Repairing model-drift in enrollment management

AE Rodriguez
University of New Haven

John Rosen
University of New Haven

ABSTRACT

The various empirical models built for enrollment management, operations, and program evaluation purposes may have lost their predictive power as a result of the recent collective impact of COVID restrictions, widespread social upheaval, and the shift in educational preferences. This statistical artifact is known as model drifting, data-shift, covariate-shift. Succinctly, these events drove changes in the stationarity of the target variable and the predictors. The result is a student body with unknown performance qualities entirely distinct from previous cohorts. This study explains and illustrates: (i) how to test for academic model drift in academe, and (ii) sets forth two methods used to repair vitiated student-body performance properties. Formally, it frames the data-drift outcome as a One Class problem which allows the deployment of two well-known One-Class algorithms: Support Vector Machines and Isolated Random Forests. The study shows their use in reconstructing a representative sample of the student-body.

Keywords: Student retention, Enrollment management, Model-shift, Data-shift, One-Class, Isolation Forests, Support Vector Machines.

"It is always dangerous to talk about probabilities without understanding the processes which generate the observed data." John Kay and Mervyn King, Radical Uncertainty, 2020

INTRODUCTION

At least three inter-related, significant events over the last couple of years have critically dented long-standing higher-education processes and procedures – in the United States. COVID restrictions, social upheaval, and a not insignificant shift in student preferences and ensuing increase in demand for non-traditional education have combined to stem the flow of conventional students (by historical standards) heading to colleges and universities.

First, the societal and economic COVID disruptions resulted in a total reset of established college practices and procedures. COVID-era proscriptions shut down colleges, dormitories, closed testing centers. COVID related turmoil appears to also have altered individual student preferences, perceptions of risk, value proposition, and expectations associated with attending college (Wingard, 2022). Obligatory masking, vaccination protocols and social-distancing directives may have influenced student opinions as to the health risks of in-person attendance.

Second, investment horizons appear to have been reduced and social discount rates increased. These shifts appear to have increased the appeal of the current vis-a-vis the future – resulting in many prospective students bypassing a traditional college education altogether. The move towards online education, skills-based training, and affordable credentials that take weeks or months to achieve rather than years appears to be increasing (ICEF Monitor, 2020). As one recent commentator noted, "there are mounting factors that dissuade prospective students from making a large investment in degrees and instead choosing to go with online alternatives to traditional higher education" (Schroeder, 2021).

Last, the George Floyd uprising and related social upheaval led to increased concern over possible social and racial inequities and unfairness built into the traditional metrics used for higher education admissions, financial aid – even social standing. To be sure, pressure on selective colleges to increase their enrollment of low-income and first-generation students has always existed. The No Child Left Behind Act of 2001 with its focus on "low-income" and "major and ethnic subgroups" propelled a long-standing debate about schooling quality and equality that continues to this day (Congress, 2001).

In short, the very expensive and lengthy higher-education path of yesteryear appears to have emerged the worse for wear (Whitford, 2022). The reality is that colleges and universities face a dwindling prospective student funnel (National Student Clearinghouse Research Center, 2022) – beyond the long-awaited and much feared "demographic cliff" (Conley, 2019) (Grawe, 2018). This downturn is poised to persist for several years.

College admissions teams have improvised and adapted. Admissions criteria were loosened, supplemented, or replaced entirely by considerations finely tuned to public sentiment and opinion over social and racial disparities. Entrance exams such as the GRE and the GMAT were rescinded entirely or rendered optional (MBA Rountable, 2022). GPA, honor courses and other conventional academic signals were de-emphasized and supplemented with a sensitivity towards less traditional, more subjective schooling experiences.[1] The changes adopted reduced costs of admissions and diminished the information content provided by test scores and GPA's. (Selingo, 2018); (Marcus, 2017).

The net impact on enrollments as result of the changes in admissions protocols is yet to be determined. What it most assuredly did do was eviscerate the historical admissions model, protocols, and processes. Internal appraisals of program performance were similarly vitiated. These systems' effectiveness relied on a relatively stable, homogenous student body, year after year (Saa, Al-Emran, & Shaalan, 2019) (Mengash, 2020). Extant admissions models, institutional analytics, academic offerings, operational rules and practices, and program evaluations were calibrated and trained – implicitly or explicitly - on well-known, stable, target variables and relied on a relatively stationary distribution of selection variables - covariates (Engler, 2021) (Ekowo, The Gift that Keeps on Giving: Why Predictive Analytics are Probably on Colleges' Wish Lists this Year, 2015). These stable features may no longer exist.

