An Adaptive Model for Predicting Course Enrollment

Christine R. Kraft and James P. Jarvis

Department of Mathematical Sciences
Clemson University
Clemson, SC USA
{stemm, jpjrv}@clemson.edu
November 14, 2005

Abstract

Predicting university course enrollment is an inherently difficult problem because of uncertainty in student retention, course pass rates, admission policies, and curriculum requirements. This study presents a procedure for creating adaptive course prediction models. We use student characteristics to identify groups of undergraduates whose course enrollment rates are significantly different than the rest of the university population and use historical enrollment rates and current information to predict enrollment for the coming semester. We then use the prediction model to aid in a system for releasing seats to new students during summer orientation sessions. The seat release system addresses how to estimate seat need each session, how to allocate seats for multiple course sections, and how to predict seat shortage and surpluses. We illustrate the course prediction modeling procedure and the seat release system using data from Clemson University.

Introduction

Course schedules are prepared and submitted well in advance of the start of the semester. At Clemson University schedules for fall semester are due in February, a time when admissions estimates, course pass rates, and student retention are still uncertain. The schedules are published in March, early registration for continuing students begins in April, spring grades are recorded in May, and freshmen and transfer orientations begin in mid-June. Adjusting the number of instructors after June is not desirable, and may not be possible. Making accurate predictions at the time the schedules are initially constructed and adjusting these predictions as soon as changes become apparent are extremely important. Although complete information is not available at the time of prediction, we have course rolls and student information such as major and enrollment status, which is a code the university uses to identify each student as first-time, transfer, continuing, or returning. This information along with analysis of historical data makes it possible to construct generic course prediction models that are robust and accurate.

Course prediction models can aid in releasing seats to new students by better predicting seat need. New student registration can be distributed over the summer preceding the fall semester through a series of orientation sessions where students may register for courses. Universities use seat release systems to give similar enrollment opportunities to students attending each session. A seat release system also hedges fall course predictions by partially filling each section over time rather than filling each section in sequence. The course prediction models aid in releasing seats to new students over summer orientation sessions by estimating seat need each session given the characteristics of the students expected to attend. The model we present also establishes how to release seats among multiple course sections and how to predict shortages and surpluses of seats.

Related Work

Most research published on enrollment modeling centers on the entire student body. There is little published work on course enrollment prediction and seat release systems. Hopkins and Massy (1975) present three categories of flow models used in university enrollment planning. They are (1) the grade progression ratio (GPR) method; (2) Markov chain models; and (3) cohort flow models. The GPR method and the Markov chain models use cross-sectional data and historical yields to predict the university population. The cohort flow models use longitudinal data to track students through the university and use historical yields for predictive purposes. They claim that the same techniques used for predicting on the aggregate level can be used at the department and course level, but do not develop the necessary models. Balachandran and Gerwin (1973) present three variable-work models for predicting course enrollments: the work model, the eligible-work model accounting for prerequisites, and the eligible-work model with program requirements. The authors claim that all three models can predict course enrollments semesters ahead using estimated probabilities, forecasts of new students each semester, and the initial distribution of students.

The work model uses the conditional probability that a student will take a course, given that he has not already, to predict how many students will enroll in that course. The eligible work model pares down the number of students eligible to take a course by accounting for prerequisites and then uses the conditional probability that they will take the course for predictive purposes. The eligible-work model with program requirements categorizes eligible and ineligible students based on the work they have completed in progress towards a degree. Conditional probabilities are used for each group to predict the total number of students who will enroll in the course in question. In testing of the models, the parameters—the conditional probabilities—were estimated using the average rates from one year, 1971. Note that each model is more detailed that the last and thus requires more data for parameter estimation. Balachandran and Gerwin (1973) test the work-model and the eligible-work model on graduate courses. Because the enrollments are low, the error rates are often high and the validity of the models is not clear. The authors concede that the models are not a replacement for intuition, but instead a complement.

The course enrollment model we present below was first presented in (Kraft 2005). Detailed information about the data and software used, and completed course models are available there.

Predicting Course Enrollment

Most of the course enrollment models currently used in the Department of Mathematical Sciences at Clemson University predict future populations using the previous semester's enrollments. There are exceptions where regression on historical aggregate data provides the model. Regression is often not appropriate because of the changing curriculum requirements on the number of hours required to graduate and the number of math courses required, as well as fluctuating fail and withdrawal rates. In order to isolate these variations we split students into groups with similar enrollment rates and identify groups whose rates have been fluctuating. In this study, we take advantage of more extensive data that includes detailed information on student characteristics and course histories that had not been used previously.

