

PROPOSTA DE UM MODELO DE SEGMENTAÇÃO BASEADO EM NÍVEIS DE DESEMPENHO DAS UNIVERSIDADES BRASILEIRAS

RAFAEL DE FREITAS SOUZA

UNIVERSIDADE DE SÃO PAULO (USP)

JULIO ARAUJO CARNEIRO DA CUNHA

UNIVERSIDADE NOVE DE JULHO (UNINOVE)

HAMILTON LUIZ CORREA

FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DA UNIVERSIDADE DE SÃO PAULO - FEA

PROPOSTA DE UM MODELO DE SEGMENTAÇÃO BASEADO EM NÍVEIS DE DESEMPENHO DAS UNIVERSIDADES BRASILEIRAS

RESUMO

Os modelos de avaliação de desempenho para as instituições de ensino superior no Brasil comumente são elaborados a partir de elementos arbitrários e subjetivos. Segmentações dos níveis de desempenho têm seus estratos definidos sem o respaldo adequado para se ter uma segregação de grupos de desempenho mais eficientes. Diante dessas dificuldades, o objetivo do trabalho foi propor um modelo de segmentação por níveis de desempenho para universidades no Brasil. A partir da coleta de dados secundários em diferentes bases de dados (MEC, CAPES, *Scopus* e *Webometrics*), 183 universidades brasileiras foram analisadas (de um universo de 197) a partir de regressão logística multinomial em cinco diferentes modelos propostos na pesquisa, incluindo duas análises para as universidades públicas (96 observações). Os resultados mostram que o modelo mais eficiente de segmentação para níveis de desempenho é o que divide as universidades em quartis, que se fortalece ainda mais quando se trata de universidades públicas. Como contribuição gerencial, propõe-se um modelo de segmentação por nível de desempenho para a avaliação das universidades brasileiras. Como contribuição teórica, coloca-se que níveis de desempenho devem considerar natureza e orientação estratégica da organização.

Palavras-chave: Educação; Níveis de Desempenho; Universidade; Brasil.

1 INTRODUÇÃO

O ensino de nível superior ganhou características de massificação ao redor de todo o mundo, o que trouxe uma visão utilitarista e quantitativa para se compreender e mensurar o desempenho das universidades (Alexander, 2000). Isso influenciou a forma como se realizam as atividades de avaliação de desempenho de universidades até os dias de hoje. No Brasil particularmente, ao longo das últimas décadas, o MEC (Ministério da Educação) tem trabalhado no desenvolvimento de indicadores quantitativos de desempenho das instituições de ensino superior, os quais intitulam de Indicadores de Qualidade. Entretanto, muitos deles apresentam problemas como, por exemplo, relativos à ponderação arbitrária de suas variáveis. Isso causa vieses e dificuldades em suas interpretações.

Por exemplo, um dos indicadores mais utilizados é o indicador da Titulação do Corpo Docente – TCD, dado por $TCD = \frac{G \times 1 + A \times 2 + E \times 3 + M \times 4 + D \times 5}{G + A + E + M + D}$ (Brasil, 2017). Nele, infere-se que um doutor vale cinco vezes mais do que um graduado, uma vez e vinte e cinco décimos a mais do que um mestre e assim sucessivamente, gerando dificuldades interpretativas e possibilitando o desenvolvimento de maneiras matemáticas que permitam um TCD razoável, mesmo não havendo uma quantidade razoável de doutores nos quadros funcionais universitários. Estatisticamente não se vislumbra as razões pelas quais tais ponderações foram utilizadas. Ou seja, foram ali colocados de maneira arbitrária. Podem se citar outros exemplos, tais como a ponderação realizada para se subdividir as cinco perspectivas de avaliação dos programas de pós-graduação no Brasil, ou mesmo a definição de níveis do IGC definidos por cálculos de médias (Brasil, 2014). Há, portanto, diversos instrumentos e modelos de avaliação de desempenho universitários cujos níveis de desempenho são definidos de forma subjetiva ou arbitrária em agências governamentais.

Paralelamente, outras iniciativas para se avaliar o desempenho universitário foram apresentadas. Regra geral, tenta-se encaixar a realidade dessas organizações educacionais, de contexto específico, dentro de modelos genéricos. Como exemplo, há estudos que aplicaram o *Balanced Scorecard* para se gerar um painel de indicadores em dimensões que não necessariamente atendem a realidade de uma instituição de ensino superior (e.g. Alani *et al.*, 2017; Chen *et al.*, 2006).

Em todos os casos citados podem-se perceber problemas relacionados às premissas e às ponderações arbitrárias de pesos nos indicadores na elaboração dos modelos – e.g.: a adoção de pesos numéricos para se diferenciar critérios de uma escala *likert* que, em essência, são variáveis qualitativas (Fávero & Belfiore, 2017); ou as ponderações arbitrárias tais como as encontradas no TCD (MEC, 2017). Muita arbitrariedade acaba gerando muitas abordagens para se entender o desempenho (Matthews, 2011), o que dificulta na identificação de um modelo amplamente aceito e se propaga, institucionalmente, diversos outros modelos arbitrários para suprir deficiências dos anteriores.

Adicionalmente, essa forma de se considerar o desempenho de forma arbitrária também traz outro problema em particular, que é a comparação de diferentes categorias de organizações, especialmente, para as organizações de ensino e universidades. Diferentes tipos de universidades acabam sendo comparados dentro das mesmas premissas e preceitos. Isso se agrava quando há natureza de atividade e objetivo estratégico diferente dentre as universidades, como é o caso de orientações distintas entre universidades públicas e privadas no Brasil.

Diante da arbitrariedade na elaboração e modelos para avaliação de desempenho em universidades, esta pesquisa tem por objetivo propor um modelo de segmentação por níveis de desempenho para universidades no Brasil. Espera-se que se possa propor um modelo que possa ser útil para a prática gerencial ao mesmo tempo em que se pretende mostrar que modelos que segmentam universidades por níveis de desempenho devem ter respaldo metodológico. Por fim, espera-se que se contribua para o corpo teórico dessa temática de forma a mostrar que estratificações de desempenho dependem, essencialmente, da natureza e orientação estratégica das organizações (no caso, a orientação pública ou privada da universidade no Brasil). Espera-se que nos estudos sobre Gestão Universitária, elementos como a orientação da universidade devem ser considerados em estratificações.

2 REFERENCIAL TEÓRICO

Para esta pesquisa, define-se por avaliação de desempenho o conjunto de mecanismos vinculados à estratégia organizacional (Bititci *et al.*, 2012) direcionados à quantificação e à medição do desempenho organizacional como um todo (Taticchi *et al.*, 2015), considerando dados financeiros e não-financeiros, e em que a ação nela respaldada leva à aferição da performance (Neely *et al.*, 2005). A partir dessa perspectiva, a avaliação de desempenho em universidades é analisada.

2.1 Desempenho organizacional das universidades brasileiras

O desempenho organizacional das universidades brasileiras, tal qual em qualquer organização, contraria o previsto pelo modelo econômico neoclássico, isto é, é algo heterogêneo. Contrário fosse, não haveria a necessidade de se pensar nas estratégias gerenciais; o foco no processo ou na operação da entrega do serviço educacional seria o suficiente (Williamson, 1991).

Ainda não se percebe de maneira parcimoniosa na literatura a razão da heterogeneidade do desempenho das organizações, porém é possível dizer que a percepção do significado do que vem a ser uma vantagem competitiva domina a explicação desse fenômeno (Kazlauskaitė & Bučiūnienė, 2008).

