



Eixo temático: Gestão e Políticas Educacionais no contexto das TDIC

## ENGENHARIA DE SOFTWARE E ESTATÍSTICA NO APOIO À EDUCAÇÃO A DISTÂNCIA

SOFTWARE ENGINEERING AND STATISTICS IN SUPPORTING DISTANCE EDUCATION

- **Eduardo Faria** (Universidade Federal de Sergipe – [faria.eduardo@academico.ufs.br](mailto:faria.eduardo@academico.ufs.br))

### Resumo:

*O presente trabalho retrata a importância da regressão múltipla, microsserviços e dados abertos para decisões em políticas públicas na área de educação a distância. Desse modo, propõe-se a implementação dessas metodologias em um cenário mais próximo à realidade. O objetivo é demonstrar que o estudo preliminar das características peculiares de discentes, dados populacionais e outras variáveis influenciam a maneira com que a ministração do conteúdo é mais assertiva. Lança-me mão de dados de áreas como educação, população e regionalidade, abordando-os através de uma visão probabilística, expondo casos manuais para em seguida automatizá-los, particularmente com o uso da tecnologia Spring Boot. A fundamentação para os procedimentos em voga é um arcabouço sólido proveniente da doutrina acadêmica, da legislação publicística e do manifesto ágil, que reforça o uso de abordagens não monolíticas. Os resultados do trabalho refletem a coexistência de diferentes cenários que requerem distintas abordagens para a facilitação do ensino.*

**Palavras-chave:** Software. Engenharia. Estatística. Distância. Educação

### Abstract:

*This work portrays the importance of multiple regression, microservices and open data for public policy decisions in the area of distance education. Therefore, it is proposed to implement these methodologies in a scenario closer to reality. The objective is to demonstrate that the preliminary study of the peculiar characteristics of students, population data and other variables influence the way in which content delivery is more assertive. Some data from areas such as education, population and regionality are used, approaching them through a probabilistic view, exposing manual cases and then automating them, particularly with the use of Spring Boot technology. The basis for the current procedures is a solid framework coming from academic doctrine, publicity legislation and the agile manifesto, which reinforces the use of non-monolithic approaches. The results of the work reflect the coexistence of different scenarios that require different approaches to facilitating teaching.*

**Keywords:** Software. Engineering. Statistics. Distance. Education.

## 1. Introdução

A educação a distância há muito passou de mera promessa para uma realidade concreta. Na década anterior, por exemplo, o contexto educacional em tecnologia desse gênero ainda era incipiente: as universidades e faculdades autorizadas pelo Ministério da Educação eram escassas e os valores dos cursos ofertados menos acessíveis que os de agora.

Não se tinha, outrora, uma solidez e segurança como hoje, a começar pelo eixo regulamentar: não existia nem mesmo uma lei de privacidade de dados efetiva. Então, como mudar tal realidade?



### 1.1. Apresentação do tema

Em 2005, o chefe do executivo, que tem a incumbência de gerir programas governamentais pela própria Constituição Federal, editou o Decreto de nº 5.622, de 19 de dezembro de 2005<sup>1</sup>, que regulamentou o ensino em ambientes virtuais em seu capítulo II e que viera a ser reformulado apenas em 2017.

Entre 2011 e 2021, o crescimento do número de ingressantes em cursos superiores na modalidade educação a distância (EAD) aumento 474% (quatrocentos e setenta e quatro por cento), de acordo com dados do Censo da Educação Superior 2021<sup>2</sup>. As tabelas inframencionadas refletem as 15 universidades com mais alunos ingressantes, entre 2013 e 2021, nas esferas privada e pública (federal, estadual e municipal):

Tabela 1. Número de egressos na graduação EAD por centro de ensino em ordem decrescente

Colocação	Nome	Novos Discentes (2013 a 2021)	Tipo	UF
1	Universidade Nove de Julho	378	Privada	SP
2	Universidade do Sul de Santa Catarina	261	Privada	SC
3	Centro Universitário Leonardo da Vinci	261	Privada	SC
4	Universidade Estácio de Sá	243	Privada	RJ
5	Centro Universitário Claretiano	243	Privada	SP
6	Universidade de Franca	243	Pública	SP
7	Universidade Anhembi Morumbi	216	Privada	SP
8	Universidade Metropolitana de Santos	198	Privada	SP
9	Universidade Pitágoras Unopar Anhanguera	171	Privada	PR
10	Universidade Luterana do Brasil	162	Privada	RS
11	Universidade de Uberaba	162	Privada	MG
12	Universidade Católica de Brasília	153	Privada	DF
13	Universidade Cesumar	144	Privada	PR
14	Centro Universitário Internacional	144	Privada	PR
15	Universidade Paulista	135	Privada	SP

Fonte: INEP<sup>3</sup> - Dados de Indicadores de fluxo da educação superior (2013 a 2021)

Diante do exposto, percebe-se claramente a superioridade em números da iniciativa privada em face das universidades públicas, as quais contam apenas com uma universidade no top 15 e ainda de cunho municipal. Ademais, não há presença de faculdades com sede na região Norte / Nordeste, o que enaltece a deficiência em políticas públicas em educação a distância.

Nesse diapasão, é premente a necessidade de busca por implementações de políticas mais assertivas e congruentes e duas grandes áreas que podem auxiliar os gestores na condução desse processo são a estatística e a engenharia de software.

Portanto, discutir-se-á a importância das variáveis em educação para fundamentar modelos de regressão múltipla que contém as principais variáveis influenciadoras do ensino em plataformas virtuais, suportado pelo fenômeno da agilidade e tecnologias de renome para o consumo de dados abertos, que darão suporte ao agente no contexto da engenharia de software.

### 1.2. Contextualização da importância da engenharia de software e estatística na educação a distância



Através de um grupo representativo de uma população, é possível tecer inferências sobre aquela. Tal premissa é sustentada pela teoria da probabilidade, utilizada quotidianamente no universo estatístico e também pela distribuição F de Snedecor<sup>4</sup>, que corrobora o modelo da equação de mínimos quadrados.

