

REGRESSÃO LOGÍSTICA APLICADA NA RELAÇÃO ENTRE O TEMPO DE PROVA E RENDIMENTO DE ALUNOS DE DISCIPLINAS RELACIONADAS A ESTATÍSTICA

Esttefani Duarte Brum*

Gilberto Rodrigues Liska**

Alisson Darós Santos***

RESUMO

O tempo com que um aluno leva para concluir uma avaliação pode influenciar no seu rendimento? Para responder essa pergunta, foi utilizado o modelo de regressão logística e utilizou-se o sistema computacional R como ferramenta de auxílio para a programação dos gráficos e cálculos estatísticos. Em seu desenvolvimento considerou-se a avaliação como forma de quantificar o desempenho de um aluno refletindo seu grau de conhecimento em um determinado conteúdo. Para tal utilizamos registros dos momentos inicial e final ao desenvolver uma avaliação. Os registros de tempos gastos foram obtidos de cinco turmas distintas de graduação, com disciplinas ministradas por mesmo docente, com igual teor teórico e em uma mesma universidade. Os resultados confirmam estatisticamente que cada minuto adicional que o aluno permaneça fazendo a prova, implica em maiores chances de obter bom rendimento, bem como diferenças de rendimento entres os gêneros femininos e masculinos, embora não estatisticamente diferentes, demonstrando que pessoas do gênero feminino possuem maiores chances de alcançar a média de 6 pontos. O modelo ainda confirma, de acordo com as razões de chances, que ao longo das avaliações o rendimento dos alunos diminui, tendo a melhor nota na primeira avaliação. Por meio das referências consultadas compreendemos que a diferença nas notas de cada aluno sofre influencias de fatores diversos, resultado de suas vivencias particulares.

Palavras-chave: Regressão logística, tempo em avaliação, razão de chances.

ABSTRACT

Can the time it takes a student to complete an assessment influence his / her performance? To answer this question, the logistic regression model was used and the computer system R was used as an aid tool for the programming of graphs and statistical calculations. Keywords: Logistic regression, time under evaluation, odds ratio. In its development, evaluation was considered as a way of quantifying the performance of a student reflecting his degree of knowledge in a given content. For this, we use records of the initial and final moments when developing an assessment. The records of time spent were obtained from five different undergraduate classes, with subjects taught by the same teacher, with the same theoretical content and at the same university. The results confirm statistically that

each additional minute that the student remains taking the test, implies in greater chances of obtaining a good performance, as well as differences in performance between the feminine and masculine genders, although not statistically different, demonstrating that the feminine gender has greater chances of reaching the average of 6 points. The model also confirms, according to the odds ratios, that throughout the evaluations the students' performance decreases, having the best score in the first evaluation. Through the references consulted, we understand that the difference in the grades of each student is influenced by different factors, the result of their particular experiences.

1. INTRODUÇÃO

As discussões a respeito de possíveis fatores que possam influenciar no rendimento de um aluno enquanto é avaliado, vem aumentando significativamente nos ambientes educacionais, onde tal preocupação percorre todos os níveis de ensino. Contudo avaliar é uma prática que se encontra na atuação de qualquer educador, sendo indispensável, pois perpassa as diferentes competências e habilidades do grupo de discentes. A área de Educação Matemática tem se preocupado com a avaliação e proposição de métodos de ensino em matemática que podem proporcionar o melhor entendimento dos conceitos envolvidos nesta ciência.

Uma das formas de quantificar o desempenho de um aluno frente a um método de avaliação é atribuir uma nota que reflita o grau de conhecimento do aluno a um determinado problema. Existem diversas formas de se avaliar o desempenho de um aluno observando sua nota, que podem ser resumidas a simples cálculos de estatísticas descritivas ou, métodos mais complexos como os usados no Exame Nacional do Ensino Médio, a teoria de resposta ao item (TRI), (ZANON et al., 2016).

Como mencionado, esses métodos levam em consideração a nota do aluno a uma determinada questão e representam, de fato, uma forma bastante coerente de avaliação de desempenho do ponto de vista estatístico. Mas, talvez, outras variáveis possam influenciar no desempenho de um estudante, tais como: o nível de ansiedade, tempo de preparação para uma avaliação, conhecimentos prévios, condições de saúde, alimentação, entre outras.

Entre essas variáveis, o tempo de realização de uma prova pode, em várias situações, determinar o sucesso ou fracasso. Para um velocista, o tempo de prova

é, obviamente, o fator determinante para a vitória. Em geral, para corridas o tempo pode ser considerado a variável mais importante no sentido de quantificar o rendimento de um atleta. Mas com relação a um aluno que será submetido a um exame avaliativo, o tempo gasto para realização do mesmo pode influenciar no seu rendimento?

Diante do exposto, o presente trabalho utilizou o modelo de regressão logística, uma vez que esse modelo descreve o comportamento entre uma variável dependente e variáveis independentes (FAVÉRO et al., 2009). Além da influência do tempo de prova, foi investigado se o gênero também se apresenta como uma variável capaz de influenciar na obtenção de uma nota média. Avaliamos a qualidade de ajuste do modelo e fornecemos razões de chances para mensurar o efeito do tempo de prova e outras variáveis na obtenção de uma nota média.

Para tal, foram avaliadas as notas e tempos gastos no desenvolvimento de avaliações de componentes curriculares relacionados a estatística, de alunos integrantes de cinco cursos distintos, ambos ministrados na Universidade Federal do Pampa - UNIPAMPA campus de Itaqui/RS. Das avaliações, foram registrados os tempos gastos para o desenvolvimento delas e as notas obtidas.

Por fim, consideramos que, enquanto docente, preocupar-se com estes fatores e refletir sobre eles, é comprometer-se com a prática pedagógica e com a própria formação profissional, já que a gente se faz educador sendo educador, permanentemente na prática e na reflexão sobre a prática (FREIRE, 1997). Nesse sentido, estudos do tipo podem viabilizar ações com respeito ao planejamento de aula, avaliações e condutas no ambiente acadêmico.

2. DESENVOLVIMENTO

Esta seção fará menção aos sistemas de avaliação e sua importância para a prática pedagógica, bem como formas que está acontecendo. Na sequência serão discutidos os métodos estatísticos e alguns aspectos da influência do tempo em avaliação e o bom rendimento.

2.1. SISTEMAS DE AVALIAÇÃO

O processo de avaliação faz parte do cotidiano de todos os educadores, a qual contribui para o diagnóstico do aprendizado e desempenho do aluno, bem como para reflexão da prática docente. Portanto, para a prática docente, avaliar não é realizar uma atividade técnica, pois não se dá no vazio conceitual, mas está referenciada em um modelo teórico de mundo, de ciência e de educação, o qual se faz na prática pedagógica (CHUIEIRE, 2008).

Essa ideia remete-se a avaliação de forma global e contínua perpassando as diferentes competências e habilidades do grupo de discentes, aplicada nos variados momentos, perante os argumentos, interesses e colocações apresentados pelos alunos em sala de aula, já que é na troca de conhecimentos que acontece a aprendizagem. Para Kraemer (2005), o termo avaliação vem do latim, e significa valor ou mérito ao objeto em pesquisa, junção do ato de avaliar ao de medir os conhecimentos adquiridos pelo indivíduo.

A autora, ainda destaca que a avaliação é um instrumento valioso no sistema escolar, já que este é capaz de descrever os conhecimentos que os alunos se apropriaram, bem como as dificuldades encontradas no processo de ensino aprendizagem (KRAEMER, 2005). Uma vez que a avaliação se consolida, é possível identificar evidências, capazes de apontar necessidades de melhorias no ensino e objetivos alcançados que foram significativos.