Frente ao exposto, pode-se verificar por meio de ranqueamentos internacionais, tais quais o *Quacquarelli Symonds World University Ranking – QS Ranking*, o *Times Higher Education World University Rankings – THE* ou o *Ranking Web of Universities – Webometrics* (utilizado nesta pesquisa para as modelagens multivariadas), quais as universidades brasileiras possuem, ao menos, indícios de uma vantagem competitiva sustentável. Repetidamente, universidades como a Universidade de São Paulo – USP, a Universidade de Campinas – UNICAMP, a Universidade Federal do Rio de Janeiro – UFRJ, a Universidade Federal de Minas Gerais – UFMG, a Universidade Estadual Paulista “Julio de Mesquita Filho” – UNESP,

a Universidade Federal do Rio Grande do Sul – UFRGS, aparecem no topo dos *rankings* internacionais que apontam as melhores universidades da América Latina. Elas conseguem essas posições independentemente de suas complexas conjunturas financeiras atuais, que são diretamente impactadas pelo atual cenário político (Kapetanidou & Lee, 2016). Essas avaliações mostram que poucas do cenário nacional se destacam. Por isso, é importante que haja um sistema de avaliação de desempenho capaz de separar as diferentes universidades nacionais em níveis para que políticas e ações sejam tomadas orientadas a grupos. Seja por necessidade de incentivo à melhoria do desempenho, seja por tornar as de alto desempenho ainda mais competitivas no cenário nacional.

Ao que parece, essa nova perspectiva já está sendo explorada pelo Estado de São Paulo. A Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP, em notícia veiculada em dezembro de 2017, direciona-se a financiar ações e a incentivar as três universidades estaduais paulistas – a USP, a UNICAMP e a UNESP – a revisar seus sistemas de avaliação de desempenho com o objetivo de incrementar métricas, estabelecer metas e melhorar competitividade internacional (Izique, 2017). Todavia, para que isso ocorra, é necessário haver um sistema que possa diferenciar as universidades conforme seus níveis de desempenho de forma a orientar as tomadas de decisões para incentivá-las a uma maior produtividade e desempenho. Daí a importância em se discutir a proposição de modelagens multivariadas preditivas aplicáveis ao caso brasileiro, direcionadas ao aumento do desempenho traduzido nos ranqueamentos internacionais. A tomada de decisões e o controle a partir de sistemas de avaliação de desempenho (Hourneaux Jr. *et al.*, 2017) parecem ficar mais facilitados se houvesse essa segmentação por níveis de desempenho.

2.2 A avaliação de desempenho de universidades por segmentação

Segmentando-se os níveis de avaliação de desempenho de universidades, alguns problemas podem acontecer em decorrência da arbitrariedade das definições desses estratos de níveis de desempenho. Porém, estudos mostraram que quando se divide os níveis de desempenho de universidades, pode haver melhor compreensão do fenômeno (e.g. Liefner, 2003). Dessa forma, cada tipo de universidade seria melhor avaliada em termos de desempenho se pudesse ter suas especificidades atendidas, respeitando assim a ideia de que modelos de avaliação de desempenho deveriam ser específicos ao setor e ao tipo de organização (Carneiro-da-Cunha & Corrêa, 2013).

Ao se criar sistemas de avaliação de desempenho, comumente, se estabelecem KPIs (*Key Performance Indicators*) para o setor da educação que são amplamente utilizados para a gestão e o controle de instituições de ensino e educação, sem que haja uma real evidenciação de sua efetividade para essas atividades (Kairuz *et al.*, 2016). Inclusive, no Brasil, há diferenças de percepções sobre os melhores indicadores para se avaliar uma instituição de ensino superior (Soares *et al.*, 2017). Quando não há modelos com segmentação por níveis de desempenho de forma não arbitrária, decisões internas e de agentes públicos podem ser tomadas para determinado grupo quando existem diferentes necessidades e perfis de universidades dentro do mesmo grupo.

Por outro lado, pode-se defender que quanto mais estratificada a segmentação para determinação de níveis de desempenho, melhor seria do ponto de vista discricionário. Isto é, maior quantidade de níveis seria interessante para diferenciar melhor cada universidade. Entretanto, não se pode considerar somente o poder de discricionariedade de um modelo, mas o quanto ele realmente pode agrupar diferentes observações dentro de um mesmo grupo homogêneo, independente de quantos grupos existem.

2.3 Universidades públicas e privadas no Brasil

Ainda que o modelo de educação de ensino superior no Brasil conte com incentivos e projetos de acesso à educação superior (Brasil, 2014), ainda assim, há universidades que se

caracterizam por serem pagas e não pagas (universidades públicas, mas que em alguns casos específicos podem ainda ter cobranças parciais, como é o caso de algumas instituições municipais). O acesso ao ensino superior privado no Brasil se iniciou a partir da Reforma Universitária de 1968 quando as vagas do ensino público não eram suficientes para atender a demanda nacional (Martins, 2009). Desde então, a proporção de ensino superior privado aumentou de forma crescente no Brasil (Mancebo *et al.*, 2015). Esse advento de instituições privadas ao mercado da educação traz à tona a produção e conhecimento mercantilizada (Mancebo *et al.*, 2016), o que determina uma diferenciação entre as orientações de produção de conhecimento entre as universidades públicas e privadas.

Entretanto, o foco e a relevância em pesquisa e desenvolvimento de novos conhecimentos científicos é algo que ainda se concentra nas universidades públicas. Dessa forma, as universidades públicas são identificadas, comumente, com desempenho superior às privadas (Hoffmann *et al.*, 2014). Uma das razões que isso ocorre é porque o professor de instituição pública tende a investir mais em sua carreira profissional (Rowe & Bastos, 2010), o que o faz ser mais orientado a produzir mais pesquisas e gerar conhecimento. Nesse sentido, reforça-se que há mais orientação para a geração de conhecimento científico em universidades públicas, o que faz com que as atividades não se concentrem exclusivamente no ensino ou no conhecimento massificado. Por isso, ao se comparar a produtividade científica, esse é um dos principais fatores que justifica o destaque em desempenho de universidades públicas.

2.4 Hipóteses de pesquisa

A forma pela qual se estabelecem quantidades de níveis de desempenho para as universidades, nem sempre têm respaldo de uma análise específica que separe bem cada um dos grupos de desempenho. Assim, pode ser que um modelo com maiores níveis de estratificação possa ser mais interessante por poder discriminar melhor os níveis de desempenho de universidades. Para sustentar essa ideia, supõe-se analisar a segmentação dos níveis de desempenho de universidades por decis (dez partes), tal como a Hipótese 1.

H₁: A segmentação por nível de desempenho é eficiente quando dividida em decis.

A partir da *H₁*, elabora-se a Hipótese 2, tentando então diminuir a quantidade de níveis de desempenho das universidades para cinco níveis. Adicionalmente, estima-se que quando há uma orientação não só ao ensino, mas também à pesquisa e geração de conhecimento científico, situação típica de universidades públicas, o modelo por quintis proposto pode se tornar ainda mais eficiente. Pelo fato de existir mais ênfase nesse foco gerador de conhecimento, espera-se que as universidades públicas tenham maior aderência a um modelo do que as universidades cujo foco está estritamente no ensino. Isto posto, sugerem-se as hipóteses:

H_{2a}: A segmentação por nível de desempenho é eficiente quando dividida em quintis.

H_{2b}: A segmentação por nível de desempenho em universidades públicas é eficiente quando dividida em quintis.

Buscando uma quantidade ainda mais restrita de níveis de desempenho, o terceiro teste a ser realizado é com quatro níveis de desempenho das universidades. Da mesma maneira como discutido nas hipóteses do bloco 2, espera-se que quando há universidades mais orientadas para a pesquisa, o modelo tende a ter um melhor ajuste e eficiência. Assim sendo, o conjunto de Hipóteses 3 traz a divisão de nível de desempenho de universidades em quartis.

H_{3a}: A segmentação por nível de desempenho é eficiente quando dividida em quartis.

H_{3b}: A segmentação por nível de desempenho em universidades públicas é eficiente quando dividida em quartis.

3 MÉTODOS

A pesquisa apresentada fundamenta-se em: a) levantamentos bibliográficos; b) na evidência empírica da bibliografia selecionada; c) na experiência dos pesquisadores em razão de suas atuações e contribuições no assunto discutido; d) na construção positivista quantitativa fruto da modelagem multivariada de dados; e) na utilização de dados financeiros e não financeiros; e f) e na utilização mútua de variáveis quantitativas e qualitativas.

