Using logistic regression for validating or invalidating initial statewide cut-off scores on basic skills placement tests at the community college level

Charles Secolsky
County College of Morris

Sathasivam 'Kris' Krishnan
The Richard Stockton College of New Jersey

Thomas P. Judd
United States Military Academy

Abstract

The community colleges in the state of New Jersey went through a process of establishing statewide cut-off scores for English and mathematics placement tests. The colleges wanted to communicate to secondary schools a consistent preparation that would be necessary for enrolling in Freshman Composition and College Algebra at the community college level. Statewide cut-off scores also allow for equity and transferability among colleges on decisions to place students out of developmental coursework in English and mathematics. Logistic regression analysis was used to produce expected probabilities of attaining a grade of C or better in the courses for given test scores. Expected probabilities were converted into statistical tests for determining the predictive accuracy of placement cut-off scores for three tests: essay, mathematics computation, and elementary algebra. Prior to the conduct of this research, passing scores were set by committees with no established standard setting method. These passing scores are evaluated here with logistic regression so that a more established standard setting methodology could be recommended to community college officials for future use.

Key Words: Placement Testing, Statewide Cut-off Scores, Logistic Regression, Validity of Cut-scores, Basic Skills Tests at Community Colleges

Some states are striving to come up with statewide cut-off scores in an effort to communicate to secondary schools the level of preparation that would be necessary to enroll in Freshman Composition and College Algebra and succeed. Furthermore, statewide cut-off scores result in equity and transferability by allowing students to test out of developmental courses with consistent preparation levels across the state. In the absence of statewide standards, a student could be deemed as passing at one college and failing at another.

The major purpose of this study was to arrive at a common passing score on essay, mathematics computation and elementary algebra placement tests in the State of New Jersey for all the community colleges that used the same placement tests. Initially, the cut-off scores for entering Freshman Composition and College Algebra were set by a statewide a committee with no established standard setting method used. The cut-off scores were determined solely on the basis of content considerations contained in the Accuplacer manual. The purpose of this paper was to validate the cut-off scores arrived at by committee and recommend whether or not there was a need to re-establish these cut-off scores in their own right using an established standard-setting methodology. Therefore, the study was an attempt to set passing scores using logistic regression and as a by-product validating the committees' initial cut-off scores.

There are a number of standard setting methods that are widely used. Popular methods include the modified Angoff method (Angoff, 1971; Livingston and Zieky, 1982), the Contrasting Groups approach (Livingston and Zieky, 1982), Borderline Group (Livingston and Zieky, 1982), the Bookmark approach (Lewis, Mitzel, and Green, 1996; Mitzel, Lewis, Patz and Green, 2001), Body of Work (Kingston, Kahl, Sweeney and Bay, 2001; Morgan and Michaelides, 2005) and logistic regression. (For a concise description of standard-setting and validation methods see Pitoniak and Morgan, 2012). Logistic regression was chosen because it is essentially used for validation. Also, there were two variables: (1) placement test scores and (2) grades. For this reason, studies aimed at validating cut scores for placement decisions using logistic regression are common. They have the advantages that they need fewer panelists, require less expense due to a much lower cost for meeting space and reduced panelist travel costs. It has the disadvantage of a total reliance on data with little consideration as to what different cut scores would mean in terms of student mastery of specific content. The reliance on classroom grades can be inherently subjective and may differ radically from one classroom to the next in terms of the level of content mastery necessary to achieve the same grade (Morgan and Michaelides, 2005). .

Logistic regression is a method that classifies students as to whether they pass or do not pass some criterion using an established cut-off score where the criterion is dichotomous. What factors predict passing or failing? In the present study, two variables were used: (1) grades that are predicted and (2) scores on placement tests that are the predictors. Statewide cut-off scores in English and mathematics were first obtained by using an aggregate of students' grades and placement test scores for four participating colleges that produced data for the study. Logistic regression, although there is some inherent subjectivity as with all of the standard setting methods, can be more accurate. Morgan and Michaelides (2005) describe the use of logistic regression as applied to establishing cut-off scores for college placement. According to these authors, "Logistic regression is a statistical method that uses binary information, e.g., the probability of success and the probability of failure, to predict success based on a piece of information, e.g. a test score (Morgan and Michaelides, 2005; p. 10)."

The Bookmark approach is most appropriate for adaptive tests where the difficulty of the items presented to the students can change depending on the correctness of the previous response. However, with the use of grading categories, the logistic regression approach is more appropriate. Nevertheless, something that is common to all standard setting approaches is the use of judgment. Logistic regression being the most empirical of all the methods described does not as much involve judgment initially, but requires more judgment in validating the passing score near the end of the process.

Method

The authors contacted institutional research officers at each of the nineteen community colleges in the state to obtain endorsement for the study. It was intended that the institutional researcher at each college that used Accuplacer would send the English and mathematics grade data as well as the results of three placement test scores to the study team for analysis.

After the data were aggregated, a table was created for the predicting the probability of success for each score interval for each placement test using the percentage of students scoring C or better in each course (Morgan and Michaelides, 2005). From the table, the proportion of successful completers in the course (C or better) tends to increase for students in the higher score intervals. However, a near perfect linear relationship does not always exist. For small sample sizes there are likely to be anomalies (Morgan and Michaelides, 2005). With the large aggregation of data like for the present study, the proportion of success is likely to be more linear.

Finally, a regression model is fitted to the data to predict success in the courses using the test scores in the case where success is defined as obtaining a grade of A, B, or C. For this model, the intercepts and slope coefficients of the equation are:

$\log_e(p_i/1-p_e)$=Intercept + (Slope)(test score), where p_i is the probability of success as given by the model (Morgan and Michaelides, 2005).

