

Explorando o ENADE com o R na Visão de Gestor de Curso

Resumo

O desenvolvimento de modelos de mensuração da qualidade de cursos superiores é assunto bastante discutido na literatura internacional dada a diversidade de dimensões e de interesses na avaliação. No Brasil uma das métricas de mensuração da qualidade do ensino superior e das instituições é o ENADE, com provas realizadas em ciclos trienais de avaliação. Este trabalho tem como principal objetivo investigar os dados de avaliação de 2015 disponibilizados pelo INEP e verificar dentre a composição do Conceito Preliminar de Cursos (CPC), as variáveis em que o gestor deverá concentrar seus esforços. Com este intuito foram utilizadas três técnicas estatísticas: Análise Exploratória de Dados, Análise do Componente Principal e Análise Fatorial Exploratória. Os resultados indicam que a Organização Didático-pedagógica, Infraestrutura e Instalações e Oportunidades de Ampliação da Formação são as variáveis de maior influência na composição do CPC.

Palavras-chave: Avaliação dos Cursos, Análise Exploratória de Dados, Análise do Componente Principal e Análise Fatorial Exploratória, R.

Introdução

A prática de ranqueamento de cursos e instituições, e sua apropriação principalmente nos mecanismos de marketing de instituições de ensino particulares é comum em todo o mundo. No Brasil, o ranqueamento é realizado com base na pontuação no ENADE, Exame Nacional de Desempenho dos Estudantes, cujo objetivo é avaliar o rendimento dos alunos dos cursos de graduação, ingressantes e concluintes, em relação aos conteúdos programáticos dos cursos em que estão matriculados. Seu principal instrumento é o CPC, Conceito Preliminar de Curso, um indicador de qualidade que avalia os cursos de graduação. Seu cálculo e divulgação ocorrem no ano seguinte ao da realização do ENADE, com base na avaliação de desempenho de estudantes, no valor agregado pelo processo formativo e em insumos referentes às condições de oferta – corpo docente, infraestrutura e recursos didático-pedagógicos (INEP, 2017). O CPC é composto por 37 variáveis. Na edição de 2015 foram avaliados 8121 cursos de 1758 instituições de ensino espalhados em 753 municípios de 27 estados brasileiros, com um percentual de participação de 81.4% dos estudantes nas provas.

Há que se investigar a relação entre a diversidade destas variáveis com o vetor qualidade. Para isso é necessário que se racionalize o universo de variáveis em um conjunto delimitado de dimensões. Diante do exposto, este trabalho é resultado da análise das variáveis que constituem o ENADE por meio da redução das dimensionalidades. Acredita-se que seja possível extrair perfis de variáveis representativos dos cursos superiores capazes de descrevê-los em arranjos significativos e de fácil compreensão.

Método

Esta pesquisa utilizou um conjunto de técnicas estatísticas distintas para a análise de dados: Análise do Componente Principal (ACP) e Análise Fatorial (AF). Utilizaram-se os dados do senso de 2015 disponíveis site do INEP. Para leitura utilizou-se o pacote readxl (WICKHAM e BRYAN, 2019), manipulação dos dados foi utilizado R, e o pacote dplyr (WICKHAM et al., 2017) e os gráficos realizados com o pacote ggplot2 (WICKHAM, 2016).

Na análise exploratória dos dados foram analisados: o número de cursos por regiões, o percentual de cursos por organização acadêmica e categoria administrativa. Uma inspeção na variável resposta, CPC, permitiu identificar que 12,97% dos cursos não obtiveram notas. Embora para variáveis preditoras as informações estivessem completas. Foi realizada a imputação dos dados utilizando o método K-NN com o pacote DMwR v0.4. (TORGO, 2013).

A padronização dos dados deve anteceder a ACP (LAROSE e LAROSE, 2015) e foi realizada com a utilização da função scales contida na instalação básica do R.

Tanto a ACP como a AF fizeram amplo uso do pacote psych (REVELLE, 2017). Dada a sua importância, este pacote constitui elemento chave para a aplicação das metodologias estatísticas desta pesquisa.

