ABSTRACT

Craig Zirbel, Advisor

Bowling Green State University, like other universities, administers a mathematics placement test to incoming students to help to identify an appropriate first math course. Students select and take one of three tests depending on their math background, the test is scored and a placement decision is made. The issue here is whether the placement decision can be improved by changing the criteria. The goal of this thesis is to accurately predict a student’s grade from a subset of available covariates for students enrolling in Math 126 at BGSU. The hope is that an accurate methodology and model can be found so that it may be considered as a placement system for students into their first math course at BGSU.

We focused on the subset of the data in which students took version B of the placement exam and enrolled in Math 126 (Business Calculus). Three regression models were used to fit the data: linear, logistic, and ordinal logistic.
For Laura and Julia
ACKNOWLEDGMENTS

I would like to thank my advisor, Craig Zirbel, for his guidance, encouragement, and generosity. I thank the BGSU mathematics department whose funding enabled me to undertake this project. Finally, I thank my wife, Laura, and daughter Julia who have patiently shared the burden of my work.
Table of Contents

CHAPTER 1: Introduction 1
  1.1 Goal ................................................... 1
  1.2 The Data Set ........................................... 2
  1.3 Initial Analysis ....................................... 5
  1.4 Related Literature .................................... 8

CHAPTER 2: Models 9
  2.1 Linear Regression ..................................... 9
  2.2 Logistic Regression .................................... 13
  2.3 Ordinal Regression .................................... 16

CHAPTER 3: Diagnostics 18
  3.1 Hypothesis Testing .................................... 18
  3.2 Variable Selection ..................................... 19
  3.3 Model Fit .............................................. 21

CHAPTER 4: Results 23
  4.1 Linear Regression Model .............................. 23
  4.2 Logistic Regression Model ............................ 24
  4.3 Ordinal Logistic Regression Model .................. 30

CHAPTER 5: Conclusions 35
# List of Figures

1.1 Scatterplots of selected variables (126B data set) .......................... 6
1.2 Graph correlation coefficients of all available variables (126 B data set) ... 7

4.1 Linear model with predictors ACTMAT, HSGPA, ACTSCI, Exponents ... 25
4.2 Linear model with predictors ACTMAT, HSGPA, ACTSCI, Exponents ... 26
4.3 Logistic model for predictors ACTMAT, HSGPA ............................... 27
4.4 Logistic model for predictors ACTMAT, HSGPA, ACTSCI, and Exponents . 29
4.5 Logistic model for predictors ACTMAT, HSGPA, ACTSCI, and Exponents . 33
4.6 Ordinal model for predictors ACTMAT, HSGPA ............................... 34
# List of Tables

1.1 Student totals for each placement exam and Math course .......................... 4

3.1 Explanation of Actual vs. Predicted Tables (Linear/Ordinal) ...................... 21
3.2 Explanation of Actual vs. Predicted Tables (Logistic) .............................. 22

4.1 Logistic results 1 ................................................................................... 24
4.2 Logistic results 2 ................................................................................... 28
4.3 Logistic results 3 ................................................................................... 28

5.1 Results of ACTMAT, HSGPA ................................................................. 35
5.2 Results of ACTMAT, HSGPA, Advanced ................................................. 36
5.3 Results of ACTMAT, ACTSCI, HSGPA, Exponents ............................... 36
5.4 Pass/Fail Results for predictors ACTMAT, HSGPA ................................. 36
5.5 Pass/Fail Results for predictors ACTMAT, HSGPA, Advanced ................. 36
5.6 Pass/Fail Results for predictors ACTMAT, ACTSCI, HSGPA, Exponents .. 36
5.7 Resulting Percentages of Predicted vs. Actual ........................................... 37
CHAPTER 1

Introduction

1.1 Goal

Bowling Green State University, like other universities, administers a mathematics placement test to incoming students to help to identify an appropriate first math course. Students select and take one of three tests depending on their math background, the test is scored and a placement decision is made. The issue here is whether the placement decision can be improved by changing the criteria. The goal of this thesis is to accurately predict a student’s grade from a subset of available covariates for students enrolling in Math 126 at BGSU. The hope is that an accurate methodology and model can be found so that it may be considered as a placement system for students into their first math course at BGSU.

There are a variety of predictive variables available. We want to use those variables to accurately predict the grade and whether the student will pass or fail. We will use logistic regression to predict the binary outcome pass/fail, and linear and ordinal regression to predict grade, and thereby, pass/fail.
1.2 The Data Set

The data set contains information from 5791 BGSU students who took the mathematics placement test online and who subsequently enrolled in a math course at BGSU. The data were provided by Institutional Research and the College of Arts and Sciences. Personally identifying information such as student ID number was stripped, and only data from the first two (math) courses was retained. This set potentially contains the following information for each student: junior year high school grade point average (HSGPA); ACT math (ACTMAT), science (ACTSCI), English (ACTENG), reading (ACTRE) scores; SAT math and verbal scores; BGSU math placement code (NPlmnt); individual answers from the placement exam, including background questions; grade received in first math course (Grade); version of placement exam. For an explanation of math courses used in this paper refer to the Appendix.

The ACT is a curriculum-based measure of college readiness. The ACT sets a college readiness benchmark for each area. For math this number is 22. A benchmark score is the minimum score needed on an ACT subject-area test to indicate a 50% chance of obtaining a B or higher or about a 75% chance of obtaining a C or higher in corresponding credit-bearing college courses which include English Composition, Algebra, Social Science, Biology (http://www.act.org/news/data/08/pdf/states/Ohio.pdf). The average math score for students graduating in 2008 in Ohio was 21.5 compared to the national average of 21.0.

Students taking the placement exam take one of three versions (A, B, C). Based on their college preparatory mathematics background, students choose which test to take. The placement exams consist of 10 background questions (common to all three exams) as well as 35 multiple choice questions. Five of the background questions ask the student if they have completed courses in high school mathematics (e.g. Geometry, Trigonometry, Calculus). Other questions ask the student the completed number of semesters of Algebra, if they can use a graphing calculator, if they have taken the Advanced Placement exam, if they are applying for transfer credits, and the version of the exam they are currently taking. The
difficulty of the 35 multiple choice questions are based on exam choice. Exam C is intended for students who have satisfactorily completed a semester in Analytic Geometry, Precalculus, Analysis, Trigonometry, or Calculus. These students can be placed into Math 126, 128, 130, 131, or 134. Those students who perform poorly on exam C are recommended to take Exam B or take Math 112 or 115. Exam B is intended for students who have completed two years of Algebra with grades of C or better. These students can be placed into Math 095, 112, 115, 122, 126, 128, or 130. Taking exam B cannot result in placement into Math 131. Students who do not match the above criteria are recommended to take exam A. These students can be placed into Math 090, 095, 112, or 115. Students who score high on exam A are encouraged to take exam B.

After the placement test is taken, the BGSU mainframe computer scores it and records a numerical placement code. It is generated from exam results as well as other predictors by the current math placement algorithm. Students may then enroll in a course that corresponds to their numerical placement. Students are not permitted to take a course higher than their numerical placement allows. Historically students who have taken courses above their placement have done worse than students who placed into the same course.

Note: The placement test itself is not included in the Appendix to protect its confidentiality. It is important to mention the data set was also censored. For example, students who did poorly on placement test B were not allowed to take Math 126. There is no information about how these weaker students would have fared in this course.

The number of students who took each version of the placement exam and course enrollment is presented in Table 1.1. For the remainder of the thesis I chose to concentrate on the subset of the data consisting of students who took placement exam B and subsequently enrolled in Math 126. There were 172 students in this subset.

