

Entrance Exam Placement Analysis

Kris Kissel

April 20, 2015

This is an analysis of student performance in math courses at Green River College based on whether they were placed into those courses by the Entrance Exams developed by faculty at the school. The primary purpose of the analysis is to advise the mathematics faculty of the college on whether the placement by Entrance Exams is working well in comparison to other placement mechanisms.

1 Abstract

The Mathematics Division at Green River College has put in place several options for entering students trying to find their appropriate placement into math classes. All incoming students take the computerized COMPASS placement test, a commercial tool employed by many schools, to identify a student's initial placement. Students who believe that they have been placed below their true ability level by this assessment then have several options to try to improve their placement: If they have taken classes from another college, they can transfer in their credits; if they have recently graduated from one of several local school districts in the GRCC vicinity, they can ask the college to evaluate their high school transcript to identify the appropriate placement; and if a student knows exactly which math course he or she wishes to take, the student may elect to take a computerized Entrance Exam designed locally by the college faculty to test mastery of prerequisite material for that course. Entrance Exams are available for all developmental (i.e. precollege-level) math classes, all college-level math classes that do not have a college-level prerequisite, as well as second-quarter precalculus and first-quarter calculus. This study examines the performance of students placed by Entrance Exams in comparison to those who are placed by other methods.

2 Summary of Results

In every course analyzed, the success rates for students (as defined in the next section) was either greater for students who were placed by the locally-designed Entrance Exams compared to the success rate for students placed by other means, or the success rate was lower but the difference was too small to be considered statistically significant. Students who gained entry to math courses via the Entrance Exams did appear to have stronger overall academic performances, as measured by their concurrent GPA in the same quarter, suggesting that because the Entrance Exams are optional, they might be attracting students who tend to have stronger academic abilities than the general population. This implies that these success rates might not extend to the entire student population if the Entrance Exams were to become the primary tool for placement. Nonetheless, the strong success rates indicate that the Entrance Exams are working as desired. Additionally, there is some evidence to suggest that the cutoff scores for passing the Entrance Exams might even be lowered somewhat without sacrificing success rates.

3 Analysis

There are a number of ways to try to define student success in math classes. For the purposes of this analysis, we mostly (see the exceptions described in the next paragraph) define a student's performance as successful if he or she earns the grade necessary to take the following course in the sequence. For example, a grade

of 2.5 (on a 4.0-scale) in Elementary Algebra is the prerequisite for enrollment in Intermediate Algebra, so we categorize an Elementary Algebra student as successful if he or she earns a grade of 2.5 or higher. The success rate for a group of students in a particular course is then defined to be the percentage of the group who are successful.

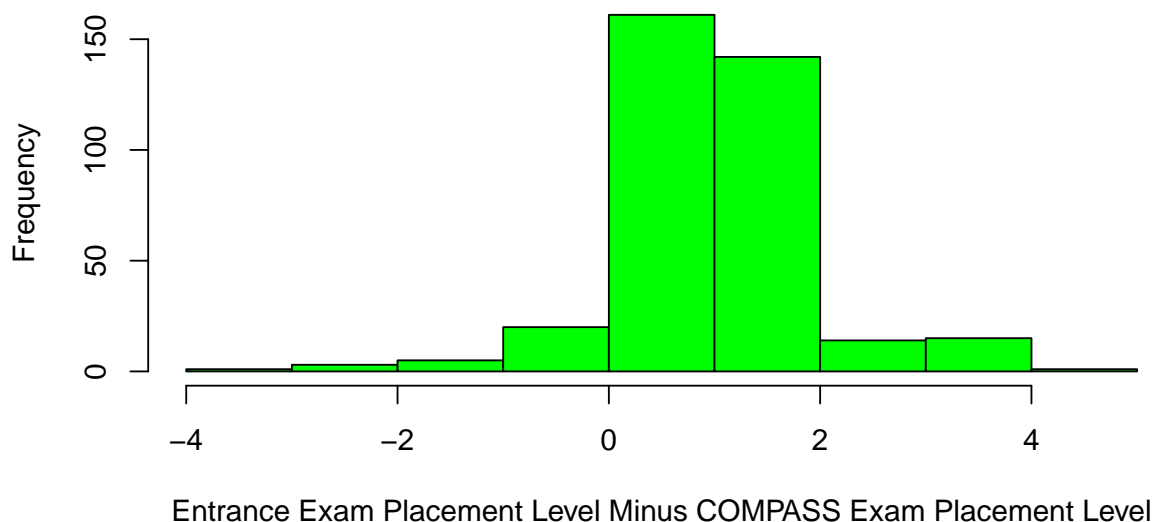
There are, however, a few complications with this definition of student success. One of the courses in the study, Math In Society, is a terminal course and is not a prerequisite for any other course; for this course, student success is defined as meeting a grade of 1.0 or better, since that is sufficient to earn college credit for the course and thereby satisfy a mathematics requirement for an Associate in Arts degree. Similarly, although Precalculus 1 and Finite Math for Business and Social Science are prerequisites for other classes, many students take these as terminal courses, intending only to earn credit and satisfy a mathematics requirement for their degree; because different subsets of students in these courses have different goals, we choose to report both the success rate at the 1.0-level (for earning college credit) and at the 2.0-level (for entrance into the next course in the sequence). Intermediate Algebra is the final developmental-level mathematics course offered at the college, and after this course there are several pathways for college-level mathematics courses from which students may choose: Precalculus 1, Number Theory for Elementary School Teachers, Finite Math for Business and Social Science, and Math in Society. Furthermore, the prerequisite grade for admission into the next course depends which sequence a student selects: the prerequisite grade in Intermediate Algebra for entrance into Precalculus 1 or Number Theory for Elementary School Teachers is 2.5, while a grade of 2.0 suffices for the other tracks. So, as with some of the other courses, we report here success rates for the two relevant cutoff grades, 2.0 and 2.5.

Students who withdrew from a course, earned an Incomplete or No-Credit grade were not considered successful; students who earned a Pass grade were only considered successful at the 1.0-grade-point level.

In this study, we examine student grades earned in the 14 academic quarters beginning Summer 2010 and ending Fall 2013 (these are all the quarters for which data was available when this study was initiated). We examined all grades earned during this period in the math classes under consideration, for a total of 25417 student grades attained by 12617 individual students. The data was obtained from the Washington State Board for Community and Technical Colleges data warehouse, via the GRC office of Institutional Excellence. (Important Disclosure: The author of this study is a faculty member in the GRC Mathematics Division but was not involved in the development of the Entrance Exams.)

Approximately 2.9% of the 12617 students in this study obtained math course placement via the college's custom Entrance Exams. Of those students, the mean number of course levels skipped by taking and passing an Entrance Exam instead of settling for their COMPASS placement was 1.48 levels. One student even managed to skip 5 levels. This histogram illustrates the difference in the number of course levels between students' Entrance Exam placement and his or her COMPASS placement. Notice that there were a few students for whom the difference was negative; because students can use the higher of the two placement recommendations, no student was actually placed lower as a result of the Entrance Exam.

Difference in Placement Levels

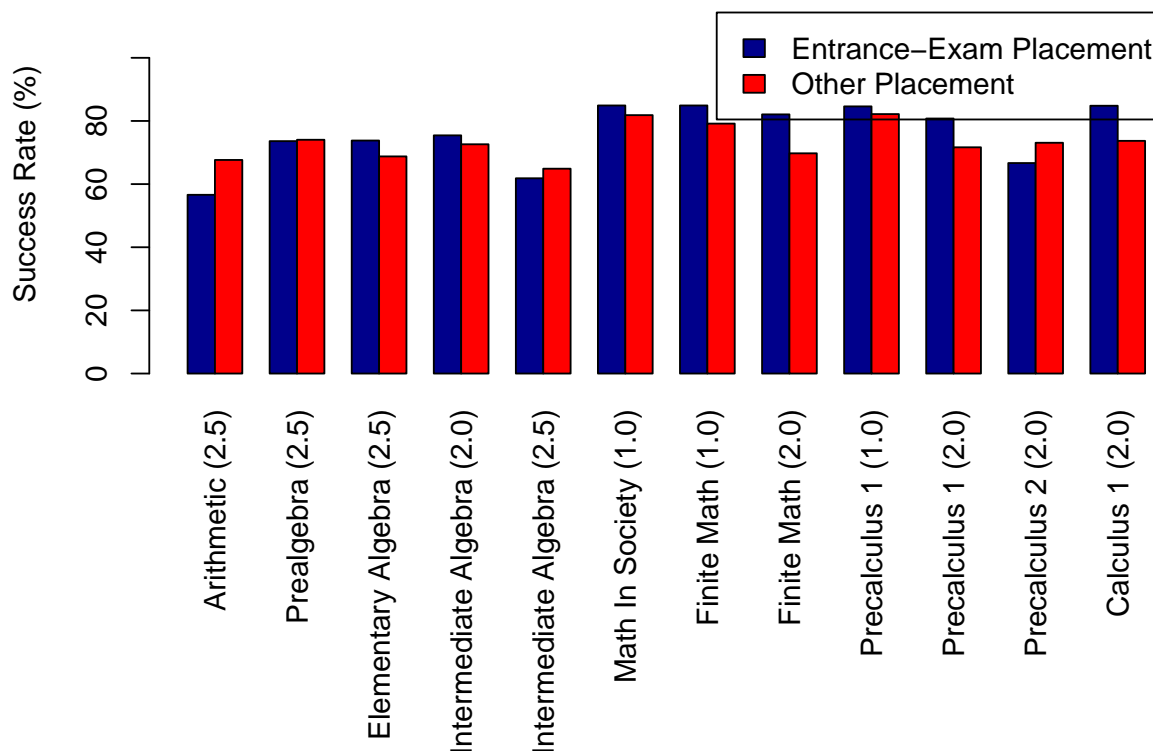


The following table lists the major results of the study. **LEGEND:** The first column of each line indicates the course name and the relevant cutoff grade (on a 4.0 scale); EESR is the success rate of students placed by the Entrance Exams, reported as a percent; OSR is the success rate of other students; #EE is the number of students attempting placement by the corresponding Entrance Exam, and #O is the number of students placed by other means. The listed P-value for each class corresponds to an exact binomial test applied to the counts of successful and unsuccessful students placed by Entrance Exams; we use the success rates of students in the corresponding course placed by other means as the parameter for the binomial distribution.

##	EESR (%)	OSR (%)	#EE	#O	P-Value
## Arithmetic (2.5)	56.60377	67.64855	53	1969	0.105
## Prealgebra (2.5)	73.61111	74.04103	72	2242	0.894
## Elementary Algebra (2.5)	73.79310	68.76813	145	5171	0.210
## Intermediate Algebra (2.0)	75.43860	72.61443	228	5156	0.373
## Intermediate Algebra (2.5)	61.84211	64.87587	228	5156	0.332
## Math In Society (1.0)	84.90566	81.84438	106	1041	0.528
## Finite Math (1.0)	84.90566	79.16236	106	1934	0.187
## Finite Math (2.0)	82.07547	69.75181	106	1934	0.006
## Precalculus 1 (1.0)	84.61538	82.17231	78	3517	0.659
## Precalculus 1 (2.0)	80.76923	71.65198	78	3517	0.079
## Precalculus 2 (2.0)	66.66667	73.08869	15	1962	0.566
## Calculus 1 (2.0)	84.81675	73.69104	191	2311	0.000

This graph illustrates the success rates listed in the table above.

Comparison of Students Success Rates by Placement Method



The only statistically significant difference in success rates occurred in the Calculus 1 course, and the success rate there was *higher* for students placed by into that course by an Entrance Exams compared to those placed by other means. (The p-value seen in the Finite Math group should not be considered significant because of the multiple-testing effect. If we use a standard significance level of $\alpha = 0.05$ for this study of 12 hypotheses, then the Bonferroni correction for multiple testing suggests that the cutoff value for significance in each individual comparison should be $\frac{0.05}{12} \approx 0.004$.)

Taken together, these results suggest that across all levels, the Entrance Exams do no worse in placing the population of students who elect to take them than other placement mechanisms available at the college.

4 Grade Distributions

An important question to ask here is whether the population of students who opt to take an Entrance Exam is somehow fundamentally different from the population of students who don't. That is to say, if the success rates differ, are there confounding variables that could account for this? For example, is it possible that students who chose to take an extra test happen to have stonger academic skills to begin with? We explore several aspects of the available data to try to answer this question.

