THE INFLUENCE OF FINANCIAL AID FACTORS ON FRESHMAN ENROLLMENT DECISIONS AT STANFORD UNIVERSITY

BY

D. CONWAY and J. VERDUCCI

TECHNICAL REPORT NO. 139
MARCH 1979

PREPARED UNDER THE AUSPICES OF
NATIONAL SCIENCE FOUNDATION GRANT
MCS 78-17730

Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
THE INFLUENCE OF FINANCIAL AID FACTORS
ON FRESHMAN ENROLLMENT DECISIONS
AT STANFORD UNIVERSITY

BY

D. CONWAY and J. VERDUCCI

TECHNICAL REPORT NO. 139
MARCH 1979

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT MCS 78-17730*

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

*This work was partially supported by a grant from the Lilly Endowment.
The Influence of Financial Aid Policies on Freshman Enrollment Decisions at Stanford University

By

D. Conway and J. Verducci *

1. INTRODUCTION AND RESEARCH OBJECTIVES

The preservation of the diversity, scope, and vitality of an undergraduate academic program depends in large measure upon an institution's continual ability to attract the most talented, motivated, and promising high school graduates. Two important developments pertaining to college costs and enrollments present new challenges to private universities in attracting the most academically qualified applicants. First, demographic projections predict a decline in the absolute number of 18 year olds by 1980. This is likely to enhance competition among major universities for the most highly qualified applicants. Second, the costs of attending a private university have been increasing very rapidly, which further enlarge differences in tuition rates between private and public universities. This suggests that the number of applicants to private universities requiring financial assistance may be increasing and that the financial aid package offered by the institution may become a more important factor in an applicant's enrollment decision. As a result, private universities face new challenges in the optimal allocation and distribution of their limited financial aid resources.

In view of these developments, the research presented in this paper was motivated by the need for Stanford University to examine carefully and understand the factors affecting the enrollment decisions of its admitted freshman applicants. Due to the higher tuition cost of attending a private university, such as Stanford, it is especially important to

---

*In collaboration with D. Hopkins, W. Masay, and I. Olkin.
A second objective of this research is to study separately and in more detail factors affecting the enrollment decisions of the highest academically qualified applicants to Stanford. Because many good schools typically compete for these applicants, they are the least likely to accept admission to Stanford. It is of particular interest to examine the enrollment decisions of this group carefully, since the quality of the undergraduate academic program at Stanford depends on attracting the most highly qualified applicants.

A final objective of this research is to examine the influence of size and composition of the financial aid offers made by Stanford on the applicants' decisions to enroll. A 1973 study at the Massachusetts Institute of Technology concluded that the total amount of financial aid offered an applicant chiefly influenced the decision to enroll at M.I.T., rather than how the aid was 'packaged' into scholarships, loans, and jobs. Since this conclusion has important financial aid policy implications to private universities, the present research verifies whether this statement also applies to Stanford applicants.

Because of the desire to study influences of financial aid policies on enrollment decisions, the research presented in this report restricts analysis to the group of admitted applicants to Stanford who also applied for financial aid. The research is also limited to the examination of measurable factors that are presently available from data in the admissions and financial aid applications. These forms comprise the only complete and reliable source of information concerning admitted aid applicants. Consequently, important factors such as reputation of departments, academic grades of applicants, and the many other personal factors that often influence the decision to enroll could not be studied.

It should be noted that throughout this report, we use the word 'effect' to describe any measurable association between factors of interest and the enrollment decision, rather than to imply any casual relationship. A previous study (Hopkins, Massy, and Curry (1973)) revealed that the effects of measurable factors from admissions and aid applications on the enrollment decision may be slight. To detect associations where they exist and to ensure that perceived effects of certain factors did not result merely by chance, five academic years of admissions and financial aid data are analyzed. In this way, factors affecting the enrollment decision are considered important only if the measurable effects of these factors are consistent and stable across all five years of data.

Finally, the approach to the research is descriptive and explanatory in nature. Models are developed which explain the enrollment decisions based on past data in a satisfactory manner, and, barring abrupt changes in policy or behavior, will fit data in the near future. However, since the data analyzed are historical rather than experimental, the research attempts only to measure and monitor the effects of key factors on the enrollment decision resulting from present financial aid policies and applicant behavior.
Each academic year, approximately 10,000 potential freshmen students apply for admission to Stanford and approximately 2,500 of these are accepted. Roughly one half of the admitted applicants also apply for financial aid. The target population for the study consists of this latter group of admitted applicants who also applied for financial aid. So as not to allow special circumstances to influence the results, we eliminated applicants from the target population who are supported by athletic scholarships as well as applicants from foreign countries. This results in a target population size of approximately 1,200 applicants for each academic year.

Both the Admissions Office and the Financial Aids Office at Stanford provided data for applicants admitted to freshmen standing during the five academic years 1973-1974 through 1977-1978. Models explaining the enrollment decisions of applicants in the target population were developed using the first four academic years of data while the last academic year served only to validate the final models.

Admissions Variables

A listing of the variables obtained from the Admissions and Financial Aids Office is given in Table 1. Although most of the variables obtained from the admission applications are self-explanatory, three variables require further description. Stanford Affiliation refers to any association the applicant's family may have with Stanford, such as a parent who is a staff or faculty member or a sibling who is a current student. Academic Rating is an index score of 1 to 5 assigned by the Admissions Office which evaluates an applicant's record of scholastic achievement.

Table 1
Variables Available for Analysis

I. Variables obtained from Admissions Office

A. Student
1. Birthdate
2. Sex
3. Minority Status
4. State of Residence
5. Type of High School (Public or Private)
6. Verbal SAT Score
7. Math SAT Score
8. Academic Rating
9. Nonacademic Rating
10. Intended Major
11. Enrollment Decision

B. Family
1. Number of Parents Living
2. Number of Parents Working
3. Parent's Marital Status
4. Number of Children in Family
5. Number of Children in College
6. Stanford Affiliation

II. Variables obtained from Financial Aids Office

A. Student
1. Student Budget
2. Student Savings
3. Summer Earnings
4. Veterans/Social Security Benefits
5. Other Resources
6. California State Scholarship
7. Outside Awards

B. Family
1. Parents' Offer
2. Parents' Contribution
3. Parents' Income
4. Parents' Assets

C. Stanford Financial Aid Award
1. Grant
2. Tuition Waiver
3. Loan
4. Job
Table 1 (Continued)

D. Derived

1. Total Resources
   (Parents' Contribution + Student
   Savings + Summer Earnings + Veterans/
   Social Security Benefits + Other
   Resources)

2. Other Scholarships
   (California State Scholarship +
   Outside Awards)

3. Student Need
   (Student Budget - Total Resources)

4. Stanford Need
   (Student Need - Other Scholarships)

5. Self-help
   (Loan + Job)

6. Scholarship
   (Grant + Tuition Waiver + Other
   Scholarships)

7. Total Aid
   (Scholarship + Self-help)

8. Aid Quality
   (Percentage of Total Aid that is
   Scholarship)

and aptitude. A score of 1 represents the highest academically rated
group. Nonacademic Rating is another index score of 1 to 5 assigned
by the Admissions Office which evaluates the applicant's achievement
in extra-curricular and leadership activities. Again, a score of 1
represents the highest rating.

The Financial Aid Process.

Before discussing the variables obtained from the Financial Aids
Office, it is helpful to give first an overview of the financial aid
application process. Families of applicants for admission to Stanford
who believe they require financial assistance provide information
regarding the family financial circumstances to the College Scholarship
Service by completing a Parent's Confidential Statement (PCS). This
agency evaluates the information pertaining to family income, assets,
and financial resources provided by the statement and calculates an
estimated parental contribution towards the costs of the freshmen year
at a particular college. Copies of the Financial Need Analysis determined
by the College Scholarship Service and the PCS are forwarded to the
colleges requested by the applicant. The Financial Aids Office at
Stanford reviews these reports for all admitted applicants and assesses
each applicant's need as the difference between the anticipated budget
for the freshmen year and the ability to meet this budget from such
resources as the parental contribution, anticipated summer earnings by
the applicant, and outside awards. In cases where an applicant's total
financial resources are insufficient to meet the anticipated budget, the
Financial Aids Office utilizes university financial aid funds to make up
the deficiency. The financial aid 'package' offered the applicant is assembled from two basic types of financial aid: (1) grants in the form of university scholarships and (2) self-help in the form of student loans and/or jobs. For those applicants having no demonstrated financial need but where the expected parental contribution is very high, a small amount of self-help may be offered the applicant to offset financial hardship to the family.

Financial Variables.

Some of the financial aid variables given in Table 1 also require further explanation. Student Budget is the anticipated cost of the freshmen year based on tuition fees, room and board, books, personal expenses, and travel home. Summer Earnings refers to the anticipated amount of money an applicant will earn during the summer just prior to entering college. Stanford Grant includes all scholarships from University funds as well as Basic Educational Opportunity Grants from the federal government. California State Scholarship is an award made by the State of California, based on academic merit and financial need, to high school graduates who are state residents. Outside Awards refer to other external grants awarded the applicant from such sources as National Merit Scholarships and various national or local service clubs. Parents' Offer is the amount the parents feel they can pay for their child's education during the freshmen year. Parents' Contribution is the expected contribution of the parents assessed by the Financial Aids Office. This amount is based upon the expected financial contribution computed by the College Scholarship Service. Parents' Income refers to the actual taxable family income of the preceding year of enrollment.

If anticipated changes in income for the current year exceed $1,000 and are well explained, Parents' Income then becomes the estimated taxable income of the current year. Parent's Assets is the net worth figure in the Financial Need Analysis of the College Scholarship Service. Finally, Tuition Waiver is a tuition credit given only to children of Stanford staff or faculty. The amount of the credit is based on the parent's length of service to the University rather than demonstrated financial need. For this reason, the target population includes only applicants who receive a Tuition Waiver and also apply for additional financial aid.

Changes in Financial Aid Awards.

It should be noted that individual data for the financial aid variables in Table 1 may change during the admissions and financial aid awards process. In particular, due to changes in the family financial circumstances or the receipt of an outside award, an applicant's financial aid package is adjusted accordingly. For example, if an applicant receives a National Merit Scholarship for $500, his Stanford grant would be reduced by $500.

On April 1st preceding the anticipated fall quarter of enrollment, the university notifies each applicant of the admission and financial aid decisions. The applicant returns his enrollment decision and acceptance or decline of any financial aid offer by May 1st. During this time, information from the admissions and financial aid applications is kept in an on-line computer file. This file is updated in June when the enrollment decisions of all admitted applicants are recorded. For those applicants who chose not to attend Stanford, no other information is recorded or changed. However, for those applicants who chose to enroll at Stanford, any changes in the financial aid package are entered by replacing the existing information. Subsequent
changes occurring during the summer are similarly recorded. At the end of August, a historical file containing all information pertaining to the freshman admissions and financial aid applications for that year is created from this on-line computer file, based on the last revision.

The historical files of admissions and financial aid data provide the primary source of data for this research. From the above discussion, it is clear that variables comprising the financial aid package in these files reflect the financial aid offer on April 1st for the non-enrolling applicants and the last revision to the original offer in August for the enrolling applicants. Ideally, any analysis of factors affecting the enrollment decision would use data approximating the information available to the applicant at the time of his decision. For this reason, we engaged in a special effort to collect the original offer data for the enrolling applicants from student files for the three academic years 1975-1976 through 1977-1978. Due to cost considerations and indications that the extent of change in financial awards occurred to a much lesser extent in the academic years 1973-1974 and 1974-1975, additional data were not collected for these two years.

Table 2 summarizes the available information on financial awards by academic year for three critical points in time: (1) first offer on April 1st, (2) last revision to offer on May 15th, and (3) final award in August. The second point in time reflects the closest approximation to the financial aid decision time of the applicant’s enrollment decision. Since these data are only available for enrolling applicants, the original offers in the last three academic years provide the closest approximation to this point in time. Of course, the first two academic years afford no choice in data to be used for the development of enrollment models. Because of the data available for analysis, a new constraint arises in analyzing factors affecting the enrollment decisions of applicants. Specifically, enrollment models should be resilient to the kinds of fluctuations that occur in the natural adjustments of financial awards, yet sensitive enough to detect any systematic effects.

Controlling for Inflation.

The effects of inflation have been substantial during the years 1973 through 1978. So as to permit comparability between financial variables across the five academic years in the study, all monetary values throughout this report are converted into 1972 constant dollars using the appropriate deflation index for each year. The text of Economic Indicators for the second session of the 95th Congress lists the GNP implicit price deflators as:
The deflators help to control for the effects of inflation during these years so that all monetary values are expressed in 1972 constant dollars. For example, $14,161 in 1977 is equivalent to $11,602 in 1974, once we control for the effects of inflation. Converting to 1972 constant dollars permits a uniform interpretation of the effects of financial variables.

3. PRELIMINARY ANALYSIS.

Preliminary analysis of the data provides a general overview of the variables and isolates key factors affecting the enrollment decision of applicants in the target population. The specific purpose of this analysis is to select from the large number of background and financial variables available those variables most related to the enrollment decision. This set of primary variables becomes the principal input to models assessing the combined and interactive effects of factors on the enrollment decision.

Several considerations guide the selection of a set of primary variables. First, those variables pertinent to specific policy questions should be included, even if their individual effects on the enrollment decision are small. This applies to such variables as parents' income and total aid. Second, variables that are strongly associated with the enrollment decision should be included. Third, variables which undergo significant changes in the computer file for the enrolling applicants should be excluded, since the data for these variables are not comparable for the enroll and not enroll groups. Finally, if several variables are highly collinear and all are measuring essentially the same quantity, only one variable from the group should be included.

Assessment of Changes in Financial Aid Awards.

Recall from the discussion of data collection procedures that financial aid awards are updated only for enrolling applicants as new information is received during May through August. It is important to assess the type and extent of changes that occur during these months.
in order to determine whether the data from the computer files for the academic years 1973-1974 and 1974-1975 are a reasonable approximation to the original offers the applicants receive in April. We used the additional data collected for the academic years 1975-1976 and 1976-1977 to assess these changes.

Changes in aid awards could be examined only for enrolling freshmen applicants who also applied for financial aid. This subpopulation consists of 794 and 772 applicants for the academic years 1975-1976 and 1976-1977, respectively. Variables related to the financial aid package are examined based on data from three points in time: (1) first offer on April 1st, (2) last revision to offer on May 15th, and (3) final award in August.

To provide a very general picture of the changes in aid awards, average amounts for each variable are given in Table 3. Notice that the average amounts of Stanford tuition waiver and outside awards are small. This is due to the fact that relatively few applicants receive these awards rather than that the size of the award is small. Less than 15 applicants from each year received a tuition waiver while 15 and 20 percent of the applicants for 1975-1976 and 1976-1977, respectively, received an outside award.