Para o aluno que está sendo avaliado, além do estudo, a administração correta do tempo no decorrer da avaliação, é fator fundamental para um bom rendimento em uma prova. De modo que o discente tenha cuidado para não permanecer tempo demais em uma única questão e extrapolar o tempo, ou se desestimular no início da prova e desistir de testar sua capacidade de raciocínio. Boff (2010) aponta que administrar bem o tempo é buscar eficiência nas atividades e conseqüentemente aumentar a produtividade.

2.2. FORMAS DE AVALIAÇÃO

A avaliação é a ferramenta mais importante que o professor (a) tem a sua disposição para chegar ao objetivo acadêmico: Justificar o avanço ou a permanência do discente em determinada disciplina. Contudo, está deve ocorrer

seguindo formas diversificadas e, acontecer no decorrer das aulas, com datas pré-estabelecidas, utilizando-se de instrumentos diversos, tais como apresentação de trabalhos, provas e resolução de exercícios.

A Educação Matemática, propõe possibilitar que todos os alunos sejam capaz de investigar e formular conjecturas matemáticas, desenvolver, avaliar, aplicar e expor argumentos, selecionar e usar diversos tipos de raciocínios e métodos de prova (VISEU et al., 2017). Assim, mesmo que cada docente adote uma forma de avaliar, é imprescindível que os significados que ficam por traz da essência do processo sejam considerados, como um meio para contemplar seus alunos e, identificar aprendizado ou dificuldade expressado pelo discente. Antunes (2002) afirma que, a tarefa do professor ao avaliar exige competência, discernimento, equilíbrio, além, é claro, de conhecimentos técnicos.

A valorização frente a produção do discente na resolução de exercícios propostos, também possibilita avaliar e constatar seu desempenho em determinada disciplina, uma vez que este compreende uma coleção intencional de trabalhos do aluno, com o objetivo de contar a história dos esforços, progressos e desempenho em uma determinada área (FREIRE, 1997).

Entretanto, quando a avaliação permeia o desenvolvimento de atividades utilizando-se de materiais para consulta, novas formas de aprender emergem, já que a aprendizagem, de fato é efetivada quando o conteúdo é posto em prática. Os trabalhos desenvolvidos com consulta “ajuda o aluno a resolver um caso, a escolher as fontes adequadas para delas retirar informações que lhes faltem” (ALTHAUS; GODOY, 2008; ZANON; ALTHAUS; BAGIO, 2018).

2.3 MÉTODOS ESTATÍSTICOS

Considerando a importância da Estatística para a compreensão de conceitos na atualidade, promover a aprendizagem efetiva torna-se cada vez mais importante, já que são inúmeras as suas aplicações. Estudos apontam que os discentes veem na Estatística um alto nível de dificuldade de compreensão, uma vez que fazem associações à aprendizagem Matemática e concluem, a priori, que também não conseguirão aprender Estatística (JACOBINI; WODEWOTZKI 2015).

É evidente que o estudo da Estatística exige habilidades básicas para o seu entendimento. Tais como capacidades de organizar dados, construir e apresentar tabelas, trabalhar com diferentes representações dos dados, entendimento de conceitos, vocabulário e símbolos e, além disso, um entendimento de probabilidade como medida de incerteza (JACOBINI; WODEWOTZKI; FERREIRA, 2011).

Nesse sentido, Wild e Pfannkuch (1999 apud DE OLIVEIRA MENDONÇA e LOPES, 2011) apresentam algumas ações importantes para o raciocínio estatístico, sendo estes:

- O reconhecimento da necessidade dos dados: consiste em perceber se a partir da análise dos dados é possível a tomada de decisão.
- A transnumeração: indica a compreensão que pode surgir ao trocar a forma de representação (gráfico, tabela).
- A percepção da variabilidade: a coleta adequada de dados e os julgamentos elaborados a partir destes requerem a compreensão da variabilidade que existe e se transmite nos dados.
- O raciocínio com modelos: Qualquer representação de dados, como um gráfico, uma tabela, uma reta de regressão ou um resumo, é um modelo representativo da realidade e permite observar o comportamento da variável em estudo, o que possibilita buscar a aproximação deste com um modelo estatístico.
- A integração da Estatística com o contexto: os dados que serão observados sob a perspectiva dos conceitos estatísticos pertencem a um contexto, e conhecê-lo é fundamental para a compreensão dos significados expressos nos dados.

Em decorrência destas considerações, deu-se a referida investigação, a qual objetiva identificar se o tempo que os alunos utilizam para desenvolver uma prova de caráter dissertativo, em disciplinas de componentes curriculares relacionados à Estatística, pode influenciar ou não nos resultados obtidos pelo discente. Considerando que estes, compõem grupos de alunos de diferentes cursos de graduação, mas que cursaram a disciplina com mesmo docente e igual teor teórico.

3. METODOLOGIA

3.1. OS DADOS

Os dados foram obtidos de avaliações de alunos que cursaram componentes curriculares relacionados à Estatística, ambas ministradas pelo mesmo docente. Os alunos pertenciam a diferentes cursos, sendo tais: Agronomia, Ciência e Tecnologia de Alimentos, Engenharia Cartográfica e de Agrimensura, Bacharelado Interdisciplinar em Ciência e Tecnologia e Licenciatura em Matemática, da Universidade Federal do Pampa - UNIPAMPA, campus de Itaqui/RS. Foram registradas 784 notas, sendo estas distribuídas em cinco cursos, três anos e três tipos de prova.

Das avaliações, foi registrado o momento inicial e de entrega de cada aluno. O tempo de prova é dado pela diferença entre o tempo de entrega e inicial, convertido em minutos. A fim de se estabelecer um limiar de desempenho de prova, adotou-se a nota de seis (6) pontos, uma vez que se o aluno obtiver essa nota em todos os sistemas de avaliação, ele é aprovado. Dessa forma, os dados foram analisados estatisticamente por meio de Regressão Logística, tendo para aplicabilidade do modelo proposto a categorização das notas da seguinte forma:

1: Notas de alunos, maior ou igual a seis (6) pontos.

0: Notas de alunos, menor que seis (6) pontos.

Convém salientar que a escolha do limiar de 6 pontos deu-se pelo interesse do aluno obter uma nota superior à média praticada na UNIPAMPA para aprovação na componente curricular e, no caso específico das componentes de Estatística, esse mesmo limiar foi aplicado para todas as avaliações dissertativas dessa componente. Esse limiar poderia, de acordo com um interesse específico, ser mudado.

3.2. MODELO ESTATÍSTICO

Nesta seção apresenta-se o contexto em que a variável resposta possui apenas duas categorias, ou seja, natureza binária ou dicotômica. O modelo de Regressão Logística faz parte da classe dos Modelos Lineares Generalizados (MLG). O MLG é especificado por três componentes: uma componente aleatória, a qual identifica a distribuição de probabilidade da variável dependente, uma

componente sistemática, que especifica uma função linear entre as variáveis independentes e uma função de ligação, que descreve a relação matemática entre a componente sistemática e o valor esperado da componente aleatória (HOSMER; LEMESHOW; STURDIVANT, 2013).

Considere a transformação logística, também chamada de *logit* (MENEZES et al., 2017), dada por

$$\text{logit} = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (1)$$

Para se analisar $\pi(x)$, tomam-se as observações independentes x_1, x_2, \dots, x_n . Nesse contexto, é razoável assumir, como suposição inicial, que $\pi(x)$ é uma função monotônica com valores entre zero e um quando x_1, x_2, \dots, x_p variam na reta real. Como $\pi(\cdot)$ varia entre zero e um, uma representação linear simples para π sobre todos os possíveis valores de x não é adequada, uma vez que os valores da forma linear estão no intervalo $(-\infty; +\infty)$. Nesse caso, uma transformação deve ser utilizada a fim de permitir que, para qualquer valor de x_1, x_2, \dots, x_p , tenha-se um valor correspondente para $\pi(\cdot)$ no intervalo $[0; 1]$. O inverso da função *logit* é a função logística, que é dada por

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \quad (2)$$

em que $\pi(x)$ varia em $[0; 1]$. No caso de termos uma única variável independente no modelo, digamos x_1 , se $\beta_1 > 0$, π é crescente e se $\beta_1 < 0$, π é decrescente. Quando x tende ao infinito, $\pi(x)$ tende a zero quando $\beta_1 < 0$ e tende a um quando $\beta_1 > 0$. Caso $\beta_1 = 0$, a variável resposta é independente da variável x , logo $\pi(x)$ é constante. O caso $\beta_0 = 0$ e $\beta_1 = 0$ corresponde a $\pi(x) = 0.5$.

