

Análise de variáveis com potencial discriminante para classificação de escolas médicas

Analysis of discriminant variables for classification of medical schools

Gabriela Furst Vaccarezza¹ | gavaccarezza@gmail.com
Erik Montagna² | erik.montagna@fmabc.br
Rita Barradas Barata¹ | rita.barradasbarata@gmail.com
Nivaldo Carneiro Junior¹ | nicarneirojr@uol.com.br

RESUMO

Introdução: Este estudo metodológico pretendeu identificar em qual medida um conjunto de variáveis características de escolas médicas possui capacidade discriminante na classificação dos cursos por meio da análise de agrupamento. Nas últimas décadas, houve um aumento expressivo do número de vagas ofertadas para cursos médicos. Esse advento impôs desafios para os programas de avaliação, tanto pela necessidade de expandir o processo como pela necessidade de implantação de novos indicadores de qualidade.

Objetivo: Esse estudo teve como objetivo propor técnicas de análise para aprimorar a capacidade discriminante na classificação de cursos médicos mediante variáveis objetivas relacionadas a aspectos estruturais e operacionais que possam ser incorporadas aos métodos já utilizados.

Método: Trata-se de um estudo descritivo, analítico-metodológico e quantitativo que utilizou dados dos cursos médicos de São Paulo, em dezembro de 2020. Foi realizada análise por agrupamentos hierárquico e não hierárquico dos cursos para identificar variáveis com capacidade discriminante em busca de padrões que cooperem para a classificação das escolas médicas. As variáveis estudadas foram: início do curso, carga horária, regime letivo, metodologia, hospital universitário, categoria administrativa da instituição e gratuidade. Para a construção dos agrupamentos, adotaram-se o método de Ward e a distância euclidiana para estimar a discriminação entre os grupos. No agrupamento não hierárquico, a definição da quantidade de grupos foi determinada pela análise da diminuição da variância. Avaliou-se a correlação entre as variáveis por meio de mapas de calor.

Resultado: As análises de agrupamento mostraram a existência de três grupos de escolas médicas por similaridade: um grupo composto por escolas mais antigas e com maior carga horária, e, nos outros dois, consideraram-se as escolas não gratuitas sem hospital universitário, diferenciando-se pela idade das escolas. Além disso, as correlações reforçaram que as variáveis adotadas cooperavam para a discriminabilidade entre grupos. Há reconhecida heterogeneidade entre os cursos de graduação no Brasil, e esse dado também se aplica aos cursos médicos que impõem desafios metodológicos para os processos de avaliação estabelecidos. Entretanto, a inclusão de variáveis requer métodos capazes de refinar a capacidade discriminante da análise.

Conclusão: A análise aqui proposta mostrou-se capaz de identificar grupos de escolas médicas por meio de indicadores objetivos e pode auxiliar o processo de avaliação das escolas médicas.

Palavras-chave: Educação Médica; Escolas Médicas; Análise por Conglomerados.

ABSTRACT

Introduction: This is a methodological study that aims to identify the extent to which a set of variables characteristic of medical schools have a discriminating capacity to classify the courses through cluster analysis. In the last two decades, there has been a significant increase of vacancies in medical courses. This advent has posed challenges for evaluation programs, both because of the need to expand the evaluation process and the need to implement new quality indicators.

Objective: To propose analysis techniques to improve the discriminating capacity to classify medical courses through variables related to structural, operational, and objective aspects that can be incorporated into the already used methods.

Method: Descriptive, analytical-methodological, quantitative study that used data from existing medical courses in December 2020, in the state of São Paulo. Analysis by hierarchical and non-hierarchical clustering of courses was performed to identify the discriminating elements that provide standards that cooperate for the classification of medical schools. The studied variables were: course start, workload, academic regime, methodology, University-Hospital, administrative category of the institution, gratuity. For the construction of the clusters, the Ward method and the Euclidean distance were used to estimate the discrimination between the groups. In the non-hierarchical clustering, the definition of the number of groups was determined by the analysis of the decrease in variance. The correlation between the variables was also evaluated through heatmaps.

Results: The cluster analysis showed the existence of three groups of medical schools by similarity, with one group consisting of older schools with greater workload, and the other two consisting of private schools without a university-hospital, differing by the course time. Furthermore, the correlations reinforced that the adopted variables cooperated for the discriminability between the groups.

Discussion: There is a known heterogeneity among undergraduate courses in Brazil and this also applies to medical courses, which poses methodological challenges for the established assessment processes. However, the inclusion of variables requires methods capable of refining the discriminant capacity of the analysis.

Conclusion: The analysis proposed here proved to be capable of identifying groups of medical schools through objective indicators that can support the evaluation process of medical schools.

Keywords: Medical Education; Medical Schools; Cluster Analysis.

¹ Faculdade de Ciências Médicas da Santa Casa de São Paulo, São Paulo, São Paulo, Brasil.

² Centro Universitário FMABC, Santo André, São Paulo, Brasil.

Editora-chefe: Rosiane Viana Zuza Diniz. | Editor associado: Fernando Almeida.

Recebido em 28/02/22; Aceito em 10/04/23. | Avaliado pelo processo de double blind review.

INTRODUÇÃO

Trata-se de um estudo metodológico que pretende identificar em qual medida um conjunto de variáveis características de escolas médicas possui capacidade discriminante na classificação de cursos do estado de São Paulo por meio da análise de *cluster*. Na análise de *cluster*, é possível utilizar um número elevado de variáveis; mas também se pode limitar a quantidade de variáveis como forma de estimar quais têm mais peso na análise¹. Por tratar-se de uma análise exploratória de dados, é lícito o uso deliberado de conjuntos limitados de variáveis como forma de verificar diferenças e semelhanças no resultado final, com consequente possibilidade de gerar agrupamentos visando à identificação de padrões nos conglomerados de acordo com as variáveis definidas².