The results in Table 3 show that for most variables the average amounts are very similar across the three time periods, and, in fact, vary by less than 10 percent. The only exception is outside awards which shows a dramatic increase in average amount from April to August. This is consistent with the fact that outside awards are often received during the summer months. Also, the increase in outside awards

<table>
<thead>
<tr>
<th>Variable</th>
<th>April 1 Average</th>
<th>May 15 Average</th>
<th>August Average</th>
<th>April 1 Average</th>
<th>May 15 Average</th>
<th>August Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Aid</td>
<td>$2656</td>
<td>$2696</td>
<td>$2641</td>
<td>$2490</td>
<td>$2535</td>
<td>$2554</td>
</tr>
<tr>
<td>Scholarship</td>
<td>$2074</td>
<td>$2095</td>
<td>$2069</td>
<td>$1815</td>
<td>$1845</td>
<td>$1869</td>
</tr>
<tr>
<td>Self-help</td>
<td>$582</td>
<td>$601</td>
<td>$572</td>
<td>$655</td>
<td>$690</td>
<td>$685</td>
</tr>
<tr>
<td>Loan</td>
<td>$308</td>
<td>$318</td>
<td>$297</td>
<td>$369</td>
<td>$380</td>
<td>$382</td>
</tr>
<tr>
<td>Job</td>
<td>$274</td>
<td>$283</td>
<td>$275</td>
<td>$296</td>
<td>$310</td>
<td>$303</td>
</tr>
<tr>
<td>Combined Stanford Scholarship</td>
<td>$1532</td>
<td>$1542</td>
<td>$1490</td>
<td>$1199</td>
<td>$1227</td>
<td>$1215</td>
</tr>
<tr>
<td>St. Grant</td>
<td>$1505</td>
<td>$1516</td>
<td>$1477</td>
<td>$1164</td>
<td>$1192</td>
<td>$1190</td>
</tr>
<tr>
<td>St. Tuition Waiver $27</td>
<td>$27</td>
<td>$27</td>
<td>$13</td>
<td>$35</td>
<td>$35</td>
<td>$25</td>
</tr>
<tr>
<td>Combined Other Scholarships</td>
<td>$542</td>
<td>$553</td>
<td>$578</td>
<td>$617</td>
<td>$618</td>
<td>$654</td>
</tr>
<tr>
<td>Cal. State</td>
<td>$502</td>
<td>$461</td>
<td>$490</td>
<td>$573</td>
<td>$553</td>
<td>$542</td>
</tr>
<tr>
<td>Outside Award</td>
<td>$42</td>
<td>$42</td>
<td>$88</td>
<td>$44</td>
<td>$46</td>
<td>$111</td>
</tr>
<tr>
<td>Aid Quality</td>
<td>0.613</td>
<td>0.620</td>
<td>0.619</td>
<td>0.587</td>
<td>0.554</td>
<td>0.566</td>
</tr>
</tbody>
</table>
explains the increase in aid quality, since aid quality measures the percentage of the aid package that is scholarship.

It is interesting that the average amounts of self-help and Stanford grant tend to increase from April 1st to May 15th and then decrease from May 15th to August. The increase is explained in part by reassessments of financial need due to requests from applicants during the period from April 1st to May 15th. The decrease in self-help after May 15th largely results from applicants’ decisions to decline a loan or job originally offered. The decrease in Stanford grant is partially explained by the fact that an applicant’s Stanford grant is reduced on a dollar-for-dollar basis upon receipt of an outside award so that the amount of total aid continues to match financial need.

To provide a more detailed picture, we examined the exact changes in aid awards occurring from April to August. For each variable, changes were measured by tabulating the total number of cases in which the amount of the award differed between two time periods as well as computing the average amount by which the awards differed. These results are summarized for several key variables in Table 4. Notice that the stability of the average amounts of total aid, scholarship, and California State Scholarships across the three time periods in Table 3 appears to result from the tendency for the number and size of increases and decreases in these awards to offset one another.

The number of increases in self-help and Stanford grant exceed the number of decreases from April 1st to May 15th. However, from April 1st to August, there are more decreases than increases. This observation helps

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total Aid Increases</th>
<th>Decreases</th>
<th>Scholarship Increases</th>
<th>Decreases</th>
<th>Self-Help Increases</th>
<th>Decreases</th>
<th>Stanford Grant Increases</th>
<th>Decreases</th>
<th>Outside Aid Increases</th>
<th>Decreases</th>
<th>Cal. State Scholar. Increases</th>
<th>Decreases</th>
</tr>
</thead>
<tbody>
<tr>
<td>April 1 and August</td>
<td>365</td>
<td>89</td>
<td>28</td>
<td>18</td>
<td>32</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>24</td>
<td>9</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>April 1 and May 15</td>
<td>357</td>
<td>97</td>
<td>26</td>
<td>4</td>
<td>24</td>
<td>7</td>
<td>8</td>
<td>3</td>
<td>19</td>
<td>1</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>April 1 and August</td>
<td>450</td>
<td>170</td>
<td>32</td>
<td>10</td>
<td>34</td>
<td>5</td>
<td>10</td>
<td>4</td>
<td>20</td>
<td>1</td>
<td>12</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4: Profile of Changes in Financial Aid Awards
explain the rising and falling behavior of the average awards for these
two variables across the three time periods. Also, the results in
Table 4 for outside awards show a consistent increase in the number of
such awards from April to August.

The total number of changes in total aid from April 1st to August
provides a reasonable measure of the extent of changes between the
April 1st offers and the final awards recorded in the computer file in
August. This is due to the fact that total aid, as the sum of all
components in the financial aid package, reflects the composite effect
of changes occurring in component variables. Changes in this variable
occurred for 21 percent of the enrolling applicants in 1973-1974 and
32 percent in 1976-1977. The higher proportion of changes in total aid
for 1976-1977 results in part from the increased number of outside
awards received during that year. In view of these proportions, it
appears that approximately 70 to 80 percent of the data in the computer
files for enrolling applicants who also applied for financial aid in
1973-1974 and 1974-1975 represent the original offers the applicants
receive on April 1st. Since only the original offers are recorded for
the non enrolling applicants, this means that approximately 80 to 85
percent of the data in the computer files reflect original offers for
the entire target population.

Further analysis of the changes in financial aid awards attempted
to detect any systematic changes. Frequency distributions of the
variables were compared across the three time periods and failed to show
any systematic changes. Also, plots of each variable for the May 15th
and August revisions to the award versus the original offer on April 1st

showed that the changes in awards were largely the result of individual
considerations. An example of such a plot for the total aid variable is
given in Graph 5 and reflects a random scattering of changes off the
diagonal where the awards agree.

Several conclusions are suggested by this analysis. First, since
approximately 80 to 85 percent of the data in the computer files reflect
the original offers the applicants receive, it is reasonable to use the
file data for academic years 1973-1974 and 1974-1975 as an approximation
to the original aid offers. Second, since substantial changes in outside
awards occur during the summer months, the effects of outside awards on
the enrollment decision cannot be measured on the basis of the file data.
Of course, the offer data for academic years 1973-1974 and 1976-1977 can
be used to assess any effects of this variable, since the data for out-
side awards on April 1st are comparable for both the enrolling and not
enrolling applicants. Finally, it appears that changes in financial aid
awards result largely from individual considerations rather than any
systematic effects.

Methods for Isolating Important Predictor Variables.

In the selection of a set of primary variables from those available,
it is important to include those variables that are strongly associated
with the enrollment decision. To isolate key predictor variables, we
use three methods of analysis to measure the individual effects of each
variable on the enrollment decision.
1. Contingency Tables

First, two-way contingency tables were developed for each variable by enrollment group. This analysis consists of determining a number of categories for the values of each variable and tabulating the numbers of enrolling and non-enrolling applicants in each category. The resultant proportions of applicants in each category can then be studied to determine if there are any differences between the enrollment groups. Examples of such tables are given by Table 6-A for the academic rating variable and Table 6-B for the parents' income variable. The chi-square statistic and its associated significance level measure departures from the hypothesis that the variables of interest and the enrollment decision are independent. In general, large chi-square statistics and small significance levels indicate that the variable is associated with the enrollment decision.

2. Two-Sample t-Tests

The second method used for measuring the effects of a variable on the enrollment decision is the two-sample t-test. This test assumes that the values of the variable for each enrollment group are normally distributed about an average value for the group. The test then compares the two means of each group controlling for the estimated variance of the difference in means. It should be noted that two difficulties arise in the application of this test to the variables available. First, since the target population consists of approximately 1200 applicants for each year, small differences between the group means are likely to be accentuated by the large sample size. Consequently, some variables may result in a statistically significant difference between the enroll and not enroll groups, whereas no real difference is perceived. Second, the t-test is sensitive to departures
<table>
<thead>
<tr>
<th>Year</th>
<th>0.0000</th>
<th>0.0000</th>
<th>0.0000</th>
<th>0.0000</th>
<th>0.0000</th>
<th>0.0000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9.99</td>
<td>7.00</td>
<td>10.99</td>
<td>8.00</td>
<td>2.99</td>
<td>9.99</td>
</tr>
<tr>
<td></td>
<td>(900')</td>
<td>(990')</td>
<td>(600')</td>
<td>(100')</td>
<td>(90')</td>
<td>(90')</td>
</tr>
<tr>
<td></td>
<td>10.00</td>
<td>10.00</td>
<td>10.00</td>
<td>10.00</td>
<td>10.00</td>
<td>10.00</td>
</tr>
<tr>
<td></td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
</tr>
<tr>
<td></td>
<td>10.00</td>
<td>10.00</td>
<td>10.00</td>
<td>10.00</td>
<td>10.00</td>
<td>10.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note: (1) Total refers to the percentage of each year.

(2) The numbers in parenthesis give the percentage of total cases for each year.

The two-way contingency table of cases, income, habitat, and employment group.

Table 6-1
from the assumption that the values of the variable are symmetrical about the group means. Certain financial variables, such as parents' income and educational aid, are not symmetrical about each group mean and are instead skewed towards lower values.

3. Frequency Distributions

To examine differences between the enrolling and not enrolling groups more carefully for the continuous variables, frequency distributions of each variable by enrollment group are examined. A method for displaying important differences between the two groups is to summarize the frequency distributions in the form of a schematic or box plot. An example of the schematic plot for total aid appears in Graph 7. The median total aid amount is given by the solid line drawn in the center of the box whereas the average amount of total aid appears as a '+' symbol. Fifty percent of the data lies within the box while approximately 95 percent of the data lies between the two outside fences. Values beyond the fences are plotted individually and suggest outliers. These plots highlight differences between the distributions of the variable for the enroll and not enroll groups as well as provide an assessment of the symmetry of the values about the median.

As a result of these three methods of analysis, the variables fall into three general categories according to their degree of association with the enrollment decision. Variables are regarded as highly associated if the measured effects on the enrollment decision are substantial and consistent across all four years of data. Variables are classified as having some association if the effects on the enrollment decision are significant across two or more years of data. Finally, variables showing no apparent differences between the two enrollment groups are classified as having little or no association. The results of this classification appear in Table 8.
Table 8
Grouping of Admission and Financial Variables by Degree of Association with Enrollment Decision

<table>
<thead>
<tr>
<th>Degree of Association</th>
<th>Admission Variables</th>
<th>Financial Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Association</td>
<td>Academic Rating</td>
<td>Total Aid</td>
</tr>
<tr>
<td></td>
<td>Verbal SAT Score</td>
<td>Student Need</td>
</tr>
<tr>
<td></td>
<td>Math SAT Score</td>
<td>Scholarship</td>
</tr>
<tr>
<td></td>
<td>Other Scholarships</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aid Quality</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total Resources</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parents’ Contribution</td>
<td></td>
</tr>
<tr>
<td>Some Association</td>
<td>Sex</td>
<td>Parents’ Income</td>
</tr>
<tr>
<td></td>
<td>Minority Status</td>
<td>Parents’ Offer</td>
</tr>
<tr>
<td></td>
<td>Stanford Affiliation</td>
<td>Student Budget</td>
</tr>
<tr>
<td></td>
<td>Region of Residence</td>
<td>Self-help</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stanford Loan</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stanford Job</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cal. State Scholarship</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Outside Awards</td>
</tr>
<tr>
<td>Little or No Association</td>
<td>Nonacademic Rating</td>
<td>Stanford Grant</td>
</tr>
<tr>
<td></td>
<td>High School Type</td>
<td>Stanford Tuition Waiver</td>
</tr>
<tr>
<td></td>
<td>Number of Parents Working</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of Children in Family</td>
<td></td>
</tr>
</tbody>
</table>

Individual Effects of Key Variables on Enrollment Decision

It is of interest to note the effects of some key variables that are strongly associated with the enrollment decision. The effects of variables pertaining to academic merit agree with the results reported in previous studies. There is a tendency for applicants in the highest academically rated group not to enroll at Stanford. The contingency table given in Table 6-A reveals that the proportion of applicants in this group enrolling at Stanford is much lower than the overall enrollment rate. Over 90 percent of the applicants in the target population score 500 or above on the Verbal SAT test and over 95 percent score 500 or above on the Math SAT test. Yet, the proportion of applicants scoring 700 or above on either test that enroll at Stanford is approximately 50 percent. This also confirms a tendency for the highest academically rated applicants not to enroll at Stanford.

The Stanford affiliation variable shows that applicants who have some affiliation with Stanford are more likely to enroll at Stanford than those who do not. Analysis of the region of residence reveals a slightly greater tendency for applicants from California to enroll and a slightly reduced tendency for applicants from the northeastern part of the U.S. to enroll. There is also a slightly greater tendency for minority applicants to enroll at Stanford. Approximately 70 percent of this group enrolled during the four academic years studied compared with the overall enrollment rate of approximately 64 percent.

The effects of total aid, scholarship, and outside awards show that large amounts of these awards are associated with high enrollment rates. The effects of student need are consistent with these awards. Applicants having the largest demonstrated need are the
most likely to enroll at Stanford. Also, there is almost a linear trend between aid quality and enrollment rates. Applicants receiving a large proportion of the aid package in scholarship are more likely to enroll than those who receive a small proportion or no scholarship. The enrollment rates are approximately 71 percent for those applicants receiving 80 percent or more scholarship in the aid package whereas for those applicants receiving no scholarship the enrollment rates range from 54 to 58 percent.

The analysis of parents' income shows a somewhat greater tendency for applicants from low income groups to enroll. From Table 6-8, approximately 70 percent of all applicants whose parents' income is $15,000 or less enroll at Stanford whereas approximately 60 percent of applicants whose parents' income is $25,000 or more enroll. No clear evidence of a middle income problem can be obtained since the enrollment rates for the middle income group vary substantially across the four years.

It is interesting that the expected parents' contribution for 40 percent of all applicants in the target population is $3000 or more. However, only 3 percent of the applicants have a parents' offer, exceeding $5000. There is a very large discrepancy between expected parents' contribution, as computed by the Financial Aid's Office, and the amount the parents feel they should contribute towards freshmen year costs. Furthermore, the amount that parents offer is relatively constant across all income groups, ranging between $0 and $5,000. The effects of parents' contribution on the enrollment rates are that applicants are more likely to enroll when the amount is small. The effects of parents' offer also show a somewhat greater likelihood of enrolling when the amount is small. However, the association between parents' offer and the enrollment decision is not nearly as strong as for parents' contribution.

We also investigated whether the difference between the parents' contribution and the parents' offer might be more strongly associated with the enrollment decision. The results revealed that this was not the case and that parents' contribution is a more key factor with regard to the enrollment decision. We observed that large differences between parents' contribution and parents' offer tend to coincide with large values of parents' contribution. Furthermore, since the amount parents offer is relatively constant across a broad range of differing family financial circumstances, it appears that differences between the expected contribution and the amount offered are largely explained by the parents' contribution variable. It should also be noted that the effects of total resources on the enrollment decision are the same as for parents' contribution. This is consistent with the observation that parents' contribution comprises such a large component of the total resources variable.

Elimination of Other Scholarship Variables.

The effects of outside awards and California State Scholarships must be interpreted carefully. The method of notifying the Financial Aid's Office about these awards tends to underestimate the number of such awards received by non-enrolling applicants. Specifically, during the academic years 1973-1974 through 1975-1976, the California State Scholarship Commission reported the amounts of awards in April to the Financial Aid's
Office for those applicants who listed Stanford as their first choice of colleges to attend. The data for those years show that approximately 230 enrolling applicants and 70 not enrolling applicants received these awards, resulting in an enrollment rate of 77 percent for the group of recipients. Beginning in 1976-1977, the Financial Aides Office introduced a new policy and assumed a California State Scholarship in the financial aid offer for applicants satisfying the eligibility requirements. The data for this year reveal that 235 enrolling applicants and 105 not enrolling applicants received this award, resulting in an enrollment rate of 68 percent for the group of recipients. Due to the notification procedure regarding California State Scholarships, it appears that the number of such awards for not enrolling applicants is underestimated in the data for academic years 1973-1974 through 1975-1976.

