



Universidade de Brasília

Instituto de Exatas - IE

Departamento de Estatística - EST

Análise de regressão logística aplicada à educação online durante a pandemia da COVID-19

Aluna

Victória da Silva Lima

Orientador

Luís Gustavo do Amaral Vinha

Brasília, 28 de Outubro de 2020

Victória da Silva Lima

Análise de regressão logística aplicada à educação online durante a pandemia da COVID-19

Orientador Luís Gustavo do Amaral Vinha

Trabalho de Conclusão de Curso apresentado ao Departamento de Estatística, Instituto de Ciências Exatas, Universidade de Brasília, como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Brasília

2021

Agradecimentos

Aos professores de Estatística da Universidade de Brasília, por todos os ensinamentos e dedicação ao longo desse anos de graduação. Em especial ao professor Luís Gustavo do Amaral Vinha por todo o apoio, conselhos e atenção durante esse último ano. Agradeço também aos professores George Freitas von Borries e Maria Teresa Leão Costa por aceitarem fazer parte da banca e dedicarem tempo para avaliar e propor melhorias ao trabalho, além de todos os ensinamentos ministrados durante a graduação.

À minha família por todo o apoio essencial ao longo desses anos de graduação. Sem esse apoio essa jornada, sem dúvidas, seria muito mais difícil. Vocês foram essenciais para essa conquista.

À ESTAT e ao Movimento Empresa Júnior pelas conexões, aprendizados e experiências vividas ao longo dos 3 anos em que tive a oportunidade de me desenvolver em habilidades pessoais e profissionais, além da estatística.

À todas as amizades que cultivei durante esses anos de graduação. Agradeço todo o apoio, aprendizados e excelentes momentos que vivi com todos.

Resumo

A pandemia da COVID-19 impactou diretamente a educação em diversos países. Inicialmente, medidas de distanciamento social foram adotadas e, com isso, muitas escolas foram fechadas, sendo a educação online adotada de forma emergencial. Com base em dados de uma pesquisa global realizada com 10.433 participantes de mais de 115 países, foram ajustados 2 modelos de regressão logística a fim de entender quais fatores influenciam a preferência dos alunos pelo ensino online ou presencial. O modelo que levou em consideração apenas variáveis referentes à utilização de tecnologias e característica individuais indicou que alunos de pós graduação apresentavam maior chance de preferir o ensino online, enquanto os alunos de graduação apresentavam uma redução nessa chance se comparado com os alunos do ensino médio. No entanto, esse modelo não apresentou uma boa discriminação. Já no segundo modelo foi incluída a variável modo de ensino online preferido, que apresentou efeito significativo. Nesse caso a utilização do Zoom e Google Meets entre outras variáveis aumentaram as chances dos alunos que preferirem o ensino online em comparação com os alunos que não as utilizaram. Esse modelo apresentou um poder de discriminação superior ao primeiro.

Palavras-chave: Regressão logística. Educação. Educação online. COVID-19.

Lista de tabelas e quadros

Lista de Tabelas

1	Frequência dos países onde os respondentes estudam	20
2	Vivência em aulas online	21
3	Nível de estudo dos respondentes	22
4	Frequência do nível de estudo por preferência	24
5	Frequência do modo de ensino vivenciado relacionado à preferência . .	29
6	Frequência do modo de ensino preferido relacionado à preferência . . .	30
7	Associação com a preferência do ensino	35
8	Associação entre o uso da plataforma e a preferência do ensino	35
9	Associação entre o método utilizado e a preferência do ensino	36
10	Associação entre a técnica utilizada e a preferência do ensino	36
11	Modelo 1 de regressão logística	37
12	Tabela de classificação para o modelo 1	38
13	Modelo 2 de regressão logística	39
14	Tabela de classificação para o modelo 2	40

Lista de Quadros

1	Variáveis do banco de dados	10
2	Variáveis do banco de dados sobre avaliação do ensino online	11
3	Variáveis do banco de dados sobre avaliação do ensino presencial . . .	12
4	Medidas de qualidade	42

Lista de figuras

1	Curva do Modelo Logístico	13
2	Preferência por modelos de estudo	22
3	Nível de estudo dos alunos dos respondentes	23
4	Nível de estudo dos alunos	24
5	Técnica utilizada	25
6	Preferência de acordo com a técnica utilizada	25
7	Método utilizado	26
8	Preferência de acordo com o método utilizado	26
9	Plataforma utilizada	27
10	Preferência de acordo com a plataforma utilizada	27
11	Modo de ensino vivenciado	28
12	Modo de ensino preferido	28
13	Modo de ensino vivenciado relacionado a preferência	29
14	Modo de ensino preferido relacionado a preferência	30
15	Avaliação do ensino online	31
16	Avaliação do ensino online relacionado à preferência dos alunos	32
17	Avaliação do ensino presencial	33
18	Avaliação do ensino presencial relacionado a preferência dos alunos	34
19	Curva ROC para o modelo 1	38
20	Gráfico Deviance (azul) x Resíduos de Pearson para o modelo 2	39
21	Curva ROC para o modelo 2	41
22	Gráfico Deviance (azul) x Resíduos de Pearson para o modelo 2	41

Sumário

1	Introdução	8
2	Metodologia	10
2.1	Banco de dados	10
2.2	Regressão logística	12
2.2.1	Interpretação dos parâmetros	13
2.3	Regressão logística múltipla	13
2.3.1	Estimação dos coeficientes	14
2.3.2	Intervalo de confiança	15
2.4	Seleção de Variáveis	15
2.4.1	Critério da Informação de Akaike (AIC)	15
2.4.2	Critério Bayesiano de Schwarz (BIC)	15
2.4.3	Teste de Wald	16
2.4.4	Métodos automáticos de seleção	16
2.5	Diagnóstico	17
2.5.1	Avaliação do ajuste	17
2.5.2	Resíduos do modelo	17
2.5.3	Curva ROC e classificação	18
3	Resultados	20
3.1	Análise descritiva	20
3.1.1	Vivência	21
3.1.2	Preferência	21
3.1.3	Nível de estudo	22
3.1.4	Nível de estudo pela preferência	23
3.1.5	Técnica	24
3.1.6	Métodos	25
3.1.7	Plataformas	27
3.1.8	Modo de ensino	28
3.1.9	Modo de ensino pela preferência	29
3.1.10	Avaliação do ensino online	30

3.1.11	Avaliação do ensino online relacionado à preferência dos alunos	31
3.1.12	Avaliação do ensino presencial	32
3.1.13	Avaliação do ensino presencial relacionado à preferência dos alunos	33
3.2	Associações	34
3.3	Modelo 1	36
3.3.1	Diagnóstico do modelo	37
3.4	Modelo 2	39
3.4.1	Diagnóstico do modelo	40
3.5	Medidas de qualidade do ajuste	41
4	Conclusão	43
5	Referências	45

1 Introdução

A pandemia da Covid-19 impôs grandes desafios para praticamente todos os setores no Brasil, assim como para todos os países do mundo. Na tentativa de conter o avanço do novo Coronavírus, medidas de distanciamento social foram adotadas em diversos países. Devido à adoção de tais medidas, praticamente todos os setores da economia tiveram que se adaptar, de forma emergencial, a essa nova realidade. Comércio foram fechados, o trabalho remoto passou a ser adotado por muitas empresas, serviços de entrega se tornaram essenciais e o uso de aplicativos para realização de compras e outros serviços cresceram significativamente.

Na educação, tais medidas implicaram o fechamento de escolas e universidades, com a paralisação das aulas presenciais. Segundo nota técnica "Ensino à Distância na Educação Básica frente à pandemia da Covid-19", divulgada pela organização da sociedade civil Todos pela Educação, 91% do total de alunos do mundo e mais de 95% da América Latina ficaram temporariamente sem acesso presencial à escola (Todos pela educação, 2020). No Brasil, esse cenário não foi diferente. Por todo o país, redes públicas e privadas interromperam o funcionamento das escolas e transferiram aulas e outras atividades pedagógicas para modelos de educação online.

Prosseguir o processo educativo na modalidade de educação a distância demandou novas habilidades. Em um levantamento feita pela maior comunidade independente de startups do Brasil, a Distrito, em seu estudo anual (Edtech Report, 2020), Marcos Boscolo, Sócio líder no Brasil no setor de educação afirma que

"No cenário pré-pandemia, parte do setor de educação já caminhava fortemente na busca de soluções tecnológicas capazes de levar conteúdo pedagógico de qualidade aos seus alunos. O uso de tecnologia no ensino, além de tornar as aulas mais dinâmicas e atrativas, pode abranger, entre outros aspectos, acompanhar o processo pedagógico - classificando os erros e acertos do aluno - identificar lacunas de aprendizagem e, assim, planejar intervenções apropriadas e individualizar o ensino."

Quando se trata do futuro da educação, muitos pesquisadores e analistas de mercado acreditam cada vez mais que ele caminha rumo à educação online. Segundo uma matéria publicada na Forbes, projeções da Global Industry Analysts para o mercado de educação online, previam que, em 2015, o mercado alcançaria 107 bilhões de dólares e essa previsão estava certa. Com isso, previsões da Research and Markets apontam que o ensino online deve subir para 325 bilhões de dólares em 2025 (TJ McCue, 2018).

A migração do ensino presencial de forma emergencial durante a pandemia mudou completamente o ambiente educacional. Esse novo ambiente de aprendizagem se baseia na implementação de tecnologia da informação e comunicação, em que uma ampla gama de ferramentas assíncronas e síncronas são necessárias (Bdiwi et al. 2019). Isso torna o ambiente de ensino online e, conseqüentemente, a experiência dos alunos diferente do ambiente de ensino presencial. Com isso, dificuldades de utilização e acessos a certas ferramentas, além do contexto social em que esse ensino foi implementado, fez com que muitos alunos não tivessem uma boa experiência com esse novo formato. A experiência no ambiente educacional dos alunos pode estar relacionada a fatores como seu desenvolvimento acadêmico, satisfação e desempenho (Genn 2001; Shirazi et al. 2014).

Nesse contexto, o presente trabalho visa entender quais são os fatores que influenciam a preferência dos alunos pelo ensino presencial ao online durante esse período de educação à distância provocado pela pandemia da COVID-19. Dessa forma será possível identificar quais fatores podem prejudicar a experiência educacional dos alunos ou quais indicam um perfil de aluno que consegue se adequar melhor a essa realidade. Com isso, este trabalho busca oferecer subsídios para tomadas de decisões referentes às possíveis melhorias e novas formas de aprimorar a educação online caso essa tendência venha de fato a se consolidar, permanecendo mesmo após o término desse período em que as escolas permanecem fechadas.

2 Metodologia

2.1 Banco de dados

O banco de dados utilizado no presente estudo é resultado de uma pesquisa global sobre a experiência de alunos em aprendizagem remota de emergência durante a pandemia da COVID-19, com 10.433 alunos participantes de 115 países diferentes. Esses dados foram obtidos pela Geneva (Minerva Schools at Keck Graduate Institute), coletados no dia 23 de maio de 2020 e publicado no repositório de banco de dados da Universidade de Harvard. O banco de dados apresenta 28 variáveis, incluindo perguntas abertas, porém para esse estudo levaremos em consideração apenas as variáveis apresentadas nos Quadros 1, 2 e 3.

Quadro 1: Variáveis do banco de dados

Variável	Nome	Descrição
Q1	Data	Data em que a pesquisa foi respondida.
Q2	Nível de estudo	Indica o nível atual de estudo do respondente.
Q3	País	Indica o país onde o respondente estuda.
Q4	Vivência	Indica se o respondente teve cursos conduzidos de forma online nos últimos meses
Q7	Modo de ensino	Indica o formato das aulas vivenciadas durante esse período, podendo ser ao vivo, gravado, por e-mail ou por fóruns.
Q8	Modo preferido	Indica qual dos modos listados na questão Q7 o respondente prefere.
Q10	Plataforma	Pergunta de múltipla escolha, com mais de uma opção de resposta, na qual os respondentes devem selecionar todas as plataformas utilizadas, estando entre as opções Teams, Google classroom, entre outras.
Q11	Métodos	Pergunta de múltipla escolha, com mais de uma opção de resposta, na qual os respondentes devem selecionar todos os métodos utilizados, estando entre as opções aulas, palestra, questionários, entre outros.

Q12	Técnicas	Pergunta de múltipla escolha, com mais de uma opção de resposta, na qual os respondentes devem selecionar todas as técnicas utilizados no período de estudo fora das aulas, estando entre as opções material de estudo, chats, entre outras opções.
Q27	Preferência	Pergunta de múltipla escolha binária sobre qual a preferência dos alunos sendo 0- ensino online 1- ensino presencial

A fim de estudar com mais profundidade cada plataforma, método, técnica, as respostas a esses questionamentos foram transformadas em novas variáveis dummy sendo 1 quando o aluno havia utilizado aquela plataforma, método ou técnica e 0 caso contrário.

