UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DE RIBEIRÃO PRETO
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

FELIPE DA SILVA BASSO

EAD funciona? Uma análise do perfil, evasão e desempenho entre os estudantes brasileiros no ensino superior

Orientador: Prof. Dr. Luiz Guilherme Dácar da Silva Scorzafave

RIBEIRÃO PRETO
2023

Prof. Dr. Carlos Gilberto Carlotti Junior
Reitor da Universidade de São Paulo

Prof. Dr. Fábio Augusto Reis Gomes
Diretor da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto

Prof. Dr. Milton Barossi Filho
Chefe do Departamento de Economia

Prof. Dr. Luciano Nakabashi
Coordenador do Programa de Pós-Graduação em Economia – Área: Economia Aplicada

Felipe da Silva Basso

EAD funciona? Uma análise do perfil, evasão e desempenho entre os estudantes brasileiros no ensino superior

Tese submetida ao Programa de Pós-Graduação em Economia - Área: Economia Aplicada da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, para a obtenção do título de Doutor em Ciências. Versão Original.

Universidade de São Paulo

Programa de Pós-Graduação em Economia - Área: Economia Aplicada da FEA-RP

Orientador: Prof. Dr. Luiz Guilherme Dácar da Silva Scorzafave

Ribeirão Preto

2023

Da Silva Basso, Felipe

EAD funciona? Uma análise do perfil, evasão e desempenho entre os estudantes brasileiros no ensino superior/ Felipe da Silva Basso – Ribeirão Preto, 2023.

154f. il.:

Tese submetida ao Programa de Pós Graduação de em Economia – Área: Economia Aplicada da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, para a obtenção do título de Doutor em Ciências – Universidade de São Paulo

Orientador: Prof. Dr. Luiz Guilherme Dácar da Silva Scorzafave

1. EAD. 2. Evasão. 3. Desempenho. 4. Ensino Superior.

AGRADECIMENTOS

Aos meus pais Carlos Basso e Solange Terezinha Rosa da Silva pelo amor e apoio incondicional. E meus irmãos Heloisa da Silva Basso e Marcos da Silva Basso pelo companheirismo, além de todos os outros familiares que sempre estiveram torcendo por mim.

A todos os professores e funcionários da FEA-RP\USP, em especial a Prof. Dr. Luiz Guilherme Dácar da Silva Scorzafave que foi um grande professor e orientador e com quem tive a oportunidade de conviver e aprender muito ao longo do doutorado e à Prof. Dr. Elaine Toldo Pazzelo que também contribuiu com meu trabalho.

Ao Prof. Dr. Edson Roberto Severnini que me acolheu como meu supervisor durante a realização do programa de doutorado sanduíche na Carnegie Mellon University.

Ao meu ex namorado Guilherme que esteve ao meu lado durante boa parte do doutorado e principalmente ao meu atual namorado Afonso que me acompanhou nessa reta final.

Aos meus colegas da Pós-graduação da FEA-RP\USP, em especial para Felipe, Fernanda, Raytza, Rafael, Paolo, Darcy e Denize que me marcaram positivamente em minha trajetória.

Também não poderia esquecer dos amigos que estiveram presentes durante esse período fornecendo amparo e suporte emocional: Rodrigo, Bruna, André, James, Lucas, Thiago e Ana Paula.

Por fim o presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

Título: EAD funciona? Uma análise do perfil, evasão e desempenho entre os estudantes brasileiros no ensino superior

Esta tese é composta por 3 artigos. No primeiro artigo nosso objetivo é identificar os principais determinantes da escolha da modalidade no ensino superior para estudantes brasileiros e comparar a incidência de políticas afirmativas entre as duas modalidades, usando modelo logit e dados do ENADE. Nossos resultados indicam que idade, trabalhar, ter concluído o ensino médio na rede pública, morar sozinho ou com cônjuge e/ou filhos em relação a morar com os pais e realizar o curso em universidade privada aumenta a probabilidade de estar matriculado no EAD e que a incidência de políticas afirmativas é maior nos programas presenciais. No segundo artigo estimamos o efeito da modalidade na taxa de evasão no ensino superior brasileiro, usando dados do Censo da Educação Superior com estimadores MQO. Os resultados indicam que não existe impacto da modalidade na taxa de evasão. O último artigo tem como objetivo estimar o efeito da modalidade na nota do ENADE, usando o método de pareamento por score de propensão e dados do ENADE. Os resultados indicam que os alunos do EAD possuem um desempenho pior no exame com relação aos estudantes do presencial.

Palavras-chave: EAD. Evasão. Desempenho. Ensino Superior.

Abstract

Title: Does distance education work? An analyses of profile, dropout and performance among Brazilians distance learners in higher education

This thesis is composed of 3 articles. In the first article our goal is to identify the main determinants of the choice of modality in higher education for Brazilian students and to compare the incidence of affirmative policies between the two modalities, using the logit model and data from ENADE. Our results indicate that age, labor, completing high school in public school, living alone or with a spouse and/or children compared to living with parents, and attend the course in a private university increases the probability of being enrolled to a distance learning course and that the incidence of affirmative policies is higher in face-to-face programs. In the second article we estimate the modality effect on the dropout rate in Brazilian higher education, using the data from Higher Education Census with OLS estimators. The results indicate that there is no impact in modality on the dropout rate. The last article aims to estimate the effect of modality on the ENADE score, using the propensity score matching method and data from ENADE. The results indicate that the distance learning students have a worse performance on the exam compared to the face-to-face students.

Keywords: Distance Learning. Dropout. Performance. Higher Education.

List of Figures

List of Tables

Contents

1 Concepts and Evidences about Distance Education

1.1 Introduction

In the last decade, there has been an expansion of higher education in distance learning courses in Brazil. According to Teixeira (2022), the proportion of those enrolled in distance learning jumped from 30.32% in 2011 (263,970 enrollments in distance education) to 77.53% in 2021 (1,255,086 enrollments in distance education). Proponents of the modality justify that the format allows the expansion of access to higher education, but for Brazil there is a lack of studies evaluating this modality as a good alternative to in-person courses.

Mainly after the pandemic of COVID-19, many systems of education and work are migrating to remote or hybrid format, and it is indispensable to analyse each case before taking a decision permanently. So to contribute with the literature, this thesis aims to evaluate the distance learning in higher education in Brazil and is composed by 3 papers. In the first paper our goal is to evaluate the profile of distance learners, establish the main determinants in the choice of the modality and also to analyse the incidence of quota and financing policies in distance education. For the second paper, we compare the drop out rate of the students in both modalities. And the last paper is about comparing the performance between students in online and face-to-face approach.

But before we start those analysis in the next chapters, in this chapter we intend: to explain what we will do in each paper and also to present a literature review to introduce the team, the concepts, definitions, to provide an overview about what was made until now, limitations, the gaps in the literature to fill and the appropriate manner to deal with the evaluation in distance learning.

In second chapter our goals are to identify the main determinants for modality choice and also evaluate the incidence of affirmative policies like quota, scholarship and financing programs between distance education and face-to-face approach. To achieve our goals we will apply logit model using data from ENADE from 2010 to 2019 and analysing the mean marginal effects. The undergraduate courses selected are: Pedagogy Degree, History Degree, Physical Education Degree, Business, Accounting, Social Work, Technology in Environmental Management, Physical Education Bachelor and Production Engineering.

In the next and third chapter of the thesis we will evaluate the effect of the modality in dropout rate in higher education institutions. For our purposes we will use data from higher education census and apply ordinary least squares in the years of 2018 and 2019. The courses analysed are: Pedagogy, History, Physical Education Degree, Business, Accounting, Social Work, Environmental Management, Physical Education Bachelor and Production Engineering.

Lastly, in the fourth chapter we intend to compare the performance in ENADES' score (a mandatory exam for higher education graduates to evaluate the quality of undergraduate courses in which the exam is composed by a general and specific component) between distance learners and in-person students using data from ENADE. We will analyse a total of 9 courses with higher number of distance students to capture the heterogeneity impact of the modality among undergraduate courses. The courses selected are: Pedagogy Degree, History Degree and Physical Education Degree in 2017; Business, Accounting and Social Work in 2018 and Technology in Environmental Management, Physical Education Bachelor and Production Engineering in 2019. As control for the selection bias we will use the propensity score matching, the nearest neighbor and inverse probability weighing algorithm, with replacement and for observations in common support and, measuring the average effect of treatment in the treated individuals (ATT). For the study we will also implement the balance and sensitivity analysis.

Besides that introduction, in this chapter in section 2 we bring the literature review on the topic of distance learning. And in the section 3 we have our final considerations and some conclusions about the debate in the literature in distance education.

1.2 Literature Review

In this section we intend to provide a theoretical basis for the proposed discussions, bringing the main contributions from the literature, national and international empirical evidence about the debate in distance education analysis. In this sense, this section serves as a guide to introduce the team, the conceptions, the definitions, to provide an overview about what was studied until now, the limitations, the gaps in the literature to fill and a start point to set the appropriate manner to deal with the evaluation in distance learning. More specifically, will be treated the following topics: Definitions and concepts; History of distance education; Brazilian legislation; Choice for distance learning and students profile; Impact of distance learning in drop out rate; Effect of distance education in students' performance.

1.2.1 Definitions and Concepts

There is an extensive literature discussing different definitions and concepts of distance education as presented in Simonson, Zvacek and Smaldino (2015a). But for our purposes we will focus in one of the definitions for the authors in which distance education can be seen as an institution based, formal education where the learning group is separated, and where interactive telecommunications systems are used to connect learners, resources, and instructors. But firstly there are some important points to establish:

- Distance education is by definition carried out through institutions; it was not self-study neither a nonacademic learning environment. The institutions might or might not offer traditional classroom-based instruction as well, but they were eligible for accreditation by the same agencies as those employing traditional methods.

-Geographic separation is inherent in distance learning, and time might also separate students and teachers. Accessibility and convenience were important advantages of this mode of education. Well-designed programs could also bridge intellectual, cultural, and social differences between students.

- Interactive telecommunications connected the learning group with each other and with the teacher. Most often, electronic communications, such as e-mail, are used, but traditional forms of communication, such as the postal system, might also play a role. Whatever the medium, interaction is essential to distance education, as it is to any education. The connections of learners, teachers, and instructional resources became less dependent on physical proximity as communications systems became more sophisticated and widely available; consequently, the Internet, cell phones, and e-mail had contributed to the rapid growth in distance education.

- Distance education, like any education, is established by learning group, sometimes called a learning community, which is composed of students, a teacher, and instructional resources like the books, sound, video, and graphic displays that allow the student to access the content of instruction.

The distance education also can be done synchronously, when the instructions are offered at the same time that the student learns, or asynchronously, which means that instruction is offered and students access it at separate times, or anytime it is convenient to them. It is also important to keep in mind another related terms to distance education as: E-learning, which usually refers to distance education in the private sector, what some also call e-training; Virtual Education/Virtual Schooling, which is used to refer to distance education in K-12 education and On-Line Learning/On-Line Education, which is the common distance education term used in higher education.

Lastly, as mentioned by Saba (2003), it is hard to imagine distance education without some elements of industrialization. To the extent that correspondence education relies on the mass production of instructional materials and involves a division of labor, it is an industrial enterprise. The author brings the definition of distance study as a rationalized method—involving the division of labor—of providing knowledge which, as a result of applying the principles of industrial organization as well as the extensive use of technology, thus facilitating the reproduction of objective teaching activity in any numbers, allows a large number of students to participate in university study simultaneously, regardless of their place of residence and occupation.

1.2.2 History of Distance Education

It is a stylized fact that the progress in distance education is intrinsically related with the development of technology and medias. Following this line of thinking Simonson, Zvacek and Smaldino (2015b) present the evolution of distance learning in four phases. The first and longest one with the courses by correspondence. The second with the use of the radio. The third with the emergence of television. And the last one and more recently with the availability of the computer and internet.

In the first phase with courses by correspondence, the authors describe some historic landmarks. For example, in 1833 an advertisement in a Swedish newspaper touted the opportunity to study "Composition through the medium of the Post.". In 1840, England's newly established penny post allowed Isaac Pitman to offer shorthand instruction via correspondence. And same type of courses started to spread across European countries and even across the Atlantic. From 1883 to 1891, academic degrees were authorized by the state of New York through the Chautauqua College of Liberal Arts. Correspondence study continued to develop in Britain with the founding of a number of correspondence institutions, such as Skerry's College in Edinburgh in 1878 and University Correspondence College in London in 1887. At the same time, the university extension movement in the United States and England promoted the correspondence method. Among the pioneers in the field were Illinois Wesleyan in 1877 and the University Extension Department of the University of Chicago in 1892.

In the beginning, according to to Moore and Aderson (2003), correspondence courses did not try to replace traditional higher education, actually they were aimed at nontraditional student populations who did not have access to higher education; and grew out of the university extension movement, not the university proper. In this sense, Pittman (2003) reinforce that the correspondence courses are not just the first method employed, but also for a long time was the only one. But after a while the method provided the funds and expertise that universities could to develop innovative and telecommunications schemes. With the expansion of distance learning programs, the International Council for Open and Distance Learning (ICDE) was created in a conference in 1938 (Bunker, 2003).

Thereby Europe experienced a steady expansion of distance education, without many changes, but with gradually more sophisticated methods and media employed like audio recordings and radio stations around 1920s. In the early 1930s, experimental television teaching programs were produced at the University of Iowa, Purdue University, and Kansas State College. However, it was not until the 1950s that college credit courses were offered via broadcast television: Western Reserve University was the first to offer a continuous series of such courses, beginning in 1951 (Simonson; Zvacek; Smaldino, 2015b).

In the third phase Simonson, Zvacek and Smaldino (2015b) highlights that over time the progress of satellite technology, developed in the 1960s and made cost effective in the 1980s, enabled the rapid spread of instructional television. Federally funded experiments in the United States and Canada, such as the Appalachian Education Satellite Project (1974−1975). The authors report that in the late 1980s and early 1990s, development of fiber-optic communication systems allowed for the expansion of live, two-way, high-quality audio and video systems in education.

Finally in the last and fourth phase, distance education opportunities have grown quickly through the use of computer mediated communications and the Internet. Courses have been offered over using the Internet since the mid-1980s, where the teachers organizes the course materials, readings, and assignments. The students read the material, view videos, listen to recordings, complete assignments, and participate in online discussions with other classmates. The advent of computer conferencing increases made possible interaction and collaborative work among the students (Simonson; Zvacek; Smaldino, 2015b).

Some of the reasons to the founding of distance teaching universities are discussed in Simonson, Zvacek and Smaldino (2015b) like: The need to increase the offerings of university education for adults; People at the job market or with family responsibilities; A large group of prospective part-time university students; A wish to serve both individuals and society by offering study opportunities to disadvantaged groups; The need found in many professions for further training at an advanced level; A wish to support educational innovation and the feasibility of an economical use of educational resources by distance education approach.

It is worth mentioning that the growth of distance learning was only possible because universities adopted the model over the 20th century and was expanded dramatically in United States, United Kingdom, Europe, Australia, New Zealand, and many other countries with the development of open universities, print-based materials, audiocassettes, and other learning resources. The increase of the modality enables learners to access university courses and degree programs at convenient times and schedules. Over the past two decades, with the rapidly growing populations and the search for higher education, many countries, specially those with more economic restrictions adopted the policy of creating universities with distance learning scheme (Hanna, 2003).

In Brazil, the emergency and development of distance education are very close to the rest of the world, and the evolution of distance learning was based on the progress and dissemination of the media as we can see in Saraiva (2008). We had the stage of teaching by correspondence, we went through radio and then television broadcasting and nowadays we use information technology up to the current processes of combined use of means -

telematics, multimedia and internet.

Initially the modality had a timid beginning in Brazil through newspaper advertisements in Rio de Janeiro, around 1900 according to Andrade and Lopes (2012). At that time the purpose was to offer professional courses by correspondence. In this sense, Litto and Formiga (2009) relates that the main goal was to expand the educational opportunities in some preparatory courses to job market, specially in basic education and the official benchmark is the installation of International Schools in 1904. Furthermore, Andrade and Lopes (2012) emphasizes that the distance learning approach suffered great resistance and prejudice, added to the idea that distance courses were aimed at disadvantaged masses, as a way to compensate them for a social disadvantage.

The first wave of distance education in Brazil and also the longest of all, as in the rest of the world, was with the development of correspondences courses. According to Andrade and Lopes (2012) there are the records date from the beginning of the 20th century and its disappearance occurred in the 90's. Posteriorly, with the emergence of radio, telephone and telegraph, the distance learning method had a great impulse in the country as stated by Mugnol (2009). Another historic landmark of the third wave of distance learning in Brazil, characterized by the use of television, was in 1974 when the Instituto Padre Reus appeared on TV Ceará begins the courses to Elementary 2, with television, printed material and monitors and in 1976 was created The National System of Teleducation, with courses through material instructional (Andrade; Lopes, 2012).

More recently, around 2000s until today we can see the use of computers, internet and other multimedia resources at the distance education approach that made possible the explosion in supply of professional, undergraduate and graduate courses replacing or completing the traditional method.

1.2.3 Brazilian regulatory

Regarding Brazilian legislation and regulation, we have 9057 Decree, of May 25, 2018, regulated art.80 of 9384 Law, of December 20, 1996, which establishes national education guidelines, presents the following definition for distance learning mode.

> Art.1 For the purposes of this Decree, distance education is considered the educational modality in which didactic-pedagogical mediation in teaching and learning processes occurs with the use of information and communication means and technologies, with qualified personnel, with policies access, with compatible monitoring and evaluation, among others, and develop educational activities by students and education professionals who are in different places and times.

The art.80 of 9394 Law, of December 20, 1996 (LDB/96) established the legal basis for distance education and provided improvements for the implementation of the modality.

> Art. 80. The Public Power will encourage the development and placement of distance learning programs, at all levels and modalities of education, and of continuing education.
>
> § 1 Distance education, organized with a special opening and regime, will be offered by institutions specifically accredited by the Union.
>
> § 2 The Federal Government will regulate the requirements for conducting exams and registering a diploma related to distance education courses.
>
> § 3 The rules for the production, control and evaluation of distance education programs and the authorization for their implementation, will fall to the respective education systems, with the possibility of cooperation and integration between the different systems.
>
> § 4 Distance education will enjoy differentiated treatment, which will include:
>
> I - reduced transmission costs in commercial sound and image and sound broadcasting channels and in other means of communication that are exploited with authorization, concession or permission from the public authority;
>
> II - concession of channels for exclusively educational purposes;
>
> III - reserve a minimum time, without burden to the Public Power, by the concessionaires of commercial channels.

Subsequently, the 10172 Law, of January 9, 2001, which approves the National Education Plan (PNE) and establishes the guidelines and goals of LDB/96, in which in PNE one of the objectives and goals is "To encourage, through resources public and private, the production of distance education programs that expand the possibilities of permanent professional education for the entire economically active population.".

1.2.4 Distance Learners' Profile

In this section we will discuss the theory behind the choice for distance education and empirical evidences about the distance learners profile. So this section is important to help us to conduct the analysis in the next chapter. Investigating the profile of these students makes it possible to improve the quality of courses offered, as it helps in the development of curricula and strategies that provide effective learning and it also contributes with information for the elaboration of public policies as mentioned in Lima, Borges and Souza (2018). Moreover, the understanding of this profile also can contribute in reducing the dropout rate of students in the distance learning modality, since the specific requirements of teaching become according to the specificities of this student, consequently decreasing demotivation and dropping out of courses.

In neoclassical microeconomic theory the choice for the distance education modality can be seen as a strategy of the agent of maximizing its usefulness. In that sense it is not surprising that the typical distance learners share broad demographic similarities as we can see in Latanich, Nonis and Hudson (2001). For example, according to Lima, Borges and Souza (2018) there is a predominance of the female audience among students of distance courses in part because this public has taken a large proportion of the labor market, and they have to reconcile their professional life with domestic activities and children. The authors argument that the vision of the role of women in society, she is expected to assume a role of multiple functions. According to them the incentives to enroll in new types of education is associated with less time spent commuting, thus facilitating the multiplicity of activities. The fact that they are more economically active and the market's need to attract more specialized professionals promote the process.

Besides, another key characteristic that distinguished the traditional distance learner from the non-distance learner was that they were place-bound, with a majority living far from the campus as in Latanich, Nonis and Hudson (2001). Distance education meets the demand not met by the offers of face-to-face courses, allowing the study of individuals in points away from large urban centers and capitals or population living in the countryside, thanks to the use of technological resources available on the internet, allowing the realization of all activities focused on learning as augmented in Ferlin, Alcântara and Paulo (2020).

It is a also relevant factor the age of the students, Latanich, Nonis and Hudson (2001) establish that distance learners are older than the typical undergraduate. One reason for this difference is that age is positively correlated with employment, marriage and family responsibilities. So, for these individuals, the ability to attend classes on campus is limited as mentioned by them. It is important mentioning that it can be expected that distance learners are more likely to be employed than non-distance learners, some of them need to keep working to pay the tuition and the method is more desirable because of schedule and place flexibility and it is a viable alternative to continue the studies, reconciling them with their professional occupations. Furthermore, Borges et al. (2016) highlights that the lower tuition fees have a strong influence on the teaching modality, because of bad financial conditions of some students.

Yet Latanich, Nonis and Hudson (2001) discuss about two of the most commonly investigated personal variables in a work or academic environment, the locus of control and motivation. Individuals who view themselves as substantially in control of events and circumstances that affect them are classified as possessing an internal locus of control (internals). Those who perceive events that influence their lives and well being as influenced largely by outside variables possess a more external locus of control (externals). To the author it seems logical to assume that those who enroll in these courses have a more self-directed orientation, and therefore exhibit a more internal locus of control. Regarding the degree of motivation, which refers to the ability to be self-starting or self directed. Students separated from their instructors need to be more self directed, so more motivated individuals are more likely to enroll in distance learning courses (Latanich; Nonis; Hudson, 2001).

After we have seen what the theory says about the distance learners profile, and the reasons why the students choose distance education approach, lets see some empirical evidences. Firstly, we have the study of Latanich, Nonis and Hudson (2001) for United States, which aims to compare key demographic and individual difference variables of distance and non-distance learners. The research concluded that both groups have significant differences, distance learners are older, there are more females, full-time workers and non-significant difference between the groups in locus control was found.

In the same vein, but for Brazil we have the paper of Spina (2013). The author describe the distance learners' profile as on average older, poorer, less white, mostly married, has children, comes more from public school, has parents with basic education, works and supports the family, has less access to the internet, uses less the computer, less knowledge of Spanish and English. He also concludes that the distance student profile reveals the strong inclusive character of distance higher education in the country. The simple confrontation between the socioeconomic profiles of distance learners with the students of face-to-face teaching of equivalent courses proves the social importance played by the institutions that implemented higher distance courses in the country. With the potential to break barriers, distance learning expands its educational targets in favor of an inclusive character and democratization of teaching, a broader characterization for the social importance of this modality.

Still for Brazil, Borges et al. (2016) found that the fact that classes take place only once a week and the lack of time to attend the on-site course were the influencing factors most pointed out by the students as determining factors for their choice and the fact that tuition is cheaper. Applying questionnaires Santos and Carneiro (2019) and for the undergraduate course of Geography at Unimontes, the authors verified the predominance of the female public and justify that with the need in search more space in the labor market, and as they most often perform a double shift, as mothers and housewives, it is difficult for women to have time available to travel to an educational institution. The distance education modality, therefore, facilitated access to higher education and the pursuit of knowledge. They also noted that the distance learners are older and argue that students who study virtually are adults who generally already have another occupation, and this kind of learning, which takes place anywhere and anytime, allows the student to continue working full-time while also paying attention to the family. The authors also found a greater presence of married individuals between distance learners and therefore, distance learning can be an excellent option for married people to study and graduate, without having a lot of time away from the family, as they can have access to the course classes they chose from their own residence. Finally, there was a greater composition of individuals who work between distance learners.

We can see from the papers analysed that distance and traditional learners are not homogeneous. There is a greater predominance in distance learning courses for women, older people, individuals who are married, have children, are in the labor force and live far from large urban centers or at rural zone. So distance education meets a demand from individuals who would not have access to higher education were it not for the possibility of remote classes, thus contributing to the democratization of higher education.

1.2.5 Higher Education Dropout

According to Bonaldo and Pereira (2016) higher education dropout is a problem that affects developed and in developing countries. The concern is because it has consequences for the individual and the society that finance those costs as stated by Lassibille and Gómez (2008). So, to the authors it is important to understand the main determinants for the withdraw to help in the reduction of the problem and maximize the allocation of the resources in education area and developing strategies to remain the students, because it is a issue with a great potential of improvement.

The main problem in this kind of analyses occurs because it is a complex topic with a lots of debate about the proper definition. For our purposes we will consider one of the definitions presented in SANTOS and Giraffa (2015), that fits with the definition of INEP (National Institute of Educational Studies and Research Anísio Teixeira), which is the situation where the student enrolled in some moment in a undergraduate program, but at one point in time he no longer enrolled.

Some discussions and evidences about the main determinants of withdraw can be seen in Lassibille and Gómez (2008) and Bonaldo and Pereira (2016). The papers discuss about the influence of age, family background, financing support and college quality in dropout rates.

In Lassibille and Gómez (2008), the paper analyses dropout rates among students in higher education in Spain using discrete time hazard models. About the results, the authors infer that: students that drop out are less motivated and skilled; the withdraw is greater in early years; age is positive correlated with dropout; students receiving financing support are less likely to give up because those programs reduce the cost of opportunity of education; students with parents with higher levels of education are less likely to finish college and students originally from another city are more likely to withdraw duo to spending time with travels and home activities.

In the same line with a research for the Brazilian case Bonaldo and Pereira (2016) use logit models to establish the main determinants of dropout rates in higher education. The work finds some evidence of age and change in marital status as positively affecting withdraw and college quality growing the rates of finishing college degree.

For dropout rate in higher distance education programs Levy (2007) and SANTOS and Giraffa (2015) have some considerations. According to SANTOS and Giraffa (2015) with the expansion of distance learning, consequently increases the challenges, so it is important to measure withdraw to develop permanence actions and retention actions. Lastly, in relation to Levy (2007), the author studied the influence of locus of control and motivation in dropout applying ANOVA and non-parametric tests. The author found no evidence about impact of locus of control in withdraw, a negative effect of motivation and college quality.

1.2.6 Effect of Distance Education in Performance

We propose in this section to discus methods and strategies applied by the literature focused in comparing the performance between distance learners and in person students. So our goal here is to reflect about differences and similarities of the approaches and analyse the appropriate manner to deal with this question. We also present results and conclusions about national and international empirical evidences.

According to Nguyen (2015) one reason why there is so much discussion around online learning are its purported benefits and uses like: Its effectiveness in educating students; Its use as professional development; Credit equivalency at the post secondary level; The possibility of providing a world class education to anyone with a broadband connection to anyone, anywhere, and anytime and its cost-effectiveness to combat the rising cost of post secondary education by spreading the cost of a class over a much larger number of students because the marginal cost of a student in an online setting is negligible relative to the traditional setting.

In this sense it is properly highlight the potential benefits of online learning as showed by Xu and Jaggars (2011) since online courses offer the flexibility of off-site asynchronous education and the potential of fully online coursework to improve and expand learning opportunities at community colleges, which educate large proportions of nontraditional students, like students with difficulties to attend on-campus courses because of employment or family commitments.

In the same line of reasoning Xu and Jaggars (2013) talks about the cost effectiveness of remotely education in relation to the traditional because online course sections are consistently less expensive and they yield fairly comparable student outcomes. But the authors also explain that studies of online course costs have not taken into account the quality or effectiveness of the courses examined, and it is possible that online courses with high completion rates and strong learning outcomes require substantial investments to design and teach.

Despite that, even with negative results in performance, they acknowledge that online learning is an important strategy to improve course access and flexibility in higher education, with benefits from both to the student and to the institutions. From the student perspective, the convenience of online learning is particularly valuable to adults with multiple responsibilities and highly scheduled lives; thus, online learning can be a opportunity to workforce development, helping adults to return to school and complete additional education that otherwise could not fit into their daily routines. From an institutional perspective, online modalities allow colleges to offer additional courses or course sections to their students, increasing student access to (and presumably progression through) required courses (Xu; Jaggars, 2013).

Furthermore, it is interesting to point out the main domains research in distance education effectiveness, as presented in Shachar and Neumann (2003): Student attitude and satisfaction regarding delivery of coursework; Interactions of students and faculty during delivery of coursework; Student outcomes in distance education coursework; Faculty satisfaction with delivery and coursework and other factors like student comfort, convenience, independence, and perceptions of effectiveness.

Although the goal here is about the outcomes, it is important reinforce the importance of other parameters to measure the quality of online courses. Therefore we need establish the issues about the objective measurements of the academic performance factors and the operationalization of these measurements. So the authors quote that it can be based on course grades, tests, and exams. However, they remember that course grades may carry some assessor subjectivity. The article also exposes the subjective measurements like attitudes, satisfaction, and evaluation of instruction factors. But as mentioned by Ni (2013) it is good to have in mind that as learning effectiveness is a complex concept with multiple dimensions and it should be assessed with multiple measures.

Moreover, Ni (2013) argues that the social communicative interactions between students and teacher and student and student are fundamental for the learning process. A student's ability to ask a question, to share an opinion, or to disagree with a point of view are important learning activities. The author claims that it is often through conversation, discourse, discussion, and debate among students and between instructors and students that a new concept is clarified, an old assumption is challenged, a skill is practiced, an original idea is formed and encouraged, and ultimately, a learning objective is achieved. The article explains that online learning requires adjustments by instructors as well as students for successful interactions to occur. Online courses try to substitute classroom interaction with discussion boards, synchronous chat, electronic bulletin boards, and e-mails. But the effectiveness of such a virtual interactive venue is not without debate. However, the advantage of online interaction may not be realized if close connection among the learners is absent (Ni, 2013).

Yet, achieving apprenticeship require a greater responsibility from the students according to Xu and Jaggars (2014), thus, a successful online student may need high levels of self-regulation, self discipline, and a related suite of meta cognitive skills, which often fall under the broad rubric of self-directed learning. But the article also discuss the obstacles that distance learners have to overcome because students in online courses may experience higher levels of dissatisfaction, interpersonal isolation, feelings of unclear direction and uncertainty, and a lack of engagement in the learning process. It is also good to know that online course is not explicitly designed to help students develop the skills they need to succeed in this new context. Accordingly, some students may struggle to perform as well in an online course as they would in a similar face-to-face course (Xu; Jaggars, 2014).

In this sense, Harmon, Lambrinos et al. (2012) explore the disadvantages of online learning like isolation and lack of media richness. Isolation encompasses insufficient peer and instructor interaction, which leads to low engagement and lower course completion rates. They suggest that media richness encompasses the attributes of face-to-face communication of verbal speech intonations and nonverbal facial and body expressions.

On the other hand, we also have to look at the advantages of the online format like a self-paced format, convenience in viewing and re-viewing lectures, and savings from reduced commuting time and expense as mentioned by Harmon, Lambrinos et al. (2012). Besides online classes can provide faster information access, it can be structured to accommodate different learning styles and student types, and it encourages students to take a more responsible, constructive role in the learning process as can be seen in Gratton-Lavoie and Stanley (2009).

So if the online classroom is at least as effective as the traditional classroom, education delivery costs per student and per concept taught could be lowered, therefore lowering the total costs of education (Gratton-Lavoie; Stanley, 2009). But Caetano et al. (2016) also explains that the evaluation of students' academic progress is very important, and should be part of public policies in understanding the processes involved in education and not only for the establishment of rankings.

Considering the advantages and disadvantages of online format, Harmon, Lambrinos et al. (2012) suggest maybe it is a good idea mixing online lectures with traditional lectures in the same course to promote equivalence between "digital"and "live"communication. Thus, a rotating format, the traditional lectures could potentially marry the advantages of the two modalities.

Once we have established the theory issues about distance learning, it is time to see some international and Brazilian empirical evidences and different approaches about analyzing the difference in performance between distance education and traditional face to face method. First we have two studies using a meta-analysis in Shachar and Neumann (2010), the authors compare the performance of students in the distance learning and face to face models and Means et al. (2013) included blended course in the analyses. According Shachar and Neumann (2010) it is perfectly possible to compare the two modalities and they noted that in 70% of the cases, distance learning students outperformed their peers in the face-to-face model. For Means et al. (2013) they also concluded that on average, students in online learning conditions performed modestly better than those receiving face-to-face instruction. But in this case, the advantage over face-to-face classes was significant just in those studies contrasting blended learning with traditional face-to-face instruction and not in the studies contrasting purely online with face-to-face conditions. According to the authors the advantage of online conditions in these recent studies stems from aspects of the treatment conditions other than the use of the Internet for delivery per se.

The research conducted by Harmon and Lambrinos (2006), Xu and Jaggars (2011), Gratton-Lavoie and Stanley (2009) and Bowen et al. (2014) aimed to compare both instructional modes, but just for one subject and not the entire undergraduate course and had different results. In Harmon and Lambrinos (2006), they found that test scores for the online format, when corrected for sample selection bias, are four points higher than for the live format, and the difference is statistically significant. One possible explanation to the authors is that there was slightly higher human capital in the classes that had the online format. A Oaxaca decomposition of this difference in grades was conducted to see how much was due to human capital and how much was due to the differences in the rates of return to human capital. The Oaxaca decomposition showed that 75% of the gap in test scores is explained by differences in returns to human capital endowments. An interpretation for them is that this 75% is explained by returns related to unobserved factors. That is it might be that online students save on commuting time and when they sit down to take the online class it is a time when they are rested, are in a distraction free environment. Whereas, perhaps students in the in-person format are distracted by fellow classmates, and are squeezing the class time in between other activities in a busy schedule after a relatively long commute.

However in Gratton-Lavoie and Stanley (2009) after controlling for selections bias the online teaching mode had a narrowly insignificant, or even negative, effect. For Xu and Jaggars (2011) with several empirical strategies used to minimize the effects of student self selection, including multilevel propensity score. The findings indicate a robust negative impact of online course taking for both subjects. Although proponents of online learning have consistently noted that it is not course modality but course quality that influences student learning, and the authors agree with this assertion, but the results suggest that designing and teaching high-quality online courses with sufficient student supports may be more difficult than doing so with face-to-face courses.

In Bowen et al. (2014), applying experiment, ols and probit approaches, by randomly assigning the students on six public university campuses to take the course in a hybrid format or a traditional format. The authors found that learning outcomes are essentially the same—that students in the hybrid format are not harmed by this mode of instruction in terms of pass rates, final exam scores, and performance on a standardized assessment of statistical literacy, so they conclude that adopting hybrid models of instruction in large introductory courses has the potential to significantly reduce instructor compensation costs in the long run.