Especificamente com relação aos cursos de Medicina, houve um aumento expressivo do número de vagas ofertadas nas duas últimas décadas. Scheffer et al.^{3,4} apontam um aumento de 124,7% no número de vagas oferecidas de 2010 até 2020, e os cursos novos foram responsáveis por 62% dessas novas vagas. Scheffer et al.⁴ também apresentam dados acerca do aumento expressivo na quantidade de escolas médicas privadas, resultando em um ensino médico eminentemente privado.

A ampliação do número de cursos de Medicina é um processo que esteve presente em vários momentos da história do Brasil. A expansão e a distribuição das escolas médicas no Brasil podem ser consideradas um reflexo de políticas públicas⁵. A política expansionista foi influenciada pela lógica da organização do sistema de saúde, pela industrialização e pela crescente urbanização. Nesse contexto, a década de 1960 foi marcada pela criação de novas instituições de ensino superior (IES) e pelo aumento do número de vagas. O ensino superior passa a compor a economia de mercado, além de possibilitar a ascensão social por meio das graduações³ distribuição e expansão dos cursos de medicina no Brasil e descreveu os processos governamentais relacionados à ampliação da oferta de vagas. Trata-se de estudo descritivo, baseado em dados disponíveis no sistema do Ministério da Educação sobre escolas médicas. Com as informações sobre os cursos, desde os primeiros implantados, estabelecem-se seis períodos de governo para análise: de 1808 a 1963 (monarquia e governos republicanos iniciais⁶). Contudo, a expansão dos cursos médicos intensificou-se após a publicação da Lei nº 12.871/2013 que instituiu o Programa Mais Médicos (PMM)^{7,8}.

A avaliação das IES e de seus cursos tem papel estratégico para o desenvolvimento de ações governamentais. Nesse sentido, o aprimoramento e a incorporação de indicadores válidos e confiáveis e de métodos analíticos representam um desafio constantemente enfrentado pela academia e por gestores públicos⁹. Entretanto, essa é uma tarefa reconhecidamente

desafiadora, sobretudo no Brasil, pelas suas dimensões, particularidades e iniquidades regionais que se refletem em um ensino superior desigual⁹. Devem-se acrescentar a esse fato a criação de novos cursos e o aumento de vagas em cursos existentes para que o processo de avaliação das IES torne-se ainda mais complexo e demande instrumentos mais sofisticados para uma avaliação justa e equânime. Esses instrumentos devem contemplar maior quantidade de variáveis e de IES, e ainda assim possuir sensibilidade analítica que aprimore a capacidade discriminante dos métodos já adotados⁷. Ademais, é razoável que se busque um método de implementação simples, baixo requerimento computacional e fácil interpretação de resultados, elementos encontrados na análise de conglomerados (ou *clusters*). Assim, o objetivo do presente trabalho é identificar em qual medida um conjunto de variáveis preditoras possui capacidade discriminante na classificação de cursos médicos no estado de São Paulo por meio da análise de *clusters*.

MÉTODO

Trata-se de estudo metodológico descritivo, de abordagem quantitativa, que utilizou dados dos 67 cursos médicos publicados em dezembro de 2020, no portal eletrônico do Ministério da Educação (e-MEC).

A opção metodológica foi a análise de *cluster* dos cursos de Medicina do estado de São Paulo com posterior agrupamento das variáveis. As escolhas do estado se deu pelo fato de ele contribuir com um quinto de todas as vagas oferecidas nas escolas médicas do país¹.

A análise de *cluster* é uma denominação genérica dada a um conjunto de diferentes técnicas que podem ser utilizadas para classificar casos em grupos⁸. Na análise de *cluster*, não há conhecimento prévio sobre quais elementos pertencem a quais *clusters*.

As técnicas de análise de *clusters* ou agrupamentos são uma forma de análise de dados exploratória. São utilizadas quando se deseja verificar semelhanças e diferenças no padrão dos dados, observações em relação a determinadas variáveis e eventual existência de características que permitam o agrupamento dessas observações, podendo ser hierárquicos ou não hierárquicos. Nesse sentido, o objetivo dessa análise é estabelecer uma quantidade relativamente pequena de agrupamentos homogêneos internamente (conjunto das IES em um mesmo *cluster*), mas que os grupos sejam heterogêneos entre si e representem o comportamento conjunto das observações a partir das variáveis definidas. Ou seja, as observações de um determinado grupo devem ser relativamente semelhantes entre si, em relação às variáveis inseridas na análise, e consideravelmente diferentes das observações de outros grupos⁹.

A respeito dos cursos médicos, coletaram-se as seguintes variáveis: categoria administrativa da instituição de ensino, gratuidade, regime letivo, metodologia de ensino e data de autorização de início do curso. As variáveis presença de hospital universitário (HU) próprio e utilização da rede do Sistema Único de Saúde (SUS) como cenários de prática foram coletadas no *website* das IES. Para as IES que têm mais de um curso de Medicina em diferentes *campi*, cada curso foi considerado como um curso independente.

A escolha das variáveis é uma das etapas mais importantes na análise de *cluster*, já que devem ser incluídas apenas as variáveis que caracterizam os objetos a serem agrupados e especificamente relacionadas aos objetivos da análise. Em outras palavras, o desenho da pesquisa deve incluir apenas variáveis consideradas relevantes para classificar os casos¹⁰.

As variáveis foram divididas nas seguintes categorias para a análise: 1. data do início do curso de Medicina; 2. carga horária: até 15% além da carga horária mínima exigida por lei para um curso médico, de 16% a 30%, de 31% a 45% e mais de 45%; 3. regime letivo: anual ou semestral; 4. metodologia de ensino: tradicional, ativa ou mista; 5. HU: presente ou ausente; 6. categoria administrativa: em IES privada com fins lucrativos não gratuita, privada sem fins lucrativos não gratuita, pública gratuita e pública não gratuita. Vale ressaltar que as IES privadas sem fins lucrativos são constituídas sob a forma de associação ou fundação e não distribuem nenhuma forma de patrimônio auferido (lucros, excedentes operacionais, dividendos e ou bonificações). As IES públicas não gratuitas correspondem às instituições educacionais criadas por lei estadual ou municipal e que não sejam mantidas exclusivamente com recursos públicos, não sendo, portanto, gratuitas¹¹.