Explorando o ENADE com o R na Visão de Gestor de Curso

Após a análise do correlograma, prejudicada pelo grande número de variáveis preditoras, partiu-se para a ACP, o critério do Scree Plot (HAIR et al., 1998) indicaria o uso de 5 fatores correspondendo a 78,7% da variabilidade.

Foi ainda adicionada uma camada de robustez a análise com a AF, que requer que as variáveis estejam minimamente correlacionadas. Neste sentido, foi realizado teste de Esfericidade de Bartlett que indicou a adequação da AF.

Discussão e conclusão

Após a aplicação da ACP e AF foi possível identificar os agrupamentos das variáveis. Verificou-se que o primeiro componente tem afinidade com a estrutura pedagógica do curso (Organização Didático-pedagógica), sua estrutura (Infraestrutura e Instalações) e perspectiva de futuro do ponto de vista do aluno (Oportunidades de Ampliação da Formação). Concluímos que esse componente apresenta um perfil do que o curso é capaz de entregar ao aluno, que é refletida nas notas que o mesmo atribui no momento da avaliação. O segundo componente observamos uma concentração dos pesos nas variáveis associadas a duas dimensões: alunos e professores. No que se refere aos alunos, pesam mais as notas avaliadas de acordo com o desempenho dos mesmos nas provas nas áreas de Formação Geral e Conhecimentos Específicos. Já em relação aos professores tem destaque as variáveis Quantidade de Mestres e Quantidade de Doutores nos cursos. O terceiro componente dá destaque as variáveis Concluintes Inscritos e Percentual de Concluintes Participantes (CP) com nota no ENEM, Exame Nacional do Ensino Médio. Portanto indica que este componente busca determinar a situação dos alunos no censo bem como a contabilização da participação dos mesmos. O quarto componente não apresenta diferença em relação ao terceiro em termos de composição.

A utilização da ACP e AF apresentaram quais variáveis são passíveis de melhor explicarem o que influencia no desempenho do curso no CPC e portanto devem ser objeto de maior preocupação por parte dos gestores os cursos e instituições.

Os gráficos, tabelas e códigos deste trabalho podem ser encontrados em https://github.com/gui-souza/ACP_AF_Aplicada.

Em áreas como Ciências Sociais por exemplo, 60% da variabilidade é considerada satisfatória dada a imprevisibilidade da natureza do comportamento humano (LAROSE; LAROSE, 2015, p. 103).

Concluímos que a identificação e análise dos três primeiros componentes são relevantes e constituem contribuição importante desta pesquisa, vez que os mesmos contabilizam 60,1% de toda a variabilidade.

Bibliografia

- HAIR, J. F. et al. Multivariate data analysis. [s.l.] Prentice hall Upper Saddle River, NJ, 1998. v. 5
- INEP. Avaliação in Loco - Glossário dos Instrumentos de Avaliação Externa. jul. 2018. JOSSE, LAROSE, D. T.; LAROSE, C. D. Data mining and predictive analytics. [s.l.] John Wiley & Sons, 2015.
- REVELLE, W. psych: Procedures for Psychological, Psychometric, and Personality Research. Evanston, Illinois: Northwestern University, 2017.
- TORGO, L. Data Mining with R, learning with case studies. [s.l.] Chapman and Hall/CRC, 2010.
- WICKHAM, H. et al. dplyr: A Grammar of Data Manipulation. [s.l.: s.n.].
- WICKHAM, H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. Disponível em: <http://ggplot2.org>.
- WICKHAM, H.; BRYAN, J. readxl: Read Excel Files. [S.l.], 2019. R package version 1.3.1. Disponível em: <https://CRAN.R-project.org/package=readxl>.
- WICKHAM, H. et al. dplyr: A Grammar of Data Manipulation. [S.l.], 2018. R package version 0.7.6. Disponível em: <https://CRAN.R-project.org/package=dplyr>.