The model provides a hypothetical example of expected probabilities of success that are plotted on a graph (see Figure 1). A curve is produced that represents the expected probability of obtaining an A, B, or C in the course given a test score (Morgan and Michaelides, 2005). College administrators use probabilities of success at each score point to determine the expected effect in terms of student success/non-success for identifying the optimal placement for the cut score (Morgan and Michaelides, 2005). Separate analyses are needed for English and mathematics courses.

Data for students from four community colleges (n=4,317) were aggregated to produce three logistic regression analyses. Although data from only four community colleges out of 19 raises the question of representativeness, the sample size is considered large enough to provide a reasonable estimate of the validity of the initial passing scores. Each analysis used a placement test and a grade from three areas: (1) an essay test and the grade in Freshman Composition, (2) a math computation test and a grade in College Algebra, and (3) an elementary algebra test score and a grade in College Algebra. The grades were treated as binary (C or better) or (below C or W).

Placement Scores from Statewide Committee Initiative

An Accuplacer score of 76 was chosen to automatically exempt a student from the developmental algebra requirement. An SAT score of 530 and above also automatically exempted a student from this developmental algebra requirement. Based upon review and discussion, the Academic Officers Association for community colleges in the State of New Jersey recommended that students shall automatically be placed into the first college-level English Composition course when they satisfy any one of the following three conditions:

- completing the developmental sequence of writing courses, or
- earning an SAT score of 540 or above on Critical Reading, or
- earning a score on college administered placement test which exceeded the threshold score delineated below.

A score of 8 or higher on the WritePlacer, or on an in-house college-developed test, was determined to be the threshold. (Colleges using the in-house test should evaluate the overall quality of writing based on the Score Point Description of the College Board WritePlacer.)

Score Point Description of 8:

> This writing sample competently communicates a message to a specified audience. Although the purpose of the writing sample may be clear, the development of supporting details may not be fully realized. The writer's organization of ideas is evident, but may lack specificity, be incomplete or not be developed in a[n] effective sequence. There is evidence of control in the use of mechanical conventions such as sentence structure, usage, spelling, and punctuation, though minor errors in the use of conventions may be present. (College Board, WritePlacer, 2006).

Logistic Regression

Logistic regression was performed using the grades for Freshman Composition and College Algebra and scores on the Writeplacer or an in-house developed test as the English essay test and Elementary Algebra as the placement test for College Algebra. In addition, the Mathematics Computation test was also used as the predictor for grades in College Algebra. "FreshComp" was designated as the Freshman Composition grade and "CollAlg" was designated as the College Algebra grade. The grades were each scored as binary variables (C or better as "1") and (below C or W-withdrawal from a course) was designated as "2". Logistic regression results for FreshComp are found in Table 1.

From Table 1, the Wald statistic was significant indicating that the WritePlacer or the in-house essay test is associated with the grading dichotomy. In addition, the maximum likelihood estimate for Freshman Composition was significant (p < /0001). However, the odds ratio shows little predictive accuracy for students getting a C or better.

The ROC (Receiver Operating Characteristic) curve is a graphical display that gives a measure of the predictive accuracy of a logistical model. For a model with high predictive accuracy, one should expect the ROC curve to rise quickly. Thus the area

under the curve is usually high. The converse is also true in that a curve that rises slowly suggests the model has low predictive accuracy. The ROC curve for Freshman Composition grades is found in Figure 2
As can be seen with the ROC curve for Freshman Composition, the estimated area under the curve is 0.53907 – not a good predictive accuracy for Freshman Composition grades. The logistic regression curve on the horizontal axis at a score of 8 (the passing score) on the writing test in Figure 3 lines up with a vertical predicted probability of between .68 and .69 that the student gets a C or better in Freshman Composition.

Classification Table

The Classification Table option on the MODEL statement in PROC LOGISTIC in SAS output classifies the input binary response observations according to whether the predicted event probabilities are above or below some cut point value z in the range (0, 1). An observation is predicted as an event if the predicted event probability exceeds z. The sensitivity value of 68.6 corresponds to a probability level of .68. The percent of false positives is 31.8%. The percent correct is 68.2%. The graph in Figure 3 and classification table for Freshman Composition (Table 2) indicate that the cut-off score of 8 does not predict accurately. The score of 8 was the score chosen by the committee for the writing placement test.

Elementary Algebra Test

From Figure 5, the predicted probability of the logistic regression for a cut-score of 76 on the elementary algebra test was .720. This implies that 72% would receive a C or better in College Algebra. The percent correct according to the classification was 55.2%. As can be seen in Table 4, the Specificity of the ROC curve is 87.3, which is fairly high, but associated with this is a very high probability of False Negatives (.602).

Mathematics Computation Test

A score on the mathematics computation placement test of 76 corresponds to a probability of between .69 and .70 in getting C or better in College Algebra. The percent correct is 62.3%. A worse result was obtained for the mathematics computation test in Table 5 than in Table 1 for the Freshman Composition grades. There was little predictive accuracy. In fact, the global hypothesis that BETA=0 was not rejected indicating no statistical association between the test and college algebra grades. However, the ROC curve had greater predictive value for this result with an estimated area of C=-0.62268/
A score on the mathematics computation placement test of 76 corresponds to a probability of between .69 and .70 of getting C or better in College Algebra. The percent correct is 62.3%. Based on Figure 6 and Table 5, there is substantial inaccuracy of prediction for the passing score of 76 on the mathematics computation test.
Findings

The cut-scores set on the Writeplacer or the in-house writing test was 8 with a probability of success of 0.68-0.69 on Freshman Composition grades. The cut-score was 76 with a probability of success in College Algebra with C or better approximately equal to 0.720 on the elementary algebra test. The cut-score for the mathematics computation test was 76 with an expected probability of - 0.730.