We aggregate students into groups instead of predicting on an individual basis because there is high variability in the latter approach. Some attributes relevant to course enrollment, such as major, divide the population into a large number of groups with few students in each. These fine divisions have too much variability to be useful predictors. We instead use broader categories to determine the significant student groupings. We use goodness of fit tests and past experience of enrollment behavior as our guides in determining which groups to use.

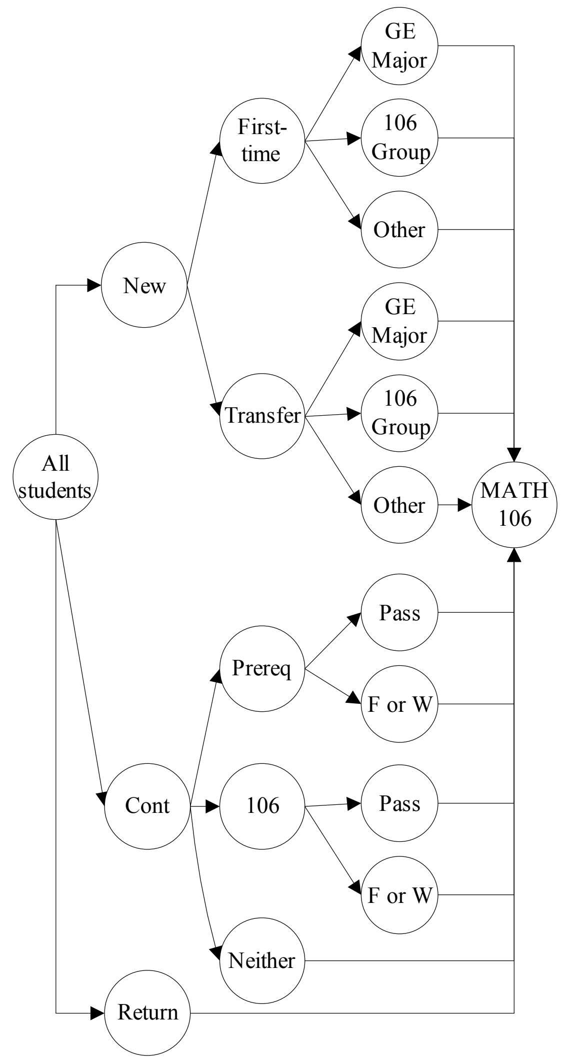

Figure 1: Template for Course Enrollment Model

Figure 1 shows the groupings of a Single Variable Calculus I (MATH106) model, where continuing students and returning students are denoted Cont and Return, respectively. The figure shows a decision tree that yields final groupings of students for predicting course enrollment. General engineering majors (GE Major), other students whose majors require MATH106 (106 Group) and all the rest of the first-time students (Other) are the final groups into which first-time students are divided for the MATH106 model. From the final groupings we record the conditional probabilities of students in the group enrolling in the course from historical data. Then we use this information to aid in estimating the conditional probabilities for the coming year. We multiply these proportions by the estimated university enrollment of the respective groups to determine total course enrollment. The number of nodes in the diagram is course dependent, and varies depending on the behavior of students enrolling in the course to be modeled.

Data

There are three sets of data required for creating the model we present. The first is the estimated new student populations, including first-time and transfer students. This information, which is usually a targeted number set by the admissions office, is not derived from any historical data. These estimates may be volatile and often change before the semester begins. For this reason we may want to use a range of estimates centered on the targeted numbers.

The second set of data we need is the current course enrollments. From these we obtain the number and characteristics of the students currently enrolled in each course, its prerequisites, and the rest of the university. Because we do not know the course pass and fail rates at the time of prediction, we must estimate these. We update our model after the grades are recorded at the end of the current semester to reflect the actual pass and fail rates and then adjust our predictions at that time.

The third and largest set of data needed for the model is historical enrollment data. There are 12,000 to 14,000 undergraduates enrolled at Clemson University (CU) in any fall or spring semester and 4,000 to 6,000 students enrolled in courses taught by the Department of Mathematical Sciences each semester. From these data we are able to determine which student characteristics and enrollments are significant predictors for modeling course enrollment. Because course descriptions, course pass rates, and program curriculums change, we make estimates for each of just a few preceding years (typically three) and combine these for the current year in light of these qualitative changes. Examples from CU of qualitative changes affecting course enrollment are the fall 2005 decrease in the required number of hours for graduation—from approximately 135 hours to just over 120 hours—and the decrease in the number of required math courses for some programs from two to one.