Visando avaliar a experiência do ensino online, foi solicitado aos respondentes que, considerando sua experiência de aprendizado remoto, indicasse o quanto concorda, é neutro ou discorda das afirmações listadas no Quadro 2.

Quadro 2: Variáveis do banco de dados sobre avaliação do ensino online

Variável	Nome	Afirmação
Q13	Simpatia	Eu gosto de ter aulas online.
Q14	Motivação	Me sinto motivado para aprender.
Q15	Satisfação	Estou satisfeito com o ensino online.
Q16	Envolvimento	Minhas aulas são envolventes.
Q17	Distração	Costumo me distrair quando faço o trabalho do curso ou quando participo das aulas.
Q18	Participação	Eu costumo fazer perguntas, comentários ou participar de discussões (ao vivo, no chat ou por e-mail).

Com o objetivo de avaliar a experiência do ensino presencial, partindo do pressuposto que todos os respondentes já o havia vivenciado, foi solicitado aos respondentes que, considerando a sua experiência prévia de aprendizagem presencial, indicasse o quanto concorda, é neutro ou discorda das afirmações apresentadas no Quadro 3.

Quadro 3: Variáveis do banco de dados sobre avaliação do ensino presencial

Variável	Nome	Afirmação
Q20	Simpatia	Eu gosto de ter aulas presenciais.
Q21	Motivação	Me sinto motivado para aprender.
Q22	Satisfação	Estou satisfeito o ensino presencial.
Q23	Envolvimento	Minhas aulas são envolventes.
Q24	Distração	Costumo me distrair quando faço o trabalho do curso ou quando participo das aulas.
Q25	Participação	Eu costumo fazer perguntas, comentários ou participar de discussões.

2.2 Regressão logística

A análise de regressão é uma metodologia estatística que utiliza a relação entre duas ou mais variáveis para que uma variável resposta possa ser predita ou explicada a partir de outras. Em uma aplicação específica de um modelo de regressão, a variável resposta apresenta apenas dois possíveis resultados qualitativos. Portanto, podem ser representados por uma variável indicadora binária, correspondendo aos valores 0 e 1. Na análise desse modelo, o valor 1 representa o "sucesso", sendo esse o interesse. No presente estudo, aluno preferir o ensino online será considerado o "sucesso". Nesse caso, é necessário um ajuste do modelo de regressão logística.

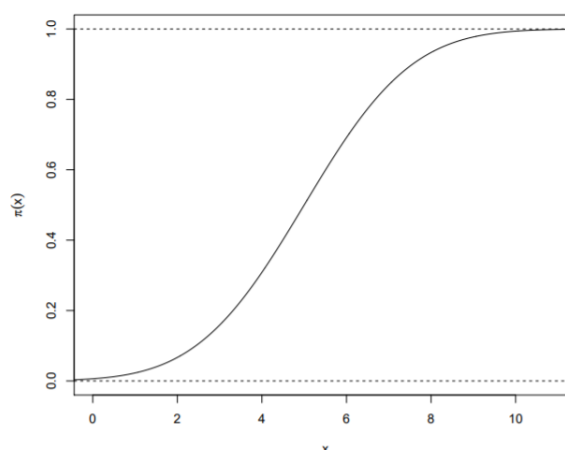
Suponha que haja uma única variável explicativa X , que é quantitativa. Para uma variável de resposta binária y_i , $\pi(x)$ denota a probabilidade de "sucesso" para um dado valor x . Essa probabilidade é o parâmetro para a distribuição binomial.

O modelo de regressão logística tem forma linear para a função logística, denotada também por logit, onde

$$\text{logit}[g(x)] = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \alpha + \beta x \quad (1),$$

sendo α e β parâmetros desconhecidos. A partir da equação (1) pode-se definir a probabilidade como $\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}$ (2). De modo a exemplificar, tem-se a Figura 1 que representa a curva do modelo logístico.

Figura 1: Curva do Modelo Logístico



2.2.1 Interpretação dos parâmetros

Uma interpretação importante do modelo de regressão logística é a razão de chances. A chance é uma medida de associação que indica a razão entre a chance de sucesso e fracasso, de forma que a probabilidade de sucesso π pode ser definida como $\frac{\pi}{1-\pi}$. Assim, para o modelo (1), as chances de sucesso são $\frac{\pi(x)}{1-\pi(x)} = \exp(\alpha + \beta x) = \epsilon^\alpha (\epsilon^\beta)^x$.

Já a razão de chance é a razão entre a chance do evento de interesse do estudo acontecer em um grupo dividido pela chance acontecer em um outro grupo. Essa razão é dada por

$$\frac{\epsilon^{\alpha+\beta_1(x+1)}}{\epsilon^{\alpha+\beta_1}} = \epsilon^{\beta_1}.$$

Esta relação exponencial fornece uma interpretação para β . As probabilidades são multiplicadas por ϵ^β para cada aumento de 1 unidade em x. Ou seja, as probabilidades no nível $x + 1$ são iguais às probabilidades em x multiplicado por ϵ^β . Quando $\beta = 0$ e $\epsilon^\beta = 1$ as chances não mudam com a mudança de x.

2.3 Regressão logística múltipla

O modelo de regressão logística explicado por apenas uma variável explicativa é categorizado como modelo de regressão logística simples. Porém, em muitos casos, são necessárias um número maior de variáveis explicativas para se obter uma descrição adequada e, nesse caso, consolida-se o modelo de regressão logística múltipla.

Com base no objetivo do presente estudo, é possível constatar que serão analisadas mais de uma variável explicativa, visando melhor compreender a variável resposta.

Para o caso de regressão logística múltipla, o modelo apresentado anteriormente para o caso simples será generalizado. Considera-se um conjunto de p variáveis independentes denotadas por $\mathbf{X}' = (x_1, \dots, x_p)$. Denota-se por $\boldsymbol{\beta} = (\alpha, \beta_1, \dots, \beta_p)^\tau$ o vetor de parâmetros desconhecidos. Dessa forma, o modelo considerando p variáveis explicativas é dado por

$$\pi(\mathbf{x}) = \frac{\exp(\alpha + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\alpha + \beta_1 x_1 + \dots + \beta_p x_p)}.$$

Considerando n observações da variável resposta y_i associadas aos valores de \mathbf{x}_i para $i \in (1, \dots, n)$, o logit, já apresentado na equação para a regressão logística simples, nesse caso será dado pela equação $\text{logit}[\pi(\mathbf{x}_i)] = \alpha + \beta_1 x_1 + \dots + \beta_p x_p$.

2.3.1 Estimação dos coeficientes

Na busca por estimar os parâmetros do modelo, pode-se utilizar o método de máxima verossimilhança. Sendo $\boldsymbol{\beta}$ o vetor de parâmetros desconhecidos relacionados à probabilidade condicional $P(Y_i = 1 | \mathbf{x}_i) = \pi(\mathbf{x}_i)$, a função de verossimilhança é dada por $L(\boldsymbol{\beta}) = \prod_{i=1}^n \pi(\mathbf{x}_i)^{y_i} [1 - \pi(\mathbf{x}_i)]^{(1-y_i)}$. Com $y_i \in (0, 1)$ o logaritmo da função de verossimilhança é dado por

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n [y_i \ln \pi(\mathbf{x}_i) + (1 - y_i) \ln(1 - \pi(\mathbf{x}_i))].$$

Para encontrar o valor de $\hat{\boldsymbol{\beta}}$ que maximiza a função é necessário derivar a função $l(\boldsymbol{\beta})$ em relação a cada parâmetro e resolver o sistema de equações.

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \alpha} = \sum_{i=1}^n y_i - \sum_{i=1}^n \pi(\mathbf{x}_i) = 0$$

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^n x_{ij} y_i - \sum_{i=1}^n x_{ij} \pi(\mathbf{x}_i) = 0$$

2.3.2 Intervalo de confiança

Para grandes amostras, o intervalo de confiança de Wald para o parâmetro β do modelo de regressão logística $\text{logit}[\pi(x)] = \alpha + \beta(x)$ é dado por $[\hat{\beta} \pm z_{\alpha/2}(SE)]$, sendo SE o erro padrão de $\hat{\beta}$.

2.4 Seleção de Variáveis

Para a construção de um modelo que busca explicar uma variável resposta, um importante passo é a escolha das variáveis. Para isso, é realizado o procedimento de seleção de variáveis. O objetivo é escolher um grupo menor de variáveis que garantem a melhor qualidade do modelo de acordo com alguns critérios.

2.4.1 Critério da Informação de Akaike (AIC)

Uma das formas de comparar n modelos é utilizando a estimativa para a log verossimilhança esperada. Assim, comparar magnitudes da função suporte maximizada $L(\hat{\theta}_i)$ é um critério para avaliar os modelos. Porém, essa forma de comparação apresenta um viés cuja a magnitude varia de acordo com a dimensão do vetor de parâmetros. Segundo Akaike, esse viés é dado assintoticamente pelo número de parâmetros a serem estimados no modelo, definindo seu critério de informação como

$$AIC = -2\log L + 2(k + s),$$

sendo k é o número de níveis de resposta menos 1 e s é o número de variáveis explicativas. No modelo com resposta binária $k + s$ é o número de parâmetros no modelo.

2.4.2 Critério Bayesiano de Schwarz (BIC)

Assim como o Critério da Informação de Akaike, o Critério Bayesiano de Schwarz também oferece penalidade para a adição de variáveis explicativas no modelo. Nesse caso, o critério foi definido por

$$BIC = -2\log L + (k + s)\log(n),$$

sendo n o número de observações.

2.4.3 Teste de Wald

Em um modelo de regressão logística é importante verificar se a probabilidade de sucesso independe da variável explicativa. Para isso, pode-se utilizar o teste de Wald, com hipótese nula $H_0 : \beta = 0$, afirmando que a probabilidade de sucesso é independente. A estatística do teste é dada por $z = \frac{\hat{\beta}}{ASE(\hat{\beta})}$, com uma distribuição normal padrão quando $\beta = 0$ ou $W = z^2 = \frac{\hat{\beta}^2}{\widehat{Var}(\hat{\beta})}$, com distribuição qui-quadrado. Esse teste também pode ser aplicado em um modelo de regressão logística múltipla para avaliar se as variáveis são estatisticamente significativas no modelo.

2.4.4 Métodos automáticos de seleção

Com o objetivo de facilitar e automatizar o processo de escolha das variáveis preditoras que farão parte do modelo e visando selecionar o modelo que melhor explica a variável resposta, pode-se utilizar os métodos automáticos.

- *Forward* O método *Forward* se baseia na estatística F e seu respectivo p-valor para selecionar a variável mais significativa. Dessa forma, partindo do modelo simples com apenas uma variável, cada nova variável que se apresenta significativa vai sendo incluída no modelo e a estatística F parcial é recalculada para analisar o efeito da entrada de cada nova variável no modelo. O processo continua de forma progressiva enquanto existirem variáveis significativas e enquanto o efeito da entrada dessas variáveis no modelo realmente estiver auxiliando na explicação da variável resposta. Caso contrário, o processo é encerrado e assim é apresentado o melhor modelo de acordo com esse método.
- *Stepwise* método similar ao método de Seleção *Forward*, uma vez que inicialmente também é ajustado um modelo mais simples, com apenas uma variável e a partir disso são calculados o F parcial e seus respectivos p-valores para determinar quais variáveis deverão entrar no modelo. A diferença em relação ao método de Seleção *Forward* que, nesse método, é incluído um teste para verificar se a variável incluída no primeiro passo deve permanecer ou ser retirada do modelo com base também no teste F. Assim, para cada variável que entra no modelo é verificado novamente se as variáveis que já foram introduzidas ao modelo devem permanecer ou sair. O método prossegue até que se alcance

o modelo final onde não há mais variáveis para serem inseridas e todas as variáveis que estão no modelo não apresentam nenhum indicativo que devam ser retiradas, conforme a estatística F.

2.5 Diagnóstico

2.5.1 Avaliação do ajuste

Uma maneira de detectar a falta de ajuste é realizar um teste de razão de verossimilhança para comparar o modelo obtido com um modelo mais complexo. Com isso, busca-se avaliar se as variáveis incluídas no modelo mais complexo são de fato significativas, quando comparado com o modelo restrito.