It is useful to discuss the nature of the variables at this time. All zeros denote missing data for HSGPA, ACT, SAT, etc.; however, NPlmnt can be ‘.’ for a missing value. HSGPA is a continuous variable between 0 and 5.11. The BGSU admissions office attempts to standard-
<table>
<thead>
<tr>
<th>Math course</th>
<th>Exam A</th>
<th>Exam B</th>
<th>Exam C</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>090X</td>
<td>28</td>
<td>7</td>
<td>1</td>
<td>36</td>
</tr>
<tr>
<td>095X</td>
<td>204</td>
<td>86</td>
<td>3</td>
<td>293</td>
</tr>
<tr>
<td>090</td>
<td>32</td>
<td>9</td>
<td>1</td>
<td>42</td>
</tr>
<tr>
<td>095</td>
<td>231</td>
<td>103</td>
<td>3</td>
<td>337</td>
</tr>
<tr>
<td>112</td>
<td>461</td>
<td>686</td>
<td>76</td>
<td>1223</td>
</tr>
<tr>
<td>115</td>
<td>308</td>
<td>658</td>
<td>241</td>
<td>1207</td>
</tr>
<tr>
<td>115R</td>
<td>12</td>
<td>26</td>
<td>11</td>
<td>49</td>
</tr>
<tr>
<td>120</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>122</td>
<td>146</td>
<td>521</td>
<td>84</td>
<td>751</td>
</tr>
<tr>
<td>126</td>
<td>12</td>
<td>172</td>
<td>158</td>
<td>342</td>
</tr>
<tr>
<td>128</td>
<td>26</td>
<td>273</td>
<td>168</td>
<td>467</td>
</tr>
<tr>
<td>129</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>130</td>
<td>1</td>
<td>31</td>
<td>61</td>
<td>93</td>
</tr>
<tr>
<td>131</td>
<td>0</td>
<td>9</td>
<td>228</td>
<td>237</td>
</tr>
<tr>
<td>131H</td>
<td>0</td>
<td>0</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>134</td>
<td>0</td>
<td>4</td>
<td>43</td>
<td>47</td>
</tr>
<tr>
<td>135</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>213</td>
<td>5</td>
<td>10</td>
<td>13</td>
<td>28</td>
</tr>
<tr>
<td>232</td>
<td>1</td>
<td>2</td>
<td>38</td>
<td>41</td>
</tr>
</tbody>
</table>

Table 1.1: Number of students who took placement test A, B, or C and enrolled in Math course listed above. Students with missing Grade data were not counted.
ize these into a range from 0 to 4, but this is apparently not always successful. ACT scores are discrete and range, in principle, from 2 to 36. The ACT consists of four sections: Math, English, Reading, Science. SAT scores are multiples of 10 from 200 to 800. The SAT has two sections, Math and Verbal. Nplmnt can take on values 11, 20, 24, 32, 41, 42, 51, 61, 71 which correspond to courses listed in the Appendix. Individual answers on the placement exam were coded 1 for correct and 0 otherwise (incorrect/blank). A variable named Raw Score was constructed by counting the correct responses. Grade in the first math class (quality points) consisted of values 0, 1, 2, 3, or 4 corresponding to F, D, C, B, A. Grades of W, WF were coded as an F. Anticipating that responses to individual questions on the placement exam may not be useful for prediction, each question of each version of the placement exam was evaluated and grouped with similar questions. This created six subsections for each version, and the total correct in each subsection was called a subscore. These subscores were hoped to be more robust predictors than individual items. What areas compose the subscores are defined in the Appendix. Which variables were ultimately chosen for each model will be described below.

Students with missing or unavailable data were ignored when the models were fitted to the data. This was accomplished by first designating which variables to use in a model. Then the students with missing data were removed before the model was fitted. Also SAT scores were unavailable for a majority of students so they were not considered as predictors.

### 1.3 Initial Analysis

Before a statistical model was fitted to the data set it was necessary to examine the data in order to understand the dependence between variables and the general characteristics of the data. In Figure 1.1 the students who took placement test B and Math 126 were used. Five variables were chosen: ACTMAT, ACTSCI, HSGPA, Raw Score, and the subscore
Figure 1.1: Scatterplots of selected variables (126B data set). Each point represents a student. Points are colored by Grade achieved in first math class where red = A, orange = B, yellow = C, cyan = D, and blue = F. Points are shifted slightly so they do not overlap Advanced. Each point corresponds to variables labeled on the axes with Grade designated by color. The color red = A, orange = B, green = C, cyan = D, and blue = F. Consider ACTMAT, HSGPA, and ACTSCI colored by course grade (Figure 1.1). The scatterplots indicate that students with higher ACT Math, HSGPA, and Raw Score got higher grades in Math 126. They also show, however that there is substantial variability in the grade earned, even among students with similar values for these predictors.

Figure 1.2 displays the degree of the linear relationship, called the correlation coefficient ($\rho$), between all variables. For example (again using 126 B data set), consider ACTMAT, HSGPA and Grade. It was expected that as ACTMAT became large HSGPA and Grade would also grow. Figure 1.2 showed that this was the case, i.e. the correlation was positive.
Figure 1.2: Scatter plot of correlation between all available variables for the Math 126 B data set. The variables are ordered by correlation to put more highly correlated variables near one another in the list. Questions 1-35 correspond to questions on the placement exam. The background questions are labeled by type of question. Any data points with an entry of 0 were omitted before calculating correlation.

Using MatLab, $\rho$($\text{ACTMAT}$, HSGPA) = 0.1658 while $\rho$($\text{ACTMAT}$, Grade) = 0.2760 and $\rho$($\text{Grade}$, HSGPA) = 0.4321. Some other observations from Figure 1.2 were that all the pairs of ACT scores were at or above $\rho = 0.5$, but were not strongly correlated with Grade. The largest ACT score correlated with Grade was ACTMAT - the remaining correlations were less than 0.2320. The greatest correlation of the subscores with Grade was Exponents at $\rho = 0.2696$.

It was surprising that the ACT scores were so positively correlated with one another. It is plausible that success in math would lead to greater success in science; however, profi-
ciency in mathematics does not guarantee success in English. It was also surprising to find that Raw score was only weakly correlated with Grade ($\rho = 0.2473$) and that some of the individual questions from the BGSU placement exam were negatively correlated with Grade.

### 1.4 Related Literature

Dorner and Hutton (2002) examined whether a mathematics placement system accurately predicted success in mathematics classes for both genders at a medium-sized independent Northwest university. They used a multiple linear regression model with predictors previous mathematics course, previous mathematics course grade, placement test score for a particular course, and SAT math. They found that multiple predictors add to the gender bias (towards women) versus SAT math alone. Course grades for men were overestimated while women’s scores were underestimated.

Naik and Ragothaman evaluated three different models – neural networks, logit, and probit to predict MBA student performance in graduate programs. They used predictors overall undergraduate GPA, junior/senior undergraduate GPA, graduate management admission test, age, undergraduate institution, undergraduate major, age, citizenship status, gender, and race. Graduate GPA was dichotomized and used as the dependent variable. They found that the neural network prediction accuracy was 89.13%, logistic was 72.83%, and probit was 73.37%.

Sheel, Renner, and Dawsey (2002) compared discriminant analysis to neural networks in predicting mathematics placement. Using a multiple regression analysis, high school GPA, SAT math, and final grade in Algebra II were found to be the best predictors of success in mathematics placement. Neural networks outperformed discriminant analysis in prediction with accuracy of 89.9% versus 67.7% respectively.
CHAPTER 2

Models

The variables ACTMAT, ACTSCI, ACTENG, ACTRE, and HSGPA were standardized before any of the models were fitted to the data. This was done to increase the ease of interpretation of the estimated coefficients.