Desse modo, combinado com o uso do modelo de mínimos quadrados, o responsável pela implementação de políticas públicas pode estudar com pormenores o comportamento de diversas dimensões, que podem influenciar os ingressantes dos ambientes virtuais: PIB per capita, quantidade de universidades públicas presentes na região do aluno e população da região.

Izabelly e Aline (2017) defendem que a metodologia de análise de software necessita de procedimentos específicos para que os detalhes sejam postos na documentação<sup>5</sup>. Isso é crucial porque, em uma abordagem que não leva em conta a documentação dos achados de forma esmerada, não há o que se falar em êxito no workflow eleito.

Dessa maneira, as tecnologias de microserviços vêm substituindo progressivamente as velhas aplicações monolíticas e uma dessas implementações é o framework Spring, que guarda uma estreita correlação com o manifesto ágil, revelando-se promissor para a elaboração de projetos nesse nicho específico, o que dará ao gestor público uma higidez robusta para o desencadear de conquistas no ensino virtual.

### 1.3. Declaração da relevância do uso de dados abertos e tecnologias ágeis na análise educacional

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), autarquia federal, possui um papel de destaque nas políticas públicas de educação e concede à sociedade dados riquíssimos sobre educação básica e superior na modalidade de dados abertos, como: indicadores financeiros educacionais, média de alunos por turma, média de horas-aula diária, nível socioeconômico, percentual de docentes com pós-graduação stricto sensu, dentre muitos outros.

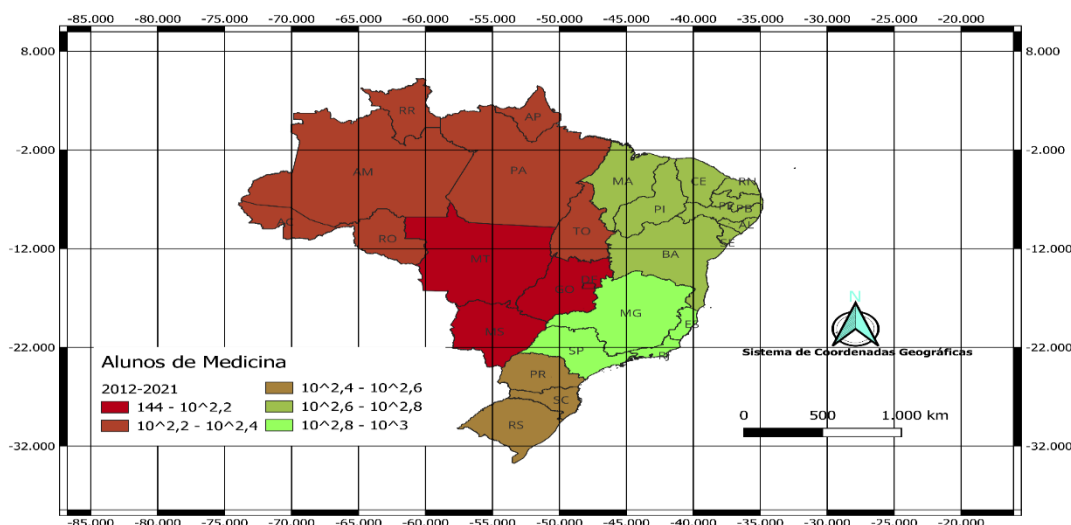


Figura 1. Ingressos de medicina na modalidade presencial de 2013 a 2021.

Fonte: Dados do Inep (fluxo da educação), com mapa de elaboração própria com o software QGIS e malha do IBGE.

## 2. A contextualização do ambiente de ensino virtual



Hoje se tem uma maior facilidade de uso da web do que em décadas passadas, devido ao barateamento de planos de internet com o advento da fibra ótica e a consolidação da tecnologia 5G, somado, obviamente, ao custo de equipamentos informáticos em declínio e à capacidade ascendente tecnológica, como computadores pessoais, tablets e smartphones com grandes capacidades de processamento, aliado ao conceito de internet das coisas, com vários dispositivos interconectados, muitos com inteligência artificial embutida.

Se tudo isso está à disposição também do gestor público, que se vê na obrigação de fomentar ainda mais a qualidade do ensino, por que os dados dos programas realizados não estão tão expressivos? Isso precisa ser averiguado mais de cerca, pois não é convincente que um país vizinho como a Colômbia tenha 30% da população com ensino superior, em detrimento dos 21% dos diplomados no Brasil<sup>6</sup>.

### **2.1. Definição de tecnologias digitais de informação e comunicação e características**

O conceito de tecnologias digitais de informação e comunicação pode se enquadrar como quaisquer facilidades postas à disposição do usuário para transmissão de informação. Na seara doutrinária, Mariana (2018) afirma também que o docente não é mais o detentor único do conhecimento, e o discente passa a ser o protagonista do saber<sup>7</sup>.

Ou seja, já vigora o empoderamento dos discentes nessa nova era tecnológica, pois cada vez mais os alunos contam com sólidas ferramentas para confrontar os ensinamentos acadêmicos, refutando-os ou ratificando-os.

### **2.2. Necessidade de compreensão da nova realidade**

Essa nova realidade deve ser entendida pelo gestor público, que deve saber como explorar esse cenário para implementar políticas mais persuasivas, usando dos recursos à sua disposição para trazer mais alunos à escola.

Para evidenciar a necessidade de mudança de foco, um levantamento de uma famosa empresa de segurança da informação mostrou que os internautas ficam mais de quatro décadas na internet, ou seja, quarenta anos<sup>8</sup>. Então, por que não investir na colonização desse espaço tecnológico com políticas públicas de inclusão de mais jovens em universidades, de menos evasão acadêmica?

### **2.3. As políticas educacionais e sua adaptabilidade em regiões diversas**

Em consideração a isso, a urgência de adaptação a diferentes cenários é gritante e, portanto, um administrador sensato precisará ter um sólido conhecimento de dados prévios e noções da particularidade de casos que se mostram em dissimilitude de região para região.