Uma vez definido o modelo, deve-se proceder à estimação de parâmetros. Seja $\boldsymbol{\beta} = (\beta_0, \beta_1)$ o vetor de parâmetros relacionado com a probabilidade condicional $P(Y_i = 1|x_i) = \pi(x_i)$, com $\pi(x_i)$ dado por

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_1 x_{i1})}{1 + \exp(\beta_0 + \beta_1 x_{i1})} \quad (3)$$

O método usual para estimar $\hat{\beta}$ é via Máxima Verossimilhança. Considerando que y_i tem distribuição Bernoulli, a contribuição de (x_i, y_i) à função de Verossimilhança é dada por

$$\pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}$$

para valores $y_i = 0$ ou $y_i = 1$, para $i \in \{1, \dots, n\}$. Como assume-se que as observações são independentes, a função de verossimilhança (L), é dada por

$$L(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (4)$$

Aplicando-se logaritmo neperiano na equação (4) tem-se a expressão (5), ($l(\beta)$),

$$l(\beta) = \ln[L(\beta)] = \sum_{i=1}^n [y_i \ln(\pi(x_i)) + (1 - y_i) \ln(1 - \pi(x_i))] \quad (5)$$

Pelo princípio da Máxima Verossimilhança, o estimador $\hat{\beta}$ é o valor que maximiza a expressão 5. Assim, deriva-se $l(\beta)$ com respeito a β_0 e β_1 e igualam-se as expressões resultantes a zero, resultando num sistema de equações não-lineares nos parâmetros e dessa forma, é requerido o uso de um procedimento iterativo numérico para encontrar $\hat{\beta}$, como por exemplo o método de Newton-Raphson (MCCULLOCH; SEARLE; NEUHAUS, 2009).

3.2.1. SELEÇÃO DE MODELO E CRITÉRIOS DE QUALIDADE DE AJUSTE

O método *Stepwise* de seleção de variáveis consiste em eliminar do modelo variáveis que não contribuem de maneira significativa para o valor esperado da variável resposta, no caso do modelo logístico, para a probabilidade de ocorrência de um evento de interesse. Nesse trabalho foi utilizado o critério de informação de Akaike (AIC) para seleção de variáveis por meio do método *Stepwise* de seleção de variáveis, que consiste dos seguintes passos:

- (1) ajustamos o modelo completo com todas variáveis independentes;
- (2) retiramos uma variável independente por vez, ajustamos o modelo e calculamos o AIC, proposto em Akaike (1974), dado por

$$AIC = -2 l(\hat{\beta}) + 2p \quad (6)$$

em que p é o número de parâmetros do modelo;

- (3) retiramos do modelo completo a variável independente que produziu o maior AIC;
- (4) reajustamos o modelo sem a variável independente retirada no passo (3);
- (5) voltamos ao passo (2) e refazemos o processo até não haver variável independente para ser retirada. Após esse procedimento e com o modelo obtido, refazemos o processo inverso do passo (1), ou seja, incluiremos, se possível, variáveis independentes no modelo utilizando o AIC.

Dado um conjunto de modelos ajustados aos dados, o modelo preferido é o que apresentar menor valor de AIC (EMILIANO; VIVANCO; DE MENEZES, 2014; LISKA et al., 2019).

A qualidade do ajuste também pode ser verificada por um gráfico de probabilidade normal para resíduos padronizados, fornecendo uma indicação gráfica de que a suposição de distribuição Bernoulli é violada. Esse gráfico, também chamado de gráfico simulado por envelope, contém bandas de confiança construídas por reamostragem e considera-se que ocorreu um ajuste adequado se todos os resíduos do modelo (ou pelo menos a maioria deles) estiverem contidos nessas bandas (CORDEIRO, 2004; GILBERTO, 1995; MORAL; HINDE; DEMÉTRIO, 2017).

3.2.2. RAZÃO DE CHANCES

Considere inicialmente o modelo definido em 3, com $\pi(x)$ a probabilidade de um aluno obter uma nota superior a seis pontos. A razão de chances quantificará o efeito de uma particular variável independente, que pode ser o tempo de prova, gênero e/ou turma do aluno, na probabilidade de o aluno obter uma nota superior a seis pontos. Dessa forma, a chance (odds) de o aluno obter nota superior a 6, para um indivíduo com presença ($x = 1$) do fator, denotada por OD_1 , fica dada por

$$OD_1 = \frac{\pi(1)}{1 - \pi(1)} = e^{\beta_0 + \beta_1}$$

enquanto que a chance de nota superior a 6 para um indivíduo com ausência ($x = 0$) do fator, denotado por OD_0 , é

$$OD_0 = \frac{\pi(0)}{1 - \pi(0)} = e^{\beta_0}$$

Logo, a razão de chances (odds ratio) de ocorrência de nota superior a 6 entre indivíduos com presença e ausência do fator fica dada por

$$OR = \frac{\pi(1)[1 - \pi(0)]}{\pi(0)[1 - \pi(1)]} = e^{\beta_1} \quad (7)$$

dependendo apenas do parâmetro β_1 . Esta é uma das grandes vantagens da regressão logística: a possibilidade de interpretação direta dos coeficientes como medidas de associação (GIANNINI; LATORRE; FERREIRA, 2012; SIO JYH LIH; ISMAIL, 2019). Esse conceito pode ser estendido para o caso múltiplo de variáveis independentes, só que, nesse caso, a razão de chances é feita entre a variável de interesse, como mencionado anteriormente, e mantendo-se fixas todas as outras variáveis explicativas, levando a equação 8 (LEITE et al., 2011).

O valor observado da variável independente no modelo logístico pode representar o valor de alguma variável quantitativa qualquer, como, por exemplo, o tempo de prova de um aluno. Nesse caso, faz sentido calcularmos a razão de chances de um indivíduo ter nota superior a 6 pontos para cada incremento que pode ser de uma unidade ou mais, no seu tempo de prova. A razão de chances de obtenção de nota superior a 6 para um incremento c , tal que $c = x^x - x$, fica dada por

$$OR_{(x^x - x)} = \frac{\pi(x^*)[1 - \pi(x)]}{\pi(x)[1 - \pi(x^*)]} = e^{\beta_1(x^* - x)} \quad (8)$$

Finalizando a metodologia, será utilizado como ferramenta de auxílio para o desenvolvimento do modelo o sistema computacional R (R CORE TEAM, 2017) e

seu ambiente integrado RStudio (RSTUDIO TEAM, 2015), os quais são softwares livres que permitem a programação para gráficos e cálculos estatísticos.

4. RESULTADOS E DISCUSSÃO

Para melhor organização dos resultados, estes serão apresentados em três cenários: influência do tempo no desempenho do aluno; adição da covariável gênero com relação ao desempenho do aluno; e por fim o tipo de prova.

4.1. INFLUÊNCIA DO TEMPO NO DESEMPENHO DO ALUNO

O processo de avaliação tem suas contribuições pautadas no diagnóstico do desempenho do aluno, onde vários fatores são capazes de causar impactos positivos e negativos, como os destacados por Zewude e Ashine (2016) que aponta que o tempo de estudo, influência dos colegas, quantidade de dinheiro recebidas da família, boa vida posterior aos anos de estudo e escolaridade dos pais são variáveis capazes de influenciar no desempenho acadêmico.

