

Institutional Typologies and Quantitative Analyses on Higher Education

Tipologias Institucionais e Análises Quantitativas sobre a Educação Superior

Tipologías Institucionales y Análisis Cuantitativas sobre la Educación Superior

Santos, Renato Augusto dos ⁽ⁱ⁾

Raposo, Bruno Marão ⁽ⁱⁱ⁾

Carrijo, Maria Clara Silva Cesar ⁽ⁱⁱⁱ⁾

⁽ⁱ⁾ Instituto Nacional de Estudos e Pesquisas Educacionais Anísio – INEP, Diretoria de Avaliação da Educação Superior – DAES, Brasília, DF, Brasil. Universidade Federal do Rio de Janeiro – UFRJ, Programa de Pós-Graduação em Sociologia e Antropologia – PPGSA, Rio de Janeiro, RJ, Brasil. <https://orcid.org/0009-0001-5909-3024>, br.rsan@gmail.com.

⁽ⁱⁱ⁾ Universidade Federal do Rio de Janeiro – UFRJ, Programa de Pós-Graduação em Sociologia e Antropologia – PPGSA, Rio de Janeiro, RJ, Brasil. Universidade de Brasília, UnB – Brasília, DF, Brasil. <https://orcid.org/0009-0005-3662-5525>, brunoraposo2@gmail.com.

⁽ⁱⁱⁱ⁾ Universidade Federal do Rio de Janeiro – UFRJ, Programa de Pós-Graduação em Sociologia e Antropologia – PPGSA, Rio de Janeiro, RJ, Brasil. Universidade de Brasília, UnB – Brasília, DF, Brasil. <https://orcid.org/0009-0009-6143-1123>, clara.carrijo@gmail.com.

Abstract

The production of scientific studies or studies that support the formulation and implementation of public policies in each Higher Education (HE) system must rely on information about its own internal dynamics. Usually, such analyses take as reference categories defined by the formal structure of HE, which are not capable of univocally guiding the strategic and pedagogical decisions of the sponsors and administrators of Higher Education Institutions (HEIs). In this regard, the application of statistical methods is addressed as another way of approaching HE dynamics, allowing for a perspective that goes beyond the structure set forth in current legislation, by unveiling analytical frameworks delimited from data that are representative of the reality of HEIs..

Keywords: Institutional Typology, Higher Education, Multivariate Analyses, Factor Analysis, Cluster Analysis.

Resumo

A produção de estudos científicos ou subsidiários à formulação e à implementação de políticas públicas em cada Educação Superior (ES) deve contar com informações da sua própria dinâmica interna. Usualmente, essas análises tomam como referência recortes definidos a partir das categorizações previstas na estrutura formal da ES, que não são capazes de direcionar de forma unívoca as decisões estratégicas e pedagógicas de mantenedores e gestores das Instituições de Educação Superior (IES). Nessa direção, aborda-se a aplicação de métodos estatísticos como outra forma de aproximação das dinâmicas da ES, permitindo um olhar que transponha a estrutura prevista na legislação vigente, ao desvelarem-se recortes de análise delimitados a partir dos dados representativos da realidade das IES.

Palavras-chave: Tipologia Institucional, Educação Superior, Análises Multivariadas, Análise Fatorial, Análise de Clusters.

Resumen

La producción de estudios científicos o que respalden a la formulación e implementación de políticas públicas en cada Educación Superior (ES) debe basarse en información sobre su propia dinámica interna. Habitualmente, estos análisis se apoyan en recortes definidos según las categorizaciones definidas para la estructura formal de la ES, que no orientan de forma unívoca las decisiones estratégicas y pedagógicas de los gestores y entidades mantenedoras de las Instituciones de Educación Superior (IES). En este sentido, se plantea la aplicación de métodos estadísticos como otra forma de aproximarse a las dinámicas de la ES, permitiendo una mirada que trascienda la estructura legal vigente, al revelar recortes analíticos delimitados por datos representativos de la realidad de las IES.

Palabras clave: Tipología Institucional, Educación Superior, Análisis Multivariantes, Análisis Factorial, Análisis de Clústeres.

Introduction

The present text aims to outline the methodological framework adopted in the research entitled “*Measuring the relationship between institutional diversity and student equity in Latin American countries*”¹, more specifically regarding the section devoted to identifying institutional types through the application of statistical methods, focusing on the functioning of Higher Education Institutions (HEIs) within the higher education (HE) systems of each country — Argentina, Brazil, Chile, Peru, and Uruguay.

The decisions made in the application of multivariate analysis (MA) statistical methods to data concerning the functioning of Brazilian HEIs are presented, covering the stages from the application of MA to data analysis. The purpose is to present the design of the quantitative methodology applied in the research, including the proper characterization of the methods used and their applications in the ongoing process of knowledge production.

In order to define which data should compose the set of variables to be used in the statistical analyses, it was important to keep in mind the overall objective of the study: to assess the extent to which different institutional arrangements and types of institutions that have emerged within HE systems, as a result of expansion and diversification, have been capable of increasing equity in educational opportunities for access and participation of certain historically disadvantaged social groups — “equity groups”.

From the defined objective, two main actions derive, reflecting a common strategy in research that analyzes the different dynamics of expansion and inclusion in HE worldwide: (1) the creation of an institutional typology that allows for analyzing the organization of HEIs, and (2) its use to examine how different national education systems have expanded and what impacts they have produced on social inequalities, through the inclusion of students from historically disadvantaged groups.

¹ This research project is carried out within the scope of the Latin American Center for Research in Higher Education (CeLapes), based at the Brazilian College of Advanced Studies (CBAE), and the Higher Education Research Laboratory (Lapes), affiliated with the Institute of Philosophy and Social Sciences (IFCS) of the Federal University of Rio de Janeiro (UFRJ), while maintaining interaction with other related projects. The project is funded by the following entities: the Society for Research into Higher Education (SRHE), the National Council for Scientific and Technological Development (CNPq), and the Carlos Chagas Filho Foundation for Research Support of the State of Rio de Janeiro (FAPERJ).

This chosen focus is related to specific research themes on typologies and higher education (Huisman et al., 2015; Teixeira et al., 2013; Van Vught, 2009), social inequalities and higher education (Alon, 2009; Lucas & Moore, 2001), and the use of typologies to analyze stratification in higher education (Barbosa & Santos, 2011; Croxford & Raffe, 2014; Fumasoli & Huisman, 2013). The methodological approach based on quantitative methods was inspired by studies conducted by members of the higher education research group (Rodrigues, 2022; Vieira, 2021), who, in turn, drew inspiration from other academic works with similar approaches.

