12</u>            | <u>550</u>   | <u>2.18</u>                 |
| <b>Total</b>      | 1558             | 23                   | 1581         | 1.45                        |

\*Disqualification other than by Not Physically Qualified

**Table 4.8** Number of Times Student Naval Aviators Took the Aviation Selection Test Battery versus Graduation or Disqualification from Aviation Preflight Indoctrination for January 1994 to March 1997.

|                   | <u>Graduates</u> | <u>Disqualified*</u> | <u>Total</u> | <u>Percent Disqualified</u> |
|-------------------|------------------|----------------------|--------------|-----------------------------|
| Once              | 426              | 11                   | 437          | 2.52                        |
| More than<br>once | <u>284</u>       | <u>22</u>            | <u>306</u>   | <u>7.19</u>                 |
| <b>Total</b>      | 710              | 33                   | 743          | 4.44                        |

\*Disqualification other than by Not Physically Qualified or Not Aeronautically Adaptable

**Table 4.9** Number of Times Student Naval Aviators Took the Aviation Selection Test Battery versus Graduation or Disqualification from Primary Flight Training for September 1993 to June 1996.

and primary flight training, respectively. Those SNA dismissed from training due to physical reasons were excluded. Those SNA in Tables 4.8 and 4.9 who were dismissed from training measure the disqualification rate due to lack of desire or academic or flying proficiency.

The GLRT statistic was computed from Table 4.8 to test the null hypothesis that the proportion of disqualifications in API for those SNA who took the ASTB once is equal to the proportion of disqualifications for those SNA who repeated it:

$$\frac{\frac{11}{1031} - \frac{12}{550}}{\sqrt{\frac{\left(\frac{11+12}{1031+550}\right)\left(1 - \frac{11+12}{1031+550}\right)(1031+550)}{(1031)(550)}}} = -1.7634.$$

This value of -1.7634 is less than  $-z_{0.05} = -1.6449$ . A z of -1.7634 corresponds to a p-value of 0.0389. The null hypothesis may be rejected at a 0.05 level of significance in favor of the alternative that the proportion of disqualifications in API for those SNA who took the ASTB once is less than the proportion of disqualifications for those SNA who repeated it. Incidentally, the null hypothesis would not be rejected if the alternative were two-sided. The significance is not strong.

A similar test was performed on the data in Table 4.9. The GLRT statistic was computed to test the null hypothesis that the proportion of disqualifications in primary flight training for those SNA who took the ASTB once is equal to the proportion of disqualifications for those SNA who repeated it:



$$\frac{\frac{11}{437} - \frac{22}{306}}{\sqrt{\frac{\left(\frac{11+22}{437+306}\right)\left(1 - \frac{11+22}{437+306}\right)(437+306)}{(437)(306)}}} = -3.0426 .$$

This value of -3.0426 is less than  $-z_{0.05} = -1.6449$ . A  $z$  of -3.0426 corresponds to a  $p$ -value of 0.0012. The null hypothesis may be rejected at a 0.05 level of significance in favor of the alternative that the proportion of disqualifications in primary flight training for those SNA who took the ASTB once is less than the proportion of disqualifications among SNA who took it more than once.

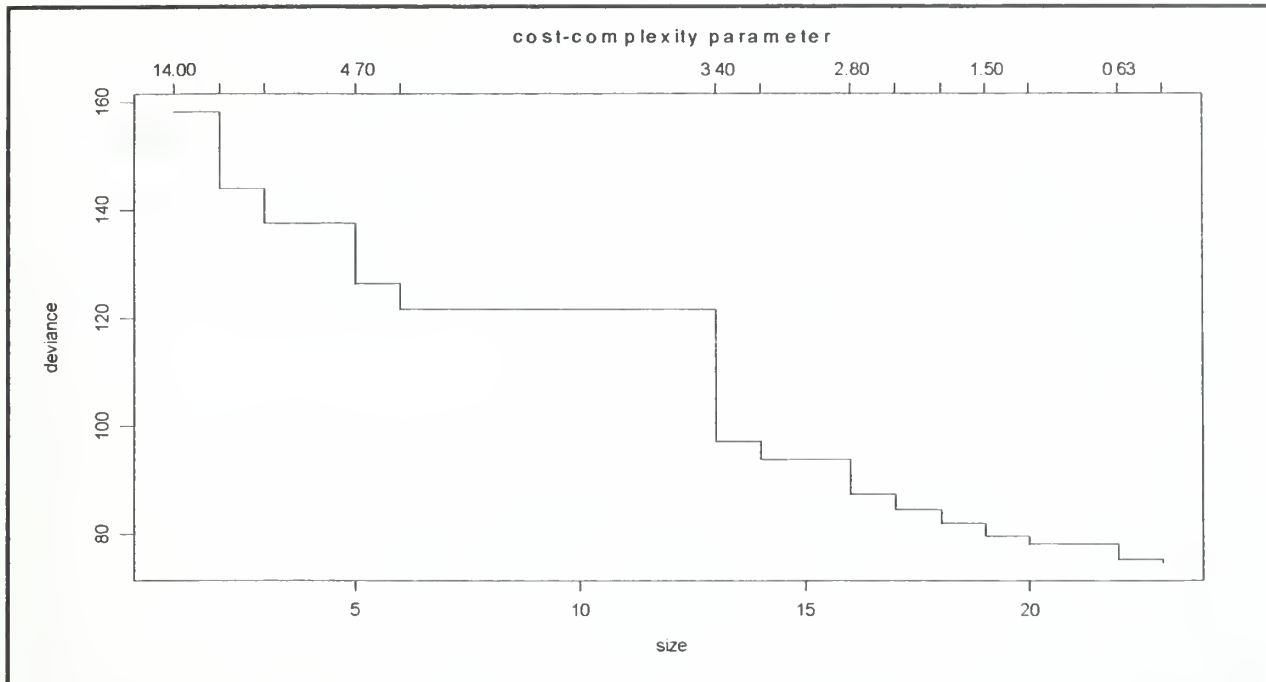
The disqualification rate is significantly lower only during primary flight training for those SNA who took the ASTB once.