Another interesting study from Deschacht and Goeman (2015) using natural experiments to measure the effect of blended learning on adult learners' academic success. They tested the impact of introducing a blended learning format within the first year in persistence and performance, using difference-in-difference research design that minimizes the potential bias resulting from the selectivity of learners enrolled in blended programs. The authors find out that blended learning improves exam results, although they observe a negative effect on the course persistence of adult learners.

In Xu and Jaggars (2013), the authors use instrumental variable technique to estimate the impact of online versus face-to-face course delivery on student course performance controlling for the self selection. They used the travel distance between each student's home and college campus served as an instrument for the likelihood of enrolling in an online section of a given course. In addition, college-by-course fixed effects controlled for within- and between-course selection bias. Analyses yield robust negative estimates for online learning in terms of both course persistence and course grade.

Focused in examining the performance gap between online and face-to-face courses but also how the size of that gap differs across student subgroups and academic subject areas, Xu and Jaggars (2014). While all types of students in the study suffered decrements in performance in online courses, those with the strongest declines were males, younger, black and students with lower grade point averages. Online performance gaps were also wider in some academic subject areas than others. After controlling for individual and peer effects, the social sciences and the applied professions (business, law, and nursing) showed the strongest online performance gaps.

For Brazil, we have the study of Nascimento, Czykiel and Figueiró (2013) which the goal is verifying the influence of distance education on student learning, an experiment was carried out with two classes of a single subject. Using as parameters of performance individual and group tasks, exam and class participation, they observed that both classes had similar results but in most of the evaluated items, the students of the distance class presented better performance. According to them despite the prejudice with distance learning approach, it can be an option with good results.

In sequence BATISTA et al. (2014) aimed to evaluate the level of student performance by distance learners and face-to-face. The results show a better performance in online students, mainly in private universities and in most states. For Caetano et al. (2016) applying OLS and Mann-Whitney non parametric techniques to compare the performance of students in both format. But in this case the authors verified that the performance of students in distance education approach is significantly lower, and according to them these results indicates the need to monitor the performance of distance learning students, given the quantitative expansion it has been achieving in Brazil and factors such as access to the library, teachers and colleagues can be an explanation for the difference in performance that has occurred.

Another Brazilian study was conducted by Oliveira and Piconez (2017). From the results, the distance learning courses do the least to avoid undergoing on-site assessment, and these numbers may be a consequence of the resistance of institutions of excellence in offering distance learning courses. Yet for Brazil, MARQUES (2020) adopted propensity score techniques for controlling the selection bias problem. The author seeks to assess the impacts of distance higher education and the results showed that it is possible to rule out the possibility of positive academic impacts, considering negative or null results observed.

From the articles analysed it is possible to see that there is no consensus about what approach is better. There are results in favor of distance education, some results show that there is no difference and some others observed that distance learning decrease the students' performance. But we can highlight some facts about the papers. First, the studies that compare the methods for one single subject may be good to decide if a particular subject is better being taught by distance or in person, but it is complicated to generalize this conclusion over a full college degree. Secondly, the hybrid format showed positive or null results compared with face to face format, so the hybrid approach could be a good alternative to take advantages of both methods like Harmon, Lambrinos et al. (2012) stated. Finally, the papers which compare a full college degree in both methods and take into account the problem of selection bias observed a negative impact on performance in distance learning modality.

1.3 Final Considerations

As we saw we had a huge growth in distance learning in higher education in Brazil in the last decade, but for Brazil there is a lack of studies evaluating this modality as a good alternative to in person courses. Mainly after the pandemic of COVID-19, many systems of education and work are migrating to remote or hybrid format, and it is indispensable to analyse each case before taking a permanent decision. So to contribute with the literature, this thesis aims to evaluate the distance learning in higher education in Brazil and is composed by 3 papers. In the first paper our goal is to evaluate the profile of distance learners and establish the main determinants in the choice of the modality. For the second paper we compare the drop out rate of the students in both modalities. And in the last paper is about comparing the performance between students in online and face-to-face approach.

But before we start those analysis in the next chapters, in this chapter we intended: to explain what we will do in each paper and also to present a literature review to introduce the team, the concepts, definitions, to provide an overview about what was made until now, limitations, the gaps in the literature to fill and the appropriate manner to deal with the evaluation in distance learning.

So in the section of literature review we provided a theoretical basis for the proposed discussions, bringing the main contributions from the literature, national and international empirical evidence about the debate in distance education analysis. In this sense, it is important to serve as a guide to introduce the team, the conceptions and the definitions and also to provide an overview about what was studied until now, the limitations, the gaps in the literature to fill and a start point to set the appropriate manner to deal with the evaluation in distance learning. More specifically, we discussed the following topics: Definitions and concepts; History of distance education; Brazilian legislation; Choice for distance learning and students profile; Impact of distance learning in drop out rate; Effect of distance education in students' performance. For the next chapters we will evaluate distance learner's profile; effect of modality in dropout rate and compare the performance between distance learners and in-person students.

2 An investigation in the profile, choice for modality and affirmative policies in higher education between distance learners and in-person students

2.1 Introduction

One of the most import steps of evaluating, implementing and improving a policy is to investigate the profile of the target audience, which is also the case for distance education. So, as stated by Lima, Borges and Souza (2018) the studies about the main characteristics of distance learners makes it possible to improve the quality of courses offered, as it helps in the development of curricula and strategies that provide effective learning and it also contributes with information for the elaboration of public policies. Moreover, the understanding of this profile also can contribute in reducing the dropout rate of students in the distance learning modality, since the specific requirements of teaching, according to the specifies of this student, consequently increases motivation and decreases the dropping out of the courses.

In neoclassical microeconomic theory the choice for the distance education modality can be seen as a strategy of the agent of maximizing its usefulness. In that sense it is not surprising that the typical distance learners share broad demographic similarities as we can see in Latanich, Nonis and Hudson (2001). For example, Lima, Borges and Souza (2018) discuss the gender aspect and the predominance of the female audience among students of distance courses. According to them as they are expected to assume a role of multiple functions and reconcile their professional life with domestic activities and children the incentives to enroll in new types of education is associated with less time spent commuting.

Another key factor is the distance from campus. Distance education meets the demand not met by the offers of face-to-face courses, allowing the study of individuals in places away from large urban centers and capitals or population living in the countryside as augmented in Ferlin, Alcântara and Paulo (2020). It is also relevant mentioning the age, in this sense Latanich, Nonis and Hudson (2001) establish that distance learners are older. To the authors the reason for this difference is that age is positively correlated with employment, marriage and family responsibilities. Distance learners are more likely to be employed than non-distance learners, some of them need to keep working to pay the tuition and the method is more desirable because of schedule and place flexibility and it is a viable alternative for continuing the studies, reconciling them with their professional occupations. Furthermore, Borges et al. (2016) highlights that the lower tuition fees have a strong influence on the teaching modality, because of some students' lack of financial conditions.

Specifically for Brazil we have these study from Spina (2013). The author describes the distance learners' profile as on average older, poorer, less white, mostly married, has children, comes more from public school, has parents with basic education, works and supports the family, has less access to the internet, uses less the computer, less knowledge of Spanish and English. He also concludes that the distance student profile reveals the strong inclusive character of distance higher education in the country. The simple confrontation between the socioeconomic profiles of distance learners with the students of face-to-face teaching of equivalent courses proves the social importance played by the institutions that implemented higher distance courses in the country. With the potential to break barriers, distance learning expands its educational targets in favor of an inclusive character and democratization of teaching, a broader characterization for the social importance of this modality.

Complementing this point of view Santos and Carneiro (2019) noted that the distance learners are older and argues that students who study virtually are adults who generally already have another occupation, and this kind of learning, which takes place anywhere and anytime, allows the student to continue working full-time while also paying attention to the family. The authors also found a greater presence of married individuals between distance learners and therefore, distance learning can be an excellent option for married people to study and graduate, without having a lot of time away from the family, as they can have access to the course classes they chose from their own residence.

We can see from the papers analysed that distance and traditional learners are not homogeneous. There is a greater predominance in distance learning courses of women, older people, individuals who are married, have children, are in the labor force and live far from large urban centers or at rural zone. So, distance education meets a demand from individuals who would not have access to higher education were it not for the possibility of remote classes, thus contributing to the democratization of higher education.

Despite these stylized facts, there still are some questions without answer or at least with insufficient discussion, such as: How has the distance learners profile changed over the years?; Are distance education programs meeting the expected public?; What about the difference in both modalities in terms of affirmative policies like quota, financing and scholarship programs? So in this study we intend to fill some gaps in the literature not just analyzing the profile of Brazilian undergraduate students over the last decade, but also apply logit estimations, using data from ENADE, to identify the main determinants of this choice, if the modality is really meeting the demand it is expected to, and the incidence of social policies between both modalities.

Particularly about quota in Brazil, racial and for students from public school, quota policies were implemented with the aim of promoting equal opportunities in accessing higher education. These policies were created in response to the historical and structural inequality that exists in the country, which makes it difficult for vulnerable groups to access education and the job market.

The racial quota policy reserves a percentage of spots in public universities for self-declared black, mixed-race, and indigenous students (PPI). The policy for public school students reserves a percentage of spots for students who completed their entire high school education in public schools.

About the financing programs, maybe the most relevant one is FIES (Student Financing Fund) a Brazilian government program that offers student-financing for low-income students who wish to pursue higher education at private institutions. The program allows students to finance their studies at low interest rates and start paying off the debt after completing the course.

Finally, in relation to scholarship programs, it is important to mention PROUNI (University for All Program), which is a Brazilian government program that offers full and partial scholarships to low-income students in private universities. The program aims to promote social inclusion and democratization of access to higher education.

Considering our results we can identify a pattern in the direction of the relation of the variables in the choice between distance education and in-person programs, but in heterogeneous levels of effect depending on undergraduate course. Its important to highlight the variables with the strongest relations with the choice for modality like: age, having a job, being from public high school, living alone or with spouse and/or child in comparison to the students living with their parents and being from privates universities have a greater probability of being enrolled in distance learning programs. Besides, in relation to the affirmative policies, like quota, scholarship and financing programs we verified a larger incidence in face-to-face modality.

So with this study, we intend to contribute with the literature: identify the main determinants of the choice for modality, the dimension of the impact and the heterogeneous effects among undergraduate courses in higher education in Brazil in the last decade; evaluate the incidence of affirmative policies like quota, scholarship and financing programs between both modalities and also to contribute with strategic policy decision-making in the sense of to improve the distance education approach taking into account the target audience.

Besides this introduction section, this chapter is also composed by section 2 with information about the data and methodology. In section 3 we bring the results of our models' estimation and in section 4 we have our final considerations.

2.2 Data and Methods

The purpose of this section is to bring information about strategies and methods used to achieve the main goal in evaluating the difference between the profile of undergraduate students and the incidence of affirmative policies. So the section includes the presentation of the database, description of the econometric techniques, definition of the model variables, descriptive analysis of the data and exposure of the models and methods applied.

2.2.1 Database and variables

Here we will present information about the database and variables used. It is intended, therefore, to describe the strategies used for the database treatment and to discuss the choice of the dependent and control variables. The database is from ENADE, provided by INEP (National Institute of Educational Studies and Research Anísio Teixeira) and contains data related to the exam results, the Educational Institution and the profile of the students. ENADE is a mandatory exam that evaluates the performance of students in higher education. The exam contains questions about the syllabus, skills and competences acquired in their training. The test has a three-year assessment periodicity for each area of knowledge. In addition, there are also data about information regarding the General Course Index (IGC) that are added to the ENADE database, where such data are also provided by INEP.

To analyze the profile of the students in both groups, we selected the 3 courses with the highest proportion of distance learners in 2017, 2018 and 2019 respectively. Thus, in 2017 we have Pedagogy Degree (which concentrates 60.35% of all students in distance education), History Degree (4.83%) and Physical Education Degree (4.67%), in 2018 we have Business (30.08%), Accounting (16.12%) and Social Work (12.54%), and in 2019 we have Physical Education Bachelor (49.36%), Technology in Environmental Management (15.11%) and Production Engineering (9.53%) . Once the courses for each year are defined, 18 models are estimated, two for each course, using the method of logistic regression in which the binary dependent variable is the modality of the undergraduate course (the variable distance takes the value of 1 if the student is enrolled in distance modality and 0 otherwise). The first sequence of 9 models are to evaluate the main characteristics of students profile and also the quota policies. The second group of 9 models are to analyse the difference between modalities in financing programs for private universities.

The control variables are shown in Table 1. The variables indicate socioeconomic characteristics of students, attributes of educational institutions and information about funding and entry policy of students.

Table 1 – Covariates

Variables	**Description**
Age	Individual's age
Sex	Dummy equals to 1 if the individual is a man.
Race	Dummy equals to 1 if the individual is white or asian. Black, brown and indigenous otherwise.
Per Capita Income	Individual's per capita income. The per capita income variable was created by dividing the midpoint of the income ranges (up to 1.5 minimum wages, between 1.5 and 3, between 3 and 4.5, between 4.5 and 6, between 6 and 10, between 10 and 30 and over 30) in which the average point used for the category over 30 was 30 minimum wages divided by the number of family members in which the categories of members in the household are 1 individual, 2, 3 , 4, 5, 6, 7 or more, in category 7 or more the value of 7 was considered. And the values are updated to the year of 2019 through the IPCA calculator
Work	The job status, in which the categories are no job, part time job and full time job.
Public School	The Dummy equals to 1 for individuals who studied at least most part of High School in public school. The individuals who studied at least the most part of the the High School abroad were excluded.
Parent's Education	Variable that assumes the level of education of the parent with the highest schooling, in which the categories are no level completed, Elementary School, Middle School, High School, College Degree or more.
Home	Situation in home, in which the categories are alone, with parents and/or relatives, with children and/or spouse and others.
State	Country's state.
Hours	Hours dedicated to studies except class hours. In which the original categories are none, between 1 and 3, between 4 and 7, between 8 and 12, more than 12. For our case we adapted for the midpoint except for the first and last categories, so the final categories are 0, 1.5, 5.5, 10 and 12.
Private University	Dummy equals to 1 if the individual studies the higher education in private university.
Scholarship	Dummy equals to 1 if the student receive a partial or full scholarship to pay the tuition in private institutions and the student does not have to pay it back. Students with partial PROUNI and FIES were excluded
Financing	Dummy equals to 1 if the student receive a financing to pay the tuition in private institutions and the student have to pay it back. Students with partial PROUNI and FIES were excluded
Quota	Dummy equals to 1 if the individuals admission was through quota.
IGC	University Score

2.2.2 Descriptive analysis

This section presents a descriptive analysis of the data of the dependent variables and its relation with the other covariates in order to contextualize and understand the profile of distance learning students compared to the traditional face-to-face students.

The Table 2 shows the total number of observations, universities and courses by modality and undergraduate course. We have a total of 1,486,287 students in all undergraduate courses analyzed, of which 1,022,092 are in traditional face-to-face format and 464,195 are in distance learning approach. In total there are 1,965 universities, in which 1,960 offers at least one of the courses in the face-to-face modality and 153 offers at least one of the courses in the distance-learning modality. In addition, we have a total of 8,555 courses offered, of which 8,125 are face-to-face and 430 in distance learning. Finally, it is worth remembering that these are all observations from the database. In the rest of the descriptive analysis, as well as in the regression models, there may be fewer observations due to missing variables.

Table 2 – Number of observations, Universities and courses by modality and undergraduate course.

	Observations			**Universities**			**Courses**		
	FTF	**Dist**	**Tot**	**FTF**	**Dist**	**Tot**	**FTF**	**Dist**	**Tot**
Pedagogy	185,095	187,575	372,670	1,030	109	1,040	1,564	128	1,692
History	40,890	14,098	54,988	298	41	311	478	43	521
Phy. Ed. DE	82,279	7,724	90,003	489	16	490	651	17	668
Business	358,743	89,913	448,656	1,552	83	1,556	2,314	85	2,399
Accounting	146,778	38,428	185,206	977	51	982	1,225	51	1,276
Social Work	71,575	91,729	163,304	374	28	382	430	28	458
Phy. Ed. BA	65,328	13,683	79,011	440	11	443	525	11	536
Tech. Env. Man	20,736	18,312	39,048	226	44	248	274	44	318
Prod. Eng.	50,668	2,733	53,401	527	23	530	664	23	687

Source: Made from ENADE 2010 to ENADE 2019 microdata, INEP.

In Figure 1 we can see the number of students students by year, undergraduate course and modality. In Figure 1 we can observe that in general the number of in-person students are higher than the number of distance learners, exceptions for specific years and for some undergraduate courses as Pedagogy Degree, Social Work and Technology in Environmental Management. Another consideration, is that generally speaking the quantity of distance learners grew over time, except for Social Work and Technology in Environmental Management, that had a fall trajectory, and Business stayed stable. We can highlight the case of Physical Education Bachelor, in which the growth was from 0 students in 2013 to 11,544 in 2019.

On the other hand, in face-to-face modality we have more cases of undergraduate courses in fall or stable trajectory, like Technology in Environmental Management, Social Work and Business falling and History Degree and Accounting stable. Business is the undergraduate in-person course which most lost students, from 145,591 students in 2012 to 91,253 students in 2018.

Another interesting point, is the case of Technology in Environmental Management which we had a reversal in the position of the curves. In 2010 the number of students enrolled at in face-to-face modality was 5,020 students and fell to 1,571 students in 2019, to distance learning modality the numbers are 4,095 and 3,533 respectively. So as we had more cases in which the number of distance learners increased, and more cases in which in-person students decreased, as a rule we can conclude that the proportion of distance learners expanded in relation to in-person students

Figure 2 exposes data on the number of students by year, undergraduate course, modality and type of university. According to data in Figure 2 we can conclude that the number of distance learners at private Universities grew up in general, specially for Physical Education Degree, Production Engineering and History, and except for Social Work and Technology in Environmental Management that dropped, and Business that stayed stable. For example, in the case of Physical Education Degree, the number of distance learners in private universities grew up from 135 in 2011 to 5,214 in 2017. At the same time we can not observe a standard for undergraduate courses at public Universities.

For in-person undergraduate courses at private universities we have more examples of reduction of the number of students, specially Technology in Environment Management, Social Work and Business. The case of Technology in Environment Management, the number of in-person students at private universities dropped from 9,596 in 2010 to 317 in 2019. On the other hand, regarding the number of in-person students in undergraduate courses at public universities we can not detect any pattern, in some undergraduate courses the number of students grew up (Pedagogy, Physical Education Degree and Bachelor and Production Engineering), some remained stable (History, Business and Accounting) and some decreased (Social Work and Technology in Environmental Management).

It is also worth mentioning that the number of in-person students at private universities are in general higher than distance learners, except for Pedagogy and History which the number of distance learners are higher. In the case of Pedagogy the number of distance learners at private universities was 63,790 in 2019 and the number of in-person students was 40,850. For History the number of in-person students at private universities was higher than distance learners in 2011, but in 2017 the number of both groups were very similar between 4,500 and 5,000 students. And in the case of Technology in Environmental Management, the number of distance learners at private universities was lower in 2010, but higher in 2019. About public universities we can see that the number of in-person students are always higher than distance learners, excepts for Social Work and before 2016.

In Figures 3, 4 and 5 are showed the distribution of the students by age, modality and undergraduate courses between 2010 and 2019. We can see in those Figures that the

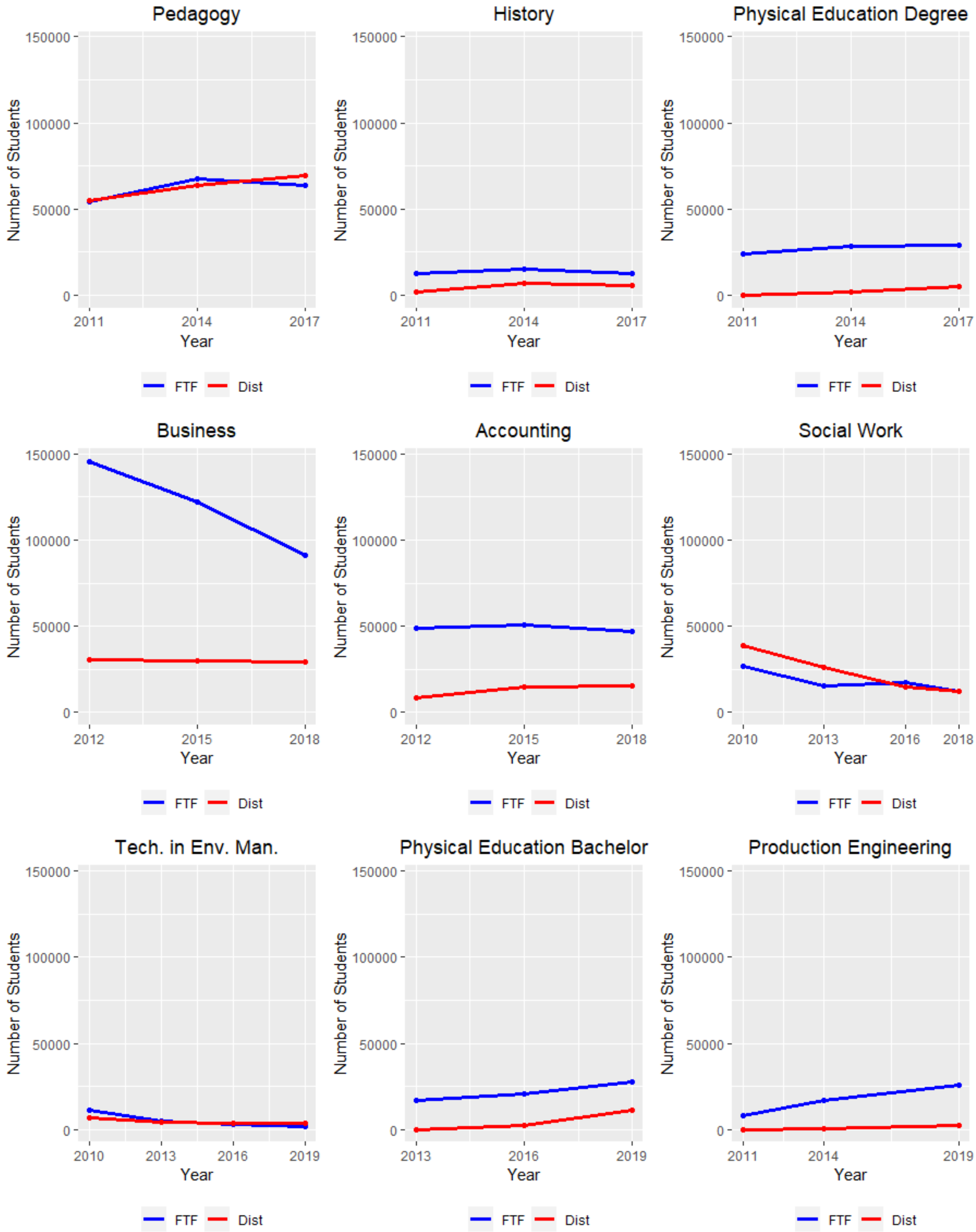

Figure 1 – Number of students by year, undergraduate course and modality.

Source: Made from ENADE 2010 to 2019 microdata, INEP.

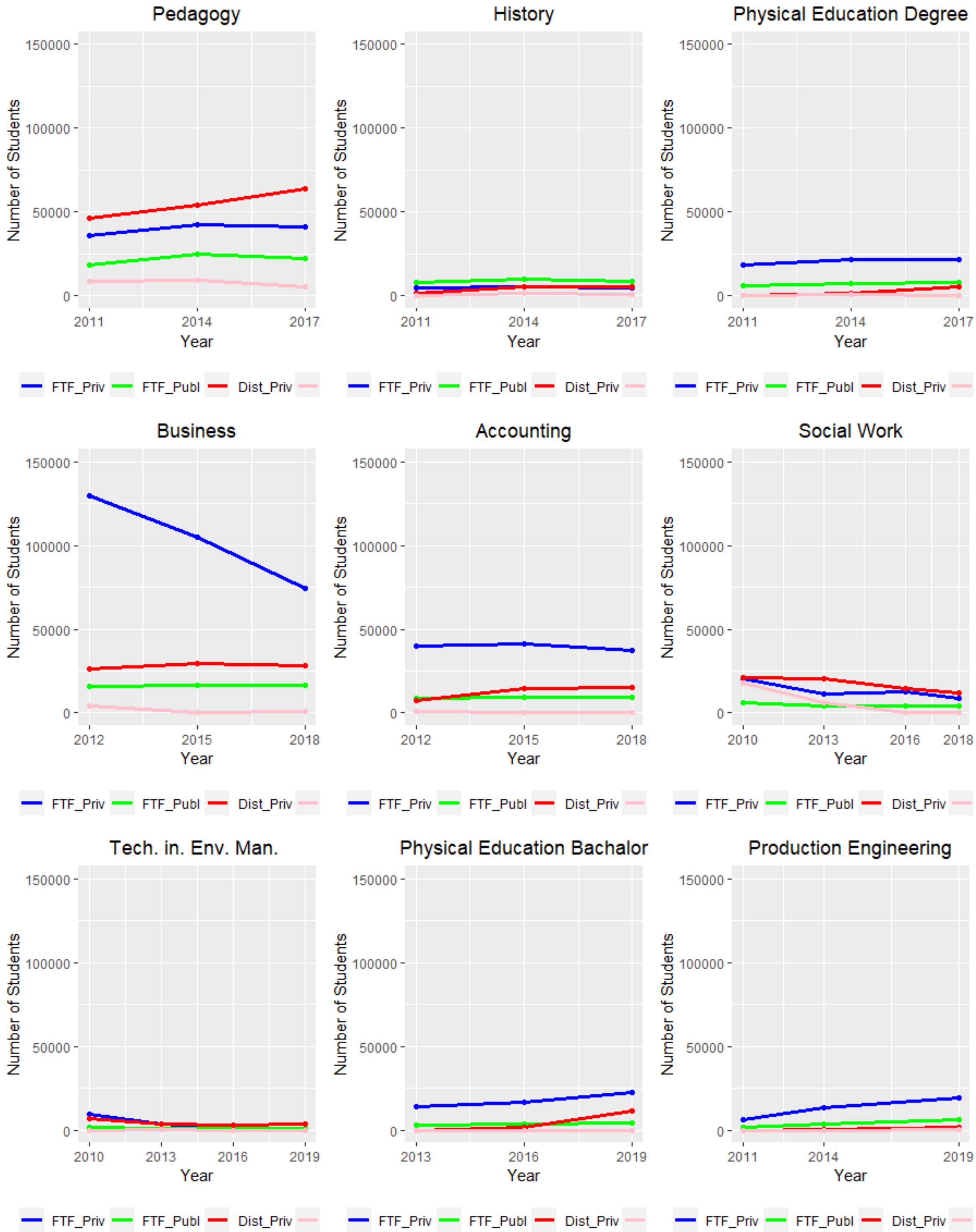

Figure 2 – Number of Students by year, undergraduate course, modality, and type of university.

Source: Made from ENADE 2010 to and 2019 microdata, INEP.

distributions for each undergraduate course did not change very much over time. Usually for the in person undergraduate courses there is a higher concentration between 20 and 30 years old, except in Technology in Environmental Management that has more dispersed distribution. About the distance undergraduates the distribution is more uniform between 20 and 60 years old. So we can conclude about the observation that the profile of distance learners are older than the in-person students.

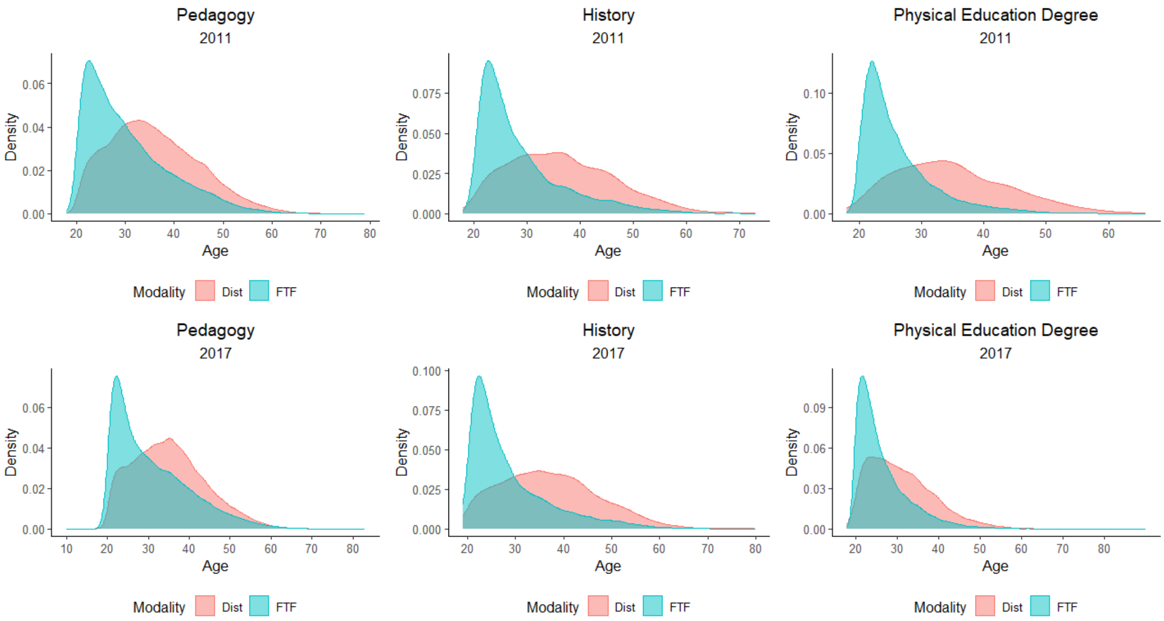

Figure 3 – Distribution of students by age, modality and undergraduate courses.

Source: Made from ENADE 2011 and 2017 microdata, INEP.

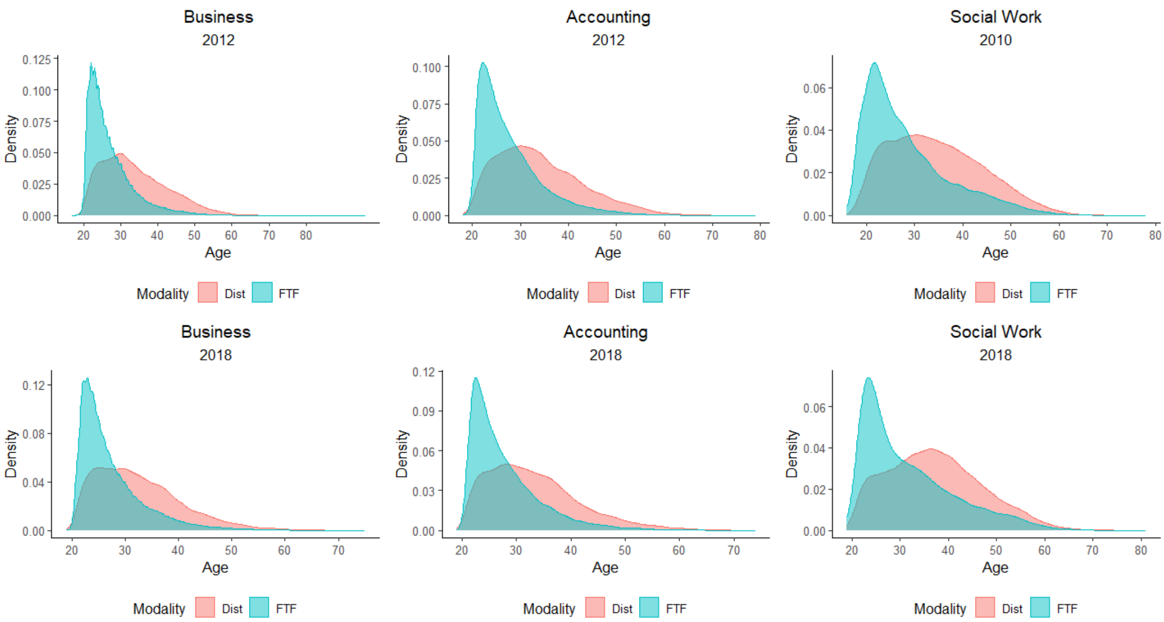

Figure 4 – Distribution of students by age, modality and undergraduate courses.

Source: Made from ENADE 2011, 2012 and 2018 microdata, INEP.

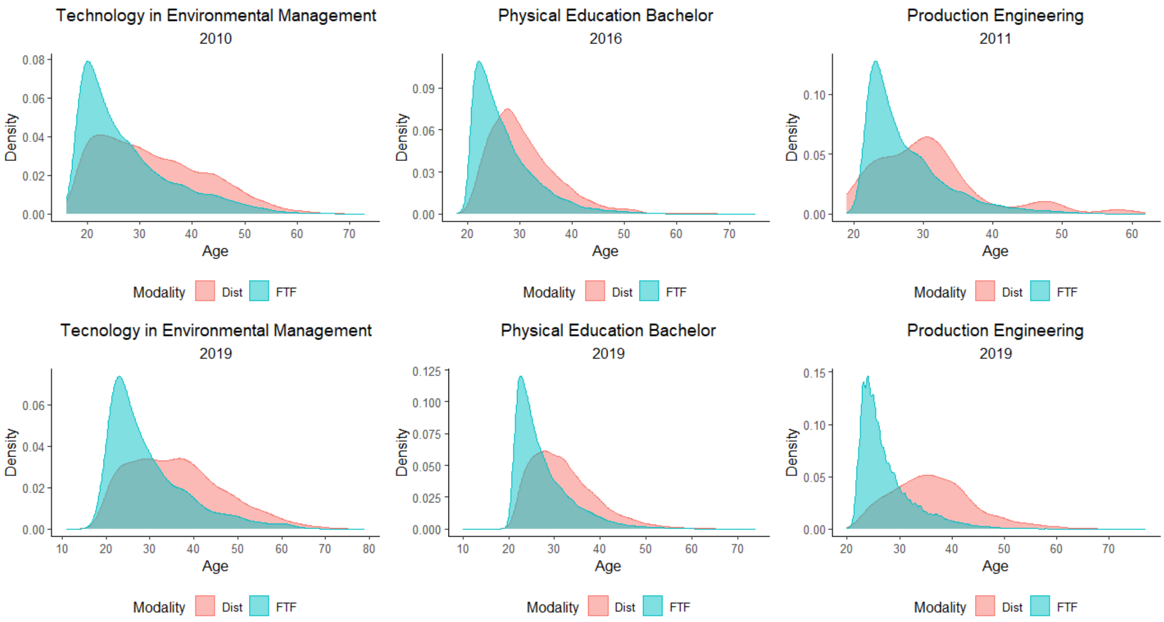

Figure 5 – Distribution of students by age, modality and undergraduate courses.
Source: Made from ENADE 2010, 2011, 2016 and 2019 microdata, INEP.

Furthermore, in Figure 6 is presented the distribution of students by gender, race, year, modality and undergraduate course. It is important to highlight that in white race are grouped white and asians, and for black race are grouped blacks, browns and indigenous. In general, we can note in Figure 6 that the proportion by gender and race vary according to undergraduate curse and year but not much with modality, with some exceptions depending on the year and undergraduate course.