Institutional Typology Defined from Empirical Data

The use of categorizations in analyses involving a country's HE becomes indispensable when the number of variables and units of observation makes it unfeasible to maintain a general and comprehensive view of the set of programs and institutions at this educational level, thus requiring systematization and grouping of information, as is the case in Brazil. In such scenarios, categorization is a recurring and indispensable element in operational and academic studies, as well as in the implementation of public policy throughout its entire lifecycle.

Normally, these categorizations are based on the formal structure and certain operational aspects established in current legislation. While this approach has proven to be quite relevant, it may hinder the identification of specific dynamics within higher education. Consequently, from the standpoint of public authorities, it may take some time before critical issues are identified and preventive or remedial public policies are proposed, with a clearly defined object of intervention, even in a favorable political context.

An institutional typology of Brazilian HEIs, supported by official categorization, can be observed in the study *"Toward a Typology of Brazilian Higher Education: Concept Testing"* (Schwartzman et al., 2021). This study uses official categorizations concerning the administrative nature of the sponsoring organizations and the types of academic organization of HEIs as a starting point to present an institutional typology. Based on the official categorizations, secondary characteristics were highlighted in order to establish subcategories or subgroups of HEIs, making use of data related to undergraduate programs, *stricto sensu* graduate programs — master's and doctoral — and enrollment figures to define the institutional size of HEIs.

It cannot be denied that aspects of the formal structure in Brazil point to factors that tend to influence how HEIs operate within the National System of Higher Education (NSHE). However, as studies and experience have shown, the types of sponsoring organizations and the academic organization of HEIs are not sufficient to ensure that these institutions perform equivalently within HE or achieve similar results in their students' educational processes.

These differences in performance occur because regulatory provisions do not impose restrictive limits on the activities of HEIs, but only set minimum requirements to be met; they also grant internal management autonomy for HEIs to define their own strategies of action, in response to surrounding social demands as well as their own institutional or sponsoring interests. As a result, they do not univocally guide the strategic and pedagogical decisions of HEI sponsors and administrators.

The effects of sponsors' autonomy on the definition of strategic directions for their HEIs can be found in the forthcoming text *"The Formation of Oligopolies in Brazilian Higher Education"* (Zuccarelli et al., in press), which presents evidence that the formal structure regarding types of HEIs does not univocally determine the strategic choices concerning the performance of the country's HEIs.

Different forms of performance can also be identified when comparing HEIs with similar formal categorizations, such as private for-profit universities, or even in comparisons involving other classifications defined by official categorizations. This is precisely why studies supported by statistical analyses can enable the identification of HE dynamics that emerge in groups different from those established in current legislation.

This approach to the object of study has the potential to highlight aspects of the phenomenon that may be attenuated or even overlooked in analyses focused on identifying similarities and differences among HEIs, given that researchers' perspectives are often shaped by their personal experiences with the object of study (Chizzotti, 2003). From this perspective, this axis of the research is carried out, with the methodological framework presented in the following section of this text, aiming to identify new forms of categorizing HEIs beyond the institutional typologies defined in legislation. Accordingly, the data used in the statistical analyses concern the operational aspects of HEIs within HE.

Certainly, this is not to suggest that the use of the formal structure or the operational aspects described in legislation loses its relevance in analyses of a country's HE system, or that

one approach is inherently better than another. What is meant to be emphasized is that there are analytical techniques more suitable to specific research problems, to certain scopes, or to particular analyses that seek new ways of approaching reality in order to identify new nuances for examination.

Methodological Aspects of Quantitative Analyses

The quantitative approach in studies considers reality as something objective and measurable, seeking to quantify phenomena and employ statistical methods to test hypotheses, generally formulated before the collection of numerical data, which is carried out through appropriate instruments (Chizzotti, 2003). This approach also allows for the testing of emerging theories, the creation of new theoretical hypotheses, and the development of new theories (Creswell, 2010; Glaser & Strauss, 1967; Yin, 2016).

It further makes it possible to identify results that demonstrate patterns or correlations that either align with or contradict existing theories, highlighting anomalies that challenge consolidated or emerging theories (Creswell, 2010; Glaser & Strauss, 1967; Yin, 2016). In light of this, there arises the need to produce new theorization to explain these new nuances of the phenomenon under analysis (Yin, 2016). Its use by the researcher may also occur in complex situations or processes where the relevant variables are not clearly defined (Yin, 2016).

In research grounded in quantitative methods, qualitative methods are also applied during the exploratory phase in order to build the conceptual foundation that reveals the categories of analysis upon which quantitative methods are structured (Luna, 1998). This becomes even more relevant in complex and multifaceted contexts (Yin, 2016), such as the one addressed in the present research. Moreover, it enables the researcher to identify relevant variables, develop hypotheses, and design data collection instruments more suitable for subsequent statistical analyses (Chizzotti, 2003; Creswell, 2010).

The process of defining an institutional typology for Brazilian HE, resulting from this research and based on HEI operational aspects, is grounded in multivariate statistical analysis (MA) methods. These serve to measure, explain, and predict the degree of relationship between statistical variables, which must be random and interrelated in such a way that their effects cannot be interpreted separately (Hair et al., 2009).

To enable the analysis and interpretation of the large number of variables that make up the database created to characterize the operational aspects of HEIs, two multivariate statistical methods were chosen, each aimed at specific and distinct objectives, although with certain points of interaction, in order to identify through:

- a) Factor Analysis (FA) – dimensions of HE behavior, revealed by groups of HEI operational characteristics derived from the correlations among variables, which may serve as mediators in interpreting the groups of HEIs identified through clustering methods.
- b) Cluster Analysis (CA) – groups of HEIs that exhibit similar operational aspects, allowing for the identification of an institutional typology defined on the basis of these attributes of HEIs.

The application of these methods is possible through computational calculations using software that enables robust statistical analyses. For the purposes of this research, the software R and the statistical packages *psych*, *mclust*, and *tidyLPA* were used. The *psych* package was employed for factor analysis. For cluster analysis, the *mclust* package was chosen, as it showed better performance in the calculation process and offered more model options to test and provide compared to the *tidyLPA* package.

These methods rely on random algorithms that generate uncorrelated random variables, which can produce different results each time the method is reapplied to the same dataset. To mitigate this behavior, it is recommended to set a random seed in the programming so that the same results can be replicated every time the calculation is repeated.

Every application of a statistical method goes through several phases, which can be listed and described in different ways. Based on some authors (Crespo, 2022; Triola, 2013) and on experiences in quantitative research, the following can be identified: definition of the research problem or hypotheses to be tested, planning of the research design, data collection, data processing, data analysis, interpretation of results, and presentation of results.

The framework presented here, covering the development of the research from its planning stage to data analysis through MA, resulted from many iterations, including revisions of the guiding dimensions of data collection, the variables used in the study, and the statistical methods and techniques employed. This latter group of revisions was closely tied to the

computational calculations required for MA, which also encompassed the verification of theoretical hypotheses derived from statistical theory.