Com hipótese nula $H_0 : \beta = 0$, a estatística da razão de verossimilhança para esse teste é o desvio do modelo. Considerando l_0 a função de verossimilhança do modelo reduzido e l_1 função de verossimilhança do modelo mais complexo, tem-se que

$$G^2(M) = -2\log(l_0/l_1).$$

Para o caso em que se utiliza a tabela de contingência, para todas as células dessa tabela, a estatística Pearson corresponde a

$$X^2(M) = \sum (\text{observado} - \text{ajustado})^2 / \text{ajustado}.$$

Outra forma de avaliar o ajuste do modelo é pelo teste de Hosmer e Lemeshow. Esse teste se baseia em agrupar as observações em grupos e depois comparar as frequências preditas com as observadas. Esse teste parte da hipótese nula que, a um nível de significância α , não existe diferença entre os valores preditos e observados, ou seja, o modelo está bem ajustado. A estatística do teste é a mesma do teste Qui quadrado de Pearson com $c - 2$ graus de liberdade.

2.5.2 Resíduos do modelo

No caso de um modelo de regressão logística, o valor da variável resposta Y dado X é dado por $y = \pi(x) + \epsilon$. Neste caso, y_i corresponde ao número de “sucessos” para n_i das i variáveis explicativas. Uma vez que $\hat{\pi}_i$ denota a probabilidade estimada

de sucesso para o ajuste do modelo, os resíduos assumem dois valores possíveis $\epsilon_i = 1 - \hat{\pi}_i$ quando $y_i = 1$ ou $\epsilon_i = -\hat{\pi}_i$ quando $y_i = 0$. Dessa forma, ϵ se aproxima de uma distribuição binomial com média zero e variância $\pi(x)[1 - \pi(x)]$ (HOSMER e LEMESHOW, 2000).

Sabendo que a média binomial estimada $n_i \hat{\pi}_i$ é o número ajustado de sucessos, para um modelo linear generalizado com componente aleatório binomial, o resíduo de Pearson, comparando y_i com seu valor ajustado, é dado por

$$r_{pj} = \frac{y_i - n_i \hat{\pi}_i}{\sqrt{n_i \hat{\pi}_i (1 - \hat{\pi}_i)}}.$$

Assim, r_{pj} divide a diferença entre uma contagem observada e seu valor ajustado pelo desvio padrão binomial estimado a partir da contagem observada.

2.5.3 Curva ROC e classificação

A fim de verificar se as observações foram classificadas corretamente de acordo com o modelo proposto, pode-se resumir essa informação em uma tabela de classificação. Com base em um ponto de corte, é possível avaliar a qualidade do modelo verificando seu percentual de acerto.

A curva ROC, ou seja, a curva de característica de operação do receptor é um gráfico de sensibilidade por 1 - especificidade, sendo um complementar do outro. Dessa forma, a curva ROC é mais informativa do que a tabela de classificação, uma vez que resume o poder preditivo para todos os possíveis π_0 (AGRESTI, 2007).

A curva ROC produz a área abaixo da curva, conhecida como AUC, que serve como uma medida de discriminação do modelo. Para uma determinada sensibilidade, a maior especificidade apresenta o melhor poder preditivo. Dessa forma, quanto maior a AUC, maior é o poder de classificação do modelo. Segundo HOSMER e LEMESHOW, um modelo pode ter sua discriminação classificada da seguinte forma (HOSMER e LEMESHOW, 2000)

- AUC = 0,5 não há discriminação;
- 0,7 < AUC < 0,8 discriminação é aceitável;
- 0,8 < AUC < 0,9 discriminação é excelente;

- AUC > 0,9 discriminação é excepcional.

3 Resultados

Nessa seção serão apresentados os resultados da análise dos dados selecionados para o trabalho.

3.1 Análise descritiva

Como apresentado anteriormente no Quadro 1, foi questionado aos respondentes o país em que os mesmos estudavam durante o período da pesquisa (maio de 2020).

Tabela 1: Frequência dos países onde os respondentes estudam

Países	Absoluta	Relativa
Reino Unido	5065	48,55%
Alemanha	554	5,31%
Estados Unidos	509	4,88%
Países Baixos	479	4,59%
Irlanda	350	3,35%
Austrália	333	3,19%
Itália	218	2,09%
Canadá	211	2,02%
Espanha	198	1,90%
França	196	1,88%
Bélgica	181	1,73%
Portugal	145	1,39%
Índia	108	1,04%
Polônia	100	0,96%
Nova Zelândia	89	0,85%
Áustria	77	0,74%
África do Sul	73	0,70%
Suíça	68	0,65%
Suécia	63	0,60%
Cingapura	60	0,58%
Emirados Árabes	60	0,58%
Peru	60	0,58%
Noruega	58	0,56%
Dinamarca	52	0,50%
República Checa	51	0,49%

Com base na Tabela 1, é possível notar que, apesar da grande variabilidade de países presentes na pesquisa, sendo no total 115 países representados, 48,55% dos respondentes estudavam no Reino Unido. Isso indica que a realidade educacional e

o combate à pandemia no Reino Unido poderão ter um grande impacto nas respostas subsequentes. Vale ressaltar que, com o objetivo de simplificar a visualização, apenas os 25 países com maior número de respondentes foram representados na Tabela 1.

3.1.1 Vivência

A pandemia da COVID-19 progrediu de maneira muito distinta em cada país. Como os dados se referem a uma pesquisa global, foi questionado aos respondentes se os mesmos tiveram aulas ministradas de forma online durante o período da pandemia. Em alguns países, essa migração para o ensino online não foi necessária ou, mesmo quando necessária, muitas vezes não foi possível de ser implementada. Esse fato será importante para avaliações que serão apresentadas em sequência uma vez que a falta de vivência no ensino online pode influenciá-las.

Tabela 2: Vivência em aulas online

Vivência online	Absoluta	Relativa
Sim	9.465	92,45%
Não	472	4,25%
Não responderam	316	3,05%
Total	10.433	100%

Ao considerar a Tabela 2, é possível notar que a grande maioria dos respondentes migraram para o ensino online durante a pandemia da COVID-19, representando 92,45% dos respondentes.

3.1.2 Preferência

Uma das perguntas mais importantes desse estudo se refere ao questionamento sobre a preferência quanto ao modelo de ensino. A pergunta foi contextualizada em um cenário onde seria possível a escolha entre o ensino online e ensino presencial, mesmo que em alguns momentos durante a pandemia, o formato presencial não fosse uma opção em diversos países. Assim, essa posição não implica necessariamente em um posicionamento sobre a abertura das escolas durante o período da pandemia.



Figura 2: Preferência por modelos de estudo

Baseado na Figura 2, é possível notar que 84,93% dos alunos indicaram a preferência pelo ensino presencial em relação ao online. Para todo o estudo, esses dados serão utilizados para avaliar quais fatores podem influenciar essa preferência.

3.1.3 Nível de estudo

Para compreender melhor o contexto ao qual os alunos reponderam à pesquisa, foi questionado o nível de estudo atual dos alunos. Como essa é uma pesquisa global, alguns ajustes foram necessários para garantir a equivalência dos níveis entre esses países.

Tabela 3: Nível de estudo dos respondentes

Nível de estudo	Absoluta	Relativa
Ensino Médio	4.973	47,66%
Graduação	4.843	46,42%
Pós Graduação	481	4,61%
Pré Universidade	115	1,10%
Ensino Técnico	16	0,15%
Ensino Fundamental	3	0,03%
Desenvolvimento profissional contínuo	2	0,01%
Total	10.433	100%

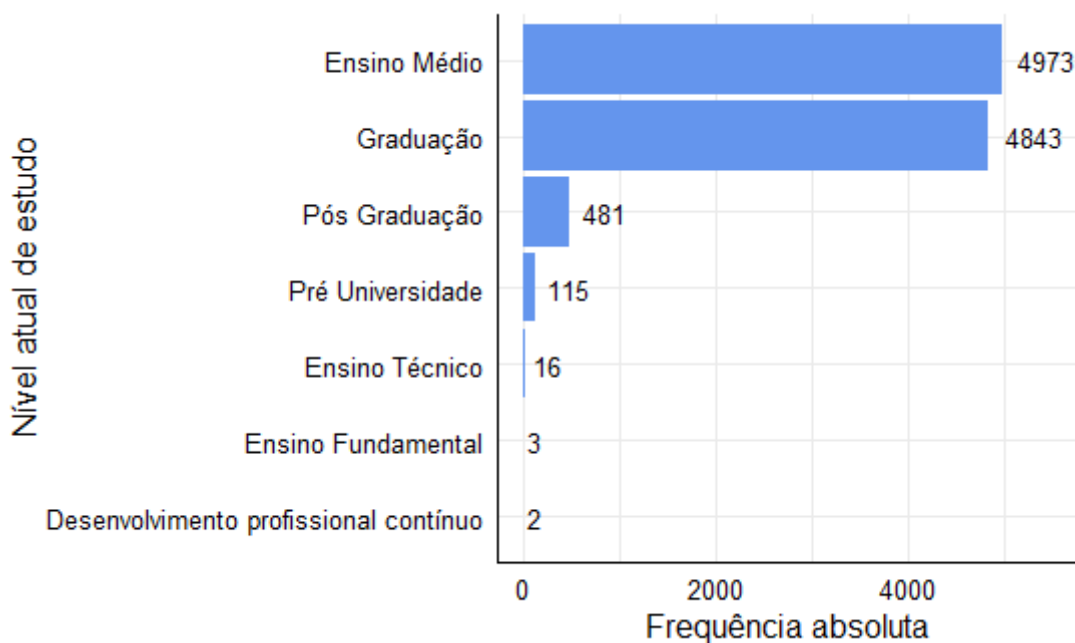


Figura 3: Nível de estudo dos alunos dos respondentes

Como é possível observar na Tabela 2 e Figura 4, a pesquisa se concentra em alunos do Ensino Médio, que representam 47,66% dos respondentes e alunos da Graduação, que representam 46,42%. No total, esses dois níveis de estudo contabilizam 94,08% de todos os respondentes.

3.1.4 Nível de estudo pela preferência

Com base na Tabela 4 e Figura 4 é possível notar que, de uma forma exploratória, não há fortes indícios que sugiram que o nível de estudo possa influenciar a preferência dos alunos. Vale ressaltar que o número de alunos que preferiam o ensino online ou presencial diferem de forma expressiva. Dessa forma, o gráfico apresenta, para cada grupo, a frequência relativa.

Tabela 4: Frequência do nível de estudo por preferência

Nível de estudo	Online		Presencial	
	N	Relativa	N	Relativa
Ensino Médio	769	48,92%	4.204	47,44%
Graduação	678	43,12%	4.165	47,00%
Pós Graduação	95	6,04%	386	4,36%
Pré Universidade	27	1,71%	88	0,99%
Ensino Técnico	3	0,18%	13	0,15%
Ensino Fundamental	0	0,00%	3	0,03%
Desenvolvimento profissional contínuo	0	0,00%	2	0,02%
Total	1.572	100%	8.861	100%

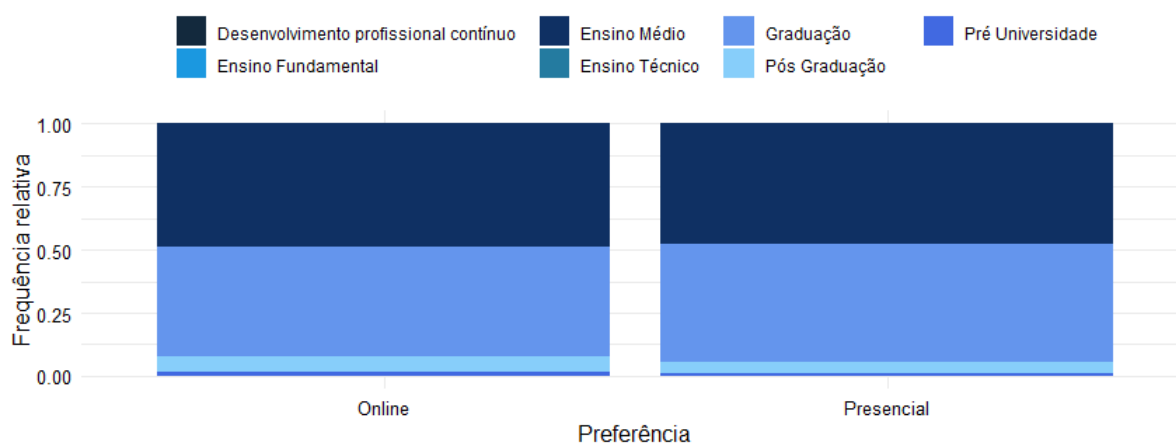


Figura 4: Nível de estudo dos alunos

3.1.5 Técnica

A partir da Figura 5 e 6 apresentadas, pode-se constatar que as técnicas mais utilizadas foram avaliações supervisionadas e vídeos. É possível notar que não há uma variação muito expressiva em relação à preferência dos alunos pelo ensino online ou presencial com base na técnica utilizada. Vale destacar que nessa pergunta, os entrevistados podiam selecionar todas as técnicas que se aplicavam à realidade vivida.

