

Title: Regional Disparities in Brazil’s Core STEM Education (2013–2023): Is There a STEM Divide?

Author: Patricia Bonini

Author: Patricia Bonini Affiliation: State University of Santa Catarina – patricia.bonini@udesc.br

Área: Crescimento Econômico e Desenvolvimento Regional

Resumo

Este artigo investiga as disparidades regionais na composição interna dos cursos de graduação em STEM (Ciência, Tecnologia, Engenharia e Matemática) no Brasil entre 2013 e 2023. A literatura mostra que as regiões Norte e Nordeste concentram menos programas em áreas STEM e apresentam inserção mais frágil no mercado de trabalho tecnológico. No entanto, este trabalho dá um passo adiante ao examinar, entre os cursos classificados como STEM, quais são mais propensos a pertencer ao grupo “core-STEM” — um subconjunto mais técnico e alinhado à inovação, que inclui Matemática, Ciências Naturais, Tecnologias da Informação e Engenharia. Utilizando microdados do Censo da Educação Superior (INEP), estimamos modelos logit para calcular a probabilidade de um curso pertencer ao grupo core-STEM. As estimativas são conduzidas separadamente para cursos presenciais e a distância (EAD), com interações temporais e variáveis institucionais (gratuidade) e demográficas (gênero, raça e tipo de ensino médio (público ou privado)). Os resultados indicam que a oferta de core-STEM é desigualmente distribuída, com predominância nas regiões Sudeste e Sul, o que reforça o padrão observado no mercado de trabalho. Além disso, observam-se associações negativas significativas entre a presença de mulheres, estudantes não brancos e egressos de escolas públicas e a probabilidade de um curso ser core-STEM, embora estas associações tenham se atenuado no caso dos cursos EAD entre 2013 e 2023. Concluímos que o Brasil não apenas enfrenta um déficit na quantidade de cursos STEM em certas regiões, mas também uma desigualdade qualitativa quanto ao tipo de formação oferecida. Esta “divisão STEM” pode reforçar desigualdades regionais de longo prazo em produtividade, inovação e inclusão no mercado de trabalho tecnológico.

Abstract

This study investigates regional disparities in the internal composition of STEM (Science, Technology, Engineering, and Mathematics) undergraduate programs in Brazil between 2013 and 2023. While previous research has shown that the North and Northeast regions have fewer STEM programs and weaker integration into the technological labor market, this study goes further by examining which existing STEM programs are more likely to belong to the “core-STEM” group—technical and innovation-driven fields including Mathematics, Natural Sciences,

Information and Communication Technologies (ICT), and Engineering (ISCED 51, 52, 53, 61). Using microdata from the Brazilian Higher Education Census (INEP), we estimate logit and probit models to assess the probability that a course belongs to the core-STEM group. Separate models are run for in-person and distance education (EAD), including time interactions and institutional (tuition-free) and demographic variables (gender, race, and high school type). Results show that core-STEM programs are disproportionately concentrated in the Southeast and South, reflecting patterns also seen in the labor market. Moreover, we find strong negative associations between the presence of women, non-white students, and public school graduates and the likelihood of a course being core-STEM—although these effects have weakened in EAD between 2013 and 2023. We conclude that Brazil faces not only a quantitative STEM gap but also a qualitative divide in the type of STEM education offered across regions. This “STEM divide” risks deepening long-term inequalities in productivity, innovation, and access to high-opportunity labor markets.

Palavras-chave: Educação Superior; STEM; Desigualdade Regional

Keywords: Higher Education; STEM; Regional Inequality

Área temática sugerida Economia da Educação / Desenvolvimento Regional / Economia do Setor Público

Classificação JEL – **I123 J24 O15**

I. Introduction

This study investigates regional disparities in the internal composition of STEM (Science, Technology, Engineering, and Mathematics) undergraduate programs across Brazilian higher education institutions between 2013 and 2023. While prior research has highlighted that the North and Northeast regions offer fewer STEM programs in absolute numbers, this research takes a step further by asking: among existing STEM programs, how likely are they to belong to the more technical and innovation-oriented subset known as core-STEM?

The goal is to assess whether regional imbalances persist or shift over time and to what extent socio-demographic and institutional factors explain these patterns. Specifically, the analysis tests whether the probability that a STEM program falls within the core-STEM group—which includes Mathematics and Statistics (ISCED 51), Natural Sciences (52), Information and Communication Technologies (53), and Engineering (61)—varies systematically across Brazil’s macro-regions and over time.

This notion of a “STEM divide” suggests that regional inequalities may manifest not only in the quantity of STEM programs offered but also in the types of STEM education available. These disparities can reinforce broader technological and

economic asymmetries, limiting the potential for inclusive innovation and regional development.

To establish context, the study begins by examining the evolution of the share of STEM programs relative to all undergraduate courses across Brazilian regions over the past decade. As shown in Figure 1, the Southeast consistently leads in the proportion of STEM offers, approaching 20% in recent years. The South and Center-West follow, while the North and, especially, the Northeast exhibit lower shares, despite modest growth from 2019 to 2023. These descriptive trends underscore the persistent asymmetries in access to scientific and technological training across Brazil.

This motivates a deeper investigation into what types of STEM programs are offered in each region. Labor market evidence indicates that graduates from core-STEM disciplines, especially in ICT and Engineering—are disproportionately concentrated in the South and Southeast, regions that also benefit from stronger innovation ecosystems and higher average incomes. By contrast, STEM programs in the North and Northeast are more heavily represented in life sciences and health fields, which, although essential, are typically less aligned with high-growth, technology-intensive sectors.

To analyze these patterns, this study employs nationally representative microdata from the INEP Higher Education Census for the years 2013 and 2023. Using logit and probit models restricted to STEM programs, the analysis evaluates whether institutional attributes (such as tuition-free status and delivery modality) and student composition (gender, race, and high school background) are significantly associated with the probability of a course being classified as core-STEM.