Given the weight of statistical methods in this research, a detailed study of the statistical methods and techniques defined for its development was undertaken. Accordingly, the study of statistical theories sought to verify the pertinence of using each method and technique according to the interests, objectives, and type of data selected for analysis, as well as to establish a solid theoretical basis for their application in the research.

Planning the Quantitative Approach of the Research

The planning stage of the study's development makes it possible to organize the process of knowledge production and to establish prior definitions regarding the measurable dimensions or characteristics of the phenomenon, the data sources, the instruments that will enable data collection, and the statistical techniques to be applied in the analyses.

In the context of the research presented here, this stage proved to be important, as it allowed for anticipating the resources required for its development and some potential operational difficulties, as well as defining the dimensions of the object of study to be used, the secondary data sources, and the statistical analyses to be applied, in line with the object of study and the research objectives.

Data Collection

Data collection can be characterized as direct, when obtained from a primary source, or indirect, when obtained from a secondary source. In this research, secondary sources were used, in which the collection involves data obtained for other purposes or from another related study conducted by another researcher or institution, through the application of instruments designed to meet their objectives (Crespo, 2002).

The choice of a given level of aggregation of the variables depends on the objectives of the study and the statistical analyses to be performed. However, when using secondary sources, it is ideal that the researcher be aware, at the time of defining the objectives and scope of the research, of what data are available and at what level of aggregation they are presented.

Otherwise, during the data collection phase, the objective and scope of the research may need to be adjusted or even completely reformulated.

From this perspective, the scope of the research, the breadth of the view of the object of study, or the focus of its statistical analyses end up being defined as a consequence of the data available or their levels of aggregation. This was the case in the present research, since not all operational aspects of HEIs are covered by the data collected or made available by the Federal Government. For example, it would have been desirable to find data of sufficient quantity and quality to represent extension activities and *lato sensu* graduate programs or of other types, beyond undergraduate and *stricto sensu* graduate programs, which ultimately was not possible.

As for the level of data aggregation, at the time of collection the data were found in official sources: (a) at their lowest level of aggregation, as microdata, and (b) at various levels of aggregation, as in the case of statistical summaries, where the data are presented according to the purposes defined for that type of data presentation. Thus, from this perspective, the research was not affected by the level of aggregation of the data available from the secondary source, making it possible to opt for the use of microdata.

The data collected from secondary sources were obtained through the Higher Education Census and the evaluation of *stricto sensu* graduate programs — master's and doctoral. Both cover the entire National System of Higher Education (NSHE) and are conducted, respectively, by the National Institute for Educational Studies and Research Anísio Teixeira (INEP), on an annual basis, and by the Coordination for the Improvement of Higher Education Personnel (CAPES), every four years, with annual updates of enrollment information.

The 2019 Higher Education Census provided enrollment and faculty data at the individual aggregation level, anonymized and presented in a way that did not allow for the reidentification of individuals without the application of more sophisticated data-mining techniques. By contrast, the most recent data from the four-year evaluation were collected at the program level of *stricto sensu* graduate programs, with enrollment information updated in 2019.

Many studies have focused on the conceptualization and measurement of institutional diversity, especially with regard to American and European higher education systems (HES). In this type of study, identifying relevant dimensions for analysis requires a conceptual understanding of the important characteristics of HEIs, which may be related to both theoretical

and epistemological considerations as well as to functional aspects of HES (Huisman et al., 2015).

To select the data to be used in the statistical analyses, dimensions of HEI functioning were defined that would allow for the identification and grouping of the available data. During the processes of literature review and quantitative data collection, other dimensions were considered and some were rearranged to better reflect the functioning of HE in the Brazilian context and in other countries.

Accordingly, based on the analysis of the literature and current legislation, and in line with the study's defined objective, six dimensions were established to guide the selection of available official data: (a) governance, (b) teaching, (c) research, (d) extension or third mission, (e) internationalization, and (f) size-related aspects of HEIs, which unfolded into subcategories of analysis.

Data Preparation

Once data collection was completed, the data preparation phase began, during which: (i) the quality of each collected or calculated variable was verified; (ii) new variables were calculated from the collected data, aggregated at the HEI level; (iii) the calculated variables to be used in MA were selected; and (iv) the variables were rescaled for the purposes of CA.

Regarding the level of data aggregation, although the intention was always to work at the HEI level, some measures were based on aspects of undergraduate programs, such as the mode of delivery or the field of knowledge of undergraduate education, according to the Cine Brasil classification (INEP, 2019).

Most of the calculated variables were based on the 2019 Higher Education Census microdata, reflecting aspects of HEIs, always with the aim of representing the three roles or obligations of HEIs in relation to teaching, extension, and research. The variables calculated from the four-year evaluation data, in turn, referred to the provision of master's or doctoral programs, enrollment figures linked to these programs, and their bibliographic output.

For the research, 156 variables were calculated, of which only 60 were selected for MA, according to the most recent revision of the variable list carried out by the researchers. In addition to criteria of relevance, original data source, and quality of the variables obtained, the

type of variable — quantitative or qualitative — was also considered. The latter criterion was due to the complications that qualitative variables introduce into the operationalization of MA². Since none of these variables were deemed indispensable for representing HEI operational aspects, all qualitative variables were excluded.

Additional decisions were also made regarding variable selection, specifically related to:

- a) Official categorizations established for HEIs, considering that: (i) as discussed earlier, these do not directly reflect how HEIs operate within HE; and (ii) their inclusion in MA could produce undesirable effects on the grouping of HEIs, as they may act as aggregating variables of the observation units due to their low variability.
- b) Indicators of inclusion of students with historically disadvantaged socioeconomic profiles, which were deemed unsuitable for defining the institutional typology targeted by this research, given the circular effect they would produce in subsequent analyses focused on the relationship between HEI functioning and equity of access and participation of such students.

The quality of the original and calculated variables was verified through descriptive statistical analyses, aimed at identifying variables that should not be included in MA due to the characteristics of their values or the absence of variability. This was done in order to mitigate unfavorable impacts on MA, in line with recommendations from the statistical literature (Alzahrani et al., 2021; Haynes, Richard & Kubany, as cited in Laros, 2012).

This step was necessary because variables that provide little information can generate noise or problems for both FA and CA, since they exhibit low variability and, consequently, low correlation with other variables. This may result in a singular correlation matrix, whose determinant is null and therefore not invertible. This issue required careful attention and ultimately led to the removal of one of the sixty variables selected for MA, namely the proportion of enrollments in the Cine Brasil field called “Services”.