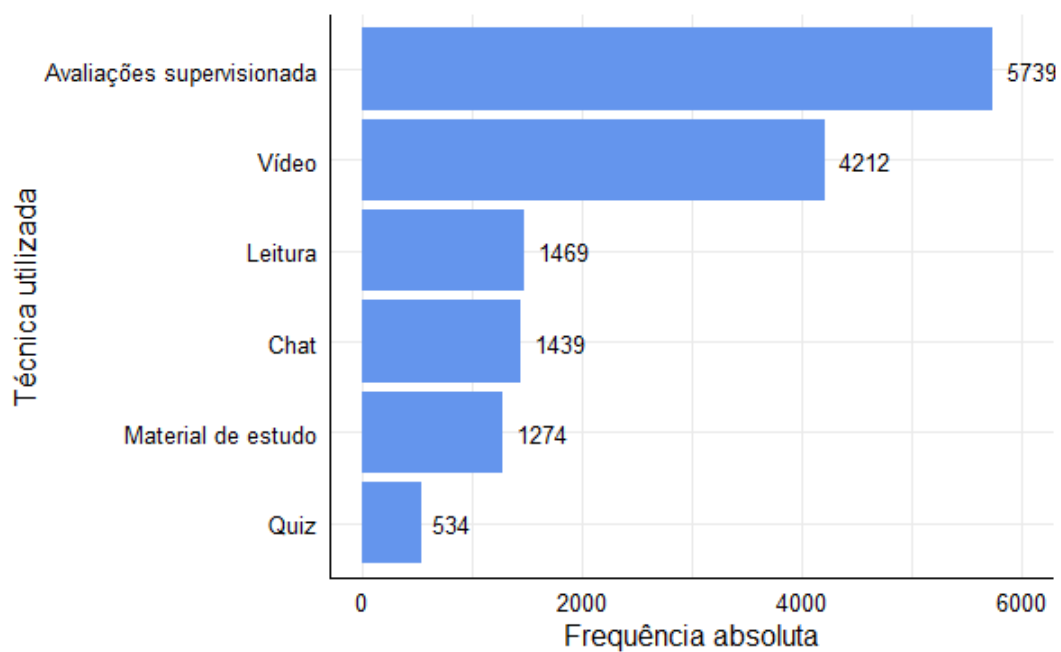


Figura 5: Técnica utilizada

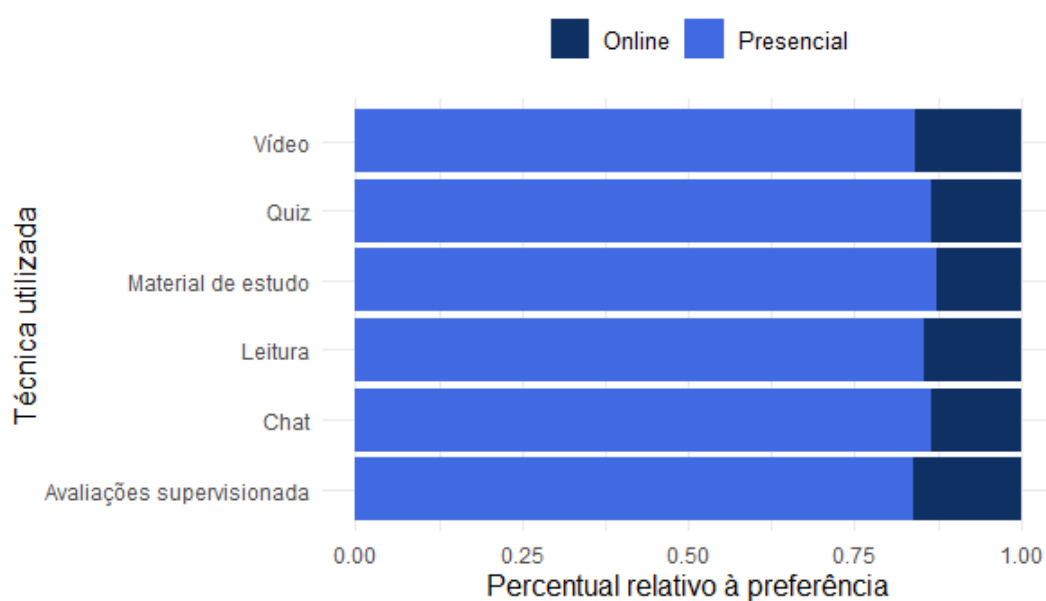


Figura 6: Preferência de acordo com a técnica utilizada

3.1.6 Métodos

Foi requisitado que os entrevistados selecionassem todas os métodos eles haviam vivenciados durante as aulas síncronas online. Dessa forma, os alunos que não vivenciaram aulas síncronas online durante esse período foram adicionados à categoria

”sem ao vivo”. A Figura 7 aponta que palestra, bate papo e quiz foram os métodos mais utilizados durante esse período.

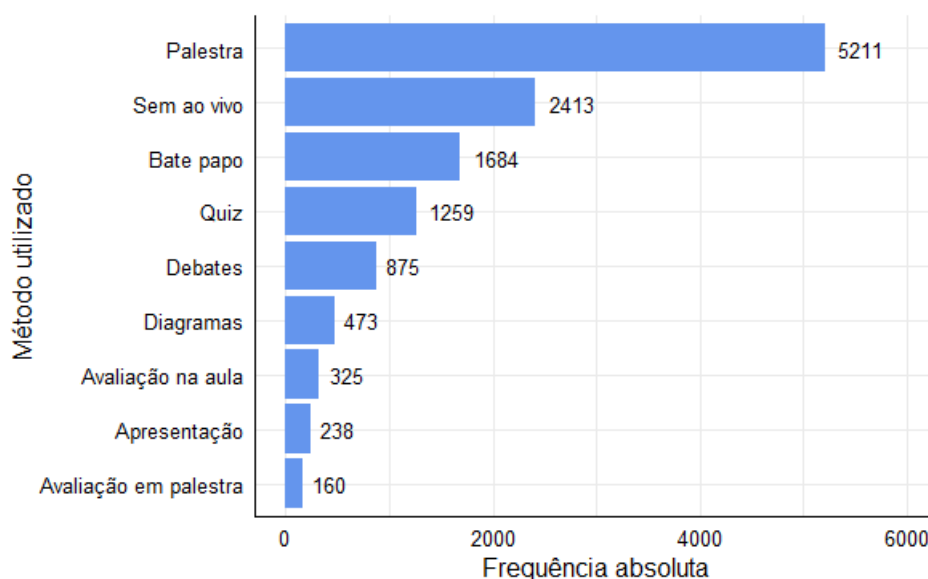


Figura 7: Método utilizado

A Figura 8 indica que não existiu grandes variações no percentual de preferência do ensino online ao presencial com base nos métodos de estudos utilizados. O Método que apresentou maior percentual relativo de preferência pelo ensino online ao presencial foi a utilização de apresentações durante as aulas síncronas.

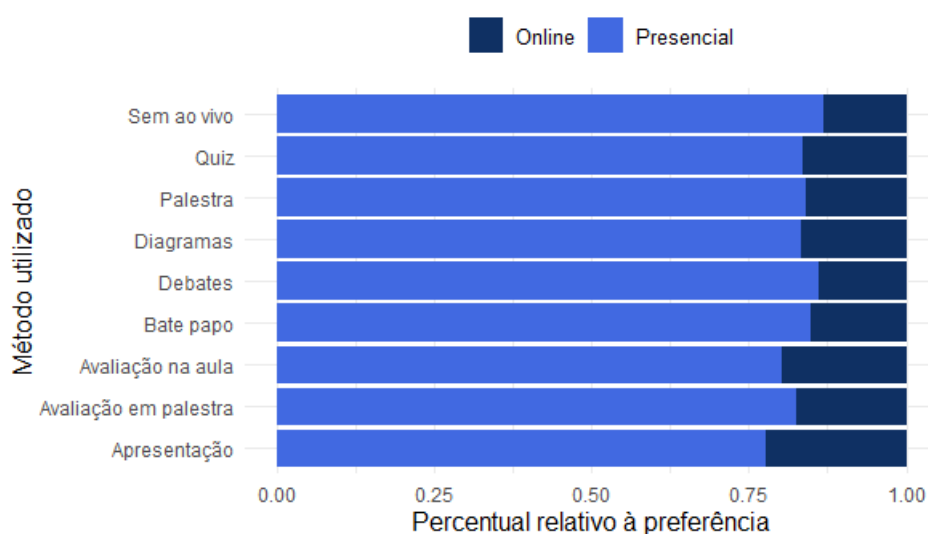


Figura 8: Preferência de acordo com o método utilizado

3.1.7 Plataformas

A Figura 9 mostra que Zoom, E-mail e Microsoft Teams foram as plataformas mais utilizadas pelos alunos durante esse período. Nessa pergunta os alunos deveriam selecionar as duas plataformas mais utilizadas.

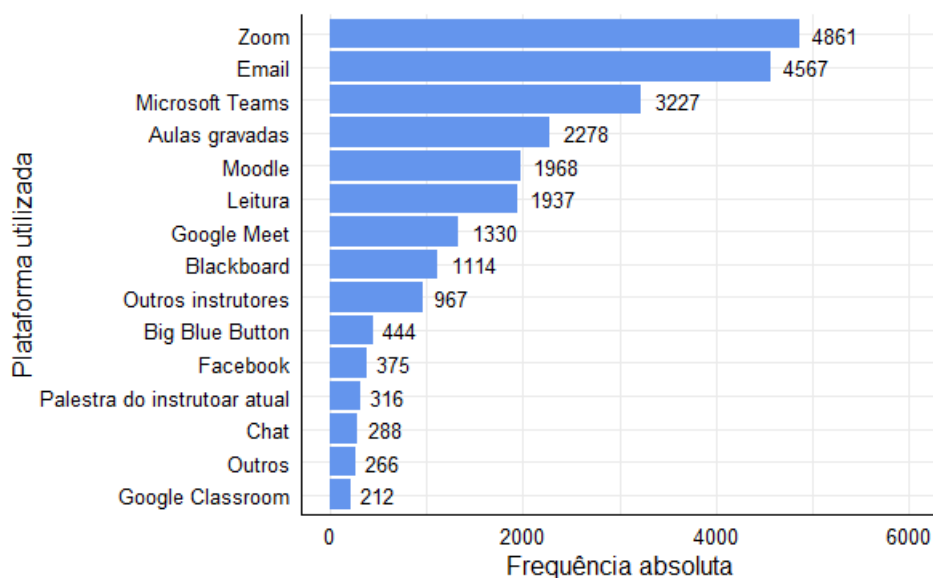


Figura 9: Plataforma utilizada

Como é possível observar na Figura 10, não houve uma diferença muito expressiva na preferência dos alunos com base na plataforma utilizada.

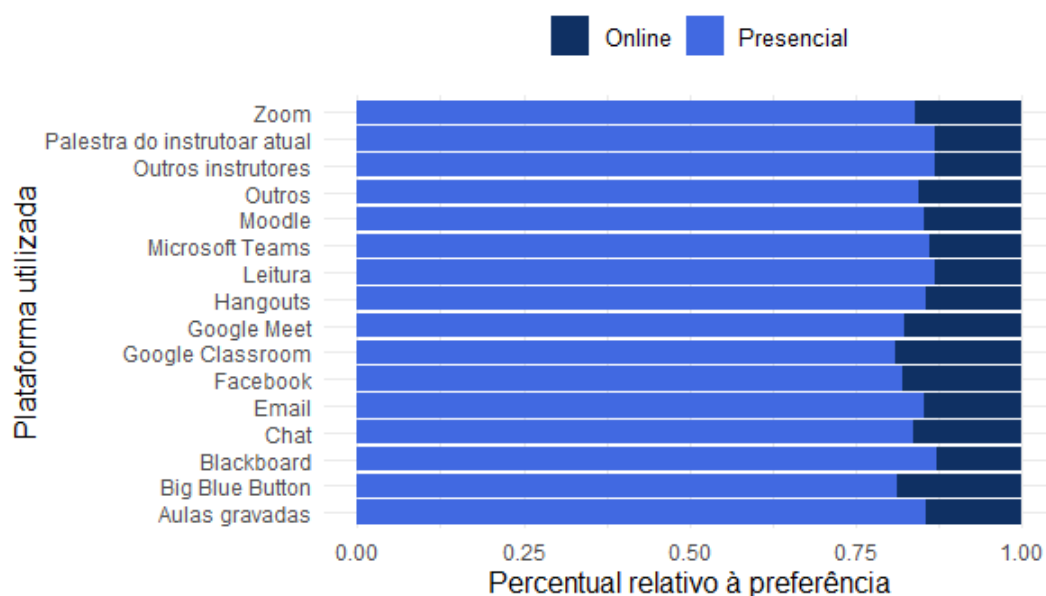


Figura 10: Preferência de acordo com a plataforma utilizada

3.1.8 Modo de ensino

Com base na Figura 11, é possível notar que 4.271 (40,94%) alunos afirmaram ter vivenciado experiências de aulas ao vivo, ou seja, aulas síncronas, além do fato de que 3.840 (36,81%) tiveram materiais enviados por e-mail durante esse período.