## 2. Analysis of Test Retakers in Primary Flight Training

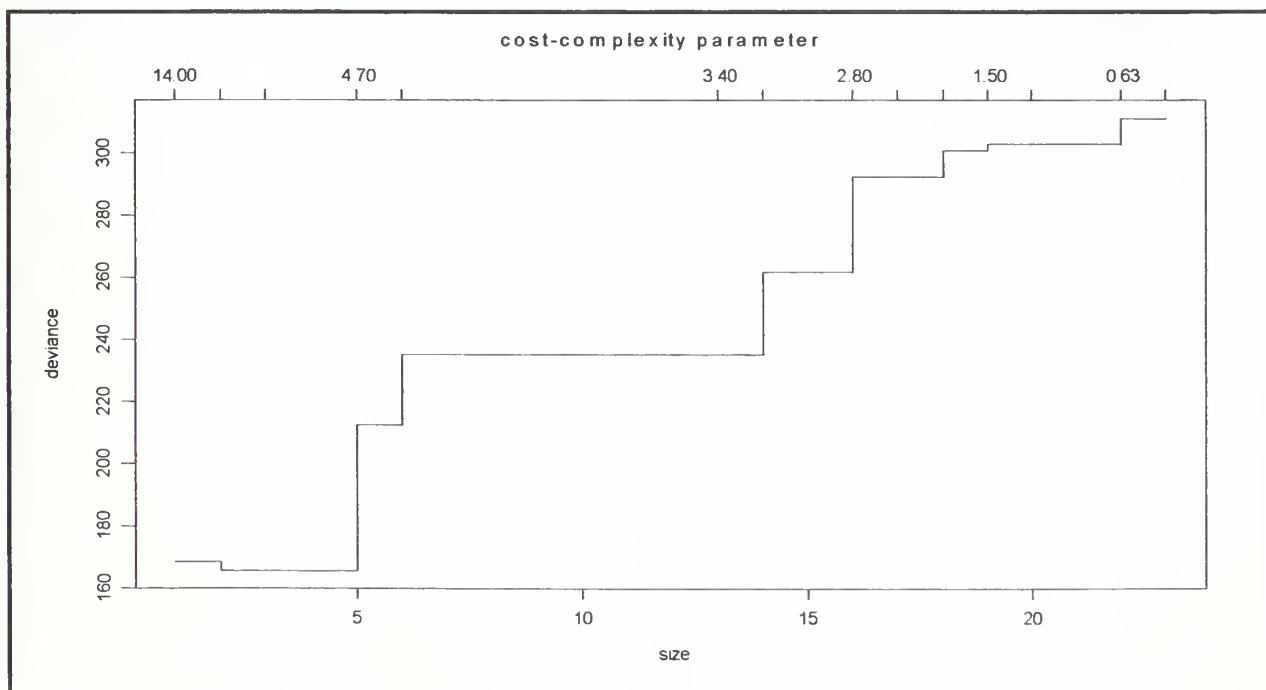
After a significant difference was shown between the two groups of test takers for primary flight training, a method of screening out those test retakers most likely to disqualify from training was developed.

The response variable for the primary flight-training data was *PRI.A.G*, whether the SNA graduated or disqualified from primary flight training. The predictor variables were *PBI*, the raw PBI score; *MVT*, the raw *MVT* score; *MCT*, the raw *MCT* score; *ANIT*, the raw *ANIT* score; *SAT*, the raw *SAT* score; *RACE*, composed of Asian, Black, Hispanic, Indian and White; and *SEX*. *API* test scores were excluded at first to study the effects of pre-flight-training predictors.

Figures 4.20 and 4.21 show the pruning and cross-validation plots, respectively, for test retakers in primary flight training. The pruning plot shows that the largest drop in deviance occurs at 13 terminal



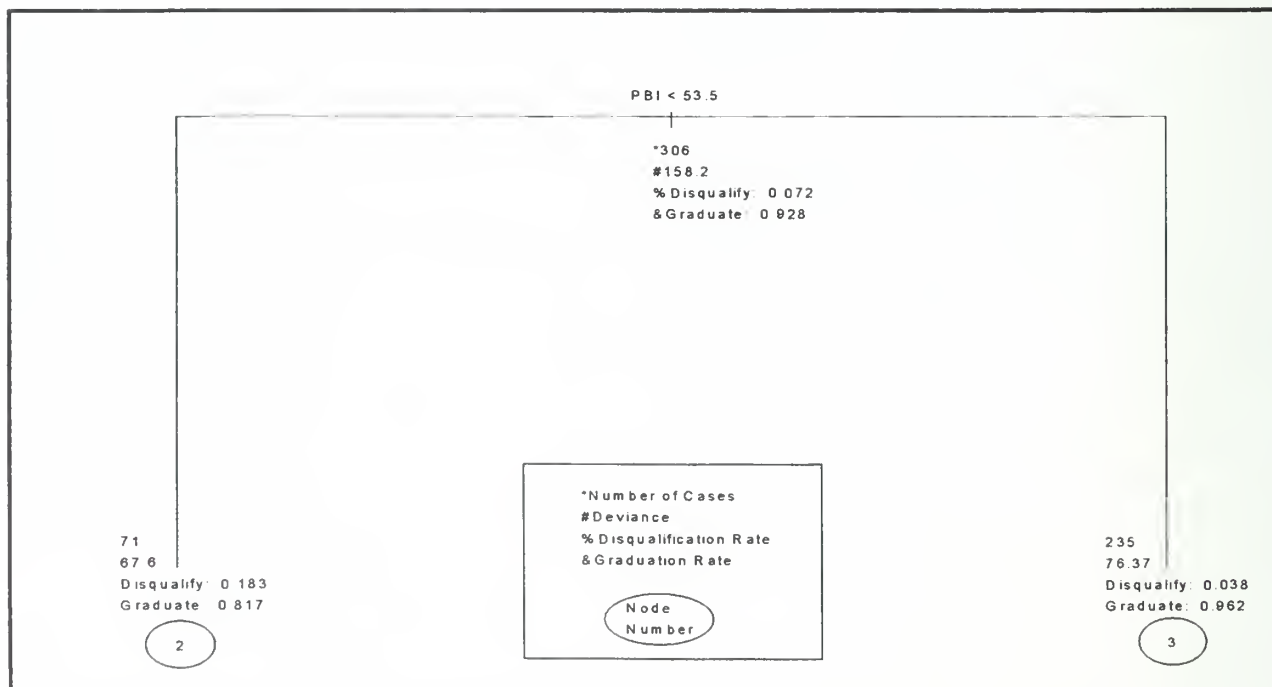
**Figure 4.20** Pruning Plot for ASTB Scores, Race and Sex as Predictors of Disqualification Among SNA Test Retakers in Primary Flight Training.



**Figure 4.21** Cross-Validation Plot for ASTB Scores, Race and Sex as Predictors of Disqualification Among SNA Test Retakers in Primary Flight Training.

nodes. However, the cross-validation plot indicates that two terminal nodes have the greatest predictability. The second largest drop in deviance in the pruning plot occurs at two terminal nodes. The ranges of the deviance, about 165 to 310, and the cost-complexity parameter, less than 0.63 to 14.00, are satisfactory in terms of good predictability for a pruned tree with two terminal nodes. Two terminal nodes were adopted.

