On the Dependence of Drop-Out Rates on Cohort Size in Public Universities

Patrick Hosein(✉) and Sabrina Sankar

The University of the West Indies, St. Augustine, Trinidad
patrick.hosein@sta.uwi.edu, sabrina.sankar@my.uwi.edu

Abstract. We investigate the relationship between the number of graduates at a university for a particular year and the size of this cohort when they entered the University. We focus on publicly funded Universities in which student fees are heavily subsidized by the Government and hence finance plays a smaller role for applicants. In fact, in some countries, fees at Government funded universities have not changed over the last decade. If such universities, consider profit as their main objective then they may accept too many candidates, and this will lead to high dropout rates. However, since these Universities are publicly funded by their respective Governments, their objective should instead be to maximize the number of students that successfully graduate (sometimes called the Goodput of a process). We investigate this relationship and propose a simple model that we believe represents it. Note that we acknowledge that many other factors play a role, but our objective is simply to illustrate that, increasing the number of students may in fact cause a reduction in graduates. In fact, we see this in the many examples that we investigated.

Keywords: Optimization · Data Science · Decision Making · Education

1 Introduction

University tuition at private universities has been rising rapidly throughout the world [4]. However, in many countries, public Universities are funded by the Government and are restricted from increasing tuition. Such Universities may instead try to increase revenue (to offset rising teaching costs) by increasing the number of students they accept. This may be done by reducing admission requirements. We conjecture that such increased intake rates may result in higher drop-out rates and hence lower graduation rates. This stems from, (a) acceptance of less capable students (who may not have been admitted in previous years) and, (b) lower teaching quality because of the increased teaching loads.

We investigate this conjecture by studying incoming and graduating rates at sample Universities. Unfortunately, since drop-out rates are typically not provided, we assume that the number of graduates of a cohort minus the number that entered equals the number who dropped out. We propose a simple function that explains the relationship between graduates of a cohort and incoming size of the cohort. We use real data to demonstrate the accuracy of the proposed model.

S. G. Yaseen (Ed.): SICB 2023, SBD 135, pp. 29–38, 2023.
https://doi.org/10.1007/978-3-031-42463-2_4

2 Related Work and Contributions

Several studies have been conducted by universities in their effort to become more accountable while increasing both admission and graduation numbers. Most have focused on increasing admission and enrolment numbers instead of graduates since their objective is profit. Empirical approaches for these studies typically involve the development of different models based on regression analysis.

In the case of publicly funded universities targeted in this study, factors such as student fees and geographical location do not play a big factor when compared with private universities on which most studies have been focused. For private universities, one of the main ways to increase student enrolment entails reducing the academic standards required for admission [11]. Benefits to universities include being more open and providing opportunities to persons that may have been unable to meet previous admission requirements as well as increasing profits associated with higher admission numbers. However, to increase graduation numbers based on this new admissions model, the standards required to complete the degree must also be reduced leading to lower quality graduates. This negatively affects the reputation of the University.

In [6, 9] a strong correlation was found between lower graduation rates and decreasing the test score for new admissions whilst maintaining the same completion standards. A study done at the University of Hawaii suggested that efforts to raise admission selection criteria to maximize first year retention would diminish admission numbers between 19.2% and 25.3% for the years 1999 to 2001 [9]. In order to increase admission rates and maintain retention they propose an alternate assessment process that was better suited to "opportunity" admissions and a faculty review process.

While many studies focused on increasing profits, [16] attempted to create a better model that focuses on increasing success rates. A Dual Admission Model (DAM) using regression analysis based on student achievement data was used to determine a student's success. This in turn was used to maximize student retention as the selection criteria was centred on the predicted GPA that would be obtained from students in their second year.

Research from [3] also studied post-entrance screening policies that have been used in some countries to balance initial enrolment and degree completion rates. A dynamic discrete choice model was developed where the enrolment outcome was uncertain. Though the admission policies for publicly funded universities vary for different countries within Europe [2, 8, 10], success was typically achieved through screening policies. Well thought out admission standards were found to lower the occurrence of unsuccessful new admissions whilst simultaneously leaving degree completion rates unaffected. More emphasis on matching new students to their right degrees showed greater promise of reducing drop-out and increasing degree completion.