Sim, o primeiro passo é factível e está facilmente disponível via dados abertos, uma vez que os gestores públicos em suas pastas respectivas podem usufruir da facilidade de obtenção de informações em matéria de educação e dados econômico-financeiros que regem a região sob análise, o que pode direcionar a resultados exitosos na alavancagem do número de pessoas com ensino superior, via um desenvolvimento sólido da educação a distância.

Nesse toar, se o Brasil, em um dado espaço temporal, apresentasse déficit no número de professores da educação básica por motivos supervenientes, como aposentadoria em massa do





corpo docente, seria congruente o investimento em uma política de cursos gratuitos na área de licenciatura nas regiões Norte e Centro-Oeste, pois, do total de cursos de licenciatura no formato EAD, entre 2013 a 2021, o Sudeste possui dezesseis entidades educacionais no top 30 de novos ingressantes, contra cinco do Nordeste e Sul e apenas duas das regiões Norte e Centro-Oeste, como mostra o gráfico a seguir:

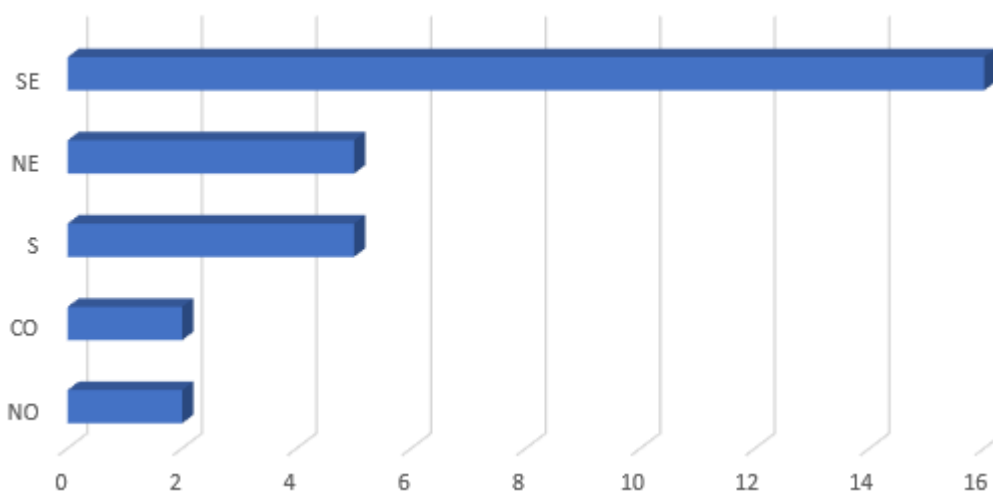


Figura 2. Regiões-sede presentes no top 30 das faculdades com maior número de ingressantes em cursos de licenciatura EAD entre 2013 a 2021.

Fonte: Dados do Inep (fluxo de educação do ensino superior). Elaboração Própria

A título de ilustração, as universidades com maiores quantidades de ingressantes nos cursos de licenciatura EAD por região foram: na região Sudeste, a Universidade de Franca, de São Paulo, com 135 ingressantes; no Sul, o Centro Universitário Leonardo da Vinci, de Santa Catarina, com 90 ingressantes; no Nordeste, a Universidade Federal da Paraíba, com 72 ingressantes; no Norte, a Universidade Federal do Pará, com ingressantes e no Centro-Oeste, a Universidade Federal de Mato Grosso do Sul, também com 45 ingressantes, de acordo com a mesma base utilizada.

### 3. Vantagens e desafios da utilização de dados abertos como suporte à decisão

A lei do governo digital<sup>9</sup> define dados abertos como dados acessíveis ao público, representados em meio digital, estruturados em formato aberto, processáveis por máquina, referenciados na internet e disponibilizados sob licença aberta que permita sua livre utilização, consumo ou tratamento por qualquer pessoa, física ou jurídica.

Depreendem-se da norma alguns conceitos que podem ser facilmente replicáveis: a facilidade de uso, já que pode ser acessado por qualquer indivíduo, independentemente do grau acadêmico ou situação social e a facilidade de leitura, ou de consumo, uma vez que o formato deve ser neutro e processável por tecnologias de processamento de máquina, não especificando o diploma legal a tecnologia a que se refere, do que se infere a generalidade de consumo da informação.

#### 3.1. A qualidade dos dados abertos em educação



No ano de 2007, trinta especialistas em dados governamentais se reuniram na Califórnia e escreveram os principais princípios de dados abertos governamentais<sup>10</sup>. Desde então, tais critérios têm sido utilizados na análise da qualidade de dados abertos em vários países, inclusive no Brasil. Trata-se, na verdade, de oito princípios básicos e sete adicionais, que norteiam as diretrizes da publicização desses dados. A imagem a seguir, retirada de uma análise de dados abertos na Indonésia<sup>11</sup>, elenca resumidamente tais princípios:

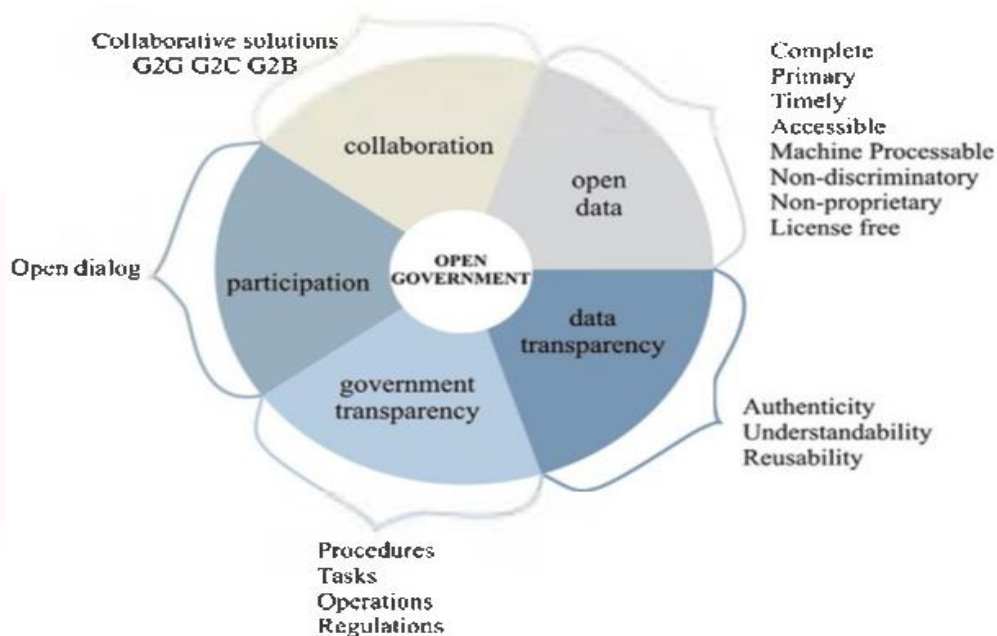


Figura 3. Resumo dos oito princípios de dados abertos na seção “open data”, com atributos adicionais.