Figure 4.22 is the pruned classification tree. The most important split occurs at a PBI score. This supports the Navy's assertion that the PBI is a good predictor of flight-training disqualification. The overall disqualification rate among test retakers is 7.2 percent, which is also reflected in Table 4.9. Those who scored less than 53.5 on the PBI suffered an 18.3 percent disqualification rate. Those who scored



**Figure 4.22** Classification Tree Pruned to Two Terminal Nodes for ASTB Scores, Race and Sex as Predictors of Disqualification Among SNA Test Retakers in Primary Flight Training.

above that figure had a 3.8 percent disqualification rate.

The GLRT statistic was computed from Table 4.10 to test the null hypothesis that the proportion of disqualifications in primary flight training for those SNA who repeated the ASTB and scored higher than 53.5 on the PBI is equal to the proportion of disqualifications for those SNA who repeated the ASTB and scored lower than 53.5 on the PBI:

$$\frac{\frac{9}{235} - \frac{13}{71}}{\sqrt{\frac{\left(\frac{9+13}{235+71}\right)\left(1 - \frac{9+13}{235+71}\right)(235+71)}{(235)(71)}}} = -4.1393.$$

|                        | Graduates | Disqualified* | Total |
|------------------------|-----------|---------------|-------|
| PBI<br>Score ><br>53.5 | 226       | 9             | 235   |
| PBI<br>Score <<br>53.5 | 58        | 13            | 71    |
| <b>Total</b>           | 284       | 22            | 306   |

\*Disqualification other than by Not Physically Qualified

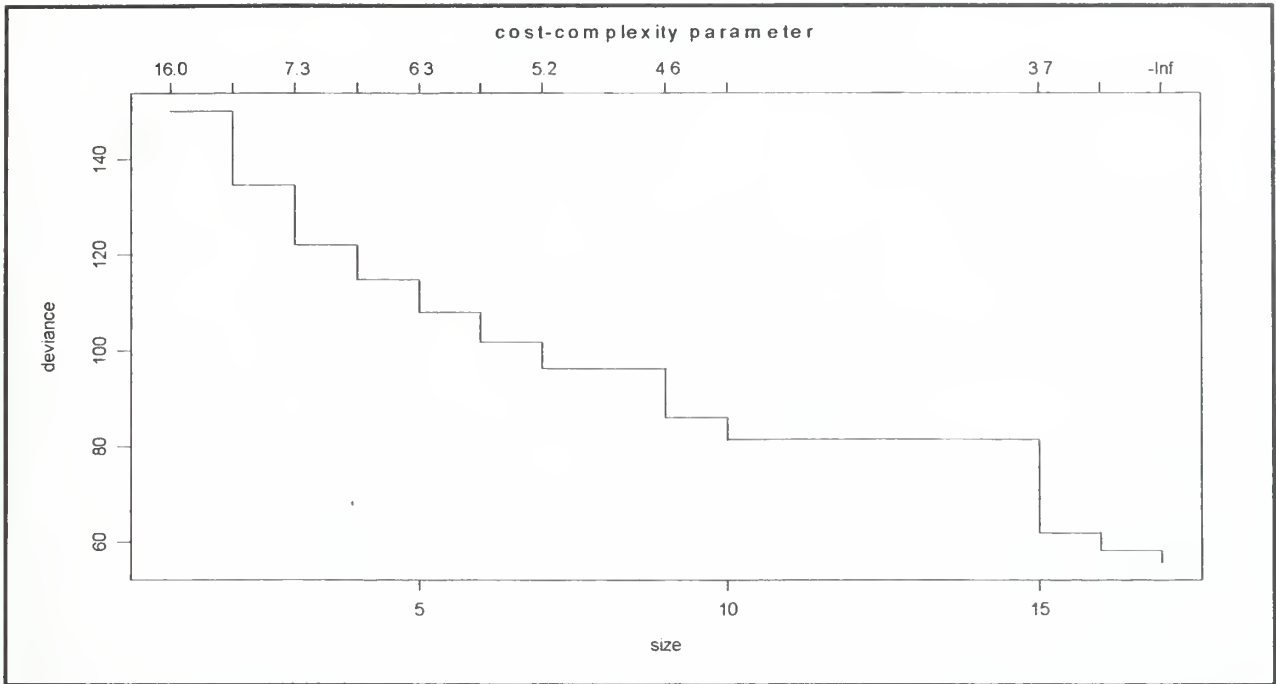
**Table 4.10** Graduation and Disqualification from Primary Flight Training for SNA Test Retakers By PBI Score.

This value of  $-4.1393$  is less than  $-z_{0.05} = -1.6449$ . A  $z$  of  $-4.1393$  corresponds to a  $p$ -value of  $0.0000$ . The null hypothesis may be rejected at a  $0.05$  level of significance in favor of the alternative that the proportion of disqualifications in primary flight training for those SNA who repeated the ASTB and scored higher than  $53.5$  on the PBI is less than the proportion of disqualifications for those SNA who repeated the ASTB and scored lower than  $53.5$ .