By comparing in-person and distance education (EAD) modalities and including interaction terms to capture temporal dynamics, the study explores whether demographic and regional inequalities have intensified, moderated, or shifted over the last decade. Ultimately, this research contributes to a better understanding of how structural disparities in higher education shape Brazil's capacity to train a diverse and regionally distributed STEM workforce capable of advancing national development and innovation goals.

II Literature review

The analysis of higher education data offers valuable insights into the training paths of professionals who will shape the STEM (Science, Technology, Engineering, and Mathematics) segment of Brazil's labor market. The quality and composition of the STEM-educated workforce are fundamental to a country's ability to generate and diffuse technological advancements, which, in turn, underpin long-run economic growth—as emphasized in the theoretical

models of Romer (1990) and Lucas Jr. (1988). In this context, the importance of STEM labor lies in the fact that these professionals are the primary agents of innovation and technological diffusion (Goldin et al., 1992).

The present investigation into regional disparities in the composition of STEM higher education programs in Brazil builds upon prior evidence of uneven labor market outcomes in this field. Bonini (2024) showed that STEM professionals are not evenly distributed across Brazilian regions: the Southeast and South concentrate a greater share of higher-paying, technology-intensive occupations, whereas the North and Northeast are more represented in STEM roles related to life sciences and agriculture—fields that typically offer lower salaries and are less tied to technological innovation.

These labor market asymmetries may reflect, and be reinforced by, regional differences in the structure and content of higher education STEM offerings. By examining the likelihood that a program is classified as "core-STEM"—a subset of fields most directly associated with scientific and technological advancement—this study tests whether Brazil's regional STEM divide is already evident at the stage of university education. In doing so, it contributes to our understanding of how educational trajectories may perpetuate or mitigate long-term regional inequalities in Brazil's knowledge economy.

The broader landscape of higher education in Brazil has undergone substantial expansion and restructuring in recent decades. Although the growth in medical education is notable—with 328 active courses and over 35,000 openings in 2020 (Pereira et al., 2021)—the majority of undergraduate enrollments remain concentrated in Education, Social Sciences, Business, and Law (Machado de Campos, 2020). STEM fields, by contrast, still account for a relatively modest share of course offerings and enrollments, with significant regional disparities in access and institutional capacity.

A complementary aspect to this discussion is the growth of Brazil's innovation ecosystem, particularly in the startup sector. Between 2015 and 2020, Brazil saw a sharp rise in unicorn companies, with nine startups reaching valuations over one billion dollars and attracting over nine billion reais in investment (Arcentales et al., 2021). This 550% increase in unicorn count and 769% rise in valuation underscore the centrality of digital innovation and technological capacity in Brazil's economic dynamism (Ferreira et al., 2021). The success of these ventures is attributed to corporate venture capital, innovation hubs, and the catalytic role of accelerators. Importantly, the emergence of unicorns in traditional sectors such as real estate (Damasceno et al., 2021) highlights how STEM competencies can drive transformation even in legacy industries.

These technological advancements, however, are constrained by structural inequalities in digital infrastructure. According to the 2023 PNAD Contínua TIC (IBGE), 88% of Brazilians aged 10 and over have internet access, but access

is uneven: while 94.1% of Southeast residents report connectivity, the rates fall to 90.4% in the North and 89.1% in the Northeast. More critically, only 22% of Brazilians experience “meaningful connectivity,” and this number drops to 11% in the Northeast, according to Cetic.br. Arretche (CEM-USP) further highlights how low-quality digital access and skill acquisition have historically limited educational outcomes in these regions.

World Bank and Ookla (2023) data on broadband performance near educational facilities show that many urban centers in the North and Northeast, such as Manaus and Fortaleza, do not meet the threshold for e-learning readiness. These infrastructural shortcomings severely hinder the potential for online STEM education, particularly in areas such as ICTs, and have disproportionately affected students from public schools and low-income backgrounds—especially during the COVID-19 pandemic, when nearly six million students were disconnected from any form of online learning.

These findings lend strong support to the hypothesis of a regional STEM divide: digitally underserved regions are less equipped to deliver quality STEM education and less likely to cultivate or retain core-STEM talent. Digital exclusion thus functions as both a symptom and a cause of educational and technological inequality, deepening pre-existing disparities across Brazil.

The present research builds directly on this context by investigating whether the unequal distribution of core-STEM labor outcomes across Brazil has its origins in higher education. By analyzing how course characteristics and student composition relate to the probability of a course being core-STEM, the study adds an important dimension to debates on regional inequality, innovation, and human capital development Brazil.

In terms of labor market outcomes, international evidence shows that the STEM wage premium varies significantly by country and demographic group. In the U.S., STEM jobs command higher wage premiums despite a relatively large STEM workforce (Even et al., 2023). In Brazil, STEM workers represent just 1.8% of formal employment, and wage premiums vary substantially by region, reflecting structural imbalances (Bonini et al., 2025). In Australia, women with STEM degrees face multiple barriers including lower employment rates and wider wage gaps compared to their male counterparts (Dockery & Bawa, 2018). These diverse findings highlight that the STEM advantage is not uniform but shaped by national contexts and intersectional inequalities factors that are highly relevant to understanding the Brazilian case.