This study identifies this artifact as a data shift problem that may have resulted in a student body with blurred qualifications and quality profiles. This condition leaves admissions, operations and institutional analytics essentially blind, handicapping planning, recruitment, assessment, retention, and other key procedures. The proposed solutions provides a means to overcome this situation and illustrates the generality of the proposed solution by simulating conditions meant to represent the situation.

WHAT IS THE DATA DRIFT PROBLEM AND WHAT ISSUES IT HAS WROUGHT?

Colleges consider analytics to be the use of data, statistical analysis, and explanatory and predictive models to gain insight and act on complex issues. Analytics are deployed to improve services and business practices across the institution and to enhance or improve student success (Yanosky & Arroway, 2015) (Ekowo & Palmer, The Promise and Peril of Predictive Analytics in Higher Education, 2016).

In the world of analytics there is an ever-present irritant known as model drift or model shift, data distribution shift, concept shift, covariate shift (Cloudera Fast Forward, 2021). The terms are interrelated and may be distinct but are often used interchangeably. These are known, generically, as "dataset-shifting." Each concept is discussed in more detail below although the use of the term dataset-shift is adopted throughout.

Conceptually, the issue is a straightforward one. If you formulated a predictive model or operational practice on specific variables (covariates) and a specific class (target variable) reflecting an implicit, or explicit, decision-making criteria, then the predictive capabilities of the model will decline as the underlying variables change or as the target variable changes (or clearly, – if both change). It is seemingly common-sensical, "Predictive models should also be updated or refreshed to reflect new campus realities and goals." (Ekowo & Palmer, Predictive Analytics in Higher Education: Five Guiding Practices for Ethical Use, 2017). What is not as straightforward to parse are (i) the seeming changes in behavioral peculiarities embedded in admissions selection programs, and (ii) the distortions in predictive capabilities because of the unintended data-shift. Both artifacts resulting from the exogenous shocks discussed here.

[1] There is evidence of considerable GPA inflation over the last few years (Sanchez & Moore, 2022). This has resulted in an unstandardized and potentially problematic way to compare students across the country and over time. Left uncorrected, this feature alone could have a similar data-shifting impact. We do not account for GPA inflation here.

The Biasing of Selection Bias

As a general matter, in a higher-education institution there are at least two potential sources of selection bias in admissions that may obscure prospective student appraisals, student performance, and other similar processes that rely on student quality and performance metrics. The issue of concern here is that one can plausibly infer that as a result of the three exogenous events described above, the *incentive* structure underlying these behavioral biases – may have changed as well. The problem is – it is not clear in which direction the shifting tilts.

First, there is a positive correlation between attending a specific university program and subsequent student performance. Students admitted into a program are likely to benefit from a college-wide or program-level success ecosystem. A success ecosystem constitutes the systems, tools, and human resources, inter alia, that a are devoted to academic tutoring and assistance, pre-program learning, retention programs, counseling, social-networking and other features deemed critical for student success. A higher-education program has every interest in the success of its students and is likely to bend over backwards to ensure accepted students make progress towards their degrees. Retention is key for academic, financial, accreditation, and reputational reasons.

As a higher number of students attend a college, it becomes costlier for the institution to finance the success ecosystem. Costlier because the institution must devote more hard and soft resources to ensure the progress and success of ostensibly a higher number of at-risk students (Westrick, Marini, & Shaw, 2021). Hard resources would be increasingly elaborate pre-program offerings, additional tutoring or remedial services; soft expenditures would be the additional faculty and staff time and effort devoted to help less talented students stay abreast. Success rates are likely to fall. The erstwhile positive correlation between program attendance and success is likely to decline.

Colleges are interested in identifying students at greater risk of failing or not completing their program of study and intervening before it happens. Administrators typically use learning analytics: statistical analysis of data gathered on students to better support education outcomes (Jia & Maloney, 2015) (Chui, Fung, Lytras, & Lam, 2020). However, the dataset-shift has blurred the historical "markers" identifying those students more likely to benefit. The correction proposed here may better profile the students and allow administrators to target attention and services.