Thus far we have focused on models used for predicting fall enrollment. However, the models we develop can be used for both semesters. The differences are slight and primarily include differing sizes of student populations. The fall semester has a large new student population with many first-time students and course enrollment depends heavily on such admissions. New student populations are much smaller in the spring. Course enrollment predictions rely more heavily on students in prerequisite courses and the population of continuing students. In developing generic models adaptable for both the fall and spring, we typically include the significant groupings needed for both semesters.

The MATH106 model differentiates admissions estimates over transfer, first-time, and first-time general engineering students. It also depends on course enrollment counts from MATH106 prerequisites and MATH106 in the present semester. We include students who are in none of the previous groups in the MATH106 model by grouping them based on semester class since MATH106 is typically a freshman course. These estimates and counts along with the historical enrollment rates of students in these groups constitute the data needed for our model.

Procedure

The procedure we have developed for creating course prediction models has four major steps: (1) identification of significant factors; (2) parameter estimation and verification; (3) historical verification, and (4) model use.

Identification of significant factors: Identifying student characteristics that are significant in predicting course enrollment is the first step in creating course enrollment models. We do this by testing for goodness of fit between the grouping variable and enrollment in the course we are modeling. To keep the model simple, we only use the factors most significant in predicting course enrollment and exclude information that increases the variability of our model's predictions. For example, we have the major codes for each student and could use these individually to estimate rates of enrollment. Because there are over 100 majors, many such groups are small and prone to high variability in their enrollment behavior. Thus, instead of using major codes individually to group students, we group students with similar enrollment behavior. For the MATH106 model we group students as general engineering majors, majors required to take MATH106, or all other majors. We do not include general engineering majors with all the other majors requiring MATH106 because the admissions office provides an estimate of first-time general engineering majors that is not derived from historical data, but set by the university. We use this target for predicting enrollment in courses highly populated by first-time general engineering students. Furthermore, this MATH106 grouping yields better estimates than grouping all the first-time students together because first-time general engineering majors enroll in MATH106 at a significantly higher rate.

We hypothesize that students who pass the prerequisite course will enroll in the course at a higher rate than those who did not, students in the course who failed would re-enroll in the course at a higher rate than those who passed, and those students whose major required that they pass with a C or higher would retake the course at a higher rate if they earned a D or below in the course previously. Because we know that the general engineering majors must earn a C or higher in MATH106 to adhere to their graduation requirements, for the MATH106 model we divide GE majors who took the course into two groups: those earning a grade of an A, B, C and those earning a D or below.

		Conditional Enrollment Rates	
		% in Group	**% in MATH 106**
New Students		**F03**	**F03**
First-time	GE Majors	25%	55%
	106 Group	29%	29%
	Everyone Else	46%	8%
Transfer	GE Majors	10%	22%
	106 Group	24%	11%
	Everyone Else	66%	2%
Continuing Students		**S03**	**F03**
In Prerequisite (Previous Term)	Pass	61%	49%
	Fail or Withdraw	39%	14%
In Course & GE Major (Previous Term)	A, B, C	44%	0%
	D, F or Withdraw	56%	16%
In Course & 106 Group (Previous Term)	Pass	58%	5%
	Fail or Withdraw	42%	20%
In Course & Other Major (Previous Term)	Pass	58%	0%
	Fail or Withdraw	42%	0%
Everyone Else (Previous Term)	Sem 1, 2, 3	12%	1%
	Sem > 3	88%	0%
Other Students		**S03**	**F03**
All Students	Other Students	2%	1%

Table 1: University and MATH106 Enrollment Rates for Designated Groups (Fall 2003)

Many model changes are made while determining which grouping variables to include in the model. We want groupings that make logical sense and show promise empirically to be good predictors. The goodness of fit tests guide us in model creation. However, there are instances where we use grouping variables despite a test result indicating independence relative to the grouping. We may want the same models for both the spring and fall semesters. Although a split is significant in one, it may not be in the other. For example, because the number of new first-time students is high in the fall and very low in the spring, the goodness of fit test does not have enough power to indicate the same groupings are significant for the spring and fall MATH106 model. However, past experience shows that these groupings are good predictors empirically and so are incorporated into both semesters' models. Table 1 shows the final groupings for the MATH106 model.