National Merit Scholarships comprise a large proportion of outside awards. The National Merit Scholarship Agency uses a similar procedure to the California State Scholarship Commission to notify the Financial Aides Office of these awards to those applicants who listed Stanford as their first choice of colleges to attend. From the offer data, the number of enrolling applicants receiving an outside award was 51 and 54 for academic years 1975-1976 and 1976-1977, respectively. However, the number of not enrolling applicants receiving an outside award was only 7 and 20 percent for academic years 1975-1976 and 1976-1977, respectively. It appears that the number of outside awards for not enrolling applicants is also underestimated in the data.

Due to the notification procedures regarding California State Scholarships and National Merit Scholarships, the effects of these awards on the enrollment decision are probably exaggerated by the data. Since Other Scholarships refer to the sum of California State Scholarships and outside awards, the effects of this variable on the enrollment decision are also exaggerated. It appears that the receipt of these other scholarships does correspond to a slightly greater likelihood of enrolling at Stanford. However, the enrollment rate for the group of recipients is probably much closer to 70 percent than the rate of 80 percent given by the data. Since the data for other scholarships are not comparable for both the enrolling and not enrolling applicants, these variables are excluded from the set of primary variables.

Assessment of Collinearity Among Predictor Variables.

Another criterion for the selection of a set of primary variables related to the enrollment decision is to exclude variables that are surrogates for another variable already in the set. To detect highly collinear variables among those highly associated with the enrollment decision, we plotted pairs of variables and looked for linear associations between them. Also, research into the financial aids process at Stanford suggested several financial variables that are causally related in a linear fashion.

Two variables that have a clear causal relationship are total aid and student need. This results from the fact that the Financial Aides Office awards total aid amounts to meet student need. Indeed, the plots of these two variables show a strong linear relationship. Since total aid is the more important variable for policy issues, student need is excluded from the set of primary variables.
Another pair of variables that are highly collinear are parents' contribution and total resources. Recall that total resources is the sum of parents' contribution, student savings, summer earnings, and other resources available to the applicant. The other components contribute relatively little to the total resources amount and parents' contribution accounts for approximately 80 percent of the total resources amount. Since parents' contribution addresses key policy questions, total resources is excluded from the set of primary variables.

Verbal and Math SAT scores are positively associated although they are clearly not collinear. There was no apparent reason to study separately the effects of these two variables on the enrollment decision. Therefore, the scores were added together and the sum of SAT scores is included in the set of primary variables.

No other pairs of variables exhibit strong linear associations. However, it is interesting to note some associations between other variables. For example, parents' contribution is closely associated with parents' income in that large amounts of parents' contribution tend to correspond to high incomes. However, there does not appear to be a clear association between parents' contribution and assets or between parents' income and assets.

Large amounts of scholarship are associated with large amounts of total aid. This affects the relationship between aid quality and total aid. The plot of aid quality versus total aid is given in Graph 9 for the academic year 1975-1976. The plot demonstrates that large amounts of total aid consist of a high proportion in scholarship. The curved lines in the plot result from a financial aid policy requiring different minimum amounts of self-help to be part of the aid package. The minimum amounts vary across...
the seven aid categories of applicants developed by the Financial Aids Office and generally range from $300 to $1,700 in the years studied.

Final Set of Primary Variables.

As a result of the preliminary data analysis, the final set of primary variables consists of two admissions and six financial variables. These variables are: Academic Rating, Sum of SAT Scores, Parents' Income, Parents' Contribution, Total Aid, Aid Quality, Scholarship, and Self-help. Most of these variables are highly associated with the enrollment decision while others are included to address key policy questions.

In addition to the set of primary variables, six other variables comprise a set of secondary variables. These include Sex, Minority Status, Region of Residence, Stanford Affiliation, Stanford Loan, and Stanford Job. All of these variables indicate some association with the enrollment decision although the degree of association is much less than for those variables in the primary set. This secondary set of variables is useful for studying the combined effects of the primary variables in greater detail. For example, analyzing the effects of Stanford loan and job awards on the enrollment decision provides additional information concerning the components of self-help.

4. METHODOLOGY

Analyzing the individual effects of factors related to the enrollment decision is useful for suggesting key factors. However, this approach is insufficient for determining the effects of a factor, such as total aid, while controlling for differences in applicants' academic achievements or family financial circumstances. It also fails to determine the combined effects of several factors on the enrollment decision.

To clarify these ideas, consider the results from the preliminary analysis of the data pertaining to the individual effects of total aid and aid quality on the enrollment decision. The results show that large amounts of total aid are associated with an increased likelihood of enrolling at Stanford. Furthermore, applicants who receive a large proportion of the aid package in scholarship are more likely to enroll than those who receive a small proportion or no scholarship. However, the plot of aid quality versus total aid reveals that high proportions of scholarship in the aid package are typically associated with large amounts of total aid. A natural question to ask is whether the effects of aid quality on the enrollment decision are simply duplicating the effects of total aid. On the other hand, once we control for the fact that large amounts of total aid are associated with higher yield rates, there may be an additional aid quality effect. By analyzing both factors simultaneously, we can detect any additional effect of aid quality on the enrollment decision in the presence of total aid.

For this reason, it is desirable to develop models whereby one can study the combined and interactive effects of primary factors related to the enrollment decision. To develop these models, three different statistical methods are used: (1) discriminant analysis, (2) logistic regression,
and (3) log-linear probability models. Each of these methods is designed to analyze the effects of a number of factors on a binary, dependent variable, which in this case is the enrollment decision. The first two methods determine the main and combined effects of a large number of factors on the dependent variable. The third method exploits structural relations among the factors and is suitable for analyzing the interactive effects of a smaller number of factors.

This section provides a general description of each method in the context of the present study. In particular, the specific model proposed by each method is given and underlying assumptions are highlighted to distinguish the three methods. To improve readability, technical details are omitted and appear in Appendix 1.

### Discriminant Analysis

Discriminant analysis is a standard statistical method used for classifying an individual into one of two mutually exclusive groups on the basis of an observed number of background variables. Specifically, one assumes that an individual with k background characteristics, summarized by the random vector \( \mathbf{X} = (x_1, x_2, \ldots, x_k) \), belongs to either Group 0 or Group 1 on the basis of the observed values of \( \mathbf{X} \). By assuming that the multivariate frequency distribution of the random vector \( \mathbf{X} \) is different for Group 0 and Group 1, it is possible to classify an individual into one of these two groups by determining whether the observed values of \( \mathbf{X} \) are more characteristic of the distribution for Group 0 or 1.

In this study, Group 0 refers to the not enrolling applicants and Group 1 refers to the enrolling applicants. The background characteristics include such information as Sum of SAT Scores, Total Aid, Parents' Income, and Aid Quality. It is helpful to let the variable \( Y \) represent the enrollment decision of an applicant. Specifically, \( Y = 1 \) if the applicant enrolls at Stanford and \( Y = 0 \) otherwise. Thus, the vector \((Y, \mathbf{X})\) summarizes the enrollment decision and background characteristics of each applicant.

The discriminant analysis procedure first assumes a certain unconditional probability of enrolling at Stanford, irrespective of the background characteristics of the applicant. That is, we assume \( P(Y=0) = \pi_0 \) and \( P(Y=1) = \pi_1 \), where \( \pi_0 + \pi_1 = 1 \). The key assumption in discriminant analysis is that the background variables \( \mathbf{X} \) have a multivariate normal distribution with a different mean vector but the same covariance matrix depending on whether \( \mathbf{X} \) characterizes an enrolling applicant or a not enrolling applicant. Specifically, the conditional distribution of \( \mathbf{X} \) given \( Y \) is

\[
\begin{align*}
\mathbf{X} | Y = 0 & \sim \mathcal{N}_k(\mu_0, \Sigma) \\
\mathbf{X} | Y = 1 & \sim \mathcal{N}_k(\mu_1, \Sigma),
\end{align*}
\]

where \( \mu_i \) denotes the \( k \times 1 \) mean vector for Group \( i \) (\( i = 0, 1 \)) and \( \Sigma \) denotes the \( k \times k \) covariance matrix.

Notice that the key assumption in discriminant analysis provides a statement about the distributional values of \( \mathbf{X} \) given that we know whether an applicant enrolls or does not enroll at Stanford. However, we can easily translate this assumption into a statement concerning the probability that an applicant enrolls at Stanford, given that we know the applicant's background characteristics \( \mathbf{X} \), by using Bayes Theorem. The result of this translation specifies that the conditional distribution of \( Y \) given \( \mathbf{X} \) is

---

37


\[ P(Y=0|x) = \frac{e^{-D(x)}}{1 + e^{-D(x)}} \]

and

\[ P(Y=1|x) = \frac{1}{1 + e^{-D(x)}} \]

where

\[ D(x) = \alpha + \beta' x, \]

\[ \alpha = \log \frac{\pi_1}{\pi_0} - \frac{1}{2} (\mu_1 - \mu_0)' \Sigma^{-1} (\mu_1 - \mu_0), \]

and

\[ \beta = (\Sigma^{-1})^{1/2} \mu_1 - \mu_0. \]

The function \( D(x) = \alpha + \beta' x \) is commonly referred to as the linear discriminant function. The discriminant analysis model specifies that the enrollment decision of an applicant is related to the background characteristics \( x \) only by the function \( D(x) \).

The explicit relationship between the function \( D(x) \) and the probability of enrolling at Stanford is given by the log odds of enrolling.

Let \( p(x) \) denote the conditional probability of enrolling at Stanford for an applicant having background characteristics \( x \). That is, \( p(x) = P(Y=1|x) \) and \( 1 - p(x) = P(Y=0|x) \). Then, some simple algebra shows that

\[ \log \frac{p(x)}{1 - p(x)} = D(x). \]

Thus, the function \( D(x) \) represents the log odds of enrolling at Stanford for an applicant having background characteristics \( x \). Large values of \( D(x) \) increase the log odds of enrolling and consequently increase the probability of enrolling at Stanford. Small values of \( D(x) \) have the opposite effect. In this way, we can use the function \( D(x) \) to classify an applicant as enrolling or not enrolling at Stanford. Specifically, we classify an applicant as enrolling if \( D(x) > 0 \) and classify the applicant as not enrolling if \( D(x) < 0 \). This method of classification minimizes the expected probability of misclassifying an applicant.

To apply the discriminant analysis procedure to the data available, it is necessary to obtain estimates of the unknown parameters \( \pi_0, \pi_1, \mu_0, \mu_1, \) and \( \Sigma \). We use the method of maximum likelihood estimation to obtain estimates of the unknown parameters and consequently the discriminant function coefficients. Also, results from asymptotic statistical theory provide estimates for the variances and correlations of the estimated coefficients. These estimates are explicit functions of the data and appear in Appendix 1.

The estimated coefficients \( \hat{\alpha} \) and \( \hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k) \) are used to derive the estimated discriminant function \( \hat{D}(x) = \hat{\alpha} + \hat{\beta}' x \). This function is equivalent to the estimated log odds of enrolling at Stanford for an applicant having background characteristics \( x \). As a result, the estimated coefficients \( \hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k) \) measure the effects of the \( k \) background variables on the enrollment decision. For example, a positive coefficient \( \hat{\beta}_1 \) indicates that the \( i \)th variable has a positive effect on the estimated log odds of enrolling and consequently on the estimated probability of enrolling at Stanford. In order to determine the size of the effect, it is helpful to compute the estimated coefficients in terms of 'standardized' background variables. To standardize a background variable, we simply subtract the sample mean and divide by the sample standard deviation. The estimated coefficients based on standardized variables measure the relative importance of each variable in the model. In particular, a large value of
a standardized coefficient implies that the associated variable has a large
effect on the enrollment decision relative to the other variables in the
model. The estimated variances and correlations of the estimated coeffi-
cients help to determine the statistical significance of the measured
effects and whether or not the effects of the background variables are
linearly related.

Once a discriminant function has been estimated for a set of $k$ back-
ground variables, it is important to determine whether any discrimination
between the enrolling and not enrolling applicants has been achieved. For
this purpose, we can use several methods for assessing the fit of the model
to the data. The first method is based on the two-sample version of
Hotelling's $T^2$ statistic, which appears in Appendix 1. This statistic
measures the distance between the sample mean vectors of the background
variables for the enrolling and not enrolling groups. Large values of the
statistic indicate that the distance between the two mean vectors $\mu_0$ and
$\mu_1$ is large and consequently there is good discriminating power in the
set of $k$ background variables. Small values of the statistic indicate
that the values of the background variables overlap for the two groups and
there is poor discriminating power. This statistic is primarily useful
for comparing the relative discriminating power of models based on different
sets of background variables.

A second method for assessing the fit of a model to the data classifies
each applicant as enrolling or not enrolling on the basis of the estimated
discriminant function. We can then determine the percentage of applicants
correctly classified by comparing the predicted classification with the
actual enrollment decision. Naturally, a high proportion of applicants
correctly classified reflects a good fit of the model to the data. Since
the same data are being used to assess the procedure as well as define it,
this method is not strictly appropriate for testing discriminatory power
in the background variables. However, it is useful for isolating sets of
background variables having very little discriminating power.

A final method for assessing the fit of a model to the data involves
computing the estimated probability of enrolling for each applicant from
the estimated discriminant function. We then group the frequency distribu-
tion of the computed $\hat{p}(x)$s into deciles. The observed numbers of
enrolling applicants in each decile can be compared with the expected
number of enrolling applicants, determined by summing the estimated
probabilities for all applicants in the decile. Striking differences
between the observed and expected numbers of applicants suggest a poor fit.

Logistic Regression

A common statistical method for expressing the relationship between a
dependent variable and a set of background variables is the regression model.
In this study, the dependent variable refers to the applicant's enrollment
decision. Since the enrollment decision has only two possible values
(0 for not enrolling and 1 for enrolling), ordinary least squares regression
is inappropriate for two reasons. First, the variance of the dependent
variable depends upon each observation. This violates the assumption of
homoskedasticity used in ordinary least squares regression. Second, the
predicted values resulting from ordinary least squares regression can have
any numerical values, which conflicts with the fact that the values of the
dependent variable are restricted to lie between 0 and 1.
Logistic regression bypasses these problems by assuming a relationship between the probability that an applicant enrolls at Stanford and the set of background variables. To state this relationship precisely, let \( Y \) denote the applicant's enrollment decision and let \( \mathbf{x} = (x_1, x_2, \ldots, x_k) \) summarize the \( k \) background characteristics of interest. The key assumption in logistic regression is the following:

\[
P(Y = 0 | \mathbf{x}) = \frac{e^{-(\alpha + \mathbf{\beta}^T \mathbf{x})}}{1 + e^{-(\alpha + \mathbf{\beta}^T \mathbf{x})}}
\]

and

\[
P(Y = 1 | \mathbf{x}) = \frac{1}{1 + e^{-(\alpha + \mathbf{\beta}^T \mathbf{x})}}
\]

where \( \alpha \) and \( \mathbf{\beta} = (\beta_1, \beta_2, \ldots, \beta_k) \) are unknown parameters to be estimated from the data. This assumption states that the conditional probability of enrolling at Stanford for an applicant having background characteristics \( \mathbf{x} \) depends only on the linear function \( \mathbf{R}(\mathbf{x}) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \).