The following subsections includes graphs that illustrate the distributions of grades for students placed by Entrance Exams and those placed by other means. In each histogram, the far-left bar represents the combined tally of the non-numerical grades earned: W (withdrawal), I (Incomplete), NC (No Credit) and N

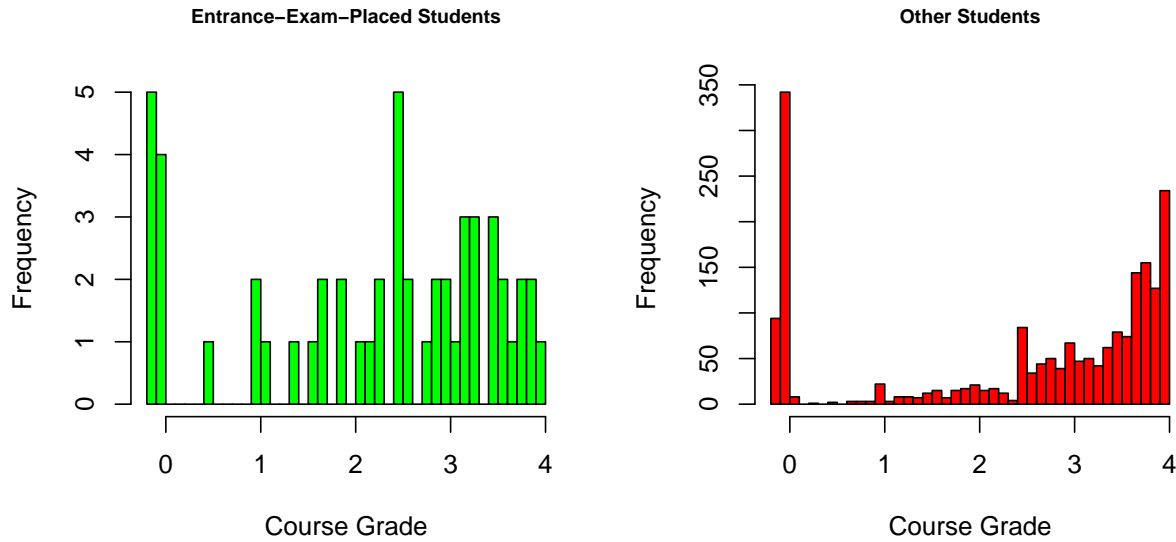
(audit); the non-numerical grades P (Pass) are included in the tally with numerical grades of 1.0.

We also examine students' concurrent GPA. It would seem ideal if we could look only at their GPA in other (non-math) classes, since the overall GPA in a given quarter is of course correlated with the performance in the math class taken at that time; however, it is not possible to extract a grade-point average in their other classes from the information available, because there is no way to determine how many credits we should be dividing by in order to compute the average. We would need to know the actual grades earned in the other courses. For example, it makes a difference if a student withdrew from another class rather than earning a 0.0 in it, because that affects the process of the grade-point calculation (credits corresponding to classes from which a student withdraws are not used in computing the average); however that information is not available in this data set. We will therefore be satisfied with examining the students' overall concurrent GPA while enrolled in a math class, including the math grades.

It is worth noting that, in every course, the mean concurrent GPA was greater among the Entrance-Exam placed students than for other students. In several of these courses, the difference was statistically significant. This suggests that there may be a difference in overall academic abilities between students who opt to take an Entrance Exam and those who don't, a possible confounding variable that makes it difficult to predict what would happen with the success rates if the current system of using Entrance Exams as an optional secondary mechanism for placement were to be substantially altered in the future.

4.1 MATH062 - Arithmetic

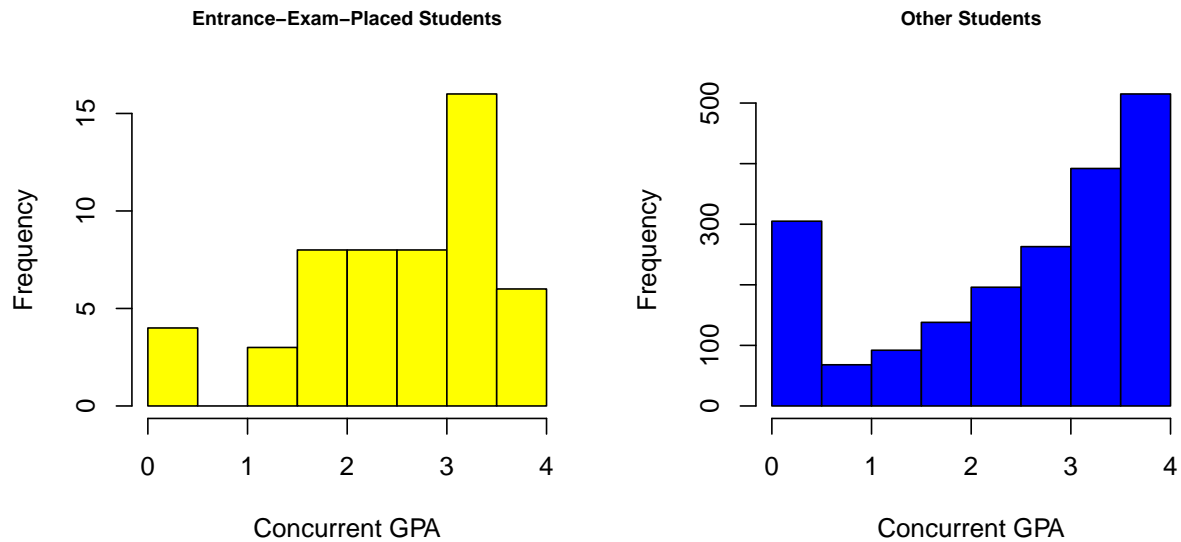
Grades Earned in MATH062



Student Success Rates at the 2.5 Grade-Point Level (%) by Entrance Exam score:

##		80	85	90	95	99
##	Unsuccessful	50.0	39.1	40.0	100.0	0.0
##	Successful	50.0	60.9	60.0	0.0	100.0

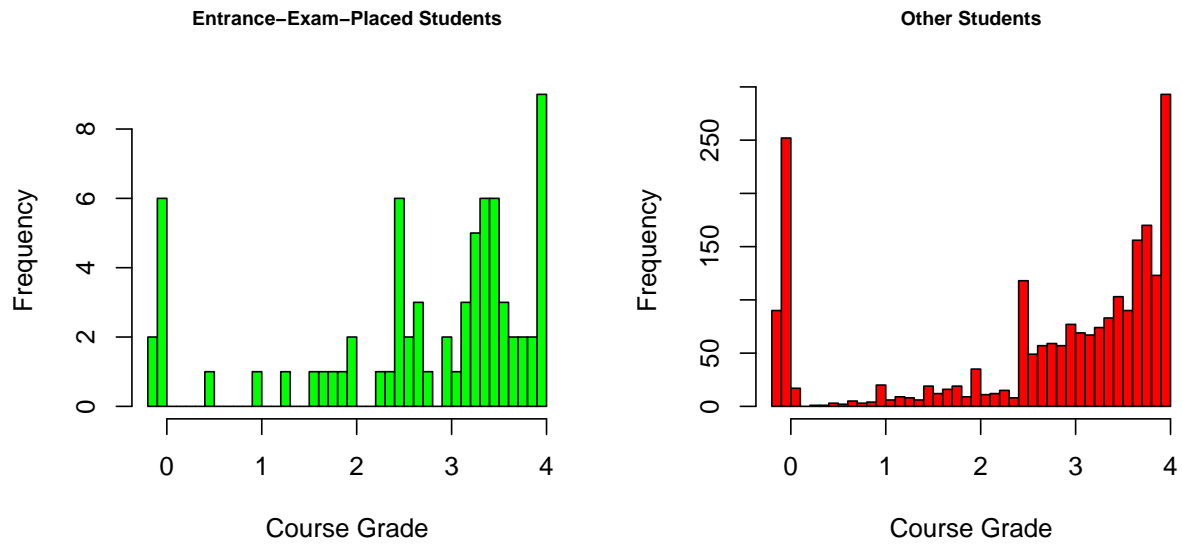
Concurrent GPA for Students in MATH062



The mean concurrent GPA of students who were placed by Entrance Exam is 2.49, and the mean concurrent GPA of students placed by other means is 2.47. A T-test of the null hypothesis that the mean concurrent GPA is the same for both groups produces a P-value of 0.846.

4.2 MATH070 - Pre-Algebra

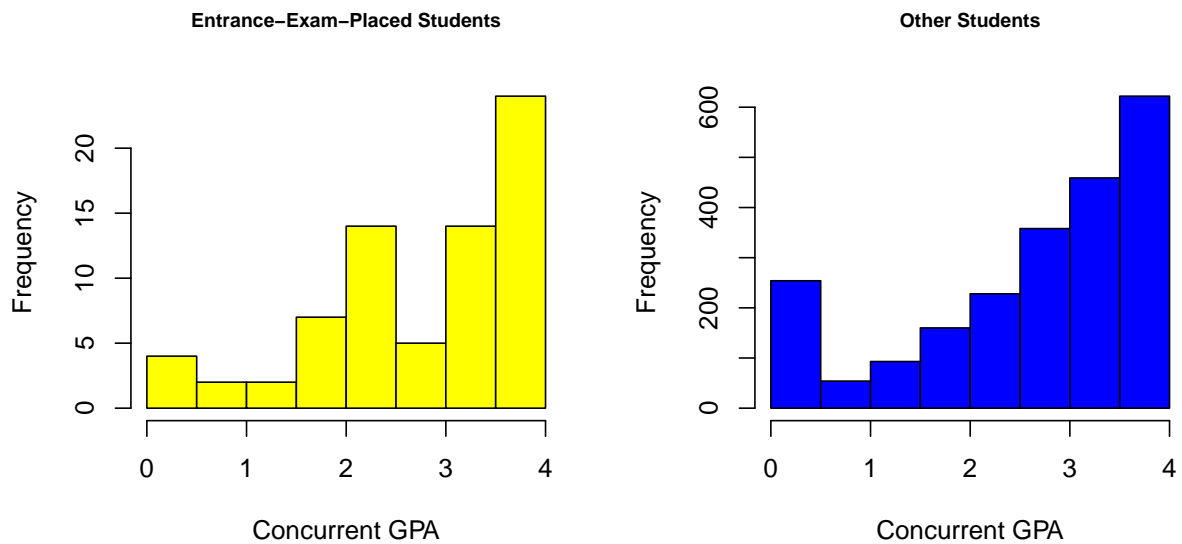
Grades Earned in MATH070



Student Success Rates at the 2.5 Grade-Point Level (%) by Entrance Exam Score:

##		180	185	190	195	199
##	Unsuccessful	37.8	100.0	4.3	0.0	42.9
##	Successful	62.2	0.0	95.7	100.0	57.1

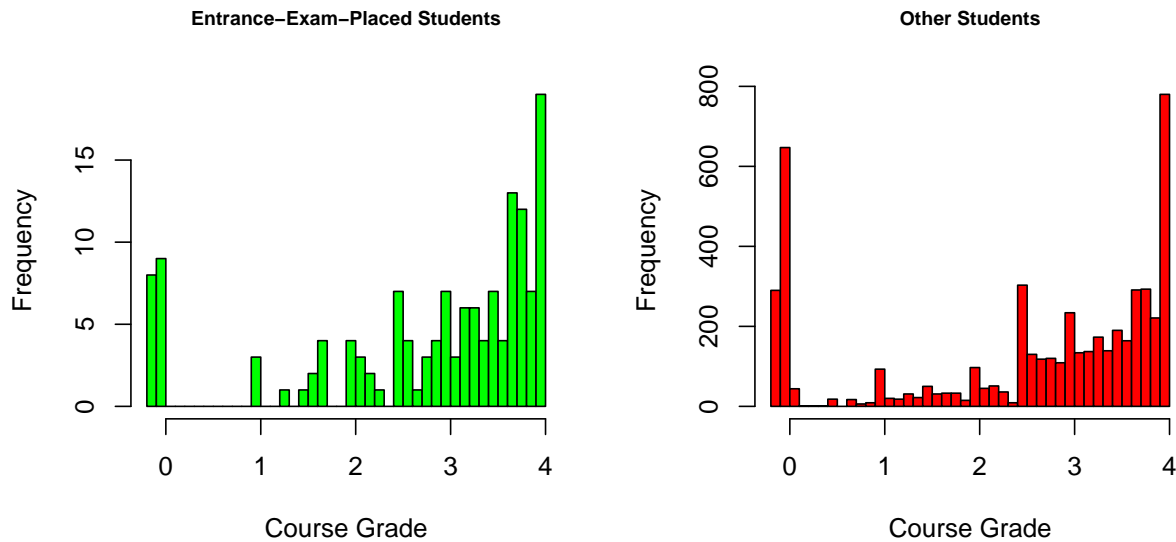
Concurrent GPA for Students in MATH070



The mean concurrent GPA of students who were placed by Entrance Exam is 2.79, and the mean concurrent GPA of students placed by other means is 2.62. A T-test of the null hypothesis that the mean concurrent GPA is the same for both groups produces a P-value of 0.201.