Parameter estimation and verification: After determining groups whose course enrollment rates are significantly different than the rest of the population, our second step is to estimate the parameters of the model. To reduce the volume of data, we base our initial prototype model solely on data from one year by recording university and math sciences course enrollment rates for the identified groupings. When we use the model in practice, we first estimate the number of students in each of the groups and then predict the number of students who will enroll in the course from each of the groups. Thus, we gather both the university and course enrollment rates from each of the final groupings for one year. Table 1 gives the MATH106 model derived from 2003 data. Note that the university and course enrollment rates are conditional on being in the group on the left. For example 25% of the first-time students are general engineering (GE) majors and 55% of the first-time GE majors enroll in MATH106.

We test the accuracy of model predictions and the robustness of the model by using the estimated parameter values to predict course enrollment for other years whose results are known. For our example, we use the parameter estimates in Table 1 for MATH106 and test the model on data from fall 2002 and fall 2004. After grouping the data for the previous semesters, we count the number of students in each group enrolled in the university. For initial testing, we use the actual numbers of students in each group to test the course prediction rates. Recall that in the final model we must estimate these groups since complete information is not available at the time course enrollments are predicted. We multiply the parameter values by the counts to get the predicted number of students in the course for that semester. Table 2 compares the predicted and observed values.

SEMESTER	**F02**	**F04**
Actual Enrollment	792	937
Predicted Enrollment	802	915
Error	-10	22
% Error	-1%	2%

Table 2: Preliminary Test Results for MATH106 Model Derived from Fall 2003 Data

We also perform goodness of fit tests to ensure that the groupings significant in the fall and spring of the given year are also significant for the other years. After some model modification with the variables, we find the final groupings that are significant relative to course enrollment.

Historical verification: The third step of the course enrollment modeling procedure is verifying the model with historical data. The final model uses university and course enrollment rates from several years to determine parameter estimates. We use data from the previous three years. We do not include less recent data because of the changes in courses, retention, and university population. These qualitative changes undermine the worth of data that is older than a few years.

The distribution of course enrollment rates is unknown. To understand better the amount of variation in the rates, we calculate the minimum, maximum, and average historical rates for university and course enrollment over the range of years we have chosen. Table 3 shows the rates of new students in MATH106 for the fall semesters of 2002, 2003, and 2004. There are noticeable differences between the rates of enrollment of GE majors, the 106 group, and the rest of the population. Notice that the percentage of transfer students who are GE majors and enrolled in MATH106 changed significantly each of the three years recorded, while the percentage of first-time GE majors enrolled in MATH106 stayed relatively constant. Grouping students and observing rates over multiple years allows us to identify these differences.

	Enrollment Rates			Variation over F02–F04		
	F02	**F03**	**F04**	**Min**	**Max**	**Avg**
First-time Students						
GE Major	56%	55%	57%	55%	57%	56%
106 Group	31%	29%	30%	29%	31%	30%
Else	7%	8%	8%	7%	8%	8%
Transfer Students						
GE Major	8%	22%	32%	8%	32%	21%
106 Group	9%	11%	8%	8%	11%	9%
Else	1%	2%	2%	1%	2%	2%

Table 3: Course Enrollment Rates for New Students in MATH106

To complete the historical verification step, we test the model using both the university and course enrollment rates derived from the recent data. We test the model on past data as it would have been used for predicting future enrollments. Thus we use only the information that we would have known at the time the schedule was submitted. We estimate university enrollment for each group in our model that we can not count or obtain at the time of prediction. For the MATH106 model we count from the historical data the numbers of first-time, transfer, and continuing students in their respective groupings, such as those who are in MATH106 and are GE majors. We then estimate the number of students in the final groupings by multiplying the average university enrollment rates derived from historical data by these counts. In the MATH106 model we estimate the number of students who pass the prerequisite by multiplying the number of students in the prerequisite by the average pass rate. Finally, we predict the course enrollment from each group by multiplying the average course enrollment rates with the group estimates. Summing the number of students expected to take the course in each group gives us our total prediction.

The MATH106 model predictions shown in Table 4 were within one section of the actual enrollment for each year we tested, with error between 10 and 19 students. If the error for a model is large at this point in the procedure, the model should be changed before continuing to the prediction step. High error may be a sign that the groups in the model do not capture the significant groups of students who enroll in the course. High error rates may also indicate that the course has changed over the years of data used and a shifting mix of students is enrolling in the course.