The probabilities selected by the representatives of each community college are judgmental and using these probabilities, possibly 19 different placement decisions (one for each college) could have been made on which cut-off scores to implement. The first time through the process is difficult because it is so judgmental. Afterwards, when statewide cut-offs are validated or invalidated considerably less judgment is required.

It is possible for 19 cut-off scores to be obtained based on the probabilities of success judged by the aggregate as deemed by the different community college academic representatives. Before cut-off scores are selected, however, representatives should be trained so that they can reach an understanding of the process of selecting the probability of success or non-success and how the generated cut-off scores on these three tests were established and validated. One problem with the procedure is that only one independent variable is used. The method used previously for arriving at the cut-off scores for each of the three tests was to use the median of the participating community college's cut-off scores. This resulted in the initial statewide cut-off scores which cannot be validated. In other words, the cut-score was reached by collective decision-making through discussion and calculating the median of eight community colleges and were not validated in this study, especially in mathematics computation.

The median value for the eight participating colleges on the essay test or WritePlacer was 7.5. The median value of the nine participating colleges on the elementary algebra test was 76. These two cut-off scores were based on committee discussion without any established standard-setting technique. Rather than basing the cut-off scores on collective decision-making and median calculations, the cut-score validation study undertaken here points to the need for the establishment of new statewide cut-off scores.

The Academic Officers Association in conjunction with the College Board had similar findings in their validity study of cut-scores for community colleges in New Jersey and opted for a decision zone for placing students into courses that affords greater latitude for the individual community college and thus gets away to some extent from the original intent of establishing one cut-off score for all 19 community colleges on each test.

Conclusion

Placement tests are high stakes tests because they inadvertently place students into developmental or credit-bearing coursework. The predictive accuracy of the placement decisions are of utmost importance if students are to succeed in coursework and, in fact, be retained in college. The goal of the present study was to evaluate the cut-off scores established for determining readiness in Freshman Composition and College Algebra. Previous efforts in setting cut-scores apparently resulted in the need for setting the standard too high with a greater number of developmental course sections having to be being offered. Logistic regression was used to invalidate the cut-scores thereby recommending new ones be re-established so that they have greater predictive accuracy.

References

Angoff, W.H. (1971).Scales, norms, and equivalent scores.In R. L. Thorndike (Ed.) *Educational Measurement* (2nd ed., pp. 508-600).Washington, DC: American Council on Education.

College Board (2006) *WritePlace*r. New York, New York

Kingston, N. M., Kahl, S., Sweeney, K., & Bay, L. (2001). *Setting performance standardsusing the body of work method*. In G. J. Cizek (Ed.), *Standard setting: Concepts, methods and perspectives* (pp. 219-248). Mahwah, NJ: Erlbaum.

Lewis. D.M., Mitzel, H.C. &Green, D.R. (1996, June).Standard setting: A bookmark approach.In D.R. Green (Chair), *IRT-based standard-setting procedures utilizing behavioral anchoring*.Symposium conducted at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ.

Livingston, S.A. & Zieky, M.J. (1982). *Passing scores: A manual for setting standards of performance on educational and occupational tests.*Princeton, NJ: Educational Testing Service.

Mitzel, H.C., Lewis, D.M, Patz, R.J. & Green, D.R. (2001).The bookmark procedure: Psychological perspectives.In G.J. Cizek (Ed.),*Setting performance standards: Concepts, methods, and perspectives*.Mahwah, NJ: Lawrence Erlbaum Associates.

Morgan, D.L. & Michaelides, M.P. (2005).*Setting Cut Scores for College Placement* Research Report No. 2005-9, New York, NY, The College Board

Pitoniak, M.J. & Morgan, D.L. (2012).Setting and validating cut scores for tests. In C. Secolsky and D.B. Denison (Eds.) Handbook on measurement, assessment, and evaluation in higher education (pp.343-366). New York: Routledge.

Appendix

Table 1:
Logistic Regression Results for Freshman Composition Grades

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Like. Ratio	18.3512	1	<.0001
Score	18.5802	1	<.0001
Wald	18.4991	1	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.4957	0.0696	50.7572	<.0001
FreshComp	1	0.0455	0.0106	18.4991	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
FreshComp	1.047	1.025	1.068

Table 2:
Classification Table for Freshman Comp Grades and Expected Probabilities

Prob Level	Correct Event	Correct Non-Event	Incorrect Event	Incorrect Non-Event	Percentages Correct	Sensitivity	Specificity	False POS	False NEG
0.620	2820	0	1317	0	68.2	100.0	0.0	31.8	.
0.640	2378	265	1052	442	63.9	84.3	20.1	30.7	62.5
0.660	2334	277	1040	486	63.1	82.8	21.0	30.8	63.7
0.680	1932	474	843	888	58.2	68.5	36.0	30.4	65.2
0.700	1094	861	456	1726	47.3	38.8	65.4	29.4	66.7
0.720	187	1255	62	2633	34.9	6.6	95.3	24.9	67.7
0.740	0	1317	0	2820	31.8	0.0	100.0	.	68.2

