

Reamostragem e detecção de outliers em seleção automática de variáveis para modelos de equações estruturais na análise de satisfação de alunos do ensino superior privado

PROPOSTA DE TRABALHO

Gabriel Torres da Rocha Pehls
gabrielpehls@hotmail.com

Curso de Sistemas de Informação
Centro Universitário Ritter Dos Reis - UNIRITTER

Matheus Bento de Pereira
matheusbentodepereira@yahoo.com.br
Curso de Analise e desenvolvimento de sistemas
Centro Universitário Ritter Dos Reis - UNIRITTER

1. Relevância do trabalho

De acordo com o Censo da Educação Superior de 2017[5, 6], o volume de ingressantes no ensino superior cresceu 27,3% na modalidade a distância, enquanto os cursos presenciais tiveram um acréscimo de 0,5%. As matrículas na rede privada cresceram 3,0%, sendo uma parcela expressiva do total de alunos de graduação (75,3%). Tendo em vista este cenário, um dos fatores de interesse das instituições é entender como se compõe a satisfação geral dos alunos.

Essa análise visa melhorar as experiências dos alunos na percepção da qualidade dos serviços, bem como priorizar quais deles tem maior impacto na percepção dos alunos. A satisfação dos estudantes é um dos grandes objetivos das universidades, tornando-se uma vantagem competitiva quando o estudante promove a instituição, elevando a lealdade, a retenção de estudantes e a imagem da mesma perante a sociedade [1].

Segundo [22], a investigação da satisfação acadêmica é um elemento importante na avaliação da eficácia institucional, possibilitando que as instituições reestruturem-se e se adaptem para atender às necessidades dos estudantes.

Muitos fatores já foram teorizados como influentes na satisfação dos estudantes, como a figura do professor e do curso[17, 15, 9], fatores relacionados a conteúdos online[16], relacionados á qualidade de serviços[12, 14], infraestrutura, que ainda podem ser subdivididos em subgrupos, como salas de aula[2], campus [7], aspectos relacionados á bibliotecas ou outras estruturas físicas[25, 9, 13], corpo docente[3, 15, 20], a administração do curso[20], oportunidades na carreira escolhida[9], serviços de apoio ao estudante, seja de origem financeira, matrícula ou de suporte, ou ainda serviços acadêmicos, levando em conta serviços utilizados pelos discentes[22], e ainda agências de viagem[13].

A noção de custo-benefício também é destacada em muitos dos estudos citados, bem como a imagem da instituição perante a sociedade, e a lealdade do estudante, seja um egresso ou um aluno corrente[1, 14, 22, 20], onde identificamos que a imagem é um fator que pode influenciar na própria satisfação geral, na lealdade do estudante e ainda em custo-benefício, enquanto que este último pode ter relação com a satisfação do estudante, formada por diversos itens comentados anteriormente.

Dentre muitas técnicas disponíveis para analisar todos estes fatores, destaca-se a Modelagem de Equações Estruturais (MEE, ou ainda, SEM, de *Structural Equation Modeling*), que é uma família de modelos estatísticos que procuram explicar as relações entre múltiplas variáveis, expressadas em uma série de equações, compostas por **constructos** (as variáveis dependentes e independentes), que são fatores não observados ou latentes representados por múltiplas variáveis (muito parecido com variáveis representando um fator na análise fatorial)[11].

Um modelo de SEM permite avaliar a relação entre as variáveis independentes, a magnitude de sua influência na variável dependente e a relação entre as variáveis externas ao modelo e as variáveis independentes, sendo muito utilizada quando se tem interesse em estudar e obter respostas a partir de conjuntos de variáveis relacionadas, como pesquisas de mercado, sendo utilizada para entender o impacto de medidas referentes ao consumidor e sua satisfação com o produto/serviço[10].

O modelo de SEM pode ser dividido entre um Modelo de medida (ou de mensuração), que representa esquematicamente as variáveis manifestas ou medidas, associando-as a cada constructo, um Modelo estrutural, que representa o conjunto de uma ou mais relações de dependência, conectando os constructos do modelo em um modelo de Regressão, que permite avaliar as relações entre os constructos[8].