SEMESTER	**F02**	**F03**	**F04**
Actual Enrollment	792	860	937
Predicted Enrollment	808	870	918
Error	-16	-10	19
% Error	-2%	-1%	2%

Table 4: Final Test Results for MATH106 Model

Model usage: The final step of the procedure is using the model for course enrollment prediction. In predicting future enrollment we follow the procedure used for final testing, except that we do not automatically use the average rate from the historical data for predictive purposes. Instead, we observed the minimum, maximum, and average historical rates multiplied by the current populations as well as any previous trend that is apparent in the absolute enrollments. We make our estimate for each group based on these factors. Table 5 shows the university and course predictions for first-time GE students in MATH106 for fall 2005. The numbers that appear below the term date are the past enrollments and enrollment rates of first-time GE students in the university and in MATH106, respectively. The numbers in the fourth through sixth columns of the university enrollment row are the past minimum, maximum, and average rates of enrollment multiplied by the admissions estimate of 2800 new first-time students. Because a target is set for the number of first-time GE students, we use that estimate of 725 rather than the average of the previous years. However, the percentage of first-time GE students who enroll in MATH106 has been relatively constant over the past three years. Our estimate for first-time GE MATH106 enrollment reflects that trend by choosing a rate similar to the historical rates.

We make estimates using the same method for each group in the course enrollment model. The sum of these estimates yields the total prediction for the course. We can also find the sum of the minimum, maximum, and average estimates to give perspective to the number we ultimately decide to use. To calculate the variance of our estimate, we collapse the decision tree for our model groupings to find the probability that a student will enroll in the course, given no information about which group the student is in. We treat whether or not a student enrolls in the course as a Bernoulli random variable. As such we can estimate the variance and construct confidence intervals for our prediction.

First-time Students (Target of 2800 for F05)						
Enrollment Rates			**Variation over F02–F04**			**Use**
F02	**F03**	**F04**	**Min**	**Max**	**Avg**	**F05**
University Enrollment (GE Major)						
660	696	734	24%	27%	25%	26%
27%	25%	24%	680	744	710	725
MATH106 Enrollment (GE Major)						
369	381	421	55%	57%	56%	57%
56%	55%	57%	397	416	406	410

Table 5: Model for University and MATH 106 Enrollment of First-time General Engineering Students

Prediction	926
Observed	889
Error	-37
% Error	-4%
80% Confidence Interval	[889, 963]
95% Confidence Interval	[870, 982]

Table 6: MATH106 Prediction and Result for Fall 2005

The prediction we made for fall 2005 MATH106 enrollment is given in Table 6 along with the 80% and 95% confidence intervals. The actual enrollment of 889 is 37 students—about four percent or one section—less than our initial prediction. The actual enrollment is captured by both the 80% and 95% confidence intervals, and is equal to the lower bound on the 80% confidence interval. The error is in a reasonable range. We used the course prediction model on three other large, primarily freshman courses and the error rates of our predictions for those courses were also low.

We have built a model to predict course enrollment that uses student characteristics as predictors of enrollment behavior. We use the model on an aggregate level to predict the total number of students enrolling in a course. The model is sufficiently detailed to predict on a smaller level given that we have student characteristics for the population. In the next section we illustrate how to apply the course prediction model to allocating course seats during orientation sessions.

Allocating Course Seats

Predicting course enrollment for fall semester does not end when schedules are submitted early spring semester. At Clemson University, early registration for continuing students begins in April, semester course grades become available in May and orientation sessions for new students begin in June. Course grades, early registration information, and information on students registered for orientation sessions increase the accuracy of the fall course predictions. We update fall enrollment predictions as pertinent information becomes available. Many courses populated largely by new students have seats reserved for them which are released gradually over the summer orientation sessions. Releasing seats is appropriate for courses that have many sections and significant new student enrollment. The release of course seats during the multiple orientation sessions depends on the fall enrollment predictions. In this section, we present a new model for seat release that addresses how to estimate seat need each session, how to allocate seats for multiple course sections, and how to predict seat shortages and surpluses.

Data

The seat release system we present requires three sets of data. The first is produced by the model we created for predicting course enrollment. The model specifies the groupings that are significant and gives conditional enrollment rates for these groups. The second set of data needed is the current enrollments in the course by section. Note that when orientation begins, continuing students will have already had the opportunity to register for courses during early registration.

The third set of data we need is the number and characteristics of students attending each orientation session. All new students are required to attend an orientation session. Most do this during the summer, although there is a late orientation session held just prior to the start of the semester. Since the attendance at each orientation session is capped, students must pre-register to attend a session. Although there are additions (walk-ins) and deletions (no-shows), these numbers are typically small. Hence, the number and characteristics of students attending an orientation is very accurate just prior to that orientation, but less so for future sessions because the orientation registration process may not be complete. Thus, after each orientation session we update the orientation and course enrollments to obtain current information on the change in orientation enrollments and the number of remaining course seats.