O universo que compõe a pesquisa conta com 197 universidades brasileiras ativas em 2017, independentemente se públicas ou privadas, pagas ou gratuitas. A amostra deriva da presença dos indivíduos que compõem a população do estudo no ranking *Webometrics* do ano de 2018 para a delimitação das variáveis dependentes e, conseqüentemente, para a proposição de cinco modelagens multivariadas das variáveis preditoras do desempenho.

Assim, as modelagens foram feitas com 183 universidades brasileiras, representando 92,89% da população. Há a presença de modelagens alternativas contando apenas com as universidades gratuitas, compondo uma amostra de 96 indivíduos, representando 48,73% do universo. O estudo debruçou-se sobre a coleta de dados secundários, em função de sua alta disponibilidade, tal qual apontado no Quadro 1.

Quadro 1 Repositório de dados sobre as universidades brasileiras

Fonte dos Dados	Tipo de Variáveis Existentes
Portais Governamentais (Portais de Transparência, Portal do Ministério da Educação – MEC e Portal do Instituto Nacional de Propriedade Industrial)	Discretas, Contínuas, Dicotômicas, Ordinais e Categóricas
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Capes	Discretas e Ordinais, regra geral.
Portais de medição da contribuição científica (Portal <i>Scopus</i>)	Discretas, regra geral.
<i>Rankings</i> Internacionais de Universidade	Catagóricas ordinais, em regra. Para a pesquisa, foi adotado o ranqueamento <i>Webometrics</i> .

Fonte: elaborado pelos autores.

Para a construção das modelagens, foi utilizado o software *Stata* em sua versão 14.0 e, em razão das variáveis independentes serem do tipo qualitativas categóricas policotômicas, adotou-se como técnica de modelagem a Regressão Logística Multinomial (Agresti, 2013; Dakin *et al.*, 2006).

3.1 Regressão Logística Multinomial

A Regressão Logística Multinomial – RLM constitui-se em uma técnica preditiva que estuda a dependência de determinadas variáveis qualitativas categóricas (Y_{ki}) em relação a variáveis explicativas propostas (X_{1i} a X_{ki}) (Hair *et al.*, 2009), indicando a probabilidade de ocorrência das primeiras, através da disposição dos dados em uma *multiway contingency table*, utilizando como critério de estimação a máxima verossimilhança (Hosmer *et al.*, 2013), que pode ser descrito como $LL = \sum_{i=1}^n \sum_{m=0}^{M-1} \left[(Y_{im}) \times \ln \left(\frac{e^{Z_{im}}}{\sum_{m=0}^{M-1} e^{Z_{im}}} \right) \right] = \text{máx.}$

Na RLM, se as Y_{ki} apresentarem M categorias de resposta, encontrar-se-á o número de $M - 1$ logitos Z_{im} e, através deles, poderá ser feita a estimação dos cálculos das probabilidades da ocorrência das categorias M .

A expressão geral dos logitos Z_{im} , com $m = (0; 1; 2; \dots; M - 1)$, pode ser dada por $Z_{im} = \alpha_m + \beta_{1m}X_{1m} + \beta_{2m}X_{2m} + (\dots) + \beta_{km}X_{ki}$, em que $Z_{i_0} = 0$ e, conseqüentemente, $e^{Z_{i_0}} = 1$ (Fávero & Belfiore, 2017).

Para o estudo, foram utilizadas como variáveis qualitativas categóricas dependentes as *labels* de ‘A’ a ‘D’, para a modelagem com a divisão da amostra em quartis; as *labels* de ‘A’ a

‘E’, para a modelagem com a divisão da amostra em quintis; e as *labels* de ‘A’ a ‘J’, para a modelagem com a divisão da amostra em decis.

As *labels* foram definidas através da evidenciação dos quartis, quintis e decis da amostra, sendo que ‘A’ representa o grupo das universidades mais bem colocadas no ranking *Webometrics* no ano de 2018, ‘B’ representa o grupo subsequente, e assim sucessivamente.

Isto posto, adotando-se, por exemplo, a categoria ‘A’ como a categoria de referência dos cálculos, poder-se-ia então descrever a probabilidade de sua ocorrência por $p_{i_A} = \frac{1}{1+e^{z_{i_B}}+e^{z_{i_C}}+\dots+e^{z_{i_M}}}$.

Assim sendo, para a categoria ‘B’, ter-se-ia $p_{i_B} = \frac{e^{z_{i_B}}}{1+e^{z_{i_B}}+e^{z_{i_C}}+\dots+e^{z_{i_M}}}$; para a categoria ‘C’, $p_{i_C} = \frac{e^{z_{i_C}}}{1+e^{z_{i_B}}+e^{z_{i_C}}+\dots+e^{z_{i_M}}}$, e assim sucessivamente até a última categoria M do estudo, de forma que, de maneira geral, poder-se-ia resumir a fórmula das probabilidades como $p_{i_m} = \frac{e^{z_{i_m}}}{\sum_{m=0}^{M-1} e^{z_{i_m}}}$.

Os intervalos de confiança considerados no estudo para o cálculo das probabilidades foram estabelecidos, regra geral, em um nível de confiança de 95% e podem ser descritos como $\alpha = \pm 1,96[s.e.(\alpha)]$ e $\beta_j = \pm 1,96[s.e.(\beta_j)]$, que para cada parâmetro β_{jm} (em que $j = (1; 2; 3; \dots; k)$) do evento de interesse.

3.2 Tratamento dos dados

Antes do tratamento dos dados de 2017, foi calculado o coeficiente de correlação Pearson (ρ) de todas as 50 variáveis quantitativas.

Para a adequação dos modelos através da RLM, as variáveis foram organizadas em ordem decrescente do valor de ρ . Após isso, tentou-se combinar algumas das 26 variáveis explicativas categóricas ao modelo.

No caso das variáveis explicativas categóricas com mais de duas categorias, estas foram transformadas em $n - 1$ *dummies* para que se evitasse o problema da ponderação arbitrária, visto não haver uma equidistância aceita entre as diferentes categorias de uma mesma variável.

Após isso, utilizou-se o método *stepwise* para a verificação da aderência das variáveis aos modelos propostos, considerando uma significância de 5%, e adotando-se para todos os casos, a *label* identificadora ‘A’ como a categoria de referência dos modelos.

Aponta-se que não houve nenhum tipo de tratamento para os *outliers* multivariados do estudo. Os pesquisadores reconheceram a sua presença através do algoritmo *Blocked Adaptive Computationally Efficient Outlier – BACON*, que utiliza a distância de Mahalanobis para todo o grupo (ver Billor *et al.*, 2000; Weber, 2012).

Assim, atentando-se ao critério proposto, afirma-se a existência de *outliers* multivariados a depender de cada modelagem, mas os pesquisadores decidiram mantê-los, pois uma das motivações do trabalho é a de contribuir para a construção de modelos preditivos da performance/qualidade de todas as universidades do caso brasileiro constantes no ranking *Webometrics* do ano de 2018.

4 APRESENTAÇÃO DE RESULTADOS

Preliminarmente, apresentam-se os resultados encontrados para a subsequente discussão e interpretação dos dados e parâmetros encontrados nas modelagens propostas.

Divida a amostra de 183 universidades brasileiras em decis, agrupadas em *labels* identificadoras de ‘A’ a ‘J’, sendo o grupo ‘A’, o representante das universidades mais bem colocadas no ranqueamento *Webometrics* do ano de 2018; e o grupo ‘J’, o representante das piores colocadas. Em relação aos quintis, apresentam-se dois resultados: a) o modelo R1, que

divide a amostra de 183 universidades brasileiras em quintis, agrupadas em *labels* identificadoras de ‘A’ a ‘E’, sendo o grupo ‘A’, o representante das universidades mais bem colocadas no ranqueamento *Webometrics* do ano de 2018; e o grupo ‘E’, o representante das piores colocadas; e b) o modelo R2, que aproveita o agrupamento proposto pela modelagem R1, excluídas as universidades não gratuitas, obtendo-se uma amostra de 96 universidades brasileiras gratuitas em quintis, agrupadas em *labels* identificadoras de ‘A’ a ‘E’, sendo o grupo ‘A’, o representante das universidades mais bem colocadas no ranqueamento *Webometrics* do ano de 2018; e o grupo ‘E’, o representante das piores colocadas.