Table 3:
Logistic Regression Results for College Algebra Grades

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Like. Ratio	48.0093	1	<.0001
Score	46.7350	1	<.0001
Wald	45.7484	1	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.1784	0.0679	6.9033	0.0086
CollAlg	1	0.00930	0.00137	45.7484	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
CollAlg	1.009	1.007	1.012

Table 4:

Classification Table for College Algebra Expected Probabilities

	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.540	1069	0	646	0	62.3	100.0	0.0	37.7	.
0.560	677	327	319	392	58.5	63.3	50.6	32.0	54.5
0.580	677	327	319	392	58.5	63.3	50.6	32.0	54.5
0.600	642	353	293	427	58.0	60.1	54.6	31.3	54.7
0.620	559	408	238	510	56.4	52.3	63.2	29.9	55.6
0.640	478	453	193	591	54.3	44.7	70.1	28.8	56.6
0.660	411	487	159	658	52.4	38.4	75.4	27.9	57.5
0.680	344	507	139	725	49.6	32.2	78.5	28.8	58.8
0.700	280	534	112	789	47.5	26.2	82.7	28.6	59.6
0.720	217	564	82	852	45.5	20.3	87.3	27.4	60.2
0.740	155	597	49	914	43.8	14.5	92.4	24.0	60.5
0.760	84	624	22	985	41.3	7.9	96.6	20.8	61.2
0.780	21	644	2	1048	38.8	2.0	99.7	8.7	61.9
0.800	0	646	0	1069	37.7	0.0	100.0	.	62.3

Table 5:

Logistic Regression Results for Math Computation

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1.3084	1	0.2527
Score	1.3024	1	0.2538
Wald	1.3015	1	0.2539

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr >ChiSq
Intercept	1	0.4485	0.0728	37.9706	<.0001
CollAlg	1	0.00096	0.000843	1.3015	0.2539

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
CollAlg	1.001	0.999	1.003

Table 6:

Classification Table from Math Computation

	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.600	1082	0	650	0	62.5	100.0	0.0	37.5	.
0.620	623	333	317	459	55.2	57.6	51.2	33.7	58.0
0.640	156	515	135	926	38.7	14.4	79.2	46.4	64.3
0.660	0	650	0	1082	37.5	0.0	100.0	.	62.5

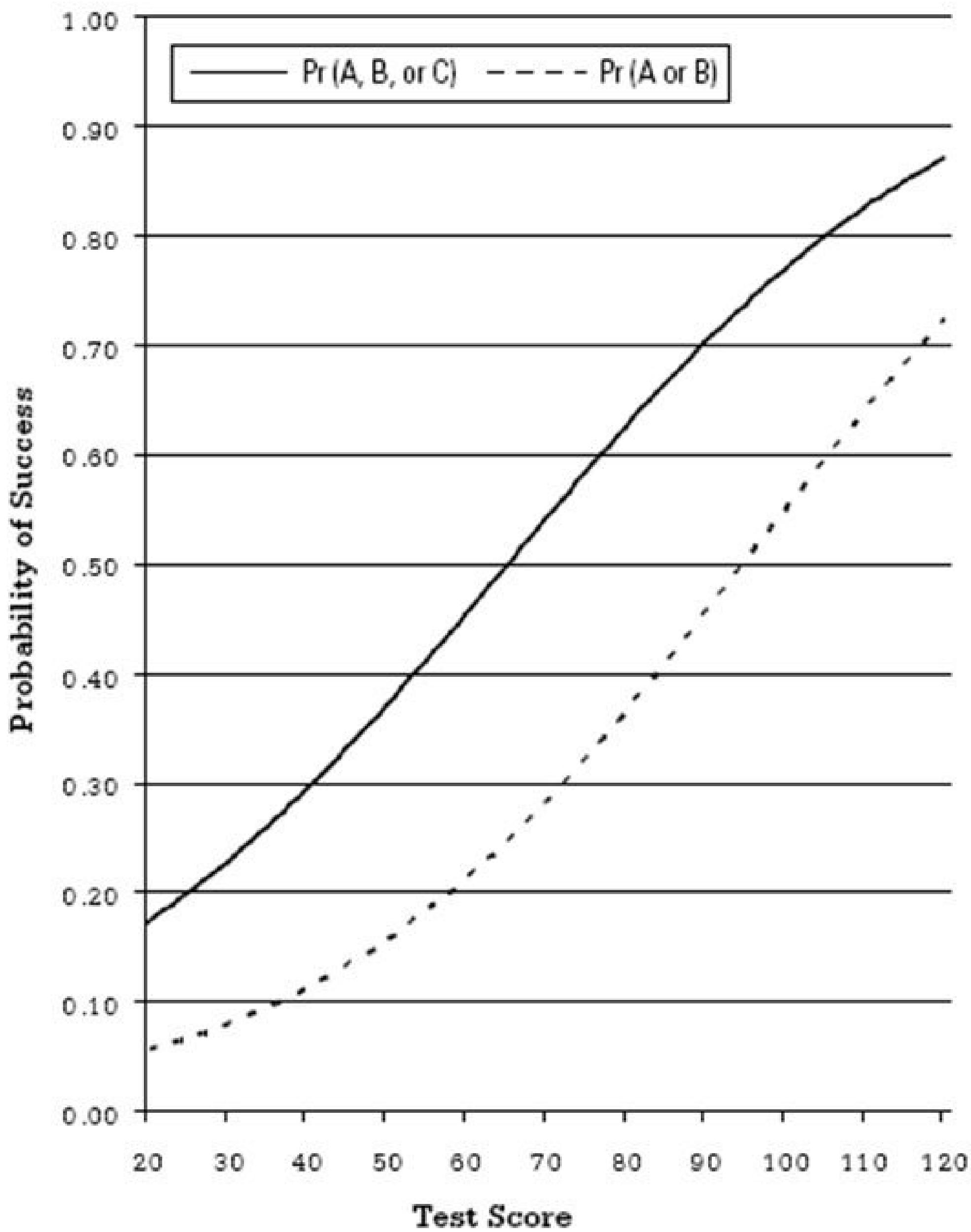

Figure 1. Graphs of predicted probability of success at each score point.