Fonte: Semantics Scholar

Diante disso, utilizou-se o relatório completo do Sistema de Seleção Unificada (Sisu) do segundo semestre de 2022<sup>12</sup>, como referência para o input dos dados abertos governamentais a fim de confrontá-los com os oito princípios analisados. Em primeiro lugar, de acordo com o princípio da completude, todos os dados devem estar disponíveis. Assim, a planilha original retornou 674.805 (seiscentos e setenta e quatro mil, oitocentos e cinco) registros, que de fato é o total de inscritos no programa naquele ano, o que atende ao primeiro princípio básico, o da completude.

De igual modo, por estar em formato CSV, a planilha é processável por máquina, seja em planilhas eletrônicas, como Excel ou LibreOffice ou até via ambientes de análise de dados como R, Python, Java (Weka) e outros. Constatou-se também a acessibilidade dos dados, pois são bem documentados e possuem legendas ilustrativas, não há licenças, não há discriminação e não é proprietário. Desse modo, com o atendimento aos princípios de dados abertos, o gestor acha facilmente a resposta para as questões formuladas:

- O total de inscritos em graduação EAD é apenas de 5295, contra 669.510 dos cursos presenciais, o que dá a efêmera percentagem de 0,78% de inscritos em cursos EAD.
- Apenas quatro universidades ofertaram cursos EAD, sendo, em ordem decrescente: UFMS, com 2199 inscritos; UFCA, com 2013 inscritos; UFPI, com 981 inscritos e UFVJM, com 102 inscritos EAD no Sisu.



- Como essas quatro faculdades se localizam em Mato Grosso do Sul, Ceará, Piauí e Minas Gerais, respectivamente, a quase totalidade dos outros estados da federação não tiveram universidades federais que ofertaram cursos EAD no Sisu.
- A média de nota de corte exigida foi de 597,15 entre as quatro universidades presentes (8 cursos no total).
- A taxa de reprovação foi muito elevada, pois apenas 802 candidatos foram aprovados dentre o total de 5295 inscritos, chegando a 84,85% de reprovação.

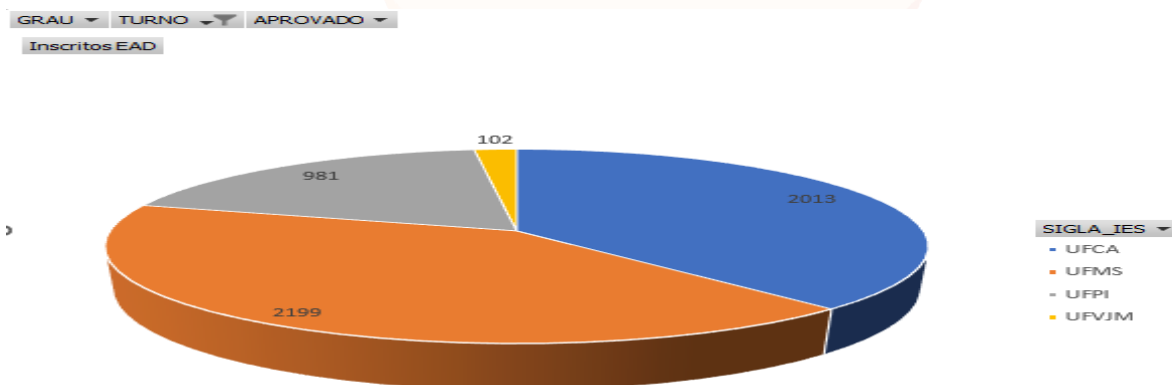


Figura 4. Únicas universidades federais ofertantes de cursos de graduação EAD no Sisu 2022.  
Fonte: Relatório Sisu – Chamada Regular 2/2022. Elaboração Própria

### 3.2. Achados identificados que necessitam de intervenção

Muitas vezes, os dados educacionais no Brasil são coletados de diferentes fontes e por diferentes entidades, resultando em falta de padronização e uniformidade. A ausência de normas comuns prejudica a integração, análise e comparabilidade dos dados, dificultando a formulação de políticas eficazes.

Assim, erros, inconsistências e falta de atualização são problemas comuns encontrados nos conjuntos de dados educacionais. Essa baixa qualidade compromete a confiabilidade das informações, afetando a tomada de decisões e a avaliação precisa do sistema educacional. Além disso, a falta de usabilidade nos dados educacionais impede que pesquisadores, educadores e formuladores de políticas aproveitem todo o potencial dessas informações e a transparência também é afetada, já que muitos conjuntos de dados não estão prontamente disponíveis para o público em formatos acessíveis e compreensíveis.

### 3.3. Uma análise de integração dos dados educacionais

A utilização de linguagens de programação como Python, R, Java, Lua ou SQL para padronizar e normalizar os dados provenientes de diversas fontes pode ser facilmente feita em arquivos com formato aberto como CSV. Isso envolve a limpeza dos dados, correção de erros e inconsistências e a organização em formatos comuns para facilitar a análise e compreensão.