The paper [14] emphasized the superiority of logistic regression over other methods. These models were also seen to perform relatively well for classification and prediction, providing good results [5] when compared to discriminant function analysis. Similar techniques have been used to predict student matriculation based on admissions data [7]. Although most of the studies mentioned are focused on student admissions and retention, the same methods can be applied to degree completion.

This paper seeks to contribute to a different perspective of university accountability where graduation rate is the primary objective. We demonstrate that graduation rates may in fact decrease as enrolment rates are increased. Therefore, one must balance profit (high enrolment rate) with education quality (low drop-out rate). We believe that Universities can use their available data to estimate the right balance and that the proposed model can help.

3 Proposed Mathematical Model

The proposed model can be applied at different levels within a university's hierarchy, at the University level, the faculty level and even the Department level. However, statistics at the lower levels are rarely made public so our examples will focus on the higher levels.

Consider a university in which N students are accepted for a school year. Naturally N will vary with entry requirements. The more stringent the requirements the smaller the value of N (lower acceptance rate). Therefore, as N increases (i.e., requirements become more relaxed) then the average quality of the student body decreases and hence the dropout rate will typically increase. This will hold true even if the student/teacher ratio remains the same (i.e., more staff are hired for the increased student intake). If the staff size has not increased, then the dropout rate is expected to increase even further since the resources per student will decrease. Hence, we expect the dropout rate to be a nonlinear function of the student intake, N. We propose a logistic function for this relationship and will provide support for this choice. We first demonstrate the approach using an illustrative example and then use real data to show that the model is valid in practice.

Let us consider a university with an intake of N students. We denote the dropout probability by $p(N)$ which means that the number of students who complete successfully is $N(1 - p(N))$. The function $p(N)$ depends on many factors and later in the paper we describe how this function, and its parameters can be estimated from historical data. As N increases, the intake will have a higher percentage of lower quality students (since we assume that the top candidates are made offers). This is the primary reason for the increase in $p(N)$ with N. Another reason is the diluted attention provided to students. If we assume that the teaching staff remains constant, then the student to staff ratio increases and hence less individualized attention results in lower student grades and hence increased likelihood of failing or dropping out. Consider the following logistic function for representing the dropout probability as a function of intake size:

$$p(N) = 1/[1 + e - ((N - a)/b)] \tag{1}$$

where $a > 0$ and $b > 0$, are parameters.

Note that when N is small $p(N)$ is small and as N goes to infinity $p(N)$ approaches 1. The parameter a is used to shift the curve horizontally while the parameter b determines the steepness of the curve. These parameters will be determined using historical data. The number of graduates (as a function of N) is then given by

$$G(N) \equiv N(1 - p(N)) = N/[1 + e(N - a)/b] \tag{2}$$

Suppose that $a = 4500$ and $b = 800$ then the resulting dropout probability curve is given in Fig. 1 (left) while the number of graduates as a function of intake is provided in Fig. 1 (right). We can take the derivative of G(N) and set it to zero to obtain the optimal student intake. Unfortunately, there is no closed form solution and hence one would need to determine the optimal value numerically. Next, we validate this model with actual data.

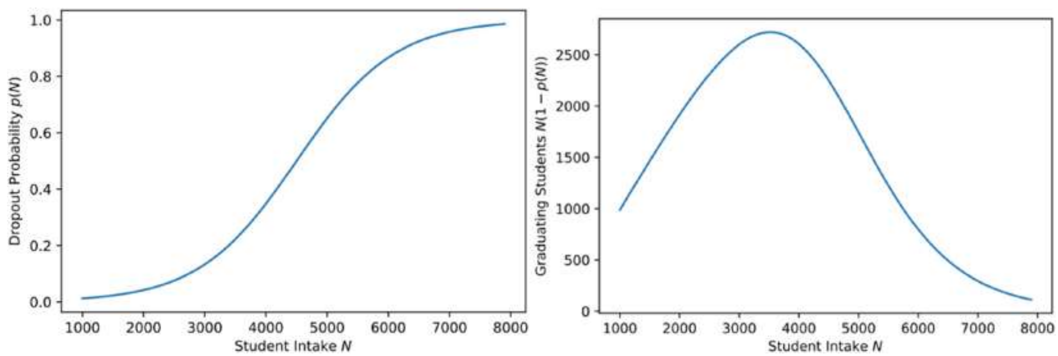

Fig. 1. Dropout probability (left) and graduating students (right) as a function of intake.