We group students attending each orientation session by their characteristics as we did in the course prediction model. The numbers of students in each group are significantly smaller than the numbers we encountered in the course prediction model since we are predicting for each of several orientation sessions.

We use data from students registered for every session, not just the current session. Although we have less confidence in the estimates for more distant sessions, the counts for the future sessions help us predict the overall number of seats we need and allow us to confirm or alter our previous predictions.

Procedure

Ideally our predictions made in the spring would be exactly realized, and we could use those predictions to reserve seats for new students. We could then set section capacities based on the seat need predictions for continuing students to allow them to register in the spring and could release seats uniformly to new students during summer orientation sessions. An example of an ideal system is given in Table 7. The total prediction for the course is 200 students—40 continuing and 160 new. During spring registration we allocate seats uniformly for 40 students among the four sections—ten per section. Since our predictions are correct and the students register promptly (since they are ideal), all continuing student have taken their seats in the course by the first orientation session and we have only to release seats for new students. In our ideal scenario, students attend orientation sessions uniformly and we therefore release seats in the same manner—eight per session per section. Table 7 shows the capacities of the sections and reflects the cumulative release of seats. (In practice section capacities are not this high and a seat release system may not be appropriate for this example because there are only four sections.)

Obviously this ideal situation is not realistic. At least four sets of imbalances prevent the system from working so smoothly: (1) we cannot allocate the exact number of seats needed since we do not know that number; (2) the attendance at orientation sessions is uncertain and therefore must be estimated; (3) our seat release is not uniform across sections because enrollments are not uniform and room capacities differ; and (4) some sections are favored by students (e.g. not eight AM). Our seat release system shows how to overcome these imbalances to obtain good estimates of seat need and allocate seats to students.

Session	**Section Capacities**			
	1	**2**	**3**	**4**
Spring	10	10	10	10
1	18	18	18	18
2	26	26	26	26
3	34	34	34	34
4	42	42	42	42
5	50	50	50	50

Table 7: Ideal Course Release, Initially 10 and Increasing 8 per Session per Section, for a Course with Four Sections over Five Orientation Sessions

Terminology: Course *reserve* is the number of seats reserved for a certain group, such as all, new, or continuing students. Course *capacity* is the number of seats currently available and is typically increased as orientations progress to make room for incoming students. The *total capacity* may not exceed the *total reserve*. *Seat release* refers to making available seats in a given course by changing section capacities.

Initialization: Preparation for summer orientation seat release begins in the spring prior to early registration. At this time we use our course enrollment predictions for new and continuing students to set course reserves. We reserve a portion of seats in each section for new students and the remaining is reserved for continuing students. During early registration we set the section capacities to make available only enough seats to accommodate the continuing students. That is, we release all the continuing student seats at one time since they have access to registration at the same time. We only release seats for continuing students so that the residual capacity is held for new students who will register later in the summer.

Before orientation sessions begin, we update our course predictions using updated admissions information and pass rates that have now been recorded. We also observe and reset, if necessary, the section capacities based on continuing student registrations. If our continuing student prediction was too high, we may need to lower the continuing student reserves and the section capacities.

With the number and characteristics of students registered for each orientation session and the course enrollment model, we group students into the groupings detailed in the model. Table 8 shows an example of first-time and transfer student orientation enrollment using the MATH106 model groupings in preparation for the third orientation session. The shaded columns represent transfer sessions. Note that we may have first-time and transfer students attending the same session, but for simplicity we will keep them separate in our example. Suppose there are three summer orientation sessions (shown in the first three columns) and there are two make-up sessions (shown in the fourth and fifth columns). Since in our example we are preparing for the seat release of the third orientation session, the registrations for the two make-up sessions—held much later in the summer—are low in comparison to the regular session enrollments because students have yet to register for an orientation session.