4.3 MATH072 - Elementary Algebra

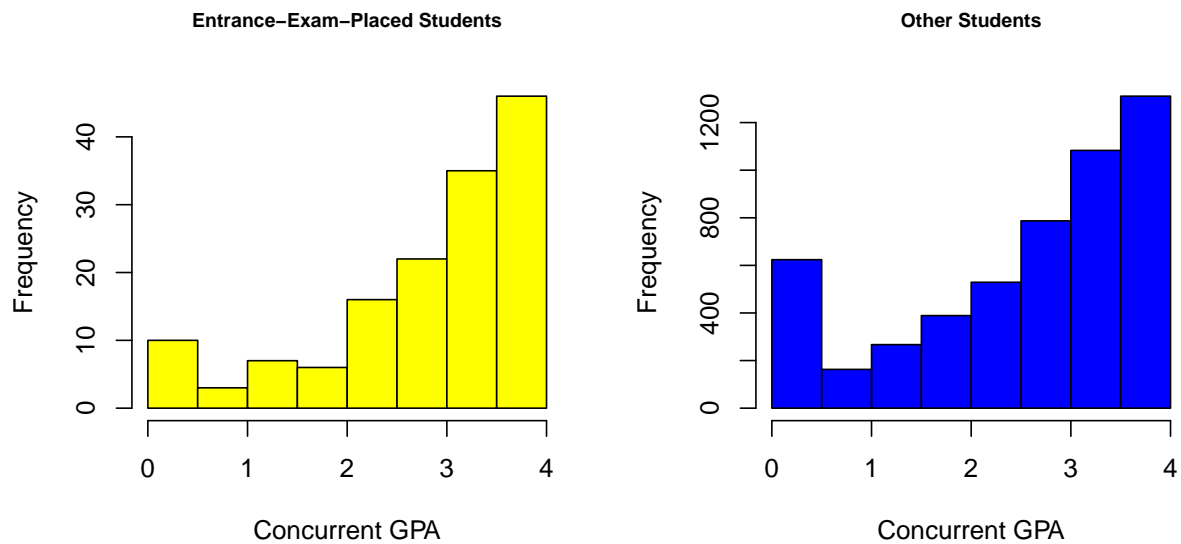
Grades Earned in MATH072



Student Success Rates at the 2.5 Grade-Point Level (%) by Entrance Exam Score:

##		280	285	290	295	299
##	Unsuccessful	20.6	12.5	28.6	37.5	15.4
##	Successful	79.4	87.5	71.4	62.5	84.6

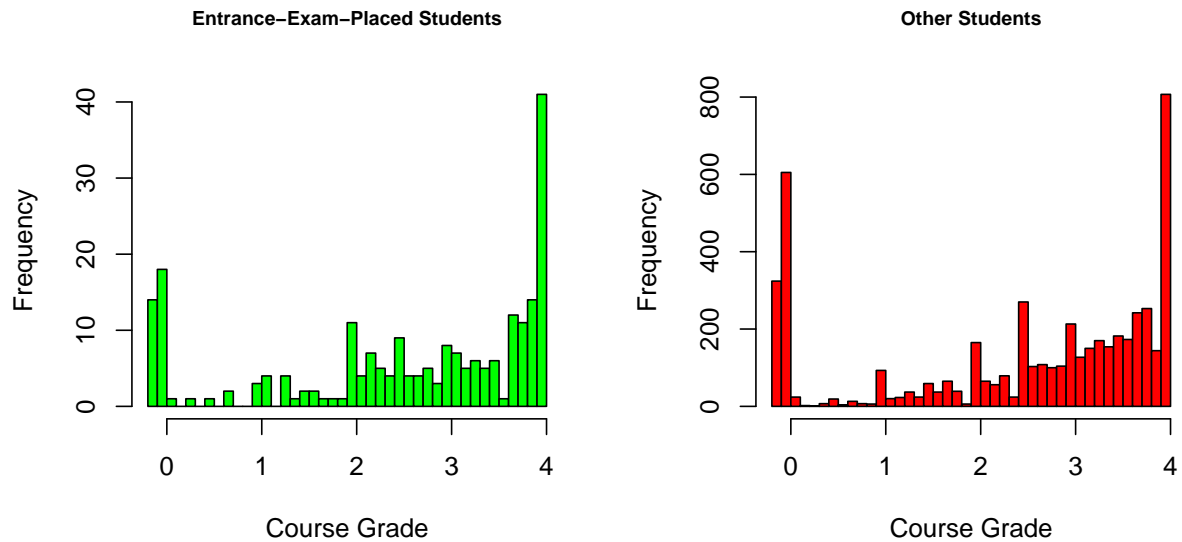
Concurrent GPA for Students in MATH072



The mean concurrent GPA of students who were placed by Entrance Exam is 2.85, and the mean concurrent GPA of students placed by other means is 2.55. A T-test of the null hypothesis that the mean concurrent GPA is the same for both groups produces a P-value of 0.001.

4.4 MATH097 - Intermediate Algebra

Grades Earned in MATH097



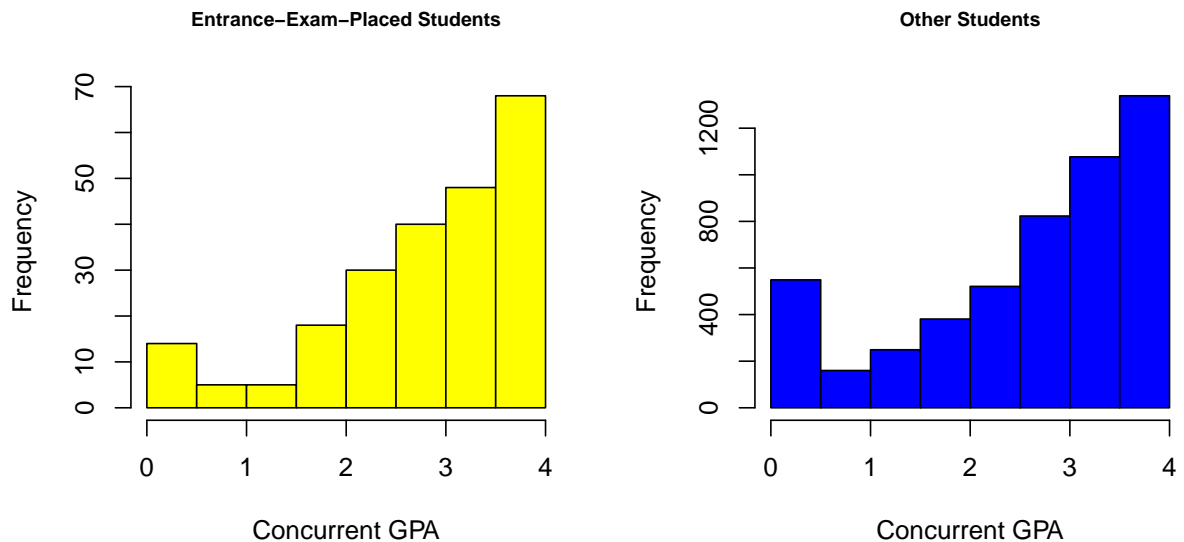
Student Success Rates at the 2.5 Grade-Point Level (%) by Entrance Exam Score:

##		380	382	383	385	388	390	395	398	399
##	Unsuccessful	43.5	33.3	0.0	46.7	100.0	26.8	32.5	50.0	0.0
##	Successful	56.5	66.7	100.0	53.3	0.0	73.2	67.5	50.0	100.0

Student Success Rates at the 2.0 Grade-Point Level (%) by Entrance Exam Score:

##		380	382	383	385	388	390	395	398	399
##	Unsuccessful	25.8	33.3	0.0	33.3	100.0	14.3	20.0	50.0	0.0
##	Successful	74.2	66.7	100.0	66.7	0.0	85.7	80.0	50.0	100.0

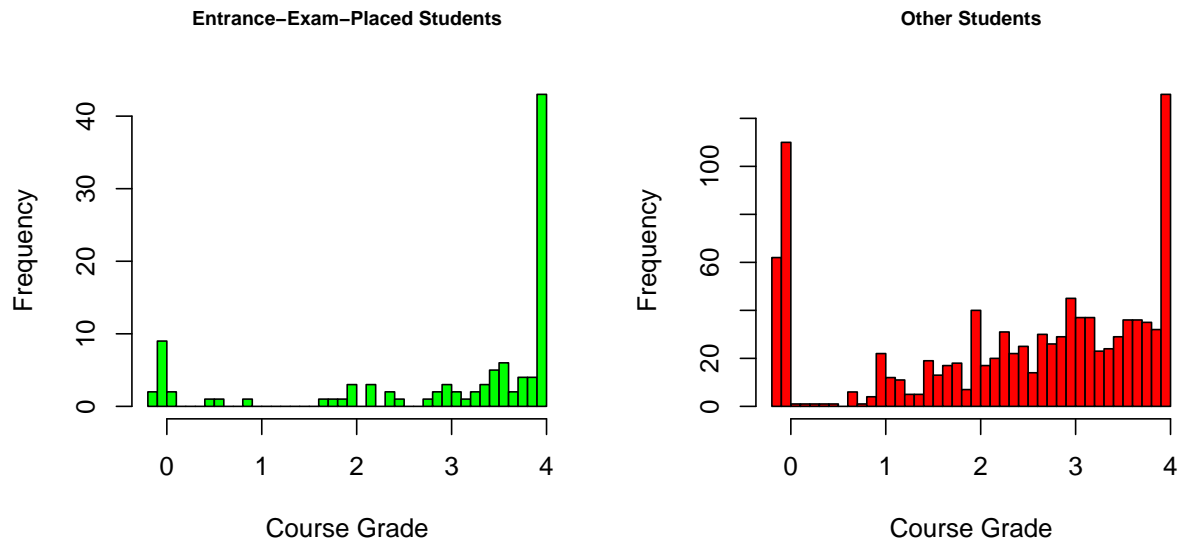
Concurrent GPA for Students in MATH097



The mean concurrent GPA of students who were placed by Entrance Exam is 2.82, and the mean concurrent GPA of students placed by other means is 2.6. A T-test of the null hypothesis that the mean concurrent GPA is the same for both groups produces a P-value of 0.003.

4.5 MATH&107 - Math in Society

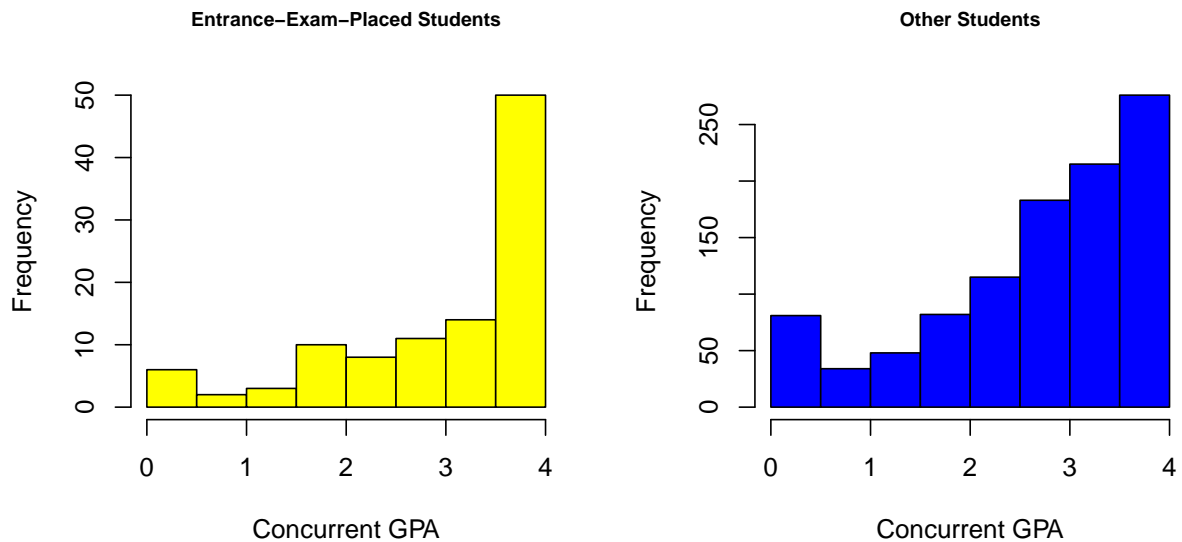
Grades Earned in MATH&107



Student Success Rates at the 1.0 Grade-Point Level (%) by Entrance Exam Score:

##											
##		475	476	478	480	481	482	483	485	486	488
##	Unsuccessful	16.7	0.0	14.3	13.3	0.0	0.0	18.2	8.3	0.0	40.0
##	Successful	83.3	100.0	85.7	86.7	100.0	100.0	81.8	91.7	100.0	60.0
##											
##		490	495								
##	Unsuccessful	16.7	25.0								
##	Successful	83.3	75.0								

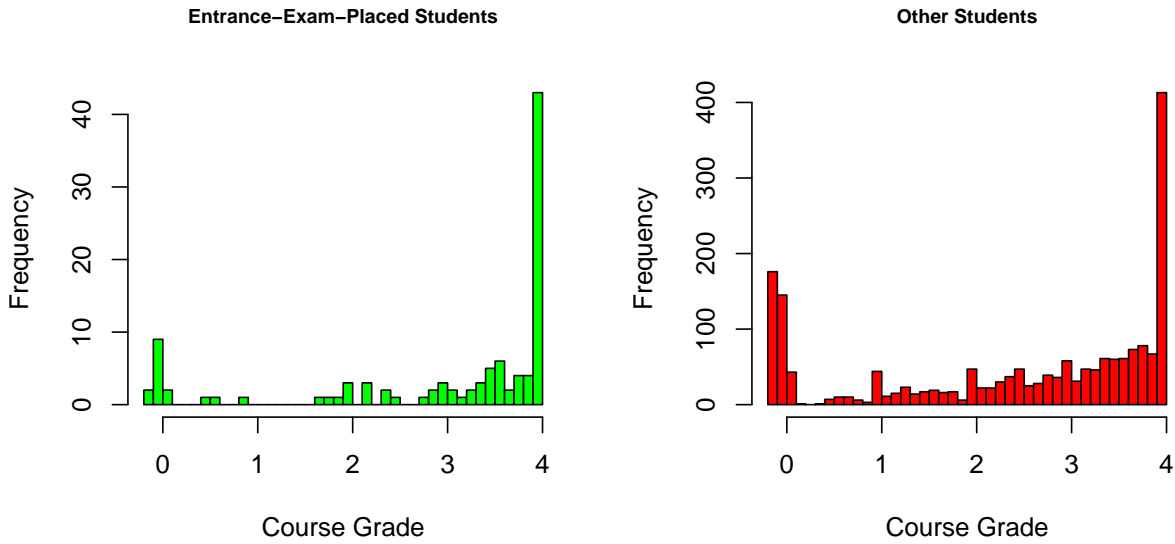
Concurrent GPA for Students in MATH&107



The mean concurrent GPA of students who were placed by Entrance Exam is 2.99, and the mean concurrent GPA of students placed by other means is 2.68. A T-test of the null hypothesis that the mean concurrent GPA is the same for both groups produces a P-value of 0.007.