2.1 Linear Regression

This thesis attempts to predict course grade given ACT scores, HSGPA and responses from a placement test. Intuitively we expect that as certain predictors (such as ACTMAT, HSGPA) increase, Grade will also increase. Linear models are convenient due to their ease of interpretation. The predictors together with the response for each student are thought of as ordered pairs \((x_i, y_i)\) which are described further in the model definition. Throughout this section \(N\) is the total number of data points, \(i\) denotes a distinct data point where \(1 \leq i \leq N\), and \(n\) is the number of predictors.

Generalized linear regression models consist of three parts:

1. A random part, where the dependent variable \(Y\) follows one of the distributions of the exponential family (e.g. normal, binomial)

2. A linear part, which describes how a function, \(\hat{Y}\), of the dependent variable, \(Y\) depends on a collection of predictors
3. A link function, which describes the transformation of the dependent variable $Y$ to $\hat{Y}$. [10]

The identity link function does not alter the dependent variable which gives the general linear model for continuous outcomes.

For simple linear regression with one predictor the regression model is defined by

$$Y_i = \alpha + \beta_1 x_i + \epsilon_i$$

where:

- $y_i$ is the observed value (Grade of the $i$th student) of the response variable $Y_i$
- the $Y_i$s are independent
- $\alpha$ and $\beta_1$ are parameters
- $x_i$ is a known constant, namely, the value of the predictor variable for the $i$th student
- $\epsilon_i$ is a random error term with mean $E[\epsilon_i] = 0$ and variance $\sigma^2[\epsilon_i] = \sigma^2$; $\epsilon_i$ and $\epsilon_j$ are uncorrelated so that their covariance is zero (i.e. $\sigma[\epsilon_i, \epsilon_j] = 0$ for all $i, j; i \neq j$)
- $i = 1, \ldots, N$.

[8]

The regression coefficients have the following interpretation. $\beta_1$ is the slope of the regression line. For every unit increase of $x$, the effect on $y$ increases by $\beta_1$. $\alpha$ is the $y$-intercept of the regression line. For our discussion this is not relevant since it provides the value of $y$ when $x = 0$, i.e. we can determine what grade a student will achieve given predictors with values of zeros; however, the intercept is still an important part of the model, as it shifts the predicted Grade up or down.

This model can be generalized to $n$ predictors by

$$y_i = \alpha + \beta_1 x_{i1} + \cdots + \beta_n x_{in} + \epsilon_i$$
where $\epsilon_1, \ldots, \epsilon_N$ are independent, identically distributed normal random variables.

In order to maintain consistent notation I have used $\alpha$ as the intercept parameter, $\beta_1$ as the parameter for $x_1$, etc. We write $\beta$ for the vector of these parameters. One method for finding estimates for $\beta$ is the method of least squares. Least squares strives to minimize the distance from $y_i$ to its expected value, $y_i - \alpha - \beta \cdot x_i$. This is consistent with maximum likelihood estimation when $\epsilon$ is normally distributed.

To accomplish this consider the sum of $N$ squared distances (for $N$ data points)

$$Q = \sum_{i=1}^{N} (Y_i - \alpha - \beta \cdot x_i)^2.$$

In order to simplify calculations initially, assume there is a single predictor $x$. To find estimates for $\alpha, \beta_1$ use calculus, i.e. find $\frac{\partial Q}{\partial \alpha}, \frac{\partial Q}{\partial \beta_1}$.

\[
\frac{\partial Q}{\partial \alpha} = 2 \sum_{i=1}^{N} (y_i - \alpha - \beta_1 x_i) (-1)
\]

\[
\frac{\partial Q}{\partial \beta_1} = 2 \sum_{i=1}^{N} (y_i - \alpha - \beta_1 x_i) (-x_i)
\]

Next we set these derivatives equal to 0. $\alpha, \beta_1$ are replaced by $b_0, b_1$ since these value minimize $Q$. Note that at times I omitted the indices on the summation when it is obvious.

\[-2 \sum (y_i - b_0 - b_1 x_i) = 0\]

\[-2 \sum X_i (y_i - b_0 - b_1 x_i) = 0\]

After simplifying,

\[\sum y_i - nb_0 - b_1 \sum x_i = 0\]

\[\sum x_i y_i - b_0 \sum x_i - b_1 \sum x_i^2 = 0\]
These are called the normal equations.

To show how to find \( b_0, b_1 \) using matrices, rewrite the normal equations.

\[
\begin{bmatrix}
 nb_0 + b_1 \sum x_i \\
 b_0 \sum x_i + b_1 \sum x_i^2
\end{bmatrix}
= \begin{bmatrix}
 \sum y_i \\
 \sum x_i y_i
\end{bmatrix}
\]

\[
\Rightarrow \begin{bmatrix}
 n \sum x_i \\
 \sum x_i \sum x_i^2
\end{bmatrix} \begin{bmatrix}
 b_0 \\
 b_1
\end{bmatrix}
= \begin{bmatrix}
 \sum y_i \\
 \sum x_i y_i
\end{bmatrix}
\]

\[
\Rightarrow xx' b = x'y
\]

where

\[
x' x = \begin{bmatrix}
 1 & 1 & \cdots & 1 \\
 x_1 & x_2 & \cdots & x_n
\end{bmatrix}
\begin{bmatrix}
 1 & x_1 \\
 1 & x_2 \\
 \vdots & \vdots \\
 1 & x_n
\end{bmatrix}
= \begin{bmatrix}
 1 & \sum x_i \\
 \sum x_i & \sum x_i^2
\end{bmatrix}
\]

Thus,

\[
\beta_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}
\]  \hspace{1cm} (2.1)

\[
\alpha = \bar{y} - \beta_1 \bar{x}.
\]  \hspace{1cm} (2.2)

Under the assumption that the distribution of the \( Y_i \)s is normal,

\[
Y_i \sim N(\alpha + \beta x_i, \sigma^2), \ i = 1, \ldots, N.
\]
The joint pdf of $Y_1, \ldots, Y_N$ is

$$f(y|\alpha, \beta, \sigma^2) = f(y_1, \ldots, y_N|\alpha, \beta, \sigma^2) = \prod_{i=1}^N f(y_i|\alpha, \beta, \sigma^2) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma}} \exp \left[ -\frac{(y_i-(\alpha+\beta x_i))^2}{2\sigma^2} \right] = \frac{1}{(2\pi)^{N/2}\sigma^N} \exp \left[ -\left( \sum_{i=1}^N (y_i - \alpha - \beta x_i) \right)^2 / (2\sigma^2) \right].$$

The likelihood function is the product of the individual densities. Maximizing likelihood over $\alpha, \beta$ coincides with minimizing the squared error, provided $\sigma^2$ is known. When the variance is unknown the likelihood function of $\alpha, \beta,$ and $\sigma^2$ is:

$$L(\alpha, \beta, \sigma^2|x, y) = \frac{1}{(2\pi)^{N/2}\sigma^N} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i - \cdots - \beta x_i)^2 \right]$$

$$\log L(\alpha, \beta, \sigma^2|x, y)) = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log \sigma^2 - \frac{\sum_{i=1}^N (y_i - \alpha - \beta x_i)}{2\sigma^2}.$$

Maximizing the log likelihood, when $\sigma^2$ is fixed, yields the same estimates for $\alpha, \beta$ that were found in equations 2.2, 2.1. [2]

