

**Survival and dropout analysis in the undergraduate mathematics course at ufv:
a statistical approach**

Survival Analysis and Dropout in the Mathematics Teaching Degree at UFV: A Statistical Approach

Dr. Aurelio de Aquino Araújo – Pontifical Catholic University of Chile

Me. Daiana Cristina Lemos – Federal University of Viçosa

Summary

This article presents a statistical analysis of student retention and dropout rates in the Bachelor's Degree in Mathematics from the Federal University of Viçosa (UFV) - Campus Florestal, using survival analysis techniques, such as the Kaplan estimator, Meier and the Cox proportional hazards model. Data from 141 students were analyzed between 2010 and 2023. The results show significant differences between the groups analyzed, with greater stability in the trajectories of students who graduated and a higher dropout rate in first years by those who dropped out or were dismissed. The relevance of public and institutional policies that consider socioeconomic inequalities in tackling evasion.

Keywords: survival analysis, school dropout, higher education, inequality educational, statistical modeling.

Abstract

This article presents a statistical analysis of student retention and dropout in the Mathematics Teaching Degree at the Federal University of Viçosa (UFV) – Florestal Campus. Using survival analysis techniques, particularly Kaplan-Meier estimation and Cox proportional hazards modeling, we examined the academic trajectories of 141 students enrolled between 2010 and 2023. The results highlight significant differences in survival time across academic status groups, with early dropout being most frequent among students who abandoned or were institutionally withdrawn. In contrast, those who graduated exhibited greater trajectory stability. The study underscores the importance of public and institutional policies that address socioeconomic

inequalities in combating dropout.

Keywords: survival analysis, dropout, higher education, educational inequality, statistical modeling.

Summary

This article presents a statistical analysis of student retention and desertion in Bachelor's degree in Mathematics from the Florestal campus of the Federal University of Viçosa (UFV). Survival analysis techniques will be used, especially the Kaplan-Meier estimation and the Cox proportional hazards model, to examine the academic trajectories of 141 students between 2010 and 2023. The results show significant differences between them analyzed groups, with greater stability among those who graduated and greater desertion in the first few years between them they abandoned or were unattached. The studio highlights it importance of public and institutional policies that consider inequalities socio-economic issues to face desertion.

Keywords: survival analysis, desertion, higher education, educational inequality, statistical modeling.

1. INTRODUCTION

Dropout in higher education is one of the biggest challenges faced by educational institutions in Brazil and around the world. This phenomenon compromises the effectiveness of policies to democratize access to education, in addition to representing a waste of resources public and frustration of students' life projects. In the case of undergraduate courses, particularly in Mathematics, dropout rates are even more significant, affecting the teacher training in strategic areas for national development (Education, 2022; Cunha and Carvalho, 2016).

Studies show that dropout is multifactorial, influenced by pedagogical aspects, institutional, socioeconomic and psychological factors. Among the most recurrent factors are di-academic difficulties, lack of institutional support, need to work, low self-esteem and devaluation of the teaching career (Tinto, 1993; Araújo and Lopes, 2018). In degrees, these causes are aggravated by the depreciative social perception of the profession teacher, which directly affects the motivation and retention of students (Furlan and Miranda, 2015).

Statistical analysis applied to education has been an important tool for

understand the students' trajectory and propose intervention actions. Among the approaches existing, survival analysis stands out for allowing the modeling of time until occurrence of events, such as evasion, considering data censorship and diversity in academic trajectories (Cleves et al., 2010; Singer and Willett, 2003). This approach enables a more accurate reading of student retention throughout the semesters.

Furthermore, the advancement of computational tools has contributed to more accurate analyses. sophisticated and reproducible, using languages such as Python and specific libraries for applied statistics, such as lifelines and pandas. The use of these technologies facilitates the monitoring of students throughout the course and the evaluation of institutional policies based on evidence quantitative (McKinney, 2021).

In this article, we propose an investigation into the patterns of permanence and evasion in Bachelor's Degree in Mathematics at the Federal University of Viçosa – Florestal Campus. Through survival analysis methods and statistical modeling, the objective is to identify factors associated with dropping out and completing the course, providing support technical and theoretical tools for formulating effective retention strategies in

scope of public educational policies.