4 Model Validation

We first consider a particular University with campuses in different countries. Each campus manages their own admission process and so can be treated separately. The data used for this university is publicly available on their web site and hence can freely be used. We considered three of the campuses for which data was available. We also found data for another University in another region of the world which we also used to illustrate the model.

4.1 University Example (STA)

We were able to obtain publicly available data from a university web site [15] for undergraduate student intake and number of graduates for each year. The degrees are designed to be completed in 3 years and so we will assume that, for a given year, the number of entering students for that year minus the number of students who graduate in 3 years' time is the number of dropouts for that class. More detailed information (dropouts in each year for example) would allow us to do a more detailed analysis but such data was not publicly available. We provide the number of new students (intake) and number of graduates for various years in Fig. 2 (top). The year denotes the first year of the academic year. Note the dip in graduating students for the academic year 2019/2020. These students would have graduated in September of 2020 but because of COVID many of them were not able to complete their degrees. This sample will be highlighted in our plots but will be treated as an outlier and hence not included in our regression analysis.

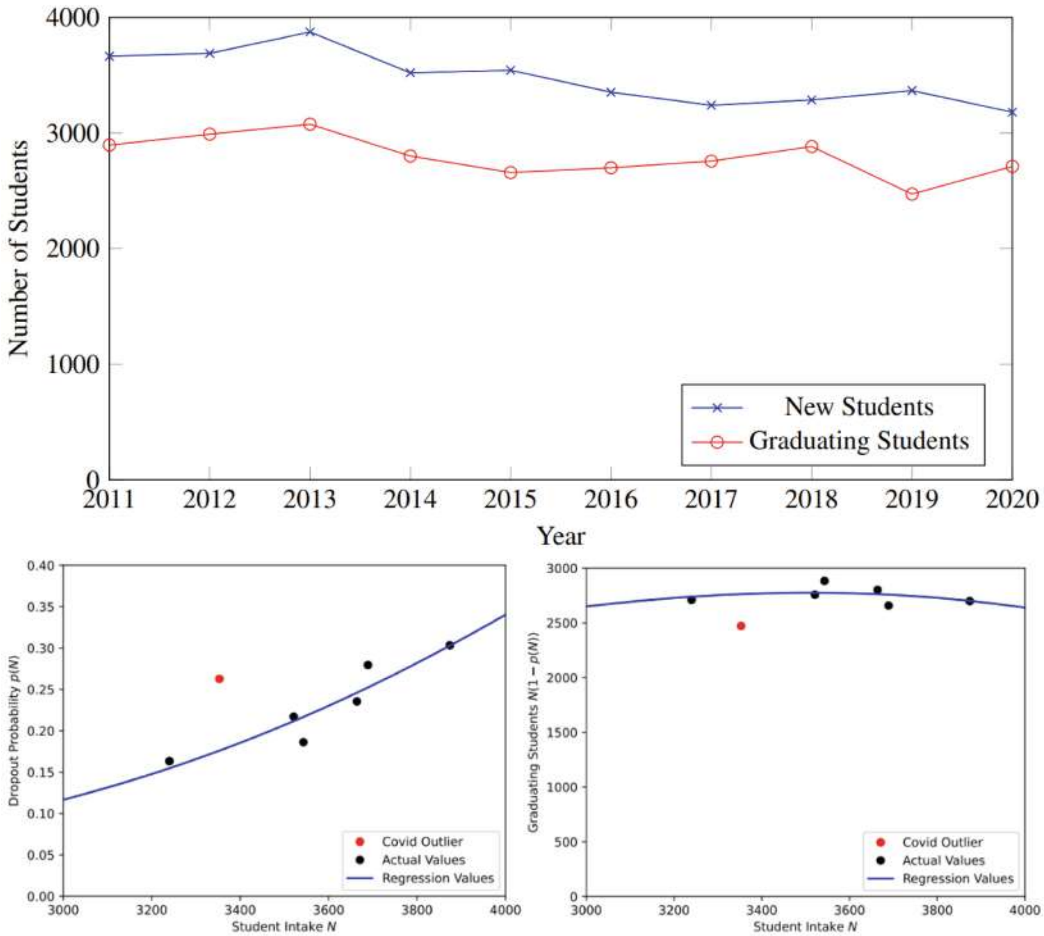

Fig. 2. STA: Intake and Graduates by year (top), Dropout Probability (bottom left) and Graduates (bottom right) vs Intake.