Para o autor estudantes que estudam mais de 48 horas semanais possuem 5,85 vezes mais chances de ter melhor desempenho acadêmico em relação aos que estudam menos horas (ZEWUDE, 2016). Outro estudo interessante também aponta a renda familiar como variável significativa para a frequência e bom rendimento acadêmico, sendo que quando a renda mensal aumenta em um a probabilidade de o aluno continuar frequentando as aulas é 1,0001 vezes maior que a probabilidade daqueles alunos cuja renda mensal é inferior a uma unidade (ABDELLATIF FADUL; OSMAN ABD ELGHAFAR MOHAMMED, 2018).

Ponderando neste sentido ao interpretarmos o coeficiente no modelo de regressão logística, constatamos estatisticamente que o tempo é capaz de influenciar no desempenho do aluno ao realizar uma avaliação (tabela 1).

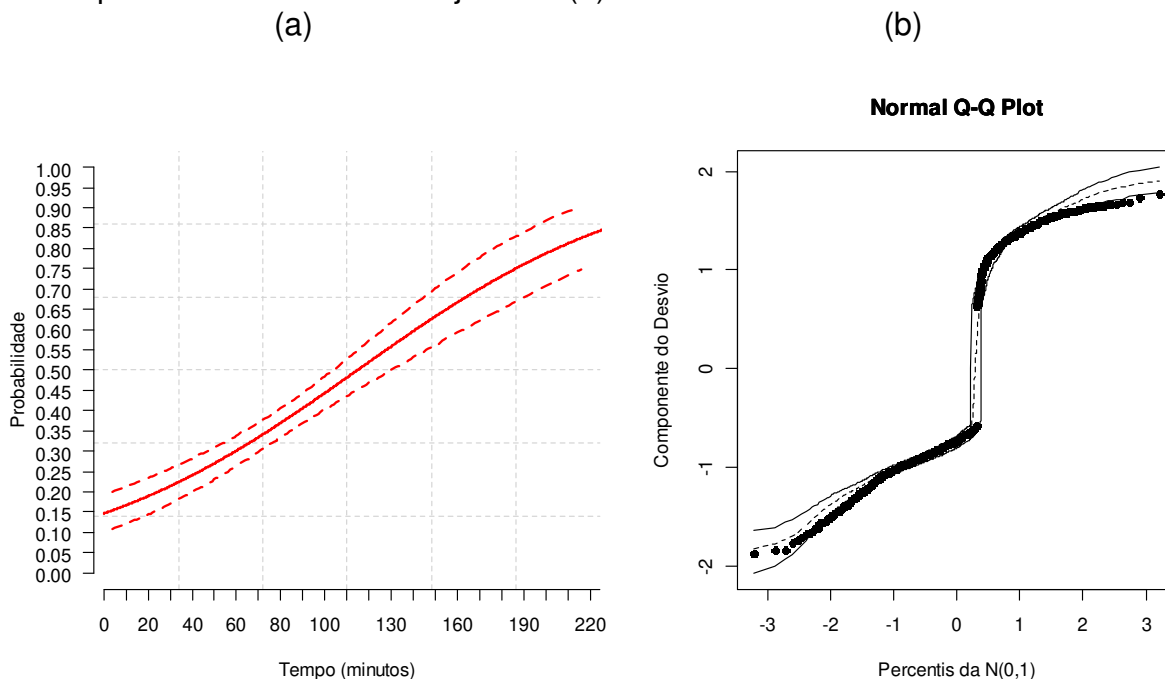
Tabela 1- Estimativas dos parâmetros do modelo de regressão logística e razão de chances (OR) para o desempenho dos alunos em relação ao tempo.

Parâmetro	Estimativa	Erro padrão	p-valor	OR ₁	OR ₁₀	AIC
β_0	-1,763370	0,184550	<0,0001	1,0155	1,116	964,73
β_1	0,015362	0,001996	<0,0001			

Fonte: A autora (2020).

Diante dessa afirmação, há uma aparente interpretação em que para cada minuto adicional que o aluno permanece fazendo a prova, a chance de sucesso atingindo a média aumenta em 1,55% (OR₁). Além disso, a cada 10 minutos que o aluno permaneça fazendo a avaliação, resulta em um aumento de 11,6 % (OR₁₀) em sua chance de atingir a média.

Figura 1- Gráfico do modelo ajustado do tempo de realização da prova (a) e envelope simulado do modelo ajustado (b)



Fonte: A autora (2020).

Assim avaliar não é simplesmente realizar uma atividade técnica, e sim um processo global e contínuo que perpassa as diferentes competências do fazer pedagógico. O fornecimento de objetivos de aprendizagens claros é um meio de contemplar positivamente os alunos, obtendo maiores chances de satisfazer os

alunos em geral com o serviço de disciplinas. Esse fato tem aumentando em 6 vezes o engajamento em relação às disciplinas não ministradas com tal objetividade (ALIJA; SNOPE; ALIU, 2016).

Estudo semelhante interpretado a partir do coeficiente do modelo de regressão, confirmou associação entre distúrbio de voz e perda de capacidade para o trabalho em professoras de São Paulo, apontando $p=0,020$, indicando que professoras têm 2,2 vezes mais chance de terem distúrbio de voz quando comparadas com outras profissionais com menos uso vocal (GIANNINI; LATORRE; FERREIRA, 2013).

Já para o presente estudo, mesmo que haja relação direta entre o tempo gasto para o desenvolvimento de uma avaliação e o desempenho do aluno, é importante atentar-se ao tempo ideal que implica maiores chances de o aluno atingir média compatível para sua aprovação. Deste modo e conforme as estimativas dos parâmetros do modelo de regressão logística na tabela 1 sugere-se estatisticamente que, para o aluno ter 50% de chance de alcançar a média, é necessário que este permaneça 114,8 minutos fazendo a prova. Tal resultado é possível ser constatado fazendo:

$$x_{50\%} = \frac{\ln\left(\frac{0,5}{1-0,5}\right) - \hat{\beta}_0}{\hat{\beta}_1} = \frac{\ln\left(\frac{0,5}{1-0,5}\right) + 1,763370}{0,015362} = 114,8 \text{ min.} \quad (9)$$

De acordo com as referências, quando trata-se de tempo gasto pelo aluno no programa de pós graduação, sempre que este aumenta, as chances de concluir o mestrado também aumentam em um fator de 2,88 % (PYKE; SHERIDAN, 1993).

Para Alija (2015), que também desenvolveu um estudo ancorado no uso de regressão logística e buscou determinar alguns dos fatores de satisfação, que influenciam positivamente a percepção dos alunos em relação a algumas práticas de ensino, constatou que mais de 90% dos alunos estão satisfeitos ou completamente satisfeitos com o ensino oferecido por seus professores, pois os dados apontam que da amostra, 67,5% dos estudantes estão completamente satisfeitos, 29,3% satisfeitos e 3,2% discordam, tendo 96,8% dos estudantes satisfeitos versus 3,2% insatisfeitos.

4.2. INFLUÊNCIA DO TEMPO E GÊNERO NO DESEMPENHO DO ALUNO

Ao analisarmos outras variáveis como a influência do tempo e gênero no desempenho dos discentes, foi possível perceber uma diferença significativa entre o gênero feminino e masculino, os quais indicam que pessoas do gênero feminino possuem maiores chances de alcançar a média, pois para cada minuto que as alunas do gênero feminino permanecem fazendo a prova, suas chances aumentam em 1,96%, enquanto que para os alunos do gênero masculino, a cada minuto adicional fazendo a avaliação é de 1,24% (Tabela 2).