A análise de *clusters* se deu de duas formas: hierárquico e não hierárquico. O *cluster* hierárquico foi gerado pelo método de Ward, que é também conhecido como método do incremento das somas de quadrados e baseado na análise de variância. Nesse método, as somas dos quadrados entre os grupos e dentro deles em relação às p variáveis são utilizadas como critério de agrupamento. O princípio do método de Ward é aglomerar os grupos de tal forma que minimizem a soma dos quadrados dentro dos grupos, ou seja, a soma dos quadrados dos erros^{2,9}.

A indicação do número de agrupamentos (*clusters*) é feita *a priori*, ou seja, é um parâmetro definido para a análise. Assim, a escolha de quantos *clusters* é o ideal depende do julgamento do analista e do significado prático da separação dos grupos. Idealmente, sugere-se a menor sobreposição possível entre os agrupamentos. Entretanto, essa decisão não é aleatória e pode ser auxiliada pelo método do cotovelo (ou

elbow), no qual o total da soma dos quadrados dentro do *cluster* mede sua homogeneidade. Em outras palavras, espera-se que dentro do *cluster* a homogeneidade seja máxima e entre os *clusters* a homogeneidade seja mínima de forma a estimar qual é a quantidade ótima de grupos⁹. Assim, é definido o número ótimo de *clusters*.

Como forma de confirmar a análise de *clusters* hierárquica, foi realizada a análise de agrupamento não hierárquica, com as distribuições das observações (escolas médicas) no plano. Essa análise prevê a redução da dimensionalidade por partir de um número elevado de variáveis e gerar um gráfico que possa mostrar a distribuição das escolas considerando os pesos das variáveis. Também se realizaram simulações com os números de pontos centrais de agrupamento no plano (*seeds*), de forma a confirmar se a quantidade de *clusters* adotada após a análise do cotovelo seria mantida.

Além disso, fez-se uma análise de correlação bivariada entre todas as variáveis do estudo, apresentada em uma matriz de correlações. Adotou-se a correlação de Spearman por tratar-se de dados categóricos. Essa análise visou identificar outros padrões entre as variáveis dentro de cada subgrupo gerado na análise de *clusters*, em busca de mais detalhes sobre características potencialmente discriminadoras para as variáveis do estudo. Nesse caso, todas as correlações foram definidas como significativas para valores de $p < 0,05$, sendo apresentadas somente as correlações significativas.

As análises foram realizadas no *software* RStudio (versão 1.4.1717), pacotes *dendextend* e *FactoExtra*, e em Python (versão 3.7) com os pacotes *scipystats*, *matplotlib* e *seaborn*.

O presente estudo foi dispensado de apreciação por Comitê de Ética de acordo com a Resolução do Conselho Nacional de Saúde (CNS) nº 510/2016, já que utilizou dados secundários, disponíveis a acesso público nos termos da Lei nº 12.527, de 18 de novembro de 2011.

RESULTADOS

O Gráfico 1A demonstra a variância dos dados em relação ao número de *clusters*. O decaimento ou “cotovelo” evidenciado no gráfico sugere que o número ótimo de *clusters* pode ser de três ou quatro. Idealmente, o valor ótimo é aquele em que é observada uma queda menos acentuada na diferença entre cada um dos valores indicados no eixo y em relação ao valor seguinte no eixo x (número de *clusters*). Entretanto, o ganho de informação será maior quanto maior for a diferença entre os *clusters*. Assim, optou-se neste estudo por três *clusters* para reforçar o caráter discriminante da análise. Dois *clusters* forneceriam grupos mais heterogêneos, e quatro *clusters* produziriam uma dupla sobreposição entre *clusters* e um grupo de apenas duas escolas (Gráfico 1B).

Gráfico 1. Representação das análises realizadas para definir o número ótimo de *clusters* para classificar os cursos médicos do estado de São Paulo, em dezembro de 2020. Método *elbow* (1A) e análise por agrupamento não hierárquica (1B).

Gráfico 1A. Valor ideal do número de *clusters* definido por meio do método de agrupamento *elbow*: relação do número total de *clusters* com a similaridade *intra-cluster*