4.6 MATH147 - Finite Math for Business and Social Science

Grades Earned in MATH147



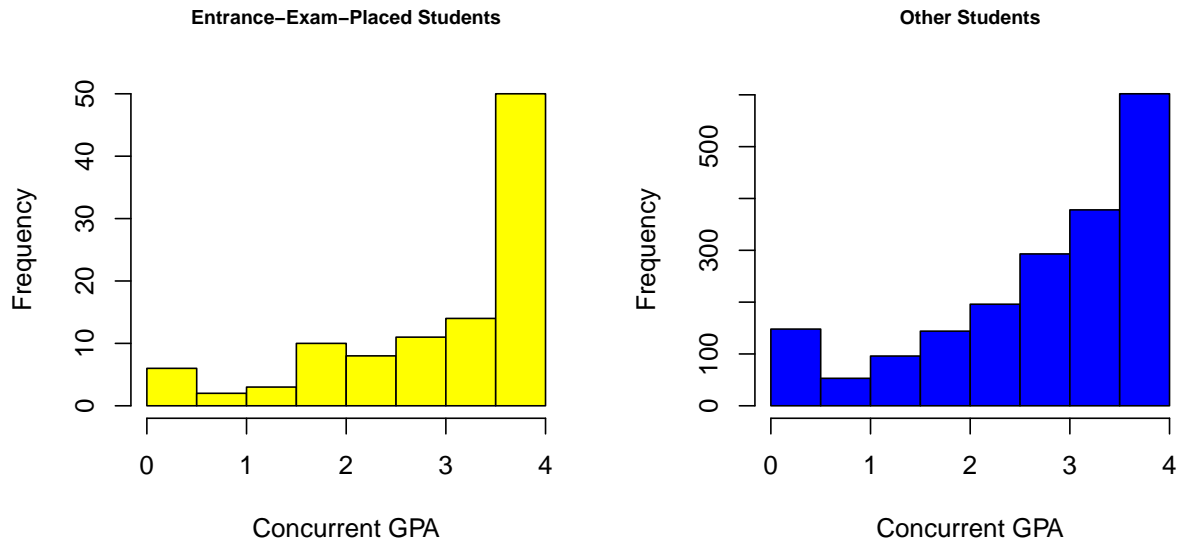
Student Success Rates at the 1.0 Grade-Point Level (%) by Entrance Exam Score:

##											
##		475	476	478	480	481	482	483	485	486	488
##	Unsuccessful	16.7	0.0	14.3	13.3	0.0	0.0	18.2	8.3	0.0	40.0
##	Successful	83.3	100.0	85.7	86.7	100.0	100.0	81.8	91.7	100.0	60.0
##											
##		490	495								
##	Unsuccessful	16.7	25.0								
##	Successful	83.3	75.0								

Student Success Rates at the 2.0 Grade-Point Level (%) by Entrance Exam Score:

##											
##		475	476	478	480	481	482	483	485	486	488
##	Unsuccessful	16.7	0.0	14.3	20.0	50.0	0.0	18.2	8.3	0.0	40.0
##	Successful	83.3	100.0	85.7	80.0	50.0	100.0	81.8	91.7	100.0	60.0
##											
##		490	495								
##	Unsuccessful	16.7	25.0								
##	Successful	83.3	75.0								

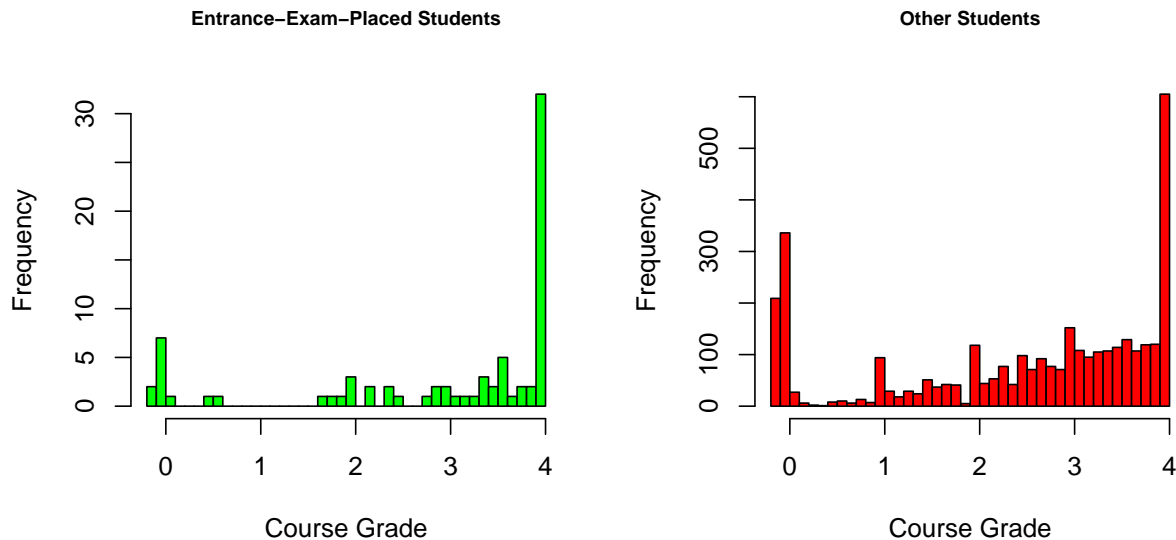
Concurrent GPA for Students in MATH147



The mean concurrent GPA of students who were placed by Entrance Exam is 2.99, and the mean concurrent GPA of students placed by other means is 2.74. A T-test of the null hypothesis that the mean concurrent GPA is the same for both groups produces a P-value of 0.028.

4.7 MATH&141 - Precalculus 1

Grades Earned in MATH&141



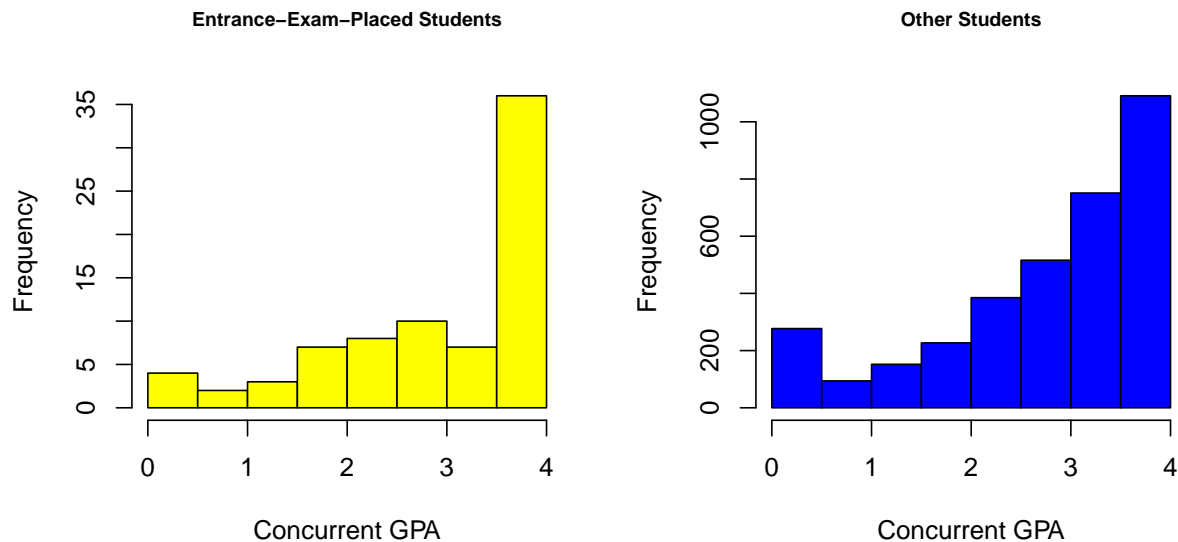
Student Success Rates at the 1.0 Grade-Point Level (%) by Entrance Exam Score:

##		480	481	482	483	485	486	488	490	495
##	Unsuccessful	13.3	0.0	0.0	18.2	8.3	0.0	40.0	16.7	25.0
##	Successful	86.7	100.0	100.0	81.8	91.7	100.0	60.0	83.3	75.0

Student Success Rates at the 2.0 Grade-Point Level (%) by Entrance Exam Score:

##		480	481	482	483	485	486	488	490	495
##	Unsuccessful	20.0	50.0	0.0	18.2	8.3	0.0	40.0	16.7	25.0
##	Successful	80.0	50.0	100.0	81.8	91.7	100.0	60.0	83.3	75.0

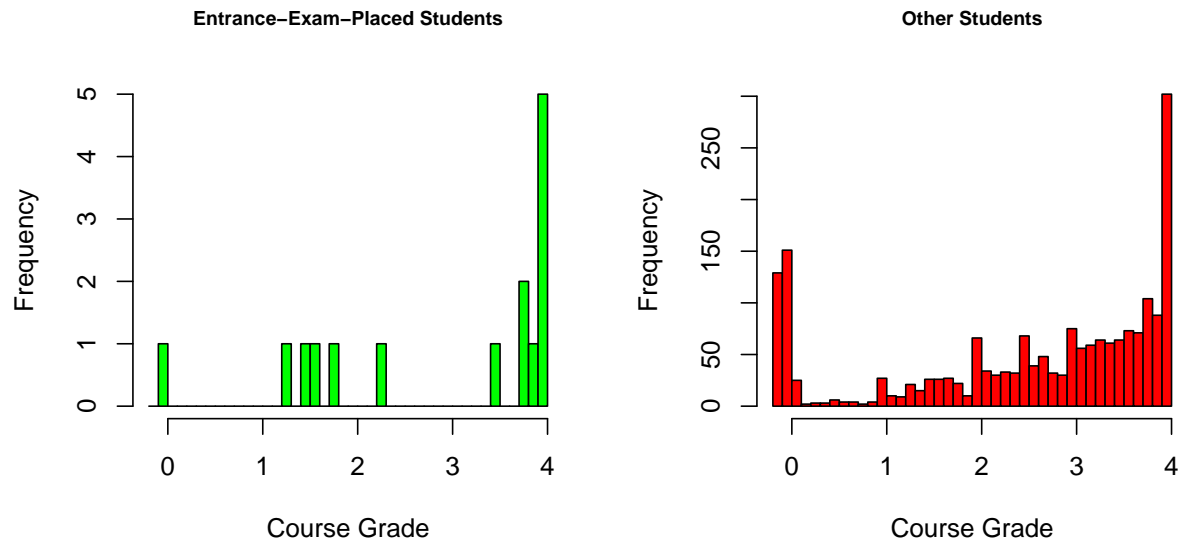
Concurrent GPA for Students in MATH&141



The mean concurrent GPA of students who were placed by Entrance Exam is 2.93, and the mean concurrent GPA of students placed by other means is 2.75. A T-test of the null hypothesis that the mean concurrent GPA is the same for both groups produces a P-value of 0.17.

4.8 MATH&142 - Precalculus 2

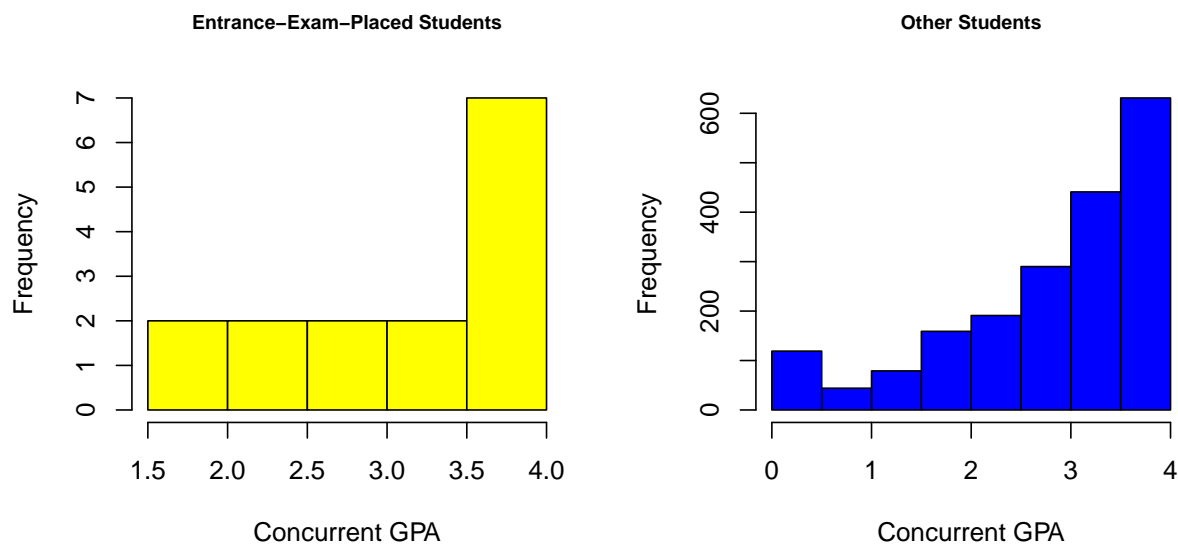
Grades Earned in MATH&142



Student Success Rates at the 2.0 Grade-Point Level (%) by Entrance Exam Score:

##					
##		575	577	578	580
##	Unsuccessful	0.0	0.0	50.0	57.1
##	Successful	100.0	100.0	50.0	42.9

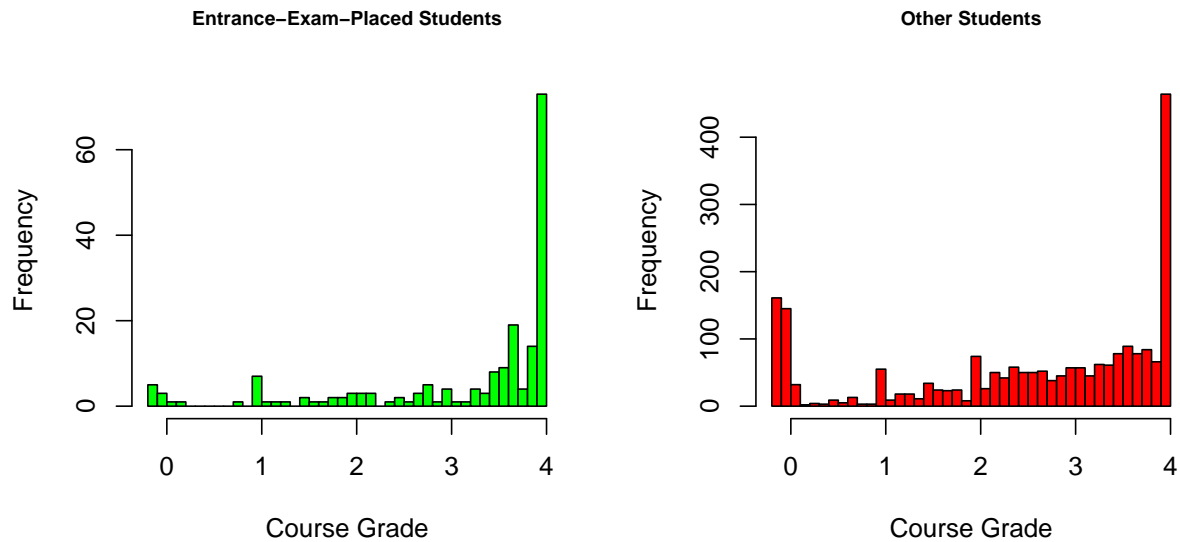
Concurrent GPA for Students in MATH&142



The mean concurrent GPA of students who were placed by Entrance Exam is 3.13, and the mean concurrent GPA of students placed by other means is 2.82. A T-test of the null hypothesis that the mean concurrent GPA is the same for both groups produces a P-value of 0.17.

4.9 MATH&151 - Calculus 1

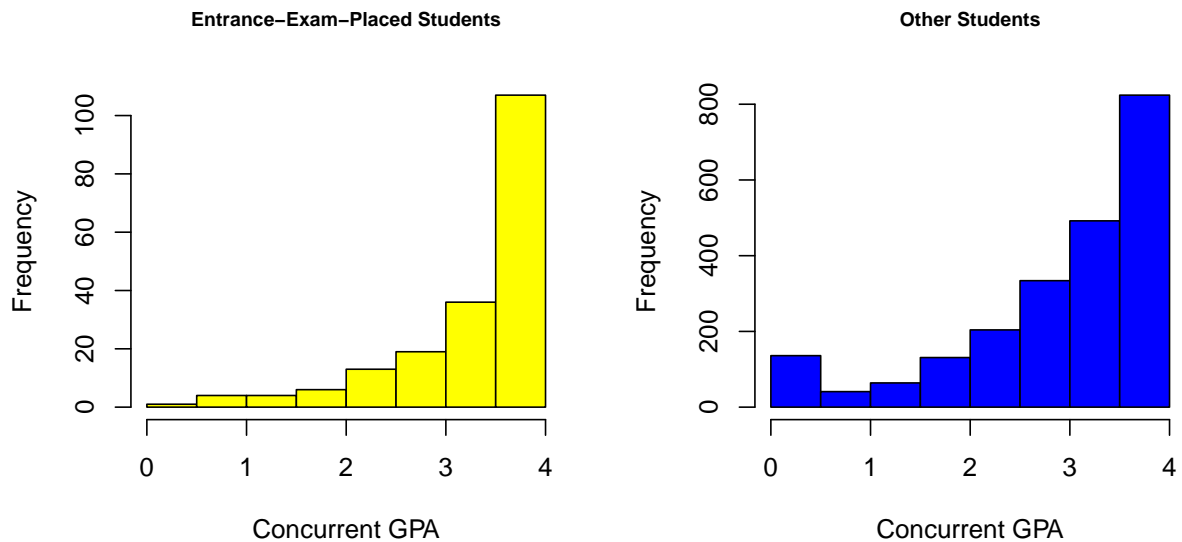
Grades Earned in MATH&151



Student Success Rates at the 2.0 Grade-Point Level (%) by Entrance Exam Score:

##											
##		675	676	677	678	679	680	682	683	685	686
##	Unsuccessful	21.3	0.0	12.5	0.0	0.0	16.7	0.0	0.0	13.6	0.0
##	Successful	78.7	100.0	87.5	100.0	100.0	83.3	100.0	100.0	86.4	100.0
##											
##		687	688	689	690	692	694	695	697		
##	Unsuccessful	0.0	0.0	33.3	21.4	0.0	25.0	12.5	0.0		
##	Successful	100.0	100.0	66.7	78.6	100.0	75.0	87.5	100.0		

Concurrent GPA for Students in MATH&151



The mean concurrent GPA of students who were placed by Entrance Exam is 3.31, and the mean concurrent GPA of students placed by other means is 2.92. A T-test of the null hypothesis that the mean concurrent GPA is the same for both groups produces a P-value of 0.

5 Cutoff Scores

Because the success rates for students placed by the Entrance Exams was often higher than that for students placed by other means, a reasonable question to ask is whether those students who narrowly missed the cutoffs might have been successful as well; if so, then it might be possible to lower the cutoff scores, thus shortening the sequence of math classes necessary for students to complete their educational programs.

This is a complicated question to answer because we don't know exactly how students would have performed in that circumstance.

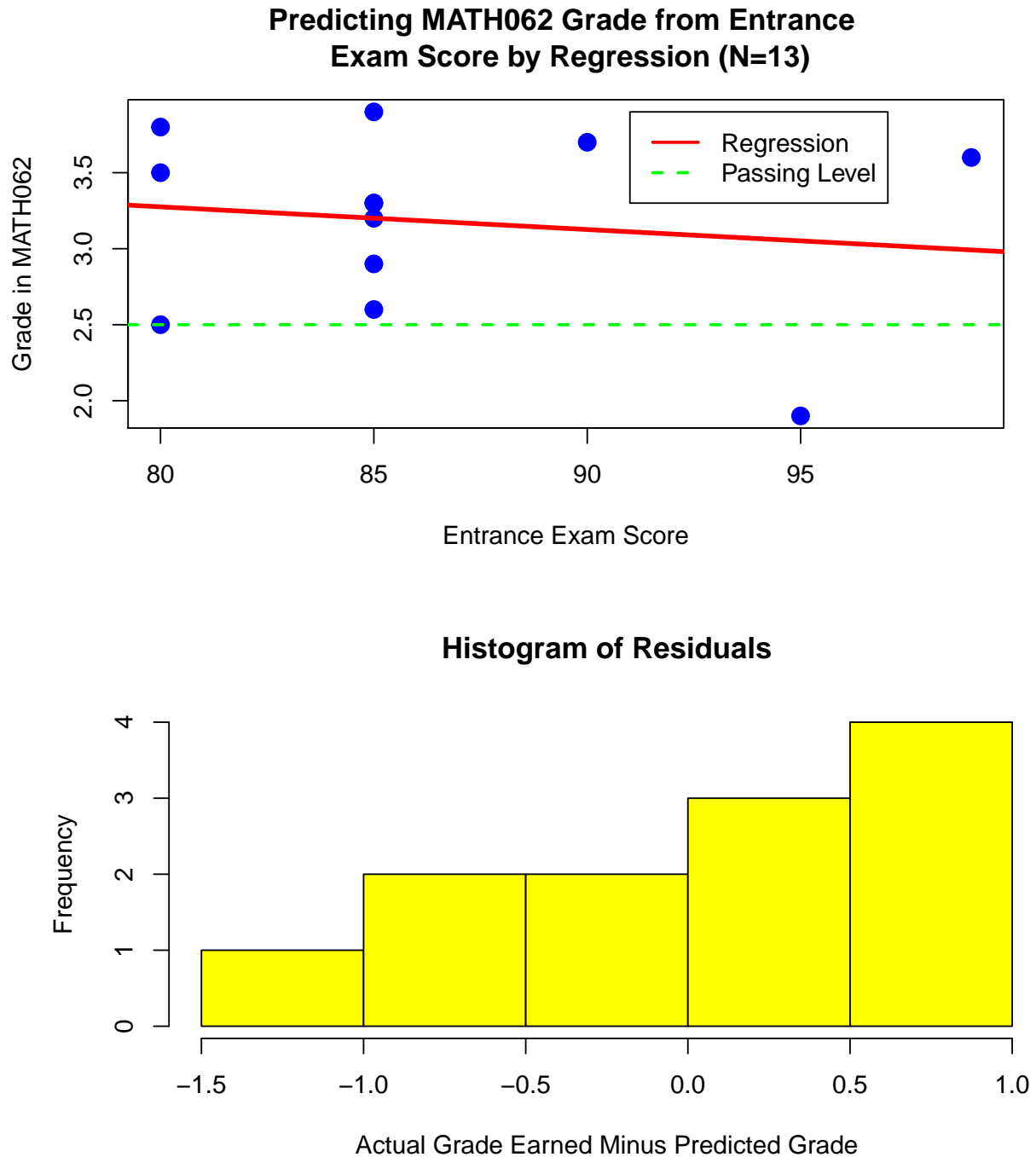
A promising line of reasoning is to use linear regression to try to predict the grades earned by students as a function of their Entrance Exam scores. The plan then would be to estimate from the regression whether students achieving a score slightly lower than the current cutoff score would be projected to earn a grade that is passing.

This approach would be absolutely acceptable if we knew that the grade earned is well-correlated with the Entrance Exam score, but as is visible in the scatter plots included here, that is not typically the case. Indeed, Entrance Exams are not designed to be predictive of the grade a student will earn. Rather, they are meant to indicate whether a student has sufficient comfort with the background material to have a reasonable chance of success in a course. In fact, in some of the regressions performed here, the correlation between Entrance Exam score and grade earned surprisingly ends up being negative. (Perhaps unsurprisingly, those negative correlations show up when the sample sizes are very small.)

On the other hand, even though the linear regression is not a perfect predictor of the grade a student will earn, our data shows that, in all of the courses studied, it is a reasonable conservative estimate. We included histograms of the residuals for the regression models to illustrate this. If the linear regression were a good predictor of the grade, we would expect to see these residuals be distributed normally, and they are not. However, in most of the courses, the majority of students performed better than the regression would have predicted from their Entrance Exam score alone.

Based on these results, it seems likely that the Green River College mathematics division could lower the cutoff scores slightly (by 5 points, an amount that permits one additional incorrect exam answer) for passing each of the Entrance Exams without substantial adverse effects on the success rates for students.

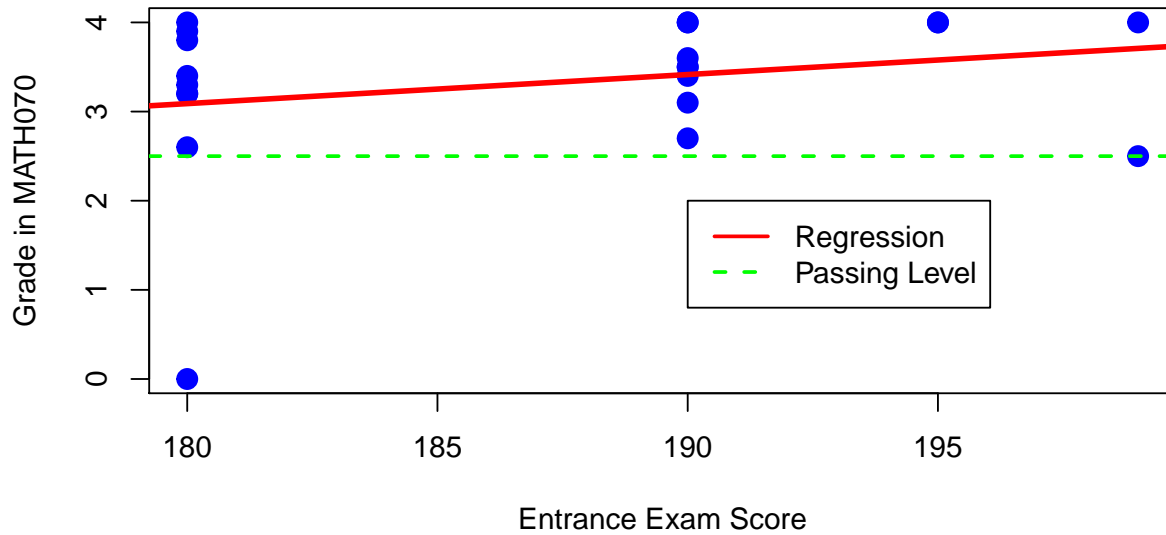
5.1 MATH062 - Arithmetic



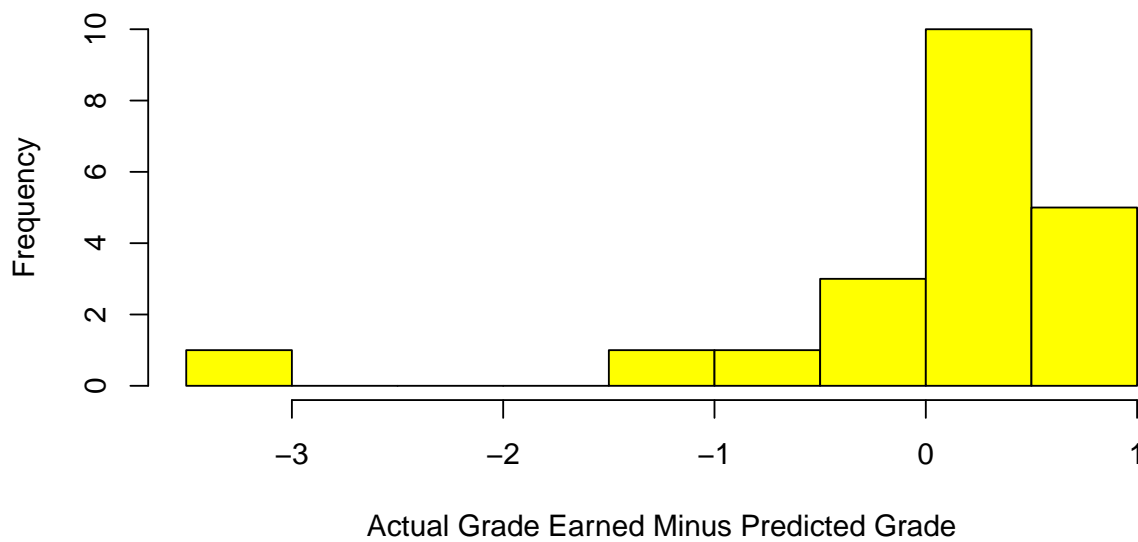
The regression line predicts a grade of 3.3 for a student earning an Entrance Exam score that is 5 points lower than the current cutoff. The negative slope of this line is likely due to the small sample size and the presence of a single outlier on the bottom right.