We obtain a sample for each year as a pair consisting of the intake size for that year and the number of those students who graduate (taken from 3 years in the future). Therefore, we can obtain samples for the years 2011–2017 which is what we use for our analysis. However, we ignore the year 2017 (with graduating year 2020) because the graduating class was particularly small because of the COVID pandemic.

We can determine the probability of graduating as the ratio of the graduating class size and the corresponding entering class size. The drop out probability is the number of students who drop out divided by the intake size and we denote this by $p(N)$. Given these samples we next determine the values a and b in 1 that provides the smallest Mean Square Error (MSE). The error is the difference between the actual value and the value obtained with the model. We do this using a coordinate descent approach as follows. We fix b and optimize (i.e., find the lowest MSE) over a by doing a simple linear search along values for a. We then fix a at this optimal value and optimize over b. We then fix b at this new value and optimize over a. This process is repeated until neither a nor b changes. When we do this, we obtain $a = 4486$ and $b = 734$.

Using these parameter values, we plot the function $p(N)$ versus N in Fig. 2 (bottom left). We include the sample point (in red) corresponding to the COVID graduating year to show how far off it is from the other points. Note that typically R^2 values are only used for checking performance for linear regression but not for non-linear regression. However, we can see that the portion of the curve within which the sample points lie is close to linear, so we did the computation to obtain $R^2 = 0.87$ which indicates a good regression approximation.

Given the function $p(N)$ we can now determine the graduating class size, and this is provided in Fig. 2 (bottom) right. The maximum graduating class of 2776 students is obtained with an intake of 3510 students. With this intake the dropout rate is 21%. These numbers suggest that this campus is already running at capacity and that any attempts to increase student intake must involve a discussion on increasing staffing and resources.

4.2 University Example (CH)

In this section we consider another campus of the same university which is on a different island with a different admission process. The student intake and graduates per year is provided in Fig. 3 (top).

We note two strange transitions. In 2013 there was a sudden drop in new students. This was due to the Government in the country stopped providing a fee subsidy to students [12]. This caused several students who could no longer afford to pay tuition to dropout leading to an unusually high dropout rate, so this sample was removed from the regression.

Next, we see a sudden rise in new students from 2018. This was since the Government changed its policy and reinstated the subsidy [1]. However, during the prior 3-year period, staff was significantly reduced due to the drop in registrations. The sudden rise in students from 2018 could not be properly managed and again this led to an unusually high dropout rate. Over time, Faculty and resources will be increased but we ignore this sample point because it violates our assumption of a stable system. In Fig. 3 (left) we plot the dropout probability, and, on the right, we plot the number of graduates, both as a function of student intake. Note that the two anomalies (red and blue) are provided for reference. In this case we obtain $a = 2571$ and $b = 609$. The optimal intake is 2050 with a dropout probability of 30%.