### 2.2 Logistic Regression

When the response variable, $Y,$ is dichotomous it is more appropriate to use a logistic regression model. In this case, the two responses will be coded as 0 and 1. Response 1 is called success, 0 is called failure. For the data set success means a Grade of passing, i.e. an A, B, or C. A Grade of D is considered by BGSU to be passing, but it’s not sufficient to go onto the next course, so for our analysis a failing Grade is a D or an F. Just as with linear regression, it is desired to know $E(Y|x).$ The notation $\pi(x) = P(Y = 1|X = x)$ will be used to denote the conditional mean of $Y$ given $x,$ i.e. $\pi(x) = E(Y|X = x) = P(Y = 1|x)$ where $x = [1 \ x_1 \ x_2 \ldots]'.$ We could then model this quantity by $E(Y|x) = \beta \cdot x.$ Thus we have a linear model. But this would be a linear model which is inappropriate. The linear regression
model discussed earlier requires outcomes to follow an approximately normal distribution (not binary). Second, the outcomes for the model will represent probabilities. This means that any estimates for the regression coefficients must restrict these probabilities to lie in [0,1]. Also, intuitively it is not expected that the probabilities will increase linearly with predictor \( x \) over the whole range of \( x \) values. For minimal values of \( x \), \( \pi(x) = 0 \) should be true, likewise for maximal values of \( x \), \( \pi(x) = 1 \). A function that does satisfy this criterion is

\[
\pi(x) = \frac{e^{\beta \cdot x}}{1 + e^{\beta \cdot x}}
\]

where \( \beta = [\alpha \quad \beta_1] \). In summary the logistic regression model assumes:

1. the outcomes are 0 or 1,
2. \( P(Y = 1|X = x) = \pi(x) = \frac{e^{\beta \cdot x}}{1 + e^{\beta \cdot x}} \),
3. the outcomes \( Y_1, \ldots, Y_N \) are independent.

The logistic regression model for multiple predictors is a generalization of the single predictor model. Given \( n \) predictors \( x_1, x_2, \ldots, x_n \) and a binary outcome \( y \) the logistic model is defined as:

\[
\pi(x) = \frac{e^{\beta \cdot x}}{1 + e^{\beta \cdot x}}
\]

where \( \beta = [1 \quad \beta_1 \quad \beta_2 \quad \cdots \quad \beta_n] \). This is equivalent to

\[
\ln \left( \frac{\pi(x)}{1 - \pi(x)} \right) = \beta \cdot x
\]

since

\[
\begin{align*}
\pi(x) &= \frac{e^{\beta \cdot x}}{1 + e^{\beta \cdot x}} \iff \pi(x) + \pi(x)(e^{\beta \cdot x}) = e^{\beta \cdot x} \\
&\implies \pi(x) = e^{\beta \cdot x} - \pi(x)e^{\beta \cdot x} \\
&\implies \pi(x) = e^{\beta \cdot x}(1 - \pi(x)) \\
&\implies \ln \left( \frac{\pi(x)}{1 - \pi(x)} \right) = \beta \cdot x.
\end{align*}
\]

The function \( p \to \ln \frac{p}{1-p} \) is called the logit transformation, which is a nonlinear function that we use to link the independent variables to \( \frac{\pi(x)}{1-\pi(x)} \) - the odds of \( P(Y=1|X=x) \). Using this
formula we can predict the log odds of \( Y = 1 \). If a comparison is needed for distinct values of a particular predictor the odds ratio (OR) may be used. For example to compare the odds of passing (a particular course) between ACTMAT scores of 21 and 22 construct the odds ratio,

\[
OR = \frac{\frac{\pi(Y=1|x=22)}{1-\pi(Y=1|x=22)}}{\frac{\pi(Y=1|x=21)}{1-\pi(Y=1|x=21)}} = \frac{e^{\beta \cdot 22}}{e^{\beta \cdot 21}} = e^{\beta(22-21)}.
\]

Odds ratios of 1 indicate there is no effect on the outcome from the predictor, or an ACTMAT score of 22 versus 21 has no greater effect on the odds of passing. Odds ratios of less than 1 indicate students with an ACTMAT score of 21 have a higher odds of success for passing. The opposite is true for odds ratios greater than 1.

The coefficients are typically estimated by maximum likelihood. To show this write the likelihood function, \( L \). \( L \) is a function of the unknown parameter, \( \beta \), and the observed values of \( x_i \) and \( y_i \).

Since the observations are assumed to be independent

\[
L(\beta; x, y) = \prod_{i=1}^{N} \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i}.
\]

Notice that when \( y_i = 1 \), \( \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i} = \pi(x_i) = P(Y = 1|X = x_i) \) and when \( y_i = 0 \), it equals \( 1 - \pi(x_i) = P(Y = 0|X = x_i) \). Now

\[
\log L(\beta) = \sum_{i=1}^{N} [y_i \pi(x_i) + (1 - y_i)(1 - \pi(x_i))].
\]

Substituting \( \pi(x) = \frac{e^{\beta \cdot x}}{1 + e^{\beta \cdot x}} \),

\[
\log L(\beta) = \sum_{i=1}^{N} \left\{ y_i [\ln e^{\beta \cdot x} - \ln(1 + e^{\beta \cdot x})] + (1 - y_i) [\ln 1 - \ln(1 + e^{\beta \cdot x})] \right\}
= \sum_{i=1}^{N} [y_i(\beta \cdot x) - y_i \ln(1 + e^{\beta \cdot x}) - (1 - y_i) \ln(1 + e^{\beta \cdot x})]
= \sum_{i=1}^{N} [y_i(\beta \cdot x) - y_i \ln(1 + e^{\beta \cdot x}) - \ln(1 + e^{\beta \cdot x}) + y_i \ln(1 + e^{\beta \cdot x})]
= \sum_{i=1}^{N} [y_i(\beta \cdot x) - \ln(1 + e^{\beta \cdot x})].
\]
For each \( j = 1, \ldots, n + 1, \)

\[
\frac{\partial L}{\partial \beta_j} = \sum x_{ij} - \frac{1}{(1 + e^{\beta x})} \cdot e^{\beta x} x_{ij}.
\] (2.3)

Equation 2.3 is analogous to the normal equations from linear regression. Equation 2.3 is nonlinear in \( \beta_0, \ldots, \beta_{N+1} \) and so require numerical methods to solve, which is beyond the scope of this thesis. The Matlab function glmfit was used to find the estimates.

### 2.3 Ordinal Regression

Suppose \( Y \) can take on values \( b_1 \) to \( b_k \) which are ordered in some meaningful way. Let \( \pi_j(x) = P(Y = b_j|x) \). We wish to model and then estimate these \( \pi_j \) functions. We reduce this to logistic regression by dichotomizing the outcomes. When the outcomes are grades, it turns out to be useful to dichotomize and look at A against B-F; A,B against C,D,F; etc, and so reduce to dichotomous cases that can be handled with logistic regression. Unlike logistic regression where we compare the probabilities of \( Y = 1 \) versus \( Y = 0 \), the ordinal logistic regression model compares the probabilities of equal or smaller responses \( Y \leq k \) to the probabilities of larger responses \( Y > k \),

\[
\text{Logit}[P(Y \leq j)] = \ln \left( \frac{P(Y \leq j|x)}{P(Y > j|x)} \right) \\
= \ln \left( \frac{\pi_0(x) + \pi_1(x) + \cdots + \pi_j(x)}{\pi_{j+1}(x) + \cdots + \pi_k(x)} \right) \\
= \ln \left( \frac{\pi_j(x)}{1 - \pi_j(x)} \right) \\
= \alpha_j + \beta_1 x_1 + \cdots + \beta_n x_n
\]

Notice that this is very similar to the logistic model with two outcomes. One difference is that the intercept parameter is unique to its response level \( j \). That is, we model the
dependence of parameters on $j$ only through $\alpha_j$. Notice that $\beta_j$ does not have a subscript that is dependent upon these levels (naturally it depends on the predictor). This is because the model assumes an identical effect of $x$ on all $j - 1$ reductions of the response into binary outcomes. [10]