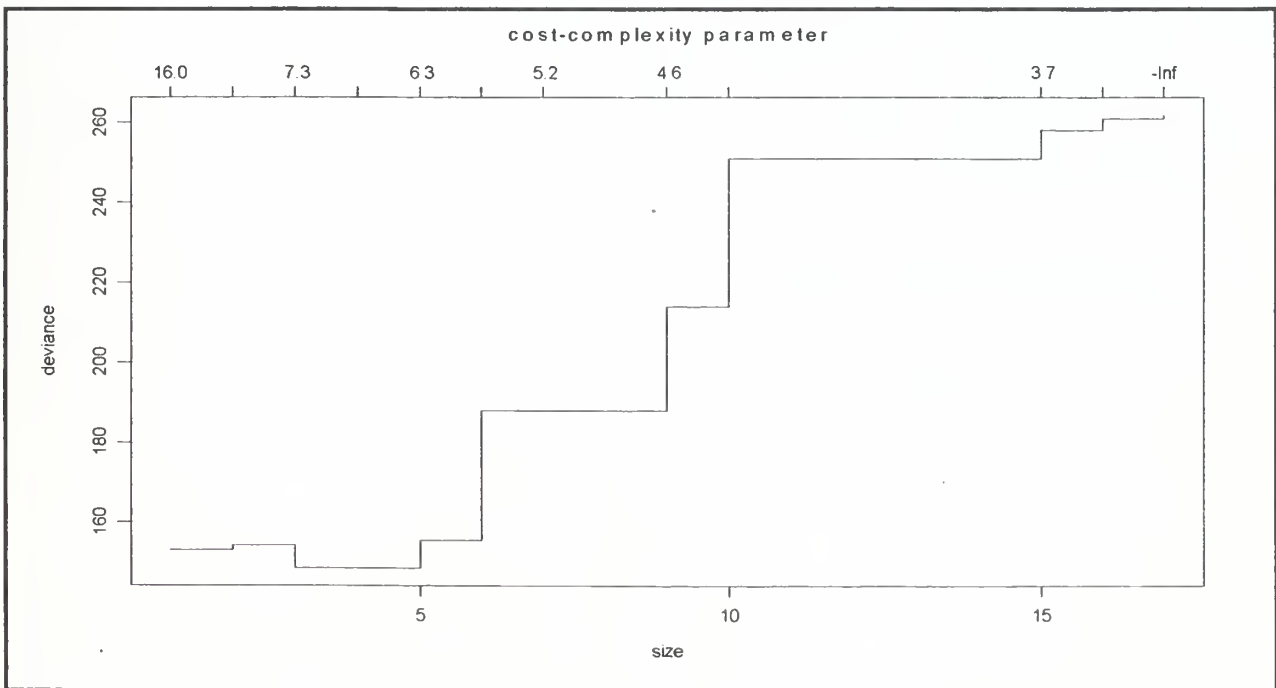
The disqualification rate in primary flight training among SNA who repeated the ASTB was significantly lower for those who scored greater than  $53.5$  on the PBI.

The analysis then included API test scores to determine API characteristics of the test retakers. The response variable and predictor variables remained the same with the addition of *PAERO*, *FAERO*, *PENGINE*, *FENGINE*, *FNAV*, *FMET* and *FFRR*. Figures 4.23 and 4.24 are the pruning and cross-validation plots, respectively. The pruning plot shows that the rate of decrease in deviance generally tapers off after three terminal nodes with large decreases occurring at nine and 15 terminal nodes. The cross-validation plot indicates that three terminal nodes have the greatest predictability with relatively large deviance and cost-complexity parameter ranges. Three terminal nodes were chosen.

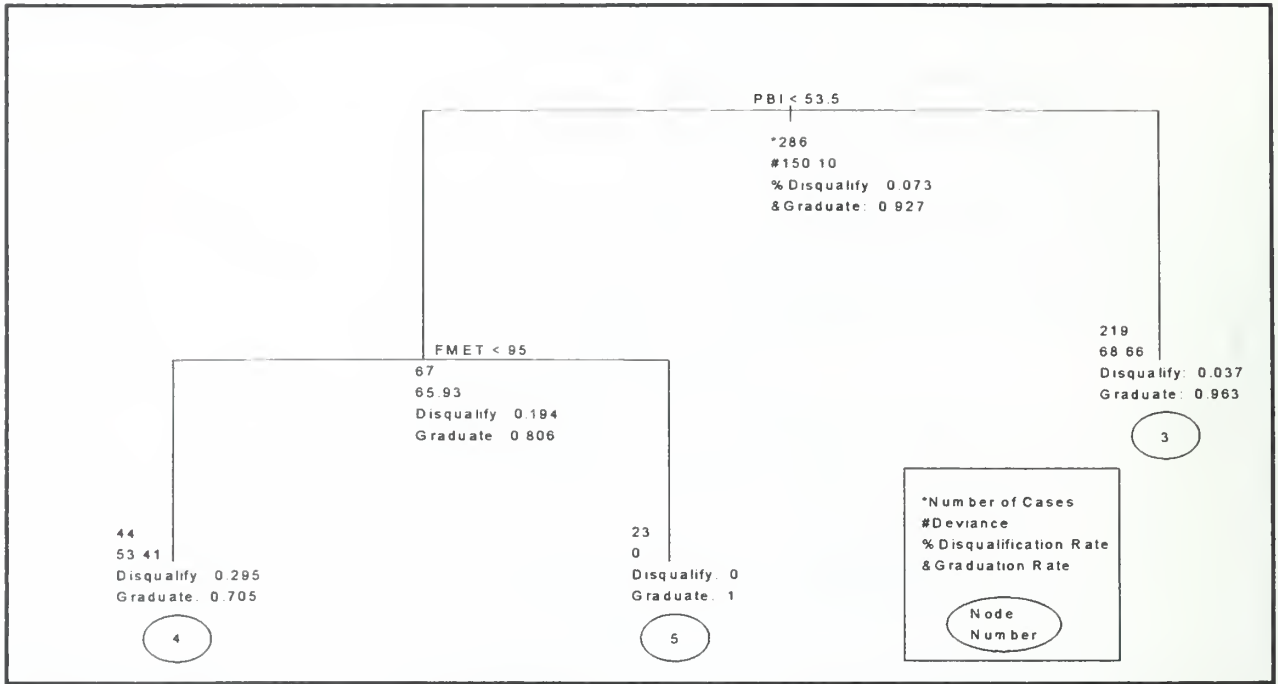
Figure 4.25 is the classification tree pruned to three terminal nodes. PBI is still the most important splitting criterion with the value remaining the same as in Figure 4.22. The other splitting criterion is the final meteorology exam in API. Node 4 contains those SNA who retook the ASTB, had a raw PBI score less than  $53.5$  and who scored less than  $95$  on the FMET. The disqualification rate among those 44 is  $29.5$  percent. All 23 SNA in node 5 graduated from primary flight training.



**Figure 4.23** Pruning Plot for ASTB Scores, API Test Scores, Race and Sex as Predictors of Disqualification Among SNA Test Retakers in Primary Flight Training.



**Figure 4.24** Cross-Validation Plot for ASTB Scores, API Test Scores, Race and Sex as Predictors of Disqualification Among SNA Test Retakers in Primary Flight Training.



**Figure 4.25** Classification Tree Pruned to Three Terminal Nodes for ASTB Scores, API Test Scores, Race and Sex as Predictors of Disqualification Among SNA Test Retakers in Primary Flight Training.

## V. COST CONSIDERATIONS

Disqualification from API and primary flight training costs the Navy several million dollars every year. The Analysis and Costing Division of the Naval Education and Training Professional Development and Technology Center in Pensacola formulated the cost data found in Table 5.1. This table is the average taxpayer cost for each API and primary flight-training completer for Fiscal Year 1996.