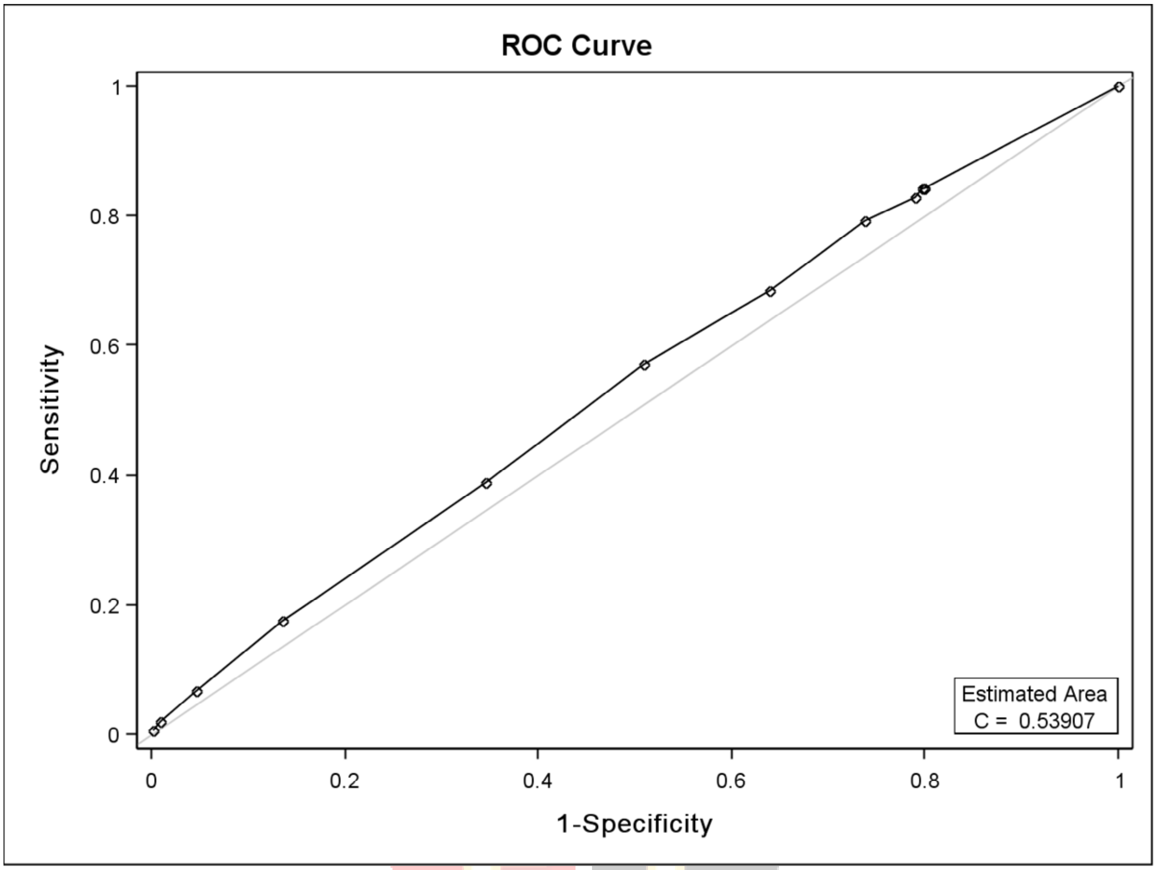

Figure 2: The ROC Curve for Freshman Composition

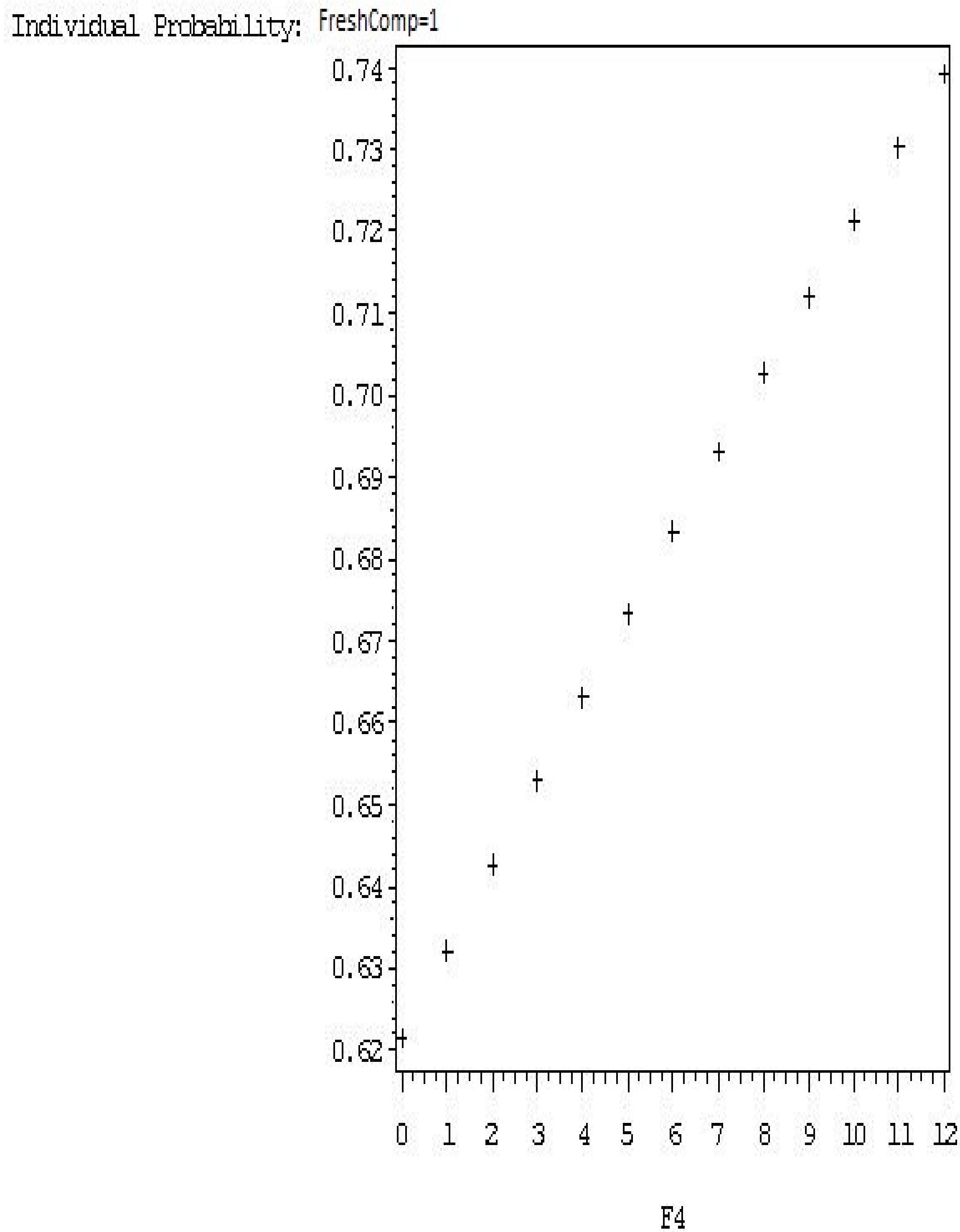

Figure 3. Plot of expected probabilities and test scores for Freshman Composition