Furthermore, the relationship between the conditional probability of enrolling and \( \mathbf{R}(\mathbf{x}) \) is described by a smooth function, referred to as the logistic function, so that large values of \( \mathbf{R}(\mathbf{x}) \) are associated with higher probabilities. Graph 10 illustrates the logistic relationship between the conditional probability of enrolling and the linear function \( \mathbf{R}(\mathbf{x}) \). Notice that the curve has an 'S' shape and is constrained to lie between 0 and 1.

To apply this model to the data available, it is necessary to estimate the unknown parameters \( \alpha \) and \( \mathbf{\beta} \). The method of maximum likelihood is used to determine the estimates \( \hat{\alpha} \) and \( \hat{\mathbf{\beta}} \) that maximize the conditional likelihood function. Unlike discriminant analysis where the unconditional likelihood function is maximized, the estimates \( \hat{\alpha} \) and \( \hat{\mathbf{\beta}} \) cannot be
written in closed form. Instead, the maximum likelihood equations must be solved iteratively, starting with an initial approximation and obtaining successive estimates. From asymptotic statistical theory, estimates for the variances and correlations of the estimated coefficients are obtained. The technical details of these methods are discussed in Appendix 1.

Notice that the maximum likelihood estimates \( \hat{\beta} \) and \( \hat{\sigma} \) also provide an estimated discriminant function, namely \( \hat{g}(x) = \hat{\alpha} + \hat{\beta}^T x \). This function can be used to classify applicants as enrolling or not enrolling in the same way as for the discriminant analysis procedure. Also, the estimated coefficients \( \hat{\beta} \) and \( \hat{\sigma} \) provide a similar interpretation of the effects of the background variables. Methods for assessing the fit of the model in discriminant analysis are similarly appropriate for logistic regression.

The primary differences between logistic regression and discriminant analysis are that logistic regression is based on a conditional procedure, and does not make any assumptions about the distribution of the background variables \( x \). Discriminant analysis begins with the assumptions that there is an unconditional probability of enrolling and that the background variables have a multivariate normal distribution. Then, the conditional probability of enrolling given the background characteristics is derived as a logistic function. In contrast, logistic regression only assumes that the conditional probability of enrolling given the background characteristics can be described by a logistic function. In this way, logistic regression imposes fewer assumptions and is a more general model, including discriminant analysis as a special case. However, in very large samples, the two methods coincide, since the multivariate normal distribution of the background variables is likely to be satisfied.

We used both discriminant analysis and logistic regression in this study for two reasons. First, since discriminant analysis is computationally simpler to use, this procedure permits investigation of a large number of models based on different sets of background characteristics. Second, logistic regression is used only to investigate key models that are particularly important for answering policy questions. For these models, we did not want to assume a normal distribution for discrete variables, such as academic rating, or highly skewed variables, such as self-help.

**Log-Linear Models**

Multidimensional contingency tables are a useful tool for analyzing relationships between categorical variables. The cells of the contingency table result from cross-classification of the variables and describe mutually exclusive combinations of attributes. The total number of individuals in the data having a specific combination of attributes can then be tabulated and entered as a count in the corresponding cell. In this way, the contingency table describes a multivariate frequency distribution of the observed values of the categorical variables.

It is helpful to illustrate our use of log-linear models in the context of the present study. To answer key policy questions, it is desirable to understand the interactive effects of academic rating, parents' income, total aid, and aid quality on the enrollment decision. For each academic year, we developed a five-way contingency table based on the enrollment decision and these four background variables. The values for the background variables were grouped into a small number of categories that are defined in Table II. This table also assigns a character label to each variable and
Table 11

Definition of Variable Categories for Five-Way Contingency Table

<table>
<thead>
<tr>
<th>Label</th>
<th>Variable</th>
<th>Total No.</th>
<th>Definition</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>Total Aid</td>
<td>4</td>
<td>Denied Aid</td>
<td>1 = 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Under $2,500</td>
<td>1 = 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$2,500 - $4,000</td>
<td>1 = 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Over $4,000</td>
<td>1 = 4</td>
</tr>
<tr>
<td>I</td>
<td>Parents' Income</td>
<td>3</td>
<td>Under $15,000</td>
<td>J = 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$15,001 - $25,000</td>
<td>J = 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Over $25,000</td>
<td>J = 3</td>
</tr>
<tr>
<td>R</td>
<td>Academic Rating</td>
<td>2</td>
<td>Low (3, 4, or 5)</td>
<td>k = 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>High (1 or 2)</td>
<td>k = 2</td>
</tr>
<tr>
<td>Q</td>
<td>Aid Quality</td>
<td>4</td>
<td>No Scholarship</td>
<td>1 = 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Less than 50% Scholarship</td>
<td>1 = 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50 - 80% Scholarship</td>
<td>1 = 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Over 80% Scholarship</td>
<td>1 = 4</td>
</tr>
<tr>
<td>E</td>
<td>Enrollment Decision</td>
<td>2</td>
<td>Enroll</td>
<td>m = 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Not Enroll</td>
<td>m = 2</td>
</tr>
</tbody>
</table>

an index for the categories that will make later references more convenient. The resultant contingency table has $4 \times 3 \times 2 \times 4 \times 2 = 192$ cells. The observed contingency table for academic year 1975-1976 is given by Table 12.

The main concern of this research is the influence of various background variables on the applicant's enrollment decision. For the five-way contingency table used in the study, it is natural to regard the information provided by the four background variables as fixed or given and the enrollment decision as a dependent variable. For this purpose, we can collapse the five-way contingency table into a table having only four dimensions by replacing the numbers of enrolling and not enrolling applicants in each cell with the yield rate. The yield rate refers to the proportion of applicants enrolling in each cell of the collapsed table. For example, the first cell of Table 12 shows that the number of enrolling applicants is 3 and the number of not enrolling applicants is 2. This results in a yield rate of $3/(3+2) = 0.60$ for that cell. In this way, the yield rates describe the conditional probabilities of enrolling at Stanford for applicants having different combinations of background attributes.

Conditional log-linear models assume that the observed yield rates result from specific effects of the four background variables on the enrollment decision. Specifically, the models assume that there exist theoretical conditional probabilities of enrolling determined by the main and interactive effects of the four background variables. For example, suppose that we hypothesize that the amount of total aid an applicant receives primarily explains his enrollment decision. The conditional log-linear model formulated from this hypothesis is
\[
\log \frac{p(\omega)}{1 - p(\omega)} = L^E + \sum_{i=1}^{4} t_i^{TE}(\omega) t_i^{TE}
\]

where \( \omega = (\omega_1, \omega_2, \omega_3, \omega_4) \) is an index variable referring to a particular cell in the Total Aid x Income x Academic Rating x Aid Quality table,

\( p(\omega) \) is the theoretical yield rate for the group of applicants in cell \( \omega \),

\( L^E \) is the 'mean effect' due to the overall yield rate for the table,

\( t_i^{TE}(\omega) \) is an 'indicator variable' which equals 1 if \( \omega_i = 1 \), and 0 otherwise, and

\( t_i^{TE} \) is a parameter to be estimated from the data representing the effect of the \( i \)th total aid category on the enrollment decision (\( i = 1, 2, 3, \)).

This specific model relates the log odds probability of enrolling given various background attributes to \( \tau \) parameters describing the main effects of the total aid variable on the enrollment decision. Consequently, the log-linear model decomposes the log odds of enrolling into specific additive effects due to the background variables in the same way as in the analysis of variance.

It should be noted that although the contingency table has four total aid categories, we describe the effect of total aid on the enrollment decision using only three \( \tau \) parameters. This is due to the normalizing constraint that the \( \tau \) parameter representing the last category of a variable is always set equal to zero.

Log-linear models of the form given above are useful for several reasons. Since the \( \tau \) parameters represent the main and interactive effects of the background variables on the enrollment decision, associations between the enrollment decision and the background variables are easily examined.
Second, there is great flexibility in fitting these models to the observed frequencies in the contingency table. By increasing the number of \( r \) parameters and incorporating more effects, the model can be chosen to fit the observed frequencies as closely as desired. In fact, by incorporating all possible associations between the background variables and the enrollment decision, we can fit the observed table exactly. Finally, log-linear models analyze the relationships between the enrollment decision and the background variables for groups of applicants rather than individual applicants. As a result, the combined effects of the parameters can be easily assessed and interpreted by referring to the estimated yield rates in the table.

It should be noted that there are several different ways to parameterize log-linear models and derive estimates of the parameters. In Chapter 10 of their book, Bishop, Feinberg, and Holland (1975) give an overview of the different approaches currently available. For this study, we used the information theory approach developed by Kullback (1978). This approach provides a systematic method for generating models and estimating the parameters. An overview of how to estimate the \( r \) parameters and their associated variances and correlations using the information theory approach is given in Appendix 1.

In Table 12, notice that some of the cells in the five-way contingency table have 'X' recorded as the observed count. These cells are special and reflect what are commonly referred to as structural zeroes in the data. Structural zero cells are necessarily empty due to the associations between the attribute variables. For example, an applicant that receives no total aid is assigned an aid quality value of 3. Therefore, the cells characterizing applicants who are denied aid and have more than 80 percent of the aid package in scholarship are necessarily empty. The method for treating structural zeroes in log-linear models is to eliminate \( r \) parameters that require estimation from structural zero cells. Since the collapsed four-way table has 96 total cells with 46 structural zero cells, the total number of parameters that can be used to develop conditional enrollment models is 50.

Once a conditional log-linear model has been developed and the \( r \) parameters estimated from the observed table, it is important to assess the fit of the model. We used three statistics to measure the fit of the model. Let \( x(a) \) denote the observed count and \( \hat{x}(a) \) denote the estimated count in the \( a \) cell of the five-way contingency table. The first statistic for assessing the fit is the information statistic given by

\[
I(x; \mathbf{x}) = \sum_a x(a) \log \frac{x(a)}{\hat{x}(a)}.
\]

The asymptotic distribution of \( -2I(x; \mathbf{x}) \) is a chi-square distribution. Consequently, we can test the hypothesis that the model fits the observed table.

However, the information statistic is more useful for comparing different models and determining their relative fits to the data. The simplest conditional log-linear model for explaining the observed yield rates assumes that the probability of enrolling is the same across all cells. This model is represented by

\[
\log \frac{p(a)}{1-p(a)} = \beta^E.
\]

Let \( I(x; \mathbf{E}) \) denote the information statistic associated with this overall yield model. We can use this as a base model and compare the information
statistics for other models with \( I(x|x^e) \). In this way, it is possible to measure how much additional information is obtained by introducing effects pertaining to the influence of the four background variables on the enrollment decision. Naturally, the more parameters a model uses, the more information can be explained and the smaller the value of the information statistic.

The following statistic is useful for comparing the information statistic from a model to \( I(x|x^e) \) since it adjusts for the number of parameters in the model:

\[
C^* = 1 - \frac{-2 \log I(x|x^e)}{(n-1) I(x|x^e)}
\]

where \( n \) refers to the total number of parameters in the model. This statistic is called the corrected proportion of explained information and ranges between 0 and 1. The interpretation of this statistic is that high values of \( C^* \) indicate a good fit of the model to the observed table whereas small values indicate a poor fit.

The second measure for assessing the fit of a model is the Pearson chi-square statistic given by

\[
\chi^2 = \sum_{\omega} \frac{(x^*(\omega) - x(\omega))^2}{x^*(\omega)}
\]

This statistic has an asymptotic chi-square distribution and is also asymptotically equivalent to \(-2 I(x|x^e)\). Small values of the chi-square statistic indicate a good fit to the data.

The final measure evaluates the sum of absolute deviations between the estimated and observed cell counts. This statistic is given by

\[
M = \frac{1}{2} \sum_{\omega} |x^*(\omega) - x(\omega)|
\]

We refer to this statistic as the number of misplaced cases since it indicates the total number of applicants incorrectly assigned as enrolling or not enrolling due to the estimated yield rates from the model. This measure is primarily useful for evaluating the relative fits of different models.

**Application of Methods to Present Research**

The three statistical methods presented all provide models explaining the enrollment decisions of applicants in terms of various background variables. Discriminant analysis and logistic regression model the individual enrollment decisions in terms of first-order effects of a large number of background variables. Log-linear models explain the enrollment decisions for groups of applicants in terms of main and interactive effects of a small number of background variables. In our search for models explaining the enrollment decision, it is important that the estimated parameters are relatively stable across the four academic years in the study. To assess the stability and provide a final measure of fit for the key models developed, we used the data from the academic year 1977-1978 for the purpose of validation.

To apply discriminant analysis and logistic regression to the data available, the set of primary variables selected from the preliminary analysis provides the background variables for the enrollment models. Since the frequency distribution of parents' income is highly skewed, it is helpful to transform this variable using a log transformation. The distribution of log income is more symmetric and consequently more suitable for the discriminant analysis procedure. Also, for those models that assess the effects of aid quality, we introduce an additional dummy variable,
reflecting whether or not an applicant received any financial aid. This variable separates those applicants who received financial aid from those who did not and provides a better assessment of any aid quality effect.

In the application of log-linear models to the data, one contingency table was selected for analysis across the four academic years. The categories developed for the background variables were selected so that key policy questions could be investigated and that the observed counts in the cells would not be too small. The resultant contingency tables show that a few cells have very small counts. For this reason, we pool several years of data and use the pooled data to develop log-linear models that have a large number of parameters.

5. RESULTS

It is convenient to summarize the key results from the study in the context of the primary objectives for the research. Recall that the first objective concerns the existence and extent of a 'middle income' issue at Stanford. The second objective focuses on specific factors affecting the enrollment decisions of the highest academically qualified applicants. Finally, the last objective examines the influence that the amount and 'packaging' of financial aid awards have on applicants' decisions to enroll. After a brief statement of the issues surrounding each objective, the results pertaining to these issues are highlighted. This is followed by a more detailed discussion of the results from specific enrollment models.

The three statistical methods discussed in the methodology section provide a systematic approach for studying the combined and interactive effects of factors related to the enrollment decision. We used these methods to examine carefully the relationships between the enrollment decision and the following primary factors: Academic Rating, Sum of SAT Scores, Parents' Income, Parents' Contribution, Total Aid, Aid Quality, Scholarship and Self-help. A large number of enrollment models were developed based on different combinations of the primary factors. Many of these models did not fit the data well, since the effects of various factors changed markedly across the years in the study. Among the models where the effects are stable, several models are distinguished as addressing the key issues of the study. These models provide the primary
results from the research and are discussed in the context of the issues
that they address. Throughout this section all financial data are given
in 1972 constant dollars.

The 'Middle Income' Issue

The 'middle income' issue refers to a concern that the rising costs
of a college education may deter applicants from moderate income families
from enrolling at a private university. This issue suggests that upper
income families are able to meet the additional costs of a private educa-
tion from their own resources, whereas financial aid awards meet the
demonstrated need of applicants from lower income families. As costs of
a college education increase, there is considerable financial pressure
placed upon families in the middle income range who feel they are not
eligible for adequate financial aid. This may result in decreased enroll-
ments of applicants from middle income families.

One objective of the present study is to assess the existence and
extent of a 'middle income' issue at Stanford. First, it is important to
determine whether there is any evidence that applicants from middle income
families are less likely to enroll at Stanford than other applicants having
similar academic backgrounds. If the data provide such evidence, we must
then determine the range of incomes affected and measure the actual decrease
in enrollments for applicants in this income range. Finally, if decreased
enrollments are substantial for applicants from the middle income group,
we must examine whether adjustments in financial aid awards are likely to
increase enrollments of these applicants.