This verification was necessary because FA and CA both rely on the correlation matrix in their calculation processes, though with different purposes: (i) FA uses it to allocate the

² Although statistical methods exist to transform qualitative variables into quantitative ones, such as optimal scaling, no application in similar studies was identified in the literature reviewed. Therefore, for all statistical procedures within the scope of this research, it was decided not to use methods and techniques beyond those previously tested in the types of multivariate analysis employed in the study.

variable to a factor, with low-information variables either failing to load onto any factor or introducing noise into the allocation process; and (ii) CA uses it to define the type of model to be applied in grouping the observation units, depending on whether variability is considered for parameters related to volume, shape, and orientation. In this case, low-information variables may result in singularity and render the use of certain models unfeasible.

This verification was necessary because FA and CA both rely on the correlation matrix in their calculation processes, but with different purposes: (i) FA uses it to allocate the variable to a factor, where low-information variables may either fail to load onto any factor or introduce noise into the allocation process; and (ii) CA uses it to determine the type of model to be applied in grouping the observation units, depending on whether variability is considered for parameters related to volume, shape, and orientation. In such cases, low-information variables may lead to singularity, rendering the use of certain models unfeasible.

Focusing further on variable quality, correlation analyses were performed among the calculated and selected variables for MA, using the Pearson method since all were quantitative and continuous. This was done because the absence of correlations or collinearity among variables can result in a singular correlation matrix, which produces undesirable consequences in both FA and CA.

The choice of correlation method depends on the type of variable selected for the study: Pearson for continuous quantitative variables, and polychoric for ordinal qualitative variables. When both quantitative and qualitative variables are present, it is recommended to exclude the qualitative variables or, if their inclusion in MA is indispensable, to transform them into dichotomous variables. In such cases, each possible category of the original variable is represented, with the attribute originally assumed coded as 1 (one) and the other attributes coded as 0 (zero), after which the Pearson method is applied (Hair et al., 2019).

Another data processing step to be considered is the rescaling of variables³, when there are exorbitant differences among the values of the data to be used in CA. Very disparate scales can generate noise in the clustering process by introducing unequal weights into the estimates, thereby hindering the formation of clusters (Bouvyron et al., 2019). Although this need is

³ In the literature, other forms of variable transformation can be found, such as logarithmic and exponential. However, as previously stated, it was decided not to use methods and techniques beyond those previously tested in the types of multivariate analysis employed in the research.

sometimes associated with FA, the idea lacks statistical grounding, since such transformation does not affect correlation analyses and, consequently, does not alter the results obtained.

When rescaling is necessary for CA, provided that the variable in question is not geospatial, among the available rescaling methods, those that result in a standard deviation of 1 (one) for the variables should be used (Bouvyron et al., 2019). This was the case in the present research, where very large differences were found among the scales of the variables selected for CA. The standardization method⁴ was therefore applied to mitigate potential noise in the clustering process.

Once the data processing steps described here were completed, it was then possible to proceed with the application of MA proper.

Factor Analysis

Factor Analysis (FA) is a method used to investigate multiple variables through their patterns and relationships, organizing them into a set of factors. These factors correspond to a minimal set of variables not directly observed, but which represent the grouping of the original variables most strongly correlated with each other and directly related to the dimensions or constructs highlighted in the object of study (Johnson & Wichern, 2007; Matos & Rodrigues, 2019).

From this perspective, FA seeks to facilitate the study of a large number of variables by reducing them into explanatory factors (Johnson & Wichern, 2007), including determining whether such reduction is possible at all. However, the dimensionality reduction resulting from FA may cause a significant loss of data variability — something undesirable in MA — making it necessary to verify the effectiveness of the reduction in representing the original variables in subsequent analyses.

The factors are linear combinations, created from the correlation matrix of the original variables, producing a score for each observation unit. To construct the factors, the method assigns factor loadings, which range from -1 to 1 , to determine the allocation of each variable

⁴ The method of variable standardization consists of transforming the original measures into values referenced by the distance between the original value and the mean of all values of the variable, with this difference divided by the standard deviation.

to its respective factor. These loadings are considered significant when their absolute value is equal to or greater than 0.35, making the factor a useful representative of the variable. Factor loadings, in turn, can be interpreted as regression weights of the measured variables for predicting the underlying construct, representing the extent to which the original variable is associated with a given factor (Laros, 2012).

There are two types of FA, differentiated by the hypotheses established a priori and their objectives (Laros, 2012; Matos & Rodrigues, 2019), as explained below:

- a) Exploratory Factor Analysis (EFA) – uses the data as a reference to define the factors (or dimensions) that group the variables, without relying on any a priori hypothesis. It is considered an exploratory approach because the researcher has no firm prior expectation regarding the factors to be created.
- b) Confirmatory Factor Analysis (CFA) – uses the data to verify the dimensions or constructs of the object of study previously defined on the basis of existing theories or empirical experience, testing a prior grouping hypothesis.

Depending on the type of FA applied, the absolute value of the factor loading has different interpretations. In EFA, when a variable does not obtain a significant factor loading on any factor or is the sole variable in a factor, it is considered unrelated to the other variables and should be excluded from the analysis set. The EFA must then be reapplied to the remaining subset, and this procedure should be repeated until all remaining variables are allocated to some factor (Laros, 2012). In CFA, however, this indicates that the previously defined hypothesis was not empirically confirmed, and the same procedure used in EFA does not apply.

The application of FA as a means of variable reduction is common in this type of study in order to: (i) enable subsequent computational calculations, when data overload prevents software from producing results; (ii) adhere to the principle of parsimony, aiming to avoid overfitting in modeling processes of a population based on samples, thereby mitigating interferences that the original set of variables may generate in subsequent MA; or (iii) make possible the analysis of results from other MAs involving a large number of variables⁶.

⁵ Other values found in the literature may be employed as analysis criteria, but the indication of the aforementioned author was used in applying this statistical method within the scope of this research.

⁶ FA can also be applied to address cases of multicollinearity among variables, when this type of relationship becomes a problem for the statistical method being used, which is not the case for CA.

None of these cases applies to this research, considering that: (a) the use of the *Mclust* package solved the issue of computational calculation, making it possible to use the entire universe of HEIs in the NSHE for cluster analysis; (b) the loss of explanatory power of the variables in the dimensionality reduction process indicated that it would be better to use the original raw dataset; (c) the use of factors, despite the loss of explanatory power, would only be acceptable if adherence to the principle of parsimony were indispensable, which is not the case here since the entire population, rather than a sample, is being used; and (d) the use of factors, despite the loss of explanatory power, with the aim of facilitating the analysis of CA results in the next stage of the statistical methodology, did not prove advantageous, given the high number of variables that were not aggregated into any factors.