5.2 MATH070 - Prealgebra

Predicting MATH070 Grade from Entrance Exam Score by Regression (N=21)

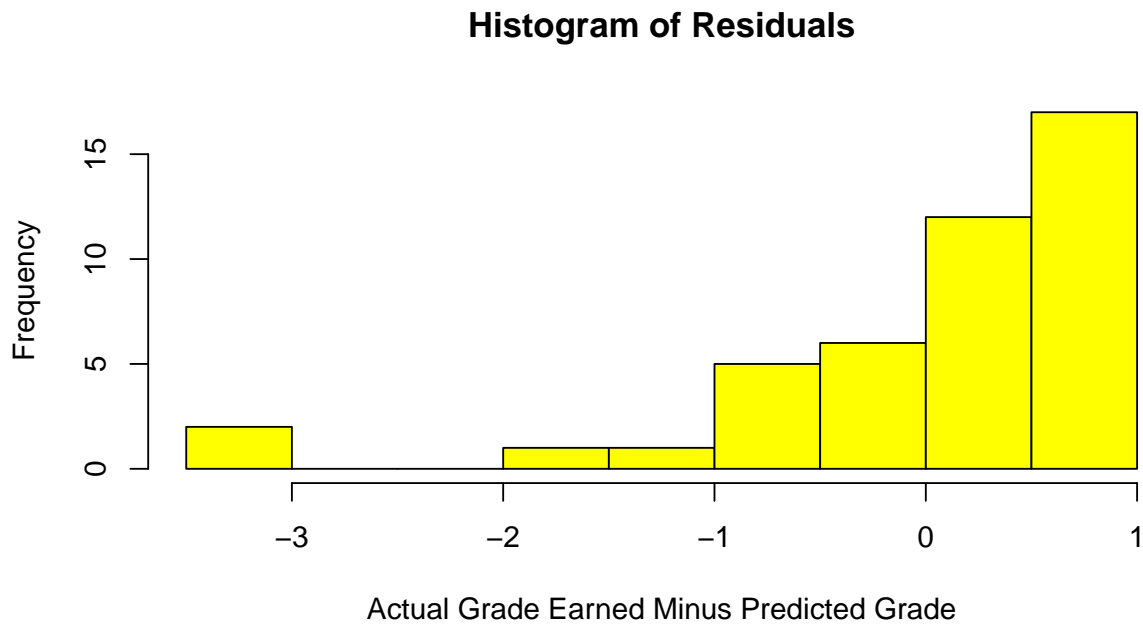
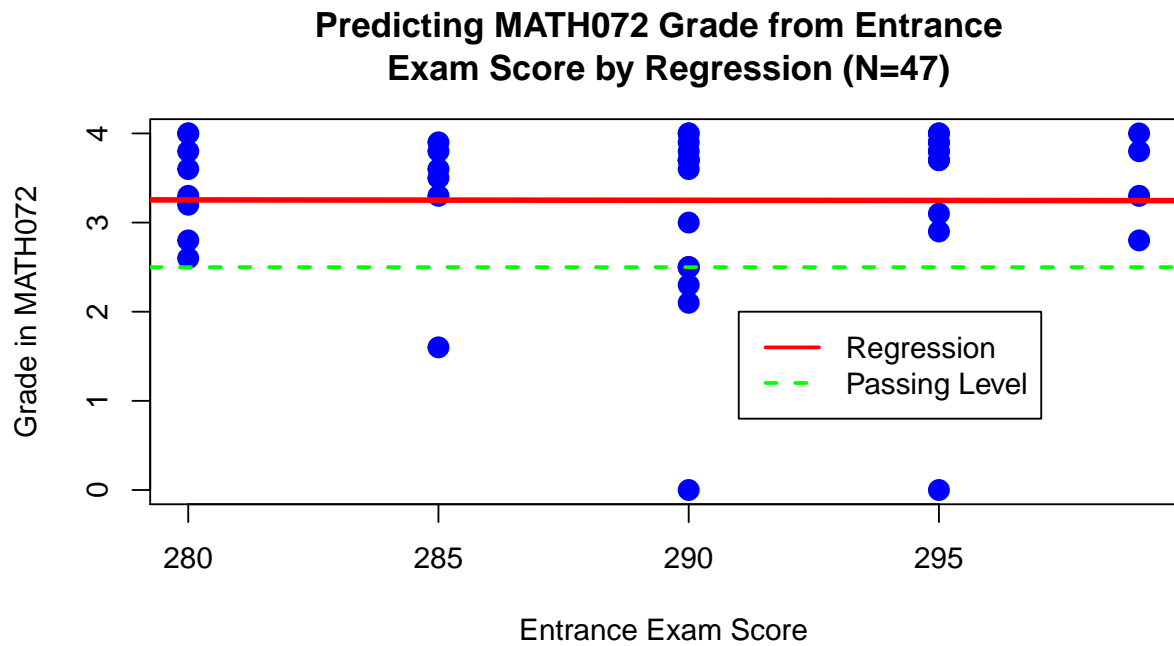


Histogram of Residuals



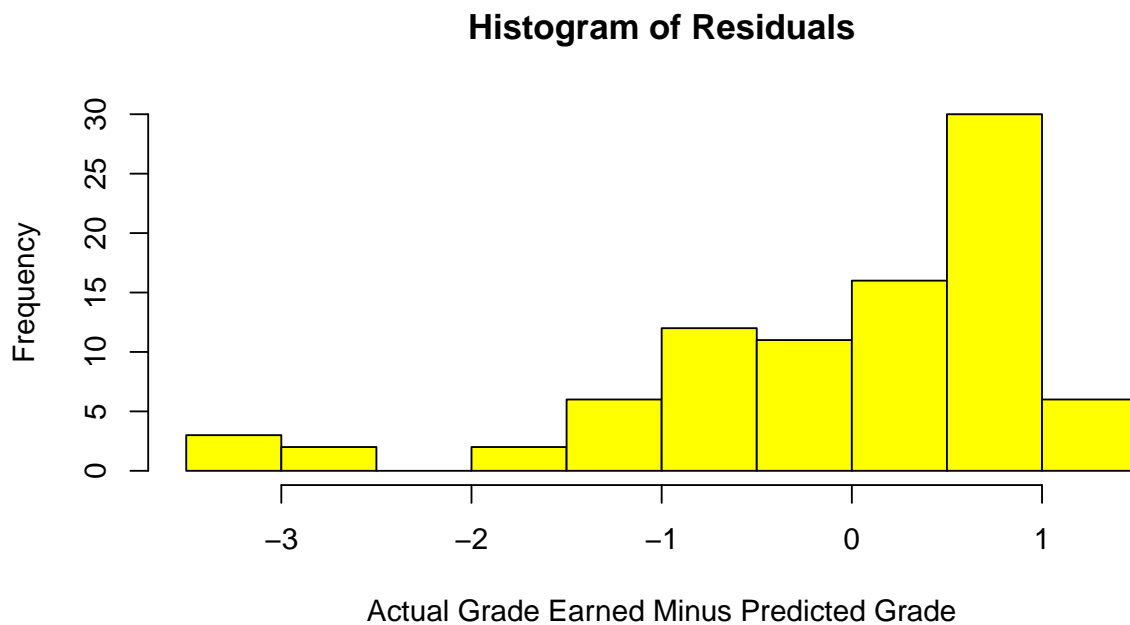
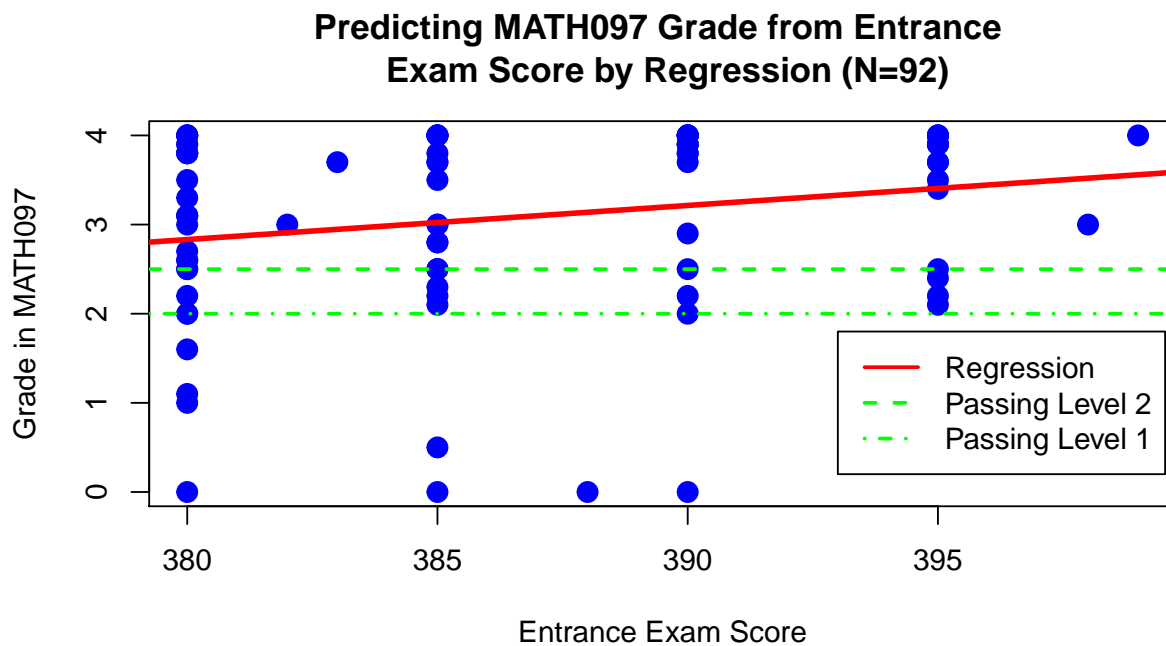
The regression line predicts a grade of 2.9 for a student earning an Entrance Exam score that is 5 points lower than the current cutoff.