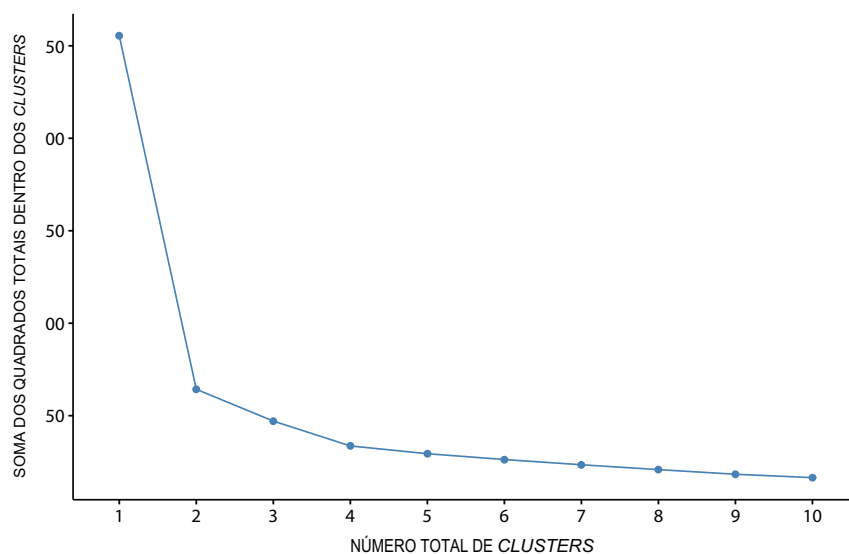
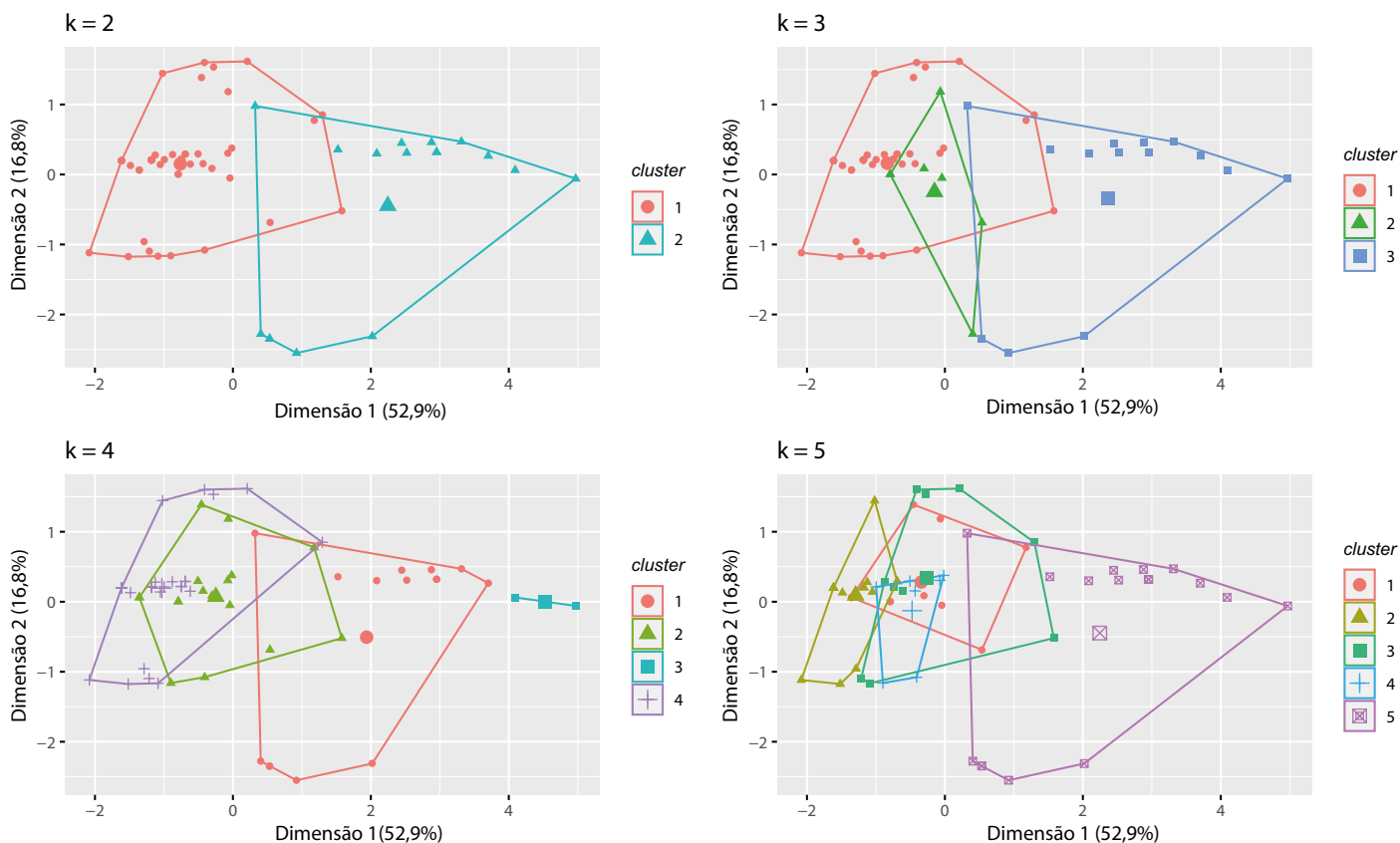


Gráfico 1B: Representação da classificação dos cursos de Medicina do estado de São Paulo, em dezembro de 2020, em dois ($k=2$), três ($k=3$), quatro ($k=4$) e cinco ($k=5$) *clusters*: análise de *cluster* não hierárquica



Fonte: Elaborado pelos autores.

Após determinado que para este estudo três *clusters* seria o número ideal, foi realizada a análise por agrupamento hierárquica. Essa análise apresenta a distância euclidiana calculada no eixo y e as observações (escolas médicas) no eixo x. A decisão quanto ao número de *clusters* depende da avaliação da existência de diferenças entre os agrupamentos e das semelhanças dentro de cada grupo de forma coerente. É importante ressaltar que a hierarquia aqui apresentada é entre as distâncias medidas e o grau de similaridades obtido. Para uma melhor leitura do dendrograma (Gráfico 2), os cursos médicos foram numerados.