The cost per completion for primary flight training is computed by adding the per capita cost for primary flight training to the total cost of attrition, divided evenly among all completers. For example, the per capita in-stage attrition cost for primary flight training is the per capita flight-hour cost multiplied by the average aircraft hours flown by a disqualified SNA, divided evenly among all completers. The per capita prior-stage attrition cost for primary flight training is the cost of completion for API multiplied by the number of SNA who disqualified from primary flight training, divided evenly among all completers. The total cost of attrition for primary flight training is the sum of the in-stage and prior-stage attrition costs multiplied by the number of completers.

For Fiscal Year 1996, the cost of a SNA who disqualified from API was \$472,322 (Table 5.1) divided by 39 disqualifiers or \$12,111. For primary flight training, the cost per disqualified SNA was \$4,292,144 (Table 5.1) divided by 52 disqualifiers or \$82,541. The total cost of primary flight-training disqualification due to disqualification by other than NPQ was \$82,541 multiplied by 37 non-NPQ disqualifiers or \$3,054,017.



| Syllabus             | Syllabus Weeks | Syllabus Flight Hours | Per Capita (Dollars) | In Stage Attrition (Dollars)* | Prior Stage Attrition (Dollars)# | Cost per Completion (Dollars) |
|----------------------|----------------|-----------------------|----------------------|-------------------------------|----------------------------------|-------------------------------|
| <b>API</b>           | 6              | N/A                   |                      |                               |                                  |                               |
| O&M (1)              |                |                       | 2,642                | 44                            | 0                                | 2,686                         |
| MP P&A (2)           |                |                       | 5,090                | 85                            | 0                                | 5,175                         |
| STU P&A (3)          |                |                       | 8,156                | 135                           | 0                                | 8,291                         |
| OTHER (4)            |                |                       | 866                  | 14                            | 0                                | 880                           |
| <b>API Total</b>     |                |                       | 16,754               | 278                           | 0                                | 17,032                        |
| <b>Primary</b>       | 22.6           | 77.4                  |                      |                               |                                  |                               |
| O&M (1)              |                |                       | 49,327               | 1,989                         | 238                              | 51,554                        |
| MP P&A (2)           |                |                       | 35,803               | 2,047                         | 459                              | 38,309                        |
| STU P&A (3)          |                |                       | 29,826               | 1,706                         | 734                              | 32,266                        |
| OTHER (4)            |                |                       | 1,090                | 61                            | 78                               | 1,229                         |
| <b>Primary Total</b> |                |                       | 116,046              | 5,803                         | 1,509                            | 123,358                       |

**Total Attrition Cost for API = \$278 x 1,699 Completers = \$472,322**

**Total Attrition Cost for Primary = (\$5,803 + \$1,509) x 587 Completers = \$4,292,144**

\*Computed as (per capita weekly cost)\*(average number of weeks at which SNA/SNFO disqualified)/(number of SNA/SNFO who completed) for API.

Computed as (per capita flight hour cost)\*(average aircraft hours flown by disqualified SNA)/(number of SNA who completed) for primary flight training.

#Computed as (per capita cost for API)\*(number of SNA who disqualified in primary flight training)/(number of SNA who completed primary flight training)

(1) Operations and Maintenance - Direct and indirect costs of instructors, support personnel, curriculum materials and development, flight gear, flight operations, simulator operations, aircraft maintenance, simulator maintenance, supplies, contracts, equipment, equipment maintenance and base support costs

(2) Military Personnel Pay and Allowances - Navy military pay and allowances for instructors and support personnel

(3) Student Pay and Allowances - Navy military pay and allowances for SNA/SNFO

(4) Other costs - Medical, housing, munitions and NAVAIR support

**Table 5.1** Average Taxpayer Cost Per API and Primary Flight Training Completion For Fiscal Year 1996. Derived from Ref. [19].

## VI. CONCLUSIONS AND RECOMMENDATIONS

### A. CONCLUSIONS

Since the dawn of naval aviation, the Navy has endeavored to minimize disqualification in its flight-training program. The aviation selection tests used since World War Two have attempted to identify attributes that characterize success or failure. Predicting academic performance has been relatively easy. Tests that demonstrate mechanical comprehension, mathematical and verbal knowledge and skills, as shown in Figure 4.16, are good predictors of academic performance in API. Figure 4.19 showed that the most important criterion for disqualification among SNA in API was race. This is likely due to different academic backgrounds between ethnic groups.

Trying to predict how a person will fare under the demands of flight training is a more difficult task. The analysis showed that individual answers to the BI had no predictive validity for SNA flight grades in primary flight training. Frank and Baisden [Ref. 6:p. 6] state that PBI and FOBI scores help predict disqualification. The data in Figure 4.13 support their belief that the PBI is the most important criterion for predicting disqualification among SNA in primary flight training. The BI is a questionnaire comprising 76 questions concerning a candidate's personal history and background. It measures a person's exposure to academics, athletics and interest in the military, particularly aviation. Thus, a person's chances of graduating from primary flight training appear to depend not on one's academic prowess, race or sex but on one's desire and motivation. A system to predict whether an individual will DOR or be an academic or flight failure in primary flight training will never be perfect because of the difficulty in quantifying desire and motivation.

Table 4.9 shows that those SNA who repeated the ASTB had a significantly higher disqualification rate (7.2 percent) in primary flight training than those who took it only once (2.5 percent). Figure 4.13 shows that the disqualification rate for those SNA who scored less than 56.5 on the PBI was 9.2 percent. For those who scored more than 56.5, it was 1.8 percent. From Figure 4.22, the disqualification rate was 18 percent among those who repeated the ASTB and scored less than 53.5 on the PBI. For those who repeated the ASTB and scored more than 53.5 on the PBI, it was 3.8 percent.

Figure 4.25 shows that for those who repeated the ASTB, scored less than 53.5 on the PBI and less than 95 on the final meteorology test in API, the disqualification rate was 30 percent. The average overall API score for the latter group was 91.8 and the average for all other SNA was 93.8. The average overall API score for all SNA was 93.7. The standard deviation of the average for all SNA was 3.35. Thus, the average overall API score for the group that repeated the ASTB, scored less than 53.5 on the PBI and less than 95 on the final meteorology test was 0.57 standard deviations below the overall API average for all SNA.

## **B. RECOMMENDATIONS**

A policy recommendation comprising three options may be made based on Figure 4.13, Table 4.9 and Figure 4.25. The first two options could be applied if a large reduction in the number of pilots needed in a fiscal year was necessary. The third option could be applied to current pilot training requirements.