Ao integrar dados educacionais, é crucial garantir a segurança e a proteção da privacidade dos dados, utilizando técnicas de criptografia, políticas de acesso restrito e práticas de segurança da informação. Isso pode proporcionar uma visão mais abrangente e detalhada do sistema



educacional, facilitando a análise, a tomada de decisões e o desenvolvimento de políticas<sup>13</sup> mais eficazes para a melhoria da educação.

#### 4. Técnicas de estatística na análise educacional

Ao empregar abordagens estatísticas que exploram a relação simultânea de várias variáveis independentes com uma variável dependente de interesse, busca-se não apenas entender o impacto individual de cada fator, mas também examinar como a combinação dessas variáveis pode afetar o desempenho educacional, auxiliando na identificação de potenciais áreas para intervenções e melhorias no sistema educativo.

##### 4.1. Regressão múltipla e utilização para compreensão de relacionamentos

A regressão múltipla é uma técnica poderosa utilizada na análise de relações complexas entre variáveis em estudos educacionais. Ela permite entender como várias variáveis independentes influenciam uma variável dependente, possibilitando a modelagem e previsão de fenômenos educacionais.

Por exemplo, ao investigar o desempenho acadêmico de estudantes, podemos empregar o método dos mínimos quadrados para encontrar o modelo que melhor se ajusta aos dados. Isso permite identificar a contribuição relativa de cada variável independente no resultado educacional, ajudando a entender o impacto individual de cada fator e suas interações.

##### 4.2. Exemplos práticos da aplicação dos conceitos estatísticos na educação a distância

A prática basilar, em resumo, consiste na análise do impacto do PIB regional<sup>14</sup>, da quantidade de universidades<sup>15</sup> com sedes na região e da população<sup>16</sup> de uma determinada localidade no desempenho educacional da política pública para medir a qualidade da educação a distância, ou seja, três variáveis para ao final estimar a porcentagem final de pessoas com nível superior<sup>17</sup>.

Assim, como a equação se baseia em uma amostra para prever cenários externos ao modelo, convém selecionar uma quantidade de estados, que pode ser de cinco unidades federativas, a depender do nível de precisão que o gestor público almeja atingir com o modelo.

Para os dados em voga, pertencentes a São Paulo, Rio Grande do Sul, Pernambuco, Pará e Distrito Federal, contemplando-se todas as regiões do Brasil, têm-se as seguintes informações:

Tabela 2. Dados representativos das três variáveis da regressão múltipla

Estado	Porcentagem de graduados (em %)	PIB per capita (em milhares)	Quantidade de Universidades	População (em milhões)
		(X)	(Y)	(Z)
São Paulo	25	58,3	882	44,4
Rio Grande do Sul	19,6	50,7	165	10,9
Pernambuco	12,9	22,8	173	9,06
Pará	13	30	92	8,12
Brasília	37	92,7	128	2,82

Fonte: IBGE e INEP - Tabela de elaboração própria





Gujarati (2017) enaltece a vantagem do cálculo da equação da reta de uma regressão múltipla pela álgebra matricial<sup>18</sup> e nesse raciocínio os coeficientes do modelo da equação se dão pela seguinte expressão:  $(X'X)^{-1} * (X'y)$ , onde são separadas as matrizes com os valores independentes (PIB per capita, universidades e população) e a matriz com a variável dependente (percentagem de graduados), calculando-se suas transpostas, quando for o caso, de acordo com o exemplo a seguir:

- A matriz das variáveis dependentes (X) compõe-se cinco linhas e quatro colunas, com valores de:  $[[1, 58.3, 882, 44.4], [1, 50.7, 165, 10.9], [1, 22.8, 173, 9.06], [1, 30, 92, 8.12], [1, 92.7, 128, 2.82]]$ .
- Em seguida, calcula-se a matriz transposta das variáveis (X'), invertendo as linhas e colunas:  $[[1, 1, 1, 1, 1], [58.3, 50.7, 22.8, 30, 92.7], [882, 165, 173, 92, 128], [44.4, 10.9, 9.06, 8.12, 2.82]]$ .
- O próximo passo é calcular o produto da transposta pela original das variáveis dependentes (X'X), a qual terá quatro linhas e quatro colunas:  $[[5, 254.5, 1440, 75.3], [254.5, 15982.51, 78356.1, 3852.73], [1440, 78356.1, 859926, 43634.68], [75.3, 3852.73, 43634.68, 2246.14]]$ .
- Calcula-se a matriz inversa da anterior (X'X<sup>-1</sup>):  $[[2.817, -0.044, 0.017, -0.367], [-0.044, 0.0008, -0.00032, 0.006248], [0.0179, -0.00032, 0.000206, -0.00406], [-0.3673, 0.006248, -0.00406, 0.08088]]$ .
- Em seguida, calcula-se o produto da matriz transposta das variáveis independentes (X') pela matriz das variáveis independente (y), que nada mais é do que os valores de y ordenados em uma única coluna com cinco linhas. O resultado é uma matriz de quatro linhas e única coluna:  $[[107.5], [6565.24], [33447.7], [1650.414]]$ .
- Por derradeiro, o produto da matriz inversa com o passo anterior formará a matriz final, com quatro linhas e uma coluna, onde o primeiro elemento é o intercepto e os seguintes representarão os valores do PIB, universidades e população:  $[[6.080829], [0.302978], [0.037606], [-0.71933]]$ .

$$F(x,y,z) = 6.081 + 0.303x + 0.038y - 0.720z$$

Onde,

X = PIB per capita da região

Y = Número de universidades na região

Z = População da região

Figura 4: Equação de regressão múltipla para previsão do índice de escolaridade superior da população em função das variáveis PIB per capita, universidades e população

Fonte: Elaboração própria

Ora, academicamente, o modelo proposto explica com 99,98% de certeza que o índice de educação superior é explicado por essas três variáveis, pois essa não é nada menos do que a informação fornecida pelo coeficiente R<sup>2</sup>, demonstrando a eficácia do modelo<sup>19</sup>.