2. OBJECTIVES

This study has the general objective of investigating, through statistical methods, the factors associated with the permanence and dropout of students in the Bachelor's Degree in Mathematics at Federal University of Viçosa (UFV), Florestal campus. By adopting survival analysis as a methodological basis, we intend to understand the time until the occurrence of events such as evasion, graduation or dismissal, considering data censorship and characteristics individual students.

The following specific objectives stand out: (i) to describe the profile of students based on temporal and institutional variables; (ii) estimate survival curves using the method of Kaplan-Meier; (iii) apply the log-rank test to verify statistical differences between the groups; and (iv) fit Cox regression models to identify variables that influence the risk of dropout, considering factors such as sex, age and type of school of origin.

3. METHODOLOGY

3.1 Data Characterization

The data were obtained from the academic coordination of the Bachelor's Degree course in Mathematics at UFV - Florestal Campus, for the period 2010 to 2023, totaling 141 student records. Information includes date of entry, final status (graduation, abandonment, lockout or shutdown), in addition to available census data such as sex, age entry and type of school of origin (public or private).

The dependent variable for survival analysis was defined as the time, in years, elapsed between entry and exit from the course. For students still active, the data was treated as right-censored, as is common practice in longitudinal studies (Kleinbaum and Klein, 2012).

3.2 Analytical Procedures

Initially, descriptive statistics and boxplot graphs were generated to assess the distribution of residence times according to the final situation. Then, the function

survival rate was estimated using the Kaplan-Meier method, which allows calculating the cumulative probability of permanence over the semesters.

To assess differences between groups, the log-rank test was used, which is appropriate for comparing survival curves in different categorical strata. Subsequently, the model was adjusted Cox, whose proportional hazard function is expressed by:

$$h(t|X) = h_0(t) \cdot \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)$$

This model allows estimating the effect of covariates on the chance of evasion, assuming proportional hazards constant over time (Hosmer et al., 2008). The following were evaluated: proportionality assumptions through Schoenfeld residuals.

3.3 Computing Environment

All analyses were conducted in Python 3.10, using the pandas libraries, seaborn, matplotlib, numpy, and lifelines. Using these tools ensures reproducibility and makes it easier visualization of results, being compatible with open science and replication practices scientific (Van Rossum and Foundation, 2020; Davidson and Silva, 2021).

4. RESULTS

The data analyzed reveal heterogeneous patterns of permanence among the different profiles of students. Figure 1 shows the distribution of length of stay, in years, according to the final enrollment status. It is noted that students who graduated (group "C") present a median duration close to 4.2 years, with low variability, which suggests a more stable academic trajectory. On the other hand, students who dropped out of the course (group "A") or were disconnected (group "D") demonstrate significantly longer median times low, less than 1.5 years, with greater dispersion in the data.

