# A Statistical Analysis of Drop Rates in Introductory Computer Science by Gender and Partial Grade

Jikai Long
Dept. of Computer and
Mathematical Sciences
University of Toronto Scarborough
jikai.long@mail.utoronto.ca

Brian Harrington
Dept. of Computer and
Mathematical Sciences
University of Toronto Scarborough
brian.harrington@utsc.utoronto.ca

## ABSTRACT

University level introductory computer science courses have a notoriously high drop rate. Many students who aim to pursue CS education fail to complete their first semester. This trend holds across a wide variety of institutions, courses, and countries. In this study, we collect data from over 4000 students enrolled in an introductory course over a 5 year period in order to analyze potential factors that affect which students are more likely to drop the course. In particular, this work focuses on gender, international student status, and course performance up to the drop date.

The data for each student includes performance on assignments and term tests as well as weekly exercises which allow us to create a temporal view of students and analyze their partial grade at the time they stopped participating in the course. We find that there are slight differences in the probabilities of dropping for lower performing students by gender, the results are not statistically significant. Furthermore, we find virtually no difference in students who, up to the drop date, are receiving passing marks by either gender or international student status.

## CCS CONCEPTS

• **Social and professional topics** → **Computing education**; **Computer science education**; **CS1**.

## KEYWORDS

CS1, drop rate, gender, international students

## 1 INTRODUCTION

Introductory computer science (CS1) courses, are often perceived by students to be more difficult than other similar courses in their

curriculum, with a correspondingly high percentage of students dropping the course in order to pursue other fields of study[5].

Many institutions allow students to withdraw from enrolled courses without academic penalty if and only if they withdraw from the course before a set drop date. Therefore, it is natural to assume that performance in the course up to this drop date would be a key factor in a student's decision to withdraw.

In this study, we track over 4000 students during 5 years worth of offerings in CS1, computing their partial grade at the drop date, in order to statistically analyze the probability that a given student will choose to drop the course given their performance up to that point.

We find across all groups that students are generally unconservative, that they are more likely to remain in a course they are likely to fail than to drop a course they are likely to pass. We find no difference between domestic and international student drop patterns (despite a large discrepancy in the fees they are paying for the course), and while there is a small non-significant difference in gender for the lower performing students (female students who are below 60% at the drop date are slightly more likely to drop than their male counterparts), for those at the higher end of the grade spectrum, there is virtually no difference by gender.

## 2 BACKGROUND

Gender performance is an interesting and much studied topic in computer science education. Multiple studies have shown that there is no difference between different genders in pass-fail rates for students completing courses. Rountree et al reported that gender is not significantly correlated to failure in computer science by applying statistical analysis on CS1 student survey responses[6]. Pappas et al also found that gender was not a factor in the decision of whether to drop CS courses from survey data.[3]

There has been some research into the relative confidence of international vs domestic students [2], but no research has shown an effect on drop rates or performance.

## 3 METHODOLOGY

[Course details removed for anonymity] is an introductory computer science course at [University removed for anonymity] servicing students intending on majoring in computer science with no prior experience. From 2012 - 2017, the course consisted of weekly or bi-weekly exercises, 1-2 term tests, 2-3 larger assignments over a 12 week period, and a final exam. Students may choose to drop the course before the end of the 10th week, by which time they have received between 36% and 45% of their final grade.

Partial grades were calculated for students who had attended all term tests and submitted code for most assignments, and at least 75% of the exercises by the drop date.

Gender was computed using self-declared honorific in official university databases. Approximate 8% of students chose not to declare an honorific, or chose a gender neutral honorific. These students were excluded from the gender-based analysis. Of the students declaring gender specific honorific, 74% of students were male and 26% were female.

Domestic or International student status was calculated from fee-status with the university. 69% of students were domestic, and 31% of students were international. International students paid on average 7.35 times more per course than domestic students.

## 4 RESULTS

Overall, there was a clear inverse correlation between partial grade and drop probability, with students above 60% (corresponding to a C- or better), being much less likely to drop than students below that threshold. One interesting anomaly is that students below 20% showed a very steep decline in drop probability. While this is a very small sample, it may merit further study.

The data for drop rate by gender can be found in Table 1. As seen in Figure 1, the trends for male and female students are very similar. A slight increase in drop probability can be seen for female students below the 60% boundary, but the results are not statistically significant at any grade range.

| Grade | 0-20 | 20-40 | 40-50 | 50-60 | 60-70 | 70-80 | 80-90 | 90-100 |
|---|---|---|---|---|---|---|---|---|
| MaleDrop | 5 | 45 | 38 | 45 | 24 | 4 | 4 | 0 |
| FemaleDrop | 0 | 19 | 20 | 21 | 7 | 3 | 1 | 0 |
| MaleComplete | 45 | 130 | 158 | 314 | 419 | 572 | 507 | 213 |
| FemaleComplete | 10 | 42 | 53 | 105 | 141 | 172 | 145 | 39 |
| MaleTotal | 50 | 175 | 196 | 359 | 443 | 576 | 511 | 213 |
| FemaleTotal | 10 | 61 | 73 | 126 | 148 | 175 | 146 | 39 |
| MaleDropPercent | 10% | 26% | 19% | 13% | 5% | 1% | 1% | 0% |
| FemaleDropPercent | 0% | 31% | 27% | 17% | 5% | 2% | 1% | 0% |