5.3 MATH072 - Elementary Algebra



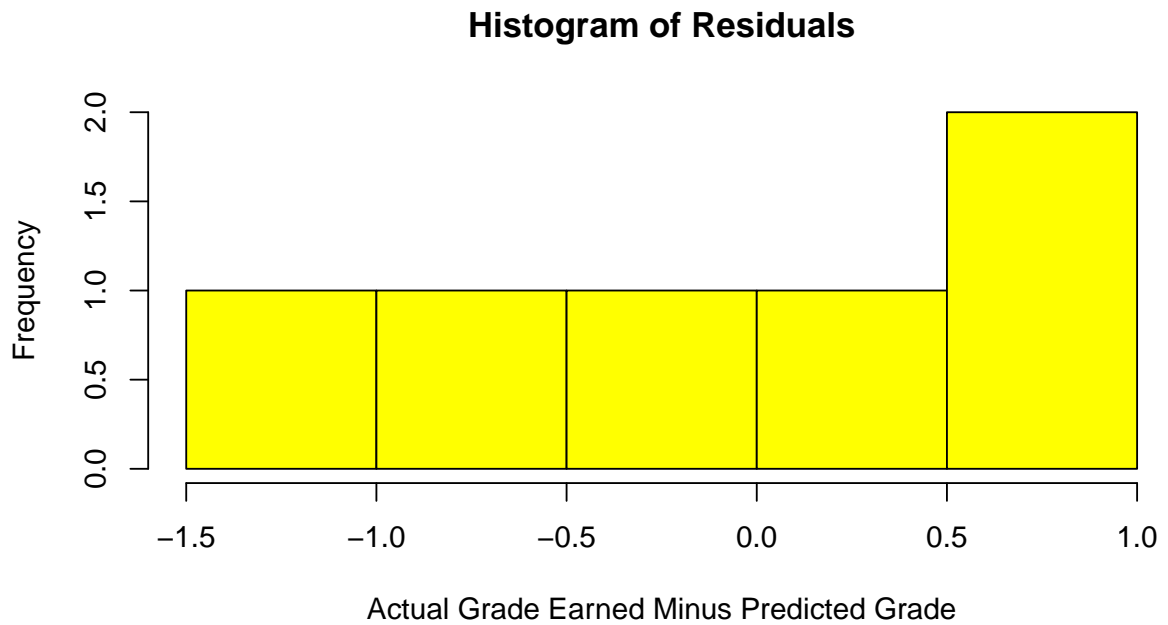
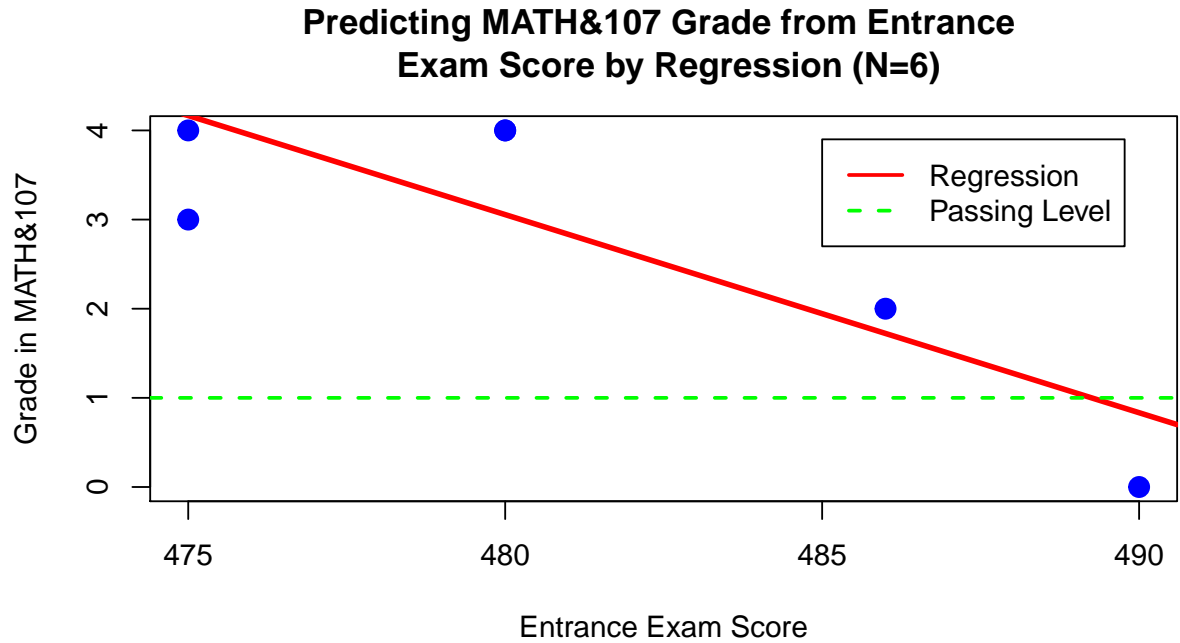
The regression line predicts a grade of 3.3 for a student earning an Entrance Exam score that is 5 points lower than the current cutoff.

5.4 MATH097 - Intermediate Algebra



The regression line predicts a grade of 2.6 for a student earning an Entrance Exam score that is 5 points lower than the current cutoff.

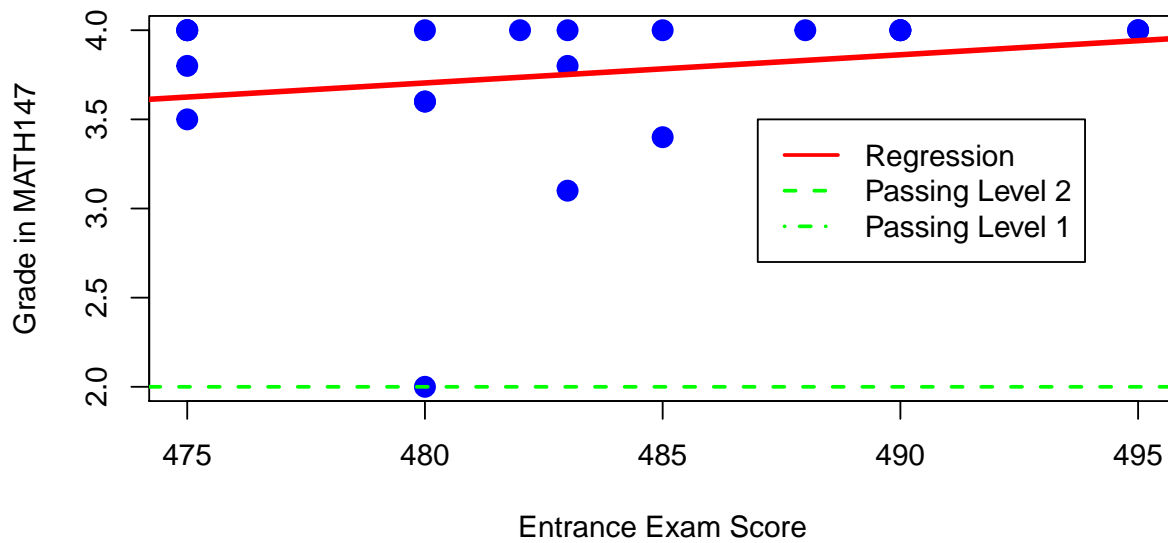
5.5 MATH&107 - Math in Society



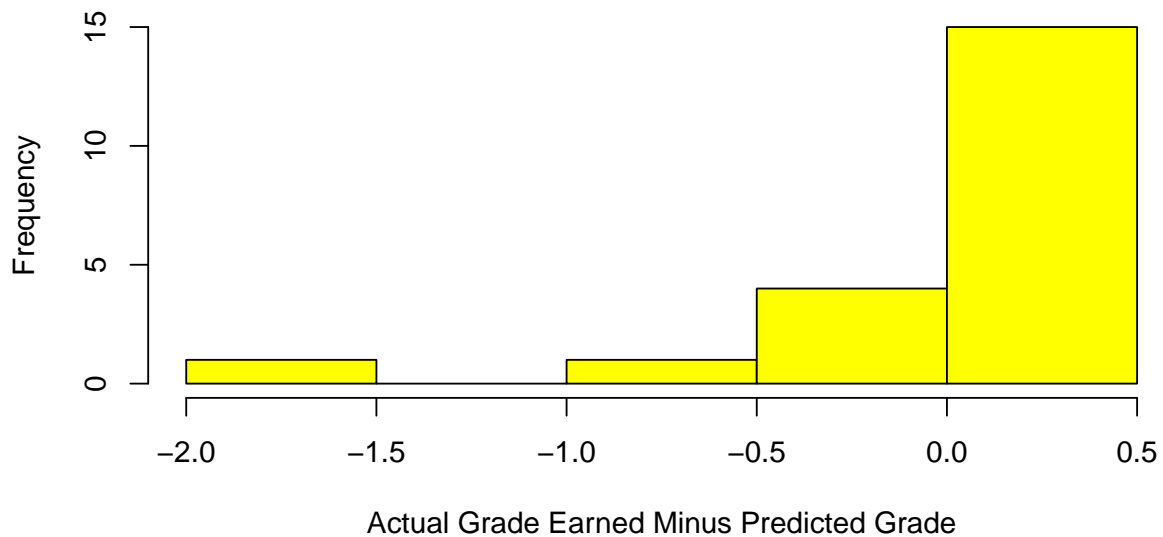
The regression line predicts a grade of 5.3 for a student earning an Entrance Exam score that is 5 points lower than the current cutoff. Of course, the maximum possible grade is 4.0. The negative slope of this line is likely due to the small sample size and the presence of a single outlier on the bottom right.

5.6 MATH&147 - Finite Math for Business and Social Science

Predicting MATH147 Grade from Entrance Exam Score by Regression (N=22)



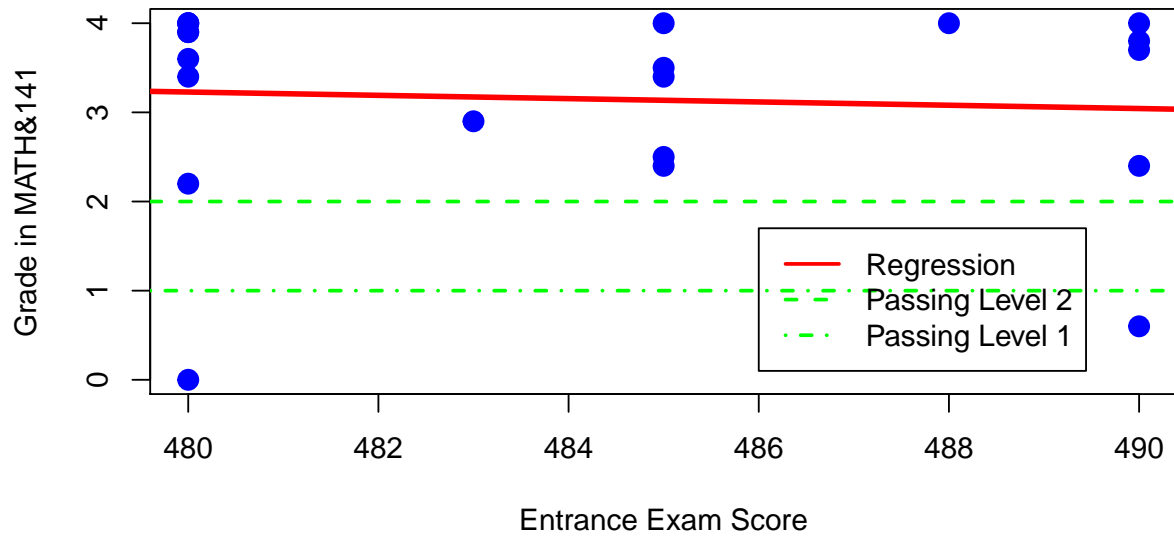
Histogram of Residuals



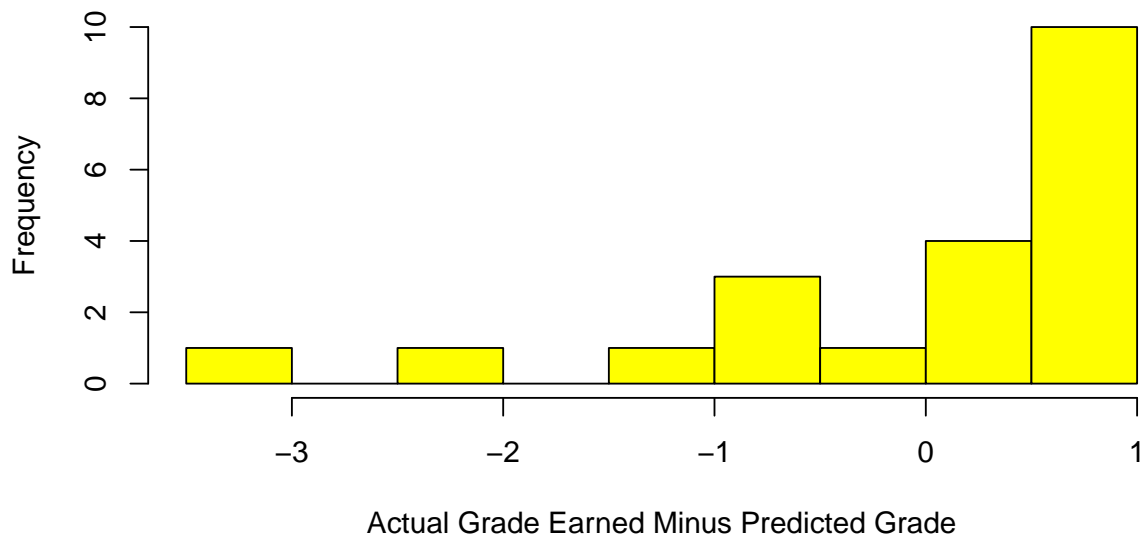
The regression line predicts a grade of 3.5 for a student earning an Entrance Exam score that is 5 points lower than the current cutoff.