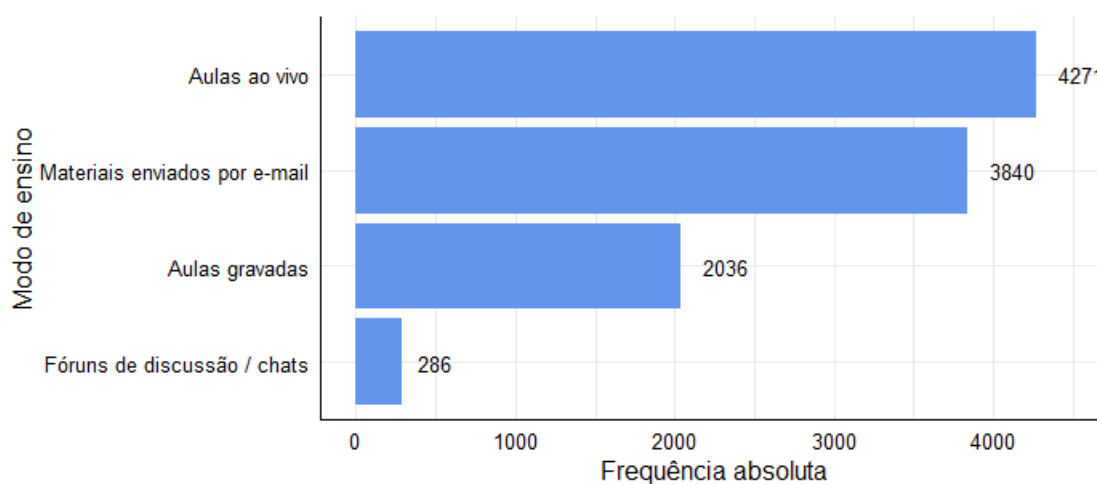


Figura 11: Modo de ensino vivenciado

Por outro lado, quando questionados sobre o modo de ensino preferencial, o percentual de alunos que afirmaram a preferência por aulas ao vivo foi ainda maior, chegando a 53,67%. Além disso, o segundo modo de ensino preferido foi distinto do vivenciado. Nesse caso, 3.063 (29,36%) respondentes disseram preferir aulas gravadas ou vídeos, mesmo que apenas 2.036 alunos tenham vivenciado esse modo de ensino durante esse período.

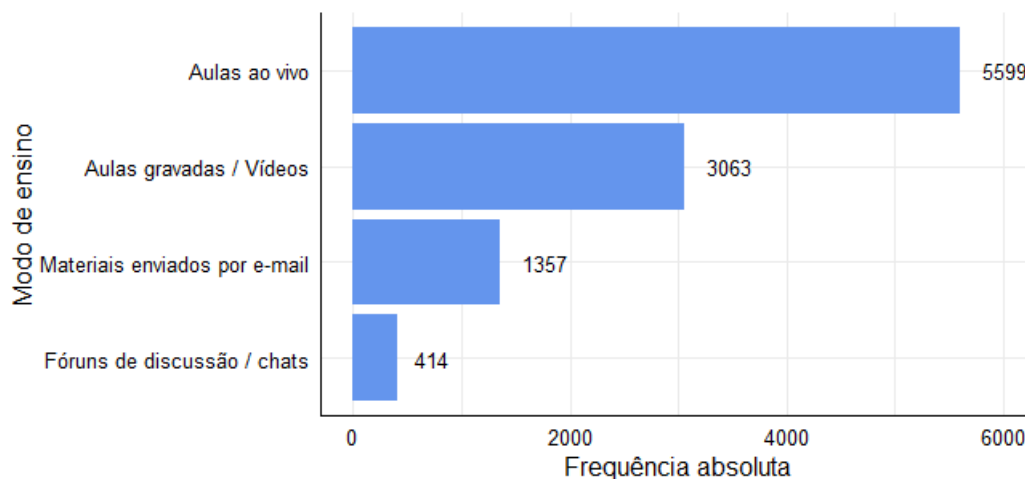


Figura 12: Modo de ensino preferido

3.1.9 Modo de ensino pela preferência

Com base na Tabela 5 e na Figura 13, pode-se observar que, de forma descritiva, não há uma grande diferença entre o modo de ensino vivenciado pelos respondentes que preferem o online e os que preferem o presencial.

Tabela 5: Frequência do modo de ensino vivenciado relacionado à preferência

Nível de estudo	Online		Presencial	
	N	Relativa	N	Relativa
Aulas ao vivo	665	42,30%	3.606	40,69%
Materiais enviados por e-mail	584	37,15%	3.256	36,74%
Aulas gravadas/Vídeos	288	18,32%	1.748	19,73%
Fóruns de discussão/chats	35	2,22%	251	2,83%
Total	1.572	100%	8.861	100%

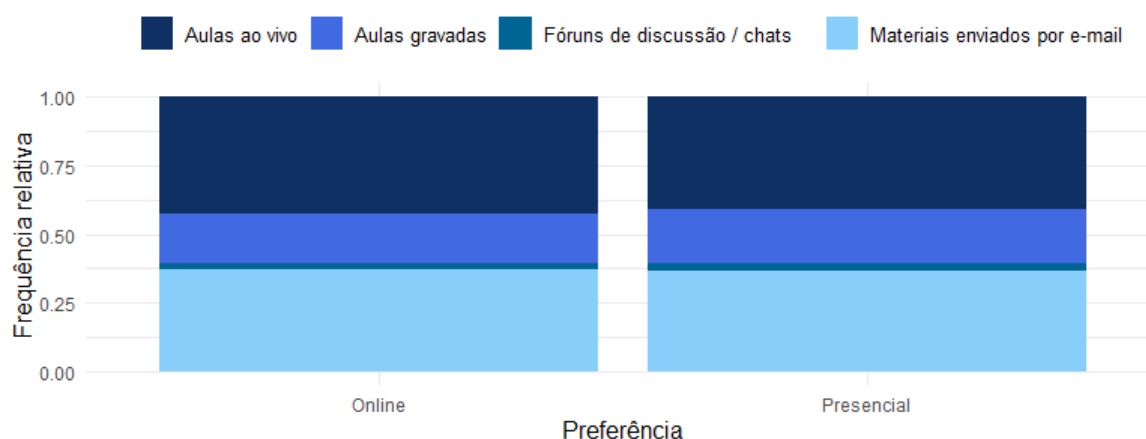


Figura 13: Modo de ensino vivenciado relacionado a preferência

Por outro lado, ao analisar a Tabela 6 e Figura 14, é possível perceber que, quando se observa o modo preferido de ensino, 38,99% dos alunos que preferem o ensino online indicaram preferir aulas gravadas ou vídeo, um percentual superior ao visto outro grupo. Já no caso dos alunos que demonstraram preferência pelo ensino presencial, a escolha por aulas ao vivo chegou a 57,08% dos alunos, contra um percentual de 34,48% no caso dos alunos que preferem o ensino online.

Tabela 6: Frequência do modo de ensino preferido relacionado à preferência

Nível de estudo	Online		Presencial	
	N	Relativa	N	Relativa
Aulas ao vivo	542	34,48%	5.057	57,08%
Aulas gravadas/Vídeos	613	38,99%	2.450	27,65%
Materiais enviados por e-mail	348	22,14%	1.009	11,39%
Fóruns de discussão/chats	69	4,38%	345	3,89%
Total	1.572	100%	8.861	100%

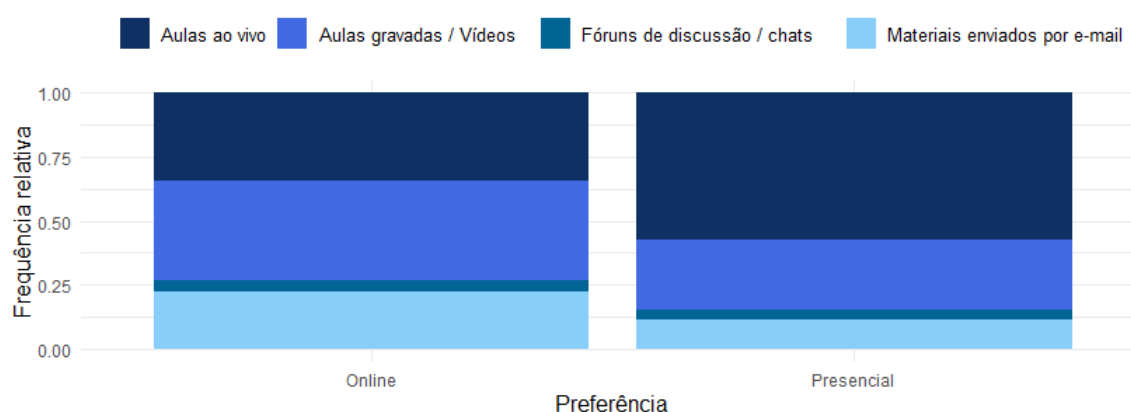
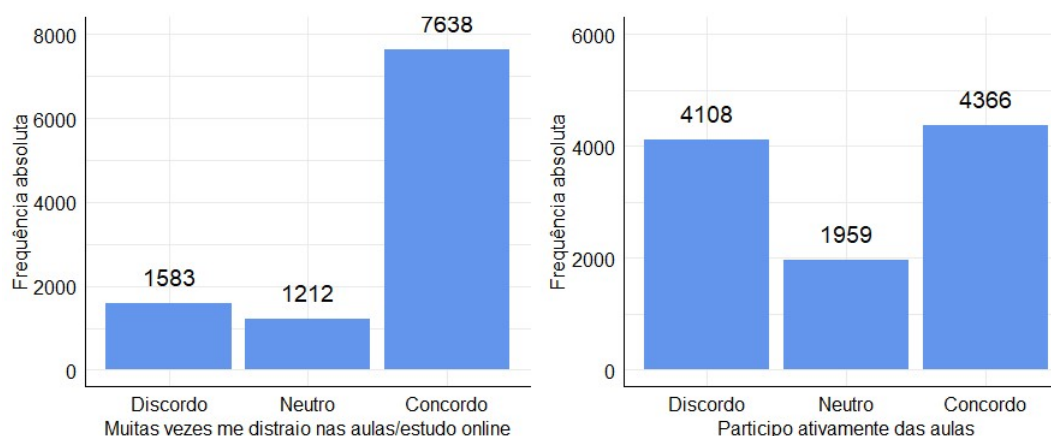


Figura 14: Modo de ensino preferido relacionado a preferência

3.1.10 Avaliação do ensino online

Ao analisar a Figura 15, pode-se observar a frequência em que os alunos concordam, são neutros ou discordam das afirmações referente à avaliação do ensino online. As afirmações "Gosto de ter aula online" e "Me sinto motivado para aprender" e "Me distraio muitas vezes nas aulas online" se destacaram com maior avaliações negativas.