Another purpose associated with EFA derives directly from the correlations embedded in the structure of this method, which allow for the identification of groups of variables that indicate trends within the set of observation units. In the context of this research, EFA has this application, making it possible to identify HE behaviors, evidenced by the groups of HEI operational characteristics highlighted from the correlations among the variables selected for the study's development.

Furthermore, the dynamics present in the NSHE, identified through the similar behavioral patterns of the variables related to HEI functioning as expressed by the factors, can serve as a framework for analyzing and interpreting the HEI groups resulting from CA. From this perspective, although it is not an application with guaranteed success — being more of an additional resource for results analysis — the use of factors can serve to more easily identify the similarities that led to the formation of HEI groups through CA.

In general, the application of FA in a study should follow the following steps, involving the dataset selected for this analysis: (1) verification of prerequisites and criteria for methodological adequacy; (2) definition of the type of factor extraction; (3) definition of the minimum number of factors; (4) execution of FA without rotation; (5) interpretation of the relationships among the created factors in order to define the type of model rotation to be applied; (6) execution of FA with the defined type of rotation; and (7) analysis of FA results.

Criteria for Method Applicability

For the application of FA, a prerequisite is the existence of correlations among the variables to be reduced into factors. These correlations should guide the correct selection of variables to be reduced, including the identification of those that should be removed for introducing noise or lacking significant correlation.

With regard to the correlation matrix, two additional tests are used to verify the feasibility of applying FA to a given dataset: the Kaiser-Meyer-Olkin (KMO) test — also known as the measure of sampling adequacy (MSA) — applied in analyses involving both samples and populations, and Bartlett's Test of Sphericity (Hongyu, 2018).

The KMO test suggests whether the proportion of variance in the items can be explained by a latent variable or factor, indicating how appropriate FA is for the dataset. The criteria for KMO indices are as follows: values below 0.5 are considered unacceptable; values between 0.5 and 0.7 are considered mediocre; values between 0.7 and 0.8 are considered good; and values above 0.8 and 0.9 are considered, respectively, excellent and outstanding⁷. Furthermore, based on the overall KMO value, it is possible to calculate a value for each variable — the MSA — which can be used as an additional criterion for analysis. Variables with an MSA below 0.6 may be removed in an attempt to increase the overall KMO value (Garson, 2022).

Bartlett's Test of Sphericity is a method that assesses the extent to which the covariance matrix resembles an identity matrix, in which the covariance terms among the variables are equal to zero. Rejection of the null hypothesis indicates that the covariance matrix contains significant values that can be grouped into factors and, therefore, that the set of variables analyzed can be used for the application of FA (Hongyu, 2018).

Definition of the Number of Factors and the Method of Factor Extraction

Once the feasibility of applying FA to the set of variables selected for the study has been verified, two methodological decisions are required, independent and not sequential: the number of factors and the method of factor extraction.

⁷ Other values found in the literature may be employed as analysis criteria, but the indication of the aforementioned author was used in applying this statistical method within the scope of this research.

To determine the number of factors needed, methods based on the eigenvalues of each factor may be used, with Horn's Parallel Analysis considered the best method for this purpose (Laros, 2012). This method employs correlation matrices calculated from uncorrelated random variables with the same dimension as the original dataset. Based on the eigenvalues of both the original matrix and the simulated matrices — represented by a critical value for each factor — two lines are plotted in a Scree Plot, where the point of intersection indicates the number of factors to be extracted, with the most suitable number being the one immediately preceding the intersection point of the lines.

Regarding the definition of the factor extraction method, two are most widely used depending on the type of data distribution: (i) maximum likelihood (ML) for normally distributed data, in which extraction is performed through parameter estimation, identifying those most likely to reproduce the observed correlation matrix when the dataset exhibits a multivariate normal distribution (IBM, 2021); and (ii) principal components, when the data do not follow a normal distribution (Costello & Osborne, 2005).

Although Principal Component Analysis (PCA) is itself a method of variable reduction, it can also be used in the factor extraction process (Laros, 2012). From this perspective, PCA seeks to explain the greatest proportion of the original information (variance) with a minimal number of factors. Thus, through spectral decomposition, factors are extracted that explain decreasing proportions of variance: the first factor explains the highest percentage, the subsequent factor explains the next highest, and so on. In this way, PCA makes it possible to obtain factor loadings in terms of eigenvalue–eigenvector relationships (Hair et al., 2009).

To reach this conclusion, the Shapiro–Wilk Test was used to verify the type of distribution of the data for each variable, with a significance level of 5%. The null hypothesis of normality was rejected for all variables (Shapiro & Wilk, 1965).

Model Rotation

In the process of generating factors, it is desirable that all have positive loadings and that the variables are not concentrated in only a few factors, in order to achieve a simple factor structure. When this does not occur, it becomes necessary to apply a transformation to the factor matrix and the factor loading matrix through rotation (Laros, 2012).

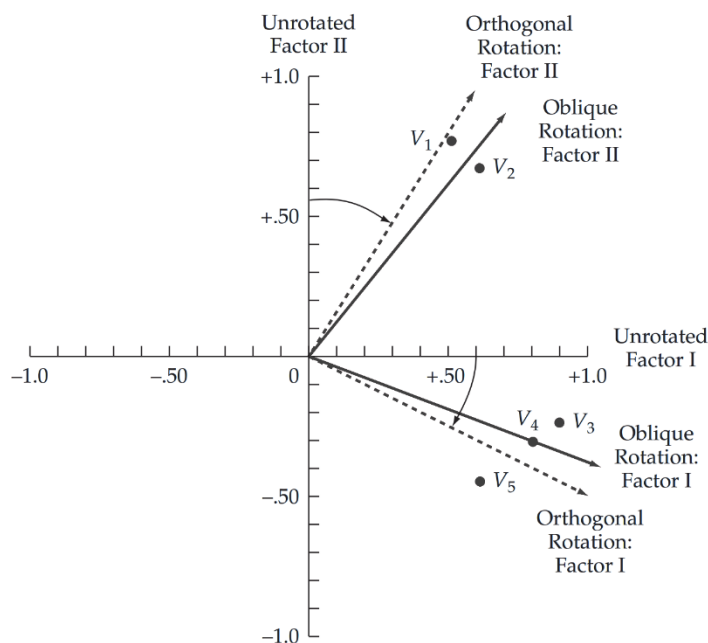
FA is a method similar to regression, where the response variable corresponds to the measured variable and the explanatory variables correspond to the factors created, whose loadings represent the weights of each factor in explaining each measured variable. Applying the same procedure to all variables yields the matrix model of FA.