III. Data and methodology

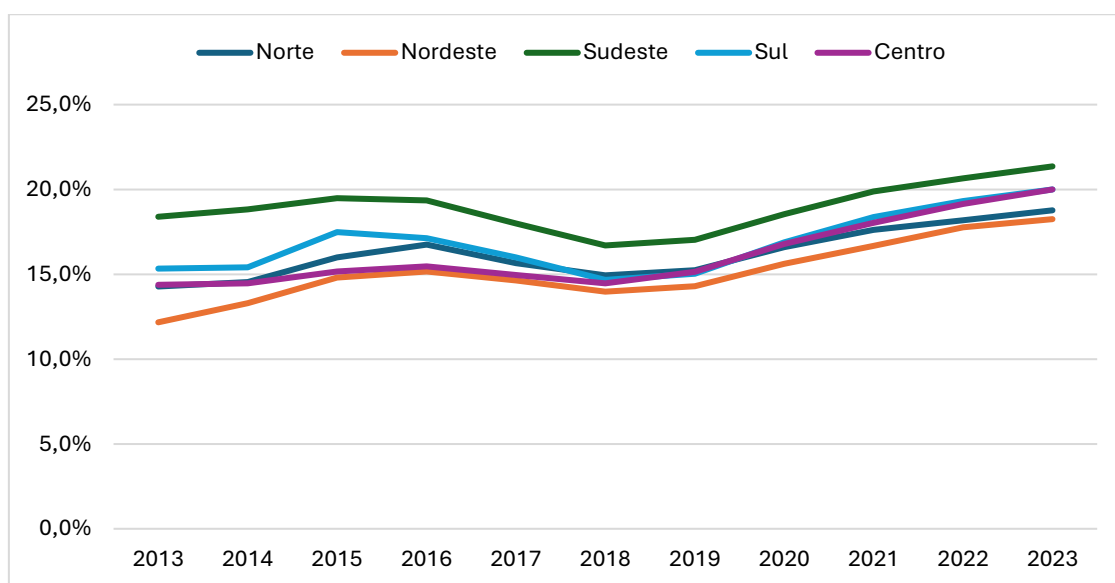
Data used in this study come from the Brazilian Higher Education Census (Censo da Educação Superior), provided by the Instituto Nacional de Estudos e

Pesquisas Educacionais Anísio Teixeira (INEP), for the years 2013 and 2023, restricted to undergraduate courses. To ensure international comparability, we adopt the International Standard Classification of Education – Fields of Education and Training (ISCED-F 2013) developed by UNESCO. The final dataset includes both in-person and distance learning (EAD) modalities and captures institutional characteristics as well as aggregated demographic profiles of enrolled students. The core STEM areas encompass science, technology, engineering, and mathematics, along with related disciplines like environmental and computer sciences - Assefa & Rorissa (2013).

The Standard Occupation Classification Policy Committee has defined two major domains for STEM occupations: core STEM fields and STEM-dependent occupations – Jones (2014). Table 1 in the Appendix briefly describes the STEM areas including STEM-dependent areas.

Chart 1 displays the evolution of the percentage of STEM programs in the total number of undergraduate in person programs across Brazil's five microregions from 2013 to 2023.

Chart 1 - STEM participation in the total courses (in persons programs)



Source: Author's preparation base on INEP

Throughout the period, the Southeast region consistently maintained the highest share of STEM courses, followed by the South and Midwest. In contrast, the Northeast and North regions had the lowest relative participation in STEM fields, although both experienced gradual growths starting around 2018.

The overall trend reveals a moderate increase in the share of STEM courses nationwide, especially after 2018. Yet the decline observed between 2016 and 2018—especially in the South and Southeast—may be linked to funding constraints, institutional restructuring, or the effects of reforms in student

financing mechanisms, such as changes to FIES and PROUNI. The convergence in regional trajectories after 2018 suggests some reduction in disparity, but important regional gaps remain.

Table I - Percentage of Distance Learning (EAD) Courses

Region	2013 STEM	2013 Non-STEM	2023 STEM	2023 Non-STEM
North	42.37%	57.15%	94.08%	94.93%
Northeast	32.04%	55.03%	93.37%	94.30%
Southeast	26.60%	53.66%	92.80%	94.88%
South	31.44%	60.73%	93.95%	95.89%
Midwest	36.25%	57.94%	93.69%	93.91%

Source: Author's preparation based on INEP

While Chart 1 refers to in person programs, the rise of distance learning courses is illustrated by Table II in the Appendix. Table II illustrates the percentage of distance learning (EAD) courses in STEM and non-STEM programs across different regions of Brazil for the years 2013 and 2023. The data reveals the varying adoption of distance learning (EAD) courses across the five regions for STEM and non-STEM programs from 2013 to 2023. The North region stands out with the highest increase in STEM EAD courses, escalating from 42.37% to an impressive 94.08%. The Northeast follows with a substantial rise from 32.04% to 93.37%. The Southeast region displays the most substantial increase as the adoption of STEM EAD courses, increased from 26.60% to 92.80%. The South and Midwest regions also show significant growth, with STEM percentages reaching 93.95% and 93.69%, respectively. In terms of non-STEM courses, all regions exhibit high percentages in 2023, with the North leading at 94.93%, closely followed by the Northeast at 94.30%. Overall, while all regions are embracing distance learning, the North region shows the most pronounced growth in STEM offerings, highlighting a shift towards increased accessibility in higher education across Brazil.

The rise of distance learning (EAD) in Brazil over the past decade has marked a significant transformation in the educational landscape of the country. This shift has been particularly accelerated by the COVID-19 pandemic, which demanded a rapid transition to online education across various institutions. As educational institutions increasingly adopted online platforms, the accessibility and flexibility of EAD courses attracted a growing number of students, particularly in the realms of STEM (Science, Technology, Engineering, and Mathematics) and non-STEM fields. This trend has been further driven by technological advancements and changes in educational policy aimed at enhancing inclusivity. Consequently, the percentage of EAD courses has surged across all regions, reflecting a broader

movement towards modernizing education and meeting the diverse needs of learners throughout Brazil.

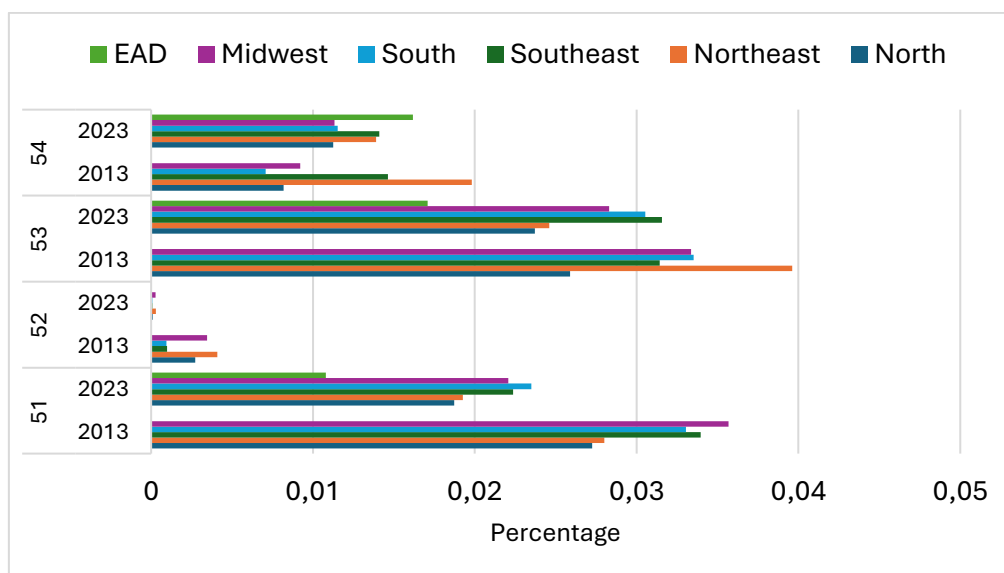
III.1 Regional Patterns and Temporal Shifts in STEM Fields (2013–2023)