Tabela 2- Estimativas dos parâmetros do modelo de regressão logística e razão de chances (OR) para o desempenho dos alunos em relação ao tempo e gênero.

Parâmetro	Estimativa	Erro padrão	p-valor	OR _{Feminino}	OR _{Masculino}	AIC
β_0	-2,086400	0,274733	<0,0001	1,0196	1,0124	
β_1	0,019371	0,003133	<0,0001			965,71
β_2	0,589650	0,374909	0,1158			
β_3	-0,007027	0,004060	0,0835			

Fonte: A autora (2020).

Para Abedalaziz (2010) essas diferenças ocorrem por que pessoas do gênero feminino tem vantagens com relação aos do gênero masculino na capacidade numérica, daí decorre a importância da avaliação se dar por meio de diversos instrumentos possibilitando contemplar ao máximo os alunos como os apresentados na seção 2.2.

De acordo com a perspectiva apresentada por Mazulo (2015 apud Hair, 1998), por se tratar de Regressão Linear Múltipla, onde objetiva-se usar as variáveis independentes que possuem valores conhecidos, para prever os valores das variáveis independentes, é possível afirmar com 5% de significância, que as variáveis β_0 e β_1 são coeficientes diferentes de zero tendo associação real com a variável resposta.

O modelo proposto por meio de regressão logística que relaciona o tempo de prova e o Gênero é dado por:

$$\pi(\text{Tempo}, \text{Gênero}) = \frac{\exp(\beta_0 + \beta_1 \text{Tempo} + \beta_2 \text{Gênero} + \beta_3 \text{Tempo} \times \text{Gênero})}{1 + \exp(\beta_0 + \beta_1 \text{Tempo} + \beta_2 \text{Gênero} + \beta_3 \text{Tempo} \times \text{Gênero})} \quad (10)$$

Como a variável Gênero é categórica e de acordo com a Tabela 2, os modelos para o gênero Feminino e Masculino são dados por

$$\pi_{\text{Feminino}}(\text{Tempo}) = \frac{\exp(\beta_0 + \beta_1 \text{Tempo})}{1 + \exp(\beta_0 + \beta_1 \text{Tempo})} = \frac{\exp(-2,0864 + 0,0194 \times \text{Tempo})}{1 + \exp(-2,0864 + 0,0194 \times \text{Tempo})} \quad (11)$$

$$\pi_{\text{Masculino}}(\text{Tempo}) = \frac{\exp(\beta_0 + \beta_2 + (\beta_1 + \beta_3) \times \text{Tempo})}{1 + \exp(\beta_0 + \beta_2 + (\beta_1 + \beta_3) \times \text{Tempo})} = \frac{\exp(-1,49675 + 0,01234 \times \text{Tempo})}{1 + \exp(-1,49675 + 0,01234 \times \text{Tempo})} \quad (12)$$

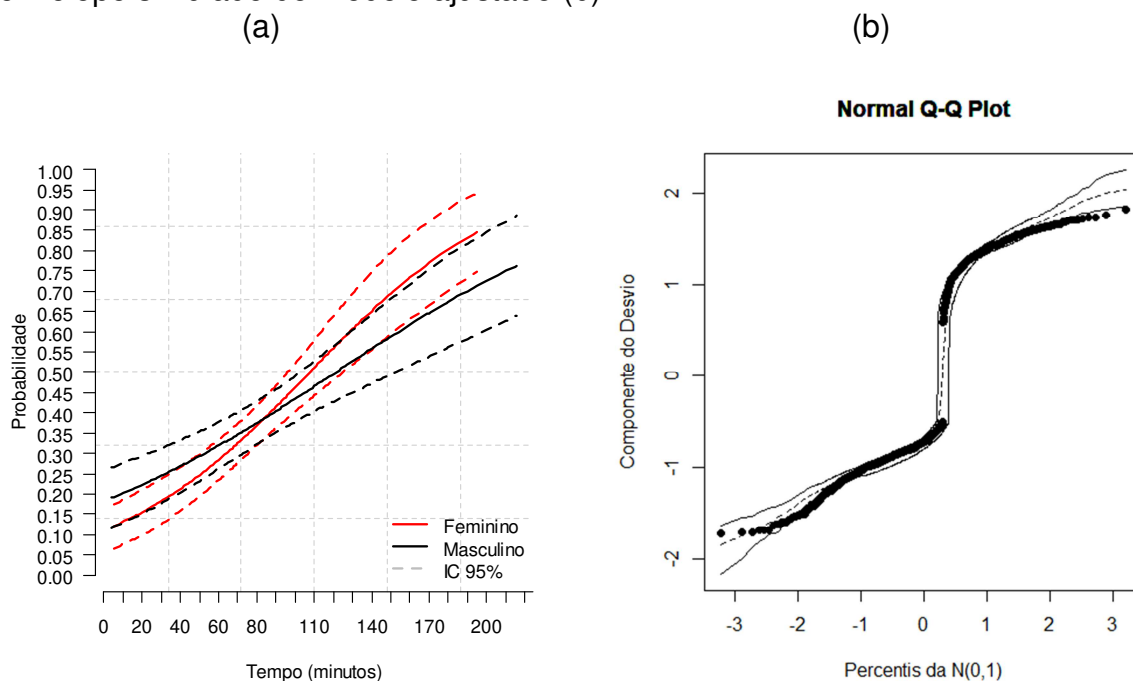
Tais afirmações acordam com as colocações feitas por Magalhães e Andrade (2002) as quais indicam que estudantes do gênero feminino obtém melhor rendimento do que estudantes do gênero masculino. Guney (2009) que também investiga a respeito da variável gênero, aponta que estudantes do gênero feminino superam os estudantes do gênero masculino no rendimento acadêmico, e Dal 'Igna (2007) completa afirmando que mesmo as atividades sendo as mesmas e os conhecimentos exigidos também os mesmos para alunos do gênero masculino e feminino, o rendimento de estudantes do gênero feminino geralmente é melhor.

Entretanto a investigação de Eratli Sirin e Sahin (2020) indicam o contrário, apontando que as chances de estudantes do gênero masculino obter sucesso é 1,427 vezes maior que a chance de sucesso para estudantes do gênero feminino, o qual ainda é justificado em razão da facilidade em terem bom relacionamento com professores que tornam o processo educacional mais ativo.

Em Daguplo (2017) o interesse em cursar ensino superior entre os alunos do ensino médio ouve diferença significativamente estatística entre gênero e idade. Estudantes mais velhas são menos propensas em seguir ensino superior em estudos realizados na cidade Irvine-Califórnia-EUA. Há também referências que apontam a decisão de jovens ao seguir no ensino superior sem influências significativas pelo variável gênero (DE DIOS JIMÉNEZ; SALAS-VELASCO, 2000). Contudo, os estudantes do gênero masculino são os que têm menor probabilidade de persistirem no curso de ensino superior (FERRÃO; ALMEIDA, 2018).

De maneira análoga a equação 9, o modelo ainda explica que, para o discente do gênero masculino ter 50% de chances de atingir a média em uma avaliação, então deve permanecer fazendo a prova, pelo menos 121,2 minutos, enquanto que, para discentes do gênero feminino, tendo como base a mesma probabilidade de chances o tempo diminui, para 107, 7 minutos.

Figura 2- Gráfico do modelo ajustado do tempo de realização da prova e gênero (a) e envelope simulado do modelo ajustado (b).



Fonte: A autora (2020).

Neste sentido Alija (2015) concorda ao analisar a variável gênero, pois afirma que as conquistas dos estudantes do gênero feminino são aproximadamente 1,022 vezes maiores que o da conquista de estudantes do gênero masculino, tal perspectiva foi apurada ao analisar a probabilidade de alunos e alunas passar no teste externo da Escola Secundária de Economia. Niu (2020) também considerou relevante a variável gênero em se tratando de desenvolvimento acadêmico, segundo o autor mulheres tem probabilidade superior em 33% aos homens de receber um diploma universitário.