Second, in yesteryear, students would self-select into opting to go to college based on an awareness of their meeting admissions thresholds such as GPA and college entrance exam scores. In other words, those students who sported good academic records were more likely to apply to college, most likely calibrating their school-choices by the various quality signals available: ACT/SAT thresholds, admissions ratios, placement rates and others. That is to say – the better students were the more likely to let themselves be considered for selection. Put differently, poor students or students with poor academic records were unlikely to consider applying to selective programs.

This self-selection process has probably been distorted - lately. The rescinding or ameliorating of admissions criteria – which historically acted as quality signals - has empowered less qualified students to apply. And perhaps simultaneously – the elimination of quality signals may have reduced the incentive of higher academic-quality students to apply. Or, the better students may choose to apply to the more selective schools who still manage to retain and enforce erstwhile admissions conditions.

The net effect of these self-selection incentives in combination with the operational changes discussed above is that all or nearly all of students currently in college programs are “one-class” students – where quality and performance are unknown. Many of those students who would otherwise have failed have been artificially shifted into the one category. Without a clear understanding of the student quality spectrum, one is confronted with a host of problems.

The unavailability of at-risk gradients confounds assessment both for purposes of the traditional student ranking and recognition and the assessment of pedagogical performance required of accreditation bodies. Another problem is that one is unable to train an unbiased admission/performance model with the existing data. Most classification algorithms require instance of both classes of a binary variable to work. Since all of students belong only to one class, one has no way of discriminating. In the following section contains a discussion of one approach to overcoming this problem – known, unsurprisingly, as the One Class problem. A comparison of two popular algorithms is used to address the issue of unlabeled classification data – as is the case here.

THE STUDENT BODY: A SIMULATION

With the elimination of entrance exams such as the ACT and the SAT, high-school GPA and life experiences became the go-to measures. A data generating process not atypical of those found in previous studies that rely on simulation is developed here (Austin & Schuster, The Performance of Different Propensity Score Methods for Estimating Absolute Effects of Treatments on Survival Outcomes: a Simulation Study, 2016), (Austin & Small, The Use of Bootstrapping When Using Propensity-Score Matching Without Replacement: a Simulation Study, 2014), (Setoguchi, Schneeweiss, Brookhart, Glynn, & Cook, 2008). For each student, two covariates, GPA, and a Personal Rating variable were generated from known data-generating processes.

Personal Rating is a student assessment that takes multiple characteristics into account to evaluate a candidate. These characteristics may include the value attributable to letters of recommendation, community-service, high-school selectivity, demographics, and other observable and unobservable factors that may vary by admissions office. And they may also include subjective factors such as “perceived leadership, maturity, integrity, reaction to setbacks, concern for others, self-confidence, likeability, helpfulness, courage, kindness, and whether the student is a ‘good person to be around.”[2] Some variant of this individualized, holistic admissions paradigm is commonplace through college campuses (Coleman & Keith, 2018).

The next step is a simulation of a hypothetical student admission pool. Admission occurs over two periods (“0” and “1”), presumably to straddle admission prior to, and after, the exogenous, causative events describe above. Both covariates were scaled; and the variables were inversely correlated – as can be seen in Figure 1 (Appendix).

Admissions propensity is a function of the two variables; the associated propensity to Succeed is determined by both variables in the first period and by GPA alone in the 2nd period. Presumably, Success is not known in the 2nd period and is therefore the variable that needs to be identified via the One-Class algorithms.

Assuming a power distribution to describe the typical student GPA admissions pool reflects the typical reality for most competitive schools where high-performing GPA students are fiercely recruited. The correlation between the variables is set arbitrarily to represent a tradeoff between higher GPA and Personal Rating; obviously, these would

[2] This characterization is based on the personal rating variable utilized by Harvard in its admissions process as set forth in Students for Fair Admissions v. Harvard College. (Keane, et al., 2022)

vary by institution. The results here are not sensitive to the correlation between the variables – as long as some positive correlation exists.