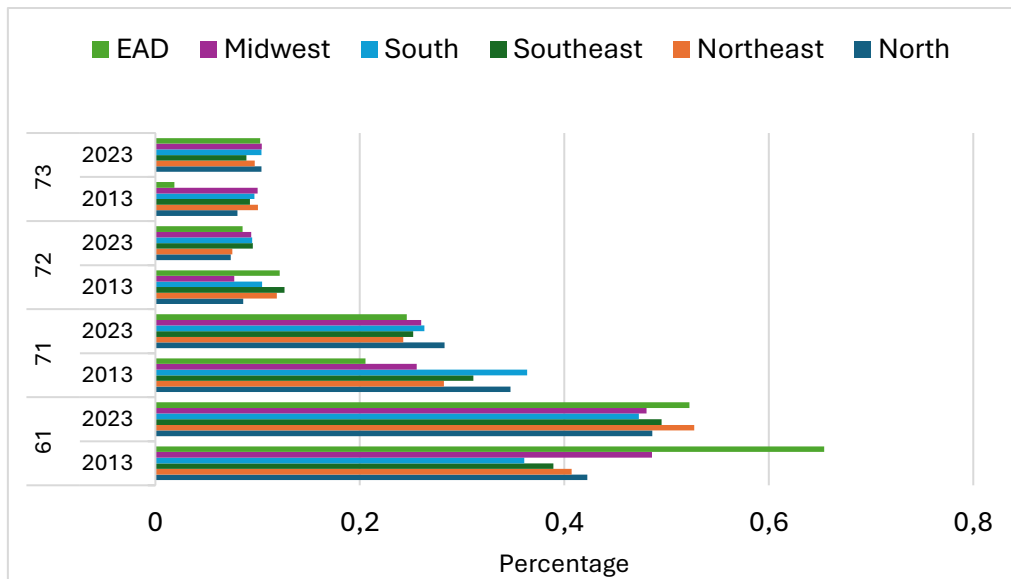
To better understand the structural composition of STEM education across Brazilian regions, this study narrows the scope to a subset of fields considered as “**core STEM.**” This classification includes disciplines that are most consistently associated with traditional science, technology, engineering, and mathematics education and labor market demands. Based on the ISCED-F 2013 classification, the following specific fields were selected as core STEM areas: 51, 52, 53, 61 and 73

These areas were chosen due to their stronger ties to industrial innovation, technological development, and scientific research, as opposed to broader or more applied domains like life sciences (code 54) or architecture and production technologies (codes 71 and 72). The data is disaggregated by Brazil's five macro-regions and the EAD modality (Distance Learning), which is treated separately due to lack of geographic attribution.

The two charts present the proportion of each STEM subfield (classified by ISCED-F codes) relative to the entire STEM group in 2013 and 2023. For better visual clarity, the fields are divided across two charts.

Chart 2 The distribution of in person coreStem and distance learning courses





Source: author's preparation base on INEP dataset

Chart 2 presents the distribution of STEM subfields corresponding to areas 51 through 54, namely: Mathematics (51), Natural Sciences (52), Information and Communication Technologies (ICT – 53), and Life Sciences (54). As anticipated, the Southeast and South regions exhibit the highest proportions of area 53 (ICT), reaffirming their established role as centers of technological development and innovation. Conversely, the North and Northeast show markedly lower participation in ICT fields—an expected outcome that reflects enduring disparities in access to digital infrastructure and limited demand for technology-intensive occupations in these regions.

Complementing this pattern, the North and Northeast maintain a comparatively higher share of Life Sciences (area 54) within their STEM composition. This regional emphasis is likely tied to local public health needs and the strategic relevance of biodiversity-related research, aligning with broader expectations regarding their scientific and developmental priorities. In contrast, Life Sciences represent a smaller proportion of STEM courses in the South and Southeast. Mathematics (51) and Natural Sciences (52) account for only marginal shares across all regions, with relatively even distributions, indicating their limited but stable presence.

Chart 2.b illustrates fields traditionally grouped under Engineering and Applied Technologies: Engineering (61), Manufacturing and Processing (71), Architecture and Construction (72), and Technical Engineering (73). Engineering (area 61) dominates the STEM landscape nationwide, with modest regional variation. In 2023, a notable rise in Engineering programs offered via distance education (EAD) is observed, raising important questions about the feasibility and pedagogical quality of delivering such technically demanding curricula remotely.

Area 71 (Manufacturing and Processing) appears slightly more concentrated in the South and Southeast, consistent with their established industrial bases. Though still modest, this field has grown incrementally in other regions over the past decade. Architecture (72) and Technical Engineering (73) remain minor but stable components of the STEM offering, with slight increases in EAD formats by 2023.

With respect to temporal dynamics in the period, Engineering (61) maintains its position as the dominant STEM area, with a gradual increase across regions—especially in EAD programs. ICT (53) is one of the fastest-growing fields, particularly in the Southeast and Midwest. Yet, this field suffered a decline in the Northeast, which suggests our hypothesis of a STEM divide.

Life Sciences (54) remains stable or slightly declines nationally, but consistently more prominent in the North and Northeast, underscoring their regional emphasis on health and environmental research. Manufacturing (71) and Architecture (72) exhibit small but positive shifts, mainly in regions with consolidated industrial and construction activity.