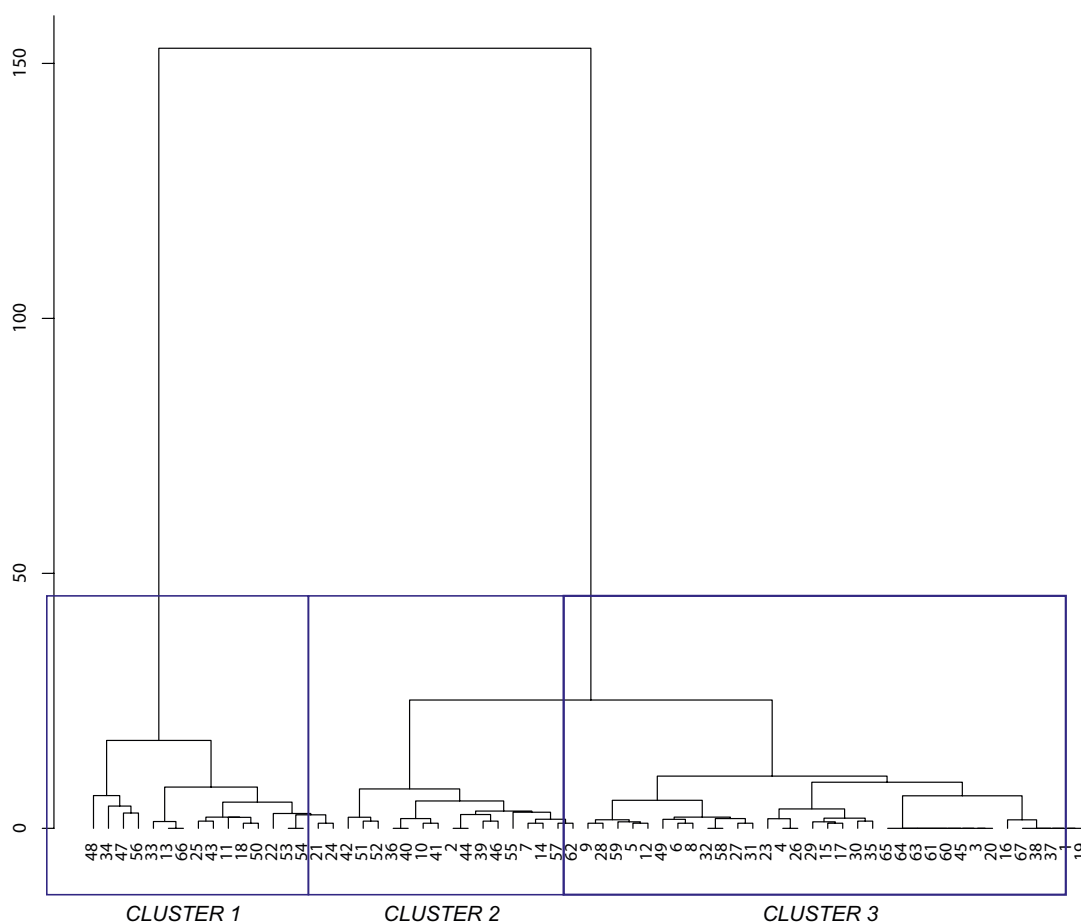
A Tabela 1 contém as estatísticas descritivas de cada variável para os cursos médicos de forma geral e para cada um dos três *clusters*. Dessa forma, é possível identificar as características em cada agrupamento.

O *cluster* 1 é formado por escolas em que o curso médico tem pelo menos 48 anos, entretanto mais da metade (53%) tem mais de 60 anos, todos com presença de HU. A maior parte das escolas (71%) utiliza metodologia tradicional. São públicas gratuitas ou privadas sem fins lucrativos. No *cluster*

2, é observado um predomínio de escolas sem HU: metade tem até 15% além do mínimo de carga horária exigida na diretriz curricular para Medicina, e 94% têm até 30%; a maior parte (88%) tem um regime semestral; mais da metade (69%) tem entre 13 e 24 anos; e 94% são escolas privadas. O *cluster* 3 representa as escolas mais jovens: todas têm até 12 anos; a maioria tem seis anos de existência (76%); não são gratuitas (97%); utilizam a metodologia ativa exclusivamente (74%) ou mista (15%); e 74% com até 15% além da carga horária mínima exigida para os cursos médicos (Tabela 1).

As variáveis se correlacionam de forma diferente dentro de cada *cluster* (Gráfico 3). Os gráficos de calor hierarquizados foram produzidos pela correlação de Pearson. As cores indicam correlação direta (tons de azul) ou inversa (tons de vermelho) entre as variáveis. A organização em *clusters* de correlação indica a proximidade entre as correlações das variáveis do estudo. Esse padrão de correlação diferente em cada *cluster* reforça a heterogeneidade dos subgrupos e mostra quais variáveis estão correlacionadas dentro de cada *cluster*.

Gráfico 2. Dendrograma da análise por agrupamento hierárquica dos cursos médicos do estado de São Paulo em dezembro de 2020.



Fonte: Elaborado pelos autores.

Tabela 1. Distribuição das frequências absoluta e relativa das variáveis: presença de HU, carga horária, regime letivo, início do curso, categoria administrativa da IES, metodologia de ensino utilizada e gratuidade dos cursos médicos, de forma geral e nos *clusters*, do estado de São Paulo em dezembro de 2020.

VARIÁVEL		GERAL		CLUSTER 1		CLUSTER 2		CLUSTER 3	
		N	%	N	%	N	%	N	%
<i>Presença de hospital universitário</i>	Não	47	70%	1	6%	13	81%	33	97%
	Sim	20	30%	16	94%	3	19%	1	3%
<i>Porcentagem da carga horária em relação ao mínimo exigido para os cursos médicos</i>	Até 15% além do mínimo	38	57%	5	29%	8	50%	25	74%
	De 16% a 30 % além do mínimo	19	28%	6	35%	7	44%	6	18%
	De 31% a 45% além do mínimo	8	12%	5	29%	1	6%	2	6%
	Mais de 45% além do mínimo	2	3%	1	6%	0	0%	1	3%
<i>Regime</i>	Anual	11	16%	8	47%	2	13%	1	3%
	Semestral	56	84%	9	53%	14	88%	33	97%
<i>Início do curso de Medicina</i>	Até 6 anos	26	39%	0	0%	0	0%	26	76%
	De 7 a 12 anos	10	15%	0	0%	2	13%	8	24%
	De 13 a 24 anos	11	16%	0	0%	11	69%	0	0%
	De 25 a 36 anos	6	9%	0	0%	6	38%	0	0%
	De 48 a 60 anos	8	12%	8	47%	0	0%	0	0%
	Mais de 60 anos	9	13%	9	53%	0	0%	0	0%
<i>Administração</i>	Pública gratuita	9	13%	7	41%	1	6%	1	3%
	Pública não gratuita	11	16%	2	12%	0	0%	9	26%
	Privada com fins lucrativos	16	24%	1	6%	5	31%	10	29%
	Privada sem fins lucrativos	31	46%	7	41%	10	63%	14	41%
<i>Metodologia</i>	Tradicional	20	30%	12	71%	4	25%	4	12%
	Ativa	35	52%	0	0%	10	63%	25	74%
	Mista	12	18%	5	29%	2	13%	5	15%
<i>Escolas gratuitas</i>	Não	58	87%	10	59%	15	94%	33	97%
	Sim	9	13%	7	41%	1	6%	1	3%
TOTAL		67	100%	17	100%	16	100%	34	100%

Fonte: Elaborada pelos autores.