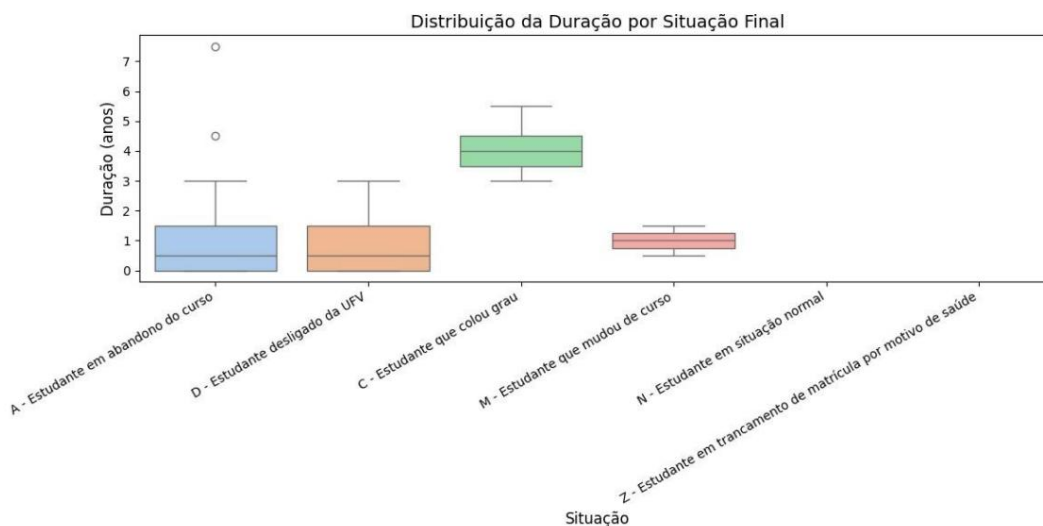


Figure 1: Distribution of length of stay by final status

These findings point to the presence of two distinct profiles: one that manages to maintain an academic trajectory until the completion of the course, and another that faces early ruptures, possibly associated with lack of adaptation, economic difficulties or lack of policies support in the first semesters. The group of students who changed courses (group “M”) also showed low permanence, suggesting trajectory reorientations in the first few years.

Figure 2 shows the survival curves estimated via Kaplan-Meier, comparing the probability of permanence over time among the different groups. Students who graduated maintain high survival rates until the fourth year, while survival curves dropout and disengagement rates show sharp declines in the first two years of the course.

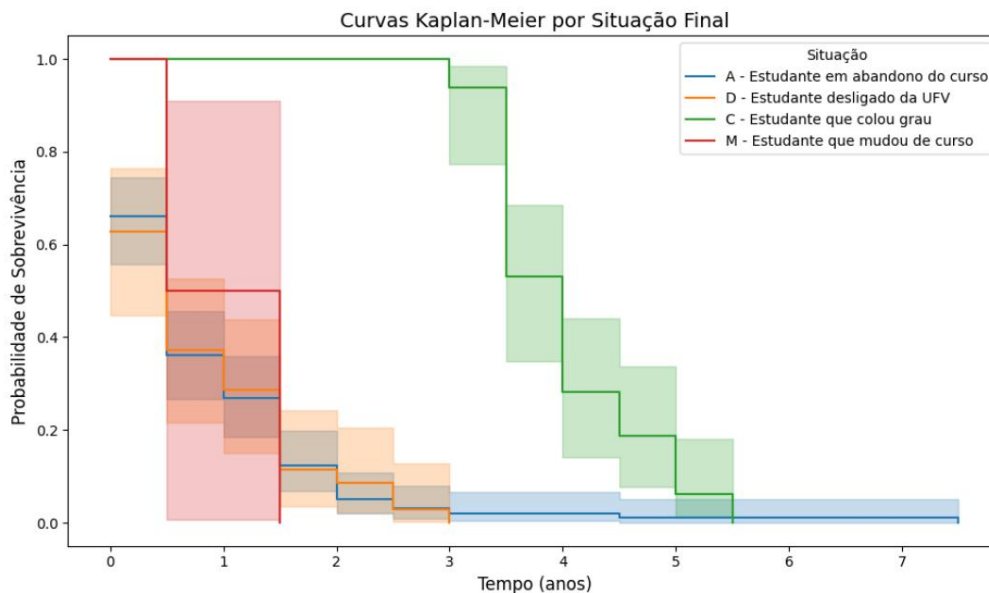


Figure 2: Kaplan-Meier survival curves by final status

The application of the log-rank test indicated statistically significant differences between the curves ($p < 0.01$), confirming that evasion patterns vary substantially according to the student's final outcome. Students who remain in the course for more than three years have increasing likelihood of graduating, while those who drop out tend to do so until the second year of enrollment.

This evidence reinforces the importance of institutional action in the initial semesters. Reception strategies, pedagogical monitoring and financial support can act in a decisive way in overcoming the initial obstacles faced by entrants (Nogueira and Silva, 2019; Maciel and Oliveira, 2022). Survival analysis, in this context, is shown to be a powerful tool for identifying critical moments and assessing the impact of institutional interventions (Cleves et al., 2010; Singer and Willett, 2003).

Finally, the data structure and visualizations reinforce the robustness of the approach adopted. The diversity of permanence patterns, combined with the statistical rigor of the methods used, provides a solid basis for the interpretation of results and the preparation of practical recommendations based on evidence.

5. DISCUSSION

Dropout rates in undergraduate courses, especially in Mathematics, are a phenomenon multifaceted that crosses pedagogical, social and structural aspects of higher education Brazilian. As pointed out by (Tinto, 1993), the decision to remain or abandon a university course is influenced by factors involving the academic and social integration of the student. In the case of undergraduate courses, this integration is often hampered by the historical devaluation of the teaching career, especially in the exact sciences (Furlan and Miranda, 2015; Silva and Nogueira, 2021).