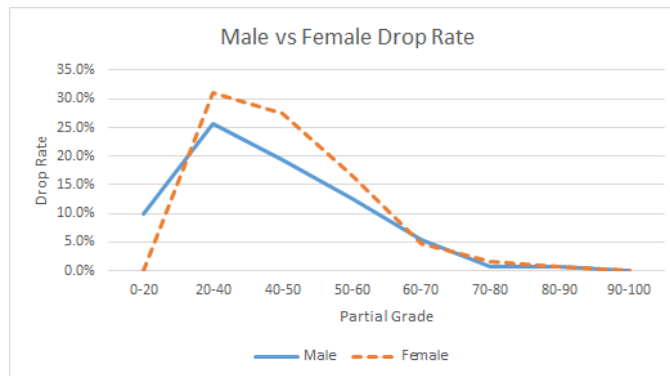**Table 1: Male vs Female Table**



**Figure 1: Male vs Female Chart**

The data for drop rate by domestic or international student status is given in Table 2. As seen in Figure 2, the trends are nearly

identical, save for those below 20% partial grade. There is no statistically significant difference between the groups except for the 0-20% range ($p = 0.03834$), but as the sample sizes are so small, this should not be treated as a particularly interesting insight.

| Grade | 0-20 | 20-40 | 40-50 | 50-60 | 60-70 | 70-80 | 80-90 | 90-100 |
|---|---|---|---|---|---|---|---|---|
| DomesticDrop | 6 | 32 | 37 | 45 | 23 | 9 | 4 | 0 |
| InternationalcDrop | 0 | 32 | 25 | 25 | 11 | 2 | 1 | 0 |
| DomesticComplete | 34 | 92 | 122 | 277 | 426 | 597 | 536 | 234 |
| InternationalComplete | 26 | 87 | 103 | 166 | 182 | 200 | 145 | 44 |
| DomesticTotal | 40 | 124 | 159 | 322 | 449 | 606 | 540 | 234 |
| InternationalTotal | 26 | 119 | 128 | 191 | 193 | 202 | 146 | 44 |
| DomesticDropPercent | 15.0% | 25.8% | 23.3% | 14.0% | 5.1% | 1.5% | 0.7% | 0.0% |
| InternationalDropPerc | 0.0% | 26.9% | 19.5% | 13.1% | 5.7% | 1.0% | 0.7% | 0.0% |

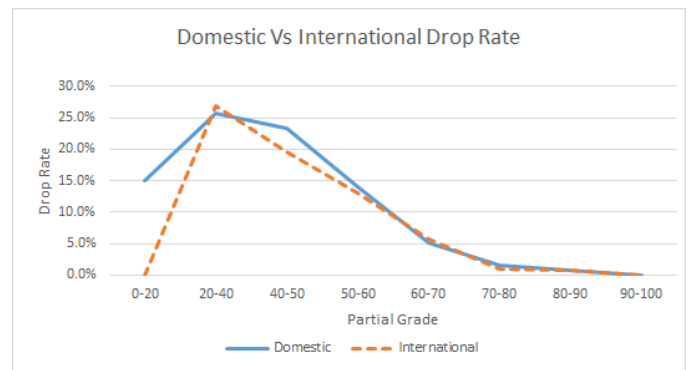**Table 2: Domestic vs International Student Table**



**Figure 2: Domestic vs International Student Chart**

## 5 CONCLUSIONS

In this paper, we have established a large scale multi-year analysis of student data in CS1. Our preliminary findings showed that there was no statistically significant difference in the probability of dropping based on gender (which is potentially interesting as it goes against common folk wisdom about the "leaky pipeline" of women in STEM careers[4]), and that there was virtually no difference in probability of dropping between domestic and international students (which is potentially interesting as it runs counter to the sunk cost fallacy[1]).

## REFERENCES

[1] Hal R Arkes and Catherine Blumer. 1985. The psychology of sunk cost. *Organizational behavior and human decision processes* 35, 1 (1985), 124–140.
[2] Brian Harrington, Jingyiran Li, Mohamed Moustafa, Marzieh Ahmadzadeh, and Nick Cheng. 2019. On the Effect of Question Ordering on Performance and Confidence in Computer Science Examinations. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. ACM, 620–626.
[3] Ilias O Pappas, Michail N Giannakos, and Letizia Jaccheri. 2016. Investigating factors influencing students' intention to dropout computer science studies. In *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education*. ACM, 198–203.
[4] Alice N Pell. 1996. Fixing the leaky pipeline: women scientists in academia. *Journal of animal science* 74, 11 (1996), 2843–2848.
[5] Andrew Petersen, Michelle Craig, Jennifer Campbell, and Anya Tafliovich. 2016. Revisiting why students drop CS1. In *Proceedings of the 16th Koli Calling International Conference on Computing Education Research*. ACM, 71–80.
[6] Nathan Rountree, Janet Rountree, and Anthony Robins. 2002. Predictors of success and failure in a CS1 course. *ACM SIGCSE Bulletin* 34, 4 (2002), 121–124.