Portanto, seguindo esse modelo, a avaliação é realizada em duas etapas, utilizando-se de uma Análise Fatorial Confirmatória (AFC, ou ainda, CFA, de *Confirmatory Factorial Analysis*), etapa que permite validar os constructos, e uma segunda etapa, que avalia os índices de ajuste do modelo da regressão, a significância e a magnitude dos coeficientes das regressões estimadas[8, 23].

A análise de equações estruturais exige, ainda, a validação de algumas suposições estatísticas para a correta análise do modelo teórico, que se ignoradas, podem comprometer os resultados das análises e as conclusões da mesma[11, 8, 18, 19, 21]:

- Independência das observações, ou seja, as observações de sujeitos diferentes devem ser independentes entre si;
- Amostragem aleatória de respondentes (também chamada de *Bootstrap*), onde se aconselha escolher uma amostragem probabilística, como por exemplo a Amostra Aleatória Simples, onde devemos ter a mesma probabilidade de escolha para todos os casos de uma população;
- Linearidade do modelo, ou seja, o modelo possui propriedades de aditividade e homogeneidade. Em geral, os métodos de análise multivariados assumem relações lineares entre as variáveis manifestas e as variáveis latentes e entre as variáveis latentes.
- Ausência de Multicolinearidade, ou seja, as variáveis independentes não podem se encontrar fortemente correlacionadas. A presença de multicolinearidade pode produzir coeficientes padronizados superiores a 1 ou inferiores a -1, ou estimativas de variâncias negativas.
- Inexistência de Outliers, o que pode inflacionar ou reduzir as covariâncias entre as variáveis, repercutindo nas estimativas das médias, desvios-padrões e covariâncias, comprometendo a qualidade do ajustamento do modelo. Podemos detectar Outliers utilizando-se a análise de pontos extremos através dos quartis, ou ainda a *Distância de Mahalanobis*, que visa estimar a distância de uma observação x à média de todas as observações de todas as variáveis. Este método pode ser aplicado, inclusive, com dados onde as suposições da normalidade não são cumpridas ou há um grande número de *missings*[24].

2. Objetivos

2.1. Objetivo geral

Este trabalho propõe um estudo sobre a aplicação da técnica de amostragem aleatória simples, em conjunto com técnicas de tratamento de bases, no contexto de equações estruturais com múltiplas variáveis latentes e com um grande número de respondentes, tendo como comparativo resultados da aplicação de equações estruturais em um estudo de satisfação de alunos de uma instituição de ensino superior brasileira, sem os métodos propostos.

2.2. Objetivos específicos

Este trabalho visa avaliar o efeito do método de reamostragem aleatória simples em equações estruturais, bem como a utilização de técnicas de pré-processamento de bases, como a detecção de outliers e remoção de *missings*. Podemos destacar, entre os objetivos específicos deste trabalho:

- Verificar a precisão, a falta de suposição e a vantagem da *generalidade* do método *Bootstrap*[4];
- Verificar o impacto da utilização do método de detecção de *outliers* pela *Distância de Mahalanobis* e pela variância em conjunto com o *Bootstrap* em dados com normalidade multivariada;
- Demonstrar a suposição de que *mais nem sempre é o melhor* e que *um tamanho não serve para todos*[26], através da análise de modelos gerados com tamanho amostral variáveis.