By looking at the log odds ratio it can be shown why this is the case. Given two values $x_1, x_2$ of $x_j$ and the estimated coefficient $\beta$ of $x_j$:

\[
\ln OR = \ln \left( \frac{P(Y \leq j | X_2)}{P(Y > j | X_2)} \right) - \ln \left( \frac{P(Y \leq j | X_1)}{P(Y > j | X_1)} \right)
\]

\[
= \text{Logit}[P(Y \leq j | X_2)] - \text{Logit}[P(Y \leq j | X_1)]
\]

\[
= \alpha_j + \beta x_2 - (\alpha_j + \beta x_1)
\]

\[
= \beta(x_2 - x_1).
\]

The log odds is used to interpret the model, and is the difference between the cumulative logits at those values of $x_j$. Thus $\beta$ is constant across all levels of reduction. This is what gives this type of ordinal regression model its name: proportional odds model. For each unit increase in $x_j$, the odds of a response being less than or equal to any given category is multiplied by $e^{\beta_j}$.
CHAPTER 3

Diagnostics

3.1 Hypothesis Testing

In the interest of being thorough, a brief explanation of hypothesis testing is included. A statistical hypothesis is a claim about the value of a population characteristic. When testing a hypothesis there are two contradictory hypotheses under consideration. The null hypothesis, $H_0$, is the claim assumed to be true. The alternative hypothesis, $H_a$, is the claim contradictory to $H_0$. The two possible outcomes are to reject $H_0$ or fail to reject $H_0$. Keep in mind then that rejecting $H_0$ does not mean accepting $H_a$. In order to test the hypotheses two other components are needed. A test statistic (T.S.) is a function of the sample data on which the decision is based. The rejection region (R.R.) is the set of all test statistic values for which $H_0$ will be rejected. Hypothesis testing is not an error free process. Type I error occurs when the null hypothesis is rejected when it is true. Type II error occurs when the null hypothesis is not rejected when it is false. These error probabilities are denoted $\alpha$ and $\beta$. $p$-values are often used when hypothesis testing. A $p$-value conveys information about the strength of evidence against $H_0$. The $p$-value is the smallest level of significance at which $H_0$ would be rejected. A $p$-value is the probability calculated, assuming $H_0$ is true, of obtaining a T.S. at least as contradictory to $H_0$ as the value that actually resulted. The
smaller the $p$-value the more contradictory is the data to $H_0$. It is customary to reject $H_0$ at level $\alpha$ if the $p$-value less than $\alpha$. In this case we call the data significant.

### 3.2 Variable Selection

It is important that the models use the correct variables. Usually this means seeking the most parsimonious model that still explains the data. In general the fewer variables used the lower the standard errors become. The more variables included in the model the more it relies on observational data. The inclusion of all variables will result in a better fit of the data, but in the loss of predictive ability.

Three different approaches were taken when selecting variables for each model. The first was to use ACTMAT and HSGPA to predict Grade. These two variables are widely used to measure students’ performance. The second was to use an intuitive approach to select variables. Having taught both Math 112 and 126 I chose variables I thought would represent what predictors would be useful in predicting Grade. The third was to use a stepwise selection algorithm. All major statistical software packages offer a stepwise option. Stepwise models are an efficient way to test variables. Stepwise methods essentially develop a sequence of regression models by adding or removing variables at each step. Variables are included/removed based on the statistical significance of their coefficient on the basis of a fixed decision rule. The statistic used in the decision rule is based on the model. In linear regression the errors follow a normal distribution so an $F$ test is used. In logistic/ordinal regression the errors are assumed to follow a binomial distribution so a chi-square test is used.

It is standard practice in most statistics textbooks or linear regression courses, to cover stepwise linear regression. Let me then explain the process for logistic regression while keeping in mind that analogous statements can be made for linear regression.
I have used the notation of Hosmer and Lemeshow for this explanation. Assume there are $n$ possible predictor variables. The algorithm begins by fitting the model using only the intercept and evaluating its log-likelihood, $L_0$. Then it fits the model for each of the $n$ variables individually and calculates their log-likelihoods, $L^{(0)}_j$ (superscripts refer to the step of the algorithm). Next the likelihood ratio test is used to compare $-2(L_0 - L^{(0)}_j)$ (the $-2$ is necessary so that the statistic follows a $\chi^2$ distribution). Call this value $G^{(0)}_j$ and its $p$-value $p^{(0)}_j$. This $p$-value is determined by $P[\chi^2(\nu) > G^{(0)}_j] = p^{(0)}_j$ where $\nu$ is the degrees of freedom.

The most significant variable is the one with the smallest $p$-value. Denote this variable, the one to be added, by $x_{e_1}$ and its $p$-value by $p^{(0)}_{e_1}$. Then $p^{(0)}_{e_1} = \min(p^{(0)}_j)$. Note that $x_{e_1}$ must have a reasonable $p$-value for it to be included in the model. Call this value $p_{IN}$. For example, if its $p$-value is 0.92 then we might consider another model since $x_{e_1}$ is not related to the response.

The next step begins like the last. The algorithm fits the model containing $x_{e_1}$ and calculates its log-likelihood, $L^{(1)}_{e_1}$. It then fits models containing $x_{e_1}$ and $x_j$ for $j = 1, \ldots, p$ with $p \neq e_1$.

Again let the log-likelihoods for these models be denoted $L^{(1)}_{e_1j}$, and the likelihood ratio chi-square statistic be $G^{(1)}_j = -2 \left(L^{(1)}_{e_1} - L^{(1)}_{e_1j}\right)$. The $p$-value is $p^{(1)}_j$. The variable evaluated for addition is $x_{e_2}$ and it is determined by $p^{(1)}_{e_2} = \left(\min p^{(1)}_{j}\right)$. If this value is less than $p_{IN}$ we include it in the model and continue.

This process is called forward selection. Another algorithm is called backward selection. It begins by including all predictors and eliminating them in a similar fashion. Lastly there are algorithms that combine both, where after a variable has been added it could be removed. To illustrate this we continue with the forward selection algorithm.

This step begins with the model containing $x_{e_1}$ and $X_{e_2}$. The algorithm will check whether $x_{e_1}$ is still important to the model. Using the likelihood ratio test we compare the log-likelihoods of the full model and the model with $x_{e_1}$ removed, $L^{(2)}_{-e_1}$, i.e. $G^{(2)}_{-e_1} = -2 \left(L^{(2)}_{-e_1} - L^{(2)}_{e_1e_2}\right)$ to obtain its $p$-value, $p^{(2)}_{-e_1}$. Variable $x_{e_1}$ would be removed if $p^{(2)}_{-e_1} > p_{OUT}$ where $p_{OUT}$ is a predetermined value.
The next step is similar to step 1 where we attempted to add variables. The algorithm ends when all variables have been included that satisfy $p_{IN}$ and excluded that exceed $p_{OUT}$.

### 3.3 Model Fit

After a statistical model has been fitted a goodness-of-fit analysis is important. Goodness-of-fit refers to how effectively the model describes the outcome variable. This is typically done by comparing the actual and predicted values of the response variable. Suppose we have observed outcome values $\mathbf{y}$ in vector form and denote $\hat{\mathbf{y}} = (y_1, y_2, \ldots, y_n)$ as the values predicted by the model. We conclude the model fits if the difference between $\mathbf{y}$ and $\hat{\mathbf{y}}$ is small for each pair $(y_i, \hat{y}_i)$ for each $i = 1, 2, \ldots, n$. To compare these vectors, tables have been created (Tables 3.1 and 3.2).

<table>
<thead>
<tr>
<th>KEY</th>
<th>Actual Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>Predicted</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>B</td>
</tr>
<tr>
<td>Grade</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>F</td>
</tr>
</tbody>
</table>

Table 3.1: Explanation of Actual versus Predicted Tables. Each cell represents the total number of students with the corresponding row/column definition. The value in the first row/first column cell is the number of students that earned an A and were predicted to earn an A. The value in the second row/fifth column cell is the students that earned an F, but were predicted to earn a B.