The results provide evidence for the existence of a 'middle income'
problem. However, the effects of income on the enrollment decisions are
small relative to the overall effects of academic achievement and financial
aid awards. There is a slight depression in enrollment rates for appli-
cants where the parents' income is between $15,000 and $25,000. The
contrast in enrollment rates between this middle income group and the
lower income group is sharp; however, there is less distinction in the
enrollment rates between this group and the adjacent higher income group.
Finally, the results strongly suggest that substantial increases in
financial aid awards to applicants in the 'middle income' group will
increase the enrollment rates for this group only by five to eight
percentage points.

To provide an overview of the existence of a 'middle income' issue
at Stanford, it is helpful to look at the enrollment rates of all
admitted applicants across the range of parents' income. Unfortunately,
data regarding parents' income are not available for admitted applicants
who did not apply for financial aid during the academic years 1973-1974
through 1975-1976. However, for the academic year 1976-1977, these data are
available from a special survey requesting the information from families
of admitted applicants who did not apply for financial aid. The results
of the survey are summarized in a report by Cohen and Hopkins (1978).
The response rate from the survey was approximatley 55 percent.

For the academic year 1976-1977, Table 13 summarizes the enrollment
rates by various income categories for admitted aid and non-aid applicants.
The last column of the table is based on combining the aid and non-aid applicants. This provides an estimate of the yield rates by income group for all admitted applicants in 1976-1977. It should be noted that these yield rates are estimates, since the data for the non-aid applicants reflect only a sample of this population and the definitions of parents' income for aid and non-aid applicants may not be comparable.

The results indicate that for 1976-1977 the yield rates are approximately 68 percent for applicants from families where the parents' income is less than $15,000 or greater than $40,000. When parents' income is between $15,000 and $40,000 the yield rate is approximately 60 percent for both aid and non-aid applicants. Thus, the results suggest a 'middle income' problem in 1976-1977 insofar as admitted applicants from families in the middle income range are less likely to enroll at Stanford.

Of course, it is impossible to make a general statement about the existence of a 'middle income' issue at Stanford from results based only on one academic year. The enrollment rates for three parents' income categories are given for aid applicants during the academic years 1973–1974 through 1975–1976 in Table 6–8 of the preliminary analysis section. These results indicate that approximately 70 percent of all applicants whose parents' income is $15,000 or less enroll at Stanford, whereas approximately 60 percent of applicants whose parents' income is $25,000 or more enroll. The enrollment rates in the middle income category range from 57 percent in 1974–1975 to 65 percent in 1975–1976. Thus, the results do not provide clear evidence of a 'middle income' problem in these four years.
To examine the 'middle income' issue more closely for aid applicants, it is necessary to control for differing academic backgrounds of the applicants. For example, we know that the highest academically qualified applicants are less likely to enroll at Stanford than applicants having lower academic ratings. Presumably, this tendency results from the observation that highly qualified applicants are more likely to have a number of choices as to which college to attend. If the highly qualified applicants are mainly from middle income families, a depression in enrollments for the middle income range could result from competition with other colleges for these applicants. Table 14 summarizes the distribution of applicants and the yield rates by three parents' income categories for applicants with high and low academic ratings. The results show consistently higher enrollment rates for applicants having low academic ratings across all income categories. Furthermore, approximately one half of the applicants having high academic ratings come from families where parents' income is between $15,000 and $25,000. Certainly, any examination of the 'middle income' issue at Stanford must control for this fact so that the 'middle income' issue is not confused with a competition issue.

Before discussing enrollment models that control for the effects of total aid and academic achievement, it is important to understand the relation between parents' income and the expected parents' contribution. Recall that parents' contribution is computed by the College Scholarship Service as the amount the parents are expected to contribute towards the costs of the freshman year. The amount computed depends on the parents' income and assets as well as number of children and other financial resources.

Graph 15 plots parents' contribution versus parents' income for the academic years 1973-1974 and 1975-1976. To highlight the relationship between the

<table>
<thead>
<tr>
<th>Parents' Income</th>
<th>Yield Rate</th>
<th>Percent of Cases</th>
<th>Yield Rate</th>
<th>Percent of Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under $5,000</td>
<td>22.6%</td>
<td>40.7%</td>
<td>22.6%</td>
<td>40.7%</td>
</tr>
<tr>
<td>$5,000 - $25,000</td>
<td>22.6%</td>
<td>40.7%</td>
<td>22.6%</td>
<td>40.7%</td>
</tr>
<tr>
<td>Over $25,000</td>
<td>30.3%</td>
<td>48.5%</td>
<td>30.3%</td>
<td>48.5%</td>
</tr>
<tr>
<td>Total cases</td>
<td>24%</td>
<td>41%</td>
<td>24%</td>
<td>41%</td>
</tr>
</tbody>
</table>

**Table 14**

Notes: The high academic rating group consists of all aid applicants having the 2 highest academic ratings. All academic rating group consists of all other aid applicants.
two variables, the figure was drawn by first sorting applicants in order of increasing parents' contribution and then plotting the average parents' contribution and the average parents' income for each successive group of 50 applicants. A smooth curve was drawn through the points to differentiate the two years. The plots for the academic years 1974-1975 and 1975-1976 are very similar to the plot for academic year 1973-1974 and consequently are omitted.

Two key points should be noted from the graph. First, expected parents' contribution as a function of parents' income increases gradually up to a contribution of $5,000. The slope of the curve becomes steeper in moving from $5,000 to $4,000 contribution and becomes very steep from $4,000 and above. The first interval where the slope of the curve changes corresponds to parents' income between $15,000 and $20,000. The second interval corresponds to parents' income between $20,000 and $25,000. Consequently, corresponding to incomes under $17,000, parents' contribution ranges from $0 to $2,000. However, corresponding to incomes between $17,000 and $25,000, parents' contribution ranges from $2,000 to $7,000. The large increases in parents' contribution across the $17,000 to $25,000 income range suggest that families in this income range are likely to feel financial pressure in meeting the expected amount. For this reason, we selected the income range $15,000 - $25,000 as a reasonable definition of the middle income group.

The second point to note is that there are substantial differences in parents' contribution across the $15,000 to $25,000 income range between the two academic years. In 1975, the College Scholarship Service reduced the expected parents' contribution for all income levels, with the greatest reductions occurring for families in the middle income group. This adjustment was a response to the 'middle income' issue at a national level and was
designed to alleviate the financial pressures on middle income families due to the rising costs of a college education. From the plot, the expected contribution from parents having an income of $20,000 is $3,000 in 1973-1974 and $2,000 in 1975-1976. In 1976, the College Scholarship Service reversed this policy and increased parents' contribution to the approximate former levels. The following table summarizes the average amounts of variables related to the financial aid package for all aid applicants and indicates the response of the Financial Aids Office at Stanford to the above changes.

<table>
<thead>
<tr>
<th>Academic Year</th>
<th>Budget</th>
<th>Total Aid</th>
<th>Stanford Scholarship</th>
<th>Other Scholarships</th>
<th>Self-help</th>
</tr>
</thead>
<tbody>
<tr>
<td>1973-1974</td>
<td>$5,116</td>
<td>$2,133</td>
<td>$1,137</td>
<td>$372</td>
<td>$684</td>
</tr>
<tr>
<td>1974-1975</td>
<td>$4,977</td>
<td>$2,066</td>
<td>$1,074</td>
<td>$551</td>
<td>$441</td>
</tr>
<tr>
<td>1975-1976</td>
<td>$5,081</td>
<td>$2,490</td>
<td>$1,459</td>
<td>$470</td>
<td>$562</td>
</tr>
<tr>
<td>1976-1977</td>
<td>$5,315</td>
<td>$2,338</td>
<td>$1,118</td>
<td>$564</td>
<td>$651</td>
</tr>
</tbody>
</table>

Notice the substantial difference in the average amount of total aid awarded between the academic years 1973-1974 and 1975-1976, despite the small change in the average budget. Adjusting for the fact that the average amount of other scholarships was approximately $100 less in 1975-1976 than in the other years, the additional average amount of Stanford scholarship offered in 1975-1976 is approximately $500-$300. Since approximately 1200 applicants apply for financial aid each academic year, this amounts to a conservative estimate of $250,000 in additional scholarship awards offered applicants in 1975-1976. The fact that the average amount of total aid in 1976-1977 did not decrease to the former levels of the first two years results from a policy introduced by the Financial Aids Office in 1976-1977. This policy awards an additional amount of self-help to those applicants having no demonstrated need but where the expected parents' contribution is very high. The substantial real increase in budget in 1976-1977 prompted this new mechanism for increasing total aid.

Two key enrollment models address the 'middle income' issue while controlling for differences in total aid awards and academic achievements among admitted aid applicants. The first model uses the discriminant analysis procedure. Since the results from the logistic regression for this model are virtually identical, only the discriminant analysis results are reported. The variables in the model consist of the following: Sum of SAT Scores, Scholarship, Self-help, Income 1 = \log(\text{Income})$, and \text{Income 2} = \left(\log(\text{Income}) - \log(20,000)\right)^2. The Income 2 variable is included to measure deviations in enrollment rates around the $20,000 income level, where the expected parents' contribution appears to increase very rapidly. The log transformation of parents' income is used to make the distribution of income more symmetric and consequently more suitable for the discriminant analysis procedure. For a similar reason, the sum of SAT scores is used in preference to Academic Rating because its distribution is approximately normal.

Table 16 summarizes the estimated coefficients for this discriminant analysis model, referred to as Model A. The estimated coefficients for the standardized variables help to determine the relative importance of the effects associated with each of the five variables in the model on the enrollment decision. For example, the large coefficients for sum of SAT scores indicate that this variable has the strongest, consistent influence on the decision to enroll. The fact that the coefficient is negative in all four academic years indicates that applicants with high SAT scores are less likely to enroll at Stanford than those applicants having lower scores. The effect of the
Table 16
Discriminant Analysis Results for Model A

1. Estimated Coefficients $\hat{a}_1$ for the Discriminant Function

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.71267</td>
<td>1.66366</td>
<td>7.67861</td>
<td>-0.42555</td>
</tr>
<tr>
<td>Sum of SAT Scores</td>
<td>-0.00236</td>
<td>-0.00233</td>
<td>-0.00073</td>
<td>-0.00234</td>
</tr>
<tr>
<td>Scholarship</td>
<td>0.00019</td>
<td>0.00017</td>
<td>0.00013</td>
<td>0.00016</td>
</tr>
<tr>
<td>Self-help</td>
<td>0.00019</td>
<td>0.00025</td>
<td>0.00004</td>
<td>0.00013</td>
</tr>
<tr>
<td>Income 1</td>
<td>0.98323</td>
<td>0.56764</td>
<td>-0.82932</td>
<td>0.86411</td>
</tr>
<tr>
<td>Income 2</td>
<td>0.46585</td>
<td>0.26955</td>
<td>-2.12640</td>
<td>1.35561</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.61535</td>
<td>0.58371</td>
<td>0.75513</td>
<td>0.60321</td>
</tr>
<tr>
<td>Sum of SAT Scores</td>
<td>-0.35874**</td>
<td>-0.34059**</td>
<td>-0.36531**</td>
<td>-0.28352**</td>
</tr>
<tr>
<td>Scholarship</td>
<td>0.29269**</td>
<td>0.29333**</td>
<td>0.19741**</td>
<td>0.24073**</td>
</tr>
<tr>
<td>Self-help</td>
<td>0.09175</td>
<td>0.18417</td>
<td>0.01848</td>
<td>0.06203</td>
</tr>
<tr>
<td>Income 1</td>
<td>0.28406**</td>
<td>0.00932</td>
<td>-0.20666</td>
<td>0.20196</td>
</tr>
<tr>
<td>Income 2</td>
<td>0.07099</td>
<td>0.04115</td>
<td>-0.50075**</td>
<td>0.19465*</td>
</tr>
</tbody>
</table>

* Significant at the .05 level
** Significant at the .01 level

2. Percent of Cases Correctly Classified

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>66.1%</td>
<td>65.3%</td>
<td>67.7%</td>
<td>65.2%</td>
</tr>
</tbody>
</table>

Note: The Percent of Cases Correctly Classified refers to the percent of applicants that the model correctly predicts will or will not enroll.

3. Prominent Correlations between Estimated Coefficients

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Income 1-Income 2</td>
<td>.614</td>
<td>.689</td>
<td>.699</td>
<td>.692</td>
</tr>
<tr>
<td>Income 1-Scholarship</td>
<td>.628</td>
<td>.686</td>
<td>.525</td>
<td></td>
</tr>
<tr>
<td>Income 1-Self-help</td>
<td>.311</td>
<td>.325</td>
<td>.292</td>
<td>.222</td>
</tr>
<tr>
<td>Income 2-Self-help</td>
<td>.244</td>
<td>.227</td>
<td>.339</td>
<td>.291</td>
</tr>
</tbody>
</table>

Note: Only those correlations less than -0.20 or greater than 0.20 are reported.

scholarship variable on the enrollment decision is also very prominent and indicates a greater tendency to enroll with increasing amounts of scholarship. The effects on the enrollment decision due to self-help are very small and probably negligible.

The estimated coefficients for Income 1 and Income 2 address the 'middle income' issue. First, except for academic year 1975-1976, the estimated coefficient for Income 1 indicates a slightly greater tendency to enroll with increasing income levels. However, in the same three years the positive coefficient for Income 2, indicates a slight decrease in enrollments for income levels around $20,000. In academic year 1975-1976, the results are reversed due to the negative coefficients of both variables. Specifically, in this year there is a tendency for a slight decrease in enrollments with increasing income levels coupled with a significant increase in enrollments for income levels around $20,000. The estimated coefficients provide some evidence for a 'middle income' problem in all years except 1975-1976 where enrollments in the middle income group increased. This is consistent with the change in expected parents' contribution during 1975-1976.

The correlations between the estimated coefficients suggest that the effects of the two income variables on the enrollment decision are associated with each other. This is not surprising since both variables assess the effects of income on the enrollment decision. The only other strong correlation occurs between Income 1 and Scholarship. This suggests that any change in the coefficient describing the effect of scholarship on the enrollment decision should be accompanied by a similar change in the coefficient describing the effect of income. In other words, if the relative effect of scholarship on the enrollment decision becomes more pronounced for a particular academic year, it is likely that the effect of income will also become slightly more pronounced.
The discriminant function coefficients measure the relative effects of each variable in the presence of the other factors in the model. To determine the extent of a "middle income" issue for the years in the study, we must look at the combined effects of all variables across different income levels. Graph 17 provides a plot of the estimated enrollment rates from the model for 1974-1975 across the range of parents' incomes. The estimated enrollment rates are derived from the estimated discriminant function and are based on the sum of SAT scores, scholarship awards, self-help award, and parents' income for each applicant. In this way, they control for differences in applicants' academic achievements and financial aid awards across different income levels. To indicate the fit of the model to the data, the figure was drawn by first sorting applicants in order of increasing parents' income and then computing the predicted enrollment rate from the model and the actual enrollment rate for each successive group of 50 applicants. A smooth curve was drawn through the modelled enrollment rates to highlight the general pattern.

The graph indicates that the model fits the data reasonably well for incomes under $40,000. However, for incomes above $40,000, the model is unreliable due to the small numbers of aid applicants from upper income families. Notice the large discrepancy between the actual and modelled enrollment rate for the group of applicants where the average parents' income is $60,000. The graphs for the other years in the study also indicate that the model fits the data reasonably well for incomes under $40,000.

The smooth curve describing the modelled enrollment rates for each of the academic years in the study is exhibited in Graph 10. Again, the fit for the upper income levels is not to be trusted due to the small numbers
The modelled yield rates for the cells of the table provide a summary and interpretation of the combined effects of academic rating, total aid, and parents' income on the enrollment decision. Table 19 gives the modelled yield rates from Model B for the academic year 1974-1975. Notice that the yield rates for applicants in cells characterizing the low academic rating group are on average 13 percentage points higher than the corresponding cells for the applicants having high academic ratings. This result is consistent across all four academic years and highlights the prominent influence of academic rating on the enrollment decision. The effect of total aid on the enrollment decision is the second strongest effect and indicates a greater tendency to enroll for applicants receiving large amounts of total aid. Notice that the modelled yield rates across the aid quality categories are constant. This results from the fact that Model B does not specify an aid quality effect on the enrollment decision.