A título de exemplo, a percentagem de graduados em SP é dada por:  $6.081 + 0.303 * 58.3 + 0.038 * 882 - 0.720 * 44.4$ , o que equivale a 24.9752 (quase 25% do dado original), enquanto que

para o estado do Pará o cálculo é descrito por:  $6.081 + 0.303 * 92.7 + 0.038 * 128 - 0.720 * 2.82$ , com resultado de 12.789 (quase 13 do dado original), com erros insignificantes. Matematicamente, o modelo é eficiente com base na tabela ANOVA a seguir, a qual revela que a estatística da distribuição F de Snedecor pertence profundamente à região crítica com três graus de liberdade no numerador e um no denominador<sup>20</sup>:

Tabela 3. Tabela ANOVA da regressão múltipla da política de dados educacionais

Fonte de Variação	Graus de Liberdade	Soma dos Quadrados	Quadrado Médio	Estatística F de Snedecor
Regressão	3	402.2240369	134.074679	1397.148511
Resíduos	1	0.095963083	0.095963083	
Total	4	402.32		

Fonte: Elaboração própria

## 5. Um ambiente de microsserviços como indutor de políticas de educação a distância

Na engenharia de software, a tecnologia de microsserviços visa a suprir as lacunas encontradas nas aplicações monolíticas tradicionais. Nas décadas passadas, até então, toda a lógica de negócio era disposta em um único local, agregando-se banco de dados, interface gráfica, procedimentos de validação e controle na mesma aplicação.

Tal conjuntura, embora trouxesse mais facilidade à rotina dos engenheiros de software, trazia consigo uma maior dificuldade na manutenção do código, na dependência enorme entre diferentes partes do sistema, em uma maior exposição a informações sensíveis presentes nos bancos de dados e em uma escalabilidade deficiente, uma vez que as modificações nos sistemas exigiam esforços cada vez maiores<sup>22</sup>.

Nesse diapasão, o desenvolvimento de software, de acordo com o padrão de microsserviços, altamente condizente com os ditames da agilidade, permite uma maior integração entre sistemas<sup>23</sup>, pois cada funcionalidade do software pode ser desenvolvida de maneira independente, com diferentes tecnologias, linguagens de programação, bancos de dados e lógicas próprias.

### 5.1. Descrição do framework utilizado e seu papel na análise de dados educacionais

Um grande exemplo desse modelo é o framework Spring Boot, desenvolvido pela empresa americana VMware, escrito em Java e com licença de código aberto. Dentre as principais características do ambiente, destacam-se:

- Configuração automática de várias partes do aplicativo.
- Servidor embutido que elimina a necessidade de outro externo.
- Injeção de dependência, que facilita a manutenção do código.
- Pré-configuração de módulos, que simplifica o processo de configuração.
- Ferramentas de métricas para monitoramento do aplicativo em tempo real.
- Facilitador de testes unitários para maior precisão de funcionalidades específicas.

Com as características que o Spring Boot possui, é exequível o seu uso como um auxiliador no emprego da gestão de políticas educacionais em tecnologias de educação a distância e, com tal afã, propõe-se a sua utilização do seguinte modo: a criação de uma URL que seja capaz de afirmar se um estado da federação está suficientemente atendido pelas políticas públicas atuais, de acordo com a meta estabelecida.



De modo complementar, como o Spring possui a arquitetura de microsserviços<sup>24</sup>, é necessário implementar as responsabilidades de maneira apartada com o estilo de arquitetura MVC, ou seja, deve haver uma camada para lidar com entidades e persistência de dados, outra para informar apenas entradas e saídas e a derradeira para conter toda a lógica de programação.

De maneira sucinta, a engenharia de software para o caso em tela propõe que o usuário da aplicação (gestor público) informe os dados específicos de uma localização (região, estado ou município) ao endereço URL em que está funcionando o servidor Spring Boot, em uma porta qualquer, hospedada em um determinado endereço governamental, por intermédio de uma requisição HTTP com o método GET, facilmente disposto em uma documentação.

Desse modo, tal ferramenta permite uma padronização na descrição das APIs, facilitando a compreensão e o uso por desenvolvedores que desejam interagir com esses serviços.

Name	Description
<b>pib-cap</b> * required number (query)	PIB per capita (em milhares) <input type="text" value="pib-cap"/>
<b>universidades</b> * required integer (query)	Número de universidades na região <input type="text" value="universidades"/>
<b>populacao</b> * required number (query)	População (em milhões) <input type="text" value="populacao"/>
<b>meta</b> * required number (query)	Índice de Graduados desejado (em %) <input type="text" value="meta"/>

Figura 5: Documentação Swagger para a URL do serviço em Spring Boot proposto.

Fonte: Elaboração própria

O funcionamento do sistema à disposição do gestor público é fácil e intuitivo. Após o fornecimento dos parâmetros informados na requisição GET, o servidor Spring Boot buscará prever a percentagem de graduados na região em estudo tendo como base a equação de regressão múltipla explorada no tópico 4.3 e comparará o valor da percentagem prevista com a meta da política deseja pelo gestor quando do início da requisição.

Como exemplo, foram fornecidos dados do estado do Amazonas, que não está incluso na tabela nos cinco estados que serviram como base da regressão. Com PIB per capita de cerca de R\$ 30,8 mil, contando com 37 universidades e com população estimada em 3,94 milhões, o índice calculado foi de 13,97%, contra 16,6% do índice real calculado por IBGE, o que demonstra a capacidade do modelo para fins de generalização.