From this Table we see that the entries along the main diagonal are the values exactly predicted by the model. Also, we can compare prediction within one letter grade, i.e. a student with an actual grade of B and a predicted grade of A would be closely predicted. These Tables are used for Linear and Ordinal models based on their ordered five values responses.

Additionally, the 5x5 Table may be collapsed into a 2x2 Table so that a comparison may be
made to the logistic model. This is done by summing the total number of students in each block designated by the double line. This is explained in Table 3.2.

<table>
<thead>
<tr>
<th>KEY</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pass</td>
</tr>
<tr>
<td>Predicted</td>
<td>Pass</td>
</tr>
</tbody>
</table>

Table 3.2: Explanation of Actual versus Predicted Tables. Each cell represents the total number of student with the represented captions. The value in the first row/first column cell is the number of students that were predicted to pass and actually passed. The value in the first row, second column cell is the number of students predicted to pass, but that actually failed.

For linear regression $\hat{y}$ was a vector of continuous values ranging from 0 to 4. Each $\hat{y}_i$ was rounded to the nearest grade level so that it could be compared to the actual grade. Predicted grades with a decimal greater than or equal to 0.50 were rounded to the next highest grade and grades less than 0.50 were rounded to the next lowest grade (e.g. 3.51 became 4.0). For logistic regression, $\hat{y}$ was a vector of probabilities between 0 and 1. They were rounded to 1 if they were greater than or equal to 0.50, and rounded to 0 otherwise (e.g. 0.71 became 1). Lastly for ordinal regression $\hat{y}$ was a matrix of probabilities. For each student the probability of earning each grade was calculated. Then the grade with the highest probability was used for comparison against the actual value. For example, consider a student with probabilities: [0.23 0.42 0.21 0.03 0.11]. This student has a 23% chance of earning an A, a 42% chance of earning a B, etc. Clearly 0.42 is the highest percentage. Thus a predicted grade of B was recorded for this student. This is discussed further in the Ordinal Regression Results section.
CHAPTER 4

Results

All estimates presented in these sections were generated using Matlab 7.4.0. (R2007a). All code can be viewed in the Appendix.

Three models (Linear Regression, Logistic Regression, Ordinal Logistic Regression) were fitted to the Math 126 B data set. The first used ACTMAT and HSGPA as variables. The second, using the stepwise procedure (and SAS) found that HSGPA, Advanced, and ACTMAT were the variables best suited for the data. Finally, the third model used ACTMAT, ACTSCI, HSGPA, and Exponents.

4.1 Linear Regression Model

The results of the linear regression were

\[
\begin{align*}
\text{Grade} &= 2.314 + 0.2631 \text{ACTMAT} + 0.4996 \text{HSGPA} \\
\text{Grade} &= 1.6469 + 0.2230 \text{ACTMAT} + 0.4825 \text{HSGPA} + 0.2101 \text{Advanced} \\
\text{Grade} &= 2.2114 + 0.1926 \text{ACTMAT} + 0.1649 \text{ACTSCI} + 0.4926 \text{HSGPA} + 0.0178 \text{Exponents}
\end{align*}
\]
Since the variables were standardized we see that HSGPA had twice the effect on Grade than any other predictor. It was clearly the most important variable. Also we see that Exponents and ACTSCI had very little effect on Grade since they had low coefficients. This can also be seen in Figures 4.1 and 4.2. As the values of the predictors changes, the graphs are still very close, implying no or little effect on Grade.

The estimates for the predictors can be interpreted in the following way. Consider the first equation where $\hat{\beta}_{HSGPA} = 0.4996$. With all other predictors fixed, a one unit increase in HSGPA will result in a Grade approximately 0.5 higher. It is important keep in mind here that since the predictors were standardized a one unit increase in HSGPA was approximately a 0.40 grade point increase in HSGPA.

Using the figures the linear models showed the effects of individual predictors, while all others were fixed, on Grade. It was seen from Figure ?? that attaining a grade of A given an ACTMAT score of 28 required approximately a 4.40 HSGPA. In the second graph of the same Figure a grade of A was almost impossible, requiring ACTMAT = 35, HSGPA = 3.80.

### 4.2 Logistic Regression Model

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>p - value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.5677</td>
<td>0.2376</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>ACTMAT</td>
<td>0.5883</td>
<td>0.2289</td>
<td>0.0102</td>
</tr>
<tr>
<td>HSGPA</td>
<td>0.9556</td>
<td>0.2198</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

Table 4.1: Results of the logistic model for predictors ACTMAT, HSGPA

Table 4.1 showed the resulting estimates for the first logistic model, thus

$$Logit(\text{Grade}) = 1.5677 + 0.5883 \text{ACTMAT} + 0.9556 \text{HSGPA}.$$  

The estimates showed that HSGPA had a greater effect of the probability of passing than did ACTMAT because the coefficient was larger. Also both variables were statistically sig-
Figure 4.1: Linear model with predictors ACTMAT, HSGPA, ACTSCI, Exponents where functions (from right to left) correspond to ACTMAT = 16, 20, 24, 28 (top), and to Exponents = 3, 4, 5, 6 (bottom). We see that ACTMAT has a greater effect on predicted Grade than Exponents.
Figure 4.2: Linear model with predictors ACTMAT, HSGPA, ACTSCI, Exponents where functions (from right to left) correspond to HSGPA = 2.30, 2.80, 3.30, and 3.80 (top), and to ACTSCI = 14, 18, 22, 26 (bottom). We see that HSGPA has a greater effect on predicted Grade than ACTSCI.
significant at the $\alpha = 0.05$ level.

The predictors in Table 4.2 were listed in the order that the stepwise algorithm added them to the model. The model was given by

$$Logit(\text{Grade}) = 0.2604 + 0.5372 \text{ACTMAT} + 1.0125 \text{HSGPA} + 0.4470 \text{Advanced}.$$
<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>p - value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.2604</td>
<td>0.5310</td>
<td>0.6239</td>
</tr>
<tr>
<td>ACTMAT</td>
<td>0.5372</td>
<td>0.2371</td>
<td>0.0235</td>
</tr>
<tr>
<td>HSGPA</td>
<td>1.0125</td>
<td>0.2332</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Advanced</td>
<td>0.4470</td>
<td>0.1723</td>
<td>0.0095</td>
</tr>
</tbody>
</table>

Table 4.2: Results of the logistic model for predictors ACTMAT, HSGPA, Advanced subscore

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>p - value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.1536</td>
<td>1.0198</td>
<td>0.0347</td>
</tr>
<tr>
<td>ACTMAT</td>
<td>0.5323</td>
<td>0.2407</td>
<td>0.0270</td>
</tr>
<tr>
<td>HSGPA</td>
<td>0.9835</td>
<td>0.2237</td>
<td>0</td>
</tr>
<tr>
<td>ACTSCI</td>
<td>0.1957</td>
<td>0.2254</td>
<td>0.3851</td>
</tr>
<tr>
<td>Exponents</td>
<td>-0.0980</td>
<td>0.1708</td>
<td>0.5662</td>
</tr>
</tbody>
</table>

Table 4.3: Results of the logistic model for predictors ACTMAT, HSGPA, ACTSCI, Exponents subscore

It was seen then that the algorithm valued ACTMAT the highest, but the estimated coefficient for ACTMAT was still half as great as HSGPA. Again all predictors were significant at the $\alpha = 0.05$ level.