Admissions Considerations

Admissions in period zero depends on GPA and Personal Rating. The school sets its GPA admissions threshold at the 75th percentile thereby admitting all students in Period 0 who meet that threshold into the pool. The admissions propensity is then determined by a weighted average of GPA and Personal Rating. The admissions decision is randomized as determined by a Bernoulli process.

In period 2, the school has to accommodate the changes brought about by the events discussed above and reacts by lowering its admissions threshold to the 25 the percentile and increases the weight ascribed to Personal Ratings in the admissions decision:

Figure 2 (Appendix) creates a visual display of the covariate displacement that occurs. The GPA distribution shifts leftward between Period 1 and 2 whereas the Personal Ratings distribution shifts towards the right.

TESTING FOR THE PRESENCE OF DATASET DRIFT

Dataset drift means that the properties of the target variable – Student Success - change over time. As a general rule, this instability causes problems because the predictions become less accurate and potentially unreliable.

Dataset drift is the change in the distribution of one or more of the independent variables or input variables of the dataset. This means that even though the relationship between covariates and target variable may remain unchanged, the distribution of the covariate itself may have changed. When statistical properties of this input data change, the same model which has been built before will not provide unbiased results. This leads to inaccurate predictions.

There are various ways to test for dataset drift. One approach is to build a classifier model to determine whether it can distinguish between the reference (Period 0) and compared distributions (Period 1). The process entails the following steps.

First, one creates a dummy variable set to 0 for the original data. And symmetrically, label the identifier dummy 1 for the new batch of students.

Second, deploy a model to discriminate between the two groupings. Here a simple naïve bayes model to discriminate between the two groups is fit (Kuhn & Johnson, 2013) (Shmueli, Bruce, Yahav, Patel, & Lichtendahl Jr., 2017). If the naïve bayes model can easily discriminate between the two sets of data, then a covariate shift has occurred.

On the other hand, if the model cannot distinguish the two sets, then it is fair to conclude that a data shift has not occurred. An accuracy of approximately 50 percent suggests an outcome no better than a random flip of a coin.

The accuracy of the Naïve Bayes test registers at 97 percent indicating a significant shift between the two periods.

REPAIRING DATASET DRIFT DISTORTIONS

It has been shown above that increasing enrollments by simultaneously lowering admissions thresholds and accepting a broader role for measures such as Personal Rating may result in a distortionary data shift. Such a shift in the distribution of features affects learned classification outcomes, reducing or completely vitiating model performance and any operations or academic program dependent on a careful appreciation of student profiles.

Take, for instance, the emphasis on retention where the key to success is to assist the students who need academic support. Retention properly trained at students who would benefit the most requires identifying these students: those who are most likely to fail. The propensity to succeed measure of the first Period cannot be used in Period two given the non-stationary data-generating process. Put differently, in the second period one doesn't know who is, and who isn't, likely to succeed.

But it is possible to approximate those likely to succeed using grouping algorithms known in unsupervised learning as One-Class algorithms (Khan & Madden, 2004). In the terminology of machine learning, statistical techniques for classification are referred to as "supervised" and "unsupervised" learning methodologies. Supervised learning can occur when the data supplies both dependent (target group or class membership identification) and independent (predictor) variable observations. Unsupervised learning classification algorithms, on the other hand, derive classification results using "unlabeled data" where no known and verified dependent variable is available.

The use and capabilities of two such unsupervised learning algorithms is demonstrated: Support Vector Machines and Isolation Forests. One-class algorithm step in when one has no information on the classification label: put differently, whether the student is likely to Succeed.

One Class Support Vector Machines

Supervised learning is not possible when data are all grouped into one class as is the case here. It is necessary to find natural clustering of the data to group or classify. Support vector machine is an unsupervised learning algorithm derived from statistical learning theory.

A SVM One Class classifier learns the boundary between groups by maximizing the margin, or distance, between class members. SVM sets forth all available data as members of its first group, C_1 and the origin as the sole member of the second group, C_2. The hyperparameter, *v*, constitutes a penalty applied to the trade-off between groups one and two. The chosen kernel is key – and they include linear (inner-product), polynomial and sigmoid. Kernels help to determine the shape of the vector, plane or hyperplane and decision boundary between the groups. In this instance the fitting process is simple and relies on a line to separate our two Classes. The basic linear kernel function is selected:

$$K(xi, xj) = xi*xj + c$$

where x and y are input vectors and "*" represents the dot product and c is a constant. The prediction utilizes the two known covariates. The classification results are compared to the previous generated Success classification scores obtained from the simulated propensity to Succeed; Success was generated via randomized Bernoulli process. The results of the SVM fit and classification results are evaluated using the R-package caret's confusion matrix function:

The results return an Accuracy score of 75 percent a relatively acceptable performance in identifying the students likely to be successful.