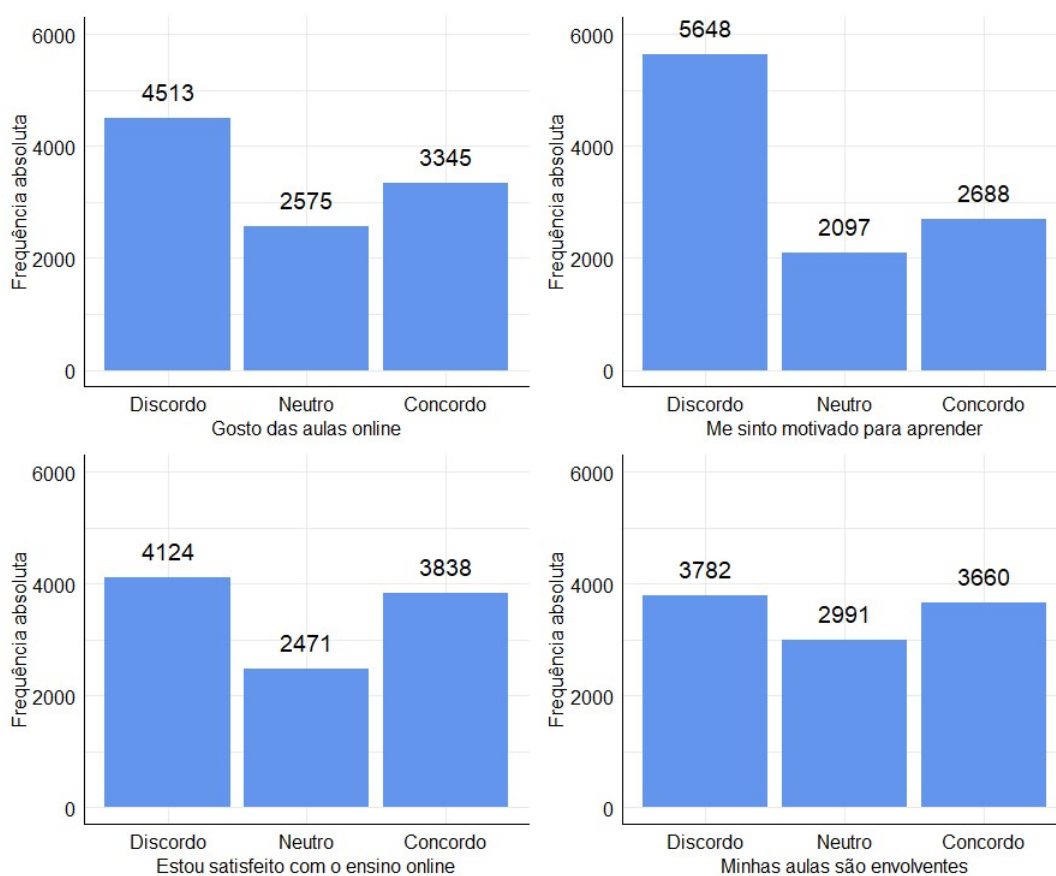


Figura 15: Avaliação do ensino online

3.1.11 Avaliação do ensino online relacionado à preferência dos alunos

A Figura 16 permite observar que os alunos que preferem o ensino online avaliaram de forma positiva essa modalidade com maior frequência do que os alunos que preferem o ensino presencial. Uma exceção ocorre em relação à participação ativa, que não apresentou diferença expressiva entre os dois grupos.

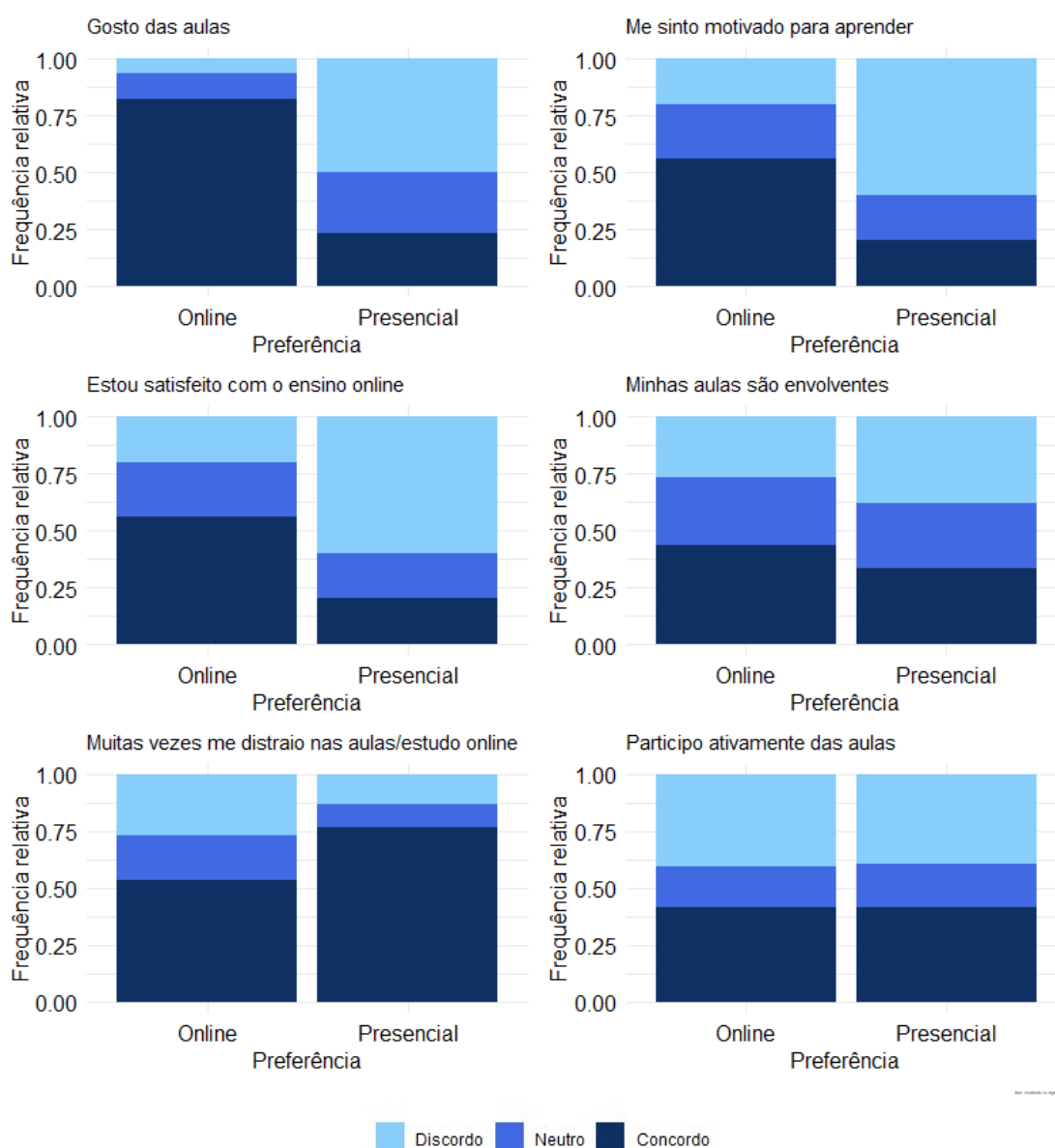


Figura 16: Avaliação do ensino online relacionado à preferência dos alunos

3.1.12 Avaliação do ensino presencial

A Figura 17 mostra que, no geral, o ensino presencial foi bem avaliado pelos respondentes. Apenas a participação ativa e a distração durante a aula apresentaram avaliações mais igualmente distribuídas entre discordo, neutro e concordo.

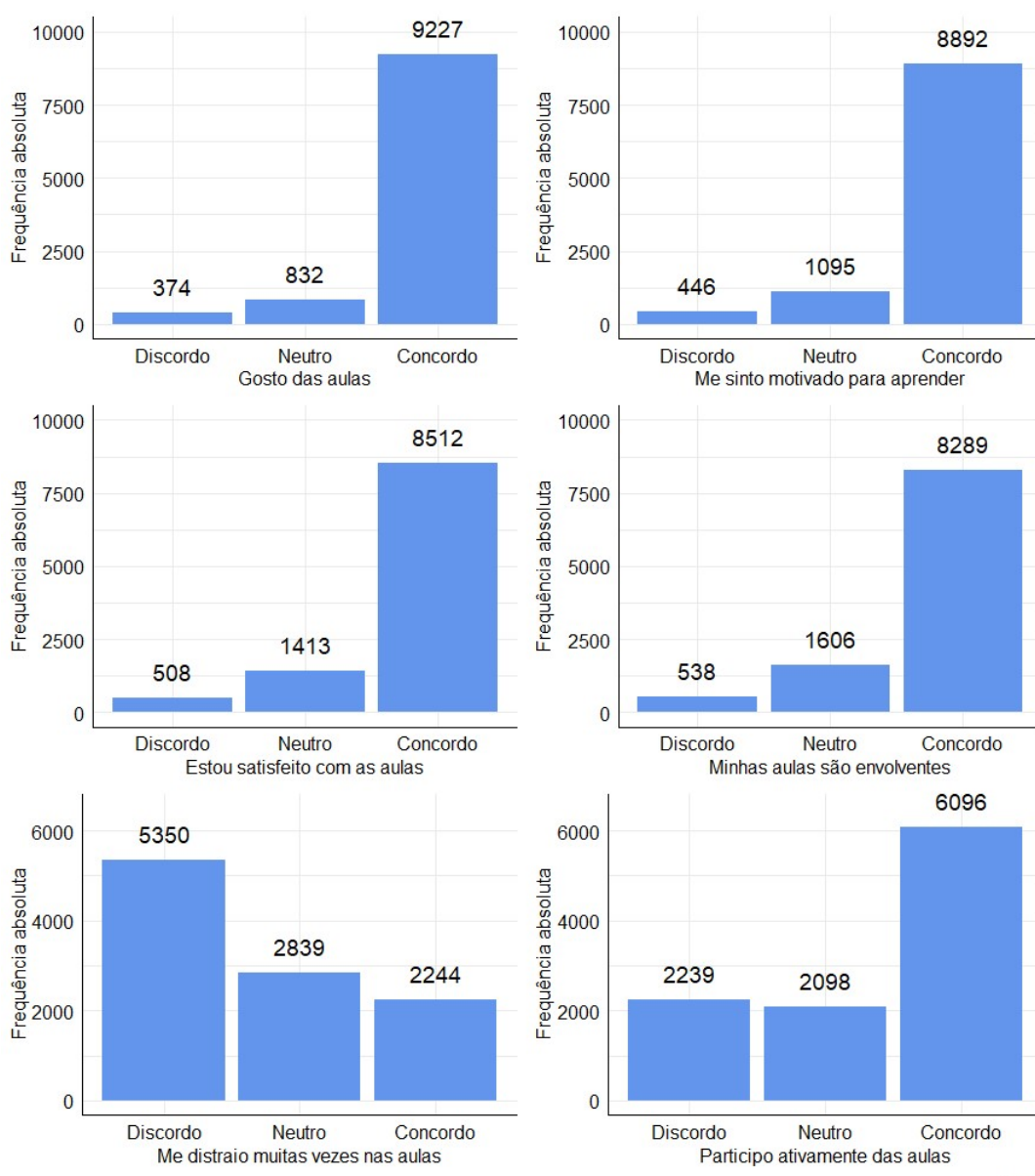


Figura 17: Avaliação do ensino presencial

3.1.13 Avaliação do ensino presencial relacionado à preferência dos alunos

Ao analisar os dados apresentados pela Figura 18, sobre a avaliação do ensino presencial, é possível notar que, proporcionalmente, os alunos que preferem o ensino presencial avaliaram de forma positiva com maior frequência do que os alunos que preferem o ensino online.