Figure 7 presents the concentration students in labor market by year, undergraduate course and modality. We can see that according to the Figure 7 the proportion of workers are higher for all undergraduate courses and in both years among distance learners. In general the gap between years did not change so much from the first year of analyses to the second one. The exception is Physical Education Degree, in which the gap decreased from 28 p.p. in 2011 to 17 p.p. in 2017. On the other hand, for Technology in Environmental Management and Production Engineering the gap increased. In the case of Technology in Environmental Degree, the gap was 14pp in 2010 and 27 p.p. in 2019.

The biggest gap occurs in Social Work, in 2010 78% of distance learners were in labor Market and 53% of in person students, and in 2018 those numbers are 70% and 49% respectively. And the smallest gaps are for Accounting and Business. For example, for Accounting, in 2012 92% of distance learners were in labor market and 88% of in-person students and in 2018 those values were 86% and 79% respectively.

Besides, Figure 8 shows the graduates concentration who attended high school in the public system by year, modality and undergraduate course. In general, it is noted that there is a slightly higher proportion of students who came from public schools in distance

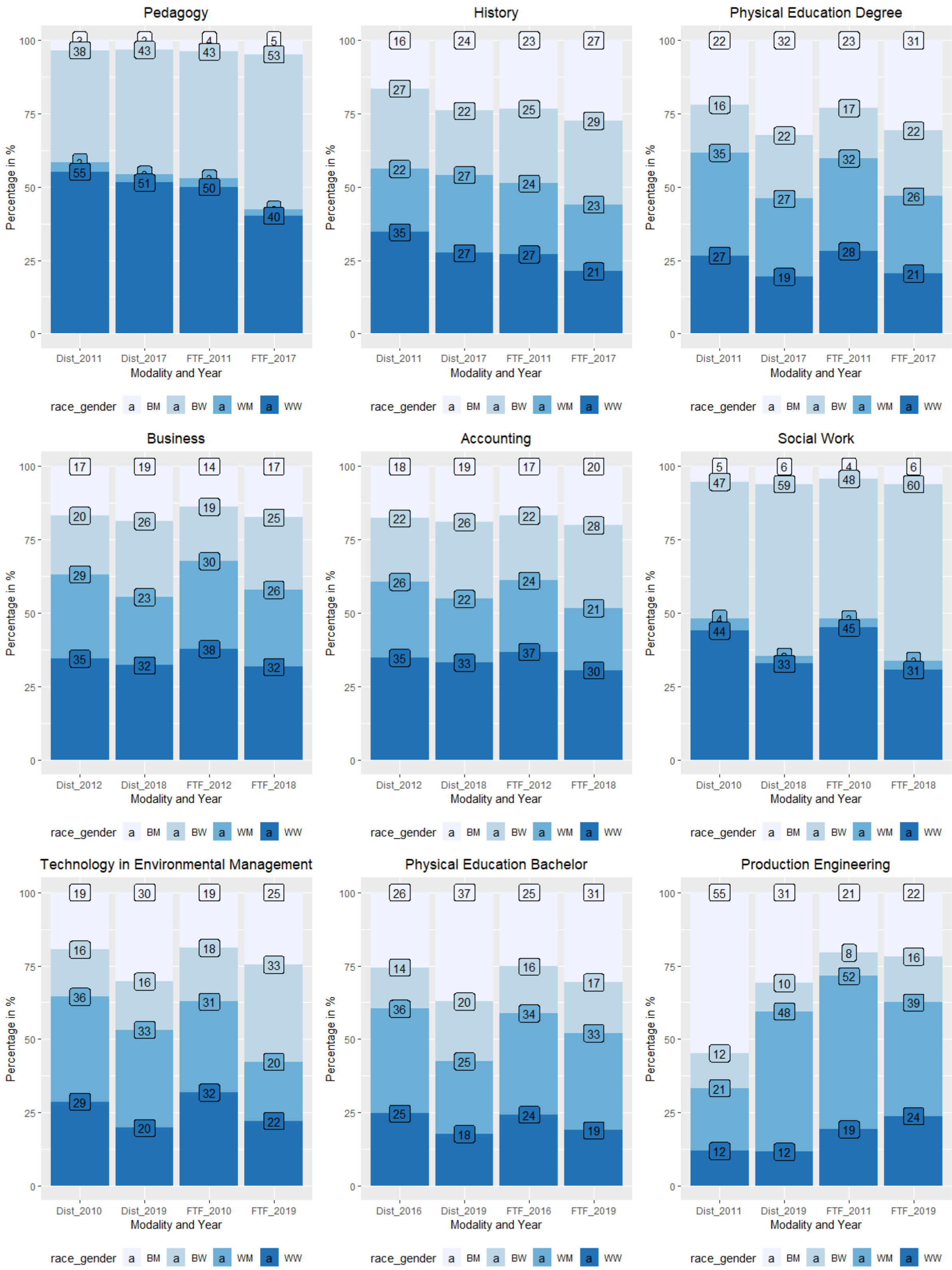

Figure 6 – Distribution of students by gender, race, year, modality and undergraduate course.

Source: Made from ENADE 2010, 2011, 2012, 2016, 2017, 2018 and 2019 microdata, INEP.

Legend note: WM means White Man, WW means White Woman, BW means Black Man and BW means Black Woman.

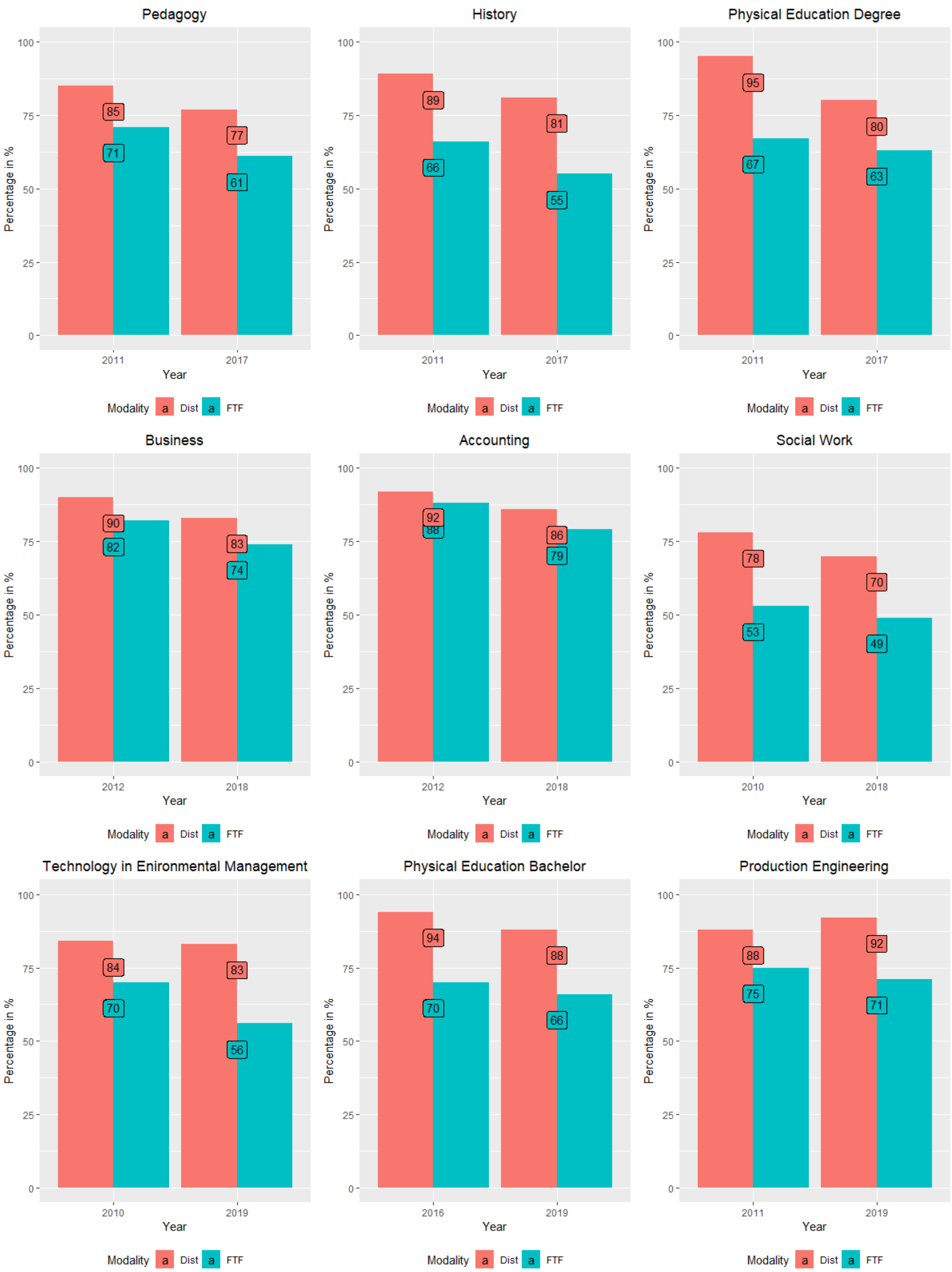

Figure 7 – Concentration in percentage of students in labor market by year, modality and undergraduate course.

Source: Made from ENADE 2010, 2011, 2012, 2016, 2017, 2018 and 2019 microdata, INEP.

learning modality compared to face-to-face modality for all years and undergraduate course. Also it's interesting to note that the gap did not change much over time. In the case of Pedagogy and Accounting the proportions are very close. And the biggest gap is in Production Engineering, in which in 2011 the proportion distance learners from public school was 83% and for in person students the proportion was 53%, in 2019 those numbers were 79% and 60% respectively.

The Figure 9 presents the distribution of students according to their situation at home by year, modality and undergraduate course. In the Figure 9 , about the legend meaning, 'alone' means that the student lives alone, 'other' means in other situation not specified, 'parents' with their parents or relatives, 'sp_ch' with spouse and/or children. According to the Figure 9, we can see that face-to-face modality serves a higher proportion of students living with their parents and, the remote one serves a higher percentage of students living alone or with spouse and/or children. Those gaps changed over time depending on the undergraduate course, but there is no standard on this change. Usually the difference between the proportion of students living alone or with spouse and/or children is 30 p.p. higher for distance learners. For example, for History, the amount of distance learners living alone or with spouse and/or children in 2011 was 70% and 34% for in person students, in 2017 those values were 72% and 33% respectively.

The Figure 10 shows the proportion of students that entered at university by quota policy, over time and by undergraduate course and modality. We can see, according to Figure 10, a big change between the years. For the first year of analysis, in general the gap is small, for some undergraduate courses the proportion is bigger in distance education, for some the proportion is the same or almost the same, and for others the proportion is bigger in face-to-face modality. But for the second year of analysis, the proportion is always bigger in face-to-face approach and the gap is considerable, in other words, the quota politics assist a larger proportion of in-person students than remote students. In the second year of analysis the gap varies from 22 p.p. for Technology in Environmental Management and 9 p.p. for Business.

The Figure[1] 11 shows the distribution of students by type of funding, year, undergraduate course and modality in private universities. In this Figure, about the legend items, fies means that the student receive FIES, fin means another funding, none means no funding, pro means PROUNI (partial or full) and sc means another scholarship. According to Figure 11 we can see that remote modality has a greater concentration of students that do not receive any external funding and less concentration of students assisted by PROUNI and FIES. The gap of students without external funding between modalities increased over time, especially because the decrease in the concentration among in-person students. This gap is higher for Physical Education Bachelor, in 2016 86% of did not receive any

[1] Students receiving partial PROUNI and FIES at the same time were excluded.

Figure 8 – Concentration in percentage of students who attended high school in public school by year undergraduate course and modality.

Source: Made from ENADE 2010, 2011, 2012, 2016, 2017, 2018 and 2019 microdata, INEP.

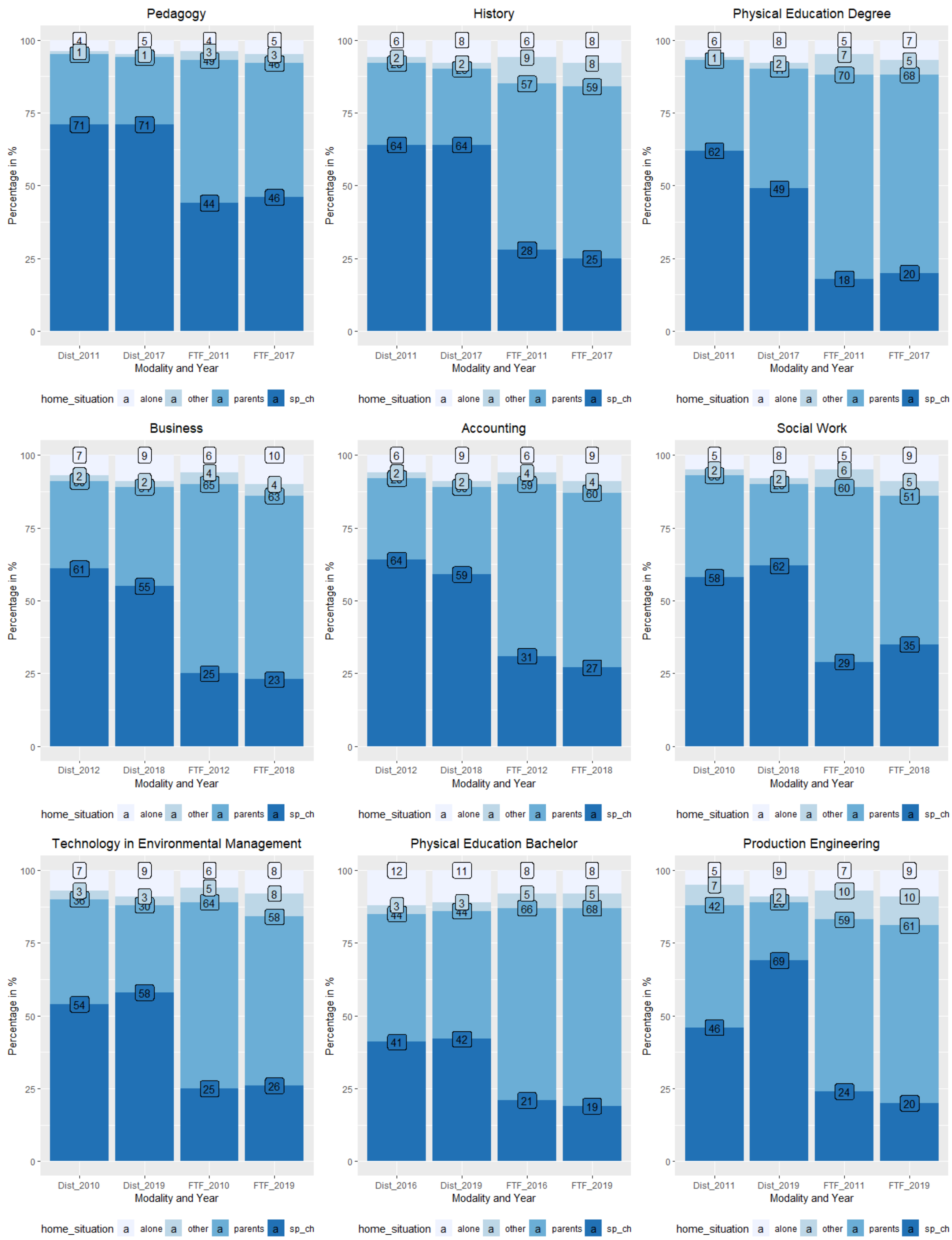

Figure 9 – Distribution of students in relation to situation at home by year, modality and undergraduate course.

Source: Made from ENADE 2010, 2011, 2012, 2016, 2017, 2018 and 2019 microdata, INEP.

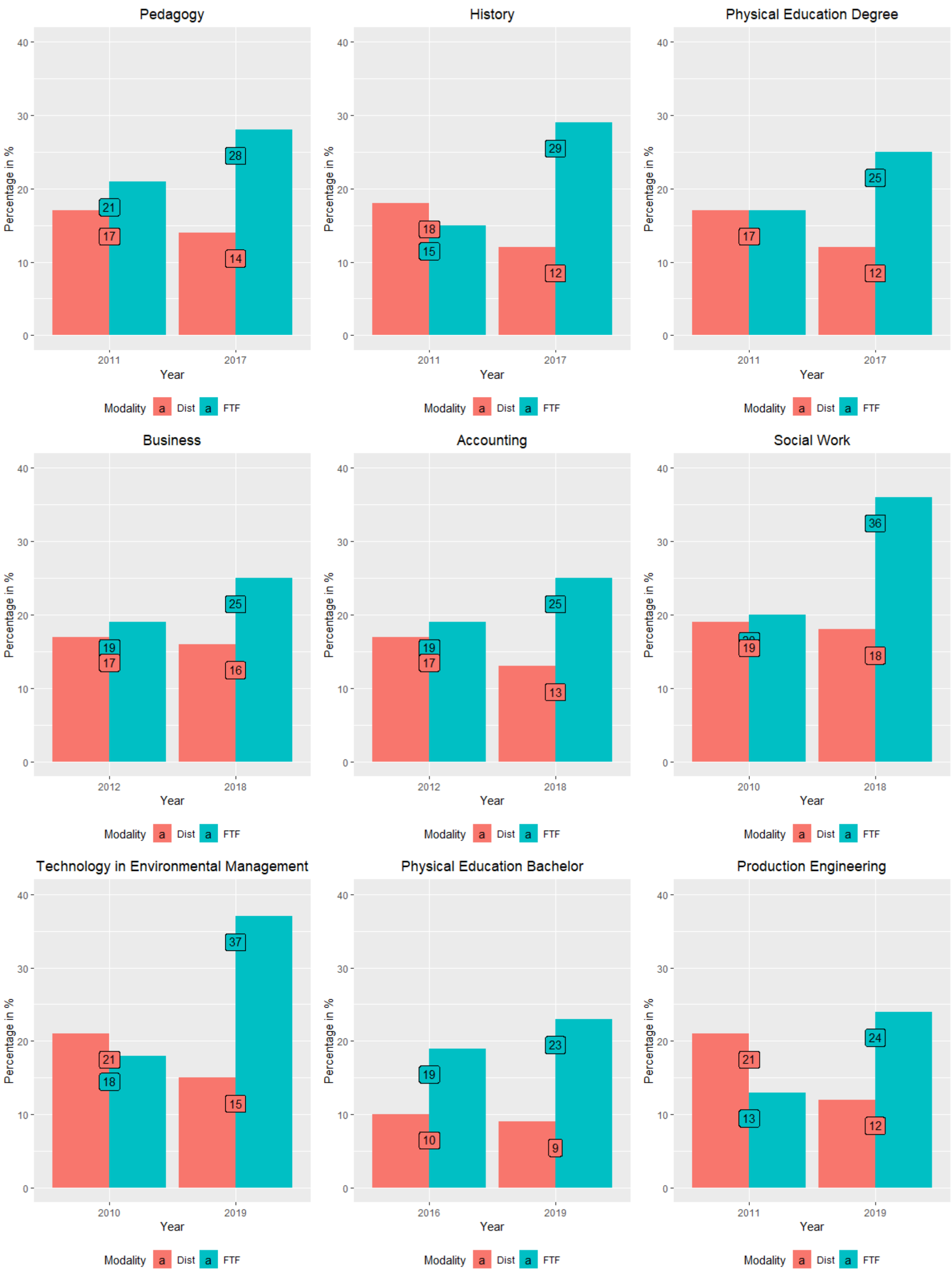

Figure 10 – Proportion of students that entered at university by quota policy, over time and by undergraduate course and modality.

Source: Made from ENADE 2010, 2011, 2012, 2016, 2017, 2018 and 2019 microdata, INEP.

external funding and 42% among in person students. In 2019 those values were 83% and 39% respectively.

About the PROUNI it is worth mentioning the case of Production Engineering, in which no student received the government's benefit in 2011 among distance learners, and 10% of in person students, in 2019 those numbers were 7% and 16% respectively. Lastly, about FIES, we can see in the second year of analysis, this type of funding assisted a high proportion of in-person students, but we can not say the same for distance learners.

2.2.3 Logit Model

In this section we will describe, based on Cameron and Trivedi (2005) the methodology approach. For our purpose we will apply the logit method, this methodology is appropriate for our case in which the dependent variable is binary, "distance modality"and "in-person"modality.

Binary dependent variable models are applied in cases in which there are only two response options for the dependent variable, for example, being enrolled in distance modality or in-person modality. In these models, the objective is to estimate the probability of the occurrence of the event according to the values of the explanatory variables, the estimated probability of the occurrence can be defined by the following equation:

$$p_i = P_r[y_i = 1|x] = F(x_i'\beta) \tag{2.1}$$

In which y_i takes the value of 1 if the event occurs and 0 otherwise. $F(.)$ is the cumulative distribution function (cdf) and guarantees that $0 \leq p \leq 1$. It is also worth mentioning that if $F(.)$ is a cdf, so this cdf is being used just to model the parameter p and it is not the cdf of y.

The model is estimated by maximum likelihood, where the likelihood for data with n binary responses is:

$$L = \prod F(x_i'\beta)^y [1 - F(x_i'\beta)]^{1-y} \tag{2.2}$$

The specific functional form is given by:

$$P(y_i = 1) = \frac{exp(x_i'\beta)}{1 + exp(x_i'\beta)} \tag{2.3}$$

Where the interpretation of the results can be given by mean marginal effect. Once the method is defined, the general model 1 is given by the following equation:

$$Distance = X\beta + u \tag{2.4}$$

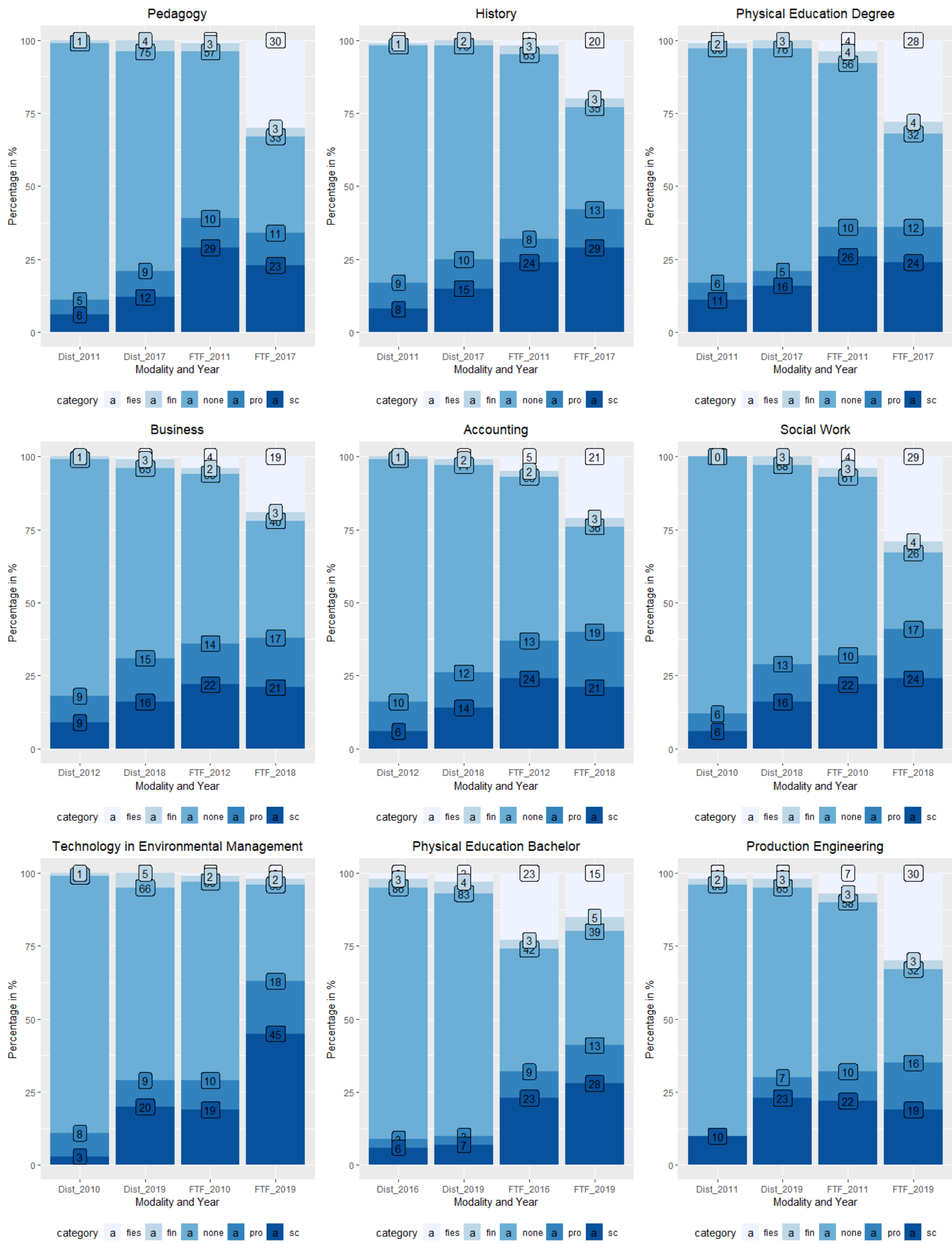

Figure 11 – Distribution of students by type of funding, year, undergraduate course and modality in private universities.

Source: Made from ENADE 2010, 2011, 2012, 2016, 2017, 2018 and 2019 microdata, INEP.

In which the dependent variable $Distance$ assumes the value of 1 if the modality is online and 0 otherwise, X is the vector with the control variables in Table 1 and u is the error term.

2.3 Results

This section aims to evaluate the difference between the profile of undergraduate students in distance and in person modality; the effect of the years in the last decade; if the public attended is the one expected to; the main determinants behind the choice of the modality, and see the heterogeneity in both groups in terms of affirmative policies in financing programs (like FIES and PROUNI) and quotas. To achieve this goal we apply the method of logit and analyse the mean marginal effects.

2.3.1 Main determinants of the choice of the modality and Quota Policies

Here we analyze the results of the logit model of the main determinants of the choice of the modality and also the quota policy for both groups. The Tables 3, 4 and 5 present the modality choice determinants and quota policy by undergraduate course. The complete tables, with the states variables are presented in Appendix A (Tables 20, 21 and 22), here we show a smaller version with only the variables of interest. It is important to mention about the reference categories of the covariates: '2017'(Pedagogy, History and Physical Education Degree), '2018' (Business, Accounting and Social Work) and '2019' (Technology in Environmental Management, Physical Education Bachelor and Production Engineering) are the category reference to year, 'no job' to work, 'live with parents and/or relatives' to situation in home and 'College Degree or more' to parent's schooling. In general most part of the coefficients are significant, and it can indicate that those variables are relevant to explain the modality choice.

First, about age variable, as we saw in descriptive analysis, age is positively related with the choice of distance learning for all undergraduate courses. But as the coefficient of age squared is negative, we can conclude that this relation is positive but with decreasing rates. So those results matches with the literature, like in Latanich, Nonis and Hudson (2001) in which the authors say that the distance learners are older then the typical undergraduate students, mainly because age is positively correlated with other responsibilities.

In relation to the male variable, in general the coefficients are negative or non significant, except for Pedagogy in which the coefficient is positive. It means that being a man at the very least is not positive related with being enrolled in an e-learning course. In History, for example, being a male is negative related of being enrolled in distance education modality in 3.6 p.p. Those results are in agreement with Lima, Borges and

Souza (2018), in which the authors verify a predominance of the female audience among students of distance courses in part because this public has taken a large proportion of the labor market, and they have to reconcile their professional life with domestic activities and children.

About the variable of work, we can see that being in the labor market is positive related with choosing the distance learning modality, and this relation is higher the greater the number of hours worked, exceptions for Accounting, Business and Technology in Environmental Management, in which those undergraduate courses do not exactly follow this standard. The coefficients are greater for Pedagogy, in which if the student has a part time job, his probability of choosing online modality is 7.6 p.p higher than non worker students, and for students with full time job the probability is 10 p.p higher. We discussed in Latanich, Nonis and Hudson (2001) paper that distance learners are more likely to be employed because some of them need to keep working to pay the tuition and the method is more desirable because of schedule and place flexibility, so it is possible reconciling the study with professional occupation.

Considering the situation at home, we can see that for the most part of the undergraduate courses, comparing students that live with parents or relatives, living alone or mainly living with spouse and/or children is positive related with choosing distance education. For Pedagogy, the students that live alone have a probability of choosing distance education in 2.5 p.p greater. and living with spouse and/or children 9.1 p.p larger. For Physical Education Degree, those coefficients are non significant for living alone and 2.7 p.p. living with spouse and/or children. Latanich, Nonis and Hudson (2001) justify that people with family responsibilities prefer distance education because of schedule flexibility.

It is also interesting look at the variable of public school, which indicates if the student had high school in public or private system. As we can see, in general the coefficient is positive, except for Social Work in which the coefficient is non significant and for Accounting in which this coefficient is negative. This means that in general students from high school in public system have a higher probability of choosing distance learning method. We can highlight History, with the higher coefficient, where students from high school in public system have greater probability of choosing distance education in 3.6 p.p. Those results indicates that as students from high school in public system have a greater odds of choosing distance education than students from private system. Considering the problems in the public education in Brazil, it is possible that in general the distance learners start college degree with a bigger lag in education.

It is also important to mention the impact of private university variable. In general the coefficient is positive, except for Physical Education Bachelor and Production engineering, in which the coefficients are insignificant. Those results mean that in comparison with the students from college in public system the students from private system has a greater probability of choosing distance education. For Social work for example, the probability is 43.8 p.p. higher. Considering the scenario of college degree in Brazil, it is a fact that distance education is more available in private institutions.

About the coefficients for quota are negative and significant for all undergraduate courses. We can see that the highest relation in module is in Pedagogy, in which if the student started college degree with quota policy he has a lower probability in 7.2 p.p. of being enrolled in a distance program than students that started college degree without quota policy, and for Technology in Environmental Management and Production Engineering the relation is 2.7 p.p. lower. In other words, it indicates that the quota policy is more likely at in-person programs than in online programs.

According to the coefficients we can make some conclusions about the profile of distance learners. They are older, predominance of feminine audience, there are a greater proportion of distance learners in labor market and have their own family nucleus living alone or with spouse and/or children. As we saw in Latanich, Nonis and Hudson (2001) for older students in the labor market and family responsibilities (mainly women), distance education is a good and maybe the only option because of the schedule flexibility. Moreover, there is a greater proportion of students that did public school in Brazilian public system, and the distance learners may have lag in education, considering the bad quality of public education in Brazil. And about quota policy we can see that the policy attends in a higher proportion the students in face-to-face approach.

Table 3 – Modality choice determinants and quota policy by undergraduate course.

	Dependent variable:		
	Distance learning		
Covariates[2]	(Ped)	(Hist)	(Phy ed de)
Age	0.036***	0.024***	0.019***
	(0.004)	(0.012)	(0.019)
Age^2	−0.0004***	−0.0002***	−0.0002***
	(0.0001)	(0.0002)	(0.0003)
Male	0.026***	−0.036***	−0.008***
	(0.023)	(0.037)	(0.041)
White	−0.010***	−0.010***	−0.032***

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Ped)	(Hist)	(Phy ed de)
	(0.012)	(0.039)	(0.043)
Per capita income	-0.00001^{***}	-0.00000^{**}	-0.00000^{***}
	(0.00001)	(0.00001)	(0.00002)
Part time job	0.076^{***}	0.044^{***}	0.011^{***}
	(0.017)	(0.056)	(0.061)
Full time job	0.100^{***}	0.071^{***}	0.033^{***}
	(0.014)	(0.047)	(0.053)
Live Alone	0.025^{***}	0.008	0.006
	(0.028)	(0.072)	(0.078)
Live with sp and/or ch	0.091^{***}	0.053^{***}	0.027^{***}
	(0.013)	(0.044)	(0.049)
Other	-0.023^{***}	-0.033^{***}	-0.029^{***}
	(0.041)	(0.101)	(0.123)
Public high sch	0.029^{***}	0.036^{***}	0.007^{**}
	(0.019)	(0.054)	(0.066)
Parent ed none	0.002	0.038^{***}	0.005
	(0.028)	(0.089)	(0.125)
Parent ed elementary 1	-0.004	0.024^{***}	-0.004
	(0.019)	(0.058)	(0.061)
Parent ed elementary 2	-0.027^{***}	-0.001	-0.017^{***}
	(0.021)	(0.065)	(0.067)
Parent ed	-0.038^{***}	-0.011^{**}	-0.016^{***}

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Ped)	(Hist)	(Phy ed de)
high sch	(0.019)	(0.056)	(0.056)
Private	0.224***	0.136***	0.011***
university	(0.019)	(0.058)	(0.064)
Quota	−0.072***	−0.052***	−0.032***
	(0.014)	(0.046)	(0.054)
IGC	0.006***	−0.049***	0.026***
	(0.015)	(0.053)	(0.056)
Hours	0.003***	0.002***	0.003***
	(0.002)	(0.005)	(0.005)
2011	0.058***	−0.159***	−0.130***
	(0.016)	(0.061)	(0.100)
2014	0.117***	−0.0004	−0.059***
	(0.014)	(0.042)	(0.055)
Constant	−1.097***	−0.581***	−0.559***
	(0.095)	(0.296)	(0.364)
Observations	231,488	36,558	51,155
Log Likelihood	−98,597.650	−10,007.110	−8,576.090
Akaike Inf. Crit.	197,291.300	20,110.220	17,248.180

Note: *p<0.1; **p<0.05; ***p<0.01.

Source: Made from ENADE microdata 2011, 2014 and 2017, INEP.

[2] Explanatory note on the reference categories of the covariates: '2017' is the category reference to year, 'no job' to work, 'live with parents and/or relatives' to situation in home and 'College Degree or more' to parent's schooling.

Table 4 – Modality choice determinants and quota policy by undergraduate course.