Isolation Forests

Isolation Forest is an ensemble learning method applicable to one Class problems. It explicitly isolates outliers rather than learn a model for normal instances.

A normal sample is hard to isolate from other samples. An outlier is more easily detected from other samples. Isolation Forests is composed of a fixed number of isolation trees each one built on a random selection of samples from the training set, the forest. From this subset of samples, an isolation tree is constructed by a random recursive partitioning, until all the samples are isolated or until a stop criterion is reached.

The partitioning is realized by the random selection of an attribute and the random choice of a pivot value in the range of the selected attribute. For an isolation tree, the sample scores are computed as the distance between the leaf node containing the sample and the root node of the tree.

The algorithm uses the number of tree splits to identify minority classes in an imbalanced data set. Intuitively, outliers take fewer splits because the density around the outliers is low. And again, caret's confusion matrix function is used to appraise the accuracy of the isolation forest algorithm. The resulting confusion matrix is presented in Table 2 (Appendix).

The results return a classification accuracy of 87 percent indicating a reasonable alternative to classify students or assign them to a preliminary ranking designed to tailor services.

CONCLUDING COMMENTS

Events over the last years vitiated the predictive models built for admissions, program evaluation, retention, and other college operations purposes. These models have lost their predictive power due to a condition known as dataset-drift or covariate-drift. At its core, this data artifact involves changes in the stationarity of the predictive model's target variable and its predictors. This study shows how to test for the presence of dataset-drift. To do so, it relies on a simulated set of data representing the features favored by college administrators for admissions and operations.

Dataset-drift has resulted in a perplexing conundrum for program administrators. Routine operations and procedures typically require an appraisal of student's performance capabilities to work optimally. Obviously, appraisals are unreliable when the predictors have shifted. This framework is designed to reconstruct appraisals of the likelihood of success of the admitted student body.

The situation can be framed as a One Class problem. This characterization allows us to deploy two well-known algorithms: Support Vector Machines and Isolation Forests to fit and identify group membership. The study shows how to use them to reconstruct a true representative sample of the student body and the associated outcomes of the admissions and success processes.

REFERENCES

Austin, P. C., & Schuster, T. (2016). The Performance of Different Propensity Score Methods for Estimating Absolute Effects of Treatments on Survival Outcomes: a Simulation Study. *Statistical Methods in Medical Research, 25*(5), 2214-2237.

Austin, P. C., & Small, D. S. (2014). The Use of Bootstrapping When Using Propensity-Score Matching Without Replacement: a Simulation Study. *Statistics in Medicing, 33*, 4306-4319.

Chui, K. T., Fung, C. D., Lytras, M. D., & Lam, M. T. (2020). Predicting At-Risk University Students in a Virtual Learning Environment Via a Machine Learning Algorithm. *Computers in Human Behavior*.

Cloudera Fast Forward. (2021). *Inferring Concept Drift Without Labeled Data.* Cloudera. Retrieved from https://www.cloudera.com/

Coleman, A. L., & Keith, J. L. (2018). *Understanding Holistic Review in Higher Education Admissions.* The College Board.

Congress, 1. (2001). *No Child Left Behind Act of 2001.*

Conley, B. (2019, September 6). The Great Enrollment Crash. *The Chronicle of Higher Education*. Retrieved from https://www.chronicle.com/article/the-great-enrollment-crash/

Ekowo, M. (2015, December 14). The Gift that Keeps on Giving: Why Predictive Analytics are Probably on Colleges' Wish Lists this Year. *EDCentral*.

Ekowo, M., & Palmer, I. (2016). *The Promise and Peril of Predictive Analytics in Higher Education.* New America.

Ekowo, M., & Palmer, I. (2017). *Predictive Analytics in Higher Education: Five Guiding Practices for Ethical Use.* New America.