Table 4.3 shows the results of the third logistic model,

$$\text{Logit(Grade)} = 2.1536 + 0.5323 \text{ACTMAT} + 0.9835 \text{HSGPA} + 0.1957 \text{ACTSCI}$$

$$= -0.0980 \text{Exponents}.$$  

The regression coefficients were interpreted using the log odds ratio. Consider $\beta_{\text{ACTMAT}} = 0.5323$ from the three variable model. A one unit increase in ACTMAT (with all other predictors fixed) increased the log odds of success in Math 126 by

$$e^{\beta_{\text{ACTMAT}}} = e^{0.5323} \approx 1.70.$$  

As with the Linear model, adding ACTSCI and Exponents to the model did not increase the
predictive accuracy of the model. The coefficient of ACTSCI was not large enough to provide a significant effect upon the log odds and the Exponents subscore had an even lower effect. Additionally, based on their $p$–values these two variables were not statistically significant.

![Logistic model for predictors ACTMAT, HSGPA, ACTSCI, and Exponents.](image)

Figure 4.4: Logistic model for predictors ACTMAT, HSGPA, ACTSCI, and Exponents. The functions moving from right to left in the top graph correspond to ACTMAT = 16, 20, etc. The bottom graph is similar. Notice the graphs are very close signifying a small change in success probabilities when Exponents increases.

Similarly, Figure 4.4 and 4.5 showed the effects for individual predictor variables on the probability of a passing grade using a four variable model consisting ACTMAT, ACTSCI, HSGPA, and Exponents.
Figure 4.3 provided a way to determine effects of increasing ACTMAT by 4 pts and HSGPA by 0.5 pts. These increments were chosen based on the variable means from the Math 126 B dataset, $\text{ACTMAT} \approx 24$ and $\text{HSGPA} \approx 3.31$.

4.3 Ordinal Logistic Regression Model

Figure 4.6 illustrates the cumulative probabilities for the proportional odds model of a five category response with two predictor variables (one was fixed so that a two dimensional picture could be used). Note that the graphs increase from left to right. This was the effect of $\beta$ for each level, $b_j$. In particular, Figure 4.6 shows the $P(\text{Grade} \leq b_j|x = \text{ACTMAT}, \text{HSGPA} = 3.30)$. The functions correspond to the grade levels (right to left), i.e.

$$P(\text{Grade} = A|x = \text{ACTMAT}, \text{HSGPA} = 3.30)$$
$$P(\text{Grade} = A \text{ or } B|x = \text{ACTMAT}, \text{HSGPA} = 3.30)$$
$$P(\text{Grade} = A, B, \text{ or } C|x = \text{ACTMAT}, \text{HSGPA} = 3.30)$$
$$P(\text{Grade} = A, B, C, \text{ or } D|x = \text{ACTMAT}, \text{HSGPA} = 3.30)$$

The notation used here did not follow from the definition of the model. This was due to the coding used by Matlab. A Grade of 4 was assigned the first level, with a Grade of 3 the second, etc. Thus when calculating $P(Y \leq b_j|x)$ for level 1 (Grade = A): $P(Y \leq 1|x)$ means $P(\text{Grade} = A|x)$, $P(Y \leq 2|x)$ means $P(\text{Grade} = A \text{ or } B|x)$, etc.

The fitted parameters for the proportional odds ordinal regression model using the 126B
data with predictors ACTMAT and HSGPA were

$$\hat{\alpha}_1 = -1.9542$$
$$\hat{\alpha}_2 = 0.0077$$
$$\hat{\alpha}_3 = 1.4208$$
$$\hat{\alpha}_4 = 2.1135$$
$$\hat{\beta}_1 = 0.4351$$
$$\hat{\beta}_2 = 0.8849.$$  

We see that HSGPA was the most important since it has the largest coefficient and subsequently the largest impact on the log odds. For each $b_j$ with HSGPA fixed (as above) the probability of achieving a particular grade was calculated for a student with an ACTMAT score of 30. Recall that these variables were standardized so ACTMAT = 30 corresponded to a value of 2.4717 (HSGPA = 3.30 was 0.03442).

\[
P(\text{Grade} = A | \text{ACTMAT} = 30, \text{HSGPA} = 3.30) = \frac{e^{-1.9542 + 0.4351 \times 2.4717 + 0.8849 \times (-0.03442)}}{1 + e^{-1.9542 + 0.4351 \times 2.4717 + 0.8849 \times (-0.03442)}} \\ \approx 0.2873
\]

\[
P(\text{Grade} = A \text{ or } B | \text{ACTMAT} = 30, \text{HSGPA} = 3.30) = \frac{e^{0.0077 + 0.4351 \times 2.4717 + 0.8849 \times (-0.03442)}}{1 + e^{0.0077 + 0.4351 \times 2.4717 + 0.8849 \times (-0.03442)}} \\ \approx 0.7413
\]

\[
P(\text{Grade} = A, B, \text{ or } C | \text{ACTMAT} = 30, \text{HSGPA} = 3.30) = \frac{e^{1.4208 + 0.4351 \times 2.4717 + 0.8849 \times (-0.03442)}}{1 + e^{1.4208 + 0.4351 \times 2.4717 + 0.8849 \times (-0.03442)}} \\ \approx 0.9217
\]

\[
P(\text{Grade} = A, B, C, \text{ or } D | \text{ACTMAT} = 30, \text{HSGPA} = 3.30) = \frac{e^{2.1135 + 0.4351 \times 2.4717 + 0.8849 \times (-0.03442)}}{1 + e^{2.1135 + 0.4351 \times 2.4717 + 0.8849 \times (-0.03442)}}
\]
So for such a student, 

\[ P(y = 4| x) = 0.2872 \]
\[ P(y = 3| x) = 0.4541 \]
\[ P(y = 2| x) = 0.1804 \]
\[ P(y = 1| x) = 0.0375 \]
\[ P(y = 0| x) = 0.0408. \]
Figure 4.5: Logistic model for predictors ACTMAT, HSGPA, ACTSCI, and Exponents labeled as in prior figure. ACTSCI (in the lower graph) behaves much like Exponents in the previous figure.
Figure 4.6: $P(Y \leq b_j | x)$ where ACTMAT = 24 (top), $P(Y \leq b_j | x)$ where HSGPA = 3.30 (bottom). Moving from right to left the functions are $P(\text{Grade} = A | \text{HSGPA}, \text{ACTMAT})$, $P(\text{Grade} = A \text{ or } B | \text{HSGPA}, \text{ACTMAT})$, $P(\text{Grade} = A, B, \text{ or } C | \text{HSGPA}, \text{ACTMAT})$, $P(\text{Grade} = A, B, C, \text{ or } D | \text{HSGPA}, \text{ACTMAT})$. 
CHAPTER 5

Conclusions

The following are tables containing actual versus predicted values for Grade for the purpose of model comparison. The table indicates the subset of the data chosen, placement exam, and course. It also states the model and variables in the model.