	Dependent variable:		
	Distance learning		
Covariates[3]	(Bus)	(Acc)	(S work)
Age	0.034***	0.029***	0.014***
	(0.006)	(0.009)	(0.010)
Age^2	−0.0003***	−0.0003***	−0.0002***
	(0.0001)	(0.0001)	(0.0001)
Male	−0.014***	−0.001	0.005
	(0.014)	(0.022)	(0.055)
White	−0.051***	−0.041***	−0.025***
	(0.014)	(0.023)	(0.033)
Per capita income	−0.00001***	0.00000**	−0.00001***
	(0.00000)	(0.00001)	(0.00002)
Part time job	0.032***	0.031***	0.036***
	(0.028)	(0.051)	(0.050)
Full time job	0.022***	0.0001	0.040***
	(0.018)	(0.032)	(0.034)
Live alone	0.046***	0.034***	−0.006
	(0.026)	(0.041)	(0.064)
Live with sp and/or ch	0.082***	0.071***	0.030***
	(0.016)	(0.025)	(0.037)
Other	0.013***	0.024***	−0.031***
	(0.042)	(0.068)	(0.092)
Public	0.034***	−0.018***	0.005

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Bus)	(Acc)	(S work)
high sch	(0.019)	(0.032)	(0.049)
Parent ed none	0.011** (0.042)	−0.0003 (0.068)	0.020*** (0.078)
Parent ed elementary 1	0.002 (0.022)	−0.011*** (0.034)	0.008* (0.053)
Parent ed elementary 2	−0.004 (0.023)	−0.018*** (0.036)	−0.002 (0.058)
Parent ed high sch	−0.004* (0.019)	−0.017*** (0.031)	0.0002 (0.052)
Private university	0.169*** (0.029)	0.313*** (0.062)	0.438*** (0.196)
Quota	−0.058*** (0.017)	−0.053*** (0.029)	−0.068*** (0.037)
IGC	−0.078*** (0.019)	−0.035*** (0.032)	−0.068*** (0.054)
Hours	0.005*** (0.002)	0.004*** (0.003)	0.0004 (0.005)
2010			−0.032*** (0.045)
2012	0.045*** (0.016)	−0.024*** (0.030)	
2015	0.046***	0.038***	

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Bus)	(Acc)	(S work)
	(0.016)	(0.025)	
2016			0.024^{***}
			(0.038)
Constant	-0.847^{***}	-0.901^{***}	-0.537^{***}
	(0.125)	(0.201)	(0.325)
Observations	217,159	100,802	57,742
Log Likelihood	−74,712.010	−29,144.520	−13,790.290
Akaike Inf. Crit.	149,520.000	58,385.030	27,676.580

Note: *p<0.1; **p<0.05; ***p<0.01.

Source: Made from ENADE microdata 2010, 2012, 2015, 2016 and 2018, INEP.

Table 5 – Modality choice determinants and quota policy by undergraduate course.

	Dependent variable:		
	Distance learning		
Covariates[4]	(Tech in env)	(Phy ed ba)	(Prod eng)
Age	0.011^{***}	0.024^{***}	0.015^{***}
	(0.024)	(0.020)	(0.027)
Age^2	-0.0001^{***}	-0.0003^{***}	-0.0001^{***}
	(0.0003)	(0.0003)	(0.0004)

Continue...

[3] Explanatory note on the reference categories of the covariates: '2018' is the category reference to year, 'no job' to work, 'live with parents and/or relatives' to situation in home and 'College Degree or more' to parent's schooling.

	Dependent variable:		
	Distance learning		
Covariates	(Tech in env)	(Phy ed ba)	(Prod eng)
Male	-0.004	-0.010^{***}	-0.004
	(0.077)	(0.043)	(0.065)
White	-0.023^{***}	-0.035^{***}	-0.023^{***}
	(0.078)	(0.043)	(0.058)
Per capita income	-0.00000^{*}	-0.00000^{**}	-0.00000^{***}
	(0.00002)	(0.00001)	(0.00001)
Part time job	-0.002	0.044^{***}	0.009
	(0.138)	(0.062)	(0.148)
Full time job	0.001	0.055^{***}	0.025^{***}
	(0.096)	(0.056)	(0.093)
Live alone	0.025^{**}	0.019^{***}	0.029^{***}
	(0.158)	(0.076)	(0.107)
Live with sp and/or ch	0.051^{***}	0.034^{***}	0.035^{***}
	(0.096)	(0.052)	(0.071)
Other	0.012	0.007	0.006
	(0.196)	(0.115)	(0.165)
Public high sch	0.018^{**}	0.008^{**}	0.009^{***}
	(0.105)	(0.056)	(0.072)
Parent ed none	0.004	-0.010	-0.007
	(0.213)	(0.149)	(0.239)
Parent ed elementary 1	0.016^{**}	-0.009^{**}	-0.011^{***}
	(0.120)	(0.068)	(0.090)

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Tech in env)	(Phy ed ba)	(Prod eng)
Parent ed elementary 2	0.005	-0.020^{***}	-0.014^{***}
	(0.129)	(0.069)	(0.095)
Parent ed high sch	0.003	-0.022^{***}	-0.006^{**}
	(0.108)	(0.053)	(0.073)
Private university	0.364^{***}	1.189	0.0002
	(0.193)	(252.161)	(0.098)
Quota	-0.027^{***}	-0.041^{***}	-0.027^{***}
	(0.091)	(0.060)	(0.081)
IGC	0.069^{***}	0.121^{***}	-0.029^{***}
	(0.126)	(0.094)	(0.074)
Hours	0.008^{***}	0.002^{***}	0.002^{***}
	(0.011)	(0.006)	(0.007)
2010	-0.193^{***}		
	(0.111)		
2011			-0.675
			(247.794)
2014			-0.024^{***}
			(0.077)
2016	-0.016^{**}	-1.122	
	(0.108)	(236.114)	
Constant	-0.682^{***}	-2.050	-0.340^{***}
	(0.617)	(252.161)	(0.541)

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Tech in env)	(Phy ed ba)	(Prod eng)
Observations	11,169	42,057	35,783
Log Likelihood	−2,455.912	−7,426.691	−4,644.783
Akaike Inf. Crit.	5,005.824	14,945.380	9,381.565

Note: *p<0.1; **p<0.05; ***p<0.01.

Source: Made from ENADE microdata 2010, 2011, 2014, 2016 and 2019 INEP.

2.3.2 Scholarship and Financing Policies in Distance Learning and Face-to-Face Programs

In this section our focus is analysing the financing programs in private universities between distance education and face-to-face modality applying logit model. Since those financing policies do not include public universities, our analyses here are just for private institutions. The complete table with all control variables are in Appendix B in Tables 23, 24 and 25. But since our goal here is to investigate the differences between distance learning and face-to-face programs in private universities in terms of scholarship and financing Policies, the Tables 6, 7 and 8 present only the coefficient of those variables.

About the coefficient of the scholarship variable we can see that it is negative and significant for all undergraduate courses, which means that if the student is supported by the program in comparison with students that are not supported the probability of being enrolled in a distance education course is smaller. For Pedagogy, the probability is 18.9 p.p. lower and for Production engineering the probability 4.7 p.p. minor.

Another variable interesting here is the financing one. The coefficient is also negative and significant for all undergraduate courses, so if the student receives a financing support in comparison with those that does no receive the probability of being a distance learner is lower. For Pedagogy, the undergraduate course with the greater relation in module, the value is negative in 51.7 p.p., and for Production Engineering, the undergraduate course with the lower relation in module, the value is negative in 12.7 p.p.

4 Explanatory note on the reference categories of the covariates: '2018' is the category reference to year, 'no job' to work, 'live with parents and/or relatives' to situation in home and 'College Degree or more' to parent's schooling.

Table 6 – Financing programs in private universities between distance education and face-to-face modality by undergraduate course.

	Dependent variable:		
	Distance learning		
Covariates	(Ped)	(Hist)	(Phy ed de)
Scholarship	-0.189^{***}	-0.154^{***}	-0.071^{***}
	(0.018)	(0.058)	(0.061)
Financing	-0.337^{***}	-0.327^{***}	-0.132^{***}
	(0.027)	(0.103)	(0.102)
Constant	-0.517^{***}	-0.448^{***}	-0.507^{***}
	(0.119)	(0.371)	(0.499)
Observations	157,784	15,032	33,409
Log Likelihood	−55,285.930	−5,311.896	−5,027.438
Akaike Inf. Crit.	110,667.900	10,703.790	10,148.880

Note: *p<0.1; **p<0.05; ***p<0.01.

Source: Made from ENADE microdata 2011, 2014 and 2017, INEP.

Table 7 – Financing programs in private universities between distance education and face-to-face modality by undergraduate course.

	Dependent variable:		
	Distance learning		
Covariates	(Bus)	(Acc)	(S work)
Scholarship	-0.136^{***}	-0.116^{***}	-0.147^{***}
	(0.017)	(0.026)	(0.042)
Financing	-0.277^{***}	-0.280^{***}	-0.267^{***}

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Bus)	(Acc)	(S work)
	(0.034)	(0.053)	(0.070)
Constant	-0.582^{***}	-0.623^{***}	-0.137^{***}
	(0.133)	(0.201)	(0.277)
Observations	171,090	75,914	41,939
Log Likelihood	−58,504.630	−24,103.730	−10,369.910
Akaike Inf. Crit.	117,105.300	48,303.460	20,835.830

Note: *p<0.1; **p<0.05; ***p<0.01.
Source: Made from ENADE microdata 2010, 2012, 2015, 2016 and 2018, INEP.

Table 8 – Financing programs in private universities between distance education and face-to-face modality by undergraduate course.

	Dependent variable:		
	Distance learning		
Covariates	(Tech in env)	(Phy ed ba)	(Prod eng)
Scholarship	-0.052^{***}	-0.122^{***}	-0.047^{***}
	(0.102)	(0.057)	(0.067)
Financing	-0.091^{***}	-0.089^{***}	-0.127^{***}
	(0.207)	(0.075)	(0.118)
Constant	-0.325^{***}	-0.831^{***}	-0.407^{***}
	(0.636)	(0.471)	(0.588)
Observations	8,000	34,378	24,560

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Tech in env)	(Phy ed ba)	(Prod eng)
Log Likelihood	−1,865.870	−6,577.816	−3,843.904
Akaike Inf. Crit.	3,795.151	12,965.200	7,579.416

Note: *p<0.1; **p<0.05; ***p<0.01.

Source: Made from ENADE microdata 2010, 2011, 2014, 2016 and 2019, INEP.

2.4 Final Considerations

As we saw, it is important to know the target audience of a public policy for evaluation, implementation and improvements, and this fact is also applied for public policy in distance higher education. So in this paper our goals were: to identify the main determinants of the choice for modality, the level of the effect and the heterogeneous impact among undergraduate courses and also the incidence of affirmative policies like quota, scholarship and financing programs between distance education and face-to-face approach.

To achieve our goals we estimated logit models to capture the mean marginal effects for Pedagogy, History, Physical Education Degree, Business, Accounting, Social Work, Technology in Environmental Management, Physical Education Bachelor and Production Engineering in Brazil, over the years of 2010 to 2019, using ENADE's database.

In relation to our results we detected the main characteristics related with the choice for modality and that this relation is heterogeneous among the undergraduate courses. Some variables as age, having a job, coming from public high school, coming from private universities and if the student lives alone or with spouse and/or child in comparison with students that live with their parents have a positive relation with the probability of being enrolled in a distance education program.

According to the coefficients we can make some conclusions about the profile of distance learners. They are older, predominance of feminine audience, there are a greater proportion of distance learners in labor market and have their own family nucleus living alone or with spouse and/or children. As we saw in Latanich, Nonis and Hudson (2001) for older students in the labor market and family responsibilities (mainly women), distance education is a good and maybe the only option because of the schedule flexibility. Moreover, there is a greater proportion of students that did public school in Brazilian public system, and the distance learners may have some lag in education, considering the bad quality of public education in Brazil. In this sense, we can infer that for the profile attended by distance education, maybe this is the only option for that public to enter in higher education.

About the quota policy we saw that the coefficients were negative and significant for all undergraduate courses. In other words, it indicates that the students with quota policy is more likely at in-person programs than in online programs. So the incidence of quota policy is higher for face-to-face programs. The same occurs in relation to scholarship and financing programs in private universities. In which those programs assist more in-person students.

Finally with this study, we intended to contribute with the literature: identifying the main characteristics related with the choice for modality, the dimension of this relation and the heterogeneity among undergraduate courses in higher education in Brazil in the last decade; evaluate the incidence of affirmative policies like quota, scholarship and financing programs between both modalities and also to contribute with strategic policy decision making in the sense of improving the distance education approach taking into account the target audience.

3 The effect of modality in dropout rate in higher education

3.1 Introduction

Dropout in higher education is a generalized problem that occurs in developed and in developing countries as stated by Bonaldo and Pereira (2016). The concern is mainly about the costs involved for the individual and also the entire society (Lassibille; Gómez, 2008). The authors argument that it is crucial to understand and identify the main determinants in order to develop strategies to reduce the issue and maximize the allocation of resources in higher education, since there is potential for improvements.

The main problems related to research in this area are about difficulties in the definition of dropout, which is still in debate, and also in establishing the proper manner to measure it. In relation to the effect of modality in withdraw, there are still not that many works applied to the Brazilian reality. Normally the papers are about dropout rate in higher education in general or focused only in distance learning.

So in this paper we intend to estimate the impact of modality in dropout rate in Brazilian universities and the heterogeneity among undergraduate courses, using data from Higher Education Census, applying the method of ordinary least squares, for the years of 2018 and 2019. The undergraduate courses selected are: Pedagogy Degree, History Degree, Physical Education Degree, Business, Accounting, Social Work, Environmental Management, Physical Education Bachelor and Production Engineering.

Some discussions and evidences about the main determinants of withdraw can be seen in Lassibille and Gómez (2008) and Bonaldo and Pereira (2016). The papers discuss about the influence of age, family background, financing support and college quality in dropout rates.

In Lassibille and Gómez (2008), the paper analyses dropout rates among students in higher education in Spain using discrete time hazard models. About the results, the authors infer that: students that drop out are less motivated and skilled; the withdraw is greater in early years; age is positive correlated with dropout; students receiving financing support are less likely to give up because those programs reduce the cost of opportunity of education; students with parents with higher levels of education are more likely to finish college and students originally from another city are more likely to withdraw duo to spending time with travels and home activities.

In the same vein, a research for Brazilian case Bonaldo and Pereira (2016) use logit models to establish the main determinants of dropout rates in higher education. The work finds some evidence of age and change in marital status as positively affecting withdraw and college quality growing the rates of finishing college degree.

For dropout rate in higher distance education programs Levy (2007) and SANTOS and Giraffa (2015) have some considerations. According to SANTOS and Giraffa (2015) with the expansion of distance learning, consequently grow the challenges, so it is important to measure withdraw to develop permanence and retention actions. Lastly, in relation to Levy (2007), the author studied the influence of locus of control and motivation in dropout, applying ANOVA and non-parametric tests. The author found no evidence about impact of locus of control in withdraw, a negative effect of motivation and college quality.

It is a fact that withdraw is a problem in higher education and with potential for improvements. But according to our results, it seems that modality has no effect in higher education dropout rate, when we control for other variables.

We expect with this research to contribute with Brazilian literature about the withdraw in Brazilian higher education by capturing the effect of modality in dropout rate in Brazilian universities and also evaluating the heterogeneity of the impact among undergraduate courses.

Besides this introduction, this chapter is also composed by the section 2 with information about data and methods. The section 3 contain the results of the OLS models. Finally in the section 4 we have the final considerations.

3.2 Data and Methods

This section contains information about strategies and methods used to achieve the main goal. In this paper we evaluate the difference between the dropout rate in higher education for undergraduate courses in distance and in-person modality. For this purpose, the section includes the presentation of the database, description of the econometric techniques, definition of the model and variables, descriptive analysis of the data and exposure of the models and methods applied.

3.2.1 Database and variables

Here we present information about the database and variables used. It is intended, therefore, to describe the strategies used for the treatment of the database and to discuss the choice of the dependent and control variables. The database is from Higher Education Census, provided by INEP (National Institute of Educational Studies and Research Anísio Teixeira). The census collects annually information on the infrastructure of higher education institutions, vacancies offered, candidates, enrollments, freshmen, graduates and professors, in the different forms of academic organization and administrative category. In addition, there are also data about the General Course Index that are added to the Higher Education Census database, where such data are also provided by INEP.

To analyse the difference between the dropout rate of higher education from distance and in-person modality we select 9 undergraduate courses: Pedagogy Degree, History Degree, Physical Education Degree, Business, Accounting, Social Work, Environmental Management, Physical Education Bachelor and Production Engineering. And we also selected the years of 2018 and 2019. Once the courses and the years are defined, 9 models are estimated, one for each course using the method of ordinary least squares, in which the dependent variable is the dropout rate. The calculus of the dropout rate in this work is a indicator constructed by the authors and based on INEP's definition of dropout, which corresponds to students with a bond situation equal to "disengaged from the course" or "transferred to another course at the same Institution". So the dropout ratio is given by:

$$DR_{imcuy} = \frac{QD_{imcuy} + QT_{imcuy}}{QD_{imcuy} + QT_{imcuy} + QE_{imcuy} + QA_{imcy}} \tag{3.1}$$

In which DR_{imcuy} is the dropout rate in Institution i, modality m, city c[1], undergraduate course u and year y. QD is the number of disengaged students. QT is the number of transferred students to another course at the same Institution. QE is the number of enrolled students. QA is the number of students with closed enrollment

The variable of interest (modality) and the other control variables are shown in Table 9. The variables indicate characteristics of the profile of the students, modality and Institutions quality.

3.2.2 Descriptive Analysis

This section presents a descriptive analysis of the dependent variables data related to the dropout rate at university in order to contextualize and understand the heterogeneity of the dropout rate between distance undergraduate courses and in-person undergraduate courses.

The Table 10 shows the total number of observations, courses and universities by modality. We have a total of 15,974 observations, of which 14,338 are in-person approach and 1636 are distance learning. In total, we have a total of 7,987 courses offered, of which 7,169 are face-to-face and 818 in distance learning. In addition, there are 1,894 Universities, in which 1,885 offer at least one of the courses in the face-to-face modality and 257 offer at

[1] As we do not have information of county for distance undergraduate courses in those cases the aggregation is in the level of Institution, undergraduate course and year

[2] The calculus of the proportions are the quantity of students in the specified category divided by total enrolled students

Table 9 – Covariates

Variables[2]	**Description**
Distance Learning	Dummy equals to 1 if the individual is enrolled in a distance learning program.
Man	Proportion of men enrolled in the undergraduate course.
Young	Proportion of students aged 29 and under.
White	Proportion of White students including Asians. The Non White students are the black, brown and indigenous students. Students who did not inform the race were were excluded.
Public School	Proportion of students who finished high school in public schools. Students who did not inform the type of school where they completed high school were excluded.
Quota	Proportion of students who entered by quota.
Social Support	Proportion of students receiving social support.
Private	Dummy equals to 1 if the undergraduate course is private.
Extracurricular Activities	Proportion of students who carry out extracurricular activities.
Year	Dummy equals to 1 if the year is 2019 and 0 for 2018.
IGC	University Score

least one of the courses in the distance-learning modality. Finally, it is worth remembering that these are all observations from the database. In the rest of the descriptive analysis, as well as in the regression models, there may be fewer observations due to missing variables.

Table 10 – Number of observations, Universities and courses by modality and undergraduate course.

	Observations			**Courses**			**Universities**		
	FTF	**Dist**	**Tot**	**FTF**	**Dist**	**Tot**	**FTF**	**Dist**	**Tot**
Pedagogy De.	2630	434	3064	1315	217	1532	1041	217	1090
History De.	682	144	826	341	72	413	250	72	282
Phy. Ed. De.	1230	100	1330	615	50	665	505	50	511
Business	3648	338	3986	1824	169	1993	1535	169	1545
Accounting	2386	258	2644	1193	129	1322	1050	129	1064
Social Work	768	96	864	384	48	432	352	48	371
Env. Man.	294	110	404	147	55	202	117	55	161
Phy. Ed. Ba.	1230	54	1284	615	27	642	545	27	548
Production Engineering	1470	102	1572	735	51	786	640	51	652

Source: Made from Higher Education Census 2018 and 2019, INEP.

3.2.2.1 Dependent Variable

Here is presented a descriptive analysis of the dependent variable which is the dropout rate for Pedagogy, History, Physical Education Degree, Business, Accounting, Social Work, Environmental Management, Physical Education Bachelor and Production

Engineering in 2018 and 2019. The Figure 12 below presents the mean and 95% confidence interval of dropout rate by undergraduate course, year and modality.

It can be seen from the Figure 12 that, in general, the dropout rate for distance courses is higher or equal to in-person courses. But the difference is not always significant for all years and all courses. It is also worth mentioning that the dropout rate is greater in 2019 than 2018 and the gap grow from one year to another.

More specifically we can see that the gap in dropout rate is not significant for Physical Education Degree, Social Work, Environmental Management and Physical Education Bachelor for both years. The gap is significant in 2019 for Pedagogy, Business, Accounting and Production Engineering. And for History, the difference is significant in both years, in 7.69 p.p. in 2018 and 7.78 p.p. in 2019.

3.2.3 Methodology

In this section we will make a brief discussion about the method and model, based in Greene (2001), to achieve our goal of estimating the effect of modality in higher education dropout rate in Brazil. So, in this sense we will estimate 9 models, one for each undergraduate course using OLS regression. A general equation for our case can be given by:

$$Dropout_{ij} = X_{ij}\beta + \varepsilon_{ij} \tag{3.2}$$

In which $Dropout_{ij}$ is the proportion of dropout of undergraduate course i at the university j, X_{ij} is the matrix of variable of interest of modality and the other control variables presented in Table 9, β is the vector of marginal effects of the covariates and ε_{ij} is the error term.

And the β estimator can be given by:

$$\beta = (X'X)^{-1}X'Y \tag{3.3}$$

3.3 Results

Here we present the results about the effect of modality in dropout rate in higher education. To achieve this goal we use OLS estimators to analyse the marginal effects. In this sense, the Tables 11, 12 and 13 present the determinants of dropout rate in higher education by undergraduate course in the years of 2018 and 2019.

Before we analyse the modality coefficient, which is our main variable of interest, lets to examine the coefficient of some of the other variables. Also, it is important to say

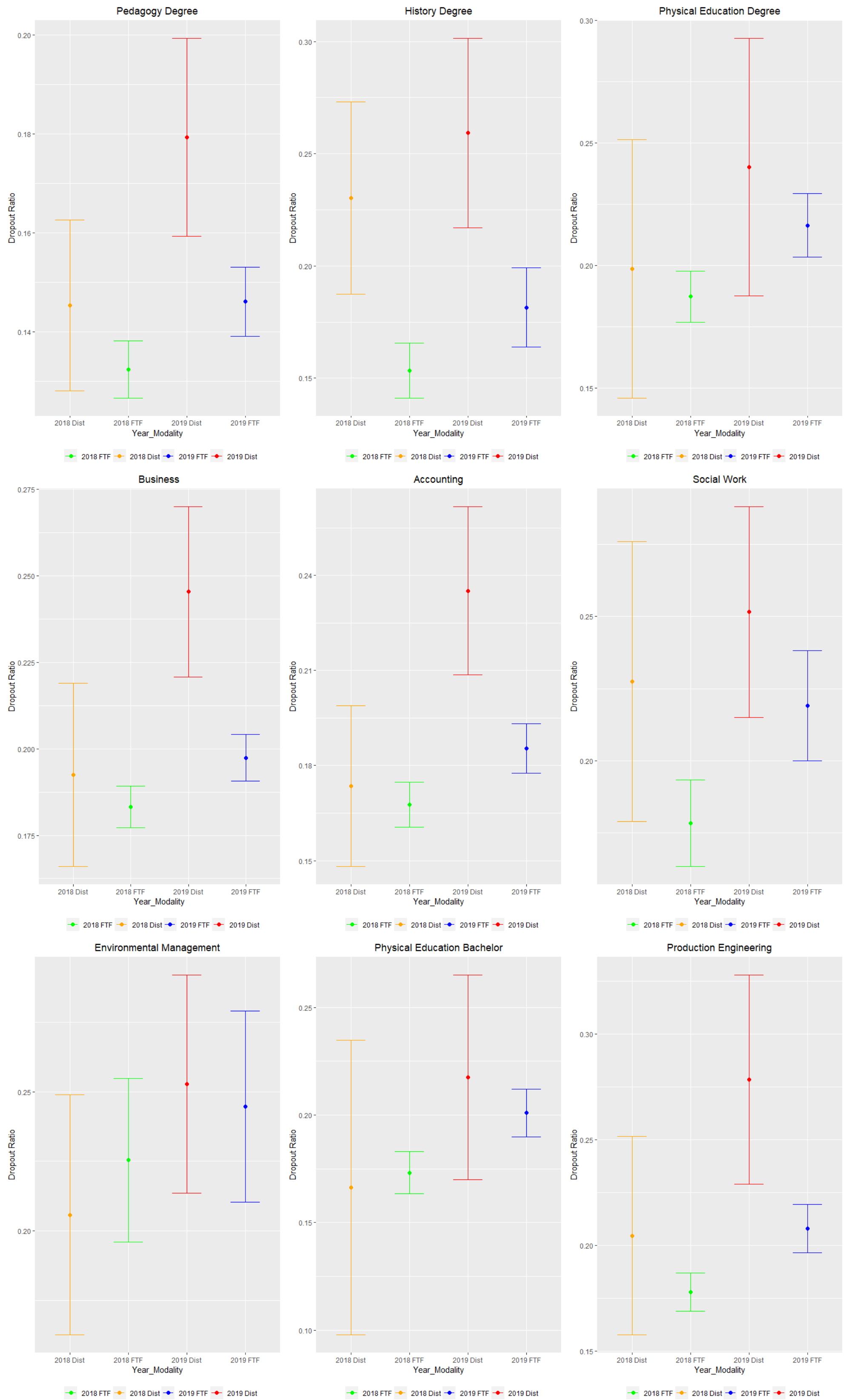

Figure 12 – Mean and 95% Confidence Interval of Dropout Rate by undergraduate course, year and modality

Source: Made from Higher Education Census 2018 and 2019, INEP.

that with our estimations we do not intend to be conclusive about the topic, but we do expect to bring some considerations about those relations and to be a starting point for future researches in this area.

Firstly about the private university coefficient we observe that is positive and significant for most undergraduate course, except for Social Work. Which indicates that private institutions have higher dropout rates, for example, for Physical Education Degree it might mean a greater rate in 7.9 p.p.

In relation to the year, we see that the coefficient for 2019 is positive and significant for all undergraduates courses, in other words, as we saw in Figure 12, we had a raise in dropout rate from 2018 to 2019. Considering the quality of the university captured by IGC variable, we verify that the impact is possibly negative or non significant, except for Pedagogy.

According to the young proportion coefficients, the values indicate that, undergraduate courses with higher proportion of young students have lower dropout rates, except for Physical Education Degree and Bachelor. Taking a look at the variable of proportion of quota students, the coefficients are negative or non significant, so we might infer that a greater proportion of quota students do not grow the dropout rate and possibly diminishes in some cases. About the proportion of students in academic activities we see a negative impact, except for History.

Finally about the variable of modality we observe that in general the coefficients are non significant, except for History where the coefficient is positive and negative for Business and Environmental Management. So, although the problem in higher education is a concern, possibly the modality does not have impact on it. In fact, the dropout rate is an issue in higher education, looking at the constant, the dropout varies from 13% in Pedagogy to about 30% in Social Work.

Table 11 – Determinants of dropout rate by undergraduate course.

	Dependent variable:		
	Dropout rate		
Covariates	(Pedagogy)	(History)	(Phy. Ed. DE.)
Distance Education	-0.010	0.039^{**}	-0.006
	(0.007)	(0.017)	(0.018)
2019	0.016^{***}	0.026^{***}	0.028^{***}

Continue...

	Dependent variable:		
	Dropout rate		
Covariates	(Pedagogy)	(History)	(Phy. Ed. DE.)
	(0.004)	(0.009)	(0.008)
Private Uni.	0.053***	0.049***	0.079***
	(0.007)	(0.014)	(0.013)
IGC	0.012***	−0.041***	−0.004
	(0.005)	(0.011)	(0.009)
Prop. Man	−0.059*	0.116***	0.029
	(0.035)	(0.040)	(0.039)
Prop. Young	−0.065***	−0.052*	−0.050
	(0.014)	(0.029)	(0.035)
Prop. White	0.022**	0.020	0.025
	(0.009)	(0.021)	(0.016)
Prop. Publ. High Sch	−0.021**	0.006	0.016
	(0.010)	(0.020)	(0.017)
Prop. Quota	−0.029	−0.004	−0.057
	(0.019)	(0.037)	(0.042)
Prop. Academic Act	−0.038***	−0.020	−0.031**
	(0.007)	(0.020)	(0.014)
Prop. Support	−0.019	0.025	−0.019
	(0.013)	(0.024)	(0.023)
Constant	0.131***	0.216***	0.152***
	(0.019)	(0.045)	(0.047)

Continue...

	Dependent variable:		
	Dropout rate		
Covariates	(Pedagogy)	(History)	(Phy. Ed. DE.)
Observations	2,778	782	1,250
R^2	0.083	0.152	0.111
Adjusted R^2	0.079	0.140	0.103
Residual Std. Error	0.112	0.132	0.138
F Statistic	22.780***	12.564***	14.025***

Note: *p<0.1; **p<0.05; ***p<0.01.

Source:Made from Higher Education Census 2018 and 2019, INEP.

Table 12 – Determinants of dropout rate by undergraduate course.

	Dependent variable:		
	Dropout rate		
Covariates	(Business)	(Accounting)	(Social Work)
Distance Education	−0.018**	−0.014	−0.014
	(0.009)	(0.010)	(0.019)
2019	0.019***	0.022***	0.032***
	(0.004)	(0.005)	(0.010)
Private Uni	0.058***	0.061***	0.019
	(0.009)	(0.012)	(0.022)
IGC	−0.006	−0.008	−0.004
	(0.005)	(0.006)	(0.012)
Prop. Man	0.067***	0.002	−0.130
	(0.023)	(0.026)	(0.082)

Continue...

	Dependent variable:		
	Dropout rate		
Covariates	(Business)	(Accounting)	(Social Work)
Prop. Young	-0.161^{***}	-0.139^{***}	-0.145^{***}
	(0.019)	(0.021)	(0.034)
Prop. White	-0.007	-0.024^{**}	-0.016
	(0.008)	(0.010)	(0.020)
Prop. Publ. High Sch	-0.012	0.001	-0.010
	(0.009)	(0.011)	(0.024)
Prop. Quota	-0.084^{***}	-0.061^{*}	-0.095^{*}
	(0.025)	(0.033)	(0.054)
Prop. Academic Act	-0.035^{***}	-0.043^{***}	-0.041^{**}
	(0.007)	(0.008)	(0.018)
Prop. Support	-0.003	0.004	-0.008
	(0.014)	(0.016)	(0.032)
Constant	0.271^{***}	0.265^{***}	0.301^{***}
	(0.024)	(0.029)	(0.047)
Observations	3,662	2,482	780
R^2	0.091	0.094	0.111
Adjusted R^2	0.088	0.090	0.098
Residual Std. Error	0.127	0.122	0.143
F Statistic	33.092^{***}	23.366^{***}	8.678^{***}

Note: *p<0.1; **p<0.05; ***p<0.01.

Source:Made from Higher Education Census 2018 and 2019, INEP.

Table 13 – Determinants of dropout rate by undergraduate course.

	Dependent variable:		
	Dropout rate		
Covariates	(Env. Man,)	(Phy. Ed. Ba)	(Prod. Eng)
Distance Education	-0.087^{***}	-0.006	0.016
	(0.027)	(0.019)	(0.016)
2019	0.033^{*}	0.024^{***}	0.025^{***}
	(0.017)	(0.007)	(0.006)
Private Uni	0.055^{*}	0.050^{***}	0.038^{***}
	(0.029)	(0.015)	(0.014)
IGC	0.010	-0.017^{**}	-0.024^{***}
	(0.023)	(0.009)	(0.008)
Prop. Man	0.108^{*}	-0.088^{*}	0.076^{**}
	(0.057)	(0.045)	(0.031)
Prop. Young	-0.099^{*}	0.044	-0.064^{**}
	(0.055)	(0.035)	(0.026)
Prop. White	-0.023	-0.014	-0.066^{***}
	(0.034)	(0.015)	(0.014)
Prop. Publ High Sch	-0.003	0.010	0.038^{***}
	(0.038)	(0.016)	(0.014)
Prop. Quota	-0.127^{*}	-0.089^{*}	-0.039
	(0.075)	(0.048)	(0.046)
Prop. Academic Act	-0.088^{***}	-0.053^{***}	-0.037^{***}
	(0.034)	(0.011)	(0.012)
Prop. Support	0.041	-0.012	-0.018

Continue...

	Dependent variable:		
	Dropout rate		
Covariates	(Env. Man,)	(Phy. Ed. Ba)	(Prod. Eng)
	(0.054)	(0.023)	(0.020)
Constant	0.219***	0.215***	0.230***
	(0.083)	(0.049)	(0.040)
Observations	358	1,238	1,456
R^2	0.108	0.087	0.142
Adjusted R^2	0.079	0.079	0.135
Residual Std. Error	0.163	0.120	0.123
F Statistic	3.802***	10.635***	21.648***

Note: *p<0.1; **p<0.05; ***p<0.01.
Source:Made from Higher Education Census 2018 and 2019, INEP.

3.4 Final Considerations

As we saw, withdraw in higher education is a problem that occurs in developed and in developing countries and generates social costs. In this sense, it is important to create manners to dimension the dropout size in order to develop strategies to reduce the problem and improve the allocation of the resources in higher education, since there is potential to make it.

The main issues about research in this area are related to difficulties in definition, since it is still in debate about the correctly manner to measure it. In addition, with the growth in distance learning also increases the challenges to deal with this modality and there is a scarcity of study comparing dropout rate between modalities, specially for Brazil.

So in this paper we intended to contribute with the literature by estimating the effect of modality on withdraw in Brazilian higher education, using OLS estimators and data from Higher Education Census, in the years of 2018 and 2019. In our paper the undergraduate courses selected were: Pedagogy, History, Physical Education Degree, Business, Accounting, Social Work, Environmental Management, Physical Education Bachelor and Production Engineering.

Moreover, it is important to reinforce that with our estimations we do not expect to be conclusive about the topic, but instead, to bring some considerations about those relations and to be a starting point for future researches in this area. So with our results we detected that in fact the levels of dropout rate in Brazilian higher education is high, but it seems that there is no effect of modality on it.

4 Do distance learners perform better? An evaluation for Brazil case in higher education

4.1 Introduction

As we saw in the first chapter we had a huge growth in distance learning in higher education in Brazil in the last decade. But we do not have many researches evaluating the quality of those undergraduate courses and its comparison with the in-person programs. It is important to highlight that there are some difficulties and challenges about this kind of analysis. Firstly, we have to decide about the appropriate parameters to measure undergraduate courses quality and academic outcomes, which might be courses grades, tests and exams. Secondly, and also the most challenging one, is the problem of endogeneity and bias selection involved. In other words, the same attributes related to the enrollment in a distance education or face-to-face undergraduate course can be the same ones relevant to determine the academic performance. So if we just compare the performance in both groups, the results may be contaminated by endogeneity.

Therefore, to try to fill the gap in Brazilian Literature we will compare the performance in ENADES' score (a mandatory exam for higher education graduates to evaluate the quality of undergraduate courses in which the exam is composed by a general and specific component) between distance learners and in-person students. We will analyse a total of 9 courses with higher number of distance students to capture the heterogeneity impact of the modality among undergraduate courses: Pedagogy Degree, History Degree and Physical Education Degree in 2017; Business, Accounting and Social Work in 2018 and Technology in Environmental Management, Physical Education Bachelor and Production Engineering in 2019. To control for the selection bias we will use the propensity score matching, the nearest neighbor and inverse probability weighing algorithm, with replacement and for observations in common support and, measuring the average effect of treatment in the treated individuals (ATT). For the study we will also implement the balance and sensitivity analysis.