Em relação aos quartis, apresentam-se dois resultados: a) o modelo S1, que divide a amostra de 183 universidades brasileiras em quartis, agrupadas em *labels* identificadoras de ‘A’ a ‘D’, sendo o grupo ‘A’, o representante das universidades mais bem colocadas no ranqueamento *Webometrics* do ano de 2018; e o grupo ‘D’, o representante das piores colocadas; e b) o modelo S2, que aproveita o agrupamento proposto pela modelagem S1, excluídas as universidades não gratuitas, obtendo-se uma amostra de 96 universidades brasileiras gratuitas em quintis, agrupadas em *labels* identificadoras de ‘A’ a ‘D’, sendo o grupo ‘A’, o representante das universidades mais bem colocadas no ranqueamento *Webometrics* do ano de 2018; e o grupo ‘D’, o representante das piores colocadas.

O Quadro 2 sintetiza os principais parâmetros das cinco modelagens trabalhadas.

Quadro 2 Síntese dos resultados dos cinco modelos.

Modelo	n	Níveis	LR χ^2	Prob > χ^2	Pseudo R ²	Log likelihood
Q	183	Decis	363.10 (18df)	0.000	0.4310	-239.653
R1	183	Quintis	306.00 (12df)	0.000	0.5195	-141.511
R2	96	Quintis	178.62 (20df)	0.000	0.6189	-54.995
S1	183	Quartis	292.52 (9df)	0.000	0.5766	-107.422
S2	96	Quartis	152.02 (9df)	0.000	0.6173	-47.120

Fonte: elaborado pelos autores.

5 ANÁLISE DOS DADOS

Para a presente pesquisa, após a realização do procedimento *stepwise*, um total de 13 variáveis aderiram aos modelos propostos, cujas descrições das respectivas composições constam no Quadro 3 a seguir.

Quadro 3 Descrição das variáveis explicativas que compõem os modelos propostos

Variáveis Escolhidas	Descrição da Composição da Variável
<i>qt_a_scopus</i>	A variável considera os valores totais retirados do sítio eletrônico <i>Scopus</i> para o ano de 2017 através de operadores booleanos.
<i>qt_doutorado_capes_total</i>	A variável aponta o somatório, por universidade, dos seus cursos de pós-graduação em nível de doutorado, recomendados pela CAPES.
<i>qt_doutorado5</i>	A variável deriva de uma curiosidade do pesquisador. Ela aponta o somatório, por universidade, dos seus cursos de pós-graduação em nível de doutorado, com conceitos iguais ou superiores a 5.
<i>qt_mestrado_capes_total</i>	A variável deriva de uma curiosidade do pesquisador. Ela aponta o somatório, por universidade, dos seus cursos de pós-graduação em nível de mestrado, recomendados pela CAPES.
<i>qt_prof_doutorado</i>	A variável aponta o quantitativo total dos Professores Universitários doutores.

<i>qt_prof_especializacao</i>	A variável aponta o quantitativo total dos Professores Universitários com a formação máxima Pós-Graduação <i>lato sensu</i> .
<i>qt_prof_mestrado</i>	A variável aponta o quantitativo total dos Professores Universitários com a formação máxima em nível de mestrado.
<i>qt_stricto_capes_total</i>	A variável aponta o somatório, por universidade, dos seus cursos de pós-graduação <i>stricto sensu</i> recomendados pela CAPES, sem distinção entre mestrados e doutorados.
<i>qt_stricto5_total</i>	A variável aponta o somatório, por universidade, dos seus cursos de pós-graduação <i>stricto sensu</i> recomendados pela CAPES, com conceitos iguais ou superiores a 5, sem distinção entre mestrados e doutorados.
<i>qt_tec_doutorado</i>	A variável aponta o quantitativo total dos Técnicos Universitários doutores.
<i>qt_tec_especializacao</i>	A variável aponta o quantitativo total dos Técnicos Universitários com a formação máxima Pós-Graduação <i>lato sensu</i> .
<i>qt_tec_medio</i>	A variável aponta o quantitativo total dos Técnicos Universitários com a formação máxima Ensino Médio Completo.
<i>qt_tec_superior</i>	A variável aponta o quantitativo total dos Técnicos Universitários com a formação máxima Ensino Superior Completo.

Para todos os modelos propostos, a estatística χ^2 , descrita por $\chi^2 = -2.(LL_0 - LL_{m\acute{a}x})$, fez com que os pesquisadores rejeitassem a H_0 do teste ao demonstrar que havia nas modelagens, ao menos um β_{jm} estatisticamente diferente de 0, ao nível de significância de 5%, visto que todos os $\chi^2_{cal} > \chi^2_c$ e, conseqüentemente, todos os *pvalue* $\chi^2 < 0,05$, conforme demonstra o Quadro 4.

Quadro 4 Análise resumo da estatística χ^2 presente nas modelagens propostas

Modelagem	Estatística χ^2	Graus de Liberdade	$\chi^2_{calc} > \chi^2_c?$	Significância
Q	363,10	18	Sim	0,0000
R1	306,00	12	Sim	0,0000
R2	178,62	20	Sim	0,0000
S1	292,52	9	Sim	0,0000
S2	152,02	9	Sim	0,0000

Fonte: elaborado pelos autores.

Nas Regressões Logísticas, por possuírem como variável dependente uma variável categórica, não faz sentido se falar nas estatísticas R^2 e R^2 Ajustado, visto não fazer sentido calcular a covariância de uma variável qualitativa (Garson, 2012).

Assim, de maneira similar, porém não idêntica, pode-se fazer o uso da estatística pseudo R^2 de McFadden, dado por $pseudo R^2 = \frac{-2LL_0 - (-2L_{m\acute{a}x})}{-2L_0}$.

O problema reside ao passo em que a estatística *pseudo* R^2 é bastante limitada e sua utilização restringe-se a uma comparação, ainda que precária, de dois ou mais modelos propostos (Ayçaguer & Utra, 2004; Azen & Walker, 2011), através da escolha do maior *pseudo* R^2 , porém essa estatística nada diz a respeito da eficiência global do modelo (Fávero, 2015) – em verdade funciona como uma pré-análise, e o exposto fica evidente quando se

compara os resultados do Quadro 2 (coluna de pseudo R^2 de McFadden) com os resultados do Quadro 6.

Isto posto, também é necessário verificar a significância estatística de cada parâmetro das modelagens através da estatística z de Wald, dada por $(\hat{\beta}_{jm}/se)^2$, que aponta se um parâmetro é ou não estatisticamente diferente de zero a determinado nível de significância. Os intervalos de confiança dos parâmetros dessa estatística podem ser descritos como $\exp[\hat{\beta}_{jm} \pm 1,96(se)]$.

Para o modelo Q, há um total de $18\beta_{jm}$, sendo que 15 deles, ou seja, 83,33%, são estatisticamente diferentes de 0, em um nível de significância de 10%. Aponta-se que 14 deles, ou seja, 77,78%, são estatisticamente diferentes de 0, em um nível de significância de 5%. Também é interessante notar que metade dos parâmetros β_{jm} são estatisticamente diferentes de 0, em um nível de significância de 1%.

No modelo R1, há um total de $12\beta_{jm}$, sendo que 8 deles, ou seja, 66,67%, são estatisticamente diferentes de 0, em um nível de significância de 10%. Aponta-se que 8 deles, ou seja, 66,67%, são estatisticamente diferentes de 0, em um nível de significância de 5%. Também é interessante notar que 7 dos parâmetros β_{jm} , ou seja, 58,33%, são estatisticamente diferentes de 0, em um nível de significância de 1%.

No modelo R2, há um total de $20\beta_{jm}$, sendo que 13 deles, ou seja, 65%, são estatisticamente diferentes de 0, em um nível de significância de 10%. Aponta-se que 11 deles, ou seja, 55%, são estatisticamente diferentes de 0, em um nível de significância de 5%. Também é interessante notar que metade dos parâmetros β_{jm} são estatisticamente diferentes de 0, em um nível de significância de 1%.