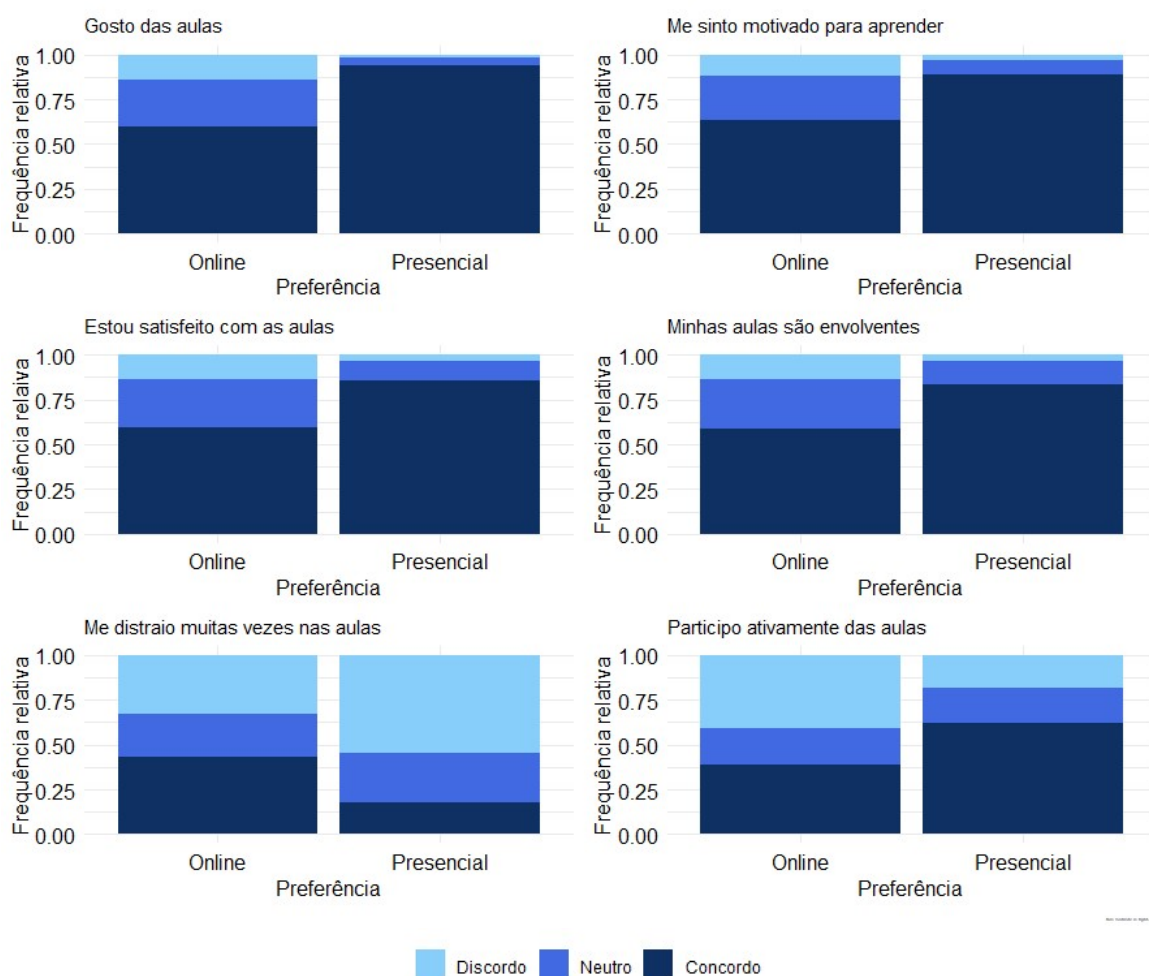


Figura 18: Avaliação do ensino presencial relacionado a preferência dos alunos

3.2 Associações

A fim de avaliar se as variáveis explicativas em estudos estão relacionadas significativamente com a preferência dos estudos pelo ensino online ou presencial, foram realizados testes de associações. Dentre todas as variáveis estudadas, as apresentadas a seguir foram as variáveis que apresentaram um p-valor menor que 5%. Por conta disso, pode-se dizer que há evidências estatísticas de que essas variáveis estão relacionadas com a preferência dos alunos.

Inicialmente, foi questionado aos alunos sobre o nível atual de ensino, país de origem, modo de ensino vivenciado e modo de ensino preferido. Dentre esses questionamentos, apenas o nível de ensino e o modo de ensino preferido apresentaram

evidências de associação significativa com a preferência do ensino online ou presencial.

Tabela 7: Associação com a preferência do ensino

Pergunta	Preferência		Total	X^2	p-valor	
	Online	Presencial				
Escolaridade	Ensino Médio	769	4.204	4.973	18.36	< 0,001
	Graduação	678	4.165			
	Pós Graduação	95	386			
	Outros	30	101			
Modo preferido de aula	Aulas ao vivo	542	5.057	5.599	304,98	< 0,001
	Aulas gravadas / Vídeos	613	2.450			
	Fóruns de discussão / chats	69	345			
	Materiais por e-mail	348	1.009			

Durante a pesquisa, foi solicitados aos alunos que selecionassem as duas plataformas mais utilizadas por seus instrutores. Dentre as mais de 20 plataformas analisadas no estudo, apenas os testes de independência para os grupos que utilizaram as plataformas Zoom, Google Meets, Teams, quadro negro online ou Big Blue Button apresentaram evidências estatísticas para rejeitar a hipótese nula de independência em relação à preferência pelo ensino online ou presencial.

Tabela 8: Associação entre o uso da plataforma e a preferência do ensino

Plataforma		Preferência		Total	X^2	p-valor
		Online	Presencial			
Zoom	Não Usou	802	4770	5.572	4,14	0,040
	Usou	770	4091			
Google Meet	Não usou	1.339	7.764	9.103	6,64	0,008
	Usou	233	1.097			
Teams	Não Usou	1.131	6.075	7.206	7,01	0,008
	Usou	441	2.786			
Quadro negro online	Não Usou	1.432	7.887	9.319	5,88	0,010
	Usou	140	974			
Big Blue Button	Não Usou	1.489	8.500	9.989	4,47	0,034
	Usou	83	361			

Por outro lado, quando questionados sobre os métodos utilizados durante as aulas, 10 métodos diferentes eram possíveis, sendo que os alunos deveriam marcar todas as opções que os mesmos haviam vivenciados durante esse período. Com base na Tabela 9, dentre esses 10 métodos, apenas para 4 há evidências para rejeitar a hipótese nula de independência.

Tabela 9: Associação entre o método utilizado e a preferência do ensino

Método		Preferência		Total	X^2	p-valor
		Online	Presencial			
Apresentação	Não teve	1.519	8.676	10.195	9,30	0,002
	Teve	53	185	238		
Palestra	Não teve	741	4.481	5.222	6,15	0,010
	Teve	831	4.380	5.211		
Ao vivo	Não teve	319	2.094	2.413	8,19	0,004
	Teve	1.253	6.767	8.020		
Avaliação na aula	Não teve	1.508	8.600	10.108	5,23	0,022
	Teve	64	261	325		

Por fim, foi questionado quais técnicas os alunos utilizavam no período fora das aulas. Em meio às 7 técnicas apresentadas no estudo, apenas em 3 houve evidências de associação significativa com a preferência dos alunos pelo ensino presencial ou online, como é possível observar na tabela 10.

Tabela 10: Associação entre a técnica utilizada e a preferência do ensino

Técnicas		Preferência		Total	X^2	p-valor
		Online	Presencial			
Video	Não teve	897	5.324	6.221	4,94	0,026
	Teve	675	3.537	4.212		
Material de estudo	Não teve	1.411	7.748	9.159	6,48	0,01
	Teve	161	1.113	1.274		
Avaliação supervisionada	Não teve	647	4.047	4.694	10,81	0,001
	Teve	925	4.814	5.739		

3.3 Modelo 1

Considerando as variáveis explicativas apresentadas na análise descritiva, foi realizado o ajuste do modelo de regressão logística para a variável resposta referente à preferência dos alunos.

Os modelos foram gerados a partir do método automático de seleção de variáveis stepwise, considerando 5% de significância para entrada e saída das variáveis. Realizando alguns ajustes para adicionar variáveis significativas para o modelo que haviam sido excluídas pelo método stepwise, as variáveis da Tabela 11 foram selecionadas.

Inicialmente foi ajustado um modelo que levava em consideração apenas variáveis referentes à utilização de tecnologias e características individuais. Assim, foram desconsideradas variáveis de avaliação e preferência dos alunos.

Ao realizar o método de seleção automática de variáveis *stepwise* e ajustar seu resultado acrescentando outras variáveis que também foram significativas para o modelo, concluiu-se que o modelo seria composto pelas variáveis nível de estudo, uso do Microsoft teams, uso de apresentação durante as aulas, realização de avaliações durante as aulas e a realização de palestras online.

Tabela 11: Modelo 1 de regressão logística

	Estimativa	Erro Padrão	P-valor	OR	IC
(Intercept)	-1,748	0,064	<0,001	-	-
Q2.Pós Graduação	0,403	0,122	<0,001	1,496	[1,172;1,893]
Q2.Graduação	-0,175	0,058	0,003	0,839	[0,749;0,940]
Q10.Microsoft Teams	-0,200	0,062	0,001	0,818	[0,724;0,923]
Q11.Apresentação	0,639	0,162	<0,001	1,895	[1,368;2,584]
Q11.Avaliação na aula	0,443	0,147	0,002	1,556	[1,159;2,061]
Q10.Palestra	0,220	0,058	<0,001	1,246	[1,112;1,395]

A partir das razões de chance apresentadas, a chance do aluno preferir o ensino online foi 49,6% maior no grupo de alunos que estavam na pós graduação, enquanto os alunos da graduação tiveram 16,1% de redução na chance de preferir o ensino online em relação aos alunos do ensino médio e outros níveis educacionais. Vale destacar que a percentual de alunos matriculados em curso de pós da graduação foi pequeno na amostra.

A utilização Microsoft Teams, por outro lado reduziu 18,2% a chance dos alunos preferirem o ensino online em relação aos alunos que não utilizaram essa plataforma. Já os alunos que tiveram apresentações, palestras e avaliações na aula apresentaram uma chance maior de preferir o ensino online.

3.3.1 Diagnóstico do modelo

Para avaliar a qualidade do modelo, foi realizado o teste de Hosmer-Lemeshow. Com base no resultado do teste, pode-se concluir que não há evidências estatísticas suficiente para se rejeitar a hipótese nula de que o modelo está bem ajustado, uma vez que o p-valor foi igual a 0,95. Isso significa que não há evidência de diferença significativa entre os valores observados e esperados no modelo, uma vez que o p-valor do teste foi superior a 5%.

Para aprofundar a investigação sobre a qualidade do modelo, foi realizado também o teste de razão de verossimilhanças e teste de Wald. Ambos os teste resultaram p-

valo menor que 0,001. Isso indica que, a um nível de significância de 5%, hipótese nula foi rejeitada para todos os testes.

A Tabela 12 indica a classificação para a modelo ajustado. Pode-se notar que a proporção de classificações corretas é 51,15%, sendo essa uma proporção baixa uma vez que se aproxima de 50%.

Tabela 12: Tabela de classificação para o modelo 1

	Presencial	Online	Total
Presencial predito	4.449	685	5.134
Online predito	4.412	887	5.299
Total	8.861	1.572	10.433

Com base na Figura 19 da curva ROC para esse primeiro modelo, é evidente seu baixo poder de discriminação. Para corroborar, a área abaixo da curva no modelo foi igual a 55,22%. Como esse valor é muito próximo de 50% a discriminação do modelo fica próximo da ausência de discriminação.

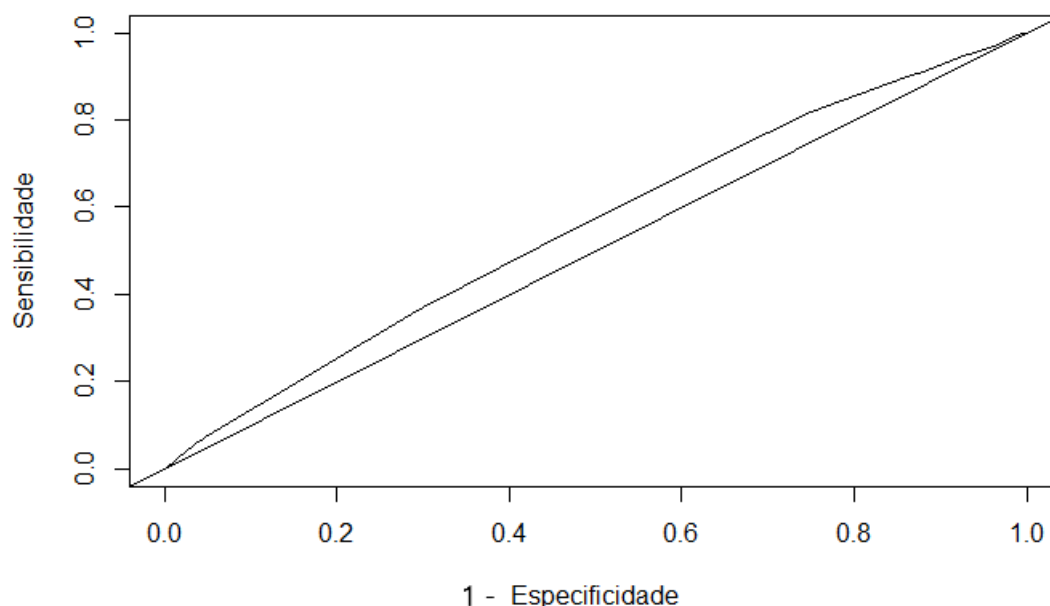


Figura 19: Curva ROC para o modelo 1

O gráfico do resíduo deviance versus o resíduo de Pearson apresentado na Figura 20 para o modelo 1 mostra que os resíduos ficaram mais concentrados entre -0,5 e 0, indicando que o modelo não está bem ajustado.