4.3. SELEÇÃO DE VARIÁVEIS

Além das variáveis previamente analisadas e considerando a inclusão do curso ao qual o aluno pertence e o tipo de prova, o modelo completo é dado por:

$$\begin{aligned} \text{logit} &= \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) \\ &= \beta_0 + \beta_1 \text{Tempo} + \beta_2 \text{Gênero} + \beta_{31} \text{Curso}_1 + \beta_{32} \text{Curso}_2 \\ &\quad + \beta_{33} \text{Curso}_3 + \beta_{34} \text{Curso}_4 + \beta_{35} \text{Curso}_5 + \beta_{36} \text{Curso}_6 + \beta_{41} \text{Prova}_1 \\ &\quad + \beta_{42} \text{Prova}_2 \end{aligned} \quad (13)$$

Tabela 3 - Estimativas dos parâmetros do modelo de regressão logística completo.

Parâmetro	Estimativa	Erro padrão	p-valor	AIC
β_0	-0,80284	0,33969	0,01811	
β_1	0,01598	0,00211	<0,0001	
β_2	0,01429	0,17258	0,93400	
β_{31}	-0,74871	0,32513	0,02129	
β_{32}	-0,68505	0,34377	0,04629	
β_{33}	-0,41212	0,34669	0,23455	920,63
β_{34}	-0,57590	0,35527	0,10501	
β_{35}	-0,53871	0,81493	0,50858	
β_{36}	0,74814	0,87400	0,39199	
β_{41}	-1,14521	0,21162	<0,0001	
β_{42}	-0,93795	0,25121	0,00019	

Fonte: A autora (2020).

Após procedimento de seleção de variáveis pelo método *Stepwise*, foi possível obter o conjunto de variáveis da tabela 3 que produzisse o modelo mais parcimonioso. Assim, uma vez que o modelo logístico ajustado se mostrou adequado (figura 3 (b)), expressar um resultado mais completo é fundamental, no geral o modelo logístico ajustado apresentou que é possível o tempo (β_1) influenciar no rendimento dos alunos, o qual configura que enquanto o aluno o utiliza corretamente para desenvolver uma avaliação este terá maiores chances de garantir sucesso.

Para expressar um resultado mais prático a Tabela 3 compara as probabilidades estimadas de rendimento na prova 2 (β_{41}) e prova 3 (β_{42}), a qual configura que os alunos apresentam diferentes rendimentos nas avaliações. A chance de o aluno tirar uma nota maior que 6 na segunda prova em relação à primeira avaliação diminui em 72%. Com relação à terceira avaliação, a chance é de 61%. Observa-se que, de acordo com as razões de chances, que o rendimento do aluno diminui ao longo das avaliações, de modo que sua maior nota é maior na primeira avaliação em relação às avaliações 2 e 3 e, é menor na segunda avaliação.

Tabela 4- Estimativas dos parâmetros do modelo de regressão logística completo com variáveis selecionadas pelo método *Stepwise*.

Parâmetro	Estimativa	Erro padrão	p-valor	OR	AIC
β_0	-1,30710	0,19726	<0,0001		
β_1	0,01596	0,00205	<0,0001	OR ₁ =1,0161 OR ₁₀ =1,1731	916,35
β_{41}	-1,25011	0,19328	<0,0001	1-OR=0,7135	
β_{42}	-0,96519	0,22481	<0,0001	1-OR=0,6191	

Fonte: A autora (2020).

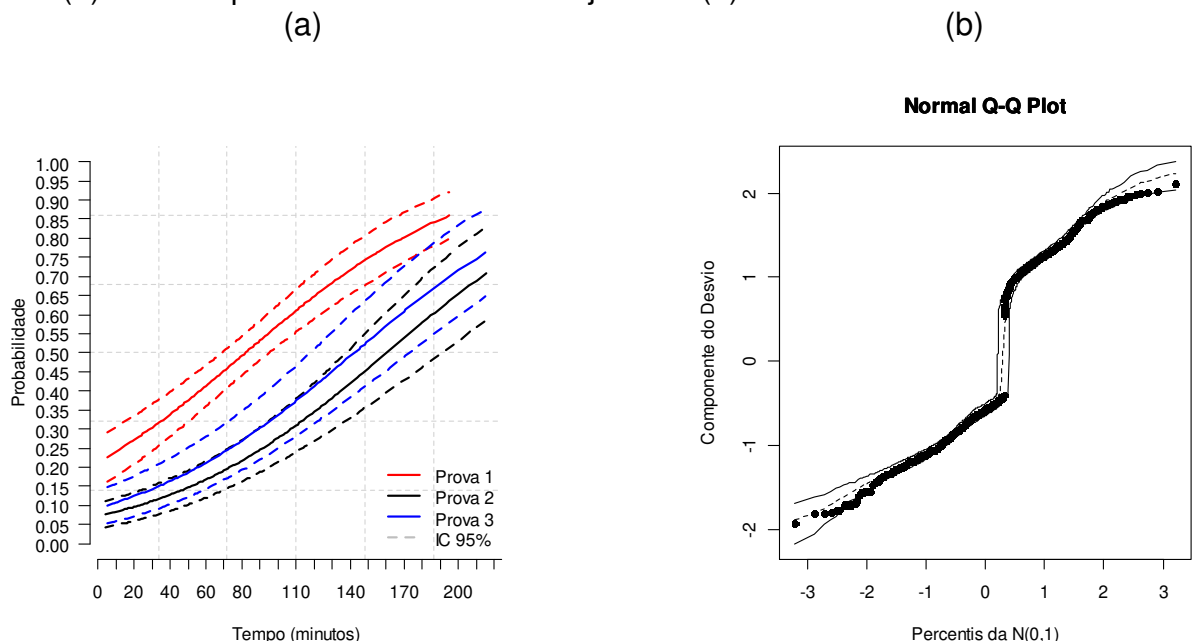
Deste modo tal estudo possibilita afirmar que na prova 2 é onde os alunos apresentam o menor rendimento, ficando a prova 1 como sendo a de melhor rendimento dos alunos. Considerando os resultados obtidos e os conceitos abordados em cada avaliação, intuimos que por termos na prova 2 conteúdos mais teóricos e conseqüentemente menos concreto ao aluno, poderá agregar menor

assimilação do conteúdo, pois nas provas 1 e 3, os conteúdos pressupõem ilustrar a realidade por meio da teoria, o que poderá gerar incentivo a criatividade e a prática investigativa, pois possibilita ao aluno interagir com os fenômenos aplicando os conceitos estudados (ANDRADE; MASSABNI, 2011).

Contudo este é o momento que permite verificar se as ações para o raciocínio estatístico foram compreendidas pelo estudante, como as destacadas por Wild e Pfannkuch (1999), apresentadas na seção 2.4. Formiga (2004) destaca que o êxito acadêmico é de responsabilidade do aluno, pois depende dele saber que tem que estudar, já que não se pode conceber que para ser bom aluno, a pessoa não venha a dedicar tempo aos estudos, pois a genialidade ou habilidade que se tem quanto a compreensão e assimilação dos conteúdos científicos não se aprende no vazio.

De tal modo acreditamos que a avaliação para o aluno vem para lhe orientar, o que resulta em um rendimento melhor na prova 3 em relação a prova 2, pois proporciona ao estudante a revisão de sua atuação, o que pode ser observado na Figura 3 (a).

Figura 3 - Gráfico do modelo ajustado do tempo de realização da prova e tipo de prova (a) e envelope simulado do modelo ajustado (b).



Fonte: A autora (2020).