No modelo S1, há um total de $9\beta_{jm}$, sendo que 7 deles, ou seja, 77,78%, são estatisticamente diferentes de 0, em um nível de significância de 10%. Aponta-se que 6 deles, ou seja, 66,67%, são estatisticamente diferentes de 0, em um nível de significância de 5%. Também é interessante notar que 4 dos parâmetros β_{jm} , ou seja, 44,44%, são estatisticamente diferentes de 0, em um nível de significância de 1%.

No modelo S2, há um total de $9\beta_{jm}$, sendo que 8 deles, ou seja, 88,89%, são estatisticamente diferentes de 0, em um nível de significância de 10%. Aponta-se que 7 deles, ou seja, 77,78%, são estatisticamente diferentes de 0, em um nível de significância de 5%. Também é interessante notar que 4 dos parâmetros β_{jm} , ou seja, 44,44%, são estatisticamente diferentes de 0, em um nível de significância de 1%.

O Quadro 5 resume o discutido sobre a estatística z de Wald.

Quadro 5 Análise resumo da estatística z de Wald presente nas modelagens propostas

Modelagem	Quantidade Total de β_{jm}	Quantidade de $\beta_{jm} \neq 0$, sig. 10%	Quantidade de $\beta_{jm} \neq 0$, sig. 5%	Quantidade de $\beta_{jm} \neq 0$, sig. 1%
Q	18	15	14	9
R1	12	8	8	7
R2	20	13	11	10
S1	9	7	6	4
S2	9	8	7	4

Fonte: elaborado pelos autores.

5.1 Apresentação das modelagens

Tal qual já discutido no item 1.1, tem-se a expressão geral das probabilidades da RLM

dada por $p_{i_m} = \frac{e^{z_{i_m}}}{\sum_{m=0}^{M-1} e^{z_{i_m}}}$.

Isto posto, tem-se para o modelo Q as seguintes expressões:

$$\begin{aligned}
 p_{iA} &= \frac{1}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + (\dots) + e^{z_{iH}} + e^{z_{iI}} + e^{z_{iJ}})} & p_{iF} &= \frac{e^{z_{iF}}}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + (\dots) + e^{z_{iH}} + e^{z_{iI}} + e^{z_{iJ}})} \\
 p_{iB} &= \frac{e^{z_{iB}}}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + (\dots) + e^{z_{iH}} + e^{z_{iI}} + e^{z_{iJ}})} & p_{iG} &= \frac{e^{z_{iG}}}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + (\dots) + e^{z_{iH}} + e^{z_{iI}} + e^{z_{iJ}})} \\
 p_{iC} &= \frac{e^{z_{iC}}}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + (\dots) + e^{z_{iH}} + e^{z_{iI}} + e^{z_{iJ}})} & p_{iH} &= \frac{e^{z_{iH}}}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + (\dots) + e^{z_{iH}} + e^{z_{iI}} + e^{z_{iJ}})} \\
 p_{iD} &= \frac{e^{z_{iD}}}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + (\dots) + e^{z_{iH}} + e^{z_{iI}} + e^{z_{iJ}})} & p_{iI} &= \frac{e^{z_{iI}}}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + (\dots) + e^{z_{iH}} + e^{z_{iI}} + e^{z_{iJ}})} \\
 p_{iE} &= \frac{e^{z_{iE}}}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + (\dots) + e^{z_{iH}} + e^{z_{iI}} + e^{z_{iJ}})} & p_{iJ} &= \frac{e^{z_{iJ}}}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + (\dots) + e^{z_{iH}} + e^{z_{iI}} + e^{z_{iJ}})}
 \end{aligned}$$

Em que o valor dos interceptos e dos betas podem ser encontrados no Quadro 2 e estão transcritos abaixo. Aponte-se que os respectivos intervalos de confiança de 95%, descritos como $\alpha = \pm 1,96[s. e. (\alpha)]$ e $\beta_j = \pm 1,96[s. e. (\beta_j)]$, que para cada parâmetro β_{jm} (em que $j = (1; 2; 3; \dots; k)$) do evento de interesse, também podem ser encontrados no Quadro 2.

$$\begin{aligned}
 Z_{iB} &= 30,500 - 0,100. qt_stricto_capes_total - 0,596. qt_stricto5_total \\
 Z_{iC} &= 34,359 - 0,164. qt_stricto_capes_total - 0,650. qt_stricto5_total \\
 Z_{iD} &= 38,397 - 0,263. qt_stricto_capes_total - 0,866. qt_stricto5_total \\
 Z_{iE} &= 39,063 - 0,266. qt_stricto_capes_total - 1,089. qt_stricto5_total \\
 Z_{iF} &= 38,681 - 0,259. qt_stricto_capes_total - 1,014. qt_stricto5_total \\
 Z_G &= 41,774 - 0,500. qt_stricto_capes_total - 1,301. qt_stricto5_total \\
 Z_{iH} &= 4,439 - 0,568. qt_stricto_capes_total - 1,769. qt_stricto5_total \\
 Z_{iI} &= 43,515 - 0,810. qt_stricto_capes_total - 1,095. qt_stricto5_total \\
 Z_{iJ} &= 44,036 - 0,906. qt_stricto_capes_total - 1,763. qt_stricto5_total
 \end{aligned}$$

Para o modelo R1, tem-se:

$$\begin{aligned}
 p_A &= \frac{1}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + e^{z_{iE}})} & p_D &= \frac{e^{z_{iD}}}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + e^{z_{iE}})} \\
 p_B &= \frac{e^{z_{iB}}}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + e^{z_{iE}})} & p_E &= \frac{e^{z_{iE}}}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + e^{z_{iE}})} \\
 p_C &= \frac{e^{z_{iC}}}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + e^{z_{iE}})}
 \end{aligned}$$

Em que o valor dos interceptos e dos betas podem ser encontrados no Quadro 3 e estão transcritos abaixo. Aponte-se que os respectivos intervalos de confiança de 95%, descritos como $\alpha = \pm 1,96[s. e. (\alpha)]$ e $\beta_j = \pm 1,96[s. e. (\beta_j)]$, que para cada parâmetro β_{jm} (em que $j = (1; 2; 3; \dots; k)$) do evento de interesse, também podem ser encontrados no Quadro 3.

$$\begin{aligned}
 Z_{iB} &= 6,0841 - 0,0033. qt_tec_especializacao - 0,2967. qt_doutorado_capes_total - 0,0004. qt_a_scopus \\
 Z_{iC} &= 7,8312 - 0,0028. qt_tec_especializacao - 0,5231. qt_doutorado_capes_total - 0,0046. qt_a_scopus \\
 Z_{iD} &= 11,3026 - 0,0084. qt_tec_especializacao - 1,1140. qt_doutorado_capes_total - 0,0232. qt_a_scopus \\
 Z_{iE} &= 13,1370 - 0,0221. qt_tec_especializacao - 1,2903. qt_doutorado_capes_total - 0,0463. qt_a_scopus
 \end{aligned}$$

Considerando-se o modelo R2, tem-se:

$$\begin{aligned}
 p_A &= \frac{1}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + e^{z_{iE}})} & p_D &= \frac{e^{z_{iD}}}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + e^{z_{iE}})} \\
 p_B &= \frac{e^{z_{iB}}}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + e^{z_{iE}})} & p_E &= \frac{e^{z_{iE}}}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + e^{z_{iE}})} \\
 p_C &= \frac{e^{z_{iC}}}{1 + (e^{z_{iB}} + e^{z_{iC}} + e^{z_{iD}} + e^{z_{iE}})}
 \end{aligned}$$

Em que o valor dos interceptos e dos betas podem ser encontrados no Quadro 4 e estão transcritos abaixo. Aponte-se que os respectivos intervalos de confiança de 95%, descritos como $\alpha = \pm 1,96[s. e. (\alpha)]$ e $\beta_j = \pm 1,96[s. e. (\beta_j)]$, que para cada parâmetro β_{jm} (em que $j = (1; 2; 3; \dots; k)$) do evento de interesse, também podem ser encontrados no Quadro 4.