Rotation is performed by transforming the factor matrix and the factor loading matrix, using the identity matrix, which mathematically does not alter the model produced from the measured matrices (Manly & Alberto, 2017). Two forms of rotation are used:

- a) *Orthogonal rotation* – appropriate for factor models in which the common factors are considered independent. The most common procedure is Varimax, which aims to maximize the variance of factor loadings for each factor by increasing high loadings and decreasing low loadings. This maximization leads to better distinction of each variable associated with each factor, ensuring that each variable has a high loading on one factor and low loadings on others, thus avoiding the undesirable situation of variables having multiple high loadings across several factors. This procedure is the default option in almost all statistical packages and produces a reasonably simple factor structure in most situations (Laros, 2012).
- b) *Oblique rotation* – appropriate for factor models in which the factors may be correlated. The most common procedure is Promax, which, among the various available procedures, yields a simpler and more convincing structure that is better interpretable than an orthogonal solution (Laros, 2012).
- c) A graphical representation of these two forms of rotation in relation to the Cartesian plane axes is shown in Figure 1.

Figure 1

Rotation of the Cartesian Plane Axes in Orthogonal and Oblique Rotations



Source: Hair et al. (2009)

Cluster Analysis

Cluster Analysis (CA) is a method used to identify meaningful groups within the analyzed variables, forming groups in which the most similar observation units, based on their common characteristics, are aggregated, while those that differ the most are separated. As a result, the groups of observation units formed are homogeneous within themselves and heterogeneous in relation to the others (Bouveyron et al., 2019).

From this perspective, the method seeks only patterns or similar groups, without paying attention to the political or conceptual meaning of the variations identified in the data (Betarelli Junior & Ferreira, 2018). However, the conceptual aspects involved in the variable selection process ultimately become an important reference point in assigning meaning to the results obtained with CA, allowing for a better characterization of the clusters defined from the selected data.

CA methods range from those that are largely heuristic to those that adopt more formal procedures based on statistical models, generally following either a hierarchical strategy or one

in which observations are reallocated among provisional groups (Fraley & Raftery, 1998). Among the different types, the one that best fits the object of study and the set of variables selected for analysis should be chosen (Betarelli Junior & Ferreira, 2018). It must be kept in mind that each method involves a certain level of computational complexity and estimation errors that affect the results produced.

The application of CA plays an important role in this research, given the objective of defining an institutional typology for a given HE system based on the operational aspects of its HEIs. Consequently, it was necessary to analyze the most commonly used types of CA methods and identify the one that best suited the objectives and the data selected for the development of the research as a whole.

In the literature, a test was found involving three types of clustering methods in order to identify the one with the lowest estimation error: k-means, single-link, and model-based. It was observed that not all are applicable for clustering purposes when the distribution of observation units within clusters shows overlap or when the cluster shape is non-spherical⁸; in such cases, model-based clustering methods are more appropriate (Fraley & Raftery, 1998).

Thus, Model-Based Cluster Analysis (MBCA) is the method that provides the greatest accuracy in its groupings and estimates, relying on probability models and statistical inference. This clustering method employs normal distributions centered on the means of the points — which represent the observation units of each cluster — and associates a probability (p) with each point, indicating the likelihood of belonging to a given cluster. An observation unit is allocated to the cluster whose mean is closest to it (Bouveyron et al., 2019).

From this, it is also possible to establish the degree of uncertainty (g) of an observation unit belonging to a given cluster, using the following calculation: $g = 1 - p$. The results for the degree of uncertainty can be presented graphically to facilitate analysis and interpretation, with the option of applying a cutoff criterion to highlight only those observation units that exceed the established threshold. In this research, only HEIs with an uncertainty level (g) above 0.25 — i.e., 25% uncertainty of cluster membership — were graphically highlighted.

This method was chosen for the research due to the following aspects: (i) the higher accuracy it provides in the results; and (ii) the use of probability in its calculation process, which

⁸ This reference to the spherical shape is due to the fact that distributions of observation units in this format are easier to cluster, whereas those with more irregular shapes add greater complexity to the clustering process.

is particularly useful for studying observation units with uncertain group membership and for identifying outliers in cluster formation.

As with any statistical analysis, MBCA generates results based on the dataset used. Therefore, the similarities and differences identified among the observation units — which are fundamental for defining and composing clusters — are established solely for the analyzed dataset.

Thus, each variable included in or removed from the analysis has the potential to alter the delimitation and composition of the set of clusters. Moreover, the use of certain variables in MBCA may introduce bias into this process if such variables reflect a common characteristic with a strong aggregating effect on the observation units, due to their low variability compared to the other variables used in applying this method.

This was the case in the first results obtained from preliminary tests, in which the use of variables related to the academic organization of HEIs led to the delimitation and composition of clusters that mirrored the same formal structure established in current legislation, with only minor differentiating nuances. This could give the impression that HEIs with the same official academic categories would tend to exhibit similar operational aspects, which was not empirically confirmed.

As a result of these preliminary findings, discussions were resumed regarding which variables should, in fact, be used in MBCA, leading to the conclusion that: (i) conceptually, the official categorization variables of HEIs should not be among those selected to represent their operational aspects; and (ii) statistically, the low variability of these variables, compared with the others, led observation units to be grouped according to the similarity expressed by the academic category of HEIs.

Variable Processing for MBCA

The processing of variables for use in MBCA was addressed in an earlier section, since part of it applied to both FA and MBCA, making it more effective to present these aspects together, with information on how to proceed and which methodological options should be considered. However, it is worth briefly revisiting this topic here, presenting a broader operational perspective that can guide the application of MBCA in any study.

Accordingly, with a focus on MBCA, the following data treatments are usually considered, to be applied when necessary depending on the characteristics of the dataset used in the analysis:

- a) *Selection of variables*, related to the scope of the phenomenon under analysis, based on the literature, the researchers' experience, and the characteristics presented by the variables, in line with the assumptions required for the use of MBCA;
- b) *Verification of data quality*, with regard to the variability of each variable and the correlation among variables, considering the undesirable impacts of lack of variability or correlation in the calculation process, which may generate noise in estimations or render MBCA or part of its estimation models unusable; and
- c) *Reduction of variables*, to mitigate computational issues related to the volume of data, avoid overfitting in processes where a sample is used to make inferences about the population, or facilitate the interpretation of clusters in studies involving a large number of variables.

Choice of Model Type in the Application of MBCA

To delimit the groupings, MBCA offers 14 estimation models based on the covariance matrix, which, as shown in Table 1, relies on three parameters: (i) *volume* – defines whether the clusters will or will not have approximately the same number of observation units; (ii) *shape* – defines whether the clusters will or will not have the same variance; and (iii) *orientation* – defines whether the clusters are constrained to lie along one or multiple axes, which may be horizontal, vertical, or diagonal (Boehmke & Greenwell, 2020; Fraley, 1999).