The yield rates for the middle income group are consistently lower than those for the other two income groups across different total aid and academic rating categories. This indicates a depression in the enrollment rates for the middle income group in 1974-1975, once we control for the effects of academic rating and total aid on the enrollment decision. To detect the extent of the depression in enrollment rates for other years, it is helpful to plot the modelled yield rates by income category for the various total aid and academic rating cells of the table. Graph 20 gives the modelled yield rates in this form for each academic year. The points referring to the three income categories in each total aid group are connected to highlight the effects of income. The plot reveals that the middle income group of applicants is less likely to enroll at Stanford in all years except 1975-1976. In 1975-1976, the graph suggests a gradual
A decline in enrollment rates with increasing incomes. Looking across years, however, it is clear that this change in the shape of the curve is due to increased enrollment in the low and middle income groups, while the enrollment rate in the upper income group remains fairly constant.

Although these modelled enrollment curves refer to only three income categories rather than a continuum of values, they are consistent with those obtained from the discriminant analysis procedure. In fact, the shape of the enrollment curves across income levels are very similar to those given in Graph 18 for Model A. Both models provide evidence for a "middle income" problem. However, they suggest that the extent of the problem is small relative to the effects on the enrollment decision of academic achievement.

Finally, the models assess the impact of the 1975-1976 adjustment in parents' contribution. The enrollment curves in Graphs 18 and 20 suggest that the enrollment rates for applicants in the "middle income" range increased by approximately 5 to 8 percentage points from the other years. Since there are approximately 500 aid applicants in the "middle income" range, the net effect is that between 25 and 40 "middle income" aid applicants, who in other years might have enrolled elsewhere, chose to enroll at Stanford in 1975-1976.

**Attracting High Academic Achievers**

The second objective of the study is to examine separately and in more detail factors affecting the enrollment decisions of the highest academically qualified applicants to Stanford. The group of high academic achievers consists of all applicants who have the two highest academic ratings. We have already seen that this group of applicants is less likely to enroll at
Stanford than the group of applicants having lower academic ratings. Because many good colleges typically compete for the most highly qualified applicants, these applicants usually have several choices as to which college to attend. This partially explains why applicants from this group are less likely to accept offers of admission to Stanford. In the context of the present study, it is important to determine the influence of financial aid factors on the enrollment decisions of high academic achievers.

The results demonstrate that academic competition is an important factor associated with the enrollment decisions of the highest academically qualified applicants. Specifically, as an applicant's academic qualifications increase, the probability of enrolling at Stanford decreases. The results also suggest that high academic achievers respond more strongly to total aid offers than low academic achievers. Large amounts of total aid are associated with higher probabilities of enrolling at Stanford for all aid applicants. However, as the amount of total aid increases, the probability of enrolling at Stanford increases more rapidly for high academic achievers than for low academic achievers.

An overview of the enrollment rates by academic rating category is given in Table 6 of the preliminary analysis section. For each academic year, the enrollment rates for the two highest categories are consistently lower than those for the other categories. For example, the enrollment rate for the lowest category is approximately 75 percent across the four academic years. However, the enrollment rate for the highest category is between 41 and 47 percent for all years except 1975-1976. In this year, the enrollment rate for the highest group is 57 percent, in striking contrast to the other years. The distribution of parents' income by academic rating, given in

Table 14, indicates that almost one half of the applicants having the two highest academic ratings are from middle income families. This suggests that the adjustment in parents' contribution for middle income families in 1975-1976 may partially explain the substantial increase in the enrollment rate for the highest academic rating category in this year.

To detect key factors affecting the enrollment decisions of the highest academically qualified applicants, we first divided the aid applicants into two groups on the basis of the academic rating variable. The High Academic Rating Group consists of all applicants having the two highest academic ratings, whereas the Low Academic Rating Group consists of all applicants having the three lowest academic ratings. Using the discriminant analysis and logistic regression procedures, we fit a number of enrollment models to each group separately. In this way, differences between the two groups are detected by comparing the estimated coefficients from the models.

One model exhibits consistent and important differences between the two groups across all four academic years. This model, referred to as Model C, includes the following variables: Sum of SAT Scores, Total Aid, Income 1 = \( \log (\text{Income}) \), and Income 2 = \([\log (\text{Income}) - \log (\$20,000)]^2\). Table 21 summarizes the estimated coefficients for the discriminant analysis model. The estimated coefficients for the logistic regression are very similar and are omitted. To interpret the relative effects on the enrollment decision of each variable, it is helpful to compare the estimated coefficients for the standardized variables.

The first result suggested by the coefficients is that the sum of SAT scores is a more important factor concerning the enrollment decision for applicants in the high academic rating group than for applicants in the low
In fact, constant differences in the sum of SAT scores have a much greater impact on the enrollment decisions of applicants in the high group than those in the low group. For example, in 1976-1977, the coefficients of the SAT factor are \(-0.00568\) for the high group and \(-0.00304\) for the low group. Suppose that there are two applicants in the low group having identical backgrounds except that the sum of SAT scores for the second applicant is 100 points higher. If the first applicant enrolls at Stanford with probability 0.70, then the probability that the second applicant enrolls is computed by

\[
\log \frac{P}{1-P} \approx \log \frac{0.70}{0.30} = (-0.00304) \times 100
\]

as \(p = 0.68\). However, for the same situation describing two applicants from the high group, the probability that the second applicant enrolls is computed as \(p = 0.57\). The 100 points difference in the sum of SAT scores has very little influence on the enrollment probability for the two applicants in the low group, whereas it has a dramatic influence for the two applicants in the high group.

This result highlights the important role that academic competition has on the enrollment decisions for the highest academically qualified applicants. Specifically, the higher the applicant's academic qualifications, the lower is the probability of enrolling at Stanford. This stems from the fact that highly qualified applicants generally have a number of choices of good colleges to attend whereas other applicants do not have as many alternative choices. Finally, the estimated coefficients indicate that competition appears to have been particularly intense for high achieving aid applicants in 1976-1977.
The other important result from Model C is that the high academic rating group is consistently more sensitive to total aid than the low academic rating group. Large amounts of total aid are associated with higher probabilities of enrolling at Stanford for both groups. However, this effect is particularly pronounced for the high academic rating group. For example, suppose that an applicant in the high academic rating group has an initial probability of enrolling at Stanford in 1973-1974 of 0.60. The estimated coefficients for that year indicate that an additional $1,000 in total aid would increase this probability to 0.67. For a similar applicant in the low academic rating group, an additional $1,000 in total aid would only increase the probability of enrolling to 0.65.

Recall that total aid is the sum of Stanford scholarship, Stanford self-help, and other scholarships. We developed several enrollment models to determine whether the total aid effect could be attributed to one of the components in the aid package. None of these models gave consistent and stable differences between the two academic rating groups across all four years. Furthermore, the total aid effect cannot be attributed to the known bias in the other scholarship variable. The average amount of other scholarships for applicants in the low academic rating is approximately $600 across the years in the study. However, for the high academic rating group, the average amount is approximately $430. The fact that more other scholarships are recorded for enrolling than for not enrolling applicants would tend to bias the total aid effect in favor of the low academic rating group, where the average amounts of these awards are higher. Therefore, the bias in recording other scholarships fails to explain the prominent total aid effect on the enrollment decisions for the high academic rating group.
To measure the impact of the total aid effect on the probability of enrolling at Stanford, we must look at the combined effects of all variables in the model across different total aid levels. Graph 22 provides a plot of the modelled enrollment rates for the high and low academic rating groups for the academic year 1973-1976. The estimated enrollment rates are based on the effects of each variable in the model and control for differences in applicants' academic achievements and family financial circumstances within each group. The figure was drawn by first sorting applicants in order of increasing amounts of total aid and then computing the enrollment rate from the model for each successive group of 20 applicants. A smooth curve was drawn through the modelled enrollment rates to highlight the general pattern for the high and low academic rating groups.

The graph shows that the estimated enrollment rates for the low academic group are substantially greater than for the high academic group. Also, applicants in the low group typically receive larger amounts of total aid as evidenced by the cluster of points for this group at the high end of the total aid range. The different slopes of the curves for the two groups indicate that high academic achievers respond more strongly to increasing amounts of total aid than do low academic achievers. Specifically, as the amount of total aid increases from $0 to $6,000, the estimated probability of enrolling at Stanford increases from 0.70 to 0.78 for the low academic group. However, for the high academic group, the estimated probability of enrolling increases from 0.48 to 0.66. The graphs of the modelled enrollment curves for the other academic years are similar.
The results from the discriminant analysis model suggest that academic competition is an important factor related to the enrollment decisions of the most highly qualified applicants. Furthermore, the model indicates that applicants in the high academic rating group respond more strongly to increasing amounts of total aid than do applicants in the low academic rating group. In view of the fact that the amount of total aid matches demonstrated financial need for all aid applicants, this latter result is not easily explained.

The 'Packaging' of Financial Aid Awards

The last objective of the study is to examine the influence of the size and composition of the financial aid offers made by Stanford on the applicants' decisions to enroll. A 1973 study at the Massachusetts Institute of Technology concluded that the total amount of financial aid offered an applicant chiefly influenced the decision to enroll at M.I.T., rather than how the aid was 'packaged' into scholarships, loans, and jobs. It is important to verify whether this statement also applies to Stanford aid applicants.

The results indicate that higher proportions of scholarships in the aid package are associated with increased enrollment rates. This effect of the quality of the aid package is still very prominent once we control for the influence of total aid, academic achievement, and parents' income on the enrollment decision. As the proportion of scholarship in the aid package increases from 0 to 1, corresponding enrollment rates increase by approximately 15 percentage points. This increase is substantial and suggests that Stanford aid applicants are influenced by the relative proportions of scholarship and self-help in the aid package.

Given the policy that the amount of total aid matches demonstrated need, it is difficult to improve yield substantially by simply increasing the proportion of the aid package that is scholarship. In this sense, the Financial Aid Office is already functioning near optimally in terms of packaging aid awards to attract the most applicants. The results suggest that some modest improvement might be obtained by increasing the aid quality of high academically rated applicants who currently receive only self-help awards.

We have already seen the important effects of scholarship and total aid amounts on the applicants' decisions to enroll. Controlling for differences in academic achievements and family financial backgrounds among applicants, we observed that large amounts of these awards are associated with higher probabilities of enrolling at Stanford. In particular, the total aid effect is particularly pronounced for highly qualified applicants. We now address the issue of whether the 'packaging' of the financial aid award in the form of scholarships, loans, and jobs has an influence on the enrollment decision.

Recall that aid quality refers to the proportion of the aid package that is scholarship. The preliminary analysis of the data revealed that high proportions of scholarship in the aid package are generally associated with large amounts of total aid. A specific question of interest is whether there is any additional aid quality effect once we control for the fact that large amounts of total aid are associated with higher enrollment rates.

To detect any additional aid quality effect, we developed a large number of log-linear probability models to investigate the main effects of aid quality on the enrollment decision as well as interactive effects between aid quality and academic rating, total aid, and parents' income.
None of the interactive effect effects were large enough to be significant.
In fact, the stability and predictive value of the estimated enrollment
decision rates decreased for complex models that incorporate many interactive effects.

For example, Table 23 lists twenty log-linear probability models that
assess the main and interactive effects of aid quality on the enrollment
decision. The models are arranged in increasing order of complexity, so
that models requiring a large number of parameters are listed at the end
of the table. To illustrate the meaning of the codes describing each
model, the ninth model, RQR-IE, includes the following effects:

a. The overall mean effect (I),
b. The main effect of academic rating on enroll (HR),
c. The main effect of parents' income on enroll (IE),
d. The main effect of aid quality on enroll (QR), and
e. The interactive effect of academic rating and aid quality
   on enroll (HRQR).

In this way, models describing interactive effects between academic rating,
total aid, parents' income, and aid quality also include the main effects
of the specified variables.

Except for the first two models, each of these models provide a good fit
to the observed contingency tables with the best fit obtained for the more
complex models. To assess the predictive value of the models, we fit each
model to the observed data and used the estimated yield rates to predict the
numbers of enrolling and not enrolling applicants in 1977-1978. The accuracy
of the predictions can be assessed by the Pearson chi-square, information, or
number of misplaced cases statistics discussed in the methodology section.

Graphs 24-A and 24-B assess the accuracy of the predictions for 1977-1978
based on the information and number of misplaced cases statistics. In each
figure, the estimated yield rates are derived from the observed data for the
previous year (1976-1977), the pooled data for the previous two years, and the
pooled data for the previous four years.

<table>
<thead>
<tr>
<th>Model Number</th>
<th>Model</th>
<th>d.f.</th>
<th>Number Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>E</td>
<td>49</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>IE</td>
<td>48</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>IE-QE</td>
<td>45</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>IE-IE-QE</td>
<td>43</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>IE-IE-QE</td>
<td>42</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>IE-QE</td>
<td>42</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>IE-IE-QE</td>
<td>41</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>IE-IE-QE</td>
<td>40</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>RQR-IE</td>
<td>40</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>RQR-IE</td>
<td>39</td>
<td>11</td>
</tr>
<tr>
<td>11</td>
<td>RQR-IE</td>
<td>39</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>RQR-IE</td>
<td>37</td>
<td>13</td>
</tr>
<tr>
<td>13</td>
<td>RQR-IE</td>
<td>37</td>
<td>12</td>
</tr>
<tr>
<td>14</td>
<td>RQR-IE</td>
<td>37</td>
<td>13</td>
</tr>
<tr>
<td>15</td>
<td>RQR-IE</td>
<td>37</td>
<td>13</td>
</tr>
<tr>
<td>16</td>
<td>RQR-IE</td>
<td>36</td>
<td>14</td>
</tr>
<tr>
<td>17</td>
<td>RQR-IE</td>
<td>36</td>
<td>14</td>
</tr>
<tr>
<td>18</td>
<td>RQR-IE</td>
<td>35</td>
<td>15</td>
</tr>
<tr>
<td>19</td>
<td>RQR-IE</td>
<td>34</td>
<td>16</td>
</tr>
<tr>
<td>20</td>
<td>RQR-IE</td>
<td>31</td>
<td>19</td>
</tr>
</tbody>
</table>

Codes T - Total Aid, I - Income, R - Academic Rating,
Q - Aid Quality, E - Enroll.

87
The first observation from the graphs is that it matters very little whether we use the 1976-1977 data to predict the yield rates for 1977-1978 or whether we pool several previous years' data. This is reflected by the fact that the three curves agree very closely across all twenty models. The second observation highlights the accuracy of the predictions. Models that predict the 1977-1978 yield rates will have small values of the information or misplaced cases statistics. Notice that the poorest predictions are given by the first model which includes only the overall yield effect. The predictions improve by introducing the effects of academic rating and aid quality on the enrollment decision. What is interesting is that more complex models do not improve the predictions substantially. This suggests that there are no prominent interactive effects between aid quality and total aid, academic rating, and parents' income on the enrollment decision.

Due to these results, we used a simple model to determine any additional effect of aid quality on the enrollment decision once we control for the fact that large amounts of total aid are associated with higher enrollment rates. This model, referred to as Model D, includes the main effects of academic rating, total aid, parents' income, and aid quality. After fitting this model to the observed tables for each academic year, we assessed that it provides a good fit to the data. Specifically, Model D explains approximately 55 percent of the information in the observed table that is not explained by the overall yield rate for each academic year. When the model is fit to the table resulting from pooling the data for the four academic years 1973-1974 through 1976-1977, Model D explains 61 percent of the information not explained by the overall yield rate for the pooled data.