The results of this research corroborate previous studies by demonstrating that students who drop out tend to leave in the first semesters of the course, confirming the importance of the initial phases for academic permanence (Nogueira and Silva, 2019; Vieira and Andrade, 2020). The abrupt drop in the survival curves of students who abandon the course reveals that evasion is related to a transition phase in which the students are still adapting to university culture. This phase is marked by shocks pedagogical, financial difficulties and, often, the lack of support effective institutional (Barros and Almeida, 2017).

Furthermore, the low retention of the group that changes course suggests that the choice of degree does not always reflect students' vocational preferences. Many enroll due to a lack of options or by the score in ENEM/SISU, which leads to demotivation and reorientation of trajectories still in the basic cycle (Dias and Souza, 2021). These findings are in line with the literature that indicates that non-vocational entry can be a strong predictor of early evasion (Franco and Silva, 2020).

The presence of students who suspend their enrollment for health reasons, even if to a lesser extent number, highlights a relevant point for the discussion: the university can also be a space of illness. According to (Santos and Oliveira, 2018), the pressure for performance, combined with the lack of emotional support, can contribute to psychological suffering, especially among students from working-class backgrounds. Social vulnerability, in this sense, enhances the impact of academic and emotional difficulties.

Survival curves and the Cox model reinforce the efficiency of the methods applied statistics. Studies such as those by (Singer and Willett, 2003) and (Cleves et al., 2010) recommend these techniques for investigating educational events over time, a since they consider both the censorship of the data and the heterogeneity of the

students. In the context of this study, these approaches allowed the identification of patterns risk and the delimitation of critical periods, contributing to more accurate diagnoses.

Thus, the findings reinforce the need for retention policies that hinder access. As pointed out by (Maciel and Oliveira, 2022) and (Nogueira, 2021), the student assistance, tutoring programs and psychoeducational support networks have an impact directly on the continuity of students in higher education. The data from this research indicate that without these policies, the probability of evasion in the first two years is significantly bigger.

The relationship between the type of school of origin and the length of stay must also be considered. Studies show that students from public schools have more vulnerable, with greater adaptation challenges, which may reflect structural inequalities historical (Paixão and Carvano, 2017; Martins and Gomes, 2018). Therefore, facing the evasion requires an intersectional reading that takes into account aspects of class, race, gender and territoriality.

The contribution of this study goes beyond the identification of statistical patterns; it proposes a critical reflection on the limits of Brazilian universities in guaranteeing equal conditions of permanence. As argued by (Frigotto, 2007) and (Saviani, 2008), it is urgent to review the model higher education which, although more accessible in terms of entry, still has structures exclusionary measures that mainly affect poor and peripheral students.

The use of computational tools such as Python, combined with statistical models robust, it also reinforces the commitment to open, replicable and socially impactful science. As discussed by (Davidson and Silva, 2021), reproducible analysis should be a goal of educational science, allowing different institutions to apply methodologies similar in their local contexts.

Therefore, evasion should not be treated as just an individual decision, but as a social phenomenon that requires the articulation between statistical diagnosis and institutional action concrete. The data presented here provide support for the formulation of public policies fairer and more effective, focused on equity and the real democratization of higher education.

6. FINAL CONSIDERATIONS

This study analyzed the retention and dropout patterns of students in the Degree in Mathematics from the Federal University of Viçosa – Florestal Campus, using survival analysis as a central statistical methodology. The application of techniques such as Kaplan-Meier, log-rank and Cox regression allowed identifying critical moments in the trajectory academic performance and factors associated with early dropout, course change and completion of training.

The results reveal that dropout occurs mainly in the first two years of course, reinforcing the need for institutional actions aimed at welcoming entrants. The overlap between socioeconomic, institutional and emotional factors highlights the complexity of the phenomenon and rejects simplistic explanations centered on performance individual academic.

The analyses indicate that students from working-class backgrounds, from public schools and who enter the degree course due to lack of options are more vulnerable to dropout. Because on the other hand, permanence is strongly associated with the existence of assistance policies student and institutional support networks that guarantee not only access, but conditions concrete steps to continue studying.