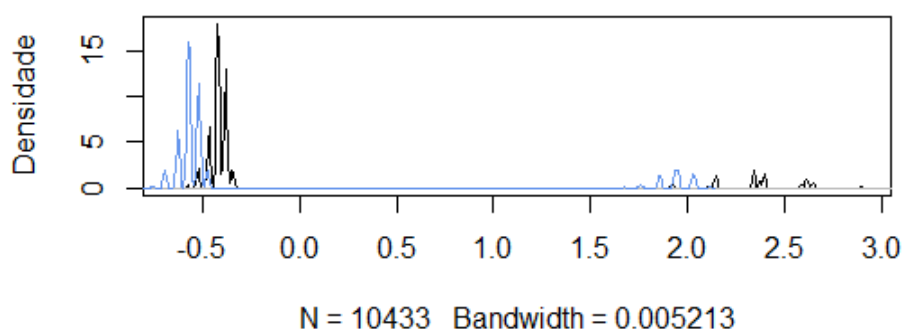


Figura 20: Gráfico Deviance (azul) x Resíduos de Pearson para o modelo 2

3.4 Modelo 2

Buscando um melhor modelo para o caso em estudo, visto que apenas a inclusão de variáveis referentes à utilização de tecnologias e características individuais não foi suficiente, será apresentado um modelo alternativo. Para esse modelo, será considerado a variável modo de ensino online preferido pelos alunos.

Tabela 13: Modelo 2 de regressão logística

	Estimativa	Erro Padrão	P-valor	OR	IC
(Intercept)	-2.580	0.066	< 0,001	0.076	[0.066;0.086]
Q2.Pós graduação	0.410	0.121	< 0,001	1.507	[1.183;1.903]
Q11.Apresentação	0.728	0.166	< 0,001	2.070	[1.483;2.848]
Q10.Zoom	0.149	0.057	< 0,001	1.161	[1.038;1.299]
Q10.Google Meet	0.214	0.081	< 0,001	1.239	[1.055;1.451]
Q11.Avaliação na aula	0.485	0.150	< 0,001	1.624	[1.203;2.165]
Q11.Palestra	0.305	0.060	< 0,001	1.357	[1.207;1.526]
Q8.Aulas gravadas	0.896	0.064	< 0,001	2.450	[2.160;2.781]
Q8.Fóruns de discussão	0.670	0.140	< 0,001	1.955	[1.475;2.559]
Q8.Materiais por e-mail	1.255	0.078	< 0,001	3.507	[3.006;4.088]

Tendo com base a Tabela 13, percebe-se que a utilização da plataforma Microsoft teams não foi mais significativo para o modelo, assim como os alunos preferirem as aulas ao vivo na questão 8. Analisando as razões de chance apresentadas é possível notar que todas as variáveis que foram significativas para entrar no modelo aumentaram a chance dos alunos preferirem o ensino online.

A utilização da plataforma Zoom foi a que menos aumentou as chances, com um aumento de apenas 16,1%. Por outro lado, o aluno preferir o envio de materiais por e-mail foi o que mais aumentou a chance da preferência pelo ensino online, O aluno estar na pós graduação também aumentou 50,7% da chance do aluno preferir o ensino online em relação aos alunos estavam no ensino médio ou na graduação.

3.4.1 Diagnóstico do modelo

Assim com no modelo 1, foi realizado o teste de Hosmer-Lemeshow. Como o p-valor para esse teste foi igual a 0,11, também pode-se concluir que, a um nível de significância de 5%, não há evidências estatísticas suficiente para se rejeitar a hipótese nula de que o modelo está bem ajustado.

Ao realizar os teste de razão de verossimilhança e o teste de Wald, também conclui-se o mesmo. Com p-valores inferiores a 0,001, há evidências estatística para se rejeitar a hipótese nula a um nível de significância de 5%. Com isso, tem-se a indicação de que o modelo não restrito se ajusta melhor aos dados.

A Tabela 14 a apresenta a classificação para o modelo 2. É possível notar que a proporção de classificações corretas é 59,18%, resultado superior ao 55,22% de classificação correta no modelo 1.

Tabela 14: Tabela de classificação para o modelo 2

	Presencial	Online	Total
Presencial predito	5.922	1.319	7.241
Online predito	2.939	253	3.192
Total	8.861	1.572	10.433

A Figura 21 representa a curva ROC para o modelo proposto, sendo a área abaixo da curva no modelo igual a 65,10%. Sabe-se que, quanto mais a curva ROC se aproxima do canto superior esquerdo do gráfico, maior o poder de discriminação do modelo.

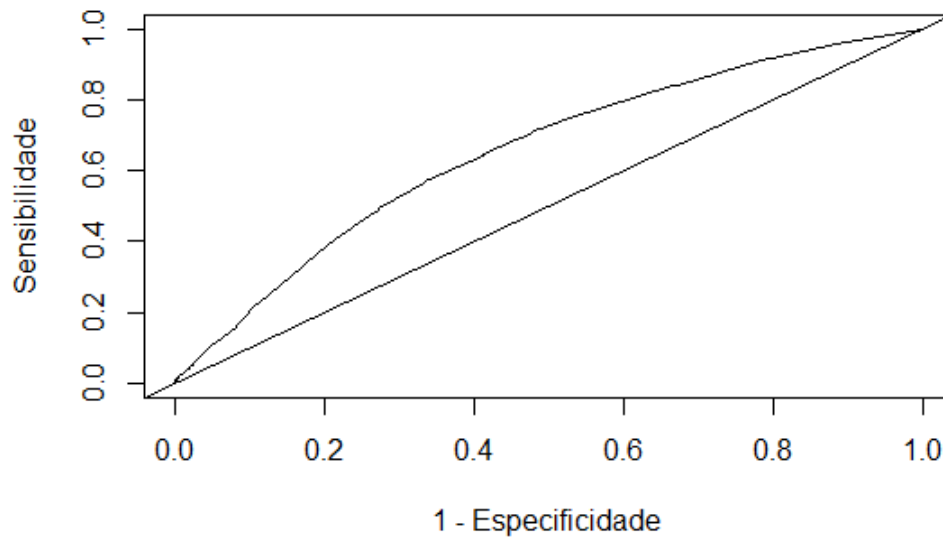


Figura 21: Curva ROC para o modelo 2

O gráfico do resíduo deviance versus o resíduo de Pearson apresentado na Figura 22 para o modelo 2 mostra que os resíduos ficaram mais concentrados entre -1 e 1, indicando um ajuste melhor do que o modelo anterior, apesar de ainda não ser considerado o ideal.

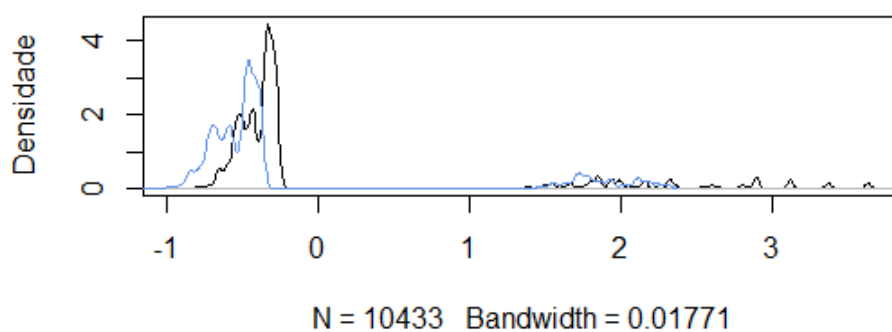


Figura 22: Gráfico Deviance (azul) x Resíduos de Pearson para o modelo 2

3.5 Medidas de qualidade do ajuste

Por fim, somando-se a todas as informações apresentadas sobre os dois modelos construídos tem-se as medidas de qualidade do modelo apresentadas no Quadro 10.

Quadro 4: Medidas de qualidade

	Deviance	AIC	BIC
Modelo 1	8.789,46	8.805,47	8.863,48
Modelo 2	8.491.89	8.471.89	8.564.42

Com base na comparação apresentada e em todos os outros resultados apresentados anteriormente, o modelo 2 parece o melhor ajustado, uma vez que esse modelo foi o que apresentou menores valores para Deviance, AIC e BIC, sendo assim o modelo escolhido.

4 Conclusão

Sendo o objetivo de entender quais variáveis influenciam a preferência dos alunos pelo ensino online, pode-se notar que as variáveis referentes a realidade vivida pelos alunos durante esse período não foram suficientes para garantir uma boa classificação do modelo. Apesar disso, foi possível observar que a utilização Microsoft Teams e o fato dos alunos estarem na graduação reduziu a chance de preferir o ensino online. Em contrapartida, o fato dos alunos estarem em uma pós graduação, o professor trazer uma apresentação pronta para as aulas, a realização de palestra e avaliações na aula, aumentaram a chance dos alunos optarem pelo ensino online.

Ao incluir a variável referentes ao modo de ensino online preferido pelos alunos, foi possível observar que esse modelo apresentou uma discriminação melhor do que o apresentado no modelo 1. Nesse modelo, o aluno estar na pós graduação, a utilização do Zoom e do Google Meets além do uso da apresentação, avaliação na aula e palestras aumentaram a chance dos alunos preferirem o ensino online.

Vale destacar que, em outras abordagens, seria possível considerar o modo de ensino online preferido assim como as variáveis de avaliação do ensino online e presencial resumidas nas componentes principais, como variável resposta. Isso pode gerar uma discussão sobre a entrada dessas variáveis em modelo que tem como resposta a preferência dos alunos pelo ensino online ou presencial.

Outro ponto importante do estudo apresentado se refere à data em que os dados foram coletados. Por ser uma pesquisa realizada em maio de 2020, fase inicial da pandemia, onde os alunos haviam vivenciado o ensino online por um tempo reduzido, é possível que novas pesquisas apontem para resultados diferentes. Além disso, por ser uma medida emergencial, a transição do ensino presencial para o online, em um primeiro momento, não ocorreu de uma forma estruturada o que também pode ter afetado a avaliação dos alunos.

O país de origem dos alunos foi desconsiderado dos modelos devido à presença de 115 países distintos, além da grande predominância de alunos do Reino Unido na pesquisa. No entanto, essa variável pode ser significativa para o modelo em outros estudos. Se tratando de um período de pandemia seria importante considerar o progresso da mesma em cada país no momento da pesquisa.

O debate sobre ensino online e presencial está apenas começando. No presente

estudo, informações demográficas com sexo, idade, renda familiar, além do acesso à internet computadores e um ambiente de estudo favorável não foram levantadas. Para futuros projetos pode ser interessante considerar essas variáveis a fim de entender melhor o contexto ao qual esses alunos estão inseridos. Outro ponto que pode ser interessante de se considerar é a opção do ensino híbrido, que vem sendo cada vez mais discutidos ao longo desses período de pandemia.

5 Referências

AGRESTI, Alan. An Introduction to Categorical Data Analysis. Second Edition. Hoboken, New Jersey John Wiley Sons, 2007.

BDIWI, R., et al., (2019) 'Smart learning environment teacher's role in assessing classroom attention', *Research in Learning Technology*, vol. 27, p. 2072.

DISTRITO, Edtech report 2020. Acesso em 17 mar. 2021.

Genn 2001; Shirazi et al;2014 'Comparison of health students' perceptions and expectations of their educational environment', *British Journal of Medicine Medical Research*, vol. 4, no. 35, pp. 5657–66.

HOSMER, D.W.; LEMESHOW S. Applied Logistic Regression. Second Edition, John Wiley Sons, Inc., Hoboken, NJ, USA, 2000.

KIM, K., Bonk, C., The Future of Online Teaching and Learning in Higher Education The Survey Says... *Educause Quarterly*, 29, 22-30, 2006.

MOUSAVI, A., Mohammadi, A., Mojtahedzadeh, R., Shirazi, M., Rashidi, H. (2020). E-Learning Educational Atmosphere Measure (EEAM) A New Instrument for Assessing E-Students' Perception of Educational Environment. *Research in Learning Technology*, 28, 17–35.

RICHARD,Arnold Johnson, Dean W. Wichern (2002) "Applied multivariate statistical analysis", 5th Edition (Prentice Hall).

STEIN, Geneva, (2020), "Student perceptions of remote learning", Disponível em <https://doi.org/10.7910/DVN/2TGOPH>, Harvard Dataverse, V1. Acesso em 23 abril de 2021.

TJ MCCUE, (2021) E Learning Climbing To \$325 Billion By 2025 UF Canvas Absorb Schoology Moodle, 2018. Disponível em <https://www.forbes.com/sites/tjmccue/2018/07/31/e-learning-climbing-to-325-billion-by-2025-uf-canvas-absorb-schoology-moodle/?sh=756792583b39>.. Acesso em 04 maio. 2021.

TODOS PELA EDUCAÇÃO, (2020) Ensino a distância na Educação Básica frente à pandemia da Covid-19. Nota Técnica - Abril 2020.