Estudos como o de Pyke e Sheridan (1993), também apresentam o uso da Regressão Logística para formular modelos que descrevem os padrões de

retenção de estudantes de mestrado pertencentes as áreas de ciências naturais, ciências sociais e humanas, onde ao selecionarem vinte e um das variáveis independentes com três possíveis termos de interação, apenas quatro variáveis mostraram-se significativas e relacionadas à retenção de alunos em nível de mestrado. Dentre elas estão o tempo no programa, quantidade média de financiamento, média acadêmica de graduação e tipo de programa.

Conforme conclusões de Batistela, Rodrigues e Bononi (2009), as chances de um aluno desistir no curso de graduação podem ser explicadas pelas variáveis “Residência”, “Renda Familiar” e “Ocupação do Pai”. Pelos resultados apresentados os autores justificam que a chance de um aluno que mora em residência fora da cidade do curso desistir é 2,29 vezes a chance de um aluno que mora na cidade em que estuda. Se referindo a variável “Renda Familiar” destacam que a chance de um aluno com renda superior a 5 salários mínimos desistir é 2,02 vezes que a chance de um aluno com renda inferior a 5 salários. Com relação a “Ocupação do Pai” ressaltam que a chance de um aluno que a ocupação do pai é empresária ou política ou profissional liberal ou funcionário público tem chance de desistir 5,29 a chance de um aluno cujo pai é militar ou agropecuarista ou assalariado ou autônomo.

No entanto o período de tempo gasto no programa foi o contribuinte mais importante do modelo de retenção no nível de mestrado, o que os leva a afirmar que, a maior probabilidade de conclusão do mestrado é representada por alunos que permanecem maior tempo no programa e conseqüentemente a menor probabilidade fica com os alunos que permanecem o menor tempo no programa, ainda é apontado que ao avaliar a influência da variável “tempo”, as descobertas não os surpreenderam tendo em vista que o tempo médio para a conclusão do mestrado equivale a 2,74 anos, já para aqueles que se retiram voluntariamente 1,39 anos e aqueles que são retirados pela universidade 1,47 anos, o que comprova uma diferença significativa na quantidade de tempo (PYKE; SHERIDAN, 1993).

Para Peng; Lee; Ingersoll (2002), ao testar a hipótese de existir ou não relação entre a probabilidade de uma criança ser recomendada para instruções de leitura corretiva e sua pontuação e gênero, foi possível constatar que para os

meninos as chances de estes serem indicado para programas de leitura corretiva é 1,9111 vezes maior que para a meninas, onde a afirmação deu-se a partir da plotagem das variáveis de cada grupo de gênero contra as variáveis de leituras e pontuações.

Portanto, o ato de avaliar perpassa todos os níveis de ensino, e nos leva a repensar e investir em estratégias e possibilidades para reduzir a cultura de reprovação ou retenção não apenas como método pedagógico, mas também para a reflexão e ressignificação da aprendizagem (LIMA et al., 2019).

Com base nos resultados obtidos ressaltamos a importância da avaliação para o processo de ensino aprendizagem, pois entendemos que na avaliação o aluno é capaz de mostrar que domina ou não o conteúdo trabalhado, de modo que consegue utilizá-lo para compreender fenômenos e atuar na sociedade (VIZZOTTO; MACKEDANZ, 2018).

Convém salientar que, para os resultados aqui apresentados, partimos do princípio de que o (a) aluno (a) ocupe seu tempo para efetivo desenvolvimento da avaliação, considerando que os resultados não podem ser considerados caso o aluno permaneça seu tempo na avaliação sem fazer atividade alguma.

5. CONCLUSÃO

O modelo de regressão logística se ajustou satisfatoriamente aos dados que relacionam o tempo de estudo com a nota obtida por alunos em uma avaliação de Estatística. Constatamos que conforme aumenta o tempo de prova do aluno, aumenta as chances de bom rendimento. Para cada minuto adicional que o aluno permanece fazendo a prova, a chance de sucesso atingindo a média aumenta em 1,55% e que em aproximadamente 115 minutos de trabalho efetivo, representa 50% de chances de o mesmo obter uma nota superior a 6 pontos.

Percebemos que as referências consultadas também indicam uma aparente influência da variável tempo em diversos aspectos relacionados à vida estudantil, se apresentando como um fator capaz de determinar sucesso ou fracasso dos discentes.

Os resultados não apontaram diferença não significativa entre os gêneros com relação ao desempenho dos discentes. No entanto, para cada minuto que as

alunas do gênero feminino permanecem fazendo a prova suas chances são ligeiramente maiores em relação aos alunos do gênero masculino. O modelo reitera, que das três avaliações propostas no decorrer da disciplina, a que os alunos apresentam melhor rendimento de acordo com as razões de chances é a prova 1, e ao longo das avaliações seus rendimentos diminuí, tendo a menor nota na prova 2.

No entanto, este estudo é conclusivamente limitado apenas quando o estudante usa seu tempo para o desenvolvimento efetivo da avaliação, na busca de soluções para as questões propostas.

Tudo isso nos faz acreditar que na educação, são inúmeros os fatores que podem causar impactos sobre o rendimento do aluno, sobretudo é importante que o docente considere, no momento de avaliar, quais significados ficaram por traz da essência do processo, como um meio para contemplar os alunos e, identificar aprendizado ou dificuldade expressado pelo discente.

REFERÊNCIAS

ABDELLATIF FADUL, A.; OSMAN ABD ELGHAFAR MOHAMMED, K. Using Logistic Regression to Model Factors Associated With Basic School Drop-Out in Kassala State, Sudan. **Global Journal of Pure and Applied Mathematics**, v. 14, n. 2, p. 301–313, 2018.

ABEDALAZIZ, N. Detecting gender related DIF using logistic Regression and Mantel-Haenszel approaches. **Procedia - Social and Behavioral Sciences**, v. 7, n. 2, p. 406–413, 2010.

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716–723, 1974.

ALIJA, S. Application of ordinal logistic regression in the study of students ' achievement in external testing. v. 8, n. 1, 2015.

ALIJA, S.; SNOPE, H.; ALIU, A. Logistic Regression for Determining Factors Influencing Students ' Perception of Course. v. 5, p. 99–106, 2016.

ALTHAUS, M. T. M.; GODOY, M. T. Lessons to pilot classes: interfaces between didactics and supervised training in biology. **Olhar de Professor**, v. 11, n. 2, p. 295–311, dez. 2008.

ANDRADE, M. L. F. DE; MASSABNI, V. G. O DESENVOLVIMENTO DE ATIVIDADES PRÁTICAS NA ESCOLA : UM DESAFIO PARA OS PROFESSORES DE CIÊNCIAS Practical activities development : a challenge to science teachers. **Ciência & Educação**, v. 17, n. 4, p. 835–854, 2011.

ANTUNES, C. **Avaliação da Aprendizagem Escolar**. 4º ed. ed. Rio de Janeiro: Vozes, 2002.

BATISTELA, G. C.; RODRIGUES, S. A.; BONONI, J. T. C. M. Estudo sobre a evasão escolar usando regressão logística : análise dos alunos do curso de Administração da Fundação Educacional de Ituverava STUDY ON THE TRUANCY USING LOGISTIC REGRESSION : ANALYSIS OF. n. October 2009, 2009.

BOFF, R. J. Administração do Tempo. **InterLink**, v. 1, n. 1, p. 59–74, 2010.