$$\begin{aligned} Z_{i_B} &= 7,575 - 0,014. qt_prof_doutorado + 0,025. qt_prof_mestrado - 0,044. qt_prof_especializacao \\ &\quad + 0,001. qt_tec_medio - 0,002. qt_tec_superior \\ Z_{i_C} &= 10,227 - 0,022. qt_prof_doutorado + 0,022. qt_prof_mestrado - 0,016. qt_prof_especializacao \\ &\quad + 0,012. qt_tec_medio - 0,011. qt_tec_superior \\ Z_{i_D} &= 12,868 - 0,032. qt_prof_doutorado + 0,027. qt_prof_mestrado - 0,016. qt_prof_especializacao \\ &\quad + 0,007. qt_tec_medio - 0,003. qt_tec_superior \\ Z_{i_E} &= 13,518 - 0,089. qt_prof_doutorado + 0,117. qt_prof_mestrado - 0,124. qt_prof_especializacao \\ &\quad + 0,028. qt_tec_medio - 0,039. qt_tec_superior \end{aligned}$$

Para o modelo S1, tem-se:

$$\begin{aligned} p_A &= \frac{1}{1 + (e^{Z_{i_B}} + e^{Z_{i_C}} + e^{Z_{i_D}})} & p_C &= \frac{e^{Z_{i_C}}}{1 + (e^{Z_{i_B}} + e^{Z_{i_C}} + e^{Z_{i_D}})} \\ p_B &= \frac{e^{Z_{i_B}}}{1 + (e^{Z_{i_B}} + e^{Z_{i_C}} + e^{Z_{i_D}})} & p_D &= \frac{e^{Z_{i_D}}}{1 + (e^{Z_{i_B}} + e^{Z_{i_C}} + e^{Z_{i_D}})} \end{aligned}$$

Em que o valor dos interceptos e dos betas podem ser encontrados no Quadro 5 e estão transcritos abaixo. Aponte-se que os respectivos intervalos de confiança de 95%, descritos como $\alpha = \pm 1,96[s. e. (\alpha)]$ e $\beta_j = \pm 1,96[s. e. (\beta_j)]$, que para cada parâmetro β_{jm} (em que $j = (1; 2; 3; \dots; k)$) do evento de interesse, também podem ser encontrados no Quadro 5.

$$\begin{aligned} Z_{i_B} &= 5,474 - 0,145. qt_mestrado_capes_total - 0,236. qt_doutorado5 - 0,001. qt_a_scopus \\ Z_{i_C} &= 7,328 - 0,024. qt_mestrado_capes_total - 0,876. qt_doutorado5 - 0,024. qt_a_scopus \\ Z_{i_D} &= 11,161 - 0,570. qt_mestrado_capes_total - 1,316. qt_doutorado5 - 0,050. qt_a_scopus \end{aligned}$$

Finalmente, para o modelo S2, tem-se:

$$\begin{aligned} p_A &= \frac{1}{1 + (e^{Z_{i_B}} + e^{Z_{i_C}} + e^{Z_{i_D}})} & p_C &= \frac{e^{Z_{i_C}}}{1 + (e^{Z_{i_B}} + e^{Z_{i_C}} + e^{Z_{i_D}})} \\ p_B &= \frac{e^{Z_{i_B}}}{1 + (e^{Z_{i_B}} + e^{Z_{i_C}} + e^{Z_{i_D}})} & p_D &= \frac{e^{Z_{i_D}}}{1 + (e^{Z_{i_B}} + e^{Z_{i_C}} + e^{Z_{i_D}})} \end{aligned}$$

Em que o valor dos interceptos e dos betas podem ser encontrados no Quadro 6 e estão transcritos abaixo. Aponte-se que os respectivos intervalos de confiança de 95%, descritos como $\alpha = \pm 1,96[s. e. (\alpha)]$ e $\beta_j = \pm 1,96[s. e. (\beta_j)]$, que para cada parâmetro β_{jm} (em que $j = (1; 2; 3; \dots; k)$) do evento de interesse, também podem ser encontrados no Quadro 6.

$$\begin{aligned} Z_{i_B} &= 5,852 + 0,004. qt_tec_superior - 0,039. qt_tec_doutorado - 0,229. qt_mestrado_capes_total \\ Z_{i_C} &= 6,560 + 0,008. qt_tec_superior - 0,162. qt_tec_doutorado - 0,289. qt_mestrado_capes_total \\ Z_{i_D} &= 11,485 + 0,014. qt_tec_superior - 0,033. qt_tec_doutorado - 1,532. qt_mestrado_capes_total \end{aligned}$$

Fávero & Belfiore (2017) apontam que, diferentemente de uma Regressão Logística Binária, que é avaliada a partir de um *cut off* definido, na RLM a classificação das observações da amostra deve ser feita tomando por base a maior probabilidade entre as calculadas. Tal procedimento refere-se à Eficiência Global do Modelo de RLM – EGM. O Quadro 6 resume o quantitativo de erros e de acertos das modelagens, evidenciando a EGM em função do ranqueamento *Webometrics* observada em 2018.

Quadro 6 Resumo da EGM do Modelo de RLM para o ranqueamento *Webometrics* em 2018

Modelo	Amostra	Acertos	Erros	EGM
Q	183	75	108	40.98%
R1	183	113	72	61.75%
R2	96	70	26	72.92%

S1	183	135	50	73.77%
S2	96	76	20	79.17%

Fonte: elaborado pelos autores.

Yerushalmy (1947), em seu seminal trabalho, aprofundou os importantes conceitos de sensibilidade e de especificidade, definindo a sensibilidade como a capacidade de acerto da probabilidade de um dado evento ser, de fato, classificado como evento; já a especificidade refere-se à capacidade de acerto da probabilidade de um não evento ser, de fato, classificado como não evento. Ambos, idealmente, deveriam ser altos (Agresti, 2013).

Isto posto, apresentam-se dois quadros a seguir que resumem a sensibilidade (*sensit*) e a especificidade (*espec*), em função da ocorrência ou não de um evento (*even* e *~even*, respectivamente) de cada modelo, individualizando-se os resultados por *label* de interesse.

Quadro 7 Resumo da sensibilidade e da especificidade dos modelos Q, R1 e R2, em função da ocorrência ou não de um evento, individualizado por *label* de interesse

	Modelo Q				Modelo R1				Modelo R2			
	even	~even	sensit	espec	even	~even	sensit	espec	even	~even	sensit	espec
Classificado como A	16	1	94.1%	99.4%	31	3	86.1%	98.0%	30	4	90.9%	93.7%
Classificado como não A	1	165			5	144			3	59		
Classificado como B	13	5	68.4%	97.0%	17	11	45.9%	92.5%	15	9	68.2%	87.8%
Classificado como não B	6	159			20	135			7	65		
Classificado como C	10	10	55.6%	93.9%	21	22	56.8%	84.9%	11	8	55.0%	89.5%
Classificado como não C	8	155			16	124			9	68		
Classificado como D	7	17	36.8%	89.6%	16	16	44.4%	89.1%	8	4	57.1%	95.1%
Classificado como não D	12	147			20	131			6	78		
Classificado como E	6	15	31.6%	90.9%	28	18	75.7%	87.7%	6	1	85.7%	98.9%
Classificado como não E	13	149			9	128			1	88		
Classificado como F	0	1	0.0%	99.4%								
Classificado como não F	17	165										
Classificado como G	2	12	10.5%	92.7%								
Classificado como não G	17	152										
Classificado como H	8	26	44.4%	84.2%								
Classificado como não H	10	139										
Classificado como I	1	4	5.3%	97.6%								
Classificado como não I	18	160										
Classificado como J	12	16	66.7%	90.3%								
Classificado como não J	6	149										

Fonte: elaborado pelos autores.

Quadro 8 Resumo da sensibilidade e da especificidade dos modelos S1 e S2, em função da ocorrência ou não de um evento, individualizado por *label* de interesse

	Modelo S1				Modelo S2			
	even	~even	sensit	espec	even	~even	sensit	espec
Classificado como A	38	5	84.4%	96.4%	38	3	92.7%	94.5%
Classificado como não A	7	133			3	52		
Classificado como B	30	12	65.2%	91.2%	16	8	64.0%	88.7%
Classificado como não B	16	125			9	63		
Classificado como C	29	18	63.0%	86.9%	12	7	63.2%	90.9%
Classificado como não C	17	119			7	70		
Classificado como D	38	13	82.6%	90.5%	10	2	90.9%	97.6%
Classificado como não D	8	124			1	83		

Fonte: elaborado pelos autores.