Distance learning (EAD) presents a distinctive profile in the STEM landscape, characterized by high concentrations of programs in Engineering (61) and ICT (53). The growing presence of distance education in fields that typically require practical training brings to the forefront important considerations about curriculum adequacy, educational quality, and workforce preparedness. In contrast, EAD remains marginal in areas such as Architecture and Natural Sciences.

III.3 The regression model

To examine whether demographic characteristics influence the probability of a course being classified as core-STEM, we opt to use graduate (completion) data as it reflects academic persistence and successful graduation. This choice allows us to assess how the representation of historically underrepresented groups evolves not only in access but also in successful outcomes, particularly in the more demanding STEM subfields. It also avoids distortions caused by high dropout rates, which are especially relevant in courses with greater academic rigor.

The first model estimates the probability that a course belongs to the core-STEM group, among all undergraduate programs. The main explanatory variables capture the composition of graduate students by gender, race, and school background, with interaction terms to test whether the effects differ between 2013 and 2023. Regional dummies are included, with the Southeast as the reference category.

$$\text{logit} (\text{Pr coreSTEM} = 1) = \alpha + \beta_1 \text{gender} + \beta_2 \text{race} + \beta_3 \text{public} + \beta_4 \text{year}_{2023} + \beta_5 (\text{gender} * \text{year}) + \beta_6 (\text{race} * \text{year}) + \beta_7 (\text{public} * \text{year}) + \beta_9 \text{Tuition} + \sum_{r \neq \text{Southeast}} \delta_r \text{region} + \varepsilon_i \quad (1)$$

Since the geographic information is unavailable for distance learning courses, Equation (1) is fitted for two sets of data, which are the in-person programs and the distance-learning courses. In this latter, the last term of the equation (1) referring to the region where the program is located is dropped from the model.

The equation (2) describes the model for the restricted sample, estimating the probabilities of a program being core-STEM withing the set of STEM programs.

$$\text{logit}((Pr \text{ coreSTEM} = 1 | \text{STEM} = 1)) = \alpha + \beta_1 \text{gender} + \beta_2 \text{race} + \beta_3 \text{public} + \beta_4 \text{year}_{2023} + \beta_5 (\text{gender} * \text{year}) + \beta_6 (\text{race} * \text{year}) + \beta_7 (\text{public} * \text{year}) + \beta_9 \text{Tuition} + \sum_{r \neq \text{Southeast}} \delta_r \text{region} + \varepsilon_i \quad (2)$$

This model is also fitted for in-person and EAD courses, separately and the variable appearing in the equation are shown in Table II.

Table II - Variables Included in the Model

Variable	Type of Variable	Description
STEM	Binary variable (0,1)	=1 for broad STEM set of “código-cine-área-especifica”
Core STEM	Binary variable (0,1)	=1 for Core-STEM áreas
Fem	Continuous variable	Proportion of female graduates in relation to total graduates in the respective year
Race	Continuous variable	Proportion of non-white graduates - those that declare themselves as blacks, pardos, yellow, indigenous or do not declare race in relation to total graduates
Public	Continuous variable	Proportion of graduates from public high schools
Year	Binary variable (0,1)	=1 if year=2023 – used for pooled regressions and interaction terms to capture temporal shifts
Tuition	Binary variable (0,1)	=0 if there is a tuition fee, =1 if the course is tuition-free
Region	Categorical variable	Indicates the macro-region where the institution is located

By estimating equation (2), we investigate disparities within the STEM domain by restricting the sample to courses already classified as STEM (i.e., STEM = 1). This restriction enables a more focused comparison of the probability that a given STEM course belongs to the more technical “core-STEM” group. The model incorporates interaction terms and regional fixed effects, along with a binary indicator for whether the course is tuition-free.

This specification provides a framework to assess whether the influence of gender, race, and high school background on the likelihood of offering core-STEM

programs has shifted over time—particularly in the context of the expansion of distance education (EAD) and enduring regional digital inequalities.

IV. Results

In this section, we discuss the main results of the logit model summed up by equations (1) and (2).

IV.1 Estimation of equation (1) for in person programs

The coefficients for regional dummies indicate that, when compared to the Southeast (the omitted reference category), courses located in the North, Northeast, and South are significantly less likely to belong to the core-STEM group. Courses in the Northeast show the largest negative coefficient (-0.370), suggesting a lower presence of core-STEM courses relative to the Southeast.

The North (-0.241) and South (-0.152) also show significantly lower probabilities, while the Center-West region does not differ significantly from the Southeast ($p = 0.592$).

These results reinforce the hypothesis of a regional STEM divide, with the Southeast maintaining a leadership position in core-STEM offerings among presential courses.

Female participation exhibits a strong and highly significant negative association (-4.20 , $p < 0.001$), indicating that courses with a higher proportion of female graduates are substantially less likely to belong to the core-STEM group.

Unexpectedly, the race characteristic - Non-White Participation rate - has a positive main effect, although not statistically significant ($p = 0.188$), suggesting no clear association in 2013. Similarly, the proportion of graduates from public high schools - Public - is not significant in the baseline year ($p = 0.506$).

The variable indicating whether the course is tuition-free (Tuition) is positive and strongly significant (0.41 , $p < 0.001$), suggesting that free public institutions are more likely to offer core-STEM programs, possibly due to their technical infrastructure and alignment with national development goals.

In terms of changes in the role of the explanatory variables over time, the time interaction terms, the female participation and year term was not statistically significant ($p = 0.387$), suggesting that the negative relationship between female representation and core-STEM classification has remained stable over the decade. Yet, the interaction term race x year is negative and significant (-0.320 , $p = 0.002$), indicating that in 2023, courses with more non-white students are less likely to be classified as core-STEM than in 2013. This marks a worsening in racial disparities within the STEM field over time.

The interaction Public High School \times Year is also negative and significant (-0.284 , $p = 0.006$), suggesting that in 2023, a higher proportion of students from public high schools reduces the likelihood of a course being core-STEM, compared to 2013. This may reflect growing inequality in access to more technical STEM fields.