Referências

- [1] R. Arambewela and J. Hall. An empirical model of international student satisfaction. *Asia Pacific journal of marketing and logistics*, 21(4):555–569, 2009. 1
- [2] B. Z. Butt and K. ur Rehman. A study examining the students satisfaction in higher education. *Procedia-Social and Behavioral Sciences*, 2(2):5446–5450, 2010. 1
- [3] H. Camgoz-Akdag and S. Zaim. Education: a comparative structural equation modeling study. *Procedia-Social and Behavioral Sciences*, 47:874–880, 2012. 1
- [4] A. S. da Silva Filho. Inferência em amostras pequenas: método bootstrap. *Revista de Ciências exatas e tecnologia*, 5(5):115–126, 2015. 2
- [5] I. I. N. de Estudos e Pesquisas Educacionais Anísio Teixeira. Censo da educação superior, 2017. disponível em : http://download.inep.gov.br/educacao_superior/censo_superior/documentos/2018/censo_da_educacao_superior_2017-notas_estatisticas2.pdf. 1
- [6] I. I. N. de Estudos e Pesquisas Educacionais Anísio Teixeira. Censo da educação superior - apresentação para a imprensa, 2017. disponível em : http://download.inep.gov.br/educacao_superior/censo_superior/documentos/2018/apresentacao_censo_superior2017F.pdf. 1
- [7] K. M. Elliott and M. A. Healy. Key factors influencing student satisfaction related to recruitment and retention. *Journal of marketing for higher education*, 10(4):1–11, 2001. 1
- [8] A. P. Fernandes Neto. Modelagem de equações estruturais na análise de dados de serviços em comunicações móveis. 2013. 2
- [9] J. B. Ford, M. Joseph, and B. Joseph. Importance-performance analysis as a strategic tool for service marketers: the case of service quality perceptions of business students in new zealand and the usa. *Journal of Services marketing*, 13(2):171–186, 1999. 1
- [10] L. G. Giordani. Um modelo de equações estruturais aplicado a dados de satisfação de alunos do ensino superior privado. 2015. 1
- [11] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson. *Multivariate data analysis: Pearson new international edition*. Essex: Pearson Education Limited, 2014. 1, 2
- [12] H. F. A. Hasan, A. Ilias, R. A. Rahman, and M. Z. A. Razak. Service quality and student satisfaction: A case study at private higher education institutions. *International Business Research*, 1(3):163–175, 2008. 1
- [13] F. M. Hill. Managing service quality in higher education: the role of the student as primary consumer. *Quality assurance in education*, 3(3):10–21, 1995. 1
- [14] A. Jiewanto, C. Laurens, and L. Nelloh. Influence of service quality, university image, and student satisfaction toward wom intention: A case study on universitas pelita harapan surabaya. *Procedia-Social and Behavioral Sciences*, 40:16–23, 2012. 1
- [15] S. D. Johnson, S. R. Aragon, and N. Shaik. Comparative analysis of learner satisfaction and learning outcomes in online and face-to-face learning environments. *Journal of interactive learning research*, 11(1):29–49, 2000. 1
- [16] Y. J. Joo, K. Y. Lim, and E. K. Kim. Online university students’ satisfaction and persistence: Examining perceived level of presence, usefulness and ease of use as predictors in a structural model. *Computers & education*, 57(2):1654–1664, 2011. 1
- [17] C. C. Lo. How student satisfaction factors affect perceived learning. *Journal of the Scholarship of Teaching and Learning*, pages 47–54, 2010. 1
- [18] N. K. MALHOTRA. *Pesquisa de marketing: foco na decisão*. São Paulo: Pearson Prentice Hall, 2011. 2
- [19] J. Marôco. *Análise de equações estruturais: Fundamentos teóricos, software & aplicações*. ReportNumber, Lda, 2010. 2
- [20] M. Marzo Navarro, M. Pedraja Iglesias, and P. Rivera Torres. A new management element for universities: satisfaction with the offered courses. *International Journal of educational management*, 19(6):505–526, 2005. 1
- [21] S. d. S. Pereira. Modelagem de equações estruturais no software r. 2013. 2
- [22] A. L. R. Schleich, S. A. J. Polydoro, and A. A. A. d. Santos. Escala de satisfação com a experiência acadêmica de estudantes do ensino superior. *Avaliação psicológica*, 5(1):11–20, 2006. 1
- [23] K. M. Vieira, F. T. Milach, and D. Huppés. Equações estruturais aplicadas à satisfação dos alunos: um estudo no curso de ciências contábeis da universidade federal de santa maria. *Revista Contabilidade & Finanças*, 19(48):65–76, 2008. 2
- [24] R. Warren, R. F. Smith, and A. K. Cybenko. Use of mahalanobis distance for detecting outliers and outlier clusters in markedly non-normal data: a vehicular traffic example. Technical report, SRA INTERNATIONAL INC DAYTON OH, 2011. 2
- [25] J. Wiers-Jenssen, B. r. Stensaker, and J. B. Gr gaard. Student satisfaction: Towards an empirical deconstruction of the concept. *Quality in higher education*, 8(2):183–195, 2002. 1
- [26] E. J. Wolf, K. M. Harrington, S. L. Clark, and M. W. Miller. Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and psychological measurement*, 73(6):913–934, 2013. 2