- CHUIEIRE, M. S. F. Concepções sobre a Avaliação Escolar. **Estudos em Avaliação Educacional**, v. 19, n. 39, p. 49, 30 abr. 2008.
- CORDEIRO, G. M. On Pearson's residuals in generalized linear models. **Statistics & Probability Letters**, v. 66, n. 3, p. 213–219, fev. 2004.
- DAL 'IGNA, M. C. Desempenho escolar de meninos e meninas: há diferença? Boys and girls' performances at school: is there any difference? p. 241–267, 2007.
- DE DIOS JIMÉNEZ, J.; SALAS-VELASCO, M. Modeling educational choices. A binomial logit model applied to the demand for Higher Education. **Higher Education**, v. 40, n. 3, p. 293–311, 2000.
- DE OLIVEIRA MENDONÇA, L. .; LOPES, C. E. . Mathematical Modeling: A learning environment for the implementation of Statistics Education in High School [Modelagem Matemática: um ambiente de aprendizagem para a implementação da Educação Estatística no Ensino Médio]. **Bolema - Mathematics Education Bulletin**, v. 24, n. 40, p. 701–724, 2011.
- EMILIANO, P. C.; VIVANCO, M. J. F.; DE MENEZES, F. S. Information criteria: How do they behave in different models? **Computational Statistics & Data Analysis**, v. 69, p. 141–153, jan. 2014.
- ERATLI ŞIRIN, Y.; ŞAHIN, M. Investigation of Factors Affecting the Achievement of University Students with Logistic Regression Analysis: School of Physical Education and Sport Example. **SAGE Open**, v. 10, n. 1, 2020.
- FAVÉRO, L. P. et al. **Análise de dados: modelagem multivariada para a tomada de decisões**. Rio de Janeiro: Elsevier, 2009.
- FERRÃO, M. E.; ALMEIDA, L. S. Multilevel modeling of persistence in higher education. **Ensaio**, v. 26, n. 100, p. 664–683, 2018.
- FORMIGA, N. S. Atribuição de causalidade e rendimento acadêmico: Predição e diferença em alunos da universidade pública e privada. **Revista de Psicologia da UNC**, v. 2, n. 1, p. 3–14, 2004.
- FREIRE, P. **Pedagogia da Autonomia: Saberes necessários à prática educativa**. 58. ed. São Paulo: Paz & Terra, 1997.
- GIANNINI, S. P. P.; LATORRE, M. DO R. D. DE O.; FERREIRA, L. P. Distúrbio de voz e estresse no trabalho docente: um estudo caso-controle. **Cadernos de Saúde Pública**, v. 28, n. 11, p. 2115–2124, nov. 2012.
- GIANNINI, S. P. P.; LATORRE, M. DO R. D. DE O.; FERREIRA, L. P. Distúrbio de voz relacionado ao trabalho docente: um estudo caso-controle. **Brazilian Journal of Otorhinolaryngology**, v. 3, n. 1, p. 129–134, 2013.
- GILBERTO, P. A. Influence and residuals in restricted generalized linear models. **Journal of Statistical Computation and Simulation**, v. 51, n. 2–4, p. 315–331, fev. 1995.
- GUNEY, Y. Exogenous and endogenous factors influencing students' performance in undergraduate accounting modules. **Accounting Education**, v. 18, n. 1, p. 51–73, 2009.
- HOSMER, D. W.; LEMESHOW, S.; STURDIVANT, R. X. **Applied logistic regression**. 3. ed. [s.l.] New York: Wiley, 2013.
- JACOBINI, O. R.; WODEWOTZKI, M. L. L. A Modelagem Matemática Aplicada no Ensino de Estatística em Cursos de Graduação. **Bolema - Boletim de Educação Matemática**, v. 14, n. 15, p. 47–68, 2015.
- JACOBINI, O. R.; WODEWOTZKI, M. L. L.; FERREIRA, D. H. L. Educação Estatística no Contexto da Educação. p. 473–494, 2011.
- KRAEMER, M. E. P. A AVALIAÇÃO DA APRENDIZAGEM COMO PROCESSO CONSTRUTIVO DE UM NOVO FAZER. **Avaliação: Revista da Avaliação da Educação Superior**, v. 10, n. 2, p. 137–147, 2005.
- LEITE, R. C. M. et al. Utilização de regressão logística simples na verificação da qualidade do ar

- atmosférico de Uberlândia. **Engenharia Sanitaria e Ambiental**, v. 16, n. 2, p. 175–180, jun. 2011.
- LIMA, N. DA C. M. et al. Associação do índice de atitudes e práticas pedagógicas ao desempenho dos estudantes na avaliação em larga escala do estado do Espírito Santo. **Educação em Revista**, v. 35, 2019.
- LISKA, G. R. et al. Machine learning based on extended generalized linear model applied in mixture experiments. **Communications in Statistics - Simulation and Computation**, p. 1–15, dez. 2019.
- MAGALHÃES, F. A. C.; ANDRADE, J. X. Exame vestibular, características demográficas e desempenho na na universidade: em busca de fatores preditivos. 2002.
- MARVIN S., D. Binary Logistic Predictive Model in Determining Students' Intention to Take Higher Education. **Asia Pacific Journal of Multidisciplinary Research**, Vol 5, Iss 4, Pp 138-143 (2017), v. 5, n. 4, p. 138, 2017.
- MAZULO, E. S. Análise da proficiência em matemática por meio de regressão linear múltipla. v. 10, p. 613–625, 2015.
- MCCULLOCH, C. E.; SEARLE, S. R.; NEUHAUS, J. M. **Generalized, Linear, and Mixed Models**. [s.l.] John Wiley Sonc Inc., 2009.
- MENEZES, F. S. et al. Data classification with binary response through the Boosting algorithm and logistic regression. **Expert Systems with Applications**, v. 69, p. 62–73, 2017.
- MORAL, R. A.; HINDE, J.; DEMÉTRIO, C. G. B. Half-Normal Plots and Overdispersed Models in R : The hnp Package . **Journal of Statistical Software**, v. 81, n. 10, nov. 2017.
- NIU, L. A review of the application of logistic regression in educational research: common issues, implications, and suggestions. **Educational Review**, v. 72, n. 1, p. 41–67, 2020.
- PENG, C. Y. J.; LEE, K. L.; INGERSOLL, G. M. An introduction to logistic regression analysis and reporting. **Journal of Educational Research**, v. 96, n. 1, p. 3–14, 2002.
- PYKE, S. W. .; SHERIDAN, P. M. . Logistic Regression Analysis of Graduate Student Retention. **The Canadian Journal of Higher Education**, v. XXIII, n. 2, p. 44–64, 1993.
- R CORE TEAM. **R: A language and environment for statistical computing**Vienna, Austria, 2017.
- RSTUDIO TEAM. **RStudio: Integrated Development for R**.BostonRStudio, Inc. , 2015.
- SIO JYH LIH, J.; ISMAIL, R. BIN. Binary Logistic Regression Analysis of Teacher Self-Efficacy Factors Influencing Literacy and Numeracy. **World Journal of Education**, v. 9, n. 1, p. 209, 2019.
- WISEU, F. et al. Concepções de Professores do Ensino Básico sobre a Prova Matemática: influência da experiência profissional. **Bolema: Boletim de Educação Matemática**, v. 31, n. 57, p. 430–453, abr. 2017.
- VIZZOTTO, P. A.; MACKEDANZ, L. F. Validação De Instrumento De Avaliação Da Alfabetização Científica Para Egressos Do Ensino Médio No Contexto Da Física Do Trânsito. **Educação em Revista**, v. 34, n. 0, 2018.
- WILD, C. J.; PFANNKUCH, M. Statistical Thinking in Empirical Enquiry. **International Statistical Review**, v. 67, n. 3, p. 223–248, dez. 1999.
- ZANON, C. et al. An application of item response theory to psychological test development. **Psicologia: Reflexao e Critica**, v. 29, n. 1, 2016.
- ZANON, D. P.; ALTHAUS, M. T. M.; BAGIO, V. A. **Didática na docência universitária em Saúde: metodologias ativas e avaliação**. 1. ed. Curitiba: Appris, 2018.
- ZEWUDE, B. T. Binary Logistic Regression Analysis in Assessment and Identifying Factors That Influence Students ' Academic Achievement : The Case of College of Natural and Computational. v. 7, n. 1, p. 1–6, 2016.