	Orientation Session					**Total**	
Student group	**1**	**2**	**3**	**4**	**5**	**First-time**	**Trans-fer**
GE Major	27	12	25	4	2	56	14
106 Group	40	16	34	18	6	92	22
Other	45	46	53	12	16	110	62
TOTAL	112	74	112	34	24	258	98

Table 8: Example of Orientation Enrollment Prior to Session 3 by Group

	Orientation Enrollment Predictions		**Estimated Number of Students Yet to Enroll**	
Student group	First-time	Transfer	First-time	Transfer
GE Major	73	23	17	9
106 Group	109	29	17	7
Other	145	94	35	32
TOTAL	327	146	69	48

Table 9: Example of Enrollment Predictions and Orientation Registration Estimates

Using our predictions for total fall enrollment, we estimate the number of students, by group, who have not yet registered for an orientation session by calculating the difference between the group prediction and the sum of orientation registrants so far in that group. For our example the expected number of first-time GE majors yet to register for orientation is 73—shown in Table 9 as the prediction for first-time GE majors—minus 56—shown in Table 8 as the total number of first-time GE majors already registered for orientation. This difference of 17 is shown in Table 9 as the estimated number of first-time GE majors yet to enroll. As orientations sessions progress, these differences will get smaller. If they become negative we have underestimated the number of students in that particular group and must make adjustments. If they remain large, we have overestimated and also need to adjust. Observing these totals will help us gauge the accuracy of our predictions and help us make adjustments as soon as corrections become apparent.

If the university using the seat release model has an extended period of time between the last summer orientation session and the make up sessions, then during that interim time updates of the number of students expected for the make-up sessions can be made. In the interim period, some transfer students may forgo the orientation process and register for courses. We need to identify these students and exclude them from the total number that we expect to attend the late orientation sessions. As these students are identified, we subtract them from the estimates of students yet to register to keep an accurate count of the students that we still expect to register for courses.

	Orientation Session					**Estimated Number of Students Yet to Enroll**	
Student group	1	2	3	4	5	First-time	Transfer
GE Major	15	3	14	2	0	10	2
106 Group	12	2	10	5	1	5	1
Other	3	1	4	1	0	3	1
Total	30	6	28	8	1	18	3
95% CI	40	13	37	14	6	24	9

Table 10: Example of expected seat need by group by session

Estimating seat need: Just before each orientation session, we update course and orientation enrollment numbers so that we have current information. Using the recorded counts of students attending each session, we estimate the number of course seats needed for each group in the current and future orientation sessions using the course enrollment model. We construct a confidence interval for our estimate using the same method used in the course enrollment model prediction. An example of these calculations is given in Table 10, where the last two columns give the expected seat need for students who have not yet registered. The last row in Table 10 gives the upper bound on the confidence interval. To hedge our estimate, we may want to release slightly more seats than we expect to need.

Adjusting section capacities: We make seats available by adjusting section capacities using our estimates of seat need and the current course and section enrollments. The course enrollments include both continuing and new students, and thus adjusting section capacities is the method for releasing seats to both groups.

Continuing students have had access to course seats since early registration and many of the seats intended for them will be filled prior to new student orientation. Because new students may take seats intended for continuing students if they are open, we take any empty seats intended for continuing students back prior to the first orientation session and then gradually release them over the summer. In this way, we are still giving continuing students opportunities to register and change sections, but we are also protecting their seats from new students.

We use the estimates we have calculated to release seats for new students and use a proportion—say session number over n, where n is the total number of sessions—to release the seats held for continuing students. Thus, at each orientation session, we reset the total capacity in each course section to allocate seats for new and continuing students.

Unlike our ideal example given in Table 7, almost certainly current section enrollments will not be uniform. Therefore, we increase the capacity in each section proportional to the number of remaining seats available. Since new and continuing students have differing numbers of seats intended for them based on the course model predictions, we calculate their releases separately and then adjust the section capacities using the two allocations. We use the estimated new student seat need divided by the total number of remaining seats for new students as the proportion of seats to release from the seats in each section remaining for new students. The expected release for new students, then, is simply their estimated seat need. To hedge against a lack of available seats for new students, we may choose to use the upper bound on an appropriate confidence interval instead of the estimated seat need.

After each orientation we update the course enrollment numbers. When we adjust the section capacities for the next release, we use these current enrollment numbers, not the capacities we previously set. That is, the allocation of seats is always based on current enrollment and predicted needs. Hence, over or under estimation of needs in a single session is not cumulative. All empty seats, whether there are more or fewer than we expected, are incorporated into future allocations through the enrollments at the time of release.

Using the seat release system: The seat release system we have presented has three main components: estimation, allocation, and recovery. We estimate seat need for each session using the course prediction model; we allocate seats using proportions of remaining seats; and we recover from hedging or unforeseen demand by using current enrollments instead of prior predictions.