## 5.2. Importância e potencial da arquitetura descentralizada na melhoria da educação a distância

Tecidas essas considerações, a base do código-fonte da aplicação pode ser analisada a partir do controlador principal, o qual traz apenas a invocação de uma outra função especializada e declarada em outro serviço, visando a respeitar o princípio da responsabilidade única de cada classe.

```
EadPolitica.java x RegressaoMultiplaEAD.java x
1
2
3
4 import org.springframework.web.bind.annotation.GetMapping;
5 import org.springframework.web.bind.annotation.RequestMapping;
6 import org.springframework.web.bind.annotation.RequestParam;
7 import org.springframework.web.bind.annotation.RestController;
8
9 @RestController
10 @RequestMapping("/politica-ead")
11 public record EadPolitica(RegressaoMultiplaEAD regressao) {
12     @GetMapping("/analise")
13     public Boolean obtencaoDePessoasComNivelSuperior
14         (
15         @RequestParam("pib-cap") Double pibPerCapita,
16         @RequestParam("universidades") Integer universidades,
17         @RequestParam("populacao") Double populacao,
18         @RequestParam("meta") Double meta
19         ) {
20         return regressao.calcularReta();
21     }
22 }
```

Figura 6: Especificação dos parâmetros, do endereço e do método da requisição. A classe usa a lógica implementada em outra classe, separando-se, assim, a visão da lógica do serviço  
Fonte: Elaboração própria

## 5.3. Exemplificação dos procedimentos realizados

Diante disso, convém sumarizar o procedimento adotado em dois cenários distintos, um mais trivial e outro mais complexo, com vistas a atender ao interesse do gestor público no cálculo da taxa de graduados em nível superior, com vistas a fomentar a oferta de cursos EAD para a alavancagem desse índice.

O primeiro, mais trivial, porém eficiente, consiste apenas em receber os dados das variáveis na URL do Spring Boot e carregar esses valores na equação da regressão múltipla calculada anteriormente em uma análise estatística pregressa, a qual possui três variáveis. Assim, basta fornecer os dados do PIB per capita, do número de universidades e da população local, nos moldes estabelecidos, calculando-se a partir de então o índice de educação superior.

O segundo, mais moroso, porém eficaz, exige que sejam consumidos diretamente os arquivos CSV do Ministério da Educação e do IBGE, a fim de gerar unicamente dados mais atualizados, protegendo-os contra a obsolescência, pois uma vez que novas universidades são





criadas, a população aumenta e o PIB per capita sofre variação, a equação tende a mudar, mas como é calculada dinamicamente, não há necessidade de uma análise estatística separada, pois esta é feita na hora da consulta, de acordo com esse tipo de análise.

## 6. Considerações finais

É indubitável a contribuição da estatística e da engenharia de software no apoio à formulação de políticas públicas em educação a distância, pois tópicos preciosos advindos dessas grandes áreas do saber, como regressão múltipla, construção da tabela ANOVA, arquitetura de microsserviços, métodos de planejamento ágeis e frameworks de programação, como o Spring Boot, podem juntos munir o gestor público de informações cruciais para o desenvolvimento do ensino à distância no país, com vistas a aumentar a quantidade de pessoas com nível superior, uma vez que tal percentagem é ainda muito reduzida, até mesmo em comparação com vizinhos latino-americanos.

Uma consideração importante é que o aumento da população é inversamente proporcional às pessoas com algum diploma, pois, no parâmetro da regressão, tal variável possui coeficiente negativo; ou seja, não basta começar a criar políticas públicas para que a taxa de graduação entre a população cresça, é necessário que esse crescimento evolua a uma taxa superior à da população entre distintas regiões, de modo a evitar a estagnação do nível educacional brasileiro.

O padrão de microsserviços sugerido, de igual modo, não deve ser utilizado como uma solução única, mas combinado com outras tecnologias de integração como banco de dados não-relacionais, containerização com Docker e suas mais variáveis formas de orquestração, assim como conceitos da filosofia DevOps para incrementar a celeridade na entrega das demandas, pois não é fácil lidar com a burocratização da gestão pública, ainda na era digital.

### 6.1. Conclusões e sugestões de futuras pesquisas e avanços nessa área

Em suma, conclui-se que a educação a distância evoluiu a passos largos e esse avanço se vincula fortemente ao desenvolvimento da legislação na área digital, com o advento de leis e regulamentos importantes, como a regulamentação do ensino à distância, o marco civil da internet, a lei geral de proteção de dados pessoais, a lei do governo digital, dentre tantas outras.

Somado a isso, o crescimento e barateamento de tecnologias de transmissão de dados como a fibra ótica e o aparecimento de internet de dados mais ágeis como o 4G e o 5G, aliado ao próprio progresso tecnológico como celulares e computadores pessoais colaborou para a procura de cursos EAD. Aliás, a própria pandemia da COVID-19 forçou as pessoas a buscarem tal modalidade de oferta de curso, os quais também passaram por extensa modernização.

Apesar disso, ainda há enormes desafios, como a baixa taxa de pessoas com ensino superior na população brasileira, a expressividade em números dos matriculados em cursos EAD em instituições privadas e a pouca oferta dessa modalidade de curso nas instituições federais, as quais, muitas vezes, limitam-se a distribuir cursos EAD em áreas restritas do conhecimento.

É necessário, assim, maior ativismo nos cursos dessa modalidade nas universidades mantidas pelo governo, uma queda mais acentuada nos valores dos cursos ofertados na esfera privada e pesquisas mais assertivas sobre preferências individuais de populações locais, com o mapeamento de preferências de estudos, de modo a realizar ações de marketing educacional endereçadas a essas pessoas que quiçá desconheçam o poder transformacional que a educação a distância representa no país.





## 7. Referências

1 BRASIL. Decreto nº 5.622, de 19 de dezembro de 2005. Regulamenta o art. 80 da Lei nº 9.394, de 20 de dezembro de 1996, que estabelece as diretrizes e bases da educação nacional. **Diário Oficial da União**: seção 1, Brasília, DF, ano 142, n. 144, p. 59-64, 20 dez. 2005.

2 ASSESSORIA DE COMUNICAÇÃO SOCIAL DO INEP. **Ensino a distância cresce 474% em uma década**. Brasília: INEP, 2022. Disponível em: <https://www.gov.br/inep/pt-br/assuntos/noticias/censo-da-educacao-superior/ensino-a-distancia-cresce-474-em-uma-decada>. Acesso em: 5 jan. 2024.

3 INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Indicadores de Fluxo da Educação Superior**. Brasília: INEP, 2023. Disponível em: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/indicadores-educacionais/indicadores-de-fluxo-da-educacao-superior>. Acesso em: 5 jan. 2024. Base de dados.