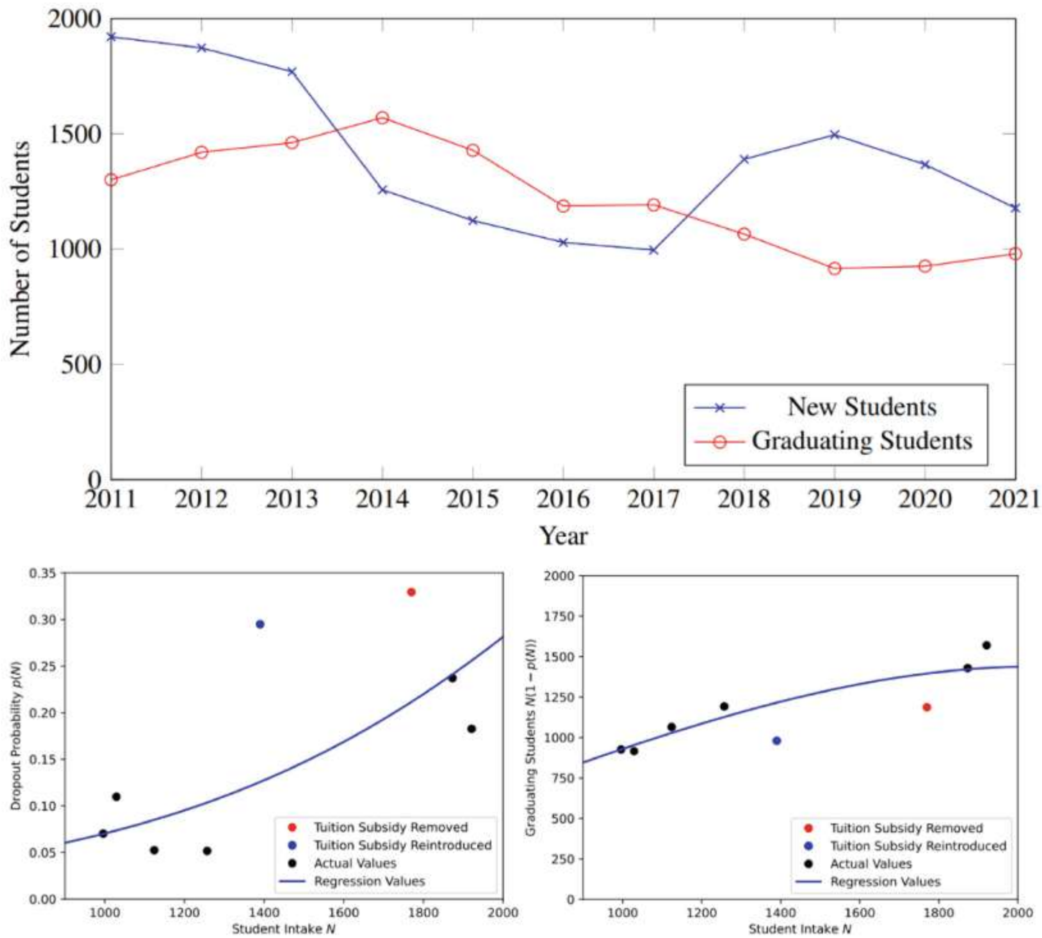

Fig. 3. CH: Intake and Graduates by year (top), Dropout Probability (bottom left) and Graduates (bottom right) vs Intake.

4.3 University Example (Mona)

Figure 4 has results for the third campus. In this case the dropout rates are high with the lowest sample dropout rate being 25%. Although the model indicates a small increase in graduates with increased intake, this will come at the cost of even higher dropout rates. Therefore, in this case the focus should be on reducing dropout rates by reducing intake and/or increasing staff and resources.