Table 1

Parameterization of the covariance matrix for application in MBCA

MODEL	VOLUME	SHAPE	ORIENTATION	IDENTIFIER IN MCLUST
1	Equal	Equal	No orientation	EII
2	Variable	Equal	No orientation	VII
3	Equal	Equal	Axes	EEI
4	Variable	Equal	Axes	VEI
5	Equal	Variable	Axes	EVI
6	Variable	Variable	Axes	VVI
7	Equal	Equal	Equal	EEE
8	Equal	Variable	Equal	EVE
9	Variable	Equal	Equal	VEE
10	Variable	Variable	Equal	VVE
11	Equal	Equal	Variable	EEV
12	Variable	Equal	Variable	VEV
13	Equal	Variable	Variable	EVV
14	Variable	Variable	Variable	VVV

Source: Boehmke e Greenwell (2020)

The choices of the best model to be used in MBCA and the optimal number of clusters can, as a statistical criterion, be based on the Bayesian Information Criterion (BIC), where the highest value of this model-fit measure for the dataset indicates the best option to adopt (Scrucca et al., 2023).

Thus, to use the BIC, the first step is to: (i) determine the maximum number of clusters (m), which should be kept as small as possible considering the research objectives and the practicality of cluster analysis, using, when possible, some reference from the literature on the phenomenon under study; and (ii) generate fit measures for the 14 estimation models within the range of 2 to m clusters (Fraley & Raftery, 1998).

Given the degree of subjectivity involved in defining the maximum number of clusters, it may be useful to test different ranges of cluster sizes for a more robust analysis, since different models and optimal cluster quantities may emerge as m is varied across ranges.

Based on the BIC values obtained for the tested cluster ranges, the next step is to analyze and identify the best estimation model to be used in MBCA and the optimal number of clusters. In this regard, the *mclust* package presents BIC results for the 14 estimation models for each tested number of clusters in a line graph, allowing, for the dataset analyzed: (i) the identification of models that are applicable in MBCA; (ii) the comparison of BIC results to identify the model that best fits the data; and (iii) the determination of the optimal number of clusters to be adopted

in MBCA. In this graph, each line corresponds to an estimation model, and the x-axis shows the tested number of clusters, with the maximum point representing the best model and the optimal number of clusters.

As explained, it was necessary to revise the list of variables selected for the research and the statistical assumptions underlying the stages of applying MA methods in the present analysis. The latest revisions stemmed mainly from the results obtained in MBCA, particularly regarding model selection and the number of clusters, given the presence of unexpected outcomes for what was considered a solid dataset, such as: (i) absence of results for several estimation models, which limited the options available for applying MBCA; and (ii) a consistently increasing trend in the BIC value, such that the higher the m — the maximum number of clusters tested —, the better the measure became, making it impossible to determine the optimal number of clusters using BIC.

Investigating this occurrence highlighted the importance of the data collection and processing stage, particularly in relation to this type of issue, emphasizing the need to verify the quality of the variables and their distributions against the assumptions and prerequisites required for applying MA in a given study.

Since the central objective of this text is to explain the methodology defined and applied in the study of institutional typologies within the ongoing research project at the Higher Education Research Laboratory (Lapes), the results of the analysis processes, discussions on the composition of the clusters and their characterizations regarding the common aspects of the HEIs that comprise them, including the analysis process and respective outcomes, will be the subject of future publications.

Closing Remarks

The methodological aspects of the quantitative analyses involved in the research were presented, addressing both theoretical aspects of statistics and methodological choices made in line with the assumptions and prerequisites of MA, as well as the characteristics of the data used to approach the operational aspects of HEIs in Brazilian HE. The analysis process and the results obtained, to be presented in future publications, will focus on: (i) the dynamics of Brazilian HE highlighted through the dimensions revealed by EFA, and (ii) the institutional

typology of Brazilian HE based on the operational aspects of its HEIs, evidenced through the clusters identified with the application of MBCA.

In the process of methodological design and the application of MA, it was possible to reinforce perspectives and hypotheses considered in the research, as well as to systematize knowledge about MA methods and their application in studies of specific phenomena. In this regard, it was emphasized that application requires knowledge that goes beyond merely replicating similar successful experiences. This is because each dataset produces partial results throughout the MA process, which may vary significantly, requiring data processing and researcher decisions guided by the assumptions and prerequisites of the methods and by other statistical knowledge.

It was also underscored that care must be taken with the stages of data collection and processing, as with every other stage of a study, since even minimal errors can invalidate all results obtained with the application of MA methods, even when these are otherwise applied correctly. Furthermore, the importance of ensuring that variable selection is grounded in the literature on the subject, in researchers' knowledge of the phenomenon, and in the characteristics of the statistical methods to be applied later was reaffirmed. This is because an inadequate variable for the study can introduce noise into the estimation processes of MA methods and compromise or bias the generation of results.

It was possible to confirm the perspective that the HE systems of the other countries included in the research tend to present institutional typologies — based on the operational aspects of HEIs — different from the one to be defined for Brazil. This results from the way in which higher education is officially structured in those countries and from the data available on the functioning of their HEIs. However, considering the characteristics of MBCA, it will still be necessary to verify the feasibility of its application to the context of the other countries.

For this reason, some decisions concerning the research's lines of action are reaffirmed, particularly those focused on cross-country comparisons, with statistical analyses constituting one of these lines. The planned comparisons, considering the formal structures of the countries and the institutional typologies defined from the operational aspects of HEIs, are as follows: (i) between countries, involving the structures and operational aspects of higher education established in legal frameworks; and (ii) within each country, involving the institutional

typologies identified from empirical data in relation to the structures and operational aspects of higher education defined in legal frameworks.

Finally, it became clear that other institutional typologies can be derived from variable selections referencing different nuances of HEI operations or from different outcome dimensions, underscoring the importance of the methodological framework presented here for application in other research focusing on the dynamics and outcomes of a given HE system.

References

- Alzahrani, A. R. R., Beh, E. J., & Stojanovska, E. (2021). Model-based clustering with *mclust* R package: Multivariate assessment of mathematics performance of students in Qatar. In *24th International Congress on Modelling and Simulation*, Sydney, NSW, Australia, 5 to 10 December 2021. mssanz.org.au/modsim2021.
- Alon, S. (2009). The evolution of class inequality in higher education: Competition, exclusion, and adaptation. *American Sociological Review*, 74(5), 731–755.
- Barbosa, M. L. de O., & Santos, C. T. (2011). A permeabilidade social das carreiras do ensino superior. *Cadernos CRH*, 24(63), 535–554.
- Betarelli Junior, A. A. , & Ferreira, S. de F. (2018). *Introdução à análise quantitativa e aos conjuntos Fuzzy (fsQCA)*. Enap.
- Boehmke, B., & Greenwell, B. M. (2020). *Hands-on machine learning with R*. CRC Press.
- Bouvyron, C., Celeux, G., Govaert, G., & Lancelot, G. (2019). *Model-based clustering and classification for data science: With applications in R*. Cambridge University Press.
- Chizzotti, A. (2003). *Pesquisa em ciências humanas e sociais* (7. ed.). Cortez.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, 10(1), 7.
- Crespo, A. A. (2002). *Estatística fácil* (18. ed.). Saraiva.
- Creswell, J. W. (2010). *Research design: Qualitative, quantitative, and mixed methods approaches* (3. ed.). Sage Publications.