The marginal effects of the proportion of female students on the probability of a course being classified as core-STEM reveal a strong and negative association in both 2013 and 2023. In 2013, a one-unit increase in the proportion of women among course graduates is associated with a 32.1 percentage point decrease in the probability that the course belongs to the core-STEM group ($dy/dx = -0.321$, $p < 0.001$). In 2023, this effect remains negative and statistically significant but is slightly attenuated, with a reduction of 27.6 percentage points ($dy/dx = -0.276$, $p < 0.001$).

This result indicates that, although the gender gap in core-STEM participation persists, its intensity may have moderated slightly over the last decade. The narrowing of the marginal effect suggests incremental progress in reducing gender-based disparities within technical STEM fields, although the overall association remains strongly negative.

The marginal effects analysis regarding racial composition reveals an important shift between 2013 and 2023. In 2013, the proportion of non-white graduates in a course did not significantly affect the probability of the course being classified as core-STEM. The average marginal effect was small and statistically insignificant ($dy/dx = 0.0074$, $p = 0.186$), suggesting no systematic racial pattern in the distribution of core-STEM courses at that time. However, by 2023, the relationship becomes negative and statistically significant. The marginal effect indicates that a one-unit increase in the proportion of non-white students in a course is associated with a 1.5 percentage point decrease in the probability that the course belongs to the core-STEM group ($dy/dx = -0.0152$, $p = 0.005$).

This result suggests that, over time, core-STEM fields have become less accessible or less representative of non-white students relative to other STEM areas. It reinforces the notion that, while inclusion may have increased in the broader STEM domain, core-STEM fields continue to present racial barriers, which may have even intensified.

The marginal effects of public-school background on the likelihood of a course being core-STEM have changed meaningfully over time. In 2013, the association between the share of students from public high schools and the probability of a course being classified as core-STEM was statistically insignificant ($dy/dx = -0.0033$, $p = 0.504$). This indicates that, at the beginning of the period, the representation of public school background graduates in a course did not significantly affect its likelihood of belonging to the core-STEM category.

However, by 2023, the relationship turned significantly negative. The marginal effect estimate suggests that a one-unit increase in the proportion of students with a public high school background is associated with a 2.2 percentage point

reduction in the likelihood that the course is in the core-STEM group ($dy/dx = -0.0221$, $p < 0.001$).

This result signals a growing social disparity within the core-STEM fields over the decade: even as access to STEM education may have broadened overall, core-STEM areas have become more exclusive with respect to socioeconomic background, as measured by public school attendance. In other words, these programs appear increasingly concentrated among students from more privileged educational trajectories.

IV.2 Estimation of equation (1) for distance learning programs

The regression results reveal key associations between student composition and the likelihood of a distance learning course being categorized as core-STEM. The model is statistically significant ($p < 0.001$), with a pseudo R^2 of 0.1286, suggesting a moderate explanatory power for a binary outcome.

The proportion of female students (variable *fem*) is strongly and negatively associated with the likelihood of a course being in the core-STEM group. The main effect is highly significant ($z = -25.17$, $p < 0.001$), and its magnitude is very large (coefficient ≈ -3.996), suggesting that courses with more women are much less likely to be core-STEM. This aligns with gender disparities traditionally seen in STEM fields. Importantly, the interaction term with “year” is positive and significant (coef = 1.494, $p < 0.001$). This indicates that the negative effect of female presence on core-STEM classification has decreased over time. In other words, although female representation still correlates negatively with core-STEM fields, the penalty is weakening between 2013 and 2023 for EAD courses.

The main effect of race (proportion of non-white students) is also significantly negative (coef = -0.365 , $p < 0.001$), suggesting that a higher share of non-white students is associated with a lower likelihood of the course being EAD core-STEM.

However, the interaction with time yields a positive and marginally significant (coef = 0.172, $p = 0.080$), indicating a potential attenuation of this negative effect over time. While still tentative, this may suggest that racial barriers to entry in core-STEM areas through EAD modalities are decreasing.

The share of students who completed public high school shows no significant effect overall ($p = 0.544$), but the interaction with year is negative and significant (coef = -0.198 , $p = 0.030$). This suggests that, over time, courses with higher public school representation have become less likely to be core-STEM in the EAD system.

The variable “year” (2023 = 1) has a positive, marginally significant effect (coef = 0.167, $p = 0.064$), implying a slight increase in the probability of a course being core-STEM in 2023. Meanwhile, the “Tuition” (whether the course is tuition-free) coefficient shows no significant effect in this subsample of EAD courses.

IV.3 Estimation of equation (2) for in person courses

In our main empirical analysis, we restrict the sample to undergraduate courses classified within the broader STEM category ($\text{STEM} == 1$). This decision is theoretically grounded and empirically advantageous for addressing the central research question: Given that a course belongs to STEM, what factors increase the likelihood that it falls within the more technical "core-STEM" group?

By focusing only on STEM courses, we eliminate noise introduced by comparisons with non-STEM fields, which may differ dramatically in structure, purpose, and student composition. This more refined comparison allows us to capture intra-STEM variation, revealing how the internal composition of STEM fields differs across regions and evolves over time.

The logistic regression results from the restricted model are substantively stronger and more aligned with the STEM divide hypothesis.

The proportion of female students is strongly and negatively associated with the likelihood of a course being core-STEM, and this negative effect becomes even stronger in 2023, as shown by the significant negative interaction term with year.

Also, the proportion of nonwhite graduates is positively associated with core-STEM in 2013, but this relationship weakens in 2023, as indicated by a statistically significant negative interaction term. The proportion of students from public high schools is positively associated with core-STEM composition, although its effect appears to diminish over time (with a marginally insignificant interaction term).

These patterns provide compelling evidence of persistent and evolving socio-demographic disparities within the STEM landscape itself, not just between STEM and non-STEM fields. The core-STEM group—composed of Mathematics and Statistics (51), Natural Sciences (52), ICT (53), and Engineering (61)—remains unequally distributed across student profiles and regions.