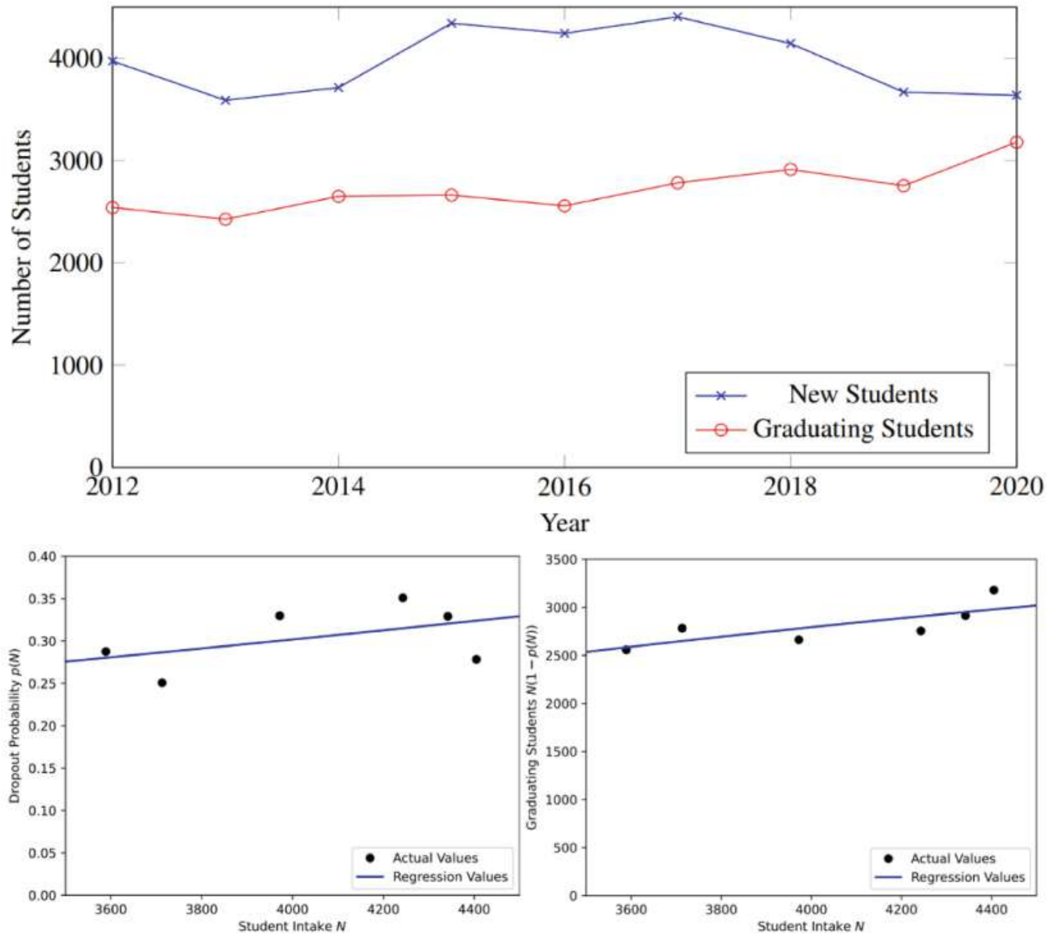

Fig. 4. MONA: Intake and Graduates by year (top), Dropout Probability (bottom left) and Graduates (bottom right) vs Intake.

4.4 University Example (Oxford)

We decided to include a university from a different region and was able to get data for Oxford University [13]. The results for this case are provided in Fig. 5. Although Oxford is considered to be a Public University, the vast majority of its revenue comes from private sources and hence maintaining its reputation is important. We therefore find that the dropout rate is very low. However, the proposed model seems to fit this Institution as well. The optimal intake is 3250 with a dropout rate of only 10%. The regression curve is almost linear for the graduation rate curve and hence we computed the R^2 value and obtained 0.92. The University appears to be operating close to its optimal capacity.