Engler, A. (2021). *Enrollment algorithms are contributing to the crises of higher education.* Washington DC: Brookings. Retrieved from https://www.brookings.edu/research/enrollment-algorithms-are-contributing-to-the-crises-of-higher-education/

Grawe, N. D. (2018). *Demographics and the Demand for Higher Education.* Baltimore: Johns Hopkins University Press.

ICEF Monitor. (2020, August). Demand for Online Learning and Alternative Credentials Surges During COVID-19. *ICEF Monitor*.

Jia, P., & Maloney, T. (2015, July). Using Predictive Modelling to Identify Students at Risk of Poor University Outcomes. *Higher Education, 70*(1), 127-149.

Keane, M. P., Campbell, J., Ching, A., Fang, H., Flinn, C. J., French, E., . . . Shum, M. S. (2022). *Brief of Economists as AMICI CUREIAE In Support of Petitioner.* Supreme Court of the United States.

Khan, S. S., & Madden, M. G. (2004). One Class Classification: Taxonomy of Study and Review of Techniques. *The Knowledge Engineering Review*, 1-24.

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling.* New York: Springer.

Marcus, J. (2017, August 16). Why Suburban Schools Are Inflating Kids' Grades. *The Atlantic*. Retrieved from https://www.theatlantic.com/education/archive/2017/08/suburban-grade-inflation/536595/

MBA Rountable. (2022). *Test-Optional Admission Policies and the Impact on Graduate Management Education.* MBA Roundtable and Wiley Research Report.

Mengash, H. A. (2020). Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems. *IEEE Access*, 55462-55470.
National Student Clearinghouse Research Center. (2022). *Current Term Enrollment Estimates.* National Student Clearinghouse. Retrieved from https://nscresearchcenter.org/current-term-enrollment-estimates/
Saa, A. A., Al-Emran, M., & Shaalan, K. (2019). Factors Affecting Students' Performance in Higher Education: A Systematic Review of Predictive Data Mining Techniques. *Technology, Knowledge and Learning*, 567-598.
Sanchez, E. I., & Moore, R. (2022). *Grade Inflation Continues to Grow in the Past Decade.* ACT.
Schroeder, R. (2021, May 19). A Second Demographic Cliff Adds to Urgency for Change. *Inside Higher Ed.*
Selingo, J. (2018, May 25). The Two Most Important College-Admissions Criteria Now Mean Less. *The Atlantic*. Retrieved from https://www.theatlantic.com/education/archive/2018/05/college-admissions-gpa-sat-act/561167/
Setoguchi, S., Schneeweiss, S., Brookhart, M. A., Glynn, R. J., & Cook, E. F. (2008, June). Evaluating Uses of Data Mining Techniques in Propensity Score Estimation: a Simulation Study. *Pharmacoepidemiol Drug Saf.*, 546-555.
Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl Jr., K. C. (2017). *Data Mining for Business Analytics.* Hoboken, NJ: John Wiley & Sons.
Westrick, P. A., Marini, J. P., & Shaw, E. J. (2021). *Using SAT Scores to Inform Academic Major-Related Decisions and Planning on Campus.* College Board.
Whitford, E. (2022, January 13). Enrollment Marches Downward. *Inside Higher Ed.*
Wingard, J. (2022). *The College Devaluation Crisis.* Stanford University Press.
Yanosky, R., & Arroway, P. (2015). *The Analytics Landscape in Higher Education.* Louisville, CO: ECAR.