	Section	**1**	**2**	**3**	**4**	**TOTAL**
Initialize	New Reserve	48	32	40	40	160
	Continuing Reserve	12	8	10	10	40
	Total Reserve	60	40	50	50	200
Update each session	New Enrollment	21	11	19	14	65
	Continuing Enrollment	8	5	7	9	29
	Total Enrollment	29	16	26	23	94

Table 11: Example of Course Enrollment Data Needed Each Orientation Session

We extend the example from Tables 8, 9, and 10 to finish the illustration of the seat release system. Table 11 shows the course data needed for initializing the seat release system and the course data needed to be updated just before each session. The initialization includes recording the reserves for new and continuing students—shown as new reserve and continuing reserve, respectively in Table 11. The sum of these makes up the total course reserve. Each session we update the current enrollment for new and continuing students. This current information helps us in allocating seats for the next orientation session and may alert us of potential problems in our course enrollment predictions.

The calculations for new and continuing student seat release are shown for our example in Table 12. There the remaining continuing seats and the remaining new seats are shown. We calculate the remaining number of seats by subtracting the current enrollment from the appropriate reserve. (Calculating remaining seats may be more involved depending on the enrollment rules at the university using the system.) Our example is for the third of five orientation sessions, so we release three fifths of the remaining continuing student seats. The proportion of remaining new student seats we release is 37—the upper bound on a 95% confidence interval shown in Table 9—divided by 95—the total number of remaining new seats, which shown in Table 12. The updated capacity is the total current enrollment plus the continuing and new student releases. Table 12 shows these sums for our example.

Section	1	2	3	4	TOTAL
Total Enrollment	33	25	30	37	125
Remaining Continuing Seats	4	3	3	1	11
Continuing Release (proportion = 3/5)	2	2	1	1	6
Remaining New Seats	27	21	21	26	95
New Release (proportion = 37/95)	11	8	8	10	37
Updated Capacity	42	26	35	34	138

Table 12: Example of Adjusted Section Capacities after New and Continuing Student Seat Release

Session	Expected Enroll-ment	Seat Release	Actual Enroll-ment	Difference
1	53	64	50	14
2	8	17	7	10
3	49	60	35	25
4	48	59	38	21
5	41	52	35	17
6	50	61	47	14
7	7	15	13	2
8	48	59	43	16
9	45	56	26	30
Interim	10	16	-2	18
10 & 11	19	25	8	17

Table 13: Simulation Results for MATH106 Seat Release Using 80% Confidence Interval

Simulation results: We simulated the seat release model on MATH106 data from the summer 2005 orientation sessions at CU. There were thirty-one sections of MATH106 offered and nearly 900 enrolled students. The new student orientations at CU begin in mid-June. The 2005 orientation schedule included nine sessions during the summer—seven for first-time students and two for transfer students. Two make-up sessions—one for transfer students and one for first-time students—were held just prior to the start of the fall semester. Table 13 shows the results of this simulation where we release seats using the upper bound on an 80% confidence interval of our estimate. The actual enrollment is the net number of students who enrolled or dropped the course. The seat release system gave good results for MATH106. Similarly good results were obtained for three other large, primarily freshman math courses.

Conclusions

The course enrollment model and seat release system address practical problems that all universities face. Predicting course enrollment is a difficult problem because of the qualitative changes in the courses and university regulations. Departments serving large portions of the university especially need accurate, robust models for predicting course enrollment. We presented an adaptive course enrollment model that uses student characteristics to identify student groups whose course enrollment rates are significantly different than the rest of the university population. The model then uses historical enrollment rates and current information to predict enrollment for the coming semester. The model has proved to be accurate. It is also robust in that it can be extended to any course by using appropriate student characteristics. The procedure we presented for creating the model is straightforward and accessible.

The seat release model we presented uses the course prediction model for estimating seat need each session and is a realistic approach to managing course seats during multiple registration sessions. The system of estimation, allocation, and recovery allows for imbalances in orientation attendance and course registration, which inevitably occur. The simulation results indicate that the model will work well in practice. These procedures will be used for the major freshman courses for the Department of Mathematical Sciences at Clemson University for the fall 2006 term.

References

Balachandran, K. R. and Gerwin, D. 1973. Variable-Work Models for Predicting Course Enrollments. *Operations Research* 21: 823-834.

Hopkins, D. P. and Massy, W. F. 1981. *Planning Models for Colleges and Universities*. Stanford University Press.

Kraft, C. R. 2005. Predicting Course Enrollment. Master's paper, Department of Mathematical Sciences, Clemson University. http://people.clemson.edu/~cstemm/course.pdf.