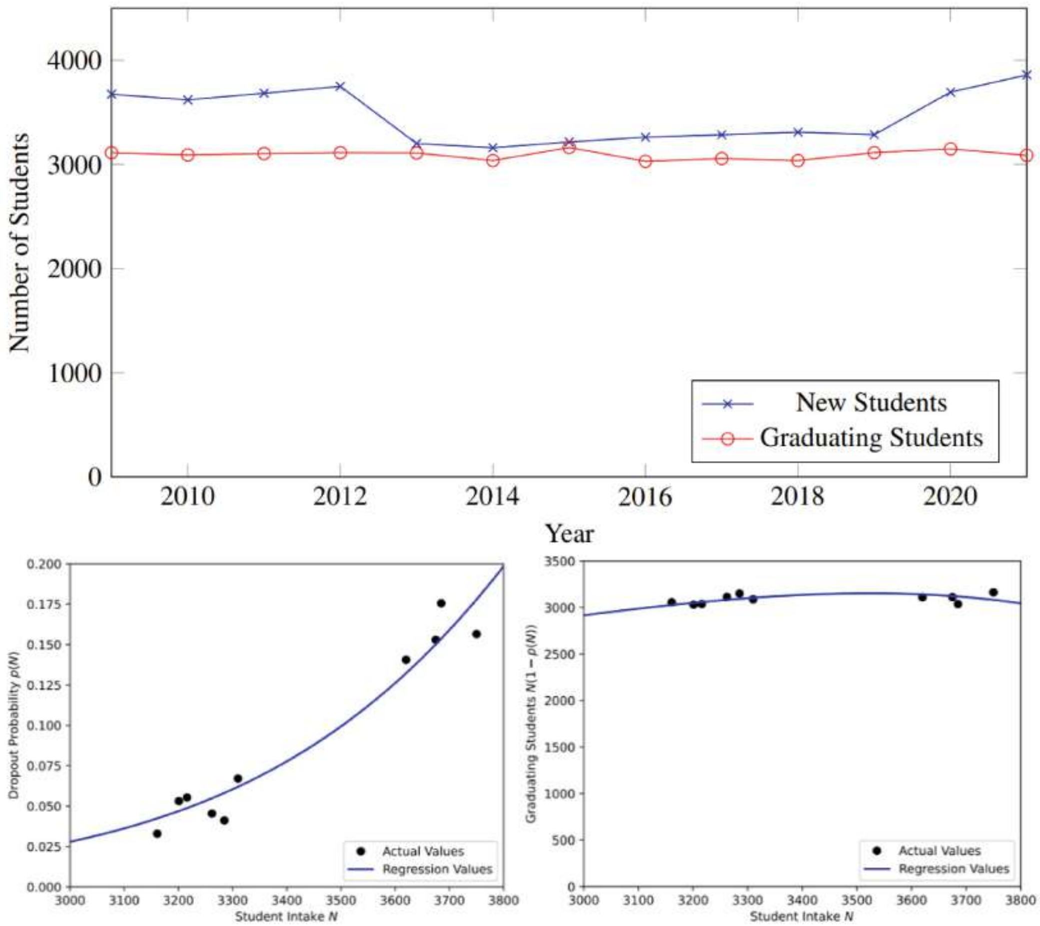

Fig. 5. Oxford: Intake and Graduates by year (top), Dropout Probability (bottom left) and Graduates (bottom right) vs Intake.

5 Conclusions

We proposed a model for estimating graduation rates based on student intake. The parameters for this model were obtained from historical data of the University. Unfortunately, since the number of samples per dataset was small, we could not use the traditional train/test validation approach. However, using real statistics from different universities we demonstrated the value of the approach. In the future we plan to perform similar analyses based on Faculty and possibly on Department level results. We believe that this work can assist Universities when making decisions on aspects such as admission requirements and staffing requirements.

References

1. Rollock, M.: Free tuition for UWI, Bcc and Erdiston Students. In: GIS (2018). https://gisbarbados.gov.bb/blog/free-tuition-for-uwi-bcc-erdiston-students/. Accessed 22 Jan 2023
2. Chen, L.: University Admission Practices – UK (2012)

3. Declercq, K., Verboven, F.: Enrollment and degree completion in higher education without admission standards. Econ. Educ. Rev. **66**, 223–244 (2018). https://doi.org/10.1016/j.econedurev.2018.08.008
4. Ehrenberg, R.G.: Tuition rising: Why college costs so much. Harvard University Press, Cambridge (2009)
5. Fan, X., Wang, L.: Comparing linear discriminant function with logistic regression for the two-group classification problem. J. Exp. Educ. **67**, 265–286 (1999). https://doi.org/10.1080/00220979909598356
6. Hurt, E.: The marketization of Higher Education. Coll. Lit. **39**, 121–132 (2012). https://doi.org/10.1353/lit.2012.0013
7. Khajuria, S.: A model to predict student matriculation from admissions data. Ohio University, Thesis (2007)
8. Kübler, D.: University Admission Practices - Germany (2019)
9. Leonard, B., de Pillis, E.: Aligning admissions standards with graduation and Retention Goals. The International Journal of Learning: Annual Review **15**, 129–140 (2008). https://doi.org/10.18848/1447-9494/cgp/v15i06/45803
10. Merllino, L.P., Nicoló, A.: University Admissions Practices - Italy (2012)
11. Mulvenon, S.W., Robinson, D.H.: The paradox of increasing both enrollment and graduation rates: Acknowledging elephants in the Ivory Tower. International Journal of Higher Education (2013). https://doi.org/10.5430/ijhe.v3n1p66
12. Simpson, S.: Students to pay uwi tuition. In: NationNews Barbados (2013). https://www.nationnews.com/2013/08/14/students-to-pay-uwi-tuition/. Accessed 22 Jan 2023
13. University of Oxford: Undergraduate Admissions Statistics. In: University of Oxford (2022). https://www.ox.ac.uk/about/facts-and-figures/admissions-statistics/undergraduate-students. Accessed 1 Jan 2023
14. Peng, C.Y.J., So, T.S.H., Stage, F.K., St. John, E.P.: Research in Higher Education **43**, 259–293 (2002). https://doi.org/10.1023/a:1014858517172
15. UWI: Student data. In: Student Data | University Office of Planning (2005). https://uwi.edu/uop/student-data. Accessed 20 Jan 2023
16. Wang, G., Shulruf, B.: Admission model and equity in Higher Education. Asia Pac. Educ. Res. **22**, 111–117 (2012). https://doi.org/10.1007/s40299-012-0002-8