### **1. First Force Reduction Option**

In Figure 4.13, the raw PBI score of 56.5 translates to a standardized score of 6. The first option is to raise the standardized minimum PBI score from 4 to 6. This would have disqualified 260 out of

702 or 37 percent of the SNA in Figure 4.13. The result would have been a disqualification rate of 1.8 percent among the remaining 442 SNA rather than 4.6 percent among the original 702. The reduction in disqualification would have been  $1 - 0.018/0.046 = 0.609$  or 61 percent. For Fiscal Year 1996, the percentage of SNA who disqualified from primary flight training for other than NPQ was 37 out of 639 or 5.8 percent.

Assuming that all 639 SNA had scored a 6 or better on the PBI, then a 61 percent reduction in disqualification would have meant a disqualification rate of 2.3 percent. The number of disqualifications due to other than NPQ would have been 15. The total cost of disqualification due to other than NPQ would then have been  $\$82,541 \times 15 = \$1,238,115$ . This would have saved American taxpayers the balance, i.e.,  $\$3,054,017 - \$1,238,115 = \$1,815,902$ .

## **2. Second Force Reduction Option**

The second option is to accept only those candidates who meet the minimum ASTB scores on their first attempt. This would have disqualified 306 out of 743 or 41 percent of the SNA in Table 4.9. The result would have been a disqualification rate of 2.5 percent among the remaining 437 SNA rather than 7.2 percent among the original 743. The reduction in disqualification would have been  $1 - 0.025/0.072 = 0.653$  or 65 percent.

Assuming that all 639 SNA from Fiscal Year 1996 had taken the ASTB only once, then a 65 percent reduction in disqualification would have meant a disqualification rate of 2.0 percent. The number of disqualifications due to other than NPQ would have been 13. The total cost of disqualification due to other than NPQ would then have been

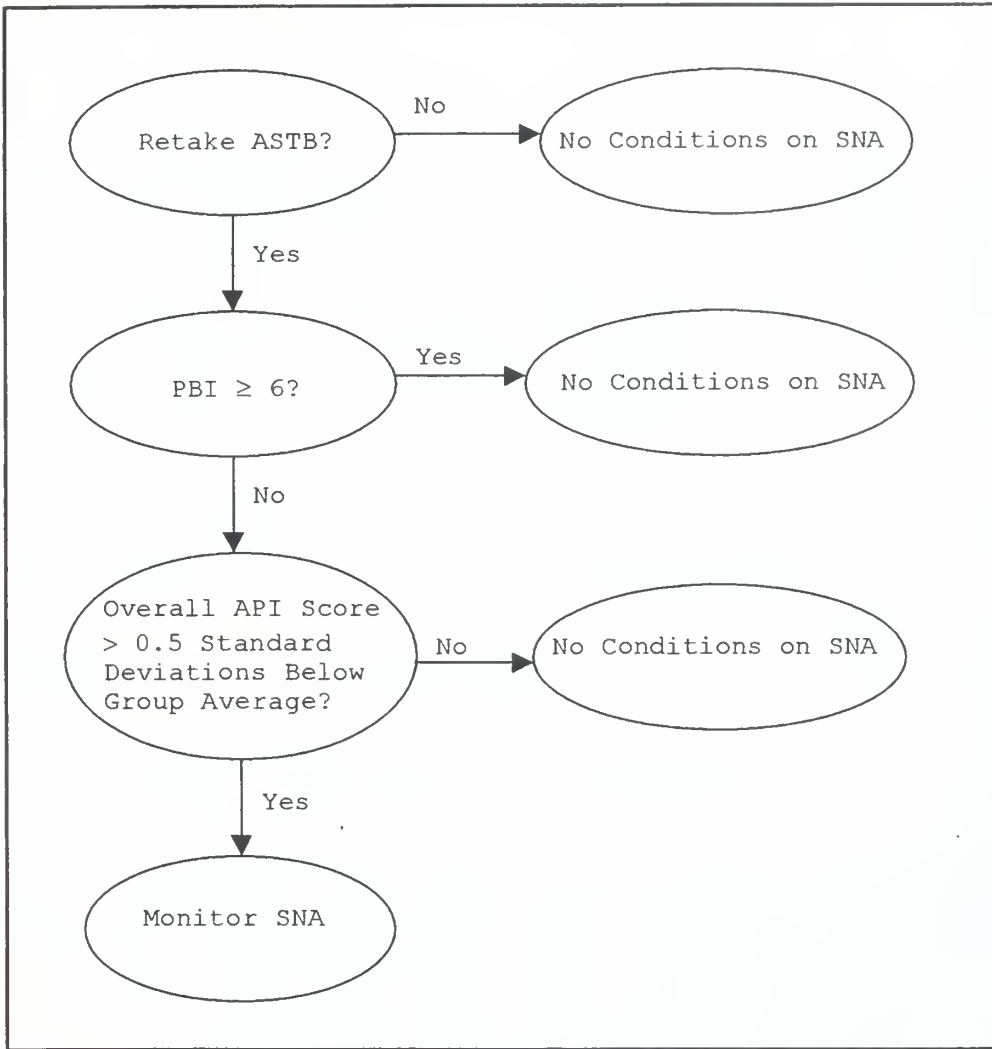
$\$82,541 \times 13 = \$1,073,033$ . This would have saved American taxpayers the balance, i.e.,  $\$3,054,017 - \$1,073,033 = \$1,980,984$ .

### 3. Third Option; Status Quo

The third option, from Figure 4.25, would have no impact on the number of SNA entering primary flight training. This thesis found no reason to alter the current qualification criteria. This option assumes that the standardized PBI score would not fluctuate from test to test for a SNA who repeated the ASTB. It provides no limit on the number of ASTB retakes allowed.

If a SNA had to retake the ASTB to meet the minimums, scored a 4 or 5 on the PBI and his or her overall API score is more than 0.5 standard deviations below the group average, then he or she is at a disproportionately high risk for disqualification from primary flight training. This SNA has demonstrated borderline motivation for aviation training and weak academic skills. However, steps could be taken that may maximize his or her likelihood of graduating. The Navy allows extra flights and a longer time for training to those SNA who are having difficulty in primary flight training. It could be wise to allow those SNA in this risk group extra flights and a longer time for training at the beginning of primary flight training, before problems become apparent. This option is presented in Figure 6.1.

The percentage of SNA who disqualified from primary flight training for other than NPQ for Fiscal Year 1996 was 37 out of 639 or 5.8 percent. If this could have been reduced to 3.7 percent (from Node 3 of Figure 4.25), the number of disqualifications due to other than NPQ would have been 24. The total cost of disqualification due to other than NPQ would then have been  $\$82,541 \times 24 = \$1,980,984$ . This would



**Figure 6.1** Flow Diagram for Third Option Based on Retaking ASTB, PBI Score and Overall API Score.

have saved American taxpayers the balance, i.e., \$3,054,017 - \$1,980,984 = \$1,073,033.