5.7 MATH&141 - Precalculus 1

Predicting MATH&141 Grade from Entrance Exam Score by Regression (N=21)



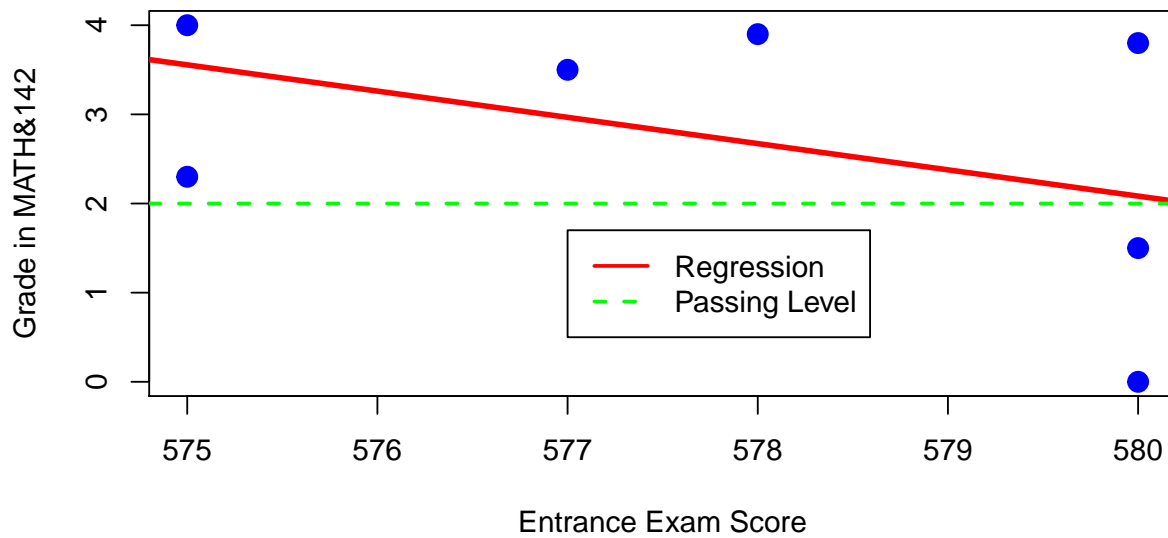
Histogram of Residuals



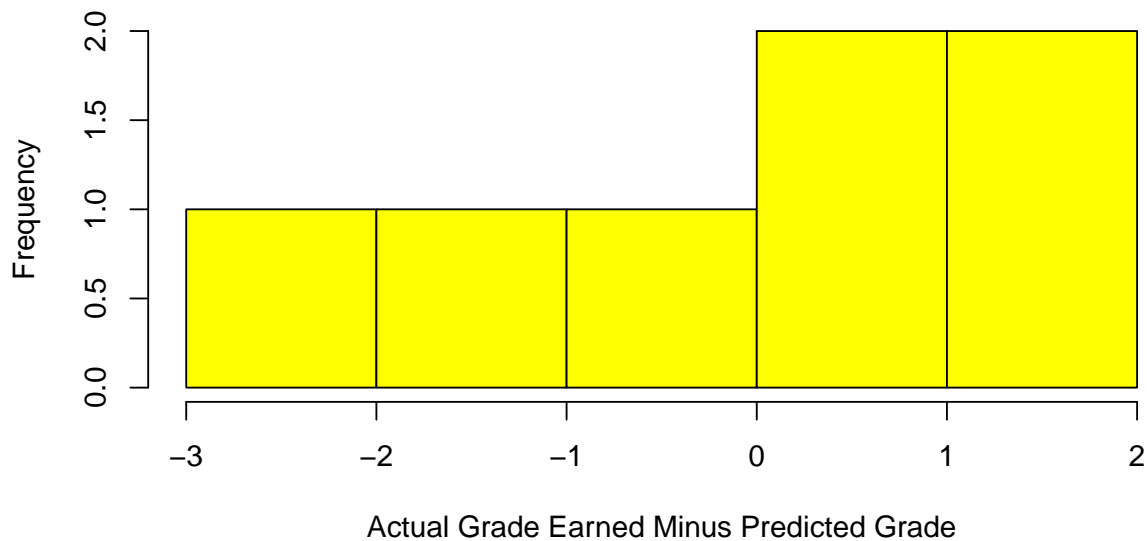
The regression line predicts a grade of 3.3 for a student earning an Entrance Exam score that is 5 points lower than the current cutoff.

5.8 MATH&142 - Precalculus 2

Predicting MATH&142 Grade from Entrance Exam Score by Regression (N=7)

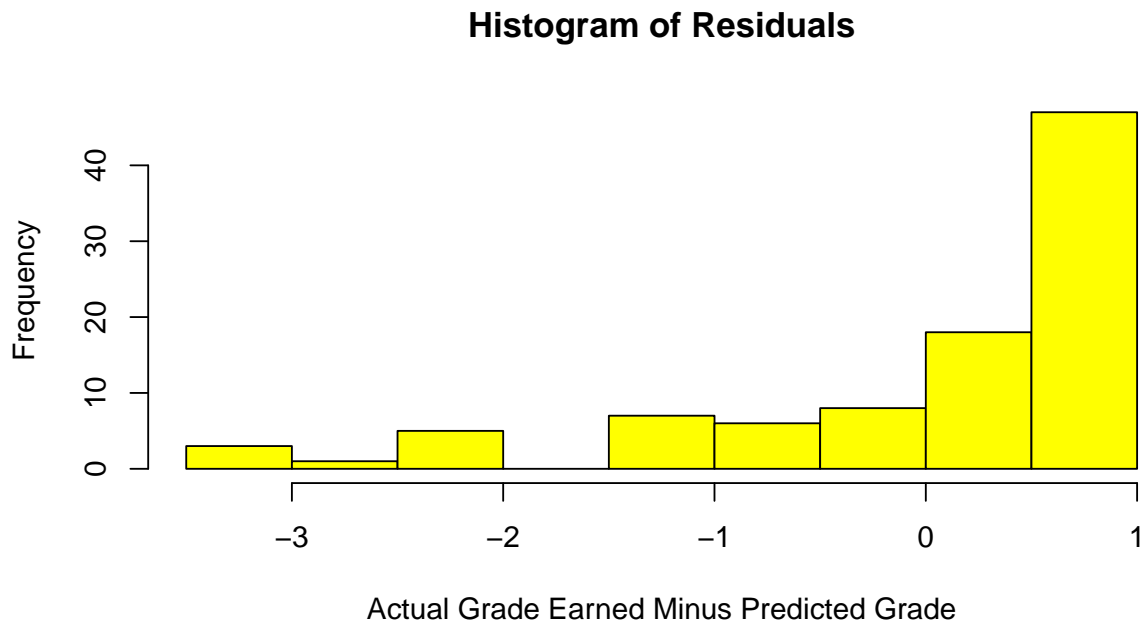
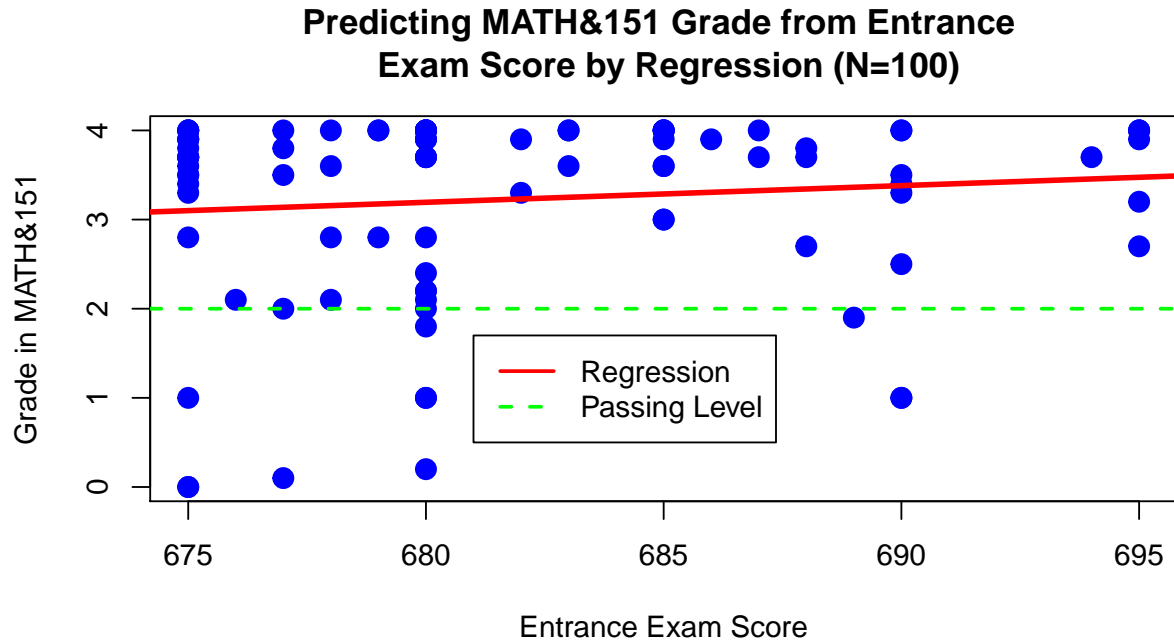


Histogram of Residuals



The regression line predicts a grade of 5 for a student earning an Entrance Exam score that is 5 points lower than the current cutoff. Of course, the maximum possible grade is 4.0. The negative slope of this line is likely due to the small sample size and the presence of a single outlier on the bottom right.

5.9 MATH&151, Calculus 1



The regression line predicts a grade of 3 for a student earning an Entrance Exam score that is 5 points lower than the current cutoff.

6 Codebook

The following information describes the data fields in the “anonymousdata.csv” file from which the analysis in this report begins.

- **X** and **X.1** - These are line numbers for the table and convey no other information.
- **Course.taken** - Indicates the math course: MATH062 is Arithmetic, MATH070 is Prealgebra, MATH072 is Elementary Algebra, MATH097 is Intermediate Algebra, MATH&107 is Math in Society, MATH147 is Finite Math for Business and Social Science, MATH&141 is Precalculus 1, MATH&142 is Precalculus 2, and MATH&151 is Calculus 1.
- **Quarter** - Indicates the quarter in which the course was taken. The first two digits indicate the academic year: 01 is 2010-2011, 12 is 2011-2012, 23 is 2012-2013, and 34 is 2013-2014. The final digit indicates the quarter: 1 is summer, 2 is fall, 3 is winter, and 4 is spring.
- **Grade.earned** - This is the grade reported on the students’ transcript. Grades are awarded on a scale of 0.0-4.0. I is Incomplete, N is Audit, NC is No Credit, P is Pass, and W is Withdrawal.
- **numericalgrade** - This is the same as Grade.earned when the assigned grade is a number in the range 0.0-4.0, and it NA otherwise.
- **Last.quarter.of.attendance** - Indicates the last quarter during which the student was enrolled; z999 indicates that the student was still enrolled at the time of data collection.
- **Drop.date** - If the student dropped the course after the 4th day of the quarter, this indicates the date of the withdrawal.
- **Concurrent.GPA** - Indicates the GPA earned during Quarter, including the grade earned in the math course.
- **Concurrent.credits.registered.for** - Indicates how many credits a student was registered for during Quarter.
- **Concurrent.credits.earned** - Indicates how many credits student successfully earned during Quarter.
- **COMPASS.Reading** - Reading score on COMPASS placement exam.
- **COMPASS.Number.Skills** - Number Skills score on COMPASS placement exam.
- **COMPASS.Elem.Alg** - Elementary Algebra score on COMPASS placement exam.
- **COMPASS.College.Alg.** - College Algebra score on COMPASS placement exam.
- **COMPASS.Trig** - Trigonometry score on COMPASS placement exam.
- **MAPS.Score** - Score on Entrance Exam. The range of the score indicates which exam was attempted: 0-99 indicates that the student tried to test into MATH062 (80 or higher is passing); 100-199 indicates MATH070 (180 or higher is passing), 200-299 indicates MATH072 (280 or higher is passing), 300-399 indicates MATH097 (380 or higher is passing), 400-499 indicates MATH&107, MATH&141, or MATH147 (475 or higher is passing for MATH&107 and MATH147, while 480 or higher is passing for MATH&141); 500-599 indicates MATH&142 (575 or higher is passing); 600-699 indicates MATH&151 (675 or higher is passing).
- **FID** - Stands for “Fake ID”, which is a surrogate for the SID (“Student ID”) number which cannot be released to the public because of federal privacy laws. Two entries in the table will have the same FID number if and only if they correspond to the same student.

7 Code and Data

The raw data for this analysis included student identification numbers, or SIDs, which cannot be released to the public due to privacy laws. Therefore, the publicly-available version of this data has had these SIDs stripped and replaced by randomly-generated fake identification numbers (labeled FID in the cleaned data set). Here is the code that was used for this pre-processing stage:

```
df1 = read.csv("rawdata/startingData.csv")
df1 = subset(df1, Grade.earned != "*")
SID = unique(df1$SID.)
FID = 1:length(SID)
set.seed(2014)
FID = sample(FID)
SIDtoFID=cbind(SID,FID)
write.csv(SIDtoFID,"cleandata/SIDtranslationtable.csv")
translator = read.csv("cleandata/SIDtranslationtable.csv")
df1$FID = numeric(length(df1$SID.))
for (k in 1:length(df1$SID.)){
  tableentry = match(df1$SID.[k],translator[,2])
  df1$FID[k] = translator[tableentry,3]
}
df1$SID. = NULL
df1$Last.Name = NULL
df1$First.Name = NULL
df1$Date.of.Birth = NULL
df1$Sex = NULL
df1$Ethnicity = NULL
write.csv(df1, "cleandata/anonymousdata.csv")
```

The second line in the code above removes the entries from the data set for which no grade was recorded (there was only one such entry in the entire study).

Two additional preprocessing steps were applied to the data. First, three columns were added to the data frame to indicate student success at the 1.0, 2.0 and 2.5 grade-point levels. Second, another column was added in which non-numerical grades were given numerical representations, to facilitate generating some of the graphs found in this report. The code for that stage of preprocessing is:

```
df1$success1.0 = logical(length(df1$FID))
df1$success1.0[df1$numericalgrade >= 1.0] = TRUE
df1$success1.0[df1$Grade.earned == "P"] = TRUE

df1$success2.0 = logical(length(df1$FID))
df1$success2.0[df1$numericalgrade >= 2.0] = TRUE

df1$success2.5 = logical(length(df1$FID))
df1$success2.5[df1$numericalgrade >= 2.5] = TRUE

df1$forcednumericalgrade = df1$numericalgrade
df1$forcednumericalgrade[df1$Grade.earned == "W"] = -0.1
df1$forcednumericalgrade[df1$Grade.earned == "N"] = -0.1
df1$forcednumericalgrade[df1$Grade.earned == "P"] = 1
df1$forcednumericalgrade[df1$Grade.earned == "NC"] = -0.1
df1$forcednumericalgrade[df1$Grade.earned == "I"] = -0.1
```

```
write.csv(df1, "cleandata/preprocessedData.csv")
```

All code, including the code used to generate this document, can be found on GitHub at <https://github.com/kriskissel>. The data files “anonymousdata.csv” and “preprocessedData.csv” can also be found there.