Conforme o demonstrado no Quadro 6, o modelo Q tem a menor EGM (40.98%) entre todas as modelagens propostas, mas deve-se considerar que é o único que divide em dez grupos toda a amostra de 183 universidades, públicas e privadas, pagas e gratuitas.

Notou-se uma alta sensibilidade do modelo para o acerto preditivo para as universidades agrupadas na *label A*, totalizando 94,1% de acertos a 95% de confiança.

As maiores limitações de acertos preditivos de eventos encontraram-se nas *labels F e I*, com capacidade de acertos desprezíveis, por outro lado, pondera-se que essas mesmas *labels* junto com *labels A e B*, detêm a maior capacidade de acerto de não eventos da modelagem *Q*, a 95% de confiança.

O modelo R1 contou com uma amostra de 183 universidades, públicas e privadas, pagas e gratuitas, divididas em cinco grupos. Sua EGM totalizou 61.75% e notou-se uma alta sensibilidade do modelo para o acerto preditivo das universidades agrupadas na *label A*, totalizando 86,1% de acertos, a 95% de confiança. A *label A* detém, também, a maior especificidade em relação às outras *labels* do modelo.

A sua maior limitação preditiva encontra-se na *label D*, em que acerta as previsões de evento 44.4% das vezes, a 95% de confiança. Por outro lado, a *label D* detém uma capacidade de acerto de não eventos de 89,1%, a 95% de confiança – o que é considerável.

O modelo R2 contou com uma amostra de 96 universidades gratuitas, divididas em cinco grupos. Sua EGM totalizou 72.92% e, igualmente às modelagens anteriores, notou-se uma alta sensibilidade do modelo para o acerto preditivo das universidades agrupadas na *label A*, totalizando 90.9% de acertos, a 95% de confiança.

A sua maior limitação preditiva com relação ao acerto de eventos encontra-se na *label C*, em que acerta as previsões 55.0% das vezes, a 95% de confiança. Mesmo assim, a *label C* detém a segunda maior capacidade de acerto de não eventos do modelo – 95,1%. A maior especificidade do modelo R2 está na *label D*, 98,9% de acertos, a 95% de confiança.

O modelo S1 contou com uma amostra de 183 universidades, públicas e privadas, pagas e gratuitas, divididas em quatro grupos. Sua EGM totalizou 73.77% e, igualmente às modelagens anteriores, notou-se uma alta sensibilidade do modelo para o acerto preditivo das universidades agrupadas na *label A*, totalizando 84.4% de acertos a 95% de confiança. A *label A* também detém a maior especificidade deste modelo, sendo 96,4%, a 95% de confiança.

A sua maior limitação preditiva relacionada à sensibilidade, encontra-se na *label C*, em que acerta as previsões 63.0% das vezes a 95% de confiança. Ainda assim, há que se ponderar que a *label C* detém uma especificidade de 86,9%, a 95% de confiança.

Por fim, o modelo S2 contou com uma amostra de 96 universidades gratuitas, divididas em cinco grupos. O modelo teve a maior EGM de todos, 79.17% e, igualmente às modelagens anteriores, notou-se uma alta sensibilidade do modelo para o acerto preditivo das universidades agrupadas na *label A*, totalizando 92.7% de acertos a 95% de confiança.

A sua maior limitação preditiva relacionada à sensibilidade também se encontra na *label C*, em que acerta as previsões 63.2% das vezes a 95% de confiança. Ainda assim, a *label C* detém uma especificidade de 90,9%, a 95% de confiança.

A maior especificidade da modelagem S2 é encontrada na *label D*, que consegue acertar os não eventos em 97,6% das vezes, a 95% de confiança.

5.2 Discussões finais

Como discussões finais e em relação às hipóteses elaboradas, pode-se dizer que, de acordo com a EGM apresentado por cada um dos modelos no Quadro 6, todas as hipóteses se confirmaram. Nota-se ainda que as H_{2b} e H_{3b} , que abordavam as universidades públicas mostraram-se mais eficientes que o modelo completo. Isso mostra que a natureza da universidade e seus objetivos estratégicos são relevantes para a eficiência do modelo. Isto é, ter a orientação para a pesquisa e a criação do conhecimento científico, faz com que elas tenham maior aderência aos objetivos das universidades no Brasil e tenham mais aderência aos modelos. Comprova-se que elaborar níveis de desempenho para se gerenciar o desempenho de universidades é adequado, conforme preceitos de Liefner (2003). Isso se justifica por haver

menos mercantilização do conhecimento em universidades públicas, tal como propõem Mancebo *et al.* (2016), e que há maior orientação por carreira que sustenta as universidades públicas (Rowe & Bastos, 2010). Portanto, é relevante que se tenha cuidado com a natureza e orientação estratégica das universidades ao se considerar os seus modelos de desempenho, tal como preconiza estudos prévios (Carneiro-da-Cunha & Corrêa, 2013). Avança-se nessa discussão quando se mostra que isso é reforçado quando a orientação estratégica das organizações se alinha ao modelo elaborado. Isto é, quando se entende que modelos de avaliação de desempenho que consideram níveis de desempenho se adequam melhor às organizações mais alinhadas com as variáveis e indicadores do modelo. Ao se definir tipos de organizações (no caso, universidades públicas e privadas), sua natureza e orientação estratégica interferem na eficiência do modelo.

6 CONCLUSÕES

O presente estudo teve como objetivo propor um modelo de segmentação dos níveis de desempenho das universidades brasileiras. Trata-se de uma alternativa para os atuais modelos de níveis de desempenho obtidos por critérios arbitrários ou subjetivos.

Como contribuição teórica, o presente estudo demonstra, na área de Gestão Estratégica de Desempenho, que estratificar níveis de desempenho deve ser de forma não arbitrária para que se possa representar os grupos de desempenho entre organizações. Adicionalmente, o presente estudo mostra que diferentes categorias ou tipos de organizações que possam trazer orientações estratégicas discrepantes requerem cuidado na sua comparação de desempenho. Modelos específicos podem ter mais aderência a um tipo específico. Assim, ao se considerar modelos que envolvam níveis de desempenho, devem ser consideradas as diferentes orientações estratégicas, o que se pôde observar quando universidades públicas têm maior adequação a um modelo que abrange a produção científica. O trabalho traz como contribuição metodológica a aplicação de um método, não antes utilizado, objetivando determinar segmentos ideais para se tratar os níveis de desempenho de universidades no Brasil. É a partir dessa segmentação quantitativa elaborada por meio de RLM que se pode ter uma segmentação mais adequada à realidade e características das universidades nacionais.

Como contribuição para a prática gerencial, tem-se cinco modelos (Q, R1, R2, S1, S2) com até 13 variáveis explicativas que pode ser utilizado para controlar o desempenho das universidades brasileiras conforme seus estratos de desempenho. Esse modelo tem ainda mais relevância quando se trata das universidades públicas. O modelo proposto maximiza as eficiências globais e pode, portanto, estabelecer níveis de desempenho de maneira mais assertiva. É um modelo que pode ser utilizado pelos gestores educacionais para controles internos e também para comparação entre diferentes instituições. Gestores públicos podem ainda utilizar-se desse modelo para avaliar o desempenho de universidades com maior ênfase e propor diferentes políticas e incentivos para as universidades conforme seus níveis de desempenho. Ou seja, maior poder de atuação para atender objetivos específicos de políticas educacionais no Brasil.

A limitação do modelo consiste que as variáveis que o compõem são dados secundários que não foram coletados exclusivamente para fins de pesquisas acadêmicas. Isso não exime a utilidade e validade deles, mas é um ponto de atenção. Adicionalmente, algumas variáveis podem ser específicas a uma determinada forma de mensurar o desempenho. Por exemplo, a produção científica (variável *qt_a_scopus*) é avaliada pelos trabalhos indexados no *Scopus* e poderia ser na base do *Web of Science*, o que pode ser mais pertinente para algumas áreas do conhecimento.