Gráfico 3. Gráficos de calor hierarquizados: análise de correlação de Pearson entre as variáveis: presença de hospital universitário (HU), carga horária (HORAS), regime letivo (REG), início do curso (TEMPO), categoria administrativa e gratuidade da instituição de ensino superior (ADM_GRAT), metodologia de ensino utilizada (METOD) entre todos os cursos médicos do estado de São Paulo, em dezembro de 2020, (Gráfico A) e nos *clusters* 1 (Gráfico B), 2 (Gráfico C) e 3 (Gráfico D).

Gráfico A: Análise de correlação das variáveis de todas as escolas médicas do estado de São Paulo

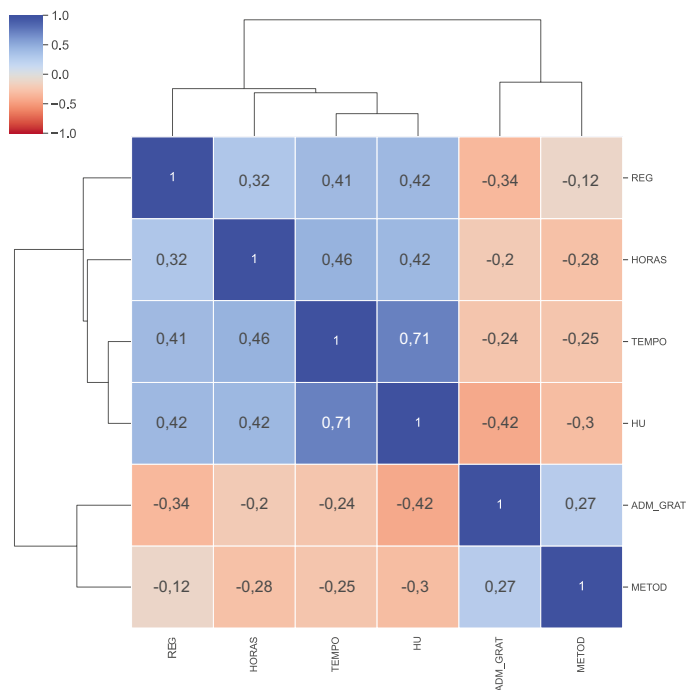


Gráfico B: Análise da correlação das variáveis no *Cluster 1*

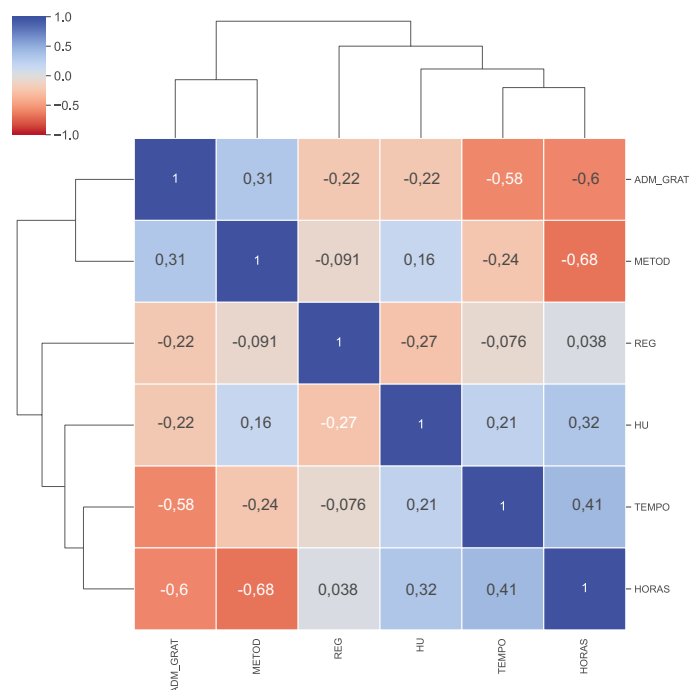


Gráfico C: Análise da correlação das variáveis no *Cluster 2*

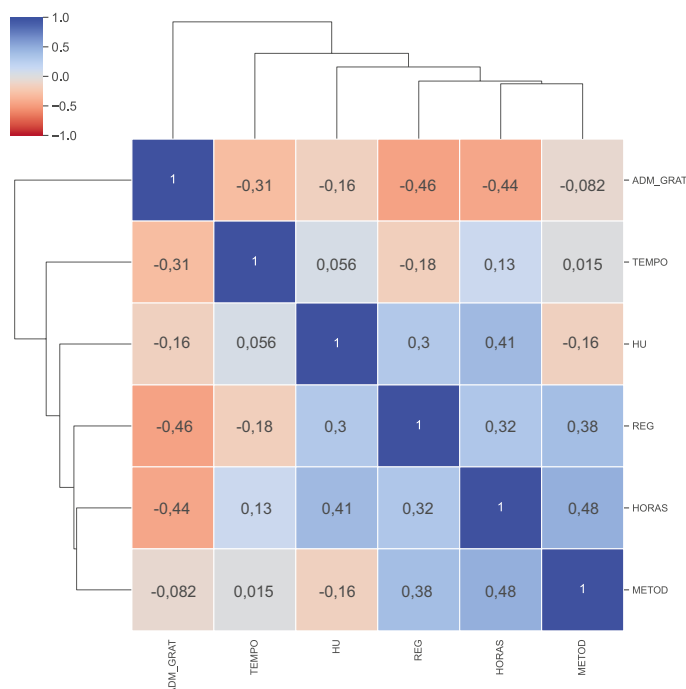
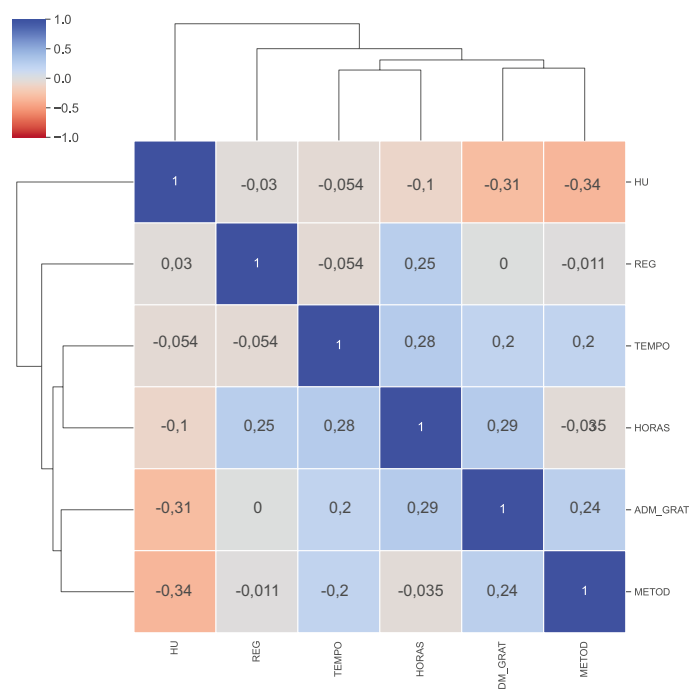