In terms of exact predictive ability none of the models greatly succeeded. However, this was most likely due to the nonlinearity and extreme variability of the data. Consider Table 5.1. The entries along the main diagonal, (6, 36, 7, 0, 7), were the correctly predicted students’ Grade. If exact prediction determined the accuracy of the model then the ordinal model failed terribly, i.e. just $\frac{56}{172} \approx 33\%$ of students’ Grade was predicted exactly. But model accuracy can be relaxed to include values adjacent to the main diagonal. These Grades were closely predicted. In other words look for occurrences when the model prediction was within one letter grade (higher or lower) of the actual value. The accuracy jumped to 80% when

<table>
<thead>
<tr>
<th></th>
<th>Ordinal</th>
<th>Linear</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>36</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5.1: Results of Predicted vs. Actual values of grade for variables ACTMAT, HSGPA
Table 5.2: Results of Predicted vs. Actual values of grade for variables ACTMAT, HSGPA, Advanced

<table>
<thead>
<tr>
<th>Math 126 B</th>
<th>Ordinal</th>
<th>Linear</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 5 0 0 0</td>
<td>3 1 0 0 0</td>
<td></td>
</tr>
<tr>
<td>16 35 28 3 5</td>
<td>19 33 12 1 0</td>
<td></td>
</tr>
<tr>
<td>5 16 13 10 12</td>
<td>6 22 29 12 16</td>
<td></td>
</tr>
<tr>
<td>0 0 0 0 0</td>
<td>1 3 4 2 8</td>
<td></td>
</tr>
<tr>
<td>1 3 4 2 7</td>
<td>0 0 0 0 0</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.3: Results of Predicted vs. Actual values of grade for variables ACTMAT, ACTSCI, HSGPA, Exponents

<table>
<thead>
<tr>
<th>Math 126 B</th>
<th>Ordinal</th>
<th>Linear</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 6 1 0 0</td>
<td>3 1 0 0 0</td>
<td></td>
</tr>
<tr>
<td>17 35 29 5 3</td>
<td>20 28 11 1 1</td>
<td></td>
</tr>
<tr>
<td>5 13 11 8 16</td>
<td>5 25 31 12 18</td>
<td></td>
</tr>
<tr>
<td>0 0 0 0 0</td>
<td>1 5 5 2 5</td>
<td></td>
</tr>
<tr>
<td>1 5 4 2 5</td>
<td>0 0 0 0 0</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4: Pass/Fail Results for predictors ACTMAT, HSGPA

<table>
<thead>
<tr>
<th>Math 126 B</th>
<th>Ordinal</th>
<th>Logistic</th>
<th>Linear</th>
</tr>
</thead>
<tbody>
<tr>
<td>121 30</td>
<td>121 31</td>
<td>121 32</td>
<td></td>
</tr>
<tr>
<td>12 9</td>
<td>12 8</td>
<td>12 7</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.5: Pass/Fail Results for predictors ACTMAT, HSGPA, Advanced

<table>
<thead>
<tr>
<th>Math 126 B</th>
<th>Ordinal</th>
<th>Logistic</th>
<th>Linear</th>
</tr>
</thead>
<tbody>
<tr>
<td>125 30</td>
<td>125 26</td>
<td>125 29</td>
<td></td>
</tr>
<tr>
<td>8 9</td>
<td>8 13</td>
<td>8 10</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.6: Pass/Fail Results for predictors ACTMAT, ACTSCI, HSGPA, Exponents
predictive exactness was relaxed. Table 5.7 shows the predictive accuracy of each model. This justifies that statement that the models poorly predict Grade exactly, i.e. the exact percentages were all lower than 39%. However, they do predict within one letter grade (higher or lower) well. All of these percentages were approximately 75%. Table 5.7 also shows that the three variable model performed the best, or the stepwise procedure increased predictive ability for each model. The main point is that all three models performed similarly. There was no standout.

<table>
<thead>
<tr>
<th>Model</th>
<th>Variables</th>
<th>% Exact</th>
<th>% Within One</th>
<th>Pass/Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>2</td>
<td>34</td>
<td>81</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>39</td>
<td>84</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>37</td>
<td>82</td>
<td>76</td>
</tr>
<tr>
<td>Logistic</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-</td>
<td>-</td>
<td>76</td>
</tr>
<tr>
<td>Ordinal</td>
<td>2</td>
<td>33</td>
<td>80</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>36</td>
<td>84</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>31</td>
<td>77</td>
<td>76</td>
</tr>
</tbody>
</table>

Table 5.7: Results (percentages) for each model in terms of prediction accuracy. The Variables column informs the reader of the predictors used in the model. They are the same predictors as in the Results sections. Exact refers to the percentage of Grades that were perfectly predicted. Within One refers to the percentage of Grades predicted to be within one letter grade (higher or lower) of the actual grade. Finally Pass/Fail refers to the collapsing of the 5x5 Table so as to compare to the Logistic model.
CHAPTER 6

Further Study

It was obvious that as the models were fit that other important predictors were missing. From the literature possibilities for these variables included gender, major, age, race, relationship status. Also as empirical power of technology increases so does the popularity of neural networks. Neural networks kept recurring when mathematics placement was mentioned. The question of placement is ongoing and shall not end while students thirst for knowledge.
Appendix A:

A.1 Placement Reference

Explanation of the Mathematics Placement Exam outcomes:
http://www.bgsu.edu/departments/math/page52942.html

A.2 BGSU Mathematics Courses

<table>
<thead>
<tr>
<th>Course</th>
<th>Subject Matter</th>
<th>Credits</th>
<th>Placement Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>MATH 090</td>
<td>Elementary Algebra</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>MATH 095</td>
<td>Intermediate Algebra</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>MATH 112</td>
<td>College Algebra I</td>
<td>3</td>
<td>24 or 27</td>
</tr>
<tr>
<td>MATH 115</td>
<td>Introduction to Statistics</td>
<td>3</td>
<td>N/A</td>
</tr>
<tr>
<td>MATH 122</td>
<td>College Algebra II</td>
<td>3</td>
<td>32</td>
</tr>
<tr>
<td>MATH 126</td>
<td>Basic Calculus</td>
<td>5</td>
<td>42</td>
</tr>
<tr>
<td>MATH 128</td>
<td>Precalculus Mathematics</td>
<td>5</td>
<td>41</td>
</tr>
<tr>
<td>MATH 130</td>
<td>Precalculus Mathematics</td>
<td>3</td>
<td>51</td>
</tr>
<tr>
<td>MATH 131</td>
<td>Calculus and Analytic Geometry</td>
<td>5</td>
<td>61</td>
</tr>
<tr>
<td>MATH 134</td>
<td>Calculus and Analytic Geometry IA</td>
<td>3</td>
<td>71</td>
</tr>
</tbody>
</table>
## A.3 Subscores

<table>
<thead>
<tr>
<th>Subscore (Exam A)</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fractions</td>
<td>1 2 10 14 15 10</td>
</tr>
<tr>
<td>Percents</td>
<td>3 7 16 17 18</td>
</tr>
<tr>
<td>Ratios</td>
<td>5 6 9 27 35</td>
</tr>
<tr>
<td>Addition/Subtraction</td>
<td>4 11 12 13 31</td>
</tr>
<tr>
<td>Solving</td>
<td>19 21 22 28 32 33 34</td>
</tr>
<tr>
<td>Advanced</td>
<td>8 20 23 24 25 26 29</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subscore (Exam B)</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fractions</td>
<td>1 4 5 7 9 14 15</td>
</tr>
<tr>
<td>Factoring</td>
<td>2 8 10 22 27</td>
</tr>
<tr>
<td>Exponents</td>
<td>3 11 13 20 23 24 31</td>
</tr>
<tr>
<td>Solving</td>
<td>6 18 19 21 26 30 32</td>
</tr>
<tr>
<td>Graphing</td>
<td>12 16 17 25</td>
</tr>
<tr>
<td>Advanced</td>
<td>28 29 33 34 35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subscore (Exam C)</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algebra 1</td>
<td>1 7 10 13 17 21</td>
</tr>
<tr>
<td>Algebra 2</td>
<td>3 8 12 22</td>
</tr>
<tr>
<td>Trigonometry 1</td>
<td>23 24 25 28 31 32 33</td>
</tr>
<tr>
<td>Trigonometry 2</td>
<td>26 27 29 30 34 35</td>
</tr>
<tr>
<td>Graphing</td>
<td>2 5 6 15 16 19</td>
</tr>
<tr>
<td>Area</td>
<td>4 11 14 18 20</td>
</tr>
</tbody>
</table>
A.4 MatLab Code

I would be happy to provide any of the functions used in this thesis upon request. Please contact me at jbarrow@bgsu.edu or barrow4007@gmail.com.
BIBLIOGRAPHY