The yield rates estimated from the model provide a summary and interpretation of the combined effects of academic rating, total aid, parents' income, and aid quality on the enrollment decision. Table 25 gives the estimated yield rates from Model D for the pooled four years of data. The estimated yield rates for the individual years differ somewhat from those based on the pooled data. However, the main effects of the variables on the enrollment decision are consistent for all years.

The yield rates in Table 25 highlight the effects of total aid and academic rating associated with the decision to enroll that we have already seen. Specifically, large amounts of total aid are associated with higher enrollment probabilities and the yield rates for the low academic rating group are substantially higher than those for the high academic rating group. The new information in the table concerns the effects of aid quality on the enrollment decision. Among the applicants who received some financial aid but less than $2,500 in total aid, the yield rates increase dramatically as the proportion of scholarship in the aid package increases. These yield rates are plotted in Graph 26 for the cells characterizing the three income categories and the two academic rating groups. The estimated yield rates in each cell are connected to make the aid quality effect more clear. As the proportion of scholarship in the aid package increases from 0 to 1, the yield rates increase by approximately 15 percentage points. Furthermore, we noticed that the estimated parameters for the model indicate that the aid quality effect is not linearly associated with the other effects of total aid, academic rating, and parents' income on the enrollment decision. Thus, we conclude that aid quality has an additional effect on the decision to enroll that is not explained by total aid.
Graph 26

Yield Rates from Model D for Second Total Aid Category
Based upon Pooled Data for 1975-1976 through 1976-1977

<table>
<thead>
<tr>
<th>Low Academic Rating</th>
<th>High Academic Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Income</td>
<td></td>
</tr>
<tr>
<td>Modelling Yield Rates</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle Income</td>
<td></td>
</tr>
<tr>
<td>Modelling Yield Rates</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>High Income</td>
<td></td>
</tr>
<tr>
<td>Modelling Yield Rates</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The upper left corner refers to the average number of applicants in the cell for each academic year. The lower right corner of each cell gives the estimated yield rate from the model.

<table>
<thead>
<tr>
<th>Aid Quality Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
</tr>
<tr>
<td>Med.</td>
</tr>
<tr>
<td>High</td>
</tr>
</tbody>
</table>

Low Income

Middle Income

High Income

Aid Quality Category
To relate the present financial aid policies with the estimated yield rates in the table, it is helpful to identify the types of applicants that characterize each cell. The Financial Aid Office at Stanford assigns each applicant in the target population to an aid category that is based on information pertaining to the applicant's academic achievements, minority status, and family financial circumstances. The minimum amounts of self-help required in every financial aid award differ across these aid categories. The aid categories pertinent to this research are the following:

1. Applicants having no demonstrated need
   - DA - Applicants who receive no financial aid award
   - SH - Applicants who receive a small amount of self-help to offset hardship imposed by a very high expected parents' contribution
   - OOT - Applicants who receive outside scholarship but do not receive any Stanford financial aid

2. Applicants having demonstrated need
   - TOP - Applicants who are in the top 5 percent of the class
   - LOW - Applicants who are classified as low income or minority
   - DIV - Applicants who are classified as diversity
   - AVE - All other applicants having demonstrated need

Table 27 gives the aid categories that primarily characterize each cell of the contingency table. Also, the percent of applicants in each cell who receive some amount of Stanford grant in the aid package is given. This table provides some interesting insights pertaining to the financial aid process. Here, it is not surprising that the applicants who receive no financial aid award do not have demonstrated financial need. Also, applicants
having no demonstrated need but who receive some additional self-help characterize the Under $2,500 in total aid and $0 aid quality cells of the table. It is interesting that low income and minority applicants characterize the high total aid and high aid quality cells of the table for the low academic rating group. Applicants classified as diversity are not concentrated in particular cells but are instead scattered throughout the top three total aid and aid quality categories. All other applicants having demonstrated need are also scattered across the cells of the table, although there are relatively few of these applicants in the cells characterizing total aid amounts over $2,500 and high aid quality.

Applicants who receive only outside scholarships are not surprisingly concentrated in high aid quality cells. Also, the amounts of the outside awards are generally less than $2,500. The small percentage of applicants receiving some Stanford grant in the under $2,500 total aid and high aid quality cells suggests that the applicants in these cells receive primarily outside awards. Finally, it is interesting to look at the concentration of the top 5 percent of the class in terms of academic qualifications. Many of these applicants receive financial aid awards in excess of $2,500 and are consequently concentrated in the high aid quality category. However, for those applicants receiving financial aid awards less than $2,500, the top 5 percent group is scattered across the low to high aid quality categories.

The average numbers of applicants in each cell for each academic year appear in Table 25. Two important conclusions are suggested by comparing the estimated yield rates in Table 25 with the types of applicants given in Table 27 for the cells of the contingency table. First, the present financial aid policies do seem to attract a diversified group of freshmen aid applicants.

In particular, the yield rates across the three categories characterizing parents' income are very similar for the pooled data. It is especially interesting that the yield rates in the cells characterizing the two highest aid quality categories are very similar. This indicates that applicants who receive over $2,500 in financial aid are probably not adversely affected by the minimum self-help requirement. Because almost one half of all aid applicants receive offers of financial aid awards in excess of $2,500, this result has particular importance.

The second conclusion pertains to the group of applicants who receive offers of some financial aid, but where the amount is less than $2,500. This is the group of applicants in the second total aid category where the yield rates change dramatically across the four aid quality categories. Although 30 percent of the applicants in the low academic rating group are in this category, 40 percent of the applicants in the high academic rating group comprise this category. Furthermore, 44 percent of the applicants classified as the top 5 percent of the class are in the second total aid category.

Of particular concern is the fact that applicants in this group who receive primarily Stanford grants and self-help are concentrated in the low and moderate aid quality categories. It appears that applicants in the high aid quality category primarily receive outside scholarships that explain the large percentage of scholarship in the financial aid awards. This observation has particular significance for applicants in the high academic rating group. The results suggest that Stanford may not provide enough scholarship incentive to attract the high academic achievers in this group.
Some modest improvements in the yield of high academic achievers in this group are suggested by the model. For example, there are approximately 20 applicants in this group who are in the top 5 percent of the class for each academic year. The estimated enrollment rates from the model under present policies indicate that approximately 15 of these applicants will enroll at Stanford. However, since this group is very special, it is not unreasonable to consider a policy whereby the self-help portions of the aid package would be replaced by scholarship amounts so that the percentage of scholarship in the aid package is at least 80 percent. The estimated enrollment rates from Model D cannot be used to determine the impact of this policy since the data analyzed are historical rather than experimental. If we assume the enrollment rates do not change dramatically from those estimated by the model, the suggested policy would result in approximately 18 of these applicants enrolling at Stanford. Admittedly, the net gain in yield is small since the numbers of applicants in this special group are small.

A second method for improving the yield of high academic rating applicants by altering the composition of the financial aid award is the following. Suppose that all applicants in second total aid category of the high academic rating group receive scholarship amounts so that all financial aid awards have at least 50 percent of scholarship in the package. There are approximately 172 applicants in this group. The estimated enrollment rates from the model under present policies indicate that approximately 90 of these applicants enroll at Stanford. Again, if we assume the enrollment rates resulting from the suggested policy do not differ substantially from those estimated by the model, the model predicts that between 98 and 101 of these applicants would enroll at Stanford.

The small improvements in yield suggested by the above two policies indicate the difficulties involved in trying to improve the yield of the highest qualified applicants by simply altering the composition of the aid package. Two important results from the research highlight the influence of financial aid factors on the decisions to enroll for high academic achievers. First, high academic achievers respond more strongly to total aid offers than low academic achievers. Second, for aid applicants in the second total aid category which includes 50 percent of the high academic rating group, large proportions of scholarship in the aid package are associated with higher enrollment rates. These two results suggest that the yield of high academic achievers may be improved by awarding an honorary scholarship or merit award to these applicants.

The issue of merit awards is complex and requires careful examination of the costs involved compared with improvement in yield. It would be of interest to know whether other colleges that Stanford regards as its competitors offer such awards to the highest qualified applicants. The impact of merit awards on the enrollment decisions of high academic achievers can only be assessed from an experiment, rather than the historical data used in this research. However, the results from this study indicate that this issue should be considered and perhaps an experiment performed. The statistical methods used in this research provide the necessary tools to determine the impact on yield rates resulting from such an experiment.
6. CONCLUSIONS

This research has attempted to provide a careful examination of the influence of financial aid factors on freshmen enrollment decisions at Stanford. Although these factors may not be the most important considerations in applicants’ decisions to enroll, the emphasis of the study has been to detect stable and consistent effects where they exist. By addressing key policy questions, we have determined three important results.

First, there is evidence that applicants from middle income families are less likely to enroll at Stanford than applicants from other income families. However, the results indicate only a slight depression in enrollment rates for applicants in the middle income group. Furthermore, substantial increases in financial aid awards to these applicants in 1975-1976 improved the enrollment rates only by five to eight percentage points.

Second, academic competition is an important factor associated with the enrollment decisions of the highest academically qualified applicants. Specifically, as an applicant’s academic qualifications increase, the probability of enrolling at Stanford decreases. The results also suggest that high academic achievers respond more strongly to total aid offers than low academic achievers. Large amounts of total aid are associated with higher probabilities of enrolling at Stanford for all aid applicants. However, as the amount of total aid increases, the probability of enrolling at Stanford increases more rapidly for high academic achievers than for low academic achievers.

Finally, higher proportions of scholarship in the aid package are associated with increased enrollment rates, even when we control for the influence of academic achievement and the amount of total aid offered on the enrollment decision. The quality of the aid package has only a small influence on the enrollment rates for applicants who receive financial aid offers in excess of $2,500. However, for those applicants who receive some financial aid offer but less than $2,500, the enrollment rates increase by approximately 15 percentage points as the proportion of scholarship in the aid package ranges from no scholarship to full scholarship. This suggests that the fixed minimum amount of self-help required in aid packages does not appreciably affect the enrollment rates of applicants receiving large financial aid offers whereas it may adversely affect the enrollment rates of those applicants offered a small amount of financial aid.

Throughout this research, the complexity of the relationships studied is very apparent. Financial aid awards are intimately related to the applicants’ family financial circumstances. Moreover, academic competition among colleges is often reflected in the enrollment rates of the most highly qualified applicants. Therefore, there is a need for cautious interpretation of policy-related conclusions suggested by the report.

With this qualification in mind, several conclusions are suggested. First, the extent of a ‘middle income’ problem at Stanford does not appear to be dramatic for admitted aid applicants. This does not preclude the possibility that the extent of a problem may appear more dramatic in the pool of all admitted applicants. Recent legislation by the U.S. Congress was introduced to provide additional financial assistance for middle income families by making government-subsidized student loans and Basic Educational Opportunity Grants more available to this group. This may have a positive impact on the slight depression in enrollment rates for applicants in the middle income group at Stanford.
The second suggested conclusion has a more direct bearing on the financial aid policies at Stanford. There is some evidence that Stanford may not provide enough scholarship incentive to attract high academic achievers who are offered small amounts of financial awards. For those applicants in this group who do receive large proportions of scholarship in the financial aid offer, the scholarship portion is often attributed to outside awards rather than Stanford grants. One possibility for improving the yields of high academic achievers in this group is to change the fixed self-help requirement to a proportional self-help requirement for these applicants. In this way, the quality of the aid package will be more comparable to applicants who receive large amounts of financial awards. Also, the results indicate that the issue of merit awards for high academic achievers should be considered. If other universities offer the highest academically qualified applicants such honorary scholarships, Stanford may not be in a competitive position to attract significant numbers of these applicants.

Perhaps, the primary benefit from the present research has been to introduce a statistical methodology for assessing the impact of financial aid policies on freshmen enrollment decisions. The methods discussed control for prominent relationships among applicant characteristics and financial aid awards. Consequently, the effects of factors on freshmen enrollment decisions are more accurately determined and relationships among factors are highlighted. The predicted enrollment rates resulting from models based on these methods are useful for assessing present policies as well as suggesting directions of change due to new policies. We hope that these methods will provide valuable tools for appraising the University's financial aid policies and the impact of those policies on freshmen enrollment rates.

APPENDIX: TECHNICAL DETAILS CONCERNING METHODOLOGY

The methodology section provides a general description of discriminant analysis, logistic regression, and log-linear models in the context of the present study. Certain technical details have been omitted, especially those pertaining to the estimation of parameters from the data available. The purpose of this appendix is to fill in those details and indicate how these three statistical methods are applied to the data in the study.

Discriminant Analysis

A. Enrollment Model

Let \( \mathbf{X} = (X_1, X_2, \ldots, X_k) \) be a random vector summarizing the \( k \) background characteristics of an applicant. Let \( Y \) represent the enrollment decision of the applicant. Specifically, \( Y = 1 \) if the applicant enrolls at Stanford and \( Y = 0 \) otherwise. We assume that \( P(Y=0) = \pi_0 \) and \( P(Y=1) = \pi_1 = 1 - \pi_0 \). The key assumption in discriminant analysis is that the random vector \( \mathbf{X} \) has a multivariate normal distribution with a different mean vector but the same covariance matrix depending on whether \( X \) characterizes an enrolling applicant or a not enrolling applicant. That is, the conditional distribution of \( \mathbf{X} \) given \( Y \) is

\[
\begin{align*}
\mathbf{X} | Y = 0 & \sim \mathcal{N}_k(\mu_0, \Sigma), \\
\mathbf{X} | Y = 1 & \sim \mathcal{N}_k(\mu_1, \Sigma)
\end{align*}
\]

where \( \mu_i \) denotes the \( k \times 1 \) mean vector for Group \( i (i = 0, 1) \) and \( \Sigma \) denotes the \( k \times k \) common covariance matrix.

We can translate this assumption into an assumption concerning the probability of enrolling at Stanford for an applicant having background
characteristics \( \mathbf{z} \). The unconditional density of \( \mathbf{z} \) is

\[
f(z) = \pi_0 f_0(z) + \pi_1 f_1(z)
\]

where

\[
r_1(z) = f(z|Y=1) = (2\pi)^{-\frac{D}{2}} |\Sigma_1|^{-\frac{1}{2}} e^{-\frac{1}{2}(z - \mu_1)'\Sigma_1^{-1}(z - \mu_1)} \quad (i = 0, 1).
\]

From Bayes Theorem, the posterior probability of enrolling at Stanford given the value of \( \mathbf{z} \) is

\[
P(Y=1|\mathbf{z}) = \frac{r_1 f_1(\mathbf{z})}{1 + r_1 f_1(\mathbf{z})}.
\]

This expression simplifies to

\[
P(Y=1|\mathbf{z}) = \frac{1}{1 + e^{-D(\mathbf{z})}}
\]

where

\[
D(\mathbf{z}) = \alpha + \mathbf{z}' \mathbf{\beta},
\]

\[
\alpha = \log \frac{\pi_1}{\pi_0} - \frac{1}{2} (\mu_1' \Sigma_1^{-1} \mu_1 - \mu_0' \Sigma_0^{-1} \mu_0),
\]

and

\[
\mathbf{\beta} = (\mu_1 - \mu_0)' \Sigma_0^{-1}.
\]

Thus, the discriminant analysis model specifies that the enrollment decision of an applicant is related to the background characteristics \( \mathbf{z} \) only by the discriminant function \( D(\mathbf{z}) \).