Gráfico D: Análise da correlação das variáveis no *Cluster 3*



ADM_GRAT: Forma de administração e gratuidade.
 TEMPO: Anos de existência do curso de Medicina.
 HU: Hospital universitário.
 REG: Regime anual ou semestral.
 HORAS: Carga horária do curso de Medicina.
 METOD: Metodologia utilizada (ativa, tradicional ou mista).

Fonte: Elaborado pelos autores.

DISCUSSÃO

No Brasil, o ensino superior tem uma grande heterogeneidade de instituições, tanto na rede pública como na privada. Gomes¹² aponta que a expressão universidade brasileira, com o objetivo de designar um conjunto de IES, deve ser aplicada com cuidado. Nesse sentido, é necessário apresentar classificações adequadas a um ensino superior tão diverso e heterogêneo quanto o brasileiro¹¹. A escolha deliberada de uma variável para determinar um padrão das escolas médicas não é suficiente. A partir das Diretrizes Curriculares Nacionais (DCN) de 2001 e 2014, o rigor de formas e o conteúdo das grades do currículo mínimo foram substituídos, possibilitando, dessa forma, a inovação e a diversificação na formação do profissional por meio da flexibilização da construção de projetos pedagógicos^{12,13}. Pereira et al.¹⁴, na cartografia das escolas médicas, descrevem a necessidade de detalhar a categoria “gratuita e paga” em razão da existência das IES de administração pública que cobram mensalidade, apontando, assim, para análises mais sofisticadas a respeito. Nessa perspectiva, é interessante observar que 82% dos cursos públicos não gratuitos estão contidos no *cluster*, três enquanto o *cluster* 1 contém 78% dos cursos gratuitos no estado de São Paulo.

As vagas em cursos médicos têm sido predominantemente ofertadas por IES privadas desde a década de 1970 e vêm aumentando progressivamente^{4,3}. A distribuição dos grupos evidencia três momentos de expansão de cursos. O *cluster* 1 representa as escolas mais antigas, a mais nova desse grupo tem 48 anos e 78% das escolas gratuitas do estado estão nesse grupo. O *cluster* 2, escolas intermediárias, tem apenas uma escola gratuita, e os cursos médicos têm idade variando de sete a 36 anos. O terceiro *cluster* contém os cursos não gratuitos (97%) com até 12 anos, entretanto a maioria (76%) tem no máximo uma turma formada e autorização que segue o edital do PMM¹⁵.

O PMM tinha como um dos objetivos a diminuição da concentração de médicos e escolas médicas em algumas regiões economicamente mais desenvolvidas. Entretanto, foi observado um processo de expansão predominantemente privado e em regiões que concentram boa parte da renda do país⁴. Praticamente metade (47%) dos cursos médicos do estado de São Paulo está nas regiões correspondentes a Grande São Paulo, Baixada Santista e Campinas, e a cidade de São Paulo responde por 28% das escolas médicas. Todas as IES utilizam o SUS como cenário de prática, e 30% têm HU próprio. No *cluster* 1, esse número sobe para 94%, no *cluster* 2 para 19% e no *cluster* 3 para 3%. Existe uma concentração de escolas com HU no *cluster* 1.

A Constituição brasileira de 1988, no artigo 207, além de definir a autonomia das IES, compromete-as com o princípio de indissociabilidade entre ensino, pesquisa e extensão^{14,16}.

A maior parte da pesquisa produzida, principalmente nas ciências básicas, é fruto do investimento público em universidades públicas¹⁷. Tomando os programas e cursos de pós-graduação em Saúde Coletiva e Medicina I, II e III, recomendados pela Capes, é possível confirmar que essa observação se repete na classificação aqui apresentada dos cursos médicos do estado de São Paulo. Todos os programas e cursos de pós-graduação em Saúde Coletiva estão no *cluster* 1. Dos oito programas, apenas dois não são em escolas gratuitas. Ao observar a área de Medicina I, II e III conjuntamente, a Capes reconhece 107 programas de pós-graduação, dos quais 95 (88%) estão no *cluster* 1. Apenas um curso em escola privada e quatro em escola pública não gratuita. O grupo 2 tem oito programas de pós-graduação. O grupo 3 não tem nenhum programa de pós-graduação. Vale ressaltar que existem quatro programas vinculados a hospitais e centro de estudos que pertencem ao mesmo grupo administrativo dos cursos médicos do *cluster* 3.

A pós-graduação não é objeto deste estudo, porém os cursos de *stricto sensu* compõem uma dimensão importante da formação em saúde e estão diretamente relacionados com a formação docente e estrutura das IES⁷. A distribuição dos programas de pós-graduação supracitada, entre os *clusters* formados, coopera no sentido de ratificar a modelagem proposta, sendo mais um elemento confirmador da capacidade de discriminação das variáveis utilizadas.