After our goals are established, it is important to discuss what the literature says about this topic. It is interesting to point out the main domains research in distance education effectiveness, as presented in Shachar and Neumann (2003): Student attitude and satisfaction regarding delivery of coursework; Interactions of students and faculty during delivery of coursework; Student outcomes in distance education coursework; Faculty satisfaction with delivery and coursework and other factors like student comfort, convenience, independence, and perceptions of effectiveness.

Although the goal here is about a objective outcome, which is the score in ENADE exam, it is good to reinforce the importance of other parameters to measure the quality of online courses. Therefore we need to establish the issues about the objective measurements of the academic performance factors and the operationalization of these measurements. So the authors quote that it can be based on course grades, tests, and exams. However, they remember that course grades may carry some assessor subjectivity. The article also exposes the subjective measurements like attitudes, satisfaction, and evaluation of instruction factors. But as mentioned by Ni (2013) it is good to have in mind that as learning effectiveness is a complex concept with multiple dimensions so it should be assessed with multiple measures.

Moreover, Ni (2013) argues that the social communicative interactions between students and teacher and student and student are fundamental for the learning process. A student's ability to ask a question, to share an opinion, or to disagree with a point of view are important learning activities. The author claims that it is often through conversation, discourse, discussion, and debate among students and between instructors and students that a new concept is clarified, an old assumption is challenged, a skill is practiced, an original idea is formed and encouraged, and ultimately, a learning objective is achieved. The article explains that online learning requires adjustments by instructors as well as students for successful interactions to occur. Online courses try to substitute classroom interaction with discussion boards, synchronous chat, electronic bulletin boards, and e-mails. But the effectiveness of such a virtual interactive venue is not without debate. However, the advantage of online interaction may not be realized if close connection among the learners is absent (Ni, 2013). Although the proponents of online learning have consistently noted that it is not course modality but course quality that influences student learning, and the authors agree with this assertion, but the results suggest that designing and teaching high-quality online courses with sufficient student supports may be more difficult than doing so with face-to-face courses (Xu; Jaggars, 2011).

Yet, achieving apprenticeship require a greater responsibility from the students according to Xu and Jaggars (2014), thus, a successful online student may need high levels of self-regulation, self discipline, and a related suite of meta cognitive skills, which often fall under the broad rubric of self-directed learning. But the article also discuss the obstacles that distance learners have to overcome because students in online courses may experience higher levels of dissatisfaction, interpersonal isolation, feelings of unclear direction and uncertainty, and a lack of engagement in the learning process. It is also good to know that online course is not explicitly designed to help students develop the skills they need to succeed in this new context. Accordingly, some students may struggle to perform as well in an online course as they would in a similar face-to-face course (Xu; Jaggars, 2014).

In this sense, Harmon, Lambrinos et al. (2012) explore the disadvantages of online learning like isolation and lack of media richness. Isolation encompasses insufficient peer and instructor interaction, which leads to low engagement and lower course completion rates. They suggest that media richness encompasses the attributes of face-to-face communication of verbal speech intonations and nonverbal facial and body expressions.

Still, we also have to look to the advantages of the online format like a self-paced format, convenience in viewing and re-viewing lectures as mentioned by Harmon, Lambrinos et al. (2012). Besides online classes can provide faster information access and also encourages the students to take a more responsible, constructive role in the learning process as can be seen in Gratton-Lavoie and Stanley (2009).

Considering the advantages and disadvantages of online format, Harmon, Lambrinos et al. (2012) suggest maybe it is a good idea mixing online lectures with traditional lectures in the same course to promote equivalence between "digital"and "live"communication. Thus, a rotating format in the traditional lectures could potentially marry the advantages of the two modalities.

The empirical evidences about this topic is very controversial, with a lot of approaches applied, there are studies defending a positive impact of distance learning, others with no effect and others with negative impact. Here we will focus in papers that take into account the selection bias problem and evaluate the effect of distance education for the entire undergraduate course.

In Xu and Jaggars (2013), in which the authors use instrumental variable technique to estimate the impact of online versus face-to-face course delivery on student course performance controlling for the self selection. They use the travel distance between each student's home and college campus served as an instrument for the likelihood of enrolling in an online section of a given course. In addition, college-by-course fixed effects controlled for within- and between-course selection bias. Analyses yield robust negative estimates for online learning in terms of both course persistence and course grade. For Brazil we have MARQUES (2020) that adopt propensity score techniques for controlling for the selection bias problem. The author seeks to assess the impacts of distance higher education and the results show that it is possible to rule out the possibility of positive academic impacts, considering negative or null results observed.

In our paper, after applying the propensity score matching we found a negative and significant impact of distance education ENADES' score. The effect is heterogeneous among undergraduate courses and varies from around 0.5 standard deviation in Social Work, Physical Education Degree and Bachelor to a little more than 0.1 standard deviation in History and Production Engineering.

About our results, it is also important to highlight that the impact is higher for the specific component than the general component, making the situation even more worrying, because the grades are worst in the final content of the undergraduate course. In the specific component the ATT varies around 0.5 standard deviation in Social Work, Technology in Environmental Management and Physical Education Degree and Bachelor, to a bit more than 0.1 standard deviation in History.

Once we have analysed our results it is important to discuss the reasons why distance learners performance is worst than in person students and Ni (2013) brings some light on the subject. The author says that the interaction between student and teacher and among students is essential for the learning process, and online courses require adjustments to promote this kind of integration. And maybe could be very difficult and expensive to create a design in distance education with a good communication among the people involved in the process, as Xu and Jaggars (2013) says online courses with high quality can demand huge investments.

So one possibility behind the negative impact is the peer effect. Winston and Zimmerman (2004) says that peer effect occurs when a person's behavior is affected by his or her interaction with peers. As communication among peers is important to the learn process, and as in distance education this effect is absent or hard to establish, it can be related to the worst performance in distance learners comparing with in-person students. The authors discuss that peer effect has an important role in academic achievements, they conclude that the peer effect exist and is strong among students. The impact of strong students tend to increase peers academic performance and weak students tend to reduce it.

Another interesting debate in this area is the effect of quality in higher education in the outcomes in labor market. Some articles that estimate this relation and control for selection bias are Dale and Krueger (2002), Heckman, Stixrud and Urzua (2006) and Saavedra (2009). Using parameters for outcome in labor market as employable, work in formal sector and earnings, those papers found a significant and positive impact of quality in education in the outcomes in labor market. Besides, according Heckman, Stixrud and Urzua (2006) those effects are over the life cycle in labor force (Heckman; Stixrud; Urzua, 2006). Even though the return fall when controlling for bias selection, the results are still significant (Dale; Krueger, 2002). The effect of quality in education is higher for the ones with disadvantaged background families (Dale; Krueger, 2002; Saavedra, 2009).

Particularly about the undergraduate courses related with professionals in the education sector, the debate of the quality in college degree is even more complex because the impact of a bad worker's training will reflect on the quality of education of children and teenagers from elementary till high school and the damage may be irreversible. This discussion is important here, because we analyse the performance in Pedagogy, History and Physical Education Degree. There are many researchers trying to investigate the role of teacher's quality in students outcomes as Darling-Hammond (2000) and CARVALHO (2018).

The authors in those studies argue that investments in quality and training of the teachers have a positive and significant improvement on the scholar outcomes of the students, despite of the fact that the intensity of the effect is lower when they control for social economic characteristics (CARVALHO, 2018). Also in CARVALHO (2018), the author says that this relationship is little studied in Brazil, but the few studies about this topic reinforce the role of teacher quality as a very important input for the learning process.

Finally, we can highlight some contributions of this paper. Our intention is to contribute with the evaluation of the quality of distance education in higher education and the heterogeneity among the undergraduate courses controlling for the bias selection. After the pandemic of COVID-19, many systems of education and work are migrating to remote or hybrid format, and it is indispensable to analyse each case before taking a decision. This kind of research can be the start to think about solutions to improve distance education modality to be closer in quality with the in-person modality. Lastly, it is essential to design policies for the future of distance education, once it has a potential impact in the outcomes in labor market and also in the quality of education professionals and scholar results.

Besides that introduction. The section 2 of that and methods brings information about the database used, a descriptive analysis of data the econometric approaches applied. Section 3 we present the results and the considerations about the models estimated. In the end in section 4 we have our final considerations.

4.2 Data and Methods

This section contains information about strategies and methods used to achieve the main objective of the paper which is to assess the impact of distance education (treatment group) versus the traditional face-to-face (control group) modality in the performance of the ENADE score of higher education. The undergraduates and years for analysis are: Pedagogy Degree, History Degree and Physical Education Degree courses in 2017, Business, Accounting and Social Work in 2018 and Technology in Environmental Management, Physical Education Bachelor and Production Engineering in 2019. For this purpose, the section includes the presentation of the database, description of the techniques adopted for the treatment of the database, definition of the model variables, descriptive analysis of the data and exposure of the models and methods applied.

4.2.1 Database and variables

Here we present information about the database and variables used. It is intended, therefore, to describe the strategies used for the treatment of the database and to discuss the choice of the dependent and control variables. The database is from ENADE, provided by INEP (National Institute of Educational Studies and Research Anísio Teixeira) and contains data related to the exam results, the Educational Institution and the student. In addition, there are also data about information regarding the General Course Index that are added to the ENADE database, where such data are also provided by INEP.

In this work we analyzed the effect of distance learning modality versus the traditional face-to-face modality in the performance of higher education students. So we estimated 36 models, four per course, using the method of propensity score matching, with three dependent variables proposed that reflect the performance of students in higher education: the general component score (value from 0 to 100), the specific component score (value from 0 to 100) and the final score (value from 0 to 100) which corresponds to the weighted average of the general component score (25%) and the specific component score (75%). But in this work we use the ENADE's Score in terms of standard deviation, which is calculated by dividing the grade by the standard deviation. In addition, for each course, we applied the technique of nearest neighbor in propensity score matching and we will also apply the inverse probability weighting algorithm, with replacement and only for observations that are within the common support.

The treatment variable and the other control variables are shown in Table 14. The variables indicate socioeconomic characteristics of individuals and also attributes of educational institutions that interfere in the choice of learning modality and also in the performance of students.

About the covariates we salient that, although the propensity score method demand to have the value of the variables before treatment, in this paper the covariates values are after the treatment. We use those variables because they are very important to the attribution to treatment and also because those variables potentially carry information before treatment.

4.2.2 Descriptive analysis

This section presents a descriptive analysis of the dependent variables related to the performance of students at ENADE in order to contextualize the difference in performance in both modalities in rough terms, without controlling for the other covariates.

The Table 15 shows the total number of observations, Universities and courses by modality and undergraduate course. We have a total of 465311 students in all undergraduate courses analyzed, of which 311017 are in traditional face-to-face and 154294 are distance

Table 14 – Covariates

Variables	**Description**
Age	Individual's age
Man	Dummy equals to 1 if the individual is a man.
White	Dummy equals to 1 if the individual is white or Asian. The other group is composed by black, brown and indigenous students.
Per Capita Income	Individual's per capita income. The per capita income variable was created by dividing the midpoint of the income ranges (up to 1.5 minimum wages, between 1.5 and 3, between 3 and 4.5, between 4.5 and 6, between 6 and 10, between 10 and 30 and over 30) in which the average point used for the category over 30 was 30 minimum wages divided by the number of family members in which the categories of members in the household are 1 individual, 2, 3 , 4, 5, 6, 7 or more, in category 7 or more the value of 7 was considered. And the values are updated to the year of 2019 through the IPCA calculator
Work	Dummy equals to 1 if the individual works.
Public School	Dummy equals to 1 if the individual studied high school in the public school. The dummy is equal to 1 for individuals who studied at least most part of High School in public school. The individuals who studied at least the most part of the the High School abroad were excluded.
Parent's Education	Variable that assumes the level of education of the parent with the highest schooling, in which the categories are no level completed, Elementary School, Middle School, High School, Higher Education or more.
Home	Situation in home, in which the categories are alone, with parents and/or relatives, with children and/or spouse and others.
State	Country's state.
Private University	Dummy equals to 1 if the individual studies the Higher Education in private University.
Scolarship	Dummy equal to 1 if the individual received some type of academic scholarship.
Quota	Dummy equal to 1 if the individual entered the university by quota policy
IGC	University Score

learning. In total we have a 6585 courses offered, of which 6223 are face-to-face and 362 in distance learning. In addition, there are 1632 Universities, in which 1626 offers at least one of the courses in the face-to-face modality and 131 offers at least one of the courses in the distance-learning modality. Finally, it is worth remembering that these are all observations from the database. In the rest of the descriptive analysis, as well as in the regression models, there may be fewer observations due to missing variables.

4.2.2.1 Dependent variables

Here is presented a descriptive analysis of the dependent variables established: the final score, the general component score and the specific component score, in which all of

Table 15 – Number of observations, Universities and courses by modality and undergraduate course.

	Observations			Courses			Universities		
	FTF	**Dist**	**Tot**	**FTF**	**Dist**	**Tot**	**FTF**	**Dist**	**Tot**
Pedagogy	63583	69171	132754	1116	96	1212	829	95	847
History	12994	5537	18531	306	35	341	223	34	237
Phy. Ed. De.	29409	5354	34763	540	13	553	430	13	432
Business	91253	29152	120405	1696	68	1764	1303	68	1306
Accounting	46852	15623	62475	1052	48	1100	850	48	855
Social Work	12474	12151	24625	306	24	330	267	24	275
Tec. Env. Man	1571	3533	5104	80	44	124	68	44	105
Phy. Ed. Ba.	27361	11544	38905	497	11	508	417	11	420
Prod. Eng.	25520	2229	27749	630	23	653	505	23	509

Source: Made from ENADE 2017, 2018 and 2019 microdata, INEP.

the grades are in terms of standard deviation, of students of Pedagogy Degree, History Degree and Physical Education Degree in 2017, Business, Accounting and Social Work in 2018 and Technology in Environmental Management, Physical Education Bachelor and Production Engineering in 2019.

In Figure 13 is presented the difference in standard deviation of the ENADE's score, by undergraduate course and type of score. It can be seen from the Figure that, for all courses and type of grade, the difference is positive, which means that the performance in ENADE for in person students is higher than for distance learners, and this difference is significant.

Particularly, regarding the final score in Figure 13, the difference in standard deviation varies from 0.25 points in History to 0.8 points in Social Work. It is also worth mentioning that most part of the dissimilarity between the modalities comes from the specific component score as we can see in Figure 13. This fact makes the gap between learning methods even more worrying, considering that the learning of specific contents is the main objective of undergraduate courses. With regard to the specific component score, the gap varies from 0.26 standard deviation in History to 0.84 points in Social Work. But the gap showed on the Figure below is without controlling for other covariates. So in the Results' section, we will see if the difference remain or change after applying the propensity score matching.

4.2.3 Methodology

This section contains a description of the methodology used to estimate the parameters and capture the effect of distance learning modality in relation to face-to-face modality in the performance of ENADE.

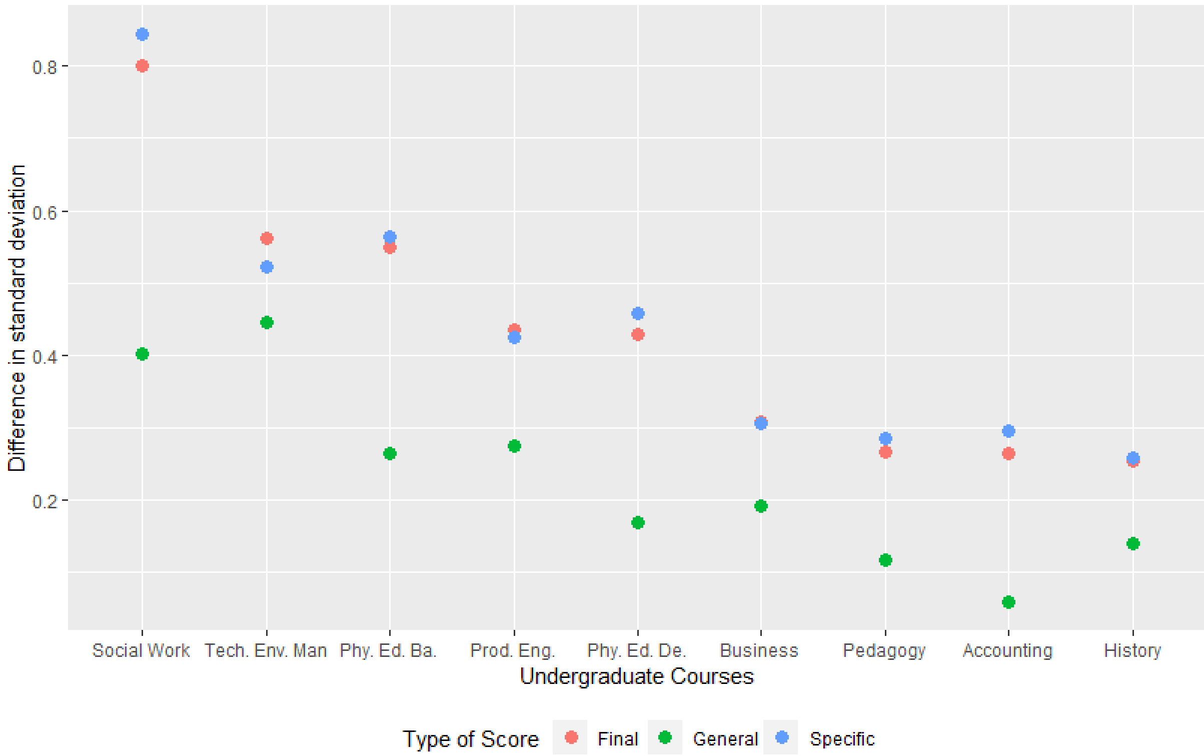

Figure 13 – Enade Score Difference in Standard Deviation by Undergraduate Course and Type of Score

Source: Made from ENADE 2017, 2018 and 2019 microdata, INEP.

4.2.3.1 Counterfactual Model

Based on Stephen and Christopher (2007) here we will describe the counterfactual model and its relation to our problem. The concept of counterfactual causality is based on the assumption that causality takes the form of a comparison between the result of a unit when that unit is treated and the result of the same unit when it is not treated. More specifically, for this study, the idea would be to compare the result of an individual in the face-to-face modality in the performance at ENADE with the result of the same individual if the distance education modality is performed.

The calculation of the treatment effect in the counterfactual model is based on the hypothesis of Stable Unit Treatment Value Assumption (SUTVA): a priori assumption that the value of Y for unit u when exposed to treatment t will be the same, regardless of the mechanism used to assign treatment t to unit u and regardless of treatments other units receive. In the case of the present study, the hypothesis of SUTVA is to consider that the performance of distance learning or face-to-face students does not depend on the number of individuals involved in each of the modalities.

In this study we are interested in the average effect of treatment on treaties (ATT). ATT would be the expected difference in performance if we could educate a student randomly selected from the distance learning system in both distance learning and face-to-face modality. Where the ATT equation is given by:

$$ATT = E(\delta|D = 1) = E(Y^1 - Y^0|D = 1) = E(Y^1|D = 1) - E(Y^0|D = 1) \tag{4.1}$$

Where Y is the performance at ENADE, D is the treatment variable and assumes a value of 1 when the individual performs the distance learning modality and assumes a value of 0 when performing the face-to-face modality.

A empirical problem of equation 4.1 is that we do not observe $E(Y^0|D = 1)$, that is, we did not observe the potential result that an student from distance learning would have if he performed the face-to-face modality. What we observe is $E(Y^0|D = 0)$ which is the performance of a face-to-face student, but if $E(Y^0|D = 1) \neq E(Y^0|D = 0)$ our ATT estimator will be inconsistent. An alternative to the problem is to use the propensity score matching method that is the purpose of this study.

4.2.3.2 Propensity Score Matching

Based on Caliendo and Kopeinig (2008) this section contains a description of the propensity score matching method and its application for our purpose. The idea of matching is to find a valid counterfactual for each individual treated, assuming that only observable characteristics are responsible for making treaties and controls "different". In this way, the control group would indicate the potential result of the treatment group if they had not been treated, that is, the only factor that differentiates the results of these individuals is the participation or not in the treatment. For this study, the matching technique will be based on the propensity score given by:

$$P(x) = Pr[D = 1|X = x] \tag{4.2}$$

Where the estimation of propensity in this work is done by logit.

Two hypotheses are necessary for the consistency of the method. The first one is the conditional independence hypothesis (CIA) which establishes that conditional on X, the attribution to treatment is independent of potential products -$(Y^0, Y^1) \perp D|X$-, that is, once we control by X (variables that are not affected by the treatment), the selection or assignment to the treatment is said to be random. In the case of propensity score matching we have that conditional on X, the attribution to treatment is independent of the potential products -$(Y^0, Y^1) \perp D|X$- we also have that conditional in $P(x)$, the attribution to treatment is independent of the potential products -$(Y^0, Y^1) \perp D|P(x)$. But for the ATT estimator we can relax the hypothesis to $Y^0 \perp D|P(x)$.

$$E(Y^0|P(x), D = 1) = E(Y^0|P(x), D = 0) = E(Y^0|P(x)) \tag{4.3}$$

Replacing 4.3 in 4.1 follows that:

$$ATT = E(Y^1|P(x), D = 1) - E(Y^0|P(x), D = 0) \tag{4.4}$$

The other hypothesis is a common support given by:

$$0 < Pr[D = 1|X] < 1 \tag{4.5}$$

Thus, for each value of X there are both treated and untreated. But for ATT we can relax the hypothesis to:

$$Pr[D = 1|x] < 1 \tag{4.6}$$

This hypothesis is necessary because if it is violated and $Pr[D = 1|X] = 1$, it will not be possible to define $E(Y^0|P(x), D = 0)$ in 4.4, because for a given $P(x)$ there will be no one in the control group. Finally, the combination of the conditional independence hypothesis with the common support hypothesis is known as strong ignorability.

For this study, the propensity score algorithm is the nearest neighbor with replacement and for observations in common support, that is, each treated i is matching with an individual from the untreated group that has the $p(x)$ closest to its own, and the same individual of control group can be use as pair for more than one individual in treatment group.

Another important step is to verify the overlap and the common support region between the treatment (distance learning modality) and the control group (face-to-face modality). Thus, it is necessary to observe whether for the propensity score values there are individuals in both the treatment group and the control group.

Furthermore, it is also important to assess the matching quality. As the conditioning is done in the propensity score and not in all covariates x, it is necessary to check if after the matching procedure the distribution of the relevant variables in both groups are similar. The basic idea is to compare the situation before and after matching and check if there is any difference after conditioning in the propensity score. If it exists, it is a sign that the propensity score equation is poorly specified. So we have to test:

$$x \perp D|p(D = 1) \tag{4.7}$$

Where the matching quality is measured by performing the t-test (Rosenbaum and Rubin (1985). In this case, before matching differences are expected and after matching the covariates must be similar between the groups and, therefore, no difference significant must be found. Once the sample is balanced, we can calculate the ATT, wich can be given by:

$$\Delta^M = \frac{1}{N_t} \sum_{i \in \{D=1\}} [y_i^1 - \sum_j w_{ij} y^0] \; where \; 0 < w_{ij} \leq 1 \; and \; \{D = 1\} \; treated \tag{4.8}$$

Where y_i^1 is the ENADE score from individual i in treatment group and w_i is the score weight of individual j at control group that is the counterfactual of individual i and y_j^0 is the performance in ENADE of individual j at the control group.

Finally, to check the robustness of ATT it is important to apply the sensitivity analysis. The description below is based on Cerulli (2015). The central idea of sensitivity analysis is to verify if the results of ATT would be different if we relax the CIA hypothesis. If the selection to treatment is duo also to a non-observable variable then two matched units should not have the same probability to be treated although balanced on observable variables.

By assuming a logistic distribution, two matching units i and j, having $x_i = x_j$, have the following odds ratio:

$$\frac{\frac{p_i}{1-p_i}}{\frac{p_j}{1-p_j}} = \frac{p_i(1-p_j)}{p_j(1-p_i)} = \frac{exp(x_i\beta + \gamma v_i)}{exp(x_j\beta + \gamma v_j)} = exp(\gamma(v_i - v_j)) \tag{4.9}$$

Showing that, as soon as $v_i \neq v_j$, the two probabilities to be treated are different, actual balancing does not hold and a hidden bias arises. Suppose that v_i and v_j take values in the interval $[0; 1]$. This implies that $-1 \leq v_i - v_j \leq 1$, so that the odds ratio is in turn bounded this way:

$$\frac{1}{e^\gamma} = \frac{p_i(1-p_j)}{p_j(1-p_i)} = e^\gamma \tag{4.10}$$

In which odds ratio are equal only when $\gamma = 0$ when no hidden bias is present because unobservables have no effect on selection. Thus, given a positive value of γ, we can depict a situation in which the odds ratio is maximum (the best case) and one in which it is minimum (the worst case). This reflects the uncertainty due to the presence of an unobservable confounder. By putting $\Gamma = e^\gamma$, we can also say that in the presence of a potential hidden bias, one unit has an odds of treatment that is up to $\Gamma \geq 1$ times greater than the odds of another unit. When randomization is allowed, however, the odds ratio is equal to one by definition and $\Gamma = 1$.

4.2.3.3 Inverse Probability Weighting Estimator

We will use the reference of StataCorp (2017) to explain the algorithm to calculate the inverse probability weighting estimator (IPW). IPW estimators use weighted averages of the observed outcome variable to estimate means of the potential outcomes. The weights account for the missing data inherent in the potential-outcome framework. Each weight is the inverse of the estimated probability that an individual receives a treatment level. Outcomes of individuals who receive a likely treatment get a weight close to one. Outcomes of individuals who receive an unlikely treatment get a weight larger than one, potentially much larger.

In other words, by giving greater weights to observations that are less likely to be treated (i.e., those that are "rare"in the treatment group), you are giving those observations more importance when estimating ATT. This helps ensure that the treatment effect is more representative of actual conditions, even for individuals who are least likely to be treated.

To see the intuition behind IPW, consider a study with observed outcome variable y, treatment variable $t \in \{0, 1\}$, and potential outcomes y_0 and y_1. As part of this process, we need to estimate the POM for treatment $t = 1$, $E(y_1)$. Using the observed data, $y_i t_i$ is y_{1i} when $t = 1$, but y_{1i} is unobserved when $t = 0$. An IPW estimator is: for $E(y_1)$ is:

$$E(y_1) = \frac{1}{N} \sum_{i=1}^{N} \frac{y_i t_i}{p(x_i)} \tag{4.11}$$

Where $p(x_i)$ is the probability that $t_i = 1$ and is a function of the covariates x_i. If y_{1i} were always observed, the weights would all equal 1. This IPW estimator places a larger weight on those observations for which y_{1i} is observed even though its observation was not likely.

And to estimate the Average Treatment Effect on the Treated (ATT), the algorithm calculate the difference in means of both groups in the outcome variable and considering the weights estimated by IPW.

4.3 Results

This section brings the estimations evaluate the effect of the distance learning modality on the performance of the ENADE score in comparison with in person students in College Degree. So we applied the method of propensity score matching and inverse probability weighting previously described.

4.3.1 Matching Quality

Here we evaluate the quality of the selection model and also about the quality of matching, comparing the treatment and control group before and after matching. As we already described the determinants of the selection model in the first paper, we do not analyse the coefficients of Logit here, but the complete regression tables are exposed in Appendix C in Tables 26, 27 and 28.

Furthermore in Appendix D in Table 29 we present some parameters about the forecasting ability of the logit model by undergraduate courses. In general the value of the parameters means that the model has great capacity of selection and it is important to highlight that the cut value is 50%, so for values of the propensity score higher than 50%, the model classify as treated, and for values smaller than 50% the model classify as untreated. About the correctly classification parameter, the values are around 80% and 90%, which implies a good forecasting ability and not disrespect the common support condition, in other words, for any value of X there are treated and untreated. The other parameters, in general, also indicate a good capacity of classification, the values for truly positives and negatives are high and the values for false positives and negatives are low.

About the quality of the matching in Figures 14, 15 and 16 we present the distribution of propensity score before and after matching by undergraduate course and modality. In those Figures we can see that before matching, the distribution of propensity score is very different between distance learning and face-to-face approach. For distance learning method there is more concentration in the higher values of propensity score and for face-to-face modality there is more concentration in the lower values of propensity score.

But we can observe that after matching the distribution of propensity score between distance education and in person approaches is very similar. Although, to have a similar distribution of propensity score after matching does not necessarily mean we have a good balance in each covariate between both groups. So it is still very important to make the balance analysis of the control variable after matching process.

So the next step, a crucial one, is to check if after the matching procedure, the sample of treatment and control groups are balanced, in other words, it is important to check if they are similar on the covariates after matching both groups. In Appendix E are presented the complete tables (Tables 30, 31 and 32) of the T-test of the control variables of the unmatched and matched sample by undergraduate course, in 2017, 2018 and 2019. Here in this section is showed in Figure 17 the graph of standardized % bias across covariates of the unmatched and matched sample by undergraduate course and year, in which are not presented the variables of state to facilitate the visualization.

In the Figure17 we can see that, in general, distance learners and in person students

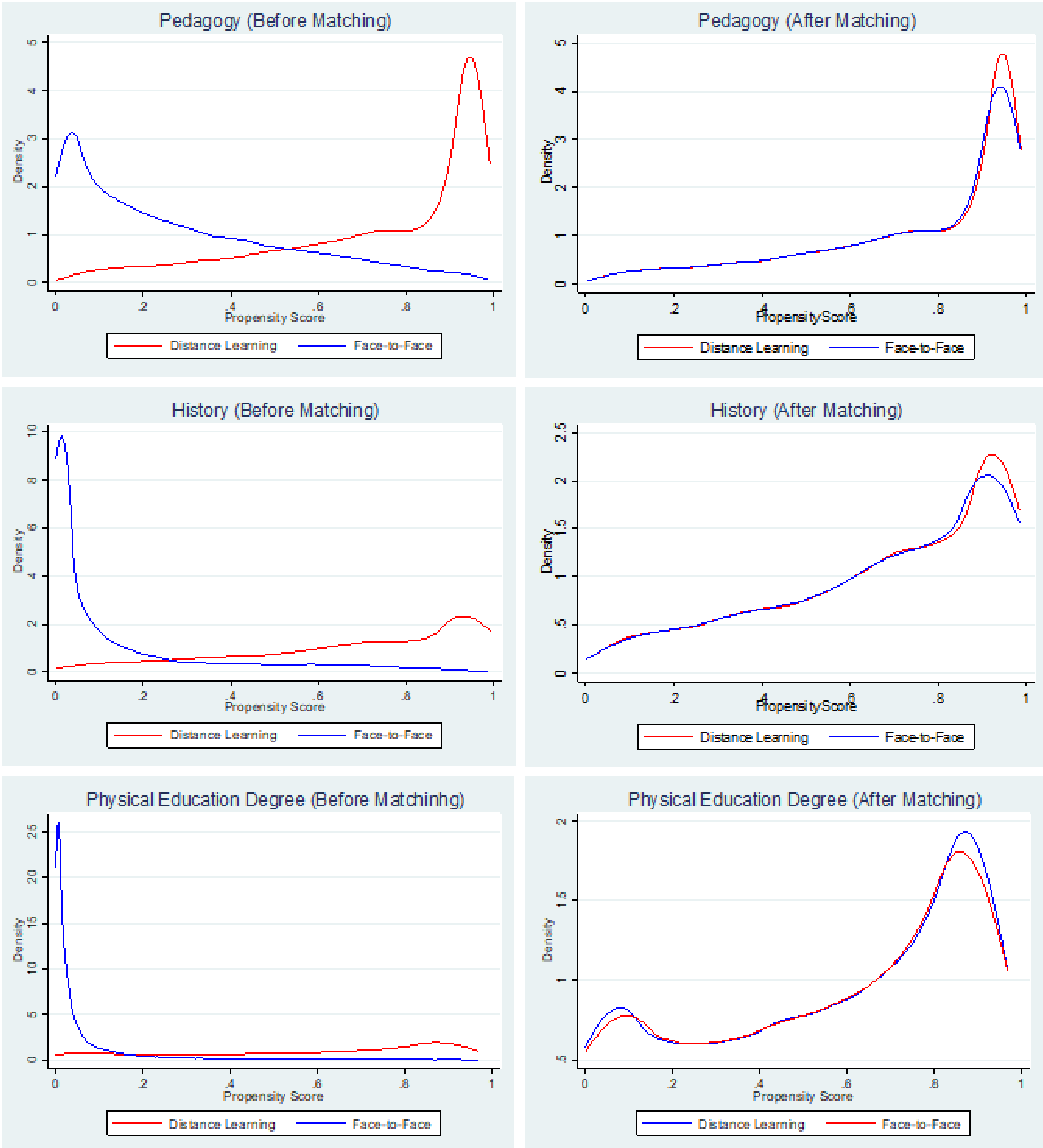

Figure 14 – Distributions of Propensity Score Before and After Matching by Undergraduate Course and Modality in 2017,

Source: Made from ENADE 2017 microdata, INEP.

were quite different, before matching. But, after matching, we can note a reduction of the standardized value % bias on the covariates, which means that treatment and control groups became more similar. Visually, the courses of History, Business, Accounting and Production Engineering, are the ones with bigger similarity between distance learners and in-person students after matching. On the other hand, the undergraduate courses of Pedagogy, Social Work, Physical Education Degree and Bachelor are the ones with bigger dissimilarity between both groups after matching.