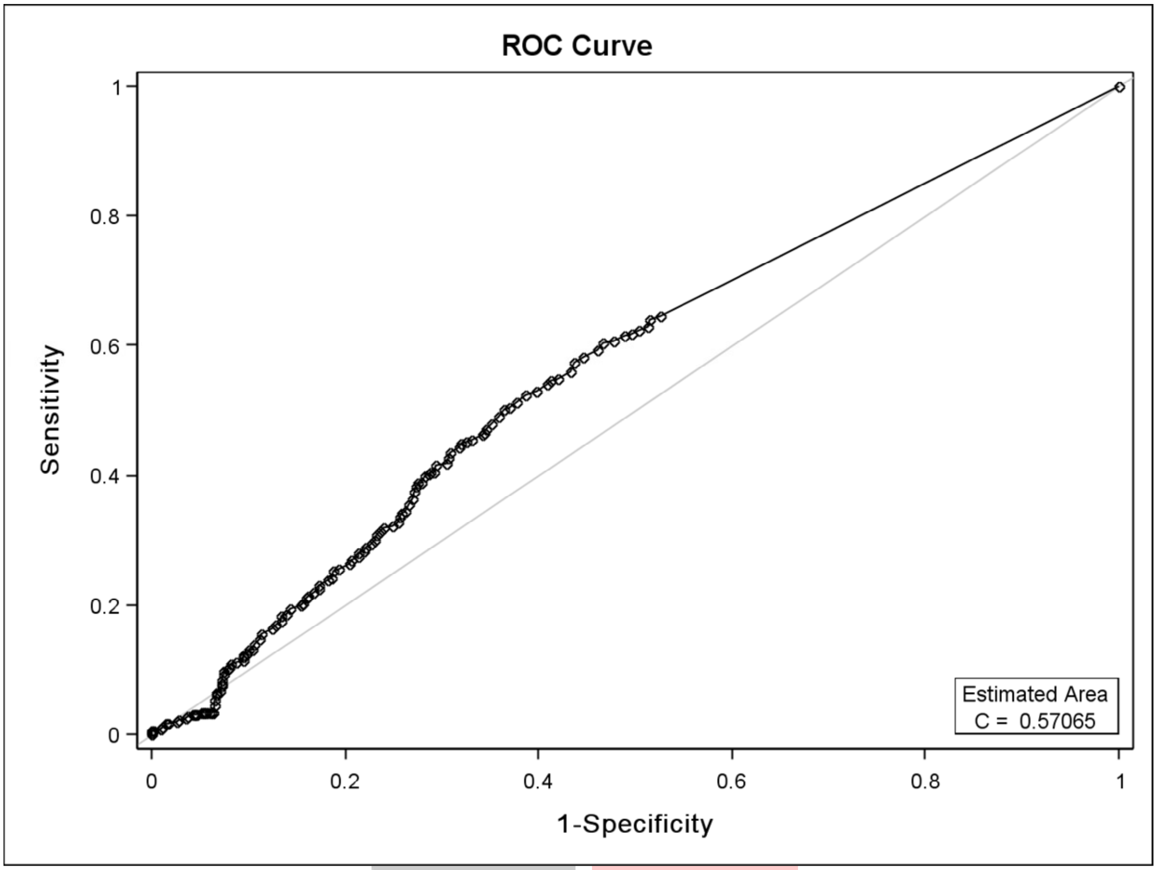

Figure 4: ROC Curve for College Algebra

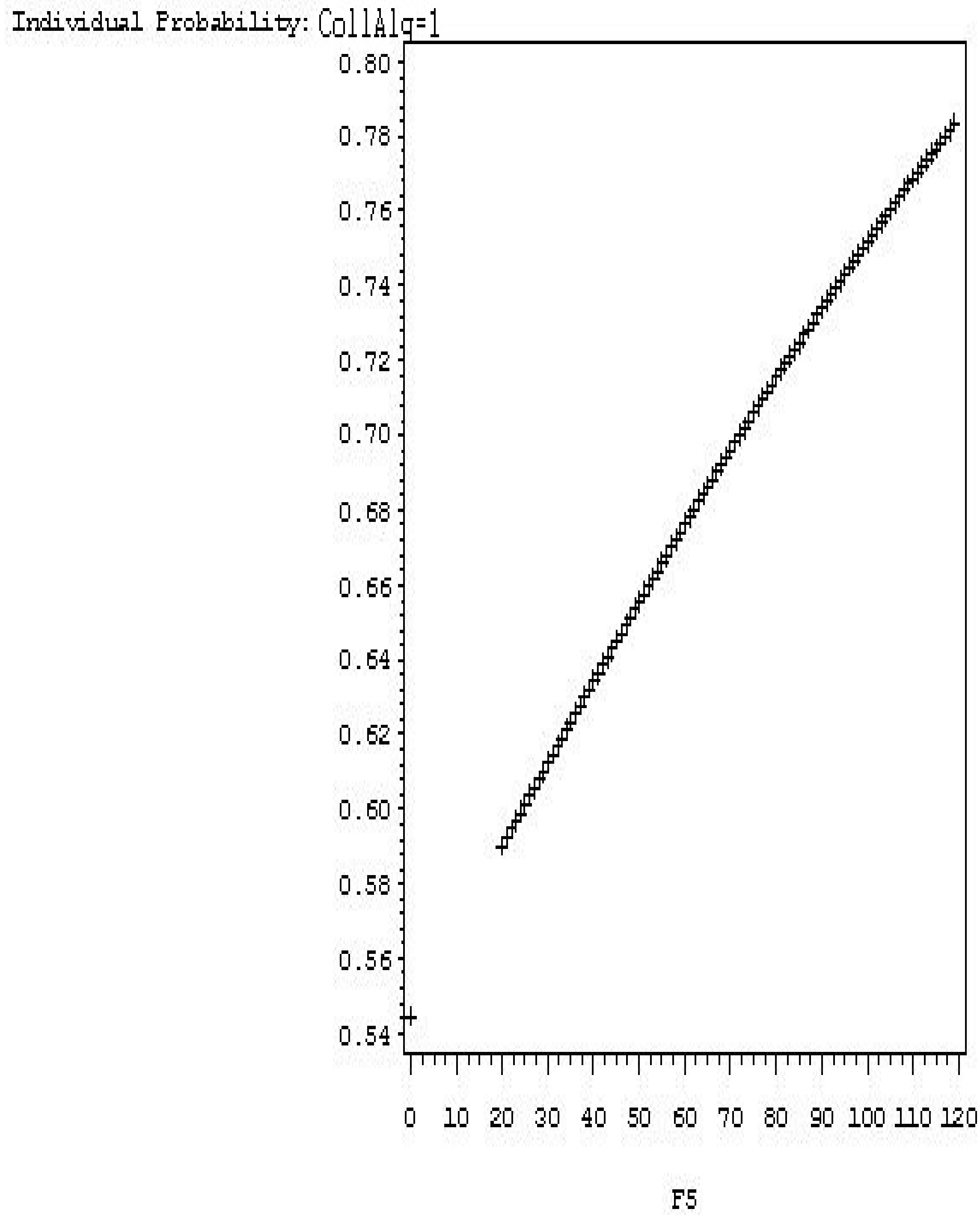