## APPENDIX. LIST OF S-PLUS COMMANDS

#Figures 4.1 and 4.2

```
>bi.first.tree tree(PRIGRADE ~ BI1 + BI2 + BI3 + BI4 + BI5 + BI6 + BI7 + BI8 +
BI9 + BI10 + BI11 + BI12 + BI13 + BI14 + BI15 + BI16 + BI17 + BI18 + BI19 +
BI20 + BI21 + BI22 + BI23 + BI24 + BI25 + BI26 + BI27 + BI28 + BI29 + BI30 +
BI31 + BI32 + BI33 + BI34 + BI35 + BI36 + BI37 + BI38 + BI39 + BI40 + BI41 +
BI42 + BI43 + BI44 + BI45 + BI46 + BI47 + BI48 + BI49 + BI50 + BI51 + BI52 +
BI53 + BI54 + BI55 + BI56 + BI57 + BI58 + BI59 + BI60 + BI61 + BI62 + BI63 +
BI64 + BI65 + BI66 + BI67 + BI68 + BI69 + BI70 + BI71 + BI72 + BI73 + BI74 +
BI75 + BI76,data=Biitem.first.conv,na.action=na.omit)
>bi.first.cv_cv.tree(bi.first.tree,FUN=prune.tree)
>plot(bi.first.cv)
>bi.first.cv_cv.tree(bi.first.tree,FUN=prune.tree)
>plot(bi.first.cv)
```

#Figure 4.3

```
>m_model.frame(bi.first.tree)
>five.fold_sample(5,length(m[[1]]),T)
>bi.first.cv_cv.tree(bi.first.tree,five.fold,FUN=prune.tree)
>plot(bi.first.cv)
```

#Figure 4.4

```
>bi.first.prune_prune.tree(bi.first.tree)
>plot(bi.first.prune)
```

#Figure 4.5

```
>bi.first.prune.best15_prune.tree(bi.first.tree,best=15)
>plot(bi.first.prune.best15)
```

#Figure 4.6

```
>bi.api.stat.tree_tree(API.STAT ~ BI1 + BI2 + BI3 + BI4 + BI5 + BI6 + BI7 +
BI8 + BI9 + BI10 + BI11 + BI12 + BI13 + BI14 + BI15 + BI16 + BI17 + BI18 +
BI19 + BI20 + BI21 + BI22 + BI23 + BI24 + BI25 + BI26 + BI27 + BI28 + BI29 +
BI30 + BI31 + BI32 + BI33 + BI34 + BI35 + BI36 + BI37 + BI38 + BI39 + BI40 +
BI41 + BI42 + BI43 + BI44 + BI45 + BI46 + BI47 + BI48 + BI49 + BI50 + BI51 +
BI52 + BI53 + BI54 + BI55 + BI56 + BI57 + BI58 + BI59 + BI60 + BI61 + BI62 +
BI63 + BI64 + BI65 + BI66 + BI67 + BI68 + BI69 + BI70 + BI71 + BI72 + BI73 +
BI74 + BI75 + BI76,data=Biitem.first.conv.cohort,na.action=na.omit)
>bi.api.stat.prune_prune.tree(bi.api.stat.tree)
>plot(bi.api.stat.prune)
```

#Figure 4.7

```
>m_model.frame(bi.api.stat.tree)
>five.fold_sample(5,length(m[[1]]),T)
>bi.api.stat.cv_cv.tree(bi.api.stat.tree,five.fold,FUN=prune.tree)
>plot(bi.api.stat.cv)
```

#Figure 4.8

```
>bi.api.stat.prune.best2_prune.tree(bi.api.stat.tree,best=2)
>plot(bi.api.stat.prune.best2)
```

#Figure 4.9

```
>pri.grade.tree_tree(PRIGRADE~TEST+PBI+MVT+MCT+ANIT+SAT+PAERO+FAERO+PENGINE+
FENGINE+FNAV+FMET+FFRR+RACE+SEX+
+PRIACAD+FLTHRS,data=Pricart,na.action=na.omit)
>pri.grade.prune_prune.tree(pri.grade.tree)
>plot(pri.grade.prune)
```



```

#Figure 4.10
>m_model.frame(pri.grade.tree)
>five.fold_sample(5,length(m[[1]]),T)
>pri.grade.cv_cv.tree(pri.grade.tree,five.fold,FUN=prune.tree)
>plot(pri.grade.cv)

#Figure 4.11
>pri.attrite.tree_tree(PRI.A.G~TEST+PBI+MVT+MCT+ANIT+SAT+PAERO+FAERO+PENGINE+
FENGINE+FNAV+FMET+FFRR+RACE+SEX,
data=Pricart,na.action=na.omit)
>pri.attrite.prune_prune.tree(pri.attrite.tree)
>plot(pri.attrite.prune)

#Figure 4.12
>m_model.frame(pri.attrite.tree)
>five.fold_sample(5,length(m[[1]]),T)
>pri.attrite.cv_cv.tree(pri.attrite.tree,five.fold,FUN=prune.tree)
>plot(pri.attrite.cv)

#Figure 4.13
>pri.attrite.prune.best2_prune.tree(pri.attrite.tree,best=2)
>plot(pri.attrite.prune.best2)

#Figure 4.14
>api.grade.tree_tree(OVERALL~TEST+PBI+MVT+MCT+ANIT+SAT+RACE+SEX,data=
apipilotcohort,na.action=na.omit)
>api.grade.prune_prune.tree(api.grade.tree)
>plot(api.grade.prune)

#Figure 4.15
>m_model.frame(api.grade.tree)
>five.fold_sample(5,length(m[[1]]),T)
>api.grade.cv_cv.tree(api.grade.tree,five.fold,FUN=prune.tree)
>plot(api.grade.cv)

#Figure 4.16
>api.grade.prune.best7_prune.tree(api.grade.tree,best=7)
>plot(api.grade.prune.best7)

#Figure 4.17
>api.attrite.tree_tree(ATTRITE~TEST+PBI+MVT+MCT+ANIT+SAT+RACE+SEX,data=
apipilotcohort,na.action=na.omit)
>api.attrite.prune_prune.tree(api.attrite.tree)
>plot(api.attrite.prune)

#Figure 4.18
>m_model.frame(api.attrite.tree)
>five.fold_sample(5,length(m[[1]]),T)
>api.attrite.cv_cv.tree(api.attrite.tree,five.fold,FUN=prune.tree)
>plot(api.attrite.cv)

#Figure 4.19
>api.attrite.prune.best2_prune.tree(api.attrite.tree,best=2)
>plot(api.attrite.prune.best2)

#Figure 4.20
>pri.retake.tree_tree(PRI.A.G~PBI+MCT+MVT+SAT+ANIT+RACE+SEX,data=Pricart,
subset=TEST=="R",na.action=na.omit)
>pri.retake.prune_prune.tree(pri.retake.tree)
>plot(pri.retake.prune)