B. Estimation of Parameters

To apply the discriminant analysis model to the data available, it is necessary to obtain estimates of the unknown parameters \( \pi_0, \pi_1, \mu_0, \mu_1, \) and \( \Sigma \). For each academic year in the study, let \( N_Y \) represent the number of enrolling applicants, \( N_X \) the number of not enrolling applicants, and \( N = N_Y + N_X \) the total number of applicants in the data. Let \( y_j \) represent the enrollment decision of the \( j \)th applicant. The, \( y_j = 1 \) if the \( j \)th applicant enrolls at Stanford and \( y_j = 0 \) otherwise. Let \( \mathbf{z}_j = (z_{j1}, z_{j2}, \ldots, z_{jk}) \) represent the observed values of the \( k \) background variables for the \( j \)th applicant. The observations \( (y_j, \mathbf{z}_j) \) for the \( j \)th applicant \( (j = 1, 2, \ldots, N) \) are assumed independent of those for all other applicants.

Anderson (1958) shows that the maximum likelihood estimates of the unknown parameters in the model are given by:

\[
\hat{\pi}_0 = \frac{N_Y}{N}, \quad \hat{\pi}_1 = \frac{N_X}{N},
\]

\[
\hat{\mu}_0 = \frac{1}{N_Y} \sum_{j=1}^{N_Y} \mathbf{z}_j, \quad \hat{\mu}_1 = \frac{1}{N_X} \sum_{j=1}^{N_X} \mathbf{z}_j,
\]

and

\[
\hat{\Sigma} = \frac{1}{N} \left( \sum_{j=1}^{N} (\mathbf{z}_j - \hat{\mu}_0)(\mathbf{z}_j - \hat{\mu}_0)' + \sum_{j=1}^{N} (\mathbf{z}_j - \hat{\mu}_1)(\mathbf{z}_j - \hat{\mu}_1) \right).
\]

Substitution of these estimates into the linear discriminant function gives the following estimated discriminant function:

\[
\hat{D}(\mathbf{z}) = \hat{\alpha} + \hat{\mathbf{z}}' \hat{\mathbf{\beta}}.
\]
where
\[
\hat{\beta} = \log \frac{N_0}{N_1} + \frac{1}{2} \left( \frac{N_0}{N_0 + N_1} \hat{\Sigma}_0 - \frac{N_1}{N_0 + N_1} \hat{\Sigma}_1 \right)
\]

and
\[
\hat{\Sigma} = (\hat{\Sigma}_0 - \hat{\Sigma}_1) \hat{\Sigma}^{-1}.
\]

Notice that the estimated discriminant function coefficients are explicit functions of the sample mean vectors and sample covariance matrix from the data.

Cornfield (1967) proposed the following approximation to the variance of \(\hat{\beta}_i\) (\(i = 1, 2, \ldots, k\)):
\[
\text{Var} \left( \hat{\beta}_i \right) \approx \hat{\Sigma}^{1/2} \left( \frac{1}{N_0} + \frac{1}{N_1} \right) \hat{\Sigma}^{1/2}
\]

where \(\hat{\Sigma}^{1/2}\) denotes the \((i,j)\) element of the matrix \(\hat{\Sigma}^{-1}\). This approximation is exact when \(\hat{\Sigma}\) is known and is quite good when the group sizes \(N_0\) and \(N_1\) are large. In a similar way, the covariance between \(\hat{\beta}_i\) and \(\hat{\beta}_j\) is approximated by
\[
\text{Cov} \left( \hat{\beta}_i, \hat{\beta}_j \right) \approx \hat{\Sigma}^{1/2} \left( \frac{1}{N_0} + \frac{1}{N_1} \right) \hat{\Sigma}^{1/2}.
\]

The estimated variances are used to construct confidence intervals for the discriminant function coefficients \(\hat{\beta}_i\); while the estimated covariances are used to detect prominent correlations among the coefficients.

C. Hotelling's \(T^2\) Statistic

One measure of the discriminant power in the set of background variables is based on Hotelling's \(T^2\) statistic. The two-sample version of this statistic is given by
\[
T^2 = N_0 N_1 \left( \hat{\Sigma}_0^{-1} \hat{\Sigma}_1 \right)^{1/2} \left( \hat{\Sigma}_0 \hat{\Sigma}_1 \right)^{-1/2}.
\]

The quantity
\[
F = \frac{N_0 + N_1 - k - 1}{(N_0 + N_1 - 2)k} T^2
\]
has an \(F\) distribution with degrees of freedom \(k\) and \(N_0 + N_1 - k - 1\) and is used to test the hypothesis that the mean vectors \(\mu_0\) and \(\mu_1\) for the two groups are equal. In the present study, this statistic is used to compare the relative discriminatory power of models based on different sets of background variables.

Logistic Regression

A. Enrollment Model

Let \(Y\) represent the enrollment decision and let \(X = (x_1, x_2, \ldots, x_k)\) summarize the \(k\) background characteristics of interest for an applicant. The key assumption in logistic regression is the following:
\[
P(Y = 0 | x) = \frac{e^{-\beta_0 - \beta' x}}{1 + e^{-\beta_0 - \beta' x}}
\]

and
\[ P(Y=1|\mathbf{x}) = \frac{1}{1+e^{-(\alpha + \beta^T \mathbf{x})}} \]

where \( \alpha \) and \( \beta = (\beta_1, \beta_2, \ldots, \beta_k) \) are unknown parameters to be estimated from the data. This assumption states that the conditional probability of enrolling at Stanford for an applicant having background characteristics \( \mathbf{x} \) depends only on the linear function \( R(x) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \).

Furthermore, the relationship between the conditional probability of enrolling and \( R(x) \) is described by a logistic function. Notice that no assumptions are made concerning the unconditional probability of enrolling at Stanford or the distribution of the background characteristics \( \mathbf{x} \).

This is in contrast to the discriminant analysis procedure which incorporates distributional assumptions for both \( \mathbf{x} \) and \( y \).

B. Estimation of Parameters

To apply the logistic regression model to the data available, it is necessary to estimate the unknown parameters \( \alpha \) and \( \beta \). For each academic year, let \( y_j \) represent the enrollment decision and \( x_{j1}, \ldots, x_{jk} \) denote the observed values of the \( k \) background variables for the \( j \)th applicant \((j = 1, 2, \ldots, N)\).

To estimate \( \alpha \) and \( \beta_j \), we use the method of maximum likelihood estimation and maximize the conditional likelihood

\[ L(y_1, \ldots, y_N|\mathbf{x}_{11}, \ldots, \mathbf{x}_{1k}) = \prod_{j=1}^{N} \frac{e^{-(\alpha + \beta^T \mathbf{x}_j - y_j)}}{1+e^{-(\alpha + \beta^T \mathbf{x}_j)}} \]

with respect to \( \alpha \) and \( \beta \). Differentiating the log of the conditional likelihood with respect to the \( \beta \) parameters, one finds that the maximum likelihood estimates satisfy the following system of \( k+1 \) equations:

\[ \frac{\partial \log L}{\partial \alpha} = \sum_{j=1}^{N} y_j \frac{1}{1+e^{-(\alpha + \beta^T \mathbf{x}_j)}} - \sum_{j=1}^{N} \frac{1}{1+e^{-(\alpha + \beta^T \mathbf{x}_j)}} = 0, \]

\[ \frac{\partial \log L}{\partial \beta_j} = \sum_{j=1}^{N} y_j x_{j1} \frac{1}{1+e^{-(\alpha + \beta^T \mathbf{x}_j)}} - \sum_{j=1}^{N} x_{j1} \frac{1}{1+e^{-(\alpha + \beta^T \mathbf{x}_j)}} = 0, \quad (j = 1, 2, \ldots, k). \]

Walker and Duncan (1967) propose several methods for solving this system of equations. We used the standard iterative solution, starting with an initial approximation to the parameters and obtaining successive estimates from the Newton-Raphson method.

Once the maximum likelihood estimates \( \hat{\alpha} \) and \( \hat{\beta} \) have been obtained, the estimated variance and covariance can be computed from statistical theory. An asymptotic estimate of the covariance matrix is given by the inverse of the information matrix. For general exponential families, the information matrix equals the expectation of the negative second partial derivatives of the log likelihood (with respect to the parameters). This expectation is easily computed and given in Cox (1970). Substituting the maximum likelihood estimates \( \hat{\alpha} \) and \( \hat{\beta} \) into the functional form provides a consistent estimate of the information matrix. Finally, the estimated variances and covariances of \( \hat{\alpha} \) and \( \hat{\beta} \) are obtained by inverting this matrix.

Log-Linear Models

A. General Log-Linear Models

Suppose that we have a multidimensional contingency table consisting of \( Q \) total cells. Let \( \mathcal{N} \) be an index of the \( Q \) cells. The first key assumption in log-linear models is that the observed frequencies in the
table comprise a random sample from a theoretical distribution, charac-
terized by cell probabilities \( P(n) \), where \( P(n) \geq 0 \) for each \( n \) and 
\( \sum P(n) = 1 \). The second key assumption in log-linear models is that the
cell probabilities \( P(n) \) are determined by the main and interaction
effects of the attribute variables. The specific relationship is the
following:

\[
\log P = \mathbf{\tau} \cdot \mathbf{I}
\]

where

\( \log P \) is a \((q \times 1)\) vector of the logarithms of the cell
probabilities

\( \mathbf{\tau} \) is a \((q \times n)\) design matrix characterizing a specific
model,

\( \mathbf{I} \) is an \((n \times 1)\) vector of parameters describing the
main and interaction effects of the attribute variables

is the model, and

\( n \) refers to the number of \( \tau \) parameters in the model.

For example, for the five-way contingency table given in the
methodology section, a simple log-linear model describing the observed
frequencies is the following:

\[
\log P(n) = L + \sum_{i=1}^{3} \tau_i T_i(n) + \sum_{j=1}^{2} \tau_j I_j(n) + \sum_{k=1}^{2} \tau_k Q_k(n) + \sum_{l=1}^{2} \tau_l R_l(n)
\]

where

\( n = (n_1, n_2, n_3, n_4, n_5) \) is an index variable referring to a
particular \( n \) in the Total Aid \times Income \times Academic Rating \times
Aid Quality \times Enroll contingency table,

\( P(n) \) is the theoretical probability of cell \( n \),

\( L \) is a constant, which serves as a reference
probability,

\( T_i(n) \) is an 'indicator variable' which equals 1 if \( n_i = 1 \), and
0 otherwise,

\( I_j(n) \) is a parameter to be estimated from the data representing the
effect of the \( i \)th total aid category for explaining the
observed probabilities in the table \((i = 1, 2, 3)\), and so forth.

This model explains the observed probabilities in the cells of the five-way
contingency table in terms of the main effects of the five variables
comprising the table. Notice that the \( \tau \) parameter representing the last
category of a variable is set equal to zero, as a normalizing constraint.
Therefore, for this model \( \log P(n, 3, 2, 4, 2) = L \).

B. Conditional Log-Linear Models

For the development of conditional models, we partition the design
matrix into two components \( \mathbf{T} = (\mathbf{T}_1, \mathbf{T}_2) \). The component \( \mathbf{T}_1 \) specifies
all possible interactions among the background variables that do not
involve the dependent variable, which is the enrollment decision for this
study. This constrains the model to fit the observed margins exactly for
the collapsed four-way table determined by total aid, income, academic
rating, and aid quality. The second component \( \mathbf{T}_2 \) specifies a model
relating the enrollment decision to the background variables. The \( \mathbf{I} \)
parameter vector is partitioned in the same way. Consequently, we obtain
from the general model

\[
\log P = \mathbf{T}_1 \mathbf{I}_1 + \mathbf{T}_2 \mathbf{I}_2
\]

Let \( \omega = (\omega_1, \omega_2, \omega_3, \omega_4) \) be an index variable referring to each cell
of the collapsed four-way contingency table determined by total aid, income,
academic rating, and aid quality. Let \( p(m) \) refer to the theoretical yield rate, or conditional probability of enrolling at Stanford, for the group of applicants in cell \( m \) of the four-way table. Then, from the general model and the normalizing constraint concerning the second category of the enrollment decision variable, we obtain

\[
\log P = \zeta + \gamma \pi_1 + \gamma \pi_2
\]

and

\[
\log(1-P) = \zeta + \gamma \pi_1 \, .
\]

Subtracting these two equations gives

\[
\log \left( \frac{P}{1-P} \right) = \gamma \pi_2 \, .
\]

which is the form of the conditional log-linear model used to develop enrollment models for the study.

Notice that these conditional models relate the log odds probability of enrolling at Stanford given various background attributed to \( \tau \) parameters describing the main and interactive effects of the background variables on the enrollment decision. Consequently, the enrollment model decomposes the log odds of enrolling into specific additive effects due to the background variables in a manner similar to the analysis of variance.

C. Estimation of Parameters

The information theory approach used in the present study provides the following method for estimating the \( \tau \) parameters in the general log-linear model. Let \( x(h) \) denote the observed count in the \( h \) cell of the contingency table. Also, let \( x^*(h) \) denote the estimated count in the \( h \) cell by a model characterized by the design matrix \( T \). The design matrix for a specific model imposes certain constraints that the margins of the estimated counts must satisfy. For example, if the model specifies that the cell probabilities \( P(h) \) are determined by only the main effects of the attribute variables, all one-way margins in the estimated table are constrained to equal the corresponding observed margins. In addition to these constraints, the \( \tau \) parameters are estimated by minimizing the discrimination information statistic given by

\[
I(x;x^*) = \sum_h x(h) \log \frac{x^*(h)}{x(h)} .
\]

Ireland and Nolan wrote the computer program CONTAB designed to compute estimates of the \( \tau \) parameters. This program is based on an iterative algorithm first published by Deming and Stephan (1940). This algorithm starts with an initial distribution for the contingency table and a set of observed margins from the data corresponding to the constraints imposed by the design matrix of the model. Typically, the uniform distribution is used as an initial approximation. The algorithm successively adjusts the initial distribution to fit the constrained margins and minimize the information statistic. The process converges rather rapidly to give the minimum discrimination estimates of the \( \tau \) parameters.

Variances of the estimated \( \tau \) parameters are provided by expressing the general log-linear model in the following exponential family form

\[
P(h) = \frac{e^{T(h)\tau}}{\sum_{h'} e^{T(h')\tau}}
\]

where \( T(h) \) refers to the \( h \)th row of the design matrix \( T \). As in logistic regression, the covariance matrix of the estimated \( \tau \) parameters is given by the inverse of the information matrix. The information matrix
can be computed in a straightforward manner and the exact form is
given in Kullback (1978).

Finally, it should be noted that the fitting algorithm to obtain
the estimated $\alpha$ parameters is sensitive to small cell counts. To
correct for this sensitivity, a value of 0.5 was added to each observed
cell count in the table before a model was estimated from the data. This
correction causes a minimal distortion to the original table and has the
effect of shrinking the yield rates towards 50 percent. In cells with
large observed counts, the amount of shrinkage is negligible.

REFERENCES


applicants to Stanford University," Academic Planning Office, Stanford
University.

35, 511-524.


Curry, P., Hopkins, D., and Main, S. (1973). "Freshmen admissions and
financial aid at Stanford," Report No. 73-1, Academic Planning Office,
Stanford University.

sampled frequency table when the expected marginal totals are known,"


Marcel Dekker Inc., New York.

multivariate logistic risk function: a comparison of the discriminant

decisions of admitted freshmen applicants," Academic Planning Office,
Stanford University.


Acknowledgments

Financial support for conducting this research was provided by a grant from the Lilly Endowment, Inc. We are grateful to Mary Roberts of the Financial Aids Office at Stanford for helpful explanations of the financial aids process and assistance in data collection. We also wish to thank Carol Link for assistance in computer programming and data collection in addition to useful statistical suggestions.