- Croxford, L., & Raffe, D. (2014). The iron law of hierarchy? Institutional differentiation in UK higher education. *Studies in Higher Education*, 40(9), 1625–1640.
- Fraley, C. (1999). Algorithms for model-based Gaussian hierarchical clustering. *SIAM Journal on Scientific Computing*, 20, 270–281.
- Fraley, C., & Raftery, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal*, 41(8), 578–588.
- Fumasoli, T., & Huisman, J. (2013). Strategic agency and system diversity: Conceptualizing institutional positioning in higher education. *Minerva*, 51(2), 155–169.
- Garson, G. D. (2022). *Factor analysis and dimension reduction in R: A social scientist's toolkit*. Routledge.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine.
- Hair, J. J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). *Análise multivariada de dados* (6. ed.). Bookman.
- Hair, J. J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8. ed.). Cengage Learning.
- Hongyu, K. (2018). *Análise fatorial exploratória: Resumo teórico, aplicação e interpretação*. *Engineering Science*.
- Huisman, J., Lepori, B., Seeber, M., Frølich, N., & Scordato, L. (2015). Measuring institutional diversity across higher education systems. *Research Evaluation*, 24(4), 369–379.
- IBM. (2021). Exploratory factor analysis: Extraction. Retrieved September 16, 2024. <https://www.ibm.com/docs/en/spss-statistics/beta?topic=analysis-exploratory-factor-extraction>
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). (2019). *Manual para classificação dos cursos de graduação e sequenciais: CINE Brasil*. Brasília, DF: INEP.
- Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6. ed.). Upper Saddle River, NJ: Prentice Hall.
- Laros, J. (2012). O uso da análise fatorial: Algumas diretrizes para pesquisadores. In Lígia Pasquali (Ed.), *Análise fatorial para pesquisadores* (cap. 7, pp. 163–184). Labore Editorial.

- Lucas, S. R., & Moore, M. R. (2001). Tracking inequality: Stratification and mobility in American high schools. *American Journal of Sociology*, 107(2), 538-540.
- Luna, S. V. (1998). *Metodologia da pesquisa: Princípios e técnicas*. Educ.
- Manly, B. F. J., & Alberto, J. A. N. (2017). *Multivariate statistical methods: A primer* (4th ed.). CRC Press.
- Matos, A.S., & Rodrigues, E. C. (2019). *Análise fatorial*. Enap.
- Rodrigues, L. A. L. (2022). *A estratificação horizontal nos cursos imperiais: os concluintes de engenharia, direito e medicina entre 2009 e 2017*. [Tese de Doutorado, Universidade Federal do Rio de Janeiro].
- Schwartzman, S., Silva Filho, R. L., & Coelho, R. R. A. (2021). Por uma tipologia do ensino superior brasileiro: teste de conceito. *Estudos Avançados*, 35(101), 153-188.
- Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2023). *Model-based clustering, classification, and density estimation using mclust in R*. CRC Press.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591-611.
- Teixeira, P., Rocha, V., Biscaia, R., & Cardoso, M. F. (2013). Competition and diversification in public and private higher education. *Applied Economics*, 45(35), 4949-4958.
- Triola, M. F. (2013). *Elementary statistics* (12. ed.). Pearson.
- Van Vught, F. A. (2009). Diversity and differentiation in higher education. In F. Van Vught (Ed.), *Mapping the higher education landscape: Towards a European classification of higher education* (pp. 1-16). Springer.
- Vieira, A. de H. P. (2021). *Estratificação no ensino superior e ingresso no mercado de trabalho no Brasil, 2009-2015*. [Tese de Doutorado, Universidade Federal do Rio de Janeiro].
- Yin, R. K. (2016). *Qualitative research from start to finish* (2. ed.). The Guilford Press.
- Zuccarelli, C., Vieira, A., Mendonça, L., & Blanco, F. (no prelo). A formação dos oligopólios da educação superior brasileira. In *Políticas de educación superior en América Latina: Expansión, diversificación y equidad*. CLACSO.

Notes

Preprint:

Previously published as a preprint on March 28th, 2025
<<https://doi.org/10.1590/SciELOPreprints.11543>>.

Submission Data:

Submitted for review on May 22nd, 2025; approved for publication on July 31st, 2025.

Corresponding Author:

Santos, Renato Augusto dos - SIG, Quadra 4, lote 327, Brasília - DF, Brasil, 70610-908.

Author Contributions

Dos Santos, Renato Augusto – Conceptualization (Lead), Formal analysis (Equal), Investigation (Equal), Methodology (Lead), Resources (Equal), Supervision (Lead), Validation (Equal), Visualization (Equal), Writing – original draft (Lead), Writing – review & editing (Equal).

Raposo, Bruno Marão – Conceptualization (Supporting), Data curation (Equal), Formal analysis (Equal), Investigation (Equal), Methodology (Equal), Resources (Equal), Software (Lead), Validation (Equal), Visualization (Equal), Writing – original draft (Supporting), Writing – review & editing (Equal).

Carrijo, Maria Clara Silva Cesar – Conceptualization (Supporting), Data curation (Equal), Formal analysis (Equal), Investigation (Equal), Methodology (Equal), Resources (Equal), Software (Supporting), Validation (Equal), Visualization (Equal), Writing – original draft (Supporting), Writing – review & editing (Equal).

Funding and Support:

National Council for Scientific and Technological Development (CNPq):
420395/2022-9

Carlos Chagas Filho Foundation for Research Support of the State of Rio de Janeiro (FAPERJ): E-26/200.863/2021

Society for Research into Higher Education: BARBOSA_RA22287SRHE

Data Availability

The authors will provide the research data upon request.

Manuscript Preparation and Text Editing:

Bibliographic normalization (APA 7th ed.), manuscript preparation and text editing in Portuguese: Vera Lúcia Fator Gouvêa Bonilha <verah.bonilha@gmail.com>

English version and proofreading: Francisco López Toledo Corrêa <francisco.toledocorrea@gmail.com>

Responsible Editor:

Helena Sampaio <<https://orcid.org/0000-0002-1759-4875>>