Há que se apontar que as modelagens preditivas possuem especial dificuldade em se explicar fenômenos; em contraponto, são ideais para a previsão de ocorrência deles (Agresti, 2013; Hair *et al.*, 2009). Assim, repousa a necessidade de se indicar para os estudos futuros o

uso de técnicas estatísticas exploratórias – e.g. análise de *clusters*, análise fatorial, análise de correspondências – para se melhor entender a explicação precisa para a aderência das variáveis adotadas.

Como estudo futuro, propõe-se, também, elaborar a mesma segmentação em universidades de outros países, com características de educação e ensino diferentes para que se possa, posteriormente, comparar os modelos e discutir as diferenças e semelhanças, especialmente sobre quais variáveis explicativas são relevantes em cada realidade. Complementarmente, novas variáveis podem ser trabalhadas a partir de coleta de dados tal como, por exemplo, a quantidade de *spin-offs* que uma universidade gera para a sociedade (O’Shea *et al.*, 2005).

REFERÊNCIAS

- Agresti, A. (2013). *Categorical Data Analysis*. (3ª ed.). Hoboken: John Wiley & Sons.
- Alani, F. S., Khan, M. F. R. & Manuel, D. F. (2017). University performance evaluation and strategic mapping using balanced scorecard (BSC): case study - Sohar University, Oman. *International Journal of Educational Management*, no prelo.
- Alexander, F. K. (2000). The changing face of accountability monitoring and assessing institutional performance in higher education. *Journal of Higher Education*, 71(4), 411-431.
- Ayçaguer, L. C. S. & Utra, I. M. B. (2004). *Regresión Logística*. Madrid: La Muralla.
- Azen, R. & Walker, C. M. (2011). *Categorical Data Analysis for the Behavioral and Social Sciences*. New York: Routledge.
- Billor, N., Hadi, A. S., Velleman, P. F. (2000). BACON: Blocked Adaptive Computational Efficient Outlier Nominator. *Computational Statistics & Data Analysis*, 34(3), 279-298.
- Bititci, U., Garengo, P., Dörfler, V. & Nudurupati, S. (2012). Performance measurement: challenges for tomorrow. *International Journal of Management Reviews*, 14(3), 305–327.
- Brasil. (2014). Saiba como funciona sistema de ensino superior no Brasil. Disponível em: <http://www.brasil.gov.br/educacao/2009/11/ensino-superior> Acesso em: 01 Mai. 2018.
- Brasil. (2017). Relatório Anual de Análise dos Indicadores de Gestão das Instituições Federais de Educação Profissional, Científica e Tecnológica. Disponível em: <http://portal.mec.gov.br/docman/novembro-2017-pdf/77151-relatorio-indicadores-da-rede-federal-2016-pdf/file> Acesso em: 01 Mai. 2018.
- Carneiro da Cunha, J. A. & Correa, H. L. (2013). Avaliação de desempenho organizacional: um estudo aplicado em hospitais filantrópicos. *Rev. de Adm. de Empresas*, 53(5), 485-499.
- Chen, S., Yang, C. & Shiau, J. (2006) The application of balanced scorecard in the performance evaluation of higher education. *The TQM Magazine*, 18(2), 190-205.
- Dakin, H. A., Devlin, N. J. & Odeyemi, I. A. O. (2006). “Yes”, “No” or “Yes, but”? Multinomial modelling of NICE decision-making. *Health Policy*, 77(3), 352–367.
- Fávero, L. P. (2015). *Análise de Dados: modelos de regressão com Excel, Stata e SPSS*. Rio de Janeiro: Elsevier.
- Fávero, L. P. & Belfiore, P. (2017). *Manual de Análise de Dados: estatística e modelagem multivariada com Excel, SPSS e Stata*. São Paulo: Elsevier.
- Garson, G. D. (2012). *Logistic Regression: binary & multinomial*. Asheboro: Statistical Associates Publishing.
- Hair Jr., J. F., Black, W. C., Babin, B. J., Anderson, R. E. & Tatham, R. L. (2009). *Análise Multivariada de Dados*. (6ª ed.). Porto Alegre: Bookman.
- Hoffmann, C., Zanini, R. R., Corrêa, A. C., Siluk, J. C. M., Schiuch Jr., C. F. & Ávila, L. V. (2014). O desempenho das universidades brasileiras na perspectiva do Índice Geral de Cursos (IGC). *Educação e Pesquisa*, 40(3), 651-666.

- Hosmer, D. W., Lemeshow, S. & Sturdivant, R. X. (2013). *Applied Logistic Regression*. (3^a ed.). New York: John Wiley & Sons.
- Hourneaux Jr, F., Carneiro-da-Cunha, J. A. & Corrêa, H. L. (2017) Performance measurement and management systems: Different usages in Brazilian manufacturing companies. *Managerial Auditing Journal*, 32(2), 148-166
- Izique, C. (2017). Universidades paulistas reavaliam seus indicadores de desempenho institucional. *Revista Amazônia*, 1-2.
- Kairuz, T., Andriés, L., Nickloes, T. & Truter, I. (2016). Consequences of KPIs and performance management in higher education. *International Journal of Educational Management*, 30(6), 881-893.
- Kapetanidou, C. & Lee, S. H. (2016). A framework for assessing the performance of universities: the case of Cyprus. *Technological Forecasting & Social Change*, 123, 169-180.
- Kazlauskaitė, R. & Buciušienė, I. (2008). The role of Human Resources and their management in the establishment of sustainable competitive advantage. *Engineering Economics*, 60(5), 78-84.
- Liefner, I. (2003). Funding, resource allocation, and performance in higher education systems. *Higher Education*, 43(4), 469-489.
- Mancebo, D., Silva Jr., J. & Schugurensky, R. (2016). A educação superior no Brasil diante da mundialização do capital. *Educação em Revista*, 32(4), 205-225.
- Mancebo, D., Vale, A. A. & Martins, T. B. (2015). Políticas de expansão da educação superior no Brasil 1995-2010. *Revista Brasileira de Educação*, 20(60), 31-50.
- Martins, C. B. (2009). A Reforma Universitária de 1968 e a abertura para o ensino superior privado no Brasil. *Educação & Sociedade*, 30(106), 15-35.
- Matthews, J. R. (2011). Assessing organizational effectiveness: the role of performance measures. *Library Quarterly*, 81(1), 83-110.
- Neely, A., Gregory, M., & Platts, K. (2005). Performance measurement system design: a literature review and research agenda. *International Journal of Operations & Production Management*, 25(12), 1228-1263.
- O'Shea, R. P., Allen, T. J., Chevalier, A. & Roche, F. (2005). Entrepreneurial orientation, technology transfer and spinoff performance of U.S. universities. *Research Policy*, 34(7), 994-1009.
- Rowe, D. E. O. & Bastos, A. V. B. (2010). Links between Career and Academic Performance: Comparing Teachers at Private and Public Higher Education Establishments in Brazil. *Revista de Administração Contemporânea*, 14(6), 1011-1030.
- Soares, T. C., Mazon, G. & Soares, J. C. (2017). Indicadores da avaliação institucional: percepção de dirigentes e integrantes das CPA. *Revista Eletrônica de Estratégia & Negócios*, 10(3), 223-247.
- Taticchi, P., Garengo, P., Nudurupati, S.S.; Tonelli, F. & Pasqualino, R. (2015). A review of decision-support tools and performance measurement and sustainable supply chain management. *International Journal of Production Research*, 53(21), 6473-6494.
- Weber, S. (2010). BACON: an effective way to detect outliers in multivariate data using Stata (and Mata). *Stata Journal*, 10(3), 331-338.
- Williamson, O. E. (1991). Strategizing, economizing and economic organization. *Strategic Management Journal*, 12, 75- 94.
- Yerushalmy, J. (1947). Statistical problems in assessing methods of medical diagnosis, with special reference to x-ray techniques. *Public Health Rep.* 62, 1432-1449.