In addition to highlighting the challenges faced by undergraduate courses, this article proposes the use of reproducible and open methodologies as a diagnostic tool for policies more effective public policies. Replication of this analytical model in other contexts could contribute to the construction of national strategies to combat school dropouts superior.

Finally, by adopting a rigorous quantitative approach combined with a critical reading of the educational context, this work is part of the study agenda that defends a inclusive, democratic and socially referenced university, capable of understanding and facing the multiple dimensions of student dropout.

References

Araújo, FP and Lopes, MS (2018). Dropout in undergraduate courses: determinants and implications for public policies. *Brazilian Journal of Politics and Administration Education*, 34(3):539–556.



- Barros, R. and Almeida, CE (2017). Dropout rates in undergraduate courses: challenges and public policies. *Brazilian Journal of Education*, 22(69):240–265.
- Cleves, M., Gould, W., and Gutierrez, R. (2010). *An Introduction to Survival Analysis Using Stata*. Stata Press, 3 edition.
- Cunha, APD and Carvalho, T. (2016). Social inequalities and school dropout rates Brazilian higher education. *Education and Society*, 37(135):685–704.
- Education, M. (2022). Higher Education Census 2022. Accessed on: April 17, 2025. Available at: <https://www.gov.br/mec>.
- Davidson, R. and Silva, J. (2021). Open science and reproducibility in educational research. *Educational Researcher*, 50(2):68–75.
- Dias, CS and Souza, FG d. (2021). Motivations for choosing a degree: a study with mathematics students. *Educar em Revista*, 37(1):1–20.
- Franco, JP and Silva, MO (2020). Factors associated with dropout in higher education: a institutional analysis. *Evaluation (Campinas)*, 25(3):599–620.
- Frigotto, G. (2007). The university in ruins: crisis and perspectives. *Education and Society*, 28(100):873–897.
- Furlan, N. and Miranda, AL (2015). Dropout in undergraduate mathematics courses: aspects social and institutional. *Educational Dialogue Journal*, 15(46):815–832.
- Hosmer, D. W., Lemeshow, S., and May, S. (2008). *Applied Survival Analysis: Regression Modeling of Time to Event Data*. Wiley-Interscience, 2nd edition.
- Kleinbaum, DG and Klein, M. (2012). *Survival Analysis: A Self-Learning Text*. Springer, 3rd edition.
- Maciel, L. and Oliveira, T. d. S. (2022). Impact of retention policies on student trajectories: a longitudinal analysis. *Education in Review*, 38:1–25.
- Martins, R. and Gomes, F. (2018). Access and permanence in public universities: the role of student assistance. *Brazilian Journal of Education*, 23:1–20.



- McKinney, W. (2021). Educational data science: The role of python in modern research. *Journal of Open Education*, 6(2):112–129.
- Nogueira, M. d. M. (2021). Student retention in Brazil: public policies and challenges contemporaries. *Exitus Magazine*, 11(1):60–80.
- Nogueira, M. d. M. and Silva, T. A. d. (2019). The retention of students of foreign origin popular in higher education: dilemmas and student assistance policies. *Brazilian Journal of Education Policy and Administration*, 35(2):325–344.
- Paixão, M. and Carvano, L. (2017). Educational inequalities and race in Brazil: reflections on access and permanence. *Research Notebooks*, 47(166):148–181.
- Santos, LF d. and Oliveira, BL (2018). Mental health and psychological distress of students university students: a case study. *Brazilian Journal of Professional Guidance*, 19(2):183–195.
- Saviani, D. (2008). Education and democracy: educational policies and the right to education in brazil. *HISTEDBR Journal*, 8(30):5–26.
- Silva, T. d. O. and Nogueira, M. d. M. (2021). Inequalities in student retention: the case of undergraduate degrees in science. *Research Notebooks*, 51(178):188–211.
- Singer, J.D. and Willett, J.B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. OxfordUniversity Press.
- Tinto, V. (1993). *LeavingCollege: Rethinking the Causes and Cures of Student Attrition*. University of Chicago Press.
- Van Rossum, G. and Foundation, P. S. (2020). Python 3.10 documentation. <https://docs.python.org/3.10/>.
- Vieira, PC and Andrade, LF d. (2020). Socioeconomic profile and access to university: permanence and social vulnerability. *Education and Society*, 41:1–24.