4 MORETTIN, P.A.; BUSSAB, W.O. **Estatística Básica**. 9. Ed. São Paulo: Saraiva, 2017. 204 p.

5 MORAIS, I. S.; ZANIN, A. **Engenharia de Software**. Porto Alegre: SAGAH, 2017. 34 p.

6 Ceará, Lianne; Amorozo, Marcos; Buono, Renata. No Brasil, proporção de adultos com nível superior é menos da metade da dos Estados Unidos. **Folha de São Paulo**, São Paulo, 14 mai. 2021. Disponível em: <https://piaui.folha.uol.com.br/no-brasil-proporcao-de-adultos-com-nivel-superior-e-menos-da-metade-da-dos-estados-unidos/>. Acesso em: 5 jan. 2024.

7 CERIGATTO, M. P.; GUIDOTTI M. **Tecnologias digitais na prática pedagógica**. Porto Alegre: SAGAH, 2018. 25 p.

8 Ramos, G. **Brasileiros passam mais da metade de suas vidas na Internet, estima pesquisa**. Rio de Janeiro: TechTudo, [2022]. Disponível em: <https://www.techtudo.com.br/noticias/2022/05/brasileiros-passam-mais-da-metade-de-suas-vidas-na-internet-estima-pesquisa.ghtml>. Acesso em: 5 jan. 2024.

9 BRASIL. Lei nº 14.129, de 29 de março de 2021. Dispõe sobre princípios, regras e instrumentos para o Governo Digital e para o aumento da eficiência pública e altera a Lei nº 7.116, de 29 de agosto de 1983, a Lei nº 12.527, de 18 de novembro de 2011 (Lei de Acesso à Informação), a Lei nº 12.682, de 9 de julho de 2012, e a Lei nº 13.460, de 26 de junho de 2017. **Diário Oficial da União**: seção 1, Brasília, DF, ano 159, n. 60, p. 3-7, 30 mar. 2021

10 Tauberer, Joshua. **The 8 Principles of Open Government Data**. Califórnia: Open Gov Data, [2007]. Disponível em: <https://opengovdata.org/>. Acesso em: 05 jan. 2024.

11 W. R. Fitriani *et al.* **Evaluating the implementation of open data principles on government websites in Indonesia**. Jakarta: Semantic Scholar, [2017]. Disponível em:



<https://www.semanticscholar.org/paper/Evaluating-the-implementation-of-open-data-on-in-Fitriani-Hidayanto/75ae64670b96e224c3a905b78550abad9cc9ef0f>. Acesso em: 6 jan. 2024

12 MEC – Ministério da Educação. **2022 – Relatório SISU – Chamada Regular 2/2022**. Brasília: MEC, 2023. Disponível em: <https://dadosabertos.mec.gov.br/sisu/item/203-2022-relatorio-sisu-chamada-regular-2-2022>. Acesso em: 6 jan. 2024. Base de dados

13 Ramiro, André; Canto, Mariana. **A importância social e econômica da Criptografia**. Recife: Coalizão Direitos na Rede, [2018]. Disponível em: <https://cartilhacriptografia.direitosnarede.org.br/index.html/>. Acesso em: 06 jan. 2024

14 IBGE – Instituto Brasileiro de Geografia e Estatística. Sistema de Contas Regionais: Brasil 2021. **Estatística**, Brasília, v. 94, p. 9, jan./dez. 2023. Disponível em: [https://biblioteca.ibge.gov.br/visualizacao/livros/liv102045\\_informativo.pdf](https://biblioteca.ibge.gov.br/visualizacao/livros/liv102045_informativo.pdf). Acesso em: 06 jan. 2024.

15 MEC – Ministério da Educação. **Instituições de Educação Superior do Brasil**. Brasília: MEC, 2022. Disponível em: <http://dadosabertos.mec.gov.br/indicadores-sobre-ensino-superior/item/181-instituicoes-de-educacao-superior-do-brasil>. Acesso em: 6 jan. 2024. Base de dados.

16 IBGE – Instituto Brasileiro de Geografia e Estatística. **Panorama do Censo 2022**. Brasília, DF: IBGE, c2022. Disponível em: <https://censo2022.ibge.gov.br/panorama/>. Acesso em: 6 jan.2024.

17 IBGE – Instituto Brasileiro de Geografia e Estatística. **Tabela 7269 – Pessoas de 25 anos ou mais de idade por sexo e nível de instrução**. Brasília, DF: IBGE, c2022. Disponível em: <https://sidra.ibge.gov.br/tabela/7269>. Acesso em: 6 jan.2024.

18 GUJARATI, D. N. **Econometria Básica**. Tradução: DURANTE, D.; ROSEMBERG, M.; ROSA, M. L. G. L. 5. ed. Porto Alegre: AMGH, 2011. 846 p.

19 OLIVEIRA, F. E. M. **Estatística e Probabilidade – Exercícios Resolvidos e Propostos**. 3. ed. Porto Alegre: AMGH, 2011. 846 p.

20 Rodrigues, Erica Castilho. Modelos de Regressão Múltipla – Parte II. **UFOP**, 25 out. 2016. Dataset. DOI sites/default/files/ericarodrigues/files/regressaolinearmultipla\_parte2.pdf. Disponível em: [http://professor.ufop.br/sites/default/files/ericarodrigues/files/regressaolinearmultipla\\_parte2.pdf](http://professor.ufop.br/sites/default/files/ericarodrigues/files/regressaolinearmultipla_parte2.pdf). Acesso em: 6 jan. 2024.

21 StatLect. **F distribution**. Geneva: StatLect, c2020. Disponível em: <https://www.statlect.com/probability-distributions/F-distribution>. Acesso em: 6 jan. 2024.

22 MASCHIETTO, L. G. *et al.* **Desenvolvimento de Software com Metodologias Ágeis**. Porto Alegre: SAGAH, 2020. 213 p.



23 FREEMAN, E. **DevOps para Leigos**. Tradução: SAMANTHA, B. Rio de Janeiro: Alta Books, 2021. 213 p.

24 CARDOSO, L.C. **Frameworks back end**. 1. ed. São Paulo: Platos Soluções Educacionais S.A., 2021. 54 p.