Furthermore, after a closer look in Tables 30, 31 and 32 in Appendix E, although the p-value indicates a persistence of the difference in the means between treatment and control, it is noticed that, with some exceptions, the size of these differences are small,

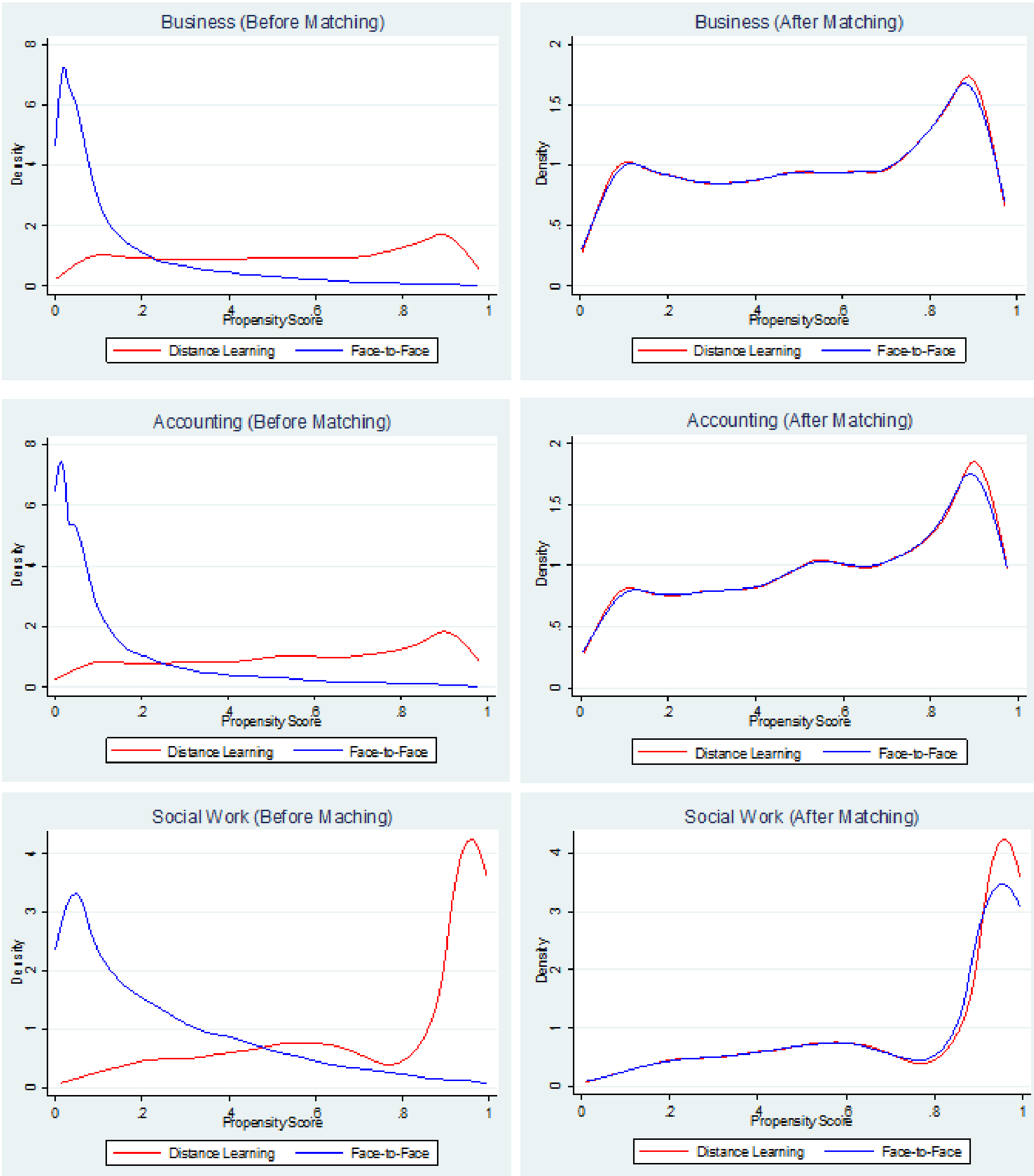

Figure 15 – Distribution of Propensity Score Before and After Matching by Undergraduate Course and Modality in 2018.

Source: Made from ENADE 2018 microdata, INEP.

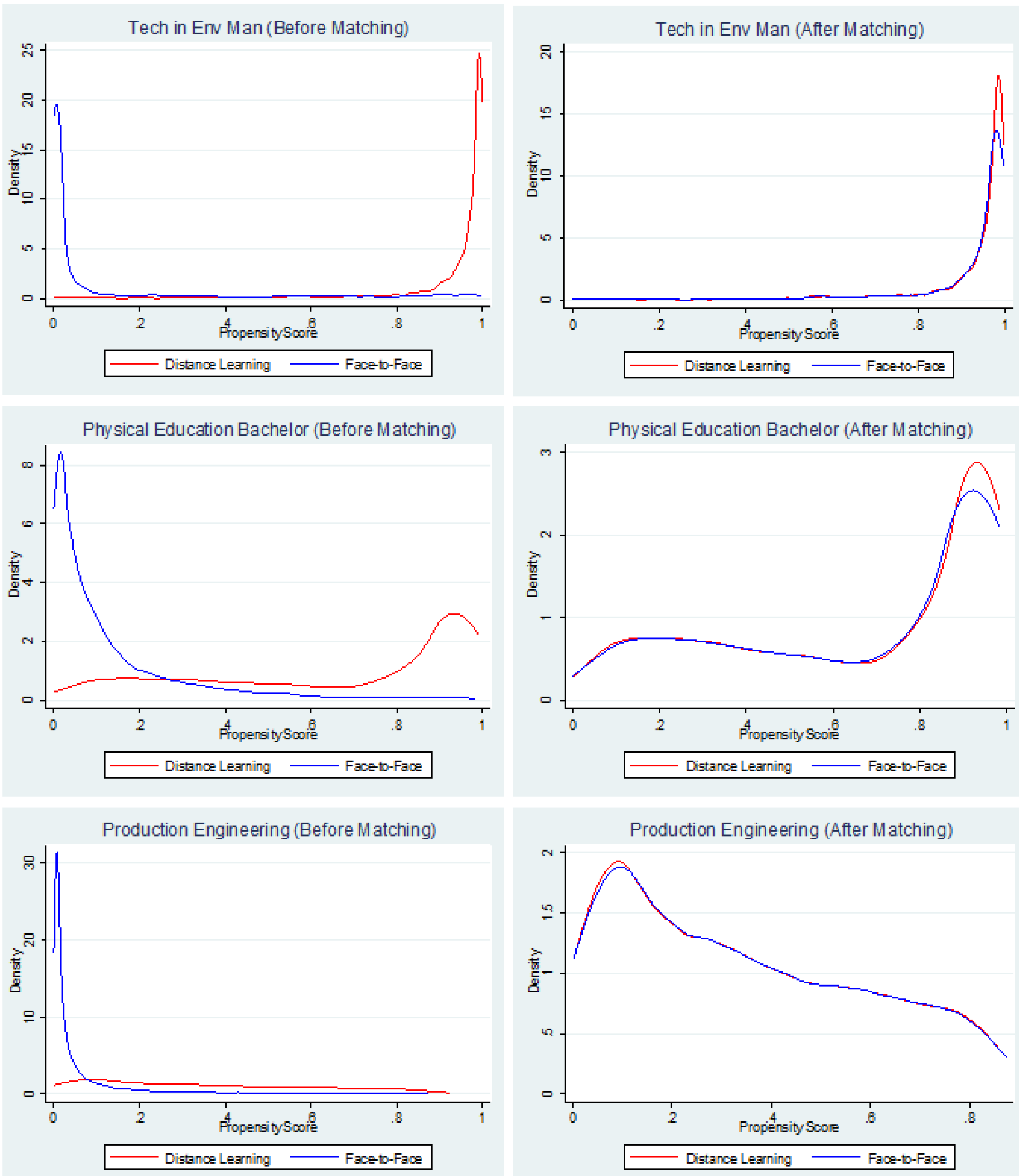

Figure 16 – Distribution of Propensity Score Before and After Matching by Undergraduate Course and Modality in 2019.

Source: Made from ENADE 2019 microdata, INEP.

that is , after matching, the means between control treatments have very close values for covariates. Most importantly, this conclusion remains for the most important treatment attribution variables that also could be related with the performance, like age, work, position in family, if the high school was concluded in public system, academic scholarship and if the university is private.

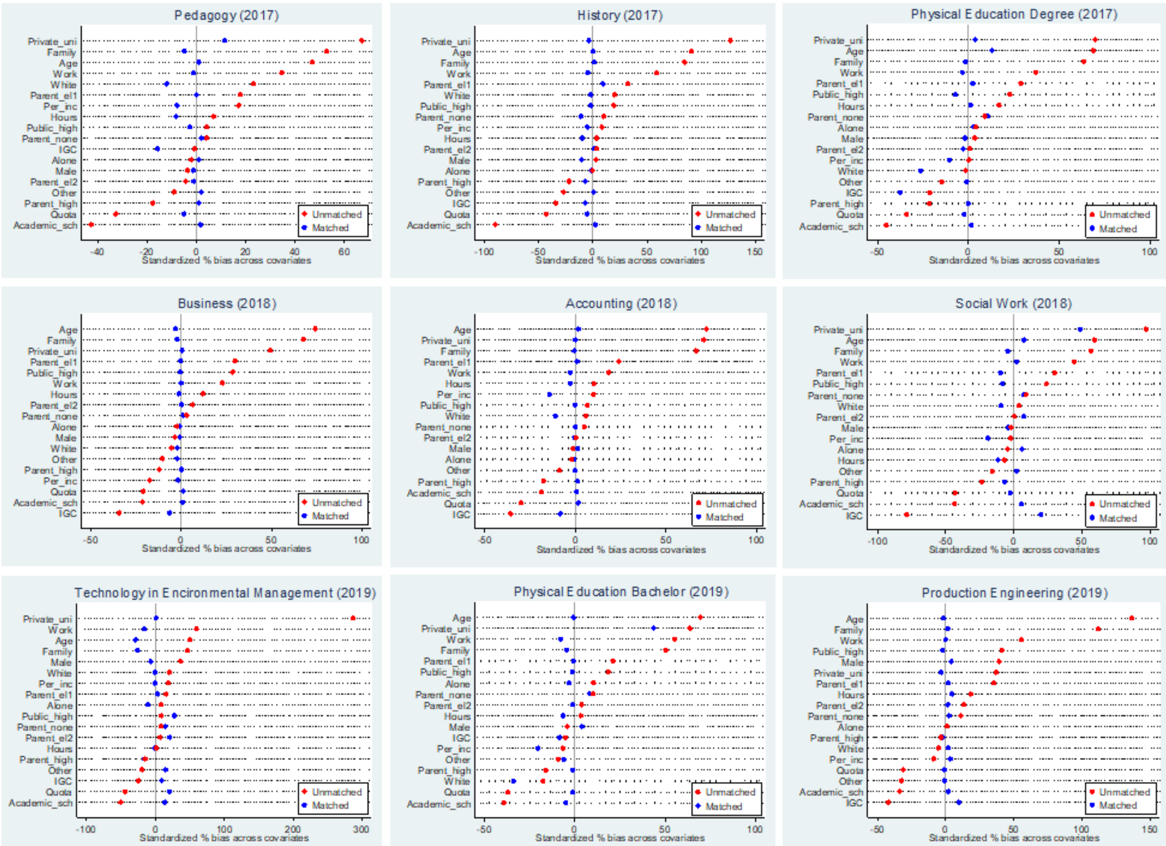

Figure 17 – Graph of standardized % bias across covariates of the unmatched and matched sample by undergraduate course and year.

Source: Made from ENADE 2017, 2018 and 2019 microdata, INEP.

4.3.2 Average Effect of Treatment for Treaties

This section presents the results of the propensity score matching model. More specifically, it is noteworthy that the model was estimated using the nearest neighbor and inverse probability weighting algorithm, with replacement, only for individuals within the common support. About the explanatory variable of performance, here is used the three types of ENADE score: final score; general component and specific component. It is also important to mention that the performance variables are in terms of standard deviation. In Appendix F, in Table 33, we also present the number of observations on and off common support by undergraduate course. In this table, it is important to highlight that we have just a few observations off common support, not more than 1% of total observations, and most important we do not lose many observations off common support among the distance learners (not more than 3% of the sample). The exception is for Technology in Environmental Management, in which about 19% of total observations are off common support, and almost 30% of the treated.

So in Table 16 are showed the results of the average effect of the treatment on treated by undergraduate course and score's type of ENADE. In this table we applied two types of algorithm, the nearest neighbor (for all score's type) with replacement and for observations on the common support and the inverse probability weight (just for final score) for observations on the common support. In general, we can see that the coefficient of ATT is negative and significant for almost every undergraduate course, type of ENADE score and type of algorithm. Those results indicate that performing the distance learning modality on average decreases student performance in comparison with in-person approach.

Looking at the final score in the first column we can note that the negative level of impact of distance education varies among the undergraduate courses. The effect is greater for Social Work, Physical Education Degree and Bachelor in the values of 0.49, 0.44 and 0.54 standard deviation respectively. On the other hand, the influence of the modality is lower for Production Engineering and History in 0.13 and 0.12 standard deviation respectively. It is also important to mention that the values of ATT in final score between nearest neighbor and inverse probability weight are very similar, with exception of Technology in Environmental Management. This fact is important to reinforce the robustness of the results.

Another interesting fact is that if we compare the gap of performance between online students and face-to-face students in ENADE before the matching process (Figure 13) and after matching (Table 16) we can conclude that the difference decreases or it remains almost the same. As we can see for Pedagogy and Physical Education Degree and Bachelor, for example, where the gap decreases. But for the other undergraduate course the gap remains almost the same, as for the other undergraduate courses. So, when we do not control for any variable, the difference is equal or underestimated (the gap is rather negative), in comparison to when we control for the variables. This fact can reinforce the hypothesis that even if we have a problem of selection bias, the problem is probably a negative selection bias, in other words, duo to unobservable variables the distance learners would have a lower performance than the in-person students even if both modalities had the same quality. But in the next section we will conduct a sensitivity analysis, to see if even with the problem of selection bias we can keep our conclusion about the ATT.

It is also important to highlight, looking at the Table 16, that the dissimilarity is higher in specific component than in general component. So the most part of heterogeneity in performance comes from the final content of undergraduate courses and makes the dissimilarity of performance between the modalities even more worrying, because it refers to the final content of the undergraduate course. Regarding the ATT for the specific component and the higher impact, we have Social Work, Technology in Environmental Management and Physical Education Degree and Bachelor, in which the values of ATT in standard deviation are 0.46, 0.48, 0.45 and 0.54 respectively. The lowest coefficient is for History of ATT of -0.11.

In addition we can take some conclusions comparing our results with some facts in the Literature. Firstly, our results are in agreement with the part of the Literature, in which the bias selection is a concern in the analysis and the distance education has a negative impact on the grades. And about, that the theory explains some disadvantages of distance learning like the necessity of self regulatory and self discipline from the student as in Xu and Jaggars (2014), the isolation and the lack of interaction that encompass the insufficient interaction with the peer and the instructor as stated by Harmon, Lambrinos et al. (2012). So those disavantages can impact negatively on the scores. According to Ni (2013), the interaction between student and professor and among students is fundamental in the learning process, and the distance education approach try to offer tools to promote that interaction. But maybe online coursers with great tools and quality could be very expensive (Xu; Jaggars, 2013).

In this sense the peer effect might be behind the negative effect of distance education in grades since communication among peers is important to the learning process, and as in distance education this effect is absent or hard to establish. As mentioned by Winston and Zimmerman (2004) the peer effect exist and is strong among students. The impact of strong students tend to increase peers academic performance and weak students tend to reduce it.

Moreover, those results can be very worrying if the lower quality in distance reflect in the quality of the work. Since online programs are in expansion and will train people for the labor market, the productivity of the workers could be impacted negatively in the long run. Some articles as Dale and Krueger (2002), Heckman, Stixrud and Urzua (2006) and Saavedra (2009) estimated this relation and found a positive impact of quality in education and in the outcomes in labor market.

Particularly about the expansion of distance education undergraduate courses related to the education are as Pedagogy, History and Physical Education Degree, it is important to think that those professionals will be responsible for the education of great part of the children in the country. So we can have a irreparable damage in education for generations. Darling-Hammond (2000) and CARVALHO (2018) inferred that investments in quality and training of the teachers have a positive and significant improvement on the scholar outcomes of the students

On the other hand, considering the advantages of distance education like the schedule flexibility (Xu; Jaggars, 2011), and the lower costs of education, where the benefits can be a good strategy to enlarge the access to the higher level and qualify the workers in the labor market. So as those analysis are a very complex topic, it can be a good option to focus on the policy of expansion of distance education for the ones who would not go to college if it were not for the distance modality. Also a good possibility could be the hybrid method alternative to take advantages of both approaches.

Table 16 – Average Treatment for Treaties by Undergraduate Course and Score Type of ENADE.

Course	**ATT (Distance Learning)**				**Obs**
	Nearest Neighbor			**IPW**	
	Final	**General**	**Specific**	**Final**	
P. Ed. BA.	-0.54***	-0.28***	-0.55***	-0.54***	30861
	(0.052)	(0.053)	(0.052)	(0.039)	
S. Work	-0.49***	-0.22***	-0.53***	-0.46***	20875
	(0.068)	(0.076)	(0.067)	(0.046)	
P. Ed. DE.	-0.44***	-0.14***	-0.48***	-0.45***	26614
	(0.044)	(0.046)	(0.043)	(0.035)	
T. Env. Man.	-0.30*	-0.12	-0.37**	-0.48*	3269
	(0.172)	(0.164)	(0.175)	(0.245)	
Pedagogy	-0.26***	-0.10***	-0.39***	-0.29***	108001
	(0.022)	(0.022)	(0.022)	(0.015)	
Accounting	-0.19***	0.03	-0.25***	-0.19***	51644
	(0.024)	(0.025)	(0.024)	(0.019)	
Business	-0.16***	-0.038**	-0.19***	-0.16***	96784
	(0.016)	(0.017)	(0.017)	(0.010)	
Prod. Eng.	-0.13***	-0.10**	-0.12***	-0.10***	23008
	(0.037)	(0.042)	(0.037)	(0.028)	
History	-0.12*	-0.11*	-0.11*	-0.14***	13220
	(0.064)	(0.067)	(0.064)	(0.037)	

Source: Made from ENADE 2017, 2018 and 2019 microdata, INEP.
Note: *p<0.1; **p<0.05; ***p<0.01.

4.3.3 Sensitivity Analysis

In this section we conduct the last step of the propensity score method, which is the sensitivity analysis. The sensitivity analysis is important to reinforce the robustness of the results. The non biased estimators of ATT in propensity score are based on the hypothesis that there are not any unobservable variables that impact the selection to treatment and also the variable of interest. The sensitivity analysis test if ATT would change in the case in which there are unobeservable variables affecting the selection to treatment. The results of the sensitivity analysis, by undergraduate course and type of score in ENADE (Final Score, General Component and Specific Component), are presented in Tables 17, 18 and 19. It is important to highlight that we conduct the analysis varying the gamma from

1 to 5 with intervals of 0.5. In Tables 17, 18 and 19 we present just the coefficient for p_mh-, which is the significance level to the assumption of underestimation of treatment effect, but the complete Tables with p_mh+, the significance level for the assumption of overestimation of treatment effect are in Appendix G. Here we focus on the coefficient of p_mh-, assuming that if there is selection bias, the direction of the bias would be underestimating the treatment effect, because the profile students that enroll in distance programs would be the profile with worst performance in ENADE.

In general, the values of p_mh-, suggest that we can reject the hypothesis of underestimation of treatment effect. So the coefficients of ATT are not sensitive to the possibility of unobservable variables impacting the selection to treatment, the coefficients.

Table 17 – Sensitivity Analysis by Undergraduate Course and Type of Score in 2017.

Gamma	**Pedagoy (p_mh-)**			**History (p_mh-)**			**Phy. Ed. DE. (p_mh-)**		
	Final	**Gen**	**Spec**	**Final**	**Gen**	**Spec**	**Final**	**Gen**	**Spec**
1	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.02	0.00
1.5	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.23	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.01
2.5	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.12
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Source: Made from ENADE 2017 microdata, INEP.

Note p_mh-: significance level (assumption: underestimation of treatment effect).

Table 18 – Sensitivity Analysis by Undergraduate Course and Type of Score in 2018.

Gamma	**Business (p_mh-)**			**Accounting (p_mh-)**			**Social Work (p_mh-)**		
	Final	**Gen**	**Spec**	**Final**	**Gen**	**Spec**	**Final**	**Gen**	**Spec**
1	0.00	0.00	0.00	0.00	0.36	0.00	0.00	0.00	0.00
1.5	0.02	0.00	0.30	0.31	0.00	0.00	0.00	0.37	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.35	0.00	0.03
3.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Source: Made from ENADE 2018 microdata, INEP.

Note p_mh-: significance level (assumption: underestimation of treatment effect).

Table 19 – Sensitivity Analysis by Undergraduate Course and Type of Score in 2019.

Gamma	T. Env. Man. (p_mh-)			Phy. Ed. BA. (p_mh-)			Prod. Eng. (p_mh-)		
	Final	**Gen**	**Spec**	**Final**	**Gen**	**Spec**	**Final**	**Gen**	**Spec**
1	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.06	0.00
1.5	0.05	0.36	0.11	0.00	0.41	0.00	0.04	0.00	0.08
2	0.49	0.01	0.35	0.00	0.00	0.00	0.00	0.00	0.00
2.5	0.08	0.00	0.04	0.09	0.00	0.09	0.00	0.00	0.00
3	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Source: Made from ENADE 2019 microdata, INEP.

Note p_mh-: significance level (assumption: underestimation of treatment effect).

4.4 Final Considerations

Brazil has gone through a process of expanding the offer of higher education in distance learning modality in the last decade, but there are still few studies evaluating this method and its quality. It is important to say that exist some barriers in this type of analysis to have satisfactory results. Firstly we need to find appropriate parameters to quality. Secondly, and main difficulty is, to correctly control for the bias selection problem. So we try to contribute with the debate in Brazil comparing the performance in ENADES's score between distance learners and in person students applying techniques to correct the endogeneity.

In this paper we used as parameter for quality for the higher educations the final ENADES' score in terms of standard deviation, and also the general and specific component. To analyse the heterogeneity effect among undergraduate courses, we selected nine courses that are the ones with concentrate the higher numbers of distance learners, three for each year of analysis. The courses selected are: Pedagogy Degree, History Degree and Physical Education Degree in 2017; Business, Accounting and Social Work in 2018 and Technology in Environmental Management, Physical Education Bachelor and Production Engineering in 2019.

We applied the propensity score matching and inverse probability weighting method with replacement and for the observations in common support. To control for bias selection we have socioeconomic variables of the student and also characteristics of the university.

About the covariates, one limitation we have is that the variables collected are about the characteristics after treatment, but those variables are very relevant also reflect the situation of the students before the treatment. Here our interest was to measure the ATT, and we also conducted balance and sensitivity analysis.

Regarding the results, we can start with the matching quality. For the logit model to selection to treatment we have a good correctly classification parameter the values are around 80% and 90%, which implies a good forecasting ability and not disrespect the common support condition. In relation to the distribution of propensity score before and after matching in Figures 13, 14 and 15 we saw that before matching the distribution of propensity score is very different but after matching the distribution is very similar. And in the balance analysis in Figure 16 in general, distance learners and in person students were quite different, before matching. But, after matching, we can note a reduction of the value of standardized % bias on the covariates, which means that treatment and control groups became more similar.

In reference to ATT in final score of ENADE in Table 11 we can note that the negative level of impact of distance education varies among the undergraduate courses. The effect is greater and around 0.5 standard deviation for Social Work, Physical Education Degree and Bachelor. In contrast, the influence of the modality is lower for Production Engineering and History a bit more than 0.1 standard deviation, such results indicate that performing the distance learning modality decreases, on average, student performance.

Furthermore the dissimilarity is higher in specific component than in general component. So the most part of heterogeneity in performance comes from the final content of undergraduate courses and makes the dissimilarity in performance between the modalities even more concerning. Regarding the ATT for the specific component and the highest impact, we have Social Work, Technology in Environmental Management and Physical Education Degree and Bachelor, in which the values of ATT are around 0.5 standard deviation. The lowest coefficient is for History of ATT of -0.11.

In addition, we can take some conclusions comparing our results with some facts in the Literature. Firstly, our results are in agreement with the part of the Literature, in which the bias selection is a concern in the analysis and the distance education has a negative impact on the grades. And about, that the theory explain some disadvantages of distance learning like the necessity of self regulatory and self discipline from the student as in Xu and Jaggars (2014), the isolation and the lack of interaction that encompass the insufficient interaction with the peer and the instructor as stated by Harmon, Lambrinos et al. (2012). So those disadvantages can impact negatively on the scores. According to Ni (2013), the interaction between student and professor and among students is fundamental in the learning process, and the distance education approach try to offer tools to promote that interaction. But maybe online coursers with great tools and quality could be very expensive (Xu; Jaggars, 2013).

Moreover, those results can be very worrying if the lower quality in distance reflect in the quality of the work. Since online programs are in expansion and will train people for the labor market, the productivity of the workers could be impacted negatively in the long run. Some articles that estimate this relation as in Dale and Krueger (2002), Heckman, Stixrud and Urzua (2006) and Saavedra (2009) and found a positive impact of quality in education and the outcomes in labor market and the is cumulative over the life cycle in labor force (Heckman; Stixrud; Urzua, 2006). Even though the return fall when controlling for bias selection, the results are still significant (Dale; Krueger, 2002). The effect of quality in education is higher for the ones with disadvantaged background families (Dale; Krueger, 2002; Saavedra, 2009).

Particularly for the expansion of distance education in undergraduate courses for professionals in education sector, it is important to think that those workers will be responsible for the education of great part of the children in the country. So we can have a irreparable damage in education for generations. There are many researchers trying to investigate the role of teacher's quality in students outcomes as Darling-Hammond (2000) and CARVALHO (2018).

The authors in those studies argue that investments in quality and training of the teachers have a positive and significant improvement on the scholar outcomes of the students, despite of the fact that the intensity of the effect is lower when they control for social economic characteristics (CARVALHO, 2018). Also in CARVALHO (2018), the author says that this relationship is little studied in Brazil, but the few studies about this topic reinforce the role of teacher quality as a very important input for the learning process.

On the other hand, considering the advantages of distance education like the schedule flexibility (Xu; Jaggars, 2011), and the lower costs of education, where the benefits can be a good strategy to enlarge the access to the higher level and qualify the workers in the labor market. So as those analysis are a very complex topic, it can be a good option to focus the policy of expansion of distance education for the ones who would not go to college if it were not for the distance modality. Also a good possibility could be the hybrid method alternative to take advantages of both approaches.

Accordingly to the sensitive analysis the values of p_mh-, suggest that we can reject the hypothesis of underestimation of treatment effect. So the coefficients of ATT are not sensitive to the possibility of unobservable variables impacting the selection to treatment, the coefficients.

The main contributions of the work are: to fill the gap of Brazilian literature in this area, by estimating the ATT controlling for bias selection and also considering the heterogeneity of the impact among the undergraduate courses; it helps with information to think about the public policies in the expansion of distance learning, maybe it is a good idea to focus on the public that really would not ingress in College Degree if they did not have this alternative and as the effect is heterogeneous among undergraduate courses, there are some courses that the modality is more appropriate than others; it can be a start to invest in strategies to improve distance education.

The debate about quality in distance education is still in progress and future researches could focus in measure the impacts of distance learning in the labor market outcomes; development of methodologies to improve distance education and analyse the effect of distance education in performance considering different profiles of students.

References

Andrade, F. V.; Lopes, A. Ead: uma história de inovações tecnológicas no brasil. In: *7º Congresso Integrado de Tecnologia da Informação*. 2012.

BATISTA, A. B.; CRUZ, N. V. S.; Andrade, C. M.; Bruni, A. L. Desempenho discente nos enades 2009 e 2012 do curso de ciências contábeis do nordeste brasileiro: uma análise comparativa entre o ensino à distância e o ensino presencial. In: *Anais do Congresso Brasileiro de Custos-ABC*. 2014.

Bonaldo, L.; Pereira, L. N. Dropout: Demographic profile of brazilian university students. *Procedia-Social and behavioral sciences*, Elsevier, v. 228, p. 138–143, 2016.

Borges, G. da R.; Mondini, V. E. D.; Domingues, M. J. C. de S.; Mondini, L. C. A relação entre o perfil dos alunos que cursam ead e os motivos de escolha desta modalidade. *Revista de Administração Unimep*, v. 14, n. 3, p. 80–101, 2016.

Bowen, W. G.; Chingos, M. M.; Lack, K. A.; Nygren, T. I. Interactive learning online at public universities: Evidence from a six-campus randomized trial. *Journal of Policy Analysis and Management*, Wiley Online Library, v. 33, n. 1, p. 94–111, 2014.

Bunker, E. L. The history of distance education through the eyes of the international council for distance education. *Handbook of distance education*, Citeseer, p. 49–66, 2003.

Caetano, C. C. R.; Cardoso, T. A. de O.; Miranda, G. J.; Freitas, S. C. de. Desempenho no enade em ciências contábeis: ensino a distância (ead) versus presencial. *Revista Universo Contábil*, v. 11, n. 4, p. 147–165, 2016.

Caliendo, M.; Kopeinig, S. Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, Wiley Online Library, v. 22, n. 1, p. 31–72, 2008.

Cameron, A. C.; Trivedi, P. K. *Microeconometrics: methods and applications.* : Cambridge university press, 2005.

CARVALHO, M. R. V. D. Formação docente e desempenho discente na educação básica. 2018.

Cerulli, G. *Econometric evaluation of socio-economic programs Theory and applications.* : Springer, 2015.

Dale, S. B.; Krueger, A. B. Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *The Quarterly Journal of Economics*, MIT Press, v. 117, n. 4, p. 1491–1527, 2002.

Darling-Hammond, L. Teacher quality and student achievement. *Education policy analysis archives*, v. 8, p. 1–1, 2000.

Deschacht, N.; Goeman, K. The effect of blended learning on course persistence and performance of adult learners: A difference-in-differences analysis. *Computers & Education*, Elsevier, v. 87, p. 83–89, 2015.

Ferlin, E. P.; Alcântara, F. C.; Paulo, L. G. de. Comparativo do perfil e do desempenho acadêmico dos estudantes nas modalidades presencial e ead: Estudo de caso do curso de engenharia da computação. *Revista de Ensino de Engenharia*, v. 39, 2020.

Gratton-Lavoie, C.; Stanley, D. Teaching and learning principles of microeconomics online: An empirical assessment. *The Journal of Economic Education*, Taylor & Francis, v. 40, n. 1, p. 3–25, 2009.

Greene, W. H. *Econometric analysis*. 7th. ed. : Prentice Hall, 2001.

Hanna, D. E. Organizational models in higher education, past and future. *Handbook of distance education*, Citeseer, p. 67–78, 2003.

Harmon, O. R.; Lambrinos, J. Online format vs. live mode of instruction: Do human capital differences or differences in returns to human capital explain the differences in outcomes? 2006.

Harmon, O. R.; Lambrinos, J. et al. *Testing the effect of hybrid lecture delivery on learning outcomes*. 2012.

Heckman, J. J.; Stixrud, J.; Urzua, S. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, The University of Chicago Press, v. 24, n. 3, p. 411–482, 2006.

Lassibille, G.; Gómez, L. N. Why do higher education students drop out? evidence from spain. *Education Economics*, Taylor & Francis, v. 16, n. 1, p. 89–105, 2008.

Latanich, G.; Nonis, S. A.; Hudson, G. I. A profile of today's distance learners: An investigation of demographic and individual difference variables of distance and non-distance learners. *Journal of Marketing for Higher Education*, Taylor & Francis, v. 11, n. 3, p. 1–16, 2001.

Levy, Y. Comparing dropouts and persistence in e-learning courses. *Computers & education*, Elsevier, v. 48, n. 2, p. 185–204, 2007.

Lima, D. d. C. B. P.; Borges, L. C.; Souza, W. C. de. Perfil dos estudantes da modalidade a distância no brasil: do levantamento bibliográfico aos dados. In: SBC. *Anais do II Seminário de Educação a Distância da Região Centro-Oeste*. 2018.

Litto, F. M.; Formiga, M. *Educação a Distancia: o estado da arte*. : Pearson, 2009. 9–13 p.

MARQUES, F. C. Avaliações de impacto do desempenho estudantil no brasil: do ensino infantil ao ensino superior. 2020.

Means, B.; Toyama, Y.; Murphy, R.; Baki, M. et al. The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record*, v. 115, n. 3, p. 1–47, 2013.

Moore, M. G.; Aderson, W. G. *Handbook of distance education*. : Routledge, 2003. 37–47 p.

Mugnol, M. A educação a distância no brasil: conceitos e fundamentos. *Revista Diálogo Educacional*, v. 9, n. 27, p. 335–349, 2009.

Nascimento, L. F.; Czykiel, R.; Figueiró, P. S. Presencial ou a distância: a modalidade de ensino influencia na aprendizagem? *Administração: Ensino e Pesquisa*, Associação Nacional dos Cursos de Graduação em Administração, v. 14, n. 2, p. 311–341, 2013.

Nguyen, T. The effectiveness of online learning: Beyond no significant difference and future horizons. *MERLOT Journal of Online Learning and Teaching*, v. 11, n. 2, p. 309–319, 2015.

Ni, A. Y. Comparing the effectiveness of classroom and online learning: Teaching research methods. *Journal of Public Affairs Education*, Taylor & Francis, v. 19, n. 2, p. 199–215, 2013.

Oliveira, É. T. d.; Piconez, S. C. B. Avaliação da educação superior nas modalidades presencial e a distância: análises com base no conceito preliminar de cursos (cpc). *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, SciELO Brasil, v. 22, n. 3, p. 833–851, 2017.

Pittman, V. V. Correspondence study in the american university: A second historiographic perspective. *Handbook of distance education*, Lawrence Erlbaum Associates, p. 21–35, 2003.

Saavedra, J. E. The learning and early labor market effects of college quality: A regression discontinuity analysis. *Investigaciones del ICFES*, 2009.

Saba, F. Distance education theory, methodology, and epistemology: Pragmatic paradigm. *Handbook of distance education*, LEA publishers, 2003.

Santos, D. P. dos; Carneiro, S. E. Curso de geografia/licenciatura da unimontes: Uma análise do perfil dos estudantes. *Revista Ciranda*, v. 3, n. 1, p. 1–10, 2019.

SANTOS, P. K. d.; Giraffa, L. Evasão na educação superior: um estudo sobre o censo da educação superior no brasil. *CLABES, III. Anais. Disponível em: http://www. alfaguia. org/www. alfa/images/ponencias/clabesIII/LT_1/ponencia_completa_200. pdf Acesso em*, v. 5, n. 03, 2015.

Saraiva, T. Educação a distância no brasil: lições da história. *Em aberto*, v. 16, n. 70, 2008.

Shachar, M.; Neumann, Y. Differences between traditional and distance education academic performances: A meta-analytic approach. *The International Review of Research in Open and Distributed Learning*, v. 4, n. 2, 2003.

Shachar, M.; Neumann, Y. Twenty years of research on the academic performance differences between traditional and distance learning: Summative meta-analysis and trend examination. *MERLOT Journal of Online Learning and Teaching*, v. 6, n. 2, 2010.

Simonson, M.; Zvacek, S. M.; Smaldino, S. Teaching and learning at a distance: Foundations of distance education 6th edition. IAP, p. 2–30, 2015.

Simonson, M.; Zvacek, S. M.; Smaldino, S. Teaching and learning at a distance: Foundations of distance education 6th edition. IAP, p. 2–30, 2015.

Spina, G. L. Razões para escolha de uma graduação a distância: uma abordagem sobre o perfil de alunos ingressos em ead. 2013.