Figure 5: Logistic Regression for College Algebra Grades

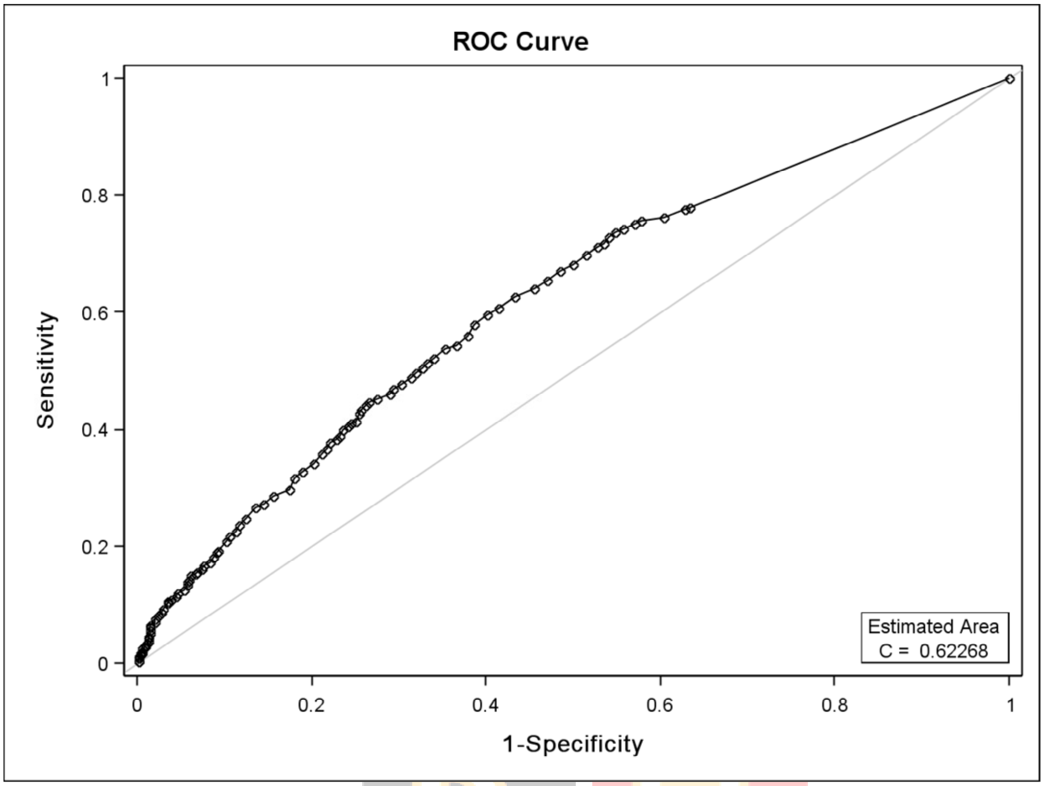

Figure 6: ROC Curve for College Algebra Using Math Computation