APPENDIX

Figure 1
Admissions Propensity Variables

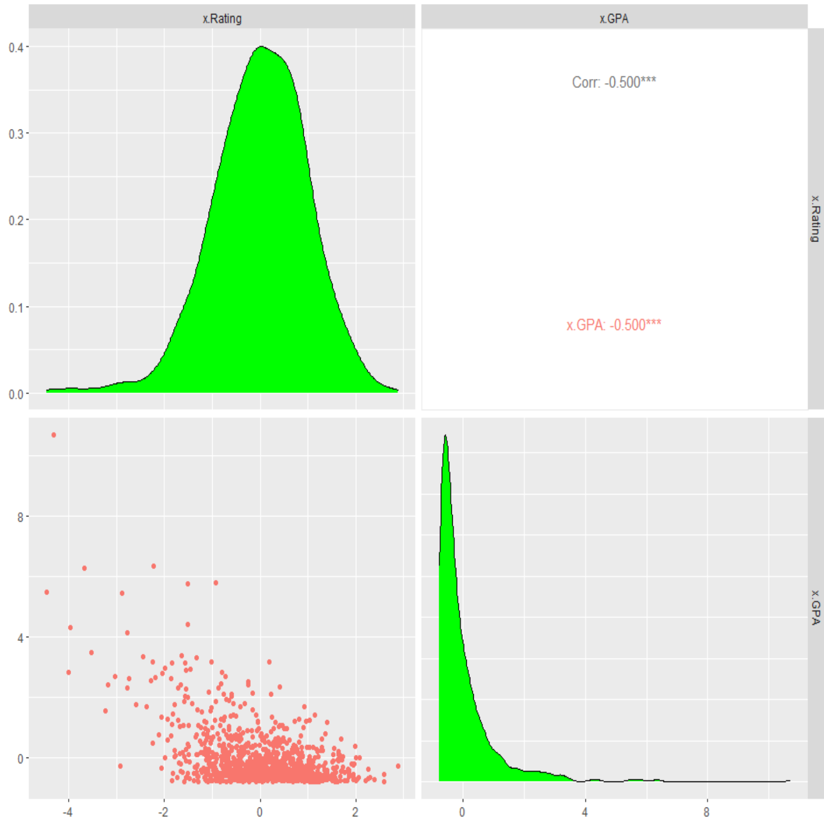

Figure 2
Data Shift Induced by Lowering Admissions Thresholds

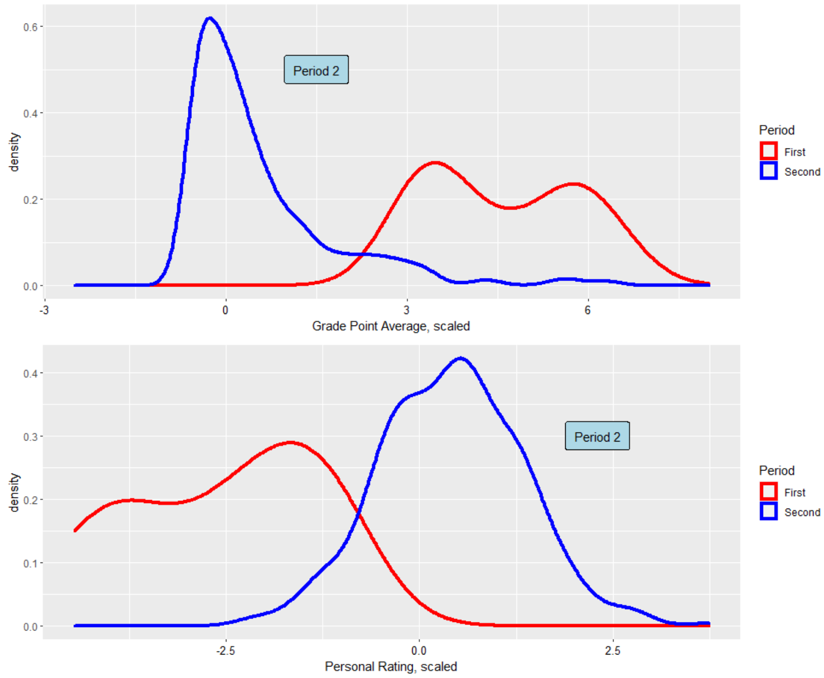

Table 1

	Reference		
		0	**1**
Prediction	**0**	13	8
	1	2	369

Accuracy : 0.974
95% CI : (0.954, 0.988)
No Information Rate : 0.962
P-Value [Acc > NIR] : 0.114
Kappa : 0.709
Mcnemar's Test P-Value : 0.114

Table 2
Confusion Matrix and Statistics

		Reference	
		0	**1**
Prediction	**0**	233	42
	1	38	10

Accuracy: 0.752
95% CI: (0.702, 0.798)
No Information Rate: 0.839
P-Value [Acc > NIR]: 1.000
Kappa: 0.054
Mcnemar's Test P-Value : 0.737