```

```

#Figure 4.21
>m_model.frame(pri.retake.tree)
>five.fold_sample(5,length(m[[1]]),T)
>pri.retake.cv_cv.tree(pri.retake.tree,five.fold,FUN=prune.tree)
>plot(pri.retake.cv)

#Figure 4.22
>pri.retake.prune.best2_prune.tree(pri.retake.tree,best=2)
>plot(pri.retake.prune.best2)

#Figure 4.23
>pri.apitest.retake.tree_tree(PRI.A.G~PBI+MCT+MVT+SAT+ANIT+PAERO+FAERO+ENGINE
+FENGINE+FNAV+FMET+FFRR+RACE+SEX,data=Pricart,subset=TEST=="R",na.action=
na.omit)
>pri.apitest.retake.prune_prune.tree(pri.apitest.retake.tree)
>plot(pri.apitest.retake.prune)

#Figure 4.24
>m_model.frame(pri.apitest.retake.tree)
>five.fold_sample(5,length(m[[1]]),T)
>pri.apitest.retake.cv_cv.tree(pri.apitest.retake.tree,five.fold,FUN=
prune.tree)
>plot(pri.apitest.retake.cv)

#Figure 4.25
>pri.apitest.retake.prune.best3_prune.tree(pri.apitest.retake.tree,best=3)
>plot(pri.apitest.retake.prune.best3)

```



## LIST OF REFERENCES

1. Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J., *Classification and Regression Trees*, Wadsworth, Inc., 1984.
2. Purcell, T. S., *The Use of Classification Trees to Characterize the Attrition Process for Army Manpower Models*, MS Thesis, Operations Research Department, Naval Postgraduate School, 1997.
3. Data Analysis Products Division, *S-PLUS 4 Guide to Statistics*, MathSoft, Inc., 1997.
4. Fought, S. O., *Evaluating Analysis*, p. 3, Rational Decision Making 9-2, August 1994.
5. North, R. A., and Griffin, G. R., *Aviator Selection 1919-1977*, p. 5, Naval Aerospace Research Laboratory, 1977.
6. Frank, L. H., and Baisden, A. G., *The 1992 Navy and Marine Corps Aviation Selection Test Battery Development*, Naval Aerospace and Operational Medical Institute, 1993.
7. Williams, Henry, code415@opmed1.med.navy.mil, Operational Psychology Department, Naval Aerospace and Operational Medical Institute, 10 September 1993, personal e-mail.
8. "What does it take to be a Naval Aviator?"  
[http://www.ncts.navy.mil/vt6/what\\_does\\_it\\_take.html](http://www.ncts.navy.mil/vt6/what_does_it_take.html), 9 November 1997, Internet.
9. Rich, VT-3 Student Control Officer, 2 October 1997, telephone conversation.
10. "How Is the Training Organized?"  
[http://www.ncts.navy.mil/vt6/how\\_does\\_training\\_work.html](http://www.ncts.navy.mil/vt6/how_does_training_work.html), 9 November 1997, Internet.
11. Schreier, Chris, Aviation Warfare Training, Chief of Naval Education and Training, 2 June 1997, facsimile.
12. "Aviation Selection Test Battery Brief," Naval Aerospace and Operational Medical Institute, 24 May 1993.
13. Hyde, Chief of Naval Air Training Office, 10 June 1997, facsimile.
14. Fleiss, J. L., *Statistical Methods for Rates and Proportions*, Second Edition, John Wiley & Sons, 1981.
15. Larsen, R. J., and Marx, M. L., *An Introduction to Mathematical Statistics and Its Applications*, Second Edition, p. 380, Prentice-Hall, 1986.

16. Lambert, Brian, nasc.0310@smtp.cnet.navy.mil, API Curriculum Development Officer, Naval Aviation Schools Command, 17 November 1997, personal e-mail.
17. "Student Naval Aviator Training and Administration Manual," Chief of Naval Air Training Instruction 1500.4E, p. VIII-1, Chief of Naval Air Training, 1993.
18. "Examiner's Manual for The United States Navy and Marine Corps Aviation Selection Tests," pp. 5-7, Naval Aerospace and Operational Medical Institute.
19. McDonald III, L. W., Analysis and Costing Division Head, Naval Education and Training Professional Development and Technology Center, 25 November 1997, personal letter.

**INITIAL DISTRIBUTION LIST**

|    |   | No. Copies |
|----|---|------------|
| 1. | Defense Technical Information Center.....<br>8725 John J. Kingman Rd., Ste 0944<br>Ft. Belvoir, VA 22060-6218                     | 2          |
| 2. | Dudley Knox Library.....<br>Naval Postgraduate School<br>411 Dyer Rd.<br>Monterey, CA 93943-5101                                  | 2          |
| 3. | Chief of Naval Education and Training.....<br>CNET 00<br>250 Dallas St.<br>Pensacola, FL 32508-5220                               | 2          |
| 4. | Director, Resource Appraisal Division, N-81.....<br>Chief of Naval Operations (N-81)<br>Navy Department<br>Washington, DC 20350   | 2          |
| 5. | Operational Psychology Department (Code 41).....<br>Naval Aerospace and Operational Medical Institute<br>Pensacola, FL 32508-1047 | 2          |
| 6. | Professor Robert R. Read.....<br>Operations Research Department, Code OR/Re<br>Naval Postgraduate School<br>Monterey, CA 93943    | 2          |
| 7. | Professor Samuel E. Buttrey.....<br>Operations Research Department, Code OR/Sb<br>Naval Postgraduate School<br>Monterey, CA 93943 | 1          |
| 8. | LT Erich J. Wahl, USN.....<br>2841 SW 171 <sup>st</sup> St.<br>Burien, WA 98166   | 2          |



DUDLEY KNOX LIBRARY  
NAVAL POSTGRADUATE SCHOOL  
MONTEREY CA 93943-5101



66 553NPS 3771  
TH  
11/99 22527-106 MULTI







DUDLEY KNOX LIBRARY



3 2768 00366349 3