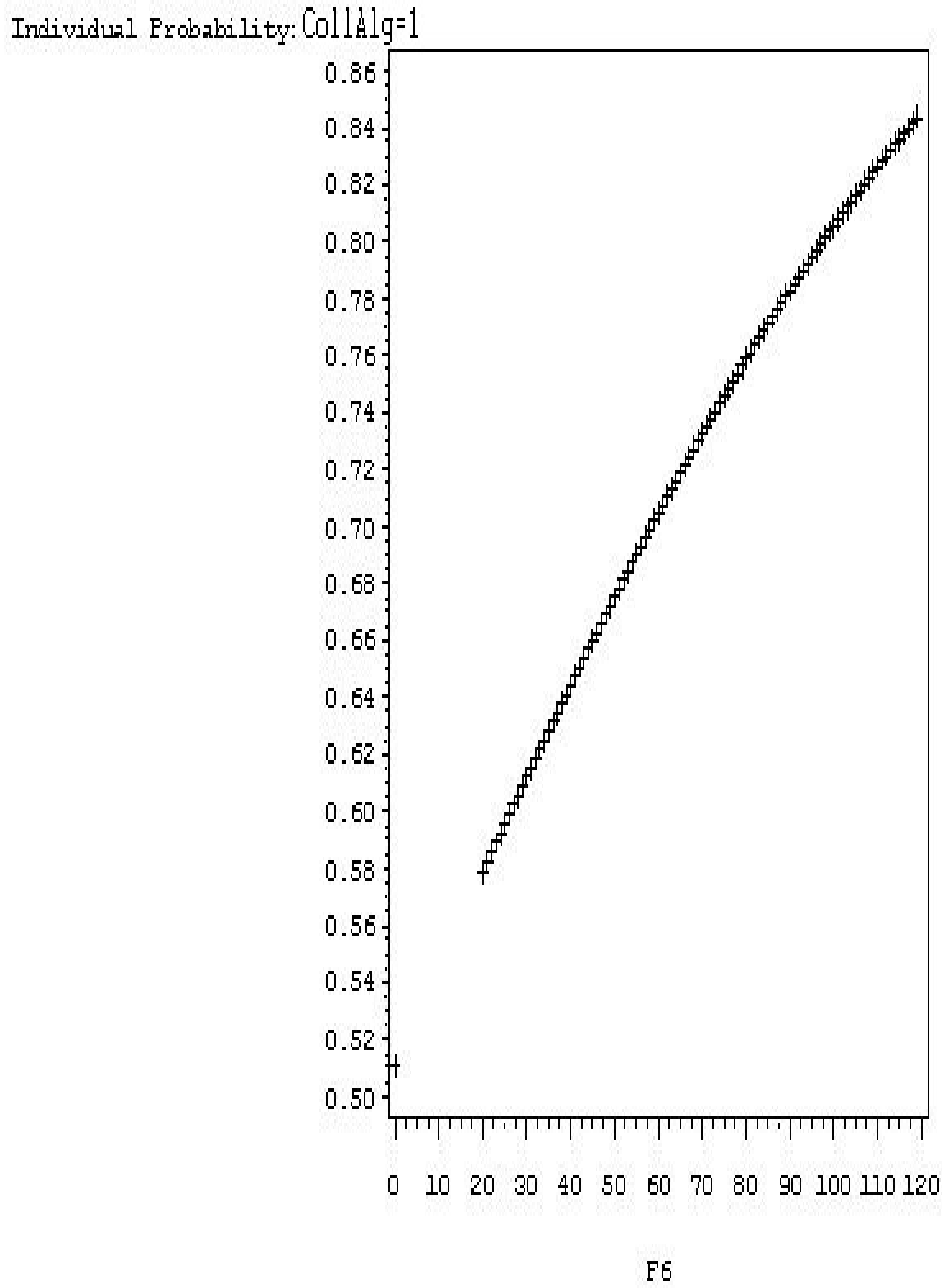

Figure 7: Logistic Regression for College Algebra Grades from Math Computation Scores