Uma limitação deste estudo é inerente ao banco de dados utilizado, isto é, foram retirados das fontes oficiais – *website* do MEC –, porém é possível que alguma IES tenha realizado alterações nos cursos sem ter alterado o cadastro. A inclusão de novas variáveis torna necessária a reaplicação da modelagem para que, eventualmente, sejam gerados novos agrupamentos⁹.

CONCLUSÃO

Este trabalho possibilitou classificar e agrupar as escolas médicas de acordo com suas semelhanças. Tal análise potencializa futuros estudos e contribui para realização deles, visando ao aprimoramento e ao conhecimento dos padrões e das relações de cada *cluster* com as distintas variáveis, identificando diferenças nas respectivas semelhanças entre os cursos de Medicina no estado de São Paulo.

CONTRIBUIÇÃO DOS AUTORES

Gabriela Furst Vaccarezza participou da concepção do estudo, da coleta e análise dos dados, e da redação do artigo. Erik Montagna participou da análise estatística e da interpretação de resultados. Rita Barradas Barata participou da concepção do estudo e da discussão e interpretação dos resultados.

Nivaldo Carneiro Junior participou da concepção do estudo, da discussão dos resultados e da revisão final do artigo.

CONFLITO DE INTERESSES

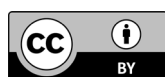
Declaramos não haver conflito de interesses.

FINANCIAMENTO

Taxa Escolar Capes: Processo nº 88887.598161/2021-00.

REFERÊNCIAS

1. Fávero LPL, Belfiore PP. Manual de análise de dados: estatística e modelagem multivariada com Excel, SPSS e Stata. Rio de Janeiro: Elsevier Brasil; 2017.
2. Aldenderfer MS, Blashfield RK. Cluster analysis. California: Thousand Oaks; 1984. doi: <https://dx.doi.org/10.4135/9781412983648>.
3. Scheffer MC, Cassenote A, Guerra A, Guilloux AGA, Brandão APD, Miotto BA, et al. Demografia médica no Brasil 2020. São Paulo: FMUSP, Conselho Federal de Medicina; 2020.
4. Scheffer MC, Dal Poz MR. The privatization of medical education in Brazil: trends and challenges. Hum Resour Health. 2015;13(1):1-10. doi: <https://doi.org/10.1186/s12960-015-0095-2>.
5. Lopes AC. A explosão numérica das escolas médicas brasileiras. Educ Med. 2018;19(supl 1):19-24. doi: <https://doi.org/10.1016/j.edumed.2018.03.004>.
6. Oliveira BLCAD, Lima SF, Pereira MUL, Pereira Jr GA. Evolução, distribuição e expansão dos cursos de medicina no Brasil (1808-2018). Trab Educ Saúde. 2019;17(1):e0018317. doi: <https://doi.org/10.1590/1981-7746-sol00183>.
7. Oliveira FP, Pinto HA, Figueiredo AM, Cyrino EG, Oliveira Neto AV, Rocha VXM. Programa Mais Médicos: avaliando a implantação do Eixo Formação de 2013 a 2015. Interface. 2019; 23(supl 1):1-17. Doi: <http://dx.doi.org/10.1590/interface.170949>.
8. Schmidt M, Ruffoni J. Interações estabelecidas pelas universidades brasileiras públicas e privadas. Rev Econômica. 2018;20(2):33-58 [acesso em 18 fev 2023]. Disponível em: <https://periodicos.uff.br/revistaeconomica/article/view/35038>.
9. Figueiredo Filho DB, Silva Jr JA, Santos Filho RP, Rocha EC, Nascimento WS, Silva MB. Happy together: como utilizar análise fatorial e análise de cluster para mensurar a qualidade das políticas públicas. Teor Soc. 2014; 22(2):123-52 [acesso em 18 fev 2023]. Disponível em: <https://bib44.fafich.ufmg.br/teoriaesociedade/index.php/rts/article/view/197/143>.
10. Hair JF, Babin BJ. Multivariate data analysis. 8th ed. Hampshire, UK: Cengage Learning Emea; 2018. 832 p.
11. Brasil. Denominações e siglas de IES [acesso em 13 jan 2022]. Disponível em: <http://portal.mec.gov.br/programa-saude-da-escola/323-secretarias-112877938/orgaos-vinculados-82187207/12914-denominacoes-e-siglas-de-ies-sp-564060814>.
12. Gomes MF. Avaliação e natureza administrativa das instituições de ensino superior. Ensaio: Aval Pol Públ Educ. 2010;18(68):589-610. doi: <https://doi.org/10.1590/S0104-40362010000300010>.
13. Brasil. Resolução CNE/CES nº 3, de 20 de junho de 2014. Institui as Diretrizes Curriculares Nacionais do Curso de Graduação em Medicina e dá outras providências. Diário Oficial da União; 23 jun 2014.
14. Pereira DVR, Fernandes DDLR, Mari JF, Lage ALDF, Fernandes APPC. Cartografia das escolas médicas: a distribuição de cursos e vagas nos municípios brasileiros em 2020. Rev Bras Educ Med. 2021;45(1):e005. doi: <https://doi.org/10.1590/1981-5271v45.1-20200282.ing>.
15. Kusakawa DHB, Antonio CA. Os eixos estruturantes das Diretrizes Curriculares Nacionais dos cursos de Medicina no Brasil. Rev Docência Ens Sup. 2017;7(1):165-84. doi: <https://doi.org/10.35699/2237-5864.2017.2245>.
16. Brasil. Constituição da República Federativa do Brasil. Brasília: Senado Federal; 1988 [acesso em 18 jan 2022]. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm.
17. Brasil. Lei nº 12.871. Institui o Programa Mais Médicos, altera as Leis nº 8.745, de 9 de dezembro de 1993, e nº 6.932, de 7 de julho de 1981, e dá outras providências. Brasília; 2013.



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.