Therefore, by narrowing the sample to STEM courses, the model delivers clearer insights into the structural barriers and inequalities that shape the internal configuration of technological education in Brazil.

IV.4 Estimation of equation (2) for distance learning courses

The logistic regression model for core-STEM classification among EAD (distance learning) courses reveals statistically significant effects for all explanatory variables included, with particularly strong results for interactions with the year 2023. The likelihood ratio chi-square test ($\text{LR } \chi^2 = 752.36, p < 0.001$) indicates that the model significantly improves upon a null model, although the pseudo R^2 (0.0262) suggests that the explanatory power is modest.

The main effect of the “year” variable is negative and significant ($\beta = -2.803$, $p < 0.001$), indicating that, holding other variables at their means, the probability of a course being classified as core-STEM decreased between 2013 and 2023. However, this result must be interpreted in combination with the interaction terms. The interaction between year and female share is strongly positive and significant ($\beta = 2.767$, $p < 0.001$), suggesting that the negative association between the proportion of women and the likelihood of a course being core-STEM was attenuated in 2023. This is confirmed by the marginal effect for “race” variable: it decreases in magnitude from -0.2277 in 2013 to -0.1854 in 2023, indicating a reduction in gender segregation within the core-STEM group over time.

Similarly, the interaction between year and the share of non-white students is positive and significant ($\beta = 1.948$, $p < 0.001$), suggesting that the initially negative association ($\beta = -2.0519$) between racial diversity and core-STEM likelihood has weakened over time. Marginal effects corroborate this, showing a smaller negative impact in 2023 ($dy/dx = -0.0143$) compared to 2013 ($dy/dx = -0.0208$).

The interaction with the share of students from public high schools is also significant ($\beta = 0.536$, $p = 0.002$), reversing the negative main effect ($\beta = -0.410$). This indicates that while in 2013 the share of students from public schools was negatively associated with core-STEM representation, this association became positive in 2023. The marginal effects reflect this shift, moving from $+0.0029$ in 2013 (not significant) to -0.0108 in 2023 (significant), signaling a complex and possibly nonlinear evolution.

The variable “Tuition”, which captures whether the course is tuition-free, has a large and statistically significant negative marginal effect ($dy/dx = -0.2662$, $p < 0.001$). This suggests that paid EAD courses are substantially less likely to be classified as core-STEM, even after controlling for demographic composition and year effects.

Overall, the results indicate an evolving structure of the core-STEM offer in distance education in Brazil. While the negative association between underrepresented groups (women, non-white students, public school graduates) and core-STEM courses persists, it weakened significantly between 2013 and 2023, particularly for gender and racial composition. These findings suggest a gradual diversification of the core-STEM field in EAD, though institutional and market-driven barriers (e.g., course tuition) still play a significant role in limiting access.

A key distinction emerges when comparing the results for in-person and distance learning (EAD) modalities. For in-person STEM programs, the variable *região* is positively and significantly associated with the likelihood that a course belongs to the core-STEM group ($\beta = 0.0549$, $p = 0.010$). This indicates that courses located in more developed regions—typically the Southeast and South—are more likely to be in technical subfields such as Engineering, ICT, and Mathematics. This

spatial pattern underscores a persistent territorial divide in access to high-value STEM education, reinforcing the uneven distribution of technological capabilities and economic opportunities across Brazilian macro-regions.

In contrast, for EAD courses, regional identifiers are unavailable due to the absence of location-specific data in the INEP microdata. As a result, spatial disparities cannot be directly assessed in this modality. Nonetheless, the comparison is telling while the EAD segment has expanded access to STEM programs nationwide, only in the in-person modality can we observe the concentration of core-STEM offerings in more industrialized regions. This highlights the role of physical infrastructure, institutional resources, and labor market linkages in shaping regional opportunities in STEM education.

IV.5 Recommendations for Future Research

Future studies should explore the effects of recent policy efforts aimed at increasing female participation in STEM fields in Brazil. Although there is not yet a single legal framework with measurable nationwide impact during the 2013–2023 period, recent initiatives—such as the National Strategy for Women in Science, Technology and Innovation (2023), targeted CNPq and CAPES programs like *Girls in STEM*, and gender-equity-oriented components of the BNCC (2017–2018)—suggest an emerging institutional commitment to reducing gender disparities in STEM education. These developments warrant future empirical evaluation. Researchers could investigate whether such programs have influenced female enrollment or completion rates in STEM programs, particularly in technical and engineering fields, and how their effects differ across regions, types of institutions, and racial or socioeconomic groups.

Moreover, as legislative and cultural shifts gain momentum, longitudinal studies may help assess the cumulative impact of these policies over time, especially if new waves of microdata allow for more detailed tracking of student characteristics and educational trajectories. Cross-country comparisons and mixed-methods approaches could further enrich the understanding of how structural, cultural, and institutional factors interact to shape women’s access to and success in STEM education.

Despite providing important insights into the composition of core-STEM courses in Brazilian higher education, this study has some limitations. First, the analysis relies on administrative data that classify students and courses by institutional records, which may not capture deeper motivational, pedagogical, or socioeconomic factors that influence course offerings and student enrollment patterns. In particular, the absence of microdata on student-level socioeconomic status, academic performance, or prior STEM exposure limits the scope of explanatory variables that could account for disparities in core-STEM participation.

Second, while the study includes variables for gender, race, and prior schooling background, it does not fully capture intersectional dynamics—such as how race and gender interact with each other and with institutional variables—to shape

access to technical STEM fields. Third, although the regional analysis uses a proxy (institutional location for in-person courses and national distribution for EAD), it does not include more granular spatial dynamics such as urban-rural differences, local labor market demands, or regional industrial policy efforts to promote technological fields.

Additionally, the categorization of courses into core-STEM vs. non-core STEM, though grounded in international standards, may overlook hybrid or interdisciplinary programs that are increasingly relevant in contemporary technological ecosystems. The models also do not account for supply-side constraints, such as faculty availability or infrastructure limitations, which might affect the regional distribution of core-STEM programs.