StataCorp, L. Stata treatment-effects reference manual: Potential outcomes/counterfactual outcomes. *(Release 15)*, Stata Press College Station, TX, 2017.

Stephen, M.; Christopher, W. *Counterfactuals and Causal Inference: Methods and Principles for Social Research.* : Cambridge University Press Cambridge, UK:, 2007.

Teixeira, I. I. N. de Estudos e P. E. A. *Censo da educação superior: notas estatísticas-2021.* : MEC Brasília, DF, 2022.

Winston, G.; Zimmerman, D. Peer effects in higher education. In: *College choices: The economics of where to go, when to go, and how to pay for it.* : University of Chicago Press, 2004. p. 395–424.

Xu, D.; Jaggars, S. S. The effectiveness of distance education across virginia's community colleges: Evidence from introductory college-level math and english courses. *Educational Evaluation and Policy Analysis*, SAGE Publications Sage CA: Los Angeles, CA, v. 33, n. 3, p. 360–377, 2011.

Xu, D.; Jaggars, S. S. The impact of online learning on students' course outcomes: Evidence from a large community and technical college system. *Economics of Education Review*, Elsevier, v. 37, p. 46–57, 2013.

Xu, D.; Jaggars, S. S. Performance gaps between online and face-to-face courses: Differences across types of students and academic subject areas. *The Journal of Higher Education*, Taylor & Francis, v. 85, n. 5, p. 633–659, 2014.

APPENDIX A – Complete tables for the models to modality choice and quota policy

Table 20 – Modality choice determinants and quota policy by undergraduate course.

	Dependent variable:		
	Distance learning		
Covariates[1]	(Ped)	(Hist)	(Phy ed de)
Age	0.036***	0.024***	0.019***
	(0.004)	(0.012)	(0.019)
Age^2	−0.0004***	−0.0002***	−0.0002***
	(0.0001)	(0.0002)	(0.0003)
Male	0.026***	−0.036***	−0.008***
	(0.023)	(0.037)	(0.041)
White	−0.010***	−0.010***	−0.032***
	(0.012)	(0.039)	(0.043)
Per capita income	−0.00001***	−0.00000**	−0.00000***
	(0.00001)	(0.00001)	(0.00002)
Part time job	0.076***	0.044***	0.011***
	(0.017)	(0.056)	(0.061)
Full time job	0.100***	0.071***	0.033***
	(0.014)	(0.047)	(0.053)
Live Alone	0.025***	0.008	0.006
	(0.028)	(0.072)	(0.078)
Live with	0.091***	0.053***	0.027***

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Ped)	(Hist)	(Phy ed de)
sp and/or ch	(0.013)	(0.044)	(0.049)
Other	−0.023***	−0.033***	−0.029***
	(0.041)	(0.101)	(0.123)
Public high sch	0.029***	0.036***	0.007**
	(0.019)	(0.054)	(0.066)
Parent ed none	0.002	0.038***	0.005
	(0.028)	(0.089)	(0.125)
Parent ed elementary 1	−0.004	0.024***	−0.004
	(0.019)	(0.058)	(0.061)
Parent ed elementary 2	−0.027***	−0.001	−0.017***
	(0.021)	(0.065)	(0.067)
Parent ed high sch	−0.038***	−0.011**	−0.016***
	(0.019)	(0.056)	(0.056)
RO	−0.114***	−1.544	−0.945
	(0.079)	(772.529)	(1,241.726)
AC	−2.277	−1.528	−0.904
	(209.593)	(680.862)	(1,137.936)
AM	−2.340	−1.615	−0.911
	(68.892)	(417.353)	(498.295)
RR	−0.162***	−1.570	−0.909
	(0.129)	(731.584)	(827.863)
PA	−0.310***	−1.577	−0.904

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Ped)	(Hist)	(Phy ed de)
	(0.063)	(299.963)	(531.880)
AP	−2.337	−1.605	0.019
	(119.449)	(655.152)	(0.253)
TO	0.303***	−1.553	−0.912
	(0.040)	(521.243)	(1,185.427)
MA	0.139***	−1.577	−0.894
	(0.042)	(302.162)	(574.384)
PI	−0.024***	−1.536	−0.932
	(0.049)	(276.372)	(556.294)
CE	−0.140***	−0.248***	−0.902
	(0.067)	(0.367)	(368.974)
RN	−0.063***	−0.064***	−0.058***
	(0.064)	(0.197)	(0.374)
PB	−0.025***	−1.472	−0.910
	(0.059)	(281.830)	(666.374)
PE	−0.202***	−0.0002	−0.889
	(0.057)	(0.126)	(530.521)
AL	−0.165***	−1.546	−0.888
	(0.063)	(373.823)	(597.570)
SE	0.102***	0.162***	−0.892
	(0.055)	(0.094)	(625.641)
BA	−0.221***	0.006	−0.022***

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Ped)	(Hist)	(Phy ed de)
	(0.036)	(0.085)	(0.113)
MG	0.073***	0.010	−0.043***
	(0.022)	(0.085)	(0.107)
ES	−0.402***	0.107***	−0.029***
	(0.138)	(0.125)	(0.165)
RJ	0.056***	−0.017***	−0.898
	(0.021)	(0.061)	(224.633)
PR	0.271***	0.161***	0.122***
	(0.019)	(0.057)	(0.055)
SC	0.249***	0.120***	0.120***
	(0.029)	(0.089)	(0.060)
RS	0.139***	−0.104***	−0.061***
	(0.024)	(0.101)	(0.133)
MS	0.307***	−0.050***	−0.060***
	(0.034)	(0.127)	(0.219)
MT	−0.140***	−1.520	−0.894
	(0.067)	(465.079)	(615.699)
GO	−2.314	−0.171***	0.020***
	(56.396)	(0.182)	(0.120)
DF	−0.318***	−1.571	0.045***
	(0.093)	(478.220)	(0.105)
Private	0.224***	0.136***	0.011***

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Ped)	(Hist)	(Phy ed de)
university	(0.019)	(0.058)	(0.064)
Quota	-0.072^{***}	-0.052^{***}	-0.032^{***}
	(0.014)	(0.046)	(0.054)
IGC	0.006^{***}	-0.049^{***}	0.026^{***}
	(0.015)	(0.053)	(0.056)
Hours	0.003^{***}	0.002^{***}	0.003^{***}
	(0.002)	(0.005)	(0.005)
2011	0.058^{***}	-0.159^{***}	-0.130^{***}
	(0.016)	(0.061)	(0.100)
2014	0.117^{***}	-0.0004	-0.059^{***}
	(0.014)	(0.042)	(0.055)
Constant	-1.097^{***}	-0.581^{***}	-0.559^{***}
	(0.095)	(0.296)	(0.364)
Observations	231,488	36,558	51,155
Log Likelihood	−98,597.650	−10,007.110	−8,576.090
Akaike Inf. Crit.	197,291.300	20,110.220	17,248.180

Note: *p<0.1; **p<0.05; ***p<0.01.
Source: Made from ENADE microdata 2011, 2014 and 2017, INEP.

[1] Explanatory note on the reference categories of the covariates: '2017' is the category reference to year, 'no job' to work, 'live with parents and/or relatives' to situation in home, 'College Degree or more' to parent's schooling and 'SP' to states.

Table 21 – Modality choice determinants and quota policy by undergraduate course.

	Dependent variable:		
	Distance learning		
Covariates[2]	(Bus)	(Acc)	(S work)
Age	0.034***	0.029***	0.014***
	(0.006)	(0.009)	(0.010)
Age^2	−0.0003***	−0.0003***	−0.0002***
	(0.0001)	(0.0001)	(0.0001)
Male	−0.014***	−0.001	0.005
	(0.014)	(0.022)	(0.055)
White	−0.051***	−0.041***	−0.025***
	(0.014)	(0.023)	(0.033)
Per capita	−0.00001***	0.00000**	−0.00001***
income	(0.00000)	(0.00001)	(0.00002)
Part	0.032***	0.031***	0.036***
time job	(0.028)	(0.051)	(0.050)
Full	0.022***	0.0001	0.040***
time job	(0.018)	(0.032)	(0.034)
Live	0.046***	0.034***	−0.006
alone	(0.026)	(0.041)	(0.064)
Live with	0.082***	0.071***	0.030***
sp and/or ch	(0.016)	(0.025)	(0.037)
Other	0.013***	0.024***	−0.031***
	(0.042)	(0.068)	(0.092)
Public	0.034***	−0.018***	0.005

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Bus)	(Acc)	(S work)
high sch	(0.019)	(0.032)	(0.049)
Parent ed	0.011**	−0.0003	0.020***
none	(0.042)	(0.068)	(0.078)
Parent ed	0.002	−0.011***	0.008*
elementary 1	(0.022)	(0.034)	(0.053)
Parent ed	−0.004	−0.018***	−0.002
elementary 2	(0.023)	(0.036)	(0.058)
Parent ed	−0.004*	−0.017***	0.0002
high sch	(0.019)	(0.031)	(0.052)
RO	−1.493	−1.612	−1.507
	(48.146)	(308.167)	(1,509.521)
AC	−1.542	−1.679	−1.483
	(103.654)	(746.456)	(1,765.695)
AM	−0.115***	−0.180***	−0.311***
	(0.084)	(0.135)	(0.260)
RR	−1.501	−1.618	−1.444
	(76.580)	(491.038)	(765.504)
PA	−0.220***	−1.670	−1.421
	(0.133)	(231.542)	(309.642)
AP	−1.582	−1.710	−1.545
	(86.013)	(857.101)	(2,475.858)
TO	0.148***	0.226***	0.582***

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Bus)	(Acc)	(S work)
	(0.084)	(0.106)	(0.222)
MA	-0.204^{***}	-0.193^{***}	-1.443
	(0.093)	(0.154)	(333.363)
PI	0.184^{***}	-1.578	-1.438
	(0.055)	(258.145)	(427.880)
CE	-0.291^{***}	-1.640	-1.434
	(0.139)	(175.968)	(249.697)
RN	-0.128^{***}	-0.112^{***}	-0.175^{***}
	(0.071)	(0.098)	(0.101)
PB	-0.283^{***}	-1.551	-1.338
	(0.203)	(210.728)	(245.159)
PE	-0.434^{***}	-0.291^{***}	-1.410
	(0.226)	(0.184)	(306.088)
AL	-0.498^{***}	-1.567	-1.290
	(0.581)	(254.333)	(301.041)
SE	-0.020^{**}	-0.096^{***}	0.028^{***}
	(0.074)	(0.149)	(0.074)
BA	-0.127^{***}	-0.140^{***}	-0.092^{***}
	(0.054)	(0.086)	(0.064)
MG	-0.081^{***}	-0.081^{***}	-0.101^{***}
	(0.031)	(0.044)	(0.085)
ES	-0.232^{***}	-0.113^{***}	-1.308

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Bus)	(Acc)	(S work)
	(0.144)	(0.133)	(458.300)
RJ	0.021***	−0.030***	−0.102***
	(0.022)	(0.038)	(0.054)
PR	0.266***	0.227***	0.226***
	(0.020)	(0.030)	(0.058)
SC	0.104***	0.064***	0.122***
	(0.027)	(0.039)	(0.084)
RS	−0.007***	−0.170***	−0.031***
	(0.024)	(0.067)	(0.065)
MS	0.295***	0.266***	0.201***
	(0.033)	(0.053)	(0.080)
MT	−0.172***	−1.622	−1.451
	(0.114)	(192.140)	(485.665)
GO	−0.374***	−1.647	−1.453
	(0.186)	(200.654)	(395.238)
DF	−0.101***	−0.154***	−0.303***
	(0.072)	(0.121)	(0.312)
Private university	0.169***	0.313***	0.438***
	(0.029)	(0.062)	(0.196)
Quota	−0.058***	−0.053***	−0.068***
	(0.017)	(0.029)	(0.037)
IGC	−0.078***	−0.035***	−0.068***

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Bus)	(Acc)	(S work)
	(0.019)	(0.032)	(0.054)
Hours	0.005***	0.004***	0.0004
	(0.002)	(0.003)	(0.005)
2010			−0.032***
			(0.045)
2012	0.045***	−0.024***	
	(0.016)	(0.030)	
2015	0.046***	0.038***	
	(0.016)	(0.025)	
2016			0.024***
			(0.038)
Constant	−0.847***	−0.901***	−0.537***
	(0.125)	(0.201)	(0.325)
Observations	217,159	100,802	57,742
Log Likelihood	−74,712.010	−29,144.520	−13,790.290
Akaike Inf. Crit.	149,520.000	58,385.030	27,676.580

Note: *p<0.1; **p<0.05; ***p<0.01.

Source: Made from ENADE microdata 2010, 2012, 2015, 2016 and 2018, INEP.

2 Explanatory note on the reference categories of the covariates: '2018' is the category reference to year, 'no job' to work, 'live with parents and/or relatives' to situation in home, 'College Degree or more' to parent's schooling and 'SP' to states.

Table 22 – Modality choice determinants and quota policy by undergraduate course.

	Dependent variable:		
	Distance learning		
Covariates[3]	(Tech in env)	(Phy ed ba)	(Prod eng)
Age	0.011***	0.024***	0.015***
	(0.024)	(0.020)	(0.027)
Age^2	−0.0001***	−0.0003***	−0.0001***
	(0.0003)	(0.0003)	(0.0004)
Male	−0.004	−0.010***	−0.004
	(0.077)	(0.043)	(0.065)
White	−0.023***	−0.035***	−0.023***
	(0.078)	(0.043)	(0.058)
Per capita income	−0.00000*	−0.00000**	−0.00000***
	(0.00002)	(0.00001)	(0.00001)
Part time job	−0.002	0.044***	0.009
	(0.138)	(0.062)	(0.148)
Full time job	0.001	0.055***	0.025***
	(0.096)	(0.056)	(0.093)
Live alone	0.025**	0.019***	0.029***
	(0.158)	(0.076)	(0.107)
Live with sp and/or ch	0.051***	0.034***	0.035***
	(0.096)	(0.052)	(0.071)
Other	0.012	0.007	0.006
	(0.196)	(0.115)	(0.165)
Public	0.018**	0.008**	0.009***

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Tech in env)	(Phy ed ba)	(Prod eng)
high sch	(0.105)	(0.056)	(0.072)
Parent ed none	0.004	−0.010	−0.007
	(0.213)	(0.149)	(0.239)
Parent ed elementary 1	0.016**	−0.009**	−0.011***
	(0.120)	(0.068)	(0.090)
Parent ed elementary 2	0.005	−0.020***	−0.014***
	(0.129)	(0.069)	(0.095)
Parent ed high sch	0.003	−0.022***	−0.006**
	(0.108)	(0.053)	(0.073)
RO	−1.028	−1.185	−0.740
	(965.707)	(2,932.172)	(1,819.715)
AC	−1.060	−1.141	
	(954.607)	(2,568.130)	
AM	−1.294	−1.126	−0.723
	(1,236.963)	(1,658.506)	(698.395)
RR	−1.469		
	(6,102.265)		
PA	−0.280***	−1.168	−0.712
	(1.077)	(1,253.984)	(811.077)
AP	−1.311	−1.122	−0.753
	(2,140.294)	(11,785.830)	(1,794.369)
TO	−1.415	−1.147	−0.697

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Tech in env)	(Phy ed ba)	(Prod eng)
	(7,372.331)	(2,848.886)	(1,924.984)
MA	-0.096^{***}	-1.139	-0.736
	(0.382)	(1,404.874)	(676.956)
PI	-1.025	-1.161	-0.705
	(804.359)	(1,266.423)	(1,082.387)
CE	-1.256	0.021^{***}	-0.723
	(842.189)	(0.076)	(532.704)
RN	-0.006	-0.096^{***}	-0.691
	(0.200)	(0.225)	(749.995)
PB	-1.166	-1.137	-0.700
	(629.025)	(869.835)	(805.116)
PE	0.193^{***}	-1.118	-0.134^{***}
	(0.241)	(637.967)	(0.365)
AL	-1.002	-1.111	-0.739
	(723.751)	(1,223.872)	(1,403.005)
SE		-0.027^{***}	-0.715
		(0.127)	(907.398)
BA	-0.128^{***}	-1.205	-0.732
	(0.286)	(1,094.119)	(467.869)
MG	-1.193	-1.119	-0.189^{***}
	(338.644)	(385.732)	(0.383)
ES	0.997	-1.163	-0.691

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Tech in env)	(Phy ed ba)	(Prod eng)
	(10,754.010)	(1,021.099)	(654.957)
RJ	−0.052***	−0.175***	−0.006**
	(0.113)	(0.148)	(0.068)
PR	0.201***	−0.329***	0.028***
	(0.114)	(0.711)	(0.079)
SC	0.171***	0.154***	−0.021***
	(0.324)	(0.051)	(0.118)
RS	−0.184***	−1.169	−0.705
	(0.172)	(494.108)	(347.684)
MS	0.032	−1.167	−0.052***
	(0.298)	(1,438.757)	(0.447)
MT	−1.059	−1.178	−0.727
	(711.723)	(2,030.012)	(876.603)
GO	−1.241	−1.140	−0.711
	(817.060)	(871.729)	(660.823)
DF	−1.416	−0.276***	−0.699
	(2,877.450)	(0.514)	(1,147.720)
Private university	0.364***	1.189	0.0002
	(0.193)	(252.161)	(0.098)
Quota	−0.027***	−0.041***	−0.027***
	(0.091)	(0.060)	(0.081)
IGC	0.069***	0.121***	−0.029***

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Tech in env)	(Phy ed ba)	(Prod eng)
	(0.126)	(0.094)	(0.074)
Hours	0.008***	0.002***	0.002***
	(0.011)	(0.006)	(0.007)
2010	−0.193***		
	(0.111)		
2011			−0.675
			(247.794)
2014			−0.024***
			(0.077)
2016	−0.016**	−1.122	
	(0.108)	(236.114)	
Constant	−0.682***	−2.050	−0.340***
	(0.617)	(252.161)	(0.541)
Observations	11,169	42,057	35,783
Log Likelihood	−2,455.912	−7,426.691	−4,644.783
Akaike Inf. Crit.	5,005.824	14,945.380	9,381.565

Note: *p<0.1; **p<0.05; ***p<0.01.

Source: Made from ENADE microdata 2010, 2011, 2014, 2016 and 2019 INEP.

[3] Explanatory note on the reference categories of the covariates: '2018' is the category reference to year, 'no job' to work, 'live with parents and/or relatives' to situation in home, 'College Degree or more' to parent's schooling and 'SP' to states.

APPENDIX B – Complete table for financing policies

Table 23 – Modality choice determinants in private universities and financing programs by undergraduate course.

	Dependent variable:		
	Distance learning		
Covariates[1]	(Ped)	(Hist)	(Phy ed de)
Age	0.028***	0.031***	0.018***
	(0.006)	(0.016)	(0.025)
Age^2	−0.0003***	−0.0003***	−0.0002***
	(0.0001)	(0.0002)	(0.0004)
Male	0.050***	−0.038***	−0.008***
	(0.035)	(0.050)	(0.053)
White	−0.019***	−0.030***	−0.038***
	(0.016)	(0.052)	(0.055)
Per capita income	−0.00001***	−0.00001***	−0.00000***
	(0.00001)	(0.00001)	(0.00002)
Part time job	0.048***	0.054***	0.005
	(0.024)	(0.078)	(0.078)
Full time job	0.062***	0.063***	0.016***
	(0.018)	(0.063)	(0.067)
Live alone	0.036***	0.028**	0.016***
	(0.038)	(0.098)	(0.100)

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Ped)	(Hist)	(Phy ed de)
Live with sp and/or ch	0.084*** (0.018)	0.088*** (0.061)	0.029*** (0.065)
Other	0.019*** (0.059)	0.007 (0.153)	0.003 (0.163)
Public high sch	0.030*** (0.026)	0.060*** (0.070)	−0.002 (0.085)
Parent ed none	−0.001 (0.040)	0.039*** (0.133)	−0.002 (0.184)
Parent ed elementary 1	−0.005* (0.025)	0.032*** (0.077)	−0.008** (0.079)
Parent ed elementary 2	−0.020*** (0.028)	−0.001 (0.086)	−0.016*** (0.085)
Parent ed high sch	−0.030*** (0.025)	−0.018** (0.073)	−0.013*** (0.070)
RO	−1.891 (186.035)		−0.829 (1,636.573)
AC	−1.864 (369.653)		−0.831 (1,635.014)
AM	−1.927 (108.481)	−2.068 (496.606)	−0.831 (549.443)
RR	0.036* (0.182)		

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Ped)	(Hist)	(Phy ed de)
PA	−1.933	−2.055	−0.831
	(132.852)	(352.500)	(861.022)
AP	−1.804		−0.802
	(259.410)		(1,562.658)
TO	0.369***		−0.835
	(0.118)		(1,379.449)
MA	−0.192***		−0.812
	(0.147)		(649.957)
PI	−1.962	−2.056	−0.832
	(154.725)	(659.730)	(3,502.043)
CE	−0.246***	−0.175***	−0.838
	(0.171)	(0.485)	(542.812)
RN	0.025**	−1.966	−0.812
	(0.108)	(636.118)	(1,562.571)
PB	−1.859		−0.820
	(467.211)		(1,762.681)
PE	−0.586***	−2.052	−0.818
	(0.293)	(462.580)	(1,004.028)
AL	−1.918		−0.816
	(217.245)		(1,235.296)
SE	0.106***	0.210***	−0.816
	(0.069)	(0.120)	(652.713)

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Ped)	(Hist)	(Phy ed de)
BA	−0.084***	−1.999	−0.818
	(0.060)	(405.336)	(634.139)
MG	0.064***	0.062***	−0.001
	(0.030)	(0.116)	(0.123)
ES	−0.290***	−1.968	−0.829
	(0.145)	(550.183)	(631.693)
RJ	0.018***	−0.041***	−0.813
	(0.028)	(0.072)	(251.751)
PR	0.277***	0.263***	0.122***
	(0.026)	(0.073)	(0.070)
SC	0.250***	0.211***	0.133***
	(0.038)	(0.111)	(0.075)
RS	0.129***	−0.100***	−0.029***
	(0.030)	(0.113)	(0.145)
MS	0.392***	0.153***	−0.126***
	(0.057)	(0.216)	(0.518)
MT	−1.882	−1.921	−0.792
	(134.891)	(1,574.625)	(931.874)
GO	−1.888	−2.005	−0.779
	(91.875)	(505.927)	(600.944)
DF	−0.251***	−2.037	−0.799
	(0.115)	(466.290)	(500.469)

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Ped)	(Hist)	(Phy ed de)
Scholarship	-0.189^{***}	-0.154^{***}	-0.071^{***}
	(0.018)	(0.058)	(0.061)
Financing	-0.337^{***}	-0.327^{***}	-0.132^{***}
	(0.027)	(0.103)	(0.102)
IGC	-0.018^{***}	-0.082^{***}	0.046^{***}
	(0.023)	(0.082)	(0.098)
Hours	0.001^{***}	0.002^{***}	0.003^{***}
	(0.002)	(0.006)	(0.007)
2011	-0.013^{***}	-0.249^{***}	-0.180^{***}
	(0.022)	(0.080)	(0.178)
2014	0.090^{***}	-0.058^{***}	-0.120^{***}
	(0.020)	(0.060)	(0.108)
Constant	-0.517^{***}	-0.448^{***}	-0.507^{***}
	(0.119)	(0.371)	(0.499)
Observations	157,784	15,032	33,409
Log Likelihood	−55,285.930	−5,311.896	−5,027.438
Akaike Inf. Crit.	110,667.900	10,703.790	10,148.880

Note: *p<0.1; **p<0.05; ***p<0.01.

Source: Made from ENADE microdata 2011, 2014 and 2017, INEP.

[1] Explanatory note on the reference categories of the covariates: '2017' is the category reference to year, 'no job' to work, 'live with parents and/or relatives' to situation in home, 'College Degree or more' to parent's schooling and 'SP' to states

Table 24 – Modality choice determinants in private universities and financing programs by undergraduate course.

	Dependent variable:		
	Distance learning		
Covariates[2]	(Bus)	(Acc)	(S work)
Age	0.035^{***}	0.032^{***}	0.014^{***}
	(0.007)	(0.010)	(0.012)
Age^2	-0.0004^{***}	-0.0003^{***}	-0.0001^{***}
	(0.0001)	(0.0001)	(0.0002)
Male	-0.016^{***}	-0.002	0.018^{***}
	(0.015)	(0.024)	(0.064)
White	-0.061^{***}	-0.052^{***}	-0.031^{***}
	(0.016)	(0.025)	(0.038)
Per capita income	-0.00001^{***}	-0.00000^{***}	-0.00001^{***}
	(0.00000)	(0.00001)	(0.00002)
Part time job	0.024^{***}	0.027^{***}	0.030^{***}
	(0.033)	(0.057)	(0.058)
Full time job	0.014^{***}	-0.004	0.039^{***}
	(0.021)	(0.036)	(0.040)
Live alone	0.064^{***}	0.051^{***}	0.007
	(0.030)	(0.047)	(0.077)
Live with sp and/or ch	0.090^{***}	0.087^{***}	0.035^{***}
	(0.018)	(0.027)	(0.043)
Other	0.045^{***}	0.045^{***}	-0.022^{***}
	(0.049)	(0.076)	(0.110)

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Bus)	(Acc)	(S work)
Public high sch	0.045***	−0.008**	0.022***
	(0.021)	(0.034)	(0.056)
Parent ed none	0.016***	0.005	0.032***
	(0.056)	(0.085)	(0.096)
Parent ed elementary 1	0.005*	−0.009**	0.015***
	(0.024)	(0.038)	(0.061)
Parent ed elementary 2	0.001	−0.018***	0.006
	(0.025)	(0.040)	(0.067)
Parent ed high sch	0.002	−0.017***	0.005
	(0.021)	(0.034)	(0.060)
RO	−1.749	−1.726	−1.462
	(194.415)	(264.221)	(857.003)
AC	−1.708	−1.728	−1.428
	(276.039)	(432.186)	(960.158)
AM	−0.167***	−0.178***	−0.304***
	(0.109)	(0.140)	(0.280)
RR	−1.747	−1.742	−1.325
	(312.468)	(450.069)	(671.125)
PA	−1.788	−1.756	−1.443
	(104.710)	(173.770)	(287.000)
AP	−1.792	−1.734	−1.471
	(244.924)	(479.737)	(1,449.262)

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Bus)	(Acc)	(S work)
TO	−1.792	−1.710	−1.430
	(247.343)	(440.545)	(1,077.181)
MA	−0.206***	−0.216***	−1.389
	(0.106)	(0.161)	(227.874)
PI	−1.758	−1.743	−1.406
	(148.495)	(267.821)	(316.463)
CE	−1.710	−1.720	−1.429
	(68.241)	(143.398)	(200.428)
RN	−0.113***	−0.103***	−0.149***
	(0.075)	(0.103)	(0.110)
PB	−1.747	−1.713	−1.417
	(133.889)	(300.600)	(397.089)
PE	−0.446***	−0.334***	−1.411
	(0.241)	(0.205)	(275.042)
AL	−1.780	−1.774	−1.472
	(158.741)	(373.919)	(756.184)
SE	−0.008	−0.090***	0.046***
	(0.081)	(0.158)	(0.085)
BA	−0.127***	−0.119***	−0.051***
	(0.064)	(0.091)	(0.075)
MG	−0.068***	−0.058***	−0.051***
	(0.034)	(0.047)	(0.094)

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Bus)	(Acc)	(S work)
ES	-0.238^{***}	-1.753	-1.441
	(0.156)	(216.348)	(631.888)
RJ	0.021^{***}	-0.015^{***}	-0.073^{***}
	(0.024)	(0.041)	(0.061)
PR	0.277^{***}	0.267^{***}	0.264^{***}
	(0.022)	(0.033)	(0.066)
SC	0.098^{***}	0.065^{***}	0.161^{***}
	(0.030)	(0.043)	(0.095)
RS	0.004	-0.174^{***}	-0.020^{***}
	(0.026)	(0.070)	(0.075)
MS	0.385^{***}	0.352^{***}	0.276^{***}
	(0.044)	(0.063)	(0.088)
MT	-1.760	-1.693	-1.371
	(111.332)	(162.881)	(461.423)
GO	-1.751	-1.738	-1.387
	(72.076)	(151.499)	(288.255)
DF	-0.065^{***}	-0.132^{***}	-0.285^{***}
	(0.076)	(0.129)	(0.323)
Scholarship	-0.136^{***}	-0.116^{***}	-0.147^{***}
	(0.017)	(0.026)	(0.042)
Financing	-0.277^{***}	-0.280^{***}	-0.267^{***}
	(0.034)	(0.053)	(0.070)

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Bus)	(Acc)	(S work)
IGC	-0.092^{***}	-0.032^{***}	-0.035^{***}
	(0.023)	(0.037)	(0.061)
Hours	0.005^{***}	0.006^{***}	0.001^{***}
	(0.002)	(0.003)	(0.005)
2010			-0.092^{***}
			(0.052)
2012	-0.009^{***}	-0.073^{***}	
	(0.019)	(0.034)	
2015	0.026^{***}	0.018^{***}	
	(0.019)	(0.028)	
2016			0.014^{***}
			(0.043)
Constant	-0.582^{***}	-0.623^{***}	-0.137^{***}
	(0.133)	(0.201)	(0.277)
Observations	171,090	75,914	41,939
Log Likelihood	−58,504.630	−24,103.730	−10,369.910
Akaike Inf. Crit.	117,105.300	48,303.460	20,835.830

Note: *p<0.1; **p<0.05; ***p<0.01.

Source: Made from ENADE microdata 2010, 2012, 2015, 2016 and 2018, INEP.

[2] Explanatory note on the reference categories of the covariates: '2018' is the category reference to year, 'no job' to work, 'live with parents and/or relatives' to situation in home, 'College Degree or more' to parent's schooling and 'SP' to states

Table 25 – Modality choice determinants in private universities and financing programs by undergraduate course.

	Dependent variable:		
	Distance learning		
Covariates[3]	(Tech in env)	(Phy ed ba)	(Prod eng)
Age	0.012***	0.024***	0.017***
	(0.029)	(0.021)	(0.030)
Age^2	−0.0001***	−0.0003***	−0.0002***
	(0.0004)	(0.0003)	(0.0004)
Male	−0.011*	−0.013***	−0.007**
	(0.089)	(0.046)	(0.072)
White	−0.026***	−0.039***	−0.033***
	(0.090)	(0.046)	(0.064)
Per capita income	−0.00000**	−0.00000***	−0.00000***
	(0.00003)	(0.00001)	(0.00001)
Part time job	−0.006	0.041***	0.009
	(0.165)	(0.067)	(0.165)
Full time job	−0.007	0.053***	0.025***
	(0.113)	(0.060)	(0.102)
Live alone	0.029**	0.025***	0.045***
	(0.188)	(0.081)	(0.118)
Live with sp and/or ch	0.055***	0.040***	0.050***
	(0.112)	(0.056)	(0.079)
Other	0.047***	0.020***	0.029***
	(0.253)	(0.126)	(0.188)

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Tech in env)	(Phy ed ba)	(Prod eng)
Public high sch	0.015* (0.124)	0.019*** (0.060)	0.016*** (0.077)
Parent ed none	−0.015 (0.242)	−0.015 (0.159)	−0.003 (0.261)
Parent ed elementary 1	0.019* (0.141)	−0.006 (0.073)	−0.010** (0.099)
Parent ed elementary 2	0.010 (0.149)	−0.017*** (0.074)	−0.014*** (0.104)
Parent ed high sch	0.007 (0.126)	−0.020*** (0.056)	−0.005 (0.079)
RO		−1.189 (1,734.031)	−0.915 (2,018.157)
AC		−1.199 (2,357.922)	
AM	−1.433 (1,977.124)	−1.155 (1,264.682)	−0.901 (974.047)
RR	−1.526 (3,691.880)		
PA	−0.333*** (1.108)	−1.166 (768.160)	−0.844 (1,016.874)
AP	−1.534 (6,522.639)	−1.100 (7,157.766)	−0.861 (2,884.911)

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Tech in env)	(Phy ed ba)	(Prod eng)
TO	−1.467	−1.210	−0.838
	(4,429.286)	(2,811.689)	(1,890.663)
MA	−0.114***	−1.157	−0.873
	(0.402)	(975.728)	(720.132)
PI		−1.183	−0.860
		(745.583)	(1,504.014)
CE	−1.506	0.020***	−0.860
	(2,834.868)	(0.083)	(639.246)
RN	−0.159***	−0.085***	−0.857
	(0.432)	(0.232)	(1,465.730)
PB	−1.479	−1.170	−0.889
	(2,819.140)	(677.296)	(1,749.517)
PE	−0.290***	−1.180	−0.151***
	(0.492)	(587.408)	(0.368)
AL		−1.161	−0.896
		(925.554)	(1,744.154)
SE		−0.041***	−0.877
		(0.134)	(1,155.207)
BA	−0.154***	−1.213	−0.886
	(0.318)	(650.068)	(542.993)
MG	−1.291	−1.152	−0.216***
	(410.110)	(294.042)	(0.414)

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Tech in env)	(Phy ed ba)	(Prod eng)
ES	0.907	−1.194	−0.846
	(6,522.639)	(785.548)	(853.376)
RJ	−0.051***	−0.171***	0.004
	(0.123)	(0.152)	(0.073)
PR	0.224***	−0.351***	0.044***
	(0.128)	(0.713)	(0.088)
SC	1.245	0.166***	−0.014**
	(327.743)	(0.056)	(0.127)
RS	−0.193***	−1.197	−0.864
	(0.186)	(336.609)	(383.991)
MS	0.034	−1.184	−0.025
	(0.312)	(834.490)	(0.497)
MT		−1.205	−0.854
		(1,341.529)	(1,175.004)
GO	−1.368	−1.157	−0.864
	(888.632)	(616.654)	(811.548)
DF	−1.388	−0.284***	−0.891
	(1,755.005)	(0.519)	(2,245.968)
Scholarship	−0.052***	−0.122***	−0.047***
	(0.102)	(0.057)	(0.067)
Financing	−0.091***	−0.089***	−0.127***
	(0.207)	(0.075)	(0.118)

Continue...

	Dependent variable:		
	Distance learning		
Covariates	(Tech in env)	(Phy ed ba)	(Prod eng)
IGC	0.089^{***}	0.119^{***}	-0.017^{***}
	(0.137)	(0.100)	(0.089)
Hours	0.009^{***}	0.003^{***}	0.002^{***}
	(0.013)	(0.007)	(0.008)
2010	-0.253^{***}		
	(0.135)		
2011			-0.850
			(341.063)
2014			-0.031^{***}
			(0.083)
2016	-0.089^{***}	-1.164	
	(0.133)	(182.639)	
Constant	-0.325^{***}	-0.831^{***}	-0.407^{***}
	(0.636)	(0.471)	(0.588)
Observations	8,000	34,378	24,560
Log Likelihood	−1,865.870	−6,577.816	−3,843.904
Akaike Inf. Crit.	3,815.740	13,247.630	7,779.808

Note: *p<0.1; **p<0.05; ***p<0.01.