Future research could address these limitations by incorporating richer data sources, such as student surveys or institutional censuses, and by developing dynamic models to capture temporal changes in access, retention, and graduation in core-STEM fields. Longitudinal studies could also help establish causal relationships, and qualitative research could complement the statistical findings by exploring how institutional cultures, market signals, or student aspirations shape the evolving STEM landscape in Brazil and comparable emerging economies.

VI. Conclusions

The descriptive results of this research highlight a clear and expected pattern of regional specialization in STEM education across Brazil. The South and Southeast concentrate on ICT and industrial-related fields, while the North and Northeast are more heavily represented in Life Sciences. Over the past decade, Engineering and ICT have shown notable expansion, though unevenly across regions. The expansion of distance learning (EAD) programs, particularly in core technical fields, calls for further research into their effectiveness and their role in shaping regional access to STEM education and labor market integration.

This study deepens the investigation into regional and demographic disparities in the composition of core-STEM courses within the broader STEM field in Brazilian higher education between 2013 and 2023. We assessed how factors such as gender, race, and public high school background influence the probability of a course being classified as core-STEM—defined as programs in mathematics, natural sciences, ICT, and engineering.

The results reveal persistent but evolving structural inequalities. In both modalities—in-person and distance learning—courses with higher proportions of women, non-white students, and public high school graduates are significantly less likely to be classified as core-STEM. However, interaction effects with the year 2023 suggest a reduction in these disparities over time, particularly in distance education. The negative association between female representation and core-STEM classification, for instance, has diminished substantially in distance

learning courses, indicating a gradual diversification of the field. Similarly, the racial gap has narrowed, though it remains statistically significant.

Regional patterns are also noteworthy. For in-person courses, we observe that the likelihood of a STEM program being part of the core-STEM subset remains higher in the South and Southeast compared to other regions, reinforcing concerns over the “STEM divide” in technological training and workforce development. These disparities are less visible in the EAD modality, although the data suggest that institutional and market constraints—such as tuition requirements—continue to limit access to technical STEM programs among underrepresented groups.

By focusing on the internal composition of STEM rather than only its presence across regions, this study advances the literature on educational inequality and offers a more nuanced understanding of how emerging economies like Brazil navigate the global technological race. The findings highlight the importance of designing targeted policies to expand access to core-STEM education—especially in distance learning contexts—as a strategic lever for inclusive development.

References

Arretche, M. (2021). *Desigualdades regionais e o acesso à internet no Brasil: Evidências da PNAD 2015–2019*. Centro de Estudos da Metrópole (CEM/USP). Retrieved from <https://www.fflch.usp.br>

Assefa, S. G. & Rorissa, A.(2013). A Bibliometric Mapping of the Structure of STEM

Education using Co-Word Analysis. *Journal of the American Society for Information Science and Technology*, 1-24

Bonini, P., Da Silva, F., & Sótero, G. (2025). Wage Profile and Gender Gap in Science and Technology: Regional Disparities in Brazil. *Sociedad y Economía*.

Campos, S. R. M., & Henriques, R. (2018). The Brazilian higher education: The undergraduate courses in light of its recent policies. *Education and New Developments*, 371-375

Cetic.br (Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação). (2023). *ICT Households Survey: Executive Summary*. São Paulo: NIC.br. Retrieved from <https://www.cetic.br>

Dockery, A.M., & Bawa, S. (2018). Labour market implications of promoting women's participation in STEM in Australia. *Australian Journal of Labour Economics*, 21, 125-151.

Even, W. E., Yamashita, T., & Cummins, P. A. (2023). The STEM wage premium across the OECD. *Journal Name*, 35(1).

IBGE (Instituto Brasileiro de Geografia e Estatística). (2023). *PNAD Contínua TIC 2023 – Acesso à internet e à televisão e posse de telefone móvel celular para uso pessoal*. Retrieved from <https://www.ibge.gov.br>

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). (2023). **Censo da Educação Superior 2023**. Retrieved from <https://www.gov.br/inep>

Jones, J.I. (2014). An Overview of Employment and Wages in Science, Technology, Engineering and Math (STEM) Groups.

Goldin, C., Katz, L. F., & Kuziemko, I. (1992). *Understanding the gender gap: An economic history of American women*. Oxford University Press.

LUCAS Jr., R. E. (1988). On the mechanics of economic development. *Journal of monetary economics*, Elsevier, v. 22, n. 1, p. 3–42.

Machado de Campos, S. R. (2020). *Brazilian higher education analysis through knowledge discovery: Annual and temporal approaches* (Tese de doutorado). NOVA Information Management School, Universidade Nova de Lisboa.

Pereira, D.V., Fernandes, D.D., Mari, J.F., Lage, A., & Fernandes, A.P. (2021). Mapping of medical schools: the distribution of undergraduate courses and annual vacancies in Brazilian cities in 2020.

ROMER, P. M. (1990). Endogenous technological change. *Journal of political Economy*, The University of Chicago Press, v. 98, n. 5, Part 2, p. S71–S102.

Yowell, J.L., Hanzel-Sello, G., Barnak, G., & Venn, M. Mathematics Success for Underrepresented Community College Students through STEM Core: A Wrap-Around Student Services Model. *2022 ASEE Annual Conference & Exposition Proceedings*.

APPENDIX

Table A.1 STEM specific code areas

Code	Area Description	Examples of Disciplines
51	Mathematics and Statistics	Pure and applied mathematics, statistics, actuarial science
52	Natural Sciences	Physics, chemistry, earth sciences, environmental science, oceanography, geology
53	Information and Communication Technologies (ICT)	Computer science, programming, software development, networks, cybersecurity
54	Life Sciences	Biology, biochemistry, biotechnology, genetics, microbiology, ecology, zoology
61	Engineering and Engineering Trades	Civil, mechanical, electrical, chemical, environmental, and other branches of engineering
71	Manufacturing and Processing	Industrial production, food processing, metallurgy, textiles, wood processing
72	Architecture and Construction	Architecture, urban planning, civil

Code	Area Description	Examples of Disciplines
73	Engineering and Related Technical Fields	construction, infrastructure design Applied engineering, technical maintenance, mechatronics, electronics, technical drawing