Source: Made from ENADE microdata 2010, 2011, 2014, 2016 and 2019, INEP.

[3] Explanatory note on the reference categories of the covariates: '2019' is the category reference to year, 'no job' to work, 'live with parents and/or relatives' to situation in home, 'College Degree or more' to parent's schooling and 'SP' to states

APPENDIX C – Complete tables of the treatment selection logit model

Table 26 – Logit regression of attribution to treatment by undergraduate course in 2017

Covariates[1]	Pedagogy	History	Phy. Ed. DE.
Distance			
Age	0.252***	0.310***	0.381***
	(0.006)	(0.018)	(0.023)
Age^2	-0.003***	-0.003***	-0.004***
	(0.000)	(0.000)	(0.000)
Male	0.140***	-0.271***	-0.111**
	(0.034)	(0.056)	(0.050)
White	0.035**	0.187***	-0.531***
	(0.017)	(0.057)	(0.051)
Per capita income	0.000***	0.000**	-0.000
	(0.000)	(0.000)	(0.000)
Work	0.563***	0.575***	0.415***
	(0.018)	(0.065)	(0.058)
Live Alone	0.154***	0.063	0.177*
	(0.040)	(0.108)	(0.093)
Live with sp and or ch	0.679***	0.759***	0.594***
	(0.020)	(0.068)	(0.060)
Other	-0.087	-0.047	-0.221
	(0.060)	(0.161)	(0.154)
Public high sch	0.071**	0.439***	0.194**
	(0.029)	(0.079)	(0.078)
Parent ed none	-0.115***	0.108	-0.214
	(0.040)	(0.144)	(0.157)
Parent ed el 1	-0.098***	0.204**	-0.199***
	(0.027)	(0.087)	(0.074)

Continue...

Parent ed 2	-0.233***	-0.028	-0.493***
	(0.030)	(0.096)	(0.080)
Parent ed high sch	-0.287***	-0.104	-0.431***
	(0.027)	(0.082)	(0.066)
RO	-0.598***		
	(0.108)		
RR	0.292**		
	(0.147)		
TO	2.630***		
	(0.060)		
MA	-1.484***		
	(0.126)		
AP			0.626
			(0.411)
PI	0.739***		
	(0.070)		
CE	-1.484***	-1.036**	
	(0.135)	(0.409)	
RN	0.643***	1.962***	0.548
	(0.079)	(0.234)	(0.385)
PB	-0.436***		
	(0.106)		
PE	-1.002***	-1.359***	
	(0.072)	(0.387)	
AL	-0.774***		
	(0.141)		
SE	0.356***	2.611***	
	(0.073)	(0.201)	
BA	-1.469***		
	(0.062)		
MG	0.613***	0.984***	-1.866***
	(0.031)	(0.120)	(0.283)

Continue...

ES	-2.714***	2.732***	
	(0.139)	(0.146)	
RJ	-0.010	0.922***	
	(0.033)	(0.079)	
PR	2.414***	2.395***	3.531***
	(0.025)	(0.088)	(0.062)
SC	2.740***	3.444***	3.851***
	(0.040)	(0.135)	(0.072)
RS	0.232***	-0.113	-1.693***
	(0.044)	(0.138)	(0.276)
MS	2.848***	1.353***	-0.119
	(0.052)	(0.158)	(0.226)
MT	-0.091		
	(0.085)		
DF	-2.986***		
	(0.144)		
Private Uni	1.889***	2.433***	2.291***
	(0.027)	(0.080)	(0.124)
IGC	1.095***	0.216***	1.024***
	(0.022)	(0.078)	(0.082)
Hours	0.064***	0.084***	0.210***
	(0.008)	(0.026)	(0.025)
Academic Scolarship	-1.489***	-1.653***	-1.510***
	(0.029)	(0.091)	(0.084)
Quota	-0.517***	-0.866***	-0.712***
	(0.020)	(0.075)	(0.066)
Constant	-10.851***	-11.314***	-15.435***
	(0.139)	(0.431)	(0.497)
Observations	114384	14267	28331

Note: *p<0.1; **p<0.05; ***p<0.01.

Source: Made from ENADE microdata 2017, INEP.

[1] Explanatory note on the reference categories of the covariates:'live with parents and/or relatives' is the category of reference to situation in home, 'College Degree or more' to parent's schooling, 'SP' to

Table 27 – Logit regression of attribution to treatment by undergraduate course in 2018

Covariates[2]	Business	Accounting	Social Work
Distance			
Age	0.310***	0.294***	0.159***
	(0.009)	(0.011)	(0.014)
Age^2	-0.003***	-0.003***	-0.002***
	(0.000)	(0.000)	(0.000)
Male	-0.143***	-0.031	0.015
	(0.019)	(0.027)	(0.069)
White	-0.283***	-0.217***	-0.180***
	(0.020)	(0.028)	(0.042)
Per capita income	-0.000***	0.000***	-0.000
	(0.000)	(0.000)	(0.000)
Work	0.157***	0.085**	0.484***
	(0.025)	(0.038)	(0.041)
Live Alone	0.339***	0.244***	-0.082
	(0.036)	(0.050)	(0.076)
Live with sp and/or ch	0.691***	0.680***	0.517***
	(0.023)	(0.031)	(0.046)
Other	0.136**	0.159*	-0.159
	(0.063)	(0.086)	(0.118)
Public high sch	0.246***	-0.114***	0.368***
	(0.027)	(0.040)	(0.064)
Parent ed none	-0.046	-0.097	0.164*
	(0.056)	(0.081)	(0.094)
Parent ed el 1	0.039	-0.119***	0.173**
	(0.031)	(0.043)	(0.067)
Parent ed el 2	-0.032	-0.205***	0.123*
	(0.033)	(0.046)	(0.074)

Continue...

states

Parent ed high sch	-0.057**	-0.230***	0.054
	(0.027)	(0.038)	(0.066)
AM	-0.814***	-0.851***	-2.460***
	(0.104)	(0.136)	(0.264)
MA	-1.286***	-0.941***	
	(0.116)	(0.165)	
PI	2.046***		
	(0.070)		
RN	-0.166	-0.287**	-1.622***
	(0.104)	(0.134)	(0.143)
PE	-3.265***	-2.297***	
	(0.227)	(0.184)	
SE	-0.339***	0.338**	0.561***
	(0.130)	(0.163)	(0.180)
BA	-1.180***	-1.137***	-0.684***
	(0.075)	(0.112)	(0.089)
MG	-0.804***	-0.512***	
	(0.048)	(0.059)	
ES	-1.649***		
	(0.144)		
RJ	0.989***	0.895***	0.225***
	(0.029)	(0.045)	(0.070)
PR	2.563***	2.829***	3.236***
	(0.027)	(0.037)	(0.064)
SC	1.607***	1.584***	3.827***
	(0.035)	(0.048)	(0.117)
RS	-0.768***	-0.882***	1.327***
	(0.050)	(0.077)	(0.108)
MS	2.902***	3.024***	3.552***
	(0.059)	(0.086)	(0.169)
DF	-0.867***	-1.651***	-1.861***
	(0.095)	(0.167)	(0.323)

Continue...

Private uni	2.012***	5.373***	
	(0.046)	(0.202)	
IGC	0.248***	0.150***	-1.363***
	(0.027)	(0.038)	(0.056)
Hours	0.154***	0.130***	-0.050**
	(0.010)	(0.013)	(0.021)
Academic Scholarship	-0.765***	-0.734***	-1.387***
	(0.037)	(0.052)	(0.077)
Quota	-0.610***	-0.615***	-1.208***
	(0.025)	(0.036)	(0.048)
Constant	-11.016***	-13.718***	-0.935***
	(0.183)	(0.314)	(0.306)
Observations	102396	54654	22026

Note: *p<0.1; **p<0.05; ***p<0.01.

Source: Made from ENADE microdata 2018, INEP.

Table 28 – Logit regression of attribution to treatment by undergraduate course in 2019

Covariates[3]	Tech. Env. Man.	Phy. Ed. BA.	Prod. Eng.
Distance			
Age	0.162***	0.382***	0.396***
	(0.045)	(0.018)	(0.028)
Age^2	-0.001**	-0.004***	-0.003***
	(0.001)	(0.000)	(0.000)
Male	0.119	-0.175***	0.088
	(0.152)	(0.038)	(0.067)
White	0.068	-0.289***	-0.078
	(0.157)	(0.037)	(0.059)

Continue...

[2] Explanatory note on the reference categories of the covariates:'live with parents and/or relatives' is the category of reference to situation in home, 'College Degree or more' to parent's schooling, 'SP' to states

Per capita income	0.000	-0.000***	-0.000
	(0.000)	(0.000)	(0.000)
Work	0.083	0.966***	0.698***
	(0.176)	(0.049)	(0.094)
Live alone	0.183	0.274***	0.462***
	(0.274)	(0.064)	(0.110)
Live with sp and/or ch	1.231***	0.575***	0.801***
	(0.198)	(0.044)	(0.074)
Other	0.772*	0.028	-0.104
	(0.395)	(0.098)	(0.168)
Public high sch	0.457**	0.408***	0.453***
	(0.211)	(0.048)	(0.075)
Parent ed none	0.473	-0.050	-0.255
	(0.465)	(0.125)	(0.251)
Parent ed el 1	0.446*	-0.088	-0.250***
	(0.246)	(0.058)	(0.093)
Parent ed el 2	-0.001	-0.279***	-0.318***
	(0.251)	(0.060)	(0.098)
Parent ed high sch	-0.018	-0.364***	-0.122
	(0.202)	(0.046)	(0.076)
MA	-2.080***		
	(0.511)		
CE		1.278***	
		(0.070)	
RN	2.429***	-1.177***	
	(0.511)	(0.220)	
PE	0.798		-1.674***
	(0.488)		(0.368)
SE		0.498***	
		(0.123)	
BA	1.390**		
	(0.645)		

Continue...

RJ	1.729***	-2.266***	1.547***
	(0.248)	(0.146)	(0.070)
PR	2.777***	-4.900***	1.849***
	(0.225)	(0.709)	(0.080)
SC	3.170***	3.703***	0.898***
	(0.525)	(0.048)	(0.124)
RS	-1.282***		
	(0.309)		
MS			0.229
			(0.449)
DF		-3.296***	
		(0.506)	
Private uni	6.373***		-0.023
	(0.279)		(0.092)
IGC	-0.183	0.484***	-0.074
	(0.278)	(0.048)	(0.073)
Hours	0.369***	0.128***	0.222***
	(0.080)	(0.020)	(0.026)
Academic Scholarship	-0.515**	-1.474***	-0.649***
	(0.262)	(0.064)	(0.102)
Qhota	-1.034***	-0.843***	-0.676***
	(0.176)	(0.053)	(0.082)
Constant	-9.802***	-11.191***	-13.211***
	(1.178)	(0.342)	(0.569)
Observations	4355	33179	24589

Note: *p<0.1; **p<0.05; ***p<0.01.

Source: Made from ENADE microdata 2019, INEP.

[3] Explanatory note on the reference categories of the covariates:'live with parents and/or relatives' is the category of reference to situation in home, 'College Degree or more' to parent's schooling, 'SP' to states

APPENDIX D – Parameters about the forecasting ability of the logit model

Table 29 – Forecasting ability of the Logit model by undergraduate course

	Ped	**Hist**	**PE. DE.**	**Bus**	**Acc**	**S W**	**TEM**	**PE. BA.**	**Eng**
Correc Classif	80%	86%	92%	84%	85%	82%	95%	87%	93%
Pr(+ D)	81%	75%	62%	56%	61%	78%	98%	67%	29%
Pr(- D)	80%	90%	97%	93%	93%	85%	87%	95%	99%
Pr(D +)	81%	77%	77%	73%	74%	83%	94%	83%	64%
Pr(D -)	80%	89%	94%	87%	88%	80%	95%	88%	94%
Pr(+ D)	20%	10%	3%	7%	7%	15%	13%	5%	1%
Pr(- D)	19%	25%	38%	44%	39%	22%	2%	33%	71%
Pr(D +)	19%	23%	23%	27%	26%	17%	6%	17%	36%
Pr(D -)	20%	11%	6%	13%	12%	20%	5%	12%	6%

Source: Made from ENADE 2017, 2018 and 2019 microdata, INEP.

APPENDIX E – Complete balancing analysis tables

Table 30 – Test-t of the control variables of the unmatched and matched sample in 2017 by undergraduate course.

Variables[1]	**Pedagogy**		**History**		**Phy. Ed. DE.**	
	PV	**Gap**	**PV**	**Gap**	**PV**	**Gap**
Age_U	4.21	0.00	8.47	0.00	4.86	0.00
Age_M	0.09	0.08	0.00	0.99	0.92	0.00
Age_2_U	271.30	0.00	578.42	0.00	299.27	0.00
Age_2_M	9.20	0.02	1.80	0.92	64.19	0.00
Male_U	-0.01	0.00	0.01	0.16	0.02	0.05
Male_M	0.00	0.03	-0.05	0.00	-0.01	0.43
White_U	0.11	0.00	0.10	0.00	-0.01	0.36
White_M	-0.06	0.00	-0.01	0.37	-0.13	0.00
Per_inc_U	177.98	0.00	145.40	0.00	4.30	0.86
Per_inc_M	-80.22	0.00	-89.10	0.03	-142.60	0.00
Work_U	0.16	0.00	0.26	0.00	0.17	0.00
Work_M	-0.01	0.04	-0.02	0.01	-0.02	0.10
Alone_U	0.00	0.00	0.00	0.80	0.01	0.02
Alone_M	0.00	0.05	0.00	0.46	0.01	0.22
Family_U	0.25	0.00	0.39	0.00	0.29	0.00
Family_M	-0.02	0.00	0.00	0.70	-0.01	0.55
Other_U	-0.01	0.00	-0.06	0.00	-0.03	0.00
Other_M	0.00	0.00	0.00	0.70	0.00	0.63
Public_high_U	0.01	0.00	0.07	0.00	0.08	0.00
Public_high_M	-0.01	0.00	-0.01	0.35	-0.02	0.00
Parent_none_U	0.01	0.00	0.02	0.00	0.01	0.00
Parent_none_M	0.01	0.00	-0.02	0.00	0.02	0.00
Parent_el1_U	0.09	0.00	0.14	0.00	0.12	0.00
Parent_el1_M	0.00	0.90	0.04	0.00	0.01	0.33
Parent_el2_U	-0.02	0.00	0.01	0.09	0.00	0.69
Parent_el2_M	0.00	0.06	0.00	0.57	-0.01	0.23
Parent_high_U	-0.08	0.00	-0.10	0.00	-0.10	0.00

Continue...

Parent_high_M	0.00	0.07	-0.03	0.00	0.00	1.00
RO_U	-0.01	0.00				
RO_M	0.00	0.00				
RR_U	0.00	0.00				
RR_M	0.00	0.03				
TO_U	0.02	0.00				
TO_M	-0.01	0.00				
MA_U	-0.02	0.00				
MA_M	0.00	0.58				
AP_U					0.00	0.00
AP_M					0.00	0.56
AL_U	-0.01	0.00				
AL_M	0.00	0.42				
SE_U	0.00	0.00				
SE_M	0.00	0.72	0.02	0.00		
PI_U	-0.01483	0.00	-0.01	0.04		
PI_M	-0.00332	0.00				
CE_U	-0.02	0.00	-0.04	0.00		
CE_M	0.00	0.44	0.00	0.20		
RN_U	-0.01	0.00	-0.01	0.00	-0.01	0.00
RN_M	0.00	0.94	0.00	0.09	0.00	0.35
PB_U	-0.01	0.00				
PB_M	0.00	0.84				
PE_U	-0.03	0.00	-0.03	0.00		
PE_M	0.00	0.22	0.00	0.37		
BA_U	-0.05	0.00				
BA_M	0.00	0.44				
MG_U	-0.03	0.00	-0.02	0.00	-0.09	0.00
MG_M	-0.01	0.00	0.00	0.48	0.00	0.12
ES_U	-0.02	0.00	0.02	0.00		
ES_M	0.00	0.03	-0.12	0.00		
RJ_U	-0.04	0.00	0.03	0.00		
RJ_M	-0.01	0.00	-0.05	0.00		
PR_U	0.32	0.00	0.24	0.00	0.39	0.00
PR_M	0.02	0.00	0.12	0.00	0.03	0.01
SC_U	0.10	0.00	0.10	0.00	0.28	0.00
SC_M	0.01	0.00	0.05	0.00	-0.04	0.00
RS_U	-0.01	0.00	-0.03	0.00	-0.06	0.00

Continue...

RS_M	0.00	0.00	0.00	0.71	0.00	0.55
MS_U	0.06	0.00	0.01	0.05	-0.01	0.00
MS_M	-0.01	0.00	0.01	0.00	0.00	0.88
MT_U	-0.01	0.00				
MT_M	0.00	0.00				
DF_U	-0.02	0.00				
DF_M	0.00	0.40				
Private_uni_U	0.26	0.00	0.52	0.00	0.23	0.00
Private_uni_M	0.04	0.00	-0.02	0.03	0.01	0.00
IGC_U	0.00	0.21	-0.15	0.00	-0.08	0.00
IGC_M	-0.07	0.00	-0.03	0.00	-0.14	0.00
Hours_U	0.07	0.00	0.04	0.09	0.16	0.00
Hours_M	-0.08	0.00	-0.10	0.00	0.01	0.60
Academic_sch_U	-0.14	0.00	-0.35	0.00	-0.16	0.00
Academic_sch_M	0.01	0.00	0.01	0.11	0.01	0.32
Quota_U	-0.13	0.00	-0.17	0.00	-0.13	0.00
Quota_M	-0.02	0.00	-0.02	0.01	-0.01	0.24

Source: Made from ENADE microdata 2017, INEP.

Table 31 – Test-t of the control variables of the unmatched and matched sample in 2018 by undergraduate course.

Variables[2]	**Business**		**Accounting**		**Social Work**	
	Gap	**PV**	**Gap**	**PV**	**Gap**	**PV**
Age_U	5.27	0.00	5.44	0.00	5.72	0.00
Age_M	-0.22	0.00	0.11	0.30	0.73	0.00
Age_2_U	340.60	0.00	359.18	0.00	396.40	0.00
Age_2_M	-12.50	0.02	9.00	0.25	54.90	0.00
Male_U	-0.02	0.00	-0.01	0.21	-0.01	0.14
Male_M	0.00	0.29	0.01	0.38	-0.01	0.00
White_U	-0.03	0.00	0.03	0.00	0.02	0.01
White_M	-0.01	0.02	-0.06	0.00	-0.05	0.00

Continue...

[1] Explanatory note on the reference categories of the covariates:'live with parents and/or relatives' is the category of reference to situation in home, 'College Degree or more' to parent's schooling, 'SP' to states

Per_inc_U	-421.10	0.00	211.50	0.00	-23.54	0.12
Per_inc_M	-43.40	0.01	-303.80	0.00	-205.55	0.00
Work_U	0.09	0.00	0.07	0.00	0.21	0.00
Work_M	0.00	0.90	-0.01	0.01	0.01	0.16
Alone_U	-0.01	0.00	-0.01	0.06	-0.01	0.00
Alone_M	0.00	0.32	0.00	0.61	0.02	0.00
Family_U	0.31	0.00	0.31	0.00	0.27	0.00
Family_M	-0.01	0.02	0.00	0.60	-0.02	0.00
Other_U	-0.02	0.00	-0.01	0.00	-0.03	0.00
Other_M	0.00	0.01	0.00	0.83	0.00	0.07
Public_high_U	0.11	0.00	0.02	0.00	0.08	0.00
Public_high_M	0.00	0.38	0.00	0.81	-0.03	0.00
Parent_none_U	0.01	0.00	0.01	0.00	0.02	0.00
Parent_none_M	0.00	0.28	0.00	0.95	0.02	0.00
Parent_el1_U	0.12	0.00	0.10	0.00	0.14	0.00
Parent_el1_M	0.00	0.64	0.00	0.51	-0.05	0.00
Parent_el2_U	0.02	0.00	0.00	0.84	0.00	0.67
Parent_el2_M	0.00	0.90	0.00	0.64	0.03	0.00
Parent_high_U	-0.06	0.00	-0.08	0.00	-0.11	0.00
Parent_high_M	0.00	0.94	0.01	0.36	-0.03	0.00
AM_U	-0.01	0.00	-0.01	0.00	-0.03	0.00
AM_M	0.00	0.59	0.00	0.44	0.00	0.73
MA_U	-0.02	0.00	-0.01	0.00		
MA_M	0.00	0.56	0.00	0.30		
PI_U	0.01	0.00				
PI_M	-0.03	0.00				
RN_U	-0.01	0.00	-0.01	0.00	-0.05	0.00
RN_M	0.00	0.95	0.00	0.93	0.00	0.05
PE_U	-0.04	0.00	-0.04	0.00		
PE_M	0.00	0.04	0.00	0.34		
SE_M	0.00	0.00	0.00	0.00	-0.01	0.00
SE_U	0.00	0.63	0.00	0.51	0.00	0.77
BA_U	-0.03	0.00	-0.04	0.00	-0.06	0.00
BA_M	0.00	0.08	0.00	0.53	0.00	0.07
MG_U	-0.09	0.00	-0.08	0.00		
MG_M	0.00	0.90	0.00	0.09		
ES_U	-0.02	0.00				
ES_M	0.00	0.48				

Continue...

RJ_U	0.03	0.00	0.03	0.00	-0.06	0.00
RJ_M	-0.01	0.00	0.00	0.37	0.00	0.54
PR_U	0.33	0.00	0.37	0.00	0.37	0.00
PR_M	0.03	0.00	0.00	0.66	0.00	0.65
SC_U	0.06	0.00	0.04	0.00	0.09	0.00
SC_M	0.01	0.01	0.01	0.04	0.05	0.00
RS_U	-0.05	0.00	-0.05	0.00	0.00	0.17
RS_M	0.00	0.62	0.00	0.96	0.01	0.00
MS_U	0.06	0.00	0.05	0.00	0.10	0.00
MS_M	0.00	0.36	0.00	0.12	-0.04	0.00
DF_U	-0.01	0.00	-0.02	0.00	-0.02	0.00
DF_M	0.00	0.33	0.00	0.17	0.00	0.55
Private_uni_U	0.15	0.00	0.20	0.00		
Private_uni_M	0.00	0.46	0.00	0.87		
IGC_U	-0.14	0.00	-0.14	0.00	-0.31	0.00
IGC_M	-0.03	0.00	-0.03	0.00	0.08	0.00
Hours_U	0.11	0.00	0.10	0.00	-0.07	0.00
Hours_M	-0.01	0.12	-0.03	0.02	-0.11	0.00
Academic_sch_U	-0.06	0.00	-0.05	0.00	-0.14	0.00
Academic_sch_M	0.00	0.42	0.00	0.56	0.02	0.00
Quota_U	-0.09	0.00	-0.12	0.00	-0.19	0.00
Quota_M	0.00	0.19	0.01	0.15	-0.01	0.04

Source: Made from ENADE microdata 2018, INEP.

Table 32 – Test-t of the control variables of the unmatched and matched sample in 2019 by undergraduate course.

Variables[3]	**Tech. Env. Man**		**Phy. Ed. BA.**		**Prod. Eng**	
	Gap	**PV**	**Gap**	**PV**	**Gap**	**PV**
Age_U	4.92	0.00	4.62	0.00	9.12	0.00
Age_M	-2.84	0.00	-0.05	0.62	-0.10	0.70
Age_2_U	342.25	0.00	288.40	0.00	614.74	0.00

Continue...

2 Explanatory note on the reference categories of the covariates:'live with parents and/or relatives' is the category of reference to situation in home, 'College Degree or more' to parent's schooling, 'SP' to states

Age_2_M	-194.10	0.00	-2.80	0.73	-12.40	0.56
Male_U	0.18	0.00	-0.02	0.00	0.18	0.00
Male_M	-0.03	0.04	0.02	0.01	0.02	0.19
White_U	0.10	0.00	-0.09	0.00	-0.03	0.03
White_M	0.00	0.83	-0.17	0.00	0.01	0.59
Per_inc_U	251.67	0.00	-121.40	0.00	-270.30	0.00
Per_inc_M	-11.30	0.78	-365.20	0.00	105.70	0.21
Work_U	0.27	0.00	0.22	0.00	0.21	0.00
Work_M	-0.07	0.00	-0.03	0.00	0.00	0.91
Alone_U	0.02	0.02	0.03	0.00	0.00	0.75
Alone_M	-0.03	0.00	-0.01	0.05	0.00	0.77
Family_U	0.22	0.00	0.22	0.00	0.48	0.00
Family_M	-0.12	0.00	-0.02	0.01	0.01	0.69
Other_U	-0.05	0.00	-0.02	0.00	-0.07	0.00
Other_M	0.03	0.00	-0.01	0.00	0.00	0.65
Public_high_U	0.03	0.02	0.08	0.00	0.19	0.00
Public_high_M	0.10	0.00	-0.01	0.32	-0.01	0.46
Parent_none_U	0.02	0.03	0.01	0.00	0.01	0.00
Parent_none_M	0.03	0.00	0.01	0.00	0.00	0.51
Parent_el1_U	0.07	0.00	0.08	0.00	0.13	0.00
Parent_el1_M	0.01	0.31	0.00	0.65	0.01	0.66
Parent_el2_U	0.03	0.06	0.01	0.00	0.04	0.00
Parent_el2_M	0.08	0.00	0.00	0.41	0.01	0.68
Parent_high_U	-0.07	0.00	-0.08	0.00	-0.02	0.15
Parent_high_M	-0.08	0.00	-0.01	0.45	-0.01	0.51
MA_U	0.00	0.35				
MA_M	0.00	0.83				
CE_U			0.03	0.00		
CE_M			0.00	0.47		
RN_U	-0.05	0.00	-0.01	0.00		
RN_M	0.00	0.30	0.00	0.65		
PE_U	-0.05	0.00			-0.03	0.00
PE_M	-0.21	0.00			0.00	1.00
SE_U			0.00	0.11		
SE_M			0.00	0.26		
BA_U	0.02	0.00				
BA_M	-0.03	0.00				
RJ_U	0.09	0.00	-0.10	0.00	0.13	0.00

Continue...

RJ_M	-0.02	0.15	0.00	0.22	-0.04	0.01
PR_U	0.35	0.00	-0.08	0.00	0.18	0.00
PR_M	0.14	0.00	0.00	0.10	0.03	0.05
SC_U	0.04	0.00	0.55	0.00	0.02	0.00
SC_M	0.05	0.00	0.00	0.57	0.00	0.83
RS_U	-0.10	0.00				
RS_M	0.00	0.73				
MS_U					-0.01	0.00
MS_M					0.00	0.78
DF_U			-0.03	0.00		
DF_M			0.00	0.71		
Private_uni_U	0.81	0.00			0.14	0.00
Private_uni_M	0.00	0.51			-0.01	0.21
IGC_U	-0.09	0.00	-0.02	0.00	-0.18	0.00
IGC_M	0.03	0.00	-0.03	0.00	0.04	0.00
Hours_U	0.01	0.68	0.03	0.01	0.19	0.00
Hours_M	-0.01	0.71	-0.06	0.00	0.05	0.18
Academic_sch_U	-0.18	0.00	-0.13	0.00	-0.11	0.00
Academic_sch_M	0.05	0.00	-0.02	0.00	0.01	0.43
Quota_U	-0.19	0.00	-0.14	0.00	-0.12	0.00
Quota_M	0.09	0.00	0.00	0.35	0.00	0.72

Source: Made from ENADE microdata 2019, INEP.

[3] Explanatory note on the reference categories of the covariates:'live with parents and/or relatives' is the category of reference to situation in home, 'College Degree or more' to parent's schooling, 'SP' to states

APPENDIX F – Analysis of Common Support

Table 33 – Number of Observations On Common Support and Off Common Support by Undergraduate Course

Course	Observations					
	Off Sup.	**On Sup.**	**Total**	**Dist. Off**	**Dist. On**	**Dist. Tot.**
Pedagoy	978	108001	108979	33	55234	55267
History	112	13220	13332	112	3794	3906
Phy. Ed. DE.	3	26614	26617	3	3746	3749
Business	4	96784	96788	4	23096	23100
Accounting	8	51644	51652	8	12517	12525
Social Work	0	20875	20875	0	10221	10221
Tech. Env. Man.	620	3269	3889	620	2035	2655
Phy. Ed. BA.	108	30861	30969	108	8233	8341
Prod. Eng.	19	23008	23027	19	1798	1817

Source: Made from ENADE 2017, 2018 and 2019 microdata, INEP.

APPENDIX G – Complete sensitivity analysis tables

Table 34 – Sensitivity Analysis for Pedagogy in 2017.

Gamma	**Final**		**General**		**Specific**	
	p_mh+	**p_mh-**	**p_mh+**	**p_mh-**	**p_mh+**	**p_mh-**
1	0.00	0.00	0.00	0.00	0.00	0.00
1.5	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00
2.5	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00
3.5	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00
4.5	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00

Source: Made from ENADE 2017 microdata, INEP.

Note p_mh+: significance level (assumption: overestimation of treatment effect).

Note p_mh-: significance level (assumption: underestimation of treatment effect).

Table 35 – Sensitivity Analysis for History in 2017.

Gamma	**Final**		**General**		**Specific**	
	p_mh+	**p_mh-**	**p_mh+**	**p_mh-**	**p_mh+**	**p_mh-**
1	0.00	0.00	0.17	0.17	0.00	0.00
1.5	0.00	0.00	0.00	0.00	0.00	0.01
2	0.00	0.00	0.00	0.00	0.00	0.00
2.5	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00
3.5	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00
4.5	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00

Source: Made from ENADE 2017 microdata, INEP.

Note p_mh+: significance level (assumption: overestimation of treatment effect).

Note p_mh-: significance level (assumption: underestimation of treatment effect).

Table 36 – Sensitivity Analysis for Physical Education Degree in 2017.

Gamma	**Final**		**General**		**Specific**	
	p_mh+	**p_mh-**	**p_mh+**	**p_mh-**	**p_mh+**	**p_mh-**
1	0.00	0.00	0.00	0.02	0.00	0.00
1.5	0.00	0.00	0.00	0.23	0.00	0.00
2	0.00	0.03	0.00	0.00	0.00	0.01
2.5	0.00	0.04	0.00	0.00	0.00	0.12
3	0.00	0.00	0.00	0.00	0.00	0.00
3.5	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00
4.5	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00

Source: Made from ENADE 2017 microdata, INEP.

Note p_mh+: significance level (assumption: overestimation of treatment effect).

Note p_mh-: significance level (assumption: underestimation of treatment effect).

Table 37 – Sensitivity Analysis for Business in 2018.

Gamma	**Final**		**General**		**Specific**	
	p_mh+	**p_mh-**	**p_mh+**	**p_mh-**	**p_mh+**	**p_mh-**
1	0.00	0.00	0.00	0.00	0.00	0.00
1.5	0.00	0.02	0.00	0.00	0.00	0.30
2	0.00	0.00	0.00	0.00	0.00	0.00
2.5	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00
3.5	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00
4.5	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00

Source: Made from ENADE 2018 microdata, INEP.

Note p_mh+: significance level (assumption: overestimation of treatment effect).

Note p_mh-: significance level (assumption: underestimation of treatment effect).

Table 38 – Sensitivity Analysis for Accounting in 2018.

Gamma	Final		General		Specific	
	p_mh+	p_mh-	p_mh+	p_mh-	p_mh+	p_mh-
1	0.00	0.00	0.36	0.36	0.00	0.00
1.5	0.00	0.31	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00
2.5	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00
3.5	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00
4.5	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00

Source: Made from ENADE 2018 microdata, INEP.

Note p_mh+: significance level (assumption: overestimation of treatment effect).

Note p_mh-: significance level (assumption: underestimation of treatment effect).

Table 39 – Sensitivity Analysis for Social Work in 2018.

Gamma	Final		General		Specific	
	p_mh+	p_mh-	p_mh+	p_mh-	p_mh+	p_mh-
1	0.00	0.00	0.00	0.00	0.00	0.00
1.5	0.00	0.00	0.00	0.37	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00
2.5	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.35	0.00	0.00	0.00	0.03
3.5	0.00	0.00	0.00	0.00	0.00	0.07
4	0.00	0.00	0.00	0.00	0.00	0.00
4.5	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00

Source: Made from ENADE 2018 microdata, INEP.

Note p_mh+: significance level (assumption: overestimation of treatment effect).

Note p_mh-: significance level (assumption: underestimation of treatment effect).

Table 40 – Sensitivity Analysis for Technology in Environmental Management in 2019.

Gamma	**Final**		**General**		**Specific**	
	p_mh+	**p_mh-**	**p_mh+**	**p_mh-**	**p_mh+**	**p_mh-**
1	0.00	0.00	0.02	0.02	0.00	0.00
1.5	0.00	0.05	0.00	0.36	0.00	0.11
2	0.00	0.49	0.00	0.01	0.00	0.35
2.5	0.00	0.08	0.00	0.00	0.00	0.04
3	0.00	0.01	0.00	0.00	0.00	0.00
3.5	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00
4.5	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00

Source: Made from ENADE 2019 microdata, INEP.

Note p_mh+: significance level (assumption: overestimation of treatment effect).

Note p_mh-: significance level (assumption: underestimation of treatment effect).

Table 41 – Sensitivity Analysis for Physical Education Bachelor in 2019.

Gamma	**Final**		**General**		**Specific**	
	p_mh+	**p_mh-**	**p_mh+**	**p_mh-**	**p_mh+**	**p_mh-**
1	0.00	0.00	0.00	0.00	0.00	0.00
1.5	0.00	0.00	0.00	0.41	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00
2.5	0.00	0.09	0.00	0.00	0.00	0.09
3	0.00	0.00	0.00	0.00	0.00	0.00
3.5	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00
4.5	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00

Source: Made from ENADE 2019 microdata, INEP.

Note p_mh+: significance level (assumption: overestimation of treatment effect).

Note p_mh-: significance level (assumption: underestimation of treatment effect).

Table 42 – Sensitivity Analysis for Production Engineering in 2019.

Gamma	**Final**		**General**		**Specific**	
	p_mh+	**p_mh-**	**p_mh+**	**p_mh-**	**p_mh+**	**p_mh-**
1	0.00	0.00	0.06	0.06	0.00	0.00
1.5	0.00	0.04	0.00	0.00	0.00	0.08
2	0.00	0.00	0.00	0.00	0.00	0.00
2.5	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00
3.5	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00
4.5	0.00	0.00	0.00	0.00	0.00	0.00

Source: Made from ENADE 2019 microdata, INEP.

Note p_mh+: significance level (assumption: overestimation of treatment effect).

Note p_mh-: significance level (assumption: underestimation of treatment effect).